1 Random Samples

Let $X_1, \ldots, X_n \sim F$. A statistic is any function $T_n = g(X_1, \ldots, X_n)$. Recall that the sample mean is

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and sample variance is

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

Let $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = \text{Var}(X_i)$. Recall that

$$\mathbb{E}(\bar{X}_n) = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \mathbb{E}(S_n^2) = \sigma^2.$$

**Theorem 1** If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ then $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

**Proof.** We know that $M_{X_i}(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}$. So,

$$M_{\bar{X}_n}(t) = \mathbb{E}(e^{t\bar{X}_n}) = \mathbb{E}(e^{\frac{t}{n} \sum_{i=1}^{n} X_i})$$

$$= (\mathbb{E}(e^{tX_i/n})^n = (M_{X_i}(t/n))^n = \left(e^{(\mu t/n) + \frac{\sigma^2 t^2}{2(n^2)}}\right)^n = \exp \left\{ \mu t + \frac{\sigma^2 t^2}{2n} \right\}$$

which is the mgf of a $N(\mu, \sigma^2/n)$. $\blacksquare$

**Example 2** Let $X_{(1)}, \ldots, X_{(n)}$ denoted the ordered values:

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

Then $X_{(1)}, \ldots, X_{(n)}$ are called the order statistics. And $T_n = (X_{(1)}, \ldots, X_{(n)})$ is a statistic.
2 Convergence

Let \( X_1, X_2, \ldots \) be a sequence of random variables and let \( X \) be another random variable. Let \( F_n \) denote the cdf of \( X_n \) and let \( F \) denote the cdf of \( X \). We are going to study different types of convergence.

**Example:** A good example to keep in mind is the following. Let \( Y_1, Y_2, \ldots, \) be a sequence of iid random variables. Let \( X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \) be the average of the first \( n \) of the \( Y_i \)'s. This defines a new sequence \( X_1, X_2, \ldots, \). In other words, the sequence of interest \( X_1, X_2, \ldots, \) might be a sequence of statistics based on some other sequence of iid random variables. **Note that the original sequence \( Y_1, Y_2, \ldots, \) is iid but the sequence \( X_1, X_2, \ldots, \) is not iid.**

1. \( X_n \) converges almost surely to \( X \), written \( X_n \overset{a.s.}{\rightarrow} X \), if, for every \( \epsilon > 0 \),
   \[
   \mathbb{P}\left( \lim_{n \rightarrow \infty} |X_n - X| < \epsilon \right) = 1.
   \]
   \( X_n \) converges almost surely to a constant \( c \), written \( X_n \overset{a.s.}{\rightarrow} c \), if, for every \( \epsilon > 0 \),
   \[
   \mathbb{P}\left( \lim_{n \rightarrow \infty} X_n = c \right) = 1.
   \]

2. \( X_n \) converges to \( X \) in probability, written \( X_n \overset{P}{\rightarrow} X \), if, for every \( \epsilon > 0 \),
   \[
   \mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0
   \]
as \( n \rightarrow \infty \). In other words, \( X_n - X = o_P(1) \).
   \( X_n \) converges to \( c \) in probability, written \( X_n \overset{P}{\rightarrow} c \), if, for every \( \epsilon > 0 \),
   \[
   \mathbb{P}(|X_n - c| > \epsilon) \rightarrow 0
   \]
as \( n \rightarrow \infty \). In other words, \( X_n - c = o_P(1) \).

3. \( X_n \) converges to \( X \) in quadratic mean (also called convergence in \( L_2 \)), written \( X_n \overset{q.m.}{\rightarrow} X \), if
   \[
   \mathbb{E}(X_n - X)^2 \rightarrow 0
   \]
as \( n \rightarrow \infty \).
   \( X_n \) converges to \( c \) in quadratic mean, written \( X_n \overset{q.m.}{\rightarrow} c \), if
   \[
   \mathbb{E}(X_n - c)^2 \rightarrow 0
   \]
as \( n \rightarrow \infty \).
4. **$X_n$ converges to $X$ in distribution**, written $X_n \rightsquigarrow X$, if

$$\lim_{n \to \infty} F_n(t) = F(t) \quad (7)$$

at all $t$ for which $F$ is continuous.

