

確率・統計学 (春学期) 期末試験

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, 以下の事実を用いてよい.

標準正規分布 $N(0, 1)$ において, $z = 1.96$ のとき $P[|Z| \leq z] = 0.95$ となる.

【問 1 (30)】 コインを投げたとき表の出る確率を p とし, n 回を投げた場合を考える.

- (1) 表が k 回出る確率を求めよ. ここで, k は $0 \leq k \leq n$ の整数である.
- (2) (1) の結果を用いて, 表が 1 回も出ない確率を求めよ.
- (3) 表が 1 回も出ない確率をポアソン分布を用いて求めよ.
- (4) 確率 p が十分に小さいとき, (2) の結果が (3) の結果と一致することを理由とともに示せ.

【問 2 (45)】 区間 $0 \leq x \leq 1$ 上に一様分布する実数を確率変数 (乱数) X とする. 以下の問いに答えよ.

- (1) X に対する確率密度 $f(x)$ および特性関数 $F(k)$ を求めよ. ここで, 特性関数 $F(k)$ は

$$F(k) = E[e^{ikX}]$$

で定義される.

- (2) X の平均 μ および分散 σ^2 を求めよ.
- (3) 生成した n 個の乱数を $X_1, X_2, \dots, X_n$ とする. $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ となる (多変数) 確率密度 $f_n(x_1, x_2, \dots, x_n)$ を $f(x)$ を用いて書き表せ.
- (4) 生成した n 個の乱数の平均を $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ としたとき, $E[\bar{X}]$ および $E[(\bar{X} - \mu)^2]$ を求めよ. ここで, μ は (2) で求めた X の平均である.
- (5) n が十分に大きいとき, $\bar{X}$ の確率密度 $h(\bar{x})$ を n, μ, σ を用いて書き表せ.
- (6) **任意 (Bonus Problem: +10)**
生成した乱数の平均 $\bar{X}$ が 95 % の確率で $0.46 < \bar{X} < 0.54$ にあるためには, 少なくとも何個の乱数を生成する必要があるか.

【問 3 (25)】 ある工場のペットボトルに水を詰める機械についての以下の問いに答えよ.

- (1) ペットボトルを 100 本無作為に取って水の量を測ったら, ペットボトル 1 個あたりに含まれる平均水量は 1000 ml, 標準偏差は 10 ml であったととする. 工場から出荷されるペットボトルの平均水量 (母平均) に対する 95 % 信頼区間を求めよ. 標本の大きさは十分に大きいと仮定してよい.
- (2) 工場の機械を調整して, 平均 1000 ml, 標準偏差 10 ml の正規分布となるように設定したとする. ペットボトルを 25 本無作為に取って水の量を測ったら, その (標本) 平均が 1006 ml であった. 機械は正しく調整されたといえるか, 帰無仮説および対立仮説を明示したうえで有意水準 5 % で検定せよ.

【問1】

- (1) 二項分布で与えられるから,

$$P[X = k] = {}_n C_k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

- (2) $k = 0$ として, $P[X = 0] = (1-p)^n$.

- (3) 平均は $\mu = np$ で与えられるから, k 回表が出る確率は

$$P[X = k] = e^{-\mu} \frac{\mu^k}{k!} = e^{-np} \frac{(np)^k}{k!}$$

$$k = 0 \text{ として, } P[X = 0] = e^{-np}.$$

- (4) p が小さいとき, $p = 0$ 近傍で (2) の式をテイラー展開すれば,

$$(1-p)^n = 1 - np + \frac{1}{2}n(n-1)p^2 - \dots \approx 1 - np$$

- (3) の式を同様に展開すれば,

$$e^{-np} = 1 - np + \frac{1}{2!}(np)^2 - \dots \approx 1 - np$$

$p \ll 1$ のとき, 二項分布はポアソン分布に移行するため両者は一致する.

