PROBABILITY 2
LECTURE NOTES

VALENTIN FÉRAY

Contents

Date: December 18, 2019.

Standard notation/terminology: throughout the script,

- r.v. stands for *random variable*: unless specified otherwise, r.v. are real-valued and usually represented by uppercase letters;
- $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space; throughout the script, ω represents an element of Ω; when there is no ambiguity, we write L^p instead of $L^p(\Omega, \mathcal{A}, \mathbb{P})$;
- $\mathbf{1}\{B\}$ is the indicator function of a measurable set B;

- $\mathbb{P}_Y$ is the distribution of the r.v. Y;
- a.s. means "almost surely", i.e. that the event we are speaking of holds with probability 1; additionally "for almost all ω" means for ω in a set of probability 1.

Part A. Conditional expectation

1. Discrete conditioning

Reminder: if A, B are measurable sets with $\mathbb{P}(B) > 0$, then we set $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$.
More generally, if $X \in L^1$ or $X \geq 0$, then we have

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)} = \sum_{k\in \mathrm{Range}(X)} k\,\mathbb{P}(X = k|B).$$

But, what if we want to condition on a *random variable* Y and not on an event B? What is the expectation of X knowing Y? This depends on the value of Y... We will define $\mathbb{E}[X|Y]$ as a function of Y.

Definition 1.1. *Consider r.v. $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to S$, where Y takes values in a* **countable set** *S. Assume $X \in L^1$ or $X \geq 0$. Then the conditional expectation of X knowing Y is defined as $\mathbb{E}[X|Y] = \varphi(Y)$, where, for y in S,*

$$\varphi(y) = \begin{cases} \mathbb{E}[X|Y = y] & \text{if } \mathbb{P}(Y = y) > 0; \\ \textit{undefined} & \text{if } \mathbb{P}(Y = y) = 0. \end{cases}$$

Warning: $\mathbb{E}[X|Y]$ is a *random variable* $\Omega \to \mathbb{R}$, like X, and not a number like $\mathbb{E}[X]$.

Example. Assume $Y \sim \mathsf{Poisson}(\lambda)$ and $X \sim \mathsf{Binomial}(Y, p)$.
If $Y = n$, then $X \sim \mathsf{Binomial}(n, p)$ and $\mathbb{E}[X|Y = n] = n\,p$. Thus $\mathbb{E}[X|Y] = Y\,p$.
Less intuitively, we can also consider $\mathbb{E}[Y|X]$ by computing $\mathbb{E}[Y|X = k]$ for all k

$$\mathbb{E}[Y|X = k] = \sum_{n\geq k} n\,\frac{\mathbb{P}(Y = n, X = k)}{\mathbb{P}(X = k)} = \frac{\sum_{n\geq k} n\,\mathbb{P}(Y = n)\mathbb{P}(X = k|Y = n)}{\mathbb{P}(X = k)}.$$

The denominator is computed as follows:

$$\mathbb{P}(X = k) = \sum_{m\geq k} \mathbb{P}(X = k|Y = m)\mathbb{P}(Y = m) = \sum_{m\geq k} \binom{m}{k} p^k(1-p)^{m^k}\,\frac{\lambda^m e^{-\lambda}}{m!} = \cdots = \frac{(\lambda p)^k e^{-\lambda p}}{k!}.$$

Simplifying we get

$$\mathbb{E}[Y|X = k] = e^{-\lambda+\lambda p} \sum_{n\geq k} n\frac{\lambda^{n-k}(1-p)^{n-k}}{(n-k)!} = \cdots = k + (1-p)\lambda.$$

Comparing with the definition of conditional expectation, we conclude that $\mathbb{E}[Y|X] = X + (1-p)\lambda$.

Remark. (i) Let $S' = \{y \in S | \mathbb{P}(Y = y) = 0\}$ be the set where $\mathbb{E}[X|Y]$ is undefined. **Since S is countable,** we have $\mathbb{P}_Y(S') = \mathbb{P}(Y \in S') = 0$. Therefore φ is well defined $\mathbb{P}_Y$ almost surely, and $\mathbb{E}[X|Y]$ is defined almost surely.

(ii) Call $(A_i)_{i\in I}$ the partition of Ω induced by Y, i.e.

$$\{A_i, i \in I\} = \{Y^{-1}(s), s \in S\}.$$

The conditional expectation $\mathbb{E}[X|Y]$ depends only on Y through the partition $(A_i)_{i\in I}$. For instance we have $\mathbb{E}[X|Y] = \mathbb{E}[X|1 - Y]$.

Proposition 1.2. *Let $X \in L^1$ and $Y : \Omega \to S$ as above. For all bounded functions $g : S \to \mathbb{R}$ we have*

$$(CP_1) \qquad \mathbb{E}\big[g(Y)\,\mathbb{E}[X|Y]\big] = \mathbb{E}\big[g(Y)\,X\big].$$

The same holds for $X \geq 0$ and all nonnegative functions $g : S \to \mathbb{R}$.

Proof. We have

$$\mathbb{E}\big[g(Y)\,\mathbb{E}[X|Y]\big] = \sum_{k\in S}\mathbb{P}(Y=k)g(k)\mathbb{E}[X|Y=k] = \sum_{k\in S} g(k)\mathbb{E}\big[X\,\mathbf{1}\{Y=k\}\big]$$
$$= \sum_{k\in S}\mathbb{E}\big[Xg(Y)\mathbf{1}\{Y=k\}\big] = \mathbb{E}\Big[Xg(Y)\big(\textstyle\sum_{k\in S}\mathbf{1}\{Y=k\}\big)\Big] = \mathbb{E}\big[g(Y)\,X\big].$$

The assumptions ensure that all sums converge (or are sums of nonnegative terms, and hence well defined). □

2. Extension to continuous setting

If Y takes value in a uncountable space, e.g., $\mathbb{R}$, it might be the case that $\mathbb{P}(Y=y)=0$ for all y. We need to define $\mathbb{E}[X|Y]$ differently. We will use (CP_1) and a detour through the theory of Hilbert spaces.

2.1. Hilbert spaces and projections.

Definition 2.1. *A (real) Hilbert space H is a complete normed real vector space, equipped with an inner scalar product s.t. $\langle x,x\rangle = \|x\|^2$ (called Hilbertian scalar product).*

Lemma 2.2 (Cauchy-Schwarz inequality)**.** *Let x,y be element in a Hilbert space H. Then $|\langle x,y\rangle| \le \|x\|\,\|y\|$.*

Hint of proof. Expand $\|x\pm y\|^2 \le (\|x\|+\|y\|)^2$ (which follows from triangular inequality), cancel out terms, and get the above inequality. □

This implies in particular that the scalar product $x,y\mapsto\langle x,y\rangle$ is continuous in both variables.

Definition 2.3. *A subspace L of a Hilbert space H is a closed linear subspace of H.*

Definition 2.4. *Let x be in H and L be a subspace of H. The distance form x to L is $d(x,L)=\inf_{y\in L}\|x-y\|$.*

Theorem 2.5. *Let x be in H and L be a subspace of H. Then there exists a unique y in L such that*

$$\|x-y\| = d(x,L). \tag{1}$$

Note: both the existence and the uniqueness of y need to be proven (no compactness for justifying the existence).

The proof of the theorem uses the following lemma.

Lemma 2.6. *Let x,y,y' be elements of a Hilbert space H and set $z=\frac{y+y'}{2}$. Then*

$$\|y-y'\|^2 = 2\|x-y\|^2 + 2\|x-y'\|^2 - 4\|x-z\|^2.$$

Idea of proof. Expand using scalar products and compare both sides. □

Proof of the theorem. Uniqueness. Let y and y' be elements of L such that $\|x-y\|=\|x-y'\|=d(x,L)$. The above lemma yields

$$\|y-y'\|^2 = 2\|x-y\|^2+2\|x-y'\|^2-4\|x-\tfrac{y+y'}{2}\|^2 = 4d(x,L)^2-4\|x-\tfrac{y+y'}{2}\|^2 \le 0,$$

since $\frac{y+y'}{2}$ is in L. Therefore, $\|y-y'\|^2=0$ and $y=y'$.

Existence. Let (y_n) be a sequence in L s.t. $\|x-y_n\|$ tends to $d(x,L)$. Using the lemma above, we have, for any $n,m\ge 1$

$$\|y_n-y_m\|^2 = 2\|x-y_n\|^2+2\|x-y_m\|^2-4\left\|x-\tfrac{y_n+y_m}{2}\right\|^2$$

Both $\|x-y_n\|$ and $\|x-y_m\|$ tends to $d(x,L)$. Moreover, $\left\|x-\frac{y_n+y_m}{2}\right\|$ is *at least* $d(x,L)$ since $\frac{y_n+y_m}{2}$ is in L. We conclude that the RHS tends to 0 (it should be nonnegative). Hence (y_n) is a Cauchy sequence and converges to some y in L (recall that Hilbert spaces are complete and that L is closed by assumption). By continuity of the norm, we have

$$\|x-y\| = \lim_{n\to\infty}\|x-y_n\| = d(x,L),$$

proving the existence of y satisfying (1). □

Proposition 2.7. *Let x in H and L be a subspace of H and define y as in the above theorem. Then y is the unique element of L such that $x-y$ is in $L^\perp$.*

Terminology: y is called the orthogonal projection of x on L.

Proof. First, we prove $x - y \in L^\perp$. Let h be in L and α in $\mathbb{R}$. We have $y + \alpha h \in L$, which implies $\|x-(y+\alpha h)\|^2 \geq \|x-y\|^2$. Expanding with scalar products, this reduces to

$$-2\alpha\langle x-y,h\rangle + \alpha^2\|h\|^2 \geq 0.$$

Since this has to hold for any α in $\mathbb{R}$, we have $\langle x-y,h\rangle = 0$. This holds for all h in L, meaning that $x-y$ is indeed in $L^\perp$.

Uniqueness: take $y' \in L$ s.t. $x-y'$ is in $L^\perp$. Then $y-y'$ is in L, but also $y-y' = (x-y')-(x-y)$ lies in $L^\perp$. This implies $y-y'=0$ i.e. $y=y'$. □

Corollary 2.8. *let L be a subspace of a Hilbert space H. The "orthogonal projection to L" map is linear.*

Proof. Take x_1, x_2 in H with orthogonal projections y_1, y_2 on L, and let α_1 and α_2 be scalars. We set

$$x_3 = \alpha_1 x_1 + \alpha_2 x_2, \quad y_3 = \alpha_1 y_1 + \alpha_2 y_2.$$

We want to prove that y_3 is the orthogonal projection of x_3 on L. It suffices to prove that y_3 is in L and $x_3 - y_3$ in $L^\perp$, which is straightforward. □

2.2. **Back to conditional expectation.** As usual, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Recall the standard definition of the L^2 space:

$$L^2 = L^2(\Omega,\mathcal{A},\mathbb{P}) = \big\{X \text{ r.v. s.t. } \mathbb{E}\big[X^2\big] < +\infty\big\}/\sim,$$

where $X \sim X'$ if $\mathbb{P}(X \neq X') = 0$ (it is standard and sometimes implicit in probability theory to identify r.v., which differ on a set of probability 0). We equip L^2 with the scalar product $\langle X,Y\rangle = \mathbb{E}[XY]$ (which is well-defined and finite for X,Y in L^2), and with the norm $\|X\|_2 = \sqrt{\mathbb{E}\big[X^2\big]}$. Then L^2 is a normed real vector space with a Hilbertian scalar product. Moreover, we know from the Probability 1 class that L^2 is complete. Hence, it is a Hilbert space.

Consider now in addition a σ-subalgebra $\mathcal{B}$ of $\mathcal{A}$. Then $L^2(\Omega,\mathcal{B},\mathbb{P})$ **is a subset of the Hilbert space** $L^2(\Omega,\mathcal{A},\mathbb{P})$. It is clearly closed by taking linear expectations. Furthermore if X_n is a sequence of $\mathcal{B}$-measurable r.v. converging in L^2 to some $\mathcal{A}$-measurable r.v. X, then a general lemma ensures that a subsequence X_{n_k} is converging a.s. to X. This implies that X is $\mathcal{B}$-measurable since it is an a.s. limit of $\mathcal{B}$-measurable r.v. We have shown that $L^2(\Omega,\mathcal{B},\mathbb{P})$ is a closed subset of the Hilbert space $L^2(\Omega,\mathcal{A},\mathbb{P})$.

Summing up it is a *subspace* in the sense of the previous subsection (closed linear subspace) and we have existence and uniqueness of projections to $L^2(\Omega,\mathcal{B},\mathbb{P})$.

We note that, in L^2, we have

$$(X-X') \perp Z \iff \mathbb{E}\big[(X-X')Z\big] = 0 \iff \mathbb{E}[XZ] = \mathbb{E}[X'Z]$$

Definition 2.9. *Let X be a r.v. in $L^2(\Omega,\mathcal{A},\mathbb{P})$ and $\mathcal{B}$ a σ-subalgebra of $\mathcal{A}$. We define the conditional expectation of X knowing $\mathcal{B}$, denoted $\mathbb{E}[X|\mathcal{B}]$, as the orthogonal projection of X on $L^2(\Omega,\mathcal{B},\mathbb{P})$.*

Equivalently, this is the unique[1] *$\mathcal{B}$-measurable X' in L^2 such that*

$$(CP_2) \qquad \text{for all } Z \text{ in } L^2(\Omega,\mathcal{B},\mathbb{P}),\ \text{we have } \mathbb{E}[XZ] = \mathbb{E}[X'Z].$$

We call (CP_2) the characterizing property.

Warning. $\mathbb{E}[X|\mathcal{B}]$ is a ($\mathcal{B}$-measurable) r.v., **not a number**. It is defined, up to equality on a set of probability 1.

Here are some first properties of the conditional expectation.

Lemma 2.10. *As usual, $(\Omega,\mathcal{A},\mathbb{P})$ is a probability space and $\mathcal{B} \subseteq \mathcal{A}$ be a σ-subalgebra.*

(i) The map $X \mapsto \mathbb{E}[X|\mathcal{B}]$ is linear;
(ii) If is a r.v. in L^2 satisfying $X \geq 0$ a.s., then $\mathbb{E}[X|\mathcal{B}] \geq 0$ a.s.;

[1]Here, and in the sequel, *unique* means up to equality on a set of probability 1.

(iii) If X_1, X_2 are r.v. in L^2 satisfying $X_1 \le X_2$ a.s., then $\mathbb{E}[X_1|\mathcal{B}] \le \mathbb{E}[X_2|\mathcal{B}]$ a.s.

Proof. Item (i) is immediate as orthogonal projections are linear. Item (iii) follows from (i) and (ii); hence, it is sufficient to prove (ii).

The only ingredient we have is the characterizing property. Set $Z = \mathbf{1}\{\mathbb{E}[X|B] < 0\}$. The r.v. is $\mathcal{B}$-measurable and bounded (hence L^2), so that

$$\mathbb{E}[XZ] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]Z\big]. \tag{2}$$

But $XZ \ge 0$ a.s., hence $\mathbb{E}[XZ] \ge 0$. On the other hand, $\mathbb{E}[X|\mathcal{B}]Z \le 0$ a.s. (for all ω, either $\mathbb{E}[X|B] < 0$ or $Z = 0$, by definition of Z). This implies $\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]Z\big] \le 0$.

We conclude that both sides of (2) are equal to 0. This implies $\mathbb{E}[X|\mathcal{B}]Z = 0$ a.s., which might happen only if $\mathbb{E}[X|\mathcal{B}] \ge 0$ a.s (see the definition of Z). □

A particular case: when $\mathcal{B}$ is the "σ-algebra generated by a r.v. Y"

Definition 2.11. *Let $\mathcal{E}$ be a σ-algebra on a base set E and $Y : \Omega \to E$ a r.v. The σ-algebra generated by Y is*

$$\sigma(Y) = \big\{A \in \mathcal{A} | \exists B \in \mathcal{E} \text{ s.t. } Y^{-1}(B) = A\big\} \subseteq \mathcal{A}.$$

Equivalently, this is the smallest σ-algebra of Ω s.t. Y is measurable.

We denote $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

Lemma 2.12. *A r.v. X is $\sigma(Y)$-measurable if and only if $X = f(Y)$ for some measurable function f.*

Proof. Admitted (easy, but technical). □

With this lemma, the definition of $\mathbb{E}[X|Y]$ rewrites as follows: $\mathbb{E}[X|Y]$ **is the unique r.v.** X' **in** L^2 **which writes** $X' = \varphi(Y)$ **and satisfies** $\mathbb{E}[\boldsymbol{XZ}] = \mathbb{E}[\boldsymbol{X'Z}]$ **for all r.v.** $Z = g(Y)$ **in** L^2. This coincides[2] with the notation in the discrete setting, see Proposition 1.2.

2.3. **Extension of nonnegative and L^1 r.v.** In the discrete setting, we have defined the conditional expectation $\mathbb{E}[X|Y]$ whenever X is in L^1 or X takes nonnegative values (same condition as for the standard expectation). In general, we have only define it when X is in L^2 (recall $L^2 \subset L^1$). We now extend the definition to X in L^1 or $X \ge 0$. There is no underlying Hilbert space anymore, so we will proceed by approximation by L^2 r.v.

Theorem 2.13. *Let $X \ge 0$ be a nonnegative r.v. and $\mathcal{B} \subseteq \mathcal{A}$ be a σ-subalgebra. There exists a unique $\mathcal{B}$-measurable r.v. $X' \ge 0$ such that*[3]

$$\text{for all } \mathcal{B}\text{-measurable r.v. } Z \ge 0, \text{ we have } \mathbb{E}[XZ] = \mathbb{E}[X'Z]. \tag{CP_+}$$

Moreover, when additionally X in L^2, we have $X' = \mathbb{E}[X|\mathcal{B}]$

Notation. This r.v. X' is denoted $\mathbb{E}[X|\mathcal{B}]$ and called the conditional expectation of X knowing $\mathcal{B}$ (also when X' is not in L^2).

Proof. Existence. Let $X_n = X \wedge n = \min(X, n)$. For fixed $n \ge 1$, the r.v. X_n is bounded, and hence in L^2. Therefore, $\mathbb{E}[X_n|\mathcal{B}]$ is well-defined.

Since $X_n \le X_{n+1}$ a.s., we have $\mathbb{E}[X_n|\mathcal{B}] \le \mathbb{E}[X_{n+1}|\mathcal{B}]$ a.s. (Lemma 2.10). Hence, for any ω, the sequence $\big(\mathbb{E}[X_n|\mathcal{B}](\omega)\big)_n$ has an almost sure limit, which we denote $X'(\omega)$.

As a pointwise limit of $\mathcal{B}$-measurable functions, X' is $\mathcal{B}$-measurable. Take $Z \ge 0$, $\mathcal{B}$-measurable, and set $Z_n = Z \wedge n$. Using the monotone convergence theorem and the caracterizing property of $\mathbb{E}[X_n|\mathcal{B}]$ we have

$$\mathbb{E}[X'Z] = \lim_{n\to\infty} \mathbb{E}\big[\mathbb{E}[X_n|\mathcal{B}]Z_n\big] = \lim_{n\to\infty} \mathbb{E}\big[X_n Z_n\big] = \exp[XZ].$$

This proves the existence part in the theorem.

[2] In fact, this does not exactly coincide: here, we have variables in L^2, while in the discrete setting, we were working with L^1/nonnegative/bounded r.v.; we address this issue in the next subsection.

[3] The expectations $\mathbb{E}[XZ]$ and $\mathbb{E}[X'Z]$ might be $+\infty$.

Uniqueness. Take X_1' and X_2' both nonnegative and $\mathcal{B}$-measurable, and assume that both satisfy (CP_+). For $0 < a < b$, Consider $Z_{a,b} = \mathbf{1}\{X_1' < a < b < X_2'\}$, which is $\mathcal{B}$-measurable. Using the characterizing property twice, we have $\mathbb{E}[X_1' Z_{a,b}] = \mathbb{E}[X Z_M] = \mathbb{E}[X_2' Z_{a,b}]$. But

$$\mathbb{E}[X_1' Z_{a,b}] \leq a\mathbb{P}(X_1' < a < b < X_2'), \text{ while } \mathbb{E}[X_2' Z_{a,b}] \geq b\mathbb{P}(X_1' < a < b < X_2').$$

Since both are equal, this is possible only if $\mathbb{P}(X_1' < a < b < X_2') = 0$. This holds for all rational numbers $a < b$, so that

$$\mathbb{P}(X_1' < X_2') = \mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}} \{X_1' < a < b < X_2'\}\right) = 0.$$

By symmetry $\mathbb{P}(X_1' > X_2') = 0$. We conclude that $X_1' = X_2'$ a.s.

Proof of $X' = \mathbb{E}[X|\mathcal{B}]$ when both $X \in L^2$ and $X \geq 0$. Let X' satisfying (CP_+) for all nonnegative $\mathcal{B}$ measurable r.v. Z. To show $X' = \mathbb{E}[X|\mathcal{B}]$, we need to show that X' also satisfies $\mathbb{E}[ZX'] = \mathbb{E}[ZX]$ it for (possibly negative) Z in L_2. We write $Z = Z^+ - Z^-$, where $Z^+ = \max(Z, 0)$ and $Z^- = -\min(Z, 0)$. Then

$$\mathbb{E}[ZX'] = \mathbb{E}[Z^+X'] - \mathbb{E}[Z^-X'] = \mathbb{E}[Z^+X] - \mathbb{E}[Z^-X] = \mathbb{E}[ZX],$$

where the middle equation uses (CP_+) for the nonnegative r.v. Z^+ and Z^-. We also note that the condition Z in L^2 ensures that $\mathbb{E}[Z^+X'], \mathbb{E}[Z^-X'] < +\infty$, so that the difference is well-defined. This concludes the proof. □

Theorem 2.14. *Let X be a r.v. in L^1 and $\mathcal{B} \subseteq \mathcal{A}$ be a σ-subalgebra. There exists a unique $\mathcal{B}$-measurable r.v. $X' \in L^1$ such that*

(CP_1) *for all* bounded *$\mathcal{B}$-measurable r.v. Z, we have $\mathbb{E}[XZ] = \mathbb{E}[X'Z]$.*

Moreover, when additionally X in L^2 or $X \geq 0$, we have $X' = \mathbb{E}[X|\mathcal{B}]$.

Notation. This r.v. X' is denoted $\mathbb{E}[X|\mathcal{B}]$ and called the conditional expectation of X knowing $\mathcal{B}$.

Proof. Existence. We write $X = X^+ - X^-$ where $X^+ = \max(X, 0) \geq 0$ and $X^- = -\min(X, 0) \geq 0$. Set $(X')^+ = \mathbb{E}[X^+|\mathcal{B}]$, which is well-defined since $X^+ \geq 0$. Using (CP_+) for $Z = 1$ we have

$$\mathbb{E}[(X')^+] = \mathbb{E}[X^+] \leq \mathbb{E}[|X|] < \infty.$$

In particular, $(X')^+$ is a.s. finite. Similarly $(X')^- := \mathbb{E}[X^-|\mathcal{B}]$ is a.s. finite. We set $X' := (X')^+ - (X')^-$ and we want to check (CP_1).

Let Z be a bounded $\mathcal{B}$-measurable random variable. We decompose $Z = Z^+ - Z^-$ as usual. Then, using four times (CP_+), we get

$$\begin{aligned}\mathbb{E}[X'Z] &= \mathbb{E}[(X')^+Z^+] - \mathbb{E}[(X')^-Z^+] - \mathbb{E}[(X')^+Z^-] + \mathbb{E}[(X')^-Z^-] \\ &= \mathbb{E}[X^+Z^+] - \mathbb{E}[X^-Z^+] - \mathbb{E}[X^+Z^-] + \mathbb{E}[X^-Z^-] = \mathbb{E}[XZ]\end{aligned}$$

This shows the existence of X' satisfying (CP_1). The uniqueness and the statement $X' = \mathbb{E}[X|\mathcal{B}]$ are proved in a similar way as in the previous theorem. □

In both cases, to prove that something is a conditional expectation, we can restrict the "tests" in the characterizing property.

Lemma 2.15. *Let X be a r.v., either nonnegative or in L^1 (recall $L^2 \subset L^1$). Then $\mathbb{E}[X|\mathcal{B}]$ is the unique $\mathcal{B}$-measurable r.v. X' such that*

$(CP_{\mathbf{1}})$ *for all B in $\mathcal{B}$, we have $\mathbb{E}[X\mathbf{1}_B] = \mathbb{E}[X'\mathbf{1}_B]$.*

Proof. Existence is a consequence of the above theorems since $(CP_{\mathbf{1}})$ is weaker than (CP_+) and (CP_1) respectively.

For uniqueness, it is easy to show that such any X' verifying Eq. $(CP_{\mathbf{1}})$ should be nonegative or in L^1, whenever X is nonnegative or in L^1. Furthermore the uniqueness proofs in the above theorem only use the case where $Z = \mathbf{1}_B$, showing uniqueness in this lemma as well. □

3. Ten computation rules for conditional expectation

Basic properties: take X, Y r.v. either nonnegative or in L^1 and $\mathcal{B} \subseteq \mathcal{A}$ be a σ-subalgebra.

(i) $\mathbb{E}[1|\mathcal{B}] = 1$ a.s. and $\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]\big] = \mathbb{E}[X]$.

(ii) $\mathbb{E}[aX + bY|\mathcal{B}] = a\mathbb{E}[X|\mathcal{B}] + b\mathbb{E}[Y|\mathcal{B}]$ a.s.

(iii) If $X \leq Y$ a.s., then $\mathbb{E}[X|\mathcal{B}] \leq \mathbb{E}[Y|\mathcal{B}]$ a.s.

(iv) If Y is $\mathcal{B}$ measurable, then $\mathbb{E}[XY|\mathcal{B}] = Y\mathbb{E}[X|\mathcal{B}]$ a.s. (assuming $XY \geq 0$ or in L^1).
In particular, $\mathbb{E}[Y|\mathcal{B}] = Y$ a.s.

(v) If X is independent from $\mathcal{B}$, then $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

(vi) If $\mathcal{C} \subseteq \mathcal{B}$ is a σ-subalgebra, then

$$\mathbb{E}[X|\mathcal{C}] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]\big|\mathcal{C}\big].$$

(This is called the *tower property.*)

Proof. (i) The r.v. X' defined by $X' = 1$ a.s. clearly satisfies $(CP_\mathbf{1})$ when $X = 1$ a.s., proving $\mathbb{E}[1|\mathcal{B}] = 1$ a.s. The second part is $(CP_\mathbf{1})$ written for $Z = 1$ (or equivalently, $B = \Omega$).

(ii) It is straightforward to check that the RHS fulfills the characterizing property for $aX + bY$.

(iii) Take $Z = \mathbf{1}\{\mathbb{E}[X|\mathcal{B}] > \mathbb{E}[Y|\mathcal{B}]\}$. We have

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]Z\big] = \mathbb{E}[XZ] \leq \mathbb{E}[YZ] = \mathbb{E}\big[\mathbb{E}[Y|\mathcal{B}]Z\big].$$

But, a.s. $Z\mathbb{E}[X|\mathcal{B}] \geq Z\mathbb{E}[Y|\mathcal{B}]$. This implies a.s. equality $Z\mathbb{E}[X|\mathcal{B}] = Z\mathbb{E}[Y|\mathcal{B}]$, i.e., $\mathbb{E}[X|\mathcal{B}] \leq \mathbb{E}[Y|\mathcal{B}]$ a.s. (using the definition of Z).

(iv) We first consider $X, Y \geq 0$. For $Z \geq 0$ $\mathcal{B}$-measurable, we have, using the characterizing property for X (recall that Y is $\mathcal{B}$ measurable, so that ZY is $\mathcal{B}$-measurable):

$$\mathbb{E}\big[ZY\mathbb{E}[X|\mathcal{B}]\big] = \mathbb{E}[ZYX]$$

Thus, $Y\mathbb{E}[X|\mathcal{B}]$ fulfills the characterizing property for YX and we have

$$Y\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[YX|\mathcal{B}].$$

The case X, Y, XY in L^1 follows by setting $X = X^+ - X^-$ and $Y = Y^+ - Y^-$.

(v) If X is independent from $\mathcal{B}$, then, for B in $\mathcal{B}$, we have

$$\mathbb{E}[\mathbf{1}_B X] = \mathbb{E}[\mathbf{1}_B]\,\mathbb{E}[X] = \mathbb{E}\big[\mathbf{1}_B\mathbb{E}[X]\big],$$

proving $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

(vi) Let $C \in \mathcal{C}$, we have, using the charactering property twice w.r.t. $\mathcal{C}$ and $\mathcal{B}$ (since $\mathcal{C} \subseteq \mathcal{B}$, we also have $C \in \mathcal{B}$)

$$\mathbb{E}\Big[\mathbf{1}_C\,\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}\big]\Big] = \mathbb{E}\big[\mathbf{1}_C\mathbb{E}[X|\mathcal{B}]\big] = \mathbb{E}[\mathbf{1}_C X],$$

proving $\mathbb{E}[X|\mathcal{C}] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]\big|\mathcal{C}\big]$. ◻

Intertwining conditional expectations and limits

(vii) (Monotone convergence theorem for conditional expectation)
Assume (X_n) is a sequence of nonnegative r.v. with $X_n \leq X_{n+1}$ a.s. for all n. This implies that $X_n(\omega)$ has a limit for almost all ω (possibly $+\infty$), which we call $X_\infty(\omega)$. Consider a σ-subalgebra $\mathcal{B}$ of $\mathcal{A}$.
We have the following a.s. convergence, as $n \to \infty$:

$$\mathbb{E}[X_n|\mathcal{B}] \to \mathbb{E}[X_\infty|\mathcal{B}]$$

(viii) (Fatou's lemma for condition expectation) Assume (X_n) is a sequence of nonnegative r.v. and $\mathcal{B}$ is a σ-subalgebra of $\mathcal{A}$. We have, for almost all ω in Ω,

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathcal{B}\right](\omega) \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{B}](\omega).$$

(ix) (Dominated convergence for conditional expectation) If X_n tends to X a.s. and there exists $Z \in L^1$ such that $|X_n| \leq Z$ a.s. for all n, then

$$\mathbb{E}[X_n|\mathcal{B}] \to \mathbb{E}[X|\mathcal{B}],$$

a.s. and in L^1.

Proof. (vii) By property (iii) above, we have $\mathbb{E}[X_n|\mathcal{B}] \leq \mathbb{E}[X_{n+1}|\mathcal{B}]$ a.s. This implies that $\mathbb{E}[X_n|\mathcal{B}](\omega)$ has a limit for almost all ω, which we denote $X'_\infty(\omega)$. We still need to identify X'_∞ and $\mathbb{E}[X_\infty|\mathcal{B}]$, i.e. to check that X'_∞ satisfies the characterizing property for X_∞ (we know that X'_∞ as a.s. limit of $\mathcal{B}$ measurable r.v.). We have, for $B \in \mathcal{B}$,

$$\mathbb{E}[\mathbf{1}_B X'_\infty] = \lim_{n\to\infty} \mathbb{E}\big[\mathbf{1}_B \mathbb{E}[X_n|\mathcal{B}]\big] = \lim_{n\to\infty} \mathbb{E}\big[\mathbf{1}_B X_n\big] = \mathbb{E}[\mathbf{1}_B X_\infty],$$

where we used the usual monotone convergence theorem (first and third equalities) and the characterizing property for X_n (second equality). This proves $X'_\infty = \mathbb{E}[X_\infty|\mathcal{B}]$ as wanted.

(viii) We apply (vii) to the increasing sequence of nonnegative r.v. $Y_k = \inf_{n\geq k} X_n$: we have, a.s., as k tends to $+\infty$

$$\mathbb{E}[Y_k|\mathcal{B}] \to \mathbb{E}[Y_\infty|\mathcal{B}],$$

where $Y_\infty(\omega) = \lim Y_k(\omega) = \liminf X_n(\omega)$. The LHS fulfills $\mathbb{E}[Y_k|\mathcal{B}] \leq \mathbb{E}[X_k|\mathcal{B}]$ a.s. (since $Y_k \leq X_k$ a.s. and using item (iii) above). We have that, for almost all ω, $\mathbb{E}[X_k|\mathcal{B}](\omega)$ dominates a sequence tending to $\mathbb{E}[\liminf X_n|\mathcal{B}](\omega)$, i.e.

$$\liminf \mathbb{E}[X_k|\mathcal{B}](\omega) \geq \mathbb{E}[\liminf X_n|\mathcal{B}](\omega),$$

as wanted.

(ix) We apply Fatou's lemma (for conditional expectation) to $Z - X_n$: for almost all ω, we have

$$\liminf \mathbb{E}[Z - X_n|\mathcal{B}](\omega) \geq \mathbb{E}[Z - X|\mathcal{B}](\omega).$$

After simplification, this gives $\limsup \mathbb{E}[X_n|\mathcal{B}](\omega) \leq \mathbb{E}[X|\mathcal{B}](\omega)$. The inequality $\liminf \mathbb{E}[X_n|\mathcal{B}](\omega) \geq \mathbb{E}[X|\mathcal{B}](\omega)$ is proved similarly by applying Fatou's lemma to $Z + X_n$. Both together imply

$$\lim \mathbb{E}[X_n|\mathcal{B}](\omega) = \mathbb{E}[X|\mathcal{B}](\omega). \qquad \square$$

This proves a.s. convergence. The L^1 convergence follows by the usual dominated convergence theorem, since $|\mathbb{E}[X_n|\mathcal{B}]| \leq \mathbb{E}\big[|Z|\,\big|\mathcal{B}\big]$ and

$$\mathbb{E}\big[\mathbb{E}\big[|Z|\,\big|\mathcal{B}\big]\big] = \mathbb{E}\big[|Z|\big] < +\infty.$$

Jensen's inequality for conditional expectations.

