5. Convergence of sequences of random variables

Throughout this chapter we assume that $\{X_1, X_2, \ldots\}$ is a sequence of r.v. and X is a r.v., and all of them are defined on the same probability space $(\Omega, \mathcal{F}, P)$.

Stochastic convergence formalizes the idea that a sequence of r.v. sometimes is expected to settle into a pattern.[1] The pattern may for instance be that:

- there is a convergence of $X_n(\omega)$ in the classical sense to a fixed value $X(\omega)$, for each and every event ω;
- the probability that the distance between X_n from a particular r.v. X exceeds any prescribed positive value decreases and converges to zero;
- the series formed by calculating the expected value of the (absolute or quadratic) distance between X_n and X converges to zero;
- the distribution of X_n may "grow" increasingly similar to the distribution of a particular r.v. X.

Just as in analysis, we can distinguish among several types of convergence (Rohatgi, 1976, p. 240). Thus, in this chapter we investigate modes of convergence of sequences of r.v.:

- almost sure convergence ($\stackrel{a.s.}{\to}$);
- convergence in probability ($\stackrel{P}{\to}$);
- convergence in quadratic mean or in L^2 ($\stackrel{q.m.}{\to}$);
- convergence in L^1 or in mean ($\stackrel{L^1}{\to}$);

[1] See http://en.wikipedia.org/wiki/Convergence_of_random_variables.

- convergence in distribution ($\stackrel{d}{\rightarrow}$).

Two laws of large numbers and central limit theorems are also stated.

It is important for the reader to be familiarized with all these modes of convergence, the way they can be related and with the applications of such results and understand their considerable significance in probability, statistics and stochastic processes.

5.1 Modes of convergence

The first four modes of convergence ($\stackrel{*}{\rightarrow}$, where $* = a.s., P, q.m., L^1$) pertain to the sequence of r.v. and to X as functions of Ω, while the fifth ($\stackrel{d}{\rightarrow}$) is related to the convergence of d.f. (Karr, 1993, p. 135).

5.1.1 Convergence of r.v. as functions on Ω

Motivation 5.1 — Almost sure convergence (Karr, 1993, p. 135)
Almost sure convergence — or convergence with probability one — is the probabilistic version of pointwise convergence known from elementary real analysis. •

Definition 5.2 — Almost sure convergence (Karr, 1993, p. 135; Rohatgi, 1976, p. 249)
The sequence of r.v. $\{X_1, X_2, \ldots\}$ is said to converge almost surely to a r.v. X if

$$P\left(\left\{w : \lim_{n\rightarrow+\infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \tag{5.1}$$

In this case we write $X_n \stackrel{a.s.}{\rightarrow} X$ (or $X_n \rightarrow X$ with probability 1). •

Exercise 5.3 — Almost sure convergence
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that $X_n \sim \text{Bernoulli}(\frac{1}{n})$, $n \in I\!N$.

Prove that $X_n \stackrel{a.s.}{\not\rightarrow} 0$, by deriving $P(\{X_n = 0, \text{ for every } m \leq n \leq n_0\})$ and observing that this probability does not converge to 1 as $n_0 \rightarrow +\infty$ for all values of m (Rohatgi, 1976, p. 252, Example 9). •

Motivation 5.4 — Convergence in probability (Karr, 1993, p. 135; http://en.wikipedia.org/wiki/Convergence_of_random_variables)
Convergence in probability essentially means that the probability that $|X_n - X|$ exceeds any prescribed, strictly positive value converges to zero.

The basic idea behind this type of convergence is that the probability of an "unusual" outcome becomes smaller and smaller as the sequence progresses. •

Definition 5.5 — Convergence in probability (Karr, 1993, p. 136; Rohatgi, 1976, p. 243)
The sequence of r.v. $\{X_1, X_2, \ldots\}$ is said to converge in probability to a r.v. X — denoted by $X_n \stackrel{P}{\to} X$ — if

$$\lim_{n \to +\infty} P\left(\{|X_n - X| > \epsilon\}\right) = 0, \tag{5.2}$$

for every $\epsilon > 0$. •

Remarks 5.6 — Convergence in probability (Rohatgi, 1976, p. 243; http://en.wikipedia.org/wiki/Convergence_of_random_variables)

- The definition of convergence in probability says nothing about the convergence of r.v. X_n to r.v. X in the sense in which it is understood in real analysis. Thus, $X_n \stackrel{P}{\to} X$ does not imply that, given $\epsilon > 0$, we can find an N such that $|X_n - X| < \epsilon$, for $n \geq N$.

