第1章　準備: 確率空間・事象・確率変数・分布・期待値

1.1　記号と定義

関数族の記号　$k,\, p$ を自然数とし, μ を $\mathbb{R}^k$ 上の測度とする.

- $\mathcal{C}_b(\mathbb{R}^k)$: $\mathbb{R}^k$ 上で定義された有界連続実数値関数全体の集合.
- $\mathcal{L}^p(\mathbb{R}^k,\, \mu)$: $\mathbb{R}^k$ 上の実数値可測関数で $\int_{\mathbb{R}^k} |f(x)|^p\, \mathrm{d}\mu(x) < \infty$ なるもの全体の集合.

指示関数 $A \subset \mathbb{R}$ に対して

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

と定める. 特に, $\mathbb{1}_{(-\infty,\, a]}(x)$ を簡単に $\mathbb{1}\{x \leq a\}$ と記すことにする.

1.2　確率と確率変数の定義

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

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

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

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

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

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

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

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

となることが直にわかる．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 1.3. $\Omega = \mathbb{R}$ 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[1].

定義 1.4. $(\Omega, \mathcal{A})$ を可測空間とする．$\mathcal{A}$ 上の関数

$$\mu :\, \mathcal{A} \ni A \mapsto \mu(A) \in [0, \infty) \cup \{\infty\}$$

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

(1) $\mu(\emptyset) = 0$ である．

(2) 互いに排反[3]な事象列 $A_n \in \mathcal{A}\,(n = 1, 2, \ldots)$ に対して

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

をみたす．

これらの 3 つの組 $(\Omega, \mathcal{A}, \mu)$ を**測度空間**という．とくに, $\mu(\Omega) = 1$ のとき, **確率測度といい**, $(\Omega, \mathcal{A}, \mu)$ を**確率空間**という．

以後では, 確率測度を $\mathsf{Pr}(\cdot)$ と記すことにする．$\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ を確率空間とする．

補題 1.5. $(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とする．このとき以下が成立する．

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

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

[2] 簡単に Ω 上の測度ともいう．

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

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

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

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

となる.

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

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

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

となる.

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

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

となる.

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

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

となる.

Proof. 節 1.6.1 を参照のこと. □

定義 1.6. $(\Omega, \mathcal{A})$ を可測空間とする. 写像 $X : \Omega \to \mathbb{R}$ は $(\Omega, \mathcal{A})$ から $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ への**可測写像**であるとは

$$X^{-1}(B) := \bigl\{\omega \in \Omega;\, X(\omega) \in B\bigr\} \in \mathcal{A} \qquad \bigl(\forall B \in \mathcal{B}(\mathbb{R})\bigr)$$

をみたすときをいう.

定義 1.7. (1) $d \ge 2\,(d \in \mathbb{N})$ とする. $X : \Omega, \longrightarrow \mathbb{R}^d$ が可測写像のとき, X は確率空間 $\bigl(\Omega, \mathcal{A}, \mathsf{Pr}\bigr)$ 上の**確率ベクトル**と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率ベクトルということもある.
(2) $d = 1$ のとき, X は確率空間 $\bigl(\Omega, \mathcal{A}, \mathsf{Pr}\bigr)$ 上の **確率変数**と呼ばれる. 定義されている確率空間に誤解がないときには, 簡単に確率変数と簡単にいうこともある. □

定理 1.8. $\big(\Omega, \mathcal{A}\big)$ を可測空間とし, $X : \Omega \to \mathbb{R}$ を写像とし, $\mathcal{C}$ を $\mathbb{R}$ の集合族とする. $\forall C \in \mathcal{C}$ に対して, $\big\{\omega \in \Omega;\ X(\omega) \in C\big\} \in \mathcal{A}$ であり, $\mathcal{C}$ が $\mathcal{B}(\mathbb{R})$ を生成する[4]とき, X は可測となる.

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

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

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

注意 1.9. (1). $\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ に対して, 定理 1.8 における $\mathcal{C}$ の選択として, $\{(-\infty, r] : r \in \mathbb{R}\}$ と $\{(-\infty, q] : q \in \mathbb{Q}\}$ などがある.
(2). $d \geq 2\,(d \in \mathbb{N})$ とする. $\big(\mathbb{X}, \mathcal{B}\big) = \big(\mathbb{R}^d, \mathcal{B}(\mathbb{R})\big)$ に対して, 定理 1.8 における $\mathcal{C}$ の選択として

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

がある. □

注意 1.10. 以下のことを証明できる.
(1) $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は確率変数とし, $f : \mathbb{R}^n \to \mathbb{R}$ は可測とする. このとき, $f(X_1, X_2, \ldots, X_n)$ は確率変数となる.
(2) $\{X_n\}_{n=1}^{\infty}$ を確率変数列とする. このとき

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

も確率変数となる. □

1.3 事象, 部分 σ 集合族, 確率変数列の独立性

定義 1.11. $A, B \in \mathcal{A}$ として, 2 つの事象 A と B が**独立**であるとは

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

が成り立つときをいう.

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

定義 1.12. $N \in \mathbb{N}$, $N \geq 3$ とする. $A_k \in \mathcal{A}\,(k = 1, 2, \dots, N)$ として, 事象の集まり $\{A_k\}_{k=1}^N$ が**独立**であるとは, 任意の $2 \leq \ell \leq N$ と任意の $1 \leq k_1 < k_2 < \cdots < k_\ell \leq N$ に対して

$$\mathsf{Pr}\biggl(\bigcap_{j=1}^{\ell} A_{k_j}\biggr) = \prod_{j=1}^{\ell} \mathsf{Pr}(A_{k_j}) \tag{1.1}$$

が成り立つときをいう.

定義 1.13. 一般に非可算な集合族 Λ によって添え字付られた事象の集まり $\{A_k\}_{k\in\Lambda}$, $A_k \in \mathcal{A}$ が**独立**であるとは, Λ の任意の有限部分集合 $\{k_1, k_2, \dots, k_\ell\} \subset \Lambda$ に対して (1.1) が成り立つときをいう.

定義 1.14. $\mathcal{A}_1$, $\mathcal{A}_2$ を $\mathcal{A}$ の**部分 σ 集合族**, すなわち

$$\text{各 } \mathcal{A}_k\,(k = 1, 2) \text{ は } \sigma \text{ 加法族で } \mathcal{A}_k \subset \mathcal{A}$$

をみたす. 部分 σ 集合族 $\mathcal{A}_1$ と $\mathcal{A}_2$ は**独立**であるとは, 任意の $A_1 \in \mathcal{A}_1$ と $A_2 \in \mathcal{A}_2$ に対して

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

が成り立つときをいう.

