第A章　補遺

A.1　Dynkin の定理

定義 A.1. Ω を空でない集合とし, $\mathcal{L}$ と $\mathcal{P}$ を Ω の部分集合族とする.
(1) 集合族 $\mathcal{P}$ が π **システム**であるとは

$$A,\, B \in \mathcal{P} \Longrightarrow A \cap B \in \mathcal{P}$$

が成立するときをいう.
(2) 集合族 $\mathcal{L}$ が λ **システム**であるとは, つぎの条件 (i) $\sim$ (iii) が成り立つときをいう.

(i) $\Omega \in \mathcal{L}$.

(ii) $A,\, B \in \mathcal{L}$ かつ $A \subset B \Longrightarrow B \setminus A \in \mathcal{L}$.

(iii) $A_j \in \mathcal{L}$ かつ $A_j \subset A_{j+1}\,(j = 1,\, 2,\, \ldots) \Longrightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$.

命題 A.2. 空でない集合 Ω の部分集合族 $\mathcal{L}$ が λ システムであるために必要十分条件は次の条件 (i) $\sim$ (iii) をすべてみたすことである.

(i)$'$ $\Omega \in \mathcal{L}$.

(ii)$'$ $A \in \mathcal{L} \Longrightarrow A^c \in \mathcal{L}$.

(iii)$'$ $A_j \in \mathcal{L}\,(j = 1,\, 2,\, \ldots)$ かつ $A_i \cap A_j \neq \varnothing\,(i \neq j) \Longrightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$.

Proof. ($\Rightarrow$): <u>(ii)$'$ の確認</u>. $A \in \mathcal{L}$ のとき, $A \subset \Omega$ かつ $\Omega \in \mathcal{L}$ なので, 定義 A.1(ii) から

$$A^c = \Omega \setminus A \in \mathcal{L}$$

となる.
<u>(iii)$'$ の確認</u>. まず, $A,\, B \in \mathcal{L}$ かつ $A \cap B \neq \varnothing$ のとき

$$A \cup B \in \mathcal{L} \tag{A.1}$$

を示す.

仮定と $(\mathrm{ii})'$ から $A^c \in \mathcal{L}$ かつ $B \subset A^c$ である．なぜならば, $\forall \omega \in B$ のとき, $A \cap B = \varnothing$ なので, $\omega \notin A$ となる．よって $\omega \in A^c$ がわかるので, $B \subset A^c$ が示せた．

このことと (ii) から $A^c \setminus B \in \mathcal{L}$ となる．さらに

$$A^c \cap B^c = A^c \setminus B \in \mathcal{L} \tag{A.2}$$

となることと $(\mathrm{ii})'$ より

$$A \cup B = \bigl(A^c \cap B^c\bigr)^c \in \mathcal{L}$$

がわかる．よって

$$A,\, B \in \mathcal{L} \implies A \cup B \in \mathcal{L} \tag{A.3}$$

がわかる．

$A_j \in \mathcal{L}\,(j = 1,\, 2, \ldots)$ と $A_i \cap A_j = \varnothing\,(i \neq j)$ とする．このとき

$$B_n := \bigcup_{j=1}^{n} A_j \quad (n \in \mathbb{N})$$

とおく．(A.3) より $B_n \in \mathcal{L}$ となる．さらに, B_n の定め方から

$$B_n \subset B_{n+1}$$

となる．よって定理 A.1(iii) から

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{n=1}^{\infty} B_n \in \mathcal{L}$$

がわかる．よって $(\mathrm{i})' \sim (\mathrm{iii})'$ が確認できた．

$(\Leftarrow)$: <u>(ii) の確認</u>. $A,\, B \in \mathcal{L}$ かつ $A \subset B$ とする．すると $(\mathrm{ii})'$ から $A^c,\, B^c \in \mathcal{L}$ となる．さらに $A \subset B$ から $A \cap B^c = \varnothing$ となる．よって $(\mathrm{iii})'$ から $A \cup B^c \in \mathcal{L}$ となる．再度, $(\mathrm{ii})'$ を用いると

$$B \setminus A = B \cap A^c = \bigl(A \cup B^c\bigr)^c \in \mathcal{L}$$

がわかる．

<u>(iii) の確認</u>. $A_n \in \mathcal{L}$ かつ $A_n \subset A_{n+1}\,(n = 1,\, 2,\, \ldots)$ とする．ここで

$$B_n := A_{n+1} \setminus A_n\,(n = 1,\, 2,\, \ldots)$$

とおくと (ii) から $B_n \in \mathcal{L}$ となる．さらに B_n の定め方から

$$B_n \cap B_m = \varnothing\,(m \neq n)$$

となる. よって (iii)$'$ から

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{n=1}^{\infty} B_n \in \mathcal{L}$$

がわかる. 以上から (i) ∼ (iii) が確認できた. □

命題 A.3. Ω を空でない集合とし, $\mathcal{C}$ を Ω の部分集合族とする. 部分集合族 $\mathcal{C}$ が π システムかつ λ システムならば, $\mathcal{C}$ は σ 集合族となる.

Proof. まず, 部分集合族 $\mathcal{C}$ は体であることを示す.
(i) $\mathcal{C}$ は λ システムなので, 命題 A.2(ii)$'$ から $\Omega \in \mathcal{C}$ である.
(ii) $\mathcal{C}$ は π システムである. $A \in \mathcal{C}$ のとき, $\Omega \in \mathcal{C}$ なので, 命題 A.2(ii)$'$ から

$$A^c = \Omega \setminus A \in \mathcal{C}$$

となる.
(iii) $A_j \in \mathcal{C}\,(j = 1, 2, \ldots)$ とする. $\mathcal{C}$ は π システムなので, 定義 A.1(1) から

$$\bigcap_{j=1}^{n} A_j \in \mathcal{C}\,(n \in \mathbb{N})$$

となる. よって, (i) ∼ (iii) から $\mathcal{C}$ は体となることがわかる.
　$\mathcal{C}$ は体であるので, あとは

$$\bigcup_{j=1}^{n} A_i \in \mathcal{C}$$

を示せばよい. $\mathcal{C}$ は体なので

$$B_n := \bigcup_{j=1}^{n} A_j \in \mathcal{C} \qquad (n \in \mathbb{N})$$

となる. さらに B_n の定め方から $B_n \subset B_{n+1}\,(n = 1, 2, \ldots)$ となる. よって, $\mathcal{C}$ は λ システムなので, 定義 A.1(iii) から

$$\bigcup_{j=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{C}$$

がわかる. 以上から $\mathcal{C}$ は σ 集合族であることがわかる. □

命題 A.4. Ω を空でない集合とする. Ω の部分集合族 $\mathcal{P}$ を π システムとし, $\mathcal{L}[\mathcal{P}]$ を $\mathcal{P}$ を含む最小の λ システムとする. このとき, $\mathcal{L}[\mathcal{P}]$ は π システムとなる.

Proof. $\sigma[\mathcal{P}]$ を $\mathcal{P}$ を含む最小の σ 加法族とする. $A \in \mathcal{L}[\mathcal{P}]$ を固定し, この A に対して

$$\mathcal{G}_A := \left\{B \in \sigma[\mathcal{P}];\, A \cap B \in \mathcal{L}[\mathcal{P}]\right\}$$

と定める. $\mathcal{G}_A$ に関する以下の主張を順に証明していく.
主張 1: $A \in \sigma[\mathcal{P}]$ のとき, $\mathcal{G}_A$ は λ システムである.

主張 1 の証明: 命題 A.2 の条件 (i)$'$ $\sim$ (iii)$'$ を確認する.
(i)$'$ $A \cap \Omega = A \in \mathcal{L}[\mathcal{P}]$ なので, $\mathcal{G}_A$ の定義から $\Omega \in \mathcal{G}_A$ がわかる.
(ii)$'$ $B \in \mathcal{G}_A$ とする. このとき

$$B^c \cap A = A \setminus (A \cap B)$$

に注意する. $B \in \mathcal{G}_A$ なので

$$A \cap B \in \mathcal{L}[\mathcal{P}] \tag{A.4}$$

である. 仮定と (A.4) から A, $A \cap B \in \mathcal{L}[\mathcal{P}]$ である. $\mathcal{L}[\mathcal{P}]$ は λ システムなので, 定義 A.1(2)(ii) から

$$B^c \cap A = A \setminus (A \cap B) \in \mathcal{L}[\mathcal{P}]$$

となる. よって $B^c \in \mathcal{L}[\mathcal{P}]$ がわかる.
(iii)$'$ $\{B_n\}_{n=1}^{\infty}$ は互いに排反な列で $B_n \in \mathcal{G}_A\,(n = 1, 2, \ldots)$ とする. このとき

$$A \cap \left(\bigcup_{j=1}^{\infty} B_j\right) = \bigcup_{j=1}^{\infty} \bigl(A \cap B_j\bigr) \tag{A.5}$$

となる. したがって $\{A \cap B_n\}_{n=1}^{\infty}$ は互いに排反な列で, (A.4) から $A \cap B_n \in \mathcal{G}_A$ となることに注意する. よって, $A \cap B_n \in \mathcal{L}[\mathcal{P}]$ かつ $\mathcal{L}[\mathcal{P}]$ は λ システムなので, 命題 A.2(ii)$'$ から

$$\bigcup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{L}[\mathcal{P}]$$

となる. このことと (A.5) から

$$\bigcup_{n=1}^{\infty} B_n \in \mathcal{G}_A$$

がわかる．

よって (i)$'$ ∼ (iii)$'$ から $\mathcal{G}_A$ は λ システムであることがわかる． □

主張 2: $A \in \mathcal{P}$ ならば $\mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$ である．

主張 2 の証明: $A \in \mathcal{P} \subset \mathcal{L}[\mathcal{P}]$ を取る．主張 1 から $\mathcal{G}_A$ は λ システムとなる．$B \in \mathcal{P}$ とする．$\mathcal{P}$ は π システムなので, 定義 A.1(1) から

$$A \cap B \in \mathcal{P} \subset \mathcal{L}[\mathcal{P}]$$

となるので, $\mathcal{G}_A$ は λ システムで $\mathcal{P}$ を含むので, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含むの最小性の λ システムなので

$$\mathcal{G}_A \supset \mathcal{L}[\mathcal{P}]$$

が示せた． □

主張 3: $A \in \mathcal{P}$ かつ $B \in \mathcal{L}[\mathcal{P}]$ ならば $A \cap B \in \mathcal{L}[\mathcal{P}]$ となる．

主張 3 の証明: 主張 2 から

$$B \in \mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$$

となる．よって，$\mathcal{G}_A$ の定義から

$$A \cap B \in \mathcal{L}[\mathcal{P}]$$

がわかる． □

主張 4: $A \in \mathcal{L}[\mathcal{P}]$ ならば $\mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$ となる．

主張 4 の証明: $B \in \mathcal{P}$ と $A \in \mathcal{L}[\mathcal{P}]$ を取る．主張 3 から

$$A \cap B \in \mathcal{L}[\mathcal{P}] \tag{A.6}$$

となる．よって, $A \in \mathcal{L}[\mathcal{P}]$ で (A.6) であることに注意して, $\mathcal{G}_A$ の定義を思い出すと $B \in \mathcal{G}_A$ となる． 以上から

$$B \in \mathcal{P} \implies B \in \mathcal{G}_A$$

がわかる．すなわち, $\mathcal{P} \subset \mathcal{G}_A$ である．さらに, 主張 1 から $\mathcal{G}_A$ は λ システムである． 以上から, $\mathcal{G}_A$ は $\mathcal{P}$ を含む λ システムとであることがわかる． よって, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含む最小の λ システムなので

$$\mathcal{L}[\mathcal{P}] \subset \mathcal{G}_A$$

がわかる. □

主張 5: $\mathcal{L}[\mathcal{P}]$ は π システムである.