  Definition 5.5 speaks only of the convergence of the sequence of probabilities $P\left(|X_n - X| > \epsilon\right)$ to zero.

- Formally, Definition 5.5 means that

$$\forall \epsilon, \delta > 0, \, \exists N_\delta : P(\{|X_n - X| > \epsilon\}) < \delta, \, \forall n \geq N_\delta. \tag{5.3}$$

- The concept of convergence in probability is used very often in statistics. For example, an estimator is called consistent if it converges in probability to the parameter being estimated.

- Convergence in probability is also the type of convergence established by the weak law of large numbers. •

Exercise 5.7 — **Convergence in probability**
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that $X_n \sim \text{Bernoulli}(\frac{1}{n})$, $n \in I\!N$.

(a) Prove that $X_n \stackrel{P}{\to} 0$, by obtaining $P(\{|X_n| > \epsilon\})$, for $0 < \epsilon < 1$ and $\epsilon \geq 1$ (Rohatgi, 1976, pp. 243–244, Example 5).

(b) Verify that $E(X_n^k) \to E(X^k)$, where $k \in I\!N$ and $X \stackrel{d}{=} 0$. •

Exercise 5.8 — **Convergence in probability does not imply convergence of kth. moments**
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that $X_n \stackrel{d}{=} n \times \text{Bernoulli}(\frac{1}{n})$, $n \in I\!N$, i.e.

$$P(\{X_n = x\}) = \begin{cases} 1 - \frac{1}{n}, & x = 0 \\ \frac{1}{n}, & x = n \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

Prove that $X_n \stackrel{P}{\to} 0$, however $E(X_n^k) \not\to E(X^k)$, where $k \in I\!N$ and the r.v. X is degenerate at 0 (Rohatgi, 1976, p. 247, Remark 3). •

Motivation 5.9 — **Convergence in quadratic mean and in L^1**
We have just seen that convergence in probability does not imply the convergence of moments, namely of orders 2 or 1. •

Definition 5.10 — **Convergence in quadratic mean or in L^2** (Karr, 1993, p. 136)
Let $X, X_1, X_2, \ldots$ belong to L^2. Then the sequence of r.v. $\{X_1, X_2, \ldots\}$ is said to converge to X in quadratic mean (or in L^2) — denoted by $X_n \stackrel{q.m.}{\to} X$ (or $X_n \stackrel{L^2}{\to} X$) — if

$$\lim_{n \to +\infty} E\left[(X_n - X)^2\right] = 0. \tag{5.5}$$

•

Exercise 5.11 — **Convergence in quadratic mean**
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that $X_n \sim \text{Bernoulli}\left(\frac{1}{n}\right)$.

Prove that $X_n \stackrel{q.m.}{\to} X$, where the r.v. X is degenerate at 0 (Rohatgi, 1976, p. 247, Example 6). •

Exercise 5.12 — Convergence in quadratic mean (bis)
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. with $P\left(\left\{X_n = \pm\frac{1}{n}\right\}\right) = \frac{1}{2}$.

Prove that $X_n \stackrel{q.m.}{\rightarrow} X$, where the r.v. X is degenerate at 0 (Rohatgi, 1976, p. 252, Example 11). •

Exercise 5.13 — Convergence in quadratic mean implies convergence of 2nd. moments (Karr, 1993, p. 158, Exercise 5.6(a))
Prove that $X_n \stackrel{q.m.}{\rightarrow} X \Rightarrow E(X_n^2) \rightarrow E(X^2)$ (Rohatgi, 1976, p. 248, proof of Theorem 8). •

Exercise 5.14 — Convergence in quadratic mean of partial sums (Karr, 1993, p. 159, Exercise 5.11)
Let $X_1, X_2, \ldots$ be pairwise uncorrelated r.v. with mean zero and partial sums $S_n = \sum_{i=1}^n X_i$.

Prove that if there is a constant c such that $V(X_i) \leq c$, for every i, then $\frac{S_n}{n^\alpha} \stackrel{q.m.}{\rightarrow} 0$ for all $\alpha > \frac{1}{2}$. •

Definition 5.15 — Convergence in mean or in L^1 (Karr, 1993, p. 136)
Let $X, X_1, X_2, \ldots$ belong to L^1. Then the sequence of r.v. $\{X_1, X_2, \ldots\}$ is said to converge to X in mean (or in L^1) — denoted by $X_n \stackrel{L^1}{\rightarrow} X$ — if

$$\lim_{n\rightarrow+\infty} E\left(|X_n - X|\right) = 0. \tag{5.6}$$

•

Exercise 5.16 — Convergence in mean implies convergence of 1st. moments (Karr, 1993, p. 158, Exercise 5.6(b))
Prove that $X_n \stackrel{L^1}{\rightarrow} X \Rightarrow E(X_n) \rightarrow E(X)$. •