定義 1.15. (1) $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_N$ を $\mathcal{A}$ の部分 σ 集合族とする. $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_N$ が**独立**であるとは, 任意の $A_k \in \mathcal{A}_k\,(k = 1, 2, \dots, N)$ に対して

$$\mathsf{Pr}\biggl(\bigcap_{k=1}^{N} A_k\biggr) = \prod_{k=1}^{N} \mathsf{Pr}(A_k)$$

が成り立つときをいう.
(2) 一般に非可算集合 Λ で添え字付けれた $\mathcal{A}$ の部分 σ 集合族の集まり $\{\mathcal{A}_\lambda\}_{\lambda\in\Lambda}$ が独立であるとは, Λ の任意の有限部分集合 $\{k_1, k_2, \dots, k_\ell\}$ に対して, $\{A_{k_j}\}_{j=1}^{\ell}$ が独立であるときをいう.
(3) $\{\mathcal{A}_\lambda\}_{\lambda\in\Lambda}$ から任意の組 $j \neq k\,(j, k \in \Lambda)$ をとるとき, $\mathcal{A}_j$ と $\mathcal{A}_k$ が独立のとき, $\{\mathcal{A}\}_{\lambda\in\Lambda}$ は**組ごとに独立**という.

定義 1.16. 確率変数 $X_1, X_2, \dots, X_N$ が**独立**であるとは, 任意の $B_k \in \mathcal{B}(\mathbb{R})\,(k = 1, 2, \dots, N)$ に対して

$$\mathsf{Pr}\biggl(\bigcap_{k=1}^{N}\bigl\{X_k \in B_i\bigr\}\biggr) = \prod_{k=1}^{N} \mathsf{Pr}\bigl(X_k \in B_k\bigr)$$

が成り立つときをいう.

(2) 一般に非可算集合 Λ で添え字付けれた確率変数の集まり $\{X_\lambda\}_{\lambda\in\Lambda}$ が**独立**であるとは, Λ の任意の有限部分集合 $\{k_1, k_2, \ldots, k_\ell\}$ と $A_j \in \mathcal{B}(\mathbb{R})\,(j = 1, 2, \ldots, \ell)$ に対して

$$\mathsf{Pr}\left(\bigcap_{j=1}^{\ell}\{X_{k_j} \in A_j\}\right) = \prod_{j=1}^{\ell}\mathsf{Pr}(X_{k_j} \in A_j)$$

が成り立つときをいう.

1.4 期待値の定義

1.4.1 期待値の定義

確率変数 X の期待値を以下の 3 つの段階, ① X の取りうる値の集合が有限である場合, ② 非負値確率変数の場合, ③ 一般の場合, の順に定義していく. ① ～ ③ の操作を行うことを**標準機械**という.

① 確率変数 X の取りうる値の集合が有限のとき

定義 1.17. 確率変数 X は単純であるとは, X の取りうる値の集合 $\big\{X(\omega); \omega \in \Omega\big\}$ が有限であることをいう.

$\{x_1, x_2, \ldots, x_n\} = \big\{X(\omega); \omega \in \Omega\big\}$ と書く. ただし, $x_1, x_2, \ldots, x_n$ は異なる値である. このとき

$$A_j := \big\{\omega \in \Omega; X(\omega) = x_j\big\} \qquad (j = 1, 2, \ldots, n)$$

とおくと

$$X(\omega) = \sum_{j=1}^{n} x_j \mathbb{1}_{A_j}(\omega) \tag{1.2}$$

と書ける. $\{A_j\}_{j=1}^{n}$ は Ω の有限な分割となっていることに注意する.

(1.2) のような形で表現される確率変数を**単純** (simple) という.

定義 1.18. 単純な確率変数 $X = \sum_{j=1}^{n} x_j \mathbb{1}_{A_j}\,(x_j \in \mathbb{R}, A_j \in \mathcal{A})$ に対して, X **の 期待値** $\mathsf{E}[X]$ を

$$\mathsf{E}[X] = \mathsf{E}\left[\sum_{j=1}^{n} x_j \mathbb{1}_{A_j}\right] = \sum_{j=1}^{n} x_j \mathsf{Pr}(A_j)$$

で定める.

命題 1.19. X, Y を単純な確率変数とし, $a, b \geq 0$ とする. このとき

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

が成り立つ.

Proof. $x_j, y_\ell \in \mathbb{R}(j = 1, \ldots, m; \ell = 1, 2, \ldots, n)$ とし, $\{A_j\}_{j=1}^m$, $\{B_\ell\}_{\ell=1}^n$ を Ω の分割とする. X, Y は単純なので

$$X = \sum_{j=1}^m x_j \mathbb{1}_{A_j}, \qquad Y = \sum_{\ell=1}^n y_\ell B_\ell$$

と書けたとする. すると $\{A_j \cap B_\ell\}_{j=1,2,\ldots,m;\,\ell=1,2,\ldots,n}$ も Ω の分割となる. このとき

$$\begin{aligned}
\mathsf{E}\big[aX + bY\big] &= \mathsf{E}\left[\sum_{j=1}^m \sum_{\ell=1}^n (ax_j + b_\ell)\mathbb{1}_{A_j \cap B_\ell}\right] \\
&= \sum_{j=1}^m \sum_{\ell=1}^n (ax_j + b_\ell)\mathsf{Pr}\big(A_j \cap B_\ell\big)\Bigg] \\
&= a\sum_{j=1}^m x_j \mathsf{Pr}(A_j) + b\sum_{\ell=1}^n y_\ell \mathsf{Pr}(B_\ell) \\
&= a\mathsf{E}[X] + b\mathsf{E}[Y]
\end{aligned}$$

となる. 以上で命題は示せた. □

命題 1.20. X, Y を単純かつ独立な確率変数とし, $a, b \in \mathbb{R}$ とする. このとき

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

が成り立つ.

Proof. 命題 1.19 の証明の記号を踏襲する. まず

$$\begin{aligned}
&X \text{ と } Y \text{ は独立} \\
&\quad \Leftrightarrow \ \mathsf{Pr}(A_j \cap B_\ell) = \mathsf{Pr}(A_j)\mathsf{Pr}(B_\ell) \qquad (\forall j = 1, \ldots, m; \ell = 1, \ldots, n)
\end{aligned}$$

であることに注意する. すると

$$\mathsf{E}[XY] = \sum_{j=1}^m \sum_{\ell=1}^n x_j y_\ell \mathsf{Pr}(A_j \cap B_\ell) \sum_{j=1}^m \sum_{\ell=1}^n x_j y_\ell \mathsf{Pr}(A_j)\mathsf{Pr}(B_\ell) = \mathsf{E}[X]\mathsf{E}[Y]$$

がわかる. □

② 非負値確率変数に対する期待値

定義 1.21. 非負値確率変数 X の期待値を

$$\mathsf{E}[X] = \sup\bigl\{\mathsf{E}[Y];\, Y \text{ は単純な確率変数で } Y \leq X\bigr\}$$

で定める. ただし, $\mathsf{E}[X] = \infty$ も許す.

