CONVERGENCE OF RANDOM VARIABLES

Contents

1 DEFINITIONS

1.1 Almost sure convergence

Definition 1. *We say that* X_n *converges to* X **almost surely** *(a.s.), and write* $X_n \stackrel{\text{a.s.}}{\to} X$*, if there is a (measurable) set* $A \subset \Omega$ *such that:*

(a) $\lim_{n\to\infty} X_n(\omega) = X(\omega)$*, for all* $\omega \in A$*;*

(b) $\mathbb{P}(A) = 1$.

Note that for a.s. convergence to be relevant, all random variables need to be defined on the same probability space (one experiment). Furthermore, the different random variables X_n are generally highly dependent.

Two common cases where a.s. convergence arises are the following.

(a) The probabilistic experiment runs over time. To each time n, we associate a nonnegative random variable Z_n (e.g., income on day n). Let $X_n = \sum_{k=1}^{n} Z_k$ be the income on the first n days. Let $X = \sum_{k=1}^{\infty} Z_k$ be the lifetime income. Note that X is well defined (as an extended real number) for every $\omega \in \Omega$, because of our assumption that $Z_k \geq 0$, and $X_n \stackrel{\text{a.s.}}{\to} X$.

(b) The various random variables are defined as different functions of a single underlying random variable. More precisely, suppose that Y is a random variable, and let $g_n : \mathbb{R} \to \mathbb{R}$ be measurable functions. Let $X_n = g_n(Y)$ [which really means, $X_n(\omega) = g_n(Y(\omega))$, for all ω]. Suppose that $\lim_{n\to\infty} g_n(y) = g(y)$ for every y. Then, $X_n \stackrel{\text{a.s.}}{\to} X$. For example, let $g_n(y) = y + y^2/n$, which converges to $g(y) = y$. We then have $Y + Y^2/n \stackrel{\text{a.s.}}{\to} Y$.

When $X_n \stackrel{\text{a.s.}}{\to} X$, we always have

$$\phi_{X_n}(t) \to \phi_X(t), \qquad \forall\, t,$$

by the dominated convergence theorem. On the other hand, the relation

$$\mathbb{E}[X_n] \to \mathbb{E}[X]$$

is not always true; sufficient conditions are provided by the monotone and dominated convergence theorems. For an example, where convergence of expectations fails to hold, consider a random variable U which is uniform on $[0, 1]$, and let:

$$X_n = \begin{cases} n, & \text{if } U \le 1/n, \\ 0, & \text{if } U > 1/n. \end{cases} \tag{1}$$

We have

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} n\mathbb{P}(U \le 1/n) = 1.$$

On the other hand, for any outcome ω for which $U(\omega) > 0$ (which happens with probability one), $X_n(\omega)$ converges to zero. Thus, $X_n \stackrel{\text{a.s.}}{\to} 0$, but $\mathbb{E}[X_n]$ does not converge to zero.

1.2 Convergence in distribution

Definition 2. *Let X and X_n, $n \in \mathbb{N}$, be random variables with CDFs F and F_n, respectively. We say that the sequence X_n converges to X* **in distribution**, *and write $X_n \stackrel{\text{d}}{\to} X$, if*

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every $x \in \mathbb{R}$ at which F is continuous.

(a) Recall that CDFs have discontinuities ("jumps") only at the points that have positive probability mass. More precisely, F is continuous at x if and only if $\mathbb{P}(X = x) = 0$.

(b) Let $X_n = 1/n$, and $X = 0$, with probability 1. Note that $F_{X_n}(0) = \mathbb{P}(X_n \leq 0) = 0$, for every n, but $F_X(0) = 1$. Still, because of the exception in the above definition, we have $X_n \stackrel{\text{d}}{\to} X$. More generally, if $X_n = a_n$ and $X = a$, with probability 1, and $a_n \to a$, then $X_n \stackrel{\text{d}}{\to} X$. Thus, convergence in distribution is consistent with the definition of convergence of real numbers. This would not have been the case if the definition required the condition $\lim_{n\to\infty} F_n(x) = F(x)$ to hold at every x.

