# Lecture Notes on Probability 

By
Omer Tamuz
California Institute of Technology

## Contents

1 Why we need measure theory ..... 6
1.1 Probability measures on countable spaces ..... 6
1.2 Probability measures on uncountable spaces ..... 6
1.3 Vitali sets ..... 7
2 Measure theory ..... 8
$2.1 \pi$-systems ..... 8
2.2 Algebras ..... 8
2.3 Sigma-algebras ..... 9
3 Constructing probability measures ..... 10
3.1 The Hahn-Kolmogorov Theorem ..... 10
3.2 Basic properties of probability measures ..... 10
3.3 Cumulative distribution functions ..... 11
4 Events and random variables ..... 13
4.1 Events ..... 13
4.2 Sub-sigma-algebras ..... 13
4.3 Random variables ..... 14
5 Independence and the Borel-Cantelli Lemmas ..... 16
5.1 Independence ..... 16
5.2 A question about a sequence of random variables ..... 16
5.3 The Borel-Cantelli Lemmas ..... 17
6 The tail sigma-algebra ..... 19
6.1 Motivating example ..... 19
6.2 The tail sigma-algebra ..... 19
6.3 The zero-one law ..... 20
7 Expectations ..... 22
7.1 Expectations in finite probability spaces ..... 22
7.2 Expectations of non-negative random variables ..... 22
7.3 Markov's inequality ..... 23
7.4 Pointwise convergence and convergence of expectations ..... 23
$7.5 \mathscr{L}^{p}$ ..... 24
$7.6 \mathscr{L}^{2}$ ..... 24
8 A strong law of large numbers and the Chernoff bound ..... 26
8.1 Expectation of a product of independent random variables ..... 26
8.2 Jensen's inequality ..... 26
8.3 SLLN in $\mathscr{L}^{4}$ ..... 26
8.4 The Chernoff bound ..... 27
9 The weak law of large numbers ..... 29
$9.1 \mathscr{L}^{2}$ ..... 29
$9.2 \mathscr{L}^{1}$ ..... 29
10 Conditional expectations ..... 31
10.1 Why things are not as simple as they seem ..... 31
10.2 Conditional expectations in finite spaces ..... 31
10.3 Conditional expectations in $\mathscr{L}^{2}$ ..... 32
10.4 Conditional expectations in $\mathscr{L}^{1}$ ..... 32
10.5 Some properties of conditional expectation ..... 33
11 The Galton-Watson process ..... 34
11.1 Definition ..... 34
11.2 The probability of extinction ..... 34
11.3 The probability generating function ..... 35
12 Markov chains ..... 37
12.1 Definition ..... 37
12.2 Irreducibility and aperiodicity ..... 37
12.3 Recurrence ..... 38
12.4 The simple random walk on $\mathbb{Z}^{d}$ ..... 39
13 Martingales ..... 41
13.1 Definition ..... 41
13.2 Examples ..... 41
13.3 Martingale convergence in $\mathscr{L}^{2}$ ..... 42
13.4 The Martingale Convergence Theorem ..... 42
14 Stopping times ..... 45
14.1 Definition ..... 45
14.2 Optional Stopping Time Theorem ..... 45
15 Harmonic and superharmonic functions ..... 47
15.1 Definition ..... 47
15.2 Harmonic functions and martingales ..... 47
15.3 Superharmonic functions and recurrence ..... 47
15.4 Bounded harmonic functions ..... 48
15.5 The shift-invariant sigma-algebra ..... 48
16 The Choquet-Deny Theorem ..... 50
16.1 The asymptotic direction of a random walk ..... 50
16.2 The Krein-Milman Theorem ..... 50
17 Basics of information theory ..... 52
17.1 Shannon entropy ..... 52
17.2 Conditional Shannon entropy ..... 53
17.3 Mutual information ..... 53
17.4 The information processing inequality ..... 54
18 Random walks on groups ..... 55
18.1 Finitely generated groups ..... 55
18.2 Random walks on finitely generated groups ..... 55
18.3 Random walk entropy ..... 55
18.4 The Kaimanovich-Vershik Theorem ..... 56
19 Characteristic functions and the Central Limit Theorem ..... 58
19.1 Convergence in distribution ..... 58
19.2 The characteristic function ..... 58
19.3 The characteristic function of normalized i.i.d. sums ..... 59
20 The Radon-Nikodym derivative and absolute continuity ..... 61
20.1 The Radon-Nikodym derivative ..... 61
20.2 Absolute continuity ..... 61
20.3 The Radon-Nikodym Theorem ..... 62
21 Large deviations ..... 64
21.1 The cumulant generating function ..... 64
21.2 The Legendre transform ..... 65
21.3 Large deviations ..... 65
22 Stationary distributions and processes ..... 68
23 Stationary processes and measure preserving transformations ..... 69
24 The Ergodic Theorem ..... 71
25 The weak topology and the simplex of invariant measures ..... 74
26 Percolation ..... 77
27 The mass transport principle ..... 79
28 Majority dynamics ..... 82
29 Scenery Reconstruction: I ..... 84
30 Scenery reconstruction: II ..... 86

## Disclaimer

This a not a textbook. These are lecture notes.

## 1 Why we need measure theory

### 1.1 Probability measures on countable spaces

We usually think of a probability measure $\mu$ on a countable set of outcomes $\Omega$ as an assignment to each $\omega \in \Omega$ of a number between 0 and 1 , with the property that these numbers sum to 1 . In this course we will think about it as a function $\mu: 2^{\Omega} \rightarrow[0,1]$ (assigning to each subset of $\Omega$ a number in $[0,1])$ with the following two properties:

1. Unit mass. $\mu(\Omega)=1$.
2. sigma-additivity. if ( $\left.A_{1}, A_{2}, \ldots\right)$ is a sequence of disjoint sets then

$$
\mu\left(\cup_{n} A_{n}\right)=\sum_{n} \mu\left(A_{n}\right) .
$$

For example, let $\Omega=\mathbb{N}=\{1,2,3, \ldots\}$, and define $\mu$ by $\mu(\{n\})=2^{-n}$. This is the distribution of the number of tosses of a fair coin until the first heads. If $E$ is the set of even numbers, then

$$
\mu(E)=\sum_{n=1}^{\infty} 2^{-2 n}=\frac{1}{4}+\frac{1}{16}+\frac{1}{64}=\frac{1}{3} .
$$

As another example, when $\Omega=\{0,1\}^{n}$, the i.i.d. fair coin toss measure can be defined by letting, for each $k \leq n$ and $a \in\{0,1\}^{k}$

$$
\begin{equation*}
\mu\left(\left\{\omega: \omega_{1}=a_{1}, \ldots, \omega_{k}=a_{k}\right\}\right)=2^{-k} . \tag{1.1}
\end{equation*}
$$

This measure has a rich symmetry structure. For $k \leq n$, let $\sigma_{k}: \Omega \rightarrow \Omega$ be the map that "flips" the $k^{\text {th }}$ entry:

$$
\sigma_{k}\left(\omega_{1}, \ldots, \omega_{n}\right)=\left(\omega_{1}, \ldots, \omega_{k-1}, 1-\omega_{k}, \omega_{k+1}, \ldots, \omega_{n}\right) .
$$

Denote by $\sigma_{k}(A)$ the $\operatorname{set}\left\{\sigma_{k}(\omega): \omega \in A\right\}$. Then

$$
\begin{equation*}
\mu\left(\sigma_{k}(A)\right)=\mu(A) \text { for all } k \text { and } A \subseteq \Omega \tag{1.2}
\end{equation*}
$$

That is, if we flip the $k^{\text {th }}$ entry in each element of a set of outcomes $A$, then we do not change its probability. In fact, it is easy to show that (1.2) implies (1.1), and thus can be taken to be the definition of $\mu$.

### 1.2 Probability measures on uncountable spaces

We would like to define the same object for a countable number of coin tosses, which makes for an uncountable set of outcomes. That is, when $\Omega=\{0,1\}^{\mathbb{N}}$, we would like to define a map $\mu: 2^{\Omega} \rightarrow[0,1]$ that satisfies unit mass and countable additivity, and additionally satisfies
(1.2). Here, $\sigma_{k}: \Omega \rightarrow \Omega$ flips the $k^{\text {th }}$ entry in the infinite sequence $\left(\omega_{1}, \omega_{2}, \ldots\right) \in \Omega$.

It turns out that no such measure exists. The next section explains why.

### 1.3 Vitali sets

Say that $\theta \in \Omega$ is basically the same as $\omega \in \Omega$ if there is some $N$ such that $\omega_{n}=\theta_{n}$ for all $n \geq N$. That is, if $\omega$ and $\theta$ agree in all but finitely many entries. Equivalently, $\theta$ is basically the same as $\omega$ if there is a finite set $\bar{k}=\left\{k_{1}, k_{2}, \ldots, k_{n}\right\} \subset \mathbb{N}$ such that $\theta=\sigma_{k_{n}} \circ \sigma_{k_{n-1}} \circ \cdots \circ \sigma_{k_{1}}(\omega)$. For example, $\theta$ is basically the same as $(0,0,0, \ldots)$ iff $\theta_{k}=1$ for only finitely many $k$.

Note that if $\theta$ is basically the same as $\omega$ then $\omega$ is basically the same as $\theta$. Note also that if $\omega$ is basically the same as $\theta$, and if $\theta$ is basically the same as $\zeta$, then $\omega$ is basically the same as $\zeta$. This makes "basically the same" an equivalence relation. Therefore, if we denote equivalence classes by

$$
[\omega]=\{\theta \text { that are basically the same as } \omega\}
$$

then $[\omega]=[\theta]$ iff $\omega$ and $\theta$ are basically the same, and the collection of equivalence classes forms a partition of $\Omega$. Denote by $K$ the (countable) set of finite subsets $\bar{k}=\left\{k_{1}, \ldots, k_{n}\right\} \subset \mathbb{N}$, and denote $\sigma_{\bar{k}}=\sigma_{k_{n}} \circ \cdots \circ \sigma_{k_{1}}$. Then

$$
[\omega]=\left\{\sigma_{\bar{k}}(\omega): \bar{k} \in K\right\}
$$

and so each equivalence class is countable.
Let $V$ be a set of representatives of these equivalence classes. That is, $V$ contains for each equivalence class $[\omega]$ a single element of $[\omega]$. Then $\sigma_{\bar{k}}(V) \neq \sigma_{\bar{\ell}}(V)$ whenever $\bar{k} \neq \bar{\ell}$, and

$$
\cup_{\bar{k} \in K} \sigma_{\bar{k}}(V)=\Omega
$$

Suppose by way of contradiction that there is a $\mu$ that has all of our desired properties. Then, by sigma-additivity and unit mass,

$$
\mu\left(\cup_{\bar{k} \in K} \sigma_{\bar{k}}(V)\right)=\mu(\Omega)=1
$$

Since $\mu\left(\sigma_{\bar{k}}(V)\right)=\mu(V)$, then the r.h.s. of the above equation is zero if $\mu(V)=0$, and infinite if $\mu(V)>0$. We have thus reached a contradiction, and no such $\mu$ exists.

## 2 Measure theory

## $2.1 \pi$-systems

Given a set $\Omega$, a $\pi$-system on $\Omega$ is a collection $\mathscr{P}$ of subsets of $\Omega$ such that if $A, B \in \mathscr{P}$ then $A \cap B \in \mathscr{P}$.

Example 2.1. Let $\Omega=\mathbb{R}$, and let

$$
\mathscr{P}=\{(-\infty, x]: x \in \mathbb{R}\} .
$$

This is a $\pi$-system because $(-\infty, x] \cap(-\infty, y]=(-\infty, \min \{x, y\}]$.
Example 2.2. Let $\Omega=\{0,1\}^{\mathbb{N}}$, and let $\mathscr{P}$ be the collection of sets $\left\{A_{S}\right\}$ indexed by finite $S \subset \mathbb{N}$ where

$$
A_{S}=\left\{\omega \in \Omega: \omega_{k}=1 \text { for all } k \in S\right\}
$$

This is a $\pi$-system because $A_{S} \cap A_{T}=A_{S \cup T}$.
Example 2.3. Let $X$ be a topological space. Then the set of closed sets in $X$ is a $\pi$-system.

### 2.2 Algebras

An algebra of subsets of $\Omega$ is a $\pi$-system $\mathscr{A}$ on $\Omega$ with the following additional properties:

1. $\Omega \in \mathscr{A}$.
2. If $A \in \mathscr{A}$ then its complement $A^{\mathrm{c}} \in \mathscr{A}$.

It is easy to see that if $\mathscr{A}$ is an algebra of subsets of $\Omega$ then

1. $\varnothing \in \mathscr{A}$.
2. If $A, B \in \mathscr{A}$ then $A \cup B \in \mathscr{A}$.

Example 2.4. Let $\Omega$ be any set. Then the collection of subsets of $\Omega$ is an algebra.
Example 2.5. Let $\Omega=\{0,1\}^{\mathbb{N}}$, and let $\mathscr{A}_{\text {clopen }}$ be the algebra of clopen sets. That is, $\mathscr{A}_{\text {clopen }}$ is the collection of finite unions of sets $A_{x}$ indexed by finite $x \in\{0,1\}^{n}$, where

$$
A_{x}=\left\{\omega \in \Omega: \omega_{k}=x_{k} \text { for all } k \leq n\right\} .
$$

Exercise 2.6. Show that $\mathscr{A}_{\text {clopen }}$ is the collection of finite disjoint unions of sets of the form $A_{x}$.

Example 2.7. Let $\Omega=\mathbb{N}$, and let $\mathscr{A}_{\infty}$ be the collection of sets $A$ such that either $A$ is finite, or else $A^{\mathrm{c}}$ is finite.

Exercise 2.8. Prove that $\mathscr{A}_{\text {clopen }}$ and $\mathscr{A}_{\infty}$ are algebras.
Given an algebra $\mathscr{A}$, a finitely additive probability measure is a function $\mu: \mathscr{A} \rightarrow[0,1]$ with the following properties:

1. $\mu(\Omega)=1$.
2. $\mu$ is additive. That is, if $A_{1}, A_{2}$ are disjoint (i.e., $A_{1} \cap A_{2}=\varnothing$ ) then

$$
\mu\left(A_{1} \cup A_{2}\right)=\mu\left(A_{1}\right)+\mu\left(A_{2}\right)
$$

Exercise 2.9. Show that $\mu(\varnothing)=0$.
Exercise 2.10. Define a finitely additive measure on the algebra $\mathscr{A}_{\infty}$ from Example 2.7.

### 2.3 Sigma-algebras

An algebra $\mathscr{F}$ of subsets of $\Omega$ is a sigma-algebra if for any sequence ( $A_{1}, A_{2}, \ldots$ ) of elements of $\mathscr{F}$ it holds that $\cup_{n} A_{n} \in \mathscr{F}$. It follows that $\cap_{n} A_{n} \in \mathscr{F}$.

Exercise 2.11. 1. Let I be a set, and let $\left\{\mathscr{F}_{i}\right\}_{i \in I}$ be a collection of sigma-algebras of subsets of $\Omega$. Show that $\cap_{i \in I} \mathscr{F}_{i}$ is a sigma-algebra.
2. Let $\mathscr{C}$ be a collection of subsets of $\Omega$. Then there exists a unique minimal (under inclusion) sigma-algebra $\mathscr{F} \supseteq \mathscr{C}$. $\mathscr{F}$ is called the sigma-algebra generated by $\mathscr{C}$, which we write as $\mathscr{F}=\sigma(\mathscr{C})$.

Exercise 2.12. Prove that $\mathscr{A}_{\infty}$ (Example 2.7) is not a sigma-algebra.
Given a topological space, the Borel sigma-algebra $\mathscr{B}$ is the sigma-algebra generated by the open sets. Hence it is also generated by any basis of the topology.

A measurable space is a pair $(\Omega, \mathscr{F})$, where $\mathscr{F}$ is a sigma-algebra of subsets of $\Omega$. A probability measure on $(\Omega, \mathscr{F})$ is a function $\mu: \mathscr{F} \rightarrow[0,1]$ with the following properties:

1. $\mu(\Omega)=1$.
2. $\mu$ is sigma additive. That is, if $\left(A_{1}, A_{2}, \ldots\right)$ is a sequence of disjoint sets (i.e., $A_{n} \cap A_{m}=\varnothing$ for all $n \neq m$ ) then

$$
\mu\left(\cup_{n} A_{n}\right)=\sum_{n} \mu\left(A_{n}\right) .
$$

## 3 Constructing probability measures

### 3.1 The Hahn-Kolmogorov Theorem

Theorem 3.1 (Hahn-Kolmogorov Theorem). Let $\mathscr{C}$ be a collection of subsets of $\Omega$, and let $\mathscr{F}=\sigma(\mathscr{C})$. Let $\mu_{0}: \mathscr{C} \rightarrow[0,1]$ be a countably additive map with $\mu(\Omega)=1$. We say that a probability measure $\mu: \mathscr{F} \rightarrow[0,1]$ extends $\mu_{0}$ if $\mu(A)=\mu_{0}(A)$ for all $A \in \mathscr{C}$.

1. If $\mathscr{C}$ is $a \pi$-system then there exists at most one probability measure $\mu$ that extends $\mu_{0}$.
2. If $\mathscr{C}$ is an algebra then there exists exactly one probability measure $\mu$ that extends $\mu_{0}$.

Example 3.2. Let $\mathscr{A}=\mathscr{A}_{\text {clopen }}$ be the algebra defined in Example 2.5. Then there is a unique map $\mu_{0}: \mathscr{A} \rightarrow[0,1]$ that is additive and satisfies

$$
\mu_{0}\left(A_{x}\right)=2^{-|x|} .
$$

Furthermore, this map is countably additive.
Hence $\mu_{0}$ has a unique extension $\mu: \mathscr{B} \rightarrow[0,1]$ (where $\mathscr{B}=\sigma(\mathscr{A})$ is the Borel sigma-algebra on $\{0,1\}^{\mathbb{N}}$, equipped with the product topology).

The probability measure $\mu$ is sometimes called the Bernoulli measure on $\{0,1\}^{\mathbb{N}}$.
Exercise 3.3. Prove that $\mu_{0}: \mathscr{A}_{\text {clopen }} \rightarrow[0,1]$ is countably additive.
Example 3.4. Let $\mathscr{P}$ be the $\pi$-system on the interval $[0,1]$ given by

$$
\mathscr{P}=\{[0, x]: x \in[0,1]\},
$$

and let and let $\mu_{0}: \mathscr{P} \rightarrow[0,1]$ be given by $\mu_{0}([0, x])=x$. Then there exists a probability measure $\mu: \mathscr{B} \rightarrow[0,1]$ (where $\mathscr{B}=\sigma(\mathscr{C})$ is the Borel sigma-algebra on $[0,1]$ ) that extends $\mu_{0}$.

Note that indeed there always exists such a $\mu$; it is called the Lebesgue measure. To prove this we naturally extend $\mu_{0}$ to the algebra generated by $\mathscr{P}$, and then show that this extension is countably additive.

### 3.2 Basic properties of probability measures

Theorem 3.5. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space.

1. If $\left(F_{1}, F_{2}, \ldots\right)$ be a sequence of sets in $\mathscr{F}$ such that $F_{n} \subseteq F_{n+1}$ then

$$
\mu\left(\cup_{n} F_{n}\right)=\lim _{n} \mu\left(F_{n}\right) .
$$

2. If $\left(F_{1}, F_{2}, \ldots\right)$ be a sequence of sets in $\mathscr{F}$ such that $F_{n} \supseteq F_{n+1}$ then

$$
\mu\left(\cap_{n} F_{n}\right)=\lim _{n} \mu\left(F_{n}\right) .
$$

Proof. 1. Let $G_{1}=F_{1}$, and for $n>1$ let $G_{n}=F_{n} \backslash F_{n-1}$. Then $\cup_{k=1}^{n} G_{n}=F_{n}, \cup_{n} F_{n}=\cup_{n} G_{n}$, and additionally the $G_{n}$ 's are disjoint. Hence

$$
\mu\left(\cup_{n} F_{n}\right)=\mu\left(\cup_{n} G_{n}\right)=\sum_{n} \mu\left(G_{n}\right)=\lim _{n} \sum_{k=1}^{n} \mu\left(G_{n}\right)=\lim _{n} \mu\left(\cup_{k=1}^{n} G_{n}\right)=\lim _{n} \mu\left(F_{n}\right) .
$$

2. Left as an exercise.

Corollary 3.6. Let $(\Omega, \mathscr{F}, \mu)$ be a probability space, and let $\left(F_{1}, F_{2}, \ldots\right)$ be a sequence of sets in $\mathscr{F}$.

1. If $\mu\left(F_{n}\right)=0$ for all $n$ then

$$
\mu\left(\cup_{n} F_{n}\right)=0 .
$$

2. If $\mu\left(F_{n}\right)=1$ for all $n$ then

$$
\mu\left(\cap_{n} F_{n}\right)=1
$$

### 3.3 Cumulative distribution functions

Let $(\mathbb{R}, \mathscr{B}, \mu)$ be a probability space. The cumulative distribution function (c.d.f.) associated with $\mu$ is

$$
F(x)=\mu((-\infty, x]) .
$$

Claim 3.7. The following holds for any cumulative distribution function $F$ :

1. $F$ is monotone non-decreasing.
2. $\sup _{x} F(x)=1$.
3. $\inf _{x} F(x)=0$.
4. Fis right-continuous.

Proof. 1. For any $x>y$ by the additivity of $\mu$ we have that

$$
F(y)=\mu((-\infty, y])=\mu((-\infty, x] \cup(x, y])=\mu((-\infty, x])+\mu((x, y))=F(x)+\mu((x, y]) \geq F(x) .
$$

2. Let $E_{n}=(-\infty, n)$. Then $E_{n} \subset E_{n+1}$ and $\cup_{n} E_{n}=\mathbb{R}$. Hence by Theorem 3.5

$$
\lim _{n} F(n)=\lim _{n} \mu\left(E_{n}\right)=\mu(\mathbb{R})=1 .
$$

Since $F$ is monotone non-decreasing it follows that $\sup _{x} F(x)=1$.
3. The proof of this is similar to the proof that $\sup _{x} F(x)=1$.
4. Fix some $x \in \mathbb{R}$. Let $E_{n}=\left(-\infty, x+\varepsilon_{n}\right)$, for any decreasing sequence $\varepsilon_{n}$ of positive numbers that converges to zero. Then $E_{n+1} \subset E_{n}$ and $\cap_{n} E_{n}=(-\infty, x]$, and so by Theorem 3.5

$$
\lim _{n} F\left(x+\varepsilon_{n}\right)=\lim _{n} \mu\left(E_{n}\right)=\lim _{n} \mu((-\infty, x])=F(x) .
$$

Claim 3.8. A probability measure $\mu$ on $(\mathbb{R}, \mathscr{B})$ is uniquely determined by its cumulative distribution function.

Proof. Let $\mathscr{P}$ be the $\pi$-system from Example 2.1, and note that $\mathscr{B}=\sigma(\mathscr{P})$. Let $\mu_{0}: \mathscr{P} \rightarrow[0,1]$ be given by

$$
\mu_{0}((-\infty, x])=F(x)
$$

so that $\mu_{0}$ is the restriction of $\mu$ to $\mathscr{P}$. Since $\mu$ is (trivially) countably additive, by the HahnKolmogorov Theorem that there is at most one probability measure that extends $\mu_{0}$ to $\mathscr{B}$. Thus $\mu$ is the unique measure with c.d.f. $F$.

One can in fact show that for any $F$ that satisfies the properties of Claim 3.7 is the cumulative distribution function of some $\mu$. The proof uses the Hahn-Kolmogorov Theorem.

## 4 Events and random variables

### 4.1 Events

Given a measurable space ( $\Omega, \mathscr{F}$ ), an event $A$ is an element of $\mathscr{F}$. We sometimes call events measurable sets.

### 4.2 Sub-sigma-algebras

A sub-sigma-algebra of $\mathscr{F}$ is a subset of $\mathscr{F}$ that is also a sigma-algebra.
Given another measurable space $(\Theta, \mathscr{G})$, a function $f: \Omega \rightarrow \Theta$ is measurable if for all $A \in \mathscr{G}$ it holds that $f^{-1}(A) \in \mathscr{F}$.

Exercise 4.1. Prove that $f$ is measurable iff the collection

$$
\begin{equation*}
\sigma(f)=\left\{f^{-1}(A): A \in \mathscr{G}\right\}=f^{-1}(\mathscr{G}) . \tag{4.1}
\end{equation*}
$$

is a sub-sigma-algebra of $\mathscr{F}$.
Hence (assuming $f$ is onto, otherwise restrict to its image), $f^{-1}: \mathscr{G} \rightarrow \sigma(f)$ is an isomorphism of sigma-algebras.

Fix a measurable space ( $\Omega, \mathscr{F}$ ), and let $f$ be a measurable function to some other measurable space. Given a sub-sigma-algebra $\mathscr{G} \subseteq \mathscr{F}$, we say that $f$ is $\mathscr{G}$-measurable if $\sigma(f)$ is a sub-sigma-algebra of $\mathscr{G}$.

We say that a sigma-algebra $\mathscr{F}$ is separable if it generated by a countable subset. That is, if there exists some countable $\mathscr{C} \subset \mathscr{F}$ such that $\mathscr{F}=\sigma(\mathscr{C})$.

We say that $\mathscr{F}$ separates points if for all $\omega_{1} \neq \omega_{2}$ there exists some $A \in \mathscr{F}$ such that $\omega_{1} \in A$ and $\omega_{2} \notin A$.

Theorem 4.2. Let $(\Omega, \mathscr{F}),\left(\Theta_{1}, \mathscr{G}_{1}\right)$ and $\left(\Theta_{2}, \mathscr{G}_{2}\right)$ be measurable spaces with sigma-algebras that separate points. Let $f: \Omega \rightarrow \Theta_{1}$ and $g: \Omega \rightarrow \Theta_{2}$ be measurable functions. Then $g$ is $\sigma(f)$-measurable iff there exists a measurable $h: \Theta_{1} \rightarrow \Theta_{2}$ such that $g=h \circ f$.

Exercise 4.3. Prove for the case that $g=h \circ f$.
Measurable functions to ( $\mathbb{R}, \mathscr{B}$ ) will be of particular interest.
Claim 4.4. Let $(\Omega, \mathscr{F})$ be a measurable space, and let $f: \Omega \rightarrow \mathbb{R}$. Then

1. If $\mathscr{C} \subset \mathscr{B}$ satisfies $\sigma(\mathscr{C})=\mathscr{B}$, and if $f^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{C}$ then $f$ is measurable.
2. For each $x \in \mathbb{R}$ let $A_{x} \subset \Omega$ be given by $A_{x}=\{\omega: f(\omega) \leq x\}$. If each $A_{x}$ is in $\mathscr{F}$ then $f$ is measurable.
3. If $\Omega$ is a topological space with Borel sigma-algebra $\mathscr{F}$, and if $f$ is continuous, then it is measurable.
4. If $g$ is a measurable function from $(\mathbb{R}, \mathscr{B})$ to itself and $f$ is measurable then $g \circ f$ is measurable.

Claim 4.5. Let $(\Omega, \mathscr{F})$ be a measurable space, and let $\left\{f_{n}\right\}$ be a sequence of measurable functions to $(\mathbb{R}, \mathscr{B})$ with $0 \leq f_{n} \leq 1$ for all $n$. Then the following are measurable:

1. $\inf _{n} f_{n}$.
2. $\liminf _{n} f_{n}$.
3. The set $\left\{\omega: \lim _{n} f_{n}(\omega)\right.$ exists $\}$.

Claim 4.6. The measurable functions $(\Omega, \mathscr{F}) \rightarrow(\mathbb{R}, \mathscr{B})$ are a vector space over the reals:

1. If $f$ is measurable then $\lambda f$ is measurable, for all $\lambda \in \mathbb{R}$.
2. If $f_{1}$ and $f_{2}$ are measurable, then $f_{1}+f_{2}$ is measurable.

### 4.3 Random variables

Given a probability space $(\Omega, \mathscr{F}, \mu)$ and a measurable space $(\Theta, \mathscr{G})$, we say that two measurable functions $f, g: \Omega \rightarrow \Theta$ are equivalent if $\mu(\{\omega: f(\omega)=g(\omega)\})=1$. A random variable is an equivalence class of measurable functions. We will often consider the case that $(\Theta, \mathscr{G})=(\mathbb{R}, \mathscr{B})$, in which case we will call $X$ a real random variable. In fact, we will do this so often that we will often refer to real random variables as just "random variables".

