Sequences of random variables

March 4, 2022

Sequences of random variables

The weak law of large numbers

Theorem (WLLN)

Let $X_1, X_2, \ldots, X_k, \ldots$ be a sequence of independent and identically distributed random variables with finite expectation m and finite variance.
Set

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad n \geqslant 1.$$

Then, for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - m\right| \geqslant \epsilon\right) = 0$$

The weak law of large numbers

Proof: We know that $\mathbb{E}\left(\overline{X}_n\right) = m$ and $\mathrm{Var}\left(\overline{X}_n\right) = \mathrm{Var}(X_1)/n$.
Applying Chebyshev's inequality,

$$\mathbb{P}\left(\left|\overline{X}_n - m\right| \geqslant \epsilon\right) \leqslant \frac{\mathrm{Var}(X_1)}{n\,\epsilon^2}$$

Therefore,

For all $\epsilon > 0$ the following limit holds:

$$\mathbb{P}\left(\left|\overline{X}_n - m\right| \geqslant \epsilon\right) \to 0 \quad \textit{as } n \to \infty$$

The sequence of random variables $\left\{\overline{X}_n : n \geqslant 1\right\}$ converges in some sense to the limit value $m = \mathbb{E}(X_1)$.

Convergence in probability

Definition

The sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ converges in probability to the random variable X if for all $\epsilon > 0$,

$$\mathbb{P}\left(|X_n - X| \geqslant \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

Notation: $X_n \xrightarrow{\mathbb{P}} X$

Therefore, the sequence of sample means $\left\{\overline{X}_n : n \geqslant 1\right\}$ converges in probability to the common expected value m:

$$\overline{X}_n \xrightarrow{\mathbb{P}} m$$

Example

Let $X_1, X_2, \ldots, X_n, \ldots$ be random variables such that

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$$

$$\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \frac{1}{2n}, \quad n \geqslant 1$$

Let us show that $X_n \xrightarrow{\mathbb{P}} 0$.

Example

We have to prove that $\mathbb{P}(|X_n| \geqslant \epsilon) \to 0$ as $n \to \infty$. Equivalently, we have to prove that

$$\mathbb{P}(|X_n| < \epsilon) \longrightarrow 1 \quad \text{as } n \to \infty$$

- If $\epsilon > 1$, then $\mathbb{P}(|X_n| < \epsilon) = 1$ for all $n \geqslant 1$.
- Otherwise, if $\epsilon \leqslant 1$, then

$$\mathbb{P}(|X_n| < \epsilon) = \mathbb{P}(X_n = 0) = 1 - \frac{1}{n} \longrightarrow 1$$

Probability as the limit of the relative frequency

An important consequence of the WLLN is the following one.

- Let A be an event with probability $\mathbb{P}(A)$.
- In each of n independent repetitions of the random experiment, we observe weather or not A occurs. More precisely, for $1 \leqslant k \leqslant n$, let X_k be the indicator of the event "A happens in the k-th repetition".

Hence, $\mathbb{E}(X_k) = \mathbb{P}(A)$, and, moreover,

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = f_n(A)$$

is the relative frequency of the event A.

Probability as the limit of the relative frequency

Corollary

The sequence of relative frequencies $\{f_n(A) : n \geqslant 1\}$ *converges in probability to* $\mathbb{P}(A)$*:*

$$f_n(A) \xrightarrow{\mathbb{P}} \mathbb{P}(A)$$

That is to say, for all $\epsilon > 0$, $\mathbb{P}(|f_n(A) - \mathbb{P}(A)| \geqslant \epsilon) \to 0$ *as* $n \to \infty$ *or, equivalently,*

$$\mathbb{P}(|f_n(A) - \mathbb{P}(A)| < \epsilon) \to 1 \quad \text{as } n \to \infty$$

The standardized sample mean

Let $X_1, X_2, \dots, X_k, \dots$ be a sequence of independent and identically distributed random variables such that $\mathbb{E}(X_k) = m$ and $\text{Var}(X_k) = \sigma^2$, and let us consider the sequence of the standardized sample means:

$$Z_n = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma} = \frac{(X_1 + \cdots + X_n) - n\,m}{\sqrt{n}\,\sigma}, \quad n \geqslant 1$$

Observe that $\mathbb{E}(Z_n) = 0$, $\text{Var}(Z_n) = 1$, and $\overline{X}_n = (\sigma/\sqrt{n})\, Z_n + m$