(c) Note that this definition just involves the marginal distributions of the random variables involved. These random variables may even be defined on different probability spaces.

(d) Let Y be a random variable whose PDF is symmetric around 0. Namely, for every real value t, $\mathbb{P}(Y \leq t) = \mathbb{P}(Y \geq -t)$. Let $X_n = (-1)^n Y$. Then, every X_n has the same distribution, so, trivially, X_n converges to Y in distribution. However, for almost all ω, the sequence $X_n(\omega)$ does not converge.

(e) If we are dealing with random variables whose distribution is in a parametric class, (e.g., if every X_n is exponential with parameter λ_n), and the parameters converge (e.g., if $\lambda_n \to \lambda > 0$ and X is exponential with parameter λ), then we usually have convergence of X_n to X, in distribution. Check this for the case of exponential distributions.

(f) It is possible for a sequence of discrete random variables to converge in distribution to a continuous one. For example, if Y_n is uniform on $\{1, \ldots, n\}$ and $X_n = Y_n/n$, then X_n converges in distribution to a random variable which is uniform on $[0, 1]$ (exercise).

(g) Similarly, it is possible for a sequence of continuous random variables to converge in distribution to a discrete one. For example if X_n is uniform on $[0, 1/n]$, then X_n converges in distribution to a discrete random variable which is identically equal to zero (exercise).

(h) If X and all X_n are continuous, convergence in distribution does not imply convergence of the corresponding PDFs. (Exercise. Find an example, by emulating the example in (f).)

(i) If X and all X_n are integer-valued, convergence in distribution turns out to be equivalent to convergence of the corresponding PMFs: $p_{X_n}(k) \to p_X(k)$, for all k.(exercise).

1.3 Convergence in probability

Definition 3. *(a) We say that a sequence of random variables X_n (not necessarily defined on the same probability space) converges* **in probability** *to a real number c, and write $X_n \stackrel{\text{i.p.}}{\to} c$, if*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0, \qquad \forall\, \epsilon > 0.$$

(b) Suppose that X and X_n, $n \in \mathbb{N}$ are all defined on the same probability space. We say that the sequence X_n converges to X, in probability, and write $X_n \stackrel{\text{i.p.}}{\to} X$, if $X_n - X$ converges to zero, in probability, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0, \qquad \forall\, \epsilon > 0.$$

(a) When X in part (b) of the definition is deterministic, say equal to some constant c, then the two parts of the above definition are consistent with each other.

(b) As we will see below convergence $X_n \stackrel{\text{i.p.}}{\to} c$ is equivalent to $X_n \stackrel{\text{d}}{\to} c$.

(c) The intuitive content of the statement $X_n \stackrel{\text{i.p.}}{\to} c$ is that in the limit as n increases, almost all of the probability mass becomes concentrated in a small interval around c, no matter how small this interval is. On the other hand, for any fixed n, there can be a small probability mass outside this interval, with a slowly decaying tail. Such a tail can have a strong impact on expected values. For this reason, convergence in probability does not have any implications on expected values. See for instance the example in Eq. (1). We have $X_n \stackrel{\text{i.p.}}{\to} X$, but $\mathbb{E}[X_n]$ does not converge to $\mathbb{E}[X]$.

(d) If $X_n \stackrel{\text{i.p.}}{\to} X$ and $Y_n \stackrel{\text{i.p.}}{\to} Y$, and all random variables are defined on the same probability space, then $(X_n + Y_n) \stackrel{\text{i.p.}}{\to} (X + Y)$ (exercise).

The following is a convenient characterization, showing that convergence in probability is very closely related to almost sure convergence.

Proposition 1. *$X_n \stackrel{\text{i.p.}}{\to} X$ iff for every subsequence X_{n_k} there exists a subsubsequence $X_{n_{k_s}} \stackrel{\text{a.s.}}{\to} X$.*

2 CONVERGENCE IN DISTRIBUTION

The following result provides insights into the meaning of convergence in distribution.

