Module 4: Point Estimation
Statistics (OA3102)

Professor Ron Fricker
Naval Postgraduate School
Monterey, California

Reading assignment:
WM&S chapter 8.1-8.4
Goals for this Module

• Define and distinguish between point estimates vs. point estimators

• Discuss characteristics of good point estimates
  – Unbiasedness and minimum variance
  – Mean square error
  – Consistency, efficiency, robustness

• Quantify and calculate the precision of an estimator via the standard error
  – Discuss the Bootstrap as a way to empirically estimate standard errors
Welcome to Statistical Inference!

• Problem:
  – We have a simple random sample of data
  – We want to use it to estimate a population quantity (usually a parameter of a distribution)
  – In point estimation, the estimate is a number

• Issue: Often lots of possible estimates
  – E.g., estimate $E(X)$ with $\bar{x}$, $\tilde{x}$, or $???$

• This module: What’s a “good” point estimate?
  – Module 5: Interval estimators
  – Module 6: Methods for finding good estimators
Point Estimation

- A point estimate of a parameter $\theta$ is a single number that is a sensible value for $\theta$
  - I.e., it’s a numerical estimate of $\theta$
  - We’ll use $\theta$ to represent a generic parameter – it could be $\mu$, $\sigma$, $p$, etc.
- The point estimate is a statistic calculated from a sample of data
  - The statistic is called a point estimator
  - Using “hat” notation, we will denote it as $\hat{\theta}$
  - For example, we might use $\bar{x}$ to estimate $\mu$, so in this case $\hat{\mu} = \bar{x}$
Definition: *Estimator*

An estimator is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.
An Example

• You’re testing a new missile and want to estimate the probability of kill (against a particular target under specific conditions)
  – You do a test with $n=25$ shots
  – The parameter to be estimated is $p_k$, the fraction of kills out of the 25 shots

• Let $X$ be the number of kills
  – In your test you observed $x=15$

• A reasonable estimator and estimate is

\[ \hat{p}_k = \frac{X}{n} \quad \text{estimate:} \quad \hat{p}_k = \frac{x}{n} = \frac{15}{25} = 0.6 \]
A More Difficult Example

• On another test, you’re estimating the mean time to failure (MTTF) of a piece of electronic equipment
• Measurements for \( n=20 \) tests (in units of 1,000 hrs):
  
  24.46  25.61  26.25  26.42  26.66  27.15  27.31  27.54  27.27  27.94  
  27.98  28.04  28.28  28.49  28.50  28.87  29.11  29.13  29.50  30.88  

• Turns out a normal distribution fits the data quite well
• So, what we want to do is to estimate \( \mu \), the MTTF
• How best to do this?
Example, cont’d

- Here are some possible estimators for $\mu$ and their values for this data set:

(1) $\hat{\mu} = \bar{X}$, so $\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 27.793$

(2) $\hat{\mu} = \tilde{X}$, so $\tilde{x} = \frac{27.94 + 27.98}{2} = 27.960$

(3) $\hat{\mu} = \bar{X}_e = \frac{\min(X_i) + \max(X_i)}{2}, \text{ so } \bar{x}_e = \frac{24.46 + 30.88}{2} = 27.670$

(4) $\hat{\mu} = \bar{X}_{tr(10)}$, so $\bar{x}_{tr(10)} = \frac{1}{16} \sum_{i=3}^{18} x_{(i)} = 27.838$

- Which estimator should you use?
  – I.e., which is likely to give estimates closer to the true (but unknown) population value?
Another Example

- In a wargame computer simulation, you want to estimate a scenario’s run-time variability ($\sigma^2$)
- The run times (in seconds) for eight runs are:
  
  44.2  43.9  44.7  44.2  44.0  43.8  44.6  43.1

- Two possible estimates:

  (1) $\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$, so $s^2 = 0.25125$

  (2) $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$, so estimate = 0.220

- Why prefer (1) over (2)?
Bias

• Definition: Let $\hat{\theta}$ be a point estimator for a $\theta$
  – $\hat{\theta}$ is an unbiased estimator if $E(\hat{\theta}) = \theta$
  – If $E(\hat{\theta}) \neq \theta$, $\hat{\theta}$ is said to be biased

• Definition: The bias of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$

• E.g.:
• Proposition: Let $X$ be a binomial r.v. with parameters $n$ and $p$. The sample proportion $\hat{p} = X/n$ is an unbiased estimator of $p$.

