条件付き期待値と十分統計量

(2024 年 11 月 22 日)

今野 良彦
大阪公立大学

概 要

十分統計量の定義から始めて, Fisher-Neyman の因子分解定理, 最小十分統計量, および完備十分統計量を議論して, Basu の定理までを解説することを目指した. 和書の数理統計学の教科書では, 数学的には厳密に書かれることがすくない証明を数学的な厳密性にも気を配りながら丁寧に解説することを試みた.

第 1 では, このノートで使われる設定と記号の解説をする. 第 2 では, 条件確率から条件付き期待値の定義を自然な形で導入する.第 3 では, 測度論的条件付き期待値の性質を利用して, Neyman $-$ Fisher の因子分解定理の証明を連続型分布に対して与える.第 4 と 5 では, 最小十分統計量と完備統計量について関係について説明し, 測度論的な証明を詳述した. 補遺では, 十分統計量の議論に必要な測度論的な積分, 期待値, 条件付き期待値について簡単に説明した. 補遺においては, 重複する内容を記号をかえて説明をしているところが散見されるが, 試作のノートということでそのまま残しておいた.

目 次

0　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.

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 $\mathsf{\Theta}$ の内部を $\mathrm{int}\,\mathsf{\Theta}$ (`\mathrm{int}\,\mathsf{\Theta}`) と書く.

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.
$\mathcal{B}(\mathbb{R})$	$\mathbb{R}$ 上の Boerel 集合族. すなわち, $\mathbb{R}$ の開集合族を含む最小の σ 集合族.
p.d.f.	確率密度関数.
p.m.f.	確率関数 (確率質量関数).
c.d.f.	分布関数.
$\big(\Omega,\, \mathcal{A},\, \mathsf{Pr}\big)$	確率空間.

1 一般的な設定と記法についての注意

基礎となる確率空間を $\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ と書くことにする. 実験で得られるデータはある空でない集合 $\mathbb{X}$ に値を取るものとする. 集合 $\mathbb{X}$ のことを**標本空間**という. $\mathbb{X}$ は d 次元 Euclid 空間 $\mathbb{R}^d$ であったり, 閉区間 $[0, 1]$ 上の実数値連続関数全体の成す集合 $C[0, 1]$ であったりする. いずれにせよ $\mathbb{X}$ は σ 加法族 $\mathcal{B}$ を持つとする. たとえば, $\mathbb{X} = \mathbb{R}$ のとき, $\mathcal{B}$ として $\mathbb{R}$ の開集合族を含む最小の σ 加法族である Borel 集合族 $\mathcal{B}(\mathbb{R})$ とすることが多い.

以上の設定において, 観測データは可測空間 $(\Omega, \mathcal{A})$ から可測空間 $(\mathbb{X}, \mathcal{B})$ への可測関数として定義される確率変数[1]X の実現値と考える. すなわち観測データを x と書いたとき, ある $\omega \in \Omega$ に対して $X(\omega) = x$ と考える.

X による $(\mathbb{X}, \mathcal{B})$ への Pr の押し出しを X **の分布**といい, P^X と書くことする. すなわち

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

と定める. すると P^X は $(\mathbb{X}, \mathcal{B})$ 上の確率測度になり, $(\mathbb{X}, \mathcal{B}, \mathsf{P}^X)$ も確率空間となる. $\mathsf{P}^X(B)$ のことを $\mathsf{P}(X \in B)$ と書いたりもする. この記法は非常に便利であり, 記号の節約になるが, 同時に誤解を誘導する表現でもある. この理由からできるだけこの記法 ($\mathsf{P}(X \in B)$ という記法) を避けるものとする.

$\mathbb{X} = \mathbb{R}^d$ とし, $\mathbb{R}^d$ 上の実数値関数 $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ を考える. 確率変数 $g(X)$ の**期待値** $\mathsf{E}[g(X)]$ を $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の積分

$$\mathsf{E}[g(X)] = \int_\Omega g(X(\omega))\,\mathrm{d}\mathsf{Pr}(\omega)$$

で定める. もちろん, , 右辺の積分が存在するとき, 期待値は定義される. 分布 P^X が確率変数 X による Pr の押し出しであることを利用すると

$$\mathsf{E}[g(X)] = \int_{\mathbb{R}^d} g(x)\,\mathrm{d}\mathsf{P}^X(x)$$

と $\mathbb{R}^d$ 上の積分としても $\mathsf{E}[g(X)]$ は表現できることがわかる.

これは以下のような議論から確認できる. $\forall B \in \mathcal{B}$ に対して指示関数 $\mathbb{1}_B(X(\omega))$ で定義される確率変数を $\mathbb{1}_B(X)$ と書くことにする. すると

$$\begin{aligned}\mathsf{E}[\mathbb{1}_B(X)] &= \int_\Omega \mathbb{1}_B(X(\omega))\,\mathrm{d}\mathsf{Pr}(\omega) = \int_{\mathbb{R}^d} \mathbb{1}_B(x)\,\mathrm{d}(\mathsf{Pr}\circ X^{-1})(x)\\ &= \int_{\mathbb{R}^d} \mathbb{1}_B(x)\,\mathrm{d}\mathsf{P}^X(x)\end{aligned}$$

[1] $\mathbb{X}$ がいろいろな空間を想定するので, 確率変数というより確率要素といったほうがよいかもしれないが, ここでこれらも確率変数ともいうことにする.

と書ける．あとは標準機械[2]を用いると

$$\int_\Omega g(X(\omega))\,\mathrm{d}\mathsf{Pr}(\omega) = \int_{\mathbb{R}^d} g(x)\,\mathrm{d}\mathsf{P}^X(x)$$

がわかる．

$\mathbb{R}^d$ 上の積分として $g(X)$ の期待値を考えているときには, このことを強調するために上の式の右辺を

$$\mathsf{E}[g(X)] := \int_{\mathbb{R}^d} g(x)\,\mathrm{d}\mathsf{P}^X(x)$$

と表記することにする．

いま, μ を可測空間 $\bigl(\mathbb{X}, \mathcal{B}\bigr)$ 上の σ 有限測度とする．連続な分布 P^X が測度 μ に関する p.d.f. $\mathsf{p}^X(x)$ を持つとき

$$\int_{\mathbb{R}^d} g(x)\,\mathrm{d}\mathsf{P}^X(x) = \int_{\mathbb{R}^d} g(x)\mathsf{p}^X(x)\,\mathrm{d}\mu(x) \tag{1.1}$$

が成り立つことがわかる．

これは以下の議論からわかる．g が階段関数のとき, すなわち

$$g(x) = \sum_{i=1}^k a_i \mathbb{1}_{B_i}(x)$$

のときを示せばよい．ここで $B_i \in \mathcal{B}\,(i = 1, 2, \dots, k)$, $a_i \geq 0$, $k \in \mathbb{N}$ である．まず

$$\int_{\mathbb{R}^d} g(x)\,\mathrm{d}\mathsf{P}^X(x) = \sum_{i=1}^k a_i \mathsf{P}^X(B)$$

となる．一方, (1.1) の右辺は

$$\begin{aligned}\int_{\mathbb{R}^d} g(x)\mathsf{p}^X(x)\,\mathrm{d}\mu(x) &= \int_{\mathbb{R}^d} \sum_{i=1}^k a_i \mathbb{1}_{B_i}(x)\mathsf{p}^X(x)\,\mathrm{d}\mu(x) = \sum_{i=1}^k a_i \int_{B_i} \mathsf{p}^X(x)\,\mathrm{d}\mu(x) \\ &= \sum_{i=1}^k a_i \mathsf{P}^X(B_i)\end{aligned}$$

となる．最後の等式は p^X が p.d.f. であることからわかる．あとは標準機械を用いればよい．

X を可測空間 $(\mathbb{X}, \mathcal{B})$ に値をとる確率変数とする．$(\mathbb{X}, \mathcal{B})$ 上の**統計的モデル**を $(\mathbb{X}, \mathcal{B})$ 上の分布の集合 $\mathcal{P}$ で定義する．とくに, $\mathcal{P}$ の要素がある

[2]注意 B.1 を参照のこと．

集合 Θ の要素 θ で添え字づけられているとき, Θ を**母数空間**といい, Θ の要素 θ を**母数**という. すなわち

$$\mathcal{P} = \left\{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta \text{ は } (\mathbb{X},\, \mathcal{B}) \text{ 上の分布}\right\}$$

である. X の独立複製を $X_1,\, X_2,\, \ldots,\, X_n$ とし, $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ とおく. すなわち $X_1,\, X_2,\, \ldots,\, X_n$ は互いに独立な確率変数で, それぞれの分布は共通の分布 P^X であるとする.

いま, $X_i \sim \mathsf{P}_\theta\,(i = 1,\, 2,\, \ldots,\, n)$ とし, $\boldsymbol{X}$ の分布を $\mathsf{P}_\theta^{\otimes n}$ と書くことにする. すると $\boldsymbol{X}$ の標本空間は $\left(\mathbb{X}^n,\, \mathcal{B}^{\otimes n}\right)$ となる.

$X \sim \mathsf{P}_\theta$ のとき, X の期待値は一般に θ にも依存するので, それを強調するために $\mathsf{E}_\theta[X]$ と書くこともある. すなわち

$$\mathsf{E}_\theta[X] = \int_{\mathbb{X}} x \, \mathrm{d}\mathsf{P}_\theta(x)$$

である. $X_1,\, X_2,\, \ldots,\, X_n$ が X の独立複製のとき, 可測関数 $g : \mathbb{X}^n \longrightarrow \mathbb{R}$ に $\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_n)$ を代入して得られる確率変数 $g(\boldsymbol{X})$ の期待値を

$$\mathsf{E}_\theta[g(\boldsymbol{X})] = \int_{\mathbb{X}^n} g(\boldsymbol{x}) \, \mathrm{d}{\mathsf{P}_\theta}^{\otimes n}(\boldsymbol{x}) \qquad \boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n)$$

とも記す.

2 条件付き確率から条件付き期待値を定義する

この節では, 条件付き確率の定義から出発して, 条件付き期待値を部分 σ 加法族に関して可測な確率変数として導入する. Radon-Nikodym の定理を利用する形ではなく, 直観的に理解のしやすい平易な形を取る.

$\left(\Omega,\, \mathcal{A},\, \mathsf{Pr}\right)$ を確率空間とする. 2 つの事象 $A,\, B \in \mathcal{A}\,(0 < \mathsf{Pr}(A),\, \mathsf{Pr}(B) < 1)$ が与えられたとき, **事象 A が与えられた** (知られた) **ときの事象 B の条件付き確率** $\mathsf{Pr}(B|\, A)$ を

$$\mathsf{Pr}(B|\, A) = \frac{\mathsf{Pr}(A \cap B)}{\mathsf{Pr}(A)} \tag{2.2}$$

で定義した. また, 事象 A と B は**確率的に独立** (簡単に独立ということにする) であるとは

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

が成立することであった. 事象 A と B が独立であるための必要十分条件は

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

である. ただし A^c は A の補事象である.

問 2.1. $\mathsf{Pr}(A \cap B) = \mathsf{Pr}(A)\mathsf{Pr}(B) \Leftrightarrow \mathsf{Pr}(B|\,A) = \mathsf{Pr}(B|\,A^c)$ を確認せよ.

条件付き確率を拡張する. A と A^c に $\varnothing$ と Ω を加えた σ 加法族 $\mathcal{F}_A$ を考える.

問 2.2. $\mathcal{F}_A$ は σ 加法族であることを確認せよ.

この σ 加法族 $\mathcal{F}_A$ に関する B の条件付き確率 $\mathsf{Pr}(B|\,\mathcal{F}_A)$ を各 $\omega \in \Omega$ に対して

$$\mathsf{Pr}(B|\,\mathcal{F}_A)(\omega) = \begin{cases} \mathsf{Pr}(B|\,A) & (\omega \in A) \\ \mathsf{Pr}(B|\,A^c) & (\omega \in A^c) \end{cases} \tag{2.3}$$

で定める. 定義から $\mathsf{Pr}(B|\,A)$ は確率変数になる.

問 2.3. $\mathsf{Pr}(B|\,A)$ が確率変数であることを確認せよ. すなわち, $\forall r \in \mathbb{R}$ に対して

$$\{\omega \in \Omega;\, \mathsf{Pr}(B|\,\mathcal{F}_A)(\omega) > r\} \in \mathcal{A}$$

を示せばよい.

さらに $\omega \in \Omega$ を固定すると

$$\mathcal{A} \ni B \mapsto \mathsf{Pr}(B|\,\mathcal{F}_A)(\omega)$$

は $\mathcal{A}$ 上の確率となる.

条件付き期待値についても同様に定める. 事象 $A \in \mathcal{A}$ の確率は定義関数 (これも確率変数になる)

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

の期待値である. すなわち

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

である. このことに注意して, 確率変数を X としたとき, **確率変数 X の事象 $A \in \mathcal{A}$ に関する条件付き期待値** $\mathsf{E}\big[X|\,A\big]$ を

$$\mathsf{E}\big[X|\,A\big] = \frac{\mathsf{E}\big[X\mathbb{1}_A\big]}{\mathsf{Pr}(A)} \tag{2.4}$$

で定めればよい. さらに σ 加法族 $\mathcal{F}_A$ に関する X の条件付き期待値 $\mathsf{E}\big[X|\,\mathcal{F}_A\big]$ を各 $\omega \in \Omega$ に対して

$$\mathsf{E}\big[X|\,\mathcal{F}_A\big](\omega) = \begin{cases} \mathsf{E}\big[X|\,A\big] & (\omega \in A) \\ \mathsf{E}\big[X|\,A^c\big] & (\omega \in A^c) \end{cases}$$

で定める. これも確率変数になることに注意せよ.

問 2.4. $\mathsf{E}\big[X|\,A\big] = a_1$, $\mathsf{E}\big[X|\,A^c\big] = a_2\,(a_1 < a_2)$ とする． このとき, $\mathsf{E}\big[X|\,\mathcal{F}_A\big]$ は確率変数となることを示せ．すなわち, $\forall r \in \mathbb{R}$ に対して

$$\{\omega \in \Omega;\, \mathsf{E}\big[X|\,\mathcal{F}_A\big](\omega) > r\} \in \mathcal{A}$$

を示せばよい．

この考え方を一般化してみよう．$\mathcal{C} = \{A_1,\, A_2,\, \ldots,\, A_n\}$ を Ω の分割とする．すなわち

$$A_i \in \mathcal{A}\,(i = 1,\, 2,\, \ldots,\, n), \quad \bigcup_{k=1}^{n} A_k = \Omega, \quad A_i \bigcap A_j = \varnothing\,(i \neq j)$$

である．さらに

$$\mathcal{F}_{\mathcal{C}} := \sigma[A_1,\, A_2,\, \ldots,\, A_n]$$

とおく．すなわち $\mathcal{C}$ を含む最小の σ 加法族である．このとき, 確率変数 X の $\mathcal{F}_{\mathcal{C}}$ に関する条件付き期待値 $\mathsf{E}\big[X|\,\mathcal{F}_{\mathcal{C}}\big]$ を各 $\omega \in \Omega$ に対して

$$\mathsf{E}\big[X|\,\mathcal{F}_{\mathcal{C}}\big](\omega) = \mathsf{E}\big[X|\,A_k\big] \qquad \big(\omega \in A_k;\, k = 1,\, 2,\, \ldots,\, n\big)$$

で定義する．特に確率変数 X を事象 B の定義関数 $\mathbb{1}_B(X)$ とすると

$$\mathsf{E}\big[\mathbb{1}_B(X)|\,\mathcal{F}_{\mathcal{C}}\big] = \mathsf{Pr}\big(X^{-1}(B)|\,\mathcal{F}_{\mathcal{C}}\big)$$

と書ける．

つぎに, より一般的な条件付き期待値を定義する．Y を確率変数とし, $g : \mathbb{R} \longrightarrow \mathbb{R}$ を関数とし, $\mathsf{E}\big[|g(X)|\big] < \infty$ をみたすとする．確率変数 $Y = y$ が与えられたときの X の関数 $g(X)$ の条件付き期待値 $\mathsf{E}\big[g(X)|\,Y = y\big]$ を定義しよう．実は一般にはこの条件付き期待値を具体的に書き下して定義するのは困難である．そのために, つぎのよう関係式をみたす y の関数 $g_Y(y)$ として定める．この $g_Y(y)$ のことを $\mathsf{E}\big[g(X)|\,Y = y\big] =: g_Y(y)$ と書く．さらに $g_Y(Y)$ のことを $\mathsf{E}\big[g(X)|\,Y\big] := g_Y(Y)$ と書くことにする．

定義 2.1. 実関数 $g_Y(y)$ が存在して, 任意の有界かつ区分的に連続な関数 $h(t)$ に対して

$$\mathsf{E}\big[h(Y)g(X)\big] = \mathsf{E}\big[h(Y)g_Y(Y)\big] \tag{2.5}$$

が成り立つとき, y の関数 $g_Y(y)$ を $\mathsf{E}\big[g(X)|\,Y = y\big] := g_Y(y)$ と書き, 条件 $Y = y$ が与えられたときの条件付き期待値という．

注意 2.2. 条件 (2.5) を証明するためには h を指示関数と限定して, (2.5) を確認することができると測度論の標準機械により, 区分的に連続な関数に対しても (2.5) が成立することがわかる．

注意 2.3. 定義 2.1 で定めた $g_Y(y)$ は y の関数としてただ一つ定まることがわかる．これは Radon-Nikodym の定理 (定理 A.15) から保証されることがわかる．

問 2.5. g_Y の一意性を直接的に (Radon-Nikodym の定理 (定理 A.15) を経由せず) 証明せよ．

例 2.4. 確率変数 (X, Y) は連続型とし, 同時 p.d.f. $\mathsf{p}^{(X,Y)}(x, y)$ を持つ場合に具体的に $\mathsf{E}\left[g(X) | Y = y\right]$ を求めてみよう．

区分的に連続な y の関数 h に対して, 定義 2.1 の等式 (2.4) の左辺は

$$\begin{aligned}
&\mathsf{E}\left[h(Y)g(X)\right] \\
&\qquad = \int_{\mathbb{R}^2} h(y)g(x)\mathsf{p}^{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&\qquad = \int_{\{y\in\mathbb{R};\,\mathsf{p}^Y(y)>0\}} h(y)\left\{\int_{-\infty}^{\infty} g(x)\frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x\right\}\cdot\mathsf{p}^Y(y)\,\mathrm{d}y \qquad (2.6)
\end{aligned}$$

となる．ただし, p^Y は確率変数 Y の周辺 p.d.f. である．一方, 右辺は

$$\mathsf{E}\left[h(Y)g_Y(Y)\right] = \int_{\{y\in\mathbb{R};\,\mathsf{p}_Y(y)>0\}} h(y)g_Y(y)\cdot\mathsf{p}^Y(y)\,\mathrm{d}y \tag{2.7}$$

である．(2.6) と (2.7) の右辺を比較すると

$$g_Y(y) = \int_{-\infty}^{\infty} g(x)\frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x \tag{2.8}$$

であることがわかる．ここで, 任意の $B \in \mathcal{B}(\mathbb{R})$ に対して, $g(x) = \mathbb{1}_B(x)$ とおくと

$$\begin{aligned}
\mathsf{Pr}(X \in B | Y = y) &= \mathsf{E}[\mathbb{1}_B(X) | Y = y] \\
&= g_Y(y) \quad \big(\because (2.5)\big) \\
&= \int_{-\infty}^{\infty} \mathbb{1}_B(x)\frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x \quad \big(\because (2.8)\big) \\
&= \int_B \frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x
\end{aligned}$$

となる． よって, 条件付き p.d.f. の定義は以下のようになる． □

定義 2.5. 条件 $Y = y$ を与えたときの確率変数 X の条件付き p.d.f. $\mathsf{p}^{X|Y}(x|y)$ を

$$\mathsf{p}^{X|Y}(x| y) := \frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^Y(y)} \qquad (\text{ただし } \mathsf{p}^Y(y) > 0 \text{ のとき})$$

で定める．

問 2.6. $a_1, a_2 \in \mathbb{R}$, $A \in \mathcal{A}$ とし, 各 $\omega \in \Omega$ に対して $Y(\omega) = a_1 \mathbb{1}_A(\omega) + a_2 \mathbb{1}_{A^c}(\omega)$ とする. このとき

$$\mathsf{E}[YX] = \mathsf{E}\big[Y\mathsf{E}[X|\,\mathcal{F}_A]\big]$$

を示せ.

