7. Convergence in Distribution

Instructor: Alessandro Rinaldo

Associated reading: Sec 2.8, 7.2 of Ash and Doléans-Dade; Sec 3.1, 3.2 of Durrett.

Overview

Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d random variables with $\mathrm{E}X_1 = 0$, $\mathrm{Var}X_1 = \sigma^2 < \infty$. In the last set of lecture notes we have shown that

$$\frac{S_n}{\sigma n^{\frac{1}{2}+\epsilon}} \xrightarrow{\text{a.s.}} 0.$$

In the Law of Iterated Logarithm we have seen that

$$\limsup_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = 1 \text{ a.s.}$$

$$\liminf_{n\to\infty} \frac{S_n}{\sigma\sqrt{2n\log(\log n)}} = -1 \text{ a.s.}$$

We shall show that $\frac{S_n}{\sigma n^{\frac{1}{2}}}$ also converges, in a different notion, to a limit. In this set of lecture notes, we introduce the notion of convergence in distribution.

1 Convergence in Distribution

Let $\mathcal{X}$ be a topological space and let $\mathcal{B}$ be the Borel σ-field. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_n : \Omega \to \mathcal{X}$ be $\mathcal{F}/\mathcal{B}$-measurable. Also, let $X : \Omega \to \mathcal{X}$ be another random quantity. Let P_n and P_X be the distribution of X_n and X, respectively. This will be the standard setup for all discussions of convergence in distribution.

Definition 1 (Convergence in Distribution). *We say that X_n* converges in distribution *to X if*

$$\lim_{n\to\infty} \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)],$$

for all bounded continuous functions $f : \mathcal{X} \to \mathbb{R}$. We denote this property $X_n \xrightarrow{\mathcal{D}} X$.

Example 2. *Let $\Omega = I\!R^\infty$ with $\mathcal{F} = \mathcal{B}^\infty$ and P being the joint distribution of a sequence of iid standard normal random variables. Let $X_n(\omega) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \omega_j$. Let $X = X_1$. Then $X_n \stackrel{\mathcal{D}}{\to} X$ in a trivial way.*

There are several conditions that are all equivalent to $X_n \stackrel{\mathcal{D}}{\to} X$.

Theorem 3 (Portmanteau theorem). *The following are all equivalent if $\mathcal{X}$ is a metric space:*

1. $\lim_{n\to\infty} \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)]$, *for all bounded continuous f,*
2. *For each closed $C \subseteq \mathcal{X}$, $\limsup_{n\to\infty} P(X_n \in C) \le P(X \in C)$.*
3. *For each open $A \subseteq \mathcal{X}$, $\liminf_{n\to\infty} P(X_n \in A) \ge P(X \in A)$.*
4. *For each $B \in \mathcal{B}$ such that $P(X \in \partial B) = 0$, $\lim_{n\to\infty} P(X_n \in B) = P(X \in B)$.*

We will not prove this whole theorem, but we will look a bit more at the four conditions. If $\mathcal{X} = I\!R$, then the fourth condition is a lot like the familiar convergence of cdf's in places where the limit is continuous. An interval $B = (-\infty, b]$ has $P_X(\partial B) = 0$ if and only if there is no mass at b, hence if and only if the cdf is continuous at b. The second condition says that we don't want any mass from the distributions of the X_n's to be able to escape from a closed set, although it could happen that mass from outside of a closed set approaches the boundary. That is why the inequality goes the way it does. Similarly, for the third condition, mass can escape from an open set but nothing should be allowed to "jump" into the open set. The first condition is related to the often overlooked fact that the distribution of a random quantity is equivalent to the means of all bounded continuous functions. The first condition is also a version of what mathematicians call *weak* convergence*, a concept that arises in the theory of normed linear spaces. Many statisticians and probabilists call convergence in distribution "weak convergence," but convergence in distribution is not quite the same as weak convergence in normed linear spaces.

Proof: First, notice that the second and third conditions are equivalent since closed sets are complements of open sets. Together the second and third conditions imply the fourth one. We will prove that the (4) $\Rightarrow$ (1) and that (1) $\Rightarrow$ (2).

(4) $\Rightarrow$ (1).

