Introduction to Probability

Nathanaël Berestycki
Lecture notes[1] - Michaelmas 2007

These notes are a list of some of the definitions and results presented in the course. Since they are free from any motivating exposition or examples, and since no proofs are given for any of the theorems, these notes should be used only as a reference to supplement the recommended texts. The class material consists, ultimately, of what was done in class and in the exmaple sheets. In what follows, phrases enclosed in brackets [. . .] are digressive or slightly outside of the syllabus of the course, but are included for completeness. A table of notation is in the appendix.

1. Basic tools

1.1. Probability spaces.

Definition 1.1. Let Ω be a set. A *sigma-field* on Ω is a non-empty set $\mathcal{F}$ of subsets of Ω such that

(1) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$,
(2) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$.

[The terms sigma-field and *sigma-algebra* are interchangeable.]

Definition 1.2. Let Ω be a set and let $\mathcal{F}$ be a sigma-field on Ω. A *probablity measure* on $\mathcal{F}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

(1) if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint then $\mathbb{P}(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mathbb{P}(A_i)$,
(2) $\mathbb{P}(\Omega) = 1$.

[The terms *disjoint* and *mutually exclusive* are interchangeable and refer to events A and B such that $A \cap B = \emptyset$.]

Definition 1.3. Let Ω be a set, $\mathcal{F}$ a sigma-field on Ω, and $\mathbb{P}$ a probability measure on $\mathcal{F}$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.*

The set Ω is called the *sample space*, and an element of Ω is called an *outcome.* A subset of Ω which is an element of $\mathcal{F}$ is called an *event.*

Let $A \in \mathcal{F}$ be an event. If $\mathbb{P}(A) = 1$ then A is called an *almost sure* event, and if $\mathbb{P}(A) = 0$ then A is called a *null* event. [The phrase "almost surely" is often abbreviated *a.s.*]

[1] Version of November 28th, 2007

1.2. **Random variables and distribution functions.**

Definition 1.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* is a function $X : \Omega \to \overline{\mathbb{R}}$ such that the set $\{\omega \in \Omega : X(\omega) \leq t\}$ is an element of $\mathcal{F}$ for all $t \in \mathbb{R}$.

Let A be a subset of $\overline{\mathbb{R}}$, and let X be a random variable. We use the notation $\{X \in A\}$ to denote the set $\{\omega \in \Omega : X(\omega) \in A\}$. A random variable X is said to *take values* in a subset $S \subseteq \overline{\mathbb{R}}$ if $X \in S$ almost surely.

The *distribution function* of X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(t) = \mathbb{P}(X \leq t)$$

for all $t \in \mathbb{R}$.

[A distribution function is called *defective* if either $\lim_{t\uparrow\infty} F_X(t) < 1$ or $\lim_{t\downarrow-\infty} F_X(t) > 0$. Unless otherwise indicated, all random variables considered here are assumed to have non-defective distribution functions.]

The *law* of X is the probability measure ν such that

$$\nu([a, b]) = \mathbb{P}(a \leq X \leq b)$$

for all $a \leq b$. The statement "the random variable X has law ν" is written $X \sim \nu$. [The measure ν is defined on the*Borel* sigma-field $\mathcal{B}$, which is defined as the smallest sigma-field on $\mathbb{R}$ containing the intervals $[a, b]$ for all $a \leq b$.]

Definition 1.5. Let A be an event in Ω. The *indicator function* of the event A is the random variable $1_A : \Omega \to \{0, 1\}$ defined by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases}$$

for all $\omega \in \Omega$.

Definition 1.6. A random variable X is called *discrete* if X takes values in a countable set.

If X is discrete, the function $p_X : \mathbb{R} \to [0, 1]$ defined by $p_X(t) = \mathbb{P}(X = t)$ is called the *mass function* of X.

Definition 1.7. Let X is a discrete random variable taking values in $\mathbb{N}$ with mass function p_X.

The random variable X is called

- *Bernoulli* with parameter p if

$$p_X(0) = 1 - p \text{ and } p_X(1) = p.$$

  where $0 \leq p \leq 1$.

- *binomial* with parameters n and p, written $X \sim \text{bin}(n,p)$, if

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for all } k = 0, 1, \dots, n$$

where $n \in \mathbb{N}$ and $0 \le p \le 1$.
- *Poisson* with parameter λ if

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for all } k = 0, 1, 2, \dots$$

where $\lambda \ge 0$.
- *geometric* with parameter p if

$$p_X(k) = p(1-p)^{k-1} \text{ for all } k = 1, 2, 3, \dots$$

where $0 \le p \le 1$.

[In the above formulae, the convention $0^0 = 1$ is used. If X is geometric with parameter $p = 0$, then $X = \infty$ almost surely.]

In practice, a random variable is specified by its law. We almost never need to construct explicitly the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated with it.

Definition 1.8. Let F_X be the distribution of a random variable X. The random variable X is *continuous* if and only if F_X is a continuous function.

The random variable X is *absolutely continuous* if and only if there exists a positive function $f_X : \mathbb{R} \to [0, \infty)$ such that

$$F_X(t) = \int_{-\infty}^{t} f_X(s) ds$$

for all $t \in \mathbb{R}$, in which case the function f_X is called the *density function* of X.

Definition 1.9. Let X is a continuous random variable with density function f_X.

The random variable X is called

- *uniform* on the interval (a,b), written $X \sim \text{unif}(a,b)$, if

$$f_X(t) = \frac{1}{b-a} \text{ for all } a < t < b$$

for some $a < b$.
- *normal* with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, if

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \text{ for all } t \in \mathbb{R}$$

for some $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

- *exponential* with rate λ, written $X \sim \exp(\lambda)$, if

$$f_X(t) = \lambda e^{-\lambda t} \text{ for all } t \geq 0$$

for some $\lambda > 0$.
- *Cauchy* if

$$f_X(t) = \frac{1}{\pi(1+t^2)} \text{ for all } t \in \mathbb{R}.$$

1.3. **Expectations and variances.**

Definition 1.10. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *expected value* of X is denoted by $\mathbb{E}(X)$ and is given by the integral

$$\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega).$$

The above integral is defined in the following cases:

- if $X \geq 0$ almost surely.
- if either $\mathbb{E}(X^+)$ or $\mathbb{E}(X^-)$ is finite, in which case $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$.

