8.3.4 Solved Problems

Problem 1

Let X_1, X_2, X_3, $\ldots$, X_n be a random sample from an exponential distribution with parameter θ, i.e.,

$$f_{X_i}(x;\theta) = \theta e^{-\theta x} u(x).$$

Our goal is to find a $(1-\alpha)100\%$ confidence interval for θ. To do this, we need to remember a few facts about the gamma distribution. More specifically, If $Y = X_1 + X_2 + \cdots + X_n$, where the X_i's are independent $Exponential(\theta)$ random variables, then $Y \sim Gamma(n, \theta)$. Thus, the random variable Q defined as

$$Q = \theta(X_1 + X_2 + \cdots + X_n)$$

has a $Gamma(n, 1)$ distribution. Let us define $\gamma_{p,n}$ as follows. For any $p \in [0,1]$ and $n \in \mathbb{N}$, we define $\gamma_{p,n}$ as the real value for which

$$P(Q > \gamma_{p,n}) = p,$$

where $Q \sim Gamma(n, 1)$.

a. Explain why $Q = \theta(X_1 + X_2 + \cdots + X_n)$ is a pivotal quantity.
b. Using Q and the definition of $\gamma_{p,n}$, construct a $(1-\alpha)100\%$ confidence interval for θ.

Solution

a. Q is a function of the X_i's and θ, and its distribution does not depend on θ or any other unknown parameters. Thus, Q is a pivotal quantity.
b. Using the definition of $\gamma_{p,n}$, a $(1-\alpha)$ interval for Q can be stated as

$$P\left(\gamma_{1-\frac{\alpha}{2},n-1} \leq Q \leq \gamma_{\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

Therefore,

$$P\left(\gamma_{1-\frac{\alpha}{2},n-1} \leq \theta(X_1 + X_2 + \cdots + X_n) \leq \gamma_{\frac{\alpha}{2},n-1}\right) = 1-\alpha.$$

Since $X_1 + X_2 + \cdots + X_n$ is always a positive quantity, the above equation is equivalent to

$$P\left(\frac{\gamma_{1-\frac{\alpha}{2},n-1}}{X_1 + X_2 + \cdots + X_n} \leq \theta \leq \frac{\gamma_{\frac{\alpha}{2},n-1}}{X_1 + X_2 + \cdots + X_n}\right) = 1 - \alpha.$$

We conclude that $\left[\frac{\gamma_{1-\frac{\alpha}{2},n-1}}{X_1+X_2+\cdots+X_n}, \frac{\gamma_{\frac{\alpha}{2},n-1}}{X_1+X_2+\cdots+X_n}\right]$ is a $(1-\alpha)100\%$ confidence interval for θ.

Problem 2

A random sample X_1, X_2, X_3, ..., X_{100} is given from a distribution with known variance $\mathrm{Var}(X_i) = 16$. For the observed sample, the sample mean is $\overline{X} = 23.5$. Find an approximate 95% confidence interval for $\theta = EX_i$.

Solution

Here, $\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ is an approximate $(1-\alpha)100\%$ confidence interval. Since $\alpha = 0.05$, we have

$$z_{\frac{\alpha}{2}} = z_{0.025} = \Phi^{-1}(1 - 0.025) = 1.96$$

Also, $\sigma = 4$. Therefore, the approximate confidence interval is

$$\left[23.5 - 1.96\frac{4}{\sqrt{100}}, 23.5 - 1.96\frac{4}{\sqrt{100}}\right] \approx [22.7, 24.3].$$

Problem 3

To estimate the portion of voters who plan to vote for Candidate A in an election, a random sample of size n from the voters is chosen. The sampling is done with replacement. Let θ be the portion of voters who plan to vote for Candidate A among all voters. How large does n need to be so that we can obtain a 90% confidence interval with 3% margin of error? That is, how large n needs to be such that

$$P\left(\overline{X} - 0.03 \leq \theta \leq \overline{X} + 0.03\right) \geq 0.90,$$

where $\overline{X}$ is the portion of people in our random sample that say they plan to vote for Candidate A.

Solution

Here,

$$\left[\overline{X} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}, \overline{X} + \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n}}\right]$$

is an approximate $(1-\alpha)100\%$ confidence interval for θ. Since $\alpha = 0.1$, we have

$$z_{\frac{\alpha}{2}} = z_{0.05} = \Phi^{-1}(1-0.05) = 1.645$$

Therefore, we need to have

$$\frac{1.645}{2\sqrt{n}} = 0.03$$

Therefore, we obtain

$$n = \left(\frac{1.645}{2 \times 0.03}\right)^2.$$

We conclude $n \geq 752$ is enough.

