昭和56年度（問　題）

1. 区間$(0, \theta)$上の一様分布

$$f(x\,;\theta)=\begin{cases}\dfrac{1}{\theta} & 0<x<\theta\\ 0 & \text{その他}\end{cases}$$

からの標本$(x_1, x_2, \cdots\cdots, x_n)$ に対して，θの２つの推定量 $2\overline{X}=\frac{2}{n}\sum_{i=1}^{n}X_i$ と $\frac{n+1}{n}\max(X_1, X_2, \cdots\cdots, X_n)$ の不偏性を示し，かつ２つの推定量の有効性を比較せよ。

2. 正規母集団 $N(\mu, 25)$ の母平均 μの検定で，帰無仮説 $H_0: \mu=0$，対立仮説 $H_1: \mu=1$ において，第１種および第２種の誤りをおかす確率をともに 0.01 にしたい。この検定に要する標本の大きさ nを求めよ。 $u(0.01)=2.33$ とする。

3. ５種類の商品を販売している会社で，販売実績から任意抽出して1,000件のサンプルをとったところ，次のような結果を得た。商品種類によるニーズ（あるいは販売姿勢）の違いはあるといえるか，有意水準５％で調べよ。

商品種類	1	2	3	4	5	合計
販売件数	210	246	150	183	211	1,000

必要あれば次の数値を用いよ。

分布値 ＼ 自由度 ϕ	$t_\phi(0.025)$	$t_\phi(0.05)$	$\chi^2_\phi(0.025)$	$\chi^2_\phi(0.05)$	$u(0.025)$	$u(0.05)$
4	2.776	2.132	11.143	9.488		
5	2.571	2.015	12.833	11.070	1.960	1.645
6	2.447	1.943	14.449	12.592		

4. ある人種のABO式による血液型のA，B，AB，Oはそれぞれ $(2+\theta)/4$，$(1-\theta)/4$，$(1-\theta)/4$，$\theta/4$ で生ずることがわかっている。ここでθは未知の母数である。n人の観察の結果これら４種の人数がそれぞれp，q，r，s（いずれも正）人になったとき，θの最尤推定値を求めよ。

5. 校正者Aによれば１頁当り平均λ_1個の誤植が残り，Bによれば１頁当り平均λ_2個の誤植が残る。ただし，λ_1，λ_2はともに未知とする。Aが校正を行なった結果20頁中に17個の誤植が残り，Bが校正を行なった結果40頁中にr個の誤植が残った。

１頁当りの平均誤植数を比較することによって，$\lambda_1=\lambda_2$といえるrの値を有意水準５％として求めよ。ここに，$u(0.05)=1.645$，$u(0.025)=1.960$ とする。

昭和56年度（解答例）

1. (1) 推定量 $2\overline{X}=\frac{2}{n}\sum_{i=1}^{n}X_i$ について

$$E(2\overline{X})=E\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right)=\frac{2}{n}\sum_{i=1}^{n}E(X_i)$$

$$=\frac{2n}{n}\cdot\int_0^\theta\frac{x}{\theta}dx=2\cdot\frac{\theta^2}{2\theta}=\theta \qquad \cdots\cdots ①$$

$$V_{ar}(2\overline{X})=V_{ar}\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right)=\frac{4}{n^2}\sum_{i=1}^{n}V_{ar}(X_i)=\frac{4n}{n^2}\left(\int_0^\theta\frac{x^2}{\theta}dx-\left(\frac{\theta}{2}\right)^2\right)$$

$$=\frac{4}{n}\left(\frac{\theta^3}{3\theta}-\frac{\theta^2}{4}\right)=\frac{\theta^2}{3n} \qquad \cdots\cdots ②$$

