一級 2017 偶数

問2

確率密度関数のパラメータの最良の推定量を選択することを題材とした問題。真面目に期待値や分散を計算する能力と、 max を含む順序統計量を扱う能力が問われる。

[1]

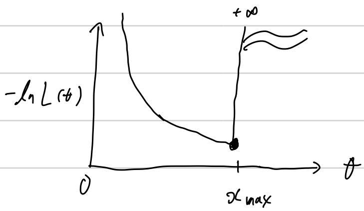
$X_1, X_2, \cdots, X_n$ の実現値 $x_1, x_2, \cdots, x_n$ のうち、最大値を x_{max} とする。
負の対数尤度 $-\ln L(\theta)$ を最小にするパラメータ $\hat{\theta}$ を求める。

$$
\begin{aligned}
-\ln L(\theta) &= -\ln \prod_{i=1}^{n} f(x_i;\theta) \\
&= \sum_{i=1}^{n} -\ln f(x_i;\theta) \\
&= \begin{cases} \sum_{i=1}^{n} \ln\theta & (x_{max} \leq \theta) \\ \ln\theta + \cdots + \underset{\sim\sim}{\infty} + \cdots + \ln\theta & (0 < \theta < x_{max}) \end{cases}
\end{aligned}
$$

少なくとも x_{max} において
$-\ln 0 = +\infty$ となる。

$$
= \begin{cases} n\ln\theta & (x_{max} \leq \theta) \\ \infty & (0 < \theta < x_{max}) \end{cases}
$$

これを最小にするのは、右図より $\theta = x_{max}$ のときである。



[2]

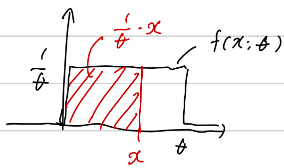
$E(\theta') = \theta$ を示す。

$$E(\theta') = E\left[\frac{2}{n}\sum_{i=1}^{n} X_i\right]$$
$$= \frac{2}{n}\sum_{i=1}^{n} E[X_i]$$
$$= \frac{2}{n}\sum_{i=1}^{n} \cdot \frac{\theta}{2} \qquad (\because 一様分布の平均)$$
$$= \frac{2}{n} \cdot n \cdot \frac{\theta}{2}$$
$$= \theta \qquad \square$$

[3]

X_{max} の累積分布関数 $F(x)$ を求め，それを微分して確率密度関数を得る。

$$F(x) = P(X_{max} \leq x)$$
$$= P(\max(X_1, X_2, \cdots, X_n) \leq x)$$
$$= P(X_1 \leq x, X_2 \leq x, \cdots, X_n \leq x)$$
$$= P(X_1 \leq x)P(X_2 \leq x) \cdots P(X_n \leq x) \qquad (\because 独立性)$$
$$= \prod_{i=1}^{n} P(X_i \leq x)$$
$$= \begin{cases} 1 & \theta < x \\ \underbrace{\left(\frac{1}{\theta} \cdot x\right)}_{右図}{}^{n} & 0 \leq x \leq \theta \\ 0 & x < 0 \end{cases}$$



確率密度関数 $g(x)$ は，

$$g(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & (x < 0 \text{ または } x > \theta) \\ \frac{n}{\theta}\left(\frac{1}{\theta}x\right)^{n-1} & (0 \leq x \leq \theta) \end{cases}$$

θ'' の不偏性は次の式より導かれる。

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

[4]

