第3章　確率と期待値の不等式

直接計算するのが困難な確率や期待値に対して上限ないしは下限を与える不等式は有効である.

3.1　確率に対する不等式

定理 3.1. (Markov の不等式) 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の非負値確率変数を X とし, $\mathsf{E}[X] < \infty$ とする. このとき, $\forall t > 0$ に対して

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

が成り立つ.

Proof. X は連続型で p.d.f. p を持つ場合を示す.

$$\begin{aligned}\mathsf{E}[X] &= \int_0^\infty x\mathsf{p}(x)\,\mathrm{d}x = \int_0^t x\mathsf{p}(x)\,\mathrm{d}x + \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \geq \int_t^\infty x\mathsf{p}(x)\,\mathrm{d}x \\ &\geq t\int_t^\infty \mathsf{p}(x)\,\mathrm{d}x = t\mathsf{Pr}(X \geq t)\end{aligned}$$

からわかる. □

系 3.2. $\lambda > 0$ とする. X を確率変数とし $\mathsf{E}\bigl[e^{\lambda X}\bigr] < \infty$ とする. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\bigl(X \geq t\bigr) \leq e^{-\lambda t}\mathsf{E}\bigl[e^{\lambda X}\bigr]$$

が成り立つ.

Proof. $e^{\lambda X}$ は非負値確率変数なので定理 3.1 を適用すれば

$$\mathsf{Pr}\bigl(X \geq t\bigr) = \mathsf{Pr}\bigl(e^{\lambda X} \geq e^{\lambda t}\bigr) \leq e^{-\lambda t}\mathsf{E}\bigl[e^{\lambda X}\bigr]$$

となる. □

系 3.3. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X] < \infty$ $(0 < \sigma < \infty)$ とする．このとき $\forall t > 0$ に対して

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

が成り立つ．

Proof. $(X - \mu)^2$ に対して Markov の不等式 (定理 3.1) を適用する．すると

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

がわかる． □

次に Hoeffding の不等式を証明するための補題を与える．

補題 3.4. $a, b \in \mathbb{R}$ は $a < 0 < b$ なる定数とする．確率変数 X は

$$\mathsf{E}[X] = 0, \quad a \leq X \leq b$$

とみたすとき $\forall \lambda > 0$ に対して

$$\mathsf{E}\big[e^{\lambda X}\big] \leq e^{\lambda^2 (b-a)^2/8}$$

となる．

Proof. X を以下のように書き直す．

$$X = \gamma b + (1 - \gamma) a, \quad \gamma = \frac{X - a}{b - a}$$

となる．すると $e^{\lambda x}$ の凸性より

$$e^{\lambda X} \leq \gamma e^{\lambda b} + (1 - \gamma) e^{\lambda a} = \frac{X - a}{b - a} e^{\lambda b} + \frac{b - X}{b - a} e^{\lambda a}$$

となる．$\mathsf{E}[X] = 0$ に注意して上の式の両辺の期待値を計算すると

$$\begin{aligned}
\mathsf{E}\big[e^{\lambda X}\big] &\leq -\frac{a}{b - a} e^{\lambda b} + \frac{b}{b - a} e^{\lambda a} \\
&= c e^{\lambda b} + (1 - c) e^{\lambda a} \quad \left(c := -\frac{a}{b - a},\ 1 - c = \frac{b}{b - a}\right) \\
&= c e^{\lambda (1-c)(b-a)} + (1 - c) e^{-\lambda c (b-a)} \\
&= e^{-\lambda c (b-a)} \big\{1 - c + c e^{\lambda (b-a)}\big\} \\
&= e^{-cu} \big\{1 - c + c e^{u}\big\} \quad (u := \lambda (b - a)) \\
&= \exp\big\{-cu + \log\big(1 - c + c e^{u}\big)\big\} =: \exp\big\{g(u)\big\}
\end{aligned}$$

がわかる. いま

$$g(0) = \dot{g}(0) = 0, \quad \ddot{g}(u) \leq \frac{1}{4}\,(u > 0) \tag{3.1}$$

に注意する. 実際

$$\begin{aligned}
\dot{g}(u) &= \frac{\mathrm{d}g}{\mathrm{d}u} = -c + \frac{ce^u}{1-c+ce^u}, \\
\ddot{g}(u) &:= \frac{\mathrm{d}^2 g}{\mathrm{d}u^2} = \frac{ce^u}{1-c+ce^u} - \frac{c^2e^{2u}}{(1-c+ce^u)^2} \\
&= \frac{ce^u(1-c)}{(1-c+ce^u)^2} \leq \frac{1}{4} \qquad \Big(\text{相加相乗平均を } 1-c \text{ と } ce^u \text{ に適用する.}\Big)
\end{aligned}$$