Recall that the boundary ∂E of a set E is a set of simultaneous limit points of E and E^c: $\partial E \triangleq [E] \cap [E^c]$, where $[\cdot]$ denotes the closure. Also recall that quantile function q of the CDF F is a right-continuous inverse of the CDF:

$$q(s) \triangleq \inf\{x : F(x) > s\}$$

Theorem 1. *Let X_n and X be random variables, $\mathbb{P}_n$ and $\mathbb{P}$ their distributions and q_n, q their quantile functions. The following are equivalent:*

(i) $X_n \stackrel{d}{\to} X$

(ii) Quantile functions $q_n(u) \to q(u)$ *for every continuity point* u *of* q.

(iii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for every bounded continuous* f.

(iv) $\mathbb{P}_n[E] \to \mathbb{P}[E]$ *for every Borel* E *with* $\mathbb{P}[\partial E] = 0$

(v) $\limsup_{n\to\infty} \mathbb{P}_n[F] \le \mathbb{P}[F]$ *for every closed* F

(vi) $\liminf_{n\to\infty} \mathbb{P}_n[U] \ge \mathbb{P}[U]$ *for every open* U.

Note: The last four statements remain equivalent for a general metric space, in which case any of them is usually taken as *definition* of weak convergence of measures.

Proof. Equivalence of the first two follows by definition of quantiles. Indeed, in the case when F is continuous and strictly monotonically increasing this is clear. The general case follows from carefully analyzing the inclusions:

$$\{(x,y) : y < F(x)\} \subseteq \{(x,y) : q(y) \le x\} \subseteq \{(x,y) : y \le F(x)\}$$

valid for any pair of a CDF and its quantile.

Next, (ii) implies (iii), (v) and (vi) by the Theorem to follow next (Skorokhod representation), since $Y_n \stackrel{\text{a.s.}}{\to} Y$ implies $f(Y_n) \stackrel{\text{a.s.}}{\to} f(Y)$ for continuous functions and $\limsup_{n\to\infty} 1_F(Y_n) \le 1_F(Y)$. The statement of (iii) and (v) then follows by the BCT and Fatou's lemma respectively.

Furthermore, (v) and (vi) are equivalent by taking complements. To show (v) and (vi) imply (iv) let $F = [E]$ and $U = \mathrm{int}E = [E^c]^c$. Then $\partial E = F \setminus U$.

Then since $U \subseteq E \subseteq F$ we have

$$\mathbb{P}[U] \le \liminf_{n\to\infty} \mathbb{P}_n[U] \le \liminf_{n\to\infty} \mathbb{P}_n[E] \tag{2}$$

$$\le \limsup_{n\to\infty} \mathbb{P}_n[E] \le \limsup_{n\to\infty} \mathbb{P}_n[F] \le \mathbb{P}[F] \tag{3}$$

Thus when $\mathbb{P}[\partial E] = 0$ we have $\mathbb{P}[F] = \mathbb{P}[U]$ and thus $\mathbb{P}_n[E] \to \mathbb{P}[E]$.

On the other hand, (iv) implies (i) by taking $E = (-\infty, x]$ for any x – point of continuity of F. So overall we have shown:

$$\text{(i)} \iff \text{(ii)} \Rightarrow \text{(v)} \iff \text{(vi)} \Rightarrow \text{(iv)} \Rightarrow \text{(i)} \Rightarrow \text{(iii)}$$

