Sample Space Ω, ω ∈ Ω

- Observing whether a single individual is male or female:

$$\Omega = \{F, M\}$$

- Pair of individuals and observed their genders in order:

$$\Omega = \{(F, F), (F, M), (M, F), (M, M)\}$$

- Select *n* people and count the number of females:

$$\Omega = \{0, \ldots, n\}$$

- For limits problems, the points in Ω are infinite sequences

Random variables are functions from Ω into the set of real numbers

$$Pr\{X \in B\} = Pr(\{\omega \in \Omega : X(\omega) \in B\}$$

Random sample $X_1(\omega), \ldots, X_n(\omega)$

$$T = T(X_1, \ldots, X_n)$$

$$T = T_n(\omega)$$

Let $n \to \infty$

To see what happens for large samples

Modes of Convergence

- Almost Sure Convergence
- Convergence in Probability
- Convergence in Distribution

Almost Sure Convergence

We say that T_n converges *almost surely* to T, and write $T_n \stackrel{a.s.}{\rightarrow}$ if

$$Pr\{\omega : \lim_{n \rightarrow \infty} T_n(\omega) = T(\omega)\} = 1.$$

Acts like an ordinary limit, except possibly on a set of probability zero.

All the usual rules apply.

Strong Law of Large Numbers

$$\overline{X}_n \xrightarrow{a.s.} \mu$$

The only condition required for this to hold is the existence of the expected value.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables; let X be a general random variable from this same distribution, and $Y=g(X)$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(X_i) &= \frac{1}{n}\sum_{i=1}^{n} Y_i \stackrel{a.s.}{\longrightarrow} E(Y) \\
&= E(g(X))
\end{aligned}
$$

So for example

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \overset{a.s.}{\longrightarrow} E(X^k)$$

$$\frac{1}{n}\sum_{i=1}^{n} U_i^2 V_i W_i^3 \overset{a.s.}{\longrightarrow} E(U^2 V W^3)$$

That is, sample moments converge almost surely to population moments.

Convergence in Probability

We say that T_n converges *in probability* to T, and write $T_n \xrightarrow{P}$ if for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\{|T_n - T| < \epsilon\} = 1$$

- Convergence in probability (say to a constant θ) means no matter how small the interval around θ, for large enough n ($n>N_1$) the probability of getting that close to θ is as close to one1 as you like.
- Almost sure convergence means no matter how small the interval around θ, for large enough n ($n>N_2$) the probability of getting that close to θ equals one.
- Almost Sure Convergence => Convergence in Probability
- Strong Law of Large Numbers => Weak Law of Large Numbers

Convergence in Distribution

Denote the cumulative distribution functions of $T_1, T_2, \ldots$ by $F_1(t), F_2(t), \ldots$ respectively, and denote the cumulative distribution function of T by $F(t)$.

We say that T_n converges *in distribution* to T, and write $T_n \xrightarrow{d} T$ if for every point t at which F is continuous,

$$\lim_{n\to\infty} F_n(t) = F(t)$$

Univariate Central Limit Theorem says

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0,1)$$

Connections among the Modes of Convergence

- $T_n \xrightarrow{a.s.} T \Rightarrow T_n \xrightarrow{P} T \Rightarrow T_n \xrightarrow{d} T.$
- If a is a constant, $T_n \xrightarrow{d} a \Rightarrow T_n \xrightarrow{P} a.$

Consistency

$T_n = T_n(X_1, \ldots, X_n)$ is a statistic estimating a parameter θ

The statistic T_n is said to be *consistent* for θ if $T_n \xrightarrow{P} \theta$.

$$\lim_{n \to \infty} P\{|T_n - \theta| < \epsilon\} = 1$$

The statistic T_n is said to be *strongly consistent* for θ if $T_n \xrightarrow{a.s.} \theta$.

Strong consistency implies ordinary consistency.

