Lecture 21.

Remark: 7.4 in the textbook will not be covered. However, I encourage you to read the material, as it is useful.

Chapter 8 – Statistical intervals for a Single Sample.

Confidence Intervals

- Suppose we have an estimator $\hat{\Theta}$ for some parameter θ.
- Given a sample $\{x_1, \dots, x_n\}$, $\hat{\Theta}(x_1, \dots, x_n) = \hat{\theta}$ is a point estimate of θ.
- We want to know how close $\hat{\theta}$ is to θ.
  - ↳ at the very least we could sample repeatedly and try to an interval where $\hat{\Theta}$ is likely to take values.

Defn: A 100(1-α)% Confidence interval (CI) for θ is an interval $[L(X_1, \dots, X_n), U(X_1, \dots, X_n)]$ such that

$$P(L(X_1, \dots, X_n) \le \theta \le U(X_1, \dots, X_n)) = 1-\alpha.$$

$\boxed{\begin{array}{l} L \text{ for "lower"} \\ U \text{ for "upper"} \end{array}}$

The term $1-\alpha$ is called the "confidence coefficient"

- usually we want 90% ($\alpha = 0.1$), 95% ($\alpha = 0.05$), 99% ($\alpha = 0.01$).

- Note that the upper and lower confidence limits, $U(X_1, \dots, X_n)$ and $L(X_1, \dots, X_n)$ are statistics of a random sample and therefore finding a CI depends heavily on the distribution of the population and known information.

Confidence Interval on the mean of a normal distribution with known variance

- In this section, we assume that we are sampling a normally distributed population with known variance σ^2.

- Though it is somewhat unrealistic to assume that the variance is known, it will illustrate the basic ideas.

- Let $\overline{X}$ be our sample mean; we want to estimate μ.

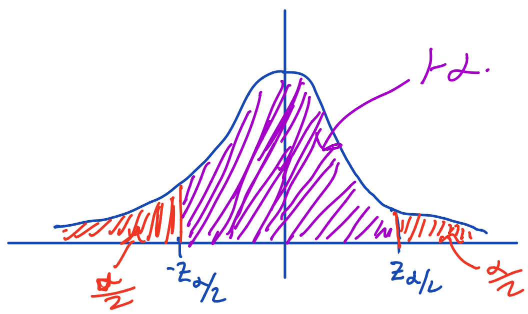
- Since we are sampling a normal, $\overline{X} = N(\mu, \sigma^2/n)$.

- we want to find a confidence interval (L, U) which is symmetric about $\overline{X}$.

- Standardizing $\overline{X}$ to $Z = \dfrac{\overline{X}-\mu}{(\sigma/\sqrt{n})}$ we find a value $z_{\alpha/2}$ (depends on α) such that

$$P\left(-z_{\alpha/2} \le \overbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}^{Z} \le z_{\alpha/2}\right) = 1-\alpha$$

* Find $z_{\alpha/2}$ in the Z-score tables!



- manipulating this expression:

$$P\Big(\underbrace{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L(\overline{X})} \le \mu \le \underbrace{\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{U(\overline{X})}\Big) = 1-\alpha.$$

- The interval $[L(\overline{X}), U(\overline{X})] = [\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}]$ is a random interval, since the endpoints involve the random variable $\overline{X}$.

- for a point estimate $\overline{X} = \overline{x}$, the $100(1-\alpha)\%$ confidence interval for μ is then
$$[\overline{x} - z_{\alpha/2}\sigma/\sqrt{n},\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}].$$

- The value $|z_{\alpha/2}\, \sigma/\sqrt{n}|$ is called the margin of error (E)
- The length of the interval, $2|z_{\alpha/2}\, \sigma/\sqrt{n}|$ is the precision $(2E)$

Ex: Find the 95% confidence interval for the mean of a normal distribution with $\sigma = 60$ and sample size 36.

Soln: $100(1-\alpha) = 95 \longrightarrow 1-\alpha = 0.95 \rightarrow \alpha = 0.05$

Now, $z_{\alpha/2}$ is by definition the value such that

$$\underbrace{P(Z \le z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975}$$

$$\text{Z-score table} \rightarrow z_{\alpha/2} = 1.96$$

Therefore, the margin of error (E) is

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 1.96\left(\frac{60}{\sqrt{36}}\right) = 1.96\left(\frac{60}{6}\right) = 19.6$$

So, given a sample mean $\bar{x}$, the $100(1-\alpha)\%$ CI for that estimate is $\bar{x} \pm 19.6$.

Choosing Sample Size:

- when estimating parameters we would like to be as precise as possible
- this means that we want to make the margin of error (E) as small as possible

Fact: Suppose that we are sampling a normal distribution with known variance σ^2. If we are approximating μ with $\bar{X}$, then our margin of error will be at most E so long as we take our sample to be of size

$$n = \left\lceil \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 \right\rceil$$

$\lceil\ \rceil$ = ceiling function, i.e. next largest integer.

<u>Ex</u>: suppose we are sampling a normal distribution with $\sigma = 60$. What sample size do we need to have a 95% confidence interval of the mean with precision 20?

<u>Soln</u> From the last example, $z_{\alpha/2} = 1.96$. Therefore, we need a sample size of at least

$$n = \left\lceil \left(\frac{(1.96)(60)}{\underbrace{10}_{E = \frac{\text{precision}}{2}}} \right)^2 \right\rceil = \left\lceil \left((1.96)\,6 \right)^2 \right\rceil = \lceil 135.49 \rceil = 136.$$

- So a sample size of 36 gave us an error of 18.6, and we need a sample size of 136 ~~for~~ $E = 10$.