定理 2.6. (i) $\mathsf{E}\big[\mathsf{E}[g(X)|\,Y]\big] = \mathsf{E}[g(X)]$.
(ii) $Y\mathsf{E}\big[g(X)|\,Y\big] = \mathsf{E}\big[Yg(X)|\,Y\big]$.
(ii) $\big|\mathsf{E}\big[g(X)|\,Y\big]\big|^2 \le \mathsf{E}\big[|g(X)|^2|\,Y\big]$.
(iv) $\mathsf{E}[g(X)] = \mu$ とおく. このとき

$$\mathsf{E}\big[\big\{\mathsf{E}[g(X)|\,Y] - \mu\big\}^2\big] \le \mathsf{E}[\{g(X) - \mu\}^2].$$

(v) X と Y が独立ならば $\mathsf{E}[g(X)|\,Y] = \mathsf{E}[g(X)]$.

注意 2.7. (ii) と (iii) の主張は確率変数の大小関係なので, 「ほとんど確実に成立する」という主張である.

Proof. (i) の証明: (2.5) において $h = 1$ とおけばよい.
(ii) の証明: 区分的に連続な関数 h に対して

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

となる. (2.5) と一意性から

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

がわかる.
(iii) の証明: 関数 g_1 と g_2 と 定数 a_1, a_2 に対して

$$\mathsf{E}\big[a_1g_1(X) + a_2g_2(X)|\,Y\big] = a_1\mathsf{E}[g_1(X)|\,Y] + a_2\mathsf{E}[g_2(X)|\,Y]$$

がわかる. また $g_1(x) \ge 0$ ならば

$$\mathsf{Pr}\big(\mathsf{E}[g_1(X)|\,Y] \ge 0\big) = 1$$

がわかる. これらのことより $t \in \mathbb{R}$ に対して

$$\begin{aligned} 0 &\le \mathsf{E}\big[\{g_1(X) + tg_2(X)\}^2|\,Y\big] \\ &= \mathsf{E}[g_1^2(X)|\,Y] + 2t\mathsf{E}\big[g_1(X)g_2(X)|\,Y\big] + t^2\mathsf{E}[g_2^2(X)|\,Y] \end{aligned}$$

を得る. これに判別式を適用すると

$$\big\{\mathsf{E}\big[g_1(X)g_2(X)|\,Y\big]\big\}^2 \le \mathsf{E}[g_1^2(X)|\,Y]\mathsf{E}[g_2^2(X)|\,Y]$$

となる．ここで $g_1 = g_2$ とすると (iii) は証明される．
(iv) の証明: (i) から $\mathsf{E}\big[\mathsf{E}[g(X)|\,Y]\big] = \mu$ であることに注意する．(iii) から

$$\begin{aligned}\mathsf{E}\Big[\big\{\mathsf{E}[g(X)|\,Y] - \mu\big\}^2\Big] &= \mathsf{E}\Big[\big\{\mathsf{E}[g(X) - \mu|\,Y]\big\}^2\Big] \\ &\leq \mathsf{E}\Big[\mathsf{E}[(g(X) - \mu)^2|\,Y]\Big] \qquad \big(\because \text{(ii)}\big) \\ &= \mathsf{E}\Big[(g(X) - \mu)^2\Big] \qquad \big(\because \text{(i)}\big)\end{aligned}$$

よりわかる．
(v) の証明: h を区分的に連続な y の関数とする．X と Y は独立であるので

$$\begin{aligned}\mathsf{E}\big[h(Y)\mathsf{E}[g(X)|\,Y]\big] &= \mathsf{E}[h(Y)g(X)] \qquad \big(\because (2.5)\big) \\ &= \mathsf{E}[h(Y)]\mathsf{E}[g(X)] \qquad \big(\because X \text{ と } Y \text{ は独立}\big) \\ &= \mathsf{E}\big[h(Y)]\mathsf{E}[g(X)]\big]\end{aligned}$$

となる．すると (2.5) と一意性から

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

を得る． □

3 十分統計量の定義とその性質

この節では, 十分統計量の動機付けを [10] に従い述べ, その定義を述べる．さらに, Fisher-Neyman の因子分解定理の測度論的な証明を与える．

連続型確率変数 X と Y は独立で同じ p.d.f.

$$\mathsf{p}_\theta(x) = \begin{cases} \theta e^{-\theta x} & (x \geq 0), \\ 0 & (x < 0) \end{cases}$$

を持つとする．ただし $\theta > 0$ である．また $\mathsf{p}_\theta(x)$ に対応する c.d.f. は

$$\mathsf{F}_\theta(x) = \int_{-\infty}^{x} \mathsf{p}_\theta(u)\,\mathrm{d}u = \begin{cases} 1 - e^{-\theta x} & (x > 0), \\ 0 & (x \leq 0) \end{cases} \tag{3.9}$$

となる．さらに, X と Y と独立な連続型確率変数 U は開区間 (0, 1) 上の一様分布に従うとする．

いま

$$T := X + Y$$

とし

$$\widetilde{X} := UT, \qquad \widetilde{Y} := (1 - U)T$$

と定める．このとき $(\widetilde{X}, \widetilde{Y})$ の同時 p.d.f. を求めたい．まず T の p.d.f. p^T は

$$\mathsf{p}^T(t) = \begin{cases} t\theta^2 e^{-\theta} & (t > 0) \\ 0 & (\text{その他の場合}) \end{cases} \tag{3.10}$$

となること[3]に注意する．

問 3.1. (3.10) を示せ．

T と U は独立なので, (T, U) の同時 p.d.f. $\mathsf{p}_\theta^{(T,U)}$ は

$$\mathsf{p}_\theta^{(T,U)}(t, u) = \begin{cases} t\theta^2 e^{-\theta t} & (t > 0,\, 0 < u < 1) \\ 0 & (\text{その他の場合}) \end{cases}$$

で与えられる．これより $B \in \mathcal{B}(\mathbb{R}^2)$ に対して

$$\begin{aligned} \mathsf{Pr}\big((\widetilde{X}, \widetilde{Y}) \in B\big) &= \mathsf{P}^{(\widetilde{X},\widetilde{Y})}(B) \\ &= \iint \mathbb{1}_B(tu,\, (1-u)t)\mathsf{p}_\theta^{(T,U)}(t, u)\,\mathrm{d}u\,\mathrm{d}t \end{aligned}$$

となる．変数変換

$$x = ut, \qquad dx = \frac{du}{t}$$

と Fubini の定理 (定理 $A.10$) を用いて積分の順序交換する．次に変数変換

$$y = t - x$$

を用いると $\big(\widetilde{X}, \widetilde{Y}\big)$ の同時 p.d.f. $\mathsf{p}_\theta^{(\widetilde{X},\widetilde{Y})}$ は

$$\mathsf{p}_\theta^{(\widetilde{X},\widetilde{Y})}(x, y) = \frac{\mathsf{p}_\theta^{(T,U)}\left(x+y,\, \frac{x}{x+y}\right)}{x+y} = \begin{cases} \theta^2 e^{-\theta(x+y)} & (x > 0,\, y > 0) \\ 0 & (\text{その他の場合}) \end{cases} \tag{3.11}$$

となること[4]がわかる．これは (X, Y) の同時 p.d.f. と同じである．この計算は, $\big(\widetilde{X}, \widetilde{Y}\big)$ の同時 p.d.f. は (X, Y) のものと同じであることを示す．

母数 θ に対する情報を考えたとき, 組 $\big(\widetilde{X}, \widetilde{Y}\big)$ と組 (X, Y) は同じ情報を持つ．しかし, $\big(\widetilde{X}, \widetilde{Y}\big)$ は $T = X + Y$ より計算されている．U の分布は母数 θ に依存しないので, U は人工的に生成できる．したがって統計量 T は母数 θ に関する情報を (X, Y) と同じだけ持つことがわかる．和

[3]問 3.1 を参照のこと．
[4]問 3.2 を参照のこと．

$T = X + Y$ は**十分統計量**と呼ばれている. T から (X, Y) と θ について同じ量の情報を持つ「偽データ」を作ることができる. なぜならば $T = t$ が与えられたときの X と Y の条件付き分布 Q_t は

$$\mathsf{Q}_t(B) = \mathsf{Pr}\big((X, Y) \in B | T = t\big) = \mathsf{Pr}\big[(Ut, (1-U)t) \in B | T = t\big]$$

と表現できるからである.

問 3.2. (3.11) を示せ.

定義 3.1. 可測空間 $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ に値をとる確率変数 X と母数空間 Θ で添え字づけられた統計的モデル

$$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}) \text{ 上の確率測度}\}$$

を考える. $X \sim \mathsf{P}_\theta\,(\mathsf{P}_\theta \in \mathcal{P})$ のとき, 統計量 $T = T(X)$ は統計的モデル $\mathcal{P}$(または $\theta \in \Theta$) に対する**十分統計量**であるとは, $T = t$ が与えられたときの θ のもとの X の条件付き分布が θ に依存しないときをいう.

可測空間 $(\mathbb{X}, \mathcal{B})$ に値を取る確率変数 X は分布 P_θ を持つとし, T は $\theta \in \Theta$ に対する十分統計量とする. $B \in \mathcal{B}$ に対して

$$\mathsf{Q}_t(B) := \mathsf{Pr}(X \in B | T = t)$$

とおく. $\mathsf{Pr}(X \in B | T) := \mathsf{Q}_T(B)$ と定めたとき

$$\mathsf{Pr}(X \in B) = \mathsf{E}\big[\mathsf{Pr}(X \in B | T)\big] = \mathsf{E}\big[\mathsf{Q}_T(B)\big]$$

となる. $T = t$ のとき, 「偽データ」$\widetilde{X}$ は分布 Q_t から生成されたとする. このとき

$$\widetilde{X} | T = t \sim \mathsf{Q}_t$$

なので

$$\mathsf{Pr}\big(\widetilde{X} \in B\big) = \mathsf{E}\big[\mathsf{Pr}(\widetilde{X} \in B | T)\big] = \mathsf{E}[\mathsf{Q}_T(B)]$$

となる. よって X は $\widetilde{X}$ と同じ分布を持つ.

定義 3.2. μ を $\big(\mathbb{X}, \mathcal{B}\big)$ 上の σ 有限測度とする. 母数空間 Θ で添え字付られた統計的モデル

$$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta \text{ は } (\mathbb{X}, \mathcal{B}) \text{ 上の分布}\}$$

は σ 有限な測度 μ に**支配されている**とき, $\forall \theta \in \Theta$ に対して, P_θ は μ に関して絶対連続のときをいう.

定理 3.3 (Fisher-Neyman の因子分解定理)**.** $(\mathbb{X}, \mathcal{B})$ を標本空間とする. 母数空間 Θ で添え字づけられている標本空間 $(\mathbb{X}, \mathcal{B})$ 上の統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta,\, \mathsf{P}_\theta$ は確率測度$\}$ は σ 有限測度 μ に支配されているとする. このとき, 統計量 T が十分であるために必要十分条件は, ある可測関数 $g_\theta \geq 0$ と $h(x) \geq 0$ が存在して, μ に関する Radon-Nikodym(定理 A.15) の微分 (p.d.f.) p_θ が

$$\mathsf{p}_\theta(x) = g_\theta\big(T(x)\big)h(x), \qquad \text{a.e. } \mu \tag{3.12}$$

と書けることである.

Proof. まず, P_θ の μ に関する p.d.f. p_θ が

$$\mathsf{p}_\theta(x) = g_\theta\big(T(x)\big)h(x)$$

と書けたとする. 補題 D.3 から, 一般性を失わずに μ を確率測度と仮定[5]してよい. E^* を $X \sim \mu$ のときの期待値を表すとする. さらに $X \sim \mu$ と $X \sim \mathsf{P}_\theta$ のときの $T(X)$ の周辺分布を G^* と G_θ とそれぞれ記すことにする. Q_t を $X \sim \mu$ のときの $T = t$ が与えられたときの X の条件付き分布とする. また, $X \sim \mu$ は θ に依存しないので, Q_t も θ に依存しないことに注意する. T の G^* に関する p.d.f. を求めるために, 可積分関数 $f(t)$ をとる. Fubini の定理 (定理 A.10) を用いて計算すると

$$\mathsf{E}_\theta\big[f(T)\big] = \int f(t)g_\theta(t)w(t)\,\mathrm{d}\mathsf{G}^*(t) \tag{3.13}$$

を得る[6]. ただし $w(t) = \int h(x)\,\mathrm{d}\mathsf{Q}_t(x)$ とした. f を指示関数とすると, G_θ は G^* に関する p.d.f. $g_\theta(t)w(t)$ を持つことがわかる. すなわち $\mathrm{d}\mathsf{G}_\theta(t) = g_\theta(t)w(t)\,\mathrm{d}\mathsf{G}^*$ である.

つぎに, $w(t) \neq 0$ なる t に対して, $\widetilde{\mathsf{Q}}_t$ は Q_t に関する p.d.f. $h(x)/w(t)$ をもつ分布とする. すなわち $\mathrm{d}\widetilde{\mathsf{Q}}_t(t) = \dfrac{h(x)}{w(t)}\,\mathrm{d}\mathsf{Q}_t(x)$ である. 定め方から $\widetilde{\mathsf{Q}}_t$ も θ に依存しないことに注意する. $w(t) = 0$ なる t に対しては $\widetilde{\mathsf{Q}}$ を任意の確率測度で定めればよい. このとき, 可積分関数 $f(x, t)$ に対して

$$\mathsf{E}_\theta\big[f(X, T)\big] = \iint f(x, t)\,\mathrm{d}\widetilde{\mathsf{Q}}_t(x)\,\mathrm{d}\mathsf{G}_\theta(t) \tag{3.14}$$

[5] 補題 $D.3$ から, ある確率測度 $\mathsf{P}_0 \in \mathcal{P}$ が存在して, $\forall \mathsf{P} \in \mathcal{P}$ に対して, $\mathsf{P} \ll \mathsf{P}_0$ となる. よって, (3.12) に注意すると

$$\frac{\mathrm{dP}}{\mathrm{dP}_0} = \frac{\mathrm{dP}}{\mathrm{d}\mu}\frac{\mathrm{d}\mu}{\mathrm{dP}_0} = g_\theta\big(T(x)\big)h(x)\frac{\mathrm{d}\mu}{\mathrm{dP}_0} =: g_\theta\big(T(x)\big)\widetilde{h}(x)$$

と書ける. 定め方から, $\widetilde{h}$ は θ に依存しないことがわかる.

[6] 問 3.3 を参照のこと.

