Statistic

- A statistic is a function of the random sample which does not depend on any unknown parameters.
- For example

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

or

$$X_{(n)} = \max(X_1, X_2, \cdots, X_n)$$

are some examples of a statistic.

✓ A statistic/estimator must be a function of a certain sample, $X_1, X_2, \ldots, X_n$ say.

✓ It does not depend on any unknown parameter. For example

★ $\sum_{i=1}^{n}(X_i - \mu)^2$ is NOT a statistic if μ is unknown; but it is a statistic if μ is assumed to be a known value.

★ $(\bar{X} - 1)/5$ is a statistic; but $(\bar{X} - \mu)/\sigma$ is not if either μ or σ is unknown.

★ $\min\{X_1, X_2, \ldots, X_n\}$ is a statistic. $\min\{X_1, X_n\}$ is a statistic. For each i, X_i is also a statistic.

✓ A statistic/estimator can be viewed either as a random variable or as a computational rule, but not a computed value. They should be distinguished from their realized value based on the observed sample. For example, $\bar{X}$ is an estimator, but $\bar{x}$ is not. Question: whether a specific value, e.g., 1, can be viewed as a statistic/estimator?

Page 6-8

Point Estimate of Mean

- Suppose μ is the population mean.
- The statistic that one uses to obtain a point estimate is called **an estimator**,

  For example, $\bar{X}$ is an estimator of μ.

  The value of $\bar{X}$, denoted by $\bar{x}$, is an estimate of μ.

We should clearly distinguish these three concepts: an estimator/statistic (e.g., $\bar{X}$), an estimate (e.g., $\bar{x}$), a (population) parameter (e.g., μ).

✓ An estimator/statistic is a computational rule. It is also a random variable. When the data (random sample) are available, it tells how to compute. For example, $\bar{X}$ is a random variable, and when we have $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ available, it tells that we should compute the mean value of them.

✓ An estimate is a computed value of the estimator based on the observed data (random sample). It is NOT a random variable.

✓ A population parameter is something about the population, but unknown.

✓ Think about which of the following probability statements is/are valid.

★ $Pr(\bar{X} \leq 1) = 0.5.$

★ $Pr(0 < \bar{x} < 2) \leq 0.8.$

★ $Pr(\bar{x} - 4 < \mu < \bar{x} + 4) = 0.95.$

★ $Pr(\bar{X} - \bar{x} < \mu - 2) = 0.90.$

★ $Pr(\bar{X} = \bar{x}) > 0.2.$

Interval Estimation

- Interval estimation is to define two statistics, say,

$$\widehat{\Theta}_L \quad \text{and} \quad \widehat{\Theta}_U, \qquad \text{where } \widehat{\Theta}_L < \widehat{\Theta}_U$$

so that $(\widehat{\Theta}_L, \ \widehat{\Theta}_U)$ constitutes a random interval for which the probability of containing the unknown parameter θ can be determined.

✓ Both interval estimation and point estimation are used to estimate a specific parameter, e.g., μ, of a population.

✓ Point estimator uses one single statistic to estimate; but interval estimator uses two statistics, which form an interval, to estimate.

✓

✓ In analogous to the point estimator, the interval estimator a RANDOM interval, so that we can assert that $Pr(\widehat{\theta}_L < \theta < \widehat{\theta}_U) = 1 - \alpha$.

✓ Similarly to the point estimation, we should distinguish the terminologies "interval estimator" and "interval estimate".

6.1.3 Unbiased Estimator

Definition 6.1 (Unbiased estimator)

- A statistic $\widehat{\Theta}$ is said to be an **unbiased estimator** of the parameter θ if

$$E\left(\widehat{\Theta}\right) = \theta.$$

We look at two criteria when comparing different estimators for the same parameter.

✓ First, we shall consider only unbiased estimators. The definition of unbiased estimator is given on this page of the lecture slide. Note that $E(\widehat{\theta}) = \theta$ needs to be true for any arbitrary θ such that the probability density function $f_X(x;\theta)$ is correctly defined.

✓ Second, among the unbiased estimators, we shall recommend the one that leads to the smallest variance.

For example, consider the sample $X_1, X_2, \ldots, X_n$ with $n > 2$. The following are candidate estimators for estimating μ: $\widehat{\mu}_1 = \bar{X}, \widehat{\mu}_2 = X_1, \widehat{\mu}_3 = \frac{X_1+X_2}{2}, \widehat{\mu}_4 = 2, \widehat{\mu}_5 = X_n - X_1$.

✓ Based on the unbiasness criterion, $\widehat{\mu}_1$, $\widehat{\mu}_2$, and $\widehat{\mu}_3$ are all unbiased estimators; but $E(\widehat{\mu}_4) = 2 \neq \mu$ and $E(X_n - X_1) = \mu - \mu = 0 \neq \mu$ indicate that they are biased. As a consequence, we should drop $\widehat{\mu}_4$ and $\widehat{\mu}_5$.

