Tutorial 8

The University of Hong Kong

1 Recap

Estimator properties

1. Confidence Intervals

2 This week

1. Confidence Intervals

2.1 Interval Estimation

- An interval estimator of θ is a random interval $[L(\mathbf{X}), U(\mathbf{X})]$, where $L(\mathbf{X}) := L\left(X_1, \ldots, X_n\right)$ and $U(\mathbf{X}) := U\left(X_1, \ldots, X_n\right)$ are two statistics such that $L(\mathbf{X}) \leq U(\mathbf{X})$ with probability 1 .

- If $\mathbf{X} = \mathbf{x}$ is observed, $[L(\mathbf{x}), U(\mathbf{x})]$ is the interval estimate of θ.

- For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of θ, the confidence coefficient $[L(\mathbf{X}), U(\mathbf{X})]$, denoted by $(1-\alpha)$ is the infimum of the coverage probabilities $\mathrm{P}(\theta \in [L(\mathbf{x}), U(\mathbf{x})])$, that is,

$$1-\alpha = \inf_{\theta} \mathrm{P}(\theta \in [L(\mathbf{x}), U(\mathbf{x})]).$$

- A random variable $Q(\mathbf{X}, \theta) = Q\left(X_1, \ldots, X_n, \theta\right)$ is a pivotal quantity if the distribution of $Q(\mathbf{X}, \theta)$ is free of θ. That is regardless of distribution of $\mathbf{X}, Q(\mathbf{X}, \theta)$ has the same distribution of all values of θ.

2.2 Confidence Intervals for Means

Let $\mathbf{X} = \{X_1, \ldots, X_n\}$ be an independent random sample from the population $N\left(\mu, \sigma^2\right)$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

- If σ^2 is known, a $1-\alpha$ confident interval of μ is

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

given that the observed value of $\bar{X}$ is $\bar{x}$.

- If σ^2 is unknown, a $1-\alpha$ confident interval of μ is

$$\left[\bar{x} - t_{\alpha/2, df=n-1}\frac{s}{\sqrt{n-1}} \quad , \quad \bar{x} + t_{\alpha/2, df=\mathrm{n}-1}\frac{s}{\sqrt{n-1}}\right]$$

given that the observed values of $\bar{X}$ and S are $\bar{x}$ and s respectively.

2.3 Distribution Theory

Let $\mathbf{X} = \{X_1, \ldots, X_n\}$ be an independent random sample from the population $N\left(\mu, \sigma^2\right)$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

- $\bar{X}$ and S^2 are independent.

- If $Z_1, \ldots, Z_k$ are k independent $N(0,1)$ random variables, then $\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$.

- $\frac{nS^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi_{n-1}^2$.

- If $Z \sim N(0,1), U \sim \chi_k^2$ and Z and U are independent, then $T = \frac{Z}{\sqrt{U/k}} \sim t_k$.

- $\frac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1}$.

- When $n \to \infty$, the distribution function of t_n tends to the distribution function of $N(0,1)$.

3 Exercise

1. A publisher wants to estimate the mean length of time (in minutes) all adults spend on reading newspaper. To determine this estimate, the publisher takes a random sample of 15 people and obtains the following results

$$11, 9, 8, 10, 10, 9, 7, 11, 11, 7, 6, 9, 10, 8, 10$$

Assume that the population of reading times is normally distributed.

a. Construct a 90% confidence interval for the population mean if the publisher assumes that σ is 1.5 minutes from past studies.

b. Construct a 90% confidence interval for the population mean if the publisher assumes that σ is unknown.

Solution:

a. Normal population with known population variance. Let μ be the population, mean length of time all adults spent on reading news paper.

$$\begin{aligned} CI_{90\%}(\mu) &= \left[\bar{x} \pm z_{0.05}\frac{\sigma}{\sqrt{n}}\right] \qquad \text{Note: } \bar{x} = 9.0667, z_{0.05} = 1.645 \\ &= \left[9.0667 \pm 1.645\frac{1.5}{\sqrt{15}}\right] \\ &= [8.4296, 9.7038] \end{aligned}$$

b. Normal population with unknown population variance.

$$\begin{aligned} CI_{90\%}(\mu) &= \left[\bar{x} \pm t_{0.05,n-1}\frac{s}{\sqrt{n-1}}\right] \qquad \text{Note: } s \approx 1.5261, t_{0.05,14} = 1.761 \\ &= \left[\bar{x} \pm t_{0.05,14}\frac{s}{\sqrt{14}}\right] \\ &= \left[9.0667 \pm 1.761 \times \frac{1.5261}{\sqrt{14}}\right] \\ &\approx [8.3484, 9.7850] \end{aligned}$$

2. An education officer wants to use the mean of a random sample of 16 students to estimate the mean score (μ) that all students of the city would get if they took a certain MAF-test. Based on experience, the officer knows that the population has a normal distribution with $\sigma = 35$.

a. What is the width of the 95% confidence interval?

b. Suppose that sample data are:

$$137, 125, 184, 116, 98, 163, 150, 145, 178, 163, 151, 103, 75, 86, 193, 141$$

Construct a 90% confidence interval for μ.

c. If the education officer wants to obtain a 99% confidence interval for with a width not exceeding 20 points, how many students must be included for sample test?