を得る[7]．よって $\widetilde{\mathsf{Q}}_t$ は P_θ のもと $T=t$ が与えらられたときの X の条件付き測度となる．$\widetilde{\mathsf{Q}}_t$ は θ に依存しないことが構成の仕方からわかる．よって T は十分統計量になる．

T は十分統計量のとき, (3.12) が成り立つことを示す．

$\mathsf{P}^* \in \mathcal{P}$ をうまく選んで固定する．すると $\mathsf{P}_\theta \ll \mathsf{P}^*$ とできることが補題 D.3 からわかる．$X \sim \mathsf{P}^*$ のときの $T=T(X)$ の周辺分布を G^* と記し, $T=t$ が与えられたときの X の条件付き分布を Q_t と書くこと[8]にする．G^* と Q_t による混合分布 $\widehat{\mathsf{P}}$ を

$$\begin{aligned}\widehat{\mathsf{P}}(B) &= \int \mathsf{Q}_t(B)\,\mathrm{d}\mathsf{G}^*(t)\\ &= \iint \mathbb{1}_B(x)\,\mathrm{d}\mathsf{Q}_t(x)\,\mathrm{d}\mathsf{G}^*(t) \qquad (B \in \mathcal{B})\end{aligned} \tag{3.15}$$

と定義する．ただし $(\mathbb{X}, \mathcal{B})$ は X が値をとる可測空間である．このとき, x の (可積分) 関数 f に対して

$$\int f(x)\,\mathrm{d}\widehat{\mathsf{P}}(x) = \iint f(x)\,\mathrm{d}\mathsf{Q}_t(x)\,\mathrm{d}\mathsf{G}^*(t) \tag{3.16}$$

と書ける．これは (3.15) から出発する標準機械 (standard machine) で証明できる[9]．

$X \sim \mathsf{P}_\theta\,(\forall\theta \in \Theta)$ のとき, $T=t$ が与えられたときの X の条件分布の G^* に関する p.d.f. を g_θ と記す[10]．このとき, 任意の $\forall B \in \mathcal{B}$ に対して,

$$\mathsf{P}_\theta(B) = \mathsf{Pr}(X \in B) = \int_B \mathsf{g}_\theta\big(T(x)\big)\,\mathrm{d}\widehat{\mathsf{P}}(x) \tag{3.17}$$

を得る[11]．よって P_θ は $\widehat{\mathsf{P}}$ に関する p.d.f. $\mathsf{g}_\theta\big(T(\,\cdot\,)\big)$ を持つことがわかる．また, 混合分布 $\widehat{\mathsf{P}}$ は μ に関して絶対連続であることがわかる．すなわち

$$\forall N \in \mathcal{B};\, \mu(N) = 0 \Rightarrow \widehat{\mathsf{P}}(N) = 0 \tag{3.18}$$

を示せばよい[12]．Radon-Nikodym の定理 (定理 A.15) を適用すると

$$\frac{d\widehat{\mathsf{P}}}{d\mu}(x) = h(x) \tag{3.19}$$

[7] 問 3.4 を参照のこと．

[8] $\mathsf{P}^* \in \mathcal{P}$ であるので, 本来ならば, Q_t^* と書くのが自然である．しかし, T が十分統計量なので, $T=t$ が与えられたときの X の条件分布は θ によらないので, Q_t^* を Q_t と書いている．

[9] 問 3.5 を参照．

[10] 問 3.6 を参照のこと．

[11] 問 3.7 を参照のこと．

[12] 問 3.8 を参照のこと．

が存在する. $\widehat{\mathsf{P}}$ は P^* を固定してつくったものなので, h は θ には依存しないことがわかる. したがって (3.18) と (3.19) から

$$\frac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\mu}(x) = \frac{\mathrm{d}\mathsf{P}_\theta}{\mathrm{d}\widehat{\mathsf{P}}}(x)\frac{\mathrm{d}\widehat{\mathsf{P}}}{\mathrm{d}\mu}(x) = \mathsf{g}_\theta\big((T(x)\big)h(x)$$

とかけることがわかる. □

問 3.3. (3.13) を示せ.

問 3.4. (3.14) を示せ.

問 3.5. (3.16) を (3.15) から出発する標準機械で証明せよ.

問 3.6. X の条件分布の G^* に関する p.d.f. g_θ が存在することを Radon-Nikodym の定理 (定理 A.15) を用いて確認せよ.

問 3.7. (3.17) を確認せよ.

問 3.8. (3.11) を示せ.

問 3.9. (3.18) を示せ.

4 最小十分統計量

この節では, [10] を参考にして, 最小十分統計量の定義とその判定定理を述べる.

$(\mathbb{X}, \mathcal{B})$ を標本空間とし, $X \in \mathbb{X}$ を確率変数とする. 統計量 $T := T(X)$ は統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta; \mathsf{P}_\theta \text{ は分布}, \theta \in \Theta\}$ に対して十分であるとき, ある関数 f があって, $T = f(\widetilde{T})$ と表現されると統計量 $\widetilde{T}$ も $\mathcal{P}$ に対して十分となる. なぜならば

$$\mathsf{E}[X|\widetilde{T}] = \mathsf{E}\big[\mathsf{E}[X|T]\big|\widetilde{T}\big]$$

と書きなおす. T は十分統計量なので, $\mathsf{E}_\theta[X|T]$ は θ に依存しないので, $\mathsf{E}[X|\widetilde{T}]$ も θ に依存しないことがわかる. このことを踏まえ, 以下では最小十分統計量の定義を与える.

定義 4.1. 統計量 T が最小十分であるとは, T は十分統計量であり, 任意の十分統計量 $\widetilde{T}$ に対して, ある関数 f が存在して $T = f(\widetilde{T})$, (a.e.$\mathcal{P}$) と書けるときをいう. ここで (a.e.$\mathcal{P}$) の意味は, 等号が成立しない集合はすべての $\mathsf{P} \in \mathcal{P}$ に対して零集合となっていることである.

注意 4.2. 最小十分統計量は常に存在するとは限らない．存在のために十分条件は以下のようにある．$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を統計的モデルとする．2 つの母数 $\theta_1,\, \theta_2\,(\in \Theta)$ の疑似測度を

$$\rho(\theta_1,\, \theta_2) = \sup_{B \in \mathcal{B}(\mathbb{X})} \left| \mathsf{P}_{\theta_1}(B) - \mathsf{P}_{\theta_2}(B) \right|$$

に関して Θ が可分[13]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 [31, pp.78-81] を参照のこと．

統計的モデル $\mathcal{P}$ が測度 μ でおさえられているとき, $\mathsf{P}_\theta\,(\theta \in \Theta)$ の μ に関する p.d.f. を $\mathsf{p}_\theta(x)$ とする．$X \sim \mathsf{P}_\theta$ のとき, 関数 $\theta \mapsto \mathsf{p}_\theta(X)$ を尤度関数とみなすことできる．十分統計量 T は尤度関数のグラフを描くために十分な情報をもっている．

定理 4.3. $\big(\mathbb{X},\, \mathcal{B}\big)$ を標本空間とし, それ上の統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ は測度 μ で支配されており, p.d.f. $\mathsf{p}_\theta(x) = g_\theta(T(x))h(x)$ を持つとする．任意の $x,\, y \in \mathbb{X}$ に対して, 関数

$$\theta \mapsto \frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$$

が定数ならば, $T(x) = T(y)$ が成り立つとき, T は最小十分統計量となる．

Proof. 厳密な証明は測度論の知識を用いることになる．ここでは証明の基本的な考え方を述べるだけにする．$\widetilde{T}$ を任意の十分統計量とする．すると Fisher-Neyman の因子分解定理から

$$\mathsf{p}_\theta(x) = \widetilde{g}_\theta(\widetilde{T}(x))\widetilde{h}(x)$$

と書ける．かりに T は $\widetilde{T}$ の関数でないとすると, あるデータ x と y があって,

$$\widetilde{T}(x) = \widetilde{T}(y) \quad かつ \quad T(x) \neq T(y) \tag{4.20}$$

とできること[14]になる．しかし

$$\frac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)} = \frac{\widetilde{g}_\theta(\widetilde{T}(x))\widetilde{h}(x)}{\widetilde{g}_\theta(\widetilde{T}(y))\widetilde{h}(y)} = \frac{\widetilde{h}(x)}{\widetilde{h}(y)}$$

[13] 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．

[14] $\widetilde{T} = f(T)$ と書けるとき,

$$T(x) = T(y) \Rightarrow \widetilde{T}(x) = \widetilde{T}(y)$$

なので, $\widetilde{T} = f(T)$ と書けない場合は, ある $x,\, y$ があって

$$T(x) = T(y) \quad かつ \quad \widetilde{T}(x) \neq \widetilde{T}(y)$$

となる．

となり, 関数 $\theta \mapsto \dfrac{\mathsf{p}_\theta(x)}{\mathsf{p}_\theta(y)}$ は定数となる. よって, T が最小十分統計量であるという仮定より $T(x) = T(y)$ となり, (4.20) と矛盾する. したがって, ある関数 f があって, $\widetilde{T} = f(\widetilde{T})$ とかける. □

例 4.4. $(\mathbb{X}, \mathcal{B})$ 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. $\mathsf{p}_\theta(x)\,(\theta \in \Theta)$ が

$$\mathsf{p}_\theta(x) = \exp\{\big(A(\theta)|\,T(x)\big) - \kappa(\theta)\}h(x)$$

と書ける. ただし $\kappa : \Theta \longrightarrow \mathbb{R}^d$ であり, $A : \mathbb{X} \longrightarrow \mathbb{R}^d$ で $T \in \mathbb{R}^d$ である. また, $(\cdot|\cdot)$ 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

$$\big(A(\theta)|\,T(x)\big) = \big(A(\theta)|\,T(y)\big) + c$$

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

$$\begin{aligned}&\big(\big[A(\theta_0) - A(\theta_1)\big]|\,T(x)\big) = \big(\big[A(\theta_0) - A(\theta_1)\big]|\,T(y)\big)\\ \Leftrightarrow\ &\big(\big[A(\theta_0) - A(\theta_1)|\,T(x) - T(y)\big) = 0\end{aligned}$$

を得る. $\mathrm{span}\Big[\{A(\theta_0) - A(\theta_1) \in \mathbb{R}^d : \theta_0,\, \theta_1 \in \Theta\}\Big] = \mathbb{R}^d$ のとき, $T(x) = T(y)$ となり, T は最小十分統計量となる. □

例 4.5. $n \geq 2\,(n \in \mathbb{N})$ 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

$$\mathsf{p}_\theta^{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2^n}\exp\bigg\{-\sum_{i=1}^n |x_i - \theta|\bigg\} \qquad \big(\boldsymbol{x} = (x_1,\, x_2,\, \ldots,\, x_n) \in \mathbb{R}^n\big)$$

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 $x = (x_1, x_2, \dots, x_n)$ と $y = (y_1, y_2, \dots, y_n)$ としたとき, $\sum_{i=1}^n |x_i - \theta|$ と $\sum_{i=1}^n |y_i - \theta|$ の差は θ に依存しない. この二つの θ の関数は区分的に連続で, 不連続点は $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ と $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ である. 関数

$$\theta \mapsto \sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |y_i - \theta|$$

が θ に関して定数であるときは, ジャンプの点が一致することである. したがって, $x_{(i)} = y_{(i)}\,(i = 1, 2, \dots, n)$ となるので, $\boldsymbol{T}$ は最小十分であることがわかる.

5 完備十分統計量

この節では, [10] を参考にして, 完備十分統計量の定義を述べる. 次に, (有界) 完備十分ならば, 最小十分であることの証明を丁寧に述べる. 最後に, 完備十分統計量と局外統計量との独立性を主張する Basu の定理を与える.

定義 5.1. $(\mathbb{X}, \mathcal{B})$ を標本空間とする. $\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}$ を標本空間 $(\mathbb{X}, \mathcal{B})$ 上の統計的モデルとする. Θ は母数空間で, P_θ は Θ の要素で添え字づけられた分布である. 統計量 T が統計的モデル $\mathcal{P}$ に対して**完備**であるとは, ある可測関数 f に対して

$$\mathsf{E}_\theta\big[f(T)\big] = 0 \qquad (\forall \theta \in \Theta))$$

ならば, $f(T) = 0\,(\text{a.e.}\,\mathcal{P})$ となるときをいう.

定義 5.2. 統計量 T は**有界完備**では, t の有界な可測関数 g に対して

$$\mathsf{E}_\theta\big[g(T)\big] = 0 \qquad (\forall \theta \in \Theta))$$

ならば, $g(T) = 0\,(\text{a.e.}\,\mathcal{P})$ となるときをいう.

注意 5.3. 完備ならば, 有界完備は明らかである. しかし, 逆は真でないことが知られている. 反例は [6, p.25] を参照のこと. □

例 5.4. $n \in \mathbb{N}$ と $\theta > 0$ とする. $X_1, X_2, \dots, X_n$ は確率空間 $(\mathbb{X}, \mathcal{B}, \mathsf{P})$ 上の独立同一に従う確率変数列で, 共通の分布は開区間 $(0, \theta)$ 上の一様分布とする. $X_1, X_2, \dots, X_n$ の同時 p.d.f. p は

$$\mathsf{p}(x_1, x_2, \dots, x_n) = \frac{\mathbb{1}_{(0,\infty)}(\min_i x_i)\mathbb{1}_{(0,\theta)}(\max_i x_i)}{\theta^n}$$

とかける. Fisher-Neyman の因子分解定理から $T = \max\{X_1, X_2, \dots, X_n\}$ は $\mathcal{P}$ に対する十分統計量である. $X_1, X_2, \dots, X_n$ の独立性から $\forall t \in (0, \theta)$ に対して

$$\begin{aligned}\mathsf{P}^T\big[(-\infty, t]\big] &= \mathsf{Pr}\big(T \leq t\big) = \mathsf{Pr}\big(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t\big)\\ &= \prod_{i=1}^{n} \mathsf{Pr}(X_i \leq t) = \left(\frac{t}{\theta}\right)^n\end{aligned}$$

となる. これを t に関して微分することで T の p.d.f. p^T は

$$\mathsf{p}^T(t) = \begin{cases} \dfrac{nt^{n-1}}{\theta^n} & (t \in (0, \theta)) \\ 0 & (\text{その他の場合}) \end{cases}$$

となる. ある定数 c と関数 f に対して

$$\mathsf{E}_\theta\big[f(T) - c\big] = 0 \qquad (\forall \theta > 0)$$

であったとする. このとき

$$\mathsf{E}_\theta\big[f(T) - c\big] = \frac{n}{\theta^n} \int_0^\theta \big\{f(t) - c\big\} t^{n-1}\,\mathrm{d}t = 0 \qquad (\forall \theta > 0)$$

となる. これは $[0, \infty)$ 上の Lebeague 測度 μ に関して

$$f(t) - c = 0 \qquad (\text{a.e. } \mu)$$

となる. したがって $\forall \theta \in \Theta$ に対して

$$f(t) - c = 0 \qquad (\text{a.e. } \mathsf{P}_\theta)$$

となる. □

問 5.1. μ を $[0, \infty)$ 上の Lebeague 測度とする. $\forall \theta > 0$ に対して

$$\mathsf{E}_\theta\big[f(T) - c\big] = \frac{n}{\theta^n} \int_0^\theta \big\{f(t) - c\big\} t^{n-1}\,\mathrm{d}t = 0 \qquad (\forall \theta > 0)$$

ならば

$$f(t) - c = 0 \qquad (\text{a.e. } \mu)$$

を示せ.

定理 5.5. $\big(\mathbb{X}, \mathcal{B}\big)$ を標本空間とする. $\boldsymbol{T}\,(\in \mathbb{R}^k)$ は可積分な統計量とする. T が標本空間 $\big(\mathbb{X}, \mathcal{B}\big)$ 上の統計的モデル $\mathcal{P} = \{\mathsf{P}_\theta; , \theta \in \Theta\}$ に対する有限次元の有界完備十分統計量ならば, $\boldsymbol{T}$ は $\mathcal{P}$ に対する最小十分統計量である.

Proof. $\boldsymbol{U}$ を任意の十分統計量とする. $\boldsymbol{T}(X) = (T_1(X), T_2(X), \ldots, T_k(X))$ とし

$$S_i(\boldsymbol{T}) = \frac{1}{1+e^{-T_i}} \qquad (i = 1, 2, \ldots, k)$$

とする. すると $S_i\ (i = 1, 2, \ldots, k)$ は有界で $\boldsymbol{S} = (S_1, S_2, \ldots, S_k)$ と $\boldsymbol{T}$ は 1 対 1 対応となる. いま

$$G_i(\boldsymbol{u}) = \mathsf{E}\big[S_i(\boldsymbol{T})|\,\boldsymbol{U} = \boldsymbol{u}\big], \tag{5.21}$$

$$H_i(\boldsymbol{t}) = \mathsf{E}\big[G_i(\boldsymbol{U})|\,\boldsymbol{T} = \boldsymbol{t}\big] \qquad (i = 1, 2, \ldots, k) \tag{5.22}$$

とおいたとき

$$S_i(\boldsymbol{T}) = G_i(\boldsymbol{U}) \qquad (\text{a.s. } \mathsf{P}_\theta) \qquad (i = 1, 2, \ldots, k) \tag{5.23}$$

を示せばよい. すると S_i は 1 対 1 対応なので

$$\boldsymbol{T} = S_i^{-1}\big(G_i(\boldsymbol{U})\big) \qquad (\text{a.s. } \mathsf{P}_\theta) \qquad (i = 1, 2, \ldots, k)$$

となる. よって, $\boldsymbol{T}$ は最小十分統計量となることがわかる.
<u>第 1 段階</u>: まず

$$S_i(\boldsymbol{T}) = H_i(\boldsymbol{T}) \qquad (\text{a.s. } \mathsf{P}_\theta) \qquad (i = 1, 2, \ldots, k) \tag{5.24}$$

を示す. そのために, (5.21) と (5.22) から

$$\mathsf{E}\big[H_i(\boldsymbol{T})\big] = \mathsf{E}\big[\mathsf{E}\big[G_i(\boldsymbol{U})|\,\boldsymbol{T}\big]\big] = \mathsf{E}\big[G_i(\boldsymbol{U})\big] = \mathsf{E}\big[\mathsf{E}\big[S_i(\boldsymbol{T})|\,\boldsymbol{U}\big]\big] = \mathsf{E}\big[S_i(\boldsymbol{T})\big]$$

となることに注意する. したがって, すべての θ に対して

$$\mathsf{E}\big[H_i(\boldsymbol{T}) - S_i(\boldsymbol{T})\big] = 0 \qquad (i = 1, 2, \ldots, k)$$

となる. T は有界完備十分かつ H_i と S_i はともに有界なので, すべての θ に対して

$$\mathsf{Pr}\big(S_i(\boldsymbol{T}) = H_i(\boldsymbol{T})\big) = 1 \qquad (i = 1, 2, \ldots, k)$$

が成り立つ.
<u>第 2 段階</u>: 次に

$$G_i(\boldsymbol{U}) = H_i(\boldsymbol{T}) \qquad \text{a.s. } \mathsf{P}_\theta \qquad (\text{a.s. } \mathsf{P}_\theta) \qquad (i = 1, 2, \ldots, k) \tag{5.25}$$

を示す. 第 1 段階の結果 (5.24) と G_i の定義から

$$\begin{aligned} &S_i(\boldsymbol{T}) = H_i(\boldsymbol{T}) \\ &\qquad \Rightarrow\ G_i(\boldsymbol{U}) = \mathsf{E}\big[S_i(\boldsymbol{T})|\,\boldsymbol{U}\big] = \mathsf{E}\big[H_i(\boldsymbol{T})|\,\boldsymbol{U}\big] \qquad (i = 1, 2, \ldots, k) \end{aligned} \tag{5.26}$$

となる. 一方

$$\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big] = 0 \qquad (i = 1,\, 2,\, \ldots,\, k) \tag{5.27}$$

となる. なぜならば, $i = 1,\, 2,\, \ldots,\, k$ に対して

$$\begin{aligned}\mathsf{Var}\big[H_i(\boldsymbol{T})\big] &= \mathsf{E}\big[\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] + \mathsf{Var}\big[\mathsf{E}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] \qquad (\because \text{ 定理 } F.5)\\ &= \mathsf{E}\big[\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] + \mathsf{Var}\big[G_i(\boldsymbol{U})\big] \quad (\because \ (5.26))\\ &= \mathsf{E}\big[\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] + \mathsf{E}\big[\mathsf{Var}\big[G_i(\boldsymbol{U})|\boldsymbol{T}\big]\big] + \mathsf{Var}\big[\mathsf{E}\big[G_i(\boldsymbol{U})|\boldsymbol{T}\big]\big]\\ &= \mathsf{E}\big[\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] + \underbrace{\mathsf{E}\big[\mathsf{Var}\big[G_i(\boldsymbol{U})|\boldsymbol{T}\big]\big]}_{=0} + \underbrace{\mathsf{Var}\big[S_i(\boldsymbol{T})\big]}_{=\mathsf{Var}[H_i(\boldsymbol{T})]}\end{aligned} \tag{5.28}$$

となる. 最後の等号は (5.26) から

$$\mathsf{E}\big[G_i(\boldsymbol{U})|\boldsymbol{T}\big] = \mathsf{E}\big[\mathsf{E}\big[S_i(\boldsymbol{T})|\boldsymbol{U}\big]|\boldsymbol{T}\big] = S_i(\boldsymbol{T})\,(\text{a.s. } \mathsf{P}_\theta)$$

となることからわかる.