A random variable X is *integrable* if $\mathbb{E}|X| < \infty$ and is *square-integrable* if $\mathbb{E}(X^2) < \infty$. The terms expected value, *expectation*, and *mean* are interchangeable.

The *variance* of an integrable random variable X, written $\mathrm{Var}(X)$, is

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

The *covariance* of square-integrable random variable X and Y, written $\mathrm{Cov}(X, Y)$, is

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

If neither X or Y is almost surely constant, then their correlation, written $\rho(X, Y)$, is

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)^{1/2}\mathrm{Var}(Y)^{1/2}}.$$

Random variables X and Y are called *uncorrelated* if $\mathrm{Cov}(X, Y) = 0$.

Theorem 1.11. *Let the function $g : \mathbb{R} \to \mathbb{R}$ be such that $g(X)$ is integrable.*

If X is a discrete random variable with probability mass function p_X taking values in a countable set S then

$$\mathbb{E}(g(X)) = \sum_{t \in S} g(t)\ p_X(t).$$

If X is an absolutely continuous integrable random variable with density function f_X then

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t)\ f_X(t)\ dt.$$

Theorem 1.12 (Cauchy–Schwarz inequality)**.** *Let X and Y square-integrable random variables. Then*

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

with equality if and only if $aX = bY$ almost surely, for some constants $a, b \in \mathbb{R}$.

1.4. **Conditional probability and expectation, independence.**

Definition 1.13. Let B be an event with $\mathbb{P}(B) > 0$. The *conditional probability* of an event A given B, written $\mathbb{P}(A|B)$, is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

The *conditional expectation* of X given B, written $\mathbb{E}(X|B)$, is

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X1_B)}{\mathbb{P}(B)}.$$

Theorem 1.14 (The law of total probability)**.** *Let $B_1, B_2, \ldots$ be disjoint, non-null events such that $\bigcup_{i=1}^{\infty} B_i = \Omega$. Then*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

for all events A.

Definition 1.15. Let X and Y be random variables. The function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ defined by

$$F_{X,Y}(s, t) = \mathbb{P}(X \leq s, Y \leq t)$$

is called their *joint distribution function.*

If both X and Y are discrete random variables, then the function $p_{X,Y} : \mathbb{R}^2 \to [0, 1]$ defined by

$$p_{X,Y}(s, t) = \mathbb{P}(X = s, Y = t)$$

is called their *joint mass function.* The *conditional mass function* of X given $Y = t$, where $p_Y(t) > 0$, is defined as

$$p_{X|Y}(s|t) = \mathbb{P}(X = s|Y = t) = \frac{p_{X,Y}(s, t)}{p_Y(t)}.$$

If there exists a function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ such that

$$F_{X,Y}(s,t) = \int_{u=-\infty}^{s} \int_{v=-\infty}^{t} f_{X,Y}(u,v)\ du\ dv$$

then X and Y are said to be *jointly absolutely continuous* and $f_{X,Y}$ is called their *joint density function*. The *conditional density function* of X given $Y = t$, where $f_Y(t) > 0$, is defined as

$$f_{X|Y}(s|t) = \lim_{\delta\downarrow 0, \epsilon\downarrow 0} \frac{1}{\delta}\mathbb{P}(|X - s| < \delta \bigm| |Y - t| < \epsilon) = \frac{f_{X,Y}(s,t)}{f_Y(t)}.$$

Theorem 1.16. *Let the function $g : \mathbb{R}^2 \to \mathbb{R}$ be such that $g(X,Y)$ is integrable.*

If X and Y are discrete and taking values in a countable set S then

$$\mathbb{E}(g(X,Y)) = \sum_{s,t\in S} g(s,t)\ p_{X,Y}(s,t).$$

If X and Y are jointly absolutely continuous then

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(s,t)\ f_{X,Y}(s,t)\ ds\ dt.$$

Definition 1.17. Random variables X and Y are *jointly normal* with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ, written

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix} \right)$$

if the joint density function is

$$f_{X,Y}(s,t) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp(-\frac{1}{2}Q(s,t))$$

where

$$Q(s,t) = \frac{1}{1-\rho^2}\left((\frac{s-\mu_X}{\sigma_X})^2 - 2\rho(\frac{s-\mu_X}{\sigma_X})(\frac{t-\mu_Y}{\sigma_Y}) + (\frac{t-\mu_Y}{\sigma_Y})^2 \right)$$

Definition 1.18. The *conditional expectation* of X given $Y = t$, written $\mathbb{E}(X|Y = t)$, is defined by either Definition 1.13, if $\mathbb{P}(Y = t) > 0$, or by the formula

$$\mathbb{E}(X|Y = t) = \lim_{\epsilon\downarrow 0} \frac{\mathbb{E}(X1_{\{|Y-t|<\epsilon\}})}{\mathbb{P}(|Y - t| < \epsilon)}$$

if $\mathbb{P}(Y = t) = 0$.