確率変数列 $\{X_n\}_{n=1}^{\infty} \nearrow X$ であるとは, $X_1 \leq X_2 \leq \cdots$ かつ $\lim_{n\to\infty} X(\omega) = X(\omega)\,(\omega \in \Omega)$ をみたすことである.

定理 1.22 (単調収束定理)**.** $\{X_n\}_{n=1}^{\infty}$ は確率変数列[5]で $\mathsf{E}[X_1] > -\infty$ かつ $\{X_n\}_{n=1}^{\infty} \nearrow X$ をみたすとする. このとき, X は確率変数で

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

が成立する.

Proof. 節 1.6.2 を参照せよ. □

定理 1.22 を用いるために, 関数 $\Psi_n(x)\colon [0,\infty) \longrightarrow [0,\infty)$ を

$$\Psi_n(x) = \min\left(n,\, \frac{1}{2^n}\lfloor 2^n x\rfloor\right)$$

で定める. ただし, $\lfloor r\rfloor$ $r \in \mathbb{R}$ を超えない最大の整数とする. すると $x \geq 0$ に対して, $\Psi_n(x) \geq 0$ かつ $\{\Psi_n(x)\}_{n=1}^{\infty} \nearrow x$ となることが直ちにわかる.

命題 1.23. X を非負値確率変数とし, $X_n = \Psi_n(X)$ とおく. このとき, $X_n \geq 0\,(n \in \mathbb{N})$, $\{X_n\} \nearrow X$ で各 X_n は単純確率変数となる.

Proof. 節 1.6.4 を参照せよ. □

③ 一般の確率変数に対する期待値: 確率変数 X に対して

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

と定める. すると X^+ と X^- はともに非負値確率変数となる. $\mathsf{E}[X^+] < \infty$ または $\mathsf{E}[X^-] < \infty$ のとき

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

と定める. $\mathsf{E}[X^+] = \infty$ かつ $\mathsf{E}[X^-] = \infty$ のとき, $\mathsf{E}[X]$ は定義されない.

すると, 一般の確率変数に対しても $\mathsf{E}[|X|] < \infty$, $\mathsf{E}[|Y|] < \infty$ のとき

$$\mathsf{E}[aX + bY] = a\mathsf{E}[X] + b\mathsf{E}[Y] \qquad (a,\, b \in \mathbb{R})$$

[5] ここでは, 非負値確率変数に対して期待値を定義したが, 一般の確率変数 X の期待値を ③ の流儀で定義することになる.

が成り立つ. さらに, X と Y が独立で $\mathsf{E}[|XY|] < \infty$ のとき

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

が成立する.

確率変数 X の分布を P^X とし, m を $\mathbb{R}$ 上の Lebesgue 測度とする. 非負値関数 p^X で $\int_{-\infty}^{\infty} \mathsf{p}^X(x)\,\mathrm{d}\mathsf{m}(x) = 1$ をみたすものが存在して

$$\mathsf{P}^X(B) = \int \mathsf{p}^X(x)\mathbb{1}_B(x)\,\mathrm{d}\mathsf{m}(x) \qquad (\forall B \in \mathcal{B}(\mathbb{R})) \tag{1.3}$$

とみたすとき, p^X を X **の確率密度関数** (p.d.f.) という.

命題 1.24 (統計学者の怠け公式)**.** 確率変数 X は, Lebesgue 測度に関する p.d.f. p^X を持つとする. 任意の可測関数 $h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して, $\mathsf{E}[h(X)]$ が定義されるとする. このとき

$$\mathsf{E}[h(X)] = \int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}\mathsf{m}(x)$$

と表現できる.

Proof. 任意の $B \in \mathcal{B}(\mathbb{R})$ に対して, $h = \mathbb{1}_B$ とおくと

$$\mathsf{E}[h(X)] = \mathsf{E}[\mathbb{1}_B(X)] = \mathsf{Pr}(B)$$

となる. 一方, (1.3) に注意すると

$$\int_{-\infty}^{\infty} h(x)\mathsf{p}^X(x)\,\mathrm{d}\mathsf{m}(x) = \int_{-\infty}^{\infty} \mathbb{1}_B(x)\mathsf{p}^X(x)\,\mathrm{d}\mathsf{m}(x) = \mathsf{P}^X(B) = \mathsf{Pr}(X \in B)$$

となり, この場合には定理の主張が正しい. あとは, 標準機械 ① ～ ③ の段階を踏めばよい. □

定義 1.25 (分散)**.** X を $\mathsf{E}[X^2] < \infty$ なる確率変数とする. このとき, X **の分散** $\mathsf{Var}[X]$ を

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

で定める.

注意 1.26. $\mathsf{E}[X^2] < \infty$ ならば, $\mathsf{E}[|X|] < \infty$ がわかる. なぜならば

$$|x| \le \max(1,\, |x|) \le 1 + x^2$$

なので

$$\mathsf{E}[|X|] \le \mathsf{E}\big[\max(1,\, |X|)\big] \le 1 + \mathsf{E}[X^2]$$

からわかる. よって, $\mathsf{E}[X^2] < \infty$ を仮定すると X の分散が定義されることがわかる. □

命題 1.27. X を非負値確率変数とする．このとき,

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

となる．

Proof. $\mathsf{Pr}(X = 0) = 1 \Rightarrow \mathsf{E}[X] = 0$ は期待値の定義から直ちにわかる．$\mathsf{E}[X] = 0 \Rightarrow \mathsf{Pr}(X = 0) = 1$ を背理法を用いて示す．そのために, $\mathsf{Pr}(X > 0) > 0$ と仮定する．するとある $\epsilon > 0$ が存在して, $\mathsf{Pr}(X > \epsilon) > 0$ となる．しかし, $X \geq \epsilon \mathbb{1}\{X > \epsilon\} := \mathbb{1}_{(\epsilon,\infty)}(X)$ より

$$0 = \mathsf{E}[X] \geq \epsilon \mathsf{E}[\mathbb{1}\{X > \epsilon\}] = \epsilon \mathsf{Pr}(X > \epsilon) > 0$$

となり, 矛盾する．よって, 命題は示せた． □

命題 1.28. X を非負値確率変数とし, $p \geq 1$ とする．このとき,

$$\mathsf{E}[X^p] = \int_0^\infty p x^{p-1} \, \mathsf{Pr}(X > x^{1/p}) \, \mathrm{d}x$$