第 1 段階の結果 (5.24) から $\mathsf{Var}\big[H_i(\boldsymbol{T})\big] = \mathsf{Var}\big[S_i(\boldsymbol{T})\big]$ である. さらに, $\boldsymbol{T}$ が既知のとき, $G_i(\boldsymbol{U})= \mathsf{E}\big[S_i(\boldsymbol{T})|\boldsymbol{U}\big]$ は既知となり, $\mathsf{E}\big[\mathsf{Var}\big[G_i(\boldsymbol{U})|\boldsymbol{T}\big]\big] = 0$ となる. これら 2 つのことを (5.28) に代入すると

$$\mathsf{E}\big[\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\big] = 0 \quad (i = 1,\, 2,\, \ldots,\, k)$$

となる. $\mathsf{Var}[H_i(\boldsymbol{T})|U] \geq 0\,(\text{a.s. } \mathsf{P}_\theta)$ から

$$\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big] = 0 \qquad (\text{a.s. } \mathsf{P}_\theta) \quad (i = 1,\, 2,\, \ldots,\, k)$$

がわかる. このことから

$$\begin{aligned}&\mathsf{Var}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big] = \mathsf{E}\big[\big\{H_i(\boldsymbol{T}) - \mathsf{E}[H_i(\boldsymbol{T})|U]\big\}^2\big|U\big] = 0\,(\text{a.s. } \mathsf{P}_\theta)\\ &\quad\Leftrightarrow \quad H_i(\boldsymbol{T}) - \mathsf{E}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big] = 0 \qquad (\text{a.s. } \mathsf{P}_\theta) \quad (i = 1,\, 2,\, \ldots,\, k)\end{aligned} \tag{5.29}$$

となる. よって, (5.26) と (5.29) から

$$G_i(\boldsymbol{U}) = H_i(\boldsymbol{T}) \qquad (\text{a.s. } \mathsf{P}_\theta) \quad (i = 1,\, 2,\, \ldots,\, k)$$

がわかる. さらに

$$\begin{aligned}&G_i(\boldsymbol{U}) = H_i(\boldsymbol{T})\,(\text{a.s. } \mathsf{P}_\theta)\\ &\qquad\Rightarrow \quad G_i(\boldsymbol{U}) = \mathsf{E}\big[G_i(\boldsymbol{U})|\boldsymbol{U}\big] = \mathsf{E}\big[H_i(\boldsymbol{T})|\boldsymbol{U}\big]\,(\text{a.s. } \mathsf{P}_\theta) \quad (i = 1,\, 2,\, \ldots,\, k)\end{aligned}$$

を得る.

よって, 第 1 段階の結果 (5.24) と第 2 段階の結果 (5.24) から

$$S_i(\boldsymbol{T}) = H_i(\boldsymbol{T})\,(\text{a.s.}\,\mathsf{P}_\theta) \quad \text{かつ} \quad G_i(\boldsymbol{U}) = H_i(\boldsymbol{T})\,(\text{a.s.}\,\mathsf{P}_\theta) \quad (i = 1,\, 2,\, \ldots,\, k)$$

なので, (5.23) が示せた. □

注意 5.6 (最小十分統計量で完備でない例). $X \sim \mathsf{Unif}(\theta,\,\theta+1)\,(\theta\in\mathbb{R})$ とする. Fisher-Neyman の因子分解定理から X は θ の自明な十分統計量であることがわかる. また, $x,\,y\in\mathbb{R}$ とある $c\in\mathbb{R}$ に対して

$$\frac{\mathbb{1}_{(\theta,\,\theta+1)}(y)}{\mathbb{1}_{(\theta,\,\theta+1)}(x)} = c \Rightarrow x = y$$

となるので, X は最小性をみたす. ところが

$$\mathsf{E}_\theta[2\pi X] = \int_\theta^{\theta+1} \sin(2\pi x)\,\mathrm{d}x = 0, \qquad (\forall\theta\in\mathbb{R})$$

かつ $\sin 2\pi X \not\equiv 0$ なので, 完備ではないことがわかる. □

定理 5.7. フルランクの k 母数指数分布族

$$\mathsf{p}_\theta(x) = \exp\{(A(\theta)|\,T(x)) - \kappa(\theta)\} \qquad (\theta\in\Theta)$$

において, T は完備十分統計量となる.

Proof. $k=1$ のときを示す. T は可測空間 $(\mathbb{T},\mathcal{C})$ に値をとるとする. $k\geq 2$ の場合については, k に関する帰納法で証明できる. まず T は $\mathbb{T}$ 上のある測度 ν^T に関して p.d.f. $\mathsf{p}_\eta^T(t) = D(\eta)e^{\eta t}$ を持つことがわかる. さらに

$$\mathsf{P}_\eta(C) = \int_C D(\eta)e^{\eta t}\,\mathrm{d}\nu^T(t) \qquad (\forall C\in\mathcal{C}),$$
$$E := \left\{\eta\in\mathbb{R};\ \int_{\mathbb{R}} D(\eta)e^{\eta t}\,\mathrm{d}\nu^T(t) < \infty\right\}$$

とおく.

g は t の可測関数で $\mathsf{E}_\eta[g(T)] = 0\,(\forall\eta)$ とする. ここで $g^+(t) := \max\{g(t),\,0\}$, $g^-(t) = \max\{-g(t),\,0\}$ とおくと

$$\int_{\mathbb{T}} g^+(t)C(\eta)e^{\eta t}\,\mathrm{d}\nu^T(t) = \int_{\mathbb{T}} g^-(t)C(\eta)e^{\eta t}\,\mathrm{d}\nu^T(t) \tag{5.30}$$

となる. $\eta_0\in E^\circ$ [15]をとる. さらに

$$Z_0 := \int_{\mathbb{T}} g^+(t)D(\eta_0)e^{\eta_0 t}\,\mathrm{d}\nu^T(t) = \int_{\mathbb{T}} g^-(t)D(\eta_0)e^{\eta_0 t}\,\mathrm{d}\nu^T(t)$$

とおき, 可測空間 $\big(\mathbb{T},\mathcal{C}\big)$ 上の測度 P^+ と P^- をそれぞれ

$$\mathsf{P}^+(C) = Z_0^{-1}\int_C g^+(t)D(\eta)e^{\eta_0 t}\,\mathrm{d}\mu^T(t) \tag{5.31}$$

$$\mathsf{P}^-(C) = Z_0^{-1}\int_C g^-(t)D(\eta)e^{\eta_0 t}\,\mathrm{d}\mu^T(t) \qquad (\forall C\in\mathcal{C}) \tag{5.32}$$

[15] E° は E の内部である.

で定める．$m_{\mathsf{P}^+}(u)$ と $m_{\mathsf{P}^-}(u)$ はそれぞれ P^+ と P^- の積率母関数とすると

$$\begin{aligned} m_{\mathsf{P}^+}(\eta-\eta_0) &= Z_0^{-1}\int e^{(\eta-\eta_0)t}g^+(t)D(\eta)e^{\eta_0 t}\,\mathrm{d}\mu^T(t) \\ &= Z_0^{-1}\int e^{\eta t}g^-(t)D(\eta)\,\mathrm{d}\mu^T(t), \\ m_{\mathsf{P}^-}(\eta-\eta_0) &= Z_0^{-1}\int e^{(\eta-\eta_0)t}g(t)D(\eta)e^{\eta_0 t}\,\mathrm{d}\mu^T(t) \\ &= Z_0^{-1}\int e^{\eta t}g^-(t)D(\eta)\,\mathrm{d}\mu^T(t) \end{aligned}$$

となる．よって (5.30) から

$$m_{\mathsf{P}^+}(\eta-\eta_0) = m_{\mathsf{P}^-}(\eta-\eta_0) \qquad (\forall\eta\in E)$$

と書ける．$u=\eta-\eta_0$ とおくと P^+ と P^- のそれぞれの積率母関数である $m_{\mathsf{P}^+}(u)$ と $m_{\mathsf{P}^-}(u)$ は $u=0$ の近傍で $m_{\mathsf{P}^+}(u)=m_{\mathsf{P}^-}(u)$ となることがわかる．したがって, 積率母関数の一意性から $\mathsf{P}^+=\mathsf{P}^-$ がわかる．この結果と (5.31) と (5.32) から $g^+(t)=g^{-1}(t)\,(\text{a.e. }\mu_T)$ となる．このことから $\mathsf{Pr}(g(T)=0)=1\,(\forall\eta)$ がわかる．よって, T は完備である． □

問 5.2. T が まず T は $\mathbb{T}$ 上のある測度 ν_T に関して p.d.f. $\exp\{\eta t - C(\eta)\}$ を持つことを示せ.

定義 5.8. $\big(\mathbb{X},\mathcal{B}\big)$ を標本空間とし, $\mathcal{P}=\{\mathsf{P}_\theta;\,\theta\in\Theta\}$ を標本空間 $\big(\mathbb{X},\mathcal{B}\big)$ 上の統計的モデルとする．統計量 V が局外であるとは, その分布が θ に依存しないときをいう．したがって, V 自体は θ の情報を含まない.

注意 5.9. 統計量 V は θ の情報を含まないが, V は十分統計量の関数となるときもある．例えば, 例 5.4 では $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ は θ の十分統計量であるが, $X_{(i)}-X_{(j)}\,(1\le j<j\le n)$ は局外統計量となる．$X_{(i)}-X_{(j)}$ は最小十分統計量の関数であるので, θ の情報を含むかもしれないが, 十分統計量が完備のときには, 局外統計量は θ の情報を含まないことが次の定理からわかる.

定理 5.10 (Basu の定理)**.** T は標本空間 $(\mathbb{X},\mathcal{B},\mathsf{P})$ 上の統計的モデル $\mathcal{P}=\{\mathsf{P}_\theta;\,\theta\in\Theta\}$ に対する完備十分統計量とし, V は局外統計量とする．このとき, T と V は任意の $\mathsf{P}_\theta\in\mathcal{P}$ のもとで独立である.

Proof. 統計量 T と V の値域の部分集合 A に対して

$$\begin{aligned} \mathsf{q}_A(T) &= \mathsf{Pr}(V\in A|\,T), \qquad (5.33) \\ \mathsf{p}_A &= \mathsf{Pr}(V\in A) \end{aligned}$$

とおく．ただし $A \in V(\mathcal{B}) := \{V(B);\, B \in \mathcal{B}\}$ である．十分性と局外性から, p_A と q_A は θ に依存しない．また, 条件付き期待値の towering property(定理 F.2) から

$$
\begin{aligned}
\mathsf{p}_A &= \mathsf{Pr}(V \in A) = \mathsf{E}[\mathbb{1}_A(V)] = \mathsf{E}\big[\mathsf{E}[\mathbb{1}_A(V)|\,T]\big] = \mathsf{E}\big[\mathsf{Pr}(V \in A|\,T)\big] \\
&= \mathsf{E}[q_A(T)]
\end{aligned}
$$

となる．T の完備性から

$$
\mathsf{q}_A(T) = \mathsf{p}_A \qquad (\text{a.e.}\,\mathcal{P}) \tag{5.34}
$$

となる．再度, 条件付き期待値の towering property を用いると任意 $C \in T(\mathcal{B})$ に対して

$$
\begin{aligned}
\mathsf{Pr}(T \in C,\, V \in A) &= \mathsf{E}\big[\mathbb{1}_C(T)\mathbb{1}_A(V)\big] \\
&= \mathsf{E}\big[\mathsf{E}_\theta[\mathbb{1}_C(T)\mathbb{1}_A(V)|\,T]\big] \qquad \big(\because \text{定理 F.2(1)}\big) \\
&= \mathsf{E}\big[\mathbb{1}_C(T)\mathsf{E}_\theta[\mathbb{1}_A(V)|\,T]\big] \\
&= \mathsf{E}\big[\mathbb{1}_C(T)\mathsf{Pr}\big(V \in A|\,T\big)\big] \\
&= \mathsf{E}\big[\mathbb{1}_C(T)\mathsf{q}_A(T)\big] \qquad \big(\because (5.33)\big) \\
&= \mathsf{E}\big[\mathbb{1}_C(T)\mathsf{p}_A\big] \qquad \big(\because (5.34)\big) \\
&= \mathsf{p}_A\mathsf{E}_\theta\big[\mathbb{1}_C(T)\big] \\
&= \mathsf{Pr}(T \in C)\mathsf{Pr}(V \in A)
\end{aligned}
$$

となる．よって任意の A と C に対して

$$
\begin{aligned}
&\mathsf{Pr}(T \in C,\, V \in A) = \mathsf{Pr}(T \in C)\mathsf{Pr}(V \in A) \\
\Leftrightarrow\ &\mathsf{P}^{(C,T)}(A \times C) = \mathsf{P}^C(A)\mathsf{P}^T(C)
\end{aligned}
$$

となるので, T と V は独立となる． □

注意 5.11. $n \geq 2\,(n \in \mathbb{N})$ とし, $X_1,\, X_2,\, \ldots,\, X_n$ は独立同一に $\mathsf{N}(\mu,\, \sigma^2)$ に従うとする．$\mathcal{P}_\sigma = \{\mathsf{N}(\mu,\, \sigma^2);\, \mu \in \mathbb{R}\}$ とする．σ は固定された値である．$\overline{x} = (1/n)(x_1 + x_2 + \cdots + x_n)$ としたとき, $\boldsymbol{X} =(X_1,\, X_2,\, \ldots,\, X_n)$ の同時 p.d.f. $\mathsf{p}^{\boldsymbol{X}}$ は

$$
\begin{aligned}
\mathsf{p}^{\boldsymbol{X}}(x_1,\, x_2,\, \ldots,\, x_n) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{n\mu}{\sigma^2}\overline{x} - \frac{n\mu^2}{\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right]
\end{aligned}
$$

となる． したがって, $\mathcal{P}_\sigma$ はフルランクの指数型分布族となり, $\overline{X} = (1/n)(X_1 + X_2 + \cdots + X_n)$ は $\mathcal{P}_\sigma$ の完備十分統計量となる．いま

$$
S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2
$$

と定義する. 統計的モデル $\mathcal{P}_\sigma$ に対して, S^2 は局外統計量となる. これを確認するために, $Y_i = X_i - \mu\,(i = 1, 2, \ldots, n)$ とおくと

$$\begin{aligned}\mathsf{Pr}(Y_i \le y) &= \mathsf{Pr}(X_i \le y + \mu)\\ &= \int_{-\infty}^{y+\mu} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]\frac{dx}{\sqrt{2\pi\sigma^2}}\\ &= \int_{-\infty}^{y} \exp\left[-\frac{1}{2\sigma^2}u^2\right]\frac{du}{\sqrt{2\pi\sigma^2}}\end{aligned}$$

となる. $Y_1, Y_2, \ldots, Y_n$ は独立同一に $\mathsf{N}(0, \sigma^2)$ に従うことがわかる. いま

$$\overline{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n) = \overline{X} - \mu, \quad X_i - \overline{X} = Y_i - \overline{Y}\,(i = 1, 2, \ldots, n)$$

となるので

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

の分布は μ に依存しない. したがって, S^2 は $\mathcal{P}_\sigma$ に対する局外統計量となる. よって, Basu の定理から $\overline{X}$ と S^2 は独立となる.

6 文献についての注意

節 2 は [21, pp.123-127] を参考にした. 節 3 は [10, pp.42-45, p.102] を借用した. 節 4 は [10, pp.106-108] を参考にした. 節 5 は [10, pp.48-50] を参考にした. 定理 **??** と **??** は [8, pp.30-31] からの借用である. 補遺 A は [4] および [33, 補遺] を参考にした. 補遺 A.4 は [15] を参考にした. 補遺 B は [4] を参考にした. 補遺 C [18] を参照した. 補遺 C.1 は [30, pp.104-105] からの借用である. 積率母関数の性質は [24, pp.348-350] からの借用である. 補遺 D は [14, p.600] からの借用である. 補遺 E は [4] を参考にした.

参考文献

[1] Bisgard, J. (2021). Analysis and Linear algebra: The Singular Value Decomposition and Applications. *Student Mathematical Library* **94**. American Mathematical Society.