**$X_n$ converges to $c$ in distribution**, written $X_n \rightsquigarrow c$, if

$$\lim_{n \to \infty} F_n(t) = \delta_c(t) \quad (8)$$

at all $t \neq c$ where $\delta_c(t) = 0$ if $t < c$ and $\delta_c(t) = 1$ if $t \geq c$.

**Theorem 3** Convergence in probability does not imply almost sure convergence.

**Proof.** Let $\Omega = [0,1]$. Let $P$ be uniform on $[0,1]$. We draw $S \sim P$. Let $X(s) = s$ and let

$$X_1 = s + I_{[0,1]}(s), \quad X_2 = s + I_{[0,1/2]}(s), \quad X_3 = s + I_{[1/2,1]}(s)$$

$$X_4 = s + I_{[0,1/3]}(s), \quad X_5 = s + I_{[1/3,2/3]}(s), \quad X_6 = s + I_{[2/3,1]}(s)$$

etc. Then $X_n \xrightarrow{P} X$. But, for each $s$, $X_n(s)$ does not converge to $X(s)$. Hence, $X_n$ does not converge almost surely to $X$. In fact, $P(\{s \in \Omega : \lim_n X_n(s) = X(s)\}) = 0$. ■

**Example 4** Let $X_n \sim N(0,1/n)$. Intuitively, $X_n$ is concentrating at 0 so we would like to say that $X_n$ converges to 0. Let’s see if this is true. Let $F$ be the distribution function for a point mass at 0. Note that $\sqrt{n}X_n \sim N(0,1)$. Let $Z$ denote a standard normal random variable. For $t < 0$,

$$F_n(t) = P(X_n < t) = P(\sqrt{n}X_n < \sqrt{n}t) = P(Z < \sqrt{n}t) \to 0$$

since $\sqrt{n}t \to -\infty$. For $t > 0$,

$$F_n(t) = P(X_n < t) = P(\sqrt{n}X_n < \sqrt{n}t) = P(Z < \sqrt{n}t) \to 1$$

since $\sqrt{n}t \to \infty$. Hence, $F_n(t) \to F(t)$ for all $t \neq 0$ and so $X_n \rightsquigarrow 0$. Notice that $F_n(0) = 1/2 \neq F(1/2) = 1$ so convergence fails at $t = 0$. That doesn’t matter because $t = 0$ is not a continuity point of $F$ and the definition of convergence in distribution only requires convergence at continuity points.

Now consider convergence in probability. For any $\epsilon > 0$, using Markov’s inequality,

$$P(|X_n| > \epsilon) = P(|X_n|^2 > \epsilon^2) \leq \frac{E(X_n^2)}{\epsilon^2} = \frac{1}{\epsilon^2} \to 0$$

as $n \to \infty$. Hence, $X_n \xrightarrow{P} 0$. 

3
The next theorem gives the relationship between the types of convergence.

**Theorem 5** The following relationships hold:

(a) \( X_n \xrightarrow{qm} X \) implies that \( X_n \xrightarrow{P} X \).

(b) \( X_n \xrightarrow{P} X \) implies that \( X_n \xrightarrow{a.s} X \).

(c) If \( X_n \xrightarrow{a.s} X \) and if \( \Pr(X = c) = 1 \) for some real number \( c \), then \( X_n \xrightarrow{P} X \).

(d) \( X_n \xrightarrow{a.s} X \) implies \( X_n \xrightarrow{P} X \).

In general, none of the reverse implications hold except the special case in (c).

