1. Exercise 4.5.8 on Page 247

Let us say the life of a tire in miles, say X, is normally distributed with mean θ and standard deviation 5000. Past experience indicates that $\theta = 30,000$. The manufacturer claims that the tires made by a new process have mean $\theta > 30,000$. It is possible that $\theta = 35,000$. Check his claim by testing $H_0 : \theta = 30,000$ against $H_1 : \theta > 30,000$. We observe n independent values of X, say $x_1, ..., x_n$, and we reject H_0 (thus accept H_1) if and only if $\bar{x} \geq c$. Determine n and c so that the power function $\gamma(\theta)$ of the test has the values $\gamma(30,000) = 0.01$ and $\gamma(35,000) = 0.98$.

Answer:

Power function is $1 - \beta$, where β is the type II error. So that we have

$$\begin{aligned} 1 - \beta &= Pr(\text{reject} \quad H_0 | H_1 \quad \text{is true}) \\ &= \gamma(\theta) \\ &= Pr(\bar{x} \geq c | \theta > 30,000) \\ &= Pr(Z \geq \frac{c - \theta}{5000/\sqrt{n}} | \theta > 30,000) \end{aligned} \tag{1}$$

Therefore, we have two equations, two unknown parameters:

$$\begin{aligned} & \gamma(30000) = Pr(Z \geq \frac{c - 30000}{5000/\sqrt{n}}) = 0.01 \\ & \gamma(35000) = Pr(Z \geq \frac{c - 35000}{5000/\sqrt{n}}) = 0.98 \\ \Rightarrow \quad & \frac{c - 30000}{5000/\sqrt{n}} = 2.326348 \\ & \frac{c - 35000}{5000/\sqrt{n}} = -2.053749 \\ \Rightarrow \quad & c = 32655.59, \qquad n = 19.18525 \end{aligned} \tag{2}$$

2. Exercise 4.6.7 on Page 253

Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles in $\mu g/m^3$. Let X and Y equal the concentration of suspended particles in $\mu g/m^3$ in the city center (commercial district) for Melbourne and Houston, respectively. Using n = 13 observations of X and $m = 16$ observations of Y , we test $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$.

(a) Define the test statistic and critical region, assuming that the unknown variances are equal. Let $\alpha = 0.05$.

(b) If $\bar{x} = 72.9, s_x = 25.6, \bar{y} = 81.7$, and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.

Answer:

As we assume the unknown variance are equal, we have the pooled variance,

$$T = \frac{\bar{X} - \bar{Y} - 0}{S_p\sqrt{1/n_1 + 1/n_2}}$$

where

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

As this is a one tail test, The Critical Region is given as :

$$\{T < t_{n_1+n_2-2,0.05}\} = t_{13+16-2,0.05} = -1.703288$$

Next, as the data given, we have

$$T = \frac{72.9 - 81.7}{\sqrt{\frac{12\times 25.6^2 + 15\times 28.3^2}{13+16-2}}\sqrt{\frac{1}{13}+\frac{1}{16}}} = -0.8685893$$

Therefore, we do not reject H_0 and we state that we do not have sufficient evidence to support that μ_x is less than μ_y.

3. Exercise 4.7.2 on Page 260

Define the sets $A_1 = \{x : -\infty < x \leq 0\}$, $A_i = \{x : i-2 < x \leq i-1\}$, $i = 2, ..., 7$, and $A_8 = \{x : 6 < x < \infty\}$. A certain hypothesis assigns probabilities p_{i0} to these sets A_i in accordance with

$$p_{i0} = \int_{A_i} \frac{1}{2\sqrt{2\pi}} \exp[-\frac{(x-3)^2}{8}]dx$$

, $i = 1, 2, ..., 7, 8$. This hypothesis (concerning the multinomial pdf with $k = 8$) is to be tested, at the 5% level of significance, by a chi-square test. If the observed frequencies of the sets $A_i, i = 1, 2, ..., 8$, are, respectively, 60, 96, 140, 210, 172, 160, 88, and 74, would H_0 be accepted at the (approximate) 5% level of significance?

Answer:

As we notice that this is a normal distribution, we can

$$p_{10} = Pr(\infty < X \leq 0) = Pr(\infty < Z \leq \frac{0-3}{2}) = Pr(\infty < Z \leq -1.5) = 0.0668072$$

$$p_{20} = Pr(0 < X \leq 1) = Pr(\frac{0-3}{2} < Z \leq \frac{1-3}{2}) = Pr(-1.5 < Z \leq -1) = 0.09184805$$

$$p_{30} = Pr(1 < X \leq 2) = Pr(\frac{1-3}{2} < Z \leq \frac{2-3}{2}) = Pr(-1 < Z \leq -0.5) = 0.1498823$$

$$p_{40} = Pr(2 < X \leq 3) = Pr(\frac{2-3}{2} < Z \leq \frac{3-3}{2}) = Pr(-0.5 < Z \leq 0) = 0.1914625$$

$$p_{50} = Pr(3 < X \leq 4) = Pr(\frac{3-3}{2} < Z \leq \frac{4-3}{2}) = Pr(0 < Z \leq 0.5) = 0.1914625$$
$$p_{60} = Pr(4 < X \leq 5) = Pr(\frac{4-3}{2} < Z \leq \frac{5-3}{2}) = Pr(0.5 < Z \leq 1) = 0.1498823$$
$$p_{70} = Pr(5 < X \leq 6) = Pr(\frac{5-3}{2} < Z \leq \frac{6-3}{2}) = Pr(1 < Z \leq 1.5) = 0.09184805$$
$$p_{70} = Pr(6 < X \leq \infty) = Pr(\frac{5-3}{2} < Z \leq \frac{6-3}{2}) = Pr(1 < Z \leq 1.5) = 0.0668072$$