✓ Based on the second criteria, we need to compare the variances of $\widehat{\mu}_1, \widehat{\mu}_2$, and $\widehat{\mu}_3$. $V(\widehat{\mu}_1) = \sigma^2/n, V(\widehat{\mu}_2) = \sigma^2$, and $V(\widehat{\mu}_3) = \sigma^2/2$. So when $n > 2$, $\widehat{\mu}_1$ is unbiased and has the minimum variance.

Unbiased Estimator (Continued)

Example 1

$\bar{X}$ is an unbiased estimator of μ. That is, $E(\bar{X}) = \mu$.

Example 2

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an **unbiased** estimator of σ^2.

That is,

$$E(S^2) = \sigma^2$$

ST2334 Probability and Statistics CYM Estimation based on Normal Distribution 6-13

Here is a derivation for $E(S^2) = \sigma^2$.

✓ Recall the formula:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2,$$

which is an algebraic formula, and is valid no matter whether X_i's are random variables and no matter what are the values for X_i.

✓ Set $Y_i = X_i - \mu$. Then $Y_1, \ldots, Y_n$ are i.i.d. with $E(Y_i) = 0$ and $V(Y_i) = V(X_i) = \sigma^2$. Furthermore $\bar{Y} = \bar{X} - \mu$, thus $E(\bar{Y}) = 0$ and $V(\bar{Y}) = V(\bar{X}) = \sigma^2/n$. We have

$$\begin{aligned} E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) &= E\left(\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right) = E\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right) \\ &= \sum_{i=1}^{n} EY_i^2 - nE(\bar{Y}^2) = \sum_{i=1}^{n} V(Y_i) - nV(\bar{Y}) = n\sigma^2 - n\sigma^2/n = (n-1)\sigma^2. \end{aligned}$$

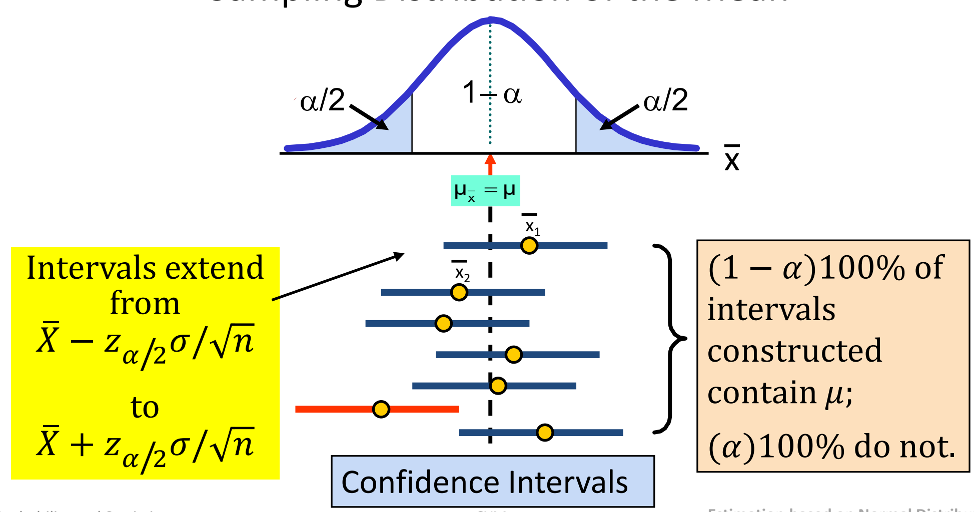
Page 6-18 & 21

Interval Estimation (Continued)

- We shall seek a random interval
$$(\widehat{\Theta}_L, \widehat{\Theta}_U)$$
containing θ with a given probability $1 - \alpha$.
- That is
$$\Pr\left(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U\right) = 1 - \alpha.$$

Intervals and Level of Confidence

Sampling Distribution of the Mean



There are two valid ways to interpret the meaning of level $1-\alpha$ confidence interval.

✓ The first is given on page 6-18 of the lecture slide; it is also how the $1-\alpha$ confidence interval is defined. Importantly, the upper and lower bounds $\widehat{\theta}_L$ and $\widehat{\theta}_U$ are random variable! For example we can claim $Pr(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$, where $\bar{X}$ is the random variable which supplies the randomness such that we can talk about "probability". But if you have the observed data $X_1 = x_1, \ldots, X_n = x_n$, then you apply the formula to get the computed value of the confidence interval $(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n})$, it is no longer valid to talk about probability. In other words, $Pr(\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}) = 1 - \alpha$ makes nonsense, as there is no random variable in the $Pr(\cdot)$ statement. More specifically, if your computed confidence interval is $(3, 6)$, it is invalid to report "the probability that μ is contained in $(3, 6)$ is 95%."