からわかる. (3.1) に注意して, Taylor 展開をすれば

$$\begin{aligned}
g(u) &= g(0) + u\,\dot{g}(0) + \frac{u^2}{2}\ddot{g}(\xi) \quad (\xi \in (0,\,u)) \\
&= \frac{u^2}{2}\,\ddot{g}(\xi) \leq \frac{u^2}{8} = \frac{\lambda^2(b-a)^2}{8}
\end{aligned}$$

がわかる. よって

$$\mathsf{E}\big[e^{\lambda X}\big] \leq e^{g(u)} \leq e^{\lambda^2(b-a)^2/8}$$

を得る. □

定理 3.5. (Hoeffding の不等式) $a_j,\, b_j\,(j = 1,\,2,\,\ldots,\,n)$ は $a_j < 0 < b_j$ なる定数とする. 確率空間 $(\Omega,\,\mathcal{A},\,\mathsf{Pr})$ 上の確率変数 $X_1,\,X_2,\,\ldots,\,X_n$ は独立で

$$\mathsf{E}[X_j] = 0, \quad a_j \leq X_j \leq b_j$$

をみたすとする. このとき, $\forall t > 0$ に対して

$$\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) \leq \exp\left\{-\frac{2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right\}$$

が成立する.

Proof. 系 3.2 を用いる. $\forall\lambda > 0$ に対して

$$\begin{aligned}
\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) &\leq e^{-\lambda t}\mathsf{E}\left[\exp\left(\lambda\sum_{j=1}^{n} X_j\right)\right] \\
&= e^{-\lambda t}\prod_{j=1}^{n}\mathsf{E}\big[\exp\{\lambda X_j\}\big] \quad (\because \text{独立性}) \\
&\leq e^{-\lambda t}\prod_{j=1}^{n}\exp\left\{\frac{\lambda^2(b_j-a_j)^2}{8}\right\} \quad (\because \text{補題 } 3.4) \\
&= \exp\left\{-\lambda t + \frac{\lambda^2}{8}\sum_{j=1}^{n}(b_j-a_j)^2\right\} \\
&= \exp\left[\frac{\sum_{j=1}^{n}(b_j-a_j)^2}{8}\left\{\lambda - \frac{4t}{\sum_{j=1}^{n}(b_j-a_j)^2}\right\}^2\right. \\
&\qquad \left. - \frac{2t^2}{\sum_{j=1}^{n}(b_j-a_j)^2}\right]
\end{aligned}$$

より

$$\mathsf{Pr}\left(\sum_{j=1}^{n} X_j \geq t\right) \leq \exp\left\{-\frac{2t^2}{\sum_{j=1}^{n}(b_j-a_j)^2}\right\}$$

がわかる. □

定理 3.6. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする. ただし $0 < \theta < 1$. このとき $\forall t > 0$ に対して

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \geq t\big) \leq 2e^{-2nt^2}$$

となる. ただし $\overline{X}_n = (1/n)\sum_{j=1}^{n} X_j$ である.

Proof. $j = 1, 2, \ldots, n$ に対して

$$Y_j = \frac{1}{n}(X_j - \theta)$$

とおけば

$$\mathsf{E}[Y_j] = 0, \quad a \leq Y_j \leq b, \quad a = -\frac{\theta}{n}, \quad b = \frac{1-\theta}{n}, \quad (b-a)^2 = \frac{1}{n^2}$$

となる. よって定理 3.5 より

$$\mathsf{Pr}\big(\overline{X}_n - \theta \geq t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \geq t\right) \leq e^{-2nt^2} \tag{3.2}$$

となる. 同様に

$$\mathsf{Pr}\big(\overline{X}_n - \theta \le -t\big) = \mathsf{Pr}\left(\sum_{j=1}^{n} Y_j \le -t\right) = \mathsf{Pr}\left(\sum_{j=1}^{n}(-Y_j) \ge t\right) \le e^{-2nt^2} \tag{3.3}$$

がわかる. よって, (3.2) と (3.3) を合わせると

$$\begin{aligned}\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \ge t\big) &= \mathsf{Pr}\big(\big\{\overline{X}_n - \theta \ge t\big\} \cup \big\{\overline{X}_n - \theta \le -t\big\}\big)\\ &\le \mathsf{Pr}\big(\overline{X}_n - \theta \ge t\big) + \mathsf{Pr}\big(\overline{X}_n - \theta \le -t\big)\\ &\le 2e^{-2nt^2}\end{aligned}$$

がわかる. □

例 3.7. $X_1, X_2, \ldots, X_{100} \overset{\text{i.i.d.}}{\sim} \mathsf{Ber}(\theta)$ とする. ただし $0 < \theta < 1$ である.
(1) $n = 100$ とし, 事象 $\big|\overline{X}_{100} - \theta\big| \ge 0.2$ の確率を Chebyshev の不等式を用いて θ に関して一様に上から評価すると

$$\mathsf{Pr}\big(\big|\overline{X}_{100} - \theta\big| \ge 0.2\big) \le 0.0625$$

がわかる.

