3 Convergence concepts and martingales

3.1 Convergence of random variables

Convergence of a numerical sequence, written $x_n \to x$ (as $n \to \infty$) or $x = \lim_{n\to\infty} x_n$ is a familiar concept from Analysis. Random variables are *functions*, for which convergence can be understood in various ways.

Let $X_n, n \in \mathbb{Z}_+$, and X be real random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We denote $\mathbb{Z}_+ := \{0, 1, 2, \dots\}$. The sequence (X_n) is said to converge to X

(i) *almost surely*, written $X_n \stackrel{\text{a.s.}}{\to} X$, if $\mathbb{P}(X_n \to X) = 1$,

(ii) *in probability*, written $X_n \stackrel{\mathbb{P}}{\to} X$, if $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$ for every $\epsilon > 0$.

(iii) *in the pth mean* ($p > 0$), written $X_n \stackrel{L^p}{\to} X$, if $\lim_{n\to\infty} \mathbb{E}\,|X_n - X|^p = 0$.

Further convergence concepts address the probability distributions rather then the random variables themselve, that is make sense also when the random variables are defined on different probability spaces. The basic type of convergence of this kind is

(iv) *the weak convergence*, also called *convergence in distribution*, written $X_n \stackrel{d}{\to} X$ (or $X_n \Rightarrow X$), which requires that $\lim_{n\to\infty} \mathbb{E}\, f(X_n) = \mathbb{E}\, f(X)$ for every bounded continuous function f. The concept generalises straightforwardly to random variables with values in arbitrary metric space.

The proximity of distributions is commonly assessed in terms of some 'probability metric' defined as

$$d(X, Y) = \sup |\mathbb{E}\, f(X) - \mathbb{E}\, f(Y)|, \quad f \in \mathcal{D},$$

where the supremum is taken over a given class of functions $\mathcal{D}$ (and the expectations may refer to different probability spaces). In particular,

(v) X_n is said to converge to X in *total variation* if $d_{TV}(X_n, X) \to 0$, where $d_{TV}(X, Y) = \sup_{|f|\le 1} |\mathbb{E}\, f(X) - \mathbb{E}\, f(Y)|$.

Limit theorems of Probability employ all these and other types of convergence.

In particular, the central limit theorem is a statement about convergence in distribution, while the strong law of large numbers asserts convergence in the sense 'almost surely'.

Example Let $X_1, X_2, \dots$ be i.i.d. random variables, with $\mathbb{E}X_1 = \mu < \infty$, $S_n = X_1 + \dots + X_n$. The strong Law of Large Numbers asserts that as $N \to \infty$

$$\frac{S_n}{n} \stackrel{\text{a.s.}}{\to} \mu,$$

which is the convergence in the sense 'almost surely'.

The weak Law of Large Numbers, in many Probability courses proved with the aid of Chebyshev inequality under the assumption $\mathrm{Var}(X_1) = \sigma^2 < \infty$, states that the convergence in probability holds

$$\frac{S_n}{n} \stackrel{\mathbb{P}}{\to} \mu.$$

The Central Limit Theorem

$$\mathbb{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \le x\right) \to \Phi(x)$$

is a result about convergence in distribution.

The types of convergence (i)-(iv) are connected as follows:

$$\begin{aligned} &\stackrel{\text{a.s.}}{\to} \;\Rightarrow\; \stackrel{\mathbb{P}}{\to}, \\ &\stackrel{L^p}{\to} \;\Rightarrow\; \stackrel{\mathbb{P}}{\to}, \\ &\stackrel{\mathbb{P}}{\to} \;\Rightarrow\; \stackrel{d}{\to}. \end{aligned}$$

For constant c, convergence $X_n \stackrel{\mathbb{P}}{\to} c$ is equivalent to $X_n \stackrel{d}{\to} c$. That the other implications do not hold is seen from the following examples.

Example Consider $[0,1]$ with Lebesgue measure, and simple r.v.'s $X_n = n^2 \cdot 1_{[0,1/n]}$. Then $X_n \to 0$ a.s., but $\mathbb{E}\, X_n = n \to \infty$. So there is convergence to $X = 0$ almost surely but not in the mean.

Example Consider $[0,1]$ with Lebesgue measure, and simple r.v.'s $X_{n,k} = 1_{[k2^{-n},(k+1)2^{-n}]}$, $0 \le k \le 2^n - 1$. We have $\mathbb{P}(X_{n,k} \ne 0) = 2^{-n}$ hence $X_{n,k} \stackrel{\mathbb{P}}{\to} 0$ as $n, k \to \infty$ but there is no almost sure convergence.

