WIKIPEDIA

Convergence of random variables

In probability theory, there exist several different notions of **convergence of random variables**. The convergence of sequences of random variables to some limit random variable is an important concept in probability theory, and its applications to statistics and stochastic processes. The same concepts are known in more general mathematics as **stochastic convergence** and they formalize the idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle down into a behavior that is essentially unchanging when items far enough into the sequence are studied. The different possible notions of convergence relate to how such a behavior can be characterized: two readily understood behaviors are that the sequence eventually takes a constant value, and that values in the sequence continue to change but can be described by an unchanging probability distribution.

Contents

Background

"Stochastic convergence" formalizes the idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle into a pattern. The pattern may for instance be

- Convergence in the classical sense to a fixed value, perhaps itself coming from a random event
- An increasing similarity of outcomes to what a purely deterministic function would produce
- An increasing preference towards a certain outcome

- An increasing "aversion" against straying far away from a certain outcome
- That the probability distribution describing the next outcome may grow increasingly similar to a certain distribution

Some less obvious, more theoretical patterns could be

- That the series formed by calculating the expected value of the outcome's distance from a particular value may converge to 0
- That the variance of the random variable describing the next event grows smaller and smaller.

These other types of patterns that may arise are reflected in the different types of stochastic convergence that have been studied.

While the above discussion has related to the convergence of a single series to a limiting value, the notion of the convergence of two series towards each other is also important, but this is easily handled by studying the sequence defined as either the difference or the ratio of the two series.

For example, if the average of n independent random variables Y_i, $i = 1, ..., n$, all having the same finite mean and variance, is given by

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i\,,$$

then as n tends to infinity, X_n converges *in probability* (see below) to the common mean, μ, of the random variables Y_i. This result is known as the weak law of large numbers. Other forms of convergence are important in other useful theorems, including the central limit theorem.

Throughout the following, we assume that (X_n) is a sequence of random variables, and X is a random variable, and all of them are defined on the same probability space $(\Omega, \mathcal{F}, \operatorname{Pr})$.

Convergence in distribution

With this mode of convergence, we increasingly expect to see the next outcome in a sequence of random experiments becoming better and better modeled by a given probability distribution.

Convergence in distribution is the weakest form of convergence typically discussed, since it is implied by all other types of convergence mentioned in this article. However, convergence in distribution is very frequently used in practice; most often it arises from application of the central limit theorem.

Examples of convergence in distribution

Dice factory

Suppose a new dice factory has just been built. The first few dice come out quite biased, due to imperfections in the production process. The outcome from tossing any of them will follow a distribution markedly different from the desired uniform distribution.

As the factory is improved, the dice become less and less loaded, and the outcomes from tossing a newly produced die will follow the uniform distribution more and more closely.

Tossing coins

Definition

A sequence X_1, X_2, ... of real-valued random variables is said to **converge in distribution**, or **converge weakly**, or **converge in law** to a random variable X if

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

for every number $x \in \mathbb{R}$ at which F is continuous. Here F_n and F are the cumulative distribution functions of random variables X_n and X, respectively.

The requirement that only the continuity points of F should be considered is essential. For example, if X_n are distributed uniformly on intervals $(0, \frac{1}{n})$, then this sequence converges in distribution to a degenerate random variable $X = 0$. Indeed, $F_n(x) = 0$ for all n when $x \le 0$, and $F_n(x) = 1$ for all $x \ge \frac{1}{n}$ when $n > 0$. However, for this limiting random variable $F(0) = 1$, even though $F_n(0) = 0$ for all n. Thus the convergence of cdfs fails at the point $x = 0$ where F is discontinuous.

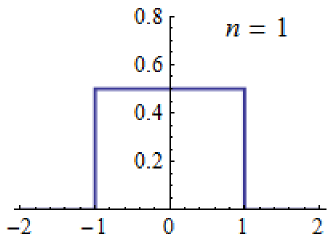
Convergence in distribution may be denoted as

Let X_n be the fraction of heads after tossing up an unbiased coin n times. Then X_1 has the Bernoulli distribution with expected value $\mu = 0.5$ and variance $\sigma^2 = 0.25$. The subsequent random variables $X_2, X_3, \ldots$ will all be distributed binomially.

