2017 理工学 奇数

問1

ガンマ分布のパラメータに関する最尤推定を行う問題。

最後の問いで相加加相乗平均の公式を使う点以外はよくある問題である。

[1]

モーメント母関数 $M_X(t)$ を元に $E[X]$, $V[X]$ を算出する。

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \, dx \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\underbrace{\left(\frac{1}{\beta}-t\right)}_{\frac{1}{\beta}\text{相当}} x} \, dx \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot \left(\frac{1}{\frac{1}{\beta}-t} \right)^\alpha \Gamma(\alpha) \\
&= \left(\frac{1}{1-\beta t} \right)^\alpha
\end{aligned}
$$

また、
$$
\begin{aligned}
M_X'(t) &= \alpha \left(\frac{1}{1-\beta t} \right)^{\alpha-1} \frac{\beta}{(1-\beta t)^2} \\
&= \alpha\beta \left(\frac{1}{1-\beta t} \right)^{\alpha+1}
\end{aligned}
$$

$$
\begin{aligned}
M_X''(t) &= \alpha\beta(\alpha+1) \left(\frac{1}{1-\beta t} \right)^{\alpha} \frac{\beta}{(1-\beta t)^2} \\
&= \alpha\beta^2(\alpha+1) \left(\frac{1}{1-\beta t} \right)^{\alpha+2}
\end{aligned}
$$

これより、
$$
\begin{aligned}
E[X] &= M_X'(0) = \alpha\beta \\
V[X] &= E[X^2] - (E[X])^2 \\
&= M_X''(0) - (M_X'(0))^2 \\
&= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 \\
&= \alpha\beta^2 \text{ 〃}
\end{aligned}
$$

[2]

$$l(\alpha, \beta : x) = \log \prod_{i=1}^{n} f(x_i)$$
$$= \sum_{i=1}^{n} \left\{ -\alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log x_i - \frac{x_i}{\beta} \right\}$$
$$= -n \left\{ \alpha \log \beta + \log \Gamma(\alpha) \right\} + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i$$

[3]

i) $\dfrac{dl}{d\beta} = 0$ と ii) $\dfrac{dl}{d\alpha} = 0$ から、次が求まる。

i) $\iff 0 = \dfrac{dl}{d\beta} = -\dfrac{n\alpha}{\beta} + \dfrac{1}{\beta^2} \displaystyle\sum_{i=1}^{n} x_i$

$\iff 0 = -n\alpha\beta + \displaystyle\sum_{i=1}^{n} x_i$

$\iff \beta = \dfrac{1}{\alpha n} \displaystyle\sum_{i=1}^{n} x_i = \dfrac{\bar{x}_n}{\alpha}$

ii) $\iff 0 = -n \log \beta - n \left(\log \Gamma(\alpha) \right)' + \displaystyle\sum_{i=1}^{n} \log x_i$

$\iff 0 = -\log \dfrac{\bar{x}}{\alpha} - \psi(\alpha) + \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \log x_i$

$\iff 0 = \log \alpha - \log \bar{x} - \psi(\alpha) + \log \left(\displaystyle\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}$

$\iff \psi(\alpha) - \log \alpha = \log \dfrac{\tilde{x}_n}{\bar{x}_n}$

[4]

まず α の推定値を求める。データより，

$$\bar{x} = 10000/10 = 1000$$

$$\log \tilde{x} = \frac{1}{n}\sum \log x_i = 63.6/10 = 6.36$$

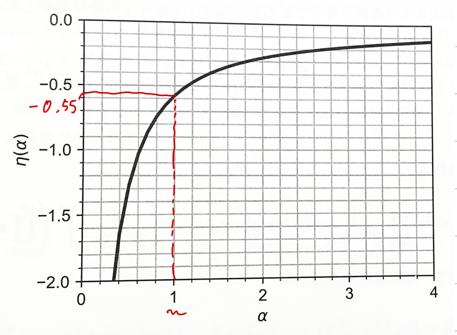
よって，

$$\psi(\alpha) - \log \alpha = 6.36 - \log 1000$$

$$= \langle 略 \rangle$$

$$= -0.55$$

グラフより，$\alpha = 1$ とわかる。



[5]