For fixed random variable X and Y, let $h : \mathbb{R} \to \mathbb{R}$ be the function defined by $h(t) = \mathbb{E}(X|Y = t)$. The *conditional expectation* of X given Y, written $\mathbb{E}(X|Y)$, is

$$\mathbb{E}(X|Y) = h(Y).$$

Theorem 1.19. *Let the function $g : \mathbb{R} \to \mathbb{R}$ be such that $g(X)$ is integrable.*

If X and Y are discrete taking values in S and $p_Y(t) > 0$, then

$$\mathbb{E}(g(X)|Y = t) = \sum_{s \in S} g(s)\ p_{X|Y}(s, t).$$

If X and Y are jointly absolutely continuous and if $f_Y(t) > 0$ then

$$\mathbb{E}(g(X)|Y = t) = \int_{-\infty}^{\infty} g(s)\ f_{X|Y}(s, t)\ ds.$$

Theorem 1.20 (The law of iterated expectation)**.** *Let X and Y be random variables and suppose X is integrable. Then*

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)).$$

Definition 1.21. Let $A_1, A_2, \ldots$ be events. If

$$\mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i)$$

for every finite subset $I \subset \mathbb{N}$ then the events are said to be *independent.*

Random variables $X_1, X_2, \ldots$ are called *independent* if the events $\{X_1 \leq t_1\}, \{X_2 \leq t_2\}, \ldots$ are independent. [The phrase "independent and identically distributed" is often abbreviated *i.i.d.*]

Theorem 1.22. *If X and Y are discrete and independent random variables then*

$$p_{X,Y}(s, t) = p_X(s)p_Y(t).$$

If X and Y are jointly absolutely continuous and independent random variables then

$$f_{X,Y}(s, t) = f_X(s)f_Y(t).$$

If X and Y are independent and integrable, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

1.5. **Probability inequalities.**

Theorem 1.23 (Markov's inequality)**.** *Let X be a nonnegative random variable with $\mathbb{E}(X) < \infty$. Then*

$$\mathbb{P}(X \geq A) \leq \frac{\mathbb{E}(X)}{A}$$

for all $A > 0$.

Corollary 1.24 (Chebychev's inequality). *Let X be a random variable with $\mathbb{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Then*

$$\mathbb{P}(|X - \mu| \geq A) \leq \frac{\sigma^2}{A^2}$$

for all $A > 0$.

1.6. **Generating and characteristic functions.**

Definition 1.25. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\{0, 1, 2, \ldots\}$. The *moment generating function* of X is the function $G_X : [0,1] \to [0,1]$ defined by

$$G_X(t) = \mathbb{E}(t^X)$$

for all $t \in (0, 1]$ and $G_X(0) = \mathbb{P}(X = 0)$.

Theorem 1.26. *Let G_X be the moment generating function of X. Then*

$$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0)$$

for all $n \in \mathbb{N}$, where $G_X^{(n)}(0)$ denotes the n-th derivative of G_X evalutated at 0.

Definition 1.27. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The *Laplace transform* of X is the function $M_X : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ defined by

$$M_X(t) = \mathbb{E}(e^{tX})$$

for all $t \in \mathbb{R}$.

Theorem 1.28. *Let M_X be the Laplace transform of a random variable X, and suppose there exists an $\epsilon > 0$ such that $M_X(t) < \infty$ for all $-\epsilon < t < \epsilon$. Then*

$$\mathbb{E}(X^n) = M_X^{(n)}(0)$$

for all $n \in \mathbb{N}$, where $M_X^{(n)}(0)$ denotes the n-th derivative of M_X evalutated at 0. (The number $\mu_n = \mathbb{E}(X^n)$ is called the n-th moment *of X.)*

Definition 1.29. The *characteristic function* of a real-valued random variable X is the function $\phi_X : \mathbb{R} \to \mathbb{C}$ defined by

$$\phi_X(t) = \mathbb{E}(e^{itX})$$

for all $t \in \mathbb{R}$, where $i = \sqrt{-1}$.

Theorem 1.30 (Uniqueness of generating and characteristic functions). *Let X and Y be real-valued random variables with distribution functions F_X and F_Y.*

- *Let ϕ_X and ϕ_Y be the characteristic functions of X and Y. Then*
$$\phi_X(t) = \phi_Y(t) \text{ for all } t \in \mathbb{R}$$
if and only if
$$F_X(t) = F_Y(t) \text{ for all } t \in \mathbb{R}.$$
- *Let X and Y be valued in $\mathbb{N}$ with moment generating functions G_X and G_Y. Then*
$$G_X(t) = G_Y(t) \text{ for all } t \in [0, 1]$$
if and only if
$$F_X(t) = F_Y(t) \text{ for all } t \in \mathbb{R}.$$
- *Let M_X and M_Y be the Laplace transforms of X and Y. If there exists an $\epsilon > 0$ such that*
$$M_X(t) = M_Y(t) < \infty \text{ for all } t \in (-\epsilon, \epsilon)$$
then
$$F_X(t) = F_Y(t) \text{ for all } t \in \mathbb{R}.$$

Theorem 1.31. *Inversion formula. Let X be a random variable with characteristic function φ. Then for any $a < b$*
$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt = \mathbb{P}(a < X < b) + \frac{1}{2}(\mathbb{P}(X = a) + \mathbb{P}(X = b))$$

2. Fundamental probability results

Definition 2.1. Let $x_1, x_2, \ldots$ be a sequence of real numbers. The *limit superior* is defined by
$$\limsup_{n\uparrow\infty} x_n = \inf_{N\in\mathbb{N}} \sup_{n\geq N} x_n$$
and *limit inferior* by
$$\liminf_{n\uparrow\infty} x_n = \sup_{N\in\mathbb{N}} \inf_{n\geq N} x_n.$$
Equivalently, if $x \in \mathbb{R}$ then
$$x = \limsup_{n\uparrow\infty} x_n \Leftrightarrow \text{ for all } \epsilon > 0 \begin{cases} \{n \in \mathbb{N} : x_n > x + \epsilon\} \text{ is finite} \\ \{n \in \mathbb{N} : x_n > x - \epsilon\} \text{ is infinite} \end{cases}$$
and
$$x = \liminf_{n\uparrow\infty} x_n \Leftrightarrow \text{ for all } \epsilon > 0 \begin{cases} \{n \in \mathbb{N} : x_n < x + \epsilon\} \text{ is infinite} \\ \{n \in \mathbb{N} : x_n < x - \epsilon\} \text{ is finite} \end{cases}$$
If $\limsup_{n\uparrow\infty} x_n = \liminf_{n\uparrow\infty} x_n = x$ then the sequence $x_1, x_2, \ldots$ is *convergent* and with *limit* $x = \lim_{n\uparrow\infty} x_n$.