(x) Let X be a r.v. and $f : \mathbb{R} \to \mathbb{R}$ a convex function. As usual $\mathcal{B} \subseteq \mathcal{A}$ is a σ-subalgebra. Assume X and $f(X)$ are either nonnegative or in L^1. Then, a.s.,

$$\mathbb{E}[f(X)|\mathcal{B}] \geq f(\mathbb{E}[X|\mathcal{B}]).$$

Proof. We admit the following fact from analysis: since f is convex, there exists sequences (a_n) and (b_n) such that, for all x in $\mathbb{R}$, we have $f(x) = \sup_{n\geq 1}(a_n x + b_n)$.

Then we have

$$\mathbb{E}[f(X)|\mathcal{B}] = \mathbb{E}[\sup_{n\geq 1}(a_n X + b_n)|\mathcal{B}] \geq \sup_{n\geq 1} \mathbb{E}[a_n X + b_n|\mathcal{B}] = \sup_{n\geq 1} a_n \mathbb{E}[X|\mathcal{B}] + b_n = f\big(\mathbb{E}[X|\mathcal{B}]\big),$$

where we used the defining property of a_n and b_n (first and last step), the monotonicity of conditional expectation (i.e. item (iii) above, in the second step) and its linearity (third step). $\square$

An important consequence. Consider the convex function $f(x) = |x|^p$ for $p \geq 1$. Jensen's inequality tells us that for $X \geq 0$ or $X \in L^1$, we have

$$\mathbb{E}[|X|^p|\mathcal{B}] \geq \big|\mathbb{E}[X|\mathcal{B}]\big|^p \text{ a.s.}$$

Taking expectation we get

$$\mathbb{E}[|X|^p] \geq \mathbb{E}\Big[\big|\mathbb{E}[X|\mathcal{B}]\big|^p\Big].$$

In particular, $X \in L^p$ implies that $\mathbb{E}[X|\mathcal{B}]$ is in L^p for any σ-subalgebra $\mathcal{B}$. Also if we have a sequence of r.v. $Y_n \to Y$ in L^p, then $\mathbb{E}[Y_n|\mathcal{B}] \to \mathbb{E}[Y|\mathcal{B}]$ in L^p (apply the above to $X = Y - Y_n$).

4. Some important examples

4.1. Conditional expectation and independence.

Proposition 4.1. *Let X and Y be independent r.v. and h a real-valued function s.t. $h(X,Y)$ is either nonnegative or in L^1. Then*
$$\mathbb{E}\big[h(X,Y)|Y\big] = H(Y),$$
where $H(y) := \mathbb{E}[h(X,y)] = \int h(x,y)\mathbb{P}_X(dx)$.

Informally, **when X and Y are independent**, conditioning on Y means taking only X at random.

Proof. Assume $h(X,T) \geq 0$ (the proof in the L^1 case being similar). Let $Z = g(Y)$ be a nonnegative $\sigma(Y)$-measurable r.v.[4], we have, using the independence of X and Y,

$$\mathbb{E}\big[h(X,Y)g(Y)\big] = \iint g(y)h(x,y)\mathbb{P}_X(dx)\mathbb{P}_Y(dy) = \int \left(\int h(x,y)\mathbb{P}_X(dx)\right) g(y)\mathbb{P}_Y(dy)$$
$$= \int H(y)g(y)\mathbb{P}_Y(dy) = \mathbb{E}\big[H(Y)g(Y)\big].$$

Hence $H(Y)$, which is $\sigma(Y)$-measurable, satisfies the characterizing property for $h(X,Y)$. We conclude that $\mathbb{E}\big[h(X,Y)|Y\big] = H(Y)$, as wanted. ☐

Remark. We did not make precise the spaces in which X and Y are taking value; it can be any spaces E and F (equipped with a σ-algebra, s.t. X,Y r.v. has a sense). Then the domain of h should be $E \times F$.

The results also holds true more generally if we condition on any σ-algebra $\mathcal{B}$ s.t. X is independent from $\mathcal{B}$ and Y $\mathcal{B}$ measurable.

4.2. Density case.

Proposition 4.2. *Let X and Y be real-valued r.v. with joint distribution $p(x,y)dxdy$. Let $q(y) = \int_{\mathbb{R}} p(x,y)dx$ denote the density of Y. For x,y in $\mathbb{R}$, we define*
$$\overline{p}(x,y) = \begin{cases} \frac{p(x,y)}{q(y)} & \text{if } q(y) > 0; \\ 0 & \text{otherwise.} \end{cases}$$
Then for any measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $h(X)$ is either nonnegative or in L^1, we have
$$\mathbb{E}\big[h(X)|Y\big] = \int_{\mathbb{R}} h(x)\overline{p}(x,Y)dx. \tag{3}$$

Informally, conditionally on Y, the r.v. X has density $\overline{p}(x,Y)$. The formula giving $\overline{p}(x,Y)$ is the same as for conditional probability in the discrete setting, replacing probabilities by densities.

Proof (in class, we assume $q(y) > 0$ for all y for simplicity). We check that the RHS of (3), satisfies the characterizing property for $h(X)$. Assuming $h(X)$ is nonnegative (the L^1 case is similar), we let $Z = g(Y)$ be a nonnegative $\sigma(Y)$-measurable random variable. On one side, we have
$$\mathbb{E}\big[h(X)g(Y)\big] = \int_{\mathbb{R}^2} h(x)g(y)p(x,y)dxdy. \tag{4}$$
On the other side, setting $\varphi(y) = \int_{\mathbb{R}} h(x)\overline{p}(x,y)dx$, we can write
$$\mathbb{E}\big[\varphi(Y)g(Y)\big] = \int_{\mathbb{R}} \varphi(y)g(y)q(y)dy = \int_{\mathbb{R}^2} h(x)\overline{p}(x,y)g(y)q(y)\mathbf{1}\{q(y)>0\}dxdy,$$
where, in the second equality, we added the indicator $\mathbf{1}\{q(y)>0\}$ (this does not change the integrand, since the integrand is zero when $q(y) = 0$ anyway) and we substituted $\varphi(y)$ by its defining formula. Furthermore, we implicitly used the fact that integrand is nonnegative to rewrite successive integrals on x and y as a joint integral on (x,y) (Fubini-Tonelli theorem). When $q(y) > 0$, we have $\overline{p}(x,y)q(y) = p(x,y)$ so that
$$\mathbb{E}\big[\varphi(Y)g(Y)\big] = \int_{\mathbb{R}^2} h(x)g(y)p(x,y)\mathbf{1}\{q(y)>0\}dxdy. \tag{5}$$

[4]From Lemma 2.12, we know that all $\sigma(Y)$-measurable r.v. are of the form $g(Y)$. We will not recall that further in the sequel.

But $q(y)=0$ implies that $p(x,y)=0$ for almost all x (see the definition of $q(y)$ and recall that $p(x,y)\geq 0$), so that

$$\int_{\mathbb{R}^2} h(x)g(y)p(x,y)\mathbf{1}\{q(y)=0\}dxdy = 0. \tag{6}$$

Comparing (4), (5) and (6), we conclude that $\mathbb{E}\big[h(X)g(Y)\big] = \mathbb{E}\big[\varphi(Y)g(Y)\big]$. Since this holds for all nonnegative $\sigma(Y)$-measurable r.v. $Z=g(Y)$ and since furthermore $\varphi(Y)$ is $\sigma(Y)$-measurable, we have $\varphi(Y)=\mathbb{E}\big[h(X)|Y\big]$, as wanted. □

4.3. **Gaussian vectors.** *Reminder/scratch course on Gaussian vectors.* Fix a dimension $d\geq 1$. Let $\vec{\mu}$ be a vector in $\mathbb{R}^d$, called *mean vector*, and K be a $d\times d$ symmetric semi-definite matrix, called *covariance matrix.* Then the Gaussian multivariate distribution $\mathcal{N}(\vec{\mu},K)$ is defined by its characteristic transform: if $\vec{X}\sim\mathcal{N}(\vec{\mu},K)$, then, for all $\vec{\zeta}$ in $\mathbb{R}^d$,

$$\mathbb{E}\big[\exp(i\langle\vec{\zeta},\vec{X}\rangle)\big] = \exp\big(i\langle\vec{\zeta},\vec{\mu}\rangle - \tfrac{1}{2}\vec{\zeta}^t K\vec{\zeta}\big).$$

In particular, for $1\leq j,\ell\leq d$,

$$\mathbb{E}[X_j]=\mu_j,\quad \mathrm{Cov}(X_j,X_\ell)=K_{j,\ell}.$$

Some properties:

- $\vec{X}$ is a multivariate Gaussian vector if and only if all linear combinations $\sum_{j=1}^d \alpha_j X_j$ ($\alpha_j\in\mathbb{R}$)are (univariate) Gaussian. In particular $\vec{X}$ is a multivariate Gaussian implies that all X_j are Gaussian but the converse is not true.
- If $\vec{X}$ is a multivariate Gaussian vector and if $\mathrm{Cov}(X_j,X_\ell)=0$ for some j and ℓ, then X_j and X_ℓ are independent. (More generally, if $\{j_1,\dots,j_s\}$ and $\{\ell_1,\dots,\ell_t\}$ are sets such that $\mathrm{Cov}(X_{j_p},X_{\ell_q})=0$ for all $p\leq s$ and $q\leq t$, then the sets of r.v. $\{X_{j_1},\dots,X_{j_s}\}$ and $\{X_{\ell_1},\dots,X_{\ell_s}\}$ are independent.)

Proposition 4.3. *Let $(Y_1,\dots,Y_d,X)$ be a centered Gaussian vector. Let $\hat{X}$ be the orthogonal projection*[5] *of X on* $\mathrm{Span}(Y_1,\dots,Y_d)$*, we have*

$$\mathbb{E}\big[X|(Y_1,\dots,Y_d)\big]=\hat{X}.$$

Proof. For any $j\leq d$, we have

$$\mathrm{Cov}(X-\hat{X},Y_j)=\mathbb{E}\big[(X-\hat{X})\,Y_j\big]=0, \tag{7}$$

where the first equality uses that the variables are all centered and the second that $\hat{X}$ is the orthogonal projection of X on $\mathrm{Span}(Y_1,\dots,Y_d)$.

Note that $(X-\hat{X},Y_1,\dots,Y_d)$ is a Gaussian vector: indeed, all linear combinations of its coordinates are linear combinations of those of $(X,Y_1,\dots,Y_d)$ and hence Gaussian. Therefore the vanishing of covariances in (7) implies that $X-\hat{X}$ is independent from $\{Y_1,\dots,Y_d\}$. We have

$$\mathbb{E}\big[X|(Y_1,\dots,Y_d)\big]=\mathbb{E}\big[X-\hat{X}|(Y_1,\dots,Y_d)\big]+\mathbb{E}\big[\hat{X}|(Y_1,\dots,Y_d)\big]=\mathbb{E}\big[X-\hat{X}\big]+\hat{X}=\hat{X},$$

where the second equality uses computation rules (iv) and (v) p. 8 (the first uses linearity of conditional expectation; the last that variables are centered). □

Remark. In fact, we can prove that the "conditional density/distribution of X knowing $(Y_1,\dots,Y_d)$" is that of a Gaussian centered in $\sum_{j=1}^d \lambda_j Y_j$. More precisely, for all $h:\mathbb{R}\to\mathbb{R}$ such that $h(X)$ is either nonnegative or in L^1,

$$\mathbb{E}\big[h(X)|(Y_1,\dots,Y_d)\big]=\int_{\mathbb{R}} h(x)\tfrac{1}{\sigma\sqrt{2\pi}}\exp\big(\tfrac{-(x-\mu)^2}{2\sigma^2}\big)dx,$$

where $\mu=\hat{X}$ and $\sigma^2=\mathbb{E}\big[(X-\hat{X})^2|(Y_1,\dots,Y_d)\big]$. The result also extends to $\mathbb{E}\big[h(X_1,\dots,X_c)|(Y_1,\dots,Y_d)\big]$, assuming that $(X_1,\dots,X_c,Y_1,\dots,Y_d)$ is a Gaussian vector.

[5] Since X and the Y_j are in L^2, we use the same scalar product as before, namely $\langle Z,Z'\rangle=\mathbb{E}(ZZ')$. Note however that, here, we project on the **linear** space spanned by the Y_j and not on the much bigger space $L^2(\Omega,\sigma(Y_1,\dots,Y_d),\mathbb{P})$ as in the general L^2 case.

Part B. Martingales

5. Basics

Definition 5.1. *A filtered probability space is a quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 1}, P)$ where $(\Omega, \mathcal{F}, P)$ is a probability space and $(\mathcal{F}_n)_{n\geq 1}$ a sequence of σ-algebra of Ω such that*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}.$$

The sequence $(\mathcal{F}_n)_{n\geq 1}$ is called a filtration and a sequence $(X_n)_{n\geq 0}$ of random variables such that each X_n is $\mathcal{F}_n$-measurable is called suited to the filtration $(\mathcal{F}_n)_{n\geq 1}$ or simply a random process (if the filtration is clear from the context).

A sequence $(X_n)_{n\geq 0}$ of random variables such that each X_n is $\mathcal{F}_{\mathbf{n-1}}$-measurable is called predictable.

Remark. This implies that $\mathcal{F}_n \supseteq \sigma(X_0, X_1, \cdots, X_n)$ (the latter is by definition the smallest σ-algebra that makes X_0,...,X_n measurable). A standard situation is $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$, but none of the results below need this assumption.

Throughout this chapter, we assume that we are working in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 1}, P)$.

Definition 5.2. *Let $(X_n)_{n\geq 0}$ be a random process such that X_n is in L^1 for all $n \geq 0$. Then X_n is called a martingale if for any $n \geq 0$, we have*

$$\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] = X_n \ \textit{a.s.} \tag{8}$$

Eq. (8) is called the martingale property.

Example. (i) Let $(Y_i)_{i\geq 1}$ be independent centered integrable r.v. Set $X_n = Y_1 + \cdots + Y_n$ and $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ for every $n \geq 1$. For each $n \geq 1$, the r.v. X_n is a sum of r.v. in L^1 and hence in L^1. We have, a.s.,

$$\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] = \mathbb{E}\big[Y_1 + \cdots + Y_n + Y_{n+1}|\mathcal{F}_n\big] = Y_1 + \cdots + Y_n + \mathbb{E}[Y_{n+1}] = X_n + 0 = X_n,$$

where in the second equality we have used that $\mathbb{E}[Y|\mathcal{B}] = Y$ is Y is $\mathcal{B}$-measurable (this applies to Y_1,...,Y_n) and that $\mathbb{E}[Y|\mathcal{B}] = \mathbb{E}[Y]$ is Y is independent from $\mathcal{B}$ (this applies to Y_{n+1}).

We conclude that X_n is a martingale.

(ii) Let $(T_i)_{i\geq 1}$ be independent unbiased coin tosses, i.e. r.v. with distribution

$$\mathbb{P}[T_i = \mathsf{head}] = \mathbb{P}[T_i = \mathsf{tail}] = 1/2.$$

We let $\mathcal{F}_n = \sigma(T_1, \ldots, T_n)$ and $(A_n)_{n\geq 1}$ be a predictable r.v. with respect to T_n. Finally, we set,

$$Y_i = A_i \mathbf{1}\{T_i = \mathsf{head}\} - A_i \mathbf{1}\{T_i = \mathsf{tail}\} \text{ and } X_n = \sum_{i=1}^{n} Y_i.$$

Gambling interpretation: at each time $i \geq 1$, a gambler bets A_i units on “head”. Then Y_i represents its (algebraic) winnings in the i-th round and X_n its cumulative winnings until time n.

At round i, the gambler can choose what he bets (A_i) depending on the outcome of the previous coin tosses T_1,..., T_{i-1}, but he does not know T_i, T_{i+1}, ...This explains the predictability hypothesis: A_i should be $\mathcal{F}_{i-1}$ measurable.

We now prove that $(X_n)_{n\geq 0}$ is a martingale. Obviously, each Y_i is in L^1 and hence X_n is in L^1 for any $n \geq 0$. We compute

$$\mathbb{E}\big[Y_{n+1}|\mathcal{F}_n\big] = A_{n+1}\mathbb{E}\big[\mathbf{1}\{T_{n+1} = \mathsf{head}\}\big] - A_{n+1}\mathbb{E}\big[\mathbf{1}\{T_{n+1} = \mathsf{tail}\}\big] = 0 \text{ a.s.},$$

which implies

$$\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] = \mathbb{E}\big[X_n|\mathcal{F}_n\big] + \mathbb{E}\big[Y_{n+1}|\mathcal{F}_n\big] = X_n + 0 = X_n.$$

This proves the martingale property.

$\rightarrow$ Martingales are models for cumulative winnings in a *fair* game.

(iii) Take $X \in L^1$ and let $X_n = \mathbb{E}[X|\mathcal{F}_n]$. Then X_n is in L^1 and satisfies the martingale property (because of the tower rule for conditional expectation). Therefore $(X_n)_{n\geq 0}$ is a martingale; martingales constructed this way are called *closed*.

Proposition 5.3. *Let $(X_n)_{n\geq 0}$ be a martingale. Then*

(i) $\mathbb{E}[X_n]$ *is independent of* n*;*
(ii) for any $n \geq p$*, we have* $\mathbb{E}[X_n|\mathcal{F}_p] = X_p$*.*

Proof. This is a straightforward application of the properties of conditional expectation. □

6. Stopping times

6.1. Definitions.

Definition 6.1. *A random variable* $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ *is called a* <u>*stopping time*</u> *if for any* $n \geq 0$*, we have* $\{T \leq n\} \in \mathcal{F}_n$*.*

Equivalently, T is a stopping time if and only if, for any $n \geq 0$, $\{T = n\}$ is in $\mathcal{F}_n$.

Gambling interpretation: T is the time at which a gambler stops playing (i.e. $T = n$ means that he stops *after* round n). The decision to stop playing at time n is taken knowing what happens until time n and not after, i.e. is $\mathcal{F}_n$-measurable.

Definition 6.2. *Let* $(X_n)_{n\geq 0}$ *be a random process and* T *a stopping time. We define the r.v.* X_T *as follows: for* ω *in* Ω*,*

$$X_T(\omega) = \begin{cases} X_{T(\omega)}(\omega) & \text{if } T(\omega) < \infty; \\ \mathsf{undefined} & \text{if } T(\omega) = \infty. \end{cases}$$

It is easy to check that X_T is indeed $\mathcal{F}$-measurable.
Gambling interpretation: X_T is the winning of the gambler when he stops playing.

6.2. Martingale and stopping times.

If n is an integer and T a stopping time, then we set $n \wedge T(\omega) = \min(n, T(\omega))$. This defines a stopping time and we have the following important stability result.

Proposition 6.3. *Let* $(X_n)_{n\geq 0}$ *be a martingale and* T *a stopping time. Then* $(X_{n\wedge T})_{n\geq 0}$ *is a martingale.*

Proof. We can write

$$X_{n\wedge T} = \sum_{i=0}^{n-1} X_i \mathbf{1}\{T = i\} + X_n \mathbf{1}\{T \geq n\},$$

showing that $X_{n\wedge T}$ is $\mathcal{F}_n$-measurable and in L^1.
We now check the martingale property. Let B be $\mathcal{F}_n$ measurable,

$$\mathbb{E}\big[X_{(n+1)\wedge T}\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_T \mathbf{1}\{T \leq n\}\mathbf{1}\{B\}\big] + \mathbb{E}\big[X_{n+1}\mathbf{1}\{T > n\}\mathbf{1}\{B\}\big].$$

Since $\{T > n\} \cap B$ is in $\mathcal{F}_n$ and since $\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] = X_n$ a.s. (martingale property), we have

$$\mathbb{E}\big[X_{n+1}\mathbf{1}\{T > n\}\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_n \mathbf{1}\{T > n\}\mathbf{1}\{B\}\big].$$

Summing up,

$$\begin{aligned} \mathbb{E}\big[X_{(n+1)\wedge T}\mathbf{1}\{B\}\big] &= \mathbb{E}\big[X_T \mathbf{1}\{T \leq n\}\mathbf{1}\{B\}\big] + \mathbb{E}\big[X_n \mathbf{1}\{T > n\}\mathbf{1}\{B\}\big] \\ &= \mathbb{E}\big[X_{n\wedge T}\mathbf{1}\{B\}\big]. \end{aligned}$$

Since this holds for any B in $\mathcal{F}_n$ and since $X_{n\wedge T}$ is $\mathcal{F}_n$ measurable, we have

$$\mathbb{E}\big[X_{(n+1)\wedge T}|\mathcal{F}_n\big] = X_{n\wedge T}.$$

This concludes the proof that $(X_{n\wedge T})_{n\geq 0}$ is a martingale. □

6.3. The optional stopping theorem.

Theorem 6.4. *Let T be a bounded stopping time, i.e. there exists $M > 0$ s.t. $\mathbb{P}(T > M) = 0$. Let $(X_n)_{n\geq 0}$ be a martingale. Then X_T is in L^1 and $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

Gambling interpretation: in a fair game and in finite time, one cannot have in expectation more at the end than at the beginning (whatever strategy one uses, and whenever one decides to stop).

Proof. Since T is bounded by M, we have $X_T = X_{M\wedge T}$ a.s. We have seen in the proof of Proposition 6.3 that $X_{M\wedge T}$ is in L^1, so that X_T is in L_1.

Applying Proposition 5.3, item (i) to the martingale $(X_{n\wedge T})_{n\geq 0}$, we have

$$\mathbb{E}[X_T] = \mathbb{E}[X_{M\wedge T}] = \mathbb{E}[X_{0\wedge T}] = \mathbb{E}[X_0]. \qquad \square$$

Remark. The hypothesis "T bounded" is necessary in the above theorem. Take $X_n = Y_1 + \cdots + Y_n$, where Y_i are i.i.d. r.v. with $\mathbb{P}[Y_i = 1] = \mathbb{P}[Y_i = -1] = 1/2$. We take the stopping time $T = \inf\{n \geq 0 \text{ s.t. } X_n = -1\}$ (why is it a stopping time? see exercises). We will see later in the lecture that $T < \infty$ a.s. Obviously $X_T = -1$ a.s., which implies $\mathbb{E}[X_T] = -1 \neq \mathbb{E}[X_0] = 0$.

7. Sub- and super-martingales

7.1. Definition and some constructions.

Definition 7.1. *A random process $(X_n)_{n\geq 0}$ with $X_n \in L^1$ (for all $n \geq 0$) is called a submartingale (resp. a supermartingale) if for any $n \geq 0$, we have*

$$\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] \geq X_n \textit{ a.s. (resp. } \mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] \leq X_n \textit{ a.s.).}$$

We note that $(X_n)_{n\geq 0}$ is a supermartingale if and only if $(-X_n)_{n\geq 0}$ is a submartingale. We will therefore restrict our attention to submartingales; all results for submartingales have analogues for supermartingales using this simple transformation.

If $(X_n)_{n\geq 0}$ is a submartingale, we have

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

Warning. I personally find the sense of the inequalities unintuitive: *sub*martigales are *nondecreasing* in expectation.

Lemma 7.2. *(i) Let $(X_n)_{n\geq 0}$ be a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. Assume $\varphi(X_n)$ is in L^1 for all $n \geq 0$. Then $(\varphi(X_n))_{n\geq 0}$ is a submartingale.*

*(ii) Let $(X_n)_{n\geq 0}$ be a sub*martingale *and $\varphi : \mathbb{R} \to \mathbb{R}$ be an* nondecreasing *convex function. Assume $\varphi(X_n)$ is in L^1 for all $n \geq 0$. Then $(\varphi(X_n))_{n\geq 0}$ is a submartingale.*

Proof. (i) Using Jensen's inequality and the martingale property for X_n, we hvave

$$\mathbb{E}\big[\varphi(X_{n+1})|\mathcal{F}_n\big] \geq \varphi\Big(\mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big]\Big) = \varphi(X_n),$$

showing that $(\varphi(X_n))_{n\geq 0}$ is a submartingale.

(ii) Similar.

$\square$

In particular, if X_n is a martingale, $|X_n|$ and $X_n^+ = \max(X_n, 0)$ are submartingales. If furthermore X_n is in L^2, then X_n^2 is a submartingale.

7.2. **Doob's decomposition.**

Theorem 7.3. *Let $(X_n)_{n\geq 0}$ be a random process with X_n in L^1 (for all $n \geq 0$). Then there exist a martingale $(M_n)_{n\geq 0}$ and a predictable process $(A_n)_{n\geq 0}$ such that*

$$X_n = X_0 + M_n + A_n \text{ and } M_0 = A_0 = 0 \text{ a.s.}$$

Moreover,

- *this decomposition is unique up to a.s. equality;*
- *A_n is a.s. nondecreasing if and only if X_n is a submartingale.*

Proof. Uniqueness. Assume that we have $(M_n)_{n\geq 0}$ and $(A_n)_{n\geq 0}$ as in the theorem. Then, for $n \geq 0$, using the martingale property for M_n and the fact that X_0 and A_{n+1} are $\mathcal{F}_n$ measurable,

$$\begin{aligned}\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_0|\mathcal{F}_n] + \mathbb{E}[M_{n+1}|\mathcal{F}_n] + \mathbb{E}[A_{n+1}|\mathcal{F}_n] \\ &= X_0 + M_n + A_{n+1} = X_n + (A_{n+1} - A_n).\end{aligned}$$

This implies $A_{n+1} - A_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n$, and therefore, since $A_0 = 0$ a.s.,

$$A_n = \sum_{k=0}^{n-1} (\mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k) \text{ a.s.}$$

Hence A_n is uniquely determined, up to a.s. equality, and so is $M_n = X_n - X_0 - A_n$.

Existence. Inspired by the above computation, we set

$$A_n = \sum_{k=0}^{n-1} (\mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k) \text{ and } M_n = X_n - X_0 - A_n.$$

Clearly, $A_0 = M_0 = 0$ a.s. and A_n is $\mathcal{F}_{n-1}$-measurable. It remains to check that $(M_n)_{n\geq 0}$ is a martingale. Clearly A_n and M_n are in L^1. Moreover,

$$\begin{aligned}\mathbb{E}\big[M_{n+1}|\mathcal{F}_n\big] &= \mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] - X_0 - A_{n+1} \\ &= \mathbb{E}\big[X_{n+1}|\mathcal{F}_n\big] - X_0 - (A_n + \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n) = X_n - X_0 - A_n = M_n.\end{aligned}$$

Second additional statement. The equivalence "$(X_n)_{n\geq 0}$ submartingale $\Leftrightarrow$ $(A_n)_{n\geq 0}$ a.s. nondecreasing" follows from the formula $A_{n+1} - A_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n] - X_n$ above. □

Corollary 7.4. *Let $(X_n)_{n\geq 0}$ be a submartingale and $S \leq T$ bounded stopping times. Then $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.*

Proof. We write the Doob's decomposition $X_n = X_0 + M_n + A_n$ of X_n, Since $(M_n)_{n\geq 0}$ is a martingale, by Theorem 6.4, we have $\mathbb{E}[M_S] = \mathbb{E}[M_T]$. Since $(X_n)_{n\geq 0}$ is a submartingale, the process $(A_n)_{n\geq 0}$ is a.s. nondecreasing and we have $A_S \leq A_T$ a.s., implying $\mathbb{E}[A_S] \leq \mathbb{E}[A_T]$. Therefore,

$$\mathbb{E}[X_S] = \mathbb{E}[X_0] + \mathbb{E}[M_S] + \mathbb{E}[A_S] \leq \mathbb{E}[X_0] + \mathbb{E}[M_T] + \mathbb{E}[A_T] = \mathbb{E}[X_T].$$ □

8. Inequalities

8.1. **Maximal inequalities.** In this section, we use the notation

$$X_n^*(\omega) = \sup_{j\leq n} |X_j(\omega)|, \quad X_\infty^*(\omega) = \sup_{j\geq 0} |X_j(\omega)|.$$

In general, X_n^* is hard to analyze: the index j maximizing $X_j(\omega)$ depends on ω. For martingales, we have a good control on it.

Theorem 8.1 (Doob's maximal inequality). *(i) Let $(X_n)_{n\geq 0}$ be a martingale or a nonnegative submartingale. Then, for $\alpha > 0$ and any $n \geq 0$,*

$$\mathbb{P}[X_n^* \geq \alpha] \leq \tfrac{1}{\alpha}\mathbb{E}\big[|X_n|\big].$$

(ii) Let $(X_n)_{n\geq 0}$ be a nonnegative supermartingale. Then, for $\alpha > 0$,

$$\mathbb{P}[X_\infty^* \geq \alpha] \leq \tfrac{1}{\alpha}\mathbb{E}\big[|X_0|\big].$$

Comment: if X_n is your (algebraic) loss at a game and you cannot exceed loss α at any time, the probability $\mathbb{P}[X_n^* \geq \alpha]$ is something you really want to know about. Markov's inequality tells you that $\mathbb{P}[X_n^* \geq \alpha] \leq \frac{1}{\alpha}\mathbb{E}[|X_n^*|]$, so that Doob's inequality "replaces" $\mathbb{E}[|X_n^*|]$ by the easier quantity $\mathbb{E}\big[|X_n|\big]$ or $\mathbb{E}\big[|X_0|\big]$.

Proof. Preliminaries: We start by some discussions common to the two cases. Set $T = \inf\{j \geq 0 : |X_j| \geq \alpha\}$. Then

$$\mathbb{P}[X_n^* \geq \alpha] = \mathbb{P}[T \leq n].$$

Besides, when $T \leq n$, we have $|X_T| = |X_{n\wedge T}| \geq \alpha$. Therefore

$$\mathbb{E}\big[|X_{n\wedge T}|\mathbf{1}\{T \leq n\}\big] \geq \alpha\mathbb{E}\big[\mathbf{1}\{T \leq n\}\big] = \alpha\mathbb{P}[T \leq n].$$

We conclude that

$$\mathbb{P}[X_n^* \geq \alpha] = \mathbb{P}[T \leq n] \leq \tfrac{1}{\alpha}\mathbb{E}\big[|X_{n\wedge T}|\mathbf{1}\{T \leq n\}\big] \leq \tfrac{1}{\alpha}\mathbb{E}\big[|X_{n\wedge T}|\big]. \tag{9}$$

item (i). In this case, $|X_n|$ is a submartingale by Lemma 7.2. Therefore, from Corollary 7.4, we have

$$\mathbb{E}[|X_{n\wedge T}|] \leq \mathbb{E}[|X_n|]. \tag{10}$$

Plugging this back into (9), we get item (i).

item (ii) Here, $|X_n| = X_n$ is a supermartingale. From Corollary 7.4, we have

$$\mathbb{E}[|X_{n\wedge T}|] \leq \mathbb{E}[|X_0|].$$

Plugging this back into (9), we get

$$\mathbb{P}[X_n^* \geq \alpha] \leq \tfrac{1}{\alpha}\mathbb{E}[|X_0|].$$

The right-hand-side is independent from n, so we can take the limit $n \to \infty$ and have

$$\mathbb{P}[X_\infty^* \geq \alpha] = \lim_{n\to\infty} \mathbb{P}[X_n^* \geq \alpha] \leq \tfrac{1}{\alpha}\mathbb{E}[|X_0|],$$

where the first limit follows from the event $\{X_\infty^* \geq \alpha\}$ being a countable increasing union of the events $\{X_n^* \geq \alpha\}$. ☐

Technical comment used in the next section. We can improve inequality (i) as follows:

$$\mathbb{P}[X_n^* \geq \alpha] \leq \tfrac{1}{\alpha}\,\mathbb{E}\big[|X_n|\mathbf{1}\{X_n^* \geq \alpha\}\big]. \tag{11}$$

Indeed, we obsere that, when $T > n$, we have $|X_{n\wedge T}| = |X_n|$, which implies

$$\mathbb{E}\big[|X_{n\wedge T}|\mathbf{1}\{T > n\}\big] = \mathbb{E}\big[|X_n|\mathbf{1}\{T > n\}\big].$$

Combining with (10) gives

$$\mathbb{E}\big[|X_{n\wedge T}|\mathbf{1}\{T \leq n\}\big] \leq \mathbb{E}\big[|X_n|\mathbf{1}\{T \leq n\}\big].$$

Plugging this back into (9) gives (11).

8.2. L^p **inequality.** In this section, we fix $p \geq 1$. The goal is now to control $\|X_n^*\|_p = \mathbb{E}\big[(X_n^*)^p\big]^{1/p}$.

Lemma 8.2. *Let Y be a nonnegative r.v. We have*

$$\mathbb{E}[Y^p] = \int_0^\infty p\,\alpha^{p-1}\,\mathbb{P}[Y \geq \alpha]d\alpha.$$

(Both sides might be finite or infinite.)

Comment. The case $p = 1$, namely $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y \geq \alpha]d\alpha$, is well-known.

Proof. Using Fubini's theorem for nonnegative functions,

$$\int_0^\infty p\,\alpha^{p-1}\,\mathbb{P}[Y \geq \alpha]d\alpha = \mathbb{E}\left[\int_0^\infty p\,\alpha^{p-1}\,\mathbf{1}\{Y \geq \alpha\}d\alpha\right] = \mathbb{E}\left[\int_0^Y p\alpha^{p-1}d\alpha\right] = \mathbb{E}\left[Y^p\right].$$ ☐

Reminder (Hölder's inequality). Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then $\mathbb{E}[|XY|] \leq \|X\|_p\|Y\|_q$.

Theorem 8.3 (Doob's L^p inequality). *Let $p>1$ and (X_n) be a martingale or a nonnegative submartingale. Then*

$$\|X_n^*\|_p \le \tfrac{p}{p-1}\|X_n\|_p.$$

($\|X_n\|_p$ or both $\|X_n\|_p$ and $\|X_n^*\|_p$ might be infinite.)