A few notes:

1. Note we will often just think of random variables as measurable functions. We will say, for example, that a real random variable is non-negative, by which we will mean that there is a non-negative function in the equivalence class. We will also define random variables by just describing one element of the equivalence class.
2. It is easy to verify that sums, products, limits etc. of random variables are well defined, in the sense that (for example) the equivalence class of $f+g$ is equal to the equivalence class of $f^{\prime}+g^{\prime}$ whenever $f$ and $f^{\prime}$ are equivalent and $g$ and $g^{\prime}$ are equivalent.
3. We will want to verify that expressions of the form

$$
\mu(\{X \in A\})=\mu(\{w: X(\omega) \in A\})
$$

for a random variable $X$ and measurable $A$ are well defined, in the sense that they are independent of the choice of representative: $X(\omega)$ can be taken to mean $f(\omega)$, where $f$ is any member of the equivalence class $X$.
Example 4.7. Let $\Omega=\{0,1\}^{\mathbb{N}}$, and let $\mathbb{P}$ be the Bernoulli measure defined in Example 3.2. Define the random variable $X: \Omega \rightarrow \mathbb{R}$ by

$$
X(\omega)=\max \left\{n \in \mathbb{N}: \omega_{k}=0 \text { for all } k \leq n\right\}
$$

Note that $X$ is not well defined at a single point in $\Omega$, the all zeros sequence. We accordingly extend $\mathbb{R}$ to include $\infty($ and $-\infty)$ and assign $X(\omega)=\infty$ in this case.

Given a random variable $X: \Omega \rightarrow \Theta$, we define the pushforward measure $v=X_{*} \mu$ on $(\Theta, \mathscr{G})$ by

$$
v(A)=\mu\left(X^{-1}(A)\right)
$$

The measure $v$ is also called the law of $X$.
Exercise 4.8. Calculate the cumulative distribution function of the random variable defined in Example 4.7.

## 5 Independence and the Borel-Cantelli Lemmas

### 5.1 Independence

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Let $\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ be sub-sigma-algebras. We say that these sigma-algebras are independent if for any $\left(A_{1}, A_{2}, \ldots\right)$ with $A_{n} \in \mathscr{F}_{n}$ and any finite sequence $n_{k}$ it holds that

$$
\begin{equation*}
\mathbb{P}\left[\cap_{k} A_{n_{k}}\right]=\prod_{k} \mathbb{P}\left[A_{n_{k}}\right] . \tag{5.1}
\end{equation*}
$$

We say that the random variables ( $X_{1}, X_{2}, \ldots$ ) are independent if ( $\sigma\left(X_{1}\right), \sigma\left(X_{2}\right), \ldots$ ) are independent.

We say that the events $\left(A_{1}, A_{2}, \ldots\right)$ are independent if their indicators functions ( $\mathbb{1}_{\left\{A_{1}\right\}}, \mathbb{1}_{\left\{A_{2}\right\}}, \ldots$ ) are independent. Note that $\sigma\left(\mathbb{1}_{\{A\}}\right)=\left\{\varnothing, A, A^{\mathrm{c}}, \Omega\right\}$.

Claim 5.1. Let the events $\left(A_{1}, A_{2}, \ldots\right)$ be independent. Then

$$
\mathbb{P}\left[\cap_{n} A_{n}\right]=\prod_{n} \mathbb{P}\left[A_{n}\right] .
$$

Proof. By independence we have that for any $m \in \mathbb{N}$

$$
\mathbb{P}\left[\cap_{n=1}^{m} A_{n}\right]=\prod_{n=1}^{m} \mathbb{P}\left[A_{n}\right] .
$$

Denote $B_{m}=\cap_{n=1}^{m} A_{n}$. Then $B_{n}$ is a decreasing sequence with $\cap_{m} B_{m}=\cap_{n} A_{n}$, and so by Theorem 3.5 we have that

$$
\mathbb{P}\left[\cap_{n} A_{n}\right]=\mathbb{P}\left[\cap_{m} B_{m}\right]=\lim _{m} \mathbb{P}\left[B_{m}\right]=\lim _{m} \prod_{n=1}^{m} \mathbb{P}\left[A_{n}\right]=\prod_{n} \mathbb{P}\left[A_{n}\right] .
$$

It turns out that to prove independence it suffices to show (5.1) for generating $\pi$-systems.

### 5.2 A question about a sequence of random variables

Theorem 5.2. Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of independent real random variables, each with the distribution $\mathbb{P}\left[X_{n}>x\right]=\mathrm{e}^{-x}$ when $x>0$ and $\mathbb{P}\left[X_{n}>x\right]=1$ when $x \leq 0$. Let

$$
L=\underset{n}{\limsup } \frac{X_{n}}{\log n} .
$$

Then $\mathbb{P}[L=1]=1$.
To prove this Theorem we will need the Borel-Cantelli Lemmas.

### 5.3 The Borel-Cantelli Lemmas

Lemma 5.3 (Borel-Cantelli Lemmas). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $\left(A_{1}, A_{2}, \ldots\right)$ be a sequence of events.

1. If $\sum_{n} \mathbb{P}\left[A_{n}\right]<\infty$ then

$$
\mathbb{P}\left[\omega \in \Omega: \omega \in A_{n} \text { for infinitely many } n\right]=0 .
$$

2. If $\sum_{n} \mathbb{P}\left[A_{n}\right]=\infty$ and $\left(A_{1}, A_{2}, \ldots\right)$ are independent then

$$
\mathbb{P}\left[\omega \in \Omega: \omega \in A_{n} \text { for infinitely many } n\right]=1 .
$$

To see why independence is needed for the second part, consider the case that all the events $A_{n}$ are equal to some event $A$ with $0<\mathbb{P}[A]<1$.

Proof of Lemma 5.3. 1. Note that

$$
\left\{\omega: \omega \in A_{n} \text { for infinitely many } n\right\}=\cap_{n} \cup_{m \geq n} A_{m} .
$$

Let $B_{n}=\cup_{m \geq n} A_{m}$, so that we want to show that $\mathbb{P}\left[\cap_{n} B_{n}\right]=0$. Note that $B_{n}$ is a decreasing sequence (i.e., $B_{n+1} \subseteq B_{n}$ ) and therefore by Theorem 3.5 we have that

$$
\mathbb{P}\left[\cap_{n} B_{n}\right]=\lim _{n} \mathbb{P}\left[B_{n}\right] .
$$

Since $B_{n}=\cup_{m \geq n} A_{m}$, we have that $\mathbb{P}\left[B_{n}\right] \leq \sum_{m \geq n} \mathbb{P}\left[A_{m}\right]$. But the latter converges to 0 , and so we are done.
2. Note that

$$
\begin{aligned}
\left\{\omega: \omega \in A_{n} \text { for infinitely many } n\right\}^{\mathrm{c}} & =\left\{\omega: \omega \in A_{n} \text { for finitely many } n\right\} \\
& =\left\{\omega: \omega \in A_{n}^{\mathrm{c}} \text { for all } n \text { large enough }\right\} \\
& =\cup_{n} \cap_{m \geq n} A_{m}^{\mathrm{c}} .
\end{aligned}
$$

We would hence like to show that $\mathbb{P}\left[\cup_{n} \cap_{m \geq n} A_{m}^{\mathrm{c}}\right]=0$.
Let $C_{n}=\cap_{m \geq n} A_{m}^{\mathrm{c}}$. Then by independence and Claim 5.1 we have that

$$
\mathbb{P}\left[C_{n}\right]=\mathbb{P}\left[\cap_{m \geq n} A_{m}^{\mathrm{c}}\right]=\prod_{m \geq n}\left(1-\mathbb{P}\left[A_{m}\right]\right)
$$

Since $1-x \leq \mathrm{e}^{-x}$ this implies that

$$
\mathbb{P}\left[C_{n}\right] \leq \exp \left(-\sum_{m \geq n} \mathbb{P}\left[A_{m}\right]\right)=0 .
$$

Finally, by Corollary 3.6, $\mathbb{P}\left[\cap_{n} C_{n}\right]=0$.

Proof of Theorem 5.2. Let $A_{n}$ be the event that $X_{n} \geq \alpha \log n$. Then

$$
\mathbb{P}\left[A_{n}\right]=n^{-\alpha},
$$

and the events $\left(A_{1}, A_{2}, \ldots\right)$ are independent (exercise!). Also, note that

$$
\sum_{n} \mathbb{P}\left[A_{n}\right] \begin{cases}=\infty & \text { if } \alpha \leq 1 \\ <\infty & \text { if } \alpha>1\end{cases}
$$

Thus, from the Borel-Cantelli Lemmas it follows that

$$
\mathbb{P}\left[X_{n} \geq \alpha \log n \text { for infinitely many } n\right]= \begin{cases}1 & \text { if } \alpha \leq 1 \\ 0 & \text { if } \alpha>1 .\end{cases}
$$

Now, note that the event $\{L \geq \alpha\}$ is identical to the event

$$
\cap_{m>0}\left\{X_{n} \geq(\alpha-1 / m) \log n \text { for infinitely many } n\right\},
$$

and so $\mathbb{P}[L \geq 1]=1$, by Corollary 3.6. It also follows that $\mathbb{P}[L \geq 1+1 / n]=0$ for any $n>0$, and so we have that $\mathbb{P}[L>1]=0$, again by Corollary 3.6. Hence $\mathbb{P}[L \leq 1]=1$, and so $\mathbb{P}[L=1]=$ 1.

## 6 The tail sigma-algebra

### 6.1 Motivating example

Consider a sequence of independent real random variables ( $X_{1}, X_{2}, \ldots$ ) such that there exists some $M \geq 0$ such that $\mathbb{P}\left[\left|X_{n}\right| \leq M\right]=1$ for all $n$. That is, the sequence is uniformly bounded.

Define the random variables

$$
Y_{n}=\frac{1}{n} \sum_{k=1}^{n} X_{n} \quad \text { and } \quad L=\limsup _{n} Y_{n}
$$

Claim 6.1. $\mathbb{P}[|L| \leq M]=1$.
Proof. Clearly $\mathbb{P}\left[\left|Y_{n}\right| \leq M\right]=1$. Hence $\mathbb{P}\left[\left|Y_{n}\right| \leq M\right.$ for all $\left.n\right]=1$, and thus $\mathbb{P}[|L| \leq M]=1$.
Define the event $A=\left\{\lim _{n} Y_{n}\right.$ exists $\}$.
Theorem 6.2. There exists some $c \in[-M, M]$ such that $\mathbb{P}[L=c]=1$, and $\mathbb{P}[A] \in\{0,1\}$.
An interesting observation is that $L$ is independent of $X_{1}$. To see this, define

$$
L^{\prime}=\limsup _{n} \frac{1}{n} \sum_{k=2}^{n} X_{n}
$$

which is clearly independent of $X_{1}$. But

$$
L=\limsup _{n} \frac{1}{n} \sum_{k=1}^{n} X_{n}=\limsup _{n} \frac{X_{1}}{n}+\frac{1}{n} \sum_{k=2}^{n} X_{n}=L^{\prime} .
$$

In fact, by the same argument, $L$ is independent of $\left(X_{1}, X_{2}, \ldots, X_{n}\right)$ for any $n$. This makes $L$ a tail random variable, as we now explain.

### 6.2 The tail sigma-algebra

For each $n \in \mathbb{N}$ define the sigma-algebra $\mathscr{T}_{n}$ by

$$
\sigma\left(X_{n}, X_{n+1}, \ldots\right),
$$

which is the smallest sigma-algebra that contains $\left(\sigma\left(X_{n}\right), \sigma\left(X_{n+1}\right), \ldots\right)$. Define the tail sigmaalgebra by

$$
\mathscr{T}=\cap_{n} \mathscr{T}_{n} .
$$

A random variable is a tail random variable if it is $\mathscr{T}$-measurable.
Claim 6.3. $L$ is a tail random variable.

Proof. Using a construction similar to the $L^{\prime}$ construction above, it is easy to see that for every $n$ there exists a function $f_{n}$ such that $L=f_{n}\left(X_{n}, X_{n+1}, \ldots\right)$. It follows that $L$ is $\mathscr{T}_{n}$-measurable. Thus for every $A \in \sigma(A)$ it holds that $L^{-1}(A) \in \mathscr{T}_{n}$, for every $n$. Thus $A \in \cap_{n} \mathscr{T}_{n}=\mathscr{T}$.

Let ( $Z_{1}, Z_{2}, \ldots$ ) be i.i.d random variables, each distributed uniformly over the set of symbols $S=\{a, b, c\}$. Let $S^{*}$ be the set of finite strings over $S$, and define the random variable $W_{n}$ taking values in $S^{*}$ as follows:

- $W_{1}=Z_{1}$.
- If $W_{n}$ is empty, or if the last symbol in $W_{n}$ is different than $Z_{n+1}$, then $W_{n+1}$ is the concatenation $W_{n} Z_{n+1}$.
- If the last symbol in $W_{n}$ is $Z_{n+1}$ then $W_{n+1}$ is equal to $W_{n}$, with this last symbol removed.

We will prove later in the course that with probability one it holds that $\lim _{n}\left|W_{n}\right|=\infty$, and hence we can define the random variable $T$ to be the eventual first symbol in all $W_{n}$ high enough. It is immediate that $T$ is measurable in the tail sigma-algebra of the sequence $\left(W_{1}, W_{2}, \ldots\right)$.

It is also easy to see that $\mathbb{P}[T=a]=1 / 3$, since by the symmetry of the definitions, $\mathbb{P}[T=a]=\mathbb{P}[T=b]=\mathbb{P}[T=c]$, and these must sum to one. By the same argument, the probability that $W_{n}$ starts with some string $w$ for all $n$ high enough is $3^{-1} \cdot 2^{-|w|}$.

### 6.3 The zero-one law

Theorem 6.4 (Kolmogorov's Zero-One Law). Let $\mathscr{T}$ be the tail sigma-algebra of a sequence of independent random variables. Then $\mathbb{P}[A] \in\{0,1\}$ for any $A \in \mathscr{T}$.

Before proving this theorem we will prove a lemma.
Lemma 6.5. Let the event $A$ be independent of itself. Then $\mathbb{P}[A] \in\{0,1\}$.
Proof. $\mathbb{P}[A]=\mathbb{P}[A \cap A]=\mathbb{P}[A] \cdot \mathbb{P}[A]$.
Proof of Theorem 6.4. Let $\mathscr{G}_{n}=\sigma\left(X_{1}, \ldots, X_{n-1}\right), \mathscr{T}_{n}=\sigma\left(X_{n}, X_{n+1}, \ldots\right)$ and $\mathscr{T}=\cap_{n} \mathscr{T}_{n}$. We first claim that $\mathscr{G}_{n}$ and $\mathscr{T}_{n}$ are independent. To see this, define $\mathscr{T}_{n}^{m}=\sigma\left(X_{n}, \ldots, X_{n+m}\right)$, and note that $\mathscr{T}_{n}^{m}$ and $\mathscr{G}_{n}$ are independent, and so $\mathbb{P}[A \cap B]=\mathbb{P}[A] \cdot \mathbb{P}[B]$ for any $A \in \mathscr{G}_{n}$ and any $B \in \mathscr{T}_{n}^{m}$. Now $\mathscr{C}_{n}=\cup_{m} \mathscr{T}_{n}^{m}$ is not a sigma-algebra, but it is a $\pi$-system. Since $\mathbb{P}[A \cap B]=\mathbb{P}[A] \cdot \mathbb{P}[B]$ for any $A \in \mathscr{G}_{n}$ and any $B \in \mathscr{C}_{n}$, it follows that $\mathscr{G}_{n}$ and $\sigma\left(\mathscr{C}_{n}\right)=\mathscr{T}_{n}$ are independent.

Since $\mathscr{T} \subset \mathscr{T}_{n}$ then $\mathscr{G}_{n}$ and $\mathscr{T}$ are independent. Hence $\mathscr{T}$ is independent of

$$
\sigma\left(\cup_{n} \mathscr{G}_{n}\right)=\sigma\left(\cup_{n} \sigma\left(X_{n}\right)\right)=\sigma\left(X_{1}, X_{2}, \ldots\right) .
$$

Since $\mathscr{T} \subset \sigma\left(X_{1}, X_{2}, \ldots\right)$ it follows that $\mathscr{T}$ is independent of $\mathscr{T}$, and so $\mathbb{P}[A] \in\{0,1\}$ for any $A \in \mathscr{T}$.

Proof of Theorem 6.2. Since $A$ is a tail random variable then $\mathbb{P}[A] \in\{0,1\}$.
For any $q \in \mathbb{Q}$ define the tail event $A_{q}=\{L \geq q\}$. By Kolmogorov's zero-one, law, the probability of each of these is either 0 or 1 , and so there is some

$$
c=\sup \left\{q: \mathbb{P}\left[A_{q}\right]=1\right\}=\inf \left\{q: \mathbb{P}\left[A_{q}\right]=0\right\} .
$$

Since $\mathbb{Q}$ is countable, $\mathbb{P}[L \geq c]=\mathbb{P}[L \leq c]=1$, and so $\mathbb{P}[L=c]=1$. Finally, $c \in[-M, M]$, since $\mathbb{P}[L \in[-M, M]]=1$.

## 7 Expectations

### 7.1 Expectations in finite probability spaces

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a finite probability space, and let $f: \Omega \rightarrow \mathbb{R}$ be any function. The expectation of $f$ is given by

$$
\mathbb{E}[f]=\sum_{\omega \in \Omega} f(\omega) \mathbb{P}[\omega] .
$$

Another way of writing this is the following:

$$
\mathbb{E}[f]=\sum_{x \in \operatorname{Im} f} x \mathbb{P}\left[f^{-1}(x)\right] .
$$

Note that this formulation does not reference points in $\Omega$. Relatedly, it has the advantage that it naturally extends to any probability space, given that $f$ has a finite (or countable) image. This is the basic idea that is behind our general definition of expectation.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. We say that a measurable non-negative $\tilde{f}$ is simple if it has a finite image, ${ }^{1}$ and define its expectation $\mathbb{E}[\tilde{f}]$ by

$$
\mathbb{E}[\tilde{f}]=\sum_{x \in \operatorname{Im} \tilde{f}} x \mathbb{P}\left[\tilde{f}^{-1}(x)\right] .
$$

### 7.2 Expectations of non-negative random variables

Given a (non-simple) non-negative real function $f$, we define its expectation by

$$
\mathbb{E}[f]=\sup \{\mathbb{E}[\tilde{f}]: \tilde{f} \text { is simple and } \tilde{f} \leq f\} .
$$

Note that this supremum may be infinite.
It is straightforward to verify that for any non-negative functions $f, g$ such that $\mathbb{P}[f=g]=$ 1 it holds that $\mathbb{E}[f]=\mathbb{E}[g]$. We can therefore define the expectation of a random variable $X$ as the expectation of any $f$ in the equivalence class. We will henceforth consider expectations of random variables.

It is likewise straightforward to verify that for any two non-negative random variables $X, Y$ :

- Linearity of expectation: For any $\lambda>0$ it holds that

$$
\mathbb{E}[X+\lambda Y]=\mathbb{E}[X]+\lambda \mathbb{E}[Y] .
$$

- If $X \geq Y$ then $\mathbb{E}[X] \geq \mathbb{E}[Y]$.

[^0]
### 7.3 Markov's inequality

Theorem 7.1 (Markov's Inequality). If $X$ is a non-negative random variable with $\mathbb{E}[X]<\infty$ then for every $\lambda>0$

$$
\mathbb{P}[X \geq \lambda] \leq \frac{\mathbb{E}[X]}{\lambda} .
$$

Proof. Let $A=\{X \geq \lambda\}$, and let $Y$ be given by

$$
Y(\omega)=\lambda \cdot \mathbb{1}_{\{A\}}(\omega)= \begin{cases}\lambda & \text { if } \omega \in A \\ 0 & \text { otherwise }\end{cases}
$$

Then $Y \leq X$, and so $\mathbb{E}[Y] \leq \mathbb{E}[X]$. Since $\mathbb{E}[Y]=\lambda \cdot \mathbb{P}[A]$, we get that

$$
\lambda \cdot \mathbb{P}[X \geq \lambda] \leq \mathbb{E}[X],
$$

and the claim follows by dividing both sides by $\lambda$.

### 7.4 Pointwise convergence and convergence of expectations

Consider the non-negative random variables ( $X_{1}, X_{2}, \ldots$ ) defined on the interval ( 0,1 ] (equipped with the Borel sigma-algebra and Lebesgue measure) which are given by

$$
X_{n}(x)= \begin{cases}n & \text { if } x \leq 1 / n \\ 0 & \text { otherwise }\end{cases}
$$

Then

1. $\mathbb{E}\left[X_{n}\right]=1$.
2. For every $x \in(0,1]$ it holds that $\lim _{n} X_{n}(x)=X(x)$, where $X$ is the constant function $X(x)=0$.
3. $\lim _{n} \mathbb{E}\left[X_{n}\right] \neq \mathbb{E}[X]$.

Hence it is not necessarily true that if $X_{n} \rightarrow X$ pointwise then $\mathbb{E}\left[X_{n}\right] \rightarrow \mathbb{E}[X]$.
Theorem 7.2 (Monotone Convergence Theorem). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of non-negative random variables such that $X_{n}(\omega)$ is increasing for every $\omega \in \Omega$. Let $X(\omega)=\lim _{n} X_{n}(\omega) \in[0, \infty]$. Then

$$
\lim _{n} \mathbb{E}\left[X_{n}\right]=\mathbb{E}[X] \in[0, \infty] .
$$

Theorem 7.3 (Dominated Convergence Theorem). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, and let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of non-negative random variables. Let $X, Y$ be a non-negative random variables with $\mathbb{E}[Y]<\infty$, and such that $\lim _{n} X_{n}(\omega)=X(\omega)$ for every $\omega \in \Omega$, and $X_{n}(\omega) \leq Y(\omega)$ for every $\omega \in \Omega$ and $n \in \mathbb{N}$. Then

$$
\lim _{n} \mathbb{E}\left[X_{n}\right]=\mathbb{E}[X] .
$$

## $7.5 \mathscr{L}^{p}$

Given a random variable $X$, we define the random variables $X^{+}$and $X^{-}$by

$$
X^{+}(\omega)=\max \{X(\omega), 0\} \quad \text { and } \quad X^{-}(\omega)=\max \{-X(\omega), 0\},
$$

so that $X^{+}$and $X^{-}$are both non-negative, and $X=X^{+}-X^{-}$. If $\mathbb{E}\left[X^{+}\right]$and $\mathbb{E}\left[X^{-}\right]$are both finite, we define

$$
\mathbb{E}[X]=\mathbb{E}\left[X^{+}\right]-\mathbb{E}\left[X^{-}\right],
$$

and say that $X \in \mathscr{L}^{1}(\Omega, \mathscr{F}, \mathbb{P})$, or just $X \in \mathscr{L}^{1}$. Note that $X \in \mathscr{L}^{1}$ iff $\mathbb{E}[|X|]<\infty$ iff $|X| \in \mathscr{L}^{1}$. For $p \geq 1$ we say that $X \in \mathscr{L}^{p}$ if $|X|^{p} \in \mathscr{L}^{1}$.

Exercise 7.4. Show that $\mathscr{L}^{p}$ is a vector space.
$X \mapsto \mathbb{E}\left[|X|^{p}\right]^{1 / p}$ defines a norm on $\mathscr{L}^{p}$.
Theorem 7.5. If $r>p \geq 1$ and $X \in \mathscr{L}^{r}$ then $X \in \mathscr{L}^{p}$ and moreover

$$
\mathbb{E}\left[|X|^{p}\right]^{1 / p} \leq \mathbb{E}\left[|X|^{r}\right]^{1 / r} .
$$

In fact, if we equip $\mathscr{L}^{p}$ with this norm, then it is a Banach space; that is, it is complete with respect to the metric induced by this norm.

Theorem 7.6. Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of random variables in $\mathscr{L}^{p}$ such that

$$
\lim _{r} \sup _{m, n \geq r}\left\{\mathbb{E}\left[\left|X_{n}-X_{m}\right|^{p}\right]\right\}=0 .
$$

Then there exists an $X \in \mathscr{L}^{p}$ such that

$$
\lim _{n} \mathbb{E}\left[\left|X_{n}-X\right|^{p}\right]=0
$$

## $7.6 \mathscr{L}^{2}$

A particularly interesting case is $p=2$. In this case we can define an inner product $(X, Y):=$ $\mathbb{E}[X \cdot Y]$, which makes $\mathscr{L}^{2}$ a Hilbert space, with completeness given by Theorem 7.6.
Theorem 7.7. Let $X, Y \in \mathscr{L}^{2}$. Then $X \cdot Y \in \mathscr{L}^{1}$.
Proof. Note first that $|X|,|Y| \in \mathscr{L}^{2}$. Since $\mathscr{L}^{2}$ is a vector space then $\mathbb{E}\left[(|X|+|Y|)^{2}\right]<\infty$, and so

$$
\mathbb{E}\left[X^{2}+2|X| \cdot|Y|+Y^{2}\right]<\infty .
$$

By the linearity of expectation

$$
\mathbb{E}\left[(|X|+|Y|)^{2}\right]=\mathbb{E}\left[X^{2}\right]+2 \cdot \mathbb{E}[|X| \cdot|Y|]+\mathbb{E}\left[Y^{2}\right],
$$

and so we have that $\mathbb{E}[|X| \cdot|Y|]<\infty$. Now,

$$
\mathbb{E}[|X \cdot Y|]=\mathbb{E}[|X| \cdot|Y|],
$$

and so $|X \cdot Y| \in \mathscr{L}^{1}$. Hence $X \cdot Y \in \mathscr{L}^{1}$.

It follows from Theorems 7.6 and 7.7 that $\mathscr{L}^{2}$ is a real Hilbert space, when equipped with the inner product $(X, Y):=\mathbb{E}[X \cdot Y]$. We can therefore immediately conclude that for any $X, Y \in \mathscr{L}^{2}$

1. $\mathbb{E}[X \cdot Y]^{2} \leq \mathbb{E}\left[X^{2}\right] \cdot \mathbb{E}\left[Y^{2}\right]$, with equality iff for some $\lambda \in \mathbb{R}$ it a.s. holds that $X=\lambda \cdot Y$. This is the Cauchy-Schwartz inequality.
2. $\mathbb{E}\left[(X+Y)^{2}\right]=\mathbb{E}\left[X^{2}\right]+\mathbb{E}\left[Y^{2}\right]$ iff $\mathbb{E}[X \cdot Y]=0$.

Given $X \in \mathscr{L}^{2}$, we define the random variable $\tilde{X}:=X-\mathbb{E}[X]$, and denote $\operatorname{Var}(X)=\mathbb{E}[\tilde{X} \cdot \tilde{X}]$ and $\operatorname{Cov}(X, Y)=\mathbb{E}[\tilde{X} \cdot \tilde{Y}]$. We say that $X$ and $Y$ are uncorrelated if $\operatorname{Cov}(X, Y)=0$. Using these definitions the facts above become

1. $\operatorname{Cov}(X, Y)^{2} \leq \operatorname{Var}(X) \cdot \operatorname{Var}(Y)$, with equality iff for some $\lambda \in \mathbb{R}$ it a.s. holds that $X=\lambda \cdot Y$.
2. $\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)$ iff $X$ and $Y$ are uncorrelated.

## 8 A strong law of large numbers and the Chernoff bound

### 8.1 Expectation of a product of independent random variables

Theorem 8.1. Let $X, Y \in \mathscr{L}^{1}$ be independent. Then $X \cdot Y \in \mathscr{L}^{1}$ and

$$
\mathbb{E}[X \cdot Y]=\mathbb{E}[X] \cdot \mathbb{E}[Y] .
$$

To prove this, we first note that it holds for indicator functions by the definition of independence, then show that it holds for simple functions, and apply the monotone convergence theorem to show that it holds in general.

### 8.2 Jensen's inequality

Theorem 8.2 (Jensen's Inequality). Let $X$ be a real random variable with $\mathbb{E}[X]=x_{0} \in \mathbb{R}$. Let $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. Then $\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X])$.

