Confidence intervals

- Suppose that $X = (X_1, \ldots, X_n)$ is a sample and the distribution of X depends on a parameter vector θ, where $\theta \in \Theta$. For $\alpha \in [0,1]$, suppose that $A(X)$ is a region determined by X such that

$$P(g(\theta) \in A(X)) \geq 1 - \alpha \text{ for all } \theta \in \Theta,$$

then $A(X)$ is called a $(1-\alpha)$ confidence set for $g(\theta)$. When $g(\theta) \in (\infty, \infty)$ and $A(X)$ is an interval, $A(X)$ is called a $(1-\alpha)$ confidence interval (信賴區間) for $g(\theta)$.

- A common approach to construct a $(1-\alpha)$ confidence set for $g(\theta)$ is given below.

  (i) Find a random quantity (a random variable or a random vector) $h(X, g(\theta))$ such that the distribution of $h(X, g(\theta))$ does not depend on θ.

  (ii) For a given α, find a set A_α so that $P(h(X, g(\theta)) \in A_\alpha) = 1 - \alpha$.

  (iii) A confidence set for $g(\theta)$ is $\{\tau : h(X, \tau) \in A_\alpha\}$.

  The random quantity $h(X, g(\theta))$ in (i) is called a pivotal quantity or a pivot. Note that a pivot does not have to be a statistic. If a pivot is a statistic, it is an ancillary statistic.

- Example 1. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Suppose that $\alpha \in (0,1)$. Let $z_{\alpha/2}$ be the quantity such that

$$P(-z_{\alpha/2} < N(0,1) < z_{\alpha/2}) = 1 - \alpha.$$

Find a $(1-\alpha)$ confidence set of (μ, σ).

Sol. Note that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1),$$

so

$$P(-z_{\alpha/2} < \sqrt{n}(\bar{X} - \mu)/\sigma < z_{\alpha/2}) = 1 - \alpha.$$

Then

$$\begin{aligned} &\{(\mu, \sigma) : -z_{\alpha/2} < \sqrt{n}(\bar{X} - \mu)/\sigma < z_{\alpha/2}, \sigma > 0\} \\ &= \{(\mu, \sigma) : \bar{X} - \sigma z_{\alpha/2}/\sqrt{n} < \mu < \bar{X} + \sigma z_{\alpha/2}/\sqrt{n}, \sigma > 0\} \end{aligned}$$

is a $(1-\alpha)$ confidence set of (μ, σ).

- Definition. For $\alpha > 0$, $\beta > 0$, the gamma distribution $\Gamma(\alpha, \beta)$ is the distribution with PDF $f_{\alpha,\beta}$, where

$$f_{\alpha,\beta}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x)$$

for $x \in (-\infty, \infty)$.

- Definition. For $r > 0$, $\Gamma(r/2, 2)$ is called the chi-squared distribution with r degrees of freedom, denoted by $\chi^2(r)$.

  Fact 1 *Suppose that Z_1, …, Z_r are IID $N(0,1)$ random variables, then $\sum_{i=1}^r Z_i^2 \sim \chi^2(r)$.*

  The fact can be established by verifying that the MGF of $\Gamma(r/2, 2) = \chi^2(r)$ and the MGF of $\sum_{i=1}^r Z_i^2$ are the same.

- Example 2. Suppose that X_1, …, X_n are IID and $X_1 \sim N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Let $\bar{X} = \sum_{i=1}^n X_i/n$. Show that

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

  Proof. Let $Z_i = (X_i - \mu)/\sigma$ for $i = 1$, …, n and let P be an $n \times n$ matrix such that the first row of P is $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ and $PP^T = I_n$, where I_n is the identity matrix of size $n \times n$. Let $Z = (Z_1, \ldots, Z_n)^T$ and let $(Y_1, \ldots, Y_n)^T = PZ$, then Y_1, …, Y_n are IID and $Y_1 \sim N(0,1)$, $Y_1 = \sum_{i=1}^n Z_i/\sqrt{n} = \sqrt{n}\bar{Z}$, where $\bar{Z} = \sum_{i=1}^n Z_i/n$, and

$$\sum_{i=1}^n Y_i^2 = (PZ)^T PZ = \sum_{i=1}^n Z_i^2,$$

  which gives

$$\sum_{i=1}^n Z_i^2 - n(\bar{Z})^2 = \left(\sum_{i=1}^n Y_i^2\right) - \left(\underbrace{\sqrt{n}\bar{Z}}_{Y_1}\right)^2 = \sum_{i=2}^n Y_i^2 \sim \chi^2(n-1).$$

  Thus

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n(\bar{Z})^2 \sim \chi^2(n-1).$$

- Example 3. Suppose that $(X_1, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Suppose that $\alpha \in (0,1)$ and $c_{1,\alpha}$ and $c_{2,\alpha}$ are two constants such that $P(c_{1,\alpha} < \chi^2(n-1) < c_{2,\alpha}) = 1 - \alpha$. Find a $(1-\alpha)$ confidence interval for σ^2.

  Ans. $(\sum_{i=1}^n (X_i - \bar{X})^2/c_{2,\alpha}, \sum_{i=1}^n (X_i - \bar{X})^2/c_{1,\alpha})$.

- Definition. Suppose that random variables Z and V are independent, $Z \sim N(0,1)$ and $V \sim \chi^2(r)$. The distribution of $Z/\sqrt{V/r}$ is called the t distribution with r degrees of freedom, denoted by $t(r)$.

  - It can be shown that $t(r)$ has a PDF f such that $f(x) = f(-x)$ for $x \in (-\infty, \infty)$.

- Example 4. Suppose that $(X_1, \dots, X_n)$ is a random sample from $N(\mu, \sigma^2)$. Show that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1),$$

where $\bar{X}$ is the sample mean and S is the sample standard deviation.

Proof. Note that

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{Z}{\sqrt{V/(n-1)}},$$

where

$$Z = \sqrt{n}(\bar{X} - \mu)/\sigma \sim N(0, 1),$$
$$V = (n-1)S^2/\sigma^2 \sim \chi^2(n-1),$$

and Z and V are independent since $\bar{X}$ and $(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$ are independent. Therefore,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} = \frac{Z}{\sqrt{V/(n-1)}} \sim t(n-1).$$

- Example 5. Suppose that $(X_1, \dots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Suppose that $\alpha \in (0, 1)$ and $t_{n-1,\alpha/2}$ is a constant such that $P(-t_{n-1,\alpha/2} < t(n-1) < t_{n-1,\alpha/2}) = 1-\alpha$. Find a $(1-\alpha)$ confidence interval for μ.

  Ans. $(\bar{X} - t_{n-1,\alpha/2}S/\sqrt{n}, \bar{X} + t_{n-1,\alpha/2}S/\sqrt{n})$.