Sample Space Ω, $\omega \in \Omega$

- Observing whether a single individual is male or female:

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

- Pair of individuals: observe 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 $X_1(\omega), \ldots, X_n(\omega)$

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

$$T = T_n(\omega)$$

$$\text{Let } n \to \infty$$

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

$$\overline{X}_n \xrightarrow{a.s.} E(X_i) = \mu$$

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

Probability is long run relative frequency

- Statistical experiment: Probability of "success" is *p*
- Carry out the experiment many times independently.
- Code the results X_i=1 if success, X_i=0 for failure, i = 1, 2, …

$$
\begin{aligned}
E(X_i) &= \sum_{x_i} x_i Pr\{X_1 = x_i\} \\
&= 0 \cdot (1-p) + 1 \cdot p \\
&= p
\end{aligned}
$$

Relative frequency is

$$
\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n \stackrel{a.s.}{\longrightarrow} p
$$

Simulation

- Weather model
- Performance of statistical methods
- Estimate almost any probability that's hard to figure out
- Confidence intervals for the estimate

A hard elementary problem

- Roll a fair die 13 times and observe the number each time.

- What is the probability that the sum of the numbers is divisible by 3?

```
> die = c(1,1,1,1,1,1)/6; die
[1] 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
> rmultinom(1,1,die)
     [,1]
[1,]    0
[2,]    0
[3,]    1
[4,]    0
[5,]    0
[6,]    0
> rmultinom(1,13,die)
     [,1]
[1,]    5
[2,]    2
[3,]    1
[4,]    2
[5,]    2
[6,]    1
```

```
> tot = sum(rmultinom(1,13,die)*(1:6))
> tot
[1] 42
> tot/3 == floor(tot/3)
[1] TRUE
> 42/3
[1] 14
```

Estimated Probability

```
> nsim = 1000 # nsim is the Monte Carlo sample size
> set.seed(9999) # So I can reproduce the numbers if desired.
> kount = numeric(nsim)
> for(i in 1:nsim)
+     {
+     tot = sum(rmultinom(1,13,die)*(1:6))
+     kount[i] = (tot/3 == floor(tot/3))
+     # Logical will be converted to numeric
+     }

> kount[1:20]
 [1] 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

> xbar = mean(kount); xbar
[1] 0.329
```

Confidence Interval

```
> z = qnorm(0.995); z
[1] 2.575829
> pnorm(z)-pnorm(-z) # Just to check
[1] 0.99

> margerror99 = sqrt(xbar*(1-xbar)/nsim)*z; margerror99
[1] 0.03827157

> cat("Estimated probability is ",xbar," with 99% margin of error ",
+      margerror99,"\n")

Estimated probability is  0.329  with 99% margin of error  0.03827157

> cat("99% Confidence interval from ",xbar-margerror99," to ",
+      xbar+margerror99,"\n")

99% Confidence interval from  0.2907284  to  0.3672716
```

Recall the Change of Variables formula: Let Y = g(X)

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Or, for discrete random variables

$$E(Y) = \sum_{y} y\, p_Y(y) = \sum_{x} g(x)\, p_X(x)$$

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

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

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

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 $T_n \xrightarrow{P} T$ 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 one 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, $\overline{X}_n \xrightarrow{a.s.} \mu$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{a.s.} E(X^2) = \sigma^2 + \mu^2$

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

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

Convergence of Random Vectors

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)
   - $\star$ $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T}$ means $P\{\omega : \lim_{n\to\infty} \mathbf{T}_n(\omega) = \mathbf{T}(\omega)\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T}$ means $\forall \epsilon > 0$, $\lim_{n\to\infty} P\{||\mathbf{T}_n - \mathbf{T}|| < \epsilon\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$ 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. $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$.
3. If $\mathbf{a}$ is a vector of constants, $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{a} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{a}$.
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 $\overline{\mathbf{X}}_n \stackrel{a.s.}{\rightarrow} E(\mathbf{X})$.
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 $f : \mathbb{R}^m \to \mathbb{R}^q$ (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 $f : \mathbb{R}^m \to \mathbb{R}^q$ (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 $g : \mathbb{R}^d \to \mathbb{R}^k$ be such that the elements of $\dot{\text{g}}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k \times d}$ 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 $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \dot{\text{g}}(\boldsymbol{\theta})\mathbf{T}$. In particular, if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathbf{Y} \sim N(\mathbf{0}, \dot{\text{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\text{g}}(\boldsymbol{\theta})')$.

An application of the Slutsky Theorems

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} ?(\mu, \sigma^2)$

By CLT, $Y_n = \sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} Y \sim N(0, \sigma^2)$

Let $\widehat{\sigma}_n$ be *any* consistent estimator of σ.

Then by 6c, $\mathbf{T}_n = \begin{pmatrix} Y_n \\ \widehat{\sigma}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Y \\ \sigma \end{pmatrix} = \mathbf{T}$

The function $f(x, y) = x/y$ is continuous except if $y = 0$ so by 6a,

$$f(\mathbf{T}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \xrightarrow{d} f(\mathbf{T}) = \frac{Y}{\sigma} \sim N(0, 1)$$

Because

- The multivariate CLT establishes convergence to a multivariate normal, and
- Vectors of MLEs are approximately normal for large samples
- We need to look at random vectors and the multivariate normal distribution.