The standardized sample mean

If $M_Z(\omega)$ is the characteristic function of $Z_1 = (X_1 - m)/\sigma$, then the characteristic function $M_n(\omega)$ of Z_n is given by

$$M_n(\omega) = \mathbb{E}\left(e^{i\omega Z_n}\right) = \mathbb{E}\left(e^{i\omega \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma}}\right) = \mathbb{E}\left(\prod_{k=1}^{n} e^{i \frac{\omega}{\sqrt{n}} \frac{X_k - m}{\sigma}}\right) = \prod_{k=1}^{n} \mathbb{E}\left(e^{i \frac{\omega}{\sqrt{n}} \frac{X_k - m}{\sigma}}\right) = \left(M_Z\left(\frac{\omega}{\sqrt{n}}\right)\right)^n$$

The standardized sample mean

The first terms of the series expansion of $M_Z(u)$ are:

$$M_Z(u) = 1 + i\,\mathbb{E}(Z)u + \frac{i^2\,\mathbb{E}(Z^2)}{2}u^2 + o(u^2) = 1 - \frac{1}{2}u^2 + o(u^2),$$

where

$$\frac{o(u^2)}{u^2} \to 0 \quad \text{as } u \to 0$$

The standardized sample mean

Therefore,

$$M_n(\omega) = \left(M_Z\left(\frac{\omega}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{1}{2}\left(\frac{\omega}{\sqrt{n}}\right)^2 + o\left(\frac{\omega}{\sqrt{n}}\right)^2\right)^n = \left(1 - \frac{1}{2}\cdot\frac{\omega^2}{n} + o\left(\frac{1}{n}\right)\right)^n$$

The standardized sample mean

Taking logarithms and using that $\ln(1+z) = z + o(|z|)$,

$$\ln M_n(\omega) = n \cdot \ln\left(1 - \frac{1}{2}\cdot\frac{\omega^2}{n} + o\left(\frac{1}{n}\right)\right) = n\left(-\frac{1}{2}\cdot\frac{\omega^2}{n} + o\left(\frac{1}{n}\right)\right) = -\frac{1}{2}\,\omega^2 + \frac{o(1/n)}{1/n} \longrightarrow -\frac{1}{2}\,\omega^2 \quad \text{as } n\to\infty$$

The standardized sample mean

(Remark)

We have proved that

$$M_n(\omega) \to e^{-\omega^2/2}, \quad \text{as } n\to\infty,$$

where the convergence is pointwise for all $\omega \in \mathbb{R}$, and the limit function $M(\omega) = e^{-\omega^2/2}$ is the characteristic function of a standard normal random variable.

This kind of convergence (in terms of characteristic functions) can also be understood in terms of distribution functions. We need a result known as the continuity theorem

The continuity theorem

Let $\{F_n(x) : n \geqslant 1\}$ be a sequence of probability distribution functions, and let $\{M_n(\omega) : n \geqslant 1\}$ be the sequence of the corresponding characteristic functions.

Theorem (Continuity Theorem)

1. *If $M(\omega) = \lim_{n\to\infty} M_n(\omega)$ exists and is continuous at $\omega = 0$, then $M(\omega)$ is the characteristic function of some distribution function $F(x)$ and $F_n(x) \to F(x)$ at each point x where the limit function $F(x)$ is continuous.*
2. *If for some distribution function $F(x)$, with characteristic function $M(\omega)$, we have that $F_n(x) \to F(x)$ at each point x where $F(x)$ is continuous, then $M_n(\omega) \to M(\omega)$ for all $\omega \in \mathbb{R}$.*

The central limit theorem

Therefore, we have the following result.

Theorem (CLT)

Let $X_1, X_2, \dots, X_k, \dots$ be a sequence of independent and identically distributed random variables such that $\mathbb{E}(X_k) = m$ and $\mathrm{Var}(X_k) = \sigma^2$. Let

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma}, \quad n \geqslant 1$$

Then, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{Z_n}(x) = F_{\mathrm{N}(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

Convergence in distribution

Definition

1. We say that a sequence $F_1, F_2, \dots, F_n, \dots$ of distribution functions converges to the distribution function F, written $F_n \to F$, if $F_n(x) \to F(x)$ at each point x where the limit function F is continuous.
2. The sequence $X_1, X_2, \dots, X_n, \dots$ of random variables converges in distribution to the random variable X if $F_{X_n} \to F_X$.