5.1.2 Convergence in distribution

Motivation 5.17 — Convergence in distribution
(http://en.wikipedia.org/wiki/Convergence_of_random_variables)
Convergence in distribution is very frequently used in practice, most often it arises from the application of the central limit theorem. •

Definition 5.18 — Convergence in distribution (Karr, 1993, p. 136; Rohatgi, 1976, pp. 240–1)
The sequence of r.v. $\{X_1, X_2, \ldots\}$ converges to X in distribution — denoted by $X_n \stackrel{d}{\rightarrow} X$ — if

$$\lim_{n \to +\infty} F_{X_n}(x) = F_X(x), \tag{5.7}$$

for all x at which F_X is continuous. •

Remarks 5.19 — Convergence in distribution
(http://en.wikipedia.org/wiki/Convergence_of_random_variables; Karr, 1993, p. 136; Rohatgi, 1976, p. 242)

- With this mode of convergence, we increasingly expect to see the next r.v. in a sequence of r.v. becoming better and better modeled by a given d.f., as seen in Exercise 5.20.

- It must be noted that it is quite possible for a given sequence of d.f. to converge to a function that is not a d.f., as shown in Exercise 5.21.

- The requirement that only the continuity points of F_X should be considered is essential, as we shall see in exercises 5.22 and 5.23.

- The convergence in distribution does not imply the convergence of corresponding p.(d.)f., as shown in Exercise 5.24. •

Exercise 5.20 — Convergence in distribution
Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v. with common p.d.f.

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{otherwise,} \end{cases} \tag{5.8}$$

where $0 < \theta < +\infty$, and $X_{(n)} = \max_{1,\ldots,n} X_i$.

Prove that $X_{(n)} \stackrel{d}{\rightarrow} \theta$ (Rohatgi, 1976, p. 241, Example 2). •

Exercise 5.21 — A sequence of d.f. converging to a non d.f.
Consider the sequence of d.f.

$$F_{X_n}(x) = \begin{cases} 0, & x < n \\ 1, & x \geq n, \end{cases} \tag{5.9}$$

where $F_{X_n}(x)$ is the d.f. of the r.v. X_n degenerate at $x = n$.

Verify that $F_{X_n}(x)$ converges to a function (that is identically equal to 0!!!) which is not a d.f. (Rohatgi, 1976, p. 241, Example 1). •

Exercise 5.22 — The requirement that only the continuity points of F_X should be considered is essential
Let $X_n \sim \text{Uniform}\left(\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right)$ and X be a r.v. degenerate at $\frac{1}{2}$.

(a) Prove that $X_n \xrightarrow{d} X$ (Karr, 1993, p. 142).

(b) Verify that $F_{X_n}\left(\frac{1}{2}\right) = \frac{1}{2}$ for each n, and these values do not converge to $F_X\left(\frac{1}{2}\right) = 1$. Is there any contradiction with the convergence in distribution previously proved? (Karr, 1993, p. 142.) •

Exercise 5.23 — The requirement that only the continuity points of F_X should be considered is essential (bis)
Let $X_n \sim \text{Uniform}\left(0, \frac{1}{n}\right)$ and X a r.v. degenerate at 0.

Prove that $X_n \xrightarrow{d} X$, even though $F_{X_n}(0) = 0$, for all n, and $F_X(0) = 1$, that is, the convergence of d.f. fails at the point $x = 0$ where F_X is discontinuous (http://en.wikipedia.org/wiki/Convergence_of_random_variables). •

Exercise 5.24 — Convergence in distribution does not imply convergence of corresponding p.(d.)f.
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. with p.f.

$$P(\{X_n = x\}) = \begin{cases} 1, & x = 2 + \frac{1}{n} \\ 0, & \text{otherwise.} \end{cases} \tag{5.10}$$

(a) Prove that $X_n \xrightarrow{d} X$, where X a r.v. degenerate at 2.

(b) Verify that none of the p.f. $P(\{X_n = x\})$ assigns any probability to the point $x = 2$, for all n, and that $P(\{X_n = x\}) \to 0$ for all x (Rohatgi, 1976, p. 242, Example 4). •

The following table condenses the definitions of convergence of sequences of r.v.

