第 A 章　確率・確率変数・期待値の復習

この章では, この講義で必要な確率・確率変数・確率分布・期待値の定義を復習する. 証明については, Moodle にある別の資料を参照のこと.

A.1　確率

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

定義 A.1. Ω を空でない集合とし, $\mathcal{A}$ を Ω の部分集合族とする. $\mathcal{A}$ が次の 3 条件をみたすとき, σ **加法族**と呼ばれる.

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

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

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

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

注意 A.2. $\mathcal{C}$ を Ω の部分集合族とする. $\mathcal{C}$ は σ 加法性をみたしてなくともよい. このとき, 集合族 $\sigma[\mathcal{C}]$ を

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

で定める. すると $\sigma[\mathcal{C}]$ は σ 加法族となる. さらに $\mathcal{G}$ を $\mathcal{A}$ を含む σ 加法族としたとき

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

[1] Ω が可算集合ならば, 事象はすべての部分集合としてよいが, Ω が連続濃度のときには, すべての部分集合を対象にすることはしない.

となる．すなわち $\sigma[\mathcal{C}]$ は $\mathcal{C}$ を含む最小 (包含関係の意味) の σ 加法族となる． □

定義 A.3. $\Omega = \mathbb{R}$ とし

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

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

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

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

注意 A.4. Ω が高々可算集合のときは，$\mathcal{F} = 2^{\Omega}$ と取る．ただし 2^{Ω} は Ω のすべての部分集合からなる集合族で冪(べき)集合という． □

定義 A.5. $(\Omega,\, \mathcal{A})$ を可測空間とする．$\mathcal{A}$ 上の関数

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

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

(1) $\mathsf{Pr}(\Omega) = 1$.

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

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

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

定義 A.6. $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ を確率空間とし，$A,\, B \in \mathcal{A}$ とする．

(1) (独立性):

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

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

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

で定める．

[2] $\sigma[\mathcal{C}] \subset \sigma[\mathcal{O}]$ は明らかであるが，逆の包含関係も示すことができる．
[3] 簡単に Ω 上の確率測度ともいう．
[4] $m \neq n$ ならば，$A_m \cap A_n = \varnothing$ が成立していること．

注意 A.7. (1) $A,\ B \in \mathcal{A}$ とし, $\mathsf{Pr}(B) > 0$ とする. A と B が独立のとき

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

が成立する.
(2) $\mathsf{Pr}(B) > 0$ のとき

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

となる. これを**乗法の公式**という.
(3) $\mathsf{Pr}(B) > 0$ のとき, 関数 $\mathsf{Pr}(\,\cdot\,|B) : \mathcal{A} \longrightarrow [0,\,1]$ は定義 A.5(1) − (3) をみたす. すなわち $(\Omega,\,\mathcal{A})$ 上の確率測度である. □

A.2 確率変数

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

定義 A.8. $(\Omega,\,\mathcal{A})$ と $(\mathbb{X},\,\mathcal{B})$ を可測空間とする. 関数 $X:\Omega \longrightarrow \mathbb{X}$ は $(\Omega,\,\mathcal{A})$ から $(\mathbb{X},\,\mathcal{B})$ への**可測写像**であるとは

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

をみたすときをいう.

注意 A.9. (i). $d \geq 2\,(d \in \mathbb{N})$ とする. $\big(\mathbb{X},\,\mathcal{B}\big) = \big(\mathbb{R}^d,\,\mathcal{B}(\mathbb{R}^d)\big)$ のとき, X は**確率ベクトル**と呼べれる.
(ii). $d = 1$ のとき, X は**確率変数**と呼べれる. □

注意 A.10. $X_1,\,X_2$ は確率変数のとき, $X_1 + X_2,\ X_1X_2$ も確率変数であることがわかる. さらに, $X_2 \neq 0$ のとき, X_1/X_2 も確率変数であることもわかる. □

定義 A.11. (1) 確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数 X に対して

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

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

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

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

$$\mathsf{F}_X(x) = \mathsf{F}_Y(x) \qquad (\forall x \in \mathbb{R})$$

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

注意 A.12. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし，F を X の c.d.f. とする．このとき F は次をみたす．

(1) F は非減少関数: $x < y \Longrightarrow \mathsf{F}(x) \le \mathsf{F}(y)$.