Notation: $X_n \xrightarrow{d} X$
Example. The Central Limit Theorem:

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma} \xrightarrow{d} \mathrm{N}(0,1) \quad \text{as } n \to \infty$$

The central limit theorem

(Remark)

The CLT is sometimes expressed by saying that $\overline{X}_n$ is asymptotically normally distributed with mean m and variance σ^2/n, in the following sense:

$$\lim_{n\to\infty} \frac{\mathbb{P}\left(\overline{X}_n \leqslant x\right)}{\frac{1}{\sqrt{2\pi(\sigma^2/n)}} \int_{-\infty}^{x} e^{-(t-m)^2/(2\sigma^2/n)}} = 1$$

The denominator is $\mathbb{P}(Y_n \leqslant x)$, being $Y_n \sim \mathrm{N}\left(m, \frac{\sigma^2}{n}\right)$.

WLLN versus CLT

Example:

Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of independent random variables, each one uniformly distributed in $[0, 1]$. Hence, $\mathbb{E}(X_n) = 1/2$ and $\mathrm{Var}(X_n) = 1/12$.

By the WLLN we know that, given $\alpha > 0$,

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{2}\right| > \frac{1}{\alpha}\right) \longrightarrow 0 \quad \text{as } n \to \infty$$

WLLN versus CLT

The CLT provides more detailed quantitative information.

$$\mathbb{P}\left(\left|\overline{X}_n - \frac{1}{2}\right| > \frac{1}{\alpha}\right) = 1 - \mathbb{P}\left(-\frac{1}{\alpha} \leqslant \overline{X}_n - \frac{1}{2} \leqslant \frac{1}{\alpha}\right)$$

$$= 1 - \mathbb{P}\left(-\frac{\sqrt{12n}}{\alpha} \leqslant \frac{\overline{X}_n - 1/2}{1/\sqrt{12n}} \leqslant \frac{\sqrt{12n}}{\alpha}\right)$$

$$\approx 2\left(1 - F_{N(0,1)}\left(\frac{\sqrt{12n}}{\alpha}\right)\right)$$

- We have taken into account that $\left(\overline{X}_n - 1/2\right)/\left(1/\sqrt{12n}\right)$ converges in distribution to a standard normal. Thus, its distribution function can be approximated by $F_{N(0,1)}$.

De Moivre-Laplace Theorem

If each X_k is a Bernoulli random variable with parameter p, then

$$S_n = X_1 + X_2 + \cdots + X_n$$

follows a $\text{Bin}(n, p)$ distribution with expected value np and variance npq.

Hence, the CLT implies

Theorem (De Moivre-Laplace)

Let $S_n \sim \text{Bin}(n, p)$. Then

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{d} \text{N}(0, 1) \quad \text{as } n \to \infty$$

A local version of the CLT

Theorem

Let $X_1, X_2, \ldots, X_k, \ldots$ be independent identically distributed random variables with zero mean and unit variance, and suppose that their common characteristic function $M_X(\omega)$ satisfies $\int_{-\infty}^{\infty} |M_X(\omega)|^r \, d\omega < \infty$ for some integer $r \geqslant 1$.

Then the density $f_n(u)$ of

$$U_n = \frac{(X_1 + X_2 + \cdots + X_n)}{\sqrt{n}}$$

exists for $n \geqslant r$ and

$$f_n(u) \to \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad \text{as } n \to \infty, \text{ uniformly on } \mathbb{R}$$

Example: sum of uniform random variables

Suppose that each variable X_i is uniform on $[-\sqrt{3}, \sqrt{3}]$ so that $\mathbb{E}(X_i) = 0$ and $\text{Var}(X_i) = 1$.