However, convergence in probability $X_n \stackrel{\mathbb{P}}{\to} X$ implies that we can choose a subsequence (n_k) to have $X_{n_k} \stackrel{\text{a.s.}}{\to} X$ as $k \to \infty$. Convergence in distribution $X_n \stackrel{d}{\to} X$ implies that it is possible to introduce a probability space and define random variables X_n', X' is such a way that $X_n' \stackrel{d}{=} X_n$ (identical in distribution), $X' \stackrel{d}{=} X$ and $X_n' \stackrel{\text{a.s.}}{\to} X'$.

Convergence in total variation is much stronger than the convergence in distribution, so sometimes it is too restrictive for applications as the following example suggests.

Example The total variation distance between discrete X and continuous Y is always 1. Thus, for S_n the number of successes in n Bernoulli trials, the distance between $(S_n - np)/\sqrt{np(1-p)}$ and the limiting normal random variable is equal to 1 for all n.

It can be shown that $X_n \stackrel{L^p}{\to} X$ implies convergence of the pth means $\mathbb{E}\,|X_n|^p \to \mathbb{E}\,|X|^p$.

Convergence almost surely implies convergence of expected values under certain conditions:

Theorem 3.1. (Monotone convergence) *Suppose that* $Y \le X_0 \le X_1 \le \ldots$, $X_n \stackrel{\text{a.s.}}{\to} X$, $\mathbb{E}\, Y > -\infty$. *Then* $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

The case $\mathbb{E}\, X = \infty$ is included.

Theorem 3.2. (Dominated convergence) *Suppose that* $X_n \stackrel{\text{a.s.}}{\to} X$, $|X_n| \le Y$ a.s. *and* $\mathbb{E}\,|Y| < \infty$. *Then* $\mathbb{E}\, X_n \to \mathbb{E}\, X$.

Conclusions of these theorems are also valid if only the convergence in probability is assumed. A more general condition is the following.

Definition 3.3. *A sequence* (X_n) *of random variables on probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called* uniformly integrable *if*

$$\sup_n \mathbb{E}[|X_n| 1_{\{|X_n| > c\}}] \to 0, \quad c \to \infty.$$

Theorem 3.4. *Suppose that* $0 \le X_n \stackrel{\text{a.s.}}{\to} X$, *and* $\mathbb{E}\, X_n < \infty$. *Then* $\mathbb{E}\, X_n \to \mathbb{E}\, X$ *if and only if the sequence* (X_n) *is uniformly integrable.*

3.2 Martingales and their relatives

Fix probability space $(\Omega, \mathcal{F}, \mathbb{P})$ along with a filtration of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$.

Definition 3.5. *A sequence* $(X_n,\ n \in \mathbb{Z}_+)$ *is called* adapted process *if each* X_n *is* $\mathcal{F}_n$*-measurable. In the case* $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ *we call the filtration* natural.

Definition 3.6. *Adapted process* $(X_n,\ n \in \mathbb{Z}_+)$ *with* $\mathbb{E}\,|X_n| < \infty$ *is called*

martingale *if* $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n, \quad n \in \mathbb{Z}_+,$

submartingale *if* $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n, \quad n \in \mathbb{Z}_+,$

supermartingale *if* $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n, \quad n \in \mathbb{Z}_+.$

Intuitively, 'on the average, by one-step look ahead' the martingale does not change, submartingale increases, supermartingale decreases.

Example Suppose $\xi_0, \xi_1, \dots$ are independent with $\mathbb{E}\,\xi_n = 0$. Then the sequence of sums $X_n = \xi_0 + \cdots + \xi_n$ (adapted to $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$) is a martingale. Indeed, using measurability of X_n

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\xi_{n+1}|\mathcal{F}_n] + \mathbb{E}[X_n|\mathcal{F}_n] = \mathbb{E}\,\xi_{n+1} + X_n = X_n.$$

Example Suppose $\xi_0, \xi_1, \dots$ are independent, with $\mathbb{E}\,\xi_n = 1$. Then the sequence of products $X_n = \prod_{j=0}^n \xi_j$ (adapted to $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$) is a martingale.