It only remains to show that (iii) implies any of the other ones. For example, we can show (iii) ⇒ (v). To that end take

$$f_\epsilon(x) = 1 - \frac{1}{\epsilon}\min(d(x,F), \epsilon)\,,$$

where $d(x,F) = \inf_{y\in F}|x-y|$ is the minimum distance between x and F. It is easy to see $d(x,F)$ is a continuous function of x which is equal to zero only on F itself. Furthermore, $f_\epsilon \searrow 1_F$ as $\epsilon \to 0$. So we have:

$$\inf_{\epsilon>0} f_\epsilon(X_n) = 1_F(X_n) \tag{4}$$

and by the BCT

$$\inf_{\epsilon>0} \mathbb{E}[f_\epsilon(X_n)] = \mathbb{P}_n[F] \tag{5}$$

From here consider the following:

$$\limsup_{n\to\infty} \mathbb{P}_n[F] \le \inf_{\epsilon>0} \limsup_{n\to\infty} \mathbb{E}[f_\epsilon(X_n)] \tag{6}$$

$$= \inf_{\epsilon>0} \mathbb{E}[f_\epsilon(X)] \tag{7}$$

$$= \mathbb{P}[F] \tag{8}$$

where (6) follows from (5) by taking the limsup and using the usual inequality $\limsup \inf \le \inf \limsup$, (7) is by the assumption (iii) and (8) by (4)-(5) applied to $\mathbb{P}$ instead of $\mathbb{P}_n$. □

The following result shows a close relation with almost sure convergence.

Theorem 2 (Skorokhod representation)**.** *Suppose that $X_n \xrightarrow{d} X$. Then, there exists a probability space and random variables Y, Y_n defined on that space with the following properties:*

(a) For every n, the random variables X_n and Y_n have the same CDF; similarly, X and Y have the same CDF.

(b) $Y_n \xrightarrow{a.s.} Y$.

For convergence in distribution, it makes no difference whether the random variables X_n are independent or not; they do not even need to be defined on the same probability space. On the other hand, almost sure convergence implies a strong form of dependence between the random variables involved. The idea in the preceding theorem is to preserve the marginal distributions, but introduce a particular form of dependence between the X_n, which then results in almost sure convergence. This dependence is introduced by generating random variables Y_n and Y with the desired distributions, using a *common random number generator*, e.g., a single random variable U, uniformly distributed on $(0, 1)$.

Proof. Recall that if q_n is the quantile function of X_n then $q_n(U) \sim X_n$, where U is uniform on $(0, 1)$. Take $Y_n = q_n(U)$ and apply Theorem 1(ii). □

2.1 Convergence to subprobability measures: Helly's theorem

It frequently turns out to be convenient to extend the concept of convergence in distribution to cases when the limiting measure is not a probability measure. For example, we may say that $\mathbb{P}_n = \delta_n$ converges in distribution to $\mu = 0$, since the sequence of corresponding CDFs $F_n(x) = 1_{[n,\infty)}(x)$ converges to $F_0(x) = 0$ at every point of continuity. Similar to Theorem 1 we have the following equivalent representations:

Proposition 2. *Let $\mathbb{P}_n$ and μ be measures on $\mathbb{R}$ with CDFs F_n and F, respectively. The following are equivalent:*

1. *For every a, b–points of continuity of F:*

$$F_n(b) - F_n(a) \to F(b) - F(a)$$

2. *For every continuous f possessing limits at infinity $f(-\infty) = f(+\infty) = 0$:*

$$\int f d\mathbb{P}_n \to \int f d\mu$$

3. *For every bounded Borel* E *with* $\mathbb{P}[\partial E] = 0$*:*

$$\mathbb{P}_n[E] \to \mathbb{P}[E]$$

In this case we say $\mathbb{P}_n$ *converges to* μ *(weakly, or in the vague topology) and write* $\mathbb{P}_n \to \mu$*.*

Note: In the case when μ is a probability measure the above definition coincides with convergence in distribution. Note however, that $\mathbb{P}_n \to \mu$ does not imply $F_n(b) \to F(b)$ or even that this limit exists. As an example consider

$$\frac{1}{2}\delta_{n(-1)^n} + \frac{1}{2}\delta_0 \to \frac{1}{2}\delta_0$$

Theorem 3 (Helly)**.** *Any (infinite) collection of probability measures on* $(\mathbb{R}, \mathcal{B})$ *contains a sequence converging in distribution to measure* μ^* *with* $\mu^*(\mathbb{R}) \leq 1$.