As n grows larger, this distribution will gradually start to take shape more and more similar to the bell curve of the normal distribution. If we shift and rescale X_n appropriately, then $Z_n = \frac{\sqrt{n}}{\sigma}(X_n - \mu)$ will be **converging in distribution** to the standard normal, the result that follows from the celebrated central limit theorem.

Graphic example

Suppose $\{X_i\}$ is an iid sequence of uniform $U(-1, 1)$ random variables. Let $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ be their (normalized) sums. Then according to the central limit theorem, the distribution of Z_n approaches the normal $N(0, \frac{1}{3})$ distribution. This convergence is shown in the picture: as n grows larger, the shape of the probability density function gets closer and closer to the Gaussian curve.



(1)

$$X_n \xrightarrow{d} X,\ \ X_n \xrightarrow{\mathcal{D}} X,\ \ X_n \xrightarrow{\mathcal{L}} X,\ \ X_n \xrightarrow{d} \mathcal{L}_X,$$
$$X_n \rightsquigarrow X,\ \ X_n \Rightarrow X,\ \ \mathcal{L}(X_n) \to \mathcal{L}(X),$$

where $\mathcal{L}_X$ is the law (probability distribution) of X. For example, if X is standard normal we can write $X_n \xrightarrow{d} \mathcal{N}(0,\,1)$.

For random vectors $\{X_1, X_2, \dots\} \subset \mathbf{R}^k$ the convergence in distribution is defined similarly. We say that this sequence **converges in distribution** to a random k-vector X if

$$\lim_{n\to\infty} \Pr(X_n \in A) = \Pr(X \in A)$$

for every $A \subset \mathbf{R}^k$ which is a continuity set of X.

The definition of convergence in distribution may be extended from random vectors to more general random elements in arbitrary metric spaces, and even to the “random variables” which are not measurable — a situation which occurs for example in the study of empirical processes. This is the “weak convergence of laws without laws being defined” — except asymptotically.[1]

In this case the term **weak convergence** is preferable (see weak convergence of measures), and we say that a sequence of random elements $\{X_n\}$ converges weakly to X (denoted as $X_n \Rightarrow X$) if

$$\mathrm{E}^*\, h(X_n) \to \mathrm{E}\, h(X)$$

for all continuous bounded functions h.[2] Here E^* denotes the *outer expectation*, that is the expectation of a “smallest measurable function g that dominates $h(X_n)$”.

Properties

- Since $F(a) = \Pr(X \le a)$, the convergence in distribution means that the probability for X_n to be in a given range is approximately equal to the probability that the value of X is in that range, provided n is sufficiently large.
- In general, convergence in distribution does not imply that the sequence of corresponding probability density functions will also converge. As an example one may consider random variables with densities $f_n(x) = (1 - \cos(2\pi nx))\mathbf{1}_{(0,1)}$. These random variables converge in distribution to a uniform *U*(0, 1), whereas their densities do not converge at all.[3]
  - However, according to *Scheffé’s theorem*, convergence of the probability density functions implies convergence in distribution.[4]
- The **portmanteau lemma** provides several equivalent definitions of convergence in distribution. Although these definitions are less intuitive, they are used to prove a number of statistical theorems. The lemma states that $\{X_n\}$ converges in distribution to X if and only if any of the following statements are true:[5]
  - $\Pr(X_n \le x) \to \Pr(X \le x)$ for all continuity points of $x \mapsto \Pr(X \le x)$;
  - $\operatorname{E} f(X_n) \to \operatorname{E} f(X)$ for all bounded, continuous functions f (where E denotes the expected value operator);
  - $\operatorname{E} f(X_n) \to \operatorname{E} f(X)$ for all bounded, Lipschitz functions f;
  - $\liminf \operatorname{E} f(X_n) \ge \operatorname{E} f(X)$ for all nonnegative, continuous functions f;
  - $\liminf \Pr(X_n \in G) \ge \Pr(X \in G)$ for every open set G;

- $\limsup \Pr(X_n \in F) \leq \Pr(X \in F)$ for every closed set F;
- $\Pr(X_n \in B) \to \Pr(X \in B)$ for all continuity sets B of random variable X;
- $\limsup \operatorname{E} f(X_n) \leq \operatorname{E} f(X)$ for every upper semi-continuous function f bounded above;
- $\liminf \operatorname{E} f(X_n) \geq \operatorname{E} f(X)$ for every lower semi-continuous function f bounded below.

