Markov Processes and Martingales

Károly Simon

Department of Stochastics
Institute of Mathematics
Technical University of Budapest
www.math.bme.hu/~simonk

D File

1 A number of boring definitions
2 Kolmogorov Extension Theorem
3 Stationary processes
4 Birkhoff pointwise Ergodic Theorem
- Invariant and ergodic measures
- Weak $*$-topology
- existence of invariant measure
- Birkhoff ET Theorem
5 Revisit Markov chains
6 Martingal CHT
- Recall: CLT for iid with finite mean and variance
- CLT for martingales with stationary L^2 differences
- CLT for Markov chains

Isomorphic probability spaces

We say that the probability spaces $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ are isomorphic if there exist

- full measure subsets $X_i' \subset X_i$, $\mu(X_i') = 1$ and
- $\varphi : X_1' \to X_2'$ invertible measure preserving map.

Theorem 1.1

Let μ be a non-atomic Borel probability measure on a complete metric space X. Then $(X, \mathcal{B}, \mu)$ is isomorphic to $([0,1], \mathcal{B}[0,1], \mathcal{L}\mathrm{eb})$. The completion of $(X, \mathcal{B}, \mu)$ is isomorphic to $([0,1], \mathcal{L}, \mathcal{L}\mathrm{eb})$, where $\mathcal{L}$ is the σ-algebra of the Lebesgue measurable subsets of $[0,1]$.

Proof: Click here in the pdf file.

Lebesgue space or Standard probability space

A probability space $(X, \mathcal{A}, \mu)$ is a Lebesgue space if it is isomorphic to the probability space which is the disjoint union of $([0,s], \mathcal{L}[0,s], \mathcal{L}\mathrm{eb})$ for some $0 \leq s \leq 1$ and at most countably many atoms. (If there are no atoms then $s = 1$.)
Here $\mathcal{L}[0,s]$ is the σ-algebra of the Lebesgue measurable subsets of the interval $[0,s]$.
Be careful: This is an ambiguous notion. The expression "Lebesgue space" may refer to some special Banach spaces called L^p spaces.

1 A number of boring definitions
2 Kolmogorov Extension Theorem
3 Stationary processes
4 Birkhoff pointwise Ergodic Theorem
- Invariant and ergodic measures
- Weak $*$-topology
- existence of invariant measure
- Birkhoff ET Theorem
5 Revisit Markov chains
6 Martingal CHT
- Recall: CLT for iid with finite mean and variance
- CLT for martingales with stationary L^2 differences
- CLT for Markov chains

Stationary Processes

Definition 2.1 (Polish space)

Polish space is a topological space that is homeomorphic to a complete metric space which is separable (has a countable dense subset).

Fact 2.2

Any two uncountable Polish spaces are Borel isomorphic.

Definitions

In this Section we always use the following notation:
Let $T \subset \mathbb{R}$. Let $\{S_t, \mathcal{B}_t\}_{t \in T}$ be a family of Polish spaces ($\mathcal{B}_t$ is their Borel σ-algebra). For each finite $F \subset T$ given a Borel probability measure μ_F on

$$(S_F, \mathcal{B}_F) := \left(\prod_{t \in F} S_t, \prod_{t \in F} \mathcal{B}_t \right)$$

For for all finite set $F \subset G \subset T$ the canonical projection from S_G to S_F is denoted by Proj_{GF}.

Kolmogorov Consistency Condition

The family $\{\mu_F : F \subset T, \text{ finite }\}$. satisfies Kolmogorov Consistency Condition if for all finite set $F \subset G \subset T$

$$\mu_F = \mu_G \left(P_{GF}^{-1} B \right), \quad \forall B \in \mathcal{B}_F. \tag{1}$$

Kolmogorov Extension Theorem

Theorem (Kolmogorov Extension Theorem)

We assume that the Kolmogorov's Consistency Condition holds for the family $\{\mu_F : F \subset T, \text{ finite }\}$. Then there is a unique probability measure μ on the product $(S_T, \mathcal{B}_T) := \left(\prod_{t\in T} S_t, \prod_{t\in T} \mathcal{B}_t\right)$ which extends each μ_F.

Kolmogorov Extension Theorem cont.

Theorem (Kolmogorov Extension Theorem) cont.

That is μ is the unique probablity measure on $(S_T, \mathcal{B}_T)$ with

(2) $$\mu_F(B) = \mu(P_{TF}^{-1}(B)),$$

where P_{TF} is the canonical projection from $(S_T, \mathcal{B}_T)$ to $(S_F, \mathcal{B}_F)$.