✓ The second is given on page 6-21. Imaging that we can sample the data infinitely many times from the population

★ Get the first sample $\left(X_1^{(1)}, X_2^{(1)}, \ldots, X_n^{(1)}\right)$ from the distribution $f_X(x; \theta)$, and compute the CI $(\widehat{\theta}_{L,1}, \widehat{\theta}_{U,1})$.

★ Get the second sample $\left(X_1^{(2)}, X_2^{(2)}, \ldots, X_n^{(2)}\right)$ from the distribution $f_X(x; \theta)$, and compute the CI $(\widehat{\theta}_{L,2}, \widehat{\theta}_{U,2})$.

★ Continue with this procedure $\ldots, \ldots, \ldots$

★ Get the Kth sample $\left(X_1^{(K)}, X_2^{(K)}, \ldots, X_n^{(K)}\right)$ from the distribution $f_X(x; \theta)$, and compute the CI $(\widehat{\theta}_{L,K}, \widehat{\theta}_{U,K})$.

★ For a sufficiently large K, the proportion of these intervals that contain the true value of θ is $1 - \alpha$.

✓ Practically, for a computed CI, we can only claim that with a certain confidence the interval will cover the true value. For example, if the computed 95% CI for μ is $(3, 6)$, we can only claim that we have 95% "confidence" that the true value of μ will be contained in $(3, 6)$.

Here is a general strategy (formula) for constructing mean related confidence intervals. Suppose we are to construct a $1-\alpha$ confidence interval for mean related parameter θ (e.g., θ could be μ, $\mu_1-\mu_2$, or other possible combinations of the population means)

✓ Step 1: look for an estimator $\widehat{\theta}$ for θ, e.g., $\bar{X}$ for θ, $\bar{X}_1-\bar{X}_2$ for $\mu_1-\mu_2$.

✓ Step 2: Derive the variance $V(\widehat{\theta})$.

✓ Step 3: Construct $(1-\alpha)$ CI to be $\widehat{\theta} \pm M\sqrt{V}$. M is called the multiplier, and V is related to $V(\widehat{\theta})$. The following is how they are determined.

★ If $V(\widehat{\theta})$ does not depend on any other parameter (e.g., in the case σ^2 is known, $V(\bar{X}) = \sigma^2/n$), $V = V(\widehat{\theta})$, and $M = z_{\alpha/2}$. Here we may need the condition that the data are normal or/and the sample size n is big.

The CIs given on pages 6-25 and 6-46 belong to this situation.

★ If the derived $V(\widehat{\theta})$ contains some other unknown parameter, e.g., σ^2, we replace the parameter with its estimator, e.g., we use S^2 to replace σ^2; this result in $\widehat{V}(\widehat{\theta})$. Then, we use $V = \widehat{V}(\widehat{\theta})$, however M has two possibilities (it has more in the literature):

(1) if the sample size n is sufficiently large, $M = z_{\alpha/2}$; the CIs given on pages 6-36, 6-52, and 6-73 belong to this situation;

(2) if the sample size n is not large, but the data are normally distributed, $M = t(df, \alpha/2)$. Here df = degrees of freedom, which is the df of the estimator for the parameter contained in $V(\widehat{\theta})$. The CIs given on pages 6-35, 6-62, and 6-72 belong to this situation.

Note: this strategy does not apply to construct variance related CIs in general. See Pages from 6-78 to 6-100 for the development of the variance related CIs.

Page 6-58

Unknown but Equal Variances (Continued)

- σ^2 can be estimated by the pooled sample variance
$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},$$
with S_1^2 and S_2^2 being the sample variances of the first and second samples respectively.

Let $X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$ be observations from population 1 and let $X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$ be observations from population 2. Then $(n_1-1)S_1^2 = \sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1)^2$ and $(n_2-1)S_2^2 = \sum_{j=1}^{n_2}(X_{2,j}-\bar{X}_2)^2$. We therefore have

$$S_p^2 = \frac{\sum_{i=1}^{n_1}(X_{1,i}-\bar{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2,i}-\bar{X}_2)^2}{n_1+n_2-2}.$$

Note that this formula intuitively makes sense: every observation contributes equally in the estimation of their common variance σ^2.

In terms of samples, it is the weighted average of the two sample variances with the weights being one less than the sample sizes.

Unknown but Equal Variances (Continued)

- Note that if the two populations are normal with the same variance σ^2, then

$$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2_{n_1-1} \quad \text{and} \quad \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_2-1},$$

Hence

$$\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

The derivation on page used the property of chi-square distribution: if $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, and Y_1 and Y_2 are independent, then $Y_1 + Y_2 \sim \chi^2(n_1+n_2)$.