LECTURE - 8

Agenda:

1. Confidence Intervals
2. Examples

CONFIDENCE INTERVALS

As we discussed previously, many people are not satisfied by a "point" estimator $\hat{\theta}$ for the unknown parameter θ. What they are more comfortable with, is an interval based on the data, such that the probability of θ lying in this interval is very high.

DEFINITION: Let $\hat{\theta}_L$ and $\hat{\theta}_U$ be functions of the sample or the data, such that

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1-\alpha.$$

Then, $[\hat{\theta}_L, \hat{\theta}_U]$ is defined as a TWO-SIDED CONFIDENCE INTERVAL for θ with CONFIDENCE COEFFICIENT $1-\alpha$.

How do we find confidence intervals? Here is a method which is often successful in forming confidence intervals.

(1) FIND A FUNCTION $g(\text{Data}, \theta)$, i.e., a function of the data and the unknown parameter θ, such that the distribution of $g(\text{Data}, \theta)$ does not depend on the parameter θ. Such a function is known as a PIVOTAL QUANTITY.

<u>VERY IMPORTANT</u>: The function g cannot involve any unknown parameter other than θ.

(2) Find a and b such that

$$P(a \le g(\text{Data}, \theta) \le b) = 1 - \alpha.$$

(3) Reduce $a \le g(\text{Data}, \theta) \le b$ to the form $\hat{\theta}_L \le \theta \le \hat{\theta}_U$. Hence,

$$P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1 - \alpha,$$

which implies that $[\hat{\theta}_L, \hat{\theta}_U]$ is a confidence interval for θ with confidence level $(1 - \alpha)$.

<u>Example</u>: Let $Y_1, Y_2, ---, Y_n$ be a random sample from a population which is Normal with mean μ and variance σ^2. Find a two-sided confidence interval for μ with confidence coefficient 95%.

How do we start looking for a pivotal quantity here? Often, the standard "point" estimator is a good place to start. We know that the standard statistical estimator for μ is the sample mean $\bar{Y}$. Also,

$$\overline{Y} \text{ is Normal}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow \sqrt{n}\left(\frac{\overline{Y}-\mu}{\sigma}\right) \text{ is Normal}(0,1).$$

(1) If σ is known, we have our pivotal quantity.

$g(\text{Data}, \mu) = \dfrac{\sqrt{n}\,(\overline{Y}-\mu)}{\sigma}$ has a standard normal distribution which is independent of μ.

(2) Find a and b such that

$$P\left(a \le \frac{\sqrt{n}\,(\overline{Y}-\mu)}{\sigma} \le b\right) = 95\%.$$

We know from the properties of the Normal $(0, 1)$ distribution that

$$P\left(-1.96 \le \frac{\sqrt{n}\,(\overline{Y}-\mu)}{\sigma} \le 1.96\right) = 95\%.$$

(3)

$$-1.96 \le \frac{\sqrt{n}\,(\overline{Y}-\mu)}{\sigma} \le 1.96$$

$$\Leftrightarrow -\frac{1.96\sigma}{\sqrt{n}} \le \overline{Y}-\mu \le \frac{1.96\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \overline{Y}+\frac{1.96\sigma}{\sqrt{n}} \le \mu \le \overline{Y}+\frac{1.96\sigma}{\sqrt{n}}.$$

Hence a 95% confidence interval for μ is given by

$$\left[\bar{Y} - \frac{1.96\sigma}{\sqrt{n}}, \ \bar{Y} + \frac{1.96\sigma}{\sqrt{n}} \right].$$

BUT WHAT IF σ is UNKNOWN? Then $\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma}$ cannot be a pivotal quantity as it is also a function of another unknown parameter σ. We will proceed in the next lecture, but let us introduce the notion of "quantiles".

DEFINITION: A number b is said to be the $(1-\alpha)^{th}$ quantile of a random variable X, if

$$P(X \le b) = 1 - \alpha.$$

Since $P(X \le b) = \int_{-\infty}^{b} f_X(x)\,dx$ (if X is a continuous random variable),

this is same as saying

Area under the density of X before b = $1 - \alpha$