Example If (X_n) is a martingale and g a convex function (for instance, twice differentiable and satisfying $g'' \geq 0$), then $Y_n = g(X_n)$ is a submartingale. Indeed, using Jensen's inequality and the martingale identity

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[g(X_{n+1})|\mathcal{F}_n] \geq g(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = g(X_n) = Y_n.$$

For instance, squared martingale (X_n^2) is a submartingale.

Example Suppose $\mathbb{E}\,|\xi| < \infty$. Then

$$X_n = \mathbb{E}[\xi|\mathcal{F}_n], n = 0, 1, \dots$$

is a martingale. Indeed, by the tower property

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[\xi|\mathcal{F}_{n+1}]|\mathcal{F}_n]] = \mathbb{E}[\xi|\mathcal{F}_n] = X_n.$$

Important: for martingales $\mathbb{E}\,X_n = \mathbb{E}\,X_0$, the expectation is constant; increases with n for submartingales, decreases for supermartingales.

Definition 3.7. *A process* $(C_n,\ n = 1, 2, \dots)$ *is called* predictable *(or* previsible*) if* C_n *is measurable with respect to* $\mathcal{F}_{n-1}$*. For* $(X_n,\ n = 0, 1, \dots)$ *martingale, the process*

$$Y_n = \sum_{k=1}^n C_k(X_k - X_{k-1}),\ \ n = 1, 2, \dots$$

is called the martingale transform *of* (X_n) *by* (C_n)*. Notation:* $(Y_n) = (C_n) \bullet (X_n)$.

The transform is similar to integral sums known from Calculus, where $X_k - X_{k-1}$ is analogous to dx, and C_k is the value of a to-be-integrated function. It is easy to check that (Y_n) satisfies the martingale identity; however this is a proper martingale in the sense of our definition only if $\mathbb{E}\,|Y_n| < \infty$ (to ensure this, it is enough to require that $|C_n| < K$ for some constant K, for all n). In general, the martingale transforms of a martingale yield processes known as *local martingales*, see below.

Example Consider a coin-tossing game, with probability p to win any given round. You can think of betting on 'red' in roulette, for instance. Mathematically, let ξ_j with values ± 1 be independent, Bernoulli(p), $X_n = \xi_1 + \cdots + \xi_n$, $X_0 = 0$. So (X_n) is a random walk; this is a martingale if $p = 1/2$ (the game is fair), submartingale if $p > 1/2$ (the game is favourable), supermartingale if $p < 1/2$ (the game is unfavourable). A betting strategy is a sequence of bet-sizes, where a bet C_n for the nth game may depend on the outcome of previous rounds, but not on ξ_n, so (C_n) is a predictable process. The capital Y_n after round n is

$$Y_n = Y_{n-1} + C_n \xi_n = \sum_{k=1}^{n} \xi_k C_k = \sum_{k=1}^{n} (X_k - X_{k-1}) C_k.$$

If the game is fair, (Y_n) is a martingale transform of the martingale (X_n). As a special case, which is 'martingale strategy' in the historical meaning of the word, consider the strategy of doubling the stake until a round is won. Let $C_1 = 1$ and $C_n = 2^{n-1}$ in the event $\xi_1 = \ldots \xi_{n-1} = -1$. Now, if $\xi_1 = \cdots = \xi_n = -1,\ \xi_{n+1} = 1$ then the gambler's capital after n rounds is

$$Y_n = -\sum_{i=1}^{n} 2^{i-1} = -(2^n - 1)$$

and after $n+1$ rounds is

$$Y_{n+1} = Y_n + C_{n+1} = -(2^n - 1) + 2^n = 1.$$

Denoting τ the duration of the game, we have $\mathbb{P}(\tau = n) = 2^{-n}$, so $\tau < \infty$ a.s. (although $\mathbb{E}\,\tau = \infty$, so the gambler should have a lot of money to play martingale!).

The study of sub-/supermartingales is largely reduced to the properties of martingales by the vurtue of the following representation.