Definition 2.2 (Modes of convergence)**.** Let $X_1, X_2, \ldots$ and X be random variables.

- $X_n \to X$ almost surely if $\mathbb{P}(X_n \to X) = 1$
- $X_n \to X$ in L_p, for $p \geq 1$, if $\mathbb{E}|X|^p < \infty$ and $\mathbb{E}|X_n - X|^p \to 0$
- $X_n \to X$ in probability if $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ for all $\epsilon > 0$
- $X_n \to X$ in distribution if $F_{X_n}(t) \to F_X(t)$ for all points $t \in \mathbb{R}$ of continuity of F_X

Theorem 2.3. *The following implications hold:*

$$\left.\begin{array}{c} X_n \to X \textit{ almost surely} \\ \textit{or} \\ X_n \to X \textit{ in } L_p, p \geq 1 \end{array}\right\} \Rightarrow X_n \to X \textit{ in probability } \Rightarrow X_n \to X \textit{ in distribution}$$

Furthermore, if $r \geq p \geq 1$ *then* $X_n \to X$ *in* $L_r \Rightarrow X_n \to X$ *in* L_p.

Definition 2.4. Let $A_1, A_2, \ldots$ be events. The term *eventually* is defined by

$$\{A_n \text{ eventually}\} = \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} A_n$$

and *infinitely often* by

$$\{A_n \text{ infinitely often}\} = \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} A_n.$$

[The phrase "infinitely often" is often abbreviated *i.o.*]

Theorem 2.5 (The first Borel-Cantelli lemma)**.** *Let $A_1, A_2, \ldots$ be a sequence of events. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

then $\mathbb{P}(A_n$ *infinitely often*$) = 0$.

Theorem 2.6 (The second Borel-Cantelli lemma)**.** *Let $A_1, A_2, \ldots$ be a sequence of* independent *events. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

then $\mathbb{P}(A_n$ *infinitely often*$) = 1$.

Theorem 2.7 (A sufficient condition for almost sure convergence)**.** *Let $X_1, X_2, \ldots$ and X be random variables. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty$$

for all $\epsilon > 0$ then $X_n \to X$ almost surely.

Theorem 2.8 (Monotone convergence theorem)**.** *Let* $X_1, X_2, \ldots$ *be positive random variables with* $X_n \le X_{n+1}$ *almost surely for all* $n \ge 1$*, and let* $X = \sup_{n\in\mathbb{N}} X_n$*. Then* $X_n \to X$ *almost surely and*

$$\mathbb{E}(X_n) \to \mathbb{E}(X).$$

Theorem 2.9 (Fatou's lemma)**.** *Let* $X_1, X_2, \ldots$ *be positive random variables. Then*

$$\mathbb{E}(\liminf_{n\uparrow\infty} X_n) \le \liminf_{n\uparrow\infty} \mathbb{E}(X_n).$$

Theorem 2.10 (Dominated convergence theorem)**.** *Let* $X_1, X_2, \ldots$ *and* X *be random variables such that* $X_n \to X$ *almost surely. If* $\mathbb{E}(\sup_{n\ge 1}|X_n|) < \infty$ *then*

$$\mathbb{E}(X_n) \to \mathbb{E}(X).$$

Definition 2.11. Let $X_1, X_2, \ldots$ be a collection of random variables, and let $S_n = X_1 + \ldots + X_n$. The collection of random variables satisfies a *weak law of large numbers* if $S_n/n \to \mu$ in probability for some constant μ. The collection satisfies a *strong law of large numbers* if $S_n/n \to \mu$ almost surely.

Theorem 2.12 (A weak law of large numbers)**.** *Let* $X_1, X_2, \ldots$ *be independent and indentically distrubuted with*

$$N\mathbb{P}(|X_i| > N) \to 0 \text{ and } \mathbb{E}(X1_{\{|X_i|\le N\}}) \to \mu \text{ as } N \uparrow \infty$$

for some $\mu \in \mathbb{R}$ *then*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ in probability.}$$

Theorem 2.13 (A strong law of large numbers)**.** *Let* $X_1, X_2, \ldots$ *be independent and identically distributed integrable random variables with common mean* $\mathbb{E}(X_i) = \mu$*. Then*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \text{ almost surely.}$$

Theorem 2.14 (Lévy's continuity theorem)**.** *Let* $X_1, X_2, \ldots$ *and* X *be random variables with characteristic functions* $\phi_1, \phi_2, \ldots$ *and* ϕ *respectively. The following are equivalent:*

- $X_n \to X$ *in distribution*
- $\phi_n(t) \to \phi(t)$ *for all* $t \in \mathbb{R}$

Theorem 2.15 (Central limit theorem)**.** *Let* $X_1, X_2, \ldots$ *be independent and identically distributed with* $\mathbb{E}(X_i) = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2$ *for each* $i = 1, 2, \ldots$*, and let*

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then $Z_n \to Z$ in distribution, where $Z \sim N(0,1)$.

3. Markov chains

Definition 3.1. A *stochastic process* is a collection $(X_i)_{i\in I}$ of random variables. If the index set I is $\mathbb{N}$, then the stochastic process is called *discrete time* and is denoted by $(X_n)_{n\geq 0}$. If the index set I is $[0,\infty)$, then the stochastic process is called *continuous time* and is denoted by $(X_t)_{t\geq 0}$.

