

§14 中心極限定理と正規近似 演習問題2 解答

📎 問題の難易度の目安【基礎】★☆☆ 【標準】★★☆ 【発展】★★★

1 (☆☆☆)(Laplace-DeMoivre の定理の証明のための準備)

$X_1, \dots, X_n, \dots$ を互いに独立かつ同じ分布に従う確率変数で, $p, q \geq 0, p + q = 1$ に対して

$$\begin{cases} \mathbb{P}[X_i = 1] = p \\ \mathbb{P}[X_i = 0] = q \end{cases}$$

をみたすとする. このとき期待値 $\mathbb{E}$ と分散 $\mathbb{V}$ について

$$\mathbb{E}[X_1 + \dots + X_n] = np,$$

$$\mathbb{V}[X_1 + \dots + X_n] = npq$$

となることを示せ.

解 各 i について, $\mathbb{E}[X_i] = 1 \cdot \mathbb{P}[X_i = 1] + 0 \cdot \mathbb{P}[X_i = 0] = p$ であるから,

$$\mathbb{E}[X_1 + \dots + X_n] = \sum_{i=1}^n \mathbb{E}[X_i] = np.$$

同様に各 i について, $\mathbb{V}[X_i] = (1 - p)^2 \cdot \mathbb{P}[X_i = 1] + (0 - p)^2 \cdot \mathbb{P}[X_i = 0] = q^2 p + p^2 q = qp$ であるから,

$$\mathbb{V}[X_1 + \dots + X_n] = \sum_{i=1}^n \mathbb{V}[X_i] = npq.$$

2 (★★★)(Laplace-DeMoivre の定理)

$X_1, \dots, X_n, \dots$ を **1** の仮定をみたす確率変数とする. $S_n := X_1 + \dots + X_n$ とおく. このとき, 任意の $-\infty < a < b < \infty$ に対して

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[a \leq \frac{S_n - np}{\sqrt{npq}} \leq b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \quad \dots \textcircled{*}$$

が成り立つことを以下の手順に従って示せ. $\textcircled{*}$ を **Laplace-DeMoivre の定理** という.

(1) S_n は 2 項分布 $B(n, p)$ に従うことを確かめよ.

(2) $\widehat{S}_n := \frac{S_n - np}{\sqrt{npq}}$ とおく. (1) より $\widehat{S}_n$ は $x_k = \frac{k - np}{\sqrt{npq}}$ ($k = 0, 1, \dots, n$) に値を持つ確率 $p_{n,k} = \binom{n}{k} p^k q^{n-k}$ の確率変数であることがわかる. ここで $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

は2項係数を表す. $n \rightarrow \infty$ のとき

$$k = np + x_k \sqrt{npq} \rightarrow \infty,$$

$$n - k = nq - x_k \sqrt{npq} \rightarrow \infty$$

であることを示せ.

(3) Stirling の公式

$$n! = e^{-n} n^n \sqrt{2\pi n} (1 + o(1))$$

を用いて,

$$p_{n,k} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} (1 + o(1)) \quad \dots \textcircled{a}$$

であることを示せ. ただし, 記号 $o(1)$ は $n \rightarrow \infty$ のとき 0 に収束する項を表す.

(4) $|x| \leq \frac{1}{2}$ に対する不等式

$$\left| \log(1+x) - x + \frac{x^2}{2} \right| \leq \frac{2}{3} |x|^3$$

を用いて,

$$\lim_{\substack{n \rightarrow \infty \\ x_k \rightarrow x}} \log \left(\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \right) = -\frac{x^2}{2} \quad \dots \textcircled{b}$$

であることを示せ.

(5) $h := \frac{1}{\sqrt{npq}}$ とおく. 等式

$$\sqrt{\frac{n}{k(n-k)}} = h(1 + o(1)) \quad \dots \textcircled{c}$$

を確かめ, 区間 $\left[-\frac{np}{\sqrt{npq}}, \frac{nq}{\sqrt{npq}}\right]$ を n 個に分割した小区間の長さが h であること, および x_k がその小区間の分点であることに着目し, 等式

$$\mathbb{P}[a \leq \widehat{S}_n \leq b] = \sum_{\substack{a \leq x_k \leq b \\ x_k = \frac{k-np}{\sqrt{npq}}}} p_{n,k}$$

から $\textcircled{*}$ を導け.

解 (1) $S_n := X_1 + \dots + X_n$ とおく. X_i の定義から確率変数 S_n は $0, 1, \dots, n$ に値を取り, 確率分布が

$$\mathbb{P}[S_n = r] = \binom{n}{r} p^r q^{n-r}, \quad r = 0, 1, \dots, n$$

で与えられるから, S_n は2項分布 $B(n, p)$ に従う.

(2) $x_k = \frac{k-np}{\sqrt{npq}} \iff k = np + x_k \sqrt{npq} \dots \textcircled{1}$. x_k の定義と, 考える確率変数 $\widehat{S}_n$ の範囲から, 各 k において $a \leq x_k \leq b$ をみたすので, x_k は有界である. したがって, $\textcircled{1}$ の第2項目も $n \rightarrow \infty$ のとき ∞ に発散する. よって $\textcircled{1}$ より $n \rightarrow \infty$ のとき $k \rightarrow \infty$. 同様にして, $n-k \rightarrow \infty$ ($n \rightarrow \infty$) も分かる.