Assume the fourth condition. Let f be bounded and continuous, $|f(x)| \le K$ for all x. Let $\epsilon > 0$. Let $v_0 < v_1 < \cdots < v_M$ be real numbers such that $v_0 < -K < K < v_M$, $v_j - v_{j-1} < \epsilon$ for all $j = 1, \ldots, M$, and $P_X(\{x : f(x) = v_j\}) = 0$ for all j. Let $F_j = \{x : v_{j-1} < f(x) \le v_j\}$. The continuity of f and the fact that $\partial F_j \subseteq \{x : f(x) \in \{v_j, v_{j-1}\}\}$ imply that

$$\{x : v_{j-1} < f(x) < v_j\} \subseteq \mathrm{int}(F_j) \subseteq \overline{F}_j \subseteq \{x : v_{j-1} \le f(x) \le v_j\}.$$

By construction

$$\begin{aligned}\left|\sum_{j=1}^{M} v_j P_n(F_j) - \mathrm{E}[f(X_n)]\right| &\leq \epsilon, \\ \left|\sum_{j=1}^{M} v_j P_X(F_j) - \mathrm{E}[f(X)]\right| &\leq \epsilon.\end{aligned}$$

By assumption $P_X(\partial F_j) = 0$ for all j and

$$\lim_{n\to\infty} \sum_{j=1}^{M} v_j P_n(F_j) = \sum_{j=1}^{M} v_j P_X(F_j).$$

Combining these yields $|\lim_{n\to\infty} \mathrm{E}[f(X_n)] - \mathrm{E}[f(X)]| < 2\epsilon$, hence the first condition holds.

(1) $\Rightarrow$ (2).

Let C be a closed set. For each m, let C_m be the set of points that are at most $1/m$ away from C. The function $f_m(x) = \max\{0, 1 - md(x, C)\}$ is bounded and continuous, equals 0 on C_m^C, equals 1 on C, and lies between 0 and 1 everywhere. We know that $\lim_{n\to\infty} \mathrm{E}(f_m(X_n)) = \mathrm{E}(f_m(X))$ for all m. Also, $P_n(C) \leq \mathrm{E}(f_m(X_n)) \leq P_n(C_m)$ for all n and m. So

$$\limsup_{n\to\infty} P_n(C) \leq \mathrm{E}(f_m(X)) \leq P_X(C_m), \tag{1}$$

for all m. Since $\{C_m\}_{m=1}^{\infty}$ is a decreasing sequence of sets whose intersection is C, we have $\lim_{m\to\infty} P_X(C_m) = P_X(C)$. Since the left side of Equation (1) doesn't depend on m, we have the result. ∎

Because convergence in distribution depends only on the distributions of the random quantities involved, we do not actually need random quantities in order to discuss convergence in distribution. Hence, we might also use notation like $\mu_n \xrightarrow{\mathcal{D}} \mu$, where μ_n and μ are probability measures on the same space. If $\mathcal{X} = \mathbb{R}$, we might refer to the cdf's and say $F_n \xrightarrow{\mathcal{D}} F$. We might even refer to the names of distributions and say that X_n converges in distribution to a standard normal distribution or some other distribution. Even if we do have random quantities, they don't even have to be defined on the same probability spaces. They do have to take values in the same space, however. For example, for each n, let $(\Omega_n, \mathcal{F}_n, P_n)$ be a probability space, and let $(\Omega, \mathcal{F}, P)$ be another one. Let $(\mathcal{X}, \mathcal{B})$ be a topological space with Borel σ-field. Let $X_n : \Omega_n \to \mathcal{X}$ and $X : \Omega \to \mathcal{X}$ be random quantities. We could then ask whether or not $X_n \xrightarrow{\mathcal{D}} X$. We won't use this last bit of added generality.

Example 4. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid standard normal random variables. Then X_n converges in distribution to standard normal, but does not converge in probability to anything.*

Some authors use the expression *converges in law* to mean "converges in distribution". They might write this $X_n \stackrel{\mathcal{L}}{\to} X$. Others use the expression *converges weakly* and might write it $X_n \stackrel{w}{\to} X$.[1]

Convergence in distribution is weaker than convergence in probability, hence it is also weaker than convergence a.s. and L^p convergence.

