PROF HONG
2021

Probability and Statistics for Economists
Chapter 8 Parameter Estimation and Evaluation

1. One observation is taken on a discrete random variable X with PMF $f(x, \theta)$ given below, where $\theta \in \Theta = \{1, 2, 3\}$. Find the MLE of θ.

x	$f(x,1)$	$f(x,2)$	$f(x,3)$
0	$\frac{1}{3}$	$\frac{1}{4}$	0
1	$\frac{1}{3}$	$\frac{1}{4}$	0
2	0	$\frac{1}{4}$	$\frac{1}{4}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$
4	$\frac{1}{6}$	0	$\frac{1}{4}$

Solution:

$$\widehat{\theta}_{MLE} = \arg\max L(\theta|x) = \arg\max f(x|\theta)$$

At $x = 0$, $\max L(\theta|0) = \max f(0|\theta) = \frac{1}{3}$ when $\theta = 1$. So $\widehat{\theta}_{MLE} = 1$;
At $x = 1$, $\max L(\theta|1) = \max f(1|\theta) = \frac{1}{3}$ when $\theta = 1$. So $\widehat{\theta}_{MLE} = 1$;
At $x = 2$, $\max L(\theta|2) = \max f(2|\theta) = \frac{1}{4}$ when $\theta = 2$ and 3. So $\widehat{\theta}_{MLE} = 2$ and 3;
At $x = 3$, $\max L(\theta|3) = \max f(3|\theta) = \frac{1}{2}$ when $\theta = 3$. So $\widehat{\theta}_{MLE} = 3$;
At $x = 4$, $\max L(\theta|4) = \max f(4|\theta) = \frac{1}{4}$ when $\theta = 3$. So $\widehat{\theta}_{MLE} = 3$.

2. Let $\mathbf{X}^n$ be an IID random sample with one of two PDF's. If $\theta = 0$, then

$$f(x, \theta) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$$

while if $\theta = 1$, then

$$f(x, \theta) = \begin{cases} 1/(2\sqrt{x}), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find the MLE of θ.

Solution:

$L(0|x) = 1$ if $0 < x_i < 1$ for $i = 1, 2, \ldots, n$. $L(1|x) = \prod_{i=1}^{n}(2\sqrt{x_i})^{-1}$ if $0 < x_i < 1$ for $i = 1, 2, \ldots, n$. And we want to maximize $L(\theta|x)$. Hence $\widehat{\theta}_{MLE} = 1$ if $\prod_{i=1}^{n}(2\sqrt{x_i})^{-1} > 1$, $\widehat{\theta}_{MLE} = 0$ if $\prod_{i=1}^{n}(2\sqrt{x_i})^{-1} < 1$, $\widehat{\theta}_{MLE} = 0$ and 1 if $\prod_{i=1}^{n}(2\sqrt{x_i})^{-1} = 1$.

3. Suppose the random variables $\{Y_1, \cdots, Y_n\}$ satisfy

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \cdots, n$$

where $x_1, \cdots, x_n$ are fixed constants, and $\{\varepsilon_i\}$ is an IID sequence from a $N(0, \sigma^2)$ distribution, with σ^2 unknown.

(1) Find a two-dimensional sufficient statistic for (β, σ^2);

(2) Find the MLE of β, and show that it is an unbiased estimator of β;

(3) Find the distribution of the MLE of β.

Solution:

(1)

$$\begin{aligned} L(\theta|y) &= \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2] \\ &= (2\pi\sigma^2)^{-n/2} \exp[-\frac{\beta^2 \sum_i x_i^2}{2\sigma^2}] \exp[-\frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i] \end{aligned}$$

By factorization theorem, $(\sum_i y_i^2, \sum_i x_i y_i)$ is a sufficient statistic for (β, σ^2).

(2)

$$\log L(\beta, \sigma^2|y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i - \frac{\beta^2}{2\sigma^2}\sum_i x_i^2$$

For a fixed value of σ^2,

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= \frac{1}{\sigma^2}\sum_i x_i y_i - \frac{\beta}{\sigma^2}\sum_i x_i^2 = 0 \\ &\Rightarrow \ \widehat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \end{aligned}$$

Also,

$$\frac{\partial^2 \log L}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_i x_i^2 < 0$$

So it is a maximum. Because $\widehat{\beta}$ does not depend on σ^2, it is the MLE. And $\widehat{\beta}$ is unbiased because

$$E\widehat{\beta} = \frac{\sum_i x_i E y_i}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot \beta x_i}{\sum_i x_i^2} = \beta$$

(3) Because $\sum_i x_i^2$ is constant and y_i is normal distribution, $\widehat{\beta}$ is normally distributed with mean β, and

$$Var(\widehat{\beta}) = Var(\frac{\sum_i x_i y_i}{\sum_i x_i^2}) = \sum_i (\frac{x_i}{\sum_i x_i^2})^2 Var(y_i) = \frac{\sum_i x_i^2}{(\sum_i x_i^2)^2}\sigma^2 = \frac{\sigma^2}{\sum_i x_i^2}$$

Thus we have $\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_i x_i^2})$.

4. One observation, X, is taken from a $N(0, \sigma^2)$ population.
(1) Find an unbiased estimation of σ^2;
(2) Find the MLE of σ;
(3) Discuss how the method of moments estimator of σ might be found.

Solution:

(1) $\mathrm{E}[X^2] = Var(X) + \mu^2 = \sigma^2$. Therefore X^2 is an unbiased estimator of σ^2.

(2)

$$L(\sigma|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}x^2]$$

$$\log L(\sigma|x) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{x^2}{2\sigma^2}$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{1}{\sigma} + \frac{x^2}{\sigma^3} = 0 \Rightarrow \widehat{\sigma} = \sqrt{x^2} = |x|$$

$$\frac{\partial^2 \log L}{\partial^2 \sigma} = -\frac{3x^2}{\sigma^4} + \frac{1}{\sigma^2} < 0 \text{ at } \widehat{\sigma} = |x|$$

Thus, $\widehat{\sigma} = |x|$ is a local maximum. Because it is the only place where the first derivative is zero, it is also a global maximum.

