

§13 標本平均, 大数の法則 演習問題1 解答

🔑 問題の難易度の目安【基礎】☆☆☆ 【標準】★★☆ 【発展】★★★

1 (☆☆☆)(Chebyshev の不等式)

X を密度関数 $f(x)$ をもつ連続型確率変数とする. X の平均を $\mu = E(X)$, 分散を $\sigma^2 = V(X)$ とする. このとき $\varepsilon > 0$ に対して

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

であることを示せ. これを ^{チェビシェフ} Chebyshev の不等式という.

解 $(x - \mu)^2 f(x) \geq 0$ に注意して,

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{|x-\mu| \geq \varepsilon} (x - \mu)^2 f(x) dx + \underbrace{\int_{|x-\mu| < \varepsilon} (x - \mu)^2 f(x) dx}_{\geq 0} \\ &\geq \int_{|x-\mu| \geq \varepsilon} (x - \mu)^2 f(x) dx \\ &\geq \varepsilon^2 \int_{|x-\mu| \geq \varepsilon} f(x) dx = \varepsilon^2 P(|X - \mu| \geq \varepsilon). \end{aligned}$$

したがって, $P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$ は成り立つ. ■

Chebyshev の不等式は離散型確率変数に対しても成り立つ:

($\because$) $p_i := P(X = a_i), i = 1, 2, \dots$ とおく.

$$\begin{aligned} \sigma^2 &= \sum_i (a_i - \mu)^2 p_i = \sum_{i: |a_i - \mu| \geq \varepsilon} (a_i - \mu)^2 p_i + \underbrace{\sum_{i: |a_i - \mu| < \varepsilon} (a_i - \mu)^2 p_i}_{\geq 0} \\ &\geq \sum_{i: |a_i - \mu| \geq \varepsilon} (a_i - \mu)^2 p_i \\ &\geq \varepsilon^2 \sum_{i: |a_i - \mu| \geq \varepsilon} p_i \\ &= \varepsilon^2 \sum_{i: |a_i - \mu| \geq \varepsilon} P(|X - \mu| \geq \varepsilon, X = a_i) = \varepsilon^2 P(|X - \mu| \geq \varepsilon) \\ &\iff P(|X - \mu| \geq \varepsilon) \leq \sigma^2 / \varepsilon^2. \end{aligned}$$

Check

2 (★★☆)(大数の法則)

$X_1, \dots, X_n$ を互いに独立で同じ分布に従う確率変数とする (これを無作為標本という). この分布の平均を $\mu = E(X_i)$, 分散を $\sigma^2 = V(X_i)$ とおく. 標本平均

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$$

について, 次の各問いに答えよ.

(1) $\bar{X}$ の平均 $E(\bar{X})$ と分散 $V(\bar{X})$ を求めよ.

(2) 任意の $\varepsilon > 0$ に対し,

$$\lim_{n \rightarrow \infty} P(|\bar{X} - E(\bar{X})| \geq \varepsilon) = 0$$

であることを示せ. これを**大数の法則** (law of large numbers) という.^a

^a正確には**大数の弱法則**とよばれる. これに対して $P\left(\lim_{n \rightarrow \infty} \bar{X} = E(\bar{X})\right) = 1$ を**大数の強法則**という.

解 (1) 標本平均 $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ に対して, 期待値 (平均) の線形性より

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu \quad \cdots \textcircled{1}.$$

また, 各 X_i が独立であるから, 期待値について

$$E((X_i - \mu)(X_j - \mu)) = E(X_i - \mu)E(X_j - \mu), \quad \forall i \neq \forall j \quad \cdots \textcircled{2}.$$

したがって①,②より

$$\begin{aligned} V(\bar{X}) &= E((\bar{X} - \mu)^2) \quad (\because \textcircled{1}) \\ &= E\left(\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right)^2\right) = \frac{1}{n^2} E\left(\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right) \\ &= \frac{1}{n^2} E\left(\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i \neq j} (X_i - \mu)(X_j - \mu)\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n E((X_i - \mu)^2) + \frac{1}{n^2} \sum_{i \neq j} E((X_i - \mu)(X_j - \mu)) \\ &= \frac{1}{n^2} \sum_{i=1}^n E((X_i - \mu)^2) + \frac{1}{n^2} \sum_{i \neq j} \underbrace{E(X_i - \mu)}_{=0} \underbrace{E(X_j - \mu)}_{=0} \quad (\because \textcircled{2}) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \end{aligned}$$