(2) 推定量 $\frac{n+1}{n}\max(X_1,X_2,\cdots\cdots,X_n)$ について

$\max(X_1,X_2,\cdots\cdots,X_n)=X_{(n)}$ の分布関数は

$$P\{\max(X_1,X_2,\cdots\cdots,X_n)<x\}=P\{X_1<x,X_2<x,\cdots,X_n<x\}$$

$$=P\{X_1<x\}\cdot P\{X_2<x\}\cdot\cdots\cdot P\{X_n<x\}=\left(\int_0^x\frac{dx}{\theta}\right)^n$$

$$=\left(\frac{x}{\theta}\right)^n \quad (0<x<\theta)$$

だから，$X_{(n)}$ の密度関数は $n\cdot\left(\frac{x}{\theta}\right)^{n-1}\cdot\frac{1}{\theta}$ $(0<x<\theta)$ であることより

$$E(X_{(n)})=\int_0^\theta x\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{dx}{\theta}=\frac{n}{\theta^n}\int_0^\theta x^n dx=\frac{n}{\theta^n}\cdot\frac{\theta^{n+1}}{n+1}=\frac{n}{n+1}\theta$$

$$V_{ar}(X_{(n)})=\int_0^\theta x^2\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{dx}{\theta}-\left(\frac{n}{n+1}\theta\right)^2=\frac{n}{\theta^n}\int_0^\theta x^{n+1}dx-\left(\frac{n}{n+1}\theta\right)^2$$

$$=\frac{n}{\theta^n}\cdot\frac{\theta^{n+2}}{n+2}-\frac{n^2}{(n+1)^2}\theta^2=n\cdot\frac{(n+1)^2-n(n+2)}{(n+2)(n+1)^2}\theta^2$$

$$=\frac{n}{(n+2)(n+1)^2}\theta^2$$

従って

$$E\left(\frac{n+1}{n}\max(X_1,X_2,\cdots\cdots,X_n)\right)=\frac{n+1}{n}E(X_{(n)})=\theta \qquad \cdots\cdots ③$$

$$V_{ar}\left(\frac{n+1}{n}\max(X_1,X_2,\cdots\cdots,X_n)\right)=\left(\frac{n+1}{n}\right)^2V_{ar}(X_{(n)})=\frac{\theta^2}{n(n+2)}$$

$\cdots\cdots$ ④

①,③ より，$2\bar{X}$，$\dfrac{n+1}{n}\max(X_1, X_2, \cdots\cdots, X_n)$ の不偏性が示され，②,④ より $\dfrac{n+1}{n}\max(X_1, X_2, \cdots\cdots, X_n)$ の方が $2\bar{X}$ より有効といえる。

2. 第1種の誤りをおかす確率を 0.01 とするため

$$P(\bar{X}>C \mid \mu=0)=P\left(\frac{\bar{X}-0}{5/\sqrt{n}}>\frac{C}{5/\sqrt{n}}\right)=\int_{C/5/\sqrt{n}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=0.01$$

また，第2種の誤りをおかす確率を 0.01 とするため，

$$P(\bar{X}\leq C \mid \mu=1)=1-P(\bar{X}>C \mid \mu=1)=0.01 \text{ より}$$

$$P(\bar{X}>C \mid \mu=1)=P\left(\frac{\bar{X}-1}{5/\sqrt{n}}>\frac{C-1}{5/\sqrt{n}}\right)=\int_{(C-1)/5/\sqrt{n}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.99$$

以上から，

$$\frac{C\sqrt{n}}{5}=2.33 \qquad \frac{(C-1)\sqrt{n}}{5}=-2.33$$

となり，　$\sqrt{n}/5=4.66$　　$n=(23.3)^2=542.89$

n は 543 以上とればよい。

3. ニーズ（あるいは販売姿勢）に違いはないとすれば，

H_0 ：　$p_i=\dfrac{1}{5}$　（p_i は i 番目の商品種類である確率）

$$\therefore\ np_i=1{,}000\times\frac{1}{5}=200$$

$$\chi^2=\sum_{i=1}^{5}\frac{(f_i-np_i)^2}{np_i} \qquad (f_i \text{ は実現個数})$$

$$=\frac{1}{200}\{(210-200)^2+(246-200)^2+(150-200)^2+(183-200)^2+(211-200)^2\}$$

$$=\frac{1}{200}(100+2{,}116+2{,}500+289+121)=5{,}126/200=25.63$$

これを自由度 $\phi=5-1=4$ の χ^2－分布上側5％点 $\chi^2_4(0.05)=9.488$ と較べると，$25.63>9.488$ で H_0 は棄却される。