主張 5 の証明: A, $B \in \mathcal{L}[\mathcal{P}]$ とする. すると主張 4 から

$$B \in \mathcal{L}[\mathcal{P}] \implies B \in \mathcal{G}_A$$

が成り立つ. $\mathcal{G}_A$ の定義から

$$A \cap B \in \mathcal{L}[\mathcal{P}]$$

となり, $\mathcal{L}[\mathcal{P}]$ は π システムであることがわかる. □

定理 A.5. Ω を空でない集合とする. Ω の部分集合族 $\mathcal{P}$ を π システムとし, $\mathcal{L}$ を λ システムかつ $\mathcal{P} \subset \mathcal{L}$ とする. このとき

$$\sigma[\mathcal{P}] \subset \mathcal{L} \tag{A.7}$$

が成立する.

Proof. 命題 A.4 から $\mathcal{L}[\mathcal{P}]$ は π システムである. よって命題 A.3 より $\mathcal{L}[\mathcal{P}]$ は σ 加法族である. よって $\sigma[\mathcal{P}]$ の最小性から

$$\mathcal{L}[\mathcal{P}] \supset \sigma[\mathcal{P}]$$

となる. さらに, $\mathcal{L}$ は $\mathcal{P}$ を含む λ システムである. 一方, $\mathcal{L}[\mathcal{P}]$ は $\mathcal{P}$ を含む最初の λ システムある. この二つのことから

$$\mathcal{L} \supset \mathcal{L}[\mathcal{P}] \supset \mathcal{P} \tag{A.8}$$

となる. (A.7) と (A.8) を合わせると

$$\mathcal{L}{=}\sigma[\mathcal{P}]$$

がわかる. □

A.1.1 Dynkin の定理の応用例

その 1

命題 A.6. P_1 と P_2 を標本空間 $\big(\mathbb{X}, \mathcal{B}\big)$ 上の確率測度とする. 集合族

$$\mathcal{L} := \big\{B \in \mathcal{B};\ \mathsf{P}_1(B) = \mathsf{P}_2(B)\big\}$$

は λ システムである.

Proof. 命題 A.2 の条件 (i)′ ∼ (iii)′ を順に確認していく.

(i)′ $\mathsf{P}_1(\mathbb{X}) = \mathsf{P}_2(\mathbb{X})$ なので, $\mathbb{X} \in \mathcal{L}$ がわかる.

(ii)′ $A \in \mathcal{L}$ とする. すなわち

$$\mathsf{P}_1(A) = \mathsf{P}_2(A)$$

である. よって

$$\mathsf{P}_1(A^c) = 1 - \mathsf{P}_1(A) = 1 - \mathsf{P}_2(A) = \mathsf{P}_2(A^c)$$

であるので, $A^c \in \mathcal{L}$ がわかる.

(iii)′ $\{A_j\}_{j=1}^{\infty}$ は $\mathcal{L}$ の互いに排反な列とする. このとき $j = 1, 2, \ldots$ に対して

$$\mathsf{P}_1(A_j) = \mathsf{P}_2(A_j)$$

となる. よって

$$\mathsf{P}_1\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathsf{P}_1(A_j) = \sum_{j=1}^{\infty} \mathsf{P}_2(A_j) = \mathsf{P}_2\left(\bigcup_{j=1}^{\infty} A_j\right)$$

となるので

$$\bigcup_{j=1}^{\infty} A_j \in \mathcal{L}$$

がわかる. □

系 A.7. P_1 と P_2 は標本空間 $\big(\mathbb{X}, \mathcal{B}\big)$ 上の確率測度とする. $\mathcal{P}$ は π システムで

$$\forall A \in \mathcal{P} \implies \mathsf{P}_1(A) = \mathsf{P}_2(A)$$

ならば

$$\forall B \in \mathcal{B} \implies \mathsf{P}_1(B) = \mathsf{P}_2(B)$$

が成り立つ.

Proof. 命題 A.6 から

$$\mathcal{L} := \big\{A \in \mathcal{B};\, \mathsf{P}_1(A) = \mathsf{P}_2(A)\big\}$$

は λ システムである. しかし $\mathcal{L} \supset \mathcal{P}$ なので, 定理 A.5 から

$$\mathcal{L} \supset \sigma[\mathcal{P}]$$

となる. □

系 A.8. $\mathbb{X} = \mathbb{R}$ とし, P_1 と P_2 を $\bigl(\mathbb{R}, \mathcal{B}(\mathbb{R})\bigr)$ 上の確率測度とし, 対応する分布関数を F_1 と F_2 は等しいとする. すなわち

$$\mathsf{F}_1(x) = \mathsf{P}_1\bigl((-\infty, x]\bigr) = \mathsf{F}_2(x) = \mathsf{P}_2\bigl((-\infty, x]\bigr)$$

である. このとき

$$\mathsf{P}_1(B) = \mathsf{P}_2(B) \qquad \bigl(\forall B \in \mathcal{B}(\mathbb{R})\bigr)$$

が成り立つ.

Proof.

$$\mathcal{P} := \bigl\{(-\infty, x];\, x \in \mathbb{R}\bigr\}$$

とおく. $\forall x, y \in \mathbb{R}$ に対して

$$(-\infty, x] \bigcap (-\infty, y] = (-\infty, x \wedge y] \in \mathcal{P}$$

となる. ここで $x \wedge y$ は x と y の大きくない方である. 以上から $\mathcal{P}$ は π システムである.

さらに $\sigma[\mathcal{P}] = \mathcal{B}(\mathbb{R})$ であるので

$$\mathsf{F}_1(x) = \mathsf{F}_2(x)\,(\forall x \in \mathbb{R}) \implies \mathsf{P}_1(A) = \mathsf{P}_2(A)\,(\forall A \in \mathcal{P})$$

である. したがって, $\mathcal{L} = \{B \in \mathcal{B}(\mathbb{R});\, \mathsf{P}_1(B) = \mathsf{P}_2(B)\}$ とすると系 A.7 から

$$\mathcal{L} \supset \sigma[\mathcal{P}]$$

となる. 一方, $\mathcal{L}$ の定義から $\mathcal{L} \subset \mathcal{B}(\mathbb{R})$ である. さらに, 定理 A.11 から $\sigma[\mathcal{P}] = \mathcal{B}(\mathbb{R})$ である. よって, $\mathcal{L} = \sigma[\mathcal{P}] = \mathcal{B}(\mathbb{R})$ となることがわかる. 以上から

$$\mathsf{P}_1(A) = \mathsf{P}_2(A) \qquad (\forall A \in \mathcal{B}(\mathbb{R}))$$

がわかる. □

その 2

$(\mathbb{X}, \mathcal{B})$ と $(\mathbb{Y}, \mathcal{C})$ を可測空間とし, μ を可測空間 $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度とする. $\mathcal{B} \otimes \mathcal{C}$ 可測関数 $f : \mathbb{X} \times \mathbb{Y} \longrightarrow \mathbb{R}$ に対して, 写像

$$y \mapsto \int_{\mathbb{X}} f(x, y)\, \mathrm{d}\mu(x) \tag{A.9}$$

は $\mathcal{C}$ 可測であることを Dynkin の補題と標準機械を落ちいて示そう.

補題 A.9. $(\mathbb{X}, \mathcal{B})$ と $(\mathbb{Y}, \mathcal{C})$ を可測空間とし, μ を可測空間 $(\mathbb{X}, \mathcal{B})$ 上の σ 有限測度とする. 各 $L \in \mathcal{B} \otimes \mathcal{C}$ に対して, 写像

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_L(x, y)\, d\mu(x) \tag{A.10}$$

は $\mathcal{C}$ 可測である.

Proof. 各 $y \in \mathbb{Y}$ に対して, 写像

$$x \mapsto 1_L(x, y) \tag{A.11}$$

は $\mathcal{B}$ 可測となる. なぜならば, 写像

$$i_y : x \mapsto (x, y)$$

は $\mathcal{B} \setminus \mathcal{B} \otimes \mathcal{C}$ 可測となること [1] がわかる. よって, (A.10) は $\mathcal{B}$ 可測となる. このことから (A.9) の右辺の積分は well-defined である.

つぎに

$$\mathcal{L} := \left\{ L \in \mathcal{B} \otimes \mathcal{C};\ y \mapsto \textstyle\int_{\mathbb{X}} \mathbb{1}_L(x, y)\, d\mu(x) \text{ は } \mathcal{C} \text{ 可測} \right\}$$

とおいたとき, $\mathcal{L} = \mathcal{B} \otimes \mathcal{C}$ を示す.

そのために, $B \in \mathcal{B}$ と $C \in \mathcal{C}$ を取る. すると

$$\begin{aligned}\int_{\mathbb{X}} \mathbb{1}_{B\times C}(x, y)\, \mathrm{d}\mu(x) &= \int_{\mathbb{X}} \mathbb{1}_B(x)\mathbb{1}_C(y)\, \mathrm{d}\mu(x) = \mathbb{1}_C(y) \int_{\mathbb{X}} \mathbb{1}_B(x)\, \mathrm{d}\mu(x) \\ &= \mathbb{1}_C(y)\mu(B)\end{aligned}$$

となる. このことから, 写像

$$y \mapsto \int_{\mathbb{X}} \mathbb{1}_{B\times C}(x, y)\, \mathrm{d}\mu(x) = \mathbb{1}_C(y)\mu(B)$$

は, 明らかに $\mathcal{C}$ 可測となる. よって

$$B \times C \in \mathcal{L} \tag{A.12}$$

である. ここで集合族 $\mathcal{P}$ を

$$\mathcal{P} := \left\{ B \times C;\, B \in \mathcal{B},\, C \in \mathcal{C} \right\}$$

で定めると (A.12) から $\mathcal{P} \subset \mathcal{L}$ であり, さらに $\mathcal{P}$ は π システムである. したがって, $\mathcal{L}$ が λ システムであることがわかると定理 A.5 から

$$\sigma[\mathcal{P}] \subset \mathcal{L}$$

[1] Hansen (2009, Lemma 8.4) を参照のこと.