二つの推定量が共に不偏推定量であるので、分散の小さい推定量がより望ましい推定量となる。

ここまでで X_i と X_{max} のp.d.f. がわかっているので、を適用

すれば積分式より分散が求まることが見通せる。

$$V[\theta'] = \left(\frac{2}{n}\right)^2 V\left[\sum_{i=1}^{n} X_i\right]$$
$$= \left(\frac{2}{n}\right)^2 \sum_{i=1}^{n} V[X_i] \qquad (\because \text{独立性})$$
$$= \left(\frac{2}{n}\right)^2 \cdot n \cdot E[(X_i - E(X_i))^2]$$
$$= \frac{4}{n} \cdot \int_0^{\theta} \left(x - \frac{\theta}{2}\right)^2 \cdot \frac{1}{\theta}\, dx$$
$$= (\text{略})$$
$$= \frac{\theta^2}{3n}$$

$$V[\theta''] = \left(\frac{n+1}{n}\right)^2 V[X_{max}]$$
$$= \left(\frac{n+1}{n}\right)^2 \left\{E[X_{max}^2] - E[X_{max}]^2\right\}$$
$$= \left(\frac{n+1}{n}\right)^2 \left\{\int_0^{\theta} x^2 \cdot \frac{n}{\theta}\left(\frac{1}{\theta}x\right)^{n-1} dx - \left(\frac{n}{n+1}\theta\right)^2\right\}$$

$\because \frac{n+1}{n} E(X_{max}) = \theta$

$$= \left(\frac{n+1}{n}\right)^2 \left\{\frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2\right\}$$
$$= (\text{略})$$
$$= \frac{\theta^2}{n(n+2)}$$

以上より、$n > 1$ で $V(\theta'') < V(\theta')$ がわかり、θ''の方が望ましいと言える。

問4

複数の確率変数のうち一つが定まった時の条件付き分布を求める問題。

どの変数が定まるかによって、求め方が異なる点に注意する。

[1] 解答参照

[2]

相関係数の定義通り求める。 $Cov(X, Y+Z) = Cov(X, Y) + Cov(X, Z)$ を使うと楽になる。

$$
\begin{aligned}
C &= \frac{Cov[X, Z]}{\sqrt{V[X]V[Z]}} = \frac{1}{\sqrt{1\cdot(k^2+1)}} Cov(X, a+kX+Y) \\
&= \frac{1}{\sqrt{k^2+1}}\left\{k\, Cov(X, X) + Cov(X, Y)\right\} \\
&= \frac{1}{\sqrt{k^2+1}}\left\{k + 0\right\} \\
&= \frac{k}{\sqrt{k^2+1}} .
\end{aligned}
$$

[3]

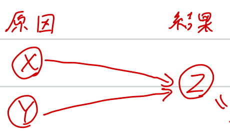
$X = a$ を与えたとき、$Z = a + kx + Y$ より、$Z \sim N(a+kx, 1)$

[4]

[3]と同じように、$X = \frac{1}{k}(z - a - Y)$ として解こうとすると間違える。

なぜだめなのかの明確な答えがわかっていないが、「原因」と「結果」のうち、

「結果」が得られたときの「原因」の分布は、ベイズの公式を使う必要があるみたい。



$= z$ として得られている。

条件付き分布 $f(x|z)$ を求めるために、その対数を考える。

（PRML 上巻 P90，(2.113)-(2.117) の結果を導くこととなる）

$$
\begin{aligned}
\log f(x|z) &= \log \frac{f(z|x)\,f(x)}{f(z)} \\
&= \log f(z|x) + \log f(x) + \text{const.} \\
&= -\frac{1}{2}\{z-(a+kx)\}^2 - \frac{1}{2}x^2 + \text{const.} \\
&= -\frac{1}{2}\{k^2x^2 - 2k(z-a)x + x^2\} + \text{const.} \\
&= -\frac{1}{2}\{(k^2+1)x^2 - 2k(z-a)x\} + \text{const.} \\
&= -\frac{1}{2/k^2+1}\left\{x^2 - \frac{2k(z-a)}{k^2+1}x\right\} + \text{const.} \\
&= -\frac{1}{2/k^2+1}\left\{x - \frac{k(z-a)}{k^2+1}\right\}^2 + \text{const.} \\
&= \log N\left(\frac{k(z-a)}{k^2+1}, \frac{1}{k^2+1}\right)
\end{aligned}
$$

よって、$N\left(\frac{k(z-a)}{k^2+1}, \frac{1}{k^2+1}\right)$ に従う。