• Proof:
Remember: Rules for Linear Combinations of Random Variables

- For random variables $X_1, X_2, \ldots, X_n$
  - Whether or not the $X_i$'s are independent
    \[
    E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n)
    \]
  - If the $X_1, X_2, \ldots, X_n$ are independent
    \[
    \text{Var}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + \cdots + a_n^2 \text{Var}(X_n)
    \]
Example 8.1

• Let \( Y_1, Y_2, \ldots, Y_n \) be a random sample with \( E(Y_i)=\mu \) and \( \text{Var}(Y_i)=\sigma^2 \). Show that

\[
S'^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2
\]

is a biased estimator for \( \sigma^2 \), while \( S^2 \) is an unbiased estimator of \( \sigma^2 \).

• Solution:
Example 8.1 (continued)
Another Biased Estimator

- Let $X$ be the reaction time to a stimulus with $X \sim \text{U}[0, \theta]$, where we want to estimate $\theta$ based on a random sample $X_1, X_2, \ldots, X_n$.
- Since $\theta$ is the largest possible reaction time, consider the estimator $\hat{\theta}_1 = \max(X_1, X_2, \ldots, X_n)$.
- However, unbiasedness implies that we can observe values bigger and smaller than $\theta$.
  - Why?
  - Thus, $\hat{\theta}_1$ must be a biased estimator.
Fixing the Biased Estimator

• For the same problem consider the estimator

\[ \hat{\theta}_2 = \left( \frac{n + 1}{n} \right) \max \left( X_1, X_2, \ldots, X_n \right) \]

• Show this estimator is unbiased
One Criterion for Choosing Among Estimators

- **Principle of minimum variance unbiased estimation**: Among all estimators of $\theta$ that are unbiased, choose the one that has the minimum variance
  - The resulting estimator $\hat{\theta}$ is called the minimum variance unbiased estimator (MVUE) of $\theta$

Estimator $\hat{\theta}_1$ is preferred to $\hat{\theta}_2$

* Figure from *Probability and Statistics for Engineering and the Sciences*, 7th ed., Duxbury Press, 2008.
Example of an MVUE

• Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with parameters $\mu$ and $\sigma$. Then the estimator $\hat{\mu} = \bar{X}$ is the MVUE for $\mu$
  – Proof beyond the scope of the class

• Note this only applies to the normal distribution
  – When estimating the population mean $E(X) = \mu$ for other distributions, $\bar{X}$ may not be the appropriate estimator
  – E.g., for Cauchy distribution $E(X) = \infty$!
How Variable is My Point Estimate?  
The Standard Error

• The precision of a point estimate is given by its standard error

• The standard error of an estimator $\hat{\theta}$ is its standard deviation

$$\sigma_{\hat{\theta}} = \sqrt{\text{Var}(\hat{\theta})}$$

• If the standard error itself involves unknown parameters whose values are estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the estimated standard error
  
  – The estimated standard error is denoted by $\hat{\sigma}_{\hat{\theta}}$ or $s_{\hat{\theta}}$
• Proposition: If $Y_1, Y_2, \ldots, Y_n$ are distributed iid with variance $\sigma^2$ then, for a sample of size $n$, $\text{Var}(\bar{Y}) = \sigma^2/n$. Thus $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$.

• Proof:
• Proposition: If $Y_i \sim \text{Bin}(n,p)$, $i=1,...,n$, then

$$\sigma_{\hat{p}} = \sqrt{pq/n}$$

where $q=1-p$ and $\hat{p} = Y/n$.