Theorem 3.8. (Doob-Meyer decomposition) *Let (X_n) be a submartingale. There exist unique martingale (M_n) and predictable process (C_n) such that*

$$X_n = M_n + C_n, \quad n \in \mathbb{Z}_+.$$

Proof. Set $M_0 = X_0, C_0 = 0$, and for $n \geq 1$

$$M_n := M_0 + \sum_{k=0}^{n-1} (X_{k+1} - \mathbb{E}[X_{k+1}|\mathcal{F}_k]), \;\; C_n := \sum_{k=1}^{n-1} (\mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k).$$

Check that (M_n) is a martingale. □

Let (X_n) be square-integrable martingale, that is with $\mathbb{E}\,X_n^2 < \infty$. The function $g(x) = x^2$ is convex, hence (X_n^2) is a submartingale admitting a Doob-Meyer decomposition

$$X_n^2 = M_n + \langle X \rangle_n,$$

where

$$\langle X \rangle_n = \sum_{k=1}^{n} \mathbb{E}[(X_k - X_{k-1})^2 | \mathcal{F}_k]$$

is the random process called *quadratic characteristic* of (X_n). For $n > m$ we have the formula

$$\mathbb{E}[(X_n - X_m)^2 | \mathcal{F}_m] = \mathbb{E}[\langle X \rangle_n - \langle X \rangle_m | \mathcal{F}_m].$$

Example Suppose $\xi_1, \xi_2, \dots$ are independent, with $\mathbb{E}\,\xi_i = 0$, $\mathbb{E}\,\xi_i^2 < \infty$. Then $X_n = \xi_1 + \cdots + \xi_n$ is a martingale, with

$$\langle X\rangle_n = \mathbb{E}\,X_n^2 = \mathrm{Var}\xi_1 + \cdots + \mathrm{Var}\xi_n,$$

which is the variance of X_n in this case.

3.3 Stopping times

In the last doubling-stake example, exiting the game immediately once in $+$ is an instance of stopping time. These are random times adapted to the history. Think of some decision strategy to stop a series of random observations, or exit from a betting game, times to sell/buy assets etc. The decision to stop may depend on the historical data but not on the future observations.

Definition 3.9. Stopping time *is a random variable* τ *with values in* $\{0, 1, 2, \dots, \infty\}$ *such that* $\{\tau = n\} \in \mathcal{F}_n,\ n \in \mathbb{Z}_+$.

Equivalently $\{\tau \le n\} \in \mathcal{F}_n,\ n \in \mathbb{Z}_+$. We say τ is finite if $\mathbb{P}(\tau = \infty) = 0$.

Example For $B \in \mathcal{B}(\mathbb{R})$, $\tau = \min\{n :\ X_n \in B\}$ is a stopping time, called *hitting time* or *the first entrance* time for set B.

If (X_n) adapted process, we define stopped variable

$$X_\tau = \sum_{k=0}^{\infty} X_k \cdot 1_{\{\tau = k\}},$$

which coincides with X_k in the event $\{\tau = k\}$. We do not exclude the possibility $\mathbb{P}(\tau = \infty) > 0$, in which case we need one more variable X_∞ to have X_τ well-defined.

Example Let $\tau_0 = 0$ and, recursively for $k = 1, 2, \dots,\ \tau_k := \min\{n : X_n > X_{\tau_{k-1}}\}$. Then τ_k is a stopping time, which is the time when the kth *record* in the sequence $X_0, X_1, \dots$ occurs. Recall that we say that a record occurs at time n is $X_n > X_0, \dots, X_n > X_{n-1}$.

Write $\wedge$ for 'minimum', e.g. $a \wedge b = \min(a, b)$. If τ_1, τ_2 stopping times, then $\tau_1 \wedge \tau_2$ is a stopping time too. Since a constant n is a stopping time, $\tau \wedge n$ also is.

Proposition 3.10. *Let* $(X_n,\ n \in \mathbb{Z}_+)$ *be a martingale (or sub-, or supermartingale), and let* τ *be a stopping time. The the stopped process*

$$(X_{\tau\wedge n},\ n \in \mathbb{Z}_+)$$

is a martingale (respectively, sub- or supermartingale).

The stopped process is frozen at the value in time τ. For example, in the event $\tau = 3$, the stopped process is $X_0, X_1, X_2, X_3, X_3, X_3, \dots$

To motivate the next definition, recall that the conditional expectation may be sometimes defined for variables with infinite expectation. Suppose X has $\mathbb{E}\,X = \infty$ (an example is X with Pareto density $1/x^2$ on $[1, \infty)$), and let ξ be independent of X, with $\mathbb{E}\,\xi < \infty$ For $Y := X + \xi.$ we have $\mathbb{E}[Y|X] = \xi + X$ is well defined, although $\mathbb{E}\,Y = \infty$.