Problem 4

a. Let X be a random variable such that $R_X \subset [a,b]$, i.e., we always have $a \leq X \leq b$. Show that

$$\mathrm{Var}(X) \leq \frac{(b-a)^2}{4}.$$

b. Let X_1, X_2, X_3, $\ldots$, X_n be a random sample from an unknown distribution with CDF $F_X(x)$ such that $R_X \subset [a,b]$. Specifically, EX and $\mathrm{Var}(X)$ are unknown. Find a $(1-\alpha)100\%$ confidence interval for $\theta = EX$. Assume that n is large.

Solution

a. Define $Y = X - \frac{a+b}{2}$. Thus, $R_Y \subset [-\frac{b-a}{2}, \frac{b-a}{2}]$. Then,

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(Y) \\ &= E[Y^2] - \mu_Y^2 \\ &\leq E[Y^2] \\ &\leq \left(\frac{b-a}{2}\right)^2 \quad \left(\text{since } Y^2 \leq \left(\frac{b-a}{2}\right)^2\right) \\ &= \frac{(b-a)^2}{4}.\end{aligned}$$

b. Here, we have an upper bound on σ, which is $\sigma_{max} = \frac{(b-a)}{2}$. Thus, the interval

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma_{max}}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma_{max}}{\sqrt{n}}\right]$$

is a $(1-\alpha)100\%$ confidence interval for θ. More specifically,

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{b-a}{2\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{b-a}{2\sqrt{n}}\right]$$

is a $(1-\alpha)100\%$ confidence interval for θ.

Problem 5

A random sample $X_1, X_2, X_3, \ldots, X_{144}$ is given from a distribution with unknown variance $\mathrm{Var}(X_i) = \sigma^2$. For the observed sample, the sample mean is $\overline{X} = 55.2$, and the sample variance is $S^2 = 34.5$. Find a 99% confidence interval for $\theta = EX_i$.

Solution

The interval

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

is approximately a $(1-\alpha)100\%$ confidence interval for θ. Here, $n = 144$, $\alpha = 0.01$, so we need

$$z_{\frac{\alpha}{2}} = z_{0.005} = \Phi^{-1}(1 - 0.005) \approx 2.58$$

Thus, we can obtain a 99% confidence interval for θ as

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right] = \left[55.2 - 2.58 \cdot \frac{\sqrt{34.5}}{12}, 55.2 + 2.58 \cdot \frac{\sqrt{34.5}}{12}\right]$$
$$\approx [53.94, 56.46].$$

Therefore, $[53.94, 56.46]$ is an approximate 99% confidence interval for θ.

Problem 6

A random sample X_1, X_2, X_3, $\ldots$, X_{16} is given from a normal distribution with unknown mean $\mu = EX_i$ and unknown variance $\mathrm{Var}(X_i) = \sigma^2$. For the observed sample, the sample mean is $\overline{X} = 16.7$, and the sample variance is $S^2 = 7.5$.

a. Find a 95% confidence interval for μ.
b. Find a 95% confidence interval for σ^2.

Solution

a. Here, the interval

$$\left[\overline{X} - t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}\right]$$

is a $(1-\alpha)100\%$ confidence interval for μ. Let $n = 16$, $\alpha = 0.05$, then

$$t_{0.025,15} \approx 2.13$$

The above value can obtained in MATLAB using the command `tinv(0.975, 15)`. Thus, we can obtain a 95% confidence interval for μ as

$$\left[16.7 - 2.13 \frac{\sqrt{7.5}}{4}, 16.7 + 2.13 \frac{\sqrt{7.5}}{4}\right] \approx [15.24, 18.16].$$

Therefore, $[15.24, 18.16]$ is a 95% confidence interval for μ.

b. Here, $\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$ is a $(1-\alpha)100\%$ confidence interval for σ^2. In this problem, $n = 16$, $\alpha = .05$, so we need

$$\chi^2_{0.025,15} \approx 27.49, \quad \chi^2_{0.975,15} \approx 6.26$$

The above values can obtained in MATLAB using the commands `chi2inv(0.975, 15)` and `chi2inv(0.025, 15)`, respectively. Thus, we can obtain a 95% confidence interval for σ^2 as

$$\begin{aligned} \left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right] &= \left[\frac{15\times 7.5}{27.49}, \frac{15\times 7.5}{6.26}\right] \\ &\approx [4.09, 17.97]. \end{aligned}$$

Therefore, $[4.09, 17.97]$ is a 95% confidence interval for σ^2.