Mode of convergence	Assumption	Defining condition
$X_n \stackrel{a.s.}{\to} X$ (almost sure)	—	$P(\{w : X_n(\omega) \to X(\omega)\}) = 1$
$X_n \stackrel{P}{\to} X$ (in probability)	—	$P(\{\lvert X_n - X\rvert > \epsilon\}) \to 0$, for all $\epsilon > 0$
$X_n \stackrel{q.m}{\to} X$ (in quadratic mean)	$X, X_1, X_2, \ldots \in L^2$	$E\left[(X_n - X)^2\right] \to 0$
$X_n \stackrel{L^1}{\to} X$ (in L^1)	$X, X_1, X_2, \ldots \in L^1$	$E(\lvert X_n - X\rvert) \to 0$
$X_n \stackrel{d}{\to} X$ (in distribution)	—	$F_{X_n}(x) \to F_X(x)$, at continuity points x of F_X

Exercise 5.25 — **Modes of convergence and uniqueness of limit** (Karr, 1993, p. 158, Exercise 5.1)
Prove that for all five forms of convergence the limit is unique. In particular:

(a) if $X_n \stackrel{*}{\to} X$ and $X_n \stackrel{*}{\to} Y$, where $* = a.s., P, q.m., L^1$, then $X \stackrel{a.s.}{\to} Y$;

(b) if $X_n \stackrel{d}{\to} X$ and $X_n \stackrel{d}{\to} Y$, then $X \stackrel{d}{\to} Y$; •

Exercise 5.26 — **Modes of convergence and the vector space structure of the family of r.v.** (Karr, 1993, p. 158, Exercise 5.2)
Prove that, for $* = a.s., P, q.m., L^1$,

$$X_n \stackrel{*}{\to} X \Leftrightarrow X_n - X \stackrel{*}{\to} 0, \tag{5.11}$$

i.e. the four function-based forms of convergence are compatible with the vector space structure of the family of r.v. •

5.1.3 Alternative criteria

The definition of almost sure convergence and its verification are far from trivial. More tractable criteria have to be stated...

Proposition 5.27 — Relating almost sure convergence and convergence in probability (Karr, 1993, p. 137; Rohatgi, 1976, p. 249)
$X_n \stackrel{a.s.}{\rightarrow} X$ iff

$$\forall \epsilon > 0, \lim_{n \to +\infty} P\left(\left\{\sup_{k \geq n} |X_k - X| > \epsilon\right\}\right) = 0, \tag{5.12}$$

i.e.

$$X_n \stackrel{a.s.}{\rightarrow} X \quad \Leftrightarrow \quad Y_n = \sup_{k \geq n} |X_k - X| \stackrel{P}{\rightarrow} 0. \tag{5.13}$$

•

Remarks 5.28 — Relating almost sure convergence and convergence in probability (Karr, 1993, p. 137; Rohatgi, 1976, p. 250, Remark 6)

- Proposition 5.27 states an equivalent form of almost sure convergence that illuminates its relationship to convergence in probability.

- $X_n \stackrel{a.s.}{\rightarrow} 0$ means that,

$$\forall \epsilon, \eta > 0, \exists n_0 \in I\!N : P\left(\left\{\sup_{k \geq n_0} |X_k| > \epsilon\right\}\right) < \eta. \tag{5.14}$$

  Indeed, we can write, equivalently, that

$$\lim_{n \to +\infty} P\left(\bigcup_{k \geq n} \{|X_k| > \epsilon\}\right) = 0, \tag{5.15}$$

  for $\epsilon > 0$ arbitrary. •

Exercise 5.29 — Relating almost sure convergence and convergence in probability
Prove Proposition 5.27 (Karr, 1993, p. 137; Rohatgi, 1976, p. 250). •

Exercise 5.30 — Relating almost sure convergence and convergence in probability (bis)
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. with $P\left(\left\{X_n = \pm\frac{1}{n}\right\}\right) = \frac{1}{2}$.

Prove that $X_n \stackrel{a.s.}{\to} X$, where the r.v. X is degenerate at 0, by using (5.15) (Rohatgi, 1976, p. 252). •

Theorem 5.31 — Cauchy criterion (Rohatgi, 1976, p. 270)

$$X_n \stackrel{a.s.}{\to} X \quad \Leftrightarrow \quad \lim_{n\to+\infty} P\left(\left\{\sup_m |X_{n+m} - X_n| \leq \epsilon\right\}\right) = 1, \forall \epsilon > 0. \tag{5.16}$$

•

Exercise 5.32 — Cauchy criterion
Prove Theorem 5.31 (Rohatgi, 1976, pp. 270–2). •

Definition 5.33 — Complete convergence (Karr, 1993, p. 138)
The sequence of r.v. $\{X_1, X_2, \ldots\}$ is said to converge completely to X if

$$\sum_{n=1}^{+\infty} P\left(\{|X_n - X| > \epsilon\}\right) < +\infty, \tag{5.17}$$

for every $\epsilon > 0$. •

The next results relate complete convergence, which is stronger than almost sure convergence, and sometimes more convenient to establish (Karr, 1993, p. 137).

