今野　(2011 年 01 月 11 日)

情報統計学の試験問題（試験時間は 80 分・裏面にヒントがあり）

解答上の注意

特別な指示がある場合を除き，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　離散型確率変数列 $X_1, X_2, \ldots, X_n$ を母数 $p\,(0 < p < 1)$ のベルヌーイ分布からの標本の大きさ 3 のランダム標本とする．すなわち，

$$\mathbb{P}(X_i = 1) = p, \qquad \mathbb{P}(X_i = 0) = 1 - p, \qquad i = 1, 2, \ldots, n.$$

(1) 確率変数 X_1 の期待値 $\mathbb{E}[X_1]$ と分散 $\mathbb{VAR}[X_1]$ を求めよ．
(2) 標本平均 $\bar{X}_n = (1/n)(X_1 + \cdots + X_n)$ の期待値を計算せよ．
(3) $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ を観測したときの p の尤度関数を書き，それを最大にする p を求めることにより，p の最尤推定値を求めよ．
(4) p の推定量である標本平均 $\bar{X}_n$ の平均自乗誤差 $\mathrm{MSE}(p, \bar{X}_n) = \mathbb{E}[(\bar{X}_n - p)^2], (0 < p < 1)$ を求め，そのグラフを描け．

配点 (1) 10 + 10 点 (2) 10 点 (3) 10 点 (4) 10 計 50 点

解答 (1)

$$\begin{aligned}
\mathbb{E}[X_1] &= \sum_{x=0,1} x\mathbb{P}(X_1 = x) = 0 \times \mathbb{P}(X_1 = 0) + 1 \times \mathbb{P}(X_1 = 1) = 0 \times (1-p) + 1 \times p = p, \\
\mathbb{VAR}[X_1] &= \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] = \mathbb{E}[(X_1 - p)^2] = \sum_{x=0,1} (x-p)^2 \mathbb{P}(X_1 = x) \\
&= (0-p)^2 \mathbb{P}(X_1 = 0) + (1-p)^2 \mathbb{P}(X_1 = p) \\
&= p^2(1-p) + (1-p)^2 p = p(1-p)\{p + (1-p)\} = p(1-p)
\end{aligned}$$

(2) $\bar{X}_n$ の期待値は以下になる：

$$E[\bar{X}_n] = (1/n)\{E[X_1] + \cdots + E[X_n]\} = (1/n) \times np = p.$$

(3)

$$L_n(p|\, x_1, x_2, \ldots, x_n) = \Pi_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n - \sum_{i=1}^n x_i}$$

よって，対数尤度関数は

$$\ell_n(p) = \log L_n(p|\, x_1, x_2, \ldots, x_n) = \sum_{i=1}^n x_i \log p + (n - \sum_{i=1}^n x_i) \log(1-p)$$

これより

$$
\begin{aligned}
\frac{d}{dp}\ell_n(p) &= \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p} = 0 \\
&\Longleftrightarrow (1-p)\sum_{i=1}^n x_i = p\left(n - \sum_{i=1}^n x_i\right) \\
&\Longleftrightarrow np = \sum_{i=1}^n x_i \\
&\Longleftrightarrow p = \frac{\sum_{i=1}^n x_i}{n}
\end{aligned}
$$

(4) $X_1, \ldots, X_n$ は互いに独立であることと (1) より

$$
\begin{aligned}
\mathrm{MSE}(p, \bar{X}_n) &= \mathbb{E}[(\bar{X}_n - p)^2] = \mathbb{VAR}[X_n] = \mathbb{VAR}\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{VAR}[X_i] = \frac{1}{n^2} n\mathbb{VAR}[X_1] = \frac{\mathbb{VAR}[X_1]}{n} = \frac{p(1-p)}{n}
\end{aligned}
$$

$\boxed{\text{問題 2}}$ 連続型確率変数列 X_1, X_2, X_3 は正規母集団 $N(\mu, 3)$（平均が 3，分散が 3 の正規分布に従う母集団分布）からの標本の大きさ 3 のランダム標本とする．次の仮説検定問題を考える：

$$
\text{帰無仮説} \quad H_0 : \mu = 0 \quad \text{vs. 対立仮説} \quad H_1 : \mu = 2
$$

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) $\bar{X}_3 = (1/3)(X_1 + X_2 + X_3)$ とする．$\bar{X}_3$ は 帰無仮説 $H_0 : \mu = 0$ のもとでどのようなものになるかを答えよ．理由も述べること．

(3) H_0 の棄却域を

$$
C(t) = \{(x_1, x_2, x_3) \in \mathbb{R}^3; x_1 + x_2 + x_3 > t\}
$$

としたとき，$C(t)$ が有意水準 0.1 の検定の棄却域になるように t をひとつ定めよ．t を求めるときには小数第 3 位を四捨五入せよ．

(4) 棄却域 $C(t)$ の検出力を 対立仮説 $H_1 : \mu = 2$ のもとで求めよ．ただし，解答は関数 $\Phi(x) = \int_{-\infty}^x (1/\sqrt{2\pi})e^{-t^2/2}\,dt$ および t を用いて表現せよ．すなわち，具体的な値を計算しなくともよい．

