EE 3070 Statistics

Homework #2 Solution

Due at 23:59, May 23, 2023
online submission to eeclass systems

Note: There are **8** questions, each worth **13** points, for a total of 104 points. The maximum score, however, will be capped at 100 points.

You may use computers, software packages, and online tools for this Homework.

1. **(Exercise 4.1.3)** Suppose the number of customers X that enter a store between the hours 9:00 a.m. and 10:00 a.m. follows a Poisson distribution with parameter θ. Suppose a random sample of the number of customers that enter the store between 9:00 a.m. and 10:00 a.m. for 10 days results in the values: 9, 7, 9, 15, 10, 13, 11, 7, 2, 12.

   Note: You may feel that this question seems familiar. Yup, it's the same one from the midterm.
   (a) Determine the maximum likelihood estimate of θ. Show that it is an unbiased estimator.
   (b) Based on these data, obtain the realization of your estimator in part **(a)**. Explain the meaning of this estimate in terms of the number of customers.

   Solution:
   (a) To compute $\widehat{\theta}$, we can use the definition of the ML estimator:

$$X \sim P(\theta) \quad P_X(x) = \frac{e^{-\theta} \cdot \theta^x}{x!}, \quad x = 0, 1, 2, \cdots$$

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} P(x_i > \theta) = \prod_{i=1}^{n} \frac{e^{-\theta} \cdot \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \cdot \theta^{x_1+x_2+\cdots+x_n}}{x_1! x_2! \cdots x_n!} \\
\Rightarrow \ln\left(L(\theta)\right) &= \ln(e^{-n\theta} \cdot \theta^{x_1+x_2+\cdots+x_n}) - \ln(x_1! x_2! \cdots x_n!) \\
&= \ln(e^{-n\theta}) + \ln(\theta^{x_1+x_2+\cdots+x_n}) - \ln(x_1! x_2! \cdots x_n!) \\
&= -n \cdot \theta + (\sum_{i=1}^{n} x_i) \cdot \ln(\theta) - \ln(\prod_{i=1}^{n} x_i!) \\
\Rightarrow \frac{\partial \ln\left(L(\theta)\right)}{\partial \theta} &= -n + \frac{1}{\theta} \sum_{i=1}^{n} x_i = 0 \\
&\Rightarrow \frac{1}{\theta} \sum_{i=1}^{n} x_i = n \\
&\Rightarrow \theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} \\
&\Rightarrow \text{ML estimator} \quad \widehat{\theta} = \bar{x}
\end{aligned}$$

To show this estimator to be unbiased, we need to find its expected value of $\bar{x}$ and show that it is equal to θ. The expected value of $\bar{x}$ is:

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n} \cdot E(\sum_{i=1}^{n} x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} \theta = \frac{1}{n} \cdot n \cdot \theta = \theta$$

Therefore, we have shown that the ML estimator $\widehat{\theta} = \bar{x}$ is unbiased.

(b)

$$\hat{\theta} = \frac{1}{10}(9+7+9+15+10+13+11+7+2+12) = \frac{95}{10} = 9.5$$

■

2. **(Exercise 4.2.1)** Let the observed value of the mean $\bar{X}$ and of the sample variance of a random sample of size 20 from a distribution that is $N(\mu, \sigma^2)$ be 81.2 and 26.5, respectively. Find respectively $90\%, 95\%$ and 99% confidence intervals for μ.
   Note: how the lengths of the confidence intervals increase as the confidence increases.