3.1. Discrete time Markov chains.

Definition 3.2. Let $(X_n)_{n\geq 0}$ be a discrete time stochastic process taking values in a countable set S. The process $(X_n)_{n\geq 0}$ is called a *Markov chain* if

$$\mathbb{P}(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1})$$

for each $n \geq 1$, and $s_0, \dots, s_n \in S$. The set S is called the *state space* of the Markov chain, and an elements of S is called a *state*. The Markov chain is called *time homogeneous* if

$$\mathbb{P}(X_n = j | X_{n-1} = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

for all $n \geq 1$ and all states $i, j \in S$.

Throughout these notes, all Markov chains are assumed to be time homogeneous.

Definition 3.3. Let $(X_n)_{n\geq 0}$ be a Markov chain on a state space S. The numbers $p_{ij} = \mathbb{P}(X_1 = j | X_0 = i)$ for $i, j \in S$ are called the *one-step transition probabilities*, and $p_{ij}(n) = \mathbb{P}(X_n = j | X_0 = i)$ for $n \geq 0$ the *n-step transition probabilities*.

The $|S| \times |S|$ matrix $P = (p_{ij})_{i,j\in S}$ is called the *transition matrix*.

We use the notation

$$\mathbb{P}_i(A) = \mathbb{P}(A | X_0 = i)$$

for any event $A \in \mathcal{F}$, and

$$\mathbb{E}_i(Z) = \mathbb{E}(Z | X_0 = i)$$

for any random variable Z for which the conditional expectation can be defined.

Theorem 3.4 (Chapman-Kolmogorov equations)**.** *Let $p_{ij}(n)$ for $i, j \in S$ and $n \geq 0$ denote the n-step transition probabilities of a Markov chain with state space S. Then*

$$p_{ik}(m+n) = \sum_{j\in S} p_{ij}(m) p_{jk}(n)$$

for all $i,k \in S$ *and* $m,n \geq 0$. *In particular, the* n*-step transition probabilities are given by*

$$p_{ij}(n) = (P^n)_{ij}$$

where P *is the transition matrix of the Markov chain.*

Definition 3.5. A $|S| \times |S|$ matrix $P = (p_{ij})_{i,j \in S}$ is called *stochastic* if $p_{ij} \geq 0$ for all $i,j \in S$ and $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$. In particular, a matrix is stochastic if and only if it is the transition matrix of a Markov chain.

Theorem 3.6. *Let* P *be an* $d \times d$ *matrix where* $d < \infty$. *Let* $\lambda_1, \ldots, \lambda_d$ *be the eigenvalues of* P. *If the eigenvalues are distinct, then there exists complex numbers* $a_{ij}^{(k)}$ *for* $i,j,k \in \{1, \ldots, d\}$ *such that*

$$(P^n)_{ij} = \sum_{k=1}^{d} a_{ij}^{(k)} \lambda_k^n$$

for all $n \in \mathbb{N}$. *More generally, if the eigenvalues are not necessarily distinct, then there exists polynomials* $a_{ij}^{(k)} : \mathbb{N} \to \mathbb{C}$ *for* $i,j,k \in \{1, \ldots, d\}$ *of degree less than the multiplicity of the eigenvalue* λ_k *such that*

$$(P^n)_{ij} = \sum_{k=1}^{d} a_{ij}^{(k)}(n) \lambda_k^n$$

for all $n \in \mathbb{N}$.

Theorem 3.7. *Let* $(X_n)_{n \geq 0}$ *be a Markov chain on a state space* S, *and let* $A \subseteq S$. *Define the random variable* H^A *valued in* $\mathbb{N} \cup \{\infty\}$ *by*

$$H^A = \inf\{n \geq 0 : X_n \in S\}.$$

[The standard convention $\inf \emptyset = \infty$ *is used throughout.] For each state* $i \in S$ *let* $h_i^A = \mathbb{P}_i(H^A < \infty)$. *Then* $(h_i^A)_{i \in S}$ *is the minimal non-negative solution to*

$$h_i^A = \begin{cases} 1 & \text{if } i \in A \\ \sum_{i \in S} p_{ij} h_j^A & \text{otherwise} \end{cases}$$

Definition 3.8. Let $p_{ij}(n)$ for $i,j \in S$ and $n \geq 0$ denote the n-step transition probabilities of a Markov chain with state space S.

States i *leads* to state j, written $i \to j$, if there exists an $n \geq 0$ such that $p_{ij}(n) > 0$. States i and j *communicate*, written $i \leftrightarrow j$, if $i \to j$ and $j \to i$. The *communicating class* containing a state i is the largest subset $C \subseteq S$ with the property that if state j is in C then states i and j communicate.

A communicating class C is called *closed* it has the property if $i \in C$ and $i \to j$ then $j \in C$. Otherwise, C is called *open* if there exists a state i in C and a state j not in C such that i leads to j.

A Markov chain is *irreducible* if all states in S communicate.

Definition 3.9. A state $i \in S$ is called *recurrent* if

$$\mathbb{P}_i(X_n = i \text{ infinitely often}) = 1,$$

and called *transient* if

$$\mathbb{P}_i(X_n = i \text{ infinitely often}) < 1.$$

Let $C \subseteq S$ be a communicating class of the Markov chain. The class is called a *transient class* if every state $i \in C$ is transient, and the class is called a *recurrent class* if every state $i \in C$ is recurrent.

[The terms recurrent and *persistent* are interchangeable.]