Proposition 5 (Relationships between convergences). *Let $(\mathcal{X}, \mathcal{B})$ be a metric space (having metric d) and its Borel σ-field. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random quantities taking values in $\mathcal{X}$ and let X be another random quantity taking values in $\mathcal{X}$.*

1. *If $\lim_{n\to\infty} X_n = X$ a.s., then $X_n \stackrel{P}{\to} X$.*
2. *If $X_n \stackrel{P}{\to} X$, then $X_n \stackrel{\mathcal{D}}{\to} X$.*
3. *If X is degenerate and $X_n \stackrel{\mathcal{D}}{\to} X$, then $X_n \stackrel{P}{\to} X$.*
4. *If $X_n \stackrel{P}{\to} X$, then there is a subsequence $\{n_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} X_{n_k} = X$, a.s.*

Proof: The first and last claims were proven earlier and are only included for completeness. For the second claim, let C be a closed set and let $C_m = \{x : d(x, C) \le 1/m\}$ for each integer $m > 0$. Then

$$\mu_{X_n}(C) \le \mu_X(C_m) + \Pr(d(X, X_n) > 1/m).$$

It follows that $\limsup_n \mu_{X_n}(C) \le \mu_X(C_m)$. Since $\lim_{m\to\infty} \mu_X(C_m) = \mu_X(C)$, we have that $X_n \stackrel{\mathcal{D}}{\to} X$ by Theorem 3. The third claim follows by approximating $I_{[c-\epsilon, c+\epsilon]}$ by a bounded continuous function, where $\Pr(X = c) = 1$. ■

2 The special case of $(\mathbb{R}^1, \mathcal{B}^1)$

Now suppose $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^1, \mathcal{B}^1)$. Consider part 4 of Theorem 3 with $B = (-\infty, x]$. The condition $P_X(\partial B) = 0$ is equivalent to $P_X(\{x\}) = 0$ and hence equivalent to that F is continuous at x. According to part 4 of Theorem 3, a necessary condition for $F_n \stackrel{\mathcal{D}}{\to} F$ is that $F_n(x) \to F(x)$ at all continuity points x of F. We will show that this is also sufficient.

The argument below uses some basic useful properties of distribution functions and their inverses. For a distribution G on $\mathbb{R}$, define the inverse of G as

$$G^{-1}(p) = \inf\{x : G(x) \ge p\}\,,\ \forall\, p \in (0,1)\,.$$

Then we have the following basic facts about G^{-1}.

[1] Convergence in distribution is not the same as weak convergence of continuous linear functionals in functional analysis. It is the same as *weak* convergence*, but we will not go into that distinction here.

Lemma 6 (Basic facts about G^{-1}). *For a distribution function G, the following are true.*

1. *G^{-1} is non-decreasing, and hence G^{-1} has at most countably many discontinuous points.*
2. *$G^{-1}(p) \le x \Leftrightarrow G(x) \ge p$, or equivalently $G^{-1}(p) > x \Leftrightarrow G(x) < p$.*
3. *$G(G^{-1}(p)) \ge p$.*
4. *$G^{-1}(G(x)) \le x$.*
5. *G^{-1} is left-continuous.*

The proof of sufficiency relies on the following theorem of Skorohod.

Lemma 7 (Skorohod). *Let $\{F_n : n \ge 1\}$, F be distribution functions on $\mathbb{R}$ such that $F_n(x) \to F(x)$ for all x at which F is continuous. Then there exist random variables $\{Y_n : n \ge 1\}$, Y defined on $((0,1), \mathcal{B}^1, \lambda)$ (λ being Lebesgue measure) such that Y_n has distribution function F_n for all n, Y has distribution function F, and $Y_n \overset{a.s.}{\to} Y$.*

Proof: Define $Y_n(\omega) = F_n^{-1}(\omega)$ and $Y(\omega) = F^{-1}(\omega)$. It is easy to see that Y_n has distribution function F_n and Y has distribution function F. For example,

$$\Pr(Y \le y) = \Pr(F^{-1}(\omega) \le y) = \Pr(\omega \le F(y)) = F(y).$$

To see that $Y_n(\omega) \to Y(\omega)$, let $\epsilon > 0$ and let $Y(\omega) - \epsilon < x < Y(\omega)$ be such that F is continuous at x. Then $F(x) < \omega$, so eventually $F_n(x) < \omega$ and eventually $Y(\omega) - \epsilon < x < Y_n(\omega)$, so $\liminf_n Y_n(\omega) \ge Y(\omega)$.