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

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

□

定義 A.13. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とし，その c.d.f. と分布をそれぞれ F と P と書く．
(1) X が高々可算個の集合 $\{x_1, x_2, \ldots\}$ 上にしか値を取らないとき，X を離散型であるという．この場合には

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

で X の分布が特徴付けられる[5]．p を X の**確率関数** (probability mass function(p.m.f.)) と呼ぶ．
(2) $\mathsf{Pr}(X = x) = 0\,(\forall x \in \mathbb{R})$ のとき，X を連続型であるという．さらにある非負値関数 p で

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

をみたすものが存在するとき，p を X の**確率密度関数** (probability density function(p.d.f.)) という．特に F がほとんどいたるところ[6]で微分可能ならば，ほとんどいたるところで

$$\dot{\mathsf{F}}(x) = \frac{d\mathsf{F}}{dx}(x) = \mathsf{p}(x)$$

となる．

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

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

定義 A.14. c.d.f. F に対して，**分位点関数** (quantile function) $\mathsf{F}^{-} : (0, 1) \longrightarrow \mathbb{R}$ を

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

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

注意 A.15. F を c.d.f. とする．このとき，$U \sim \mathsf{U}(0, 1)$ に対して

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

となる． □

A.3 主な 1 次元分布

A.3.1 離散型確率変数

Bernoulli 分布

$0 \leq \theta \leq 1$ とする．確率変数 X は**母数** θ **の** Bernoulli **分布**に従うとは，X の p.m.f. $\mathsf{p}(\,\cdot\,|\,\theta)$ が

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

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

2 項分布

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

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

のときをいう．ただし

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

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

幾何分布

$0 < \theta < 1$ とする. 確率変数 X は (X は**母数 θ の幾何分布**に従うとは, X の p.m.f. $\mathsf{p}(\cdot\,|\,\theta)$ が

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

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

Poisson 分布

$\theta > 0$ とする. 確率変数 X は**母数 θ の Poisson 分布**に従うとは, X の p.m.f. $\mathsf{p}(\cdot\,|\,\theta)$ が

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

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

A.3.2 連続型確率変数

正規分布

$\mu \in \mathbb{R}$, $0 < \sigma < \infty$ とする. 確率変数 X は**平均 μ, 分散 σ^2 の正規分布**に従うとは, X の p.d.f. $\mathsf{p}(\cdot\,|\,\mu,\, \sigma^2)$ が

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

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

ガンマ分布

$\alpha > 0$, $\beta > 0$ とする. 確率変数 X は**母数 $(\alpha,\, \beta)$ のガンマ分布**に従うとは, X の p.d.f. $\mathsf{p}(\cdot\,|\,\alpha,\, \beta)$ が

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

のときをいう. ただし

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

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

A.4 2 次元の分布

A.4.1 同時確率関数と密度関数

定義 A.16. 確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数 X, Y は連続型とする. $\mathbb{R}^2$ 上の非負値実数値関数 p が確率ベクトル $(X,\, Y)$ の **同時確率密度関数** (同時 p.d.f.) であるとは, 次の条件をみたすときをいう.