Hence, the χ^2 test:

$$\sum_{i=1}^{8} \frac{(X_i - np_{i0})^2}{np_{i0}} = 6.919$$

When $\alpha = 0.05, df = 7, \chi^2(7, 0.95) = 14.06714$, we do not reject H_0 and conclude that there is no evidence to say there are significant differences between the distributions.

4. Exercise 5.3.2 on Page 313

Let X denote the mean of a random sample of size 128 from a gamma distribution with $\alpha = 2$ and $\beta = 4$. Approximate $P(7 < X < 9)$.

Answer:

$$\mathbb{E}(X) = \alpha\beta = 8 \qquad \mathbb{V}(X) = \alpha\beta^2 = 32$$

Then, by CLT,

$$P(7 < \bar{X} < 9) = P(\frac{7-8}{1/2} < Z < \frac{9-8}{1/2}) = \Phi(2) - \Phi(-2) = 0.9544997 \tag{3}$$

5. Exercise 6.1.2 on Page 326

Let $X_1, X_2, ..., X_n$ represent a random sample from each of the distributions having the following pdfs:

(a) $f(x;\theta) = \theta x^{\theta-1}$,$0 < x < 1, 0 < \theta < \infty$, zero elsewhere.

(b) $f(x;\theta) = e^{-(x-\theta)}$, $\theta \leq x < \infty, -\infty < \theta < \infty$, zero elsewhere. Note this is a nonregular case. In each case find the mle $\hat{\theta}$ of θ.

Answer:

(a)

$$L(\theta; x_i) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ell(\theta; x_i) = n\log(\theta) + (\theta - 1)\sum_{i=1}^{n} \log(x_i)$$

$$\Rightarrow \qquad \frac{\partial \ell(\theta; x_i)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(x_i) = 0 \tag{4}$$

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log(x_i)}$$

(b)

$$\begin{aligned} L(\theta; x_i) &= \prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-(\sum_{i=1}^{n} x_i - n\theta)} \\ \ell(\theta; x_i) &= -(\sum_{i=1}^{n} x_i - n\theta) \\ \Rightarrow \quad \frac{\partial \ell(\theta; x_i)}{\partial \theta} &= n > 0 \end{aligned} \tag{5}$$

Therefore, as $\theta \le x_i$, $i = 1, 2, ..., n$. Then

$$-\infty < \theta < \min(X_i)$$

$$\hat{\theta} = \min(X_i)$$

6. Exercise 6.2.8 on Page 340

Let X be $N(0, \theta)$, $0 < \theta < \infty$.

(a) Find the Fisher information $I(\theta)$.

(b) If $X_1, X_2, ..., X_n$ is a random sample from this distribution, show that the mle of θ is an efficient estimator of θ.

(c) What is the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$?

Answer:

(a)

$$\begin{aligned} \log f(x;\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2\theta} \\ \frac{\partial \log f(x;\theta)}{\partial \theta} &= -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} \\ \frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} &= \frac{1}{2\theta^2} - \frac{x^2}{\theta^3} \\ I(\theta) &= -\mathbb{E}(\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2}) = \frac{\mathbb{E}(X^2)}{\theta^3} - \frac{1}{2\theta^2} = \frac{1}{2\theta^2} \\ \text{where} \quad & \mathbb{E}(X^2) = 0^2 + \theta = \theta \\ & \text{or} \\ I(\theta) &= \mathbb{E}[(\frac{\partial \log f(x;\theta)}{\partial \theta})^2] \end{aligned} \tag{6}$$

(b)As for the mle we have

$$\begin{aligned} L(\theta; x_i) &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= (2\pi\theta)^{-n/2} e^{-\sum_{i=1}^{n} x_i^2/2\theta} \end{aligned} \tag{7}$$

$$\frac{\partial \log(L(\theta))}{\partial \theta} = -\frac{n}{2\theta} + \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} = 0 \tag{8}$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i^2}{n} \tag{9}$$

$$\begin{aligned} Var(\hat{\theta}) &= Var(\frac{\sum_{i=1}^{n} x_i^2}{n}) \\ &= \frac{1}{n^2} Var(\sum_{i=1}^{n} x_i^2) \\ &= \frac{1}{n} Var(X^2) \\ &= \frac{1}{n}(E(X^4) - (E(X^2))^2) \\ &= \frac{2\theta^2}{n} \end{aligned} \tag{10}$$

where

$$E(X^4) = 3\theta^2 \qquad \text{You may use the MGF to derive} \tag{11}$$

which attains the Rao-Cramér bounds, thus, mle is an efficient estimator.

(c)As

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \frac{1}{I(\theta)})$$

then,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, 2\theta^2)$$