**Proof.** We start by proving (a). Suppose that \( X_n \xrightarrow{qm} X \). Fix \( \epsilon > 0 \). Then, using Markov’s inequality,

\[
\Pr(|X_n - X| > \epsilon) = \Pr(|X_n - X|^2 > \epsilon^2) \leq \frac{\mathbb{E}|X_n - X|^2}{\epsilon^2} \to 0.
\]

Proof of (b). Fix \( \epsilon > 0 \) and let \( x \) be a continuity point of \( F \). Then

\[
F_n(x) = \Pr(X_n \leq x) = \Pr(X_n \leq x, X \leq x + \epsilon) + \Pr(X_n \leq x, X > x + \epsilon) \\
\leq \Pr(X \leq x + \epsilon) + \Pr(|X_n - X| > \epsilon) \\
= F(x + \epsilon) + \Pr(|X_n - X| > \epsilon).
\]

Also,

\[
F(x - \epsilon) = \Pr(X \leq x - \epsilon) = \Pr(X \leq x - \epsilon, X_n \leq x) + \Pr(X \leq x - \epsilon, X_n > x) \\
\leq F_n(x) + \Pr(|X_n - X| > \epsilon).
\]

Hence,

\[
F(x - \epsilon) - \Pr(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) \leq F(x) + \Pr(|X_n - X| > \epsilon).
\]

Take the limit as \( n \to \infty \) to conclude that

\[
F(x - \epsilon) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x + \epsilon).
\]

This holds for all \( \epsilon > 0 \). Take the limit as \( \epsilon \to 0 \) and use the fact that \( F \) is continuous at \( x \) and conclude that \( \lim_n F_n(x) = F(x) \).

Proof of (c). Fix \( \epsilon > 0 \). Then,

\[
\Pr(|X_n - c| > \epsilon) = \Pr(X_n < c - \epsilon) + \Pr(X_n > c + \epsilon) \\
\leq \Pr(X_n \leq c - \epsilon) + \Pr(X_n > c + \epsilon) \\
= F_n(c - \epsilon) + 1 - F_n(c + \epsilon) \\
\to F(c - \epsilon) + 1 - F(c + \epsilon) \\
= 0 + 1 - 1 = 0.
\]
Proof of (d). Omitted.

Let us now show that the reverse implications do not hold.

**Convergence in probability does not imply convergence in quadratic mean.** Let $U \sim \text{Unif}(0,1)$ and let $X_n = \sqrt{nI_{(0,1/n)}(U)}$. Then $P(|X_n| > \epsilon) = P(\sqrt{nI_{(0,1/n)}(U)} > \epsilon) = P(0 \leq U < 1/n) = 1/n \to 0$. Hence, $X_n \xrightarrow{p} 0$. But $E(X_n^2) = n \int_0^{1/n} du = 1$ for all $n$ so $X_n$ does not converge in quadratic mean.

**Convergence in distribution does not imply convergence in probability.** Let $X \sim N(0,1)$.

Let $X_n = -X$ for $n = 1, 2, 3, \ldots$; hence $X_n \sim N(0,1)$. $X_n$ has the same distribution function as $X$ for all $n$ so, trivially, $\lim_n F_n(x) = F(x)$ for all $x$. Therefore, $X_n \xrightarrow{d} X$. But $P(|X_n - X| > \epsilon) = P(2|X| > \epsilon) = P(|X| > \epsilon/2) \neq 0$. So $X_n$ does not converge to $X$ in probability. ■

The relationships between the types of convergence can be summarized as follows:

<table>
<thead>
<tr>
<th>q.m.</th>
<th>↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.s.</td>
<td>→ prob</td>
</tr>
</tbody>
</table>

**Example 6** One might conjecture that if $X_n \xrightarrow{p} b$, then $E(X_n) \to b$. This is not true. Let $X_n$ be a random variable defined by $P(X_n = n^2) = 1/n$ and $P(X_n = 0) = 1 - (1/n)$. Now, $P(|X_n| < \epsilon) = P(X_n = 0) = 1 - (1/n) \to 1$. Hence, $X_n \xrightarrow{p} 0$. However, $E(X_n) = [n^2 \times (1/n)] + [0 \times (1 - (1/n))] = n$. Thus, $E(X_n) \to \infty$.

