今野　(2010 年 01 月 05 日)

情報統計学の試験問題（試験時間は 85 分）

解答上の注意

特別な指示がある場合を除き，答えが合っているかどうかよりも解答の途中過程の論理的展開を重視して採点をします．以下の点に留意して解答を作成すること．

(1) 解答の途中過程は丁寧に記述すること．

(2) 設問中で証明することを求めている場合をのぞき，講義で述べたことは証明なしに用いてよい．しかし，なにをどのように用いたかを可能な限り明示すること．

(3) 各自が理解していることを採点者にわかるように解答を作成することを心がけること．

(4) 等号の使い方に注意すること．

(5) どの問題を解答しているかが明示されていれば，解答は問題番号順でなくともよい．

問題 1　$\theta > 0$ とする．連続型確率変数 X は確率密度関数

$$f_X(x|\,\theta) = \begin{cases} \theta x^{\theta-1}, & (0 < x < 1); \\ 0 & (\text{その他}) \end{cases}$$

からの大きさ 1 のランダム標本とする．

(1) 連続型確率変数 X の期待値 $\mathbb{E}[X]$ を求めよ．

(2) 連続型確率変数 X の分布関数 $F_X(x) = \mathbb{P}(X \le x)\,(x \in \mathbb{R})$ を求めよ．

(3) $X = x\,(0 < x < 1)$ を観測したのときの尤度関数を述べよ (答えのみでよい).

(4) θ の最尤推定値を求めよ．

配点 (1) 10 点 (2) 10 点 (3) 10 点 (4) 10 点 合計 40 点

Solutions.

(1)　期待値の定義に従えば，

$$\mathbb{E}[X] \;=\; \int_{-\infty}^{\infty} x f_X(x|\,\theta) = \theta \int_0^x x^\theta\,dx = \theta\left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

(2) $x \le 0$ のとき，$f_X(x|\,\theta) = 0$ なので，

$$F_X(x) = \int_{-\infty}^{x} f_X(t|\,\theta)\,dt = 0.$$

$0 < x < 1$ のとき，

$$F_X(x) = \int_{-\infty}^{x} f_X(t|\,\theta)\,dt = \int_0^x \theta t^{\theta-1}\,dt = [t^\theta]_0^x = x^\theta$$

$x \geq 1$ のとき，$f_X(x|\,\theta) = 0$ なので，

$$F_X(x) = \int_{-\infty}^{x} f_X(t|\,\theta)\,dt = F_X(1) = 1$$

よって，

$$F_X(x|\,\theta) = \begin{cases} 0 & (x \leq 0); \\ x^\theta & (0 < x < 1); \\ 1 & (x \geq 1). \end{cases}$$

(3) $L(\theta\,|\,x) = \theta x^{\theta-1}\mathbb{1}_{(0,\,1)}(x)$．ただし，$\mathbb{1}_{(0,\,1)}(x) = \begin{cases} 1 & (0 < x < 1); \\ 0 & (\text{その他}) \end{cases}$

(4) $\theta > 0$ と $0 < x < 1$ であるので，対数尤度関数は，

$$\ell(\theta) = \log L(\theta\,|\,x) = \log\theta + (\theta - 1)\log x$$

となる．したがって，尤度方程式は，

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{1}{\theta} + \log x = 0$$

よって，最尤推定値は，

$$\hat{\theta}(x) = -\frac{1}{\log x}.$$

$\boxed{\text{問題 2}}$　連続型確率変数列 $X_1,\, X_2,\, \ldots,\, X_{16}$ は正規分布 $N(\mu,\, 3^2)$ からのランダム標本とする．ただし，$-\infty < \mu < \infty$ とする．
(1) 標本平均 $\bar{X}_{16} = (1/16)(X_1 + \cdots + X_{16})$ の期待値と分散を計算することにより，$\bar{X}_{16}$ の分布を述べよ．
(2) $a > 0,\, b$ を定数とし，$Z = a\bar{X}_{16} + b$ としたとき，Z の分布が標準正規分布になるように定数 $a,\, b$ を定めよ．
(3) 信頼係数 90% の μ の信頼区間を構成せよ．
$\boxed{\text{配点}}$ (1) 5 + 5 点 (2) 10 点 (3) 10 点　合計 30 点