① 「$x_1 = x_2 = \cdots = x_n$ でない $\Rightarrow 0 > \log \dfrac{\tilde{x}_n}{\bar{x}_n}$」を示す。

すると，$\psi(\alpha) - \log \alpha$ は α が $(0, \infty)$ の範囲で $(-\infty, 0)$ の値をとる単調増加関数であることから，$\psi(\alpha) - \log \alpha = \log \dfrac{\tilde{x}_n}{\bar{x}_n}$ を満たす α が必ず唯一つ存在する。よって $\beta = \dfrac{\bar{x}_n}{\alpha}$ も唯一つ存在する。

[① の理由]

相加相乗平均の公式より，

$$\bar{x}_n \geq \tilde{x}_n \quad (等号は\ x_1 = x_2 = \cdots = x_n\ のときのみ成立)$$

$x_1 = x_2 = \cdots = x_n$ は成立しないので，

$$\bar{x}_n > \tilde{x}_n$$

$$\Leftrightarrow 1 > \frac{\tilde{x}_n}{\bar{x}_n}$$

$$\Leftrightarrow 0 > \log \frac{\tilde{x}_n}{\bar{x}_n}$$

問3

故障までの時間が指数分布に従うとして、最尤推定をする問題。

指数分布の無記憶性を使えるかが試されるが、それ以外は特に難しいところはない。

[1]

$$
\begin{aligned}
E[T] &= \int_0^\infty t \cdot \frac{1}{\mu} e^{-\frac{t}{\mu}} dt \\
&= \left[t (-1) e^{-\frac{t}{\mu}} \right]_0^\infty - \int_0^\infty (-1) e^{-\frac{t}{\mu}} dt \\
&= 0 + \left[-\mu e^{-\frac{t}{\mu}} \right]_0^\infty \\
&= \mu
\end{aligned}
$$

指数分布には無記憶性があるため、次の式が成り立つ。

$$P(T = t+s \mid T > t) = P(T = s)$$

これより、

$$
\begin{aligned}
\xi &= E[T \mid T > t] \\
&= \int_0^\infty u \cdot P(T = u \mid T > t)\, du \\
&= \int_0^t u \cdot 0\, du + \int_t^\infty u\, P(T = u \mid T > t)\, du \\
&= \int_0^\infty (s+t)\, P(T = s+t \mid T > t)\, ds \qquad (s = u+t \text{ で置換}) \\
&= \int_0^\infty s \cdot P(T = s)\, ds + t \cdot \int_0^\infty P(T = s)\, ds \\
&= \mu + t \;_{\text{//}}
\end{aligned}
$$

[2]

$$l_1 = \log\{f(t_1)\cdot f(t_2)\}$$

$$= -\log\mu - \frac{t_1}{\mu} - \log\mu - \frac{t_2}{\mu}$$

$$= -2\log\mu - \frac{t_1+t_2}{\mu}$$

$0 = \frac{d}{d\mu}l_1(\mu)$ を満たす $\hat{\mu}$ を求める。

$$0 = -\frac{2}{\hat{\mu}} + \frac{t_1+t_2}{\hat{\mu}^2}$$

$$\therefore \hat{\mu} = \frac{t_1+t_2}{2}$$

[3]

$T_1=t_1$ の尤度は $P(T_1=t_1)$ で. $T_2>t_2$ の尤度は $P(T_2>t_2)$ である点に注意.