1. A number of boring definitions
2. Kolmogorov Extension Theorem
3. Stationary processes
4. Birkhoff pointwise Ergodic Theorem
   - Invariant and ergodic measures
   - Weak $*$-topology
   - existence of invariant measure
   - Birkhoff ET Theorem
5. Revisit Markov chains
6. Martingal CHT
   - Recall: CLT for iid with finite mean and variance
   - CLT for martingales with stationary L^2 differences
   - CLT for Markov chains

Some conventions used in the Section

Assume that we are given sequence $\{\xi_k\}_{k\in\mathcal{I}}$ of $(S, \mathcal{B})$-valued r.v. , where the index set $\mathcal{I}$ is either $\mathbb{N}$ or $\mathbb{Z}$. Further, for all k and $n_1, \dots, n_k$ given the joint distribution of

$$(\xi_{n_1}, \dots, \xi_{n_k}).$$

Moreover, we assume that the conditions of the Kolmogorov Extension Theorem hold. Hence, we can consider the infinite product measurable space $(\Omega, \mathcal{F})$, where

(3) $$\Omega := S^{\mathcal{I}}, \quad \mathcal{F} := \mathcal{B}^{\mathcal{I}}.$$

and on it the infinite product σ-algebra $\mathcal{F}$ form a measurable space.

Some conventions used in the Section

On the measurable space $(\Omega, \mathcal{F})$ we can define the coordinate maps. Namely, for every $i \in \mathcal{I}$ and $\boldsymbol{\omega} \in \Omega$ let

$$X_i(\boldsymbol{\omega}) := \omega_i.$$

We say that $\{X_i\}_{i\in\mathcal{I}}$ is the canonical process for $(\Omega, \mathcal{F})$.

Some conventions used in the Section (Cont.)

On the measurable space $(\Omega, \mathcal{F})$ we denote by $\mathbb{P}$ the measure whose existence in guaranteed by the Kolmogorov Extension Theorem. In general $\mathbb{P}$ is NOT a product measure (only if ξ_i are independent).

Corollary 3.1 (Corollary of Kolmorov Extention Theorem)

The n-th coordinate ω_n of a random element $\boldsymbol{\omega} \in \Omega$ is equal to ξ_n in distribution.

Stationary Processes

Here we follow Varadhan's book [24].

Definition 3.2 (Stationary Processes)

Let $(S, \mathcal{B})$ be a Polish space equipped with the σ-algebra of its Borel sets $\mathcal{B}$. We say that the sequence $\{\xi_i : i \in \mathbb{Z}\}$ of S-valued random variables is *stationary* if for every $k \geq 1$, $n, n_1, \dots n_k \in \mathbb{Z}$ we have

(4) $$(\xi_{n_1}, \dots, \xi_{n_k}) \stackrel{d}{=} (\xi_{n_1+n}, \dots, \xi_{n_k+n}).$$

The left shift σ

Let us define σ the left shift on Ω by

(5) $$\sigma(\boldsymbol{\omega})_n := \omega_{n+1},$$

that is the n-th coordinate of $\sigma(\boldsymbol{\omega})$ is equal to the $n+1$-th coordinate of $\boldsymbol{\omega}$. That is for a stationary process:

(6) $$\mathbb{P}\left(\sigma^{-1}E\right) = \mathbb{P}(E), \quad \forall E \in \mathcal{F}.$$

Definition 3.3

For a sequence $\{\xi_k\}$ the dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ constructed on the last three slides is called the canonical dynamical system.

Stationary Processes

Example 3.4

- If $X = (X_n)$ is iid then it is stationary.
- Let $X = \{X_n\}_{n=0}^{\infty}$ be a Markov chain with stationary distribution π. If we choose the initial element X_0 according to π then X is stationary.
- Let $Y = \{Y_n\}_n \in \mathbb{Z}$ and let $c_1, \dots, c_k \in \mathbb{R}$. Then for $X_n := \sum_{\ell=1}^{k} c_\ell Y_{n-\ell}$ is a stationary process called moving averages with $c_1, \dots, c_k$.

1. A number of boring definitions
2. Kolmogorov Extension Theorem
3. Stationary processes
4. Birkhoff pointwise Ergodic Theorem
   - Invariant and ergodic measures
   - Weak $*$-topology
   - existence of invariant measure
   - Birkhoff ET Theorem
5. Revisit Markov chains
6. Martingal CHT
   - Recall: CLT for iid with finite mean and variance
   - CLT for martingales with stationary L^2 differences
   - CLT for Markov chains

Recall: Push forward measure

Let X, Y be metric spaces. Given a measure μ on X and a map $T : X \to Y$. The push forward measure is

(7) $$T_*\mu(A) := \mu\left(T^{-1}A\right) \text{ for } A \subset Y.$$

Clearly, $T_*\mu$ is a measure and if T and μ are Borel then $T_*\mu$ is also a Borel measure.

Recall: Push forward measure (cont.)

Theorem 4.1

Let $T : X \to Y$ be a Borel mapping, μ is a Borel measure on X, and g is a non-negative Borel function on Y. Then

(8) $$\int g dT_*\mu = \int (g \circ T) d\mu.$$

Invariant measure

Consider the special case when $X = Y$:

Definition 4.2

Let $T : X \to X$, $\mathcal{B}$ a σ-algebra on X finally, let and μ be a probability measure on the measurable space $(X, \mathcal{B})$. We say that μ is an invariant measure if $T^{-1}A \in \mathcal{B}$ for every $A \in \mathcal{B}$ and

(9) $T*_\mu = \mu$ that is $\mu(A) = \mu(T^{-1}A), \quad \forall A \in \mathcal{B}$.

Definitions

Then it follows from (8) that

(10) $$\int g(x)d\mu(x) = \int g(T(x))d\mu(x),$$

where g is a non-negative Borel function.

Definition 4.3

Let $\mathcal{M}(X)$ be the set of probability measures on X and we write $\mathcal{M}_T(X)$ for the subset of invariant probability measures.

Example 4.4

Let $X_1 := S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}_1$ be the Borel σ-algebra on S^1. Let μ_1 be the normalized length on S^1.

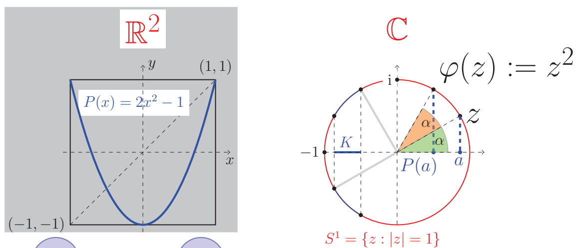
Let $X_2 := [-1, 1]$, $\mathcal{A}_2$ be the Borel σ-algebra on $[-1, 1]$

$$\mu_2(K) := \frac{\operatorname{arclength}(h^{-1}(K))}{2\pi} = \int_K \frac{1}{\pi\sqrt{1-x^2}}dx$$

where $h(z) := \mathrm{Re}(z)$. Further, let $\varphi : S^1 \to S^1$, $P : [-1, 1] \to [-1, 1]$:

$$\varphi(z) := z^2 \text{ and } P(x) := 2x^2 - 1.$$

Then P is a factor of φ. (See the Figure on the next slide.)



$$\begin{tikzcd} S^1 \arrow[r, "\varphi"] \arrow[d, "h"'] \arrow[dr, phantom, "\text{commutes}"] & S^1 \arrow[d, "h"] \\ {[-1,1]} \arrow[r, "P"] & {[-1,1]} \end{tikzcd}$$

$$2(\cos\alpha)^2 - 1 = \cos(2\alpha)$$

$$P(\mathrm{Re}(z)) = \mathrm{Re}(z^2)$$

$$h(z) := \mathrm{Re}(z)$$

for $K \subset [-1, 1]$: $\mu_2(K) := \dfrac{\operatorname{arclength}(h^{-1}(K))}{2\pi}$

Weak $*$-topology on $C(X)^*$

As usual we write $C(X)$ for the Banach space of continuous real valued functions on the compact metric space X endowed with the sup norm. We denote the dual space of $C(X)$ by $C(X)^*$. That is $C(X)^*$ is the set of continuous linear functionals on $\alpha : C(X) \to \mathbb{R}$. The weak $*$-topology on $C(X)^*$ is generated by the sets:

$$U(f, \varepsilon, \alpha_0) := \{\alpha \in C(X)^* : |\alpha(f) - \alpha_0(f)| < \varepsilon\},$$

where $f \in C(X)$, $\varepsilon > 0$.

Weak $*$-topology on $C(X)^*$ II

The reason that we (very much) like the weak $*$-topology is that:

Lemma 4.5

If $C(X)$ is separable then any closed ball in $C(X)^*$ is compact.

Theorem 4.6 (Riesz Repezentation Theorem)

We write $\alpha_\mu(f) := \int f d\mu$ for a $\mu \in \mathcal{M}(X)$. Then $\mu \leftrightarrow \alpha_\mu$ is a bijection between $\mathcal{M}(X)$ and

$$\{\alpha \in C(X)^* : \alpha(1) = 1 \text{ and } \alpha(f) \geq 0 \text{ if } f \geq 0\}.$$

Using this identification we often write μ when we mean α_μ. The topology on $\mathcal{M}(X)$ generated by the weak $*$-topology on $C(X)^*$ is called weak topology on $\mathcal{M}(X)$. It is easy to verify that

Lemma 4.7

Let $\{f_n\}_{n=1}^\infty$ be a dense subset of $C(X)$. Then the weak topology on $\mathcal{M}(X)$ is equivalent to the topology defined by the metric

$$d(\mu, \nu) := \sum_{n=1}^{\infty} 2^{-n} \|f_n\|^{-1} \left| \int f_n d\mu - \int f_n d\nu \right|.$$

Using Cantor diagonal method and the previous lemma, one can easily see that

Lemma 4.8

Let X be compact. Then $\mathcal{M}(X)$ is compact, convex in the weak topology.