Definition 3.10. Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S. A *stopping time* T is a random variable taking values in $\mathbb{N}\cup\{\infty\}$ having the property that for each $n \in \mathbb{N}$ the event $\{T = n\}$ is determined by $X_0, \ldots, X_n\}$ in the sense that there exists a function $f : S^n \to \{0, 1\}$ such that

$$1_{\{T=n\}} = f(X_0, \ldots, X_n).$$

Theorem 3.11 (The strong Markov property)**.** *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S and let T be a stopping time. Then conditional on the events $\{T < \infty\}$ and $X_T = i$, the random process $(X_{T+n})_{n\geq 0}$ is a Markov chain starting at i, independent of $X_0, \ldots, X_T$.*

Theorem 3.12. *Let $(X_n)_{n\geq 0}$ be a Markov chain with state space S. For each state $i \in S$ define the random variable T_i taking values in $\mathbb{N} \cup \{\infty\}$ by*

$$T_i = \inf\{n \geq 1 : X_n = i\}.$$

The following are equivalent:

- *State i is recurrent.*
- $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$.
- $\mathbb{P}_i(T_i < \infty) = 1$

Moreover, the following are equivalent:

- *State i is transient.*
- $\mathbb{P}_i(X_n = i$ *infinitely often*$) = 0$.
- $\mathbb{P}_i(T_i < \infty) < 1$.
- $\sum_{n=1}^{\infty} p_{ii}(n) = \frac{1}{1-\mathbb{P}_i(T_i<\infty)} < \infty$.

Definition 3.13. Let $(X_n)_{n\geq 0}$ be a Markov chain and let $T_i = \inf\{n \geq 1 : X_n = i\}$. The state i is called *positive recurrent* if $\mathbb{E}_i(T_i) < \infty$ and is called *null recurrent* otherwise. [The terms positive and *non-null* are interchangeable in the context of recurrent Markov chains.]

Theorem 3.14 (Recurrence, transience, and positive recurrence are class properties)**.** *Let i and j be communicating states of a Markov chain. Then i is recurrent if and only if j is recurrent. Equivalently, i is transient if and only if j is transient. Also, i is positive recurrent if and only if j is positive recurrent.*

Definition 3.15. Let P be the transition matrix of a Markov chain. An *invariant distribution* is a row vector $\pi = (\pi_i)_{i\in S}$ such that $\pi_i \geq 0$ for all $i \in S$, $\sum_{i\in S} \pi_i = 1$ and

$$\pi P = \pi.$$

[The terms invariant distribution and *stationary distribution* are interchangeable.]

Theorem 3.16. *Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain with transition matrix P. Then $(X_n)_{n\geq 0}$ is positive recurrent if an only if there exists an invariant distribution for P. If the Markov chain is positive recurrent then the invariant distribution is unique and is given by the formula*

$$\pi_i = \frac{1}{\mathbb{E}_i(T_i)}$$

where $T_i = \inf\{n \geq 1 : X_n = i\}$.

Theorem 3.17. *Let $(X_n)_{n\geq 0}$ be an irreducible recurrent Markov chain on S with transition matrix P. Let $k \in S$ be a fixed state and let $T_k = \inf\{n \geq 1 : X_n = k\}$. For each $i \in S$ let*

$$\gamma_i^{(k)} = \mathbb{E}_k \sum_{n=1}^{T_k} 1_{\{X_n = i\}}$$

be the expected number of times, conditional starting from state k, the chain visits state i before returning to k, and let $\gamma^{(k)}$ be the row vector $(\gamma_i^{(k)})_{i\in S}$. Then $\gamma^{(k)}$ is satisfies the equation

$$\gamma^{(k)} P = \gamma^{(k)}$$

with $\gamma_k^{(k)} = 1$. If $(X_n)_{n\geq 0}$ is positive recurrent with invariant distribution π then

$$\gamma_i^{(k)} = \frac{\pi_i}{\pi_k}.$$

Definition 3.18. Let $(p_{ij}(n))_{i,j\in S}$ be the n-step transition probabilities for a Markov chain. The *period* d_i of the state i is given by

$$d_i = \gcd\{n \geq 1 : p_{ii}(n) > 0\}.$$

A state i is *aperiodic* if $d_i = 1$.

A communicating class C is has *period* d if $d = d_i$ for all $i \in C$. A communicating class is *aperiodic* if it has period $d = 1$.

Theorem 3.19 (Periods are class properties)**.** *If i and j are communicating states of a Markov chain, then they have the same period $d_i = d_j$.*

Definition 3.20. A Markov chain is *ergodic* if it is irreducible, positive recurrent, and aperiodic.

Theorem 3.21 (Ergodic theorems)**.** *Let $(X_n)_{n\geq 0}$ be an ergodic Markov chain on S with n-step transition probabilites $p_{ij}(n)$ and let $T_i = \inf\{n \geq 1 : X_n = i\}$. Let π be the unique invariant distribution.*

- $$p_{ij}(n) \to \pi_j$$
  for all $i, j \in S$ where π.
- $$\frac{1}{n}\sum_{k=0}^{n} 1_{\{X_k=j\}} \to \pi_j \text{ almost surely}$$
  for all $j \in S$, independently of the distribution of X_0.

3.2. **Continuous time Markov chains.**

Definition 3.22. Let $(X_t)_{t\geq 0}$ be a continuous time stochastic process taking values in a countable set S such that $t \mapsto X_t(\omega)$ is right-continuous for almost all $\omega \in \Omega$. Then $(X_t)_{t\geq 0}$ is a Markov chain if

$$\mathbb{P}(X_{t_n} = i_n | X_{t_0} = i_0 \ldots, X_{t_{n-1}} = i_{n-1}) = \mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1})$$

for each $n \geq 1$, and $i_0, \ldots, i_n \in S$. The Markov chain is called *time homogeneous* if

$$\mathbb{P}(X_t = j | X_s = i) = \mathbb{P}(X_{t-s} = j | X_0 = i)$$

for all $0 \leq s \leq t$ and all states $i, j \in S$.

As before, all Markov chains in these notes are time homogeneous.

Notation 3.23. Let $(X_t)_{t\geq 0}$ be a Markov chain on a state space S. Let $p_{ij}(t) = \mathbb{P}_i(X_t = j)$ for $i, j \in S$ denote the transition probabilities, and let $P(t) = (p_{ij}(t))_{i,j\in S}$ denote the $|S| \times |S|$ transition matrix.

