Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 7
Weak Convergence

The definition

In addition to the modes of convergence we introduced so far (a.s.-convergence, convergence in probability and $\mathcal{L}^p$-convergence), there is another one, called **weak convergence** or **convergence in distribution**. Unlike the other three, whether a sequence of random variables (elements) converges in distribution or not depends only on their *distributions*. In addition to its intrinsic mathematical interest, convergence in distribution (or, equivalently, the weak convergence) is precisely the kind of convergence we encounter in the central limit theorem.

We take the abstraction level up a notch and consider sequences of probability measures on $(S, \mathcal{S})$, where (S, d) is a metric space and $\mathcal{S} = \mathcal{B}(d)$ is the Borel σ-algebra there. In fact, it will always be assumed that S is a metric space and $\mathcal{S}$ is the Borel σ-algebra on it, throughout this chapter.

Definition 7.1. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $(S, \mathcal{S})$. We say that μ_n converges **weakly**[1] to a probability measure μ on $(S, \mathcal{S})$ - and write $\mu_n \stackrel{w}{\to} \mu$ - if

$$\int f\, d\mu_n \to \int f\, d\mu,$$

for all $f \in C_b(S)$, where $C_b(S)$ denotes the set of all continuous and bounded functions $f : S \to \mathbb{R}$.

[1] It would be more in tune with standard mathematical terminology to use the term *weak-*$*$ convergence instead of weak convergence. For historical reasons, however, we omit the $*$.

Definition 7.2. A sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables (elements) is said to **converge in distribution** to the random variable (element) X, denoted by $X_n \stackrel{\mathcal{D}}{\to} X$, if $\mu_{X_n} \stackrel{w}{\to} \mu_X$.

For the uniqueness of limits, we need a simple approximation result:

Problem 7.1. Let F be a closed set in S. Show that for any $\varepsilon > 0$ there exists a Lipschitz and bounded function $f_{F;\varepsilon} : S \to \mathbb{R}$ such that

1. $0 \leq f_{F;\varepsilon}(x) \leq 1$, for all $x \in \mathbb{R}$,
2. $f_{F;\varepsilon}(x) = 1$ for $x \in F$, and
3. $f_{F;\varepsilon}(x) = 0$ for $d(x,F) \geq \varepsilon$, where $d(x,F) = \inf\{d(x,y) : y \in F\}$.

Hint: Show first that the function $x \mapsto d(x,F)$ is Lipschitz. Then argue that $f_{F;\varepsilon}(x) = h(d(x,F))$ has the required properties for a well-chosen function $h : [0,\infty) \to [0,1]$.

Proposition 7.3. *Suppose that $\{\mu_n\}_{n\in\mathbb{N}}$ is a sequence of probability measures on $(S,\mathcal{S})$ such that $\mu_n \overset{w}{\to} \mu$ and $\mu_n \overset{w}{\to} \mu'$. Then $\mu = \mu'$.*

Proof. By the very definition of weak convergence, we have

$$\int f\, d\mu = \lim_n \int f\, d\mu_n = \int f\, d\mu', \tag{7.1}$$

for all $f \in C_b(S)$. Let F be a closed set, and let $\{f_k\}_{k\in\mathbb{N}}$ be as in Problem 7.1, with $f_k = f_{F;\varepsilon}$ corresponding to $\varepsilon = 1/k$. If we set $F_k = \{x \in S : d(x,F) \leq 1/k\}$, then F_k is a closed set (why?) and we have $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$. By (7.1), we have

$$\mu(F) \leq \int f_k\, d\mu = \int f_k d\mu' \leq \mu'(F_k),$$

and, similarly, $\mu'(F) \leq \mu(F_k)$, for all $k \in \mathbb{N}$. Since $F_k \searrow F$ (why?), we have $\mu(F_k) \searrow \mu(F)$ and $\mu'(F_k) \searrow \mu'(F)$, and it follows that $\mu(F) = \mu'(F)$.

It remains to note that the family of all closed sets is a π-system which generates the σ-algebra $\mathcal{S}$ to conclude that $\mu = \mu'$. □

Out next task if to give a useful operational characterization of weak convergence. Before we do that, we need a simple observation; remember that ∂A denotes the topological boundary $\partial A = \operatorname{Cl} A \setminus \operatorname{Int} A$ of a set $A \subseteq S$.