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

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

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

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

注意 A.17. $(X,\, Y)$ の同時累積分布関数 (同時 c.d.f.) F を

$$\begin{aligned} \mathsf{F}(x,\, y) &:= \mathsf{Pr}(X \leq x,\, Y \leq y) \qquad \text{(A.1)} \\ &:= \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \leq x,\, Y(\omega) \leq y\}) \qquad (\forall (x,\, y) \in \mathbb{R}^2) \end{aligned}$$

で定義する. $(X,\, Y)$ の同時分布 P を

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

で定義する. どの確率変数の c.d.f. と分布であるかを明示したいときには, $\mathsf{F}_{(X,Y)}$ または $\mathsf{P}_{(X,Y)}$ と記す. 細かなことであるが, (A.1) の F によって $(\mathbb{R}^2,\, \mathcal{B}(\mathbb{R}^2))$ 上の確率測度 P を一意的に定めていることができることが知られている. この事実の証明に関しては測度論の知識が必要となる. □

A.4.2 周辺分布

定義 A.18. (X, Y) を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率ベクトルとする．(1) (X, Y) が離散型で同時 p.m.f. p を持つとする．X の**周辺確率関数** (周辺 p.m.f.) を

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

で定義する．ただし

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

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

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

で定義し，Y の**周辺確率密度関数** (周辺 p.d.f.) を

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

で定義する．

注意 A.19. 連続型確率ベクトル (X, Y) に対して

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

となるので，X の周辺 p.d.f. と X の p.d.f. は同じである． □

注意 A.20. F を確率ベクトル (X, Y) の同時 c.d.f. とし，F_X を X の周辺 c.d.f. とする．このとき $x \in \mathbb{R}$ に対して

$$\mathsf{F}_X(x) = \lim_{y \to \infty} \mathsf{F}(x, y); \ \mathsf{F}_Y(y) = \lim_{x \to \infty} \mathsf{F}(x, y)$$

が成立する． □

A.4.3 独立性な分布と条件付き分布

定義 A.21. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の 2 つの確率変数 X と Y は**独立**であるとは, $\forall A, B \in \mathcal{B}(\mathbb{R})$ に対して

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

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

注意 A.22. p を確率ベクトル (X, Y) の同時 p.d.f. とし, p_X と p_Y を X と Y それぞれの周辺 p.d.f. とする. このとき

$$X \text{ と } Y \text{ は独立} \iff \mathsf{p}(x, y) = \mathsf{p}_X(x)\mathsf{p}_Y(y) \quad (\forall x, y \in \mathbb{R})$$

が成り立つ.

定義 A.23. (1) 離散型確率ベクトル (X, Y) は同時 p.m.f. $\mathsf{p}(x, y)$ を持つとする. $\mathsf{p}_Y(y) > 0$ なる y に対して, $Y = y$ を与えたときの X の**条件付き確率関数** (条件付き p.m.f.) を

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

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

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

で定義する.

注意 A.24. (1) (X, Y) が連続型確率ベクトルのとき, $\mathsf{p}_Y(y) > 0$ なる $y \in \mathbb{R}$ と $A \in \mathcal{B}(\mathbb{R})$ に対して, 条件付き確率 $\mathsf{Pr}(X \in A|Y = y)$ を

$$\mathsf{Pr}(X \in A|Y = y) := \int_A \mathsf{p}_{X|Y}(x|y)\,dx \qquad \big(\forall A \in \mathcal{B}(\mathbb{R})\big) \qquad \text{(A.3)}$$

と形式的に定義する.
(2) (A.3) が成り立つとき,

$$X|Y = y \sim \mathsf{p}_{X|Y}(x|y)$$

と書くことにする. □

A.5 多次元分布と i.i.d. 標本

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

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

と書く. $\boldsymbol{X}$ を確率ベクトルという.

定義 A.25. $X_1, X_2, \ldots, X_n$ が独立であるとは, すべての Borel 集合 $A_1, A_2, \ldots, A_n \in \mathcal{B}(\mathbb{R})$ に対して

$$\mathsf{Pr}(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{j=1}^{n} \mathsf{Pr}(X_j \in A_j) \tag{A.4}$$

が成立するときである.

$\boldsymbol{X}$ の同時 p.d.f. を $\mathsf{p}(x_1, x_2, \ldots, x_n)$ と書き, 各 $X_j\,(j = 1, 2, \ldots, n)$ の周辺 p.d.f. を p_{X_j} と書くことにする.