Solution:

a. Normal population with known population variance.

$$CI_{95\%}(\mu) = \left[\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}}\right]$$

Width of $CI = 2z_{0.025}\frac{\sigma}{\sqrt{n}} = 2 \times 1.96 \times \frac{35}{\sqrt{16}} = 34.3$

b. Normal population with known population variance.

$$\begin{aligned} CI_{90\%}(\mu) &= \left[\bar{x} \pm z_{0.05}\frac{\sigma}{\sqrt{n}}\right] \\ &= \left[138 \pm 1.645 \times \frac{35}{\sqrt{16}}\right] \approx [123.6063; 152.3938] \quad \bar{x} = 138, \quad z_{0.05} = 1.645. \end{aligned}$$

c.

$$
\begin{aligned}
CI &= 2z_{0.005}\frac{\sigma}{\sqrt{n}} \leqslant 20\\
&\Rightarrow n \geqslant \left(\frac{2z_{0.005}\sigma}{20}\right)^2 \approx 81.2883\\
&\therefore n = 82
\end{aligned}
$$

3. Consider a random sample $T_1, T_2, \ldots, T_n$ drawn from an exponential distribution with rate $\lambda > 0$.

a. Show that $2\lambda \sum T_i$ has a Chi-squared distribution with $2n$ degrees of freedom. (Hint: Chi-squared distribution with n degrees of freedom is the same as Gamma $(0.5n, 0.5)$, the Gamma distribution with the shape parameter equal to $0.5n$ and the rate parameter equal to 0.5 . The pdf of Gamma(α, β) is $f(x) = \beta^\alpha x^{\alpha-1}\mathrm{e}^{-\beta x}/\Gamma(\alpha)$.)

b. Find an equal-tailed confidence interval for λ with α level.

Solution:

a.

$$T_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda), \quad \lambda > 0 \tag{1}$$

$$\therefore M_{T_i}(t) = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda \tag{2}$$

$$\text{Let } Y = 2\lambda \sum_{i=1}^{n} T_i \tag{3}$$

Let's prove the mgf of Y is the same as the mgf of $\Gamma\left(\frac{2n}{2}, \frac{1}{2}\right)$, i.e., χ^2_{2n}.

$$
\begin{aligned}
&M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t2\lambda\Sigma T_i}\right) = E\left(e^{2\lambda t T_1}e^{2\lambda t T_2}\cdots e^{2\lambda t T_n}\right)\\
&\overset{\text{ind.}}{=} E\left(e^{2\lambda t T_1}\right)E\left(e^{2\lambda t T_2}\right)\cdots eE\left(e^{2\lambda t T_n}\right)\\
&\overset{\text{identical}}{=} \left[M_{Ti}(2\lambda t)\right]^n\\
&= \left[\frac{\lambda}{\lambda - 2\lambda t}\right]^n = \left(\frac{1}{1-2t}\right)^n = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^n\\
&\therefore Y \sim \Gamma\left(n, \frac{1}{2}\right) \equiv \Gamma\left(\frac{2n}{2}, \frac{1}{2}\right) \equiv \chi^2_{2n}
\end{aligned}
$$

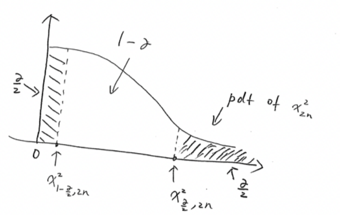
b.

$$\because Y = 2\lambda \sum_{i=1}^{n} T_i \sim \chi^2_{2n}$$

$$\therefore 1-\alpha = P\left(\chi^2_{1-\frac{\alpha}{2},2n} < 2\lambda \sum_{i=1}^{n} T_i < \chi^2_{\frac{\alpha}{2},2n}\right)$$

$$= P\left(\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i} < \lambda < \frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}\right)$$

$$\therefore CI_{(1-\alpha)}(\lambda) = \left[\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}, \frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}\right]$$



4. Suppose that we obtained a single observation Y from an exponential distribution with mean $\theta, f(y) = \frac{1}{\theta}\mathrm{e}^{-y/\theta} I(y \geq 0)$. Use Y to form a 90% confidence interval for θ.

Solution:

First of all, let's find the cdf of Y.

$$F_Y(y) = \int_0^y f_Y(t)dt = \int_0^y \frac{1}{\theta} e^{-\frac{t}{\theta}} dt = \frac{1}{\theta}\left[e^{-\frac{t}{\theta}} \cdot (-\theta)\right]_0^y \tag{4}$$

$$= 1 - e^{-\frac{y}{\theta}} \quad \text{for } y > 0. \tag{5}$$

Now, let's find $[a, b]$ s.t. $p(Y < a) = 0.05$ and $P(Y < b = 0.95)$.

$$0.05 = P(Y < a) = F_Y(a) = 1 - e^{-\frac{a}{\theta}} \to e^{-\frac{a}{\theta}} = 0.95 \tag{6}$$

$$\to -\frac{a}{\theta} = \ln 0.95 \to a = -\theta \ln 0.95. \tag{7}$$

$$0.95 = p(Y < b) = F_Y(b) = 1 - e^{-\frac{b}{\theta}} \to e^{-\frac{b}{\theta}} = 0.05 \tag{8}$$

$$\to -\frac{b}{\theta} = \ln 0.05 \to b = -\theta \ln 0.05. \tag{9}$$

$$\therefore 0.90 = P(a < Y < b) \tag{10}$$

$$= P(\theta \ln 0.95 < Y < -\theta \ln 0.05) \tag{11}$$

$$= P(\frac{Y}{-\ln 0.05} < \theta < \frac{Y}{-\ln 0.95}) \tag{12}$$

$$= P(\frac{Y}{2.9957} < \theta < \frac{Y}{0.0513}). \tag{13}$$

$\therefore CI_{90\%}(\theta) = [\frac{y}{2.9957}, \frac{y}{0.0513}]$, where y is the single observation of Y.