Proposition 5.34 — Relating almost sure convergence and complete convergence (Karr, 1993, p. 138)

$$\sum_{n=1}^{+\infty} P\left(\{|X_n - X| > \epsilon\}\right) < +\infty, \forall \epsilon > 0 \quad \Rightarrow \quad X_n \stackrel{a.s.}{\to} X. \tag{5.18}$$

•

Remark 5.35 — Relating almost sure convergence and complete convergence (Karr, 1993, p. 138)
$X_n \stackrel{P}{\to} X$ iff the probabilities $P\left(\{|X_n - X| > \epsilon\}\right)$ converge to zero, while $X_n \stackrel{a.s.}{\to} X$ if (but not only if) the convergence of probabilities $P\left(\{|X_n - X| > \epsilon\}\right)$ is fast enough that their sum, $\sum_{n=1}^{+\infty} P\left(\{|X_n - X| > \epsilon\}\right)$, is finite. •

Exercise 5.36 — **Relating almost sure convergence and complete convergence**
Prove Proposition 5.34, by using the (1st.) Borel-Cantelli lemma (Karr, 1993, p. 138). •

Theorem 5.37 — **Almost sure convergence of a sequence of independent r.v.** (Rohatgi, 1976, p. 265)
Let $\{X_1, X_2, \ldots\}$ be a sequence of independent r.v. Then

$$X_n \stackrel{a.s.}{\to} 0 \quad \Leftrightarrow \quad \sum_{n=1}^{+\infty} P\left(\{|X_n| > \epsilon\}\right) < +\infty, \forall \epsilon > 0. \tag{5.19}$$

•

Exercise 5.38 — **Almost sure convergence of a sequence of independent r.v.**
Prove Theorem 5.37 (Rohatgi, 1976, pp. 265–6). •

The definition of convergence in distribution is cumbersome because of the proviso regarding continuity points of the limit d.f. F_X. An alternative criterion follows.

Theorem 5.39 — **Alternative criterion for convergence in distribution** (Karr, 1993, p. 138)
Let $\mathbf{C}$ be the set of bounded, continuous functions $f : \mathbb{R} \to \mathbb{R}$. Then

$$X_n \stackrel{d}{\to} X \quad \Leftrightarrow \quad E[f(X_n)] \to E[f(X)], \forall f \in \mathbf{C}. \tag{5.20}$$

•

Remark 5.40 — **Alternative criterion for convergence in distribution** (Karr, 1993, p. 138)
Theorem 5.39 provides a criterion for convergence in distribution which is superior to the definition of convergence in distribution in that one need not to deal with continuity points of the limit d.f. •

Exercise 5.41 — **Alternative criterion for convergence in distribution**
Prove Theorem 5.39 (Karr, 1993, pp. 138–139). •

Since in the proof of Theorem 5.39 the continuous functions used to approximate indicator functions can be taken to be arbitrarily smooth we can add a sufficient condition that guarantees convergence in distribution.

Corollary 5.42 — Sufficient condition for convergence in distribution (Karr, 1993, p. 139)
Let:

- k be a fixed non-negative integer;
- $\mathbf{C}^{(k)}$ be the space of bounded, $k-$times uniformly continuously differentiable functions $f : I\!R \to I\!R$.

Then

$$E[f(X_n)] \to E[f(X)], \forall\, f \in \mathbf{C}^{(k)} \quad \Rightarrow \quad X_n \stackrel{d}{\to} X. \tag{5.21}$$

•

The next table summarizes the alternative criteria and sufficient conditions for almost sure convergence and convergence in distribution of sequences of r.v.