For the other direction, let ω be a continuity point of Y. For any $\omega' > \omega$ and $\epsilon > 0$ choose $Y(\omega') < x < Y(\omega') + \epsilon$ with x an continuity point of F. Then $F_n(x) \to F(x) \ge F(Y(\omega')) \ge \omega' > \omega$. As a result, for n large enough we have $Y_n(\omega) \le x \le Y(\omega') + \epsilon$. Thus $\limsup_n Y_n(\omega) \le Y(\omega') + \epsilon$. Let $\omega' \downarrow \omega$ and $\epsilon \downarrow 0$ we have $\limsup_n Y_n(\omega) \le Y(\omega)$.

Now we have shown that $Y_n(\omega) \to Y(\omega)$ for all ω at which $Y(\omega)$ is continuous. The desired result follows because the set of ω at which Y is discontinuous is at most countable. ∎

The following result says that the usual definition of convergence in distribution in one dimension is equivalent to what we have stated above.

Lemma 8 (Portmanteau in $\mathbb{R}^1$). *Let $(\mathcal{X}, \mathcal{B}) = (I\!R, \mathcal{B}^1)$. Let F_n be the cdf of X_n and let F be the cdf of X. Then $X_n \overset{\mathcal{D}}{\to} X$ if and only if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous.*

Proof: The proof of the "only if" direction is direct from Theorem 3 because F is continuous at x if and only if $\mu_X(\{x\}) = 0$ and $\{x\}$ is the boundary of $(-\infty, x]$. For the "if" part, construct Y_n and Y as in the proof of Lemma 7. It then follows from the dominated convergence theorem that $\mathrm{E}(f(Y_n)) \to \mathrm{E}(f(Y))$ for all bounded continuous f. ■

Example 9. *Let Φ be the standard normal cdf, and let*

$$F_n(x) = \begin{cases} 0 & \text{if } x < -n, \\ \frac{\Phi(x)-\Phi(-n)}{\Phi(n)-\Phi(-n)} & \text{if } -n \le x < n, \\ 1 & \text{if } x \ge n. \end{cases}$$

Then, we see that $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for all x. Each F_n gives probability 1 to a bounded set, but the limit distribution does not.

Example 10. *Let Φ be the standard normal cdf, and let*

$$F_n(x) = \begin{cases} 0 & \text{if } x < -n, \\ \Phi(x) & \text{if } -n \le x < n, \\ 1 & \text{if } x \ge n. \end{cases}$$

Then, we see that $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for all x. Each F_n is neither discrete nor continuous, but the limit is continuous.

Example 11. *Enumerate the dyadic rationals in this sequence: 1/2, 1/4, 3/4, 1/8, 3/8, 5/8, 7/8, 1/16, 3/16, ... Let μ_n be the measure that puts mass $1/n$ on each of the first n in the list. Then the subsequence $\{\mu_{2^n-1}\}_{n=1}^{\infty}$ converges in distribution to the uniform distribution on $[0,1]$, but the whole sequence does not converge. Consider the subsequence $\{\mu_{2^{n+2}-2^n-1}\}_{n=1}^{\infty}$, which converges to a distribution with twice as much probability on $[0,1/2]$ as on $(1/2,1]$.*

Example 12. *Let F_n be the cdf of the uniform distribution on $[-n,n]$. No subsequence of F_n converges in distribution even though each cdf gives probability 1 to a bounded set.*

Examples 9 and 12 illustrate a necessary and sufficient condition for a sequence of distributions to have a convergent (in distribution) subsequence. Even though the F_n in both examples assign probability to 1 to the same intervals, the probability moves out to infinity at different rates in the two examples. In Definition 22, we will see a condition on how fast probability can move out to infinity and still allow subsequences to converge in distribution.

3 Continuous Mapping

If f is a continuous function and $X_n \xrightarrow{\mathcal{D}} X$, then $f(X_n) \xrightarrow{\mathcal{D}} f(X)$. Indeed, even if f is not continuous, so long as μ_X assigns 0 probability to the set of discontinuities, the result still holds.