注意 A.26. (A.4) を示すためには

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

を示せばよい. □

定義 A.27. $X_1, X_2, \ldots, X_n$ は独立で, 各 $X_j\,(j = 1, 2, \ldots, n)$ は同じ c.d.f. F を持つとき, $X_1, X_2, \ldots, X_n$ は**独立同一分布に従う** (i.i.d. =identically and independently distributed) といい

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

と書く[7]. $X_1, X_2, \ldots, X_n$ は累積分布関数 F から標本の大きさが n のランダム標本ともいう.

A.5.1 多変量正規分布

$$\boldsymbol{Z} := \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_d \end{pmatrix}, \quad Z_1, Z_2, \ldots, Z_d \sim \mathsf{N}(0, 1)$$

[7] p.d.f. p を使い

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

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

とする[8]. $\boldsymbol{Z}$ の同時 p.d.f. は

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

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

$$\int \cdots \int \mathsf{p}_{\boldsymbol{Z}}(z_1,\, z_2,\, \ldots,\, z_d)\, dz_1 dz_2 \cdots dz_d = 1$$

となっていることもわかる.

次に

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

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

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

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

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

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

$$\mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}|\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}\mathrm{Det}\,(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (\boldsymbol{x} \in \mathbb{R}^d) \tag{A.5}$$

で与えられる. この分布を**平均ベクトル $\boldsymbol{\mu}$, 分散共分散行列 Σ の d 変量正規分布**という.

注意 A.28. (A.5) の導出は以下のように行うことができる. 一般に $\boldsymbol{Z} = (Z_1,\, Z_2,\, \ldots,\, Z_d)^\top$ を d 次元確率ベクトルとし, その同時 p.d.f. を $p_{\boldsymbol{Z}}$ と書くことにする. さらに $\mathbb{X} := \{\boldsymbol{x} \in \mathbb{R}^d;\, \mathsf{p}_{\boldsymbol{Z}}(\boldsymbol{x}) > 0)\}$, 関数

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

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

は 1 対 1 とし

$$X_j = h_j(Z_1, Z_2, \ldots, Z_d) \quad (j = 1, 2, \ldots, d), \boldsymbol{X} \quad = (X_1, X_2, \ldots, X_d)^\top$$

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

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

が存在して, $\boldsymbol{Z} = \boldsymbol{h}^{-1}(\boldsymbol{X})$ となる. $\boldsymbol{X}$ の同時 p.d.f. を求めるために, $\boldsymbol{h}^{-1}(\boldsymbol{x})$ の Jacobian $\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})$ を

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

で定める. このとき

$$\mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}) = \left|\mathbf{J}_{\boldsymbol{h}^{-1}}(\boldsymbol{x})\right| \mathsf{p}_{\boldsymbol{Z}}(h^{-1}(\boldsymbol{x})) \tag{A.6}$$

となるが知られている.

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

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

を (A.6) に代入すれば, (A.5) はわかる. 以上の議論から

$$\int\int\cdots\int \mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x})\,dx_1 dx_2 \cdots dx_d = 1$$

となっていることもわかる. □

A.6 確率変数の期待値

定義 A.29. $g : \mathbb{R} \longrightarrow \mathbb{R}$ を可測[9]とする. $g(X)$ の期待値 $\mathsf{E}[g(X)]$ を次のように定義する.

(1) $g \geq 0$ のとき

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

と定義する. 右辺は ∞ を許せば, 必ず存在する.

[9] g が可測関数であるとは, $\forall B \in \mathcal{B}(\mathbb{R})$ に対して, $f^{-1}(B) = \{x \in \mathbb{R}; f(x) \in B\} \in \mathcal{B}(\mathbb{R})$ をみたすときをいう.

(2) 一般の g に対して

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

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

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

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

注意 A.30. X を確率変数とする. 可測関数 $g, h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して, $\mathsf{E}[|g(X)|] < \infty$, $\mathsf{E}[|h(X)|] < \infty$ を仮定する.