- The **continuous mapping theorem** states that for a continuous function g, if the sequence $\{X_n\}$ converges in distribution to X, then $\{g(X_n)\}$ converges in distribution to $g(X)$.
  - Note however that convergence in distribution of $\{X_n\}$ to X and $\{Y_n\}$ to Y does in general *not* imply convergence in distribution of $\{X_n + Y_n\}$ to $X + Y$ or of $\{X_n Y_n\}$ to XY.
- **Lévy's continuity theorem:** the sequence $\{X_n\}$ converges in distribution to X if and only if the sequence of corresponding characteristic functions $\{\varphi_n\}$ converges pointwise to the characteristic function φ of X.
- Convergence in distribution is metrizable by the Lévy–Prokhorov metric.
- A natural link to convergence in distribution is the Skorokhod's representation theorem.

Convergence in probability

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

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 quantity being estimated. Convergence in probability is also the type of convergence established by the weak law of large numbers.

Examples of convergence in probability

Height of a person

This example should not be taken literally. Consider the following experiment. First, pick a random person in the street. Let X be his/her height, which is *ex ante* a random variable. Then ask other people to estimate this height by eye. Let X_n be the average of the first n responses. Then (provided there is no systematic error) by the law of large numbers, the sequence X_n will converge in probability to the random variable X.

Definition

A sequence $\{X_n\}$ of random variables **converges in probability** towards the random variable X if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\big(|X_n - X| > \varepsilon\big) = 0.$$

Formally, let P_n be the probability that X_n is outside the ball of radius ε centered at X. Then X_n is said to converge in probability to X if for any $\varepsilon > 0$ and any $\delta > 0$ there exists a number N (which may depend on ε and δ) such that for all $n \geq N$, $P_n < \delta$.

Convergence in probability is denoted by adding the letter p over an arrow indicating convergence, or using the "plim" probability limit operator:

$$X_n \xrightarrow{p} X, \quad X_n \xrightarrow{P} X, \quad \underset{n\to\infty}{\operatorname{plim}} X_n = X. \qquad \textbf{(2)}$$

For random elements $\{X_n\}$ on a separable metric space (S, d), convergence in probability is defined similarly by[6]

$$\forall\varepsilon > 0, \Pr\big(d(X_n, X) \geq \varepsilon\big) \to 0.$$

Properties

- Convergence in probability implies convergence in distribution.[proof]

- In the opposite direction, convergence in distribution implies convergence in probability when the limiting random variable X is a constant.[proof]
- Convergence in probability does not imply almost sure convergence.[proof]
- The continuous mapping theorem states that for every continuous function $g(\cdot)$, if $X_n \xrightarrow{p} X$, then also $g(X_n) \xrightarrow{p} g(X)$.
- Convergence in probability defines a topology on the space of random variables over a fixed probability space. This topology is metrizable by the *Ky Fan metric*:[7]

$$d(X,Y) = \inf\big\{\varepsilon > 0 :\ \Pr\big(|X-Y| > \varepsilon\big) \le \varepsilon\big\}$$

or alternately by this metric

$$d(X,Y) = \mathbb{E}\left[\min(|X-Y|, 1)\right].$$

Almost sure convergence

This is the type of stochastic convergence that is most similar to pointwise convergence known from elementary real analysis.

Definition

To say that the sequence X_n converges **almost surely** or **almost everywhere** or **with probability 1** or **strongly** towards X means that

$$\Pr\!\Big(\lim_{n\to\infty} X_n = X\Big) = 1.$$

This means that the values of X_n approach the value of X, in the sense (see almost surely) that events for which X_n does not converge to X have probability 0. Using the probability space $(\Omega, \mathcal{F}, \Pr)$ and the concept of the random variable as a function from Ω to **R**, this is equivalent to the statement

$$\Pr\Big(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big) = 1.$$

Using the notion of the limit inferior of a sequence of sets, almost sure convergence can also be defined as follows:

$$\Pr\Big(\liminf_{n\to\infty}\big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\big\}\Big) = 1 \quad \text{for all} \quad \varepsilon > 0.$$

