Chapter 3

Strong convergence

As pointed out in the Chapter 2, there are multiple ways to define the notion of convergence of a sequence of random variables. That chapter defined convergence in probability, convergence in distribution, and convergence in $k$th mean. We now consider a fourth mode of convergence, almost sure convergence or convergence with probability one, which will be shown to imply both convergence in probability and convergence in distribution. It is for this reason that we attach the term “strong” to almost sure convergence and “weak” to the other two; these terms are not meant to indicate anything about their usefulness. In fact, the “weak” modes of convergence are used much more frequently in asymptotic statistics than the strong mode, and thus a reader new to the subject may wish to skip this chapter; most of the rest of the book may be understood without a grasp of strong convergence.

3.1 Definition of almost sure convergence

This form of convergence is in some sense the simplest to understand, since it depends only on the concept of limit of a sequence of real numbers, Definition 1.1. Since a random variable like $X_n$ or $X$ is a function on a sample space (say, $\Omega$), if we fix a particular element of that space (say, $\omega$), we obtain the real numbers $X_n(\omega)$ and $X(\omega)$. We may then ask, for each $\omega \in \Omega$, whether $X_n(\omega)$ converges to $X(\omega)$ as a sequence of real numbers.

Definition 3.1 Suppose $X$ and $X_1, X_2, \ldots$ are random variables defined on the same sample space $\Omega$ (and as usual $P$ denotes the associated probability measure). If

$$P \left( \{ \omega \in \Omega : X_n(\omega) \to X(\omega) \} \right) = 1,$$

then $X_n$ is said to converge almost surely (or with probability one) to $X$, denoted $X_n \overset{\text{a.s.}}{\to} X$ or $X_n \to X$ a.s. or $X_n \to X$ w.p. 1.
In other words, convergence with probability one means exactly what it sounds like: The probability that $X_n$ converges to $X$ equals one.

Later, in Theorem 3.3, we will formulate an equivalent definition of almost sure convergence that makes it much easier to see why it is such a strong form of convergence of random variables. Yet the intuitive simplicity of Definition 3.1 makes it the standard definition.

As in the case of convergence in probability, we may replace the limiting random variable $X$ by any constant $c$, in which case we write $X_n \overset{a.s.}{\to} c$. In the most common statistical usage of convergence to a constant, the random variable $X_n$ is some estimator of a particular parameter, say $g(\theta)$:

**Definition 3.2** If $X_n \overset{a.s.}{\to} g(\theta)$, $X_n$ is said to be **strongly consistent** for $g(\theta)$.

As the names suggest, strong consistency implies consistency, a fact to be explored in more depth below.

### 3.1.1 Strong Consistency versus Consistency

As before, suppose that $X$ and $X_1, X_2, \ldots$ are random variables defined on the same sample space, $\Omega$. For given $n$ and $\epsilon > 0$, define the events

$$A_n = \{ \omega \in \Omega : |X_k(\omega) - X(\omega)| < \epsilon \text{ for all } k \geq n \}$$

(3.1) and

$$B_n = \{ \omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon \}.$$ 

(3.2)

Evidently, both $A_n$ and $B_n$ occur with high probability when $X_n$ is in some sense close to $X$, so it might be reasonable to say that $X_n$ converges to $X$ if $P(A_n) \to 1$ or $P(B_n) \to 1$ for any $\epsilon > 0$. In fact, Definition 2.1 is nothing more than the latter statement; that is, $X_n \overset{P}{\to} X$ if and only if $P(B_n) \to 1$ for any $\epsilon > 0$.

Yet what about the sets $A_n$? One fact is immediate: Since $A_n \subset B_n$, we must have $P(A_n) \leq P(B_n)$. Therefore, $P(A_n) \to 1$ implies $P(B_n) \to 1$. By now, the reader might already have guessed that $P(A_n) \to 1$ for all $\epsilon$ is equivalent to $X_n \overset{a.s.}{\to} X$:

**Theorem 3.3** With $A_n$ defined as in Equation (3.1), $P(A_n) \to 1$ for any $\epsilon > 0$ if and only if $X_n \overset{a.s.}{\to} X$.

The proof of Theorem 3.3 is the subject of Exercise 3.1. The following corollary now follows from the preceding discussion:

**Corollary 3.4** If $X_n \overset{a.s.}{\to} X$, then $X_n \overset{P}{\to} X$. 

51
The converse of Corollary 3.4 is not true, as the following example illustrates.

**Example 3.5** Take $\Omega$ to be the half-open interval $(0, 1]$, and for any interval $J \subset \Omega$, say $J = (a, b]$, take $P(J) = b - a$ to be the length of that interval. Define a sequence of intervals $J_1, J_2, \ldots$ as follows:

$$J_1 = (0, 1]$$

$J_2$ through $J_4 = (0, \frac{1}{2}], (\frac{1}{2}, \frac{3}{4}], (\frac{3}{4}, 1]$

$J_5$ through $J_9 = (0, \frac{1}{3}], (\frac{1}{3}, \frac{2}{3}], (\frac{2}{3}, \frac{4}{5}], (\frac{4}{5}, 1]$

$$\vdots$$

$J_{m^2+1}$ through $J_{(m+1)^2} = \left(0, \frac{1}{2m+1}\right], \ldots, \left(\frac{2m}{2m+1}, 1\right]$

Note in particular that $P(J_n) = 1/(2m + 1)$, where $m = \lfloor \sqrt{n - 1} \rfloor$ is the largest integer not greater than $\sqrt{n - 1}$. Now, define $X_n = I\{J_n\}$ and take $0 < \epsilon < 1$. Then $P(|X_n - 0| < \epsilon)$ is the same as $1 - P(J_n)$. Since $P(J_n) \to 0$, we conclude $X_n \overset{P}{\to} 0$ by definition.

However, it is *not* true that $X_n \overset{a.s.}{\to} 0$. Since every $\omega \in \Omega$ is contained in infinitely many $J_n$, the set $A_n$ defined in Equation (3.1) is empty for all $n$. Alternatively, consider the set $S = \{\omega : X_n(\omega) \to 0\}$. For any $\omega$, $X_n(\omega)$ has no limit because $X_n(\omega) = 1$ and $X_n(\omega) = 0$ both occur for infinitely many $n$. Thus $S$ is empty. This is not convergence with probability one; it is convergence with probability zero!

### 3.1.2 Multivariate Extensions

We may extend Definition 3.1 to the multivariate case in a completely straightforward way:

**Definition 3.6** $X_n$ is said to converge almost surely (or with probability one) to $X$ ($X_n \overset{a.s.}{\to} X$) if

$$P\left(X_n \to X \text{ as } n \to \infty\right) = 1.$$ 

Alternatively, since the proof of Theorem 3.3 applies to random vectors as well as random variables, we say $X_n \overset{a.s.}{\to} X$ if for any $\epsilon > 0$,

$$P\left(\|X_k - X\| < \epsilon \text{ for all } k \geq n\right) \to 1 \text{ as } n \to \infty.$$ 

52
We saw in Theorems 2.24 and 2.30 that continuous functions preserve both convergence in probability and convergence in distribution. Yet these facts were quite difficult to prove. Fortunately, the analogous result for convergence with probability one is quite easy to prove. In fact, since almost sure convergence is defined in terms of convergence of sequences of real (not random) vectors, the following theorem may be proven using the same method of proof used for Theorem 1.16.

**Theorem 3.7** Suppose that $f : S \to \mathbb{R}^\ell$ is a continuous function defined on some subset $S \subset \mathbb{R}^k$, $X_n$ is a $k$-component random vector, and $P(X \in S) = 1$. If $X_n \overset{\text{a.s.}}{\to} X$, then $f(X_n) \overset{\text{a.s.}}{\to} f(X)$.

We conclude this section with a simple diagram summarizing the implications among the modes of convergence defined so far. In the diagram, a double arrow like $\Rightarrow$ means “implies”. Note that the picture changes slightly when convergence is to a constant $c$ rather than a random vector $X$.

\[
X_n \overset{\text{qm}}{\to} X \quad \Rightarrow \quad X_n \overset{\text{Ps}}{\to} X \quad \Rightarrow \quad X_n \overset{d}{\to} X \quad \Rightarrow \quad X_n \overset{\text{a.s.}}{\to} X \quad \Rightarrow \quad X_n \overset{\text{Ps}}{\to} c \quad \Rightarrow \quad X_n \overset{d}{\to} c
\]

**Exercises for Section 3.1**

**Exercise 3.1** Prove Theorem 3.3.

**Hint:** Note that the sets $A_n$ are increasing in $n$, so that by the lower continuity of any probability measure (which you may assume without proof), $\lim_n P(A_n)$ exists and is equal to $P(\bigcup_{n=1}^\infty A_n)$.

**Exercise 3.2** Prove Theorem 3.7.

### 3.2 The Strong Law of Large Numbers

Some of the results in this section are presented for univariate random variables and some are presented for random vectors. Take note of the use of bold print to denote vectors.

**Theorem 3.8** *Strong Law of Large Numbers:* Suppose that $X_1, X_2, \ldots$ are independent and identically distributed and have finite mean $\mu$. Then $\overline{X}_n \overset{\text{a.s.}}{\to} \mu$. 

53
It is possible to use fairly simple arguments to prove a version of the Strong Law under more restrictive assumptions than those given above. See Exercise 3.4 for details of a proof of the univariate Strong Law under the additional assumption that $X_n^4 < \infty$. To aid the proof of the Strong Law in its full generality, we first establish a useful lemma.

**Lemma 3.9** If $\sum_{k=1}^{\infty} P(\|X_k - X\| > \epsilon) < \infty$ for any $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$.

**Proof:** The proof relies on the *countable subadditivity* of any probability measure, an axiom stating that for any sequence $A_1, A_2, \ldots$ of events,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{i=k}^{\infty} P(A_k). \quad (3.3)$$

To prove the lemma, we must demonstrate that $P(\|X_k - X\| \leq \epsilon$ for all $k \geq n) \to 1$ as $n \to \infty$, which (taking complements) is equivalent to $P(\|X_k - X\| > \epsilon$ for some $k \geq n) \to 0$. Letting $A_k$ denote the event that $\|X_k - X\| > \epsilon$, countable subadditivity implies

$$P(\text{for some } k \geq n) = P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k),$$

and the right hand side tends to 0 as $n \to \infty$ because it is the tail of a convergent series. ■

Lemma 3.9 is nearly the same as a famous result called the First Borel-Cantelli Lemma, or sometimes simply the Borel-Cantelli Lemma; see Exercise 3.3. The utility of Lemma 3.9 is illustrated by the following useful result, which allows us to relate almost sure convergence to convergence in probability (see Theorem 2.24, for instance).

**Theorem 3.10** $X_n \xrightarrow{P} X$ if and only if each subsequence $X_{n_1}, X_{n_2}, \ldots$ contains a further subsequence that converges almost surely to $X$.

The proof of Theorem 3.10, which uses Lemma 3.9, is the subject of Exercise 3.7.

### 3.2.1 Independent but not identically distributed variables

Here, we generalize the univariate version of the Strong Law to a situation in which the $X_n$ are assumed to be independent and satisfy a second moment condition:

**Theorem 3.11** *Kolmogorov’s Strong Law of Large Numbers:* Suppose that $X_1, X_2, \ldots$ are independent with mean $\mu$ and

$$\sum_{i=1}^{\infty} \frac{\text{Var } X_i}{i^2} < \infty.$$

Then $\overline{X}_n \xrightarrow{a.s.} \mu$.  

54
Note that there is no reason the $X_i$ in Theorem 3.11 must have the same means: If $E X_i = \mu_i$, then the theorem as written implies that $(1/n) \sum_i (X_i - \mu_i) \overset{a.s.}{\to} 0$.

Theorem 3.11 may be proved using Kolmogorov’s inequality from Exercise 1.31; this proof is the focus of Exercise 3.6. In fact, Theorem 3.11 turns out to be very important because it may be used to prove the Strong Law, Theorem 3.8. The key to completing this proof is to introduce truncated versions of $X_1, X_2, \ldots$ as in the following lemma.

**Lemma 3.12** Suppose that $X_1, X_2, \ldots$ are independent and identically distributed and have finite mean $\mu$. Define $X_i^* = X_i I\{|X_i| \leq i\}$. Then

$$\sum_{i=1}^{\infty} \frac{\text{Var} X_i^*}{i^2} < \infty \quad (3.4)$$

and $\overline{X}_n - \overline{X}_n^* \overset{a.s.}{\to} 0$.

Under the assumptions of Lemma 3.12, we see immediately that $\overline{X}_n = \overline{X}_n^* + (\overline{X}_n - \overline{X}_n^*) \overset{a.s.}{\to} \mu$, because Equation (3.4) implies $\overline{X}_n^* \overset{a.s.}{\to} \mu$ by Theorem 3.11. This proves the univariate version of Theorem 3.8; the full multivariate version follows because $X_n \overset{a.s.}{\to} \mu$ if and only if $X_{nj} \overset{a.s.}{\to} \mu_j$ for all $j$ (Lemma 1.31).

A proof of Lemma 3.12 is the subject of Exercise 3.5.

**Exercises for Section 3.2**

**Exercise 3.3** Let $B_1, B_2, \ldots$ denote a sequence of events. Let $B_n$ i.o., which stands for $B_n$ infinitely often, denote the set

$$B_n \overset{\text{def}}{=} \{ \omega \in \Omega : \text{for every } n, \text{ there exists } k \geq n \text{ such that } \omega \in B_k \}.$$ 

Prove the first Borel-Cantelli Lemma, which states that if $\sum_{n=1}^{\infty} P(B_n) < \infty$, then $P(B_n \text{ i.o.)} = 0$.

**Hint:** Argue that

$$B_n \overset{\text{i.o.}}{=} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k,$$

then adapt the proof of Lemma 3.9.

**Exercise 3.4** Use the hint below to prove that if $X_1, X_2, \ldots$ are independent and identically distributed and $E X_1^4 < \infty$, then $\overline{X}_n \overset{a.s.}{\to} E X_1$. You may assume without loss of generality that $E X_1 = 0$. 

55
**Hint:** Use Markov’s inequality (1.22) with \( r = 4 \) to put an upper bound on

\[ P\left( |\bar{X}_n| > \epsilon \right) \]

involving \( E\left( X_1 + \ldots + X_n \right)^4 \). Expand \( E\left( X_1 + \ldots + X_n \right)^4 \) and then count the nonzero terms. Finally, argue that the conditions of Lemma 3.9 are satisfied.

**Exercise 3.5** Lemma 3.12 makes two assertions about the random variables \( X_i^* = X_i I\{|X_i| \leq i\} \):

(a) Prove that

\[ \sum_{i=1}^{\infty} \frac{\text{Var} \ X_i^*}{i^2} < \infty. \]

**Hint:** Use the fact that the \( X_i \) are independent and identically distributed, then show that

\[ X_i^2 \sum_{k=1}^{\infty} \frac{1}{k^2} I\{|X_1| \leq k\} \leq 2|X_1|, \]

perhaps by bounding the sum on the left by an easy-to-evaluate integral.

(b) Prove that \( \bar{X}_n - \bar{X}_n^* \xrightarrow{a.s.} 0 \).

**Hint:** Use Lemma 3.9 and Exercise 1.32 to show that \( \bar{X}_n - \bar{X}_n^* \xrightarrow{a.s.} 0 \). Then use Exercise 1.3.

**Exercise 3.6** Prove Theorem 3.11. Use the following steps:

(a) For \( k = 1, 2, \ldots \), define

\[ Y_k = \max_{2^{k-1} \leq n < 2^k} |\bar{X}_n - \mu|. \]

Use the Kolmogorov inequality from Exercise 1.31 to show that

\[ P(Y_k \geq \epsilon) \leq \frac{4 \sum_{i=1}^{2^k} \text{Var} \ X_i}{4^k \epsilon^2}. \]

(b) Use Lemma 3.9 to show that \( Y_k \xrightarrow{a.s.} 0 \), then argue that this proves \( \bar{X}_n \xrightarrow{a.s.} \mu \).

**Hint:** Letting \([\log_2 i]\) denote the smallest integer greater than or equal to \( \log_2 i \) (the base-2 logarithm of \( i \)), verify and use the fact that

\[ \sum_{k=[\log_2 i]}^{\infty} \frac{1}{4^k} \leq \frac{4}{3i^2}. \]
Exercise 3.7  Prove Theorem 3.10.

**Hint:** To simplify notation, let $Y_k = X_{n_k}$ denote an arbitrary subsequence. If $P(Y_k \overset{p}{\to} X)$, Show that there exist $k_1, k_2, \ldots$ such that

$$P(\|Y_{k_j} - X\| > \epsilon) < \frac{1}{2^j},$$

then use Lemma 3.9.

On the other hand, if $X_n$ does not converge in probability to $X$, argue that there exists a subsequence $Y_1 = X_{n_1}, Y_2 = X_{n_2}, \ldots$ and $\epsilon > 0$ such that

$$P(\|Y_k - X\| > \epsilon) > \epsilon$$

for all $k$. Then use Corollary 3.4 to argue that $Y_n$ does not have a subsequence that converges almost surely.

### 3.3 The Dominated Convergence Theorem

We now consider the question of when $Y_n \overset{d}{\to} Y$ implies $E Y_n \to E Y$. This is not generally the case: Consider contaminated normal distributions with distribution functions

$$F_n(x) = \left(1 - \frac{1}{n}\right) \Phi(x) + \frac{1}{n} \Phi(x - 37n). \quad (3.5)$$

These distributions converge in distribution to the standard normal $\Phi(x)$, yet each has mean 37.

However, recall Theorem 2.25, which guarantees that $Y_n \overset{d}{\to} Y$ implies $E Y_n \to E Y$ if all of the $Y_n$ and $Y$ are uniformly bounded — say, $|Y_n| < M$ and $|Y| < M$ — since in that case, there is a bounded, continuous function $g(y)$ for which $g(Y_n) = Y_n$ and $g(Y) = Y$: Simply define $g(y) = y$ for $-M < y < M$, and $g(y) = My/|y|$ otherwise.

To say that the $Y_n$ are *uniformly* bounded is a much stronger statement than saying that each $Y_n$ is bounded. The latter statement implies that the bound we choose is allowed to depend on $n$, whereas the uniform bound means that the same bound must apply to all $Y_n$.

When there are only finitely many $Y_n$, then boundedness implies uniform boundedness since we may take as a uniform bound the maximum of the bounds of the individual $Y_n$. However, in the case of an infinite sequence of $Y_n$, the maximum of an infinite set of individual bounds might not exist.
The intuition above, then, is that some sort of uniform bound on the \( Y_n \) should be enough to guarantee \( EY_n \to EY \). The most common way to express this idea is the Dominated Convergence Theorem, given later in this section as Theorem 3.17.

The proof of the Dominated Convergence Theorem that we give here relies on a powerful technique that is often useful for proving results about convergence in distribution. This technique is called the Skorohod Representation Theorem, which guarantees that convergence in distribution implies almost sure convergence for a possibly different sequence of random variables. More precisely, if we know \( X_n \overset{d}{\to} X \), the Skorohod Representation Theorem guarantees the existence of \( Y_n \overset{d}{=} X_n \) and \( Y \overset{d}{=} X \) such that \( Y_n \overset{a.s.}{\to} Y \), where \( \overset{d}{=} \) means “has the same distribution as”.

Construction of such \( Y_n \) and \( Y \) will depend upon inverting the distribution functions of \( X_n \) and \( X \). However, since not all distribution functions are invertible, we first generalize the notion of the inverse of a distribution function by defining the quantile function.

**Definition 3.13** If \( F(x) \) is a distribution function, then we define the quantile function \( F^- : (0, 1) \to \mathbb{R} \) by

\[
F^-(u) \overset{\text{def}}{=} \inf\{x \in \mathbb{R} : u \leq F(x)\}.
\]

With the quantile function thus defined, we may prove a useful lemma:

**Lemma 3.14** \( u \leq F(x) \) if and only if \( F^-(u) \leq x \).

**Proof:** Using the facts that \( F \) and \( F^- \) are nondecreasing and \( F^-[F(x)] \leq x \),

\[
\begin{align*}
u \leq F(x) & \quad \Rightarrow \quad F^-(u) \leq F^-[F(x)] \leq x \\
& \quad \Rightarrow \quad F[F^-(u)] \leq F(x) \\
& \quad \Rightarrow \quad u \leq F(x),
\end{align*}
\]

where the first implication follows because \( F^-[F(x)] \leq x \) and the last follows because \( u \leq F[F^-(u)] \) (the latter fact requires right-continuity of \( F \)).

Now the construction of \( Y_n \) and \( Y \) proceeds as follows. Let \( F_n \) and \( F \) denote the distribution functions of \( X_n \) and \( X \), respectively, for all \( n \). Take the sample space \( \Omega \) to be the interval \((0, 1)\) and adopt the probability measure that assigns to each interval subset \((a, b) \subset (0, 1)\) its length \((b - a)\). (There is a unique probability measure on \((0, 1)\) with this property, a fact we do not prove here.) Then for every \( \omega \in \Omega \), define

\[
Y_n(\omega) \overset{\text{def}}{=} F_n^-(\omega) \quad \text{and} \quad Y(\omega) \overset{\text{def}}{=} F^-(\omega).
\] (3.6)

The random variables \( Y_n \) and \( Y \) are exactly the random variables we need, as asserted in the following theorem.
**Theorem 3.15**  *Skorohod representation theorem:* Assume $F_n \overset{d}{\to} F$. The random variables defined in expression (3.6) satisfy

1. $P(Y_n \leq t) = F_n(t)$ for all $n$ and $P(Y \leq t) = F(t)$;
2. $Y_n \overset{a.s.}{\to} Y$.

Before proving Theorem 3.15, we first state a technical lemma, a proof of which is the subject of Exercise 3.8(a).

**Lemma 3.16** Assume $F_n \overset{d}{\to} F$ and let the random variables $Y_n$ and $Y$ be defined as in expression (3.6). Then for any $\omega \in (0, 1)$ and any $\epsilon > 0$ such that $\omega + \epsilon < 1$,

$$Y(\omega) \leq \liminf_n Y_n(\omega) \leq \limsup_n Y_n(\omega) \leq Y(\omega + \epsilon).$$  (3.7)

**Proof of Theorem 3.15:** By Lemma 3.14, $Y \leq t$ if and only if $\omega \leq F(t)$. But $P(\omega \leq F(t)) = F(t)$ by construction. A similar argument for $Y_n$ proves the first part of the theorem.

For the second part of the theorem, letting $\epsilon \to 0$ in inequality (3.7) shows that $Y_n(\omega) \to Y(\omega)$ whenever $\omega$ is a point of continuity of $Y(\omega)$. Since $Y(\omega)$ is a nondecreasing function of $\omega$, there are at most countably many points of discontinuity of $\omega$; see Exercise 3.8(b). Let $D$ denote the set of all points of discontinuity of $Y(\omega)$. Since each individual point in $\Omega$ has probability zero, the countable subadditivity property (3.3) implies that $P(D) = 0$. Since we have shown that $Y_n(\omega) \to Y(\omega)$ for all $\omega \not\in D$, we conclude that $Y_n \overset{a.s.}{\to} Y$. 

Note that the points of discontinuity of $Y(\omega)$ mentioned in the proof of Theorem 3.15 are not in any way related to the points of discontinuity of $F(x)$. In fact, “flat spots” of $F(x)$ lead to discontinuities of $Y(\omega)$ and vice versa.

Having thus established the Skorohod Represenation Theorem, we now introduce the Dominated Convergence Theorem.

**Theorem 3.17**  *Dominated Convergence Theorem:* If for some random variable $Z$, $|X_n| \leq |Z|$ for all $n$ and $E|Z| < \infty$, then $X_n \overset{d}{\to} X$ implies that $E X_n \to E X$.

**Proof:** Fatou’s Lemma (see Exercise 3.9) states that

$$E \liminf_n |X_n| \leq \liminf_n E |X_n|. \quad (3.8)$$

A second application of Fatou’s Lemma to the nonnegative random variables $|Z| - |X_n|$ implies

$$E |Z| - E \limsup_n |X_n| \leq E |Z| - \limsup_n E |X_n|.$$
Because $E|Z| < \infty$, subtracting $E|Z|$ preserves the inequality, so we obtain

$$\limsup_n E|X_n| \leq E\limsup_n |X_n|.$$  \hfill (3.9)

Together, inequalities (3.8) and (3.9) imply

$$E\liminf_n |X_n| \leq \liminf_n E|X_n| \leq \limsup_n E|X_n| \leq E\limsup_n |X_n|.$$ 

Therefore, the proof would be complete if $|X_n| \xrightarrow{a.s.} |X|$. This is where we invoke the Skorohod Representation Theorem: Because there exists a sequence $Y_n$ that does converge almost surely to $Y$, having the same distributions and expectations as $X_n$ and $X$, the above argument shows that $EY_n \rightarrow EY$, hence $EX_n \rightarrow EX$, completing the proof. 

**Exercises for Section 3.3**

**Exercise 3.8** This exercise proves two results used to establish theorem 3.15.

(a) Prove Lemma 3.16.

**Hint:** For any $\delta > 0$, let $x$ be a continuity point of $F(t)$ in the interval $(Y(\omega) - \delta, Y(\omega))$. Use the fact that $F_n \xrightarrow{d} F$ to argue that for large $n$, $Y(\omega) - \delta < Y_n(\omega)$. Take the limit inferior of each side and note that $\delta$ is arbitrary. Similarly, argue that for large $n$, $Y_n(\omega) < Y(\omega + \epsilon) + \delta$.

(b) Prove that any nondecreasing function has at most countably many points of discontinuity.

**Hint:** If $x$ is a point of discontinuity, consider the open interval whose endpoints are the left- and right-sided limits at $x$. Note that each such interval contains a rational number, of which there are only countably many.

**Exercise 3.9** Prove Fatou’s lemma:

$$E\liminf_n |X_n| \leq \liminf_n E|X_n|.$$  \hfill (3.10)

**Hint:** Argue that $E|X_n| \geq E\inf_{k \geq n} |X_k|$, then take the limit inferior of each side. Use the monotone convergence property on page 25.

**Exercise 3.10** If $Y_n \xrightarrow{d} Y$, a sufficient condition for $EY_n \rightarrow EY$ is the **uniform integrability** of the $Y_n$.
**Definition 3.18** The random variables $Y_1, Y_2, \ldots$ are said to be uniformly integrable if

$$\sup_n E (|Y_n| I\{|Y_n| \geq \alpha\}) \to 0 \text{ as } \alpha \to \infty.$$ 

Prove that if $Y_n \xrightarrow{d} Y$ and the $Y_n$ are uniformly integrable, then $E Y_n \to E Y$.

**Exercise 3.11** Prove that if there exists $\epsilon > 0$ such that $\sup_n E |Y_n|^{1+\epsilon} < \infty$, then the $Y_n$ are uniformly integrable.

**Exercise 3.12** Prove that if there exists a random variable $Z$ such that $E |Z| = \mu < \infty$ and $P(|Y_n| \geq t) \leq P(|Z| \geq t)$ for all $n$ and for all $t > 0$, then the $Y_n$ are uniformly integrable. You may use the fact (without proof) that for a nonnegative $X$,

$$E(X) = \int_0^\infty P(X \geq t) \, dt.$$ 

**Hints:** Consider the random variables $|Y_n| I\{|Y_n| \geq t\}$ and $|Z| I\{|Z| \geq t\}$. In addition, use the fact that

$$E |Z| = \sum_{i=1}^\infty E (|Z| I\{|Z| \leq i\})$$

to argue that $E (|Z| I\{|Z| < \alpha\}) \to E |Z|$ as $\alpha \to \infty$. 