Proof. We assume w.l.o.g. $\|X_n\|_p<\infty$, i.e. X_n in L^p. This implies $\mathbb{E}[X_n|\mathcal{F}_j]\in L^p$ for any $j\le n$. But we have either $X_j=\mathbb{E}[X_n|\mathcal{F}_j]$ (martingale case) or $|X_j|=X_j\le\mathbb{E}[X_n|\mathcal{F}_j]$ (nonnegative submartingale case), so that X_j is also in L^p. Since $X_n^*\le|X_0|+\cdots+|X_n|$, we know that X_n^* is also in L^p, i.e. $\|X_n^*\|_p<\infty$. We have

$$\begin{aligned}
&\text{(Lemma 8.2)} && \mathbb{E}\big[(X_n^*)^p\big] = \int_0^{+\infty} p\alpha^{p-1}\mathbb{P}[X_n^*\ge\alpha]d\alpha\\
&\text{(Eq. (11))} && \le \int_0^{+\infty} p\alpha^{p-1}\frac{\mathbb{E}\big[|X_n|\mathbf{1}\{X_n^*\ge\alpha\}\big]}{\alpha}d\alpha\\
&\text{(Fubini)} && \le p\,\mathbb{E}\left[|X_n|\int_0^{+\infty}\alpha^{p-2}\mathbf{1}\{X_n^*\ge\alpha\}d\alpha\right]\\
&\text{(computations)} && \le p\,\mathbb{E}\left[|X_n|\int_0^{X_n^*}\alpha^{p-2}d\alpha\right] = \tfrac{p}{p-1}\mathbb{E}\left[|X_n|(X_n^*)^{p-1}\right]\\
&\text{(Hölder)} && \le \tfrac{p}{p-1}\,\|X_n\|_p\,\mathbb{E}\big[(X_n^*)^{(p-1)q}\big]^{1/q}.
\end{aligned}$$

But $(p-1)q=p$, so that dividing the above equality by $\mathbb{E}\big[(X_n^*)^p\big]^{1/q}<\infty$ gives

$$\|X_n^*\|_p = \mathbb{E}\big[(X_n^*)^p\big]^{1-1/q} \le \tfrac{p}{p-1}\|X_n\|_p. \qquad \square$$

Comment. With the hypothesis of the theorem, $|X_n|^p$ is a submartingale so that $\|X_j\|_p\le\|X_n\|_p$. This implies

$$\|X_n^*\|_p \le \|X_0\|_p+\cdots+\|X_n\|_p \le (n+1)\|X_n\|_p.$$

The nice aspect of Doob's L^p inequality is that the pre-factor is independent of n. In particular, it yields the following.

Corollary 8.4. *Let $(X_n)_{n\ge0}$ be a martingale or a nonnegative submartingale bounded in L^p (i.e. we assume $\sup_{n\ge0}\|X_n\|_p<+\infty$). Then X_∞^* is in L^p and*

$$\|X_\infty^*\|_p \le \frac{p}{p-1}\sup_{n\ge0}\|X_n\|_p.$$

Proof. By definition X_n^* is a.s. nondecreasing (we take a supremum over more and more variables) and tends a.s. to X_∞^*. Thus, by monotone convergence,

$$\|X_\infty^*\|_p = \lim_{n\to\infty}\|X_n^*\|_p.$$

The corollary then follows from Doob's L^p inequality. $\square$

9. Almost sure convergence

9.1. **The upcrossing inequality.** Here, we prove a technical result which will be useful for proving convergence theorem in the next section.

Upcrossings. Let $\underline{x}=(x_n)_{n\ge0}$ be a deterministic real-valued sequence and $a<b$ be real numbers. We set $t_0=0$ and for $j\ge0$,

$$\begin{aligned}
s_{j+1}^{(a,b)}(\underline{x}) &= \inf\{n\ge t_j : x_n\le a\};\\
t_{j+1}^{(a,b)}(\underline{x}) &= \inf\{n\ge s_{j+1} : x_n\le b\};\\
u_n^{(a,b)}(\underline{x}) &= \max\{j : t_j^{(a,b)}(\underline{x})\le n\};\\
u_\infty^{(a,b)}(\underline{x}) &= \max\{j : t_j^{(a,b)}(\underline{x})<+\infty\}.
\end{aligned}$$

Informally, $u_n^{(a,b)}(\underline{x})$ (resp. $u_\infty^{(a,b)}(\underline{x})$) is the number of times the sequence $(x_n)_{n\ge0}$ has crossed the strip (a,b) from bottom to top until time n (resp. until time ∞); see Fig. 1.

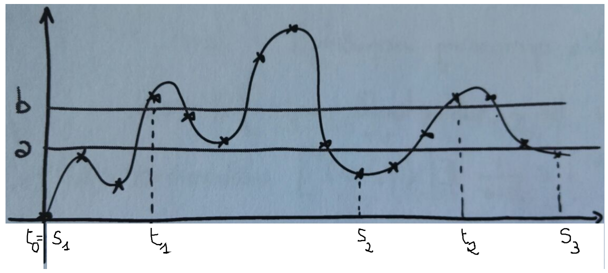


FIGURE 1. Illustration of the definition of upcrossings. The crosses indicate the dots of coordinates (n, x_n). The curve does not represent anything concrete, it is only here to make the definition of upcrossings more intuitive.

Lemma 9.1. *A real-valued sequence $(x_n)_{n\geq 0}$ has a limit ℓ in $\mathbb{R} \cup \{-\infty, +\infty\}$ if and only if, for all rational numbers $a < b$, we have $u_\infty^{(a,b)}(\underline{x}) < +\infty$.*

Proof. Admitted (easy Analysis I exercise). □

We know consider a random process $\underline{X} := (X_n)_{n\geq 0}$. Then, for $j \geq 0$ and n in $\mathbb{N} \cup \{\infty\}$, the quantity

$$S_j^{(a,b)} = s_j^{(a,b)}(\underline{X}), T_j^{(a,b)} = t_j^{(a,b)}(\underline{X}) \text{ and } U_n^{(a,b)} = u_n^{(a,b)}(\underline{X})$$

are random variables. It is easy to see that $S_j^{(a,b)}$ and $T_j^{(a,b)}$ are stopping times and the $U_n^{(a,b)}$ is $\mathcal{F}_n$ measurable.

Notation: $y^+ = \max(y, 0)$.

Theorem 9.2 (Doob's upcrossing inequality)**.** *Let $(X_n)_{n\geq 0}$ be a submartingale and fix $a < b$. Then*

$$\mathbb{E}\big[U_n^{(a,b)}\big] \leq \tfrac{1}{b-a}\mathbb{E}\big[(X_n - a)^+\big]$$

Proof. We write $U_n = U_n^{(a,b)}$ for simplicity. Let $Y_n = (X_n - a)^+$. Then $(Y_n)_{n\geq 0}$ is a submartingale since the function $x \mapsto (x-a)^+$ is convex and nondecreasing. We can write

$$Y_n = Y_{S_1} + (Y_{T_1} - Y_{S_1}) + (Y_{S_2} - Y_{T_1}) + \dots + \begin{cases} (Y_{T_j} - Y_{S_j}) + (Y_n - T_{T_j}); \\ (Y_{S_{j+1}} - Y_{T_j}) + (Y_n - Y_{S_{j+1}}), \end{cases}$$

where j is chosen such that $T_j \leq n < S_{j+1}$ (first case) or $S_{j+1} \leq n < Tj + 1$ (second case). In both cases, $j = U_n$. We can rewrite in a unified way as

$$Y_n = Y_{S_1 \wedge n} + \sum_{i=1}^{n} (Y_{T_i \wedge n} - Y_{S_i \wedge n}) + + \sum_{i=1}^{n} (Y_{S_{i+1} \wedge n} - Y_{T_i \wedge n}) \tag{12}$$

For $i \leq U_n$, we have

$$Y_{T_i \wedge n} = Y_{T_i} \geq b - a \quad (\text{since } X_{T_i} \geq b); \qquad Y_{S_i \wedge n} = Y_{S_i} = 0 \quad (\text{since } X_{S_i} \leq a).$$

This implies $Y_{T_i \wedge n} - Y_{S_i \wedge n} \geq b - a$. Other terms in the first sum of (12) (for $i > U_n$) are nonnegative (either $Y_{S_i \wedge n} = Y_{S_i} = 0$ or $Y_{S_i \wedge n} = Y_n = Y_{T_i \wedge n}$), so that this sum is at least $(b-a)U_n$.

For the second sum in (12), we observe that $(T_i \wedge n) \leq (S_{i+1} \wedge n)$ are both bounded stopping time. Since Y_n is a submartingale, from Corollary 7.4, we have

$$\mathbb{E}\big[Y_{T_i \wedge n}\big] \leq \mathbb{E}\big[Y_{S_{i+1} \wedge n}\big].$$

Therefore the second term in (12) has a nonnegative expectation.

Finally, the first term in (12) is $Y_{S_1 \wedge n}$ which is nonnegative a.s.

We conclude, taking expectation (12) and using the previous discussions, that

$$\mathbb{E}[Y_n] \geq \mathbb{E}\left[\sum_{i=1}^{n} (Y_{T_i \wedge n} - Y_{S_i \wedge n})\right] \geq (b-a)\mathbb{E}[U_n].$$

□

9.2. The almost sure convergence theorem.

Theorem 9.3 (submartingale a.s. convergence theorem). *Let $(X_n)_{n\geq 0}$ be a submartingale. We assume $\sup_{n\geq 0}\mathbb{E}[X_n^+]<\infty$. Then there exists $X_\infty\in L^1$ such that $X_n \xrightarrow{a.s.} X_\infty$.*

Comments:

- Submartingale are assumed to be in L^1. This implies that, for each $n\geq 0$, we have $\mathbb{E}[X_n^+]<\infty$. What is important in the above hypothesis is that we have a bound *independent of* n.
- Here are some stronger conditions (hence sufficient to imply a.s. convergence of submartingales): $\sup\mathbb{E}[|X_n|]<\infty$ or "there exists $M>0$ s.t., for all $n\geq 0$, $X_n\leq M$ a.s.". Again, the important point is that the bound is independent of n.

  In particular, **nonnegative martingales and supermartingales cv a.s.**
- For all $n\geq 0$, X_n lies in L^1 (definition of submartingale). The theorem states that the a.s. limit X_∞ is in L^1 as well. However, X_n *does not necessarily converge to* X_∞ *in* L^1. It may also happen that $\mathbb{E}[X_n]$ does not converge to $\mathbb{E}[X_\infty]$.

Proof. Let $a<b$ be rational numbers. Using the notation of Section 9.1, we have
$$\mathbb{E}\big[U_\infty^{(a,b)}\big]=\lim_{n\to\infty}\mathbb{E}\big[U_n^{(a,b)}\big]\leq \liminf n\to\infty \tfrac{1}{b-a}\mathbb{E}\big[(X_n-a)^+\big]\leq \tfrac{1}{b-a}\big(\mathbb{E}[(X_n)^+]+|a|\big)<+\infty,$$
where the first equality comes from the monotone convergence theorem and the following inequality from Theorem 9.2. This implies
$$\mathbb{P}\big[U_\infty^{(a,b)}<\infty\big]=1.$$
This holds for any rational numbers $a<b$. Since $\mathbb{Q}$ is countable, we have
$$\mathbb{P}\left[\bigcap_{a,b\in\mathbb{Q},\,a<b}\{U_\infty^{(a,b)}<\infty\}\right].$$
From Lemma 9.1, this implies, that for ω in a set of probability 1, $X_n(\omega)$ has a limit in $\mathbb{R}\cup\{-\infty,+\infty\}$, which we denote $X_\infty(\omega)$.

This defines X_∞ on a set of probability 1. By construction $X_n\xrightarrow{\text{a.s.}}X_\infty$. It remains to check that X_∞ is in L_1. We have, for all $n\geq 0$,
$$\mathbb{E}[|X_n|]=\mathbb{E}[X_n^+]+\mathbb{E}[X_n^-]=2\mathbb{E}[X_n^+]-\mathbb{E}[X_n]\leq 2\mathbb{E}[X_n^+]-\mathbb{E}[X_0],$$
where we used in the middle the inequality $\mathbb{E}[X_n]\geq\mathbb{E}[X_0]$, true for any submartingale. This implies, in combination with Fatou's lemma,
$$\mathbb{E}[|X_\infty|]\leq\liminf_{n\geq 0}\mathbb{E}[|X_n|]\leq 2(\sup_{n\geq 0}\mathbb{E}[X_n^+])-\mathbb{E}[X_0]<+\infty. \qquad\square$$

Note: when we construct X_∞ in the proof, it might a priori take values $\pm\infty$. However, a posteriori, from $\mathbb{E}[|X_\infty|]<+\infty$, we know that X_∞ takes finite values a.s.

<u>An important example.</u> Take $X_n=Y_1+\cdots+Y_n$, where Y_i are i.i.d. r.v. with $\mathbb{P}[Y_i=1]=\mathbb{P}[Y_i=-1]=1/2$. We take the stopping time $T=\inf\{n\geq 0 \text{ s.t. } X_n=1\}$ and set $M_n=X_{n\wedge T}$, which is a martingale by Proposition 6.3. We note that $M_n\leq 1$ a.s. (either $T>n$, then $M_n=X_n$ and X_n has not reach 1 yet, so that $X_n<1$, or $T\leq n$ and then $M_n=X_T=1$). Applying Theorem 9.3, we know that M_n converges a.s. to some M_∞.

Let ω be an element of the probability space Ω. Assume that $T(\omega)=+\infty$. Then $M_n(\omega)=X_n(\omega)$ for all $n\geq 0$. But $X_n(\omega)$ does not converge (for all $n\geq 1$, we have $|X_n(\omega)-X_{n-1}(\omega)|=|Y_n(\omega)|=1$), therefore $M_n(\omega)$ does not converge.

We set
$$C=\{\omega\in\Omega: M_n(\omega)\text{ converges}\}.$$
we have proved above that, on the one hand $\mathbb{P}[C]=1$ and on the other hand $\{T=\infty\}\subseteq\overline{C}$. This implies $\mathbb{P}\big[T=\infty\big]=0$. In other words, with probability 1, there exists a time n, at which $X_n=1$.

(This is not easy to prove directly!)

In the above example, we have $M_\infty=X_T=1$ a.s., while $M_0=X_0=0$. In particular $\mathbb{E}[M_\infty]\neq\mathbb{E}[M_0]$, while $\mathbb{E}[M_n]=\mathbb{E}[M_0]$ for all $n\geq 0$ (since M_n is a martingale). In summary, we have a.s. convergence $M_n\to M_\infty$, but not $\mathbb{E}[M_n]\to\mathbb{E}[M_\infty]$ (and a fortiori not convergence in L^1).

10. Uniform integrable (sub)martingales and L^1 convergence

10.1. L^1 convergence theorems.

Uniform integrability is a necessary and sufficient condition for a sequence of random variables assumed to converge in probability to converge also in L^1. A short account on uniform integrability is given in Appendix A[6].

Theorem 10.1 (Submartingale L^1 convergence theorem)**.** *Let $(X_n)_{n\geq 0}$ be a submartingale. The following assertions are equivalent:*

(i) $(X_n)_{n\geq 0}$ *is u.i.;*

(ii) There exists X_∞ *in* L^1 *such that* $X_n \overset{a.s.,\, L^1}{\longrightarrow} X_\infty$.

Proof. First assume (i). By Proposition A.2 item (i), we have $\sup_{i\in I} \mathbb{E}\big[|X_i|\big] < \infty$. We can therefore apply the submartingale a.s. convergence theorem, and we know that there exists X_∞ in L^1 such that $X_n \overset{\text{a.s.}}{\longrightarrow} X_\infty$. A fortiori, X_n converges to X_∞ in probability; since X_n is u.i., by Proposition A.2, item (iii), we conclude that X_n converges to X_∞ in L^1.

Conversely, let us assume (ii). Then (i) follows from Proposition A.2, item (ii). □

For martingales, there is a third equivalent statement.

Theorem 10.2 (Martingale L^1 convergence theorem)**.** *Let $(X_n)_{n\geq 0}$ be a martingale. The following assertions are equivalent:*

(i) $(X_n)_{n\geq 0}$ *is u.i.;*

(ii) There exists X_∞ *in* L^1 *such that* $X_n \overset{a.s.,\, L^1}{\longrightarrow} X_\infty$*;*

(iii) There exists X *in* L^1 *such that, for all* $n \geq 0$*, we have* $X_n = \mathbb{E}[X|\mathcal{F}_n]$.

Moreover, when these statements hold, we have $X_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$*, where* $\mathcal{F}_\infty = \sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$.

Comments:

- Recall that martingales as in item (iii) are called *closed*. The theorem asserts that a martingale is u.i. if and only if it is closed.
- The r.v. X from item (iii) is not unique. We will see in the proof that we can take $X = X_\infty$. Moreover, when $\mathcal{F}_\infty = \mathcal{F}$, the additional statement says that necessarily $X = X_\infty$, so that we have uniqueness in this case. It is however good to keep in mind that it is not always the case.

Proof. We already proved (i) ⇔ (ii) in the more general setting of submartingales (Theorem 10.1). Moreover, (iii) ⇒ (i) is a direct consequence of Corollary A.4.

Proof of (ii) ⇒ (iii). We will prove that $X_n = \mathbb{E}\big[X_\infty|\mathcal{F}_n\big]$, yielding the existence of X as wanted (take $X = X_\infty$). To this end, we check the characterizing property of conditional expectation. Take B in $\mathcal{F}_n$. For $p \geq n$, we have $\mathbb{E}\big[X_p|\mathcal{F}_n\big] = X_n$ (martingale property), implying $\mathbb{E}\big[X_p\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_n\,\mathbf{1}\{B\}\big]$. Since X_p converges to X_∞ in L^1, we have

$$\lim_{p\to\infty} \mathbb{E}\big[X_p\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_\infty\,\mathbf{1}\{B\}\big].$$

Therefore $\mathbb{E}\big[X_\infty\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_n\,\mathbf{1}\{B\}\big]$. Since this holds for any B in $\mathcal{F}_n$ and since X_n is $\mathcal{F}_n$ measurable, we have $X_n = \mathbb{E}\big[X_\infty|\mathcal{F}_n\big]$, concluding the proof.

Proof of the extra statement. For any $n \geq 0$, X_n is $\mathcal{F}_n$-measurable and hence $\mathcal{F}_\infty$ measurable. The limit of measurable functions is measurable, so X_∞ is also $\mathcal{F}_\infty$ measurable. To prove $X_\infty = \mathbb{E}[X|\mathcal{F}_\infty]$, we need to prove that it satisfy the characterizing property.

Let B be in F_n for some $n \geq 0$. Then we have

$$\mathbb{E}\big[X\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_n\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_\infty\,\mathbf{1}\{B\}\big],$$

where the first equality follows from $X = \mathbb{E}[X_n|\mathcal{F}_n]$ (asumption of item (iii) above), while the second follows from $X_\infty = \mathbb{E}[X_n|\mathcal{F}_n]$ (proved in (ii) ⇒ (iii) above). Therefore the equality $\mathbb{E}\big[X\,\mathbf{1}\{B\}\big] = \mathbb{E}\big[X_\infty\,\mathbf{1}\{B\}\big]$

[6] The material in Appendix A will be presented in the lecture; we have chosen to put it in appendix in the lecture notes, because this material can be of interest, independently of its use in martingale theory.

holds for all B in $\bigcup_{n\geq 0}\mathcal{F}_n$. By the monotone class theorem, see, e.g., [JP04, Theorem 6.2], it holds for any B in $\sigma\left(\bigcup_{n\geq 0}\mathcal{F}_n\right)=\mathcal{F}_\infty$. □

10.2. **An example of application.** As above, we consider $X_n = Y_1+\cdots+Y_n$, where Y_i are i.i.d. r.v. with $\mathbb{P}[Y_i=1]=\mathbb{P}[Y_i=-1]=1/2$. Additionally, let a and b be two positive integers and set

$$T=\inf\{n\geq 0 \text{ s.t. } X_n=-a \text{ or } X_n=b\}.$$

Obviously, T is a stopping time. Therefore $X_{n\wedge T}$ is a martingale (Proposition 6.3). We note that $X_{n\wedge T}$ is a.s. in the set $\{-a,-a+1,\cdots,b\}$, so that it is a bounded martingale, and in particular u.i. Therefore $X_{n\wedge T}$ converges a.s. and in L^1 to X_T. The same argument as at the end of Section 9.2 tells us that $T<+\infty$ a.s., which implies that a.s. X_T takes values in $\{-a,b\}$. We have

$$\mathbb{E}[X_T]=-a\,\mathbb{P}[X_T=-a]+b\mathbb{P}[X_T=b]=\mathbb{E}[X_0]=0.$$

Since $\mathbb{P}[X_T=-a]+\mathbb{P}[X_T=b]=1$, we can solve and get

$$\mathbb{P}[X_T=-a]=\tfrac{b}{a+b},\quad \mathbb{P}[X_T=-a]=\tfrac{a}{a+b}.$$

We determined the probabilities of winning and losing in the strategy "I stop the game when I have lost amount a or won amount b".

10.3. **The unbounded optional stopping theorem.** In the following, we consider a **u.i.** martingale $(X_n)_{n\geq 0}$ and a stopping time T. Since X_n has an a.s. limit X_∞, we can define X_T regardless of T being finite or not, namely

$$X_T(\omega)=\begin{cases}X_{T(\omega)}(\omega) & \text{if } T(\omega)<\infty;\\ X_\infty(\omega) & \text{if } T(\omega)=+\infty.\end{cases}$$

Theorem 10.3. *Let $(X_n)_{n\geq 0}$ be a* **u.i.** *martingale and T be a stopping time. Then $\mathbb{E}[X_T]=\mathbb{E}[X_0]$.*

Comparison with optional stopping theorem (Theorem 6.4). Here, T is not assumed to be bounded, not even to be a.s. finite. The cost is that we have to assume X_n u.i.

Gambling interpretation. If we assume the winnings (and losses) of the gambler to be bounded (quite reasonable assumption!), then *even unbounded time strategies* cannot achieve a better expectation of winnings than just keeping the initial amount X_0.

To prove the theorem, we first prove the following lemma.

Lemma 10.4. *Let $(X_n)_{n\geq 0}$ be a u.i. martingale. Then family (X_S), where the index S runs over all stopping times S, is u.i.*

Proof. Recall that, as a u.i. martingale, X_n has an a.s. and L^1 limit X_∞ and that $X_n=\mathbb{E}[X_\infty|\mathcal{F}_n]$ (see Theorem 10.2 and the discussion below).

$$\begin{aligned}X_S &= X_\infty\,\mathbf{1}\{S=+\infty\}+\sum_{n\geq 0}X_n\,\mathbf{1}\{S=n\}\\ &= X_\infty\mathbf{1}\{S=+\infty\}+\sum_{n\geq 0}\mathbb{E}[X_\infty|\mathcal{F}_n]\,\mathbf{1}\{S=n\},\end{aligned}$$

The single r.v. family (X_∞) is uniformly inegrable so that, by Proposition A.3, there exists a convex nondecreasing function φ with $\lim_{x\to\infty}\frac{\varphi(x)}{x}=+\infty$ and $\mathbb{E}[\varphi(|X_\infty|)]<\infty$. We apply $x\to\varphi(|x|)$ to the above display, observing that there is always only one non-zero term on the right-hand side:

$$\varphi\big(|X_S|\big)=\varphi\big(|X_\infty|\big)\,\mathbf{1}\{S=+\infty\}+\sum_{n\geq 0}\varphi\big(|\mathbb{E}[X_\infty|\mathcal{F}_n]|\big)\,\mathbf{1}\{S=n\}.$$

Now we use Jensen's inequality for the convex function $x\to\varphi(|x|)$:

$$\varphi\big(|X_S|\big)\leq\ \varphi\big(|X_\infty|\big)\,\mathbf{1}\{S=+\infty\}+\sum_{n\geq 0}\mathbb{E}\big[\varphi(|X_\infty|)|\mathcal{F}_n\big]\,\mathbf{1}\{S=n\}.$$

Taking expectation, and using the characterizing property, we get

$$\mathbb{E}\Big[\varphi\big(|X_S|\big)\Big] \leq \mathbb{E}\Big[\varphi(|X_\infty|)\mathbf{1}\{S=+\infty\}\Big] + \sum_{n\geq 0} \mathbb{E}\Big[\varphi(|X_\infty|)\mathbf{1}\{S=n\}\Big] = \mathbb{E}\big[\varphi(|X_\infty|)\big].$$

The right-hand side is finite (by construction of φ) and the bound is uniform on all stopping times S. From Proposition A.3, this proves that the family (X_S), where S runs over all stopping times, is u.i. □

Proof of Theorem 10.3. From Proposition 6.3, we know that $(X_{n\wedge T})_{n\geq 0}$ is a martingale. It is easy to check from the definition that this martingale converges a.s. to X_T, as n tends to $+\infty$.

Since for each $n \geq 0$, the quantity $n \wedge T$ is a stopping time, the family $(X_{n\wedge T})_{n\geq 0}$ is a subfamily of the family (X_S), where S runs over all stopping times. From Lemma 10.4, this family is u.i. Therefore the martingale $X_{n\wedge T}$ converges to X_T also in L^1. Consequently

$$\mathbb{E}[X_T] = \lim_{n\to\infty} \mathbb{E}[X_{n\wedge T}].$$

But, from the optional stopping theorem for bounded stopping times (Theorem 6.4), we know that for every $n \geq 0$, we have $\mathbb{E}[X_{n\wedge T}] = \mathbb{E}[X_0]$. We conclude that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, as wanted. □

10.4. L^p **convergence.** In this section, we discuss L^p convergence of martingales. For background and motivation for L^p convergence of r.v. in general, we refer the reader to Appendix B[7]

Theorem 10.5 (Submartingale L^p convergence theorem)**.** *Let $(X_n)_{n\geq 0}$ be a submartingale and $p > 1$. Then the following assertions are equivalent:*

(i) $\sup_{n\geq 0} \|X_n\|_p < \infty$*;*
(ii) $(X_n)_{n\geq 0}$ *converges to some* X_∞ *in* L^p*.*

Comments

- The implication (ii) ⇒ (i) is true for any sequence $(X_n)_{n\geq 0}$ of r.v.; see Proposition B.1. Only the implication (i) ⇒ (ii) uses the martingale assumption and needs to be proven.
- (i) holds in particular for any p for bounded martingales.

Proof of (i) ⇒ (ii). From Proposition A.3 used with $\varphi(x) = x^p$, the hypothesis $\sup_{n\geq 0} \|X_n\|_p < \infty$ implies that the martingale $(X_n)_{n\geq 0}$ is u.i. Therefore, from Theorem 10.2, it has some a.s. and L^1 limit X_∞. We need to prove that X_∞ is in L^p and that the convergence $X_n \to X_\infty$ holds in L^p as well.

We recall Doob's L^p inequality (or rather Corollary 8.4): if $X^*_\infty = \sup_{j\geq 0} |X_j|$, then

$$\|X^*_\infty\|_p \leq \tfrac{p}{p-1} \sup_{n\geq 0} \|X_n\|_p.$$

Assuming (i), we have X^*_∞ is in L^p. Since $|X_\infty| \leq X^*_\infty$ a.s., X_∞ is in L^p as well.

We finally want to prove $\mathbb{E}[|X_n - X_\infty|^p] \to 0$. We will apply the dominated convergence theorem.

- Since $X_n \xrightarrow{\text{a.s.}} X_\infty$, we have $|X_n - X_\infty|^p$ tends to 0 a.s.
- Moreover, a.s., we have

$$|X_n - X_\infty|^p \leq (|X_n| + |X_\infty|)^p \leq 2^{p-1}(|X_n|^p + |X_\infty|^p) \leq 2^{p-1}(|X^*_\infty|^p + |X_\infty|^p),$$

  where the middle inequality uses the convexity of the fonction $x \mapsto |x|^p$. The upper bound is independent of n and integrable (since X^*_∞ and X_∞ are both in L^p).

We conclude that $\mathbb{E}[|X_n - X_\infty|^p] \to 0$, i.e. X_n converges to X_∞ in L^p, as wanted. □

11. Backwards martingales

In this section, we consider *martingale* indexed by **negative** integers. For simplicity, we do not discuss generalizations, as *backward submartingales.*

[7]The material of this appendix will be presented in class.

Definition 11.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, then a negatively indexed sequence $(\mathcal{F}_n)_{\boldsymbol{n\le 0}}$ is a backward filtration if each $\mathcal{F}_n$ is a σ-algebra and we have*

$$\cdots \subseteq \mathcal{F}_{-2} \subseteq \mathcal{F}_{-1} \subseteq \mathcal{F}_0 \subseteq \mathcal{F}.$$

Moreover a negatively indexed sequence $(X_n)_{\boldsymbol{n\le 0}}$ of r.v. in L^1 is a backward martingale if each X_n is $\mathcal{F}_n$-measurable and, for every $n<0$,

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

As for usual martingales, we have more generally $\mathbb{E}\big[X_p|\mathcal{F}_n\big] = X_n$ for $p \ge n$. In particular $\mathbb{E}\big[X_0|\mathcal{F}_n\big] = X_n$, i.e. all backward martingales are "closed" (borrowing the terminology we introduced for usual martingales). Consequently, from Corollary A.4, *all backward martingales are u.i.* This observation is used in the following **convergence theorem**; note there is **no assumptions**, besides being a backward martingale.

Theorem 11.2. *Let $(X_n)_{n\le 0}$ be a backward martingale. Then there exists a r.v. $X_{-\infty}$ in L^1 s.t., as n tends to $-\infty$, $X_n \overset{a.s.,\ L^1}{\longrightarrow} X_{-\infty}$*

Proof. The proof uses Doob's upcrossing inequality. Since this inequality is stated for usual martingales (and not backward martingales), we need to construct a usual martingale. Fix $N \ge 0$ and define, for $k \ge 0$,

$$\mathcal{G}_k^N = \mathcal{F}_{\min(k-N,0)}, \qquad Y_k^N = X_{\min(k-N,0)}.$$

For each fixed $N \ge 0$, we have a filtration $(\mathcal{G}_k^N)_{k\ge 0}$ and a martingale $\underline{Y}^k := (Y_k^N)_{k\ge 0}$ w.r.t. that filtration (straightforward to check).

Doob's upcrosisng inequality (Theorem 9.2) applied to the martingale $(-Y_k^N)_{k\ge 0}$ at time $k=N$ tells us that, for any rational numbers $a<b$, we have

$$\mathbb{E}\big[U_N^{(a,b)}(-\underline{Y}^k)\big] \le \tfrac{1}{b-a}\mathbb{E}\big[(-Y_N^N - a)_+\big].$$

But $-Y_N^N = X_0$ and one can check easily that $U_N^{(a,b)}(-\underline{Y}^k) = U_N^{(-b,-a)}(\underline{X}')$, where $X'_n := X_{-n}$ is the positively-indexed version of X_n. Summing up, we have

$$\mathbb{E}\big[U_N^{(-b,-a)}(\underline{X}')\big] \le \tfrac{1}{b-a}\mathbb{E}\big[(-X_0 - a)_+\big],$$

where we note that the right-hand-side is independent of N and therefore gives an upper bound for $\mathbb{E}\big[U_\infty^{(-b,-a)}(\underline{X}')\big] = \lim_{N\to\infty} \mathbb{E}\big[U_N^{(-b,-a)}(\underline{X}')\big]$ (this limit holds by monotone convergence).

The proof is now similar to the forward case. The probability that $U_\infty^{(-b,-a)}(\underline{X}') < \infty$ is one. Since this holds for any rational numbers $a<b$, we know that a.s., the sequence $\underline{X}'$ satisfies: for all $a' := -b < b' := -a$, $U_\infty^{(a',b')}(\underline{X}') < \infty$. By the upcrossing lemma (Lemma 9.1), this implies, that for almost all ω, the sequence $X'_n(\omega)$ as n tends to $+\infty$, or equivalently $X_n(\omega)$ as n tends to $-\infty$, has a limit which we denote $X_{-\infty}(\omega)$. This defines a r.v. $X_{-\infty}$ and we have $X_n \overset{\text{a.s.}}{\longrightarrow} X_{-\infty}$ as n tends to $-\infty$.

Using Fatou's lemma, we have

$$\mathbb{E}\big[|X_{-\infty}|\big] \le \liminf_{n\to-\infty} \mathbb{E}\big[|X_n|\big] < +\infty,$$

the last bound coming from the fact that $(X_n)_{n\le 0}$ is u.i. (see the discussion before the theorem) and hence bounded in L^1 (Proposition A.2, item (i)). We conclude that $X_{-\infty}$ is in L_1.

Finally, recall that $X_n \overset{\text{a.s.}}{\longrightarrow} X_{-\infty}$. Since $(X_n)_{n\le 0}$ is u.i., from Proposition A.2, item (iii), we have $X_n \to X_{-\infty}$ in L^1 as well. ∎

12. Some applications

12.1. The strong law of large numbers.

Theorem 12.1. *Let $(Y_i)_{i\ge 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}\big[|Y_1|\big] < \infty$. Then*

$$\frac{Y_1 + \cdots + Y_n}{n} \overset{a.s.,\ L^1}{\longrightarrow} \mathbb{E}[Y_1].$$

(The right-hand side is the constant random variable X_∞ s.t. $X_\infty(\omega) = \mathbb{E}[Y_1]$ for (almost) all ω.)

The proof of the theorem uses backward martingales and the two following properties of conditional expectation (left as exercises to the reader)

(i) Let X be a r.v. defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{B}$ and $\mathcal{C}$ be σ-subalgebra of $\mathcal{F}$. We assume that $(X, \mathcal{B})$ is independent from $\mathcal{C}$. Then

$$\mathbb{E}\big[X|\sigma(\mathcal{B},\mathcal{C})\big] = \mathbb{E}\big[X|\mathcal{B}\big]$$

(ii) Take $(Y_i)_{i\geq 1}$ as above and $S_n = Y_1 + \cdots + Y_n$. Then, for i in $\{1, \cdots, n\}$, we have $\mathbb{E}\big[Y_i|S_n\big] = \frac{S_n}{n}$ a.s.