4.　4次元の確率変数（X_1，X_2，X_3，X_4）を考え，そのとり得る値は，

$$(1,0,0,0),\ (0,1,0,0),\ (0,0,1,0),\ (0,0,0,1)$$

とし，それぞれの値をとる確率を

$$(2+\theta)/4,\quad (1-\theta)/4,\quad (1-\theta)/4,\quad \theta/4$$

とする。（x_1，x_2，x_3，x_4）はこれら4つの点のいずれかとすれば，その密度は

$$f(x_1,\ x_2,\ x_3,\ x_4\ ;\ \theta) \doteqdot (2+\theta)^{x_1}\ (1-\theta)^{x_1+x_2}\ \theta^{x_4}/4$$

いま，これら4つの点が，n回のうちでそれぞれp，q，r，s回観察されたとすれば，尤度関数は，

$$L(\theta) = (2+\theta)^p(1-\theta)^{q+r}\theta^s/4^n$$

となる。これより

$$\frac{\partial \log L}{\partial \theta} = \frac{p}{2+\theta} - \frac{q+r}{1-\theta} + \frac{s}{\theta}$$

$$= \{p(1-\theta)\cdot\theta - (q+r)(2+\theta)\cdot\theta + s(2+\theta)(1-\theta)\}/(2+\theta)(1-\theta)\theta$$

$$= \frac{-n\theta^2 + (p-2q-2r-s)\theta + 2s}{\theta(1-\theta)(2+\theta)} \quad \cdots\cdots\cdots\cdots \quad ①$$

$g(\theta) = -n\theta^2 + (p-2q-2r-s)\theta + 2s = 0$とおいて$\theta$を求めれば，正負2つの根が生ずる。$\theta > 0$であるから，正根を$\hat{\theta}$とする。

$$g(0) = 2s > 0,\quad g(1) = -3q-3r < 0$$

であるから，$0 < \hat{\theta} < 1$となる。

①の分母は $0 < \theta < 1$ のとき正，分子は $0 < \theta < \hat{\theta}$ のとき正で，$\hat{\theta} < \theta < 1$ のとき負となるからLは$\hat{\theta}$で最大となり，$\hat{\theta}$はθの最尤推定量となる。

5.　$x_1 = 17$はPoisson分布$P_o(20\lambda_1)$にしたがう確率変数X_1の実現値，$x_2 = r$は$P_o(40\lambda_2)$にしたがう確率変数X_2の実現値と考えられる。

$20\lambda_1$，$40\lambda_2$はともに大きいと想像されるので，

$$(X_1 - 20\lambda_1)/\sqrt{20\lambda_1},\quad (X_2 - 40\lambda_2)/\sqrt{40\lambda_2}$$

の分布はいずれも$N(0,1)$にちかい。

そこで，$\lambda_1 = \lambda_2$であれば，この共通の値をλとおくと，

$$\left(\frac{X_1}{20}-\frac{X_2}{40}\right)\Big/\sqrt{\left(\frac{1}{20}+\frac{1}{40}\right)\lambda}$$

の分布は$N(0,1)$にちかい。

さらに，λのかわりに推定量$(X_1+X_2)/(20+40)$を用いると，

$$U=\left(\frac{X_1}{20}-\frac{X_2}{40}\right)\Big/\sqrt{\left(\frac{1}{20}+\frac{1}{40}\right)\frac{X_1+X_2}{20+40}}$$

の分布は$N(0,1)$にちかい。(注参照)

$X_1=17$として$|U|<1.960$よりrを求めると，

$$\left(\frac{17}{20}-\frac{r}{40}\right)^2<(1.960)^2\left(\frac{1}{20}+\frac{1}{40}\right)\cdot\frac{17+r}{60}$$

$$(17\times 2-r)^2<(1.96)^2\times 2\times(17+r)$$

$$r^2-75.6832r+1,025.3856<0$$

$$18\leq r\leq 58$$

(注) A. 確率変数列(X_1, Y_1)，(X_1, Y_2)，……において，X_1，X_2，……の分布はFに収束し，Y_1，Y_2，……は1に確率収束するものとする。このとき，X_1/Y_1，X_2/Y_2，……の分布はFに収束する。

