9.0 Lesson Plan

- Answer Questions
- Distribution of Linear Combinations
- Point Estimation
- Unbiased Estimators
- Minimum Variance Unbiased Estimators
- The Variance of a Uniform RV
A linear combination of random variables $X_1, \ldots, X_n$ is a new random variable $Y$ such that

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

where the $a_i$ are known constants.

Some important linear combinations include:

- The sample mean, $\bar{X}$, in which each $a_i$ equals $1/n$.

- A difference, $X_1 - X_2$, in which $a_1 = 1$ and $a_2 = -1$. This is helpful when deciding whether, say, one brand of light bulb outlasts another brand, or whether one company outperforms another.
Let $X_i$ have mean $\mu_i$ and variance $\sigma_i^2$. Then

$$\mathbb{E}[Y] = \mathbb{E}[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] = \sum_{i=1}^{n} a_i \mu_i.$$ 

This holds even when the $X_i$ are dependent. It follows because integration is a linear operator: $\int a_i x f_i(x) \, dx = a_i \int x f_i(x) \, dx = a_i \mu_i$. Also,

$$\text{Var}[Y] = \text{Var}[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}[X_i, X_j].$$

Why? $\text{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$ and $Y^2 = (a_1 X_1 + \ldots + a_n X_n) \ast (a_1 X_1 + \ldots + a_n X_n)$, which generates the cross-product terms that define the $\text{Cov}[X_i, X_j]$. It takes some algebra.

**Theorem:** If the $X_i$ have (possibly different, possibly correlated) normal distributions, then $Y$ is normally distributed.
If one looks at the definitions, one sees that \( \text{Cov}[X_i, X_i] \) is just the variance \( \sigma_i^2 \). So one can write

\[
\text{Var}[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j \text{Cov}[X_i, X_j].
\]

In the special case when the random variables are independent, then the covariances are all zero and this simplifies to

\[
\text{Var}[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2.
\]

In particular, \( \mathbb{E}[X_1 - X_2] = \mu_1 - \mu_2 \). And if \( X_1 \) and \( X_2 \) are independent, then \( \text{Var}[X_1 - X_2] = \sigma_1^2 + \sigma_2^2 \).
9.2 Point Estimation

Statisticians provide two things:

- a point estimate of some quantity of interest, and
- a statement of the uncertainty in that estimate.

Other disciplines only provide the point estimate.

A **parameter** is some property of a distribution function, such as the mean, median, standard deviation, and so forth. A **point estimate** for a parameter is some statistic $h(X_1, \ldots, X_n)$ which, when evaluated for a random sample, gives a sensible approximation to the parameter.
The Central Limit Theorem indicates one of many approaches. If the parameter of interest is the population mean $\mu$, then the statistic $h(X_1, \ldots, X_n) = \bar{X}$ provides a sensible estimate of $\mu$.

In particular, we know the uncertainty in that estimate: $\sigma/\sqrt{n}$.

There are several properties one can want from a point estimate:

- unbiasedness
- minimum variance (i.e., minimum uncertainty)
- minimum mean squared error.

We discuss these in the context of several estimation strategies.
Besides the mean, other point estimates for common parameters are:

- the sample proportion $X/n$ for the population proportion $p$.
- the sample variance,
  \[ \hat{\sigma}^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]
  for the population variance.
- the average squared deviation,
  \[ s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]
  for the population variance.
- the 10% trimmed sample mean for the population mean; this is the average of the sample after removing the largest 5% of the values and the smallest 5% of the values.
9.3 Unbiased Estimates

A point estimate \( \hat{\theta} = h(X_1, \ldots, X_n) \) is said to be an **unbiased estimator** for a population parameter \( \theta \) if \( \mathbb{E}[\hat{\theta}] = \theta \).

The **bias** in a point estimate is \( \mathbb{E}[\hat{\theta}] - \theta = \text{bias}(\hat{\theta}) \). For unbiased estimates, this is zero.

The **mean squared error** of a point estimate is \( \text{Var}[\hat{\theta}] + \text{bias}^2(\hat{\theta}) \).

The book does not cover the mean squared error, but it is has many attractive features. In particular, it is sometimes possible to trade-off a small bias for a large reduction in variance, and this leads to better accuracy.
Recall from the previous lecture how bias and variance contribute differently to total error. The left target has small variance and small bias. The right target has large bias and large variance, the worst of both worlds.
If $X$ has the Bin$(n, p)$ distribution, then the sample proportion $X/n$ is an unbiased estimate for the parameter $p$.

$$\mathbb{E}[X/n] = \frac{1}{n} \mathbb{E}[X] = \frac{1}{n} np = p.$$

The sample mean $\bar{X}$ of a random sample is unbiased for the population mean $\mu$. We know this from the properties of linear combinations:

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} (n\mu) = \mu.$$

In this case we also know the variance of the estimator:

$$\text{Var} [\bar{X}] = \left( \frac{1}{n} \right)^2 \sum_{i=1}^{n} \sigma^2 = \left( \frac{1}{n} \right)^2 (n\sigma^2) = \frac{\sigma^2}{n}.$$
**Example: Second-Price Auctions.** Let $X$ be uniformly distributed on the interval $[0, \theta]$ where $\theta$ is the unknown parameter. You have a random sample $X_1, \ldots, X_n$ of losing bids and use the statistic $\hat{\theta}_1 = \max\{X_1, \ldots, X_n\}$ as an estimate of $\theta$, the winning bid. Let

$$F(z) = \mathbb{P}[X \leq z] = \int_0^z 1/\theta \, dx = z/\theta.$$  

Then let $G(z) = \mathbb{P}[\hat{\theta}_1 \leq z] = \mathbb{P}[\max\{X_1, \ldots, X_n\} \leq z]$, and note that

$$\mathbb{P}[\max\{X_1, \ldots, X_n\} \leq z] = \mathbb{P}[X_1 \leq z \text{ and } \cdots \text{ and } X_n \leq z]$$

$$= \prod_{i=1}^n \mathbb{P}[X_i \leq z]$$

The $\prod$ symbol multiplies its arguments just as the $\sum$ symbol adds them. We have shown that

$$G(z) = \prod_{i=1}^n \mathbb{P}[X_i \leq z] = \prod_{i=1}^n z/\theta = (z/\theta)^n.$$
So the distribution of the sample maximum is \( G(z) = (z/\theta)^n \) for 
\( 0 \leq z \leq \theta \) and thus the probability density function of the maximum is 
\( g(z) = n(1/\theta)^n z^{n-1} \) on \( 0 \leq z \leq \theta \).

Since we know the density, we can find the expected value of \( Z \), where \( Z \) 
is the sample maximum:

\[
\mathbb{E}[Z] = \int_0^\theta z \ast g(z) \, dz = \int_0^\theta z \ast \frac{n}{\theta^n} z^{n-1} \, dz = \frac{n}{\theta^n} \frac{1}{n+1} \theta^{n+1} \bigg|_0^\theta
\]

\[
= \frac{n}{n+1} \theta.
\]

So the estimator \( \hat{\theta}_1 \) of \( \theta \) has a small bias: 
\( \frac{n}{n+1} \theta - \theta = -\theta/(n + 1) \).

One can \textbf{unbias} this \( \hat{\theta}_1 \) estimator by using the new estimator 
\( \hat{\theta}_2 = (n + 1)/n \ast \hat{\theta}_1 = (n + 1)/n \ast \max\{X_1, \ldots, X_n\} \).
9.4 Minimum Variance Unbiased Estimators

The book holds that the first requirement for a good estimator of a parameter is that it be unbiased. When there are several unbiased estimators, one should use the one that has smallest variance.

This is not the only way to frame the problem of selecting an estimator. For example, one might want the estimator which:

- minimized the mean squared error,
- had the largest probability of being within some fixed distance from the true value,
- was unbiased and minimized something more practical than the variance.
Consider again the case of a random sample from Unif(0, \theta) distribution. Clearly, \( \mathbb{E}[\bar{X}] = \theta/2 \) so \( \hat{\theta}_3 = 2\bar{X} \) is an unbiased estimator of \( \theta \).

We now have two candidate estimators:

\[ \hat{\theta}_2 = (n + 1)/n \ast \max\{X_1, \ldots, X_n\} \text{ and } \hat{\theta}_3 = 2\bar{X}. \]

Which has the smaller variance?

Since \( \hat{\theta}_3 \) is a linear combination, we know that its variance is \( 4 \ast (\sigma^2/n) \) where \( \sigma^2 \) is the variance of the Unif(0, \theta) distribution. And we know* that the variance of the Unif(0, \theta) distribution is \( \theta^2/12 \). Thus

\[ \text{Var} [\hat{\theta}_3] = \frac{\theta^2}{3n}. \]
To find the variance of \( \hat{\theta}_2 \) we first find

\[
\mathbb{E}[Z^2] = \int_0^\theta z^2 g(z) \, dz = \int_0^\theta z^2 \frac{n}{\theta} z^{n-1} \, dz = \frac{n}{n+2} \theta^2.
\]

Since \( \text{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 \), we have

\[
\text{Var}[Z] = \frac{n}{n+2} \theta^2 - \left[ \frac{n}{n+1} \theta \right]^2 = \left[ \frac{n}{(n+2)(n+1)^2} \right] \theta^2.
\]

Since \( \hat{\theta}_2 = (n+1)/n * Z \), then

\[
\text{Var}[\hat{\theta}_2] = \left( \frac{n+1}{n} \right)^2 * \left[ \frac{n}{(n+2)(n+1)^2} \right] \theta^2 = \frac{1}{n(n+2)} \theta^2.
\]

A little algebra shows that \( n(n+2) > 3n \) for all \( n > 1 \), so \( \hat{\theta}_2 \) is better than \( \hat{\theta}_3 \).
Let $X \sim \text{Unif}(0, \theta)$. Then $\mathbb{E}[X] = \theta/2$. To find the variance of $X$ we first calculate $\mathbb{E}[X^2]$, and then use $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\mathbb{E}[X^2] = \int_0^\theta x^2 f(x) \, dx = \int_0^\theta x^2 \cdot 1/\theta \, dx = \frac{1}{\theta} x^3 / 3 \bigg|_0^\theta = \frac{\theta^2}{3}$$

so

$$\text{Var}[X] = \frac{\theta^2}{3} - \left( \frac{\theta}{2} \right)^2 = \frac{\theta^2}{12}.$$ 

Quod erat demonstrandum.