今野　(2007 年 1 月 9 日)

情報統計学の試験問題（試験時間は 75 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) 講義で述べたことは設問中で証明することを求めていない場合に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．

答案の返却の日時は掲示する．

問題 1　離散型確率変数 X は，確率関数

$$f(x|\,\theta) = \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}, \qquad x = -1, 0, 1$$

からの標本の大きさが 1 のランダム標本とする．ただし，$0 < \theta < 1$ とする．ふたつの統計量

$$S = S(X) = |X|, \qquad T = T(X) = \begin{cases} 2, & (X = 1) \\ 0, & \text{その他} \end{cases}$$

を考える．

(1) S の確率分布を求めよ[1]．

(2) T の確率分布を求めよ．

(3) S が与えられたときの X の条件付確率分布を求め[2]，S は θ の十分統計量かどうかを調べよ．

(4) S は θ の不偏推定量かどうかを調べよ．

(5) T は θ の不偏推定量かどうかを調べよ．

(6) S と T の平均 2 乗誤差 $\mathrm{MSE}_S(\theta) = \mathbb{E}[(S-\theta)^2]$ と $\mathrm{MSE}_T(\theta) = \mathbb{E}[(T-\theta)^2]$ を求めよ．S と T の平均 2 乗誤差 $\mathrm{MSE}_S(\theta)$ と $\mathrm{MSE}_T(\theta)$ の大小の比較をせよ．(横軸を θ とし，縦軸を MSE の値として，S と T の平均 2 乗誤差 $\mathrm{MSE}_S(\theta)$ と $\mathrm{MSE}_T(\theta)$ のグラフを描き比較すること)

解答と配点　(1) 5 (2) 5 (3) 10 (4) 5 (5) 5 (10) 10 計 40

(1) S の取り得る値は 0, 1 である．

$$\{S = 1\} \Longleftrightarrow \{X = 1\} \cup \{X = -1\}$$

と事象 $\{X = 1\}$ と $\{X = -1\}$ は排反なので，

$$\mathbb{P}(S=1) = \mathbb{P}(\{X=1\}\cup\{X=-1\}) = \mathbb{P}(X=1)+\mathbb{P}(X=-1) = \left(\frac{\theta}{2}\right)^{|-1|}(1-\theta)^{1-|-1|} + \left(\frac{\theta}{2}\right)^{|1|}(1-\theta)^{1-|1|} = \theta$$

つぎに，

$$\{S = 0\} \Longleftrightarrow \{X = 0\}$$

[1] 確率関数または確率分布表を求めること．

[2] 条件付き確率関数を求めるか，X が与えられた値ごとの確率分布表を求めればよい．

より

$$\mathbb{P}(S=0)=\mathbb{P}(X=0)=\left(\frac{\theta}{2}\right)^{|0|}(1-\theta)^{1-|0|}=1-\theta$$

したがって，

$$\mathbb{P}(S=s)=\begin{cases}1-\theta & (s=0)\\ \theta & (s=1)\\ 0 & (\text{その他})\end{cases}$$

となる．

(2) T の取り得る値は $0,\,2$ である．

$$\{T=0\}\Longleftrightarrow\{X=0\}\cup\{X=-1\}$$

と事象 $\{X=0\}$ と $\{X=-1\}$ は排反なので，

$$\begin{aligned}\mathbb{P}(T=0)=\mathbb{P}(\{X=0\}\cup\{X=-1\}) &= \mathbb{P}(X=0)+\mathbb{P}(X=-1)=\left(\frac{\theta}{2}\right)^{|0|}(1-\theta)^{1-|0|}+\left(\frac{\theta}{2}\right)^{|-1|}(1-\theta)^{1-|-1|}\\ &= 1-\theta+\frac{\theta}{2}=1-\frac{\theta}{2}\end{aligned}$$

　つぎに，

$$\{T=2\}\Longleftrightarrow\{X=1\}$$

より

$$\mathbb{P}(T=2)=\mathbb{P}(X=1)=\left(\frac{\theta}{2}\right)^{|1|}(1-\theta)^{1-|1|}=\frac{\theta}{2}$$

したがって，

$$\mathbb{P}(T=t)=\begin{cases}1-\dfrac{\theta}{2} & (t=0)\\ \dfrac{\theta}{2} & (t=2)\\ 0 & (\text{その他})\end{cases}$$

となる．

(3) $S=0$ のとき，

$$\mathbb{P}(X=0|\,S=0)=\frac{\mathbb{P}(\{X=0\}\cap\{S=0\})}{\mathbb{P}(S=0)}=\frac{\mathbb{P}(X=0)}{\mathbb{P}(S=0)}=\frac{\left(\frac{\theta}{2}\right)^{|0|}(1-\theta)^{1-|0|}}{1-\theta}=1$$

よって，$\{S=0\}\Longleftrightarrow\{X=0\}$ なので，

$$\mathbb{P}(X=x|\,S=0)=\begin{cases}1 & (x=0)\\ 0 & (\text{その他})\end{cases}$$

$S=1$ のとき，

$$\mathbb{P}(X=1|\,S=1)=\frac{\mathbb{P}(\{X=1\}\cap\{S=1\})}{\mathbb{P}(S=1)}=\frac{\mathbb{P}(X=1)}{\mathbb{P}(S=1)}=\frac{\left(\frac{\theta}{2}\right)^{|1|}(1-\theta)^{1-|1|}}{\theta}=\frac{1}{2}$$