• Proof:
# Expected Values and Standard Errors of Some Common Point Estimators

<table>
<thead>
<tr>
<th>Target Parameter $\theta$</th>
<th>Sample Size(s)</th>
<th>Point Estimator $\hat{\theta}$</th>
<th>$E(\hat{\theta})$</th>
<th>Standard Error $\sigma_{\hat{\theta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$n$</td>
<td>$\bar{Y}$</td>
<td>$\mu$</td>
<td>$\sigma \sqrt{\frac{1}{n}}$</td>
</tr>
<tr>
<td>$p$</td>
<td>$n$</td>
<td>$\hat{p} = \frac{Y}{n}$</td>
<td>$p$</td>
<td>$\sqrt{\frac{pq}{n}}$</td>
</tr>
<tr>
<td>$\mu_1 - \mu_2$</td>
<td>$n_1$ and $n_2$</td>
<td>$\bar{Y}_1 - \bar{Y}_2$</td>
<td>$\mu_1 - \mu_2$</td>
<td>$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$</td>
</tr>
<tr>
<td>$p_1 - p_2$</td>
<td>$n_1$ and $n_2$</td>
<td>$\hat{p}_1 - \hat{p}_2$</td>
<td>$p_1 - p_2$</td>
<td>$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$</td>
</tr>
</tbody>
</table>

If populations are independent.
However, Unbiased Estimators Aren’t Always to be Preferred

• Sometimes an estimator with a small bias can be preferred to an unbiased estimator

• Example:

• More detailed discussion beyond scope of course – just know unbiasedness isn’t necessarily required for a good estimator
Mean Square Error

- **Definition:** The mean square error (MSE) of a point estimator $\hat{\theta}$ is
  \[ \text{MSE}(\hat{\theta}) = E\left[ (\hat{\theta} - \theta)^2 \right] \]

- MSE of an estimator $\hat{\theta}$ is a function of both its variance and its bias
  - I.e., it can be shown (extra credit problem) that
    \[ \text{MSE}(\hat{\theta}) = E\left[ (\hat{\theta} - \theta)^2 \right] = \text{Var}(\hat{\theta}) - \left[ B(\hat{\theta}) \right]^2 \]
  - So, for unbiased estimators $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$
• **Definition:** The **error of estimation** $\varepsilon$ is the distance between an estimator and its target parameter: $\varepsilon = |\hat{\theta} - \theta|$
  
  – Since $\hat{\theta}$ is a random variable, so it the error of estimation, $\varepsilon$
  
  – But we can bound the error:

\[
\Pr\left( |\hat{\theta} - \theta| < b \right) = \Pr\left( -b < \hat{\theta} - \theta < b \right) = \Pr\left( \theta - b < \hat{\theta} < \theta + b \right)
\]
• **Tchebyssheff’s Theorem.** Let $Y$ be a random variable with finite mean $\mu$ and variance $\sigma^2$. Then for any $k > 0$, $\Pr\left(|Y - \mu| < k\sigma\right) \geq 1 - 1/k^2$

  – Note that this holds for any distribution
  – It is a (generally conservative) bound
  – E.g., for any distribution we’re guaranteed that the probability $Y$ is within 2 standard deviations of the mean is at least 0.75

• So, for unbiased estimators, a good bound to use on the error of estimation is $b = 2\sigma_{\hat{\theta}}$
Example 8.2

• In a sample of $n=1,000$ randomly selected voters, $y=560$ are in favor of candidate Jones.

• Estimate $p$, the fraction of voters in the population favoring Jones, and put a 2-s.e. bound on the error of estimation.

• Solution:
Example 8.3

- Car tire durability was measured on samples of two types of tires, \( n_1 = n_2 = 100 \). The number of miles until wear-out were recorded with the following results:

\[
\bar{y}_1 = 26,400 \text{ miles} \quad \bar{y}_2 = 25,100 \text{ miles}
\]

\[
s_1^2 = 1,440,000 \text{ miles}^2 \quad s_2^2 = 1,960,000 \text{ miles}^2
\]

- Estimate the difference in mean miles to wear-out and put a 2-s.e. bound on the error of estimation
Example 8.3

- Solution:
Other Properties of Good Estimators

• An estimator is **efficient** if it has a small standard deviation compared to other unbiased estimators.