(3) Stirling の公式 $n! = e^{-n} n^n \sqrt{2\pi n} (1 + o(1))$ を $n!$, $k!$ および $(n-k)!$ に用いて,

$$\begin{aligned} p_{n,k} &= \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \frac{e^{-n} n^n \sqrt{2\pi n} p^k q^{n-k}}{e^{-k} k^k \sqrt{2\pi k} e^{-(n-k)} (n-k)^{(n-k)} \sqrt{2\pi(n-k)}} (1 + o(1)) \\ &= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} (1 + o(1)) \end{aligned}$$

となり, $\textcircled{a}$ を得る.

(4) (2) より $k = np + x_k \sqrt{npq}$ であるから,

$$\begin{aligned} \log \left(\frac{np}{k}\right)^k &= -k \log \left(\frac{k}{np}\right) = -k \log \left(1 + \sqrt{\frac{q}{np}} x_k\right) \\ &= -(np + x_k \sqrt{npq}) \log \left(1 + \sqrt{\frac{q}{np}} x_k\right). \end{aligned}$$

ここで, 不等式 $\left| \log(1+x) - x + \frac{x^2}{2} \right| \leq \frac{2}{3} |x|^3$ を $x = \sqrt{\frac{q}{np}} x_k$ として適用すれば*,

$$\left| \log \left(1 + \sqrt{\frac{q}{np}} x_k\right) - \sqrt{\frac{q}{np}} x_k + \frac{q}{np} \frac{x_k^2}{2} \right| \leq \frac{2}{3} \left| \sqrt{\frac{q}{np}} x_k \right|^3$$

すなわち,

$$\begin{aligned} \left| -\log \left(\frac{np}{k}\right)^k + \underbrace{(np + x_k \sqrt{npq}) \left(-\sqrt{\frac{q}{np}} x_k + \frac{q}{np} \frac{x_k^2}{2}\right)}_{=A} \right| &\leq \frac{2}{3} (np + x_k \sqrt{npq}) \left| \sqrt{\frac{q}{np}} x_k \right|^3 \\ &\rightarrow 0 \quad (n \rightarrow \infty) \quad \dots \textcircled{2}. \end{aligned}$$

同様にして, 上の $\log$ に関する不等式を $x = -\sqrt{\frac{p}{nq}} x_k$ として適用すれば

$$\begin{aligned} \left| -\log \left(\frac{nq}{n-k}\right)^{n-k} + \underbrace{(nq - x_k \sqrt{npq}) \left(\sqrt{\frac{p}{nq}} x_k + \frac{p}{nq} \frac{x_k^2}{2}\right)}_{=B} \right| &\leq \frac{2}{3} |nq - x_k \sqrt{npq}| \left| \sqrt{\frac{p}{nq}} x_k \right|^3 \\ &\rightarrow 0 \quad (n \rightarrow \infty) \quad \dots \textcircled{3}. \end{aligned}$$

ここで, $p+q=1$ を用いると $A+B$ に関して

$$A+B = (np + x_k \sqrt{npq}) \left(-\sqrt{\frac{q}{np}} x_k + \frac{q}{np} \frac{x_k^2}{2}\right) + (nq - x_k \sqrt{npq}) \left(\sqrt{\frac{p}{nq}} x_k + \frac{p}{nq} \frac{x_k^2}{2}\right)$$

* n は十分大きいので, $\left| \sqrt{\frac{q}{np}} x_k \right| \leq \frac{1}{2}$ と考えてよい.

$$\begin{aligned}
 &= -\sqrt{npq}x_k + \frac{q}{2}x_k^2 - qx_k^2 + \frac{q^{3/2}}{2\sqrt{n}\sqrt{p}}x_k^3 + \sqrt{npq}x_k + \frac{p}{2}x_k^2 - px_k^2 + \frac{p^{3/2}}{2\sqrt{n}\sqrt{q}}x_k^3 \\
 &= -\frac{x_k^2}{2} + \frac{1}{2\sqrt{n}} \left(\frac{q^{2/3}}{\sqrt{p}} + \frac{p^{2/3}}{\sqrt{q}} \right) x_k^3.
 \end{aligned}$$

したがって、これと②, ③を合わせて

$$\begin{aligned}
 &\left| \log \left(\left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} \right) + \frac{x_k^2}{2} \right| \\
 &= \left| \log \left(\frac{np}{k} \right)^k + \log \left(\frac{nq}{n-k} \right)^{n-k} - (A+B) - \frac{1}{2\sqrt{n}} \left(\frac{q^{2/3}}{\sqrt{p}} + \frac{p^{2/3}}{\sqrt{q}} \right) x_k^3 \right| \\
 &\leq \left| \log \left(\frac{np}{k} \right)^k - A \right| + \left| \log \left(\frac{nq}{n-k} \right)^{n-k} - B \right| + \frac{1}{2\sqrt{n}} \left(\frac{q^{2/3}}{\sqrt{p}} + \frac{p^{2/3}}{\sqrt{q}} \right) |x_k|^3 \\
 &\rightarrow 0 + 0 + 0 = 0 \quad (n \rightarrow \infty, x_k \rightarrow x),
 \end{aligned}$$