一方 Hoeffding の不等式を用いて θ に関して一様に評価すると

$$\mathsf{Pr}\big(\big|\overline{X}_{100} - \theta\big| \ge 0.2\big) \le 2e^{-2(100)(0.2)^2} = 0.00067$$

となる.

(2) $0 < \alpha < 1$ を固定する. いま

$$t = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

とおく. すると Hoeffding の不等式より

$$\mathsf{Pr}\left(\big|\overline{X}_n - \theta\big| \ge \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \le \alpha$$

となる. これより

$$\mathsf{Pr}\left(\big|\overline{X}_n - \theta\big| \le \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right) \ge 1 - \alpha$$

を得る. よって

$$C = \left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

とおけば

$$\mathsf{Pr}(\theta \in C) \geq 1-\alpha$$

を得る．すなわち信頼係数 $(1-\alpha)$ の θ の信頼区間

$$\left[\overline{X}_n - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}, \overline{X}_n + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}\right]$$

を得る．

$\alpha = 0.05$, $\theta = 1/2$ として, $n = 10, 20, 50, 100$ の信頼区間の比較の計算を行う．CLT の方は数値計算で．

□

注意 3.8. 信頼区間については節 8.6 を参照のこと．

定理 3.9. (Mill の不等式) $Z \sim \mathsf{N}(0,1)$ とする．このとき $\forall t > 0$ に対して

$$\mathsf{Pr}(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t}$$

が成り立つ．

Proof. まず

$$\begin{aligned}
\mathsf{Pr}(Z \geq t) &= \frac{1}{\sqrt{2\pi}}\int_t^\infty \exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z \\
&\leq \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{z}{t}\exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\int_t^\infty \left\{-\frac{\partial}{\partial z}\exp\left(-\frac{z^2}{2}\right)\right\}\,\mathrm{d}z \\
&= \frac{1}{t\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right) \qquad (3.4)
\end{aligned}$$

が成り立つことに注意する．さらに, $\mathsf{N}(0,1)$ の p.d.f. が偶関数であることに注意すると

$$\begin{aligned}
\mathsf{Pr}(|Z| \geq t) &= \mathsf{Pr}(\{Z \geq t\} \cup \{Z \leq -t\}) \\
&\leq \mathsf{Pr}(Z \geq t) + \mathsf{Pr}(Z \leq -t) \\
&= 2\mathsf{Pr}(Z \geq t)
\end{aligned}$$

からわかる．この不等式と (3.4) を合わせると定理の不等式が得られる．
□

3.2 期待値に対する不等式

定理 3.10. (Cauchy-Schwarz の不等式) 確率変数 X と Y は 2 次の有限な期待値を持つとき

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

となる.

Proof. $\mathsf{E}[X^2] = \mathsf{E}[Y^2] = 0$ のとき, 補題 2.3 から $\mathsf{Pr}(X = 0) = \mathsf{Pr}(Y = 0) = 1$ となるので, 不等式は自明である. このことを踏まえ, $\mathsf{E}[X^2] \neq 0$ として証明を進める.

いま, $g(t) := \mathsf{E}\big[(tX - Y)^2\big]$ とおく. 期待値の中を展開して期待値の線型性を用いると

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

となる. これから

$$\begin{aligned} g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right) &= \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2}{\mathsf{E}[X^2]} \ge 0 \\ &\Leftrightarrow\ \mathsf{E}[X^2]\mathsf{E}[Y^2] - \{\mathsf{E}[XY]\}^2 \ge 0 \\ &\Leftrightarrow\ \Big|\mathsf{E}[XY]\Big| \le \sqrt{\mathsf{E}[X^2]\mathsf{E}[Y^2]} \end{aligned} \tag{3.5}$$

を得る. (3.5) において, X, Y を $|X|, |Y|$ と置き換えると定理の不等式は示される. 等号が成立するのは $g(t) = 0$ が重解を持つときである. 重解を c とおけば

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

となる [1]. □

[1] 非負値確率変数 X に対して

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

であることに注意せよ. 実際 $\mathsf{Pr}(X > 0) > 0$ と仮定する. するとあるの $\epsilon > 0$ 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.

