今野　(2005 年 12 月 22 日作成)

2004 年度情報統計学の試験問題の解説 (2005 年 1 月 11 日実施)

注意　急いで作成したのでタイプミスがあると思うので注意してください！

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) なに (どのような定義、定理，および性質) をどのように用いたかをできるだけ明示すること．

(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　確率変数 $X_1, X_2, \ldots, X_n\,(n \geq 2)$ は独立同一に母数 $\lambda\,(\lambda > 0)$ のポアソン分布に従うとする．すなわち，各 $X_i\,(i = 1, 2, \ldots, n)$ は確率関数

$$f_X(x) = \mathbb{P}(X_i = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, \ldots$$

をもち，$(X_1, X_2, \ldots, X_n)$ は同時確率関数

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

を持つ．このとき，以下の問いに答えよ．

(a) $X_1, \ldots, X_n$ の標本平均を $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ とする．$\bar{X}_n$ の平均 $\mathbb{E}[\bar{X}_n]$ と分散 $\mathbb{VAR}[\bar{X}_n]$ を求めよ．

(b) $S = n\bar{X}_n$ とおく．$S = s$ が与えられたときの $(X_1, X_2, \ldots, X_n)$ の条件付き確率は λ に依存しないことを示せ．

(c) ネイマンの因子分解定理（ある統計量が十分であるための必要十分条件）を用いて，S が λ の十分統計量であることを示せ．

問題 1 で証明なしで用いてよいこと

(1a) X_1 の平均と分散は $\mathbb{E}[X_1] = \lambda$, $\mathbb{VAR}[X_1] = \lambda$ で与えられる．

(1b) $X_1 + X_2$ の分布は母数 2λ のポアソン分布に従う．

(1c) 2 次の積率が有限で独立な確率変数の 1 次結合に対する期待値と分散の公式

(1d) ネイマンの因子分解定理（講義の定理 5.1）

$\boxed{\text{解答}}$ (a) ヒント (1a)，平均の線形性，独立同一分布に従っていることより $\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \lambda$. また，ヒント (1a)，分散の性質，独立同一分布より $\mathbb{VAR}[\bar{X}_n] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{VAR}[X_i] = \frac{1}{n}\lambda$.
(b) ヒント (1b) を繰り返し適用することから $S = \sum_{i=1}^n X_i$ は平均 $n\lambda$ のポアソン分布に従う．したがって，$\mathbb{P}(S = s) = \frac{(n\lambda)^s}{s!}e^{-n\lambda}$, $s = 0, 1, 2, \ldots$ となる．一方，$(X_1, X_2, \ldots, X_n)$ の同時確率密度は $\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$, $(x_1, \ldots, x_n = 0, 1, \ldots$ である．したがって，$S = s$ が与えられたときの $(X_1, X_2, \ldots, X_n)$ の条件付確率は

$$\begin{aligned}\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid S = s) &= \frac{\mathbb{P}(S = s, X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)}{\mathbb{P}(S = s)} \\ &= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \Big/ \frac{(n\lambda)^s}{s!}e^{-n\lambda} = \frac{s!}{x_1!x_2!\cdots x_n!}\frac{1}{n^s}\end{aligned}$$

となり θ に依存しなくなる．したがって，S は θ の十分統計量である．
(c)

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{1}{x_1!x_2!\cdots x_n!}\lambda^{\sum_{i=1}^n x_i}e^{-n\lambda}$$

となるので，$S = \sum_{i=1}^n X_i$ はネイマンの因子分解定理より θ の十分統計量になる．

$\boxed{\text{問題 2}}$ X_1, X_2, X_3, X_4 を確率密度関数

$$f_X(x) = \begin{cases} 2x & (0 < x < 1), \\ 0 & (\text{その他の場合}) \end{cases}$$

からの標本の大きさが 4 のランダム標本とし，$X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ をその順序統計量とする．

(a) X_1 の分布関数 $F_X(x) = \mathbb{P}(X_1 \le x)$ を求めよ．注意：解答は，F_X は実数上に定義された関数であることを明示すること．

(b) 最小値 $X_{(1)}$ と最大値 $X_{(4)}$ のそれぞれの周辺確率密度関数を求めよ．

$\boxed{\text{解答}}$ (a)

$$F_x(x) = \begin{cases} 0 & (x \le 0), \\ x^2 & (0 < x < 1), \\ 1 & (x \ge 1). \end{cases}$$