$$l_2(\mu) = \log\{P(T_1=t_1)\cdot P(T_2>t_2)\}$$

$$= \log\left(\frac{1}{\mu}e^{-\frac{t_1}{\mu}}\right) + \log\left\{\int_{t_2}^{\infty}\frac{1}{\mu}e^{-\frac{t}{\mu}}dt\right\}$$

$$= -\log\mu - \frac{t_1}{\mu} + \log\left\{\left[-e^{-\frac{t}{\mu}}\right]_{t_2}^{\infty}\right\}$$

$$= -\log\mu - \frac{t_1}{\mu} - \frac{t_2}{\mu}$$

$$= -\log\mu - \frac{t_1+t_2}{\mu}$$

$0 = \frac{d}{d\mu}l_2(\mu)\Big|_{\mu=\tilde{\mu}}$ より.

$$\tilde{\mu} = t_1+t_2$$

[4]

式の操作が何を意味しているかはナゾだが、機械的に処理して解くことができる。

$\mu^{(k)}$ に関する漸化式を得たい。$\mu^{(k+1)}$ は $T_1 = t_1$, $T_2 = t_2$ が得られたときの最尤推定量なので、[2] より、

$$\mu^{(k+1)} = \frac{t_1 + \xi^{(k)}}{2} = \frac{t_1}{2} + \frac{1}{2}\left(\mu^{(k)} + t\right) \quad (\because [1])$$

$$= \frac{t_1 + t}{2} + \frac{1}{2}\mu^{(k)}$$

[5]

[4] $\Leftrightarrow$ $$\mu^{(k)} - (t_1 + t) = \frac{1}{2}\left\{\mu^{(k-1)} - (t_1 + t)\right\}$$
$$= \cdots$$
$$= \left(\frac{1}{2}\right)^k \left\{\mu^{(0)} - (t_1 + t)\right\}$$
$$\xrightarrow{k \to \infty} 0$$

これより、$\mu^{(k)} \xrightarrow{k \to \infty} t_1 + t = \tilde{\mu}$ //

問5

前半は独立な施行からなる割り当て方法と特殊な割り当て方法を比較する問題で、後半は二項分布のパラメータの信頼区間を求める問題。

特殊な知識を要さないので冷静になれば解けるが、計算量が多いので覚えられる公式は覚えて臨みたい問題である。

[1]

X は $Bin(5, \frac{1}{2})$ に従うことから,

$P(X=3) = {}_5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{5}{16}$

$E[X] = \frac{5}{2}$

$V[X] = 5\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right) = \frac{5}{4}$

[2]

最初の2人は必ずAとBに1人ずつ割り合てられる。3人目以降について, Aに a 人, Bに b 人の状態を (a, b) のように書く。

$Y=3$ となるのは3通りの場合がある.

i) $(2,1) \to (3,1) \to (3,2)$ となる場合, 確率は $\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{8}$

ii) $(2,1) \to (2,2) \to (3,2)$ となる場合, 確率は $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{4} = \frac{1}{6}$

iii) $(1,2) \to (2,2) \to (3,2)$ となる場合, 確率は $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{4} = \frac{1}{6}$

したがって, $P(Y=3) = \frac{1}{8} + \frac{1}{6} + \frac{1}{6} = \frac{11}{24}$

同様に, $P(Y=4)$ は $(2,1) \to (3,1) \to (4,1)$ の場合のみなので,

$P(Y=4) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{24}$

また, AとBは条件が対称であるので,

$P(Y=1) = P(Y=4) = \frac{1}{24}$

$P(Y=2) = P(Y=3) = \frac{11}{24}$

これより、全ての場合の確率が求まったので、期待値と分散を出せる。

$$E[Y] = 1\cdot\frac{1}{24} + 2\cdot\frac{11}{24} + 3\cdot\frac{11}{24} + 4\,\frac{1}{24}$$
$$= \frac{5}{2}$$
$$V[Y] = \left(1-\frac{5}{2}\right)^2\cdot\frac{1}{24} + \left(2-\frac{5}{2}\right)^2\frac{11}{24} + \left(3-\frac{5}{2}\right)^2\frac{11}{24} + \left(4-\frac{5}{2}\right)^2\frac{1}{24}$$
$$= \frac{5}{12}$$