[2] Borwein, J.M., Lewis, A.S. (2006). ConvexAnalysis and Nonlinear Optimization Theory and Examples, 2nd. edition. Spriner.

[3] Boyd, S., Vandenberghe, L. (2018). Introduction to Applied Linear algebra. Cambridge University Press.

[4] Durrett, R. (2019). Probability, 2nd. edition. *Cambridge Series in Statistical and Probalistic Mathematics.*

[5] Gallier, J. (2001). Geometrical Methods and Applications. *Texts in Applied Mathematics* **38**. Springer.

[6] Fraser, D.A.S. (1957) Nonparametric Methods in Statistics. *A Wiley Publication in Mathematics and Statistics.*

[7] Hansen, E. (2009). Measure Theory. Fourth edtion. *Department of Mathematical Sciences, University of Copenhagen.*

[8] 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).

[9] Kaipio, J., Somerslo, E. (2005). Statistical and Computational Inverse Problems. *Applied Mathematical Sciences* **60**. Springer.

[10] Keener, R.W. (2010). Theoretical Statistics. *Springer Texts in Statistics.* Springer.

[11] Lederer, J. (2022). Fundamentals of High-Dimensional Statistics with Exercies and R Labs. *Springer Text in Statistics.* Springer.

[12] Nyblom, J. (2023+). Note on Legendre 's Method of Least Squares. To appear in *Stattistical Science.*

[13] Rigollet, P., Hütter, J.-C. (2017). High Dimensionnal Statistics. Available at `https://math.mit.edu/~rigollet/PDFs/RigNotes17.pdf`(2023/05/17 accessed).

[14] Schervish, M. J. (1997). Theory of Statistics. *Springer Series in Statistics*, Springer.

[15] Schilling, R. L. (2017). Measures, Integrals and Martingales. Second Edtion. Cambridge University Press.

[16] Stiegler, S. M. (1981). Gauss and the invention of least squares. *Annals of Statistics* **9** 465-474.

[17] 新井仁之 (2006). 線形代数　基礎と応用 (第 1 刷). 日本評論社.

[18] 笠原勇二 (2013). 明解 確率論入門 (第 1 版第 2 刷). 日本評論社.

[19] 工藤弘吉 (1957). 数理統計学 (改訂版第 1 刷) 共立出版.

[20] 久保川達也 (2017). 現代数理統計学の基礎 (初版第 1 刷) 共立出版

[21] 河野敬雄 (2013). 確率概論 (初版第 5 刷). 京都大学学術出版.

[22] 駒木文保・清智也 (2020). 確率・統計 III(初版第 5 刷). 丸善出版.

[23] 清水泰隆 (2019). 統計学への確率論, その先へ (第 2 版第 1 刷). 内田老確圃.

[24] 高橋礼二 (2019). 線型代数講義 (第 1 版第 2 刷). 日本評論社.

[25] 高橋陽一郎 (1996). 実関数と Fourier 解析１ (初版第 2 刷).

[26] 高橋陽一郎 (1998). 実関数と Fourier 解析２ (初版第 2 刷).

[27] 竹村彰通 (2020). 現代数理統計学解析入門 (初版第 1 刷).

[28] 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.

[29] 永原正章 (2020). スパースモデリング (第 1 版第 4 刷). コロナ社.

[30] 野村隆昭 (2016)(初版第 1 刷) 複素関数論講義

[31] 鍋谷清治 (1989). 数理統計学 (初版版第 3 刷). コロナ社.

[32] 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.

[33] 吉田伸生 (2023). 関数解析の基礎解析 (第 1 版第 1 刷). 裳華房.

[34] 柳田英二 (2022). 解析入門 (第 1 版第 1 刷). 裳華房.

A　補遺: 積分の定義

この節では, 積分の定義と積分の収束定理などを証明なしに述べる. 証明は [4, 7] を参考にするとよい.

A.1　積分の定義

定義 A.1. $\bigl(\mathbb{X}, \mathcal{A}\bigr)$, $\bigl(\mathbb{Y}, \mathcal{B}\bigr)$ を可測空間とし, $f : \mathbb{X} \longrightarrow \mathbb{Y}$ を関数とする. 任意の $B \in \mathcal{B}$ に対して

$$f^{-1}\bigl(B\bigr) := \bigl\{x \in \mathbb{X};\ f(x) \in B\bigr\} \in \mathcal{A}$$

のとき, f は $\mathcal{A}\diagdown\mathcal{B}$ **可測**であるといわれる. さらに, $\mathcal{A}\diagdown\mathcal{B}$ 可測関数全体の集合を $\mathcal{M}\bigl((\mathbb{X}, \mathcal{A}), (\mathbb{Y}, \mathcal{B})\bigr)$ と記す. 誤解のない場合には, 簡単に $\mathcal{M}\bigl(\mathbb{X}, \mathbb{Y}\bigr)$ とも書く. 特に, $\mathcal{M}\bigl((\mathbb{X}, \mathcal{A}), (\mathbb{R}, \mathcal{B}(\mathbb{R})\bigr)$ の元を Borel 可測関数という.

注意 A.2. 以下の事実が知られている.
(i) 連続関数は Borel 可測関数である.
(ii) f, g が Borel 可測関数のとき, 和 $f + g$, 積 fg は Borel 可測である. ただし, $f(x)+g(x) = \infty+(-\infty)$ または $(-\infty)+\infty$, $f(x)g(x) = 0\cdot(\pm\infty)$ または $f(x)g(x) = (\pm\infty)\cdot 0$ の場合を除く.
(iii) $\{f_n\}_{n=1}^{\infty}$ が Borel 可測関数列のとき

$$\sup_{n\in\mathbb{N}} f_n, \quad \inf_{n\in\mathbb{N}} f_n, \quad \limsup_{n\to\infty} f_n, \quad \liminf_{n\to\infty} f_n$$

も Borel 可測である. さらに $\lim_{n\to\infty} f_n$ が存在[16]すれば, $\lim_{n\to\infty} f_n$ も Borel 可測である. □

$E \in \mathcal{A}$ と $f \in \mathcal{M}\bigl(\mathbb{X}, \mathbb{R}\cup\{\pm\infty\}\bigr)$ に対して, 積分 $\displaystyle\int_E f\,\mathrm{d}\mu$ を以下のステップ ① ～ ③ で定める.
① f は単関数とする. すなわち, $n \in \mathbb{N}$ とし, 互いに排反な $\{A_i\}_{i=1}^{n} \subset \mathcal{A}$ で $\mu(A_i) < \infty\,(i = 1, 2, \ldots, n)$ と実数 $a_i \in \mathbb{R}\,(i = 1, 2, \ldots, n)$ があって

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x)$$

[16] $\lim_{n\to\infty} f_n$ が存在するとは

$$\mu\bigl(x \in \mathbb{X};\ \liminf_{n\to\infty} f_n(x) < \limsup_{n\to\infty} f_n(x)\bigr) = 0$$

のときをいう.

と書ける. このとき, 積分 $\displaystyle\int_E f\,\mathrm{d}\mu$ を

$$\int_E f\,\mathrm{d}\mu = \sum_{i=1}^{n} a_i\mu(A_i)$$

で定める.
② 可測関数 $f:\mathbb{X}\longrightarrow[0,\infty)\cup\{\infty\}$ の E 上積分を定める. そのために, $n\in\mathbb{N}$, $i=1,2,\ldots,n2^n$ に対し, $A_i^{(n)}\in\mathcal{A}$ を

$$\begin{aligned}
A_i^{(n)} &= \left\{x\in E;\ \frac{i-1}{2^n}\leq f(x)<\frac{i}{2^n}\right\}(i=1,2,\ldots,n2^n-1),\\
A_{n2^n}^{(n)} &= \{x\in E;\ f(x)\geq n\}
\end{aligned}$$

で定める. 積分 $\displaystyle\int_E f\,\mathrm{d}\mu$ を

$$\int_E f\,\mathrm{d}\mu = \lim_{n\to\infty}\sum_{i=1}^{n2^n}\frac{i-1}{2^n}\mu\bigl(A_i^{(n)}\bigr)$$

で定める.
③ 最後に, $f\in\mathcal{M}\bigl(\mathbb{X},\mathbb{R}\cup\{\pm\}\bigr)$ の積分 $\displaystyle\int_E f\,\mathrm{d}\mu$ を定める. そのために

$$f^+=\max\{0,f\},\quad f^-=\max\{0,-f\}$$

と書く. $\displaystyle\int_E f^+\,\mathrm{d}\mu<\infty$ または $\displaystyle\int_E f^-\,\mathrm{d}\mu<\infty$ のいずれかが成り立つとき, f の E 上の積分は確定するといい, 積分 $\displaystyle\int_E f\,\mathrm{d}\mu$ を

$$\int_E f\,\mathrm{d}\mu = \int_E f^+\,\mathrm{d}\mu - \int_E f^-\,\mathrm{d}\mu$$

で定める. 特に, $\displaystyle\int_E |f|\,\mathrm{d}\mu<\infty$ のとき, f は E 上で可積分という.

注意 A.3. $f,g\in\mathcal{M}\bigl(\mathbb{X},\mathbb{R}\cup\{\pm\infty\}\bigr)$ と $E\in\mathcal{A}$ に対して, ① ～ ③ の手続きで積分を定義すると以下のような性質が成り立つ.
(i) $E\in\mathcal{A}$ とする. $\mu(E)=0$ のとき, f は E 上可積分で $\displaystyle\int_E f\,\mathrm{d}\mu=0$ となる.
(ii) $E\in\mathcal{A}$ とし, f は A 上積分確定とする. このとき, $f\mathbb{1}_E$ は $\mathbb{X}$ 積分確定で $\displaystyle\int_E f\,\mathrm{d}\mu=\int_{\mathbb{X}} f\mathbb{1}_E\,\mathrm{d}\mu$ となる.
(iii) f はともに E 上積分確定とし, $a\in\mathbb{R}$ とする. このとき

$$\int_E af\,\mathrm{d}\mu = a\int_E f\,\mathrm{d}\mu$$

となる.
(iv) f は E 上積分確定とする. このとき

$$\left|\int_E f\,\mathrm{d}\mu\right| \leq \int_E |f|\,\mathrm{d}\mu$$

となる.
(v) f, g は E 上で積分確定で, A 上で　$f \leq g\,(\mu - \text{a.e.})$ とする. すなわち, $\mu\bigl(\{x \in E;\, f(x) > g(x)\}\bigr) = 0$ である. このとき

$$\int_E f\,\mathrm{d}\mu \leq \int_E g\,\mathrm{d}\mu$$

となる.
(vi) $\displaystyle\int_A f^+\,\mathrm{d}\mu + \int_A g^+\,\mathrm{d}\mu < \infty$ または $\displaystyle\int_A f^-\,\mathrm{d}\mu + \int_A g^-\,\mathrm{d}\mu < \infty$ とする. このとき, $f+g$ も A 上積分確定で

$$\int_A (f+g)\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu + \int_A g\,\mathrm{d}\mu \tag{A.35}$$

となる. さらに, f, g が A 上可積分のとき, $f+g$ も A 上可積分で $(A.35)$ が成立する. □

A.2　積分の収束定理

定理 A.4. $0 \leq f_n \leq f_{n+1}\,(n \in \mathbb{N})$ をみたす Borel 可測集合 $E \subset \mathbb{R}^d$ 上で定義された Borel 可測関数列 $\{f_n\}_{n=1}^\infty$ が与えられているとする. 関数 $f\colon E \longrightarrow \mathbb{R}$ を各点 $x \in E$ に対して $\displaystyle\lim_{n\to\infty} f_n(x) = f(x)$ で定めると

$$\lim_{n\to\infty} \int_E f_n(x)\,\mathrm{d}\mu(x) = \int_E f(x)\,\mathrm{d}\mu(x)$$

が成り立つ.

定理 A.5 (Fatou の補題)**.** Borel 集合 $E \subset \mathbb{R}^d$ 上で定義さらた正値 Borel 可測関数列 $\{f_n\}_{n=1}^\infty$ に対して

$$\int_E \liminf_{n\to\infty} f_n(x)\,\mathrm{d}\mu(x) \leq \liminf_{n\to\infty} \int_e f_n(x)\,\mathrm{d}\mu(x)$$

が成り立つ.

定理 A.6 (Lebeague の収束定理)**.** $E \subset \mathbb{R}^d$ を Borel 可測集合とする. $f_n\colon E \longrightarrow \mathbb{R} \cup \{\pm\infty\}\,(n = 1, 2, \ldots)$ を関数 f に各点で収束する Borel

可測関数とする. 非負値可積分関数 $g\colon E \longrightarrow [0, \infty) \cup \{+\infty\}$ が存在し, $|f_n(x)| \leq g(x)\,(\forall x \in E, \forall n \in \mathbb{N})$ をみたしていれば

$$\lim_{n\to\infty}\int_E f_n(x)\,\mathrm{d}\mu(x) = \int_E f(x)\,\mathrm{d}\mu(x)$$

となる.

系 A.7. Borel 可測集合 $E \subset \mathbb{R}^d$ 上の可積分 Borel 可測関数列 $\{f_n\}_{n=1}^{\infty}$ が $\displaystyle\sum_{n=1}^{\infty}\int_E |f_n(x)|\,\mathrm{d}\mu(x) < \infty$ をみたすとする.

$$F = \left\{x \in E;\ \textstyle\sum_{n=1}^{\infty}|f_n(x)| \text{ は収束}\right\}$$

とおいたとき

$$\mu\bigl(E \setminus F\bigr) = 0$$

で

$$\sum_{n=1}^{\infty}\int_E f_n(x)\,\mathrm{d}\mu(x) = \int_F \sum_{n=1}^{\infty} f_n(x)\,\mathrm{d}\mu(x)$$

が成り立つ.

定理 A.8. $E \subset \mathbb{R}^d$ を Borel 可測集合とする. $a, b \in \mathbb{R}(a < b)$ とし, 関数 $f\colon (a, b) \times E \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ は次の条件をみたしているとする.

(1) $x \in \mathbb{R}^d$ を止めるごとに $f(\,\cdot\,, x)$ は連続関数となる.

(2) 非負値可積分関数 $g\colon E \longrightarrow [0, \infty)\cup\{+\infty\}$ が存在して $|f(t\,\cdot)| \leq g$ となる.

このとき $t_0 \in (a, b)$ に対して

$$\lim_{t\to t_0}\int_E f(t, x)\,\mathrm{d}\mu(x) = \int_E f(t_0, x)\,\mathrm{d}\mu(x)$$

が成り立つ.

定理 A.9. $E \subset \mathbb{R}^d$ を Borel 可測集合とする. 関数 $f\colon (a, b) \times E \longrightarrow \mathbb{R}$ は次の条件をみたすとする.

(1) $\forall t \in (a, b)$ と $\forall x \in \mathbb{R}^d$ に対して $\dfrac{\partial}{\partial t}f(t,x)$ は存在する.

(2) $f(t,\ \cdot)$ は可積分である.

(3) 非負値可積分 $g\colon E \longrightarrow (0, \infty)$ が存在して

$$\left|\frac{\partial}{\partial t}f(t, x)\right| \leq g(x) \qquad (x \in E)$$

となる.

このとき
$$\frac{d}{dt}\int_E f(t,\,x)\,\mathrm{d}\mu(x) = \int_E \frac{\partial}{\partial t} f(t,\,x)\,\mathrm{d}\mu(x)$$
が成り立つ.

A.3 Fubini の定理

$(\mathbb{X},\,\mathcal{A},\,\mu)$ と $(\mathbb{Y},\,\mathcal{B},\,\nu)$ を測度空間とし,
$$\mathcal{D} := \{A\times B;\ A\in\mathcal{A},\ B\in\mathcal{B}\} \tag{A.36}$$
とする. 任意の $A\times B\in\mathcal{D}$ に対して
$$\rho(A\times B) := \mu(A)\times\nu(B) \tag{A.37}$$
と定める. ただし $0\cdot\infty = \infty\cdot 0 = 0$ と約束する. すると ρ は $\mathcal{A}\otimes\mathcal{B}$ 上の σ 加法的測度に一意的に拡張でることが知られている. ρ のことを $\mu\times\nu$ とも記すことにする.

定理 A.10. $(\mathbb{X},\,\mathcal{A},\,\mu)$ と $(\mathbb{Y},\,\mathcal{B},\,\nu)$ を σ 有限測度空間とし, $f:\mathbb{X}\times\mathbb{Y}\longrightarrow[0,\,\infty)\cup\{+\infty\}$ は $\mathcal{A}\otimes\mathcal{B}$ 可測関数または $f\in\mathcal{L}^1(\mathbb{X}\times\mathbb{Y},\,\mathcal{A}\otimes\mathcal{B},\,\mu\times\nu)$ [17]とする.
(i) さらに, ほとんど至るところの $y\in\mathbb{Y}$ に対して $\int f(x,\,y)\,\mathrm{d}\mu(x)$ は定義され, $y\mapsto\int f(x,\,y)\,\mathrm{d}\mu(x)$ は $\mathcal{B}\diagdown\mathcal{B}(\mathbb{R})$ 可測となる. また, ほとんど至るところの $x\in\mathbb{X}$ に対して $\int f(x,\,y)\,\mathrm{d}\nu(y)$ は定義され, $x\mapsto\int f(x,\,y)\,\mathrm{d}\nu(y)$ は $\mathcal{A}\diagdown\mathcal{B}(\mathbb{R})$ 可測となる.
(ii) 以下の等式が成立する.
$$\begin{aligned}\int f(x,\,y)\,\mathrm{d}(\mu\times\nu)(x,\,y) &= \iint f(x,\,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y)\\ &= \iint f(x,\,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\end{aligned}$$
が成り立つ.

[17]すなわち
$$\int |f|\,\mathrm{d}(\mu\times\nu) < \infty$$
である.

A.4 Jacobi の変換公式

定義 A.11. $U, V \subset \mathbb{R}^n$ を開集合とする．写像 $T : U \longrightarrow V$ は C^1 級微分同相であるとは次の条件をみたすときをいう．

(1) T は全単射.

(2) T と T^{-1} は連続微分可能.

