1 Samples and Sampling Distributions

Definition. We say $X_1, \ldots, X_n$ is a random sample of size $n$ if each $X_i$ is drawn independently from the same pdf, $f(x_i, \theta)$.

Notes:

1. $\{X_i\}_{i=1}^n$ is sometimes said to be an independent and identically distributed (i.i.d.) random sample.

2. $\theta$ is a vector of parameters (e.g., $\theta = (\mu, \sigma^2)$).

3. Three data types: time series, cross sectional, and panel.

1.1 Descriptive Statistics

Definition. A function of one or more random variables that does not depend on any unknown parameters is a statistic.

1. Measures of Central Tendency.

   - **Mean.** $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.
   
   - **Median.** Let $Y_1, ..., Y_n$ be the reordering of $X_1, ..., X_n$ from smallest to largest. $Y_i$ is called the $i^{th}$ order statistic of $X_1, ..., X_n$. The median is defined as $Y_{(n+1)/2}$.
   
   - **Mode.** Most frequent $X_i$.


   - $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.
   
   - $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$.


   - **Covariance.** $s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$.
   
   - **Correlation.** $r_{xy} = s_{xy}/(s_x s_y)$ where $-1 \leq r_{xy} \leq 1$. 

1
1.2 Sampling Distribution

Definition. A statistic (e.g., $Y$, $\bar{X}$ and $s_{xy}$) is a random variable with a distribution called a sampling distribution.

Example. If $X_1, ..., X_n$ are a random sample with mean $\mu$ and variance $\sigma_x^2$, then $\bar{X}$ is a random variable with a sampling distribution that has mean $\mu$ and variance $\sigma_x^2/n$.

Proof.
1. $E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n}(n\mu) = \mu$.
2. $Var(\bar{X}) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}(n\sigma_x^2) = \frac{1}{n}\sigma_x^2$.

See MATLAB example #6 for the sampling distributions of $\bar{X}$ where $X_i \sim N(0, 1)$ with $n = 3, 10, 100$.

2 Finite Sample Estimation

Definition. An estimator is a rule for using the sample data to form either a point (i.e., single value) or interval (i.e., range of values) estimate.

2.1 Estimation Criterion

1. Unbiasedness. An estimator is unbiased if $E(\hat{\theta}) = \theta$.

Examples.

- $\bar{X}$ is an unbiased estimator of $\mu$.
- The statistic $Z = X + 1000$ if coin is “heads”, $Z = X - 1000$ if coin is “tails” is an unbiased estimator of $\mu$.

2. Efficient Unbiasedness. An unbiased estimator $\hat{\theta}_1$ is efficient if there is no $\hat{\theta}_i$ such that $var(\hat{\theta}_i) < var(\hat{\theta}_1), i \neq 1$.

Example continued.

- $var(\bar{X}) = \sigma_x^2/n$  
- $var(Z) = 0.5E(X + 1000 - \mu)^2 + 0.5E(X - 1000 - \mu)^2$.

3. Mean-Square Error. The mean-square error (MSE) of $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = var(\hat{\theta}) + bias(\hat{\theta})^2.$$ 

Notes:
1. Given some regularity conditions, the \( \text{var}(\hat{\theta}) \) will never be smaller than the Cramer-Rao lower bound.

2. A minimum variance unbiased estimator (MVUE) is an efficient unbiased estimator among all linear and nonlinear estimators.

3. A minimum variance linear unbiased estimator (or sometimes called best linear unbiased estimator, BLUE) is an efficient estimator among all linear estimators.

4. Attaining the Cramer-Rao lower bound \( \Rightarrow \) efficiency. However, efficiency \( \not\Rightarrow \) attaining the Cramer-Rao lower bound.

5. A linear estimator is one that is a linear function of the data.

### 2.2 \( s^2 \text{ versus } \hat{s}^2 \). Which is a better estimator?

- Is \( s^2 \) unbiased?

\[
E(s^2) = E\left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} [(X_i - \mu) - (\bar{X} - \mu)]^2\right) = \frac{1}{n-1} \left[E \sum_{i=1}^{n} (X_i - \mu)^2 - 2E \sum_{i=1}^{n} (X_i - \mu)(\bar{X} - \mu) + E \sum_{i=1}^{n} (\bar{X} - \mu)^2\right] = \frac{1}{n-1} \left[\sum_{i=1}^{n} E(X_i - \mu)^2 - 2nE(\bar{X} - \mu)^2 + \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 + nE(\bar{X} - \mu)^2\right] = \frac{1}{n-1} \left[\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right] = \frac{1}{n-1} \left[n\sigma^2 - \frac{n^2}{n}\right] = \sigma^2.
\]

Yes, \( s^2 \) is an unbiased estimator of \( \sigma^2 \).

