11. Weak Convergence

Definition. A sequence of distribution functions $F_1, F_2, ...$ *converges weakly* to a distribution function F_∞ (written $F_n \Rightarrow F_\infty$) if $\lim_{n\to\infty} F_n(x) = F_\infty(x)$ for all x at which F is continuous.
Random variables $X_1, X_2, ...$ *converge weakly* (or *converge in distribution*) to a random variable X_∞ (written $X_n \Rightarrow X_\infty$) if $F_n \Rightarrow F_\infty$ where $F_n(x) = P(X_n \le x)$ for $1 \le n \le \infty$.

Note that since the definition of weak convergence of random variables depends only on their distribution functions, one can speak of a sequence $X_1, X_2, ...$ converging weakly even if the $X_n's$ are not defined on the same probability space. This is not the case with the other modes of convergence we have discussed.

Also, since distribution functions are right-continuous and have only countably many discontinuities, we see that F_∞ is uniquely determined by its values at continuity points.

Example 11.1. As a trivial example, suppose that X has distribution function F and let $\{a_n\}_{n=1}^\infty$ be any sequence of real numbers which decreases to 0. Then $X_n = X - a_n$ has distribution function $F_n(x) = P(X - a_n \le x) = F(x + a_n)$, hence $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$ (since F is right-continuous) and thus $X_n \Rightarrow X$.
On the other hand, $X_n = X + a_n$ has distribution function $F_n(x) = F(x - a_n)$, which converges to F only at continuity points. Thus we still have $X_n \Rightarrow X$, but the distribution functions do not necessarily converge pointwise.

For some more interesting examples, we first need some elementary facts:

Fact 11.1. *If $c_j \to 0$, $a_j \to \infty$, and $a_j c_j \to \lambda$, then $(1 + c_j)^{a_j} \to e^\lambda$.*

Proof. $\lim_{x\to 0} \frac{\log(1+x)}{x} = \lim_{x\to 0} \frac{1}{1+x} = 1$ by L'Hospital's rule, so $a_j \log(1 + c_j) = (a_j c_j) \frac{\log(1+c_j)}{c_j} \to \lambda$, hence $(1 + c_j)^{a_j} = e^{a_j \log(1+c_j)} \to e^\lambda$. □

Fact 11.2. *If $\lim_{n\to\infty} \max_{1\le j\le n} |c_{j,n}| = 0$, $\lim_{n\to\infty} \sum_{j=1}^n c_{j,n} = \lambda$, and $\sup_n \sum_{j=1}^n |c_{j,n}| < \infty$, then $\lim_{n\to\infty} \prod_{j=1}^n (1 + c_{j,n}) = e^\lambda$.*

Proof. (Homework) It suffices to show that $\sum_{j=1}^n \log(1 + c_{j,n}) \to \lambda$ since then

$$\prod_{j=1}^n (1 + c_{j,n}) = \prod_{j=1}^n e^{\log(1+c_{j,n})} = e^{\sum_{j=1}^n \log(1+c_{j,n})} \to e^\lambda.$$

To this end, note that the first condition ensures that we can choose n large enough that $|c_{j,n}| < 1$, hence $\log(1 + c_{j,n}) = -\sum_{m=1}^\infty (-1)^m \frac{c_{j,n}^m}{m}$ and thus $|\log(1 + c_{j,n}) - c_{j,n}| < \frac{(c_{j,n})^2}{2}$ by standard results for alternating series. It follows that

$$\left|\sum_{j=1}^n \log(1 + c_{j,n}) - \lambda\right| \le \left|\sum_{j=1}^n \log(1 + c_{j,n}) - \sum_{j=1}^n c_{j,n}\right| + \left|\sum_{j=1}^n c_{j,n} - \lambda\right|$$
$$\le \frac{1}{2}\sum_{j=1}^n (c_{j,n})^2 + \left|\sum_{j=1}^n c_{j,n} - \lambda\right| \le \max_{1\le j\le n} |c_{j,n}| \sum_{j=1}^n |c_{j,n}| + \left|\sum_{j=1}^n c_{j,n} - \lambda\right|$$

$$\leq \max_{1\leq j\leq n} |c_{j,n}| \sup_n \sum_{j=1}^{n} |c_{j,n}| + \left| \sum_{j=1}^{n} c_{j,n} - \lambda \right|.$$