Alternative criterion or sufficient condition		**Mode of convergence**
$\forall \epsilon > 0, \lim_{n\to+\infty} P\left(\left\{\sup_{k\geq n} \lvert X_k - X\rvert > \epsilon\right\}\right) = 0$	$\Leftrightarrow$	$X_n \stackrel{a.s.}{\to} X$
$Y_n = \sup_{k\geq n} \lvert X_k - X\rvert \stackrel{P}{\to} 0$	$\Leftrightarrow$	$X_n \stackrel{a.s.}{\to} X$
$\lim_{n\to+\infty} P\left(\{\sup_m \lvert X_{n+m} - X_n\rvert \leq \epsilon\}\right) = 1, \forall \epsilon > 0$	$\Leftrightarrow$	$X_n \stackrel{a.s.}{\to} X$
$\sum_{n=1}^{+\infty} P\left(\{\lvert X_n - X\rvert > \epsilon\}\right) < +\infty, \forall\, \epsilon > 0$	$\Rightarrow$	$X_n \stackrel{a.s.}{\to} X$
$E[f(X_n)] \to E[f(X)], \forall\, f \in \mathbf{C}$	$\Leftrightarrow$	$X_n \stackrel{d}{\to} X$
$E[f(X_n)] \to E[f(X)], \forall\, f \in \mathbf{C}^{(k)}$ for a fixed $k \in I\!N_0$	$\Rightarrow$	$X_n \stackrel{d}{\to} X$

5.2 Relationships among the modes of convergence

Given the plethora of modes of convergence, it is natural to inquire how they can be always related or hold true in the presence of additional assumptions (Karr, 1993, pp. 140 and 142).

5.2.1 Implications always valid

Proposition 5.43 — Almost sure convergence implies convergence in probability (Karr, 1993, p. 140; Rohatgi, 1976, p. 250)

$$X_n \stackrel{a.s.}{\to} X \quad \Rightarrow \quad X_n \stackrel{P}{\to} X. \tag{5.22}$$

•

Exercise 5.44 — Almost sure convergence implies convergence in probability
Prove Proposition 5.43 (Karr, 1993, p. 140; Rohatgi, 1976, p. 251). •

Proposition 5.45 — Convergence in quadratic mean implies convergence in L^1 (Karr, 1993, p. 140)

$$X_n \stackrel{q.m.}{\to} X \quad \Rightarrow \quad X_n \stackrel{L^1}{\to} X. \tag{5.23}$$

•

Exercise 5.46 — Convergence in quadratic mean implies convergence in L^1
Prove Proposition 5.45, by applying Cauchy-Schwarz's inequality (Karr, 1993, p. 140). •

Proposition 5.47 — Convergence in L^1 implies convergence in probability (Karr, 1993, p. 141)

$$X_n \stackrel{L^1}{\to} X \quad \Rightarrow \quad X_n \stackrel{P}{\to} X. \tag{5.24}$$

•

Exercise 5.48 — Convergence in L^1 implies convergence in probability
Prove Proposition 5.47, by using Chebyshev's inequality (Karr, 1993, p. 141). •

Proposition 5.49 — Convergence in probability implies convergence in distribution (Karr, 1993, p. 141)

$$X_n \stackrel{P}{\to} X \quad \Rightarrow \quad X_n \stackrel{d}{\to} X. \tag{5.25}$$

•

Exercise 5.50 — Convergence in probability implies convergence in distribution

Prove Proposition 5.49, (Karr, 1993, p. 141). •

Figure 5.1 shows that convergence in distribution is the weakest form of convergence, since it is implied by all other types of convergence studied so far.

$$\begin{array}{ccccccc} X_n \stackrel{q.m.}{\to} X & \Rightarrow & X_n \stackrel{L^1}{\to} X & & & & \\ & & & \Rightarrow & & & \\ & & & & X_n \stackrel{P}{\to} X & \Rightarrow & X_n \stackrel{d}{\to} X \\ & & & \Rightarrow & & & \\ & & X_n \stackrel{a.s.}{\to} X & & & & \end{array}$$

Figure 5.1: Implications always valid between modes of convergence.

5.2.2 Counterexamples

Counterexamples to all implications among the modes of convergence (and more!) are condensed in Figure 5.2 and presented by means of several exercises.

$$\begin{array}{ccccccc} X_n \stackrel{q.m.}{\to} X & \not\Leftarrow & X_n \stackrel{L^1}{\to} X & & & & \\ & & & \not\Leftarrow & & & \\ & & & & X_n \stackrel{P}{\to} X & \not\Leftarrow & X_n \stackrel{d}{\to} X \\ & & & \not\Leftarrow & & & \\ & & X_n \stackrel{a.s.}{\to} X & & & & \end{array}$$

Figure 5.2: Counterexamples to implications among the modes of convergence.

