STAT 410

Practice Problems 12

Fall 2020
A. Stepanov

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with probability mass function

$$P(X_i = 1) = \frac{\theta}{3+\theta}, \quad P(X_i = 2) = \frac{2}{3+\theta}, \quad P(X_i = 3) = \frac{1}{3+\theta}, \qquad \theta > 0.$$

a) Find a sufficient statistic for θ.

b) Obtain the method of moments estimator $\tilde{\theta}$ of θ.

c) Obtain the maximum likelihood estimator $\hat{\theta}$ of θ.

2. Let $\lambda > 0$ and let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a double exponential distribution. That is,

$$f(x;\lambda) = \frac{\lambda}{2} e^{-\lambda|x|}, \qquad -\infty < x < \infty.$$

a) Obtain the maximum likelihood estimator of λ, $\hat{\lambda}$.

b) Find a closed-form expression for $E(X^k)$ for positive integer k.

3. Let $\theta \in \mathbb{R}$ and let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with probability density function

$$f(x;\theta) = \frac{1}{2} e^{-|x-\theta|}, \qquad x \in \mathbb{R}.$$

a) Find a method of moments estimator $\tilde{\theta}$ of θ.

b) Find the maximum likelihood estimator $\hat{\theta}$ of θ.

4. a) Let $X_1, X_2, \ldots$ be i.i.d. with p.d.f. $f(x) = \frac{1}{3} e^{-x/3}$, $x > 0$.

Prove (show) that $P(X_1 + X_2 + \ldots + X_n > n) \to 1$ as $n \to \infty$.

b) Let $Y_1, Y_2, \ldots$ be i.i.d. with p.d.f. $f(y) = 3e^{-3y}$, $y > 0$.

Prove (show) that $P(Y_1 + Y_2 + \ldots + Y_n > n) \to 0$ as $n \to \infty$.

Hint: Use the WLLN.

5. **4.4.31** (7th edition) **5.4.13** (6th edition)

Let $Y_1 < Y_2 < \ldots < Y_n$ denote the order statistics of a random sample of size n from a distribution that has p.d.f.

$$f_X(x;\theta) = \frac{3}{\theta^3} \cdot x^2, \qquad 0 < x < \theta, \qquad \text{zero elsewhere.}$$

a) Show that $P\left(c < \frac{Y_n}{\theta} < 1 \right) = 1 - c^{3n}$, where $0 < c < 1$.

b) If $n = 4$ and if the observed value of Y_4 is 2.3, what is a 95% confidence interval for θ?

6. **6.2.9** (7th and 6th edition)

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with pdf

$$f(x;\theta) = \frac{3\theta^3}{(x+\theta)^4}, \qquad 0 < x < \infty, \qquad 0 < \theta < \infty.$$

Show that $Y = 2\overline{X}$ is an unbiased estimator of θ and determine its efficiency.

7.* **6.2.2** (7th and 6th edition)

Given $f(x;\theta) = \frac{1}{\theta}$, $0 < x < \theta$, zero elsewhere, with $\theta > 0$, formally compute the reciprocal of

$$n\,\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f(\mathrm{X};\theta)\right]^2\right\}.$$

Compare this with the variance of $\frac{n+1}{n}\cdot \mathrm{Y}_n$, where Y_n is the largest observation of a random sample of size n from this distribution. Comment.

8. Bert and Ernie find a coin on the sidewalk on Sesame Street. They wish to estimate p, the probability of Heads. Bert got X Heads in N coin tosses (N is fixed, X is random). Ernie got Heads for the first time on the Y^{th} coin toss (Y is random). They decide to combine their information in hope of a better estimate. (Assume independence.)

a) What is the likelihood function $\mathrm{L}(p) = \mathrm{L}(p;\mathrm{X},\mathrm{N},\mathrm{Y})$?

b) Obtain the maximum likelihood estimator for p.

c) Explain intuitively why your estimator makes good sense.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with probability mass function

$$P(X_i = 1) = \frac{\theta}{3+\theta}, \quad P(X_i = 2) = \frac{2}{3+\theta}, \quad P(X_i = 3) = \frac{1}{3+\theta}, \qquad \theta > 0.$$

a) Find a sufficient statistic for θ.

$$f(x_1;\theta)\, f(x_2;\theta)\, \ldots\, f(x_n;\theta) = \frac{1}{(3+\theta)^n} \cdot \theta^{(\text{\# of 1's})} \cdot 2^{(\text{\# of 2's})} \cdot 1^{(\text{\# of 3's})}.$$

$\Rightarrow$ $Y = (\text{\# of 1's})$ is a sufficient statistic for θ.

b) Obtain the method of moments estimator $\tilde{\theta}$ of θ.

$$E(X) = 1 \times \frac{\theta}{3+\theta} + 2 \times \frac{2}{3+\theta} + 3 \times \frac{1}{3+\theta} = \frac{\theta+7}{3+\theta}.$$

$$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{x} = \frac{\tilde{\theta}+7}{3+\tilde{\theta}}. \qquad 3\bar{x} + \tilde{\theta}\,\bar{x} = \tilde{\theta} + 7.$$

$$\Rightarrow \quad \tilde{\theta} = \frac{7-3\bar{x}}{\bar{x}-1}.$$

c) Obtain the maximum likelihood estimator $\hat{\theta}$ of θ.

$$L(\theta) = \frac{1}{(3+\theta)^n} \cdot \theta^{(\text{\# of 1's})} \cdot 2^{(\text{\# of 2's})} \cdot 1^{(\text{\# of 3's})}.$$

$$\ln L(\theta) = -n \ln(3+\theta) + (\text{\# of 1's}) \ln(\theta) + (\text{\# of 2's}) \ln(2) + (\text{\# of 3's}) \ln(1).$$

$$(\ln L(\theta))' = -\frac{n}{3+\theta} + \frac{(\text{\# of 1's})}{\theta} = 0 \qquad \Rightarrow \qquad \hat{\theta} = \frac{3 \cdot (\text{\# of 1's})}{n - (\text{\# of 1's})}.$$

2. Let $\lambda > 0$ and let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a double exponential distribution. That is,

$$f(x;\lambda) = \frac{\lambda}{2}\, e^{-\lambda|x|}, \qquad -\infty < x < \infty.$$

a) Obtain the maximum likelihood estimator of λ, $\hat{\lambda}$.

$$L(\lambda) = \frac{\lambda^n}{2^n} \exp\left\{-\lambda \cdot \sum_{i=1}^{n} |x_i|\right\}. \qquad \ln L(\lambda) = n \ln\lambda - n\ln 2 - \lambda \cdot \sum_{i=1}^{n} |x_i|.$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} |x_i| = 0. \qquad \Rightarrow \qquad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} |X_i|}.$$

b) Find a closed-form expression for $E(X^k)$ for positive integer k.

$$E(X^k) = \int_{-\infty}^{\infty} x^k \cdot \frac{\lambda}{2}\, e^{-\lambda|x|}\, dx = \int_{-\infty}^{0} x^k \cdot \frac{\lambda}{2}\, e^{\lambda x}\, dx + \int_{0}^{\infty} x^k \cdot \frac{\lambda}{2}\, e^{-\lambda x}\, dx = \ldots$$

k odd $\qquad \ldots = 0.$

k even $\qquad \ldots = 2 \cdot \int_{0}^{\infty} x^k \cdot \frac{\lambda}{2}\, e^{-\lambda x}\, dx = \int_{0}^{\infty} \lambda \cdot x^k\, e^{-\lambda x}\, dx$

$$= \frac{\Gamma(k+1)}{\lambda^k} \cdot \int_{0}^{\infty} \frac{\lambda^{k+1}}{\Gamma(k+1)}\, x^{k+1-1}\, e^{-\lambda x}\, dx = \frac{\Gamma(k+1)}{\lambda^k} = \frac{k!}{\lambda^k}.$$

3. Let $\theta \in \mathbb{R}$ and let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with probability density function

$$f(x;\theta) = \frac{1}{2} e^{-|x-\theta|}, \qquad x \in \mathbb{R}.$$

a) Find a method of moments estimator $\tilde{\theta}$ of θ.

$f(x;\theta)$ is symmetric about θ.

$\Rightarrow \quad E(X) = \theta \quad$ (balancing point) $\qquad \tilde{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) Find the maximum likelihood estimator $\hat{\theta}$ of θ.

$$L(\theta) = \frac{1}{2^n}\exp\left\{-\sum_{i=1}^{n}|x_i - \theta|\right\}.$$

$\Rightarrow \quad$ To maximize $L(\theta)$, we need to minimize $\sum_{i=1}^{n}|x_i - \theta|$.

Let $Y_1 < Y_2 < \ldots < Y_n$ denote the corresponding order statistics.

If $\theta \in (y_k, y_{k+1})$, $\qquad \frac{d}{d\theta}\sum_{i=1}^{n}|x_i - \theta| = k - (n-k) = 2k - n,$

$$\frac{d}{d\theta}\sum_{i=1}^{n}|x_i - \theta| < 0 \text{ if } k < \frac{n}{2}, \qquad \frac{d}{d\theta}\sum_{i=1}^{n}|x_i - \theta| > 0 \text{ if } k > \frac{n}{2}.$$

If n is odd, $\hat{\theta} = Y_{\frac{n+1}{2}}$ (the middle value in the data set).

If n is even, $\hat{\theta} \in [Y_{\frac{n}{2}}, Y_{\frac{n}{2}+1}]$ (any value between the middle two).

For example, $\hat{\theta}$ = sample median.

4. a) Let $X_1, X_2, \ldots$ be i.i.d. with p.d.f. $f(x) = \frac{1}{3} e^{-x/3}, \quad x > 0.$

Prove (show) that $P(X_1 + X_2 + \ldots + X_n > n) \to 1$ as $n \to \infty.$

Hint: Use the WLLN.

By WLLN, $\overline{X} \xrightarrow{P} \mu = 3.$

$1 \geq P(X_1 + X_2 + \ldots + X_n > n) = P(\overline{X} > 1) \geq P(|\overline{X} - 3| < 0.1).$

Since $\overline{X} \xrightarrow{P} 3, \qquad P(|\overline{X} - 3| < 0.1) \to 1$ as $n \to \infty.$

$\Rightarrow \qquad P(X_1 + X_2 + \ldots + X_n > n) \to 1$ as $n \to \infty.$

Note: $\varepsilon = 0.1$ is chosen arbitrarily. Any $0 < \varepsilon < 2$ would have worked.

b) Let $Y_1, Y_2, \ldots$ be i.i.d. with p.d.f. $f(y) = 3e^{-3y}, \quad y > 0.$

Prove (show) that $P(Y_1 + Y_2 + \ldots + Y_n > n) \to 0$ as $n \to \infty.$

By WLLN, $\overline{Y} \xrightarrow{P} \mu = \frac{1}{3}.$

$0 \leq P(Y_1 + Y_2 + \ldots + Y_n > n) = P(\overline{Y} > 1) \leq P(|\overline{Y} - \frac{1}{3}| \geq 0.1).$

Since $\overline{Y} \xrightarrow{P} \frac{1}{3}, \qquad P(|\overline{Y} - \frac{1}{3}| \geq 0.1) \to 0$ as $n \to \infty.$

$\Rightarrow \qquad P(Y_1 + Y_2 + \ldots + Y_n > n) \to 0$ as $n \to \infty.$

Note: $\varepsilon = 0.1$ is chosen arbitrarily. Any $0 < \varepsilon < \frac{2}{3}$ would have worked.

5. **4.4.31** (7th edition) **5.4.13** (6th edition)

Let $Y_1 < Y_2 < \ldots < Y_n$ denote the order statistics of a random sample of size n from a distribution that has p.d.f.

$$f_X(x;\theta) = \frac{3}{\theta^3} \cdot x^2, \qquad 0 < x < \theta, \qquad \text{zero elsewhere.}$$

a) Show that $P\left(c < \frac{Y_n}{\theta} < 1 \right) = 1 - c^{3n}$, where $0 < c < 1$.

$F(x) = (x/\theta)^3, \quad 0 < x < \theta.$ $\qquad F_{Y_n}(x) = (x/\theta)^{3n}, \quad 0 < x < \theta.$

$$P\left(c < \frac{Y_n}{\theta} < 1 \right) = P(c\theta < Y_n < \theta) = F_{Y_n}(\theta) - F_{Y_n}(c\theta) = 1 - c^{3n}.$$

b) If $n = 4$ and if the observed value of Y_4 is 2.3, what is a 95% confidence interval for θ?

$$1 - c^{3n} = P\left(c < \frac{Y_n}{\theta} < 1 \right) = 0.95. \qquad \Rightarrow \qquad c = 0.05^{1/3n}.$$

$$n = 4 \qquad \Rightarrow \qquad c = 0.05^{1/12} \approx 0.7791.$$

$$P\left(c < \frac{Y_n}{\theta} < 1 \right) = 0.95 \qquad \Rightarrow \qquad P\left(Y_n < \theta < \frac{Y_n}{c} \right) = 0.95$$

95% confidence interval for θ: $\left(Y_n, \frac{Y_n}{c} \right)$

$n = 4, \; Y_4 = 2.3$

$\Rightarrow$ 95% confidence interval for θ: $\mathbf{(2.3, 2.9522)}$

6. **6.2.9** (7th and 6th edition)

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with pdf

$$f(x;\theta) = \frac{3\theta^3}{(x+\theta)^4}, \qquad 0 < x < \infty, \qquad 0 < \theta < \infty.$$

Show that $Y = 2\overline{X}$ is an unbiased estimator of θ and determine its efficiency.

$$\mu = E(X) = \int_0^\infty x \cdot \frac{3\theta^3}{(x+\theta)^4}\, dx = \int_\theta^\infty (y-\theta) \cdot \frac{3\theta^3}{y^4}\, dy = \int_\theta^\infty \left(\frac{3\theta^3}{y^3} - \frac{3\theta^4}{y^4} \right) dy$$

$$= \left(-\frac{3\theta^3}{2y^2} + \frac{3\theta^4}{3y^3} \right) \Bigg|_\theta^\infty = \frac{3}{2}\theta - \theta = \frac{\theta}{2}.$$

$$E(Y) = E(2\overline{X}) = 2\mu = \theta.$$

$$E(X^2) = \int_0^\infty x^2 \cdot \frac{3\theta^3}{(x+\theta)^4}\, dx = \int_\theta^\infty (y-\theta)^2 \cdot \frac{3\theta^3}{y^4}\, dy = \int_\theta^\infty \left(\frac{3\theta^3}{y^2} - \frac{6\theta^4}{y^3} + \frac{3\theta^5}{y^4} \right) dy$$

$$= \left(-\frac{3\theta^3}{y} + \frac{6\theta^4}{2y^2} - \frac{3\theta^5}{3y^3} \right) \Bigg|_\theta^\infty = \theta^2.$$

$$\sigma^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{3\theta^2}{4}.$$

$$\mathrm{Var}(Y) = \mathrm{Var}(2\overline{X}) = 4\,\frac{\sigma^2}{n} = \frac{3\theta^2}{n}.$$

$$\ln f(x;\theta) = \ln 3 + 3\ln\theta - 4\ln(x+\theta)$$

$$\frac{\partial}{\partial\theta} \ln f(x;\theta) = \frac{3}{\theta} - \frac{4}{x+\theta}$$

$$\frac{\partial^2}{\partial\theta^2} \ln f(x;\theta) = -\frac{3}{\theta^2} + \frac{4}{(x+\theta)^2}$$

$$\mathrm{E}\left[\frac{4}{(x+\theta)^2}\right] = \int_0^\infty \frac{12\,\theta^3}{(x+\theta)^6}\,dx = \frac{2.4}{\theta^2}.$$

$$\mathrm{I}(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} \ln f(x;\theta)\right] = \frac{0.6}{\theta^2}.$$

Rao-Cramer lower bound for Y: $\dfrac{1}{n\,\mathrm{I}(\theta)} = \dfrac{\theta^2}{0.6\,n}.$

Y is NOT an efficient estimator of θ.

$$(\text{efficiency of Y}) = \frac{\dfrac{\theta^2}{0.6\,n}}{\dfrac{3\,\theta^2}{n}} = \frac{1}{1.8} = \frac{5}{9}.$$

7.* **6.2.2** (7th and 6th edition)

Given $f(x;\theta) = \frac{1}{\theta}$, $0 < x < \theta$, zero elsewhere, with $\theta > 0$, formally compute the reciprocal of

$$n\,\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f(\mathrm{X};\theta)\right]^2\right\}.$$

Compare this with the variance of $\frac{n+1}{n}\cdot \mathrm{Y}_n$, where Y_n is the largest observation of a random sample of size n from this distribution. Comment.

$f(x;\theta) = \frac{1}{\theta}$ $\qquad$ $\ln f(x;\theta) = -\ln\theta$ $\qquad$ $\frac{\partial}{\partial\theta}\ln f(x;\theta) = -\frac{1}{\theta}$

$$\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\mathrm{X};\theta)\right)^2\right] = \mathrm{E}\left[\left(-\frac{1}{\theta}\right)^2\right] = \frac{1}{\theta^2}.$$

Recall (Examples for 07/12/2019 (2)):

$$\mathrm{Var}\left(\frac{n+1}{n}\max \mathrm{X}_i\right) = \left(\frac{n+1}{n}\right)^2\cdot \mathrm{Var}\left(\max \mathrm{X}_i\right) = \frac{\theta^2}{(n+2)\cdot n}.$$

$\mathrm{Var}\left(\frac{n+1}{n}\max \mathrm{X}_i\right) < \frac{\theta^2}{n}$ $\qquad$ Rao-Cramér Lower Bound does NOT hold.

The p.d.f.s do NOT have common support for all θ.

8. Bert and Ernie find a coin on the sidewalk on Sesame Street. They wish to estimate p, the probability of Heads. Bert got X Heads in N coin tosses (N is fixed, X is random). Ernie got Heads for the first time on the Y^{th} coin toss (Y is random). They decide to combine their information in hope of a better estimate. (Assume independence.)

a) What is the likelihood function $\text{L}(p) = \text{L}(p;\, \text{X}, \text{N}, \text{Y})$?

X has a Binomial(N, p) distribution. Y has a Geometric(p) distribution.

$$\text{L}(p) = \binom{\text{N}}{\text{X}} p^{\text{X}} (1-p)^{\text{N}-\text{X}} \times (1-p)^{\text{Y}-1}\, p = \binom{\text{N}}{\text{X}} p^{\text{X}+1} (1-p)^{\text{N}-\text{X}+\text{Y}-1}.$$

b) Obtain the maximum likelihood estimator for p.

$$\ln \text{L}(p) = \ln \binom{\text{N}}{\text{X}} + (\text{X}+1)\ln p + (\text{N}-\text{X}+\text{Y}-1)\ln(1-p).$$

$$\frac{d}{dp} \ln \text{L}(p) = \frac{\text{X}+1}{p} - \frac{\text{N}-\text{X}+\text{Y}-1}{1-p} = \frac{\text{X}+1-\text{X}p-p-\text{N}p+\text{X}p-\text{Y}p+p}{p(1-p)}$$

$$= \frac{\text{X}+1-\text{N}p-\text{Y}p}{p(1-p)} = 0.$$

$$\Rightarrow \qquad \hat{p} = \frac{\text{X}+1}{\text{N}+\text{Y}}.$$

c) Explain intuitively why your estimator makes good sense.

Bert:	N attempts,	X "successes"
Ernie:	Y attempts,	1 "success"

$$\hat{p} = \frac{\text{X}+1}{\text{N}+\text{Y}} = \frac{\text{total number of "successes"}}{\text{total number of attempts}}.$$