The common characteristic function is

$$M_X(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f_X(x) \, dx$$

$$= \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} e^{i\omega x} \, dx = \frac{\sin\left(\sqrt{3}\,\omega\right)}{\sqrt{3}\,\omega}$$

Example: sum of uniform random variables

It can be shown that

$$\int_{-\infty}^{\infty} |M_X(\omega)|^2 \; d\omega = \int_{-\infty}^{\infty} \left(\frac{\sin\left(\sqrt{3}\,\omega\right)}{\sqrt{3}\,\omega} \right)^2 \; d\omega = \frac{\pi}{\sqrt{3}} < \infty,$$

Hence, the sufficient condition of the theorem holds for $r = 2$.
Thus, the density f_n of

$$U_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

exists for all $n \geqslant 1$ and

$$f_n(u) \to \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2}$$

Example: sum of uniform random variables

For instance, let

$$U_3 = \frac{X_1 + X_2 + X_3}{\sqrt{3}}$$

The density function of the variable U_3 can be expressed as

$$f_3(u) = \sqrt{3}\, f(\sqrt{3}\, u),$$

where

$$f(s) = f_X(s) * f_X(s) * f_X(s)$$

Example: sum of uniform random variables

The calculation of $f_3(u)$ gives

$$f_3(u) = \begin{cases} 0, & u < -3 \\ (u+3)^2/16, & -3 \leqslant u < -1 \\ (3-u^2)/8, & -1 \leqslant u < 1 \\ (3-u)^2/16, & 1 \leqslant u < 3 \\ 0, & u \geqslant 3 \end{cases}$$

Example: sum of uniform random variables

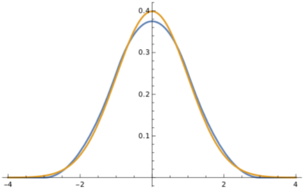

Figure: $f_3(u)$, $\frac{1}{\sqrt{2\pi}} e^{-u^2/2}$

Poisson's theorem

Another application of the continuity theorem is the following result.

Theorem (Poisson)

If $X_n \sim \text{Bin}(n, \lambda/n)$, *then* $X_n \xrightarrow{d} \text{Po}(\lambda)$.

Proof:

$$M_{X_n}(\omega) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} e^{i\omega}\right)^n$$
$$= \left(1 + \frac{\lambda(e^{i\omega} - 1)}{n}\right)^n \to e^{\lambda(e^{i\omega}-1)} \quad \text{as } n \to \infty$$

Convergence in distribution: a result

Theorem

Let $X_1, X_2, \ldots X_n \ldots$ *and* X *be random variables taking nonnegative integer values. A necessary and sufficient condition for* $X_n \xrightarrow{d} X$ *is*

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k) \quad \textit{for all } k \geqslant 0$$

For example, by Poisson's theorem and this result, we have that for all $k \geqslant 0$, if $n \to \infty$, then

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

Convergence in mean square

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let us consider the vector space $\mathcal{H}$ whose elements (vectors) are the random variables X such that $\mathbb{E}(X^2) < \infty$ (i.e., with a finite second order moment).

- We can define an inner product in $\mathcal{H}$ by $\langle X, Y \rangle = \mathbb{E}(XY)$.
- The norm induced by this inner product is $\| X \| = \sqrt{\langle X, X \rangle}$, that is, $\| X \| = \sqrt{\mathbb{E}(X^2)}$
- Moreover, a distance between points (random variables) of $\mathcal{H}$ can be considered:

$$d(X, Y) = \| X - Y \| = \sqrt{\mathbb{E}\left((X - Y)^2\right)}$$

Convergence in mean square

Definition

The sequence $X_1, X_2, \ldots, X_n, \ldots$ converges in mean square to the random variable X if $d(X_n, X) \to 0$ as $n \to \infty$.

Equivalently, $$\mathbb{E}\left((X_n - X)^2\right) \to 0 \quad \text{as } n \to \infty$$

Notation: $X_n \xrightarrow{2} X$

Example: The sequence of sample means $\overline{X}_n$ converges in mean square to the expected value m, because

$$\mathbb{E}\left(\left(\overline{X}_n - m\right)^2\right) = \text{Var}\left(\overline{X}_n\right) = \frac{\text{Var}(X_1)}{n} \to 0$$

Convergence in r-mean

More generally,

Definition

The sequence $X_1, X_2, \ldots, X_n, \ldots$ converges in r-mean to the random variable X if

$$\mathbb{E}\left(|X_n - X|^r\right) \to 0 \quad \text{as } n \to \infty$$

Notation: $X_n \xrightarrow{r} X$

Example

Consider the sequence $X_1, X_2, \ldots, X_n, \ldots$, such that

$$\begin{cases} \mathbb{P}(X_n = 0) = 1 - \dfrac{1}{n} \\ \mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \dfrac{1}{2n} \end{cases}, \qquad n = 1, 2, 3, \ldots$$

Let us prove that $X_n \xrightarrow{r} 0$ for any $r > 0$.