すなわち、これは⑥を意味する。

(5) ⑥より

$$\lim_{\substack{n \rightarrow \infty \\ x_k \rightarrow x}} \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} = e^{-\frac{x^2}{2}} \quad \dots \textcircled{d}.$$

さらに、 $h = 1/\sqrt{npq}$ および $k = np + x_k\sqrt{npq}$, $n - k = nq - x_k\sqrt{npq}$ を用いて、

$$\begin{aligned}
 \sqrt{\frac{n}{k(n-k)}} &= \sqrt{\frac{n}{(np + x_k\sqrt{npq})(nq - x_k\sqrt{npq})}} \\
 &= \frac{1}{\sqrt{npq}} \sqrt{\frac{n}{(x_k + \frac{\sqrt{np}}{q})(\frac{\sqrt{nq}}{p} - x_k)}} \\
 &= \frac{1}{\sqrt{npq}} \sqrt{\frac{1}{\frac{1}{\sqrt{pq}} + \left(-\frac{\sqrt{p}}{q\sqrt{n}} + \frac{\sqrt{q}}{p\sqrt{n}} \right) x_k - \frac{x_k^2}{n}}} \\
 &= h(1 + o(1))
 \end{aligned}$$

となって、③を得る。③, ③より、

$$\begin{aligned}
 p_{n,k} &= \frac{1}{\sqrt{2\pi}} h \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} (1 + o(1)) \\
 &= \frac{1}{\sqrt{2\pi}} \left(\frac{np}{k} \right)^k \left(\frac{nq}{n-k} \right)^{n-k} (1 + o(1))(x_{k+1} - x_k) \quad \dots \textcircled{e}.
 \end{aligned}$$

したがって、④, ⑤を用いて Riemann 和

$$\mathbb{P}[a \leq \widehat{S}_n \leq b] = \sum_{\substack{a \leq x_k \leq b \\ x_k = \frac{k-np}{\sqrt{npq}}}} p_{n,k}$$

において $n \rightarrow \infty$ の極限に移行すれば、所望の等式 $\star$ を得る。 ■

- **1** より

$$\frac{S_n - np}{\sqrt{npq}} = \frac{S_n - \mathbb{E}[S_n]}{\mathbb{V}[S_n]^{1/2}}$$

Check

であるから, Laplace-DeMoivre の定理より $\frac{S_n - \mathbb{E}[S_n]}{\mathbb{V}[S_n]^{1/2}}$ は近似的に標準正規分布 $N(0, 1)$ に従うことがわかる. すなわち, S_n は近似的に $N(\mathbb{E}[S_n], \mathbb{V}[S_n])$ に従う.

- Laplace-DeMoivre の定理を, より一般化したものが**中心極限定理**である.

3 (★★☆)(Laplace-DeMoivre の定理の応用)

X が整数値のみをとる離散確率変数で, $\tilde{X}$ を連続確率変数とする. 離散変数 X の分布を連続変数 $\tilde{X}$ の分布で近似するとき, $\mathbb{P}[a \leq X \leq b]$ を実際に計算するときは, $\mathbb{P}\left[a - \frac{1}{2} \leq \tilde{X} \leq b + \frac{1}{2}\right]$ を計算する方が精度がよくなることが多い. これを**半数補正近似**という.

どの面も等確率に出るサイコロを 100 回独立に振るとき, 6 が出る回数を S_{100} とする. 確率 $\mathbb{P}[7 \leq S_{100} \leq 11]$ を半数補正近似と Laplace-DeMoivre の定理を用いて近似的に何 % か求めよ. ただし, 近似計算にあたり正規分布表を用いよ.

解 S_{100} は 2 項分布 $B(100, \frac{1}{6})$ に従うので, 半数補正近似を用いて, 求める確率 $\mathbb{P}[7 \leq S_{100} \leq 11]$ を $\mathbb{P}\left[\frac{13}{2} \leq S_{100} \leq \frac{23}{2}\right]$ で近似すると考える. ここで, $\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{x^2}{2}} dx$ とおく. このとき, Laplace-DeMoivre の定理より

$$\begin{aligned} \mathbb{P}\left[\frac{13}{2} \leq S_{100} \leq \frac{23}{2}\right] &= \mathbb{P}\left[\frac{\frac{13}{2} - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq \frac{S_{100} - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq \frac{\frac{23}{2} - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right] \\ &\cong \mathbb{P}\left[-0.27 \leq \frac{S_{100} - \frac{100}{6}}{\sqrt{100 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq -0.14\right] \\ &= \Phi(0.27) - \Phi(0.14) \quad (\because \text{対称性}) \\ &\cong 0.106 - 0.056 \quad (\because \text{正規分布表}) \\ &= 0.05 = 5\%. \end{aligned}$$

■