The first and third assumptions ensure that the first term goes to zero and the second assumption ensures that the second term goes to zero. □

Example 11.2. Let X_p be the number of trials until the first success in a sequence of independent Bernoulli trials with success probability $p \in (0,1)$. (That is X_p is geometric with parameter p.)

Then $P(X_p > n) = (1-p)^n$, so Fact 11.1 shows that

$$P(pX_p > x) = P\left(X_p > \frac{x}{p}\right) = (1-p)^{\lfloor \frac{x}{p} \rfloor} \to e^{-x}$$

as $p \searrow 0$, hence pX_p converges to the rate 1 exponential distribution.

Example 11.3. Let $X_1, X_2, \ldots$ be independent and uniformly distributed over $\{1, 2, \ldots, N\}$, and let $T_N = \min\{n : X_n = X_m \text{ for some } m < n\}$.

We have

$$P(T_N > n) = \prod_{k=2}^{n} \left(1 - \frac{k-1}{N}\right) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{N}\right).$$

When $N = 365$, this is the probability that no two people in a group of size n have a common birthday. Using Fact 11.2 (with $n = \left\lfloor x\sqrt{N} \right\rfloor - 1$, $c_{j,n} = -j/\left(\frac{n+1}{x}\right)^2$, and $\lambda = -\frac{x^2}{2}$) and the observation that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{\lfloor x\sqrt{N} \rfloor - 1} j = \lim_{N\to\infty} \frac{1}{N} \frac{\left\lfloor x\sqrt{N} \right\rfloor \left(\left\lfloor x\sqrt{N} \right\rfloor - 1\right)}{2} = \frac{x^2}{2},$$

we see that

$$P\left(\frac{T_N}{\sqrt{N}} > x\right) = \prod_{k=1}^{\lfloor x\sqrt{N} \rfloor - 1} \left(1 - \frac{k}{N}\right) \to e^{-\frac{x^2}{2}}$$

as $N \to \infty$.

The approximation $P\left(T_N > x\sqrt{N}\right) \approx e^{-\frac{x^2}{2}}$ with $N = 365$ yields $P\left(T_{365} > 22\right) \approx e^{-0.663} \approx 0.515$ and $P\left(T_{365} > 23\right) \approx e^{-0.725} \approx 0.484$.

This is the *birthday paradox* that in a room of 23 or more people, it is more likely than not that two share a birthday.

Though distributional convergence is defined in terms of distribution functions, it is often convenient to be able to work with random variables when proving theorems.

Theorem 11.1 (Skorokhod Representation)**.** *If $F_n \Rightarrow F_\infty$, then there are random variables Y_n, $1 \leq n \leq \infty$, on a common probability space $(\Omega, \mathcal{F}, P)$ such that F_n is the distribution of Y_n and $Y_n \to Y_\infty$ a.s.*

Proof.

Let $\Omega = (0,1)$, $\mathcal{F} =$ Borel sets, $P =$ Lebesgue measure, and set $Y_n(\omega) = F_n^{-1}(\omega) := \inf\{y : F_n(y) \geq \omega\}$.

We have seen that Y_n has c.d.f. F_n. Also, we know that $\mathcal{D} = \{y : F_\infty \text{ is discontinuous at } y\}$ is countable, so given $\varepsilon > 0$ and $\omega \in (0,1)$, there is some $x \in \mathcal{D}^C$ with $Y_\infty(\omega) - \varepsilon < x < Y_\infty(\omega)$.