**Example 7** Let $X_1, \ldots, X_n \sim \text{Uniform}(0,1)$. Let $X_{(n)} = \max_i X_i$. First we claim that $X_{(n)} \xrightarrow{p} 1$. This follows since

$$P(|X_{(n)} - 1| > \epsilon) = P(X_{(n)} \leq 1 - \epsilon) = \prod_i P(X_i \leq 1 - \epsilon) = (1 - \epsilon)^n \to 0.$$  

Also

$$P(n(1 - X_{(n)}) \leq t) = P(X_{(n)} \geq 1 - (t/n)) = 1 - (1 - t/n)^n \to 1 - e^{-t}.$$  

So $n(1 - X_{(n)}) \sim \text{Exp}(1)$.

Some convergence properties are preserved under transformations.

**Theorem 8** Let $X_n, X, Y_n, Y$ be random variables.

(a) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $X_n + Y_n \xrightarrow{p} X + Y$.

(b) If $X_n \xrightarrow{qm} X$ and $Y_n \xrightarrow{qm} Y$, then $X_n + Y_n \xrightarrow{qm} X + Y$.

(c) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $X_nY_n \xrightarrow{p} XY$.  


In general, \( X_n \Rightarrow X \) and \( Y_n \Rightarrow Y \) does not imply that \( X_n + Y_n \Rightarrow X + Y \). But there are cases when it does:

**Theorem 9 (Slutzky’s Theorem)** If \( X_n \Rightarrow X \) and \( Y_n \Rightarrow c \), then \( X_n + Y_n \Rightarrow X + c \). Also, if \( X_n \Rightarrow X \) and \( Y_n \Rightarrow c \), then \( X_n Y_n \Rightarrow cX \).

**Theorem 10 (The Continuous Mapping Theorem)** Let \( X_n, X, Y_n, Y \) be random variables. Let \( g \) be a continuous function.

- (a) If \( X_n \overset{P}{\to} X \), then \( g(X_n) \overset{P}{\to} g(X) \).
- (b) If \( X_n \Rightarrow X \), then \( g(X_n) \Rightarrow g(X) \).

**Exercise:** Prove the continuous mapping theorem.

### 3 The Law of Large Numbers

The law of large numbers (LLN) says that the mean of a large sample is close to the mean of the distribution. For example, the proportion of heads of a large number of tosses of a fair coin is expected to be close to 1/2. We now make this more precise.

Let \( X_1, X_2, \ldots \) be an iid sample, let \( \mu = \mathbb{E}(X_1) \) and \( \sigma^2 = \text{Var}(X_1) \). Recall that the sample mean is defined as \( \overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i \) and that \( \mathbb{E}(\overline{X}_n) = \mu \) and \( \text{Var}(\overline{X}_n) = \sigma^2 / n \).

**Theorem 11 (The Weak Law of Large Numbers (WLLN))** If \( X_1, \ldots, X_n \) are iid, then \( \overline{X}_n \overset{P}{\to} \mu \). Thus, \( \overline{X}_n - \mu = o_P(1) \).

**Interpretation of the WLLN:** The distribution of \( \overline{X}_n \) becomes more concentrated around \( \mu \) as \( n \) gets large.

**Proof.** Assume that \( \sigma < \infty \). This is not necessary but it simplifies the proof. Using Chebyshev’s inequality,

\[
P \left( \left| \overline{X}_n - \mu \right| > \epsilon \right) \leq \frac{\text{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}
\]

which tends to 0 as \( n \to \infty \).

**Theorem 12 The Strong Law of Large Numbers.** Let \( X_1, \ldots, X_n \) be iid with mean \( \mu \). Then \( \overline{X}_n \overset{a.s.}{\to} \mu \).