Definition 3.11. *Adapted process* $(X_n,\ n \in \mathbb{Z}_+)$ *is called* local martingale *if there exists an increasing infinite sequence of finite stopping times* $\tau_1 < \tau_2 < \dots$ *such that each stopped process*

$$(X_{\tau_k\wedge n},\ n \in \mathbb{Z}_+)$$

is a martingale.

Of course every martingale is a local martingale: just take $\tau_k = k$ for 'localising' stopping times, but in general the concept is more general, covering the case when $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ with $\mathbb{E}\,|X_n| = \infty$. In fact:

Theorem 3.12. *Every local martingale (Y_n) can be represented as a martingale transform $(C_n)\bullet(X_n)$ of some martingale (X_n).*

Let τ be a finite stopping time, (X_n) a martingale, so $(X_{\tau\wedge n},\ n \in \mathbb{Z}_+)$ is a martingale, hence $\mathbb{E}\,X_{\tau\wedge n} = \mathbb{E}\,X_0$. Sending $n \to \infty$ we get $\tau \wedge n \overset{\text{a.s.}}{\to} \tau$ thus $X_{\tau\wedge n} \overset{\text{a.s.}}{\to} X_\tau$. But this does not imply $\mathbb{E}\,X_\tau = \mathbb{E}\,X_0$.

Example In the fair coin-tossing game let $\tau = \min\{n : X_n = 1\}$. We now that the random walk (X_n) will reach 1 with certainty, that is τ is finite, and so $X_\tau = 1,\ \mathbb{E}\,X_\tau = 1 \neq 0 = \mathbb{E}\,X_0$. Conditions that guarantee $\mathbb{E}\,X_\tau = \mathbb{E}\,X_0$ are given in the next theorem.

Theorem 3.13. (Doob's optional sampling theorem) *Let (X_n) be a supermartingale, τ a stopping time. Then $\mathbb{E}\,X_\tau \leq \mathbb{E}\,X_0$ if any of the following conditions hold:*

(i) *τ is bounded ($\mathbb{P}(\tau < K) = 1$ for some constant K),*

(ii) *$|X_n| < K$ for all n, and some constant K,*

(iii) *$\mathbb{E}\,\tau < \infty$ and $\mathbb{E}[|X_{n+1} - X_n||\mathcal{F}_n] < K$ for all n, and some constant K,*

(iv) *$\sup_n \mathbb{E}(|X_n| \cdot 1_{\{|X_n|>K\}}) \to 0$ as $K \to \infty$ (uniform integrability).*

If (X_n) is a martingale, any of these conditions ensures that $\mathbb{E}\,X_\tau = \mathbb{E}\,X_0$. Under any of these conditions, if σ is another stopping time with $\sigma \leq \tau$ then $\mathbb{E}\,X_\tau \leq \mathbb{E}\,X_\sigma$ (with equality in the case of martingale).

Example (Gambler's ruin) Consider the coin-tossing game with probability p to win, $q = 1 - p$ to lose a round. A gambler with initially A pounds plays agains a bank with initially B pounds until one of them is ruined. Each round the stake is one pound. In terms of the random walk (X_n), with $X_0 = 0$, the question is about the stopping time $\tau = \min\{n : X_n \in \{-A, B\}\}$. If $X_\tau = -A$ the gambler is ruined, and if $X_\tau = B$ the bank is ruined.

Suppose first the game is fair, $p = q = 1/2$, then (X_n) is a martingale. By the optional sampling theorem, $0 = \mathbb{E}\,X_0 = \mathbb{E}\,X_\tau = -A\,\mathbb{P}(X_\tau = -A) + B\,\mathbb{P}(X_\tau = B)$. Together with $\mathbb{P}(X_\tau = -A) + \mathbb{P}(X_\tau = B) = 1$ this gives

$$\mathbb{P}(X_\tau = -A) = \frac{B}{A+B}, \quad \mathbb{P}(X_\tau = B) = \frac{A}{A+B}.$$

If $p \neq q$, we may consider the martingale

$$Y_n := \left(\frac{q}{p}\right)^{X_n}.$$

By the optional sampling theorem,

$$\mathbb{P}(X_\tau = -A)\left(\frac{q}{p}\right)^{-A} + \mathbb{P}(X_\tau = B)\left(\frac{q}{p}\right)^{B} = 1,$$

whence the probability of gambler's/bank's ruin is

$$\mathbb{P}(X_\tau = -A) = \frac{\left(\frac{p}{q}\right)^{A+B} - \left(\frac{p}{q}\right)^{A}}{\left(\frac{p}{q}\right)^{A+B} - 1}, \quad \mathbb{P}(X_\tau = B) = \frac{\left(\frac{p}{q}\right)^{A} - 1}{\left(\frac{p}{q}\right)^{A+B} - 1}. \tag{1}$$