Before proceeding with exercises, recall exercises 5.3 and 5.7 which pertain to the sequence of r.v. $\{X_1, X_2, \ldots\}$, where $X_n \sim \text{Bernoulli}(\frac{1}{n})$, $n \in I\!N$. In the first exercise we proved that $X_n \stackrel{a.s.}{\not\to} 0$, whereas in the second one we concluded that $X_n \stackrel{P}{\to} 0$. Thus, combining these results we can state that $X_n \stackrel{P}{\to} 0 \not\Rightarrow X_n \stackrel{a.s.}{\to} 0$.

Exercise 5.51 — Almost sure convergence does not imply convergence in quadratic mean

Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that

$$P(\{X_n = x\}) = \begin{cases} 1 - \frac{1}{n}, & x = 0 \\ \frac{1}{n}, & x = n \\ 0, & \text{otherwise.} \end{cases} \tag{5.26}$$

Prove that $X_n \stackrel{a.s.}{\to} 0$, and, hence, $X_n \stackrel{P}{\to} 0$ and $X_n \stackrel{d}{\to} 0$, but $X_n \stackrel{L^1}{\not\to} 0$ and $X_n \stackrel{q.m.}{\not\to} 0$ (Karr, 1993, p. 141, Counterexample a)). •

Exercise 5.52 — Almost sure convergence does not imply convergence in quadratic mean (bis)

Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that

$$P(\{X_n = x\}) = \begin{cases} 1 - \frac{1}{n^r}, & x = 0 \\ \frac{1}{n^r}, & x = n \\ 0, & \text{otherwise,} \end{cases} \tag{5.27}$$

where $r \geq 2$.

Prove that $X_n \stackrel{a.s.}{\to} 0$, but $X_n \stackrel{q.m.}{\not\to} 0$ (Rohatgi, 1976, p. 252, Example 10). •

Exercise 5.53 — Convergence in quadratic mean does not imply almost sure convergence

Let $X_n \sim \text{Bernoulli}\left(\frac{1}{n}\right)$.

Prove that $X_n \stackrel{q.m.}{\to} 0$, but $X_n \stackrel{a.s.}{\not\to} 0$ (Rohatgi, 1976, p. 252, Example 9). •

Exercise 5.54 — Convergence in L^1 does not imply convergence in quadratic mean

Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that

$$P(\{X_n = x\}) = \begin{cases} 1 - \frac{1}{n}, & x = 0 \\ \frac{1}{n}, & x = \sqrt{n} \\ 0, & \text{otherwise.} \end{cases} \tag{5.28}$$

Prove that $X_n \stackrel{a.s.}{\to} 0$ and $X_n \stackrel{L^1}{\to} 0$, however $X_n \stackrel{q.m.}{\not\to} 0$ (Karr, 1993, p. 141, Counterexample b)). •

Exercise 5.55 — Convergence in probability does not imply almost sure convergence

For each positive integer n there exists integers m and k (uniquely determined) such that

$$n = 2^k + m,\ m = 0, 1, \ldots, 2^k - 1,\ k = 0, 1, 2, \ldots \tag{5.29}$$

Thus, for $n = 1$, $k = m = 0$; for $n = 5$, $k = 2$, m=1; and so on.

Define r.v. X_n, for $n = 1, 2, \ldots$, on $\Omega = [0, 1]$ by

$$X_n(\omega) = \begin{cases} 2^k, & \frac{m}{2^k} \leq w < \frac{m+1}{2^k} \\ 0, & \text{otherwise.} \end{cases} \tag{5.30}$$

Let the probability distribution of X_n be given by $P(\{I\})$ =length of the interval $I \subset \Omega$. Thus,

$$P(\{X_n = x\}) = \begin{cases} 1 - \frac{1}{2^k}, & x = 0 \\ \frac{1}{2^k}, & x = 2^k \\ 0, & \text{otherwise.} \end{cases} \tag{5.31}$$

Prove that $X_n \stackrel{P}{\rightarrow} 0$, but $X_n \stackrel{a.s.}{\not\rightarrow} 0$ (Rohatgi, 1976, pp. 251–2, Example 8). •

Exercise 5.56 — Convergence in distribution does not imply convergence in probability

Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. such that

$$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2} - \frac{1}{n}, & 0 \leq x < 1 \\ 1, & x \geq 1, \end{cases} \tag{5.32}$$

i.e. $X_n \sim \text{Bernoulli}\left(\frac{1}{2} + \frac{1}{n}\right)$.

Prove that $X_n \stackrel{d}{\rightarrow} X$, where $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, but $X_n \stackrel{P}{\not\rightarrow} X$ (Karr, 1993, p. 142, Counterexample d)). •

Exercise 5.57 — Convergence in distribution does not imply convergence in probability (bis)

Let $X, X_1, X_2, \ldots$ be i.i.d. r.v. and let the joint p.f. of (X, X_n) be $P(\{X = 0, X_n = 1\}) = P(\{X = 1, X_n = 0\}) = \frac{1}{2}$.

