

確率・統計学 (春学期) 期末試験

【問 1】、【問 2】、【問 3】をそれぞれ別々の解答用紙 (合計 3 枚) に解答せよ。
 解答用紙に学籍番号と名前の記入を忘れないようにすること。
 単純ミスがあっても途中点を与えられるように、考え方の筋道が分かるように解答すること。
 必要に応じて、裏面の表を用いてよい。

【問 1】 (30) 確率変数 X の確率密度を

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in (-\infty, \infty)$$

と表したとき、以下の間に答えよ。

- (1) 特性関数 $F(k) = E[e^{ikX}]$ を求めよ。
- (2) 特性関数 $F(k)$ を用いて、 $f(x)$ のもとでの平均 μ と分散 σ^2 を求めよ。
- (3) 新たな確率変数 $Y = X^2$ が従う確率分布 (確率密度) $g(y)$ を求めよ。
- (4) (3) で求めた $g(y)$ を用いて、平均 $E[Y]$ を求めよ。

【問 2】 (40) 以下の間に答えよ。

- (1) あるサイコロを 10 回振って出た目の平均を求めたところ 3.9 であった。一方、標本平均 \bar{X} と母平均 μ の関係は $E[\bar{X}] = \mu$ で与えられる。 $\mu = 3.5$ であるから、この値は、うえの値 3.9 とかなり異なる。その理由を述べよ。
- (2) ある部品の長さを測るときの誤差が、平均 0、標準偏差 0.2 mm の正規分布に従うとする。この測定器による 1 回の測定誤差が 0.5 mm 以上となる確率を求めよ。
- (3) 母平均が μ 、母分散が σ^2 の母集団から大きさ n の標本を抽出した。標本平均 \bar{X} の分散の期待値と標本分散 S^2 の期待値を n と σ^2 を用いて表せ。
- (4) 母集団の確率分布が平均 θ のポアソン分布に従うとする。母集団から無作為に大きさ 4 の標本を抽出したところ、その値が x_1, x_2, x_3, x_4 であったとする。母平均の最尤推定量を求めよ。
- (5) 母分散 $\sigma^2 = 9$ の正規母集団から大きさ n の標本を抽出して、母平均 μ の信頼水準 95% の信頼区間を求めたい。その区間の幅を 1 以下にするには、標本の大きさ n をいくつ以上にしなければいけないか。

【問 3】 (30) 中心極限定理に関する以下の間に答えよ。

- (1) 中心極限定理を述べよ。
- (2) 以下の 2 項分布

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

に関して、つぎの定理が成立する。ここで、 $q = 1 - p$ である。

定理 1 n が十分に大きければ、2 項分布は平均 np 、分散 npq の正規分布に近づく。

この定理を用いて、2 項分布を与える確率過程を (講義で説明した) 別の表現で明示したうえで、(1) の中心極限定理を証明せよ。

【問1】

(1) 特性関数 $F(k)$ は

$$\begin{aligned} F(k) &= E[e^{ikX}] = \int_{-\infty}^{\infty} dx f(x)e^{ikx} = \int_{-\infty}^0 dx \frac{1}{2}e^x e^{ikx} + \int_0^{\infty} dx \frac{1}{2}e^{-x} e^{ikx} \\ &= \frac{1}{1 - (ik)^2} \end{aligned}$$

(2) $k = 0$ 近傍でテイラー展開すれば、

$$F(k) = 1 + (ik)^2 + (ik)^4 + \dots$$

となるから、平均は、

$$\mu = E[X] = \left. \frac{\partial F(k)}{\partial(ik)} \right|_{k=0} = 0$$

分散は、

$$\sigma^2 = E[X^2] - \mu^2 = E[X^2] = \left. \frac{\partial^2 F(k)}{\partial(ik)^2} \right|_{k=0} = 2$$

(3) 確率変数の変換だから、

$$y = x^2 \Rightarrow x = \pm\sqrt{y}$$

より、

$$\begin{aligned} g(y) &= f(\sqrt{y}) \left| \frac{dx}{dy} \right|_{x=\sqrt{y}} + f(-\sqrt{y}) \left| \frac{dx}{dy} \right|_{x=-\sqrt{y}} = \frac{1}{2}e^{-\sqrt{y}} \frac{1}{2\sqrt{y}} + \frac{1}{2}e^{-\sqrt{y}} \frac{1}{2\sqrt{y}} \\ &= \frac{e^{-\sqrt{y}}}{2\sqrt{y}} \end{aligned}$$

よって、

$$g(y) = \begin{cases} \frac{e^{-\sqrt{y}}}{2\sqrt{y}} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

(4)

$$E[Y] = \int_{-\infty}^{\infty} dy g(y)y = \int_0^{\infty} dy y \frac{e^{-\sqrt{y}}}{2\sqrt{y}} = 2 \quad (= E[X^2])$$

となって、(2)の結果と一致する。

【問2】

- (1) 母集団のもとでの標本平均 \bar{X} の期待値とは、サイコロを 10 回振ってその平均を求める操作を無限回繰り返したときに得られる \bar{X} の平均値を意味し、それは母平均となる。従って、ある特定の標本の平均値が母平均と異なるのは自然である。一方、標本の大きさが大きい場合や、平均を取る操作を数多く繰り返せば、標本平均の（平均）値は母平均の値に近づく。
- (2) 測定器の誤差 x mm の従う確率分布は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