(1) $a, b \in \mathbb{R}$ に対して

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

が成り立つ.

(2) $g(x) \leq h(x)$ $(\forall x \in \mathbb{R})$ ならば

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

が成り立つ.

□

定義 A.31. (1) $k = 1, 2, \ldots$ に対して, $\mathsf{E}[|X|^k] < \infty$ のとき, $\mathsf{E}[X^k]$ を X の k **次モーメント** (または**積率**) という.

(2) $\mathsf{E}[|X|] < \infty$ のとき $\mathsf{E}[X]$ を X **の平均値**[10]という.

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

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

で定義する.

(4) $A \subset \mathbb{R}$ に対して

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

を A の**指示関数**[11]という. すると $A \in \mathcal{B}(\mathbb{R})$ のとき

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

となる.

[10]簡単に「平均」ともいう.

[11]指示関数は $\mathbb{R}$ の任意の部分集合に定義できる.

A.7 確率ベクトルの期待値

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

(X, Y) が離散型のときその同時 p.m.f. を $\mathsf{p}(x, y)$ とし, 連続型のときその同時 p.d.f. を $\mathsf{p}(x, y)$ とする．

定義 A.32. 可測関数 $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ に対して $g(X, Y)$ の期待値を次のように定義する．

(1) $g \geq 0$ のとき

$$\mathsf{E}[g(X, Y)] = \begin{cases} \displaystyle\sum_{x,y} g(x, y)\mathsf{p}(x, y) & (\text{離散型}) \\ \displaystyle\iint_{\mathbb{R}^2} g(x, y)\mathsf{p}(x, y)\,dxdy & (\text{連続型}). \end{cases}$$

(2) 一般の g に対して

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

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

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

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

注意 A.33. 3 つ以上の確率変数 $X_1, X_2, \ldots, X_n$ に対しても期待値を定義 A.5 と同様に定義する． □

注意 A.34. $X_1, X_2, \ldots, X_n$ を確率変数とし, 各 $X_j\,(j = 1, 2, \ldots, n)$ の期待値は有限とする．$a_1, a_2, \ldots, a_n$ を定数としたとき

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

が成り立つ． □

A.8 分散と共分散

定義 A.35. X を確率変数とし, $\mathsf{E}[X^2] < \infty$ とする．X **の分散**を

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

で定義する．ただし $\mu = \mathsf{E}[X]$ と書いた．さらに $\sqrt{\mathsf{Var}[X]}$ を X **の標準偏差**という．

注意 A.36. 分散 $\mathsf{Var}[X]$ は X の分布の平均 μ まわりの散らばりを測る量である．分散がおおきいほど分布は広がっていることになる． □

注意 A.37. 以下の確率変数は有限の 2 次の積率を持つとする．このとき, 次が成立する．

(1) $\mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2$ が成り立つ．

(2) 定数 $a, b \in \mathbb{R}\,(a \neq 0)$ に対して

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

が成り立つ．

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

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

が成り立つ．

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

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

が成り立つ．

□

注意 A.38. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n$ は i.i.d. 確率変数列とし

$$\mathsf{E}[X_1] = \mu, \quad \mathsf{Var}[X_1] = \sigma^2, \quad 0 < \sigma < \infty$$

とする．$X_1, X_2, \ldots, X_n$ に基づく**標本平均**を

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

で定義し, **標本** (不偏) **分散**を

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

で定義する．このとき

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

が成り立つ． □

定義 A.39. X と Y は確率変数とし

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

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

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

で定義し，**X と Y の** (Pearson) **の相関係数**を

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

で定義する．

定義 A.40. $X_1,\, X_2,\, \ldots,\, X_d$ を有限な 2 次の積率を持つ確率変数とし

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

と書く．このとき確率ベクトル $\boldsymbol{X}$ の期待値を

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

で定義する．$\boldsymbol{X}$ の共分散を

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

で定義する[12]．ただし $\boldsymbol{\mu} = (\mu_1,\, \mu_2,\, \ldots,\, \mu_d)^\top := \mathsf{E}[\boldsymbol{X}]$ である．これは