【問2】

- (1) 確率密度は $x \in [0, 1]$ で一様となるから, $f(x) = c$ と書ける. ここで, c は定数. 正規化条件から, $c = 1$ となるから,

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x < 0, x > 1 \end{cases}$$

特性関数 $F(k)$ は

$$F(k) = E[e^{ikX}] = \int_{-\infty}^{\infty} dx f(x) e^{ikx} = \int_0^1 dx e^{ikx} = \frac{e^{ik} - 1}{ik}$$

- (2)

$$F(k) = 1 + \frac{ik}{2!} + \frac{(ik)^2}{3!} + \dots$$

となるから,

$$\mu = \left. \frac{\partial F(k)}{\partial (ik)} \right|_{k=0} = \frac{1}{2}, \quad \sigma^2 = E[X^2] - \mu^2 = \left. \frac{\partial^2 F(k)}{\partial (ik)^2} \right|_{k=0} - \left(\frac{1}{2} \right)^2 = \frac{1}{12}$$

- (3) 乱数の発生は独立過程だから,

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n) = \begin{cases} 1 & 0 \leq x_i \leq 1 \quad \forall i = 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

(4)

$$\begin{aligned} E[\bar{X}] &= \int_{-\infty}^{\infty} dx_1 dx_2 \cdots dx_n f_m(x_1, x_2, \dots, x_n) \bar{x} \\ &= \int_{-\infty}^{\infty} dx_1 dx_2 \cdots dx_n f(x_1) f(x_2) \cdots f(x_n) \frac{x_1 + x_2 + \cdots + x_n}{n} \\ &= \frac{1}{n} \left\{ \int_{-\infty}^{\infty} dx_1 f(x_1) x_1 + \int_{-\infty}^{\infty} dx_2 f(x_2) x_2 + \cdots + \int_{-\infty}^{\infty} dx_n f(x_n) x_n \right\} \\ &= \frac{1}{n} n E[X] = \mu = \frac{1}{2} \end{aligned}$$

同様にして,

$$E[(\bar{X} - \mu)^2] = \frac{1}{n^2} E[\{(X_1 - \mu) + \cdots + (X_n - \mu)\}^2] = \frac{1}{n^2} n [(X - \mu)^2] = \frac{\sigma^2}{n} = \frac{1}{12n}$$

(5) n が十分に大きいときは中心極限定理により, $\bar{X}$ の確率分布 $h(\bar{x})$ は正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に近づく. よって,

$$h(\bar{x}) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{n}}} \exp \left[-\frac{(\bar{x} - \mu)^2}{2 \frac{\sigma^2}{n}} \right]$$

(6) 任意 (Bonus Problem: +10)
標準化変換は

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}}$$

となるから,

$$-1.96 \leq Z \leq 1.96 \quad \Rightarrow \quad -1.96 \leq \frac{\bar{X} - \frac{1}{2}}{\frac{1}{\sqrt{12n}}} \leq 1.96$$

よって,

$$\frac{1.96}{\sqrt{12n}} < 0.04$$

となればよい. $n > 200.1$ となるから, 201 個以上の乱数の生成が必要.

【問3】

(1) 標本の大きさは $n = 100$ で十分に大きいから, 標本平均と標本標準偏差はそれぞれ $\bar{X} = 1000$, $S = 10 \approx \sigma$ とみなせる. ペットボトルの平均水量 $\bar{X}$ に対する標準化変換は

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z = \frac{1000 - \mu}{1}$$

となる. よって, 95 % 信頼区間は

$$-1.96 \leq Z = \frac{1000 - \mu}{1} \leq 1.96 \quad \Rightarrow \quad 998 \leq \mu \leq 1002$$

(2) 機械が正しく設定されているときの平均を帰無仮説に選ぶ. よって,

$$H_0: \mu = 1000 \text{ ml}$$

平均からのずれが問題となるので, 両側検定である. よって, 対立仮説は

$$H_1: \mu \neq 1000 \text{ ml}$$

標本の大きさは $n = 25$ で, 母平均と母標準偏差はそれぞれ $\mu = 1000$, $\sigma = 10$ だから, 有意水準 5 % の採択域は

$$-1.96 \leq Z = \frac{\bar{X} - 1000}{2} \leq 1.96 \quad \Rightarrow \quad 996 \leq \bar{X} \leq 1004$$

$\bar{X} = 1006 \text{ ml}$ は棄却域に入るから, 有意水準 5 % で帰無仮説は棄却され, 対立仮説が採択される. つまり, 機械は正しく設定されていないといえる.