Proof: We have to show that $\mathbb{E}\left(|X_n|^r\right) \to 0$ as $n \to \infty$.

Indeed,

$$\begin{aligned} \mathbb{E}(|X_n|^r) &= 0 \cdot \mathbb{P}(X_n = 0) + 1 \cdot \left(\mathbb{P}(X_n = -1) + \mathbb{P}(X_n = 1)\right) \\ &= \frac{1}{2n} + \frac{1}{2n} \longrightarrow 0 \end{aligned}$$

Almost surely convergence

- Given a sequence $X_1, X_2, \ldots, X_n, \ldots$ of random variables, the sequence $\{X_n(\omega) : n \geqslant 1\}$ is, for each outcome ω of the random experiment, a numeric sequence of real numbers.
- If X is a random variable and $X(\omega)$ is the numerical value taken by X if the outcome is ω, then it makes sense to consider the event

$$\{\omega \in \Omega : X_n(\omega) \to X(\omega) \quad \text{as } n \to \infty\}$$

Almost surely convergence

Definition

The sequence $X_1, X_2, \ldots, X_n, \ldots$ converges almost surely (or with probability 1) to the random variable X if

$$\mathbb{P}\left(\{\omega \in \Omega : X_n(\omega) \to X(\omega) \quad \text{as } n \to \infty\}\right) = 1$$

Notation: $X_n \xrightarrow{a.s.} X$

(Remark)

$X_n \xrightarrow{a.s.} X$ if and only if the set

$$N = \{\omega \in \Omega : X_n(\omega) \not\to X(\omega) \quad \text{as } n \to \infty\}$$

is a null event, that is to say, $\mathbb{P}(N) = 0$.

Strong laws of large numbers

Theorem (SLLN)

Let $X_1, X_2, \ldots, X_k, \ldots$ be a sequence of independent and identically distributed random variables and set

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad n \geqslant 1$$

Then

$$\overline{X}_n \xrightarrow{a.s.} m \quad \text{as } n \to \infty$$

for some constant m, if and only if $\mathbb{E}(|X_1|) < \infty$. In this case, $E(X_1) = m$.

Uniqueness

Theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables. If the sequence converges

- *almost surely,*
- *in probability,*
- *in r-mean,*
- *in distribution,*

then the limiting random variable (distribution) is unique

Uniqueness

For example, let us prove the uniqueness of the limit in the case of almost sure convergence.

Suppose that $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$ and let

$$N_X = \{\omega : X_n(\omega) \not\to X(\omega) \text{ as } n \to \infty\}$$

$$N_Y = \{\omega : X_n(\omega) \not\to Y(\omega) \text{ as } n \to \infty\}$$

So we have that

$$\mathbb{P}(N_X) = \mathbb{P}(N_Y) = 0$$

Uniqueness

Let $\omega \in \overline{N_X} \cap \overline{N_Y} = \overline{N_X \cup N_Y}$. Then

$$|X(\omega) - Y(\omega)| \leqslant |X(\omega) - X_n(\omega)| + |X_n(\omega) - Y(\omega)| \longrightarrow 0$$

So, if $\omega \in \overline{N_X \cup N_Y}$, then $X(\omega) = Y(\omega)$; hence, if $X(\omega) \neq Y(\omega)$, then $\omega \in N_X \cup N_Y$.

Thus

$$\mathbb{P}(X \neq Y) \leqslant \mathbb{P}(N_X \cup N_Y) \leqslant \mathbb{P}(N_X) + \mathbb{P}(N_Y) = 0$$

That is,

$$X = Y \quad \text{with probability } 1$$

Relations between the convergence concepts

Theorem

The following implications hold:

- $\left(X_n \xrightarrow{a.s.} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathbb{P}} X\right)$
- $\left(X_n \xrightarrow{r} X\right) \Longrightarrow \left(X_n \xrightarrow{\mathbb{P}} X\right)$ *for any* $r \geqslant 1$.
- $\left(X_n \xrightarrow{\mathbb{P}} X\right) \Longrightarrow \left(X_n \xrightarrow{d} X\right)$
- *If* $r > s \geqslant 1$, *then* $\left(X_n \xrightarrow{r} X\right) \Longrightarrow \left(X_n \xrightarrow{s} X\right)$

All implications are strict.

