Lecture Note 9 *
Interval Estimation and Confidence Intervals

MIT 14.30 Spring 2006
Herman Bennett

22 Interval Estimation

Interval estimation is another approach for estimating a parameter θ. Interval estimation consists in finding a random interval that contains the true parameter θ with probability $(1-\alpha)$. Such an interval is called confidence interval and the probability $(1-\alpha)$ is called the confidence level.

$$P\Big(A(X_1, ..., X_n) \leq \theta \leq B(X_1, ...X_n)\Big) = 1 - \alpha \tag{71}$$

Example 22.1. Assume a random sample from a $N(\mu, \sigma^2)$ population, where both parameters are unknown. Find the 90% (symmetric) confidence interval of σ^2.

*Caution: These notes are not necessarily self-explanatory notes. They are to be used as a complement to (and not as a substitute for) the lectures.

Example 22.2. Assume a random sample from a $N(\mu, \sigma^2)$ population, where the parameter μ is unknown and $\sigma = 2$. Find the 95% (symmetric) confidence interval of μ. How would your answer change if σ is unknown?

- The random interval $\left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right)$ contains the true parameter θ with 95% probability. It is **wrong** to say that θ lies in the interval with 95% probability...θ is not a RV!

23 Useful results

23.1 *t*-student distribution

A RV X is said to have a <u>t-student distribution</u> with parameter $v > 0$ (degrees of freedom) if the pdf of X is:

$$X \sim t_{(v)}: \quad f(x|v) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}} \frac{1}{(1 + (\frac{x^2}{v}))^{(v/2+0.5)}} \quad , \tag{72}$$

for $-\infty < x < \infty$ and v positive integer.

Let $X \sim N(0,1)$ and $Z \sim \chi^2_n$ be independent RVs. Then, the RV H is distritbuted t-student with n degrees of freedom.

$$H = \frac{X}{\sqrt{Z/n}} \sim t_{(n)} \tag{73}$$

- Symmetric distribution around 0, which implies that $t_{\alpha/2,n} = -t_{1-\alpha/2,n}$.
- As $n \to \infty$, $t(n) \to N(0,1)$.

(See attached graph and table).

Example 23.1. Important result. Assume a normally distributed random sample. Then, $\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{(n-1)}$. Prove this result.

23.2 *F* distribution

A RV X is said to have a F distribution with parameters $v_1 > 0$ and $v_2 > 0$ if the pdf of X is:

$$X \sim F_{(v_1,v_2)}: \quad f(x|v_1,v_2) = \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}\left(\frac{v_1}{v_2}\right)^{(v_1/2)} \frac{x^{(v_1/2-1)}}{(1+(\frac{v_1}{v_2})x)^{(v_1/2+v_2/2)}} \quad , \tag{74}$$

for $0 < x < \infty$ and v_i positive integer.

Let $X \sim \chi^2_n$ and $Z \sim \chi^2_m$ be independent RVs. Then, the RV G is distritbuted F with n and m degrees of freedom.

$$G = \frac{X/n}{Z/m} \sim F_{(n,m)} \tag{75}$$

24 Constructing Confidence Intervals for θ

In what follows, we consider 5 possible cases of limited information about the parameter(s) θ. For each case we study how to construct confidence intervals.

24.1 Case 1: $\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$ and $Var(\hat{\theta})$ known

We just saw an example of this case in Example 22.2. Note that in this example $\theta = \mu$, $\hat{\theta} = \bar{X}$, and the $Var(\hat{\theta})$ is known since σ is known.

24.2 Case 2: $\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$ and $Var(\hat{\theta})$ unknown

Example 24.1. Assume as in Example 22.2 a normal random sample, but now both μ and σ_2 are unknown. Construct a 95% confidence interval of μ.

24.3 Case 3: $\hat{\theta}$ not $\sim N(\ )$ but pmf/pdf known

We just saw an example of this case in Example 22.1. Note that in this example $\theta = \sigma^2$, $\hat{\theta} = S^2$, and the pdf of $\hat{\theta}$ (a function of $\hat{\theta}$ in this case) is known and depends only on one parameter, σ^2.

24.4 Case 4: $\hat{\theta} \sim$ not Normal, pmf/pdf unknown, and $n > 30$

Example 24.2. Assume a random sample of size n from a population $f(x)$, which is unknown. Construct a 99% confidence interval of $\mu = E(X_i)$.

24.5 Case 5: $\hat{\theta} \sim$ not Normal, pmf/pdf unknown, and $n < 30$