$\boxed{\textbf{Solutions}}$
(1) $\bar{X}_{16}$ の期待値と分散は以下になる：

$$\begin{aligned} E[\bar{X}_{16}] &= (1/16)\{E[X_1] + \cdots + E[X_{16}]\} = (1/16) \times 16\mu = \mu. \\ VAR[\bar{X}_{16}] &= (1/16)^2 = VAR[X_1 + \cdots X_{16}] \\ &= (1/16)^2\{VAR[X_1] + \cdots + VAR[X_{16}]\} = (1/16)^2 \times 16 \times 9 = 9/16 \end{aligned}$$

したがって，正規分布に従う確率変数の一次結合の分布は正規分布なので，$\bar{X}_{16}$ の分布は $N(\mu,\, 9/16)$ となる．
(2) Z は標準正規分布に従うとき，$E[Z] = 0,\, VAR[Z] = 1$ となるので，

$$\begin{aligned} 0 &= E[Z] = aE[\bar{X}_{16}] + b = a\mu + b \\ 1 &= VAR[Z] = a^2 VAR[\bar{X}_{16}] = a^2(9/16) \end{aligned}$$

よって，

$$\begin{cases} a\mu + b = 0 \\ a^2(9/16) = 1 \end{cases}$$

より，$a = 4/3, b = -(4/3)\mu$. よって，

$$Z = \frac{4}{3}(\bar{X}_{16} - \mu)$$

(3) $\{-1.65 \le Z \le 1.65\} = \{-1.65 \le Z\} \cap \{Z \le 1.65\}$ から

$$\{-1.65 \le Z \le 1.65\}^c = \{-1.65 \le Z\}^c \cup \{Z \le 1.65\}^c = \{-1.65 > Z\} \cup \{Z > 1.65\}$$

$\{-1.65 > Z\} \cap \{Z > 1.65\} = \emptyset$ なので，

$$\begin{aligned} P(-1.65 \le Z \le 1.65) &= 1 - P[\{-1.65 \le Z \le 1.65\}^c] \\ &= 1 - \{P[\{-1.65 > Z\} \cup \{Z > 1.65\}]\} \\ &= 1 - \{P(-1.65 > Z) + P(Z > 1.65)\} \end{aligned}$$

正規分布の確率密度関数は偶関数であることより，

$$P(-1.65 > Z) = P(Z > 1.65) = 0.05$$

なので，

$$P(-1.65 \le Z \le 1.65) = 1 - 0.05 - 0.05 = 0.9$$

ここで，$\frac{4}{3}(\bar{X}_{16} - \mu)$ は標準正規分布に従うことを用いると

$$P\left(-1.65 \le \frac{4}{3}(\bar{X}_{16} - \mu) \le 1.65\right) = 0.9$$

さらに，

$$\begin{aligned} -1.65 \le \frac{4}{3}(\bar{X}_{16} - \mu) \le 1.65 &\iff -1.65 \le \frac{4}{3}(\bar{X}_{16} - \mu) \quad \text{かつ} \quad \frac{4}{3}(\bar{X}_{16} - \mu) \le 1.65 \\ &\iff -1.65 \times \frac{3}{4} \le \bar{X}_{16} - \mu \quad \text{かつ} \quad \bar{X}_{16} - \mu \le 1.65 \times \frac{3}{4} \\ &\iff \mu - 1.65 \times \frac{3}{4} \le \bar{X}_{16} \quad \text{かつ} \quad \bar{X}_{16} \le \mu + 1.65 \times \frac{3}{4} \\ &\iff \mu \le \bar{X}_{16} + 1.65 \times \frac{3}{4} \quad \text{かつ} \quad \bar{X}_{16} - 1.65 \times \frac{3}{4} \le \mu \\ &\iff \bar{X}_{16} - 1.65 \times \frac{3}{4} \le \mu \le \bar{X}_{16} + 1.65 \times \frac{3}{4} \end{aligned}$$

なので，

$$0.9 = P\left(-1.65 \le \frac{4}{3}(\bar{X}_{16} - \mu) \le 1.65\right) = P\left(\bar{X}_{16} - 1.65 \times \frac{3}{4} \le \mu \le \bar{X}_{16} + 1.65 \times \frac{3}{4}\right)$$