   Solution:

$$\bar{x} - (t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{N}}) \leq \mu \leq \bar{x} + (t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{N}}), \quad \frac{S}{\sqrt{N}} = \sqrt{\frac{26.5}{20}} \sim 1.1511$$

   (1) The 90% confidence interval for μ is:

$$\begin{aligned} \alpha = 1 - 0.90 = 0.10 \quad &\Rightarrow \quad t_{\frac{\alpha}{2},n-1} = t_{0.05,19} = 1.729 \\ \Rightarrow \ 81.2 - (1.729{\times}1.1511) &\leq \mu \leq 81.2 + 1.9903 \\ \Rightarrow \qquad 79.2097 &\leq \mu \ \leq 83.1903 \end{aligned}$$

   (2) The 95% confidence interval for μ is:

$$\begin{aligned} \alpha = 1 - 0.95 = 0.05 \quad &\Rightarrow \quad t_{\frac{\alpha}{2},n-1} = t_{0.025,19} = 2.093 \\ \Rightarrow \ 81.2 - (2.093{\times}1.1511) &\leq \mu \leq 81.2 + 2.4100 \\ \Rightarrow \qquad 78.79 &\leq \mu \ \leq 83.61 \end{aligned}$$

   (3) The 99% confidence interval for μ is:

$$\begin{aligned} \alpha = 1 - 0.99 = 0.01 \quad &\Rightarrow \quad t_{\frac{\alpha}{2},n-1} = t_{0.005,19} = 2.861 \\ \Rightarrow \ 81.2 - (2.861{\times}1.1511) &\leq \mu \leq 81.2 + 3.2933 \\ \Rightarrow \qquad 77.9067 &\leq \mu \ \leq 84.4933 \end{aligned}$$

■

3. **(Exercise 4.2.9)** Let $\bar{X}$ denote the mean of a random sample of size n from a distribution that has mean μ and variance $\sigma^2 = 10$. Find n so that the probability is approximately 0.954 that the random interval $(\bar{X} - \frac{1}{2}, \bar{X} + \frac{1}{2})$ includes μ.

   Solution:

$$\begin{aligned} &\bar{x} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S}{N}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S}{N}} \\ &P(\bar{X} - \frac{1}{2} < \mu < \bar{X} + \frac{1}{2}) = 0.954 \\ &z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{S}{N}} = \frac{1}{2} = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}, \quad \alpha = 1 - 0.954 = 0.046 \end{aligned}$$

$\Rightarrow z_{\frac{\alpha}{2}} = z_{\frac{0.046}{2}} = z_{0.023} \sim 1.995$

$\Rightarrow 1.995 \times \sqrt{\frac{10}{n}} = \frac{1}{2} \rightarrow (1.995 \times 2)^2 = (\sqrt{\frac{n}{10}})^2 \rightarrow (3.99)^2 \times 10 = 159.201 \sim 160$

$\therefore n = 160$

■

4. **(Exercise 4.4.7)** Let $f(x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6$, zero elsewhere, be the pmf of a distribution of the discrete type. Show that the pmf of the smallest observation of a random sample of size 5 from this distribution is

$$g_1(y_1) = \left(\frac{7-y_1}{6}\right)^5 - \left(\frac{6-y_1}{6}\right)^5, y_1 = 1, 2, \cdots, 6, \text{ zero elsewhere.}$$

Note that in this exercise the random sample is from a distribution of the discrete type. All formulas in the text were derived under the assumption that the random sample is from a distribution of the continuous type and are not applicable. Why?

Solution:
Since the distribution is of the discrete type, we cannot use the formulas in the textbook. However,

$$P(Y_1 = y_1) = P(\text{all} \geq y_1) - P(\text{all} \geq y_1 + 1) = (\frac{7-y_1}{6})^5 - (\frac{6-y_1}{6})^5.$$