Problem 7.2. Let $(F_\gamma)_{\gamma\in\Gamma}$ be a partition of S into (possibly uncountably many) measurable subsets. Show that for any probability measure μ on $\mathcal{S}$, $\mu(F_\gamma) = 0$, for all but countably many $\gamma \in \Gamma$

Hint: For $n \in \mathbb{N}$, define $\Gamma_n = \{\gamma \in \Gamma : \mu(F_\gamma) \geq \frac{1}{n}\}$. Argue that Γ_n has at most n elements.

Definition 7.4. A set $A \in \mathcal{S}$ with the property that $\mu(\partial A) = 0$, is called a μ-**continuity set**

Theorem 7.5 (Portmanteau Theorem). *Let μ, $\{\mu_n\}_{n\in\mathbb{N}}$ be probability measures on $\mathcal{S}$. Then, the following are equivalent:*

1. $\mu_n \overset{w}{\to} \mu$,
2. $\int f\, d\mu_n \to \int f\, d\mu$, *for all bounded,* Lipschitz *continuous* $f : S \to \mathbb{R}$,

3. $\limsup_n \mu_n(F) \leq \mu(F)$, *for all closed* $F \subseteq S$,
4. $\liminf_n \mu_n(G) \geq \mu(G)$, *for all open* $G \subseteq S$,
5. $\lim_n \mu_n(A) = \mu(A)$, *for all* μ*-continuity sets* $A \in \mathcal{S}$.

Note: Here is a way to remember whether closed sets go together with the lim inf or the lim sup: take a convergent sequence $\{x_n\}_{n\in\mathbb{N}}$ in S, with $x_n \to x$. If μ_n is the Dirac measure concentrated on x_n, and μ the Dirac measure concentrated on x, then clearly $\mu_n \stackrel{w}{\to} \mu$ (since $\int f\, d\mu_n = f(x_n) \to f(x) = \int f\, d\mu$). Let F be a closed set. It can happen that $x_n \notin F$ for all $n \in \mathbb{N}$, but $x \in F$ (think of x on the boundary of F). Then $\mu_n(F) = 0$, but $\mu(F) = 1$ and so $\limsup \mu_n(F) = 0 < 1 = \mu(F)$.

Proof. 1. ⇒ 2.: trivial.

2. ⇒ 3.: given a closed set F and let $F_k = \{x \in S : d(x, F) \leq 1/k\}$, $f_k = f_{F;1/k}$, $k \in \mathbb{N}$, be as in the proof of Proposition 7.3. Since $\mathbf{1}_F \leq f_k \leq \mathbf{1}_{F_k}$ and the functions f_k are Lipschitz continuous, we have

$$\limsup_n \mu_n(F) = \limsup_n \int \mathbf{1}_F\, d\mu_n \leq \lim_n \int f_k\, d\mu_n = \int f_k\, d\mu \leq \mu(F_k),$$

for all $k \in \mathbb{N}$. Letting $k \to \infty$ - as in the proof of Proposition 7.3 - yields 3.

3. ⇒ 4.: follows directly by taking complements.

4. ⇒ 1.: Pick $f \in C_b(S)$ and (possibly after applying a linear transformation to it) assume that $0 < f(x) < 1$, for all $x \in S$. Then, by Problem 5.10, we have $\int f\, d\nu = \int_0^1 \nu(f > t)\, dt$, for any probability measure on $\mathcal{B}(\mathbb{R})$. The set $\{f > t\} \subseteq S$ is open, so by 3., $\liminf_n \mu_n(f > t) \geq \mu(f > t)$, for all t. Therefore, by Fatou's lemma,

$$\begin{aligned}\liminf_n \int f\, d\mu_n &= \liminf_n \int_0^1 \mu_n(f > t)\, dt \geq \int_0^1 \liminf_n \mu_n(f > t)\, dt \\ &\geq \int_0^1 \mu(f > t)\, dt = \int f\, d\mu.\end{aligned}$$

We get the other inequality - $\int f\, d\mu \geq \limsup_n \int f\, d\mu_n$, by repeating the procedure with f replaced by $-f$.