The proof is beyond the scope of this course.
4 The Central Limit Theorem

The law of large numbers says that the distribution of $\bar{X}_n$ piles up near $\mu$. This isn’t enough to help us approximate probability statements about $\bar{X}_n$. For this we need the central limit theorem.

Suppose that $X_1, \ldots, X_n$ are iid with mean $\mu$ and variance $\sigma^2$. The central limit theorem (CLT) says that $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ has a distribution which is approximately Normal with mean $\mu$ and variance $\sigma^2/n$. This is remarkable since nothing is assumed about the distribution of $X_i$, except the existence of the mean and variance.

**Theorem 13 (The Central Limit Theorem (CLT))** Let $X_1, \ldots, X_n$ be iid with mean $\mu$ and variance $\sigma^2$. Let $Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\text{Var}(\bar{X}_n)}} \sim N(0, 1)$. In other words,

$$\lim_{n \to \infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$ 

Interpretation: Probability statements about $\bar{X}_n$ can be approximated using a Normal distribution. It’s the probability statements that we are approximating, not the random variable itself.

**Remark:** We often write $\bar{X}_n \approx N\left( \mu, \frac{\sigma^2}{n} \right)$ as short form for $\sqrt{n}(\bar{X}_n - \mu) \sim N(0, 1)$.

Recall that if $X$ is a random variable, its moment generating function (mgf) is $\psi_X(t) = \mathbb{E}e^{tX}$. Assume in what follows that the mgf is finite in a neighborhood around $t = 0$.

**Lemma 14** Let $Z_1, Z_2, \ldots$ be a sequence of random variables. Let $\psi_n$ be the mgf of $Z_n$. Let $Z$ be another random variable and denote its mgf by $\psi$. If $\psi_n(t) \to \psi(t)$ for all $t$ in some open interval around 0, then $Z_n \sim Z$.

**Proof of the central limit theorem.** Let $Y_i = (X_i - \mu)/\sigma$. Then, $Z_n = n^{-1/2} \sum_i Y_i$. Let $\psi(t)$ be the mgf of $Y_i$. The mgf of $\sum_i Y_i$ is $(\psi(t))^n$ and mgf of $Z_n$ is $[\psi(t/\sqrt{n})]^n \equiv \xi_n(t)$. Now $\psi'(0) = \mathbb{E}(Y_1^2) = 0$, $\psi''(0) = \mathbb{E}(Y_1^4) = \text{Var}(Y_1) = 1$. So,

$$\psi(t) = \psi(0) + t\psi'(0) + \frac{t^2}{2!}\psi''(0) + \frac{t^3}{3!}\psi'''(0) + \cdots$$

$$= 1 + 0 + \frac{t^2}{2} + \frac{t^3}{3!}\psi''(0) + \cdots$$

$$= 1 + \frac{t^2}{2} + \frac{t^3}{3!}\psi''(0) + \cdots$$
Now,

\[ \xi_n(t) = \left[ \psi \left( \frac{t}{\sqrt{n}} \right) \right]^n \]
\[ = \left[ 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{3/2}} \psi'''(0) + \cdots \right]^n \]
\[ = \left[ 1 + \frac{t^2}{2} + \frac{t^3}{3!n^{3/2}} \psi'''(0) + \cdots \right]^n \]
\[ \to e^{t^2/2} \]

which is the mgf of a N(0,1). The result follows from Lemma 14. In the last step we used the fact that if \( a_n \to a \) then

\[ \left( 1 + \frac{a_n}{n} \right)^n \to e^a. \]

The central limit theorem tells us that \( Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma \) is approximately N(0,1). However, we rarely know \( \sigma \). We can estimate \( \sigma^2 \) from \( X_1, \ldots, X_n \) by

\[ S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \]

This raises the following question: if we replace \( \sigma \) with \( S_n \), is the central limit theorem still true? The answer is yes.