3.4 Wald's identities

You roll a die until the first 6. What is the expected total of the rolls? Each roll the score is uniformly distributed on $\{1, 2, 3, 4, 5, 6\}$, with mean 3.5. However, if you know that the first 6 occurs at, say, 8th roll, the first 7 scores are uniform on $\{1, 2, 3, 4, 5\}$. Computing the expected total with the formula of total expectation is a challenging task. By Wald's identity, the answer $21 = 3.5 \cdot 6$ is straightforward; here, 6 is the mean of the geometric distribution with parameter $1/6$.

Theorem 3.14. (Wald's identities) *Let $\xi_1, \xi_2, \dots$ be independent, identically distributed random variables with $\mathbb{E}\,|\xi_i| < \infty$, and let τ be a stopping time with $\mathbb{E}\,\tau < \infty$. Then*

$$\mathbb{E}[\xi_1 + \cdots + \xi_\tau] = \mathbb{E}\,\xi_1\,\mathbb{E}\,\tau.$$

If also $\mathbb{E}\,\xi_i^2 < \infty$ then

$$\mathbb{E}[(\xi_1 + \cdots + \xi_\tau) - \tau\,\mathbb{E}\,\xi_1]^2 = \mathrm{Var}\xi_1 \cdot \mathbb{E}\,\tau.$$

Proof. We will prove only the first identity. Consider martingale

$$X_n = \xi_1 + \cdots + \xi_n - n\,\mathbb{E}\,\xi_1$$

with respect to filtration $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, with $X_0 = 0$. By the optional sampling theorem

$$0 = \mathbb{E}\,X_\tau = \mathbb{E}(\xi_1 + \cdots + \xi_\tau) - \mathbb{E}\,\xi_1\,\mathbb{E}\,\tau,$$

Indeed, condition (ii) of the theorem applies, because by independence

$$\mathbb{E}[|X_{n+1} - X_n||\mathcal{F}_n] = \mathbb{E}[|\xi_{n+1} - \mathbb{E}\,\xi_1||\mathcal{F}_n] = \mathbb{E}\,|\xi_{n+1} - \mathbb{E}\,\xi_1| \le 2\,\mathbb{E}\,|\xi_1| < \infty.$$

□

Example Let's apply to the duration of the game in gambler's ruin problem. This is found from $\mathbb{E}\,X_\tau = \mathbb{E}\,\xi_1\,\mathbb{E}\,\tau$ where $\mathbb{E}\,X_\tau$ derives from (1) and $\mathbb{E}\,\xi_1 = p - q$. In the fair game case $\mathbb{E}\,\tau = AB$.

Example In fair coin-tossing (symmetric random walk) let $\tau = \min\{n : X_n = 1\}$ be the first time when the gambler has more rounds won than lost. A martingale argument yields the distribution of τ. Clearly, τ may only assume odd values $1, 3, \dots$ The moment generating function of X_n is

$$\mathbb{E}[e^{\lambda X_n}] = \left(\mathbb{E}[e^{\lambda \xi_1}]\right)^n = \left(\frac{e^\lambda + e^{-\lambda}}{2}\right)^n = (\cosh\lambda)^n,$$

hence

$$M_n := \frac{e^{\lambda X_n}}{(\cosh\lambda)^n}$$

is a martingale. From $\mathbb{E}\,M_{\tau\wedge n} = 1$ and $M_{\tau\wedge n} < e^\lambda$ (for $\lambda > 0$) we get by the optional sampling theorem $\mathbb{E}\,M_\tau = 1$, so

$$\mathbb{E}\left[\frac{e^{\lambda X_\tau}}{(\cosh\lambda)^\tau}\right] = 1.$$

Change variable:

$$z := \frac{1}{\cosh\lambda}, \quad e^{-\lambda} = \frac{1 - \sqrt{1 - z^2}}{z}$$

and recall that $X_\tau = 1$ by definition of τ. Thus we obtain the probability generating function of τ,

$$\mathbb{E}[z^\tau] = \sum_{n=0}^{\infty} \mathbb{P}(\tau = n) z^n = \frac{1 - \sqrt{1 - z^2}}{z}.$$

Expanding the numerator in right-hand side in binomial series yields

$$\mathbb{P}(\tau = 2m-1) = (-1)^{m+1}\binom{1/2}{m} = (-1)^{m+1}\frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-m+1)}{m!}$$

(recall the definition of generalised binomial coefficients).