3., 4. ⇒ 5.: Let A be a μ-continuity set, let $\operatorname{Int} A$ be its interior and $\operatorname{Cl} A$ its closure. Then, since $\operatorname{Int} A$ is open and $\operatorname{Cl} A$ is closed, we have

$$\begin{aligned}\mu(\operatorname{Int} A) &\leq \liminf_n \mu_n(\operatorname{Int} A) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(A) \\ &\leq \limsup_n \mu_n(\operatorname{Cl} A) \leq \mu(\operatorname{Cl} A).\end{aligned}$$

Since $0 = \mu(\partial A) = \mu(\operatorname{Cl} A \setminus \operatorname{Int} A) = \mu(\operatorname{Cl} A) - \mu(\operatorname{Int} A)$, we conclude that all inequalities above are, in fact, equalities so that $\mu(A) = \lim_n \mu_n(A)$

5. ⇒ 3.: For $x \in S$, consider the family $\{B_F(r) : r \geq 0\}$, where

$$B_F(r) = \{x \in S : d(x, F) \leq r\},$$

of closed sets.

Claim: There exists a countable subset R of $[0,\infty)$ such that $B_F(r)$ is a μ-continuity set for all $r \in [0,\infty) \setminus R$.

Proof. For $r \geq 0$ define $C_F(r) = \{x \in S : d(x,F) = r\}$, so that $\{C_F(r) : r \geq 0\}$ forms a measurable partition of S. Therefore, by Problem, 7.2, there exists a countable set $R \subseteq [0,\infty)$ such that $\mu(C_F(r)) = 0$ for $r \in [0,\infty) \setminus R$. It is not hard to see that $\partial B_F(r) \subseteq C_F(r)$ (btw, the inclusion may be strict), for each $r \geq 0$. Therefore, $\mu(\partial B_F(r)) = 0$, for all $r \in [0,\infty) \setminus R$. □

The above claim implies that there exists a sequence $r_k \in [0,\infty) \setminus R$ such that $r_k \searrow 0$. By 5. and the Claim above, we have $\mu_n(B_F(r_k)) \to \mu(B_F(r_k))$ for all $k \in \mathbb{N}$. Hence, for $k \in \mathbb{N}$,

$$\mu(B_F(r_k)) = \lim_n \mu_n(B_F(r_k)) \geq \limsup_n \mu_n(F).$$

By continuity of measure we have $\mu(B_F(r_k)) \searrow \mu(F)$, as $k \to \infty$, and so $\mu(F) \geq \limsup_n \mu_n(F)$. □

As we will soon see, it is sometimes easy to prove that $\mu_n(A) \to \mu(A)$ for all A in some subset of $\mathcal{B}(\mathbb{R})$. Our next result has something to say about cases when that is enough to establish weak convergence:

Proposition 7.6. *Let $\mathcal{I}$ be a collection of open subsets of S such that*

1. *$\mathcal{I}$ is a π-system,*
2. *Each open set in S can be represented as a finite or countable union of elements of $\mathcal{I}$.*

If $\mu_n(I) \to \mu(I)$, for each $I \in \mathcal{I}$, then $\mu_n \xrightarrow{w} \mu$.

Proof. For $I_1, I_2 \in \mathcal{I}$, we have $I_1 \cap I_2 \in \mathcal{I}$, and so

$$\begin{aligned}\mu(I_1 \cup I_2) &= \mu(I_1) + \mu(I_2) - \mu(I_1 \cap I_2)\\ &= \lim_n \mu_n(I_1) + \lim_n \mu_n(I_2) - \lim_n \mu_n(I_1 \cap I_2)\\ &= \lim_n \Big(\mu_n(I_1) + \mu_n(I_2) - \mu_n(I_1 \cap I_2)\Big) = \lim_n \mu_n(I_1 \cup I_2).\end{aligned}$$

Therefore, we can assume, without loss of generality that $\mathcal{I}$ is closed under finite unions.

For an open set G, let $G = \cup_k I_k$ be a representation of G as a union of a countable family in $\mathcal{I}$. By continuity of measure, for each $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that $\mu(G) \leq \mu(\cup_{k=1}^K I_k) + \varepsilon$. Since $\cup_{k=1}^K I_K \in \mathcal{I}$, we have

$$\mu(G) - \varepsilon \leq \mu(\cup_{k=1}^K I_k) = \lim_n \mu_n(\cup_{k=1}^K I_k) \leq \liminf \mu_n(G).$$

Given that $\varepsilon > 0$ was arbitrary, we get $\mu(G) \leq \liminf_n \mu_n(G)$. □