**Theorem 15** Assume the same conditions as the CLT. Then,

\[ T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \Rightarrow N(0,1). \]

**Proof.** Here is a brief proof. We have that

\[ T_n = Z_n W_n \]

where

\[ Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \]

and

\[ W_n = \frac{\sigma}{S_n}. \]

Now \( Z_n \Rightarrow N(0,1) \) and \( W_n \to 1 \). The result follows from Slutzky’s theorem. ■
Here is an extended proof.

**Step 1.** We first show that $R_n^2 \xrightarrow{P} \sigma^2$ where $$R_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - X_n)^2.$$ Note that $$R_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2.$$ Define $Y_i = X_i^2$. Then, using the LLN (law of large numbers) $$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{P} \frac{1}{n} \sum_{i=1}^{n} Y_i = \mathbb{E}(Y_i) = \mathbb{E}(X_i^2) = \mu^2 + \sigma^2.$$ Next, by the LLN, $$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$ Since $g(t) = t^2$ is continuous, the continuous mapping theorem implies that $$\left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 \xrightarrow{P} \mu^2.$$ Thus $$R_n^2 \xrightarrow{P} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.$$

**Step 2.** Note that $$S_n^2 = \left( \frac{n}{n-1} \right) R_n^2.$$ Since, $R_n^2 \xrightarrow{P} \sigma^2$ and $n/(n-1) \to 1$, we have that $S_n^2 \xrightarrow{P} \sigma^2$.

**Step 3.** Since $g(t) = \sqrt{t}$ is continuous, (for $t \geq 0$) the continuous mapping theorem implies that $S_n \xrightarrow{P} \sigma$.

**Step 4.** Since $g(t) = t/\sigma$ is continuous, the continuous mapping theorem implies that $S_n/\sigma \xrightarrow{P} 1$.

**Step 5.** Since $g(t) = 1/t$ is continuous (for $t > 0$) the continuous mapping theorem implies that $\sigma/S_n \xrightarrow{P} 1$. Since convergence in probability implies convergence in distribution, $\sigma/S_n \xrightarrow{D} 1$. 

9
Step 5. Note that

\[ T_n = \left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \right) \left( \frac{\sigma}{S_n} \right) = V_n W_n. \]

Now \( V_n \to Z \) where \( Z \sim N(0, 1) \) by the CLT. And we showed that \( W_n \to 1 \). By Slutzky’s theorem, \( T_n = V_n W_n \to Z \times 1 = Z \).

The next result is very important. It tells us how close the distribution of \( \bar{X} \) is to the Normal distribution.

**Theorem 16 (Berry-Esseen Theorem)** Let \( X_1, \ldots, X_n \sim P \). Let \( \mu = \mathbb{E}[X_i] \) and \( \sigma^2 = \text{Var}[X_i] \). Assume that \( \mu_3 = \mathbb{E}[|X_i - \mu|^3] < \infty \). Let

\[ F_n(z) = \mathbb{P} \left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z \right). \]

Then

\[ \sup_z |F_n(z) - \Phi(z)| \leq \frac{33}{4} \frac{\mu_3}{\sigma^3 \sqrt{n}}. \]

There is also a multivariate version of the central limit theorem. Recall that \( X = (X_1, \ldots, X_k)^T \) has a multivariate Normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \) if

\[ f(x) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right). \]

In this case we write \( X \sim N(\mu, \Sigma) \).

**Theorem 17 (Multivariate central limit theorem)** Let \( X_1, \ldots, X_n \) be iid random vectors where \( X_i = (X_{1i}, \ldots, X_{ki})^T \) with mean \( \mu = (\mu_1, \ldots, \mu_k)^T \) and covariance matrix \( \Sigma \). Let \( \bar{X} = (\bar{X}_1, \ldots, \bar{X}_k)^T \) where \( \bar{X}_j = n^{-1} \sum_{i=1}^n X_{ji} \). Then,

\[ \sqrt{n}(\bar{X} - \mu) \to N(0, \Sigma). \]

**Remark:** There is also a multivariate version of the Berry-Esseen theorem but it is more complicated than the one-dimensional version.