と

$$\mathbb{P}(X=-1|\,S=1)=\frac{\mathbb{P}(\{X=-1\}\cap\{S=1\})}{\mathbb{P}(S=1)}=\frac{\mathbb{P}(X=-1)}{\mathbb{P}(S=1)}=\frac{\left(\frac{\theta}{2}\right)^{|-1|}(1-\theta)^{1-|-1|}}{\theta}=\frac{1}{2}$$

となる．$\{S=1\}\Longleftrightarrow\{X=1\}\cup\{X=-1\}$ より，

$$\mathbb{P}(X=x|\,S=1)=\begin{cases}\dfrac{1}{2} & (x=-1,\,1)\\ 0 & (\text{その他})\end{cases}$$

よって，S を与えたときの条件付確率 $\mathbb{P}(X=x|\,S=1)$ と $\mathbb{P}(X=x|\,S=0)$ は θ に依存したので，S は θ の十分統計量となる．

(4) S の確率関数より

$$\mathbb{E}[S]=\sum_{s=0,\,1}s\times\mathbb{P}(S=s)=0\times(1-\theta)+1\times\theta=\theta,\qquad {}^{\forall}\theta\in(0,\,1)$$

より S は θ の不偏推定量となる．
(5) S の確率関数より

$$\mathbb{E}[T] = \sum_{t=0,\,2} t \times \mathbb{P}(T = t) = 0 \times (1 - (\theta/2)) + 2 \times (\theta/2) = \theta, \qquad {}^{\forall}\theta \in (0,\,1)$$

より T は θ の不偏推定量となる．
(6) S は θ の不偏推定量なので，

$$\begin{aligned} \mathrm{MSE}_S(\theta) &= \mathbb{E}[(S-\theta)^2] = \mathbb{VAR}[S] = \mathbb{E}[S^2] - \{\mathbb{E}[S]\}^2 = \sum_{s=0,\,1} s^2 \times \mathbb{P}(S = s) - \theta^2 \\ &= 0^2 \times (1-\theta) + 1^2 \times \theta - \theta^2 = \theta - \theta^2 \end{aligned}$$

となる．一方，T も θ の不偏推定量なので，

$$\begin{aligned} \mathrm{MSE}_T(\theta) &= \mathbb{E}[(T-\theta)^2] = \mathbb{VAR}[T] = \mathbb{E}[T^2] - \{\mathbb{E}[T]\}^2 = \sum_{t=0,\,2} t^2 \times \mathbb{P}(T = t) - \theta^2 \\ &= 0^2 \times (1-(\theta/2)) + 2^2 \times (\theta/2) - \theta^2 = 2\theta - \theta^2 \end{aligned}$$

となる．

$$\mathrm{MSE}_T(\theta) - \mathrm{MSE}_S(\theta) = 2\theta - \theta^2 - (\theta - \theta^2) > 0 \qquad {}^{\forall}\theta \in (0,\,1)$$

$\boxed{\text{問題 2}}$ $X_1,\, X_2,\, X_3,\, X_4$ 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．

(1) X_1 の分布関数 $F_X(x) = \mathbb{P}(X_1 \le x)$ を求めよ[3]．

(2) 最小値 $X_{(1)}$ と最大値 $X_{(4)}$ の周辺確率密度関数 $f_{X_{(1)}}(x)$ と $f_{X_{(4)}}(x)$ を求めよ．

(3) $\int_{-\infty}^{\infty} f_{X_{(1)}}(x)\,dx = 1$ と $\int_{-\infty}^{\infty} f_{X_{(4)}}(x)\,dx = 1$ を確認せよ..

$\boxed{\text{解答と配点}}$ (1) 5 (2) 20 (3) 10 計 35
(1)

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

(2) $0 < x < 1$ 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，

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

(3)

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X_{(1)}}(x)\,dx &= \int_0^1 8x^7\,dx = \left[x^8\right]_0^1 = 1 \\ \int_{-\infty}^{\infty} f_{X_{(4)}}(x)\,dx &= \int_0^1 8x(1-x^2)^3\,dx = \left[-(1-x^2)^4\right]_0^1 = 1 \end{aligned}$$

[3] 解答は，F_X は実数上に定義された関数であることに注意せよ．

$\boxed{\text{問題 3}}$　$X_1, X_2, \ldots, X_n\ (n \geq 2)$ を平均 μ，分散 $\sigma^2\ (0 < \sigma < \infty)$ 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．このとき，以下の問いの答えよ．

(1) $S_n^2 \xrightarrow{P} \sigma^2$，すなわち S_n^2 が σ^2 に確率収束することを示せ．

(2) 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 $(1/\sqrt{2\pi})e^{-x^2/2}$ を用いて記述すること．

(3) つぎのことを示せ．

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{S_n^2}} \xrightarrow{d} N(0,\, 1).$$

ただし，$\xrightarrow{d}$ は分布収束を示す．

$\boxed{\text{問題 3}}$ で証明なしで用いてよいこと

(a) 大数の法則と中心極限定理

(b) $Y_n \xrightarrow{P} Y$ 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．

(c) $g : \mathbb{R} \to \mathbb{R}$ 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．

(d) $Y_n \xrightarrow{d} Y$ 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．

$\boxed{\text{解答と配点}}$　(1) 10 (2) 5 (3) 10 計 25

(1) $\mathbb{E}[X_1] = \mu$ 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 (b) 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．

(2) 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．

(3) $\mathbb{E}[X_1] = \mu$ 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，(1) と (b) または (c) 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

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

を得る．