Consistency of the Sample Variance

$$\begin{aligned}\widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\end{aligned}$$

By SLLN, $\overline{X}_n \stackrel{a.s.}{\rightarrow} \mu$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \stackrel{a.s.}{\rightarrow} E(X^2) = \sigma^2 + \mu^2$

Because the function $g(x, y) = x - y^2$ is continuous,

$$\widehat{\sigma}_n^2 = g(\frac{1}{n}\sum_{i=1}^{n} X_i^2, \overline{X}_n) \stackrel{a.s.}{\rightarrow} g(\sigma^2 + \mu^2, \mu) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Consistency of the Sample Covariance

$$\widehat{\sigma}_{1,2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}_n \overline{Y}_n$$

By SLLN, $\overline{X}_n \stackrel{a.s.}{\rightarrow} E(X)$, $\overline{Y}_n \stackrel{a.s.}{\rightarrow} E(Y)$, and $\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \stackrel{a.s.}{\rightarrow} E(XY)$

Because the function $g(x, y, z) = x - yz$ is continuous,

$$\begin{aligned}
\widehat{\sigma}_{1,2} &= g\left(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i, \overline{X}_n, \overline{Y}_n\right) \stackrel{a.s.}{\rightarrow} g\left(E(XY), E(X), E(Y)\right) \\
&= E(XY) - E(X)E(Y) = Cov(X, Y) \\
&= \sigma_{1,2}
\end{aligned}$$

Convergence of Random Vectors

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)
   - $\star$ $\mathbf{T}_n \stackrel{a.s.}{\to} \mathbf{T}$ means $P\{\omega : \lim_{n\to\infty} \mathbf{T}_n(\omega) = \mathbf{T}(\omega)\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{P}{\to} \mathbf{T}$ means $\forall\epsilon > 0$, $\lim_{n\to\infty} P\{||\mathbf{T}_n - \mathbf{T}|| < \epsilon\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$ means for every continuity point $\mathbf{t}$ of $F_{\mathbf{T}}$, $\lim_{n\to\infty} F_{\mathbf{T}_n}(\mathbf{t}) = F_{\mathbf{T}}(\mathbf{t})$.
2. $\mathbf{T}_n \stackrel{a.s.}{\to} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{P}{\to} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$.
3. If $\mathbf{a}$ is a vector of constants, $\mathbf{T}_n \stackrel{d}{\to} \mathbf{a} \Rightarrow \mathbf{T}_n \stackrel{P}{\to} \mathbf{a}$.
4. Strong Law of Large Numbers (SLLN): Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be independent and identically distributed random vectors with finite first moment, and let $\mathbf{X}$ be a general random vector from the same ditribution. Then $\overline{\mathbf{X}}_n \stackrel{a.s.}{\to} E(\mathbf{X})$.
5. Central Limit Theorem: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors with expected value vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then $\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu})$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

6. Slutsky Theorems for Convergence in Distribution:

   (a) If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and if $f : \mathbb{R}^m \longrightarrow \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \xrightarrow{d} f(\mathbf{T})$.

   (b) If $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \xrightarrow{P} 0$, then $\mathbf{Y}_n \xrightarrow{d} \mathbf{T}$.

   (c) If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $\mathbf{Y}_n \xrightarrow{d} \mathbf{c}$, then

$$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{T} \\ \mathbf{c} \end{pmatrix}$$

7. Slutsky Theorems for Convergence in Probability:

   (a) If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and if $f : \mathbb{R}^m \longrightarrow \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \xrightarrow{P} f(\mathbf{T})$.

   (b) If $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \xrightarrow{P} 0$, then $\mathbf{Y}_n \xrightarrow{P} \mathbf{T}$.

   (c) If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and $\mathbf{Y}_n \xrightarrow{P} \mathbf{Y}$, then

$$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mathbf{T} \\ \mathbf{Y} \end{pmatrix}$$

8. Delta Method (Theorem of Cramér, Ferguson p. 45): Let $g : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ be such that the elements of $\dot{\mathrm{g}}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k \times d}$ are continuous in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^d$. If $\mathbf{T}_n$ is a sequence of d-dimensional random vectors such that $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{T}$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \dot{\mathrm{g}}(\boldsymbol{\theta})\mathbf{T}$. In particular, if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \dot{\mathrm{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\theta})')$.