5 The Delta Method

If \( Y_n \) has a limiting Normal distribution then the delta method allows us to find the limiting distribution of \( g(Y_n) \) where \( g \) is any smooth function.
Theorem 18 (The Delta Method) Suppose that
\[ \frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0, 1) \]
and that \( g \) is a differentiable function such that \( g'(\mu) \neq 0 \). Then
\[ \frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\sigma} \rightsquigarrow N(0, 1). \]
In other words,
\[ Y_n \approx N\left( \mu, \frac{\sigma^2}{n} \right) \text{ implies that } g(Y_n) \approx N\left( g(\mu), \frac{(g'(\mu))^2\sigma^2}{n} \right). \]

Example 19 Let \( X_1, \ldots, X_n \) be iid with finite mean \( \mu \) and finite variance \( \sigma^2 \). By the central limit theorem, \( \frac{\sqrt{n}(X_n - \mu)}{\sigma} \rightsquigarrow N(0, 1) \). Let \( W_n = e^{X_n} \). Thus, \( W_n = g(X_n) \) where \( g(s) = e^s \). Since \( g'(s) = e^s \), the delta method implies that \( W_n \approx N(e^\mu, e^{2\mu}\sigma^2/n) \).

There is also a multivariate version of the delta method.

Theorem 20 (The Multivariate Delta Method) Suppose that \( Y_n = (Y_{n1}, \ldots, Y_{nk}) \) is a sequence of random vectors such that
\[ \sqrt{n}(Y_n - \mu) \rightsquigarrow N(0, \Sigma). \]
Let \( g : \mathbb{R}^k \rightarrow \mathbb{R} \) and let
\[ \nabla g(y) = \left( \frac{\partial g}{\partial y_1}, \ldots, \frac{\partial g}{\partial y_k} \right). \]
Let \( \nabla \mu \) denote \( \nabla g(y) \) evaluated at \( y = \mu \) and assume that the elements of \( \nabla \mu \) are nonzero. Then
\[ \sqrt{n}(g(Y_n) - g(\mu)) \rightsquigarrow N\left( 0, \nabla^T \mu \Sigma \nabla \mu \right). \]

Example 21 Let
\[ \left( \begin{array}{c} X_{11} \\ X_{21} \end{array} \right), \left( \begin{array}{c} X_{12} \\ X_{22} \end{array} \right), \ldots, \left( \begin{array}{c} X_{1n} \\ X_{2n} \end{array} \right) \]
be iid random vectors with mean \( \mu = (\mu_1, \mu_2)^T \) and variance \( \Sigma \). Let
\[ \bar{X}_1 = \frac{1}{n} \sum_{i=1}^{n} X_{1i}, \quad \bar{X}_2 = \frac{1}{n} \sum_{i=1}^{n} X_{2i} \]
and define \( Y_n = \bar{X}_1 \bar{X}_2 \). Thus, \( Y_n = g(\bar{X}_1, \bar{X}_2) \) where \( g(s_1, s_2) = s_1 s_2 \). By the central limit theorem,
\[ \sqrt{n} \left( \frac{\bar{X}_1 - \mu_1}{\bar{X}_2 - \mu_2} \right) \rightsquigarrow N(0, \Sigma). \]
Now
\[ \nabla g(s) = \left( \frac{\partial g}{\partial s_1}, \frac{\partial g}{\partial s_2} \right) = \left( \begin{array}{c} s_2 \\ s_1 \end{array} \right) \]

and so
\[ \nabla^T \Sigma \nabla = (\mu_2, \mu_1) \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix} = \mu_2^2 \sigma_{11} + 2 \mu_1 \mu_2 \sigma_{12} + \mu_1^2 \sigma_{22}. \]

Therefore,
\[ \sqrt{n}(X_1 X_2 - \mu_1 \mu_2) \rightsquigarrow N \left( 0, \mu_2^2 \sigma_{11} + 2 \mu_1 \mu_2 \sigma_{12} + \mu_1^2 \sigma_{22} \right). \]