Theorem 13 (Continuous mapping, 1). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random quantities, and let X be another random quantity all taking values in the same metric space $\mathcal{X}$. Suppose that $X_n \xrightarrow{\mathcal{D}} X$. Let $\mathcal{Y}$ be a metric space and let $g : \mathcal{X} \to \mathcal{Y}$. Define*

$$C_g = \{x : \ g \text{ is continuous at } x\}.$$

Suppose that $\Pr(X \in C_g) = 1$. *Then $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.*

The proof of Theorem 13 together with the proof of Theorem 15 both rely on the second part of Theorem 3, and they resemble the part of the proof of Proposition 5 that we already did.

Proof: Let Q_n be the distribution of $g(X_n)$ and let Q be the distribution of $g(X)$. Let R_n be the distribution of X_n and let R be the distribution of X. Let B be a closed subset of $\mathcal{Y}$. If $x \in \overline{g^{-1}(B)}$ but $x \notin g^{-1}(B)$, then g is not continuous at x because otherwise there exists a sequence $\{x_n : n \geq 1\} \subseteq g^{-1}(B)$ such that $g(x_n) \to g(x) \in B$ since B is closed and all $g(x_n) \in B$. It follows that $\overline{g^{-1}(B)} \subseteq g^{-1}(B) \cup C_g^C$. Now write

$$\begin{aligned}
\limsup_{n\to\infty} Q_n(B) &= \limsup_{n\to\infty} R_n(g^{-1}(B)) \leq \limsup_{n\to\infty} R_n(\overline{g^{-1}(B)}) \\
&\leq R(\overline{g^{-1}(B)}) \leq R(g^{-1}(B)) + R(C_g^C) \\
&= R(g^{-1}(B)) = Q(B),
\end{aligned}$$

and the result now follows from the Theorem 3. ■

Example 14. *If $(S_n - n\mu)/[\sqrt{n}\sigma]$ converges in distribution to standard normal, then $(S_n - n\mu)^2/(n\sigma^2)$ converges in distribution to χ^2 with one degree of freedom.*

Theorem 15 (Continuous mapping, 2). *Let $\{X_n\}_{n=1}^{\infty}$, X, and $\{Y_n\}_{n=1}^{\infty}$ be random quantities taking values in a metric space with metric d. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and $d(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.*

Proof: Let Q_n be the distribution of Y_n, let R_n be the distribution of X_n and let R be the distribution of X. Let B be an arbitrary closed set. According to Theorem 3, it suffices to show that $\limsup Q_n(B) \leq R(B)$. Then

$$\{Y_n \in B\} \subseteq \{d(X_n, B) \leq \epsilon\} \cup \{d(X_n, Y_n) > \epsilon\}.$$

Define $C_\epsilon = \{x : d(x, B) \leq \epsilon\}$, which is a closed set. So,

$$\begin{aligned}
Q_n(B) &= P_n(Y_n \in B) \\
&\leq P_n(d(X_n, B) \leq \epsilon) + P_n(d(X_n, Y_n) > \epsilon) \\
&= R_n(C_\epsilon) + P_n(d(X_n, Y_n) > \epsilon).
\end{aligned}$$

We have assumed that $\lim_{n\to\infty} P_n(d(X_n, Y_n) > \epsilon) = 0$ and that $X_n \xrightarrow{\mathcal{D}} X$, so we conclude $\limsup_{n\to\infty} Q_n(B) \le \limsup_{n\to\infty} R_n(C_\epsilon) \le R(C_\epsilon)$. Since B is closed, $\lim_{\epsilon\to 0} R(C_\epsilon) = R(B)$. It follows then that

$$\limsup_{n\to\infty} Q_n(B) \le R(B),$$

hence $Y_n \xrightarrow{\mathcal{D}} X$. ■

The most common use of this theorem is the following. If the difference between two sequences converges to 0 in probability and if one of the two sequences converges in distribution to X, then so does the other one. A related result is the following.