By construction, we have that $F_\infty(x) < \omega$, so, since $F_n(x) \to F_\infty(x)$, there is an $N \in \mathbb{N}$ such that $F_n(x) < \omega$ and thus $Y_\infty(\omega) - \varepsilon < x < Y_n(\omega)$ for all $n \geq N$. Accordingly, $\liminf_{n\to\infty} Y_n(\omega) \geq Y_\infty(\omega)$.

Now for any $\omega' > \omega$, there is a $y \in \mathcal{D}^C$ such that $Y_\infty(\omega') < y < Y_\infty(\omega') + \varepsilon$, hence $\omega < \omega' \leq F_\infty(y)$. It follows that for n large enough, $\omega < F_n(y)$ and thus $Y_n(\omega) \leq y < Y_\infty(\omega') + \varepsilon$, so that $\limsup_{n\to\infty} Y_n(\omega) \leq Y_\infty(\omega')$ for all $\omega' > \omega$.

If Y_∞ is continuous at ω, then letting $\omega' \searrow \omega$ gives $\limsup_{n\to\infty} Y_n(\omega) \leq Y_\infty(\omega)$, so $\lim_{n\to\infty} Y_n(\omega) = Y_\infty(\omega)$. Of course, Y_∞ is nondecreasing in ω by construction, so it has only countably many discontinuities, and we conclude that the convergence is almost sure. □

Note that if $Y_n \to Y_\infty$ a.s., then $\mathcal{D} = \{\omega : Y_n(\omega) \nrightarrow Y_\infty(\omega)\}$ has probability zero. Since modifying a random variable on a null set does not change its distribution, we can define $Z_n(\omega) = \begin{cases} Y_n(\omega), & \omega \notin \mathcal{D} \\ 0, & \omega \in \mathcal{D} \end{cases}$ for $1 \leq n \leq \infty$. Then $Z_n =_d Y_n$ and $Z_n(\omega) \to Z_\infty(\omega)$ for all ω, thus almost sure convergence can be replaced by sure convergence in the above theorem.

Our next result gives an equivalent definition of weak convergence. The basic idea is that $C_b(\mathbb{R})$, the space of bounded continuous functions from $\mathbb{R}$ to $\mathbb{R}$ equipped with the supremum norm, is a Banach space. It follows from the Riesz representation theorem that its (continuous) dual $C_b(\mathbb{R})^*$, the space of continuous linear functionals on $C_b(\mathbb{R})$, may be identified with the space of finite and finitely additive signed Radon measures. From this perspective, weak convergence of probability measures corresponds to weak-* convergence in $C_b(\mathbb{R})^*$:

Theorem 11.2. *$X_n \Rightarrow X_\infty$ if and only if for every bounded continuous function g we have $E[g(X_n)] \to E[g(X_\infty)]$.*

Proof. First suppose that $X_n \Rightarrow X_\infty$. Theorem 11.1 shows that there exist random variables with $Y_n =_d X_n$ and $Y_n \to Y_\infty$ a.s. If g is bounded and continuous, then $g(Y_n) \to g(Y_\infty)$ a.s. and bounded convergence gives

$$E[g(X_n)] = E[g(Y_n)] \to E[g(Y_\infty)] = E[g(X_\infty)].$$

To prove the converse, define for each $x \in \mathbb{R}$, $\varepsilon > 0$,

$$g_{x,\varepsilon}(y) = \begin{cases} 1, & y \leq x \\ 0, & y \geq x+\varepsilon \\ 1 - \frac{y-x}{\varepsilon}, & x < y < x+\varepsilon \end{cases}.$$

Since $g_{x,\varepsilon}$ is continuous with $1_{(-\infty,x]} \leq g_{x,\varepsilon} \leq 1_{(-\infty,x+\varepsilon)}$ pointwise, we have

$$\begin{aligned} \limsup_{n\to\infty} P(X_n \leq x) &= \limsup_{n\to\infty} E\left[1_{(-\infty,x]}(X_n)\right] \leq \limsup_{n\to\infty} E\left[g_{x,\varepsilon}(X_n)\right] \\ &= E\left[g_{x,\varepsilon}(X_\infty)\right] \leq E\left[1_{(-\infty,x+\varepsilon]}(X_\infty)\right] = P(X_\infty \leq x+\varepsilon). \end{aligned}$$

Letting $\varepsilon \searrow 0$ gives $\limsup_{n\to\infty} P(X_n \leq x) \leq P(X_\infty \leq x)$.