Prove that $X_n \stackrel{d}{\rightarrow} X$, but $X_n \stackrel{P}{\rightarrow} X$ (Rohatgi, 1976, p. 247, Remark 2). •

5.2.3 Implications of restricted validity

Proposition 5.58 — Convergence in distribution to a constant implies convergence in probability (Karr, 1993, p. 140; Rohatgi, 1976, p. 246)
Let $\{X_1, X_2, \ldots\}$ be a sequence of r.v. and $c \in I\!\!R$. Then

$$X_n \stackrel{d}{\to} c \quad \Rightarrow \quad X_n \stackrel{P}{\to} c. \tag{5.33}$$

•

Remark 5.59 — Convergence in distribution to a constant is equivalent to convergence in probability (Rohatgi, 1976, p. 246)
If we add to the previous result the fact that $X_n \stackrel{P}{\to} c \Rightarrow X_n \stackrel{d}{\to} c$, we can conclude that

$$X_n \stackrel{P}{\to} c \quad \Leftrightarrow \quad X_n \stackrel{d}{\to} c. \tag{5.34}$$

•

Exercise 5.60 — Convergence in distribution to a constant implies convergence in probability
Prove Proposition 5.58 (Karr, 1993, p. 140). •

Definition 5.61 — Uniform integrability (Karr, 1993, p. 142)
A sequence of r.v. $\{X_1, X_2, \ldots\}$ is uniformly integrable if $X_n \in L^1$ for each $n \in I\!\!N$ and if

$$\lim_{a \to +\infty} \sup_n E(|X_n|; \{|X_n| > a\}) = 0. \tag{5.35}$$

Recall that the expected value of a r.v. X over an event A is given by $E(X; A) = E(X \times \mathbf{1}_A)$. •

Proposition 5.62 — Alternative criterion for uniform integrability (Karr, 1993, p. 143)
A sequence of r.v. $\{X_1, X_2, \ldots\}$ is uniformly integrable iff

- $\sup_n E(|X_n|) < +\infty$ and
- $\{X_1, X_2, \ldots\}$ is uniformly absolutely continuous: for each $\epsilon > 0$ there is $\delta > 0$ such that $\sup_n E(|X_n|; A) < \epsilon$ whenever $P(A) > \delta$. •

Proposition 5.63 — Combining convergence in probability and uniform integrability is equivalent to convergence in L^1 (Karr, 1993, p. 144)
Let $X, X_1, X_2, \ldots \in L^1$. Then

$$X_n \stackrel{P}{\to} X \text{ and } \{X_1, X_2, \ldots\} \text{ is uniformly integrable} \quad \Leftrightarrow \quad X_n \stackrel{L^1}{\to} X. \tag{5.36}$$

•

Exercise 5.64 — Combining convergence in probability and uniform integrability is equivalent to convergence in L^1
Prove Proposition 5.63 (Karr, 1993, p. 144). •

Exercise 5.65 — Combining convergence in probability of the sequence of r.v. and convergence of sequence of the means implies convergence in L^1 (Karr, 1993, p. 160, Exercise 5.16)
Let $X, X_1, X_2, \ldots$ be positive r.v.

Prove that if $X_n \stackrel{P}{\to} X$ and $E(X_n) \to E(X)$, then $X_n \stackrel{L^1}{\to} X$. •

Exercise 5.66 — Increasing character and convergence in probability combined imply almost sure convergence (Karr, 1993, p. 160, Exercise 5.15)
Prove that if $X_1 \leq X_2 \leq \ldots$ and $X_n \stackrel{P}{\to} X$, then $X_n \stackrel{a.s.}{\to} X$. •

Exercise 5.67 — Strictly decreasing and positive character and convergence in probability combined imply almost sure convergence (Rohatgi, 1976, p. 252, Theorem 13)
Let $\{X_1, X_2, \ldots\}$ be a strictly decreasing sequence of positive r.v.

Prove that if $X_n \stackrel{P}{\to} 0$ then $X_n \stackrel{a.s.}{\to} 0$. •

References

- Karr, A.F. (1993). *Probability.* Springer-Verlag.
- Resnick, S.I. (1999). *A Probability Path.* Birkhäuser. (QA273.4-.67.RES.49925)
- Rohatgi, V.K. (1976). *An Introduction to Probability Theory and Mathematical Statistics.* John Wiley & Sons. (QA273-280/4.ROH.34909)