Theorem 16 (Continuous mapping, 3). *Let X_n take values in a metric space and let Y_n take values in a metric space. Suppose that $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c$, then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$.*

Proof: Let d_1 be the metric in the space where X_n takes values and let d_2 be the metric in the space where Y_n takes values. Then

$$d((x_1, y_1), (x_2, y_2)) = d_1(x_1, x_2) + d_2(y_1, y_2),$$

defines a metric in the product space and the product σ-field is the Borel σ-field. First note that $(X_n, c) \xrightarrow{\mathcal{D}} (X, c)$ since every bounded continuous function of (X_n, c) is a bounded continuous function of X_n alone. Next, note that $d((X_n, Y_n), (X_n, c)) = d_2(Y_n, c)$ and $P_n(d_2(Y_n, c) > \epsilon) \to 0$ for all $\epsilon > 0$, so $d((X_n, Y_n), (X_n, c)) \xrightarrow{P} 0$. By Theorem 15, $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$. ■

A simple and useful consequence of the continuous mapping theorem is the Slutsky's Theorem.

Theorem 17 (Slutsky's Theorem). *If $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \xrightarrow{\mathcal{D}} c$ then*

1. $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$.
2. $X_n Y_n \xrightarrow{\mathcal{D}} cX$.
3. $X_n / Y_n \xrightarrow{\mathcal{D}} X/c$ *provided* $c \neq 0$.

Example 18. *Suppose that $U_n = (S_n - n\mu)/(\sqrt{n}\sigma)$ converges in distribution to standard normal. Suppose also, that $T_n \xrightarrow{P} \sigma$. Then $(U_n, T_n) \xrightarrow{\mathcal{D}} (Z, \sigma)$, where $Z \sim N(0, 1)$. Consider the continuous function $g(z, s) = z\sigma/s$. It follows that*

$$g(U_n, T_n) = \frac{S_n - n\mu}{\sqrt{n} T_n} \xrightarrow{\mathcal{D}} Z.$$

Example 19 (Delta method). *Suppose that* $\lim_{n\to\infty} r_n = \infty$ *and* $r_n(X_n - a) \xrightarrow{\mathcal{D}} Y$. *Then* $X_n \xrightarrow{P} a$. *Suppose that* g *is a function that has a derivative* $g'(a)$ *at* a. *Define*

$$h(x) = \frac{g(x) - g(a)}{x - a} - g'(a).$$

We know that $\lim_{x\to a} h(x) = 0$, *so we can make* h *continuous at* a *by setting* $h(a) = 0$. *Also* $g(x) - g(a) = (x - a)g'(a) + (x - a)h(x)$. *So,*

$$r_n[g(X_n) - g(a)] = r_n(X_n - a)g'(a) + r_n(X_n - a)h(X_n).$$

It follows from Theorems 13 and 5 that $h(X_n) \xrightarrow{P} 0$. *By Theorem 17,* $r_n(X_n - a)h(X_n) \xrightarrow{P} 0$ *and* $r_n(X_n - a)g'(a) \xrightarrow{\mathcal{D}} g'(a)Y$. *By Theorem 17,* $r_n[g(X_n) - g(a)] \xrightarrow{\mathcal{D}} g'(a)Y$. *After we see the central limit theorem, there will be many examples of the use of this result.*

If $g'(a) = 0$ in the above example, there may still be hope if a higher derivative is nonzero.

Example 20. *Let* $\{X_n\}_{n=1}^{\infty}$ *be iid with exponential distribution with parameter 2. That is, the density is* $2\exp(-2x)$ *for* $x > 0$. *Let* $Y_n = \min\{X_1, \ldots, X_n\}$. *Then* Y_n *has an exponential distribution with parameter* $2n$. *So* $n(Y_n - 0) \xrightarrow{\mathcal{D}} X_1$. *Let* $g(y) = \cos(y)$ *so that* $g'(y) = -\sin(y)$. *Then* $n[\cos(Y_n) - 1] \xrightarrow{\mathcal{D}} 0$. *But* $g(y) - 1 = 0 - y^2/2 + o(y^2)$ *as* $y \to 0$. *So,*

$$n^2[g(Y_n) - 1] = \frac{n^2}{2} Y_n^2 + Z_n \xrightarrow{\mathcal{D}} \frac{1}{2} X_1^2,$$

where $Z_n \xrightarrow{P} 0$.