Similarly,

$$\liminf_{n\to\infty} P(X_n \leq x) \geq \liminf_{n\to\infty} E\left[g_{x-\varepsilon,\varepsilon}(X_n)\right] = E\left[g_{x-\varepsilon,\varepsilon}(X_\infty)\right] \geq P(X_\infty \leq x-\varepsilon),$$

so $\liminf_{n\to\infty} P(X_n \leq x) \geq P(X_\infty < x)$.

This completes the proof since $P(X_\infty < x) = P(X_\infty \leq x)$ if x is a continuity point of F_∞. □

We now show that weak convergence is preserved under (almost) continuous functions.

Theorem 11.3. *Let $g : \mathbb{R} \to \mathbb{R}$ be measurable and set $\mathcal{D}_g = \{x : g \text{ is discontinuous at } x\}$. If $X_n \Rightarrow X_\infty$ and $P(X_\infty \in \mathcal{D}_g) = 0$, then $g(X_n) \Rightarrow g(X_\infty)$. If g is bounded as well, then $E[g(X_n)] \to E[g(X_\infty)]$.*

Proof. Let $Y_n =_d X_n$ with $Y_n \to Y_\infty$ a.s. If f is continuous, then $\mathcal{D}_{f\circ g} \subseteq \mathcal{D}_g$, so $P(Y_\infty \in \mathcal{D}_{f\circ g}) = 0$ and thus $f(g(Y_n)) \to f(g(Y_\infty))$ a.s.

If f is bounded as well, then bounded convergence implies

$$E[f(g(X_n))] = E[f(g(Y_n))] \to E[f(g(Y_\infty))] = E[f(g(X_\infty))].$$

As this is true for all $f \in C_b(\mathbb{R})$, Theorem 11.2 shows that $g(X_n) \Rightarrow g(X_\infty)$.

The second assertion follows by noting that $g(Y_n) \to g(Y_\infty)$ a.s. and likewise applying bounded convergence. □

At this point, we have characterized weak convergence in terms of convergence of distribution functions at continuity points and as weak-* convergence when probability measures are viewed as living in the dual of $C_b(\mathbb{R})$. Here are some further useful definitions.

Theorem 11.4 (Portmanteau Theorem)**.** *The following statements are equivalent:*

- **(i):** $X_n \Rightarrow X_\infty$
- **(ii):** *For all open sets U, $\liminf_{n\to\infty} P(X_n \in U) \geq P(X_\infty \in U)$*
- **(iii):** *For all closed sets K, $\limsup_{n\to\infty} P(X_n \in K) \leq P(X_\infty \in K)$*
- **(iv):** *For all sets A with $P(X_\infty \in \partial A) = 0$, $\lim_{n\to\infty} P(X_n \in A) = P(X_\infty \in A)$*
  (Such an A is called a continuity set *for the distribution of X_∞.)*

Proof. We establish equivalence by showing (i) implies (ii); (ii) implies (iii); (ii) and (iii) imply (iv); and (iv) implies (i).

Suppose that (i) holds. Then there exist random variables $Y_n \to Y_\infty$ on a common probability space $(\Omega, \mathcal{F}, P)$ with $Y_n =_d X_n$ and $Y_n \to Y_\infty$ pointwise.

Now let $\omega \in \Omega$ be such that $Y_\infty(\omega) \in U$. Since $Y_n(\omega) \to Y_\infty(\omega)$, for every open set $V \ni Y_\infty(\omega)$, there is an $N_V \in \mathbb{N}$ with $Y_n(\omega) \in V$ whenever $n \geq N_V$. In particular, there is an $N_U \in \mathbb{N}$ such that $Y_n(\omega) \in U$ for all $n \geq N_U$.

In other words, for all $\omega \in \Omega$ with $1_U(Y_\infty(\omega)) = 1$, we have $1_U(Y_n(\omega)) = 1$ for n sufficiently large, hence $\liminf_{n\to\infty} 1_U(Y_n(\omega)) = 1$. It follows that $\liminf_{n\to\infty} 1_U(Y_n) \geq 1_U(Y_\infty)$ pointwise and thus, by Fatou's lemma and monotonicity, we have

$$\begin{aligned}\liminf_{n\to\infty} P(X_n \in U) &= \liminf_{n\to\infty} P(Y_n \in U) = \liminf_{n\to\infty} E[1_U(Y_n)] \\ &\geq E\left[\liminf_{n\to\infty} 1_U(Y_n)\right] \geq E[1_U(Y_\infty)] = P(Y_\infty \in U) = P(X_\infty \in U),\end{aligned}$$

which is statement (ii).

To see that (ii) implies (iii), observe that if K is closed, then K^C is open, so (ii) implies that $\liminf_{n\to\infty} P(X_n \in K^C) \geq P(X_\infty \in K^C)$, and thus

$$\begin{aligned}P(X_\infty \in K) &= 1 - P(X_n \in K^C) \geq 1 - \liminf_{n\to\infty} P(X_n \in K^C) \\ &= \limsup_{n\to\infty} \left[1 - P(X_n \in K^C)\right] = \limsup_{n\to\infty} P(X_n \in K).\end{aligned}$$

Now assume both (ii) and (iii) are true and let A be such that $P(X_\infty \in \partial A) = 0$. Since $\partial A = \overline{A} \setminus A^\circ$, this means that $P(X_\infty \in \overline{A}) = P(X_\infty \in A^\circ)$. Because $A^\circ \subseteq A \subseteq \overline{A}$, monotonicity implies that the common value is equal to $P(X_\infty \in A)$. Applying (ii) to $A^\circ \subseteq A$ and (iii) to $\overline{A} \supseteq A$ gives

$$\liminf_{n\to\infty} P(X_n \in A) \geq \liminf_{n\to\infty} P\left(X_n \in A^\circ\right) \geq P\left(X_\infty \in A^\circ\right) = P(X_\infty \in A),$$
$$\limsup_{n\to\infty} P(X_n \in A) \leq \limsup_{n\to\infty} P\left(X_n \in \overline{A}\right) \leq P\left(X_\infty \in \overline{A}\right) = P(X_\infty \in A),$$

and (iv) follows.

Finally, suppose that (iv) holds and let F_n denote the distribution function of X_n. Let x be any continuity point of F_∞. Then $P\left(X_\infty \in \{x\}\right) = 0$, so, since $\{x\} = \partial(-\infty, x]$, we have

$$F_n(x) = P(X_n \in (-\infty, x]) \to P(X_\infty \in (-\infty, x]) = F_\infty(x),$$

hence $X_n \Rightarrow X_\infty$. □

Our next set of theorems form a sort of compactness result for certain families of probability measures. We begin with

Theorem 11.5 (Helly's Selection Theorem). *If $\{F_n\}_{n=1}^\infty$ is any sequence of distribution functions, then there is a subsequence $\{F_{n(m)}\}_{m=1}^\infty$ and a nondecreasing, right-continuous function F with $\lim_{m\to\infty} F_{n(m)}(x) = F(x)$ at all continuity points x of F.*

Proof.

