

01:XXX:XXX - Homework n

Pranav Tikkawar

November 15, 2024

1 Chapter 11: Confidence Intervals

Given a α such that $0 < \alpha < 1$, an Interval Estimation strategy provides two statistics (r.v) L and R s.t $P(L < \theta < R) = 1 - \alpha$. The interval $[L, R]$ is called a $(1 - \alpha)\%$ confidence interval for θ .

We can say that if you repeat the experiment N times and get N intervals, then $(1 - \alpha)\%$ of the intervals will contain the true value of θ .

Definition (Confidence Interval). A confidence interval with $(1-\alpha)$ confidence level are two statistics L and R such that $P(L < \theta < R) = 1 - \alpha$.

Remark. If someone says after an experiment that $2 < \lambda < 2.1$ with 90% confidence, it means that on average that 90% of the intervals will contain the true value of λ .

We want CI to be symmetric about \bar{X}

CI so far: $N(\mu, \sigma^2)$ population

$$1. \sigma^2 \text{ known: } \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$2. \sigma^2 \text{ unknown: } \bar{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

We can think of $z_{\alpha/2}$ in the normal curve as the shaded area in the tails.

New Context: Two pops $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and sample from both n_1 and n_2 from each

$$X_{11} \dots X_{1n_1} \sim N(\mu_1, \sigma_1^2)$$

$$X_{21} \dots X_{2n_2} \sim N(\mu_2, \sigma_2^2)$$

These are independent but not necessarily identically distributed.

Let \bar{X}_1 and \bar{X}_2 be the sample means and S_1^2 and S_2^2 be the sample variances.

Want CI for $\mu_1 - \mu_2$

Case 1: σ_1^2 and σ_2^2 are known.

We can use point estimators:

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Case 2: σ_1^2 and σ_2^2 are unknown. but $\sigma_1^2 = \sigma_2^2$

we can define a pooled sample variance:

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

Also

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

Remark. This is a weighted average of the sample variances. with weights $n_1 - 1$ and $n_2 - 1$

Remark.

$$S_p^2 := \frac{\sum^{n_1} (X_{1i} - \bar{X}_1)^2 + \sum^{n_2} (X_{2i} - \bar{X}_2)^2}{n_1 + n_2 - 2} \sim \chi_{n_1 + n_2 - 2}^2$$

Remark. Then the r.v T is

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

Consider making a CI for σ^2

Remark. We know that S^2 is a point estimator for σ^2

$$K = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Consider the graph of the χ^2 distribution. It is defined from $0, \infty$ and we can trap an α area in the tails.

We can call this $(\chi_{n-1}^2)_{\alpha/2}$ and $(\chi_{n-1}^2)_{1-(\alpha/2)}$

Notice the non symmetry of the χ^2 distribution.

We can say that $P\left((\chi_{n-1}^2)_{\alpha/2} < K < (\chi_{n-1}^2)_{1-(\alpha/2)}\right) = 1 - \alpha$

In a more insightful form we have

$$P\left(\frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2}\right) = 1 - \alpha$$

Definition (F distribution). F distribution with $\nu_1 > 0$ and $\nu_2 > 0$ degrees of freedom is defined by

$$F = \frac{\frac{U}{\nu_1}}{\frac{V}{\nu_2}}$$

where $U \sim \chi_{\nu_1}^2$ and $V \sim \chi_{\nu_2}^2$

ie $F = \frac{U\nu_2}{V\nu_1}$

Remark. Order is important

Example. $\{X_i\}_{i \in \text{range}(5)}$ independent standard normal r.v.

Then $\frac{\alpha(Z_1^2 + Z_2^2)}{Z_3^2 + Z_4^2 + Z_5^2} \sim F_{2,3}$ where $\alpha = 3/2$

This is due to the fact that $Z_i^2 \sim \chi_1^2$ and $\sum_i^n Z_i^2 \sim \chi_n^2$

Theorem 1. PDF of f_{ν_1, ν_2} is

$$g(x) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{\frac{\nu_1}{2}-1} \left(1 + \frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1 + \nu_2}{2}}$$

If we consider two normal pops $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ if

$$\frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2 \text{ and } \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$$

then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

This is useful for a CI for ratio of variances.

Hypothesis Testing

Definition (Hypothesis Testing). It is an assertion about the theoretical distribution. for example about parameters.

Example. In a coin toss experiment $C \sim \text{Ber}(p)$ "coin is unfair": $p \neq 0.5$

It is simple if it completely determines the distribution
It is not simple if it is composite

Example. Coin is fair (ie $p = 0.5$)
is a simple hypothesis

Definition (Statistical Test). it is a criterion used to decide whether to reject (or accept) one hypothesis called the null hypothesis H_0 in favor of an alternative hypothesis H_1 .

Chapter 12

Example. $N(\mu, \sigma^2 = 4)$ sample to get $\{X_i\}$ want a test to find $\bar{X} < C$

$H_0 : \mu = 4$

$H_1 : \mu = 1$

$\alpha = P(\text{reject } H_0 | H_0 \text{ is true})$

$\beta = P(\text{accept } H_0 | H_1 \text{ is true})$

We found that if α is small, then β is not if n is fixed.

Increasing n will decrease β

Example. Consider $\text{Exp}(\lambda)$ $H_0 : \lambda = 4$

$H_1 : \lambda = 1$

Suppose we get $\{X_i\}$

$\bar{X} < C$

Find α and β for a fixed $C > 0$ Assume sample size is 1

PDF is $f(x) = e^{-x/\lambda}/\lambda$

Thus $\alpha = P(\bar{X} < C | \lambda = 4)$

$\beta = P(\bar{X} > C | \lambda = 1)$

We can use the CDF of the exponential distribution to find these values.