がわかる．さらに，$\mathcal{B}\otimes\mathcal{C}$ の定義を思い出すと

$$\sigma[\mathcal{P}]=\mathcal{B}\otimes\mathcal{C}$$

であるので

$$\mathcal{L}=\mathcal{B}\otimes\mathcal{C}$$

がわかる．よって，(A.9) で定義された写像は $\mathcal{C}$ 可測であることがわかる．
□

補題 A.10.

$$\mathcal{L}:=\left\{L\in\mathcal{B}\otimes\mathcal{C};\ \ y\mapsto \textstyle\int_{\mathbb{X}}\mathbb{1}_L(x,\,y)\,\mathrm{d}\mu(x)\ \text{は}\ \mathcal{C}\ \text{可測}\ \right\}$$

は λ システムである．

Proof. 定義 A.1 の (i) $\sim$ (iii) を確認する．
(i): $\mathbb{X}\times\mathbb{Y}\in\mathcal{L}$ であることは明らか．
(ii): $L_1,\,L_2\in\mathcal{L}$ かつ $L_1\subset L_2$ に対して

$$\mathbb{1}_{L_1}(x,\,y)+\mathbb{1}_{L_2\setminus L_1}(x,\,y)=\mathbb{1}_{L_2}(x,\,y)\qquad\big((x,\,y)\in\mathbb{X}\times\mathbb{Y}\big)$$

である．$L_1,\,L_2\in\mathcal{L}$ なので，写像

$$y\mapsto\int_{\mathbb{X}}\mathbb{1}_{L_1}(x,\,y)\,\mathrm{d}\mu(x)\quad \text{と}\quad y\mapsto\int_{\mathbb{X}}\mathbb{1}_{L_2}(x,\,y)\,\mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測である．写像

$$y\mapsto\int_{\mathbb{X}}\mathbb{1}_{L_2\setminus L_1}(x,\,y)\,\mathrm{d}\mu(x)$$

は $\mathcal{C}$ 可測関数の差なので，$\mathcal{C}$ 可測となる．よって，$L_2\setminus L_1\in\mathcal{L}$ となる．
$L_n\in\mathcal{L}\,(n=1,\,2,\,\ldots)$ かつ $L_1\subset L_2\subset\cdots$ とする．$L=\bigcup_{n=1}^{\infty}L_n$ とおいたとき

$$\mathbb{1}_{L_n}(x,\,y)\nearrow\mathbb{1}_L(x,\,y)\qquad\big((x,\,y)\in\mathbb{X}\times\mathbb{Y}\big)$$

となる．よって，単調収束定理から

$$\lim_{n\to\infty}\int_{\mathbb{X}}\mathbb{1}_{L_n}(x,\,y)\,\mathrm{d}\mu(x)=\int_{\mathbb{X}}\mathbb{1}_L(x,\,y)\,\mathrm{d}\mu(x)$$

となる．よって，写像

$$y\mapsto\int_{\mathbb{X}}\mathbb{1}_L(x,\,y)\,\mathrm{d}\mu(x)$$

も $\mathcal{C}$ 可測となる. よって, $L \in \mathcal{L}$ がわかる.

(i) ∼ (iii) から $\mathcal{L}$ は λ システムであることが示せた. □

(A.9) の可測性の証明: あとは標準機械を用いる. すなわち, 以下の手続きをふくことで証明をする.

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 一般の可測関数 f を非負の部分 f^+ と負の部分 f^- にわけたものに ③ を適用し, $f = f^+ - f^-$ 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.

すなわち, 補題 A.10 から出発して, ① ∼ ④ の議論を実行することにより, (A.9) によって定義される写像は $\mathcal{C}$ であることがわかる. □

$\mathbb{R}$ の位相についての注意

定理 A.11. $U \subset \mathbb{R}$ が開集合ならば, 互いに排反な開区間 $\{I_j\}_{j=1}^{\infty}$ があって

$$U = \bigcup_{j=1}^{\infty} I_j$$

と書ける.

Proof. U は開集合なので, 各 $x \in U$ に対して $y > x$ なる $y \in \mathbb{R}$ が存在して,

$$(x, y) \subset U$$

とできる. このとき

$$a := \inf\{z; (z, x) \subset U\}, \quad b := \sup\{y; (x, y) \subset U\}$$

と定める. a と b は x に依存することに注意する. このとき

$$x \in (a, b) \quad \text{かつ} \quad I_x = (a, b) \text{ は開区間}$$

である. これらを踏まえて, 次の主張を示していく.

(i) $U = \bigcup\{I_x;\, x \in U\}$.

(ii) $\forall x,\, y \in U$ に対して

$$x \neq y \implies I_x \cap I_y = \varnothing.$$

(iii) $\#\{I_x;\, x \in U\} \leq \aleph_0$. ただし $I_x = I_y$ なるものを除いてである.

まず, 次のことを注意する. $x \in U$ に対して, $I_x = (a,\, b)$ とすると $I_x \subset U$ かつ $a,\, b \notin U$ である. なぜならば $w \in (a,\, b)$ とすると $y > w$ なる $y \in \mathbb{R}$ が存在して, $(x,\, y) \subset U$ となる. よって $w \in U$ である. いま $b \in U$ と仮定するとある $\epsilon > 0$ が存在して

$$(b - \epsilon,\, b + \epsilon) \subset U$$

とできる. ϵ を十分小さくすると

$$\left(x,\, b - \frac{\epsilon}{2}\right) \subset U$$

となる. 結局

$$(x,\, b + \epsilon) \subset U$$

となり, b の定義と矛盾する. よって $b \notin U$ が示せた.

(i) 各 $x \in U$ に対して, $x \in I_x \subset U$ なので

$$\bigcup_{x \in U} I_x \subset U$$

となる. 逆は明らかなので

$$\bigcup_{x \in U} I_x = U$$

がわかる.

(ii) $x,\, y \in U$ とし, $I_x = (a,\, b)$ と $I_y = (c,\, d)$ と定める. $I_x \cap I_y \neq \varnothing$ と仮定する. このとき $a < d$ かつ $c < b$ となる. $c \notin U$ なので $c \in (a,\, b)$ である. $c < b$ なので

$$c \leq a \tag{A.13}$$

である. $a \notin U$ なので

$$a \notin (c,\, d)$$

である．よって $a < d$ なので

$$a \leq c \tag{A.14}$$

がわかる．(A.13) と (A.14) から

$$a = c$$

がわかる．同様に

$$b = d \tag{A.15}$$

を示すことができる．よって

$$I_x \cap I_y \neq \varnothing \implies I_x = I_y$$

がわかる．これの対偶を取ると

$$I_x \neq I_y \implies I_x \cap I_y = \varnothing$$

がわかる．

(iii) 各 $I_x\,(x \in U)$ に対して

$$q_x \in I_x \cap \mathbb{Q}$$

を取る．$I_x \neq I_y$ ならば $I_x \cap I_y = \varnothing$ なので

$$q_x \neq q_y$$

がわかる．よって $\#(\mathbb{Q}) \leq \aleph_0$ なので, (iii) がわかる． □

A.1.2　文献についての注釈

節 3.2 は [?] の第 2 章を借用した．節 A.1.1 は [?] の第 2 章を借用した．節 A.1.1 は [?](pp.173-179) を借用した．節 A.1.1 は [?](p.4) を借用した．

A.2　Radon-Nikodym の定理と条件付き期待値

定義 A.12. $(\mathbb{X}, \mathcal{B})$ 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 A.13 (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

$$\nu(B) = \int_B \mathsf{p}(x)\, \mathrm{d}\mu(x) \tag{A.16}$$

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

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

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 $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$ と書く. ν が σ 有限のとき μ a.e. で f は有限となる.

$(\Omega, \mathcal{A}, \mathsf{Pr})$ 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 A.14. $\mathcal{F}$ 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 A.15. Y を定義 $A.14$ 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

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

よって

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

からわかる. □

補題 A.16. Y は一意.

Proof. Y' も定義 A.14 の条件 (i)(ii) をみたすとする. すると (ii) より

$$\int_A Y(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \int_A Y'(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad (\forall A \in \mathcal{F}).$$

$\epsilon > 0$ を取り, $A := \{\omega \in \Omega;\, Y(\omega) - Y'(\omega) \geq \epsilon\}$ とする. $A \in \mathcal{F}$ であることから (ii) から

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

となり, $\mathsf{Pr}(A) = 0$ がわかる.

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

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

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

がわかる. よって

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

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

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

よって

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

となる. □

定義 A.17. 定義 $A.14$ 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 A.18. $\mathsf{E}\big[X|\,\mathcal{F}\big](\omega)$ 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

$$\begin{aligned}\int_A \mathsf{E}\big[X|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mu(\omega) &= \int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) = \nu(A)\\ &= \int_A \frac{d\nu}{d\mu}(\omega)\,\mathrm{d}\mu(\omega) \qquad (A \in \mathcal{F}) \end{aligned} \tag{A.17}$$

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

$$\begin{aligned}\int_A X(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \int_A X^+(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A X^-(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A Y_1(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) - \int_A Y_2(\omega)\,\mathrm{d}\mathsf{Pr}(\omega) \qquad \big(\because (\mathrm{A}.17)\big)\\ &= \int_A \Big\{Y_1(\omega) - Y_2(\omega)\Big\}\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \int_A \big(Y_1 - Y_2\big)(\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\end{aligned}$$

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 A.19. 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

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

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 **??**)) の定理から

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

となる．よって

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

なので

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

がわかる． □

A.3 補遺: 条件付き期待値の性質

定理 A.20. $X,\, \{X_n\}_{n=1}^{\infty},\, Y$ を確率空間 $(\Omega,\, \mathcal{A},\, \mathsf{Pr})$ 上の確率変数列とする．
(1) $\mathsf{E}\bigl[|X|\bigr] < \infty$ かつ $\mathsf{E}\bigl[|Y|\bigr] < \infty$ とする．このとき $a,\, b \in \mathbb{R}$ に対して

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

と成り立つ．
(2) さらに $X \le Y$ のとき

$$
\mathsf{E}\bigl[X|\,\mathcal{F}\bigr](\omega) \le \mathsf{E}\bigl[Y|\,\mathcal{F}\bigr](\omega)
$$

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

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

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

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

を得る. $\forall\epsilon>0$ に対して

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

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

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

となる. よって

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

となるから

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

となる.