となる．

Proof. 簡単のために, X は Lebesgue 測度に関する p.d.f. p^X を持つとして証明を与える．Fubini の定理を用いると

$$\begin{aligned}
\int_0^\infty p x^{p-1} \, \mathsf{Pr}(X > x^{1/p}) \, \mathrm{d}x &= \int_0^\infty p x^{p-1} \left\{ \int_{x^{1/p}}^\infty \mathsf{p}^X(t) \, \mathrm{d}t \right\} \mathrm{d}t \\
&= \int_0^\infty \mathsf{p}^X(t) \left\{ \int_0^t p x^{p-1} \, \mathrm{d}x \right\} \mathrm{d}t \\
&= \int_0^\infty t^p \mathsf{p}^X(t) \, \mathrm{d}t \\
&= \mathsf{E}[X^p]
\end{aligned}$$

よりわかる． □

1.4.2 期待値の不等式

命題 1.29 (Markov の不等式)**.** X は非負値確率変数とする．このとき

$$\mathsf{Pr}\big(X \geq t) \leq \frac{\mathsf{E}[X]}{t} \qquad (\forall t > 0)$$

が成立する．

Proof. $\mathsf{E}[X]=\infty$ のときは, 不等式は自明なので, $\mathsf{E}[X]<\infty$ として証明すればよい.

$$X = X\mathbb{1}_{[t,\infty)}(X) + X\mathbb{1}_{[0,t)}(X)$$

と書けることに注意する．すると

$$\begin{aligned}\mathsf{E}[X] &= \mathsf{E}\big[X\mathbb{1}_{[t,\infty)}(X)\big] + \underbrace{\mathsf{E}\big[X\mathbb{1}_{[0,t)}(X)\big]}_{\geq 0} \\ &\geq \mathsf{E}[X\mathbb{1}_{[t,\infty)}(X)] \geq t\mathsf{E}[\mathbb{1}_{[t,\infty)}(X)] = t\,\mathsf{Pr}(X\geq t)\end{aligned}$$

よりわかる. □

系 1.30. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X] < \infty\,(0<\sigma<\infty)$ とする．このとき $\forall t>0$ に対して

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

が成り立つ.

Proof. $(X-\mu)^2$ に対して Markov の不等式 (定理 1.29) を適用する．すると

$$\mathsf{Pr}\big(|X-\mu|\geq t\big) = \mathsf{Pr}\big((X-\mu)^2\geq t^2\big) \leq \frac{\mathsf{E}[(X-\mu)^2]}{t^2}$$

がわかる. □

定理 1.31. (Cauchy-Schwarz の不等式) 確率変数 X と Y は 2 次の有限な期待値を持つとき

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

となる.

Proof. 節 1.6.4 を参照せよ. □

定義 1.32. 関数 $g:\mathbb{R}\to\mathbb{R}$ が凸であるとは各 $x, y\in\mathbb{R}$ と $0\leq\forall t\leq 1$ に対して

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

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

定理 1.33. (Jensen の不等式) X を有限な期待値を持つ確率変数とする.
(1) 関数 $g : \mathbb{R} \to \mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

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

となる.
(2) g が concave のとき

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

となる.

Proof. $\forall x \in \mathbb{R}$ に対してある定数 $r \in \mathbb{R}$ が存在[6]して

$$g(\mathsf{E}[X]) + r\{x - \mathsf{E}[X]\} \leq g(x)$$

となる. x に X を代入して上の不等式の両辺の期待値を取れば

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

がわかる. □

系 1.34. (Young の不等式) $p, q > 1$ とし

$$\frac{1}{p} + \frac{1}{q} = 1$$

をみたすとする. このとき $\forall a, b > 0$ に対して

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

となる.

Proof. 関数 g を凸とし, $Y \sim \mathsf{Unif}(0, 1)$ とする. 可積分関数 $h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して $X = h(Y)$ とすれば

$$g\left(\int_0^1 h(y)\,\mathrm{d}y\right) = g(\mathsf{E}[X]) \leq \mathsf{E}[g(X)] = \int_0^1 g(h(y))\,\mathrm{d}y$$

を得る. ここで

$$g(x) = e^x, \quad h(y) = \begin{cases} p\log a & \left(0 \leq y < \frac{1}{p}\right) \\ q\log b & \left(\frac{1}{p} \leq y \leq 1\right) \end{cases}$$

[6]定理 A.42 を参照.

とおけば g は凸なので

$$
\begin{aligned}
ab &= \exp\left\{\frac{p\log a}{p} + \frac{q\log b}{q}\right\} = \exp\left\{\int_0^1 h(y)\,\mathrm{d}y\right\} \le \int_0^1 \exp\bigl\{h(y)\bigr\}\,\mathrm{d}y \\
&= \frac{1}{p}\exp\bigl\{p\log a\bigr\} + \frac{1}{q}\exp\bigl\{q\log b\bigr\} = \frac{a^p}{p} + \frac{b^q}{q}
\end{aligned}
$$

がわかる. □

定理 1.35. (1)(Hölder の不等式) p, q は $1 \le p \le +\infty$, $1 \le q \le +\infty$ と $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\bigl[|X|^p\bigr] < \infty$, $\mathsf{E}\bigl[|Y|^q\bigr] < \infty$ をみたすとする. このとき, $\mathbb{E}\bigl[|XY|\bigr] < \infty$ で

$$
\mathsf{E}\bigl[|XY|\bigr] \le \bigl\{\mathsf{E}\bigl[|X|^p\bigr]\bigr\}^{1/p}\bigl\{\mathsf{E}\bigl[|Y|^q\bigr]\bigr\}^{1/q}
$$

となる.
(2)(Minkowski の不等式) $1 \le p \le +\infty$ で $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\bigl[|X|^p\bigr] < \infty$, $\mathsf{E}\bigl[|Y|^p\bigr] < \infty$ をみたすとする. このとき, $\mathbb{E}\bigl[|X+Y|^p\bigr] < \infty$ で

$$
\bigl\{\mathsf{E}\bigl[|X+Y|^p\bigr]\bigr\}^{1/p} \le \bigl\{\mathsf{E}\bigl[|X|^p\bigr]\bigr\}^{1/p} + \bigl\{\mathsf{E}\bigl[|Y|^p\bigr]\bigr\}^{1/p}
$$

となる.

Proof. 節 1.6.5 を参照せよ. □

1.4.3 収束定理

定理 1.36 (Fatou の補題)**.** $C(> -\infty)$ を定数とし, $\{X_n\}_{n=1}^{\infty}$ を確率変数列とし, $X_n \ge C$ $(\forall n \in \mathbb{N})$ とする. このとき

$$
\mathsf{E}\left[\liminf_{n\to\infty} X_n\right] \le \liminf_{n\to\infty} \mathsf{E}\bigl[X_n\bigr]
$$

が成り立つ. ただし, 不等式は両辺が $+\infty$ の場合も含めている.