The following result is the convergence in probability version of Theorem 13.

Theorem 21 (In-probability version of CMT). *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of random quantities, and let* X *be another random quantity all taking values in the same metric space* $\mathcal{X}$ *with metric* d_1. *Suppose that* $X_n \xrightarrow{P} X$. *Let* $\mathcal{Y}$ *be a metric space with metric* d_2 *and let* $g : \mathcal{X} \to \mathcal{Y}$. *Define*

$$C_g = \{x : \ g \text{ is continuous at } x\}.$$

Suppose that $\Pr(X \in C_g) = 1$. *Then* $g(X_n) \xrightarrow{P} g(X)$.

4 Tightness and Helly-Bray Selection

For the discussions of tight sequences and characteristic functions, we will take $\mathcal{X} = \mathbb{R}^p$ for some p with Borel σ-field $\mathcal{B}^p$.

Definition 22 (Tightness and Relative Sequential Compactness). *A collection* $\{P_\alpha : \alpha \in \aleph\}$ *of probability measures on* $(\mathbb{R}^p, \mathcal{B}^p)$ *is called* tight *if for every* $\epsilon > 0$*, there exists a compact set* A *such that* $P_\alpha(A) > 1 - \epsilon$ *for all* α*. The collection is* relatively sequentially compact *if every sequence* $\{P_{\alpha_n}\}_{n=1}^\infty$ *from the collection has a subsequence that converges in distribution.*

Examples 9 and 11 are tight sequences. Example 12 is not a tight sequence. It is fairly easy to see that relative sequential compactness implies tightness. Theorem 23 provides most of the proof that tightness implies relative sequential compactness.[2]

Theorem 23 (Helly-Bray Selection). *Let* $\{P_n\}_{n=1}^\infty$ *be a tight sequence of probability measures on* $(\mathbb{R}^p, \mathcal{B}^p)$*. Then there exists a subsequence that converges in distribution. Also, if every convergent subsequence converges in distribution to the same probability* P*. Then* $P_n \xrightarrow{\mathcal{D}} P$*.*

Proof: Each P_n is equivalent to a cdf F_n and P is equivalent to a cdf F. We will use the following notation: $x \le y$ will mean that $x_i \le y_i$ for $i = 1, \ldots, p$ for vectors $x, y \in \mathbb{R}^p$. Similarly, $x < y$ means $x_i < y_i$ for all i. For each rectangle A of the form $\{x : a_i < x_i \le b_i, \text{ for } i = 1, \ldots, p\}$ and function $H : \mathbb{R}^p \to \mathbb{R}$, define

$$H(A) = \sum_{\text{All corners } r} (-1)^{c(r)} H(r),$$

where $c(r) = 1$ if r has an odd number of a_i and $c(r) = 0$ if r has an even number of a_i. When H is the cdf of X, $H(A) = \Pr(X \in A)$.

The proof is in six parts:

1. Show that every subsequence of $\{F_n\}_{n=1}^\infty$ has a further subsequence F_{n_k} that converges at points with rational coordinates to some function G^* that is nondecreasing.
2. Show that G^* can be modified to a function G defined on $\mathbb{R}^p$ that is non-decreasing and continuous from above.
3. Show that F_{n_k} converges to G at continuity points of G.
4. Show that $G(A) \ge 0$ for each rectangle A.
5. Show that G is a cdf (using tightness).
6. If every convergent subsequence converges to the same G then show that $G = F$, and the whole sequence converges to F.

[2] In more general spaces than $\mathbb{R}^p$, the concepts of tightness and relative sequential compactness are not equivalent.

First, let $\{H_n\}_{n=1}^{\infty}$ be a subsequence of $\{F_n\}_{n=1}^{\infty}$. Let $\{q_n\}_{n=1}^{\infty}$ be an enummeration of the points with all rational coordinates. Let $\{H_{n_k^1}\}_{k=1}^{\infty}$ be a subsequence such that $\lim_{k\to\infty} H_{n_k^1}(q_1) = G^*(q_1)$ exists. For $i > 1$, let $\{H_{n_k^i}\}_{k=1}^{\infty}$ be a subsequence of $\{H_{n_k^{i-1}}\}_{k=1}^{\infty}$ such that $H_{n_k^i}(q_i) = G^*(q_i)$ exists. Our final subsequence is $\{n_k\}_{k=1}^{\infty}$ with $n_k = n_k^k$. For this subsequence $\lim_{k\to\infty} H_{n_k}(q_i) = G^*(q_i)$ for all i. Because F_{n_k} are distribution function, it is straightforward to check that G^* is non-decreasing on points with all rational numbers.

Second, for all x, define $G(x) = \inf\{G^*(q_i) : q_i > x\}$. Then G is non-decreasing and continuous from above. To show that G is continuous from above, let $z_n \downarrow x$. Then $G(z_n) \downarrow b \geq G(x)$ for some b. If $b > G(x)$, then we can find a rational $q > x$ such that $G(x) \leq G^*(q) < b$. For n large enough, we have $x < z_n < q$, so $G(z_n) \leq G^*(q) < b$. Thus $\lim G(z_n) < b$, a contradiction.

Third, we show that $F_{n_k}(x) \to G(x)$ if G is continuous at x. For $\delta > 0$, let q have rational coordinates such that $x < q$ and $G^*(q) \leq G(x) + \delta$. Then $\limsup F_{n_k}(x) \leq \limsup F_{n_k}(q) = G^*(q) < G(x) + \delta$, hence $\limsup F_{n_k}(x) \leq G(x)$. On the other hand, for any $x' < x$, let q have rational coordinates such that $x' < q < x$. We have $G(x') \leq G^*(q) = \lim F_{n_k}(q) \leq \liminf F_{n_k}(x)$. Let $G(x^-) = \sup\{G(x') : x' < x\}$. Then $G(x^-) \leq \liminf F_{n_k}(x)$. Because G is continuous at x, $G(x) = G(x^-)$ and hence $\lim F_{n_k}(x) = G(x)$.

Fourth, let $A = \{x : a_i < x_i \leq b_i, \text{ for } i = 1, \ldots, p\}$ be a rectangle. We now show that $G(A) \geq 0$. Let $\epsilon > 0$. Let $y_1, \ldots, y_{2^p}$ be the corners of A. Let $q_i > a_i$ and $s_i > b_i$ be rational numbers and let $y'_1, \ldots, y'_{2^p}$ be the corresponding corners of $A' = \{x : q_i < x_i \leq s_i, \text{ for } i = 1, \ldots, p\}$. We can choose q_i, s_i close enough to a_i, b_i such that $G(y_r) \leq G^*(y_{r'}) < G(y_r) + \epsilon/2^{p+1}$ for all $r = 1, ..., 2^p$ and hence $|G(A) - G^*(A')| \leq \epsilon/2$. On the other hand, when k is large enough we have $|F_{n_k}(y_{r'}) - G^*(y_{r'})| \leq \epsilon/2^{p+1}$ for all $r = 1, ..., 2^p$. Thus, $|G(A) - F_{n_k}(A')| \leq |G(A) - G^*(A')| + |G^*(A') - F_{n_k}(A')| \leq \epsilon$. As a result, $G(A) \geq |F_{n_k}(A')| - \epsilon \geq -\epsilon$. Let $\epsilon \to 0$ we have $G(A) \geq 0$.

Fifth, let $\epsilon > 0$, and let $A = [-q, q]^p$ with a rational number q. For a rational number $s > q$ define $A' = [-s, s]^p$. We can chose s close enough to q such that $G(A) \geq G * (A') - \epsilon/2$. Also, when q is large enough, s is also large enough such that $F_{n_k}(A') \geq 1 - \epsilon/2$ for all k. Therefore, $G(A) \geq 1 - \epsilon$ and hence G is a probability measure in $\mathbb{R}^p$.

Sixth, suppose that there exists a continuity point of F such that $\lim_{n\to\infty} F_n(x) \neq F(x)$. So, there exists $\epsilon > 0$ and a sequence of integers $\{m_k\}_{k=1}^{\infty}$ such that $|F_{m_k}(x) - F(x)| > \epsilon$ for all k. Then no subsequence of $\{F_{m_k}(x)\}_{k=1}^{\infty}$ can converge to $F(x)$, which contradicts what we proved in the first five parts. ∎