Proof. We set $S_n = Y_1 + \cdots + Y_n$, $\mathcal{G}_{-n} = \sigma(S_n, S_{n+1}, \cdots)$ and $X_{-n} = \frac{S_n}{n}$ for $n \geq 1$.

Claim: $(X_p)_{p\leq -1}$ is a backward martingale w.r.t to the filtration $(\mathcal{G}_p)_{p\leq 1}$.

Proof of the claim: clearly X_p is in L^1 since the Y_i's are in L^1 and X_p is $(\mathcal{G}_p)_{p\leq 1}$-measurable. We also note that $\mathcal{G}_{-n} = \sigma(S_n, Y_{n+1}, Y_{n+2}, \cdots)$. Using the above properties (i) and (ii), we have, for $p = -n < 1$

$$\begin{aligned}\mathbb{E}\big[X_{p+1}|\mathcal{G}_p\big] = \mathbb{E}\big[\tfrac{S_{n-1}}{n-1}|S_n\big] = \frac{1}{n-1}\big(\mathbb{E}\big[Y_1|S_n\big] + \cdots + \mathbb{E}\big[Y_{n-1}|S_n\big]\big)& \\ = \frac{1}{n-1}\big(\tfrac{S_n}{n} + \cdots + \tfrac{S_n}{n}\big) = \tfrac{S_n}{n} = X_p.&\end{aligned}$$

Therefore the martingale property is verified and $(X_p)_{p\leq -1}$ is a backward martingale, as claimed.

From Theorem 11.2, we know that there exists $X_{-\infty}$ in L^1 s.t. $X_p \overset{\text{a.s., } L^1}{\longrightarrow} X_{-\infty}$ as p tends to $-\infty$. Since $X_{-n} = \frac{S_n}{n}$, this translates as $\frac{S_n}{n} \overset{\text{a.s., } L^1}{\longrightarrow} X_{-\infty}$ as n tends to $+\infty$ and the only thing left to prove is $X_{-\infty} = \mathbb{E}[Y_1]$ a.s.

As the convergence $X_p \to X_{-\infty}$ holds in L^1, we also have convergence in expectation and

$$\mathbb{E}[X_{-\infty}] = \lim_{n\to\infty} \mathbb{E}[\tfrac{S_n}{n}] = \mathbb{E}[Y_1]. \tag{13}$$

Besides, for each $N \geq 1$, $X_{-\infty}$ is $\sigma(Y_{N+1}, Y_{N+2}, \cdots)$ measurable. Indeed, setting $S_n^N = \sum_{i=N+1}^n Y_i$, we have

$$\lim_{n\to\infty} \tfrac{S_n}{n} = \lim_{n\to\infty} \tfrac{S_n^N}{n}$$

and the RHS is $\sigma(Y_{N+1}, Y_{N+2}, \cdots)$ measurable. We conclude that $X_{-\infty}$ is

$$\mathcal{H}_\infty := \bigcap_{N\geq 1} \sigma(Y_{N+1}, Y_{N+2}, \cdots)$$

measurable. But $\mathcal{H}_\infty$ is known as Kolmogorov's tail algebra and Kolmogorov 0-1 law states that events in this algebra have probability 0 or 1 [JP04, Theorem 10.6] or equavalently, r.v. which are $\mathcal{H}_\infty$-measurable are a.s. constant. Therefore $X_{-\infty} = C$ a.s. for some real number C. But $C = \mathbb{E}[X_{-\infty}] = \mathbb{E}[Y_1]$ by Eq. (13) above, which concludes the proof. $\square$

12.2. **Hewitt-Savage zero-one law.** Let $(Y_i)_{i\geq 1}$ be i.i.d. random variables and $X = F(Y_1, Y_2, \cdots)$ be some random variable. The function F is called symmetric if, for any permutation $\pi : \mathbb{N} \to \mathbb{N}$ with *finite support* and any variables x_1,x_2, ..., we have

$$F(x_{\pi(1)}, x_{\pi(2)}, \cdots) = F(x_1, x_2, \cdots)$$

Theorem 12.2. *Let $(Y_i)_{i\geq 1}$ be i.i.d. random variables and F be a symmetric function as above. Then the r.v. $X = F(Y_1, Y_2, \cdots)$ is a.s. constant.*

Example. We take

$$F(x_1, x_2, \cdots) = \mathbf{1}\big\{x_1 + \cdots + x_n > 10 \text{ for infinitely many } n\big\}.$$

Then $F(Y_1, Y_2, \cdots) = 1$ a.s. or $F(Y_1, Y_2, \cdots) = 0$ a.s. (this depends on the distribution of Y), i.e. either the partial sum sequence goes infinitely many times above 10 a.s., or it stays ultimately below 10 a.s.

This extends to any dimension and any Borel set B instead of the interval $(10, +\infty)$.

Proof. W.l.o.g., assume F bounded (otherwise replace F by $\varphi \circ F$ where φ is a bounded injective function); this implies that X is in L^1. We set

$$\mathcal{F}_n = \sigma(Y_1, \cdots, Y_n), \qquad \mathcal{G}_n = \sigma(Y_{n+1}, Y_{n+2}, \cdots),$$

and let $X_n = \mathbb{E}[X|\mathcal{F}_n]$ and $Z_n = \mathbb{E}[X|\mathcal{G}_n]$. The sequences X_n and Z_n defines usual martingale, resp. backward martingales with respect to the filtration $\mathcal{F}_n$, resp. backward filtration $\mathcal{G}_n$. Therefore, from Theorems 10.2 and 11.2 (note that X_n is a closed martingale by definition), we know that X_n and Z_n converge a.s. and in L^1 to X_∞ and Z_∞, respectively. Furthermore, we have

$$X_\infty = \mathbb{E}\Big[X|\sigma\big(\bigcup_{n\geq 0} \mathcal{F}_n\big)\Big] = X.$$

On the other hand, Z_∞ is $\bigcap_{n\geq 0} \mathcal{G}_n$ measurable, and hence a.s. equal to a constant C, by Kolmogorov 0-1 law. Since we have L^1 convergence, we have

$$C = \mathbb{E}[Z_\infty] = \lim_{n\to\infty} \mathbb{E}[Z_n] = \mathbb{E}[X],$$

where the third inequality follows from Z_n being a backward martingale. To sum up, Z_n converges to $\mathbb{E}[X]$, a.s and in L^1.

But X_n is $\mathcal{F}_n$ measurable, and hence, $X_n = g(Y_1, \dots, Y_n)$ for some measurable function g. Recall that X_n tends in L^1 to $X = F(Y_1, Y_2, \dots)$. Since $(Y_1, Y_2, \dots)$ has the same law as $(Y_{n+1}, \dots, Y_{2n}, Y_1, \dots, Y_n, Y_{2n+1}, \dots)$ (these are i.i.d. random variables, changing the order does not change their joint law), we have that $g(Y_{n+1}, \dots, Y_{2n})$ tends in L^1 to

$$F(Y_{n+1}, \dots, Y_{2n}, Y_1, \dots, Y_n, Y_{2n+1}, \dots) = F(Y_1, Y_2, \dots) = X,$$

where we used the symmetry of F. Concretely, we have

$$\lim_{n\to\infty} \mathbb{E}\Big[\big|g(Y_{n+1}, \dots, Y_{2n}) - X\big|\Big] = 0.$$

Since taking conditional expectation is a L^1 contraction (Jensen's inequality), we also have

$$\lim_{n\to\infty} \mathbb{E}\Big[\big|\mathbb{E}\big[g(Y_{n+1}, \dots, Y_{2n})|\mathcal{G}_n\big] - \mathbb{E}\big[X|\mathcal{G}_n\big]\big|\Big] = 0.$$

But $g(Y_{n+1}, \dots, Y_{2n})$ is $\mathcal{G}_n$ measurable and $\mathbb{E}\big[X|\mathcal{G}_n\big] = Z_n$ by definition, so that

$$\lim_{n\to\infty} \|g(Y_{n+1}, \dots, Y_{2n}) - Z_n\|_1 = 0.$$

However, we have proved that $g(Y_{n+1}, \dots, Y_{2n})$ tends to X in L^1 while Z_n tends to $\mathbb{E}[X]$. This forces $X = \mathbb{E}[X]$ a.s., which is what we wanted to prove. □

12.3. **Galton-Watson processes.** The two applications above are "theoretical" applications of martingale theory. There are also many applications in the analysis of probabilistic models: the idea is to find a martingale in the model, and use convergence results for this martingale to understand the behaviour of the models. A few examples of this kind are given in the exercise sheet (gamble models, urn models, ...). We give here a standard one: Galton-Watson processes.

Let μ be a probability measure on $\{0, 1, 2, \cdots\}$ with $m := \sum_{i\geq 0} i\mu(i) < \infty$. We assume that $0 < \mu(\{0\}) < 1$ (otherwise, the model is trivial).

Let $(\xi_k^{(n)})_{k,n\geq 1}$ be a collection of i.i.d. random variables with law μ We set $X_0 = 1$ and for $n \geq 0$,

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_k^{(n+1)}.$$

(The number of terms in the sum defining X_{n+1} is X_n.

Interpretation: this models the number of individuals in a population with asexual reproduction (or considering only males in a population; the model was initially introduced to estimate the probability of survival of last names of Lords in England). Then X_n is the number of individual at generation n; $\xi_k^{(n+1)}$ is the number of children of the k-th individual of generation n; this is a random variable following a distribution μ, which is called the *offspring distribution*. See Fig. 2.

Main question: Will there be **extinction** or **survival** of the population? i.e. what is the probability that $X_{n_0} = 0$ for some n_0 (which implies $X_n = 0$ for $n \geq n_0$)?

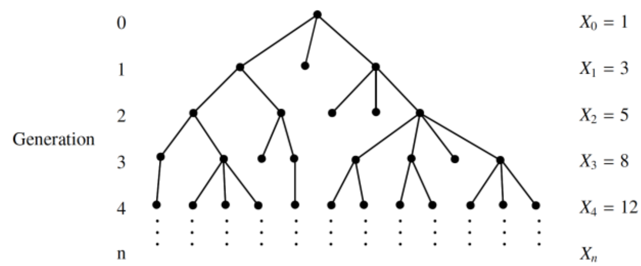


FIGURE 2. The first 4 generation of a Galton-Watson process (this shows one possible realization of the process). Taken from this url

.

A martingale: we consider the filtration

$$\mathcal{F}_n = \sigma\left((\xi_k^{(p)})_{\boldsymbol{p\le n}, k \ge 0} \right).$$

(This is the σ-algebra containing what happens until generation n.) Clearly, X_n is $\mathcal{F}_n$-measurable and

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}\left[\sum_{k=1}^{X_n} \xi_k^{(n+1)} \big| \mathcal{F}_n \right] = \mathbb{E}\left[\sum_{k=1}^{\infty} \mathbf{1}[k \le X_n] \xi_k^{(n+1)} \big| \mathcal{F}_n \right]$$

We have replace the upper bound X_n in the summation index by ∞ and added $\mathbf{1}[k \le X_n]$ in order not to manipulate a sum with a random summation set. We now use that X_n is $\mathcal{F}_n$-measurable, while $\xi_k^{(n+1)}$ is independent from $\mathcal{F}_n$: we get

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \sum_{k=1}^{\infty} \mathbf{1}[k \le X_n] \mathbb{E}\big[\xi_k^{(n+1)}\big] = \sum_{k=1}^{\infty} (\mathbf{1}[k \le X_n] m) = m\, X_n.$$

Dividing by m^{n+1}, this gives

$$\mathbb{E}\big[\tfrac{X_{n+1}}{m^{n+1}} \big| \mathcal{F}_n\big] = \tfrac{X_n}{m^n}.$$

This proves that $M_n := \frac{X_n}{m^n}$ is a martingale (why is it in L^1?). Since it is nonnegative, Theorem 9.3 (applied to $-M_n$) implies the existence of Z in L^1 such that $M_n \xrightarrow{\text{a.s.}} Z$ (but we might not have convergence in L^1).

How to use this martingale convergence? We now need to distinguish 3 cases

$\boldsymbol{m < 1}$**:** Observing that either $M_n = 0$ or $M_n \ge \frac{1}{m^n} \to +\infty$, we see that M_n can only converge if $M_n = 0$ for n large enough. Therefore, a.s., there exists n_0 (depending on ω) s.t. $M_n = 0$ for $n \ge n_0$. This implies that $X_n = 0$ for $n \ge n_0$, i.e. we have almost sure extinction.

Note: consequently in this case, $Z = 0$. On the other hand, for all $n \ge 0$, we have $\mathbb{E}[M_n] = \mathbb{E}[M_0] = 1$ (martingales have constant expectation). In particular, we note that $\mathbb{E}[Z] = 0 \ne \lim \mathbb{E}[M_n] = 1$, so that the a.s. convergence $M_n \to Z$ does not hold in L^1.

$\boldsymbol{m = 1}$**:** In this case, $M_n = X_n$. Note that, for $n, k > 0$, $\mathbb{P}\big[X_{n+1} = k | X_n = k\big] < 1$ so that for fixed k and n_0,

$$\mathbb{P}\big[X_n = k \text{ for all } n \ge n_0\big] = 0.$$

Taking a countable union, we conclude that the probability that X_n converges (or equivalently stabilizes) to some $k > 0$ is equal to 0. Since X_n converges a.s. to Z, we conclude that necessarily $Z = 0$ a.s. This implies that a.s., $X_n = 0$ for $n \ge n_0$ (n_0 depends on ω), i.e. we have almost sure extinction.

***m* > 1:** In this case, one can prove that there is a nonzero probability of survival, i.e.

$$\mathbb{P}\big[X_n > 0 \text{ for all } n \geq 0\big] > 0.$$

For simplicity we only prove this assuming that $m_2 := \sum_{k\geq 0} k^2 \mu(\{k\})$ is finite.

We first prove that M_n is bounded in L^2. We have

$$\begin{aligned}\mathbb{E}\big[X_{n+1}^2 | \mathcal{F}_n\big] &= \mathbb{E}\left[\sum_{j,k\geq 1} \mathbf{1}[j \leq X_n]\mathbf{1}[k \leq X_n]\xi_j^{(n+1)}\xi_k^{(n+1)} \big| \mathcal{F}_n\right] \\ &= \sum_{j,k\geq 1} \mathbf{1}[j \leq X_n]\mathbf{1}[k \leq X_n]\,\mathbb{E}\left[\xi_j^{(n+1)}\xi_k^{(n+1)}\right],\end{aligned}$$

where we used that X_n is $\mathcal{F}_n$-measurable, while $\xi_j^{(n+1)}$ and $\xi_k^{(n+1)}$ are independent from $\mathcal{F}_n$. The expectation in the last line is m_2 when $j = k$ and m^2 when $j \neq k$ (in this case $\xi_j^{(n+1)}$ and $\xi_k^{(n+1)}$ are independent one from the other). We get

$$\mathbb{E}\big[X_{n+1}^2 | \mathcal{F}_n\big] = X_n\, m_2 + (X_n^2 - X_n)m^2 = (m_2 - m^2)X_n + m^2 X_n^2.$$

Dividing by m^{2n+2} and taking expectation, we have

$$\mathbb{E}\big[M_{n+1}^2\big] = \frac{m_2 - m^2}{m^{n+2}}\mathbb{E}[M_n] + \mathbb{E}\big[M_n^2\big].$$

But $\mathbb{E}[M_n] = \mathbb{E}[M_0] = 1$ (since M_n is a martingale) and $\mathbb{E}[M_0^2] = 1$, so that

$$\mathbb{E}\big[M_n^2\big] = 1 + \sum_{k=0}^{n-1} \frac{m_2 - m^2}{m^{k+2}},$$

which is bounded since $\frac{m_2 - m^2}{m^{k+2}}$ is the general term of a convergent geometric series (the common ratio is $m^{-1} < 1$). This proves that M_n is bounded in L^2 and therefore the a.s. convergence $M_n \to Z$ holds in L^2 as well (Theorem 10.5). We therefore have $\mathbb{E}[Z] = \lim_{n\to\infty} \mathbb{E}[M_n] = \mathbb{E}[M_0] = 1$. This proves $\mathbb{P}[Z > 0] > 0$. But $Z > 0$ implies that X_n is nonzero for n sufficiently large (and hence for all n, since $X_{n_0} = 0$ implies $X_n = 0$ for all $n \geq n_0$). We have proved that there is a positive probability of survival.

Note: it is easy to show that the probability of extinction is also nonzero, indeed

$$\mathbb{P}\big[X_n = 0 \text{ for some } n \geq 0\big] \leq \mathbb{P}[X_1 = 0] = \mu(\{0\}) > 0.$$

Part C. Markov chains

13. Basics

Framework. As in the previous chapter, we will consider some random processes $(X_n)_{n\geq 0}$ on a (filtered) probability space. We additionally make an important assumption throughout Part C: X_n takes value in a *countable space* S (called "state space"). This implies:

- for fixed N, the law of X_N is determined by the individual probabilities $(\mathbb{P}[X_N = s])_{s\in S}$.
- the law of the whole process $(X_n)_{n\geq 0} : \Omega \to S^{\mathbb{N}}$ is determined by the *cylinder probabilities*

$$\big(\mathbb{P}\big[X_0 = s_0, X_1 = s_1, \cdots, X_N = s_N\big]\big)_{N\geq 0, s_0,\dots,s_N\in S}.$$

(This is a consequence of Kolmogorov extension theorem.)

13.1. The Markov property.

Definition 13.1. *A random process $(X_n)_{n\geq 0}$ with values in a countable space S has the Markov property if one has*

$$\mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_0 = s_0, \cdots, X_n = s_n\big] = \mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_n = s_n\big],$$

for all $n \geq 0$, $s_0, s_1, \cdots, s_n, s_{n+1} \in S$ such that $\mathbb{P}[X_0 = s_0, \cdots, X_n = s_n] > 0$.

A random process $(X_n)_{n\geq 0}$ with the Markov property is called a Markov process.

In words, the law of X_{n+1} knowing $(X_0, \cdots, X_n)$ is the same as that knowing X_n. This means that if, at time n, you want to predict the next state X_{n+1} and if you know the current state X_n, then the previous states $X_0, \cdots, X_{n-1}$ will not bring you extra information. Such models are sometimes called *memoryless*: knowing the present, the future is independent from the past.

The Markov property in fact implies more generally the following:

Lemma 13.2. *Let (X_n) be a Markov process. Then, for any $A \subseteq S^n$ and $B \subseteq S^p$, we have*

$$\mathbb{P}\big[(X_{n+1}, \cdots, X_{n+p}) \in B \big| X_n = s_n, (X_0, \cdots, X_{n-1}) \in A\big] = \mathbb{P}\big[(X_{n+1}, \cdots, X_{n+p}) \in B \big| X_n = s_n, \tag{14}$$

as soon as the conditioning events have nonzero probability.

We will use the above lemma repeatedly saying only "using the Markov property".

Proof. We first want to prove, that, for any $n \geq 0$, s_n, s_{n+1} in S and $A \subseteq S^n$, we have

$$\mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_n = s_n, (X_0, \cdots, X_{n-1}) \in A\big] = \mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_n = s_n\big] \tag{15}$$

The Markov property corresponds to (15) when A is a singleton (such that the conditioning event is nonzero). Assume (15) holds for two disjoint sets A_1 and A_2: then, we have

$$\begin{aligned}\frac{\mathbb{P}\big[X_{n+1} = s_{n+1}, X_n = s_n, (X_0, \cdots, X_{n-1}) \in A_1\big]}{\mathbb{P}\big[X_n = s_n, (X_0, \cdots, X_{n-1}) \in A_1\big]} &= \mathbb{P}\big[X_{n+1} = s_{n+1} | X_n = s_n\big] \\ &= \frac{\mathbb{P}\big[X_{n+1} = s_{n+1}, X_n = s_n, (X_0, \cdots, X_{n-1}) \in A_2\big]}{\mathbb{P}\big[X_n = s_n, (X_0, \cdots, X_{n-1}) \in A_2\big]}\end{aligned}$$

Using the easy fact that $\frac{a}{b} = X = \frac{c}{d}$ implies $\frac{a+c}{b+d} = X$ as well, we have

$$\frac{\mathbb{P}\big[X_{n+1} = s_{n+1}, X_n = s_n, (X_0, \cdots, X_{n-1}) \in (A_1 \uplus A_2)\big]}{\mathbb{P}\big[X_n = s_n, (X_0, \cdots, X_{n-1}) \in (A_1 \uplus A_2)\big]} = \mathbb{P}\big[X_{n+1} = s_{n+1} | X_n = s_n\big].$$

This proves that Eq. (15) holds for $A_1 \uplus A_2$. We conclude that it holds for all $A \subseteq S^n$ (such that the conditioning event is nonzero).

Now, taking in addition s_{n+2} in S, we have, using the definition of conditional probabilities

$$\begin{aligned}&\mathbb{P}\big[X_{n+1} = s_{n+1}, X_{n+2} = s_{n+2} \big| (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big] \\ &\qquad = \mathbb{P}\big[X_{n+2} = s_{n+2} \big| X_{n+1} = s_{n+1}, (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big] \\ &\qquad\qquad \cdot \mathbb{P}\big[X_{n+1} = s_{n+1} \big| (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big].\end{aligned}$$

We can apply twice (15) (once with n and A, and once with $n+1$ and $A \times s_n$) and get

$$\begin{aligned}&\mathbb{P}\big[X_{n+1} = s_{n+1}, X_{n+2} = s_{n+2} \big| (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big] \\ &\qquad = \mathbb{P}\big[X_{n+2} = s_{n+2} \big| X_{n+1} = s_{n+1}\big] \mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_n = s_n\big]\end{aligned}$$

But applying again (15), we have

$$\mathbb{P}\big[X_{n+2} = s_{n+2} \big| X_{n+1} = s_{n+1}\big] = \mathbb{P}\big[X_{n+2} = s_{n+2} \big| X_{n+1} = s_{n+1}, X_n = s_n\big].$$

This implies

$$\begin{aligned}&\mathbb{P}\big[X_{n+1} = s_{n+1}, X_{n+2} = s_{n+2} \big| (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big] \\ &\qquad = \mathbb{P}\big[X_{n+2} = s_{n+2} \big| X_{n+1} = s_{n+1}, X_n = s_n\big] \cdot \mathbb{P}\big[X_{n+1} = s_{n+1} \big| X_n = s_n\big] \\ &\qquad = \mathbb{P}\big[X_{n+1} = s_{n+1}, X_{n+2} = s_{n+2} \big| X_n = s_n\big],\end{aligned}$$

where the last equality is a simple manipulation of conditional expectation.

An easy induction on p with similar arguments proves

$$\begin{aligned}&\mathbb{P}\big[X_{n+1} = s_{n+1}, \cdots, X_{n+p} = s_{n+p} \big| (X_0, \cdots, X_{n-1}) \in A, X_n = s_n\big] \\ &\qquad = \mathbb{P}\big[X_{n+1} = s_{n+1}, \cdots, X_{n+p} = s_{n+p} \big| X_n = s_n\big].\end{aligned}$$

Summing over tuples $(s_{n+1}, \cdots, s_{n+p})$ in some set B we get (14). $\square$

Warning: however, taking $C \subset S$, in general we have

$$\mathbb{P}\big[X_{n+1} = x_{n+1}\big|X_n \in C, X_{n-1} = x_{n-1}\big] \neq \mathbb{P}\big[X_{n+1} = x_{n+1}\big|X_n \in C\big]$$

Example: let $X_n = Y_1 + \cdots + Y_n$ be a simple random walk on $\mathbb{Z}$ (i.e. the Y_i are i.i.d. uniform in $\{-1, +1\}$). It is a Markov chain (see Section 14.4.2 later). Note that

$$(X_2 \in \{-2, 2\}) \wedge (X_1 = 1) \Rightarrow X_2 = 2,$$

so that

$$\mathbb{P}\big[X_3 = 3|X_2 \in \{-2, 2\}, X_1 = 1\big] = \mathbb{P}\big[X_3 = 3|X_2 = 2, X_1 = 1\big] = \mathbb{P}\big[X_3 = 3|X_2 = 2\big] = 1/2.$$

On the other hand

$$\mathbb{P}\big[X_3 = 3|X_2 \in \{-2, 2\}\big] = \frac{\mathbb{P}\big[X_3 = 3 \wedge X_2 \in \{-2, 2\}\big]}{\mathbb{P}[X_2 \in \{-2, 2\}]} = \frac{\mathbb{P}\big[X_3 = 3 \wedge X_2 = 2\big]}{\mathbb{P}[X_2 = -2] + \mathbb{P}[X_2 = 2]} = \frac{1/8}{1/2} = 1/4.$$

13.2. **Transition matrices.** Let $(X_n)_{n\geq 0}$ be a random process with values in S. For x, y in S and $i \geq 0$, we denote

$$Q_i(x, y) = \begin{cases} \mathbb{P}[X_{i+1} = y|X_i = x] & \text{when } \mathbb{P}[X_i = x] > 0; \\ \text{any value} & \text{otherwise.} \end{cases} \tag{16}$$

Q_i is a function $S \times S \to \mathbb{R}$. We will think at Q_i as a matrix (possibly of infinite size, whose rows and columns are indexed by S) and call Q_i the transition matrix at time i.

Lemma 13.3. *If $(X_n)_{n\geq 0}$ has the Markov property, then*

$$\mathbb{P}\big[X_0 = s_0, X_1 = s_1, \cdots, X_N = s_N\big] = \mathbb{P}[X_0 = s_0]Q_0(s_0, s_1)\cdots Q_{N-1}(s_{N-1}, s_N)$$

In particular the distribution of a Markov process (X_n) is determined by its initial distribution $(\mathbb{P}[X_0 = s_0])_{s_0\in S}$ and the transition matrices $(Q_i)_{i\geq 0}$.

Proof. Assuming the probability of the conditioning events are non-zero, we have

$$\begin{aligned}\mathbb{P}\big[X_0 = s_0, X_1 = s_1, \cdots, X_N = s_N\big] &= \mathbb{P}\big[X_N = s_N|X_0 = s_0, \cdots, X_{N-1} = s_{N-1}\big]\,\mathbb{P}\big[X_0 = s_0, \cdots, X_{N-1} = s_{N-1}\big] \\ &= \mathbb{P}\big[X_N = s_N|X_{N-1} = s_{N-1}\big]\,\mathbb{P}\big[X_0 = s_0, \cdots, X_{N-1} = s_{N-1}\big] \\ &= Q(s_{N-1}, s_N)\,\mathbb{P}\big[X_0 = s_0, \cdots, X_{N-1} = s_{N-1}\big],\end{aligned}$$

where we used the Markov property in the second line. Note that the last equality holds also in the case $\mathbb{P}\big[X_0 = s_0, \cdots, X_{N-1} = s_{N-1}\big] = 0$. Iterating it, we get the formula in the lemma. □

Definition 13.4. *Let S be a countable set. A "matrix" $Q : S \times S \to \mathbb{R}$ is called* stochastic *if it has nonnegative entries and row sums equal to* 1.

In equation, $Q(x, y) \geq 0$ for all x, y in S, and for x in S,

$$\sum_{y\in S} Q(x, y) = 1.$$

Lemma 13.5. *The transition matrices Q_i defined in* (16) *can be chosen as stochastic matrices.*

From now on, we assume that choices in (16) have been done such that Q_i is stochastic matrix.

Proof. It is clear that rows corresponding to some x with $\mathbb{P}[X_i = x] > 0$ (those whose entries are well-defined) have nonnegative entries, which sum up to 1 (they are conditional probabilities). The rows corresponding to some x with $\mathbb{P}[X_i = x] = 0$ can be filled with arbitrary values without any constraints: it is therefore straightforward to choose nonnegative values with sum 1. □

Given two $S \times S$ matrices R_1 and R_2, we want to define their product as

$$R_1R_2(x, y) = \sum_{z\in S} R_1(x, z)R_2(z, y). \tag{17}$$

However, when $|S| = \infty$, the sum might not be defined. This problem never occurs with stochastic matrices, as stated in the following lemma.

Lemma 13.6. *If R_1 and R_2 are stochastic matrices, then the product R_1R_2 is well-defined (in the sense that the RHS of* (17) *is convergent) and is a stochastic matrix.*

Proof. Assuming that R_1 and R_2 are stochastic matrices, the RHS of (17) is a sum of nonnegative terms and therefore is defined as an element of $\mathbb{R}_+ \cup \{+\infty\}$. Moreover, since entries of R_2 are all at most 1, we have

$$R_1R_2(x,y) = \sum_{z\in S} R_1(x,z)R_2(z,y) \leq \sum_{z\in S} R_1(x,z) = 1,$$

so that the product has in fact finite entries in $[0,1]$. Moreover, for x in S,

$$\sum_{y\in S} R_1R_2(x,y) = \sum_{y,z\in S} R_1(x,z)R_2(z,y) = \sum_{z\in S} R_1(x,z)\left(\sum_{y\in S} R_2(z,y)\right) = \sum_{z\in S} R_1(x,z) = 1,$$

where we used the fact that all summands are nonnegative to change the order of summation. □

13.3. **Chapman-Kolmogorov equations.** Let (X_n) be a random process. For $i \leq j$, we denote

$$Q_i^j(x,y) = \begin{cases} \mathbb{P}[X_j = y | X_i = x] & \text{if } \mathbb{P}[X_i = x] > 0; \\ \text{arbitrary value} & \text{otherwise.} \end{cases} \tag{18}$$

Proposition 13.7 (Chapman-Kolmogorov equations)**.** *When $(X_n)_{n\geq 0}$ is a Markov process, the arbitrary values in* (18) *can be chosen so that, for all $i < j < k$, we have*

$$Q_i^k = Q_i^j\, Q_j^k.$$

Proof. To simplify, we consider only the case where the relevant events have positive probability. For $x, y \in S$, we have

$$Q_i^k(x,y) = \mathbb{P}[X_k = y | X_i = x] = \sum_{z\in S} \mathbb{P}[X_k = y | X_j = z, X_i = x]\mathbb{P}[X_j = z | X_i = x],$$

where we use the law of total probability for conditional probabilities. Using the Markov property, we have $\mathbb{P}[X_k = y | X_j = z, X_i = x] = \mathbb{P}[X_k = y | X_j = z]$. Therefore

$$Q_i^k(x,y) = \sum_{z\in S} \mathbb{P}[X_k = y | X_j = z]\mathbb{P}[X_j = z | X_i = x] = \sum_{z\in S} Q_j^k(z,y)Q_i^j(x,z) = Q_i^jQ_j^k(x,z),$$

where we used the definition of the $Q_\cdot^\cdot$ and of a product of stochastic matrices. □

13.4. **Homogeneity.**

Definition 13.8. *A Markov process $(X_n)_{n\geq 0}$ is (time-)homogeneous if one can choose arbitrary values in* (16) *such that Q_i is independent of i.*

From now on, we only consider *homogeneous Markov processes*, which we call Markov chain. We denote Q the common values of the Q_i. We take it as a stochastic matrix, called the transition matrix of the Markov chain.

Important facts:

- The distribution of a Markov chain $(X_n)_{n\geq 0}$ is determined by the stochastic matrix Q and by the distribution of X_0 (called initial distribution). Conversely, given a probability distribution ν on a countable set S and an $S \times S$ stochastic matrix Q, one can construct a Markov chain $(X_n)_{n\geq 0}$ with initial distribution ν and transition matrix Q (admitted).
- Chapman-Kolmogorov equations for Markov chains give: for $i \leq j$,

  $$Q_i^j = Q^{j-i}, \text{ i.e., for all } x, y \text{ in } S, \text{ one has } \mathbb{P}[X_j = y | X_i = x] = Q^{j-i}(x,y)$$

  (whenever $\mathbb{P}[X_i = x] > 0$).

13.5. **Examples.** From now on, we use the matrix notation $Q = (Q_{x,y})_{x,y\in S}$ for transition matrices (instead of the above function notation $Q(x,y)$).

General pattern: when X_{n+1} is defined as a function of X_n and of some additional r.v. independent from $X_0, \cdots, X_n$, then (X_n) has a Markov property. If furthermore, this function and the law of the additional r.v. do not depend on n, then we have a Markov chain.

13.5.1. *I.i.d. random variables.* Let $(X_i)_{i\geq 0}$ be a sequence of i.i.d. r.v. on a countable set S with common distribution μ. Then $(X_i)_{i\geq 0}$ is a Markov chain with initial distribution μ and transition matrix $Q_{x,y} = \mu(y)$.

13.5.2. *Random walk on $\mathbb{Z}^d$.* Let $S = \mathbb{Z}^d$ and $(Y_i)_{i\geq 1}$ be i.i.d. r.v. on $\mathbb{Z}^d$ with common distribution μ. Set $X_n = Y_1 + \cdots + Y_n$ for $n \geq 0$. Then $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix $Q_{x,y} = \mu(\{y - x\})$. Indeed, the Markov property is clearly satisfied and

$$\mathbb{P}[X_{n+1} = y|X_n = x] = \mathbb{P}[Y_{n+1} = y - x|X_n = x] = \mathbb{P}[Y_{n+1} = y - x] = \mu(\{y - x\}),$$

where the second equality uses the independence of Y_{n+1} and X_n.

13.5.3. *Simple random walk on a graph.* Let $G = (V, E)$ be a graph. We assume V is countable and that for every v in V, one has $0 < \deg(v) < \infty$. Fix a vertex v_0 and set $\nu = \delta_{v_0}$, the Dirac measure in v_0. For x, y in V, we set

$$Q_{x,y} = \begin{cases} \frac{1}{\deg(x)} & \text{if } \{x, y\} \in E; \\ 0 & \text{otherwise.} \end{cases}$$

Fig. 3 shows these transition probabilities for a concrete graph G.

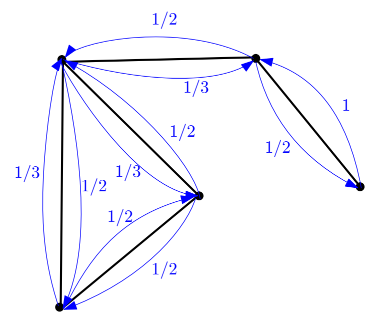


FIGURE 3. Transition probabilities of the simple random walk on a 5-vertex graph G

For any graph G, the associated Q is a stochastic matrix. The Markov chain $(X_n)_{n\geq 0}$ with initial distribution $\nu = \delta_{v_0}$ and transition matrix Q is called the *simple random walk on G starting at v_0*.