Proof. $Y_n := \inf_{k\ge n} X_k$ とし

$$
Y = \lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n
$$

とおく. このとき, $Y_n \ge C$ で $\{Y_n\}_{n=1}^{\infty} \nearrow Y$ となる. さらに, $Y_n \le X_n$ である. 期待値の順序保存性と単調収束定理から

$$
\liminf_{n\to\infty} \mathsf{E}\bigl[X_n\bigr] \ge \liminf_{n\to\infty} \mathsf{E}\bigl[Y_n\bigr] = \mathsf{E}[Y]
$$

がわかる. よって, 補題の主張は証明された. □

定理 1.37. X, X_1, X_2, ... を確率変数列とし, $X_n \overset{\text{a.s.}}{\to} X$ とする. さらにある確率変数 Y が存在して, $|X_n| \le Y$ $(\forall n \in \mathbb{N})$ かつ $\mathsf{E}[Y] < \infty$ とする. このとき

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

が成り立つ.

Proof. 節 1.6.6 を参照せよ. □

定理 1.38 (Fubini の定理)**.** X を確率空間 $\big(\Omega_1 \times \Omega_2,\, \mathcal{A}_1 \times \mathcal{A}_2,\, \mathsf{Pr}_1 \times \mathsf{Pr}_2\big)$ 上の非負値確率変数とする. このとき, 以下が成立する.
(1) 関数 $\Omega_1 \ni \omega_1 \mapsto Y_1(\omega_1) = \mathsf{E}_{\mathsf{Pr}_2}\big[X(\omega_1,\, \cdot\,)\big]$ は確率空間 $\big(\Omega_1,\, \mathsf{A}_1,\, \mathsf{Pr}_1\big)$ 上の確率変数である.
(2) 関数 $\Omega_2 \ni \omega_2 \mapsto Y_2(\omega_2) = \mathsf{E}_{\mathsf{Pr}_1}\big[X(\,\cdot\,,\, \omega_2)\big]$ は確率空間 $\big(\Omega_2,\, \mathsf{A}_2,\, \mathsf{Pr}_2\big)$ 上の確率変数である.
(3) $\mathsf{E}_{\mathsf{Pr}_1}[Y_1] = \mathsf{E}_{\mathsf{Pr}_1 \times \mathsf{Pr}_2}[X] = \mathsf{E}_{\mathsf{Pr}_2}[Y_2]$.

Proof. 条件 (1) ∼ (3) をみたす非負値確率変数の集合は Dynkin 族であることを示し, あとは標準機械を用いればよい. □

定理 1.39 (Fubini の定理)**.** X を確率空間 $\big(\Omega_1 \times \Omega_2,\, \mathcal{A}_1 \times \mathcal{A}_2,\, \mathsf{Pr}_1 \times \mathsf{Pr}_2\big)$ 上の確率変数で $\mathsf{E}_{\mathsf{Pr}_1 \times \mathsf{Pr}_2}[|X|] < \infty$ とする. このとき, 以下が成立する.
(1) $\mathsf{Pr}_1 -$ a.s. な ω_1 に対して, 確率変数 $X(\omega_1,\, \cdot\,)$ は $\mathsf{E}_{\mathsf{Pr}_2}[|X(\omega_1,\, \cdot\,)|] < \infty$ をみたし

$$Y_1(\omega_1) = \begin{cases} \mathsf{E}_{\mathsf{Pr}_2}[X(\omega_1,\, \cdot\,)] & (X(\omega_1,\, \cdot\,) \in L^1(\mathsf{Pr}_1)) \\ 0 & (\text{その他の場合}) \end{cases}$$

は $L^1(\mathsf{Pr}_1)$ に属する確率変数である.
(1) $\mathsf{Pr}_2 -$ a.s. な ω_2 に対して, 確率変数 $X(\omega_1,\, \cdot\,)$ は $\mathsf{E}_{\mathsf{Pr}_1}[|X(\,\cdot\,, \omega_2)|] < \infty$ をみたし

$$Y_2(\omega_2) = \begin{cases} \mathsf{E}_{\mathsf{Pr}_1}[X(\,\cdot\,, \omega_2)] & (X(\,\cdot\,, \omega_2) \in L^1(\mathsf{Pr}_2)) \\ 0 & (\text{その他の場合}) \end{cases}$$

は $L^1(\mathsf{Pr}_2)$ に属する確率変数である.
(3) $\mathsf{E}_{\mathsf{Pr}_1}[Y_1] = \mathsf{E}_{\mathsf{Pr}_1 \times \mathsf{Pr}_2}[X] = \mathsf{E}_{\mathsf{Pr}_2}[Y_2]$.

Proof. 証明は定理 1.38 と同じ方針である. □

1.5 条件付き期待値の定義とその基本的性質

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.

$\big(\Omega, \mathcal{A}, \mathsf{Pr}\big)$ 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

$$\mathsf{Pr}(B|\,A) = \frac{\mathsf{Pr}(A\cap B)}{\mathsf{Pr}(A)} \tag{1.4}$$

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 1.1. $\mathsf{Pr}(A\cap B) = \mathsf{Pr}(A)\mathsf{Pr}(B) \;\Leftrightarrow\; \mathsf{Pr}(B|\,A) = \mathsf{Pr}(B|\,A^c)$ 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 1.2. $\mathcal{F}_A$ 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

$$\mathsf{Pr}(B|\,\mathcal{F}_A)(\omega) = \begin{cases} \mathsf{Pr}(B|\,A) & (\omega\in A) \\ \mathsf{Pr}(B|\,A^c) & (\omega\in A^c) \end{cases} \tag{1.5}$$

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

問 1.3. $\mathsf{Pr}(B|\,A)$ 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

$$\mathsf{E}\big[X|\,A\big] = \frac{\mathsf{E}\big[X\mathbb{1}_A\big]}{\mathsf{Pr}(A)} \tag{1.6}$$

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 1.4. $\mathsf{E}\big[X|\,A\big] = a_1$, $\mathsf{E}\big[X|\,A^c\big] = a_2\,(a_1 < a_2)$ 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

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

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 1.40. 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

$$\mathsf{E}\big[h(Y)g(X)\big] = \mathsf{E}\big[h(Y)g_Y(Y)\big] \tag{1.7}$$

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 1.41. 条件 (1.7) を証明するためには h を指示関数と限定して, (1.7) 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 (1.7) が成立することがわかる.

注意 1.42. 定義 1.40 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.13) から保証されることがわかる.

問 1.5. g_Y の一意性を直接的に (Radon-Nikodym の定理 (定理 A.13) を経由せず) 証明せよ.