別解 分散の性質を用いることによって、次のように簡単に $V(\bar{X})$ を求めることができる：

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} V(X_1 + \cdots + X_n) \\ &= \frac{1}{n^2} (V(X_1) + \cdots + V(X_n)) \\ &= \frac{\sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

(2) **1** の Chebyshev の不等式 および (1) の結果より、任意の $\varepsilon > 0$ に対して

$$0 \leq P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \frac{\sigma^2}{n^2} \rightarrow 0 \quad (\text{as } n \rightarrow \infty).$$

■

3 (★☆☆)(大数の法則の応用 1)

すべての目が等確率で出るサイコロを 300 回投げるとき、1 の目は概ね 50 回程度出ると考えて良いか。大数の法則に基づいて判定せよ。

解 確率変数 X_i を

$$X_i := \begin{cases} 1 & \dots i \text{ 回目に 1 が出る} \\ 0 & \dots i \text{ 回目に 1 が出ない} \end{cases}$$

と定めると、 $\{X_i\}_{i=1}^{300}$ は互いに独立かつ同じ分布に従い、各 $1 \leq i \leq 300$ に対して

$$\begin{aligned} E(X_i) &= 1 \cdot \frac{1}{6} + 0 \cdot \frac{5}{6} = \frac{1}{6}, \\ V(X_i) &= \left(1 - \frac{1}{6}\right)^2 \cdot \frac{1}{6} + \left(0 - \frac{1}{6}\right)^2 \cdot \frac{5}{6} = \frac{5}{36} \end{aligned}$$

をみたら、ゆえに、標本平均を $\bar{X} = \frac{1}{300} \sum_{i=1}^{300} X_i$ とすると、 $E(\bar{X}) = \frac{1}{6}$ であるから、大数の法則により任意の $\varepsilon > 0$ に対し、

$$\begin{aligned} &P(|300 \text{ 回中 1 の目が出る回数} - 50| > \varepsilon) \\ &= P\left(\left|\sum_{i=1}^n X_i - 50\right| > \varepsilon\right) \\ &= P\left(\left|\bar{X} - \frac{1}{6}\right| > \frac{\varepsilon}{300}\right) \approx 0. \quad (\because \text{大数の法則}) \end{aligned}$$

したがって、1 の目は概ね 50 回程度出るといえる。

■

4 (★★☆)(大数の法則の応用 2)

1 枚のコインを n 回投げるとき、表が出る回数を X とする. n が大きくなるとき, $\frac{X}{n}$ はいくらくに近づくか.

解 確率変数 X_i を

$$X_i := \begin{cases} 1 & \dots i \text{ 回目} \text{ が表} \\ 0 & \dots i \text{ 回目} \text{ が裏} \end{cases}$$

と定めると, $\{X_i\}_{i=1}^n$ は互いに独立かつ同じ分布に従い, 各 i に対して

$$E(X_i) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2},$$

$$V(X_i) = \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

をみtas. ゆえに, 標本平均 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X}{n}$ について, **2** の結果より

$$E(\bar{X}) = \frac{1}{2}, \quad V(\bar{X}) = \frac{1}{4n}.$$

よって, 大数の法則より任意の $\varepsilon > 0$ に対して

$$P\left(\left|\bar{X} - \frac{1}{2}\right| > \varepsilon\right) = P\left(\left|\frac{X}{n} - \frac{1}{2}\right| > \varepsilon\right) \rightarrow 0 \quad (\text{as } n \rightarrow \infty).$$

ゆえに $\frac{X}{n}$ は $\frac{1}{2}$ に近づく. ■

5 (★★★)(中心極限定理と大数の法則)

以下の問いに答えよ.

(1) $z \in \mathbb{R}$ に対して, $\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}x^2} dx$ とおく. このとき, $\Phi(z) + \Phi(-z) = 1$ であることを示せ.