• An estimator is **robust** if it is not sensitive to outliers, distributional assumptions, etc.
  – That is, robust estimators work reasonably well under a wide variety of conditions.

• An estimator $\hat{\theta}_n$ is **consistent** if

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty$$

➢ For more detail, see Chapter 9.1-9.5
A Useful Aside: Using the Bootstrap to Empirically Estimate Standard Errors

Draw multiple \((R)\) samples from the population, where
\[ x_i = \{x_{1i}, x_{2i}, \ldots, x_{ni}\} \]

Calculate multiple parameter estimates

Estimate s.e. of the parameter using the std. dev. of the estimates

\[
\text{s.e.}[\theta(X)] = \sqrt{\frac{1}{R-1} \sum_{i=1}^{R} \left[ \hat{\theta}(x_i) - \hat{\theta}(X) \right]^2}
\]
The Bootstrap

• The “hard way” is either not possible or is wasteful in practice

• **Bootstrap** is:
  – Useful when you don’t know or, worse, simply cannot analytically derive sampling distribution
  – Provides a computer-intensive method to empirically estimate sampling distribution

• Only feasible recently with the widespread availability of significant computing power
Plug-in Principle

- We’ve been doing this throughout the class
- If you need a parameter for a calculation, simply “plug in” the equivalent statistic
- For example, we defined
  \[ \text{Var}(X) = E\left[ (X - E(X))^2 \right] \]
  and then we sometimes did the calculation using \( \bar{X} \) for \( E(X) \)
- Relevant for the bootstrap as we will “plug in” the empirical distribution in place of the population distribution
Empirically Estimating Standard Errors Using the Bootstrap

\[ \mathbf{x} = (x_1, x_2, \ldots, x_n) \sim \hat{F} \]

Draw multiple \((B)\) resamples from the data, where
\[ \mathbf{x}_i^* = \{x_{1i}^*, x_{2i}^*, \ldots, x_{ni}^*\} \]

Calculate multiple bootstrap estimates

Estimate s.e. from bootstrap estimates

\[
S_\hat{\theta} = \sqrt{\frac{1}{B-1} \sum_{i=1}^{B} \left[ \hat{\theta}(\mathbf{x}_i^*) - \bar{\hat{\theta}}^* \right]^2}
\]

where
\[
\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}(\mathbf{x}_i^*)
\]

Revision: 1-12

38
Some Key Ideas

- Bootstrap samples are drawn with replacement from the empirical distribution
  - So, observations can actually occur in the bootstrap sample more frequently than they occurred in the actual sample

- Empirical distribution substitutes for the actual population distribution
  - Can draw lots of bootstrap samples from the empirical distribution to calculate the statistic of interest
    - Make $B$ as big as can run in a reasonable timeframe
    - Bootstrap resamples are of same size as original sample ($n$)

- Because this is all empirical, don’t need to analytically solve for the sampling distribution of the statistic of interest
• Defined and distinguished between point estimates vs. point estimators
• Discussed characteristics of good point estimates
  – Unbiasedness and minimum variance
  – Mean square error
  – Consistency, efficiency, robustness
• Quantified and calculated the precision of an estimator via the standard error
  – Discussed the Bootstrap as a way to empirically estimate standard errors
Homework

- **WM&S chapter 8.1-8.4**
  - Required exercises: 2, 8, 21, 23, 27
  - Extra credit: 1, 6

- **Useful hints:**
  - Problem 8.2: Don’t just give the obvious answer, but show why it’s true mathematically
  - Problem 8.8: Don’t do the calculations for the $\hat{\Theta}_4$ estimator
  - Extra credit problem 8.6: The $a$ term is a constant with $0 \leq a \leq 1$