〔3〕

$E[X] = E[Y]$ については確めることができた。$V[X] > V[Y]$ を一般の $n \geq 2$ について証明するのは、数式を使う場合非常に骨が折れる。
(解答のように「平均値から離れた値を取る確率が方法1より小さくなる。」と論じて良いならいいのだが…)

n 人のときの A の人数を Y_n とすると、

$$P(Y_{n+1} = y) = P(Y_n = y-1)\cdot\frac{n-(y-1)}{n} + P(Y_n = y)\cdot\frac{y}{n}$$

また、$E[Y_{n+1}] = \frac{n+1}{2}$　(全数の半分となる)

$$E[Y_{n+1}^2] = \sum_{y=1}^{n} y^2 P(Y_{n+1} = y)$$
$$= \sum_{y=1}^{n} y^2\left(1-\frac{y-1}{n}\right)P(Y_n = y-1) + \sum_{y=1}^{n} y^2\left(\frac{y}{n}\right)P(Y_n = y)$$

↓ $y' = y-1$ で置換

$$= \sum_{y'=0}^{n-1} (y'+1)^2\left(1-\frac{y'}{n}\right)P(Y_n = y') + \sum_{y=1}^{n-1} y^2\left(\frac{y}{n}\right)P(Y_n = y)$$
$$= \underbrace{\sum_{y'=1}^{n-1}}_{P(Y_n=0)=0\text{ より}} (y'^2+2y'+1)\left(1-\frac{y'}{n}\right)P(Y_n = y') + \sum_{y=1}^{n-1} \frac{y^3}{n}P(Y_n = y)$$
$$= \sum_{y=1}^{n-1} \left(y^2+2y+1-\frac{y^3}{n}-\frac{2}{n}y^2-\frac{y}{n}+\frac{y^3}{n}\right)P(Y_n = y)$$
$$= \sum_{y=1}^{n-1} \left\{\left(1-\frac{2}{n}\right)y^2 + \left(2-\frac{1}{n}\right)y + 1\right\} P(Y_n = y)$$

$$= \left(1-\frac{2}{n}\right) \sum_{y=1}^{n-1} y^2 P(Y_n = y) + \left(2 - \frac{1}{n}\right) \sum_{y=1}^{n-1} y P(Y_n = y) + \sum_{y=1}^{n-1} P(Y_n = y)$$

$$= \left(\frac{n-2}{n}\right) E[Y_n^2] + \left(2 - \frac{1}{n}\right) E[Y_n] + 1$$

$$= \left(\frac{n-2}{n}\right) \left(V[Y_n] + E[Y_n]^2\right) + \left(2 - \frac{1}{n}\right) \frac{n}{2} + 1$$

$$= \left(\frac{n-2}{n}\right) V[Y_n] + \left(1 - \frac{2}{n}\right)\left(\frac{n}{2}\right)^2 + n - \frac{1}{2} + 1$$

$$= \left(\frac{n-2}{n}\right) V[Y_n] + \frac{n^2}{4} - \frac{n}{2} + n + \frac{1}{2}$$

$$= \left(\frac{n-2}{n}\right) V[Y_n] + \frac{n^2 + 2n + 2}{4}$$

よって

$$V[Y_{n+1}] = E[Y_{n+1}^2] - E[Y_{n+1}]^2$$

$$= \left(\frac{n-2}{n}\right) V[Y_n] + \frac{n^2+2n+2}{4} - \left(\frac{n+1}{2}\right)^2$$

$$= \left(\frac{n-2}{n}\right) V[Y_n] + \frac{1}{4}$$

また、2人目までは、必ず $Y_2 = 1$ なので、

$$V[Y_2] = 0$$

次に X について、

$$V[X_{n+1}] = (n+1)\frac{1}{4} = n \cdot \frac{1}{4} + \frac{1}{4} = V[X_n] + \frac{1}{4}$$

$$V[X_2] = \frac{1}{2}$$

これらの数列を比べると、$V[X] > V[Y]$ がわかる。

(4)

