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 X を非負値確率変数とし, $\mathsf{E}[X] < \infty$ とする．このとき, $\forall t > 0$ に対して

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

が成り立つ.

Proof. X は連続型で p.d.f. p を持つ場合を示す.

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

□

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

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

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

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

□

系 3.3. (Chebyshev の不等式) X を確率変数とし $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X]$ $(0 < \sigma < \infty)$ 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．

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

□

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

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

とみたすとき $\forall \lambda > 0$ に対して

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

Proof. X を以下のように書き直す．

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

すると $e^{\lambda x}$ の凸性より

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

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

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

すると

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

となる．実際

$$\dot{g}(u) = \frac{dg}{du} = -\lambda + \frac{ee^u}{1-c+ce^u},$$
$$\ddot{g}(u) := \frac{d^2g}{du^2} = \frac{ce^u}{1-c+ce^u} - \frac{c^2e^{2u}}{(1-c+ce^u)^2}$$
$$= \frac{ce^u(1-c)}{(1-c+ce^u)^2} \leq \frac{1}{4} \qquad \text{（相加相乗平均を } 1-c \text{ と } ce^u \text{ に適用する．）}$$

からわかる．Taylor 展開をすれば

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

よって

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

□

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

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

をみたすとき $\forall t > 0$ に対して

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

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

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

より

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

□

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

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

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

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

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

同様に

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

よって

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

□

例 3.7. (1) $X_1, X_2, \ldots, X_{100} \sim$ i.i.d. $\mathsf{Ber}(\theta)$ とする．ただし $0 < \theta < 1$ である．事象 $\big|\overline{X}_n - \theta\big| \geq 0.2$ 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．

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \geq 0.2\big) \leq 0.0625.$$

一方 Hoeffding の不等式を用いて θ に関して一様に評価すると

$$\mathsf{Pr}\big(\big|\overline{X}_n - \theta\big| \geq 0.2\big) \leq 2e^{-2(100)(0.2)^2} = 0.00067$$

となる．
(2) $0 < \alpha < 1$ を固定する．

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

とおけば Hoeffding の不等式より

$$\mathsf{Pr}\left(\left|\overline{X}_n - \theta\right| \geq \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}\right) \leq \alpha.$$

これより

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

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

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

を得る．すなわち信頼係数 $(1-\alpha)$ の p 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． □

注意 3.8. 信頼区間については節 8.4 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

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

Proof.

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

2 番目の不等式は $\mathsf{N}(0,1)$ の p.d.f. が偶関数であることからわかる. 3 番目の不等式は

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

からわかる. □

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

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

Proof. $\mathsf{E}[X^2] = \mathsf{E}[Y^2] = 0$ 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. $g(t) = \mathsf{E}\big[(tX - Y)^2\big]$ 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

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

となる.

$$g\left(\frac{\mathsf{E}[XY]}{\mathsf{E}[X^2]}\right) = \frac{\mathsf{E}[X^2]\mathsf{E}[Y^2] - \mathsf{E}[XY]}{\mathsf{E}[X^2]} \geq 0 \Longleftrightarrow \mathsf{E}[X^2]\mathsf{E}[Y^2] - \mathsf{E}[XY] \geq 0$$

より不等式は示された. 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

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

となる[1]. □

[1]非負値確率変数 X に対して

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

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 $X \geq \epsilon \mathbb{1}\{X > \epsilon\}$ 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 $\mathsf{Pr}(X > 0) > 0 \Longleftrightarrow \mathsf{Pr}(X = 0) = 1$ がわかる.

定義 3.11. 関数 $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ が凸であるとは各 $x,\, y \in \mathbb{R}$ と $0 \le \forall t \le 1$ に対して

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

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

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 $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ は凸で $g(X)$ の期待値は有限のとき

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

(2) g が concave のとき

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

Proof. $\forall x \in \mathbb{R}$ に対してある定数 $r \in \mathbb{R}$ が存在[2]して

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

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

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

□

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

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

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

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

を得る．ここで

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

[2]定理 D.9 を参照.

とおけば g は凸なので

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

□

3.3　章末注釈と参考文献

3.4　演習問題