Theorem 3.24. *The collection $(P(t))_{t\geq 0}$ of transition matrices for a continuous time Markov chain has the following properties:*

- *For all $t \geq 0$ the matrix $P(t)$ is stochastic.*
- *$P(0) = I$ where I is the $|S| \times |S|$ identity matrix*
- *The Chapman-Kolmogorov equation $P(s+t) = P(s)P(t)$ holds for $s, t \geq 0$. Equivalently, $p_{ik}(s+t) = \sum_{j\in S} p_{ij}(s)p_{jk}(t)$ for all $i, k \in S$.*

Definition 3.25. The collection $(P(t))_{t\geq 0}$ of transition matrices is *uniform* if the following conditions hold:

- There exist finite constants $g_{ij} \geq 0$ such that
$$\lim_{t\downarrow 0} \frac{p_{ij}(t)}{t} = g_{ij}$$
for all $i \neq j$.
- There exist finite constants $g_{ii} \leq 0$ such that
$$g_{ii} = -\sum_{j\neq i} g_{ij}$$
for all i.
- $\inf_i g_{ii} > -\infty$

The matrix $G = (g_{ij})_{i,j\in S}$ is called the *generator* of the Markov process.

All Markov chains considered here have uniform transition matrices.

Theorem 3.26. *Let $(X_t)_{t\geq 0}$ be a Markov chain with generator G, and for each state i, let*
$$U_i = \inf\{t > 0, X_t \neq i\}.$$
Then conditional on $X_0 = i$, the random variable U_i is exponential with rate $-g_{ii}$; that is
$$\mathbb{P}_i(U_i > t) = e^{g_{ii}t}.$$
Furthermore,
$$\mathbb{P}_i(X_U = j) = \frac{g_{ij}}{-g_{ii}}.$$

Theorem 3.27. *Let G be the generator of a Markov chain with transition matrices $(P(t))_{t\geq 0}$. Then for all $t \geq 0$*

- *$\sum_{j\in S} g_{ij} = 0$, or in matrix notation, $G\mathbf{1} = 0$ where $\mathbf{1} = (1, 1, \ldots)$*
- *$P(t) = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n G^n}{n!}$*
- *(Kolmogorov's forward equation) $P'(t) = P(t)G$*
- *(Kolmogorov's backward equation) $P'(t) = GP(t)$*

where $P'(t)$ denotes the matrix $(p'_{ij}(t))_{i,j\in S}$ and p'_{ij} denotes the derivative of p_{ij}.

Definition 3.28. An *invariant distribution* for a Markov chain with transition matrices $(P(t))_{t\geq 0}$ is a row vector $\pi = (\pi_i)_{i\in S}$ such that $\pi P(t) = \pi$ for all $t \geq 0$.

Theorem 3.29. *Let G be the generator of a Markov chain. Then a row vector π is an invariant distribution if and only if*

$$\pi G = 0.$$

Definition 3.30. Let $(p_{ij}(t))_{i,j\in S}$ be the transition probabilities of a Markov chain. The Markov chain is *irreducible* if $p_{ij}(t) > 0$ for all $t > 0$.

Theorem 3.31. *Let $(p_{ij}(t))_{i,j\in S}$ be the transition probabilities of an irreducible Markov chain. If there exists an invariant distribution π, the invariant distribution is unique and*

$$p_{ij}(t) \to \pi_j \text{ as } t \uparrow \infty$$

for all $i, j \in S$. Otherwise, if no invariant distribution exists, then

$$p_{ij}(t) \to 0.$$

4. Martingales

Theorem 4.1. *Let X be an integrable random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sigma-field of $\mathcal{F}$. Then there exists an integrable $\mathcal{G}$-measureable random variable Y such that*

$$\mathbb{E}(1_G Y) = \mathbb{E}(1_G X)$$

for all $G \in \mathcal{G}$. Furthermore, if there exists another $\mathcal{G}$-measureable random variable Y' such that $\mathbb{E}(1_G Y') = \mathbb{E}(1_G X)$ for all $G \in \mathcal{G}$, then $Y = Y'$ almost surely.

Definition 4.2. Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sigma-field. The *conditional expectation* of X given $\mathcal{G}$, written $\mathbb{E}(X|\mathcal{G})$, is a $\mathcal{G}$-measurable random variable with the property that

$$\mathbb{E}\left[1_G \mathbb{E}(X|\mathcal{G})\right] = \mathbb{E}(1_G X)$$

for all $G \in \mathcal{G}$.

Theorem 4.3. *Let X and Y be integrable random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be sigma-algebras, and $a, b \in \mathbb{R}$ be constants.*

- $\mathbb{E}(X|\{\emptyset, \Omega\}) = \mathbb{E}(X)$
- $\mathbb{E}(X|\mathcal{F}) = X$.
- $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$

- *If $X \geq 0$ almost surely, then $\mathbb{E}(X|\mathcal{G}) \geq 0$ almost surely*
- $\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{H}] = \mathbb{E}[\mathbb{E}(X|\mathcal{H})|\mathcal{G}] = \mathbb{E}(X|\mathcal{H})$
- *If Y is independent of $\mathcal{G}$ (the events $\{Y \leq t\}$ and G are indendent for each $t \in \mathbb{R}$ and $G \in \mathcal{G}$) then $\mathbb{E}(Y|\mathcal{G}) = \mathbb{E}(Y)$*
- *If Y is $\mathcal{G}$-measurable (the events $\{Y \leq t\}$ are in $\mathcal{G}$ for each $t \in \mathbb{R}$) then $\mathbb{E}(XY|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G})$*

Theorem 4.4. *Let X be an integrable random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$.*

- *Let $B_1, B_2, \dots$ be a sequence of disjoint non-null events with $\bigcup_n B_n = \Omega$. Let $\mathcal{G}$ be the smallest sigma-field containing $\{B_1, B_2, \dots, ...\}$. Then*
$$\mathbb{E}(X|\mathcal{G}) = \frac{\mathbb{E}(X 1_{B_n})}{\mathbb{P}(B_n)} \text{ if } \omega \in B_n.$$
- *Let Y be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G}$ be the smallest sigma-field containing the events $\{Y \leq t\}$ for all $t \in \mathbb{R}$. Then*
$$\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X|Y)$$
where the right side is defined by Definition 1.18.