We begin with a diagonalization argument: Let $q_1, q_2, ...$ be an enumeration of $\mathbb{Q}$. Since the sequence $\{F_n(q_1)\}_{n=1}^\infty$ is contained in the compact set $[0,1]$, it has a convergence subsequence by the Bolzano-Weierstrass theorem. That is, there exist $n_1(1) < n_1(2) < \cdots$ such that $\left\{F_{n_1(m)}(q_1)\right\}_{m=1}^\infty$ converges to some value $G(q_1) \in [0,1]$. Similarly, the sequence $\left\{F_{n_1(m)}(q_2)\right\}_{m=1}^\infty$ has a subsequence $\left\{F_{n_2(m)}(q_2)\right\}_{m=1}^\infty$ which converges to $G(q_2)$.

In general, we can find a subsequence $\{n_{k+1}(m)\}_{m=1}^\infty$ of $\{n_k(m)\}_{m=1}^\infty$ such that $\lim_{m\to\infty} F_{n_{k+1}(m)}(q_{k+1}) = G(q_{k+1})$ for each $k \geq 1$.

Define the subsequence $\{F_{n(m)}\}_{m=1}^\infty$ by $F_{n(m)} = F_{n_m(m)}$ (so that $\left\{F_{n(k+j)}(q_k)\right\}_{j=1}^\infty$ is a subsequence of $\left\{F_{n_k(m)}(q_k)\right\}_{m=1}^\infty$ for all $k \geq 1$).

By construction, $\lim_{m\to\infty} F_{n(m)}(q) = G(q)$ for all $q \in \mathbb{Q}$.

Also, if $r, s \in \mathbb{Q}$ with $r < s$, then $F_{n(m)}(r) \leq F_{n(m)}(s)$ for all m, hence $G(r) \leq G(s)$.

Now define the function $F : \mathbb{R} \to [0,1]$ by

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\}.$$

To see that F is nondecreasing, note that for any $x < y$, there is some $r \in \mathbb{Q}$ with $x < r < y$.

Since $G(r) \leq G(s)$ for all rational $r < s$, we have

$$F(x) = \inf\{G(q) : q \in \mathbb{Q}, q > x\} \leq G(r) \leq \inf\{G(s) : s \in \mathbb{Q}, s > r\} \leq \inf\{G(s) : s \in \mathbb{Q}, s > y\} = F(y).$$

Now for each $x \in \mathbb{R}$, $\varepsilon > 0$, there is some rational $q > x$ such that $G(q) \leq F(x) + \varepsilon$. Thus if $x \leq y < q$, then $F(y) \leq G(q) \leq F(x) + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we see that F is right-continuous as well.

Finally, suppose that F is continuous at x. Then there exist $r_1, r_2, s \in \mathbb{Q}$ with $r_1 < r_2 < x < s$ such that

$$F(x) - \varepsilon < F(r_1) \leq F(r_2) \leq F(x) \leq F(s) < F(x) + \varepsilon.$$

Since $F_{n(m)}(r_2) \to G(r_2) \geq F(r_1)$ and $F_{n(m)}(s) \to G(s) \leq F(s)$ as $m \to \infty$, we see that for m sufficiently large,

$$F(x) - \varepsilon < F_{n(m)}(r_2) \leq F_{n(m)}(x) \leq F_{n(m)}(s) < F(x) + \varepsilon,$$

hence $F_{n(m)}(x) \to F(x)$. □

It should be noted that the subsequential limit F from Theorem 11.5 is not necessarily a distribution function since the boundary conditions may not hold. When a sequence of distribution functions converges to a nondecreasing right-continuous function at its continuity points, the sequence is said to exhibit *vague convergence*, which will be denoted by $\Rightarrow_v$.

In these terms, Helly's selection theorem says that every sequence of distribution functions has a vaguely convergent subsequence.

Because all of the distribution functions take values in $[0, 1]$, their limit must as well. The limit is thus a Stieltjes measure function for some *subprobability measure* on $\mathbb{R}$ - a positive measure ν with $\nu(\mathbb{R}) \leq 1$.