点 $x \in U$ における微分

$$DT(x) := \left(\frac{\partial}{\partial x_j} T_j(x)\right)_{i,\,j=1}^{n}$$

を**ヤコビアン (Jacobian)** という．ただし $x = (x_1, x_2, \ldots, x_n), T = (T_1, T_2, \ldots, T_n)$ である．

定理 A.12 (Rudin (1976, pp.221-228))**.** $U \subset \mathbb{R}^n$ を開集合とする．写像

$$T : U \longrightarrow V = T(U)$$

が C^1 級微分同相であるための必要十分条件は次の (a), (b), (c) すべてが成立することである．

(a) $T : U \longrightarrow \mathbb{R}^n$ は単射.

(b) $T : U \longrightarrow \mathbb{R}^n$ は連続微分可能.

(c) $DT(x)$ は各 $x \in U$ で可逆.

さらに (a), (b), (c) が成り立つとき

$$D(T^{-1})(y) = (DT)^{-1}\big(T^{-1}(y)\big) \qquad (\forall y \in V)$$

が成り立つ．

定理 A.13. U, V を $\mathbb{R}^n$ の開集合とし, $T : U \longrightarrow V$ は C^1 級微分同相写像とする．$W \subset \mathbb{R}^n$ と可測空間 $(W, \mathcal{B}(W))$ 上の Lebeague 測度を λ_W を

$$\lambda_W(A) = \lambda_n(A \cap W) \qquad (A \in \mathcal{B}(\mathbb{R}^n)) \tag{A.38}$$

で定める．ただし λ_n は $\mathbb{R}^n$ 上の Lebeague 速度である．このとき

$$\lambda_V = T\big(|\det DT(\,\cdot\,)|\lambda_U\big) \tag{A.39}$$

となる．すなわち $f : V \longrightarrow [0, \infty)$ に対して

$$\int_V f(y)\,\mathrm{d}\lambda_V(y) = \int_U f\big(T(x)\big)|\det DT(x)|\,\mathrm{d}\lambda_U(x) \tag{A.40}$$

である．

A.5 Radon-Nikodym の定理

定義 A.14. $(\mathbb{X}, \mathcal{A})$ を可測空間とし, μ と ν を $(\mathbb{X}, \mathcal{A})$ 上の測度とする. $\forall A \in \mathcal{A}$ に対して

$$\mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0$$

のとき ν は μ に関して**絶対連続**であるといい, $\nu \ll \mu$ と書く.

定理 A.15 (Radon-Nikodym の定理)**.** μ と ν を可測空間 $(\mathbb{X}, \mathcal{A})$ 上の測度とし, μ は σ 有限とする. このとき可測関数 $\mathsf{p} : \mathbb{X} \longrightarrow [0, \infty) \cup \{+\infty\}$ が存在して, $\forall A \in \mathcal{A}$ に対して

$$\nu(A) = \int_A \mathsf{p}(x)\, \mathrm{d}\mu(x) \tag{A.41}$$

と書ける. さらに $g : \mathbb{X} \longrightarrow \mathbb{R}$ は ν 可積分とする. このとき

$$\int g(x)\, \mathrm{d}\nu(x) = \int g(x)\mathsf{p}(x)\, \mu(x)$$

が成立する. 関数 p を μ に関する ν の Radon-Nikodym **の微分**といい, $\mu - a.e.$ で一意的である. p のことを $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ と書く. ν が σ 有限のとき $\mu - a.e.$ で f は有限となる.

B 補遺: 期待値の定義

$\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ を確率空間とする. $\mathcal{A}$ の元のことを**事象**と呼ぶ. 任意の事象 $A \in \mathcal{A}$ に対して, 次の関数 (確率変数)

$$\mathbb{1}_A(\omega) := \begin{cases} 1 & (\omega \in A), \\ 0 & (\omega \notin A) \end{cases} \tag{B.42}$$

を事象 A の**定義関数**という.

問 B.1. 事象 $A, B \in \mathcal{A}$ に対する次の等式を確かめよ.

(i) $\mathbb{1}_{A\cap B}(\omega) = \mathbb{1}_A(\omega)\mathbb{1}_B(\omega)$.

(ii) $\mathbb{1}_{A\cup B}(\omega) = \max\big\{\mathbb{1}_A(\omega),\, \mathbb{1}_B(\omega)\big\}$.

(iii) $\mathbb{1}_{A^c}(\omega) = 1 - \mathbb{1}_A(\omega)$.

$\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 上で定義された実数値確率変数 X の期待値をつぎの ① ～ ④ の段階を踏み定義する.

① 事象 $A \in \mathcal{A}$ の指示関数 $\mathbb{1}_A$ とする．すなわち $(B.42)$ で与えられる $\{0,\,1\}$ 値確率変数である．$X = \mathbb{1}_A$ の期待値を

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

で定義する．

② $N \in \mathbb{N}$ とする．事象 $A_1,\,A_2,\,\ldots,\,A_N$ と非負の定数 $a_1,\,a_2,\,\ldots,\,a_N$ に対して, $X = \sum_{i=1}^{N} a_i \mathbb{1}_{A_i}$ とする．このとき, X の期待値を

$$\mathsf{E}[X] = \mathsf{E}\biggl[\sum_{i=1}^{N} a_i \mathbb{1}_{A_i}\biggr] := \sum_{i=1}^{N} a_i \mathsf{Pr}(A_i)$$

で定める．

③ 非負値確率変数 X に対して期待値を定める．そのために, $n = 1,\,\ldots$ と $i = 1,\,2,\,\ldots,\,n2^n$ に対して

$$A_i^{(n)} := \left\{\omega \in \Omega;\; \frac{i-1}{2^n} < X(\omega) \leq \frac{i}{2^n}\right\}, \quad a_i^{(n)} := \frac{i-1}{2^n}$$

とする．$X_n := \sum_{i=1}^{n2^n} a_i^{(n)} \mathbb{1}_{A_i}^{(n)}$ とし, X の期待値を

$$\mathsf{E}[X] := \lim_{n\to\infty} \mathsf{E}[X_n] = \lim_{n\to\infty} \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathsf{Pr}(A_i^{(n)})$$

で定める．

$$\mathsf{E}[X_n] := \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathsf{Pr}\bigl(A_i^{(n)}\bigr)$$

なので,

$$\begin{aligned}
\mathsf{E}[X_{n+1}] - \mathsf{E}[X_n] &= \sum_{i=1}^{(n+1)2^{(n+1)}} \frac{i-1}{2^{n+1}} \mathsf{Pr}\bigl(A_i^{(n+1)}\bigr) - \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathsf{Pr}\bigl(A_i^{(n)}\bigr) \\
&= \sum_{i=1}^{n2^n} \frac{1}{2^{n+1}} \mathsf{Pr}\bigl(A_{2i}^{(n+1)}\bigr) + \sum_{i=1}^{(n+1)2^{n+1}} \frac{i-1}{2^{n+1}} \mathsf{Pr}\bigl(A_i^{(n+1)}\bigr)
\end{aligned}$$

となり, $\{\mathsf{E}[X_n]\}_{n=1}^{\infty}$ は単調非減少列になる．よって $+\infty$ も含めて

$$\lim_{n\to\infty} \mathsf{E}[X_n] =: \mathsf{E}[X]$$

を定義できる．

④ 一般の確率変数 X に対して期待値を定義する．そのために

$$X^+(\omega) := \max\bigl\{X(\omega),\,0\bigr\}, \quad X^-(\omega) := \max\bigl\{X(\omega),\,0\bigr\}$$

と定める. X の期待値を

$$\mathsf{E}[X] = \begin{cases} \mathsf{E}[X^+] - \mathsf{E}[X^-] & \bigl(\mathsf{E}[X^+] < \infty \text{ かつ } \mathsf{E}[X^-] < \infty\bigr), \\ +\infty & \bigl(\mathsf{E}[X^+] = \infty \text{ かつ } \mathsf{E}[X^-] < \infty\bigr), \\ -\infty & \bigl(\mathsf{E}[X^+] < \infty \text{ かつ } \mathsf{E}[X^-] = \infty\bigr), \\ \text{定義しない} & \bigl(\mathsf{E}[X^+] = \infty \text{ かつ } \mathsf{E}[X^-] = \infty\bigr) \end{cases}$$

とする.

以上の段階を経て確率変数 X の期待値を定義することにする.

$X(\omega) = \mathbb{1}_A(\omega)$, $Y(\omega) = \mathbb{1}_B(\omega)\,(A,\,B \in \mathcal{A})$ として,

$$\mathsf{E}[\mathbb{1}_A + \mathbb{1}_B] = \mathsf{E}[\mathbb{1}_A] + \mathsf{E}[\mathbb{1}_B]$$

がわかる. したがって, $a,\,b \in \mathbb{R}$ と一般の確率変数 X と Y の期待値が有限のとき, ① から ④ のステップを経ると

$$\mathsf{E}[aX + bY] = a\mathsf{E}[X] + b\mathsf{E}[Y]$$

がわかる. また $X(\omega) \geq Y(\omega)$ のとき

$$\mathsf{E}[X] \geq \mathsf{E}[Y]$$

もわかる. さらに

$$\bigl|\mathsf{E}[X]\bigr| \leq \mathsf{E}[|X|]$$

もわかる.

注意 B.1. いろいろな性質を示すために以下の手続きをふくことを**標準機械**とよぶことにする.

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 $A_1,\,A_2,\,\ldots,\,A_n \in \mathcal{A}$ で互いに排反なものと異なる $a_1,\,a_2,\,\ldots,\,a_n$ 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 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.

④ 一般の確率変数 X を非負の部分 X^+ と負の部分 X^- にわけたものに ③ を適用し, $X = X^+ - X^-$ にたいしてその結果を拡張して, X についての性質を確認する.

□

命題 B.2. (統計家の怠け公式) 確率変数 $X : \Omega \longrightarrow \mathbb{R}$ の分布を P^X とし, Lebeague 測度 μ に関する確率密度関数 p^X を持つとする. このとき, 任意の可積分関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ に対して

$$\mathsf{E}\big[g(X)\big] = \int_{\mathbb{R}} g(x)\mathsf{p}(x)\,\mathrm{d}\mu(x) \tag{B.43}$$

が成立する.

Proof. $A \subset \mathbb{R}$ を Borel 可測集合とする. このとき

$$\int \mathbb{1}_A(x)\mathsf{p}^X(x)\,\mathrm{d}\mu(x) = \int \mathbb{1}_A(x)\,\mathrm{d}\mathsf{P}^X = \mathsf{P}^X(A) = \mathsf{Pr}(X \in A) = \mathsf{E}[\mathbb{1}_A(X)]$$

となり, g が指示関数のとき $(B.43)$ は成立する. 一般の可積分関数 g に対しても $(B.43)$ は成立することは標準機械 ② ～ ④ を用いればよい. □

C　補遺: 特性関数と積率母関数

定義 C.1. (1) $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の分布 P に対して

$$\varphi(t) = \int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\,\mathrm{d}\mathsf{P}(x) = \int_{-\infty}^{\infty} \cos(tx)\,\mathrm{d}\mathsf{P}(x) + \sqrt{-1}\int_{-\infty}^{\infty} \sin(tx)\,\mathrm{d}\mathsf{P}(x),$$
$$(-\infty < t < \infty)$$

を P の**特性関数** (characteristic function) という.
(2) また $X \sim \mathsf{P}$ のとき, X の特性関数 φ^X を

$$\varphi^X(t) = \mathsf{E}\big[e^{\sqrt{-1}tX}\big], \qquad (-\infty < t < \infty)$$

と書く.

注意 C.2. (1) 「確率変数の特性関数」と「分布の特性関数」は実質的に同じであるので, 主として確率変数の特性関数の場合のみを書く.
(2) また $\left|e^{\sqrt{-1}tx}\right| \leq 1$ なので, 特性関数は常に存在する.
(3) この積分の計算の仕方は, X の p.m.f. または p.d.f. を $\mathsf{p}(x)$ と書く. すると離散型の場合は

$$\varphi^X(t) = \sum_x e^{\sqrt{-1}tx}\mathsf{p}(x), \qquad (-\infty < t < \infty)$$

となり, 連続の場合は

$$\varphi^X(t) = \int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\mathsf{p}(x)\,\mathrm{d}x, \qquad (-\infty < t < \infty)$$

となる.

定理 C.3. 確率変数 X の特性関数 φ^X は次の性質をみたす.
(1) $\left|\varphi^X(t)\right| \le 1$, $\varphi^X(0) = 1$.
(2) $\varphi^X(-t) = \overline{\varphi^X(t)}$.
(3) $\varphi^X(t)$ は t について連続.

Proof. (1), (2) は定義と複素数の性質からわかる. (3) のみ証明を書いておく. $\epsilon \neq 0$ に対して

$$\left|e^{\sqrt{-1}(t+\epsilon)x} - e^{\sqrt{-1}tx}\right| = \left|e^{\sqrt{-1}tx}\right| \cdot \left|e^{\sqrt{-1}\epsilon x} - 1\right| \le 2$$

となるので, Lebeague の優収束定理より

$$\begin{aligned}\lim_{\epsilon \to 0}\bigl\{\varphi^X(t+\epsilon) - \varphi^X(t)\bigr\} &= \lim_{\epsilon \to 0}\left\{\int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\bigl\{e^{\sqrt{-1}\epsilon} - 1\bigr\}\,\mathsf{dP}(x)\right\} \\ &= \int_{-\infty}^{\infty} e^{\sqrt{-1}tx}\lim_{\epsilon \to 0}\bigl\{e^{\sqrt{-1}\epsilon} - 1\bigr\}\,\mathsf{dP}(x)\Bigr\} \\ &= 0\end{aligned}$$

となるので, $\varphi^X(t)$ の t における連続性は示せた. □

定理 C.4. $a\,(\neq 0)$, b は定数とし, 確率変数 X と $aX+b$ の特性関数をそれぞれ $\varphi^X(t)$, $\varphi_{aX+b}(t)$ とおくと次の関係が成り立つ.

$$\varphi_{aX+b}(t) = e^{\sqrt{-1}tb}\varphi^X(at).$$

Proof. 特性関数の定義からわかる. □

注意 C.5. (1) $X \sim \mathsf{Bino}(n,\,p)\,(n \in \mathbb{N},\,0 \le p \le 1)$ のとき

$$\varphi^X(t) = \bigl(e\,\sqrt{-1}tp + (1-p)\bigr)^n, \qquad (-\infty < t < \infty).$$

(2) $X \sim \mathsf{Pois}(\lambda)\,(\lambda > 0)$ のとき

$$\varphi^X(t) = \exp\bigl\{\lambda(e^{\sqrt{-1}t} - 1)\bigr\}, \qquad (-\infty < t < \infty).$$

(3) $X \sim \mathsf{Exp}(n,\,p)\,(n \in \mathbb{N},\,0 \le p \le 1)$ のとき

$$\varphi^X(t) = \frac{\lambda}{\lambda - \sqrt{-1}t}, \qquad (-\infty < t < \infty).$$

(4) $X \sim \mathsf{N}(\mu,\,\sigma^2)\,(\mu \in \mathbb{R},\,0 < \sigma < \infty)$ のとき

$$\varphi^X(t) = \exp\left\{\sqrt{-1}\mu t - \frac{1}{2}\sigma^2 t^2\right\}, \qquad (-\infty < t < \infty).$$

□

定理 C.6. 確率空間 $(\Omega, \mathcal{F}, \mathsf{Pr})$ 上の確率変数 X の特性関数を φ^X とする．$\mathsf{Pr}(X=a)=\mathsf{Pr}(X=b)=0$ なる $a,\, b\,(a<b)$ について次が成りたつ.

$$\mathsf{Pr}\big(a<X\leq b\big)=\mathsf{P}^X((a,\,b])=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\varphi^X(t)\,\mathrm{d}t.$$

ただし P^X は X の分布である.