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

である．

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

注意 A.41. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ は確率ベクトルで各成分は有限な 2 次の積率を持つとし

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

とする．ただし $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ は $d \times d$ の半正定値行列[13]である．
(1) 任意の定数ベクトル $\boldsymbol{a} \in \mathbb{R}^d$ に対して

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

が成り立つ．
(2) $\in \mathbb{N}$ とする．任意の定数の $k \times d$ 次正方行列 $\mathbf{A}$ に対して

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

が成り立つ． □

A.9 条件付き期待値

定義 A.42. (1) X と Y を確率変数とし, 条件付き p.d.f.(条件付き p.m.f.) を $\mathsf{p}_{X|Y}(p_{X|Y})$ とする．$Y = y$ を与えたときの X の条件付き期待値を

$$\mathsf{E}[X|\,Y = y] = \begin{cases} \displaystyle\sum_x x\mathsf{p}_{X|Y}(x|\,y) & (\text{離散型の場合}) \\ \displaystyle\int x\mathsf{p}_{X|Y}(x|\,y)\,dx & (\text{連続型の場合}) \end{cases}$$

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

$$\mathsf{E}[g(X, Y)|\,Y = y] = \begin{cases} \displaystyle\sum_x g(x, y)\mathsf{p}_{X|Y}(x|\,y) & (\text{離団型の場合}) \\ \displaystyle\int g(x, y)\mathsf{p}_{X|Y}(x|\,y)\,dx & (\text{連続型の場合}) \end{cases}$$

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

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

注意 A.43. (1) $\mathsf{E}[X]$ は定数であるが, $\mathsf{E}[X|\,Y=y]$ は一般に y の関数である.

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

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

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

と定義する. したがって $\omega \in \Omega$ に対して, $y = Y(\omega)$ と書けば

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

は可測となる.
(2) 任意の $B \in \mathcal{B}(\mathbb{R})$ に対して, 条件付き確率 $\mathsf{Pr}(X \in B|\,Y)$ を

$$\mathsf{Pr}(X \in B|\,Y) = \mathsf{E}[\mathbb{1}_B(X)|\,Y]$$

で定める. □

注意 A.44. (1) 有限な期待値を持つ確率変数 X, Y と確率変数 Z に対して

$$\mathsf{E}[X+Y|\,Z] = \mathsf{E}[X|\,Z] + \mathsf{E}[Y|\,Z]$$

が成り立つ.
(2) 有限な期待値を持つ確率変数 X, Y に対して

$$\mathsf{E}[\mathsf{E}[Y|\,X]] = \mathsf{E}[Y], \quad \mathsf{E}[\mathsf{E}[X|\,Y]] = \mathsf{E}[X]$$

が成り立つ.
(3) 一般の (Borel 可測) 関数 $g:\mathbb{R}^2 \longrightarrow \mathbb{R}$ を考える. $\mathsf{E}[|g(X,Y)|] < \infty$ のとき

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

が成り立つ.
(4) $\mathsf{E}[XY|\,Y] = Y\mathsf{E}[X|\,Y]$ が成り立つ. □

注意 A.45. 条件付き p.d.f. から出発して, 条件付き期待値を定義した. 逆に, 条件付き期待値の定義から出発して, 条件付き p.d.f. を導入する流儀[14]もある. $\mathsf{E}\big[|g(X,Y)|\big] < \infty$ なる関数 $g:\mathbb{R}^2 \longrightarrow \mathbb{R}$ に対して, 条件付き期待値 $\mathsf{E}\big[g(X,Y)|Y\big]$ を以下のように定義してもよい. $h:\mathbb{R} \longrightarrow \mathbb{R}$

[14] 測度論的確率論の流儀である. 実は, こちらの定式化のが, 直観的に理解するのは難しいが, 数学的には一貫した定義になる.