(b) $0 < x < 1$ に対して，$\mathbb{P}(X_{(1)} \le x) = \mathbb{P}(X_1 \le x_1, \ldots, X_4 \le x) = \prod_{i=1}^4 \mathbb{P}(X_i \le x) = x^8$ となる．よって，$f_{X_{(1)}}(x) = (d/dx)\mathbb{P}(X_{(1)} \le x) = 8x^7$ となる．また，$\mathbb{P}(X_{(4)} \le x) = 1 - \mathbb{P}(X_{(4)} > x) = 1 - \mathbb{P}(X_1 > x_1, \ldots, X_4 > x) = 1 - (1 - x^2)^4$ となる．したがって，$f_{X_{(4)}}(x) = (d/dx)\mathbb{P}(X_{(4)} \le x) = 8(1 - x^2)^3 x$ となる．故に，

$$f_{X_{(1)}}(x) = \begin{cases} 0 & (x \le 0), \\ 8x^7 & (0 < x < 1), \\ 0 & (x \ge 1), \end{cases} \qquad f_{X_{(4)}}(x) = \begin{cases} 0 & (x \le 0), \\ 8x(1 - x^2)^3 & (0 < x < 1), \\ 0 & (x \ge 1). \end{cases}$$

$\boxed{\text{問題 3}}$ $X_1, X_2, \ldots, X_n\,(n \ge 2)$ を平均 μ，分散 $\sigma^2\,(\sigma^2 < \infty)$ の分布からの標本の大きさが n のランダム標本とし，

$$\bar{X}_n = \frac{1}{n}\sum_{\ell=1}^n X_\ell, \qquad S_n^2 = \frac{1}{n-1}\sum_{\ell=1}^n (X_\ell - \bar{X}_n)^2$$

とする．ただし，$\sigma > 0$ とする．このとき，以下の問いの答えよ．
(a) $S_n^2 \xrightarrow{P} \sigma^2$ を示せ．

(b) 確率変数列 $\{T_n\}_{n=1}^\infty$ が標準正規分布に分布収束するとは，どのようなことをみたしているかを述べよ．ただし，標準正規分布の確率密度関数 $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ を用いて記述すること．

(c) つぎのことを示せ．$\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{S_n^2}} \xrightarrow{d} N(0, 1)$．ただし，$\xrightarrow{d}$ は分布収束を示す．

問題 3 で証明なしで用いてよいこと

(3a) 大数の法則と中心極限定理

(3b) $Y_n \xrightarrow{P} Y$ と $Z_n \xrightarrow{P} Z$ ならば，$Y_n + Z_n \xrightarrow{P} Y + Z$ と $Y_n Z_n \xrightarrow{P} YZ$ が成り立つ．

(3c) $g : \mathbb{R} \to \mathbb{R}$ を連続関数としたとき，$Z_n \xrightarrow{P} Z$ ならば，$g(Z_n) \xrightarrow{P} g(Z)$ となる．

(3d) $Y_n \xrightarrow{d} Y$ と $Z_n \xrightarrow{P} c$ （c は零でない定数) ならば，$Y_n Z_n \xrightarrow{d} cY$ となる．

ここで，$\{Z_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ を確率変数列とし，Z, Y を確率変数とする．

解答 (a) $\mathbb{E}[X_1] = \mu$ と $\mathbb{E}[|X_1|] \leq \mathbb{E}[X_1^2] = \sigma^2 + \mu^2 < \infty$ となるので，大数の法則を用いると $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu$ と $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \sigma^2 + \mu^2$ となる．さらに，ヒント (3b) から $\frac{1}{n-1}\sum_{i=1}^n = \frac{1}{n-1}\{\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\} = \frac{n}{n-1}\{\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2\} \xrightarrow{P} (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$ がわかる．
(b) 標準正規分布の分布関数は実数上で連続なので，すべての実数 x に対して，

$$\lim_{n\to\infty} \mathbb{P}(T_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

をみたすことである．
(d) $\mathbb{E}[X_1] = \mu$ と $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ に注意して中心極限定理を用いると

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,\, 1)$$

となる．さらに，(a) と (3b) または (3c) を用いると $S_n^2/\sigma^2 \xrightarrow{P} 1$ となる．最後に (3d) を用いると

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{S_n^2}} = \left(1 \middle/ \sqrt{\frac{S_n^2}{\sigma^2}}\right) \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,\, 1)$$

を得る．