Proof. 補題 $C.15$ から

$$\int_0^\infty \frac{\sin xt}{t}\,\mathrm{d}t=\begin{cases}\dfrac{\pi}{2} & (x>0),\\ 0 & (x=0),\\ -\dfrac{\pi}{2} & (x<0)\end{cases} \tag{C.44}$$

となる．X の分布は P^X とおく．φ^X の定義に注意すると $T>0$ と $a<b$ に対して

$$\begin{aligned}
&\int_{-T}^{T}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}\varphi^X(t)\,\mathrm{d}t\\
&=\int_{-T}^{T}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}\left\{\int_{-\infty}^{\infty}e^{\sqrt{-1}tx}\,\mathrm{d}\mathsf{P}^X(x)\right\}\mathrm{d}t\\
&=\int_{-T}^{T}\left\{\int_{-\infty}^{\infty}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}e^{\sqrt{-1}tx}\,\mathrm{d}\mathsf{P}^X(x)\right\}\mathrm{d}t\\
&=\int_{-\infty}^{\infty}\left\{\int_{-T}^{T}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}e^{\sqrt{-1}tx}\,\mathrm{d}t\right\}\mathrm{d}\mathsf{P}^X(x) \qquad (\because \text{Fubini(定理 } A.10\text{)))}\\
&=\int_{-T}^{T}\left\{\int_{-\infty}^{\infty}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}e^{\sqrt{-1}tx}\,\mathrm{d}t\right\}\mathrm{d}\mathsf{P}^X(x)\\
&=\int_{-\infty}^{\infty}\left\{\int_{-T}^{T}\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}e^{\sqrt{-1}tx}\,\mathrm{d}t\right\}\mathrm{d}\mathsf{P}^X(x)\\
&=\int_{-\infty}^{\infty}\left\{\int_{-T}^{T}\frac{e^{\sqrt{-1}t(x-a)}-e^{\sqrt{-1}t(x-b)}}{\sqrt{-1}t}\,\mathrm{d}t\right\}\mathrm{d}\mathsf{P}^X(x)
\end{aligned} \tag{C.45}$$

を得る．Euler の公式から $y\in\mathbb{R}$ に対して

$$\begin{aligned}
\int_{-T}^{T}\frac{e^{\sqrt{-1}ty}}{\sqrt{-1}t}\,\mathrm{d}t&=\int_{-T}^{T}\frac{\cos ty+\sqrt{-1}\sin ty}{\sqrt{-1}t}\,\mathrm{d}t\\
&=\int_{-T}^{T}\frac{\cos ty}{\sqrt{-1}t}\,\mathrm{d}t+\sqrt{-1}\int_{-T}^{T}\frac{\sin ty}{\sqrt{-1}t}\,\mathrm{d}t=2\int_0^T\frac{\sin ty}{t}\,\mathrm{d}t
\end{aligned}$$

であるので

$$\int_{-T}^{T}\frac{e^{\sqrt{-1}t(x-a)}-e^{\sqrt{-1}t(x-b)}}{\sqrt{-1}t}\,\mathrm{d}t=2\int_0^T\left\{\frac{\sin(x-a)t}{t}-\frac{\sin(x-b)t}{t}\right\}\mathrm{d}t \tag{C.46}$$

となる. $(C.46)$ を $(C.45)$ に代入すると

$$
\begin{aligned}
&\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\varphi^X(t)\,\mathrm{d}t\\
&\quad=2\lim_{T\to\infty}\int_{-\infty}^{\infty}\left\{\int_0^T\frac{\sin(x-a)t}{t}\,\mathrm{d}t-\int_0^T\frac{\sin(x-b)t}{t}\,\mathrm{d}t\right\}\mathsf{P}^X(x)\\
&\quad=2\lim_{T\to\infty}\int_{-\infty}^{\infty}\left\{\int_0^\infty\frac{\sin(x-a)t}{t}\,\mathrm{d}t-\int_0^\infty\frac{\sin(x-b)t}{t}\,\mathrm{d}t\right\}\mathsf{P}^X(x)
\end{aligned}
$$

となる. 最後は

$$
\left|\frac{e^{-\sqrt{-1}at}-e^{-\sqrt{-1}bt}}{\sqrt{-1}t}\varphi^X(t)\right|
$$

は有界であるので, Lebeague の収束定理よりわかる. ここで

$$
f(x):=\int_0^\infty\frac{\sin(x-a)t}{t}\,\mathrm{d}t-\int_0^\infty\frac{\sin(x-b)t}{t}\,\mathrm{d}t
$$

とおくと $(C.44)$ から

$$
f(x)=\begin{cases}\pi & (a<x<b),\\ \dfrac{\pi}{2} & (x=a,\,b),\\ 0 & (x<a\ \text{または}\ x>b)\end{cases}
$$

がわかる. よって

$$
\begin{aligned}
\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-\sqrt{-1}ta}-e^{-\sqrt{-1}tb}}{\sqrt{-1}t}\varphi^X(t)\,\mathrm{d}t&=2\int_{-\infty}^{\infty}f(x)\,\mathrm{d}\mathsf{P}(x)\\
&=2\pi\mathsf{P}\big((a,\,b)\big)+\mathsf{P}\big(\{a\}\cup\{b\}\big)
\end{aligned}
$$

がわかる. □

系 C.7. 分布はその特性関数から一意的に定まる.

Proof. 定理 $C.6$ からわかる. □

定理 C.8. (1) $\mathsf{E}\big[|X|\big]<\infty$ のとき $\varphi^X(t)$ は連続微分可能であり

$$
\dot{\varphi}^X(0)=\frac{d\varphi^X}{dt}(t)\bigg|_{t=0}=\sqrt{-1}\mathsf{E}\big[X\big]
$$

となる.
(2) $\mathsf{E}\big[|X^2|\big]<\infty$ のとき $\varphi^X(t)$ は 2 回連続微分可能であり

$$
\ddot{\varphi}^X(0)=\frac{d^2\varphi^X}{dt^2}(t)\bigg|_{t=0}=-\mathsf{E}\big[X^2\big]
$$

となる.
(3) したがって

$$\mathsf{E}\big[X\big] = -\sqrt{-1}\varphi^X(0),\ \mathsf{Var}\big[X\big] = -\ddot{\varphi}^X(0) + \big\{\dot{\varphi}^X(0)\big\}^2$$

となる.

Proof. □

定義 C.9. 分布 P について

$$\mathsf{m}^{\mathsf{P}}(t) = \int_{-\infty}^{\infty} e^{tx}\,\mathrm{d}\mathsf{P}(x)$$

が原点の近傍 $|t| < h\,(h > 0)$ で収束するとき, $\mathsf{m}^{\mathsf{P}}(t)$ を P の**積率母関数**という. また確率変数 X の積率母関数とは

$$\mathsf{m}^X(t) = \mathsf{E}\big[e^{tX}\big] = \int_{-\infty}^{\infty} e^{tx}\,d\mathsf{P}^X(x)$$

のことである.

定理 C.10. 確率変数 X と Y の分布関数と積率母関数をそれぞれF^X, F^Y と m^X, m^Y とする. このとき, ある $\delta > 0$ が存在して

$$\mathsf{m}^X(t) = \mathsf{m}^Y(t) \quad (|t| < \delta) \quad \Rightarrow \quad \mathsf{F}^X(t) = \mathsf{F}^Y(t) \quad (\forall t \in \mathbb{R})$$

となる.

Proof. X の特性関数は $\varphi^X(\xi) = \mathsf{E}\big[e^{\sqrt{-1}\xi X}\big]$ であった. したがって, 特性関数は積率母関数 $\mathsf{m}^X(t) = \mathsf{E}\big[e^{tX}\big]$ において t を $\sqrt{-1}\xi$ を代入したものであり, 両者の関係は

$$\varphi^X(\xi) = \mathsf{m}^X(\sqrt{-1}\xi)$$

となる. このことから, 積率母関数が与えられると特性関数が定まり, 分布も一意的に定める. □

注意 C.11.

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad (-\infty < x < \infty)$$

なので

$$\mathsf{m}^X(t) = \mathsf{E}\big[e^{tX}\big] = \mathsf{E}\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{\mathsf{E}[X^k]}{k!} t^k$$

となる. すなわち, 積率母関数は「その係数が積率 $\mathsf{E}[X^k]$ を与える整級数」である. □

定義 C.12. $[0, \infty)$ 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 C.13. 区間 $[0, 1]$ に集中した分布の分布関数 G_1, G_2 に対して

$$\int_0^1 x^n\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 x^n\,\mathrm{d}\mathsf{G}_2(x) \qquad (n = 0, 1, 2, \ldots) \tag{C.47}$$

が成り立てば

$$\mathsf{G}_1 = \mathsf{G}_2$$

である.

Proof. Weierstrass の多項式近似定理により, $(C.47)$ が成り立てば, 任意の $[0, 1]$ 上の連続関数 f に対して

$$\int_0^1 f(x)\,\mathrm{d}\mathsf{G}_1(x) = \int_0^1 f(x)\,\mathrm{d}\mathsf{G}_2(x) \qquad (n = 0, 1, 2, \ldots)$$

が成り立つ.

とくに $0 \leq a < 1$ として, $n \geq 1$ に対して

$$f_n(x) = \begin{cases} 1 & (0 \leq x \leq a) \\ n(x-a) & (a \leq x \leq a + 1/n) \\ 0 & (n + 1/n \leq x \leq 1) \end{cases}$$

と定めると, $\mathsf{G}_1(0)$= $\mathsf{G}_2(0) = 0$ より

$$\begin{aligned} G_1(a) &= \int_0^a \mathrm{d}\mathsf{G}_1 = \lim_{n\to\infty} \int_0^1 f_n(x)\,\mathrm{d}\mathsf{G}_1(x) = \lim_{n\to\infty} \int_0^1 f_n(x)\,\mathrm{d}\mathsf{G}_2(x) \\ &= \int_0^a \mathrm{d}\mathsf{G}_2 = G_2(a) \end{aligned}$$

を得る．また $\mathsf{G}_1(1) = \mathsf{G}_2(1)$ だから, すてべの $a \in [0, 1]$ に対して $\mathsf{G}_1(a) = \mathsf{G}_2(a)$ がわかる. □

定理 C.14. $[0, \infty)$ に集中した確率をもつ 2 つの分布関数 F_1, 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. $\mathsf{F}\colon [0, \infty) \longrightarrow \mathbb{R}$ を確率分布関数とする. $x = e^{-t}$, $\mathsf{G}(x) = 1 - \mathsf{F}(t)$ とすると

$$\int_0^\infty e^{-\lambda t}\,\mathrm{d}\mathsf{F}(t) = \int_0^1 x^\lambda\,\mathrm{d}\mathsf{G}(x) \qquad (\lambda \geq 0)$$

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 C.13 を用いれば, 定理は証明される. □

C.1 積分の補題

補題 C.15.

$$\int_0^\infty \frac{\sin t}{t}\,\mathrm{d}t = \frac{\pi}{2}.$$

Proof. $R > 0$ とし,

$$f(t) := \frac{e^{\sqrt{-1}t} - 1}{t}$$

とおく. Cauchy の積分定理から

$$\begin{aligned}
&\int_{-R}^R f(t)\,\mathrm{d}t + \sqrt{-1}R\int_0^\pi f\bigl(Re^{\sqrt{-1}\theta}\bigr)e^{\sqrt{-1}\theta}\,\mathrm{d}\theta \\
&= \int_{-R}^R \frac{e^{\sqrt{-1}t} - 1}{t}\,\mathrm{d}t + \sqrt{-1}R\int_0^\pi \frac{e^{Re^{\sqrt{-1}\theta}} - 1}{Re^{\sqrt{-1}\theta}}e^{\sqrt{-1}\theta}\,\mathrm{d}\theta \\
&= \int_{-R}^R \frac{e^{\sqrt{-1}t} - 1}{t}\,\mathrm{d}t + \sqrt{-1}\int_0^\pi \Bigl\{e^{Re^{\sqrt{-1}\theta}} - 1\Bigr\}\,\mathrm{d}\theta \\
&= \int_{-R}^R \frac{e^{\sqrt{-1}t} - 1}{t}\,\mathrm{d}t + \sqrt{-1}\int_0^\pi e^{Re^{\sqrt{-1}\theta}}\,\mathrm{d}t - \sqrt{-1}\pi \\
&= 0
\end{aligned}$$

から

$$\int_{-R}^R \frac{e^{\sqrt{-1}t} - 1}{t}\,\mathrm{d}t - \sqrt{-1}\pi = -\sqrt{-1}\int_0^\pi e^{Re^{\sqrt{-1}\theta}}\,\mathrm{d}t \tag{C.48}$$

を得る. Euler の公式を適用すると

$$\begin{aligned}
\int_{-R}^R \frac{e^{\sqrt{-1}t} - 1}{t}\,\mathrm{d}t &= \int_{-R}^R \frac{\cos t + \sqrt{-1}\sin t - 1}{t}\,\mathrm{d}t \\
&= \sqrt{-1}\int_{-R}^R \frac{\sin t}{t}\,\mathrm{d}t \qquad \text{(C.49)}
\end{aligned}$$

と変形できる. $(C.48)$ と $(C.49)$ を合わせると

$$\int_{-R}^{R} \frac{\sin t}{t}\,\mathrm{d}t - \pi = -\int_0^{\pi} e^{\sqrt{-1}Re^{\sqrt{-1}\theta}}\,\mathrm{d}\theta$$

を得る. $t \mapsto \dfrac{\sin t}{t}$ は偶関数であることと Euler の公式に注意して

$$\begin{aligned}
\left|\int_0^R \frac{\sin t}{t}\,\mathrm{d}t - \frac{\pi}{2}\right| &\leq \frac{1}{2}\int_0^{\pi} \left|e^{\sqrt{-1}Re^{\sqrt{-1}\theta}}\right|\mathrm{d}\theta = \frac{1}{2}\int_0^{\pi} \left|e^{\sqrt{-1}R\left\{\cos\theta + \sqrt{-1}\sin\theta\right\}}\right|\mathrm{d}\theta \\
&= \frac{1}{2}\int_0^{\pi} \left|e^{\sqrt{-1}R\cos\theta}\right| e^{-R\sin\theta}\,\mathrm{d}\theta = \frac{1}{2}\int_0^{\pi} e^{-R\sin\theta}\,\mathrm{d}\theta \\
&= \int_0^{\pi/2} e^{-R\sin\theta}\,\mathrm{d}\theta \leq \int_0^{\pi/2} e^{-R\theta/\pi}\,\mathrm{d}\theta \\
&= \left[-\frac{\pi}{2R}e^{-2R\theta/\pi}\right]_0^{\pi/2} = \frac{\pi}{2R}.
\end{aligned}$$

がわかる. 最後の不等式は $0 \leq \theta \leq \pi/2$ に対して, $\sin\theta \geq \dfrac{2}{\pi}\theta$ であることからわかる.

□

D 補遺: Radon-Nikodym の定理の再訪問

定義 D.1. $(\mathbb{X}, \mathcal{B})$ を可測空間とし, μ と ν を $(\mathbb{X}, \mathcal{B})$ 上の測度とする. $\forall B \in \mathcal{B}$ に対して

$$\mu(B) = 0 \quad \Rightarrow \quad \nu(B) = 0$$

のとき ν は μ に関して**絶対連続**であるといい, $\nu \ll \mu$ と書く.

定理 D.2 (Radon-Nikodym の定理)**.** μ と ν を可測空間 $(\mathbb{X}, \mathcal{B})$ 上の測度とし, μ は σ 有限とする. $\nu \ll \mu$ のとき可測関数 $\mathsf{p} : \mathbb{X} \longrightarrow [0, \infty) \cup \{+\infty\}$ が存在して, $\forall B \in \mathcal{B}$ に対して

$$\nu(B) = \int_B \mathsf{p}(x)\,\mathrm{d}\mu(x) \tag{D.50}$$

と書ける. さらに $g : \mathbb{X} \longrightarrow \mathbb{R}$ は ν 可積分とする. このとき

$$\int g(x)\,\mathrm{d}\nu(x) = \int g(x)\mathsf{p}(x)\,\mu(x)$$

が成立する. 関数 f を μ に関する ν の Radon-Nikodym **の微分**といい, μ a.e. で一意的である. p のことを $\dfrac{d\nu}{d\mu}(x)$ と書く. ν が σ 有限のとき μ a.e. で f は有限となる.

補題 D.3. μ を可測空間 $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度とし, $\mathcal{P}$ を $(\mathbb{X}, \mathcal{B})$ 上の測度の集合で, $\forall \mathsf{P} \in \mathcal{P}$ に対して, $\mathsf{P} \ll \mu$ 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 $\sum_{n=1}^{\infty} c_n = 1$ なるものと $\{\mathsf{P}_n\}_{n=1}^{\infty} \subset \mathcal{P}$ が存在して $\forall \mathsf{P} \in \mathcal{P}$ に対して $\mathsf{P} \ll \sum_{n=1}^{\infty} c_n \mathsf{P}_n$ となる.

Proof. $\mathcal{P}$ が可算集合ならば定理の主張は自明であるので, $\mathcal{P}$ 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. $\forall B \in \mathcal{B}$ 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 $\mathsf{P} \ll \lambda \, (\forall \mathsf{P} \in \mathcal{N})$ となる[18]. ここで

$$\mathcal{Q} := \left\{ \sum_{n=1}^{\infty} c_n \mathsf{P}_n ;\ \sum_{n=1}^{\infty} c_n = 1, c_n \in \mathbb{R}, \text{ かつ } \{\mathsf{P}_n\}_{n=1}^{\infty} \subset \mathcal{P} \right\}$$

とする. 明らかに

$$\beta \in \mathcal{Q} \ \Rightarrow\ \beta \ll \lambda$$

である.

つぎに

$$\mathcal{D} := \left\{ C \in \mathcal{A};\ \exists \mathsf{Q} \in \mathcal{Q} \text{ s.t. } \lambda\left(\left\{ x \in C;\ \frac{d\mathsf{Q}}{d\lambda}(x) = 0 \right\} \right) = 0 \text{ かつ } \mathsf{Q}(C) > 0 \right\}$$

とする. まず $\mathcal{D} \neq \varnothing$ を確認する. そのために $\mathsf{P} \in \mathcal{P}$ で $\mathsf{P} \neq 0$ なるものを取り

$$C := \left\{ x \in \mathbb{X};\ \frac{d\mathsf{P}}{d\lambda}(x) > 0 \right\}$$

[18] $\lambda(B) = 0$ なる $B \in \mathcal{B}$ をとる. すると $\mu(B \cap \mathbb{X}_n) = 0 \, (\forall n \in \mathbb{N})$ となる. したがって

$$\mu(B) = \sum_{n=1}^{\infty} \mu(B \cap \mathbb{X}_n) = 0$$

となる. $\mathsf{P} \ll \mu$ から $\mathsf{P}(B) = 0$ となる. よって $\mathsf{P} \ll \lambda$ がわかる.

とする．$\mathsf{Q} = \mathsf{P}$ としたとき

$$\left\{x \in C;\ \frac{d\mathsf{P}}{d\lambda}(x) = 0\right\} = \varnothing$$

となり

$$\mathsf{Q}(C) = \mathsf{P}(C) = \mu(\mathbb{X}) > 0$$

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

$$\mathsf{Q}_n(C_n) > 0 \text{ かつ } \lambda\left(\left\{x \in C_n;\ \frac{d\mathsf{Q}_n}{d\lambda}(x) = 0\right\}\right) = 0$$

なるもの[19]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

$$\frac{d\mathsf{Q}_0}{d\lambda} = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d\mathsf{Q}_n}{d\lambda}$$

かつ

$$\left\{x \in C_0;\ \frac{d\mathsf{Q}_0}{d\lambda}(x) = 0\right\} \subset \bigcup_{n=1}^{\infty} \left\{x \in C_n;\ \frac{d\mathsf{Q}_n}{d\lambda}(x) = 0\right\}$$

となる．このことより

$$\lambda\left(\left\{x \in C_0;\ \frac{d\mathsf{Q}_0}{d\lambda}(x) = 0\right\}\right) \leq \sum_{n=1}^{\infty} \lambda\left(\left\{x \in C_n;\ \frac{d\mathsf{Q}_n}{d\lambda}(x) = 0\right\}\right)$$

[19] $C_n \mathcal{D}$ なのでこのような $Q_n \in \mathcal{Q}$ が取れる．

となるで, $C_0 \in \mathcal{D}$ がわかる. さらに

$$\lambda(C_0) = \lambda\left(\bigcup_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} \lambda(C_n) = c$$

がわかる.