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

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

はほとんど確実に単調減少列なので

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

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 A.21. $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}$ 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 A.22. X は $\mathcal{F}$ 可測とし, $\mathsf{E}\big[|X|\big] < \infty$ かつ $\mathsf{E}\big[|XY|\big] < \infty$ とする. このとき

$$\mathsf{E}\big[XY \big| \mathcal{F}\big](\omega) = X(\omega)\mathsf{E}\big[Y \big| \mathcal{F}\big](\omega) \tag{A.18}$$

が成立する.

Proof. (i). $X(\omega) = \mathbb{1}_B(\omega)\,(\forall B \in \mathcal{F})$ のとき (A.18) 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 (A.18) 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

$$\begin{aligned}\int_A X(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &= \lim_{n\to\infty}\int_A X_n(\omega)\mathsf{E}\big[Y|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega)\\ &= \lim_{n\to\infty}\int_A \mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &&\big(\because \text{(i) の結果}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[X_nY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &&\big(\because \text{単調収束定理}\big)\\ &= \int_A \lim_{n\to\infty}\mathsf{E}\big[XY|\,\mathcal{F}\big](\omega)\,\mathrm{d}\mathsf{Pr}(\omega) &&\big(\because \text{定理 A.20(3)}\big)\end{aligned}$$

がわかり, $X,\,Y \geq 0$ のとき $(A.18)$ 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 A.23. 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

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

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 A.24. $X,\,Y$ 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.

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

しかし

$$
\begin{aligned}
\mathsf{E}\bigl[\{X-\mathsf{E}[X|\,Y]\}^2\bigr] &= \mathsf{E}\bigl[X^2-2X\mathsf{E}[X|\,Y]+\{\mathsf{E}[X|\,Y]\}\bigr] \\
&= \mathsf{E}\Bigl[\mathsf{E}\bigl[X^2-2X\mathsf{E}[X|\,Y]+\{\mathsf{E}[X|\,Y]\}^2\bigr]\Big|\,Y\Bigr] \\
&\quad (\text{定理 A.21(1)}) \\
&= \mathsf{E}\Bigl[\mathsf{E}\bigl[X^2\big|\,Y\bigr]-2\mathsf{E}[X|\,Y]\,\mathsf{E}[X|\,Y]+\{\mathsf{E}[X|\,Y]\}^2\Bigr] \\
&= \mathsf{E}\Bigl[\mathsf{E}[X^2|\,Y]-\{\mathsf{E}[X|\,Y]\}^2\Bigr] \\
&= \mathsf{E}\Bigl[\mathsf{Var}[X|\,Y]\Bigr], \\
\mathsf{E}\Bigl[\{\mathsf{E}[X|\,Y]-\mathsf{E}[X]\}^2\Bigr] &= \mathsf{E}\Bigl[\bigl\{\mathsf{E}[X|\,Y]-\underbrace{\mathsf{E}[\mathsf{E}[X|\,Y]]}_{=\mathsf{E}[X]}\bigr\}^2\Bigr] \qquad (\text{定理 A.21(1)}) \\
&= \mathsf{Var}\Bigl[\mathsf{E}[X|\,Y]\Bigr], \\
\mathsf{E}\bigl[\{X-\mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y]-\mathsf{E}[X]\}\bigr] &= \mathsf{E}\Bigl[\mathsf{E}\bigl[\{X-\mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y]-\mathsf{E}[X]\}\Big|\,Y\Bigr] \\
&\quad (\text{定理 A.21(1)}) \\
&= \mathsf{E}\Bigl[\{\mathsf{E}[X|\,Y]-\mathsf{E}[X]\}\,\underbrace{\mathsf{E}\bigl[X-\mathsf{E}[X|\,Y]\big|Y\bigr]}_{=0}\Bigr] \\
&= 0.
\end{aligned}
$$

最後から 2 番目の等号は定理 A.21(1) を用いた. □

A.4 Shannon エントロピーと Kullback-Leibler 偏差

$\mathbb{X}$ を可算集合とする. X を $\mathbb{X}$ に値を取る確率変数とし, その分布を

$$
\mathsf{Pr}(X=x)=\mathsf{p}(x) \qquad (x\in\mathbb{X})
$$

とする.

定義 A.25. 確率変数 X の Shannon **エントロピー** $\mathsf{H}(X)$ を

$$\mathsf{H}(X) := \mathsf{E}\big[-\log \mathsf{p}(X)\big] = -\sum_{x\in\mathbb{X};\, \mathsf{p}(x)>0} \mathsf{p}(x)\log\mathsf{p}(x)$$

で定める. ただし, $0\log 0 = 0$ と約束する.

注意 A.26. Shannon のエントロピーの記号は伝統的なものを用いた. しかし, 以下の注意が必要である. $\mathsf{H}(X)$ は確率変数 X の関数ではなく, X の分布の関数である. □

定義 A.27. P と Q を可算集合 $\mathbb{X}$ 上の確率分布とし

$$\mathsf{P}(\{x\}) = \mathsf{p}(x), \quad \mathsf{Q}(\{x\}) = \mathsf{q}(x) \qquad (x\in\mathbb{X})$$

とする. このとき, 分布 P と Q の Kullback-Leibler **偏差** $\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q})$ を

$$\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) = \begin{cases} \displaystyle\sum_{x\in\mathbb{X}} \mathsf{p}(x)\log\frac{\mathsf{p}(x)}{\mathsf{q}(x)} & (\mathsf{P}\ll\mathsf{Q}) \\ \infty & (\text{その他の場合}) \end{cases}$$

で定める. ただし, $\mathsf{P}\ll\mathsf{Q}$ は P は Q に関して絶対連続, すなかち, $\forall x\in\mathbb{X}$ に対して

$$\mathsf{q}(x) = 0 \;\Rightarrow\; \mathsf{p}(x) = 0$$

を意味する.

Kullback-Leibler 偏差の重要な性質は

(1) $\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) \geq 0$,

(2) $\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) = 0 \;\Leftrightarrow\; \mathsf{P} = \mathsf{Q}$

である.

<u>(1) の証明</u>: 実際

$$\log x \leq x - 1 \qquad (x>0)$$

に注意すると

$$\begin{aligned}\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) &= -\sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\log\frac{\mathsf{q}(x)}{\mathsf{p}(x)} \\ &\geq -\sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\left(\frac{\mathsf{q}(x)}{\mathsf{p}(x)} - 1\right) \\ &= -\underbrace{\sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{q}(x)}_{\leq 1} + \underbrace{\sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)}_{=1} \\ &\geq 0\end{aligned}$$

からわかる.
(2) の証明:

$$\begin{aligned}\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) \;&\Leftrightarrow \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\big\{\log\mathsf{p}(x) - \log\mathsf{q}(x)\big\}\\ &\Leftrightarrow\; \mathsf{p}(x) = \mathsf{q}(x) \qquad (\forall x\,\mathrm{s.t.}\,\mathsf{p}(x)>0)\end{aligned}$$

となる. しかし

$$1 = \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x) = \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{q}(x)$$

なので, $\forall x \in \mathbb{X}$ に対して

$$\mathsf{p}(x) \;\Rightarrow\; \mathsf{q}(x)$$

がわかる. 一方, $\mathsf{P} \ll \mathsf{Q}$ から, $\forall x \in \mathbb{X}$ に対して

$$\mathsf{q}(x) \;\Rightarrow\; \mathsf{p}(x)$$

である. 以上から $\mathsf{P} = \mathsf{Q}$ がわかる. □

例 A.28. $\mathbb{X}$ を有限集合とする. Q を $\mathbb{X}$ 上の一様分布とする. すなわち

$$\mathsf{Q}(\{x\}) = \frac{1}{\#(\mathbb{X})} \qquad (x \in \mathbb{X})$$

である. ただし, $\#(\mathbb{X})$ は有限集合 $\mathbb{X}$ の要素の個数である. X を確率変数とし, $X \sim \mathsf{P}$ としたとき

$$\begin{aligned}\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) &= \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\log\frac{\mathsf{p}(x)}{\mathsf{q}(x)}\\ &= \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\log\mathsf{p}(x) - \sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)\log\mathsf{q}(x)\\ &= -\mathsf{H}(X) + \log\#(\mathbb{X})\underbrace{\sum_{x\in\mathbb{X};\,\mathsf{p}(x)>0} \mathsf{p}(x)}_{=1}\\ &= \log\#(\mathbb{X}) - \mathsf{H}(X)\end{aligned}$$

がわかる. よって, $\mathsf{KL}(\mathsf{P}\,\|\,\mathsf{Q}) \geq 0$ から

$$\mathsf{H}(X) \leq \log\#(\mathbb{X})$$

を得る. さらに, 上の不等式の等号成立条件は P が $\mathbb{X}$ 上の一様分布のときに限ることもわかる. □

A.5 直積空間のエントロピーと連鎖律

$\mathbb{X}$, $\mathbb{Y}$ を可算集合とする．直積集合 $\mathbb{X}\times\mathbb{Y}$ 上の Shannon のエントロピーを考える．確率変数の組 (X, Y) は直積空間 $\mathbb{X}\times\mathbb{Y}$ に値を取るとき，(X, Y) の同時エントロピー $\mathsf{H}(X, Y)$ を確率変数の組 (X, Y) のエントロピーで定義する．

いま，(X, Y) の p.m.f. を $\mathsf{p}^{(X,Y)}$ と書くことする．すなわち

$$\mathsf{Pr}(X = x, Y = y) = \mathsf{p}^{(X,Y)}(x, y) \qquad ((x, y) \in \mathbb{X}\times\mathbb{Y})$$

である．さらに，X と Y の周辺 p.m.f. をそれぞれ p^X と p^Y と記すことにする．このとき

$$\mathsf{H}(X) + \mathsf{H}(Y) - \mathsf{H}(X, Y) = \sum_{x,y} \mathsf{p}^{(X,Y)}(x, y) \log \frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^X(x)\,\mathsf{p}^Y(y)} \tag{A.19}$$