(3) Since the first moment $E[X] = 0$ is given, we only need to match the second second moment: $E[X^2] = \sigma^2 = X^2 \Rightarrow \widehat{\sigma} = |X|$.

5. Suppose $f(x, \theta)$ is a PDF model and $f(x, \theta)$ is continuously differentiable with respect to $\theta \in \Theta$, where θ is an interior point in parameter space Θ.

Then for all θ in the interior of Θ,

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(x,\theta)}{\partial \theta} f(x,\theta)dx = 0$$

Solution:

See Lemma 8.3 on page 426 of textbook.

6. [Information Matrix Equality]: Suppose a PDF model $f(x,\theta)$ is twice continuously differentiable with respect to $\theta \in \Theta$, where θ is an interior point in parameter space Θ. Define

$$I(\theta) = \int_{-\infty}^{\infty} \left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2 f(x,\theta)dx$$
$$H(\theta) = \int_{-\infty}^{\infty} \left[\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right] f(x,\theta)dx$$

Then show that for all θ in the interior of Θ,

$$I(\theta) + H(\theta) = 0.$$

Solution:

See Lemma 8.4 on page 427 of textbook.

7. Let $W_1, \ldots, W_k$ be unbiased estimators of a parameter θ with $\text{var}(W_i) = \sigma_i^2$ and $\text{Cov}(W_i, W_j) = 0$ if $i \neq j$

(1) Show that, of all estimators of the form $\sum_{i=1}^k a_i W_i$, where the a_i s are constants and $E_\theta\left(\sum_{i=1}^k a_i W_i\right) = \theta$, the estimator $W^* = \frac{\sum_{i=1}^k W_i/\sigma_i^2}{\sum_{i=1}^k 1/\sigma_i^2}$ has minimum variance.

(2) Show that $\text{Var}(W^*) = \frac{1}{\sum_{i=1}^k 1/\sigma_i^2}$.

Solution:

(1)

$$\min_{a_i} Var(\sum_{i=1}^k a_i W_i) \text{ s.t. } E(\sum_{i=1}^k a_i W_i) = \theta$$

Note that $Var(\sum_{i=1}^k a_i W_i) = \sum_{i=1}^k a_i^2 Var(W_i)$ for W_i are uncorrelated. Also note that $E(\sum_{i=1}^k a_i W_i) = \theta \Leftrightarrow \sum_{i=1}^k a_i = 1$. Therefore, the question becomes

$$\min_{a_i} \sum_{i=1}^k a_i^2 Var(W_i) \text{ s.t. } \sum_{i=1}^k a_i = 1$$

$L = \sum_{i=1}^{k} a_i^2 Var(W_i) + \lambda(1 - \sum_{i=1}^{k} a_i)$
FOC: $\frac{\partial L}{\partial a_i} = 0 \Rightarrow 2\sigma_i^2 a_i^* = \lambda^*$. $\frac{\partial L}{\partial \lambda} = 0 \Rightarrow \sum_{i=1}^{k} a_i = 1$.
So $a_i^* = \frac{\lambda^*}{2\sigma_i^2}$ and plug it into $\sum_{i=1}^{k} a_i = 1 \Rightarrow \lambda^* = \frac{2}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}$ and thus $a_i^* = \frac{\frac{2}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}}{2\sigma_i^2} = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}$.

Note that SOC is satisfied. Hence $W^* = a_i^* W_i = \frac{\sum_{i=1}^{k} \frac{1}{\sigma_i^2} W_i}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}$ has the minimum variance.

(2) $Var(W^*) = \sum_{i=1}^{k} a_i^{*2} Var(W_i) = \sum_{i=1}^{k} \frac{\left(\frac{1}{\sigma_i^2}\right)^2 \sigma_i^2}{\left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2}\right)^2} = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}}$.

8. Suppose $\{X_1, X_2, \ldots, X_n\}$ is an i.i.d. random sample from some population with unknown mean μ and variance σ^2. Define parameter $\theta = (\mu - 2)^2$.

(1) Suppose $\hat{\theta} = \left(\bar{X}_n - 2\right)^2$ is an estimator for θ, where $\bar{X}_n$ is the sample mean. Show that $\hat{\theta}$ is not unbiased for θ. [Hint: $\bar{X}_n - 2 = \bar{X}_n - \mu + \mu - 2$.]

(2) Find an unbiased estimator for θ.

Solution:

(1) To show $E(\widehat{\theta}) \neq \theta$:

$$\begin{aligned} E(\widehat{\theta}) &= E(\bar{X}_n - 2)^2 = E(\bar{X}_n - \mu + \mu - 2)^2 \\ &= E(\bar{X}_n - \mu)^2 + 2(\mu - 2)E(\bar{X}_n - \mu) + (\mu - 2)^2 \\ &= Var(\bar{X}_n) + 2(\mu - 2)(E\bar{X}_n - \mu) + (\mu - 2)^2 \\ &= Var(\bar{X}_n) + (\mu - 2)^2 > (\mu - 2)^2 = \theta \end{aligned}$$

So $\widehat{\theta}$ is not unbiased for θ.

(2) From (1) $E(\widehat{\theta}) = Var(\bar{X}_n) + \theta = \frac{\sigma^2}{n} + \theta$, and we know $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ is an unbiased estimator for σ^2. So We can define $\widehat{\theta'} = \widehat{\theta} - \frac{S_n^2}{n}$, then $E(\widehat{\theta'}) = E(\widehat{\theta}) - E(\frac{S_n^2}{n}) = E(\widehat{\theta}) - \frac{\sigma^2}{n} = \theta$. $\widehat{\theta'}$ is unbiased.

9. A random sample, $X_1, \ldots, X_n$, is taken from an i.i.d. population with (μ, σ^2). Consider the following estimator of μ :

$$\hat{\mu} = \frac{2}{n(n+1)} \sum_{i=1}^{n} i \cdot X_i = \frac{2}{n(n+1)} \left(X_1 + 2X_2 + 3X_3 + \cdots + nX_n\right)$$

(1) Show $\hat{\mu}$ is unbiased for μ.