- Is \( \hat{s}^2 \) unbiased?

\[
E(\hat{s}^2) = E\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2\right) = \frac{n-1}{n} E(s^2) = \frac{n-1}{n} \frac{n}{n} \sigma^2.
\]

No, \( \hat{s}^2 \) is not an unbiased estimator of \( \sigma^2 \). However, the bias clearly shrinks as \( n \) grows.

- Assuming a normal distribution, what is the variance of \( s^2 \)?

\[
\text{var}(s^2) = \frac{2\sigma^4}{(n-1)} = \text{MSE}(s^2).
\]

- What is the variance of \( \hat{s}^2 \)?

\[
\text{var}(\hat{s}^2) = \text{var}\left(\frac{n-1}{n} s^2\right) = \left(\frac{n-1}{n}\right)^2 \text{var}(s^2) < \text{var}(s^2).
\]
• Which estimator has a smaller MSE?

\[
MSE(\hat{\sigma}^2) = \text{var}(\hat{\sigma}^2) + \text{bias}(\hat{\sigma}^2)^2 = \left(\frac{n-1}{n}\right)^2 \text{var}(s^2) + \left(-\frac{1}{n}\sigma^2\right)^2
\]

\[
= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{(n-1)} + \frac{1}{n^2}\sigma^4
\]

\[
= \frac{(2n-1)\sigma^4}{n^2} = \frac{2\sigma^4}{n} - \frac{\sigma^4}{n^2} < MSE(s^2).
\]

Therefore, \(\hat{\sigma}^2\) has a smaller MSE than \(s^2\) under normality.

See MATLAB example #7 for an example of the sampling distribution for the estimator of variance.

3 Large-Sample Distribution Theory

Large-sample distribution theory is important because the small-sample distribution of random variables are often unknown.

3.1 Convergence in Probability

Definition. Let \(X_n\) be a random variable whose distribution depends on \(n\). We say \(X_n\) converges in probability to \(c\) (or \(\text{plim } X_n = c\)) if \(\lim_{n \to \infty} \Pr(|X_n - c| > \epsilon) = 0\) for every \(\epsilon > 0\). If \(X_n\) has mean \(\mu_n\) and variance \(\sigma_n^2\) with limits \(c\) and 0, then \(X_n\) convergence in mean square to \(c\).

Notes:

1. \(\hat{\theta}\) is a consistent estimator of \(\theta\) iff \(\text{plim}(\hat{\theta}) = \theta\).

2. Convergence in mean square \(\implies\) convergence in probability. Convergence in probability \(\not\implies\) convergence in mean square.

3. Slutsky’s Theorem. If \(g(X)\) is a continuous function not in \(n\), \(\text{plim}(g(X)) = g(\text{plim}(X))\). For example, \(E(X_n^2) = \mu^2\) but \(\text{plim}(\bar{X}_n^2) = \text{plim}(\bar{X}_n)^2 = \mu^2\).

4. Jensen’s Inequality. If \(g(X_n)\) is concave in \(X_n\), \(g(E(X_n)) \geq E(g(X_n))\).

5. Using Slutsky’s theorem where \(\text{plim } X_n = c\) and \(\text{plim } X_n = d\),

(a) \(\text{plim}(X_n + Y_n) = c + d\).

(b) \(\text{plim}(X_nY_n) = cd\).

(c) \(\text{plim}(X_n/Y_n) = c/d, d \neq 0\).
### 3.2 Convergence in Distribution

**Definition.** $X_n$ is said to converge in distribution to $F(x)$ if $\lim_{n \to \infty} F_n(x) = F(x)$ at all continuity points of $F(x)$.

Notes:

1. Converge in distribution: $X_n \overset{d}{\to} X$.
2. $F(x)$ is the limiting distribution of $X_n$.
3. The mean and variance of $F(x)$ are called the limiting mean and limiting variance.
4. Rules when $X_n \overset{d}{\to} X$ and $Y_n \overset{p}{\to} c$.
   