と書ける[2]．上の式の左辺は $\mathsf{p}^{(X,Y)}$ と $\mathsf{p}^X(x)\mathsf{p}^Y(y)$ の Kullback-Leibler 偏差であることに注意すると Shannon のエントロピーの劣加法性がわかる．すなわち

$$\mathsf{H}(X, Y) \le \mathsf{H}(X) + \mathsf{H}(Y)$$

が成立する．

注意 A.29. $\mathsf{H}(X) + \mathsf{H}(Y) - \mathsf{H}(X, Y)$ を確率変数 X と Y の相互情報量という． □

定義 A.30. 条件付きエントロピー $\mathsf{H}(X|Y)$ を

$$\mathsf{H}(X|Y) = \mathsf{H}(X, Y) - \mathsf{H}(Y) \tag{A.20}$$

で定義する．

---

[2]

$$\begin{aligned}
&\sum_{x,y} \mathsf{p}^{(X,Y)}(x, y) \log \frac{\mathsf{p}^{(X,Y)}(x, y)}{\mathsf{p}^X(x)\,\mathsf{p}^Y(y)} \\
&\quad= \sum_{x,y} \mathsf{p}^{(X,Y)}(x, y) \log \mathsf{p}^{(X,Y)}(x, y) - \sum_{x,y} \mathsf{p}^{(X,Y)}(x, y) \log \mathsf{p}^X(x) \\
&\qquad - \sum_{x,y} \mathsf{p}^{(X,Y)}(x, y) \log \mathsf{q}^X(x) \\
&\quad= -\mathsf{H}(X, Y) - \sum_x \log \mathsf{p}^X(x) \underbrace{\sum_y \mathsf{p}^{(X,Y)}(x, y)}_{=\mathsf{p}^X(x)} - \sum_y \log \mathsf{p}^Y(y) \underbrace{\sum_x \mathsf{p}^{(X,Y)}(x, y)}_{=\mathsf{p}^Y(y)} \\
&\quad= -\mathsf{H}(X, Y) + \mathsf{H}(X) + \mathsf{H}(Y)
\end{aligned}$$

からわかる．

注意 A.31. $Y = y$ を与えたときの X の条件付き p.m.f. を $\mathsf{p}^{X|Y}(x|\,y)$ と書いたとき

$$\begin{aligned}\mathsf{H}(X|\,Y) &= -\sum_{(x,\,y)\in\mathbb{X}\times\mathbb{Y}} \mathsf{p}^{(X,\,Y)}(x,\,y)\log\mathsf{p}^{(X,\,Y)}(x,\,y) + \sum_{y\in\mathbb{Y}}\mathsf{p}^{Y}(y)\log\mathsf{p}^{Y}(y)\\ &= \sum_{y\in\mathbb{Y}}\mathsf{p}^{Y}(y)\Bigg(-\sum_{x\in\mathbb{X}}\underbrace{\frac{\mathsf{p}^{(X,\,Y)}(x,\,y)}{\mathsf{p}^{Y}(y)}}_{=\mathsf{p}^{X|\,Y}(x|\,y)}\log\mathsf{p}^{(X,\,Y)}(x,\,y) + \underbrace{\log\mathsf{p}^{Y}(y)}_{=\log\mathsf{p}^{Y}(y)\sum_{x\in\mathbb{X}}\mathsf{p}^{X|\,Y}(x|\,y)}\Bigg)\\ &= \sum_{y\in\mathbb{Y}}\mathsf{p}^{Y}(y)\Bigg(-\sum_{x\in\mathbb{X}}\mathsf{p}^{X|Y}(x|\,y)\log\mathsf{p}^{X|Y}(x|\,y)\Bigg)\\ &= \mathsf{E}\big[-\log\mathsf{p}^{X|Y}(X|\,Y)\big]\end{aligned}$$

を得る. したがって, 条件付きエントロピーは条件付き分布の Shannon のエントロピーの期待値になるので, $\mathsf{H}(X|\,Y)\geq 0$ となることがわかる. □

注意 A.32. $(X,\,Y)$ の同時分布 $\mathsf{P}^{(X,\,Y)}$ と記し, X と Y の周辺分布をそれぞれ P^{X} と P^{Y} と記す. $\mathsf{P}^{(X,\,Y)}$ と $\mathsf{P}^{X}\otimes\mathsf{P}^{Y}$ の Kullback-Leibler 偏差は, $(A.19)$ から

$$\begin{aligned}\mathsf{KL}(\mathsf{P}^{(X,\,Y)}\,\|\mathsf{P}^{X}\otimes\mathsf{P}^{Y}) &= \mathsf{H}(X) + \underbrace{\mathsf{H}(Y) - \mathsf{H}(X,\,Y)}_{=-\mathsf{H}(X|\,Y)\quad \because(A.20)}\\ &= \mathsf{H}(X) - \mathsf{H}(X|\,Y)\end{aligned}$$

がわかる. したがって

$$\mathsf{H}(X)\geq\mathsf{H}(X|\,Y)$$

が成立することがわかる. □

注意 A.33. (1) $X,\,Y,\,Z$ を (離散型) 確率変数としたとき

$$\mathsf{H}(X,\,Y|\,Z) = \mathsf{H}(Y|\,Z) + \mathsf{H}(X|\,Y,\,Z)$$

が成立する. 実際, $(A.20)$ から

$$\begin{aligned}\mathsf{H}(X,\,Y|\,Z) &= \mathsf{H}(X,\,Y,\,Z) - \mathsf{H}(Z)\\ &= \mathsf{H}(X,\,Y,\,Z) - \mathsf{H}(Y,\,Z) + \mathsf{H}(Y,\,Z) - \mathsf{H}(Z)\\ &= \mathsf{H}(X|\,Y,\,Z) + \mathsf{H}(Y|\,Z)\end{aligned}$$

となることがわかる.

(2) $X_1, X_2, \ldots, X_n$ を (離散型) 確率変数列としたとき

$$
\begin{aligned}
&\mathsf{H}(X_1, X_2, \ldots, X_n)\\
&\quad = \underbrace{\mathsf{H}(X_1, X_2, \ldots, X_n) - \mathsf{H}(X_1, X_2, \ldots, X_{n-1})}_{=H(X_n|\, X_1, X_2, \ldots, X_{n-1})} + \mathsf{H}(X_1, X_2, \ldots, X_{n-1})\\
&\quad = H(X_n|\, X_1, X_2, \ldots, X_{n-1})\\
&\qquad + \underbrace{\mathsf{H}(X_1, X_2, \ldots, X_{n-1}) - \mathsf{H}(X_1, X_2, \ldots, X_{n-2})}_{=H(X_{n-1}|\, X_1, X_2, \ldots, X_{n-2})} + \mathsf{H}(X_1, X_2, \ldots, X_{n-2})\\
&\quad = \cdots\\
&\quad = H(X_n|\, X_1, X_2, \ldots, X_{n-1}) + H(X_{n-1}|\, X_1, X_2, \ldots, X_{n-2}) + \cdots\\
&\qquad + \mathsf{H}(X_2|\, X_1) + \mathsf{H}(X_1)
\end{aligned}
$$

がわかる. □

A.6 凸関数とその性質

この節では, 凸関数の定義をまず述べる. 凸関数のこの定義は幾何的なものであり, 凸関数がどのようなものであるかを直観的に理解するが用意である. しかし, 解析的な議論をするためには, この定義では不十分[3]である. 凸関数であるための必要十分条件が重要になってくる. まず, ① 閉区間上の凸関数はその内部で絶対連続であることを示す. さらに, 凸関数であるための必要十分条件を 2 つ述べる. ひとつは, ② 凸関数の平均変動は単調増加であることであり, もうひとつは, ③下から支える直線の存在することである. これらは, 凸関数を特徴付けていることに注意する.

以下, m を $\mathbb{R}$ 上の Lebeague 測度とする.

定義 A.34. $a, b \in \mathbb{R}, a < b$ とする. 関数 $f : [a, b] \to \mathbb{R}$ が $x, y \in [a, b], 0 \leq \lambda \leq 1$ に対して

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y)$$

をみたすとき, 関数 f は**凸関数**であるという. とくに $x \neq y, 0 < \lambda < 1$ のとき

$$f\big(\lambda x + (1-\lambda)y\big) < \lambda f(x) + (1-\lambda) f(y)$$

が成立する場合は, 関数 f は**狭義凸関数**であるという.

注意 A.35. $-f$ が凸関数 (狭義凸関数) のとき, 関数 f は**凹関数** (**狭義凹関数**) であるという. □

[3]たとえば, この定義の直接的に利用では Jensen の不等式の証明できない.

定理 A.36. 区間 $[a,\,b]$ 上の凸関数は開区間 $(a,\,b)$ 上の各点で連続である.

Proof. まず, $\forall x \in [a,\,b]$ は

$$x = \lambda a + (1-\lambda)b \quad (0 \le \lambda \le 1)$$

と表せるから, 凸関数の定義より

$$f(x) \le \lambda f(a) + (1-\lambda) f(b) \le \max\{f(a),\, f(b)\} =: K$$

となることに注意する. よって, 凸関数 f は上に有界である. $\forall x_0 \in (a,\,b)$ とし, $f(x)$ は $x = x_0$ で連続であることを示すために関数 g を

$$g(t) := f(x_0 + t) - f(x_0) \qquad (a - x_0 < t < b - x_0)$$

と定める. このとき $g(t)$ は $(a - x_0,\, b - x_0) \ni \{0\}$ 上で凸関数であり, $g(0) = 0$ である. いま $\forall \epsilon > 0$ に対して正数 ϵ' を

$$0 < \epsilon' < \min\left\{\frac{\epsilon}{K + |f(x_0)|},\, 1\right\} \tag{A.21}$$

をみたすように選ぶ. さらに正数 ρ と δ を

$$0 < \rho < \min\{x_0 - a,\, b - x_0\}, \quad 0 < \delta < \rho\epsilon' \tag{A.22}$$

をみたすように定める. $0 < \epsilon' \le 1$ だから

$$|t| < \delta \;\Rightarrow\; |t| < \rho$$

となる. したがって $|t| < \delta$ に対して, $g(t)$ は定義されている. しかも

$$|t| < \delta \;\Rightarrow\; \left|\frac{t}{\epsilon'}\right| < \rho \;\Rightarrow\; x_0 + \frac{t}{\epsilon'} \in (a,\, b)$$

となること [4] がわかる. よって, (A.23) から

$$\begin{aligned} g\left(\frac{t}{\epsilon'}\right) &= f\left(x_0 + \frac{t}{\epsilon'}\right) - f(x_0) \le \left|f\left(x_0 + \frac{t}{\epsilon'}\right)\right| + |f(x_0)| \\ &\le K + |f(x_0)| =: K_1 \end{aligned}$$

を得る. ところで

$$t = (1-\epsilon') \times 0 + \epsilon'\left(\frac{t}{\epsilon'}\right)$$

[4] $\left|\dfrac{t}{\epsilon'}\right| < \rho$ と (A.22) から $-x_0 + a < -\rho < \dfrac{t}{\epsilon'} < \rho < b - x_0$ となるので $a < \dfrac{t}{\epsilon'} + x_0 < b$ がわかる.