(2) Which estimator, $\hat{\mu}$ or $\bar{X}_n$, is more efficient? Explain. [Hint: $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ and $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.]

Solution:

(1) $E(\widehat{\mu}) = \frac{2}{n(n+1)}\sum_{i=1}^n iE(X_i) = \mu\frac{2}{n(n+1)}\sum_{i=1}^n i = \mu\frac{2}{n(n+1)}\frac{n(n+1)}{2} = \mu$, so $\widehat{\mu}$ is unbiased for μ.

(2) Both $\widehat{\mu}$ and $\bar{X}$ are unbiased to μ,

$$\begin{aligned}
MSE(\widehat{\mu}) &= Var(\widehat{\mu}) = Var(\frac{2}{n(n+1)}\sum_{i=1}^n iX_i) \\
&= \frac{4}{n^2(n+1)^2}\sum_{i=1}^n i^2\sigma^2 \quad (X_i \text{ is i.i.d.}) \\
&= \sigma^2\frac{4}{n^2(n+1)^2}\frac{n(n+1)(2n+1)}{6} \\
&= \frac{2(2n+1)}{3n(n+1)}\sigma^2 \\
MSE(\bar{X}) &= Var(X) = \frac{\sigma^2}{n} \\
MSE(\widehat{\mu}) - MSE(\bar{X}) &= \frac{2(2n+1)}{3n(n+1)}\sigma^2 - \frac{\sigma^2}{n} \\
&= \frac{n-1}{3n(n+1)}\sigma^2 > 0 \text{ for } n > 1
\end{aligned}$$

So $\bar{X}$ is more efficient for $n > 1$.

10. Suppose $(X_1, X_2, \ldots, X_n)$ is an i.i.d. $N(\mu, \sigma^2)$ random sample. Define

$$S_n^2 = (n-1)^{-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$; and

$$\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2.$$

(1) Are S_n^2 and $\hat{\sigma}_n^2$ unbiased estimators for σ^2?

(2) Show which estimator is more efficient. Give your reasoning.

Solution:

(1) For X_i ~i.i.d $N(\mu, \sigma^2)$, S_n^2 is unbiased. And $\widehat{\sigma}_n^2 = \frac{(n-1)S_n^2}{n}$, then $E(\widehat{\sigma}_n^2) = \frac{n-1}{n}E(S_n^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$.

So $\widehat{\sigma}_n^2$ is not unbiased estimator for σ^2.

(2) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2 \Rightarrow Var(\frac{(n-1)S_n^2}{\sigma^2}) = 2(n-1) \Rightarrow Var(S_n^2) = \frac{2\sigma^4}{n-1}$

$$MSE(S_n^2) = Var(S_n^2) = \frac{2\sigma^4}{n-1}$$

$Bias(\widehat{\sigma}_n^2)^2 = (E(\widehat{\sigma}_n^2) - \sigma^2)^2 = \frac{1}{n^2}\sigma^4$. $\frac{n\widehat{\sigma}_n^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2 \Rightarrow Var(\frac{n\widehat{\sigma}_n^2}{\sigma^2}) = 2(n-1) \Rightarrow Var(\widehat{\sigma}_n^2) = \frac{2(n-1)\sigma^4}{n^2}$

So

$$\begin{aligned}
MSE(\widehat{\sigma}_n^2) &= Var(\widehat{\sigma}_n^2) + Bias(\widehat{\sigma}_n^2)^2 \\
&= \frac{2(n-1)\sigma^4}{n^2} + \frac{1}{n^2}\sigma^4 \\
&= \frac{2n-1}{n^2}\sigma^4 \\
MSE(S_n^2) - MSE(\widehat{\sigma}_n^2) &= (\frac{2}{n-1} - \frac{2n-1}{n^2})\sigma^4 \\
&= \frac{3n-1}{n(n-1)}\sigma^4 > 0 \text{ for } n > 1
\end{aligned}$$

Therefore, $\widehat{\sigma}_n^2$ is more efficient.

11. Let $X_1, \cdots, X_n$ be an IID random sample from the following distribution:

$$P(X=-1) = \frac{1-\theta}{2}, \quad P(X=0) = \frac{1}{2}, \quad P(X=1) = \frac{\theta}{2}.$$

(1) Find the MLE of θ and check whether it is unbiased estimator;

(2) Find the method of moments estimator of θ;

(3) Calculate the Cramer-Rao lower bound for the variance of an unbiased estimator of θ.

Solution:

(1) X_i follows categorical distribution with three outcomes. This is a generalization of Bernoulli distribution. Given the iid sample, we can write the joint PMF as

$$f_{X^n}(x^n|\theta) = \prod_{i=1}^{n}(\frac{1-\theta}{2})^{\mathbf{1}(x_i=-1)}(\frac{1}{2})^{\mathbf{1}(x_i=0)}(\frac{\theta}{2})^{\mathbf{1}(x_i=1)},$$

where $\mathbf{1}(\cdot)$ is an indicator function. Then we can further write the log-likelihood function as

$$\begin{aligned} lnL(\theta|x^n) &= ln(\frac{1-\theta}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=-1) \\ &\quad +ln(\frac{1}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=0) + ln(\frac{\theta}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=1). \end{aligned}$$

Check the FOC:

$$\frac{\partial lnL(\theta|x^n)}{\partial\theta} = \frac{-1}{1-\theta}\sum_{i=1}^{n}\mathbf{1}(x_i=-1) + \frac{1}{\theta}\sum_{i=1}^{n}\mathbf{1}(x_i=1) = 0$$

Check the SOC:

$$\frac{\partial^2 lnL(\theta|x^n)}{\partial^2\theta} = \frac{-1}{(1-\theta)^2}\sum_{i=1}^{n}\mathbf{1}(x_i=-1) - \frac{1}{\theta^2}\sum_{i=1}^{n}\mathbf{1}(x_i=1) < 0.$$

By FOC, we have $\widehat{\theta_{MLE}} = \frac{\sum_{i=1}^n \mathbf{1}(X_i=1)}{\sum_{i=1}^n \mathbf{1}(X_i=1)+\sum_{i=1}^n \mathbf{1}(X_i=-1)} = \frac{\sum_{i=1}^n \mathbf{1}(X_i=1)}{n-\sum_{i=1}^n \mathbf{1}(X_i=0)}$.