$\sigma^2 = 20^2$, $n_A = 96$, $n_B = 104$ とおく。

μ_A の推定量 $\hat{\mu}_A$ は、Aのi番目の生徒の成績 X_{Ai} を用いて、

$$\hat{\mu}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} X_{Ai} \sim N\left(\mu_A, \frac{\sigma^2}{n_A}\right)$$

と表される。求めたい L_A は、次を満たす。

$$L_A = \alpha_+ - \alpha_- \quad (\text{ただし、} P(\alpha_- < \hat{\mu}_A < \alpha_+) = 0.95)$$

ここで、$P(\alpha_- < \hat{\mu}_A < \alpha_+) = 0.95$ を地道に展開すれば良いが、

次の式は良く表れるので覚えておいて良さそう。

$$\begin{cases} \alpha_+ = \mu_A + z_{0.025}\sqrt{\dfrac{\sigma^2}{n_A}} \\ \alpha_- = \mu_A - z_{0.025}\sqrt{\dfrac{\sigma^2}{n_A}} \end{cases}$$

したがって、

$$L_A = 2 \cdot 1.96\sqrt{\frac{20^2}{96}} = \cdots = 8.0$$

同様に、Bについても、$L_B = 2 \cdot z_{0.025}\sqrt{\dfrac{\sigma^2}{n_B}}$ が成り立つので、

$$\frac{L_A}{L_B} = \sqrt{\frac{n_B}{n_A}} = \sqrt{\frac{104}{96}} = 1.04$$ 。

[5]

$L_A = 2z_{0.025}\sqrt{\frac{\sigma^2}{n_A}}$ と求めたが、n_A を改めて確率変数として扱うので、L_A も確率変数である。問題の条件を式に書き下す。

$$P(L_A \le 8) \ge 0.8 \quad — \quad (*)$$

これを n について解けば良い。

$$(*) \iff P\left(2z_\alpha\sqrt{\frac{\sigma^2}{n_A}} \le 8\right) \ge 0.8$$

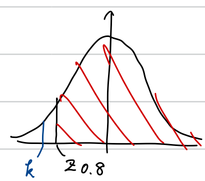
$$P\left\{n_A \ge \left(\frac{2z_\alpha\sqrt{\sigma^2}}{8}\right)^2\right\} \ge 0.8$$

↑ n_A は二項分布だが、[4] より n, n_A 共に大きいことから、正規分布の近似を使うことを考える。

$$P\left\{\frac{n_A - \frac{n}{2}}{\sqrt{n\frac{1}{4}}} \ge \frac{\left(\frac{2z_\alpha\sqrt{\sigma^2}}{8}\right)^2 - \frac{n}{2}}{\sqrt{\frac{n}{4}}}\right\} \ge 0.8$$

($\because$ $n_A \sim Bin(n, \frac{1}{2})$)

$$P\left\{z \ge \underbrace{\frac{\left(\frac{2z_\alpha\sqrt{\sigma^2}}{8}\right)^2 - \frac{n}{2}}{\sqrt{\frac{n}{4}}}}_{=k}\right\} \ge 0.8$$



これより、

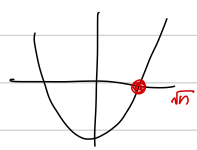
$$\frac{\left(\frac{2z_\alpha\sqrt{\sigma^2}}{8}\right)^2 - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \le z_{0.8}$$

$$\iff \left(\frac{2\cdot 1.96\cdot 20}{8}\right)^2 \le -0.84\cdot\frac{\sqrt{n}}{2} + \frac{n}{2}$$

$$\iff n - 0.84\sqrt{n} - (96.04)\cdot 2 \ge 0$$

$$\iff n - 0.84\sqrt{n} - 192 \ge 0$$

$\sqrt{n} = x$ としてこれを解き、正の解を得ると、



$$\sqrt{n} = x^+ = \frac{0.84 + \sqrt{(0.84)^2 + 4\cdot 192}}{2}$$

$$\fallingdotseq 14.28$$

よって、 $n \geq (14.28)^2$

$= 203.9$

n は 204 以上であれば良い。