LECTURE - 11

Agenda:

1. Revisiting the central limit theorem.
2. Approximate large sample confidence intervals.

REVISITING THE CENTRAL LIMIT THEOREM

Recall the following result from STA 4321.

Result: If $Y_1, Y_2, Y_3, \ldots\ldots$ is a sequence of independent random variables with a common distribution which has mean μ and variance σ^2, then the distribution function of $\dfrac{\sqrt{n}\,(\overline{Y}_n - \mu)}{\sigma}$ $\left(\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i\right)$ converges to the distribution function of a Normal $(0, 1)$ random variable. In particular,

$$P\left(\frac{\sqrt{n}\,(\overline{Y}_n - \mu)}{\sigma} \le x\right) \longrightarrow P(Z \le x) \text{ as } n \to \infty.$$

(Z is Normal $(0,1)$ random variable)

This result is immensely useful in situations where it is hard to construct pivotal quantities in order to get confidence intervals.

APPROXIMATE LARGE SAMPLE CONFIDENCE INTERVALS

Consider the situation where we have a random sample $Y_1, Y_2, ---, Y_n$ from a population with mean μ and variance σ^2. Suppose we have no further information about the distribution of $Y_1, Y_2, ---, Y_n$.

<u>Task:</u> Obtain a $(1-\alpha)$-confidence interval for μ.

Since we have no idea about the distribution of $Y_1, Y_2, ---, Y_n$, it is not feasible to construct a pivotal quantity, and we cannot obtain a confidence interval using such a method.

However, IF n is <u>LARGE ENOUGH</u>, then by the central limit theorem, the distribution of $\dfrac{\sqrt{n}\,(\bar{Y}_n - \mu)}{\sigma}$ is <u>APPROXIMATELY</u> Normal$(0,1)$. Hence, if σ is known, $\dfrac{\sqrt{n}\,(\bar{Y} - \mu)}{\sigma}$ qualifies to be an <u>APPROXIMATE PIVOTAL QUANTITY.</u>

<u>Step 2:</u> Find a and b such that

$$P\left(a \le \frac{\sqrt{n}\,(\bar{Y}_n - \mu)}{\sigma} \le b\right) = 1-\alpha.$$

Note that $P\left(a \le \dfrac{\sqrt{n}\,(\bar{Y}_n - \mu)}{\sigma} \le b\right) \underset{\text{approximately}}{\approx} P\big(a \le \text{Normal}(0,1) \le b\big)$

RESULT: Let $z_{1-\alpha}$ denote the $(1-\alpha)^{th}$ quantile of the Normal(0,1) distribution. Then,

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \text{Normal}(0,1) \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

It follows that

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) \underset{\text{approximately}}{\approx} 1-\alpha.$$

Step 3: Express $\left\{-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right\}$ as $\left\{\hat{\mu}_L \leq \mu \leq \hat{\mu}_U\right\}$.

Note that

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{Y}_n-\mu)}{\sigma} \leq z_{1-\frac{\alpha}{2}}$$

$$\Leftrightarrow \frac{-z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq \bar{Y}_n-\mu \leq \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \bar{Y}_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y}_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}$$

Hence, an APPROXIMATE LARGE SAMPLE interval for μ is given by $\left[\bar{Y}_n - \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \bar{Y}_n + \frac{z_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right]$.

The phrase "LARGE SAMPLE" means that the approximation used to obtain the interval is valid only if n is large enough. How large enough? That depends on the specific scenario, but typically we would want n atleast to be around 50 (or higher). The punchline is "higher n, better approximation".

WHAT ARE THE OPTIONS IF σ is UNKNOWN?

The most obvious solution is: Since we are anyway using an approximation, why not replace σ by an estimate, such as S. Note that S^2 is ~~a~~ the standard estimate for σ^2, hence S would be a viable estimate for σ. Hence, the proposed APPROXIMATE LARGE SAMPLE confidence interval for μ, when σ is unknown is given by

$$\left[\bar{Y}_n - \frac{z_{1-\frac{\alpha}{2}} S}{\sqrt{n}}, \bar{Y}_n + \frac{z_{1-\frac{\alpha}{2}} S}{\sqrt{n}}\right].$$

<u>FACT:</u> A nice result which further justifies using the above confidence interval says that if $Y_1, Y_2, ---, Y_n$ is a random sample from a population with mean μ and variance σ^2, then the distribution function of $\frac{\sqrt{n}(\bar{Y}_n - \mu)}{S}$

converge to the distribution function of a Normal$(0,1)$ random variable.