例 1.43. 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 $\mathsf{E}\big[g(X)|\,Y = y\big]$ 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 1.40 の等式 (1.6) の左辺は

$$\begin{aligned}
&\mathsf{E}\big[h(Y)g(X)\big] \\
&\quad = \int_{\mathbb{R}^2} h(y)g(x)\mathsf{p}^{(X,Y)}(x,\,y)\,\mathrm{d}x\,\mathrm{d}y \\
&\quad = \int_{\{y\in\mathbb{R};\,\mathsf{p}^Y(y)>0\}} h(y)\left\{\int_{-\infty}^{\infty} g(x)\frac{\mathsf{p}^{(X,Y)}(x,\,y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x\right\}\cdot \mathsf{p}^Y(y)\,\mathrm{d}y
\end{aligned} \tag{1.8}$$

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

$$\mathsf{E}\big[h(Y)g^Y(Y)\big] = \int_{\{y\in\mathbb{R};\, \mathsf{p}_Y(y)>0\}} h(y)g^Y(y)\cdot \mathsf{p}^Y(y)\,\mathrm{d}y \tag{1.9}$$

である. (1.8) と (1.9) の右辺を比較すると

$$g^Y(y) = \int_{-\infty}^{\infty} g(x)\frac{\mathsf{p}^{(X,Y)}(x,\,y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x \tag{1.10}$$

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

$$\begin{aligned}
\mathsf{Pr}(X\in B|\,Y=y) &= \mathsf{E}[\mathbb{1}_B(X)|\,Y=y]\\
&= g_Y(y) \quad \big(\because (1.7)\big)\\
&= \int_{-\infty}^{\infty}\mathbb{1}_B(x)\frac{\mathsf{p}^{(X,Y)}(x,\,y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x \quad \big(\because (1.10)\big)\\
&= \int_B \frac{\mathsf{p}^{(X,Y)}(x,\,y)}{\mathsf{p}^Y(y)}\,\mathrm{d}x
\end{aligned}$$

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 1.44. 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 1.6. $a_1,\,a_2\in\mathbb{R}$, $A\in\mathcal{A}$ 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 1.45. (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 $\mathsf{E}[g(X)|\,Y] = \mathsf{E}[g(X)]$.

Proof. 節 1.6.7 を参照のこと. □

注意 1.46. (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.

1.6 証明

1.6.1 補題 1.5 の証明

(1) $F_1 := \Omega$, $F_n := \varnothing\,(n \geq 2)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となる．定義 1.7(2) を用いると

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

を得る．よって

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

がわかる．

(2) $F_1 := A$, $F_2 := A^c$, $F_n = \varnothing\,(n \geq 3)$ とおく．すると $\bigcup_{n=1}^{\infty} F_n = \Omega$ かつ $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列となるので, 定義 1.7(1), (2) を用いると

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

がわかる．よって, (2) は示せた．

(3) $F_i := A_i\,(i = 1, 2, \ldots, N)$ と $F_i = \varnothing\,(i \geq N+1)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = \bigcup_{i=1}^{N} A_i$ となる．定義 1.7(2) を用いると

$$\begin{aligned}\mathsf{Pr}\left(\bigcup_{i=1}^{N} A_i\right) &= \mathsf{Pr}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{i=1}^{\infty} \mathsf{Pr}(F_i) = \sum_{i=1}^{N} \mathsf{Pr}(A_i) + \sum_{i=N+1}^{\infty} \mathsf{Pr}(\varnothing) \\ &= \sum_{i=1}^{N} \mathsf{Pr}(A_i)\end{aligned}$$

を得る．よって, (3) は示された．

(4) $B \setminus A = B \cap A^c$ かつ $B = (B \cap A^c) \cup A$ である．このことに注意して, $F_1 := B \cap A^c$, $F_2 = A$, $F_n = \varnothing\,(n \geq 3)$ とおくと $\{F_n\}_{n=1}^{\infty}$ は互いに排反な事象列で $\bigcup_{n=1}^{\infty} F_n = B$ となる．定義 1.7(2) を用いると

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

がわかる．よって, (4) は示せた．

(5) $F_1 := A_1$, $F_2 = A_2 \setminus A_1$, ..., $F_{n+1} := A_{n+1} \setminus A_n$ とおく. $\{F_n\}_{n=1}^{\infty}$ は互いに排反であり

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

となる. よって

$$\begin{aligned}
\mathsf{Pr}(A) &= \mathsf{Pr}\biggl(\bigcup_{i=1}^{\infty} F_i\biggr) = \sum_{i=1}^{\infty} \mathsf{Pr}(F_i) \quad (\because \text{定義 } 1.7(2)) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{Pr}(F_i) = \lim_{n\to\infty} \mathsf{Pr}\biggl(\bigcup_{i=1}^{n} F_i\biggr) \quad (\because (3)) \\
&= \lim_{n\to\infty} \mathsf{Pr}(A_n)
\end{aligned}$$

を得る. よって, (6) は示せた.

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

$$\bigcup_{n=1}^{\infty} (A_1 \setminus A_n) = \bigcup_{n=1}^{\infty} (A_1 \cap A_n^c) = A_1 \cap \biggl(\bigcap_{n=1}^{\infty} A_n\biggr)^c = A_1 \cap A^c = A_1 \setminus A$$

となる. 次に (4) と (6) を用いると

$$\begin{aligned}
\mathsf{Pr}(A_1) - \mathsf{Pr}(A) &= \mathsf{Pr}(A_1 \setminus A) \quad (A_1 \supset A \text{ なので, (4) を用いた}) \\
&= \lim_{n\to\infty} \mathsf{Pr}(A_1 \setminus A_n) \quad (\because (6) \text{ を用いた}) \\
&= \lim_{n\to\infty} \Bigl(\mathsf{Pr}(A_1) - \mathsf{Pr}(A_n)\Bigr) \\
&= \mathsf{Pr}(A_1) - \lim_{n\to\infty} \mathsf{Pr}(A_n)
\end{aligned}$$

がわかる.