$\boxed{\text{配点}}$ (1) 5 + 5 (2) 10+10 (3) 10 (4) 10 計 50

$\boxed{\text{解答}}$ (1) H_0 は単純，H_1 も単純.
(2) $\mu = 0$ のとき，

$$
\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \mathbb{E}[\frac{1}{3}(X_1 + X_2 + X_3)] = \frac{1}{3}\{\mathbb{E}[\mathbb{E}[X_1] + \mathbb{E}[\mathbb{E}[X_2] + \mathbb{E}[\mathbb{E}[X_3]\} = 0 \\
\mathbb{VAR}[\bar{X}_n] &= \mathbb{VAR}\left[\frac{1}{3}(X_1 + X_2 + X_3)\right] = \frac{1}{9}\{\mathbb{E}[\mathbb{VAR}[X_1] + \mathbb{E}[\mathbb{VAR}[X_2] + \mathbb{E}[\mathbb{VAR}[X_3]\} \\
&= \frac{3 \times 3}{9} = 1
\end{aligned}
$$

よって，$\bar{X}_3$ は標準正規分布 $N(0,\,1)$ に従う．
(3) $\mu=0$ のとき，$\bar{X}_3$ は標準正規分布 $N(0,\,1)$ に従うので，

$$\begin{aligned}
\mathbb{P}_{\mu=0}((X_1,\,X_2,\,X_3)\in C(t)) &= \mathbb{P}_{\mu=0}\left(X_1+X_2+X_3>t\right)\\
&= \mathbb{P}_{\mu=0}\left(\bar{X}_3>3t\right)\\
&= 1-\mathbb{P}\left(\bar{X}_3\le 3t\right)=1-\Phi(3t)=0.10\\
&\iff \Phi(3t)=0.90\\
&\iff 3t=1.28 \iff t=0.43
\end{aligned}$$

(4) $\mu=2$ のとき，$\bar{X}_3-2$ が標準正規分布 $N(0,\,1)$ に従うことから，

$$\begin{aligned}
\mathbb{P}_{\mu=2}((X_1,\,X_2,\,X_3)\in C(t)) &= \mathbb{P}_{\mu=2}(\bar{X}_3>3t)=\mathbb{P}_{\mu=2}(\bar{X}_3-2>3t-2)=\Phi(3t-2)\\
&= 1-\Phi(3t-1)
\end{aligned}$$

ヒント：以下のことは証明なしで用いてよい

- 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：$\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$.
  - $\mathbb{VAR}[X]=\mathbb{E}[(X-\mathbb{E}[X])^2]$.
  - 分散：$\mathbb{VAR}[aX+bX]=a^2\mathbb{VAR}[X]+b^2\mathbb{VAR}[Y]+2ab\mathbb{COV}[X,\,Y]$.
  - 共分散：X と Y が独立のとき，$\mathbb{COV}[X,\,Y]=0$.
- 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．

解答

成績について

得点	0 ～ 19	20 ～ 39	40 ～ 59	60 ～ 84	85 ～ 100
成績	D	C	B	A	A^+

得点分布　平均点 = 26.3，中央値 = 27、標準偏差 = 21.1

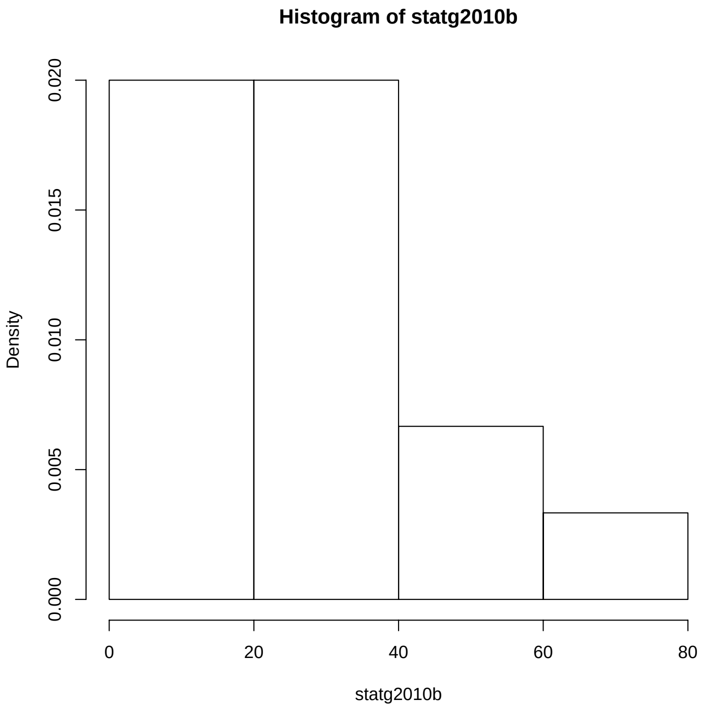

Figure 1: This is a figure