3.5 Inequalities and martingale convergence

Recall two inequalities from your courses. Markov's inequality: for nonnegative r.v. X and $\lambda > 0$ it holds that $\mathbb{P}(X \geq \lambda) \leq \lambda^{-1}\,\mathbb{E}\,X$. Chebyshev's inequality: for r.v. X with $\mathbb{E}\,X^2 < \infty$ it holds that $\mathbb{P}(|X - \mathbb{E}\,X| \geq \lambda) \leq \lambda^{-2}\mathrm{Var}X$. These generalise to martingales (sub-, super-) in the form of *maximal* inequalities involving the maximum $\max_{k\leq n} X_k$.

It will be convenient tro write $1(A)$ (instead of 1_A) for indicator of event $A \in \mathcal{F}$.

Proposition 3.15. *If (X_n) is a submartingale then for $\lambda > 0$*

$$\mathbb{P}(\max_{k\leq n} X_k \geq \lambda) \leq \lambda^{-1}\,\mathbb{E}[X_n^+ \cdot 1(\max_{k\leq n} X_k \geq \lambda)] \leq \lambda^{-1}\,\mathbb{E}\,X_n^+.$$

If (X_n) is a martingale then

$$\mathbb{P}(\max_{k\leq n} |X_k| \geq \lambda) \leq \lambda^{-2}\,\mathbb{E}\,X_n^2.$$

Proof. For the first inequality, consider stopping time $\tau = \min\{k \leq n : X_k \geq \lambda\}$, with the convention that $\tau = n$ in the event $\max_{k\leq n} X_k < \lambda$. By the optional sampling theorem

$$\begin{aligned}\mathbb{E}\,X_n \geq \mathbb{E}\,X_\tau = \mathbb{E}[X_\tau \cdot 1(\max_{k\leq n} X_k \geq \lambda)] + \mathbb{E}[X_n \cdot 1(\max_{k\leq n} X_k < \lambda)] \geq \\ \lambda\,\mathbb{P}(\max_{k\leq n} X_k \geq \lambda) + \mathbb{E}[X_n \cdot 1(\max_{k\leq n} X_k < \lambda)],\end{aligned}$$

whence

$$\lambda\,\mathbb{P}(\max_{k\leq n} X_k \geq \lambda) \leq \mathbb{E}\,X_n - \mathbb{E}[X_n \cdot 1(\max_{k\leq n} X_k < \lambda)] = \mathbb{E}[X_n \cdot 1(\max_{k\leq n} X_k \geq \lambda)] \leq \mathbb{E}\,X_n^+.$$

For the second inequality, with martingale (X_n), use that (X_n^2) is a submartingale, and apply the first inequality. □

The martingale $X_n = \mathbb{E}[\xi|\mathcal{F}_n]$ can be thought of as a sequence of approximations to random variable ξ, which are updated as the information about ξ increases. Suppose ξ is measurable with respect to $\mathcal{F}_\infty := \sigma(\cup_{n=0}^\infty \mathcal{F}_n)$. It is natural to expect then that the sequence X_n converges to ξ, because ξ is known exactly when the whole infinite information flow gets processed. This intuition is made precise by the martingale convergence theorems.

Theorem 3.16. (Doob's martingale convergence theorem.) *Let $(X_n,\ n \in \mathbb{Z}_+)$ be a submartingale with $\sup_n \mathbb{E}\,|X_n| < \infty$. Then there exists a random variable X_∞ with $\mathbb{E}\,|X_\infty| < \infty$ and such that*

$$X_n \stackrel{\text{a.s.}}{\to} X_\infty, \text{ as } n \to \infty.$$

If the submartingale is uniformly integrable, then also convergence in the mean holds:

$$X_n \stackrel{L^1}{\to} X_\infty, \text{ as } n \to \infty.$$

Example If $\mathbb{E}\,|\xi| < \infty$ then $\mathbb{E}[\xi|\mathcal{F}_n]$ converges to $\mathbb{E}[\xi|\mathcal{F}_\infty]$ almost surely and in the mean.