あとは $\mathsf{Q}_0 \in \mathcal{Q}$ なので $\forall \mathsf{P} \in \mathcal{P}$ に対して $\mathsf{P} \ll \mathsf{Q}_0$ を示せばよい. そのために $B \in \mathcal{B}$ で $\mathsf{Q}_0(B) = 0$ なるものと $\forall \mathsf{P} \in \mathcal{P}$ を取る. $x \in C_0$ に対して

$$\mathsf{Q}_0(B \cap C_0) \le \mathsf{Q}_0(B) = 0 \text{ かつ } \frac{d\mathsf{Q}_0}{d\lambda}(x) > 0 \, (\forall x \in C_0)$$

となる[20]ので

$$\lambda(B \cap C_0) = 0 \Rightarrow \mathsf{P}(B \cap C_0) = 0 \qquad (\because \mathsf{P} \ll \lambda)$$

となる. ここで

$$C := \left\{x \in \mathbb{X};\, \frac{d\mathsf{P}}{d\lambda}(x) > 0\right\}$$

とおく. すると

$$x \in C^c \ \Rightarrow\ \frac{d\mathsf{P}}{d\lambda}(x) = 0$$

なので

$$\mathsf{P}(B \cap C_0^c \cap C^c) \le \int_{C^c} \frac{d\mathsf{P}}{d\lambda}(x)\, \mathrm{d}\lambda(x) = 0$$

となる. さらに $D := B \cap C_0^c \cap C$ 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

$$\mathsf{Q}_0\left(\left\{x \in D;\, \frac{dQ_0}{d\lambda}(x) = 0\right\}\right) = 0$$

---

[20] $x \in C_0$ のときある $n \in \mathbb{N}$ があって $x \in C_n$ となる. すると

$$\lambda\left(\left\{x \in C_n;\, \frac{d\mathsf{Q}_n}{d\lambda}(x) = 0\right\}\right) = 0$$

なので

$$\frac{d\mathsf{Q}_0}{d\lambda}(x) > 0$$

となる.

となる[21]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

$$\lambda(C_0 \cup D) > \lambda(C_0) = c$$

となるので, c の定義[22]と矛盾する. しがたって

$$\lambda(D) = 0 \text{ かつ } \mathsf{P}(D) = 0$$

となる. このことより

$$\begin{aligned}
\mathsf{P}(B) &= \mathsf{P}(B \cap C_0) + \mathsf{P}(B \cap C_0^c) \\
&= \mathsf{P}(B \cap C_0) + \mathsf{P}(B \cap C_0^c \cap C^c) + \mathsf{P}(B \cap C_0^c \cap C) \\
&= \mathsf{P}(B \cap C_0) + \mathsf{P}(B \cap C_0^c \cap C^c) + \mathsf{P}(D) = 0
\end{aligned}$$

がわかる. よって定理が証明された. □

E 補遺: 条件付き期待値

$(\Omega, \mathcal{A}, \mathsf{Pr})$ 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 E.1. $\mathcal{F}$ 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

$$\int_A X(\omega)\, \mathrm{d}\mathsf{Pr}(\omega) = \int_A Y(\omega)\, \mathsf{Pr}(\omega).$$

まず条件付き期待値の存在と一意性を確認する.

補題 E.2. Y を定義 $E.1$ 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

$$\int_A Y(\omega)\, \mathrm{d}\mathsf{Pr}(\omega) = \int_A X(\omega)\, \mathrm{d}\mathsf{Pr}(\omega) \leq \int_A |X(\omega)|\, \mathrm{d}\mathsf{Pr}(\omega),$$
$$\int_{A^c} \big(-Y(\omega)\big)\, \mathrm{d}\mathsf{Pr}(\omega) = \int_{A^c} \big(-X(\omega)\big)\, \mathrm{d}\mathsf{Pr}(\omega) \leq \int_{A^c} |X(\omega)|\, \mathrm{d}\mathsf{Pr}(\omega).$$

[21]

$$\mathsf{Q}_0 \in \mathcal{Q} \;\Rightarrow\; \mathsf{Q}_0 \ll \lambda$$

であることからわかる.

[22] $c = \sup_{C \in \mathcal{D}} \lambda(C)$ である.

よって

$$\begin{aligned}\int |Y(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} \bigl(-Y(\omega)\bigr)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &\le \int_A |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega) + \int_{A^c} |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int |X(\omega)|\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

からわかる. □

補題 E.3. Y は一意.

Proof. Y' も定義 E.1 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, $A := \{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \ge \epsilon\}$ とする. $A \in \mathcal{F}$ であることから (ii) から

$$\begin{aligned}0 &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \bigl\{Y(\omega) - Y'(\omega)\bigr\}\,\mathrm{d}\mathsf{Pr}(\omega)\\ &\ge \epsilon \int_A \mathrm{d}\mathsf{Pr}(\omega) = \epsilon \mathsf{Pr}(A)\end{aligned}$$

となり, $\mathsf{Pr}(A) = 0$ がわかる.

$$\Bigl\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\Bigr\} = \bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \ge \frac{1}{n}\right\}$$

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

$$\begin{aligned}\mathsf{Pr}\Bigl(\Bigl\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) > 0\Bigr\}\Bigr) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} \left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \ge \frac{1}{n}\right\}\right)\\ &= \lim_{n\to\infty} \mathsf{Pr}\left(\left\{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \ge \frac{1}{n}\right\}\right)\\ &= 0\end{aligned}$$

がわかる. よって

$$\mathsf{Pr}\bigl(\omega \in \Omega;\, Y(\omega) \le Y'(\omega)\bigr) = 1$$

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 E.4. 定義 $E.1$ 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 E.5. $\mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ 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 $\dfrac{d\nu}{d\mu}$ が存在し

$$\begin{aligned}\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mu(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \nu(A) \\ &= \int_A \frac{d\nu}{d\mu}(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F}) \end{aligned} \tag{E.51}$$

となる. $A = \Omega$ とすると $\dfrac{d\nu}{d\mu} \geq 0$ は可積となる. Radon-Nikodym の一意性から

$$\mathsf{E}\big[X|\,\mathcal{F}\big](\omega) = \frac{d\nu}{d\mu}(\omega)$$

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 $Y_2(\omega) = \mathsf{E}\big[X^{-1}|\,\mathcal{F}\big](\omega)$ 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

$$\begin{aligned}\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X^+(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X^-(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A Y_1(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y_2(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because \text{(E.51)}\big) \\ &= \int_A \Big\{Y_1(\omega) - Y_2(\omega)\Big\}\,\mathrm{d}\mathsf{Pr}(\omega) \\ &= \int_A \big(Y_1 - Y_2\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

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 E.6. 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) $A,\, B \in \mathcal{A}$ 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

$$\sigma(Y) \colon = \sigma\Big[\big\{\omega \in \Omega;\, Y(\omega) \le r\,(\forall r \in \mathbb{R})\big\}\Big]$$

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. $p(x,\, y)$ 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

$$\int p(x,\, y)\,\mathrm{d}x > 0$$

とする．

いま関数 $g:\mathbb{R}\longrightarrow\mathbb{R}$ で $\mathsf{E}\big[|g(X)|\big]<\infty$ なるものと考える．このとき

$$\mathsf{E}\big[g(X)|\,Y\big](\omega)=h(Y(\omega)),\qquad h(y)=\frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x}$$

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(定理 A.10)) の定理から

$$\begin{aligned}
\mathsf{E}\big[h(Y)\mathbb{1}_A\big]&=\mathsf{E}\big[h(Y)\mathbb{1}_B(Y)\big]\qquad\big(\because Y\in B\Leftrightarrow\omega\in A\big)\\
&=\int_B\int h(y)p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y\\
&=\int_B\int\left\{\frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x}\right\}p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y\qquad\big(\because h\text{ の定義を代入}\big)\\
&=\int_B\left\{\int\frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x}p(x,\,y)\,\mathrm{d}x\right\}\mathrm{d}y\qquad\big(\because \text{Fubini の定理 (定理 A.10)}\big)\\
&=\int_B\frac{\int g(x)p(x,\,y)\,\mathrm{d}x}{\int p(x,\,y)\,\mathrm{d}x}\left\{\int p(x,\,y)\,\mathrm{d}x\right\}\mathrm{d}y\\
&=\int_B\int g(x)p(x,\,y)\,\mathrm{d}x\,\mathrm{d}y\\
&=\mathsf{E}\big[g(X)\mathbb{1}_A\big]\qquad\big(\because Y\in B\Leftrightarrow\omega\in A\big)
\end{aligned}$$

となる．よって

$$\int_A h\big(Y(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega)=\int_A g\big(X(\omega)\big)\,\mathrm{d}\mathsf{Pr}(\omega)$$

なので

$$\mathsf{E}\big[g(X)|\,\mathcal{F}\big](\omega)=h\big(Y(\omega)\big)$$

がわかる． □

F 補遺: 条件付き期待値の性質

定理 F.1. $X,\,\{X_n\}_{n=1}^{\infty},\,Y$ を確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数列とする．
(1) $\mathsf{E}\big[|X|\big]<\infty$ かつ $\mathsf{E}\big[|Y|\big]<\infty$ とする．このとき $a,\,b\in\mathbb{R}$ に対して

$$\mathsf{E}\big[aX+bY|\,\mathcal{F}\big](\omega)=a\mathsf{E}\big[X|\,\mathcal{F}\big](\omega)+b\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)$$

と成り立つ.
(2) さらに $X \leq Y$ のとき

$$\mathsf{E}\big[X|\mathcal{F}\big](\omega) \leq \mathsf{E}\big[Y|\mathcal{F}\big](\omega)$$

が成り立つ.
(3) $X_n \geq 0\,(n = 1,\, 2,\, \ldots)$ かつ $X_n \uparrow X\,(n \to \infty)$ で $\mathsf{E}\big[X\big] < \infty$ のとき

$$\mathsf{E}\big[X_n|\mathcal{F}\big](\omega) \uparrow \mathsf{E}\big[Y|\mathcal{F}\big](\omega) \quad (n \to \infty)$$

が成り立つ.

Proof. (1) あきらかに $a\mathsf{E}\big[X|\mathcal{F}\big] + b\mathsf{E}\big[Y|\mathcal{F}\big]$ は $\mathcal{F}$ 可測である. $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}
&\int_A \big\{a\mathsf{E}\big[X|\mathcal{F}\big](\omega) + b\mathsf{E}\big[Y|\mathcal{F}\big](\omega)\big\}\,\mathrm{d}\mathsf{Pr}(\omega) \\
&\qquad = a\int_A \mathsf{E}\big[X|\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A \mathsf{E}\big[Y|\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\
&\qquad = a\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) + b\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\
&\quad = \int_A \big(aX + bY\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}$$

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

$$\begin{aligned}
\int_A \mathsf{E}\big[X|\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \leq \int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \\
&= \int_A \mathsf{E}\big[Y|\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)
\end{aligned}$$

を得る. $\forall \epsilon > 0$ に対して

$$A := \big\{\omega \in \Omega;\, \mathsf{E}\big[X|\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\mathcal{F}\big](\omega) \geq \epsilon\big\}$$

とおくと $\mathsf{Pr}(A) = 0$ となる. ϵ は任意だったので

$$\begin{aligned}
&\Big\{\omega \in \Omega;\, \mathsf{E}\big[X|\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\mathcal{F}\big](\omega)\Big\} \\
&\qquad = \bigcup_{n=1}^{\infty}\Big\{\omega \in \Omega;\, \mathsf{E}\big[X|\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\mathcal{F}\big](\omega) \geq \frac{1}{n}\Big\}
\end{aligned}$$

となる. よって

$$\begin{aligned}
&\mathsf{Pr}\bigg(\Big\{\omega \in \Omega;\, \mathsf{E}\big[X|\mathcal{F}\big](\omega) > \mathsf{E}\big[Y|\mathcal{F}\big](\omega)\Big\}\bigg) \\
&\quad = \lim_{n\to\infty} \mathsf{Pr}\bigg(\Big\{\omega \in \Omega;\, \mathsf{E}\big[X|\mathcal{F}\big](\omega) - \mathsf{E}\big[Y|\mathcal{F}\big](\omega) \geq \frac{1}{n}\Big\}\bigg) = 0
\end{aligned}$$

となるから

$$\mathsf{Pr}\Big(\omega \in \Omega;\ \mathsf{E}\big[X \big| \mathcal{F}\big](\omega) \leq \mathsf{E}\big[Y \big| \mathcal{F}\big](\omega)\Big) = 1$$

となる.
(3) $Y_n := X - X_n\,(n \in \mathbb{N})$ とおくと $\mathsf{E}\big[Y_n \big| \mathcal{F}\big] \downarrow 0\,(n \to \infty)$ を示せばよい. $Y_n \downarrow 0\,(n \to \infty)$ なので (2) の結果から

$$Z_n(\omega) := \mathsf{E}\big[Y_n \big| \mathcal{F}\big](\omega)$$

はほとんど確実に単調減少列なので

$$Z_\infty(\omega) := \lim_{n\to\infty} \mathsf{E}\big[Y_n \big| \mathcal{F}\big](\omega)$$

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 $Z_\infty \equiv 0$ となる. □

定理 F.2. $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}$ とする. このとい

(1) $\mathsf{E}\big[\mathsf{E}\big[X \big| \mathcal{F}_1\big] \big| \mathcal{F}_2\big](\omega) = \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)$.

(2) $\mathsf{E}\big[\mathsf{E}\big[X \big| \mathcal{F}_2\big] \big| \mathcal{F}_1\big](\omega) = \mathsf{E}\big[X \big| \mathcal{F}_1\big](\omega)$.

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 F.3. 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

$$\mathsf{E}\big[XY|\,\mathcal{F}\big](\omega) = X(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega) \tag{F.52}$$

が成立する.

Proof. (i). $X(\omega) = \mathbb{1}_B(\omega)\,(\forall B \in \mathcal{F})$ のとき (F.52) が成り立つことを示す. $\forall A \in \mathcal{F}$ に対して

$$\begin{aligned}\int_A \mathbb{1}_B(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_{A\cap B} \mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_{A\cap B} Y(\omega)\,\mathsf{Pr}(\omega)\\ &= \int_A \mathbb{1}_B(\omega)Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

となる. よって $X(\omega) = \mathbb{1}_B(\omega)$ のとき (F.52) は 成立する.

(ii). 次に $X,\,Y \geq 0$ とし, X_n は階段関数とし, $X_n \uparrow X\;(n \to \infty)$ とする. 単調収束定理から

$$\begin{aligned}\int_A X(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \lim_{n\to\infty}\int_A X_n(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \lim_{n\to\infty}\int_A \mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) && \big(\because \text{(i) の結果}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) && \big(\because \text{単調収束定理}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[XY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) && \big(\because \text{定理 F.1(3)}\big)\end{aligned}$$

がわかり, $X,\,Y \geq 0$ のとき $(F.52)$ は成立する.

(iii). 最後に一般の $X,\,Y$ に対して

$$\begin{aligned}&X^+ = \max\{X,\,0\}, \quad X^{-1} = \max\{-X,\,0\},\\ &Y^+ = \max\{Y,\,0\}, \quad Y^{-1} = \max\{-Y,\,0\}\end{aligned}$$

として, 上の結果を用いればよい. □

定義 F.4. X は有限の 2 次の積率を持つとする. Y を与えたときの条件付き分散を各 $\omega \in \Omega$ に対して

$$\begin{aligned}\mathsf{Var}\big[X|\,Y\big](\omega) :=&\mathsf{E}\big[X^2|\,Y\big](\omega) - \big\{\mathsf{E}\big[X|\,Y\big](\omega)\big\}^2\\ =&\mathsf{E}\big[\big\{X - \mathsf{E}[X|\,Y](\omega)\big\}^2|Y\big](\omega)\end{aligned}$$

で定義する. このことを

$$\mathsf{Var}\big[X|\,Y\big] = \mathsf{E}\big[X^2|\,Y\big] - \big\{\mathsf{E}\big[X|\,Y\big]\big\}^2 = \mathsf{E}\big[\big\{X - \mathsf{E}\big[X|\,Y\big]\big\}^2|\,Y\big]$$

とも書く.

定理 F.5. 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.

Proof.

$$\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}\big[\{X - \mathsf{E}[X]\}^2\big] \\
&= \mathsf{E}\Big[\big\{X - \mathsf{E}[X|\,Y] + \mathsf{E}[X|\,Y] - \mathsf{E}[X]\big\}^2\Big] \\
&= \mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}^2\big] + \mathsf{E}\big[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}^2\big] \\
&\quad + 2\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big].
\end{aligned}$$

しかし

$$\begin{aligned}
\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}^2\big] &= \mathsf{E}\big[X^2 - 2X\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}\big] \\
&= \mathsf{E}\left[\mathsf{E}\big[X^2 - 2X\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}^2\big]\,\middle|\, Y\right] \\
&\qquad (\text{定理 F.2(1)}) \\
&= \mathsf{E}\left[\mathsf{E}\big[X^2\big|\,Y\big] - 2\mathsf{E}[X|\,Y]\,\mathsf{E}[X|\,Y] + \{\mathsf{E}[X|\,Y]\}^2\right] \\
&= \mathsf{E}\left[\mathsf{E}[X^2|\,Y] - \{\mathsf{E}[X|\,Y]\}^2\right] \\
&= \mathsf{E}\left[\mathsf{Var}[X|\,Y]\right], \\
\mathsf{E}\left[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}^2\right] &= \mathsf{E}\left[\Big\{\mathsf{E}[X|\,Y] - \underbrace{\mathsf{E}[\mathsf{E}[X|\,Y]]}_{=\mathsf{E}[X]}\Big\}^2\right] \qquad (\text{定理 F.2(1)}) \\
&= \mathsf{Var}\left[\mathsf{E}[X|\,Y]\right], \\
\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big] &= \mathsf{E}\left[\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\,\middle|\, Y\right] \\
&\qquad (\text{定理 F.2(1)}) \\
&= \mathsf{E}\left[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\underbrace{\mathsf{E}\big[X - \mathsf{E}[X|\,Y]\big|Y\big]}_{=0}\right] \\
&= 0.
\end{aligned}$$

最後から 2 番目の等号は定理 F.2(1) を用いた. □