と表されることと (A.21) から

$$g(t) \leq (1-\epsilon')g(0) + \epsilon' g\left(\frac{t}{\epsilon'}\right) = \epsilon' g\left(\frac{t}{\epsilon'}\right) \leq \epsilon' K_1 < \epsilon \tag{A.23}$$

となることに注意をする. 他方

$$0 = \frac{\epsilon'}{1+\epsilon'}\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1+\epsilon'} \cdot t$$

と表されるので

$$\begin{aligned} 0 \leq g(0) &\leq \frac{\epsilon'}{1+\epsilon'} g\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1+\epsilon'} g(t) \leq \frac{\epsilon' K_1}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t) \\ &< \frac{\epsilon}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t) \end{aligned}$$

となり

$$-\epsilon < g(t) \tag{A.24}$$

を得る. よって (A.23) と (A.24) より

$$|t| < \delta \Rightarrow |g(t)| < \epsilon$$

である. $t = x - x_0$ と $g(t)$ の定義より

$$|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon$$

となり, $f(x)$ は点 x_0 で連続となる. □

定理 A.37. 関数 $\varphi\colon [a, b] \to \mathbb{R}$ が単調増加ならば

$$f(x) := C + \int_a^x \varphi(t)\,\mathsf{dm}(t) \qquad (a \leq x \leq b)$$

は凸関数である. ただし C は定数である.

Proof. まず, $x, y \in [a, b]$, $x < y$, $0 \leq \lambda \leq 1$ のとき

$$x \leq \lambda x + (1-\lambda)y \leq y$$

であることに注意する. ここで, 関数 φ の単調性に注意すれば

$$\int_{\lambda x+(1-\lambda)y}^{y} \varphi(t)\,\mathsf{dm}(t) \geq \{y - \lambda x - (1-\lambda)y\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr), \tag{A.25}$$

$$\int_{x}^{\lambda x+(1-\lambda)y} \varphi(t)\,\mathsf{dm}(t) \leq \{\lambda x + (1-\lambda)y - x\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \tag{A.26}$$

がわかる. f の凸性を示すために, 以下の式が非負であることがわかればよい.

$$
\begin{aligned}
&\lambda f(x)+(1-\lambda)f(y)-f\bigl(\lambda x+(1-\lambda)y\bigr)\\
&\quad=(1-\lambda)\bigl\{f(y)-f\bigl(\lambda x+(1-\lambda)y\bigr)\bigr\}-\lambda\bigl\{f\bigl(\lambda x+(1-\lambda)y\bigr)-f(x)\bigr\}\\
&\quad=(1-\lambda)\int_{\lambda x+(1-\lambda)y}^{y}\varphi(t)\,\mathsf{dm}(t)-\lambda\int_{x}^{\lambda x+(1-\lambda)y}\varphi(t)\,\mathsf{dm}(t)\\
&\quad\geq(1-\lambda)\{y-\lambda x-(1-\lambda)y\}\varphi\bigl(\lambda x+(1-\lambda)y\bigr)\qquad(\because\ \text{(A.25)})\\
&\qquad-\lambda\{\lambda x+(1-\lambda)y-x\}\varphi\bigl(\lambda x+(1-\lambda)y\bigr)\qquad(\because\ \text{(A.26)})\\
&\quad=0
\end{aligned}
$$

となる. 以上の議論から

$$\lambda f(x)+(1-\lambda)f(y)\geq f\bigl(\lambda x+(1-\lambda)y\bigr)$$

が示せた. □

定理 A.38. 関数　f が区間 $[a,\,b]$ 上の凸のとき, 開区間 $(a,\,b)$ の各点 x で右微分係数 $\overset{\bullet}{f}_+(x)$ が存在し, 関数

$$\overset{\bullet}{f}_+\colon\ (a,\,b)\ni x\mapsto\overset{\bullet}{f}_+(x)\in\mathbb{R}$$

は単調増加である.

同様に左微分係数 $\overset{\bullet}{f}_-(x)$ が存在し, 関数

$$\overset{\bullet}{f}_-\colon\ (a,\,b)\ni\mapsto\overset{\bullet}{f}_-(x)\in\mathbb{R}$$

も単調増加である. さらに各点 $x\in(a,\,b)$ において

$$\overset{\bullet}{f}_-(x)\leq\overset{\bullet}{f}_+(x)$$

が成立する.

Proof. 任意の $x\in(a,\,b)$ に対して実数 η と ξ を

$$0<\eta<\min\{x-a,\,b-x\},\quad 0<\xi<\eta$$

をみたすように選ぶ. このとき

$$0<\frac{\xi}{\eta}<1,\quad x+\xi=\frac{\xi}{\eta}(x+\eta)+\left(1-\frac{\xi}{\eta}\right)x$$

と書けることに注意する. 関数 f は凸であることから

$$f(x+\xi)\leq\frac{\xi}{\eta}f(x+\eta)+\left(1-\frac{\xi}{\eta}\right)f(x)=\frac{\xi}{\eta}\bigl\{f(x+\eta)-f(x)\bigr\}+f(x)$$

となる. つまり $0 < \xi < \eta$ のとき

$$\frac{f(x+\xi)-f(x)}{\xi} \leq \frac{f(x+\eta)-f(x)}{\eta}$$

となり, $\xi \searrow 0$ とすれば

$$\overset{\bullet}{f}_+(x) \leq \frac{f(x+\eta)-f(x)}{\eta} \tag{A.27}$$

がわかる. 同様に

$$x-\xi = \frac{\xi}{\eta}(x-\eta) + \left(1-\frac{\xi}{\eta}\right)x$$

と表せるから

$$f(x-\xi) \leq \frac{\xi}{\eta}f(x-\eta) + \left(1-\frac{\xi}{\eta}\right)f(x) = \frac{\xi}{\eta}\big\{f(x-\eta)-f(x)\big\} + f(x)$$

である. したがって $0 < \xi < \eta$ のとき

$$\frac{f(x)-f(x-\eta)}{\eta} \leq \frac{f(x)-f(x-\xi)}{\xi}$$

となり, $\xi \searrow 0$ とすれば

$$\overset{\bullet}{f}_-(x) \geq \frac{f(x)-f(x-\eta)}{\eta} \tag{A.28}$$

を得る. さらに

$$f(x) = f\left(\frac{(x+\xi)+(x-\xi)}{2}\right) \leq \frac{f(x+\xi)+f(x-\xi)}{2}$$

となり

$$f(x)-f(x-\xi) \leq f(x+\xi)-f(x)$$

である. だから

$$\frac{f(x)-f(x-\xi)}{\xi} \leq \frac{f(x+\xi)-f(x)}{\xi}$$

ががわかる. ここで, $\xi \searrow 0$ とすると

$$\overset{\bullet}{f}_-(x) \leq \overset{\bullet}{f}_+(x) \tag{A.29}$$

がわかる. したがって, (A.27) – (A.29) を考慮すれば

$$\frac{f(x)-f(x-\eta)}{\eta} \leq \overset{\bullet}{f}_-(x) \leq \overset{\bullet}{f}_+(x) \leq \frac{f(x+\eta)-f(x)}{\eta} \tag{A.30}$$

がわかる. よって $\dot{f}_-(x)$ と $\dot{f}_+(x)$ は有界であることがわかる.

次に, 単調性を示す. そのために $a < x_1 < x_2 < b$ とする. (A.30) の

$$f_-(x) \leq f_+(x) \leq \frac{f(x+\eta) - f(x)}{\eta}$$

において, $x = x_1$, $x + \eta = x_2$ とおくと

$$f_-(x_1) \leq f_+(x_1) \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

を得る. 同様に (A.30) の

$$\frac{f(x) - f(x-\eta)}{\eta} \leq f_-(x) \leq f_+(x)$$

において, $x = x_2$, $x - \eta = x_1$ とおくと

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f_-(x_2) \leq f_+(x_2)$$

を得る. これらの式を合わせると

$$\dot{f}_-(x_1) \leq \dot{f}_+(x_1) \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \dot{f}_-(x_2) \leq \dot{f}_+(x_2) \tag{A.31}$$

となる. よって

$$\dot{f}_-(x_1) \leq \dot{f}_-(x_2), \quad \dot{f}_+(x_1) \leq \dot{f}_+(x_2)$$

が得られる. □

定理 A.39. 区間 (a, b) 上で定義された凸関数 f は, 単調増加で右連続関数 g と点 $c \in (a, b)$ によって

$$f(x) = f(c) + \int_c^x g(t)\,\mathrm{dm}(t) \qquad (a < x < b)$$

と表せる.