Informally, we start at v_0, i.e. $X_0 = v_0$ a.s. At each time $n \geq 0$, we choose X_{n+1} uniformly at random among the neighbours of X_n.

13.5.4. *Galton-Watson (GW) branching process.* Recall that the GW process with offspring distribution μ is defined by $X_0 = 1$ a.s. and $n \geq 0$,

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_k^{(n+1)},$$

where the $(\xi_k^{(n)})_{k,n\geq 1}$ are i.i.d. random variables with law μ.

This is a Markov chain, with initial distribution δ_1 and transition matrix

$$Q_{x,y} = \mathbb{P}\big[Z_1 + \cdots + Z_x = y\big] = \mu^{*x}(y),$$

where in the first expression, $(Z_1, \cdots, Z_x)$ are i.i.d. random variables of law μ and in the second one, μ^{*x} is the x-th convolution power of the μ.

14. Classification of states

Setting: we fix a transition matrix Q (=stochastic matrix) on a countable set S (=state space). For any initial distribution ν, there is a Markov chain $(X_n)_{n\geq 0}$ with initial distribution ν and transition matrix Q. We will denote probabilities/expectations with respect to this Markov chain as $\mathbb{P}_\nu$ and $\mathbb{E}_\nu$. For x in S, we simplify $\mathbb{P}_{\delta_x}$ and $\mathbb{E}_{\delta_x}$ to $\mathbb{P}_x$ and $\mathbb{E}_x$. Informally the index x means that the chain starts in x. (Some authors also write $\mathbb{P}[\cdots|X_0 = x]$.)

14.1. **Transience/recurrence.**

Definition 14.1. *A state x in S is called*

- recurrent *(or* persistent*) if* $\mathbb{P}_x\big[\exists n \geq 0 : X_n = x\big] = 1$*;*
- transient *if* $\mathbb{P}_x\big[\exists n \geq 0 : X_n = x\big] < 1$*;*

In words, $\mathbb{P}_x\big[\exists n \geq 0 : X_n = x\big]$ is the probability that, when starting at x, the chains comes back to x at some time (not specified in advance).

Goal: given a transition matrix Q, determine which states are transient/recurrent.

14.2. **The generating function approach.** We introduce additional notation: for x, y in S

$$F^{(n)}_{x,y} = \mathbb{P}_x\big[X_n = y \wedge (\forall k \in \{1, \cdots, n-1\},\, X_k \neq y)\big].$$

In words, $F^{(n)}_{x,y}$ is the probability that, starting at x, the *first "visit"* in state y occurs at time n (if $y = x$, we do not consider $n = 0$ as the first visit time; we will speak of the first "return" at x). Summing over n, we get the probability that the chain ever visit y (resp. ever returns at x), i.e.

$$F_{x,y} := \sum_{n \geq 0} F^{(n)}_{x,y} = \mathbb{P}_x\big[\exists n > 0, X_n = y\big].$$

By definition x is recurrent if and only if $F_{x,x} = 1$.

We would like a relation between Q and F (to get a recurrence criterion in terms of Q). To this end, we need to introduce *generating series*: for $t \in [0, 1)$, we set

$$Q_{x,y}(t) = \sum_{n \geq 0} t^n Q^n_{x,y};$$
$$F_{x,y}(t) = \sum_{n \geq 0} t^n F^{(n)}_{x,y}.$$

Since $Q^n_{x,y}$ and $F^{(n)}_{x,y}$ are smaller than 1, both series converge for $t < 1$.

Why generating series? For this to be interesting, we need two ingredients:

(1) One can find relations between the generating series;
(2) one can extract information on the object we are interested in (here $F_{x,x}$) from the generating series.

Proposition 14.2. *For t in $(0, 1)$ and x, y in S with $x \neq y$, we have*

(i) $Q_{x,x}(t) = 1 + F_{x,x}(t)Q_{x,x}(t)$
(ii) $Q_{x,y}(t) = F_{x,y}(t)Q_{y,y}(t)$.

Proof. Take x, y in S possibly equal. For $m, r \geq 1$, we write $A_m = \{X_m = y\}$ and

$$B_r = \{X_r = y,\, X_1 \neq y, \cdots,\, X_{r-1} \neq y\}.$$

In words, A_m means that the Markov chain visits y at time m, while B_r means that the Markov chain visits y *for the first time* at time r. Using $A_m = B_1 \uplus \cdots \uplus B_m$, we have, for $m \geq 1$,

$$Q^m_{x,y} = \mathbb{P}_x[X_m = y] = \mathbb{P}_x[A_m] = \sum_{r=1}^{m} \mathbb{P}_x[A_m \cap B_r] = \sum_{r=1}^{m} \mathbb{P}_x[A_m | B_r]\mathbb{P}_x[B_r]$$

Let us consider the conditional probability $\mathbb{P}_x[A_m|B_r]$; assume first $r < m$ and imagine we are at time r. The event B_r contains the information on the state at time r ($X_r = y$) and some information on the past (we never visited y before), while A_m is an event about the future ($X_m = x$ recall that $m > r$). By the Markov property, we can forget about the past and we have

$$\mathbb{P}_x[A_m|B_r] = \mathbb{P}[X_m = x | X_r = y] = Q^{m-r}_{y,y},$$

where the last equality is an application of the Chapman-Kolmogorov equations. On the other hand, by definition $\mathbb{P}_x[B_r] = F^{(r)}_{x,y}$, so that we get

$$Q^m_{x,y} = \sum_{r=1}^{m} Q^{m-r}_{y,y} F^{(r)}_{x,y}.$$

We multiply both sides by t^m and sum over $m \geq 1$,

$$Q_{x,y}(t) - Q^0_{x,y} = \sum_{m\geq 1} Q^m_{x,y} t^m = \sum_{m\geq 1}\left(\sum_{r=1}^{m} Q^{m-r}_{y,y} t^{m-r}\, F^{(r)}_{x,y} t^r\right)$$

Setting $p = m - r$ the sum over $m \geq 1$ and $r \in \{1, \cdots, m\}$ can be rewritten as a double sum over $r \geq 1$ and $p \geq 0$:

$$Q_{x,y}(t) - Q^0_{x,y} = \left(\sum_{p\geq 0} Q^p_{y,y} t^p\right)\left(\sum_{r\geq 1} F^{(r)}_{x,y} t^r\right) = Q_{y,y}(t)\, F_{x,y}(t).$$

Using $Q^0_{x,y} = \mathbf{1}[x = y]$, this proves both (i) and (ii). $\square$

The following lemma addresses item 2. above, i.e. how to extract the relevant information from the generating series.

Lemma 14.3 (Abel)**.** *Assume $a_n \in [0, 1]$ for all $n \geq 0$. Then*

$$\lim_{s\uparrow 1}\left(\sum_{n\geq 0} a_n s^n\right) = \sum_{n\geq 0} a_n.$$

The notation $s \uparrow 1$ means that we take the limit as s increases and tends to 1, i.e. with $s < 1$; as said above the sum $\sum_{n\geq 0} a_n s^n$ is absolutely convergent for $s < 1$ and hence $\sum_{n\geq 0} a_n s^n$ is a finite real number. Since this is an increasing function of s (recall $a_n \geq 0$), the limit as $s \uparrow 1$ exists, as an extended real number (it might be $+\infty$). Similarly, as an infinite sum of nonnegative real numbers, the RHS always exists as an extended real number.

Proof. Assume first $\sum_{n\geq 0} a_n < +\infty$. Then the series $\sum_{n\geq 0} a_n s^n$ of functions on $[0, 1]$ is normally convergent, so that the sum $\sum_{n\geq 0} a_n s^n$ is a continuous function on $[0, 1]$. This proves the lemma in this case.

We consider now the case $\sum_{n\geq 0} a_n = +\infty$. Fix $M > 0$, there exists $N > 0$ s.t. $\sum_{n=0}^{N} a_n \geq M$. We can then choose $\varepsilon > 0$ s.t. $(1-\varepsilon)^N \geq \frac{1}{2}$. For $s \geq 1 - \varepsilon$, it holds that

$$\sum_{n\geq 0} a_n s^n \geq \sum_{n=0}^{N} a_n s^N \geq \tfrac{1}{2}\sum_{n=0}^{N} a_n \geq \tfrac{1}{2}M.$$

Such an $\varepsilon > 0$ can be found for each $M > 0$, proving $\lim_{s\uparrow 1}\left(\sum_{n\geq 0} a_n s^n\right) = +\infty$. $\square$

Theorem 14.4. *Let Q be a transition matrix on S and x, y be states in S. Then*

(i) x is recurrent if and only if $\sum_{n\geq 0} Q^n_{x,x} = +\infty$;
(ii) if y is transient, then $\sum_{n\geq 0} Q^n_{x,y} < +\infty$ (in particular $Q^n_{x,y} \to 0$, as n tends to ∞);
(iii) if y is recurrent and $F_{x,y} > 0$, then $\sum_{n\geq 0} Q^n_{x,y} = +\infty$.

Proof. By Proposition 14.2, for t in $(0, 1)$, we have

$$Q_{x,x}(t) = \frac{1}{1 - F_{x,x}(t)}.$$

Using Abel's lemma, as $t \uparrow 1$, the quantity $F_{x,x}(t)$ tends to $F_{x,x}$. and $Q_{x,x}(t)$ tends to $\sum_{n\geq 0} Q^n_{x,x}$. Therefore

$$\sum_{n\geq 0} Q^n_{x,x} = \frac{1}{1 - F_{x,x}},$$

with the convention $\frac{1}{0} = +\infty$. We recall that x is recurrent if and only if $F_{x,x} = 1$. This implies that x is recurrent if and only if $\sum_{n\geq 0} Q^n_{x,x} = +\infty$, proving (i).

(ii) and (iii) follow easily using again Abel's lemma and Proposition 14.2, item (ii). $\square$

14.3. **Relations between states.**

Definition 14.5. *Let x, y be in S with $x \neq y$, we say that*

- *x communicates with y if $F_{x,y} > 0$ or equivalently $Q^n_{x,y} > 0$ for some $n \geq 0$ (notation: $x \longrightarrow y$);*
- *x and y intercommunicate if $x \longrightarrow y$ and $y \longrightarrow x$ (notation $x \longleftrightarrow y$).*

By convention, $x \longrightarrow x$ and $x \longleftrightarrow x$.

Informally, $x \longrightarrow y$ means that we can go from x to y. The relation $x \longrightarrow y$ is transitive, while $x \longleftrightarrow y$ is an equivalence relation. Equivalence classes are in many examples of Markov chains immediate to determine.

Lemma 14.6. *If x is recurrent and $x \longrightarrow y$, then y is recurrent and necessarily, $y \longrightarrow x$.*

Proof. Since $x \longrightarrow y$, there exists $p \geq 0$ such that $Q^p_{x,y} \geq 0$. We set $C = Q^p_{x,y}$. For $n \geq 0$, we have

$$Q^{n+p}_{x,y} = \sum_{z \in S} Q^n_{x,z}\, Q^p_{z,y} \geq Q^n_{x,x}\, Q^p_{x,y} = C\, Q^n_{x,x}.$$

This implies

$$\sum_{n \geq 0} Q^n_{x,y} \geq \sum_{n \geq 0} Q^{n+p}_{x,y} \geq C \left(\sum_{n \geq 0} Q^n_{x,x} \right).$$

But the latter sum is infinite (Theorem 14.4). We conclude that $\sum_{n \geq 0} Q^n_{x,y} = +\infty$, proving that y is recurrent (contrapositive of Theorem 14.4, item (ii)).

We now prove that $y \longrightarrow x$ (wlog, we assume $x \neq y$). To this end, take n_0 to be the minimal $n > 0$ such that $Q^n_{x,y} > 0$ (such an n exists since $X \longrightarrow y$). For i in $\{1, \cdots, n_0 - 1\}$, we have

$$\mathbb{P}_x\big[X_i = x, X_{n_0} = y\big] = \mathbb{P}_x\big[X_i = x\big]\mathbb{P}_x\big[X_{n_0} = y | X_i = x\big].$$

Using the Markov property, the second factor is $Q^{n_0-i}_{x,y}$. The latter is 0 by minimality of n_0. Thus, for any i in $\{1, \cdots, n_0 - 1\}$, we have $\mathbb{P}_x\big[X_i = x, X_{n_0} = y\big] = 0$. On the other hand $\mathbb{P}_x\big[X_{n_0} = y\big] = Q^{n_0}_{x,y} > 0$. Both statements together imply

$$\mathbb{P}_x\big[X_{n_0} = y, X_1 \neq x, \cdots, X_{n_0-1} \neq x\big] > 0,$$

i.e. we hit y (at time n_0) before coming back to x with positive probability.

Assume now for the sake of contradiction, that we cannot go from y to x (i.e. $y \not\rightarrow x$). Then, whenever we hit y before coming back to x, we never go back to x. Formally, for $j > n_0$, $\mathbb{P}_x\big[X_{n_0} = y, X_j = x\big] = Q^{j-n_0}_{x,y} = 0$. Combining with the above, we have

$$\mathbb{P}_x\big[X_{n_0} = y, X_1 \neq x, \cdots, X_{n_0-1} \neq x, X_{n_0+1} \neq x, X_{n_0+2} \neq x, \cdots\big] > 0.$$

Consequently, with positive probability, we never go back to x. This contradicts x being recurrent. We conclude that $y \longrightarrow x$. ☐

The lemma has the two following consequences, useful to identify transient/recurrent states.

Corollary 14.7. *If $x \rightarrow y$ but $y \not\rightarrow x$, then x is transient.*

Corollary 14.8. *If $x \longleftrightarrow y$, then x and y are either both recurrent or both transient.*

Definition 14.9. *A subset C of S is called:*

- closed *if for any x in C and y in S such that $x \longrightarrow y$, we have $y \in C$.*
- irreducible *if for any x, y in C, we have $x \longleftrightarrow y$.*

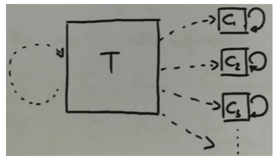
When the whole space state S is irreducible, we simply say that the Markov chain is irreducible.

Theorem 14.10 (decomposition theorem for Markov chains)**.** *Let Q be a transition matrix on S. The state space S can be uniquely partitioned as*

$$S = T \uplus C_1 \uplus C_2 \uplus \cdots, \tag{19}$$

where T is the set of transient states, and the C_i are closed irreducible subsets of recurrent states (there can be finitely many or infinitely many of them).

Here is a schematic illustration of the theorem, where the dotted arrow represent possible transitions and plain arrows are sure transitions (when you are in some C_i, you stay in C_i with probability 1).



We note that the transient part can be one or several equivalence classes. It might be possible to jump from one to another of these classes.

Proof. Let $(C_i)_{1\le i\le k}$ be the equivalence classes for $\longleftrightarrow$ on the set of recurrent states ($k \in \mathcal{N}_0 \cup \{+\infty\}$). By definition, each C_i is irreducible and (19) holds. We need to show that the C_i's are closed.

Take x in C_i and y in S such that $x \longrightarrow y$. From Lemma 14.6, we know that y is recurrent and that $y \longrightarrow x$. Therefore, we have $x \longleftrightarrow y$, and y is in the same $\longleftrightarrow$ equivalence class as x, i.e. in C_i. This concludes the proof. $\square$

We finish by a simple lemma, useful in some examples

Lemma 14.11. *Let C be a closed finite subset of S. Then C contains at least a recurrent state.*

Proof. For the sake of contradiction, we assume that all states in C are transient. Fix x in C and consider $Q^n_{x,y}$ for all y and large n:

- if y is not in C, then $Q^n_{x,y} = 0$ since C is closed;
- if y is in C, then y is transient and $Q^n_{x,y} \to 0$.

We conclude that $\sum_y Q^n_{x,y} \to 0$ (each of the finitely many non-zero terms tends to 0). But the $\sum_y Q^n_{x,y} = 1$ for each n. We have reached a contradiction, proving that C contains at least one recurrent state. $\square$

14.4. **Examples.** We discuss recurrence and transience in the examples of the previous chapter.

14.4.1. *I.i.d. r.v.* Here (X_n) is a sequence of i.i.d. r.v. of distribution μ. We have $Q^n_{x,y} = \mu(y)$ for all x, y in S and $n \ge 0$. Then $x \longrightarrow y$ if and only if $\mu(y) > 0$ and, similarly, y is recurrent if and only if $\mu(y) > 0$.

We have the decomposition $S = T \uplus C_1$, where the set of transitive states $T = \{y \in S : \mu(y) = 0\}$ is a union of singleton equivalence class, while the set of recurrent states consists of a unique close irreducible component $C_1 = \{y \in S : \mu(y) > 0\}$.

14.4.2. *Random walk on $\mathbb{Z}^d$.* Here $X_n = Y_1 + \cdots + Y_n$, where the Y_i's are i.i.d. r.v. of distribution μ. We recall that X_n is a Markov chain with transition matrix $Q_{x,x'} = \mu(\{x' - x\})$.

We have $x \longrightarrow x'$ if there exists n and $y_1, \cdots, y_n$ such that

$$\mu(y_1), \cdots, \mu(y_n) > 0 \text{ and } x' = x + y_1 + \cdots + y_n.$$

Defining $\mathrm{Supp}(\mu) = \{y \in \mathbb{Z}^d : \mu(y) > 0\}$, this happens if and only if $x' - x$ is in the semigroup G_μ generated by $\mathrm{Supp}(\mu)$.

Case 1: G is not a group, i.e there exists g in G such that $-g$ is not in G (since G is a semi-group, this is the only group axiom that might fail). Then, for any x in $\mathbb{Z}^d$, we have $x \longrightarrow x + g$ but $x + g \not\longrightarrow x$. From Lemma 14.6, this proves that x is transient.

Case 2: G is a group. There is no general answer, for some d and μ, the random walk is transient (i.e. all states are transient), in other cases it is recurrent (i.e. all states are recurrent); see Section 14.5 for examples.

14.4.3. *Random walk on a graph.* Here, states are vertices, and $x \longrightarrow y$ if there is a path from x to y. In particular $x \longrightarrow y$, is a symmetric relation and the closed and irreducible subsets of S are the same. The $\longrightarrow$ equivalence classes are the vertex-sets of the *connected components* of the graphs.

We claim that when $|V| < \infty$, all states are recurrent. Indeed, by Lemma 14.11, each connected component contains at least one recurrent state (since it's finite and closed). But since all states in a connected components are equivalent, every state in the component is recurrent (Corollary 14.8). This is valid for each component, so for each state of the Markov chain.

For infinite graphs, we can have both transient or recurrent connected components (see, again, Section 14.5, for examples).

14.4.4. *Branching processes.* Here, X_n is a Markov chain with transition matrix $Q_{x,y} = \mu^{*x}(y)$. In particular for any $x > 0$, we have $Q_{x,0} = \mu(0)^x > 0$ (we assumed $\mu(0) > 0$ so that extinction can happen). On the other hand $Q_{0,0} = 1$, and $Q_{0,x} = 0$ for any $x > 0$ (such a state is called an *absorbing* state). This shows that any $x > 0$ is transient; 0 is trivially recurrent.

14.5. **The simple random walk on $\mathbb{Z}^d$.** Determining recurrence or transience is in general a difficult problem and might need important computations. We study here the classical setting of the simple random walk on $\mathbb{Z}^d$. Calling $(e_j)_{j\le d}$ the unit vectors of $\mathbb{Z}^d$, we let $(Y_i)_{i\ge 1}$ be i.i.d. r.v. with $\mathbb{P}[Y_i = \pm e_j] = (2d)^{-1}$ and set $X_n = Y_1 + \dots + Y_n$.

Theorem 14.12. *[Pólya recurrence theorem, 1921] The simple random walk on $\mathbb{Z}^d$ is recurrence for $d = 1$ or $d = 2$, and transient for $d \ge 3$.*

Note: this is one of the most well-known theorem in probability theory. The change of qualitative behaviour of the simple random walk depending on the dimension is remarkable.

Proof. Let $\mathbf{0}_d = (0, \dots, 0) \in \mathbb{Z}^d$. We say that $(x_1, \cdots, x_{2n}$ is a path of length $2n$ starting at $\mathbf{0}_d$ if, for each $i \le 2n$, $|x_i - x_{i-1}| = 1$ (setting $x_0 = \mathbf{0}_d$). For such a path, we have

$$\mathbb{P}_{\mathbf{0}_d}\big[X_1 = x_1, \cdots, X_{2n} = x_{2n}\big] = (2d)^{-2n},$$

while this probability is 0 for other lists $(x_1, \cdots, x_{2n})$. Therefore we have

$$Q^{2n}_{\mathbf{0}_d,\mathbf{0}_d} = \mathbb{P}_{\mathbf{0}_d}\big[X_{2n} = \mathbf{0}^d\big] = \sum_{x_1,\dots,x_{2n-1}\in\mathbb{Z}^d} \mathbb{P}_{\mathbf{0}_d}\big[X_1 = x_1, \cdots, X_{2n} = \mathbf{0}^d\big] = (2d)^{-2n} p_{2n}(d),$$

where $p_{2n}(d)$ is the number of paths of length $2n$ starting *and ending* at $\mathbf{0}_d$. (For obvious parity reasons, $Q^{2n+1}_{\mathbf{0}_d,\mathbf{0}_d} = 0$ for all n.)

We count such paths as follows. A path starting at $\mathbf{0}^d$ is a word in the alphabet $\{I_1, D_1, \cdots, D_d, D_d\}$, a letter I_j (resp. D_j) at position i indicating that $x_i - x_{i-1} = e_j$ (resp. $x_i - x_{i-1} = -e_j$). The path ends at $\mathbf{0}^d$, if and only if, for all $j \le d$, there are as many I_j as D_j. We call a_j this number. Then $a_1, \cdots, a_d$ satisfies $a_1 + \dots + a_d = n$: indeed the total number of letters is $2a_1 + \ldots + 2a_d$, and is also the length of the path, i.e. $2n$. Given $a_1, \cdots, a_d$ with $a_1 + \dots + a_d = n$, the number of words with a_1 times the letter I_1, a_1 times the letter D_1, a_2 times the letter I_2, and so on, is the multinomial

$$\binom{2n}{a_1, a_1, a_2, a_2, \dots, a_d, a_d} = \frac{(2n)!}{(a_1!)^2 \cdots (a_d!)^2}.$$

Since the numbers $a_1, \cdots, a_d$ are not prescribed in advance, we should sum over their possible values to get the total number of paths, i.e.

$$p_{2n}(d) = \sum_{\substack{a_1,\dots,a_d\ge 0\\ a_1+\dots+a_d=n}} \frac{(2n)!}{(a_1!)^2 \cdots (a_d!)^2}.$$

We now analyze the convergence of $\sum_{n\ge 0} Q^{2n}_{\mathbf{0}_d,\mathbf{0}_d} = \sum_{n\ge 0} (2d)^{-2n} p_{2n}(d)$, splitting the discussion depending on the dimension.

***d* = 1**: in this case, we have

$$p_{2n}(1) = \frac{(2n)!}{(n!)^2}, \quad Q^{2n}_{0,0} = 2^{-2n}\frac{(2n)!}{(n!)^2}.$$

Using Stirling formula we get $Q^{2n}_{0,0} \sim (\pi n)^{1/2}$. Since the latter is the general term of a divergent series, so is $Q^{2n}_{0,0}$. This implies that 0 is recurrent (Theorem 14.4). Using the irreducibility (or the translation invariance) of the chain, we know that all states are recurrent.

***d* = 2**: in this case we can prove that $p_{2n}(2)$ simplifies to

$$p_{2n}(2) = \left(\frac{(2n)!}{(n!)^2}\right)^2.$$

This implies $Q^{2n}_{\mathbf{0}_2,\mathbf{0}_2} \sim (\pi n)^{-1}$. This is the term of a divergent series, and we conclude as for $d = 1$, that the chain is recurrent.

***d* ≥ 3**: here, it is not easy to find an asymptotic equivalent for $p_{2n}(d)$, but we will give an upper bound (since we want to prove transient here, i.e. that $\sum_{n\geq 0}(2d)^{-2n}p_{2n}(d)$, we need only an upper bound). We start by the case where n is divisible by d, i.e. $n = dr$. Observe that if $a_1 + \cdots + a_d = dr$, we have the inequality $a_1! \cdots a_d! \geq (r!)^d$ (to see that, compare the factors on both sides). We therefore have

$$p_{2dr}(d) \leq \frac{(2dr)!}{(r!)^d} \sum_{\substack{a_1,\dots,a_d\geq 0\\ a_1+\dots+a_d=dr}} \frac{1}{(a_1!)\cdots(a_d!)}.$$

But the sum is easily seen to be equal to $d^{dr}/(dr)!$. We therefore have

$$p_{2dr}(d) \leq \frac{(2dr)!\, d^{dr}}{(r!)^d (dr)!} \sim_{n\to\infty} (2d)^{2dr} \frac{2}{\sqrt{2\pi r}^d}.$$

This implies that the sum $\sum_{r\geq 0} Q^{2dr}_{\mathbf{0}_d,\mathbf{0}_d}$ is convergent for $d \geq 3$.

To generalize this to the sum of all values $p_{2n}(d)$ of n (not only multiple of d), we observe that $p_{2n}(d)$ is increasing in n (appending I_1D_1 at the end of a path of length $2n$ gives a path of length $2n+2$ with the same ending point, and defines an injective procedure). Thus if n is in $\{(r-1)d+1,\cdots,rd\}$, we have

$$p_{2n}(d) \leq p_{2dr}(d), \quad Q^{2n}_{\mathbf{0}_d,\mathbf{0}_d} \leq (2d)^{2dr-2n} Q^{2dr}_{\mathbf{0}_d,\mathbf{0}_d} \leq (2d)^{2d} Q^{2dr}_{\mathbf{0}_d,\mathbf{0}_d}.$$

To each value of r correspond d values of n, i.e. n in $\{(r-1)d+1,\cdots,rd\}$. We get

$$\sum_{n\geq 0} Q^{2n}_{\mathbf{0}_d,\mathbf{0}_d} \leq d \sum_{r\geq 0} (2d)^{2d} Q^{2dr}_{\mathbf{0}_d,\mathbf{0}_d} < +\infty.$$

We conclude that $\mathbf{0}_d$ is transient (Theorem 14.4). Since the chain is irreducible (or using translation invariance), we know that all states are transient. □

15. The strong Markov property

15.1. **Statement.** Let $(X_n)_{n\geq 0}$ be a Markov chain with transition Q and initial distribution μ. Fix some time $k \geq 0$ and a state s such that $\mathbb{P}[X_k = s] > 0$. The Markov property and the time homogeneity implies the following:

> Conditionally on $\{X_k = s\}$, the time-shifted process $(X_{k+p})_{p\geq 0}$ has distribution $\mathbb{P}_s$ and is independent from $(X_n)_{n<p}$.

This is sometimes referred to as the weak Markov property.

We would like to replace the time n by a random time; as for martingale, we use the notion of stopping time We denote the standard filtration associated to (X_n) as $\mathcal{F}_n = \sigma(X_0,\dots,X_n)$. Given a stopping time T, we define $\mathcal{F}_T$ as the σ-algebra generated by

$$\big\{\{T = n\} \cap A,\ n \in \mathbb{N} \cup \{+\infty\} \text{ and } A \in \mathcal{F}_n\big\}.$$

Informally, this is the σ-algebra corresponding to everything that happens before the (random) time T.

Proposition 15.1 (strong Markov property)**.** *Let x be a state in S and T be a stopping time such that $T < +\infty$ implies $X_T = x$. Then conditionally on $T < +\infty$, we have $(X_{T+p})_{p\geq 0}$ has distribution $\mathbb{P}_x$ and is independent from $\mathcal{F}_T$.*

Note: the statement does not depend of the initial distribution μ.

Proof. Since $X_0,\dots,X_n$ takes values in the countable space S, events in $\mathcal{F}_n$ are exactly the disjoint unions of *atomic events*

$$\{X_0 = x_0,\dots,X_n = x_n\}, \text{ for } x_0,\dots,x_n \in S.$$

We are interested in events A in $\mathcal{F}_T$ such that $A \subseteq \{T < +\infty\}$ Such events are disjoint unions of events in the family

$$A_{n,x_0,\dots,x_n} := \{T = n, X_0 = x_0,\dots,X_n = x_n\}, \text{ for } n \geq 0,\ x_0,\dots,x_n \in S.$$

With this observation, the statement in the theorem rewrites as: for any $A_{n,x_0,\dots,x_n}$ with nonzero probability and any $s_0,\dots,s_k$, we have

$$\mathbb{P}_\mu\big(X_T = s_0,\dots,X_{T+k} = s_k | A_{n,x_0,\dots,x_n}\big) = \mathbb{P}_x(X_0 = s_0,\dots,X_k = s_k).$$

Fix an event $A_{n,x_0,\dots,x_n}$ with nonzero probability. We recall that since T is a stopping time, $\{T=n\}$ is in $\mathcal{F}_n$. Using the above description of events in $\mathcal{F}_n$, we know that

- either $\{X_0=x_0,\dots,X_n=x_n\} \subseteq \{T=n\}$;
- or $\{X_0=x_0,\dots,X_n=x_n\} \cap \{T=n\} = \emptyset$.

The second case cannot occur since $P_\mu(A_{n,x_0,\dots,x_n})>0$. We are therefore in the first case and

$$A_{n,x_0,\dots,x_n} = \{X_0=x_0,\dots,X_n=x_n\}.$$

(The information $T=n$ in $A_{n,x_0,\dots,x_n}$ is redundant.) Moreover since $T<\infty$ implies $X_T=x$, we must have $x_n=x$, otherwise $A_{n,x_0,\dots,x_n}$ would have probability 0 Therefore, using the Markov property,

$$\mathbb{P}_\mu\big(X_T=s_0,\dots,X_{T+k}=s_k|A_{n,x_0,\dots,x_n}\big) = \mathbb{P}_\mu\big(X_n=s_0,\dots,X_{n+k}=s_k|X_0=x_0,\dots,X_n=x\big)$$
$$= \mathbb{P}_\mu\big(X_n=s_0,\dots,X_{n+k}=s_k|X_n=x\big) = \delta_{s_0,x}Q(s_0,s_1)\dots Q(s_{k-1},s_k) = \mathbb{P}_x\big(X_0=s_0,\dots,X_k=s_k\big). \quad \square$$

Comment on the terminology: While these are called weak and strong Markov property, it relies on the time homogeneity and not only on the Markov property.

15.2. **First application: visiting recurrent states.**

Proposition 15.2. *Let Q be a transition matrix on S and x,y be intercommunicating recurrent states in S. Then*

$$\mathbb{P}_x[\exists n>0 : X_n=y]=1.$$

Proof. Assume w.l.o.g. that $x\neq y$. Let T_x be, as usual, the first return time of the chain at x: $T_x<+\infty$ a.s. since x is recurrent. We set

$$p=\mathbb{P}_x\big[\exists n>0 : X_n=y\big], \quad q=\mathbb{P}_x\big[\exists n<T_x : X_n=y\big].$$

We have seen in the proof of Lemma 14.6 that $q>0$. Moreover, we have

$$p-q=\mathbb{P}_x\big[\exists n>T_x : X_n=y,\, \nexists n<T_x : X_n=y\big] = \mathbb{P}_x\big[\exists n>T_x : X_n=y \,|\, \nexists n<T_x : X_n=y\big]\cdot(1-q).$$

The event $\exists n>T_x : X_n=y$ can be rewritten in term of $\tilde{X}_p=X_{T_x+p}$ as $\exists p>0 : \tilde{X}_p=y$. On the other hand, the conditioning event $\nexists n<T_x : X_n=y$ is in $\mathcal{F}_{T_x}$. By Proposition 15.1, we know that $(\tilde{X}_p)_{p\geq 0}$ has the same distribution as $(X_n)_{n\geq 0}$ and is independent from $\mathcal{F}_{T_x}$. We conclude that

$$p-q=\mathbb{P}_x[\exists n>0 : X_n=y]\cdot(1-q)=p(1-q)=p-pq.$$

Since $q>0$, this implies $p=1$, which is what we wanted to prove. $\square$

15.3. **Second application: numbers of visits.** Let $(X_n)_{n\geq 0}$ be a Markov cahin on S. For y in S, we define

$$N_y=\big|\{n>0 : X_n=y\}\big| = \sum_{n>0}\mathbf{1}\{X_n=y\} \ \in \mathbb{N}_0\cup\{+\infty\}.$$

In words, this is the number of visits of the process in the state y. Its expectation under $\mathbb{P}_x$ is given by

$$\mathbb{E}_x[N_y]=\mathbb{E}_x\left[\sum_{n>0}\mathbf{1}\{X_n=y\}\right] = \sum_{n>0}\mathbb{P}_x[X_n=y] = \sum_{n>0}Q^n_{x,y}.$$

- If y is transient, we know that $\sum_{n>0}Q^n_{x,y}<+\infty$ (Theorem 14.4), i.e. $\mathbb{E}_x[N_y]<+\infty$. This implies $N_y<+\infty$ a.s.
- If y is recurrent and $x\longrightarrow y$, we know that $\sum_{n>0}Q^n_{x,y}=+\infty$ (again by Theorem 14.4), i.e. $\mathbb{E}_x[N_y]=+\infty$. It might still be the case that $N_y<+\infty$ a.s. We will prove with the strong Markov property that this is not the case.

We start by the case $y=x$.

Proposition 15.3. *If x is a recurrent state, then, $\mathbb{P}_x$-a.s., we have $N_x=+\infty$.*

By definition of *recurrence*, under distribution $\mathbb{P}_x$, we have $N_x\geq 1$ a.s. The proposition tells us that in fact $N_x=+\infty$ a.s.