Relations between the convergence concepts

With additional hypothesis some converse implications also hold.

Theorem

- *If* c *is a constant, then* $\left(X_n \xrightarrow{d} c\right) \Longrightarrow \left(X_n \xrightarrow{\mathbb{P}} c\right)$
- *If* $X_n \xrightarrow{\mathbb{P}} X$ *and there exists a constant* C *such that* $\mathbb{P}(|X_n| \leqslant C) = 1$ *for all* n, *then* $X_n \xrightarrow{r} X$ *for all* $r \geqslant 1$.
- *If* $P_n(\epsilon) = \mathbb{P}(|X_n - X| > \epsilon)$ *satisfies* $\sum_n P_n(\epsilon) < \infty$ *for all* $\epsilon > 0$, *then* $X_n \xrightarrow{a.s.} X$

Relations between the convergence concepts

For instance, let us consider the proof of the implication

$$\left(X_n \xrightarrow{d} c\right) \Longrightarrow \left(X_n \xrightarrow{\mathbb{P}} c\right)$$

Hence, assume that $X_n \xrightarrow{d} X$ where $X = c$ is a "constant" random variable with distribution function

$$F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geqslant c \end{cases}$$

Relations between the convergence concepts

If $\epsilon > 0$ is a fixed number, we have that

$$\begin{aligned} P(|X_n - c| > \epsilon) &= 1 - \mathbb{P}(c - \epsilon \leqslant X_n \leqslant c + \epsilon) \\ &= 1 - (F_{X_n}(c+\epsilon) - F_{X_n}(c-\epsilon) + \mathbb{P}(X_n = c - \epsilon)) \\ &\leqslant 1 - F_{X_n}(c+\epsilon) + F_{X_n}(c-\epsilon) \longrightarrow 0, \end{aligned}$$

because

$$F_{X_n}(c+\epsilon) \to F_X(c+\epsilon) = 1, \quad F_{X_n}(c-\epsilon) \to F_X(c-\epsilon) = 0$$

Therefore,

$$X_n \xrightarrow{\mathbb{P}} c$$

Example

Consider the sequence $X_1, X_2, \dots, X_n, \dots$ such that

$$\mathbb{P}(X_1 = 1) = 1,$$

$$\mathbb{P}(X_n = 1) = 1 - \frac{1}{n^2}, \quad \mathbb{P}(X_n = n) = \frac{1}{n^2} \quad n \geqslant 2$$

Let us prove that the sequence converges almost surely to 1.

Example

We have that

$$P_n(\epsilon) = \mathbb{P}(|X_n - 1| > \epsilon) = \begin{cases} 0, & n = 1 \\ 1/n^2, & n \geqslant 2 \end{cases}$$

Therefore

$$\sum_{n=1}^{\infty} P_n(\epsilon) = \sum_{n=2}^{\infty} \frac{1}{n^2} < \infty$$

Hence

$$X_n \xrightarrow{a.s.} 1$$

Operations with limits

Theorem

- *If* $X_n \xrightarrow{a.s.} X$ *and* $Y_n \xrightarrow{a.s.} Y$, *then* $X_n + Y_n \xrightarrow{a.s.} X + Y$.
- *If* $X_n \xrightarrow{\mathbb{P}} X$ *and* $Y_n \xrightarrow{\mathbb{P}} Y$, *then* $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.
- *If* $X_n \xrightarrow{r} X$ *and* $Y_n \xrightarrow{r} Y$ *for some* $r > 0$, *then* $X_n + Y_n \xrightarrow{r} X + Y$.

Cramer's Theorem

Theorem (Cramer)

Suppose that $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{\mathbb{P}} a$, *where* a *is a constant. Then*

- $X_n + Y_n \xrightarrow{d} X + a$.
- $X_n - Y_n \xrightarrow{d} X - a$.
- $X_n \cdot Y_n \xrightarrow{d} X \cdot a$.
- $X_n / Y_n \xrightarrow{d} X / a$, *for* $a \neq 0$.

Operations with limits

Theorem

Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$. If X_n and Y_n are independent random variables for all n and, moreover, X and Y are independent, then $X_n + Y_n \xrightarrow{d} X + Y$.