Almost sure convergence is often denoted by adding the letters *a.s.* over an arrow indicating convergence:

$$X_n \xrightarrow{\text{a.s.}} X. \qquad (3)$$

For generic random elements $\{X_n\}$ on a metric space (S, d), convergence almost surely is defined similarly:

$$\Pr\Big(\omega \in \Omega :\ d\big(X_n(\omega), X(\omega)\big) \underset{n\to\infty}{\longrightarrow} 0\Big) = 1$$

Examples of almost sure convergence

Example 1
Consider an animal of some short-lived species. We record the amount of food that this animal consumes per day. This sequence of numbers will be unpredictable, but we may be *quite certain* that one day the number will become zero, and will stay zero forever after.
Example 2
Consider a man who tosses seven coins every morning. Each afternoon, he donates one pound to a charity for each head that appeared. The first time the result is all tails, however, he will stop permanently. Let X_1, X_2, … be the daily amounts the charity received from him. We may be *almost sure* that one day this amount will be zero, and stay zero forever after that. However, when we consider *any*

finite number of days, there is a nonzero probability the terminating condition will not occur.

Properties

- Almost sure convergence implies convergence in probability (by Fatou's lemma), and hence implies convergence in distribution. It is the notion of convergence used in the strong law of large numbers.
- The concept of almost sure convergence does not come from a topology on the space of random variables. This means there is no topology on the space of random variables such that the almost surely convergent sequences are exactly the converging sequences with respect to that topology. In particular, there is no metric of almost sure convergence.

Sure convergence or pointwise convergence

To say that the sequence of random variables (X_n) defined over the same probability space (i.e., a random process) converges **surely** or **everywhere** or **pointwise** towards X means

$$\lim_{n\to\infty} X_n(\omega) = X(\omega), \ \forall \omega \in \Omega.$$

where Ω is the sample space of the underlying probability space over which the random variables are defined.

This is the notion of pointwise convergence of a sequence of functions extended to a sequence of random variables. (Note that random variables themselves are functions).

$$\left\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = \Omega.$$

Sure convergence of a random variable implies all the other kinds of convergence stated above, but there is no payoff in probability theory by using sure convergence compared to using almost sure convergence. The difference between the two only exists on sets with probability zero. This is why the concept of sure convergence of random variables is very rarely used.

Convergence in mean

Given a real number $r \geq 1$, we say that the sequence X_n converges **in the *r*-th mean** (or **in the L^r-norm**) towards the random variable X, if the r-th absolute moments $\mathrm{E}(|X_n|^r)$ and $\mathrm{E}(|X|^r)$ of X_n and X exist, and

$$\lim_{n\to\infty} \mathrm{E}(|X_n - X|^r) = 0,$$

where the operator E denotes the expected value. Convergence in r-th mean tells us that the expectation of the r-th power of the difference between X_n and X converges to zero.

This type of convergence is often denoted by adding the letter L^r over an arrow indicating convergence:

$$X_n \xrightarrow{L^r} X. \tag{4}$$

The most important cases of convergence in r-th mean are:

- When X_n converges in r-th mean to X for r = 1, we say that X_n converges **in mean** to X.
- When X_n converges in r-th mean to X for r = 2, we say that X_n converges **in mean square** to X.

Convergence in the r-th mean, for $r \geq 1$, implies convergence in probability (by Markov's inequality). Furthermore, if $r > s \geq 1$, convergence in r-th mean implies convergence in s-th mean. Hence, convergence in mean square implies convergence in mean.

It is also worth noticing that if

$$X_n \xrightarrow{L^r} X, \tag{4}$$

then

$$\lim_{n\to\infty} E[|X_n|^r] = E[|X|^r]$$

Properties

Provided the probability space is complete:

- If $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{\text{a.s.}} X$ and $X_n \xrightarrow{\text{a.s.}} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{L^r} X$ and $X_n \xrightarrow{L^r} Y$, then $X = Y$ almost surely.
- If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $aX_n + bY_n \xrightarrow{p} aX + bY$ (for any real numbers a and b) and $X_nY_n \xrightarrow{p} XY$.
- If $X_n \xrightarrow{\text{a.s.}} X$ and $Y_n \xrightarrow{\text{a.s.}} Y$, then $aX_n + bY_n \xrightarrow{\text{a.s.}} aX + bY$ (for any real numbers a and b) and $X_nY_n \xrightarrow{\text{a.s.}} XY$.
- If $X_n \xrightarrow{L^r} X$ and $Y_n \xrightarrow{L^r} Y$, then $aX_n + bY_n \xrightarrow{L^r} aX + bY$ (for any real numbers a and b).
- None of the above statements are true for convergence in distribution.