Proof. Let $(X_n)_{n\geq 0}$ be a Markov chain on S with transition matrix Q and initial distribution δ_x. We set $T = \min\{n > 0 : X_n = x\}$. In words, T is the first return time in x. Since x is recurrent, we have $T < +\infty$ a.s. Furthermore, we note that $X_T = x$ a.s.

We consider the random sequence $(X_{T+p})_{p\geq 0}$ and denote

$$\tilde{N}_x = \big|\{p > 0 : X_{T+p} = x\}\big|$$

the number of visits of this process in x. We have $N_x = \tilde{N}_x + 1$ a.s. Indeed, both count the visits of (X_n) at state x, except that, for $\tilde{N}_x$, we do not count the first visit (which occurs at time T, by definition).

On the other hand, the strong Markov property (Proposition 15.1) implies that $(X_{T+p})_{p\geq 0}$ has distribution $\mathbb{P}_x$, i.e. the same as (X_n). This implies $N_x \stackrel{d}{=} \tilde{N}_x$.

Bringing everything together we have $N_x \stackrel{d}{=} \tilde{N}_x \stackrel{d}{=} N_x - 1$. We claim that this implies $N_x = +\infty$ a.s. Indeed, we first write

$$\mathbb{P}_x[N_x \geq k+1] = \mathbb{P}_x[N_x - 1 \geq k] = \mathbb{P}_x[N_x \geq k].$$

Iterating, we get that for all $k \geq 0$, it holds that $\mathbb{P}_x[N_x \geq k] = \mathbb{P}_x[N_x \geq 1] = 1$. Since this is valide for all k, this yields

$$\mathbb{P}_x[N_x = +\infty] = \lim_{k\to+\infty} \mathbb{P}_x[N_x \geq k] = 1.$$

The first equality follows from $\{N_x = +\infty\}$ being the countable intersection of $\{N_x \geq k\}$ (for $k \geq 1$). $\square$

In the general case where y is possibly different from x, we have the following result.

Proposition 15.4. *If y is a recurrent state, then*

$$\mathbb{P}_x\big[N_y = 0 \text{ or } N_y = +\infty\big] = 1.$$

Note: combining with Proposition 15.2, if we assume x recurrent as well and $x \leftrightarrow y$, we have $\mathbb{P}_x\big[N_y = +\infty\big] = 1$.

Proof. We let $T = \min\{n > 0 : X_n = y\}$ be the first visit time in y. If $T = +\infty$ (which might happen with positive probability), we have $N_y = 0$.

We now work conditionally on $T < +\infty$ (assuming that this has positive probability, otherwise the proposition is trivial). We have $X_T = y$. Applying the strong Markov property (Proposition 15.1) tells us that $(X_{T+p})_{p\geq 0}$ has distribution $\mathbb{P}_y$. Using Proposition 15.3, we know that a.s., $(X_{T+p})_{p\geq 0}$ visits y infinitely many times. Therefore, conditionally on $T < +\infty$, we have $N_y = +\infty$. $\square$

A consequence. We consider a random walk on $\mathbb{Z}$, i.e. $X_n = Y_1 + \cdots + Y_n$, where the $(Y_i)_{i\geq 1}$ are i.i.d. r.v. Assume that $\mathbb{E}[Y_1] = m \neq 0$. Then, X_n/n tends to m a.s. (strong law of large numbers), which implies $|X_n| \to +\infty$ a.s. In particular, a.s., $X_n = 0$ for finitely many n, i.e. $N_0 < +\infty$ $\mathbb{P}_0$-a.s. From the above results, this implies that 0 is transient. Using the invariance by translation of the model (if X_n has distribution $\mathbb{P}_0$, then $X_n + a$ has distribution $\mathbb{P}_a$), we deduce that all states are transient.

16. Stationary distribution, expected return times and null-recurrence

16.1. **The distribution of X_n.** Let μ_n be the distribution of X_n. Since X_n takes values in the countable set S, this distribution is determined by the point probabilities $(\mu_n(y))_{y\in S}$. We will think at this as a row vector.

Using the law of total probability and the definition of the transition matrix Q, we have

$$\mu_{n+1}(y) = \mathbb{P}[X_{n+1} = y] = \sum_{x\in S} \mathbb{P}\big[X_{n+1} = y | X_n = x\big]\mathbb{P}[X_n = x] = \sum_{x\in S} Q(x,y)\mu_n(x) = (\mu_n \cdot Q)(y).$$

(In the last expression, the product of a row vector and a matrix is defined as usual, i.e. $(v \cdot M)(y) = \sum_{x\in S} v(x)M(x,y)$; if the matrix is stochastic and the vector has nonnegative coordinates summing to 1, then the sum is convergent and the resulting vector has coordinates summing to 1.)

The above relation simply writes $\mu_{n+1} = \mu_n \cdot Q$. Iterating this relation, we get

Lemma 16.1. *The distribution μ_n of X_n is given by $\mu_n = \mu_0 Q^n$.*

16.2. **Stationary distribution.** Assume for a moment that μ_n tends to some limiting distribution μ_∞ ($\mu_n(x) \to \mu_\infty(x)$ for all x in S). Then the relation $\mu_{n+1} = \mu_n \cdot Q$ implies $\mu_\infty = \mu_\infty \cdot Q$ (at least heuristically, but one can justify the exchange of limits and of the infinite sum contained in the definition of Q).

This motivates the following definition:

Definition 16.2. *Let Q be a transition matrix on S and μ be a measure on S. Assume μ is nonzero (i.e. $\mu(S) > 0$) and locally finite (i.e. $\mu(x) < \infty$ for all x in S). Then μ is called stationary (or invariant) if we have $\mu = \mu \cdot Q$.*

Note: $\mu = \mu \cdot Q$ implies $\mu = \mu \cdot Q^n$ for all $n \geq 0$. We will use this naive remark repeatedly.

Though we are merely interested in stationary probability distribution (for the reason explained above), we define the notion of being stationary not only for probability distributions, but for all (locally finite) measures on S (possibly infinite, i.e. of infinite mass).

If μ is a stationary measure and α a positive real number, then $\alpha\mu$ is also a stationary measure. In particular, if there exists a stationary measure μ which is finite (i.e. of finite mass), then $\frac{1}{\mu(S)}\mu$ is stationary probability distribution.

It turns out that the existence of stationary measures is related to transience/recurrence.

Theorem 16.3. *Let Q be an irreducible transition matrix, then*

a. if states are transient, there are no finite stationary measures;
b. if states are recurrent, there exists a stationary measure which is unique up to multiplication by a constant.

Comments:

- In the transient case (a.), we have no information about the existence of infinite stationary measures. There are examples of transient irreducible Markov chains, for which no such measures exist, some for which there are several non-equivalent ones (i.e. which cannot be obtained one from another by multiplication by a constant).
- In the recurrent case (b.), we do not know whether the unique stationary measure is finite or infinite. Again, both might happen.
- The theorem can be used to prove transience/recurrence in some cases. If there is a finite stationary measure, then the states are necessarily recurrent. On the other hand, if one can find two non-equivalent stationary distribution, then states are necessarily transient. (In both cases, we assume Q irreducible, or apply the criterion within a close irreducible component.)

Proof. a. We proceed by contradiction: assume that there exists a finite stationary measure μ. By definition, we have $\mu = \mu Q$, which implies that for all $n \geq 0$, $\mu = \mu Q^n$. Unwrapping notation, we have that, for all y in S,

$$\mu(y) = \sum_{x \in S} \mu(x) Q^n_{x,y}.$$

But since states are transient, for any x, y in S, we have $\lim_{n\to\infty} Q^n_{x,y} = 0$ (Theorem 14.4), item (ii)). Using the dominated convergence theorem, we have

$$\lim_{n\to\infty} \sum_{x \in S} \mu(x) Q^n_{x,y} = \sum_{x \in S} \big(\lim_{n\to\infty} \mu(x) Q^n_{x,y} \big) = 0.$$

Indeed, using the inequality $|Q^n_{x,y}| \leq 1$, we see that the summand $x \mapsto \mu(x) Q^n_{x,y}$ is dominated by $x \mapsto \mu(x)$, which an integrable (or here rather a summable) function independent of n; hence we can apply the dominated convergence theorem.

From the two above displays, we conclude that $\mu(y) = 0$. This holds for all y in S, implying $\mu(S) = 0$. This is a contradiction since we excluded the null measure in the definition of stationary measures.

b. *Existence.* We fix x in S and consider, under distribution $\mathbb{P}_x$, the first return time T_x in x. In formula, $T_x = \min\{n > 0 : X_n = x\}$. We recall that $T_x < +\infty$ a.s. since x is recurrent.

Now, for y in S, we define

$$\nu^{(x)}(y) = \sum_{i=0}^{\infty} \mathbb{P}_x \left[X_i = y, i < T_x \right].$$

We claim that $\nu^{(x)}$ is a stationary measure on S.

First, we observe that $\nu^{(x)}$ is nonzero since trivially, $\nu^{(x)}(x) = 1$ (the summand for $i = 0$ is 1, other summands are 0). We postpone the proof that $\nu^{(x)}$ is locally finite since we will prove this as a consequence of the identity $\nu^{(x)} = \nu^{(x)}Q$. So let us prove $\nu^{(x)} = \nu^{(x)}Q$.

Fix y in S and consider

$$\sum_{z\in S} Q_{z,y}\,\nu^{(x)}(z) = \sum_{z\in S}\sum_{i=0}^{\infty} \mathbb{P}_x\big[X_i = z, i < T_x\big]\mathbb{P}_x\big[X_{i+1} = y | X_i = z\big] \tag{20}$$

Note that $T_x > i$ gives information on the Markov chain until time i (it has not yet returned to x). Using the Markov property, we have $\mathbb{P}_x\big[X_{i+1} = y|X_i = z\big] = \mathbb{P}_x\big[X_{i+1} = y|X_i = z, i < T_x\big]$, so that the above product of probabilities simplifies as

$$\begin{aligned}\mathbb{P}_x\big[X_i = z, i < T_x\big]\mathbb{P}_x\big[X_{i+1} = y|X_i = z\big] &= \mathbb{P}_x\big[X_i = z, i < T_x\big]\mathbb{P}_x\big[X_{i+1} = y|X_i = z, i < T_x\big] \\ &= \mathbb{P}_x\big[X_i = z, i < T_x, X_{i+1} = y\big].\end{aligned}$$

Putting this back into Eq. (20), exchanging the summation order (everything is nonnegative) and using the law of total probability with the partition $\biguplus_{z\in S}\{X_i = z\} = \Omega$, we have

$$\sum_{z\in S} Q_{z,y}\nu^{(x)}(z) = \sum_{i=0}^{\infty}\sum_{z\in S} \mathbb{P}_x\big[X_i = z, i < T_x, X_{i+1} = y\big] = \sum_{i=0}^{\infty} \mathbb{P}_x\big[i < T_x, X_{i+1} = y\big] \tag{21}$$

For $y = x$, we have the equality of events $\{i < T_x, X_{i+1} = y\} = \{T_x = i + 1\}$. But, when i goes from 0 to $+\infty$, the latter is a partition of the probability space Ω (recall that $T_x < +\infty$ a.s.). It follows that the probabilities in (21) sum up to 1 and that

$$\sum_{z\in S} Q_{z,x}\nu^{(x)}(z) = 1 = \nu^{(x)}(x).$$

For $y \neq x$, we cannot have both $X_{i+1} = y$ and $T_x = i + 1$, so that we can replace $i < T_x$ by $i + 1 < T_x$ in (21). Setting $j = i + 1$, we have

$$\sum_{z\in S} Q_{z,y}\nu^{(x)}(z) = \sum_{j=1}^{\infty} \mathbb{P}_x\big[j < T_x, X_j = y\big].$$

But $\mathbb{P}_x\big[X_0 = y\big] = 0$ so that starting the sum at $j = 0$ or $j = 1$ does not change its value. Up to this irrelevant change, the right hand side is the definition of $\nu^{(x)}(y)$, so that we have

$$\sum_{z\in S} Q_{z,y}\nu^{(x)}(z) = \nu^{(x)}(y) \tag{22}$$

also when $y \neq x$.

It remains to prove that $\nu^{(x)}$ is locally finite. As already explained, (22) rewrites as $\nu^{(x)} = \nu^{(x)}Q$ and thus implies $\nu^{(x)} = \nu^{(x)}Q^n$ for all n. Looking at the x-coefficient, we get

$$1 = \nu^{(x)}(x) = \sum_{z\in S} Q^n_{z,x}\nu^{(x)}(z).$$

Fix now z_0 in S. Since the Markov chain is irreducible, one can find n such that $Q^n_{z_0,x} > 0$. For such n, we have

$$Q^n_{z_0,x}\nu^{(x)}(z_0) \leq \sum_{z\in S} Q^n_{z,x}\nu^{(x)}(z) = 1,$$

implying $\nu^{(x)}(z_0) \leq 1/Q^n_{z_0,x} < +\infty$. Since this holds for any z_0 in S, this proves that $\nu^{(x)}$ is locally finite, as wanted.

Uniqueness. Let μ be a stationary measure and x be an element of S. We want to prove that $\mu = \mu(x)\cdot\nu^{(x)}$.

We first prove the inequality: for any y in S

$$\mu(y) \geq \mu(x)\cdot\nu^{(x)}(y). \tag{23}$$

To this end, we prove by induction on p the following statement:

$$(H_p) \qquad \mu(y) \geq \mu(x) \left[\sum_{i=0}^{p} \mathbb{P}_x[X_i = y, i < T_x] \right].$$

For $y = x$ and $i > 0$, we cannot have both $X_i = y = x$ and $T_x > i$. Hence, for $y = x$, summands corresponding to $i > 0$ above are zero and the sum in the bracket is 1. The inequality is trivially verified (for all p).

We now assume that $y \neq x$, implying in particular $\mathbb{P}_x[X_0 = y] = 0$. Then the RHS of (H_0) is 0 and (H_0) trivially holds. By induction we assume that (H_p) holds for all y and we want to prove (H_{p+1}). (The following arguments are similar to the existence proof.) Using the invariance of μ, the induction hypothesis for $\mu(z)$ and the definition of Q, we have

$$\mu(y) = \sum_{z \in S} \mu(z) Q_{z,y} \geq \sum_{z \in S} \mu(x) \left(\sum_{i=0}^{p} \mathbb{P}_x\{X_i = z, i < T_x\} \mathbb{P}_x\big[X_{i+1} = y | X_i = z\big] \right).$$

Using the Markov property, we have $\mathbb{P}_x\big[X_{i+1} = y | X_i = z\big] = \mathbb{P}_x\big[X_{i+1} = y | X_i = z, i < T_x\big]$, so that the above inequality rewrites

$$\mu(y) \geq \mu(x) \left(\sum_{i=0}^{p} \sum_{z \in S} \mathbb{P}_x\big[X_{i+1} = y, X_i = z, i < T_x\big] \right).$$

Note that since $x \neq y$ we cannot have both $X_{i+1} = y$ and $T_x = i + 1$. Therefore we can replace $i < T_x$ by $i + 1 < T_x$ in the above probability. We also use that $\biguplus_{z \in S} \{X_i = z\} = \Omega$ to simplify the above expression to

$$\mu(y) \geq \mu(x) \left(\sum_{i=0}^{p} \mathbb{P}_x\big[X_{i+1} = y, i + 1 < T_x\big] \right)$$

Setting $j = i + 1$, we get

$$\mu(y) \geq \mu(x) \left(\sum_{j=1}^{p+1} \mathbb{P}_x\big[X_j = y, j < T_x\big] \right).$$

Since $y \neq x$ and $X_0 = x$, we have $\mathbb{P}_x\big[X_0 = y, 0 < T_x\big] = 0$ and starting the above sum at $j = 0$ or $j = 1$ does not change its value. in the above sum would be 0 and we can add it for free, This proves (H_{p+1}) and concludes the induction: (H_p) holds for all p.

Making p go to infinity in (H_p) shows that for any y in S

$$\mu(y) \geq \mu(x) \left[\sum_{i=0}^{+\infty} \mathbb{P}_x[X_i = y, i < T_x] \right] \geq \mu(x) \cdot \nu^{(x)}(y). \tag{24}$$

(The sum is finite, as proved in the existence part.)

We still need to prove the reverse inequality: using the stationarity of μ and $\nu^{(x)}$ and the above inequality, we have the following: for any $n \geq 0$

$$\mu(x) = \sum_{y \in S} \mu(y) Q^n_{y,x} \geq \mu(x) \sum_{y \in S} \nu^{(x)}(y) Q^n_{y,x} = \mu(x) \nu^{(x)}(x) = \mu(x).$$

Thus the inequality above is in fact an equality: this means that for all y in S (and all $n \geq 0$), we have

$$\mu(y) Q^n_{y,x} = \mu(x) \nu^{(x)}(y) Q^n_{y,x}.$$

For any y in S we can find n such that $Q^n_{y,x} \neq 0$ (since the Markov chain is irreducible); the above equation then simplifies to $\mu(y) = \mu(x) \nu^{(x)}(y)$. □

16.3. **Null and nonnull recurrent states.** Consider a recurrent irreducible Markov chain Q. The above proof of existence of a stationary measure gives in fact more information since we have an explicit formula for it. In particular, the stationary measure $\nu^{(x)}$ constructed in the proof has total mass

$$\nu^{(x)}(S) = \sum_{y\in S}\sum_{i=0}^{\infty} \mathbb{P}_x\left[X_i = y, i < T_x\right] = \sum_{i=0}^{\infty} \mathbb{P}_x\left[i < T_x\right] = \mathbb{E}_x[T_x].$$

Two cases might occur:

- either the chain admit a stationary probability distribution μ. By uniqueness, all stationary measures are finite and proportional to it; in particular, for all x in S, we have $\nu^{(x)}(S) = \mathbb{E}_x[T_x] < +\infty$ and $\mu = \frac{1}{\mathbb{E}_x[T_x]}\nu^{(x)}$. In particular, evaluating at x, we have the **important relation**
$$\mu(x) = \big(\mathbb{E}_x[T_x]\big)^{-1}.$$
We call such chains or, equivalently, their states, <u>nonnull recurrent</u>.
- or the chain has no stationary probability distribution and hence no finite stationary measures. In particular, for all x in S, the measure $\nu^{(x)}$ is infinite, i.e. $\nu^{(x)}(S) = \mathbb{E}_x[T_x] = +\infty$. We call such chains or, equivalently, their states, <u>null recurrent</u>.

We generalize the definition beyond the *irreducible* case.

Definition 16.4. *Let Q be a transition matrix on S. Then a* recurrent *state x in S is*

- <u>*null recurrent*</u> *if $\mathbb{E}_x[T_x] = +\infty$;*
- <u>*nonnull recurrent*</u> *if $\mathbb{E}_x[T_x] < +\infty$.*

Lemma 16.5. *Let x, y be recurrent states with $x \leftrightarrow y$. Then x is null recurrent if and only if y is null recurrent.*

Proof. In the discussion above, we have seen that, for irreducible recurrent Markov chains, either all states are null recurrent or all are nonnull recurrent.

For the general case, we consider the decomposition theorem Theorem 14.10 and write

$$S = T \uplus C_1 \uplus C_2 \uplus \cdots ,$$

where T is the set of transient states and the C_i are closed irreducible sets of recurrent states. Then for each i, $Q/(C_i\times C_i)$ is a stochastic matrix (since C_i is closed). The corresponding Markov chain on C_i is irreducible and its states are recurrent. We also note that the return time in x starting from x for the original Markov chain or for the one restricted to C_i have the same distribution (since we never leave C_i when starting in C_i).

From the irreducible case, either all x in X_i fulfill $\mathbb{E}_x[T_x] = +\infty$ or all fulfill $\mathbb{E}_x[T_x] < +\infty$. In other words, either all x in C_i are all null recurrent or they are all nonnull recurrent. This holds for all closed irreducible components C_i, proving the lemma. □

Finite case: let Q be a transition matrix on a finite state space S. Then for each closed irreducible component C_i of recurrent states, the Markov chain $Q/(C_i \times C_i)$ has a stationary measure. But locally finite measures on a finite set are necessarily finite; so this stationary measure is finite and elements of C_i are nonnull recurrent.

We conclude that finite Markov chains have only transient or nonnull recurrent states. Recall in addition (Lemma 14.11) that such chains always have at least one recurrent state (which is thus nonnull recurrent). Consequently, all states of *irreducible* finite Markov chains are nonnull recurrent.

16.4. **Reversible Markov chains.** Finding stationary distributions (and checking that they are indeed stationary) is not always easy. We give here an easy sufficient condition (but not necessary, see Section 16.5.2).

Definition 16.6. *Let Q be a transition matrix on S. A nonzero locally finite measure μ on S is called* <u>*reversible*</u> *with respect to Q if, for all x, y in S, we have $\mu(y)\,Q_{y,x} = \mu(x)Q_{x,y}$.*

Proposition 16.7. *If μ is reversible with respect to Q, then μ is stationary.*

Proof. Let y be in S, we have

$$(\mu Q)_y = \sum_{x \in S} \mu(x) Q_{x,y} = \sum_{x \in S} \mu(y)\, Q_{y,x} = \mu(y) \sum_{x \in S} Q_{y,x} = \mu(y),$$

where the second equality use the reversibility of μ and the last one the fact that Q is a stochastic matrix. Since the above holds for all y in S, we have $\mu Q = \mu$, i.e. μ is stationary. □

16.5. **Examples.**

16.5.1. *I.i.d. r.v.* Let $(X_n)_{n \geq 0}$ be i.i.d. random variables with distribution. This is a Markov chain with transition matrix $Q_{x,y} = \mu(y)$. We recall that s in S is recurrent if and only if $\mu(s) > 0$. We assume $\mu(s) > 0$ for all s in S so that the chain is irreducible and consists only of recurrent states.

It is easy to check that μ is a stationary measure for Q. Since it is a probability distribution, the states are all nonnull recurrent. Furthermore, for all x in S, we have $\mathbb{E}_x[T_x] = 1/\mu(x)$, which can either be seen as a consequence of the general theory, or directly in this case, since T_x has the distribution of a geometric random variable of parameter $\mu(x)$.

16.5.2. *Random walk on $\mathbb{Z}^d$.* Let $(Y_i)_{i \leq 1}$ be i.i.d. random $\mathbb{Z}^d$ valued r.v. with distribution μ and $X_n = Y_1 + \dots + Y_n$. Recall that $(X_n)_{n \geq 0}$ is a Markov chain with transition matrix $Q_{x,y} = \mu(y - x)$. For simplicity, we assume the chain to be irreducible.

Consider the counting measure π on $\mathbb{Z}^d$, i.e. $\pi(x) = 1$ for all x in $\mathbb{Z}^d$. This is a stationary distribution: indeed, for any y in $\mathbb{Z}^d$, we have

$$\sum_{x \in \mathbb{Z}^d} \pi(x) Q_{x,y} = \sum_{x \in \mathbb{Z}^d} \mu(y - x) = \sum_{z \in \mathbb{Z}^d} \mu(z) = 1 = \pi(y),$$

where, in the second equality, we set $z = y - x$, and in the third, we use that μ is a probability distribution. We note that π is reversible if and only if μ is symmetric ($\mu(-z) = \mu(z)$ for all z in $\mathbb{Z}^d$), providing examples of stationary non-reversible measures (when μ is not symmetric). Since π is infinite, the states are either null recurrent or transient: in fact, both cases can append (recall Polya's recurrence theorem for $\mathbb{Z}^d$). In both cases, we have $\mathbb{E}_x[T_x] = +\infty$.

We now focus on the case $d = 1$ and $\mathbb{P}[Y_1 = 1] = 2/3 = 1 - \mathbb{P}[Y_1 = -1]$. Let ρ be the measure on $\mathbb{Z}^d$ giving weight $\rho(k) = 2^k$ to any integer k. We claim that ρ is reversible. Indeed, if $y = x + 1$

$$\rho(y) Q_{y,x} = 2^{x+1} 1/3 = 2^x 2/3 = \rho(x) Q_{x,y};$$

when $y = x - 1$, the equality $\rho(y) Q_{y,x} = \rho(x) Q_{x,y}$ also holds by exchanging x and y; finally, when $|y - x| \neq 1$, both sides equal 0.

Consequently, ρ is a stationary measure. We therefore have two non-proportional stationary measures π and ρ. We conclude that states are transient (which we had already proved with another argument in Section 15.3).

16.5.3. *Random walk on graphs.* Here, $G = (V, E)$ is a (locally finite) graph (without isolated vertices) and we consider the Markov chain $(X_n)_{n \geq 0}$ on $S = V$ with initial distribution δ_{x_0} (for some x_0 in V) and

$$Q_{x,y} = \tfrac{\mathbf{1}[\{x,y\} \in E]}{\deg(x)}.$$

Let μ be the measure on V giving weight $\mu(x) = \deg(x)$ to any x in V. Clearly $\mu(x) Q_{x,y} = \mu(y) Q_{y,x} = \mathbf{1}[\{x, y\} \in E]$, i.e. μ is reversible. Hence, it is stationary.

We now assume that G is connected, implying that the Markov chain is irreducible.

- if $|V| < +\infty$ (implying E finite, since the graph is locally finite), then $\mu(V) = \sum_{x \in V} \deg(x) = 2|E|$ is finite and the chain has a stationary probability distribution $\pi = \frac{1}{2|E|}\, \mu$. Hence all states are nonnull recurrent and we have $\mathbb{E}_x[T_x] = 1/\pi(x) = (2|E|)/\deg(x)$.
- If $|V| = +\infty$ (implying E infinite as well, since we have no isolated vertices), then $\mu(V) = \sum_{x \in V} \deg(x) = 2|E|$ is infinite. Hence states are null recurrent or transient (both cases indeed happen).

16.5.4. *A Markov chain without stationary measure.* Let $S = \{0, 1, 2, \cdots\}$ and fix a sequence $(p_i)_{i\geq 0}$ in $(0, 1)$ with $\prod_{i=0}^{\infty} p_i > 0$. We consider the transition matrix given by

$$Q_{i,j} = \begin{cases} p_i & \text{if } j = i+1; \\ 1 - p_i & \text{for } j = 0; \\ 0 & \text{otherwise.} \end{cases}$$

We prove by contradiction that there is no stationary measure. Let $(\pi(x))_{x\geq 0}$ be one. Then we have, for $j > 0$,

$$\pi(j) = \sum_i Q_{i,j}\pi(i) = p_{j-1}\pi(j-1).$$

This implies $\pi(j) = \pi(0)\prod_{i=1}^{j-1} p_i$. On the other hand,

$$\pi(0) = \sum_{i=0}^{\infty} Q_{i,0}\pi(i) = \sum_{i=0}^{\infty} \pi(i)(1 - p_i) = \sum_{i=0}^{\infty} (\pi(i) - \pi(i+1)) = \pi(0) - \lim_{i\to\infty} \pi(i).$$

But $\lim_{i\to\infty} \pi(i) = \pi(0)\prod_{i=0}^{\infty} p_i \neq 0$, which leads to a contradiction.

17. The limit theorem

In this section, we are interested in a limit in distribution for Markov chains. I.e. given a transition matrix Q and states x, y, does $Q^n_{x,y} = \mathbb{P}_x[X_n = y]$ converge?

17.1. **Aperiodicity.** We first see an example where it does not converge: take

$$S = \{1, 2, 3\}, \quad Q = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad x = y = 1.$$

In words: if the Markov chain is in state 1 at some time n, it choses uniformly at random between states 2 and 3 at time $n+1$; if it is in state 2 or 3, it goes back to 1. Starting at time 1, we will clearly come back to state 1 at even times, and be either in 2 or 3 at odd times. Therefore we have $Q^n_{1,1} = \mathbb{P}_1[X_n = 1] = \mathbf{1}[n \text{ is even}]$. In particular $Q^n_{1,1}$ does not have a limit.

Let us note that this Markov chain is irreducible. Since the state space is finite, states are nonnull recurrent, and there is a unique stationary probability distribution π, with $\pi(1) = 1/2$ and $\pi(2) = \pi(3) = 1/4$.

Conclusion: to get a limit theorem, we need some aperiodicity assumption.

Definition 17.1. *For x in S, we define the period of x as follows:*

$$P(x) = \gcd(A_x), \text{ where } A_x = \{n \geq 0 : Q^n_{x,x} > 0\}.$$

Note: $Q^n_{x,x} > 0$ means $\mathbb{P}_x[X_n = x] > 0$, i.e. it's possible to be back at state x in n steps. This is how we think at it to determine A_x in examples (we don't need to compute $Q^n_{x,x}$).

In the above example, we have $A_1 = A_2 = A_3 = 2\mathbb{N}_0$, giving $P(1) = P(2) = P(3) = 2$.

Definition 17.2. *A state x in S is aperiodic if $P(x) = 1$. A Markov chain (or equivalently, its transition matrix) is aperiodic if all its states are aperiodic.*

A sufficient condition for aperiodicity is that $Q_{x,x} > 0$. This is however not necessary, as shown in the following example. Take

$$Q = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

We have $1 \notin A_x$ (since $Q_{1,1} = 0$), but one easily checks that $\{2, 3, 4\} \subset A_x$, implying nevertheless $P(x) = 1$.

As for recurrence/transient, the following helps finding the periods of states.

Proposition 17.3. *If $x \leftrightarrow y$, then $P(x) = P(y)$.*

Proof. We have to show that $\gcd(A_x) = \gcd(A_y)$. This follows from the two following claims (using that A_x and A_y always contain 0):

i) if S is a subset of $\mathbb{Z}$ containing 0, then $\gcd(S) = \gcd(S - S)$, where $S - S$ denotes the set of differences among elements in S, i.e. $S - S = \{s_1 - s_2, \text{ for } s_1, s_2 \text{ in } S\}$.
ii) $A_x - A_x = A_y - A_y$

Proof of i). Since 0 is in S, we have $S \subseteq S - S$, implying that $\gcd(S - S)$ divides $\gcd(S)$. Conversely, $\gcd(S)$ divides any element of S, hence any element of $S - S$, and therefore it divides $\gcd(S - S)$

Proof of ii). By symmetry it is enough to prove $A_x - A_x \subseteq A_y - A_y$.

The assumption $x \leftrightarrow y$ tells us that there exists p and r such that $Q^p_{x,y} > 0$ and $Q^r_{y,x} > 0$. Then for any n in A_x, we have

$$Q^{r+n+p}_{y,y} \geq Q^r_{y,x} Q^n_{x,x} Q^p_{x,y} > 0, \tag{25}$$

proving that $r + n + p$ is in A_y.

Take a generic element $n_1 - n_2$ in $A_x - A_x$, where n_1 and n_2 are in A_x. From the above, we know that $r + n_1 + p$ and $r + n_2 + p$ is in A_y. Therefore $(r + n_1 + p) - (r + n_2 + p) = n_1 - n_2$ is $A_y - A_y$. This proves $A_x - A_x \subseteq A_y - A_y$, as wanted. □

We end this section with a useful lemma.

Lemma 17.4. *Let y be an aperiodic element in S and x be such that $x \longrightarrow y$. Then the set $A_{x,y} = \{n \geq 0 : Q^n_{x,y} > 0\}$ contains all integers bigger than some threshold value $n_0(x, y)$.*

Proof. We start with the case $x = y$; then $A_{x,y} = A_y$. Since y is aperiodic, the subgroup $A_y - A_y$ of $\mathbb{Z}$ is equal to $\mathbb{Z}$ itself. In particular there exists n_1 such that both n_1 and $n_1 + 1$ are in A_1. It is easy to see that every integer bigger than n_1^2 is a nonnegative integer combination of n_1 and $n_1 + 1$ and hence belong to A_1. This proves the case $x = y$.

In general, since $x \longrightarrow y$, we know that there exists p such that $Q^p_{x,y} > 0$. With the same argument as above, we have

$$A_{x,y} \subseteq p + A_{y,y}.$$

The lemma follows directly. □

17.2. **The main limit theorem.**

Theorem 17.5. *Let Q be an irreducible aperiodic transition matrix. Then for all states x,y in S, we have*

$$\lim_{n\to\infty} \mathbb{P}_x[X_n = y] = \frac{1}{\mathbb{E}_y[T_y]},$$

where as usual, $T_y = \min\{n > 0 : X_n = y\}$.

We note that the RHS is also (maybe better) described as:

- 0 in the transient and null recurrent cases;
- $\pi(y)$, where π is the unique stationary probability distribution, in the nonnull recurrent case.

An interesting remark is that the limit is independent from the starting point x. Informally, when run during a long time, an irreducible aperiodic chain forgets about its initial condition.

Irreducible nonnull recurrent aperiodic Markov chains are sometimes called ergodic.

Proof. We consider the three cases (transient, null recurrent and nonnull recurrent) separately.

Transient case. Already proved (see Theorem 14.4, item ii)).

Nonnull recurrent case We let as usual $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix Q and initial distribution δ_x. In addition, let $(X'_n)_{n\geq 0}$ be independent from X_n and be a Markov chain with transition matrix Q and stationary initial distribution: i.e. the distribution μ_0 of X'_0 is the stationary probability distribution π of Q (we know that it exists and is unique since Q is irreducible and nonnull recurrent). The distribution of X'_n is $\pi Q^n = \pi$, see Section 16.1, i.e. $\mathbb{P}[X'_n = y] = \pi(y)$.

The idea of the proof is to show that $\mathbb{P}[X_n = y]$ is close to $\mathbb{P}[X'_n = y]$.