の正規分布で与えられる。ここで、 $\sigma = 0.2$ mm である。
求める確率は、

$$P[|X| \geq 0.5] = 2 \int_{0.5}^{\infty} dx f(x) = 2 \int_{2.5}^{\infty} dz \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = 2 \times 0.00621 = 0.01242$$

ここで、標準化変換 $z = x/\sigma = x/0.2$ を行ったうえで、数表を用いた。

- (3) 標本平均 \bar{X} の分散の期待値は

$$V[\bar{X}] = E[(\bar{X} - \mu)^2] = \frac{1}{n^2} E[\{(X_1 - \mu) + \dots + (X_n - \mu)\}^2] = \frac{\sigma^2}{n}$$

標本分散の期待値は、

$$\begin{aligned} E[S^2] &= E\left[\frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2\right] = \frac{1}{n} E\left[\sum_{k=1}^n \left\{(X_k - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_k - \mu)(\bar{X} - \mu)\right\}\right] \\ &= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \end{aligned}$$

- (4) 標本値 x_1, x_2, x_3, x_4 が選ばれる確率は、

$$L(\theta) = e^{-\theta} \frac{\theta^{x_1}}{x_1!} \cdot e^{-\theta} \frac{\theta^{x_2}}{x_2!} \cdot e^{-\theta} \frac{\theta^{x_3}}{x_3!} \cdot e^{-\theta} \frac{\theta^{x_4}}{x_4!} = e^{-4\theta} \frac{\theta^{x_1+x_2+x_3+x_4}}{x_1!x_2!x_3!x_4!}$$

このとき、この確率が最大（極値）を持つとすれば、

$$\frac{\partial L(\theta)}{\partial \theta} = e^{-4\theta} \frac{\theta^{x_1+x_2+x_3+x_4}}{x_1!x_2!x_3!x_4!} \left(-4 + \frac{x_1+x_2+x_3+x_4}{\theta}\right) = 0$$

よって、母平均の最尤推定量は

$$\theta = \frac{x_1 + x_2 + x_3 + x_4}{4}$$

で与えられ、標本平均と一致する。

- (5) 標準正規分布のもとでの信頼水準 99% の信頼区間は、数表から $[-1.96, 1.96]$ である。よって、信頼区間は、

$$-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96 \quad \Rightarrow \quad \bar{X} - 1.96 \frac{3}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{3}{\sqrt{n}}$$

この区間が 1 以下であるためには、

$$2 \times 1.96 \frac{3}{\sqrt{n}} \leq 1 \quad \Rightarrow \quad n \geq 138.3$$

$n \in \mathbb{N}$ だから、 $n \geq 139$ 。

【問3】

- (1) n 個の独立な確率変数を X_1, X_2, \dots, X_n とする。このとき、 $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ で定義される確率変数 \bar{X} の確率分布は、 n が十分に大きいとき、正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に近づく。ここで、 μ と σ^2 はそれぞれ X_i ($i = 1, \dots, n$) の平均と分散である。
- (2) 2項分布を与える確率過程として、以下のような過程を考える。事象 A が起こる確率を p とし、このとき確率変数は $X = 1$ を取るとする。事象 B の起こる確率を $q = 1 - p$ とし、このときは $X = 0$ とする。新たに確率変数 $Y = X_1 + X_2 + \dots + X_n$ で定義したとき、 Y は n 回繰り返したときに事象 A の起こる回数を表す。従って、確率変数 Y に対する確率分布は、以下の2項分布で与えられる。

$$f(y) = \frac{n!}{y!(n-y)!} p^y q^{n-y}$$

問題文の定理から、 n が十分に大きければ、 $f(y)$ は平均 np 、分散 npq の正規分布で近似できるから、

$$f(y) = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{(y-np)^2}{2npq}}$$

$\bar{X} = Y/n$ とすれば、確率変数 \bar{X} に対する確率分布 $h(\bar{x})$ は、

$$h(\bar{x}) = f(y) \left| \frac{dy}{d\bar{x}} \right| = \frac{n}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{(n\bar{x}-np)^2}{2npq}} = \frac{1}{\sqrt{2\pi}\sqrt{\frac{pq}{n}}} e^{-\frac{(\bar{x}-p)^2}{2\frac{pq}{n}}}$$

これは、平均 p 、分散 pq/n の正規分布である。

一方、うえで述べた確率過程において、確率変数 X に対する平均 μ は、

$$\mu = E[X] = 1p + 0q = p$$

となる。分散 σ^2 は、 $E[X^2] = 1^2p + 0^2q = p$ から、

$$\sigma^2 = E[X^2] - E[X]^2 = p - p^2 = p(1-p) = pq$$

つまり、 n が十分に大きければ、

$$h(\bar{x}) = \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{n}}} e^{-\frac{(\bar{x}-\mu)^2}{2\frac{\sigma^2}{n}}}$$

となって、中心極限定理の意味することと等価になる。