The Hong Kong University of Science & Technology

MATH243 – Statistical Inference

Midterm Examination – Fall 02/03

Answer ALL questions | **Date: 1 November 2002 (Fri)**

All Equal Marks | **Time allowed: 2 Hours**

---

1. Let $X_1, X_2, \ldots, X_n$ denote a random sample from $B_i(1, p)$. $f_{X_i}(x_i, p) = p^{x_i}(1-p)^{1-x_i}$

   (a) Find the maximum likelihood estimator of p. $\bar{X}$

   (b) Is the estimator in (a) an unbiased estimator of p? Yes, $E(\bar{X}) = p$

   (c) Find Cramer Rao lower bound for the variance of an unbiased estimator for p. $\frac{p(1-p)}{n}$

   (d) Find $Var(\bar{X})$. $\frac{p(1-p)}{n}$

   (e) Find $E(\bar{X}(1-\bar{X}))$. Then, find the value of c so that $c\bar{X}(1-\bar{X})$ is an unbiased estimator of $Var(\bar{X})$. $E(\bar{X}^2) = \frac{p(1-p)}{n} + p^2 \Rightarrow E(\bar{X}(1-\bar{X})) = p(1-p)\frac{n-1}{n} \Rightarrow c = \frac{1}{n-1}$

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$. $f_{X_i}(x_i, \theta) = 1 \quad \theta - \frac{1}{2} \le X_i \le \theta + \frac{1}{2}$

   (a) Find the method of moments estimator for θ. $\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

   (b) Is the estimator in (a) an unbiased estimator for θ? Yes, $E(\bar{X}) = \theta$

   (c) Find the variance of the estimator in (a). $Var(\bar{X}) = \frac{1}{12n}$

   (d) Prove that the mid-range $Y = \frac{1}{2}(Y_1 + Y_n)$ where $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$ is an unbiased estimator for θ.

   (e) **OPTIONAL (4 marks)** Find the variance of the estimator Y in (d). Is the variance in (c) greater?

   Hint: $f_{Y_1,Y_n}(y_1,y_n) = n(n-1)(y_n - y_1)^{n-2}$ for $\theta - 1/2 \le y_1 \le y_n \le \theta + 1/2$

e.g. Let $(X_1, ..., X_n)$ be a r.s. from the continuous uniform distribution in the interval $[\theta-\frac{1}{2}, \theta+\frac{1}{2}]$, then

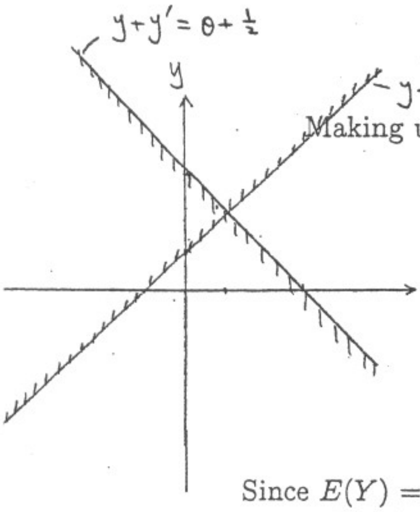
$E(x) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} x\,dx = \frac{1}{2} x^2 \Big|_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} = \theta$

$E(X_i) = \theta,\ Var(X_i) = \frac{1}{12} \quad (i = 1, ..., n)$ $\quad Var(x) = \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} x^2\,dx - \theta^2 = \frac{1}{3} x^3 \Big|_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} - \theta^2 = \frac{1}{12}$

$\Rightarrow\ E(\bar{X}) = \theta,\ Var(\bar{X}) = \frac{1}{12n}$ $\quad F(x) = x - (\theta - \frac{1}{2})$

If instead, we use the sample mid-range $Y = \frac{1}{2}(Y_1 + Y_n)$ to estimate θ, where $Y_1 = \min(X_i)$

$Y_n = \max(X_i)$ $\quad f_{Y_\alpha, Y_\beta}(x, y) = \frac{n!}{(\alpha-1)!(\beta-\alpha-1)!(n-\beta)!} [F(x)]^{\alpha-1} [F(y) - F(x)]^{\beta-\alpha-1} [1 - F(y)]^{n-\beta} f(x) f(y)$



$$f(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2} \qquad \theta - \frac{1}{2} \le y_1 \le y_n \le \theta + \frac{1}{2}$$

$\theta-\frac{1}{2} \le y-y' \le y+y' \le \theta+\frac{1}{2}$

$\Rightarrow y' \le y - (\theta - \frac{1}{2}),\ y' \le (\theta + \frac{1}{2}) - y$

$\Rightarrow y - y' \ge \theta - \frac{1}{2},\ y + y' \le \theta + \frac{1}{2}$

Making use of the transformation

$$(Y_1, Y_n) \longrightarrow \frac{1}{2}(Y_n + Y_1) = y \Rightarrow y_1 = y - y', \quad y_n = y + y' \qquad \frac{\partial(y_1, y_n)}{\partial(y, y')} = 2$$

$$(Y_1, Y_n) \longrightarrow \frac{1}{2}(Y_n - Y_1) = y'$$

$\theta-\frac{1}{2} \le y-y' \le y+y' \le \theta+\frac{1}{2}$

$f(y, y') = 2^{n-1} n (n-1) (y')^{n-2}$

$$\text{Thus,}\quad f(y) = \begin{cases} n2^{n-1}(y - \theta + \frac{1}{2})^{n-1} & y \in [\theta - \frac{1}{2}, \theta] \\ n2^{n-1}(\theta + \frac{1}{2} - y)^{n-1} & y \in [\theta, \theta + \frac{1}{2}] \end{cases}$$

$f(y) = \int_0^{y-(\theta-\frac{1}{2})} f(y, y')\,dy' \quad y \in [\theta - \frac{1}{2}, \theta]$

$= \int_0^{(\theta+\frac{1}{2})-y} f(y, y')\,dy' \quad y \in [\theta, \theta + \frac{1}{2}]$

Since $E(Y) = \theta$, $Y = \frac{1}{2}(Y_1 + Y_n)$ is unbiased.