(2) $X_1, \dots, X_n, \dots$ を互いに独立で同じ確率分布に従う確率変数列とし,

$$\mu = E(X_i), \quad \sigma^2 = V(X_i), \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

とする. このとき中心極限定理

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq b\right) = \Phi(b) - \Phi(a) \quad (a \leq b)$$

を用いて, 大数の法則を導け.

解 (1) $\Phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-\frac{1}{2}x^2} dx$ において, $x = -y$ の変数変換を施すと

$$\Phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^z e^{-\frac{1}{2}y^2} (-dy) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-\frac{1}{2}y^2} dy = \lim_{R' \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_z^{R'} e^{-\frac{1}{2}x^2} dx$$

となる. したがって, これより

$$\begin{aligned} \Phi(z) + \Phi(-z) &= \lim_{R \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-R}^z e^{-\frac{1}{2}x^2} dx + \lim_{R' \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_z^{R'} e^{-\frac{1}{2}x^2} dx \\ &= \lim_{\substack{R \rightarrow \infty \\ R' \rightarrow \infty}} \frac{1}{\sqrt{2\pi}} \int_{-R}^{R'} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = 1. \end{aligned}$$

(2) $a = -c/\sigma$, $b = c/\sigma$ とする ($c > 0$ はあとで選ぶ定数). 中心極限定理より

$$\lim_{n \rightarrow \infty} P\left(-\frac{c}{\sqrt{n}} \leq \bar{X}_n - \mu \leq \frac{c}{\sqrt{n}}\right) = \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(-\frac{c}{\sigma}\right) \quad \dots \textcircled{1}$$

が成り立つ. $\varepsilon > 0$ および $\nu > 0$ を任意にとり, 定数 $c > 0$ を

$$\Phi\left(-\frac{c}{\sigma}\right) \leq \frac{\nu}{3} \quad \dots \textcircled{2}$$

をみたすようにとる. 実際, $\Phi(z) \rightarrow 0$ ($z \rightarrow -\infty$) であるから, このように定数 $c > 0$ を選ぶことができる. さらに, 自然数 N を $c/\sqrt{N} \leq \varepsilon \dots \textcircled{3}$ をみたすように十分大きくとる. このとき $\forall n \geq N$ に対して $\textcircled{2}, \textcircled{3}$ より

$$\begin{aligned} P\left(|\bar{X}_n - \mu| \leq \frac{c}{\sqrt{n}}\right) &\leq P\left(|\bar{X}_n - \mu| \leq \frac{c}{\sqrt{N}}\right) \\ &\leq P(|\bar{X}_n - \mu| \leq \varepsilon) \quad \dots \textcircled{4}. \end{aligned}$$

また, $\textcircled{1}$ より $\forall n \geq N$ に対して,

$$\left|P\left(|\bar{X}_n - \mu| \leq \frac{c}{\sqrt{n}}\right) - \Phi\left(\frac{c}{\sigma}\right) + \Phi\left(-\frac{c}{\sigma}\right)\right| < \frac{\nu}{3}$$

が成り立つので, これから

$$P\left(|\bar{X}_n - \mu| \leq \frac{c}{\sqrt{n}}\right) > \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(-\frac{c}{\sigma}\right) - \frac{\nu}{3} \quad \dots \textcircled{5}.$$

また (1) より $\Phi\left(\frac{c}{\sigma}\right) + \Phi\left(-\frac{c}{\sigma}\right) = 1 \dots \textcircled{6}$ であるから, $\textcircled{4}, \textcircled{5}, \textcircled{6}$ より $\forall \varepsilon > 0, \forall \nu > 0, \forall n \geq N$ に対し,

$$P(|\bar{X}_n - \mu| \leq \varepsilon) > \Phi\left(\frac{c}{\sigma}\right) - \Phi\left(-\frac{c}{\sigma}\right) - \frac{\nu}{3}$$

$$= 1 - 2\Phi\left(-\frac{c}{\sigma}\right) - \frac{\nu}{3} \quad (\because \textcircled{6})$$

$$> 1 - \frac{2\nu}{3} - \frac{\nu}{3} = 1 - \nu$$

i.e.,

$$P(|\bar{X}_n - \mu| > \varepsilon) = 1 - P(|\bar{X}_n - \mu| \leq \varepsilon) < \nu$$

に逢着する. これは $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$, 大数の法則に他ならない. ■