10. Weak convergence.

Given Borel probability measures $\mu, \mu_1, \mu_2, \ldots$ on $\mathbf{R}$, we shall write $\mu_n \Rightarrow \mu$, and say that $\{\mu_n\}$ *converges weakly* to μ, if $\int_{\mathbf{R}} f\, d\mu_n \to \int_{\mathbf{R}} f\, d\mu$ for all bounded continuous functions $f : \mathbf{R} \to \mathbf{R}$.

This is a rather natural* definition, though we draw the reader's attention to the fact that this convergence need hold only for *continuous* functions f (as opposed to all Borel-measurable f; cf. Proposition 3.1.8). That is, the "topology" of $\mathbf{R}$ is being used here, not just its measure-theoretic properties.

10.1. Equivalences of weak convergence.

We now present a number of equivalences of weak convergence (see also Exercise 10.3.8). For condition (2), recall that the *boundary* of a set $A \subseteq \mathbf{R}$ is $\partial A = \{x \in \mathbf{R};\ \forall\, \epsilon > 0,\ A \cap (x-\epsilon, x+\epsilon) \neq \emptyset,\ A^C \cap (x-\epsilon, x+\epsilon) \neq \emptyset\}$.

Theorem 10.1.1. *The following are equivalent.*
(1) $\mu_n \Rightarrow \mu$ (i.e., $\{\mu_n\}$ converges weakly to μ);
(2) $\mu_n(A) \to \mu(A)$ for all measurable sets A such that $\mu(\partial A) = 0$;
(3) $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all $x \in \mathbf{R}$ such that $\mu\{x\} = 0$;
(4) (Skorohod's Theorem) there are random variables $Y, Y_1, Y_2, \ldots$ defined jointly on some probability triple, with $\mathcal{L}(Y) = \mu$ and $\mathcal{L}(Y_n) = \mu_n$ for each $n \in \mathbf{N}$, such that $Y_n \to Y$ with probability 1.
(5) $\int_{\mathbf{R}} f\, d\mu_n \to \int_{\mathbf{R}} f\, d\mu$ for all bounded Borel-measurable functions $f : \mathbf{R} \to \mathbf{R}$ such that $\mu(D_f) = 0$, where D_f is the set of points where f is discontinuous.

Proof. (5) $\Longrightarrow$ (1): Immediate.

(5) $\Longrightarrow$ (2): This follows by setting $f = \mathbf{1}_A$, so that $D_f = \partial A$, and $\mu(D_f) = \mu(\partial A) = 0$. Then $\mu_n(A) = \int f\, d\mu_n \to \int f\, d\mu = \mu(A)$.

(2) $\Longrightarrow$ (3): Immediate, since the boundary of $(-\infty, x]$ is $\{x\}$.

(1) $\Longrightarrow$ (3): Let $\epsilon > 0$, and let f be the function defined by $f(t) = 1$ for $t \le x$, $f(t) = 0$ for $t \ge x + \epsilon$, with f linear on the interval $(x, x+\epsilon)$ (see Figure 10.1.2 (a)). Then f is continuous, with $\mathbf{1}_{(-\infty,x]} \le f \le \mathbf{1}_{(-\infty,x+\epsilon]}$. Hence,

$$\limsup_{n\to\infty} \mu_n((-\infty, x]) \le \limsup_{n\to\infty} \int f d\mu_n = \int f d\mu \le \mu((-\infty, x+\epsilon])\ .$$

In fact, it corresponds to the weak ("weak-star") topology from functional analysis, with $\mathcal{X}$ the set of all continuous functions on $\mathbf{R}$ vanishing at infinity (cf. Exercise 10.3.8), with norm defined by $\|f\| = \sup_{x\in\mathbf{R}} |f(x)|$, and with dual space $\mathcal{X}^*$ consisting of all finite signed Borel measures on $\mathbf{R}$. The Helly Selection Principle below then follows from Alaoglu's Theorem. See e.g. pages 161–2, 205, and 216 of Folland (1984).