(7) まず

$$\mathsf{Pr}(A_1 \cup A_2) \le \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \tag{1.11}$$

を示せす. そのために

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

であることに注意する. (3) から

$$\begin{aligned}
\mathsf{Pr}\bigl(A_1 \cup A_2) &= \mathsf{Pr}(A_1) + \mathsf{Pr}\Bigl(A_2 \setminus (A_1 \cap A_2)\Bigr) \\
&= \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) - \mathsf{Pr}(A_1 \cap A_2) \\
&\qquad \bigl(\because A_2 \supset A_1 \cap A_2 \text{ なので (4) を用いた}\bigr) \\
&\le \mathsf{Pr}(A_1) + \mathsf{Pr}(A_2) \qquad \bigl(\because \mathsf{Pr}(A_1 \cap A_2) \ge 0\bigr)
\end{aligned}$$

がわかる. (1.11) の操作を繰り返せば, $N \in \mathbb{N}$ に対して

$$\mathsf{Pr}\left(\bigcup_{n=1}^{N} A_n\right) \le \sum_{n=1}^{N} \mathsf{Pr}(A_n) \tag{1.12}$$

がわかる. $\{\cup_{n=1}^{N} A_n\}_{N=1}^{\infty}$ は増加列なので, (5) と (1.12) から

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

がわかる. □

1.6.2 定理 1.22 の証明

① X は確率変数であることの証明: $\forall x \in \mathbb{R}$ に対して

$$\{X \le x\} = \bigcap_{n=1}^{\infty} \{X_n \le x\} \in \mathsf{A}$$

となるので, X は確率変数である.
② $\lim_{n\to\infty} E[X_n]$ の存在の証明: 単調性から $\mathsf{E}[X_1] \le \mathsf{E}[X_2] \le \cdots \mathsf{E}[X]$ となる. よって, $\lim_{n\to\infty} E[X_n]$ は $+\infty$ も含めて存在する.
② $\lim_{n\to\infty} E[X_n] = \mathsf{E}[X]$ の存在の証明: $\lim_{n\to\infty} \mathsf{E}[X_n] \ge \mathsf{E}[X]$ を示せばよい. $\mathsf{E}[X_1] = \infty$ のときは, 自明なので, $\mathsf{E}[X_1] < \infty$ と仮定して証明すればよい. $Z_n := X_n - X_1$ と $Z := X - X_1$ とおく. 仮定より Z は非負値確率変数となる. よって, $Y \le Z$ なる単純確率変数 Y に対して

$$\lim_{n\to\infty} \mathsf{E}[Z_n] \ge \mathsf{E}[Y]$$

を示せばよい. いま, $Y = \sum_{j=1}^{n} y_j \mathbb{1}_{A_j}$ $(x_j \in \mathbb{R},\ A_j \in \mathcal{A})$ と書く. すると, $Z(\omega) \ge y_j$ $(\omega \in A_j)$ のとき

$$\mathsf{E}[Z_n] \ge \sum_{j=1}^{n} y_j \mathsf{Pr}(A_j)$$

を示せばよい.

任意の $\epsilon > 0$ に対して

$$A_{jn} := \left\{\omega \in A_j;\ X_n(\omega) \ge y_j - \epsilon\right\}$$

と定める. $n \to \infty$ のとき, $\{A_{jn}\} \nearrow A_j$ となる. さらに

$$\mathsf{E}[X_n] \ge \sum_{j=1}^{n} (y_j - \epsilon) \mathsf{Pr}(A_{jn})$$

である. 確率の連続性から

$$\lim_{n\to\infty}\sum_{j=1}^{n}(y_j-\epsilon)\mathsf{Pr}(A_{jn})=\sum_{j=1}^{n}(y_j-\epsilon)\mathsf{Pr}(A_j)$$

となる. したがって

$$\mathsf{E}[X_n]\ge\sum_{j=1}^{n}y_j\mathsf{Pr}(A_j)-\epsilon$$

となる. この不等式は, 任意の $\epsilon>0$ に対して成立するので

$$\mathsf{E}[X_n]\ge\sum_{j=1}^{n}y_j\mathsf{Pr}(A_j)$$

がわかる. よって, ③ が証明できた. □

1.6.3 命題 1.23 の証明

X, Y を非負値確率変数とし, a, $b\ge 0$ とする. このとき, 命題 1.19, 1.23 から

$$\begin{aligned}\mathsf{E}\big[aX+bY\big]&=\lim_{n\to\infty}\mathsf{E}\big[\Psi_n(aX+bY)\big]=\lim_{n\to\infty}\mathsf{E}\big[a\Psi_n(X)+b\Psi_n(Y)\big]\\&=\lim_{n\to\infty}\Big\{a\mathsf{E}[\Psi_n(X)]+b\mathsf{E}[\Psi_n(Y)]\Big\}\\&=a\lim_{n\to\infty}\mathsf{E}[\Psi_n(X)]+b\lim_{n\to\infty}\mathsf{E}[\Psi_n(Y)]=a\mathsf{E}[X]+b\mathsf{E}[Y]\end{aligned}$$

がわかる. 同様に, X と Y が独立のとき

$$\mathsf{E}[XY]=\lim_{n\to\infty}\mathsf{E}[\Psi_n(XY)]=\lim_{n\to\infty}\mathsf{E}[\Psi_n(X)]\mathsf{E}[\Psi_n(Y)]=\mathsf{E}[X]\mathsf{E}[Y]$$

がわかる.

X_1, X_2, ... を非負値確率変数列としたとき, 単調収束定理から

$$\begin{aligned}\mathsf{E}\big[\sum_{k=1}^{\infty}X_k\big]&=\lim_{n\to\infty}\mathsf{E}\left[\Psi_n\left(\sum_{k=1}^{\infty}X_k\right)\right]=\lim_{n\to\infty}\left\{\sum_{k=1}^{\infty}\mathsf{E}[\Psi_n(X_k)]\right\}\\&=\sum_{k=1}^{\infty}\lim_{n\to\infty}\mathsf{E}[\Psi_n(X_k)]=\sum_{k=1}^{\infty}\mathsf{E}[X_k]\end{aligned}$$

がわかる.

□

1.6.4 定理 1.31 の証明

$\mathsf{E}[X^2] = \mathsf{E}[Y^2] = 0$ のとき不等式は自明である. $\mathsf{E}[X^2] \neq 0$ として証明を進める. いま, $g(t) = \mathsf{E}\big[(tX - Y)^2\big]$ とおく. 期待値の中を展開して期待値の線型性を用いると

$$
\begin{aligned}
0 \le g(t) &= \mathsf{E}\big[t^2X^2 - 2tXY + Y^2\big] \\
&= \mathsf{E}[X^2]\left\{t - \frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right\}^2 + \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]}
\end{aligned}
$$

となる.

$$
\begin{aligned}
g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right) = \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]} \ge 0 \;&\Leftrightarrow\; \mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2 \ge 0 \\
&\Leftrightarrow\; \Big|\mathsf{E}[XY]\Big| \le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]}
\end{aligned}
\tag{1.13}
$$

を得る. (1.13) において, X, Y を $|X|, |Y|$ と置き換えると定理の不等式は示される. 等号が成立するのは $g(t) = 0$ が重解を持つときである. 重解を c とおけば

$$
g(t) = \mathsf{E}[(cX - Y)^2] = 0 \;\Leftrightarrow\; \mathsf{Pr}(Y = cX) = 1
$$

となる[7]. □

1.6.5 定理 1.35 の証明

(1) の証明: $1 < p < \infty$, $1 < q < \infty$ の場合についてのみ証明を与える. $p = 1$, $q = \infty$ の場合は Cohen (2010, pp.93-94) を参照のこと.