Corollary 7.7. *Suppose that $S = \mathbb{R}$, and let μ_n be a family of probability measures on $\mathcal{B}(\mathbb{R})$. Let $F(x) = \mu((-\infty, x])$ and $F_n(x) = \mu_n((-\infty, x])$, $x \in \mathbb{R}$ be the corresponding cdfs. Then, the following two statements are equivalent:*

1. *$F_n(x) \to F(x)$ for all x such that F is continuous at x, and*
2. *$\mu_n \overset{w}{\to} \mu$.*

Proof. 2. $\Rightarrow$ 1.: Let C be the set of all x such that F is continuous at x; eqivalently, $C = \{x \in \mathbb{R} : \mu(\{x\}) = 0\}$. The sets $(-\infty, x]$ are μ-continuity sets for $x \in C$, so the Portmanteau theorem (Theorem 7.5) implies that $F_n(x) = \mu_n((-\infty, x]) \to \mu((-\infty, x]) = F(x)$, for all $x \in C$.

1. $\Rightarrow$ 2.: The set C^c is at most countable (why?) and so the family

$$\mathcal{I} = \{(a, b) : a < b, a, b \in C\},$$

satisfies the the conditions of Proposition 7.6. To show that $\mu_n \overset{w}{\to} \mu$, it will be enough to show that $\mu_n(I) \to \mu(I)$, for all $a, b \in \mathcal{I}$. Since $\mu((a, b)) = F(b-) - F(a)$, where $F(b-) = \lim_{x \nearrow b} F(x)$, it will be enough to show that

$$F_n(x-) \to F(x),$$

for all $x \in C$. Since $F_n(x-) \leq F_n(x)$, we have $\limsup_n F_n(x-) \leq \lim F_n(x) = F(x)$. To prove the other inequality, we pick $\varepsilon > 0$, and, using the continuity of F at x, find $\delta > 0$ such that $x - \delta \in C$ and $F(x - \delta) > F(x) - \varepsilon$. Since $F_n(x - \delta) \to F(x - \delta)$, there exists $n_0 \in \mathbb{N}$ such that $F_n(x - \delta) > F(x) - 2\varepsilon$ for $n \geq n_0$, and, by increase of F_n, $F_n(x-) > F(x) - 2\varepsilon$, for $n \geq n_0$. Consequently $\liminf_n F_n(x-) \geq F(x) - 2\varepsilon$ and the statement follows. □

One of the (many) reasons why weak convergence is so important, is the fact that it possesses nice compactness properties. The central result here is the theorem of Prohorov which is, in a sense, an analogue of the Arzelá-Ascoli compactness theorem for families of measures. The statement we give here is not the most general possible, but it will serve all our purposes.

Definition 7.8. A subset $\mathcal{M}$ of probability measures on $\mathcal{S}$ is said to be

1. **tight**, if for each $\varepsilon > 0$ there exists a compact set K such that

$$\sup_{\mu \in \mathcal{M}} \mu(K^c) \leq \varepsilon.$$

2. **relatively (sequentially) weakly compact** if any sequence $\{\mu_n\}_{n \in \mathbb{N}}$ in $\mathcal{M}$ admits a weakly-convergent subsequence $\{\mu_{n_k}\}_{k \in \mathbb{N}}$.

Theorem 7.9 (Prohorov). *Suppose that the metric space (S, d) is complete and separable, and let $\mathcal{M}$ be a set of probability measures on $\mathcal{S}$. Then $\mathcal{M}$ is relatively weakly compact if and only if it is tight.*

Proof. (Tight $\Rightarrow$ relatively weakly compact): Suppose that $\mathcal{M}$ is tight, and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}$. Let Q be a countable and dense subset of $\mathbb{R}$, and let $\{q_k\}_{k\in\mathbb{N}}$ be an enumeration of Q. Since all $\{\mu_n\}_{n\in\mathbb{N}}$ are probability measures, the sequence $\{F_n(q_1)\}_{n\in\mathbb{N}}$, where $F_n(x) = \mu_n((-\infty, x])$ is bounded. Consequently, it admits a convergent subsequence; we denote its indices by $n_{1,k}$, $k \in \mathbb{N}$. The sequence $\{F_{n_{1,k}}(q_2)\}_{k\in\mathbb{N}}$ is also bounded, so we can extract a further subsequence - let's denote it by $n_{2,k}$, $k \in \mathbb{N}$, so that $F_{n_{2,k}}(q_2)$ converges as $k \to \infty$. Repeating this procedure for each element of Q, we arrive to a sequence of increasing sequences of integers $n_{i,k}$, $k \in \mathbb{N}$, $i \in \mathbb{N}$ with the property that $n_{i+1,k}$, $k \in \mathbb{N}$ is a subsequence of $n_{i,k}$, $k \in \mathbb{N}$ and that $F_{n_{i,k}}(q_j)$ converges for each $j \le i$. Therefore, the diagonal sequence $m_k = n_{k,k}$, is a subsequence of each $n_{i,k}$, $k \in \mathbb{N}$, $i \in \mathbb{N}$, and can define a function $\tilde{F} : Q \to [0, 1]$ by

Note: In addition to the fact that the stated version of the theorem is not the most general available, we only give the proof for the so-called *Helly's selection theorem*, i.e., the special case $S = \mathbb{R}$. The general case is technically more involved, but the key ideas are similar.

$$\tilde{F}(q) = \lim_{k\to\infty} F_{m_k}(q).$$

Each F_n is non-decreasing and so if $\tilde{F}$. As a matter of fact the "right-continuous" version

$$F(x) = \inf_{q<x, q\in Q} \tilde{F}(q),$$

is non-decreasing and right-continuous (why?), with values in $[0, 1]$.

Our next task is to show that $F_{m_k}(x) \to F(x)$, for each $x \in C_F$, where C_F is the set of all points where F is continuous. We pick $x \in C_F$, $\varepsilon > 0$ and $q_1, q_2 \in Q$, $y \in \mathbb{R}$ such that $q_1 < q_2 < x < y$ and

$$F(x) - \varepsilon < F(q_1) \le F(q_2) \le F(x) \le F(y) < F(x) + \varepsilon.$$

Since $F_{m_k}(q_2) \to \tilde{F}(q_2) \ge F(q_1)$ and $F_{m_k}(s) \to \tilde{F}(s) \le F(s)$ (why is $\tilde{F}(s) \le F(s)$?), we have, for large enough $k \in \mathbb{N}$

$$F(x) - \varepsilon < F_{m_k}(q_2) \le F_{m_k}(x) \le F_{m_k}(s) < F(x) + \varepsilon,$$

which implies that $F_{m_k}(x) \to F(x)$.

It remains to show - thanks to Corollary 7.7 - that $F(x) = \mu((-\infty, x])$, for some probability measure μ on $\mathcal{B}(\mathbb{R})$. For that, in turn, it will be enough to show that $F(x) \to 1$, as $x \to \infty$ and $F(x) \to 0$, as $x \to -\infty$. Indeed, in that case, we would have all the conditions needed to use Problem 6.24 to construct a probability space and a random variable X on it so that F is the cdf of X; the required measure μ would be the distribution $\mu = \mu_X$ of X.

To show that $F(x) \to 0, 1$ as $x \to \pm\infty$, we use tightness (note that this is the only place in the proof where it is used). For $\varepsilon > 0$, we pick

$M > 0$ such that $\mu_n([-M, M]) \geq 1 - \varepsilon$, for all $n \in \mathbb{N}$. In terms of corresponding cdfs, this implies that

$$F_n(-M) \leq \varepsilon \text{ and } F_n(M) \geq 1 - \varepsilon \text{ for all } n \in \mathbb{N}.$$

We can assume that $-M$ and M are continuity points of F (why?), so that

$$F(-M) = \lim_k F_{m_k}(-M) \leq \varepsilon \text{ and } F(M) = \lim_k F_{m_k}(M) \geq 1 - \varepsilon,$$

so that $\lim_{x\to\infty} F(x) \geq 1 - \varepsilon$ and $\lim_{x\to-\infty} F(x) \leq \varepsilon$. The claim follows from the arbitrariness of $\varepsilon > 0$.

(Relatively weakly compact $\Rightarrow$ tight): Suppose to the contrary, that $\mathcal{M}$ is relatively weakly compact, but not tight. Then, there exists $\varepsilon > 0$ such that for each $n \in \mathbb{N}$ there exists $\mu_n \in \mathcal{M}$ such that $\mu_n([-n, n]) < 1 - \varepsilon$, and, consequently,

$$\mu_n([-M, M]) < 1 - \varepsilon \text{ for } n \geq M. \tag{7.2}$$

The sequence $\{\mu_n\}_{n\in\mathbb{N}}$ admits a weakly-convergent subsequence, denoted by $\{\mu_{n_k}\}_{k\in\mathbb{N}}$. By (7.2), we have

$$\limsup_k \mu_{n_k}([-M, M]) \leq 1 - \varepsilon, \text{ for each } M > 0,$$

so that $\mu([-M, M]) \leq 1 - \varepsilon$ for all $M > 0$. Continuity of probability implies that $\mu(\mathbb{R}) \leq 1 - \varepsilon$ - a contradiction with the fact that μ is a *probability* measure on $\mathcal{B}(\mathbb{R})$. □

The following problem cases tightness in more operational terms:

Problem 7.3. Let $\mathcal{M}$ be a non-empty set of probability measures on $\mathbb{R}$. Show that $\mathcal{M}$ is tight if and only if there exists a non-decreasing function $\varphi : [0, \infty) \to [0, \infty)$ such that

1. $\varphi(x) \to \infty$ as $x \to \infty$, and
2. $\sup_{\mu\in\mathcal{M}} \int \varphi(|x|)\, \mu(dx) < \infty$.

Prohorov's theorem goes well with the following problem (it will be used soon):

Problem 7.4. Let μ be a probability measure on $\mathcal{B}(\mathbb{R})$ and let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$ with the property that every subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mu_n\}_{n\in\mathbb{N}}$ has a (further) subsequence $\{\mu_{n_{k_l}}\}_{l\in\mathbb{N}}$ which converges towards μ. Show that $\{\mu_n\}_{n\in\mathbb{N}}$ is convergent.

Hint: If $\mu_n \overset{w}{\not\to} \mu$, then there exists $f \in C_b$ and a subsequence $\{\mu_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mu_n\}_{n\in\mathbb{N}}$ such that $\int f\, d\mu_{n_k}$ converges, but not to $\int f\, d\mu$.

We conclude with a comparison between convergence in distribution and convergence in probability.

Proposition 7.10 (Relation between $\stackrel{\mathbb{P}}{\to}$ and $\stackrel{\mathcal{D}}{\to}$)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Then $X_n \stackrel{\mathbb{P}}{\to} X$ implies $X_n \stackrel{\mathcal{D}}{\to} X$, for any random variable X. Conversely, $X_n \stackrel{\mathcal{D}}{\to} X$ implies $X_n \stackrel{\mathbb{P}}{\to} X$ if there exists $c \in \mathbb{R}$ such that $\mathbb{P}[X = c] = 1$.*

Proof. Assume that $X_n \stackrel{\mathbb{P}}{\to} X$. To show that $X_n \stackrel{\mathcal{D}}{\to} X$, the Portmanteau theorem guarantees that it will be enough to prove that $\limsup_n \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F]$, for all closed sets F. For $F \subseteq \mathbb{R}$, we define $F^\varepsilon = \{x \in \mathbb{R} : d(x, F) \leq \varepsilon\}$. Therefore, for a closed set F, we have

$$\begin{aligned}\mathbb{P}[X_n \in F] &= \mathbb{P}[X_n \in F, |X - X_n| > \varepsilon] + \mathbb{P}[X_n \in F, |X - X_n| \leq \varepsilon] \\ &\leq \mathbb{P}[|X - X_n| > \varepsilon] + \mathbb{P}[X \in F_\varepsilon].\end{aligned}$$

because $X \in F_\varepsilon$ if $X_n \in F$ and $|X - X_n| \leq \varepsilon$. Taking a lim sup of both sides yields

$$\limsup \mathbb{P}[X_n \in F] \leq \mathbb{P}[X \in F_\varepsilon] + \limsup_n \mathbb{P}[|X - X_n| > \varepsilon] = \mathbb{P}[X \in F_\varepsilon].$$

Since $\cap_{\varepsilon>0} F^\varepsilon = F$, the statement follows.