For this we consider the pair $Z_n = (X_n, X'_n)$. It is a Markov chain on $S \times S$ with initial distribution $\delta_x \otimes \pi$ ($\otimes$ is the direct product of measure) and transition matrix R defined by

$$R_{(x,x'),(y,y')} = Q_{x,y} Q_{x',y'}.$$

We claim that R is irreducible. To prove this, take (x, x') and (y, y') in $S \times S$, we want to prove that there exists $n \geq 0$ s.t. $(R^n)_{(x,x'),(y,y')} > 0$. Using the definition of product of matrices, we see easily that $(R^n)_{(x,x'),(y,y')} = (Q^n)_{x,y}(Q^n)_{x',y'}$. We consider

$$A_{x,y} = \{n \geq 0 : Q^n_{x,y} > 0\}$$

and want to show that $A_{x,y} \cap A_{x',y'} \neq$. From Lemma 17.4, $A_{x,y}$ and $A_{x',y'}$ contains all integers bigger than some $n_0(x, y)$ and $n_0(x', y')$, respectively. In particular, $A_{x,y} \cap A_{x',y'}$ contains all integers bigger than $\max\big(n_0(x, y), n_0(x', y')\big)$ and hence is nonempty. This concludes the proof of the claim, i.e. of the irreducibility of R.

A direct computation shows that $\pi \otimes \pi$ is a stationary probability measure for R. This implies that the chain Z_n is (nonnull) recurrent.

We fix some $x_\bullet$ in S and consider the state $(x_\bullet, x_\bullet)$ in $S \times S$. Let $T_{(x_\bullet,x_\bullet)} = \inf\{n \geq 0 | Z_n = (x_\bullet, x_\bullet)\}$ be the first passage time in $(x_\bullet, x_\bullet)$. By Proposition 15.2, we have $T_{(x_\bullet,x_\bullet)} < +\infty$ (Proposition 15.2 in fact works with a Dirac initial distribution, but this generalizes to any distribution by conditioning on the value of X_0).

For $p < n$, let us consider $\mathbb{P}\big[X_n = y \,|\, T_{(x_\bullet,x_\bullet)} = p\big]$. The event $T_{(x_\bullet,x_\bullet)} = p$ contains complete information about X_p, namely $X_p = x_\bullet$. It contains also information on $(X_i)_{i<p}$ and Y. But X_n is independent from $(X'_i)_{i\geq 0}$, and, knowing X_p, it is also independent from $(X_i)_{i<p}$. Therefore knowing X_p, it is independent from $\{(X'_i)_{i\geq 0}, (X_i)_{i<p}\}$. We therefore have

$$\mathbb{P}\big[X_n = y \,|\, T_{(x_\bullet,x_\bullet)} = p\big] = \mathbb{P}\big[X_n = y \,| X_p = x_\bullet\big] = Q^{n-p}_{x_\bullet,y},$$

where the last equality is an application of Chapman-Kolmogorov equation. Similarly,

$$\mathbb{P}\big[X'_n = y \,|\, T_{(x_\bullet,x_\bullet)} = p\big] = \mathbb{P}\big[X'_n = y \,| Y_p = x_\bullet\big] = Q^{n-p}_{x_\bullet,y}.$$

We conclude that

$$\mathbb{P}\big[X_n = y,\, T_{(x_\bullet,x_\bullet)} = p\big] = \mathbb{P}\big[X'_n = y,\, T_{(x_\bullet,x_\bullet)} = p\big] = Q^{n-p}_{x_\bullet,y}\mathbb{P}[T_{(x_\bullet,x_\bullet)} = p].$$

Summing over $p > n$, we have

$$\mathbb{P}\big[X_n = y,\, T_{(x_\bullet,x_\bullet)} < n\big] = \mathbb{P}\big[X'_n = y,\, T_{(x_\bullet,x_\bullet)} < n\big]. \tag{26}$$

But

$$\Big|\mathbb{P}\big[X_n = y\big] - \mathbb{P}\big[X_n = y,\, T_{(x_\bullet,x_\bullet)} < n\big]\Big| = \Big|\mathbb{P}\big[X_n = y,\, T_{(x_\bullet,x_\bullet)} \geq n\big]\Big| \leq \Big|\mathbb{P}\big[T_{(x_\bullet,x_\bullet)} \geq n\big]\Big|.$$

As n tends to $+\infty$ the right-hand side tends to $\mathbb{P}[T_{(x_\bullet,x_\bullet)} = +\infty] = 0$. Similarly

$$\lim_{n\to\infty} \mathbb{P}\big[X'_n = y\big] - \mathbb{P}\big[X'_n = y,\, T_{(x_\bullet,x_\bullet)} < n\big] = 0.$$

Going back to eq. (26), we have

$$\lim_{n\to\infty} \mathbb{P}\big[X_n = y\big] - \mathbb{P}\big[X'_n = y\big] = 0.$$

We conclude the proof using that $\mathbb{P}\big[X'_n = y\big] = \pi(y)$, as observed at the beginning of the proof.

Null recurrent case. In this case, we use a similar strategy as before, comparing two Markov chains, both with transition matrix Q but different initial distribution. Since there is no stationary measure, we use initial distributions δ_{x_0} and $\delta_{x'_0}$, with $x \neq x'_0$. Formally we let X_n and X'_n be Markov chains on S with initial distributions δ_{x_0} and $\delta_{x'_0}$, respectively, independent from each other. Let $Z_n = (X_n, X'_n)$. As before Z_n is an irreducible Markov chain on $S \times S$ with transition matrix R.

Again, if π is a stationary measure for Q (since we assumed Q null recurrent, there exists such a measure, which is infinite), then $\pi \otimes \pi$ is stationary for R. But $\pi \otimes \pi$ is an infinite measure. This implies that R is either transient or null recurrent.

Subcase where R is transient: in this subcase, we know (Theorem 14.4, item ii)) that, for any (x, y) and (x', y') in $S \times S$, we have

$$\lim_{n\to\infty} R^n_{(x,x'),(y,y')} = 0.$$

Taking $x' = x$ and $y' = y$, we have $R^n_{(x,x'),(y,y')} = (Q^n_{x,y})^2$, so that we have

$$\lim_{n\to\infty} Q^n_{x,y} = 0,$$

which is what we wanted to prove.

Subcase where R is null recurrent: Copying the proof in the nonnull recurrent case, we get that, for all y,

$$\lim_{n\to\infty} \mathbb{P}\big[X_n = y\big] - \mathbb{P}\big[Y_n = y\big] = 0.$$

i.e. $\lim_{n\to\infty} Q^n_{x_0,y} - Q^n_{x'_0,y} = 0$.

We want to prove that for any x, y, we have $\lim_{n\to+\infty} Q^n_{x,y} = 0$. We proceed by contradiction. Assume that there exists $(x^\star, y^\star)$ such that $Q^n_{x^\star,y^\star}$ does not tend to 0. Since $Q^n_{x^\star,y^\star}$ takes value in $[0,1]$, this implies that there exists a subsequence $Q^{n_k}_{x^\star,y^\star}$ which converges to a nonzero value $\alpha_{x^\star,y^\star}$. Taking another pair $(x^\bullet, y^\bullet)$ in $S\times S$, since $Q^{n_k}_{x^\bullet,y^\bullet}$ takes value in $[0,1]$, we can further extract a subsequence $Q^{n_{k_j}}_{x^\bullet,y^\bullet}$, which converges to some value $\alpha_{x^\bullet,y^\bullet}$ (possibly zero). Using the diagonal extraction principle, we can find a subsequence, that we abusively denote n_k for simplicity, so that $Q^{n_k}_{x,y}$ converges to some value $\alpha_{x,y}$ for all (x,y) in $S\times S$ (since $S\times S$ is countable). Furthermore the vector $(\alpha_{x,y})_{x,y\in S}$ is not identically zero because $\alpha_{x^\star,y^\star} \neq 0$ (however, other coordinates can be zero).

Recall that $Q^n_{x_0,y} - Q^n_{x'_0,y}$ tends to 0 for any X_0, x'_0 and y. This implies $\alpha_{x_0,y} = \alpha_{x'_0,y}$, *i.e.* $\alpha_{x,y}$ is in fact independent of x; we denote it β_y. The vector $(\beta_y)_{y\in S}$ is not identically zero and can be seen as a (locally finite) measure on S.

We claim that $(\beta_y)_{y\in S}$ is a finite stationary measure for Q, leading to a contradiction, since Q was assumed to be null recurrent.

Proof of the claim.

Finiteness: We have, using the construction of β and Fatou's lemma,

$$\sum_{y\in S} \beta_y = \sum_{y\in S}\left(\lim_{k\to\infty} Q^{n_k}_{x,y}\right) \le \liminf_{k\to\infty}\left(\sum_{y\in S} Q^{n_k}_{x,y}\right).$$

But for any k, the sum $\sum_{y\in S} Q^{n_k}_{x,y}$ is identically 1 (Q^{n_k} is a stochastic matrix), so that the RHS is 1. We have $\sum_{y\in S}\beta_y \le 1$ proving that $(\beta_y)_{y\in S}$ is a finite measure.

Stationarity: Using again the construction of β and Fatou's lemma,

$$(\beta\cdot Q)_z = \sum_{y\in S}\beta_y Q_{y,z} = \sum_{y\in S}\left(\lim_{k\to\infty} Q^{n_k}_{x,y}\right) Q_{y,z} \le \liminf_{k\to\infty}\left(\sum_{y\in S} Q^{n_k}_{x,y} Q_{y,z}\right) = \liminf_{k\to\infty} Q^{n_k+1}_{x,z}$$

But we can also write

$$Q^{n_k+1}_{x,z} = \sum_{y\in S} Q_{x,y} Q^{n_k}_{y,z}.$$

We know that each summand $Q_{x,y}Q^{n_k}_{y,z}$ has a limit $Q_{x,y}\beta_z$ as n tends to $+\infty$. In addition $Q_{x,y}Q^{n_k}_{y,z}$ is bounded independently on k by $Q_{x,y}$, which is summable on y for x fixed ($\sum_{y\in S} Q_{x,y} = 1$ since Q is a stochastic matrix). We therefore have

$$\lim_{k\to\infty} Q^{n_k+1}_{x,z} = \sum_{y\in S} Q_{x,y}\beta_z = \beta_z\left(\sum_{y\in S} Q_{x,y}\right) = \beta_z.$$

We conclude that, for all z in S,

$$(\beta\cdot Q)_z \le \beta_z \tag{27}$$

But, on the other hand, using that Q is a stochastic matrix, we have

$$\sum_{z\in S}(\beta\cdot Q)_z = \sum_{y,z\in S}\beta_y Q_{y,z} = \sum_{y\in S}\beta_y\left(\sum_{z\in S} Q_{y,z}\right) = \sum_{y\in S}\beta_y,$$

where the exchange of infinite sums is justified by the nonnegativity of the summands. This implies that the inequalities in (27) are equalities, i.e. β is stationary. □

Remark. In the nonnull-recurrent case, with a small easy change at the end of the proof, we can show that, for any x in S,

$$\lim_{n\to+\infty}\sum_{y\in S}|\mathbb{P}_x[X_n=y]-\pi(y)|=0.$$

(This is called convergence in total variation distance.)

17.3. **Periodic chains.** Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain, and let d be the common period to all states (all states have the same period from Proposition 17.3). If $d>1$, the above theorem does not apply. However, we can consider $Y_n := X_{dn}$, which is a Markov chain of period 1. Note however that Y_n is not necessarily irreducible (see next subsection for a limit theorem for possibly reducible Markov chains).

We also have the following theorem for Cesaró's means

Theorem 17.6. *Let Q be an irreducible transition matrix on S. Then, for all x,y in S, we have*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n}Q^k_{x,y}=\frac{1}{\mathbb{E}_y[T_y]}.$$

Note that there is no aperiodicity condition here. In the aperiodic case ($d=1$), this is a consequence of Theorem 17.5 (indeed, if a sequence converges, its Cesaro's means converge to the same limit).

We skip the proof in these notes, see exercises.

17.4. **Reducible chains.**

Proposition 17.7. *Let Q be a transition matrix on S. Let X,y be in S. If y is aperiodic, then*

$$\lim_{n\to\infty}Q^n_{x,y}=\frac{F_{x,y}}{\mathbb{E}_y[T_y]},$$

where we recall that $F_{x,y}$ is the probability to reach y (at some point) starting at x.

Proof. If y is transient, then $Q^n_{x,y}$ tends to 0 (Theorem 14.4, item ii)) and there is nothing to prove since $\mathbb{E}_y[T_y]=+\infty$.

So let us assume w.l.o.g. y recurrent. We have

$$Q^n_{x,y}=\mathbb{P}_x[X_n=y]=\mathbb{P}_x[X_n=y,T_y=+\infty]+\sum_{k=1}^{\infty}\mathbb{P}_x[X_n=y,T_y=k].$$

If $T_y>n$ (including $T_y=+\infty$), we necessarily have $X_n\neq y$. Therefore the first summand and all summands in the sum corresponding to $k>n$ are 0.

On the other hand, for $k\leq n$, we have

$$\mathbb{P}_x[X_n=y,T_y=k]=\mathbb{P}_x[T_y=k]\mathbb{P}_x\big[X_n=y\,|\,T_y=k\big]=\mathbb{P}_x[T_y=k]\mathbb{P}_x\big[X_n=y\,|X_k=y\big]=\mathbb{P}_x[T_y=k]Q^{n-k}_{y,y},$$

where the middle equality uses the Markov property and that $T_y=k$ means $X_k=y,X_1\neq k,\dots,X_{k-1}\neq k$. To sum up, we have

$$Q^n_{x,y}=\sum_{k=1}^{\infty}\mathbb{P}_x[T_y=k](\mathbf{1}[n\geq k]Q^{n-k}_{y,y}) \tag{28}$$

We claim that $\lim_{n\to\infty}Q^n_{y,y}=\frac{1}{\mathbb{E}_y[T_y]}$. Indeed, let C be the closed irreducible component of recurrent states containing y (see Theorem 14.10). Then $R:=Q/(C\times C)$ is a stochastic matrix (since C is closed), and it is irreducible.

We easily check that $R^n_{y,y}=Q^n_{y,y}$. In particular, y is aperiodic for R. The first return time T_y to y under $\mathbb{P}_y$ has the same distribution for the original transition matrix Q, and for the restricted one R (since in the original chain, the chain always stays in C).

Therefore applying Theorem 17.5, we get

$$\lim_{n\to\infty}Q^n_{y,y}=\frac{1}{\mathbb{E}_y[T_y]}.$$

This implies that for any fixed k, the quantity $Q^{n-k}_{y,y}$ (and hence $\mathbf{1}[n \ge k]Q^{n-k}_{y,y}$) tends to $\frac{1}{\mathbb{E}_y[T_y]}$ as n tends to $+\infty$. Using (28) the dominated convergence theorem, we have

$$\lim_{n\to\infty} Q^n_{x,y} = \sum_{k=1}^{\infty} \mathbb{P}_x[T_y = k] \left(\lim_{n\to\infty} \mathbf{1}[n \ge k] Q^{n-k}_{y,y} \right) = \sum_{k=1}^{\infty} \mathbb{P}_x[T_y = k] \frac{1}{\mathbb{E}_y[T_y]}.$$

We can indeed exchange the infinite sum and the limit since $\mathbb{P}_x[T_y = k]\mathbf{1}[n \ge k]Q^{n-k}_{y,y}$ is bounded by $\mathbb{P}_x[T_y = k]$ independently of n and the sum $\sum_{k=1}^{\infty} \mathbb{P}_x[T_y = k] = \mathbb{P}[T_y < +\infty] \le 1$ is finite. The RHS of the above display can be simplified and we get

$$\lim_{n\to\infty} Q^n_{x,y} = \frac{1}{\mathbb{E}_y[T_y]} \left(\sum_{k=1}^{\infty} \mathbb{P}_x[T_y = k] \right) = \frac{1}{\mathbb{E}_y[T_y]} \mathbb{P}_x[T_y < +\infty] = \frac{F_{x,y}}{\mathbb{E}_y[T_y]},$$

concluding the proof. □

18. Finite Markov chains and Perron-Frobenius theorem

We focus here on the case $|S| < \infty$; the transition matrix Q is then a standard square matrix and we can use linear algebra to study Markov chains.

Theorem 18.1 (Special case of Perron-Frobenius theorem)**.** *Let Q be an aperiodic irreducible transition matrix, then 1 is a simple eigenvalue and all other eigenvalues have modulus smaller than 1.*

Here, *simple eigenvalue* means that 1 is a *simple* root of the characteristic polynomial of Q (this implies that, but is not equivalent to, $\dim\big(\mathrm{Ker}(Q - \mathrm{Id})\big) = 1$).

Proof. Since Q row sums equal to 1, we have $Q\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the column vector with all entries equal to 1. Hence 1 is a eigenvalue of Q.

Let λ be a possibly complex eigenvalue of Q and u be the associated eigenvector. We consider the maximal coordinate of u (in absolute value) that is x_0 in S such that

$$|u_{x_0}| = \max_{x \in S} |u_x|.$$

We have, for any n,

$$|\lambda^n u_{x_0}| = |(Q^n u)_{x_0}| = \left| \sum_{y\in S} Q^n_{x_0,y} u_y \right| \le \sum_{y\in S} Q^n_{x_0,y} |u_y| \le \sum_{y\in S} Q^n_{x_0,y} |u_{x_0}| = |u_{x_0}|,$$

where we used that Q^n is a stochastic matrix in the last equality. We conclude that $|\lambda| \le 1$.

Moreover, to have equality, we need that

a. all $Q^n_{x_0,y} u_y$ have the same complex argument;
b. for all y, $Q^n_{x_0,y}|u_y| = Q^n_{x_0,y}|u_{x_0}|$.

The second inequality means that $|u_y| = |u_{x_0}|$ as soon as $Q^n_{x_0,y} \ne 0$ for some n. By irreducibility, such an n exists for any y, meaning that $|u_y| = |u_{x_0}|$ for all y in S. We recall that aperiodicity implies that for a given y, $Q^n_{x_0,y} \ne 0$ for n large enough (Lemma 17.4). We can choose n such that $Q^n_{x_0,y} \ne 0$ for all y. Then, by *a.*, all u_y's have the same argument as u_{x_0} as soon as $Q_{x_0,y} \ne 0$ (or $Q^n_{x_0,y} \ne 0$ for some n). Hence $u = C\mathbf{1}$, for some $C \ne 0$, showing that 1 is the only eigenvalue of modulus 1 and that $\dim\big(\mathrm{Ker}(Q - \mathrm{Id})\big) = 1$.

We still have to prove that the block corresponding to 1 in the Jordan reduction of Q has size 1. Otherwise, there would be a vector $u^{(2)}$ such that $Q\,u^{(2)} = u^{(2)} + \mathbf{1}$. This implies $Q^n u^{(2)} = u^{(2)} + n\,\mathbf{1}$. This is impossible since Q^n is stochastic and has therefore entries in $[0, 1]$. □

Consequences.

Q has a unique *left-eigenvector* corresponding to $\lambda = 1$. Such a vector is a stationary measure (π with $\pi = \pi Q$), so we recover existence and uniqueness of the stationary measure (irreducible finite Markov chains are always nonnull recurrent).

Furthermore, let λ_1, λ_2, ...be the eigenvalues of Q (with multiplicities) ordered in such a way that

$$1 = |\lambda_1| > |\lambda_2| \ge |\lambda_3| \ge \cdots$$

Assume for simplicity that Q is diagonalizable. We consider the associated left eigenvectors $\tau^{(1)} = \pi$ (the stationary measure), $\tau^{(2)}$,...Fix an initial distribution μ_0 and decompose it on this basis

$$\mu_0 = \sum_{i=1}^{|S|} \alpha_i \tau^{(i)}.$$

Multiplying by Q^n, we get

$$\mu_n = \mu_0 Q^n = \sum_{i=1}^{|S|} \alpha_i \lambda_i^n \tau^{(i)} = \alpha_1 \pi + O(\lambda_2^n).$$

Since both μ_n and π have sum 1, this forces $\alpha_1 = 1$. We therefore get that for any initial distribution μ_0 and y,

$$\mathbb{P}_{\mu_0}[X_n = y] = \pi(y) + O(\lambda_2^n).$$

This recovers the limit theorem in the finite case with a strong error bound: the error is exponentially small and its rate is the second eigenvalue of Q ($1 - \lambda_2$ is sometimes called the spectral gap of Q).

Part D. **An introduction to Brownian motion**

19. Basics

19.1. **Definition.** *Motivation:* Brownian motion $(B_t)_{t\in\mathbb{R}_+}$ models the trajectory of a particle in a very rough unpredictable environment.

(1) Since it is a trajectory, $t \mapsto B_t$ should be a *continuous* function from $\mathbb{R}_+ \to \mathbb{R}^d$;
(2) it should be a random function (since it is unpredictable);
(3) the rough environment makes the particle change speed/direction at all times.

To simplify, we will assume that the direction/speed taken at time t is independent from the past and from the position.

It is thus a continuous analogue from a random walk. In particular, Brownian motion is an example of "continuous Markov chain" (continuous both in time and space).

Mathematical formalization. We let $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ be the space of continuous function from $\mathbb{R}_+ \to \mathbb{R}^d$, equipped with the topology of uniform convergence on compact sets. We can consider the Borel σ-algebra associated to this topology, making $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ a measured space.

Definition 19.1. *A d-dimensional Brownian motion is a $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$-valued random variable $t \mapsto B_t$ such that*

(P): *$B_0 = 0$ a.s., and for each $p \geq 0$ and each $t_0 < t_1 < \cdots < t_p$, the random variables $(B_{t_j} - B_{t_{j-1}})_{1\leq j\leq p}$ are independent and are distributed as Gaussian vectors of covariance matrices $(t_j - t_{j-1})\,\mathrm{Id}_d$ (for $1 \leq j \leq p$).*

Theorem 19.2. *Fix $d \geq 1$. There exists a d-dimensional Brownian motion on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Moreover, the distribution of (B_t) is unique.*

The distribution of B_t is called the (d-dimensional) Wiener measure. It is a measure on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$

Proof. Skipped; see, e.g. [LG06, Section 14.1]. □

Here are some examples of

- one-dimensional Brownian motion: link. Here the horizontal axis is the time t and the vertical axis is the value of B_t.
- two-dimensional Brownian motion: link. Here t does not appear, we see, for each t, B_t plotted in the plane (B_t lives in $\mathbb{R}^2$).

19.2. **Invariance properties.**

Proposition 19.3. *Let B_t be a (d-dimensional) Brownian motion. Then*

(i) $-B_t$ *is also a Brownian motion. More generally, for any isometry φ of $\mathbb{R}^d$ fixing 0, $\varphi(B_t)$ is also a Brownian motion.*
(ii) *(self-similarity property): for $\gamma > 0$, $B_t^\gamma := \frac{1}{\gamma} B_{\gamma^2 t}$ defines a Brownian motion.*
(iii) *(simple Markov property): for $s > 0$, $B_t^{(s)} := B_{s+t} - B_s$ defines a Brownian motion independent from $(B_u)_{u \le s}$.*
(iv) *The coordinates of a d-dimensional Brownian motion are* independent *1-dimensional Brownian motion.*

Proof. We'll prove (ii), other items are similar. To simplify notation, we suppose $d = 1$.

It is clear that $t \mapsto \frac{1}{\gamma} B_{\gamma^2 t}$ is a continuous function a.s. Let $0 \le t_0 < \cdots < t_p$. For $j \in \{1, \dots, p\}$,

$$B_{t_j}^\gamma - B_{t_{j-1}}^\gamma = \tfrac{1}{\gamma}\big(B_{\gamma^2 t_j} - B_{\gamma^2 t_{j-1}}\big) \sim \tfrac{1}{\gamma}\mathcal{N}(0, \gamma^2 t_j - \gamma^2 t_{j-1}) = N(0, t_j - t_{j-1}).$$

Moreover, when j runs over $\{1, \dots, p\}$, these random variables are independent. Therefore $t \mapsto B_t^\gamma$ satisfies the properties defining a Brownian motion. □

19.3. **Blumenthal $0-1$ law and irregularity.** Notation: $\mathcal{F}_s = \sigma(B_u; u \le s)$, $\mathcal{F}_{0+} = \bigcap_{\varepsilon > 0} \mathcal{F}_\varepsilon$.

$\mathcal{F}_{0+}$ represents events that can be determined looking at any neighbourhood of 0, e.g. "B_t is differentiable in 0".

Proposition 19.4 (Blumenthal $0-1$ law)**.** *Let $A \in \mathcal{F}_{0+}$. Then $\mathbb{P}[A] \in \{0, 1\}$.*

Proof. We fix $\varepsilon > 0$ and $t > \varepsilon$. Then $B_t - B_\varepsilon$ is independent from $\mathcal{F}_\varepsilon$ (simple Markov property) and therefore from $\mathcal{F}_{0+}$. Making ε tends to 0, we get that B_t is independent from $\mathcal{F}_{0+}$. Since this is true for any t, $\mathcal{F}_{0+}$ is independent from $\mathcal{F}_1$. But $A \in \mathcal{F}_{0+} \subseteq \mathcal{F}_1$, so A is independent from itself, i.e. $\mathbb{P}[A] \in \{0, 1\}$. □

Corollary 19.5. *Let $(B_t)_{t \ge 0}$ be a 1-dimensional Brownian motion. A.s. we have*

$$\forall \varepsilon > 0, \inf_{s \le \varepsilon} B_s < 0 \text{ and } \sup_{s \le \varepsilon} B_s > 0. \tag{29}$$

Informally, this says that a Brownian motion crosses infinitely many times the x-axis in any neighborhood of 0.

Proof. Call $A = \bigcap_{p \ge 1} \{\inf_{s \le 1/p} B_s < 0\}$. For any p_0, the event is the same if we restrict the intersection to $p \ge p_0$. Therefore $A \in \mathcal{F}_{1/p_0}$. Since this holds for any p_0, we have that A is in $\mathcal{F}_{0+}$. In particular $\mathbb{P}[A]$ is in $\{0, 1\}$.

Since A is a countable intersection of decreasing events, we have

$$\mathbb{P}[A] = \lim_{n \to \infty} \mathbb{P}[\{\inf_{s \le 1/p} B_s < 0\} \ge \limsup_{n \to \infty} \mathbb{P}[B_{1/p} < 0] = 1/2.$$

The middle inequality uses that $\inf_{s \le 1/p} B_s < 0$ certainly holds when $B_{1/p} < 0$. The last equality is that a Gaussian random variable is negative with probability 1/2 (regardless of its variance).

We conclude that $\mathbb{P}[A] = 1$. This implies the inf statement in the corollary. The sup statement follows by symmetry (recall that $-B_t$ is a Brownian motion). □

Combining this with the simple Markov property we get that for fixed $t > 0$,

$$\mathbb{P}\Big[\forall \varepsilon > 0 : \sup\{B_s, s \in (t, t+\varepsilon)\} > B_t\Big] = 1.$$

Taking the intersection over rational t, we see that, with probability 1, there is no interval on which (B_t) is increasing. The same statement holds with decreasing by symmetry.

Conclusion: a.s., there is no interval on which (B_t) is monotone.

19.4. **Asymptotic behaviour.**

Proposition 19.6. *Let (B_t) be a one dimensional Brownian motion. The following holds a.s.*

$$\limsup_{t\to+\infty} B_t = +\infty, \quad \liminf_{t\to+\infty} B_t = -\infty.$$

Proof. We first prove that, for any $A > 0$,

$$\mathbb{P}\Big[\sup_{t\geq 0} B_t \geq A\Big] = 1.$$

Using self-similarity (Proposition 19.3, item ii), for any $\delta > 0$, we have

$$\mathbb{P}\Big[\sup_{t\geq 0} B_t \geq A\Big] = \mathbb{P}\Big[\sup_{t\geq 0} \tfrac{\delta}{A} B_t \geq \delta\Big] = \mathbb{P}\Big[\sup_{t\geq 0} B_{\frac{\delta^2}{A^2}t} \geq \delta\Big] = \mathbb{P}\Big[\sup_{u\geq 0} B_u \geq \delta\Big].$$

We then a limit as δ tends to 0

$$\mathbb{P}\Big[\sup_{t\geq 0} B_t \geq A\Big] = \lim_{\delta\to 0} \mathbb{P}\Big[\sup_{u\geq 0} B_u \geq \delta\Big] = \mathbb{P}\Big[\sup_{u\geq 0} B_u \geq 0\Big] = 1,$$

where the last equality follows from Corollary 19.5. We conclude that for any $A > 0$,

$$\mathbb{P}\Big[\sup_{t\geq 0} B_t \geq A\Big] = 1.$$

Taking $A = n$ and building a countable intersection of such events, we have

$$\mathbb{P}\Bigg[\bigcap_{n\geq 1} \Big\{\sup_{t\geq 0} B_t \geq n\Big\}\Bigg] = 1,$$

i.e., $\limsup_{t\to+\infty} B_t = +\infty$ a.s. The $\liminf$ statement is proved with similar arguments (or follows by symmetry). ☐

20. Donsker's theorem

The goal here is to prove that rescaled random walks on $\mathbb{Z}$ (or more generally on $\mathbb{R}^d$) converge to the Brownian motion.

Let $(Y_i)_{i\geq 1}$ be a sequence of i.i.d. r.v. in $\mathbb{R}^d$ with mean 0 and covariance matrix Id_d. We set $X_n = Y_1 + \cdots + Y_n$. Furthermore, define $W^{(n)}$ to be the (random) continuous function on $[0,1]$ such that

1. if $t = i/n$ for some integer i in $\{0, 1, \ldots, n\}$, then $W^{(n)}(t) = \frac{1}{\sqrt{n}} X_{nt}$.
2. $W^{(n)}$ is linear on the interval $[i/n, (i+1)/n]$ for every i in $\{0, 1, \ldots, n-1\}$.

In the following, we work in the space $\mathcal{C}([0,1], \mathbb{R})$ of continuous functions on $[0,1]$, endowed with the supremum norm.

Theorem 20.1 (Donsker)**.** *With the above notation, $W^{(n)}$ converges in distribution to the d-dimensional Brownian motion $(B_t)_{0\leq t\leq 1}$ on $[0,1]$.*

Note: all simulations of the Brownian motion are in fact realizations of $W^{(n)}$ for large n.

Incomplete proof. To simplify the notation we assume $d = 1$. We will prove that for each $p \geq 0$ and each $t_1 < \cdots < t_p \leq 1$, we have the following convergence in distribution

$$(W^{(n)}(t_j))_{1\leq j\leq p} \longrightarrow (B_{t_j})_{1\leq j\leq p}. \tag{30}$$

This is called convergence of the finite dimensional distributions of $W^{(n)}$ to B_t. This is **not enough** to prove the convergence of $W^{(n)}$ to (B_t) in the space $\mathcal{C}([0,1], \mathbb{R})$. We need some additional argument (called *tightness*), which is outside the scope of this lecture.

Fix $p \geq 0$ and $t_1 < \cdots < t_p \leq 1$, and let us prove (30). It is equivalent to prove

$$\big(W^{(n)}(t_j) - W^{(n)}(t_{j-1})\big)_{1\leq j\leq p} \longrightarrow (B_{t_j} - B_{t_{j-1}})_{1\leq j\leq p}, \tag{31}$$

where we set $t_0 = 0$ (by construction, a.s., $W^{(n)}(0) = B_0 = 0$). To simplify again notation, we assume that $i_1 := nt_1, \cdots, i_p := nt_p$ are integers. Then

$$W^{(n)}(t_j) - W^{(n)}(t_{j-1}) = \frac{1}{\sqrt{n}} \sum_{i=i_{j-1}+1}^{i_j} X_i.$$

In particular, for different values of j, the variables $W^{(n)}(t_j) - W^{(n)}(t_{j-1})$ are independent since they are renormalized sums of disjoint sets of the independent variables $(X_i)_{i\geq 0}$. Moreover, using the central limit theorem,

$$\frac{1}{\sqrt{t_j - t_{j-1}}}\big(W^{(n)}(t_j) - W^{(n)}(t_{j-1})\big) = \frac{1}{\sqrt{i_j - i_{j-1}}} \sum_{i=i_{j-1}+1}^{i_j} X_i$$

converges in distribution towards a standard Gaussian random variable. Thus, $W^{(n)}(t_j) - W^{(n)}(t_{j-1})$ tends towards $\mathcal{N}(0, t_j - t_{j-1})$. Using the independence for different values of j, we deduce that

$$\big(W^{(n)}(t_j) - W^{(n)}(t_{j-1})\big)_{1\leq j\leq p}$$

tends towards a vector of independent centered Gaussian random variables of variance $t_j - t_{j-1}$ respectively. By definition of the Brownian motion (Definition 19.1), this is the distribution of $(B_{t_j} - B_{t_{j-1}})_{1\leq j\leq p}$. We therefore have proved (31), or equivalently, (30). □

21. Strong Markov property and reflection principle

We recall the simple Markov property (Proposition 19.3, item ii): for $s > 0$, $B_t^{(s)} := B_{s+t} - B_s$ defines a Brownian motion independent from $(B_u)_{u\leq s}$. We want to replace s by a random stopping time.

Definition 21.1. *A random variable $T : \Omega \to \mathbb{R}_+ \cup \{+\infty\}$ is a stopping time if, for all t in $\mathbb{R}_+$, the event $\{T \leq t\}$ is in $\mathcal{F}_t$. Its associated σ-algebra is*

$$\mathcal{F}_T = \big\{A \in \mathcal{F} :\ A \cap \{T \leq t\} \in \mathcal{F}_t\big\}.$$

Proposition 21.2 (Strong Markov property)**.** *Let (B_t) be a Brownian motion and T be a stopping time with $\mathbb{P}[T < +\infty] = 1$. Then $B_t^{(T)} := B_{T+t} - B_T$ is a Brownian motion independent from $\mathcal{F}_T$.*

Proof. Let (B'_u) be another Brownian motion. We need to prove that for integrable $X \in \mathcal{F}_T$, $p > 0$, $0 \leq t_0 < t_1 < \cdots < t_p$, we have

$$\big(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)}\big) \overset{\text{law}}{=} \big(B'_{t_1}, \cdots, B'_{t_p}\big)$$

and the LHS is independent from X. Since the distribution of a random variable Y is determined by $\mathbb{E}[F(Y)]$ for bounded continuous function F (Riesz theorem), we have to prove that

$$\mathbb{E}\Big[X\, F\big(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)}\big)\Big] = \mathbb{E}[X]\mathbb{E}\Big[F\big(B'_{t_1}, \cdots, B'_{t_p}\big)\Big].$$

Up to writing $X = X_1 - X_2$, $F = F_1 - F_2$, we can assume X and F nonnegative.