Thus vague convergence means that the distribution functions converge to a "distribution function" of a subprobability measure, whereas weak convergence means that they converge to the distribution function of a probability measure.

More generally, just as weak convergence is weak-* convergence with respect to $C_b(\mathbb{R})$, vague convergence is weak-* convergence with respect to the subspaces $C_K(\mathbb{R})$ or $C_0(\mathbb{R})$, the spaces of continuous functions with compact support or which vanish at infinity.

This distinction between these notions is illustrated in the following example.

Example 11.4. Choose any $a, b, c > 0$ with $a + b + c = 1$ and any distribution function $G(x)$, and define

$$F_n(x) = a1(x \geq n) + b1(x \geq -n) + cG(x).$$

One easily checks that the $F_n's$ are distribution functions and $F_n(x) \to F(x) := b + cG(x)$. However,

$$\lim_{x \to -\infty} F(x) = b \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1 - a,$$

so F is not a distribution function.

In words, an amount of mass a escapes to $+\infty$ and mass b escapes to $-\infty$.

Intuitively, the test functions in C_K or C_0 that define vague convergence can't detect mass lost to infinity, whereas C_b test functions can.

An immediate question is "Under which conditions do the two definitions coincide?" or "Is there a property of a vaguely convergent sequence of distribution functions which prevents mass from being lost in the limit?"

Definition. A sequence of distribution functions is *tight* if for every $\varepsilon > 0$, there is an $M_\varepsilon > 0$ such that

$$\limsup_{n \to \infty} 1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon) \leq \varepsilon.$$

Theorem 11.6. *A sequence of distribution functions is tight if and only if every subsequential limit is a distribution function.*

Proof. Suppose the sequence is tight and $F_{n(m)} \Rightarrow_v F$. Given $\varepsilon > 0$, let $r < -M_\varepsilon$ and $s > M_\varepsilon$ be continuity points of F.

Since $F_{n(m)}(r) \to F(r)$ and $F_{n(m)}(s) \to F(s)$, we have

$$1 - F(s) + F(r) = \lim_{m\to\infty} \left(1 - F_{n(m)}(s) + F_{n(m)}(r)\right) \leq \varepsilon.$$

As ε was arbitrary, we see that F is indeed a distribution function.

On the other hand, suppose that $\{F_n\}_{n=1}^\infty$ is not tight. Then there is an $\varepsilon > 0$ and a subsequence $\{F_{n(m)}\}_{m=1}^\infty$ with

$$1 - F_{n(m)}(m) + F_{n(m)}(-m) \geq \varepsilon$$

for all m.

Helly's theorem says that there is a further subsequence $\{F_{n(m_k)}\}_{k=1}^\infty$ which converges vaguely to F.

Let $r < 0 < s$ be continuity points of F. Then

$$\begin{aligned} 1 - F(s) + F(r) &= \lim_{k\to\infty} \left(1 - F_{n(m_k)}(s) + F_{n(m_k)}(r)\right) \\ &\geq \liminf_{k\to\infty} \left(1 - F_{n(m_k)}(m_k) + F_{n(m_k)}(-m_k)\right) \geq \varepsilon, \end{aligned}$$

so letting $r \to -\infty$ and $s \to \infty$ along continuity points of F shows that F is not a distribution function. □

Roughly, we have that weak convergence equals vague convergence plus tightness.

We conclude with a sufficient condition for tightness.

Theorem 11.7. *If there is a nonnegative function ϕ such that $\phi(x) \to \infty$ as $x \to \pm\infty$ and*

$$C = \sup_n \int \phi(x) dF_n(x) < \infty,$$

then $\{F_n\}_{n=1}^\infty$ is tight.

Proof. If the assumptions hold, then for every n

$$1 - F_n(M) + F_n(-M) = \int_{|x|\geq M} dF_n(x) \leq \frac{C}{\inf\limits_{|x|\geq M} \phi(x)},$$

which goes to 0 as $M \to \infty$ by assumption. □