Definition 4.5. A *filtration* $(\mathcal{F}_n)_{n\geq 0}$ on Ω is a collection of sigma-fields on Ω such that
$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots.$$

Definition 4.6. A *martingale* relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$ is a stochastic process with the following properties:

- $\mathbb{E}|M_n| < \infty$ for all $n \geq 0$
- $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ for all $n \geq 0$.

Theorem 4.7 (The L_2 martingale convergence theorem)**.** *Let $(M_n)_{n\geq 0}$ be a martingale relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$. If $\sup_{n\geq 0} \mathbb{E}(S_n^2) < \infty$ then there exists a random variable M such that $\mathbb{E}(M^2) < \infty$ and $M_n \to M$ in L_2. Furthermore $M_n = \mathbb{E}(M|\mathcal{F}_n)$.*

Definition 4.8. A *stopping time* for a filtration $(\mathcal{F}_n)_{n\geq 0}$ is a random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ with the property that $\{T = n\} \in \mathcal{F}_n$ for each $n \geq 0$.

Theorem 4.9 (The optional sampling theorem)**.** *Let $(M_n)_{n\geq 0}$ be a martingale relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$, let T a bounded stopping time (there exists a real constant $C > 0$ such that $T(\omega) < C$ for almost all $\omega \in \Omega$.) Then*
$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$

Theorem 4.10 (The optional stopping theorem)**.** *Let $(M_n)_{n\geq 0}$ be a martingale relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$, let T be almost surely finite stopping time. (that is $\mathbb{P}(T<\infty)=1$.) If there is $C>0$ such that*

$$|M_n| \leq C \text{ if } n \leq \tau$$

then

$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$

Theorem 4.11 (The optional stopping theorem)**.** *Let $(M_n)_{n\geq 0}$ be a martingale relative to a filtration $(\mathcal{F}_n)_{n\geq 0}$, let T be almost surely finite stopping time. (that is $\mathbb{P}(T<\infty)=1$.) If $\sup_{n\geq 0}\mathbb{E}(M_n^2)<\infty$ then*

$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$

As an application:

Theorem 4.12. *Let $(S_n, n \geq 0)$ be simple random walk on , i.e., $S_n = \sum_{i=1}^n X_i$ with $X_i = \pm 1$ with probability 1/2, i.i.d. Let $a, b \geq 1$ and let $\tau_a = \inf\{n \geq 1 : X_n = a\}$. Then*

$$\mathbb{P}(\tau_a \leq \tau_{-b}) = \frac{b}{a+b}$$

and if $\tau = \tau_a \wedge \tau_{-b}$, then

$$\mathbb{E}(\tau) = ab$$

$\mathbb{R}$	the set of real numbers
$\overline{\mathbb{R}}$	the set of extended real numbers $\{-\infty\} \cup \mathbb{R} \cup \{\infty\}$
$\mathbb{N}$	the set of natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{C}$	the set of complex numbers
$\mathbb{Z}$	the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
$\mathcal{P}(\Omega)$	the set of all subsets of Ω, called the power set of Ω
A^c	the complement of the set A
$A \subset B$	A is a proper subset of B
$A \subseteq B$	A is a subset of B, and the possibility that $A = B$ is allowed
F_X	the distribution function of a random variable X
p_X	the mass function of a discrete random variable X
f_X	the density function of an absolutely continuous random variable X
$F_{X,Y}$	the joint distribution function of X and Y
$p_{X,Y}$	the joint mass function of X and Y
$f_{X,Y}$	the joint density function of X and Y
$p_{X\|Y}$	the conditional mass function of X given Y
$f_{X\|Y}$	the conditional density of X given Y
G_X	the probability generating function of X
M_X	the moment generating function of X
ϕ_X	the characteristic function of X
$\mathbb{E}(X)$	the expected value of the random variable X
$\mathrm{Var}(X)$	the variance of X
$\mathrm{Cov}(X, Y)$	the covariance of X and Y
$\mathbb{E}(X\|B)$	the conditional expectation of X given the event B
$\mathbb{E}(X\|Y = t)$	the conditional expectation of X given the (possibly null) event $Y = t$
$\mathbb{E}(X\|Y)$	the conditional expectation of X given the random variable Y
$\mathbb{E}_i(Z)$	the conditional expection $\mathbb{E}(Z\|X_0 = i)$, where $(X_n)_{n \geq 0}$ is a Markov chain
$\mathbb{E}(X\|\mathcal{G})$	the conditional expectation of X given the sigma-field $\mathcal{G}$
a^+	$\max\{a, 0\}$
a^-	$\max\{-a, 0\}$
$X \sim \nu$	the random variable X is distributed as the probability measure ν
1_A	the indicator function of the event A
$N(\mu, \sigma^2)$	the normal probability law with mean μ and variance σ^2
$\mathrm{bin}(n, p)$	the binomial probability law with parameters n and p
$\exp(\lambda)$	the exponential probability law with parameter λ
$\mathrm{unif}(a, b)$	the uniform probability law on the interval (a, b)
$\limsup_{n\uparrow\infty} x_n$	the limit superior of the sequence $x_1, x_2, \ldots$
$\liminf_{n\uparrow\infty} x_n$	the limit inferior of the sequence $x_1, x_2, \ldots$
L_p	the set of random variables X with $\mathbb{E}\|X\|^p < \infty$
$\gcd(A)$	the greatest common divisor of the set $A \subseteq \mathbb{N}$

TABLE 1. Notation