We can also list all possible cases as below,

$$\begin{aligned}
P(Y_1 = 6) &= P(Y_1 = 6, Y_2 = 6, \cdots, Y_5 = 6) = P(Y_1 = 6) \cdot P(Y_2 = 6) \cdots P(Y_5 = 6) \\
&= (\frac{1}{6})^5 - 0^5 = (\frac{7-6}{6})^5 - (\frac{6-6}{6})^5 \\
P(Y_1 = 5) &= P(Y_i \geq 5,\ i = 1, \cdots, 5) - P(Y_1 = 6) = (P(Y_i \geq 5))^5 - P(Y_1 = 6) \\
&= (\frac{2}{5})^5 - (\frac{1}{6})^5 = (\frac{7-5}{6})^5 - (\frac{6-5}{6})^5 \\
P(Y_1 = 4) &= P(Y_i \geq 4,\ i = 1, \cdots, 5) - P(Y_1 = 5) - P(Y_1 = 6) = (\frac{7-4}{6})^5 - (\frac{6-4}{6})^5 \\
P(Y_1 = 3) &= P(Y_i \geq 3,\ i = 1, \cdots, 5) - \sum_{i=4}^{6} P(Y_1 = i) = (\frac{7-3}{6})^5 - (\frac{6-3}{6})^5 \\
P(Y_1 = 2) &= P(Y_i \geq 2,\ i = 1, \cdots, 5) - \sum_{i=3}^{6} P(Y_1 = i) = (\frac{7-2}{6})^5 - (\frac{6-2}{6})^5 \\
P(Y_1 = 1) &= P(Y_i \geq 1,\ i = 1, \cdots, 5) - \sum_{i=2}^{6} P(Y_1 = i) = (\frac{7-1}{6})^5 - (\frac{6-1}{6})^5 \\
&\therefore g_1(y_1) = (\frac{7-y_1}{6})^5 - (\frac{6-y_1}{6})^5,\ \ y_1 = 1, 2, \cdots, 6.
\end{aligned}$$

■

5. **(Exercise 4.4.20)** Let the joint pdf of X and Y be $f(x,y) = \frac{12}{7} \cdot x(x+y), 0 < x < 1, 0 < y < 1$, zero elsewhere. Let $U = \min(X,Y)$ and $V = \max(X,Y)$. Find the joint pdf of U and V.

Solution:
(1) $U = \min(X,Y) = X, U = \max(X,Y) = Y$

$$\begin{aligned} f_{U,V}(u,v) &= f(x=u, y=v) \cdot |\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}| \\ &= \frac{12}{7} \cdot u(u+v) \cdot |\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}| \\ &= \frac{12}{7} \cdot u(u+v) \cdot 1 = \frac{12}{7} \cdot u(u+v) \end{aligned}$$

(2) $U = \min(X,Y) = Y, U = \max(X,Y) = X$

$$\begin{aligned} f_{U,V}(u,v) &= f(x=v, y=u) \cdot |\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}| \\ &= \frac{12}{7} \cdot v(u+v) \cdot |\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}| \\ &= \frac{12}{7} \cdot v(u+v) \cdot |(-1)| = \frac{12}{7} \cdot v(u+v) \end{aligned}$$

$$f_{U,V}(u,v) = \begin{cases} \frac{12}{7} \cdot u(u+v), & 0 < u < v < 1, \ \ u = x, \ v = y \\ \frac{12}{7} \cdot v(u+v), & 0 < u < v < 1, \ \ u = y, \ v = x \end{cases}$$

■

6. **(Exercise 4.5.3)** Let X have a pdf of the form $f(x;\theta) = \theta \cdot x^{(\theta-1)}, 0 < x < 1$, zero elsewhere, where $\theta \in \{\theta : \theta = 1, 2\}$. To test the simple hypothesis $H_0 : \theta = 1$ against the alternative simple hypothesis $H_1 : \theta = 2$, use a random sample X_1, X_2 of size $n = 2$ and define the critical region to be $C = \{(x_1, x_2) : \frac{3}{4} \leq x_1 \cdot x_2\}$. Find the power function of the test.

Solution:

$$\begin{aligned} g(x_1, x_2) &= \begin{cases} n! \ f(x_1) \cdot f(x_2), & a < x_1 < x_2 < b \\ \quad 0 \quad , & \text{elsewhere} \end{cases} \\ &= \begin{cases} 2! \ \theta \cdot x_1^{\theta-1} \cdot \theta \cdot x_2^{\theta-1}, & 0 < x_1 < x_2 < 1 \\ \quad 0 \quad , & \text{elsewhere} \end{cases} \\ &= \begin{cases} 2 \cdot \theta^2 \cdot (x_1 \cdot x_2)^{\theta-1}, & 0 < x_1 < x_2 < 1 \\ \quad 0 \quad , & \text{elsewhere} \end{cases} \end{aligned}$$