(証明)

$\dfrac{X_n}{Y_n}=X_n+X_n\cdot\dfrac{1-Y_n}{Y_n}$となるので，$X_n\cdot\dfrac{1-Y_n}{Y_n}$が0に確率収束することを証明すれば，次に帰着する。

B. 確率変数列(X_1, Y_1)，(X_2, Y_2)，……においてX_1，X_2の分布はFに収束し，Y_1，Y_2，……は0に確率収束するものとする。このときX_1+Y_1，X_2+Y_2，……の分布はFに収束する。

$X_n\cdot\dfrac{1-Y_n}{Y_n}$が0に確率収束することを証明するには，任意の正数$\varepsilon$と$\delta$に対して，$n$を十分大きくとれば

$$|P(|X_n(1-Y_n)/Y_n|\geqq\varepsilon)<\delta$$

となることをいえばよい。

さて，X_nの分布はFに収束する。そこで十分大きい正数aをとって，$\pm a$が$F(x)$の連続点で，しかも$F(a)-F(-a)>1-\dfrac{\delta}{4}$となるようにする。こ

のとき n_1 を適当にとれば，$n \geq n_1$ なるすべての n に対して

$$P(-a \leq X_n \leq a) > 1 - \delta/2$$

となる。

つぎに $Y_1, Y_2, \cdots\cdots$ は 1 に確率収束するので，$(1-Y_n)/Y_n$ は $n\to\infty$ のとき 0 に確率収束する（証明略）。よって，n_2 を適当にとれば，$n \geq n_2$ なるすべての n に対して

$$P\left(\left|\frac{1-Y_n}{Y_n}\right| < \frac{\varepsilon}{a}\right) > 1 - \frac{\delta}{2}$$

となる。

したがって，$n \geq \max(n_1, n_2)$ なるすべての n に対して

$$P\left(\left|X_n\cdot\frac{1-Y_n}{Y_n}\right| \geqq \varepsilon\right) \leqq P(|X_n| > a) + P\left(\left|\frac{1-Y_n}{Y_n}\right| \geq \frac{\varepsilon}{a}\right) < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

これは，$X_n(1-Y_n)/Y_n$ が 0 に確率収束することを示している。

Bの証明

$F(x)$ の任意の連続点 a において

$$P(X_n + Y_n \leq a) \longrightarrow F(a) \qquad (n\to\infty)$$

となることをいえばよい。

いま正数 ε に対して，$F(x)$ の連続点 $a\pm\delta$ をとって，

(1)　$0 \leq F(a) - F(a-\delta) < \varepsilon/3,\quad 0 \leq F(a+\delta) - F(a) < \varepsilon/3$

ならしめる。a は $F(x)$ の連続点であり，また $F(x)$ の不連続点は可不番個しかないから，このことはいつも可能である。

このとき，任意の自然数 n について

(2)　$P(X_n \leq a-\delta) - P(|Y_n| > \delta) \leq P(X_n + Y_n \leq a) \leq P(X_n \leq a+\delta)$
$+ P(|Y_n| > \delta)$

仮定により，$n\to\infty$ とすると，

$$P(X_n \leq a-\delta) \to F(a-\delta),\quad P(X_n \leq a+\delta) \to F(a+\delta),$$
$$P(|Y_n| > \delta) \to 0$$

となるから，n を十分大きくとることにより，

(3)　$F(a-\delta) - \varepsilon/3 \leq P(X_n \leq a-\delta),\ P(X_n \leq a+\delta) < F(a+\delta) + \varepsilon/3,$

$$P(|Y_n| > \delta) < \varepsilon/3$$

とすることができる。

(2),(3)から，このような n に対して，

$$F(a-\delta) - 2\varepsilon/3 < P(X_n + Y_n \leq a) < F(a+\delta) + 2\varepsilon/3$$

さらに(1)を用いれば，

$$F(a) - \varepsilon < P(X_n + Y_n \leq a) < F(a) + \varepsilon$$

となる。これは

$$P(X_n + Y_n \leq a) \to F(a)$$

を示している。