Proof:
It suffices to proof that $M_{X_n+Y_n}(\omega) \to M_{X+Y}(\omega)$ as $n \to \infty$ (pointwise). Thus, it suffices to proof that

$$M_{X_n}(\omega)M_{Y_n}(\omega) \to M_X(\omega)M_Y(\omega)$$

But this is a simple consequence of $M_{X_n}(\omega) \to M_X(\omega)$ and $M_{Y_n}(\omega) \to M_Y(\omega)$.

Continuous functions

Theorem

Let $X_n \xrightarrow{\mathbb{P}} a$, where a is a constant. Suppose, further, that g is a continuous function at point a. Then

$$g(X_n) \xrightarrow{\mathbb{P}} g(a) \quad \text{as } n \longrightarrow \infty.$$

Continuous functions

Proof:
Given $\epsilon > 0$ there exists $\delta > 0$ such that $|g(x) - g(a)| < \epsilon$ if $|x - a| < \delta$. Hence,

$$\{|g(X_n) - g(a)| \geqslant \epsilon\} \subseteq \{|X_n - a| \geqslant \delta\}$$

and so

$$\mathbb{P}(|g(X_n) - g(a)| \geqslant \epsilon) \leqslant \mathbb{P}(|X_n - a| \geqslant \delta)$$

But $\mathbb{P}(|X_n - a| \geqslant \delta) \to 0$, because $X_n \xrightarrow{\mathbb{P}} a$. Therefore $\mathbb{P}(|g(X_n) - g(a)| \geqslant \epsilon) \to 0$, that is

$$g(X_n) \xrightarrow{\mathbb{P}} g(a) \quad \text{as } n \longrightarrow \infty$$

Delta method

Theorem

Let

(a) *$\{a_n : n \geqslant 1\}$ be a sequence of real numbers such that $a_n \to \infty$ as $n \to \infty$, and $a_n \neq 0$ for all n,*

(b) *$\{X_n : n \geqslant 1\}$ be a sequence of random variables an θ be a real number such that $a_n(X_n - \theta) \xrightarrow{d} N(0, \sigma)$,*

(c) *g be a real function with a continuous derivative in an interval that contains θ and such $g'(\theta) \neq 0$.*

Then

$$a_n (g(X_n) - g(\theta)) \xrightarrow{d} \mathrm{N}\left(0, |g'(\theta)|\,\sigma\right)$$

Shorokhod's representation theorem

Theorem

Let $X_1, X_2, \ldots, X_k, \ldots$ and X be random variables such that $X_n \xrightarrow{d} X$ as $n \to \infty$.

Then there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random variables $Y_1, Y_2, \ldots, Y_k, \ldots$ and Y, which map Ω' into $\mathbb{R}$, such that:

- *$Y_1, Y_2, \ldots, Y_k, \ldots$ and Y have the same distribution functions that $X_1, X_2, \ldots, X_k, \ldots$ and X, respectively.*
- *$Y_n \xrightarrow{a.s} Y$.*

Borel-Cantelli lemmas

Given a sequence of events $A_1, A_2, \ldots, A_n, \ldots$, let

$$B_n = \bigcup_{k=n}^{\infty} A_k, \quad n \geqslant 1$$

Notice that $B_1 \supseteq \cdots \supseteq B_n \supseteq B_{n+1} \supseteq \cdots$ is a decreasing sequence of events.

The limit of the sequence $B_1, B_2, \ldots, B_n, \ldots$ is

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Borel-Cantelli lemmas

Definition

Given a sequence of events $A_1, A_2, \ldots, A_n, \ldots$, the event $A^\star$ defined by

$$A^\star = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

is called the limit superior of the sequence and is denoted by $\limsup_n A_n$.

- Notice that $\omega \in A^\star$ if and only if ω belongs to infinitely many of the events A_n. That is, the event $A^\star = \limsup_n A_n$ happens if and only if infinitely many of the events A_n occur.

Borel-Cantelli lemmas

Theorem (Borel-Cantelli lemmas)

Let $A_1, A_2, \ldots, A_n, \ldots$ be a sequence of events and $A^\star$ its limit superior. Then

- *$\mathbb{P}(A^\star) = 0$ if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.*
- *$\mathbb{P}(A^\star) = 1$ if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and the events $A_1, A_2, \ldots$ are independent.*

Corollary (zero-one law)

Let $A_1, A_2, \ldots, A_n, \ldots$ be a sequence of independent events and let $A^\star$ be its limit superior. Then either $\mathbb{P}(A^\star) = 0$ or $\mathbb{P}(A^\star) = 1$ according as $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges or diverges, respectively.