$$\begin{aligned} P(\frac{3}{4} \leq x_1 \cdot x_2) &= \int_{\frac{3}{4}}^{1} \int_{\frac{3}{4 \cdot x_1}}^{1} 2 \cdot \theta^2 \cdot x_1^{\theta-1} \cdot x_2^{\theta-1} \ dx_2 dx_1 \\ &= 2 \cdot \theta^2 \int_{\frac{3}{4}}^{1} x_1^{\theta-1} \cdot (\frac{1}{\theta} \cdot x_2^{\theta-1} \Big|_{\frac{3}{4x_1}}^{1}) \ dx_1 \\ &= \frac{2 \cdot \theta^2}{\theta} \int_{\frac{3}{4}}^{1} x_1^{\theta-1} \cdot (1 - (\frac{3}{4x_1})^{\theta}) \ dx_1 \\ &= \frac{2 \cdot \theta^2}{\theta} \int_{\frac{3}{4}}^{1} x_1^{\theta-1} - (\frac{3}{4})^{\theta} \cdot \frac{1}{x_1} \ dx_1 \end{aligned}$$

$$\begin{aligned}
\Rightarrow P(\frac{3}{4} \le x_1 \cdot x_2) &= 2\theta \cdot (\frac{1}{\theta} \cdot x_1^\theta - (\frac{3}{4})^\theta \cdot \log x_1)\Big|_{\frac{3}{4}}^{1} \\
&= 2\theta \cdot (\frac{1}{\theta} - (\frac{3}{4})^\theta \cdot \log 1 - \frac{1}{\theta} \cdot (\frac{3}{4})^\theta + (\frac{3}{4})^\theta \cdot \log(\frac{3}{4})) \\
&= 2\theta \cdot (\frac{1}{\theta} - 0 - \frac{1}{\theta} \cdot (\frac{3}{4})^\theta + (\frac{3}{4})^\theta \cdot \log(\frac{3}{4})) \\
&= 2 \cdot (1 - (\frac{3}{4})^\theta + \theta \cdot (\frac{3}{4})^\theta \cdot \log(\frac{3}{4})), \quad \theta = 1, 2.
\end{aligned}$$

Here's another approach, for a general θ, the probability of rejecting H_0 is

$$\gamma(\theta) = \int_{3/4}^{1} \int_{3/4x_1}^{1} \theta^2 \cdot x_1^{\theta-1} \cdot x_2^{\theta-1}\, dx_2 dx_1 = 1 - (\frac{3}{4})^\theta + \theta \cdot (\frac{3}{4})^\theta \cdot \log(\frac{3}{4})$$

where $\gamma(1)$ is the significance level and $\gamma(2)$ is the power when $\theta = 2$.

■

7. **(Exercise 4.6.2)** Consider the power function $\gamma(\mu)$ and its derivative $\gamma'(\mu)$ given by equations (4.6.5) and (4.6.6), on page 249. Show that $\gamma'(\mu)$ is strictly negative for $\mu < \mu_0$ and strictly positive for $\mu > \mu_0$.

Solution:

$$\begin{aligned}
\gamma(\mu) &= P_\mu(\bar{x} \le \mu_0 - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}) + P_\mu(\bar{x} \ge \mu_0 + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{N}}) \\
&= \phi(\frac{\sqrt{N} \cdot (\mu_0 - \mu)}{\sigma} - z_{\frac{\alpha}{2}}) + 1 - \phi(\frac{\sqrt{N} \cdot (\mu_0 - \mu)}{\sigma} + z_{\frac{\alpha}{2}}) \\
\gamma'(\mu) &= \frac{\sqrt{N}}{\sigma} \cdot \left(\phi(\frac{\sqrt{N} \cdot (\mu_0 - \mu)}{\sigma} + z_{\frac{\alpha}{2}}) - \phi(\frac{\sqrt{N} \cdot (\mu_0 - \mu)}{\sigma} - z_{\frac{\alpha}{2}})\right)
\end{aligned}$$

$\therefore \gamma'(\mu)$ is strictly negative for $\mu < \mu_0$ and strictly positive for $\mu > \mu_0$.