Proof. $a < c < x < b$ とし, 閉区間 $[c, x]$ の分割を

$$c = x_0 < x_1 < \cdots < x_n = x$$

とする. このとき (A.31) より, $k = 1, 2, \ldots, n$ に対し

$$\dot{f}_-(x_{k-1}) \leq \dot{f}_+(x_{k-1}) \leq \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \leq \dot{f}_-(x_k) \leq \dot{f}_+(x_k)$$

と書ける．これらに $x_k - x_{k-1}$ を掛ければ

$$\begin{aligned}\overset{\bullet}{f}_-(x_{k-1})(x_k - x_{k-1}) &\leq \overset{\bullet}{f}_+(x_{k-1})(x_k - x_{k-1}) \leq f(x_k) - f(x_{k-1}) \\ &\leq \overset{\bullet}{f}_-(x_k)(x_k - x_{k-1}) \leq \overset{\bullet}{f}_+(x_k)(x_k - x_{k-1})\end{aligned}$$

が成立し，上の不等式の辺々について，1 から n まで k について和をとると

$$\begin{aligned}\sum_{k=1}^n \overset{\bullet}{f}_-(x_{k-1})(x_k - x_{k-1}) \leq \sum_{k=1}^n \overset{\bullet}{f}_+(x_{k-1})(x_k - x_{k-1}) &\leq f(x) - f(c) \\ \leq \sum_{k=1}^n \overset{\bullet}{f}_-(x_k)(x_k - x_{k-1}) &\leq \sum_{k=1}^n \overset{\bullet}{f}_+(x_k)(x_k - x_{k-1})\end{aligned}$$

が成立する．$\overset{\bullet}{f}_-$ と $\overset{\bullet}{f}_+$ は単調増加だから Riemann 積分可能である．よって $\max_{1\leq k\leq n}\{x_k - x_{k-1}\} \to 0\,(n \to \infty)$ とすれば

$$\begin{aligned}\sum_{k=1}^n \overset{\bullet}{f}_-(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-(t)\,\mathsf{dm}(t) \\ \sum_{k=1}^n \overset{\bullet}{f}_+(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+(t)\,\mathsf{dm}(t) \\ \sum_{k=1}^n \overset{\bullet}{f}_-(x_k)(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-(t)\,\mathsf{dm}(t) \\ \sum_{k=1}^n \overset{\bullet}{f}_+(x_k)(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+(t)\,\mathsf{dm}(t)\end{aligned}$$

となり

$$f(x) - f(c) = \int_c^x \overset{\bullet}{f}_-(t)\,\mathsf{dm}(t) = \int_c^x \overset{\bullet}{f}_+(t)\,\mathsf{dm}(t)$$

がわかる．他方 $a < x < b$ のときも閉区間 $[x,\, c]$ について同様の論理を展開して

$$f(c) - f(x) = \int_x^c \overset{\bullet}{f}_-(t)\,\mathsf{dm}(t) = \int_x^c \overset{\bullet}{f}_+(t)\,\mathsf{dm}(t)$$

を示すことができる．したがって

$$g(x) = \overset{\bullet}{f}_+(x+0)$$

とおけば，g は右連続で単調増加で

$$f(x) = f(c) + \int_c^x g(t)\,\mathsf{dm}(t)$$

となる． □

定理 A.40. 関数 f が区間 (a, b) 上の凸であるための必要十分条件は, $\forall x_0 \in (a, b)$ について関数

$$(a, b) \setminus \{x_0\} \ni x \mapsto \frac{f(x) - f(x_0)}{x - x_0} \in \mathbb{R} \tag{A.32}$$

が単調増加となることである.

Proof. <u>必要性</u>: 関数 f が開区間 (a, b) 上で凸ならば, 定理 A.39 により, 右連続な単調増加関数 g が存在して

$$f(x) - f(x_0) = \int_{x_0}^{x} g(t)\,\mathrm{dm}(t)$$

と表せる. よって $x \neq x_0$ のとき

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{1}{x - x_0} \int_{x_0}^{x} g(t)\,\mathrm{dm}(t)$$

は $(a, b) \setminus \{x_0\}$ 上で単調増加である.

<u>十分性</u>: (A.32) が成立したと仮定する. $\forall x, y \in (a, b)$, $x < y$, $0 < \lambda < 1$ に対し

$$x_0 = \lambda x + (1 - \lambda) y$$

とおく. (A.32) より $x < x_0 < y$ のとき

$$\frac{f(x_0) - f(x)}{x_0 - x} \leq \frac{f(y) - f(x_0)}{y - x_0}.$$

この式に $x_0 = \lambda x + (1 - \lambda) y$ を代入すると

$$\frac{f\bigl(\lambda x + (1 - \lambda) y\bigr) - f(x)}{(1 - \lambda)(y - x)} \leq \frac{f(y) - f\bigl(\lambda x + (1 - \lambda) y\bigr)}{\lambda (y - x)}$$

を得る. 上の不等式の辺々に $\lambda(1 - \lambda)(y - x)$ を掛けと

$$\lambda \Bigl\{ f\bigl(\lambda x + (1 - \lambda) y\bigr) - f(x) \Bigr\} \leq (1 - \lambda) \Bigl\{ f(y) - f\bigl(\lambda x + (1 - \lambda) y\bigr) \Bigr\}$$

を得る. これを整理すれば

$$f\bigl(\lambda x + (1 - \lambda) y\bigr) \leq \lambda f(x) + (1 - \lambda) f(y)$$

がわかる. よって, $\lambda = 0, 1$ のときは自明だから, 関数 f は (a, b) 上で凸であることがわかる. □

定義 A.41. 関数 f を (a, b) 上の凸とする. 点 $x_0 \in (a, b)$ において適当に実数 m を定めると $\forall x \in (a, b)$ に対して

$$S(x) := m(x - x_0) + f(x_0) \leq f(x) \tag{A.33}$$

が成立するとき, 直線 $y = S(x)$ を点 x_0 において f を下から支える直線という.

定理 A.42. 関数 f が区間 (a, b) 上の凸であるための必要十分条件は, (a, b) の各点 x_0 において f を下から支える直線 $y = S(x)$ が存在することである.

Proof. 必要性: 関数 f は開区間 (a, b) 上の凸で $x_0 \in (a, b)$ なので, 定理 A.39 より, ある右連続な単調増加関数 g が存在して

$$f(x) = f(x_0) + \int_{x_0}^{x} g(t)\,\mathrm{dm}(t)$$

と書ける. このとき

$$f(x) \geq f(x_0) + g(x_0)\int_{x_0}^{x} \mathrm{dm}(t) = f(x_0) + g(x_0)(x - x_0)$$

となることがわかる. したがって, 直線

$$S(x) := f(x_0) + g(x_0)(x - x_0)$$

は点 x_0 で $f(x)$ を下から支えることがわかる.

十分性: $x, y \in (a, b)$, $x < y$, $0 \leq \lambda \leq 1$ とし

$$x_0 = \lambda x + (1 - \lambda)y$$

とおく. 仮定より x_0 において凸関数 f を下から支える直線 $S(x)$ が存在する. すると (A.33) から $S(x) \leq f(x)$ とである. さらに, S の線型性から

$$\begin{aligned} f\big(\lambda x + (1-\lambda)y\big) &= f(x_0) = S(x_0) = \lambda S(x) + (1-\lambda)S(y) \\ &\leq \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

となる. よって, 関数 f は (a, b) 上の凸である. □

A.7 共役関数と Young の不等式

$\varphi\colon [0, \infty) \longrightarrow [0, \infty)$ を単調増加で右連続な関数とし

$$\varphi(0) = 0, \quad \varphi(t) > 0 \quad (t > 0), \quad \varphi(\infty) = \infty \tag{A.34}$$

をみたすと仮定する. このとき, φ に**右逆関数** $\varphi^{\leftarrow}$ を

$$\varphi^{\leftarrow}(x) := \sup\{y \geq 0;\, \varphi(y) \leq x\}$$

で定義する.

注意 A.43. $x > 0$ に対して, 左逆関数は

$$\varphi(x)^{\leftarrow} = \inf\{y > 0; \varphi(y) > x\}$$

とも表現できる. また, F を分布関数としてとき, F の quantile 関数を

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

で定めている. F^{-} を F の左逆関数とも呼ぶことがある. □

補題 A.44. 関数 φ は $(A.34)$ をみたすとする. このとき, φ の右逆関数 $\varphi^{\leftarrow}$ は右連続で

$$\varphi^{\leftarrow}(0) = 0, \quad \varphi^{\leftarrow}(t) > 0 \quad (t > 0), \quad \varphi^{\leftarrow}(\infty) = \infty$$

である.

Proof. ① $\varphi^{\leftarrow}(0) = 0$ の証明: $\varphi(0) = 0$ かつ $\varphi(t) > 0\,(t > 0)$ だから

$$\{y \geq 0; \varphi(y) \leq 0\} = \{0\}$$

となる. このことから $\varphi^{\leftarrow}(0) = 0$ がわかる.
② $\varphi^{\leftarrow}$ 単調性の証明: $x_1 < x_2$ とすると

$$\varphi(y) \leq x_1 \;\Rightarrow\; \varphi(y) \leq x_2$$

である. よって

$$\{y \in \mathbb{R}; \varphi(y) \leq x_1\} \subset \{y \in \mathbb{R}; \varphi(y) \leq x_x\}$$

となるので

$$\varphi^{\leftarrow}(x_1) = \sup\{y \in \mathbb{R}; \varphi(y) \leq x_1\} \leq \sup\{y \in \mathbb{R}; \varphi(y) \leq x_x\} = \varphi^{\leftarrow}(x_2)$$

がわかる.
③ $\varphi^{\leftarrow}(\infty) = \infty$ の証明: 背理法で証明するために, $\varphi^{\leftarrow}(\infty) < \infty$ と仮定する. 背理法の仮定から, $y \geq 0$ を

$$\varphi^{\leftarrow}(\infty) < y < \infty$$

をみたすように取れる. するとすべての $x \geq 0$ に対して

$$\varphi(x) \leq y$$

となる. ここで, $x \to \infty$ とすると

$$\varphi(\infty) \leq y$$

となり, $\varphi(\infty) = \infty$ と矛盾する. よって, $\varphi^{\leftarrow}(\infty) = \infty$ が示せた.
④ $\varphi^{\leftarrow}$ 右連続性の証明: これを示すために, $x \geq 0$ を固定し, $0 \leq x < x_n,\, x_n \downarrow x\,(n \to \infty)$ とする. $\varphi^{\leftarrow}$ は単調増加だから

$$\varphi^{\leftarrow}(x) \leq \varphi^{\leftarrow}(x_n) \downarrow \varphi^{\leftarrow}(x+0)$$

である. 仮に

$$\varphi^{\leftarrow}(x) < \varphi^{\leftarrow}(x+0) \tag{A.35}$$

と仮定して矛盾を導こう. $(A.35)$ から, ある $y \geq 0$ があって

$$\varphi^{\leftarrow}(x) < y < \varphi^{\leftarrow}(x+0) \tag{A.36}$$

とできる. すると $\forall n \in \mathbb{N}$ に対して

$$y < \varphi^{\leftarrow}(x_n) = \sup\{y \geq 0;\, \varphi(y) \leq x_n\}$$

と φ の単調増加性から

$$\varphi(y) \leq x_n$$

が成り立つ. ここで, $n \to \infty$ とすると

$$\varphi(y) \leq x$$

を得る. よって

$$y \in \{s \geq 0;\, \varphi(s) \leq x\} \;\Rightarrow\; y \leq \sup\{s \geq 0;\, \varphi(s) \leq x\} = \varphi^{\leftarrow}(x)$$

から

$$\mathrm{y}\ \leq \varphi^{\leftarrow}(x)$$

となる. これは $(A.36)$ と矛盾する. よって, $\varphi^{\leftarrow}$ は右連続であることが示せた. □

補題 A.45. $x \geq 0$ に対して

$$x \leq \varphi\big(\varphi^{\leftarrow}(x)\big), \quad x \leq \varphi^{\leftarrow}\big(\varphi(x)\big)$$

である. さらに, $\forall \epsilon > 0$ に対して

$$\varphi\big(\varphi^{\leftarrow}(x) - \epsilon\big) \leq x, \quad \varphi^{\leftarrow}\big(\varphi(x) - \epsilon\big) \leq x$$

が成り立つ.

Proof. $x \le \varphi\bigl(\varphi^{\leftarrow}(x)\bigr)$ を示す. $\varphi\bigl(\varphi^{\leftarrow}(0)\bigr) = 0$ なので, $x > 0$ に対して示せばよい.

$$\varphi^{\leftarrow}(x) = \inf\{y > 0;\, \varphi(y) > x\}$$

とも表現できることに注意する. これより

$$\forall y > \varphi^{\leftarrow}(x) \;\Rightarrow\; \varphi(y) \ge x$$

である. $y \downarrow \varphi^{\leftarrow}(x)$ とすると, φ の右連続性より

$$\varphi\bigl(\varphi^{\leftarrow}(y)\bigr) = \lim_{y \downarrow \varphi^{\leftarrow}(x)} \varphi(y) \ge x \tag{A.37}$$

がわかる. また, $(A.37)$ から, 直ちに

$$\varphi\bigl(\varphi^{\leftarrow}(x) - \epsilon\bigr) \le x$$

がわかる. のこりの不等式も同様に示せばよい. □

$(A.34)$ をみたす単調増加かつ右連続関数 φ とその右逆関数 $\varphi^{\leftarrow}$ に対して

$$\Phi(x) := \int_0^x \varphi(t)\,\mathrm{d}\mathsf{m}(t), \quad \Phi(x)^{\vee} := \int_0^x \varphi^{\leftarrow}(t)\,\mathrm{d}\mathsf{m}(t) \qquad (x \ge 0)$$

と定める. 定理 $A.37$ から Φ と $\Phi^{\vee}$ はともに凸関数になる. $\Phi^{\vee}$ を凸関数 Φ の**共役凸関数**とよぶ.

例 A.46. $1 < p < \infty$, $p^{-1} + q^{-1} = 1$ とする.

$$\varphi(x) = x^{p-1} \qquad (0 \le x < \infty)$$

とおくと

$$\varphi^{\leftarrow}(x) = \sup\{y \ge 0;\, \varphi(y) \le x\} = \sup\{y \ge 0;\, y^{p-1} \le x\} = x^{1/(p-1)}$$

となる. よって

$$\begin{aligned}
\Phi(x) &= \int_0^{\infty} \varphi(t)\,\mathrm{d}\mathsf{m}(t) = \int_0^{\infty} t^{p-1}\,\mathrm{d}\mathsf{m}(t) = \frac{x^p}{p}, \\
\Phi(x)^{\vee} &= \int_0^x \varphi^{\leftarrow}(t)\,\mathrm{d}\mathsf{m}(t) = \int_0^x t^{1/(p-1)}\,\mathrm{d}\mathsf{m}(t) = \frac{x^{1+1/(p-1)}}{1+1/(p-1)} = \frac{x^q}{q}
\end{aligned}$$

となる. □

定理 A.47. $\forall x,\, y \geq 0$ に対して

$$xy \leq \Phi(x) + \Phi^{\vee}(y) \tag{A.38}$$

が成り立つ. とくに

$$x\varphi(x) = \Phi(x) + \Phi^{\vee}(\varphi(x)) \tag{A.39}$$

$$y\varphi^{\leftarrow}(y) = \Phi(\varphi^{\leftarrow}(y)) + \Phi^{\vee}(y) \tag{A.40}$$

が成り立つ.

Proof. 図を書けばよい. □

系 A.48.

$$\Phi^{\vee}(y) = \sup_{x \geq 0}\{xy - \Phi(x)\}$$

が成り立つ.

Proof. $y \geq 0$ に対して, (*A*.37) から

$$xy - \Phi(x) \leq \Phi^{\vee}(y)$$

がわかる. したがって

$$\sup_{x \geq 0}\{xy - \Phi(x)\} \leq \Phi^{\vee}(y)$$

を得る. $x = \varphi^{\leftarrow}(y)$ とおくと (*A*.40 から

$$xy = y\varphi^{\leftarrow}(y) = \Phi\big(\varphi^{\leftarrow}(y)\big) + \Phi^{\vee}(y)$$

となるので

$$\Phi^{\vee}(y) = \sup_{x \geq 0}\{xy - \Phi(x)\}$$

がわかる.

例 A.49 (例 *A*.46 の続き)**.** 系 *A*.48 を用いて, $\Phi^{\vee}$ を求めてみよう.

$$\Phi^{\vee}(y) = \sup_{x \geq 0}\left\{xy - \frac{x^p}{p}\right\} =: \sup_{x \geq 0} h(x)$$

であった. すると

$$\dot{h}(x) = y - x^{p-1} = 0$$

なので, $x = y^{-(p-1)}$ である. よって

$$\Phi^{\vee}(y) = y^{-(p-1)}y - \frac{y^{p/(p-1)}}{p} = y^{1+(p-1)^{-1}} - \frac{y^q}{p} = y^q - \frac{y^q}{p} = \left(1 - \frac{1}{p}\right)y^q$$
$$= \frac{y^q}{q}$$

となる.

A.8 距離空間とノルム空間

$\mathcal{M}$ を空でない集合とする. $\mathcal{M}$ 上の距離 d とは, 写像 $d : \mathcal{M}\times\mathcal{M} \longrightarrow [0, \infty)$ で以下の性質をもつものである: 任意の $x, y, z \in \mathcal{M}$ に対して
(i) $d(x, y) = d(y, x)$;
(ii) $d(x, z) \le d(x, y) + d(y, z)$;
(iii) $d(x, y) \Longleftrightarrow x = y$.
(i) と (ii) をみたし, (iii) を $d(x, x) = 0$ にかえたものを擬 (準) 距離 (pseudo-metric [5]) という. 組 $(\mathcal{M}, d)$ を距離空間という.

開球とは, 任意の $x \in \mathcal{M}$ と任意の $r > 0$ に対し,

$$U(x; r) := \{y \in \mathcal{M}; d(x, y) < r\}$$

の形で作れたものをいう. $A \subset \mathcal{M}$ が開集合であるとは, 開球の和集合となるときをいう. A が閉集合とは, A の補集合が開集合のときをいう.

$\mathcal{M}$ の点列 $\{x_n\}_{n=1}^{\infty}$ が $\mathcal{M}$ の点 x に収束すうるとは,

$$d(x_n, x) \to 0\,(n \to \infty)$$

のときをいう. これを $x_n \longrightarrow x\,(n \to \infty)$ と記す.

$A \subset \mathcal{M}$ の閉包とは, A を含む最小（包含関係の意味）の閉集合のことをいい, これを $\overline{A}$ と記す. A の内部とは, A に含まれる最大の開集合のことをいい, A° と記す.

$\mathcal{M}_1, \mathcal{M}_2$ を距離空間としたとき, 函数 $f : \mathcal{M}_1 \to \mathcal{M}_2$ が連続であるとは, すべての $x_n \to x\,(n \to \infty)$ なる $\mathcal{M}_1$ の数列 $\{x_n\}_{n=1}^{\infty}$ に対して, $f(x_n) \to f(x)\,(n \to \infty)$ が成り立つときをいう. $\mathcal{M}$ の空でない部分集合 D が稠密であるとは, $\overline{D} = \mathcal{M}$ のときをいう. 距離空間が可分であるとは, $\mathcal{M}$ が稠密可算部分集合をもつことをいう. $\mathcal{M}$ の部分集合 K がコンパクトであるとは, K は閉集合であって, K の任意の点列は収束する部分列をもつときをいう. 部分集合 K が全有界であるとは, 任意の $\epsilon > 0$ に対して, 有限個の半径　ϵ の開球で K を覆うことができるときをいう.

半距離 d が完備であるとは, 任意のコーシー列が $\mathcal{M}$ のある点に収束するときをいう. $\mathcal{M}$ が完備な準距離をもつとき, $\mathcal{M}$ は完備であるという. 完備な準距離空間 $\mathcal{M}$ の部分集合 K がコンパクトであるための必要十分条件は, K が全有界かつ閉集合であることである.

ノルム空間 $\mathcal{M}$ とは, ノルムをもつベクトル空間のことをいう. ノルムとは, 写像 $\|\cdot\| : \mathcal{M} \to [0, \infty)$ であって, すべての $x, y \in \mathcal{M}$ と $\alpha \in \mathbb{R}$ に対し,
(i) $\|x + y\| \le \|x\| + \|y\|$;
(ii) $\|\alpha x\| = |\alpha|\|x\|$;

[5] スードウメトリック

(iii) $\|x\| = 0 \Longleftrightarrow x = 0$.
(i) と (ii) をみたすが, (iii) をみたさないとき, 半ノルムという. ノルム $\|\cdot\|$ が与えらると, $d(x, y) = \|x - y\|$ によって距離が定まる.

定義 A.50. 距離空間 $\mathcal{M}$ のボレル σ 加法族とは, $\mathcal{M}$ の開集合族を含む最小の σ 加法族である. これを $\mathcal{B}(\mathcal{M})$ と記す. $\mathcal{M}_1$, $\mathcal{M}_2$ を距離空間とする. 写像 $f : \mathcal{M}_1 \to \mathcal{M}_2$ がボレル可測であるとは, 任意の $B \in \mathcal{B}(\mathcal{M}_2)$ に対して, $f^{-1}(B) \in \mathcal{B}(\mathcal{M}_1)$ のときをいう. 確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上のボレル可測写像 $X : \Omega \to \mathcal{M}$ を $\mathcal{M}$ 値確率要素または $\mathcal{M}$ 値確率変数という.

補題 A.51. 連続写像はボレル可測である.

Proof. van der Vaart (1998, p.296) を参照のこと. □

$\mathbb{T}$ を空でない集合とし, $\ell^\infty(\mathbb{T})$ を有界函数 $f : \mathbb{T} \to \mathbb{R}$ の全体のなす集合とする. $f, g \in \ell^\infty(\mathbb{T})$ に対して, 和 $f+g$ を $(f+g)(t) = f(t)+g(t)$ $(t \in \mathbb{T})$ で定めると, $f + g \in \ell^\infty(\mathbb{T})$ となる. 一様ノルムを

$$\|f\|_{\mathbb{T}} := \sup_{t\in\mathbb{T}} |f(t)|$$

で定める. 一様ノルムをもつ $\ell^\infty(\mathbb{T})$ が可分であるための必要十分条件は $\mathbb{T}$ が高々可算集合となることである. これは厳しい条件である.