For the second part, without loss of generality, we assume $c = 0$. Given $m \in \mathbb{N}$, let $f_m \in C_b(\mathbb{R})$ be a continuous function with values in $[0, 1]$ such that $f_m(0) = 1$ and $f_m(x) = 0$ for $|x| > 1/m$. Since $f_m(x) \leq \mathbf{1}_{[-1/m,1/m]}(x)$, we have

$$\mathbb{P}[|X_n| \leq 1/m] \geq \mathbb{E}[f_m(X_n)] \to f_m(0) = 1,$$

for each $m \in \mathbb{N}$. □

Remark 7.11. It is not true that $X_n \stackrel{\mathcal{D}}{\to} X$ implies $X_n \stackrel{\mathbb{P}}{\to} X$ in general. Here is a simple example: take $\Omega = \{1, 2\}$ with uniform probability, and define $X_n(1) = 1$ and $X_n(2) = 2$, for n odd and $X_n(1) = 2$ and $X_n(2) = 1$, for n even. Then all X_n have the same distribution, so we have $X_n \stackrel{\mathcal{D}}{\to} X_1$. On the other hand $\mathbb{P}[|X_n - X_1| \geq \frac{1}{2}] = 1$, for n even. In fact, it is not hard to see that $X_n \stackrel{\mathbb{P}}{\not\to} X$ for any random variable X.

Additional Problems

Problem 7.5 (Total-variation convergence)**.** A sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of probability measures on $\mathcal{B}(\mathbb{R})$ is said to converge to the probability measure μ **in (total) variation** if

$$\sup_{A\in\mathcal{B}(\mathbb{R})} |\mu_n(A) - \mu(A)| \to 0 \text{ as } n \to \infty.$$

Compare convergence in variation to weak convergence: if one implies the other, prove it. Give counterexamples, if they are not equivalent.

Problem 7.6 (Scheffé's Theorem)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be absolutely continuous random variables with densities f_{X_n}, such that $f_{X_n}(x) \to f(x)$, λ-a.e., where f is the density of the absolutely-continuous random variable X. Show that X_n converges to X in total variation (defined in Problem 7.5), and, therefore, also in distribution.

Hint: Show that $\int_{\mathbb{R}} |f_{X_n} - f|\, d\lambda \to 0$ by writing the integrand in terms of $(f - f_{X_n})^+ \leq f$.

Problem 7.7 (Convergence of moments)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables with a common uniform bound, i.e., such that

$$\exists\, M > 0,\ \forall\, n \in \mathbb{N},\ |X_n| \leq M,\ |X| \leq M, \text{ a.s.}$$

Hint: Use the Weierstrass approximation theorem: *Given $a < b \in \mathbb{R}$, a continuous function $f : [a, b] \to \mathbb{R}$ and $\varepsilon > 0$ there exists a polynomial P such that* $\sup_{x\in[a,b]} |f(x) - P(x)| \leq \varepsilon$.

Show that the following two statements are equivalent:

1. $X_n \overset{\mathcal{D}}{\to} X$ (where $\overset{\mathcal{D}}{\to}$ denotes convergence in distribution), and
2. $\mathbb{E}[X_n^k] \to \mathbb{E}[X^k]$, as $n \to \infty$, for all $k \in \mathbb{N}$.

Problem 7.8 (Convergence of Maxima)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence of standard normal ($N(0,1)$) random variables. Define the sequence of *up-to-date-maxima* $\{M_n\}_{n\in\mathbb{N}}$ by

$$M_n = \max(X_1, \dots, X_n).$$

Show that

1. Show that $\lim_{x\to\infty} \frac{\mathbb{P}[X_1 > x]}{x^{-1}\exp(-\frac{1}{2}x^2)} = (2\pi)^{-\frac{1}{2}}$ by establishing the following inequality

   Hint: Integration by parts.

$$\frac{1}{x} \geq \frac{\mathbb{P}[X_1 > x]}{\phi(x)} \geq \frac{1}{x} - \frac{1}{x^3},\ x > 0, \tag{7.3}$$

   where, $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$ is the density of the standard normal.

2. Prove that for any $\theta \in \mathbb{R}$, $\lim_{x\to\infty} \frac{\mathbb{P}[X_1 > x + \frac{\theta}{x}]}{\mathbb{P}[X_1 > x]} = \exp(-\theta)$.
3. Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers with the property that $\mathbb{P}[X_1 > b_n] = 1/n$. Show that

$$\mathbb{P}[b_n(M_n - b_n) \leq x] \to \exp(-e^{-x}).$$

4. Show that $\lim_n \frac{b_n}{\sqrt{2\log n}} = 1$.
5. Show that $\frac{M_n}{\sqrt{2\log n}} \to 1$ in probability.