Example (Galton-Watson Branching process) Let ξ_{nk} be i.i.d. with some discrete distribution on $\mathbb{Z}_+$, let

$$\mu = \mathbb{E}\,\xi_{nk}, \quad \mathrm{Var}\,\xi_{nk} < \infty.$$

The Galton-Watson branching process is defined recursively by setting $Z_0 = 1$

$$Z_n = \xi_{n1} + \xi_{n2} + \cdots + \xi_{n,Z_n}.$$

Think of ξ_{nk} as offspring number of individual k from generation n, then Z_n is the population size in generation n.

Using Wald's identity,

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \mu Z_n,$$

where $\mathcal{F}_n = \sigma(Z_0, \cdots, Z_n)$, and the first equality holds since this is a Markov process. Thus

$$\frac{Z_n}{\mu^n}, \quad n \in \mathbb{Z}_+$$

is a martingale, hence there exists a limit

$$\frac{Z_n}{\mu^n} \stackrel{\text{a.s.}}{\to} W.$$

If $\mu > 1$ (supercritical case) the limit W is a random variable with $W = 0$ in the event of extinction $\cup_{n=1}^\infty \{Z_n = 0\}$, and $W > 0$ on the event $\{Z_n \to \infty\}$. If $\mu \le 1$ the extinction is certain and $W = 0$ a.s.

Exercises

1. Suppose $X_n \stackrel{d}{\to} c$ for constant c. Show that $X_n \stackrel{\mathbb{P}}{\to} c$. Hint: use functions
$$f_\epsilon(x) = (1 - |x - c|\epsilon^{-1})_+$$
to estimate $\mathbb{P}(|X_n - c| \le \epsilon)$ from below.

2. Suppose $\sum_{n=1}^\infty \mathbb{E}\,|X_n| < \infty$. Using Chebyshev inequality and Borel-Cantelli lemma show that $X_n \stackrel{\text{a.s.}}{\to} 0$.

3. Suppose that $X_n \stackrel{\mathbb{P}}{\to} X$. Show that there exists subsequence (n_k) such that $X_{n_k} \stackrel{\text{a.s.}}{\to} X$.

4. Suppose $X_n \stackrel{\mathbb{P}}{\to} Y$ and $X_n \stackrel{\mathbb{P}}{\to} Z$. Prove that $\mathbb{P}(Y \ne Z) = 0$.

5. Show that $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ holds if and only if $\mathbb{E}[X_{n+1} \cdot 1_A] = \mathbb{E}[X_n \cdot 1_A]$ for every $A \in \mathcal{F}_n$.

6. For martingale (X_n), show that $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ for $n \ge m$. What are the analogues for sub- and supermartingales?

7. Let $\xi_1, \xi_2, \ldots$ be r.v.'s with $\mathbb{E}\,|\xi_j| < \infty$ and $\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] = 0$. Show that $X_n = \sum_{k=1}^n \xi_k$ is a martingale (with $X_0 = 0$).

8. Let τ, σ be stopping times. Which of the following random variables are stopping times: $\sigma \vee \tau, \sigma \wedge \tau, \sigma + \tau, \sigma + \tau + 1, \sigma \cdot \tau, \sigma - \tau$? We use $\vee$ for $\max$, $\wedge$ for $\min$.

9. Two dice are rolled until a sum of 7 is thrown. Find the expectation of the sum of scores over all rolls.

10. Let $\xi_1, \xi_2, \ldots$ be i.i.d. with $\mathbb{E}\,\xi_j = 0$, $X_n = \xi_1 + \cdots + \xi_n$, $\tau = \min\{n : X_n \geq 0\}$. Prove that $\mathbb{E}\,\tau = \infty$.

11. Let $X_1, X_2, \ldots$ be independent with $\mathbb{E}\,X_j = 0$. Show that $Y_n = \sum_{1 \leq i < j \leq n} X_i X_j$ is a martingale.

12. Show that for submartingales $(X_n), (Y_n)$ also $(X_n \wedge Y_n)$ is a submartingale.

13. Consider the coin-tossing game with probabilities p and q, starting with $X_0 = 0$. For $\tau = \min\{n : X_n = 1\}$ (where $\tau = \infty$ if no such n exists) find $\mathbb{P}(\tau < \infty)$.

Literature

1. D. Williams, Probability with martingales, CUP 1991.

2. A. Shiryaev, Probability, Springer, 1996.