$E(y) = n2^{n-1} \int_{\theta-\frac{1}{2}}^{\theta} y (y - \theta + \frac{1}{2})^{n-1} dy + n2^{n-1} \int_{\theta}^{\theta+\frac{1}{2}} y (\theta + \frac{1}{2} - y)^{n-1} dy$

$= 2^{n-1} [y (y - \theta + \frac{1}{2})^n \Big|_{\theta-\frac{1}{2}}^{\theta} - \int_{\theta-\frac{1}{2}}^{\theta} (y - \theta + \frac{1}{2})^n dy - y (y - \theta + \frac{1}{2})^n \Big|_{\theta}^{\theta+\frac{1}{2}} + \int_{\theta}^{\theta+\frac{1}{2}} (\theta + \frac{1}{2} - y)^{n-1} dy] = \theta$

$$\begin{aligned} Var(Y) &= E((Y-\theta)^2) \\ &= E(W^2) \end{aligned}$$

$$\text{p.d.f. of } W \text{ is} = \begin{cases} n2^{n-1}(w + \frac{1}{2})^{n-1} & w \in [-\frac{1}{2}, 0] \\ n2^{n-1}(\frac{1}{2} - w)^{n-1} & w \in [0, \frac{1}{2}] \end{cases}$$

$$\begin{aligned} \therefore\ Var(Y) &= \int_{-\frac{1}{2}}^{0} w^2 f(w)\,dw + \int_0^{1/2} w^2 f(w)\,dw \\ &= \frac{1}{2(n+1)(n+2)} \end{aligned}$$

$\lim_{n\to\infty} Var(Y) = 0 \quad \Rightarrow Y$ is consistent

Efficiency: $\dfrac{1}{2(n+1)(n+2)} < \dfrac{1}{12n}$ $\quad$ for $n > 2$

$\Rightarrow$ Y is more efficient.

Note that Y involves only min (X_i) and max (X_i) and ignores immediate sample value.

$Y_1 : n(\theta + \frac{1}{2} - y_1)^{n-1}$

$\theta - \frac{1}{2} + \frac{1}{n+1}$

$\shortparallel$

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

$Y_n : n(y_n - \theta + \frac{1}{2})^{n-1}$

$\theta + \frac{1}{2} - \frac{1}{n+1}$

$\shortparallel$

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

$Y_1 : (\theta - \frac{1}{2})^2 + \frac{2}{n+1}(\theta - \frac{1}{2}) + \frac{2}{(n+1)(n+2)}$

$\Downarrow$

$\frac{n}{(n+1)(n+2)}$

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

$\Downarrow$

$\frac{n}{(n+1)(n+2)}$

3. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Define

$$\overline{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i ,$$

$$\overline{X}_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i ,$$

and

$$S_k^2 = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - \overline{X}_k)^2 ,$$

$$S_{n-k}^2 = \frac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \overline{X}_{n-k})^2 .$$

Answer the following:

(a) What is the distribution of $\left((k-1)S_k^2 + (n-k-1)S_{n-k}^2\right)/\sigma^2$? $\frac{(k-1)S_k^2}{\sigma^2} \sim \chi^2_{k-1}$, $\frac{(n-k-1)S_{n-k}^2}{\sigma^2} \sim \chi^2_{n-k-1}$ $\Rightarrow \sim \chi^2_{n-2}$

(b) What is the distribution of $\left(\overline{X}_k + \overline{X}_{n-k}\right)/2$? $\sim N\left(\mu, \frac{1}{4}\left(\frac{\sigma^2}{k} + \frac{\sigma^2}{n-k}\right)\right)$

(c) What is the distribution of $\sigma^{-2}(X_i - \mu)^2$? $\sim \chi^2_1$

(d) What is the distribution of S_k^2 / S_{n-k}^2 ? $\sim F_{k-1,\, n-k-1}$

(e) What is the distribution of $(\overline{X}_n - \mu)/(S_n/\sqrt{n})$? $\sim t_{n-1}$

If $\mu = 0$ and $\sigma^2 = 1$,

(f) What is the distribution of $k\overline{X}_k^2 + (n-k)\overline{X}_{n-k}^2$? $k\overline{X}_k^2 \sim \chi^2_1$, $(n-k)\overline{X}_{n-k}^2 \sim \chi^2_1$ $\Rightarrow \sim \chi^2_2$

(g) What is the distribution of X_1^2 / X_2^2 ? $\sim F_{1,1}$

(h) What is the distribution of X_1 / X_n ? $\sim t_1$

(i) What is the distribution of $(X_2 + X_1)^2 / (X_2 - X_1)^2$? $\frac{(X_2+X_1)^2}{2} \sim \chi^2_1$, $\frac{(X_2-X_1)^2}{2} \sim \chi^2_1$,

$Cov(X_2 + X_1, X_2 - X_1) = Var(X_2) - Var(X_1) =$

Note: Write your answer as: $S_n^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$.

$\Rightarrow \sim F_{1,1}$

****** **END** ******