Concepts of Point Estimate

- A **point estimate** of some population parameter $\theta$ is a numerical value calculated from the sample.

- A **point estimator** is a formula or rule that tells us how to calculate a numerical estimate, denoted by $\hat{\theta}(y_1, y_2, \ldots, y_n)$.

- The **bias** $B$ of an estimator $\hat{\theta}$ is equal to 
  \[ B = \mathbb{E}(\hat{\theta}) - \theta \]

- An estimator $\hat{\theta}$ is **unbiased** if $\mathbb{E}(\hat{\theta}) = \theta$, i.e., $B = 0$.

- The **mean squared error** of a point estimator is equal to 
  \[ \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2] = \text{Var}(\hat{\theta}) + B^2 \]

**Method of Moment**

- Let $y_1, y_2, \ldots, y_n$ represent a random sample of size $n$ from some distribution.

  kth population moment: \[ \mathbb{E}(y^k) \]

  kth sample moment:
  \[ m^k = \frac{y_1^k + y_2^k + \cdots + y_n^k}{n} \]

- Suppose the population distribution has parameters $\theta_1, \ldots, \theta_m$. Then the **moment estimators**, $\hat{\theta}_1, \ldots, \hat{\theta}_m$, are obtained by equating the first $m$ sample moments to the corresponding first $m$ population moments and solving the resulting equations for the unknown parameters.
Examples: Find the moment estimators.

- $y_1, y_2, \ldots, y_n \sim \text{Exp}(eta)$. ($\hat{\beta} = \bar{y}$)
- $y_1, y_2, \ldots, y_n \sim \text{Poisson}({\lambda})$. ($\hat{\lambda} = \bar{y}$)
- $y_1, y_2, \ldots, y_n \sim \text{Gamma}(\alpha, \beta)$. (see example 8.3 on page 346)
- $y_1, y_2, \ldots, y_n \sim \text{N}(\mu, \sigma^2)$.

$$\hat{\mu} = \bar{y} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}.$$  

Note: The moment estimator of $\sigma^2$ is not an unbiased estimator.

Method of Maximum Likelihood

- Suppose we randomly select a sample of $n$ observations, $y_1, \ldots, y_n$ from a distribution $p(y \mid \theta)$, where $\theta$ is an unknown parameter. Then the likelihood of the sample is

$$L(\theta) = p(y_1 \mid \theta) \cdot p(y_2 \mid \theta) \cdots p(y_n \mid \theta)$$

Note:

- The likelihood is the joint probability function $p(y_1, \ldots, y_n \mid \theta)$ when $y_i$’s are discrete r.v.
- The likelihood is the joint density function $f(y_1, \ldots, y_n \mid \theta)$ when $y_i$’s are continuous r.v.
- Once we have observed $y_i$’s, the likelihood is a function of only the unknown parameter $\theta$. 
• Ronald A. Fisher (1890-1962):

One should choose as an estimate of $\theta$ the value of $\theta$ that maximizes the likelihood $L(\theta)$.

• MLE $\hat{\theta} = \arg\max_\theta L(\theta)$

• MLE $\hat{\theta} = \arg\max_\theta \log L(\theta)$

$$L(\theta) = \prod_{i=1}^{n} p(y_i | \theta)$$

$$\log L(\theta) = \sum_{i=1}^{n} \log p(y_i | \theta)$$

How to Find MLE

• MLE $\hat{\theta} = \arg\max_\theta L(\theta)$.

• Solve

$$\frac{dL}{d\theta} = 0 \quad \text{or} \quad \frac{d\log L}{d\theta} = 0$$

and check that the resulting solution is a maximum.

• Examples:

(1) (example 8.4) Find the MLE of Exp($\beta$).

(2) (example 8.5) Find the MLE of Normal($\mu, \sigma^2$).
Complications in Using MLE

- Maximum occurs at a discontinuous point.

  Example: Suppose \( y_1, \ldots, y_n \sim \text{Unif}(0, \theta) \). Find the MLE of \( \theta \).

- Close-form solution does not exist

  Example: Suppose \( y_1, \ldots, y_n \sim \text{Gamma}(\alpha, \beta) \). Find the MLE.

Confidence Interval

- Aim: How to use the sample to calculate two numbers that define an interval that will enclose the unknown parameter with certain probability (confidence).

- The resulting random interval is called a confidence interval.

- The probability that the interval contains the unknown parameter is called its confidence coefficient.

\[
P(\text{LCL} \leq \theta \leq \text{UCL}) = 1 - \alpha
\]

\( \text{LCL}(y_1, \ldots, y_n) \): lower confidence limit

\( \text{UCL}(y_1, \ldots, y_n) \): upper confidence limit
Case 1: Normal with Known Variance

Suppose \( \tilde{\mu} \sim N(\mu, \sigma^2) \) with \( \sigma^2 \) known. Define
\[
z = \frac{\tilde{\mu} - \mu}{\sigma} \sim N(0, 1).
\]
Locate values \( z_{\alpha/2} \) and \( -z_{\alpha/2} \) that place a probability of \( \alpha/2 \) in each tail of \( N(0, 1) \). For example, \( z_{.025} = 1.96 \).