Proof. Since $\varphi$ is convex we can find $a, b \in \mathbb{R}$ such that $\varphi(x) \geq a x+b$ for all $x \in \mathbb{R}$, and $\varphi\left(x_{0}\right)=a x_{0}+b$. Hence

$$
\mathbb{E}[\varphi(X)] \geq \mathbb{E}[a X+b]=a \mathbb{E}[x]+b=a x_{0}+b=\varphi\left(x_{0}\right)=\varphi(\mathbb{E}[X]) .
$$

### 8.3 SLLN in $\mathscr{L}^{4}$

Theorem 8.3. Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of independent random variables uniformly bounded in $\mathscr{L}^{4}$ (so that $\mathbb{E}\left[X_{n}^{4}\right]<K$ for all $n$ and some $K>0$ ), and with $\mathbb{E}\left[X_{n}\right]=0$. Let

$$
Y_{n}=\frac{1}{n} \sum_{k \leq n} X_{n} .
$$

Then $\lim _{n} Y_{n}=0$ a.s.
Proof. By independence

$$
\mathbb{E}\left[X_{k} \cdot X_{\ell}^{3}\right]=\mathbb{E}\left[X_{k} \cdot X_{\ell}^{2} \cdot X_{m}\right]=0,
$$

and so, by linearity we have that

$$
\mathbb{E}\left[Y_{n}^{4}\right]=\mathbb{E}\left[\left(\frac{1}{n} \sum_{k=1}^{n} X_{k}\right)^{4}\right]=\frac{1}{n^{4}} \sum_{k \leq n} \mathbb{E}\left[X_{k}^{4}\right]+\frac{6}{n^{4}} \sum_{k<\ell \leq n} \mathbb{E}\left[X_{k}^{2} \cdot X_{\ell}^{2}\right] .
$$

Again applying independence we get

$$
\mathbb{E}\left[Y_{n}^{4}\right]=\frac{1}{n^{4}} \sum_{k \leq n} \mathbb{E}\left[X_{k}^{4}\right]+\frac{6}{n^{4}} \sum_{k<\ell \leq n} \mathbb{E}\left[X_{k}^{2}\right] \cdot \mathbb{E}\left[X_{\ell}^{2}\right] .
$$

By Jensen's inequality $\mathbb{E}\left[X_{k}^{2}\right]^{2}<K$, and so

$$
\mathbb{E}\left[Y_{n}^{4}\right] \leq \frac{K}{n^{3}}+\frac{6 K}{n^{2}} \leq \frac{7 K}{n^{2}}
$$

It follows from Markov's inequality that for any $\varepsilon>0$

$$
\mathbb{P}\left[Y_{n}^{4} \geq \varepsilon^{4}\right] \leq \frac{7 K}{\varepsilon^{4} n^{2}},
$$

and so, by Borel-Cantelli, $\limsup _{n}\left|Y_{n}\right| \leq \varepsilon$ for any $\varepsilon>0$. Intersecting these probability one events for $\varepsilon=1 / 2,1 / 3,1 / 4, \ldots$ yields that $\limsup _{n}\left|Y_{n}\right|=0$ and thus $\lim _{n} Y_{n}=0$.

### 8.4 The Chernoff bound

With a little additional effort we can prove that if $\mathbb{E}\left[X_{n}\right]=\mu$ then $\lim _{n} Y_{n}=\mu$. A natural question is: what is the probability that $Y_{n}$ is significantly far from $\mu$, for finite $n$ ? For example, for $\eta>\mu$, what is the probability that $Y_{n} \geq \eta$ ?

Theorem 8.4 (Chernoff Bound). Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of i.i.d. random variables in $\mathscr{L}^{\infty}$, and with $\mathbb{E}\left[X_{n}\right]=\mu$. Then for every $\eta>\mu$ there is an $r>0$ such that

$$
\mathbb{P}\left[Y_{n} \geq \eta\right] \leq \mathrm{e}^{-r \cdot n}
$$

Proof. Denote $p_{n}=\mathbb{P}\left[Y_{n} \geq \eta\right]$; we want to show that $p_{n} \leq \mathrm{e}^{-r \cdot n}$.
Note that the event $\left\{Y_{n} \geq \eta\right\}$ is identical to the event $\left\{\mathrm{e}^{t \cdot n \cdot Y_{n}} \geq \mathrm{e}^{t \cdot n \cdot \eta}\right\}$, for any $t>0$. Since $\mathrm{e}^{t \cdot n \cdot Y_{n}}$ is a positive random variable, by the Markov inequality we have that

$$
p_{n}=\mathbb{P}\left[\mathrm{e}^{t \cdot n \cdot Y_{n}} \geq \mathrm{e}^{t \cdot n \cdot \eta}\right] \leq \frac{\mathbb{E}\left[\mathrm{e}^{t \cdot n \cdot Y_{n}}\right]}{\mathrm{e}^{t \cdot n \cdot \eta}} .
$$

Now,

$$
\mathbb{E}\left[\mathrm{e}^{t \cdot n \cdot Y_{n}}\right]=\mathbb{E}\left[\prod_{k \leq n} \mathrm{e}^{t \cdot X_{k}}\right]=\prod_{k \leq n} \mathbb{E}\left[\mathrm{e}^{t \cdot X_{k}}\right],
$$

where the penultimate equality uses independence. Let $X$ be a random variable with the same distribution as each $X_{k}$. Then we have shown that

$$
\mathbb{E}\left[\mathrm{e}^{t \cdot n \cdot Y_{n}}\right]=\mathbb{E}\left[\mathrm{e}^{t \cdot X}\right]^{n} .
$$

We now define the moment generating function of $X$ by $M_{X}(t):=\mathbb{E}\left[\mathrm{e}^{t X}\right]$. The name comes from the fact that

$$
\begin{equation*}
M_{X}(t)=\sum_{n=0}^{\infty} \frac{t^{n}}{n!} \mathbb{E}\left[X^{n}\right] . \tag{8.1}
\end{equation*}
$$

Note that this means that $M_{X}^{\prime}(0)=\mathbb{E}[X]$. Using $M_{X}$ we can write

$$
p_{n} \leq \exp \left(-\left(t \cdot \eta-\log M_{x}(t)\right) \cdot n\right)
$$

If we define the cumulant generating function of $X$ by $K_{X}(t):=\log M_{X}(t)$, then

$$
p_{n} \leq \exp \left(-\left(t \cdot \eta-K_{X}(t)\right) \cdot n\right) .
$$

Since $K_{X}^{\prime}(0)=M_{X}^{\prime}(0) / M_{X}(0)=\mathbb{E}[X]$, and since $K_{X}$ is smooth (as it turns out), it follows that for $t>0$ small enough,

$$
t \cdot \eta-K_{X}(t)=t \cdot \eta-t \cdot \mu-O\left(t^{2}\right)>0
$$

Hence, if we define

$$
r=\sup _{t}\left\{t \cdot \eta-K_{X}(t)\right\}
$$

we get that $r>0$ and

$$
p_{n} \leq \mathrm{e}^{-r \cdot n} .
$$

It turns out that the Chernoff bound is asymptotically tight. We show this in §21.

## 9 The weak law of large numbers

## $9.1 \mathscr{L}^{2}$

Theorem 9.1. Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of independent real random variables in $\mathscr{L}^{2}$, let $\mathbb{E}\left[X_{n}\right]=\mu, \operatorname{Var}\left(X_{n}\right) \leq \sigma^{2}$, and let $Y_{n}=\frac{1}{n} \sum_{k \leq n} X_{n}$. Then for every $\varepsilon>0$ and $n \in \mathbb{N}$

$$
\mathbb{P}\left[\left|Y_{n}-\mu\right| \geq \varepsilon\right] \leq \frac{\sigma^{2}}{n \varepsilon},
$$

and in particular

$$
\lim _{n} \mathbb{P}\left[\left|Y_{n}-\mu\right| \geq \varepsilon\right]=0 .
$$

In this case we say that $Y_{n}$ converges in probability to $\mu$. More generally, we say that a sequence of real random variables $Y_{n}$ converges in probability to a real random variable $Y$ if

$$
\lim _{n} \mathbb{P}\left[\left|Y_{n}-Y\right| \geq \varepsilon\right]=0 .
$$

Exercise 9.2. Does convergence in probability imply pointwise convergence? Does pointwise convergence imply convergence in probability?

To prove this Theorem we will need Chebyshev's inequality, which is just Markov's inequality in disguise.

Lemma 9.3 (Chebyshev's Inequality). For every $X \in \mathscr{L}^{2}$ and for every $\lambda>0$ it holds that

$$
\mathbb{P}[|X-\mathbb{E}[X]| \geq \lambda] \leq \frac{\operatorname{Var}(X)}{\lambda^{2}}
$$

Proof of Theorem 9.1. Note that $\mathbb{E}\left[Y_{n}\right]=\mu$, and that, by independence,

$$
\operatorname{Var}\left(Y_{n}\right)=\operatorname{Var}\left(\frac{1}{n} \sum_{k \leq n} X_{k}\right)=\frac{1}{n^{2}} \operatorname{Var}\left(\sum_{k \leq n} X_{k}\right)=\frac{1}{n^{2}} \sum_{k \leq n} \operatorname{Var}\left(X_{k}\right) \leq \frac{\sigma^{2}}{n} .
$$

Hence Chebyshev's inequality yields that for every $\lambda>0$ we have that

$$
\mathbb{P}\left[\left|Y_{n}-\mu\right| \geq \varepsilon\right] \leq \frac{\sigma^{2}}{n \varepsilon}
$$

## $9.2 \mathscr{L}^{1}$

We can relax the assumption $X \in \mathscr{L}^{2}$ to $X \in \mathscr{L}^{1}$ and still prove the weak law of large numbers. In fact, even the strong law holds in this setting (for i.i.d. random variables), but we will leave the proof of that for after we prove the Ergodic Theorem.

Theorem 9.4. Let $\left(X_{1}, X_{2}, \ldots\right)$ be a sequence of i.i.d. real random variables in $\mathscr{L}^{1}$, let $\mathbb{E}\left[X_{n}\right]=$ $\mu$, and let $Y_{n}=\frac{1}{n} \sum_{k \leq n} X_{n}$. Then for every $\varepsilon>0$

$$
\lim _{n} \mathbb{P}\left[\left|Y_{n}-\mu\right| \geq \varepsilon\right]=0
$$

Proof. We assume $\mu=0$; the reduction is straightforward.
Let $X=X_{1}$. For $N \in \mathbb{N}$, and a r.v. $X$ denote

$$
X^{\leq N}=X \cdot \mathbb{1}_{\{|X| \leq N\}} \quad \text { and } \quad X^{>N}=X \cdot \mathbb{1}_{\{|X|>N\}},
$$

so that $X=X^{\leq N}+X^{>N}$. By the Dominated Convergence Theorem

$$
\begin{equation*}
\mathbb{E}\left[\left|X^{>N}\right|\right] \rightarrow 0 \quad \text { and } \quad \mathbb{E}\left[X^{\leq N}\right] \rightarrow \mathbb{E}[X]=0 \tag{9.1}
\end{equation*}
$$

since both are dominated by $|X|$.
Fix $\varepsilon, \delta>0$. To prove the claim (under our assumption that $\mu=0$ ) we show that $\mathbb{P}\left[\left|Y_{n}\right| \geq \varepsilon\right]<\delta$ for all $n$ large enough. For any $N \in \mathbb{N}$ we can write $Y_{n}$ as

$$
Y_{n}=\frac{1}{n} \sum_{k \leq n} X_{k}^{\leq N}+X_{k}^{>N}=Y_{n}^{\leq}+Y_{n}^{>},
$$

where

$$
Y_{n}^{\leq}:=\frac{1}{n} \sum_{k \leq n} X_{k}^{\leq N} \quad \text { and } \quad Y_{n}^{>}=\frac{1}{n} \sum_{k \leq n} X_{k}^{>N} .
$$

Note that $Y_{n}^{\leq}$is not the same as $Y_{n}^{\leq N}$; we will not need the latter. Likewise, $Y_{n}^{>}$is not the same as $Y_{n}^{>N}$.

Choose $N$ large enough so that $\mathbb{E}\left[\left|X^{>N}\right|\right]<\varepsilon \cdot \delta / 4$; this is possible by (9.1). Now,

$$
\mathbb{E}\left[\left|Y_{n}^{>}\right|\right]=\mathbb{E}\left[\frac{1}{n}\left|\sum_{k \leq n} X_{k}^{>N}\right|\right] \leq \mathbb{E}\left[\frac{1}{n} \sum_{k \leq n}\left|X_{k}^{>N}\right|\right]=\mathbb{E}\left[\left|X^{>N}\right|\right]<\varepsilon \cdot \delta / 4
$$

Therefore, by Markov's inequality, we have that

$$
\mathbb{P}\left[\left|Y_{n}^{>}\right| \geq \varepsilon / 2\right]<\delta / 2
$$

Since $X_{k}^{\leq N}$ is bounded it is in $\mathscr{L}^{2}$. Therefore, by independence,

$$
\operatorname{Var}\left(Y_{n}^{\leq}\right)=\frac{\operatorname{Var}\left(X_{k}^{\leq N}\right)}{n} \leq \frac{N^{2}}{n}
$$

By linearity of expectations $\mathbb{E}\left[Y_{n}^{\leq}\right]=\mathbb{E}\left[X_{n}^{\leq N}\right]$, and thus tends to zero, by (9.1). It thus from Chebyshev's inequality that for $n$ large enough $\mathbb{P}\left[\left|Y_{n}^{\leq}\right| \geq \varepsilon / 2\right]<\delta / 2$. Since $\mathbb{P}\left[\left|Y_{n}\right| \geq \varepsilon\right] \leq$ $\mathbb{P}\left[\left|Y_{n}^{\leq}\right| \geq \varepsilon / 2\right.$ and $\left.\left|Y_{n}^{>}\right| \geq \varepsilon / 2\right]$, the claim follows by the union bound.

## 10 Conditional expectations

### 10.1 Why things are not as simple as they seem

Consider a point chosen uniformly from the surface of the (idealized, spherical) earth, so that the probability of falling on a set is proportional to its area.

Say we condition on the point falling on the equator. What is the conditional distribution? It obviously has to be uniform: by symmetry, there cannot be a reason that it is more likely to be in one time zone than another.

Say now that we condition on the point falling on a particular meridian $m$. By the same reasoning, the conditional distribution is uniform, and so, for example, the probability that we are within 2 meters of the north pole is the same as the probability that we are within 1 meter from the equator. Integrating over $m$ we get that regardless of the meridian, the probability of being 2 meters from the north pole is the same as the probability of being 1 meter from the equator. But the area within 2 meters of the north pole is about $4 \mathrm{\pi m}^{2}$, whereas the area within 1 meter of the equator is about $80000 \mathrm{~m}^{2}$.

### 10.2 Conditional expectations in finite spaces

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $|\Omega|<\infty, \mathscr{F}=2^{\Omega}$, and $\mathbb{P}[\omega]>0$ for all $\omega \in \Omega$. Let $\Omega=\{1, \ldots, n\}^{2}$, let $Z$ be the random variable given by $Z\left(\omega_{1}, \omega_{2}\right)=\omega_{1}$, and let $\mathscr{G}=\sigma(Z)$ be the sigma-algebra generated by the sets $A_{k}=\{k\} \times\{0, \ldots, n\}$. Let $X$ be a real random variable. Then the usual definition is the $\mathbb{E}[X \mid Z]$ is the random variable $\Omega \rightarrow \mathbb{R}$ given by

$$
\mathbb{E}[X \mid Z](\omega)=\frac{\sum_{\omega^{\prime} \in Z^{-1}(\omega)} X\left(\omega^{\prime}\right) \mathbb{P}\left[\omega^{\prime}\right]}{\sum_{\omega^{\prime} \in Z^{-1}(\omega)} \mathbb{P}\left[\omega^{\prime}\right]}
$$

This notation can be confusing $-\mathbb{E}[X \mid Z]$ is a random variable and not a number! But given $A \in \mathscr{F}$ with $\mathbb{P}[A]>0$, we denote by $\mathbb{E}[X \mid A]$ the number

$$
\mathbb{E}[X \mid A]=\frac{1}{\mathbb{P}[A]} \mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right] .
$$

As $\mathscr{G}=\sigma(Y)$, it will often be less confusing to write instead $\mathbb{E}[X \mid \mathscr{G}]$, which denotes the same random variable.

Exercise 10.1. Let $Y=\mathbb{E}[X \mid \mathscr{G}]$.

1. $Y=\operatorname{argmin}_{W \in \mathscr{L}^{2}(\Omega, \mathscr{G}, \mathbb{P})} \mathbb{E}\left[(X-W)^{2}\right]$.
2. $Y$ is $\mathscr{G}$-measurable.
3. If $A \in \mathscr{G}$ with $\mathbb{P}[A]>0$ then $\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right]=\mathbb{E}\left[Y \cdot \mathbb{1}_{\{A\}}\right]$.

### 10.3 Conditional expectations in $\mathscr{L}^{2}$

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Given a sub-sigma-algebra $\mathscr{G} \subseteq \mathscr{F}$, we know by Theorem 7.6 that the subspace $\mathscr{L}^{2}(\Omega, \mathscr{G}, \mathbb{P}) \subseteq \mathscr{L}^{2}(\Omega, \mathscr{F}, \mathbb{P})$ is closed. We can therefore define the projection operator

$$
P_{\mathscr{G}}: \mathscr{L}^{2}(\Omega, \mathscr{F}, \mathbb{P}) \rightarrow \mathscr{L}^{2}(\Omega, \mathscr{G}, \mathbb{P})
$$

by

$$
P_{\mathscr{G}}(X)=\underset{Y \in \mathscr{L}^{2}(\Omega, \mathscr{G}, \mathbb{P})}{\operatorname{argmin}} \mathbb{E}\left[(X-Y)^{2}\right] .
$$

Some immediate observations:

1. $P_{\mathscr{G}}(X)$ is $\mathscr{G}$-measurable.
2. If $Y \in \mathscr{L}^{2}(\Omega, \mathscr{F}, \mathbb{P})$ then $\mathbb{E}\left[\left(X-P_{\mathscr{G}}(X)\right) \cdot Y\right]=0$, or $\mathbb{E}[X \cdot Y]=\mathbb{E}\left[P_{\mathscr{G}}(X) \cdot Y\right]$. Thus given $A \in \mathscr{G}$ with $\mathbb{P}[A]>0$ we have that $\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right]=\mathbb{E}\left[P_{\mathscr{G}}(X) \cdot \mathbb{1}_{\{A\}}\right]$.

### 10.4 Conditional expectations in $\mathscr{L}^{1}$

Theorem 10.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with a r.v. $X \in \mathscr{L}^{1}$ and a sub-sigmaalgebra $\mathscr{G} \subseteq \mathscr{F}$. Then there exists a unique random variable $Y$ with the following properties:

1. $Y \in \mathscr{L}^{1}(\Omega, \mathscr{G}, \mathbb{P})$.
2. For every $A \in \mathscr{G}$ it holds that $\mathbb{E}\left[Y \cdot \mathbb{1}_{\{A\}}\right]=\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right]$.

We denote $\mathbb{E}[X \mid \mathscr{G}]:=Y$. For $A \in \mathscr{F}$ with $\mathbb{P}[A]>0$ we denote $\mathbb{E}[X \mid A]=\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right] / \mathbb{P}[A]$.
Proof. We first prove uniqueness. Let $Y$ and $Z$ both satisfy the two conditions in the theorem, and assume by contradiction that $\mathbb{P}[Y>Z]>0$. Then there is some $\varepsilon>0$ such that $\mathbb{P}[Y-\varepsilon>Z]>0$. Let $A=\{Y-\varepsilon>Z\}$, and note that $A \in \mathscr{G}$. Then

$$
\begin{aligned}
\mathbb{E}\left[Y \cdot \mathbb{1}_{\{A\}}\right] & =\mathbb{E}\left[(Y-\varepsilon) \cdot \mathbb{1}_{\{A\}}\right]+\varepsilon \mathbb{P}[A] \\
& >\mathbb{E}\left[Z \cdot \mathbb{1}_{\{A\}}\right]+\varepsilon \cdot \mathbb{P}[A] \\
& >\mathbb{E}\left[Z \cdot \mathbb{1}_{\{A\}}\right] .
\end{aligned}
$$

But since $A \in \mathscr{G}$ we have that both $\mathbb{P}\left[Y \cdot \mathbb{1}_{\{A\}}\right]$ and $\mathbb{P}\left[Z \cdot \mathbb{1}_{\{A\}}\right]$ are equal to $\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right]$ - contradiction.

We prove the reminder under the assumption that $X \geq 0$; the reduction is straightforward. Let $X_{n}=X \cdot \mathbb{1}_{\{X \leq n\}}$. Then $X_{n}$ is bounded, and in particular is in $\mathscr{L}^{2}$. Let $Y_{n}=P_{\mathscr{G}}\left(X_{n}\right)$. We claim that $Y$ is non-negative. To see this, assume by contradiction that $\mathbb{P}\left[Y_{n}<-\varepsilon\right]>0$ for some $\varepsilon>0$, and let $A=\left\{Y_{n}<-\varepsilon\right\}$. Then $\mathbb{E}\left[Y_{n} \cdot \mathbb{1}_{\{A\}}\right]<-\varepsilon \cdot \mathbb{P}[A]<0$, but $\mathbb{E}\left[Y_{n} \cdot \mathbb{1}_{\{A\}}\right]=$ $\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right] \geq 0$.

Now, $Y_{n}$ is a monotone increasing sequence. To see this, note that $X_{n}$ is monotone increasing, and that $P_{\mathscr{G}}$ is a linear operator, and so $Y_{n+1}-Y_{n}=P_{\mathscr{G}}\left(X_{n+1}-X_{n}\right)$ is non-negative, by the same proof as above.

Since $Y_{n}$ is monotone increasing then so is $Y_{n} \cdot \mathbb{1}_{\{A\}}$, for any $A \in \mathscr{G}$. Therefore, if we define $Y=\lim _{n} Y_{n}$, then $\mathbb{E}\left[Y_{n} \cdot \mathbb{1}_{\{A\}}\right] \rightarrow \mathbb{E}\left[Y \cdot \mathbb{1}_{\{A\}}\right]$. But $\mathbb{E}\left[Y_{n} \cdot \mathbb{1}_{\{A\}}\right]=\mathbb{E}\left[X_{n} \cdot A\right]$, and, since $X_{n} \cdot \mathbb{1}_{\{A\}}$ is also monotone increasing with $X \cdot \mathbb{1}_{\{A\}}=\lim _{n} X_{n} \cdot \mathbb{1}_{\{A\}}$, we have that

$$
\mathbb{E}\left[Y \cdot \mathbb{1}_{\{A\}}\right]=\lim _{n} \mathbb{E}\left[Y_{n} \cdot A\right]=\lim _{n} \mathbb{E}\left[X_{n} \cdot \mathbb{1}_{\{A\}}\right]=\mathbb{E}\left[X \cdot \mathbb{1}_{\{A\}}\right] .
$$

Finally, each $Y_{n}$ is $\mathscr{G}$-measurable by construction, and therefore so is $Y$.

### 10.5 Some properties of conditional expectation

Exercise 10.3. 1. If $X$ is $\mathscr{G}$-measurable (i.e., $\sigma(X) \subseteq \mathscr{G}$ ) then $\mathbb{E}[X \mid \mathscr{G}]=X$.
2. The Law of Total Expectation. If $\mathscr{G}_{2} \subseteq \mathscr{G}_{1}$ then $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathscr{G}_{1}\right] \mid \mathscr{G}_{2}\right]=\mathbb{E}\left[X \mid \mathscr{G}_{2}\right]$. In particular $\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]]=\mathbb{E}[X]$.
3. If $Z \in \mathscr{L}^{\infty}(\Omega, \mathscr{G}, \mathbb{P})$ then

$$
\mathbb{E}[Z \cdot X \mid \mathscr{G}]=Z \cdot \mathbb{E}[X \mid \mathscr{G}] .
$$

## 11 The Galton-Watson process

### 11.1 Definition

Consider an asexual organism (in the original work these were Victorian men) whose number of offspring $X_{0}$ is chosen at random from some distribution on $\mathbb{N}_{0}=\{0,1,2, \ldots\}$. Each of its descendants $i \in\left\{1, \ldots, X_{0}\right\}$ has $X_{i}$ offspring, with the random variables ( $X_{0}, X_{1}, X_{2}, \ldots$ ) distributed independently and identically. An interesting question is: what is the probability that the organism's progeny will live forever, and what is the probability that there will be a last one to its name? ${ }^{2}$

Formally, consider generations $\{1,2, \ldots\}$, and to each generation $n$ associate an infinite sequence of random variables ( $X_{n, 1}, X_{n, 2}, \ldots$ ), with all the random variables ( $X_{n, i}$ ) independent and identically distributed on $\mathbb{N}_{0}$. We will, to simplify some expressions, define $X=X_{1,1}$. We assume that

$$
0<\mathbb{E}[X]<\infty,
$$

and denote $\mu=\mathbb{E}[X]$. We also assume that $\mathbb{P}[X=0]>0$.
To each generation $n$ we define the number of organisms $Z_{n}$, which is also a random variable. It is defined recursively by $Z_{1}=1$ and $Z_{n+1}=\sum_{i=1}^{Z_{n}} X_{n, i}$. Clearly $Z_{n}=0$ implies $Z_{n+1}=0$.

### 11.2 The probability of extinction

We are interested in the event that $Z_{n}=0$ for some $n$, or that, equivalently, $Z_{n}=0$ for all $n$ large enough. This is again equivalent to the event $\sum_{n} Z_{n}<\infty$, since each $Z_{n}$ is an integer. We denote this event by $E$ (for extinction), and denote $E_{n}=\left\{Z_{n}=0\right\}$.

Note that the sequence $E_{n}$ is increasing and $E=\cup_{n} E_{n}$. Therefore, by Theorem 3.5,

$$
\mathbb{P}[E]=\lim _{n} \mathbb{P}\left[Z_{n}=0\right] .
$$

We first calculate the expectation of $Z_{n+1}$. Since $Z_{n}$ is independent of ( $X_{n, 1}, X_{n, 2}, \ldots$ ), it holds that

$$
\begin{aligned}
\mathbb{E}\left[Z_{n+1}\right] & =\mathbb{E}\left[\sum_{i=1}^{Z_{n}} X_{n, i}\right] \\
& =\mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{Z_{n}} X_{n, i} \mid Z_{n}\right]\right] \\
& =\mathbb{E}\left[Z_{n} \cdot \mathbb{E}\left[X \mid Z_{n}\right]\right] \\
& =\mathbb{E}\left[Z_{n}\right] \cdot \mathbb{E}[X],
\end{aligned}
$$

[^1]and so
$$
\mathbb{E}\left[Z_{n+1}\right]=\mu^{n} .
$$

Claim 11.1. If $\mu<1$ then $\mathbb{P}[E]=1$.
Proof 1. By Markov's inequality, $\mathbb{P}\left[Z_{n} \geq 1\right] \leq \mu^{n}$. Thus by the Borel-Cantelli Lemma w.p. 1 there will be some $n$ with $Z_{n}<1$, and thus $Z_{n}=0$.

Proof 2. Note that $\mathbb{E}\left[\sum_{n} Z_{n}\right]=\sum_{n} \mathbb{E}\left[Z_{n}\right]<\infty$, and so $\mathbb{P}\left[\sum_{n} Z_{n}=\infty\right]=0$.
It is also true that $\mathbb{P}[E]=1$ when $\mu=1$. Note that in this case

$$
\mathbb{E}\left[Z_{n+1} \mid Z_{1}, Z_{2}, \ldots, Z_{n}\right]=\mathbb{E}\left[Z_{n+1} \mid Z_{n}\right]=Z_{n} \cdot \mathbb{E}[X]=Z_{n} .
$$

The first equality makes $Z_{n}$ a Markov chain. The second makes it a Martingale; we will discuss both concepts formally. By the Martingale Convergence Theorem we have that $Z_{n}$ converges almost surely to some r.v. $Z_{\infty}$. But clearly $Z_{n}$ cannot converge to anything but 0 , and so $\mathbb{P}[E]=1$.

Note that the event $E$ is equal to the union of the event that $X_{1,1}=0$ with the event that $X_{1,1}>0$ but each of the sub-tree of the $Z_{2}$ offspring goes extinct. Since the process on each subtree is identical, and since the probability that all of such $k$ offspring trees goes extinct is $\mathbb{P}[E]^{k}$, we have that $\mathbb{P}[E]$ must satisfy

$$
\begin{equation*}
\mathbb{P}[E]=\sum_{k \in \mathbb{N}_{0}} \mathbb{P}[X=k] \mathbb{P}[E]^{k} . \tag{11.1}
\end{equation*}
$$

### 11.3 The probability generating function

We accordingly define $f:[0,1] \rightarrow[0,1]$, the probability generating function of $X$, by

$$
f(t)=\sum_{k \in \mathbb{N}_{0}} \mathbb{P}[X=k] \cdot t^{k}=\mathbb{E}\left[t^{X}\right],
$$

where we take $0^{0}=1$. Then (11.1) is equivalent to observing that $\mathbb{P}[E]$ is a fixed point of $f$. Note that 1 is always a fixed point, but in general there might be more.

Some observations:

1. $f(0)=\mathbb{P}[X=0]$ and $f(1)=1$.
2. $f^{\prime}(t)=\sum_{k \in \mathbb{N}} \mathbb{P}[X=k] \cdot k \cdot t^{k-1}=\mathbb{E}\left[X \cdot t^{X-1}\right]$. Hence $f^{\prime}(1)=\mathbb{E}[X]=\mu$. Note also that $f^{\prime}(t) \geq 0$.
3. Likewise, the $k^{\text {th }}$ derivative of $f$ is non-negative. Thus $f$ is convex.

Let $f_{n}(t)=\mathbb{E}\left[t^{Z_{n}}\right]$ be the generating function of $Z_{n}$. Then

$$
\begin{aligned}
f_{n+1}(t) & =\mathbb{E}\left[t^{Z_{n+1}}\right] \\
& =\mathbb{E}\left[\mathbb{E}\left[t^{Z_{n+1}} \mid Z_{n}\right]\right] \\
& =\mathbb{E}\left[\mathbb { E } \left[t^{\left.\left.\Sigma_{k=1}^{Z_{n} X_{n, k}} \mid Z_{n}\right]\right]}\right.\right. \\
& =\mathbb{E}\left[\mathbb{E}\left[t^{X}\right]^{Z_{n}}\right] \\
& =\mathbb{E}\left[f(t)^{Z_{n}}\right] \\
& =f_{n}(f(t)),
\end{aligned}
$$

where we again used the fact that $Z_{n}$ is independent of ( $X_{n, 1}, X_{n, 2}, \ldots$ ). Since $f_{1}(t)=t, f_{n+1}$ is the $n$-fold composition of $f$ with itself:

$$
f_{n+1}=f \circ f \circ \cdots \circ f
$$

Now $\mathbb{P}\left[Z_{n}=0\right]=f_{n}(0)$. Since $f$ is analytic,

$$
\mathbb{P}[E]=\lim _{n} \mathbb{P}\left[E_{n}\right]=\lim _{n} f_{n}(0)
$$

will be the fixed point of $f$ that one converges to by applying $f$ repeatedly to 0 . Furthermore, $f(0)=\mathbb{P}[X=0]>0, f(1)=1$, and $f$ is increasing and convex. Thus $f$ will have a unique fixed point. Finally, since $f^{\prime}(1)=\mu$, this fixed point will be 1 iff $\mu \leq 1$.

## 12 Markov chains

### 12.1 Definition

Let the state space $S$ be a countable or finite set. A sequence of $S$-valued random variables ( $X_{0}, X_{1}, X_{2}, \ldots$ ) is said to be a Markov chain if for all $x \in S$ and $n>0$

$$
\mathbb{P}\left[X_{n}=x \mid X_{0}, X_{1}, \ldots, X_{n-1}\right]=\mathbb{P}\left[X_{n}=x \mid X_{n-1}\right] .
$$

A Markov chain is said to be time homogeneous if $\mathbb{P}\left[X_{n}=x \mid X_{n-1}\right]$ does not depend on $n$. In this case it will be useful to study the associated stochastic $S$-indexed matrix $P(x, y)=$ $\mathbb{P}\left[X_{n+1}=y \mid X_{n}=x\right]$. It is easy to see that $\mathbb{P}\left[X_{n+m}=y \mid X_{n}=x\right]=P^{m}(x, y)$, where $P^{m}$ denotes the usual matrix exponentiation. We call $P$ the transition matrix of the Markov chain.

In the context of a transition matrix $P$, we will denote by $\mathbb{P}^{x}$ the measure of the Markov chain for which $\mathbb{P}\left[X_{0}=x\right]=1$.

The next claim is needed to formally apply the Markov property.
Claim 12.1. Let $\left(X_{0}, X_{1}, \ldots\right)$ be a time homogeneous Markov chain. Fix some measurable $f: S^{\mathbb{N}} \rightarrow \mathbb{R}$ and denote

$$
Y_{n}=f\left(X_{n}, X_{n+1}, \ldots\right)
$$

Then for any $n, m \in \mathbb{N}$ and $x \in S$ such that $\mathbb{P}\left[X_{n}=x\right]>0$ and $\mathbb{P}\left[X_{m}=x\right]>0$ it holds that

$$
\mathbb{E}\left[Y_{n+1} \mid X_{n}=x\right]=\mathbb{E}\left[Y_{m+1} \mid X_{m}=x\right] .
$$

Example: let $S=\mathbb{Z}$, let $X_{0}=0$, and let $P(x, y)=\frac{1}{2} \mathbb{1}_{\{|x-y|=1\}}$. This is called the simple random walk on $\mathbb{Z}$. More generally (in some direction), one can consider a graph $G=(S, E)$ with finite positive out-degrees $d(x)=|E \cap\{x\} \times S|$ and let

$$
P(x, y)=\frac{\mathbb{1}_{\{(x, y) \in E\}}}{d(x)} .
$$

The lazy random walk on $\mathbb{Z}$ has transition probabilities $P(x, y)=\frac{1}{3} \mathbb{1}_{\{|x-y| \leq 1\}}$.

### 12.2 Irreducibility and aperiodicity

We say that a (time homogeneous) Markov chain is irreducible if for all $x, y \in S$ there exists some $m$ so that $P^{m}(x, y)>0$. We say that an irreducible chain is aperiodic if for some (equivalently, every) $x \in S$ it holds that $P^{m}(x, x)>0$ for all $m$ large enough.

Exercise 12.2. Show that if an irreducible chain is not aperiodic then for every $x \in S$ there is a $k \in \mathbb{N}$ so that $P^{m}(x, x)=0$ for all $m$ not divisible by $k$.

Exercise 12.3. 1. Show that the simple random walk on $\mathbb{Z}$ is irreducible but not aperiodic.
2. Show that the lazy random walk on $\mathbb{Z}$ is irreducible and aperiodic.
3. Show that the simple random walk on a directed graph is irreducible iff the graph is strongly connected.
4. Show that the simple random walk on a connected, undirected graph is aperiodic iff the graph is not bipartite.

### 12.3 Recurrence

We define the hitting time to $x \in S$ by

$$
T_{x}=\min \left\{n>0: X_{n}=x\right\} .
$$

This is a random variable taking values in $\mathbb{N} \cup\{\infty\}$. An irreducible Markov chain is said to be recurrent if $\mathbb{P}\left[T_{x}<\infty\right]=1$ whenever $\mathbb{P}\left[X_{0}=x\right]>0$. A non-recurrent random walks is called transient.

Theorem 12.4. Fix an irreducible Markov chain with $\mathbb{P}\left[X_{0}=x\right]>0$ for all $x \in S$. Then the following are equivalent.

1. The Markov chain is recurrent.
2. For some (all) $x \in X$ it holds that

$$
\mathbb{P}\left[X_{n}=x \text { i.o. }\right]=1 .
$$

3. For some (all) $x \in X$ it holds that $\sum_{m} P^{m}(x, x)=\infty$.

Proof. Choose any $x \in S$. Since $\mathbb{P}\left[T_{x}<\infty\right]=1$, and since $\mathbb{P}\left[X_{0}=y\right]>0$ for any $y \in S$, we have that $\mathbb{P}\left[T_{x}<\infty \mid X_{0}=y\right]=1$, or that

$$
\mathbb{P}\left[X_{n}=x \text { for some } n>0 \mid X_{0}=y\right]=1 .
$$

By irreducibility we have that $\mathbb{P}\left[X_{m}=y\right]>0$ for any $m$, and so by the Markov property it follows that

$$
\mathbb{P}\left[X_{n}=x \text { for some } n>m \mid X_{m}=y\right]=1 .
$$

Summing over $y$ yields that

$$
\mathbb{P}\left[X_{n}=x \text { for some } n>m\right]=1,
$$

and so

$$
\mathbb{P}\left[X_{n}=x \text { i.o. }\right]=1 .
$$

We have thus shown that (1) implies (2).

Note that $P^{m}(x, x)=\mathbb{P}\left[X_{m}=x \mid X_{0}=x\right]$. Now, (2) implies that

$$
\mathbb{P}\left[X_{n}=x \text { i.o. } \mid X_{0}=x\right]=1
$$

and so, by Borel-Cantelli, (2) implies (3).
Finally, to show that (3) implies (1), assume that the Markov chain is transient. Then $\mathbb{P}\left[T_{x}<\infty\right]<1$, and so $\mathbb{P}\left[T_{x}<\infty \mid X_{0}=x\right]<1$. Denote the latter by $p$. Hence, by the Markov property,

$$
p=\mathbb{P}\left[X_{n}=x \text { for some } n>m \mid X_{m}=x\right] .
$$

Therefore, conditioned on $X_{0}=x$, the probability that $x$ is visited $k$ more times is $p^{k}(1-p)$. In particular the expected number of visits is finite, and since this expectation is equal to $\sum_{m} P^{m}(x, x)$, the proof is complete.

Exercise 12.5. Prove that every irreducible Markov chain over a finite state space is recurrent.
Exercise 12.6. Let $P$ be the transition matrix of a Markov chain over $S$, and for $\varepsilon>0$ let $P_{\varepsilon}=(1-\varepsilon) P+\varepsilon I$, where $I$ is the identity matrix. Thus $P_{\varepsilon}$ is the $\varepsilon$-lazified version of $P$. Consider two Markov chains over $S$ : both with $X_{0}=x$, and one with transition matrix $P$ and the other with transition matrix $P_{\varepsilon}$. Prove that either both are recurrent or both are transitive.

### 12.4 The simple random walk on $\mathbb{Z}^{d}$

Corollary 12.7. The simple random walk on $\mathbb{Z}$ is recurrent.
Proof. Note that $\mathbb{P}\left[X_{2 n+1}=0\right]=0$ and that

$$
\mathbb{P}\left[X_{2 n}=0\right]=2^{-2 n}\binom{2 n}{n}
$$

By Stirling

$$
\binom{2 n}{n} \geq \frac{2^{2 n-1}}{\sqrt{n}}
$$

and so

$$
\mathbb{P}\left[X_{2 n}=0\right] \geq \frac{1}{2 \sqrt{n}} .
$$

Hence

$$
\sum_{m} P^{m}(0,0) \geq \frac{1}{2 \sqrt{m}}=\infty
$$

and the claim follows by Theorem 12.4.

Consider now a random walk with a drift on $\mathbb{Z}$. For example, let $P(x, y)=p$ if $y=x+1$ and $P(x, y)=1-p$ if $y=x-1$. In this case, assuming $X_{0}=0, X_{n}=\sum_{k \leq n} Y_{n}$ where the $Y_{n}$ are i.i.d. r.v. with $\mathbb{P}\left[Y_{n}=1\right]=p$ and $\mathbb{P}\left[Y_{n}=-1\right]=1-p$. It follows from the strong law of large numbers that a.s. $\lim _{n} X_{n} / n=2 p-1>0$, and so in particular $\lim _{n} X_{n}=\infty$, and the random walk is transient. The same argument holds whenever the transition probabilities correspond to an $\mathscr{L}^{1}$ random variable with non-zero expectation, by the same argument (although we have yet to prove an $\mathscr{L}^{1}$ SLLN).

Exercise 12.8. Prove that the simple random walk on $\mathbb{Z}^{2}$ (given by $P(x, y)=\frac{1}{4} \mathbb{1}_{\{|x-y|=1\}}$ ) is recurrent, but that the simple random walk on $\mathbb{Z}^{d}$ (given by $P(x, y)=\frac{1}{d} \mathbb{1}_{\{|x-y|=1\}}$ ) is transient for all $d \geq 3$.

## 13 Martingales

### 13.1 Definition

A filtration $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ is a sequence of increasing sigma-algebras $\mathscr{F}_{1} \subseteq \mathscr{F}_{2} \subseteq \cdots$. A natural (and in some sense only) example is the case that $\mathscr{F}_{n}=\sigma\left(Y_{1}, \ldots, Y_{n}\right)$ for some sequence of random variables ( $Y_{1}, Y_{2}, \ldots$ ).

A process $\left(X_{1}, X_{2}, \ldots\right)$ is said to be adapted to $\Phi$ if each $X_{n}$ is $\mathscr{F}_{n}$-measurable. A sequence of real random variables ( $X_{1}, X_{2}, \ldots$ ) that is adapted to $\Phi$ and is in $\mathscr{L}^{1}$ is called a martingale with respect to $\Phi$ if for all $n \geq 1$

$$
\mathbb{E}\left[X_{n+1} \mid \mathscr{F}_{n}\right]=X_{n} .
$$

It is called a supermartingale if

$$
\mathbb{E}\left[X_{n+1} \mid \mathscr{F}_{n}\right] \leq X_{n} .
$$

Note that if $\left(X_{1}, X_{2}, \ldots\right)$ is a martingale then $\mathbb{E}\left[X_{n}\right]=\mathbb{E}\left[X_{1}\right]$ and by subtracting the constant $\mathbb{E}\left[X_{1}\right]$ from all $X_{n}$ 's we get that ( $X_{0}, X_{1}, \ldots$ ) is a martingale with $X_{0}=0$. A similar statement holds for supermartingales.

As a first example, let $W_{n}$ be i.i.d. r.v. with $\mathbb{P}\left[W_{n}=+1\right]=\mathbb{P}\left[W_{n}=-1\right]=1 / 2$, let $X_{n}=$ $\sum_{k \leq n} W_{n}$, and let $\mathscr{F}_{n}=\sigma\left(X_{1}, \ldots, X_{n}\right)$. Then $X_{n}$ is the amount of money made in $n$ fair bets (or the locations of a simple random walk on $\mathbb{Z}$ ) and is a martingale with respect to $\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$. If we set $\mathbb{P}\left[W_{n}=+1\right]=1 / 2-\varepsilon$ and $\mathbb{P}\left[W_{n}=-1\right]=1 / 2+\varepsilon$ for some $\varepsilon>0$ then $X_{n}$ is a supermartingale.

### 13.2 Examples

As a second example we introduce Pólya's urn. Consider an urn in which there are initially a single black ball and a single white ball. In each time period we reach in, pull out a ball, and then put back two balls of the same color. Formally, let $\left(Y_{1}, Y_{2}, \ldots\right)$ be i.i.d. random variables distributed uniformly over [0,1], and let the number of black balls at time $n$ be $B_{n}$, given by $B_{1}=1$ and

$$
B_{n+1}=B_{n}+\mathbb{1}_{\left\{Y_{n}<B_{n} /(n+1)\right\}} .
$$

Denote by $R_{n}=B_{n} /(n+1)$ the fraction of black balls. Then

$$
\mathbb{E}\left[R_{n+1} \mid B_{1}, \ldots, B_{n}\right]=\mathbb{E}\left[R_{n+1} \mid B_{n}\right],
$$

since the process ( $B_{1}, B_{2}, \ldots$ ) is a Markov chain. Furthermore

$$
\begin{aligned}
\mathbb{E}\left[R_{n+1} \mid B_{n}\right] & =\frac{1}{n+2} \mathbb{E}\left[B_{n+1} \mid B_{n}\right] \\
& =\frac{1}{n+2}\left(B_{n}+\frac{B_{n}}{n+1}\right) \\
& =\frac{B_{n}}{n+1} \\
& =R_{n},
\end{aligned}
$$

and so $R_{n}$ is a martingale with respect to $\mathscr{F}_{n}=\sigma\left(B_{1}, \ldots, B_{n}\right)$.
As a third example let $X \in \mathscr{L}^{2}$ (e.g., $X$ is a standard Gaussian), let $Y_{1}, Y_{2}, \ldots$ be i.i.d. in $\mathscr{L}^{2}$, and let $Z_{n}=X+Y_{n}$. This can be thought of as a model of independent measurements $\left(Z_{n}\right)$ with noise $\left(Y_{n}\right)$ of a physical quantity $(X)$. Under this interpretation,

$$
\tilde{X}_{n}=\mathbb{E}\left[X \mid Y_{1}, \ldots, Y_{n}\right]
$$

is a natural estimator of $X$. By the law of total expectations

$$
\mathbb{E}\left[\tilde{X}_{n+1} \mid Y_{1}, \ldots, Y_{n}\right]=\tilde{X}_{n}
$$

and thus $\tilde{X}_{n}$ is also a martingale.

### 13.3 Martingale convergence in $\mathscr{L}^{2}$

Theorem 13.1 (Martingale Convergence in $\left.\mathscr{L}^{2}\right)$. Let $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ be a filtration, and let $\left(X_{1}, X_{2}, \ldots\right)$ be a martingale w.r.t. $\Phi$. Furthermore, assume that there exists a $K$ such that $\mathbb{E}\left[X_{n}^{2}\right]<K$ for all $n$. Then there exists a random variable $X \in \mathscr{L}^{2}$ such that $\mathbb{E}\left[\left(X-X_{n}\right)^{2}\right] \rightarrow 0$.

Proof. Set $X_{0}=0$, and for $n \geq 1$ let $Y_{n}=X_{n}-X_{n-1}$. Since $X_{n-1}=\mathbb{E}\left[X_{n} \mid \mathscr{F}_{n-1}\right]$, we have that $Y_{n}$ is orthogonal to any $\mathscr{F}_{n-1}$-measurable r.v., and in particular is orthogonal to $Y_{m}$ for any $m<n$. Now,

$$
\sum_{k \leq n} Y_{n}=X_{n}
$$

and so by the orthogonality of the $Y_{n}$ 's it follows that

$$
\sum_{k \leq n} \mathbb{E}\left[Y_{k}^{2}\right]=\mathbb{E}\left[X_{n}^{2}\right]<K .
$$

Thus

$$
\sum_{k} \mathbb{E}\left[Y_{n}^{2}\right]<K
$$

and we have that $X_{n}$ is a Cauchy sequence in $\mathscr{L}^{2}$. Therefore, since $\mathscr{L}^{2}$ is complete (Theorem 7.6) there exists some $X \in \mathscr{L}^{2}$ such that $\mathbb{E}\left[\left(X-X_{n}\right)^{2}\right] \rightarrow 0$.

The next theorem shows that, in fact, convergence is pointwise, and an $\mathscr{L}^{1}$ assumption suffices.

### 13.4 The Martingale Convergence Theorem

Theorem 13.2 (Martingale Pointwise Convergence). Let $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ be a filtration, and let $\left(X_{1}, X_{2}, \ldots\right)$ be a supermartingale w.r.t. $\Phi$. Furthermore, assume that there exists a $K$ such that $\mathbb{E}\left[\left|X_{n}\right|\right]<K$ for all $n$. Then there exists a random variable $X \in \mathscr{L}^{1}$ such that almost surely $\lim _{n} X_{n}=X$.

Before proving this theorem we will need the following lemmas.
Lemma 13.3. Let $\left(X_{0}, X_{1}, X_{2}, \ldots\right)$ be a supermartingale w.r.t. $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ with $X_{0}=0$, let $B_{n}$ be $\{0,1\}$-values random variables adapted to $\Phi$, and let $Y_{n}=\sum_{k \leq n} B_{k-1} \cdot\left(X_{k}-X_{k-1}\right)$. Then $Y_{n}$ is a supermartingale and $\mathbb{E}\left[Y_{n}\right] \leq 0$.

The idea behind this lemma is the following: imagine that you are gambling at a casino with non-positive expected wins from every gamble. Say that you have some system for deciding when to gamble and when to stay out (i.e., the $B_{n}$ 's). Then you do not expect to win more than you would have if you stayed in the game every time.

Proof.

$$
\begin{aligned}
\mathbb{E}\left[Y_{n+1} \mid \mathscr{F}_{n}\right] & =\mathbb{E}\left[\sum_{k \leq n+1} B_{k-1} \cdot\left(X_{k}-X_{k-1}\right) \mid \mathscr{F}_{n}\right] \\
& =\mathbb{E}\left[Y_{n}+B_{n} \cdot\left(X_{n+1}-X_{n}\right) \mid \mathscr{F}_{n}\right] \\
& =Y_{n}+B_{n} \mathbb{E}\left[X_{n+1}-X_{n} \mid \mathscr{F}_{n}\right] \\
& =Y_{n}+B_{n}\left(\mathbb{E}\left[X_{n+1} \mid \mathscr{F}_{n}\right]-X_{n}\right) \\
& \leq Y_{n} .
\end{aligned}
$$

Thus $Y_{n}$ is a supermartingale, and by induction $\mathbb{E}\left[Y_{n}\right] \leq 0$.
Lemma 13.4. Let $\left(X_{0}, X_{1}, X_{2}, \ldots\right)$ be a supermartingale w.r.t. $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ with $X_{0}=0$. Fix some $a<b$, and let $B_{n}$ be defined as follows: $B_{0}=0$, and $B_{n+1}$ is the indicator of the union of the events

1. $B_{n}=1$ and $X_{n} \leq b$.
2. $B_{n}=0$ and $X_{n}<a$.

Let $U_{n}^{a, b}$ be the number of $k \leq n$ such that $B_{k}=0$ and $B_{k-1}=1$. Then

$$
\mathbb{E}\left[U_{n}^{a, b}\right] \leq \frac{\mathbb{E}\left[\left(X_{n}-a\right)^{-}\right]}{b-a}
$$

Proof. By picture, it is clear that for

$$
Y_{n}=\sum_{k \leq n} B_{n-1} \cdot\left(X_{n}-X_{n-1}\right)
$$

it holds that

$$
Y_{n} \geq(b-a) U_{n}^{a, b}-\left(X_{n}-a\right)^{-} .
$$

By Lemma 13.3 we have that $\mathbb{E}\left[Y_{n}\right] \leq 0$, and so the claim follows by taking expectations.

Proof of Theorem 13.2. For a given $a<b$, let $U_{\infty}^{a, b}=\lim _{n} U_{n}^{a, b}$. The limit exists since this is a monotone increasing sequence, and it also follows that

$$
\mathbb{E}\left[U_{\infty}^{a, b}\right]=\lim _{n} \mathbb{E}\left[U_{n}^{a, b}\right] \leq \lim _{n} \frac{\mathbb{E}\left[\left(X_{n}-a\right)^{-}\right]}{b-a} \leq \frac{|a|+K}{b-a}<\infty .
$$

Thus $\mathbb{P}\left[U_{\infty}^{a, b}<\infty\right]=1$, and it follows that with probability zero it occurs that $\limsup _{n} X_{n} \geq b$ and $\liminf _{n} X_{n} \leq a$. Applying this to a countable dense set of pairs ( $a, b$ ) we get that with probability zero $\limsup X_{n} X_{n}>\liminf _{n} X_{n}$, and so $\limsup { }_{n} X_{n}=\liminf _{n} X_{n}$ almost surely.

Exercise 13.5. Let $R_{n}$ be the fraction of black balls in Pólya's urn. Show that $\lim _{n} R_{n}$ is distributed uniformly on $(0,1)$. Hint: calculate the distribution of $R_{n}$.

## 14 Stopping times

### 14.1 Definition

Let $\Phi=\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ be a filtration, and let $\left(X_{0}, X_{1}, X_{2}, \ldots\right)$ be a supermartingale w.r.t. $\Phi$. Denote $\mathscr{F}_{\infty}=\sigma\left(\cup_{n} \mathscr{F}_{n}\right)$.

A random variable $T$ taking values in $\mathbb{N} \cup\{\infty\}$ is called a stopping time if for all $n \leq \infty$ it holds that the event $\{T \leq n\}$ is $\mathscr{F}_{n}$-measurable.

Example: $\left(X_{1}, X_{2}, \ldots\right)$ is a Markov chain over the state space $S$, and $T_{x}$ is the hitting time to $x \in S$ given by

$$
T_{x}=\min \left\{n>0: X_{n}=x\right\} .
$$

Example: $\left(X_{1}, X_{2}, \ldots\right)$ is the simple random walk on $\mathbb{Z}$, and $T$ is the first $n \geq 3$ such that $X_{n}<X_{n-1}<X_{n-2}$.

Given a stopping time $T$, we define the stopped process

$$
\left(X_{1}^{T}, X_{2}^{T}, \ldots\right)=\left(X_{1}, X_{2}, \ldots, X_{T-1}, X_{T}, X_{T}, \ldots\right)
$$

That is, $X_{n}^{T}=X_{n}$ if $n \leq T$, and $X_{n}^{T}=X_{T}$ if $n \geq T$. Equivalently, $X_{n}^{T}=X_{\min \{T, n\}}$. Intuitively, the stopped process corresponds to the process of a gambler's bank account, when the gambler decides stopping at time $T$.

Theorem 14.1. If $\left(X_{0}, X_{1}, X_{2}, \ldots\right)$ is a (super)martingale (with $X_{0}=0$ ) then ( $X_{0}^{T}, X_{1}^{T}, X_{2}^{T}, \ldots$ ) is a (super)martingale.

Proof. We prove for the case of supermartingales; the proof for martingales is identical.
Let $B_{n}=\mathbb{1}_{\{T \geq n\}}$ and $Y_{n}=\sum_{k \leq n} B_{k-1} \cdot\left(X_{k}-X_{k-1}\right)$. Then by Lemma 13.3 we have that $Y_{n}$ is a supermartingale. But $Y_{n}=X_{n}^{T}$.

So the gambler's bank account is still a martingale, no matter what the stopping time is, and in particular $\mathbb{E}\left[X_{n}^{T}\right] \leq 0$ (with equality for martingales). However, consider a simple random walk on $\mathbb{Z}$, with stopping time $T_{1}$. That is, the gambler stops once she has earned a dollar. Then clearly $\mathbb{E}\left[X_{T_{1}}\right]=1$. The following theorem gives conditions for when $\mathbb{E}\left[X_{T}\right]=0$.

### 14.2 Optional Stopping Time Theorem

Theorem 14.2 (Doob's Optional Stopping Time Theorem). Let $\left(X_{0}, X_{1}, \ldots\right)$ be a supermartingale, and let $T$ be a stopping time. Assume that $\mathbb{P}[T=\infty]=0$, and that one of following holds:

1. $\exists N$ s.t. $\mathbb{P}[T \leq N]=1$.
2. $\exists K$ s.t. $\mathbb{P}\left[\left|X_{n}\right| \leq K\right.$ for all $\left.n\right]=1$.
3. $\mathbb{E}[T]<\infty$ and $\exists K$ s.t. $\mathbb{P}\left[\left|X_{n+1}-X_{n}\right| \leq K\right.$ for all $\left.n\right]=1$.
4. $X_{n}$ is non-negative.

Then $\mathbb{E}\left[X_{T}\right] \leq \mathbb{E}\left[X_{0}\right]$, with equality if $\left(X_{0}, X_{1}, \ldots\right)$ is a martingale.
To prove this theorem we will need the following important lemma.
Lemma 14.3 (Fatou's Lemma). Let $\left(Z_{1}, Z_{2}, \ldots\right)$ be a sequence of non-negative real random variables. Then

$$
\mathbb{E}\left[\liminf _{n} Z_{n}\right] \leq \liminf _{n} \mathbb{E}\left[Z_{n}\right] .
$$

Recall from the Galton-Watson example that indeed this may be a strict inequality.
Exercise 14.4. Prove Fatou's Lemma. Hint: use the Monotone Convergence Theorem.
Proof. We prove that $\mathbb{E}\left[X_{T}\right] \leq 0$; the equality in case of the martingales follows easily.
Note that $\mathbb{E}\left[X_{n}^{T}\right] \leq 0$, by Theorem 14.1. Also $\lim _{n} X_{n}^{T}=X_{T}$, since $\mathbb{P}[T<\infty]=1$ under all conditions.

1. $X_{T}=X_{N}^{T}$.
2. By the Bounded Convergence Theorem $\mathbb{E}\left[X^{T}\right]=\lim _{n} \mathbb{E}\left[X_{n}^{T}\right] \leq 0$.
3. 

$$
\left|X_{n}^{T}\right|=\left|\sum_{k=1}^{\min \{T, n\}} X_{k}-X_{k-1}\right| \leq K \cdot T
$$

Hence by the Dominated Convergence Theorem $\mathbb{E}\left[X_{T}\right]=\mathbb{E}\left[X_{n}^{T}\right]$.
4. By Fatou's Lemma,

$$
\mathbb{E}\left[X_{T}\right] \leq \liminf _{n} \mathbb{E}\left[X_{n}^{T}\right] \leq 0
$$

Corollary 14.5. Let $T_{1}$ be the hitting time to 1 of the simple random walk on $\mathbb{Z}$. Then $\mathbb{E}\left[T_{1}\right]=\infty$.

## 15 Harmonic and superharmonic functions

### 15.1 Definition

Let ( $X_{0}, X_{1}, \ldots$ ) be a Markov chain over the state space $S$ with transition matrix $P$. We say that a function $f: S \rightarrow \mathbb{R}$ is $P$-harmonic if $P f=f$. Here $P f: S \rightarrow \mathbb{R}$ is

$$
[P f](x)=\sum_{y \in S} P(x, y) f(y)
$$

We say that $f$ is $P$-superharmonic if $[P f](x) \leq f(x)$ for all $x \in S$.

### 15.2 Harmonic functions and martingales

Claim 15.1. Assume that $P$ is irreducible, so that for all $x$ there exists an $n$ such that $\mathbb{P}\left[X_{n}=x\right]>0$. Let $Z_{n}=f\left(X_{n}\right)$. Then $Z_{n}$ is a (super)martingale iff $f$ is (super)harmonic.

Proof. We prove for the (super) case:

$$
\mathbb{E}\left[f\left(X_{n+1}\right) \mid X_{0}, \ldots, X_{n}\right]=\mathbb{E}\left[f\left(X_{n+1}\right) \mid X_{n}\right]=\sum_{y \in S} P\left(X_{n}, y\right) f(y) \leq f\left(X_{n}\right)
$$

iff $f$ is superharmonic.

### 15.3 Superharmonic functions and recurrence

Theorem 15.2. Let $P$ be irreducible. Then the following are equivalent.

1. Every Markov chain with transition matrix $P$ is recurrent.
2. Some Markov chain with transition matrix $P$ is recurrent.
3. Every non-negative $P$-superharmonic function is constant.

Proof. The equivalence of (1) and (2) follows easily from Theorem 12.4.
To see that (1) implies (3), let $T_{y}$ be the hitting time to $y$, and note that $\mathbb{P}_{x}\left[T_{y}<\infty\right]=1$, by recurrence. Let $f$ be a non-negative superharmonic function, and let $Z_{n}=f\left(X_{n}\right)$. Then we can apply the Optional Stopping Time Theorem to $Z_{n}^{T_{y}}$ to get that

$$
\mathbb{E}_{x}\left[Z_{T_{y}}\right] \leq \mathbb{E}_{x}\left[Z_{0}\right] .
$$

The l.h.s. is equal to $f(y)$ and the r.h.s. is equal to $f(x)$, and so $f$ is constant.

Assume (3), and note that

$$
\begin{aligned}
\mathbb{P}_{x}\left[T_{y}<\infty\right] & =\mathbb{P}_{x}\left[X_{1}=y, T_{y}<\infty\right]+\mathbb{P}_{x}\left[X_{1} \neq y, T_{y}<\infty\right] \\
& =\mathbb{P}_{x}\left[X_{1}=y\right]+\sum_{z \neq y} \mathbb{P}_{x}\left[X_{1}=z, T_{y}<\infty\right] \\
& =\mathbb{P}_{x}\left[X_{1}=y\right]+\sum_{z \neq y} \mathbb{P}_{x}\left[T_{y}<\infty \mid X_{1}=z\right] \cdot \mathbb{P}_{x}\left[X_{1}=z\right] \\
& =P(x, y)+\sum_{z \neq y} P(x, z) \cdot \mathbb{P}_{z}\left[T_{y}<\infty\right] \\
& \geq \sum_{z} P(x, z) \cdot \mathbb{P}_{z}\left[T_{y}<\infty\right] .
\end{aligned}
$$

Hence $f(x)=\mathbb{P}_{x}\left[T_{y}<\infty\right]$ is superharmonic, and thus constant by assumption. Say $p=$ $\mathbb{P}_{x}\left[T_{y}<\infty\right]$. By irreducibility $p>0$. Hence, by the Markov property, for every $N$ the expected number of visits at times $n>N$ is at least $p$, and so the expected number of visits is infinite. Thus the random walk is recurrent.

### 15.4 Bounded harmonic functions

The following claim is a direct consequence of Claim 15.1 and the Martingale Convergence Theorem.

Claim 15.3. Let $f: S \rightarrow \mathbb{R}$ be bounded and superharmonic. Then $Z_{n}=f\left(X_{n}\right)$ is a bounded supermartingale and therefore converges almost surely to $Z:=\lim _{n} Z_{n}$.

Recall that $\mathscr{T}_{n}=\sigma\left(X_{n}, X_{n+1}, \ldots\right)$ and that

$$
\mathscr{T}=\cap_{n} \mathscr{T}_{n}
$$

is the tail sigma-algebra. We think of our probability space as being $(\Omega, \mathscr{F}, \mathbb{P})$ with $\Omega=S^{\mathbb{N}}$ and $\mathscr{F}$ the Borel sigma-algebra of the product of the discrete topologies. Then $A \in \mathscr{T}_{n}$ iff $A$ is of the form $S^{n} \times B$ for some measurable $B \in \mathscr{F}$. Equivalently, $A \in \mathscr{T}_{n}$ iff for every $\left(x_{0}, x_{1}, \ldots\right) \in A$, and $\left(y_{0}, \ldots, y_{n-1}\right) \in S^{n}$ it holds that

$$
\left(y_{0}, \ldots, y_{n-1}, x_{n}, x_{n+1}, \ldots\right) \in A
$$

### 15.5 The shift-invariant sigma-algebra

Another important sigma-algebra is the shift-invariant sigma-algebra $\mathscr{I}$. To define it, let $\varphi: S^{\mathbb{N}} \rightarrow S^{\mathbb{N}}$ be the shift map given by $\varphi\left(x_{0}, x_{1}, x_{2}, \ldots\right)=\left(x_{1}, x_{2}, \ldots\right)$. Then $\mathscr{I}$ is the $\varphi$-invariant subsets of $S^{\mathbb{N}}$. That is,

$$
\mathscr{I}=\left\{A \subset S^{\mathbb{N}}: \varphi^{-1}(A)=A\right\} .
$$

Equivalently, $A \in \mathscr{I}$ iff for every $\left(x_{0}, x_{1}, \ldots\right) \in A$, and $y \in S$ it holds that

$$
\left(y, x_{1}, x_{2}, \ldots\right) \in A
$$

and

$$
\left(x_{2}, x_{3}, \ldots\right) \in A .
$$

Exercise 15.4. Show that $\mathscr{I} \subset \mathscr{T}$, but that the two are not equal.
Exercise 15.5. Find an irreducible Markov chain on the state space $\mathbb{N}$ that has a random variable that is $\mathscr{T}$-measurable but not $\mathscr{I}$-measurable.

Claim 15.6. Let $Z=\lim _{n} f\left(X_{n}\right)$ for some bounded harmonic $f$. Then $Z$ is $\mathscr{I}$-measurable.
The proof is a (perhaps tedious) application of the definition.
Since $Z$ is bounded we have that $Z \in \mathscr{L}^{\infty}(\mathscr{I})$.
Theorem 15.7. For every $Z \in \mathscr{L}^{\infty}(\mathscr{I})$ there is a bounded harmonic function $f$ such that $Z=\lim _{n} f\left(X_{n}\right)$.

Proof. For $x \in S$ choose any $n$ such that $\mathbb{P}\left[X_{n}=x\right]>0$, and let $f(x)=\mathbb{E}\left[Z \mid X_{n}=x\right]$. This is well defined (i.e., independent of the choice of $n$ ) because $Z$ is $\mathscr{\mathscr { L }}$-measurable. It is straightforward to check that $f$ is bounded.

If $\mathbb{P}\left[X_{n}=x\right]>0$ then $\mathbb{P}\left[X_{n+1}=y\right]>0$ for all $y$ such that $P(x, y)>0$, and so

$$
\begin{aligned}
{[P f](x) } & =\sum_{y} P(x, y) f(y) \\
& =\sum_{y} \mathbb{P}\left[X_{n+1}=y \mid X_{n}=x\right] \cdot \mathbb{E}\left[Z \mid X_{n+1}=y\right] \\
& =\sum_{y} \mathbb{P}\left[X_{n+1}=y \mid X_{n}=x\right] \cdot \mathbb{E}\left[Z \mid X_{n+1}=y, X_{n}=x\right] \\
& =\mathbb{E}\left[Z \mid X_{n}=x\right] \\
& =f(x),
\end{aligned}
$$

where the equality before last follows from the Markov property. Thus $f$ is harmonic.
To see that $Z=\lim _{n} f\left(X_{n}\right)$, note that, by the martingale convergence theorem,

$$
Z=\lim _{n} \mathbb{E}\left[Z \mid X_{1}, \ldots, X_{n}\right] .
$$

By the Markov property

$$
\mathbb{E}\left[Z \mid X_{1}, \ldots, X_{n}\right]=\mathbb{E}\left[Z \mid X_{n}\right]=f\left(X_{n}\right),
$$

and so $Z=\lim _{n} f\left(X_{n}\right)$.
To summarize, denote by $h^{\infty}(S, P) \subset \ell^{\infty}(S)$ the bounded $P$-harmonic functions. If $f \in$ $h^{\infty}(S, P)$, then $Z=\lim _{n} f\left(X_{n}\right)$ is in $\mathscr{L}^{\infty}(\mathscr{I})$. Conversely, if $Z \in \mathscr{L}^{\infty}(\mathscr{I})$ then $f=\mathbb{E}\left[Z \mid X_{n}=x\right]$ is in $h^{\infty}(S, P)$.

It turns out that the map $\Phi: \mathscr{L}^{\infty}(\mathscr{T}) \rightarrow h^{\infty}(S, P)$ given by $\Phi: Z \mapsto f$ is a linear isometry.

## 16 The Choquet-Deny Theorem

### 16.1 The asymptotic direction of a random walk

As motivation, consider the simple random walk $\left(X_{1}, X_{2}, \ldots\right)$ on $\mathbb{Z}^{3}$. Let $P_{n}=X_{n} /\left|X_{n}\right|$ be the projection of $X_{n}$ to the unit sphere (and assume $P_{n}=0$ whenever $X_{n}=0$ ). Since this random walk is transient, it is easy to deduce that $\lim _{n}\left|X_{n}\right|=\infty$. It follows that $\lim _{n}\left|P_{n+1}-P_{n}\right|=0$; that is, the projection moves more and more slowly. A natural question is: does $P_{n}$ converge?

Theorem 16.1 (Choquet-Deny Theorem). Let $\left(Y_{1}, Y_{2}, \ldots\right)$ be i.i.d. random variables taking values in some countable abelian group $G$. Let $X_{n}=\sum_{k \leq n} Y_{n}$. Then ( $X_{1}, X_{2}, \ldots$ ) is a time homogeneous Markov chain over the state space $G$. If ( $X_{1}, X_{2}, \ldots$ ) is also irreducible then every $W \in \mathscr{L}^{\infty}(\mathscr{I})$ is constant.

For the proof of this theorem we will need an important classical result from convex analysis.

### 16.2 The Krein-Milman Theorem

Theorem 16.2 (Krein-Milman Theorem). Let X be a Hausdorff locally convex topological space. A point $x \in C$ is extreme if whenever $x$ is equal to the non-trivial convex combination $\alpha y+(1-\alpha) z$ then $y=z$.

Let $C$ be compact convex subset of $X$. Then every $x \in C$ can be written as the limit of convex combinations of extreme points in $C$.

Proof of Theorem 16.1. Denote by $P$ the transition matrix of ( $X_{1}, X_{2}, \ldots$ ), and let $\mu(g)=$ $\mathbb{P}\left[Y_{n}=g\right]$. Then $P(g, k)=\mu(k-g)$. Thus, if $f$ is $P$-harmonic then

$$
f(g)=\sum_{k \in G} f(k) P(g, k)=\sum_{k \in G} f(k) \mu(k-g)=\sum_{k \in G} f(g+k) \mu(k) .
$$

Let $H=h^{[0,1]}(G, P)$ be the set of all $P$-harmonic functions with range in $[0,1]$. We note that harmonicity is invariant to multiplication by a constant and addition, and so if we show that every $f \in h^{[0,1]}(G, P)$ is constant then we have shown that every $f \in H$ is constant. It then follows that every $W \in \mathscr{L}^{\infty}(\mathscr{I})$ is constant, by the fact that $\Phi$ is an isometry.

We state three properties of $H$ that are easy to verify.

1. $H$ is invariant to the $G$ action: for any $f \in H$ and $g \in G$, the function $f^{g}: G \rightarrow \mathbb{R}$ given by $\left[f^{g}\right](k)=f(k-g)$ is also in $H$.
2. $H$ is compact in the topology of pointwise convergence.
3. $H$ is convex.

As a convex compact space, $H$ is the closed convex hull of its extreme points; this is the Krein-Milman Theorem. Thus $H$ has extreme points. Let $f \in H$ be an extreme point. Then, since $f$ is harmonic,

$$
f(g)=\sum_{k \in G} f(g+k) \mu(k)=\sum_{k \in G} f^{-k}(g) \mu(k) .
$$

By the first property of $H$ each $f^{-k}$ is also in $H$, and thus we have written $f$ as a convex combination of functions in $H$. But $f$ is extreme, and so $f=f^{-k}$ for all $k$ in the support of $\mu$. But since the Markov chain is irreducible, the support of $\mu$ generates $G$. Hence $f$ is invariant to the $G$-action, and therefore constant.

An immediate corollary of the Choquet-Deny Theorem is that every event in $\mathscr{I}$ has probability either 0 or 1 . As an application, consider the question on the simple random walk on $\mathbb{Z}^{3}$. We would like to show that $P_{n}$ does not converge pointwise. Note that the event that $P_{n}$ converges is a shift-invariant event, and therefore has measure in $\{0,1\}$. Assume by contradiction that it has measure 1, and let $P=\lim _{n} P_{n}$. For each Borel subset $B$ of the sphere, the event that $P \in B$ is shift-invariant, and therefore has measure in $\{0,1\}$. It follows that there is some $p$ such that $P=p$ almost surely. But by the symmetry of the problem the probability that $P=p$ is the same as he probability that $P=-p$, which is impossible.

Exercise 16.3. Derive Kolmogorov's zero-one law from Theorem 16.1.

## 17 Basics of information theory

### 17.1 Shannon entropy

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $X$ be a simple random variable taking values in some measurable space $(\Theta, \mathscr{G})$. We define the Shannon entropy of $X$ by

$$
H(X)=-\sum_{\theta \in \Theta} \mathbb{P}[X=\theta] \log \mathbb{P}[X=\theta],
$$

where we use the convention $0 \log 0=0$. Since $X$ is simple, there is some finite subset of $\Theta$, which we will denote by $\operatorname{supp} X$, for which $\mathbb{P}[X=\theta]>0$. Furthermore, $\sum_{\theta \text { in }} \mathbb{P}[X=\theta]=1$.

Denote by $\mathbb{P}[X]$ the random variable given by $\mathbb{P}[X](\omega)=\mathbb{P}[X=X(\omega)]$. Then we can write the entropy as

$$
H(X)=\mathbb{E}[-\log \mathbb{P}[X]] .
$$

Exercise 17.1. Show that if $|\operatorname{supp} X|=n$ then $H(X) \leq \log n$, with equality iff $X$ is distributed uniformly on its support. Hint: use Jensen's inequality, and the $\ell^{1}-\ell^{2}$ inequality, which states that for every $z \in \mathbb{R}^{n}$ it holds that $\|x\|_{1} \leq \sqrt{n}\|x\|_{2}$.

The first important property of Shannon entropy is the following form of monotonicity:
Claim 17.2. Let $X, Y$ be simple random variables. Suppose $Y$ is $\sigma(X)$-measurable (i.e, $Y=f(X)$ for some function $f)$. Then $H(Y) \leq H(X)$.

Proof. Note that $\mathbb{P}[Y] \leq \mathbb{P}[X]$ almost surely. Hence

$$
H(Y)=\mathbb{E}[-\log \mathbb{P}[Y]] \leq \mathbb{E}[-\log \mathbb{P}[X]]=H(X) .
$$

Given two random variables $X$ and $X^{\prime}$ taking values in $\Theta, \Theta^{\prime}$, we can consider the pair ( $X, X^{\prime}$ ) as a single random variable taking values in $\Theta \times \Theta^{\prime}$. We denote the entropy of this random variable as $H\left(X, X^{\prime}\right)$. The second important property of Shannon entropy is additivity with respect to independent random variables.

Claim 17.3. Let $X, Y$ be independent simple random variables. Then $H(X, Y)=H(X)+H(Y)$.
Proof. By independence, $\mathbb{P}[X, Y]=\mathbb{P}[X] \cdot \mathbb{P}[Y]$. Hence

$$
H(X, Y)=\mathbb{E}[-\log \mathbb{P}[X, Y]]=\mathbb{E}[-\log \mathbb{P}[X]-\log \mathbb{P}[Y]]=H(X)+H(Y) .
$$

### 17.2 Conditional Shannon entropy

Let $\mathscr{G}$ be a sub-sigma-algebra of $\mathscr{F}$. For a simple random variable $X$, define the random variable $\mathbb{P}[X \mid \mathscr{G}](\omega)=\mathbb{P}[X=X(\omega) \mid \mathscr{G}](\omega)$, and denote the conditional Shannon entropy by

$$
H(X \mid \mathscr{G})=\mathbb{E}[-\log \mathbb{P}[X \mid \mathscr{G}]] .
$$

For a simple random variable $X$ and any random variable $Y$, we denote $H(X \mid Y)=H(X \mid \sigma(Y))$.
Claim 17.4. $H(X \mid \mathscr{G}) \leq H(X)$, with equality if and only if $X$ is independent of $\mathscr{G}$.
Proof. By the law of total expectation, $\mathbb{P}[X \mid \mathscr{G}]=\mathbb{E}[\mathbb{P}[X] \mid \mathscr{G}]$. Since $x \mapsto-\log (x)$ is a convex function, it follows from Jensen's inequality that

$$
\begin{aligned}
H(X \mid \mathscr{G}) & =\mathbb{E}[-\log \mathbb{P}[X \mid \mathscr{G}]] \\
& =\mathbb{E}[-\log \mathbb{E}[\mathbb{P}[X] \mid \mathscr{G}]] \\
& \leq \mathbb{E}[\mathbb{E}[-\log \mathbb{P}[X] \mid \mathscr{G}]] \\
& =\mathbb{E}[-\log \mathbb{P}[X]] \\
& =H(X) .
\end{aligned}
$$

When $X$ is independent of $\mathscr{G}, \mathbb{P}[X]=\mathbb{P}[X \mid \mathscr{G}]$, and we therefore have equality. It thus remains to be shown if $X$ is not independent of $\mathscr{G}$ then the inequality is strict. Indeed, in that case $\mathbb{P}[X] \neq \mathbb{P}[X \mid \mathscr{G}]$ with positive probability, and thus Jensen's inequality is strict with positive probability, from which it follows that our inequality is also strict.

The same proof shows more generally that if $\mathscr{G}_{1} \subseteq \mathscr{G}_{2}$ then $H\left(X \mid \mathscr{G}_{1}\right) \geq H\left(X \mid \mathscr{G}_{2}\right)$.
Exercise 17.5. Suppose $\mathscr{G}=\cap_{i=n}^{\infty} \mathscr{G}_{n}$, and $\mathscr{G}_{n+1} \subseteq \mathscr{G}_{n}$. Then

$$
H(X \mid \mathscr{G})=\lim _{n} H\left(X \mid \mathscr{G}_{n}\right)=\sup _{n} H\left(X \mid \mathscr{G}_{n}\right) .
$$

### 17.3 Mutual information

We denote the mutual information of $X$ and $\mathscr{G}$ by $I(X ; \mathscr{G})=H(X)-H(X \mid \mathscr{G})$. By the above, $I$ is non-negative, and is equal to 0 if and only if $X$ is independent of $\mathscr{G}$. For two random variables $X, Y$, we denote $I(X ; Y)=I(X ; \sigma(Y))$.

Claim 17.6. Let $X, Y$ be simple random variables. Then

$$
I(X ; Y)=H(X)+H(Y)-H(X, Y)=I(Y ; X)
$$

Proof. By definition,

$$
I(X ; Y)=\mathbb{E}[-\log \mathbb{P}[X]]-\mathbb{E}[-\log \mathbb{P}[X \mid Y]]
$$

By Bayes' Law, $\mathbb{P}[X \mid Y] \mathbb{P}[Y]=\mathbb{P}[X, Y]$. Hence $\log \mathbb{P}[X \mid Y]=\log \mathbb{P}[X, Y]-\log \mathbb{P}[Y]$, and

$$
\begin{aligned}
I(X ; Y) & =\mathbb{E}[-\log \mathbb{P}[X]]-\mathbb{E}[-\log \mathbb{P}[X, Y]+\log \mathbb{P}[Y]] \\
& =\mathbb{E}[-\log \mathbb{P}[X]]-\mathbb{E}[-\log \mathbb{P}[X, Y]]+\mathbb{E}[-\log \mathbb{P}[Y]] \\
& =H(X)-H(X, Y)+H(Y) .
\end{aligned}
$$

It follows that

$$
H(X \mid Y)=H(X)-I(X ; Y)=H(X)-I(Y ; X)=H(X)+H(Y \mid X)-H(Y)
$$

and so

$$
\begin{equation*}
H(X \mid Y)=H(Y \mid X)-H(Y)+H(X) \tag{17.1}
\end{equation*}
$$

### 17.4 The information processing inequality

Let $X_{1}, X_{2}, X_{3}, \ldots$ be a Markov chain, with each $X_{n}$ simple.
Claim 17.7. $I\left(X_{3} ; X_{1}, X_{2}\right)=I\left(X_{3} ; X_{2}\right)$. Likewise, for $m>n, I\left(X_{n} ; \sigma\left(X_{m}, X_{m+1}, \ldots\right)\right)=I\left(X_{n} ; X_{m}\right)$.
The claim is a consequence of the fact that by the Markov property, $\mathbb{P}\left[X_{3} \mid X_{1}, X_{2}\right]=$ $\mathbb{P}\left[X_{3} \mid X_{2}\right]$.

## 18 Random walks on groups

### 18.1 Finitely generated groups

Let $G$ be a group. We will denote the group operation by $g \cdot h$, or just $g h$. A subset $S$ of $G$ is said to generate $G$ if each $g \in G$ is equal to a product of elements in $S$. We say that $G$ is finitely generated if it admits a finite generating set $S$. Note that finitely generated groups are countable.

Example 18.1. $\quad G=\mathbb{Z}$, where the operation is addition, and $S=\{-1,1\}$.

- $\mathrm{SL}(2, \mathbb{Z})$ : the integer matrices with determinant 1 , with the operation of matrix multiplication. It is not immediate that the following is a generating set:

$$
\left(\begin{array}{cc}
0 & -1 \\
1 & 0
\end{array}\right) \quad\left(\begin{array}{cc}
0 & 1 \\
-1 & 0
\end{array}\right) \quad\left(\begin{array}{ll}
1 & 1 \\
0 & 1
\end{array}\right) \quad\left(\begin{array}{cc}
1 & -1 \\
0 & 1
\end{array}\right)
$$

- The free group generated by $\left\{a, b, a^{-1}, b^{-1}\right\}$.


### 18.2 Random walks on finitely generated groups

Let $G$ be finitely generated by $S$. Let $X_{1}, X_{2}, \ldots$ be i.i.d. random variables taking value in $G$, and such that $\mathbb{P}\left[X_{n}=g\right]>0$ iff $g \in S$. Denote by $\mu(g)=\mathbb{P}\left[X_{n}=g\right]$ the distribution of $X_{n}$. Let $Z_{n}=X_{1} \cdot X_{2} \cdots X_{n}$. Note that $Z$ is a Markov chain, with transition matrix $P(g, h)=\mu\left(g^{-1} h\right)$.

### 18.3 Random walk entropy

Let $h_{n}=\frac{1}{n} H\left(X_{1}, \ldots, X_{n}\right)$.
Claim 18.2. $H\left(Z_{n+m}\right) \leq H\left(Z_{n}\right)+H\left(Z_{m}\right)$.
Proof.

$$
Z_{n+m}=\left(X_{1} \cdots X_{n}\right) \cdot\left(X_{n+1} \cdots X_{n+m}\right),
$$

and so

$$
H\left(Z_{n+m}\right) \leq H\left(X_{1} \cdots X_{n}, X_{n+1} \cdots X_{n+m}\right)
$$

These two random variables are independent, and so

$$
H\left(Z_{n+m}\right) \leq H\left(X_{1} \cdots X_{n}\right)+H\left(X_{n+1} \cdots X_{n+m}\right) .
$$

The distribution of $Z_{m}=X_{1} \cdots X_{m}$ is identical to that of $X_{n+1} \cdots X_{n+m}$, and so

$$
H\left(Z_{n+m}\right) \leq H\left(Z_{n}\right)+H\left(Z_{m}\right) .
$$

This claim shows that the sequence $H\left(Z_{n}\right)$ is subadditive. Fekete's Subadditive Lemma states that if $\left(a_{n}\right)_{n}$ is a subadditive sequence then $\frac{a_{n}}{n}$ converges, and that furthermore

$$
\lim _{n} \frac{a_{n}}{n}=\lim _{n} a_{n+1}-a_{n} .
$$

. We accordingly define the random walk entropy $h_{\mu}$ by

$$
h(\mu)=\lim _{n \rightarrow \infty} \frac{1}{n} H\left(Z_{n}\right) .
$$

Note that $\frac{1}{n} H\left(Z_{n}\right) \leq \frac{1}{n} H\left(X_{1}, \ldots, X_{n}\right)=H\left(X_{1}\right)$, and thus $h(\mu)$ is finite.

### 18.4 The Kaimanovich-Vershik Theorem

Theorem 18.3. The Markov chain $Z_{1}, Z_{2}, \ldots$ has a trivial tail sigma-algebra if and only if $h(\mu)=0$.

Proof. We calculate the mutual information $I\left(Z_{1} ; \mathscr{T}\right)$, where $\mathscr{T}$ is the tail sigma-algebra. Recall that $\mathscr{T}=\cap_{n} \mathscr{T}_{n}$, where $\mathscr{T}_{n}=\sigma\left(Z_{n}, Z_{n+1}, \ldots\right)$. Hence, by Exercise 17.5,

$$
H\left(Z_{1} \mid \mathscr{T}\right)=\lim _{n} H\left(Z_{1} \mid Z_{n}, Z_{n+1}, \ldots\right) .
$$

By the Markov property it follows that

$$
H\left(Z_{1} \mid \mathscr{T}\right)=\lim _{n} H\left(Z_{1} \mid Z_{n}\right) .
$$

By (17.1)

$$
H\left(Z_{1} \mid \mathscr{T}\right)=\lim _{n} H\left(Z_{n} \mid Z_{1}\right)-H\left(Z_{n}\right)+H\left(Z_{1}\right) .
$$

Now, $Z_{1}=X_{1}$, and $Z_{n}=X_{1} \cdots X_{n}$, and so

$$
H\left(Z_{1} \mid \mathscr{T}\right)=\lim _{n} H\left(X_{1} \cdots X_{n} \mid X_{1}\right)-H\left(Z_{n}\right)+H\left(Z_{1}\right) .
$$

Note that conditioned on $X_{1}=g$, the distribution of $X_{1} \cdots X_{n}$ is identical to the distribution of $g X_{1} \cdots X_{n-1}$, which has the same entropy as $X_{1}, \ldots, X_{n-1}=Z_{n-1}$. Hence $H\left(X_{1} \cdots X_{n} \mid X_{1}\right)=$ $H\left(Z_{n-1}\right)$, and we get that

$$
H\left(Z_{1} \mid \mathscr{T}\right)=\lim _{n} H\left(Z_{n-1}\right)-H\left(Z_{n}\right)+H\left(Z_{1}\right) .
$$

Thus

$$
I\left(Z_{1} ; \mathscr{T}\right)=\lim _{n} H\left(Z_{n}\right)-H\left(Z_{n-1}\right)=h(\mu) .
$$

It follows that if $h(\mu)>0$ then $\mathscr{T}$ is not independent of $Z_{1}$, and in particular $\mathscr{T}$ is non-trivial.
For the other direction, a calculation similar to the one above shows that $I\left(Z_{1}, \ldots, Z_{n} ; \mathscr{T}\right)=$ $n h(\mu)$. Thus, if $h(\mu)=0$, then $\mathscr{T}$ is independent of $\left(Z_{1}, \ldots, Z_{n}\right)$ for all $n$, and, as in the proof of Kolmogorov's zero-one law, is trivial.

For a finitely generated group $G$ with generating set $S$, we denote by $|g|$ the minimum number of elements of $S$ whose product is equal to $g$. It is easy to see that $|g h| \leq|g|+|h|$. We denote $B_{r}=\{g \in G:|g| \leq r\}$. We say that $G$ has subexponential growth if $\left|B_{r}\right|$ is smaller than any exponent. That is, if $\lim _{r} \frac{1}{r} \log \left|B_{r}\right|=0$.

Corollary 18.4. If $G$ has subexponential growth then $\mathscr{T}$ is trivial.
Proof. Since $Z_{n}$ is supported on $B_{r}, H\left(Z_{n}\right) \leq \log \left|B_{n}\right|$. Hence

$$
h(\mu)=\lim _{n} \frac{1}{n} H\left(Z_{n}\right) \leq \lim _{n} \frac{1}{n} \log \left|B_{n}\right| .
$$

Hence if $G$ is subexponential then $h(\mu)=0$ and $\mathscr{T}$ is trivial.

## 19 Characteristic functions and the Central Limit Theorem

### 19.1 Convergence in distribution

Let ( $X, X_{1}, X_{2}, \ldots$ ) be real random variables. Denote their c.d.f.s by $F, F_{1}, F_{2}, \ldots$ We say that $X_{n}$ converges in distribution to $X$ if $F_{n}(x)$ converges to $F(x)$ for every $x$ that is a continuity point of $F$.

Claim 19.1. The following are equivalent.

1. $X_{n}$ converges in distribution to $X$.
2. $\lim _{n} \mathbb{E}\left[h\left(X_{n}\right)\right]=\mathbb{E}[h(X)]$ for every bounded continuous $h: \mathbb{R} \rightarrow \mathbb{R}$.

### 19.2 The characteristic function

Let $X$ be a real random variables. The characteristic function $\varphi_{X}: \mathbb{R} \rightarrow \mathbb{C}$ of $X$ is given by

$$
\varphi_{X}(t)=\mathbb{E}\left[\mathrm{e}^{i t X}\right]=\mathbb{E}[\cos (t X)]+i \cdot \mathbb{E}[\sin (t X)] .
$$

This expectation exists for any real random variable $X$ and any real $t$, since the sine and cosine functions are bounded.

Note that

$$
\begin{aligned}
\varphi_{a X+b}(t) & =\mathbb{E}\left[\mathrm{e}^{i t(a X+b)}\right] \\
& =\mathbb{E}\left[\mathrm{e}^{i t a X} \cdot \mathrm{e}^{i t b}\right] \\
& =\varphi_{a X} \cdot \mathrm{e}^{i t b} .
\end{aligned}
$$

Exercise 19.2. $\varphi_{X}$ is continuous, and is differentiable $n$ times if $X \in \mathscr{L}^{n}$. In this case $\varphi_{X}^{(n)}(0)=i^{n} \cdot \mathbb{E}\left[X^{n}\right]$.

If $X$ and $Y$ are independent, then

$$
\begin{aligned}
\varphi_{X+Y}(t) & =\mathbb{E}\left[\mathrm{e}^{i t(X+Y)}\right] \\
& =\mathbb{E}\left[\mathrm{e}^{i t X}\right] \cdot \mathbb{E}\left[\mathrm{e}^{i t Y}\right] \\
& =\varphi_{X}(t) \cdot \varphi_{Y}(t) .
\end{aligned}
$$

A real random variable $X$ is said to have a probability distribution function (or p.d.f.) $f_{X}: \mathbb{R} \rightarrow \mathbb{R}$ if for any measurable $h: \mathbb{R} \rightarrow \mathbb{R}$ it holds that

$$
\mathbb{E}[h(X)]=\int_{-\infty}^{\infty} h(x) f_{X}(x) \mathrm{d} x,
$$

whenever the l.h.s. exists. In this case the cumulative distribution function

$$
F(x)=\mathbb{P}[X \leq x]=\int_{-\infty}^{x} f_{X}(x) \mathrm{d} x
$$

so that $f$ is the derivative of $F$. Also,

$$
\varphi_{X}(t)=\int_{-\infty}^{\infty} \mathrm{e}^{i t x} f_{X}(x) \mathrm{d} x
$$

So that $\varphi_{X}$ is the Fourier transform of $f_{X}$.
We saw in Claim 3.8 that $F$ uniquely determines the distribution of $X$.
Theorem 19.3 (Lévy's Inversion Formula). Let $X$ be a real random variable. For every $b>a$ such that $\mathbb{P}[X=a]=\mathbb{P}[X=b]=0$ it holds that

$$
F(b)-F(a)=\lim _{T \rightarrow \infty} \int_{-T}^{T} \frac{\mathrm{e}^{-i t a}-\mathrm{e}^{-i t b}}{i t} \varphi_{X}(t) \mathrm{d} t .
$$

Since there are at most countably many $c \in \mathbb{R}$ such that $\mathbb{P}[X=c]>0, F$ is determined by $\varphi_{X}$.
Theorem 19.4 (Lévy's Continuity Theorem). Let ( $X, X_{1}, X_{2}, \ldots$ ) be real random variables. Then the following are equivalent:

1. $X_{n}$ converges in distribution to $X$.
2. $\varphi_{X_{n}}(t)$ converges to $\varphi_{X}(t)$ for every $t \in \mathbb{R}$.

### 19.3 The characteristic function of normalized i.i.d. sums

Let $X$ be a standard Gaussian (or normal) random variable. This is a real random variable with p.d.f. $f_{X}(x)=\frac{1}{2 \pi} \mathrm{e}^{-x^{2} / 2}$. It is easy to calculate that

$$
\varphi_{X}(t)=\mathrm{e}^{-\frac{1}{2} t^{2}}
$$

Thus if $X_{1}$ and $X_{2}$ are independent standard Gaussian then

$$
\varphi_{\left(X_{1}+X_{2}\right) / \sqrt{2}}(t)=\mathrm{e}^{-\frac{1}{2} t^{2}}
$$

and more generally the same holds for $\left(X_{1}+\cdots+X_{n}\right) / \sqrt{n}$.
If ( $X_{1}, X_{2}, \ldots$ ) are (not necessarily Gaussian) i.i.d. and $Y_{n}=\sum_{k \leq n} X_{k}$ then

$$
\varphi_{Y_{n}}(t)=\varphi_{X_{n}}(t)^{n} .
$$

If we define

$$
Z_{n}=\frac{1}{\sqrt{n}} Y_{n}=\frac{1}{\sqrt{n}} \sum_{k \leq n} X_{n}
$$

then

$$
\varphi_{Z_{n}}(t)=\varphi_{X}(t / \sqrt{n})^{n} .
$$

Now, let $\mathbb{E}[X]=0$ and $\mathbb{E}\left[X^{2}\right]=1$. Since $X \in \mathscr{L}^{2}$ then $\varphi_{X}$ is twice differentiable and

$$
\varphi_{X}(0)=1 \quad \varphi^{\prime}(0)=0 \quad \varphi^{\prime \prime}(0)=1 .
$$

It is an exercise to show that it follows that

$$
\varphi_{X}(t)=1-\frac{1}{2} t^{2}+o\left(t^{2}\right)
$$

where here we mean by $o\left(t^{2}\right)$ that as $t \rightarrow 0$ it holds that

$$
\left|\varphi_{X}(t)-1-\frac{1}{2} t^{2}\right| \cdot t^{2} \rightarrow 0
$$

Thus we have that

$$
\varphi_{Z_{n}}(t)=\varphi_{X}(t / \sqrt{n})^{n}=\left(1-\frac{1}{2} t^{2} / n+o\left(t^{2} / n^{2}\right)\right)^{n}
$$

and thus

$$
\lim _{n} \varphi_{Z_{n}}(t)=\mathrm{e}^{-\frac{1}{2} t^{2}}
$$

As we know, $\mathrm{e}^{-\frac{1}{2} t^{2}}$ is the characteristic function of a standard Gaussian. Thus we have proved that if $G$ is a standard Gaussian then for any $t \in \mathbb{R}$ it holds that

$$
\mathbb{E}\left[\mathrm{e}^{i t Z_{n}}\right] \rightarrow \mathbb{E}\left[\mathrm{e}^{i t G}\right] .
$$

Hence $\varphi_{Z_{n}}$ converges pointwise to $\varphi_{G}$. Using Lévy's Continuity Theorem, we have thus proved the Central Limit Theorem:

Theorem 19.5. Let $\left(X_{1}, X_{2}, \ldots\right)$ be i.i.d. real random variables with $\mathbb{E}\left[X_{n}\right]=0$ and $\mathbb{E}\left[X_{n}^{2}\right]=1$. Then the sequence $Z_{n}=\frac{1}{\sqrt{n}} \sum_{k \leq n} X_{n}$ converges in distribution to a standard Gaussian.

## 20 The Radon-Nikodym derivative and absolute continuity

### 20.1 The Radon-Nikodym derivative

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Given a non-negative r.v. $X$ with $\mathbb{E}[X]=1$, we can define a the measure $\mathbb{Q}=X \cdot \mathbb{P}$ by

$$
\mathbb{Q}[A]=\mathbb{E}\left[\mathbb{1}_{\{A\}} \cdot X\right] .
$$

It is easy to show that $X$ is the unique r.v. such that $\mathbb{Q}=X \cdot \mathbb{P}$.
In this case we call $X$ the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$, and denote

$$
\frac{\mathrm{d} \mathbb{Q}}{\mathrm{dP}}(\omega)=X(\omega) .
$$

### 20.2 Absolute continuity

Note that

$$
\begin{equation*}
\mathbb{P}[A]=0 \quad \text { implies } \quad \mathbb{Q}[A]=0, \tag{20.1}
\end{equation*}
$$

so that not every measure $\mathbb{Q}$ can be written as $X \cdot \mathbb{P}$ for some $X$. When $\mathbb{Q}$ and $\mathbb{P}$ satisfy (20.1) then we say that $\mathbb{Q}$ is absolutely continuous relative to $\mathbb{P}$.

Example 20.1. - The uniform distribution on $[0,1]$ is absolutely continuous relative to the uniform distribution on $[0,2]$.

- If $\mathbb{P}[A]>0$ then $\mathbb{P}[\cdot \mid A]$ is absolutely continuous relative to $\mathbb{P}$.
- The point mass $\delta_{1 / 2}$ is not absolutely continuous relative to the uniform distribution on [0,1].
- The i.i.d. $q$ measure on $\{0,1\}^{\mathbb{N}}$ is not absolutely continuous relative to the i.i.d. p measure on $\{0,1\}^{\mathbb{N}}$, unless $p=q$.

Lemma 20.2. If $\mathbb{Q}$ is absolutely continuous relative to $\mathbb{P}$, then for each $\varepsilon>0$ there exists a $\delta>0$ such that, for every measurable $A, \mathbb{P}[A]<\delta$ implies $\mathbb{Q}[A]<\varepsilon$.

Proof. Assume the contrary, so that there is some $\varepsilon$ and a sequence of events $\left(A_{1}, A_{2}, \ldots\right)$ with $\mathbb{P}\left[A_{n}\right]<2^{-n}$ and $\mathbb{Q}\left[A_{n}\right] \geq \varepsilon$. Let $A=\cap_{n} \cup_{m>n} A_{m}$ be the event that infinitely many of these events occur. Then by Borel-Cantelli $\mathbb{P}[A]=0$. On the other hand $\mathbb{Q}[A] \geq \varepsilon$, in contradiction to absolute continuity.

### 20.3 The Radon-Nikodym Theorem

Recall that $\mathscr{F}$ is separable if it generated by a countable subset $\left\{F_{1}, F_{2}, \ldots\right\}$. We can assume w.l.o.g. that this subset is a $\pi$-system.

Theorem 20.3 (Radon-Nikodym Theorem). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with $\mathscr{F}$ separable, and let $\mathbb{Q}$ be absolutely continuous relative to $\mathbb{P}$. Then there exists a r.v. $X$ such that $\mathbb{Q}=X \cdot \mathbb{P}$.

Proof. Let $\mathscr{F}_{n}=\sigma\left(F_{1}, \ldots, F_{n}\right)$. Then $\mathscr{F}_{n}$ is a finite sigma-algebra, and as such is the set of all possible unions of $\left\{B_{1}^{n}, \ldots, B_{k}^{n}\right\}$, a finite partition of $\Omega$. Define the $\mathscr{F}_{n}$-measurable r.v. $X_{n}$ as follows. For a given $\omega \in \Omega$ there is a unique $B^{n} \in\left\{B_{1}^{n}, \ldots, B_{k}^{n}\right\}$ such that $\omega \in B^{n}$. Set

$$
X_{n}(\omega)=X_{n}\left(B^{n}\right)=\frac{\mathbb{Q}\left[B^{n}\right]}{\mathbb{P}\left[B^{n}\right]},
$$

where we take $0 / 0=0$. It is easy to verify that $\mathbb{E}\left[X_{n}\right]=1$, and that for every $B \in \mathscr{F}_{n}$ it holds that

$$
\mathbb{Q}[B]=\mathbb{E}\left[\mathbb{1}_{\{B\}} \cdot X_{n}\right],
$$

so that on $\mathscr{F}_{n}$ it holds that $X_{n}$ is the Radon-Nikodym derivative $\mathrm{d} \mathbb{Q} / \mathrm{d} \mathbb{P}$.
Now, since $\left(\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots\right)$ is a filtration, each element in $B^{n}$ is the disjoint union of (at most) two sets $B_{i}^{n+1}$ and $B_{j}^{n+1}$. Hence

$$
\begin{aligned}
\mathbb{E}\left[X_{n+1} \mid \mathscr{F}_{n}\right](\omega) & =\frac{X_{n+1}\left(B_{i}^{n+1}\right) \cdot \mathbb{P}\left[B_{i}^{n+1}\right]+X_{n+1}\left(B_{j}^{n+1}\right) \cdot \mathbb{P}\left[B_{j}^{n+1}\right]}{\mathbb{P}\left[B_{i}^{n+1}\right]+\mathbb{P}\left[B_{j}^{n+1}\right]} \\
& =\frac{\frac{\mathbb{P}\left[B_{i}^{n+1}\right]}{\mathbb{P}\left[B_{i}^{n+1}\right]} \cdot \mathbb{P}\left[B_{i}^{n+1}\right]+\frac{\mathbb{Q}\left[B_{j}^{n+1}\right]}{\mathbb{P}\left[B_{j}^{n+1}\right]} \cdot \mathbb{P}\left[B_{j}^{n+1}\right]}{\mathbb{P}\left[B^{n}\right]} \\
& =\frac{\mathbb{Q}\left[B_{i}^{n+1}\right]+\mathbb{Q}\left[B_{j}^{n+1}\right]}{\mathbb{P}\left[B^{n}\right]} \\
& =\frac{\mathbb{Q}\left[B^{n}\right]}{\mathbb{P}\left[B^{n}\right]} \\
& =X_{n}(\omega),
\end{aligned}
$$

and thus ( $X_{1}, X_{2}, \ldots$ ) is a martingale w.r.t. the filtration ( $\mathscr{F}_{1}, \mathscr{F}_{2}, \ldots$ ). Since it is non-negative then it converges almost surely to some r.v. $X$.

We now claim that ( $X_{1}, X_{2}, \ldots$ ) are uniformly integrable, in the sense that for every $\varepsilon$ there exists a $K$ such that for all $n$ it holds that

$$
\mathbb{E}\left[X_{n} \cdot \mathbb{1}_{\left\{X_{n}>K\right\}}\right]<\varepsilon .
$$

To see this, recall that $\mathbb{E}\left[X_{n}\right]=1$, note that $X_{n}$ is non-negative, and apply Markov's inequality to arrive at

$$
\mathbb{P}\left[X_{n}>K\right]<\frac{1}{K}
$$

Now, by Lemma 20.2, if we choose $K$ large enough then this implies that $\mathbb{Q}\left[X_{n}>K\right]<\varepsilon$. But the event $\left\{X_{n}>K\right\}$ is in $\mathscr{F}_{n}$, since $X_{n}$ is $\mathscr{F}_{n}$-measurable. Hence

$$
\mathbb{E}\left[X_{n} \cdot \mathbb{1}_{\left\{X_{n}>K\right\}}\right]=\mathbb{Q}\left[X_{n}>K\right]<\varepsilon .
$$

This proves that ( $X_{1}, X_{2}, \ldots$ ) are uniformly integrable. An important result (which is not hard but which we will not prove) is that if $X_{n} \rightarrow X$ almost surely, then uniform integrability implies that this convergence is also in $\mathscr{L}^{1}$, in the sense that $\mathbb{E}\left[\left|X_{n}-X\right|\right] \rightarrow 0$. It follows that for any $F_{i} \in\left\{F_{1}, F_{2}, \ldots\right\}$

$$
\lim _{n} \mathbb{E}\left[\mathbb{1}_{\left\{F_{i}\right\}} \cdot\left(X_{n}-X\right)\right] \leq \lim _{n} \mathbb{E}\left[\mathbb{1}_{\left\{F_{i}\right\}} \cdot\left|X_{n}-X\right|\right]=0
$$

and thus

$$
\mathbb{E}\left[\mathbb{1}_{\left\{F_{i}\right\}} \cdot X\right]=\lim _{n} \mathbb{E}\left[\mathbb{1}_{\left\{F_{i}\right\}} \cdot X_{n}\right]=\mathbb{Q}\left[F_{i}\right]
$$

Thus the measure $X \cdot \mathbb{P}$ agrees with $\mathbb{Q}$ on the generating algebra $\left\{F_{1}, F_{2}, \ldots\right\}$, and thus $\mathbb{Q}=X \cdot \mathbb{P}$.

## 21 Large deviations

Let $\left(X_{1}, X_{2}, \ldots\right)$ be i.i.d. real random variables. Denote $X=X_{1}$ and let $\mu=\mathbb{E}[X]$. Let

$$
Z_{n}=\frac{1}{n} \sum_{k=1}^{n} X_{k} .
$$

By the law of large numbers we expect that $Z_{n}$ should be close to $\mu$ for large $n$. What is the probability that it is larger than some $\eta>\mu$ ? We already proved the Chernoff lower bound. We here prove an asymptotically matching upper bound.

### 21.1 The cumulant generating function

Recall that the moment generating function of $X$ is

$$
M(t)=\mathbb{E}\left[e^{t X}\right]
$$

and that its cumulant generating function is

$$
K(t)=\log M(t)=\log \mathbb{E}\left[\mathrm{e}^{t X}\right] .
$$

Of course, these may be infinite for some $t$. Let $I$, the domain of both, be the set on which they are finite, and note that $0 \in I$.

Claim 21.1. $I$ is an interval, and $K$ is convex on $I$.
For the proof of this claim we will need Hölder's inequality. For $p \in[1, \infty]$ and a real r.v. $X$ denote

$$
|X|_{p}=\mathbb{E}\left[|X|^{p}\right]^{1 / p}
$$

Lemma 21.2 (Hölder's inequality). For any $p, q \in[1, \infty]$ with $1 / p+1 / q=1$ and r.v.s $X, Y$ it holds that

$$
|X \cdot Y|_{1} \leq|X|_{p} \cdot|Y|_{q}
$$

Exercise 21.3. Prove Hölder's inequality. Hint: use Young's inequality, which states that for every real $x, y \geq 0$ and $p, q>1$ with $1 / p+1 / q=1$ it holds that

$$
x y \leq \frac{x^{p}}{p}+\frac{y^{q}}{q} .
$$

Proof of Claim 21.1. Assume $a, b \in I$. Then for any $r \in(0,1)$

$$
K(r a+(1-r) b)=\log \mathbb{E}\left[\mathrm{e}^{(r a+(1-r) b) X}\right]=\log \mathbb{E}\left[\left(\mathrm{e}^{a X}\right)^{r}\left(\mathrm{e}^{b X}\right)^{1-r}\right]
$$

By Hölder's inequality

$$
\begin{aligned}
K(r a+(1-r) b) & \leq \log \mathbb{E}\left[\left(\mathrm{e}^{a X}\right)^{r}\right]^{1 / r}+\log \mathbb{E}\left[\left(\mathrm{e}^{b X}\right)^{1-r}\right]^{1 /(1-r)} \\
& =\log \mathbb{E}\left[\mathrm{e}^{a X}\right]^{r}+\log \mathbb{E}\left[\mathrm{e}^{b X}\right]^{1-r} \\
& =r \log \mathbb{E}\left[\mathrm{e}^{a X}\right]+(1-r) \log \mathbb{E}\left[\mathrm{e}^{b X}\right] \\
& =r K(a)+(1-r) K(b) .
\end{aligned}
$$

Since $K$ is non-negative it follows that it is finite on $r a+(1-r) b$, and thus $I$ is an interval on which it is convex.

Applying the Dominated Convergence Theorem inductively can be used to show that $M$ and $K$ are smooth (i.e., infinitely differentiable) on the interior of $I$.

### 21.2 The Legendre transform

Let the Legendre transform of $K$ be given by

$$
K^{\star}(\eta)=\sup _{t>0}(t \eta-K(t)) .
$$

It turns out that the fact that $K$ is smooth and convex implies that $K^{\star}$ is also smooth and convex. Therefore, if the supremum in this definition is obtained at some $t$, then $K^{\prime}(t)=\eta$. Conversely, if $K^{\prime}(t)=\eta$ for some $t$, then this $t$ is unique and $K^{\star}(\eta)=t \eta-K(t)$.

Theorem 21.4 (Chernoff bound).

$$
\mathbb{P}\left[Z_{n} \geq \eta\right] \leq \mathrm{e}^{-K^{\star}(\eta) n}
$$

Proof. For any $t \geq 0$

$$
\begin{aligned}
\mathbb{P}\left[Z_{n} \geq \eta\right] & \leq \mathbb{P}\left[t Z_{n} \geq t \eta\right] \\
& =\mathbb{P}\left[\mathrm{e}^{t \sum_{k \leq n} X_{k}} \geq \mathrm{e}^{t \eta n}\right] \\
& \leq \frac{\mathbb{E}\left[\mathrm{e}^{\sum_{k \leq n} t X_{k}}\right]}{\mathrm{e}^{t \eta n}} \\
& =\mathrm{e}^{-(t \eta-K(t)) n} .
\end{aligned}
$$

Optimizing over $t$ yields the claim.

### 21.3 Large deviations

Theorem 21.5. If $\eta=K^{\prime}(t)$ for some $t$ in the interior of I then

$$
\mathbb{P}\left[Z_{n} \geq \eta\right]=\mathrm{e}^{-K^{\star}(\eta) n+o(n)}
$$

Proof. One side is given by the Chernoff bound. It thus remains to prove the upper bound. Let $Z_{n}=\sum_{k=1}^{n} X_{k}$. We want to prove that

$$
\mathbb{P}\left[Z_{n} \geq \eta n\right]=\mathrm{e}^{-K^{\star}(\eta) n+o(n)} .
$$

For a given $n$ and $t \geq 0$, define the measure $\tilde{\mathbb{P}}$ by

$$
\frac{\mathrm{d} \tilde{\mathbb{P}}}{\mathrm{dP}}=\mathrm{e}^{n t Z_{n}-n K(t)} .
$$

Under this measure $\left(X_{1}, \ldots, X_{n}\right)$ are still i.i.d., and that their distribution does not depend on $n$ :

$$
\begin{aligned}
\tilde{\mathbb{P}}\left[X_{1} \in A_{1}, \ldots, X_{n} \in A_{n}\right] & =\tilde{\mathbb{E}}\left[\prod_{i=1}^{n} \mathbb{1}_{\left\{X_{i} \in A_{i}\right\}}\right] \\
& =\mathbb{E}\left[\mathrm{e}^{n t Z_{n}-n K(t)} \prod_{i=1}^{n} \mathbb{1}_{\left\{X_{i} \in A_{i}\right\}}\right] \\
& =\mathbb{E}\left[\mathrm{e}^{t\left(X_{1}+\cdots+X_{n}\right)-n K(t)} \prod_{i=1}^{n} \mathbb{1}_{\left\{X_{i} \in A_{i}\right\}}\right] \\
& =\mathbb{E}\left[\prod_{i=1}^{n} \mathbb{1}_{\left\{X_{i} \in A_{i}\right\}} \mathrm{e}^{t X_{i}-K(t)}\right] \\
& =\prod_{i=1}^{n} \mathbb{E}\left[\mathbb{1}_{\left\{X_{i} \in A_{i}\right\}} \mathrm{e}^{t X_{i}-K(t)}\right] .
\end{aligned}
$$

Using the fact that the expectation of a random variable is equal to the derivative at zero of its cumulant generating function, a simple calculation shows that

$$
\tilde{\mathbb{E}}\left[X_{1}\right]=\tilde{\mathbb{E}}\left[Z_{n}\right]=K^{\prime}(t) .
$$

Choose any $\eta<\bar{\eta}$ and $t$ such that $K^{\prime}(t) \in(\eta, \bar{\eta})$, so that $\tilde{\mathbb{E}}\left[Z_{n}\right]=K^{\prime}(t) \in(\eta, \bar{\eta})$. Then

$$
\begin{aligned}
\mathbb{P}\left[\eta \leq Z_{n}\right] & \geq \mathbb{P}\left[\eta \leq Z_{n} \leq \bar{\eta}\right] \\
& =\mathbb{E}\left[\mathbb{1}_{\left\{\eta \leq Z_{n} \leq \bar{\eta}\right\}}\right] \\
& =\tilde{\mathbb{E}}\left[\mathbb{1}_{\left\{\eta \leq Z_{n} \leq \bar{\eta}\right\}} \mathrm{e}^{-\left(n t Z_{n}-n K(t)\right)}\right] \\
& \geq \tilde{\mathbb{E}}\left[\mathbb{1}_{\left\{\eta \leq Z_{n} \leq \bar{\eta}\right\}} \mathrm{e}^{-(n t \bar{\eta}-n K(t))}\right] \\
& =\mathrm{e}^{-(t \bar{\eta}-K(t)) n} \tilde{\mathbb{E}}\left[\mathbb{1}_{\left\{\eta \leq Z_{n} \leq \bar{\eta}\right\}}\right] \\
& =\mathrm{e}^{-(t \bar{\eta}-K(t)) n} \tilde{\mathbb{P}}\left[\eta \leq Z_{n} \leq \bar{\eta}\right] .
\end{aligned}
$$

Since $\tilde{\mathbb{E}}\left[Z_{n}\right] \in(\eta, \bar{\eta})$, and since $\left(X_{1}, \ldots, X_{n}\right)$ are i.i.d. under $\tilde{\mathbb{P}}$ with a distribution that does not depend on $n$, by the law of large numbers,

$$
\tilde{\mathbb{P}}\left[\eta \leq Z_{n} \leq \bar{\eta}\right] \rightarrow_{n} 1,
$$

and so

$$
\lim _{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left[\eta \leq Z_{n}\right] \geq-(t \bar{\eta}-K(t))
$$

Since this holds for any $\bar{\eta}>\eta$ and $\bar{\eta}>K^{\prime}(t)>\eta$, it also holds for $\bar{\eta}=\eta$ and $t^{*}$ such that $K^{\prime}\left(t^{*}\right)=\eta$. So

$$
\lim _{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left[\eta \leq Z_{n}\right] \geq-\left(t^{*} \eta-K\left({ }^{*} t\right)\right)
$$

or

$$
\mathbb{P}\left[\eta \leq Z_{n}\right] \geq \mathrm{e}^{-\left(t^{*} \eta-K\left(t^{*}\right)\right) n+o(n)} .
$$

Finally, since $K$ is convex and smooth, and since $K^{\prime}\left(t^{*}\right)=\eta$, then $t$ is the maximizer of $t \eta-K(t)$, and thus $t^{*} \eta-K\left(t^{*}\right)=K^{\star}(\eta)$. We have thus shown that

$$
\mathbb{P}\left[\eta \leq Z_{n}\right] \geq \mathrm{e}^{-K^{\star}(\eta) n+o(n)} .
$$

## 22 Stationary distributions and processes

Given a transition matrix $P$ on some state space $S$, and given a Markov chain ( $X_{1}, X_{2}, \ldots$ ) over this $P$, the law of $X_{2}$ is given by

$$
\begin{aligned}
\mathbb{P}\left[X_{2}=t\right] & =\sum_{s} \mathbb{P}\left[X_{1}=s, X_{2}=t\right] \\
& =\sum_{s} \mathbb{P}\left[X_{1}=s\right] \mathbb{P}\left[X_{2}=t \mid X_{1}=s\right] \\
& =\sum_{s} \mathbb{P}\left[X_{1}=s\right] P(s, t) .
\end{aligned}
$$

Thus, if we think of the distributions of $X_{1}$ and $X_{2}$ as vectors $v_{1}, v_{2} \in \ell^{1}(S)$, then we have that $v_{2}=v_{1} P$.

A non-negative left eigenvector of $P$ is called a stationary distribution of $P$. It corresponds to a distribution of $X_{1}$ that induces the same distribution on $X_{2}$. By the Perron-Frobenius Theorem, if $S$ is finite then $P$ has a stationary distribution. Furthermore, if $P$ is also irreducible then this distribution is unique.

Exercise 22.1. The uniform distribution on $\mathbb{Z} / n \mathbb{Z}$ is the unique stationary distribution of the $\mu$ random walk (recall that $\mu$ is generating).

Let $\left(Y_{1}, Y_{2}, \ldots\right)$ be a general process. We say that this process is stationary (or shiftinvariant $)$ if its law is the same as the law of $\left(Y_{2}, Y_{3}, \ldots\right)$. Equivalently, for every $n$, the law of $\left(Y_{k+1}, \ldots, Y_{k+n}\right)$ is independent of $k$.

Exercise 22.2. Show that the two definitions are indeed equivalent.
Claim 22.3. If $\left(Y_{1}, Y_{2}, \ldots\right)$ is a Markov chain, and if the distribution of $Y_{1}$ is stationary, then ( $Y_{1}, Y_{2}, \ldots$ ) is a stationary process.

Returning to our scenery reconstruction problem, we can use what we learned above to deduce that ( $Z_{1}, Z_{2}, \ldots$ ) is a stationary process. It easily follows that

$$
\left(F_{1}, F_{2}, \ldots\right)=\left(f\left(Z_{1}\right), f\left(Z_{2}\right), \ldots\right)
$$

is also a stationary process.

## 23 Stationary processes and measure preserving transformations

We say that a stationary process $\left(Y_{1}, Y_{2}, \ldots\right)$ is ergodic if its shift-invariant sigma-algebra is trivial. That is, if for every shift-invariant event $A$ it holds that $\mathbb{P}[A] \in\{0,1\}$.

Some examples:

- An i.i.d. process is obviously stationary. By Kolmogorov's zero-one law its tail sigmaalgebra is trivial, and so its shift-invariant sigma-algebra is also trivial. Thus it is ergodic.
- Let $\left(Y_{1}, Y_{2}, \ldots\right)$ be binary random variables such that

$$
\mathbb{P}\left[\left(Y_{1}, Y_{2}, \ldots\right)=(1,1, \ldots)\right]=1 / 2
$$

and

$$
\mathbb{P}\left[\left(Y_{1}, Y_{2}, \ldots\right)=(0,0, \ldots)\right]=1 / 2 .
$$

This process is stationary but not ergodic; the event $\lim _{n} Y_{n}=1$ is shift-invariant and has probability $1 / 2$.

- Let $\left(Y_{1}, Y_{2}, \ldots\right)$ be binary random variables such that

$$
\mathbb{P}\left[\left(Y_{1}, Y_{2}, \ldots\right)=(1,0,1,0, \ldots)\right]=1 / 2
$$

and

$$
\mathbb{P}\left[\left(Y_{1}, Y_{2}, \ldots\right)=(0,1,0,1, \ldots)\right]=1 / 2 .
$$

This process is stationary and ergodic.

- Let $P$ be chosen uniformly over $[0,1]$, and let $\left(Y_{1}, Y_{2}, \ldots\right)$ be binary random variables, which conditioned on $P$ are i.i.d. Bernoulli with parameter $P$. This process is stationary but not ergodic. For example, the event that

$$
\lim _{n} \frac{1}{n} \sum_{k \leq n} Y_{k} \leq 1 / 2
$$

is a shift-invariant event that has probability $1 / 2$.

- Let $\left(Y_{1}, Y_{2}, \ldots\right)$ be a Markov chain, with the distribution of $Y_{1}$ equal to some stationary distribution. Then this process is stationary. It is ergodic iff the distribution of $Y_{1}$ is not a non-trivial convex combination of two different stationary distributions.
- Let $Y_{1}$ be distributed uniformly on [0,1). Fix some $0<\alpha<1$, and let $Y_{n+1}=Y_{n}+\alpha$ $\bmod 1$. This is a stationary process, and it is ergodic iff $\alpha$ is irrational. We will show this later.

A generalization of the last example is the following. Let $(\Omega, \mathscr{F}, v)$ be a probability space, and let $T: \Omega \rightarrow \Omega$ be a measurable transformation that preserves $v$. That is, $v(A)=v\left(T^{-1}(A)\right)$ for all $A \in \mathscr{F}$. We say that $A \in \mathscr{F}$ is $T$-invariant if $T^{-1}(A)=A$, and note that the collection of $T$-invariant sets is a sub-sigma-algebra. Let $Y_{1}$ have law $v$, and let each $Y_{n+1}=T\left(Y_{n}\right)$. Then $\left(Y_{1}, Y_{2}, \ldots\right)$ is a stationary process.
Claim 23.1. $\left(Y_{1}, Y_{2}, \ldots\right)$ is ergodic iff for every $T$-invariant $A \in \mathscr{F}$ it holds that $v(A) \in\{0,1\}$.
Proof. The map $\pi: \Omega \rightarrow \Omega^{\mathbb{N}}$ given by $\pi(\omega)=\left(\omega, T(\omega), T^{2}(\omega), \ldots\right)$ is a bijection that pushes the measure $v$ to the law $\mathbb{P}$ of $\left(Y_{1}, Y_{2}, \ldots\right)$, and thus these two probability spaces are isomorphic. Furthermore, if we denote the shift by $\sigma: \Omega^{\mathbb{N}} \rightarrow \Omega^{N}$, then $\pi$ is equivariant, in the sense that $\pi \circ T=\sigma \circ \pi$. It follows that the $T$-invariant sigma-algebra is mapped to the shift-invariant sigma-algebra, and thus one is trivial iff the other is trivial.

Of course, if we have a process $\left(Y_{1}, Y_{2}, \ldots\right)$ taking values in $\Omega^{n}$, then stationarity is precisely invariance w.r.t. the shift transformation $T: \Omega \rightarrow \Omega$ given by $T\left(x_{1}, x_{2}, x_{3}, \ldots\right)=$ $\left(x_{2}, x_{3}, \ldots\right)$. Thus stationary processes and measure preserving transformations are two manifestations of the same object. We say that $T$ is ergodic if the $T$-invariant sigma-algebra is trivial. That is, if for every measurable $A$ such that $T^{-1}(A)=A$ it holds that $A$ has measure in $\{0,1\}$.
Claim 23.2. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, with $T: \Omega \rightarrow \Omega$ an ergodic measure preserving transformation.

If $Z: \Omega \rightarrow \mathbb{R}$ is a T-invariant random variable (i.e., $Z(\omega)=Z(T(\omega))$ for almost every $\omega \in \Omega$ ) then there is some $z \in \mathbb{R}$ such that $\mathbb{P}[Z=z]=1$.

Exercise 23.3. Prove this claim. Hint: If $Z$ is $T$-invariant then for any $a<b \in \mathbb{R}$, the event $Z \in[a, b]$ is T-invariant, and thus has measure either 0 or 1 .

Consider the $\operatorname{map} R_{\alpha}: S^{1} \rightarrow S^{1}$ given by $R_{\alpha}\left(\mathrm{e}^{2 \pi i z}\right)=\mathrm{e}^{2 \pi i(z+\alpha)}$. This is a measure preserving transformation of $S^{1}$, equipped with the uniform measure.
Proposition 23.4. $R_{\alpha}$ is ergodic iff $\alpha$ is irrational.
Proof. If $\alpha=k / m$ is rational, then the set $\left\{\mathrm{e}^{2 \pi i z}: z \in \cup_{n=0}^{m-1}[n / m, n / m+1 / 2 m]\right\}$ is $R_{\alpha}$-invariant and has measure $1 / 2$. Hence $R_{\alpha}$ is not ergodic.

If $\alpha$ is irrational, let $f: S^{1} \rightarrow\{0,1\}$ be the indicator of an $R_{\alpha}$-invariant set. We can use the Fourier transform to write $f$ as

$$
f(z)=\sum_{k \in \mathbb{Z}} c_{n} \mathrm{e}^{2 \pi i k z}
$$

for some coefficients ( $c_{n}$ ). Then $\left[R_{\alpha} f\right](z)$ is

$$
\left[R_{\alpha} f\right](z)=\sum_{k \in \mathbb{Z}} c_{k} \mathrm{e}^{2 \pi i k(z-\alpha)}=\sum_{k \in \mathbb{Z}} d_{k} \mathrm{e}^{2 \pi i k z},
$$

where $d_{k}=c_{k} \mathrm{e}^{-2 \pi i k \alpha}$. Since $A$ is $R_{\alpha}$-invariant then $R_{\alpha} f=f$, and so $c_{k}=d_{k}$. Since $\alpha$ is irrational, $\mathrm{e}^{-2 \pi i k \alpha} \neq 1$ unless $k=0$, and so we have that $c_{k}=0$ unless $k=0$. Thus $f$ is constant, and so it must be the indicator of a set of measure either 0 or 1 .

## 24 The Ergodic Theorem

Theorem 24.1 (The Pointwise Ergodic Theorem). Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, with $T: \Omega \rightarrow \Omega$ a measure preserving transformation. If $T$ is ergodic then for every $X \in \mathscr{L}^{1}(\Omega, \mathscr{F}, \mathbb{P})$ it holds that for $v$-almost every $\omega \in \Omega$

$$
\lim _{n} \frac{1}{n} \sum_{k=0}^{n-1} X\left(T^{k}(\omega)\right)=\mathbb{E}[X]
$$

In the language of stationary processes, one can say that if $\left(Y_{1}, Y_{2}, \ldots\right)$ is a stationary process with trivial shift-invariant sigma-algebra, and if $f\left(Y_{1}, Y_{2}, \ldots\right) \in \mathscr{L}^{1}$, then almost surely

$$
\lim _{n} \frac{1}{n} \sum_{k=1}^{n} f\left(Y_{k}, Y_{k+1}, \ldots\right)=\mathbb{E}\left[f\left(Y_{1}, Y_{2}, \ldots\right)\right] .
$$

This Theorem was originally proved by Birkhoff [?]. We give a proof due to Katznelson and Weiss [?].

Proof. We assume without loss of generality that $X$ is non-negative; otherwise apply the proof separately to $X^{+}$and $X^{-}$. Define $X^{*}: \Omega \rightarrow \Omega$ by

$$
X^{*}(\omega)=\lim _{n} \frac{1}{n} \sum_{k=0}^{n-1} X\left(T^{k}(\omega)\right)
$$

whenever this limit exists. We want to show that it exists w.p. 1 , and that $\mathbb{P}\left[X^{*}=\mathbb{E}[X]\right]=1$.
Define $\bar{X}: \Omega \rightarrow \mathbb{R}$ by

$$
\bar{X}(\omega)=\limsup _{n} \frac{1}{n} \sum_{k=0}^{n-1} X\left(T^{k}(\omega)\right)
$$

and likewise

$$
\underline{X}(\omega)=\liminf _{n} \frac{1}{n} \sum_{k=0}^{n-1} X\left(T^{k}(\omega)\right) .
$$

Note that both are $T$-invariant, and so there are some $\bar{x}$ and $\underline{x}$ such that

$$
\mathbb{P}[\bar{X}=\bar{x}, \underline{X}=\underline{x}]=1 .
$$

Proving that

$$
\begin{equation*}
\bar{x} \leq \mathbb{E}[X] \leq \underline{x} \tag{24.1}
\end{equation*}
$$

will thus finish the proof.

Fix some $\varepsilon>0$. Let $N(\omega)$ be the first positive integer such that

$$
\begin{equation*}
\frac{1}{N(\omega)} \sum_{k=0}^{N(\omega)-1} X\left(T^{k}(\omega)\right)+\varepsilon \geq \bar{x} \tag{24.2}
\end{equation*}
$$

Since $N(\omega)$ is a.s. finite, there is some $K \in \mathbb{N}$ such that the set $A=\{\omega: N(\omega)>K\}$ has measure less than $\varepsilon / \bar{x}$. Define

$$
\tilde{X}(\omega)= \begin{cases}X(\omega) & \omega \notin A \\ \max \{X(\omega), \bar{x}\} & \omega \in A\end{cases}
$$

and also

$$
\tilde{N}(\omega)= \begin{cases}N(\omega) & \omega \notin A \\ 1 & \omega \in A .\end{cases}
$$

Note that in analogy to (24.2) we have that

$$
\frac{1}{\tilde{N}(\omega)} \sum_{k=0}^{\tilde{N}(\omega)-1} \tilde{X}\left(T^{k}(\omega)\right)+\varepsilon \geq \bar{x}
$$

or, rearranging, that

$$
\begin{equation*}
\sum_{k=0}^{\tilde{N}(\omega)-1} \tilde{X}\left(T^{k}(\omega)\right) \geq \tilde{N}(\omega)(\bar{x}-\varepsilon) . \tag{24.3}
\end{equation*}
$$

Now $X$ and $\tilde{X}$ only differ on $A$, and when they do differ then it is at most by $\bar{x}$, since $X$ is non-negative. Hence

$$
\begin{align*}
\mathbb{E}[\tilde{X}] & =\mathbb{E}[X+(\tilde{X}-X)] \\
& =\mathbb{E}[X]+\mathbb{E}[\tilde{X}-X] \\
& =\mathbb{E}[X]+\mathbb{E}\left[(\tilde{X}-X) \cdot \mathbb{1}_{\{A\}}\right] \\
& \leq \mathbb{E}[X]+\mathbb{E}\left[\bar{x} \cdot \mathbb{1}_{\{A\}}\right] \\
& \leq \mathbb{E}[X]+\bar{x} \cdot \varepsilon / \bar{x} \\
& =\mathbb{E}[X]+\varepsilon . \tag{24.4}
\end{align*}
$$

Now, let $L=K \bar{x} / \varepsilon$. For each $\omega \in \Omega$, let $\omega_{0}=\omega$ and let

$$
\omega_{j+1}=T^{\tilde{N}\left(\omega_{j}\right)}\left(\omega_{j}\right)
$$

It follows that

$$
\omega_{j}=T^{\tilde{N}\left(\omega_{0}\right)+\tilde{N}\left(\omega_{1}\right)+\cdots+\tilde{N}\left(\omega_{j-1}\right)}(\omega) .
$$

Let $J(\omega)$ be the maximal $j$ such that

$$
\tilde{N}\left(\omega_{0}\right)+\tilde{N}\left(\omega_{1}\right)+\cdots+\tilde{N}\left(\omega_{j}\right)<L
$$

and let

$$
\tilde{N}_{L}(\omega)=\tilde{N}\left(\omega_{0}\right)+\tilde{N}\left(\omega_{1}\right)+\cdots+\tilde{N}\left(\omega_{J(\omega)}\right) .
$$

Note that $\tilde{N}_{L}(\omega)>L-K$. Then we can write

$$
\sum_{k=0}^{L-1} \tilde{X}\left(T^{k}(\omega)\right)=\sum_{k=0}^{\tilde{N}\left(\omega_{0}\right)} \tilde{X}\left(T^{k}\left(\omega_{0}\right)\right)+\cdots+\sum_{k=0}^{\tilde{N}\left(\omega_{J(\omega)}\right)} \tilde{X}\left(T^{k}\left(\omega_{J(\omega)}\right)\right)+\sum_{k=\tilde{N}_{L}(\omega)}^{L-1} \tilde{X}\left(T^{k}(\omega)\right)
$$

Applying (24.3) to each term but the last yields

$$
\sum_{k=0}^{L-1} \tilde{X}\left(T^{k}(\omega)\right) \geq \tilde{N}_{L}(\omega)(\bar{x}-\varepsilon)+\sum_{k=\tilde{N}_{L}(\omega)}^{L-1} \tilde{X}\left(T^{k}(\omega)\right),
$$

and using the fact that $X$ is non-negative means

$$
\sum_{k=0}^{L-1} \tilde{X}\left(T^{k}(\omega)\right) \geq \tilde{N}_{L}(\omega)(\bar{x}-\varepsilon)
$$

Since $\tilde{N}_{L}(\omega)>L-K$ we can apply this estimate too, and, rearranging, arrive at

$$
\frac{1}{L} \sum_{k=0}^{L-1} \tilde{X}\left(T^{k}(\omega)\right) \geq \bar{x}-\frac{K}{L} \bar{x}-\varepsilon
$$

which by the choice of $L$ we can write as

$$
\frac{1}{L} \sum_{k=0}^{L-1} \tilde{X}\left(T^{k}(\omega)\right) \geq \bar{x}-2 \varepsilon
$$

Now, by $T$-invariance the expectation of the l.h.s. is just equal to the expectation of $\tilde{X}$. Hence

$$
\mathbb{E}[\tilde{X}] \geq \bar{x}-2 \varepsilon
$$

Putting this together with (24.4) yields

$$
\bar{x} \leq \mathbb{E}[\tilde{X}]+2 \varepsilon \leq \mathbb{E}[X]+3 \varepsilon,
$$

and taking $\varepsilon$ to zero yields $\bar{x} \leq \mathbb{E}[X]$. This completes the first half of the proof of (24.1); the second follows by a similar argument.

Exercise 24.2. Use the Ergodic Theorem to prove the strong law of large numbers.

## 25 The weak topology and the simplex of invariant measures

Let $X$ be a compact metrizable topological space. By the Riesz Representation Theorem we can identify $\mathscr{P}(X)$, the set of probability measures on $X$, with the positive bounded linear functionals on $C(X)$ that assign 1 to the constant function 1. The space $X^{*}$ of bounded linear functionals on $C(X)$ comes equipped with the compact, metrizable weak* topology, under which $\varphi_{n} \rightarrow \varphi$ if $\varphi_{n}(f) \rightarrow \varphi(f)$ for all $f \in C(X)$. The restriction of this topology to the (closed) set of probability measures yields what probabilists call the weak topology on the probability measures on $X$.

In the important case that $X=\{0,1\}^{\mathbb{N}}$ we have that $v_{n} \rightarrow v$ weakly if for every clopen $A$ it holds that $v_{n}(A) \rightarrow v(A)$. In the case $X=\mathbb{R} \cup\{-\infty, \infty\}$ we have that $v_{n} \rightarrow v$ if $\lim \sup _{n} v_{n}(A) \leq$ $v(A)$ for all closed $A$, or if $\liminf _{n} v_{n}(A) \geq v(A)$ for all open $A$.

Let $X=\{0,1\}^{\mathbb{Z}}$, and denote by $\mathscr{I}(X)$ the set of stationary (or shift-invariant) probability measures on $X$.

Claim 25.1. $\mathscr{I}(X)$ is a closed subset of $\mathscr{P}(X)$.
Proof. Denote the shift by $\sigma: X \rightarrow X$. Assume that $v_{n}$ is a sequence in $\mathscr{I}(X)$ that converges to some $v \in \mathscr{P}(X)$. We prove the claim by showing that $v$ is stationary.

Let $A$ be a clopen subset of $X$. Then

$$
v(A)=\lim _{n} v_{n}(A)=\lim _{n} v_{n}(\sigma(A))=v(\sigma(A)),
$$

where the last equality follows from the fact that $A$ being clopen implies that $\sigma(A)$ is clopen. Thus $v$ is invariant on a generating sub-algebra of the sigma-algebra, and by a standard argument it is invariant.

Clearly, $\mathscr{I}(X)$ is a convex set. The next proposition shows (a more general claim which implies) that its extreme points $\mathscr{I}_{e}(X)$ are the ergodic measures.

Proposition 25.2. A T-invariant measure $v$ on $(\Omega, \mathscr{F})$ is ergodic iff it is extreme.
Proof. Assume that $v$ is not ergodic. Then there is some $T$-invariant $A \in \mathscr{F}$ such that $p:=v(A) \in(0,1)$. Let $v_{1}$ be given by $v_{1}(B)=v(B \mid A)=\frac{1}{p} v(B \cap A)$, and let $v_{2}(B)=v\left(B \mid A^{\mathrm{c}}\right)$. Then

$$
\begin{aligned}
v_{1}\left(T^{-1} B\right) & \left.=\frac{1}{p} v\left(\left(T^{-1}(B)\right) \cap A\right)\right) \\
& \left.=\frac{1}{p} v\left(\left(T^{-1}(B)\right) \cap T^{-1}(A)\right)\right) \\
& =\frac{1}{p} v\left(\left(T^{-1}(B \cap A)\right)\right) \\
& =\frac{1}{p} v(B \cap A) \\
& =v_{1}(B) .
\end{aligned}
$$

And thus $v_{1}$ is $T$-invariant. The same argument applies to $v_{2}$, since $A^{\mathrm{c}}$ is also $T$-invariant. Finally, $v=p v_{1}+(1-p) v_{2}$.

For the other direction, assume $v=p v_{1}+(1-p) v_{2}$ for some $p \in(0,1)$. Clearly, $v_{1}$ is absolutely continuous relative to $v$, and so we write $v_{1}=X \cdot v$ for some $X \in \mathscr{L}^{1}(v)$.

We now claim that $X$ is $T$-invariant; we prove this for the case that $T$ is invertible (although it is true in general). In this case, for any $A \in \mathscr{F}$

$$
\begin{aligned}
v_{1}(A)=v_{1}(T(A)) & =\int_{\Omega} \mathbb{1}_{\{A\}}\left(T^{-1}(\omega)\right) \cdot X(\omega) \mathrm{d} v(\omega) \\
& =\int_{\Omega} \mathbb{1}_{\{A\}}\left(T^{-1}(\omega)\right) \cdot X(\omega) \mathrm{d} v\left(T^{-1}(\omega)\right) \\
& =\int_{\Omega} \mathbb{1}_{\{A\}}(\omega) \cdot X(T(\omega)) \mathrm{d} v(\omega),
\end{aligned}
$$

and so $X \circ T$ is also a Radon-Nikodym derivative $\mathrm{d} v_{1} / \mathrm{d} v$. But by the uniqueness of this derivative $X$ and $X \circ T$ agree almost everywhere. It is a now nice exercise to show that there exists some $X^{\prime}$ that is equal to $X$ almost everywhere and is $T$-invariant. It then follows by Claim 23.2, and by the fact that $\mathbb{E}[X]=1$, that $\mathbb{P}[X=1]=1$, and thus $v=v_{1}$.

This Theorem has an interesting consequence.
Exercise 25.3. Assume $v, \mu$ are both T-invariant ergodic measures on $(\Omega, \mathscr{F})$. Show that there exist two disjoint set $A, B \in \mathscr{F}$ such that $v(A)=0$ and $\mu(A)=1$, while $v(B)=1$ and $\mu(B)=0$.

Thus $v$ and $\mu$ "live in different places."
In fact, it is possible to show that there is a map $\beta: \mathscr{I}_{e}(X) \rightarrow \mathscr{F}$ with the properties that

1. $\mu\left(\beta_{\mu}\right)=1$ for all $\mu \in \mathscr{\mathscr { C }}_{e}(X)$.
2. For all $\mu \neq v \in \mathscr{I}_{e}(X)$ it holds that $\beta_{\mu} \cap \beta_{v}=\varnothing$.

Using this, it is possible to show that $\mathscr{I}(X)$ is in fact a simplex: a compact convex set in which there is a unique way to write each element as a convex integral of the extreme points.

Proposition 25.4. The ergodic measures $\mathscr{I}_{e}(X)$ are dense in $\mathscr{I}(X)$.
Thus the simplex $\mathscr{I}(X)$ has the interesting property that its extreme points are dense. It turns out that there is only one such simplex (up to affine homeomorphisms), which is called the Poulsen simplex.

Proof. It suffices to show that for $v, \mu \in \mathscr{I}_{e}(X)$ and $\theta=\frac{1}{2} v+\frac{1}{2} \mu$ it is possible to find $\theta_{n} \in \mathscr{I}_{e}(X)$ s.t. $\lim _{n} \theta_{n}=\theta$.

To this end, fix $n$ and define $\theta_{n}$ as follows. Let the law of the r.v.s $\left(X_{k}\right)_{k \in \mathbb{Z}}$ be $\mu$, and the law of $\left(Y_{k}\right)_{k \in \mathbb{Z}}$ be $v$. For $m \in \mathbb{Z}$, let $\left(X_{0}^{m}, \ldots, X_{n-1}^{m}\right)$ be independent of all previously defined random variables, and with law equal to that of ( $X_{0}, \ldots, X_{n-1}$ ). Define ( $Y_{0}^{m}, \ldots, Y_{n-1}^{m}$ ) analogously.

Define $\left(W_{k}\right)_{k \in \mathbb{Z}}$ by

$$
W_{k}= \begin{cases}X_{k}^{\lfloor k / n\rfloor} & \text { if }\lfloor k / n\rfloor \text { is even } \\ Y_{k \bmod n}^{\lfloor k / n\rfloor} & \text { if }\lfloor k / n\rfloor \text { is odd } .\end{cases}
$$

Finally, choose $N$ uniformly at random from $\{0,1, \ldots, 2 n-1\}$, and define $\left(Z_{k}\right)_{k \in \mathbb{Z}}$ by

$$
Z_{k}=W_{k+N}
$$

Let $\theta_{n}$ be the law of $\left(Z_{k}\right)$.
It is straightforward (if tedious) to verify that $\left(Z_{k}\right)$ is stationary. We leave it as an exercise to show that it is ergodic. Thus to finish the proof we have to show that $\lim _{n} \theta_{n}=\theta$.

Fix $M \in \mathbb{N}$, and consider the event that $N \in\{1, \ldots, M\}$. As $n$ tends to infinity, the probability of this event tends to zero. Thus, if we condition on $N$, with probability that tends to $1 / 2$ we have that the law of $\left(Z_{1}, \ldots, Z_{M}\right)$ is equal to the law of $\left(X_{1}, \ldots, X_{M}\right)$, and likewise for $\left(Y_{1}, \ldots, Y_{M}\right)$. This completes the proof.

## 26 Percolation

Let $V$ be a countable set, and let $\mathscr{G}=(V, \mathscr{E})$ be a locally finite, simple symmetric graph. That is, $\mathscr{E}$ is a symmetric relation on $\mathscr{V}$ with $\mathscr{E} \cap\{v\} \times V$ finite for each $v \in \mathcal{V}$. We also assume that $\mathscr{G}$ is connected, so that the transitive closure of $\mathscr{E}$ is $V \times V$.

The i.i.d. p percolation measure on $\{0,1\}^{\mathscr{E}}$ is simply the product Bernoulli measure, in which we choose each edge independently with probability $p$. We will denote this measure by $\mathbb{P}_{p}[\cdot]$, and will denote by $E$ the random edge set with this law. $G=(\mathcal{V}, E)$ will be the corresponding random graph.

Note that $G$ will in general not be connected. For each $v \in \mathscr{V}$ we denote by $K(v)$ the (random) connected component that $v$ belongs to in $G$. We denote by $\{v \leftrightarrow \infty\}$ the event that $K(v)$ is infinite. We denote by $K_{\infty}$ the event that there is some $v$ for which $K(v)$ is infinite.

Claim 26.1. The probability of $K_{\infty}$ is either 0 or 1. In the former case, for every $v \in \mathcal{V}$, $\mathbb{P}_{p}[v \leftrightarrow \infty]=0$, while in the latter $\mathbb{P}_{p}[v \leftrightarrow \infty]>0$.

Proof. Enumerate $\mathscr{E}=\left(e_{1}, e_{2}, \ldots\right)$, and let $A_{n}=\left\{e_{i} \in E\right\}$. Then $\left(A_{1}, A_{2}, \ldots\right)$ is an i.i.d. sequence. Clearly $K_{\infty}$ is $\sigma\left(A_{1}, A_{2}, \ldots\right)$-measurable, and also clearly it is a tail event. Hence the first part of the claim follows by Kolmogorov's 0-1 law.

Since the event $K_{\infty}$ contains $\{v \leftrightarrow \infty\}$, it is immediate that $\mathbb{P}_{p}\left[K_{\infty}\right]=0$ implies $\mathbb{P}_{p}[v \leftrightarrow \infty]=$ 0 . Assume now that $\mathbb{P}_{p}\left[K_{\infty}\right]=1$. Then there is some $w \in \mathcal{V}$ such that, with positive probability, $\mathbb{P}_{p}[w \leftrightarrow \infty]$. Let $P=\left(e_{1}, e_{2}, \ldots, e_{n}\right)$ be a path between $v$ and $w$.

Consider the random variable $\tilde{E}$ taking values in $\{0,1\}^{\mathscr{E}}$ defined as follows: for every edge $e \notin P$, we set $e \in \tilde{E}$ iff $e \in E$. And we set $e \in \tilde{E}$ for all $e \in P$. We (you) prove in the exercise below that the law of $\tilde{E}$ is absolutely continuous relative to the law of $E$.

Denote $\tilde{G}=(\mathcal{V}, \tilde{E})$, and denote by $\tilde{K}(v)$ the connected component of $v$ in $\tilde{G}$. Now, $\tilde{K}(v)=$ $\tilde{K}(w)$, since $v$ and $w$ are connected in $\tilde{G}$. Also, $\tilde{K}(w)$ contains $K(w)$, since $\tilde{E}$ contains $E$. Hence the event $\{|\tilde{K}(w)|=\infty\}$ occurs with positive probability, and so the same holds for $\tilde{K}(v)=\tilde{K}(w)$. Finally, by absolute continuity, the same holds for $K(v)$, and so $\mathbb{P}_{p}[v \leftrightarrow \infty]>0$.

Exercise 26.2. Prove that the law of $\tilde{E}$ is absolutely continuous relative to the law of $E$.
Claim 26.3. If $q>p$ then $\mathbb{P}_{q}\left[K_{\infty}\right] \geq \mathbb{P}_{p}\left[K_{\infty}\right]$.
To prove this claim we prove a stronger theorem, and in the process introduce the technique of coupling. Let $\Omega=\{0,1\}^{\mathbb{N}}$, with $\mathscr{F}$ the Borel sigma-algebra. We consider the natural partial order on $\Omega$ given by $\omega \geq \omega^{\prime}$ if $\omega_{n} \geq \omega_{n}^{\prime}$ for all $n \in \mathbb{N}$. We say that $A \in \mathscr{F}$ is increasing if for all $\omega \geq \omega^{\prime}$ it holds that $\omega^{\prime} \in A$ implies $\omega \in A$. Let $\mathbb{P}_{p}[\cdot]$ denote the i.i.d. $p$ measure on $\Omega$.

Theorem 26.4. If $A$ is increasing then $q>p$ implies $\mathbb{P}_{q}[A] \geq \mathbb{P}_{p}[A]$.
Proof. Let ( $X_{1}, X_{2}, \ldots$ ) be i.i.d. random variables, each distributed uniformly on [0,1]. For each $n$ let $Q_{n}=\mathbb{1}_{\left\{X_{n} \leq q\right\}}$ and $P_{n}=\mathbb{1}_{\left\{X_{n} \leq p\right\}}$. Note that $\mathbb{P}\left[Q_{n}=1\right]=q$ and $\mathbb{P}\left[P_{n}=1\right]=p$, and that
$\left(Q_{1}, Q_{2}, \ldots\right)$ is i.i.d., as is $\left(P_{1}, P_{2}, \ldots\right)$. Hence the law of $\left(Q_{1}, Q_{2}, \ldots\right)$ (resp., $\left(P_{1}, P_{2}, \ldots\right)$ ) is $\mathbb{P}_{q}[\cdot]$ (resp., $\mathbb{P}_{p}[\cdot]$ ). Note also that

$$
\left(Q_{1}, Q_{2}, \ldots\right) \geq\left(P_{1}, P_{2}, \ldots\right),
$$

since $q>p$. Hence for any increasing event $A \subset\{0,1\}^{\mathbb{N}}$ it holds that $\left(P_{1}, P_{2}, \ldots\right) \in A$ implies $\left(Q_{1}, Q_{2}, \ldots\right) \in A$, and thus $\mathbb{P}_{q}[A] \leq \mathbb{P}_{p}[A]$.

The construction in this proof is an example of coupling. Formally, a coupling of two probability spaces $(\Omega, \mathscr{F}, \mathbb{P})$ and $\left(\Omega^{\prime}, \mathscr{F}^{\prime}, \mathbb{P}^{\prime}\right)$, is a probability space $\left(\Omega \times \Omega^{\prime}, \sigma\left(\mathscr{F} \times \mathscr{F}^{\prime}\right), \mathbb{Q}\right)$ such that the projections on the two coordinates pushes $\mathbb{Q}$ forward to $\mathbb{P}$ and $\mathbb{P}^{\prime}$.

Since $\mathbb{P}_{p}\left[K_{\infty}\right] \in\{0,1\}$, since $\mathbb{P}_{p}\left[K_{\infty}\right]$ is weakly increasing in $p$, we are interested in the critical percolation probability

$$
p_{c}=\sup \left\{p: \mathbb{P}_{p}\left[K_{\infty}\right]=0\right\} .
$$

An interesting (and often hard) question is whether $\mathbb{P}_{p_{c}}\left[K_{\infty}\right]$ is zero or one.
Let $\mathscr{G}$ be the infinite $k$-ary tree with root $o$. In this case we can calculate $p_{c}$, by noting that the event $\{o \leftrightarrow \infty\}$ can be thought of as the event that the Galton-Watson tree with children distribution $B(k, p)$ is infinite. We know that this happens with positive probability iff $p>1 / k$. Hence in this case $p_{c}=1 / k$, and $\mathbb{P}_{p_{c}}\left[K_{\infty}\right]=0$.

## 27 The mass transport principle

Let $\mathscr{G}=(\mathcal{V}, \mathscr{E})$ be a locally finite, countable graph. A graph automorphism is a bijection $f: V) V \mathcal{V}$ such that $(v, w) \in \mathscr{E}$ iff $(f(v), f(w)) \in \mathscr{E}$. The automorphisms of a graph form the group $\operatorname{Aut}(\mathscr{G})$ under composition. We say that $\mathscr{G}$ is transitive if its automorphism group acts on it transitively. That is, if for all $v, w \in \mathscr{V}$ there is a graph automorphism $f$ s.t. $f(v)=w$. Intuitively, this means that the geometry of the graph "looks the same" from the point of view of every vertex.

An important example is when $\Gamma$ is finitely generated by a symmetric finite subset $S$, and $\mathscr{G}=(\mathcal{V}, \mathscr{E})$ with $\mathcal{V}=\Gamma$ is the corresponding Cayley graph. In this case it is easy to see that the $\Gamma$ action on itself is an action by graph automorphisms, which is furthermore already transitive. We will restrict our discussion to this setting, even though it all extends to unimodular transitive graphs; these are graphs with a unimodular automorphism group.

A map $f: \Gamma \times \Gamma \rightarrow[0, \infty)$ is a mass-transport if it is invariant under the diagonal $\Gamma$-action:

$$
f(h, k)=f(g h, g k)
$$

for all $g, h, k \in \Gamma$. It is useful to think about $f$ as indicating how much "mass" is passed from $h$ to $k$, where the amount passed can depend on identities of $h$ and $k$, but in a way that (in some sense) only depends on the geometry of the graph and not on their names.

Theorem 27.1 (Mass Transport Principle for Groups). For every mass transport $f: \Gamma \times \Gamma \rightarrow$ $[0, \infty)$ and $g \in \Gamma$ it holds that

$$
\sum_{k \in G} f(g, k)=\sum_{k \in G} f(k, g) .
$$

That is, the total mass flowing out of $g$ is equal to the total mass flowing in.
Proof. By invariance

$$
\sum_{k \in G} f(g, k)=\sum_{k \in G} f\left(k^{-1} g, e\right) .
$$

Changing variables to $h=k^{-1} g$ yields

$$
=\sum_{h \in G} f(h, e) .
$$

Applying invariance again yields

$$
\sum_{h \in G} f(g h, g)
$$

and again changing variables to $k=g h$ yields the desired result.


Figure 1: Spanning forest.

As an application, consider the following random subgraph $E$ of the standard Cayley graph of $\mathbb{Z}^{2}$. For each $z, w_{1}, w_{2} \in \mathbb{Z}^{2}$ such that $w_{1}=z+(0,1)$ and $w_{2}=z+(1,0)$, we independently set $\left(z, w_{1}\right) \in E,\left(z, w_{2}\right) \notin E$ w.p. $1 / 2$, and $\left(z, w_{1}\right) \notin E,\left(z, w_{2}\right) \in E$ w.p. $1 / 2$.

For distinct $z, w \in \mathbb{Z}^{2}$, we say that $w$ is a descendant of $z$ (and $z$ is an ancestor of $w$ ) in $E$ if there is a path between $w$ and $z$, and if $w \leq z$ in both coordinates.

Note that, by construction,

1. $E$ has no cycles, and each node is adjacent to at least one edge, and so $E$ is a spanning forest.
2. Each $w$ has infinitely many ancestors.
3. If $w \leq z$ then the number of descendants of $w$ is independent of the number of descendants of $z$.

Proposition 27.2. The number of descendants of each $w \in \mathbb{Z}$ is almost surely finite, with infinite expectation.

Proof. Let $f(w, z)$ equal the probability that $z$ is an ancestor of $w$, and note that by the invariance of the definitions $f$ is a mass transport.

The sum $\sum_{z} f(w, z)$ is the expected number of ancestors of $w$, which is infinite, since $w$ a.s. has infinitely many ancestors. It follows by the mass transport principle that $\sum_{z} f(z, w)$, the expected number of descendants of $w$, is likewise infinite.

It is easy to see that the expected number of direct descendants of any $w$ is 1 . By the independence property mentioned above, the process ( $N_{1}, N_{2}, \ldots$ ) - where $N_{k}$ is the number of descendants at distance $k$ from $w$ - is a non-negative martingale. It thus converges, and moreover must converge to an integer, and so converges to 0 .

Note that we can define the same process on $\mathbb{Z}^{d}$, where now each vertex has $d$ potential ancestors, with the proof applying as is. We can further generalize to groups that have a set $S$ such that $S \cup S^{-1}$ generates $\Gamma$ and the graph induced by $S$ has no cycles.

Exercise 27.3. On $\mathbb{Z}^{d}$, prove that $E$ is a spanning tree iff $d \leq 3$.

## 28 Majority dynamics

Let $\mathscr{G}=(\mathcal{V}, \mathscr{E})$ be a locally finite, countable (or finite) graph, and denote by $N(v)$ the neighbors of $v \in \mathcal{V}$. We would like $|N(v)|$ to be odd, and so we either add or remove $v$ to $N(v)$ to achieve this.

Let $\left(V_{1}, V_{2}, \ldots\right)$ be random subsets of $\mathcal{V}$, with the property that if $v, w \in V_{n}$ then $(v, w) \notin E$; that is, each $V_{n}$ is an independent set.

Let $\Theta=\{-1,+1\}^{\vee}$, and consider the following sequence of random variables ( $X_{1}, X_{2}, \ldots$ ), each taking values in $\Theta$. First, let $X_{1}$ be chosen i.i.d. 1/2. Given $X_{n}$, we define $X_{n+1}$ as follows. For $v \notin V_{n}$ let $X_{n+1}(v)=X_{n}(v)$. For $v \in V_{n}$ let

$$
X_{n+1}(v)=\operatorname{sgn} \sum_{w \in N(v)} X_{n}(w) .
$$

Note that since $|N(v)|$ is odd then the sum is never 0 , and there is no ambiguity with taking the signum. This process is called majority dynamics, or zero temperature Glauber dynamics.

Proposition 28.1. If $\mathscr{G}$ is finite then $X_{n}$ converges (hence stabilizes) almost surely.
Proof. Let

$$
H_{n}=\sum_{(v, w) \in E} X_{n}(v) \cdot X_{n}(w),
$$

so that $H_{n}$ is the number of edges in the graph along which there is agreement, minus the number of edges along which there is disagreement.

Note that in majority dynamics, whenever a node changes its label and non of its neighbors do, the total number of disagreements in the graph strictly decreases, and thus $H_{n}$ decreases. Since $V_{n}$ is an independent set, changes at a node are always done while keeping its neighbors constant, and thus $H_{n}$ decreases by at least 2 with each change in $X_{n}$. Since $H_{n}$ is bounded from below by $-|E|, X_{n}$ must stabilize.

In fact, this proof can be generalized to the case that $\mathscr{G}$ is infinite, but with bounded degrees and subexponential growth. This is no longer true on general graphs:

Exercise 28.2. Prove that on the 3-regular tree $X_{n}$ does not in general stabilize.
It is also not true if the $V_{n}$ are not independent sets.
Let $\mathscr{G}$ be a Cayley graph of a finitely generated group $\Gamma$, and choose ( $V_{1}, V_{2}, \ldots$ ) from a distribution that is invariant to the $\Gamma$-action. For example, to choose $V_{n}$ one can choose an independent uniform number for each node, and include in $V_{n}$ only those nodes whose numbers are higher than all of their neighbors'.

Theorem 28.3. In this setting $X_{n}$ converges pointwise almost surely.

Proof. Let

$$
h_{n}=\sum_{g \in N(e)} \mathbb{E}\left[X_{n}(e) \cdot X_{n}(g)\right]
$$

Now,

$$
\begin{aligned}
h_{n+1}-h_{n} & =\sum_{g \in N(e)} \mathbb{E}\left[X_{n+1}(e) \cdot X_{n+1}(g)\right]-\sum_{g \in N(e)} \mathbb{E}\left[X_{n}(e) \cdot X_{n}(g)\right] \\
& =\sum_{g \in N(e)} \mathbb{E}\left[X_{n+1}(e) \cdot X_{n+1}(g)-X_{n}(e) \cdot X_{n}(g)\right]
\end{aligned}
$$

If $e \notin V_{n}$ and also $g \notin V_{n}$ then the term in the expectation is zero. Hence

$$
\begin{aligned}
=\sum_{g \in N(e)} \mathbb{E} & {\left[X_{n+1}(e) \cdot X_{n+1}(g)-X_{n}(e) \cdot X_{n}(g) \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}\right] } \\
& +\mathbb{E}\left[X_{n+1}(e) \cdot X_{n+1}(g)-X_{n}(e) \cdot X_{n}(g) \cdot \mathbb{1}_{\left\{g \in V_{n}\right\}}\right],
\end{aligned}
$$

since the two conditioned events $\left\{e \in V_{n}\right\}$ and $\left\{g \in V_{n}\right\}$ are mutually exclusive. Furthermore, this implies

$$
\begin{aligned}
=\sum_{g \in N(e)} \mathbb{E} & {\left[X_{n+1}(e) \cdot X_{n}(g)-X_{n}(e) \cdot X_{n}(g) \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}\right] } \\
& +\mathbb{E}\left[X_{n}(e) \cdot X_{n+1}(g)-X_{n}(e) \cdot X_{n}(g) \cdot \mathbb{1}_{\left\{g \in V_{n}\right\}}\right],
\end{aligned}
$$

which by the mass transport principle

$$
=2 \sum_{g \in N(e)} \mathbb{E}\left[X_{n+1}(e) \cdot X_{n}(g)-X_{n}(e) \cdot X_{n}(g) \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}\right] .
$$

Rearranging yields

$$
=2 \mathbb{E}\left[\left(X_{n+1}(e)-X_{n}(e)\right) \cdot \sum_{g \in N(e)} X_{n}(g) \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}\right],
$$

and since, conditioned on $e \in V_{n}, X_{n+1}(e)=\operatorname{sgn} \sum_{g \in N(e)} X_{n}(g)$, then

$$
=2 \mathbb{E}\left[2 \mathbb{1}_{\left\{X_{n+1}(e) \neq X_{n}(e)\right\}} \cdot\left|\sum_{g \in N(e)} X_{n}(g)\right| \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}\right],
$$

Now, by definition $\mathbb{1}_{\left\{X_{n+1}(e) \neq X_{n}(e)\right\}} \cdot \mathbb{1}_{\left\{e \in V_{n}\right\}}=\mathbb{1}_{\left\{X_{n+1}(e) \neq X_{n}(e)\right\}}$. Also, since $|N(e)|$ is odd, $\left|\sum_{g \in N(e)} X_{n}(g)\right| \geq$ 1 , and so

$$
\geq 4 \mathbb{P}\left[X_{n+1}(e) \neq X_{n}(e)\right] .
$$

Hence $h_{n}$ is non-decreasing. Since it is bounded by $|N(e)|$ it converges to some $h_{\infty}<\infty$. Furthermore

$$
h_{\infty}-h_{1} \geq 4 \sum_{n=1}^{\infty} \mathbb{P}\left[X_{n+1}(e) \neq X_{n}(e)\right],
$$

and so the expected number of $n$ such that $X_{n+1}(e) \neq X_{n}(e)$ is finite, and in particular $X_{n}(e)$ stabilizes w.p. 1. By invariance this holds for every $X_{n}(g)$, and we have proved our claim.

## 29 Scenery Reconstruction: I

Fix $n$, and let ( $X=X_{1}, X_{2}, \ldots$ ) be i.i.d. random variables on the abelian group $\mathbb{Z} / n \mathbb{Z}$. Denote by $\mu(k)=\mathbb{P}[X=k]$ their law. Let $X_{0}$ be uniformly distributed on $\mathbb{Z} / n \mathbb{Z}$, and let $Z_{n}=\sum_{k=0}^{n} X_{n}$ be the corresponding random walk. We assume throughout that the support of $\mu$ generates $\mathbb{Z} / n \mathbb{Z}$.

Some important examples to keep in mind:

- $\mu(1)=1$.
- $\mu(1)=1-\varepsilon, \mu(2)=\varepsilon$.

Fix some $f \in\{0,1\}^{n}$, and let $F_{n}=f\left(Z_{n}\right)$. The law of $\left(F_{1}, F_{2}, \ldots\right)$ depends on $f$; we think of these distributions as a family indexed by $f$. We denote by $\mathbb{P}_{f}[\cdot]$ the distribution when we fix a particular $f$. Note that $\mathbb{P}_{f}[\cdot]$ does not change if we shift $f$.

## Exercise 29.1. Prove this.

Denote by $[f]$ the equivalence class of $f$ under shifts. That is, $f^{\prime} \in[f]$ if there is some $k \in \mathbb{Z} / n \mathbb{Z}$ such that for every $m \in \mathbb{Z} / n \mathbb{Z}$ it holds that $f^{\prime}(k+m)=f(m)$.

The question of scenery reconstruction is the following: is it possible to determine [ $f$ ] given ( $F_{1}, F_{2}, \ldots$ )? In particular we say that we can reconstruct $f$ if there is some measurable

$$
\hat{f}:\{0,1\}^{\mathbb{N}} \rightarrow\{0,1\}^{n}
$$

such that for every $f \in\{0,1\}^{n}$ it holds that

$$
\begin{equation*}
\mathbb{P}_{f}\left[\hat{f}\left(F_{1}, F_{2}, \ldots\right) \in[f]\right]=1 \tag{29.1}
\end{equation*}
$$

Equivalently, if

$$
\mathbb{P}\left[\hat{f}\left(f\left(Z_{1}\right), f\left(Z_{2}\right), \ldots\right) \in[f]\right]=1 .
$$

In statistics, $\hat{f}$ is called an estimator of $f$, and the existence of such an $\hat{f}$ is called identifiability (of $f$ ). This clearly depends on $\mu$, and so we say that $\mu$ is reconstructive if this holds.

One can reformulate (29.1) in finitary terms. It is equivalent to the existence of a sequence $\left(\hat{f}_{1}, \hat{f}_{2}, \ldots\right)$ with $\hat{f}_{k}$ being $\sigma\left(F_{1}, \ldots, F_{k}\right)$-measurable and with

$$
\lim _{k} \mathbb{P}_{f}\left[\hat{f}_{k}\left(F_{1}, \ldots, F_{k}\right) \in[f]\right]=1
$$

for all $f \in\{0,1\}^{n}$.
A very interesting question is how quickly does this converge to one (when it does), for $\mu$ chosen uniformly over $n$; for example for $\mu(1)=0.99, \mu(2)=0.01$.

Question 29.2. Let $N(n, \varepsilon)$ be the smallest $k$ such that there is an $\hat{f}_{k}:\{0,1\}^{k} \rightarrow\{0,1\}^{n}$ with

$$
\mathbb{P}_{f}\left[\hat{f}_{k}\left(F_{1}, \ldots, F_{k}\right) \in[f]\right] \geq 1-\varepsilon
$$

for all $f$. For fixed $\varepsilon$ (say $1 / 3$ ), how does $N(n, \varepsilon)$ grow with $n$ ?
This is not known; it is not even known if $N(\cdot, \varepsilon)$ is exponential or polynomial. The question of whether a given $\mu$ is reconstructive is much better understood.

Theorem 29.3. Let $n$ be a prime $>5$, and let $\mu \in \mathbb{Q}^{n}$. Then $\mu$ is reconstructive iff $\varphi_{\mu}(k) \neq \varphi_{\mu}(m)$ for all $k \neq m$. Here $\varphi_{\mu}$ is given by

$$
\varphi_{\mu}(k)=\varphi_{X}(k)=\mathbb{E}\left[\mathrm{e}^{\frac{2 \pi i}{n} \cdot k \cdot X}\right]=\sum_{\ell \in \mathbb{Z} / n \mathbb{Z}} \mathrm{e}^{\frac{2 \pi i}{n} \cdot k \cdot X} \mu(k) .
$$

where $k \cdot X$ is multiplication $\bmod n$.
The first direction (the case that $\varphi_{\mu}(k) \neq \varphi_{\mu}(m)$ for all $k \neq m$ ) does not require the extra assumptions on $n$ and $\mu$. This is due to Matzinger and Lember [?].

To prove this theorem we will need to study a few new concepts.

## 30 Scenery reconstruction: II

Fix $n \in \mathbb{N}, f \in\{0,1\}^{n}$ and $\mu$ a generating probability measure on $\mathbb{Z} / n \mathbb{Z}$, and recall our process in which $X_{0}$ is uniform on $\mathbb{Z} / n \mathbb{Z},\left(X_{1}, X_{2}, \ldots\right)$ are i.i.d. with law $\mu, Z_{n}=X_{0}+X_{1}+\cdots+X_{n}$ and $F_{n}=f\left(Z_{n}\right)$. Recall also that are we are interested in guessing (correctly, almost surely) what [ $f$ ] is (the equivalence class of functions that are shifts of $f$ ) from a single random instance of ( $F_{1}, F_{2}, \ldots$ ).

Define the $a: \mathbb{Z} / n \mathbb{Z} \rightarrow \mathbb{R}$, autocorrelation of $f$ by

$$
a(k)=\frac{1}{n} \sum_{m=0}^{n} f(m) \cdot f(m+k),
$$

and note that $a$ is the same for any $f^{\prime} \in[f]$. Imagine that we are willing to settle on reconstructing $a$ rather than [ $f$ ]. We will show that if the values of the characteristic function $\varphi_{\mu}$ are unique then we can reconstruct the $a(k)$ 's.

To this end, we define $A: \mathbb{N} \rightarrow \mathbb{R}$, the autocorrelation of $F$ by

$$
\alpha_{k}=\mathbb{E}\left[F_{T} \cdot F_{T+k}\right]
$$

for some $T \in \mathbb{N}$; by stationarity, the choice of $T$ is immaterial. We will show that if we know the $\alpha_{k}$ 's the we can infer the $a_{k}$ 's. But this will not help us, unless there some measurable $\hat{\alpha}_{k}:\{0,1\}^{\mathbb{N}} \rightarrow \mathbb{R}$ such that

$$
\mathbb{P}_{f}\left[\hat{\alpha}_{k}\left(F_{1}, F_{2}, \ldots\right)=\alpha_{k}\right]=1 .
$$

A natural candidate for $\hat{\alpha}_{k}$ is the empirical average; we take limsup rather than lim to make sure $\hat{\alpha}_{k}$ is well defined:

$$
\hat{\alpha}_{k}=\limsup _{m} \frac{1}{m} \sum_{T=1}^{m} F_{T} \cdot F_{T+k} .
$$

A statement such as " $\hat{\alpha}_{k}=\alpha_{k}$ almost surely" sounds a lot like the strong law of large numbers. We will show later that this is indeed true, and that it follows from the Ergodic Theorem, which is a generalization of the SLLN.

Let $\mu * \mu$ be the convolution of $\mu$ with itself, which is given by

$$
[\mu * \mu](k)=\sum_{m} \mu(k-m) \cdot \mu(m) .
$$

This is a probability distribution which is exactly the law of $X_{1}+X_{2}$. Define analogously the $k$-fold convolution $\mu^{(k)}$, which is the law of $X_{1}+\cdots+X_{k}$.

Claim 30.1. For every $k \in \mathbb{N}$ it holds that

$$
\alpha_{k}=\sum_{m} \mu^{(k)}(m) \cdot a_{m} .
$$

Proof. We set $T=0$, condition on $X_{0}$ and $Z_{k}$ and thus

$$
\begin{aligned}
\alpha_{k} & =\mathbb{E}\left[f\left(Z_{0}\right) \cdot f\left(Z_{k}\right)\right] \\
& =\sum_{m, \ell} \mathbb{E}\left[f\left(X_{0}\right) \cdot f\left(Z_{k}\right) \mid X_{0}=\ell, Z_{k}=\ell+m\right] \cdot \mathbb{P}\left[X_{0}=\ell, Z_{k}=\ell+m\right] \\
& =\sum_{m, \ell} f(\ell) \cdot f(\ell+m) \cdot \frac{1}{n} \cdot \mu^{(k)}(m) \\
& =\sum_{m} a_{m} \cdot \mu^{(k)}(m) .
\end{aligned}
$$

It follows that if we denote by $\alpha$ the column vector ( $\alpha_{0}, \ldots, \alpha_{n-1}$ ), by $a$ the column vector $\left(a_{0}, \ldots, a_{n-1}\right)$, and by $M$ the $n \times n$ matrix $M_{k, m}=\mu^{(k)}(m)$ then $\alpha=M a$. Assuming (as we will show later) that we can determine $\alpha$, it follows that we can determine $a$ if $M$ is invertible.

Claim 30.2. $M$ is invertible iff the values of the characteristic function $\varphi_{\mu}$ are unique.
Proof. We apply the Fourier transform to each row of $M$. Since the Fourier transform is an orthogonal linear transformation, the resulting matrix $\hat{M}$ is invertible iff $M$ is invertible.

Now, over $\mathbb{Z} / n \mathbb{Z}$ the Fourier transform is identical to the characteristic function. Since the $k^{\text {th }}$ row of $M$ is the law of $X_{1}+\cdots+X_{k}$, the $k^{\text {th }}$ row of $\hat{M}$ is given by

$$
\varphi_{X_{1}+\cdots+X_{k}}(m)=\mathbb{E}\left[\mathrm{e}^{\frac{2 \pi i}{n} \cdot m \cdot\left(X_{1}+\cdots+X_{k}\right)}\right]=\varphi_{X}(m)^{k} .
$$

Thus $\hat{M}$ is a Vandermonde matrix, and is invertible iff $\varphi_{X}$ has unique values.
Recall that we are interested in reconstructing [ $f$ ] rather than $a$. To this end we need to define the two-fold autocorrelation

$$
a_{k, \ell}=\frac{1}{n} \sum_{m=0}^{n} f(m) \cdot f(m+k) \cdot f(m+k+\ell)
$$

and its analogue

$$
\alpha_{k, \ell}=\mathbb{E}\left[F_{T} \cdot F_{T+k} \cdot F_{T+k+\ell}\right] .
$$

It is then easy to show that there is also a linear relation between these two objects, with the corresponding matrix being $M \otimes M$, the tensor product of $M$ with itself. This is invertible iff $M$ is invertible, and so we get the same result. However, this still does not suffice, and we need to add still more indices and calculate $n$-fold autocorrelations. The appropriate matrices are again invertible iff $M$ is, and moreover [ $f$ ] is uniquely determined by the $n$-fold autocorrelation.


[^0]:    ${ }^{1}$ Note that this definition is slightly different than the standard one.

[^1]:    ${ }^{2}$ Neither of the names Galton and Watson have died out (as of 2018), although Galton is rather rare: https://forebears.io/surnames/galton.