Denote $a = \sum_{i=1}^n \mathbf{1}(X_i=-1)$, $b = \sum_{i=1}^n \mathbf{1}(X_i=0)$, and $c = \sum_{i=1}^n \mathbf{1}(X_i=1)$, then we have $\widehat{\theta_{MLE}} = \frac{c}{a+c} = \frac{n-a-b}{n-b}$. We can show that (a,b,c) follows trinomial distribution with the following density

$$f(a,b,c) = \frac{n!}{a!b!c!}(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^c.$$

Since we want to calculate the expectation of $\widehat{\theta_{MLE}}$, we need to calculate $E(\frac{n-a-b}{n-b})$.

$$\begin{aligned} E(\frac{n-a-b}{n-b}) &= \sum_{b=0}^{n}\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}\frac{n!}{a!b!(n-a-b)!}(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^{n-a-b} \\ &= \sum_{b=0}^{n}\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}C_n^b C_{n-b}^a(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^{n-a-b} \\ &= \sum_{b=0}^{n}C_n^b(\frac{1}{2})^b\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}C_{n-b}^a(\frac{1-\theta}{2})^a(\frac{\theta}{2})^{n-a-b} \end{aligned}$$

For the summation of a, when $a = n-b$, the term $\frac{n-a-b}{n-b} C_{n-b}^{a} (\frac{1-\theta}{2})^{a} (\frac{\theta}{2})^{n-a-b} = 0$. So we can change the upper bound of the summation from $n-b$ to $n-b-1$, we have

$$
\begin{aligned}
E(\frac{n-a-b}{n-b}) &= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{n-a-b}{n-b} C_{n-b}^{a} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{n-a-b}{n-b} \frac{(n-b)!}{(n-b-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{(n-b-1)!}{(n-b-1-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{(n-b-1)!}{(n-b-1-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b-1} \times \frac{\theta}{2} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} C_{n-b-1}^{a} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-b-1-a} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1-\theta}{2} + \frac{\theta}{2})^{n-b-1} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b-1} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b} \times 2 \\
&= \theta \times \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b} \\
&= \theta \times (\frac{1}{2} + \frac{1}{2})^n \\
&= \theta,
\end{aligned}
$$

where we have make use of the following equality

$$
(p+q)^n = \sum_{i=0}^{n} C_n^i p^i q^{n-i}.
$$

Thus, the MLE is unbiased.

Many of you show the unbiasedness by arguing

$$E\widehat{\theta_{MLE}} = E\left(\frac{\sum_{i=1}^n \mathbf{1}(X_i = 1)}{n - \sum_{i=1}^n \mathbf{1}(X_i = 0)}\right) = E(\frac{A}{B}) = \frac{EA}{EB} = \frac{E\sum_{i=1}^n \mathbf{1}(X_i = 1)}{n - E\sum_{i=1}^n \mathbf{1}(X_i = 0)} = \theta,$$

where we let $A = \sum_{i=1}^n \mathbf{1}(x_i = 1)$ **and** $B = n - \sum_{i=1}^n \mathbf{1}(x_i = 0)$**. This argument is WRONG! Since in general,** $E(\frac{A}{B}) \neq \frac{EA}{EB} = \theta$**. Therefore, you need to calculate the expectation using definition.**

(2) For the Method of Moment Estimator, we just need to match the first moment because $E(X) = -1 \times \frac{1-\theta}{2} + 1 \times \frac{\theta}{2} = \theta - \frac{1}{2}$. So we have

$$\frac{1}{n}\sum_{i=1}^n X_i = \widehat{\theta_{MME}} - \frac{1}{2}.$$

Thus we have $\widehat{\theta_{MME}} = \frac{1}{n}\sum_{i=1}^n X_i + \frac{1}{2}$. It is unbiased since $E(\widehat{\theta_{MME}}) = E(X_i) + 1/2 = \theta - 1/2 + 1/2 = \theta$.

(3) Since only the $\widehat{\theta_{MME}}$ is unbiased, so we just need to calculate the Cramer-Rao lower bound for unbiased estimator. By definition, the Cramer-Rao lower bound $B_n(\theta)$

$$B_n(\theta) = \frac{\left[\frac{dE_\theta(\widehat{\theta_{MME}})}{d\theta}\right]^2}{E_\theta\left[\frac{\partial lnf_{X^n}(x^n,\theta)}{\partial\theta}\right]^2}.$$

The numerator is 1 given the MME estimator is unbiased. For the denominator, given i.i.d. and information matrix equality, we have

$$B_n(\theta) = \frac{1}{-nE_\theta\left[\frac{\partial^2 lnf_{X_i}(x_i,\theta)}{\partial^2\theta}\right]}.$$

The log of marginal PMF is given by

$$lnf_{X_i}(x_i,\theta) = \mathbf{1}(x_i = -1)ln(\frac{1-\theta}{2}) + \mathbf{1}(x_i = 0)ln(\frac{1}{2}) + \mathbf{1}(x_i = 1)ln(\frac{\theta}{2}).$$

Then we have

$$\frac{\partial^2 lnf_{X_i}(x_i,\theta)}{\partial^2\theta} = \frac{-1}{(1-\theta)^2}\mathbf{1}(x_i = -1) - \frac{1}{\theta^2}\mathbf{1}(x_i = 1).$$

By definition of indicator function, $E(\mathbf{1}(x_i = 1)) = P(X_i = 1) = \frac{\theta}{2}$ and $E(\mathbf{1}(x_i = -1)) = P(X_i = -1) = \frac{1-\theta}{2}$. Then we have

$$E\left[\frac{\partial^2 ln f_{X_i}(x_i, \theta)}{\partial^2 \theta}\right] = \frac{-1}{2(1-\theta)} - \frac{1}{2\theta} = \frac{-1}{2\theta(1-\theta)}.$$

Finally, we have the Cramer-Rao lower bound given by

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

We can see that the MME estimator doesn't achieve this lower bound since $Var(\widehat{\theta_{MME}}) = \frac{1}{2n} \geq B_n(\theta) = \frac{2\theta(1-\theta)}{n}$.

12. Suppose $X^n = (X_1, \cdots, X_n)$ is an IID random sample from a Poisson(α) distribution with probability mass function

$$f_X(x) = e^{-\alpha} \frac{\alpha^x}{x!} \text{ for } x = 0, 1, 2, \cdots,$$

where α is unknown.

(1) Find the MLE for α;

(2) Is the MLE for α the best unbiased estimator for α? Give your reasoning.

Solution:

(1) Given the i.i.d. data X^n, we can write the log-likelihood function as

$$lnL(\alpha|x^n) = -n\alpha - \sum_{i=1}^{n} ln(x_i!) + ln\alpha \sum_{i=1}^{n} x_i.$$

By FOC, we have

$$\frac{\partial lnL(\alpha|x^n)}{\partial \alpha} = -n + \frac{1}{\alpha} \sum_{i=1}^{n} x_i = 0$$

we can have $\widehat{\alpha_{MLE}} = \frac{1}{n} \sum_{i=1}^{n} x_i$. Checking the SOC we have $\frac{\partial^2 lnL(\alpha|x^n)}{\partial^2 \alpha} = \frac{-1}{\alpha^2} \sum_{i=1}^{n} x_i < 0$. Thus we have $\widehat{\alpha_{MLE}}$ is a global maximizer.

(2) To check if the MLE is the best unbiased estimator for α, we need to calculate the Cramer-Rao lower bound and compare it to the variance

of the MLE estimator. By similar argument, we just need to calculate the expectation of the second order derivative of the log density function.

$$\frac{\partial^2 lnf_X(x)}{\partial^2\alpha} = \frac{-1}{\alpha^2}x.$$

Then we have the denominator is given by

$$-nE\left[\frac{\partial^2 lnf_X(x)}{\partial^2\alpha}\right] = -n\frac{-1}{\alpha^2}EX = \frac{n}{\alpha}.$$

And we have the Cramer-Rao lower bound $B_n(\alpha) = \frac{\alpha}{n}$. Next, we need to calculate $Var(\widehat{\alpha_{MLE}})$:

$$\begin{aligned} Var(\widehat{\alpha_{MLE}}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}Var(X) \\ &= \frac{\alpha}{n}. \end{aligned}$$

Since $Var(\widehat{\alpha_{MLE}}) = B_n(\alpha)$, then the MLE achieves the Cramer-Rao lower bound and is the best unbiased estimator for α.

13. Suppose Assumptions M.1-M.6 hold except that the density model $f(x, \theta)$ may not be correctly specified for the unknown population density $f_X(x)$, i.e., there exists no $\theta \in \Theta$ such that $f_X(x) = f(x, \theta)$ for all x. The MLE $\hat{\theta} = \arg\max_{\theta\in\Theta} \ln \hat{L}(\theta \mid \mathbf{X}^n)$ is called the Qausi-MLE. (1) Show $\hat{\theta} \to \theta_0$ almost surely as $n \to \infty$; (2) derive the asymptotic distribution of $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$ and compare it with the result of Theorem 8.5.

Solution:

(1) The result still holds because nothing is changed in the proof of Theorem 8.4. Note that here θ_0 is still the unique maximizer of $Q(\theta) = E[lnf(X_i, \theta)]$ and $\widehat{Q}_n(\theta) \overset{a.s.}{\to} Q(\theta)$ no matter what the true distribution is. (2) First, Lemma 8.2 does not hold if the model is misspecified. That is, mean of score function $E[S_i(\theta)] \neq 0$.[**Correction: Lemma 8.2 does not hold for any θ, but $E[S_i(\theta_0)] = 0$ since θ_0 is the optimizer to $Q(\theta) = E[lnf(X_i, \theta)]$. If you interchange the expectation and differentiation, you will see that.**] Second, Information Matrix Equality

(Lemma 8.3) no longer holds. This will not change the structure of the proof of Theorem 8.5 but make some of the simplification impossible. Note that now $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial lnf(X_i,\theta_0)}{\partial\theta} \xrightarrow{d} N(E[S_i(\theta_0)], I(\theta_0))$.

Therefore, $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N([-H(\theta_0)]^{-1}E[S_i(\theta_0)], [H(\theta_0)]^{-1}I(\theta_0)[H(\theta_0)]^{-1})$ [**Correction: this should be** $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N([0, [H(\theta_0)]^{-1}I(\theta_0)[H(\theta_0)]^{-1})$ **based on the correction above.**]

14. Prove Theorem 8.9 and discuss the implication of the theorem.

Solution:

See page 445 of the textbook. The implication is that GMM estimator achieve the lowest variance (or most efficient) by choosing the weighting matrix to be V. In practice, everything about GMM is to get an estimate of this V under different specification. You will spend a lot more time on this in ECON 6200.

15. Suppose $\hat{\theta}_1, \hat{\theta}_2$ and $\hat{\theta}_3$ are estimators of θ, and we know that $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$, $E(\hat{\theta}_3) \neq \theta$, $\text{var}(\hat{\theta}_1) = 12$, $\text{var}(\hat{\theta}_2) = 10$, and $E(\hat{\theta}_3 - \theta)^2 = 6$. Which estimator is the best in terms of MSE criterion?

Solution:

$MSE(\widehat{\theta}_1) = var(\widehat{\theta}_1) = 12$,$MSE(\widehat{\theta}_2) = var(\widehat{\theta}_2) = 10$. $MSE(\widehat{\theta}_3) = 6$. So $3 > 2 > 1$.

16. Let $\mathbf{X}^n$ be an IID $U[0,\theta]$ random sample, where θ is unknown. Define two estimators of θ :

$$\begin{aligned}\hat{\theta}_1 &= \frac{n+1}{n}\max_{1\le i\le n} X_i,\\ \hat{\theta}_2 &= \frac{2}{n}\sum_{i=1}^{n} X_i.\end{aligned}$$

(1) Show $P(\max_{1\le i\le n} X_i \le t) = [F_X(t)]^n$, where $F_X(\cdot)$ is the CDF of the population distribution $U[0,\theta]$;

(2) Compute $E_\theta(\hat{\theta}_1)$ and $\text{var}_\theta(\hat{\theta}_1)$;

(3) Show $\hat{\theta}_1$ converges to θ in probability;

(4) Compute $E_\theta(\hat{\theta}_2)$ and $\text{var}_\theta(\hat{\theta}_2)$;

(5) Which estimator, $\hat{\theta}_1$ or $\hat{\theta}_2$, is more efficient? Explain.

Solution:

(1) $P(max_i X_i \le t) = P(X_i \le t, i = 1, ..., n) = F_X(t)^n$

(2) Fist notice $F_{max_i X_i}(t) = F_X(t)^n$. Then $f_{max_i X_i}(t) = F'_{max_i X_i}(t) = n(\frac{t}{\theta})^{n-1}\frac{1}{\theta}$. Then easy to calculate $E(max_i X_i) = \frac{n}{n+1}\theta$. Therefore $E_\theta(\widehat{\theta}_1) = \theta$ and $var_\theta(\widehat{\theta}_1) = \frac{1}{n(n+2)}\theta^2$.

(3) $MSE(\widehat{\theta}_1) = \frac{1}{n(n+2)}\theta^2 \to 0$ as $n \to \infty$. Hence $\widehat{\theta}_1 \xrightarrow{p} \theta$

(4) $E_\theta(\widehat{\theta}_2) = \frac{2}{n}\sum E(X_i) = \frac{2}{n}n\frac{\theta}{2} = \theta$ and $var(\widehat{\theta}_2) = \frac{1}{3n}\theta^2$

(5) $MSE(\widehat{\theta}_1) < MSE(\widehat{\theta}_2)$. Therefore $\widehat{\theta}_1$ is better.

17. An IID random sample $\mathbf{X}^n$ is taken from a population with mean μ and variance σ^2. Consider the following estimator of μ :

$$\hat{\mu} = \frac{2}{n(n+1)}\sum_{i=1}^{n} i \cdot X_i.$$

(1) Show $\hat{\mu}$ is unbiased for μ;

(2) Which estimator, $\hat{\mu}$ or $\bar{X}_n$, is more efficient? Explain. [Hint: $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ *and* $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.]

Solution:

(1) $E(\widehat{\mu}) = \frac{2}{n(n+1)}\sum_{i=1}^n iE(X_i) = \mu\frac{2}{n(n+1)}\sum_{i=1}^n i = \mu\frac{2}{n(n+1)}\frac{n(n+1)}{2} = \mu$, so $\widehat{\mu}$ is unbiased for μ.

(2) Both $\widehat{\mu}$ and $\bar{X}$ are unbiased to μ,

$$\begin{aligned}
MSE(\widehat{\mu}) &= Var(\widehat{\mu}) = Var(\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i) \\
&= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i^2\sigma^2 \quad (X_i \text{ is i.i.d.}) \\
&= \sigma^2\frac{4}{n^2(n+1)^2}\frac{n(n+1)(2n+1)}{6} \\
&= \frac{2(2n+1)}{3n(n+1)}\sigma^2 \\
MSE(\bar{X}) &= Var(X) = \frac{\sigma^2}{n} \\
MSE(\widehat{\mu}) - MSE(\bar{X}) &= \frac{2(2n+1)}{3n(n+1)}\sigma^2 - \frac{\sigma^2}{n} \\
&= \frac{n-1}{3n(n+1)}\sigma^2 > 0 \text{ for } n > 1
\end{aligned}$$

So $\bar{X}$ is more efficient for $n > 1$.

18. Suppose $\mathbf{X}^n$ is an IID $N(0, \sigma^2)$ random sample. Define

$$S_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

where $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$; and

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} X_i^2.$$

Which estimator is more efficient? Give your reasoning.
Solution:

$MSE(S_n^2) = Var(S_n^2) = \frac{2\sigma^4}{n-1}$.

$MSE(\widehat{\sigma}^2) = Bias^2 + var(\widehat{\sigma}^2) = var(\widehat{\sigma}^2) = \frac{2\sigma^4}{n}$

Therefore,$\widehat{\sigma}^2$ is more efficient.

19. Let $X_1, \cdots, X_n$ be an IID random sample from the following distribution:

$$P(X = -1) = \frac{1-\theta}{2}, \quad P(X = 0) = \frac{1}{2}, \quad P(X = 1) = \frac{\theta}{2}.$$

(1) Find the MLE of θ and check whether it is unbiased estimator;
(2) Find the method of moments estimator of θ;
(3) Calculate the Cramer-Rao lower bound for the variance of an unbiased estimator of θ.
Solution:

(1) X_i follows categorical distribution with three outcomes. This is a generalization of Bernoulli distribution. Given the iid sample, we can write the joint PMF as

$$f_{X^n}(x^n|\theta) = \prod_{i=1}^{n} (\frac{1-\theta}{2})^{\mathbf{1}(x_i=-1)} (\frac{1}{2})^{\mathbf{1}(x_i=0)} (\frac{\theta}{2})^{\mathbf{1}(x_i=1)},$$

where $\mathbf{1}(\cdot)$ is an indicator function. Then we can further write the log-likelihood function as

$$\begin{aligned} lnL(\theta|x^n) &= ln(\frac{1-\theta}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=-1) \\ &\quad +ln(\frac{1}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=0)+ln(\frac{\theta}{2})\sum_{i=1}^{n}\mathbf{1}(x_i=1). \end{aligned}$$

Check the FOC:

$$\frac{\partial lnL(\theta|x^n)}{\partial\theta}=\frac{-1}{1-\theta}\sum_{i=1}^{n}\mathbf{1}(x_i=-1)+\frac{1}{\theta}\sum_{i=1}^{n}\mathbf{1}(x_i=1)=0$$

Check the SOC:

$$\frac{\partial^2 lnL(\theta|x^n)}{\partial^2\theta}=\frac{-1}{(1-\theta)^2}\sum_{i=1}^{n}\mathbf{1}(x_i=-1)-\frac{1}{\theta^2}\sum_{i=1}^{n}\mathbf{1}(x_i=1)<0.$$

By FOC, we have $\widehat{\theta_{MLE}}=\frac{\sum_{i=1}^{n}\mathbf{1}(X_i=1)}{\sum_{i=1}^{n}\mathbf{1}(X_i=1)+\sum_{i=1}^{n}\mathbf{1}(X_i=-1)}=\frac{\sum_{i=1}^{n}\mathbf{1}(X_i=1)}{n-\sum_{i=1}^{n}\mathbf{1}(X_i=0)}$.

Denote $a=\sum_{i=1}^{n}\mathbf{1}(X_i=-1)$, $b=\sum_{i=1}^{n}\mathbf{1}(X_i=0)$, and $c=\sum_{i=1}^{n}\mathbf{1}(X_i=1)$, then we have $\widehat{\theta_{MLE}}=\frac{c}{a+c}=\frac{n-a-b}{n-b}$. We can show that (a,b,c) follows trinomial distribution with the following density

$$f(a,b,c)=\frac{n!}{a!b!c!}(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^c.$$

Since we want to calculate the expectation of $\widehat{\theta_{MLE}}$, we need to calculate $E(\frac{n-a-b}{n-b})$.

$$\begin{aligned} E(\frac{n-a-b}{n-b}) &= \sum_{b=0}^{n}\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}\frac{n!}{a!b!(n-a-b)!}(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^{n-a-b} \\ &= \sum_{b=0}^{n}\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}C_n^bC_{n-b}^a(\frac{1-\theta}{2})^a(\frac{1}{2})^b(\frac{\theta}{2})^{n-a-b} \\ &= \sum_{b=0}^{n}C_n^b(\frac{1}{2})^b\sum_{a=0}^{n-b}\frac{n-a-b}{n-b}C_{n-b}^a(\frac{1-\theta}{2})^a(\frac{\theta}{2})^{n-a-b} \end{aligned}$$

For the summation of a, when $a = n-b$, the term $\frac{n-a-b}{n-b} C_{n-b}^a (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} = 0$. So we can change the upper bound of the summation from $n-b$ to $n-b-1$, we have

$$\begin{aligned}
E(\frac{n-a-b}{n-b}) &= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{n-a-b}{n-b} C_{n-b}^a (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{n-a-b}{n-b} \frac{(n-b)!}{(n-b-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{(n-b-1)!}{(n-b-1-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b} \\
&= \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} \frac{(n-b-1)!}{(n-b-1-a)!a!} (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-a-b-1} \times \frac{\theta}{2} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \sum_{a=0}^{n-b-1} C_{n-b-1}^a (\frac{1-\theta}{2})^a (\frac{\theta}{2})^{n-b-1-a} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1-\theta}{2} + \frac{\theta}{2})^{n-b-1} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b-1} \\
&= \frac{\theta}{2} \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b} \times 2 \\
&= \theta \times \sum_{b=0}^{n} C_n^b (\frac{1}{2})^b \times (\frac{1}{2})^{n-b} \\
&= \theta \times (\frac{1}{2} + \frac{1}{2})^n \\
&= \theta,
\end{aligned}$$

where we have make use of the following equality

$$(p+q)^n = \sum_{i=0}^{n} C_n^i p^i q^{n-i}.$$

Thus, the MLE is unbiased.

Many of you show the unbiasedness by arguing

$$E\widehat{\theta_{MLE}} = E\left(\frac{\sum_{i=1}^n \mathbf{1}(X_i = 1)}{n - \sum_{i=1}^n \mathbf{1}(X_i = 0)}\right) = E(\frac{A}{B}) = \frac{EA}{EB} = \frac{E\sum_{i=1}^n \mathbf{1}(X_i = 1)}{n - E\sum_{i=1}^n \mathbf{1}(X_i = 0)} = \theta,$$

where we let $A = \sum_{i=1}^n \mathbf{1}(x_i = 1)$ **and** $B = n - \sum_{i=1}^n \mathbf{1}(x_i = 0)$**. This argument is WRONG! Since in general,** $E(\frac{A}{B}) \neq \frac{EA}{EB} = \theta$**. Therefore, you need to calculate the expectation using definition.**

(2) For the Method of Moment Estimator, we just need to match the first moment because $E(X) = -1 \times \frac{1-\theta}{2} + 1 \times \frac{\theta}{2} = \theta - \frac{1}{2}$. So we have

$$\frac{1}{n}\sum_{i=1}^n X_i = \widehat{\theta_{MME}} - \frac{1}{2}.$$

Thus we have $\widehat{\theta_{MME}} = \frac{1}{n}\sum_{i=1}^n X_i + \frac{1}{2}$. It is unbiased since $E(\widehat{\theta_{MME}}) = E(X_i) + 1/2 = \theta - 1/2 + 1/2 = \theta$.

(3) Since only the $\widehat{\theta_{MME}}$ is unbiased, so we just need to calculate the Cramer-Rao lower bound for unbiased estimator. By definition, the Cramer-Rao lower bound $B_n(\theta)$

$$B_n(\theta) = \frac{\left[\frac{dE_\theta(\widehat{\theta_{MME}})}{d\theta}\right]^2}{E_\theta\left[\frac{\partial lnf_{X^n}(x^n,\theta)}{\partial\theta}\right]^2}.$$

The numerator is 1 given the MME estimator is unbiased. For the denominator, given i.i.d. and information matrix equality, we have

$$B_n(\theta) = \frac{1}{-nE_\theta\left[\frac{\partial^2 lnf_{X_i}(x_i,\theta)}{\partial^2\theta}\right]}.$$

The log of marginal PMF is given by

$$lnf_{X_i}(x_i,\theta) = \mathbf{1}(x_i = -1)ln(\frac{1-\theta}{2}) + \mathbf{1}(x_i = 0)ln(\frac{1}{2}) + \mathbf{1}(x_i = 1)ln(\frac{\theta}{2}).$$

Then we have

$$\frac{\partial^2 lnf_{X_i}(x_i,\theta)}{\partial^2\theta} = \frac{-1}{(1-\theta)^2}\mathbf{1}(x_i = -1) - \frac{1}{\theta^2}\mathbf{1}(x_i = 1).$$

By definition of indicator function, $E(\mathbf{1}(x_i = 1)) = P(X_i = 1) = \frac{\theta}{2}$ and $E(\mathbf{1}(x_i = -1)) = P(X_i = -1) = \frac{1-\theta}{2}$. Then we have

$$E\left[\frac{\partial^2 ln f_{X_i}(x_i, \theta)}{\partial^2 \theta}\right] = \frac{-1}{2(1-\theta)} - \frac{1}{2\theta} = \frac{-1}{2\theta(1-\theta)}.$$

Finally, we have the Cramer-Rao lower bound given by

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

We can see that the MME estimator doesn't achieve this lower bound since $Var(\widehat{\theta_{MME}}) = \frac{1}{2n} \geq B_n(\theta) = \frac{2\theta(1-\theta)}{n}$.

20. Let $X_1, \cdots, X_n$ be an IID random sample from the population with PMF

$$f(x, \theta) = \begin{cases} \theta & \text{if } x = 1 \\ 1 - \theta & \text{if } x = 0, \end{cases}$$

where $0 < \theta < 1$.
(1) Find the MLE $\hat{\theta}$ of θ;
(2) Is $\hat{\theta}$ the best unbiased estimator of θ?

Solution:

(1) Let X be the number of X_i with value equals to 1. Then $n - X$ is the number of X_i with value equals to 0. Since X_i follow Bernoulli(p), then X follows $B(n, \theta)$. Furthermore, $L(\theta|x^n) = \theta^X(1-\theta)^{n-X}$. FOC implies $\widehat{\theta} = X/n$. (2) $E(\widehat{\theta}) = E(X)/n = n\theta/n = \theta$. So it is unbiased.

21. Put $\theta = (\mu, \sigma^2)$. A random variable X with PDF

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \ 0 < x < \infty,$$

is called a Lognormal (μ, σ^2) random variable because its logarithm $\ln X$ is $N(\mu, \sigma^2)$, namely,

$$\ln X \sim N(\mu, \sigma^2).$$

Suppose $\{X_1, \cdots, X_n\}$ is an IID random sample from a Lognormal (μ, σ^2) population.

(1) Find the Maximum Likelihood Estimator (MLE) for (μ, σ^2);

(2) Denote the MLE estimator for μ by $\hat{\mu}$. Is $\hat{\mu}$ the best unbiased estimator of μ?

Solution:

(1) The only difference between the estimators here and MLE for normal distribution is substituting X_i by lnX_i. Therefore, $\widehat{\mu} = \sum lnX_i/n$ and $\widehat{\sigma}^2 = \sum(lnX_i - \widehat{\mu})^2/n$

(2) First easy to see $\widehat{\mu}$ is unbiased. $var(\widehat{\mu}) = \sigma^2/n$. Check the Cramer-Rao Lower Bound by Corollary 8.1:

$$B_n(\theta) = \frac{1}{-nH(\theta)} = \frac{1}{-n(-\frac{1}{\sigma^2})} = \sigma^2/n$$

So it is the best unbiased estimator.

22. Suppose $X^n = (X_1, \cdots, X_n)$ is an IID random sample from a Poisson(α) distribution with probability mass function

$$f_X(x) = e^{-\alpha}\frac{\alpha^x}{x!} \text{ for } x = 0, 1, 2, \cdots,$$

where α is unknown.

(1) Find the MLE for α;

(2) Is the MLE for α the best unbiased estimator for α? Give your reasoning.

Solution:

(1) Given the i.i.d. data X^n, we can write the log-likelihood function as

$$lnL(\alpha|x^n) = -n\alpha - \sum_{i=1}^{n} ln(x_i!) + ln\alpha \sum_{i=1}^{n} x_i.$$

By FOC, we have

$$\frac{\partial lnL(\alpha|x^n)}{\partial\alpha} = -n + \frac{1}{\alpha}\sum_{i=1}^{n} x_i = 0$$

we can have $\widehat{\alpha_{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Checking the SOC we have $\frac{\partial^2 lnL(\alpha|x^n)}{\partial^2\alpha} = \frac{-1}{\alpha^2}\sum_{i=1}^{n} x_i < 0$. Thus we have $\widehat{\alpha_{MLE}}$ is a global maximizer.

(2) To check if the MLE is the best unbiased estimator for α, we need to calculate the Cramer-Rao lower bound and compare it to the variance of the MLE estimator. By similar argument, we just need to calculate the expectation of the second order derivative of the log density function.

$$\frac{\partial^2 lnf_X(x)}{\partial^2\alpha} = \frac{-1}{\alpha^2}x.$$

Then we have the denominator is given by

$$-nE\left[\frac{\partial^2 lnf_X(x)}{\partial^2\alpha}\right] = -n\frac{-1}{\alpha^2}EX = \frac{n}{\alpha}.$$

And we have the Cramer-Rao lower bound $B_n(\alpha) = \frac{\alpha}{n}$. Next, we need to calculate $Var(\widehat{\alpha_{MLE}})$:

$$\begin{aligned} Var(\widehat{\alpha_{MLE}}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}Var(X) \\ &= \frac{\alpha}{n}. \end{aligned}$$

Since $Var(\widehat{\alpha_{MLE}}) = B_n(\alpha)$, then the MLE achieves the Cramer-Rao lower bound and is the best unbiased estimator for α.