Here's another approach, suppose $\mu > \mu_0$. Then

$$\Big|\frac{\sqrt{n} \cdot (\mu_0 - \mu)}{\sigma} + z_{\alpha/2}\Big| < \Big|\frac{\sqrt{n} \cdot (\mu_0 - \mu)}{\sigma} - z_{\alpha/2}\Big|$$

Hence,

$$\phi(\Big|\frac{\sqrt{n} \cdot (\mu_0 - \mu)}{\sigma} + z_{\alpha/2}\Big|) > \phi(\Big|\frac{\sqrt{n} \cdot (\mu_0 - \mu)}{\sigma} - z_{\alpha/2}\Big|)$$

Because $\phi(t)$ is symmetric to 0, $\phi(t) = \phi(|t|)$. This observation plus the last inequality shows that $\gamma'(\mu)$ is increasing, for $\mu > \mu_0$. Likewise for $\mu < \mu_0$, $\gamma'(\mu)$ is decreasing.

■

8. **(Exercise 6.1.1)** Let $X_1, X_2, \cdots, X_n$ be a random sample from a $\Gamma(\alpha = 3, \beta = \theta)$ distribution, $0 < \theta < \infty$. Determine the mle of θ.

Solution:

$$L(\theta, x) = \prod_{i=1}^{n} f(x_i; \theta), \quad x \in \Omega$$

$$X_1, X_2, \cdots, X_n \sim \text{iid}, \quad \theta \text{ is unknown.}$$

$$\hat{\theta} = \max_{\theta \in \Omega} L(\theta; x)$$

$$
\begin{aligned}
f(x;\alpha,\beta) &= \frac{x^{\alpha-1}}{\Gamma(\alpha)\cdot\beta^{\alpha}}\cdot e^{-\frac{x}{\beta}} = \frac{1}{\Gamma(3)\cdot\theta^{3}}\cdot x^{2}\cdot e^{-\frac{x}{\theta}}, \quad 0<x<1. \\
\Rightarrow \quad L(\theta) &= \prod_{i=1}^{n} f(x_i\ ;\ 3,\theta) = \prod_{i=1}^{n}\frac{1}{\Gamma(3)\cdot\theta^{3}}\cdot x_i^{2}\cdot e^{-\frac{x_i}{\theta}} \\
&= (\frac{1}{\Gamma(3)\cdot\theta^{3}})^{n}\cdot(\prod_{i=1}^{n} x_i^{2}\cdot e^{-\frac{x_i}{\theta}}) \\
&= (\frac{1}{\Gamma(3)\cdot\theta^{3}})^{n}\cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i}\cdot\prod_{i=1}^{n}x_i^{2} \\
\Rightarrow \quad l(\theta) &= \ln(L(\theta)) = \ln\left((\frac{1}{\Gamma(3)\cdot\theta^{3}})^{n}\cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i}\cdot\prod_{i=1}^{n}x_i^{2}\right) \\
&= -n\ln(\Gamma(3)) - 3n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n}x_i + 2\sum_{i=1}^{n}\ln(x_i) \\
\Rightarrow \quad & \frac{\partial}{\partial\theta}l(\theta) = 0 \\
\Rightarrow \quad & \frac{\partial}{\partial\theta}\left(-n\ln(\Gamma(3)) - 3n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n}x_i + 2\sum_{i=1}^{n}\ln(x_i)\right) = 0 \\
\Rightarrow & -\frac{3n}{\theta} + \frac{1}{\theta^{2}}\sum_{i=1}^{n}x_i = 0 \\
\Rightarrow \quad & \frac{3n}{\theta} = \frac{1}{\theta^{2}}\sum_{i=1}^{n}x_i \quad \rightarrow \quad 3n\theta = \sum_{i=1}^{n}x_i \\
\therefore \quad & \hat{\theta} = \frac{1}{3n}\sum_{i=1}^{n}x_i = \frac{1}{3}\bar{x}
\end{aligned}
$$

$\blacksquare$