Caution: Theorem does not imply that a sequence of probability measures contains a subsequence converging to a *probability* measure. Necessary and sufficient conditions for the latter will be discussed in the next Section.

Proof. Let $\{r_j, j = 1, \ldots\}$ be enumeration of rationals on $\mathbb{R}$. Let $\{\mu_s, s \in S\}$ be the collection of probability measures and F_s the respective CDFs. For each r_j the values taken by $F_s(r_j)$ belong to $[0, 1]$. By compactness of $[0, 1]$ it follows that for every j there is a sequence $s_{j,n}$ indexed by n such that

$$F_{s_{j,n}}(r_j) \to F(r_j)\,.$$

Furthermore, we may arrange the choice so that $s_{j,\cdot}$ is a subsequence of $s_{j-1,\cdot}$ etc. Then define

$$F_n \triangleq F_{s_{n,n}}$$

(Cantor's diagonal process). Since $s_{n,n}$ is a subsequence of $s_{j,\cdot}$ for every j we have

$$F_n(r_j) \to F(r_j) \qquad \forall r_j \in \mathbb{Q}\,.$$

Finally, define

$$F^*(x) = \inf_{r > x} F(r)\,.$$

One easily verifies that F^* is a right-continuous, non-decreasing function on $\mathbb{R}$ with

$$0 \leq F^*(-\infty) \leq F(+\infty) \leq 1\,.$$

Thus there is a unique measure μ^* on $(\mathbb{R}, \mathcal{B})$ so that

$$\mu^*((a,b]) = F^*(b) - F^*(a)\,.$$

The proof completes by showing that $F_n(x) \to F^*(x)$ at every point of continuity of F^*.

First, notice that for every rational r we have

$$F^*(r) \ge F(r) \triangleq \lim_{n\to\infty} F_n(r)\,.$$

Thus for every $r > x$ we have by monotonicity of F_n:

$$F^*(r) \ge \lim_{n\to\infty} F_n(r) \ge \limsup_{n\to\infty} F_n(x)$$

Taking limit as $r \searrow x$ and using right-continuity of F^* we obtain

$$F^*(x) \ge \limsup_{n\to\infty} F_n(x) \qquad \forall x \in \mathbb{R} \tag{9}$$

Conversely, for every $x_1 < x$ and some rational r between them we have

$$F^*(x_1) \le \lim_{n\to\infty} F_n(r) \le \liminf_{n\to\infty} F_n(x)$$

by monotonicity of F_n. Thus, taking the limit as $x_1 \nearrow x$ we get:

$$F^*(x-) \le \liminf_{n\to\infty} F_n(x)\,. \tag{10}$$

Together (9) and (10) establish convergence at the points of continuity since $F^*(x-) = F^*(x)$. □

2.2 Convergence to probability measures: tightness

Definition 4. *A collection of probability measures $\{\mathbb{P}_s, s \in S\}$ on $(\mathbb{R}, \mathcal{B})$ is called **tight** if for every ϵ there exists a compact set $K = [-A, A]$ such that*

$$\sup_{s\in S} \mathbb{P}_s(K^c) \le \epsilon\,.$$

In words, a collection is tight if there is no "escaping of mass to infinity", similar to the case of $\mathbb{P}_n = \delta_n$.

Theorem 4 (Prokhorov's criterion)**.** *A collection of probability measures $\{\mathbb{P}_s, s \in S\}$ on $(\mathbb{R}, \mathcal{B})$ is tight if and only if every sequence contains a subsequence converging to a probability measure.*

Proof. If collection is tight, then every sequence contains a convergent subsequence by Helly's theorem: $\mathbb{P}_n \to \mu_*$. Assuming without loss of generality that $\mu^*(\{n\} \cup \{-n\}) = 0$ (otherwise just shift these slightly) and since $\mathbb{P}_n([-n,n]^c) \le \epsilon$ we have

$$\mu_*([-n,n]^c) = \lim_{k\to\infty} \mathbb{P}_k([-n,n]^c) \le \epsilon$$

for each n and thus $\mu_*(\mathbb{R}) = 1$. Conversely, if collection is not tight, then there exist $\epsilon_0 > 0$ and measures $\mathbb{P}_n$ such that

$$\mathbb{P}_n([-n,n]^c) \ge \epsilon_0 > 0$$

for all n. If there is a subsequence $\mathbb{P}_{n_k} \to \mu_*$ then $\mu^*(\mathbb{R}) \le 1 - \epsilon_0$ and cannot be a probability measure. ☐

3 THE HIERARCHY OF CONVERGENCE CONCEPTS

Theorem 5. *We have*

$$[X_n \overset{\text{a.s.}}{\to} X] \;\Rightarrow\; [X_n \overset{\text{i.p.}}{\to} X] \;\Rightarrow\; [X_n \overset{\text{d}}{\to} X] \;\iff\; [\phi_{X_n}(t) \to \phi_X(t),\, \forall\, t].$$

(The first two implications assume that all random variables be defined on the same probability space.)

Proof:

(a) $[X_n \overset{\text{a.s.}}{\to} X] \;\Rightarrow\; [X_n \overset{\text{i.p.}}{\to} X]$:
We give a short proof, based on the DCT, but more elementary proofs are also possible. Fix some $\epsilon > 0$. Let

$$Y_n = I_{\{|X_n - X| \ge \epsilon\}}.$$

If $X_n \overset{\text{a.s.}}{\to} X$, then $Y_n \overset{\text{a.s.}}{\to} 0$. By the DCT, $\mathbb{E}[Y_n] \to 0$. On the other hand,

$$\mathbb{E}[Y_n] = \mathbb{P}\big(X_n - X| \ge \epsilon\big).$$

This implies that $\mathbb{P}\big(|X_n - X| \ge \epsilon\big) \to 0$, and therefore, $X_n \overset{\text{i.p.}}{\to} X$.

(b) $[X_n \overset{\text{i.p.}}{\to} X] \;\Rightarrow\; [X_n \overset{\text{d}}{\to} X]$:
Since the magnitude of derivative of the function $a \mapsto \cos(ta)$ is bounded by t, we have that

$$|\cos(ta) - \cos(tb)| \le t|a - b|\,.$$

Notice, however, that when $t|a-b| > 2$ this bound is not good, so overall we get:

$$|\cos(ta) - \cos(tb)| \leq \left\{ t\epsilon, \quad |a-b| \leq 2\epsilon/t\,, 2, |a-b| > 2\epsilon/t \right.$$

Using this with $a = X_n$ and $b = X$ and taking the expectation we get

$$\mathbb{E}[\cos(tX_n) - \cos(tX)] \leq t\epsilon\mathbb{P}[|X_n - X| \leq \epsilon/t] + 2\mathbb{P}[|X_n - X| > 2\epsilon/t]\,.$$

The second term converges to zero as $n \to \infty$ for any t, ϵ, whereas the first term is bounded by $t\epsilon$. Thus, first taking $\lim_{n\to\infty}$ and then $\lim_{\epsilon\to 0}$ we obtain

$$\mathbb{E}[\cos(tX_n)] \to \mathbb{E}[\cos(tX)]$$

for every $t \in \mathbb{R}$. Similar proof shows

$$\mathbb{E}[\sin(tX_n)] \to \mathbb{E}[\sin(tX)]\,.$$

Thus, characteristic functions $\phi_{X_n} \to \phi_X$ and from the last part we get the claimed result.

(c) $[X_n \overset{\text{d}}{\to} X] \;\Rightarrow\; [\phi_{X_n}(t) \to \phi_X(t),\ \forall\, t]$:
Suppose that $X_n \overset{\text{d}}{\to} X$. Let Y_n and Y be as in Theorem 2, so that $Y_n \overset{\text{a.s.}}{\to} Y$. Then, for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \phi_{X_n}(t) = \lim_{n\to\infty} \phi_{Y_n}(t) = \lim_{n\to\infty} \mathbb{E}[e^{itY_n}] = \mathbb{E}[e^{itY}] = \phi_Y(t) = \phi_X(t),$$

where we have made use of the facts $Y_n \overset{\text{a.s.}}{\to} Y$, $e^{itY_n} \overset{\text{a.s.}}{\to} e^{itY}$, and the DCT. Finally, the converse direction will be established in the next lecture. □

Exercise 1 (Smoothing method)**.** Show that for every $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B})$ there exist a sequence $\mathbb{P}_{X_n} \overset{\text{d}}{\to} \mathbb{P}_X$ such that every $\mathbb{P}_{X_n}$ has continuous, bounded, infinitely-differentiable PDF. Steps:

1. Show $X_\epsilon = X + \epsilon Z \overset{\text{d}}{\to} X$ as $\epsilon \to 0$.
2. Let $X \perp\!\!\!\perp Z \sim \mathcal{N}(0,1)$ and show that CDF of X_ϵ is continuous (*Hint:* BCT) and differentiable (*Hint: Fubini*) with derivative

$$f_{X_\epsilon}(a) = \mathbb{E}\left[f_Z\left(\frac{a-X}{\epsilon}\right)\frac{1}{\epsilon}\right]$$

3. Show that $a \mapsto f_{X_\epsilon}(a)$ is continuous.
4. Conclude the proof (*Hint:* derivatives of f_Z are uniformly bounded on $\mathbb{R}$.)

At this point, it is natural to ask whether the converses of the implications in Theorem 5 hold. For the first two, the answer is, in general, “no”, although we will also note some exceptions.

3.1 Convergence almost surely versus in probability

$[X_n \stackrel{\text{i.p.}}{\to} X]$ **does not imply** $[X_n \stackrel{\text{a.s.}}{\to} X]$:
Let X_n be equal to 1, with probability $1/n$, and equal to zero otherwise. Suppose that the X_n are independent. We have $X_n \stackrel{\text{i.p.}}{\to} 0$. On the other hand, by the Borel-Cantelli lemma, the event $\{X_n = 1, \text{ i.o}\}$ has probability 1 (check this). Thus, for almost all ω, the sequence $X_n(\omega)$ does not converge to zero.

Nevertheless, a weaker form of the converse implication turns out to be true. If $X_n \stackrel{\text{i.p.}}{\to} X$, then there exists an increasing (deterministic) sequence n_k of integers, such that $\lim_{k\to\infty} X_{n_k} = X$, a.s. (We omit the proof.)

For an illustration of the last statement in action, consider the preceding counterexample. If we let $n_k = k^2$, then we note that $\mathbb{P}(X_{n_k} \neq 0) = 1/k^2$, which is summable. By the Borel-Cantelli lemma, the event $\{X_{n_k} \neq 0\}$ will occur for only finitely many k, with probability 1. Therefore, X_{n_k} converges, a.s., to the zero random variable.

3.2 Convergence in probability versus in distribution

The converse turns out to be false in general, but true when the limit is deterministic.
$[X_n \stackrel{\text{d}}{\to} X]$ **does not imply** $[X_n \stackrel{\text{i.p.}}{\to} X]$:
Let the random variables X, X_n be i.i.d. and nonconstant random variables, in which case we have (trivially) $X_n \stackrel{\text{d}}{\to} X$. Fix some ϵ. Then, $\mathbb{P}(|X_n - X| \geq \epsilon)$ is positive and the same for all n, which shows that X_n does not converge to X, in probability.
$[X_n \stackrel{\text{d}}{\to} c]$ **implies** $[X_n \stackrel{\text{i.p.}}{\to} c]$:
The proof is very simple: by definition we have

$$\mathbb{P}[X_n \leq c - \epsilon] \to 0\,, \qquad \mathbb{P}[X_n > c + \epsilon] \to 0$$

for any $\epsilon > 0$. Thus

$$\mathbb{P}[|X_n - c| > \epsilon] \to 0$$

for any ϵ as well.