We would like to use the law of total probability summing over possible values of T, but it takes uncountably many values. The idea is therefore to use a discrete approximation and to write that, a.s.,

$$F\big(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)}\big) = \lim_{n\to+\infty} \sum_{k\geq 1} \mathbf{1}\big\{\tfrac{k-1}{n} \leq T < \tfrac{k}{n}\big\} F\big(B_{t_1+k/n} - B_{k/n}, \dots, B_{t_p+k/n} - B_{k/n}\big).$$

Since X is in $\mathcal{F}_T$, the random variable $X\,\mathbf{1}\big\{\frac{k-1}{n} \leq T < \frac{k}{n}\big\}$ belongs to $\mathcal{F}_{k/n}$. By the simple Markov property $(B_{t_1+k/n} - B_{k/n}, \dots, B_{t_p+k/n} - B_{k/n})$ has the same distribution as $(B'_{t_1}, \cdots, B'_{t_p})$ and is independent from $\mathcal{F}_{k/n}$. So, using the dominated convergence theorem and exchanging infinite sum and expectation of nonnegative terms,

$$\mathbb{E}\Big[X\, F\big(B_{t_1}^{(T)}, \cdots, B_{t_p}^{(T)}\big)\Big] = \lim_{n\to+\infty} \sum_{k\geq 1} \mathbb{E}\Big[X\,\mathbf{1}\big\{\tfrac{k-1}{n} \leq T < \tfrac{k}{n}\big\}\Big]\mathbb{E}\Big[F(B'_{t_1}, \cdots, B'_{t_p})\Big].$$

But the second expectation can be factorized out of the sum and, clearly we have

$$\sum_{k\geq 1} \mathbb{E}\Big[X\,\mathbf{1}\big\{\tfrac{k-1}{n} \leq T < \tfrac{k}{n}\big\}\Big] = \mathbb{E}[X].$$

This concludes the proof. □

Comment. In the strong Markov property for Markov chains, we needed to assume $X_T = x$ a.s.; For Brownian motion, we do not need this because of the homogeneity in space; the distribution of $B_{T+t} - B_T$ is independent of B_T.

As a first application to the strong Markov property, we can compute the distribution of the supremum of a one-dimensional Brownian motion until time t. This uses a trick called reflection principle: we can reflect the Brownian motion after a stopping time, without changing its distribution. We note that the reflection principle can also be used in the discrete world to count families of lattice paths not crossing a given line, such as Dyck paths; see, e.g., this wikipedia page.

Proposition 21.3. *Let $(B_s)_{s\geq 1}$ be a 1-dimensional Brownian motion. We set, for $t > 0$, $S_t = \sup_{s\leq t} B_s$. Then for $a \geq 0$ and $b \leq a$, we have*

$$\mathbb{P}\big[S_t \geq a, B_t \leq b\big] = \mathbb{P}\big[B_t \geq 2a - b\big].$$

Consequently S_t has the same distribution as $|Z|$, where $Z \sim \mathcal{N}(0,t)$.

Proof. Consider the stopping time $T_a = \inf\{t \geq 0, B_t = a\}$. We know (using Proposition 19.6 and the continuity of $t \mapsto B_t$) that $T_a < +\infty$ a.s. We also observe that $S_t \geq a$ if and only if $T_a \leq t$. Therefore,

$$\mathbb{P}\big[S_t \geq a, B_t \leq b\big] = \mathbb{P}\big[T_a \leq t, B_t \leq b\big] = \mathbb{P}[T_a \leq t]\,\mathbb{P}\big[B^{(T_a)}_{t-T_a} \leq b - a \,\big|\, T_a \leq t\big],$$

where, using the same notation as above, $B^{(T_a)}_s = B_{s+T_a} - B_{T_a} = B_{s+T_a} - a$. But the strong Markov property (Proposition 21.2) asserts that $B^{(T_a)}_s$ is distributed as a Brownian motion B'_s and is independent from $\mathcal{F}_{T_a}$, hence in particular from T_a. We get

$$\mathbb{P}\big[S_t \geq a, B_t \leq b\big] = \mathbb{P}[T_a \leq t]\,\mathbb{P}\big[B'_{t-T_a} \leq b - a\big].$$

Here comes the reflection trick: we now use the symmetry $\mathrm{Law}(B'_s) = \mathrm{Law}(-B'_s)$, which implies

$$\mathbb{P}\big[B'_{t-T_a} \leq b - a\big] = \mathbb{P}\big[B'_{t-T_a} \geq a - b\big].$$

Making the same reasoning as above in the other direction we have

$$\begin{aligned}\mathbb{P}\big[S_t \geq a, B_t \leq b\big] = \mathbb{P}[T_a \leq t]\,\mathbb{P}\big[B'_{t-T_a} \geq a - b\big] = \mathbb{P}[T_a \leq t]\,\mathbb{P}\big[B^{(T_a)}_{t-T_a} \geq a - b \,\big|\, T_a \leq t\big]\\ = \mathbb{P}\big[T_a \leq t, B_t \geq 2a - b\big] = \mathbb{P}\big[B_t \geq 2a - b\big],\end{aligned}$$

where in the last equality we used that $B_t \geq 2a - b$ implies $T_a \leq t$ (recall that $a \geq b$ and hence $2a - b \geq a$). This proves the first part of the proposition.

For the second part, we write (noting that $B_t = a$ is a zero probability event)

$$\mathbb{P}\big[S_t \geq a\big] = \mathbb{P}\big[S_t \geq a, B_t \geq a\big] + \mathbb{P}\big[S_t \geq a, B_t \leq a\big].$$

In the first term, $S_t \geq a$ is superfluous ($B_t > a$ implies $S_t \geq a$), so that the first term simplifies to $\mathbb{P}\big[B_t \geq a\big]$. For the second term, we apply the formula above and get again $\mathbb{P}\big[B_t \geq a\big]$. Finally,

$$\mathbb{P}[S_t \geq a\big] = 2\mathbb{P}\big[B_t \geq a\big] = \mathbb{P}\big[|B_t| \geq a\big].$$ □

22. Harmonic functions, recurrence and transient

Question: does the Brownian motion come back near 0 with probability 1?

- In dimension 1, yes (as a consequence of Proposition 19.6);
- we will see that it does in dimension 2 (we say that Brownian motion is *recurrent*), but not in dimension 3 and higher (we will see that Brownian motion is *transient*).

Comment. This result is a continuous analogue of Pólya's recurrence theorem (Theorem 14.12). Interestingly, his proof involves partial difference equations.

Throughout this section, we admit some proofs, in particular of statements of analytic natures. Complete proofs can be found e.g. in [LG06], where the presentation is closed to the one used here.

22.1. **Dirichlet problem: using Brownian motion to solve a partial differential equation.** Some notation: we write $D(x,r)$ (resp. $\overline{D}(x,r)$ and $\partial D(x,r)$) for the open ball (resp. closed ball and circle) of center x and radius r.

A *domain* Ω is an open connected subset of $\mathbb{R}^d$

Definition 22.1. *Let Ω be a domain in $\mathbb{R}^d$. A function $h : \Omega \to \mathbb{R}$ is called* harmonic *if, for all x in D and $r > 0$ such that $\overline{D}(x,r) \subset \Omega$, we have*

$$h(x) = \int_{\overline{D}(x,r)} h(y)\sigma_{x,r}(dy),$$

where $\sigma_{x,r}$ is the unique probability measure[8] *on $\partial D(x,r)$ invariant by isometries of $\mathbb{R}^d$ fixing x.*

In dimension 2, identifying $\mathbb{R}^2$ and $\mathbb{C}$, the RHS above rewrites as

$$\int_{\overline{D}(x,r)} h(y)\sigma_{x,r}(dy) = \frac{1}{2\pi}\int_0^{2\pi} h(x + r\exp(i\theta))d\theta.$$

Proposition 22.2. *h is harmonic on Ω if and only if h is of class C^∞ and*

$$\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} h = 0.$$

Proof. Admitted. □

Note: the differential operator $\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ is usually denoted Δ and called *Laplacian* operator on Ω.

A standard problem in partial differential equations, called *Dirichlet problem* is the following: given a bounded domain Ω and a continuous function $g : \partial\Omega \to \mathbb{R}$, find a function $h : \overline{\Omega} \to \mathbb{R}$ such that

- h is harmonic on Ω;
- h is continuous on $\overline{\Omega}$;
- $h \equiv g$ on $\partial\Omega$.

Lemma 22.3. *Fix Ω and g as above. If there is a solution to Dirichlet's problem, then it is unique.*

Brownian motion will help us construct a solution, in some cases. For x in $\mathbb{R}^d$, we denote $\widetilde{B}_t := x + B_t$ be a "Brownian motion starting at x". The starting point x does not appear in the notation $\widetilde{B}_t$, but we will use $\mathbb{P}_x$ and $\mathbb{E}_x$ to indicate it (as for Markov chains).

Theorem 22.4. *Fix Ω and g as above. For x in Ω, we consider, under $\mathbb{P}_x$, the stopping time $T = \inf\{t > 0 : \widetilde{B}_t \notin \Omega\}$ and we define*

$$h(x) = \mathbb{E}_x\big[g(\widetilde{B}_T)\big].$$

Then h is harmonic on Ω

Before proving the theorem, let us justify that h is well-defined: for B_T to make sense, we need $T < \infty$, and for $g(\widetilde{B}_T)$ to make sense, we need $\widetilde{B}_T$ to be in $\partial\Omega$. But, Ω is a bounded domain, while $\widetilde{B}_t$ is unbounded with probability 1 (its projection are 1-dimension Brownian motion, which are a.s. unbounded by Proposition 19.6). Hence $\widetilde{B}_t$ eventually leaves Ω with probability 1, i.e. $T < \infty$ a.s. Furthermore, when $T < \infty$, $\widetilde{B}_T$ is in $\overline{\Omega}$: indeed, it is the limit of $(\widetilde{B}_t; t \to T, t < T)$, which takes value in Ω by minimality of T. On the other hand there is a sequence t_n tending to T such that B_t is not in Ω (possibly $t_n = T$). Since the complement of Ω is closed, this implies that $B_T = \lim B_{t_n}$ is not in Ω. We conclude that $\widetilde{B}_T$ is in $\partial\Omega$, so that $g(\widetilde{B}_T)$ is well-defined as soon as $T < \infty$, which happens a.s. Finally since g is a continuous function on the bounded closed set $\partial\Omega$, g is necessarily bounded (we are in $\mathbb{R}^d$, bounded closed sets are compact). Therefore $g(\widetilde{B}_T)$ is bounded, and hence integrable; its expectation $h(x)$ is well-defined.

[8] We admit existence and uniqueness of this measure

Proof. Take x in Ω and $r>0$ such that $\overline{D}(x,r)\subset\Omega$. We need to prove that

$$h(x)=\int_{\overline{D}(x,r)} h(y)\sigma_{x,r}(dy). \tag{32}$$

As above, denote by $\widetilde{B}_t$ a Brownian motion starting at x. We define $S=\inf\{t>0:\widetilde{B}_t\notin D(x,r)\}$. This is a stopping time and we have $S<T<\infty$ a.s.

We claim that $\widetilde{B}_S$ belong to $\partial D(x,r)$ a.s. and has distribution $\sigma_{x,r}$. The proof that is belongs to $\partial D(x,r)$ is similar to the proof that $\widetilde{B}_T$ belongs to $\partial\Omega$. If φ is a isometry of $\mathbb{R}^d$ fixing x, $\varphi(\widetilde{B}_t)$ has the same distribution as $\widetilde{B}_t$. We conclude that $\varphi(\widetilde{B}_S)$ has the same distribution as $\widetilde{B}_S$ (since applying φ does not change the stopping time S). But $\sigma_{x,r}$ is the unique probability distribution on $\partial D(x,r)$ which is invariant by all isometries fixing x. We conclude that the distribution of $\widetilde{B}_S$ is $\sigma_{x,r}$, as claimed.

We know compute $h(x)=\mathbb{E}_x\big[g(\widetilde{B}_T)\big]$ by conditioning w.r.t. $\widetilde{B}_S$, namely we write

$$h(x)=\mathbb{E}\Big[\mathbb{E}_x\big[g(\widetilde{B}_T)|\widetilde{B}_S\big]\Big].$$

By the strong Markov property, conditionally on $\widetilde{B}_S$ the process $(B_{S+t}-B_S)_{t\ge 0}$ is a Brownian motion, and hence $\hat{B}_t=\widetilde{B}_{S+t}$ is a Brownian motion starting at $\widetilde{B}_S$. This Brownian motion leaves Ω at time $\hat{T}=T-S$, taking value $\hat{B}_{\hat{T}}=\widetilde{B}_T$ at that time. Hence

$$\mathbb{E}_x\big[g(\widetilde{B}_T)|\widetilde{B}_S\big]=\mathbb{E}_{B_S}\big[g(\hat{B}_{\hat{T}})\big]=h(B_S).$$

We conclude that $h(x)=\mathbb{E}\big[h(B_S)\big]$, proving (32) since $\widetilde{B}_S$ has distribution $\sigma_{x,r}$. □

We can extend the above function h by setting $h(y)=g(y)$ for y in $\partial\Omega$. *If* the resulting h is continuous on $\overline{\Omega}$, then it is a solution of Dirichlet's problem. The continuity holds under reasonable asumptions on Ω (but not in general), e.g. when Ω contains an open cone in the neighbourhood of each of its boundary point (we admit this result here).

22.2. **Using harmonic function to analyse the recurrence of Brownian motion.**

Proposition 22.5. *Let x in $\mathbb{R}^d\setminus\{0\}$ and $(B_t)_{t\ge0}$ a BM in $\mathbb{R}^d$. Set $\widetilde{B}_t=x+B_t$ as above and fix ε,R such that $0<\varepsilon<\|x\|<R$. Then*

$$\mathbb{P}_x\left[\begin{array}{c}\widetilde{B}_t \text{ visits } \partial D(0,\varepsilon)\\ \text{before } \partial D(0,R)\end{array}\right]=\begin{cases}\frac{\log(R)-\log\|x\|}{\log(R)-\log(\varepsilon)} & \text{if } d=2;\\ \frac{\|x\|^{2-d}-R^{2-d}}{\varepsilon^{2-d}-R^{2-d}} & \text{if } d\neq 2.\end{cases} \tag{33}$$

Proof. We set $\Omega=\{y\in\mathbb{R}^d:\varepsilon<\|y\|<R\}$ and T be the first exit time from Ω of the Brownian motion $\widetilde{B}_t$ starting at x. The probability in the LHS above can be rewritten as $\mathbb{E}_x\big[g(\widetilde{B}_T)\big]$, where $g(y)=\mathbf{1}\{\|y\|=\varepsilon\}$ (i.e. $g\equiv1$ on $\partial D(0,\varepsilon)$ and 0 on $\partial D(0,R)$). Using the result of the previous section $h(x)=\mathbb{E}_x\big[g(\widetilde{B}_T)\big]$ is a solution of the Dirichlet problem associated to Ω and g (the domain Ω satisfies the cone condition, so that h is indeed a solution of the Dirichlet problem). But one can check that the RHS of (33) is also a solution to that Dirichlet problem (by computing its Laplacian). We conclude by uniqueness. □

Corollary 22.6.
- *In dimension 2, for any fixed $\varepsilon>0$ and fixed starting point x, the Brownian motion $\widetilde{B}_t$ visits $\partial D(0,\varepsilon)$ with probability 1.*
- *In dimension 3 of higher, for any fixed $\varepsilon>0$ and fixed starting point x, the Brownian motion $\widetilde{B}_t$ visits $\partial D(0,\varepsilon)$ with probability smaller than 1.*
- *Assume $x\neq0$. In dimension 2 or higher, the probability that $\widetilde{B}_t$ visits 0 is 0.*

Proof. Since trajectories are continuous, $\widetilde{B}_t$ visits $\partial D(0,\varepsilon)$ if and only if there exists $R>\|x\|$ such that it visits $\partial D(0,\varepsilon)$ before $\partial D(0,R)$. Therefore

$$\mathbb{P}_x\big[\widetilde{B}_t \text{ visits } \partial D(0,\varepsilon)\big]=\lim_{R\to\infty}\mathbb{P}_x\left[\begin{array}{c}\widetilde{B}_t \text{ visits } \partial D(0,\varepsilon)\\ \text{before } \partial D(0,R)\end{array}\right]$$

Using the formula given in (33), it is immediate that the limit in the RHS is 1 for $d\in\{1,2\}$ and 0 for $d\ge3$.

It remains to prove that $\widetilde{B}_t$ visits 0 with probability 0 in dimension 2 and higher. We start again form (33) and make ε tend to 0 (R being fixed). We have

$$\mathbb{P}_x\left[\begin{array}{c}\widetilde{B}_t \text{ visits } 0\\ \text{before } \partial D(0,R)\end{array}\right] = \lim_{\varepsilon\to 0}\mathbb{P}_x\left[\begin{array}{c}\widetilde{B}_t \text{ visits } \partial D(0,\varepsilon)\\ \text{before } \partial D(0,R)\end{array}\right] = 0. \qquad \square$$

Note: using Markov property and similar argument as for Markov chains, we can prove that the Brownian motion visits a fixed neighbourhood $\partial D(0,\varepsilon)$ of 0 infinitely often in 2D, but only finitely many often in dimension 3 and higher (counting two visits as different if the Brownian motion leaves the disk $\partial D(0,2\varepsilon)$ in-between).

Part E. **Appendices**

Appendix A. Uniform integrability

Definition A.1. *Let $(X_i)_{i\in I}$ be a family of r.v. in L^1. $(X_i)_{i\in I}$ is uniformly integrable (u.i. for short) if*

$$\lim_{c\to\infty}\left(\sup_{i\in I}\mathbb{E}\big[|X_i|\mathbf{1}\{|X_i|\ge c\}\big]\right) = 0.$$

Comment: Let X be in L^1. As c tends to ∞, $|X|\mathbf{1}\{|X|\ge c\}$ tends a.s. to 0. Moreover, it is a.s. bounded by the integrable r.v. $|X|$ (uniformly on c). From the dominated convergence theorem, we have

$$\lim_{c\to\infty}\mathbb{E}\big[|X|\mathbf{1}\{|X|\ge c\}\big] = 0.$$

In summary, a family (X) restricted to a single r.v. is always u.i. Consequently, finite families ($|I|<\infty$) are always u.i. as well.

The interest of uniform integrability lies in its relation with L^1 convergence.

Proposition A.2. *Let $(X_i)_{i\in I}$ be a family of r.v. in L^1.*

(i) If $(X_i)_{i\in I}$ is u.i., then $\sup_{i\in I}\mathbb{E}\big[|X_i|\big]<\infty$;
(ii) If $(X_n)_{n\ge 0}$ is a sequence of r.v. that converges in L^1, then $(X_n)_{n\ge 0}$ is u.i.
(iii) If $(X_n)_{n\ge 0}$ is a sequence of r.v. that converges in probability to Z and if $(X_n)_{n\ge 0}$ is u.i., then X_n converges to Z in L^1.

Proof. (i) Since $(X_i)_{i\in I}$ is u.i., there exists $C>0$ such that

$$\sup_{i\in I}\mathbb{E}\big[|X_i|\mathbf{1}\{|X_i|\ge C\}\big]\le 1.$$

Then, for any i in I,

$$\mathbb{E}\big[|X_i|\big]\le\mathbb{E}\big[|X_i|\mathbf{1}\{|X_i|\ge C\}\big]+\mathbb{E}\big[|X_i|\mathbf{1}\{|X_i|< C\}\big]\le 1+C,$$

showing that $(\mathbb{E}[|X_i|])_{i\in I}$ is bounded, as wanted.

(ii) Fix $\varepsilon>0$. Call Z the limit of X_n in L^1. There exists $N>0$ s.t. $n\ge N \Rightarrow \|X_n-Z\|_1\le\frac{\varepsilon}{3}$. For $n\ge N$ and an arbitrary value $A>0$, we write

$$\begin{aligned}\mathbb{E}\big[|X_n|\,\mathbf{1}\{|X_n|\ge C\}\big] &\le \mathbb{E}\big[|X_n-Z|\,\mathbf{1}\{|X_n|\ge C\}\big]+\mathbb{E}\big[|Z|\,\mathbf{1}\{|X_n|\ge C\}\big]\\ &\le \mathbb{E}\big[|X_n-Z|\big]+\mathbb{E}\big[|Z|\,\mathbf{1}\{|X_n|\ge C\}\,\mathbf{1}\{|Z|\ge A\}\big]+\mathbb{E}\big[|Z|\,\mathbf{1}\{|X_n|\ge C\}\,\mathbf{1}\{|Z|< A\}\big]\\ &\le \tfrac{\varepsilon}{3}+\eta(A)+A\mathbb{P}[|X_n|\ge C],\end{aligned}$$

where $\eta(A):=\mathbb{E}\big[|Z|\mathbf{1}\{|Z|\ge A\}\big]$ tends to 0 when A tends to infinity (indeed, the single r.v. family (Z) is u.i.). We choose A_0 such that $\eta(A_0)\le\frac{\varepsilon}{3}$ and C_1 such that $\mathbb{P}[|X_n|\ge C_1]\le\frac{\varepsilon}{3A_0}$. Then for $n\ge N$ and $C\ge C_1$, the above inequality specialized to $A=A_0$ implies

$$\mathbb{E}\big[|X_n|\,\mathbf{1}\{|X_n|\ge C\}\big]\le\varepsilon.$$

Furthermore, since the finite family $(X_n)_{n<N}$ is u.i., there exists C_2 s.t., for $C\ge C_2$ and $n<N$, we have

$$\mathbb{E}\big[|X_n|\,\mathbf{1}\{|X_n|\ge C\}\big]\le\varepsilon.$$

Summing up, for $C \geq C_0 := \max(C_1, C_2)$, we have

$$\sup_{n\geq 0} \mathbb{E}\big[|X_n|\, \mathbf{1}\{|X_n| \geq C\}\big] \leq \varepsilon.$$

Since this holds for any $\varepsilon > 0$ (with a threshold value C_0 depending on ε), we have

$$\lim_{C\to\infty} \left(\sup_{n\geq 0} \mathbb{E}\big[|X_n|\, \mathbf{1}\{|X_n| \geq C\}\big] \right) = 0,$$

i.e. $(X_n)_{n\geq 0}$ is u.i.

(iii) Fix $\varepsilon > 0$. We write, for an arbitrary $A > 0$,

$$\begin{aligned}\mathbb{E}\Big[|X_n - Z|\Big] \leq \mathbb{E}\Big[|X_n - Z|\, \mathbf{1}\{|X_n - Z| < \tfrac{\varepsilon}{5}\}\Big] \\ + \mathbb{E}\Big[|X_n|\, \mathbf{1}\{|X_n - Z| \geq \tfrac{\varepsilon}{5}\}\mathbf{1}\{|X_n| \geq A\}\Big] + \mathbb{E}\Big[|X_n|\, \mathbf{1}\{|X_n - Z| \geq \tfrac{\varepsilon}{5}\}\mathbf{1}\{|X_n| < A\}\Big] \\ + \mathbb{E}\Big[|Z|\, \mathbf{1}\{|X_n - Z| \geq \tfrac{\varepsilon}{5}\}\mathbf{1}\{|Z| \geq A\}\Big] + \mathbb{E}\Big[|Z|\, \mathbf{1}\{|X_n - Z| \geq \tfrac{\varepsilon}{5}\}\mathbf{1}\{|Z| < A\}\Big].\end{aligned} \tag{34}$$

The terms of the right-hand-side are bounded as follows. The first term is smaller than $\frac{\varepsilon}{5}$. The second and fourth terms are at most $\mathbb{E}\big[|X_n|\mathbf{1}\{|X_n| \geq A\}\big]$ and $\mathbb{E}\big[|Z|\mathbf{1}\{|Z| \geq A\}\big]$, respectively. Since the families $(X_n)_{n\geq 0}$ and (Z) are u.i., we can choose $A = A_0$ such that each of these terms is at most $\frac{\varepsilon}{5}$. When $A = A_0$, the third and fifth terms are each smaller than $A_0\, \mathbb{P}\big[|X_n - Z| \geq \frac{\varepsilon}{5}\big]$. Since X_n tends in probability to Z, one can choose N s.t. $n \geq N$ implies

$$\mathbb{P}\big[|X_n - Z| \geq \tfrac{\varepsilon}{5}\big] \leq \tfrac{\varepsilon}{5A_0},$$

i.e. such that the third and fifth terms are $\frac{\varepsilon}{5}$. Summing up, when $n \geq N$ and when A is specialized to A_0, each term on the right-hand side of (34) is at most $\frac{\varepsilon}{5}$. We conclude that $n \geq N$ implies

$$\mathbb{E}\big[|X_n - Z|\big] \leq \varepsilon.$$

Since this holds for any $\varepsilon > 0$ (with a threshold value N depending on ε), we have $X_n \xrightarrow{L^1} Z$, as wanted. □

Proposition A.3. *A family $(X_i)_{i\in I}$ is u.i. if and only if there exists a function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, such that*

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = +\infty \quad \text{and} \quad \sup_{i\in I} \mathbb{E}\big[\varphi(|X_i|)\big] < +\infty.$$

Moreover, if such a function exists, it can be chosen convex and nondecreasing.

Proof. First assume the existence of φ as in the statement and write $M = \sup_{i\in I} \mathbb{E}\big[\varphi(|X_i|)\big]$. Fix $\varepsilon > 0$. Since $\lim_{x\to\infty} \frac{x}{\varphi(x)} = 0$, there exists C_ε s.t. $x \geq C_\varepsilon$ implies $x \leq \frac{\varepsilon}{M}\varphi(x)$. We have, for any i in I and $C > C_\varepsilon$,

$$\mathbb{E}\big[|X_i|\, \mathbf{1}\{|X_i| \leq C\}\big] \leq \tfrac{\varepsilon}{M}\, \mathbb{E}\big[|\varphi(X_i)|\, \mathbf{1}\{|X_i| \leq C\}\big] \leq \varepsilon.$$

This upper bound is uniform in I. Since it holds for any $\varepsilon > 0$, this proves that $(X_i)_{i\in I}$ is u.i.

Conversely, we assume that $(X_i)_{i\in I}$ is u.i. For each $m \geq 0$, we can find C_m such that

$$\mathbb{E}\big[|X_i|\, \mathbf{1}\{|X_i| \geq C_m\}\big] \leq 2^{-m},$$

for all $i \in I$. Define now $\varphi(x) = \sum_{m\geq 0}(x - C_m)_+$. It is clearly convex and nondecreasing.

On the one hand,

$$\tfrac{\varphi(x)}{x} = \sum_{m\geq 0} \big(1 - \tfrac{C_m}{x}\big)_+$$

tends by the monotone convergence theorem to $\sum_{m\geq 0} 1 = +\infty$.

On the other hand, using again monotone convergence, for all i in I, we have

$$\mathbb{E}\big[\varphi(|X_i|)\big] = \sum_{m\geq 0} \mathbb{E}\big[(|X_i| - C_m)_+\big] \leq \sum_{m\geq 0} \mathbb{E}\big[|X_i|\mathbf{1}\{|X_i| \geq C_m\}\big] \leq \sum_{m\geq 0} 2^{-m} = 2.$$

In particular, $\sup_{i\in I} \mathbb{E}\big[\varphi(|X_i|)\big]$ is finite as wanted. □

Proposition A.3 is a convenient way to prove uniform integrability, you only have to exhibit such a function φ. In particular taking $\varphi(x) = x^p$ for $p > 1$, we have the following important criterion:

$$\sup_{i\in I} \|X_i\|_p < \infty \text{ for some } p > 1 \implies (X_i)_{i\in I} \text{ is u.i.}$$

Additionally, Proposition A.3 is useful for theoretical purposes, for example in the following corollary.

Corollary A.4. *Let Y be a r.v. in L^1. Then the family $(\mathbb{E}[Y|\mathcal{G}])$, where $\mathcal{G}$ runs over all σ-subalgebra $\mathcal{G}$ of $\mathcal{F}$, is u.i.*

Proof. The single-r.v. family (Y) is u.i., so that, by Proposition A.3, there exists a convex nondecreasing function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = +\infty \quad\text{and}\quad \mathbb{E}\big[\varphi(|Y|)\big] < +\infty.$$

For any σ-subalgebra $\mathcal{G}$ of $\mathcal{F}$, we write $Y_\mathcal{G} = \mathbb{E}[Y|\mathcal{G}]$. Using Jensen's inequality for the convex function $x \mapsto \varphi(|x|)$, we have

$$\varphi\big(|Y_\mathcal{G}|\big) \le \mathbb{E}\big[\varphi(|Y|)|\mathcal{G}\big] \text{ a.s.}$$

Taking expectation we have

$$\mathbb{E}\Big[\varphi\big(|Y_\mathcal{G}|\big)\Big] \le \mathbb{E}\Big[\mathbb{E}\big[\varphi(|Y|)|\mathcal{G}\big]\Big] = \mathbb{E}\big[\varphi(|Y|)\big],$$

where we use the standard rule for the expectation of the conditional expectation. The right-hand side is finite by construction of φ, and the bound is uniform on all σ-subalgebra $\mathcal{G}$ of $\mathcal{F}$. By Proposition A.3, this implies that the family $(Y_\mathcal{G})_{\mathcal{G}\subseteq\mathcal{F}}$ is u.i., which is what we wanted to prove. ☐

Appendix B. L^p convergence of random variables

Question. A standard question arising often in probability theory is the following: we know that $X_n \to X$ a.s. (or in probability), can we conclude that $\mathbb{E}[X_n]$ tends to $\mathbb{E}[X]$? or more generally that $\mathbb{E}[f(X_n)]$ tends to $\mathbb{E}[f(X)]$ for some function f?

Answer.

- Yes, if f is a **bounded** function (since convergence a.s./in probability implies convergence in distribution).
- No, in general. A sufficient condition for $\lim_{n\to\infty} \mathbb{E}[X_n] \to \mathbb{E}[X]$ is that X_n tends to X in L^1 (indeed, $|\mathbb{E}[X_n] - \mathbb{E}[X]| \le \mathbb{E}[|X_n - X|] = \|X_n - X\|_1$). Since we assumed X_n to converges a.s./in probability, this happens if and only if X_n is u.i.
- Take more generally $f(x) = x^p$ for $p \ge 1$, i.e. we are asking whether $\mathbb{E}[X_n^p]$ tends to $\mathbb{E}[X^p]$ (assume either p integer or X_n, X nonnegative to define the relevant quantities).

  A sufficient condition is that X_n tends to X in L^p. Indeed, this implies $\|X_n\|_p \to \|X\|_p$, and, thus, taking the p-th power, $\mathbb{E}[X_n^p] \to \mathbb{E}[X^p]$.

This raises the following problem: find conditions (necessary and/or sufficient) for L^p convergence?

Proposition B.1. *Let $(X_n)_{n\ge0}$ be a sequence of r.v. and fix $q > p \ge 1$.*

(i) If X_n converges to X_∞ in L^p, then $\sup_{n\ge0} \|X_n\|_p < +\infty$.
(ii) If $\sup_{n\ge0} \|X_n\|_q < +\infty$ and if X_n tends to some X_∞ in probability, then X_n tends to X_∞ in L^p.

In particular L^q convergence implies L^p convergence (this can also be proved directly with Hölder's inequality).

Proof. (i) We have $\|X_n\|_p \le \|X_n - X_\infty\|_p + \|X_\infty\|_p$. The first term tends to 0, while the second does not depend on n. The sum is therefore bounded.

(ii) We want to prove that $\mathbb{E}[|X_n - X_\infty|^p]$ tends to 0. Since $|X_n - X_\infty|^p$ tends to 0 in probability, it is enough to prove that the sequence $(|X_n - X_\infty|^p)_{n\ge0}$ is u.i. We consider the convex increasing function $\varphi(x) = x^{q/p}$, for $x \ge 0$. We have $\lim_{x\to+\infty} \frac{\varphi(x)}{x} = +\infty$. Moreover,

$$\mathbb{E}\big[|\varphi(|X_n - X_\infty|^p)\big] = \mathbb{E}\big[|X_n - X_\infty|^q\big] = \|X_n - X_\infty\|_q^q \le (\|X_n\|_q + \|X_\infty\|_q)^q.$$

But $\|X_n\|_q$ is assumed to be bounded. Using Fatou's lemma on a subsequence X_{n_k} converging a.s. to X_∞, we have

$$\|X_\infty\|_q \leq \liminf_{k\to\infty} \|X_{n_k}\|_q \leq \sup_{n\geq 0} \|X_n\|_q < +\infty.$$

We conclude that $\mathbb{E}\big[|\varphi(|X_n - X_\infty|^p)\big]$ is bounded. From Proposition A.3, this proves that the sequence $(|X_n - X_\infty|^p)_{n\geq 0}$ is u.i. $\square$

Acknowledgements

The main sources used for preparing this lecture have been

- Jacod–Protter's book *Probability essentials* [JP04] (from Chapter 22 on) for the conditional expectation and martingale parts;
- Grimmett–Stirzaker's book *Probability and Random Precesses* [GS01], Chapter 6, for the Markov chain part;
- Le Gall's lecture notes on *Intégration, Probabilités et Processus Aléatoires* (in French) [LG06], as a complement for the first three parts, and as a main source for the last part on Brownian motion.

References

[GS01] G. Grimmett, D. Stirzaker, *Probability and Random Precesses*, third edn, Oxford University Press, 2001.
[JP04] J. Jacop, Ph. Protter, *Probability Essentials*, Universitytext, Springer, 2004.
[LG06] J.F. Le Gall, *Intégration, Probabilités et Processus Aléatoires*, unpublished lecture notes available online, 2006.

University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland