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 \(F\) of subsets of Ω such that

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

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

Definition 1.2. Let Ω be a set and let \(F\) be a sigma-field on Ω. A *probability measure* on \(F\) is a function \(P : F \to [0,1]\) such that

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

[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, \(F\) a sigma-field on Ω, and \(P\) a probability measure on \(F\).

The triple \((\Omega, F, P)\) is called a *probability space*.

The set \(\Omega\) is called the *sample space*, and an element of \(\Omega\) is called an *outcome*. A subset of \(\Omega\) which is an element of \(F\) is called an *event*.

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

1.2. Random variables and distribution functions.

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

Let \(A\) be a subset of \(\mathbb{R}\), 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 \(S \subseteq \mathbb{R}\) if \(X \in S\) almost surely.

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

\[
F_X(t) = P(X \leq t)
\]

for all \(t \in \mathbb{R}\).

[A distribution function is called *defective* if either \(\lim_{t \to \infty} F_X(t) < 1\) or \(\lim_{t \to -\infty} F_X(t) > 0\). Unless otherwise indicated, all random variables considered here are assumed to have non-defective distribution functions.]
The law of $X$ is the probability measure $\nu$ 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 $\nu$” is written $X \sim \nu$. [The measure $\nu$ 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 $\Omega$. 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

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

  where $n \in \mathbb{N}$ and $0 \leq p \leq 1$.

- **Poisson** with parameter $\lambda$ if

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

  where $\lambda \geq 0$.

- **geometric** with parameter $p$ if

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

  where $0 \leq p \leq 1$.

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

**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 $X$ be 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 $\mu$ and variance $\sigma^2$, written $X \sim \mathcal{N}(\mu, \sigma^2)$, if
  \[ f_X(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \text{ for all } t \in \mathbb{R} \]
  for some $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- **exponential** with rate $\lambda$, written $X \sim \text{exp}(\lambda)$, 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 $E(X)$ and is given by the integral
\[ E(X) = \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \]

The above integral is defined in the following cases:
- if $X \geq 0$ almost surely.
- if either $E(X^+)$ or $E(X^-)$ is finite, in which case $E(X) = E(X^+) - E(X^-)$.

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

The **variance** of an integrable random variable $X$, written $\text{Var}(X)$, is
\[ \text{Var}(X) = E(X^2) - E(X)^2. \]

The **covariance** of square-integrable random variable $X$ and $Y$, written $\text{Cov}(X, Y)$, is
\[ \text{Cov}(X, Y) = E(XY) - E(X)E(Y). \]

If neither $X$ or $Y$ is almost surely constant, then their correlation, written $\rho(X, Y)$, is
\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{Var}(X)^{1/2}\text{Var}(Y)^{1/2}}. \]

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

**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
\[ E(g(X)) = \sum_{t \in S} g(t) \ p_X(t). \]
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
\[
\mathbb{E}(X|B) = \frac{\mathbb{E}(X 1_B)}{\mathbb{P}(B)}.
\]

**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 $F_{X,Y}: \mathbb{R}^2 \rightarrow [0, 1]$ 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 $p_{X,Y}: \mathbb{R}^2 \rightarrow [0, 1]$ 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 $f_{X,Y}: \mathbb{R}^2 \rightarrow [0, \infty)$ 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
\[
f_{X|Y}(s|t) = \lim_{\delta \downarrow 0, \epsilon \downarrow 0} \frac{1}{\delta} \mathbb{P}(|X - s| < \delta \mid |Y - t| < \epsilon) = \frac{f_{X,Y}(s, t)}{f_Y(t)}.
\]
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 $\mu_X$ and $\mu_Y$, variances $\sigma_X^2$ and $\sigma_Y^2$, and correlation $\rho$, written $(X,Y) \sim N((\mu_X, \mu_Y), (\sigma_X^2, \sigma_Y^2, \rho))$ if the joint density function is
\[
f_{X,Y}(s,t) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} Q(s,t)\right)
\]
where
\[
Q(s,t) = \frac{1}{1-\rho^2} \left(\frac{(s-\mu_X)^2}{\sigma_X^2} - 2\rho \frac{(s-\mu_X)(t-\mu_Y)}{\sigma_X \sigma_Y} + \frac{(t-\mu_Y)^2}{\sigma_Y^2}\right)
\]

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
\[
\mathbb{E}(X|Y = t) = \lim_{\epsilon \downarrow 0} \frac{\mathbb{E}(X \mathbb{1}_{\{|Y-t| < \epsilon\}})}{\mathbb{P}(|Y-t| < \epsilon)}
\]
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
\[
P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)
\]
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 \text{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
\[
E(XY) = E(X)E(Y).
\]

1.5. Probability inequalities.

Theorem 1.23 (Markov’s inequality). Let $X$ be a positive random variable. Then
\[
P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}
\]
for all $\epsilon > 0$.

Corollary 1.24 (Chebychev’s inequality). Let $X$ be a random variable with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$. Then
\[
P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}
\]
for all $\epsilon > 0$.

1.6. Generating and characteristic functions.

Definition 1.25. Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$ taking values in \{0, 1, 2, \ldots\}. The **probability generating function** of $X$ is the function $G_X : [0, 1] \rightarrow [0, 1]$ defined by
\[
G_X(t) = E(t^X)
\]
for all $t \in (0, 1]$ and $G_X(0) = P(X = 0)$.

Theorem 1.26. Let $G_X$ be the probability generating function of $X$. Then
\[
P(X = n) = \frac{1}{n!} G_X^{(n)}(0)
\]
for all $n \in \mathbb{N}$, where $G_X^{(n)}(0)$ denotes the $n$-th derivative of $G_X$ evaluated at 0.

Definition 1.27. Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$. The **moment generating function** of $X$ is the function $M_X : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ defined by
\[
M_X(t) = E(e^{tX})
\]
for all $t \in \mathbb{R}$. 

6
Theorem 1.28. Let $M_X$ be the moment-generating function 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 $M_X$ evaluated 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 $\phi_X$ and $\phi_Y$ be the characteristic functions of $X$ and $Y$. Then $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbb{R}$ if and only if $F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}$.

- Let $X$ and $Y$ be valued in $\mathbb{N}$ with probability generating functions $G_X$ and $G_Y$. Then $G_X(t) = G_Y(t)$ for all $t \in [0, 1]$ if and only if $F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}$.

- Let $M_X$ and $M_Y$ be the moment generating functions of $X$ and $Y$. If there exists an $\epsilon > 0$ such that $M_X(t) = M_Y(t) < \infty$ for all $t \in (-\epsilon, \epsilon)$ then $F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}$.

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

$$\limsup_{n \to \infty} x_n = \inf_N \sup_{n \geq N} x_n$$

and limit inferior by

$$\liminf_{n \to \infty} x_n = \sup_N \inf_{n \geq N} x_n.$$ 

Equivalently, if $x \in \mathbb{R}$ then

$$x = \limsup_{n \to \infty} x_n \iff \text{for all } \epsilon > 0 \left\{ \begin{array}{ll} \{n \in \mathbb{N} : x_n > x + \epsilon\} & \text{is finite} \\ \{n \in \mathbb{N} : x_n > x - \epsilon\} & \text{is infinite} \end{array} \right.$$ 

and

$$x = \liminf_{n \to \infty} x_n \iff \text{for all } \epsilon > 0 \left\{ \begin{array}{ll} \{n \in \mathbb{N} : x_n < x + \epsilon\} & \text{is infinite} \\ \{n \in \mathbb{N} : x_n < x - \epsilon\} & \text{is finite} \end{array} \right.$$ 

If $\limsup_{n \to \infty} x_n = \liminf_{n \to \infty} x_n = x$ then the sequence $x_1, x_2, \ldots$ is convergent and with limit $x = \lim_{n \to \infty} x_n$. 

7
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 $\mathbb{E}|X_n - X|^p \to 0$ for all $\epsilon > 0$
- $X_n \to X$ 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:

$$X_n \to X \text{ almost surely} \quad \text{or} \quad X_n \to X \text{ in } L_p, p \geq 1 \quad \Rightarrow \quad X_n \to X \text{ in probability} \quad \Rightarrow \quad X_n \to X \text{ in distribution}$$

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

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

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 $X_n \leq X_{n+1}$ almost surely for all $n \geq 1$, 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

$$\mathbb{E}(\liminf_{n \to \infty} X_n) \leq \liminf_{n \to \infty} \mathbb{E}(X_n).$$
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 $\mathbb{E}(\sup_{n \geq 1} |X_n|) < \infty$ 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 \textit{weak law of large numbers} if $S_n/n \to \mu$ in probability for some constant $\mu$. The collection satisfies a \textit{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 identically distributed with
$$
N\mathbb{P}(|X_i| > N) \to 0 \text{ and } \mathbb{E}(X 1_{\{|X_i| \leq N\}}) \to \mu \text{ as } N \to \infty
$$
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 $\phi$ 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 $\text{Var}(X_i) = \sigma^2$ 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 \textit{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 \textit{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 \textit{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 \textit{Markov chain} if
$$
\mathbb{P}(X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1})
$$
for each $n \geq 1$, and $s_0, \ldots, s_n \in S$. The set $S$ is called the \textit{state space} of the Markov chain, and an elements of $S$ is called a \textit{state}. The Markov chain is called \textit{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 \(\lambda_k\) 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 \(h^A_i = \mathbb{P}_i(H^A < \infty)\). Then \((h^A_i)_{i \in S}\) is the minimal non-negative solution to
\[
h^A_i = \begin{cases} 
1 & \text{if } i \in A \\
\frac{1}{\sum_{j \in S} p_{ij} h^A_j} & \text{otherwise}
\end{cases}
\]
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 \( i \rightarrow j \), if there exists an \( n \geq 0 \) such that \( p_{ij}(n) > 0 \). States \( i \) and \( j \) communicate, written \( i \leftrightarrow j \), if \( i \rightarrow j \) and \( j \rightarrow i \). 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 \( i \rightarrow j \) 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 \( i \rightarrow j \).

A Markov chain is irreducible if all states in \( S \) communicate.

Definition 3.9. A state \( i \in S \) is called recurrent if
\[
P_i(X_n = i \text{ infinitely often}) = 1,
\]
and called transient if
\[
P_i(X_n = i \text{ infinitely often}) < 1.
\]

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 \( f : S^n \rightarrow \{0, 1\} \) such that
\[
\mathbb{1}_{\{T = n\}} = f(X_0, \ldots, X_n).
\]

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.
- \( \sum_{n=1}^{\infty} p_{ii}(n) = \infty \).
- \( P_i(T_i < \infty) = 1 \)

Moreover, the following are equivalent:

- State \( i \) is transient.
- \( P_i(X_n = i \text{ infinitely often}) = 0 \).
- \( P_i(T_i < \infty) < 1 \).
- \( \sum_{n=1}^{\infty} p_{ii}(n) = \frac{1}{1-P_i(T_i < \infty)} < \infty \).

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 \( E_i(T_i) < \infty \) 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 and 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 
\[
\gamma_i^{(k)} = \frac{1}{\mathbb{E}_k(T_k)} \sum_{n=1}^{T_k} 1\{X_n = i\}
\]

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 \( \pi \) 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 probabilities \( p_{ij}(n) \) and let \( T_i = \inf\{n \geq 1 : X_n = i\} \). Let \( \pi \) be the unique invariant distribution.
\[ p_{ij}(n) \to \pi_j \]
for all \( i, j \in S \) where \( \pi \).

\[ \frac{1}{n} \sum_{k=0}^{n} \mathbb{1}_{\{X_k=j\}} \to \pi_j \text{ almost surely} \]
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 
\[ P_i(U_i > t) = e^{-g_{ii}t}. \]
Furthermore, 
\[ P_i(X_U = j) = \frac{g_{ij}}{-g_{ii}}. \]

Theorem 3.27. Let \(G\) be the generator of a Markov chain with transition matrices \((P(t))_{t \geq 0}\). Then for all \(t \geq 0\)
\[ \sum_{j \in S} g_{ij} = 0, \text{ or in matrix notation, } G \mathbf{1} = 0 \text{ where } \mathbf{1} = (1, 1, \ldots) \]
\[ P(t) = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n G^n}{n!} \]
\[ (\text{Kolmogorov’s forward equation}) \quad P'(t) = P(t)G \]
\[ (\text{Kolmogorov’s backward equation}) \quad P'(t) = GP(t) \]
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 
\[ \pi P(t) = \pi \text{ for all } t \geq 0. \]

Theorem 3.29. Let \(G\) be the generator of a Markov chain. Then a row vector \(\pi\) 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 \(\pi\), 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 \(\mathcal{G}\)-measurable 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 \(\mathcal{G}\)-measurable random variable \(Y'\) such that 
\[ \mathbb{E}(1_G Y') = \mathbb{E}(1_G X) \text{ for all } G \in \mathcal{G}, \]
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 
\[ \mathbb{E}[1_G \mathbb{E}(X|\mathcal{G})] = \mathbb{E}(1_G X) \]
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 independent for each $t \in \mathbb{R}$ and $G \in \mathcal{G}$) 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 $B_1, B_2, \ldots$ be a sequence of disjoint non-null events with $\bigcup_n B_n = \Omega$. Let $\mathcal{G}$ be the smallest sigma-field containing $\{B_1, B_2, \ldots, \}\}$. Then
  $$\mathbb{E}(X|\mathcal{G}) = \frac{\mathbb{E}(X1_{B_n})}{\mathbb{P}(B_n)} \text{ if } \omega \in B_n.$$  

- 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 $\Omega$ is a collection of sigma-fields on $\Omega$ such that
$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots.$$  

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 $\sup_{n \geq 0} \mathbb{E}(M_n^2) < \infty$ then
$$\mathbb{E}(M_T) = \mathbb{E}(M_0).$$
\( \mathbb{R} \) 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 \( \Omega \), called the power set of \( \Omega \)
\( A^c \) 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 \( X \)
\( \phi_X \) the characteristic function of \( X \)

\( E(X) \) the expected value of the random variable \( X \)
\( \text{Var}(X) \) the variance of \( X \)
\( \text{Cov}(X,Y) \) the covariance of \( X \) and \( Y \)
\( E(X|B) \) the conditional expectation of \( X \) given the event \( B \)
\( E(X|Y = t) \) the conditional expectation of \( X \) given the (possibly null) event \( Y = t \)
\( E(X|Y) \) the conditional expectation of \( X \) given the random variable \( Y \)
\( E_i(Z) \) the conditional expectation \( E(Z|X_0 = i) \), where \((X_n)_{n \geq 0}\) is a Markov chain
\( E(X|\mathcal{G}) \) 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 \( \nu \)
\( 1_A \) the indicator function of the event \( A \)
\( N(\mu, \sigma^2) \) the normal probability law with mean \( \mu \) and variance \( \sigma^2 \)
\( \text{bin}(n, p) \) the binomial probability law with parameters \( n \) and \( p \)
\( \exp(\lambda) \) the exponential probability law with parameter \( \lambda \)
\( \text{unif}(a, b) \) the uniform probability law on the interval \((a, b)\)

\( \limsup_{n \to \infty} x_n \) the limit superior of the sequence \( x_1, x_2, \ldots \)
\( \liminf_{n \to \infty} x_n \) the limit inferior of the sequence \( x_1, x_2, \ldots \)
\( L_p \) the set of random variables \( X \) with \( E|X|^p < \infty \)
\( \gcd(A) \) the greatest common divisor of the set \( A \subseteq \mathbb{N} \)

**Table 1. Notation**