を任意の区分的に有界連続な関数とする．このとき, ある $\widetilde{g}: \mathbb{R} \longrightarrow \mathbb{R}$ があって

$$\mathbb{E}\big[h(Y)\widetilde{g}(Y)\big] = \mathbb{E}\big[h(Y)g(X, Y)\big]$$

が成り立つとき, 条件付き期待値 $\mathsf{E}\big[g(X, Y)|\, Y\big]$ を $\widetilde{g}(Y) = \mathsf{E}[g(X, Y)|\, Y]$ で定めるとしてもよい. □

定義 A.46. X は有限の 2 次の積率を持つとする．$Y = y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}_{X|Y}$(p.m.f. $\mathsf{p}_{X|Y}$) が定義できる y を考える．このとき, $Y = y$ を与えたときの条件付き分散を

$$\mathsf{Var}[X|\, Y = y] = \begin{cases} \displaystyle\sum_x \{x - \mu(y)\}^2 \mathsf{p}_{X|Y}(x|\, y) & (\text{離散型の場合}) \\ \displaystyle\int \{x - \mu(y)\}^2 \mathsf{p}_{X|Y}(x|\, y)dx & (\text{連続型の場合}) \end{cases}$$

で定義する．ただし $\mu(y) = \mathsf{E}[X|\, Y = y]$ である．これは

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

とも書ける.

注意 A.47. X, Y を確率変数とし $\mathsf{E}[X^2] < \infty$ とする．このとき

$$\mathsf{Var}[X] = \mathsf{E}[\mathsf{Var}[X|\, Y]] + \mathsf{Var}[\mathsf{E}[X|\, Y]]$$

が成立する．ただし $h(y) = \mathsf{Var}[X|\, Y = y]$ としたとき $\mathsf{Var}[X|\, Y] := h(Y)$ と定義した. □

A.10 積率母関数

定義 A.48. X を確率変数とし，ある $t_0 > 0$ が存在して，$\mathsf{E}[e^{tX}] < \infty\,(\forall |t| < t_0)$ とする．このとき，X **の積率母関数** (Moment Generating Function (MGF)) を

$$\mathsf{m}_X(t) := \mathsf{E}[e^{tX}] \quad (-t_0 < t < t_0)$$

と定義する.

注意 A.49. (1) $a, b \in \mathbb{R}\,(a \neq 0)$ とする．$Y = aX + b$ としたとき

$$\mathsf{m}_Y(t) = e^{tb}\mathsf{m}_X(at)$$

が成り立つ.

(2) $X_1, X_2, \ldots, X_d$ は独立とし $Y = \sum_{j=1}^{d} Y_j$ とする．このとき

$$\mathsf{m}_Y(t) = \prod_{j=1}^{d} \mathsf{m}_{X_j}(t)$$

が成り立つ. □

注意 A.50. X と Y を確率変数とする．ある数 $t_0 > 0$ が存在して

$$\mathsf{m}_X(t) = \mathsf{m}_Y(t) \quad (|t| < t_0)$$

ならば

$$X \overset{d}{=} Y$$

となる．ただし X と Y の c.d.f. を F_X と F_Y としたとき

$$X \overset{d}{=} Y \iff \mathsf{F}_X(x) = \mathsf{F}_Y(x) \qquad (\forall x \in \mathbb{R})$$

である. □

A.11 確率に対する不等式

注意 A.51. (Markov の不等式) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の X を非負値確率変数とし, $\mathsf{E}[X] < \infty$ とする．このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\big(X \geq t\big) \leq \frac{\mathsf{E}[X]}{t}$$

が成り立つ.

