数理統計学序説[a]

大阪公立大学 大学院 理学研究科 数学専攻

今野　良彦

2026 年 9 月 19 日

[a] 立教大学 2026 年度数理統計学序論 1
立教大学 大学院　2026 年度数理統計学 1
用講義録

加筆予定事項

- 統計学の歴史を Samworth and Shah (2026) を参考にして加筆.
- Banach 空間の定義と基本的な事項を補遺に書き込むこと.
- 最尤推定量の漸近分布 (母数空間が多次元の場合) の証明 (定理 12.11) の加筆
  - 指数型分布族での最尤推定量の漸近分布の証明は
- 関連図書の形式を成形すること
- 12 章を書くこと.
- 節 9.5: 統計的理論からみた Bayes 法のところを加筆.
- 節 10.2 のところを全面的に書き直すこと.
- 節 10.4 のところを Keener を読みながら書き直すこと.

目 次

はじめに **ix**

第 I 部　宴の支度編: 確率・確率変数・確率分布・標本分布 **1**

第 0 章　統計学歴史と現代の数理統計学 **3**

第 1 章　確率・確率変数の基本事項 **5**
1.1 基礎的な確率規則 5
1.2 確率変数 13
1.3 分位点関数 21
1.4 主な 1 次元分布 25
1.4.1 離散型確率変数 25
1.4.2 連続型確率変数 27
1.5 2 次元の分布 28
1.5.1 同時確率関数と確率密度関数 28
1.5.2 周辺分布 29
1.5.3 独立な分布と条件付き分布 30
1.6 多次元分布と i.i.d. 標本 33
1.6.1 重要な多次元分布モデル 34
1.7 正規分布に関連した分布 40
1.8 ガンマ分布 40
1.9 χ^2 分布 43
1.10 F 分布 44
1.11 t 分布 46
1.12 章末注釈と参考文献 47
1.13 演習問題 47

第 2 章　期待値の基礎事項 **53**
2.1 期待値 53
2.2 確率ベクトルの期待値 62
2.3 分散と共分散 64
2.4 条件付き期待値 70

2.5 積率母関数 75
2.6 章末注釈と参考文献 79
2.7 演習問題 79

第 3 章 確率と期待値の不等式 87
3.1 確率に対する不等式 87
3.2 期待値に対する不等式 92
3.3 章末注釈と参考文献 96
3.4 演習問題 96

第 4 章 確率変数列と分布列の収束 99
4.1 確率変数列の収束のタイプ 100
4.2 大数の法則 110
4.3 中心極限定理 114
4.4 デルタ法 115
4.5 延期された証明 118
4.5.1 定理 4.17 の証明 118
4.5.2 補題 4.18 の証明 121
4.5.3 定理 4.9 の証明 122
4.5.4 定理 4.12 の証明 123
4.5.5 定理 4.14 の証明 126
4.5.6 定理 4.25 の証明 126
4.6 章末注釈と参考文献 128
4.7 演習問題 128

第 II 部 宴編: 統計的推測 131

第 5 章 統計的推測論の枠組み 133
5.1 統計的実験と母数モデル 133
5.2 統計的決定問題 138
5.3 章末注釈と参考文献 142
5.4 演習問題 143

第 6 章 母数モデル 145
6.1 正則母数モデルと Fisher 情報量 145
6.2 指数型分布族 150
6.2.1 1 変数の場合 150
6.3 指数型分布族の正準表示 153
6.4 十分統計量 157
6.5 統計量の最小十分性と完備性 160

6.5.1 十分統計量の最小性 160
6.5.2 十分統計量の完備性 164
6.6 補遺: Laplace 変換の一意性 166
6.7 章末注釈と参考文献 167
6.8 演習問題 167

第 7 章 推定 **169**
7.1 モーメント法 169
7.2 最尤法 171
7.3 不偏推定と情報不等式 176
7.4 章末注釈と参考文献 188
7.5 演習問題 188

第 8 章 検定と信頼区間 **193**
8.1 仮説検定の考え方 193
8.2 Neyman-Pearson の定理 196
8.3 検定統計量の導出方法 201
8.3.1 尤度比検定統計量 201
8.3.2 スコア検定 202
8.3.3 Wald 検定 204
8.4 検定の一様最適性 204
8.4.1 単調尤度比 204
8.4.2 不偏検定 209
8.5 多重比較と FDR 212
8.6 区間推定の考え方 212
8.7 信頼区間の構成法 213
8.7.1 検定方式の反転 213
8.7.2 枢軸量 (pivotal quantity) 214
8.8 章末注釈と参考文献 215
8.9 演習問題 215

第 9 章 Bayes 的推測 **219**
9.1 Bayes 的推測の考え方 219
9.2 Bayes 的推測手法 220
9.3 事前分布の選択について 224
9.3.1 一様事前分布 225
9.3.2 Jeffreys 事前分布 226
9.3.3 共役事前分布 227
9.4 1 次元の van Trees 不等式 229
9.5 統計的決定理論からみた Bayes 法 235

9.6 章末注釈と参考文献 235
9.7 演習問題 235

第 10 章 計算機集約的手法 237
10.1 乱数の発生法 237
10.1.1 ベルヌーイ分布 238
10.1.2 2 項分布 238
10.1.3 標準正規分布 238
10.1.4 正規分布 239
10.1.5 確率の積分変換 239
10.1.6 受容・棄却法 240
10.2 Markov 連鎖 Monte Carlo 法 (MCMC 法) 242
10.2.1 基本的な考え方 243
10.2.2 推移核の Metropolis–Hastings による構成 245
10.2.3 Gibbs サンプリング法 251
10.3 ブートストラップ法 251
10.4 EM アルゴリズム 253

第 11 章 モデル選択 261
11.1 モデル選択の目標 261
11.2 罰則法: AIC 261
11.3 Bayes モデル選択 261
11.4 交差検証法 261
11.5 章末注釈と参考文献 261
11.6 演習問題 261

第 12 章 大標本理論 263
12.1 ランダム関数の確率収束 263
12.2 最尤推定量の一致性 271
12.3 最尤推定量の漸近正規性 278
12.4 尤度比検定統計量の漸近分布 282
12.5 章末注釈と参考文献 285
12.6 演習問題 285

第 III 部 宴の始末編: 付録 287

第 A 章 補遺: 集合・写像・位相の復習 289
A.1 集合論の言葉使い 289
A.1.1 数の集合 289
A.1.2 集合の記号 289

A.2 写像 291
A.2.1 写像の定義と性質 291
A.2.2 像と逆像 292
A.2.3 合成写像 293
A.3 可算集合・非可算集合・濃度 294
A.4 距離と位相 294
A.4.1 距離空間 294
A.4.2 位相空間 296
A.5 $\mathbb{R}$ の拡張 297
A.6 章末注釈と参考文献 298

第 B 章 補遺 **299**
B.1 大数の強法則の証明 299
B.1.1 証明のための準備の補題 299
B.1.2 大数の強法則 (定理 4.19) の証明 302
B.2 中心極限定理の証明 304
B.3 Dynkin 族の議論 308
B.4 Dynkin の定理の応用例 314
B.4.1 その 1 314
B.4.2 その 2 316
B.4.3 $\mathbb{R}$ の位相についての注意 321
B.5 絶対連続関数 323
B.6 凸関数の性質 327
B.7 Radon-Nikodym の定理と条件付き期待値の議論 333
B.8 条件付き期待値の性質 337
B.9 Neyman-Fisher の因子分解定理の証明 340
B.10 Dini の定理 347

はじめに

この講義録は, 数理統計学の基本的な事項についてまとめたものである. 第 I 部では, 統計推測理論に必要な確率・確率変数・期待値等の概念と性質を丁寧に説明した. 性質についてはできるだけ証明を与えておいた. ただし, 証明に測度論の知識が必要な個所については, 補遺を参照のこと. 第 II 部では, 統計的推測理論の基本的な事項について説明をした. 第 III 部は補遺である.

2026 年 9 月 19 日更新
2025 年 3 月 22 日更新
2024 年 1 月 22 日更新

記号について

ここに掲げる規則は Jordan Stoyanov (2023, IMS Bulletin **52**, Issue 4, 25-26) の借用である．また, 記号の字体等は Notation list-Cambridge International Mathematics Qualifications[1] を参考にした．

3 つの基本規則

- 規則 1: ひとつの項目や対象に対してひとつの文字か記号を使用すること．
- 規則 2: 対象の同じグループに対しては同じ字体を使用すること．
- 規則 3: 対象の異なるグループに対しては異なる字体を使用すること．

派生規則

- 確率, 確率分布, 期待値, 分散, 共分散には `\mathsf` を使用する．すなわち, Pr, P, E, Var, Cov のように記す．
- 重要な空間には `\mathbb` を使用する．
- 集合族には `\mathcal` を使用する．
- 略称/略語は「p.m.f.」等のように書く．
- 分布名は `\mathsf` を使用する．たとえば, $\mathsf{Poi}(\lambda)$ で母数 $\lambda > 0$ の Poisson 分布を表す．
- 母数と母数空間等の未知のものは Greek letters を使用する．
- ベクトルと行列は小文字と大文字の `\mathbf` を使用する．
- 内部等の数学の概念は `\mathrm` を使用する．たとえば, 母数空間 Θ の内部を $\mathrm{int}\,\Theta$ (`\mathrm{int}\,\mathsf{\Theta}`) と書く．
- Θ を (`\Textsf{\THeta}`) とも書く．

[1] https://www.cambridgeinternational.org/Images/420009-mathematics-notation-list-.pdf

記号一覧

記号	説明
$\mathbb{N}$	自然数全体の集合
$\mathbb{Z}$	整数全体の集合
$\mathbb{Q}$	有理数全体の集合
$\mathbb{R}$	実数全体の集合
$\mathbb{R}^n$	有限 n 個の実数空間の直積集合
$\mathbb{C}$	複素全体の集合
$\mathbb{R}^n$	有限 n 個の実数空間の直積集合
$\mathrm{Mat}(n, m; \mathbb{R})$	n 行 m 列の実行列全体のなす集合. $\mathrm{Mat}(n, n; \mathbb{R})$ を $\mathrm{Mat}(n; \mathbb{R})$ と書く.
$\mathrm{Sym}(n, \mathbb{R})$	n 行 n 列の対称行列全体のなす集合.
$\mathrm{Sym}^+(n, \mathbb{R})$	n 行 n 列の半正値対称行列全体のなす集合.
$\mathrm{Sym}^{++}(n, \mathbb{R})$	n 行 n 列の正値対称行列全体のなす集合.
$\mathrm{Herm}(n, \mathbb{C})$	n 行 n 列のエルミート行列全体のなす集合.
$\mathrm{Herm}^+(n, \mathbb{C})$	n 行 n 列の半正定値エルミート行列全体のなす集合.
$\mathrm{Herm}^{++}(n, \mathbb{C})$	n 行 n 列の正定値エルミート行列全体のなす集合.
$\mathbf{I}_n$	n 次の単位行列
$\mathbf{0}_n$	$\mathbb{R}^n$ の零ベクトル
$\mathbf{A}^\top$	m 行 n 列の行列 $\mathbf{A}$ の転置行列.
$\mathbf{A}^{-1}$	正方行列 $\mathbf{A}$ の逆行列
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	σ 集合族.

第I部

宴の支度編: 確率・確率変数・確率分布・標本分布

第0章　統計学歴史と現代の数理統計学

第1章　確率・確率変数の基本事項

この章では, 統計的推測理論を理解する上で必要な確率論の基礎事項を簡潔に述べる. したがって, 証明はしない場合がある. また, 測度論の内容には立ち入らない. 数学的な厳密性を犠牲にして, 直観的な理解を目指す記述を心掛けた. 節 1.1 では, 確率に関わる基本的な概念を導入する. 節 1.2 では, 抽象的な確率空間と数量の世界を結ぶ確率変数を導入し, 確率変数・確率分布・累積分布関数の基本的な性質を説明する. 節 1.3 では, 累積分布関数の逆関数のようなものである分位点関数を導入する. 節 1.4 では, 1 変数確率変数の確率モデルを説明する. 節 1.5 では, 2 次元確率変数の同時分布, 周辺分布, 条件付き分布を説明する. 節 1.6 では, 多変数確率変数の確率モデルを導入する. 節 1.7 では, 正規分布から誘導される重要な分布であるガンマ分布, χ^2 分布, t 分布, F 分布を導入する.

1.1　基礎的な確率規則

確率論はランダムな現象を扱う数学である. 確率論で扱う行為を **試行** という. 試行のあり得る結果すべてを集めた集合を **標本空間** といい, Ω と記すことにする. Ω の部分集合[1] を **事象** という. 事象には標本空間 Ω と **空事象** $\varnothing$(何も起こらないという事象) も含めることにする. 事象を集めた部分集合族を $\mathcal{A}$ と記す. この部分集合族 $\mathcal{A}$ は以下で述べる **σ 加法性**(完全加法性) をみたすことにする.

定義 1.1. Ω を空でない集合とし, $\mathcal{A}$ を Ω の部分集合族とする. $\mathcal{A}$ が次の 3 条件をみたすとき, **σ 加法族** と呼ばれる.

(1) $\Omega \in \mathcal{A}$.

(2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

(3) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

[1] Ω が可算集合ならば, σ 加法族をすべての部分集合としてよいが, Ω が連続濃度のときには, すべての部分集合とすると不都合が生じることになる. これについては [2, p.21] を参照のこと.

ただし $A^c = \{\omega \in \Omega;\, \omega \notin A\}$ である. Ω と $\mathcal{A}$ の組 $(\Omega, \mathcal{A})$ を 可測空間 と呼ぶ

補題 1.2. $(\Omega, \mathcal{A})$ を可測空間とする. このとき以下が成立する.

(1) $\varnothing \in \mathcal{A}$.

(2) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots) \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Proof. (1) $\varnothing = \Omega^c \in \mathcal{A}$.
(2) $A_n^c \in \mathcal{A}\,(n = 1, 2, \ldots)$ と De Morgan の法則から

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} (A_n^c)^c = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{A}.$$

最後は定義 1.1(2)(3) を用いた. □

注意 1.3. $\mathcal{C}$ を Ω のある部分集合族とする. 部分集合族 $\mathcal{C}$ は σ 加法性をみたしてなくともよい. このとき, 集合族 $\sigma[\mathcal{C}]$ を

$$\sigma[\mathcal{C}] := \bigcap \{ \mathcal{A};\, \mathcal{A} \supset \mathcal{C},\, \mathcal{A} \text{ は } \sigma \text{ 加法族} \}$$

で定める. すると $\sigma[\mathcal{C}]$ は σ 加法族となることを確かめることができる. さらに $\mathcal{G}$ を $\mathcal{C}$ を含む σ 加法族としたとき

$$\sigma[\mathcal{C}] \subseteq \mathcal{G}$$

となることが直ちにわかる. すなわち $\sigma[\mathcal{C}]$ は $\mathcal{C}$ を含む最小の σ 加法族 (包含関係の意味で最小) となる. □

問 1.1. 集合族 $\mathcal{C}$ に対して

$$\sigma[\mathcal{C}] = \bigcap \{ \mathcal{A};\, \mathcal{A} \supset \mathcal{C},\, \mathcal{A} \text{ は } \sigma \text{ 加法族} \}$$

と定める. このとき, 以下を証明せよ.
(1) $\sigma[\mathcal{C}]$ は σ 加法族になることを示せ.
(2) $\mathcal{G}$ を $\mathcal{C}$ を含む任意の σ 加法族としたとき

$$\sigma[\mathcal{C}] \subset \mathcal{G}$$

を示せ.

定義 1.4. $\Omega = \mathbb{R}$ とし

$$\mathcal{O} = \{ O \subset \mathbb{R};\, O \text{ は } \mathbb{R} \text{ の開集合} \}$$

とする. $\sigma[\mathcal{O}]$ を $\mathbb{R}$ の **Borel 集合族** と呼び, $\mathcal{B}(\mathbb{R})$ と記す. また

$$\mathcal{C} = \{(-\infty, x) \subset \mathbb{R};\, x \in \mathbb{R}\}$$

とする. このとき $\mathcal{C}$ は $\mathcal{O}$ の真部分集合であるが, $\sigma[\mathcal{C}] = \sigma[\mathcal{O}]$ となる [2].

注意 1.5. Ω が高々可算集合のときは, $\mathcal{A} = 2^{\Omega}$ と取る. ただし 2^{Ω} は Ω のすべての部分集合からなる集合族で冪(べき)集合という. □

定義 1.6. $(\Omega, \mathcal{A})$ を可測空間とする. $\mathcal{A}$ 上の関数

$$\mathsf{Pr} : \mathcal{A} \ni A \mapsto \mathsf{Pr}(A) \in [0, 1]$$

が次の 2 条件をみたすとき, $(\Omega, \mathcal{A})$ 上の **確率**[3] と呼ばれる.

(1) $\mathsf{Pr}(\Omega) = 1$ である.

(2) 互いに排反[4] な事象列 $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(A_n) := \lim_{N\to\infty} \sum_{n=1}^{N} \mathsf{Pr}(A_n)$$

をみたす.

これらの 3 つの組 $(\Omega, \mathcal{A}, \mathsf{Pr})$ を **確率空間** という.

補題 1.7. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とする. このとき以下が成立する.

(1) $\mathsf{Pr}(\varnothing) = 0$ である.

(2) $A \in \mathcal{A}$ に対して, $\mathsf{Pr}(A^c) = 1 - \mathsf{Pr}(A)$ となる.

(3) $N \in \mathbb{N}$ とする. $\{A_n\}_{n=1}^{N} \subset \mathcal{A}$ が互いに排反ならば

$$\mathsf{Pr}\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \mathsf{Pr}(A_n)$$

となる.

(4) $A, B \in \mathcal{A}$, $A \subset B \Rightarrow \mathsf{Pr}(B \setminus A) = \mathsf{Pr}(B) - \mathsf{Pr}(A)$ となる. よって, $A \subset B \Rightarrow \mathsf{Pr}(A) \leq \mathsf{Pr}(B)$ が成立する.

---

[2] $\sigma[\mathcal{C}] \subset \sigma[\mathcal{O}]$ は明らかであるが, 逆の包含関係も示すことができる. この逆の証明は, 定理 1.16 と Euclid 位相の事実を用いて証明ができる.

[3] 簡単に Ω 上の確率ともいう.

[4] $m \neq n$ ならば, $A_m \cap A_n = \varnothing$ が成立していること.

(5) $A_n \in \mathcal{A}$ が $A_n \subset A_{n+1}\,(n = 1, 2, \ldots)$ をみたすならば

$$\lim_{n\to\infty} \mathsf{Pr}(A_n) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right)$$

となる.

(6) $A_n \in \mathcal{A}$ が $A_n \supset A_{n+1}\,(n = 1, 2, \ldots)$ をみたすならば

$$\lim_{n\to\infty} \mathsf{Pr}(A_n) = \mathsf{Pr}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

となる.

(7) (Boole の定理/ユニオン・バウンド) $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathsf{Pr}(A_n)$$

となる.

Proof. (1) $F_1 := \Omega$, $F_n := \varnothing\,(n \geq 2)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となる. 定義 1.6(2) を用いると

$$\mathsf{Pr}(\Omega) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(\Omega) + \sum_{n=2}^{\infty} \mathsf{Pr}(\varnothing)$$

を得る. よって

$$\mathsf{Pr}(\varnothing) = 0$$

がわかる.
(2) $F_1 := A$, $F_2 := A^c$, $F_n = \varnothing\,(n \geq 3)$ とおく. すると $\bigcup_{n=1}^{\infty} F_n = \Omega$ かつ $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となるので, 定義 1.6(1), (2) を用いると

$$\begin{aligned} 1 = \mathsf{Pr}(\Omega) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(A) + \mathsf{Pr}(A^c) + \sum_{n=3}^{\infty} \mathsf{Pr}(\varnothing) \\ &= \mathsf{Pr}(A) + \mathsf{Pr}(A^c) \qquad (\because (1)) \end{aligned}$$

がわかる. よって, (2) は示せた.
(3) $F_n := A_n\,(n = 1, 2, \ldots, N)$ と $F_n = \varnothing\,(n \geq N+1)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{N} A_n$ となる. 定義 1.6(2) を用いると

$$\mathsf{Pr}\left(\bigcup_{n=1}^{N} A_i\right) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \sum_{n=1}^{N} \mathsf{Pr}(A_n) + \sum_{n=N+1}^{\infty} \mathsf{Pr}(\varnothing)$$

$$= \sum_{n=1}^{N} \mathsf{Pr}(A_n)$$

を得る. よって, (3) は示された.
(4) $B \setminus A = B \cap A^c$ かつ $B = (B \cap A^c) \cup A$ である. このことに注意して, $F_1 := B \cap A^c$, $F_2 = A$, $F_n = \varnothing\,(n \geq 3)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = B$ となる. 定義 1.6(2) を用いると

$$\begin{aligned}\mathsf{Pr}(B) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) = \mathsf{Pr}(B \cap A^c) + \mathsf{Pr}(A) + \sum_{n=3}^{\infty} \mathsf{Pr}(\varnothing) \\ &= \mathsf{Pr}(B \cap A^c) + \mathsf{Pr}(A)\end{aligned}$$

がわかる. よって, (4) は示せた.
(5) $F_1 := A_1$, $F_2 = A_2 \setminus A_1$, $\ldots$, $F_{n+1} := A_{n+1} \setminus A_n$ とおく. $\{F_n\}_{n=1}^{\infty}$ は互いに排反であり

$$A_n = \bigcup_{i=1}^{n} F_i \;\Rightarrow\; A := \bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} A_n$$

となる. よって

$$\begin{aligned}\mathsf{Pr}(A) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathsf{Pr}(F_n) \quad (\because \text{定義 } 1.6(2)) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathsf{Pr}(F_n) = \lim_{N\to\infty} \mathsf{Pr}\left(\bigcup_{n=1}^{N} F_n\right) \quad (\because (3)) \\ &= \lim_{N\to\infty} \mathsf{Pr}(A_N)\end{aligned}$$

を得る. よって, (5) は示せた.
(6) $A = \bigcap_{n=1}^{\infty} A_n$ とおく. すると $\{A_1 \setminus A_n\}_{n=1}^{\infty}$ は $\varnothing = (A_1 \setminus A_1) \subset (A_1 \setminus A_2) \subset (A_1 \setminus A_3) \subset \cdots$ となる. このことに注意して分配法則を用いると

$$\bigcup_{n=1}^{\infty}(A_1 \setminus A_n) = \bigcup_{n=1}^{\infty}(A_1 \cap A_n^c) = A_1 \cap \left(\bigcap_{n=1}^{\infty} A_n\right)^c = A_1 \cap A^c = A_1 \setminus A \tag{1.1}$$

となる. 次に (4) と (6) を用いると

$$\begin{aligned}\mathsf{Pr}(A_1) - \mathsf{Pr}(A) &= \mathsf{Pr}(A_1 \setminus A) \quad (A_1 \supset A \text{ なので, } (4) \text{ を用いた}) \\ &= \lim_{n\to\infty} \mathsf{Pr}(A_1 \setminus A_n) \quad (\because (1.1) \text{ から } (5) \text{ を用いた}) \\ &= \lim_{n\to\infty} \Big(\mathsf{Pr}(A_1) - \mathsf{Pr}(A_n)\Big) \qquad \big(\because (4)\big)\end{aligned}$$

$$= \mathsf{Pr}(A_1) - \lim_{n\to\infty} \mathsf{Pr}(A_n)$$

から (6) は示せた.
(7) まず

$$\mathsf{Pr}(A_1 \cup A_2) \leq \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \tag{1.2}$$

を示す. そのために

$$A_1 \cup A_2 = A_1 \cup \Big(A_2 \setminus (A_1 \cap A_2) \Big) \quad \text{かつ} \quad A_1 \bigcap \Big(A_2 \setminus (A_1 \cap A_2) \Big) = \varnothing$$

であることに注意する. (3) から

$$\begin{aligned} \mathsf{Pr}\big(A_1 \cup A_2\big) &= \mathsf{Pr}(A_1) + \mathsf{Pr}\Big(A_2 \setminus (A_1 \cap A_2) \Big) \\ &= \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) - \mathsf{Pr}(A_1 \cap A_2) \\ &\qquad \big(\because A_2 \supset A_1 \cap A_2 \text{ なので } (4) \text{ を用いた} \big) \\ &\leq \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \qquad \big(\because \mathsf{Pr}(A_1 \cap A_2) \geq 0 \big) \end{aligned}$$

がわかる. (1.2) の操作を繰り返せば, $N \in \mathbb{N}$ に対して

$$\mathsf{Pr}\Big(\bigcup_{n=1}^{N} A_n \Big) \leq \sum_{n=1}^{N} \mathsf{Pr}(A_n) \tag{1.3}$$

がわかる. $\{\cup_{n=1}^{N} A_n\}_{N=1}^{\infty}$ は増加列なので, (5) と (1.3) から

$$\mathsf{Pr}\Big(\bigcup_{n=1}^{\infty} A_n \Big) = \lim_{N\to\infty} \mathsf{Pr}\Big(\bigcup_{n=1}^{N} A_n \Big) \leq \lim_{N\to\infty} \sum_{n=1}^{N} \mathsf{Pr}(A_n) = \sum_{n=1}^{\infty} \mathsf{Pr}(A_n)$$

がわかる. □

注意 1.8. 補題 1.7(6)(7) において, $\{\mathsf{Pr}(A_n)\}_{n=1}^{\infty}$ は有界な非減少列もしくは非増加列なので, 左辺は必ず収束することに注意せよ.

定義 1.9. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, $A, B \in \mathcal{A}$ とする.

(1) (独立性):

$$A \text{ と } B \text{ は } \underline{\text{独立}} \Leftrightarrow \mathsf{Pr}(A \cap B) = \mathsf{Pr}(A)\mathsf{Pr}(B).$$

(2) (条件付き確率) $\mathsf{Pr}(B) > 0$ のとき, B を与えたときの A の <u>条件付き確率</u> $\mathsf{Pr}(A|\, B)$ を

$$\mathsf{Pr}(A|\, B) := \frac{\mathsf{Pr}(A \cap B)}{\mathsf{Pr}(B)}$$

で定める.

(3) $m \geq 3$ とし, $A_j \in \mathcal{A}\,(j = 1, 2, \ldots, m)$ とする. 事象 $A_1, A_2, \ldots, A_m$ は 互いに独立 であるとは, $2 \leq \forall k \leq m$ と $\forall 1 \leq j_1 < j_2 < \cdots < j_k \leq m$ に対して

$$\mathsf{Pr}\big(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_k}\big) = \mathsf{Pr}(A_{j_1})\mathsf{Pr}(A_{j_2}) \cdots \mathsf{Pr}(A_{j_k})$$

が成り立つときをいう.

注意 1.10. (1) $A, B \in \mathcal{A}$ とし, $\mathsf{Pr}(B) > 0$ とする. A と B が独立のとき

$$\mathsf{Pr}(A|\, B) = \mathsf{Pr}(A)$$

が成立する.
(2) $\mathsf{Pr}(B) > 0$ のとき

$$\mathsf{Pr}(A \cap B) = \mathsf{Pr}(B)\mathsf{Pr}(A|\, B)$$

となる. これを 乗法の公式 という.
(3) $\mathsf{Pr}(B) > 0$ のとき, 関数 $\mathsf{Pr}(\,\cdot\,|B) : \mathcal{A} \to [0, 1]$ は定義 1.6(1) − (2) をみたす. すなわち $(\Omega, \mathcal{A})$ 上の確率である.
(4) $m \geq 3$ とする. 事象列 $A_1, A_2, \ldots, A_m$ は 対独立 であるとは

$$\mathsf{Pr}\big(A_i \cap A_j\big) = \mathsf{Pr}\big(A_i\big)\,\mathsf{Pr}\big(A_j\big) \qquad (i \neq j,\, i,\, j \in \{1, 2, \ldots, m\})$$

をみたすときをいう.
(5) 対独立でも独立でないことがある. 以下はその例である.

$$\Omega = \{b_1, b_2, b_3, b_4\}, \qquad \mathsf{Pr}(\{b_j\}) = \frac{1}{4} \qquad (j = 1, 2, 3, 4)$$

とする. いま

$$A_j = \{b_j, b_4\} \qquad (j = 1, 2, 3)$$

とする. すると

$$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \{b_4\}$$

から

$$\mathsf{Pr}(A_1) = \mathsf{Pr}(A_2) = \mathsf{Pr}(A_3) = \frac{1}{2},$$
$$\mathsf{Pr}(A_1 \cap A_2) = \mathsf{Pr}(A_1 \cap A_3) = \mathsf{Pr}(A_2 \cap A_3) = \mathsf{Pr}(A_1 \cap A_2 \cap A_3) = \frac{1}{4}$$

となる. したがって

$$\mathsf{Pr}(A_1 \cap A_2) = \mathsf{Pr}(A_1)\mathsf{Pr}(A_2);\ \mathsf{Pr}(A_1 \cap A_3) = \mathsf{Pr}(A_1)\mathsf{Pr}(A_3);$$

$$\mathsf{Pr}(A_2 \cap A_3) = \mathsf{Pr}(A_2)\mathsf{Pr}(A_3)$$

である. しかし

$$\frac{1}{4} = \mathsf{Pr}(A_1 \cap A_2 \cap A_3) \neq \mathsf{Pr}(A_1)\mathsf{Pr}(A_2)\mathsf{Pr}(A_3) = \frac{1}{8}$$

である. □

問 1.2. 注意 1.10(1)(3) を証明せよ.

補題 1.11. (全確率の法則) $N \in \mathbb{N}$ とする. $A_1, A_2, \ldots, A_N \in \mathcal{A}$ は Ω の分割とする. すなわち, すべての $m \neq n\,(m, n \in \{1, 2, \ldots, N\})$ に対し, $A_m \cap A_n = \varnothing$ で, $\bigcup_{n=1}^{N} A_n = \Omega$ が成立している. このとき任意の $B \in \mathcal{A}$ に対して

$$\mathsf{Pr}(B) = \mathsf{Pr}\left(\bigcup_{n=1}^{N}(A_n \cap B)\right) = \sum_{n=1}^{N}\mathsf{Pr}(A_n \cap B)$$

となる.

Proof. $B = B \cap \Omega = B \cap \left(\bigcup_{n=1}^{N} A_n\right) = \bigcup_{n=1}^{N}(A_n \cap B)$ と $\{A_i \cap B\}_{i=1}^{n}$ は互いに排反であることに注意して, 定義 1.4(2) を用いればよい. □

定理 1.12. (Bayes の定理) (1) $N \in \mathbb{N}$ とする. $A_1, A_2, \ldots, A_N, B \in \mathcal{A}$ とし, $A_1, A_2, \ldots, A_N$ は Ω の**分割**とする. $j = 1, 2, \ldots, N$ に対して, $\mathsf{Pr}(B) > 0$, $\mathsf{Pr}(A_j) > 0$ のとき

$$\mathsf{Pr}(A_j|\, B) = \frac{\mathsf{Pr}(A_j)\mathsf{Pr}(B|\, A_j)}{\sum_{n=1}^{N}\mathsf{Pr}(A_n)\mathsf{Pr}(B|\, A_n)}$$

が成立する.
(2) $A, B \in \mathcal{A}$ とし, $\mathsf{Pr}(A) > 0$, $\mathsf{Pr}(B) > 0$ と仮定する. このとき

$$\mathsf{Pr}(A|\, B)\mathsf{Pr}(B) = \mathsf{Pr}(B|\, A)\mathsf{Pr}(A)$$

が成立する.

Proof. 条件付き確率の定義, 補題 1.11, 定義 1.9(2) を用いればよい. □

例 1.13. 病気 D に対する検査の結果を $+$ と $-$ とし, 確率が以下であるとする.

$$\mathsf{Pr}(+|\, D) = 0.9, \quad \mathsf{Pr}(-|\, D^c) = 0.9, \quad \mathsf{Pr}(D) = 0.01.$$

Bayes の定理を用いて, 検査で + の人が本当に D である確率を求めると

$$\begin{aligned}\mathsf{Pr}(D|+) &= \frac{\mathsf{Pr}(+\cap D)}{\mathsf{Pr}(+)} = \frac{\mathsf{Pr}(D)\mathsf{Pr}(+|\,D)}{\mathsf{Pr}(D)\mathsf{Pr}(+|\,D)+\mathsf{Pr}(D^c)\mathsf{Pr}(+|\,D^c)}\\ &= \frac{0.01\times 0.9}{0.01\times 0.9+(1-0.01)\times\{1-0.9\}} \approx 0.083\end{aligned}$$

となる.

1.2 確率変数

前節では, 確率と事象を記述する数学的なモデルを導入した. しかし, 統計学が扱う現象は, 事象には直接結びつかないかもしれない数量的な情報である. 以下で定義する確率変数は, 事象と数量の間の橋渡しをする.

定義 1.14. $(\Omega, \mathcal{A})$ と $(\mathbb{X}, \mathcal{B})$ を可測空間とする. 写像 $X : \Omega \to \mathbb{X}$ は $(\Omega, \mathcal{A})$ から $(\mathbb{X}, \mathcal{B})$ への **可測写像** であるとは

$$X^{-1}(B) := \big\{\omega \in \Omega : X(\omega) \in B\big\} \in \mathcal{A} \qquad \big(\forall B \in \mathcal{B}\big)$$

をみたすときをいう.

注意 1.15. (1) $d = 1$ のとき, X は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の **確率変数** と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率変数と簡単にいうこともある.
(2) $d \geq 2\,(d \in \mathbb{N})$ とする. $\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ のとき, X は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の **確率ベクトル** と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率ベクトルということもある.

□

定理 1.16. $\big(\Omega, \mathcal{A}\big)$ と $\big(\mathbb{X}, \mathcal{B}\big)$ を可測空間とし, $X : \Omega \to \mathbb{X}$ を写像とし, $\mathcal{C}$ を $\mathbb{X}$ の集合族とする. $\forall C \in \mathcal{C}$ に対して, $\big\{\omega \in \Omega;\, X(\omega) \in C\big\} \in \mathcal{A}$ であり, $\mathcal{C}$ が $\mathcal{B}$ を生成する[5] とき, X は可測となる.

Proof. $B \in \mathcal{B}$ に対して, $\big\{\omega \in \Omega : X(\omega) \in B\big\}$ を $\{X \in B\}$ と書くことにする. $\{B\}, \{B_n\}_{n=1}^{\infty} \subset \mathcal{B}$ に対して

$$\Big\{X \in \bigcup_{n=1}^{\infty} B_n\Big\} = \bigcup_{n=1}^{\infty}\{X \in B_n\}$$
$$\big\{X \in B^c\big\} = \big\{X \in B\big\}^c$$

[5] $\mathcal{B}$ は $\mathcal{C}$ を含む最小の σ 集合族.

となる. したがって, 集合族 $\mathcal{D} := \{B \in \mathcal{B};\, \{X \in B\} \in \mathcal{A}\}$ は σ 加法族 (必要ならば, $\mathbb{X}$ も加える) となる. よって, $\mathcal{C} \subset \mathcal{D}$ であり, $\mathcal{C}$ は $\mathcal{B}$ を生成するので, 最小性から $\mathcal{B} \subset \mathcal{D}$ となる. よって, $\forall B \in \mathcal{B}$ に対して, $\{X \in B\} \in \mathcal{A}$ となることがわかる. 以上から X は確率変数であることが示せた. □

注意 1.17. (1). $\big(\mathbb{X},\, \mathcal{B}\big) = \big(\mathbb{R},\, \mathcal{B}(\mathbb{R})\big)$ に対して, 定理 1.16 における $\mathcal{C}$ の選択として, $\{(-\infty,\, r] : r \in \mathbb{R}\}$ と $\{(-\infty,\, q] : q \in \mathbb{Q}\}$ などがある. ただし, これらの集合族が $\mathcal{B}(\mathbb{R})$ を生成することは Dynkin 定理を用いる必要がある. 補遺 B.4.1 を参照のこと.
(2). $d \geq 2\,(d \in \mathbb{N})$ とする. $\big(\mathbb{X},\, \mathcal{B}\big) = \big(\mathbb{R}^d,\, \mathcal{B}(\mathbb{R})\big)$ に対して, 定理 1.16 における $\mathcal{C}$ の選択として

$$\{(a_1,\, b_1) \times (a_2,\, b_2) \times \cdots \times (a_d,\, b_d) : -\infty < a_i < b_i < \infty\,(i = 1,\, 2,\, \ldots,\, d)\}$$

がある. □

定理 1.18. $\big(\Omega,\, \mathcal{A}\big)$, $\big(\mathbb{X},\, \mathcal{B}\big)$, $\big(\mathbb{Y},\, \mathcal{C}\big)$ を可測空間とする. $X : \Omega \to \mathbb{X}$ と $f : \mathbb{X} \to \mathbb{Y}$ は可測写像のとき, $f(X) : \Omega \to \mathbb{Y}$ は可測写像となる.

Proof. $C \in \mathcal{C}$ に対して

$$\big\{\omega \in \Omega :\, f(X(\omega)) \in C\big\} = \big\{\omega \in \Omega :\, X(\omega) \in f^{-1}(C)\big\} \in \mathcal{A}$$

となる. ただし, $f^{-1}(C) = \{x \in \mathbb{X};\, f(x) \in C\}$ である. なぜならば, f の可測性から $f^{-1}(C) \in \mathcal{B}$ となることからわかる. □

定理 1.19. $n \in \mathbb{N}$ とし, $X_1,\, X_2,\, \ldots,\, X_n$ は確率変数とし, $f : \mathbb{R}^n \to \mathbb{R}$ は可測とする. このとき, $f(X_1,\, X_2,\, \ldots,\, X_n)$ は確率変数となる.

Proof. 定理 1.18 から $(X_1,\, X_2,\, \ldots,\, X_n)$ は確率ベクトルであることを示せばよい. $A_1,\, A_2,\, \ldots,\, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\big\{(X_1,\, X_2,\, \ldots,\, X_n) \in A_1 \times A_2 \times \cdots \times A_n\big\} = \bigcap_{i=1}^{n} \{X_i \in A_i\} \in \mathcal{A}$$

である. さらに, 集合族 $\{A_1 \times A_2 \times \cdots \times A_n;\, A_i \in \mathcal{B}(\mathbb{R})\,(i = 1,\, 2, \ldots,\, n)\}$ は $\mathcal{B}(\mathbb{R}^n)$ を生成[6] する. したがって, 定理 1.16 から $\big(\mathbb{X},\, \mathcal{B}\big) = \big(\mathbb{R}^n,\, \mathcal{B}(\mathbb{R}^n)\big)$ に対して, $(X_1,\, X_2,\, \ldots,\, X_n)$ は確率ベクトルであることがわかる. □

定理 1.20. $X_1,\, X_2,\, \ldots,\, X_n$ は確率変数のとき, $X_1 + X_2 + \cdots + X_n$ も確率変数となる.

[6] この事実の証明には, 測度論の議論が必要なので, 信じることにしよう.

Proof. 定理 1.19 から $\mathbb{R}^n$ 上の実数値関数 $f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$ が可測であることを示せばよい. $\forall r \in \mathbb{R}$ に対して

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n < r\}$$

は $\mathbb{R}^n$ の開集合となるので

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n < r\} \in \mathcal{B}(\mathbb{R}^n)$$

となる. 集合族 $\{(-\infty, r) : r \in \mathbb{R}\}$ は $\mathcal{B}(\mathbb{R})$ を生成するので[7], f は可測である. □

注意 1.21. 同様の議論から, X_1, X_2 は確率変数のとき, X_1X_2 も確率変数であることがわかる. さらに, $X_2 \neq 0$ のとき, X_1/X_2 も確率変数であることもわかる. □

定理 1.22. $X_1, X_2, \ldots$ は確率変数列のとき

$$\inf_n X_n \quad \sup_n X_n \quad \limsup_n X_n \quad \liminf_n X_n$$

も確率変数となる.

Proof. $r \in \mathbb{R}$ に対して

$$\left\{\omega \in \Omega; \inf_n X_n(\omega) < r\right\} = \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega; X_n(\omega) < r\right\} \in \mathcal{A}$$

から $\inf_n X_n$ は確率変数であることがわかる. 同様に

$$\left\{\omega \in \Omega; \sup_n X_n(\omega) < r\right\} = \bigcap_{n=1}^{\infty} \left\{\omega \in \Omega; X_n(\omega) < r\right\}$$

から $\sup_n X_n$ も確率変数であることがわかる. 次に

$$\liminf_n X_n = \sup_n \left(\inf_{m \geq n} X_m \right)$$

に注意する. $Y_n := \inf_{m \geq n} X_m$ は各 $n \in \mathbb{N}$ に対して確率変数なので, $\sup_n Y_n$ も確率変数となる. 同様に

$$\limsup_n X_n = \inf_n \left(\sup_{m \geq n} X_m \right)$$

から $\limsup_n X_n$ も確率変数であることはわかる. □

[7]定理 B.17 から, 開集合は開区間により生成されるので, 集合族 $\{(-\infty, r); r \in \mathbb{R}\}$ は $\mathcal{B}(\mathbb{R})$ を生成することがわかる.

注意 1.23. (1). $X_1, X_2, \ldots\ldots$ を確率変数列とする. 定理 1.22 と 定理 1.19 から

$$\begin{aligned}\Omega_0 :=&\big\{\omega \in \Omega :\ \lim_{n\to\infty} X_n(\omega) \text{ は存在 } \big\}\\ =&\big\{\omega \in \Omega :\ \limsup_{n\to\infty} X_n(\omega) - \liminf_{n\to\infty} X_n(\omega) = 0\big\}\end{aligned}$$

は可測集合となる. $\mathsf{Pr}(\Omega_0) = 1$ のとき, 確率変数列 $\{X_n\}_{n=1}^{\infty}$ は**ほとんど確実に収束**するといい, a.s. と記す. $\mathbb{R}\cup\{\pm\infty\}$ に値を取る確率変数を X_∞ とおくと

$$X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega) \qquad (\text{a.s.})\,\omega \in \Omega$$

と書ける.
(2). 任意の $a, b \in \mathbb{R}\,(a < b)$ とし, $\{-\infty\} \cup (-\infty, a)$, (a, b), $(b, \infty) \cup \{+\infty\}$ によって生成される σ 加法族を $\mathcal{B}(\mathbb{R}\cup\{\pm\infty\})$ と書く. 集合 $D \subset \Omega$ が定義域で, $\mathbb{R} \cup \{\pm\infty\}$ を終域ともつ関数 X は可測であるとは, $\forall B \in \mathcal{B}(\mathbb{R} \cup \{\pm\infty\})$ に対して

$$X^{-1}(B) = \{\omega \in \Omega :\ X(\omega) \in B\} \in \mathcal{A}$$

をみたすときをいう. □

定義 1.24. (1) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X に対して

$$\mathsf{F}^X(x) := \mathsf{Pr}(X \le x) := \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \le x\}) \qquad (x \in \mathbb{R})$$

を X **の累積分布関数**(cumulative distribution function(c.d.f.)) という. また

$$\mathsf{P}^X(B) := \mathsf{Pr}(X \in B) := \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \in B\}) \qquad (B \in \mathcal{B}(\mathbb{R}))$$

を X の**分布**という. したがって, $\mathsf{P}^X((-\infty, x]) = \mathsf{F}^X(x)$ である.
(2) 確率変数 X, Y を $(\Omega, \mathcal{A}, \mathsf{Pr})$ の確率変数とし, それぞれの c.d.f. を F^X, F^Y とする. このとき

$$\mathsf{F}^X(x) = \mathsf{F}^Y(x) \qquad (\forall x \in \mathbb{R})$$

が成立するとき, X **と** Y **の分布は同じ**である[8] という. これを $X \overset{d}{=} Y$ と書く.
(3) 確率変数 X が c.d.f. F を持つとき, $X \sim \mathsf{F}$ と書く.

[8] $\mathsf{F}^X(x) = \mathsf{F}^Y(x)\,(\forall x \in \mathbb{R}) \Leftrightarrow \mathsf{P}^X(B) = \mathsf{P}^Y(B)\,(\forall B \in \mathcal{B}(\mathbb{R})$ となることが知られている. 注意 1.26(2) を参照のこと.

注意 1.25. 文脈から, どの確率変数の確率分布または累積分布関数であることがわかる場合には, 簡単に P または F と書くこともある. □

注意 1.26. (1) P^X は可測空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度である. すなわち P^X は定義 1.6 をみたすことがわかる.
(2) c.d.f. F が与えられると

$$\mathsf{F}(x) = \mathsf{P}((-\infty, x]) \qquad (x \in \mathbb{R})$$

をみたす $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 P が一意的に定まることが知られている. このことにより X の c.d.f. と分布を同一視する. さらに $X \sim \mathsf{F}$ とし, c.d.f. F から定まる確率測度を P としたとき, $X \sim \mathsf{P}$ とも書く.
(3) 確率変数 X と Y の分布が同じとき

$$\mathsf{P}^X(B) = \mathsf{P}^Y(B) \qquad (\forall B \in \mathcal{B}(\mathbb{R}))$$

も成立する.

問 1.3. 注意 1.26(1) を示せ. 逆像の性質を利用すること.

定理 1.27. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし, F を X の c.d.f. とする. このとき F は次をみたす.

(1) F は非減少関数; $x < y \Rightarrow \mathsf{F}(x) \leq \mathsf{F}(y)$.

(2) $\lim_{x\to\infty} \mathsf{F}(x) = 1$; $\lim_{x\to-\infty} \mathsf{F}(x) = 0$.

(3) F は右連続関数; $\lim_{y\to x+0} \mathsf{F}(y) = \mathsf{F}(x)$.

(4) F の不連続点は高々可算個.

Proof. 確率変数 X の分布を P とする. すなわち $\mathsf{P}(B) := \mathsf{Pr}(X \in B)\ \big(\forall B \in \mathcal{B}(\mathbb{R})\big)$ である.
(1) $(-\infty, x] \subset (-\infty, y]$ に注意して, 補題 1.7(4) を P に適用すればよい.
(2) $A_n = (-\infty, n]$ と $A_n = (-\infty, -n]$ として, P に補題 1.7(5)(6) を適用すればよい.
(3) $A_n = \left(-\infty, x + \dfrac{1}{n}\right]$ として, P に補題 1.7(7) を適用すればよい.
(4) $n \in \mathbb{N}$ に対して

$$A_n := \left\{ r \in \mathbb{R};\, \mathsf{Pr}(X = r) > \frac{1}{n} \right\}$$

とおく. $\#A_n \leq n$ となる. なぜならば, n 個を越えてあると仮定すると, その中の n 個を $x_1 < x_2 < \cdots < x_n$ となるように選ぶことができる. すると

$$\mathsf{F}^X(x_n) = \mathsf{Pr}(X = x_n) + \lim_{x\to x_n-0} \mathsf{F}^X(x) > \frac{1}{n} + \mathsf{F}^X(x_{n-1})$$

$$> \frac{1}{n} + \mathsf{Pr}(X = x_{n-1}) + \lim_{x \to x_{n-1}-0} \mathsf{F}^X(x)$$
$$> \cdots > \frac{n-1}{n} + \mathsf{F}^X(x_1) > 1$$

となり, $0 \leq \mathsf{F}^X(x) \leq 1\,(x \in \mathbb{R})$ に矛盾する.

$$A := \{r \in \mathbb{R};\, \mathsf{Pr}(X = r) > 0\} = \bigcup_{n=1}^{\infty} A_n$$

なので, $\#A$ は高々可算となる. □

問 1.4. 固定した $x \in \mathbb{R}$ に対して

$$(-\infty,\, x] = \bigcap_{n=1}^{\infty} \left(-\infty,\, x + \frac{1}{n}\right)$$

を示せ.

問 1.5. 定理 1.27(1) – (3) の証明を詳しく書き下せ.

定理 1.28. 確率変数 X を確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数とし, X の c.d.f. を F とする. このとき次の (1) – (3) は同値である.

(1) F は $\mathbb{R}$ 上の連続関数.

(2) $\mathsf{F}(x) = \mathsf{F}(x-)\,(\forall x \in \mathbb{R})$. ただし, $\mathsf{F}(x-) := \sup\{\mathsf{F}(y);\, y < x\}$ とした.

(3) $\mathsf{Pr}(X = x) = 0\,(\forall x \in \mathbb{R})$.

Proof. ① で (1) ⇒ (2), ② で (2) ⇒ (3), ③ で (3) ⇒ (1) を証明する.
① (1) ⇒ (2) の証明: $x_n = x - \dfrac{1}{n}\,(n = 1,\, 2,\, \ldots)$ とすれば

$$\lim_{n \to \infty} \mathsf{F}(x_n) = F(x-)$$

である. この式から, F が連続ならば, $\mathsf{F}(x) = \mathsf{F}(x-)$ となることがわかる.
② (2) ⇒ (3) の証明:

$$\{X \leq x_n\} \subset \{X \leq x_{n+1}\}\,(n = 1,\, 2, \ldots), \quad \bigcup_{n=1}^{\infty} \{X \leq x_n\} = \{X < x\}$$

であるので, 補題 1.7(6) より

$$\mathsf{F}(x-) = \lim_{n \to \infty} \mathsf{F}(x_n) = \lim_{n \to \infty} \mathsf{Pr}(X \leq x_n) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} \{X \leq x_n\}\right)$$

$$= \mathsf{Pr}(X < x)$$

を得る. さらに補題 1.7(4) より

$$\mathsf{Pr}(X = x) = \mathsf{Pr}\Big(\{X \le x\} \setminus \{X < x\} \Big) = \mathsf{Pr}(X \le x) - \mathsf{Pr}(X < x)$$
$$= F(x) - F(x-) \tag{1.4}$$

であることからわかる. よって, F の右連続性から $\mathsf{Pr}(X = x) = 0$ となる.
③ (3) ⇒ (1) の証明: (1.4) から

$$\mathsf{Pr}(X = x) \Rightarrow \mathsf{F}(x) = \mathsf{F}(x-)$$

である. このことと F は右連続であることから F は連続となる. 以上から 3 つの条件は同値であることが示せた. □

問 1.6. 定理 1.28 の記号を踏襲する. 固定した $x \in \mathbb{R}$ に対して

$$\bigcup_{n=1}^{\infty} \left\{ X \le x - \frac{1}{n} \right\} = \bigcup_{n=1}^{\infty} \left\{ \omega \in \Omega;\, X(\omega) \le x - \frac{1}{n} \right\} = \{\omega \in \Omega;\, X(\omega) < x\}$$
$$= \{X < x\}$$

を示せ.

注意 1.29. F が $\mathbb{R}$ 上で連続のとき, $\mathsf{Pr}(X = x) = 0\,(\forall x \in \mathbb{R})$ なので, $a < b$ に対して

$$\mathsf{Pr}(a \le X \le b) = \mathsf{Pr}(a < X < b) = \mathsf{Pr}(a < X \le b) = \mathsf{Pr}(a \le X < b)$$

である. □

定義 1.30. 確率変数 X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし, その c.d.f. と分布を F と P と書く.
(1) X が高々可算個の集合 $\{x_1, x_2, \ldots\}$ 上にしか値を取らないとき, X を **離散型** であるという. この場合には

$$\mathsf{p}(x) := \mathsf{Pr}(X = x) = \mathsf{F}(x) - \mathsf{F}(x-) \qquad (x \in \{x_1, x_2, \ldots\})$$

で X の分布が特徴付けられる [9]. その p を X の **確率関数**(probability mass function(p.m.f.)) と呼ぶ.

[9] $\mathsf{p}(x) = 0\,(x \notin \{x_1, x_2, \ldots\})$ となるので, p は $\mathbb{R}$ 上の関数であり, $0 \le \mathsf{p}(x) \le 1$ となる.

(2) $\mathsf{Pr}(X = x) = 0\,(\forall x \in \mathbb{R})$ のとき, X を 連続型 であるという. さらにある非負値関数 p で

$$\mathsf{F}(x) = \int_{-\infty}^{x} \mathsf{p}(t)\,\mathrm{d}t \qquad (\forall x \in \mathbb{R})$$

をみたすものが存在するとき, p を X の 確率密度関数(probability density function(p.d.f.)) という. 特に F がほとんどいたるところ [10] で微分可能ならば, ほとんど至るところで

$$\overset{\bullet}{\mathsf{F}}(x) = \frac{\mathrm{d}\mathsf{F}}{\mathrm{d}x}(x) = \mathsf{p}(x)$$

となる.

注意 1.31. 確率変数 X の分布を P とする.
(1) X を離散型とし, $S = \{x \in \mathbb{R};\, \mathsf{p}(x) > 0\}$ とおく [11]. すると任意の Borel 集合 $B \subset \mathbb{R}$ に対して

$$\mathsf{P}(B) = \mathsf{Pr}(X \in B) = \sum_{x \in B \cap S} \mathsf{p}(x)$$

が成り立つことを示すことができる.
(2) X を連続型とし, その p.d.f. p が定義されるとする. このとき, 任意の Borel 集合 $B \subset \mathbb{R}$ に対して

$$\mathsf{P}(B) = \mathsf{Pr}(X \in B) = \int_B \mathsf{p}(x)\,\mathrm{d}x$$

が成り立つことを示すことができる. また,

$$\mathsf{Pr}(B) = \int_B \mathsf{p}_1(x)\,\mathrm{d}x = \int_B \mathsf{p}_2(x)\,\mathrm{d}x \qquad \big(\forall B \in \mathcal{B}(\mathbb{R})\big)$$

ならば, 高々可算個の点を除いて, $\mathsf{p}_1 = \mathsf{p}_2$ となることが知られている. したがって, 1.30(2) の p.d.f. は存在するならば, 高々可算個の点を除いて, 同じ値を取るものになる. これらの主張の厳密な証明は測度論の知識が必要となるので, 証明は省略する.

例 1.32. (1) $\Omega = \{0, 1\}$, $\mathcal{A} = 2^{\Omega}$, $\mathsf{Pr}(\{0\}) = \mathsf{Pr}(\{1\}) = 1/2$ とし

$$X : \Omega \ni \omega \mapsto X(\omega) = \omega \in \mathbb{R}$$

と定義すれば

$$\mathsf{Pr}(X = 0) = \mathsf{Pr}(X = 1) = \frac{1}{2}$$

[10] この授業では, $\mathbb{R}$ から可算個の点を除いた集合上で微分可能と理解して差し支えない.

[11] 累積分布関数 F は有界で非減少なので, その不連続点は高々可算個であるから, 下記の式の右辺の和記号が正当化される.

となる. このとき, X の c.d.f. F と p.m.f. p はそれぞれ

$$\mathsf{F}(x) = \begin{cases} 0 & (x < 0) \\ \frac{1}{2} & (0 \le x < 1) \\ 1 & (x \ge 1) \end{cases} ; \qquad \mathsf{p}(x) = \begin{cases} \frac{1}{2} & (x = 0,\, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる.

(2) $\Omega = [0,\, 1]$, $\mathcal{A} = \mathcal{B}(\mathbb{R}) \cap [0,\, 1] := \{B \cap [0,\, 1];\, B \in \mathcal{B}(\mathbb{R})\}$ とし

$$\mathsf{Pr}((a,\, b]) = b - a \qquad (0 \le a < b \le 1)$$

とする. さらに

$$X : \Omega \ni \omega \mapsto X(\omega) = \omega \in \mathbb{R}$$

と定義すれば, X の c.d.f. F と p.d.f. p はそれぞれ

$$\mathsf{F}(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \le x \le 1) \\ 1 & (x > 1) \end{cases} ; \quad \mathsf{p}(x) = \begin{cases} 1 & (0 \le x \le 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる.

1.3 分位点関数

定義 1.33. c.d.f. F に対して, 分位点関数(quantile function) $\mathsf{F}^- : (0,\, 1) \to \mathbb{R}$ を

$$\mathsf{F}^-(y) := \inf\bigl\{x \in \mathbb{R} :\, \mathsf{F}(x) \ge y\bigr\} \qquad (0 < y < 1)$$

で定義する. また, $y \in (0,\, 1)$ に対して, $\mathsf{F}^{-1}(y)$ を F の y 分位点 とよぶ. 1/2 分位点を メディアン (median) と呼ぶ.

注意 1.34. c.d.f. F が連続かつ $\operatorname{supp}\mathsf{F} = \{x \in \mathbb{R};\, 0 < \mathsf{F}(x) < 1\}$ 上で狭義単調増加のとき, F^- は F の逆関数になる. $\operatorname{supp}\mathsf{F}$ は累積分布関数が $\mathsf{F}(x) > 0$ かつ $\mathsf{F}(x) < 1$ となる点 x の全体である. 関数の台はこの集合の閉包で定義するので, 違いに注意せよ. □

注意 1.35. F^- の定義から, 任意の $0 < y < 1$ に対して, 単調減少列 $\{x_n\}_{n=1}^{\infty}$ が存在して

$$\mathsf{F}(x_n) \ge y \quad \text{かつ} \quad \lim_{n\to\infty} x_n = \mathsf{F}^{-1}(y)$$

とできる. このとき, $\{x_n\}_{n=1}^{\infty}$ の単調減少性と F の右連続性から

$$\mathsf{F}\bigl(\mathsf{F}^-(y)\bigr) = \mathsf{F}\bigl(\lim_{n\to\infty} x_n\bigr) = \lim_{n\to\infty} \mathsf{F}(x_n) \ge y \tag{1.5}$$

となることがわかる. よって, F^- の定義における inf は達成される. □

注意 1.36. c.d.f. F が連続のとき,

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y \qquad (0 < y < 1) \tag{1.6}$$

となることは以下の議論からわかる.
① $\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \le y$ の確認: c.d.f. F は連続とする. c.d.f. の性質

$$\lim_{n\to\infty} \mathsf{F}(x) = 1 \quad \text{かつ} \quad \lim_{x\to-\infty} \mathsf{F}(x) = 0$$

と F の連続性から, 中間値の定理を用いると任意の $0 < y < 1$ に対して, ある $x \in \mathbb{R}$ が存在して

$$\mathsf{F}(x) = y$$

をみたす. ここで, F^{-} の定義から

$$\mathsf{F}^{-}(y) \le x \tag{1.7}$$

となる. なぜならば, $\mathsf{F}^{-}(y)$ は条件 $\mathsf{F}(z) \ge y$ をみたす $z \in \mathbb{R}$ の inf である.
② $\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \ge y$ の確認: 一方, $\mathsf{F}(x) = y$ だから $x \in \{z \in \mathbb{R} : \mathsf{F}(z) \ge y\}$ なので, $x \ge \mathsf{F}^{-}(y)$ がわかる. (1.7) と F の非減少性から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \le \mathsf{F}(x) = y \tag{1.8}$$

となる. 一方, F^{-} の定義から

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \ge y \tag{1.9}$$

となる.
③ (1.6) の確認: (1.8) と (1.9) を合わせると

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y$$

を得る. したがって

$$\mathsf{F} \text{ が連続 } \Rightarrow \mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y\,(0 < y < 1)$$

となる.

□

注意 1.37. F が不連続ならば

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) = y$$

とは限らない. 反例は, 次の例 1.39 を参照せよ. □

例 1.38. $a < b\,(a, b \in \mathbb{R})$ とし, $X \sim \mathsf{U}(a, b)$ とする. すると X の c.d.f. は

$$\mathsf{F}(x) = \begin{cases} 0 & (x \leq a) \\ \dfrac{x-a}{b-a} & (a < x < b) \\ 1 & (x \geq b) \end{cases}$$

である. よって, F の分位点関数は

$$\mathsf{F}^{-}(y) = a + (b-a)y \qquad (y \in (0,\, 1))$$

となる. □

例 1.39. $\mathsf{Pr}(X = 0) = \mathsf{Pr}(X = 1) = \dfrac{1}{2}$ とする. このとき, X の c.d.f. は

$$\mathsf{F}(x) = \begin{cases} 0 & (x < 0) \\ \dfrac{1}{2} & (0 \leq x < 1) \\ 1 & (x \geq 1) \end{cases}$$

となる. よって, F の分位点関数は

$$\mathsf{F}^{-}(y) = \begin{cases} 0 & \left(0 < y \leq \dfrac{1}{2}\right) \\ 1 & \left(\dfrac{1}{2} < y < 1\right) \end{cases}$$

となる. この場合, $y \neq \dfrac{1}{2}$ のとき

$$\mathsf{F}\big(\mathsf{F}^{-}(y)\big) \neq y$$

となる. □

定理 1.40. 分位点関数は以下の性質 (1) ∼ (3) をもつ.
(1) F^{-} は非減少である.
(2) F^{-} は左連続である. すなわち, $\{y_n\}_{n=1}^{\infty}$ を任意の非減少列で $\lim\limits_{n\to\infty} y_n = y$ としたとき

$$\lim_{n\to\infty} \mathsf{F}^{-}(y_n) = \mathsf{F}^{-}(y)$$

が成立する.
(3) $\mathsf{F}^{-}(y) \leq x \Leftrightarrow y \leq \mathsf{F}(x)$ が成立する.

Proof. (1) $y_1 < y_2\,(y_1, y_2 \in (0, 1))$ に対して

$$\big\{x \in \mathbb{R} : \mathsf{F}(x) \geq y_1\big\} \supset \big\{x \in \mathbb{R} : \mathsf{F}(x) \geq y_2\big\}$$

となる. 両辺の inf をとると

$$\mathsf{F}^-(y_1) = \inf\big\{x \in \mathbb{R} : \mathsf{F}(x) \geq y_1\big\} \leq \inf\big\{x \in \mathbb{R} : \mathsf{F}(x) \geq y_2\big\} = \mathsf{F}^-(y_2)$$

となる. よって, F^- の非減少性が証明できた.
(2) $x_n := \mathsf{F}^-(y_n)\,(n \in \mathbb{N})$ とおく. 数列 $\{y_n\}_{n=1}^{\infty}$ は非減少なので, (1) から $\{x_n\}_{n=1}^{\infty}$ も非減少列となる. さらに, $y_n \leq y$ から

$$x_n = \mathsf{F}^-(y_n) \leq \mathsf{F}^-(y) =: x_0 \tag{1.10}$$

となる. よって, (1.10) の両辺の極限を取ると

$$x := \lim_{n\to\infty} x_n \leq x_0 \tag{1.11}$$

となる. ここで, $x < x_0$ を仮定して矛盾を導く. そのために

$$\epsilon := \frac{1}{2}(x_0 - x) > 0 \;\Leftrightarrow\; x = x_0 - 2\epsilon$$

とおく. F^- の定義から導出された (1.9) に注意すると

$$y_n \leq \mathsf{F}\big(\mathsf{F}^-(y_n)\big) \leq \mathsf{F}(x_n) \leq \mathsf{F}(x + \epsilon) = \mathsf{F}(x_0 - \epsilon)$$

となる. すると

$$y = \lim_{n\to\infty} y_n \leq \mathsf{F}(x_0 - \epsilon)$$

となる. しかし, F^- の定義から

$$\mathsf{F}(x_0 - \epsilon) < y$$

となる. なぜならば, $\mathsf{F}^-(y) = x_0$ なので, x_0 は条件 $\mathsf{F}(z) \geq y$ をみたす z の inf である. $x_0 > x_0 - \epsilon$ だから, $x_0 - \epsilon \notin \{z \in \mathbb{R} : \mathsf{F}(z) \geq y\}$ となる. よって, $\mathsf{F}(x_0 - \epsilon) < y$ となることがわかる. 以上から $y \leq \mathsf{F}(x_0 - \epsilon)$ かつ $y > \mathsf{F}(x_0 - \epsilon)$ となるので, 矛盾. よって

$$\mathsf{F}^-(y) = x_0 = x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} \mathsf{F}^-(y_n)$$

となるので, F^- の左連続性がわかる.
(3) F^- の非減少性から

$$y \leq \mathsf{F}(x) \;\Rightarrow\; \mathsf{F}^-(y) \leq \mathsf{F}^-\big(\mathsf{F}(x)\big) \leq x$$

となる. 最後の不等号は

$$\mathsf{F}^{-}\big(\mathsf{F}(x)\big) = \inf\big\{z \in \mathbb{R} : \mathsf{F}(z) \geq \mathsf{F}(x)\big\}$$

である. しかし, $x \in \big\{z \in \mathbb{R} : \mathsf{F}(z) \geq \mathsf{F}(x)\big\}$ であるので, $x \geq \mathsf{F}^{-}\big(\mathsf{F}(x)\big)$ がわかる.

逆に, $\mathsf{F}^{-}(y) \leq x$ ならば, (1.9) と F の非減少性から

$$y \leq \mathsf{F}\big(\mathsf{F}^{-}(y)\big) = \mathsf{F}(x)$$

がわかる. よって, (3) は証明された. □

系 1.41. F を c.d.f. とする. このとき, $U \sim \mathsf{U}(0, 1)$ に対して

$$X := \mathsf{F}^{-}(U) \sim \mathsf{F}$$

となる.

Proof. 定理 1.40(3) から, $\forall x \in \mathbb{R}$ に対して

$$\{X \leq x\} \Leftrightarrow \{U \leq \mathsf{F}(x)\}$$

となる. よって

$$\mathsf{Pr}(X \leq x) = \mathsf{Pr}(U \leq \mathsf{F}(x)) = \mathsf{F}(x)$$

を得る. □

1.4 主な 1 次元分布

1.4.1 離散型確率変数

Bernoulli 分布

$0 \leq \theta \leq 1$ とする. 確率変数 X は **母数 θ の Bernoulli 分布** に従うとは, X の p.m.f. p_θ が

$$\mathsf{p}_\theta(x) = p(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & (x = 0, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう. このことを $X \sim \mathsf{Ber}(\theta)$ と記す.

注意 1.42. $\theta = 0$ のとき, $0^0 = 1$ と定めていることに注意せよ.

2 項分布

$n \in \mathbb{N}$, $0 \le \theta \le 1$ とする. 確率変数 X は **母数 (n, θ) の 2 項分布** に従うとは, X の p.m.f. p_θ が

$$\mathsf{p}_{\theta,n}(x) = \mathsf{p}(x) = \begin{cases} \binom{n}{x} \theta^x (1-\theta)^{n-x} & (x = 0, 1, \ldots, n) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう. ただし

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}, \qquad 0! = 1$$

である. このことを $X \sim \mathsf{Bino}(n, \theta)$ と記す.

問 1.7. $\mathsf{p}_{\theta,n}(x)$ を 2 項分布 $\mathsf{Bino}(n, \theta)$ $(0 < \theta < 1, n \in \mathbb{N})$ p.m.f. としたとき, $\sum_{x=0}^{n} \mathsf{p}_{\theta,n}(x) = 1$ を確認せよ.

幾何分布

$0 < \theta < 1$ とする. 確率変数 X は **母数 θ の幾何分布** に従うとは, X の p.m.f. p_θ が

$$\mathsf{p}_\theta(x) = \mathsf{p}(x) = \begin{cases} \theta(1-\theta)^{x-1} & (x = 1, 2, \ldots) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう. このことを $X \sim \mathsf{Geo}(\theta)$ と記す.

問 1.8. $0 < \theta < 1$ とする. p_θ を幾何分布 $\mathsf{Geo}(\theta)$ の p.m.f. としたとき, $\displaystyle\sum_{x=0}^{\infty} \mathsf{p}_\theta(x) = 1$ を確認せよ.

Poisson 分布

$\theta > 0$ とする. 確率変数 X は **母数 θ の Poisson 分布** に従うとは, X の p.m.f. p_θ が

$$\mathsf{p}_\theta(x) = \mathsf{p}(x) = \begin{cases} e^{-\theta} \dfrac{\theta^x}{x!} & (x = 0, 1, 2, \ldots) \\ 0 & (\text{その他の場合}) \end{cases}$$

のときをいう. このことを $X \sim \mathsf{Po}(\theta)$ と記す.

問 1.9. $\theta > 0$ とする. $\mathsf{p}_\theta(x)$ を Poisson 分布 $\mathsf{Po}(\theta)$ の p.m.f. としたとき, $\displaystyle\sum_{x=0}^{\infty} \mathsf{p}_\theta(x) = 1$ を確認せよ.

1.4.2 連続型確率変数

正規分布

$\mu \in \mathbb{R}$, $0 < \sigma < \infty$ とする. 確率変数 X は平均 μ, 分散 σ^2 の正規分布[12] に従うとは, X の p.d.f. $\mathsf{p}_{\mu,\sigma^2}$ が

$$\mathsf{p}_{\mu,\sigma^2}(x) = \mathsf{p}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (-\infty < x < \infty)$$

のときをいう. ただし $\exp(x) = e^x$ である. このことを $X \sim \mathsf{N}(\mu, \sigma^2)$ と記す. $\mu = 0$, $\sigma = 1$ のときの分布を標準正規分布という.

注意 1.43. p.d.f. $\mathsf{p}_{\mu,\sigma^2}(x)$ のグラフは単峰型であり, μ に関して左右対称となる. さらに, μ が p.d.f. $\mathsf{p}_{\mu,\sigma^2}(x)$ のグラフの峰の位置に対応することがわかる. また, σ が大きくなると p.d.f. $\mathsf{p}_{\mu,\sigma^2}(x)$ のグラフの裾が長くなることがわかる.

注意 1.44. よく知られている事実として

$$\int_0^\infty e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}$$

がある. $\mathbb{R} \ni t \mapsto e^{-t^2} \in (0, \infty)$ は偶関数であることに注意して, $t = z/\sqrt{2}$ と変換すると

$$\int_{-\infty}^\infty e^{-z^2/2}\,\mathrm{d}z = \sqrt{2\pi}$$

となることがわかる. さらに, $z = (x-\mu)/\sigma$ と変数変換をすると

$$\int_{-\infty}^\infty \mathsf{p}_{\mu,\sigma^2}(x)\,\mathrm{d}x = 1$$

となることが確認できる. □

ガンマ分布

$\alpha > 0$, $\beta > 0$ とする. 確率変数 X は母数 (α, β) のガンマ分布に従うとは, X の p.d.f. $\mathsf{p}_{\alpha,\beta}$ が

$$\mathsf{p}_{\alpha,\beta}(x) = \mathsf{p}(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

[12] 母数 (μ, σ^2) を平均と分散となぜ呼ぶかは第 2 章で判明する.

のときをいう. ただし

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x \qquad (\alpha > 0)$$

である. このことを $X \sim \mathsf{Ga}(\alpha, \beta)$ と記す. $\lambda > 0$ とし, $\alpha = 1$, $\beta = \dfrac{1}{\lambda}$ とおく. このとき, $\mathsf{Ga}(1, 1/\lambda)$ を 母数 λ の指数分布 といい, $\mathsf{Exp}(\lambda)$ と書く. また, $p \in \mathbb{N}$ とし, $\alpha = \dfrac{p}{2}$, $\beta = 2$ とおく. このとき, $\mathsf{Ga}(p/2, 2)$ を 自由度 p の χ^2 分布 といい, χ^2_p と記す.

問 1.10. $\mathsf{p}(x|\,\alpha, \beta)$ をガンマ分布 $\mathsf{Ga}(\alpha, \beta)\,(\alpha > 0, \beta > 0)$ の p.d.f. としたとき, $\int_0^\infty \mathsf{p}_{\alpha,\beta}(x)\,\mathrm{d}x = 1$ を示せ.

1.5　2 次元の分布

1.5.1　同時確率関数と確率密度関数

確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の離散型確率変数 (X, Y) の対に対して, 同時確率関数(同時 p.m.f.) p を

	$\mathrm{Y}=0$	$\mathrm{Y}=1$	
$\mathrm{X}=0$	$1/9$	$2/9$	$1/3$
$\mathrm{X}=1$	$2/9$	$4/9$	$2/3$
	$1/3$	$2/3$	

で定める. たとえば

$$\mathsf{p}(1, 1) = \mathsf{Pr}(X = 1, Y = 1) = \frac{4}{9}$$

である.

定義 1.45. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X, Y は連続型とする. $\mathbb{R}^2$ 上の非負値実数値関数 p が確率ベクトル (X, Y) の 同時確率密度関数(同時 p.d.f.) であるとは, 次の条件をみたすときをいう.

(1) $\mathsf{p}(x, y) \geq 0\,(\forall (x, y) \in \mathbb{R}^2)$.

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}x\mathrm{d}y = 1$.

(3) $\forall A \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\mathsf{Pr}((X, Y) \in A) = \int\int_A \mathsf{p}(x, y)\,\mathrm{d}x\mathrm{d}y.$$

注意 1.46. (X, Y) の同時累積分布関数 (同時 c.d.f.) F を

$$\begin{aligned}\mathsf{F}(x, y) :&= \mathsf{Pr}(X \leq x, Y \leq y) \qquad (1.12)\\ &:= \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \leq x, Y(\omega) \leq y\}) \qquad (\forall (x, y) \in \mathbb{R}^2)\end{aligned}$$

で定義する. (X, Y) の同時分布 P を

$$\mathsf{P}(A) = \mathsf{Pr}\big((X, Y) \in A\big) \quad (\forall A \in \mathcal{B}(\mathbb{R}^2)) \qquad (1.13)$$

で定義する. どの確率変数の c.d.f. または確率分布であるかを明示したいときには, $\mathsf{F}^{(X,Y)}$ または $\mathsf{P}^{(X,Y)}$ と記す.

細かなことであるが, (1.12) の F によって $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ 上の確率測度 P を一意的に定めていることができる. この点に関しては補遺 B.4.1 章の測度論の知識が必要となる. □

例 1.47. 連続型確率変数のベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}(x, y) = \begin{cases} x + y & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき集合 $\{(x, y) \in \mathbb{R}^2;\, \mathsf{p}(x, y) > 0\}$ に注意すれば

$$\begin{aligned}\int_0^1 \int_0^1 \mathsf{p}(x, y)\, \mathrm{d}x\mathrm{d}y &= \int_0^1 \left[\int_0^1 x\, \mathrm{d}x\right] \mathrm{d}y + \int_0^1 \left[\int_0^1 y\, \mathrm{d}y\right] \mathrm{d}x \\ &= \int_0^1 \frac{1}{2}\, \mathrm{d}y + \int_0^1 \frac{1}{2}\, \mathrm{d}x = 1\end{aligned}$$

となる. □

1.5.2 周辺分布

定義 1.48. (X, Y) を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率ベクトルとする.
(1) (X, Y) が離散型で同時 p.m.f. p を持つとする. X の <u>周辺確率関数</u>(周辺 p.m.f.) を

$$\mathsf{p}^X(x) = \mathsf{Pr}(X = x) = \sum_{y \in S_Y} \mathsf{Pr}(X = x, Y = y) = \sum_{y \in S_Y} \mathsf{p}(x, y) \qquad (\forall x \in S_X)$$

で定義する. ただし

$$\begin{aligned}S_X &:= \{x \in \mathbb{R};\, \text{ある } y \in \mathbb{R} \text{ に対して}, \mathsf{p}(x, y) > 0\} \\ S_Y &:= \{y \in \mathbb{R};\, \text{ある } x \in \mathbb{R} \text{ に対して}, \mathsf{p}(x, y) > 0\}\end{aligned}$$

である.
(2) (X, Y) は連続型とし, 同時 p.d.f. p を持つとする. このとき X の周辺確率密度関数(周辺 p.d.f.) を

$$\mathsf{p}^X(x) = \int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}y \qquad (\forall x \in \mathbb{R})$$

で定義し, Y の 周辺確率密度関数(周辺 p.d.f.) を

$$\mathsf{p}^Y(x) = \int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}x \qquad (\forall y \in \mathbb{R})$$

で定義する.

注意 1.49. 連続型確率ベクトル (X, Y) に対して

$$\begin{aligned}\mathsf{F}^X(x) := \mathsf{Pr}(X \le x) &= \int\int_{(s,t)\in\mathbb{R}^2;\, s\le x} \mathsf{p}(s, t)\,\mathrm{d}s\mathrm{d}t \\ &= \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} \mathsf{p}(s, t)\,\mathrm{d}t \right\} \mathrm{d}s = \int_{-\infty}^{x} \mathsf{p}^X(s)\,\mathrm{d}s\end{aligned}$$

となるので, X の周辺 p.d.f. と X の p.d.f. は高々可算個の点を除いて同じ値を取ることがわかる. □

注意 1.50. F を確率ベクトル (X, Y) の同時 c.d.f. とし, F^X を X の周辺 c.d.f. とする. このとき $x \in \mathbb{R}$ と $y \in \mathbb{R}$ に対して

$$\mathsf{F}^X(x) = \lim_{y\to\infty} \mathsf{F}(x, y); \qquad \mathsf{F}^Y(y) = \lim_{x\to\infty} \mathsf{F}(x, y)$$

が成立する. □

問 1.11. 注意 1.50 の式を証明せよ.

1.5.3 独立な分布と条件付き分布

定義 1.51. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の 2 つの確率変数 X と Y は 独立 であるとは, $\forall A, B \in \mathcal{B}(\mathbb{R})$ に対して

$$\mathsf{Pr}(X \in A, Y \in B) = \mathsf{Pr}(X \in A)\mathsf{Pr}(Y \in B)$$

が成り立つときをいう. 独立でないとき X と Y は 従属 であるという.

定理 1.52. p を確率ベクトル (X, Y) の同時 p.d.f. とし, p^X と p^Y を X と Y それぞれの周辺 p.d.f. とする. このとき

$$X \text{ と } Y \text{ は独立} \Leftrightarrow \mathsf{p}(x, y) = \mathsf{p}^X(x)\mathsf{p}^Y(y) \quad (\forall x, y \in \mathbb{R}).$$

となる.

Proof. 独立ならば, 同時 p.d.f. が周辺 p.d.f. の積で表現されることの証明は易しい. 逆はこの講義の範囲 (測度の拡張の議論が必要となる!) を超えるので, 信じることにする. □

注意 1.53. (1) 連続型の場合には, 測度 0 の集合を除いて[13]p.d.f. は定義されるので, 定理 1.52 の書き方はやや数学的な厳密性に欠ける.
(2) 離散型確率変数の場合には, p.d.f. を p.m.f. に置き換えればよい. □

例 1.54. 連続型確率変数 X と Y は独立で同時 p.d.f.

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. このとき確率 $\mathsf{Pr}(X + Y \leq 1)$ を求めてみよう. 独立性より (X, Y) の同時 p.d.f. は

$$\mathsf{p}^{(X,Y)}(x, y) = \begin{cases} 4xy & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. したがって

$$\begin{aligned} \mathsf{Pr}(X + Y \leq 1) &= \int\int_{x+y\leq 1} \mathsf{p}^{(X,Y)}(x, y)\,\mathrm{d}x\mathrm{d}y = 4\int_0^1 x\left\{\int_0^{1-x} y\,\mathrm{d}y\right\}\mathrm{d}x \\ &= 4\int_0^1 x\frac{(1-x)^2}{2}\,\mathrm{d}x = \frac{1}{6} \end{aligned}$$

となる. □

定理 1.55. 連続型確率ベクトル (X, Y) は同時 p.d.f. $\mathsf{p}(x, y)$ を持つとする. ただし $\{(x, y) \in \mathbb{R}^2; \mathsf{p}(x, y) > 0\}$ は矩形[14] とする. このときある非負値関数 q と r が存在して

$$\mathsf{p}(x, y) = q(x)r(y)$$

と書けるとき, X と Y は独立である.

Proof. 積分をして確かめればよい. □

注意 1.56. 定理 1.55 の主張も「ほとんど至る所」でよい. □

[13] この用語は定義されていない. 「有限個の点を除いて」と理解してもこの講義の内容の範囲では問題ない.
[14] 有界でなくともよい.

例 1.57. 連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}(x, y) = \begin{cases} 2e^{-(x+y)} & (x > 0, y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. $\{(x, y) \in \mathbb{R}^2; \mathsf{p}(x, y) > 0\} = (0, \infty)^2$ なので

$$q(x) = \begin{cases} 2e^{-x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases};$$

$$r(y) = \begin{cases} e^{-y} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

と取れば

$$\mathsf{p}(x, y) = q(x)r(y)$$

となるので, X と Y は独立である. □

定義 1.58. (1) 離散型確率ベクトル (X, Y) は同時 p.m.f. $\mathsf{p}(x, y)$ を持つとする. $\mathsf{p}_Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の<u>条件付き確率関数</u>(条件付き p.m.f.) を

$$\mathsf{p}^{X|Y}(x|\, y) = \mathsf{Pr}(X = x|\, Y = y) = \frac{\mathsf{Pr}(X = x, Y = y)}{\mathsf{Pr}(Y = y)} = \frac{\mathsf{p}(x, y)}{\mathsf{p}^Y(y)} \quad (x \in \mathbb{R})$$

で定義する.
(2) 連続型確率ベクトル (X, Y) は同時 p.d.f. $\mathsf{p}(x, y)$ を持つとする. $\mathsf{p}^Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の<u>条件付き確率密度関数</u>(条件付き p.d.f.) を

$$\mathsf{p}^{X|Y}(x|\, y) = \frac{\mathsf{p}(x, y)}{\mathsf{p}^Y(y)} \quad (x \in \mathbb{R})$$

で定義する.

注意 1.59. (X, Y) が連続型確率ベクトルのとき, $\mathsf{p}^Y(y) > 0$ なる $y \in \mathbb{R}$ に対して, $Y = y$ を与えたときの事象 $\{X \in A\}$ の条件付き確率を

$$\mathsf{Pr}(X \in A|\, Y = y) := \int_A \mathsf{p}^{X|Y}(x|\, y)\, \mathrm{d}x \qquad \big(\forall A \in \mathcal{B}(\mathbb{R})\big)$$

と形式的に定義する. □

例 1.60. 連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}(x, y) = \begin{cases} x + y & (0 < x < 1, 0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする．このとき $\mathsf{Pr}(X \leq 1/4 | Y = 1/3)$ を求めてみよう．まず $0 < y < 1$ に対して

$$\mathsf{p}^Y(y) = \int_{-\infty}^{\infty} \mathsf{p}(x, y)\,\mathrm{d}x = y + \int_0^1 x\,\mathrm{d}x = y + \frac{1}{2}$$

となる．したがって

$$\mathsf{p}^Y(y) = \begin{cases} y + \frac{1}{2} & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．よって $0 < y < 1$ と $0 < x < 1$ に対して

$$\mathsf{p}^{X|Y}(x|\,y) = \frac{\mathsf{p}(x, y)}{\mathsf{p}^Y(y)} = \frac{x + y}{y + \frac{1}{2}}$$

となる．これより $0 < y < 1$ のとき

$$\mathsf{p}^{X|Y}(x|\,y) = \begin{cases} \dfrac{x + y}{y + \frac{1}{2}} & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．注意 1.59 より

$$\mathsf{Pr}\left(X \leq \frac{1}{4} \middle| Y = \frac{1}{3}\right) = \int_0^{1/4} \mathsf{p}^{X|Y}\left(x \middle| \frac{1}{3}\right) \mathrm{d}x = \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}}\,\mathrm{d}x = \frac{11}{80}$$

となる． □

1.6 多次元分布と i.i.d. 標本

$X_1, X_2, \ldots, X_n$ を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$$

と書く．$\boldsymbol{X}$ を確率ベクトルという．

注意 1.61. 本講義録では，ベクトルは縦ベクトルとする． □

定義 1.62. $X_1, X_2, \ldots, X_n$ が 独立 であるとは，すべての Borel 集合 $A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\mathsf{Pr}(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{j=1}^{n} \mathsf{Pr}(X_j \in A_j) \tag{1.14}$$

が成立するときである．

注意 1.63. $\boldsymbol{X}$ の同時 p.d.f. を $\mathsf{p}(x_1, x_2, \ldots, x_n)$ と書き, 各 $X_j\,(j = 1, 2, \ldots, n)$ の周辺 p.d.f. を p^{X_j} と書くことにする. (1.14) を示すためには

$$\mathsf{p}(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} \mathsf{p}^{X_j}(x_j) \quad (\forall x_1, x_2, \ldots, x_n \in \mathbb{R})$$

を示せばよい. □

定義 1.64. $X_1, X_2, \ldots, X_n$ は独立で, 各 $X_j\,(j = 1, 2, \ldots, n)$ は同じ c.d.f. F を持つとき, $X_1, X_2, \ldots, X_n$ は独立同一分布に従う(i.i.d. =identically and independently distributed) といい

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{F}$$

と書く[15]. $X_1, X_2, \ldots, X_n$ は累積分布関数 F からの標本の大きさが n のランダム標本ともいう.

1.6.1 重要な多次元分布モデル

この本で取り上げる代表的な多次元分布モデルをあげておく. 離散型分布は多項分布, 連続型分布は多変量正規分布である.

多項分布

2 項分布を多変量にしたものが多項分布である. $d, n \in \mathbb{N}$ とし, $\boldsymbol{p} = (p_1, p_2, \ldots, p_d)$, $p_j\,(0 \le p_j \le 1; j = 1, 2, \ldots, d)$, $\sum_{j=1}^{d} p_j = 1$ とする. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ は母数 $(d, \boldsymbol{p})$ の多項分布に従うとは, $\boldsymbol{X}$ の p.m.f. が

$$\mathsf{p}_{n,d,\boldsymbol{p}}(\boldsymbol{x}) = \begin{cases} \begin{pmatrix} d \\ x_1, x_2, \ldots, x_d \end{pmatrix} p_1^{x_1} p_2^{x_2} \times \cdots \times p_d^{x_d} & \begin{pmatrix} x_j = 0, 1, \ldots, n; \\ j = 1, 2, \ldots, d \end{pmatrix} \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられたときをいう. ただし, $\sum_{j=1}^{d} x_j = n$ で

$$\begin{pmatrix} n \\ x_1, x_2, \ldots, x_d \end{pmatrix} = \frac{n!}{x_1! x_2! \cdots x_d!}, \quad 0! = 1$$

[15] p.d.f. p を使い

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}$$

と書く. $\sim$ の右側には, 確率測度/確率分布/p.d.f./p.m.f./$\mathsf{N}(0, 1)$ など分布を特定するものを書いてよいことにする.

である. これを $\boldsymbol{X} \sim \mathsf{Multi}_d(n, \boldsymbol{p})$ と記す.

補題 1.65. $\boldsymbol{X} \sim \mathsf{Multi}_d(n, \boldsymbol{p})$ とする. ただし, $\boldsymbol{p} = (p_1, p_2, \ldots, p_d)$ と $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ とする. このとき $X_j\,(j = 1, 2, \ldots, d)$ の周辺分布は $\mathsf{Bino}(n, p_j)$ である.

Proof. x_2 から x_d について同時 p.m.f. の和をとればよい. □

多変量正規分布

$Z_1, Z_2, \ldots, Z_d \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, 1)$ に対して

$$\boldsymbol{Z} := \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_d \end{pmatrix} \tag{1.15}$$

と定める [16]. すると $\boldsymbol{Z}$ の同時 p.d.f. は

$$\begin{aligned} \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z}) = \mathsf{p}^{\boldsymbol{Z}}(z_1, z_2, \ldots, z_d) &= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z_j^2\right\} \\ &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2} \sum_{j=1}^{d} z_j^2\right\} \\ &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2} \boldsymbol{z}^\top \boldsymbol{z}\right\} \quad (\boldsymbol{z} \in \mathbb{R}^d) \end{aligned}$$

と書ける. ただし $\boldsymbol{z} = (z_1, z_2, \ldots, z_d)^\top$ である. これを $\boldsymbol{Z} \sim \mathsf{N}_d(\boldsymbol{0}, \boldsymbol{I}_d)$ と記す. ただし $\boldsymbol{I}_d$ は d 次単位行列である. さらに定義より

$$\int \cdots \int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{Z}}(z_1, z_2, \ldots, z_d)\, \mathrm{d}z_1 \mathrm{d}z_2 \cdots dz_d = 1$$

となっていることもわかる.

次に

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix} \in \mathbb{R}^d, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{pmatrix}, \quad \begin{array}{l} \sigma_{ij} = \sigma_{ji} \\ (i, j = 1, 2, \ldots, d) \end{array}$$

[16] この節ではベクトルは縦とする. 下の式では縦ベクトルと横ベクトルを混ぜて表現している. これは記号の乱用であるが, $f((z_1, z_2, \ldots, z_d)^\top)$ などと書くのは煩わしい.

とする. ただし $\boldsymbol{\Sigma}$ は正定値対称行列[17] とする.　このとき, ある $d \times d$ の正則行列 $\mathbf{A}$ が存在して

$$(1)\ \mathbf{A} \text{ は対称行列}, (2)\ \boldsymbol{\Sigma} = \mathbf{A}^2$$

と取れる. これを $\boldsymbol{\Sigma}$ の平方根といい, $\boldsymbol{\Sigma}^{1/2}$ と書くことにする. これを用いて

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} \tag{1.16}$$

と定めたとき, $\boldsymbol{X}$ の分布を $\mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ と記す. このとき $\boldsymbol{X}$ の同時 p.d.f. は

$$\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (\boldsymbol{x} \in \mathbb{R}^d) \tag{1.17}$$

で与えられる. ただし, $\det(\cdot)$ は行列式を表す. この分布を **平均ベクトル $\boldsymbol{\mu}$, 分散共分散行列 $\boldsymbol{\Sigma}$ の d 変量正規分布** という.

注意 1.66. (1.17) の導出は以下のように行うことができる. 一般に

$$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_d)^\top$$

を d 次元確率ベクトルとし, その同時 p.d.f. を $\mathsf{p}^{\boldsymbol{Z}}$ と書くことにする. さらに $\mathbb{Z} := \{\boldsymbol{z} \in \mathbb{R}^d;\ \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z}) > 0)\}$ とし, 関数

$$\boldsymbol{h}(\cdot) = (h_1(\cdot), h_2(\cdot), \ldots, h_d(\cdot))^\top : \mathbb{Z} \to \boldsymbol{h}(\mathbb{Z})$$

は 1 対 1 とし, $j = 1, 2, \ldots, d$ に対して

$$X_j = h_j(Z_1, Z_2, \ldots, Z_d); \quad \boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$$

とおく. $\boldsymbol{h}$ は 1 対 1 なので, $\boldsymbol{h}$ の逆写像を

$$\boldsymbol{h}^{-1} = (h_1^{-1}, h_2^{-1}, \ldots, h_d^{-1})^\top : \qquad \boldsymbol{h}(\mathbb{Z}) \to \mathbb{Z}$$

が存在すると仮定する. よって, $\boldsymbol{h}^{-1}\big(\boldsymbol{h}(\boldsymbol{z})\big) = \boldsymbol{z}$ なので, $\boldsymbol{Z} = \boldsymbol{h}^{-1}(\boldsymbol{X})$ となる. ここで, すこし記号の乱用をしている. すなわち, h_j^{-1} は $\boldsymbol{h}^{-1}$ の第 j 成分の表記であって, h_1 の逆関数ではないことに注意せよ. 変換 $\boldsymbol{h}$ の Jacobian を

$$J_{\boldsymbol{h}}(\boldsymbol{z}) = \det \begin{pmatrix} \frac{\partial h_1(\boldsymbol{z})}{\partial z_1} & \cdots & \frac{\partial h_1(\boldsymbol{z})}{\partial z_d} \\ \vdots & \ddots & \vdots \\ \frac{h_d(\boldsymbol{z})}{\partial z_1} & \cdots & \frac{\partial h_d(\boldsymbol{z})}{\partial z_d} \end{pmatrix}$$

[17] $\forall \boldsymbol{x} \in \mathbb{R}^d\,(\boldsymbol{x} \neq \mathbf{0}_d)$ に対して, $\boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} > 0$ が成立する. したがって, 逆行列 $\boldsymbol{\Sigma}^{-1}$ が存在する.

とおく. さらに, $\boldsymbol{X}$ の同時 p.d.f. を求めるために, $\boldsymbol{h}^{-1}(\boldsymbol{x})$ の Jacobian $\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})$ を

$$\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) = \det \begin{pmatrix} \frac{\partial h_1^{-1}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_1^{-1}(\boldsymbol{x})}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{h_d^{-1}(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial h_d^{-1}(\boldsymbol{x})}{\partial x_d} \end{pmatrix}$$

で定める.

$\mathbf{J}_{\boldsymbol{h}}(\boldsymbol{z}) \neq 0\,(\boldsymbol{z} \in \mathbb{Z})$ [18] が成立するとき

$$\mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{Z}}\big(\boldsymbol{h}^{-1}(\boldsymbol{x})\big)}{J_{\boldsymbol{h}}\big(\boldsymbol{h}^{-1}(\boldsymbol{x})\big)} = \big|\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})\big|\mathsf{p}^{\boldsymbol{Z}}\big(\boldsymbol{h}^{-1}(\boldsymbol{x})\big) \tag{1.18}$$

となることが知られている. これは, 標準的な多変数の積分の変換公式からわかる.

すると $\boldsymbol{h}(\boldsymbol{z}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{z} \Leftrightarrow \boldsymbol{h}^{-1}(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ より

$$\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) = \det(\boldsymbol{\Sigma})^{-1/2} = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}}$$

を (1.18) に代入すれば, (1.17) はわかる. 以上の議論から

$$\int\int\cdots\int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_d = \int\int\cdots\int_{\mathbb{R}^d} \mathsf{p}^{\boldsymbol{Z}}(\boldsymbol{z})\,\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_d = 1$$

となっていることもわかる. □

問 1.12. $\boldsymbol{h}^{-1}(\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ のとき

$$J_{\boldsymbol{h}^{-1}}(\boldsymbol{x}) = \frac{1}{\sqrt{\det\boldsymbol{\Sigma}}}$$

を示せ.

注意 1.67. $\boldsymbol{\Sigma}$ を半正定値としたときに, 多変量正規分布を (1.15) と変換 (1.16) を用いて定義することがことができる. しかし, $\det(\boldsymbol{\Sigma}) = 0$ のときは, $\forall B \in \mathcal{B}(\mathbb{R}^d)$ に対して, 確率 $\mathsf{Pr}(\boldsymbol{X} \in B)$ は定義できるが, $\boldsymbol{X}$ は同時 p.d.f. を持たないことが知られている. このような分布を退化した多変量正規分布という. □

定理 1.68. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top \sim \mathsf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とする. ただし $\boldsymbol{\mu} \in \mathbb{R}^d$ で, $\boldsymbol{\Sigma}$ は $d \times d$ の正定値行列である. このとき次が成立する.

(1) $X_j\,(j = 1, 2, \ldots, d) \sim \mathsf{N}(\mu_j, \sigma_{jj})$. ただし μ_j は $\boldsymbol{\mu}$ の第 j 成分, σ_{jj} は $\boldsymbol{\Sigma}$ の第 j 対角成分である.

[18] 実は, a.e $\mathbb{X}$ でよい.

(2) $X_\ell = x_\ell\,(\ell = 1, 2, \ldots, d)$ を与えたときの $X_j\,(j \neq \ell)$ の条件付き分布は

$$X_j|\, X_\ell = x_\ell \sim \mathsf{N}\left(\mu_j + \frac{\sigma_{j\ell}}{\sigma_{jj}}(x_\ell - \mu_\ell),\, \sigma_{jj} - \frac{\sigma_{j\ell}\sigma_{\ell j}}{\sigma_{\ell\ell}}\right)$$

となる.

(3) 定数ベクトル $\boldsymbol{c} \in \mathbb{R}^d\,(\boldsymbol{c} \neq \mathbf{0}_d)$ に対して

$$\boldsymbol{c}^\top \boldsymbol{X} \sim \mathsf{N}(\boldsymbol{c}^\top \boldsymbol{\mu},\, \boldsymbol{c}^\top \boldsymbol{\Sigma} \boldsymbol{c})$$

となる. ただし, $\mathbf{0}_d$ は $\mathbb{R}^d$ の零ベクトルである.

(4) $V := (\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_d$ である. なお, 自由度 d の χ^2 分布 χ^2_d は定義 1.74 で導入する.

Proof. (1) $j = 1$ として示す. $\boldsymbol{\Sigma}$ の $(i, j)\,(i, j = 1, 2, \ldots, d)$ 成分を σ_{ij} と書き, $(d-1) \times (d-1)$ 行列 $\boldsymbol{\Sigma}_1$ と $(d-1) \times 1$ ベクトル $\boldsymbol{\sigma}_1$ をそれぞれ

$$\boldsymbol{\Sigma}_1 = (\sigma_{ij})_{i,\, j=2,3,\ldots,d}, \quad \boldsymbol{\sigma}_1 = (\sigma_{21},\, \sigma_{31},\, \ldots,\, \sigma_{d1})^\top$$

で定める. すると

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} \sigma_{11} & \boldsymbol{\sigma}_1^\top \\ \hline \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}_2 \end{array}\right)$$

となる. このとき

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline \sigma_{11}^{-1}\boldsymbol{\sigma}_1 & \boldsymbol{I}_{d-1} \end{array}\right) \left(\begin{array}{c|c} \sigma_{11} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top \end{array}\right) \left(\begin{array}{c|c} 1 & \sigma_{11}^{-1}\boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{I}_{d-1} \end{array}\right)$$

と書ける. ただし $\boldsymbol{I}_{d-1}$ は $(d-1)$ 次の単位行列で, $\mathbf{0}_{d-1}$ は $(d-1) \times 1$ の零ベクトルである. $\boldsymbol{\Sigma}$ が正定値ならば, $\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top$ も正定値であるので

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} = &\left(\begin{array}{c|c} 1 & -\sigma_{11}^{-1}\boldsymbol{\sigma}_1^\top \\ \hline \mathbf{0}_{d-1} & \boldsymbol{I}_{d-1} \end{array}\right) \left(\begin{array}{c|c} \sigma_{11}^{-1} & \mathbf{0}_{d-1}^\top \\ \hline \mathbf{0}_{d-1} & \left\{\boldsymbol{\Sigma}_2 - \sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1} \end{array}\right) \\ &\times \left(\begin{array}{c|c} 1 & \mathbf{0}_{d-1}^\top \\ \hline -\sigma_{11}^{-1}\boldsymbol{\sigma}_1 & \boldsymbol{I}_{d-1} \end{array}\right) \end{aligned} \tag{1.19}$$

となる. $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)^\top$ と $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top$ と書いたときに, $\boldsymbol{\mu}_1 = (\mu_2, \mu_3, \ldots, \mu_d)^\top$ と $\boldsymbol{x}_1 = (x_2, x_3, \ldots, x_d)^\top$ とおく. (1.19) から

$$\mathrm{tr}\,(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

$$= \frac{1}{\sigma_{11}}(x_1-\mu_1)^2 + \operatorname{tr}\left[\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)^\top \left\{\boldsymbol{\Sigma}_2-\sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1} \times\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)\right]$$

$$=: \frac{1}{\sigma_{11}}(x_1-\mu_1)^2 + g_{d-1}(\boldsymbol{x}_1) \tag{1.20}$$

となる. ただし

$$g_{d-1}(\boldsymbol{x}_1) = \operatorname{tr}\left[\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)^\top \left\{\boldsymbol{\Sigma}_2-\sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\right\}^{-1} \times\left(\boldsymbol{x}_1-\boldsymbol{\mu}_1-\frac{x_1-\mu_1}{\sigma_{11}}\boldsymbol{\sigma}_1\right)\right]$$

である. このことから

$$\begin{aligned}&\int_{\mathbb{R}^{d-1}} \exp\left[-\frac{1}{2}\operatorname{tr}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]\mathrm{d}x_2\,\mathrm{d}x_3\,\times\,\mathrm{d}x_d\\ &\quad= \exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]\int_{\mathbb{R}^{d-1}}\exp\left[-\frac{1}{2}g_{d-1}\right]\mathrm{d}x_2\,\mathrm{d}x_3\,\times\,\mathrm{d}x_d\\ &\quad= \exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]\det\left(2\pi\big(\boldsymbol{\Sigma}_2-\sigma_{11}^{-1}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1^\top\big)\right)^{1/2}\end{aligned}$$

がわかる. よって

$$\int_{\mathbb{R}^{d-1}} \mathsf{p}^{\boldsymbol{X}}\big(\boldsymbol{x}|\,\boldsymbol{\mu},\,\boldsymbol{\Sigma}\big)\,\mathrm{d}x_2\,\mathrm{d}x_3,\cdots\mathrm{d}x_d = \frac{1}{\sqrt{2\pi\sigma_{11}}}\exp\left[\frac{(x_1-\mu_1)^2}{2\sigma_{11}}\right]$$

がわかる.

(2) (1.20) からわかる.

(3) $\boldsymbol{c}^\top\boldsymbol{X}$ の 第 2 章で定義する積率を計算すればよい.

(4) (3) から

$$\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X}-\boldsymbol{\mu})\ \sim\ \mathsf{N}_d(\boldsymbol{0}_d,\,\boldsymbol{I}_d)$$

となる. あとは χ^2 分布の定義からわかる. □

(1.18) を用いた例題を述べておく

例 1.69. 連続型確率ベクトル $(X_1,\,X_2)$ は同時 p.d.f.

$$\mathsf{p}(x_1,\,x_2) = \begin{cases} 1 & (0<x_1<1,\,0<x_2<1)\\ 0 & (\text{その他の場合})\end{cases}$$

を持つとする. さらに

$$y_1 = h_1(x_1,\,x_2) = x_1+x_2,\ y_2 = h_2(x_1,\,x_2) = x_2$$

$$\Leftrightarrow x_1 = h_1^{-1}(y_1, y_2),\ x_2 = h_2^{-1}(y_1, y_2)$$

とし, $\boldsymbol{Y} = (h_1(X_1, X_2), h_2(X_1, X_2))^\top$, $\boldsymbol{y} = (y_1, y_2)^\top$ とする. なお, 上式では記号の乱用をしている. h_1^{-1}, h_2^{-1} は単に $\boldsymbol{h}^{-1}$ の成分を表し, h_1, h_2 の逆写像ではないことに注意せよ. このとき

$$J_{\boldsymbol{h}^{-1}}(\boldsymbol{y}) = \det \begin{pmatrix} \frac{\partial h_1^{-1}(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_1^{-1}(\boldsymbol{y})}{\partial y_2} \\ \frac{\partial h_2^{-1}(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_2^{-1}(\boldsymbol{y})}{\partial y_2} \end{pmatrix} = \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1$$

より

$$\begin{aligned} \mathsf{p}^{\boldsymbol{Y}}(\boldsymbol{y}) &= \begin{cases} 1 & (0 < y_1 - y_2 < 1,\ 0 < y_2 < 1) \\ 0 & (\text{その他の場合}) \end{cases} \\ &= \begin{cases} 1 & (y_2 < y_1 < 1 + y_2,\ 0 < y_2 < 1) \\ 0 & (\text{その他の場合}) \end{cases} \end{aligned}$$

を得る. また

$$\int\int_{\mathbb{R}^2} \mathsf{p}^{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}y_1\mathrm{d}y_2 = 1$$

となっていることに注意せよ. □

1.7　正規分布に関連した分布

正規分布に関連した分布であるガンマ分布, χ^2 分布, F 分布, t 分布を定義し, これらの基本的な性質を述べる. これらの分布は正規母集団から標本に基づく統計量の分布 (いわゆる標本分布) の議論で重要な役割を担う. それぞれの分布の確率密度関数の導出は自己完結な形で行った.

1.8　ガンマ分布

まず, ガンマ関数を導入する. $a > 0$ に対してガンマ関数を

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,\mathrm{d}x$$

で定義する. $a > 1$ のとき, 部分積分により

$$\begin{aligned} \Gamma(a) &= \int_0^\infty x^{a-1}e^{-x}\,\mathrm{d}x = -x^{a-1}e^{-x}\Big|_0^\infty + (a-1)\int_0^\infty x^{a-2}e^{-x}\,\mathrm{d}x \\ &= (a-1)\Gamma(a-1) \end{aligned}$$

となる. $\Gamma(1) = \int_0^\infty e^{-x}\,\mathrm{d}x = 1$ なので, 帰納法により

$$\Gamma(k) = (k-1)! \qquad (k = 1,\, 2,\, \ldots)$$

となる.

定義 1.70. (ガンマ分布) $a,\, b > 0$ とする. 連続型確率変数 X が母数 $(a,\, b)$ のガンマ分布に従うとは, X の p.d.f. が

$$\mathsf{p}_{a,\,b}(x) = \begin{cases} \dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられるときをいう. これを $X \sim \mathsf{Ga}(a,\, b)$ と記す.

補題 1.71. $k \geq 2$ は自然数とする. $X_1,\, X_2,\, \ldots,\, X_k$ を独立な確率変数, 各 $X_j \sim \mathsf{Ga}(a_j,\, b)\,(j = 1,\, 2,\, \ldots,\, k,\, a_j > 0,\, b > 0$ としたとき

$$X_1 + X_2 + \cdots + X_k \sim \mathsf{Ga}(a_1 + a_2 + \cdots + a_k,\, b)$$

となる.

Proof. 証明は次の補題を用いてする. □

補題 1.72. 連続型確率変数 $X,\, Y$ は独立とし, それぞれの p.d.f. を $\mathsf{p}^X,\, \mathsf{p}^Y$ とする. このとき $X + Y$ の p.d.f. は

$$\mathsf{p}^{X+Y}(x) = \int_{-\infty}^{\infty} \mathsf{p}^X(x-y)\mathsf{p}^Y(y)\,\mathrm{d}y$$

で与えられる.

Proof. 確率ベクトル $(X,\, Y)$ の同時 p.d.f. は $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ になることに注意する. $t \in \mathbb{R}$ に対して

$$\begin{aligned} \mathsf{Pr}\big(X + Y \leq t\big) &= \int\int_{x+y \leq t} \mathsf{p}^X(x)\mathsf{p}^Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathsf{p}^Y(y) \left\{ \int_{-\infty}^{t-y} \mathsf{p}^X(x)\,\mathrm{d}x \right\} \mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathsf{p}^Y(y) \left\{ \int_{-\infty}^{t} \mathsf{p}^X(z-y)\,\mathrm{d}z \right\} \mathrm{d}y \\ &\qquad (x\,\big|_{-\infty}^{t-y} = z - y\,\big|_{-\infty}^{t} \text{ と変換}) \\ &= \int_{-\infty}^{t} \left\{ \int_{-\infty}^{\infty} \mathsf{p}^X(z-y)\mathsf{p}^Y(y)\,\mathrm{d}y \right\} \mathrm{d}z \qquad (\because \text{Fubini の定理}^{19}) \end{aligned}$$

となる. よって微積分の基本定理より

$$\mathsf{p}^{X+Y}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(X+Y \le t\big) = \int_{-\infty}^{\infty} \mathsf{p}^{X}(t-y)\mathsf{p}^{Y}(y)\,\mathrm{d}y$$

を得る. □

補題 1.71 の証明. $X_1 \sim \mathsf{Ga}(a_1, b)$, $X_2 \sim \mathsf{Ga}(a_2, b)$ とし, X_1 と X_2 は独立とする. $\mathsf{Ga}(a, b)$ の p.d.f. を $\mathsf{p}a, b(\,\cdot\,)$ と書く. 補題 1.72 より

$$\begin{aligned}
\mathsf{p}^{X_1+X_2}(x) &= \int_{-\infty}^{\infty} \mathsf{p}_{a_1,b}(x-y)\mathsf{p}_{a_2,b}(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{b^{a_1}}{\Gamma(a_1)}(x-y)^{a_1-1}e^{-b(x-y)} \cdot \frac{b^{a_2}}{\Gamma(a_2)}y^{a_2-1}e^{-by}\mathrm{d}y \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)}e^{-bx}\int_0^x (x-y)^{a_1-1}y^{a_2-1}\,\mathrm{d}y \\
&\quad \big(x-y>0 \text{ かつ } y>0 \text{ より } 0<y<x \text{ となる.}\big) \\
&= \frac{b^{a_1+a_2}}{\Gamma(a_1)\Gamma(a_2)}x^{a_1+a_2-1}e^{-bx}\int_0^1 (1-z)^{a_1-1}z^{a_2-1}\,\mathrm{d}z \\
&\quad (y = xz \text{ と変換})
\end{aligned}$$

を得る. ここで $\mathsf{p}^{X_1+X_2}$ は p.d.f. であることに注意すると

$$\begin{aligned}
1 &= \int_0^{\infty} \mathsf{p}^{X_1+X_2}(x)\,\mathrm{d}x \\
&= \frac{1}{\Gamma(a_1)\Gamma(a_2)}\left(\int_0^1 (1-z)^{a_1-1}z^{a_2-1}\,\mathrm{d}z\right) b^{a_1+a_2}\int_0^{\infty} x^{a_1+a_2-1}e^{-bx}\,\mathrm{d}x \\
&= \frac{1}{\Gamma(a_1)\Gamma(a_2)}\left(\int_0^1 (1-z)^{a_1-1}z^{a_2-1}\,\mathrm{d}z\right) \underbrace{\int_0^{\infty} w^{a_1+a_2-1}e^{-w}\,\mathrm{d}w}_{=\Gamma(a_1+a_2)} \\
&\quad (w = bx \text{ と変換}) \\
&= \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}\int_0^1 (1-z)^{a_1-1}z^{a_2-1}\,\mathrm{d}z \qquad (\text{ガンマ関数の定義より})
\end{aligned}$$

より

$$\int_0^1 (1-z)^{a_1-1}z^{a_2-1}\,\mathrm{d}z = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)} \tag{1.21}$$

を得る. よって

$$X_1 + X_2 \sim \mathsf{Ga}(a_1+a_2, b)$$

がわかる. あとはこのことを繰り返せばよい.

注意 1.73. $a_1 > 0$, $a_2 > 0$ に対して

$$B(a_1,\, a_2) := \int_0^1 (1-z)^{a_1-1} z^{a_2-1}\, \mathrm{d}z$$

を ベータ関数 とよぶ. すると (1.21) から

$$B(a_1,\, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+a_2)}$$

という関係式が成り立つことがわかる. □

1.9 χ^2 分布

定義 1.74. $Z_1, Z_2, \ldots, Z_k \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\, 1)$ とする.

$$S = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

の分布を 自由度 k の χ^2 分布 といい, $S \sim \chi_k^2$ と記す.

$Z \sim \mathsf{N}(0,\, 1)$ のとき $Y = Z^2$ の p.d.f. は

$$\mathsf{p}^Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases} \tag{1.22}$$

となる. 演習問題 1.3 を参照せよ. さらに

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

であったこと [20] を思い出すと

$$\mathsf{p}^Y(y) = \begin{cases} \dfrac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)} y^{(1/2)-1} e^{-y/2} & (y > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

[20] $\mathsf{N}(0,\, 1/2)$ の p.d.f. から

$$\sqrt{\pi} = \sqrt{2\pi(1/2)} = \int_{-\infty}^{\infty} e^{-x^2}\, \mathrm{d}x = 2\int_0^{\infty} e^{-x^2}\, \mathrm{d}x$$

である. ここで, $t = x^2$ とおく. すると $\mathrm{d}t = 2x\mathrm{d}x$ から $\mathsf{d}x = \dfrac{1}{2\sqrt{t}}\mathrm{d}t$ となるので

$$\sqrt{\pi} = 2\int_0^{\infty} e^{-x^2} \underbrace{\mathrm{d}x}_{=\frac{1}{2}t^{(1/2)-1}\mathrm{d}t} = \int_0^{\infty} t^{(1/2)-1} e^{-t}\, \mathrm{d}t = \Gamma\left(\frac{1}{2}\right)$$

がわかる.

がわかる. これは $\mathsf{Ga}(1/2,\, 1/2)$ の p.d.f. である. さらに補題 1.71 より

$$\chi^2_k = \mathsf{Ga}\left(\frac{k}{2},\, \frac{1}{2}\right)$$

がわかる. よって χ^2_k の p.d.f. は

$$\mathsf{p}_k(x) = \begin{cases} \dfrac{\left(\dfrac{1}{2}\right)^{k/2}}{\Gamma\left(\dfrac{k}{2}\right)} x^{(k/2)-1} e^{-x/2} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

である.

1.10　F 分布

定義 1.75. $k,\, m \in \mathbb{N}$ とする. X と Y は独立で $X \sim \chi^2_k$ と $Y \sim \chi^2_m$ とする. このとき

$$F = \frac{X/k}{Y/m}$$

の分布を 自由度 $(k,\, m)$ の F 分布 といい, $\mathsf{F} \sim \mathsf{F}(k,\, m)$ と記す.

注意 1.76. 上の定義では, F 分布の記号を $\mathsf{F}(k,\, m)$ と書いた. 分布関数の記号と同じ F の字体 F を使用しているが, 文脈から判断できるので記号の乱用をする. □

補題 1.77. $k,\, m \in \mathbb{N}$ とする. $F \sim \mathsf{F}(k,\, m)$ のとき, F の p.d.f. $\mathsf{p}(x|\, k,\, m)$ は

$$\mathsf{p}_{k,\,m}(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{k+m}{2}\right)}{\Gamma\left(\dfrac{k}{2}\right)\Gamma\left(\dfrac{m}{2}\right)} \dfrac{k^{k/2} m^{m/2} x^{(k/2)-1}}{(m+kx)^{(k+m)/2}} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる.

Proof. 証明は次の補題を用いて行う. □

補題 1.78. $X,\, Y$ を独立な正値の連続型確率変数とし, それぞれの p.d.f. を $\mathsf{p}^X(x)$ と $\mathsf{p}^Y(x)$ とする. このとき

$$Z = \frac{X}{Y}$$

の p.d.f. は

$$\mathsf{p}^Z(z) = \begin{cases} \displaystyle\int_0^\infty y\,\mathsf{p}^X(zy)\mathsf{p}^Y(y)\,\mathrm{d}y & (z>0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる.

Proof. 仮定から確率ベクトル (X, Y) の同時 p.d.f. は $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ となる. よって $t>0$ に対して

$$\begin{aligned} \mathsf{Pr}\big(Z\le t\big) &= \int_0^\infty \left\{ \int_0^{ty} \mathsf{p}^X(x)\mathsf{p}^Y(y)\,\mathrm{d}x \right\} \mathrm{d}y \\ &= \int_0^\infty \left\{ \int_0^{t} \mathsf{p}^X(zy)\mathsf{p}^Y(y)y\,\mathrm{d}z \right\} \mathrm{d}y \quad (x=zy \text{ と変換}) \\ &= \int_0^t \left\{ \int_0^\infty \mathsf{p}^X(zy)\mathsf{p}^Y(y)y\,\mathrm{d}y \right\} \mathrm{d}z \qquad (\because \text{Fubini の定理}) \end{aligned}$$

となる. ここで微積分の基本定理を用いると

$$\mathsf{p}^Z(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(Z\le t\big) = \int_0^\infty y\,\mathsf{p}^X(ty)\mathsf{p}^Y(y)\,\mathrm{d}y$$

となる. □

補題 1.77 の証明　補題 1.78 を用いる. $z>0$ のとき, $Z=X/Y$ の p.d.f. は

$$\begin{aligned} \mathsf{p}^Z(z) &= \int y\,\mathsf{p}^X(zy)\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \frac{\left(\frac{1}{2}\right)^{k/2}\left(\frac{1}{2}\right)^{m/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} \int_0^\infty y(zy)^{(k/2)-1}e^{-(zy)/2}y^{(m/2)-1}e^{-y/2}\,\mathrm{d}y \\ &= \frac{\left(\frac{1}{2}\right)^{(k+m)/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} z^{(k/2)-1} \int_0^\infty y^{(k+m)/2-1}e^{-(z+1)y/2}\,\mathrm{d}y \\ &= \frac{\left(\frac{1}{2}\right)^{(k+m)/2}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} z^{(k/2)-1} \int_0^\infty \left(\frac{2x}{z+1}\right)^{(k+m)/2-1} e^{-x}\frac{2}{z+1}\,\mathrm{d}x \\ &\qquad \big(x=\tfrac{z+1}{2}y \text{ と変換}\big) \end{aligned}$$

$$= \frac{1}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} z^{(k/2)-1}(1+z)^{-(k+m)/2} \int_0^\infty x^{(k+m)/2-1} e^{-x}\,\mathrm{d}x$$
$$= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} z^{(k/2)-1}(1+z)^{-(k+m)/2}$$

となる. $F = \dfrac{m}{k}Z$ なので

$$\mathsf{p}_{k,\,m}(x) = \mathsf{p}^Z\left(\frac{k}{m}x\right)\frac{k}{m}$$
$$= \frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} (kx/m)^{(k/2)-1}(1+(kx/m))^{-(k+m)/2}$$

となる. この式を整理すると主張は証明される.

1.11　t 分布

定義 1.79. $Z_0,\, Z_1,\, Z_2,\, \ldots,\, Z_k \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ のとき

$$T = \frac{Z_0}{\sqrt{\dfrac{1}{k}\left(Z_1^2 + Z_2^2 + \cdots + Z_k^2\right)}}$$

を 自由度 k の t 分布 といい, これを $T \sim t_k$ と記す.

補題 1.80. T の p.d.f. は

$$\mathsf{p}_k^T(x) = \begin{cases} \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{k}{2}\right)} \dfrac{1}{\sqrt{k}} \left(1 + \dfrac{x^2}{k}\right)^{-(k+1)/2} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる

Proof. 定義より

$$T^2 \sim \mathsf{F}(1,\,k)$$

であることをまず思い出す. T の分布は対称なので

$$\mathsf{p}_k^T(x) = \mathsf{p}_k^T(-x) \qquad (x \geq 0)$$

となる. よって $x > 0$ に対して

$$2\mathsf{Pr}\big(0 \leq T \leq x\big) = \mathsf{Pr}\big(-x \leq T \leq x\big) = \mathsf{Pr}\big(T^2 \leq x^2\big)$$

となる. 上の式から $\mathsf{F}(1,\, k)$ の p.d.f. を $\mathsf{p}_{1,\,k}(\,\cdot\,)$ と書けば

$$2\int_0^x \mathsf{p}_k^T(t)\,\mathrm{d}t = \int_0^{x^2} \mathsf{p}_{1,\,k}(t)\,\mathrm{d}t \tag{1.23}$$

となる. さらに, $t \mapsto t^2$ と変換すると

$$\int_0^{x^2} \mathsf{p}_{1,\,k}(t)\,\mathrm{d}t = \int_0^x \mathsf{p}_{1,\,k}(t^2)2t\,\mathrm{d}t \tag{1.24}$$

となる. よって, (1.23) と (1.24) を合わせると

$$\int_0^x \mathsf{p}_k^T(t)\,\mathrm{d}t = \int_0^x \mathsf{p}_{1,\,k}(t^2)t\,\mathrm{d}t \tag{1.25}$$

を得る. ここで補題 1.77 から

$$\mathsf{p}_{1,\,k}(t) = \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{k}{2}\right)} \frac{k^{k/2}t^{-1/2}}{(k+t)^{(k+1)/2}}$$

であることを思い出す. この式を (1.25) の右辺に代入して, x について両辺を微分して, 微積分の基本定理を用いると

$$\mathsf{p}_k^T(x) = \mathsf{p}_{1,\,k}(x^2)\underbrace{x}_{=t^{1/2}} = \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{k}{2}\right)} \frac{1}{\sqrt{k}} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}$$

となる. □

1.12 章末注釈と参考文献

第 1.1 節は [28] を参照した. 第 1.2 節は [8] を参照した. 第 1.3 節から第 1.6 節は [42] を参照した. 第 1.7 節は [23] からの借用である.

1.13 演習問題

演習問題 1.1. 確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の離散型確率変 X は以下の p.m.f. を持つ :

$$\mathsf{p}(x) = \begin{cases} \begin{pmatrix} 3 \\ x \end{pmatrix} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x} & (x = 0,\, 1,\, 2,\, 3) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. ただし

$$\begin{pmatrix} 3 \\ x \end{pmatrix} = \frac{3!}{x! \times (3-x)!}, \qquad 0! = 1$$

である.
(1) $\mathsf{p}(0)$, $\mathsf{p}(1)$, $\mathsf{p}(2)$, $\mathsf{p}(3)$ の値を計算せよ.
(2) $\mathsf{Pr}(0 < X < 3)$ を求めよ.

演習問題 1.2. $\Omega := \{1, 2, 3, 4\}$ 上の部分集合族 $\mathcal{F} = \big\{\{1\}, \{2, 3\}, \{4\}\big\}$ を考える.
(1) $\mathcal{F}$ は Ω 上の σ 加法族でないことを示せ.
(2) $\mathcal{F}$ を含む最小の σ 加法族 を求めよ.
(3) 写像 $\mathsf{Pr} : \mathcal{F} \mapsto \mathbb{R}$ を

$$\mathsf{Pr}(\{1\}) = \frac{1}{2}, \mathsf{Pr}(\{2, 3\}) = \frac{1}{3}, \mathsf{Pr}(\{4\}) = \frac{1}{6}$$

で定める. (2) で求めた σ 加法族を $\mathcal{A}$ と記したとき, 写像 Pr が可測空間 $\big(\Omega, \mathcal{A}\big)$ 上の確率測度[21] となるように $\mathsf{Pr}(A)$ の値を定めよ. ただし, $\forall A \in \mathcal{A} \setminus \mathcal{F}$ である.

演習問題 1.3. $Z \sim \mathsf{N}(0, 1)$ とする. Z の p.d.f. を

$$\mathsf{p}(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad (-\infty < z < \infty)$$

としたとき, 以下の問いに答えよ.
(1) 任意の $y > 0$ に対して, $\mathsf{Pr}(Z^2 \leq y)$ を p も用いて表現せよ.
(2) $Y := Z^2$ とし, Y の p.d.f. を p^Y と書く. このとき, $y > 0$ に対して

$$\mathsf{p}^Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} \mathsf{Pr}(Y \leq y)$$

となることを利用して, Y の p.d.f. が (1.22) で与えられることを示せ.

演習問題 1.4. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2, \infty)}(X) := \begin{cases} 1 & \big(X > \frac{1}{2}\big) \\ 0 & \big(X \leq \frac{1}{2}\big) \end{cases}$$

[21] 定まる確率測度は一意であることも証明できるが, それは信じてよい.

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の c.d.f. を求めよ. なお, X の c.d.f. を F と書くことにする.
(2) 確率変数 Y の p.m.f. を求めよ. なお, X の p.m.f. を p^Y と書くことにする.
(3) 確率 $\mathsf{Pr}\left(0 < X \leq \frac{1}{2}\right)$, $\mathsf{Pr}(Y = 0)$ を求めよ.
(4) 確率 $\mathsf{Pr}\left(0 < X \leq \frac{1}{2}, Y = 0\right)$ を求めよ.
(5) 確率変数 X と Y は独立か従属かを判定せよ.

演習問題 1.5. 大小 2 つのサイコロを投げたとき, それぞれの出る目を X と Y とおく. これらを確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されている確率変数と考える. この空間上の確率変数 Z と W を

$$Z(\omega) = \min\{X(\omega), Y(\omega)\}, \quad W(\omega) = \max\{X(\omega), Y(\omega)\} \quad (\forall \omega \in \Omega)$$

で定める. このとき, 以下の問いに答えよ.
(1) 確率ベクトル (Z, W) の同時 p.m.f. $\mathsf{p}^{(Z,W)}$ を求めよ.
(2) 確率 $\mathsf{Pr}(Z = W)$ と条件付き確率 $\mathsf{Pr}(Z = 2 | W < 5)$ を求めよ.
(3) 確率変数 Z の周辺 p.m.f. p^Z を求めよ.
(4) 確率変数 W の周辺 p.m.f. p^W を求めよ.
(5) 確率変数 Z と W は独立であるかどうかを判定せよ.

演習問題 1.6. 標準正規分布に独立に従う 2 つの確率変数 X と Y に対して

$$Z = \min\{X, Y\}, \qquad W = \begin{cases} 1 & (Z = X) \\ 0 & (Z = Y) \end{cases}$$

と定める. また, 標準正規分布の p.d.f と c.d.f. を p と F と書くことにする.
(1) $z \in \mathbb{R}$, $w = 0, 1$ に対して

$$\mathsf{Pr}(Z \leq z, W = w)$$

を求めよ.
(2) Z の周辺 c.d.f. を求めよ.
(3) 事象 $\{Z \leq z\}$ が与えらたときの $W = 1$ となる条件付き確率

$$\mathsf{Pr}(W = 1 \mid Z \leq z)$$

を求めよ.
(4) Z と W は独立かどうかを示せ.

演習問題 1.7. p を自然数とする. 連続型確率変数 T は p.d.f.

$$\mathsf{p}^T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)} \frac{1}{\sqrt{p\pi}} \frac{1}{(1+t^2/p)^{(p+1)/2}}, \qquad -\infty < t < \infty$$

を持つとする. ただし, $\Gamma(\cdot)$ の Euler のガンマ関数で

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, \mathrm{d}x \qquad (a > 0)$$

で定義する.
(1) $\displaystyle\int_{-\infty}^{\infty} \mathsf{p}^T(t)\, \mathrm{d}t$ を計算せよ.
(2) 連続型確率変数 U と V は独立で, U は自由度 p の χ^2 分布 χ^2_p に従い, V は $\mathsf{N}(0, 1)$ に従うとする.

$$T = \frac{V}{\sqrt{\dfrac{U}{p}}}$$

とおいたとき, T の p.d.f. p^T を求めよ.

ヒント 確率ベクトル (U, V) の同時 p.d.f. $\mathsf{p}^{(U,V)}$ は

$$\mathsf{p}^{(U,V)}(u, v) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} e^{-v^2/2} \dfrac{1}{\Gamma(p/2) 2^{p/2}} u^{(p/2)-1} e^{-u/2} & \left(\begin{array}{l} -\infty < v < \infty; \\ 0 < u < \infty \end{array} \right) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. さらに

$$T = \frac{V}{\sqrt{\dfrac{U}{p}}}; \quad W = U$$

として, (1.18) を利用して, 同時 p.d.f. $\mathsf{p}^{(T,W)}$ を求め

$$\mathsf{p}^T(t) = \int_0^\infty \mathsf{p}^{(T,W)}(t, w)\, \mathrm{d}w$$

を計算すればよい.

演習問題 1.8. 確率変数 U_1, U_2 は開区間 $(0, 1)$ 上の一様分布に互いに独立に従うとし

$$X = \sqrt{-2 \log U_1} \cos(2\pi U_2),$$
$$Y = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

とする．さらに

$$R = \sqrt{X^2 + Y^2},$$
$$W = 2\pi U_2$$

とする．
(1) R の周辺 p.d.f. $\mathsf{p}^R(r)$ を求めよ．
(2) (R, W) の同時 p.d.f. $\mathsf{p}^{(R,W)}(r,\, w)$ を求めよ．
(3) $\mathbb{R}^2$ の部分集合 $\{(x, y) \in \mathbb{R}^2 \mid x \neq 0\}$ から $\mathbb{R}^2$ への写像を

$$\begin{pmatrix} r \\ w \end{pmatrix} = \begin{pmatrix} \sqrt{x^2 + y^2} \\ \tan^{-1}(\frac{y}{x}) \end{pmatrix}$$

で定義する．このとき, Jacobi 行列式 $\dfrac{\partial(r,\, w)}{\partial(x,\, y)}$ を求めよ．ただし, $\tan^{-1}$ は tan の逆関数である．
(4) (X, Y) の同時 p.d.f. $\mathsf{p}^{(X,Y)}(x,\, y)$ を求めよ．

第2章　期待値の基礎事項

この章では, 確率変数の期待値と関連するものを導入し, それらの基本的な性質を述べる. 節 2.1 では, 確率変数に対する期待値を定義する. 節 2.2 では, 期待値を確率ベクトルに対して定義する. 節 2.3 では, 分布のばらつきを測る量である分散と共分散を導入し, 基本的な性質を説明する. 節 2.4 では, 条件付き分布に対する期待値を導入する. 節 2.5 では, 分布を特定する量である積率母関数を導入し, その基本的な性質を説明する.

2.1　期待値

X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とする. X が離散型のとき, その p.m.f. を p とし, X の取り得る値を $x_1, x_2, \ldots$ とする. 連続型のとき, その p.d.f. も p と書くことにする.

定義 2.1. $g : \mathbb{R} \to \mathbb{R}$ を可測[1] 関数とする. 確率変数 $g(X)$ の 期待値 $\mathsf{E}[g(X)]$ を次のように定義する.

(1) $g(x) \geq 0\,(x \in \mathbb{R})$ のとき

$$\mathsf{E}[g(X)] = \begin{cases} \displaystyle\sum_{n=1}^{\infty} g(x_n)\mathsf{p}(x_n) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\mathsf{p}(x)\,\mathrm{d}x & \text{(連続型の場合)} \end{cases}$$

と定義する. 右辺は ∞ を許せば, 必ず存在する.

(2) 一般の可測関数 g に対して

$$g^+(x) = \max\{g(x), 0\}, \quad g^-(x) = \max\{-g(x), 0\}$$

と定義すれば, $g^+(x) \geq 0$, $g^-(x) \geq 0\,(x \in \mathbb{R})$ となる. $\mathsf{E}[g^+(X)]$ または $\mathsf{E}[g^-(X)]$ のいずれかが有限ならば

$$\mathsf{E}[g(X)] := \mathsf{E}[g^+(X)] - \mathsf{E}[g^-(X)]$$

[1]関数 $f : \mathbb{R} \longrightarrow \mathbb{R}$ は可測であるとは, 任意の $r \in \mathbb{R}$ に対して, $\{x \in \mathbb{R};\, f(x) > r\} \in \mathcal{B}(\mathbb{R})$ であるときをいう.

と定義する．$\mathsf{E}[g^+(X)] = \mathsf{E}[g^-(X)] = \infty$ のときは，$g(X)$ の期待値は定義されない．$\mathsf{E}[g^+(X)] < \infty$ かつ $\mathsf{E}[g^-(X)] < \infty$ のとき，$\mathsf{E}[g(X)]$ は有限となる．

補題 2.2. X を確率変数とする．関数 $g, h : \mathbb{R} \to \mathbb{R}$ を可測とする．

(1) $h(X)$ は有限の期待値を持つ[2]とする．$0 \le g(x) \le h(x)\,(\forall x \in \mathbb{R})$ ならば

$$\mathsf{E}[g(X)] \le \mathsf{E}[h(X)]$$

となる．

(2) $\mathsf{E}[|g(X)|] < \infty$, $\mathsf{E}[|h(X)|] < \infty$ を仮定する．$a, b \in \mathbb{R}$ に対して

$$\mathsf{E}[ag(X) + bh(X)] = a\,\mathsf{E}[g(X)] + b\,\mathsf{E}[h(X)]$$

となる．

Proof. 連続型確率変数の場合について証明を与えておく．離散型の場合には，積分記号を和の記号に書き換えればよい．
(1) の証明: $h(x) - g(x) \ge 0\,(x \in \mathbb{R})$ であるので，定義 2.1(1) と積分の性質から

$$\begin{aligned}\mathsf{E}[h(X)] &= \int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \{g(x) + (h(x) - g(x))\}\mathsf{p}^X(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x)\mathsf{p}^X(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} \underbrace{\big\{h(x) - g(x)\big\}\mathsf{p}^X(x)}_{\ge 0}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} g(x)\mathsf{p}^X(x)\,\mathrm{d}x = \mathsf{E}[g(X)]\end{aligned}$$

となる．よって，(1) は示せた．
(2) の証明: 確率変数 X が p.d.f. p^X を持つ場合を示す．$\big|ag(x)+bh(x)\big| \le |a||g(x)| + |b||h(x)|$ と $g(X)$ と $h(X)$ の期待値が有限なので，(1) から $ag(X) + bh(X)$ は有限な期待値をもつ．このことから，定義 2.1 のステップは省略[3]できる．あとは，積分の線型性から

$$\begin{aligned}\mathsf{E}\big[ag(X) + bh(X)\big] &= \int_{-\infty}^{\infty} \big\{ag(x) + bh(x)\big\}\mathsf{p}^X(x)\,\mathrm{d}x \\ &= a\int_{-\infty}^{\infty} g(x)\mathsf{p}^X(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}x\end{aligned}$$

[2] g, h を共に非負値関数としているので，$\mathsf{E}[|h(X)|] < \infty$ だけを仮定すればよい．もし，g, h を可測関数とすると $\mathsf{E}[|g(X)|] < \infty$ の仮定も必要になる． なぜならば，$g(x) \le h(x)\,(x \in \mathbb{R})$ と $\mathsf{E}[|h(X)|] < \infty$ だけだと $\mathsf{E}[h(X)]$ が定義できない場合があるからである．なので，非負値関数に関する主張とした方が実用的である．

[3] 関数を非負と非正の部分に分けて，それぞれが可積であることは自明になるので，これらのステップを飛ばして，積分を考えてよいことになる．

$$= a\mathsf{E}[g(X)] + b\mathsf{E}[h(X)]$$

がわかる. □

補題 2.3. X を非負値確率変数とする. このとき

$$\mathsf{E}[X] = 0 \Rightarrow \mathsf{Pr}(X = 0) = 1$$

となる.

Proof. まず

$$A := \big\{\omega \in \Omega;\, X(\omega) > 0\big\} = \{X > 0\},$$
$$A_n := \left\{\omega \in \Omega;\, X(\omega) > \frac{1}{n}\right\} \qquad (n = 1,\, 2, \ldots)$$

とおく. すると $A_1 \subset A_2 \subset \cdots$ かつ $\bigcup_{n=1}^{\infty} A_n = A$ となる. よって, 補題 1.7(5) から

$$\lim_{n\to\infty} \mathsf{Pr}(A_n) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathsf{Pr}(A) \tag{2.1}$$

となる. 一方, $A_n \subset A$ なので, $0 \le \mathbb{1}_{A_n}(\omega) \le \mathbb{1}_A(\omega)\,(\omega \in \Omega)$ で [4] ある. このことに注意して, 補題 2.2(1) を用いると

$$0 = \mathsf{E}[X] = \mathsf{E}\big[\underbrace{X\mathbb{1}_{A^c}}_{\ge 0} + X\mathbb{1}_A\big] \ge \mathsf{E}\big[X\mathbb{1}_A\big] \ge \mathsf{E}\big[X\mathbb{1}_{A_n}\big]$$
$$\ge \frac{1}{n}\mathsf{E}[\mathbb{1}_{A_n}] = \frac{1}{n}\mathsf{Pr}(A_n)$$

を得る [5]. したがって, $\mathsf{Pr}(A_n) = 0\,(\forall n \in \mathbb{N})$ がわかる. (2.1) と合わせると $\mathsf{Pr}(A) = 0$ となる. よって

$$1 = \mathsf{Pr}(A^c) = \mathsf{Pr}\Big(\big(\{X \le 0\} \cap \{X \ge 0\}\big) \cup \big(\{X \le 0\} \cap \{X < 0\}\big)\Big)$$
$$= \mathsf{Pr}(X = 0) + \underbrace{\mathsf{Pr}(X < 0)}_{=0} = \mathsf{Pr}(X = 0)$$

が示せた. □

---

[4] $\mathbb{1}_A$ は $\mathbb{1}_A : \Omega \ni \omega \mapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & (\omega \in A) \\ 0 & (\omega \in A^c) \end{cases}$ なる確率変数である.

[5]

補題 2.4. $0 < q < r$ に対して

$$\mathsf{E}[|X|^r] < \infty \Rightarrow \mathsf{E}[|X|^q] < \infty$$

となる.

Proof. Young の不等式 (系 3.13)

$$ab \leq \frac{a^s}{s} + \frac{b^t}{t}; \quad a, b > 0; \quad s, t > 0; \quad \frac{1}{t} + \frac{1}{s} = 1$$

において

$$a = |x|^q; s = \frac{r}{q} > 1, \quad b = 1; t = \frac{r}{r-q} = \frac{1}{1 - 1/s} > 1$$

とおくと

$$|x|^q \leq \frac{|x|^r}{s} + \frac{1}{t}$$

を得る. よって, 補題 2.2(1) からわかる. □

問 2.1. 連続確率変数 X が p.d.f. p をもつとき, 補題 2.4 の証明を具体的に書け. $q < r$ なので, $|x|^{q-r} \leq 1\,(|x| > 1)$ となることを使えばよい.

定理 2.5. $t > 0$ とする.

$$\mathsf{E}[|X|^t] < \infty$$

ならば

$$\lim_{n\to\infty} n^t \,\mathsf{Pr}(|X| > n) = 0$$

となる.

Proof. 連続型の場合について証明を与えておく. X の p.d.f. を p と書いたとき

$$\infty > \int_{-\infty}^{\infty} |x|^t \mathsf{p}(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{-n}^{n} |x|^t \mathsf{p}(x)\,\mathrm{d}x$$

から

$$\lim_{n\to\infty} \int_{|x|>n} |x|^t \mathsf{p}(x)\,\mathrm{d}x = 0$$

となる. しかし

$$0 = \lim_{n\to\infty} \int_{|x|>n} |x|^t \mathsf{p}(x)\,\mathrm{d}x \geq \lim_{n\to\infty} n^t \int_{|x|>n} \mathsf{p}(x)\,\mathrm{d}x = \lim_{n\to\infty} n^t \,\mathsf{Pr}(|X| > n)$$

がわかる. □

定理 2.6. X を確率変数とし, ある $t > 0$ に対して

$$\lim_{n\to\infty} n^t \,\mathsf{Pr}(|X| > n) = 0$$

をみたすとする. このとき, $0 < \forall s < t$ に対して

$$\mathsf{E}[|X|^s] < \infty$$

となる.

注意 2.7. 定理 2.6 は $\mathsf{E}[|X|^t] < \infty$ までは主張していないことに注意せよ. □

この定理の証明のために, 以下の補題と系を準備する.

補題 2.8. X を非負値確率変数で分布関数 F をもつとする. このとき, $\mathsf{E}[X] < \infty$ または $\int_0^\infty \{1 - \mathsf{F}(x)\}\,\mathrm{d}x < \infty$ のいずれかが成り立つならば, もう片方も有限で

$$\mathsf{E}[X] = \int_0^\infty \{1 - \mathsf{F}(x)\}\,\mathrm{d}x \tag{2.2}$$

が成り立つ.

Proof. X は連続型確率変数で p.d.f. p をもつ場合のみの証明を与えておく. いま, 分布関数を F と記すことにする.

① $\mathsf{E}[X] < \infty \Rightarrow \int_0^\infty \{1 - \mathsf{F}(x)\}\,\mathrm{d}x < \infty$ の証明: まず

$$\mathsf{E}[X] = \int_0^\infty x\mathsf{p}(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_0^n x\mathsf{p}(x)\,\mathrm{d}x$$

に注意する. 上式の右辺を部分積分を用いて書き換えると

$$\begin{aligned}
\int_0^n x\mathsf{p}(x)\,\mathrm{d}x &= \big\{x\mathsf{F}(x)\big\}\Big|_{x=0}^{x=n} - \int_0^n \mathsf{F}(x)\,\mathrm{d}x \\
&= n\,\mathsf{F}(n) - 0 \cdot \mathsf{F}(0) - \int_0^n \mathsf{F}(x)\,\mathrm{d}x \\
&= n\,\mathsf{F}(n) - n + \underbrace{n}_{=\int_0^n \mathrm{d}x} - \int_0^n \mathsf{F}(x)\,\mathrm{d}x \\
&= -n\big[1 - \mathsf{F}(n)\big] + \int_0^n \big[1 - \mathsf{F}(x)\big]\,\mathrm{d}x \\
&= -n\,\mathsf{Pr}(X > n) + \int_0^n \big[1 - \mathsf{F}(x)\big]\,\mathrm{d}x
\end{aligned}$$

となる. よって

$$\mathsf{E}[X] = \lim_{n\to\infty}\left(-n\mathsf{Pr}(X>n) + \int_0^n \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x\right) \tag{2.3}$$

$$\overset{\because\text{定理 2.5}}{=} 0 + \lim_{n\to\infty}\int_0^n \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x \tag{2.4}$$

$$= \int_0^\infty \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x \tag{2.5}$$

がわかる. よって

$$\mathsf{E}[X]<\infty \;\Rightarrow\; \int_0^n \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x<\infty$$

となり, (2.2) が成り立つことがわかる.
② $\int_0^\infty \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x<\infty \;\Rightarrow\; \mathsf{E}[X]<\infty$ の証明: $\forall n\in\mathbb{N}$ に対して, 部分積分を用いると

$$\int_0^n |x|\mathsf{p}(x)\,\mathrm{d}x = \int_0^n x\mathsf{p}(x)\,\mathrm{d}x = -n\big[1-\mathsf{F}(n)\big] + \int_0^n \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x$$

となることに注意する. $-n\big[1-\mathsf{F}(n)\big]\le 0$ なので

$$\int_0^n |x|\mathsf{p}(x)\,\mathrm{d}x \le \int_0^n \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x \le \int_0^\infty \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x \qquad (\forall n\in\mathbb{N})$$

を得る. したがって

$$\lim_{n\to\infty}\int_0^n |x|\mathsf{p}(x)\,\mathrm{d}x = \int_0^\infty |x|\mathsf{p}(x)\,\mathrm{d}x \le \int_0^\infty \big[1-\mathsf{F}(x)\big]\,\mathrm{d}x<\infty$$

から $\mathsf{E}[X]<\infty$ がわかる.
③ (2.2) の証明: さらに, (2.5) から (2.2) もわかる. □

系 2.9. $s>0$ とする. $\mathsf{E}[|X|^s]<\infty$ または $s\int_0^\infty y^{s-1}\mathsf{Pr}\big(|X|>y\big)\,\mathrm{d}y<\infty$ のいずれかが成り立つとき, もう片方も有限で

$$\mathsf{E}[|X|^s] = s\int_0^\infty y^{s-1}\mathsf{Pr}\big(|X|>y\big)\,\mathrm{d}y$$

が成り立つ.

Proof. $|X|^s$ の分布関数を $\mathsf{F}^{|X|^s}$ と記す. 補題 2.8 から

$$\mathsf{E}[|X|^s] \overset{\because\text{補題 2.8}}{=} \int_0^\infty \big[1-\mathsf{F}^{|X|^s}(z)\big]\,\mathrm{d}z = \int_0^\infty \mathsf{Pr}\big(|X|^s>z\big)\,\mathrm{d}z$$

となる. ここで $z = y^s$ と変換すると $\dfrac{\mathrm{d}z}{\mathrm{d}y} = sy^{s-1}$ と $\mathrm{d}z = sy^{s-1}\mathrm{d}y$ となる. したがって

$$\begin{aligned}\int_0^\infty \mathsf{Pr}\big(|X|^s > z\big)\,\mathrm{d}z &= \int_0^\infty \mathsf{Pr}\big(|X|^s > y^s\big)\,sy^{s-1}\,\mathrm{d}y \\ &= s\int_0^\infty y^{s-1}\mathsf{Pr}\big(|X|^s > y^s\big)\,\mathrm{d}y \\ &= s\int_0^\infty y^{s-1}\mathsf{Pr}\big(|X| > y\big)\,\mathrm{d}y\end{aligned}$$

となることがわかる. よって, 系 2.9 から, $\mathsf{E}[|X|^s] < \infty$ となる. □

定理 2.6 の証明: 任意の $\epsilon > 0$ に対して, N をうまく選ぶと

$$\mathsf{Pr}\big(|X| > n\big) < \frac{\epsilon}{n^t} \qquad (\forall n \geq N) \tag{2.6}$$

が成立する. さらに, $c < -1$ のとき

$$\int_N^\infty y^c\,\mathrm{d}y = \left[\frac{1}{c+1}y^{c+1}\right]_N^\infty = -\frac{1}{c+1}N^{c+1} < \infty \tag{2.7}$$

となること注意をする. すると

$$\begin{aligned}&s\int_0^\infty y^{s-1}\mathsf{Pr}\big(|X| > y\big)\,\mathrm{d}y \\ &\quad= s\int_0^N y^{s-1}\underbrace{\mathsf{Pr}\big(|X| > y\big)}_{\leq 1}\,\mathrm{d}y + s\int_N^\infty y^{s-1}\underbrace{\mathsf{Pr}\big(|X| > y\big)}_{<\epsilon/y^t \because (2.6)}\,\mathrm{d}y \\ &\quad\leq \int_0^N sy^{s-1}\,\mathrm{d}y + s\epsilon\int_N^\infty y^{s-1}\frac{1}{y^t}\,\mathrm{d}y \\ &\quad= N^s + s\epsilon\int_N^\infty y^{s-1-t}\,\mathrm{d}y \\ &\quad< \infty\end{aligned}$$

がわかる. 最後の不等号は $s - 1 - t < -1$ と (2.7) からわかる. よって, $0 < \forall s < t$ に対して

$$\mathsf{E}[|X|^s] < \infty$$

がわかる. □

定義 2.10. (1) $k = 1, 2, \ldots$ に対して, $\mathsf{E}[|X|^k] < \infty$ のとき, $\mathsf{E}[X^k]$ を X の k **次モーメント** または k **次の積率** という.

(2) $\mathsf{E}[|X|] < \infty$ のとき $\mathsf{E}[X]$ を X の平均値[6] という.

(3) $\mathsf{E}[X^2] < \infty$ のとき X の分散 を

$$\mathsf{Var}[X] := \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

で定義する.

(4) 部分集合 $A \subset \mathbb{R}$ に対して

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

を A の指示関数[7] という. $A \in \mathcal{B}(\mathbb{R})$ のとき

$$\mathsf{E}[\mathbb{1}_A(X)] = \mathsf{Pr}(X \in A)$$

となる.

注意 2.11. 分散 $\mathsf{Var}[X]$ は X の分布の平均 μ まわりの散らばりを測る量である. 分散がおおきいほど分布は広がっていることになる. □

注意 2.12. (1) X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の離散型確率変数とする. p^X を X の p.m.f. とし

$$S_X := \{x \in \mathbb{R};\, \mathsf{p}^X(x) > 0\}$$

とする. 関数 $g : \mathbb{R} \to \mathbb{R}$ に対して

$$Y = g(X)$$

とおく. $y \in \mathbb{R}$ に対して

$$A_y := \{x \in S_X;\, g(x) = y\}$$

とおき

$$S_Y := \{y \in \mathbb{R};\, A_y \neq \varnothing\}$$

とする. $y \in S_Y$ に対して

$$\mathsf{p}^Y(y) := \mathsf{Pr}(Y = y) = \mathsf{Pr}(X \in A_y) = \sum_{x \in A_y} \mathsf{Pr}(X = x) = \sum_{x \in A_y} \mathsf{p}^X(x)$$

[6] 簡単に「平均」ともいう.
[7] 指示関数は $\mathbb{R}$ の任意の部分集合に定義ができることに注意をせよ.

と書ける. 正項級数は項の順番を並び替えてもその値は変わらないので

$$\sum_{y\in S_Y}|y|\mathsf{p}^Y(y)=\sum_{y\in S_Y}|y|\sum_{x\in A_y}\mathsf{p}^X(x)=\sum_{y\in S_Y}\sum_{x\in A_y}|g(x)|\mathsf{p}^X(x)$$
$$=\sum_{x\in S_X}|g(x)|\mathsf{p}^X(x)$$

となる. さらにいずれかの和が有限ならば

$$\sum_{y\in S_Y}y\mathsf{p}^Y(y)=\sum_{y\in S_Y}y\sum_{x\in A_y}\mathsf{p}^X(x)=\sum_{y\in S_Y}\sum_{x\in A_y}g(x)\mathsf{p}^X(x)=\sum_{x\in S_X}g(x)\mathsf{p}^X(x)$$

となる. よって

$$\mathsf{E}[Y]=\mathsf{E}[g(X)]$$

となる.

(2) 確率変数 X は連続型とする. p^X を X の p.d.f. とし

$$S_X:=\{x\in\mathbb{R};\,\mathsf{p}^X(x)>0\}$$

とする. 関数 $g\colon S_X\to\mathbb{R}$ は狭義単調増加かつ C^1 級とする. このとき

$$Y=g(X),\quad S_Y:=\{y\in\mathbb{R};\,y=g(x)\,(\exists x\in S_X)\}$$

とする. ここで $g\colon S_X\to S_Y$ と制限[8]すれば g の逆関数 $g^{-1}\colon S_Y\to S_X$ が存在し

$$\mathsf{F}^Y(y):=\mathsf{Pr}(Y\le y)=\mathsf{Pr}\big(g^{-1}(Y)\le g^{-1}(y)\big)=\mathsf{Pr}(X\le g^{-1}(y))$$
$$=\mathsf{F}^X(g^{-1}(y))$$

となる. さらに $g(g^{-1}(y))=y$ より

$$\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)=\frac{1}{\overset{\bullet}{g}\,(g^{-1}(y))}$$

となる. ただし $\overset{\bullet}{g}\,(y)=\dfrac{\mathrm{d}g}{\mathrm{d}y}(y)$ である. これらより, $y\in S_Y$ に対して

$$\mathsf{p}^Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}\mathsf{F}^X(g^{-1}(y))=\overset{\bullet}{\mathsf{F}}{}^X\,(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$$
$$=\mathsf{p}^X(g^{-1}(y))\frac{1}{\overset{\bullet}{g}\,\big(g^{-1}(y)\big)}$$

[8]制限したものを同じ g を用いて表現するのは, 記号の乱用である. 記号が煩雑になるので, 記号を乱用した.

となる. よって

$$\mathsf{p}^Y(y) = \mathsf{p}^X\big(g^{-1}(y)\big)\frac{1}{\overset{\bullet}{g}\,(g^{-1}(y))}$$

となる. g が狭義単調減少の場合もふくめると

$$\mathsf{p}^Y(y) = \mathsf{p}^X\big(g^{-1}(y)\big)\frac{1}{\big|\overset{\bullet}{g}\,\big(g^{-1}(y)\big)\big|}$$

となる.

以上のことから

$$\begin{aligned}
\mathsf{E}[Y] &= \int_{S_Y} y\mathsf{p}^Y(y)\,\mathrm{d}y \\
&= \int_{S_Y} y\mathsf{p}^X(g^{-1}(y))\frac{1}{\overset{\bullet}{g}\,\big(g^{-1}(y)\big)}\,\mathrm{d}y \\
&= \int_{S_X} g(x)\mathsf{p}^X(x)\,\mathrm{d}x \qquad \big(\because x = g^{-1}(y) \Leftrightarrow y = g(x);\ \mathrm{d}y = \overset{\bullet}{g}\,(x)\mathrm{d}x\big) \\
&= \mathsf{E}[g(X)]
\end{aligned}$$

がわかる. □

2.2　確率ベクトルの期待値

確率変数 X, Y を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数とする. (X, Y) を確率ベクトルという.

確率ベクトル (X, Y) が離散型のときその同時 p.m.f. を $\mathsf{p}(x, y)$ とし, 連続型のときその同時 p.d.f. も $\mathsf{p}(x, y)$ と書くことにする.

定義 2.13. 可測関数 $g : \mathbb{R}^2 \to \mathbb{R}$ に対して $\underline{g(X, Y) \text{ の期待値}}$ を次のように定義する.

(1) $g \geq 0$ のとき

$$\mathsf{E}[g(X, Y)] = \begin{cases} \displaystyle\sum_{x,y} g(x, y)\mathsf{p}(x, y) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\mathsf{p}(x, y)\,\mathrm{d}x\mathrm{d}y & \text{(連続型の場合)} \end{cases}$$

で定義する.

(2) 一般の g に対して

$$g^+(x, y) := \max\{g(x, y), 0\}, \quad g^-(x, y) := \max\{-g(x, y), 0\}$$

と定義すれば $g^+,\, g^- \geq 0$ となる. $\mathsf{E}[g^+(X,\, Y)]$ または $\mathsf{E}[g^-(X,\, Y)]$ のいずれかが有限ならば

$$\mathsf{E}[g(X,\, Y)] := \mathsf{E}[g^+(X,\, Y)] - \mathsf{E}[g^-(X,\, Y)]$$

で定義する. $\mathsf{E}[g^+(X,\, Y)] = \mathsf{E}[g^-(X,\, Y)] = \infty$ のときは, $g(X,\, Y)$ の期待値は定義されない. $\mathsf{E}[g^+(X,\, Y)] < \infty$ かつ $\mathsf{E}[g^-(X,\, Y)] < \infty$ のとき, $\mathsf{E}[g(X,\, Y)]$ は有限の値を取る.

注意 2.14. 3 つ以上の確率変数 $X_1,\, X_2,\, \ldots,\, X_n$ に対しても期待値を定義 2.13 と同様に定義する. □

定理 2.15. $X_1,\, X_2,\, \ldots,\, X_n$ を確率変数とし, 各 $X_j\,(j = 1,\, 2,\, \ldots,\, n)$ の期待値は有限とする. $a_1,\, a_2,\, \ldots,\, a_n$ を定数としたとき

$$\mathsf{E}\left[\sum_{j=1}^{n} a_j X_j\right] = \sum_{j=1}^{n} a_j \mathsf{E}[X_j]$$

となる.

Proof. まず, $\left|\sum_{j=1}^{n} a_j x_j\right| \leq \sum_{j=1}^{n} |a_j| \cdot |x_j|$ なので, 補題 2.2(1) から $\sum_{j=1}^{n} a_j X_j$ は有限な期待値をもつ. $\boldsymbol{X} := (X_1,\, X_2,\, \ldots,\, X_n)^\top$ とおき, $\boldsymbol{X}$ は同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}$ を持つとする. すると定義 2.13 から 補題 2.2(2) と同様の期待値の線型性から

$$\begin{aligned}
\mathsf{E}\left[\sum_{j=1}^{n} a_j X_j\right] &= \int_{\mathbb{R}^n} \left\{\sum_{j=1}^{n} a_j x_j\right\} \mathsf{p}^{\boldsymbol{X}}(x_1,\, x_2,\, \ldots,\, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \sum_{j=1}^{n} a_j \int_{\mathbb{R}^n} x_j \mathsf{p}^{\boldsymbol{X}}(x_1,\, x_2,\, \ldots,\, x_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \sum_{j=1}^{n} a_j \mathsf{E}\big[X_j\big]
\end{aligned}$$

がわかる. □

例 2.16. $0 < p < 1$ と $j = 1,\, 2, \ldots,\, n$ に対して, $X_j \sim \mathsf{Ber}(p)$ とする. このとき

$$\mathsf{E}[X_j] = \sum_{x=0}^{1} x \mathsf{Pr}(X = x) = 0 \times (1 - p) + 1 \times p = p$$

である. $S = \sum_{j=1}^{n} X_j$ としたとき

$$\mathsf{E}[S] = \sum_{j=1}^{n} \mathsf{E}[X_j] = np$$

がわかる. □

2.3 分散と共分散

X を確率変数とし, $\mathsf{E}[X^2] < \infty$ とする. X の分散 を

$$\mathsf{Var}[X] := \mathsf{E}[(X-\mu)^2]$$

で定義した. ただし $\mu = \mathsf{E}[X]$ と書いた. さらに, 分散の定義から $\mathsf{Var}[X] \geq 0$ となることに注意する. $\sqrt{\mathsf{Var}[X]}$ を X の標準偏差 という.

定理 2.17. 以下の確率変数は有限の 2 次の積率を持つとする. このとき, 次が成立する.

(1) $\mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2$ となる.

(2) 定数 $a, b \in \mathbb{R}$ に対して

$$\mathsf{Var}[aX+b] = a^2\mathsf{Var}[X]$$

となる.

(3) X と Y は独立で $\mathsf{E}[|XY|] < \infty$ とする. このとき

$$\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y]$$

となる.

(4) $X_1, X_2, \ldots, X_n$ は独立とし, $\mathsf{E}[X_j^2] < \infty\,(j = 1, 2, \ldots, n)$ とする. $a_1, a_2, \ldots, a_n$ は定数としたとき

$$\mathsf{Var}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1^2\,\mathsf{Var}[X_1] + a_2^2\,\mathsf{Var}[X_2] + \cdots + a_n^2\,\mathsf{Var}[X_n]$$

となる.

Proof. ① (1), (2) の証明: 分散を期待値の記号を用いて表現し, 期待値の線型性を用いて計算すればよい.
② (3) の証明: については, 連続型の場合を示す. (X, Y) の同時 p.d.f. は, X と Y の周辺 p.d.f. p^X と p^Y を用いて $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ という形でかけるので

$$\begin{aligned}\mathsf{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\mathsf{p}^X(x)\mathsf{p}^Y(y)\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} x\mathsf{p}^X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} y\mathsf{p}^Y(y)\,\mathrm{d}y \\ &= \mathsf{E}[X]\mathsf{E}[Y]\end{aligned}$$

となること[9] がわかる.
③ (4) の証明: 分散を期待値の記号を用いて表現し, (3) に注意する. 期待値の線型性を用いて計算すればよい. □

[9] $\mathsf{Var}[X] < \infty$, $\mathsf{Var}[Y] < \infty$ なので, Cauchy-Schwarz の不等式から $\mathsf{E}[|XY|] < \infty$ が確認できるので, Fubini 定理から積分順序の入れ替えが保証されることがわかる.

例 2.18. (例 2.16 の続き) 例 2.16 の設定に加えて $X_1, X_2, \ldots, X_n$ は互いに独立とする. 定理 2.15 に注意すれば

$$\begin{aligned}\mathsf{Var}[X_j] &= \mathsf{E}[X_j^2] - \{\mathsf{E}[X_j]\}^2 = \sum_{x=0}^{1} x^2 \mathsf{Pr}(X=x) - p^2 \\ &= 0 \times (1-p) + 1 \times p - p^2 = p(1-p)\end{aligned}$$

となる. これと定理 2.17(4) から

$$\mathsf{Var}[S] = \sum_{j=1}^{n} \mathsf{Var}[X_j] = np(1-p)$$

がわかる. □

定理 2.19. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n$ は i.i.d. 確率変数列とし

$$\mathsf{E}[X_1] = \mu, \quad \mathsf{Var}[X_1] = \sigma^2$$

とする. ただし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ である. $X_1, X_2, \ldots, X_n$ に基づく 標本平均 を

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$$

で定義し, 標本 (不偏) 分散 を

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} \big(X_j - \overline{X}_n\big)^2$$

で定義する. このとき

$$(1)\, \mathsf{E}\big[\overline{X}_n\big] = \mu; \qquad (2)\, \mathsf{Var}\big[\overline{X}_n\big] = \frac{\sigma^2}{n}; \qquad (3)\, \mathsf{E}\big[S_n^2\big] = \sigma^2$$

となる.

Proof. (1), (2) は定理 2.15, 2.17(4) よりわかる. (3) を証明するために

$$\sum_{j=1}^{n} \big(X_j - \overline{X}_n\big)^2 = \sum_{j=1}^{n} (X_j - \mu)^2 - n\big(\overline{X}_n - \mu\big)^2$$

に注意する. 定理 2.17(1) より

$$\mathsf{E}\left[\sum_{j=1}^{n} (X_j - \mu)^2\right] = n\sigma^2$$

がわかる. さらに, (1) と (2) より

$$\mathsf{E}\big[\big(\overline{X}_n-\mu\big)^2\big]=\mathsf{Var}\big[\overline{X}_n\big]=\frac{\sigma^2}{n}$$

が示される. したがって

$$\begin{aligned}\mathsf{E}[S_n^2]&=\frac{1}{n-1}\Big\{\sum_{j=1}^n\mathsf{E}\big[(X_j-\mu)^2\big]-n\mathsf{E}\big[\big(\overline{X}_n-\mu\big)^2\big]\Big\}\\&=\frac{1}{n-1}\big\{n\sigma^2-\sigma^2\big\}=\sigma^2\end{aligned}$$

がわかる. □

定義 2.20. X と Y は確率変数とし

$$\mathsf{E}[X]=\mu_X,\quad \mathsf{Var}[X]=\sigma_X^2,\quad \mathsf{E}[Y]=\mu_Y,\quad \mathsf{Var}[Y]=\sigma_Y^2$$

とする. ただし $\mu_X, \mu_Y \in \mathbb{R}, 0 < \sigma_X, \sigma_Y < \infty$ とする. このとき X と Y の共分散 を

$$\mathsf{Cov}[X,\,Y]=\mathsf{E}\big[(X-\mu_X)(Y-\mu_Y)\big]$$

で定義し, X と Y の (Pearson) の相関係数 を

$$\rho:=\rho[X,\,Y]:=\frac{\mathsf{Cov}[X,\,Y]}{\sigma_X\,\sigma_Y}$$

で定義する.

定理 2.21. X, Y, Z は 2 次の積率が有限な確率変数とする.
(1) 共分散は

$$\mathsf{Cov}[X,\,Y]=\mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y]$$

と書き直せる.
(2) $\mathsf{Cov}[X,\,Y]=\mathsf{Cov}[Y,\,X]$ である.
(3) 定数 a, b に対して

$$\mathsf{Cov}[aX+bY,\,Z]=a\mathsf{Cov}[X,\,Z]+b\mathsf{Cov}[Y,\,Z]$$

となる.
(4) 相関係数は

$$-1\le\rho[X,\,Y]\le 1$$

をみたす.
(5) ある定数 $a, b\,(a\neq 0)$ が存在して $Y=aX+b$ となったとき

$$a>0\Rightarrow\rho[X,\,Y]=1,$$

$$a < 0 \Rightarrow \rho[X, Y] = -1$$

である.
(6) X と Y が独立のとき

$$\mathsf{Cov}[X, Y] = 0$$

となる.

Proof. ① (1) の証明: 期待値の線型性からわかる.
② (2), (3) の証明: (2)(3) は共分散の定義と期待値の線型性よりわかる.
③ (4) の証明: $\mathsf{Var}[X] = \mathsf{Var}[Y] = 0$ の場合は $\mathsf{Cov}[X, Y] = 0$ で自明である. よって $\mathsf{Var}[X] > 0$ として証明する. 補題 2.2(2) より

$$\mathsf{Var}[X]t^2 - 2\mathsf{Cov}[X, Y]t + \mathsf{Var}[Y] = \mathsf{E}[\big\{t(X - \mathsf{E}[X]\big) - \big(Y - \mathsf{E}[Y]\big)\big\}^2] \geq 0$$

となる. ここで, 判別式をとれば

$$\{\mathsf{Cov}[X, Y]\}^2 - \mathsf{Var}[X]\mathsf{Var}[Y] \leq 0 \tag{2.8}$$

がわかる.
④ (5) の証明: (2), (3) および定理 2.17(2) からわかる.
⑤ (6) の証明: 定理 2.17(3) からわかる. □

注意 2.22. (2.8) において, $\mathsf{E}[X] = \mathsf{E}[Y] = 0$ の場合を考えると Cauchy-Schwarz の不等式

$$\big|\mathsf{E}[XY]\big| \leq \sqrt{\mathsf{E}[X^2]}\sqrt{\mathsf{E}[Y^2]}$$

と呼ばれる重要なものを得る.

注意 2.23. (6) の逆は一般に正しくない. 反例は下の問いをみよ. □

問 2.2. Z を $[0, 2\pi]$ 上の一様分布とし

$$X = \cos Z, \quad Y = \sin Z$$

とおく.

(1) $\mathsf{E}[X] = 0, \mathsf{E}[Y] = 0, \mathsf{E}[XY] = 0$ を確かめよ.

(2) $\mathsf{Pr}(X \leq 1/2, Y \leq 1/2) = \mathsf{Pr}((\pi/3) \leq Z \leq (5/6)\pi)$ を確かめよ.

(3) $\mathsf{Pr}(X \leq 1/2) = \mathsf{Pr}((\pi/3) \leq Z \leq (7/3)\pi)$ を確かめよ [10].

[10] これらより, $\mathsf{Pr}(X \leq 1/2, Y \leq 1/2) \neq \mathsf{Pr}(X \leq 1/2)\mathsf{Pr}(Y \leq 1/2)$ となり, $\mathsf{Cov}[X, Y] = 0$ だが, X と Y は従属であることがわかる.

定理 2.24. (1) X, Y は有限な 2 次の積率を持つ確率変数とする. このとき $\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X, Y]$ となる.
(2) $d \geq 2$ とする. $X_1, X_2, \ldots, X_d$ は有限な 2 次の積率を持つ確率変数とする. このとき

$$\mathsf{Var}\left[\sum_{j=1}^{d} a_j X_j\right] = \sum_{j=1}^{d} a_j^2 \mathsf{Var}[X_j] + 2\sum_{j=1}^{d}\sum_{\ell=j+1}^{d} a_j a_\ell \mathsf{Cov}[X_j, X_\ell]$$

となる.

Proof. 共分散を期待値で表現し, 展開して期待値の線型性を用いればよい. □

問 2.3. 定理 2.24(1) の証明を具体的に書け.

定義 2.25. $X_1, X_2, \ldots, X_d$ を有限な 2 次の積率を持つ確率変数とし

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}$$

と書く. このとき確率ベクトル $\boldsymbol{X}$ の期待値を

$$\mathsf{E}[\boldsymbol{X}] := \begin{pmatrix} \mathsf{E}[X_1] \\ \mathsf{E}[X_2] \\ \vdots \\ \mathsf{E}[X_d] \end{pmatrix}$$

で定義する. $\boldsymbol{X}$ の共分散を

$$\mathsf{Var}[\boldsymbol{X}] := \mathsf{E}\big[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\big]$$

で定義する[11]. ただし $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)^\top := \mathsf{E}[\boldsymbol{X}]$ である. これは

$$\mathsf{Var}[\boldsymbol{X}] = \begin{pmatrix} \mathsf{Var}[X_1] & \mathsf{Cov}[X_1, X_2] & \cdots & \mathsf{Cov}[X_1, X_d] \\ \mathsf{Cov}[X_2, X_1] & \mathsf{Var}[X_2] & \cdots & \mathsf{Cov}[X_2, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_d, X_1] & \mathsf{Cov}[X_d, X_2] & \cdots & \mathsf{Var}[X_d] \end{pmatrix}$$

である.

[11]確率ベクトルと同様に確率変数を成分とする行列を確率行列という. 確率行列の期待値はそれぞれの成分の期待値を取ったものを配置した行列と定義している.

注意 2.26. 定義から $\mathsf{Var}[\boldsymbol{X}]$ は半正定値対称行列となる. なぜならば, 期待値の線型性と補題 2.2(2) から, 任意の $\boldsymbol{a} \in \mathbb{R}^d$ に対して

$$\begin{aligned}\boldsymbol{a}^\top \mathsf{Var}[\boldsymbol{X}]\boldsymbol{a} &= \mathsf{E}\big[\boldsymbol{a}^\top(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{a}\big] \\ &= \mathsf{E}\big[\big\{\boldsymbol{a}^\top(\boldsymbol{X}-\boldsymbol{\mu})\big\}^2\big] \geq 0\end{aligned}$$

となる. 最後の不等号は補題 2.2(2) からわかる. よって $\mathsf{Var}[\boldsymbol{X}]$ は半正定値であることが示せた. □

補題 2.27. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ は確率ベクトルで各成分は有限な 2 次の積率を持つとし

$$\mathsf{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{X}] = \boldsymbol{\Sigma}$$

とする. ただし $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ は $d \times d$ の半正定値行列[12] である.
(1) 任意の定数ベクトル $\boldsymbol{a} \in \mathbb{R}^d$ に対して

$$\mathsf{E}[\boldsymbol{a}^\top\boldsymbol{X}] = \boldsymbol{a}^\top\boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{a}^\top\boldsymbol{X}] = \boldsymbol{a}^\top\boldsymbol{\Sigma}\boldsymbol{a}$$

となる.
(2) $k \in \mathbb{N}$ とする. 任意の定数の $k \times d$ 行列 $\mathbf{A}$ に対して

$$\mathsf{E}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{E}[\boldsymbol{X}], \quad \mathsf{Var}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{Var}[\boldsymbol{X}]\mathbf{A}^\top$$

となる.

Proof. (1) の証明: 期待値の線型性を用いて計算すればよい.
(2) の証明:

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_k^\top \end{bmatrix}; \qquad \boldsymbol{a}_j \in \mathbb{R}^d\,(j = 1, 2, \ldots, k)$$

とおくと

$$\mathsf{E}\big[\mathbf{A}\boldsymbol{X}\big] = \mathsf{E}\begin{bmatrix} \boldsymbol{a}_1^\top\boldsymbol{X} \\ \boldsymbol{a}_2^\top\boldsymbol{X} \\ \vdots \\ \boldsymbol{a}_k^\top\boldsymbol{X} \end{bmatrix} = \begin{bmatrix} \mathsf{E}\big[\boldsymbol{a}_1^\top\boldsymbol{X}\big] \\ \mathsf{E}\big[\boldsymbol{a}_2^\top\boldsymbol{X}\big] \\ \vdots \\ \mathsf{E}\big[\boldsymbol{a}_k^\top\boldsymbol{X}\big] \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^\top\boldsymbol{\mu} \\ \boldsymbol{a}_2^\top\boldsymbol{\mu} \\ \vdots \\ \boldsymbol{a}_k^\top\boldsymbol{\mu} \end{bmatrix} = \mathbf{A}\boldsymbol{\mu}$$

[12] $d \times d$ の対称行列 $\mathbf{A}$ が半正定値であるとは, $\forall \boldsymbol{a} \in \mathbb{R}^d$ に対して $\boldsymbol{a}^\top\mathbf{A}\boldsymbol{a} \geq 0$ が成立するときをいう. また $d \times d$ の対称行列 $\mathbf{A}$ が正定値であるとは, $\forall \boldsymbol{a} \in \mathbb{R}^d\,(\boldsymbol{a} \neq \mathbf{0})$ に対して $\boldsymbol{a}^\top\mathbf{A}\boldsymbol{a} > 0$ が成立するときをいう.

がわかる. つぎに, 共分散の定義と上の結果から

$$\begin{aligned}\mathsf{Var}\big[\mathbf{A}\boldsymbol{X}\big] &= \mathsf{E}\big[\big\{\mathbf{A}\boldsymbol{X} - \mathsf{E}\big[\mathbf{A}\boldsymbol{X}\big]\big\}\big\{\mathbf{A}\boldsymbol{X} - \mathsf{E}\big[\mathbf{A}\boldsymbol{X}\big]\big\}^\top\big] \\ &= \mathsf{E}\big[\big\{\mathbf{A}\boldsymbol{X} - \mathbf{A}\mathsf{E}\big[\boldsymbol{X}\big]\big\}\big\{\mathbf{A}\boldsymbol{X} - \mathbf{A}\mathsf{E}\big[\boldsymbol{X}\big]\big\}^\top\big] \\ &= \mathsf{E}\big[\mathbf{A}\big\{\boldsymbol{X} - \mathsf{E}\big[\boldsymbol{X}\big]\big\}\big\{\boldsymbol{X} - \mathsf{E}\big[\boldsymbol{X}\big]\big\}^\top\mathbf{A}^\top\big] \\ &= \mathbf{A}\mathsf{E}\Big[\big\{\boldsymbol{X} - \mathsf{E}\big[\boldsymbol{X}\big]\big\}\big\{\boldsymbol{X} - \mathsf{E}\big[\boldsymbol{X}\big]\big\}^\top\Big]\mathbf{A}^\top \\ &= \mathbf{A}\mathsf{Var}[\boldsymbol{X}]\mathbf{A}^\top\end{aligned}$$

からわかる. □

問 2.4. $d = 2$ として, 補題 2.27 を確認せよ.

系 2.28. 確率ベクトル $\boldsymbol{X}$ は, 任意の定数ベクトル $\boldsymbol{a}(\neq \mathbf{0})$ に対して

$$\mathsf{Pr}(\boldsymbol{a}^\top\boldsymbol{X} = 0) = 0$$

をみたす. このとき, $\boldsymbol{\Sigma}$ は正定値である.

注意 2.29. $\boldsymbol{X}$ の共分散行列 $\boldsymbol{\Sigma}$ は, 補題 2.27(1) から半正定値であることがわかる. 系 2.28 の仮定は, 確率ベクトル $\boldsymbol{X}$ が d 次元空間より次元の低い空間に集中しないことを意味している. したがって, 確率ベクトル $\boldsymbol{X}$ の値を取る空間が退化していなければ, その共分散行列は正定値となることがわかる. □

2.4 条件付き期待値

定義 2.30. (1) X と Y を確率変数とし, 条件付き p.d.f.(または条件付き p.m.f.) を $\mathsf{p}^{X|Y}$(または $\mathsf{p}^{X|Y}$) とする. $Y = y$ を与えたときの X の条件付き期待値 を

$$\mathsf{E}[X|\,Y = y] = \begin{cases} \displaystyle\sum x\mathsf{p}^{X|Y}(x|\,y) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{x\infty} x\mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x & \text{(連続型の場合)} \end{cases}$$

で定義する. ただし考えている $Y = y$ で条件付き p.d.f. または p.m.f. は定義され, さらに $\mathsf{E}[|X|] < \infty$ とする.
(2) (Borel 可測) 関数 $g : \mathbb{R}^2 \to \mathbb{R}$ に対して, $Y = y$ を与えたときの $g(X, Y)$ の条件付き期待値 を

$$\mathsf{E}[g(X, Y)|\,Y = y] = \begin{cases} \displaystyle\sum g(x, y)\mathsf{p}^{X|Y}(x|\,y) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{x\infty} g(x, y)\mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x & \text{(連続型の場合)} \end{cases}$$

で定義する. ただし, 考えている $Y = y$ での条件付き p.d.f. または p.m.f. は定義され, $\mathsf{E}[|g(X, Y)|] < \infty$ とする.

注意 2.31. 定義 2.1 と同様に関数 g を g^+ と g^- の部分に分けて条件付き期待値を定義をするべきであろうが, $\mathsf{E}[|g(X)|] < \infty$ の仮定のもとでは, $\mathsf{E}[g^+(X)|\, Y = y] < \infty$, $\mathsf{E}[g^-(X)|\, Y = y] < \infty$ が保証されるので, 上記のように簡単に定義した.

注意 2.32. $\mathsf{E}[X]$ は定数であるが, $\mathsf{E}[X|\, Y = y]$ は一般に y の関数である. このことから

$$h(y) := \mathsf{E}[X|\, Y = y]$$

とおいたときに $h(y)$ に Y を代入したもの $h(Y)$ は確率変数 [13] になる. これを

$$\mathsf{E}[X|\, Y] := h(Y)$$

と記すことにする. したがって $\omega \in \Omega$ に対して, $y = Y(\omega)$ と書けば

$$\mathsf{E}[X|\, Y] : \Omega \ni \omega \mapsto \mathsf{E}[X|\, Y(\omega)] = \mathsf{E}[X|\, Y = y] \in \mathbb{R}$$

は可測となる.

測度論的確率論の教科書では, Radon-Nikodym の定理から条件付き期待値を定義する. これから条件付き p.d.f. を定義することになる. この点については [8, pp.181 − 182] を参照のこと. □

例 2.33. 連続型確率変数 Y は p.d.f.

$$\mathsf{p}^Y(y) = \begin{cases} 1 & (0 < y < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする. $Y = y\,(0 < y < 1)$ を観測したとき

$$X|\, Y = y \sim \mathsf{Unif}(y, 1)$$

とする. すなわち $0 < y < 1$ のとき

$$\mathsf{p}^{X|Y}(x|\, y) = \begin{cases} \dfrac{1}{1-y} & (y < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. よって

$$\mathsf{E}[X|\, Y = y] = \int_y^1 x\mathsf{p}^{X|Y}(x|\, y)\,\mathrm{d}x = \frac{1}{1-y}\int_y^1 x\,\mathrm{d}x = \frac{1+y}{2}$$

[13] $h \mapsto h(y)$ の可測性は測度論の知識が必要となる. このことは Fubini の定理から保証されることがわかる.

となる. これより

$$\mathsf{E}[X|\,Y] = \frac{1+Y}{2}$$

となる. □

定理 2.34. (1) 有限な期待値を持つ確率変数 X, Y と確率変数 Z に対して

$$\mathsf{E}[X+Y|\,Z] = \mathsf{E}[X|\,Z] + \mathsf{E}[Y|\,Z]$$

となる.
(2) 有限な期待値を持つ確率変数 X, Y に対して

$$\mathsf{E}[\mathsf{E}[Y|\,X]] = \mathsf{E}[Y], \quad \mathsf{E}[\mathsf{E}[X|\,Y]] = \mathsf{E}[X]$$

となる.
(3) 一般の (Borel 可測) 関数 $g : \mathbb{R}^2 \to \mathbb{R}$ を考える. $\mathsf{E}[|g(X, Y)|] < \infty$ のとき

$$\mathsf{E}[\mathsf{E}[g(X, Y)|\,Y]] = \mathsf{E}[g(X, Y)]$$

となる.
(4) $\mathsf{E}[XY|\,Y] = Y\mathsf{E}[X|\,Y]$ となる.

Proof. ① (1) の証明: 期待値の線型性よりわかる.
② (2) の証明: 連続型の場合について (2) の第 1 番目の等式を示す. 他の場合もほとんど同じように証明できる. p を (X, Y) の同時 p.d.f. とする. このとき

$$\mathsf{p}(x, y) = \mathsf{p}^X(x)\mathsf{p}^{Y|X}(y|\,x)$$

となる. ただし p^X は X の周辺 p.d.f. で, $\mathsf{p}^{Y|X}$ は $X = x$ を与えたときの Y の条件付き p.d.f. である. $\mathsf{E}[Y|\,X = x] =: h(x)$ とおいたとき, $\mathsf{E}[Y|\,X] := h(X)$ と書いたことを思い出す. また, $S_X = \{x \in \mathbb{R};\, \mathsf{p}^X(x) > 0\}$ とおいたとき

$$\begin{aligned}
\mathsf{E}[\mathsf{E}[Y|\,X]] = \mathsf{E}[h(X)] &= \int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}x = \int_{S_X} h(x)\mathsf{p}^X(x)\,\mathrm{d}x \\
&= \int_{S_X} \mathsf{E}[Y|\,X = x]\mathsf{p}^X(x)\,\mathrm{d}x \\
&= \int_{S_X} \left\{ \int_{-\infty}^{\infty} y\mathsf{p}^{Y|X}(y|\,x)\,\mathrm{d}y \right\} \mathsf{p}^X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \int_{S_X} y\mathsf{p}^{Y|X}(y|\,x)\mathsf{p}^X(x)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \left\{ \int_{S_X} y\,\frac{\mathsf{p}(x, y)}{\mathsf{p}^X(x)}\,\mathsf{p}^X(x)\,\mathrm{d}x \right\} \mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \left\{ \int_{S_X} y\, \mathsf{p}(x,\, y)\, \mathrm{d}x \right\} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} y \left\{ \int_{-\infty}^{\infty} \mathsf{p}(x,\, y)\, \mathrm{d}x \right\} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} y \mathsf{p}^{Y}(y)\, \mathrm{d}y \\
&= \mathsf{E}[Y]
\end{aligned}
$$

となる[14]. 積分の順序交換は $g(X,\, Y)$ が有限の期待値を持つことから保証されること[15]が知られている.

③ (3), (4) の証明: 同様に証明できる. □

問 2.5. 定理 2.34(3)(4) の証明を書け.

例 2.35. (例 2.33 の続き)

$$
\mathsf{E}[X] = \mathsf{E}\left[\frac{1+Y}{2} \right] = \frac{3}{4}
$$

となる. 一方 $0 < y < x < 1$ に対して

$$
\mathsf{p}(x,\, y) = \mathsf{p}^{X|\,Y}(x|\, y) \mathsf{p}^{Y}(y) = \frac{1}{1-y}
$$

となる. よって

$$
\begin{aligned}
\mathsf{E}[X] &= \int_0^1 \left\{ \int_y^1 x\, \frac{1}{1-y}\, \mathrm{d}x \right\} \mathrm{d}y \\
&= \int_0^1 \frac{1}{1-y} \left[\frac{x^2}{2} \right]_y^1 \mathrm{d}y \\
&= \int_0^1 \frac{1}{1-y} \frac{1-y^2}{2}\, \mathrm{d}y = \left[\frac{(1+y)^2}{4} \right]_0^1 = \frac{3}{4}
\end{aligned}
$$

となる. □

定義 2.36. X は有限の 2 次の積率を持つとする. $Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|\,Y}$(p.m.f. $\mathsf{p}^{X|\,Y}$) が定義できる y を考える. このとき, $Y = y$ **を与えたときの条件付き分散** を

$$
\mathsf{Var}[X|\, Y = y] = \begin{cases} \displaystyle\sum_{x} \{x - \mu(y)\}^2 \mathsf{p}^{X|\,Y}(x|\, y) & \text{(離散型の場合)} \\ \displaystyle\int_{-\infty}^{\infty} \{x - \mu(y)\}^2 \mathsf{p}^{X|\,Y}(x|\, y)\, \mathrm{d}x & \text{(連続型の場合)} \end{cases}
$$

[14] h は $S_Y := \{y \in \mathbb{R};\, \mathsf{p}^{Y}(y) > 0\}$ においてのみ定義されるが, 形式的に $h(y) = 0\, (y \notin S_Y)$ と考えて上の式を理解すればよい.

[15] Fubini の定理からわかる.

で定義する. ただし $\mu(y) = \mathsf{E}[X|Y=y]$ である. これは

$$\mathsf{Var}[X|Y] = \mathsf{E}[X^2|Y] - \{\mathsf{E}[X|Y]\}^2$$

とも書ける.

定理 2.37. X, Y を確率変数とし $\mathsf{E}[X^2] < \infty$ とする. このとき

$$\mathsf{Var}[X] = \mathsf{E}[\mathsf{Var}[X|Y]] + \mathsf{Var}[\mathsf{E}[X|Y]]$$

が成立する. ただし $h(y) := \mathsf{Var}[X|Y=y]$ としたとき $\mathsf{Var}[X|Y] := h(Y)$ と定義した.

Proof. まず

$$\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}\big[\{X - \mathsf{E}[X]\}^2\big] \\
&= \mathsf{E}\big[\big\{X - \mathsf{E}[X|Y] + \mathsf{E}[X|Y] - \mathsf{E}[X]\big\}^2\big] \\
&= \mathsf{E}\big[\{X - \mathsf{E}[X|Y]\}^2\big] + \mathsf{E}\big[\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}^2\big] \\
&\quad + 2\mathsf{E}\big[\{X - \mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}\big] \qquad (2.9)
\end{aligned}$$

と書き直す. しかし, (2.9) の最右辺の各項は以下のように評価できる.

$$\begin{aligned}
\mathsf{E}\big[\{X - \mathsf{E}[X|Y]\}^2\big] &= \mathsf{E}\big[X^2 - 2X\mathsf{E}[X|Y] + \{\mathsf{E}[X|Y]\}^2\big] \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2 - 2X\mathsf{E}[X|Y] + \{\mathsf{E}[X|Y]\}^2\big]\Big|Y\Big] \\
&\quad (\because \text{定理 } 2.34(3)) \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2\big|Y\big] - 2\mathsf{E}[X|Y]\,\mathsf{E}[X|Y] + \{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\Big[\mathsf{E}[X^2|Y] - \{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\big[\mathsf{Var}[X|Y]\big], \\
\mathsf{E}\Big[\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}^2\Big] &= \mathsf{E}\Big[\big\{\mathsf{E}[X|Y] - \underbrace{\mathsf{E}[\mathsf{E}[X|Y]]}_{=\mathsf{E}[X]}\big\}^2\Big] \\
&\quad (\because \text{定理 } 2.34(3)) \\
&= \mathsf{Var}\Big[\mathsf{E}[X|Y]\Big], \\
\mathsf{E}\big[\{X - \mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}\big] &= \mathsf{E}\Big[\mathsf{E}\big[\{X - \mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}\big|Y\big]\Big] \\
&\quad (\because \text{定理 } 2.34(3)) \\
&= \mathsf{E}\Big[\{\mathsf{E}[X|Y] - \mathsf{E}[X]\}\,\underbrace{\mathsf{E}\big[X - \mathsf{E}[X|Y]\big|Y\big]}_{=0}\Big]
\end{aligned}$$

$$= 0$$

となる. 最後から 2 番目の等号は定理 2.34(3) を用いた. これらの結果を (2.9) の最右辺の各項に代入すれば, 定理は証明される. □

2.5 積率母関数

定義 2.38. X を確率変数とし, ある $t_0 > 0$ が存在して, $\mathsf{E}[e^{tX}] < \infty\,(\forall |t| < t_0)$ とする. このとき, X の積率母関数(Moment Generating Function (m.g.f.)) を

$$\mathsf{m}^X(t) := \mathsf{E}[e^{tX}] \quad (-t_0 < t < t_0)$$

で定義する.

定理 2.39. ある $\delta > 0$ に対して, $\mathsf{M}^X(t) < \infty\,(\forall |t| < \delta)$ が成り立つとき, $\mathsf{M}^X(t)$ は開区間 $(-\delta, \delta)$ で何回も微分可能で, 任意の非負値整数 m と $t \in (-\delta, \delta)$ に対して, $\big(\mathsf{M}^X\big)^{(m)}(t) = \mathsf{E}\big[X^m e^{tX}\big]$ となる. よって, $\mathsf{E}\big[X^m\big] = \big(\mathsf{M}^X\big)^{(m)}(0)$ となる. ただし, $\big(\mathsf{M}^X\big)^{(m)}(t)$ は $\mathsf{M}^X(t)$ の m 次導関数である.

Proof. 期待値記号 E と微分記号 $\dfrac{\mathrm{d}}{\mathrm{d}t}$ の交換の正当化のために以下の記号を準備する. $a > 0$ と $m \in \mathbb{N}$ に対して, $C_{a,\,m} = \left(\dfrac{m}{a}\right)^m e^{-m}$ とおく. さらに, $t \in (-\delta, \delta)$ と $m = 0, 1, \ldots$ に対して, $\mathsf{L}_m^X(t) := \mathsf{E}\big[X^m e^{tX}\big]$ と $\mathsf{M}^{|X|}(t) = \mathsf{E}\big[e^{t|X|}\big]$ と形式的に定める. 以下のことを示していく. ① $x^m e^{-ax} \le C_{a,\,m}\,(x \ge 0)$ が成り立つこと, ② $-\delta < t < \delta$ に対して, $\mathsf{M}^{|X|}(t) < \infty$ であること, ③ 任意の $m = 0, 1, \ldots$ と $t \in (-\delta, \delta)$ に対して, $\big|\mathsf{L}_m^X(t)\big| < \infty$ であること, ④ $x \in \mathbb{R}$, $t \in \mathbb{R}$, $\Delta t \neq 0$, $m = 0, 1, \ldots$ に対して

$$\left|\frac{x^m e^{(t+\Delta t)x} - x^m e^{tx}}{\Delta t} - x^{m+1} e^{tx}\right| \le \frac{1}{2}\big|\Delta t\big| e^{(|t|+|\Delta t|)|x|}\big|x\big|^{m+2}$$

が成り立つこと, ⑤ $t \in (-\delta, \delta)$ と $m = 0, 1, \ldots$ に対して

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{L}_m^X(t) = \mathsf{L}_{m+1}^X(t)$$

が成り立つことを示していく. 以上から, $\mathsf{M}^X(t) = \mathsf{L}_0^X(t)$ と ⑤ を用いると

$$\frac{\mathsf{d}}{\mathrm{d}t}\mathsf{M}^X(t) = \frac{\mathsf{d}}{\mathrm{d}t}\mathsf{L}_0^X(t) = \mathsf{L}_1^X(t)$$

がわかる. 以上の操作を繰り返せば

$$\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathsf{M}^X(0) = \mathsf{L}_m^X(0) = \mathsf{E}\big[X^m\big]$$

がわかる.
① $g(x) \le C_{a,,m}$ の証明: $g(x) := x^m e^{-ax}\,(x \ge 0)$ とおく. すると

$$\overset{\bullet}{g}\,(x) = mx^{m-1}e^{-ax} - ax^m e^{-ax} = x^{m-1}e^{-ax}\big\{m - ax\big\}$$

から $g(x) \le g(m/a) = C_{a,\,m}$ となることがわかる.
② $\mathsf{M}^{|X|}(t) < \infty\,(|t| < \delta)$ の証明: 任意の $x,\, t \in \mathbb{R}$ に対して

$$0 \le e^{t|x|} \le e^{tx} + e^{-tx}$$

となる. x に X を代入して, 両辺の期待値を取ると

$$\mathsf{E}\big[e^{t|X|}\big] \le \mathsf{E}\big[e^{tX}\big] + \mathsf{E}\big[e^{tX}\big] < \infty \qquad (-\delta < t < \delta)$$

がわかる. さらに, $e^{t|x|} \ge \dfrac{|x|^m}{m!} \ge 0$ から $\mathsf{E}\big[|X|^m\big] < \infty$ もわかる.
③ $\mathsf{L}_m^{|X|}(t) < \infty$ の証明: $|t| < \delta$ なので, $|t| < t' < \delta$ なる t' を選ぶことができる. さらに, $a = t' - |t|$ とおくと $a > 0$ となる. ① の結果から, $|x|^m e^{-a|x|} \le C_{a,\,m}$ なので

$$\begin{aligned}\big|\mathsf{L}_m^X(t)\big| &= \big|\mathsf{E}\big[X^m e^{tX}\big]\big| \le \mathsf{E}\big[|X|^m\big|e^{tX}\big|\big] \le \mathsf{E}\big[|X|^m e^{|t|\,\cdot\,|X|}\big] \\ &= \mathsf{E}\big[|X|^m e^{-a|X|}e^{\{|t|+a\}|X|}\big] \le \mathsf{E}\big[C_{a,\,m}e^{\{|t|+a\}|X|}\big] = C_{a,\,m}\mathsf{E}\big[e^{t'|X|}\big] \\ &= C_{a,\,m}\mathsf{M}^{|X|}(t') < \infty \qquad (t' < \delta)\end{aligned}$$

となることがわかる.
④ $\left|\frac{x^m e^{(t+\Delta t)x} - x^m e^{tx}}{\Delta t} - x^{m+1}e^{tx}\right| \le \frac{1}{2}\big|\Delta t\big|e^{(|t|+|\Delta t|)|x|}|x|^{m+2}$ の証明: Maclaurin 展開により, ある $0 < \theta < 1$ が存在して

$$e^x = 1 + x + \frac{1}{2}e^{\theta x}x^2$$

と書ける. これより

$$\begin{aligned}x^m e^{(t+\Delta t)x} - x^m e^{tx} &= x^m e^{tx}\big\{e^{\Delta tx} - 1\big\} \\ &= x^m e^{tx}\left\{1 + \Delta tx + \frac{1}{2}e^{\theta\Delta tx}(\Delta tx)^2 - 1\right\} \\ &= x^{m+1}e^{tx}\Delta t + \frac{1}{2}x^{m+2}e^{(t+\theta\Delta t)x}(\Delta t)^2\end{aligned}$$

となる. したがって

$$\begin{aligned}
&\left|\frac{x^m e^{(t+\Delta t)x} - x^m e^{tx}}{\Delta t} - x^{m+1}e^{tx}\right| \\
&\quad= \left|\frac{x^{m+1}e^{tx}\Delta t + \frac{1}{2}x^{m+2}e^{(t+\theta\Delta t)x}(\Delta t)^2}{\Delta t} - x^{m+1}e^{tx}\right| \\
&\quad= \frac{1}{2}\left|x^{m+2}e^{(t+\theta\Delta t)x}\Delta t\right| \\
&\quad\le \frac{1}{2}\left|\Delta t\right| \cdot \left|x\right|^{m+2} e^{(|t|+|\Delta t|)|x|}
\end{aligned}$$

となる.

⑤ $\underline{\left(\mathsf{M}^X\right)^{(m)}(t) = \mathsf{E}\left[X^m e^{tX}\right] \text{ の証明:}}$ ④ で得た不等式で x に X を代入して, 両辺の期待値を取ると

$$\begin{aligned}
&\left|\frac{\mathsf{L}_m^X(t+\Delta t) - \mathsf{L}_m^X(t)}{\Delta t} - \mathsf{L}_{m+1}^X(t)\right| \\
&\quad= \left|\mathsf{E}\left[\frac{X^m e^{(t+\Delta t)X} - X^m e^{tX}}{\Delta t} - X^{m+1}e^{tX}\right]\right| \\
&\quad\le \mathsf{E}\left[\left|\frac{X^m e^{(t+\Delta t)X} - X^m e^{tX}}{\Delta t} - X^{m+1}e^{tX}\right|\right] \\
&\quad\le \frac{1}{2}\left|\Delta t\right|\mathsf{E}\left[\left|X\right|^{m+2} e^{(|t|+|\Delta t|)|X|}\right]
\end{aligned}$$

を得る. $|t| < \delta$ なので, $|t| < t' < \delta$ なる t' をを取ることができる. さらに, $a := \dfrac{t' - |t|}{2}$ とおくと, $a > 0$ となる. $|\Delta t| < a$ をみたすような Δt を考えと $\underbrace{|t|}_{=t'-2a} + |\Delta t| + a = t' - a + |\Delta t| < t'$ となることに注意する. すると $t' \in (-\delta,\, \delta)$ なので

$$\begin{aligned}
\mathsf{E}\left[|X|^{m+2} e^{(|t|+|\Delta t|)|X|}\right] &= \mathsf{E}\left[|X|^{m+2} e^{-a|X|} e^{(|t|+|\Delta t|+a)|X|}\right] \le C_{a,\,m+2}\mathsf{E}\left[e^{t'|X|}\right] \\
&= C_{a,\,m+2}\mathsf{M}^{|X|}(t') < \infty
\end{aligned}$$

がわかる. よって

$$\begin{aligned}
\left|\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{L}_m^X(t) - \mathsf{L}_{m+1}^X(t)\right| &= \left|\lim_{\Delta t\to 0}\frac{\mathsf{L}_m^X(t+\Delta t) - \mathsf{L}_m^X(t)}{\Delta t} - \mathsf{L}_{m+1}^X(t)\right| \\
&= \lim_{\Delta t\to 0}\left|\frac{\mathsf{L}_m^X(t+\Delta t) - \mathsf{L}_m^X(t)}{\Delta t} - \mathsf{L}_{m+1}^X(t)\right| \\
&\le \frac{1}{2}\lim_{\Delta t\to 0}\left|\Delta t\right|\mathsf{E}\left[\left|X\right|^{m+2} e^{(|t|+|\Delta t|)|X|}\right] = 0
\end{aligned}$$

となる. よって

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{L}_m^X(t) = \mathsf{L}_{m+1}^X(t)$$

が示せた.

□

例 2.40. $X \sim \mathsf{Exp}(1)$ とする. $t < 1$ に対して

$$\mathsf{m}^X(t) = \mathsf{E}\left[e^{tX}\right] = \int_0^\infty e^{tx}e^{-x}\,\mathrm{d}x = \int_0^\infty e^{-(1-t)x}\,\mathrm{d}x = \frac{1}{1-t}$$

となる. $t \geq 1$ のときは, e^{tX} の期待値は発散する. したがって

$$\mathsf{m}^X(t) = \frac{1}{1-t} \qquad (t < 1)$$

となる. 簡単な計算から

$$\overset{\bullet}{\mathsf{m}}{}^X(0) = 1, \quad \overset{\bullet\bullet}{\mathsf{m}}{}^X(0) = 2$$

なので

$$\mathsf{E}[X] = 1, \quad \mathsf{E}[X^2] = 2, \quad \mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2 = 1$$

となる. □

補題 2.41. (1) $a, b \in \mathbb{R}\,(a \neq 0)$ とする. $Y = aX + b$ としたとき

$$\mathsf{m}^Y(t) = e^{tb}\mathsf{m}^X(at)$$

となる.
(2) $X_1, X_2, \ldots, X_d$ は独立とし $Y = \sum_{j=1}^d X_j$ とする. このとき

$$\mathsf{m}^Y(t) = \prod_{j=1}^d \mathsf{m}^{X_j}(t)$$

となる.

Proof. (1) は指数関数と期待値の性質よりわかる. (2) は $X_j\,(j = 1, 2, \ldots, d)$ の独立性と指数関数の性質に注意して, 定理 2.17(3) を適用するとわかる. □

定理 2.42. X と Y を確率変数とする. ある数 $t_0 > 0$ が存在して

$$\mathsf{m}^X(t) = \mathsf{m}^Y(t) \quad (|t| < t_0)$$

ならば

$$X \overset{d}{=} Y$$

となる. ただし X と Y の c.d.f. を F^X と F^Y としたとき

$$X \overset{d}{=} Y \;\Leftrightarrow\; \mathsf{F}^X(x) = \mathsf{F}^Y(x)\,(\forall x \in \mathbb{R})$$

である.

Proof. これは信じることにする. □

注意 2.43. 定理 2.42 の証明は, 特性関数と分布の間に一対一対応があることを利用して証明する. $\mathsf{M}^X(t) < \infty\,(|t| < \delta)$ とする. 複素数 $z = t + \sqrt{-1}\,s\,(t,\, s \in \mathbb{R})$ に対して, X の積率母関数の複素領域への拡張 $\mathsf{E}\left[e^{zX}\right]\,(\left|\mathrm{Re}\,z\right| = |t| < \delta)$ のとき, 正則であることがわかる. 一致の定理と Levi の反転公式 (特性関数と分布が一対一対応すること) を用いて, 証明する. 阪本 (2025, pp.166-167) を参照のこと. □

例 2.44. $n_1,\, n_2 \in \mathbb{N}, 0 < p < 1$ とし $X_1 \sim \mathsf{Bino}(n_1,\, p)$, $X_2 \sim \mathsf{Bino}(n_2,\, p)$ は独立とする. $Y = X_1 + X_2$ としたとき

$$\mathsf{m}^Y(t) = \mathsf{m}^{X_1}(t)\mathsf{m}^{X_2}(t) = \left(pe^t + q\right)^{n_1}\left(pe^t + q\right)^{n_2} = (pe^t + q)^{n_1+n_2}$$

となる. ただし $q = 1 - p$ である. よって $Y \sim \mathsf{Bino}(n_1 + n_2,\, p)$ となる. □

例 2.45. $\lambda_1,\, \lambda_2 > 0$ とし, $X_1 \sim \mathsf{Po}(\lambda_1)$, $X_2 \sim \mathsf{Po}(\lambda_2)$ は独立とする. $Y = X_1 + X_2$ としたとき

$$\mathsf{m}^Y(t) = \mathsf{m}^{X_1}(t)\mathsf{m}^{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

となる. したがって, $Y \sim \mathsf{Po}(\lambda_1 + \lambda_2)$ がわかる. □

2.6 章末注釈と参考文献

この章は [42] を参考にした.

2.7 演習問題

演習問題 2.1. (1) 非負の正数 n に対して

$$(1+x)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) x^k$$

を帰納法を用いて証明せよ. ただし

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}, \qquad 0! = 1$$

である.

(2)

$$\left(\begin{array}{c} n \\ 0 \end{array}\right) + \left(\begin{array}{c} n \\ 1 \end{array}\right) + \cdots + \left(\begin{array}{c} n \\ n \end{array}\right) = 2^n \qquad (n \geq 0)$$

を示せ.

(3)

$$\left(\begin{array}{c} n \\ 0 \end{array}\right) - \left(\begin{array}{c} n \\ 1 \end{array}\right) + \left(\begin{array}{c} n \\ 2 \end{array}\right) - \left(\begin{array}{c} n \\ 3 \end{array}\right) \cdots + (-1)^n \left(\begin{array}{c} n \\ n \end{array}\right) = 0 \qquad (n \geq 1)$$

を示せ.

(4)

$$\sum_{k=1}^{n} k \cdot \left(\begin{array}{c} n \\ k \end{array}\right) = n2^{n-1} \qquad (n \geq 1)$$

を示せ.

(5)

$$\sum_{k=2}^{n} k(k-1) \cdot \left(\begin{array}{c} n \\ k \end{array}\right) = n(n-1)2^{n-2} \qquad (n \geq 2)$$

を示せ.

演習問題 2.2. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は

$$\mathsf{Pr}(X = c) = 1; \qquad c \text{ は定数}$$

をみたすとする. このとき, X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

演習問題 2.3. (1) $X \sim \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ のとき

$$\mathsf{E}[X] = \theta, \quad \mathsf{Var}[X] = \theta(1-\theta)$$

を示せ.

(2) $X_1, X_2, \ldots, X_n \sim$ i.i.d.$\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. $S = \sum_{j=1}^{n} X_j$ とおいたとき, $S \sim \mathsf{Bino}(n, \theta)$ となることを示せ.

(3) $S \sim \mathsf{Bino}(n, \theta)\,(n \in \mathbb{N}, 0 < \theta < 1)$ のとき

$$\mathsf{E}[S] = n\theta, \quad \mathsf{Var}[S] = n\theta(1-\theta)$$

を示せ.

(4) $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. このとき, X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

(5) $U \sim \mathsf{U}(0, 1)$ のとき

$$\mathsf{E}[U] = \frac{1}{2}, \quad \mathsf{Var}[U] = \frac{1}{12}$$

を示せ.
(6) $X \sim \mathsf{Ex}(\theta)\,(\theta > 0)$ とする. このとき

$$\mathsf{E}[X] = \frac{1}{\theta}, \quad \mathsf{Var}[X] = \frac{1}{\theta^2}$$

を示せ.
(7) $X \sim \mathsf{N}(\mu, \sigma^2)\,(\mu \in \mathbb{R}, \sigma > 0)$ とする. このとき

$$\mathsf{E}[X] = \mu, \quad \mathsf{Var}[X] = \sigma^2$$

を示せ.
(8) $X \sim \chi^2_n$ のとき

$$\mathsf{E}[X] = n, \quad \mathsf{Var}[X] = 2n$$

を示せ.

演習問題 2.4. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は $\mathsf{Var}[X] = \sigma^2\,(0 < \sigma < \infty)$ とする. $c\,(\neq 0)$ を定数としたとき, $\mathsf{Var}[X + c]$ と $\mathsf{Var}[cX]$ を σ と c を用いて表せ.

演習問題 2.5. (1) $X \sim \mathsf{Ga}(a, \beta)$ とする. X の k 次の積率は以下で与えられることを示せ.

$$\mathsf{E}[X^k] = \frac{(a + k - 1)(a + k - 2) \cdots a}{\beta^k}.$$

(2) 上の問いの結果から

$$\mathsf{E}[X] = \frac{a}{\beta}, \quad \mathsf{E}[X^2] = \frac{(a + 1)a}{\beta^2}, \quad \mathsf{Var}[X] = \frac{a}{\beta^2}$$

を示せ.
(3) X の積率母関数は以下で与えられることを示せ.

$$\mathsf{m}^X(t) = \left(\frac{1}{1 - \beta t}\right)^a \quad (t < 1/\beta).$$

演習問題 2.6. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2,\,\infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \leq \frac{1}{2}\right) \end{cases}$$

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(2) 確率変数 Y の平均 $\mathsf{E}[Y]$ と分散 $\mathsf{Var}[Y]$ を求めよ.
(3) 確率変数 X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.

演習問題 2.7. X_1, X_2 を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし,

$$X_1 \sim \mathsf{N}(x_0 + \mu, \sigma^2), \qquad X_2 | \, X_1 = x_1 \sim \mathsf{N}(x_1 + \mu, \sigma^2)$$

とする. ただし, $x_0, x_1, \mu \in \mathbb{R}$ である. このとき, 以下の問いに答えよ. ただし, 正規分布 $\mathsf{N}(\mu, \sigma^2)$ に従う確率変数 Z の期待値と分散が $\mathsf{E}[Z] = \mu$, $\mathsf{Var}[Z] = \sigma^2$ となることは証明なしで用いてよい. また, 期待値, 分散, 共分散, 条件付き期待値に係る資料に書いてある性質も証明なしで用いてよい. なお, どの性質を用いたかは明示すること.
(1) 確率変数 $X_2 - X_1$ の期待値 $\mathsf{E}[X_2 - X_1]$ を求めよ.
(2) X_2 の分散 $\mathsf{Var}[X_2]$ を求めよ.
(3) X_1 と X_2 の共分散 $\mathsf{Cov}[X_1, X_2]$ を求めよ.
(4) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を $\mathsf{Var}[X_1]$, $\mathsf{Var}[X_2]$, $\mathsf{Cov}[X_1, X_2]$ で表現せよ.
(5) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を求めよ.

演習問題 2.8. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, X をこの空間上で定義された確率変数とする. X は開区間 $(0, 1)$ 上の一様分布に従うとし

$$\boldsymbol{X} = \begin{bmatrix} X \\ X^2 \end{bmatrix}$$

とおく. このとき, 以下の問いに答えよ.
(1) 確率ベクトル $\boldsymbol{X}$ の期待値 $\mathsf{E}[\boldsymbol{X}]$ を求めよ.
(2)

$$\mathsf{Var}[\boldsymbol{X}] = \mathsf{E}[\boldsymbol{X}\boldsymbol{X}^\top] - \mathsf{E}[\boldsymbol{X}]\big\{\mathsf{E}[\boldsymbol{X}]\big\}^\top$$

を示せ.
(3) 確率ベクトル $\boldsymbol{X}$ の分散共分散行列 $\mathsf{Var}[\boldsymbol{X}]$ を求めよ.
(4) $\det\bigl[\mathsf{Var}[\boldsymbol{X}]\bigr]$ を求めよ.

演習問題 2.9. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}^{(X,Y)}(x, y) = \begin{cases} \dfrac{3}{8} \times y & (0 \le x \le y \le 2) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つとする.
(1) X の周辺 p.d.f. $\mathsf{p}^X(x)$ を求め,

$$\int_{-\infty}^{\infty} \mathsf{p}^X(x)\,\mathrm{d}x = 1$$

を確認せよ.
(2) Y の周辺 p.d.f. $\mathsf{p}^Y(y)$ を求め,

$$\int_{-\infty}^{\infty} \mathsf{p}^Y(y)\,\mathrm{d}y = 1$$

を確認せよ.
(3) X と Y の期待値 $\mathsf{E}[X]$ と $\mathsf{E}[Y]$ を求めよ.
(4) XY の期待値 $\mathsf{E}[XY]$ を求め, X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.
(5) $\mathsf{p}^Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|Y}(x|\,y)$ を求めよ. $\mathsf{p}^{X|Y}(x|\,y) > 0$ となる x の範囲を書いたうえで,

$$\int_{-\infty}^{\infty} \mathsf{p}^{X|Y}(x|\,y)\,\mathrm{d}x = 1$$

を確認せよ.

演習問題 2.10. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された連続型確率ベクトル (X, Y) は同時 p.d.f.

$$\mathsf{p}^{(X,Y)}(x, y) = \begin{cases} e^{-(x+y)} & (0 < x < \infty, 0 < y < \infty) \\ 0 & (\text{その他の場合}) \end{cases}$$

を持つ.

$$U = 2X, \qquad V = X + Y$$

とおく.
(1) 確率ベクトル (U, V) の同時 p.d.f. $\mathsf{p}^{(U,V)}(u, v)$ を求めよ. さらに

$$T = \{(u, v) \in \mathbb{R}^2 : \mathsf{p}^{(U,V)}(u, v) > 0\}$$

を図示せよ.
(2) U の周辺 p.d.f. $\mathsf{p}^U(u)$ を求めよ.

演習問題 2.11. $\lambda > 0$ とし, W は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数で, 母数 λ の指数分布 $\mathsf{Exp}(\lambda)$ に従うとする. すなわち, W は連続型確率変数で, p.d.f.

$$\mathsf{p}^W(w) = \begin{cases} \lambda e^{-\lambda w} & (w > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

をもつ. 同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数 X と Y を

$$X = \lfloor W \rfloor, \quad Y = W - X$$

で定める. ただし, $\lfloor x \rfloor$ は実数 x 以下の最大の整数とする. たとえば, $\lfloor 1.5 \rfloor = 1$ となる. このとき, 以下の問いに答えよ.
(1) 非負の整数 x に対して

$$\mathsf{Pr}(X = x)$$

を求めることで X の p.m.f. p^X を求めよ.
(2) $\displaystyle\sum_{x=0}^{\infty} \mathsf{p}^X(x)$ を計算せよ.
(3) x を非負の整数とし, $0 < y < 1$ とする. $X = x$ を与えたときの $\{Y \le y\}$ の条件付き確率

$$\mathsf{Pr}(Y \le y \,|\, X = x)$$

を求めよ.
(4) x を非負の整数とする. $y \in \mathbb{R}$ に対して, $\mathsf{F}^{Y|X}(y|\,x) = \mathsf{Pr}(Y \le y \,|\, X = x)$ と定める. $\mathsf{F}^{Y|X}(y|\,x)$ を実数上で定義された関数として書き表せ.
(5) x を非負の整数とする. $X = x$ を与えたときの Y の条件付き p.d.f. $\mathsf{p}^{Y|X}(y|\,x)$ を求めよ.
(6) 期待値 $\mathsf{E}[Y]$ を求めよ.

補足 $\mathsf{Pr}(0 < Y < 1) = 1$ から $0 < \mathsf{E}[Y] < 1$ となる. 得られた結果について, 不等式 $0 < \mathsf{E}[Y] < 1$ となっているかを確認するとよい.

演習問題 2.12. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2,\infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \leq \frac{1}{2}\right) \end{cases}$$

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(2) 確率変数 Y の平均 $\mathsf{E}[Y]$ と分散 $\mathsf{Var}[Y]$ を求めよ.
(3) 確率変数 X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.

演習問題 2.13. (U, Y) の同時分布を以下のように定める. $U \sim \mathsf{Unif}(0, 1)$ とし, $U = u\,(0 < u < 1)$ が与えられたときの Y の条件付き分布は $\mathsf{N}(\theta, \sigma^2)$ とする. ただし,

$$\theta = 2 - 6u, \quad \sigma^2 = 25$$

とする.
(1) U の期待値 $\mathsf{E}[U]$ を計算せよ.
(2) U の分散 $\mathsf{Var}[U]$ を計算せよ.
(3) $U = u\,(0 < u < 1)$ を与えたときの Y の条件付き期待値 $\mathsf{E}[Y \,|\, U = u]$ を答えよ.
(4) Y の期待値 $\mathsf{E}[Y]$ を計算せよ.
(5) Y の分散 $\mathsf{Var}[Y]$ を求めよ.
(6) $U = u\,(0 < u < 1)$ を与えたときの条件付き分散 $\mathsf{Var}[Y \,|\, U = u]$ を答えよ.
(7) $\mathsf{E}\big[\mathsf{Var}[Y \,|\, U]\big] + \mathsf{Var}\big[\mathsf{E}[Y \,|\, U]\big]$ を計算せよ.
(8) 連続関数 $g : [0, 1] \to \mathbb{R}$ に対して

$$\mathsf{E}\big[\big(Y - g(U)\big)^2\big] \geq 25$$

となること示せ.

演習問題 2.14. $X \sim \mathsf{N}(0, 1)$ とする. $\mathsf{Var}[g(X)] < \infty$ なる 絶対連続関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ に対して

$$\mathsf{Var}[g(X)] \leq \mathsf{E}\big[\{\overset{\bullet}{g}(X)\}^2\big]$$

が成立することを示せ.

演習問題 2.15. 確率変数 X と Y は独立同一に共通の分布関数 F と p.d.f. p を持つ分布に従うとする.
(1) $V = \max\{X, Y\}$ は分布関数 $\mathsf{F}^V(v) = \{\mathsf{F}(v)\}^2\,(v \in \mathbb{R})$ と p.d.f. $\mathsf{p}^V(v) = 2\mathsf{p}(v)\mathsf{F}(v)$ を持つことを示せ.

(2) $U = \min\{X, Y\}$ の分布関数 F^U と p.d.f. p^U を求めよ.
(3) X と Y は独立同一に $[0, 1]$ 上の一様分布に従うとき, $\mathsf{E}[U]$ と $\mathsf{Cov}[U, V]$ を求めよ.
(4) X と Y は独立同一に指数分布

$$\mathsf{p}(x) = \begin{cases} e^{-x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

に従うとき, U の分布を求めよ. さらに, $\mathsf{E}[V]$ と $\mathsf{Var}[V]$ を求めよ.

演習問題 2.16. $a_i > 0\,(i = 1, 2, 3)$ は定数とする．確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 Y_1, Y_2, Y_3 は独立で，$i = 1, 2, 3$ に対する Y_i の確率密度関数 f_i は

$$\mathsf{p}_i(y) = \begin{cases} \dfrac{1}{\Gamma(a_i)}\, y^{a_i - 1} e^{-y} & (y > 0) \\ 0 & (y \leq 0) \end{cases}$$

であるとする．ただし, $\Gamma(a) = \int_0^\infty s^{a-1} e^{-s} ds\,(a > 0)$ である．確率変数 S と確率変数 $X_i\,(i = 1, 2)$ をそれぞれ

$$S = Y_1 + Y_2 + Y_3, \qquad X_i = \frac{Y_i}{S} \qquad (i = 1, 2)$$

と定める．
(1) S の平均 $\mathsf{E}[S]$ と分散 $\mathsf{Var}[S]$ をそれぞれ求めよ.
(2) (X_1, X_2) の同時確率密度関数を求めよ.
(3) X_1 と X_2 の (Pearson の) 相関係数を求めよ.

第3章　確率と期待値の不等式

直接計算するのが困難な確率や期待値に対して上限ないしは下限を与える不等式は有効である. 節 3.1 では, Markov の不等式と Hoeffding 不等式等を与える. 節 3.2 では Cauchy–Schwarz などの期待値を評価する不等式を与える.

3.1　確率に対する不等式

定理 3.1. (Markov の不等式) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の非負値確率変数を X とし, $\mathsf{E}[X] < \infty$ とする. このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\big(X \geq t\big) \leq \frac{\mathsf{E}[X]}{t}$$

が成り立つ.

Proof. X は連続型で p.d.f. p を持つ場合を示す.

$$\begin{aligned}\mathsf{E}[X] &= \int_0^\infty x\mathsf{p}(x)\,\mathrm{d}x = \int_0^t x\mathsf{p}(x)\,\mathrm{d}x + \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \geq \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \\ &\geq t\int_t^\infty \mathsf{p}(x)\,\mathrm{d}x = t\mathsf{Pr}(X \geq t)\end{aligned}$$

からわかる. □

系 3.2. $\lambda > 0$ とする. X を確率変数とし $\mathsf{E}\big[e^{\lambda X}\big] < \infty$ とする. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(X \geq t\big) \leq e^{-\lambda t}\mathsf{E}\big[e^{\lambda X}\big]$$

が成り立つ.

Proof. $e^{\lambda X}$ は非負値確率変数なので定理 3.1 を適用すれば

$$\mathsf{Pr}\big(X \geq t\big) = \mathsf{Pr}\big(e^{\lambda X} \geq e^{\lambda t}\big) \leq e^{-\lambda t}\mathsf{E}\big[e^{\lambda X}\big]$$

となる. □

系 3.3. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X] < \infty\,(0 < \sigma < \infty)$ とする. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(|X - \mu| \geq t\big) \leq \frac{\sigma^2}{t^2}$$

が成り立つ.

Proof. $(X-\mu)^2$ に対して Markov の不等式 (定理 3.1) を適用する. すると

$$\mathsf{Pr}\big(|X - \mu| \geq t\big) = \mathsf{Pr}\big((X-\mu)^2 \geq t^2\big) \leq \frac{\mathsf{E}[(X-\mu)^2]}{t^2}$$

がわかる. □

次に Hoeffding の不等式を証明するための補題を与える.

補題 3.4. $a, b \in \mathbb{R}$ は $a < 0 < b$ なる定数とする. 確率変数 X は

$$\mathsf{E}[X] = 0, \quad a \leq X \leq b$$

とみたすとき $\forall \lambda > 0$ に対して

$$\mathsf{E}\big[e^{\lambda X}\big] \leq e^{\lambda^2 (b-a)^2/8}$$

となる.

Proof. X を以下のように書き直す.

$$X = \gamma b + (1-\gamma)a, \quad \gamma = \frac{X-a}{b-a}$$

となる. すると $e^{\lambda x}$ の凸性より

$$e^{\lambda X} \leq \gamma e^{\lambda b} + (1-\gamma)e^{\lambda a} = \frac{X-a}{b-a}e^{\lambda b} + \frac{b-X}{b-a}e^{\lambda a}$$

となる. $\mathsf{E}[X] = 0$ に注意して上の式の両辺の期待値を計算すると

$$\begin{aligned}
\mathsf{E}\big[e^{\lambda X}\big] &\leq -\frac{a}{b-a}e^{\lambda b} + \frac{b}{b-a}e^{\lambda a} \\
&= ce^{\lambda b} + (1-c)e^{\lambda a} \quad \left(c := -\frac{a}{b-a},\, 1-c = \frac{b}{b-a}\right) \\
&= ce^{\lambda(1-c)(b-a)} + (1-c)e^{-\lambda c(b-a)} \\
&= e^{-\lambda c(b-a)}\big\{1 - c + ce^{\lambda(b-a)}\big\} \\
&= e^{-cu}\big\{1 - c + ce^{u}\big\} \quad (u := \lambda(b-a)) \\
&= \exp\big\{-cu + \log\big(1 - c + ce^{u}\big)\big\} =: \exp\big\{g(u)\big\}
\end{aligned}$$

がわかる. いま

$$g(0) = \overset{\bullet}{g}(0) = 0, \quad \overset{\bullet\bullet}{g}(u) \le \frac{1}{4}\,(u > 0) \tag{3.1}$$

に注意する. 実際

$$\begin{aligned}
\overset{\bullet}{g}(u) &= \frac{\mathrm{d}g}{\mathrm{d}u} = -c + \frac{ce^u}{1-c+ce^u},\\
\overset{\bullet\bullet}{g}(u) &:= \frac{\mathrm{d}^2 g}{\mathrm{d}u^2} = \frac{ce^u}{1-c+ce^u} - \frac{c^2e^{2u}}{(1-c+ce^u)^2}\\
&= \frac{ce^u(1-c)}{(1-c+ce^u)^2} \le \frac{1}{4} \qquad \big(\text{ 相加相乗平均を } 1-c \text{ と } ce^u \text{ に適用する. }\big)
\end{aligned}$$

からわかる. (3.1) に注意して, Taylor 展開をすれば

$$\begin{aligned}
g(u) &= g(0) + u\,\overset{\bullet}{g}(0) + \frac{u^2}{2}\,\overset{\bullet\bullet}{g}(\xi) \quad (\xi \in (0,\, u))\\
&= \frac{u^2}{2}\,\overset{\bullet\bullet}{g}(\xi) \le \frac{u^2}{8} = \frac{\lambda^2(b-a)^2}{8}
\end{aligned}$$

がわかる. よって

$$\mathsf{E}\big[e^{\lambda X}\big] \le e^{g(u)} \le e^{\lambda^2(b-a)^2/8}$$

を得る. □

定理 3.5. (Hoeffding の不等式) $a_j,\, b_j\,(j = 1,\, 2,\, \ldots,\, n)$ は $a_j < 0 < b_j$ なる定数とする. 確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数 $X_1,\, X_2,\, \ldots,\, X_n$ は独立で

$$\mathsf{E}[X_j] = 0, \quad a_j \le X_j \le b_j$$

をみたすとする. このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \ge t\right) \le \exp\left\{-\frac{2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right\}$$

が成立する.

Proof. 系 3.2 を用いる. $\forall \lambda > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \ge t\right) &\le e^{-\lambda t}\mathsf{E}\left[\exp\left(\lambda \sum_{j=1}^{n} X_j\right)\right]\\
&= e^{-\lambda t}\prod_{j=1}^{n}\mathsf{E}\big[\exp\{\lambda X_j\}\big] \quad (\because \text{独立性})\\
&\le e^{-\lambda t}\prod_{j=1}^{n}\exp\left\{\frac{\lambda^2(b_j-a_j)^2}{8}\right\} \quad (\because \text{補題 3.4})
\end{aligned}$$

$$= \exp\left\{-\lambda t + \frac{\lambda^2}{8}\sum_{j=1}^{n}(b_j - a_j)^2\right\}$$
$$= \exp\left[\frac{\sum_{j=1}^{n}(b_j - a_j)^2}{8}\left\{\lambda - \frac{4t}{\sum_{j=1}^{n}(b_j - a_j)^2}\right\}^2 - \frac{2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right]$$

より

$$\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) \leq \exp\left\{-\frac{2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right\}$$

がわかる. □

定理 3.6. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする. ただし $0 < \theta < 1$ である. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \geq t\big) \leq 2e^{-2nt^2}$$

となる. ただし $\overline{X}_n = (1/n)\sum_{j=1}^{n} X_j$ である.

Proof. $j = 1, 2, \ldots, n$ に対して

$$Y_j = \frac{1}{n}(X_j - \theta)$$

とおけば

$$\mathsf{E}[Y_j] = 0, \quad a \leq Y_j \leq b, \quad a = -\frac{\theta}{n}, \quad b = \frac{1-\theta}{n}, \quad (b-a)^2 = \frac{1}{n^2}$$

となる. よって定理 3.5 より

$$\mathsf{Pr}\big(\overline{X}_n - \theta \geq t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \geq t\right) \leq e^{-2nt^2} \tag{3.2}$$

となる. 同様に

$$\mathsf{Pr}\big(\overline{X}_n - \theta \leq -t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \leq -t\right) = \mathsf{Pr}\left(\sum_{j=1}^{n}(-Y_j) \geq t\right) \leq e^{-2nt^2} \tag{3.3}$$

がわかる. よって, (3.2) と (3.3) を合わせると

$$\begin{aligned}\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \geq t\big) &= \mathsf{Pr}\big(\big\{\overline{X}_n - \theta \geq t\big\} \cup \big\{\overline{X}_n - \theta \leq -t\big\}\big) \\ &\leq \mathsf{Pr}\big(\overline{X}_n - \theta \geq t\big) + \mathsf{Pr}\big(\overline{X}_n - \theta \leq -t\big) \\ &\leq 2e^{-2nt^2}\end{aligned}$$

がわかる. □

例 3.7. $X_1, X_2, \ldots, X_{100} \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする. ただし $0 < \theta < 1$ である.
(1) $n = 100$ とし, 事象 $\left|\overline{X}_{100} - \theta\right| \geq 0.2$ の確率を Chebyshev の不等式を用いて θ に関して一様に上から評価すると

$$\mathsf{Pr}\left(\left|\overline{X}_{100} - \theta\right| \geq 0.2\right) \leq 0.0625$$

がわかる.

一方 Hoeffding の不等式を用いて θ に関して一様に評価すると

$$\mathsf{Pr}\left(\left|\overline{X}_{100} - \theta\right| \geq 0.2\right) \leq 2e^{-2(100)(0.2)^2} = 0.00067$$

となる.
(2) $0 < \alpha < 1$ を固定する. いま

$$t = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

とおく. すると Hoeffding の不等式より

$$\mathsf{Pr}\left(\left|\overline{X}_n - \theta\right| \geq \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \leq \alpha$$

となる. これより

$$\mathsf{Pr}\left(\left|\overline{X}_n - \theta\right| \leq \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \geq 1 - \alpha$$

を得る. よって

$$C = \left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

とおけば

$$\mathsf{Pr}(\theta \in C) \geq 1 - \alpha$$

を得る. すなわち信頼係数 $(1 - \alpha)$ の θ の信頼区間

$$\left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

を得る.

$\alpha = 0.05, \theta = 1/2$ として, $n = 10, 20, 50, 100$ の信頼区間の比較の計算を行う. CLT の方は数値計算で.

□

注意 3.8. 信頼区間については節 8.6 を参照のこと.

定理 3.9. (Mill の不等式) $Z \sim \mathsf{N}(0,1)$ とする. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$

が成り立つ.

Proof. まず

$$\begin{aligned}
\mathsf{Pr}(Z \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \\
&\leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{z}{t} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}} \int_t^\infty \left\{ -\frac{\partial}{\partial z} \exp\left(-\frac{z^2}{2}\right) \right\} \mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \qquad (3.4)
\end{aligned}$$

が成り立つことに注意する. さらに, $\mathsf{N}(0,1)$ の p.d.f. が偶関数であることに注意すると

$$\begin{aligned}
\mathsf{Pr}(|Z| \geq t) &= \mathsf{Pr}(\{Z \geq t\} \cup \{Z \leq -t\}) \\
&\leq \mathsf{Pr}(Z \geq t) + \mathsf{Pr}(Z \leq -t) \\
&= 2\mathsf{Pr}(Z \geq t)
\end{aligned}$$

からわかる. この不等式と (3.4) を合わせると定理の不等式が得られる. □

3.2 期待値に対する不等式

定理 3.10. (Cauchy-Schwarz の不等式) 確率変数 X と Y は 2 次の有限な期待値を持つとき

$$\mathsf{E}\big[|XY|\big] \leq \sqrt{\mathsf{E}[X^2]\,\mathsf{E}[Y^2]}$$

となる.

Proof. $\mathsf{E}[X^2] = 0$ のとき, 補題 2.3 から $\mathsf{Pr}(X = 0) = 1$ となるので, 不等式は自明である. このことを踏まえ, $\mathsf{E}[X^2] \neq 0$ として証明を進める.

いま, $g(t) := \mathsf{E}\big[(tX - Y)^2\big]$ とおく. 期待値の中を展開して期待値の線型性を用いると

$$\begin{aligned} 0 \le g(t) &= \mathsf{E}\big[t^2X^2 - 2tXY + Y^2\big] \\ &= \mathsf{E}[X^2]\left\{t - \frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right\}^2 + \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]} \end{aligned}$$

となる. これから

$$\begin{aligned} g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right) = \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]} &\ge 0 \\ \Leftrightarrow\ \mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2 &\ge 0 \\ \Leftrightarrow\ \Big|\mathsf{E}[XY]\Big| &\le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]} \end{aligned} \tag{3.5}$$

を得る. (3.5) において, X, Y を $|X|, |Y|$ と置き換えると定理の不等式は示される. 等号が成立するのは $g(t) = 0$ が重解を持つときである. 重解を c とおけば

$$g(t) = \mathsf{E}[(cX - Y)^2] = 0 \ \Leftrightarrow\ \mathsf{Pr}(Y = cX) = 1$$

となる [1]. □

定義 3.11. 関数 $g : \mathbb{R} \to \mathbb{R}$ が凸であるとは各 $x, y \in \mathbb{R}$ と $\forall t \in [0, 1]$ に対して

$$g(tx + (1-t)y) \le tg(x) + (1-t)g(y)$$

が成立するときをいう. さらに $-g$ が凸のとき g は **concave** であるという.

定理 3.12. (Jensen の不等式) X を有限な期待値を持つ確率変数とする. (1) 関数 $g : \mathbb{R} \to \mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

$$\mathsf{E}[g(X)] \ge g(\mathsf{E}[X])$$

となる.

[1] 非負値確率変数 X に対して

$$\mathsf{E}[X] = 0 \ \Leftrightarrow\ \mathsf{Pr}(X = 0) = 1$$

であることに注意せよ. 実際 $\mathsf{Pr}(X > 0) > 0$ と仮定する. するとある $\epsilon > 0$ が存在して $\mathsf{Pr}(X > \epsilon) > 0$ となる. しかし $X \ge \epsilon \mathbb{1}\{X > \epsilon\}$ より

$$0 = \mathsf{E}[X] \ge \epsilon\mathsf{E}[\mathbb{1}\{X > \epsilon\}] = \epsilon\mathsf{Pr}(X > \epsilon) > 0$$

となり矛盾する. よって $\mathsf{E}[X] = 0 \Leftrightarrow \mathsf{Pr}(X = 0) = 1$ がわかる.

(2) g が concave のとき

$$\mathsf{E}[g(X)] \leq g(\mathsf{E}[X])$$

となる.

Proof. ある定数 $r \in \mathbb{R}$ が存在[2] して

$$g(\mathsf{E}[X]) + r\{x - \mathsf{E}[X]\} \leq g(x) \qquad (\forall x \in \mathbb{R})$$

となる. x に X を代入して上の不等式の両辺の期待値を取れば

$$g(\mathsf{E}[X]) \leq \mathsf{E}[g(X)]$$

がわかる. □

系 3.13. (Young の不等式) $p,\, q > 1$ とし

$$\frac{1}{p} + \frac{1}{q} = 1$$

をみたすとする. このとき $\forall a,\, b > 0$ に対して

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

となる.

Proof. 関数 g を凸とし, $Y \sim \mathsf{Unif}(0,\, 1)$ とする. 可積分関数 $h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して $X = h(Y)$ とすれば

$$g\left(\int_0^1 h(y)\,\mathrm{d}y\right) = g(\mathsf{E}[X]) \leq \mathsf{E}[g(X)] = \int_0^1 g(h(y))\,\mathrm{d}y$$

を得る. ここで

$$g(x) = e^x, \quad h(y) = \begin{cases} p\log a & \left(0 \leq y < \frac{1}{p}\right) \\ q\log b & \left(\frac{1}{p} \leq y \leq 1\right) \end{cases}$$

とおけば g は凸なので

$$\begin{aligned} ab &= \exp\left\{\frac{p\log a}{p} + \frac{q\log b}{q}\right\} = \exp\left\{\int_0^1 h(y)\,\mathrm{d}y\right\} \leq \int_0^1 \exp\big\{h(y)\big\}\,\mathrm{d}y \\ &= \frac{1}{p}\exp\big\{p\log a\big\} + \frac{1}{q}\exp\big\{q\log b\big\} = \frac{a^p}{p} + \frac{b^q}{q} \end{aligned}$$

がわかる. □

[2] このような直線を下から支える直線という. 関数が凸であることとその関数上の任意の点において下から支える直線が存在することが同値になることが知られている. 詳しくは補遺の B.6 章を参照.

定理 3.14. (1)(Hölder の不等式) p, q は $1 \leq p \leq +\infty$, $1 \leq q \leq +\infty$ と $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\big[|X|^p\big] < \infty$, $\mathsf{E}\big[|Y|^q\big] < \infty$ をみたすとする. このとき, $\mathsf{E}\big[|XY|\big] < \infty$ で

$$\mathsf{E}\big[|XY|\big] \leq \big\{\mathsf{E}\big[|X|^p\big]\big\}^{1/p}\big\{\mathsf{E}\big[|Y|^q\big]\big\}^{1/q}$$

となる. ただし, $p = +\infty$ または $q = +\infty$ のときは, $|X|^\infty$ または $|Y|^\infty$ の本質的上限を用いる.

(2)(Minkowski の不等式) $1 \leq p \leq +\infty$ で $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\big[|X|^p\big] < \infty$, $\mathsf{E}\big[|Y|^p\big] < \infty$ をみたすとする. このとき, $\mathbb{E}\big[|X+Y|^p\big] < \infty$ で

$$\big\{\mathsf{E}\big[|X+Y|^p\big]\big\}^{1/p} \leq \big\{\mathsf{E}\big[|X|^p\big]\big\}^{1/p} + \big\{\mathsf{E}\big[|Y|^p\big]\big\}^{1/p}$$

となる.

Proof. (1) の証明: $1 < p < \infty$, $1 < q < \infty$ の場合についてのみ証明を与える. $p = 1$, $q = \infty$ の場合は Cohen (2010, pp.93-94) を参照のこと.

系 3.13 から

$$|XY| \leq \frac{|X|^p}{p} + \frac{|Y|^q}{q} \tag{3.6}$$

である. まず, $\mathsf{E}[|X|^p] = 1$, $\mathsf{E}[|Y|^q] = 1$ とする. (3.6) の両辺の期待値を取ると

$$\mathsf{E}[|XY|] \leq \frac{\mathsf{E}[|X|^p]}{p} + \frac{\mathsf{E}[|Y|^q]}{q} = \frac{1}{p} + \frac{1}{q} = 1 \tag{3.7}$$

を得る. $\mathsf{E}[|X|^p] \neq 1$ または $\mathsf{E}[|Y|^q] \neq 1$ のとき, (3.7) において, X, Y を $X/\{\mathsf{E}[|X|^p]\}^{1/p}$, $Y/\{\mathsf{E}[|Y|^q]\}^{1/q}$ と置き換えると

$$\frac{\mathsf{E}[|XY|]}{\{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|Y|^q]\}^{1/q}} \leq 1$$

を得る. よって, $1 < p < \infty$, $1 < q < \infty$ の場合について (1) は証明された.

(2) の証明: $1 < p < \infty$, $1 < q < \infty$ の場合についてのみ証明を与える. $p = 1$, $q = \infty$ の場合は Cohen (2010, pp.94-95) を参照のこと. $\mathsf{E}[|X+Y|^p] = 0$ のときは, 不等式は自明なので, $\mathsf{E}[|X+Y|^p] \neq 0$ を仮定しても一般性を失わない. このとき

$$\begin{aligned}\mathsf{E}[|X+Y|^p] &\leq \mathsf{E}\big[\big(|X| + |Y|\big)|X+Y|^{p-1}\big] \\ &= \mathsf{E}\big[|X| \times |X+Y|^{p-1}\big] + \mathsf{E}\big[|Y| \times |X+Y|^{p-1}\big]\end{aligned}$$

$$
\begin{aligned}
&\leq \{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \\
&\quad + \{\mathsf{E}[|Y|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \qquad (\because (1) \text{ を用いた}) \\
&= \{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^p]\}^{1/q} + \{\mathsf{E}[|Y|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^p]\}^{1/q} \\
&= \left\{\{\mathsf{E}[|X|^p]\}^{1/p} + \{\mathsf{E}[|Y|^p]\}^{1/p}\right\}\{\mathsf{E}[|X+Y|^p]\}^{1/q}
\end{aligned}
$$

を得る. 上の不等式の最左辺と最右辺を $\{\mathsf{E}[|X+Y|^p]\}^{1/q}$ で割れば, $1 < p < \infty$, $1 < q < \infty$ の場合の (2) が示せた. □

3.3 章末注釈と参考文献

この章で扱った凸関数は統計的推測理論では重要な役割を果たす. Hoeffding の不等式は, 統計的機械学習理論で用いられる基本不等式である. この章は [32, 4 章] を参考にした.

3.4 演習問題

演習問題 3.1. $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. このとき, 以下の問いに答えよ.
(1) X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(2) Chebyshev の不等式を用いて

$$\mathsf{Pr}\big(X \geq 2\theta\big) \leq \frac{1}{\theta}$$

を示せ.

演習問題 3.2. $n \in \mathbb{N}$ とする. $1 < p < \infty$, $p^{-1} + q^{-1} = 1$ のとき [3], 任意の実数 x_j, $y_j\,(j = 1, 2, \ldots, n)$ に対して

$$\sum_{j=1}^{n} |x_j y_j| \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}\left(\sum_{j=1}^{n} |y_j|^q\right)^{1/q}$$

が成り立つことを Young の不等式を利用して示せ. この不等式を Hölder の不等式という.

演習問題 3.3. $s > t > 0$ とする. 確率変数 X が $\mathsf{E}\big[|X|^s\big] < \infty$ ならば, $\mathbb{E}\big[|X|^t\big] < \infty$ となることを Young の不等式を用いて示せ.

[3] 問題を易しくするために, $p = 1$ の場合を除いていることに注意せよ.

演習問題 3.4. $X \sim \mathsf{N}(0,\,1)$ とする．$\mathsf{Var}[g(X)] < \infty$ なる 可微分関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ に対して

$$\mathsf{Var}[g(X)] \leq \mathsf{E}\big[\{\overset{\bullet}{g}\,(X)\}^2\big]$$

が成立することを示せ．ただし，$\overset{\bullet}{g}\,(x) = \dfrac{\mathrm{d}g}{\mathrm{d}x}(x)$ である．

第4章　確率変数列と分布列の収束

確率論の最も重要な側面のひとつは確率変数列の挙動に関することである. 確率論のこの部分のことを「大標本論」,「極限論」,「漸近論」と数理統計学では呼んでいる. 大標本論の基本的な問いは次である. 確率変数列 $X_1, X_2, \ldots$ の極限の振る舞いについて言えることは何だろうか? 統計学はデータの収集にかかわる学問である. したがってデータを集めれば集めるほど何が起こるかを調べることは重要である.

すこし実解析の話題を復習する. 実数列 $\{x_n\}$ が点 x に収束するとは, 任意の $\epsilon > 0$ に対してある正の整数 $N \in \mathbb{N}$ が存在して

$$\forall n > N \Rightarrow |x_n - x| < \epsilon$$

が成り立つことである. このことを $\lim_{n\to\infty} x_n = x$ と書いた. たとえば $x_n = x\,(\forall n \in \mathbb{N})$ ならば, 明らかに $\lim_{n\to\infty} x_n = x$ となる. これと同じことを確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の確率変数列について考えてみる.

$$X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, 1)$$

とする. すなわちこの確率変数列の任意の有限個の確率変数列は独立同一に $\mathsf{N}(0, 1)$ に従う. さらに別の確率変数 X も $\mathsf{N}(0, 1)$ に従うとする. このとき X_n は X に「収束」するとしたい. しかし

$$\mathsf{Pr}\big(X_n \neq X\big) = 1 \qquad (\forall n \in \mathbb{N})$$

である.

別の例をあげる. $X_1, X_2, \ldots$ は独立で

$$X_n \sim \mathsf{N}\left(0, \frac{1}{n}\right) \qquad (n = 1, 2, \ldots)$$

とする. 直観的には, 十分大きな n に対して X_n は 0 の近辺に集中すると予想するだろう. しかし X_n は連続型確率変数なので, すべての n に対して

$$\mathsf{Pr}\big(X_n = 0\big) = 0$$

である. これらの例から, 確率変数列の収束については, 実数列の収束とは異なる道具立てが必要になることがわかる. この章では確率変数列の収束の定義を述べ, これに関わる基本的な事項をまとめたうえで次の重要な事項を説明する.

(1) 大数の法則. $X_1, X_2, \ldots$ を i.i.d. 確率変数列とし

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad \mu = \mathsf{E}[X_1]\,(-\infty < \mu < \infty)$$

とする. このとき $\overline{X}_n$ は高い確率で μ の近くにいることを保証する定理である.

(2) 中心極限定理. $\mathsf{Var}[X_1] = \sigma^2\,(0 < \sigma < \infty)$ とする. n が十分大きいとき $\sqrt{n}\big(\overline{X}_n - \mu\big)$ の分布は正規分布で近似できることを保証する定理である.

以上の事項について証明なしで主張をまず紹介する. 証明については補遺 B にまとめることにする.

4.1 確率変数列の収束のタイプ

定義 4.1. 確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の $X, X_1, X_2, \ldots$ を確率変数列とする. さらに, 各 $X_n\,(n = 1, 2, \ldots)$ の c.d.f. を F_n, X の c.d.f. を F とする.

(1) 確率変数列 $\{X_n\}$ は確率変数 X に 確率収束 するとは, $\forall \epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) = 0$$

が成り立つときをいう. このことを $X_n \xrightarrow{\mathrm{P}} X\,(n \to \infty)$ と書く.

(2) 確率変数列 $\{X_n\}$ は確率変数 X に 概収束 するとは

$$\mathsf{Pr}\left(\omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

が成り立つときをいう. このことを $X_n \xrightarrow{\text{a.s.}} X\,(n \to \infty)$ と書く.

(3) 確率変数列 $\{X_n\}$ は確率変数 X に 分布収束 または 法則収束 するとは, F のすべての連続点 x において

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

が成り立つときをいう. このことを $X_n \rightsquigarrow X\,(n \to \infty)$ と書く. また, $X_n \rightsquigarrow X\,(n \to \infty)$ かつ $X \sim \mathsf{N}(0, 1)$ のときには $X_n \rightsquigarrow \mathsf{N}(0, 1)\,(n \to \infty)$ とも書くことがある.

注意 4.2. (1) c を定数とする. $\mathsf{Pr}(X=c)=1$ かつ $X_n \overset{\mathrm{P}}{\to} X$ とき $X_n \overset{\mathrm{P}}{\to} c$ と記す.
(2) $\mathsf{Pr}(X=c)=1$ かつ $X_n \rightsquigarrow X\,(n\to\infty)$ とき $X_n \rightsquigarrow c\,(n\to\infty)$ と記す. □

定義 4.3. $X_1, X_2, \ldots$ を確率変数列とし X を別の確率変数とする. さらに $\mathsf{E}\big[X_n^2\big] < \infty, \mathsf{E}[X^2] < \infty\,(n=1, 2, \ldots)$ とする. 確率変数列 $\{X_n\}$ が確率変数 X に 平均 2 乗の意味で収束 するとは

$$\lim_{n\to\infty} \mathsf{E}\big[(X_n - X)^2\big] = 0$$

が成り立つときをいう. このことを $X_n \overset{\text{q.m.}}{\to} X\,(n\to\infty)$ と記す

定理 4.4. (1) $X_n \overset{\text{a.s.}}{\to} X\,(n\to\infty)$ ならば, $X_n \overset{\mathrm{P}}{\to} X\,(n\to\infty)$ である.
(2) $X_n \overset{\text{q.m.}}{\to} X\,(n\to\infty)$ ならば, $X_n \overset{\mathrm{P}}{\to} X\,(n\to\infty)$ である.

Proof. (1) 集合 $A := \{\omega\in\Omega;\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ は以下のように表現できる. $m, j\in\mathbb{N}$ に対して

$$A_{m,j} = \left\{\omega\in\Omega;\, |X_m(\omega) - X(\omega)| \le \frac{1}{j}\right\} \tag{4.1}$$

とおく. すると

$$A = \bigcap_{j=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j}$$

と表現できる. $X_n \overset{\text{a.s.}}{\to} X\,(n\to\infty)$ のとき, $\mathsf{Pr}(A)=1$ である. $j_0\in\mathbb{N}$ に対して

$$\bigcap_{j=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j} \supset \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j_0}$$

であるので

$$1 = \mathsf{Pr}(A) \le \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j_0}\right) \Rightarrow \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j_0}\right) = 1$$

がわかる. $B_{n,j_0} = \bigcap_{m=n}^{\infty} A_{m,j_0}$ とおくと $B_1\subset B_2\subset\cdots$ である. よって, 確率測度の連続性から

$$\lim_{n\to\infty}\mathsf{Pr}\big(B_{n,j_0}\big) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} B_{n,j_0}\right) = \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_{m,j_0}\right) = 1$$

となる. さらに, $B_{n,\,j_0} \subset A_{n,\,j_0}$ だから, 任意の $j_0 = 1, 2, \ldots$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}\big(A_{n,\,j_0}\big) \geq \lim_{n\to\infty} \mathsf{Pr}\left(|X_n - X| \leq \frac{1}{j_0}\right) = 1$$

であることがわかった. さて, 任意の $\epsilon > 0$ が与えられたとき, j_0 を十分大きくとり, $\dfrac{1}{j_0} < \epsilon$ とすると

$$\left\{|X_n - X| \leq \frac{1}{j_0}\right\} \subset \{|X_n - X| \leq \epsilon\}$$

となる. したがって

$$\lim_{n\to\infty} \mathsf{Pr}\big(|X_n - X| \leq \epsilon\big) = 1$$

が得られる. これの補集合をとると (1) の結論が証明できた.
(2) Chebyshev の不等式 (系 3.3) から

$$\mathsf{Pr}\big(|X_n - X| > \epsilon\big) \leq \frac{1}{\epsilon^2}\mathsf{E}[|X_n - X|^2]$$

であることに注意すればよい. □

注意 4.5 (定理 4.4(1) の別証明)**.** 任意の $\epsilon > 0$ を取る. $\mathbb{1}_{(\epsilon,\,\infty)}(|X_n - X|) \leq 1$ と $X_n \overset{\text{a.s.}}{\to} X\,(n \to \infty)$ であるので, $\lim_{n\to\infty} \mathbb{1}_{(\epsilon,\,\infty)}(X_n - X) = 0$ a.s. であることに注意して, Lebesgue の有界収束定理を用いると

$$\begin{aligned}\lim_{n\to\infty} \mathsf{Pr}\big(|X_n - X| > \epsilon\big) &= \lim_{n\to\infty} \mathsf{E}\big[\mathbb{1}_{(\epsilon,\,\infty)}(X_n - X)\big] \\ &= \mathsf{E}\big[\lim_{n\to\infty} \mathbb{1}_{(\epsilon,\,\infty)}(X_n - X)\big] \\ &= 0\end{aligned}$$

がわかる. □

注意 4.6. 確率収束と概収束の違いを説明する. X, X_1, X_2, $\ldots$ を同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数列とし

$$M_n(\omega) := \sup_{k\geq n} |X_k(\omega) - X(\omega)| \qquad (\omega \in \Omega)$$

とおく. すると

$$X_n \overset{\text{a.s.}}{\to} X \;\Leftrightarrow\; M_n \overset{\text{P}}{\to} 0\,(n \to \infty)$$

となることが以下の議論からわかる.

まず, $\forall \ell \in \mathbb{N}$ に対して, ある十分大きな $n_0 \in \mathbb{N}$ があって

$$\forall n \geq n_0 \ \Rightarrow \ M_n(\omega) < \frac{1}{\ell}$$

ならば

$$M_n(\omega) \overset{n\to\infty}{\longrightarrow} 0$$

となることを示す. そのために, $\ell \in \mathbb{N}$ に対して

$$B_\ell := \bigcup_{n=n_0}^{\infty} \left\{ \omega \in \Omega;\ M_n(\omega) < \frac{1}{\ell} \right\}$$

と定めると

$$\left\{ \omega \in \Omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega) \right\} = \bigcap_{\ell=1}^{\infty} B_\ell$$

と書けるので

$$X_n \overset{\text{a.s.}}{\to} X \ (n \to \infty) \ \Leftrightarrow \ \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell \right) = 1$$

がわかる.

次に, $\ell_1 < \ell_2 \, (\ell_1,\, \ell_2 \in \mathbb{N})$ に対し

$$B_{\ell_1} = \bigcup_{n=1}^{\infty} \left\{ \omega \in \Omega;\ M_n(\omega) < \frac{1}{\ell_1} \right\} \supset \bigcup_{n=1}^{\infty} \left\{ \omega \in \Omega;\ M_n(\omega) < \frac{1}{\ell_2} \right\} = B_{\ell_2}$$

なので

$$B_1 \supset B_2 \supset \cdots \supset B_\ell \supset \cdots$$

となる. したがって, 補題 1.7(4) から

$$\mathsf{Pr}(B_1) \geq \mathsf{Pr}(B_2) \geq \cdots \geq \mathsf{Pr}(B_\ell) \geq \cdots$$

がわかる. このことと補題 1.7(7) から

$$\mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell \right) = \lim_{\ell\to\infty} \mathsf{Pr}(B_\ell) \leq \mathsf{Pr}(B_\ell) \leq 1 \, (\forall \ell \in \mathbb{N})$$

となる. よって

$$1 = \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell \right) \ \Leftrightarrow \ \mathsf{Pr}(B_\ell) = 1 \, (\forall \ell \in \mathbb{N})$$

がわかる. さらに, $n_1 < n_2\,(n_1,\, n_2 \in \mathbb{N})$ に対して

$$M_{n_1}(\omega) := \sup_{k \geq n_1} |X_k(\omega) - X(\omega)| \geq \sup_{k \geq n_2} |X_k(\omega) - X(\omega)|$$

から

$$\left\{\omega \in \Omega;\, M_{n_1}(\omega) < \frac{1}{\ell}\right\} \subset \left\{\omega \in \Omega;\, M_{n_2}(\omega) < \frac{1}{\ell}\right\}$$

がわかる. よって, $\left\{\omega \in \Omega;\, M_n(\omega) < \dfrac{1}{\ell}\right\}_{n=1}^{\infty}$ は増大列となるので, 補題 1.7(6) から

$$\begin{aligned}\mathsf{Pr}(B_\ell) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty}\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right) \\ &= \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right)\end{aligned}$$

となる. よって

$$\begin{aligned}1 = \mathsf{Pr}\left(\bigcap_{\ell=1}^{\infty} B_\ell\right) &\Leftrightarrow \mathsf{Pr}(B_\ell) = 1\,(\forall \ell \in \mathbb{N}) \\ &\Leftrightarrow \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) < \frac{1}{\ell}\right\}\right) = 1\,(\forall \ell \in \mathbb{N}) \\ &\Leftrightarrow \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, M_n(\omega) \geq \frac{1}{\ell}\right\}\right) = 0\,(\forall \ell \in \mathbb{N})\end{aligned}$$

がわかる. したがって

$$X_n \overset{\text{a.s.}}{\to} X\,(n \to \infty) \;\Leftrightarrow\; M_n \overset{\text{P}}{\to} 0\,(n \to \infty) \tag{4.2}$$

となる. 以上の議論から, 概収束は n ステップ後の $|X_k - X|\,(k \geq n)$ が最大 (極大) になるものが 0 に確率収束することと同値である. □

例 4.7. $X_1, X_2, \ldots$ は同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された独立な確率変数列で

$$X_n \;\sim\; \mathsf{Ber}(p_n)\,(n \in \mathbb{N})$$

とする. ただし, $0 < p_n < 1$ とする. すると

$$X_n \overset{\text{P}}{\to} 0\,(n \to \infty) \;\Leftrightarrow\; p_n \overset{n\to\infty}{\longrightarrow} 0$$

である. 実際, 十分小さな任意の $\epsilon > 0$ に対して, $\mathsf{Pr}(X_n \leq -\epsilon) = 0$ と $X_n(\omega) \geq \epsilon \;\Leftrightarrow\; X_n(\omega) = 1\,(\omega \in \Omega)$ であることに注意すると

$$\mathsf{Pr}(|X_n| \geq \epsilon) = \mathsf{Pr}\big(\{X_n \leq -\epsilon\} \cup \{X_n \geq \epsilon\}\big) = \mathsf{Pr}(X_n \geq \epsilon) = \mathsf{Pr}(X_n = 1)$$

$$= p_n$$

がわかる.

次に

$$M_n = \sup_{k \geq n} |X_k - 0|$$

とし, $\pi_n = \mathsf{Pr}(M_n = 1)$ とおく. すると $M_n(\omega) \in \{0, 1\}\,(\omega \in \Omega)$ なので

$$\begin{aligned} 1 - \pi_n = \mathsf{Pr}(M_n = 0) &= \mathsf{Pr}\left(\bigcap_{k=n}^{\infty} \{X_k = 0\}\right) \\ &= \prod_{k=n}^{\infty} \mathsf{Pr}(X_k = 0) \qquad (\because \{X_n\}_{n=1}^{\infty} \text{ は独立}) \\ &= \prod_{k=n}^{\infty} (1 - p_k) \end{aligned}$$

を得る. すると (4.2) と $X_n \overset{\text{a.s.}}{\to} 0\,(n \to \infty)$ から $\pi_n \overset{n \to \infty}{\longrightarrow} 0$ がわかる. ここで

$$\prod_{k=n}^{\infty} (1 - p_k) \overset{n \to \infty}{\longrightarrow} 1 \Leftrightarrow \sum_{n=1}^{\infty} p_n < \infty\,(n \to \infty)$$

であること [1] に注意すると

$$X_n \overset{\text{a.s.}}{\to} 0\,(n \to \infty) \Leftrightarrow \sum_{n=1}^{\infty} p_n < \infty$$

がわかる.

たとえば, $p_n = \dfrac{1}{n}\,(n \in \mathbb{N})$ とすると

$$X_n \overset{\text{P}}{\to} 0 \qquad (n \to \infty)$$

だが

$$\mathsf{Pr}\left(\left\{\omega \in \Omega;\ \lim_{n \to \infty} X_n(\omega) = 0\right\}\right) < 1 \Leftrightarrow \ X_n \overset{\text{a.s.}}{\to} 0\,(n \to \infty)$$

がわかる. □

注意 4.8. $n = 1, 2, \ldots$ に対して $X_n \sim \mathsf{N}(0, 1/n)$ とする. $\{X_n\}_{n=1}^{\infty}$ は 0 に収束することが期待される.

[1] たとえば, [33, pp.91 − 92] を参照のこと.

まず分布収束について確認する. そのために次のような c.d.f. を考える.

$$\mathsf{F}(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

と定義する. すなわち $\mathsf{Pr}(X = 0) = 1$ をみたす確率変数 X の c.d.f. である. $\mathsf{F}(x)$ は $x = 0$ で不連続であることに注意する. したがって $x = 0$ 以外での各点収束を言えばよい.

そのために $\sqrt{n}X_n \sim \mathsf{N}(0,\, 1)\,(n = 1,\, 2,\, \ldots)$ であることに注意する. ただし, Z は標準正規分布 $\mathsf{N}(0,\, 1)$ に従う確率変数とする. $x < 0$ のとき

$$\begin{aligned} \mathsf{F}_n(x) &= \mathsf{Pr}\big(X_n \leq x\big) = \mathsf{Pr}\big(\sqrt{n}X_n \leq \sqrt{n}x\big) = \mathsf{Pr}(Z \leq \sqrt{n}x) \\ &= \int_{-\infty}^{\sqrt{n}x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z \to 0 \quad (n \to \infty) \end{aligned}$$

となる. 最後の極限は $\sqrt{n}x \overset{n\to\infty}{\longrightarrow} -\infty$ より $\mathsf{F}(x) = 0\,(x < 0)$ がわかる.

つぎに $x > 0$ のとき

$$\begin{aligned} \mathsf{F}_n(x) &= \mathsf{Pr}\big(X_n \leq x\big) = \mathsf{Pr}\big(\sqrt{n}X_n \leq \sqrt{n}x\big) = \mathsf{Pr}(Z \leq \sqrt{n}x) \\ &= \int_{-\infty}^{\sqrt{n}x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z \to 1 \quad (n \to \infty) \end{aligned}$$

となる. 最後の極限は $\sqrt{n}x \overset{n\to\infty}{\longrightarrow} \infty$ よりわかる. よって, $\mathsf{F}(x) = 1\,(x > 0)$ である.

以上の議論から $x \neq 0$ のとき

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

がわかる.

しかし, $x = 0$ は $\mathsf{F}(x)$ の不連続点なので, $\mathsf{F}_n(0) = 1/2 \neq 1 = F(0)$ であることは問題ない.

最後に, 確率収束を示そう. Chebyshev の不等式 (系 3.3) より, $\forall\epsilon > 0$ に対して

$$\mathsf{Pr}\big(\big|X_n\big| \geq \epsilon\big) = \mathsf{Pr}\big(X_n^2 \geq \epsilon^2\big) \leq \frac{\mathsf{E}\big[X_n^2\big]}{\epsilon^2} = \frac{1}{n\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

となる. 最後の等号は, $X_n \sim \mathsf{N}(0,\, 1/n)$ なので, $\mathsf{E}[X_n^2] = \dfrac{1}{n}$ となることを用いた. よって $X_n \xrightarrow{\mathrm{P}} 0\,(n \to \infty)$ である. □

次に収束のタイプ間の関係について述べる.

定理 4.9. 次に関係が成立する.

(1) $X_n \overset{\text{q.m.}}{\to} X \Rightarrow X_n \overset{\text{P}}{\to} X\,(n \to \infty)$ である.

(2) $X_n \overset{\text{P}}{\to} X \Rightarrow X_n \rightsquigarrow X\,(n \to \infty)$ である.

(3) $X_n \rightsquigarrow X$ かつある定数 c があって $\mathsf{Pr}(X = c) = 1$ のとき $X_n \overset{\text{P}}{\to} c\,(n \to \infty)$ である.

Proof. 節 4.5.3 で示す. □

注意 4.10. 定理 4.9(1)(2) の逆は一般に成立しない.
① 定理 4.9(1) の逆の反例: $U \sim \mathsf{Unif}(0,\,1)$, $X_n = \sqrt{n}\mathbb{1}_{(0,\,1/n)}(U)\,(n = 1,\,2,\,\ldots)$, $X = 0$ とおく. このとき $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}\big(\big|X_n\big| \geq \epsilon\big) = \mathsf{Pr}\big(\sqrt{n}\mathbb{1}_{(0,\,1/n)}(U) \geq \epsilon\big) = \mathsf{Pr}\left(0 < U < \frac{1}{n}\right)$$
$$= \mathsf{Pr}\left(0 \leq U \leq \frac{1}{n}\right) = \frac{1}{n} \overset{n\to\infty}{\longrightarrow} 0$$

となる. よって $X_n \overset{\text{P}}{\to} X\,(n \to \infty)$ である. しかし

$$\mathsf{E}\big[X_n^2\big] = n\int_0^{1/n} \mathrm{d}u = 1$$

なので, $X_n \overset{\text{q.m.}}{\to} X\,(n \to \infty)$ は成立しない.
② 定理 4.9(2) の逆の反例: $X \sim \mathsf{N}(0,\,1)$ とし, $X_n = -X\,(n = 1,\,2,\,\ldots)$ とする. したがって, $X_n \sim \mathsf{N}(0,\,1)$ である. 明らかに

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

である. よって $X_n \rightsquigarrow X\,(n \to \infty)$ となる. しかし $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}\big(\big|X_n - X\big| > \epsilon\big) = \mathsf{Pr}(|2X| > \epsilon) = \mathsf{Pr}\left(|X| > \frac{\epsilon}{2}\right) \neq 0$$

である. よって $X_n \overset{\text{P}}{\to} X\,(n \to \infty)$ は成立しない. □

注意 4.11. 以上の結果をまとめると

$$\begin{array}{ccccc} X_n \overset{\text{a.s.}}{\to} X\,(n\to\infty) & \Rightarrow & X_n \overset{\text{P}}{\to} X\,(n\to\infty) & \Rightarrow & X_n \rightsquigarrow X\,(n\to\infty) \\ & & \Uparrow & & \\ & & X_n \overset{\text{q.m.}}{\to} X\,(n\to\infty) & & \end{array}$$

となる.
「$X_n \rightsquigarrow X\,(n \to \infty) \Rightarrow X_n \overset{\text{q.m.}}{\to} X\,(n \to \infty)$」も「$X_n \overset{\text{q.m.}}{\to} X\,(n \to \infty) \Rightarrow X_n \overset{\text{a.s.}}{\to} X\,(n \to \infty)$」も一般には真ではない. 反例は演習問題 4.4 と 4.5 を参照のこと. □

定理 4.12 (Portmanteau の補題)**.** 確率変数列 $X_n\,(n = 1, 2, \ldots)$, X に対し, X_n と X の分布関数を F_n と F とそれぞれ書く. このとき, 以下の (1) $\sim$ (7) は同値である.
(1) $X_n \rightsquigarrow X$ である.
(2) $\mathbb{R}$ 上の任意の有界連続関数 g に対して

$$\lim_{n\to\infty} \mathsf{E}[g(X_n)] = \mathsf{E}[g(X)]$$

である.
(3) $\mathbb{R}$ 上の任意の有界 Lipschitz 連続関数 g に対して c

$$\lim_{n\to\infty} \mathsf{E}[g(X_n)] = \mathsf{E}[g(X)]$$

である.
(4) $\mathbb{R}$ 上の任意の非負値連続関数 g に対して

$$\liminf_{n\to\infty} \mathsf{E}[g(X_n)] \geq \mathsf{E}[g(X)]$$

である.
(5) $\mathbb{R}$ の任意の開集合 O に対して

$$\liminf_{n\to\infty} \mathsf{Pr}(X_n \in O) \geq \mathsf{Pr}(X \in O)$$

である.
(6) $\mathbb{R}$ の任意の閉集合 C に対して

$$\limsup_{n\to\infty} \mathsf{Pr}(X_n \in C) \leq \mathsf{Pr}(X \in C)$$

である.
(7) $\mathbb{R}$ の任意の Borel 集合 B が $\mathsf{Pr}(X \in \partial B) = 0$ ならば

$$\lim_{n\to\infty} \mathsf{Pr}(X_n \in B) = \mathsf{Pr}(X \in B)$$

である. ただし, ∂B は B の境界である.

Proof. 節 4.5.4 で示す. ただし, 証明には積分の収束定理の知識が必要である. □

注意 4.13. 定理 4.12 を有限次元の確率ベクトルに拡張することができる. 記号が煩雑にはなるが, 証明は本質的に同じである. □

定理 4.14. X_n, X, Y_n, $Y\,(n = 1, 2, \ldots)$ は確率変数列とする. $g : \mathbb{R} \to \mathbb{R}$ を連続関数とし, c を定数とする.

(1) $X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y \Rightarrow X_n + Y_n \xrightarrow{\mathrm{P}} X + Y\,(n \to \infty)$ である.

(2) $X_n \xrightarrow{\mathrm{q.m.}} X$ かつ $Y_n \xrightarrow{\mathrm{q.m.}} Y \Rightarrow X_n + Y_n \xrightarrow{\mathrm{q.m.}} X + Y\,(n \to \infty)$ である.

(3) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{\mathrm{P}} c \Rightarrow X_n + Y_n \rightsquigarrow X + c\,(n \to \infty)$ である.

(4) $X_n \xrightarrow{\mathrm{P}} X$ かつ $Y_n \xrightarrow{\mathrm{P}} Y \Rightarrow X_n Y_n \xrightarrow{\mathrm{P}} XY\,(n \to \infty)$ である.

(5) $X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{\mathrm{P}} c \Rightarrow X_n Y_n \rightsquigarrow cX\,(n \to \infty)$ である.

(6) $X_n \xrightarrow{\mathrm{P}} X \Rightarrow g(X_n) \xrightarrow{\mathrm{P}} g(X)\,(n \to \infty)$ である.

(7) $X_n \rightsquigarrow X \Rightarrow g(X_n) \rightsquigarrow g(X)\,(n \to \infty)$ である.

Proof. 節 4.5.5 で示す. □

定義 4.15. (1) $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ を確率変数列とし, すべての $n \geq 1$ に対して, $\mathsf{Pr}(Y_n > 0) = 1$ とする. $X_n = o_P(Y_n)$ であるとは

$$\frac{X_n}{Y_n} \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

が成り立つときをいう.
(2) $X_n = O_p(Y_n)$ であるとは, 任意の $\epsilon > 0$ に対して, ある $K_\epsilon > 0$ が存在して

$$\mathsf{Pr}\left(\left|\frac{X_n}{Y_n}\right| > K_\epsilon\right) < \epsilon$$

が成り立つときをいう.

注意 4.16. $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$, $\{Z_n\}_{n=1}^{\infty}$ は確率変数列とする. 次の関係式が成り立つことが証明できる.
(1) $X_n = O_P(1)$ かつ $Y_n = O_P(1)$ ならば, $X_n + Y_n = O_P(1)$ である.
(2) $X_n = o_P(1)$ かつ $Y_n = o_P(1)$ ならば, $X_n + Y_n = o_P(1)$ である.
(3) $X_n = O_P(1)$ かつ $Z_n = O_P(1)$ ならば, $X_n \cdot Z_n = O_P(1)$ である.
(4) $X_n = O_P(1)$ かつ $Z_n = o_P(1)$ ならば, $X_n \cdot Z_n = o_P(1)$ である.
(5) $h : \mathbb{R} \longrightarrow \mathbb{R}$ を連続関数とする. このとき, $X_n = O_P(1)$ ならば, $h(X_n) = O_P(1)$ である.

問 4.1. 注意 4.16(1) – (5) を証明せよ.

4.2 大数の法則

定理 4.17. (大数の弱法則) $X_1, X_2, \ldots$ は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の i.i.d. 確率変数列とする. $\mathsf{E}\big[\big|X_1\big|\big] < \infty$ のとき

$$\overline{X}_n := \frac{1}{n}\big(X_1 + X_2 + \cdots + X_n\big) \xrightarrow{\mathrm{P}} \mu, \qquad (n \to \infty)$$

が成立する. ただし, $\mu = \mathsf{E}[X_1]$ である.

Proof. より強い条件 $\mathsf{E}\big[X_1^2\big] < \infty$ のもとで定理の主張を証明する. 定理の仮定のもとの証明は節 4.5.1 で与える. $\mathsf{Var}[X_1] = \sigma^2$ とおく. Chebyshev の不等式 (系 3.3) より

$$\mathsf{Pr}\big(\big|\overline{X}_n - \mu\big| \geq \epsilon\big) \leq \frac{\mathsf{Var}\big[\overline{X}_n\big]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

よりわかる.

仮定のもとでの証明は, 節 4.5.1 を参照せよ. □

大数の強法則を述べる前に, 定理の証明に必要な補題を述べる.

補題 4.18. $X_1, X_2, \ldots$ は非負値確率変数列で, 条件

$$\sum_{n=1}^{\infty} \mathsf{E}[X_n] < \infty \tag{4.3}$$

をみたすとする. このとき

$$\mathsf{Pr}\left(\omega \in \Omega; \sum_{n=1}^{\infty} X_n(\omega) < \infty\right) = 1$$

が成り立つ.

Proof. 証明は節 4.5.2 で行う. □

定理 4.19. (大数の強法則) $X_1, X_2, \ldots$ は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の i.i.d. 確率変数列とし, $\mathsf{E}\big[\big|X_1\big|^4\big] < \infty$ とする. このとき

$$\overline{X}_n := \frac{1}{n}\big(X_1 + X_2 + \cdots + X_n\big) \xrightarrow{\text{a.s.}} \mu, \qquad (n \to \infty) \tag{4.4}$$

が成立する. ただし, $\mu = \mathsf{E}[X_1]$ である.

Proof. まず

$$\mathsf{E}\big[\big|X_1\big|^4\big] =: K < \infty, \quad T_n = X_1 + X_2 + \cdots + X_n, \quad \overline{X}_n = \frac{T_n}{n}$$

とおく.

最初に, $\mu = 0$ として (4.4) を示す. 以下の事象の包含関係に注意する. すなわち

$$\left\{\omega \in \Omega;\, \sum_{n=1}^{\infty}\big(\overline{X}_n(\omega)\big)^4 < \infty\right\} \subset \left\{\omega \in \Omega;\, \lim_{n\to\infty} \overline{X}_n(\omega) = 0\right\}$$

が成立する. このことより

$$\mathsf{Pr}\left(\sum_{n=1}^{\infty} \overline{X}_n^4 < \infty\right) = 1 \tag{4.5}$$

がわかると

$$1 = \mathsf{Pr}\left(\sum_{n=1}^{\infty} \overline{X}_n^4 < \infty\right) \le \mathsf{Pr}\left(\lim_{n\to\infty} \overline{X}_n = 0\right)$$

となり, $\mu = 0$ のときに (4.4) がわかる.

以下では $\mu = 0$ として (4.5) を示す. そのために多項定理を用いて, T_n を展開する. すると

$$\begin{aligned}&\big(X_1 + \cdots + X_n\big)^4 \\ &\quad = \sum_{\substack{\ell_1+\cdots+\ell_n=4 \\ \ell_j\ge 0\,(j=1,\ldots,n)}} \frac{4!}{\ell_1! \times \ell_2! \times \cdots \times \ell_n!} X_1^{\ell_1} \times X_2^{\ell_2} \times \cdots \times X_n^{\ell_n}\end{aligned}$$

となる. ただし $\ell_1, \ell_2, \ldots, \ell_n$ は 0 以上の整数である. ここで, $X_1^{\ell_1}, \ldots, X_n^{\ell_n}$ は独立であることに注意すると

$$\mathsf{E}\big[T_n^4\big] = \sum_{\substack{\ell_1+\ell_2+\cdots+\ell_n=4 \\ \ell_j\ge 0\,(j=1,2,\ldots,n)}} \frac{4!}{\ell_1!\ell_2! \times \cdots \times \ell_n!} \mathsf{E}\big[X_1^{\ell_1}\big] \times \mathsf{E}\big[X_2^{\ell_2}\big] \times \cdots \times \mathsf{E}\big[X_n^{\ell_n}\big]$$

がわかる. さらに, $\mathsf{E}\big[X_k\big] = \mu = 0\,(k = 1, 2, \ldots, n)$ と $\dfrac{4!}{2!2!} = 6$ を用いると

$$\mathsf{E}\big[T_n^4\big] = \sum_{k=1}^{n} \mathsf{E}\big[X_k^4\big] + 6 \sum_{1\le k<\ell\le n} \mathsf{E}\big[X_k^2\big]\,\mathsf{E}\big[X_\ell^2\big]$$

を得る. ここで $\mathsf{E}\big[X_k^4\big] = K\,(k = 1, 2, \ldots, n)$ と Cauchy-Schwarz の不等式 (定理 3.10) から

$$\mathsf{E}\big[X_k^2\big] \le \sqrt{\mathsf{E}\big[X_k^4\big]} = \sqrt{K}$$

となることに注意すると

$$\mathsf{E}\big[T_n^4\big] \le nK + 6 \sum_{1\le k<\ell\le n} \sqrt{K}\,\sqrt{K} = nK + 3n(n-1)K \le 3Kn^2$$

を得る [2]. よって

$$\sum_{n=1}^{\infty} \mathsf{E}\big[\overline{X}_n^4\big] = \sum_{n=1}^{\infty} \frac{1}{n^4}\mathsf{E}\big[T_n^4\big] \le 3K \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

となる. したがって 補題 4.18 から (4.5) が成立することがわかる.

つぎに $\mu \neq 0$ の場合を示す. $Y_k = X_k - \mu\,(k = 1, 2, \ldots, n)$ とおくと

$$\mathsf{E}\big[Y_k^4\big] \le \mathsf{E}\big[\big\{|X_k| + |\mu|\big\}^4\big] \le 8\mathsf{E}\big[|X_k|^4 + |\mu|^4\big] \le 8\big(K + \mu^4\big) < \infty$$

が成り立つ. したがってこの定理の証明の前半部分で得られた結果から

$$\mathsf{Pr}\left(\lim_{n\to\infty} \frac{1}{n}\big(Y_1 + Y_2 + \cdots + Y_n\big) = 0\right) = 1$$

となる. 最後に, $\forall\omega \in \Omega$ に対して

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n}\big(Y_1(\omega) + Y_2(\omega) + \cdots + Y_n(\omega)\big) = 0 \\ &\qquad \Leftrightarrow \lim_{n\to\infty} \frac{1}{n}\big(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)\big) = \mu \end{aligned}$$

に注意すればよい. □

注意 4.20. 大数の法則は $\mathsf{E}\big[|X_1|\big] < \infty$ で成立する. この条件下での大数の強法則の証明は節 B.1.2 に記した. □

系 4.21 (Weierstrass の近似定理)**.** 閉区間 $[0, 1]$ 上の任意の連続関数 f は多項式の極限として表すことができる. 特に, $f(0) = f(1) = 0$ のとき [3], $n \in \mathbb{N}$ に対して

$$B_n(x) = \sum_{j=1}^{n} f\left(\frac{j}{n}\right)\left(\begin{array}{c} n \\ j \end{array}\right) x^j (1-x)^{n-j}$$

とおいたとき

$$f(x) = \lim_{n\to\infty} B_n(x) \tag{4.6}$$

が成立する.

---

[2] 上式の最初の等号は $\sum_{1\le k<\ell\le n}(\,\cdot\,)$ で足し合わせる項数が $\dfrac{n(n-1)}{2}$ 項あることからわかる.

[3] $f(0) = a$, $f(1) = b$ に対して, $\widetilde{f}(x) = f(x) - (b-a)x - a$ とおくと $\widetilde{f}$ は $[0, 1]$ 上の連続関数で $\widetilde{f}(0) = \widetilde{f}(1) = 0$ となるので, 左記の設定は一般性を失わない仮定である.

Proof. $x = 0, 1$ のとき (4.6) は明らかなので, $0 < x < 1$ に対して, (4.6) が成立することを示す. X_j $(j = 1, 2, \ldots)$ は i.i.d. 確率変数列で

$$\mathsf{Pr}(X_j = 1) = x = 1 - \mathsf{Pr}(X_j = 0)$$

をみたすとする. いま

$$\mathsf{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = x$$

に注意して, Chebyshev の不等式 (系 3.3) を用いると, 任意の $\epsilon > 0$ に対して

$$\begin{aligned}
&\mathsf{Pr}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - x\right| \geq \epsilon\right) \\
&\quad \leq \frac{\mathsf{Var}\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)}{\epsilon^2} \\
&\quad = \frac{\dfrac{1}{n^2}\big(\mathsf{Var}(X_1) + \mathsf{Var}(X_2) + \cdots + \mathsf{Var}(X_n)\big)}{\epsilon^2} \\
&\quad = \frac{x(1-x)}{n\epsilon^2}
\end{aligned}$$

となる. したがって

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} x \qquad (n \to \infty)$$

がわかる. さらに, 定理 4.9(2) より

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \rightsquigarrow x \qquad (n \to \infty)$$

を得る. f は有界連続関数なので, 定理 4.12(2) から

$$\lim_{n\to\infty} \mathsf{E}\left[f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)\right] = f(x)$$

となる. しかし, $S := X_1 + X_2 + \cdots + X_n \sim \mathsf{Bino}(n, x)$ なので

$$\begin{aligned}
\mathsf{E}\left[f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)\right] &= \mathsf{E}\left[f\left(\frac{S}{n}\right)\right] \\
&= \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\binom{n}{j} x^j (1-x)^{n-j} \\
&= B_n(x)
\end{aligned}$$

がわかる. よって, 主張は証明された. □

4.3 中心極限定理

定理 4.22. $X_1, X_2, \dots$ を i.i.d. 確率変数列とし, $\mathsf{E}[X_1] = \mu$, $\mathsf{Var}[X_1] = \sigma^2\,(0 < \sigma < \infty)$ とする. このとき

$$Z_n := \frac{\overline{X}_n - \mu}{\sqrt{\mathsf{Var}\big[\overline{X}_n\big]}} = \frac{\sqrt{n}\big(\overline{X}_n - \mu\big)}{\sigma} \rightsquigarrow Z \qquad (n \to \infty)$$

が成り立つ. ただし $\overline{X}_n = \dfrac{1}{n}\big(X_1 + X_2 + \cdots + X_n\big)$, $Z \sim \mathsf{N}(0,\, 1)$ である.

Proof. 中心極限定理の証明は補遺 B.2 でする. □

例 4.23. $X_1, X_2, \dots$ を i.i.d. 確率変数列とし, $\mathsf{E}\big[X_1\big] = \mu$, $\mathsf{Var}\big[X_1\big] = \sigma^2\,(0 < \sigma < \infty)$ とする. このとき中心極限定理 (定理 4.22) より

$$Z_n = \frac{\sqrt{n}\big(\overline{X}_n - \mu\big)}{\sigma} \rightsquigarrow \mathsf{N}(0,\, 1) \qquad (n \to \infty) \tag{4.7}$$

が成立する. ここで

$$S_n^2 := \frac{1}{n-1}\sum_{j=1}^{n}\big(X_j - \overline{X}_n\big)^2, \quad S_n = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\big(X_j - \overline{X}_n\big)^2}$$

とおく. すると

$$S_n^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{j=1}^{n}\big(X_j - \mu\big)^2 - \big(\overline{X}_n - \mu\big)^2\right\}$$

と書き直せる. 設定から $\mathsf{E}[(X_1 - \mu)^2] = \sigma^2 < \infty$ なので, 大数の法則 (定理 4.17) より

$$\frac{1}{n}\sum_{j=1}^{n}\big(X_j - \mu\big)^2 \xrightarrow{\mathrm{P}} \mathsf{E}\big[\big(X_1 - \mu\big)^2\big] = \sigma^2, \tag{4.8}$$

$$\overline{X}_n \xrightarrow{\mathrm{P}} \mu \qquad (n \to \infty) \tag{4.9}$$

である. (4.9) に対して, 定理 4.14(6)($g(x) = (x - \mu)^2$) を用いると

$$\big(\overline{X}_n - \mu\big)^2 \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{4.10}$$

となる. さらに, (4.8), (4.10) と定理 4.14(3)(5) より

$$S_n^2 \xrightarrow{\mathrm{P}} \sigma^2 \qquad (n \to \infty)$$

である. 再度, 定理 4.14(6) を用いると

$$\frac{S_n}{\sigma} \xrightarrow{\mathrm{P}} 1 \qquad (n \to \infty) \tag{4.11}$$

がわかる. 最後に (4.7), (4.11) と定理 4.14(5) より

$$\frac{\sqrt{n}\big(\overline{X}_n - \mu\big)}{S_n} \rightsquigarrow \mathsf{N}(0,\, 1) \qquad (n \to \infty)$$

を得る.

□

定理 4.24. $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ を i.i.d. 確率ベクトル列とする. ただし $j = 1, 2, \ldots$ に対して

$$\boldsymbol{X}_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{dj} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix} = \begin{pmatrix} \mathsf{E}[X_{1j}] \\ \mathsf{E}[X_{2j}] \\ \vdots \\ \mathsf{E}[X_{dj}] \end{pmatrix}$$

とし,

$$\mathsf{Var}[\boldsymbol{X}_1] = \boldsymbol{\Sigma} = \begin{pmatrix} \mathsf{Var}[X_{11}] & \mathsf{Cov}[X_{11},\, X_{21}] & \cdots & \mathsf{Cov}[X_{11},\, X_{d1}] \\ \mathsf{Cov}[X_{21},\, X_{11}] & \mathsf{Var}[X_{21}] & \cdots & \mathsf{Cov}[X_{21},\, X_{d1}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_{d1},\, X_{11}] & \mathsf{Cov}[X_{d1},\, X_{21}] & \cdots & \mathsf{Var}[X_{d1}] \end{pmatrix}$$

は正定値とする. このとき任意の $\boldsymbol{c} \in \mathbb{R}^d$ に対して

$$\sqrt{n}\boldsymbol{c}^\top\big(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}\big(0,\, \boldsymbol{c}^\top\boldsymbol{\Sigma}\boldsymbol{c}\big) \qquad (n \to \infty)$$

が成立する. このことを

$$\sqrt{n}\big(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}_d\big(\boldsymbol{0},\, \boldsymbol{\Sigma}\big) \qquad (n \to \infty)$$

と書くことにする.

Proof. 証明は後で行う. □

4.4 デルタ法

$\{Y_n\}$ を確率変数列とする. Y_n の極限分布が正規分布のとき滑らかな実数値関数 $g : \mathbb{R} \to \mathbb{R}$ に対して $g(Y_n)$ の極限分布を求めよう.

定理 4.25. (デルタ法) $Y_1, Y_2, \ldots$ を確率変数列とし

$$\frac{\sqrt{n}\big(Y_n - \mu\big)}{\sigma} \rightsquigarrow \mathsf{N}(0,\, 1) \qquad (n \to \infty)$$

とし, g は $x = \mu$ の近傍で連続微分可能な関数で $\overset{\bullet}{g}(\mu) \neq 0$ とする. ただし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$, $\overset{\bullet}{g}(t) = \dfrac{\mathrm{d}g}{\mathrm{d}t}(t)$ である. このとき

$$\frac{\sqrt{n}\big(g(Y_n) - g(\mu)\big)}{\big|\overset{\bullet}{g}(\mu)\big|\sigma} \rightsquigarrow \mathsf{N}(0,\, 1)$$

が成立する. すなわち

$$Y_n \approx \mathsf{N}\left(\mu,\, \frac{\sigma^2}{n}\right) \;\Rightarrow\; g(Y_n) \approx \mathsf{N}\left(g(\mu),\, \big(\overset{\bullet}{g}(\mu)\big)^2\frac{\sigma^2}{n}\right)$$

である. ただし「≈」は「分布が近似できる」の意味である.

Proof. 節 4.5.5 で証明をする. □

例 4.26. $X_1, X_2, \ldots$ を i.i.d. 確率変数列とし, $\mathsf{E}\big[X_1\big] = \mu$, $\mathsf{Var}\big[X_1\big] = \sigma^2$ $(0 < \sigma < \infty)$ とする. このとき中心極限定理 (定理 4.22) より

$$\frac{\sqrt{n}\big(\overline{X}_n - \mu\big)}{\sigma} \rightsquigarrow \mathsf{N}(0,\, 1) \qquad (n \to \infty)$$

となる. いま

$$W_n = e^{\overline{X}_n}$$

とおく. したがって $g(x) = e^x$ $(x \in \mathbb{R})$ とすれば $\overset{\bullet}{g}(x) = e^x$ となる. よってデルタ法 (定理 4.25) より

$$\frac{\sqrt{n}\big(W_n - e^{\mu}\big)}{e^{\mu}\sigma} \rightsquigarrow \mathsf{N}(0,\, 1) \qquad (n \to \infty)$$

となる. よって

$$W_n \approx \mathsf{N}\left(e^{\mu},\, \frac{e^{2\mu}\sigma^2}{n}\right)$$

がわかる. □

定理 4.27. (多次元デルタ法) $\boldsymbol{Y}_n = (Y_{1n}, Y_{2n}, \ldots, Y_{dn})^\top$ $(n = 1, 2, \ldots)$ を確率ベクトル列とし

$$\sqrt{n}\big(\boldsymbol{Y}_n - \boldsymbol{\mu}\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}_d,\, \boldsymbol{\Sigma}\big) \qquad (n \to \infty)$$

とする．ただし $\boldsymbol{\mu} \in \mathbb{R}^d$ で $\boldsymbol{\Sigma}$ は $d \times d$ の正定値対称行列とする．関数 $g : \mathbb{R}^d \to \mathbb{R}$ は滑らかで

$$\nabla g(\boldsymbol{y}) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \frac{\partial g}{\partial y_2} \\ \vdots \\ \frac{\partial g}{\partial y_d} \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix}, \quad \nabla_{\boldsymbol{\mu}} := \nabla g(\boldsymbol{y})\Big|_{\boldsymbol{y}=\boldsymbol{\mu}}$$

とする．このとき

$$\sqrt{n}\big(g(\boldsymbol{Y}_n) - g(\boldsymbol{\mu})\big) \rightsquigarrow \mathsf{N}\big(0,\, \nabla_{\boldsymbol{\mu}}^{\top} \boldsymbol{\Sigma} \nabla_{\boldsymbol{\mu}}\big) \qquad (n \to \infty)$$

となる．

例 4.28.

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix}$$

は i.i.d. 確率ベクトル列で

$$\mathsf{E}\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{\mu},$$
$$\mathsf{Var}\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \boldsymbol{\Sigma}$$

とする．ただし $\boldsymbol{\Sigma}$ は 2×2 の正定値対称行列である．

$$\overline{X}_1 = \frac{1}{n}\sum_{j=1}^{n} X_{1j}, \quad \overline{X}_2 = \frac{1}{n}\sum_{j=1}^{n} X_{2j}$$

とし

$$Y_n = \overline{X}_1 \overline{X}_2, \quad g(s_1,\, s_2) = s_1 s_2 \,(s_1,\, s_2 \in \mathbb{R})$$

とおく．中心極限定理 (定理 4.24) より

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \overline{X}_2 - \mu_2 \end{pmatrix} \rightsquigarrow \mathsf{N}_2\big(\mathbf{0}_2,\, \boldsymbol{\Sigma}\big) \qquad (n \to \infty)$$

となる．いま

$$\nabla g(\boldsymbol{s}) = \begin{pmatrix} \dfrac{\partial g}{\partial s_1} \\ \dfrac{\partial g}{\partial s_2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_1 \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$

であるので

$$\nabla_{\boldsymbol{\mu}}^{\top} \boldsymbol{\Sigma} \nabla_{\boldsymbol{\mu}} = \mu_2^2 \sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2 \sigma_{22}$$

である. したがって

$$\sqrt{n}\big(\overline{X}_1\overline{X}_2 - \mu_1\mu_2\big) \rightsquigarrow \mathsf{N}\big(0,\, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}\big) \qquad (n \to \infty)$$

がわかる. □

4.5 延期された証明

4.5.1 定理 4.17 の証明

任意の $\delta > 0$ と n に対して

$$Y_j := Y_j(n) = \begin{cases} X_j & (|X_j| \le \delta n) \\ 0 & (|X_j| > \delta n) \end{cases}$$

$$Z_j := Z_j(n) = \begin{cases} 0 & (|X_j| \le \delta n) \\ X_j & (|X_j| > \delta n) \end{cases} \quad ; \quad (j = 1,\, 2,\, \ldots,\, n)$$

と定める. 明らかに

$$X_j = Y_j + Z_j, \qquad (j = 1,\, 2,\, \ldots,\, n)$$

である. 以下では, $X_j\,(j = 1,\, 2,\, \ldots,\, n)$ は連続型確率変数とし, 共通の p.d.f. p を持つとして, ① ∼ ④ の段階を経て議論を進めていく.
① $\mathbb{E}[Y_1] \xrightarrow{n\to\infty} \mu$ の証明: このとき

$$\begin{aligned} \mathsf{Var}[Y_j] = \mathsf{Var}[Y_1] &= \mathsf{E}[Y_1^2] - \big(\mathsf{E}[Y_1]\big)^2 \le \mathsf{E}[Y_1^2] \\ &= \mathsf{E}\big[X_1^2 \mathbb{1}_{[-\delta n,\, \delta n]}(X_1)\big] \\ &= \int_{-\infty}^{\infty} x^2 \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x = \int_{-\delta n}^{\delta n} x^2 \mathsf{p}(x)\,\mathrm{d}x \\ &\le \delta n \int_{-\delta n}^{\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x \le \delta n \int_{-\infty}^{\infty} |x|\mathsf{p}(x)\,\mathrm{d}x \\ &= \delta n \mathsf{E}\big[|X_1|\big] \end{aligned}$$

となる. すなわち

$$\mathsf{Var}[Y_j] \le \delta n \mathsf{E}[|X_1|] \qquad (j = 1,\, 2,\, \ldots,\, n) \tag{4.12}$$

である. 次に

$$\mathsf{E}[Y_j] = \mathsf{E}[Y_1] = \mathsf{E}[X_1 \mathbb{1}_{[-\delta n,\, \delta n]}(X_1)] = \int_{-\infty}^{\infty} x \mathbb{1}_{[-\delta n,\, \delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x$$

と書ける. ここで

$$\left|x\mathbb{1}_{[-\delta n,\,\delta n]}(x)\mathsf{p}(x)\right| \le |x|\mathsf{p}(x), \qquad x\mathbb{1}_{[-\delta n,\,\delta n]}(x)\mathsf{p}(x) \overset{n\to\infty}{\longrightarrow} x\mathsf{p}(x)$$

かつ

$$\int_{-\infty}^{\infty} |x|\mathsf{p}(x)\,\mathrm{d}x < \infty$$

であることに注意して, 優収束定理を用いると

$$\mathsf{E}[Y_1] = \int_{-\infty}^{\infty} x\mathbb{1}_{[-\delta n,\,\delta n]}(x)\mathsf{p}(x)\,\mathrm{d}x \overset{n\to\infty}{\longrightarrow} \int_{-\infty}^{\infty} x\mathsf{p}(x)\,\mathrm{d}x$$

となることがわかる. したがって

$$\mathsf{E}[Y_1] \overset{n\to\infty}{\longrightarrow} \mu \tag{4.13}$$

である.

② $\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \ge 2\epsilon\right) \le \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|]$ の証明: 以上のことを踏まえると任意の $\epsilon > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \ge \epsilon\right) &= \mathsf{Pr}\left(\left|\sum_{j=1}^{n} Y_j - \mathsf{E}\left[\sum_{j=1}^{n} Y_j\right]\right| \ge n\epsilon\right) \\
&\le \frac{1}{n^2\epsilon^2}\mathsf{Var}\left[\sum_{j=1}^{n} Y_j\right] \qquad (\because \text{系 } 3.3) \\
&= \frac{n}{n^2\epsilon^2}\mathsf{Var}[Y_1] \le \frac{n\delta n\mathsf{E}[|X_1|]}{n^2\epsilon^2} \qquad (\because (4.12)) \\
&= \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|]
\end{aligned}$$

となる. したがって

$$\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \ge \epsilon\right) \le \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] \tag{4.14}$$

となる. よって, (4.13) と (4.14) から, 十分大きな n に対して

$$\begin{aligned}
\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \ge 2\epsilon\right) &= \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1] + \mathsf{E}[Y_1] - \mu\right| \ge 2\epsilon\right) \\
&\le \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| + \left|\mathsf{E}[Y_1] - \mu\right| \ge 2\epsilon\right) \\
&\le \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mathsf{E}[Y_1]\right| \ge \epsilon\right) + \mathsf{Pr}\left(\left|\mathsf{E}[Y_1] - \mu\right| \ge \epsilon\right)
\end{aligned}$$

$$\leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|]$$

となる [4]. したがって

$$\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \geq 2\epsilon\right) \leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] \tag{4.15}$$

となる.

③ $\mathsf{Pr}\left(\sum_{j=1}^{n} Z_j \neq 0\right) \leq \delta$ の証明: さらに, $\mathsf{E}[|X_1|] < \infty$ なので, 任意の $\delta > 0$ と十分大きな n に対して

$$\int_{|x|>\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x < \delta^2$$

とできることに注意する. このことより, $j = 1, 2, \ldots, n$ に対して

$$\begin{aligned}\mathsf{Pr}(Z_j \neq 0) = \mathsf{Pr}(|Z_j| > \delta n) = \mathsf{Pr}(|X_j| > \delta n) &= \int_{-\infty}^{-\delta n} \mathsf{p}(x)\,\mathrm{d}x + \int_{\delta n}^{\infty} \mathsf{p}(x)\,\mathrm{d}x \\ &= \int_{|x|>\delta n} \mathsf{p}(x)\,\mathrm{d}x < \int_{|x|>\delta n} \frac{|x|}{\delta n}\mathsf{p}(x)\,\mathrm{d}x = \frac{1}{\delta n}\int_{|x|>\delta n} |x|\mathsf{p}(x)\,\mathrm{d}x \\ &< \frac{\delta^2}{\delta n} = \frac{\delta}{n}\end{aligned} \tag{4.16}$$

となる.

④ $\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| > 4\epsilon\right) \overset{n\to\infty}{\longrightarrow} 0$ の証明: したがって, (4.16) から

$$\mathsf{Pr}\left(\sum_{j=1}^{n} Z_j \neq 0\right) \leq \mathsf{Pr}\left(\bigcup_{j=1}^{n}\{Z_j \neq 0\}\right) \leq n\mathsf{Pr}(Z_1 \neq 0) \leq \delta \tag{4.17}$$

を得る. よって, (4.15) と (4.17) から, 十分大きな n に対して

$$\begin{aligned}\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \geq 4\epsilon\right) &= \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j + \frac{1}{n}\sum_{j=1}^{n} Z_j - \mu\right| \geq 4\epsilon\right) \\ &\leq \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| + \left|\frac{1}{n}\sum_{j=1}^{n} Z_j\right| \geq 4\epsilon\right) \\ &\leq \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \geq 2\epsilon\right) + \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Z_j\right| \geq 2\epsilon\right) \\ &\leq \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| \geq 2\epsilon\right) + \mathsf{Pr}\left(\sum_{j=1}^{n} Z_j \neq 0\right)\end{aligned}$$

[4] (4.13) から n を十分大きくとると $|\mathsf{E}[Y_1] - \mu| \geq \epsilon$ となるので, $\mathsf{Pr}(|\mathsf{E}[Y_1] - \mu| \geq \epsilon) = 0$ となる.

$$\leq \frac{\delta}{\epsilon^2}\mathsf{E}[|X_1|] + \delta$$

となる. ここで, δ を ϵ^3 とおきなおせば, 十分大きな n に対して

$$\mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \geq 4\epsilon\right) \leq \epsilon\mathsf{E}[|X_1|] + \epsilon^3$$

となる. 上式は任意の $\epsilon > 0$ で成立したので, 定理の主張は証明された. □

4.5.2 補題 4.18 の証明

背理法で証明する. そのために

$$\mathsf{Pr}\left(\sum_{k=1}^{\infty} X_k < \infty\right) < 1$$

を仮定する. さらに事象 F と N を

$$F := \left\{\omega \in \Omega;\, \sum_{k=1}^{\infty} X_k(\omega) < \infty\right\}, \qquad N := \left\{\omega \in \Omega;\, \sum_{k=1}^{\infty} X_k(\omega) = \infty\right\}$$

で定める. このとき

$$\Omega = F \cup N \quad \text{かつ} \quad F \cap N = \varnothing$$

が成り立つ. したがって

$$1 = \mathsf{Pr}(\Omega) = \mathsf{Pr}\big(F \cup N\big) = \mathsf{Pr}(F) + \mathsf{Pr}(N) \qquad (\because \mathsf{Pr}\text{ の加法性})$$

がわかる. この関係式と背理法の仮定 $\mathsf{Pr}(F) < 1$ から $\mathsf{Pr}(N) > 0$ となる. 一方, $\omega \in N$ のとき

$$\sum_{k=1}^{\infty} X_k(\omega) = \infty$$

であり, $\omega \in N^c$ のとき

$$0 \leq \sum_{k=1}^{\infty} X_k(\omega) < \infty$$

である. したがって, 任意の非負の実数 $r \geq 0$ に対して

$$r\mathbb{1}_N(\omega) \leq \sum_{k=1}^{\infty} X_k(\omega) \qquad (\forall \omega \in \Omega) \tag{4.18}$$

が成り立つ. (4.18) の両辺の期待値をとる. すると 定義 2.10(4), (4.3) と (4.18) から

$$r \times \mathsf{Pr}(N) = \mathsf{E}[r\mathbb{1}_N] \leq \mathsf{E}\left[\sum_{k=1}^{\infty} X_k\right] = \sum_{k=1}^{\infty} \mathsf{E}\big[X_k\big] < \infty \tag{4.19}$$

を得る. (4.19) の最後の等号は単調収束定理からわかる. ここで $\mathsf{Pr}(N) > 0$ であり, $r \geq 0$ は任意の実数だったので, $r \to \infty$ とすれば, (4.19) の最左辺は $+\infty$ となるので, 矛盾が生じる. したがって, $\mathsf{Pr}(F) = 1$ が成り立つ. □

4.5.3 定理 4.9 の証明

(1) $X_n \overset{qm}{\to} X$ とする. $\epsilon > 0$ を固定する. このとき Markov の不等式 (定理 3.1) より

$$\mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) = \mathsf{Pr}\big(\big|X_n - X\big|^2 \geq \epsilon^2\big) \leq \frac{\mathsf{E}\big[\big|X_n - X\big|^2\big]}{\epsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

がわかる.
(2) $X_n \overset{\mathrm{P}}{\to} X$ とする. $\epsilon > 0$ を固定し, x と $x \pm \epsilon$ を X の c.d.f. $\mathsf{F}(x)$ の連続点[5] とする. このとき

$$\begin{aligned}
\mathsf{F}_n(x) &= \mathsf{Pr}\big(X_n \leq x\big) \\
&= \mathsf{Pr}\big(X_n \leq x,\, X < x + \epsilon\big) + \mathsf{Pr}\big(X_n \leq x,\, X \geq x + \epsilon\big) \\
&\leq \mathsf{Pr}\big(X \leq x + \epsilon\big) + \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) && (4.20) \\
&= \mathsf{F}(x + \epsilon) + \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) && (4.21)
\end{aligned}$$

となる. 同様に

$$\begin{aligned}
\mathsf{F}(x - \epsilon) &= \mathsf{Pr}\big(X \leq x - \epsilon\big) \\
&= \mathsf{Pr}\big(X_n \leq x - \epsilon,\, X_n < x\big) + \mathsf{Pr}\big(X \leq x - \epsilon,\, X_n \geq x\big) \\
&\leq \mathsf{Pr}\big(X_n \leq x\big) + \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) && (4.22) \\
&= \mathsf{F}_n(x) + \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) && (4.23)
\end{aligned}$$

となる. (4.21) と (4.23) を合わせると

$$\mathsf{F}(x - \epsilon) - \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big) \leq \mathsf{F}_n(x) \leq \mathsf{F}(x + \epsilon) + \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big)$$

がわかる. 上の式の辺々で $n \to \infty$ とすれば

$$\mathsf{F}(x - \epsilon) \leq \liminf_{n\to\infty} \mathsf{F}_n(x) \leq \limsup_{n\to\infty} \mathsf{F}_n(x) \leq \mathsf{F}(x + \epsilon)$$

[5] X の分布関数の不連続点は高々加算個なので, このように点を取れることに注意せよ.

となる. ここで $\epsilon \to 0$ とし x を $\mathsf{F}(x)$ の連続点とすれば

$$\lim_{n\to\infty} \mathsf{F}_n(x) = \mathsf{F}(x)$$

がわかる.
(3) $X_n \rightsquigarrow X$ かつある定数 c があって $\mathsf{Pr}(X = c) = 1$ とする. $c \pm \epsilon$ を $\mathsf{F}(x)$ の連続点になるようにして $\epsilon > 0$ を固定する [6]. このとき

$$\begin{aligned}
\mathsf{Pr}\big(\big|X_n - c\big| \geq \epsilon\big) &= \mathsf{Pr}\big(X_n \leq c - \epsilon\big) + \mathsf{Pr}\big(X_n \geq c + \epsilon\big) \\
&= \mathsf{Pr}\big(X_n \leq c - \epsilon\big) + \mathsf{Pr}\big(X_n \geq c + \epsilon\big) \\
&= \mathsf{F}_n(c - \epsilon) + 1 - \mathsf{F}_n(c + \epsilon) \\
&\to \mathsf{F}(c - \epsilon) + 1 - \mathsf{F}(c + \epsilon) \quad (n \to \infty) \\
&= 0 + 1 - 1 = 0
\end{aligned}$$

となる. □

問 4.2. (4.20) と (4.22) を確認せよ. すなわち

$$\begin{aligned}
\mathsf{Pr}\big(X_n \leq x,\, X \geq x + \epsilon\big) &\leq \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big), \\
\mathsf{Pr}\big(X \leq x - \epsilon,\, X_n \geq x\big) &\leq \mathsf{Pr}\big(\big|X_n - X\big| \geq \epsilon\big)
\end{aligned}$$

を示せ.

4.5.4 定理 4.12 の証明

①(1) ⇒ (2), ② (2) ⇒ (3), ③(2) ⇒ (4), ④(4) ⇒ (2)⑤ (3) ⇒ (5)⑥ (5) ⇔ (6)⑦(5) + (6) ⇒ (7) ⑧(7) ⇒ (1) の順に証明をしていく.
①(1) ⇒ (2) の証明: g を任意の有界連続関数とする. 一般性を失わずに $|g(x)| \leq 1/2\,(x \in \mathbb{R})$ と仮定してよい. 任意の $\epsilon > 0$ に対して, 十分大きな有界閉区間 K をとると

$$\mathsf{Pr}(X \in K^c) \leq \epsilon \tag{4.24}$$

とできる. そこで, この K を $m\,(m \in \mathbb{N})$ 個の互いに素な区間 $K_1 = [a_1, b_1]$, $K_j = (a_j, b_j]\,(j = 2, 2, \ldots, m); a_j, b_j \in \mathbb{R})$ に直和分解する. すなわち, $K = \bigcup_{j=1}^{m} K_j$ となっている. このとき, m を十分大きく取ると

$$\max_{j\in\{1,2,\ldots,m\}} \max_{x,\,y\in K_j} |g(x) - g(y)| \leq \epsilon \tag{4.25}$$

[6] $\mathsf{F}(x)$ は有界非減少関数なので, $\mathsf{F}(x)$ の不連続点は高々可算個である. よってこのように ϵ を取ることができる.

となるように $\{K_j\}_{j=1}^m$ をとることができる. さらに, $\{a_j, b_j\}_{j=1}^m$ は X の分布関数 F の連続点[7] に取る. このとき, (1) より

$$\mathsf{Pr}(X_n \in K) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}(X \in K) \tag{4.26}$$

となる. つぎに, 各 K_j から 1 点 x_j を任意に選び, 関数 g_ϵ を

$$g_\epsilon(x) = \sum_{j=1}^m g(x_j)\mathbb{1}_{K_j}(x) \qquad (x \in \mathbb{R})$$

と定める. すると

$$\max_{x\in K} \big|g(x) - g_\epsilon(x)\big| \le \epsilon \tag{4.27}$$

となる. したがって, (4.24) と (4.27) に注意すると

$$\begin{aligned}
&\big|\mathsf{E}[g(X)] - \mathsf{E}[g_\epsilon(X)]\big| \\
&\quad= \big|\mathsf{E}[\mathbb{1}_K(X)g(X)] + \mathsf{E}[\mathbb{1}_{K^c}(X)g(X)] - \mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)] \\
&\qquad - \mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\big| \\
&\quad\le \big|\mathsf{E}[\mathbb{1}_K(X)g(X)] - \mathsf{E}[\mathbb{1}_K(X)g_\epsilon(X)]\big| + \big|\mathsf{E}[\mathbb{1}_{K^c}(X)g(X)] \\
&\qquad - \mathsf{E}[\mathbb{1}_{K^c}(X)g_\epsilon(X)]\big| \\
&\quad\le \mathsf{E}\big[\mathbb{1}_K(X)\big|g(X) - g_\epsilon(X)\big|\big] + \mathsf{E}\big[\mathbb{1}_{K^c}(X)\big|g(X) - g_\epsilon(X)\big|\big] \\
&\quad\le \underbrace{\max_{x\in K}|g(x) - g_\epsilon(x)|}_{\le\epsilon \quad \because (4.27)} \mathsf{E}\big[\mathbb{1}_K(X)\big] + \mathsf{E}\big[\mathbb{1}_{K^c}(X)\big|g(X) - g_\epsilon(X)\big|\big] \\
&\quad\le \epsilon\mathsf{Pr}(X \in K) + \mathsf{Pr}(X \in K^c) \\
&\quad\le 2\epsilon
\end{aligned} \tag{4.28}$$

となる. (4.26) から, n を十分大きく取ると

$$\big|\underbrace{\mathsf{Pr}(X_n \in K^c)}_{=1-\mathsf{Pr}(X_n\in K)} - \underbrace{\mathsf{Pr}(X \in K^c)}_{=1-\mathsf{Pr}(X\in K)}\big| = \big|\mathsf{Pr}(X \in K) - \mathsf{Pr}(X \in K)\big| \le \epsilon$$

とできるので

$$\mathsf{Pr}(X_n \in K^c) \le \mathsf{Pr}(X \in K^c) + \big|\mathsf{Pr}(X_n \in K^c) - \mathsf{Pr}(X \in K^c)\big| \le 2\epsilon$$

がわかる. 同様に

$$\big|\mathsf{E}[g(X_n)] - \mathsf{E}[g_\epsilon(X_n)]\big| \le \epsilon\mathsf{Pr}(X_n \in K) + \mathsf{Pr}(X_n \in K^c) \le 3\epsilon \tag{4.29}$$

[7] F の不連続点は高々可算個なので, このように取れることがわかる.

を得る. g_ϵ の作りかたから n を十分大きくとると

$$\begin{aligned}
&\big|\mathsf{E}[g_\epsilon(X)] - \mathsf{E}[g_\epsilon(X_n)]\big| \\
&\qquad \le \sum_{j=1}^{n} \big|\mathsf{Pr}(X \in K_j) - \mathsf{Pr}(X_n \in K_j)\big| \max_{j=\{1,2,\dots,m\}} |g(x_j)| \\
&\qquad \le \frac{\epsilon}{2} < \epsilon \qquad (4.30)
\end{aligned}$$

となる. よって, (4.28) − (4.30) を合わせると

$$\begin{aligned}
\big|\mathsf{E}[g(X_n)] - \mathsf{E}[g(X)]\big| &\le \underbrace{\big|\mathsf{E}[g(X_n)] - \mathsf{E}[g_\epsilon(X_n)]\big|}_{\le 3\epsilon} + \underbrace{\big|\mathsf{E}[g_\epsilon(X_n)] - \mathsf{E}[g_\epsilon(X)]\big|}_{\le \epsilon} \\
&\quad + \underbrace{\big|\mathsf{E}[g_\epsilon(X)] - \mathsf{E}[g(X)]\big|}_{\le 2\epsilon} \\
&\le 6\epsilon
\end{aligned}$$

がわかるので, 主張は証明された.

② (2) ⇒ (3) の証明: 明らか.

③ (2) ⇒ (4) の証明: 非負値連続関数 g と $M > 0$ に対して, $g_M(x) = \min\{g(x), M\} \ge 0\,(x \in \mathbb{R})$ と定める g_M は有界連続となる. よって

$$\liminf_{n\to\infty} \mathsf{E}[g(X_n)] \ge \liminf_{n\to\infty} \mathsf{E}[g_M(X_n)] = \mathsf{E}[g_M(X)] \qquad \big(\because (2)\big)$$

となる. 両辺で $M \to \infty$ とすることにより, 有界単調収束定理から (4) を得る.

④ (4) ⇒ (2) の証明: g を有界連続関数とする. するとある $M > 0$ が存在して, $|g| \le M$ とできる. すると $M \pm g \ge 0$ は非負値連続関数となるので, (4) を用いると (2) を得る.

⑤ (3) ⇒ (5) の証明: 開集合 O と $x \in \mathbb{R}$ に対して $g_M(x) \uparrow \mathbb{1}_O(x)\,(M \to \infty)$　となるような Lipschitz 連続関数列 $g_M \ge 0$ を取ることができるので, (3) から

$$\liminf_{n\to\infty} \mathsf{Pr}(X_n \in O) \ge \liminf_{n\to\infty} \mathsf{E}[g_M(X_n)] = \mathsf{E}[g_M(X)]$$

となる. この両辺で $M \to \infty$ とすると, 単調収束定理により (5) を得る.

⑥ (5) ⇔ (6) の証明: お互いの補集合をとればよい.

⑦ (5) + (6) ⇒ (7) の証明: B° を B の内部, $\mathrm{cl}(B)$ を B の閉包とすると $\partial B = \mathrm{cl}(B) \setminus B^\circ$ である. このことに注意すると

$$\begin{aligned}
\mathsf{Pr}\big(X \in B^\circ\big) &\le \liminf_{n\to\infty} \mathsf{Pr}\big(X_n \in B^\circ\big) \le \limsup_{n\to\infty} \mathsf{Pr}\big(X_n \in \mathrm{cl}(B)\big) \\
&\le \mathsf{Pr}\big(X_n \in \mathrm{cl}(B)\big)
\end{aligned}$$

となる. $\mathsf{Pr}(X \in \partial B) = 0$ より, 上式の右辺と左辺は等しいので, (7) が得られる.

⑧ (7) ⇒ (1) の証明: $B = (-\infty, x]$ ととればよい. □

4.5.5　定理 4.14 の証明

(6) の証明: 関数 g は $x_0\,(\in \mathbb{R})$ で連続なので, 任意の $\epsilon > 0$ に対して, $\exists\delta > 0$ があって

$$\forall x \in \mathbb{R}:\ |x - x_0| < \delta \ \Rightarrow\ |g(x) - g(x_0)| < \epsilon$$

である. よって

$$\{\omega \in \Omega:\ |X(\omega) - X_n(\omega)| < \delta\} \subset \{\omega \in \Omega:\ |g\big(X(\omega)\big) - g\big(X_n(\omega)\big)| < \epsilon\}$$

となる. 上式の補事象を取り, 補題 1.7(4) の後半の主張を適用すると

$$\mathsf{Pr}\big(|X - X_n| \geq \delta\big) \geq \mathsf{Pr}\big(|g(X) - g(X_n)| \geq \epsilon)\big)$$

がわかる. よって, $g(X_n) \xrightarrow{P} g(X)\,(n \to \infty)$ が示せた.
(7) の証明: $g \circ f$ が有界連続関数となるように有界連続関数 f を取る. $X_n \rightsquigarrow X\,(n \to \infty)$ と定理 4.12 から

$$\lim_{n\to\infty} \mathsf{E}[f\big(g(X_n)\big)] = \lim_{n\to\infty} \mathsf{E}[(f \circ g)(X_n)] = \mathsf{E}[(f \circ g)(X)] = \mathsf{E}[f\big(g(X)\big)]$$

から (7) は示される.
(1) の証明: $X_n \xrightarrow{P} X\,(n \to \infty)$ かつ $Y_n \xrightarrow{P} Y\,(n \to \infty)$ から $(X_n,\, Y_n) \xrightarrow{P} (X,\, Y)\,(n \to \infty)$ がわかるので, (6) の関数を $g(x,\, y) = x + y$ ととればよい.
(2) の証明: 実数 $x,\, y,\, a,\, b \in \mathbb{R}$ に対して, 不等式 $\{(x + y) - (a + b)\}^2 \leq 2\{(x - a)^2 + (y - b)^2\}$ を使えばよい.
(3) の証明: $Y_n \xrightarrow{P} c\,(n \to \infty)$ なので, 注意 4.2(2) から $Y_n \rightsquigarrow c\,(n \to \infty)$ となる. このことから $(X_n,\, Y_n) \rightsquigarrow (X,\, c)\,(n \to \infty)$ となることが証明[8] できる. あとは (7) において, $g(x,\, y) = x + y$ とすればよい.
(4) の証明: (6) を用いればよい.
(5) の証明: (7) を用いればよい. □

4.5.6　定理 4.25 の証明

$g(x)$ は $x = \mu$ の近傍で連続微分可能である. このことからある $\delta_1 > 0$ が存在して $|x - \mu| < \delta_1$ なる任意の x に対して

$$g(x) - g(\mu) = (x - \mu) \int_0^1 \dot{g}\,\big(\mu + t(x - \mu)\big)\,\mathrm{d}t \tag{4.31}$$

[8] これは特性関数の収束と分布収束が同値である事実からわかる. あとは特性関数に対して三角不等式を用いればよい. 詳しくは [48, p.135] を参照のこと.

が成立する. また $\overset{\bullet}{g}(x)$ は $x=\mu$ で連続なので, 任意の $\epsilon>0$ に対してある $\delta_2>0$ があって

$$|x-\mu|<\delta_2 \Rightarrow \left|\overset{\bullet}{g}(x)-\overset{\bullet}{g}(\mu)\right|<\epsilon$$

となる. よって $|x-\mu|<\min(\delta_1,\delta_2)=:\delta$ なる任意の x に対して

$$\begin{aligned}\left|\int_0^1 \overset{\bullet}{g}\big(\mu+t(x-\mu)\big)\,dt-\overset{\bullet}{g}(\mu)\right| &\le \int_0^1 \left|\overset{\bullet}{g}\big(\mu+t(x-\mu)\big)-\overset{\bullet}{g}(\mu)\right| dt \\ &< \epsilon\int_0^1 dt=\epsilon \end{aligned} \tag{4.32}$$

となる. (4.32) より

$$\mathsf{Pr}\left(\left|\int_0^1 \overset{\bullet}{g}\big(\mu+t(Y_n-\mu)\big)\,dt-\overset{\bullet}{g}(\mu)\right|\ge\epsilon\right)\le\mathsf{Pr}\big(|Y_n-\mu|\ge\delta\big) \tag{4.33}$$

が成立する. $\sqrt{n}\big(Y_n-\mu\big)/\sigma \rightsquigarrow \mathsf{N}(0,1)$ なので $Y_n\xrightarrow{\mathrm{P}}\mu$ である. (4.33) から

$$\int_0^1 \overset{\bullet}{g}\big(\mu+t(Y_n-\mu)\big)\,\mathrm{d}t\xrightarrow{\mathrm{P}}\overset{\bullet}{g}(\mu) \tag{4.34}$$

が成立する. $|Y_n-\mu|<\delta$ が起こったとき, (4.31) に $x=Y_n$ を代入すれば

$$\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma}=\frac{\sqrt{n}\big(Y_n-\mu\big)}{\sigma}\int_0^1 \overset{\bullet}{g}\big(\mu+t(Y_n-\mu)\big)\,\mathrm{d}t \tag{4.35}$$

を得る. さらに, $\sqrt{n}(Y_n-\mu)/\sigma \rightsquigarrow \mathsf{N}(0,1)$ と (4.34) に注意して定理 4.14(5) を (4.35) に適用すれば

$$\begin{aligned}&\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma} \\ &=\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma}\mathbb{1}\{|Y_n-\mu|<\delta\}+\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma}\mathbb{1}\{|Y_n-\mu|\ge\delta\} \\ &=\frac{\sqrt{n}(Y_n-\mu)}{\sigma}\underbrace{\int_0^1 \overset{\bullet}{g}\big(\mu+t(Y_n-\mu)\big)\,\mathrm{d}t\mathbb{1}\{|Y_n-\mu|<\delta\}}_{\overset{\bullet}{g}(\mu)+o_P(1)} \\ &\quad+\underbrace{\frac{\sqrt{n}\big(g(Y_n)-g(\mu)\big)}{\sigma}\mathbb{1}\{|Y_n-\mu|\ge\delta\}}_{=o_P(1)} \\ &\rightsquigarrow\overset{\bullet}{g}(\mu)\mathsf{N}(0,1)\end{aligned}$$

を得る. よって定理は示された. □

4.6 章末注釈と参考文献

第 4.1 節 は [17, 48] を参考にした. 注意 4.6 は [17] からの借用である. 第 4.2 節 は [40] を借用した. 第 4.4 節 は [17, 48] を借用した. 定理 4.17 の証明は [27] からの借用である. 定理 4.25 の証明は [9] からの借用である.

4.7 演習問題

演習問題 4.1. $n \in \mathbb{N}$ とする. 離散型確率変数 X_n は $\left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ 上の一様分布に従うとする. すなわち

$$\mathsf{Pr}\left(X_n = \frac{k}{n}\right) = \begin{cases} \frac{1}{n} & (k = 1, 2, \ldots, n) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. 連続型確率変数 X は $(0, 1)$ 上の一様分布に従うとする. すなわち, X は p.d.f.

$$\mathsf{p}(x) = \begin{cases} 1 & (0 < x < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

をもつ. このとき, 以下の問いに答えよ.
(1) $\frac{k}{n} \leq x < \frac{(k+1)}{n}\,(k = 1, \ldots, n)$ に対して

$$\mathsf{Pr}(X_n \leq x)$$

を求めることで, $x \in \mathbb{R}$ に対して $\mathsf{Pr}(X_n \leq x)$ を求めよ.
(2) $x \in \mathbb{R}$ に対して

$$\mathsf{Pr}(X \leq x)$$

を求めよ.
(3)

$$\lim_{n\to\infty} \left|\mathsf{Pr}(X_n \leq x) - \mathsf{Pr}(X \leq x)\right|$$

を評価することで $X_n \rightsquigarrow X\,(n \to \infty)$ を示せ.

演習問題 4.2. $\{X_n\}_{n=1}^{\infty}$ を確率変数列とし, $\mathsf{E}[X_n^2] < \infty\,(n \in \mathbb{N})$ とする.
(1) $\mu \in \mathbb{R}$ を定数とする.

$$\mathsf{E}[(X_n - \mu)^2] = \mathsf{Var}[X_n] + \left\{\mathsf{E}[X_n] - \mu\right\}^2$$

が成り立つことを示せ.
(2) 次の条件は同値であることを示せ.

(a) $\lim_{n\to\infty} \mathsf{E}[(X_n - \mu)^2] = 0.$

(b) $\lim_{n\to\infty} \mathsf{E}[X_n] = \mu$ かつ $\lim_{n\to\infty} \mathsf{Var}[X_n] = 0$

演習問題 4.3. $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$, X, Y を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ で定義された確率変数列とする. 確率ベクトル (X_n, Y_n) が確率ベクトル (X, Y) に確率収束するとは, 任意の $\epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}\Big(\sqrt{(X_n - X)^2 + (Y_n - Y)^2} \geq \epsilon\Big) = 0$$

が成り立つこととする. これを $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)\,(n \to \infty)$ と記すことにする. このとき, 以下の問いを答えよ.
(1) $X_n \xrightarrow{\mathrm{P}} X\,(n \to \infty)$ かつ $Y_n \xrightarrow{\mathrm{P}} Y\,(n \to \infty)$ ならば, $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)\,(n \to \infty)$ が成り立つことを示せ.
(2) $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)\,(n \to \infty)$ ならば, $X_n \xrightarrow{\mathrm{P}} X\,(n \to \infty)$ かつ $Y_n \xrightarrow{\mathrm{P}} Y\,(n \to \infty)$ が成り立つことを示せ.

演習問題 4.4. $X_1, X_2, \ldots$ は互いに独立な確率変数列で

$$\mathsf{Pr}(X_n = 1) = \frac{1}{n}, \quad \mathsf{Pr}(X_n = 0) = 1 - \frac{1}{n} \qquad (n = 1, 2, \ldots)$$

とする. このとき, $X_n \xrightarrow{\text{q.m.}} 0\,(n \to \infty)$ だが, $X_n \not\xrightarrow{\text{a.s.}} 0\,(n \to \infty)$ であることを示せ.

ヒント Borel-Cantelli の第 2 補題を用いるとよい.

演習問題 4.5. $\Omega = (0, 1]$, $\mathcal{A} = \mathcal{B}((0, 1])$ とし

$$\mathsf{Pr}((a, b]) = b - a \qquad (a, b \in \Omega)$$

とする.

$$X_n(\omega) = \sqrt{n}\mathbb{1}_{(0,\,1/n]}(\omega) \qquad (n = 1, 2, \ldots)$$

と定義したとき, $X_n(\omega) \xrightarrow{n\to\infty} 0$ だが, $\lim_{n\to\infty} \mathsf{E}[X_n^2] \neq 0$ を示せ.

第II部

宴編: 統計的推測

第5章　統計的推測論の枠組み

この章では統計的推測の枠組みについて述べる. 節 5.1 では, 統計的推測理論で重要な役割を果たす統計的モデルを説明する. 節 5.2 では, 統計的推測理論を統一的な枠組みで議論する統計的決定理論を簡単に説明する.

5.1　統計的実験と母数モデル

$\mathbb{X}$ を距離空間とし, $\mathcal{B}(\mathbb{X})$ を $\mathbb{X}$ 上の Borel 集合族[1] とする. すなわち, $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ は可測空間である.

いま, $X_1, X_2, \ldots, X_n$ は確率空間 $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上の $\mathbb{X}$ 値確率要素[2] とする. さらに, $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\top$ と書くことにする. 典型的な例は $\mathbb{X} = \mathbb{R}^d\,(d \in \mathbb{N})$ 等である.

この講義では $X_1, X_2, \ldots, X_n$ は独立同一の分布に従うものとする. このような場合, $X_1, X_2, \ldots, X_n$ を **ランダム標本** と呼ぶ. また, 可測空間 $\big(\mathbb{X}^n, \big(\mathcal{B}(\mathbb{X})\big)^{\otimes n}\big)$ を **標本空間** と呼ぶ.

可測空間 $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の確率測度 P を

$$\mathsf{P}(B) := \mathsf{Pr}(X_1 \in B) \qquad (B \in \mathcal{B}(\mathbb{X}))$$

で定義する. 確率変数 X_1 によって誘導された $\mathbb{X}$ 上の確率 P を **X_1 の確率分布**(簡単に分布ともいうことがある) といい, $X_1 \sim \mathsf{P}$ と表記することにする. さらに, $\boldsymbol{X}$ の分布を $\mathsf{P}^{\otimes n}$ または $\mathsf{P}^{\mathbf{X}}$ と書くことにする. したがって, $B_j \in \mathcal{B}(\mathbb{X})\,(j = 1, 2, \ldots, n)$ に対して

$$\mathsf{P}^{\otimes n}(B_1 \times B_2 \times \cdots \times B_n) = \mathsf{Pr}\big(\mathbf{X} \in B_1 \times B_2 \times \cdots \times B_n\big)$$

となる[3].

注意 5.1. $B \in \mathcal{B}(\mathbb{X})$ に対して, $\mathsf{Pr}(X_1 \in B)$ と書いたとき

$$\mathsf{Pr}(X_1 \in B) = \mathsf{Pr}\big(\{\omega \in \Omega;\, X_1(\omega) \in B\}\big)$$

[1] 開集合族を含む最小の σ 加法族

[2] 確率変数, 確率ベクトル, 確率行列の総称を確率要素という.

[3] この記法 $\mathsf{P}^{\otimes n}$ にも測度論的な背景がある. すなわち, $\mathsf{P}^{\otimes n}$ は P の n 個の測度の直積測度になっている.

の意味である. □

確率論的アプローチでは, 観測データを生成するメカニズムを表現する確率分布 P は既知であり, 確率要素 $\boldsymbol{X}$ の分布論的な特徴を調べる. 一方, 統計的推測のアプローチでは観測データを生成するメカニズムを表現する確率測度 P は観測者には未知であり, 観測データ $\boldsymbol{X}$ に基づいて未知の確率測度 P を回復することを目指す. 統計的推測は確率論的アプローチの逆問題と言える.

統計的推測の考え方は統計的実験という概念を基礎に組み立てられる. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}^*$ とする. ここで, P^* はデータを生成する**真の確率分布**とした. すなわち, $\boldsymbol{X} := (X_1, \ldots, X_n) \sim \left(\mathsf{P}^*\right)^{\otimes n}$ である. 統計的推測では観測データを生成するメカニズムを表現する未知の確率分布 P^* を観測データから回復することを目的とする. $\left(\mathbb{X}, \mathcal{B}(\mathbb{X})\right)$ 上の確率分布全体の集合から真の分布を探すのは多くの困難を伴うので, 候補となる確率分布の適当な集まりを設定する. この確率分布族の特定の集まりを $\mathcal{P}$ と書いたとき, この $\mathcal{P}$ を**統計的モデル**という. そして, 真の確率分布 P^* は設定した統計的モデル $\mathcal{P}$ に含まれるとこの講義では仮定[4] する.

さらに統計的モデルの各要素はある集合 Θ の要素 θ で添え字付けられると仮定する. すなわち, 添え字集合全体と統計的モデルとの対応である写像

$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$

を想定[5] する. これを統計的モデルの**母数化**といい, Θ を**母数空間**, その要素 θ を**母数**という. そして真の確率分布 P^* に対応するある母数 $\theta^* \in \Theta$ が存在[6] すると仮定する. この θ^* を**真の母数**ということにする. すなわち

$$\mathsf{P}^* = \mathsf{P}_{\theta^*}$$

である. もちろん $\theta^* \in \Theta$ ではあるが, 真の母数 θ^* は観測者には未知である. したがって, 統計的推測では未知の真の母数 θ^* を観測データ $\boldsymbol{X}$ から回復することが目的[7] となる.

次に統計的モデルのなかでも最も基本的な正則母数モデルを定義する.

[4] もちろん, 真の確率分布 P^* が想定した統計的モデルに含まれないことを想定した議論もある. しかし, 統計的モデルに真の分布が含まれない場合には, より高度な議論が必要となるので, この講義録ではより基本的な設定を考えている.

[5] この写像がどのような性質を持つかを数学的に定義することはできる. しかし, 確率分布の集まりの空間に位相をいれることになり, 数学的に議論が高度になるので, 漠然とした写像と理解しておくことにしよう.

[6] 母数化が識別可能 (この概念は後に定義) であれば, この母数は一意である..

[7] 統計的機械学習では, 真の母数 θ^* の回復を目標とするより, 真の確率測度と同等のメカニズムから生成される未来の観測 $\widetilde{X}_{n+1}$ を観測データ $\boldsymbol{X}$ に基づいて予測することを目標にしている. 統計的機械学習では, 統計的モデルの役割が相対的に低くなる.

定義 5.2. 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta\}$ が **正則母数モデル**(regular parametric model) であるとは, 次の条件をみたすときをいう.

(1) 母数空間 Θ は 有限次元 Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である. ただし $d \in \mathbb{N}$ である.

(2) 写像

$$\Theta \ni \theta \mapsto \mathsf{P}_\theta \in \mathcal{P}$$

は「滑らか」である [8].

条件 (2) を母数化の正則性という. 正則母数モデルを簡単に母数モデルということもある.

さらにこの講義では次の条件も仮定する.

(3) $\forall \theta_1, \theta_2 \in \Theta$ に対して

$$\theta_1 \neq \theta_2 \Rightarrow \mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}$$

をみたす.

すなわち母数化を与える写像は単射である. このような母数化を **識別可能** (identifiable) であるという.

注意 5.3. 定義 5.2 は数学的には厳密ではない表現である. 数学的により厳密な母数モデルの定義については Bickel *et al.* (1993, pp.11-13) を参照のこと.

注意 5.4. 母数空間 Θ が有限次元ではないような統計的モデルを考えることも重要である. Θ が無限次元の統計的モデルのことを「ノンパラメトリック・モデル」と統計学では慣例的に呼んでいることが多い. 母数空間が無限次元とはいえ, 統計的モデルは母数化されているので, 「非母数モデル」と呼ぶのは奇異である. しかし歴史的にこの用語が使用されてきたので, この講義録では統計学の歴史的慣例に従うことにする. この言葉使いを嫌う数理統計学者は「ノンパラメトリック・モデル」のことを「無限次元統計的モデル」と呼んでいる. さすがに「無限次元母数モデル」とは言わないようである. さらに母数空間が有限次元の母数空間と無限次元の母数空間の直積で表現され, 有限次元の母数を回復の対象とするような統計的モデルを「セミパラメトリック・モデル [9]」という. これも言葉の意味としては奇異であるが, 統計学の習慣に従うことにする.

[8] $\mathcal{P}$ は測度の集合なので位相をどのようにいれるかはすこし難しい議論になる. $\mathcal{P}$ が Radon 測度の集まりならば, weak-star 位相を入れることができる. この議論は Tojo and Yoshino (2021) を参照のこと.

[9] 英語読みをすれば, 「セマイパラメトリックモデル」という. 「semi-parametric model」の最初の「i」は長母音であることに注意が必要である.

生存データ解析で広く使用される Cox の比例ハザード・モデルはセミパラメトリック・モデルの最高傑作であろう. 20 世紀の数理統計学の到達点のひとつであるセミパラメトリック・モデルに対する統計的推測理論については Bickel *et al.* (1993), van der Vaart and Wellner (1996), van der Vaart (1998) , Kosorok (2007), 久保木・鈴木 (2015) を参照のこと. □

定義 5.5. (1) 可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ と母数モデル

$$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}(\mathbb{X})) \text{ 上の確率分布 }\}$$

の組を 統計的実験(statistical experiment) といい

$$\left(\mathbb{X}^n,\, \mathcal{B}(\mathbb{X})^{\otimes n},\, \{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}\right)$$

と書く.
(2) 観測データを生成するメカニズムを表現する真の確率測度 P^* に対応する母数を $\theta^* \in \Theta$ と書くことにする. すなわち, θ^* は真の母数で

$$\mathsf{P}^* = \mathsf{P}_{\theta^*}$$

である.

この講義では $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ と書いたとき, $X_1, X_2, \ldots, X_n$ は独立に同一分布 P^* に従うことを仮定する. 上記の仮定をおいた観測データ $\boldsymbol{X}$ のことを 標本の大きさが n のランダム標本 という.

以上の議論から, この講義で扱う統計的実験をまとめると下記のようになる.

この講義で仮定する統計的実験

(1) 観測データを

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$$

と書き, 可測空間 $\left(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\right)$ に値をとる.
(2) 観測データは真の分布 P^* からの標本である. すなわち

$$\boldsymbol{X} \sim \left(\mathsf{P}^*\right)^{\otimes n}$$

である. ただし P^* は $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の確率測度である.
(3) 統計的モデル $\mathcal{P} = \left\{\mathsf{P}_\theta;\, \theta \in \Theta\right\}$ を設定する. ただし P_θ も P^* と同じ可測空間 $\left(\mathbb{X}, \mathcal{B}(\mathbb{X})\right)$ 上の確率測度である.
(4) 母数空間 Θ は Euclid 空間 $\mathbb{R}^d$ の「よい」部分集合である. ただし $d \in \mathbb{N}$ である.

(5) $\Theta \ni \theta$ から $\mathsf{P}_\theta \in \mathcal{P}$ への写像は「滑らか」かつ単射 (母数化の識別可能性を仮定).
(6) ある $\theta^* \in \Theta$ が唯一あって $\mathsf{P}^* = \mathsf{P}_{\theta^*}$ と書ける.

これらをまとめて統計的実験といい

$$\left(\mathbb{X}^n,\, \mathcal{B}(\mathbb{X})^{\otimes n},\, \{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}\right)$$

と書き, $\mathcal{P} = \left\{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\right\}$ を正則母数モデルという. ランダム標本の場合には, 簡単に $\mathcal{P} = \left\{\mathsf{P}_\theta;\, \theta \in \Theta\right\}$ とも書くことにする. 以後は単に母数モデルということにする. そして統計的推測の目標は真の確率分布 P^* からのランダム標本 $(X_1, X_2, \ldots, X_n)$ に基づき θ^* を回復することである.

例 5.6. (1) $X_1, X_2, \ldots, X_n$ を正規分布 $\mathsf{N}(\mu, \sigma^2)$ から標本の大きさが n のランダム標本とする. ただし $\mu, \sigma\,(0 < \sigma < \infty)$ が共に未知とする. このとき統計的実験

$$\left(\mathbb{R}^n,\, \mathcal{B}(\mathbb{R}^n),\, \left\{\mathsf{p}_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)(-\infty < x < \infty);\right.\right.$$
$$\left.\left.\theta := (\mu, \sigma) \in \Theta := \mathbb{R} \times (0, \infty)\right\}\right)$$

を想定していることになる. 統計的モデルは確率分布が特定できる表現でよいので, この場合には p.d.f. で表現している.
(2) $X_1, X_2, \ldots, X_n$ を Bernoulli 分布 $\mathsf{Ber}(\theta)\,(0 < \theta < 1)$ から標本の大きさが n のランダム標本とする. ただし θ が未知のときには, 統計的実験

$$\left(\{0, 1\}^n,\, 2^{\{0,1\}^n},\, \left\{\mathsf{p}_\theta(\boldsymbol{x}) = \theta^x(1-\theta)^{1-x}\,(x = 0, 1);\, \theta \in \Theta = (0, 1)\right\}\right)$$

を想定していることになる. □

注意 5.7. $X_1, X_2, \ldots, X_n$ を分布 P^* から標本の大きさが n のランダム標本といったときには,

$$\left(\mathbb{R}^n,\, \mathcal{B}(\mathbb{R}^n),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\}\right)$$

のような統計的実験を仮定し, ある $\theta^* \in \Theta$ があって, $\mathsf{P}_{\theta^*} = \mathsf{P}^*$ であることを想定している. ことさらに統計的実験という用語を今後は用いないことにする. □

5.2 統計的決定問題

統計的推測理論には多くのアプローチがある. その中で代表的なアプローチが二つある. 一つは頻度論的なもので, もう一方はベイズ論的なものである. 以下では頻度論的推測理論の枠組みを説明することにする. Bayes 論的推測理論は第 9 章で説明する. 以下では, 頻度論的推測理論の枠組みを統計的決定理論[10] の言葉を使って説明する.

標本空間を $\mathbb{X}^n$ とし, 観測データを $\boldsymbol{X}$ とする.

(1) まず観測データに基づき行う行動のすべてを集めた集合を 行動空間 といい $\mathbb{A}$ で記す. この講義では $\mathbb{A} = \mathbb{R}$ や $\mathbb{A} = \{0, 1\}$ などである. 観測者が観測データに基づき行動 $\mathbb{A}$ の要素を選択するルールを決定関数といい

$$d : \mathbb{X}^n \ni \boldsymbol{x} \mapsto d(\boldsymbol{x})$$

で記す. 決定関数の集まりを 決定空間 といい, $\mathbb{D}$ と記す. したがって観測者は合理的な行動 d が存在すればありがたいわけである.

(2) 次に行動を評価するための道具として直積空間 $\Theta \times \mathbb{A}$ 上の非負値実数値関数

$$\mathsf{L} : \Theta \times \mathbb{A} \mapsto [0, \infty) \cup \{\infty\}$$

を用意する[11]. この関数を 損失関数 といい, $\mathsf{L}(\theta, a)$ の値が小さいほど望ましい行動である. 決定関数の「よさ」を評価するには観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ と真の母数 θ^* における損失関数の値 $\mathsf{L}\big(\theta^*, d(\boldsymbol{x})\big)$ がわかればよい. したがって決定関数 d の「よさ」の評価に $\mathsf{L}\big(\theta^*, d(\boldsymbol{X})\big)$ を使えばよいのだが, これは用いることができない. これはランダムな量であり, 未知の母数 $\theta^* \in \Theta$ がわからないので, $\mathsf{L}\big(\theta^*, d(\boldsymbol{x})\big)$ は知ることができない量である. そこで母数 θ に対して観測データが $\mathsf{P}_\theta^{\otimes n}$ によって生成されたと仮定し, 損失関数 $\mathsf{L}\big(\theta, d(\boldsymbol{X})\big)$ を $\mathsf{P}_\theta^{\otimes n}$ に関して期待値を取ったもの

$$\mathsf{R}(\theta, d) := \mathsf{E}_\theta\big[\mathsf{L}\big(\theta, d(\boldsymbol{X})\big)\big]$$

を考える. ただし, $\mathsf{E}_\theta[\,\cdot\,]$ は $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta)$ のもとでの期待値である. これを決定関数 d の 危険関数 という. なお, 危険関数は $\theta \in \Theta$ の関数であることに注意せよ.

(3) 決定関数 d の「よさ」の評価は真の母数 θ^* のもとで行いたいところである. しかしこれは未知である. 危険関数の $\theta \in \Theta$ に関する

[10] 統計的決定理論はゲーム論の概念を借用して, 統計的推測論の枠組みと最適理論を定式化 (言語化) したものである.

[11] ただし, 区間推定のばあいには, $\mathsf{L} : \mathbb{A} \times \Theta \mapsto [-1, \infty) \cup \{\infty\}$ とすることもある.

なんらかの一様な評価が必要になってくる. このことから危険関数の母数空間に関する一様な評価が統計的推測論の深みと困難の淵源である. またこれが統計的推測理論のわかりにくさの原因でもあろう. 前節で説明した統計的実験に行動空間, 決定空間, そして損失関数を加えた組

$$\big(\mathbb{X}^n,\, \mathcal{B}\big(\mathbb{X}^n\big),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\},\, \mathbb{A},\, \mathbb{D},\, \mathsf{L}\big)$$

を統計的決定問題[12] という.

(4) 決定空間 $\mathbb{D}$ は $\mathbb{X}^n$ から $\mathbb{A}$ への可測関数全体とすることもできる. しかし目標は危険関数の Θ に関する一様な評価であるので, $\mathbb{D}$ には可測性以外の合理的な制限[13] を設けるのが一般的である.

点推定問題

真の母数 θ^* を観測データ $\boldsymbol{X}$ に基づいて 1 点で回復するのが点推定である. したがって $\mathbb{A} = \Theta$ となる. 点推定の場合には決定関数 $d(\boldsymbol{X})$ を**推定量**といい, 観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ における推定量の値 $d(\boldsymbol{x})$ を**推定値**という. $\Theta = \mathbb{R}$ ならば損失関数として

$$\mathsf{L}\big(a,\, \theta\big) = \big(a - \theta\big)^2, \quad \mathsf{L}\big(a,\, \theta\big) = \big|a - \theta\big|$$

を取るのが代表的なアプローチである. 上記の損失関数それぞれに対応する危険関数

$$\mathsf{R}(\theta,\, d) = \mathsf{E}_\theta\big[\mathsf{L}\big(\theta,\, d(\boldsymbol{X})\big)\big]$$

を**平均 2 乗誤差**と**平均絶対誤差**という. したがって平均 2 乗誤差を Θ に関してなんらかの意味で一様に評価することで考えている推定量の族 $\mathbb{D}$ の中から「最適」な推定量ないしは合理的な観点から正当化される推定量をみつけたいわけである.

検定問題

母数空間を 2 つの排反な部分集合に分ける. すなわち

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \varnothing$$

である. 行動は真の母数 θ^* が Θ_0 に属するか, Θ_1 に属するかを判断する. したがって行動空間は $\mathbb{A} = \{0,\, 1\}$ と書ける. すなわち, $\mathbb{A}$ の元 0 は

[12] 本来であれば, どこで可測であるかを考える必要があるので,

$$\Big(\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\big),\, \{\mathsf{P}_\theta;\, \theta \in \Theta\},\, \big(\mathbb{A},\, \mathcal{B}(\mathbb{A})\big),\, \big(\mathbb{D},\, \mathcal{B}(\mathbb{D})\big),\, \mathsf{L}\Big)$$

と書くべきである.

[13] 合理的な制限の概念として不変性や不偏性などがある. また尤度に基づく方法に限定するといった考え方もある.

$\theta^* \in \Theta_0$ であると判断したことを表し, 1 は $\theta^* \in \Theta_1$ であると判断したことを表す. 決定関数 $d : \mathbb{X}^n \longrightarrow \mathbb{A}$ は標本空間 $\mathbb{X}^n$ の部分集合 W に対して

$$d(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \notin W) \end{cases}$$

で定めること[14]ができる. 検定問題では d のことを検定関数という. 損失関数としては

$$\mathsf{L}\big(\theta,\, 0\big) = \begin{cases} 0 & (\theta \in \Theta_0) \\ 1 & (\theta \in \Theta_1) \end{cases} ; \qquad \mathsf{L}\big(\theta,\, 1\big) = \begin{cases} 1 & (\theta \in \Theta_0) \\ 0 & (\theta \in \Theta_1) \end{cases}$$

と取る. すなわち

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	0	1
$d = 1$	1	0

である.

通常 $H_0 : \theta \in \Theta_0$ のことを帰無仮説とよぶ. さらに, $d(\boldsymbol{x})$ に形式的に $\boldsymbol{X}$ を代入した $d(\boldsymbol{X})$ を観測データに基づく検定の決定を表す. 簡単に, これを検定統計量という. $H_1 : \theta \in \Theta_1$ のことを対立仮説という. 危険関数 $\mathsf{R}\big(\theta,\, d\big) = \mathsf{E}_\theta\big[\mathsf{L}\big(\theta,\, d(\boldsymbol{X})\big)\big]$ は以下のようになる. ここで第 1 種の誤

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$d = 0$	正しい判断	第 2 種の誤り
$d = 1$	第 1 種の誤り	正しい判断

りの確率と第 2 種の誤りの確率はトレード・オフの関係になっていることが鍵である. すなわち, 同時には二つの確率を小さくできない. 実は

$$(\text{第 1 種の誤りの確率}) + (\text{第 2 種の誤りの確率}) \geq \text{下限}$$

ということになっているのである[15]. そこで $\theta \in \Theta_0$ のとき

$$\beta(\theta) := \mathsf{R}\big(\theta,\, d\big)$$

とし, $\theta \in \Theta_1$ のとき

$$\beta(\theta) := 1 - \mathsf{R}\big(\theta,\, d\big)$$

[14]正確には確率化決定関数を考える必要があるが, 議論を簡単にするためにこれは考えないことにする.

[15]このことは第 7 章で説明する Neyman-Pearson の補題からわかる.

と定義したものを検出力関数という. 仮説検定では与えられた数 $\alpha\,(0 < \alpha < 1)$ に対して

$$\sup_{\theta\in\theta_0} \mathsf{R}\big(\theta,\, d\big) \leq \alpha$$

をみたす検定関数の中から $\theta \in \Theta_1$ において $\beta(\theta)$ を大きくするもの, すなわち $\mathsf{R}\big(\theta,\, d\big)$ を小さくするものと選ぶことを目指す. ちなみに α のことを有意水準という. $\sup_{\theta\in\Theta_0} \mathsf{R}\big(\theta,\, d\big)$ を検定関数 d のサイズという. したがってサイズが有意水準より小さい検定関数の中から $\theta \in \Theta_1$ において検出力関数の値が一様に大きなもの [16] を探したいわけである.

区間推定

議論を簡単にするために $\Theta = \mathbb{R}$ とする. 区間推定において行動は $\mathbb{R}$ の区間となる. したがって決定空間は観測データから区間への対応となる. 観測データの実現値 $\boldsymbol{X} = \boldsymbol{x}$ に基づく母数 θ の推定区間 $[\ell(\boldsymbol{x}),\, u(\boldsymbol{x})]$ に対して損失関数として

$$\mathsf{L}\big(\theta,\, [\ell,\, u]\big) = (u - \ell) - \mathbb{1}\big\{\theta \in [\ell,\, u]\big\}$$

などが考えられる. ただし, $\mathbb{1}\big\{\theta \in [\ell,\, u]\big\} = \begin{cases} 1 & (\theta \in [\ell,\, u]) \\ 0 & (\theta \notin [\ell,\, u]) \end{cases}$ である. この場合には, L は負の値を取ることもある. 決定関数

$$d(\boldsymbol{X}) = [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]$$

に対して危険関数は

$$\mathsf{R}\big(\theta,\, d\big) = \mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big] - \mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big)$$

となる. ただし, $\forall \mathcal{B}(\mathbb{X}^n)$ と $\boldsymbol{X} \sim \mathsf{P}^{\otimes n}$ に対して, $\mathsf{Pr}_\theta(\boldsymbol{X} \in B) := \mathsf{E}_\theta\big[\mathbb{1}_B(\boldsymbol{X})\big]$ で定義した.

実数 $\alpha\,(0 < \alpha < 1)$ が与えられたとき

$$\mathsf{Pr}_\theta\big(\theta \in [\ell(\boldsymbol{X}),\, u(\boldsymbol{X})]\big) \geq 1 - \alpha$$

のもとで区間の長さの期待値 $\mathsf{E}_\theta\big[u(\boldsymbol{X}) - \ell(\boldsymbol{X})\big]$ を短くする区間が望ましい区間といえよう. $1 - \alpha$ を信頼係数とよぶ.

以上のように統計的決定問題の枠組みで統計的推測の問題である点推定, 区間推定, および検定を統一的に扱うことができる.

次に決定空間の元の間に順序 $\prec$ を導入しよう. 決定関数 $\forall d_1,\, d_2 \in \mathbb{D}$ に対して

$$d_1 \prec d_2$$

[16] これは究極の目標であり, 一様に検出力関数の値を最大にする検定統計量は存在しないかもしれない.

$$\Leftrightarrow \mathsf{R}(\theta,\, d_1) \leq \mathsf{R}(\theta,\, d_2)\,(\forall\theta \in \Theta) \text{ かつ } \mathsf{R}(\theta_0,\, d_1) < \mathsf{R}(\theta_0,\, d_2)\,(\exists\theta_0 \in \Theta)$$

定める. すると決定空間 $\mathbb{D}$ を標本空間 $\mathbb{X}$ から行動空間 $\mathbb{A}$ への可測関数すべてから成る集合とすれば順序 $\prec$ は半順序になる. すなわち順序 $\prec$ の意味で一番よいものは存在しない.

たとえば $X \sim \mathsf{N}(\mu,\, 1)$ によって μ を推定する問題を損失関数 $\mathsf{L}\big(\mu,\, \widehat{\mu}\big) = \big(\widehat{\mu} - \mu\big)^2$ のもとで考える. ただし $\widehat{\mu}$ は μ の推定量である. このとき

$$\widehat{\mu}_0(X) = 0$$

なる推定量より, 母数 ν に関して一様によい推定量は存在しない. なぜならば $\mu = 0$ において $\widehat{\mu}_0$ の危険関数の値は 0 となるので, $\widehat{\mu}_0$ より一様によい推定量は存在しないこと [17] がわかる.

最良の決定関数が存在しない場合には決定関数を比較するための別の観点の導入が必要となる. 主なもので次の二つがある.

(1) 決定関数の最適性について別の概念を導入する. 代表的なものとしてミニマックス基準と Bayes 基準がある.

(2) 考察する決定関数を制限し, その中で危険関数を母数 Θ に関して一様に小さくする決定関数をみつける. たとえば不偏性, 不変性などを導入して, 考察する決定関数を制限する方法がある. また Neyman-Pearson の補題による議論がある.

さらに決定空間のなかからよい決定関数をみつけるのではなく, 一定の原理によって導かれる決定関数を考えて, それについてなんらかの合理性を証明する方針がある. 統計的決定問題の枠組みからははずれるが, ある原理に基づきなんらかのかたちで合理的な正当化ができる決定関数を導出することが考えられてきた. 導出の原理として推定ではモーメント法, 最尤法 (第 7 章) などが知られている. 検定法では尤度比検定, スコア検定, Wald 検定, Rao 検定 (第 8 章) がある. 区間推定では検定統計量の反転, ピボット法 (第 8 章) などがある.

5.3　章末注釈と参考文献

この章の話題を数学的に厳密に議論をするためには, 数学の進んだ道具立てが必要となるので, 直観的な説明に留めた.

[17] ある推定量に対して、それより一様によい推定量が存在しないとき、その推定量を許容的という.

5.4 演習問題

演習問題 5.1. $0 < \theta < 1$ とし, $X \sim \mathsf{Bino}(100, \theta)$ とする. 母数 θ に対する損失関数 $\mathsf{L}(\theta, d) = (\theta - d)^2$ を考える. さらに, θ の推定量

$$d_0(X) = \frac{X}{100}, \quad d_1(X) = \frac{X+3}{100}, \quad d_2(X) = \frac{X+3}{106}$$

を考える.
(1) 推定量 d_0 の危険関数 $\mathsf{R}(\theta, d_0)$ を求めよ.
(2) 推定量 d_1 の危険関数 $\mathsf{R}(\theta, d_1)$ を求めよ.
(3) 推定量 d_2 の危険関数 $\mathsf{R}(\theta, d_2)$ を求めよ.
(4) 危険関数 $\mathsf{R}(\theta, d_0)$ と $\mathsf{R}(\theta, d_1)$ の大小を比較せよ.
(5) 危険関数 $\mathsf{R}(1/2, d_0)$ と $\mathsf{R}(1/2, d_2)$ および $\mathsf{R}(1/8, d_0)$ と $\mathsf{R}(1/8, d_2)$ の大小を比較せよ.

第6章　母数モデル

節 6.1 では, 統計的モデルのなかでも最も基本的な正則母数モデルとそのもとで Fisher 情報量を導入する. 節 6.2 と節 6.3 では, 母数モデルのなかでも非常によい性質と数学的な美しさをもつ指数型分布族とその基本的な性質を説明する. 節 6.4 では, 十分統計量を定義し, 十分性の利便性の高い判定条件を述べる. 節 6.5 では, 最小十分統計量と完備十分統計量を定義し, それらに関わる重要な定理を述べる. いくつかの主張の証明で Lebesgue 積分の積分の収束定理を用いた. この点について慣れていない読者は証明を読み飛ばしても全体の流れを理解するのには困難が伴わないようにした.

6.1　正則母数モデルと Fisher 情報量

定義 6.1. $d \in \mathbb{N}$ とする. $\mathbb{X}(\subset \mathbb{R})$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ が与えられているとする. 分布 $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f.(または p.m.f.) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ と表記する. このとき母数モデル $\mathcal{P}$ が 正則 であるとは次の条件をみたすときをいう.

(1) $\Theta \subset \mathbb{R}^d$ は開集合.

(2) $\mathcal{P}$ に属する分布の p.d.f.(または p.m.f.) は同じ台をもつ. すなわち, 集合 $\{x \in \mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}}(x) > 0\}$ は $\boldsymbol{\theta} \in \Theta$ に依存しない.

(3) 任意の $\boldsymbol{\theta} \in \Theta$ とする. $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の $\boldsymbol{\theta}$ の 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ に関して連続である.

(4) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の $\boldsymbol{\theta}$ に関する 1 次と 2 次の偏導関数は $x(\in \mathbb{X})$ の関数として可積分である.

(5) $\mathsf{p}_{\boldsymbol{\theta}}(x)$ の 1 次と 2 次の偏導関数は $\boldsymbol{\theta}$ の微分記号と x の積分記号と交換が可能である.

注意 6.2. (1) 定義 6.1(5) において, $\boldsymbol{\theta}$ の微分記号と x の積分記号が交換できることは無条件で数学的には成立することでないことを注意しておく.

(2) さらに, この仮定は以下で定義する Fisher 情報量 (下記で定義する) の基本性質を担保するために使われることにも注意せよ.
(3) 6.2 で導入する指数型分布族は定義 6.1 の条件をみたすことがわかる. 補題 6.11 を参照のこと. □

$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき Θ 上の実数値関数 ℓ_n を

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j) \qquad (\boldsymbol{x} = (x_1, x_2, \ldots, x_n))$$

で定義[1] する.

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のときこの分布に関する期待値, 分散および共分散を

$$\mathsf{E}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Var}_{\boldsymbol{\theta}}[\,\cdot\,], \qquad \mathsf{Cov}_{\boldsymbol{\theta}}[\,\cdot,\,\cdot\,]$$

と表記する. すなわち, 可積分な関数 $h : \mathbb{X}^n \to \mathbb{R}$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}}[h(\boldsymbol{X})] := \begin{cases} \sum_{\boldsymbol{x}\in\mathbb{X}^n} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j) & \text{(離散型の場合)} \\ \int_{\mathbb{X}^n} h(\boldsymbol{x}) \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j)\,\mathrm{d}\boldsymbol{x} & \text{(連続型の場合)} \end{cases}$$

と定める. ただし $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$ である. ここで, 各 $X_j\,(j = 1, \ldots, n)$ が離散型のとき, $\mathbb{X}$ は高々可算個の集合であることに注意する. さらに, $\boldsymbol{X}$ を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率ベクトルとしたとき, 任意の $B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 可測空間 $(\Omega, \mathcal{A})$ 上の確率測度 $\mathsf{Pr}_{\boldsymbol{\theta}}$ を

$$\mathsf{Pr}_{\boldsymbol{\theta}}\big(\boldsymbol{X}^{-1}(B)\big) := \mathsf{Pr}_{\boldsymbol{\theta}}(\boldsymbol{X} \in B) := \mathsf{E}_{\boldsymbol{\theta}}\big[\mathbb{1}_B(\boldsymbol{X})\big]$$

で定める. ただし, $\boldsymbol{X}^{-1}(B) := \{\omega \in \Omega;\, \boldsymbol{X}(\omega) \in B\} \in \mathcal{A}$ である.

さらに, $i, j = 1, 2, \ldots, d$ に対して

$$\overset{\bullet}{\ell}_{n,i} := \overset{\bullet}{\ell}_{n,i}\,(\boldsymbol{\theta}|\,\boldsymbol{x}) = \frac{\partial \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial \theta_i}, \quad \overset{\bullet\bullet}{\ell}_{n,ij} := \overset{\bullet\bullet}{\ell}_{n,ij}\,(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial \theta_i \partial \theta_j},$$

$$\overset{\bullet}{\boldsymbol{\ell}}_n = (\overset{\bullet}{\ell}_{n,1}, \overset{\bullet}{\ell}_{n,2}, \ldots, \overset{\bullet}{\ell}_{n,d})^\top, \quad \overset{\bullet\bullet}{\boldsymbol{\ell}}_n = \big(\overset{\bullet\bullet}{\ell}_{n,ij}\big), \quad \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$$

と定義する. $\boldsymbol{X} \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ のとき Fisher **情報量** $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ は $d \times d$ 行列で $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})$ の (i, j) 成分 $\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij}$ は

$$\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij} = \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\ell}_{n,i}\,(\boldsymbol{\theta} \mid \boldsymbol{X})\ \overset{\bullet}{\ell}_{n,j}\,(\boldsymbol{\theta} \mid \boldsymbol{X})\big] \qquad (1 \le i, j \le d)$$

で定義される.

[1] 対数 尤度 (ゆうど) 関数 と呼ぶ.

定理 6.3. 標本空間 $\mathbb{X}$ 上の母数モデル $\mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ は正則であるとする. $\boldsymbol{X} = (X_1, \ldots, X_n) \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ としたとき以下が成立する.

(1) $\forall \boldsymbol{\theta} \in \Theta$ に対して $\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = \mathbf{0}_d$ となる.

(2) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = \mathsf{Cov}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta} \mid \boldsymbol{X})\big] = \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\ \overset{\bullet}{\boldsymbol{\ell}}_n^{\top}\,(\boldsymbol{\theta}|\,\boldsymbol{X})\big]$ が成り立つ.

(3) $\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta}) = -\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\boldsymbol{\ell}}_n\,(\boldsymbol{\theta}|\,\boldsymbol{X})\big]$ が成り立つ.

Proof. 証明は連続型分布の場合を示す. 離散型の場合は積分を和の記号に置き替えればよい. $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f. を $\mathsf{p}_{\boldsymbol{\theta}}$ と表記する. さらに, $\widetilde{\mathbb{X}} = \{x \in \mathbb{X};\, \mathsf{p}_{\boldsymbol{\theta}}(x) > 0\}$ とおく. 母数モデルの正則性から微分記号と積分記号の交換が保証されているので, $i = 1, 2, \ldots, d$ に対して

$$
\begin{aligned}
\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\ell}_{n,i}\,(\boldsymbol{\theta}|\,\boldsymbol{X})\big] &= \sum_{k=1}^{n} \int_{\mathbb{X}} \left(\frac{\partial}{\partial \theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k) \right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k \\
&= \sum_{k=1}^{n} \left\{ \int_{\widetilde{\mathbb{X}}} \left(\frac{\partial}{\partial \theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k) \right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k \right. \\
&\qquad \left. + \int_{\mathbb{X} \setminus \widetilde{\mathbb{X}}} \left(\frac{\partial}{\partial \theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x_k) \right) \underbrace{\mathsf{p}_{\boldsymbol{\theta}}(x_k)}_{=0}\, \mathrm{d}x_k \right\} \\
&= \sum_{k=1}^{n} \int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x_k)} \left(\frac{\partial}{\partial \theta_i} \mathsf{p}_{\boldsymbol{\theta}}(x_k) \right) \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k \\
&= \sum_{k=1}^{n} \int_{\widetilde{\mathbb{X}}} \frac{\partial}{\partial \theta_i} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \left\{ \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k + \int_{\mathbb{X} \setminus \widetilde{\mathbb{X}}} \underbrace{\mathsf{p}_{\boldsymbol{\theta}}(x_k)}_{=0}\, \mathrm{d}x_k \right\} \\
&= \sum_{k=1}^{n} \frac{\partial}{\partial \theta_i} \underbrace{\int_{\mathbb{X}} \mathsf{p}_{\boldsymbol{\theta}}(x_k)\, \mathrm{d}x_k}_{=1} \\
&= 0
\end{aligned}
$$

となる. よって (1) は示された.
(2) は明らか.
(3) を示すために次に注意する. $i, j = 1, 2, \ldots, d$ と $x \in \widetilde{\mathbb{X}}$ に対して

$$
\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial \theta_i \partial \theta_j} = \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)} \frac{\partial^2 \mathsf{p}(x|\,\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} - \left(\frac{\partial}{\partial \theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x) \right) \left(\frac{\partial}{\partial \theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(x) \right)
$$

となる. このとき

$$\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)} \left(\frac{\partial^2 \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i \partial\theta_j} \right) \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x = \frac{\partial^2}{\partial\theta_i\partial\theta_j} \left\{ \int_{\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x + \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \right\}$$
$$= \frac{\partial^2}{\partial\theta_i\partial\theta_j} \underbrace{\int_{\mathbb{X}} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x}_{=1} = 0$$

と

$$\int_{\widetilde{\mathbb{X}}} \left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(x) \right) \left(\frac{\partial}{\partial\theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(x) \right) \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x$$
$$= \mathsf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial\theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \right]$$

となる. このことより

$$\mathsf{E}_{\boldsymbol{\theta}} \left[\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(X_1)}{\partial\theta_i\partial\theta_j} \right] = \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x$$
$$+ \int_{\mathbb{X}\setminus\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x$$
$$= \int_{\widetilde{\mathbb{X}}} \frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j} \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x$$
$$= \underbrace{\int_{\widetilde{\mathbb{X}}} \frac{1}{\mathsf{p}_{\boldsymbol{\theta}}(x)} \left(\frac{\partial^2 \mathsf{p}_{\boldsymbol{\theta}}(x)}{\partial\theta_i\partial\theta_j} \right) \mathsf{p}_{\boldsymbol{\theta}}(x)\,\mathrm{d}x}_{=0}$$
$$- \mathsf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial\theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \right]$$
$$= -\mathsf{E}_{\boldsymbol{\theta}} \left[\left(\frac{\partial}{\partial\theta_i} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \left(\frac{\partial}{\partial\theta_j} \log \mathsf{p}_{\boldsymbol{\theta}}(X_1) \right) \right] \tag{6.1}$$

がわかる. よって

$$\{\mathcal{F}_{\boldsymbol{X}}(\boldsymbol{\theta})\}_{ij} = \mathsf{E} \left[\left(\sum_{k_1=1}^{n} \frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i} \right) \left(\sum_{k_2=1}^{n} \frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j} \right) \right]$$
$$= \sum_{k=1}^{n} \mathsf{E} \left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i} \frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_j} \right]$$
$$+ \sum_{k_1\neq k_2} \mathsf{E} \left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i} \frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j} \right]$$
$$= \sum_{k=1}^{n} \mathsf{E} \left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i} \frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_j} \right]$$
$$+ \sum_{k_1\neq k_2} \underbrace{\mathsf{E} \left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_1})}{\partial\theta_i} \right]}_{=0} \mathsf{E} \left[\frac{\partial \log \mathsf{p}_{\boldsymbol{\theta}}(X_{k_2})}{\partial\theta_j} \right]$$

$$= -\sum_{k=1}^{n} \mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log \mathsf{p}_{\boldsymbol{\theta}}(X_k)}{\partial\theta_i \partial\theta_j}\right] \qquad \left(\because (6.1)\right)$$
$$= -\mathsf{E}\left[\frac{\partial^2 \ell_n(\boldsymbol{\theta} \mid \boldsymbol{X})}{\partial\theta_i \partial\theta_j}\right] = -\mathsf{E}\big[\overset{\bullet\bullet}{\ell}_{n,ij}(\boldsymbol{\theta} \mid \boldsymbol{X})\big]$$

がわかる. □

例 6.4. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ で $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする. このとき, 統計的モデルは $\mathcal{P} = \{\mathsf{N}(\mu, \sigma^2);\, \mu \in \mathbb{R},\, \sigma > 0\}$ で与えられる. したがって

$$\Theta = \{\boldsymbol{\theta} = (\mu, \sigma) \in \mathbb{R}^2;\, \mu \in \mathbb{R},\, \sigma > 0\} = \mathbb{R} \times (0, \infty)$$

となる. このとき

$$\ell_n(\boldsymbol{\theta} | \boldsymbol{x}) = -\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{2\sigma^2} - n\log\sigma - n\log(\sqrt{2\pi}); \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

となる. よって

$$\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta} | \boldsymbol{x}) = \frac{\partial \ell_n(\boldsymbol{\theta} | \boldsymbol{x})}{\partial\mu} = \sum_{j=1}^{n} \frac{x_j - \mu}{\sigma^2},$$
$$\overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta} | \boldsymbol{x}) = \frac{\partial \ell_n(\boldsymbol{\theta} | \boldsymbol{x})}{\partial\sigma} = \sum_{j=1}^{n} \frac{(x_j - \mu)^2}{\sigma^3} - \frac{n}{\sigma}$$

となる. これらの期待値を取ると

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\{\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta} | \boldsymbol{X})\}^2\big] = \frac{n}{\sigma^2},$$
$$\mathsf{E}_{\boldsymbol{\theta}}\big[\{\overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta} | \boldsymbol{X})\}^2\big] = \frac{3n}{\sigma^2} - \frac{2n}{\sigma^2} + \frac{n}{\sigma^2} = \frac{2n}{\sigma^2},$$
$$\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet}{\ell}_{n,1}(\boldsymbol{\theta} | \boldsymbol{X})\, \overset{\bullet}{\ell}_{n,2}(\boldsymbol{\theta} | \boldsymbol{X})\big] = 0$$

となるので

$$\mathcal{F}_X(\boldsymbol{\theta}) = \frac{n}{\sigma^2}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

を得る.

一方

$$\overset{\bullet\bullet}{\ell}_{n,11} = \frac{\partial^2 \ell_n(\boldsymbol{\theta} | \boldsymbol{x})}{\partial\mu^2} = \frac{\partial}{\partial\mu}\left(\sum_{j=1}^{n} \frac{x_j - \mu}{\sigma^2}\right) = -\frac{n}{\sigma^2},$$
$$\overset{\bullet\bullet}{\ell}_{n,22} = \frac{\partial^2 \ell_n(\boldsymbol{\theta} | \boldsymbol{x})}{\partial\sigma^2} = \frac{\partial}{\partial\sigma}\left(\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{\sigma^3} - \frac{n}{\sigma}\right) = -3\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{\sigma^4} + \frac{n}{\sigma^2},$$

$$\overset{\bullet\bullet}{\ell}_{n,12} = \frac{\partial^2 \ell(\boldsymbol{\theta}|\,\boldsymbol{x})}{\partial\mu\partial\sigma} = \frac{\partial}{\partial\sigma}\left[\sum_{j=1}^{n}\frac{x_j-\mu}{\sigma^2}\right] = -2\sum_{j=1}^{n}\frac{x_j-\mu}{\sigma^3}$$

となる. よって

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,11}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = -\frac{n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,22}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = -\frac{2n}{\sigma^2},\ \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,12}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] = 0$$

なので

$$\begin{bmatrix} \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,11}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,12}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] \\ \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,12}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] & \mathsf{E}_{\boldsymbol{\theta}}\big[\overset{\bullet\bullet}{\ell}_{n,22}(\boldsymbol{\theta}|\,\boldsymbol{X})\big] \end{bmatrix} = -\mathcal{F}_X(\boldsymbol{\theta})$$

となる. □

6.2 指数型分布族

この節では, 統計的モデルでもっとも重要なものを導入し, そのモデルの性質を説明する.

6.2.1 1 変数の場合

定義 6.5. $d \in \mathbb{N}$ とし, $\mathbb{X} \subset \mathbb{R}^d$ を空でない部分集合とする. さらに $\Theta \subset \mathbb{R}$ も空でない部分集合とする. $\mathbb{X}$ 上の統計的モデル $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は 1 母数指数型分布族であるとは, Θ 上の実数値関数 A と $\mathsf{K}^\vee$, $\mathbb{X}$ 上の実数値関数 T と h が存在して, P_θ の p.d.f. または p.m.f. $\mathsf{p}_\theta^\vee$ [2] が

$$\mathsf{p}_\theta^\vee(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x})\exp\big\{A(\theta)T(\boldsymbol{x}) - \mathsf{K}^\vee(\theta)\big\} & (\boldsymbol{x} \in \mathbb{X} \subset \mathbb{R}^d) \\ 0 & (\boldsymbol{x} \notin \mathbb{X}) \end{cases} \tag{6.2}$$

の形で書けるときをいう. ただし

$$\mathsf{K}^\vee(\theta) = \begin{cases} \log\left(\displaystyle\sum_{\boldsymbol{x}\in\mathbb{X}} h(\boldsymbol{x})\exp\big\{A(\theta)T(\boldsymbol{x})\big\}\right) & (\text{離散型の場合}) \\ \log\left(\displaystyle\int_{\mathbb{X}} h(\boldsymbol{x})\exp\big\{A(\theta)T(\boldsymbol{x})\big\}\,\mathrm{d}\boldsymbol{x}\right) & (\text{連続型の場合}) \end{cases}$$

である.

[2] 後に別の母数化である自然指数型分布族を導入するときに記号を簡単にするために, ここでは「$\mathsf{p}^\vee$ と $\mathsf{K}^\vee$」という少し不自然な記号を用いた.

注意 6.6. (1) 関数 A, h, T の表現は一意ではない.
(2) あとの都合のために, p_θ に「$^\vee$」の記号をつけておく. □

例 6.7. (Poisson 分布族) $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする. その p.m.f. は

$$\mathsf{p}_\theta^\vee(x) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}\exp\bigl\{x\log\theta - \theta\bigr\} \qquad (x = 0,\, 1,\, \ldots)$$

である. したがって $\{\mathsf{Po}(\theta);\, \theta \in (0,\, \infty)\}$ は

$$d = 1,\, A(\theta) = \log\theta,\, \mathsf{K}^\vee(\theta) = \theta,\, T(x) = x,\, h(x) = \frac{1}{x!}$$

によって生成される 1 母数指数型分布族である. □

例 6.8. (2 項分布族) $X \sim \mathsf{Bino}(n,\, \theta)$ とする. ただし $n \in \mathbb{N}$, $0 < \theta < 1$ である. このとき X の p.m.f. は

$$\begin{aligned}\mathsf{p}_\theta^\vee(x) &= \left(\begin{array}{c} n \\ x \end{array}\right)\theta^x(1-\theta)^{n-x} \\ &= \left(\begin{array}{c} n \\ x \end{array}\right)\exp\left\{x\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\} \qquad (x = 0,\, 1,\, \ldots,\, n)\end{aligned}$$

と書ける. ただし

$$\left(\begin{array}{c} n \\ x \end{array}\right) = \frac{n!}{x!\times(n-x)!}, \qquad 0! = 1$$

である. したがって $\{\mathsf{Bino}(n,\, \theta);\, \theta \in (0,\, 1)\}$ は

$$d = 1,\, A(\theta) = \log\left(\frac{\theta}{1-\theta}\right),\, \mathsf{K}^\vee(\theta) = -n\log(1-\theta),\, T(x) = x,$$
$$h(x) = \left(\begin{array}{c} n \\ x \end{array}\right)$$

によって生成される 1 母数指数型分布族である. □

例 6.9. $\boldsymbol{X} = (Y,\, Z)^\top$, $Z,\, W \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\, 1)$ とし, $Y = Z + \theta W\,(\theta > 0)$ とする. このとき $\boldsymbol{X}$ の同時 p.d.f. は次で与えられる. $\boldsymbol{x} = (y,\, z)$ としたとき

$$\begin{aligned}\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \mathsf{p}_\theta^{\boldsymbol{X}}(y,\, z) &= \mathsf{p}^Z(z)\mathsf{p}^{Y|\,Z}(y|\,z) = \varphi(z)\theta^{-1}\varphi\left(\frac{y-z}{\theta}\right) \\ &= \frac{1}{2\pi\theta}\exp\left\{-\frac{1}{2}\left(z^2 + \frac{(y-z)^2}{\theta^2}\right)\right\}\end{aligned}$$

$$= \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}\exp\left\{-\frac{(y-z)^2}{2\theta^2} - \log\theta\right\}$$

と書ける. ただし p_Z は Z の p.d.f. とし, $\mathsf{p}_{Y|Z}$ は $Z = z$ を与えたときの Y の条件付き p.d.f., φ は $\mathsf{N}(0, 1)$ の p.d.f. で

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2} \qquad (z \in \mathbb{R})$$

である. したがって $\boldsymbol{X}$ の分布は

$$d = 2,\ A(\theta) = -\frac{1}{2\theta^2},\ \mathsf{K}^\vee(\theta) = \log\theta,\ T(\boldsymbol{x}) = (y-z)^2,$$
$$h(\boldsymbol{x}) = \frac{1}{2\pi}\exp\left\{-\frac{z^2}{2}\right\}$$

で生成される 1 母数指数型分布族に属する. □

1 母数指数型分布族に属する分布からのランダム標本の分布も 1 母数指数型分布族に属する. 特に, $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_\theta$ とする. ただし, $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は (6.2) で与えられる指数型分布族とする. すると $\{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}$ は $\boldsymbol{X}^{(n)} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m)$ に対する指数型分布族となる. ただし, $\mathbb{R}^{dn}$ 上の確率測度を

$$\mathsf{P}_\theta^{\otimes n} = \underbrace{\mathsf{P}_\theta \times \mathsf{P}_\theta \times \cdots \times \mathsf{P}_\theta}_{n\text{個}}$$

と定めた. このとき, $\mathsf{P}_\theta^{\otimes n}$ の同時 p.d.f. $\prod_{j=1}^n \mathsf{p}_\theta^\vee(\,\cdot\,)$ は次で与えられる. すると

$$\prod_{j=1}^n \mathsf{p}_\theta^\vee(\boldsymbol{x}_j) = \prod_{j=1}^n h(\boldsymbol{x}_j)\exp\left\{A(\theta)T(\boldsymbol{x}_j) - \mathsf{K}^\vee(\theta)\right\}$$
$$= \left[\prod_{j=1}^n h(\boldsymbol{x}_j)\right]\exp\left\{A(\theta)\sum_{j=1}^n T(\boldsymbol{x}_j) - n\mathsf{K}^\vee(\theta)\right\}$$

となる. したがって $\{\mathsf{P}_\theta^{\otimes n};\, \theta \in \Theta\}$ は

$$A(\theta), \quad \sum_{j=1}^n T(\boldsymbol{x}_j), \quad n\mathsf{K}^\vee(\theta), \quad \prod_{j=1}^n h(\boldsymbol{x}_j)$$

で生成される $\mathbb{R}^{dn}$ 上の 1 母数指数型分布族となる.

統計量 $\sum_{j=1}^n T(\boldsymbol{X}_j)$ は θ の自然十分統計量[3] となる.

[3] 十分統計量の定義は定義 6.13 を参照のこと.

6.3 指数型分布族の正準表示

(6.2) で表現された指数型分布族を θ でなく $\eta\big(=A(\theta)\big)$ で添え字付けることを考える. 1 母数指数型分布族は

$$\mathsf{p}_\eta(\boldsymbol{x}) = \begin{cases} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} & (\boldsymbol{x}\in\mathbb{X}\subset\mathbb{R}^d) \\ 0 & (\boldsymbol{x}\notin\mathbb{X}) \end{cases} \tag{6.3}$$

と書くことができる. ただし

$$\mathsf{K}(\eta) = \begin{cases} \log\Big(\displaystyle\sum_{\boldsymbol{x}\in\mathbb{X}} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x})\big\}\Big) & (\text{離散型の場合}) \\ \log\Big(\displaystyle\int_{\mathbb{X}} h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x})\big\}\,\mathrm{d}\boldsymbol{x}\Big) & (\text{連続型の場合}) \end{cases}$$

である. K の定義より, 直ちに $\mathsf{K}(\eta) = \mathsf{K}(A(\theta)) = \mathsf{K}^\vee(\theta)$ がわかる. さらに

$$\mathcal{E} := \{\eta\in\mathbb{R};\, \mathsf{K}(\eta) < \infty\}$$

とおく. $\mathcal{E}$ は $\mathbb{R}$ の区間となる. すると確率分布族

$$\Big\{\mathsf{p}_\eta(\boldsymbol{x}) = h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\};\, \eta\in\mathcal{E}\Big\}$$

は確率分布族

$$\Big\{\mathsf{p}_\theta(\boldsymbol{x}) = h(\boldsymbol{x})\exp\big\{A(\theta)T(\boldsymbol{x}) - \mathsf{K}(\theta)\big\};\, \theta\in\Theta\Big\}$$

を含む. 確率分布族 $\{\mathsf{p}_\eta(\boldsymbol{x});\, \eta\in\mathcal{E}\}$ は T, h によって生成された 正準(せいじゅん)指数型分布族 または 自然指数型分布族 といい, $\mathcal{E}$ を 自然母数空間 といい, T を自然十分統計量という.

例 6.10. (例 6.7 の続き) Poisson 分布族は

$$\begin{aligned} \mathsf{p}_\eta(x) &= \frac{1}{x!}\exp\big\{\eta x - e^\eta\big\} \qquad (x = 0,\, 1,\, 2,\ldots), \\ \eta &= \log\theta,\, h(x) = \frac{1}{x!},\, T(x) = x, \\ \exp\big(\mathsf{K}(\eta)\big) &= \sum_{x=0}^{\infty} h(x)\exp\big(\eta T(x)\big) = \sum_{x=0}^{\infty}\frac{\exp(\eta x)}{x!} = \sum_{x=0}^{\infty}\frac{\big(e^\eta\big)^x}{x!} = \exp\big(e^\eta\big), \\ \mathcal{E} &= \mathbb{R} \end{aligned}$$

となる. □

補題 6.11. [4]

$$\mathcal{E} := \left\{ \eta \in \mathbb{R};\, \mathsf{K}(\eta) := \log\Big(\mathsf{E}[\exp\{\eta T(\boldsymbol{X})\}] \Big) < \infty \right\}$$

とする. このとき, 関数 $\mathsf{K}(\eta)$ は $\mathcal{E}^\circ$ 上で無限回微分可能である. ただし $\mathcal{E}^\circ$ は $\mathcal{E}$ の内部である. さらに積分記号と微分記号の交換は可能である.

Proof. $\eta \in \mathcal{E}^\circ$ なので, ある $\epsilon > 0$ が存在して $[\eta - 2\epsilon,\, \eta + 2\epsilon] \subset \mathcal{E}^\circ$ とできる.

$$\begin{aligned}
\frac{d}{d\eta} \exp\big\{\mathsf{K}(\eta)\big\} &= \lim_{n\to\infty} \frac{\exp\left\{ \mathsf{K}\left(\eta + \frac{\epsilon}{n} \right) \right\} - \exp\big\{\mathsf{K}(\eta)\big\}}{\epsilon/n} \\
&= \lim_{n\to\infty} \int \frac{\exp\left\{ \left(\eta + \frac{\epsilon}{n} \right) T(\boldsymbol{x}) \right\} - \exp\big\{\mathsf{K}(\eta)\big\}}{\epsilon/n} h(\boldsymbol{x})\, \mathrm{d}x \\
&= \lim_{n\to\infty} \int \frac{\exp\big\{\eta T(\boldsymbol{x})\big\} \left[\exp\left\{ \frac{\epsilon\, T(\boldsymbol{x})}{n} \right\} - 1 \right]}{\epsilon/n} h(\boldsymbol{x})\, \mathrm{d}x \\
&=: \lim_{n\to\infty} \int \exp\big\{\eta T(\boldsymbol{x})\big\} f_n(\boldsymbol{x})\, \mathrm{d}x
\end{aligned}$$

と書ける. ただし

$$f_n(\boldsymbol{x}) = \frac{\exp\big\{\epsilon\, T(\boldsymbol{x})/n\big\} - 1}{\epsilon/n} h(\boldsymbol{x})$$

である. すると

$$\lim_{n\to\infty} f_n(\boldsymbol{x}) =: f(\boldsymbol{x}) = T(\boldsymbol{x}) h(\boldsymbol{x})$$

となる. ここで

$$|e^t - 1| \le |t| e^{|t|}; \qquad |t| \le e^{|t|} \qquad (t \in \mathbb{R})$$

に注意する. これらの不等式を用いると

$$\begin{aligned}
\left| \frac{\exp\big\{\epsilon\, T(\boldsymbol{x})/n\big\} - 1}{\epsilon/n} \right| &\le \left\{ \left| \frac{\epsilon\, T(\boldsymbol{x})}{n} \Big/ \frac{\epsilon}{n} \right| \right\} \exp\left\{ \left| \frac{\epsilon\, T(\boldsymbol{x})}{n} \right| \right\} \\
&\le \frac{1}{\epsilon} \big| \epsilon\, T(\boldsymbol{x}) \big| \exp\left\{ \left| \frac{\epsilon\, T(\boldsymbol{x})}{n} \right| \right\} \\
&\le \frac{1}{\epsilon} \exp\big\{ \big| \epsilon\, T(\boldsymbol{x}) \big| \big\}\, \exp\big\{ \big| \epsilon\, T(\boldsymbol{x}) \big| \big\}
\end{aligned}$$

[4] 証明には, 優収束定理を用いている.

$$
\begin{aligned}
&= \frac{1}{\epsilon}\exp\bigl\{\bigl|2\epsilon\, T(\boldsymbol{x})\bigr|\bigr\} \\
&\leq \frac{1}{\epsilon}\Bigl\{\exp\bigl(2\epsilon T(\boldsymbol{x})\bigr) + \exp\bigl(-2\epsilon T(\boldsymbol{x})\bigr)\Bigr\}
\end{aligned}
$$

を得る. よって

$$
|f_n(\boldsymbol{x})| \leq \frac{1}{\epsilon}\Bigl\{\exp\bigl(2\epsilon T(\boldsymbol{x})\bigr) + \exp\bigl(-2\epsilon T(\boldsymbol{x})\bigr)\Bigr\}h(\boldsymbol{x})
$$

となるので

$$
\begin{aligned}
\Bigl|\exp\bigl\{\eta T(\boldsymbol{x})\bigr\}f_n(\boldsymbol{x})\Bigr| &\leq \frac{1}{\epsilon}\Bigl\{\exp\bigl((\eta+2\epsilon)T(\boldsymbol{x})\bigr) + \exp\bigl((\eta-2\epsilon)T(\boldsymbol{x})\bigr)\Bigr\}h(\boldsymbol{x}) \\
&=: g(\boldsymbol{x})
\end{aligned}
$$

を得る. $[\eta-2\epsilon,\, \eta+2\epsilon] \subset \mathcal{E}^\circ$ から

$$
\int g(\boldsymbol{x})\,\mathrm{d}x < \infty
$$

となる. よって優収束定理から

$$
\begin{aligned}
\frac{d}{d\eta}\exp\bigl\{\mathsf{K}(\eta)\bigr\} &= \lim_{n\to\infty}\int \exp\bigl\{\eta T(\boldsymbol{x})\bigr\}f_n(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int \exp\bigl\{\eta T(x)\bigr\}T(\boldsymbol{x})h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}
\end{aligned}
$$

がわかる. この操作を繰り返して行けば, $\mathsf{K}(\eta)$ は無限回微分可能なことがわかる. □

定理 6.12. X は (6.3) で与えられた自然指数型分布族に属する分布 $\mathsf{p}_\eta(\boldsymbol{x})$ に従うとする. η は $\mathcal{E}$ の内点としたとき, $T(\boldsymbol{X})$ の積率母関数は原点の近傍で存在し

$$
M_T(s) = \exp\bigl[\mathsf{K}(s+\eta) - \mathsf{K}(\eta)\bigr]
$$

で与えられる. ただし, s は 0 のある近傍に含まれるとする. さらに

$$
\mathsf{E}[T(\boldsymbol{X})] = \dot{\mathsf{K}}\,(\eta); \qquad \mathsf{Var}[T(\boldsymbol{X})] = \ddot{\mathsf{K}}\,(\eta) \tag{6.4}
$$

である. ただし

$$
\dot{\mathsf{K}}\,(\eta) = \frac{\mathrm{d}\mathsf{K}}{\mathrm{d}\eta}(\eta), \qquad \ddot{\mathsf{K}}\,(\eta) = \frac{\mathrm{d}^2\mathsf{K}}{\mathrm{d}\eta^2}(\eta)
$$

である.

Proof. 連続型分布の場合について証明する. 離散型の場合には積分を和に変更すればよい. K の定義に注意すると

$$\begin{aligned} M(s) &= \mathsf{E}\big[\exp(sT(\boldsymbol{X}))\big] \\ &= \int_{\mathbb{X}} h(\boldsymbol{x}) \exp\big[(s+\eta)T(\boldsymbol{x}) - \mathsf{K}(\eta)\big] \,\mathrm{d}\boldsymbol{x} \\ &= \exp\big[\mathsf{K}(s+\eta) - \mathsf{K}(\eta)\big] \int_{\mathbb{X}} h(x) \exp\big[(s+\eta)T(\boldsymbol{x}) - \mathsf{K}(s+\eta)\big] \,\mathrm{d}\boldsymbol{x} \\ &= \exp\big[\mathsf{K}(s+\eta) - \mathsf{K}(\eta)\big] \end{aligned}$$

から 1 番目の主張はわかる. 最後の等号は $s+\eta \in \mathcal{E}^\circ$ となるように s をとると

$$\int_{\mathbb{X}} h(\boldsymbol{x}) \exp\big[(s+\eta)T(\boldsymbol{x}) - \mathsf{K}(s+\eta)\big] \,\mathrm{d}\boldsymbol{x} = 1$$

となることよりわかる. 次に, $\mathsf{K}(\,\cdot\,)$ の定義と補題 6.11 から

$$\begin{aligned} \frac{\mathrm{d}\mathsf{K}}{\mathrm{d}\eta}(\eta) &= \frac{\mathrm{d}}{\mathrm{d}\eta} \log\left(\int_{\mathbb{X}} \exp\big\{\eta T(\boldsymbol{x})\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \right) \\ &= \frac{\dfrac{\mathrm{d}}{\mathrm{d}\eta} \displaystyle\int_{\mathbb{X}} \exp\big\{\eta T(\boldsymbol{x})\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\displaystyle\int_{\mathbb{X}} \exp\big\{\eta T(\boldsymbol{x})\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}} \\ &= \frac{\displaystyle\int_{\mathbb{X}} \frac{\mathrm{d}}{\mathrm{d}\eta} \exp\big\{\eta T(\boldsymbol{x})\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\exp\big\{\mathsf{K}(\eta)\big\}} \\ &= \frac{\displaystyle\int_{\mathbb{X}} T(\boldsymbol{x}) \exp\big\{\eta T(\boldsymbol{x})\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x}}{\exp\big\{\mathsf{K}(\eta)\big\}} \\ &= \int_{\mathbb{X}} T(\boldsymbol{x}) \exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &= \mathsf{E}\big[T(\boldsymbol{X})\big] \end{aligned}$$

から (6.4) の 1 番目の等号が示せた. 同様に

$$\begin{aligned} \frac{\mathrm{d}^2\mathsf{K}}{\mathrm{d}\eta^2}(\eta) &= \frac{\mathrm{d}}{\mathrm{d}\eta} \int_{\mathbb{X}} T(\boldsymbol{x}) \exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} h(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{X}} T(\boldsymbol{x}) h(\boldsymbol{x}) \frac{\mathrm{d}}{\mathrm{d}\eta} \exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} \,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{X}} T(\boldsymbol{x}) h(\boldsymbol{x}) \exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} \frac{\mathrm{d}}{\mathrm{d}\eta} \Big(\eta T(\boldsymbol{x}) - \mathsf{K}(\eta) \Big) \,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{X}} T^2(\boldsymbol{x}) h(\boldsymbol{x}) \exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\big\} \,\mathrm{d}\boldsymbol{x} \end{aligned}$$

$$
\begin{aligned}
&\quad - \underbrace{\frac{\mathrm{d}}{\mathrm{d}\eta}\mathsf{K}(\eta)}_{=\mathsf{E}\left[T(\boldsymbol{X}\right]} \int_{\mathbb{X}} T(\boldsymbol{x})h(\boldsymbol{x}) \exp\bigl\{\eta T(\boldsymbol{x}) - \mathsf{K}(\eta)\bigr\}\,\mathrm{d}\boldsymbol{x} \\
&= \mathsf{E}\bigl[T^2(\boldsymbol{X})\bigr] - \bigl\{\mathsf{E}\bigl[T(\boldsymbol{X})\bigr]\bigr\}^2 \\
&= \mathsf{Var}\bigl[T(\boldsymbol{X})\bigr]
\end{aligned}
$$

から (6.4) の 2 番目の等号が示せた. □

6.4 十分統計量

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の $\mathbb{X}$ 値確率変数列で, 互いに独立で各々同一の確率分布を持つとする. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ とおき $B \in \mathcal{B}(\mathbb{X}^n)$ に対し

$$
\begin{aligned}
\mathsf{P}^{\boldsymbol{X}}(B) &= \mathsf{Pr}\bigl\{\boldsymbol{X}^{-1}(B)\bigr\} \\
&= \mathsf{Pr}\biggl(\bigl\{\omega \in \Omega;\, \boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \in B\bigr\}\biggr)
\end{aligned}
$$

とおく. さらに $B \in \mathcal{B}(\mathbb{X})$ に対し

$$
\mathsf{P}^{X_j}(B) = \mathsf{Pr}\biggl(\bigl\{\omega \in \Omega;\, X_j(\omega) \in B\bigr\}\biggr) \qquad (j = 1, 2, \ldots, n)
$$

と書いたとき

$$
\mathsf{P}^{\boldsymbol{X}} = \mathsf{P}^{X_1} \times \mathsf{P}^{X_2} \times \cdots \times \mathsf{P}^{X_n}
$$

と書ける. $\boldsymbol{X}$ により誘導された確率空間を $(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n), \mathsf{P}^{\boldsymbol{X}})$ と書く.

$(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 上の統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ と書く. ただし θ は母数で Θ は母数空間である. さらにある $\theta^* \in \Theta$ が存在して

$$
\mathsf{P}^{X_1} = \mathsf{P}_{\theta^*}
$$

とする.

次に統計量 $\boldsymbol{T}(\boldsymbol{X})$ を

$$
\boldsymbol{T} :\, (\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))
$$

なる可測関数で θ に依存しないものとする. ただし $k \in \mathbb{N}$ である.

定義 6.13. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta)$ とする. $\forall B \in \mathcal{B}(\mathbb{X}^n)$ に対して, 統計量 $\boldsymbol{T}(\boldsymbol{X})$ を与えたときの $\boldsymbol{X}$ の条件付き確率

$$
\mathsf{P}_\theta^{\otimes n}(B|\,\boldsymbol{T})
$$

が $\theta \in \Theta$ に無関係であるとき, $\boldsymbol{T}$ は $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ に対する<u>十分統計量</u>(sufficient statistic) であるという.

定理 6.14. (Fisher-Neyman の因子分解定理) 統計的モデルを $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ とする．ただし θ は母数で Θ は母数空間とする．$\boldsymbol{X}$ は確率空間 $\left(\mathbb{X}^n,\, \mathcal{B}(\mathbb{X}^n),\, \mathsf{P}_\theta^{\otimes n}\right)$ 上の大きさが n の標本とし, 同時 p.d.f.(または p.m.f.) $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ [5] を持つとする．このとき統計量 $\boldsymbol{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^k$ が統計的モデル $\mathcal{P}$ に対する十分統計量であるための必要十分条件は, $\boldsymbol{X}$ の同時 p.d.f.(または同時 p.m.f.) $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ が

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = g_\theta\{\boldsymbol{T}(\boldsymbol{x})\}h(\boldsymbol{x}) \qquad (\boldsymbol{x} \in \mathbb{X}^n) \tag{6.5}$$

の形で表されるときである．ここで $g_\theta\{\boldsymbol{T}(\boldsymbol{x})\}$ と $h(\boldsymbol{x})$ は非負値関数で $h(\boldsymbol{x})$ は θ に無関係な関数で, g_θ は $\boldsymbol{T}$ を通してのみ $\boldsymbol{x}$ に依存する．

Proof. $\boldsymbol{X}$ が離散型のときのみの証明を与える．連続型の場合については, 注意 6.15 を参考のこと．

$\boldsymbol{T}$ は十分統計量とする．$\boldsymbol{T}$ の p.m.f. を $\mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})\,(\boldsymbol{t} \in \mathbb{R}^k)$ と書く．$\boldsymbol{T} = \boldsymbol{t}$ が与えられたときの $\boldsymbol{X}$ の条件付き p.m.f. を $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ とする．仮定から $\boldsymbol{T}$ は十分統計量なので $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t})$ は θ に依存しない．条件付き p.m.f. の定義から

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t}) \tag{6.6}$$

となる．(6.6) において

$$h(\boldsymbol{x}) = \mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}(\boldsymbol{x}|\,\boldsymbol{T} = \boldsymbol{t}), \quad g_\theta\{\boldsymbol{T}(\boldsymbol{x})\} = \mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})$$

とおけば, (6.5) の形になる．

次に (6.5) と書けたときに $\boldsymbol{T}$ は十分統計量であることを示す．まず

$$\begin{aligned}
\mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t}) &= \sum_{\boldsymbol{y} \in \mathbb{X}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} \mathsf{p}^{\boldsymbol{X}}(\boldsymbol{y}|\,\theta) \\
&= \sum_{\boldsymbol{y} \in \mathbb{X}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} g_\theta\big\{\boldsymbol{T}(\boldsymbol{y})\big\}h(\boldsymbol{y}) \\
&= g_\theta(\boldsymbol{t}) \sum_{\boldsymbol{y} \in \mathbb{X}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})
\end{aligned}$$

に注意する．(6.5) から $\boldsymbol{T} = \boldsymbol{t}$ を与えたときの $\boldsymbol{X}$ の条件付き p.m.f. は

$$\begin{aligned}
\frac{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_\theta^{\boldsymbol{T}}(\boldsymbol{t})} &= \frac{g_\theta(\boldsymbol{t})h(\boldsymbol{x})}{g_\theta(\boldsymbol{t}) \sum_{\boldsymbol{y} \in \mathbb{X}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})} \\
&= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y} \in \mathbb{X}^n;\, \boldsymbol{T}(\boldsymbol{y}) = \boldsymbol{t}} h(\boldsymbol{y})}
\end{aligned}$$

となり, $\mathsf{p}^{\boldsymbol{X}|\boldsymbol{T}}$ は θ に依存しない．よって $\boldsymbol{T}$ は $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ に対する十分統計量である．

□

[5] 離散型のときは同時 p.m.f. $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ をもつ．

注意 6.15. Fisher-Neyman の因子分解定理の測度論的な証明には Radon-Nikodym の微分に関わる知識が必要になる．これについては定理 B.49 を参照のこと．

例 6.16. $n \in \mathbb{N}$ とし，$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 \leq \theta \leq 1)$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.m.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j}\mathbb{1}_{\{0,1\}^n}(\boldsymbol{x}),$$
$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n,$$
$$\mathbb{1}_{\{0,1\}^n}(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in \{0,1\}^n) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる．したがって

$$T(\boldsymbol{x}) = \sum_{j=1}^n x_j, \quad g_\theta\{T(\boldsymbol{x})\} = \theta^{T(\boldsymbol{x})}(1-\theta)^{n-T(\boldsymbol{x})}, \quad h(\boldsymbol{x}) = \mathbb{1}_{\{0,1\}^n}(\boldsymbol{x})$$

とおけば，定理 6.14 から $T(\boldsymbol{X}) = X_1 + X_2 + \cdots + X_n$ は

$$\mathcal{P} = \{\mathsf{p}_\theta(x) = \theta^x(1-\theta)^{1-x}\mathbb{1}_{\{0,1\}}(x);\ 0 \leq \theta \leq 1\}$$

の十分統計量であることがわかる． □

例 6.17. $\theta \in \mathbb{R}$ とし，$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, 1)$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}\sum_{j=1}^n (x_j - \theta)^2\right\} \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

である．$T(\boldsymbol{x}) = n^{-1}\sum_{j=1}^n x_j$ とすると

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \exp\left\{nT(\boldsymbol{x})\theta - \frac{n\theta^2}{2}\right\} \times \left(\frac{1}{2\pi}\right)^{n/2} \times \exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

となる．よって (6.5) で

$$g_\theta\big\{T(\boldsymbol{x})\big\} = \exp\left\{nT(\boldsymbol{x})\theta - \frac{n\theta^2}{2}\right\}, \quad h(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{n/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^n x_j^2\right\}$$

とおけば，定理 6.14 から $T(\boldsymbol{X}) = (X_1 + X_2 + \cdots + X_n)/n$ は $\mathcal{P} = \{\mathsf{N}(\theta, 1);\ \theta \in \mathbb{R}\}$ の十分統計量であることがわかる． □

例 6.18. $n \in \mathbb{N}\,(n \geq 2)$ とし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ とする. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2\right\}$$

となる. ただし $\theta = (\mu, \sigma^2)$ である. ここで

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad s_n^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x}_n)^2$$

とおけば

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}\big(s_n^2 + (\overline{x}_n - \mu)^2\big)\right\}$$

と書き直せる. (6.5) において

$$T(\boldsymbol{x}) = (\overline{x}_n, s_n^2), \quad g_\theta\big\{T(\boldsymbol{x})\big\} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{n}{2\sigma^2}\big(s_n^2 + (\overline{x}_n - \mu)^2\big)\right\},$$

$h(\boldsymbol{x}) = 1$

とすれば, 定理 6.14 から

$$\boldsymbol{T}(\boldsymbol{X}) = \left(\frac{1}{n}\sum_{j=1}^{n} X_j (=: \overline{X}_n), \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2\right)$$

は $\mathcal{P} = \{\mathsf{N}^{\otimes n}(\mu, \sigma^2);\, \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ の十分統計量である. ただし, $\mathsf{N}^{\otimes n}(\mu, \sigma^2)$ は $\mathsf{N}(\mu, \sigma^2)$ の n 個の直積分布である. □

6.5 統計量の最小十分性と完備性

6.5.1 十分統計量の最小性

与えられた統計的モデル $\mathcal{P}$ に対して, たくさんの十分統計量が存在する. $\boldsymbol{S}$ と $\boldsymbol{T}$ が $\mathsf{P} \in \mathcal{P}$ に対する十分統計量で, S の値域上で定義された可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S})$$

と表現できたとする. すると

$$\sigma(\boldsymbol{T}) \subset \sigma(\boldsymbol{S})$$

となる. このことから $\boldsymbol{T}$ は統計的モデル $\mathcal{P}$ の情報を失くことなく, データの情報を S より縮約していることがわかる. したがって T は S よりも有用な統計量となる. この観点を定式化してみよう.

定義 6.19. すべての $\mathsf{P} \in \mathcal{P}$ に対して, $\mathsf{P}(A) = 0$ となる事象 A を除いて, ある命題 M が成立しているとき

$$M \qquad \text{a.s.}\,\mathcal{P}$$

と記すことにする.

定義 6.20. (最小十分性) $\mathcal{P}$ を統計的モデルとし, 統計量 $\boldsymbol{T}$ を $\mathsf{P} \in \mathcal{P}$ に対する十分統計量とする. 統計量 T は $\mathsf{P} \in \mathcal{P}$ に関して **最小十分** であるとは, 任意の他の十分統計量 S に対して可測関数 ψ が存在して

$$\boldsymbol{T} = \psi(\boldsymbol{S}), \qquad \text{a.s.}\,\mathcal{P}$$

と表現できるときをいう. もちろん ψ は $\boldsymbol{S}$ ごとに定まればよい.

注意 6.21. $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を統計的モデルとする. 2 つの母数 $\theta_1, \theta_2\, (\in \Theta)$ の疑似距離を

$$\rho(\theta_1, \theta_2) = \sup_{B \in \mathcal{B}(\mathbb{X})} \left|\mathsf{P}_{\theta_1}(B) - \mathsf{P}_{\theta_2}(B)\right|$$

に関して Θ が可分[6] であるとき, $\mathcal{P}$ に対する最小十分統計量が存在することが知られている. 証明は [45, pp.78-81] を参照のこと. □

例 6.22. $\mathcal{P} = \left\{\mathsf{P} \text{ は開区間 } (\theta, \theta+1) \text{ 上の一様分布};\ \theta \in \mathbb{R}\right\}$ とする. $n \geq 2\, (\in \mathbb{N})$ とし

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P} \in \mathcal{P}$$

とする. Lebesgue 測度に関する $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \Pi_{j=1}^n \mathbb{1}_{(\theta, \theta+1)}(x_j) = \mathbb{1}_{(x_{(n)}-1,\, x_{(1)})}(\theta),$$
$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

となる. ただし $x_{(1)}, x_{(n)}$ は $x_1, x_2, \ldots, x_n$ の最小値と最大値である.

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}, \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

とおくと Fisher-Neyman の因子分解定理から $\boldsymbol{T} = (X_{(1)}, X_{(n)})$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量である.

次に, 十分統計量 $\boldsymbol{T}$ が最小性もみたすことを示す.

$$\begin{aligned} &\theta < x_j < \theta + 1 \quad (\forall j \in \{1, 2, \ldots, n\}) \\ &\Leftrightarrow\ \theta < x_{(1)} < x_{(n)} < \theta + 1 \end{aligned} \tag{6.7}$$

[6] Θ の可算無限集合 C が存在して, 任意の θ と ϵ に対して, $\theta_c \in C$ をうまく選ぶと $\rho(\theta, \theta_c) < \epsilon$ とできることである.

となることに注意する. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき

$$x_{(1)} = \sup\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}, \quad x_{(n)} = 1 + \inf\{\theta \in \mathbb{R};\, \mathsf{p}(\boldsymbol{x}|\,\theta) > 0\}$$

となる. ここで統計量 $\boldsymbol{S}$ は $\mathsf{P} \in \mathcal{P}$ に対して十分であると仮定する. Fisher-Neyman の因子分解定理から可測関数 g_θ と h が存在して

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = g_\theta\big(\boldsymbol{S}(\boldsymbol{x})\big)h(\boldsymbol{x})$$

と書ける. $A := \{\boldsymbol{x} \in \mathbb{R}^n;\, h(\boldsymbol{x}) > 0\}$ 上で (6.7) が成立するので, ある可測関数 ψ が存在して A 上で

$$\boldsymbol{T} := (x_{(1)},\, x_{(n)}) = \psi\big(\boldsymbol{S}(\boldsymbol{x})\big)$$

と表現できる. さらに $\mathsf{Pr}(\boldsymbol{X} \in A) = 0$ なので $\boldsymbol{T}$ は最小十分統計量となる. ここでは, 最小十分統計量であることを直接的に確認したが, 次の定理 6.23 の判定条件を用いると簡単に最小性を確認できる. □

定理 6.23. $\big(\mathbb{X},\, \mathcal{B}\big)$ を標本空間とし, その上の統計的モデル

$$\mathcal{P} = \{\, \mathsf{P}_\theta;\, \theta \in \Theta \,\}$$

は測度 μ で支配されているとする. また, P_θ の μ に関する p.d.f. または p.m.f. $\mathsf{p}_\theta(x)$ は

$$\mathsf{p}_\theta(x) > 0 \qquad (x \in \mathbb{X},\ \theta \in \Theta)$$

をみたすと仮定する. T が十分統計量であり, 任意の $x, y \in \mathbb{X}$ に対して

$$\theta \longmapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

が θ に依存しないならば

$$T(x) = T(y)$$

が成り立つとする. このとき, T は最小十分統計量である.

Proof. $\widetilde{T}$ を任意の十分統計量とする. T が最小十分統計量であることを示すためには, T が $\widetilde{T}$ の関数として表されること, すなわち, ある関数 f が存在して

$$T = f \circ \widetilde{T}$$

と書けることを示せばよい. $\widetilde{T}$ は十分統計量であるから, Fisher–Neyman の因子分解定理より

$$\mathsf{p}_\theta(x) = \widetilde{g}_\theta\big(\widetilde{T}(x)\big)\widetilde{h}(x)$$

と書くことができる. ここで

$$\widetilde{T}(x) = \widetilde{T}(y)$$

を満たす任意の $x, y \in \mathbb{X}$ をとる. すると

$$\begin{aligned}\frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)} &= \frac{\widetilde{g}_\theta\big(\widetilde{T}(x)\big)\widetilde{h}(x)}{\widetilde{g}_\theta\big(\widetilde{T}(y)\big)\widetilde{h}(y)} \\ &= \frac{\widetilde{h}(x)}{\widetilde{h}(y)}\end{aligned}$$

となる. したがって

$$\theta \longmapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

は θ に依存しない. 定理の仮定から

$$T(x) = T(y)$$

が成り立つ. 以上により

$$\widetilde{T}(x) = \widetilde{T}(y) \quad \Longrightarrow \quad T(x) = T(y)$$

である. したがって, T は $\widetilde{T}$ の値だけによって一意に定まり, ある関数 f が存在して

$$T = f \circ \widetilde{T}$$

と書くことができる. $\widetilde{T}$ は任意の十分統計量であったから, T は任意の十分統計量の関数として表される. よって T は最小十分統計量である. □

例 6.24. $\big(\mathbb{X}, \mathcal{B}\big)$ を標本空間とし, $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を $(\mathbb{X}, \mathcal{B})$ 上の d 母数指数型分布族とする. すなわち p.d.f. $\mathsf{p}(x|\,\theta)\,(\theta \in \Theta)$ が

$$\mathsf{p}_\theta(x) = \exp\{\big(\eta(\theta)|\,T(x)\big) - \mathsf{K}(\theta)\}h(x)$$

と書ける. ただし $\mathsf{K} : \Theta \longrightarrow \mathbb{R}$ および $\eta : \Theta \longrightarrow \mathbb{R}^d$ であり, $T \in \mathbb{R}^d$ で, $(\,\cdot\,|\,\cdot\,)$ は $\mathbb{R}^d$ の Euclid 内積である. Fisher-Neyman の因子分解定理から T は統計的モデル $\mathcal{P}$ に対する十分統計量となる. データ x と y は, 関数

$$\theta \mapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

が定数となるものとする. するとある θ に依存しない定数 c (x と y に依存してよい) があって

$$\big(\eta(\theta)|\,T(x)\big) = \big(\eta(\theta)|\,T(y)\big) + c$$

と書けることになる. 任意の点 $\theta_0, \theta_1 \in \Theta\,(\theta_0 \neq \theta_1)$ に対して

$$\big(\big[\eta(\theta_0) - \eta(\theta_1)\big]|\,T(x)\big) = \big(\big[\eta(\theta_0) - \eta(\theta_1)\big]|\,T(y)\big)$$

$$\Leftrightarrow \big(\big[\eta(\theta_0)-\eta(\theta_1)\big|\,T(x)-T(y)\big)=0$$

を得る. $\mathrm{span}\Big[\{\eta(\theta_0)-\eta(\theta_1)\in\mathbb{R}^d:\,\theta_0,\,\theta_1\in\Theta\}\Big]=\mathbb{R}^d$ のとき, $T(x)=T(y)$ となり, T は最小十分統計量となる. □

例 6.25. $n\geq 2\,(n\in\mathbb{N})$ と $\theta\in\mathbb{R}$ とし, $X_1,\,X_2,\,\ldots,\,X_n$ は独立同一に分布する確率変数列で, 共通の p.d.f.

$$\mathsf{p}_\theta(x)=\frac{1}{2}e^{-|x-\theta|}\qquad(x\in\mathbb{R})$$

を持つとする. $\boldsymbol{X}:=(X_1,\,X_2,\,\ldots,\,X_n)$ の同時 p.d.f. $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{2^n}\exp\Big\{-\sum_{i=1}^n|x_i-\theta|\Big\}\qquad\big(\boldsymbol{x}=(x_1,\,x_2,\,\ldots,\,x_n)\in\mathbb{R}^n\big)$$

となる. $X_{(1)}\leq X_{(2)}\leq\cdots\leq X_{(n)}$ を順序統計量とする. Fisher-Neyman の因子分解定理から $\boldsymbol{T}=(X_{(1)},\,X_{(2)},\,\ldots,\,X_{(n)})$ は十分統計量となる. データを $x=(x_1,\,x_2,\,\ldots,\,x_n)$ と $y=(y_1,\,y_2,\,\ldots,\,y_n)$ としたとき, 関数

$$\theta\mapsto\sum_{i=1}^n|x_i-\theta|-\sum_{i=1}^n|y_i-\theta|\tag{6.8}$$

は θ に依存しない. この二つの θ の関数は区分的に連続でり, その微分が不連続となる点は $x_{(1)},\,x_{(2)},\,\ldots,\,x_{(n)}$ と $y_{(1)},\,y_{(2)},\,\ldots,\,y_{(n)}$ である. (6.8) で与えられた関数が θ に関して定数であるときは, 微分のジャンプの位置が一致する. したがって, $x_{(i)}=y_{(i)}\,(i=1,\,2,\,\ldots,\,n)$ となるので, $\boldsymbol{T}$ は最小十分であることがわかる.

6.5.2　十分統計量の完備性

定義 6.26. $k\in\mathbb{N}$ とし, $\mathbb{T}\subset\mathbb{R}^k$ とする. $\mathcal{P}$ を統計的モデルとし, $X_1,\,X_2,\,\ldots,\,X_n\overset{\text{i.i.d.}}{\sim}\mathsf{P}\in\mathcal{P}$ とする. $\boldsymbol{T}(\boldsymbol{X})\big(\in\mathbb{T}\big)$ を $\mathsf{P}\in\mathcal{P}$ に対する十分統計量とする. ただし, $\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_n)^\top$ と書いた. 統計量 $\boldsymbol{T}$ は $\mathsf{P}\in\mathcal{P}$ に対して <u>完備十分</u> であるとは, 任意の可測関数 $f:\mathbb{T}\to\mathbb{R}$ に対して

$$\mathsf{E}\big[f(\boldsymbol{T})\big]=0\,(\forall\mathsf{P}\in\mathcal{P})\;\Rightarrow\;f(\boldsymbol{T})=0,\,\text{a.s.}\,\mathcal{P}$$

が成り立つときをいう.

次の命題は, 指数型分布族の十分統計量は完備であることを主張している. 本講義録では証明していない Fubini の定理と Laplace 変換の一意性を用いてこの命題を証明することができる.

命題 6.27. $k, n \in \mathbb{N}$ とする. $\boldsymbol{\eta} \in \mathbb{R}^k$, $\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とし

$$\mathsf{p}_{\boldsymbol{\eta}}(\boldsymbol{x}) = h(\boldsymbol{x}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{T}\bigl(x\bigr) - \mathsf{K}(\boldsymbol{\eta})\bigr\} \qquad (\boldsymbol{x} \in \mathbb{X}^n)$$

とする. ただし

$$\mathcal{E} := \left\{\boldsymbol{\eta} \in \mathbb{R}^k;\, \mathsf{K}(\boldsymbol{\eta}) := \int_{\mathbb{X}^n} \exp\{\boldsymbol{\eta}^\top \boldsymbol{T}(\boldsymbol{x})\} h(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} < \infty\right\}$$

は $\mathbb{R}^k$ の開集合を含み

$$\boldsymbol{T} : \mathbb{X}^n \to \mathbb{R}^k, \quad h : \mathbb{X}^n \to \mathbb{R}, \quad \mathsf{K} : \mathcal{E} \to \mathbb{R}$$

は可測関数とする. $\boldsymbol{X} \sim \mathsf{P}\,(\mathsf{P} \in \mathcal{P} := \{\mathsf{p}_{\boldsymbol{\eta}}(\boldsymbol{x});\, \boldsymbol{\eta} \in \mathcal{E}\})$ としたとき, $\boldsymbol{T}(\boldsymbol{X})$ は $\mathsf{P} \in \mathcal{P}$ に対する完備かつ十分統計量である.

Proof. Fisher-Neyman の因子分解定理から $\boldsymbol{T}$ は $\mathsf{P} \in \mathcal{P}$ に対する十分統計量であることは明らか. つぎに $f : \mathbb{R}^k \to \mathbb{R}$ を可測関数とし

$$\mathsf{E}\bigl[f(\boldsymbol{T})\bigr] = 0 \quad (\forall \boldsymbol{\eta} \in \mathcal{E})$$

とする. $\mathsf{E}\bigl[|f(\boldsymbol{T})|\bigr] < \infty$ なので

$$\int_{\mathbb{T}} f(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t} - \mathsf{K}(\boldsymbol{\eta})\bigr\}\, \mathrm{d}\nu(\boldsymbol{t}) = 0$$

と書ける. ただし $\mathbb{T} = \{\boldsymbol{T}(\boldsymbol{x}) \in \mathbb{R}^k;\, \boldsymbol{x} \in \mathbb{X}\}$ で, $\nu(B) := \displaystyle\int_{\boldsymbol{T}^{-1}(B)} h(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}\, (B \in \mathcal{B}(\mathbb{R}^d)$ である. $\boldsymbol{\eta}_0 \in \mathcal{E}$ の内点とし, $\epsilon > 0$ を十分小さく取ると

$$\int f_+(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t}\bigr\}\, \mathrm{d}\nu(\boldsymbol{t}) = \int f_-(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}^\top \boldsymbol{t}\bigr\}\, \mathrm{d}\nu(\boldsymbol{t})$$
$$\bigl(\forall \boldsymbol{\eta} \in N(\boldsymbol{\eta}_0) := \{\boldsymbol{\eta} \in \mathcal{R}^k;\, |\boldsymbol{\eta} - \boldsymbol{\eta}_0|_{2,k} < \epsilon\bigr)$$

となる. ここで

$$f_+(\boldsymbol{t}) := \max\{f(\boldsymbol{t}), 0\}, \quad f_-(\boldsymbol{t}) := \max\{-f(\boldsymbol{t}), 0\}$$

で, $|\cdot|_{2,k}$ は $\mathbb{R}^k$ 上の Euclid の距離である. 特に

$$\int f_+(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\bigr\}\, \mathrm{d}\nu(\boldsymbol{t}) = \int f_-(\boldsymbol{t}) \exp\bigl\{\boldsymbol{\eta}_0^\top \boldsymbol{t}\bigr\}\, \mathrm{d}\nu(\boldsymbol{t}) =: c$$

となる. $c = 0$ のとき, $f(\boldsymbol{t}) = 0\,(\text{a.s.})$ となるので, この場合は明らか. $c > 0$ のとき, Laplace 変換の一意性[7] から

$$f_+(\boldsymbol{t}) = f_-(\boldsymbol{t}), \quad \text{a.s.}$$

となる. よって $f(\boldsymbol{t}) = f_+(\boldsymbol{t}) - f_-(\boldsymbol{t}) = 0$ がわかる. □

注意 6.28. 完備十分統計量は最小十分統計量であることが知られている. しかし最小十分統計量で完備でないものの存在が知られている. □

[7] Laplace 変換の一意性の証明は次章の定理 6.31 を参照のこと.

6.6 補遺: Laplace 変換の一意性

定義 6.29. $[0, \infty)$ 上に集中した分布 P について

$$f(s) = \int_{[0,\infty)} e^{-sx} \, \mathrm{d}\mathsf{P}(x)$$

を P の Laplace 変換という. また, 非負値確率変数 X の Laplace 変換とは

$$f(s) = \mathsf{E}\big[e^{-sX}\big] = \int_{[0,\infty)} e^{-sx} \, \mathrm{d}\mathsf{P}(x) \qquad (s \geq 0)$$

のことである.

補題 6.30. 区間 $[0, 1]$ に集中した分布の分布関数 G_1, G_2 に対して

$$\int_0^1 x^n \, \mathrm{d}\mathsf{G}_1(x) = \int_0^1 x^n \, \mathrm{d}\mathsf{G}_2(x) \qquad (n = 0, 1, 2, \ldots) \tag{6.9}$$

が成り立てば

$$\mathsf{G}_1 = \mathsf{G}_2$$

である.

Proof. Weierstrass の多項式近似定理 (系 **??**) により, (6.9) が成り立てば, 任意の $[0, 1]$ 上の連続関数 f に対して

$$\int_0^1 f(x) \, \mathrm{d}\mathsf{G}_1(x) = \int_0^1 f(x) \, \mathrm{d}\mathsf{G}_2(x) \tag{6.10}$$

が成り立つ.

いま, $0 \leq a < 1$ と $n \geq 1$ に対して

$$f_n(x) = \begin{cases} 1 & (0 \leq x \leq a), \\ 1 - n(x - a) & (a \leq x \leq a + 1/n), \\ 0 & (a + 1/n \leq x \leq 1) \end{cases}$$

と定める. 優収束定理より

$$\begin{aligned} \mathsf{G}_1(a) &= \int_0^a \mathrm{d}\mathsf{G}_1(x) = \int_0^1 \mathbb{1}_{[0,a]}(x) \, \mathrm{d}\mathsf{G}_1(x) = \int_0^1 \lim_{n\to\infty} f_n(x) \, \mathrm{d}\mathsf{G}_1(x) \\ &= \lim_{n\to\infty} \int_0^1 f_n(x) \, \mathrm{d}\mathsf{G}_1(x) \overset{\because (6.10)}{=} \lim_{n\to\infty} \int_0^1 f_n(x) \, \mathrm{d}\mathsf{G}_2(x) \\ &= \int_0^1 \lim_{n\to\infty} f_n(x) \, \mathrm{d}\mathsf{G}_2(x) = \int_0^1 \mathbb{1}_{[0,a]}(x) \, \mathrm{d}\mathsf{G}_2(x) = \int_0^a \mathrm{d}\mathsf{G}_2(x) \\ &= \mathsf{G}_2(a) \end{aligned}$$

を得る. また $\mathsf{G}_1(1) = \mathsf{G}_2(1)$ だから, すべての $a \in [0, 1]$ に対して $\mathsf{G}_1(a) = \mathsf{G}_2(a)$ がわかる. □

定理 6.31. $[0,\,\infty)$ に集中した確率をもつ 2 つの分布関数 $\mathsf{F}_1,\,\mathsf{F}_2$ に対して, それぞれの Laplace 変換を $f_1,\,f_2$ とする. このとき, f_1 と f_2 が一致すれば, $\mathsf{F}_1 = \mathsf{F}_2$ である.

Proof. $\mathsf{F} : [0,\,\infty) \longrightarrow \mathbb{R}$ を確率分布関数とする. $x = e^{-t}$, $\mathsf{G}(x) = \mathsf{F}(t)$ とすると

$$\int_0^\infty e^{-\lambda t}\,\mathrm{d}\mathsf{F}(t) = \int_0^1 x^{\lambda+1}\,\mathrm{d}\mathsf{G}(x) \qquad (\lambda \geq 0)$$

となる. よって $f_1 = f_2$ ならば,

$$\int_0^1 x^n\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 x^n\,\mathrm{d}\mathsf{G}_2(x) \qquad (n = 0,\,1,\,2,\ldots)$$

となる. あとは補題 6.30 を用いれば, 定理は証明される. □

6.7 章末注釈と参考文献

節 6.1 は $[1, \mathrm{pp}.69-71]$ を借用した. 補題 6.11 は $[17, \mathrm{pp}.28-30]$ を借用した.

6.8 演習問題

演習問題 6.1. $n \in \mathbb{N}$, $0 < \theta < 1$ とし,

$$\mathsf{p}_\theta(x) = \mathsf{Pr}(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x} & (x = 0,\,1,\,\ldots,\,n) \\ 0 & (\text{その他の場合}) \end{cases},$$

$$\begin{pmatrix} n \\ x \end{pmatrix} = \frac{n!}{x!\,(n-x)!}, \qquad 0! = 1$$

とする. 母数モデル

$$\mathcal{P} = \left\{\mathsf{p}_\theta;\, 0 < \theta < 1\right\}$$

を考える.
(1) $X \sim \mathsf{Bino}(n,\,\theta)$ のとき

$$\mathsf{m}^X(t) = \mathsf{E}\left[e^{tX}\right] = \left\{\theta e^t + 1 - \theta\right\}^n \qquad (t \in \mathbb{R})$$

を示せ.

(2) $\dfrac{\mathrm{d}}{\mathrm{d}t}\mathsf{m}^X(t)\Big|_{t=0}$ を計算することで $\mathsf{E}[X]$ を求めよ.
(3) $\dfrac{\mathrm{d}^2}{\mathrm{d}t^t}\mathsf{m}^X(t)\Big|_{t=0}$ を計算することで $\mathsf{E}[X^2]$ を求めよ.
(4) $\mathsf{Var}[X]$ を計算せよ.
(5) $X \sim \mathsf{p}_\theta$ とする. $0 < \forall\theta < 1$ に対して, $\mathsf{E}_\theta\big[\dot{\ell}_1\ (X|\,\theta)\big]$ を求めよ. ただし, $X = x$ を観測したとき, $\ell_1(\theta|\,x) = \log \mathsf{p}_\theta(x)$ とし, $\dot{\ell}_1\ (\theta|\,x) = \dfrac{\mathrm{d}}{\mathrm{d}\theta}\ell_1(\theta|\,x)$ と書いた.
(6) $\mathsf{E}_\theta\Big[\big\{\dot{\ell}_1\ (\theta|\,X)\big\}^2\Big]$ を計算せよ.
(7) $-\mathsf{E}_\theta\Big[\ddot{\ell}_1\ (\theta|\,X)\Big]$ を計算せよ. ただし, $\ddot{\ell}_1\ (\theta|\,x) = \dfrac{\mathrm{d}^2}{\mathrm{d}\theta^2}\ell_1(\theta|\,x)$ A と書いた.

演習問題 6.2. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Po}(\theta)\,(\theta > 0)$ とする. $S := X_1 + X_2 + \cdots + X_n$ は $\mathcal{P} = \{\mathsf{Po}(\theta);\, \theta > 0\}$ の十分統計量であることを示せ.

演習問題 6.3.

第7章　推定

この章では, 点推定法における推定量の導出原理と推定量の精度評価についての考え方の説明を行う. 節 7.1 では, モーメント法について説明する. 節 7.2 では, 最尤法について説明する. 最尤推定量の漸近分布は, 統計的モデルが指数型分布族の場合について示している. 節 7.3 では, 推定量のひとつの性質である不偏推定の概念を説明し, それに関係する情報不等式を与える.

7.1　モーメント法

この方法で得られた推定量は最適性を持たないことがある. しかし求めるのが容易であるので, 陽には式が与えられない別の推定量の値を繰り返し計算で求めるときの初期値として用いることができる.

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $d, n \in \mathbb{N}$ とし, 統計的実験を $\big(\mathbb{X}^n, \{\mathsf{P}_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}\big)$ を考え

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*} \quad (\boldsymbol{\theta}^* \in \Theta)$$

とする. ただし, $\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ と $\mathsf{P}_{\boldsymbol{\theta}}$ は, それぞれ測度空間 $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\big)$ と $\big(\mathbb{X}, \mathcal{B}(\mathbb{X})\big)$ 上の分布であり, 母数モデル $\mathcal{P}$ は正則であると仮定する.

$n \geq d$ とする. $1 \leq j \leq d$ に対して

$$\alpha_j := \alpha_j(\boldsymbol{\theta}) := \mathsf{E}\big[X_1^j\big], \quad \widehat{\alpha}_j := \frac{1}{n}\sum_{\ell=1}^{n} X_\ell^j$$

と定める.

定義 7.1. $n \geq d$ とし $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*}\ \big(\boldsymbol{\theta}^* \in \Theta\big)$ とする. このとき $\boldsymbol{\theta}^*$ の <u>モーメント法推定量</u> $\widehat{\boldsymbol{\theta}}_n$ とは次をみたす解 (存在すれば) である.

$$\alpha_1(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_1,\ \alpha_2(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_2, \ldots,\ \alpha_d(\widehat{\boldsymbol{\theta}}_n) = \widehat{\alpha}_d \tag{7.1}$$

注意 7.2. $n \geq 2$ とする. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}\big(\mu^*, \big(\sigma^*\big)^2\big)$ とする. ただし $\boldsymbol{\theta}^* := (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty) =: \Theta$ で (μ^*, σ^*) は未知とする. このとき

$$\alpha_1 = \mathsf{E}\big[X_1\big] = \mu^*, \quad \alpha_2 = \mathsf{E}\big[X_1^2\big] = \mathsf{Var}\big[X_1\big] + \big\{\mathsf{E}\big[X_1\big]\big\}^2 = (\sigma^*)^2 + (\mu^*)^2$$

である. したがって $\widehat{\boldsymbol{\theta}}_n = (\widehat{\mu}_n, \widehat{\sigma}_n)$ とすれば

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma}_n^2 + \{\widehat{\mu}_n\}^2 = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell^2$$

を解くと (μ^*, σ^*) のモーメント法推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{\ell=1}^{n} X_\ell, \quad \widehat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{\ell=1}^{n}(X_\ell - \widehat{\mu}_n)^2}$$

となる. □

定理 7.3. $d, n \in \mathbb{N}$ は $n \geq d$ とする.

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_{\boldsymbol{\theta}^*}\ (\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d), \quad \mathsf{E}\big[\big|X_1\big|^{2d}\big] < \infty$$

とする. (7.1) で定義した $\boldsymbol{\alpha}(\boldsymbol{\theta}) := \big(\alpha_1(\boldsymbol{\theta}), \ldots, \alpha_d(\boldsymbol{\theta})\big)^\top$ は, $\boldsymbol{\theta}^*$ の近傍で局所逆写像 $\boldsymbol{g} := \boldsymbol{\alpha}^{-1}$ をもち, $\boldsymbol{g}$ は $\boldsymbol{\alpha}^{-1}(\boldsymbol{\theta}^*)$ の近傍で全微分可能とする. このとき適当な条件のもとで以下が成立する.

(1) $\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\ \widehat{\boldsymbol{\theta}}_n$ は存在 $\big) \overset{n\to\infty}{\longrightarrow} 1$ である.

(2) $\forall \epsilon > 0$ に対して

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\big(\big|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big|_{2,\,d} \geq \epsilon\big) \overset{n\to\infty}{\longrightarrow} 0$$

である. ただし $\big|\ \cdot\ \big|_{2,\,d}$ は $\mathbb{R}^d$ の Euclid ノルムである.

(3) $\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0},\, \boldsymbol{\Sigma}\big)$ である. ただし

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{G}\big(\boldsymbol{\theta}^*\big)\mathsf{Cov}_{\boldsymbol{\theta}^*}\big[\boldsymbol{Y}\big]\boldsymbol{G}^\top\big(\boldsymbol{\theta}^*\big), \\
\boldsymbol{G}\big(\boldsymbol{\theta}\big) &= \big(\dot{\boldsymbol{g}}_1\ (\boldsymbol{\theta}),\, \dot{\boldsymbol{g}}_2\ (\boldsymbol{\theta}),\, \ldots,\, \dot{\boldsymbol{g}}_d\ (\boldsymbol{\theta})\big), \\
\dot{\boldsymbol{g}}_j\ (\boldsymbol{\theta}) &= \left(\frac{\partial \alpha_j^{-1}}{\partial \theta_1}(\boldsymbol{\theta}),\, \frac{\partial \alpha_j^{-1}}{\partial \theta_2}(\boldsymbol{\theta}),\, \ldots,\, \frac{\partial \alpha_j^{-1}}{\partial \theta_d}(\boldsymbol{\theta})\right)^\top (j = 1, 2, \ldots, d), \\
\boldsymbol{Y} &= \big(X_1, X_1^2, \ldots, X_1^d\big)^\top
\end{aligned}$$

が成立する.

Proof. 証明は加筆予定である. □

7.2 最尤法

$\mathbb{X}$ を $\mathbb{R}$ の空でない部分集合とし, $d, n \in \mathbb{N}$ とする. 統計的実験 $\big(\mathbb{X}^n, \{\mathsf{p}_{\boldsymbol{\theta}}(x);\, \boldsymbol{\theta} \in \Theta\}\big)$ を考える. ただし $\mathsf{p}(x|\,\boldsymbol{\theta})$ は p.d.f. (または p.m.f.) で $\Theta \subset \mathbb{R}^d$ である. $\boldsymbol{\theta}^* \in \Theta$ とし

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(x)$$

とする.

定義 7.4. $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき $\boldsymbol{\theta}$ の 尤(ゆう) 度関数 $\mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を

$$\mathsf{lik}_n\big(\boldsymbol{\theta}|\,\boldsymbol{x}\big) = \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j)$$

で定義し, 対数尤度 $\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を

$$\ell_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$$

で定義する. ただし, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である.

注意 7.5. 尤度関数は

$$\mathsf{lik}_n(\,\cdot\,|\,\boldsymbol{x}) : \Theta \ni \boldsymbol{\theta} \mapsto \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x}) \in [0, \infty)$$

である. □

記法: 関数 $g(x)$ の最大値を取る点を表す集合を

$$\arg\max_{x \in \mathbb{R}} g(x)$$

と書く. たとえば $g(x) = -(x-1)^2$ のとき

$$\arg\max_{x \in \mathbb{R}} g(x) = \{1\}$$

となる. $g(x) = \sin x$ のとき

$$\arg\max_{0 \le x \le 4\pi} g(x) = \{\pi/2, 5\pi/2\}$$

となる.

定義 7.6. $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したとき $\boldsymbol{\theta}^*$ の 最尤推定値(maximum likelihood estimate) を $\mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$ を最大にする値 $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ で定義する. すなわち

$$\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x}) \in \arg\max_{\boldsymbol{\theta} \in \Theta} \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$$

である. $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ に $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を代入したもの $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ を $\boldsymbol{\theta}^*$ の 最尤推定量(maximum likelihood estimator=m.l.e.) という.

例 7.7. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta^*)$ とする. ただし $(0, 1) =: \Theta \ni \theta^*$ は未知とする. すなわち

$$\mathsf{p}(x|\,\theta^*) = \begin{cases} (\theta^*)^x\big(1-\theta^*\big)^{1-x} & (x = 0, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. $X_j = x_j\,(j = 1, 2, \ldots, n)$ を観測したとき

$$\mathsf{lik}_n\big(\theta|\,\boldsymbol{x}\big) = \prod_{j=1}^{n} \mathsf{p}(x_j|\,\theta) = \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j} = \theta^{t_n}(1-\theta)^{n-t_n}$$

となる. ただし, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ と $t_n = \sum_{j=1}^{n} x_j$ である. よって対数尤度は

$$\ell_n(\theta|\,\boldsymbol{x}) = t_n \log\theta + (n - t_n)\log(1-\theta) \qquad (0 < \theta < 1)$$

となる. このことから, $0 < t_n < n$ のとき

$$\frac{t_n}{n} \in \arg\max_{\theta\in(0,\,1)} \ell_n(\theta|\,\boldsymbol{x})$$

がわかる. したがって, $0 < t_n < n$ のとき θ^* の最尤推定値は $\widehat{\theta}_n = \dfrac{\sum_{j=1}^{n} x_j}{n}$ となり, $t_n = 0$ または $t_n = n$ のとき, 最尤推定値は存在しない. □

注意 7.8. 定義 7.6 は [3] の流儀に従った. 一方, 母数空間 Θ の閉包 $\mathrm{cl}(\Theta)$ を考え, $\mathrm{cl}(\Theta)$ に尤度関数 lik_n の定義域を拡張して

$$\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x}) \in \arg\max_{\boldsymbol{\theta}\in\mathrm{cl}(\Theta)} \mathsf{lik}_n(\boldsymbol{\theta}|\,\boldsymbol{x})$$

と最尤推定値を定義する流儀もある. [29, 52] を参照のこと. 例 7.7 を踏まえると, こちらの定義の方が数学的に扱いやすいようにみえる.

例 7.9. $n \geq 2\,(n \in \mathbb{N})$ とする. $\boldsymbol{\theta}^* = (\mu^*, \sigma^*) \in \mathbb{R} \times (0, \infty)$ とし

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}\big(\mu^*, (\sigma^*)^2\big)$$

とする. $X_j = x_j\,(j = 1, 2, \ldots, n)$ を観測したとき, 尤度関数は

$$\begin{aligned} \mathsf{lik}_n(\mu, \sigma|\,\boldsymbol{x}) &= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_j-\mu)^2}{2\sigma^2}\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left\{-\frac{ns_n^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\overline{x}_n-\mu)^2}{2\sigma^2}\right\} \end{aligned} \tag{7.2}$$

となる. ただし

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad s_n^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x}_n)^2$$

である. さらに, $\displaystyle\sum_{j=1}^{n}(x_j - \overline{x}_n)^2 > 0$ を仮定する. なお, (7.2) の最後の等号は

$$\sum_{j=1}^{n}(x_j - \mu)^2 = ns_n^2 + n(\overline{x}_n - \mu)^2 \tag{7.3}$$

からわかる. 対数尤度は

$$\ell_n(\mu,\, \sigma|\, \boldsymbol{x}) = -n\log\sigma - \frac{ns_n^2}{2\sigma^2} - \frac{n(\overline{x}_n - \mu)^2}{2\sigma^2} + (\text{定数項})$$

となる. よって

$$\begin{cases} \dfrac{\partial \ell_n}{\partial \mu}(\mu,\, \sigma|\, \boldsymbol{x}) = \dfrac{n(\overline{x}_n - \mu)}{\sigma^2} = 0 \\ \dfrac{\partial \ell_n}{\partial \sigma}(\mu,\, \sigma|\, \boldsymbol{x}) = -\dfrac{n}{\sigma} + \dfrac{ns_n^2}{\sigma^3} + \dfrac{n(\overline{x}_n - \mu)^2}{\sigma^3} = 0 \end{cases}$$

を解くと

$$\mu = \overline{x}_n, \qquad \sigma = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\bigl(x_j - \overline{x}_n\bigr)^2}$$

となる. $\ell_n(\mu,\, \sigma|\, \boldsymbol{x})$ の Hessian を求める.

$$\boldsymbol{H} := \begin{pmatrix} \dfrac{\partial \ell_n^2}{\partial \mu^2}\bigl(\overline{x}_n,\, s_n|\, \boldsymbol{x}\bigr) & \dfrac{\partial \ell_n^2}{\partial \mu \partial \sigma}\bigl(\overline{x}_n,\, s_n|\, \boldsymbol{x}\bigr) \\ \dfrac{\partial \ell_n^2}{\partial \sigma \partial \mu}\bigl(\overline{x}_n,\, s_n|\, \boldsymbol{x}\bigr) & \dfrac{\partial \ell_n^2}{\partial \sigma^2}\bigl(\overline{x}_n,\, s_n|\, \boldsymbol{x}\bigr) \end{pmatrix} = \begin{pmatrix} -\dfrac{n}{s_n^2} & 0 \\ 0 & -\dfrac{2n}{s_n^2} \end{pmatrix} \tag{7.4}$$

より, $-\boldsymbol{H}$ は正定値行列となる. したがって, 関数

$$\Theta \ni (\mu,\, \sigma) \mapsto \ell_n(\mu,\, \sigma|\, \boldsymbol{x})$$

は $(\mu,\, \sigma) = (\overline{x}_n,\, s_n)$ で最大となる. 以上の議論から $(\mu^*,\, \sigma^*)$ の最尤推定量は

$$\widehat{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} X_j =: \overline{X}_n, \quad \widehat{s}_n = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\bigl(X_j - \overline{X}_n\bigr)^2}$$

となる. □

注意 7.10. $g : \mathbb{R} \times [0, \infty) \longrightarrow [0, \infty)$ は 1 対 1 とする. すると $\widehat{\boldsymbol{\theta}}_n$ が $\boldsymbol{\theta}^*$ の最尤推定量ならば, $g\big(\widehat{\boldsymbol{\theta}}_n\big)$ は $g(\boldsymbol{\theta}^*)$ の最尤推定量であることがわかる. すなわち, 最尤推定量は母数の変換に関して不変である. このことに注意すれば, 例 7.9 の結果から $(\widehat{\mu}_n,\, s_n^2)$ は $(\mu^*,\, (\sigma^*)^2)$ の最尤推定量になることがわかる. □

問 7.1. (7.3) と (7.4) を確認せよ.

定理 7.11.

$$X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_\eta(x) = h(x) \exp\{\eta^* T(x) - \mathsf{K}(\eta^*)\}\, (\eta^* \in \mathcal{E}^o)$$

とする. ただし $\mathcal{E}^o$ は自然母数空間 $\mathcal{E} \subset \mathbb{R}$ の内部である. さらに $\ddot{\mathsf{K}}\,(\eta^*) > 0\, (\eta^* \in \mathcal{E}^o)$ を仮定する. このとき η^* の最尤推定量 $\widehat{\eta}_n$ は十分大きな n に対して確率 1 で一意的に存在 [1] して

$$\widehat{\eta}_n \xrightarrow{\text{a.s.}} \eta^*,$$
$$\sqrt{n}\big(\widehat{\eta}_n - \eta^*\big) \rightsquigarrow \mathsf{N}\left(0,\, \frac{1}{\ddot{\mathsf{K}}\,(\eta^*)}\right)$$

が成立する.

Proof. $X_1 = x_1,\, X_2,\, \ldots,\, X_n = x_n$ を観測したときの尤度関数 lik_n は

$$\mathsf{lik}_n(\eta|\, \boldsymbol{x}) = \left\{\prod_{j=1}^n h(x_j)\right\} \exp\big[\eta \sum_{j=1}^n T(x_j) - n\mathsf{K}(\eta)\big]$$

となる. このことから対数尤度関数は

$$\ell_n(\eta|\, \boldsymbol{x}) = \log\big(\mathsf{lik}_n(\eta|\, \boldsymbol{x})\big) = n\big\{\eta \overline{T}_n - \mathsf{K}(\eta)\big\} + (\text{constant})$$

となる. ただし $\overline{T}_n = n^{-1} \sum_{j=1}^n T(x_j)$ である. したがって

$$\dot{\ell}_n\,(\eta|\, \boldsymbol{x}) = \frac{d\ell_n}{d\eta}(\eta|\, \boldsymbol{x}) = n\big(\overline{T}_n - \dot{\mathsf{K}}\big) = 0 \;\Leftrightarrow\; \overline{T}_n = \dot{\mathsf{K}}\,(\eta) = \mathsf{E}[T(X_1)]$$

となる. 最後の等号は (6.4) からわかる.

[1] これは不明瞭な表現である. 意味するところは, $A_n = \{\widehat{\eta}_n$ は存在 $\}$ としたとき

$$\lim_{n\to\infty} \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = 1$$

が成立することである. これは「確率 1 で、ある時点以降のすべての n について存在する」を表す.

以後, 対数尤度関数の $x_j\,(j=1,2,\ldots,n)$ のところに X_j を代入して, 確率変数にしたものを考える. $T(X_1)$ は有限の期待値をもつので, 大数の法則 (定理 4.17) から

$$\overline{T}_n=\frac{1}{n}\sum_{j=1}^{n}T(X_j)\xrightarrow{\text{a.s.}}\mathsf{E}[T(X_1)]=\overset{\bullet}{\mathsf{K}}(\eta^*)\quad(n\to\infty)\tag{7.5}$$

がわかる. η^* の適当な近傍を $U(\eta^*)$ と書き, $\overset{\bullet}{\mathsf{K}}(U(\eta^*)):=\{\overset{\bullet}{\mathsf{K}}(\eta);\,\eta\in U(\eta^*)\}$ とおく. (7.5) から

$$\mathsf{Pr}\Big(\lim_{n\to\infty}\overline{T}_n\in\overset{\bullet}{\mathsf{K}}(U(\eta^*))\Big)=1$$

となる. さらに, $\overset{\bullet\bullet}{\mathsf{K}}(\eta^*)>0$ なので, $\overset{\bullet\bullet}{\mathsf{K}}(\eta)>0\,(\eta\in U(\eta^*))$ となる. したがって, 関数 $\mathcal{E}^o\ni\eta\mapsto\overset{\bullet}{\mathsf{K}}(\eta)$ は近傍 $U(\eta^*)$ 上で狭義単調増加関数である. このことより

$$\overline{T}_n=\overset{\bullet}{\mathsf{K}}(\eta)$$

をみたす $\eta\in U(\eta^*)\subset\mathcal{E}^o$ が存在する. したがって, 陰関数の定理から, 近傍 $\mathsf{K}(U(\eta^*))$ 上で $\overset{\bullet}{\mathsf{K}}$ の逆写像 $\overset{\bullet}{\mathsf{K}}^{-1}$ が存在する. このことを踏まえて

$$\widehat{\eta}_n=\overset{\bullet}{\mathsf{K}}^{-1}\big(\overline{T}_n\big)$$

とおく. さらに

$$\overline{T}_n\xrightarrow{\text{a.s.}}\overset{\bullet}{\mathsf{K}}(\eta^*)\quad\text{かつ}\quad\overset{\bullet}{\mathsf{K}}^{-1}(t)\text{ は近傍 }\overset{\bullet}{\mathsf{K}}\big(U(\eta^*)\big)\text{ 上で連続}$$

なので定理 4.14(6) から

$$\widehat{\eta}_n=\overset{\bullet}{\mathsf{K}}^{-1}\big(\overline{T}_n\big)\xrightarrow{\text{a.s.}}\overset{\bullet}{\mathsf{K}}^{-1}\big(\overset{\bullet}{\mathsf{K}}(\eta^*)\big)=\eta^*$$

がわかる. 一方, (6.4) から $\mathsf{E}[T(X_1)]=\overset{\bullet}{\mathsf{K}}(\eta^*)$, $\mathsf{Var}[T(X_1)]=\overset{\bullet}{\mathsf{K}}(\eta^*)$ となることに注意して, 中心極限定理 (定理 4.22) を用いると

$$\sqrt{n}\big(\overline{T}_n-\overset{\bullet}{\mathsf{K}}(\eta^*)\big)\rightsquigarrow\mathsf{N}\big(0,\,\overset{\bullet\bullet}{\mathsf{K}}(\eta^*)\big)$$

がわかる. 以上のことを踏まえて, $\widehat{\eta}_n=\overset{\bullet}{\mathsf{K}}^{-1}\big(\overline{T}_n\big)$ に対して, デルタ法 (定理 4.25) を適用すると

$$\sqrt{n}\big(\widehat{\eta}_n-\eta^*\big)\rightsquigarrow\mathsf{N}\left(0,\,\frac{1}{\overset{\bullet\bullet}{\mathsf{K}}(\eta^*)}\right)$$

がわかる. □

7.3　不偏推定と情報不等式

$\mathbb{X} \subset \mathbb{R}$ を空でない部分集合とする. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とし, $X_1, X_2, \ldots, X_n$ をこの空間上の独立同一分布に従う $\mathbb{X}$ 値確率変数列 (ランダム標本) とする. また $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ と記し, $\boldsymbol{X}$ の値域 (標本空間) を $\mathbb{X}^n \subset \mathbb{R}^n$ と表記する. さらに

$$\mathsf{P} = \mathsf{Pr} \circ X_1^{-1}, \quad \mathsf{P}^{\otimes n} = \mathsf{Pr} \circ \boldsymbol{X}^{-1}$$

とおく. すると $\mathsf{P}^{\otimes n} = \underbrace{\mathsf{P} \times \mathsf{P} \times \cdots \times \mathsf{P}}_{n \text{ 個}}$ となっている.

$\mathcal{P} = \{\mathsf{P}_\theta; \theta \in \Theta \subset \mathbb{R}\}$ を $\mathbb{X}$ 上の正則母数モデルとする. 以下では, $\Theta \subset \mathbb{R}$ とし, P_θ に関する期待値と分散を $\mathsf{E}_\theta[\,\cdot\,]$, $\mathsf{Var}_\theta[\,\cdot\,]$ と表記する.

定義 7.12. $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする. 統計量 $T(\boldsymbol{X})$ が $\forall \theta \in \Theta$ に対して

$$\mathsf{E}_\theta\big[T(\boldsymbol{X})\big] = \theta$$

をみたすとき $T(\boldsymbol{X})$ は θ の <u>不偏推定量</u>(unbiased estimator) という.

例 7.13. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, 1)\,(\theta \in \mathbb{R})$ とする. このとき $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^n X_j$ は θ の不偏推定量である. □

定理 7.14 (Rao-Blackwell の定理)**.** $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ で $T = T(\boldsymbol{X})$ は θ の十分統計量とする. $\widehat{\theta}_n(\boldsymbol{X})$ は θ の任意の不偏推定量で, 有限の分散 $\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]$ をもつものとする. 推定量 $\widetilde{\theta}_n(T)$ を

$$\widetilde{\theta}_n(T) := \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X}) \big| T\big]$$

により定めたとき, 次の (1), (2) が成立する.

(1) $\widetilde{\theta}_n(T)$ は θ の不偏推定量である.

(2) $\mathsf{Var}_\theta\big[\widetilde{\theta}_n(T)\big] \leq \mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]$ $(\forall \theta \in \Theta)$ が成立する.

Proof. (1) T は十分統計量なので, $\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X}) \big| T\big]$ は θ に依存しないので $\widetilde{\theta}_n(T)$ は推定量となる. また定理 2.34(2) より

$$\mathsf{E}_\theta\big[\widetilde{\theta}_n(T)\big] = \mathsf{E}_\theta\Big[\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X}) \big| T\big]\Big] = \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \theta \qquad (\theta \in \Theta)$$

となり $\widetilde{\theta}_n(T)$ は θ の不偏推定量であることが示せた.
(2) 定理 2.34(4) より

$$\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\widetilde{\theta}_n(T)\big|T\big] = \widetilde{\theta}_n(T)\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)|T\big]$$

$$
\begin{aligned}
&= \widetilde{\theta}_n(T)\Big(\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})|T\big] - \widetilde{\theta}_n(T)\Big) \\
&= \widetilde{\theta}_n(T)\big(\widetilde{\theta}_n(T) - \widetilde{\theta}_n(T)\big) = 0
\end{aligned}
$$

となり

$$
\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\widetilde{\theta}_n(T)\big] = \mathsf{E}_\theta\big[\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\widetilde{\theta}_n(T)\big|T\big]\big] = 0 \tag{7.6}
$$

を得る. 一方

$$
\begin{aligned}
\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}_n(\boldsymbol{X}) - \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\big\}^2\big] \\
&= \mathsf{E}_\theta\big[\big\{\widehat{\theta}_n(\boldsymbol{X}) - \theta\big\}^2\big] \\
&= \mathsf{E}_\theta\big[\big\{\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big) + \big(\widetilde{\theta}_n(T) - \theta\big)\big\}^2\big] \\
&= \mathsf{E}_\theta\big[\big\{\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big\}^2\big] + \mathsf{E}_\theta\big[\big\{\widetilde{\theta}_n(T) - \theta\big\}^2\big] \\
&\quad + 2\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\big(\widetilde{\theta}_n(T) - \theta\big)\big]
\end{aligned}
$$

である. しかし, (7.6) および $\widehat{\theta}_n(\boldsymbol{X})$ と $\widetilde{\theta}_n(T)$ は θ の不偏推定量であることに注意すると

$$
\begin{aligned}
&\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\big(\widetilde{\theta}_n(T) - \theta\big)\big] \\
&\quad = \underbrace{\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big)\widetilde{\theta}_n(T)\big]}_{=0 \quad \because (7.6)} - \underbrace{\theta\big\{\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] - \mathsf{E}_\theta\big[\widetilde{\theta}_n(T)\big]\big\}}_{=0} \\
&\quad = 0
\end{aligned}
$$

がわかる. 以上から

$$
\begin{aligned}
\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] &= \mathsf{E}_\theta\big[\big\{\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{\theta}_n(T)\big\}^2\big] + \mathsf{Var}_\theta\big[\widetilde{\theta}_n(T)\big] \\
&\geq \mathsf{Var}_\theta\big[\widetilde{\theta}_n(T)\big] \qquad (\theta \in \Theta)
\end{aligned}
$$

を得る. □

定義 7.15. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}$ $(\theta \in \Theta \subset \mathbb{R})$ とする. θ の任意の不偏推定量 $\widetilde{\theta}_n(\boldsymbol{X})$ に対して

$$
\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \leq \mathsf{Var}_\theta\big[\widetilde{\theta}_n(\boldsymbol{X})\big] \qquad (\theta \in \Theta)
$$

をみたす θ の不偏推定量 $\widehat{\theta}_n(\boldsymbol{X})$ が存在するとき, $\widehat{\theta}_n(\boldsymbol{X})$ を θ の **一様最小分散不偏推定量**(uniformly minimum variance unbiased estimator: UMVUE) という.

注意 7.16. 存在すれば, 一意的である. 演習問題 7.7 を参照のこと.

可測関数 $g : \mathbb{R} \to \mathbb{R}$ は連続微分可能とする. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}$ に基づき $g(\theta)$ の推定問題を考える.

定理 7.17. $\{\mathsf{P}_\theta;\, \theta \in \Theta \subset R\}$ は正則母数モデルとする. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\, (\theta \in \Theta \subset \mathbb{R})$ とし $\boldsymbol{X}$ は同時 p.d.f. または p.m.f. $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ をもつとする. $\widehat{\theta}_n = \widehat{\theta}_n(\boldsymbol{X})$ を $g(\theta)$ の任意の不偏推定量とし

$$A(\widetilde{\theta},\, \theta) := \mathsf{Var}_\theta\left[\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})}\right] \qquad (\widetilde{\theta},\, \theta \in \Theta) \tag{7.7}$$
$$A(\widetilde{\theta},\, \theta) > 0 \qquad (\widetilde{\theta} \in \Theta,\, \widetilde{\theta} \neq \theta)$$

とおく. このとき

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \geq \sup_{\widetilde{\theta} \in \Theta} \frac{\{g(\widetilde{\theta}) - g(\theta)\}^2}{A(\widetilde{\theta},\, \theta)}$$

が成り立つ.

Proof. $\boldsymbol{X}$ は連続型確率変数の場合の証明を与える. $\boldsymbol{X}$ の値域を $\mathbb{X}^n$ と書く. さらに, $\widetilde{\mathbb{X}}^n = \{\boldsymbol{x} \in \mathbb{X}^n;\, \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) > 0\}$ とおく. ここで, 正則母数モデルの仮定から, $\widetilde{\mathbb{X}}^n$ は $\theta \in \Theta$ に依存しないことに注意する. また, 離散型の場合は, 積分記号を和の記号に変更すればよい. 推定量 $\widehat{\theta}_n$ は $g(\theta)$ に対して不偏なので

$$\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x}) \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = g(\theta) \qquad (\theta \in \Theta)$$

となる. よって

$$\begin{aligned}
\mathsf{E}_\theta\left[\widehat{\theta}_n(\boldsymbol{X})\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right] &= \int_{\widetilde{\mathbb{X}}^n} \widehat{\theta}_n(\boldsymbol{x})\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})} - 1\right\}\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x}) \mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} - \int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x}) \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\
&= g(\widetilde{\theta}) - g(\theta)
\end{aligned} \tag{7.8}$$

となる. 一方

$$\mathsf{E}_\theta\left[\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})}\right] = \int_{\widetilde{\mathbb{X}}^n} \frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})} \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{X}^n} \mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = 1 \tag{7.9}$$

と (7.8) に注意すると

$$\begin{aligned}
g(\widetilde{\theta}) - g(\theta) &= \mathsf{E}_\theta\left[\widehat{\theta}_n(\boldsymbol{X})\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right] \\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}_n(\boldsymbol{X}) - \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] + \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}_n(\boldsymbol{X}) - \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right] \\
&\quad + \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\underbrace{\mathsf{E}_\theta\left[\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right]}_{=0} \\
&= \mathsf{E}_\theta\left[\left\{\widehat{\theta}_n(\boldsymbol{X}) - \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\right\}\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right]
\end{aligned}
$$

を得る. Cauchy-Schwarz の不等式 (定理 3.10) を上の式の最右辺に適用すると

$$
\begin{aligned}
\big|g(\widetilde{\theta}) - g(\theta)\big| &= \left|\mathsf{E}_\theta\left[\widehat{\theta}_n(\boldsymbol{X})\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}\right]\right| \\
&\leq \sqrt{\left\{\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X}) - \mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\big\}^2\right]}\sqrt{\mathsf{E}_\theta\left[\left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} - 1\right\}^2\right]} \\
&= \sqrt{\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]}\sqrt{\mathsf{Var}_\theta\left[\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})}\right]} \qquad (7.10)
\end{aligned}
$$

を得る. 最後の等号は (7.9) を用いた. したがって

$$
\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big]\mathsf{Var}_\theta\left[\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})}\right] \geq \{g(\widetilde{\theta}) - g(\theta)\}^2 \qquad (7.11)
$$

を得る. したがって (7.7) と (7.11) から $\forall\theta \in \Theta$ に対して

$$
\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \geq \frac{\{g(\widetilde{\theta}) - g(\theta)\}^2}{A\big(\widetilde{\theta},\, \theta\big)}
$$

となる. 上式の $\widetilde{\theta}$ は任意だったので, 上式の左辺において $\widetilde{\theta} \in \Theta \setminus \{\theta\}$ に関する上限を取ると定理の主張はわかる. □

定理 7.18. (Cramér-Rao の不等式) 次の条件を仮定する.

(1) $\{\mathsf{P}_\theta;\, \theta \in \Theta \subset R\}$ は正則母数モデルとする.

(2) $g : \mathbb{R} \longrightarrow \mathbb{R}$ は連続微分可能でその導関数は $\overset{\bullet}{g}(\theta) \neq 0\,(\theta \in \Theta)$ をみたす. さらに

$$
\lim_{\widetilde{\theta}\to\theta} \frac{A\big(\widetilde{\theta},\, \theta\big)}{(\widetilde{\theta} - \theta)^2} = J(\theta) > 0
$$

が存在する. ただし, $A\big(\widetilde{\theta},\, \theta\big)$ は (7.7) で与えたものである.

(3) $\forall\theta \in \Theta$ に対して十分小さな $\epsilon > 0$ をとると任意の $\widetilde{\theta} \in \{\psi \in \Theta;\ |\psi - \theta| < \epsilon\}$ に対して, ある関数 $G_\theta(\boldsymbol{x})$ が存在して

$$\left|\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x}) - \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})}{(\widetilde{\theta}-\theta)\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})}\right| < G_\theta(\boldsymbol{x}) \quad \text{かつ} \quad \mathsf{E}_\theta\big[G_\theta^2(\boldsymbol{X})\big] < \infty$$

をみたす.

このとき $g(\theta)$ の任意の不偏推定量 $\widehat{\theta}_n(\boldsymbol{X})$ に対して

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \qquad (\theta \in \Theta) \tag{7.12}$$

が成り立つ. ただし

$$\mathcal{F}_{\boldsymbol{X}}(\theta) = \mathsf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})\right\}^2\right]$$

である.

注意 7.19. (7.12) を Cramér-Rao の不等式といい, その右辺を Cramér-Rao の下限という. □

定理 7.18 の証明: $\boldsymbol{X}$ の値域を $\mathbb{X}^n$ とし, $\widetilde{\mathbb{X}}^n = \{\boldsymbol{x} \in \mathbb{X}^n;\ \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) > 0\}$ とおく. 定理 7.17 と仮定 (2) から

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \geq \lim_{\widetilde{\theta}\to\theta} \frac{\left\{\dfrac{g(\widetilde{\theta}) - g(\theta)}{\widetilde{\theta}-\theta}\right\}^2}{\dfrac{A\big(\widetilde{\theta},\,\theta\big)}{(\widetilde{\theta}-\theta)^2}} = \frac{\{\dot{g}(\theta)\}^2}{J(\theta)} \tag{7.13}$$

を得る. 次に仮定 (3) に注意して Lebesgue の優収束定理を用いると

$$\begin{aligned}
J(\theta) &= \lim_{\widetilde{\theta}\to\theta} \frac{A\big(\widetilde{\theta},\,\theta\big)}{\big(\widetilde{\theta}-\theta\big)^2} \\
&= \lim_{\widetilde{\theta}\to\theta} \int_{\widetilde{\mathbb{X}}^n} \left\{\frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})} - 1\right\}^2 \frac{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})}{\big(\widetilde{\theta}-\theta\big)^2}\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\widetilde{\mathbb{X}}^n} \left\{\lim_{\widetilde{\theta}\to\theta} \frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x}) - \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})}{\big(\widetilde{\theta}-\theta\big)\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})}\right\}^2 \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{\widetilde{\mathbb{X}}^n} \left\{\frac{\partial}{\partial\theta}\log \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\right\}^2 \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \mathcal{F}_{\boldsymbol{X}}(\theta)
\end{aligned}$$

となる. これと (7.13) を合わせると

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] \geq \frac{\{\dot{g}(\theta)\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \tag{7.14}$$

を得る.

□

注意 7.20. 指数型分布族が定理 7.18 の正則条件 (2) と (3) をみたしているかを確認する.

簡単な設定で確認をするために, 以下の記号を導入する. $\mathbb{X} \subset \mathbb{R}$ とし

$$\mathcal{P} = \Big\{ \mathsf{p}_\theta(x) = h(x) \exp\big\{\theta T(x) - \mathsf{K}(\theta)\big\} \Big\} \qquad (\theta \in \Theta \subset \mathbb{R})$$

とする. ただし, $h : \mathbb{X} \longrightarrow \mathbb{R}$, $T : \mathbb{X} \longrightarrow \mathbb{R}$ とし

$$\Theta = \left\{ \theta \in \mathbb{R};\, \mathsf{K}(\theta) = \log\left(\int_{\mathbb{X}} h(x) \exp\big\{\theta T(x)\big\}\, \mathrm{d}x \right) < \infty \right\}$$

である. さらに

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^n \mathsf{p}_\theta(x_i) = \prod_{i=1}^n h(x_i) \exp\left\{ \theta \sum_{i=1}^n T(x_i) - n\mathsf{K}(\theta) \right\}$$
$$(\boldsymbol{x} = (x_1, \ldots, x_n))$$

と記す.

① (2) の確認: $\widetilde{\theta}, \theta \in \Theta$ とし, $\delta := \widetilde{\theta} - \theta$ とおく. すると

$$L := \frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{X})} = \exp\left(\delta \sum_{i=1}^n T(X_i) - n\big\{\mathsf{K}(\theta+\delta) - \mathsf{K}(\theta)\big\} \right)$$

となる. $\boldsymbol{X} \sim \mathsf{p}_\theta^{\boldsymbol{X}}$ なので, $\mathsf{E}_\theta[L] = 1$ となる. したがって

$$\begin{aligned} A(\widetilde{\theta}, \theta) &= \mathsf{Var}_\theta\big[L\big] \\ &= \mathsf{E}_\theta\big[L^2\big] - \big\{\mathsf{E}_\theta\big[L\big]\big\}^2 \\ &= \mathsf{E}_\theta\big[L^2\big] - 1 \end{aligned}$$

である. ここで

$$\begin{aligned} \mathsf{E}_\theta\big[L^2\big] &= \exp\Big(-2n\big\{\mathsf{K}(\theta+\delta) - \mathsf{K}(\theta)\big\} \Big) \mathsf{E}_\theta\left[\exp\left(2\delta \sum_{i=1}^n T(X_i) \right) \right] \\ &= \exp\Big(-2n\big\{\mathsf{K}(\theta+\delta) - \mathsf{K}(\theta)\big\} \Big) \prod_{i=1}^n \mathsf{E}_\theta\left[\exp\Big(2\delta T(X_i) \Big) \right] \end{aligned}$$

となることに注意する. しかし

$$\mathsf{E}_\theta\left[\exp\Big(2\delta T(X_i) \Big) \right]$$

$$
\begin{aligned}
&= \int_{\mathbb{X}} \exp\Big(2\delta T(x)\Big) h(x) \exp\big\{\theta T(x) - \mathsf{K}(\theta)\big\}\,\mathrm{d}x \\
&= \exp\Big(\mathsf{K}(\theta+2\delta) - \mathsf{K}(\theta)\Big) \underbrace{\int_{\mathbb{X}} h(x) \exp\big\{\big(\theta+2\delta\big)T(x) - \mathsf{K}(\theta+2\delta)\big\}\,\mathrm{d}x}_{=1} \\
&= \exp\Big(\mathsf{K}(\theta+2\delta) - \mathsf{K}(\theta)\Big)
\end{aligned}
$$

となる. 以上から

$$
A(\theta+\delta,\,\theta) = \exp\Big(n\big\{\mathsf{K}(\theta+2\delta) - 2\mathsf{K}(\theta+\delta) + \mathsf{K}(\theta)\big\}\Big) - 1
$$

となる. ここで

$$
\mathsf{K}(\theta+2\delta) - 2\mathsf{K}(\theta+\delta) + \mathsf{K}(\theta) = \int_0^\delta \int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t
$$

と書けるので

$$
\begin{aligned}
\lim_{\delta\to 0} \frac{A(\theta+\delta,\,\theta)}{\delta^2} &= \lim_{\delta\to 0} \frac{\exp\Big(n\displaystyle\int_0^\delta\int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t\Big) - 1}{\delta^2} \\
&= \lim_{\delta\to 0} \frac{\exp\Big(n\displaystyle\int_0^\delta\int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t\Big) - 1}{n\displaystyle\int_0^\delta\int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t} \\
&\qquad \times \lim_{\delta\to 0} \frac{n\displaystyle\int_0^\delta\int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t}{\delta^2} \\
&= 1 \cdot n\,\ddot{\mathsf{K}}\,(\theta) \\
&= n\,\mathsf{Var}_\theta[T(X_1)] > 0
\end{aligned}
$$

となる. 最後から 2 番目の等号は $\ddot{\mathsf{K}}$ が θ に関して連続ならば

$$
\begin{aligned}
\frac{1}{\delta^2}\int_0^\delta\int_0^\delta \ddot{\mathsf{K}}\,(\theta+s+t)\,\mathrm{d}s\,\mathrm{d}t &= \int_0^1\int_0^1 \ddot{\mathsf{K}}\,\big(\theta+\delta(s+t)\big)\,\mathrm{d}s\,\mathrm{d}t \\
&\xrightarrow{\delta\to 0} \ddot{\mathsf{K}}\,(\theta)\int_0^1\int_0^1 \mathrm{d}s\,\mathrm{d}t = \ddot{\mathsf{K}}\,(\theta)
\end{aligned}
$$

と $\displaystyle\lim_{u\to 0}\frac{\{e^u-1\}}{u} = 1$ からわかる.

② (3) の確認: $\theta \in \Theta$ を固定する. θ は Θ の内点とする. すると十分小さい $\epsilon > 0$ をとって

$$
[\theta - 3\epsilon,\,\theta + 3\epsilon] \subset \Theta
$$

となるようにできる. また

$$S(\boldsymbol{x}) := \sum_{i=1}^{n} T(x_i)$$

とおく. $|u| < \epsilon$ に対して

$$\begin{aligned} R_u(\boldsymbol{x}) &:= \frac{\mathsf{p}_{\theta+u}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \exp\Big(uS(\boldsymbol{x}) - n\big\{\mathsf{K}(\theta+u) - \mathsf{K}(\theta)\big\} \Big) \end{aligned}$$

と定義する. このとき $R_0(\boldsymbol{x}) = 1$ であり

$$\frac{\partial}{\partial u} R_u(\boldsymbol{x}) = R_u(\boldsymbol{x})\big\{S(\boldsymbol{x}) - n\,\dot{\mathsf{K}}\,(\theta+u)\big\}$$

となる. $\delta := \widetilde{\theta} - \theta$ とすると, 平均値の定理から 0 と δ の間のある u_δ が存在して

$$\begin{aligned} \left| \frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x}) - \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})}{(\widetilde{\theta}-\theta)\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} \right| &= \left| \frac{R_\delta(\boldsymbol{x}) - R_0(\boldsymbol{x})}{\delta} \right| \\ &= R_{u_\delta}(\boldsymbol{x}) \left| S(\boldsymbol{x}) - n\,\dot{\mathsf{K}}\,(\theta+u_\delta) \right| \end{aligned}$$

となる. ここで

$$\begin{aligned} M_\theta &:= \sup_{|u|\le\epsilon} \left| \dot{\mathsf{K}}\,(\theta+u) \right|, \\ C_\theta &:= \sup_{|u|\le\epsilon} |\mathsf{K}(\theta+u) - \mathsf{K}(\theta)| \end{aligned}$$

とおく. K および $\dot{\mathsf{K}}$ の連続性から $M_\theta < \infty$ および $C_\theta < \infty$ である. また $|u_\delta| \le \epsilon$ なので

$$R_{u_\delta}(\boldsymbol{x}) \le \exp\big\{\epsilon|S(\boldsymbol{x})| + nC_\theta\big\}$$

と評価できる. したがって

$$G_\theta(\boldsymbol{x}) := e^{nC_\theta} e^{\epsilon|S(\boldsymbol{x})|}\big\{|S(\boldsymbol{x})| + nM_\theta\big\}$$

とおけば

$$\left| \frac{\mathsf{p}_{\widetilde{\theta}}^{\boldsymbol{X}}(\boldsymbol{x}) - \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})}{(\widetilde{\theta}-\theta)\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} \right| \le G_\theta(\boldsymbol{x})$$

が成り立つ. さらに, ある定数 $C > 0$ に対して

$$(|y| + nM_\theta)^2 \le Ce^{\epsilon|y|} \qquad (y \in \mathbb{R})$$

とできる [2] ので

$$\begin{aligned} G_\theta(\boldsymbol{x})^2 &\le Ce^{2nC_\theta} e^{3\epsilon|S(\boldsymbol{x})|} \\ &\le Ce^{2nC_\theta}\{e^{3\epsilon S(\boldsymbol{x})} + e^{-3\epsilon S(\boldsymbol{x})}\} \end{aligned}$$

となる. ところが

$$\mathsf{E}_\theta\left[e^{aS(\boldsymbol{X})}\right] = \exp\Big(n\big\{\mathsf{K}(\theta + a) - \mathsf{K}(\theta)\big\}\Big)$$

であり, $\theta \pm 3\epsilon \in \Theta$ であるから

$$\begin{aligned} \mathsf{E}_\theta\big[G_\theta(\boldsymbol{X})^2\big] &\le Ce^{2nC_\theta}\bigg[\exp\Big(n\big\{\mathsf{K}(\theta + 3\epsilon) - \mathsf{K}(\theta)\big\}\Big) \\ &\qquad + \exp\Big(n\big\{\mathsf{K}(\theta - 3\epsilon) - \mathsf{K}(\theta)\big\}\Big)\bigg] \\ &< \infty \end{aligned}$$

がわかる. よって正則条件 (3) がみたされる. □

定義 7.21. $\boldsymbol{X} \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\widehat{\theta}_n(\boldsymbol{X})$ は R 有効 [3] であるとは

$$\mathsf{Var}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\big] = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)} \qquad (\theta \in \Theta)$$

をみたすときをいう.

定理 7.22. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset \mathbb{R})$ とする. θ の不偏推定量 $\widehat{\theta}_n$ が R 有効であるための必要十分条件は $\mathbb{R}$ 上の $\mathbb{R}$ 値関数 $T(x)$, $A(\theta)$, $\mathsf{K}(\theta)$, $g(x)$ があって

$$\widehat{\theta}_n(\boldsymbol{X}) = \frac{1}{n}\sum_{j=1}^{n} T(X_j) \tag{7.15}$$

---

[2] 指数関数は任意の多項式より速く増大するので

$$\sup_{y\in\mathbb{R}}\big(|y| + nM_\theta\big)^2 e^{-\epsilon|y|} < \infty$$

である. したがって, ある定数 $C > 0$ が存在して

$$\big(|y| + nM_\theta\big)^2 \le Ce^{\epsilon|y|} \qquad (y \in \mathbb{R})$$

となる.

[3] Rao 有効の意味であろう.

$$\log \mathsf{p}_\theta(x) = A(\theta)T(x) - \mathsf{K}(\theta) + g(x), \tag{7.16}$$

$$\int_{\mathbb{X}} T(x)\mathsf{p}_\theta(x)\,\mathrm{d}x = \theta \tag{7.17}$$

をみたすときである. ただし X_1 の p.d.f.(または p.m.f.) と値域をそれぞれ $\mathsf{p}_\theta(x)$ と $\mathbb{X}(\subset \mathbb{R})$ と書いた.

Proof. 連続型分布に対する証明を与える. 離散型分布に対しては, 積分記号を和の記号に替えればよい. 表現 (7.15) – (7.17) が成立したとする. $\ell_n(\theta|\,\boldsymbol{x}) = \sum_{j=1}^n \log \mathsf{p}_\theta(x_j)$ $\big(\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)\big)$ とおく. すると

$$\dot{\ell}_n\,(\theta|\,\boldsymbol{x}) = \frac{\partial}{\partial\theta}\ell_n(\theta|\,\boldsymbol{x}) = \dot{A}\,(\theta)\sum_{j=1}^n T(x_j) - n\,\dot{\mathsf{K}}\,(\theta),$$
$$\dot{A}\,(\theta) = \frac{\partial A}{\partial\theta}(\theta), \quad \dot{\mathsf{K}}\,(\theta) = \frac{\partial \mathsf{K}}{\partial\theta}(\theta)$$

となる. よって

$$\mathsf{E}_\theta\big[\dot{\ell}_n\,(\theta|\,\boldsymbol{X})\big] = 0 \;\Leftrightarrow \dot{A}\,(\theta)\mathsf{E}_\theta\big[T(X_1)\big] = \dot{\mathsf{K}}\,(\theta)$$

と (7.17) とあわせると

$$\theta\,\dot{A}\,(\theta) = \dot{\mathsf{K}}\,(\theta) \tag{7.18}$$

となる. さらに

$$0 = \int_{\mathbb{X}} \{T(x) - \theta\}\mathsf{p}_\theta(x)\,\mathrm{d}x = \int_{\mathbb{X}} \{T(x) - \theta\}\exp\{\log \mathsf{p}(x|\,\theta)\}\,\mathrm{d}x$$

を (7.16) と 7.18) に注意して, θ に関して微分すると

$$0 = -\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}x + \int_{\mathbb{X}} \{T(x) - \theta\}\big\{\dot{A}\,(\theta)T(x) - \underbrace{\dot{\mathsf{K}}\,(\theta)}_{=\theta\dot{A}(\theta)}\big\}\mathsf{p}_\theta(x)\,\mathrm{d}x$$
$$= -\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}x + \int_{\mathbb{X}} \{T(x) - \theta\}\big\{\dot{A}\,(\theta)T(x) - \theta\,\dot{A}\,(\theta)\big\}\mathsf{p}_\theta(x)\,\mathrm{d}x$$

となることがわかる. よって, $\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}x = 1$ に注意して, 上式を整理すると

$$1 = \dot{A}\,(\theta)\mathsf{E}_\theta\big[\{T(X_1) - \theta\}^2\big]$$

を得る. さらに $\mathsf{E}_\theta[T(X_1)] = \theta$ に注意すれば

$$\mathsf{Var}_\theta\big[T(X_1)\big] = \frac{1}{\dot{A}\,(\theta)} \tag{7.19}$$

となる. したがって

$$\begin{aligned}\mathcal{F}_{\boldsymbol{X}}(\theta) &= \mathsf{Var}_\theta\bigl[\dot{\ell}_n\,(\theta|\,\boldsymbol{X})\bigr] \qquad (7.20)\\ &= \mathsf{Var}_\theta\biggl[\sum_{j=1}^n\{\dot{A}\,(\theta)T(X_j) - \dot{\mathsf{K}}\,(\theta)\}\biggr]\\ &= \mathsf{Var}_\theta\biggl[\sum_{j=1}^n \dot{A}\,(\theta)\bigl\{T(X_j)\bigr\}\biggr] \qquad (\because \text{分散は平行移動に関して不変})\\ &= \sum_{j=1}^n\{\dot{A}\,(\theta)\}^2\mathsf{Var}\bigl[T(X_j)\bigr] \qquad (\because X_1,\,X_2,\,\ldots,\,X_n \text{ は独立})\\ &= n\{\dot{A}\,(\theta)\}^2\mathsf{Var}\bigl[T(X_1)\bigr] \qquad (7.21)\\ &= n\,\dot{A}\,(\theta) \qquad (\because (7.19))\end{aligned}$$

を得る. また, $\widehat{\theta}_n(\boldsymbol{X}) = \dfrac{1}{n}\sum_{j=1}^n T(X_j)$ なので

$$\mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\bigr] = \theta$$

となり $\widehat{\theta}_n$ は θ の不偏推定量である. 再度, (7.19) に注意すると

$$\mathsf{Var}_\theta\bigl[\widehat{\theta}_n\bigr] = \mathsf{Var}_\theta\biggl[\frac{1}{n}\sum_{j=1}^n T(X_j)\biggr] = \frac{1}{n^2}\sum_{j=1}^n \mathsf{Var}_\theta\bigl[T(X_j)\bigr] = \frac{1}{n\,\dot{A}\,(\theta)} = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

となり $\widehat{\theta}_n$ は R 有効推定量となる.

次に逆を示す. P_θ の同時 p.d.f. $\prod_{j=1}^n \mathsf{p}_\theta(x_j)$ に対して, その対数尤度関数を $\ell_n(\theta|\,\boldsymbol{x}) = \log\biggl(\prod_{j=1}^n \mathsf{p}_\theta(x_j)\biggr)$ と書くことにする. すると

$$\int_{\mathbb{X}^n} \exp\{\ell_n(\theta|\,\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x} = 1$$

である. この式を θ に関して微分すると

$$0 = \int_{\mathbb{X}^n} \dot{\ell}_n\,(\theta|\,\boldsymbol{x})\exp\{\ell_n(\theta|\,\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x} = \mathsf{E}_\theta\bigl[\dot{\ell}_n\,(\theta|\,\boldsymbol{X})\bigr] \qquad (7.22)$$

となる. 同様に $\mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\bigr] = \theta$ より

$$\int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x})\exp\{\ell_n(\theta|\,\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x} = \theta$$

となる. この式を θ に関して微分すると

$$1 = \int_{\mathbb{X}^n} \widehat{\theta}_n(\boldsymbol{x})\,\dot{\ell}_n\,(\theta|\,\boldsymbol{x})\exp\{\ell_n(\theta|\,\boldsymbol{x})\}\,\mathrm{d}\boldsymbol{x} = \mathsf{E}_\theta\bigl[\widehat{\theta}_n(\boldsymbol{X})\,\dot{\ell}_n\,(\theta|\,\boldsymbol{X})\bigr] \qquad (7.23)$$

を得る. (7.22) と (7.23) を合わせると

$$\begin{aligned}\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X})-\theta)\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big] &= \underbrace{\mathsf{E}_\theta\big[\widehat{\theta}_n(\boldsymbol{X})\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1} - \theta\underbrace{\mathsf{E}_\theta\big[\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=0}\\ &= 1 \qquad (7.24)\end{aligned}$$

を得る. いま

$$Y = \frac{\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)}$$

とおく. (7.20) に注意すると

$$\mathsf{E}_\theta\big[Y\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big] = \mathsf{E}_\theta\left[\frac{\big\{\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big\}^2}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\right] = \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\mathsf{Var}[\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})] = 1$$

となる. これと (7.24) を合わせると

$$\begin{aligned}&\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X})-\theta-Y)Y\big]\\ &\quad= \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\big[(\widehat{\theta}_n(\boldsymbol{X})-\theta)\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1} - \frac{1}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\underbrace{\mathsf{E}_\theta\big[Y\overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\big]}_{=1}\\ &\quad= 0\end{aligned}$$

を得る. 上の式を用いると

$$\begin{aligned}\mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big)^2\big] &= \mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta-Y+Y\big)^2\big]\\ &= \mathsf{E}_\theta\big[\big(\widehat{\theta}_n(\boldsymbol{X})-\theta-Y\big)^2\big] + \mathsf{E}_\theta\big[Y^2\big]\\ &\geq \mathsf{E}_\theta\big[Y^2\big]\end{aligned}$$

となる. よって等号成立は

$$\begin{aligned}\widehat{\theta}_n(\boldsymbol{X})-\theta-Y=0 \;&\Leftrightarrow\; \widehat{\theta}_n(\boldsymbol{X})-\theta = \frac{\overset{\bullet}{\ell}(\theta|\,\boldsymbol{X})}{\mathcal{F}_{\boldsymbol{X}}(\theta)}\\ &\Leftrightarrow\; \mathcal{F}_{\boldsymbol{X}}(\theta)\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big) = \overset{\bullet}{\ell}_n(\theta|\,\boldsymbol{X})\end{aligned}$$

となる. このことより, $\overline{\theta}$, $\underline{\theta}\in\mathbb{X}$ とすると

$$\begin{aligned}\ell_n(\overline{\theta}|\,\boldsymbol{X}) - \ell_n(\underline{\theta}|\,\boldsymbol{X}) &= \int_{\underline{\theta}}^{\overline{\theta}}\overset{\bullet}{\ell}_n(\boldsymbol{X}|\,\theta)\,\mathrm{d}\theta = \int_{\underline{\theta}}^{\overline{\theta}}\mathcal{F}_{\boldsymbol{X}}(\theta)\big(\widehat{\theta}_n(\boldsymbol{X})-\theta\big)\,\mathrm{d}\theta\\ &= \widetilde{A}(\overline{\theta},\,\underline{\theta})\widehat{\theta}_n(\boldsymbol{X}) - \widetilde{B}(\overline{\theta},\,\underline{\theta})\end{aligned}$$

と書ける. ただし

$$\widetilde{A}(\overline{\theta},\,\underline{\theta}) := \int_{\underline{\theta}}^{\overline{\theta}}\mathcal{F}_{\boldsymbol{X}}(\theta)\,\mathrm{d}\theta; \qquad \widetilde{B}(\overline{\theta},\,\underline{\theta}) := \int_{\underline{\theta}}^{\overline{\theta}}\theta\mathcal{F}_{\boldsymbol{X}}(\theta)\,\mathrm{d}\theta$$

である. ここで $n = 1$, $\overline{\theta} = \theta$, $\underline{\theta} = 1$, $X_1 = x$ とおくと

$$\ell_1(\theta|\, x) = T(x)A(\theta) - B(\theta) + g(x)$$

と書ける. ただし

$$T(x) = \widehat{\theta}_1(x), \quad g(x) = \ell_1(1|\, x), \quad A(\theta) = \widetilde{A}(\theta,\, 1), \quad B(\theta) = \widetilde{B}(\theta,\, 1)$$

と書ける. □

注意 7.23. 定理 7.22 から, R 有効な推定量が存在する統計的モデルは指数型分布族となることがわかる. □

7.4 章末注釈と参考文献

定理 7.11 の証明は $[7, \mathrm{pp}.241 - 242]$ を借用した. 定理 7.14 の証明は $[50, \mathrm{pp}.35 - 41]$ を借用した. 定理 7.22 は $[30, \mathrm{pp}.39 - 40]$ を借用した.

7.5 演習問題

演習問題 7.1. 離散型確率変数 X は, p.m.f.

$$\mathsf{p}_\theta(x) = \begin{cases} \left(\dfrac{\theta}{2}\right)^{|x|} (1-\theta)^{1-|x|} & (x = -1,\, 0,\, 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

からの標本の大きさが 1 のランダム標本とする. ただし, $0 < \theta < 1$ で, X は確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上で定義された確率変数とする. ふたつの統計量

$$S = S(X) = |X|, \qquad T = T(X) = \begin{cases} 2 & (X = 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を考える.
(1) S の確率分布を求めよ [4].
(2) T の確率分布を求めよ.
(3) S が与えられたときの X の条件付確率分布を求め [5], S は θ の十分統計量かどうかを調べよ.
(4) S は θ の不偏推定量かどうかを調べよ.
(5) T は θ の不偏推定量かどうかを調べよ.

[4] 確率関数または確率分布表を求めること.
[5] 条件付き p.m.f. を求めるか, X が与えられた値ごとの確率分布表を求めればよい.

(6) S と T の平均 2 乗誤差 $\mathsf{MSE}_S(\theta) = \mathsf{E}[(S-\theta)^2]$ と $\mathsf{MSE}_T(\theta) = \mathsf{E}[(T-\theta)^2]$ を求めよ. S と T の平均 2 乗誤差 $\mathsf{MSE}_S(\theta)$ と $\mathsf{MSE}_T(\theta)$ の大小の比較をせよ. (横軸を θ とし, 縦軸を MSE の値として, S と T の平均 2 乗誤差 $\mathsf{MSE}_S(\theta)$ と $\mathsf{MSE}_T(\theta)$ のグラフを描き比較すること.)

演習問題 7.2. $\theta > 0$ とする. 連続型確率変数 X は p.d.f.

$$\mathsf{p}_\theta^X(x) = \begin{cases} \theta x^{\theta-1} & (0 < x < 1) \\ 0 & (\text{その他}) \end{cases}$$

からの大きさ 1 のランダム標本とする. ただし, X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数とする.
(1) 連続型確率変数 X の期待値 $\mathsf{E}[X]$ を求めよ.
(2) 連続型確率変数 X の分布関数 $\mathsf{F}_\theta^X(x) = \mathsf{Pr}(X \leq x)\,(x \in \mathbb{R})$ を求めよ.
(3) $X = x\,(0 < x < 1)$ を観測したのときの尤度関数 $\mathsf{lik}(\theta|\, x)$ を述べよ (答えのみでよい).
(4) θ の最尤推定値を求めよ.

演習問題 7.3. $m, n \geq 2$ を整数とし

$$\begin{aligned} X_1, X_2, \ldots, X_m &\overset{\text{i.i.d.}}{\sim} N(\mu_1, \sigma^2) \\ Y_1, Y_2, \ldots, Y_n &\overset{\text{i.i.d.}}{\sim} N(\mu_2, \sigma^2) \end{aligned}$$

とする. さらに

$$\begin{aligned} \overline{X}_m &= \frac{1}{m}\sum_{i=1}^m X_i \\ \overline{Y}_n &= \frac{1}{n}\sum_{i=1}^n Y_i \\ \widehat{\sigma}^2 &= \frac{1}{m+n-2}\left\{\sum_{i=1}^m (\overline{X}_m - X_i)^2 + \sum_{i=1}^n (\overline{Y}_n - Y_i)^2\right\} \end{aligned}$$

とおく. これらの確率変数は同じ確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されたものとする. このとき, 以下の問いに答えよ.
(1)

$$\overline{X}_m - \overline{Y}_n \sim \mathsf{N}\left(\mu_1 - \mu_2,\, \frac{m+n}{mn}\sigma^2\right)$$

となることを説明 (証明) せよ.
(2)

$$\mathsf{E}[\widehat{\sigma}^2] = \sigma^2$$

となることを説明（証明）せよ.

演習問題 7.4. $n \geq 2$ を整数とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Po}(\lambda)$ とする [6]. ただし, $\lambda > 0$ で, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されているとする.
(1) $\sum_{i=1}^{n} X_i$ の分布は $\mathsf{Po}(n\lambda)$ となることを積率母関数を計算することで示せ.
(2) λ の任意の不偏推定量の分散についてのその下限（Cramér-Rao の下限）を求めよ.
(3) λ の最尤推定量を求め, その分散が Cramér-Rao の下限に到達することを確認せよ.

演習問題 7.5. $X_1, X_2, \ldots, X_n$ を母集団分布 (平均は θ, 分散は 1^2 の正規分布)

$$\mathsf{p}_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right)$$

からの大きさ n のランダム標本とする. ただし, 母数 $\theta\,(-\infty < \theta < \infty)$ は未知とする. このとき, 以下の問いに答えよ.
(1) $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したときの θ の尤度関数 $\mathsf{lik}_n(\theta|\, x_1, x_2, \ldots, x_n)$ と対数尤度関数 $\mathsf{lik}_n(\theta|\, x_1, x_2, \ldots, x_n)$ を書け.
(2) θ の最尤推定値 $\widehat{\theta}_n(x_1, x_2, \ldots, x_n)$ を求めよ.
(3) θ の最尤推定量 $\widehat{\theta}_n(X_1, X_2, \ldots, X_n)$ の平均 $\mathsf{E}[\widehat{\theta}_n(X_1, X_2, \ldots, X_n)]$ を求めよ.
(4) θ の最尤推定量 $\widehat{\theta}_n(X_1, X_2, \ldots, X_n)$ の分散 $\mathsf{Var}[\widehat{\theta}_n(X_1, X_2, \ldots, X_n)]$ を求めよ.
(5) 任意の正数 ϵ に対して,

$$\lim_{n\to\infty} \mathsf{Pr}\big(|\widehat{\theta}_n(X_1, X_2, \ldots, X_n) - \theta| \geq \epsilon\big) = 0$$

を示せ.

演習問題 7.6. 確率空間 $(\Omega, \mathcal{A}, P)$ 上の確率変数 X_1, X_2, X_3 は独立同一に平均が μ, 分散が σ^2 の正規分布 $\mathsf{N}(\mu, \sigma^2)$ に従うとする. ただし, $\mu \in \mathbb{R}$, $0 < \sigma < \infty$ である. このとき, 以下の問に答えよ.
(1) X_1 の積率母関数 $M(t) = \mathsf{E}[e^{tX_1}]$ が

$$M(t) = \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right)$$

---

[6]

$$\mathsf{Pr}(X_1 = x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad (x = 0, 1, \ldots)$$

である. ただし, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されたものとする. その積率母関数は

$$\mathsf{M}_{X_1}(t) = \mathsf{E}[e^{tX_1}] = e^{(e^t - 1)\lambda}$$

である.

で与えられることを示せ. ただし, 任意の $\mu \in \mathbb{R}$ と $\sigma > 0$ に対して

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x = 1$$

を証明なしで用いてよい.
(2) X_1 の期待値 $\mathsf{E}[X_1]$ と 2 次の積率 $\mathbb{E}[X_1^2]$ を (1) で求めた積率母関数を利用して求めよ.
(3) a, b, c, $d \in \mathbb{R}$ に対して, σ^2 の推定量の族

$$T := g(X_1, X_2, X_3) = a\big\{X_1^2 + X_2^2 + X_3^2\big\} + b\big\{X_1X_2 + X_1X_3 + X_2X_3\big\} + c\big\{X_1 + X_2 + X_3\big\} + d$$

を考える.

$$\mathsf{E}[T] = \sigma^2$$

をみたす a, b, c, d を求めよ.
(4) (3) で求めた a, b, c, d に対する T は

$$T = \frac{1}{2}\left((X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + (X_3 - \overline{X}2)^2\right); \quad \overline{X} = \frac{1}{3}\big(X_1 + X_2 + X_3\big)$$

の形に変形できることを確認せよ.

演習問題 7.7. T_1 と T_2 を母数 θ の最小分散不偏推定量とする. このとき, 以下の問いに答えよ.
(1) $\mathsf{E}\big[\frac{1}{2}(T_1 + T_2)\big]$ を計算せよ.
(2) $\mathsf{Var}\big[\frac{1}{2}(T_1 + T_2)\big]$ を $\mathsf{Var}[T_1]$, $\mathsf{Cov}[T_1, T_2]$ で表現せよ.
(3) $\mathsf{Cov}[T_1, T_2] = \mathsf{Var}[T_1]$ を証明せよ.
(4) $\mathsf{Var}[T_1 - T_2]$ の値を求めよ.
(5) $\mathsf{Pr}(T_1 = T_2) = 1$ を示せ.

ヒント (3) T_1 は最小分散不偏推定量であることと Cauchy-Schwarz の不等式を $\mathsf{Cov}[T_1, T_2]$ に適用するとよい.
(5) 非負値確率変数 W に対して

$$\mathsf{E}[W] = 0 \;\Rightarrow\; \mathsf{Pr}(W = 0) = 1$$

を証明なしで用いてよい.

演習問題 7.8.

演習問題 7.9.

第8章　検定と信頼区間

節 8.1 では, 仮説検定問題の枠組みと考え方を説明する. 節 8.2 では, 帰無仮説も対立仮説も単純であるとき, 検定統計量の最適定理である Neyman-Pearson の補題を説明する. 節 8.3 では, 検定統計量の導出原理を説明する. 節 8.4 では様々な検定方法をまとめる. 節 8.6 では, 区間推定量の考え方を説明する. 節 8.7 では, 区間推定量の構成法の代表的なものを説明する.

8.1　仮説検定の考え方

母集団分布を特徴付ける母数について想定したある仮説の真偽を標本に基づいて調べることを **仮説検定**(hypothesis test) という.

いま

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}^*}^{\otimes n} \, (\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$$

とする. ただし $d, n \in \mathbb{N}$ で, Θ は母数空間である. また, $\mathbb{X}^n \, (\subset \mathbb{R}^n)$ を $\boldsymbol{X}$ の値域としたとき, $\mathsf{P}_{\boldsymbol{\theta}^*}^{\otimes n}$ は $\big(\mathbb{X}^n, \mathcal{B}(\mathbb{X}^n)\big)$ 上の確率測度 ($\boldsymbol{X}$ の分布) である. $\Theta_0 \subset \Theta$ は空でない Θ の真部分集合とし, $\boldsymbol{\theta}^*$ が Θ_0 に入るか否かを調べたいとき, 仮説

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta_1 := \Theta \setminus \Theta_0 \tag{8.1}$$

を考える. H_0 または H_1 のいずれかが正しいかを判断することを「H_0 を H_1 に対して検定 (test) する」という. H_0 を **帰無仮説**(null hypothesis) といい, H_1 を **対立仮説**(alternative hypothesis) という. Θ_0 が Θ の 1 つの元から成るとき H_0 を **単純仮説**(simple hypothesis) という. そうでないとき H_0 を **複合仮説**(composite hypothesis) という. 言葉を乱用して, 「Θ_0 は単純仮説である」等ということもある.

仮説 (8.1) に対して, 以下のように検定方式を定めることができる. $\boldsymbol{X}$ の取り得るすべての値の集合を $\mathbb{X}^n (\subset \mathbb{R}^n)$ と表す. $\mathbb{X}^n$ を 2 つの排反で空でない部分集合 W と W^c に分割する. すなわち $W \neq \varnothing$, $W^c \neq \varnothing$ で $W \cup W^c = \mathbb{X}^n$ かつ $W \cap W^c = \varnothing$ である. $\boldsymbol{X}$ の実現値を $\boldsymbol{x}$ と書いたとき, 検定方式は

$$\boldsymbol{x} \in W \;\Rightarrow\; \text{帰無仮説 } H_0 \text{ を棄却し, 対立仮説 } H_1 \text{ を採択,}$$

$$\boldsymbol{x} \in W^c \Rightarrow \text{帰無仮説 } H_0 \text{ を受容}$$

と表現できる. このとき W を **棄却域**(critical region) といい, W^c を **受容域**(acceptance region) という.

上のように定めた検定方式には 2 つのタイプの誤りが起こる可能性がある. (1) 帰無仮説 H_0 が正しいにもかかわらず標本の実現値 $\boldsymbol{x}$ に基づいて検定した結果, H_0 を棄却してしまうこと. 逆に, (2) 対立仮説 H_1 が正しいにもかかわらず標本の実現値 $\boldsymbol{x}$ に基づいて検定した結果, H_0 を受容してしまうことである. (1) の誤判断を **第 1 種の誤り** といい, (2) の誤判断を **第 2 種の誤り** とそれぞれ呼ぶ. 一般に一方の誤りが起こる確率を小さくする検定方式は, 他方の誤りを起こす確率を大きくする. すなわち, 両者の誤りが起こる確率を同時に小さくする検定方式はないことが知られている.

以下では「よい」検定方式を一般的な形で定式化することを考える. 関数 $\phi\colon \mathbb{X}^n \to [0, 1]$ は可測関数とする. この関数 ϕ を用いて次のように検定方式を定める. $\boldsymbol{X} = \boldsymbol{x}$ を観測したとき, 確率 $\phi(\boldsymbol{x})$ で帰無仮説 H_0 を棄却する検定方式を考える. この ϕ を **検定関数**(test function) という. 関数 ϕ が $\mathbb{X}$ の空でない部分集合の定義関数のとき, この検定関数 ϕ で定まる検定方式を **非確率化検定**(nonrandomized test) という. すなわち

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \in W^c) \end{cases}$$

とすると棄却域 W をもつ非確率化検定が定まる. そうでない検定方式を **確率化検定**(randomized test) という.

以後 $\phi(\boldsymbol{X})$ を **検定統計量**(test statistic) ということにする. さらに検定統計量 ϕ によって定まる検定方式を単に検定ということにする. 検定統計量 $\phi(\boldsymbol{X})$ の第 1 種の誤りの確率は

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_0)$$

となり, 第 2 種の誤りの確率は

$$1 - \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_1)$$

となる. ただし

$$\mathsf{E}_{\boldsymbol{\theta}}[\phi(\boldsymbol{X})] = \begin{cases} \sum_{x \in \mathbb{X}} \phi(\boldsymbol{x}) \mathsf{p}_{\boldsymbol{\theta}}(\boldsymbol{x}) & \text{(離散型の場合)} \\ \int_{\mathbb{X}^n} \phi(\boldsymbol{x}) \mathsf{p}_{\boldsymbol{\theta}}(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} & \text{(連続型の場合)} \end{cases}$$

と定めた. 「よい」検定として, まず第 1 種の誤りの確率の Θ_0 上の上限を $\alpha\,(0 < \alpha < 1)$ 以下にするような検定を考える. すなわち

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \le \alpha \tag{8.2}$$

である. (8.2) をみたす検定関数または検定 ϕ を **有意水準 α の検定**(level α test) という. つぎに有意水準 α の検定 ϕ の中で, 第 2 種の誤りの確率

$$1 - \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_1) \tag{8.3}$$

を Θ_1 上で最小にするものを見つけることを目指す. すなわち

$$\mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_1)$$

を最大にするものである. この確率を $\boldsymbol{\theta} \in \Theta_1$ の関数とみて

$$\beta\big(\boldsymbol{\theta}\big) := \mathsf{E}_{\boldsymbol{\theta}}\big[\phi(\boldsymbol{X})\big] \qquad (\boldsymbol{\theta} \in \Theta_1)$$

と表記する. これを **検出力関数**(power function) または **検出力** という. したがって「よい」検定は次のように定義される.

定義 8.1. 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定 ϕ で検出力を任意の $\boldsymbol{\theta} \in \Theta_1$ に対して最大にするものを有意水準 α の **一様最強力検定**(uniformly most powerful test= u.m.p. 検定) という. 特に帰無仮説と対立仮説が単純仮説であるとき, u.m.p. 検定を単に有意水準 α の **最強力検定**(m.p. 検定) という.

最後に, P 値を定義する. 検定問題 (8.1) において, 関数 $T : \mathbb{X}^n \longrightarrow \mathbb{R}$ と定数 $c \in \mathbb{R}$ により棄却域 $W \subset \mathbb{X}^n$ が

$$W := \big\{\boldsymbol{x} \in \mathbb{X}^n;\, T(\boldsymbol{x}) > c\big\}$$

で定まる検定関数 ϕ を考える. すなわち, $\phi(\boldsymbol{x}) = \mathbb{1}_W(\boldsymbol{x})\,(\boldsymbol{x} \in \mathbb{X}^n)$ である. ただし, 定数 c は

$$\sup_{\theta \in \Theta_0} \mathsf{E}_\theta\big[\phi(\boldsymbol{X})\big] \le \alpha$$

をみたすものである.

定義 8.2.

$$p(\boldsymbol{x}) := \sup_{\theta \in \Theta_0} \mathsf{Pr}_\theta\big(T(\boldsymbol{X}) \ge T(\boldsymbol{x})\big) = \sup_{\theta \in \Theta_0} \mathsf{E}_\theta\big[\mathbb{1}_{[W(\boldsymbol{x}),\,\infty)}\big(T(\boldsymbol{X})\big)\big]$$

を実現値 $\boldsymbol{X} = \boldsymbol{x}$ に対する P **値** (P 値) という.

定理 8.3. $p(\,\cdot\,)$ を P 値とする. このとき, すべての $\theta \in \Theta_0$ と $\alpha\,(0 < \alpha < 1)$ に対して

$$\mathsf{Pr}_\theta\big(p(\boldsymbol{X}) \le \alpha\big) \le \alpha$$

が成り立つ.

Proof. $\theta_0 \in \Theta_0$ を固定して

$$p_{\theta_0}(\boldsymbol{x}) := \mathsf{Pr}_{\theta_0}\big(T(\boldsymbol{X}) \geq T(\boldsymbol{x})\big)$$

とおき, $-T(\boldsymbol{X})$ の分布関数を F_{θ_0} と書く. すなわち

$$F_{\theta_0}(t) = \mathsf{Pr}_{\theta_0}\big(-T(\boldsymbol{X}) \leq t\big) \qquad (t \in \mathbb{R})$$

である. すると

$$p_{\theta_0}(\boldsymbol{x}) = \mathsf{Pr}_{\theta_0}\big(T(\boldsymbol{X}) \geq T(\boldsymbol{x})\big) = \mathsf{Pr}_{\theta_0}\big(-T(\boldsymbol{X}) \leq -T(\boldsymbol{x})\big) = \mathsf{F}_{\theta_0}\big(-T(\boldsymbol{x})\big)$$

と書き直せる. このことから, $\mathsf{F}_{\theta_0}\big(-T(\boldsymbol{X})\big)$ は $-T(\boldsymbol{X})$ の分布関数 F_{θ_0} に $-T(\boldsymbol{X})$ を代入したものであるに注意する. すると命題 1.41 から

$$\mathsf{F}_{\theta_0}\big(-T(\boldsymbol{X})\big) \sim \mathsf{U}(0, 1) \tag{8.4}$$

となることがわかる. $p_{\theta_0}(\boldsymbol{X}) = \mathsf{F}_{\theta_0}\big(-T(\boldsymbol{X})\big)$ と (8.4) から, $0 < \alpha < 1$ に対して

$$\mathsf{Pr}_{\theta_0}\big(p_{\theta_0}(\boldsymbol{X}) \leq \alpha\big) = \mathsf{Pr}_{\theta_0}\big(\mathsf{F}_{\theta_0}\big(-T(\boldsymbol{X})\big) \leq \alpha\big) \leq \alpha$$

を得る. さらに

$$p(\boldsymbol{x}) = \sup_{\theta \in \Theta_0} p_\theta(\boldsymbol{x}) \geq p_{\theta_0}(\boldsymbol{x})$$

であるから

$$\mathsf{Pr}_{\theta_0}\big(p(\boldsymbol{X}) \leq \alpha\big) \leq \mathsf{Pr}_{\theta_0}\big(p_{\theta_0}(\boldsymbol{X}) \leq \alpha\big) \leq \alpha$$

がわかる. よって, 定理の主張は証明された. □

8.2 Neyman-Pearson の定理

まず m.p. 検定を求める最も基本的定理を述べる. 以下では, 簡単のために $\boldsymbol{X}$ は同時 p.d.f. $\mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})$ をもつとして議論を進めていく. 離散型確率変数のときは同時 p.m.f. を考え, 積分を和の記号に替えればよい.

定理 8.4. (Neyman-Pearson の定理) 母数空間は $\mathbb{R}^d$ の異なる 2 点から成るとする. $\Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ である. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}\,(\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d)$ とし, $\mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})\,(\boldsymbol{\theta} \in \Theta, \boldsymbol{x} \in \mathbb{R}^n)$ を $\boldsymbol{X}$ の同時 p.d.f. または p.m.f. とする. 検定問題

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$$

に対する有意水準 $\alpha\,(0<\alpha<1)$ の m.p. 検定 ϕ_0 は以下で与えられる.

$$\phi_0(\boldsymbol{x})=\begin{cases}1 & (\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})>c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\\ \gamma & (\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})=c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\\ 0 & (\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})<c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\end{cases} \tag{8.5}$$

である. ただし $\gamma,\,c\,(0\le\gamma\le1,\,c>0)$ は

$$\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi_0(\boldsymbol{X})\big]=\alpha \tag{8.6}$$

から定まる定数である.

Proof. まず

$$\begin{aligned}B_1&=\{\boldsymbol{x}\in\mathbb{X}^n;\,\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})>c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\},\\ B_2&=\{\boldsymbol{x}\in\mathbb{X}^n;\,\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})=c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\},\\ B_3&=\{\boldsymbol{x}\in\mathbb{X}^n;\,\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})<c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\}\end{aligned}$$

とする. ϕ は有意水準 α の任意の検定とする. すなわち

$$\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi(\boldsymbol{X})\big]\le\alpha \tag{8.7}$$

をみたす. 一方 (8.5) より

$$\begin{aligned}&\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi_0(\boldsymbol{X})\big]-\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi(\boldsymbol{X})\big]\\ &\quad=\int_{\mathbb{X}^n}\phi_0(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}-\int_{\mathbb{X}^n}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\quad=\int_{B_1}\underbrace{\phi_0(\boldsymbol{x})}_{=1}\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+\int_{B_2}\underbrace{\phi_0(\boldsymbol{x})}_{=\gamma}\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+\int_{B_3}\underbrace{\phi_0(\boldsymbol{x})}_{=0}\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\qquad-\int_{B_1}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}-\int_{B_2}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}-\int_{B_3}\phi(\boldsymbol{x})\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\quad=\int_{B_1}\{1-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}_{>c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,(\boldsymbol{x}\in B_1)}+\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}}_{=c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,(\boldsymbol{x}\in B_2)}\\ &\qquad+\int_{B_3}\{-\phi(\boldsymbol{x})\}\underbrace{\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_1}(\boldsymbol{x})}_{<c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,(\boldsymbol{x}\in B_3)}\,\mathrm{d}\boldsymbol{x}\\ &\quad\ge\int_{B_1}\{1-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\qquad+\int_{B_3}\{-\phi(\boldsymbol{x})\}c\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &\quad=c\int_{B_1}\{1-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}+c\int_{B_2}\{\gamma-\phi(\boldsymbol{x})\}\mathsf{p}^{\boldsymbol{X}}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

$$
\begin{aligned}
&\quad + c\int_{B_3}\{-\phi(\boldsymbol{x})\}\mathsf{p}_{\boldsymbol{\theta}_0}^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= c\int_{\mathbb{X}^n}\{\phi_0(\boldsymbol{x})-\phi(\boldsymbol{x})\}\mathsf{p}_{\boldsymbol{\theta}_0}^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= c\Big\{\mathsf{E}_{\theta_0}[\phi_0(\boldsymbol{X})]-\mathsf{E}_{\theta_0}[\phi(\boldsymbol{X})]\Big\}\qquad(\because (8.5))\\
&= c\Big\{\alpha-\mathsf{E}_{\boldsymbol{\theta}_0}\big[\phi(\boldsymbol{X})\big]\Big\}\geq 0\qquad(\because (8.7)\text{ より})
\end{aligned}
$$

を得る. したがって

$$\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi_0(\boldsymbol{X})\big]\geq\mathsf{E}_{\boldsymbol{\theta}_1}\big[\phi(\boldsymbol{X})\big]$$

となるので, ϕ_0 は有意水準 α の m.p. 検定となる. □

例 8.5. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, \sigma^2)$ で $\sigma^2\,(\sigma>0)$ は既知とする. このとき検定問題

$$H_0:\theta=\theta_0\quad\text{vs.}\quad H_1:\theta=\theta_1\qquad(\theta_1>\theta_0)$$

に対する m.p. 検定を求める. まず $\boldsymbol{X}=(X_1, X_2, \ldots, X_n)^\top$ の同時 p.d.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n(x_j-\theta)^2\right],\qquad \boldsymbol{x}=(x_1, x_2, \ldots, x_n)^\top$$

で与えられることに注意をする. 簡単な計算から

$$
\begin{aligned}
\log\left[\frac{\mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x})}\right]&=-\frac{1}{2\sigma^2}\left[\sum_{j=1}^n(x_j-\theta_1)^2-\sum_{j=1}^n(x_j-\theta_0)^2\right]\\
&=\frac{n(\theta_1-\theta_0)}{\sigma^2}\left(\overline{x}_n-\frac{\theta_0+\theta_1}{2}\right)
\end{aligned}
\tag{8.8}
$$

となる. ただし

$$\overline{x}_n=\frac{1}{n}\sum_{j=1}^n x_j$$

である. ここで

$$
\begin{aligned}
\mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x})>c\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x})&\Leftrightarrow\log\left[\frac{\mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x})}\right]>\log c\\
&\Leftrightarrow\overline{x}_n>c'\qquad(\because (8.8)\text{ より})
\end{aligned}
$$

である. $\boldsymbol{X}$ は連続型確率変数なので, $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j = c'$ である確率は 0 となるので, m.p. 検定の形は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & (\overline{x}_n > c') \\ 0 & (\overline{x}_n \le c') \end{cases}$$

となる. 定数 c' は

$$\alpha = \mathsf{Pr}_{\theta_0}\{\overline{X}_n > c'\} \tag{8.9}$$

から定まる. ただし, $B \in \mathcal{B}(\mathbb{R}^n)$ に対して

$$\mathsf{Pr}_\theta(\boldsymbol{X} \in B) = \mathsf{E}_\theta[\mathbb{1}_B(\boldsymbol{X})] = \int_B \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad (\theta \in \Theta)$$

と定めた. (8.9) は

$$\alpha = \mathsf{Pr}_{\theta_0}\left\{\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\sigma} > \frac{\sqrt{n}(c' - \theta_0)}{\sigma}\right\} \tag{8.10}$$

と書き直せ, $\theta = \theta_0$ のもとで $\sqrt{n}(\overline{X}_n - \theta_0)/\sigma \sim \mathsf{N}(0,\,1)$ であるので

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x$$

とすると (8.10) は

$$\alpha = 1 - \Phi\left[\frac{\sqrt{n}(c' - \theta_0)}{\sigma}\right]$$

となる. 標準正規分布の上側 $100 \times \alpha\%$ を z_α とすると

$$\frac{\sqrt{n}(c' - \theta_0)}{\sigma} = z_\alpha \;\Leftrightarrow\; c' = \theta_0 + \frac{z_\alpha \sigma}{\sqrt{n}}$$

を得る. よって有意水準 α の m.p. 検定は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & \left(\overline{x}_n > \theta_0 + \dfrac{z_\alpha\sigma}{\sqrt{n}}\right) \\ 0 & \left(\overline{x}_n \le \theta_0 + \dfrac{z_\alpha\sigma}{\sqrt{n}}\right) \end{cases}$$

となる. 次に ϕ_0 の検出力は次のようになる.

$$\begin{aligned} \beta_{\phi_0}(\theta_1) &= \mathsf{E}_{\theta_1}\big[\phi_0(\boldsymbol{X})\big] = \mathsf{Pr}_{\theta_1}\left\{\overline{X}_n > \theta_0 + \frac{z_\alpha\sigma}{\sqrt{n}}\right\} \\ &= \mathsf{Pr}_{\theta_1}\left\{\frac{\sqrt{n}\big(\overline{X}_n - \theta_1\big)}{\sigma} > z_\alpha - \frac{\sqrt{n}\big(\theta_1 - \theta_0\big)}{\sigma}\right\} \end{aligned} \tag{8.11}$$

となる. $\theta = \theta_1$ のとき $\sqrt{n}\big(\overline{X}_n - \theta_1\big)/\sigma \sim \mathsf{N}(0,\,1)$ なので, (8.11) より仮説間の平均の差 $\theta_1 - \theta_0 (> 0)$ が大きいほど検出力は大きくなる. また標本 n が大きくなっても検出力が大きくなることがわかる. □

例 8.6. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. このとき検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1 \qquad (\theta_1 > \theta_0) \tag{8.12}$$

に対する m.p. 検定を求める. まず $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ の同時 p.m.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=1}^n \theta^{x_j}(1-\theta)^{1-x_j} \qquad \left(\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top\right) \tag{8.13}$$

で与えられるので

$$\log \frac{\mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x})} = \left\{\sum_{j=1}^n x_j\right\} \log\left\{\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}\right\} + n\log\left\{\frac{1-\theta_1}{1-\theta_0}\right\}$$

となる. $\theta_1 > \theta_0$ としたので, $\dfrac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0} > 1$ であることに注意すると

$$\begin{aligned} \mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x}) > c\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x}) &\Leftrightarrow \log\left[\frac{\mathsf{p}_{\theta_1}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta_0}^{\boldsymbol{X}}(\boldsymbol{x})}\right] > c' \\ &\Leftrightarrow \sum_{j=1}^n x_j > c'' \end{aligned}$$

と書きかえることができる. よって有意水準 α の m.p. 検定は

$$\phi_0(\boldsymbol{x}) = \begin{cases} 1 & \left(\sum_{j=1}^n x_j > c''\right) \\ \gamma & \left(\sum_{j=1}^n x_j = c''\right) \\ 0 & \left(\sum_{j=1}^n x_j < c''\right) \end{cases} \tag{8.14}$$

の形になる. ここで γ と c'' を

$$\begin{aligned} \alpha &= \mathsf{E}_{\theta_0}\left[\phi_0(\boldsymbol{X})\right] \\ &= \mathsf{Pr}_{\theta_0}\left\{\sum_{j=1}^n X_j > c''\right\} + \gamma\mathsf{Pr}_{\theta_0}\left\{\sum_{j=1}^n X_j = c''\right\} \end{aligned} \tag{8.15}$$

から定まる. $\sum_{j=1}^n X_j$ は $\theta = \theta_0$ のとき二項分布 $\mathsf{Bino}(n, \theta_0)$ に従うので, (8.15) は

$$\alpha = \sum_{j=c''+1}^n \binom{n}{j} \theta_0^j(1-\theta_0)^{n-j} + \gamma \binom{n}{c''} \theta_0^{c''}(1-\theta_\theta)^{n-c''}$$

となる. まずは c'' を

$$\sum_{j=c''+1}^n \binom{n}{j} \theta_0^j(1-\theta_0)^{n-j} \le \alpha < \sum_{j=c''}^n \binom{n}{j} \theta_0^j(1-\theta_0)^{n-j}$$

をみたす整数を定める. これから c_0 と書くことにする. すると γ は

$$\gamma = \left[\alpha - \sum_{j=c_0+1}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) \theta_0^j (1-\theta_0)^{n-j}\right] \Big/ \left(\begin{array}{c} n \\ c_0 \end{array} \right) \theta_0^{c_0} (1-\theta_0)^{n-c_0}$$

で定められる. □

8.3 検定統計量の導出方法

$\mathbb{X}^n$ を標本空間とし, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n} (\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d)$ とする. 検定関数 $\phi : \mathbb{X}^n \to [0, 1]$ によって定まる検定方式は, 以下のように定まることがある. ある統計量 $S : \mathbb{X}^n \longrightarrow \mathbb{R}$ と定数 c が存在して

$$S(\boldsymbol{x}) \leq c \Rightarrow \phi(\boldsymbol{x}) = 1$$
$$S(\boldsymbol{x}) > c \Rightarrow \phi(\boldsymbol{x}) = 0$$

となる. この場合, $S(\boldsymbol{X})$ のことも検定統計量と呼ぶことにする.

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_{\boldsymbol{\theta}}^{\otimes n} (\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d)$ とする. ただし, $\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ は $\mathbb{R}^n$ 上の確率測度である. $\mathsf{P}_{\boldsymbol{\theta}}^{\otimes n}$ は同時 p.d.f. $\mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{x})$ をもつとする. また, 母数空間 Θ は Θ_0 と Θ_0 に分割されたとする. すなわち $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \varnothing$, $\Theta_0 \neq \varnothing$, $\Theta_1 \neq \varnothing$ である.

8.3.1 尤度比検定統計量

定義 8.7. 検定問題

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta} \in \Theta_1$$

を検定するための 尤度比検定統計量(likelihood ratio statistic=l.r. 統計量) は

$$\lambda(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}{\sup_{\boldsymbol{\theta} \in \Theta} \mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{X})}$$

で与えられる. このとき正の定数 C が存在して H_0 の棄却域が

$$W = \{\boldsymbol{x} \in \mathbb{X}^n;\, \lambda(\boldsymbol{x}) \leq C\}$$

で与えられる検定を 尤度比検定(likelihood ratio test=l.r.t.) という. すなわち

$$\phi(\boldsymbol{x}) = \mathbb{1}_W(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in W) \\ 0 & (\boldsymbol{x} \notin W) \end{cases}$$

となる.

注意 8.8. $\widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ を母数空間 Θ での $\boldsymbol{\theta}$ の最尤推定量とし, $\widehat{\boldsymbol{\theta}}_0(\boldsymbol{X})$ を母数空間を Θ_0 に制限したときの $\boldsymbol{\theta}$ の最尤推定量とする. このとき

$$\lambda(\boldsymbol{X}) = \frac{\mathsf{p}^{\boldsymbol{X}}_{\widehat{\boldsymbol{\theta}}_0(\boldsymbol{X})}(\boldsymbol{X})}{\mathsf{p}^{\boldsymbol{X}}_{\widehat{\boldsymbol{\theta}}(\boldsymbol{X})}(\boldsymbol{X})}$$

と表現できる.

定理 8.9. $\boldsymbol{X} \sim \mathsf{P}^{\otimes n}_{\boldsymbol{\theta}^*}\,(\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d)$ とする. Θ の次元を d, Θ_0 の次元を $r\,(r < d)$ とし, $\Theta_1 := \Theta \setminus \Theta_0$ とする. 検定問題

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta_1$$

に対する尤度比検定統計量を $\lambda(\boldsymbol{X})$ とする. このとき H_0 のもとで次が成り立つ.

$$-2\log\lambda(\boldsymbol{X}) \rightsquigarrow \chi^2_{d-r}$$

が成立する.

Proof. 節 12.4 の定理 12.12 で証明を与える. □

注意 8.10. 定理 8.9 の結果を用いると尤度比検定の棄却域は

$$W = \left\{\boldsymbol{x} \in \mathbb{X}^n;\, -2\log\lambda(\boldsymbol{x}) > \chi^2_{d-r,\,\alpha}\right\}$$

で与えられる. ただし $\chi^2_{d-r,\,\alpha}$ は自由度 $d-r$ の χ^2 分布の上側 $100\times\alpha\%$ 点である. したがって, 検定手続きは

$$\boldsymbol{x} \in W \Rightarrow H_0 \text{ は棄却}$$

となる. □

8.3.2 スコア検定

分布 $\mathsf{P}_{\boldsymbol{\theta}}$ の p.d.f./p.m.f. を $\mathsf{p}_{\boldsymbol{\theta}}$ と書いたとき, スコア関数を

$$\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}}\,(x) := \nabla_{\boldsymbol{\theta}} \log \mathsf{p}_{\boldsymbol{\theta}}(x)$$

で定める. ただし, $\nabla_{\boldsymbol{\theta}} = \left(\dfrac{\partial}{\partial\theta_1}, \dfrac{\partial}{\partial\theta_2}, \ldots, \dfrac{\partial}{\partial\theta_d}\right)^\top$ である. $X \sim \mathsf{P}_{\boldsymbol{\theta}}$ のとき, 適当な正則条件のもとで

$$\mathsf{E}_{\boldsymbol{\theta}}\left[\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}}\,(X)\right] = \mathbf{0}_d$$

となる. 実現値 $X = x$ に対するスコア関数の実現値 $\overset{\bullet}{\ell}_{\boldsymbol{\theta}_0}(x)$ が 0 からある程度離れているとき, $\boldsymbol{\theta}_0$ は真の母数でないことを示唆することになる.

つぎに,「ある程度 0 から離れている」を定量化する. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ をランダム標本とする. すると $\boldsymbol{X}$ の同時 p.d.f./p.m.f. は

$$\mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=1}^{n} \mathsf{p}_{\boldsymbol{\theta}}(x_j)$$

となる. このことから, 帰無仮説 $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ に対するスコア関数は

$$\nabla_{\boldsymbol{\theta}_0} \log \mathsf{p}_{\boldsymbol{\theta}}^{\boldsymbol{X}}(\boldsymbol{X}) = \sum_{j=1}^{n} \overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(X_j)$$

の形で表現できることがわかる. ここで, $\overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}$ はひとつのデータに対するスコア関数である. 帰無仮説 H_0 のもとでは, $\mathsf{E}_{\boldsymbol{\theta}_0}\big[\overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(X_j)\big] = \mathbf{0}_d\, (j = 1, 2, \ldots, n)$ である. したがって, 中心極限定理 4.24 から, 帰無仮説 H_0 のもと

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(X_j) \rightsquigarrow \mathsf{N}_d\big(\mathbf{0}_d, \mathcal{F}(\boldsymbol{\theta}_0)\big) \qquad (n \to \infty)$$

となる. ただし

$$\mathcal{F}(\boldsymbol{\theta}_0) = \mathsf{E}_{\boldsymbol{\theta}_0}\big[\overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(X_1)\, \overset{\bullet}{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}^{\top}(X_1)^{\top}\big]$$

である. よって, $\boldsymbol{S}_n := \sum_{j=1}^{n} \overset{\bullet}{\boldsymbol{\ell}}_n(X_j)$ とおき,

$$\frac{1}{n} \boldsymbol{S}_n^{\top} \mathcal{F}^{-1}(\boldsymbol{\theta}_0) \boldsymbol{S}_n > \chi_d^2(\alpha) \;\Rightarrow\; H_0 \text{ を棄却} \tag{8.16}$$

とする. ただし, $\chi_d^2(\alpha)$ は自由度 d の χ 自乗分布の上側 α 点である. (8.16) で与えられる検定方式を スコア検定 という.

次に, 複合帰無仮説に対するスコア検定統計量を導出しよう. 一般性を失くことなく, Θ_0 は

$$\Theta_0 = \big\{\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^d;\, \boldsymbol{\theta}_1 \in \mathbb{R}^r,\, \boldsymbol{\theta}_2 = \mathbf{0}_{d-r}\big\}$$

と表現されるとする. ただし, $0 < r < d$ である. この分割に対応させてスコア関数

$$\overset{\bullet}{\ell}_{\boldsymbol{\theta}}^{\top} = \big(\overset{\bullet}{\ell}_{\boldsymbol{\theta},1}^{\top}, \overset{\bullet}{\ell}_{\boldsymbol{\theta},2}^{\top}\big)$$

と分割する. H_0 のもとの $\boldsymbol{\theta}_0$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_0^\top = \big(\widehat{\boldsymbol{\theta}}_{0,1},\, \mathbf{0}_{d-r}^\top\big)$ は尤度方程式

$$\sum_{j=1}^{n} \overset{\bullet}{\ell}_{\widehat{\boldsymbol{\theta}}_{0,\,1}} (X_j) = \mathbf{0}_r$$

をみたす. このことから

$$\sum_{j=1}^{n} \overset{\bullet}{\ell}{}^{\top}_{\widehat{\boldsymbol{\theta}}_0} (X_j) = \left(\mathbf{0}_r^\top,\, \sum_{j=1}^{n} \overset{\bullet}{\ell}_{\widehat{\boldsymbol{\theta}}_{0,\,2}} (X_j) \right)$$

と表現できる. したがって, (8.16) は

$$\frac{1}{n} \left(\sum_{j=1}^{n} \overset{\bullet}{\ell}_{\widehat{\boldsymbol{\theta}}_{0,\,2}} (X_j) \right)^{\top} \left(\mathcal{F}^{-1}_{\widehat{\boldsymbol{\theta}}_0} \right)_{2,\,2} \left(\sum_{j=1}^{n} \overset{\bullet}{\ell}_{\widehat{\boldsymbol{\theta}}_{0,\,2}} (X_j) \right) > \chi^2_{d-r}(\alpha)$$

となる. ここで, $d \times d$ の正値対称行列 $\mathbf{A}$ に対して, $(\mathbf{A}^{-1})_{2,\,2}$ は $\mathbf{A}^{-1}$ の $(d-r) \times (d-r)$ の右下ブロック行列である.

8.3.3　Wald 検定

$\widehat{\boldsymbol{\theta}}$ をフルモデルの最尤推定量とし, $\widehat{\boldsymbol{\theta}}_0$ を帰無仮説 H_0 のもとでの最尤推定量としたとき, 帰無仮説 H_0 のもとで

$$n\big(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_0\big)^\top \mathcal{F}_{\widehat{\boldsymbol{\theta}}_0} \big(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_0\big) > \chi^2_{d-r}(\alpha) \;\Rightarrow\; H_0 \text{ を棄却}$$

とする検定方式を **Wald 検定** という.

8.4　検定の一様最適性

8.4.1　単調尤度比

$n \in \mathbb{N}$ とし, $\mathbb{X}^n$ を標本空間とする. $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ をランダム標本とし, その同時 p.d.f./p.m.f. を $\mathsf{p}_\theta^{\boldsymbol{X}}$ とする. ただし, $\theta \in \Theta \subset \mathbb{R}$ である. 以下では, 集合 $\big\{\boldsymbol{x} \in \mathbb{X}^n;\, \mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) > 0\big\}$ は θ に依存しないと仮定する.

定義 8.11. 関数 $U : \mathbb{X}^n \longrightarrow \mathbb{R}$ があって, $\theta < \theta'$ なる任意の $\theta,\, \theta' \in \Theta$ に対して

$$\frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})}$$

は $U(\boldsymbol{x})$ の関数であり, $U(\boldsymbol{x})$ に関して単調増加であるとき, すなわち, 任意の $\forall \boldsymbol{x}, \widetilde{\boldsymbol{x}} \in \mathbb{X}^n$ に対して

$$U(\boldsymbol{x}) < U(\widetilde{\boldsymbol{x}}) \Rightarrow \frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} \leq \frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\widetilde{\boldsymbol{x}})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\widetilde{\boldsymbol{x}})}$$

が成り立つとき, $\{\mathsf{p}_\theta;\, \theta \in \Theta\}$ は U に関して **単調尤度比**(monotone likelihood ratio=m.l.r.) を持つという. U に関して強い意味の単調であるとき, U に関して強い意味での単調尤度比を持つという.

定理 8.12. 標本空間 $\mathbb{X}^n$ 上の確率変数列 $X_1, X_2, \ldots, X_n$ は指数型分布族

$$\mathsf{p}_\theta(x) = h(x) \exp\bigl\{A(\theta)T(x) - \mathsf{K}^\vee(\theta)\bigr\} \mathbb{1}_{\mathbb{X}}(x)$$

からのランダム標本とする. ただし, $\theta \in \Theta \subset \mathbb{R}$ で, $A : \Theta \longrightarrow [0, \infty)$, $T : \mathbb{X} \longrightarrow [0, \infty)$, $\mathsf{K}^\vee : \Theta \longrightarrow [0, \infty)$ である. さらに, A は θ の単調増加関数とし, $U(\boldsymbol{x}) = \sum_{j=1}^n T(x_j)$ ($\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とおく. このとき, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f./p.m.f. の分布族 $\{\mathsf{p}_\theta^{\boldsymbol{X}};\, \theta \in \Theta\}$ は U に関して単調尤度比を持つ.

Proof. まず

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{j=1}^n \mathsf{p}_\theta(x_j) = \left(\prod_{j=1}^n h(x_j)\right) \exp\left\{A(\theta) \sum_{j=1}^n T(x_j) - n\mathsf{K}^\vee(\theta)\right\}$$

と書けることに注意する. $\theta < \theta'$ とすると

$$\frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} = \exp\left\{-n\bigl\{\mathsf{K}^\vee(\theta') - \mathsf{K}^\vee(\theta)\bigr\}\right\} \exp\left\{\bigl(A(\theta') - A(\theta)\bigr) \sum_{j=1}^n T(x_j)\right\}$$

となる. A は単調増加と仮定しているので, $A(\theta') - A(\theta) \geq 0$ である. よって, $\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})/\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})$ は $\sum_{j=1}^n T(x_j)$ に関して単調増加であることがわかるので, 定理の主張は示された. □

定理 8.13. 確率ベクトル $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f/p.m.f. を $\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x})$ ($\theta \in \Theta \subset \mathbb{R}$) とする. 分布族 $\{\mathsf{p}_\theta^{\boldsymbol{X}};\, \theta \in \Theta\}$ は U に関して強い意味で単調尤度比を持つとする. $\theta_0 \in \Theta$ を定めて, $\Theta_0 = \Theta \cap (-\infty, \theta_0]$ とする. 検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

に対する有意水準 α ($0 < \alpha < 1$) の検定関数

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & (U(\boldsymbol{x}) > c) \\ \gamma & (U(\boldsymbol{x}) = c) \\ 0 & (U(\boldsymbol{x}) < c) \end{cases}$$

は u.m.p. 検定となる. ただし, c, γ は $\mathsf{E}_{\theta_0}\bigl[\phi^*(\boldsymbol{X})\bigr] = \alpha$ から定める定数である.

Proof. $\theta' \in \Theta \setminus \Theta_0$ とする. まず, 検定問題

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta'$$

を考える. この場合, 帰無仮説 H_0 と対立仮説 H_1 はともに単純なので, 定理 8.4 より

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) > c_0 \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \\ \gamma_0 & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) = c_0 \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \\ 0 & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) < c_0 \mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \end{cases}$$

は有意水準 α の m.p. 検定である. ただし, c_0 と γ_0 は

$$\mathsf{E}_{\theta_0}\big[\phi^*(\boldsymbol{X})\big] = \alpha$$

となるように定められた定数である. 仮定から, 強い意味での単調増加関数 $\tau : \mathbb{R} \longrightarrow \mathbb{R}$ が存在して

$$\frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} = \tau\big(U(\boldsymbol{x})\big)$$

と書ける. このとき

$$\frac{\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x})}{\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})} > c_0 \;\Leftrightarrow\; \tau\big(U(\boldsymbol{x})\big) > c_0 \;\Leftrightarrow\; U(\boldsymbol{x}) > \tau^{-1}(c_0)$$

が成立する. また, 上の同値関係は $>$ を $=$ と $<$ としても成り立つ. ここで $c := \tau^{-1}(c_0)$ と $\gamma_0 = \gamma$ とおくと

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & (U(\boldsymbol{x}) > c) \\ \gamma & (U(\boldsymbol{x}) = c) \\ 0 & (U(\boldsymbol{x}) < c) \end{cases}$$

と表現できる. さらに

$$\begin{aligned} \alpha &= \mathsf{E}_{\theta_0}\big[\phi^*(\boldsymbol{X})\big] \\ &= \mathsf{Pr}_{\theta_0}\big(\mathsf{p}_{\theta'}^{\boldsymbol{X}} > c_0 \mathsf{p}_{\theta}^{\boldsymbol{X}}\big) + \gamma_0 \mathsf{Pr}_{\theta_0}\big(\mathsf{p}_{\theta'}^{\boldsymbol{X}} = c_0 \mathsf{p}_{\theta}^{\boldsymbol{X}}\big) \\ &= \mathsf{Pr}_{\theta_0}\big(U(\boldsymbol{X}) > c\big) + \gamma \mathsf{Pr}_{\theta_0}\big(U(\boldsymbol{X}) = c\big) \end{aligned}$$

から, c と γ は θ_0 にのみ依存する. すなわち, θ' には無関係である. したがって, あらゆる $\theta' \in \Theta \setminus \Theta_0$ に対する検定問題

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta'$$

に対して, 検定関数 ϕ^* は m.p. 検定となる. すなわち, 任意の検定関数 ϕ に対して

$$\mathsf{E}_{\theta_0}\big[\phi(\boldsymbol{X})\big] \leq \alpha \;\Rightarrow\; \mathsf{E}_{\theta'}\big[\phi^*(\boldsymbol{X})\big] \geq \mathsf{E}_{\theta'}\big[\phi(\boldsymbol{X})\big] \qquad (\theta' \in \Theta \setminus \Theta_0)$$

が成り立つ.

以上のことから, すべての $\theta \in \Theta_0$ に対して

$$\mathsf{E}_\theta\big[\phi^*(\boldsymbol{X})\big] \le \alpha \tag{8.17}$$

が成り立つことを示せば, 定理の主張は証明される. ここで, $\theta' \in \Theta_0 \setminus \{\theta_0\}$ に対して

$$\alpha(\theta') := \mathsf{E}_{\theta'}\big[\phi^*(\boldsymbol{X})\big] \tag{8.18}$$

とおく. 検定問題

$$H_0 : \theta = \theta' \quad \text{vs.} \quad H_1 : \theta = \theta_0$$

に対する有意水準 $\alpha(\theta')$ の m.p. 検定は

$$\widetilde{\phi}(\boldsymbol{x}) = \begin{cases} 1 & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) > \widetilde{c}\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \\ \widetilde{\gamma} & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) = \widetilde{c}\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \\ 0 & \big(\mathsf{p}_{\theta'}^{\boldsymbol{X}}(\boldsymbol{x}) < \widetilde{c}\mathsf{p}_{\theta}^{\boldsymbol{X}}(\boldsymbol{x})\big) \end{cases}$$

で与えられることが定理 8.4 からわかる. ただし, $\widetilde{c}$ と $\widetilde{\gamma}$ は

$$\mathsf{E}_{\theta'}\big[\widetilde{\phi}(\boldsymbol{X})\big] = \alpha(\theta')$$

となるように定めた. さらに, これは

$$\widetilde{\phi}(\boldsymbol{x}) = \begin{cases} 1 & \big(U(\boldsymbol{x}) > c'\big) \\ \widetilde{\gamma} & \big(U(\boldsymbol{x}) = c'\big) \\ 0 & \big(U(\boldsymbol{x}) < c'\big) \end{cases}$$

とも表すことができる. ただし, $c' = \tau^{-1}(\widetilde{x})$ である. $\widetilde{\phi}$ の定め方から

$$\alpha(\theta') = \mathsf{E}_{\theta'}\big[\widetilde{\phi}(\boldsymbol{X})\big] = \mathsf{Pr}_{\theta'}\big(U(\boldsymbol{X}) > c'\big) + \widetilde{\gamma}\mathsf{Pr}_{\theta'}\big(U(\boldsymbol{X}) = c'\big) \tag{8.19}$$

が成立している. 一方, (8.18) から

$$\alpha(\theta') = \mathsf{E}_{\theta'}\big[\phi^*(\boldsymbol{X})\big] = \mathsf{Pr}_{\theta'}\big(U(\boldsymbol{X}) > c\big) + \gamma\mathsf{Pr}_{\theta'}\big(U(\boldsymbol{X}) = c\big) \tag{8.20}$$

が成立している. (8.19) と (8.20) を比較すると

$$c = c', \quad \gamma = \widetilde{\gamma}$$

と c' と $\widetilde{\gamma}$ を取ることができる. したがって, $\widetilde{\phi}(\boldsymbol{X}) = \phi^*(\boldsymbol{X})$ である. いま, $\alpha(\theta') \equiv \phi(\boldsymbol{x})$ なる検定関数を考えると

$$\mathsf{E}_{\theta'}\big[\phi(\boldsymbol{X})\big] = \mathsf{E}_{\theta'}\big[\alpha(\theta')\big] = \alpha(\theta')$$

であるから, 検定関数 ϕ は検定問題

$$H_0 : \theta = \theta' \quad \text{vs.} \quad H_1 : \theta = \theta_0$$

の有意水準 $\alpha(\theta')$ 検定である. 検定関数 $\widetilde{\phi}$ はこの検定問題に対する m.p. 検定であることと (8.18) から, $\forall \theta' \in \Theta_0 \setminus \{\theta_0\}$ に対して

$$\alpha = \mathsf{E}_{\theta_0}\big[\phi^*(\boldsymbol{X})\big] = \mathsf{E}_{\theta_0}\big[\widetilde{\phi}(\boldsymbol{X})\big] \geq \mathsf{E}_{\theta_0}\big[\phi(\boldsymbol{X})\big] = \alpha(\theta') = \mathsf{E}_{\theta'}\big[\phi^*(\boldsymbol{X})\big]$$

が成立することがわかる. 以上から (8.17) が確認できた. 以上の議論から, $\forall \theta \in \Theta_0$ に対して

$$\mathsf{E}_{\theta}\big[\phi^*(\boldsymbol{X})\big] \leq \alpha$$

が成り立つので, ϕ^* は有意水準 α の検定となる. さらに, 任意の有意水準 ϕ の検定と $\forall \theta \in \Theta \setminus \Theta_0$ に対して

$$\mathsf{E}_{\theta}\big[\phi^*(\boldsymbol{X})\big] \geq \mathsf{E}_{\theta}\big[\phi(\boldsymbol{X})\big]$$

となることが示せたので, 検定関数 ϕ^* は u.m.p. 検定であることが示せた. □

例 8.14. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. $\theta_0 \in (0, 1)$ を固定して, 検定問題

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0 \tag{8.21}$$

に対する有意水準 $\alpha\,(0 < \alpha < 1)$ の u.m.p. 検定を求めよう. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.m.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \theta^{\sum_{j=1}^n x_j}(1-\theta)^{n-\sum_{j=1}^n x_j} \prod_{j=1}^n \mathbb{1}_{\{0,1\}}(x_j) \qquad (\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

であるから

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \exp\left\{ n\log(1-\theta) + \sum_{j=1}^n x_j \log\left(\frac{\theta}{1-\theta}\right)\right\} \prod_{j=1}^n \mathbb{1}_{\{0,1\}}(x_j)$$

と書き直せるので, 1 母数指数型分布族で

$$A(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$$

は θ の強い意味での単調増加関数である. したがって, 定理 8.12 から $U(\boldsymbol{x}) := \sum_{j=1}^n x_j$ に関して強い意味での単調尤度比を持つ. したがって, 検定問題 (8.21) において, 有意水準 α の u.m.p. 検定が存在し

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & (\sum_{j=1}^n x_j > c) \\ \gamma & (\sum_{j=1}^n x_j = c) \\ 0 & (\sum_{j=1}^n x_j < c) \end{cases}$$

が u.m.p. 検定となる. ただし, c と γ は

$$\alpha = \mathsf{E}_{\theta_0}\big[\phi^*(\boldsymbol{X})\big]$$

から定まる定数である. □

例 8.15. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, \theta)\,(0 < \theta < \infty)$ とする. $\theta_0 \in (0, \infty)$ を固定して, 検定問題

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0 \tag{8.22}$$

に対する有意水準 $\alpha\,(0 < \alpha < 1)$ の u.m.p. 検定を求めよう. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時 p.d.f. は

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left\{-\frac{1}{2\theta}\sum_{j=1}^n x_j^2\right\} \quad (\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$$

であるから, 1 母数指数型分布族で

$$A(\theta) = -\frac{1}{2\theta}$$

は θ の強い意味での単調増加関数である. したがって, 定理 8.12 から $U(\boldsymbol{x}) := \sum_{j=1}^n x_j^2$ に関して強い意味での単調尤度比を持つ. したがって, 検定問題 (8.22) において, 有意水準 α の u.m.p. 検定が存在し

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & (\sum_{j=1}^n x_j^2 > c) \\ \gamma & (\sum_{j=1}^n x_j^2 = c) \\ 0 & (\sum_{j=1}^n x_j^2 < c) \end{cases}$$

が u.m.p. 検定となる. ただし, c と γ は

$$\alpha = \mathsf{E}_{\theta_0}\big[\phi^*(\boldsymbol{X})\big]$$

から定まる定数である. □

8.4.2 不偏検定

節 8.4.1 では, 1 次元の母数の片側検定問題おいて, u.m.p. 検定が構成できる場合があることを示した. しかし, 単調尤度比の手法は両側検定に対しては適用できない. そこで, 両側検定問題の場合において, 自然と思われる検定方式の正当化のための概念を導入する.

まず, 簡単な例をみてみよう.

例 8.16. $\mu^* \in \mathbb{R}$ とし, $X \sim \mathsf{N}(\mu, 1)$ とする. 観測 X に基づいて, 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定問題

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu \neq 0 \tag{8.23}$$

を考える.

u.m.p. 検定が存在しないことを確認しよう. そのために, $\mu_1 \in \mathbb{R} \setminus \{0\}$ を固定して, 検定問題

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu = \mu_1$$

を考える. 帰無仮説も対立仮説も単純であるので, 定理 8.4 を用いて m.p. 検定が構成できることに注意する. $\mu_1 > 0$ の場合には

$$\phi_1(x) = \begin{cases} 1 & (X > c') \\ 0 & (X \leq c') \end{cases}$$

は m.p. 検定となる. ただし, c' は $\mathsf{E}_{\mu=0}\big[\phi_1(X)\big] = \alpha$ となるように定めた定数である. 一方, $\mu_1 < 0$ の場合には

$$\phi_2(x) = \begin{cases} 1 & (X < -c') \\ 0 & (X \geq -c') \end{cases}$$

は m.p. 検定となる. u.m.p. 検定はどんな対立仮説 $H_1 : \mu = \mu_1$ に対しても同時に m.p. 検定でなければならない. しかし, $\mu_1 > 0$ と $\mu_1 < 0$ の場合では, m.p. 検定の形が異なるので, u.m.p. 検定は存在しないことがわかる.

検定問題 (8.23) において, 自然な検定関数は

$$\phi_3(x) = \begin{cases} 1 & (|X| > c) \\ 0 & (|X| \leq c) \end{cases}$$

である. ただし, c は $\mathsf{E}_{\mu=0}\big[\phi_3(X)\big] = \alpha$ をみたすように定めた定数である. この検定関数は, 以下で定める不偏検定の族の中で m.p. となることがわかる. □

定義 8.17. Θ を母数空間とし, $\Theta_0, \Theta_1 \subset \Theta$ は $\Theta_0 \neq \emptyset$, $\Theta_1 \neq \emptyset$, $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta = \Theta_0 \cup \Theta_1$ をみたすとする. 母数 $\theta \in \Theta$ に対する有意水準 $\alpha\,(0 < \alpha < 1)$ の検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1$$

を考える. さらに, X を母数 θ の母集団分布から観測とする. 有意水準 α の検定関数 ϕ は 不偏(unbiased) であるとは

$$\beta_\phi(\theta) = \mathsf{E}_\theta\big[\phi(X)\big] \geq \alpha \qquad (\forall \theta \in \Theta_1)$$

が成り立つことである.

定義 8.18. 定義 8.17 の設定を踏襲する. 有意水準 α の不偏検定の族の中で, すべての対立仮説について検出力を最大にする検定関数 ϕ^* が存在すれば, ϕ^* を 一様最強力不偏検定(uniformly most powerful unbiased =u.m.p.u.) という. すなわち, 有意水準 α の任意の不偏検定関数 ϕ に対して

$$\beta_{\phi^*}(\theta) \geq \beta_{\phi}(\beta) \qquad (\forall \theta \in \Theta_1)$$

が成り立つことである.

定理 8.19. 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定問題において, u.m.p. 検定は u.m.p.u. 検定である.

Proof. 定義 8.17 の設定を踏襲して, 証明を与える. 有意水準 α の u.m.p. 検定関数を ϕ^* とおく. すると

$$\mathsf{E}_\theta\big[\phi^*(X)\big] \leq \alpha \qquad (\theta \in \Theta_0)$$

である.

いま, $\widetilde{\phi}(x) = \alpha$ なる検定関数を考える. すると

$$\mathsf{E}_\theta\big[\widetilde{\phi}(X)\big] = \mathsf{E}_\theta\big[\alpha\big] = \alpha \qquad (\theta \in \Theta)$$

であることがわかる. 一方, ϕ^* は u.m.p. 検定関数なので

$$\mathsf{E}_\theta\big[\phi^*(X)\big] \geq \mathsf{E}_\theta\big[\widetilde{\phi}(X)\big] = \alpha \qquad (\theta \in \Theta_1)$$

である. したがって, ϕ^* は不偏検定である. ϕ^* は u.m.p. なので, 有意水準 α の任意の検定関数 ϕ に対して

$$\beta_{\phi^*}(\theta) \geq \beta_{\phi}(\theta) \qquad (\forall \theta \in \Theta_1)$$

が成立する. したがって, ϕ^* は u.m.p.u. 検定である. □

定理 8.20. 標本空間 $\mathbb{X}^n$ 上の確率変数列 $X_1, X_2, \ldots, X_n$ は指数型分布族

$$\mathsf{p}_\theta(x) = h(x) \exp\big\{A(\theta)T(x) - \mathsf{K}^\vee(\theta)\big\} \mathbb{1}_{\mathbb{X}}(x)$$

からのランダム標本とする. ただし, $\theta \in \Theta \subset \mathbb{R}$ で, $A : \Theta \longrightarrow [0, \infty)$, $T : \mathbb{X} \longrightarrow [0, \infty)$, $\mathsf{K}^\vee : \Theta \longrightarrow [0, \infty)$ である. $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ に対して,

$$\Theta_0 = \Theta \cap [\theta_1, \theta_2]$$

と定め, 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定問題

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

を考える. さらに, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とおく. 検定関数 ϕ^* を

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & \left(\sum_{j=1}^n T(x_j) < c_1 \text{ または } \sum_{j=1}^n T(x_j) > c_2\right) \\ \gamma & \left(\sum_{j=1}^n T(x_j) = c_1 \text{ または } \sum_{j=1}^n T(x_j) = c_2\right) \\ 0 & (\text{その他の場合}) \end{cases}$$

で定めたとき, 検定関数 ϕ^* は u.m.p.u. 検定となる. ただし, γ, c_1, c_2 は $\mathsf{E}_{\theta_j}\left[\phi^*(\boldsymbol{X})\right] = \alpha\,(j = 1, 2)$ から定まる定数である.

Proof. 草間 (1975, pp.107-108) では証明は省略されている. Lehmann-Romano (2022, pp.126-128) を参照すること. □

8.5　多重比較と FDR

8.6　区間推定の考え方

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \sim \mathsf{P}_\theta^{\otimes n}\,(\theta \in \Theta \subset R)$ とする. $0 < \alpha < 1$ を固定する. 母数 θ に依存しない区間 $\left[L(\boldsymbol{X}), U(\boldsymbol{X})\right] \subset \Theta$ が $\forall\theta \in \Theta$ に対して

$$\mathsf{Pr}_\theta\left\{L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})\right\} \geq 1 - \alpha \tag{8.24}$$

をみたすとき区間 $\left[L(\boldsymbol{X}), U(\boldsymbol{X})\right]$ を **信頼係数** $(1-\alpha)$ **の** θ **の信頼区間** (confidence interval) という. $L(\boldsymbol{X})$, $U(\boldsymbol{X})$ を **信頼限界**(confident limits) という. 通常 α として 0.05, 0.01, 0.1 等が用いられる. (8.24) の関係式は, たとえば 100 組の実現値を発生させると 100α 回程度は信頼区間に真の母数 θ は含まれないと考える.

例 8.21. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\theta, \sigma^2)$ とする. ただし $\theta \in \mathbb{R}$ で $\sigma^2\,(\sigma > 0)$ の値は既知である. このとき標本平均 $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j$ は

$$\overline{X}_n \sim \mathsf{N}\left(\theta, \frac{\sigma^2}{n}\right)$$

となる. このことより

$$S_\theta(\boldsymbol{X}) = \frac{\sqrt{n}\left(\overline{X}_n - \theta\right)}{\sigma} \sim \mathsf{N}(0, 1)$$

となる. ここで $z_{\alpha/2}$ を標準正規分布 $\mathsf{N}(0,\,1)$ の上側 $100(\alpha/2)\%$ 点とると

$$\begin{aligned} 1-\alpha &= \mathsf{Pr}_\theta\left\{-z_{\alpha/2} \le S_\theta(\boldsymbol{X}) \le z_{\alpha/2}\right\} \\ &= \mathsf{Pr}_\theta\left\{\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \theta \le \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right\} \end{aligned}$$

なることがわかる. したがって

$$\big[L(\boldsymbol{X}),\, U(\boldsymbol{X})\big] = \left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\, \overline{X}_n + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right]$$

は信頼係数 $(1-\alpha)$ の θ の信頼区間となる. □

8.7 信頼区間の構成法

8.7.1 検定方式の反転

$\Theta \subset \mathbb{R}$ とし, $\theta_0 \in \Theta$ を取る[1]. 標本 $\boldsymbol{X} \in \mathbb{R}^n$ に基づき, 検定問題

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

を考える. 有意水準 $\alpha\,(0 < \alpha < 1)$ の検定の受容域を $A(\theta_0) \subset \mathbb{R}^n$ とおく. すなわち

$$\mathsf{Pr}_{\theta_0}\big\{\boldsymbol{X} \in A(\theta_0)\big\} \ge 1-\alpha$$

が成り立っている. そこで $\boldsymbol{x} \in A(\theta_0)$ を θ_0 に関して解くことによって

$$C(\boldsymbol{x}) = \{\theta \in \Theta;\, \boldsymbol{x} \in A(\theta)\} \subset \mathbb{R}$$

が得られる. 一般に $C(\boldsymbol{X})$ は連結区間になるという保証はない. しかし, $C(\boldsymbol{X})$ が連結区間となるときには

$$\mathsf{Pr}_\theta\big\{\theta \in C(\boldsymbol{X})\big\} \ge 1-\alpha$$

となるので, $C(\boldsymbol{X})$ は信頼係数 $(1-\alpha)$ の信頼区間となる.

例 8.22. $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とする. $\theta_0 \in (0,\,1)$ と固定し, 検定問題

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

[1] θ_0 を θ^* と書くべきであろうが, 後の表記上の都合でこの記号を採用した.

を考える. この検定問題に対する尤度比検定統計量は

$$\lambda(\boldsymbol{X}) = \frac{\prod_{j=1}^{n} \theta_0^{X_j}(1-\theta_0)^{1-X_j}}{\prod_{j=1}^{n} \overline{X}_n^{X_j}(1-\overline{X}_n)^{1-X_j}} = \prod_{j=1}^{n}\left(\frac{\theta_0}{\overline{X}_n}\right)^{X_j}\left(\frac{1-\theta_0}{1-\overline{X}_n}\right)^{1-X_j}$$

となる. これより受容域 $A(\theta_0)$ は

$$A(\theta_0) = \Bigg\{ \boldsymbol{x} \in \{0,\,1\}^n; \\ -2\log\lambda(\boldsymbol{x}) = 2n\overline{x}_n \log\left(\frac{\overline{x}_n}{\theta_0}\right) + 2n(1-\overline{x}_n)\log\left(\frac{1-\overline{x}_n}{1-\theta_0}\right) \le \chi^2_{1,\,\alpha} \Bigg\}$$

となる. ただし $\overline{x}_n = n^{-1}\sum_{j=1}^{n} x_j$, $\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ である. よって

$$C(\boldsymbol{X}) := \left\{ \theta;\, \overline{X}_n \log\left(\frac{\overline{X}_n}{\theta}\right) + (1-\overline{X}_n)\log\left(\frac{1-\overline{X}_n}{1-\theta}\right) \le \frac{\chi^2_{1,\,\alpha}}{2n} \right\}$$

となる. ただし信頼限界 $L(\boldsymbol{X})$, $U(\boldsymbol{X})$ を陽に求めることはできない. □

8.7.2 枢軸量 (pivotal quantity)

一般に $Q(\boldsymbol{X},\,\theta)$ の分布が θ に依存しないとき, $Q(\boldsymbol{X},\,\theta)$ を 枢軸量(pivotal quantity) という. このとき

$$\mathsf{Pr}_\theta\Big(a \le Q(\boldsymbol{X},\,\theta) \le b \Big) = 1-\alpha$$

をみたす a, b を定めて, $a \le Q(\boldsymbol{X},\,\theta) \le b$ を θ に関して解くことにより, 信頼係数 $(1-\alpha)$ の θ の信頼区間

$$C(\boldsymbol{X}) = \big\{ \theta \in \Theta;\, a \le Q(\boldsymbol{X},\,\theta) \le b \big\}$$

が得られる. もちろん $C(\boldsymbol{X})$ が連結区間になる保証は一般的にはないが, うまく連結区間になれば, 信頼区間として使用できる.

例 8.23. $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(\theta)\,(\theta > 0)$ とする. すなわち, p.d.f. は

$$\mathsf{p}_\theta(x) = \begin{cases} \theta e^{-\theta x} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる. すると $2\theta\sum_{j=1}^{n} X_j$ は自由度 $2n$ の χ^2 分布に従うので

$$Q(\boldsymbol{X},\,\theta) = 2\theta\sum_{j=1}^{n} X_j$$

とおく. よって $\chi^2_{2n,\,\alpha}$ を自由度 $2n$ の χ^2 分布の上側 $100\alpha\%$ 点とすると

$$\mathsf{Pr}_\theta\left(\chi^2_{2n,\,1-\alpha/2} \leq Q(\boldsymbol{X},\,\theta) \leq \chi^2_{2n,\,\alpha/2}\right) = 1-\alpha$$

となるので,

$$\big[L(\boldsymbol{X}),\,U(\boldsymbol{X})\big] = \left[\frac{\chi^2_{2n,\,1-\alpha/2}}{2\sum_{j=1}^n X_j},\,\frac{\chi^2_{2n,\,\alpha/2}}{2\sum_{j=1}^n X_j}\right]$$

は信頼係数 $(1-\alpha)$ の θ の信頼区間となる. □

問 8.1. $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Exp}(\theta)\,(\theta>0)$ のとき, $2\theta\sum_{j=1}^n X_j \sim \chi^2_{2n}$ となることを示せ.

8.8 章末注釈と参考文献

節 8.4 は $[50,\,\text{pp.}62-67]$ を借用した.

8.9 演習問題

演習問題 8.1. 確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上で定義された確率変数 $X_1,\,X_2,\,X_3$ は独立同一に母数 $\theta\,(0<\theta<1)$ の Bernoulli 分布 $\mathsf{Ber}(\theta)$ に従っているとする. 検定問題

$$\mathrm{H}_0:\,\theta=\frac{1}{2}\quad\text{vs.}\quad \mathrm{H}_1:\,\theta\neq\frac{1}{2}$$

の検定問題に対して, 検定方式

$$\begin{aligned} W_1 &= \{(x_1,\,x_2,\,x_3)\in S;\,\sum_{i=1}^3 x_i \in \{0,\,3\}\} \\ W_2 &= \{(x_1,\,x_2,\,x_3)\in S;\,\sum_{i=1}^3 x_i \in \{2,\,3\}\} \end{aligned}$$

を考える. ただし, $S=\{(x_1,\,x_2,\,x_3);\,x_i\in\{0,\,1\}\,(i=1,\,2,\,3)\}$ である. 検定方式 $W_1,\,W_2$ のサイズを求めよ. すなわち

$$\mathsf{Pr}_{\theta=1/2}((X_1,\,X_2,\,X_3)\in R_1)\quad\text{および}\quad \mathsf{Pr}_{\theta=1/2}((X_1,\,X_2,\,X_3)\in R_2)$$

の確率である.

記号 $\mathsf{p}_\theta(x)$ を $\mathsf{Ber}(\theta)$ の p.m.f. とする. Borel 集合 $A \subset \mathbb{R}^3$ に対して

$$\mathsf{Pr}_{\theta=1/2}\big((X_1, X_2, X_3) \in A\big) = \sum_{(x_1,x_2,x_3)\in A\cap S} \mathsf{p}_{1/2}\big(x_1\big)\mathsf{p}_{1/2}\big(x_2\big)\mathsf{p}_{1/2}\big(x_3\big)$$

と定めている.

演習問題 8.2. $\theta \in \mathbb{R}$ とし, 連続型確率変数 X は p.d.f.

$$\mathsf{p}_\theta(x) = \frac{1}{\pi\{1+(x-\theta)^2\}} \qquad (-\infty < x < \infty)$$

を持つとする. 仮説検定問題

$$\text{帰無仮説 } H_0 : \theta = 0 \quad \text{vs.} \quad \text{対立仮説 } H_1 : \theta = 1$$

を X に基づいて検定する. この検定問題に対して, 棄却域

$$W := \{x \in \mathbb{R} : 1 < x < 3\}$$

を考える. 以下の問いに答えよ. ただし, $\arctan 2 = 1.107$, $\arctan 3 = 1.249$, $\pi = 3.1416$ として計算せよ.
(1) 積分

$$\int_{-\infty}^{\infty} \mathsf{p}_\theta(x)\,\mathrm{d}x$$

を計算せよ.
(2) W で定まる検定のサイズ (第 1 種の誤りの確率) α の値を小数第 3 位まで求めよ.
(3) W で定まる検定の検出力 $1-\beta$ (ただし, β は第 2 種の誤りの確率) を小数第 3 位まで求めよ.
(4) 尤度比

$$\Lambda(x) = \frac{\mathsf{p}_1(x)}{\mathsf{p}_0(x)}$$

の $x = 0$ と $x = 1$ における値を求めよ.
(5) 不等式 $\Lambda(x) > 2$ をみたす領域を求めることにより, 棄却域 W で定まる検定は (2) で求めた α を有意水準とする検定の中で最強力検定となることを Neyman-Pearson の補題を用いて証明せよ.

演習問題 8.3. 連続型確率変数列 $X_1, X_2, \ldots, X_{16}$ は正規分布 $\mathsf{N}(\mu, 3^2)$ からのランダム標本とする. ただし, $-\infty < \mu < \infty$ で, これらの確率変数は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義されたものとする.

(1) 標本平均 $\overline{X}_{16} = (1/16)(X_1 + \cdots + X_{16})$ の期待値と分散を計算することにより, $\overline{X}_{16}$ の分布を述べよ.
(2) $a > 0$, b を定数とし, $Z = a\overline{X}_{16} + b$ としたとき, Z の分布が標準正規分布になるように定数 a, b を定めよ.
(3) 信頼係数 90% の μ の信頼区間を構成せよ.

演習問題 8.4. 連続型確率変数列 X_1, X_2, X_3 は正規母集団 $\mathsf{N}(\mu, 3)$ からの標本の大きさ 3 のランダム標本とする. 次の仮説検定問題を考える:

$$H_0 : \mu = 0 \quad \text{vs.} \quad H_1 : \mu = 2$$

このとき, 以下の問いに答えよ.
(1) H_0 と H_1 は単純仮説か複合仮説かを答えよ.
(2) $\overline{X}_3 = (1/3)(X_1 + X_2 + X_3)$ とする. $\bar{X}_3$ は 帰無仮説 $H_0 : \mu = 0$ のもとでどのようなものになるかを答えよ. 理由も述べること.
(3) H_0 の棄却域を

$$C(t) = \{(x_1, x_2, x_3) \in \mathsf{R}^3;\, x_1 + x_2 + x_3 > t\}$$

としたとき, $C(t)$ が有意水準 0.1 の検定の棄却域になるように t をひとつ定めよ. t を求めるときには小数第 3 位を四捨五入せよ.
(4) 棄却域 $C(t)$ の検出力を 対立仮説 $H_1 : \mu = 2$ のもとで求めよ. ただし, 解答は関数 $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) e^{-t^2/2}\, \mathrm{d}t$ および t を用いて表現せよ. すなわち, 具体的な値を計算しなくともよい.

演習問題 8.5. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 $X_1,\ X_2, \ldots, X_n$ は独立で同じ分布に従い, 平均 $\mathsf{E}[X_1] = \mu$, 分散 $\mathsf{Var}[X_1] = \sigma^2$ $(0 < \sigma < \infty)$, 4 次中心積率 $\mathsf{E}[(X_1 - \mu)^4] = \sigma^4 + 1$ とする. 任意の実数 a と $0 < p < 1$ に対して, 関数 $r(\cdot)$ と $z(\cdot)$ はそれぞれ

$$r(a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \qquad p = \int_{z(p)}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

をみたすものとする. このとき, 以下の各問いに答えよ. ただし, 中心極限定理は証明なしで用いてよい. また, (3), (4) は関数 $z(\cdot)$ または $r(\cdot)$ を用いて表せ.
(1) $Y_i = (X_i - \mu)^2$ $(i = 1, 2, \ldots, n)$ の平均 $\mathsf{E}[Y_i]$ および分散 $\mathsf{Var}[Y_i]$ を求めよ.
(2) $\overline{X} = \sum_{i=1}^n X_i/n$ とおく. このとき, 任意の正数 $\epsilon > 0$ に対して

$$\lim_{n\to\infty} \mathsf{Pr}(|\overline{X} - \mu| \geq \epsilon) = 0$$

が成り立つことを示せ.

(3) 次の値を求めよ.

$$\lim_{n\to\infty} \mathsf{Pr}\left(\sigma^2 - \frac{1}{\sqrt{n}} \leq \frac{(X_1-\mu)^2 + \cdots + (X_n-\mu)^2}{n} \leq \sigma^2 + \frac{1}{\sqrt{n}}\right)$$

(4) 平均 μ を既知, 分散 σ^2 は未知とする. $\overline{Y} = \sum_{i=1}^n Y_i/n$ とする. $\overline{Y}$ の実現値 $\overline{y}$ に基づく分散 σ^2 の 95% 近似信頼区間で有界なものをひとつ与えよ. ただし, 95% 近似信頼区間とは, 区間 $\big(\ell(\overline{y}), u(\overline{y})\big)$ で

$$\lim_{n\to\infty} \mathsf{Pr}\Big(\ell(\overline{Y}) < \sigma^2 < u(\overline{Y})\Big) \geq 0.95$$

を満たすものである.

第9章　Bayes的推測

9.1　Bayes 的推測の考え方

前節までの統計的手法は頻度論的手法 (frequentist methods) と呼ばれるものである. 頻度論的アプローチの考え方は以下のようにまとめられる.

【F1】確率は大量に観測されたデータの度数の極限と考える.

【F2】真の母数は未知だが, 固定した値であると考える. したがって真の母数は変動しないし, 真の母数に対する意味のある確率的な主張はない.

【F3】推測手法は大量にデータが観測されれば, うまく機能する保証があるように設計されている. たとえば信頼係数 0.95 の信頼区間は, 大量に観測されたデータの極限的な頻度において, 0.95 の信頼度が保証されている.

この考え方とは異なるアプローチの 1 つが Bayes 的推測である. Bayes 的アプローチは以下のような考え方に基づいている.

【B1】確率は極限的な頻度ではなく, 信頼の程度を表現するものであると考える. データはある確率変数の実現値と考えるだけでなく, それ以外の色々なものも確率変数の実現値と考える.

【B2】真の母数は固定された定数にも関わらず, 真の母数の確率的な主張をする.

【B3】真の母数に対して確率分布を想定し, 未知の母数の推測を行う. 点推定値や区間推定値もこの分布からの実現値と考える.

確率に主観的な見方を導入する Bayes 的アプローチには, 頻度論的立場の主流派 (古典的な) 統計学者からの大きな批判がある. しかし, 統計学の隣接分野である機械学習やデータマイニングの分野では Bayes 的なアプローチが広く採用されている.

哲学的な議論は脇において, Bayes 的推測がどのような形式で行われるかをこの節ではみていく.

記法に関する注意　母数 θ で添え字付けられた p.d.f./p.m.f. を $\mathsf{p}_\theta(\cdot)$ と前の章まで記してきた. この章では, 条件付きであることを強調するために, $\mathsf{p}(\cdot \mid \theta)$ のように記すことにする. □

9.2 Bayes 的推測手法

Bayes 的推測は以下のステップに従い行われる. 母数モデルを $\{\mathsf{P}_\theta;\, \theta \in \Theta\}$ とする.

1. 事前分布 (prior distribution) と呼ばれる母数 θ についての分布 $\mathsf{p}^{\Theta}(\theta)$ を想定する.

2. 母数 θ が与えられたときにデータの分布を表現する条件付きの分布 (頻度論的な立場では真の分布) $\mathsf{p}^{\boldsymbol{X}|\,\Theta}(\boldsymbol{x}|\,\theta)$ を想定する.

3. 観測されたデータ $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ に基づき, θ の信頼度をアップデートする. すなわち, **事後分布** (posterior distribution)　$\mathsf{p}^{\Theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})$ を求める. ただし $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とした.

そして, 事後分布に基づき, 母数の推測を行う. 3 番目のステップがどのように行われるかを θ の事前分布とデータ $\boldsymbol{X}$ の分布が共に離散型の場合で説明する.

真の母数 θ は確率変数 Θ の実現値と考える. いまは離散型確率変数の設定なので

$$\begin{aligned}\mathsf{Pr}\big(\Theta = \theta|\,\boldsymbol{X} = \boldsymbol{x}\big) &= \frac{\mathsf{Pr}\big(\Theta = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\mathsf{Pr}(\boldsymbol{X} = \boldsymbol{x})} \\ &= \frac{\mathsf{Pr}\big(\Theta = \theta,\, \boldsymbol{X} = \boldsymbol{x}\big)}{\sum_\theta \mathsf{Pr}\big(\boldsymbol{X} = \boldsymbol{x}\big|\,\Theta = \theta\big)\mathsf{Pr}\big(\Theta = \theta\big)}\end{aligned}$$

となる. 最後の等号は全確率の法則 (補題 1.11) を用いた. これを連続型分布の場合に形式的に書き直すと

$$\mathsf{p}^{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}|\,\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}|\,\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

と書ける. ただし $\mathsf{p}^{\boldsymbol{\Theta}|\,\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x})$ は $\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの $\boldsymbol{\Theta}$ の条件付き p.d.f. であり, θ の事後分布である. $\mathsf{p}^{\boldsymbol{X}|\,\Theta}(\boldsymbol{x}|\,\theta)$ は真の母数が θ のときのデータの分布の p.d.f. で, $\mathsf{p}^{\Theta}(\theta)$ は Θ の分布の p.d.f. である.

n 個のランダム標本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ に対して, 真の母数が θ のときの $\boldsymbol{X}$ の同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}|\Theta}(\,\cdot\,|\,\theta)$ とし, $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$ を観測したときの尤度関数 $\mathsf{lik}_n(\theta)$ を

$$\mathsf{lik}_n(\theta) = \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)$$

と書くことにする. よって

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta} =: \frac{\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)}{c_n} \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)$$

と表すことができる. ただし

$$c_n := \int \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta$$

は正規化定数と呼ばれる値である. c_n は θ に依存せず, データの実現値に依存した値である. よって事後分布の条件付き p.d.f. は事前分布と尤度関数の積の定数倍である.

$$\mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta).$$

c_n を無視してもよいのだろうか? 必要なときは求めることができるので問題ない. 事後分布の平均やモード (最頻値) を推測に用いることが多い. たとえば,

$$\int \{y - \theta\}^2 \mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

を最小にする y の値を $\overline{\theta}$ と書くと

$$\overline{\theta} = \int \theta \mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \frac{\displaystyle\int \theta\,\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}{\displaystyle\int \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

と表現できる. また信頼区間であれば, $0 < \alpha < 1/2$ に対して

$$\int_{-\infty}^{a} \mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \int_{b}^{\infty} \mathsf{p}^{\boldsymbol{\Theta}|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = \frac{\alpha}{2}$$

をみたす a, b を求めると開区間 $C = (a, b)$ は

$$\mathsf{Pr}\big(\theta \in C\big) = 1 - \alpha$$

をみたす事後信頼区間となる.

例 9.1. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0<\theta<1)$ とし, 事前分布は $(0,\,1)$ 上の一様分布とする. すなわち

$$\mathsf{p}^{\Theta}(\theta) = \begin{cases} 1 & (0<\theta<1) \\ 0 & (\text{その他の場合}) \end{cases}$$

である. すると

$$\begin{aligned} \mathsf{p}^{\Theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \propto \mathsf{p}^{\Theta}(\theta)\mathsf{lik}_n(\theta) &= \prod_{j=1}^{n} \theta^{x_j}(1-\theta)^{1-x_j}\mathbb{1}_{\{0,1\}}(x_j) \\ &= \theta^{s_n}(1-\theta)^{n-s_n}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(s_n) \end{aligned}$$

となる. ただし $s_n = \sum_{j=1}^{n} x_j$ である. したがって

$$\Theta|\,\boldsymbol{X} = \boldsymbol{x} \;\sim\; \mathsf{Beta}(s_n+1,\,n-s_n+1)$$

となることがわかる.

一方 $\alpha>0,\,\beta>0$ とし

$$\mathsf{p}^{\Theta}(\theta|\,\alpha,\,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (0<\theta<1) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. ただし $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}\,\mathrm{d}x$ である. すなわち θ の事前分布は $\mathsf{B}(\alpha,\,\beta)$ である. このとき

$$\begin{aligned} &\mathsf{p}^{\Theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x}) \\ &\quad= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s_n+\alpha)\Gamma(n-s_n+\beta)}\theta^{(s_n+\alpha)-1}(1-\theta)^{(n-s+\beta)-1}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(s_n) \end{aligned} \tag{9.1}$$

となる. したがって

$$\Theta|\,\boldsymbol{X} = \boldsymbol{x} \;\sim\; \mathsf{B}(s_n+\alpha,\,n-s_n+\beta)$$

がわかる. このとき

$$\overline{\theta} = \int \theta\mathsf{p}^{\Theta|\,\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta$$

とおくと

$$\overline{\theta} = \frac{s_n+\alpha}{n+\alpha+\beta}$$

を得る. $\widehat{\theta} = \dfrac{s_n}{n},\,\widetilde{\theta} = \dfrac{1}{2}$ とおくと

$$\overline{\theta} = \lambda_n\widehat{\theta} + (1-\lambda_n)\widetilde{\theta}, \quad \lambda_n = \frac{n}{n+\alpha+\beta}$$

と書ける.

信頼係数 0.95 の θ の事後信頼区間 $C = (a, b)$ は

$$\int_a^b \mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x})\,\mathrm{d}\theta = 0.95$$

をみたす a, b を数値計算で求めればよい. □

問 9.1. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)\,(0 < \theta < 1)$ とし,

$$\mathsf{p}^{\Theta}(\theta|\,\alpha,\,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & (x > 0) \\ 0 & (\text{その他の場合}) \end{cases}$$

とする. このとき, $X_1 = x_1$, $X_2 = x_2$, …, $X_n = x_n$ が与えられたときの Θ の条件付き p.d.f. が (9.1) で与えられることを示せ.

問 9.2. $X \sim \mathsf{B}(\alpha,\,\beta)$ のとき, $\mathsf{E}[X] = \dfrac{\alpha}{\alpha+\beta}$ となることを示せ.

例 9.2. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu,\,\sigma^2)$ とする. ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする. 事前分布として

$$\Theta \sim \mathsf{N}(a,\,b^2)$$

を仮定する. ただし $a \in \mathbb{R}$, $0 < b < \infty$ である. すると

$$\begin{aligned} &\Theta|\,\boldsymbol{X} = \boldsymbol{x} \sim \mathsf{N}(\overline{\theta},\,\tau^2), \qquad (9.2)\\ &\overline{\theta} = w\overline{x}_n + (1-w)a, \quad \overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j, \\ &w = \frac{\frac{1}{\mathsf{se}^2}}{\frac{1}{\mathsf{se}^2} + \frac{1}{b^2}}, \quad \frac{1}{\tau^2} = \frac{1}{\mathsf{se}^2} + \frac{1}{b^2}, \quad \mathsf{se} = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

となる. 問 9.3 を参照のこと. ただし, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とした.

このように事前分布と事後分布が同じ母数モデルに属するとき, 事前分布はこのモデルに随伴する(conjugate) という. または, このような事前分布を随伴事前分布という.

$n \to \infty$ のとき $w \to 1$ かつ $\dfrac{\tau}{\mathsf{se}} \to 1$ となる. 標本数が大きいとき

$$\Theta|\,\boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\overline{x}_n,\,\mathsf{se}^2\big)$$

となる. また n を固定する. $b \to \infty$ としたとき

$$\Theta|\,\boldsymbol{X} = \boldsymbol{x} \approx \mathsf{N}\big(\overline{x}_n,\,\mathsf{se}^2\big)$$

となる. これは一様な事前分布に対応するものである.

区間 $C = (c, d)$ は

$$\mathsf{Pr}\big(\theta \in C \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.95$$

をみたすものとする. したがって

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025, \quad \mathsf{Pr}\big(\theta > d \big| \boldsymbol{X} = \boldsymbol{x}\big) = 0.025$$

となるように c, d を選べばよい. このことから c を以下をみたすように選べばよい.

$$\mathsf{Pr}\big(\theta < c \big| \boldsymbol{X} = \boldsymbol{x}\big) = \mathsf{Pr}\left(\frac{\theta - \overline{\theta}}{\tau} < \frac{c - \overline{\theta}}{\tau} \middle| \boldsymbol{X} = \boldsymbol{x}\right) = \mathsf{Pr}\left(Z < \frac{c - \overline{\theta}}{\tau}\right).$$

ただし $Z \sim \mathsf{N}(0, 1)$ である. さらに

$$\mathsf{Pr}(Z < -1.96) = 0.025$$

なので

$$\frac{c - \overline{\theta}}{\tau} = -1.96$$

とすればよい. よって

$$c = \overline{\theta} - 1.96\tau$$

とする. 同様に

$$d = \overline{\theta} + 1.96\tau$$

を得る. これらのことから信頼係数 0.95 の Bayes 的信頼区間は $(\overline{\theta} - 1.96\tau, \overline{\theta} + 1.96\tau)$ となる. さらに $\widehat{\theta} \approx \widetilde{\theta}$ かつ $\tau \approx \mathsf{se}$ なので信頼係数 0.95 の Bayes 的信頼区間は近似的に $(\widehat{\theta} - 1.96\mathsf{se}, \widehat{\theta} + 1.96\mathsf{se})$ となり, 頻度論的信頼区間と同じになる.

問 9.3. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(\mu, \sigma^2)$ とする. ただし $\mu \in \mathbb{R}$ で σ^2 は既知とする. さらに, 事前分布として

$$\Theta \sim \mathbf{N}(a, b^2)$$

を仮定する. ただし $a \in \mathbb{R}, 0 < b < \infty$ である. このとき, $\boldsymbol{X} = \boldsymbol{x}$ が与えられたときの Θ の条件付き分布が (9.2) で与えられることを確かめよ.

9.3　事前分布の選択について

Bayes 的推測における重要な疑問は, 事前分布 π の選択である. ある学派は, 母数 θ に関する主観的な意見を反映させて事前分布を選択すれ

ばよいと主張している. ある場合には, このやり方は実行可能かもしれない. しかし, 多変数の母数を含む複雑な問題に対しては, 実行可能性に欠けることもあるだろう. さらに, 客観的かつ科学的な決定を目標とするデータ解析において, 主観的な意見を入り込ませることは相容れないという見方ができるであろう.

別の方策として, ある種の無情報事前分布を用いるやり方がある. たとえば, Bernoulli 試行において, 母数 $\theta\,(0<\theta<1)$ に対する事前分布 π を

$$\mathsf{p}^{\Theta}(\theta)\equiv 1$$

と定める. このとき, 形式的な計算から

$$\Theta|\,\mathcal{D}_n\ \sim\ \mathsf{Beta}(s_n+1,\,n-s_n+1)$$

を得る. ただし, $\mathcal{D}_n=\{x_1,\,x_2,\,\ldots,\,x_n\}$ で $s_n=\sum_{j=1}^{n}x_j$ である.

9.3.1 一様事前分布

$X\ \sim\ \mathsf{N}(\theta,\,\sigma^2)$ とする. ただし, σ は既知とする. いま, $\mathcal{D}_n$ を, $\mathsf{N}(\theta,\,\sigma^2)$ からの標本の大きさが n 個のランダム標本の実現値とする. θ の事前分布 p^{Θ} を

$$\mathsf{p}^{\Theta}(\theta)\ \propto\ c$$

を採用する. ただし, $c>0$ は定数である. このとき

$$\int \mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta=\infty$$

となるので, p^{Θ} は p.d.f. ではない. しかし, 形式的な計算により

$$\mathsf{p}^{\Theta|\,\boldsymbol{X}}(\theta|\,\mathcal{D}_n)\ \propto\ \mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)\ \propto\ \mathsf{lik}_n(\theta)$$

となることがわかる. $\overline{x}_n=(1/n)\sum_{j=1}^{n}x_j$ とおいたとき

$$\Theta|\,\mathcal{D}_n\ \sim\ \mathsf{N}\left(\overline{x}_n,\,\frac{\sigma^2}{n}\right)$$

となることがわかる. なぜならば

$$\begin{aligned}\mathsf{lik}_n(\theta)\mathsf{p}^{\Theta}(\theta)&=c\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_j-\theta)^2}{2\sigma^2}\right\}\\&=c\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{n}\exp\left\{-\frac{\sum_{j=1}^{n}(x_j-\overline{x}_n)^2}{2\sigma^2}-\frac{(\overline{x}_n-\theta)^2}{2\sigma^2/n}\right\}\end{aligned}$$

$$\propto \exp\left\{-\frac{(\overline{x}_n-\theta)^2}{2\sigma^2/n}\right\}$$

であるからである. 以上の議論から Bayes 的推定量は頻度論的な推定量である標本平均と同じになる.

しかし, 一様事前分布は母数の変換に関して不変ではない. たとえば, $X \sim \mathsf{Ber}(\theta)\,(0<\theta<1)$ とし, 事前分布 p^Θ として

$$\mathsf{p}^\Theta(\theta) = \begin{cases} 1 & (0<\theta<1) \\ 0 & (\text{その他の場合}) \end{cases}$$

を採用する. いま, 母数の変換 Ψ を

$$\Psi := \Psi(\theta) = \log\left(\frac{\theta}{1-\theta}\right) \Leftrightarrow \Psi := \log\left(\frac{\Theta}{1-\Theta}\right)$$

と定める. すると

$$\begin{aligned}\mathsf{p}^\Psi(\psi) &= \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}^\Psi(\Psi\le\psi) = \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}^\Psi\left(\log\left(\frac{\Theta}{1-\Theta}\right)\le\psi\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}\psi}\mathsf{Pr}\left(\Theta\le\frac{e^\psi}{1+e^\psi}\right) = \frac{\mathrm{d}}{\mathrm{d}\psi}\left(\frac{e^\psi}{1+e^\psi}\right) = \frac{e^\psi}{(1+e^\psi)^2} \qquad (\psi\in\mathbb{R})\end{aligned}$$

となるので, Θ の分布は一様分布であるので, Θ についての事前情報はない. Θ に関して無知であれば, Θ の変換である Ψ に関しても事前情報はないはずである. しかし, Ψ の分布は一様分布でないので, Ψ についての事前情報があることになる. したがって, Ψ についても事前情報がないことと p^Ψ の事前分布の間には齟齬が生じることになる.

9.3.2 Jeffreys 事前分布

母数の変換に関して不変な事前分布を定めるために, **Jeffreys 事前分布** を

$$\mathsf{p}^\Theta(\theta) \propto \sqrt{|\mathcal{F}(\theta)|}$$

で定める. ここで, $\mathcal{F}(\theta)$ は Fisher 情報行列で

$$\mathcal{F}(\theta) = \mathsf{E}\left[\left(\frac{\partial}{\partial\theta}\log\mathsf{lik}_n(\theta|\,\boldsymbol{X})\right)\left(\frac{\partial}{\partial\theta}\log\mathsf{lik}_n(\theta|\,\boldsymbol{X})\right)^\top\middle|\,\Theta=\theta\right]$$

で定義され, $|\cdot|$ は行列式を表す. また, 期待値は $\boldsymbol{X}$ の分布に関して取っていることに注意せよ.

定理 9.3. Jeffreys 事前分布は母数の変換に対して不変である.

Proof. □

例 9.4. Bernoulli 分布を考える. すると

$$\begin{aligned}\mathcal{F}(\theta) &= \mathsf{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\log\bigl(\theta^X(1-\theta)^{1-X}\bigr)\right)^2\middle|\Theta=\theta\right]\\ &= \mathsf{E}\left[\left(\frac{X}{\theta}-\frac{1-X}{1-\theta}\right)^2\middle|\Theta=\theta\right]\\ &= \mathsf{E}\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\middle|\Theta=\theta\right] = \frac{1}{\theta(1-\theta)}\end{aligned}$$

となる. よって, Jeffreys 事前分布 p^{Θ} は

$$\mathsf{p}^{\Theta}(\theta) \propto \sqrt{|\mathcal{F}(\theta)|} = \theta^{-1/2}(1-\theta)^{-1/2}$$

となる. これは $\mathsf{Beta}(1/2, 1/2)$ の p.d.f. であり, 両端の 0 と 1 の付近で密度が高い U 字型の分布である. □

9.3.3 共役事前分布

定義 9.5. 事前分布は**共役**であるとは, 事後分布が事前分布の分布族に属すときをいう. すなわち, 事前分布の分布族を $\mathcal{P}$ としたとき

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\,\cdot\,|\,\boldsymbol{x}) \in \mathcal{P}$$

が成立することである. ただし, $\mathsf{p}_\theta^{\boldsymbol{X}}$ を標本分布の p.d.f. で

$$\mathsf{p}^{\Theta|\boldsymbol{X}}(\theta|\,\boldsymbol{x}) = \frac{\mathsf{p}_\theta^{\boldsymbol{X}|\Theta}(\boldsymbol{x})\mathsf{p}^{\Theta}(\theta)}{\int \mathsf{p}_\theta^{\boldsymbol{X}|\Theta}(\boldsymbol{x})\mathsf{p}^{\Theta}(\theta)\,\mathrm{d}\theta}$$

である.

指数型分布族に対する共役事前分布を導出してみよう. σ 有限な測度 μ に関する p.d.f. $\mathsf{p}^{\boldsymbol{X}|\Theta}(\,\cdot\,|\,\theta)$ は

$$\mathsf{p}^{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\,\boldsymbol{\theta}) = \exp\bigl\{\boldsymbol{\theta}^\top\boldsymbol{x} - \mathsf{K}(\boldsymbol{x})\bigr\} \tag{9.3}$$

で与えられたとする. ただし, $\boldsymbol{\theta} \in \mathbb{R}^d$ で, 母数空間 $\Theta \subset \mathbb{R}^d$ は部分集合

$$\left\{\boldsymbol{\theta} \in \mathbb{R}^d;\, \mathsf{K}(\boldsymbol{\theta}) = \log\left(\int \exp\bigl\{\boldsymbol{\theta}^\top\boldsymbol{x}\bigr\}\,\mathrm{d}\mu(\boldsymbol{x})\right) < \infty\right\}$$

の内部で定義する.

定理 9.6. (9.3) で与えられる指数型分布族に対する共役事前分布の p.d.f. は

$$\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta}) = \frac{\exp\Big(n_0\boldsymbol{x}_0^{\top}\boldsymbol{\theta} - n_0\mathsf{K}(\boldsymbol{\theta})\Big)}{\displaystyle\int \exp\Big(n_0\boldsymbol{x}_0^{\top}\boldsymbol{\theta} - n_0\mathsf{K}(\boldsymbol{\theta})\Big)\,\mathrm{d}\mu(\boldsymbol{\theta})}$$

で与えられる. ただし, $\boldsymbol{x}_0 \in \mathbb{R}^d$, $n_0 \in \mathbb{R}$ である.

Proof. 以下の計算からわかる.

$$\begin{aligned}
\mathsf{p}^{\Theta|\,\boldsymbol{X}}(\boldsymbol{x}|\,\boldsymbol{\theta})\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta}) &\propto \exp\Big(\boldsymbol{\theta}^{\top}\boldsymbol{x} - \mathsf{K}(\boldsymbol{\theta})\Big)\exp\Big(n_0\boldsymbol{x}_0^{\top}\boldsymbol{\theta} - n_0\mathsf{K}(\boldsymbol{\theta})\Big)\\
&= \exp\Big((\boldsymbol{x}+n\boldsymbol{x}_0)^{\top}\boldsymbol{\theta} - (1+n_0)\mathsf{K}(\boldsymbol{\theta})\Big)\\
&= \exp\left\{(1+n_0)\left(\frac{1}{1+n_0}\boldsymbol{x} + \frac{n_0}{1+n_0}\boldsymbol{x}_0\right)^{\top} - (1+n_0)\mathsf{K}(\boldsymbol{\theta})\right\}\Big)\\
&\propto \mathsf{p}^{\Theta}_{\frac{1}{1+n_0}\boldsymbol{x}+\frac{n_0}{1+n_0}\boldsymbol{x}_0,\,1+n_0}(\boldsymbol{\theta}).
\end{aligned}$$

□

$\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ を (9.3) からの標本の大きさが n のランダム標本とする. すると $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ の同時 p.d.f. は

$$\prod_{j=1}^{n}\mathsf{p}^{\boldsymbol{X}|\,\Theta}(\boldsymbol{x}_j|\,\boldsymbol{\theta}) = \exp\Big(n\boldsymbol{\theta}^{\top}\overline{\boldsymbol{x}}_n - n\mathsf{K}(\boldsymbol{\theta})\Big), \qquad \overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j$$

となるので

$$\begin{aligned}
&\mathsf{p}^{\Theta|\,(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_n)}\big(\boldsymbol{\theta}|\,\boldsymbol{x}_1,\,\boldsymbol{x}_2,\,\ldots,\,\boldsymbol{x}_n\big)\\
&\quad\propto \exp\Big(n\boldsymbol{\theta}^{\top}\overline{\boldsymbol{x}}_n - n\mathsf{K}(\boldsymbol{\theta})\Big)\exp\Big(n_0\boldsymbol{x}_0^{\top}\boldsymbol{\theta} - n_0\mathsf{K}(\boldsymbol{\theta})\Big)\\
&\quad= \exp\left((n+n_0)\left(\frac{n}{n+n_0}\overline{\boldsymbol{x}}_n + \frac{n_0}{n+n_0}\boldsymbol{x}\right) - (n+n_0)\mathsf{K}(\boldsymbol{\theta})\right)
\end{aligned}$$

となる. よって, 事後分布の p.d.f. は $\mathsf{p}^{\Theta}_{\frac{n}{n+n_0}\overline{\boldsymbol{x}}_n+\frac{n_0}{n+n_0}\boldsymbol{x}_0,\,n+n_0}(\boldsymbol{\theta})$ で与えられることがわかる.

いま, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^{\top}$ と $\nabla = \left(\dfrac{\partial}{\partial\theta_1}, \dfrac{\partial}{\partial\theta_2}, \ldots, \dfrac{\partial}{\partial\theta_d}\right)^{\top}$ としたとき

$$\nabla\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta}) = n_0\big(\boldsymbol{x}_0 - \nabla\mathsf{K}(\boldsymbol{\theta})\big)\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\,n_0}(\boldsymbol{\theta})$$

となる. 一方

$$\int \nabla \mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\, n_0}(\boldsymbol{\theta})\, \mathrm{d}\mu(\boldsymbol{\theta}) = \nabla \underbrace{\int \mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\, n_0}(\boldsymbol{\theta})\, \mathrm{d}\mu(\boldsymbol{\theta})}_{=1} = \mathbf{0}_d$$

となる. これらの 2 つの等式から

$$\begin{aligned}
\mathsf{E}\big[\nabla \mathsf{K}(\boldsymbol{\theta})\big] &= \mathsf{E}\left[\boldsymbol{x}_0 - \frac{1}{n_0} \nabla \mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\, n_0}(\boldsymbol{\theta})\right] \\
&= \boldsymbol{x}_0 - \frac{1}{n_0} \int \nabla \mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\, n_0}(\boldsymbol{\theta})\, \mathrm{d}\mu(\boldsymbol{\theta}) \\
&= \boldsymbol{x}_0
\end{aligned}$$

を得る. ただし, 期待値は $\mathsf{p}^{\Theta}_{\boldsymbol{x}_0,\, n_0}(\boldsymbol{\theta})$ に関してである. もっと一般的には

$$\mathsf{E}\big[\nabla \mathsf{K}(\boldsymbol{\theta}) \big| \boldsymbol{X}_1,\, \boldsymbol{X}_2,\, \ldots,\, \boldsymbol{X}_n\big] = \frac{n}{n+n_0} \overline{\boldsymbol{X}}_n + \frac{n_0}{n+n_0} \boldsymbol{x}_0$$

となる.

9.4 1 次元の van Trees 不等式

母数空間 Θ を $\mathbb{R}$ の開部分集合とする状態空間 $(\mathbb{X},\, \mathcal{B})$ 上の統計的モデル

$$\mathcal{P} := \big\{\mathsf{P}_\theta;\, \theta \in \Theta, \mathsf{P}_\theta \text{ は } (\mathbb{X},\, \mathcal{B}) \text{ 上の確率分布で, } \mathsf{P}_\theta \ll \mu\big\}$$

を考える. ただし, μ は $(\mathbb{X},\, \mathcal{B})$ 上の σ 有限な測度である. さらに

$$\mathsf{p}_\theta := \frac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mu}$$

とし, 写像

$$\Theta \times \mathbb{X} \ni (\theta,\, x) \mapsto \mathsf{p}_\theta(x)$$

は可測[1] と仮定する. さらに

$$\sqrt{\mathsf{p}_\theta} \in L^2(\mu)$$

と仮定する. 以降, $\|\cdot\|_\mu$ は $L^2(\mu)$ における L^2 ノルムとする. すなわち, $L^2(\mu)$ の元 $g : \mathbb{X} \longrightarrow \mathbb{R}$ に対して

$$\|g\|_\mu = \sqrt{\int_{\mathbb{X}} g^2(x)\, \mathrm{d}\mu(x)}$$

と定める. また, $\mathbb{X}$ のノルムを $\|\cdot\|$ と記すことにする.

[1] $\|\cdot\|_\mu$ によって誘導される位相における開集合族の Borel 集合族に関して可測と考えるのが自然であろう.

定義 9.7 (L^2 における可微分性)**.** σ 有限な測度 μ によって支配される統計的モデル $\mathcal{P}$ は $\theta_0 \in \Theta$ で $\underline{L^2(\mu) \text{ に関して微分可能}}$ であるとは, 任意の $\sqrt{\mathsf{p}_\theta} \in L^2(\mu)$ に対して, ある $\dot{\sqrt{\mathsf{p}}}_{\theta_0} \in L^2(\mu)$ が存在して,

$$\left\| \sqrt{\mathsf{p}_\theta} - \sqrt{\mathsf{p}_{\theta_0}} - (\theta - \theta_0)\, \dot{\sqrt{\mathsf{p}}}_{\theta_0} \right\|_\mu = o(|\theta - \theta_0|) \qquad (\theta \to \theta_0)$$

をみたすときをいう. ここで, $\dot{\sqrt{\mathsf{p}}}_{\theta_0}$ を $\underline{\theta_0 \text{ における } L^2(\mu) \text{ 微分}}$ とよぶ.

注意 9.8. 統計的モデル $\mathcal{P}$ が $L^2(\mu)$ に関して微分可能のとき, $\theta_0\,(\in \Theta)$ における Fisher 情報量 $\mathcal{F}_{\mathcal{P}}(\theta_0)$ は

$$\begin{aligned}\mathcal{F}_{\mathcal{P}}(\theta_0) &= \int_{\mathbb{X}} \left\{ \frac{\mathrm{d}}{\mathrm{d}\theta} \log \mathsf{p}_\theta(x) \bigg|_{\theta=\theta_0} \right\}^2 \mathsf{p}_{\theta_0}(x)\, \mathrm{d}\mu(x) \\ &= 4 \int_{\mathbb{X}} \left\{ \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x) \right\}^2 \mathrm{d}\mu(x) = \int_{\mathbb{X}} \left\{ \frac{\dot{\mathsf{p}}_\theta(x)}{\mathsf{p}_\theta(x)} \right\}^2 \mathrm{d}\mu(x)\end{aligned}$$

と表現できる.

定義 9.9. 開集合 $\Theta\,(\subset \mathbb{R})$ 上に集中する確率分布 Q は $\underline{\text{よい振る舞いの事前分布}}$ であるとは, Q は $\mathbb{R}$ 上の Lebesgue 測度に関して絶対連続な密度関数 q を持ち, さらにほとんど至る所存在する導関数 $\dot{\mathsf{q}}$ を持ち

$$\mathcal{F}_{\mathsf{Q}} := \int_\Theta \left\{ \dot{\mathsf{q}}(\theta) \right\}^2 \frac{\mathbb{1}\{\mathsf{q}(\theta) > 0\}}{\mathsf{q}(\theta)}\, \mathrm{d}\mathsf{m}(\theta) < \infty$$

をみたすときをいう. ただし, m は $\mathbb{R}$ 上の Lebesgue 測度である.

統計量 $T : \mathbb{X} \longrightarrow \mathbb{R}$ は関数 $\psi(\theta)$ の推定量とする. ここで, ψ は絶対連続で, ほとんど至る所導関数 $\dot{\psi}$ をもつとする. E_θ は P_θ に関する期待値としたとき, van Trees 不等式は

$$\int_\Theta \mathsf{E}_\theta\left[\{T - \psi(\theta)\}^2 \right] \mathrm{d}\mathsf{Q}(\theta) \geq \frac{\left(\int_\Theta \dot{\psi}(\theta)\, \mathrm{d}\mathsf{Q}(\theta) \right)^2}{\mathcal{F}_{\mathsf{Q}} + \displaystyle\int_\Theta \mathcal{F}_{\mathcal{P}}(\theta)\, \mathrm{d}\mathsf{Q}(\theta)} \tag{9.4}$$

と表現できる. この不等式のために以下の仮定をおく.

仮定 9.10. (1) Θ は $\mathbb{R}$ の開集合.
(2) 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度 μ によって支配される.
(3) μ に関する P_θ の密度を $\mathsf{p}_\theta = \dfrac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mu}$ と書いたとき, 写像

$$\Theta \times \mathbb{X} \ni (\theta, x) \mapsto \mathsf{p}_\theta(x)$$

は可測である.
(4) $\mathcal{P}$ は $\Theta \cap \mathrm{supp}(\mathsf{q})$ 上のほとんど至る所で可微分である. ただし, $\mathrm{supp}(\mathsf{q}) := \left\{\theta \in \Theta;\ \dfrac{\mathrm{d}\mathsf{Q}}{\mathrm{dm}} = \mathsf{q}(\theta) > 0\right\}$ であり, m は $\mathbb{R}$ 上の Lebesgue 測度である.
(5) 関数 $\psi : \Theta \longrightarrow \mathbb{R}$ は絶対連続である.
(6) 関数 ψ^2 と $\dot{\psi}$ は Q 可積である.
(7)

$$\int_{\Theta} \mathsf{E}_\theta[T^2]\,\mathrm{d}\mathsf{Q}(\theta) < \infty; \qquad \int_{\Theta} \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta) < \infty$$

である.
(8) 任意の $B \in \mathcal{B}$ に対して, 関数

$$\Theta \cap \mathrm{supp}(\mathsf{q}) \ni \theta \mapsto \mathsf{P}_\theta(B)$$

は絶対連続である.
(9) θ が Θ の有限な境界に近づくとき, $\mathsf{q}(\theta) \to 0$ が成立する.

定理 9.11. 仮定 9.10 が成立し, $\mathcal{F}_{\mathsf{Q}} > 0$ とする. このとき, van Trees 不等式 (9.4) が成立する.

定理 9.11 の証明のために次の補題を準備する.

補題 9.12. 統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\ \theta \in \Theta,\ \mathsf{P}_\theta \ll \mu\}$ は $\theta_0 \in \Theta$ において, $L^2(\mu)$ に関して可微分とする. 有界な統計量 $T : \mathbb{X} \longrightarrow \mathbb{R}$ を考える. すなわち, ある $M > 0$ が存在して, $\mu\big\{|T| > M\big\} = 0$ である. このとき

$$\gamma_T : \Theta \ni \theta \mapsto \mathsf{E}_\theta[T]$$

は θ_0 において微分可能で, 導関数

$$\dot{\gamma}_T\,(\theta_0) = 2\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}\,(x)\sqrt{\mathsf{p}}_{\theta_0}(x)T(x)\,\mathrm{d}\mu(x) \tag{9.5}$$

をもつ.

Proof. 導関数を求めるために

$$\mathsf{p}_\theta - \mathsf{p}_{\theta_0} - 2(\theta - \theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0}$$

を評価する. そのために, $r_\theta = \sqrt{\mathsf{p}}_\theta - \sqrt{\mathsf{p}}_{\theta_0} - (\theta - \theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}$ とおく. すると

$$\mathsf{p}_\theta - \mathsf{p}_{\theta_0} - 2(\theta - \theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0} = \big\{\sqrt{\mathsf{p}}_\theta\big\}^2 - \big\{\sqrt{\mathsf{p}}_{\theta_0}\big\}^2 - 2(\theta - \theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0}$$

$$
\begin{aligned}
&= \Big\{\sqrt{\mathsf{p}}_{\theta_0} + (\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0} + r_\theta\Big\}^2 - \Big\{\sqrt{\mathsf{p}}_{\theta_0}\Big\}^2 \\
&\qquad - 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\,\sqrt{\mathsf{p}}_{\theta_0} \\
&= (\theta-\theta_0)^2\Big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\Big\}^2 + r_\theta^2 + 2\sqrt{\mathsf{p}}_{\theta_0} r_\theta \\
&\qquad + 2(\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\, r_\theta \qquad (9.6)
\end{aligned}
$$

となる. 最右辺の各項を積分したものを評価する.

$$
\begin{aligned}
\int_{\mathbb{X}} (\theta-\theta_0)^2\Big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\Big\}^2 \mathrm{d}\mu &= O(|\theta-\theta_0|^2), \qquad (\because \mathcal{P} \text{ の可微分性}) \\
\int_{\mathbb{X}} r_\theta^2 \,\mathrm{d}\mu &= o(|\theta-\theta_0|^2), \qquad (\because \mathcal{P} \text{ の可微分性}) \\
\left|\int_{\mathbb{X}} \sqrt{\mathsf{p}}_{\theta_0} r_\theta \,\mathrm{d}\mu\right| &\le \underbrace{\sqrt{\int_{\mathbb{X}} \mathsf{p}_{\theta_0}\,\mathrm{d}\mu}}_{=1} \sqrt{\int_{\mathbb{X}} r_\theta^2\,\mathrm{d}\mu} = o(|\theta-\theta_0|) \\
\left|\int_{\mathbb{X}} (\theta-\theta_0)\,\dot{\sqrt{\mathsf{p}}}_{\theta_0}\, r_\theta \,\mathrm{d}\mu\right| &\le |\theta-\theta_0| \sqrt{\int_{\mathbb{X}} \Big\{\dot{\sqrt{\mathsf{p}}}_{\theta_0}\Big\}^2 \mathrm{d}\mu} \sqrt{\int_{\mathbb{X}} r_\theta^2\,\mathrm{d}\mu} \\
&= o(|\theta-\theta_0|^2)
\end{aligned}
$$

となる. 以上から

$$
\begin{aligned}
&\left|\gamma_S(\theta) - \gamma_S(\theta_0) - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x)\sqrt{\mathsf{p}}_\theta(x) S(x)\,\mathrm{d}\mu(x)\right| \\
&\quad = \left|\int_{\mathbb{X}} S(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{X}} S(x)\mathsf{p}_{\theta_0}(x)\,\mathrm{d}\mu(x)\right. \\
&\qquad\quad \left. - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x) r_\theta(x) S(x)\,\mathrm{d}\mu(x)\right| \\
&\quad \le M\left|\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}\mu(x) - \int_{\mathbb{X}} \mathsf{p}_{\theta_0}(x)\,\mathrm{d}\mu(x)\right. \\
&\qquad\quad \left. - 2(\theta-\theta_0)\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}}}_{\theta_0}(x) r_\theta(x)\,\mathrm{d}\mu(x)\right| \\
&\quad = o(|\theta-\theta_0|)
\end{aligned}
$$

がわかる. 以上から, $\gamma_S(\theta_0)$ は θ_0 で可微分で, (9.5) で与えられる導関数 $\dot{\gamma}_S(\theta_0)$ をもつことがわかる. □

定理 9.11 の証明: いま

$$
\Delta : \mathbb{X}\times\Theta \ni (x,\,\theta) \mapsto \dot{\mathsf{q}}(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{p}_\theta(x)}
$$

$$+\sqrt{\mathsf{q}(\theta)}\,\dot{\sqrt{\mathsf{p}_\theta}}\,(x) \tag{9.7}$$

とおく. ただし, ほとんど至る所の $\theta \in \Theta_q := \Theta \cap \mathrm{supp}(\mathsf{q})$ 上では, Δ は well-defined であり, $\theta \notin \mathrm{supp}(\mathsf{q})$ 上では, $\Delta = 0$ となる. m を $\mathbb{R}$ 上の Lebesgue 測度としたとき, 以下の主張を証明する.

主張

$$2\int_{\Theta\times\mathbb{X}} \Delta(x,\theta)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}\big\{T(x)-\psi(\theta)\big\}\,\mathrm{d}\mathsf{m}(\theta) = \int_\Theta \dot{\psi}\,(\theta)\,\mathrm{d}\mathsf{Q}(\theta) \tag{9.8}$$

が成り立つ.

主張の証明: (9.4) もしくは (9.8) を T が指示関数の有限和である場合について証明し, あとは標準機械の議論を経由して, 可積な T に対しても成り立つことを示せばよい. さらに, $n \in \mathbb{N}$ に対して, $\psi_n = \max\{-n, \min\{\psi, n\}\}$ とおくと, ほとんど至るところで, $\psi_n \overset{n\to\infty}{\longrightarrow} \psi$ かつ $\dot{\psi}_n \overset{n\to\infty}{\longrightarrow} \dot{\psi}$ である. 有界な ψ_n と $\dot{\psi}_n$ に対して (9.4) もしくは (9.8) が成立すれば, 優収束定理により, 一般の ψ と $\dot{\psi}$ に対しても成立することがわかる. 以上の議論から, 一般性を失うことなく, T, ψ, $\dot{\psi}$ は有界と仮定してよい.

(9.8) の左辺の項ごとに評価していく:

$$\begin{aligned}
&2\int_{\Theta_q} \dot{\mathsf{q}}\,(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{q}(\theta)}T(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= \int_{\Theta_q\times\mathbb{X}} \dot{\mathsf{q}}\,(\theta)\Big\{\int_{\mathbb{X}} T(x)\mathsf{p}_\theta(x)\,\mathrm{d}\mu(x)\Big\}\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= \int_{\Theta_q} \dot{\mathsf{q}}\,(\theta)\gamma_T(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.9)\\
&-2\int_{\Theta_q\times\mathbb{X}} \dot{\mathsf{q}}\,(\theta)\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{2\sqrt{\mathsf{q}(\theta)}}\sqrt{\mathsf{q}(\theta)}\psi(\theta)\mathsf{p}_\theta(x)\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= -\int_{\Theta_q} \dot{\mathsf{q}}\,(\theta)\psi(\theta)\Big\{\int_{\mathbb{X}} \mathsf{p}_\theta(x)\,\mathrm{d}\mu(x)\Big\}\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= -\int_{\Theta_q} \dot{\mathsf{q}}\,(\theta)\gamma_T(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.10)\\
&2\int_{\Theta_q\times\mathbb{X}} \sqrt{\mathsf{q}(\theta)}\,\dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}T(x)\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= 2\int_{\Theta_q} \mathsf{q}(\theta)\Big\{\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{p}_\theta(x)}T(x)\,\mathrm{d}\mu(x)\Big\}\,\mathrm{d}\mathsf{m}(\theta)\\
&\quad= \int_{\Theta_q} \mathsf{q}(\theta)\,\dot{\gamma}_T\,(\theta)\,\mathrm{d}\mathsf{m}(\theta), \qquad (9.11)
\end{aligned}$$

$$
\begin{aligned}
&2\int_{\Theta_q\times\mathbb{X}} \sqrt{\mathsf{q}(\theta)}\ \dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{q}(\theta)}\sqrt{\mathsf{p}_\theta(x)}\psi(\theta)\,\mathrm{d}\mathsf{m}(\theta)\\
&\qquad = -2\int_{\Theta_q}\psi(\theta)\mathsf{q}(\theta)\left\{\dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{p}_\theta(x)}\,\mathrm{d}\mu(x)\right\}\mathrm{d}\mathsf{m}(\theta)\\
&\qquad = 0
\end{aligned}
$$

となる. 4 番目の式の最後の等号は, 補題 9.12 において, $T=1$ とおくと $\gamma_T(\theta)=1$ となるので, $\dot{\gamma}_T\,(\theta)=2\int_{\mathbb{X}}\dot{\sqrt{\mathsf{p}_\theta}}\,\sqrt{\mathsf{p}_\theta}\,\mathrm{d}\mu=0$ となることからわかる.

次に, Θ_q は可算個の互いに排反な区間 $(a_r,\,b_r)\,(r\in R)$ の和集合として表現できることに注意する. Θ_q の有限な各境界は, Θ の有限な境界, Θ の内点, $\mathrm{supp}(\mathsf{q})$ の有限な境界のいずれかである. ψ の連続性から, $\mathrm{supp}(\mathsf{q})$ の境界に θ が近づくと ψ と $\dot{\psi}$ の値は 0 に収束する. γ_T と ψ の有界性から, θ が a_r または b_r に近づくとき, $\gamma_T(\theta)\mathsf{q}(\theta)$ と $\psi(\theta)\mathsf{q}(\theta)$ の値は 0 に近づく. さらに, $\lim_{\theta\to\pm\infty}\mathsf{q}(\theta)=0$ である.

したがって, 閉区間 $[c,\,d]\subset\Theta_q$ に対して, (9.9) + (9.11) は

$$
\begin{aligned}
&\int_{[c,\,d]}\left\{\dot{\mathsf{q}}\,(\theta)\gamma_T(\theta)+\mathsf{q}(\theta)\,\dot{\gamma}_T\,(\theta)\right\}\mathrm{d}\mathsf{m}(\theta)=\left[\mathsf{q}(\theta)\gamma_T(\theta)\right]_c^d,\\
&\int_{[c,\,d]}\psi(\theta)\,\dot{\mathsf{q}}\,(\theta)\,\mathrm{d}\mathsf{m}(\theta)=\left[\psi(\theta)\mathsf{q}(\theta)\right]_c^d-\int_{[c,\,d]}\dot{\psi}\,(\theta)\mathsf{q}(\theta)\,\mathrm{d}\mathsf{m}(\theta)
\end{aligned}
$$

となる. よって, $c\to a_r$ と $d\to b_r$ とすれば

$$
\begin{aligned}
&\int_{a_r}^{b_r}\left\{\dot{\mathsf{q}}\,(\theta)\gamma_T(\theta)+\mathsf{q}(\theta)\,\dot{\gamma}_T\,(\theta)\right\}\mathrm{d}\mathsf{m}(\theta)=0,\\
&\int_{a_r}^{b_r}\psi(\theta)\,\dot{\mathsf{q}}\,(\theta)\,\mathrm{d}\mathsf{m}(\theta)=-\int_{a_r}^{b_r}\dot{\psi}\,(\theta)\mathsf{q}(\theta)\,\mathrm{d}\mathsf{m}(\theta)
\end{aligned}
$$

がわかる. 最後に, $r\in R$ について和を取れば, (9.8) は示せた.

(9.4) の証明: Cauchy-Schwarz の不等式を (9.8) に適用すると

$$
\begin{aligned}
&4\int_{\Theta\times\mathbb{X}}\Delta(x,\,\theta)^2\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\int_{\Theta\times\mathbb{X}}\left\{T(x)-\psi(\theta)\right\}^2\mathrm{d}\mathsf{P}_\theta(x)\,\mathrm{d}\mathsf{Q}(\theta)\\
&\qquad\ge\left(\int_\Theta\dot{\psi}\,(\theta)\,\mathrm{d}\mathsf{Q}(\theta)\right)^2
\end{aligned}
$$

となる. さらに

$$
\begin{aligned}
&4\int_{\Theta\times\mathbb{X}}\Delta(x,\,\theta)^2\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)\\
&\qquad=4\int_{\Theta\times\mathbb{X}}\left\{\dot{\mathsf{q}}\,(\theta)\right\}^2\frac{\mathbb{1}\{\mathsf{q}(\theta)>0\}}{4\mathsf{q}(\theta)}\mathsf{p}_\theta(x)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + 4\int_{\Theta\times\mathbb{X}} \Big\{\dot{\sqrt{\mathsf{p}_\theta}}\,(x)\Big\}^2 \mathsf{q}(\theta)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x) \\
&\quad + 4\int_{\Theta\times\mathbb{X}} \dot{\mathsf{q}}\,(\theta)\mathbb{1}\{\mathsf{q}(\theta)>0\}\,\dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{p}_\theta}(x)\,\mathrm{d}\mathsf{m}(\theta)\,\mathrm{d}\mu(x) \\
&= \mathcal{F}_{\mathsf{Q}} + \int_\Theta \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta) \\
&\quad + \int_\Theta \dot{\mathsf{q}}\,(\theta)\mathbb{1}\{\mathsf{q}(\theta)>0\} \underbrace{\left\{\int_{\mathbb{X}} \dot{\sqrt{\mathsf{p}_\theta}}\,(x)\sqrt{\mathsf{p}_\theta}(x)\,\mathrm{d}\mu(x)\right\}}_{=0\ (\because\text{補題 9.12 で } S=1\ )} \mathrm{d}\mu(x) \\
&= \mathcal{F}_{\mathsf{Q}} + \int_\Theta \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta)
\end{aligned}
$$

から

$$
\int_{\Theta\times\mathbb{X}} \big\{T(x)-\psi(\theta)\big\}^2 \mathrm{d}\mathsf{P}_\theta(x)\,\mathrm{d}\mathsf{Q}(\theta) \geq \frac{\left(\int_\Theta \dot{\psi}\,(\theta)\,\mathrm{d}\mathsf{Q}(\theta)\right)^2}{\mathcal{F}_{\mathsf{Q}} + \int_\Theta \mathcal{F}_{\mathcal{P}}(\theta)\,\mathrm{d}\mathsf{Q}(\theta)}
$$

がわかる. よって, 定理の不等式は証明された. □

9.5 統計的決定理論からみた Bayes 法

竹村 (2020, pp.326-334) と草間 (1975, pp.115-119) から借用すること.

9.6 章末注釈と参考文献

この章は [32, pp.175 – 192] を借用した.

9.7 演習問題

演習問題 9.1. $0<\theta<1$ と $n\in\mathbb{N}$ に対して, $X \sim \mathsf{Bino}(n,\,\theta)$ とする. θ の事前分布を $\Theta \sim \mathsf{B}(\alpha,\,\beta)$ とする. ただし, $\alpha>0$, $\beta>0$ である. すなわち, Θ の p.d.f. と $\Theta=\theta$ が与えられたときの X の条件付き p.m.f. はそれぞれ

$$
\begin{aligned}
\mathsf{p}^{\Theta}(\theta) &= \frac{1}{B(\alpha,\,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{1}_{(0,\,1)}(\theta), \\
\mathsf{p}^{X|\,\Theta}(x|\,\theta) &= \binom{n}{x}\theta^x(1-\theta)^{n-x}\mathbb{1}_{\{0,\,1,\,\ldots,\,n\}}(x)
\end{aligned}
$$

である. ただし

$$B(\alpha,\,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x$$

である. このとき, 以下の問いに答えよ.
(1) X の周辺 p.m.f. $\mathsf{p}^X(\,\cdot\,)$ を求めよ.
(2) $x=0,\,1,\,\dots\,n$ に対して, $X=x$ を与えたときの Θ の条件付き分布 (事後分布) は $B(\alpha+x,\,\beta+n-x)$ となることを示せ.
(3) $d^{\mathrm{Bayes}}(X)=\mathsf{E}[\Theta|\,X]$ とおいたとき

$$d^{\mathrm{Bayes}}(X)=\frac{\alpha+\beta}{\alpha+\beta+n}\cdot\frac{\alpha}{\alpha+\beta}+\frac{n}{\alpha+\beta+n}\cdot\frac{X}{n}$$

であることを示せ.
(4)

$$\mathsf{R}_{d^{\mathrm{Bayes}}}(\theta):=\mathsf{E}\big[\{d^{\mathrm{Bayes}}(X)-\theta\}^2\big]$$

を計算し, $\mathsf{R}_{d^{\mathrm{Bayes}}}(\theta)$ が θ に依存しない定数関数になるための条件は

$$\alpha=\beta=\frac{\sqrt{n}}{2}$$

であることを示せ.
(5) $\alpha=\sqrt{n}/2$, $\beta=\sqrt{n}/2$ のときの d^{Bayes} を d^{M} とする. すなわち

$$d^{\mathrm{M}}=\frac{1}{1+\sqrt{n}}\cdot\frac{1}{2}+\frac{\sqrt{n}}{1+\sqrt{n}}\cdot\frac{X}{n}$$

である. このとき

$$\mathsf{R}_{d^{\mathrm{M}}}(\theta)=\frac{1}{4(1+\sqrt{n})^2}<\sup_{\theta\in(0,\,1)}\mathsf{R}_{X/n}(\theta)$$

を示せ. ただし

$$\mathsf{R}_{X/n}(\theta):=\mathsf{E}\left[\left(\frac{X}{n}-\theta\right)^2\right]$$

である.

ヒント $\alpha>0$, $\beta>0$ に対して

$$B(\alpha,\,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},\quad \Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x,\quad \Gamma(\alpha+1)=\alpha\Gamma(\alpha)$$

である.

演習問題 9.2.

第10章　計算機集約的手法

節 10.1 では, 基本的な乱数発生法を述べる.　節 10.2 では, Markov 連鎖 Monte Carlo 法について述べる. 節 10.3 では, ブートストラップ法について説明する. 節 10.4 では, 指数型分布族の仮定のもとで不完全データに基づく最尤推定値を求めるアルゴリズムである EM アルゴリズムを説明する.

10.1　乱数の発生法

Monte Carlo 法は乱数生成法のことである. 区間 (0, 1) 上の一様乱数の列を発生させる. この列は決定的なアルゴリズムにより生成されるので真の意味ではランダム性を持たないので, 疑似乱数とよばれるものである. R では高速で安定した一様疑似乱数を発生するアルゴリズムである Mersenne twister 法を利用している. 一様乱数を使って代表的な分布に従う疑似乱数を生成する. ここでは, 一様乱数からある分布に従う (疑似) 乱数をどのように生成させるかを説明する. まず, c.d.f や p.d.f.(p.m.f.) の性質を直接利用して, 目標の分布に従う乱数の生成法 (確率積分法とよぶことにする) をまず説明する. つぎに, 本質的に計算機を利用すること (計算機集中的な手法) で目標の疑似乱数を生成する方法である受容・棄却法やマルコフ連鎖法を説明する.

以下の議論では, 開区間 (0, 1) 上の一様疑似乱数は容易に生成し, そのように目的の分布に従う疑似乱数を生成するかを説明していく. 以下では, (0, 1) 上の一様分布に対する疑似乱数のことを簡単に一様疑似乱数とよぶことにする. すなわち, 分布 P の疑似乱数を作成したとき, 一様疑似乱数から数 $x_1, x_2, \ldots, x_n$ を作成する. もちろん, これらは互いに関連しない数[1] である. これらが任意の開集合 $A \subset \mathbb{R}$ に対して

$$\mathsf{P}(A) \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(x_i)$$

となるようにする. 上記のような関係をみたす数 $x_1, x_2, \ldots, x_n$ を P の疑似乱数ということにする.

[1] 確率変数の言葉でいえば, 独立となるが, 数に対して独立を定義できないであろう.

10.1.1　ベルヌーイ分布

$0 < \theta < 1$ なる θ に対して, $\mathsf{Ber}(\theta)$ に従う乱数を発生させる場合には

- 一様乱数 u を発生させる.
- $0 < u \leq \theta \Longrightarrow x = 1$, $\theta < u < 1 \Longrightarrow x = 0$ とおく.

このとき x は $\mathsf{Ber}(\theta)$ の疑似乱数となる.

10.1.2　2 項分布

$n \in \mathbb{N}$, $0 < \theta < 1$ とする.

- $\mathsf{Ber}(\theta)$ の疑似乱数 $x_1, x_2, \ldots, x_n$ を発生させる.
- $s = x_1 + x_2 + \cdots + x_n$ とする.

このとき s は $\mathsf{Bino}(n, \theta)$ の疑似乱数となる.

10.1.3　標準正規分布

Box-Müller 法を用いる.

- 一様乱数 u_1, u_2 を発生させる.
- $r = \sqrt{-2 \log u_1}$, $\theta = 2\pi u_2$ とする.
- $x = r\cos\theta$, $y = r\sin\theta$ とする.

このとき x, y は　$\mathsf{N}(0, 1)$ の 2 つの疑似乱数となる.

これは以下のことからわかる. 大文字が確率変数で小文字はその実現値とする. たとえば, r は確率変数 R の実現値とみる. まず R の p.d.f. は

$$\mathsf{p}^R(r) = \begin{cases} r\exp\left(-\dfrac{r^2}{2}\right) & (r > 0) \\ 0 & (r \leq 0) \end{cases}$$

となることに注意する. したがって (R, Θ) の同時 p.d.f. は

$$\mathsf{p}^{(R,\Theta)}(r, \theta) = \begin{cases} \dfrac{1}{2\pi} r\exp\left(-\dfrac{r^2}{2}\right) & (r > 0,\ 0 < \theta < 2\pi) \\ 0 & (\text{その他の場合}) \end{cases} \tag{10.1}$$

である. ここで変換

$$\boldsymbol{h} = (h_1,\, h_2) : (r,\, \theta) \mapsto (r\cos\theta,\, r\sin\theta) = (x,\, y)$$

を考える.

$$r = \sqrt{x^2+y^2},\, \cos\theta = \frac{x}{\sqrt{x^2+y^2}},\, \sin\theta = \frac{y}{\sqrt{x^2+y^2}}$$

である. ただし, $0 \leq \theta < 2\pi$ とする. 変換 $\boldsymbol{h}$ の Jacobian は

$$J_{\boldsymbol{h}}(r,\, \theta) = \det \begin{pmatrix} \dfrac{\partial h_1}{\partial r} & \dfrac{\partial h_1}{\partial \theta} \\ \dfrac{\partial h_2}{\partial r} & \dfrac{\partial h_2}{\partial \theta} \end{pmatrix} = \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r$$

となるので

$$J_{\boldsymbol{h}^{-1}}(x,\, y) = \frac{1}{r} = \frac{1}{\sqrt{x^2+y^2}}.$$

(10.1) に対して p.d.f. の変換公式を用いれば

$$\begin{aligned} \mathsf{p}^{(X,Y)}(x,\, y) &= p^{(R,\Theta)}\left(\sqrt{x^2+y^2},\, \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right)\right)\frac{1}{\sqrt{x^2+y^2}} \\ &= \frac{1}{2\pi}\sqrt{x^2+y^2}\exp\left(-\frac{x^2+y^2}{2}\right)\frac{1}{\sqrt{x^2+y^2}} \\ &= \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right) \end{aligned}$$

となることよりわかる.

10.1.4 正規分布

$-\infty < \mu < \infty,\, 0 < \sigma < \infty$ とする. $\mathsf{N}(\mu,\, \sigma^2)$ に従う疑似乱数を生成させるには

- $\mathsf{N}(0,\, 1)$ の疑似乱数 z を発生させる.
- $x = \mu + \sigma z$ とおく.

このとき x は $\mathsf{N}(\mu,\, \sigma^2)$ の疑似乱数となる.

10.1.5 確率の積分変換

連続型分布の c.d.f. F に対してその逆関数 F^{-1} が陽にわかれば F に従う疑似乱数を発生させることができる.

- 一様疑似乱数 u を発生させる.
- $0 < u < 1$ に対して, $x = \mathsf{F}^{-1}(u) := \inf\{x \in \mathbb{R};\, \mathsf{F}(x) \geq u\}$ とおく. ただし, c.d.f. F は, $\mathsf{supp}(\mathsf{F}) := \overline{\{x \in \mathbb{R};\, \mathsf{F}(x) > 0\}}$ 上で連続かつ狭義単調増加であることを仮定している.

このとき, x は F の疑似乱数となる. なぜならば, $X \sim \mathsf{F}$ としたとき, $\forall x \in \{x \in \mathbb{R};\, \mathsf{F}(x) > 0\}$ に対して

$$\mathsf{Pr}\big(X \leq x\big) = \mathsf{Pr}\big(\mathsf{F}^{-1}(U) \leq x\big) = \mathsf{Pr}\big(U \leq \mathsf{F}(x)\big) = \mathsf{F}(x)$$

よりわかる.

10.1.6　受容・棄却法

疑似乱数をこれまでに説明してきた確率積分法等で発生させるのが難しい (疑似) 乱数の生成法を考える. 生成した乱数が従う分布の p.d.f. を p とする. さらに, 容易に乱数が生成できる分布の p.d.f. を q とする. この q はつぎのように取ることができたとする.

ある定数 $C \geq 1$ が存在して $\forall x$ に対して

$$\mathsf{p}(x) \leq C\mathsf{q}(x)$$

をみたす. この q に従う疑似乱数を用いて p に従う疑似乱数を発生させることができる.

- q の疑似乱数 x を発生させる.
- $\mathsf{Unif}\big(0,\, C\mathsf{q}(x)\big)$ の疑似乱数 u を発生させる.
- $\mathsf{p}(x) \geq u \Rightarrow \widetilde{x} = x$, $\mathsf{p}(x) < u \Rightarrow x$ を捨てる.

このとき $\widetilde{x}$ は p の疑似乱数となる.

このことは以下からわかる. まず

$$X \sim \mathsf{q}, \quad U|\, X = x \sim \mathsf{Unif}\big(0,\, C\mathsf{q}(x)\big)$$

に注意する. $-\infty < x < \infty$, $0 < u < C\mathsf{q}(x)$ のとき, $(X,\, U)$ の同時 p.d.f. $\mathsf{p}^{(X,U)}$ は

$$\mathsf{p}^{(X,U)}(x,\, u) = \mathsf{p}_{U|\,X}(u|\, x)\mathsf{q}(x) = \frac{1}{C\mathsf{q}(x)}\mathsf{q}(x) = \frac{1}{C}$$

で与えられることに注意する. よって

$$\mathsf{p}^{(X,U)}(x,\, u) = \begin{cases} \dfrac{1}{C} & (-\infty < x < \infty,\, 0 < u < C\mathsf{q}(x)) \\ 0 & (\text{その他の場合}) \end{cases}$$

である[2]. $t \in \mathsf{supp}(\mathsf{p}) \subset \mathbb{R}$ に対して

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(\widetilde{X} \le t\big) &= \frac{d}{dt}\mathsf{Pr}\big(X \le t \big|\, \mathsf{p}(X) \ge U\big) \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(X \le t,\, \mathsf{p}(X) \ge U\big)}{\mathsf{Pr}\big(\mathsf{p}(X) \ge U\big)} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t}\left\{\int_0^{\mathsf{p}(x)} \mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}u\right\}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\mathsf{p}(x)} \mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}u\right\}\mathrm{d}x} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t}\left\{\int_0^{\mathsf{p}(x)} \frac{1}{C}\,\mathrm{d}u\right\}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\mathsf{p}(x)} \frac{1}{C}\,\mathrm{d}u\right\}\mathrm{d}x} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t}\frac{\mathsf{p}(x)}{C}\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\frac{\mathsf{p}(x)}{C}\,\mathrm{d}x} \\
&= \frac{\dfrac{\mathsf{p}(t)}{C}}{\dfrac{1}{C}\displaystyle\int_{-\infty}^{\infty}\mathsf{p}(x)\,\mathrm{d}x} = \mathsf{p}(t)
\end{aligned}$$

からわかる.

注意 10.1. $C \ge 1$ としているが, $1/C$ の割合で X を採用しているので, C は 1 に近いほど乱数生成の計算効率はよいことがわかる. □

例 10.2. $\alpha = 1.3$, $\lambda = 5.6$ としたときの $\mathsf{Ga}(\alpha,\, \lambda)$ に従う乱数を発生させよう. すなわち

$$\mathsf{p}(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & (x \ge 0) \\ 0 & (x < 0) \end{cases}$$

に従う疑似乱数を生成する. ただし $\Gamma(\alpha) = \displaystyle\int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$ である. 簡単

[2]もちろん, $\mathsf{p}^{(X,U)}$ は同時 p.d.f. の条件をみたしている. なぜならば

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}x\,\mathrm{d}u = \frac{1}{C}\int_{-\infty}^{\infty}\left\{\int_0^{C\mathsf{q}(x)}\mathrm{d}u\right\}\mathrm{d}x = \int_{-\infty}^{\infty}\mathsf{q}(x)\,\mathrm{d}x = 1$$

からわかる.

に生成できる乱数の p.d.f. q として $\mathsf{Exp}(4)$ を用いる. すなわち

$$\mathsf{q}(x) = \begin{cases} 4e^{-4x} & (x \geq 0) \\ 0 & (x < 0). \end{cases}$$

である. $C = 1.2$ とすれば $x \geq 0$ のとき

$$\mathsf{p}(x) \leq C\mathsf{q}(x)$$

となることが計算機で確認することができる. $\mathsf{Exp}(4)$ の c.d.f. G は

$$\mathsf{G}(x) = \begin{cases} 1 - e^{-4x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

となるので, $0 < u < 1$ に対して

$$\mathsf{G}^{-1}(u) = -\frac{\log(1-u)}{4}$$

となる. さらに $1 - U \sim \mathsf{Unif}(0, 1)$ であることに注意すれば

$$U \sim \mathsf{Unif}(0, 1) \Rightarrow X = -\frac{\log U}{4} \sim \mathsf{Exp}(4)$$

となることがわかる.

よって, 以上の議論を踏まえると分布 p の疑似乱数の生成アルゴリズムは以下のようになる.

- 一様疑似乱数 v_1 を発生させる.
- $x = -\dfrac{\log v_1}{4}$ とおく.
- 別の一様疑似乱数 v_2 を発生させる. $u = C\mathsf{q}(x)v_2$ とおく.
- $u \leq \mathsf{p}(x) \Rightarrow \widetilde{x} = x$ とする. $u > \mathsf{p}(x)$ のときには x を捨てる.

この操作を繰り返せば, p に従う疑似乱数列を得る. □

10.2 Markov 連鎖 Monte Carlo 法 (MCMC 法)

以下は不正確なので全面的に書き直すこと. (2026/09/15).

μ を $\mathbb{R}^n$ 上の確率測度とする. すなわち $\mu : \mathcal{B}(\mathbb{R}^n) \ni A \mapsto \mu(A) \in [0, 1]$ は以下の条件をみたす.

- $\mu(\mathbb{R}^n) = 1$.
- $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R}^n)$, $A_i \cap A_j = \varnothing\ (i \neq j)$ に対して

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j).$$

ただし, $\mathcal{B}(\mathbb{R}^n)$ は $\mathbb{R}^n$ 上の Borel 集合族である.

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$ は μ 可積分関数としたとき積分

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\, \mathrm{d}\mu(\boldsymbol{x})$$

を求めた.

求積法では求積点 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^N$ と重み $w_1, w_2, \ldots, w_N\ (N \in \mathbb{N})$ をうまく選んで

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\, \mathrm{d}\mu(\boldsymbol{x}) \approx \sum_{j=1}^{N} w_j f(\boldsymbol{x}_j)$$

と近似することを目指す.

Monte Carlo 積分では, 台 $\boldsymbol{x}_j$ を確率測度 μ からランダムに生成させ, 重み w_j は分布 μ によって決める. X を μ に従う確率ベクトルとしたとき

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\, \mathrm{d}\mu(\boldsymbol{x}) = \mathsf{E}[f(X)] \approx \frac{1}{N} \sum_{j=1}^{N} f(\boldsymbol{x}_j) \tag{10.2}$$

となることが期待される. ただし $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ は μ からのランダム標本の実現値である. MCMC 法では (10.2) をみたすような標本を生成する体系的な方法である.

10.2.1 基本的な考え方

$\mathcal{B} := \mathcal{B}(\mathbb{R}^d)$ を $\mathbb{R}^d$ 上の Borel 集合族とする. 写像

$$P : \mathbb{R}^d \times \mathcal{B} \longrightarrow [0, 1]$$

は次の条件をみたす 確率推移核(probability transition kernel) とよぶ.

(1) 各 $B \in \mathcal{B}$ に対して写像

$$\mathbb{R}^d \ni \boldsymbol{x} \mapsto P(\boldsymbol{x}, B) \in [0, 1]$$

は可測関数.

(2) 各 $\boldsymbol{x} \in \mathbb{R}^d$ に対して写像

$$\mathcal{B} \ni B \mapsto P(\boldsymbol{x},\, B) \in [0,\, 1]$$

は確率測度.

離散時間確率過程は確率変数 $\boldsymbol{X}_j \in \mathbb{R}^d\,(j = 1,\, 2,\, \ldots)$ の添え字集合 $\mathbb{N}$ で順序付けられた集合 $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ である. 推移核 P をもつ時間等質的な Markov 連鎖 とは確率過程 $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ で次の性質を持つものである.

$$\mu^{\boldsymbol{X}_{j+1}}(B_{j+1}|\,\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j) = \mu^{\boldsymbol{X}_{j+1}}(B_{j+1}|\,\boldsymbol{x}_j) = P(\boldsymbol{x}_j,\, B_{j+1})$$
$$(B_{j+1} \in \mathcal{B},\, \boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j \in \mathbb{R}^d).$$

ただし $\boldsymbol{X}_j$ の確率測度を $\mu_{\boldsymbol{X}_j}$ と書き, $\boldsymbol{X}_1 = \boldsymbol{x}_1,\, \boldsymbol{X}_2 = \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{X}_j = \boldsymbol{x}_j$ を与えたときの $\boldsymbol{X}_{j+1}$ の条件付き確率測度を $\mu_{\boldsymbol{X}_{j+1}}(\,\cdot\,|\,\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j)$ と書いた. すなわち $\boldsymbol{X}_1 = \boldsymbol{x}_1,\, \boldsymbol{X}_2 = \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{X}_j = \boldsymbol{x}_j$ を与えたときの $\boldsymbol{X}_{j+1}$ の条件付き分布は $\boldsymbol{X}_j = \boldsymbol{x}_j$ にのみ依存する. そして $\boldsymbol{X}_{j+1}$ の分布 $\mu_{\boldsymbol{X}_{j+1}}$ は時間に関して等質である. すると $\boldsymbol{X}_{j+1}$ の分布 $\mu_{\boldsymbol{X}_{j+1}}$ は次のように表現される.

$$\mu^{\boldsymbol{X}_{j+1}}(B_{j+1}) = \int_{\mathbb{R}^d} P(\boldsymbol{x}_j,\, B_{j+1})\mu_{\boldsymbol{X}_{j+1}}(\mathrm{d}\boldsymbol{x}_j) =: \mu_{\boldsymbol{X}_{j+1}}P(B_{j+1}).$$

さらに 3 つの概念を導入する. $\mu P(\,\cdot\,) = \int_{\mathbb{R}^d} P(\boldsymbol{x},\, \cdot\,)\, d\mu(\boldsymbol{x})$ とおく.

- 測度 μ は推移核 $P(\boldsymbol{x}_j,\, B_{j+1})$ の 不変測度 であるとは

$$\mu P = \mu \iff \int_{\mathbb{R}^d} P(\boldsymbol{x},\, B)\,\mathrm{d}\mu(\boldsymbol{x}) = \mu(B) \qquad (B \in \mathcal{B})$$

をみたすときをいう.

- 推移核 P は μ に関して 既約 であるとは各 $\boldsymbol{x} \in \mathbb{R}^d$ と $B \in \mathcal{B}$ で $\mu(B) > 0$ なるものに対して, ある正の整数 k が存在して

$$P^{(k)}(\boldsymbol{x},\, B) > 0$$

をみたすときをいう. ただし

$$P^{(k)}(\boldsymbol{x}, B) = \int_{\mathbb{R}^d} P(\boldsymbol{y},\, B)P^{(k-1)}(\boldsymbol{x},\, \mathrm{d}\boldsymbol{y})$$
$$P^{(1)}(\boldsymbol{x},\, B) = P(\boldsymbol{x},\, B) \quad (\boldsymbol{x} \in \mathbb{R}^d,\, B \in \mathcal{B})$$

である.

- P は既約な推移核とする. P は **周期的** であるとはある整数 $m \geq 2$ が存在して互いに異なる空でない集合 E_j $(j = 1, 2, \ldots, m)$ と $\boldsymbol{x} \in E_j$ に対して
$$P(\boldsymbol{x}, E_{j+1(\mathrm{mod}\ m)}) = 1 \qquad (\forall x \in E_j)$$
であるときをいう. P が周期的でないとき, **非周期的** であるという.

命題 10.3.

ChatGPT のコメント (2026/09/15).
一般状態空間の Markov 連鎖では, 「不変測度・既約・非周期」だけでは, すべての初期値からの全変動収束や強法則までは保証されない. 主張する収束の種類に応じて, Harris 再帰性, 正 Harris 再帰性などを明記する必要あり.

μ は $\mathbb{R}^d$ 上の確率測度とし, $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ は推移核 P を持つ時間等質な Markov 連鎖とする. さらに μ は推移核 P に関して不変測度とし, P は既約で非周期的とする. このとき $\forall \boldsymbol{x} \in \mathbb{R}^d$ に対し

$$\lim_{N\to\infty} P^{(N)}(\boldsymbol{x}, B) = \mu(B) \qquad (B \in \mathcal{B}) \tag{10.3}$$

と μ 可積分な関数 f に対して

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} f(\boldsymbol{X}_j) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\, \mathrm{d}\mu(\boldsymbol{x})$$

がほとんど確実に成立する.

Proof. 信じることにする. □

注意 10.4. 生成する乱数の従う分布を μ と考えることになる. 既約で非周期的な推移確率核 P がわかると μ は (10.3) でつくることができるわけである. P の構成の仕方として, 次の節で詳解する Metropolis–Hasting 法が有名である. □

10.2.2 推移核の Metropolis–Hastings による構成

μ は $\mathbb{R}^d$ 上の確率測度とし $\mathbb{R}^d$ 上の Lebesgue 測度 $\mathsf{m}(\,\cdot\,)$ に関して絶対連続と仮定する. このことより $\mathbb{R}^d$ 上の非負関数 $\pi(\,\cdot\,)$ が存在して

$$\frac{\mathrm{d}\mu}{\mathrm{dm}}(\boldsymbol{x}) = \pi(\boldsymbol{x})$$

と書ける.

P は任意の推移核とする. 点 $\boldsymbol{x} \in \mathbb{R}^d$ が与えられたとき推移核 P は点 $\boldsymbol{x}$ を異なる点 $\boldsymbol{y}$ に移動させるか, そのまま留める. このことから推移核 P は以下のように分解できると仮定する.

$$P(\boldsymbol{x},\, B) = \int_B K(\boldsymbol{x},\, \boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y}) + r(\boldsymbol{x}) \mathbb{1}_B(\boldsymbol{x}).$$

ただし

$$\mathbb{1}_B(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in B) \\ 0 & (\boldsymbol{x} \notin B) \end{cases}$$

である. 核 $K(\boldsymbol{x},\, \boldsymbol{y}) : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0,\, \infty)$ に対して, $\boldsymbol{x}$ を $\boldsymbol{y}$ を含む微小な領域 $\mathrm{dm}(\boldsymbol{y})$ に移す確率は $K(\boldsymbol{x},\, \boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y})$ とし, $\boldsymbol{x}$ に留めおく確率を $r(\boldsymbol{x})$ である. 条件 $P(\boldsymbol{x},\, \mathbb{R}^d) = 1$ から

$$r(\boldsymbol{x}) = 1 - \int_{\mathbb{R}^d} K(\boldsymbol{x},\, \boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y}) \tag{10.4}$$

を得る.

次に μ が P に関して不変になるための条件を K と π で表現してみよう.

まず $B \in \mathcal{B}$ に対して

$$\begin{aligned}
\mu P(B) &= \int_{\mathbb{R}^d} P(\boldsymbol{x},\, B)\, \mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} P(\boldsymbol{x},\, b) \frac{\mathrm{d}\mu}{\mathrm{dm}}(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} P(\boldsymbol{x},\, B)\, \pi(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} \left\{ \int_B K(\boldsymbol{x},\, \boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y}) + r(\boldsymbol{x}) \mathbb{1}_B(\boldsymbol{x}) \right\} \pi(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} \left\{ \int_B K(\boldsymbol{x},\, \boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y}) \right\} \pi(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \\
&\qquad + \int_{\mathbb{R}^d} r(\boldsymbol{x}) \pi(\boldsymbol{x}) \mathbb{1}_B(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \\
&= \int_B \left\{ \int_{\mathbb{R}^d} K(\boldsymbol{x},\, \boldsymbol{y}) \pi(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x}) \right\} \mathrm{dm}(\boldsymbol{y}) \\
&\qquad + \int_B r(\boldsymbol{y}) \pi(\boldsymbol{y})\, \mathrm{dm}(\boldsymbol{y}) \\
&\qquad (2 \text{ 項目の積分の変数を } \boldsymbol{x} \text{ から } \boldsymbol{y} \text{ に変更}) \\
&= \int_B \left\{ \int_{\mathbb{R}^d} K(\boldsymbol{x},\, \boldsymbol{y}) \pi(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} + r(\boldsymbol{y}) \pi(\boldsymbol{y}) \right\} \mathrm{dm}(\boldsymbol{y})
\end{aligned} \tag{10.5}$$

と書きかえられることに注意する. すると $\pi(\boldsymbol{x})\, \mathrm{dm}(\boldsymbol{x})$ が P に関して不変測度になるためには

$$\mu P = \mu \Longleftrightarrow \mu P(B) = \mu(B) \qquad (\forall B \in \mathcal{B})$$

$$
\begin{aligned}
&\iff \int_B \left\{ \int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) + r(\boldsymbol{y})\pi(\boldsymbol{y}) \right\} \mathsf{dm}(\boldsymbol{y}) \\
&\qquad = \int_B \pi(\boldsymbol{y})\,\mathsf{dm}(\boldsymbol{y}) \qquad (\forall B \in \mathcal{B}) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) + r(\boldsymbol{y})\pi(\boldsymbol{y}) = \pi(\boldsymbol{y}) \qquad (\forall \boldsymbol{y} \in \mathbb{R}^d) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) = \pi(\boldsymbol{y})(1 - r(\boldsymbol{y})) \qquad (\forall \boldsymbol{y} \in \mathbb{R}^d)
\end{aligned}
\tag{10.6}
$$

となる. (10.4) を (10.6) の最右辺に代入 ($\boldsymbol{x}$ を $\boldsymbol{y}$ に変更した式) すると

$$
\begin{aligned}
&\int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) = \pi(\boldsymbol{y}) \int_{\mathbb{R}^d} K(\boldsymbol{y}, \boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})\,\mathsf{dm}(\boldsymbol{x}) = \int_{\mathbb{R}^d} K(\boldsymbol{y}, \boldsymbol{x})\pi(\boldsymbol{y})\,\mathsf{dm}(\boldsymbol{x})
\end{aligned}
\tag{10.7}
$$

を得る. この式を 釣合方程式(balance equation) とよぶ. K が

$$
\pi(\boldsymbol{x})K(\boldsymbol{x}, \boldsymbol{y}) = \pi(\boldsymbol{y})K(\boldsymbol{y}, \boldsymbol{x}) \quad (\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d) \tag{10.8}
$$

をみたすとき K は 詳細的釣合方程式(detailed balance equation) をみたすという.

Metropolis-Hastings アルゴリズムでは (10.8) をみたす推移核 K を構成する.

$q : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0, \infty)$ は任意の関数で

$$
\int q(\boldsymbol{x}, \boldsymbol{y})\,\mathsf{dm}(\boldsymbol{y}) = 1
$$

をみたす. この関数 q を用いて推移核 Q を

$$
Q(\boldsymbol{x}, B) = \int_B q(\boldsymbol{x}, \boldsymbol{y})\,\mathsf{dm}(\boldsymbol{y}) \qquad (\forall B \in \mathcal{B})
$$

で定める. q が詳細的釣合方程式をみたし

$$
K(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{x}, \boldsymbol{y}), \quad r(\boldsymbol{x}) = 0
$$

と仮定する. みたしていなければ乗数倍で修正すればよいので, 一般性を失わずに上記のことを仮定してよいことが以下の考察からわかる. まず

$$
K(\boldsymbol{x}, \boldsymbol{y}) = \alpha(\boldsymbol{x}, \boldsymbol{y})q(\boldsymbol{x}, \boldsymbol{y}) \tag{10.9}
$$

とおく. この α の選び方はいかのようにする. ある $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ に対して

$$
\pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x}) < \pi(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{y})
$$

ならば

$$\pi(\boldsymbol{y})\alpha(\boldsymbol{y},\,\boldsymbol{x})q(\boldsymbol{y},\,\boldsymbol{x}) = \pi(\boldsymbol{x})\alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y})$$

となるように α を選ぶ. これは

$$\alpha(\boldsymbol{y},\,\boldsymbol{x}) = 1, \qquad \alpha(\boldsymbol{x},\,\boldsymbol{y}) = \frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})} < 1$$

とすればよいことがわかる. したがって, $\boldsymbol{x}$ と $\boldsymbol{y}$ を入れ替えることで

$$\begin{aligned}\alpha(\boldsymbol{x},\,\boldsymbol{y}) &= \min\left\{1,\,\frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})}\right\},\\ K(\boldsymbol{x},\,\boldsymbol{y}) &= \alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y})\end{aligned}$$

とすると K は詳細的釣合方程式をみたすことがわかる. 実際, q が釣合方程式を満たしているときは, $\alpha = 1$ なので自明. そうでないときは $\alpha(\boldsymbol{x},\,\boldsymbol{y}) = \dfrac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})}$ なので

$$\begin{aligned}\pi(\boldsymbol{x})K(\boldsymbol{x},\,\boldsymbol{y}) &= \pi(\boldsymbol{x})\alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y})\\ &= \pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})\frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})}\\ &= \pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})\end{aligned}$$

からわかる.

候補分布 q が対称とする. すなわち

$$q(\boldsymbol{x},\,\boldsymbol{y}) = q(\boldsymbol{y},\,\boldsymbol{x}) \qquad (\forall \boldsymbol{x},\,\boldsymbol{y} \in \mathbb{R}^d)$$

とする. すると

$$\alpha(\boldsymbol{x},\,\boldsymbol{y}) = \min\left\{1,\,\frac{\pi(\boldsymbol{y})q(y,\,x_k)}{\pi(\boldsymbol{x})q(x_k,\,y)}\right\} \tag{10.10}$$

となる. 対称な q の例としては

$$\begin{aligned}&q(\boldsymbol{x},\,\boldsymbol{y}) = g(\boldsymbol{x}-\boldsymbol{y}),\\ &g : \mathbb{R}^d \longrightarrow [0,\,\infty) \text{ で } g(\boldsymbol{x}) = g(-\boldsymbol{x})\end{aligned}$$

をみたすものがある.

以上の議論から繰り返し回数 K のアルゴリズムは以下のようになる.

Step 1 : 初期値 $\boldsymbol{x}_1 \in \mathbb{R}^d$ を選び, $k = 1$ とする.

Step 2 : $q(\boldsymbol{x}_k, \boldsymbol{y})$ に従う疑似乱数 $\boldsymbol{y}$ を生成する.

Step 3 : $\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{1, \dfrac{\pi(\boldsymbol{y})q(\boldsymbol{y}, \boldsymbol{x}_k)}{\pi(\boldsymbol{x}_k)q(\boldsymbol{x}_k, \boldsymbol{y})}\right\}$ を計算する.

Step 4 : $\mathsf{Unif}(0, 1)$ に従う疑似乱数 t を生成する.

Step 5 :

$$\alpha(\boldsymbol{x}_k, \boldsymbol{y}) \geq t \Rightarrow \boldsymbol{x}_{k+1} = \boldsymbol{y}.$$
$$\alpha(\boldsymbol{x}_k, \boldsymbol{y}) < t \Rightarrow \boldsymbol{x}_{k+1} = \boldsymbol{x}_k$$

とおく.

Step 6 : $k = K$ ならば $\boldsymbol{x}_k$ を求めたい乱数として出力する. $k < K$ ならば $k \longrightarrow k+1$ として, Step 2 に戻る.

をみたすものがある.

注意 10.5.

ChatGPT からのコメント (2026/09/15).
提案値が棄却された場合には現在値 x_k に留まるため, 遷移分布には点質量が含まれます：

$$\mathsf{Pr}(x_k, \mathrm{d}y) = \alpha(x_k, y)\mathsf{q}x_k, y)\,\mathrm{d}y + r(x_k)\delta_{x_k}(\mathrm{d}y),$$
$$r(x_k) = 1 - \int_{\mathbb{R}} \alpha(x_k, y)\mathsf{q}(x_k, y)\,\mathrm{d}y.$$

上記アルゴリズムの Step 5 で得られた $\boldsymbol{X}_{k+1}$ の分布を確認する. 以後は, 簡単のために $d = 1$ とする. $X_k = x_k$ のときの条件付き p.d.f. を $\mathsf{q}(y|\, x_k)$ とする. 以下では $d\mathsf{m}(y)$ を簡単に dy と書くことにする. $Y \sim \mathsf{q}(y|\, x_k)$ かつ $U \sim \mathsf{Unif}(0, 1)$ で Y と U は独立とする. このとき $t \in \mathbb{R}$ に対して

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}(X_{k+1} \leq t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(Y \leq t \big|\, \alpha(X_k, Y) \geq U\big) \qquad (\because \text{Step 5 より}) \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\big(Y \leq t,\, \alpha(X_k, Y) \geq U)\big)}{\mathsf{Pr}\big(\alpha(X_k, Y) \geq U\big)} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t}\left\{\int_0^{\alpha(x_k, y)} \mathrm{d}u\right\}q(x_k, y)\,\mathrm{d}y}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\alpha(x_k, y)} \mathrm{d}u\right\}q(x_k, y)\,\mathrm{d}y}
\end{aligned}$$

$$= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y}{\displaystyle\int_{-\infty}^{\infty} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y}$$
$$= \frac{\alpha(x_k,\, t) q(x_k,\, t)}{\displaystyle\int_{-\infty}^{\infty} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y}$$

となり, (10.9) から, $X_k = x_k$ を与えたときの X_{k+1} の条件付き分布は

$$X_{k+1} \sim \frac{K(x_k,\, y)}{\displaystyle\int_{-\infty}^{\infty} K(x_k,\, y)\,\mathrm{d}y}$$

となることがわかる. □

例 10.6. p.d.f.

$$\pi(x) = Ce^{-4}\bigl(1 + |x|^3\bigr) \qquad (x \in \mathbb{R}) \tag{10.11}$$

から乱数を発生させることを考える. ただし

$$C^{-1} = \int_{-\infty}^{\infty} e^{-4}\bigl(1 + |x|^3\bigr)\,\mathrm{d}x$$

である. 候補 p.d.f. として

$$Y \,|\, x_k \sim \mathsf{N}(x_k,\, 1)$$

を採用する. すなわち

$$q(x,\, y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y - x)^2\right\}$$

である. ここで π は円周率である. すると (10.10) と (10.11) から

$$\alpha(x,\, y) = \min\left\{1,\, e^{-y^4 + x^4}\frac{1 + |y|^3}{1 + |x|^3}\right\}$$

となる.

以上の議論から繰り返し回数 K のアルゴリズムは以下のようになる.

Step 1 : 初期値 $x_1 \in \mathbb{R}$ を選び, $k = 1$ とする.

Step 2 : $\mathsf{N}(x_k,\, 1)$ からの疑似乱数 y を生成する.

Step 3 : $\alpha(x_k,\, y) = \min\left\{1,\, e^{-y^4 + x_k^4}\dfrac{1 + |y|^3}{1 + |x_k|^3}\right\}$ を計算する.

Step 4 : $\mathsf{Unif}(0,\, 1)$ からの疑似乱数 t を生成する.

Step 5 :

$$\alpha(x_k,\, y) \geq t \Longrightarrow x_{k+1} = y.$$
$$\alpha(x_k,\, y) < t \Longrightarrow x_{k+1} = x_k$$

とおく.

Step 6 : $k = K$ ならば $\boldsymbol{x}_k$ を求めたい乱数として出力する. $k < K$ ならば $k \longrightarrow k+1$ として, Step 2 に戻る.

□

10.2.3 Gibbs サンプリング法

10.3 ブートストラップ法

ブートストラップ法は推定量の標準誤差を推定する方法である. F を未知の c.d.f. とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{F}$ とする. 未知の母数には依存しないある関数 g があって $S_n = g(X_1, X_2, \ldots, X_n)$ と書く. すなわち, S_n は統計量である. このときに

$$T(\mathsf{F}) = \mathsf{Var}_{\mathsf{F}}\big[S_n\big]$$

を推定したいとする. 一般にこの量は未知の分布 F に依存する. たとえば $S_n = \overline{X}_n$ のとき,

$$\mathsf{Var}_{\mathsf{F}}\big[S_n\big] = \frac{\sigma^2}{n}$$

であり, 未知の母数 σ に依存する. ただし $\sigma^2 = \mathsf{E}\big[(X_1 - \mu)^2\big]$, $\mu = \mathsf{E}[X_1]$ である. $T(\mathsf{F})$ の推定量として, 差し込み推定量を考える.

$$T(\widehat{\mathsf{F}}_n) = \mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big].$$

ここで 大数の法則を思い出す. $B \in \mathbb{N}$ とし, G を c.d.f. とする. $Y_1, Y_2, \ldots, Y_B \overset{\text{i.i.d.}}{\sim} \mathsf{G}$ とする. $B \to \infty$ のとき

$$\overline{Y}_B = \frac{1}{B}\sum_{j=1}^{B} Y_j \xrightarrow{P} \mathsf{E}[Y]$$

となる. ただし $Y \sim \mathsf{G}$ である. さらに, 関数 h は $\mathsf{E}\big[|h(Y)|\big] < \infty$ をみたすとする. すると $B \to \infty$ のとき

$$\frac{1}{B}\sum_{j=1}^{B} h(Y_j) \xrightarrow{P} \mathsf{E}\big[h(Y)\big].$$

$\mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big]$ をシミュレーションで近似することを考える. $\mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big]$ は「データが分布 $\widehat{\mathsf{F}}_n$ に独立同一にしたがっているときの S_n の分散」と考えることができる. $X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ とし, $S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$ を計算する.

- 現実の世界: $\mathsf{F} \Longrightarrow X_1, X_2, \ldots, X_n \Longrightarrow S_n = g(X_1, X_2, \ldots, X_n)$.
- ブートストラップの世界: $\widehat{\mathsf{F}}_n \Longrightarrow X_1^*, X_2^*, \ldots, X_n^* \Longrightarrow S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$.

では, $X_1^*, X_2^*, \ldots, X_n^*$ を $\widehat{\mathsf{F}}_n$ からどのように発生させるのだろうか? $\widehat{\mathsf{F}}_n$ は各 $X_j\,(j = 1, 2, \ldots, n)$ に確率 $1/n$ をのせたものであったことを思い出そう.

$\widehat{\mathsf{F}}_n$ に従う確率変数 $X_j^*\,(j = 1, 2, \ldots, n)$ は元のデータ $X_1, X_2, \ldots, X_n$ から, 各データが等確率 $1/n$ で抽出したものと同じ条件付き分布をもつ

したがって $X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ を発生させるためには, $X_1, X_2, \ldots, X_n$ からデータを復元抽出すればよい. 手続きをまとめると以下のようになる.

- $X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ を発生させる. $X_1, X_2, \ldots, X_n$ から復元抽出をするだけでよい.
- $S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$ を計算する.
- 上のふたつのステップを B 回繰り返し, $S_{n,1}^*, S_{n,2}^*, \ldots, S_{n,B}^*$ を求める.
- $$\mathsf{v}_{\text{boot}} = \frac{1}{B}\sum_{j=1}^{B}\left(S_{n,j}^* - \frac{1}{B}\sum_{k=1}^{B} S_{n,k}^*\right)^2$$
  を計算する.

このとき

$$\mathsf{Var}_F\big(S_n\big) \overset{\text{そん色ない}}{\approx} \mathsf{Var}_{\widehat{F}_n}\big(S_n\big) \overset{\text{小さい}}{\approx} \mathsf{v}_{\text{boot}}.$$

$0 < \alpha < 1$ に対して, $z_{\alpha/2}$ を標準正規分布の両側 α 点とすれば, $\theta(\mathsf{F}) = \mathsf{E}\big[S_n\big]$ とし

$$\widehat{\mathsf{se}}_{\mathrm{boot}} := \sqrt{\mathsf{v}_{\mathrm{boot}}}$$

としたとき

$$S_n \pm z_{\alpha/2}\widehat{\mathsf{se}}_{\mathrm{boot}}$$

は $\theta(\mathsf{F})$ の近似 $100(1-\alpha)\%$ 信頼区間となる. すなわち

$$\mathsf{Pr}\Big(S_n - z_{\alpha/2}\widehat{\mathsf{se}}_{\mathrm{boot}} \le \theta(\mathsf{F}) \le S_n + z_{\alpha/2}\widehat{\mathsf{se}}_{\mathrm{boot}}\Big) \approx 1-\alpha$$

となることが期待される. 実際, 適当な仮定のもとで上記の主張を理論的に正当化できることが知られている.

10.4 EM アルゴリズム

EM アルゴリズムは不完全データから最尤推定値を求めるための再帰的繰り返し計算法である. この節では, 実際にはそのすべてが観測されない完全データ $\boldsymbol{X}$ は指数型分布族に含まれる分布から生成されたものと仮定して, 議論を進めていく. $\boldsymbol{X}$ は未観測であるが, その代わり $\boldsymbol{X}$ の関数である $\boldsymbol{Y}$ を観測したと仮定する. すなわち, 単射ではない関数 g があって, $\boldsymbol{Y} = g(\boldsymbol{X})$ と表現できると仮定する. 以下では, $\boldsymbol{X}$ は測度 μ に関する p.d.f.

$$\mathsf{p}(\boldsymbol{x}) = h(\boldsymbol{x})\exp\big\{\eta T(\boldsymbol{x}) - \mathsf{K}^{\vee}(\eta)\big\}$$

を持つとする. ここで, $\mathbb{X}$ と E をそれぞれ $\boldsymbol{X}$ の標本空間と母数空間とし, $T : \mathbb{X} \to \mathbb{R}$, $\mathsf{K}^{\vee} : E \to \mathbb{R}$ とする.

EM アルゴリズムは, 観測データが不完全なとき, 有効である. たとえば, $X_1, X_2, \ldots, X_n$ は i.i.d. ランダム標本とし, $Y_j\,(j = 1, 2, \ldots, n)$ は X_j を小数第 1 位を四捨五入したもの等である. また, EM アルゴリズムは, 欠損データに対してもうまく機能することが知られている. たとえば, ある調査で設問が 2 つあり, おのおの設問では, いくつかの選択肢から回答を選ぶものとする. $\boldsymbol{X}$ は 2 つの設問に対する回答としたとき, Y は 1 つ目の質問をスキップして, 2 つ目だけ回答をしたデータ等である.

いま, $\mathbb{X}$, $\mathbb{Y}$ をそれぞれ $\boldsymbol{X}$ と $\boldsymbol{Y}$ の標本空間とする. $\forall \boldsymbol{y} \in \mathbb{Y}$ に対して, 切断面 $\mathbb{X}(\boldsymbol{y})$ を

$$\mathbb{X}(\boldsymbol{y}) = \big\{\boldsymbol{x} \in \mathbb{X};\, g(\boldsymbol{x}) = \boldsymbol{y}\big\}$$

で定める. このとき

$$\boldsymbol{Y} = \boldsymbol{y} \;\Leftrightarrow\; \boldsymbol{X} \in \mathbb{X}(\boldsymbol{y})$$

である.

命題 10.7. μ を標本空間 $\mathbb{X}$ 上の測度とし, ν を誘導された標本空間 $\mathbb{Y}$ 上の計数測度とする. X は測度 μ に関する p.d.f.

$$\mathsf{p}(x) = h(x)\exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\}$$

をもち, $Y = g(X)$ は離散確率変数とする. ただし, $T : \mathbb{X} \to \mathbb{R}$, $h : \mathbb{X} \to \mathbb{R}$, $\mathsf{K} : E \to \mathbb{R}$ である. このとき, $\mu\otimes\nu$ に関する (X, Y) の同時 p.d.f. は

$$\mathsf{p}^{(X,Y)}(x, y) = \mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\}$$

で与えられる.

Proof. $f : \mathbb{X}\times\mathbb{Y} \to [0, \infty)$ を任意の非負値関数とする. このとき

$$f(X, Y) = \sum_{y\in\mathbb{Y}} f(X, y)\mathbb{1}\{Y = y\}$$

と書き直す. 期待値作用素の線型性と $\mathbb{1}\{g(x) = y\} = \mathbb{1}_{\mathbb{X}(y)}(x)$ であることから

$$\begin{aligned}
\mathsf{E}\bigl[f(X, Y)\bigr] &= \sum_{y\in\mathbb{Y}} \mathsf{E}\bigl[f(X, y)\mathbb{1}\{g(X) = y\}\bigr] \\
&= \sum_{y\in\mathbb{Y}} \int_{\mathbb{X}} f(x, y)\mathbb{1}\{g(x) = y\}h(x)\exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x) \\
&= \sum_{y\in\mathbb{Y}} \int_{\mathbb{X}} f(x, y)\mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\}\,\mathrm{d}\mu(x)
\end{aligned}$$

となる. このことから命題の主張は示せた. □

アルゴリズムを導入するために, 完全データに基づく最尤推定量は $\psi(T)$ であることに注意する. ただし, ψ は $\overset{\bullet}{\mathsf{K}}{}^{\vee} = \dfrac{\mathrm{d}\mathsf{K}^{\vee}}{\mathrm{d}\eta}$ の逆関数である. さらに

$$e(y, \eta) := \mathsf{E}_{\eta}\bigl[T(X)\,|\,Y = y\bigr]$$

とおく. これは $Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}^{X|Y}(\,\cdot\,|\,y)$ に関する積分で計算できる. 命題 10.7 から

$$\begin{aligned}
\mathsf{p}^{X|Y}(x|\,y) &= \frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^{Y}(y|\,\eta)} \\
&= \frac{\mathbb{1}_{\mathbb{X}(y)}(x)h(x)\exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\}}{\mathsf{p}^{Y}(y|\,\eta)}
\end{aligned}$$

である. ただし

$$\mathsf{p}^{Y}(y|\,\eta) = \mathsf{Pr}(Y = y) = \mathsf{Pr}\bigl(X \in \mathbb{X}(y)\bigr)$$

$$= \int_{\mathbb{X}(y)} h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \, \mathrm{d}\mu(x) \tag{10.12}$$

である.

アルゴリズムは, 最尤推定値 $\widehat{\eta}$ に対する初期値 $\widehat{\eta}_0$ から始める. $T(X)$ の条件付き期待値を

$$T_1 = e(Y, \widehat{\eta}_0)$$

と定める. これを **E-Step** と呼ぶ. 更新された値を $\widehat{\eta}_1 = \psi(T_1)$ で定める. これを **M-Step** という. E-Step と M-Step を交互に繰り返し行い, 更新される値が収束するまでこれらの 2 つの Steps を交互に繰り返す.

指数型分布族が正準形ではなく

$$h(x) \exp\bigl\{\eta(\theta) T(x) - \mathsf{K}(\theta)\bigr\} \qquad (\theta \in \Theta = \overset{\bullet}{\mathsf{K}}{}^{\vee}(E))$$

で表現されるとする. すると E-Step は正準形のときと同じで

$$T_{k+1} = \mathsf{E}_{\widehat{\Theta}_k}\bigl[T(X) \,|\, Y = y\bigr]$$

となり, M-Step は関数

$$\Theta \ni \theta \mapsto \eta(\theta) T_{k+1} - \mathsf{K}(\theta)$$

を最大にする値 $\widehat{\theta}_{k+1}$ $(k = 1, 2, \ldots)$ である.

EM アルゴリズムの収束を調べる. そのために, EM アルゴリズムが $\widetilde{\eta}$ に収束したとき, $\widetilde{\eta}$ は関係式

$$\widetilde{\eta} = \psi\bigl(e(Y, \widetilde{\eta})\bigr) \tag{10.13}$$

をみたす. なぜならば, (10.12) から, $\mathsf{p}^Y(y|\,\eta) > 0$ なる y に対して,

$$\begin{aligned}
\frac{\partial}{\partial \eta} \log \mathsf{p}^Y(y|\,\eta) &= \frac{\dfrac{\partial}{\partial \eta} \displaystyle\int_{\mathbb{X}(y)} h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \, \mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(y)} \frac{\partial}{\partial \eta} h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \, \mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \frac{\displaystyle\int_{\mathbb{X}(y)} \bigl\{T(x) - \overset{\bullet}{\mathsf{K}}{}^{\vee}(\eta)\bigr\} h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \, \mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)} \\
&= \int_{\mathbb{X}(y)} T(x) \underbrace{\frac{h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \, \mathrm{d}\mu(x)}{\mathsf{p}^Y(y|\,\eta)}}_{= \mathsf{p}^{X|\,Y}(x|\,y)}
\end{aligned}$$

$$
\begin{aligned}
&- \overset{\bullet}{\mathsf{K}}^{\vee}(\eta) \frac{\displaystyle\int_{\mathbb{X}(y)} h(x) \exp\bigl\{\eta T(x) - \mathsf{K}^{\vee}(\eta)\bigr\} \,\mathrm{d}\mu(x)}{\mathsf{p}^{Y}(y|\,\eta)} \\
&= \mathsf{E}_{\eta}\bigl[T(X)|\, Y = y\bigr] - \overset{\bullet}{\mathsf{K}}^{\vee}(\eta) \\
&= e(y,\, \eta) - \overset{\bullet}{\mathsf{K}}^{\vee}(\eta)
\end{aligned}
$$

となることからわかる.

さらに, ψ は $\overset{\bullet}{\mathsf{K}}^{\vee}$ の逆写像であることに注意すると (10.13) は

$$
\overset{\bullet}{\mathsf{K}}^{\vee}\bigl(\widetilde{\eta}\bigr) = e\bigl(Y,\, \widetilde{\eta}\bigr) \tag{10.14}
$$

と同値になる.

例 10.8. $X_1, X_2, \ldots, Xn \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\eta}$ とする. ただし

$$
\mathsf{p}_{\eta}(x) = \begin{cases} \eta e^{-\eta x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}
$$

である. いま

$$
Y_j = \lfloor X_j \rfloor \qquad (j = 1,\, 2,\, \ldots,\, n)
$$

とする. すなわち, Y_j は X_j を超えない最大の整数である. $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)^{\top}$ の同時 p.d.f. は正準母数 η と十分統計量

$$
T := T(\boldsymbol{X}) = -(X_1 + X_2 + \cdots + X_n)
$$

の指数型分布族を作る. $\mathsf{K}^{\vee}(\eta) = -n \log \eta$ であることに注意すると, $\psi(x) = -\dfrac{n}{x}$ となる. よって, $\boldsymbol{X}$ に基づく η の最尤推定量は $\psi(T) = -\dfrac{n}{T}$ である. $\boldsymbol{Y} = (Y_1,\, Y_2,\, \ldots,\, Y_n)^{\top}$ と $\boldsymbol{y} = (y_1,\, y_2,\, \ldots,\, y_n)^{\top}$ とおく. 命題 10.7 から $j = 1,\, 2,\, \ldots,\, n$ に対して

$$
\begin{aligned}
\mathsf{E}_{\eta}\bigl[X_j|\, \boldsymbol{Y} = \boldsymbol{y}\bigr] &= \mathsf{E}_{\eta}\bigl[X_j|\, y_j \leq X_j < y_j + 1\bigr] \\
&= \frac{\displaystyle\int_{y_j}^{y_j+1} x\eta e^{-\eta x} \,\mathrm{d}x}{\displaystyle\int_{y_j}^{y_j+1} \eta e^{-\eta x} \,\mathrm{d}x} \\
&= \frac{\Bigl[-xe^{-\eta x}\Bigr]_{y_j}^{y_j+1} + \displaystyle\int_{y_j}^{y_j+1} \eta e^{-\eta x} \,\mathrm{d}x}{\Bigl[-e^{-\eta x}\Bigr]_{y_j}^{y_j+1}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{e^{-\eta y_j}\{y_j - (1+y_j)e^{-\eta} - \dfrac{e^{-\eta y_j}}{\eta}\{e^{-\eta}-1\}}{e^{-\eta y_j}\{1-e^{-\eta}\}} \\
&= \frac{y_j\{1-e^{-\eta}\} - e^{-\eta} - \dfrac{e^{-\eta}-1}{\eta}\{e^{-\eta}-1\}}{1-e^{-\eta}} \\
&= y_j + \frac{1-e^{-\eta}-\eta e^{-\eta}}{\eta\{1-e^{-\eta}\}} \\
&= y_j + \frac{e^{\eta}-1-\eta}{\eta\{e^{\eta}-1\}}
\end{aligned}
$$

となる. さらに, 独立性から

$$
\mathsf{E}_\eta\big[X_j \big| \boldsymbol{Y} = \boldsymbol{y}\big] = \mathsf{E}_\eta\big[X_j \big| Y_j = y_j\big]
$$

となるので

$$
\begin{aligned}
e(\boldsymbol{y}, \eta) &= \mathsf{E}_\eta\big[T(\boldsymbol{X}) \big| \boldsymbol{Y} = \boldsymbol{y}\big] \\
&= -\sum_{j=1}^{n} \mathsf{E}_\eta\big[X_j \big| Y_j = y_j\big] \\
&= -\sum_{j=1}^{n} \left\{ y_j + \frac{e^{\eta}-1-\eta}{\eta\{e^{\eta}-1\}} \right\} \\
&= -n\left\{ \overline{y}_n + \frac{e^{\eta}-1-\eta}{\eta\{e^{\eta}-1\}} \right\}
\end{aligned}
$$

となる. ただし, $\overline{y}_n = \dfrac{1}{n}\sum_{j=1}^{n} y_j$ である. よって, k-Step($k = 1, 2, \ldots$) の推定値を $\widehat{\eta}_k$ と書いたとき, EM アルゴリズムは

$$
\widehat{\eta}_{k+1} = -\frac{n}{T_{k+1}} \quad \text{かつ} \quad T_{k+1} = -n\left\{ \overline{y}_n + \frac{e^{\widehat{\eta}_k}-1-\widehat{\eta}_k}{\widehat{\eta}_k\{e^{\widehat{\eta}_k}-1\}} \right\} \tag{10.15}
$$

である. □

注意 10.9. 例 10.8 では, Y_j の p.m.f. を陽に計算できる. 実際, $y = 0, 1, 2, \ldots$ に対して

$$
\begin{aligned}
\mathsf{Pr}_\eta\big(Y_j = y\big) &= \mathsf{Pr}_\eta\big(y \le X_j < y+1\big) \\
&= \int_y^{y+1} \eta e^{-\eta x}\,\mathrm{d}x \\
&= \Big[-e^{-\eta x}\Big]_y^{y+1} = -e^{-\eta(y+1)} + e^{-\eta y}
\end{aligned}
$$

$$= \big(1 - e^{-\eta}\big)\big(e^{-\eta}\big)^{y}$$

となる. したがって, $Y_1, Y_2, \ldots, Y_n$ は母数 $\pi := 1 - e^{-\eta}$ の幾何分布からのランダム標本とみなせる. よって, π の最尤推定量は

$$\widehat{\pi} = \frac{1}{1 + \overline{Y}_n}; \qquad \overline{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

で与えられる. $\eta = -\log(1-\pi)$ なので, η の最尤推定量は

$$\widehat{\eta} = -\log\big(1 - \widehat{\pi}\big) = \log\left(1 + \frac{1}{\overline{Y}_n}\right)$$

である.

EM アルゴリズムの繰り返し $\{\widehat{\eta}_k\}_{k=1}^{\infty}$ の $\widehat{\eta}$ への収束のスピードを調べてみよう. そのために, $\widehat{\eta}_k = \widehat{\eta} + \epsilon$ と書く. すると $\widehat{\eta}$ は繰り返し計算の固定点なので

$$-\frac{n}{\widehat{\eta}} = -n\left\{\overline{y}_n + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}\right\}$$

をみたしていること [3] に注意する. ここで, 次の Taylor 展開を考える.

$$\frac{1}{\widehat{\eta} + \epsilon} = \frac{1}{\widehat{\eta}} - \frac{\epsilon}{\widehat{\eta}^2} + O(\epsilon^2),$$
$$\frac{1}{e^{\widehat{\eta}+\epsilon} - 1} = \frac{1}{e^{\widehat{\eta}} - 1} - \frac{e^{\widehat{\eta}}\epsilon}{\big(e^{\widehat{\eta}} - 1\big)^2} + O(\epsilon^2).$$

これらより

$$\begin{aligned} T_{k+1} &= -n\left\{\overline{y}_n + \frac{e^{\widehat{\eta}_k} - 1 - \widehat{\eta}_k}{\widehat{\eta}_k\big(e^{\widehat{\eta}_k} - 1\big)}\right\} \\ &= -n\left\{\overline{y}_n + \frac{1}{\widehat{\eta}_k} - \frac{1}{e^{\widehat{\eta}_k} - 1}\right\} \\ &= -n\left\{\underbrace{\overline{y}_n + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}}_{=\frac{1}{\widehat{\eta}}} - \frac{\epsilon}{\widehat{\eta}^2} + \frac{\epsilon e^{\widehat{\eta}}}{\big(e^{\widehat{\eta}} - 1\big)^2} + O(\epsilon^2)\right\} \end{aligned}$$

---

[3] $\lim_{k\to\infty} T_k = T_\infty$ とおくと $\widehat{\eta} = -\frac{n}{T_\infty}$ となる. したがって, (10.15) から

$$-\frac{n}{\widehat{\eta}} = \lim_{k\to\infty} T_k = -n\left\{\overline{y}_n + \frac{\exp\big\{\lim_{k\to\infty}\widehat{\eta}_k\big\} - 1 - \lim_{k\to\infty}\widehat{\eta}_k}{\lim_{k\to\infty}\widehat{\eta}_k\left\{\exp\big\{\lim_{k\to\infty}\widehat{\eta}_k\big\} - 1\right\}}\right\} = -n\left\{\overline{y}_n + \frac{e^{\widehat{\eta}} - 1 - \widehat{\eta}}{\widehat{\eta}\big\{e^{\widehat{\eta}} - 1\big\}}\right\}$$
$$= -n\left\{\overline{y}_n + \frac{1}{\widehat{\eta}} - \frac{1}{e^{\widehat{\eta}} - 1}\right\}$$

からわかる.

$$= -\frac{n}{\widehat{\eta}}\left\{1 - \left(\frac{\epsilon}{\widehat{\eta}} - \frac{\epsilon\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right) + O(\epsilon^2)\right\}$$

を得る. この式から

$$\begin{aligned}\widehat{\eta}_{k+1} = -\frac{n}{T_{k+1}} &= \frac{\widehat{\eta}}{1 - \epsilon\left(\frac{1}{\widehat{\eta}} - \frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right) + O(\epsilon^2)} \\ &= \widehat{\eta}\left\{1 + \epsilon\left(\frac{1}{\widehat{\eta}} - \frac{\widehat{\eta}e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right) + O(\epsilon^2)\right\} \\ &= \widehat{\eta} + \epsilon\left\{1 - \frac{\widehat{\eta}^2e^{\widehat{\eta}}}{\left(e^{\widehat{\eta}}-1\right)^2}\right\} + O(\epsilon^2)\end{aligned}$$

を得る. 特に, $\widehat{\eta}_k = \widehat{\eta}$ のとき, $\epsilon = 0$ なので, $\widehat{\eta}_{k+1} = \widehat{\eta}$ となる. したがって, $\widehat{\eta}$ は繰り返しの固定点であることがわかる. □

注意 10.10. 一般に EM アルゴリズムは安定しており, 信頼される. このアルゴリズムの良い性質は, 繰り返しをするごとに尤度関数を増加させることである. すなわち, EM アルゴリズムの $k - Step$ の推定値を $\widehat{\eta}_k\,(k = 1, 2, \ldots)$ としたとき

$$\prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_{k+1}) \geq \prod_{j=1}^{n} \mathsf{p}^Y(y_j|\,\widehat{\eta}_k)$$

となることである. これは, EM アルゴリズムが MM アルゴリズムの特別な場合であることからわかる. この点については Lange (2014) を参照のこと. しかし, 収束は一般に保証されない. たとえば, 尤度関数のモードが複数ある場合には, EM アルゴリズムは収束しないことがある. EM アルゴリズムが収束するための十分条件は Wu (1983) を参照のこと. EM アルゴリズムは安定しているが, 収束のスピードは遅くことが知られている. □

第11章　モデル選択

11.1　モデル選択の目標

11.2　罰則法: AIC

11.3　Bayes モデル選択

11.4　交差検証法

11.5　章末注釈と参考文献

11.6　演習問題

第12章　大標本理論

この章では, 正則母数モデルにおける最尤推定量の一致性と漸近正規性について説明する. 節 12.1 では, 準備として, ランダム関数の確率収束について説明する. 尤度関数ないし対数尤度関数を未知母数の関数とみなして議論をするためのものである. 節 12.2 では, 前節の結果を利用して, 最尤推定量の一致性を証明する. 節 12.3 では, 最尤推定量の漸近正規性を証明する. 節 12.4 では, 帰無仮説のもとで尤度比検定統計量の対数の -2 倍が χ^2 分布に分布収束することを主張する Wilks の定理を証明する.

12.1　ランダム関数の確率収束

$K \subset \mathbb{R}^d$ をコンパクト集合として, $\{X_n\}_{n=1}^{\infty}$ を i.i.d. 確率変数列とする. さらに, $h\colon K \times \mathbb{R} \to \mathbb{R}$ を各 $x \in \mathbb{R}$ に対して

$$h(\,\cdot\,, x)\colon K \ni \boldsymbol{t} \mapsto h(\boldsymbol{t}, x) \in \mathbb{R}$$

が連続となるような関数とする. つぎに, $n = 1, 2, \ldots$ に対して

$$W_n(\boldsymbol{t}) := h(\boldsymbol{t}, X_n) \qquad (\boldsymbol{t} \in K)$$

と定めると $\{W_n\}_{n=1}^{\infty}$ は K 上の i.i.d. ランダム連続関数列となる. したがって, $W_n\,(n \in \mathbb{N})$ は $C(K)$ 値ランダム要素とみなすことができる. ただし, $C(K)$ は K 上の連続関数全体が成す集合である.

集合 $C(K)$ に値を取るランダム要素はベクトルと同じように和, 差とスカラー倍が定義される. このような性質を持つ空間を **線型空間** または **ベクトル空間** という. さらに, 線型空間 $C(K)$ に収束の概念も導入できる. $w \in C(K)$ に対して, w のノルム $\|w\|_\infty$ を

$$\|w\|_\infty := \sup_{\boldsymbol{t} \in K} |w(\boldsymbol{t})|$$

で定める. これを w の **sup ノルム** と呼ぶ. 関数列 $w, w_n \in C(K)\,(n = 1, 2, \ldots)$ に対して, $\{w_n\}_{n=1}^{\infty}$ は w に収束するとは

$$\|w_n - w\|_\infty \overset{n\to\infty}{\longrightarrow} 0$$

で定める. 線型空間 $C(K)$ はノルム $\|\cdot\|_\infty$ に関して完備[1] となることが知られている. 完備な線型ノルム空間を Banach 空間 と呼ぶ. Banach 空間 $C(K)$ の最後のよい性質は可分性である. すなわち, 線型ノルム空間 $C(K)$ は稠密[2] な可算部分集合を持つ.

補題 12.1. $K \subset \mathbb{R}^d$ をコンパクトな部分集合とし, W を $C(K)$ 値ランダム要素とし

$$\mu(\boldsymbol{t}) = \mathsf{E}[W(\boldsymbol{t})] \qquad (\boldsymbol{t} \in K)$$

と定める. このとき, 以下のことが成立する.
(1) $\mathsf{E}[\|W\|_\infty] < \infty$ のとき, μ は K 上の連続関数となる.
(2) $\mathsf{E}[\|W\|_\infty] < \infty$ とする. このとき, 任意の $\epsilon > 0$ に対して

$$\sup_{\boldsymbol{t}\in K} \mathsf{E}\left[\sup_{\boldsymbol{s}\in K;\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\epsilon} |W(\boldsymbol{s}) - W(\boldsymbol{t})|\right] \xrightarrow{\epsilon\to 0} 0$$

となる. ただし, $|\cdot|_{2,\,d}$ は $\mathbb{R}^d$ の Euclid ノルムである.

Proof. (1) の証明: $\boldsymbol{t}_n, \boldsymbol{t} \in K$ $(n = 1, 2, \ldots)$ とし

$$\lim_{n\to\infty} \boldsymbol{t}_n = \boldsymbol{t}$$

とする. ランダム関数 W は連続だから

$$W(\boldsymbol{t}_n) \xrightarrow{\text{a.s.}} W(\boldsymbol{t}) \qquad (n \to \infty)$$

となる. さらに, $|W(\boldsymbol{t}_n)| \le \|W\|_\infty$ かつ $\mathsf{E}[\|W\|_\infty] < \infty$ であることに注意して, 優収束定理を用いると

$$\lim_{n\to\infty} \mu(\boldsymbol{t}_n) = \lim_{n\to\infty} \mathsf{E}[W(\boldsymbol{t}_n)] = \mathsf{E}\Big[\lim_{n\to\infty} W(\boldsymbol{t}_n)\Big] = \mathsf{E}[W(\boldsymbol{t})] = \mu(\boldsymbol{t})$$

となる. 点列 $\{\boldsymbol{t}_n\}_{n=1}^\infty$ は任意だったので, μ は $\boldsymbol{t}$ で連続であることがわかる. さらに, $\boldsymbol{t} \in K$ も任意だったので, μ は K 上で連続であることがわかる.
(2) の証明: W の見本路を 1 つ固定し

$$w(\boldsymbol{t}) := W(\omega)(\boldsymbol{t}) \qquad (\omega \in \Omega, \boldsymbol{t} \in K)$$

[1] $C(K)$ の任意の Cauchy 列は $C(K)$ のある元に必ず収束することである.
[2] $C(K)$ の可算部分集合 $\{w_i;\, i \in \mathbb{N}\}$ で, 任意の $\epsilon > 0$ と $w \in C(K)$ に対して, ある $i_0 \in \mathbb{N}$ が存在して

$$\|w - w_{i_0}\|_\infty < \epsilon$$

となることをいう.

と記す. $w \in C(K)$ であり, K はコンパクトなので, w は K 上で一様連続である. $\epsilon > 0$ に対して

$$\tau_w(\epsilon) := \sup_{\boldsymbol{u},\, \boldsymbol{v} \in K;\, |\boldsymbol{u}-\boldsymbol{v}|_{2,\,d} < \epsilon} \bigl| w(\boldsymbol{u}) - w(\boldsymbol{v}) \bigr|$$

と定める. このとき, w は一様連続だから, $\tau_w(\epsilon) \downarrow 0\, (\epsilon \downarrow 0)$ となる. 一方, $M_\epsilon(\boldsymbol{t}) = \sup_{\boldsymbol{s} \in K;\, |\boldsymbol{t}-\boldsymbol{s}|_{2,\,d} < \epsilon} \bigl| W(\boldsymbol{s}) - W(\boldsymbol{t}) \bigr|$ だったので, 固定した $\boldsymbol{t} \in K$ に対して

$$M_\epsilon(\boldsymbol{t}) \leq \tau_W(\epsilon)$$

である. したがって

$$\sup_{\boldsymbol{t} \in K} \mathsf{E}\bigl[M_\epsilon(\boldsymbol{t}) \bigr] \leq \mathsf{E}\bigl[\tau_W(\epsilon) \bigr]$$

である. また

$$0 \leq \tau_W(\epsilon) \leq 2 \| W \|_\infty$$

である. 仮定から $\mathsf{E}\bigl[\| W \|_\infty \bigr] < \infty$ だったので

$$0 \leq \mathsf{E}\bigl[\tau_W(\epsilon) \bigr] \leq 2 \mathsf{E}\bigl[\| W \|_\infty \bigr] < \infty$$

である. さらに, 各見本路について

$$\tau_{W(\omega)}(\epsilon) \downarrow 0\, (\epsilon \to 0) \qquad (\omega \in \Omega)$$

である. したがって, 優収束定理から

$$\mathsf{E}\bigl[\tau_W(\epsilon) \bigr] \xrightarrow{\epsilon \to 0} 0$$

がわかる. よって

$$\sup_{\boldsymbol{t} \in K} \mathsf{E}\left[\sup_{\boldsymbol{s} \in K;\, |\boldsymbol{s}-\boldsymbol{t}|_{2,\,d} < \epsilon} |W(\boldsymbol{s}) - W(\boldsymbol{t})| \right] \xrightarrow{\epsilon \to 0} 0$$

が証明された.

□

以下の定理は, i.i.d. ランダム関数の標本平均がその期待値に一様に収束に関する議論をする. 各点の収束は大数の弱法則から直ちにわかる. さらに, 次の定理は, コンパクト集合 K 上でランダム標本の平均がその期待値に一様に確率収束することを主張している.

定理 12.2. $K \subset \mathbb{R}^d$ をコンパクト部分集合とする. $W, W_1, W_2, \ldots$ を $C(K)$ 値 i.i.d. ランダム要素とする. さらに

$$\mu(\boldsymbol{t}) = \mathsf{E}\big[W(\boldsymbol{t})\big] \qquad (\boldsymbol{t} \in K)$$

とおき

$$\mathsf{E}\big[\|W\|_\infty\big] < \infty$$

を仮定する. ただし, $\|W\|_\infty = \sup_{\boldsymbol{t}\in K}|W(\boldsymbol{t})|$ である. $n \in \mathbb{N}$ に対して

$$\overline{W}_n := \frac{1}{n}\big\{W_1 + W_2 + \cdots + W_n\big\}$$

とおいたとき

$$\big\|\overline{W}_n - \mu\big\|_\infty \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{12.1}$$

が成り立つ.

Proof. $\epsilon > 0$ を固定する. $\delta > 0$ と $j = 1, 2, \ldots, n$ に対して

$$M_{\delta,\,j}(\boldsymbol{t}) := \sup_{\boldsymbol{s}\in K:\,|\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\delta} \big|W_j(\boldsymbol{s}) - W_j(\boldsymbol{t})\big|; \qquad \lambda_{\delta,\,j}(\boldsymbol{t}) := \mathsf{E}\big[M_{\delta,\,j}(\boldsymbol{t})\big]$$

と定める. 補題 12.1 から δ を十分小さくすると

$$\lambda_{\delta,\,j}(\boldsymbol{t}) = \mathsf{E}\bigg[\sup_{\boldsymbol{s}\in K:\,|\boldsymbol{s}-\boldsymbol{t}|_{2,\,d}<\delta} \big|W_j(\boldsymbol{s}) - W_j(\boldsymbol{t})\big|\bigg] \le \epsilon \qquad (\boldsymbol{t} \in K)$$

とできる. ここで

$$B(\boldsymbol{t};\,\delta) = \big\{\boldsymbol{s} \in K;\,|\boldsymbol{s} - \boldsymbol{t}|_{2,\,d} < \delta\big\}$$

とおく. すると K はコンパクトなので, K の開被覆 $\big\{B(\boldsymbol{t};\,\delta)\big\}_{\boldsymbol{t}\in K}$ から有限被覆を取ることができる. すなわち, ある $m \in \mathbb{N}$ と $\{\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_m\} \subset K$ が存在して

$$K \subset \bigcup_{j\in\{1,\,2,\,\ldots,\,m\}} B(\boldsymbol{t}_j;\,\delta)$$

となる. $O_j = B(\boldsymbol{t}_j;\,\delta)\,(j = 1, 2, \ldots, m)$ とおく. このとき

$$\begin{aligned}\big\|\overline{W}_n - \mu\big\|_\infty &= \max_{j\in\{1,\,2,\,\ldots,\,m\}}\Big\{\sup_{\boldsymbol{t}\in O_j}\big|\overline{W}_n(\boldsymbol{t}) - \mu(\boldsymbol{t})\big|\Big\} \\ &\le \max_{j\in\{1,\,2,\,\ldots,\,m\}}\Big\{\sup_{\boldsymbol{t}\in O_j}\big|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_j)\big| + \big|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\big|\end{aligned}$$

$$
\begin{aligned}
&\qquad + \left|\mu(\boldsymbol{t}_j) - \mu(\boldsymbol{t})\right| \Bigg\} \\
&\leq \max_{j\in\{1,2,\ldots,m\}} \left\{ \sup_{\boldsymbol{t}\in O_j} \left|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_j)\right| \right\} \\
&\qquad + \max_{j\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\right| + \epsilon \qquad (\mu \text{ の連続性})
\end{aligned}
$$

となる. さらに

$$
\begin{aligned}
\sup_{\boldsymbol{t}\in O_j} \left|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_j)\right| &= \frac{1}{n} \sup_{\boldsymbol{t}\in O_j} \left| \sum_{k=1}^{n} \bigl(W_k(\boldsymbol{t}) - W_k(\boldsymbol{t}_j)\bigr) \right| \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \sup_{\boldsymbol{t}\in O_j} \left|W_k(\boldsymbol{t}) - W_k(\boldsymbol{t}_j)\right| \\
&= \frac{1}{n} \sum_{k=1}^{n} M_{\delta,\,k}(\boldsymbol{t}_j) =: \overline{M}_{\delta,\,j}
\end{aligned}
$$

となる. $j = 1, 2, \ldots, m$ に対して, 大数の弱法則から

$$
\overline{M}_{\delta,\,j} \xrightarrow{\mathrm{P}} \lambda_\delta(\boldsymbol{t}_j) < \epsilon \qquad (n \to \infty)
$$

が成立する. よって

$$
\mathsf{Pr}\left(\overline{M}_{\delta,\,j} < \epsilon \right) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.2}
$$

がわかる. さらに, 大数の弱法則から

$$
\mathsf{Pr}\left(\max_{i\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)\right| < \epsilon \right) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.3}
$$

したがって, $\bigcap_{i=1}^{n} \{\overline{M}_{\delta,\,i} < \epsilon\}$ と $\sup_{i\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_i) - \mu(\boldsymbol{t}_i)\right| < \epsilon$ が起こったとき

$$
\begin{aligned}
\left\|\overline{W}_n - \mu\right\|_\infty &\leq \max_{i\in\{1,2,\ldots,m\}} \Bigg\{ \underbrace{\sup_{\boldsymbol{t}\in O_i} \left|\overline{W}_n(\boldsymbol{t}) - \overline{W}_n(\boldsymbol{t}_i)\right|}_{\leq \overline{M}_{\delta,\,i}(\boldsymbol{t}_i)} \Bigg\} \\
&\qquad + \max_{j\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\right| + \epsilon \\
&\leq 3\epsilon
\end{aligned}
$$

が成り立つ. したがって, (12.2) と (12.3) と合わせると

$$
\mathsf{Pr}\left(\left\|\overline{W}_n - \mu\right\|_\infty \leq 3\epsilon \right)
$$

$$\geq \mathsf{Pr}\left(\{\overline{M}_{\delta,i} < \epsilon\} \bigcap \left\{ \sup_{j\in\{1,2,\ldots,m\}} \left|\overline{W}_n(\boldsymbol{t}_j) - \mu(\boldsymbol{t}_j)\right| < \epsilon \right\}\right)$$
$$\overset{n\to\infty}{\longrightarrow} 1$$

となる. よって

$$\mathsf{Pr}\left(\left\|\overline{W}_n - \mu\right\|_\infty > 3\epsilon \right) \overset{n\to\infty}{\longrightarrow} 0$$

がわかる. □

注意 12.3. 大数の強法則から

$$\left\|\overline{W}_n - \mu\right\|_\infty \overset{\text{a.s.}}{\to} 0 \qquad (n \to \infty)$$

も同様に証明できる. □

定理 12.4. $K \subset \mathbb{R}^d$ をコンパクト部分集合とし, $\{G_n\}_{n=1}^\infty$ を $C(K)$ 値ランダム要素とする. ある非ランダムな関数 $g\,(\in C(K))$ が存在して

$$\left\|G_n - g\right\|_\infty \overset{\text{P}}{\to} 0 \qquad (n \to \infty) \tag{12.4}$$

が成立すると仮定する. このとき, 以下が成立する.
(1) 確率ベクトル列 $\{\boldsymbol{t}_n\}_{n=1}^\infty \subset K$ はある非ランダムベクトル $\boldsymbol{t}^*\,(\in K)$ に確率収束するとする. このとき

$$G_n(\boldsymbol{t}_n) \overset{\text{P}}{\to} g(\boldsymbol{t}^*) \qquad (n \to \infty)$$

が成立する.
(2) 非ランダムな関数 g は $\boldsymbol{t}^*\,(\in K)$ で唯一の最大値を取るとする. さらに, 各 n に対して, 確率ベクトル $\boldsymbol{t}_n\,(n = 1, 2, \ldots)$ で G_n は最大を取るとする. すなわち

$$G_n(\boldsymbol{t}_n) = \sup_{\boldsymbol{t}\in K} G_n(\boldsymbol{t})$$

である. このとき

$$\boldsymbol{t}_n \overset{\text{P}}{\to} \boldsymbol{t}^* \qquad (n \to \infty)$$

が成立する.
(3) $K \subset \mathbb{R}$ とし, $t^* \in K$ は方程式 $g(t) = 0$ の唯一の解とする. 各 n に対して, 確率変数 t_n は方程式 $G_n(t_n) = 0$ の解とする. このとき

$$t_n \overset{\text{P}}{\to} t^* \qquad (n \to \infty)$$

が成立する.

Proof. (1) の証明: まず

$$\begin{aligned}\left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| &\leq \left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}_n)\right| + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \\ &\leq \left\|G_n - g\right\|_\infty + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \end{aligned} \tag{12.5}$$

となることに注意する. $\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t}^*\,(n \to \infty)$ と g の連続性から

$$g(\boldsymbol{t}_n) \xrightarrow{\mathrm{P}} g(\boldsymbol{t}^*) \qquad (n \to \infty)$$

となる. したがって, (12.5) から $\forall \epsilon > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\left(\left|G_n(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \geq \epsilon\right) &\leq \mathsf{Pr}\left(\left\|G_n - g\right\|_\infty + \left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \geq \epsilon\right) \\ &\leq \mathsf{Pr}\left(\left\|G_n - g\right\|_\infty \geq \frac{\epsilon}{2}\right) + \mathsf{Pr}\left(\left|g(\boldsymbol{t}_n) - g(\boldsymbol{t}^*)\right| \geq \frac{\epsilon}{2}\right) \\ &\xrightarrow{n\to\infty} 0\end{aligned}$$

がわかる. よって, (1) が示された.
(2) の証明: $\epsilon > 0$ を固定して

$$K_\epsilon := K \setminus B(\boldsymbol{t}^*;\, \epsilon) = K \bigcap \left\{B(\boldsymbol{t}^*;\, \epsilon)\right\}^c \tag{12.6}$$

とおく. ただし, $B(\boldsymbol{t}^*;\, \epsilon) = \left\{\boldsymbol{t} \in \mathbb{R}^d;\, \left|\boldsymbol{t} - \boldsymbol{t}^*\right|_{2,\,d} < \epsilon\right\}$ である. K は有界なので, K_ϵ も有界である. また, K_ϵ は 2 つの閉集合の共通部分なので, 閉集合である. これらのことから K_ϵ はコンパクトであることがわかる. ここで

$$M := g(\boldsymbol{t}^*) = \sup_{\boldsymbol{t}\in K} g(\boldsymbol{t}); \qquad M_\epsilon := \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t})$$

とおく. K_ϵ のコンパクト性と g の連続性から, ある $\boldsymbol{t}_\epsilon^* \in K_\epsilon$ が存在して

$$M_\epsilon = g(\boldsymbol{t}_\epsilon^*)$$

とできる. g は K 上で唯一の点 $\boldsymbol{t}^*$ で最大値を取るので, $\boldsymbol{t}_\epsilon^* \neq \boldsymbol{t}^*$ である. よって

$$M_\epsilon < M$$

となる. ここで

$$\delta := M - M_\epsilon > 0$$

とおく. $\left\|G_n - g\right\|_\infty < \dfrac{\delta}{2}$ 上では

$$\sup_{\boldsymbol{t}\in K_\epsilon} G_n(\boldsymbol{t}) \leq \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t}) + \sup_{\boldsymbol{t}\in K_\epsilon} \left|G_n(\boldsymbol{t}) - g(\boldsymbol{t})\right|$$

$$\leq \sup_{\boldsymbol{t}\in K_\epsilon} g(\boldsymbol{t}) + \left\|G_n - g\right\|_\infty$$
$$\leq M_\epsilon + \frac{\delta}{2} = M - \frac{\delta}{2} \tag{12.7}$$

となる. 一方, $\left\|G_n - g\right\|_\infty < \dfrac{\delta}{2}$ 上では

$$\sup_{\boldsymbol{t}\in K} G_n(\boldsymbol{t}) \geq G_n(\boldsymbol{t}^*) = g(\boldsymbol{t}^*) + G_n(\boldsymbol{t}^*) - g(\boldsymbol{t}^*)$$
$$\geq g(\boldsymbol{t}^*) - \sup_{\boldsymbol{t}\in K}\left|G_n(\boldsymbol{t}) - g(\boldsymbol{t})\right|$$
$$\geq g(\boldsymbol{t}^*) - \left\|G_n - g\right\|_\infty \tag{12.8}$$
$$> g(\boldsymbol{t}^*) - \frac{\delta}{2} = M - \frac{\delta}{2} \tag{12.9}$$

となる. したがって, (12.7) と (12.9) から

$$\left\|G_n - g\right\|_\infty < \frac{\delta}{2} \Rightarrow \sup_{\boldsymbol{t}\in K} G_n(\boldsymbol{t}) > \sup_{\boldsymbol{t}\in K_\epsilon} G_n(\boldsymbol{t})$$

となる. このことと K_ϵ の定義 (12.6) から

$$\left\|G_n - g\right\|_\infty < \frac{\delta}{2} \Rightarrow \boldsymbol{t}_n \in B(\boldsymbol{t}^*;\, \epsilon)$$

がわかる. よって

$$\mathsf{Pr}\left(\left\|G_n - g\right\|_\infty < \frac{\delta}{2}\right) \leq \mathsf{Pr}\left(\left|\boldsymbol{t}_n - \boldsymbol{t}^*\right|_{2,\,d} < \epsilon\right)$$

がわかる. 上の不等式の両辺の事象の補事象を取り, $\left\|G_n - g\right\|_\infty \xrightarrow{\mathrm{P}} 0$ に注意すると

$$\mathsf{Pr}\left(\left|\boldsymbol{t}_n - \boldsymbol{t}^*\right|_{2,\,d} \geq \epsilon\right) \leq \mathsf{Pr}\left(\left\|G_n - g\right\|_\infty \geq \frac{\delta}{2}\right) \overset{n\to\infty}{\longrightarrow} 0$$

がわかる. よって, $\boldsymbol{t}_n \xrightarrow{\mathrm{P}} \boldsymbol{t}^*$ が示せた.
(3) の証明: ⑶ⓐ 準備の補題: 任意の $\epsilon > 0$ に対して

$$\inf_{t\in K;\,|t-t^*|\geq\epsilon}\left|g(t)\right| > 0 \tag{12.10}$$

となることを確認する. $K_\epsilon := \{t \in K;\, |t - t^*| \geq \epsilon\}$ とおく. K はコンパクトなので, K_ϵ もコンパクトである. 関数 $|g|$ は K_ϵ 上でも連続なので, K_ϵ 上で最小値をもつ. しかし, $g(t) = 0$ をみたす唯一の零点 $t^* \notin K_\epsilon$ なので, (12.10) が示せた.

㉟ (3) の証明: 固定した $\epsilon > 0$ に対して, $c_\epsilon := \inf_{t\in K;\,|t-t^*|\geq\epsilon} \bigl|g(t)\bigr| > 0$ とおく. $G_n(t_n) = 0$ なので

$$\bigl|g(t_n)\bigr| = \bigl|g(t_n) - G_n(t_n)\bigr| \leq \sup_{t\in K} \bigl|g(t) - G_n(t)\bigr| = \|G_n - g\|_\infty$$

となる. よって

$$|t - t^*| \geq \epsilon \;\Rightarrow\; \|G_n - g\|_\infty \geq c_\epsilon$$

がわかる. すなわち, $\{t_n - t^*\} \subset \{\|G_n - g\|_\infty \geq c_\epsilon\}$ となる. 補題 1.7(4) と定理の仮定 (12.4) から

$$\mathsf{Pr}\bigl(|t_n - t^*| \geq \epsilon\bigr) \leq \mathsf{Pr}\bigl(\|G_n - g\|_\infty \geq c_\epsilon\bigr) \overset{n\to\infty}{\longrightarrow} 0$$

がわかる. よって, (3) の主張は証明された. □

12.2 最尤推定量の一致性

$\Theta \subset \mathbb{R}^d$ をコンパクト部分集合とし, 正則母数モデル $\mathcal{P} = \{\mathsf{p}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta\}$ を考える. ただし, $\mathsf{p}_{\boldsymbol{\theta}}$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. $\boldsymbol{\theta}^* \in \Theta$ を固定して

$$X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}$$

とする. $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n) = \boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$ を観測したときの対数尤度 $\ell_n(\boldsymbol{\theta})$ を

$$\ell_n(\boldsymbol{\theta}) = \log\left(\prod_{i=1}^n \mathsf{p}_{\boldsymbol{\theta}}(x_i)\right) = \sum_{i=1}^n \log \mathsf{p}_{\boldsymbol{\theta}}(x_i)$$

で定める. このとき, 関数 $\Theta \ni \boldsymbol{\theta} \mapsto \ell_n(\boldsymbol{\theta})$ を最大にする $\boldsymbol{\theta}$ の値 $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{x})$ を <u>最尤推定値</u> という. 変数ベクトルに $\boldsymbol{X}$ を形式的に代入したものを $\widehat{\boldsymbol{\theta}}_n := \widehat{\boldsymbol{\theta}}_n(\boldsymbol{X})$ を <u>最尤推定量</u> という. 統計的モデルの正則性から, ほとんど至るところの $x \in \mathbb{R}$ に対して, 写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}_{\boldsymbol{\theta}}(x) \in [0,\, \infty)$$

は連続である.

定義 12.5. $\mathsf{P} = \{\mathsf{p}_{\boldsymbol{\theta}};\, \boldsymbol{\theta} \in \Theta\}$ を統計的モデルとする. $\boldsymbol{\theta}^* \in \Theta$ を真の母数とする. $\boldsymbol{\theta} \in \Theta$ に対する <u>Kullback–Leibler 情報量</u> $\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta})$ を

$$\mathsf{KL}(\boldsymbol{\theta}^*,\, \boldsymbol{\theta}) := \mathsf{E}_{\boldsymbol{\theta}^*}\left[\log\left\{\frac{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}{\mathsf{p}_{\boldsymbol{\theta}}(X)}\right\}\right]$$

で定める.

補題 12.6. $\forall \boldsymbol{\theta} \in \Theta$ に対して, $\mathsf{P}_{\boldsymbol{\theta}^*} \neq \mathsf{P}_{\boldsymbol{\theta}}$ ならば, $\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > 0$ である.

Proof. Jensen の不等式 (定理 3.12) から

$$\begin{aligned}
-\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) &= \mathsf{E}_{\boldsymbol{\theta}^*}\left[\log\left\{\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right\}\right] \\
&\leq \log\left\{\mathsf{E}_{\boldsymbol{\theta}^*}\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right]\right\} \\
&= \log\left\{\int_{\{x\in\mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)>0\}} \frac{\mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)\, \mathrm{d}x\right\} \\
&= \log\left\{\int_{\{x\in\mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)>0\}} \mathsf{p}_{\boldsymbol{\theta}}(x)\, \mathrm{d}x\right\} \\
&\leq \log\left\{\int_{-\infty}^{\infty} \mathsf{p}_{\boldsymbol{\theta}}(x)\, \mathrm{d}x\right\} \\
&= \log(1) = 0
\end{aligned}$$

となる. 上の不等式の等号成立条件は $\dfrac{\mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\, \mathsf{p}_{\boldsymbol{\theta}^*}(x)$ がほとんど至る所の $x \in \mathbb{R}$ において定数のときである. この場合, $\mathsf{P}_{\boldsymbol{\theta}} = \mathsf{P}_{\boldsymbol{\theta}^*}$ となるので

$$\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) > 0$$

がわかる. □

記号: $\boldsymbol{\theta}^* \in \Theta$ を固定する. $X \sim \mathsf{p}_{\boldsymbol{\theta}^*}$ に対して

$$W(\boldsymbol{\theta}) := \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right] \qquad (\boldsymbol{\theta} \in \Theta)$$

と定める.

定理 12.7. 以下を仮定する.
(a) $\Theta(\subset \mathbb{R}^d)$ はコンパクト部分集合.
(b) $\mathsf{E}_{\boldsymbol{\theta}^*}\left[\|W\|_\infty\right] < \infty$ である. ただし, $\|W\|_\infty = \sup_{\boldsymbol{\theta}\in\Theta}\left|W(\boldsymbol{\theta})\right|$ である.
(c) ほとんど至ところの $x \in \mathbb{R}$ で写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}_{\boldsymbol{\theta}}(x) \in [0, \infty)$$

は連続である.
(d) $\forall \boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow \mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$$

である.

このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ に対して

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

が成り立つ.

Proof. $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}$ とする. $j = 1, 2, \ldots, n$ に対して

$$W_j(\boldsymbol{\theta}) := \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X_j)}\right] \qquad (\boldsymbol{\theta} \in \Theta)$$

とおく. このとき, $\{W_j(\boldsymbol{\theta})\}_{j=1}^n$ は $C(\Theta)$ 値 i.i.d. 確率要素である. さらに

$$\mu(\boldsymbol{\theta}) := \mathsf{E}\big[W_j(\boldsymbol{\theta})\big] = -\mathsf{KL}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) \qquad (\boldsymbol{\theta} \in \Theta)$$

と書く. 仮定 (d) と補題 12.6 から

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \ \Rightarrow\ \mu(\boldsymbol{\theta}) < 0$$

であり, $\mu(\boldsymbol{\theta}^*) = 0$ となる. よって, μ は $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ で唯一の最大値を取る. また

$$\overline{W}_n(\boldsymbol{\theta}) := \frac{1}{n}\sum_{j=1}^n W_j(\boldsymbol{\theta}) = \frac{\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}^*)}{n}$$

なので, $\overline{W}_n$ は最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ で最大となる. よって, 定理 12.2 から

$$\big\|W_n - \mu\big\|_\infty \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

となる. よって, 定理 12.4(2) から

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

がわかる. □

定理 12.7 では母数空間をコンパクトとした. いくつかの仮定を加えることにより, コンパクトでない母数空間をもつ統計的モデルにおいても最尤推定量の一致性は保証されることがわかる.

定理 12.8. 統計的モデル $\mathcal{P} = \{\mathsf{p}_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$ を考える. ただし, $\Theta = \mathbb{R}^d$ もしくは有界でなくともよい. また, $\mathsf{p}_{\boldsymbol{\theta}}$ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. さらに, $\mathsf{p}_{\boldsymbol{\theta}}$ に対応する $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ と $(\Omega, \mathcal{A})$ それぞれの上の測度を $\mathsf{P}_{\boldsymbol{\theta}}$ と $\mathsf{Pr}_{\boldsymbol{\theta}}$ と書く. $\boldsymbol{\theta}^* \in \Theta$ とし

$$X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}(\,\cdot\,),$$

$$W_j(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X_j)}\right] \qquad (j = 1, 2, \ldots, n),$$
$$W(\boldsymbol{\theta}) = \log\left[\frac{\mathsf{p}_{\boldsymbol{\theta}}(X)}{\mathsf{p}_{\boldsymbol{\theta}^*}(X)}\right]$$

とする. 以下を仮定する.
(a) ほとんど至ところの $x \in \mathbb{R}$ で写像

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}_{\boldsymbol{\theta}}(x) \in [0, \infty)$$

は連続である.
(b) $\forall \boldsymbol{\theta} \in \Theta$ に対して

$$\boldsymbol{\theta} \neq \boldsymbol{\theta}^* \Rightarrow \mathsf{P}_{\boldsymbol{\theta}} \neq \mathsf{P}_{\boldsymbol{\theta}^*}$$

である.
(c) ほとんど至るところの $x \in \mathbb{R}$ に対して, $\left|\boldsymbol{\theta}\right|_{2,\,d} \to \infty$ のとき, $\mathsf{p}_{\boldsymbol{\theta}}(x) \to 0$ となる. ただし, $|\cdot|_{2,\,d}$ は $\mathbb{R}^d$ の Euclid ノルムである.
(d) Θ の任意のコンパクト部分集合 K に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\|\mathbb{1}_K W\|_\infty\right] < \infty$$

である. ただし, $g \in C(\Theta)$ に対して, $\|g\|_\infty = \sup_{\boldsymbol{\theta}\in\Theta}\left|g(\boldsymbol{\theta})\right|$ である.
(e) ある $M > 0$ が存在して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\left|\sup_{|\boldsymbol{\theta}|_{2,\,d} > M} W(\boldsymbol{\theta})\right|\right] < \infty$$

である.
このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ に対して

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

が成り立つ.

Proof. $|\boldsymbol{\theta}|_{2,\,d} \to \infty$ のとき, $\mathsf{p}_{\boldsymbol{\theta}}(x) \to 0$ なので, 固定した x と任意の $\epsilon > 0$ に対して, ある $M_\epsilon > 0$ が存在して

$$|\boldsymbol{\theta}|_{2,\,d} > M_\epsilon \Rightarrow \mathsf{p}_{\boldsymbol{\theta}}(x) < \epsilon$$

となる. したがって

$$\sup_{|\boldsymbol{\theta}|_{2,\,d} > M} \mathsf{p}_{\boldsymbol{\theta}}(x) \overset{M\to\infty}{\longrightarrow} 0 \tag{12.11}$$

となる. $\mathsf{p}_{\boldsymbol{\theta}^*}(x) > 0$ なる $x \in \mathbb{R}$ に対して

$$\sup_{|\boldsymbol{\theta}|_{2,\,d}>M} \log\left(\frac{\mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\right) = \log\left(\frac{\sup_{|\boldsymbol{\theta}|_{2,\,d}>M} \mathsf{p}_{\boldsymbol{\theta}}(x)}{\mathsf{p}_{\boldsymbol{\theta}^*}(x)}\right)$$
$$\overset{M\to\infty}{\longrightarrow} -\infty \qquad \big(\because (12.11)\big)$$

となる. よって

$$\mathsf{Pr}\left(\lim_{M\to\infty} \sup_{|\boldsymbol{\theta}|_{2,\,d}>M} W(\boldsymbol{\theta}) = -\infty\right) = 1 \tag{12.12}$$

となる. いま, 条件 (e) をみたす $M_1 > 0$ を固定し

$$Z_M := \sup_{|\boldsymbol{\theta}|_{2,\,d}>M} W(\boldsymbol{\theta}) \qquad (M \geq M_1)$$

とする. $M \to \infty$ のとき, $Z_M \to -\infty$ a.s. なので, ほとんと確実に $Z_{M_1} - Z_M \uparrow +\infty\,(M \to \infty)$ となる. よって, (12.12) に注意して単調収束定理を用いると

$$\lim_{M\to\infty} \mathsf{E}_{\boldsymbol{\theta}^*}\big[Z_{M_1} - Z_M\big] = +\infty$$

となる. さらに, $\mathsf{E}_{\boldsymbol{\theta}^*}[Z_{M_1}] < \infty$ なので

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M} W(\boldsymbol{\theta})\right] = \mathsf{E}_{\boldsymbol{\theta}^*}\big[Z_M\big] = \mathsf{E}_{\boldsymbol{\theta}^*}\big[Z_{M_1}\big] - \mathsf{E}_{\boldsymbol{\theta}^*}\big[Z_{M_1} - Z_M\big] \overset{M\to\infty}{\longrightarrow} -\infty$$

がわかる. ここで, 十分大きな $M_0 > 0$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W(\boldsymbol{\theta})\right] < 0 \tag{12.13}$$

とできる. $W(\boldsymbol{\theta}^*) = 0$ なので

$$|\boldsymbol{\theta}^*|_{2,\,d} < M_0$$

である. ここで,

$$Y_j := \sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W_j(\boldsymbol{\theta}) \qquad (j = 1, 2, \ldots)$$

とおく. 仮定 (e) を満たす $M_1 \leq M_0$ に対して

$$Y_j \leq \sup_{|\boldsymbol{\theta}|_{2,\,d}>M_1} W_j(\boldsymbol{\theta})$$

であるから

$$\mathsf{E}_{\boldsymbol{\theta}^*}[Y_j^+] < \infty, \qquad Y_j^+ := \max\{Y_j,\, 0\}$$

である. $L > 0$ に対して

$$Y_j^{(L)} := \max\{Y_j,\, -L\}$$

とおく. このとき

$$|Y_j^{(L)}| \leq Y_j^+ + L$$

であるから

$$\mathsf{E}_{\boldsymbol{\theta}^*}\left[|Y_j^{(L)}|\right] < \infty$$

となる. また

$$Y_j^{(L)} \downarrow Y_j \qquad (L \to \infty)$$

である. さらに,

$$\begin{aligned} Y_1^{(L)} &= \max\{Y_1,\, -L\} = \max\{Y_1^+ - Y_1^{-1},\, -L\} \\ &= \max\{Y_1^+,\, -L\} + \max\{-Y_1^-,\, -L\} = Y_1^+ - \min\{Y_1^{-1},\, L\} \end{aligned}$$

である. したがって, $L \to \infty$ のとき, 単調収束定理より

$$\begin{aligned} \mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1^{(L)}\right] &= \mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1^+\right] - \mathsf{E}_{\boldsymbol{\theta}^*}\left[\min\{Y_1^{-1},\, L\}\right] \downarrow \mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1^+\right] - \mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1^-\right] \\ &= \mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1\right] \end{aligned}$$

となる. ただし, $\mathsf{E}_{\boldsymbol{\theta}^*}\left[Y_1\right] = -\infty$ の場合も許していることに注意せよ. 以上から, 十分大きな $L_0 > 0$ に対して

$$\mathsf{E}_{\boldsymbol{\theta}^*}[Y_1^{(L_0)}] < 0$$

とできる. このとき, $Y_1^{(L_0)}, Y_2^{(L_0)}, \ldots$ は i.i.d. かつ可積分であるから, 大数の強法則より

$$\frac{1}{n}\sum_{j=1}^{n} Y_j^{(L_0)} \xrightarrow{\text{a.s.}} \mathsf{E}_{\boldsymbol{\theta}^*}[Y_1^{(L_0)}] < 0$$

となる. 一方

$$\sup_{|\boldsymbol{\theta}|_{2,\,d} > M_0} \overline{W}_n(\boldsymbol{\theta}) = \sup_{|\boldsymbol{\theta}|_{2,\,d} > M_0} \frac{1}{n}\sum_{j=1}^{n} W_j(\boldsymbol{\theta}) \leq \frac{1}{n}\sum_{j=1}^{n} \sup_{|\boldsymbol{\theta}|_{2,\,d} > M_0} W_j(\boldsymbol{\theta})$$

$$= \frac{1}{n}\sum_{j=1}^{n} Y_j \leq \frac{1}{n}\sum_{j=1}^{n} Y_j^{(L_0)}$$

となる. したがって,

$$\mathsf{Pr}_{\boldsymbol{\theta}^*}\left(\sup_{|\boldsymbol{\theta}|_{2,d}>M_0} \overline{W}_n(\boldsymbol{\theta}) \geq 0\right) \leq \mathsf{Pr}_{\boldsymbol{\theta}^*}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j^{(L_0)} \geq 0\right) \overset{n\to\infty}{\longrightarrow} 0$$

となるので

$$\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} \overline{W}_n(\boldsymbol{\theta}) \leq \frac{1}{n}\sum_{j=1}^{n} \sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W_j(\boldsymbol{\theta})$$
$$\overset{\mathrm{P}}{\to} \mathsf{E}_{\boldsymbol{\theta}^*}\left[\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} W(\boldsymbol{\theta})\right] < 0 \qquad (n\to\infty)$$

となるので

$$\mathsf{Pr}\left(\sup_{|\boldsymbol{\theta}|_{2,\,d}>M_0} \overline{W}_n(\boldsymbol{\theta}) \geq 0\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{12.14}$$

となることがわかる. いま, K を $\boldsymbol{\theta}^*$ を中心とする半径 M_0 の閉球, すなわち

$$K := \{\boldsymbol{\theta}\in\Theta;\ |\boldsymbol{\theta}-\boldsymbol{\theta}^*|_{2,d} \leq M_0\}$$

と[3]する. さらに,

$$A_n := \left\{\sup_{\boldsymbol{\theta}\in\Theta;\,|\boldsymbol{\theta}-\boldsymbol{\theta}^*|_{2,d}>M_0} \overline{W}_n(\boldsymbol{\theta}) < 0\right\}$$

とおく. (12.25) から

$$\mathsf{Pr}(A_n) \longrightarrow 1 \qquad (n\to\infty)$$

である. 一方

$$\overline{W}_n(\boldsymbol{\theta}^*) = 0$$

であり, $\widehat{\boldsymbol{\theta}}_n$ は $\overline{W}_n$ の最大化点なので

$$\overline{W}_n\left(\widehat{\boldsymbol{\theta}}_n\right) \geq \overline{W}_n(\boldsymbol{\theta}^*) = 0$$

である. したがって, 事象 A_n 上では

$$\widehat{\boldsymbol{\theta}}_n \in K$$

[3] $\Theta = \mathbb{R}^d$ のときは, 自動的に K はコンパクトとなる. しかし, $\Theta \neq \mathbb{R}^d$ で有界でないときには, K がコンパクトになるように M_0 を注意してとる必要がある.

となる. ここで, $\widetilde{\boldsymbol{\theta}}_n$ を K 上で $\overline{W}_n(\boldsymbol{\theta})$ を最大にする確率ベクトルとする. ただし, A_n 上では

$$\widetilde{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n$$

となるように選ぶ. 定理 12.7 から

$$\widetilde{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

となる. また,

$$\mathsf{Pr}\left(\widehat{\boldsymbol{\theta}}_n \neq \widetilde{\boldsymbol{\theta}}_n\right) \leq \mathsf{Pr}(A_n^c) \overset{n\to\infty}{\longrightarrow} 0$$

である. したがって, 任意の $\epsilon > 0$ に対して

$$\mathsf{Pr}\left(\left|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\right|_{2,\,d} > \epsilon\right) \leq \mathsf{Pr}\big(\widehat{\boldsymbol{\theta}}_n \neq \widetilde{\boldsymbol{\theta}}_n\big) + \mathsf{Pr}\big(\left|\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\right|_{2,d} > \epsilon\big) \overset{n\to\infty}{\longrightarrow} 0$$

がわかる. よって

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^* \qquad (n \to \infty)$$

が示された. □

12.3　最尤推定量の漸近正規性

定理 12.9. 統計的モデル $\mathcal{P} = \{\mathsf{p}_\theta;\, \theta \in \Theta\}$ を考える. ただし, $\Theta \subset \mathbb{R}$ で p_θ は $\mathbb{R}$ 上の p.d.f. または p.m.f. である. 以下を仮定する.
(a) $X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\theta^*}$ とする. ただし, $\theta^* \in \Theta \subset \mathbb{R}$ である.
(b) $A := \{x \in \mathbb{R};\, \mathsf{p}_\theta(x) > 0\}$ は　$\theta(\in \Theta)$ に依存しない.
(c) すべての $x \in A$ において, $\mathsf{p}_\theta(x)$ は Θ に関する 2 回連続微分可能な関数である.
(d) $X \sim \mathsf{p}_{\theta^*}$ に対して

$$W(\theta) = \log \mathsf{p}_\theta(X) \qquad (\theta \in \Theta)$$

と定める. 1 つの標本に基づく Fisher 情報量 $\mathcal{F}_X(\theta^*)$ は有界で

$$\mathcal{F}_X(\theta^*) = \mathsf{Var}\left[\dot{W}\,(\theta^*)\right], \qquad \mathcal{F}_X(\theta^*) = -\mathsf{E}_{\theta^*}\left[\ddot{W}\,(\theta^*)\right]$$

と表現できるとする. ただし, $\dot{W}\,(\theta) = \dfrac{\mathrm{d}W}{\mathrm{d}\theta}(\theta)$ と $\ddot{W}\,(\theta) = \dfrac{\mathrm{d}^2W}{\mathrm{d}\theta^2}(\theta)$ と定めた.

(e) Θ のすべての内点 θ^* に対して, ある $\epsilon > 0$ が存在して

$$\mathsf{E}_{\theta^*}\left[\left\| \mathbb{1}_{[\theta^*-\epsilon,\,\theta^*+\epsilon]}\, \ddot{W} \right\|_\infty\right] < \infty$$

とする. ただし, Θ 上の実数値関数 g に対して, $\|g\|_\infty = \sup_{\theta\in\Theta} |g(\theta)|$ と定めた.
(f) θ^* の最尤推定量 $\widehat{\theta}_n$ は一意性をもつ.
このとき, Θ の任意の内点 θ^* に対して

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) \rightsquigarrow \mathsf{N}\left(0,\, \frac{1}{\mathcal{F}_X(\theta^*)}\right) \qquad (n\to\infty)$$

が成り立つ.

この定理を証明するために次の補題を利用する.

補題 12.10. Y, $\{Y_n\}_{n=1}^\infty$ を確率変数列とし, $Y_n \rightsquigarrow Y\,(n\to\infty)$ とする. $\{B_n\}_{n=1}^\infty$ を事象の列とし, $\mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 1$ とする. このとき, 任意の確率変数列 $\{Z_n\}_{n=1}^\infty$ に対して

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}_{B_n^c} \rightsquigarrow Y \qquad (n\to\infty)$$

が成り立つ. ただし, $\mathbb{1}_{B_n} = \begin{cases} 1 & (B_n \text{ が起こったとき}) \\ 0 & (\text{その他の場合}) \end{cases}$ である.

Proof. $\forall\epsilon > 0$ に対して

$$\mathsf{Pr}\big(|Z_n \mathbb{1}_{B_n^c}| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となる. したがって

$$Z_n \mathbb{1}_{B_n^c} \overset{\mathrm{P}}{\to} 0 \qquad (n\to\infty) \tag{12.15}$$

がわかる. さらに

$$\mathsf{Pr}\big(|\mathbb{1}_{B_n} - 1| > \epsilon\big) \le \mathsf{Pr}(B_n^c) = 1 - \mathsf{Pr}(B_n) \overset{n\to\infty}{\longrightarrow} 0$$

となるので

$$\mathbb{1}_{B_n} \overset{\mathrm{P}}{\to} 1 \qquad (n\to\infty) \tag{12.16}$$

がわかる. (12.15) と (12.16) に注意して Slusky の定理 (定理 4.14(3)(4)) を用いると

$$Y_n \mathbb{1}_{B_n} + Z_n \mathbb{1}_{B_n^c} \rightsquigarrow Y \qquad (n\to\infty)$$

がわかる. □

定理 12.9 の証明: 仮定 (e) から $\epsilon > 0$ をうまく取って

$$[\theta^* - \epsilon,\, \theta^* + \epsilon] \subset \Theta \quad \text{かつ} \quad \mathsf{E}_{\theta^*}\left[\left\| \mathbb{1}_{[\theta^*-\epsilon,\,\theta^*+\epsilon]}\, \overset{\bullet\bullet}{W} \right\|_\infty\right] < \infty$$

となるようにできる. 事象 B_n を

$$B_n := \left\{\widehat{\theta}_n \in [\theta^* - \epsilon,\, \theta^* + \epsilon]\right\}$$

と定める. $\widehat{\theta}_n$ の一致性から

$$\mathsf{Pr}_{\theta^*}(B_n) \overset{n\to\infty}{\longrightarrow} 1 \tag{12.17}$$

となる. さらに, B_n 上では, $\widehat{\theta}_n$ は $n\overline{W}_n(\theta) = \ell_n(\theta)\,(\theta \in \Theta)$ を最大にするので

$$\overset{\bullet}{\overline{W}}_n\,(\widehat{\theta}_n) = \frac{\mathrm{d}\overline{W}}{\mathrm{d}\theta}(\theta)\bigg|_{\theta=\widehat{\theta}_n} = 0$$

となる. ここで, θ^* のまわりで $\overset{\bullet}{\overline{W}}_n\,(\theta^*)$ を Taylor 展開すると

$$\overset{\bullet}{\overline{W}}_n\,(\theta^*) = \overset{\bullet}{\overline{W}}_n\,(\widehat{\theta}_n) + \overset{\bullet\bullet}{\overline{W}}_n\,(\widetilde{\theta}_n)(\theta^* - \widehat{\theta}_n) \tag{12.18}$$

を得る. ただし, $\widetilde{\theta}_n$ は θ^* と $\widehat{\theta}_n$ を結ぶ直線上の確率変数である. $\overset{\bullet}{\overline{W}}_n\,(\widehat{\Theta}) = 0$ なので

$$\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) = \frac{\sqrt{n}\,\overset{\bullet}{\overline{W}}_n\,(\theta^*)}{-\overset{\bullet\bullet}{\overline{W}}_n\,(\widetilde{\theta}_n)}$$

となる. 仮定 (d) から $\overset{\bullet}{\overline{W}}_n\,(\theta^*)$ は平均 0, 分散 $\mathcal{F}_X(\theta^*)$ の正規分布に従う i.i.d. 確率変数の標本平均であることに注意して, 中心極限定理 (定理 4.22) を用いると

$$\sqrt{n}\,\overset{\bullet}{\overline{W}}_n\,(\theta^*) \rightsquigarrow Z \sim \mathsf{N}\big(0,\, \mathcal{F}_X(\theta^*)\big) \qquad (n \to \infty) \tag{12.19}$$

となる. B_n 上では

$$|\widetilde{\theta}_n - \theta^*| \le |\widehat{\theta}_n - \theta^*|$$

であることと $\widehat{\theta}_n$ の一致性から

$$\mathsf{Pr}\big(|\widetilde{\theta}_n - \theta^*| < \epsilon\big) \ge \mathsf{Pr}\big(|\widehat{\theta}_n - \theta^*| < \epsilon\big) \overset{n\to\infty}{\longrightarrow} 1$$

となることがわかる. よって

$$\widetilde{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n \to \infty) \tag{12.20}$$

となる. 定理 12.2 から

$$\left\| \mathbb{1}_{[\theta^*-\epsilon,\, \theta^*+\epsilon]} \big(\overset{\bullet\bullet}{\overline{W}}_n - \mu \big) \right\|_\infty \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{12.21}$$

となる. ただし, $\mu(\theta) = \mathsf{E}_{\theta^*}\big[\overset{\bullet\bullet}{\overline{W}}_n(\theta)\big]$ $(\theta \in \Theta)$ である. さらに, (12.20) と (12.21) に注意して, 定理 12.4(1) を用いると

$$\overset{\bullet\bullet}{\overline{W}}_n(\widetilde{\theta}_n) \xrightarrow{\mathrm{P}} \mu(\theta^*) = -\mathcal{F}_X(\theta^*) \tag{12.22}$$

となる. B_n^c 上では $\widehat{\theta}_n$ の振る舞いは関係ない. 最後に, (12.17), (12.19) および (12.22 に注意して補題 12.10 を用いると

$$\begin{aligned}
\sqrt{n}\big(\widehat{\theta}_n - \theta^*\big) &= \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n} + \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n^c} \\
&= \frac{\sqrt{n}\,\overset{\bullet}{\overline{W}}_n(\theta^*)}{-\overset{\bullet\bullet}{\overline{W}}_n(\widetilde{\theta}_n)}\mathbb{1}_{B_n} + \sqrt{n}\big(\widehat{\theta}_n - \theta^*\big)\mathbb{1}_{B_n^c} \\
&\rightsquigarrow \frac{Z}{\mathcal{F}_X(\theta^*)} \sim \mathsf{N}\left(0,\, \frac{1}{\mathcal{F}_X(\theta^*)}\right) \qquad (n \to \infty)
\end{aligned}$$

がわかる. □

定理 12.11 (Camér の定理)**.** $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*}$ とする. ただし, $\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d$ とする. さらに, 以下の条件 (a) – (f) を仮定する.
(a) Θ は $\mathbb{R}^d$ の開部分集合である.
(b) $A := \{x \in \mathbb{R};\, \mathsf{p}_{\boldsymbol{\theta}}(x) > 0\}$ は $\boldsymbol{\theta}(\in \Theta)$ に依存しない.
(c) ほとんど至るところの $x \in A$ について, 関数 $\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{p}_{\boldsymbol{\theta}}(x)$ は 2 回連続微分可能である. さらに, その 2 回の微分演算と積分記号の交換ができると仮定する.
(d) $\mathbb{R}$ 上の関数 K が存在して, $\mathsf{E}_{\boldsymbol{\theta}^*}\big[|K(X)|\big] < \infty$ であり

$$\overset{\bullet\bullet}{\boldsymbol{W}}(\boldsymbol{\theta}) := \left(\frac{\partial^2}{\partial\theta_j \partial\theta_k} \log \mathsf{p}_{\boldsymbol{\theta}}(X) \right)_{j,\,k=1,\,2,\ldots,\,d}; \qquad \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$$

のどの成分の絶対値も $\boldsymbol{\theta}^*$ の近傍において, 一様に K で上からおさえられる.
(e) $\boldsymbol{\mathcal{F}}_X(\boldsymbol{\theta}^*) = -\mathsf{E}_{\boldsymbol{\theta}^*}\big[\overset{\bullet\bullet}{\boldsymbol{W}}(\boldsymbol{\theta}^*)\big]$ は正定値である.
(f) $\mathsf{p}_{\boldsymbol{\theta}} = \mathsf{p}_{\boldsymbol{\theta}^*} \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^*$.
このとき, $\boldsymbol{\theta}^*$ の最尤推定量 $\widehat{\boldsymbol{\theta}}_n$ は

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\big) \rightsquigarrow \mathsf{N}\Big(\mathsf{0}_d,\, \big\{\boldsymbol{\mathcal{F}}_X(\boldsymbol{\theta}^*)\big\}^{-1}\Big) \qquad (n \to \infty)$$

Proof. 証明は Fergusson (2017, pp.135-136) を参照のこと.
【コメント】 定理 12.9 の条件を確認すればよい. □

12.4 尤度比検定統計量の漸近分布

正則母数モデル $\mathcal{P} = \{\mathsf{p}_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d\}$ を考える. ただし, $\mathsf{p}_{\boldsymbol{\theta}}$ は $\mathbb{R}^d$ 上の p.d.f. または p.m.f. である. いま

$$X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{p}_{\boldsymbol{\theta}^*} \qquad (\boldsymbol{\theta}^* \in \Theta)$$

を仮定する. 標本 $\boldsymbol{X} := (X_1, X_2, \ldots, X_n)^\top$ に基づき, 仮説

$$H_0 : \boldsymbol{\theta}^* \in \Theta_0 \quad \text{v.s.} \quad H_1 : \boldsymbol{\theta}^* \in \Theta \setminus \Theta_0$$

を検定する. ここで Θ_0 は Θ の空でない真部分集合である. 汎用的検定統計量である尤度比検定統計量

$$\lambda_n(\boldsymbol{X}) := \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \prod_{j=1}^n \mathsf{p}_{\boldsymbol{\theta}}(X_j)}{\sup_{\boldsymbol{\theta} \in \Theta} \prod_{j=1}^n \mathsf{p}_{\boldsymbol{\theta}}(X_j)} = \frac{\mathsf{lik}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)}{\mathsf{lik}_n\big(\widehat{\boldsymbol{\theta}}_n\big)} \tag{12.23}$$

を考える. ここで, $\widehat{\boldsymbol{\theta}}_n$ と $\widehat{\boldsymbol{\theta}}_{n,0}$ はそれぞれ Θ と Θ_0 における最尤推定量である. $\lambda_n(\boldsymbol{X})$ の実現値が小さい値のとき, H_0 は棄却されるという検定方式である.

帰無仮説 H_0 のもとでの尤度比検定統計量の漸近分布を求めるために, Θ_0 を含む最小の affine 平面の次元 $\dim \Theta_0$ が $d - r\,(1 \leq r < d)$ であると仮定する. さらに, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ と書いたとき, 帰無仮説 H_0 が

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_r = 0$$

と書けたと仮定する.

定理 12.12 (Wilks の定理)**.** 定理 12.11 の仮定のもとで真の母数 $\boldsymbol{\theta}^*$ が帰無仮説 H_0 を満足しているとする. このとき

$$-2 \log \lambda_n(\boldsymbol{X}) \rightsquigarrow \chi^2_r \qquad (n \to \infty)$$

が成立する. ただし, χ^2_r は自由度 r の χ^2 分布を表す.

Proof. (12.23) から

$$-2 \log \lambda_n(\boldsymbol{X}) = 2\big\{\ell_n\big(\widehat{\boldsymbol{\theta}}_n\big) - \ell_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)\big\}$$

と書き直せることに注意する. ただし, $\ell_n(\boldsymbol{\theta}) = \log \mathsf{lik}_n(\boldsymbol{\theta})$ である. $\ell_n\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr)$ を $\widehat{\boldsymbol{\theta}}_n$ のまわりで 2 次の項まで Taylor 展開すると

$$\begin{aligned}\ell_n\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr) = \ell_n\bigl(\widehat{\boldsymbol{\theta}}_n\bigr) &+ \overset{\bullet}{\ell}_n\,\bigl(\widehat{\boldsymbol{\theta}}_n\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) \\ &- n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^{\top}\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\bigl(\widetilde{\boldsymbol{\theta}}_{n,0}\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)\end{aligned}$$

と書ける. ただし

$$\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\bigl(\widetilde{\boldsymbol{\theta}}_{n,0}\bigr) = -\frac{1}{n}\int_0^1\int_0^1 v\,\overset{\bullet\bullet}{\ell}_n\,\bigl\{\widehat{\boldsymbol{\theta}}_n + uv\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)\bigr\}\,\mathrm{d}u\,\mathrm{d}v$$

とした. さらに

$$\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\bigl(\widetilde{\boldsymbol{\theta}}_{n,0}\bigr) \overset{\text{a.s.}}{\to} \frac{1}{2}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*) \qquad (n \to \infty) \tag{12.24}$$

となることがわかる. ここで $\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\,\cdot\,)$ は定理 12.11(e) で与えられた Fisher 情報量である. $\overset{\bullet}{\ell}_n\,\bigl(\widehat{\boldsymbol{\theta}}_n\bigr) = 0$ が成り立つので, 十分大きな n に対して

$$\begin{aligned}-2\log\lambda_n(\boldsymbol{X}) &= 2n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^{\top}\widetilde{\boldsymbol{\mathcal{F}}}_{\boldsymbol{X}}\bigl(\widetilde{\boldsymbol{\theta}}_{n,0}\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) \\ &= n\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr)^{\top}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\bigl(\boldsymbol{\theta}^*\bigr)\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) + o_P(1)\end{aligned} \tag{12.25}$$

が成り立つ. ただし, 確率変数列 A_n, $B_n\,(n \in \mathbb{N})$ に対して, $A_n = B_n + o_P(1)$ は $A_n - B_n \overset{\mathrm{P}}{\to} 0\,(n \to \infty)$ を意味する.

帰無仮説 H_0 が単純のとき, $\widehat{\boldsymbol{\theta}}_{n,0} = \boldsymbol{\theta}^*$ なので

$$\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\bigr) \rightsquigarrow \mathsf{N}_d\Bigl(\mathbf{0}_d,\,\bigl\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)\bigr\}^{-1}\Bigr) \qquad (n \to \infty)$$

となるので, 証明は終わりである.

帰無仮説 H_0 が複合のとき, $\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_{n,0}\bigr)$ の漸近分布を求めるために, $\overset{\bullet}{\ell}_n\,\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr)$ を $\widehat{\boldsymbol{\theta}}_n$ のまわりで Taylor 展開する.

$$\begin{aligned}&\frac{1}{\sqrt{n}}\,\overset{\bullet}{\ell}_n\,\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr) \\ &\quad= \frac{1}{\sqrt{n}}\,\overset{\bullet}{\ell}_n\,\bigl(\widehat{\boldsymbol{\theta}}_{n,0}\bigr) + \frac{1}{n}\Bigl(\int_0^1 \overset{\bullet\bullet}{\ell}_n\,\bigl\{\widehat{\boldsymbol{\theta}}_n + v\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr\}\,\mathrm{d}v\Bigr)\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) \\ &\quad= -\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\bigl(\boldsymbol{\theta}^*\bigr)\sqrt{n}\bigl(\widehat{\boldsymbol{\theta}}_{n,0} - \widehat{\boldsymbol{\theta}}_n\bigr) + o_P(1)\end{aligned}$$

となる. ただし, $\boldsymbol{\theta} = (\Theta_1, \Theta_2, \ldots, \Theta_d)^{\top}$ と書いたとき

$$\begin{aligned}\overset{\bullet}{\ell}_n\,(\boldsymbol{\theta}) &= \Bigl(\frac{\partial\ell_n}{\partial\Theta_1}(\boldsymbol{\theta}),\,\frac{\partial\ell_n}{\partial\Theta_2}(\boldsymbol{\theta}),\,\ldots,\,\frac{\partial\ell_n}{\partial\Theta_d}(\boldsymbol{\theta})\Bigr)^{\top}, \\ \overset{\bullet\bullet}{\ell}_n\,(\boldsymbol{\theta}) &= \Bigl(\frac{\partial^2\ell_n}{\partial\Theta_j\partial\Theta_k}(\boldsymbol{\theta})\Bigr)_{j,\,k=1,\,2,\ldots,\,d}\end{aligned}$$

と定めた. したがって

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\widehat{\boldsymbol{\theta}}_n\big)=\big\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)\big\}^{-1}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)+o_P(1)$$

を得る. このことから

$$-2\log\lambda_n=\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)^\top\big\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)\big\}^{-1}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)+o_P(1)$$

と表現できる. $\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)$ の漸近分布を求めるために $\boldsymbol{\theta}^*$ の周りでこれを Taylor 展開する.

$$\begin{aligned}&\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)\\&\quad=\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\big(\boldsymbol{\theta}^*\big)+\frac{1}{n}\bigg(\int_0^1\overset{\bullet\bullet}{\ell}_n\big\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\big\}\,\mathrm{d}v\bigg)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\end{aligned}\tag{12.26}$$

を得る. ここで, $d\times d$ 行列 $\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)$ を 4 つのブロック行列に分解する.

$$\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}(\boldsymbol{\theta}^*)=\begin{pmatrix}\mathbf{G}_1&\mathbf{G}_2\\\mathbf{G}_2^\top&\mathbf{G}_3\end{pmatrix}.$$

ただし, $\mathbf{G}_1$ は $r\times r$ 行列である. さらに

$$\mathbf{H}=\begin{pmatrix}\mathbf{0}_{r\times r}&\mathbf{0}_{r\times(d-r)}\\\mathbf{0}_{r\times r}^\top&\mathbf{G}_3^{-1}\end{pmatrix}$$

とする. $\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)$ の後半の $(d-r)$ 個の要素は 0 なので

$$\mathbf{H}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)=\mathbf{0}_d\tag{12.27}$$

となる. よって, (12.26) と (12.27) から

$$\begin{aligned}\mathbf{H}\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\big(\boldsymbol{\theta}^*\big)&=\mathbf{H}\bigg\{\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\big(\widehat{\boldsymbol{\theta}}_{n,0}\big)-\frac{1}{n}\bigg(\int_0^1\overset{\bullet\bullet}{\ell}_n\big\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\big\}\,\mathrm{d}v\bigg)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\bigg\}\\&=-\mathbf{H}\frac{1}{n}\bigg(\int_0^1\overset{\bullet\bullet}{\ell}_n\big\{\boldsymbol{\theta}^*+v\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\big\}\,\mathrm{d}v\bigg)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)\\&=-\mathbf{H}\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\big(\boldsymbol{\theta}^*\big)\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=-\begin{pmatrix}\mathbf{0}_{r\times r}&\mathbf{0}_{r\times(d-r)}\\\mathbf{0}_{r\times r}^\top&\mathbf{G}_3^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{G}_1&\mathbf{G}_2\\\mathbf{G}_2^\top&\mathbf{G}_3\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=-\begin{pmatrix}\mathbf{0}_{r\times r}&\mathbf{0}_{r\times(d-r)}\\\mathbf{G}_3^{-1}\mathbf{G}_2^\top&\mathbf{I}_{d-r}\end{pmatrix}\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\\&=\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_{n,0}-\boldsymbol{\theta}^*\big)+o_P(1)\end{aligned}\tag{12.28}$$

となる. 最後の等号は $\widehat{\boldsymbol{\theta}}_{n,0}$ と $\boldsymbol{\theta}^*$ の最初の r 個の成分は 0 であることからわかる. (12.28) を (12.27) に代入すると

$$\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right) = \left\{\mathbf{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right\}\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right) + o_P(1)$$

を得る. 中心極限定理 (定理 4.24) から

$$\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\boldsymbol{\theta}^*\right) \rightsquigarrow \mathsf{N}_d\Big(\mathbf{0}_d,\, \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\Big) \qquad (n\to\infty)$$

が成り立つので

$$\frac{1}{\sqrt{n}}\overset{\bullet}{\ell}_n\left(\widehat{\boldsymbol{\theta}}_{n,0}\right) \rightsquigarrow \left(\boldsymbol{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right)\boldsymbol{Y} \qquad (n\to\infty),$$

$$\boldsymbol{Y} \sim \mathsf{N}_d\Big(\mathbf{0}_d,\, \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\Big)$$

である. これより

$$\begin{aligned}
-2\log\lambda_n &= \boldsymbol{Y}^\top\left(\boldsymbol{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right)^\top\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{-1}\left(\boldsymbol{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right)\boldsymbol{Y} + o_P(1)\\
&= \boldsymbol{Y}^\top\left[\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{-1} - \mathbf{H}\right]\boldsymbol{Y} + o_P(1)\\
&= \boldsymbol{Z}^\top\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{1/2}\left[\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{-1} - \mathbf{H}\right]\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{1/2}\boldsymbol{Z} + o_P(1)
\end{aligned}$$

となる. ただし, $\boldsymbol{Z} \sim \mathsf{N}_d(\mathbf{0}_d, \mathbf{I}_d)$ である. ここで, 行列

$$\mathbf{P} = \left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{1/2}\left[\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{-1} - \mathbf{H}\right]\left\{\boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\right\}^{1/2}$$

は

$$\mathbf{P}^2 = \mathbf{P};\quad \mathbf{P}^\top = \mathbf{P};\quad \mathrm{rank}(\mathbf{P}) = \mathrm{tr}\,(\mathbf{P}) = \mathrm{tr}\left(\mathbf{I}_d - \boldsymbol{\mathcal{F}}_{\boldsymbol{X}}\left(\boldsymbol{\theta}^*\right)\mathbf{H}\right) = r$$

をみたす. $\mathbf{P}$ はランク r の正射影となる. しがたって

$$-2\log\lambda_n(\boldsymbol{X}) \rightsquigarrow \boldsymbol{Z}^\top\mathbf{P}\boldsymbol{Z} \sim \chi^2_r \qquad (n\to\infty)$$

を得る. □

(12.24) の証明 □

12.5 章末注釈と参考文献

12.6 演習問題

演習問題 12.1.

演習問題 12.2.

第III部

宴の始末編：付録

第 A 章　補遺: 集合・写像・位相の復習

この章では, 集合・写像・位相の基本的な事項をまとめている. 詳しくは [35, 10, 20, 39] を参照のこと. 節 A.1 では, 集合の言葉使いと数の集合の基礎事項をまとめている. 節 A.2 では, 写像の性質の基礎事項をまとめている. 節 A.3 では, 可算集合と非可算集合の基礎事項をまとめている. 節 A.4 では, 距離空間と位相空間の定義と基礎事項をまとめている. 節 A.5 では, 実数の拡張である $\mathbb{R} \cup \{-\infty, \infty\}$ における演算の規則についてまとめた. これらのことをよく理解している場合には, この章を読み飛ばしてもよい.

A.1　集合論の言葉使い

A.1.1　数の集合

数の集合に対して, 以下の記号を用いることにする.

$\mathbb{N}$: 正の整数の集合 (0 を除く),
$\mathbb{Z}$: 整数の集合,
$\mathbb{Q}$: 有理数の集合,
$\mathbb{R}$: 実数の集合,
$\mathbb{C}$: 複素数の集合

とする.

A.1.2　集合の記号

空集合を $\varnothing$ と記し, 集合 X のすべての部分集合の族を 2^X と記す. すなわち

$$2^X := \{E;\, E \subset X\}$$

である.

$\mathcal{E}$ を集合 X の部分集合族としたとき

$$\bigcup_{E\in\mathcal{E}} E := \{x \in X;\ \text{ある } E \in \mathcal{E} \text{ が存在して}, x \in E\},$$
$$\bigcap_{E\in\mathcal{E}} E := \{x \in X;\ \text{すべての } E \in \mathcal{E} \text{ に対して}, x \in E\}$$

と定める. 通常, 添え字集合を用いて, $\mathcal{E} = \{E_\alpha;\ \alpha \in A\} = \{E_\alpha\}_{\alpha\in A}$ と書いたとき

$$\bigcup_{\alpha\in A} E_\alpha; \qquad \bigcap_{\alpha\in A} E_\alpha$$

と表す . $\alpha \neq \beta\,(\alpha, \beta \in A)$ に対して, $E_\alpha \cap E_\beta = \varnothing$ のとき, 集合族 $\{E_\alpha\}_{\alpha\in A}$ は 互いに排反 という. 部分集合族が $\mathbb{N}$ で添え字付けられるとする. すなわち, $\{E_n\}_{n\in\mathbb{N}}$ を考える. このとき, この集合族の 上極限 と 下極限 をそれぞれ

$$\limsup E_n := \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} E_k; \qquad \liminf E_n := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} E_k$$

と定める. すると

$$\limsup E_n = \{x \in X;\ \text{無限個の } n \text{ に対して}, x \in E_n\}$$
$$\liminf E_n = \{x \in X;\ \text{有限個の } n \text{ を除いたすべてに対して}, x \in E_n\}$$

と書き直せることがわかる.

部分集合 $E, F \subset X$ のとき, $E \setminus F$ によって, それらの差を表す. すなわち

$$E \setminus F = \{x \in X;\ x \in E \text{ かつ } x \notin F\}$$

である. 集合 $E \subset X$ の 補集合 を E^c で表す. すなわち

$$E^c = X \setminus E$$

である. このとき, **De Morgan の法則**

$$\left(\bigcup_{\alpha\in A} E_\alpha\right)^c = \bigcap_{\alpha\in A} E_\alpha^c; \qquad \left(\bigcap_{\alpha\in A} E_\alpha\right)^c = \bigcup_{\alpha\in A} E_\alpha^c \tag{A.1}$$

が成立する.

集合 X と Y の 直積 を $X \times Y$ と書く. すなわち

$$X \times Y = \{(x, y);\ x \in X, y \in Y\}$$

である. 有限個の集合 $X_1, X_2, \ldots, X_n$ の直積

$$X_1 \times X_2 \times \cdots \times X_n \left(\prod_{j=1}^{n} X_j \text{ とも表す}\right)$$

も同様に定義する. 特に, $X_1 = X_2 = \cdots = X_n = X$ のとき, $X_1 \times X_2 \times \cdots \times X_n$ を X^n とも書く.

A.2 写像

A.2.1 写像の定義と性質

定義 A.1. (1) X, Y を空でない集合とする. このとき

- X の任意の元に対して
- その元に対応する Y のある元がただ 1 つ与えられる

とする. このことを

$$f : X \to Y$$

と表し, f を X から Y への **写像** という.
(2) X を写像 f の **定義域**, Y を写像 f の **終域** という.

注意 A.2. (1) X, Y を空でない集合とし, $y_0 \in Y$ を 1 つ選んで固定しておく. このとき, 写像 $f : X \to Y$ を

$$f(x) = y_0 \qquad (x \in X)$$

により定める. この f を **定値写像** という.
(2) X を空でない集合とし, $A \subset X$ とする. このとき, X から $\{0, 1\}$ への写像 $\mathbb{1}_A : X \to \{0, 1\}$ を

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \in X \setminus A) \end{cases}$$

により定める. この $\mathbb{1}_A$ を **定義関数** または **特性関数** という.
(3) X, Y を空でない集合とし, $X \subset Y$ とする. このとき, 写像 $\imath : X \to Y$ を

$$\imath(x) = x \qquad (x \in X)$$

により定める. $\imath$ を **包含写像** という. 特に, $X = Y$ のとき, $\imath$ を id_X と表し, **恒等写像** という.
(4) X, Y を空でない集合, $f : X \to Y$ を写像とし, $A \subset X$, $A \neq \varnothing$ とする. このとき, 写像 $f\big|_A : A \to Y$ を

$$f\big|_A(x) = f(x) \qquad (x \in A)$$

により定める. この $f\big|_A$ を f の A への **制限** または **制限写像** という.

定義 A.3. f, g を写像とする. f と g の定義域が等しく, f と g の終域も等しく, さらに f, g の定義域の任意の元 x に対して, $f(x) = g(x)$ がなりたつとき, $f = g$ と表し, 写像 f と g は **等しい** という. また, 写像 f と g が等しくないとき, $f \neq g$ と表す.

定義 A.4. X, Y を空でない集合, $f : X \to Y$ を写像とする. このとき, $G(f) \subset X \times Y$ を

$$G(f) := \big\{\big(x, f(x)\big);\, x \in X\big\}$$

により定め, これを写像 f の **グラフ** という.

A.2.2　像と逆像

定義 A.5. X, Y を空でない集合とし, $f : X \to Y$ を写像とする.
(1) X の部分集合 A に対して, $\big\{f(x);\, x \in A\big\}$ を写像 f による部分集合 A の **像** といい, $f(A)$ と書く. $f(X)$ を写像 f の **値域** という.
(2) Y の部分集合 B に対して, 写像 f による B の **逆像** $f^{-1}(B)$ を

$$f^{-1}(B) = \{x \in X;\, f(x) \in B\}$$

で定める. 略記して, $\{f \in B\}$ とも書く.
(3) $f(X) = Y$ のとき, f は **全射** という.
(4) $x, x' \in X$, $x \neq x'$ ならば, $f(x) \neq f(x')$ であるとき, f は **単射** という.
(5) f が全射かつ単射のとき, **全単射** という.

命題 A.6. X, Y を空でない集合とし, $f : X \to Y$ を写像とし, $A, A_1, A_2 \subset X$, $B, B_1, B_2 \subset Y$ とする. このとき, 次の (1) ∼ (10) がなりたつ.
(1) $A_1 \subset A_2$ ならば, $f(A_1) \subset f(A_2)$.
(2) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.
(3) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$.
(4) $f(A_1 \setminus A_2) \supset f(A_1) \setminus f(A_2)$.
(5) $B_1 \subset B_2$ ならば, $f^{-1}(B_1) \subset f^{-1}(B_2)$.
(6) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.
(7) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.
(8) $f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$.
(9) $f^{-1}\big(f(A)\big) \supset A$.
(10) $f\big(f^{-1}(B)\big) \subset B$.

Proof. 証明は [35, pp.32-33] を参照のこと. □

命題 A.6(3)(4)(9)(10) において, 等号が成立しない例をあげる.

注意 A.7. 写像 $f : \mathbb{R} \to \mathbb{R}$ を

$$f(x) = x^2 \qquad (x \in \mathbb{R})$$

により定める.
(1) $A_1 = (-1, 0]$, $A_2 = [0, 1)$ とすると

$$f(A_1 \cap A_2) = \{0\}; \qquad f(A_1) \cap f(A_2) = [0, 1);$$
$$f(A_1 \setminus A_2) = (0, 1); \qquad f(A_1) \setminus f(A_2) = \varnothing$$

となる.
(2) $A = [0, 1)$, $B = (-1, 1)$ とすると

$$f^{-1}\big(f(A)\big) = (-1, 1); \qquad f\big(f^{-1}(B)\big) = [0, 1)$$

となる. □

命題 A.8. X, Y を空でない集合とし, $f : X \to Y$ を写像とし, A, A_1, $A_2 \subset X$, $B \subset Y$ とする. このとき, 次の (1) ∼ (4) がなりたつ.
(1) f が単射ならば, $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$.
(2) f が単射ならば, $f(A_1 \setminus A_2) = f(A_1) \setminus f(A_2)$.
(3) f が単射ならば, $f^{-1}\big(f(A)\big) = A$.
(4) f が全射ならば, $f\big(f^{-1}(B)\big) = B$.

Proof. 証明は [35, p.37] を参照のこと. □

A.2.3 合成写像

定義 A.9. X, Y, Z を空でない集合とし, $f : X \to Y$, $g : Y \to Z$ を写像とする. このとき, X から Z への写像 $g \circ f$ を

$$(g \circ f)(x) = g\big(f(x)\big) \qquad (x \in X)$$

により定める. $g \circ f$ を f と g の **合成** または **合成写像** という.

命題 A.10. X, Y, Z, W を空でない集合, $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ を写像とする. このとき

$$h \circ (g \circ f) = (h \circ g) \circ f$$

である. 特に, $h \circ (g \circ f)$ と $(h \circ g) \circ f$ をともに $h \circ g \circ f$ と表す.

Proof. 証明は [35, p.40] を参照のこと. □

命題 A.11. X, Y, Z を空でない集合, $f : X \to Y$, $g : Y \to Z$ を写像とする. このとき, (1) $\sim$ (3) がなりたつ.
(1) f, g が全射ならば, $g \circ f$ も全射.
(2) f, g が単射ならば, $g \circ f$ も単射.
(3) f, g が全単射ならば, $g \circ f$ も全単射.

Proof. 証明は [35, p.40] を参照のこと. □

A.3　可算集合・非可算集合・濃度

定義 A.12. (1) X, Y を空でない集合とする. X から Y への全単射が存在するとき, X と Y は **濃度が等しい** という.
(2) 有限の元から構成される集合を **有限** 集合という. 無限の元から構成される集合を **無限** 集合という.
(3) 自然数全体の集合 $\mathbb{N}$ と濃度が等しい集合を **可算** 集合という.
(4) 集合 X が有限集合または可算集合のとき, X は **高々可算集合** という.
(5) 可算集合でない無限集合を **非可算** 集合という.

命題 A.13. (1) 整数全体の集合 $\mathbb{Z}$ と有理数全体の集合 $\mathbb{Q}$ は可算集合である.
(2) 自然数 d に対して, $\mathbb{N}^d$, $\mathbb{Z}^d$, $\mathbb{Q}^d$ は可算集合である.
(3) 実数全体の集合 $\mathbb{R}$ と $\mathbb{N}$ の冪集合 $2^{\mathbb{N}}$ は濃度が等しい.
(4) X を空でない集合とする. このとき, X から X の冪集合 2^X への全射は存在しない.
(5) d を自然数とする. このとき, $\mathbb{R}$ と $\mathbb{R}^d$ は濃度が等しい.

Proof. 証明は [35, pp.68-80] を参照のこと. □

A.4　距離と位相

A.4.1　距離空間

定義 A.14. $\mathbb{X}$ を空でない集合とする. 写像 $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ が次の 3 条件をみたすとき, d を $\mathbb{X}$ 上の **距離関数** といい, 組 $(\mathbb{X}, d)$ を **距離空間** という.
(1) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) \geq 0$ である. 特に, $d(x, y) = 0 \Leftrightarrow x = y$ である.
(2) 任意の $x, y \in \mathbb{X}$ に対して, $d(x, y) = d(y, x)$.
(3) 任意の $x, y, z \in \mathbb{X}$ に対して, $d(x, z) \leq d(x, y) + d(y, z)$.

注意 A.15. d を自然数とし, $\mathbb{X} = \mathbb{R}^d$ とする. 距離関数の例として, Euclid ノルム

$$|\boldsymbol{x}|_{2,\,d} = \sqrt{\sum_{j=1}^{d} x_j^2} \qquad (\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_d)^\top \in \mathbb{R}^d)$$

から定まる Euclid の距離 $d(\boldsymbol{x},\, \boldsymbol{y})$ を

$$d(\boldsymbol{x},\, \boldsymbol{y}) = |\boldsymbol{x} - \boldsymbol{y}|_{2,\,d}\, (\boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^d)$$

で定める. □

定義 A.16. $(\mathbb{X},\, d)$ を距離空間とする.
(1) $x \in \mathbb{X}$, $r > 0$ に対して

$$B(x;\, r) = \{y \in \mathbb{X};\, d(x,\, y) < r\}$$

を, 中心 x, 半径 r の **開球** という.
(2) $\mathbb{X}$ の部分集合 A が **開集合** であるとは, 任意の $x \in A$ に対して, $\exists r > 0$ を選んで, $B(x;\, r) \subset A$ とできること [1] である.
(3) 開集合の補集合を **閉集合** という.
(4) A が **有界** であるとは, $\exists x \in \mathbb{X}$ と $\exists r > 0$ を選んで $A \subset B(x;\, r)$ とできることである.

定義 A.17. $(\mathbb{X},\, d)$ を距離空間とし, $\{x_n\}_{n\in\mathbb{N}}$ を $\mathbb{X}$ の点列とする.
(1) $x \in \mathbb{X}$ とする. 点列 $\{x_n\}_{n\in\mathbb{N}}$ が x に収束するとは

$$\lim_{n\to\infty} d(x_n,\, x) = 0$$

がなりたつことである.
(2) 点列 $\{x_n\}_{n\in\mathbb{N}}$ が Cauchy 列であるとは, 任意の $\epsilon > 0$ に対して, ある $N \in \mathbb{N}$ が存在して, $m \geq N$, $n \geq N$ ならば, $d(x_m,\, x_n) < \epsilon$ となることである.
(3) Cauchy 列が常に収束するとき, 距離空間 $\mathbb{X}$ は **完備** であるという.

定義 A.18. (1) $(\mathbb{X},\, d_X)$ と $(\mathbb{Y},\, d_Y)$ を距離空間とする. 写像 $f : \mathbb{X} \to \mathbb{Y}$ が点 $x \in \mathbb{X}$ で **連続** であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ が存在して, 任意の $x' \in \mathbb{X}$ に対して

$$d_X(x,\, x') < \delta \;\Rightarrow\; d_Y(f(x),\, f(x')) < \epsilon$$

[1] C, D を命題とする. そのとき, C が成立しないならば, 命題 「$C \;\Rightarrow\; D$」 は真と約束する. したがって, $A = \varnothing$ の場合には, 任意の $x \in A$ が取れないので, 空集合も開集合である. 本によっては, 上記のような論理を踏まずに空集合は開集合であることを約束する場合もあるようだ. この論法を使うと空集合は閉集合でもあることがわかる. Euclid 位相空間において, 空集合と全体集合は常に開集合でもあり, 閉集合である.

をみたすときをいう. 任意の $x \in \mathbb{X}$ で f が連続ならば, 単に f は連続という.
(2) 写像 f が一様連続であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ が存在して, 任意の $x, x' \in \mathbb{X}$ に対して

$$d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon$$

がなりたつときをいう.

定義 A.19. X を集合, $(\mathbb{Y}, d)$ を距離空間とする. $f, f_1, f_2, \ldots, f_n, \ldots$ を X から $\mathbb{Y}$ への写像の列とする.
(1) $\{f_n\}_{n\in\mathbb{N}}$ が f に 各点収束 するとは, 任意の $x \in X$ に対して

$$\lim_{n\to\infty} d(f_n(x), f(x)) = 0$$

であることをいう.
(2) X の部分集合 $A\,(\neq \varnothing)$ に対して, $\{f_n\}_{n\in\mathbb{N}}$ が f に A 上で 一様収束 するとは

$$\lim_{n\to\infty} \sup_{x\in A} d(f_n(x), f(x)) = 0$$

がなりたつときをいう.

A.4.2　位相空間

定義 A.20. $\mathbb{X}$ を空でない集合とする. $\mathbb{X}$ 上の集合族 $\mathcal{O} \subset 2^{\mathbb{X}}$ が次の条件 (1) ∼ (3) をみたすとき, $\mathcal{O}$ を $\mathbb{X}$ の 位相 という. 組 $(\mathbb{X}, \mathcal{O})$ を 位相空間 という. $\mathcal{O}$ の元を $\mathbb{X}$ の開集合という.
(1) $\varnothing, \mathbb{X} \in \mathcal{O}$.
(2) $A, B \in \mathcal{O}$ ならば, $A \cap B \in \mathcal{O}$.
(3) $\{A_\lambda\}_{\lambda\in\Lambda}$ を　$\mathcal{O}$ の元から成る集合族としたとき

$$\bigcup_{\lambda\in\Lambda} A_\lambda \in \mathcal{O}$$

である.

定義 A.21. $(\mathbb{X}, \mathcal{O})$ を位相空間とする.
(1) $A \subset \mathbb{X}$ とする. A のすべての部分開集合からなる族を $\{O_\lambda\}_{\lambda\in\Lambda}$ とする. このとき, A の 内部 A° を

$$A^\circ = \bigcup_{\lambda\in\Lambda} O_\lambda$$

により定める. また, A° の元を A の **内点** という.
(2) $A \subset \mathbb{X}$ とする. A を含むすべての $\mathbb{X}$ の部分閉集合から成る族を $\{C_\lambda\}_{\lambda\in\Lambda}$ とする. このとき, A の **閉包** $\mathrm{cl}(A)$ を

$$\mathrm{cl}(A) = \bigcap_{\lambda\in\Lambda} C_\lambda$$

により定める. また, $\mathrm{cl}(A)$ の元を A の **触点** という.
(3) $A \subset \mathbb{X}$ とする. $\partial A = \mathrm{cl}(A) \setminus A^\circ$ とし, A の **境界** という.
(4) $A \subset \mathbb{X}$ とする. $x \in A^\circ$ のとき, A は x の **近傍** という.
(5) $\varnothing \neq A \subset \mathbb{X}$ とする. $\mathcal{O}_A = \{O \cap A;\, O \in \mathcal{O}\}$ とすると, $\mathcal{O}_A$ は A の位相となる. これを A 上の **相対位相** という.
(6) $A \subset \mathbb{X}$ とする. A が $\mathbb{X}$ において **稠密** であるとは, $\mathrm{cl}(A) = \mathbb{X}$ となることである.
(7) $\mathbb{X}$ が **可分** であるとは, $\mathbb{X}$ の高々可算な部分集合 A で, $\mathbb{X}$ において稠密なものが存在するときである.
(8) $\mathbb{X}$ が **コンパクト** であるとは, $\mathbb{X}$ の任意の **開被覆** $\{O_\lambda\}_{\lambda\in\Lambda}$, すなわち, $\mathbb{X} = \bigcup_{\lambda\in\Lambda} O_\lambda$ に対して, Λ の有限部分集合 Λ_0 が存在して, $\mathbb{X} = \bigcup_{\lambda\in\Lambda_0} O_\lambda$ となるときをいう.

定義 A.22. $\mathbb{X}$, $\mathbb{Y}$ を位相空間とする.
(1) 写像 $f : \mathbb{X} \to \mathbb{Y}$ が **連続** であるとは, $\mathbb{Y}$ の任意の開集合 B に対して, $f^{-1}(B)$ は $\mathbb{X}$ の開集合となるときをいう. この定義は距離空間の連続写像の定義と整合している.
(2) 全単射な連続写像 $f : \mathbb{X} \to \mathbb{Y}$ で逆写像 f^{-1} も連続のとき, $\mathbb{X}$ と $\mathbb{Y}$ は **同相** であるという. また, f を **同相写像** という.

命題 A.23. (1) $\mathbb{X}$, $\mathbb{Y}$ を位相空間とし, $f : \mathbb{X} \to \mathbb{Y}$ を連続写像とする. $\mathbb{X}$ のコンパクト部分集合 A に対して, $f(A)$ は $\mathbb{Y}$ のコンパクト部分集合となる.
(2) $(\mathbb{X}, d_X)$ をコンパクト距離空間, $(\mathbb{Y}, d_Y)$ を距離空間とする. $\mathbb{X}$ から $\mathbb{Y}$ への連続写像は一様連続である.

Proof. 証明は省略. □

A.5 $\mathbb{R}$ の拡張

集合 $\mathbb{R} \cup \{-\infty, \infty\}$ を $\overline{\mathbb{R}}$ と書くことにする. $\overline{\mathbb{R}}$ での演算を以下のように定める.

- 実数に関する演算は通常通り.
- $a \in \mathbb{R}$ に対して, $-\infty < a < \infty$.

- $a \in \mathbb{R}$ に対して, 以下のように定める.
  - $a + (\pm\infty) = \pm\infty$, $\pm\infty + a = \pm\infty$.
  - $a > 0$ のとき, $a \times (\pm\infty) = \pm\infty$, $\pm\infty \times a = \pm\infty$.
  - $a + (\pm\infty) = \pm\infty$, $\pm\infty + a = \pm\infty$.
  - $a < 0$ のとき, $a \times (\pm\infty) = \mp\infty$, $\pm \times a = \mp\infty$.
  - $0 \times (\pm\infty) = 0$, $\pm\infty \times 0 = 0$.
  - $a/(\pm\infty) = 0\,(a \in \mathbb{R})$.
- $(+\infty) + (+\infty) = +\infty$, $(-\infty) + (-\infty) = -\infty$, $(+\infty) - (-\infty) = +\infty$, $(-\infty) - (+\infty) = -\infty$.
- $(\pm\infty) \times (\pm\infty) = +\infty$, $(\pm\infty) \times (\mp\infty) = -\infty$.
- $(\pm\infty)/(\pm\infty)$, $(\pm\infty)/(\mp\infty)$ は定義されないものとする.
- $|\pm\infty| = +\infty$.
- $\overline{\mathbb{R}}$ の部分集合 $A\,(\neq \varnothing)$ に対して, 上限 $\sup A$ と下限 $\inf A$ が定まり, $\overline{\mathbb{R}}$ に値をとる.
- $a, b \in \overline{\mathbb{R}}$ について, 閉区間 $[a, b]$ と開区間 (a, b) を

  $$[a, b] = \{x \in \overline{\mathbb{R}};\, a \le x \le b\}, \qquad (a, b) = \{x \in \overline{\mathbb{R}};\, a < x < b\}$$

  により定める.

なお, 以下の 4 つは定義しない.

$$(+\infty) + (-\infty),\ (-\infty) + (+\infty),\ (+\infty) - (+\infty),\ (-\infty) - (-\infty).$$

A.6　章末注釈と参考文献

節 A.1 は [39] を借用した. 節 A.2 は [35, pp.26 – 29] を借用した. 節 A.3 は [39] を借用した. 節 A.4 は [39] と [35] を参照した. 節 A.5 は [39, pp.18 – 19] を借用した.

第B章　補遺

節 B.1 では大数の強法則の証明を述べる. 節 B.2 では大数の強法則の証明を述べる. 節 B.3 では Dynkin の補題の証明を述べる. 節 B.4 では Dynkin の補題の応用例をいくつか述べる. 節 B.5 では絶対連続関数の性質をまとめた. 節 B.6 では凸関数の性質をまとめた. 節 B.7 では Radon–Nikodym により条件付期待値を導入する. 節 B.8 では Radon–Nikodym の定理により条件付き期待値を導入する. 節 B.9 では Radon–Nikodym の定理により Neyman–Fisher の因子分解定理の証明を述べる. 節 B.10 では Dini の定理の証明を述べる.

B.1　大数の強法則の証明

B.1.1　証明のための準備の補題

大数の強法則を証明するための Borel-Cantelli の補題を述べる.

補題 B.1 (Borel-Cantelli の補題)**.** $\sum_{n=1}^{\infty} \mathsf{Pr}(E_n) < \infty$ なる任意の事象列 $\{E_n\}_{n=1}^{\infty}$ に対して

$$\mathsf{Pr}\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} E_n\right) = \mathsf{Pr}(E_n,\ \text{i.o.}) = 0$$

となる.

Proof. $G_k := \bigcup_{n=k}^{\infty} E_n$, $G := \limsup_{n\to\infty} E_n = \bigcap_{k=1}^{\infty} G_k$ とおく. すると $G_k \supset G_{k+1}\,(k = 1, 2, \ldots)$ で $G = \bigcap_{k=1}^{\infty} G_k$ である. 任意の $k \in \mathbb{N}$ に対して

$$\begin{aligned}\mathsf{Pr}(G) &\le \mathsf{Pr}(G_k) \qquad (\because \text{補題 } 1.7(4))\\ &\le \sum_{n=k}^{\infty} \mathsf{Pr}(E_n). \qquad (\because \text{補題 } 1.7(7))\end{aligned}$$

となる. $\sum_{n=1}^{\infty} \mathsf{Pr}(E_n) < \infty$ なので, $\sum_{n=k}^{\infty} \mathsf{Pr}(E_n) \overset{k\to\infty}{\longrightarrow} 0$ となるので, $\mathsf{Pr}(G_k) \overset{k\to\infty}{\longrightarrow} 0$ である. ここで, 補題 1.7(6) を用いると

$$\mathsf{Pr}(G) = \mathsf{Pr}\left(\bigcap_{k=1}^{\infty} G_k\right) = \lim_{k\to\infty} \mathsf{Pr}(G_k) = 0$$

がわかる. □

補題 B.2 (Borel-Cantelli の逆)**.** $\sum_{n=1}^{\infty}\mathsf{Pr}(E_n)=\infty$ なる任意の独立な事象列 $\{E_n\}_{n=1}^{\infty}$ に対して

$$\mathsf{Pr}\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}E_n\Big)=\mathsf{Pr}(E_n,\text{ i.o.})=1$$

となる.

Proof. $p_n:=\mathsf{Pr}(E_n)\,(n=1,\,2,\,\ldots)$ とおく. 独立性から

$$\mathsf{Pr}\Big(\bigcap_{n=k}^{\infty}E_n^c\Big)=\prod_{n=k}^{\infty}(1-p_n)\qquad(\forall k\in\mathbb{N})$$

となる. これは, $r\geq k$ に対して

$$\mathsf{Pr}\Big(\bigcap_{n=k}^{r}E_n^c\Big)=\prod_{n=k}^{r}(1-p_n)$$

となる. 実際

$$\mathsf{Pr}\Big(\bigcap_{n=k}^{\infty}E_n^c\Big)=\lim_{r\to\infty}\mathsf{Pr}\Big(\bigcap_{n=k}^{r}E_n^c\Big)=\lim_{r\to\infty}\prod_{n=k}^{r}(1-p_n)=\prod_{n=k}^{\infty}(1-p_n)$$

がわかる. つぎに, $1-x\leq e^{-x}\,(x\geq 0)$ に注意すると, 任意の $k\in\mathbb{N}$ に対して

$$0\leq\lim_{r\to\infty}\prod_{n=k}^{r}(1-p_n)\leq\lim_{r\to\infty}\prod_{n=k}^{r}e^{-p_n}=\lim_{r\to\infty}e^{-\sum_{n=k}^{r}p_n}=e^{-\sum_{n=k}^{\infty}p_n}=0$$

となる. よって

$$\mathsf{Pr}\Big(\bigcap_{n=k}^{\infty}E_n^c\Big)=0\tag{B.1}$$

がわかる. しかし

$$\big(\limsup E_n\big)^c=\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}E_n\Big)^c=\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}E_n^c$$

であることと補題 1.7(5)　に注意すると

$$\begin{aligned}\mathsf{Pr}\Big(\big(\limsup E_n\big)^c\Big)&=\mathsf{Pr}\Big(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}E_n^c\Big)=\lim_{k\to\infty}\mathsf{Pr}\Big(\bigcap_{n=k}^{\infty}E_n^c\Big)\\&=0\qquad(\because\text{(B.1)})\end{aligned}$$

が示せた.

つぎに, 大数法則を示すための補題をふたつ述べる.

補題 B.3. $n \in \mathbb{N}$ に対して $Y_n = X_n \mathbb{1}\{|X_n| \leq n\}$ とする. このとき

$$S_n := X_1 + X_2 + \cdots + X_n,$$
$$T_n := Y_1 + Y_2 + \cdots + Y_n$$

とすると

$$\frac{T_n}{n} \overset{\text{a.s.}}{\to} \mu\,(n \to \infty) \;\Rightarrow\; \frac{S_n}{n} \overset{\text{a.s.}}{\to} \mu\,(n \to \infty)$$

が成立する.

Proof. Y_n の定義と系 2.9 から

$$\sum_{n=1}^{\infty} \mathsf{Pr}\big(X_n \neq Y_n\big) = \sum_{n=1}^{\infty} \mathsf{Pr}\big(|X_n| > n\big) \leq \int_0^{\infty} \mathsf{Pr}\big(|X_1| > t\big)\,\mathrm{d}t = \mathsf{E}\big[|X_1|\big] < \infty$$

となる. よって Borel-Cantelli の第 1 の補題 (補題 B.1) から

$$\mathsf{Pr}\Big(X_n \neq Y_n, \text{ i.o.}\Big) = 0$$

がわかる. すると $\omega \in \Omega$ に対して有限個の n に対して $X_n \neq Y_n$ なので

$$\lim_{n\to\infty} \frac{|S_n(\omega) - T_n(\omega)|}{n} = 0$$

となる. よって定理の主張は証明された. □

補題 B.4. 定理 4.19 と補題 B.3 の仮定のもと

$$\sum_{n=1}^{\infty} \frac{\mathsf{Var}\big(Y_n\big)}{n^2} \leq 4\mathsf{E}\big[|X_1|\big] < \infty$$

が成立する.

Proof. 系 2.9 を用いて, $Y_n = X_n \mathbb{1}\{|X_n| \leq n\}$ と $\{X_n\}_{n=1}^{\infty}$ は i.i.d. 確率変数列であること思い出すと

$$\begin{aligned}
\mathsf{Var}\big[Y_n\big] \leq \mathsf{E}\big[Y_n^2\big] &= \int_0^{\infty} 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t = \int_0^{n} 2t\,\mathsf{Pr}\big(|Y_n| > t\big)\,\mathrm{d}t \\
&= \int_0^{n} 2t\,\mathsf{Pr}\big(|X_n| > t\big)\,\mathrm{d}t = \int_0^{n} 2t\,\mathsf{Pr}\big(|X_1| > t\big)\,\mathrm{d}t \\
&= \int_0^{\infty} \mathbb{1}\{t < n\} 2t\,\mathsf{Pr}\big(|X_1| > t\big)\,\mathrm{d}t
\end{aligned}$$

となる. よって, Tonelli の定理より

$$\sum_{n=1}^{\infty} \frac{\mathsf{E}\big[Y_n^2\big]}{n^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} \int_0^{\infty} \mathbb{1}\{t < n\} 2t\,\mathsf{Pr}\big(|X_1| > t\big)\,\mathrm{d}t$$

$$= \int_0^\infty \left\{ \sum_{n=1}^\infty \frac{\mathbb{1}\{t < n\}}{n^2} \right\} 2t\, \mathsf{Pr}\big(|X_1| > t\big)\, \mathrm{d}t$$

となる. 最後の等号は, 各項が非負なので, Fubini の定理を用いることで和と積分の記号の交換が保証されることがわかる. よって後は $\forall t \geq 0$ に対して

$$2t \sum_{n>t} \frac{1}{n^2} \leq 4 \tag{B.2}$$

を示せばよい. $m \geq 2\,(m \in \mathbb{N})$ に対して

$$\sum_{n \geq m} \frac{1}{n^2} \leq \int_{m-1}^\infty \frac{\mathrm{d}x}{x^2} = \frac{1}{m-1}$$

となる. $t \geq 1$ のとき, (B.2) の和は $n = [t] + 1 \geq 2$ で始まるので

$$2t \sum_{n>t} \frac{1}{n^2} \leq \frac{2t}{[t]} \leq 4 \tag{B.3}$$

となる [1]. $0 \leq t \leq 1$ の場合

$$2t \sum_{n>t} \frac{1}{n^2} \leq 2\left(1 + \sum_{n=2}^\infty \frac{1}{n^2}\right) \leq 4 \tag{B.4}$$

となる. よって (B.3), (B.4) から (B.2) がわかる. □

B.1.2 大数の強法則 (定理 4.19) の証明

$n \in \mathbb{N}$ に対して

$$X_n^+ := \max\{X_n,\, 0\}, \qquad X_n^- := \max\{-X_n,\, 0\}$$

とおくと X_n^+ と X_n^- は定理 4.19 の仮定をみたし, $X_n = X_n^+ - X_n^-$ である. したがって一般性を失うことなく $X_n \geq 0\,(n \in \mathbb{N})$ と仮定してよい.

$\alpha > 1$ をとり

$$k(n) = [\alpha^n]$$

と定める. $\forall \epsilon > 0$ を取る. Chebyshev の不等式から

$$\sum_{n=1}^\infty \mathsf{Pr}\left(\big|T_{k(n)} - \mathsf{E}\big[T_{k(n)}\big]\big| > \epsilon k(n) \right) \leq \frac{1}{\epsilon^2} \sum_{n=1}^\infty \frac{\mathsf{Var}\big[T_{k(n)}\big]}{\{k(n)\}^2}$$

[1] なぜならば

$$\frac{t}{[t]} \leq 2 \qquad (t \geq 1)$$

であることからわかる.

$$= \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{1}{\{k(n)\}^2} \sum_{m=1}^{k(n)} \mathsf{Var}\big[Y_m\big]$$
$$= \frac{1}{\epsilon^2} \sum_{m=1}^{\infty} \mathsf{Var}\big[Y_m\big] \sum_{n:\, k(n)\geq m} \frac{1}{\{k(n)\}^2} \tag{B.5}$$

を得る. 最後の等号は, すべての項が非負なので Fubini の定理を用いて和の順番を交換した. $n \geq 1$ に対して, $k(n) = [\alpha^n]$ とおく. このとき, 十分大きな n に対して, $[\alpha^n] \geq \frac{\alpha^n}{2}$ となるので

$$\sum_{n:\, \alpha^n \geq m} \frac{1}{[\alpha^n]^2} \leq 4 \sum_{n:\, \alpha^n \geq m} \frac{1}{\alpha^{2n}} \leq \frac{4}{(1-\alpha^{-2})m^2} \tag{B.6}$$

となる. (B.5) と (B.6) を合わせると

$$\sum_{n=1}^{\infty} \mathsf{Pr}\bigg(\big| T_{k(n)} - \mathsf{E}\big[T_{k(n)}\big] \big| > \epsilon k(n) \bigg) \leq \frac{4}{(1-\alpha^{-2})\epsilon^2} \sum_{m=1}^{\infty} \frac{\mathsf{E}[Y_m^2]}{m^2}$$
$$< \infty \qquad (\because \text{補題 B.4})$$

となる. ϵ は任意だったので

$$\frac{T_{k(n)} - \mathsf{E}[T_{k(n)}]}{k(n)} \overset{\text{a.s.}}{\to} 0 \qquad (n \to \infty)$$

を得る. $\{X_n\}_{n=1}^{\infty}$ は i.i.d. 確率変数列で $Y_n = X_n \mathbb{1}\{|X_n| \leq n\}$ であったことを思います. $\mathsf{E}[|X_1|] < \infty$ のとき, すると $Y_n - X_n \overset{\text{a.s.}}{\to} 0\,(n \to \infty)$ となる [2] ので, 有界収束定理から

$$\lim_{n\to\infty} \mathsf{E}[Y_n] = \lim_{n\to\infty} \mathsf{E}[X_n + (Y_n - X_n)] = \lim_{n\to\infty} \underbrace{\mathsf{E}[X_n]}_{=\mathsf{E}[X_1]} + \lim_{n\to\infty} \mathsf{E}[Y_n - X_n] = \mathsf{E}[X_1]$$

[2] 実際, $X_n \neq Y_n$ と $|X_n| > n$ は同値である. $\{X_n\}_{n=1}^{\infty}$ は i.i..d. 確率変数列なので

$$\sum_{n=1}^{\infty} \mathsf{Pr}(|X_n| > n) = \sum_{n=1}^{\infty} \mathsf{Pr}(|X_1| > n)$$

である. $\mathbb{E}[|X_1|] < \infty$ と系 2.9 から

$$\sum_{n=1}^{\infty} \mathsf{Pr}(|X_1| > n) < \infty$$

となる. したがって, Borel–Cantelli の補題から

$$\mathsf{Pr}(|X_n| > n \,\text{i.o.}) = 0$$

となる. したがって, $X_n - Y_n \overset{\text{a.s.}}{\to} 0\,(n \to \infty)$ がわかる.

である. よって

$$\frac{\mathsf{E}\big[T_{k(n)}\big]}{k(n)} \overset{n\to\infty}{\longrightarrow} \mathsf{E}[X_1]$$

を得る. これより

$$\frac{T_{k(n)}}{k(n)} \overset{\text{a.s.}}{\to} \mathsf{E}[X_1] \qquad (n\to\infty)$$

がわかる. $Y_n \geq 0\,(\forall n\in\mathbb{N})$ なので, $k(n)\leq m < k(n+1)\,(m\in\mathbb{N})$ に対して

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)}$$

である. $k(n)=[\alpha^n]$ だったので,

$$\frac{k(n+1)}{k(n)} \overset{n\to\infty}{\longrightarrow} \alpha$$

である. よって

$$\frac{1}{\alpha}\mathsf{E}[X_1] \leq \liminf_{m\to\infty}\frac{T_m}{m} \leq \limsup_{m\to\infty}\frac{T_m}{m} \leq \alpha\mathsf{E}[X_1]$$

となる. $\alpha>1$ は任意だったので

$$\frac{T_m}{m} \overset{\text{a.s.}}{\to} \mathsf{E}[X_1] \qquad (m\to\infty)$$

がわかる. □

B.2 中心極限定理の証明

まず $\mathsf{E}[X_1^3]<\infty$ を仮定して証明をする.

$$Y_j = \frac{X_j-\mu}{\sigma}, \qquad (j=1,\,2,\,\ldots,\,n)$$

とおくと $\mathsf{E}[Y_1]=0,\,\mathsf{E}[Y_1^2]=1,\,\mathsf{E}[|Y_1|^3]<\infty$ となることに注意する. $Z_1,\,Z_2,\,\ldots,\,Z_n \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,\,1)$ とし

$$S_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Y_j, \quad S = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_j \sim \mathsf{N}(0,\,1)$$

とおく.

$\forall x \in \mathbb{R}$ に対して

$$\mathsf{Pr}\big(S_n \le x\big) \overset{n\to\infty}{\longrightarrow} \mathsf{Pr}\big(S \le x\big) \tag{B.7}$$

を示せばよい. ここで

$$g(y) = \mathbb{1}_{(-\infty,\, x]}(y) \qquad (y \in \mathbb{R})$$

とおくと (B.7) を

$$\mathsf{E}\big[g(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[g(S)\big] \tag{B.8}$$

と書きかえることができる. さらに $\forall \delta > 0$ に対して関数 $g_{\delta,\, x} : \mathbb{R} \longrightarrow [0,\, 1]$ を以下のような性質をもつ関数とする.

- $g_{\delta,\, x}(y) = 1\,(y \le x)$.
- $g_{\delta,\, x}(y) = 0\,(y \ge x + \delta)$.
- $g_{\delta,\, x}$ は有界な 3 次の導関数をもつ. 3 次の導関数の絶対値の上界を C と書くことにする.

さらに $g_{\delta,\, x-\delta}(y) = g_{\delta,\, x}(y + \delta)$ と定める. すると

$$g_{\delta,\, x-\delta}(y) \le g(y) \le g_{\delta,\, x}(y), \qquad (y \in \mathbb{R})$$

となる. このことから $\forall x \in \mathbb{R}$ に対して

$$\begin{aligned}
\mathsf{Pr}\big(S_n \le x\big) &\le \mathsf{E}\big[g_{\delta,\, x}(S_n)\big] \\
&= \mathsf{E}\big[g_{\delta,\, x}(S_n)\big] - \mathsf{E}\big[g_{\delta,\, x}(S)\big] + \mathsf{E}\big[g_{\delta,\, x}(S)\big] \\
&\le \mathsf{Pr}\big(S \le x + \delta\big) + \big|\mathsf{E}\big[g_{\delta,\, x}(S_n)\big] - \mathsf{E}\big[g_{\delta,\, x}(S)\big]\big|
\end{aligned}$$

を得る. 同様に

$$\mathsf{Pr}\big(S_n \le x\big) \ge \mathsf{Pr}\big(S \le x - \delta\big) - \big|\mathsf{E}\big[g_{\delta,\, x-\delta}(S_n)\big] - \mathsf{E}\big[g_{\delta,\, x-\delta}(S)\big]\big|$$

を得る. これらの不等式を合わせると

$$\begin{aligned}
&\mathsf{Pr}\big(S \le x + \delta\big) + \big|\mathsf{E}\big[g_{\delta,\, x}(S_n)\big] - \mathsf{E}\big[g_{\delta,\, x}(S)\big]\big| \\
&\quad \ge \mathsf{Pr}\big(S_n \le x\big) \\
&\quad \ge \mathsf{Pr}\big(S \le x - \delta\big) - \big|\mathsf{E}\big[g_{\delta,\, x-\delta}(S_n)\big] - \mathsf{E}\big[g_{\delta,\, x-\delta}(S)\big]\big|
\end{aligned}$$

となる. δ を 0 に近づけると $\mathsf{Pr}\big(S \le x \pm \delta\big)$ は $\mathsf{Pr}\big(S \le x\big)$ に近づく. したがって

$$\mathsf{E}\big[g_{\delta,\, x}(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[g_{\delta,\, x}(S)\big] \quad \text{と} \quad \mathsf{E}\big[g_{\delta,\, x-\delta}(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[g_{\delta,\, x-\delta}(S)\big] \tag{B.9}$$

を示せばよい. よって $h : \mathbb{R} \to [0, 1]$ を有界な 3 次の導関数をもつ関数として (B.9) を示す.

$1 \le m \le n+1$ に対して

$$T_m = \frac{1}{\sqrt{n}}\big(Y_1 + \cdots + Y_{m-1} + Z_m + \cdots + Z_n\big)$$

とおく. さらに, $T_1 = S_n$ と $T_{n+1} = S$ とする. このとき

$$\begin{aligned}\big|\mathsf{E}\big[h(S_n)\big] - \mathsf{E}\big[h(S)\big]\big| &= \Big|\sum_{m=1}^{n}\big\{\mathsf{E}\big[h(T_m)\big] - \mathsf{E}\big[h(T_{m+1})\big]\big\}\Big| \\ &\le \sum_{m=1}^{n}\big|\mathsf{E}\big[h(T_m) - h(T_{m+1})\big]\big| \qquad \text{(B.10)}\end{aligned}$$

と書きかえる. 次に $m = 1, 2, \ldots, n$ に対して

$$U_m = \frac{1}{\sqrt{n}}\big(Y_1 + \cdots + Y_{m-1} + Z_{m+1} + \cdots + Z_n\big)$$

とおくと

$$T_m = U_m + \frac{1}{\sqrt{n}}Z_m, \qquad T_{m+1} = U_m + \frac{1}{\sqrt{n}}Y_m$$

となる. Taylor 展開を用いると

$$\begin{aligned}h(T_m) - h(U_m) &= (T_m - U_m)\,\dot{h}\,(U_m) + \frac{1}{2}(T_m - U_m)^2\,\ddot{h}\,(U_m) \\ &\quad + \frac{1}{6}(T_m - U_m)^3\,\dddot{h}\,(U_m^*), \qquad \text{(B.11)} \\ h(T_{m+1}) - h(U_m) &= (T_{m+1} - U_m)\,\dot{h}\,(U_m) + \frac{1}{2}(T_{m+1} - U_m)^2\frac{1}{2}\,\dot{h}\,(U_m) + \ddot{h}\,(U_m) \\ &\quad + \frac{1}{6}(T_{m+1} - U_m)^3\,\dddot{h}\,(U_m^{**}), \qquad \text{(B.12)}\end{aligned}$$

を得る. ただし $\dot{h}$, $\ddot{h}$, $\dddot{h}$ は h の 1 次, 2 次, 3 次の導関数で, U_m^*, U_m^{**} は T_m と U_m の間と T_{m+1} と U_m の間の数である. U_m は Y_m と Z_m の両方とも独立であるので

$$\begin{aligned}&\mathsf{E}\big[\dot{h}\,(U_m)Z_m\big] = \mathsf{E}\big[\dot{h}\,(U_m)Y_m\big] = 0, \\ &\mathsf{E}\big[\ddot{h}\,(U_m)Z_m^2\big] = \mathsf{E}\big[\ddot{h}\,(U_m)\big] = \mathsf{E}\big[\ddot{h}\,(U_m)Y_m^2\big], \\ &\big|\mathsf{E}\big[\dddot{h}\,(U_m^*)\big]\big| \le C,\ \big|\mathsf{E}\big[\dddot{h}\,(U_m^{**})\big]\big| \le C\end{aligned}$$

となる. 上の期待値と (B.11) と (B.12) を合わせると

$$\Big|\mathsf{E}\big[h(T_m) - h(T_{m+1})\big]\Big| = \Big|\mathsf{E}\big[\big(h(T_m) - h(U_m)\big) - \big(h(T_{m+1}) - h(U_m)\big)\big)\big]\Big|$$

$$\leq \left| \frac{C}{6n^{3/2}} \mathsf{E}\big[|Z_m|^3 + |Y_m|^3\big] \right|$$

を得る. この式を (B.10) に代入すると

$$\left| \mathsf{E}\big[h(S_n)\big] - \mathsf{E}\big[h(S)\big] \right| \leq \frac{C}{6\sqrt{n}} \mathsf{E}\big[|Z_1|^3 + |Y_1|^3\big] \overset{n\to\infty}{\longrightarrow} 0$$

を得る. よって Y_1 が有界な 3 次の積率をもつとき

$$S_n \rightsquigarrow S \sim \mathsf{N}(0,\, 1) \qquad (n \to \infty)$$

が示せた.

最後に $\mathsf{E}\big[Y_1^3\big] < \infty$ の仮定を除くための議論を行う. そのために任意の $\epsilon > 0$ とし, $\sup_{x\in\mathbb{R}} |\overset{\bullet\bullet}{h}(y)| \leq C'$, $\sup_{x\in\mathbb{R}} |\overset{\bullet\bullet\bullet}{h}(y)| \leq C$ をみたす定数 C をとる. $m = 1, 2, \ldots, n$ に対して, $|Y_m| \leq \epsilon\sqrt{n}$ のとき

$$\begin{aligned}
&\left| h(T_m) - h(U_m) - \frac{\overset{\bullet}{h}(U_m)Y_m}{\sqrt{n}} - \frac{\overset{\bullet\bullet}{h}(U_m)Y_m^2}{2n} \right| \mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\} \\
&\quad \leq C \left| \frac{\overset{\bullet\bullet\bullet}{h}(U_m^*)Y_m^3}{6n^{3/2}} \right| \mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\} \\
&\quad \leq \frac{C\epsilon}{6n} Y_m^2 \mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\}
\end{aligned}$$

となる. 一方 $|Y_m| > \epsilon\sqrt{n}$ のとき

$$\left| h(T_m) - h(U_m) - \frac{\overset{\bullet}{h}(U_m)Y_m}{\sqrt{n}} - \frac{\overset{\bullet\bullet}{h}(U_m)Y_m^2}{2n} \right| \mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}$$

$$\leq \left\{ \left| h(T_m) - h(U_m) - \frac{\overset{\bullet}{h}(U_m)Y_m}{\sqrt{n}} \right| + \left| \frac{\overset{\bullet\bullet}{h}(U_m)Y_m^2}{2n} \right| \right\} \mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}$$

$$\begin{aligned}
&\leq \left\{ \left| \frac{\overset{\bullet\bullet}{h}(U_m^{***})Y_m^2}{2n} \right| + \left| \frac{\overset{\bullet\bullet}{h}(U_m)Y_m^2}{2n} \right| \right\} \mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\} \\
&\leq \frac{C'Y_m^2}{n} \mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}
\end{aligned}$$

となる. ただし U_m^{***} は T_m と U_m の間にあるものである. これらの不等式を合わせると

$$\left| h(T_m) - h(U_m) - \frac{\overset{\bullet}{h}(U_m)Y_m}{\sqrt{n}} - \frac{\overset{\bullet\bullet}{h}(U_m)Y_m^2}{2n} \right|$$

$$\leq \frac{C\epsilon}{6n} Y_m^2 \mathbb{1}\{|Y_m| \leq \epsilon\sqrt{n}\} + \frac{C' Y_m^2}{n} \mathbb{1}\{|Y_m| > \epsilon\sqrt{n}\}$$

となる. この式と (B.12) を合わせると

$$\begin{aligned}
\big|\mathsf{E}\big[h(S_n)\big] - \mathsf{E}\big[h(S)\big]\big| &\leq \frac{C}{\sqrt{n}} \mathsf{E}\big[Z_1^3\big] + \frac{C}{6}\epsilon \mathsf{E}\big[Y_1^2 \mathbb{1}\{|Y_1| \leq \epsilon\sqrt{n}\}\big] \\
&\qquad + C' \mathsf{E}\big[Y_1^2 \{Y_1 > \epsilon\sqrt{n}\}\big] \\
&\leq \frac{C}{\sqrt{n}} \mathsf{E}\big[Z_1^3\big] + \frac{C\epsilon}{6} + \frac{C'}{6} \mathsf{E}\big[Y_1^2 \{Y_1 > \epsilon\sqrt{n}\}\big] \\
&\overset{n\to\infty}{\longrightarrow} \frac{C'}{6}\epsilon
\end{aligned}$$

を得る. 上の式で $\epsilon \to 0$ とすれば

$$\mathsf{E}\big[h(S_n)\big] \overset{n\to\infty}{\longrightarrow} \mathsf{E}\big[h(S)\big]$$

が証明できた. □

B.3 Dynkin 族の議論

定義 B.5. Ω を空でない集合とし, $\mathcal{L}$ と $\mathcal{P}$ を Ω の部分集合族とする.
(1) 集合族 $\mathcal{P}$ が π システム であるとは

$$A,\, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$$

が成立するときをいう.
(2) 集合族 $\mathcal{L}$ が λ システム であるとは, つぎの条件 (i) $\sim$ (iii) が成り立つときをいう.

(i) $\Omega \in \mathcal{L}$.

(ii) $A,\, B \in \mathcal{L}$ かつ $A \subset B \implies B \setminus A \in \mathcal{L}$.

(iii) $A_j \in \mathcal{L}$ かつ $A_j \subset A_{j+1}\, (j = 1,\, 2,\, \ldots) \implies \bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$.

次に述べる命題は λ システムであるための同値条件である.

命題 B.6. 空でない集合 Ω の部分集合族 $\mathcal{L}$ が λ システムであるために必要十分条件は次の条件 (i) $\sim$ (iii) をすべてみたすことである.

(i)$'$ $\Omega \in \mathcal{L}$.

(ii)$'$ $A \in \mathcal{L} \implies A^c \in \mathcal{L}$.

(iii)$'$ $A_j \in \mathcal{L}\, (j = 1,\, 2,\, \ldots)$ かつ $A_i \cap A_j = \varnothing\, (i \neq j) \implies \bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$.

Proof. ① $(\Rightarrow)$ の証明: 定義 B.5(2) の (i) $\sim$ (iii) を仮定して, (ii)$'$ と (iii)$'$ を確認すればよい.
⑴ⓐ (ii)$'$ の確認: $A \in \mathcal{L}$ のとき, $A \subset \Omega$ かつ $\Omega \in \mathcal{L}$ なので, 定義 B.5(ii) から

$$A^c = \Omega \setminus A \in \mathcal{L}$$

となる.
⑴ⓑ (iii)$'$ の確認: まず, A, $B \in \mathcal{L}$ かつ $A \cap B = \varnothing$ のとき

$$A \cup B \in \mathcal{L} \tag{B.13}$$

を示す.
仮定と (ii)$'$ から $A^c \in \mathcal{L}$ かつ $B \subset A^c$ である. なぜならば, $\forall \omega \in B$ のとき, $A \cap B = \varnothing$ なので, $\omega \notin A$ となる. よって $\omega \in A^c$ がわかるので, $B \subset A^c$ が示せた.
このことと (ii) から $A^c \setminus B \in \mathcal{L}$ となる. さらに

$$A^c \cap B^c = A^c \setminus B \in \mathcal{L} \tag{B.14}$$

となることと (ii)$'$ より

$$A \cup B = \left(A^c \cap B^c\right)^c \in \mathcal{L}$$

がわかる. よって

$$A,\, B \in \mathcal{L} \Longrightarrow A \cup B \in \mathcal{L} \tag{B.15}$$

がわかる.
$A_j \in \mathcal{L}\,(j = 1, 2, \ldots)$ と $A_i \cap A_j = \varnothing\,(i \neq j)$ とする. このとき

$$B_n := \bigcup_{j=1}^{n} A_j \quad (n \in \mathbb{N})$$

とおく. (B.15) より $B_n \in \mathcal{L}$ となる. さらに, B_n の定め方から

$$B_n \subset B_{n+1}$$

となる. よって定義 B.5(iii) から

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{n=1}^{\infty} B_n \in \mathcal{L}$$

がわかる. よって (i)$'$ $\sim$ (iii)$'$ が確認できた.

($\Leftarrow$): ② (ii) の確認: 命題 B.6(i)$'$ $\sim$ (iii)$'$ を仮定して, 定義 B.5(2) の (ii) と (iii) を確認すればよい.
②ⓐ (ii) の確認: $A, B \in \mathcal{L}$ かつ $A \subset B$ とする. すると (ii)$'$ から $A^c, B^c \in \mathcal{L}$ となる. さらに $A \subset B$ から $A \cap B^c = \varnothing$ となる. よって (iii)$'$ から $A \cup B^c \in \mathcal{L}$ となる. 再度, (ii)$'$ を用いると

$$B \setminus A = B \cap A^c = \big(A \cup B^c\big)^c \in \mathcal{L}$$

がわかる.
②ⓑ (iii) の確認: $A_n \in \mathcal{L}$ かつ $A_n \subset A_{n+1}\,(n = 1, 2, \ldots)$ とする. ここで

$$B_1 := A_1, \qquad B_n := A_{n+1} \setminus A_n\,(n \geq 2)$$

とおくと (ii) から $B_n \in \mathcal{L}$ となる. さらに B_n の定め方から

$$B_n \cap B_m = \varnothing\,(m \neq n)$$

となる. よって (iii)$'$ から

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{n=1}^{\infty} B_n \in \mathcal{L}$$

がわかる. 以上から (i) $\sim$ (iii) が確認できた. □

命題 B.7. Ω を空でない集合とし, $\mathcal{C}$ を Ω の部分集合族とする. 部分集合族 $\mathcal{C}$ が π システムかつ λ システムならば, $\mathcal{C}$ は σ 集合族となる.

Proof. ①$\mathcal{C}$ は体であることの証明: そのために, ⓐ$\Omega \in \mathcal{C}$ ⓑ$A \in \mathcal{C} \Rightarrow A^c \in \mathcal{C}$, ⓒ$A, B \in \mathcal{C} \Rightarrow A \cup B \in \mathcal{C}$ を示せばよい.
ⓐ $\Omega \in \mathcal{C}$ の証明: $\mathcal{C}$ は λ システムなので, 命題 B.6(ii)$'$ から $\Omega \in \mathcal{C}$ である.
ⓑ $A \in \mathcal{C} \Rightarrow A^c \in \mathcal{C}$ の証明: $A \in \mathcal{C}$ とする. ⓐ から $\Omega \in \mathcal{C}$ であり, $A \subset \Omega$ である. よって, $\mathcal{C}$ は λ システムであるので, 定義 B.5(2) の (ii) から

$$A^c = \Omega \setminus A \in \mathcal{C}$$

となる.
ⓒ $A, B \in \mathcal{C} \Rightarrow A \cup B \in \mathcal{C}$ の証明: ⓑ から $A^c, B^c \in \mathcal{C}$ である. さらに, $\mathcal{C}$ は π システムなので, $A \cap B^c$, $A^c \cap B$, $A \cap B \in \mathcal{C}$ であり, 互いに排反である. よって, 命題 B.6 の (iii)$'$ から

$$A \cup B = \big(A \cap B^c\big) \cup \big(A^c \cap B\big) \cup \big(A \cap B\big) \in \mathcal{C}$$

がわかる.

⑴ⓐ ∼ ⑴ⓒ より, $\mathcal{C}$ は体であることが確認できた.

② $A_j \in \mathcal{C}\,(j = 1, 2, \ldots) \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{C}$ の証明: $A_j \in \mathcal{C}\,(j = 1, 2, \ldots)$ とする. $\mathcal{C}$ は体なので

$$B_n := \bigcup_{j=1}^{n} A_j \in \mathcal{C} \qquad (n \in \mathbb{N})$$

となる. さらに B_n の定め方から $B_n \subset B_{n+1}\,(n = 1, 2, \ldots)$ となる. よって, $\mathcal{C}$ は λ システムなので, 定義 B.5(2) の (iii) から

$$\bigcup_{j=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{C}$$

がわかる. 以上から $\mathcal{C}$ は σ 集合族であることがわかる. □

命題 B.8. Ω を空でない集合とする. Ω の部分集合族 $\mathcal{P}$ を π システムとし, $\mathcal{L}[\mathcal{P}]$ を $\mathcal{P}$ を含む最小の λ システムとする. このとき, $\mathcal{L}[\mathcal{P}]$ は π システムとなる.

Proof. $A \in \mathcal{L}[\mathcal{P}]$ と $C \in \mathcal{P}$ に対して

$$\mathcal{G}_A := \big\{B \in \sigma[\mathcal{P}];\, A \cap B \in \mathcal{L}[\mathcal{P}]\big\}, \qquad \mathcal{G}_C := \big\{B \in \sigma[\mathcal{P}];\, C \cap B \in \mathcal{L}[\mathcal{P}]\big\},$$

とそれぞれ定める. ただし, $\sigma[\mathcal{P}]$ を $\mathcal{P}$ を含む最小の σ 加法族とする. このとき, 以下の主張を順に証明していく. $\mathcal{G}_A$ に関する以下の主張を

① $A \in \mathcal{L}[\mathcal{P}] \Rightarrow \mathcal{G}_A$ は λ システム,

② $C \in \mathcal{P} \Rightarrow \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_C$

③ $C \in \mathcal{P}$ かつ $A \in \mathcal{L}[\mathcal{P}] \Rightarrow A \cap B \in \mathcal{L}[\mathcal{P}]$

④ $A \in \mathcal{L}[\mathcal{P}] \Rightarrow \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$

を順に証明していき

⑤ $\mathcal{L}[\mathcal{P}]$ は π システム,

を示す.

① $A \in \mathcal{L}[\mathcal{P}] \Rightarrow \mathcal{G}_A$ は λ システム の証明: $A \in \mathcal{P}[\mathcal{P}]$ とする. $\mathcal{G}_A$ は λ システムであることを示すために, 命題 B.6 の条件 (i)′ $\Omega \in \mathcal{G}_A$, (ii)′ $B \in \mathcal{G}_A \Rightarrow B^c \in \mathcal{G}_A$ (iii)′ $B_j \in \mathcal{G}_A\,(j = 1, 2, \ldots,)$ かつ $B_i \cap B_j = \varnothing\,(i \neq j) \Rightarrow \bigcup_{j=1}^{\infty} B_j \in \mathcal{G}_A$, を順に確認していく.

⑴a $\Omega \in \mathcal{G}_A$ の確認: $A \cap \Omega = A \in \mathcal{L}[\mathcal{P}]$ なので, $\mathcal{G}_A$ の定義から $\Omega \in \mathcal{G}_A$ がわかる.
⑴b $B \in \mathcal{G}_A \Rightarrow B^c \in \mathcal{G}_A$ の確認: $B \in \mathcal{G}_A$ とする. このとき

$$B^c \cap A = A \setminus (A \cap B)$$

に注意する. $B \in \mathcal{G}_A$ なので, $\mathcal{G}_A$ の定義から

$$A \cap B \in \mathcal{L}[\mathcal{P}] \tag{B.16}$$

である. 仮定と (B.16) から A, $A \cap B \in \mathcal{L}[\mathcal{P}]$ かつ $A \cap B \subset A$ である. $\mathcal{L}[\mathcal{P}]$ は λ システムなので, 定義 B.5(2) の (ii) から

$$B^c \cap A = A \setminus (A \cap B) \in \mathcal{L}[\mathcal{P}]$$

となる. よって $A \cap B^c \in \mathcal{L}[\mathcal{P}]$ がわかる. したがって, $\mathcal{G}_A$ の定義から $B^c \in \mathcal{G}_A$ となる.
⑴c $B_j \in \mathcal{G}_A\,(j = 1, 2, \ldots,)$ かつ $B_i \cap B_j = \varnothing\,(i \neq j) \Rightarrow \bigcup_{j=1}^{\infty} B_j \in \mathcal{G}_A$ の確認: まず, $\{B_n\}_{n=1}^{\infty}$ は互いに排反な列で $B_n \in \mathcal{G}_A\,(n = 1, 2, \ldots)$ とする. このとき, $\{A \cap B_n\}_{n=1}^{\infty}$ は互いに排反な列であることに注意する. さらに, $A \in \mathcal{L}[\mathcal{P}]$ かつ $B_n \in \mathcal{G}_A$ なので, B.16 から $A \cap B_n \in \mathcal{L}[\mathcal{P}]$ となる. $\{A \cap B_n\}_{n=1}^{\infty} \subset \mathcal{L}[\mathcal{P}]$ であり, $\mathcal{L}[\mathcal{P}]$ は λ システムである. すると, 分配法則と命題 B.6(iii)$'$ から

$$A \cap \left(\bigcup_{j=1}^{\infty} B_j \right) = \bigcup_{j=1}^{\infty} (A \cap B_j) \in \mathcal{L}[\mathcal{P}] \tag{B.17}$$

となることがわかる. ここで, $\mathcal{G}_A$ の定義から, $B_j \in \mathcal{G}_A\,(j = 1, 2, \ldots)$ から $B_j \in \sigma[\mathcal{P}]$ であるので, $\bigcup_{j=1}^{\infty} B_j \in \sigma[\mathcal{P}]$ である. よって, $\mathcal{G}_A$ の定義と (B.17) から

$$\bigcup_{n=1}^{\infty} B_n \in \mathcal{G}_A$$

がわかる.

よって ⑴a $\sim$ ⑴c から $\mathcal{G}_A$ は λ システムであることがわかる. □

② $C \in \mathcal{P} \Rightarrow \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_C$ の証明: $C \in \mathcal{P} \subset \mathcal{L}[\mathcal{P}]$ を取る. いま, $B \in \mathcal{P} \subset \sigma[\mathcal{P}]$ をとる. 命題の仮定から $\mathcal{P}$ は π システムであることを思い出す. よって, 定義 B.5(1) から

$$C \cap B \in \mathcal{P} \subset \mathcal{L}[\mathcal{P}]$$

となる. $\mathcal{G}_C$ の定義を思い出すと $B \in \mathcal{G}_C$ である. よって, $B \in \mathcal{P} \Rightarrow B \in \mathcal{G}_C$ が示せたので, $\mathcal{P} \subset \mathcal{G}_C$ である. 一方, ① の主張から $\mathcal{G}_C$ は λ システムであるので, $\mathcal{G}_C$ は λ システムで $\mathcal{P}$ を含む λ システムとなる. ここで, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含むの最小の λ システムであることを思い出すと

$$\mathcal{G}_C \supset \mathcal{L}[\mathcal{P}]$$

がわかる.

③ $C \in \mathcal{P}$ かつ $A \in \mathcal{L}[\mathcal{P}] \Rightarrow A \cap C \in \mathcal{L}[\mathcal{P}]$ の証明: ② の主張から

$$A \in \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_C \subset \sigma[\mathcal{P}]$$

となる. よって, $\mathcal{G}_C$ の定義から

$$A \cap C \in \mathcal{L}[\mathcal{P}]$$

がわかる.

④ $A \in \mathcal{L}[\mathcal{P}] \Rightarrow \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$ の証明: $C \in \mathcal{P} \subset \sigma[\mathcal{P}]$ と $A \in \mathcal{L}[\mathcal{P}]$ を取る. ③ の主張から

$$A \cap C = C \cap A \in \mathcal{L}[\mathcal{P}] \tag{B.18}$$

となる. よって, $A \in \mathcal{L}[\mathcal{P}]$ で (B.18) であることに注意して, $\mathcal{G}_A$ の定義を思い出すと $C \in \mathcal{G}_A$ となる. 以上から

$$C \in \mathcal{P} \Longrightarrow C \in \mathcal{G}_A$$

がわかる. すなわち, $\mathcal{P} \subset \mathcal{G}_A$ である. さらに, ① の主張から $\mathcal{G}_A$ は λ システムである. 以上から, $\mathcal{G}_A$ は $\mathcal{P}$ を含む λ システムであることがわかる. よって, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含む最小の λ システムなので

$$\mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$$

がわかる. □

⑤ $\mathcal{L}[\mathcal{P}]$ は π システム: $A, B \in \mathcal{L}[\mathcal{P}]$ とする. ④ の主張から, $B \in \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A \subset \sigma[\mathcal{P}]$ である. $\mathcal{G}_A$ の定義から

$$A \cap B \in \mathcal{L}[\mathcal{P}]$$

となり, $\mathcal{L}[\mathcal{P}]$ は π システムであることがわかる. □

定理 B.9. Ω を空でない集合とする. Ω の部分集合族 $\mathcal{P}$ を π システムとし, $\mathcal{L}$ を λ システムかつ $\mathcal{P} \subset \mathcal{L}$ とする. このとき

$$\sigma[\mathcal{P}] \subset \mathcal{L} \tag{B.19}$$

が成立する.

Proof. 命題 B.8 から $\mathcal{L}[\mathcal{P}]$ は π システムである. よって命題 B.7 より $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含む σ 加法族である. よって $\sigma[\mathcal{P}]$ は $\mathcal{P}$ を含む最小の σ 加法族だから

$$\mathcal{L}[\mathcal{P}] \supset \sigma[\mathcal{P}] \tag{B.20}$$

となる. さらに, 定理の仮定から, $\mathcal{L}$ は $\mathcal{P}$ を含む λ システムである. 一方, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含む最小の λ システムから

$$\mathcal{L} \supset \mathcal{L}[\mathcal{P}] \tag{B.21}$$

となる. (B.20) と (B.21) を合わせると

$$\mathcal{L} \supset \sigma[\mathcal{P}]$$

がわかる. □

B.4 Dynkin の定理の応用例

B.4.1 その 1

補題 B.10. P_1 と P_2 を標本空間 $\big(\mathbb{X}, \mathcal{B}\big)$ 上の確率測度とする. 集合族

$$\mathcal{L} := \big\{B \in \mathcal{B};\, \mathsf{P}_1(B) = \mathsf{P}_2(B)\big\} \tag{B.22}$$

は λ システムである.

Proof. 命題 B.6 の条件 $(\mathrm{i})' \sim (\mathrm{iii})'$ を順に確認していく.

$(\mathrm{i})'$ $\mathsf{P}_1(\mathbb{X}) = \mathsf{P}_2(\mathbb{X})$ なので, $\mathbb{X} \in \mathcal{L}$ がわかる.

$(\mathrm{ii})'$ $A \in \mathcal{L}$ とする. すなわち

$$\mathsf{P}_1(A) = \mathsf{P}_2(A)$$

である. よって

$$\mathsf{P}_1(A^c) = 1 - \mathsf{P}_1(A) = 1 - \mathsf{P}_2(A) = \mathsf{P}_2(A^c)$$

であるので, $A^c \in \mathcal{L}$ がわかる.

$(\mathrm{iii})'$ $\{A_j\}_{j=1}^{\infty}$ は $\mathcal{L}$ の互いに排反な列とする. このとき, $\mathcal{L}$ の定義 (B.22) から

$$\mathsf{P}_1(A_j) = \mathsf{P}_2(A_j) \qquad (j = 1, 2, \ldots)$$

となる. よって

$$\mathsf{P}_1\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \sum_{j=1}^{\infty} \mathsf{P}_1(A_j) = \sum_{j=1}^{\infty} \mathsf{P}_2(A_j) = \mathsf{P}_2\Big(\bigcup_{j=1}^{\infty} A_j\Big)$$

となるので

$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$$

がわかる. □

補題 B.11. P_1 と P_2 は標本空間 $\big(\mathbb{X}, \mathcal{B}\big)$ 上の確率測度とする. $\mathcal{P}$ は π システムで, $\sigma[\mathcal{P}] = \mathcal{B}$ とする. このとき

$$\forall A \in \mathcal{P} \implies \mathsf{P}_1(A) = \mathsf{P}_2(A) \tag{B.23}$$

ならば

$$\forall B \in \mathcal{B} \implies \mathsf{P}_1(B) = \mathsf{P}_2(B) \tag{B.24}$$

が成り立つ.

Proof. 補題 B.10 から

$$\mathcal{L} := \big\{A \in \mathcal{B};\, \mathsf{P}_1(A) = \mathsf{P}_2(A)\big\}$$

は λ システムである. しかし, (B.23) から $\mathcal{L} \supset \mathcal{P}$ である. よって, 定理 B.9 と定理の仮定[3] から

$$\mathcal{L} \supset \sigma[\mathcal{P}] = \mathcal{B}$$

となる. よって, (B.23) ならば, (B.24) がわかる. □

命題 B.12. $\mathbb{X} = \mathbb{R}$ とし, P_1 と P_2 を $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ 上の確率測度とし, 対応する分布関数を F_1 と F_2 が等しいとする. すなわち

$$\mathsf{F}_1(x) = \mathsf{P}_1\big((-\infty, x]\big) = \mathsf{F}_2(x) = \mathsf{P}_2\big((-\infty, x]\big)$$

である. このとき

$$\mathsf{P}_1(B) = \mathsf{P}_2(B) \qquad \big(\forall B \in \mathcal{B}(\mathbb{R})\big)$$

が成り立つ.

[3] $\sigma[\mathcal{P}] = \mathcal{B}$ のこと.

Proof.

$$\mathcal{P} := \big\{(-\infty,\, x];\, x \in \mathbb{R}\big\}$$

とおく. $\forall x,\, y \in \mathbb{R}$ に対して

$$(-\infty,\, x] \bigcap (-\infty,\, y] = (-\infty,\, x \wedge y] \in \mathcal{P}$$

となる. ここで $x \wedge y$ は x と y の大きくない方である. 以上から $\mathcal{P}$ は π システムである. さらに

$$\mathsf{F}_1(x) = \mathsf{F}_2(x)\,(\forall x \in \mathbb{R}) \;\Leftrightarrow\; \mathsf{P}_1(A) = \mathsf{P}_2(A)\,(\forall A \in \mathcal{P}) \tag{B.25}$$

である. いま

$$\mathcal{L} = \{B \in \mathcal{B}(\mathbb{R});\, \mathsf{P}_1(B) = \mathsf{P}_2(B)\}$$

とする. すると, $\mathcal{L}$ の定義から $\mathcal{L} \subset \mathcal{B}(\mathbb{R})$ である. 一方, 補題 B.10 から, $\mathcal{L}$ は λ システムである. さらに, $\sigma[\mathcal{P}] = \mathcal{B}(\mathbb{R})$ であること [4] と (B.25) に注意をして, 補題 B.11 を用いると

$$\forall B \in \mathcal{B}(\mathbb{R}) \;\Rightarrow\; \mathsf{P}_1(B) = \mathsf{P}_2(B)$$

となる. したがって, $\mathcal{B}(\mathbb{R}) \subset \mathcal{L}$ となる. 一方, $\mathcal{L}$ の定義から

$$\mathcal{L} \supset \sigma[\mathcal{P}]$$

となる. よって, $\mathcal{L} = \sigma[\mathcal{P}] = \mathcal{B}(\mathbb{R})$ となることがわかる. 以上から

$$\mathsf{P}_1(A) = \mathsf{P}_2(A) \qquad \big(\forall A \in \mathcal{B}\big(\mathbb{R}\big)\big)$$

がわかる. □

B.4.2 その 2

$(\mathbb{X},\, \mathcal{B})$ と $(\mathbb{Y},\, \mathcal{C})$ を可測空間とし, μ を可測空間 $(\mathbb{X},\, \mathcal{B})$ 上の σ 有限測度とする. $\mathcal{B} \otimes \mathcal{C}$ 可測の非負値関数 $f : \mathbb{X} \times \mathbb{Y} \longrightarrow \mathbb{R}$ に対して, 写像

$$y \mapsto \int_{\mathbb{X}} f(x,\, y)\, \mathrm{d}\mu(x) \tag{B.26}$$

は $\mathcal{C}$ 可測であることを Dynkin の補題と標準機械を用いて示そう. 標準機械を用いるためには, 任意の $L \in \mathcal{B} \otimes \mathcal{C}$ に対して

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\, \mathrm{d}\mu(x)$$

が $\mathcal{C}$ 可測であることを示そう. そのために, 準備の補題から証明していく.

[4] この主張は定理 B.17 から直ちにわかる.

補題 B.13. $(\mathbb{X}, \mathcal{B})$ と $(\mathbb{Y}, \mathcal{C})$ を可測空間とし, μ を $(\mathbb{X}, \mathcal{B})$ 上の有限測度とする. このとき,

$$\mathcal{L}_\mu := \left\{ L \in \mathcal{B} \otimes \mathcal{C};\ y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x, y)\, \mathrm{d}\mu(x) \text{ は } \mathcal{C} \text{ 可測} \right\}$$

は λ システムである.

Proof. λ システムの定義の (i) $\sim$ (iii) を確認する.
① (i) の確認: $\mathbb{X} \times \mathbb{Y} \in \mathcal{B} \otimes \mathcal{C}$ であり,

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_{\mathbb{X}\times\mathbb{Y}}(x, y)\, \mathrm{d}\mu(x) = \mu(\mathbb{X})$$

は定数関数であるから $\mathcal{C}$ 可測である. したがって,

$$\mathbb{X} \times \mathbb{Y} \in \mathcal{L}_\mu$$

がわかる.
② (ii) の確認: $L_1, L_2 \in \mathcal{L}_\mu$ かつ $L_1 \subset L_2$ とする. このとき

$$\mathbb{1}_{L_2 \setminus L_1}(x, y) = \mathbb{1}_{L_2}(x, y) - \mathbb{1}_{L_1}(x, y)$$

であるから

$$\int_{\mathbb{X}} \mathbb{1}_{L_2\setminus L_1}(x, y)\, \mathrm{d}\mu(x) = \int_{\mathbb{X}} \mathbb{1}_{L_2}(x, y)\, \mathrm{d}\mu(x) - \int_{\mathbb{X}} \mathbb{1}_{L_1}(x, y)\, \mathrm{d}\mu(x)$$

となる. ここで $\mu(\mathbb{X}) < \infty$ であるから, 右辺の二つの積分はいずれも有限である. また, $L_1, L_2 \in \mathcal{L}_\mu$ なので, 右辺は二つの $\mathcal{C}$ 可測関数の差である. したがって,

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_{L_2\setminus L_1}(x, y)\, \mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測である. よって,

$$L_2 \setminus L_1 \in \mathcal{L}_\mu$$

が示せた.
③ (iii) の確認: $L_n \in \mathcal{L}_\mu$ $(n = 1, 2, \ldots)$ かつ

$$L_1 \subset L_2 \subset \cdots$$

とし

$$L := \bigcup_{n=1}^{\infty} L_n$$

とおく. このとき

$$\mathbb{1}_{L_n}(x,\, y) \nearrow \mathbb{1}_L(x,\, y) \qquad \big((x,\, y) \in \mathbb{X} \times \mathbb{Y}\big).$$

したがって, 単調収束定理により, 各 $y \in \mathbb{Y}$ に対して

$$\int_{\mathbb{X}} \mathbb{1}_{L_n}(x,\, y)\, \mathrm{d}\mu(x) \nearrow \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\, \mathrm{d}\mu(x)$$

となる.

各 n に対して左辺は y の $\mathcal{C}$ 可測関数であるから, その各点収束極限

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\, \mathrm{d}\mu(x)$$

も $\mathcal{C}$ 可測である. したがって

$$L \in \mathcal{L}_\mu$$

である. 以上から, $\mathcal{L}_\mu$ は λ システムである. □

命題 B.14. μ を $(\mathbb{X},\, \mathcal{B})$ 上の有限測度とする. このとき, 任意の $L \in \mathcal{B} \otimes \mathcal{C}$ に対して

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\, \mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測である.

Proof.

$$\mathcal{P} := \big\{B \times C;\, B \in \mathcal{B},\, C \in \mathcal{C}\big\}$$

とおく. $\mathcal{P}$ は π システムであり

$$\sigma[\mathcal{P}] = \mathcal{B} \otimes \mathcal{C}$$

である. また, $B \in \mathcal{B}$ および $C \in \mathcal{C}$ に対して

$$\int_{\mathbb{X}} \mathbb{1}_{B \times C}(x,\, y)\, \mathrm{d}\mu(x) = \mu(B) \mathbb{1}_C(y)$$

である. 右辺は y の $\mathcal{C}$ 可測関数であるから

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_{B \times C}(x,\, y)\, \mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測となる. よって

$$B \times C \in \mathcal{L}_\mu$$

である. したがって

$$\mathcal{P} \subset \mathcal{L}_\mu$$

がわかる. 補題 B.13 から $\mathcal{L}_\mu$ は λ システムである. よって, Dynkin の π-λ 定理 (定理 B.9) から

$$\mathcal{B} \otimes \mathcal{C} = \sigma[\mathcal{P}] \subset \mathcal{L}_\mu$$

である. 一方, $\mathcal{L}_\mu$ の定義から

$$\mathcal{L}_\mu \subset \mathcal{B} \otimes \mathcal{C}$$

であるので

$$\mathcal{L}_\mu = \mathcal{B} \otimes \mathcal{C}$$

となる. □

命題 B.15. μ を $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度とする. このとき, 任意の $L \in \mathcal{B} \otimes \mathcal{C}$ に対して

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x, y)\, \mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測である.

Proof. μ は σ 有限であるから, $X_n \in \mathcal{B}$ $(n = 1, 2, \ldots)$ で

$$X_1 \subset X_2 \subset \cdots, \qquad \bigcup_{n=1}^{\infty} X_n = \mathbb{X}, \qquad \mu(X_n) < \infty$$

を満たすものをとることができる. 各 n に対して

$$\mu_n(B) := \mu(B \cap X_n) \qquad \big(B \in \mathcal{B}\big)$$

とおく. このとき μ_n は有限測度である. したがって, 命題 B.14 から

$$\begin{aligned} g_n(y) &:= \int_{\mathbb{X}} \mathbb{1}_L(x, y)\, \mathrm{d}\mu_n(x) \\ &= \int_{\mathbb{X}} \mathbb{1}_L(x, y) \mathbb{1}_{X_n}(x)\, \mathrm{d}\mu(x) \end{aligned}$$

は $\mathcal{C}$ 可測である. 一方,

$$\mathbb{1}_L(x,\, y)\mathbb{1}_{X_n}(x) \nearrow \mathbb{1}_L(x,\, y)$$

であるから, 単調収束定理により

$$g_n(y) \nearrow \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\,\mathrm{d}\mu(x)$$

となる. 任意の $r \in \mathbb{R}$ に対して

$$\begin{aligned}&\left\{y \in \mathbb{Y};\, \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\,\mathrm{d}\mu(x) > r\right\}\\ &\qquad = \bigcup_{n=1}^{\infty}\left\{y \in \mathbb{Y};\, \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\,\mathrm{d}\mu_n(x) > r\right\} \in \mathcal{C}\end{aligned}$$

となる. したがって

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x,\, y)\,\mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測関数列の各点収束極限であるから, $\mathcal{C}$ 可測である. □

命題 B.16. μ を $(\mathbb{X}, \mathcal{C}$ 上の σ 有限な測度とする. このとき, $\mathbb{X} \times \mathbb{Y}$ 上の非負値可測関数 f に対して

$$y \mapsto \int_{\mathbb{X}} f(x,\, y)\,\mathrm{d}\mu(x)$$

が $\mathcal{C}$ 可測である.

Proof. 標準機械を用いて示そう. そのために, 以下の手続きを踏む.

① 指示関数 $\mathbb{1}_A$ に対してある性質を証明する.

② 有限個の互いに排反な $A_1, A_2, \ldots, A_n \in \mathcal{A} \in \mathcal{B} \otimes \mathcal{C}$ と異なる $a_1, a_2, \ldots, a_n$ に対して単関数 $\sum_{k=1}^{n} a_k \mathbb{1}_A$ に対して ① で示した性質を確認する.

③ 非負単関数の極限として非負の確率変数を定義し, ② で示した性質を非負値確率変数に対して証明する.

すなわち, 命題 B.15 から出発して, ① ∼ ③ の議論を実行することにより, 写像 $y \mapsto \int_{\mathbb{X}} f(x,\, y)\,\mathrm{d}\mu(x)$ は $\mathcal{C}$ 可測であることがわかる. □

B.4.3　$\mathbb{R}$ の位相についての注意

定理 B.17. 空でない開集合 $U \subset \mathbb{R}$ は高々可算個の互いに排反な開区間の和として表される. すなわち, 互いに排反な開区間 $\{I_j\}_{j=1}^{\infty}$ があって

$$U = \bigcup_{j=1}^{\infty} I_j$$

と書ける.

Proof. U は開集合なので, 各 $x \in U$ に対して $y > x$ なる $y \in \mathbb{R}$ が存在して,

$$(x,\, y) \subset U$$

とできる. このとき

$$a := \inf\{z;\, (z,\, x) \subset U\}, \quad b := \sup\{y;\, (x,\, y) \subset U\}$$

と定める. a と b は x に依存することに注意する. このとき

$$x \in (a,\, b) \quad \text{かつ} \quad I_x = (a,\, b) \text{ は開区間}$$

である. これらを踏まえて, 次の主張を示していく.

(i) $U = \bigcup\{I_x;\, x \in U\}$.

(ii) $\forall x,\, y \in U$ に対して

$$I_x \neq I_y \implies I_x \cap I_y = \varnothing$$

である. すなわち, 同一の連結区間に属する異なる $x,\, y$ に対して, $I_x = I_y$ となる.

(iii) $\#\{I_x;\, x \in U\} \leq \aleph_0$. ただし $I_x = I_y$ なるものを除いてである.

まず, 次のことを注意する. $x \in U$ に対して, $I_x = (a,\, b)$ とすると $I_x \subset U$ かつ $a,\, b \notin U$ である. なぜならば $w \in (a,\, b)$ とすると $y > w$ なる $y \in \mathbb{R}$ が存在して, $(x,\, y) \subset U$ となる. よって $w \in U$ である. いま $b \in U$ と仮定するとある $\epsilon > 0$ が存在して

$$(b - \epsilon,\, b + \epsilon) \subset U$$

とできる. ϵ を十分小さくすると

$$\left(x,\, b - \frac{\epsilon}{2}\right) \subset U$$

となる. 結局

$$(x,\, b+\epsilon) \subset U$$

となり, b の定義と矛盾する. よって $b \notin U$ が示せた.

(i) 各 $x \in U$ に対して, $x \in I_x \subset U$ なので

$$\bigcup_{x \in U} I_x \subset U$$

となる. 逆は明らかなので

$$\bigcup_{x \in U} I_x = U$$

がわかる.

(ii) $x,\, y \in U$ とし, $I_x = (a,\, b)$ と $I_y = (c,\, d)$ と定める. $I_x \cap I_y \neq \varnothing$ と仮定する. このとき $a < d$ かつ $c < b$ となる. $c \notin U$ なので $c \in (a,\, b)$ である. $c < b$ なので

$$c \leq a \tag{B.27}$$

である. $a \notin U$ なので

$$a \notin (c,\, d)$$

である. よって $a < d$ なので

$$a \leq c \tag{B.28}$$

がわかる. (B.27) と (B.28) から

$$a = c$$

がわかる. 同様に

$$b = d \tag{B.29}$$

を示すことができる. よって

$$I_x \cap I_y \neq \varnothing \implies I_x = I_y$$

がわかる. これの対偶を取ると

$$I_x \neq I_y \implies I_x \cap I_y = \varnothing$$

がわかる.

(iii) 各 $I_x\,(x \in U)$ に対して

$$q_x \in I_x \cap \mathbb{Q}$$

を取る. $I_x \neq I_y$ ならば $I_x \cap I_y = \varnothing$ なので

$$q_x \neq q_y$$

がわかる. よって $\#(\mathbb{Q}) \leq \aleph_0$ なので, (iii) がわかる. □

B.5 絶対連続関数

可測空間 $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ 上の測度 μ と Lebesgue 測度 m を考える. $\boldsymbol{x} \in \mathbb{R}^d$ および $r > 0$ に対し，中心 $\boldsymbol{x}$，半径 r の開球を

$$B(\boldsymbol{x};\, r) := \big\{\boldsymbol{v} \in \mathbb{R}^d;\, |\boldsymbol{v} - \boldsymbol{x}|_{2,\,d} < r\big\}$$

で定める. いま

$$G(\boldsymbol{x}) := \lim_{r \to 0} \frac{\mu\big(B(\boldsymbol{x};\, r)\big)}{\mathsf{m}\big(B(\boldsymbol{x};\, r)\big)} \qquad (\boldsymbol{x} \in \mathbb{R}^d)$$

とする. ただし, 右辺の極限が存在した場合に, $\boldsymbol{x}$ における G の値を定義する.

定義 B.18. (1) 可測関数 $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ は局所可積分(loc) であるとは, すべての有界 Borel 集合 $K \subset \mathbb{R}^d$ に対して

$$\int_K f(\boldsymbol{x})\, \mathrm{d}\mathsf{m}(\boldsymbol{x}) < \infty$$

が成り立つときをいう. $\mathbb{R}^d$ 上の局所可積分関数全体の集合を L^1_{loc} と記す. (2) $f \in L^1_{\mathrm{loc}}$ とする. $\boldsymbol{x} \in \mathbb{R}^d$ と $r > 1$ に対して

$$A_r f(\boldsymbol{x}) = \frac{1}{\mathsf{m}\big(B(\boldsymbol{x};\, r)\big)} \int_{B(\boldsymbol{x};\, r)} f(\boldsymbol{v})\, \mathrm{d}\mathsf{m}(\boldsymbol{v})$$

と定める.

補題 B.19. $f \in L^1_{\mathrm{loc}}$ とし, $r > 0$ と $\boldsymbol{x} \in \mathbb{R}^d$ とする. このとき, $A_r f(\boldsymbol{x})$ は r と $\boldsymbol{x}$ に関して連続である.

Proof.

証明は加筆予定 (2026/05/29).

□

定理 B.20. $f \in L^1_{\mathrm{loc}}$ とする. このとき, ほとんど至るところの $\boldsymbol{x} \in \mathbb{R}^d$ において

$$\lim_{r \to 0} A_r f(\boldsymbol{x}) = f(\boldsymbol{x})$$

が成立[5] する.

Proof.

[5] すなわち, $\mathsf{m}_d\big\{\boldsymbol{x} \in \mathbb{R}^d;\, \lim_{n \to \infty} A_r f(\boldsymbol{x}) \neq f(\boldsymbol{x})\big\} = 0$ のことである. ただし, m_d は $\mathbb{R}^d$ 上の Lebesgue 測度である.

証明は加筆予定 (2026/05/29).

□

定義 B.21. (1) $f \in L^1_{\mathrm{loc}}$ の Lebesgue 集合 L_f を

$$L_f := \left\{ \boldsymbol{x} \in \mathbb{R}^d;\ \lim_{r\to 0} \frac{1}{\mathsf{m}\big(B(\boldsymbol{x};\, r)\big)} \int_{B(\boldsymbol{x};\, r)} \big|f(\boldsymbol{v}) - f(\boldsymbol{x})\big|\, \mathrm{d}\mathsf{m}(\boldsymbol{v}) = 0 \right\}$$

で定義する.
(2) $r > 0$ で添え字づけらた $\mathbb{R}^d$ の Borel 集合族 $\{E_r\}_{r>0}$ は $\boldsymbol{x} \in \mathbb{R}^d$ にうまく縮む とは次の条件 (2a) と (2b) をみたすときをいう.

(2a) 各 r に対して, $E_r \subset B(\boldsymbol{x};\, r)$.

(2b) ある定数 $\alpha > 0$ (r とは無関係に取れる) が存在して, $\mathsf{m}(E_r) > \alpha \mathsf{m}\big(B(\boldsymbol{x};\, r)\big)$ となる.

定理 B.22. $f \in L^1_{\mathrm{loc}}$ と $\boldsymbol{x} \in L_f$ とする. 任意の $\boldsymbol{x} \in \mathbb{R}^d$ にうまく縮む Borel 部分集合族 $\{E_r\}_{r>0}$ に対して

$$\lim_{r\to 0} \frac{1}{\mathsf{m}(E_r)} \int_{E_r} \big|f(\boldsymbol{v}) - f(\boldsymbol{x})\big|\, \mathrm{d}\mathsf{m}(\boldsymbol{v}) = 0$$

となる.

Proof.

証明は加筆予定 (2026/05/29).

□

定義 B.23. (1) 関数 $F : \mathbb{R} \longrightarrow \mathbb{R}$ と $x \in \mathbb{R}$ に対して

$$T_F(x) := \sup\left\{ \sum_{k=1}^{n} \big|F(x_k) - F(x_{k-1})\big|;\ n \in \mathbb{N},\ -\infty < x_0 < \cdots < x_n = x \right\}$$

と定める. T_F を関数 F の 全変動 という.
(2) $\lim_{x\to\infty} T_F(x) < \infty$ のとき, F は 有界変動 という. 有界変動関数全体の集合を BV と記す.
(3) BV の部分集合 NBV を

$$NBV := \big\{ F \in BV;\ F \text{ は右連続で } F(-\infty) = \lim_{x\to-\infty} F(x) = 0 \big\}$$

で定める.

注意 B.24. (1) $F : \mathbb{R} \longrightarrow \mathbb{R}$ が有界かつ非減少ならば, $F \in BV$ である.
(2) $F, G \in BV$ のとき, $aF + bG \in BV$ である. ただし, $a, b \in \mathbb{R}$ である.
(3) F は $\mathbb{R}$ 上で微分可能で, $\dot{F}$ は有界のとき, $-\infty < a < b < \infty$ に対して, $F \in BV([a, b])$ である. □

補題 B.25. (1) $F \in BV$ とする. このとき, $T_F + F$ と $T_F - F$ は非減少関数である.
(2) $F : \mathbb{R} \longrightarrow \mathbb{R}$ とする. このとき

$$F \in BV \Leftrightarrow F \text{ は 2 つの非減少関数の差として表現できる.}$$

(3) $F \in BV$ のとき, F の右極限と左極限は存在する.
(4) $F \in BV$ とする. このとき, F の不連続点は高々可算個である.
(5) $F \in BV$ とし, $G(x) = F(x+)\,(x \in \mathbb{R})$ とする. このとき, $\dot{F}$ と G はほとんど至るところで存在し, $\dot{F}$=$\dot{G}$ (a..e.) である.
(6) $F \in BV$ とする. このとき, $T_F(-\infty) = 0$ となる. さらに, F が右連続ならば, T_F も右連続である.

Proof.

証明は加筆予定 (2026/05/29).

□

定理 B.26. NBV を定義 B.23(3) で定めた関数族とする. このとき, 以下のことが成立する.
(1) μ を $\mathbb{R}$ 上の Borel 測度とし, $F(x) = \mu\big((-\infty, x]\big)\,(x \in \mathbb{R})$ が有限であるとする. このとき, $F \in NBV$ となる.
(2) $F \in NBV$ のとき, $\mathbb{R}$ 上の Lebesgue–Stieltjes 測度 μ_F が一意的に存在して, $F(x) = \mu_F\big((-\infty, x]\big)\,(x \in \mathbb{R})$ となる.

Proof.

証明は加筆予定 (2026/05/29).

□

命題 B.27. NBV を定義 B.23(4) で定めた関数族とする. $F \in NBV$ に対し, μ_F を定理 B.26(2) で定めた $\mathbb{R}$ 上の Lebesgue–Stieltjes 測度とする. このとき, $\dot{F} \in L^1(\mathsf{m})$ である. ただし, $\dot{F}$ は F の右連続に修正された導

関数[6] である. さらに $\mu_F \ll \mathsf{m}$ [7] ならば,

$$F(x) = \int_{-\infty}^{x} \dot{F}(t)\,\mathrm{d}\mathsf{m}(t)$$

となる.

Proof.

証明は加筆予定 (2026/05/29).

□

定義 B.28. 関数 $F : \mathbb{R} \longrightarrow \mathbb{R}$ は閉区間 $[a, b]$ 上で絶対連続であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ と任意の互いに素な有限個の開区間 $\{(a_j, b_j)\}_{j=1}^{N}\,(N \in \mathbb{N})$ で, $a \leq a_j < b_j \leq b\,(j = 1, \ldots, N)$ をみたすものに対して

$$\sum_{k=1}^{N}(b_k - a_k) < \delta \Rightarrow \sum_{k=1}^{N}\left|F(b_k) - F(a_k)\right| < \epsilon$$

となるときをいう.

命題 B.29. $F \in NBV$ とする. このとき

$$F \text{ は絶対連続} \Leftrightarrow \mu_F \ll \mathsf{m}$$

である.

Proof.

証明は加筆予定 (2026/05/29).

□

系 B.30. (1) $f \in L^1(\mathsf{m})$ とし, 関数 $F : \mathbb{R} \longrightarrow \mathbb{R}$ を

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}\mathsf{m}(t)$$

で定める. このとき, $F \in NBV$ で, F は絶対連続である.
(2) $F \in NBV$ で, さらに F は絶対連続ならば, $\dot{F} \in L^1(\mathsf{m})$ となる. さらに, $F(x) = \int_{-\infty}^{x} \dot{F}(t)\,\mathrm{d}t$ となる.

---

[6] 補題 B.25(5) のようにして, 右連続に修正したもの.
[7] 任意の Borel 部分集合 $A \subset \mathbb{R}$ に対して

$$\mathsf{m}(A) = 0 \Rightarrow \mu_F(A) = 0$$

が成立することである.

Proof.

証明は加筆予定 (2026/05/29).

□

B.6 凸関数の性質

凸関数はおおくとも 2 つの単調増加関数の差として表現できるので, 補題 B.25(2) から BV となる. よって, 前節の結果と微分可能であれば, 導関数は非負であることからこの節の主定理の定理 B.36 の主張は用意に想像できる. この定理から凸関数の基本的な性質を証明することができる. しかし, 煩雑ではあるが, 初等的な議論により凸関数の基本的な性質を証明する.

定義 B.31. $a, b \in \mathbb{R}$, $a < b$ とする. 関数 $f : [a, b] \to \mathbb{R}$ が $x, y \in [a, b]$, $0 \leq \lambda \leq 1$ に対して

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y)$$

をみたすとき, 関数 f は凸関数 であるという. とくに $x \neq y$, $0 < \lambda < 1$ のとき

$$f\big(\lambda x + (1-\lambda)y\big) < \lambda f(x) + (1-\lambda) f(y)$$

が成立する場合は, 関数 f は 狭義凸関数 であるという.

注意 B.32. $-f$ が凸関数 (狭義凸関数) のとき, 関数 f は凹関数(狭義凹関数) であるという. □

定理 B.33. 区間 $[a, b]$ 上の凸関数は開区間 (a, b) 上の各点で連続である.

Proof. まず, $\forall x \in [a, b]$ は

$$x = \lambda a + (1-\lambda) b \quad (0 \leq \lambda \leq 1)$$

と表せるから, 凸関数の定義より

$$f(x) \leq \lambda f(a) + (1-\lambda) f(b) \leq \max\{f(a),\, f(b)\} =: K \tag{B.30}$$

となること[8] に注意する. よって, 凸関数 f は上に有界である. $\forall x_0 \in (a, b)$ とし, $f(x)$ は $x = x_0$ で連続であることを示すために関数 g を

$$g(t) := f(x_0 + t) - f(x_0) \qquad (a - x_0 < t < b - x_0)$$

[8] K は負であってもよい.

と定める[9]. このとき $g(t)$ は $(a-x_0, b-x_0) \ni \{0\}$ 上で凸関数であり, $g(0)=0$ であることに注意する. いま $\forall \epsilon > 0$ に対して正数 ϵ' を

$$0 < \epsilon' < \min\left\{ \frac{\epsilon}{K + |f(x_0)| + 1}, 1 \right\} \tag{B.31}$$

をみたすように選ぶ. さらに正数 ρ と δ を

$$0 < \rho < \min\{x_0 - a, b - x_0\}, \quad 0 < \delta < \rho\epsilon' \tag{B.32}$$

をみたすように定める. (B.31) より $0 < \epsilon' \leq 1$ である. したがって

$$|t| < \delta \Rightarrow |t| < \rho$$

となる. したがって $|t| < \delta$ に対して, $g(t)$ は定義されている. しかも

$$|t| < \delta \Rightarrow \left| \frac{t}{\epsilon'} \right| < \rho \Rightarrow x_0 + \frac{t}{\epsilon'} \in (a, b)$$

となること[10]がわかる. よって, (B.30) から

$$\begin{aligned} g\left(\pm\frac{t}{\epsilon'}\right) = f\left(x_0 \pm \frac{t}{\epsilon'}\right) - f(x_0) &\leq \left| f\left(x_0 \pm \frac{t}{\epsilon'}\right) \right| + |f(x_0)| \\ &\leq K + |f(x_0)| =: K_1 \end{aligned} \tag{B.33}$$

を得る. ((B.31) から

$$K_1\epsilon' \leq \epsilon \tag{B.34}$$

となっていることに注意せよ. ところで

$$t = (1-\epsilon') \times 0 + \epsilon'\left(\frac{t}{\epsilon'}\right)$$

と表されることと (B.33) と (B.34) から

$$g(t) \leq (1-\epsilon') \underbrace{g(0)}_{=0} + \epsilon' g\left(\frac{t}{\epsilon'}\right) = \epsilon' g\left(\frac{t}{\epsilon'}\right) \leq \epsilon' K_1 < \epsilon \tag{B.35}$$

となることに注意をする. 他方

$$0 = \frac{\epsilon'}{1+\epsilon'}\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1+\epsilon'} \cdot t$$

[9] $x_0 \in (a, b)$ だったので, $x_0 - a > 0$ かつ $b - x_0 > 0$ である. よって, $\min\{x_0 - a, b - x_0\} > 0$ となることに注意せよ.

[10] $\left|\frac{t}{\epsilon'}\right| < \rho$ と (B.32) から $-x_0 + a < -\rho < \frac{t}{\epsilon'} < \rho < b - x_0$ となるので $a < \frac{t}{\epsilon'} + x_0 < b$ がわかる.

と表されるので, (B.33) と (B.34) から

$$0 \leq g(0) \leq \frac{\epsilon'}{1+\epsilon'} g\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1+\epsilon'} g(t) \leq \frac{\epsilon' K_1}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t)$$
$$< \frac{\epsilon}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t)$$

となり

$$-\epsilon < g(t) \tag{B.36}$$

を得る. よって (B.35) と (B.36) より

$$|t| < \delta \Rightarrow |g(t)| < \epsilon$$

である. $t = x - x_0$ と $g(t)$ の定義より

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon$$

となり, $f(x)$ は点 x_0 で連続となる. □

定理 B.34. 関数 f が区間 $[a, b]$ 上で凸のとき, 開区間 (a, b) の各点 x で右微分係数 $\overset{\bullet}{f}_+(x)$ が存在し, 関数

$$\overset{\bullet}{f}_+ \colon (a, b) \ni x \mapsto \overset{\bullet}{f}_+(x) \in \mathbb{R}$$

は単調増加である.

同様に左微分係数 $\overset{\bullet}{f}_-(x)$ が存在し, 関数

$$\overset{\bullet}{f}_- \colon (a, b) \ni x \mapsto \overset{\bullet}{f}_-(x) \in \mathbb{R}$$

も単調増加である. さらに各点 $x \in (a, b)$ において

$$\overset{\bullet}{f}_-(x) \leq \overset{\bullet}{f}_+(x)$$

が成立する.

Proof. 任意の $x \in (a, b)$ に対して実数 η と ξ を

$$0 < \eta < \min\{x - a, b - x\}, \quad 0 < \xi < \eta$$

をみたすように選ぶ. このとき

$$0 < \frac{\xi}{\eta} < 1, \quad x + \xi = \frac{\xi}{\eta}(x + \eta) + \left(1 - \frac{\xi}{\eta}\right) x$$

と書けることに注意する. 関数 f は凸であることから

$$f(x+\xi) \leq \frac{\xi}{\eta} f(x+\eta) + \left(1 - \frac{\xi}{\eta}\right) f(x) = \frac{\xi}{\eta}\big\{f(x+\eta) - f(x)\big\} + f(x)$$

となる. つまり $0 < \xi < \eta$ のとき

$$\frac{f(x+\xi) - f(x)}{\xi} \leq \frac{f(x+\eta) - f(x)}{\eta}$$

となり, $\xi \searrow 0$ とすれば

$$\overset{\bullet}{f}_{+}(x) \leq \frac{f(x+\eta) - f(x)}{\eta} \tag{B.37}$$

がわかる. 同様に

$$x - \xi = \frac{\xi}{\eta}(x - \eta) + \left(1 - \frac{\xi}{\eta}\right) x$$

と表せるから

$$f(x-\xi) \leq \frac{\xi}{\eta} f(x-\eta) + \left(1 - \frac{\xi}{\eta}\right) f(x) = \frac{\xi}{\eta}\big\{f(x-\eta) - f(x)\big\} + f(x)$$

である. したがって $0 < \xi < \eta$ のとき

$$\frac{f(x) - f(x-\eta)}{\eta} \leq \frac{f(x) - f(x-\xi)}{\xi}$$

となり, $\xi \searrow 0$ とすれば

$$\overset{\bullet}{f}_{-}(x) \geq \frac{f(x) - f(x-\eta)}{\eta} \tag{B.38}$$

を得る. さらに

$$f(x) = f\left(\frac{(x+\xi) + (x-\xi)}{2}\right) \leq \frac{f(x+\xi) + f(x-\xi)}{2}$$

となり

$$f(x) - f(x-\xi) \leq f(x+\xi) - f(x)$$

である. だから

$$\frac{f(x) - f(x-\xi)}{\xi} \leq \frac{f(x+\xi) - f(x)}{\xi}$$

がわかる. ここで, $\xi \searrow 0$ とすると

$$\overset{\bullet}{f}_{-}(x) \leq \overset{\bullet}{f}_{+}(x) \tag{B.39}$$

がわかる. したがって, (B.37) – (B.39) を考慮すれば

$$\frac{f(x)-f(x-\eta)}{\eta} \leq \overset{\bullet}{f}_{-}(x) \leq \overset{\bullet}{f}_{+}(x) \leq \frac{f(x+\eta)-f(x)}{\eta} \tag{B.40}$$

がわかる. よって $\overset{\bullet}{f}_{-}(x)$ と $\overset{\bullet}{f}_{+}(x)$ は有界であることがわかる.

次に, 単調性を示す. そのために $a < x_1 < x_2 < b$ とする. (B.40) の

$$f_{-}(x) \leq f_{+}(x) \leq \frac{f(x+\eta)-f(x)}{\eta}$$

において, $x = x_1$, $x + \eta = x_2$ とおくと

$$f_{-}(x_1) \leq f_{+}(x_1) \leq \frac{f(x_2)-f(x_1)}{x_2-x_1}$$

を得る. 同様に (B.40) の

$$\frac{f(x)-f(x-\eta)}{\eta} \leq f_{-}(x) \leq f_{+}(x)$$

において, $x = x_2$, $x - \eta = x_1$ とおくと

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq f_{-}(x_2) \leq f_{+}(x_2)$$

を得る. これらの式を合わせると

$$\overset{\bullet}{f}_{-}(x_1) \leq \overset{\bullet}{f}_{+}(x_1) \leq \frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \overset{\bullet}{f}_{-}(x_2) \leq \overset{\bullet}{f}_{+}(x_2) \tag{B.41}$$

となる. よって

$$\overset{\bullet}{f}_{-}(x_1) \leq \overset{\bullet}{f}_{-}(x_2), \quad \overset{\bullet}{f}_{+}(x_1) \leq \overset{\bullet}{f}_{+}(x_2)$$

が得られる. □

系 B.35. 関数 F は (a, b) 上で凸のとき, 任意の s, s', t, t' で $s \leq s' < t'$ かつ $s < t \leq t'$ なるものに対して

$$\frac{F(t)-F(s)}{t-s} \leq \frac{F(t')-F(s')}{t'-s'}$$

となる.

Proof. 定理 B.34 から直ちにわかる. □

定理 B.36. 関数 F は (a, b) 上で凸のとき, 関数 F は (a, b) のすべてのコンパクト部分区間において絶対連続で, $\overset{\bullet}{F}$ は非減少 (存在する場合) である.

Proof. $[c,\, d] \subset (a,\, b)$ をコンパクト区間[11] とする. 定理 B.34 から, ある $M > 0$ が存在して, 任意の $x,\, y \in [c,\, d]$ かつ $x < y$ に対して

$$-M \leq \frac{F(y) - F(x)}{y - x} \leq M$$

となる. したがって

$$\big|F(y) - F(x)\big| \leq M|y - x|$$

となる. よって, F は $[c,\, d]$ 上で Lipschitz 連続である. したがって, F は絶対連続である. よって, ほとんど至るところで $\dot{F}$ が存在して

$$F(y) - F(x) = \int_x^y \dot{F}\,(t)\,\mathrm{d}\mathsf{m}(t)$$

と書ける. さらに, 定理 B.34 から $\dot{F}$ は非減少であることがわかる. □

定理 B.37. 関数 F は $(a,\, b)$ 上で凸とする. このとき, $t_0 \in (a,\, b)$ に対して, ある $\beta \in \mathbb{R}$ が存在して, $F(t) - F(t_0) \geq \beta(t - t_0)\,\big(t \in (a,\, b)\big)$ となる.

Proof. 定理 B.34 から, $t_0 \in (a,\, b)$ において, 右微分 $\dot{F}_-\,(t_0)$ と左微分 $\dot{F}_+\,(t_0)$ が存在して, $\dot{F}_-\,(t_0) \leq \dot{F}_+\,(t_0)$ となる. ここで, $\beta \in [\dot{F}_-\,(t_0),\, \dot{F}_+\,(t_0)]$ をとる. すると

$$\frac{F(t) - F(t_0)}{t - t_0} \leq \beta \qquad (t \in (a,\, b),\, t \neq t_0)$$

となる. よって

$$F(t) \geq F(t_0) + \beta(t - t_0) \qquad (t \in (a,\, b))$$

がわかる. □

定理 B.38 (Jensen の不等式)**.** $\mathbb{X} \subset \mathbb{R}$ とする. 測度空間 $\big(\mathbb{R},\, \mathcal{B}(\mathbb{X})\big)$ 上の測度 μ は $\mu(\mathbb{X}) = 1$ とする. $g : \mathbb{X} \longrightarrow (a,\, b)$ は $\int_{\mathbb{X}} |g(x)|\,\mathrm{d}\mu(x) < \infty$ とし, 関数 F は $(a,\, b)$ 上で凸とする. このとき

$$F\left(\int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x)\right) \leq \int_{\mathbb{X}} F \circ g(x)\,\mathrm{d}\mu(x)$$

となる.

[11] すなわち, 有界区間である.

Proof. $m = \int_{\mathbb{X}} g(x)\,\mathrm{d}\mu(x)$ とおく. すると m において, ある $\beta \in \mathbb{R}$ が存在して

$$F(t) \geq F(m) + \beta(t-m) \qquad \big(t \in (a,\, b)\big)$$

となる. $t = g(x)$ として, 積分をすると

$$\int_{\mathbb{X}} F \circ g(x)\,\mathrm{d}\mu(x) \geq \int_{\mathbb{X}} \big\{F(m) + \beta(g(x)-m)\big\}\,\mathrm{d}\mu(x) = F(m)$$

となる. よって, 定理は証明された. □

B.7 Radon-Nikodym の定理と条件付き期待値の議論

$(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ を確率空間とし, $\mathcal{F} \subset \mathcal{A}$ も σ 加法族とする. X は $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数で $\mathsf{E}\big[|X|\big] < \infty$ とする.

定義 B.39. $\mathcal{F}$ を与えたときの X の**条件付き期待値**を次の条件をみたす任意の確率変数 Y とする.

(i) Y は $\mathcal{F}$ 可測.

(ii) 任意の $A \in \mathcal{F}$ に対して

$$\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega).$$

まず条件付き期待値の存在と一意性を確認する.

補題 B.40. Y を定義 B.39 の条件 (i)(ii) をみたすとき Y は可積である.

Proof. $A := \{\omega \in \Omega;\, Y(\omega) > 0\}$ とおくと, $A,\, A^c \in \mathcal{F}$ となる. (ii) を用いると

$$\begin{aligned}
\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \leq \int_A |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega), \\
\int_{A^c} \big(-Y(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_{A^c} \big(-X(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega) \leq \int_{A^c} |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega).
\end{aligned}$$

よって

$$\begin{aligned}
\int |Y(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} \big(-Y(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega) \\
&\leq \int_A |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) \\
&= \int |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}$$

からわかる. □

補題 B.41. Y は Pr a.s. に一意である.

Proof. Y' も定義 B.39 の条件 (i)(ii) をみたすとする. すると (ii) より

$$\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad (\forall A \in \mathcal{F}).$$

$\epsilon > 0$ を取り, $A := \{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \epsilon\}$ とする. $A \in \mathcal{F}$ であることから (ii) から

$$\begin{aligned} 0 &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A \big\{Y(\omega) - Y'(\omega)\big\}\,\mathrm{d}\mathsf{Pr}(\omega) \\ &\geq \epsilon \int_A \mathrm{d}\mathsf{Pr}(\omega) = \epsilon \mathsf{Pr}(A) \end{aligned}$$

となり, $\mathsf{Pr}(A) = 0$ がわかる.

$$\left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\right\} = \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\right\}$$

と表現できることとすべての $\epsilon > 0$ に対して $\mathsf{Pr}(A) = 0$ なので

$$\begin{aligned} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\right\}\right) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\right\}\right) \\ &= \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \frac{1}{n}\right\}\right) \\ &= 0 \end{aligned}$$

がわかる. よって

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) \leq Y'(\omega)\big) = 1$$

となる. 次に $A := \{\omega \in \Omega;\, Y'(\omega) - Y(\omega) \geq \epsilon\}$ として, 同じ議論を行えば

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) \geq Y'(\omega)\big) = 1.$$

よって

$$\mathsf{Pr}\big(\omega \in \Omega;\, Y(\omega) = Y'(\omega)\big) = 1$$

となる. □

定義 B.42. 定義 B.39 で定めた Y を

$$Y(\omega) = \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$$

と記すことにする.

補題 B.43. $\mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ は存在する.

Proof. $\mu = \mathsf{Pr}$ と書くことにする. まず $X \geq 0$ とする. $\mathcal{F}$ 上の測度 ν を

$$\nu(A) = \int_A X(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F})$$

と定める. 単調収束定理を用いると ν は $\mathcal{F}$ 上の測度となる. さらに $\nu \ll \mu$ である. よって Radon-Nikodym の定理より, $\mathcal{F}$ 可測関数 $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ が存在し

$$\begin{aligned}\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mu(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \nu(A)\\ &= \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F}) \qquad \text{(B.42)}\end{aligned}$$

となる. $A = \Omega$ とすると $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \geq 0$ は可積となる. Radon-Nikodym の一意性から

$$\mathsf{E}\big[X|\,\mathcal{F}\big](\omega) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(\omega)$$

となる.

一般の場合については

$$\begin{aligned}&X(\omega) = X^+(\omega) - X^-(\omega),\\ &X^+(\omega) = \max\{X(\omega),\,0\}, \quad X^-(\omega) = \max\{-X(\omega),\,0\}\end{aligned}$$

とする. $Y_1(\omega) = \mathsf{E}\big[X^+|\,\mathcal{F}\big](\omega)$ と $Y_2(\omega) = \mathsf{E}\big[X^-|\,\mathcal{F}\big](\omega)$ とおくと $Y_1 - Y_2$ は $\mathcal{F}$ 可測である. さらに $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X^+(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X^-(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A Y_1(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y_2(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{(B.42)}\big)\\ &= \int_A \big\{Y_1(\omega) - Y_2(\omega)\big\}\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \big(Y_1 - Y_2\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

となる. よって $\big(Y_1 - Y_2\big)(\omega) = \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ がわかる. □

注意 B.44. 以下の記法を導入する.
(1) $A \in \mathcal{A}$ に対して

$$\mathsf{Pr}\big(A|\,\mathcal{F}\big)(\omega) := \mathsf{E}\big[\mathbb{1}_A|\,\mathcal{F}\big](\omega)$$

と定める.
(2) A, $B \in \mathcal{A}$ かつ $\mathsf{Pr}(B) > 0$ に対して

$$\mathsf{Pr}\big(A|\,B\big) := \frac{\mathsf{Pr}\big(A \cap B\big)}{\mathsf{Pr}\big(B\big)}$$

と定める.
(3) X と Y を $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数としたとき

$$\mathsf{E}\big[X|\,Y\big](\omega) := \mathsf{E}\big[X|\,\sigma(Y)\big](\omega)$$

とする. ただし $\sigma(Y)$ は Y によって生成された σ 加法族である. すなわち

$$\sigma(Y) := \sigma\left[\big\{\omega \in \Omega;\, Y(\omega) \leq r\,(\forall r \in \mathbb{R})\big\}\right]$$

である.
(4) X と Y を $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とする. さらに (X, Y) は同時 p.d.f. $\mathsf{p}(x, y)$ を持つとする. 議論を簡単にするために $\forall y \in \mathbb{R}$ に対して

$$\int \mathsf{p}(x, y)\,\mathrm{d}x > 0$$

とする.

いま関数 $g : \mathbb{R} \to \mathbb{R}$ で $\mathsf{E}\big[|g(X)|\big] < \infty$ なるものと考える. このとき

$$\mathsf{E}\big[g(X)|\,Y\big](\omega) = h(Y(\omega)), \qquad h(y) = \frac{\int g(x)\mathsf{p}(x, y)\,\mathrm{d}x}{\int \mathsf{p}(x, y)\,\mathrm{d}x}$$

となる. これを示すために $A \in \sigma(Y)$ を取る. このときある $B \in \mathcal{B}(\mathbb{R})$ が存在して

$$A := \big\{\omega \in \Omega;\, Y(\omega) \in B\big\}$$

と書ける. したがって Fubini の定理から

$$\begin{aligned}\mathsf{E}\big[h(Y)\mathbb{1}_A\big] &= \mathsf{E}\big[h(Y)\mathbb{1}_B(Y)\big] \qquad \big(\because Y \in B \Leftrightarrow \omega \in A\big)\\ &= \int_B \int h(y)p(x, y)\,\mathrm{d}x\,\mathrm{d}y\end{aligned}$$

$$
\begin{aligned}
&= \int_B \int \left\{ \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} \right\} p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \qquad \bigl(\because h \text{ の定義を代入}\bigr) \\
&= \int_B \left\{ \int \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} p(x,\,y)\,\mathrm{d}x \right\} \mathrm{d}y \qquad \bigl(\because \text{Fubini の定理}\bigr) \\
&= \int_B \frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x} \left\{ \int p(x,\,y)\,\mathrm{d}x \right\} \mathrm{d}y \\
&= \int_B \int g(x)p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \mathsf{E}\bigl[g(X)\mathbb{1}_A\bigr] \qquad \bigl(\because Y \in B \Leftrightarrow \omega \in A\bigr)
\end{aligned}
$$

となる. よって

$$\int_A h\bigl(Y(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A g\bigl(X(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega)$$

なので

$$\mathsf{E}\bigl[g(X)|\,\mathcal{F}\bigr](\omega) = h\bigl(Y(\omega)\bigr) \qquad \bigl(\mathsf{Pr}\text{ a..s}\bigr)$$

がわかる. □

B.8　条件付き期待値の性質

定理 B.45. X, Y, $\{X_n\}_{n=1}^{\infty}$ を確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数列とする.
(1) $\mathsf{E}\bigl[|X|\bigr] < \infty$ かつ $\mathsf{E}\bigl[|Y|\bigr] < \infty$ とする. このとき $a,\,b \in \mathbb{R}$ に対して

$$\mathsf{E}\bigl[aX + bY|\,\mathcal{F}\bigr](\omega) = a\mathsf{E}\bigl[X|\,\mathcal{F}\bigr](\omega) + b\mathsf{E}\bigl[Y|\,\mathcal{F}\bigr](\omega) \qquad \bigl(\mathsf{Pr}\text{ a.s.}\bigr)$$

と成り立つ.
(2) さらに $X \leq Y$ のとき

$$\mathsf{E}\bigl[X|\,\mathcal{F}\bigr](\omega) \leq \mathsf{E}\bigl[Y|\,\mathcal{F}\bigr](\omega) \qquad \bigl(\mathsf{Pr}\text{ a.s.}\bigr)$$

が成り立つ.
(3) $X_n \geq 0\,(n = 1,\,2,\,\ldots)$ かつ $X_n \uparrow X\,(n \to \infty)$ で $\mathsf{E}\bigl[X\bigr] < \infty$ のとき

$$\mathsf{E}\bigl[X_n|\,\mathcal{F}\bigr](\omega) \uparrow \mathsf{E}\bigl[X|\,\mathcal{F}\bigr](\omega) \quad (n \to \infty) \qquad \bigl(\mathsf{Pr}\text{ a.s.}\bigr)$$

が成り立つ.

Proof. (1) あきらかに $a\mathsf{E}\bigl[X|\,\mathcal{F}\bigr] + b\mathsf{E}\bigl[Y|\,\mathcal{F}\bigr]$ は $\mathcal{F}$ 可測である. $\forall A \in \mathcal{F}$ に対して

$$\int_A \bigl\{a\mathsf{E}\bigl[X|\,\mathcal{F}\bigr](\omega) + b\mathsf{E}\bigl[Y|\,\mathcal{F}\bigr](\omega)\bigr\}\,\mathrm{d}\mathsf{Pr}(\omega)$$

$$
\begin{aligned}
&= a\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&= a\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&= \int_A \big(aX+bY\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}
$$

を得る.

(2) 条件付き期待値の定義より $\forall A \in \mathcal{F}$ に対して

$$
\begin{aligned}
\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \le \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\
&= \int_A \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}
$$

を得る. $\forall \epsilon > 0$ に対して

$$
A := \big\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \epsilon\big\}
$$

とおくと $\mathsf{Pr}(A) = 0$ となる. ϵ は任意だったので

$$
\begin{aligned}
&\left\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\right\}\\
&\qquad = \bigcup_{n=1}^{\infty}\left\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \frac{1}{n}\right\}
\end{aligned}
$$

となる. よって

$$
\begin{aligned}
&\mathsf{Pr}\left(\left\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\right\}\right)\\
&\quad = \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \ge \frac{1}{n}\right\}\right) = 0
\end{aligned}
$$

となるから

$$
\mathsf{Pr}\left(\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}\big](\omega) \le \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\right) = 1
$$

となる.

(3) $Y_n := X - X_n\,(n \in \mathbb{N})$ とおくと $\mathsf{E}\big[Y_n|\,\mathcal{F}\big] \downarrow 0\,(n \to \infty)$ を示せばよい. $Y_n \downarrow 0\,(n \to \infty)$ なので (2) の結果から

$$
Z_n(\omega) := \mathsf{E}\big[Y_n|\,\mathcal{F}\big](\omega)
$$

はほとんど確実に単調減少列なので

$$
Z_\infty(\omega) := \lim_{n\to\infty} \mathsf{E}\big[Y_n|\,\mathcal{F}\big](\omega)
$$

はほとんど確実に存在する. $A \in \mathcal{F}$ に対して

$$\int_A Z_n(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y_n(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)$$

である. $n \to \infty$ のとき $Y_n(\omega) \downarrow 0\,(n \to \infty)$ なので, 単調収束定理から

$$\int_A Z_\infty(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = 0 \qquad (\forall A \in \mathcal{F})$$

となる. よって, $A \in \mathcal{F}$ かつ $Z_\infty \geq 0$ なので, $Z_\infty \equiv 0\,\big(\mathsf{Pr}\text{ a.s.}\big)$ となる. □

定理 B.46. $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}$ とする. このとき

(1) $\mathsf{E}\big[\mathsf{E}\big[X\big|\,\mathcal{F}_1\big]\big|\,\mathcal{F}_2\big](\omega) = \mathsf{E}\big[X\big|\,\mathcal{F}_1\big](\omega)\,\big(\mathsf{Pr}\text{ a.s.}\big)$ となる.

(2) $\mathsf{E}\big[\mathsf{E}\big[X\big|\,\mathcal{F}_2\big]\big|\,\mathcal{F}_1\big](\omega) = \mathsf{E}\big[X\big|\,\mathcal{F}_1\big](\omega)\,\big(\mathsf{Pr}\text{ a.s.}\big)$ となる.

Proof. (1) $\mathcal{F}_1 \subset \mathcal{F}_2$ なので, $\mathsf{E}\big[X\big|\,\mathcal{F}_1\big]$ は $\mathcal{F}_2$ 可測でもある. よって $A \in \mathcal{F}_2$ に対して

$$\int_A \mathsf{E}\big[X\big|\,\mathcal{F}_1\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A \mathsf{E}\Big[\mathsf{E}\big[X\big|\,\mathcal{F}_1\big]\Big|\,\mathcal{F}_2\Big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)$$

がわかる.
(2) $\mathsf{E}\big[X\big|\,\mathcal{F}_1\big]$ は $\mathcal{F}_1$ 可測である. $A \in \mathcal{F}_1 \subset \mathcal{F}_2$ に対して

$$\begin{aligned}
\int_A \mathsf{E}\big[X\big|\,\mathcal{F}_1\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_1\big) \\
&= \int_A \mathsf{E}\big[X\big|\,\mathcal{F}_2\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_2\big) \\
&= \int_A \Big\{\mathsf{E}\Big[\mathsf{E}\big[X\big|\,\mathcal{F}_2\big]\Big|\,\mathcal{F}_1\Big](\omega)\Big\}\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because A \in \mathcal{F}_1\big)
\end{aligned}$$

より

$$\mathsf{E}\big[X\big|\,\mathcal{F}_1\big](\omega) = \mathsf{E}\Big[\mathsf{E}\big[X\big|\,\mathcal{F}_2\big]\Big|\,\mathcal{F}_1\Big](\omega)$$

がわかる.

定理 B.47. X は $\mathcal{F}$ 可測とし, $\mathsf{E}\big[|X|\big] < \infty$ かつ $\mathsf{E}\big[|XY|\big] < \infty$ とする. このとき

$$\mathsf{E}\big[XY\big|\,\mathcal{F}\big](\omega) = X(\omega)\mathsf{E}\big[Y\big|\,\mathcal{F}\big](\omega) \tag{B.43}$$

が成立する.

Proof. $X(\omega) = \mathbb{1}_B(\omega)\,(\forall B \in \mathcal{F})$ のとき (B.43) が成り立つことを示す. $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}\int_A \mathbb{1}_B(\omega)\mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_{A\cap B} \mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_{A\cap B} Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \mathbb{1}_B(\omega)Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

となる. よって $X(\omega) = \mathbb{1}_B(\omega)$ のとき (B.43) は 成立する.

次に $X, Y \geq 0$ とし, X_n は非負値の階段関数とし, $X_n \uparrow X\,(n \to \infty)$ とする. 条件付き単調収束定理 (定理 **??**(3)) を用いて

$$\begin{aligned}\int_A X(\omega)\mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \lim_{n\to\infty}\int_A X_n(\omega)\mathsf{E}\big[Y\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \lim_{n\to\infty}\int_A \mathsf{E}\big[X_nY\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because (1)\text{ の結果}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[X_nY\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{単調収束定理}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[XY\,|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{定理 B.45(3)}\big)\end{aligned}$$

がわかり, $X, Y \geq 0$ のとき (B.43) は成立する.

最後に一般の X, Y に対して

$$\begin{aligned}X^+ &= \max\{X, 0\}, \quad X^- = \max\{-X, 0\},\\ Y^+ &= \max\{Y, 0\}, \quad Y^- = \max\{-Y, 0\}\end{aligned}$$

として, 上の結果を用いればよい. □

B.9 Neyman-Fisher の因子分解定理の証明

Radon-Nikodym の定理から導かれることを用いて Fisher-Neyman の因子分解定理の測度論的な証明を与える.

こ節 B.9 の校正ははもう一度やり直すこと. (2026/09/15).

補題 B.48. μ を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の σ 有限測度とし, $\mathcal{N}$ を $(\mathbb{X}, \mathcal{A})$ 上の測度の集合で, $\forall \nu \in \mathcal{N}$ に対して, $\nu \ll \mu$ をみたしているとする. このとき非負の数列 $\{c_n\}_{n=1}^\infty$ で $c_n \geq 0\,(n \in \mathbb{N})$ かつ $\sum_{n=1}^\infty c_n = 1$ なるものと $\{\nu_n\}_{n=1}^\infty \subset \mathcal{N}$ が存在して $\forall \nu \in \mathcal{N}$ に対して $\nu \ll \sum_{n=1}^\infty c_n\nu_n$ となる.

Proof. $\mathcal{N}$ が可算集合ならば定理の主張は自明であるので, $\mathcal{N}$ は非可算とする.

μ は有限測度の場合は

$$\lambda = \mu$$

とおく. μ が有限測度でない場合には $\mathbb{X}$ の可算個の部分集合 $\{\mathbb{X}_n\}_{n=1}^{\infty}$ をうまくとると

$$\mathbb{X} = \bigcup_{n=1}^{\infty} \mathbb{X}_n; \qquad 0 < \mu(\mathbb{X}_n) = d_n < \infty$$

とできる. $\forall B \in \mathcal{A}$ に対して

$$\lambda(B) := \sum_{n=1}^{\infty} \frac{\mu(B \cap \mathbb{X}_n)}{2^n d_n}$$

とおく. いずれの場合でも λ は有限測度で $\nu \ll \lambda\,(\forall \nu \in \mathcal{N})$ となる [12]. ここで

$$\mathcal{Q} := \left\{ \sum_{n=1}^{\infty} c_n \nu_n;\ \sum_{n=1}^{\infty} c_n = 1\, c_n \in \mathbb{R},\ \text{かつ}\ \{\nu_n\}_{n=1}^{\infty} \subset \mathcal{N} \right\}$$

とする. 明らかに

$$\beta \in \mathcal{Q} \ \Rightarrow\ \beta \ll \lambda$$

である.

つぎに

$$\mathcal{D} := \left\{ C \in \mathcal{A};\ \exists \mathsf{Q} \in \mathcal{Q}\ \text{s.t.}\ \lambda\left(\left\{ x \in C;\ \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\lambda}(x) = 0 \right\} \right) = 0\ \text{かつ}\ \mathsf{Q}(C) > 0 \right\}$$

とする. まず $\mathcal{D} \neq \varnothing$ を確認する. そのために $\nu \in \mathcal{N}$ で $\nu \neq 0$ なるものを取り

$$C := \left\{ x \in \mathbb{X};\ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) > 0 \right\}$$

とする. $\mathsf{Q} = \nu$ としたとき

$$\left\{ x \in C;\ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) = 0 \right\} = \varnothing$$

[12] $\lambda(A) = 0$ なる $A \in \mathcal{A}$ をとる. すると $\mu(A \cap \mathbb{X}_n) = 0\,(\forall n \in \mathbb{N})$ となる. したがって

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap \mathbb{X}_n) = 0$$

となる. $\nu \ll \mu$ から $\nu(A) = 0$ となる. よって $\lambda \ll \mu$ がわかる.

となり

$$Q(C) = \nu(C) = \mu(\mathbb{X}) > 0$$

となる. よって $C \in \mathcal{D}$ となる. λ は有限測度なので

$$\sup_{C \in \mathcal{D}} \lambda(C) =: c < \infty$$

となる. よって $\{C_n\}_{n=1}^{\infty} \subset \mathcal{D}$ かつ $C_1 \subset C_2 \subset \cdots$ をうまくとると

$$\lim_{n \to \infty} \lambda(C_n) = c$$

とできる. ここで

$$C_0 := \bigcup_{n=1}^{\infty} C_n$$

とおき, $\mathsf{Q}_n \in \mathcal{Q}\,(n \in \mathbb{N})$ で

$$\mathsf{Q}_n(C_n) > 0 \text{ かつ } \lambda\left(\left\{x \in C_n;\ \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}\right) = 0$$

なるもの [13] を取る. さらに

$$\mathsf{Q}_0 := \sum_{n=1}^{\infty} \frac{\mathsf{Q}_n}{2^n} \in \mathcal{Q} \qquad \left(\because \sum_{n=1}^{\infty} \frac{1}{2^n} = 1\right)$$

とおく. すると

$$\frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda} = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}$$

かつ

$$\left\{x \in C_0;\ \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) = 0\right\} \subset \bigcup_{n=1}^{\infty} \left\{x \in C_n;\ \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}$$

となる. このことより

$$\lambda\left(\left\{x \in C_0;\ \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) = 0\right\}\right) \leq \sum_{n=1}^{\infty} \lambda\left(\left\{x \in C_n;\ \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0\right\}\right)$$

となるで, $C_0 \in \mathcal{D}$ がわかる. さらに

$$\lambda(C_0) = \lambda\left(\bigcup_{n=1}^{\infty} C_n\right) = \lim_{n \to \infty} \lambda(C_n) = c$$

[13] $C_n \in \mathcal{D}$ なのでこのような $\mathsf{Q}_n \in \mathcal{Q}$ が取れる.

がわかる.

あとは $\mathsf{Q}_0 \in \mathcal{Q}$ なので $\forall \nu \in \mathcal{N}$ に対して $\nu \ll Q_0$ を示せばよい. そのために $A \in \mathcal{A}$ で $\mathsf{Q}_0(A) = 0$ なるものと $\forall \nu \in \mathcal{N}$ を取る. $x \in C_0$ に対して

$$\mathsf{Q}_0(A \cap C_0) \leq \mathsf{Q}_0(A) = 0 \text{ かつ } \frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) > 0 \, (\forall x \in C_0)$$

となる [14] ので

$$\lambda(A \cap C_0) = 0 \Rightarrow \nu(A \cap C_0) = 0 \qquad (\because \nu \ll \lambda)$$

となる. ここで

$$C := \left\{ x \in \mathbb{X};\ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) > 0 \right\}$$

とおく. すると

$$x \in C^c \ \Rightarrow \ \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) = 0$$

なので

$$\nu(A \cap C_0^c \cap C^c) \leq \int_{C^c} \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}(x) \, \mathrm{d}\lambda(x) = 0$$

となる. さらに $D := A \cap C_0^c \cap C$ とおくと $D \cap C_0 = \varnothing$ となる. $\lambda(D) > 0$ のとき

$$\lambda(C_0 \cap D) > \lambda(D)$$

であり

$$Q_0\left(\left\{ x \in D;\ \frac{\mathrm{d}Q_0}{\mathrm{d}\lambda}(x) = 0 \right\} \right) = 0$$

となる [15] となるので, $D \in \mathcal{D}$ となる. 明らかに $C_0 \cup D \in \mathcal{D}$ であり

$$\lambda(C_0 \cup D) > \lambda(C_0) = c$$

[14] $x \in C_0$ のときある $n \in \mathbb{N}$ があって $x \in C_n$ となる. すると

$$\lambda\left(\left\{ x \in C_n;\ \frac{\mathrm{d}\mathsf{Q}_n}{\mathrm{d}\lambda}(x) = 0 \right\} \right) = 0$$

なので

$$\frac{\mathrm{d}\mathsf{Q}_0}{\mathrm{d}\lambda}(x) > 0$$

となる.

[15]

$$\mathsf{Q}_0 \in \mathcal{Q} \ \Rightarrow \ \mathsf{Q}_0 \ll \lambda$$

であることからわかる.

となるので, c の定義[16] と矛盾する. しがたって

$$\lambda(D) = 0 \text{ かつ } \nu(D) = 0$$

となる. このことより

$$\begin{aligned}\nu(A) &= \nu(A \cap C_0) + \nu(A \cap C_0^c) \\ &= \nu(A \cap C_0) + \nu(A \cap C_0^c \cap C^c) + \nu(A \cap C_0^c \cap C) \\ &= \nu(A \cap C_0) + \nu(A \cap C_0^c \cap C^c) + \nu(D) = 0\end{aligned}$$

がわかる. よって定理が証明された. □

定理 B.49. ν を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の σ 有限測度とし,

$$\mathcal{P} = \{\mathsf{P};\ \mathsf{P} \text{ は } (\mathbb{X}, \mathcal{A}) \text{ 上の確率分布で, } \mathsf{P} \ll \nu\}$$

を統計的モデルとする. X を母集団分布 $\mathsf{P} \in \mathcal{P}$ からの標本とし, $T(X)$ を統計量とする. このとき $T(X)$ は $\mathcal{P}$ に対する十分統計量であるための必要十分条件は, $(\mathbb{X}, \mathcal{A})$ 上の可測関数 h (これは $\mathsf{P} \in \mathcal{P}$ に依存しない) と T の値域上の可測関数 g_P (これは $\mathsf{P} \in \mathcal{P}$ に依存する) が存在して

$$\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x) = g_{\mathsf{P}}(T(x))h(x), \qquad \mathsf{P} - a.s. \tag{B.44}$$

と書けることである.

Proof. ① $(\Rightarrow)$ の証明: T は $\mathcal{P}$ に対する十分統計量とする. 補題 B.48 から, $\{\mathsf{P}_n\}_{n=1}^{\infty} \subset \mathcal{P}$ と数列 $\{c_n\}_{n=1}^{\infty}$ $(\sum_{n=1}^{\infty} c_n = 1)$ をうまく取れば

$$\mathsf{P} \ll \mathsf{Q} = \sum_{n=1}^{\infty} c_n \mathsf{P}_n$$

とできる. ここで $\dfrac{d\mathsf{Q}}{d\nu}$ は $\mathsf{P} \in \mathcal{P}$ に依存しないことに注意する. $\mathsf{E}[\cdot]$ と $\mathsf{E}[\cdot\,|\,T]$ を P に関する期待値と条件付き期待値とし, $\mathsf{E}_{\mathsf{Q}}[\cdot]$, $\mathsf{E}_{\mathsf{P}_n}[\cdot]$ と $\mathsf{E}_{\mathsf{Q}}[\cdot\,|\,T]$, $\mathsf{E}_{\mathsf{P}_n}[\cdot\,|\,T]$ を Q, P_n それぞれにに関する期待値と条件付き期待値とする. T は十分統計量であるので T を与えたときの X の条件付き分布は $\mathsf{P} \in \mathcal{P}$ に依存しないので

$$\mathsf{E}[\mathbb{1}_A(X)|\,T] = \mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_A(X)|\,T] \qquad (\forall A \in \mathcal{A}) \tag{B.45}$$

が成立することに注意する. $\forall A \in \mathcal{A}$ と $B \in \sigma(T)$ に対して

$$\mathsf{Q}(A \cap B) = \sum_{n=1}^{\infty} c_n \mathsf{P}_n(A \cap B) = \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A\cap B}(X)]$$

[16] $c = \sup_{C \in \mathcal{D}} \lambda(C)$ である.

$$
\begin{aligned}
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_{A\cap B}(X)|\,T]\big] \qquad (\because \text{条件付き期待値の性質}) \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_A(X)\mathbb{1}_B(X)|\,T]\big] \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[\mathbb{1}_B(X)\mathsf{E}_{\mathsf{P}_n}[\mathbb{1}_A(X)|\,T]\big] \qquad (\because \mathbb{1}_B(X) \text{ は } \sigma(T) \text{ 可測}) \\
&= \sum_{n=1}^{\infty} c_n \mathsf{E}_{\mathsf{P}_n}\big[1_B(X)\mathsf{E}[\mathbb{1}_A(X)|\,T]\big] \qquad (\because \text{(B.45)}) \\
&= \sum_{n=1}^{\infty} c_n \int_B \Big\{\mathsf{E}[\mathbb{1}_A(X)|\,T](x)\Big\}\,\mathrm{d}\mathsf{P}_n \\
&= \int_B \Big\{\mathsf{E}[\mathbb{1}_A(X)|\,T](x)\Big\} \sum_{n=1}^{\infty} c_n\,\mathrm{d}\mathsf{P}_n \qquad (\because \text{有界収束定理}) \\
&= \int_B \Big\{\mathsf{E}[\mathbb{1}_A(X)|\,T](x)\Big\}\,\mathrm{d}\mathsf{Q}(x)
\end{aligned}
$$

となる. 一方

$$
\mathsf{Q}(A\cap B) = \int_B \mathbb{1}_A(x)\,\mathrm{d}\mathsf{Q}(x) = \int_B \Big\{\mathsf{E}_Q[\mathbb{1}_A(X)|\,T](x)\Big\}\,\mathrm{d}\mathsf{Q}(x)
$$

となるので

$$
\mathsf{E}[\mathbb{1}_A(X)|\,T](x) = \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x), \qquad \mathsf{Q}-a.s.
$$

を得る. いま

$$
\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(x) =: g_P\big(T(x)\big)
$$

とおくと $\forall A \in \mathcal{A}$ に対して

$$
\begin{aligned}
\mathsf{P}(A) = \mathsf{E}[\mathbb{1}_A(X)] = \mathsf{E}\big[\mathsf{E}[\mathbb{1}_A(X)|\,T]\big] &= \int \Big\{\mathsf{E}[\mathbb{1}_A(X)|\,T](x)\Big\}\,\mathrm{d}\mathsf{P}(x) \\
&= \int \Big\{\mathsf{E}_Q[\mathbb{1}_A(X)|\,T](x)\,\frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(x)\Big\}\,\mathrm{d}\mathsf{Q}(x) \\
&= \int \Big\{\mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x)\,g_{\mathsf{P}}(T(x))\Big\}\,\mathrm{d}\mathsf{Q}(x) \\
&= \int \Big\{\mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)g_{\mathsf{P}}(T(X))|\,T](x)\Big\}\,\mathrm{d}\mathsf{Q}(x) \\
&\qquad (\because g_{\mathsf{P}}(T(x)) \text{ は } \sigma(T) \text{ 可測}) \\
&= \mathsf{E}_{\mathsf{Q}}\big[\mathbb{1}_A(X)g_{\mathsf{P}}(T(X))\big] \qquad (\because \text{条件付き期待値の定義}) \\
&= \int \Big\{\mathbb{1}_A(x)g_{\mathsf{P}}(T(x))\Big\}\,\mathrm{d}\mathsf{Q}(x)
\end{aligned}
$$

$$= \int_A \left\{ g_{\mathsf{P}}(T(x)) \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}(x) \right\} \mathrm{d}\nu(x)$$

となる. よって

$$h = \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}$$

ととればよいことがわかる.

② ($\Leftarrow$) の証明: (B.44) が成り立つとする. T が $\mathcal{P}$ に対する十分統計量であることを示すために

$$\begin{aligned} &\mathsf{P}(A|\,T) = \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T] \qquad \mathsf{P}-a.s. \\ &\Leftrightarrow \int_B \left\{ \mathsf{P}(A|\,T)(x) \right\} \mathrm{d}\mathsf{P}(x) = \int_B \left\{ \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) \qquad \text{(B.46)} \\ &\qquad \big(\forall B \in \sigma(T)\big) \end{aligned}$$

を示せばよい. Radon-Nikodym の微分の性質と (B.44) から

$$\begin{aligned} \frac{d\mathsf{P}}{d\mathsf{Q}}(x) &= \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x) \Big/ \frac{\mathrm{d}\mathsf{Q}}{\mathrm{d}\nu}(x) = \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\nu}(x) \Big/ \sum_{n=1}^{\infty} c_n \frac{\mathrm{d}\mathsf{P}_n}{\mathrm{d}\nu}(x) \\ &= g_{\mathsf{P}}(T(x))h(x) \Big/ \sum_{n=1}^{\infty} c_n g_{\mathsf{P}_n}(T(x))h(x) \\ &= g_{\mathsf{P}}(T(x)) \Big/ \sum_{n=1}^{\infty} c_n g_{\mathsf{P}_n}(T(x)) \end{aligned}$$

となることに注意する. このことより $\dfrac{d\mathsf{P}}{d\mathsf{Q}}(x)$ は $\sigma(T)$ 可測であることがわかる. $\forall B \in \sigma(T)$ に対して

$$\begin{aligned} \int_B \left\{ \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) &= \int_B \left\{ \mathsf{E}_{\mathsf{Q}}[\mathbb{1}_A(X)|\,T](x) \right\} \mathrm{d}\mathsf{Q}(x) \\ &= \int_B \left\{ \mathsf{E}_{\mathsf{Q}}\left[\mathbb{1}_A(X) \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(X) \middle| T \right](x) \right\} \mathrm{d}\mathsf{Q}(x) \\ &= \int_B \left\{ \mathbb{1}_A(x) \frac{\mathrm{d}\mathsf{P}}{\mathrm{d}\mathsf{Q}}(x) \right\} \mathrm{d}\mathsf{Q}(x) \\ &\qquad \big(\text{条件付き期待値の定義}\big) \\ &= \int_B \mathbb{1}_A(x)\, \mathrm{d}\mathsf{P}(x) = \int_B \left\{ \mathsf{E}[\mathbb{1}_A|\,T](x) \right\} \mathrm{d}\mathsf{P}(x) \\ &= \int_A \left\{ \mathsf{P}(A|\,T)(x) \right\} \mathrm{d}\mathsf{P}(x) \end{aligned}$$

となる. よって (B.46) が示されたので, T は $\mathcal{P}$ に対する十分統計量となることがわかる. □

B.10 Dini の定理

関連図書

[1] Almudevar, A. (2022). Theory of Statistical Inference. CRC Press.

[2] Axler, S. (2020). Measure, Integration & Real Analysis. *Springer*.

[3] Bickel, P.J., Doksum, K.A. Mathematical Statistics: Basic Ideas and Selected Topics, vol. 1. 2nd. edition. CRC Press.

[4] Bisgard, J. (2021). Analysis and Linear algebra: The Singular Value Decomposition and Applications. *Student Mathematical Library* **94**. American Mathematical Society.

[5] Borwein, J.M., Lewis, A.S. (2006). ConvexAnalysis and Nonlinear Optimization Theory and Examples, 2nd. edition. Springer.

[6] Boyd, S., Vandenberghe, L. (2018). Introduction to Applied Linear algebra. Cambridge University Press.

[7] DasGupta, A. (2008). Asymptotic Theory of Statistics and Probability. Springer.

[8] Durrett, R. (2019). Probability, 2nd. edition. *Cambridge Series in Statistical and Probabilistic Mathematics*.

[9] Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall.

[10] Folland, G. F. (1999). *Real Analysis*, 2nd edition. John Wiley & Sons, INC.

[11] Gallier, J. (2001). Geometrical Methods and Applications. *Texts in Applied Mathematics* **38**. Springer.

[12] Gassiat, E. Stoltz, G. (2024+). The van Trees inequalities in the spirit of Hájek and Le Cam. To appear in *Statistical Science*.

[13] Hansen, E. (2009). Measure Theory. Fourth edition. Department of Mathematical Statistics, University of Copenhagen.

[14] Huber, P.J. (1981). Robust Statistics. Wiley Series in Probability and Mathematical Statistics.

[15] Hult, H. (2015). Lecture note on SF3961 Graduate course in Statistical inference. https://www.math.kth.se/matstat/gru/Statistical\%20inference/literature.html (accessed at 2023/06/03).

[16] Kaipio, J., Somerslo, E. (2005). Statistical and Computational Inverse Problems. *Applied Mathematical Sciences* **60**. Springer.

[17] Keener, R.W. (2010). Theoretical Statistics. *Springer Texts in Statistics*. Springer.

[18] Komaba, A. Johno, H. Nakamoto, K. (2024). Extension of the one-sample Kolmogorov-Smirnov test. https://doi.org/10.48550/arXiv.2408.10612

[19] Lederer, J. (2022). Fundamentals of High-Dimensional Statistics with Exercises and R Labs. *Springer Text in Statistics*. Springer.

[20] Munkres, J. R. (2000). Topology. Second edition. Pearson.

[21] Neveu, J. (1965). Mathematical Foundations of the Calculus of Probability(English translated by A. Feinstein). Holden-Day, San Francisco.

[22] Nyblom, J. (2023+). Note on Legendre's Method of Least Squares. To appear in *Statistical Science*.

[23] Panchenko, D. (2019). Introduction to Probability Theory. ISBN-13: 978-1-9994190-2-8.

[24] Panchenko, D. (2019). Lecture notes on Probabilty Theory.
https://sites.google.com/site/panchenkomath/ (最終閲覧日 2024 年 9 月 26 日)

[25] Pollard, D. (2014). The van Trees Inequality. Lecture notes from Courses 610.fall2014 at Yale University. http://www.stat.yale.edu/~pollard/Courses/610.fall2014/Handouts/vanTrees.pdf (最終閲覧日 2024 年 10 月 12 日)

[26] Rigollet, P., Hütter, J.-C. (2017). High Dimensional Statistics. Available at https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf(2023/05/17 accessed).

[27] Roussas, G.G. (1997). A Course in Mathematical Statistics, 2nd. edition. Academic Press.

[28] Schilling, R. L. (2017). Measures, Integrals and Martingales. *Cambridge*.

[29] Shao, J. (2003). Mathematical Statistics. 2nd. edition. Springer

[30] Spokoiny, V. and Dickhaus, T. (2015). Basics of Modern Mathematical Statistics. Springer.

[31] Stigler, S. M. (1981). Gauss and the invention of least squares. *Annals of Statistics* **9** 465-474.

[32] Wasserman, L. (2004). All of Statistics. Springer.

[33] E. アルティン (2002). ガンマ関数入門 (第 1 版第 1 刷・上野健爾訳). 日本評論社.

[34] 新井仁之 (2006). 線形代数　基礎と応用 (第 1 刷). 日本評論社.

[35] 藤岡敦 (2020). 『集合と位相』, 裳華房 (第 1 版第 1 刷).

[36] 藤岡敦 (2021). 『入門情報幾何』, 共立出版.

[37] 福水健次 (2010). 『カーネル法入門』, 朝倉書店.

[38] 原啓介 (2017). 『測度・確率・ルベーグ積分』, 講談社.

[39] 日野正訓 (2023). ルベーグ積分の基礎 (初版第 1 刷).　共立出版.

[40] 石谷謙介 (2021). ガイダンス 確率統計 (初版).

[41] 笠原勇二 (2013). 明解 確率論入門 (第 1 版第 2 刷). 日本評論社.

[42] 加藤賢吾 (2015),『数理統計　講義ノート』, Available at

`https://drive.google.com/file/d/0B7C_CufYq6j6ekFINVhTS09lZzQ/view?resourcekey=0-zaYo9XudjhK_K6kMfpq-XQ(accessed 2022/09/22).`

[43] 河野敬雄 (2013). 確率概論 (初版第 5 刷). 京都大学学術出版.

[44] 森棟公夫 (2010). 地球の大きさと最小 2 乗法. 社会とマネジメント **8** 111-130.

[45] 鍋谷清治 (1989). 数理統計学 (初版版第 3 刷). 共立出版.

[46] 永原正章 (2020). スパースモデリング (第 1 版第 4 刷). コロナ社.

[47] 阪本雄二 (2025). 数理統計学の基礎 (初版第 1 刷). 近代科学社

[48] 清水康隆 (2021). 統計学への確率論, その先へ (第 2 版). 内田老鶴圃
.

[49] 清水康隆 (2023). 統計学の漸近論, その先は (第 1 版). 内田老鶴圃.

[50] 谷口正信 (2005). 数理統計・時系列・金融工学 (初版第 1 刷). 朝倉書店.

[51] 高橋礼二 (2019). 線型代数講義 (第 1 版第 2 刷). 日本評論社.

[52] 綿森葉子・田中秀和・田中潮 (2023). 測度論からの数理統計学 (初版第 1 刷). 共立出版.

[53] 柳田英二 (2022). 解析入門 (第 1 版第 1 刷). 裳華房.

[54] 吉田伸生 (2021). 『ルベーグ積分入門』, 日本評論社.