The chain of implications between the various notions of convergence are noted in their respective sections. They are, using the arrow notation:

$$\begin{matrix} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{d} \end{matrix}$$

These properties, together with a number of other special cases, are summarized in the following list:

- Almost sure convergence implies convergence in probability:[8][proof]

$$X_n \xrightarrow{\text{a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$$

- Convergence in probability implies there exists a sub-sequence (k_n) which almost surely converges:[9]

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_{k_n} \xrightarrow{\text{a.s.}} X$$

- Convergence in probability implies convergence in distribution:[8][proof]

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

- Convergence in *r*-th order mean implies convergence in probability:

$$X_n \xrightarrow{L^r} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$$

- Convergence in *r*-th order mean implies convergence in lower order mean, assuming that both orders are greater than or equal to one:

$$X_n \xrightarrow{L^r} X \quad \Rightarrow \quad X_n \xrightarrow{L^s} X,$$ provided $r \geq s \geq 1$.

- If X_n converges in distribution to a constant *c*, then X_n converges in probability to *c*:[8][proof]

$$X_n \xrightarrow{d} c \quad \Rightarrow \quad X_n \xrightarrow{p} c,$$ provided *c* is a constant.

- If X_n converges in distribution to *X* and the difference between X_n and Y_n converges in probability to zero, then Y_n also converges in distribution to *X*:[8][proof]

$$X_n \xrightarrow{d} X,\ \ |X_n - Y_n| \xrightarrow{p} 0 \quad \Rightarrow \quad Y_n \xrightarrow{d} X$$

- If X_n converges in distribution to *X* and Y_n converges in distribution to a constant *c*, then the joint vector (X_n, Y_n) converges in distribution to (X, c):[8][proof]

$$X_n \xrightarrow{d} X,\ \ Y_n \xrightarrow{d} c \quad \Rightarrow \quad (X_n, Y_n) \xrightarrow{d} (X, c)$$ provided *c* is a constant.

  Note that the condition that Y_n converges to a constant is important, if it were to converge to a random variable *Y* then we wouldn't be able to conclude that (X_n, Y_n) converges to (X, Y).

- If X_n converges in probability to *X* and Y_n converges in probability to *Y*, then the joint vector (X_n, Y_n) converges in probability to (X, Y):[8][proof]

$$X_n \xrightarrow{p} X,\ \ Y_n \xrightarrow{p} Y \quad \Rightarrow \quad (X_n, Y_n) \xrightarrow{p} (X, Y)$$

- If X_n converges in probability to *X*, and if $\mathbf{P}(|X_n| \leq b) = 1$ for all *n* and some *b*, then X_n converges in *r*th mean to *X* for all $r \geq 1$. In other words, if X_n converges in probability to *X* and all random variables X_n are almost surely bounded above and below, then X_n converges to *X* also in any *r*th mean.
- **Almost sure representation**. Usually, convergence in distribution does not imply convergence almost surely. However, for a given sequence $\{X_n\}$ which converges in distribution to X_0 it is always possible to find a new probability space (Ω, *F*, P) and random variables $\{Y_n, n = 0, 1, ...\}$ defined on it such that Y_n is equal in distribution to X_n for each $n \geq 0$, and Y_n converges to Y_0 almost surely.[10][11]
- If for all *ε* > 0,

$$\sum_n \mathbb{P}\left(|X_n - X| > \varepsilon\right) < \infty,$$

  then we say that X_n *converges almost completely*, or *almost in probability* towards *X*. When X_n converges almost completely towards *X* then it also converges almost surely to *X*. In other words, if X_n converges in probability to *X* sufficiently quickly (i.e. the above sequence of tail probabilities is summable for all $\varepsilon > 0$), then X_n also converges almost surely to X. This is a direct implication from the Borel–Cantelli lemma.

- If S_n is a sum of *n* real independent random variables:

$$S_n = X_1 + \cdots + X_n$$

then S_n converges almost surely if and only if S_n converges in probability.