系 1.34 から

$$
|XY| \le \frac{|X|^p}{p} + \frac{|Y|^q}{q} \tag{1.14}
$$

[7]非負値確率変数 X に対して

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

であることに注意せよ. 実際 $\mathsf{Pr}(X > 0) > 0$ と仮定する. するとあるの $\epsilon > 0$ が存在して $\mathsf{Pr}(X > \epsilon) > 0$ となる. しかし $X \ge \epsilon \mathbb{1}\{X > \epsilon\}$ より

$$
0 = \mathsf{E}[X] \ge \epsilon \mathsf{E}[\mathbb{1}\{X > \epsilon\}] = \epsilon \mathsf{Pr}(X > \epsilon) > 0
$$

となり矛盾する. よって $\mathsf{E}[X] = 0 \Leftrightarrow \mathsf{Pr}(X = 0) = 1$ がわかる.

である．まず, $\mathsf{E}[|X|^p] = 1$, $\mathsf{E}[|Y|^q] = 1$ とする．(1.14) の両辺の期待値を取ると

$$\mathsf{E}[|XY|] \leq \frac{\mathsf{E}[|X|^p]}{p} + \frac{\mathsf{E}[|Y|^q]}{q} = \frac{1}{p} + \frac{1}{q} = 1 \tag{1.15}$$

を得る．$\mathsf{E}[|X|^p] \neq 1$ または $\mathsf{E}[|Y|^q] \neq 1$ のとき, (1.15) において, X, Y を $X/\{\mathsf{E}[|X|^p]\}^{1/p}$, $Y/\{\mathsf{E}[|Y|^q]\}^{1/q}$ と置き換えると

$$\frac{\mathsf{E}[|\mathsf{XY}|]}{\{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|Y|^q]\}^{1/q}} \leq 1$$

を得る．よって, $1 < p < \infty$, $1 < q < \infty$ の場合について (1) は証明された.

(2) の証明: $1 < p < \infty$, $1 < q < \infty$ の場合についてのみ証明を与える．$p = 1$, $q = \infty$ の場合は Cohen (2010, pp.94-95) を参照のこと. $\mathsf{E}[|X + Y|^p] = 0$ のときは, 不等式は自明なので, $\mathsf{E}[|X + Y|^p] \neq 0$ を仮定しても一般性を失わない．このとき

$$\begin{aligned}
\mathsf{E}[|X+Y|^p] &\leq \mathsf{E}\big[\big(|X|+|Y|\big)|X+Y|^{p-1}\big] \\
&= \mathsf{E}\big[|X| \times |X+Y|^{p-1}\big] + \mathsf{E}\big[|Y| \times |X+Y|^{p-1}\big] \\
&\leq \{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \\
&\qquad + \{\mathsf{E}[|Y|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^{q(p-1)}]\}^{1/q} \qquad (\because (1) \text{ を用いた}) \\
&= \{\mathsf{E}[|X|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^p]\}^{1/q} + \{\mathsf{E}[|Y|^p]\}^{1/p}\{\mathsf{E}[|X+Y|^p]\}^{1/q} \\
&= \Big\{\{\mathsf{E}[|X|^p]\}^{1/p} + \{\mathsf{E}[|Y|^p]\}^{1/p}\Big\}\{\mathsf{E}[|X+Y|^p]\}^{1/q}
\end{aligned}$$

を得る．上の不等式の最左辺と最右辺を $\{\mathsf{E}[|X + Y|^p]\}^{1/q}$ で割れば, $1 < p < \infty$, $1 < q < \infty$ の場合の (2) が示せた. □

1.6.6 定理 1.37 の証明

Z すべての $n \in \mathbb{N}$ に対して, $Y + X_n \geq 0$ となることに注意する．確率変数列 $\{Y + X_n\}_{n=1}^{\infty}$ に Fatou の補題 (定理 1.36) を適用すると

$$\mathsf{E}[Y] + \mathsf{E}[X] = \mathsf{E}[Y + X] \leq \liminf_{n\to\infty} \mathsf{E}\big[Y + X_n\big] = \mathsf{E}[Y] + \liminf_{n\to\infty} \mathsf{E}\big[X_n\big]$$

がわかる．したがって,

$$\mathsf{E}[X] \leq \liminf_{n\to\infty} \mathsf{E}\big[X_n\big] \tag{1.16}$$

がわかる.

同様に $Y - X_n \geq 0\,(\forall n \in \mathbb{N})$ に注意して, Fatou の補題を $\{Y - X_n\}_{n=1}^{\infty}$ に適用すると

$$\mathsf{E}[Y] - \mathsf{E}[X] \leq \mathsf{E}[Y - X] = \mathsf{E}[Y] + \liminf_{n\to\infty} \mathsf{E}\big[-X_n\big] = \mathsf{E}[Y] - \limsup_{n\to\infty} \mathsf{E}\big[X_n\big]$$

がわかる. がわかる. よって

$$\mathsf{E}[X] \geq \limsup_{n\to\infty} \mathsf{E}\big[X_n\big] \tag{1.17}$$

がわかる. (1.16) と (1.17) を合わせると

$$\limsup_{n\to\infty} \mathsf{E}\big[X_n\big] \leq \liminf_{n\to\infty} \mathsf{E}\big[X_n\big]$$

がわかる. しかし

$$\limsup_{n\to\infty} \mathsf{E}\big[X_n\big] \geq \liminf_{n\to\infty} \mathsf{E}\big[X_n\big]$$

なので

$$\limsup_{n\to\infty} \mathsf{E}\big[X_n\big] = \liminf_{n\to\infty} \mathsf{E}\big[X_n\big] = \mathsf{E}[X]$$

がわかる. よって, 定理は証明された. □

1.6.7 定理 1.45 の証明

Proof. (i) の証明: (1.7) 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. (1.7) 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 $g_1(x) \geq 0$ ならば

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

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

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

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

$$
\begin{aligned}
\mathsf{E}\big[h(Y)\mathsf{E}[g(X)| Y]\big] &= \mathsf{E}[h(Y)g(X)] \qquad \big(\because (1.7)\big) \\
&= \mathsf{E}[h(Y)]\mathsf{E}[g(X)] \qquad \big(\because X \text{ と } Y \text{ は独立}\big) \\
&= \mathsf{E}\big[h(Y)]\mathsf{E}[g(X)]\big]
\end{aligned}
$$

となる．すると (1.7) と一意性から

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

を得る． □