This is true for any $\epsilon > 0$, so we conclude that $\limsup_{n\to\infty} \mu_n((-\infty, x]) \le \mu((-\infty, x])$.

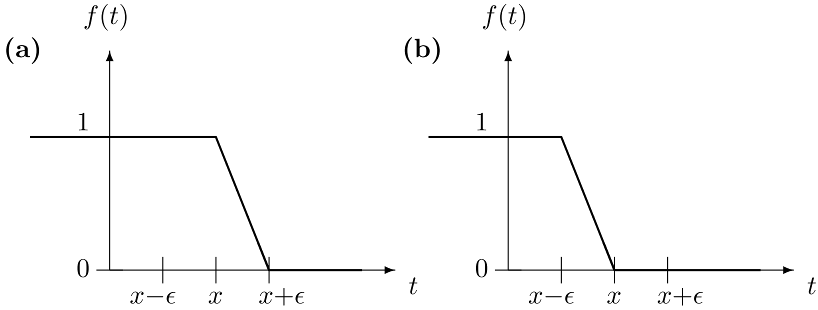

Figure 10.1.2. Functions used in proof of Theorem 10.1.1.

Similarly, if we let f be the function defined by $f(t) = 1$ for $t \le x - \epsilon$, $f(t) = 0$ for $t \ge x$, with f linear on the interval $(x - \epsilon, x)$ (see Figure 10.1.2 (b)), then $\mathbf{1}_{(-\infty, x-\epsilon]} \le f \le \mathbf{1}_{(-\infty, x]}$, and we obtain that

$$\liminf_{n\to\infty} \mu_n((-\infty, x]) \ge \liminf_{n\to\infty} \int f d\mu_n = \int f d\mu \ge \mu((-\infty, x - \epsilon]) .$$

This is true for any $\epsilon > 0$, so we conclude that $\liminf_{n\to\infty} \mu_n((-\infty, x]) \ge \mu((-\infty, x))$.

But if $\mu\{x\} = 0$, then $\mu((-\infty, x]) = \mu((-\infty, x))$, so we must have

$$\limsup_{n\to\infty} \mu_n((-\infty, x]) = \liminf_{n\to\infty} \mu_n((-\infty, x]) = \mu((-\infty, x]) ,$$

as claimed.

(3) $\Longrightarrow$ (4): We first define the cumulative distribution functions, by $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \mu((-\infty, x])$. Then, if we let $(\Omega, \mathcal{F}, \mathbf{P})$ be Lebesgue measure on $[0, 1]$, and let $Y_n(\omega) = \inf\{x; F_n(x) \ge \omega\}$ and $Y(\omega) = \inf\{x; F(x) \ge \omega\}$, then as in Lemma 7.1.2 we have $\mathcal{L}(Y_n) = \mu_n$ and $\mathcal{L}(Y) = \mu$. Note that if $F(z) < a$, then $Y(a) \ge z$, while if $F(z) \ge b$, then $Y(b) \le z$.

Since $\{F_n\} \to F$ at most points, it seems reasonable that $\{Y_n\} \to Y$ at most points. We will prove that $\{Y_n\} \to Y$ at points of continuity of Y. Then, since Y is non-decreasing, it can have at most a countable number of discontinuities: indeed, it has at most $m\big(Y(1-\frac{1}{n})-Y(\frac{1}{n})\big)<\infty$ discontinuities of size $\geq 1/m$ within the interval $(\frac{1}{n},1-\frac{1}{n})$, then take countable union over m and n. Since countable sets have Lebesgue measure 0, this implies that $\{Y_n\}\to Y$ with probability 1, proving (4).

Suppose, then, that Y is continuous at ω, and let $y=Y(\omega)$. For any $\epsilon>0$, we claim that $F(y-\epsilon)<\omega<F(y+\epsilon)$. Indeed, if we had $F(y-\epsilon)=\omega$, then setting $w=y-\epsilon$ and $b=\omega$ above, this would imply $Y(\omega)\leq y-\epsilon=Y(\omega)-\epsilon$, a contradiction. Or, if we had $F(y+\epsilon)=\omega$, then setting $z=y+\epsilon$ and $a=\omega+\delta$ above, this would imply $Y(\omega+\delta)\geq y+\epsilon=Y(\omega)+\epsilon$ for all $\delta>0$, contradicting the continuity of Y at ω. So, $F(y-\epsilon)<\omega<F(y+\epsilon)$ for all $\epsilon>0$.

Next, given $\epsilon>0$, find ϵ' with $0<\epsilon'<\epsilon$ such that $\mu\{y-\epsilon'\}=\mu\{y+\epsilon'\}=0$. Then $F_n(y-\epsilon')\to F(y-\epsilon')$ and $F_n(y+\epsilon')\to F(y+\epsilon')$, so $F_n(y-\epsilon')<\omega<F_n(y+\epsilon')$ for all sufficiently large n. This in turn implies (setting first $z=y-\epsilon'$ and $a=\omega$ above, and then $w=y+\epsilon'$ and $b=\omega$ above) that $y-\epsilon'\leq Y_n(\omega)\leq y+\epsilon'$, i.e. $|Y_n(\omega)-Y(\omega)|\leq\epsilon'<\epsilon$ for all sufficiently large n. Hence, $Y_n(\omega)\to Y(\omega)$.

(4) $\Longrightarrow$ (5): Recall that if f is continuous at x, and if $\{x_n\}\to x$, then $f(x_n)\to f(x)$. Hence, if $\{Y_n\}\to Y$ and $Y\notin D_f$, then $\{f(Y_n)\}\to f(Y)$. It follows that $\mathbf{P}[\{f(Y_n)\}\to f(Y)]\geq\mathbf{P}[\{Y_n\}\to Y$ and $Y\notin D_f]$. But by assumption, $\mathbf{P}[\{Y_n\}\to Y]=1$ and $\mathbf{P}[Y\notin D_f]=\mu(D_f^C)=1$, so also $\mathbf{P}[\{f(Y_n)\}\to f(Y)]=1$. If f is bounded, then from the bounded convergence theorem, $\mathbf{E}\,[f(Y_n)]\to\mathbf{E}\,[f(Y)]$, i.e. $\int f\,d\mu_n\to\int f\,d\mu$, as claimed. ∎

For a first example, let μ be Lebesgue measure on $[0,1]$, and let μ_n be defined by $\mu_n\left(\frac{i}{n}\right)=\frac{1}{n}$ for $i=1,2,\ldots,n$. Then μ is purely continuous while μ_n is purely discrete; furthermore, $\mu(\mathbf{Q})=0$ while $\mu_n(\mathbf{Q})=1$ for each n. On the other hand, for any $0\leq x\leq 1$, we have $\mu\left((-\infty,x]\right)=x$ while $\mu_n\left((-\infty,x]\right)=\lfloor nx\rfloor/n$. Hence, $|\mu_n\left((-\infty,x]\right)-\mu\left((-\infty,x]\right)|\leq\frac{1}{n}\to 0$ as $n\to\infty$, so we do indeed have $\mu_n\Rightarrow\mu$. (Note that $\partial\mathbf{Q}=[0,1]$ so that $\mu(\partial\mathbf{Q})\neq 0$.)

For a second example, suppose $X_1,X_2,\ldots$ are i.i.d. with finite mean m, and $S_n=\frac{1}{n}(X_1+\ldots+X_n)$. Then the weak law of large numbers says that for any $\epsilon>0$ we have $\mathbf{P}\left(S_n\leq m-\epsilon\right)\to 0$ and $\mathbf{P}\left(S_n\leq m+\epsilon\right)\to 1$ as $n\to\infty$. It follows that $\mathcal{L}(S_n)\Rightarrow\delta_m(\cdot)$, a point mass at m. Note that it is *not* necessarily the case that $\mathbf{P}\left(S_n\leq m\right)\to\delta_m\left((-\infty,m]\right)=1$, but this is no contradiction since the boundary of $(-\infty,m]$ is $\{m\}$, and $\delta_m\{m\}\neq 0$.

10.2. Connections to other convergence.

We now explore a sufficient condition for weak convergence.

Proposition 10.2.1. *If $\{X_n\}\to X$ in probability, then $\mathcal{L}(X_n)\Rightarrow\mathcal{L}(X)$.*

Proof. For any $\epsilon > 0$, if $X > z + \epsilon$ and $|X_n - X| < \epsilon$, then we must have $X_n > z$. That is, $\{X > z + \epsilon\} \cap \{|X_n - X| < \epsilon\} \subseteq \{X_n > z\}$. Taking complements, $\{X \le z + \epsilon\} \cup \{|X_n - X| \ge \epsilon\} \supseteq \{X_n \le z\}$. Hence, by the order-preserving property and subadditivity, $\mathbf{P}(X_n \le z) \le \mathbf{P}(X \le z + \epsilon) + \mathbf{P}(|X - X_n| \ge \epsilon)$. Since $\{X_n\} \to X$ in probability, we get that $\limsup_{n\to\infty} \mathbf{P}(X_n \le z) \le \mathbf{P}(X \le z + \epsilon)$. Letting $\epsilon \searrow 0$ gives $\limsup_{n\to\infty} \mathbf{P}(X_n \le z) \le \mathbf{P}(X \le z)$.

Similarly, interchanging X and X_n and replacing z with $z - \epsilon$ in the above gives $\mathbf{P}(X \le z - \epsilon) \le \mathbf{P}(X_n \le z) + \mathbf{P}(|X - X_n| \ge \epsilon)$, or $\mathbf{P}(X_n \le z) \ge \mathbf{P}(X \le z - \epsilon) - \mathbf{P}(|X - X_n| \ge \epsilon)$, so $\liminf \mathbf{P}(X_n \le z) \ge \mathbf{P}(X \le z - \epsilon)$. Letting $\epsilon \searrow 0$ gives $\liminf \mathbf{P}(X_n \le z) \ge \mathbf{P}(X < z)$.

If $\mathbf{P}(X = z) = 0$, then $\mathbf{P}(X < z) = \mathbf{P}(X \le z)$, so we must have $\liminf \mathbf{P}(X_n \le z) = \limsup \mathbf{P}(X_n \le z) = \mathbf{P}(X \le z)$, as claimed. ∎

Remark. We sometimes write $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ simply as $X_n \Rightarrow X$, and say that $\{X_n\}$ converges weakly (or, *in distribution*) to X.

We now have an interesting near-circle of implications. We already knew (Proposition 5.2.3) that if $X_n \to X$ almost surely, then $X_n \to X$ in probability. We now see from Proposition 10.2.1 that this in turn implies that $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$. And from Theorem 10.1.1(4), this implies that there are random variables Y_n and Y having the same laws, such that $Y_n \to Y$ almost surely.

Note that the converse to Proposition 10.2.1 is clearly false, since the fact that $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$ says nothing about the underlying relationship between X_n and X, it only says something about their laws. For example, if $X, X_1, X_2, \ldots$ are i.i.d., each equal to ± 1 with probability $\frac{1}{2}$, then of course $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$, but on the other hand $\mathbf{P}\left(|X - X_n| \ge 2\right) = \frac{1}{2} \not\to 0$, so X_n does *not* converge to X in probability or with probability 1. However, if X is *constant* then the converse to Proposition 10.2.1 does hold (Exercise 10.3.1).

Finally, we note that Skorohod's Theorem may be used to translate results involving convergence with probability 1 to results involving weak convergence (or, by Proposition 10.2.1, convergence in probability). For example, we have

Proposition 10.2.2. *Suppose $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$, with $X_n \ge 0$. Then $\mathbf{E}(X) \le \liminf \mathbf{E}(X_n)$.*

Proof. By Skorohod's Theorem, we can find random variables Y_n and Y with $\mathcal{L}(Y_n) = \mathcal{L}(X_n)$, $\mathcal{L}(Y) = \mathcal{L}(X)$, and $Y_n \to Y$ with probability 1. Then, from Fatou's Lemma,

$$\mathbf{E}(X) = \mathbf{E}(Y) = \mathbf{E}\left(\liminf Y_n\right) \le \liminf \mathbf{E}(Y_n) = \liminf \mathbf{E}(X_n)\,. \qquad ∎$$

For example, if $X \equiv 0$, and if $\mathbf{P}(X_n = n) = \frac{1}{n}$ and $\mathbf{P}(X_n = 0) = 1 - \frac{1}{n}$, then $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$, and $0 = \mathbf{E}(X) \le \liminf \mathbf{E}(X_n) = 1$. (In fact, here $X_n \to X$ in probability, as well.)

Remark 10.2.3. We note that most of these weak convergence concepts have direct analogues for higher-dimensional distributions, not considered here; see e.g. Billingsley (1995, Section 29).

10.3. Exercises.

Exercise 10.3.1. Suppose $\mathcal{L}(X_n) \Rightarrow \delta_c$ for some $c \in \mathbf{R}$. Prove that $\{X_n\}$ converges to c in probability.

Exercise 10.3.2. Let $X, Y_1, Y_2, \ldots$ be independent random variables, with $\mathbf{P}(Y_n = 1) = 1/n$ and $\mathbf{P}(Y_n = 0) = 1 - 1/n$. Let $Z_n = X + Y_n$. Prove that $\mathcal{L}(Z_n) \Rightarrow \mathcal{L}(X)$, i.e. that the law of Z_n converges weakly to the law of X.

Exercise 10.3.3. Let $\mu_n = N(0, \frac{1}{n})$ be a normal distribution with mean 0 and variance $\frac{1}{n}$. Does the sequence $\{\mu_n\}$ converge weakly to some probability measure? If yes, to what measure?

Exercise 10.3.4. Prove that weak limits, if they exist, are *unique*. That is, if $\mu, \nu, \mu_1, \mu_2, \ldots$ are probability measures, and $\mu_n \Rightarrow \mu$, and also $\mu_n \Rightarrow \nu$, then $\mu = \nu$.

Exercise 10.3.5. Let μ_n be the **Poisson**(n) distribution, and let μ be the **Poisson**(5) distribution. Show explicitly that each of the five conditions of Theorem 10.1.1 are violated.

Exercise 10.3.6. Let $a_1, a_2, \ldots$ be any sequence of non-negative real numbers with $\sum_i a_i = 1$. Define the discrete measure μ by $\mu(\cdot) = \sum_{i \in \mathbf{N}} a_i \delta_i(\cdot)$, where $\delta_i(\cdot)$ is a point-mass at the positive integer i. Construct a sequence $\{\mu_n\}$ of probability measures, each having a density with respect to Lebesgue measure, such that $\mu_n \Rightarrow \mu$.

Exercise 10.3.7. Let $\mathcal{L}(Y) = \mu$, where μ has continuous density f. For $n \in \mathbf{N}$, let $Y_n = \lfloor nY \rfloor / n$, and let $\mu_n = \mathcal{L}(Y_n)$.
(a) Describe μ_n explicitly.
(b) Prove that $\mu_n \Rightarrow \mu$.
(c) Is μ_n discrete, or absolutely continuous, or neither? What about μ?

Exercise 10.3.8. Prove that the following are equivalent.
(1) $\mu_n \Rightarrow \mu$.
(2) $\int f\,d\mu_n \to \int f\,d\mu$ for all *non-negative* bounded continuous $f : \mathbf{R} \to \mathbf{R}$.
(3) $\int f\,d\mu_n \to \int f\,d\mu$ for all non-negative continuous $f : \mathbf{R} \to \mathbf{R}$ with *compact support*, i.e. such that there are finite a and b with $f(x) = 0$ for all $x < a$ and all $x > b$.
(4) $\int f\,d\mu_n \to \int f\,d\mu$ for all continuous $f : \mathbf{R} \to \mathbf{R}$ with compact support.
(5) $\int f\,d\mu_n \to \int f\,d\mu$ for all non-negative continuous $f : \mathbf{R} \to \mathbf{R}$ which *vanish at infinity*, i.e. such that $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$.
(6) $\int f\,d\mu_n \to \int f\,d\mu$ for all continuous $f : \mathbf{R} \to \mathbf{R}$ which vanish at infinity.
[Hints: You may assume the fact that all continuous functions on $\mathbf{R}$ which have compact support or vanish at infinity are bounded. Then, showing that (1) $\Longrightarrow$ each of (4)–(6), and that each of (4)–(6) $\Longrightarrow$ (3), is easy. For (2) $\Longrightarrow$ (1), note that if $|f| \le M$, then $f + M$ is non-negative. For (3) $\Longrightarrow$ (2), note that if f is non-negative bounded continuous and $m \in \mathbf{Z}$, then $f_m \equiv f\,\mathbf{1}_{[m,m+1)}$ is non-negative bounded with compact support and is "nearly" continuous; then recall Figure 10.1.2, and that $f = \sum_{m\in\mathbf{Z}} f_m$.]

Exercise 10.3.9. Let $0 < M < \infty$, and let $f, f_1, f_2, \ldots : [0,1] \to [0, M]$ be Borel-measurable functions with $\int_0^1 f\,d\lambda = \int_0^1 f_n\,d\lambda = 1$. Suppose $\lim_n f_n(x) = f(x)$ for each fixed $x \in [0,1]$. Define probability measures $\mu, \mu_1, \mu_2, \ldots$ by $\mu(A) = \int_A f\,d\lambda$ and $\mu_n(A) = \int_A f_n\,d\lambda$, for Borel $A \subseteq [0,1]$. Prove that $\mu_n \Rightarrow \mu$.

Exercise 10.3.10. Let $f : [0,1] \to (0,\infty)$ be a continuous function such that $\int_0^1 f\,d\lambda = 1$ (where λ is Lebesgue measure on $[0,1]$). Define probability measures μ and $\{\mu_n\}$ by $\mu(A) = \int_0^1 f\,\mathbf{1}_A\,d\lambda$ and $\mu_n(A) = \sum_{i=1}^n f(i/n)\,\mathbf{1}_A(i/n) \,/\, \sum_{i=1}^n f(i/n)$.
(a) Prove that $\mu_n \Rightarrow \mu$. [Hint: Recall Riemann sums from calculus.]
(b) Explicitly construct random variables Y and $\{Y_n\}$ so that $\mathcal{L}(Y) = \mu$, $\mathcal{L}(Y_n) = \mu_n$, and $Y_n \to Y$ with probability 1. [Hint: Remember the proof of Theorem 10.1.1.]

10.4. Section summary.

This section introduced the notion of weak convergence. It proved equivalences of weak convergence in terms of convergence of expectations of bounded continuous functions, convergence of probabilities, convergence of cumulative distribution functions, and the existence of corresponding random variables which converge with probability 1. It proved that if random variables converge in probability (or with probability 1), then their laws converge weakly.

Weak convergence will be very important in the next section, including allowing for a precise statement and proof of the Central Limit Theorem (Theorem 11.2.2 on page 134).

This page intentionally left blank