よって，信頼係数 90% の μ の信頼区間は

$$\left[\bar{X}_{16} - 1.65 \times \frac{3}{4}, \bar{X}_{16} + 1.65 \times \frac{3}{4}\right]$$

$\boxed{\text{問題 3}}$ 連続型確率変数 X は正規母集団 $N(\mu, 2^2)$ からの標本の大きさ 1 のランダム標本とする．次の仮説検定問題を考える：

$$\text{帰無仮説} \quad H_0: \mu = 0 \quad \text{vs. 対立仮説} \quad H_1: \mu > 0$$

このとき，以下の問いに答えよ．

(1) H_0 と H_1 は単純仮説か複合仮説かを答えよ.

(2) H_0 の棄却域として,

$$C(t) = \{x;\, x > t\}$$

としたとき,有意水準 0.1 の検定になるように t をひとつ定めよ.

(3) $\mu = \mu_0\,(\mu_0 > 0)$ のとき,前問で求めた有意水準 0.1 の棄却域 $C(t)$ の検出力を μ_0 と関数 $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) e^{-t^2/2}\, dt$ を用いて表せ.ただし,(2) が解答できない場合には,t を用いてよい.

配点 (1) 5 + 5 点 (2) 10 点 (3) 10 点　合計 30 点

Solutions

(1) H_0 は単純,H_1 は複合.

(2) $\mu = 0$ のとき,$X/2$ は標準正規分布 $N(0, 1)$ に従うので,

$$\begin{aligned}
\mathbb{P}_{\mu=0}(X \in C(t)) &= \mathbb{P}_{\mu=0}\left(\frac{X}{2} > \frac{t}{2}\right) = 1 - \mathbb{P}\left(Z \leq \frac{t}{2}\right) = 0.10 \\
&\Longrightarrow \mathbb{P}\left(Z \leq \frac{t}{2}\right) = 0.90 \Longrightarrow \frac{t}{2} = 1.68.
\end{aligned}$$

(3) $\mu_0 > 0$ のとき,$(X - \mu_0)/2$ が標準正規分布 $N(0, 1)$ に従うことから,

$$\begin{aligned}
\mathbb{P}_{\mu=\mu_0}(X \in C(t)) &= \mathbb{P}_{\mu=\mu_0}(X > t) = \mathbb{P}_{\mu=\mu_0}(\frac{X-\mu_0}{2} > \frac{t-\mu_0}{2}) = 1 - \mathbb{P}\left(Z \leq \frac{t-\mu_0}{2}\right) \\
&= 1 - \Phi\left(\frac{t-\mu_0}{2}\right) = 1 - \Phi\left(\frac{3.36-\mu_0}{2}\right)
\end{aligned}$$

ヒント 以下のことは用いてよい:

- 確率変数 X は確率関数または確率密度関数　$f_X(x)$ を持つとし,$g(x)$ を $\mathbb{R}$ 上の実数値関数とする.このとき,$g(X)$ の期待値を

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x) f_X(x), & \text{(離散型)} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\, dx, & \text{(連続型)} \end{cases}$$

  で定義する.ただし,$\sum_x$ は $\{x \in \mathbb{R} : f_X(x) > 0\}$ 上の和とする.$\sum_x |g(x)| f_X(x) < \infty$ もしくは $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$ のとき,$g(X)$ の期待値を定義することにする.期待値が定義されるとき,$g(X)$ の期待値が存在するという.

- a, b を定数とする.確率変数 X と Y は有限の分散を持つとき,次が成立する:
  - 期待値:$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
  - 分散:$VAR[aX + bX] = a^2 VAR[X] + b^2 VAR[Y] + 2abCOV[X, Y]$.
  - 共分散:X と Y が独立のとき,$COV[X, Y] = 0$.

- 標準正規分布の分布関数を $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) e^{-t^2/2}\, dt$ としたとき,次をみたす:$\Phi(1.28) = 0.90$, $\Phi(1.65) = 0.95$, $\Phi(2.33) = 0.99$,

- 正規分布に従う確率変数の線形結合の分布も正規分布に従う.