定義 3.11. 関数 $g : \mathbb{R} \to \mathbb{R}$ が凸であるとは各 $x, y \in \mathbb{R}$ と $0 \leq \forall t \leq 1$ 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 3.12. (Jensen の不等式) X を有限な期待値を持つ確率変数とする．
(1) 関数 $g : \mathbb{R} \to \mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

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

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[2]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 3.13. (Young の不等式) $p, q > 1$ とし

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

をみたすとする．このとき $\forall a, b > 0$ に対して

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

となる．

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

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

[2]定理 **??** を参照．

を得る. ここで

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

とおけば g は凸なので

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

がわかる. □

定理 3.14. (1)(Hölder の不等式) p, q は $1 \leq p \leq +\infty$, $1 \leq q \leq +\infty$ と $\dfrac{1}{p} + \dfrac{1}{q} = 1$ とし, 確率変数 X, Y は $\mathsf{E}\big[|X|^p\big] < \infty$, $\mathsf{E}\big[|Y|^q\big] < \infty$ をみたすとする. このとき, $\mathsf{E}\big[|XY|\big] < \infty$ で

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

となる.

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

$$\big\{\mathsf{E}\big[|X+Y|^p\big]\big\}^{1/p} \leq \big\{\mathsf{E}\big[|X|^p\big]\big\}^{1/p} + \big\{\mathsf{E}\big[|Y|^p\big]\big\}^{1/p}$$

となる.

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

系 3.13 から

$$|XY| \leq \frac{|X|^p}{p} + \frac{|Y|^q}{q} \tag{3.6}$$

である. まず, $\mathsf{E}[|X|^p] = 1$, $\mathsf{E}[|Y|^q] = 1$ とする. (3.6) の両辺の期待値を取ると

$$\mathsf{E}[|XY|] \leq \frac{\mathsf{E}[|X|^p]}{p} + \frac{\mathsf{E}[|Y|^q]}{q} = \frac{1}{p} + \frac{1}{q} = 1 \tag{3.7}$$

を得る．$\mathsf{E}[|X|^p] \neq 1$ または $\mathsf{E}[|Y|^q] \neq 1$ のとき, (3.7) において, X, Y を $X/\{\mathsf{E}[|X|^p]\}^{1/p}$, $Y/\{\mathsf{E}[|Y|^q]\}^{1/q}$ と置き換えると

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

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

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

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. □

3.3 章末注釈と参考文献

この章で扱った凸関数は統計的推測理論では重要な役割を果たす. Hoeffding の不等式では, 統計的機械学習理論で用いられる基本不等式である．この章は [32, 4 章] を参考にした.

3.4 演習問題

演習問題 3.1. $X \sim \mathsf{Po}(\theta)\,(\theta > 0)$ とする．このとき, 以下の問いに答えよ.

(1) X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

(2) Chebyshev の不等式を用いて

$$\mathsf{Pr}\big(X \geq 2\theta\big) \leq \frac{1}{\theta}$$

を示せ.

演習問題 3.2. $n \in \mathbb{N}$ とする. $1 < p < \infty$, $p^{-1} + q^{-1} = 1$ のとき[3], 任意の実数 x_j, y_j $(j = 1, 2, \ldots, n)$ に対して

$$\sum_{j=1}^{n} |x_j y_j| \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} \left(\sum_{j=1}^{n} |y_j|^q\right)^{1/q}$$

が成り立つことを Young の不等式を利用して示せ. この不等式を Hölder の不等式という.

演習問題 3.3. $s > t > 0$ とする. 確率変数 X が $\mathsf{E}\big[|X|^s\big] < \infty$ ならば, $\mathbb{E}\big[|X|^t\big] < \infty$ となることを Young の不等式を用いて示せ.

演習問題 3.4. $X \sim \mathsf{N}(0, 1)$ とする. $\mathsf{Var}[g(X)] < \infty$ なる 可微分関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ に対して

$$\mathsf{Var}[g(X)] \leq \mathsf{E}\big[\{\overset{\bullet}{g}\,(X)\}^2\big]$$

が成立することを示せ. ただし, $\overset{\bullet}{g}\,(x) = \dfrac{\mathrm{d}g}{\mathrm{d}x}(x)$ である.

[3]問題を易しくするために, $p = 1$ の場合を除いていることに注意せよ.