注意 A.52. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X]\,(0 < \sigma < \infty)$ とする．このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(|X - \mu| \geq t\big) \leq \frac{\sigma^2}{t^2}$$

が成り立つ. □

A.12 期待値に対する不等式

注意 A.53. (Cauchy-Schwarz の不等式) 確率変数 X と Y は 2 次の有限な期待値を持つとき

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

が成り立つ. □

定義 A.54. 関数 $g:\mathbb{R}\longrightarrow\mathbb{R}$ が凸であるとは各 $x,\,y\in\mathbb{R}$ と $0\le\forall t\le 1$ に対して

$$g(tx+(1-t)y)\le tg(x)+(1-t)g(y)$$

が成立するときをいう．さらに $-g$ が凸のとき g は concave であるという．

注意 A.55. (Jensen の不等式) X を有限な期待値を持つ確率変数とする．
(1) 関数 $g:\mathbb{R}\longrightarrow\mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

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

が成り立つ．
(2) g が concave のとき

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

が成り立つ． □

A.12.1 測度論的な積分の定義

定義 A.56. $\big(\mathbb{R},\,\mathcal{B}(\mathbb{R})\big)$ 上の測度とは, $\mu:\mathcal{B}(\mathbb{R})\longrightarrow[0,\,\infty)\cup\{+\infty\}$ を以下の条件 (i), (ii) をみたすものである．
(i) $\mu(\varnothing)=0$.
(ii) $A_j\in\mathcal{B}(\mathbb{R})\,(j=1,\,2,\,\ldots)$ は互いに排反で

$$B:=\bigcup_{j=1}^{\infty}A_j\in\mathcal{B}(\mathbb{R})$$

とする．このとき

$$\mu(B)=\sum_{j=1}^{\infty}\mu(A_j)$$

をみたす．

注意 A.57. 任意の $a<b\,(a,\,b\in\mathbb{R}))$ に対して, 測度 m で

$$\mathsf{m}\big((a,\,b]\big)=b-a$$

をみたすものを $\mathbb{R}$ 上の Lebeague **測度**という．さらに, $d\ge 2\,(d\in\mathbb{N}$ としたとき, $\mathbb{R}^d$ 上の Lebeague 測度 m_d も同じように定義できることが知られている．

定義 A.58. $(\mathbb{X}, \mathcal{A})$, $(\mathbb{Y}, \mathcal{B})$ 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

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

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

注意 A.59. 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[15]すれば，$\lim_{n\to\infty} f_n$ も Borel 可測である． □

$E \in \mathcal{A}$ と $f \in \mathcal{M}\big(\mathbb{X}, \mathbb{R} \cup \{\pm\infty\}\big)$ に対して，積分 $\displaystyle\int_E f\,d\mu$ 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

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

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

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

で定める．

[15] $\lim_{n\to\infty} f_n$ が存在するとは

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

のときをいう．

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

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

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

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

で定める.

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

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

と書く. $\displaystyle\int_E f^+\, d\mu < \infty$ または $\displaystyle\int_E f^-\, d\mu < \infty$ 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 $\displaystyle\int_E f\, d\mu$ を

$$\int_E f\, d\mu = \int_E f^+\, d\mu - \int_E f^-\, d\mu$$

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

注意 A.60. $f, g \in \mathcal{M}\bigl(\mathbb{X}, \mathbb{R} \cup \{\pm\infty\}\bigr)$ 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 $\displaystyle\int_E f\, d\mu = 0$ 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 $\displaystyle\int_E f\, d\mu = \int_{\mathbb{X}} f\mathbb{1}_E\, d\mu$ 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

$$\int_E af\, d\mu = a\int_E f\, d\mu$$

となる.

(iv) f は E 上積分確定とする. このとき

$$\left| \int_E f\, d\mu \right| \leq \int_E |f|\, d\mu$$

となる．

(v) f, g は E 上で積分確定で，A 上で　$f \le g\,(\mu-\text{a.e.})$ 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

$$\int_E f\,d\mu \le \int_E g\,d\mu$$

となる．

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

$$\int_A (f+g)\,d\mu = \int_A f\,d\mu + \int_A g\,d\mu \tag{A.7}$$

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 $(A.7)$ が成立する． □