- The dominated convergence theorem gives sufficient conditions for almost sure convergence to imply L^1-convergence:

$$\left.\begin{matrix} X_n \xrightarrow{\text{a.s.}} X \\ |X_n| < Y \\ \mathrm{E}(Y) < \infty \end{matrix}\right\} \quad \Rightarrow \quad X_n \xrightarrow{L^1} X \tag{5}$$

- A necessary and sufficient condition for L^1 convergence is $X_n \xrightarrow{P} X$ and the sequence (X_n) is uniformly integrable.

See also

- Proofs of convergence of random variables
- Convergence of measures
- Continuous stochastic process: the question of continuity of a stochastic process is essentially a question of convergence, and many of the same concepts and relationships used above apply to the continuity question.
- Asymptotic distribution
- Big O in probability notation
- Skorokhod's representation theorem
- The Tweedie convergence theorem
- Slutsky's theorem

Notes

1. Bickel et al. 1998, A.8, page 475
2. van der Vaart & Wellner 1996, p. 4
3. Romano & Siegel 1985, Example 5.26
4. Durrett, Rick (2010). *Probability: Theory and Examples*. p. 84.
5. van der Vaart 1998, Lemma 2.2
6. Dudley 2002, Chapter 9.2, page 287
7. Dudley 2002, p. 289
8. van der Vaart 1998, Theorem 2.7
9. Gut, Allan (2005). *Probability: A graduate course*. Theorem 3.4: Springer. ISBN 978-0-387-22833-4.
10. van der Vaart 1998, Th.2.19
11. Fristedt & Gray 1997, Theorem 14.5

References

- Bickel, Peter J.; Klaassen, Chris A.J.; Ritov, Ya'acov; Wellner, Jon A. (1998). *Efficient and adaptive estimation for semiparametric models*. New York: Springer-Verlag. ISBN 978-0-387-98473-5.

- Billingsley, Patrick (1986). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics (2nd ed.). Wiley.
- Billingsley, Patrick (1999). *Convergence of probability measures* (https://archive.org/details/convergenceofpro0000bill) (2nd ed.). John Wiley & Sons. pp. 1–28 (https://archive.org/details/convergenceofpro0000bill/page/1). ISBN 978-0-471-19745-4.
- Dudley, R.M. (2002). *Real analysis and probability*. Cambridge, UK: Cambridge University Press. ISBN 978-0-521-80972-6.
- Fristedt, Bert; Gray, Lawrence (1997). *A Modern Approach to Probability Theory*. New York: Springer Science+Business Media. doi:10.1007/978-1-4899-2837-5 (https://doi.org/10.1007%2F978-1-4899-2837-5). ISBN 978-1-4899-2837-5.
- Grimmett, G.R.; Stirzaker, D.R. (1992). *Probability and random processes* (2nd ed.). Clarendon Press, Oxford. pp. 271–285. ISBN 978-0-19-853665-9.
- Jacobsen, M. (1992). *Videregående Sandsynlighedsregning (Advanced Probability Theory)* (3rd ed.). HCØ-tryk, Copenhagen. pp. 18–20. ISBN 978-87-91180-71-2.
- Ledoux, Michel; Talagrand, Michel (1991). *Probability in Banach spaces*. Berlin: Springer-Verlag. pp. xii+480. ISBN 978-3-540-52013-9. MR 1102015 (https://www.ams.org/mathscinet-getitem?mr=1102015).
- Romano, Joseph P.; Siegel, Andrew F. (1985). *Counterexamples in probability and statistics*. Great Britain: Chapman & Hall. ISBN 978-0-412-98901-8.
- van der Vaart, Aad W.; Wellner, Jon A. (1996). *Weak convergence and empirical processes*. New York: Springer-Verlag. ISBN 978-0-387-94640-5.
- van der Vaart, Aad W. (1998). *Asymptotic statistics*. New York: Cambridge University Press. ISBN 978-0-521-49603-2.
- Williams, D. (1991). *Probability with Martingales*. Cambridge University Press. ISBN 978-0-521-40605-5.
- Wong, E.; Hájek, B. (1985). *Stochastic Processes in Engineering Systems*. New York: Springer–Verlag.

- https://www.ma.utexas.edu/users/gordanz/notes/weak.pdf