   (a) $X_n Y_n \overset{d}{\to} cX$, $X_n + Y_n \overset{d}{\to} X + c$ and $X_n / Y_n \overset{d}{\to} X / c$.
   
   (b) If $g(X_n)$ is a continuous function, $g(X_n) \overset{d}{\to} g(X)$.
   
   (c) If $\text{plim}(X_n - Y_n) = 0$, then $Y_n \overset{d}{\to} X$, provided a limiting distribution for $Y_n$ exists.

**Example.** The pdf of the $n^{th}$ order statistic from the random sample $X_1, \ldots, X_n$, where

$$f(x) = \frac{1}{\theta}, \ 0 < x \leq \theta; \ 0 < \theta < \infty$$

(and zero elsewhere) is

$$g_n(y) = \frac{ny^{n-1}}{\theta^n}, \ 0 < y \leq \theta$$

and zero elsewhere. Find the limiting distribution $G(y)$.

**Answer.** First, we need to find $G_n(y)$.

$$G_n(y) = \int_0^y \frac{nz^{n-1}}{\theta^n} dz = \left(\frac{z}{\theta}\right)^n \bigg|_{z=0}^{y} = \left(\frac{y}{\theta}\right)^n, \ 0 < y < \theta$$

$$= 1, \ y \geq \theta.$$

Now the limiting distribution is

$$G(y) = \lim_{n \to \infty} G_n(y) = 0, \ 0 < y < \theta$$

$$= 1, \ y \geq \theta.$$ 

Therefore, $G(y)$ is a degenerate cdf with all the mass at $Y = \theta$.

### 3.3 Central Limit Theorem

**Question.** What is the limiting distribution of $\bar{X}_n$?
Answer. A spike at \( \mu \).

Consider a stabilizing transformation of \( \bar{X}_n \):

\[
Y = \sqrt{n}(\bar{X}_n - \mu).
\]

Definition. Let \( X_1, \ldots, X_n \) denote a random sample from any distribution with finite mean \( \mu \) and finite variance \( \sigma^2 \). Then

\[
\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).
\]

See MATLAB example \#8 for an example of the CLT in action.

### 3.4 Asymptotic Distributions

Definition. An asymptotic distribution is used to approximate a true (and possibly unknown) finite-sample distribution.

Notes:

1. The mean and variance of an asymptotic distribution are called the asymptotic mean and asymptotic variance.

2. \( \hat{\theta} \) is said to be asymptotically efficient if \( \text{asy.var}(\hat{\theta}) \) is less than or equal to the asymptotic variance of any other consistent estimator.

3. Occasionally you will hear the term asymptotically unbiased: \( \lim_{n \to \infty} E(\hat{\theta}) = \theta \).

Example \#1. Consider the random variable

\[
\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).
\]

We say that

\( \bar{X}_n \xrightarrow{asy} N(\mu, \sigma^2/n) \).

Example \#2. Find the asymptotic distribution of \( Z_n = n(1 - Y_n) \), where \( Y_n \) is the \( n^{th} \) order statistic from the \( \text{uniform}[0, 1] \) random sample \( X_1, \ldots, X_n \).

Answer. Start by finding the limiting distribution of \( Y_n \):

\[
G(y_n) = \begin{cases} 
0, & 0 \leq y_n < 1 \\
1, & y_n = 1.
\end{cases}
\]

Therefore, \( Y_n \) has a degenerate limiting pdf with all the mass at \( Y_n = 1 \).
The pdf for $Z_n$ can be found by the change of variable technique:

$$h_n(z) = (1 - z/n)^{n-1}, \quad 0 < z < n$$

and zero elsewhere. The cdf for $Z_n$ is

$$H_n(z) = 0, \quad z < 0$$
$$= \int_0^z (1 - w/n)^{n-1} dw = 1 - (1 - z/n)^n, \quad 0 \leq z < n$$
$$= 1, \quad z \geq n$$

and its limiting distribution $H(z)$ is

$$\lim_{n \to \infty} H_n(z) = 0, \quad z < 0$$
$$= 1 - e^{-z}, \quad 0 \leq z < \infty.$$ 

Therefore, $Z_n \xrightarrow{asy} \text{exponential}(\lambda = 1)$. See MATLAB example #9 for an example of an asymptotic exponential sampling distribution.