Borel-Cantelli lemmas

For example, the proof of the first Borel-Cantelli lemma is:

$$\mathbb{P}(A^\star) = \mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0$$

because $\sum_n \mathbb{P}(A_n)$ converges by hypothesis.

Examples

Example 1.

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables such that

$$\mathbb{P}(X_n = 0) = \frac{1}{n^2}, \quad \mathbb{P}(X_n = 1) = 1 - \frac{1}{n^2} \quad n \geqslant 1$$

Let $A_n = \{X_n = 0\}$, $n \geqslant 1$.

- Since $\sum_n \mathbb{P}(A_n) < \infty$, the first Borel-Cantelli lemma implies that $\mathbb{P}(A^\star) = 0$. Thus, there is a 0 probability that $\{X_n = 0\}$ happens infinitely often.
- Therefore $\mathbb{P}(X_n = 1$ for all n suficiently large$) = 1$. We deduce that $\lim_{n\to\infty} X_n = 1$ (with probability 1).

Examples

Example 2.

Now, let $X_1, X_2, \ldots, X_n, \ldots$ be independent random variables such that

$$\mathbb{P}(X_n = 0) = \frac{1}{n}, \quad \mathbb{P}(X_n = 1) = 1 - \frac{1}{n} \quad n \geqslant 1$$

Let $A_n = \{X_n = 0\}$, $n \geqslant 1$.

- Since $\sum_n \mathbb{P}(A_n) = \infty$ and the events A_n are independent, the second Borel-Cantelli lemma implies that, with probability 1, infinitely many of the events $A_n = \{X_n = 0\}$ occur.
- Analogously, the probability that $\overline{A}_n = \{X_n = 1\}$ occurs infinitely often is also 1, because $\sum_n \mathbb{P}(\overline{A}_n) = \infty$ and the events $\overline{A}_n$ are independent.
- Hence, with probability 1, $\lim_{n\to\infty} X_n$ does not exist.

Examples

Example 3.

Let a fair coin be tossed infinitely many times and denote by $A^\star$ the event "heads occurs infinitely often". Thus, $A^\star = \limsup_n A_n$, where A_n is the event "the n-th toss lands heads".

- Since the events A_n are independent, the second Borel-Cantelly lemma implies $\mathbb{P}(A^\star) = 1$, because $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} (1/2) = \infty$.
- What happens if $\mathbb{P}(\text{heads}) = \epsilon$, being $\epsilon > 0$ a very small number. Is it still true that "heads will occur infinitely often"?

  Again, $\mathbb{P}(A^\star) = 1$, because $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \epsilon = \infty$.

Examples

Example 4.

Now, let $\mathcal{P}_k$ be a given pattern (of heads and tails) of length k, for instance $\mathcal{P}_6 = \text{HHTTHT}$.

Fact

If a coin with $\mathbb{P}(\text{heads}) = p$, $0 < p < 1$, is tossed infinitely many times, the probability of observing $\mathcal{P}_k$ infinitely often is 1, no matters how large is k.

Examples

Proof: Let A_n be the event "the pattern $\mathcal{P}_k$ is observed between the $[(n-1)k+1]$-th and the $[nk]$-th toss, $n \geqslant 1$".

- If A_n happens infinitely often, then it is true that the pattern $\mathcal{P}_k$ is observed infinitely often.
- The events A_n are independent, because if $i \neq j$, then A_i and A_j correspond to two non-ovelapping sequences of tosses.
- $\mathbb{P}(A_n) = \epsilon$, where $\epsilon = p^r(1-p)^{k-r}$, being r the number of heads in the pattern $\mathcal{P}_k$. If k is large, then ϵ is very small, but we have $\epsilon > 0$.
- By the second Borel-Cantelli lemma, we conclude that $\mathbb{P}(\limsup_n A_n) = 1$, because $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \epsilon = \infty$. So, with probability 1, the pattern $\mathcal{P}_k$ occurs infinitely often.

The infinite monkey theorem

Theorem

A monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type any given text, such as the complete works of William Shakespeare.

In fact, the monkey would almost surely type every possible finite text an infinite number of times.

Proof: see Wiquipedia:

`https://en.wikipedia.org/wiki/Infinite_monkey_theorem`