\[
1 - \alpha = P\left( z_{\alpha/2} \leq z \leq z_{\alpha/2}\right)
\]
\[
= P\left(-z_{\alpha/2} \leq \frac{\tilde{\mu} - \mu}{\sigma} \leq z_{\alpha/2}\right)
\]
\[
= P\left(-z_{\alpha/2} \sigma \leq \tilde{\mu} - \mu \leq z_{\alpha/2} \sigma\right)
\]
\[
= P\left(\tilde{\mu} - z_{\alpha/2} \sigma \leq \mu \leq \tilde{\mu} + z_{\alpha/2} \sigma\right)
\]

**Theorem 8.2** Let \( \tilde{\mu} \sim N(\mu, \sigma^2) \). Then a \((1 - \alpha)100\%\) confidence interval for \( \mu \) is
\[
\tilde{\mu} - z_{\alpha/2} \sigma \text{ to } \tilde{\mu} + z_{\alpha/2} \sigma
\]

**Example**: ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch(CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (\( J \)) on specimens of A238 steel cut at \( 60^\circ\)C are as follows: 64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2 and 64.3. Assume that impact energy is normally distributed with \( \sigma = 1.0 J \).

Find a 95% CI for \( \mu \), the mean impact energy.

\[
\bar{y} \sim N(\mu, \sigma^2/n)
\]
\[
n = 10, \quad \sigma = 1, \quad \alpha_{.025} = 1.96
\]
\[
\bar{y} - z_{.025} \sigma \bar{y} \leq \mu \leq \bar{y} + z_{.025} \sigma \bar{y}
\]
\[
64.46 \quad 1.96 \frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96 \frac{1}{\sqrt{10}}
\]
\[
63.84 \leq \mu \leq 65.08
\]

How many specimens must be tested to ensure that the 95% CI of \( \mu \) has a length of at most \( 1.0 J \)?

\[
n = \left[ \frac{(1.96)1.0}{0.5} \right]^2 = 15.37.
\]
Interpreting a CI

- Can we conclude: The true mean $\mu$ is within the interval (63.84, 65.08) with probability 0.95? - NO

- The statement 63.84 $\leq$ $\mu$ $\leq$ 65.08 is either correct (true with probability 1) or incorrect (false with probability 1).

- Remember that a CI is a random interval and the correct interpretation of a $100(1 - \alpha)$% CI should depend on the relative frequency view of probability.

- We conclude:
  
  If we were to repeatedly collect a sample of size $n$ and construct a 95% CI for each sample, then we expect 95% of the intervals to enclose the true parameter $\mu$. 

m=20; n=10;
y = normrnd(0, 0.5, m,n);
y_mean = mean(y,2);
e = norminv(0.95)*0.5/sqrt(n)*ones(m,1)
errorbar(1:m, y_mean, e);
h = line([1, m], [0, 0]);
set(h, 'color', [1 0 0]);

LCL = y_mean - e;
UCL = y_mean + e;
sum(LCL > 0) + sum(UCL < 0)

ans =
  2
• If $\bar{y}$ is the sample mean of a random sample of size $n$ from $N(\mu, \sigma^2)$, the $(1 - \alpha)100\%$ CI is

$$\bar{y} - z_{\alpha / 2} \sigma / \sqrt{n} \leq \mu \leq \bar{y} + z_{\alpha / 2} \sigma / \sqrt{n}$$

• The length of the CI is equal to $2z_{\alpha / 2} \sigma / \sqrt{n}$.

• What’s the relationship between the length of a CI and
  • the confidence coefficient $(1 - \alpha)100\%$?
  • the sample size $n$?

Q: How many sample size we should choose in order to have the length of CI less than $l_0$.

$$2z_{\alpha / 2} \sigma / \sqrt{n} \leq l_0$$

$$n \geq \left( \frac{z_{\alpha / 2} \sigma}{l_0 / 2} \right)^2$$

---

**Sampling Dist Related to Normal**

A random sample $y_1, y_2, \ldots, y_n$ is drawn from $N(\mu, \sigma^2)$.

- sample mean $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

- sample var $s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n - 1}$

- Recall that $\chi^2(\nu) = z_1^2 + z_2^2 + \cdots + z_\nu^2$, where each $z_i \sim N(0,1)$.

- $(n - 1)s^2 / \sigma^2 \sim \chi^2(n - 1)$

- Let $z$ be a standard normal and $\chi^2$ be a chi-square with $\nu$ degrees of freedom, If $z$ and $\chi^2$ are independent, then

$$t = \frac{z}{\sqrt{\chi^2 / \nu}}$$

has a **Student’s t distribution** with $\nu$ degree of freedom.
Case 2: Normal with Unknown Variance
(Example 8.6) Let \( \bar{y} \) and \( s^2 \) be the sample mean and variance based on a random sample of \( n \) normal(\( \mu, \sigma^2 \)) observations

Define

\[
t = \frac{\bar{y} - \mu}{s / \sqrt{n}} = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \sqrt{\frac{(n-1)s^2}{\sigma^2(n-1)}}
\]

~ Student’s t distribution

\[
1 - \alpha = P(-t_{\alpha/2,n-1} \leq t \leq t_{\alpha/2,n-1})
\]

\[
= P(-t_{\alpha/2,n-1} \leq \frac{\bar{y} - \mu}{s / \sqrt{n}} \leq t_{\alpha/2,n-1})
\]

\[
= P(\bar{y} - t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{\alpha/2,n} \frac{s}{\sqrt{n}})
\]

Take a look of Example 8.7