Theorem 4.9

Let X be compact and let $T : X \to X$ be continuous. Then $\mathcal{M}_T(X)$ is a non-empty compact convex subset of $\mathcal{M}(X)$ (in the weak topology).

Proof.

One can easily see that $T_* : \mathcal{M} \to \mathcal{M}$ is a homeomorphism. (Recall: We always mean that $\mathcal{M}(X)$ is a topological space endowed with the weak topology.) Clearly, $\mathcal{M}_T(X) = \{\mu \in \mathcal{M}(X) : T_*\mu = \mu\}$. For an arbitrary $\mu \in \mathcal{M}(X)$ let

$$\mu_n := \frac{1}{n}\left(\mu + T_*\mu + \cdots + T_*^{n-1}\mu\right).$$

By compactness we can find a convergent subsequence $\mu_{n_k} \to \mu'$. It is straightforward that , $\mu' \in \mathcal{M}_T(X)$. □

Definition 4.10

We say that a measure $\mu \in \mathcal{M}_T$ is ergodic if the following holds:

(11) If $T^{-1}A = A$ then $\mu(A)$ is either 0 or 1.

We write $\mathcal{E}_T$ for the set of ergodic measures.

Homework 1

Prove that this definition is equivalent to the following one:
A measure μ is ergodic if for every $f \in L^1$ the fact that f is constant on μ-a.a. orbits $\{T^n(x)\}_{n=0}^\infty$ is equivalent to the fact that f is constant for μ-a.a. x.

Another useful characterization of the ergodic measures is as follows:

$\mathcal{E}_T$ consists of the extremal points of $\mathcal{M}_T$.

Namely, $\mathcal{M}_T$ is a convex set. An invariant measure μ is ergodic iff it cannot be presented as a non-trivial convex combination of invariant measures. Further, it follows from Choquet Theorem that there exists an ergodic decomposition of invariant measures. It says that invariant measures can be presented as convex combinations of ergodic measures (in the sense of (12)). For a nice account about this click here in the pdf file.

Corollary 4.11 (Corollary of Choquet Theorem)

There is a probability measure η on $\mathcal{M}_T$ (with the σ-algebra generated by the weak topology) such that

(a) *For every $f : X \to \mathbb{R}$ continuous function:*
(12)
$$\int f(x) d\mu(x) = \int_{\nu \in \mathcal{M}_T} \left(\int_{x \in X} f(x) d\nu(x) \right) d\eta(\nu)$$

(b) $\eta(\mathcal{E}_T) = 1$.

Equivalent formulations of ergodicity

Let $(X, \mathcal{A}, m)$ be a probability space and let $T : X \to X$ be measurable. Then the following are equivalent:

1. T is ergodic
2. If f is measurable and $\forall x$: $f \circ T(x) = f(x)$ holds then f is constant.
3. If f is measurable and $\forall x$: $f \circ T(x) = f(x)$ holds a.e x then f is constant a.e. x.
4. If $f \in L^2$ and $f \circ T(x) = f(x)$ holds $\forall x$ then f is constant a.e.
5. If $f \in L^2$ and $f \circ T(x) = f(x)$ holds for a.e. x then f is constant a.e.

See Walters' book [25, p. 28] and see Homeworks # 39.

Birkhoff ET

Theorem 4.12

Assume that μ is ergodic and $f \in L^1(X, \mu)$. Then
(13)
$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \longrightarrow \int_X f(x)d\mu(x) \quad \text{for } \mu \text{ a.a. } x. \text{ and in } L^1$$

Proof (Katzneson, B. Weiss)

First we assume that $f \in L^\infty$ and we prove only the a.s. convergence.

$$\underline{f}(x) := \liminf_{n\to\infty} \frac{1}{n} S_n f(x) \text{ and } \overline{f}(x) := \liminf_{n\to\infty} \frac{1}{n} S_n f(x),$$

Proof of the Birkhoff ET cont.

Proof (Cont.)

where $S_n f(x) := f(x) + f(Tx) + \cdots + f(T^{n-1}x)$. Using that both $\underline{f}$ and $\overline{f}$ are constant on every orbit, that is

$$\underline{f}(x) = \underline{f}(Tx) \text{ and } \overline{f}(x) = \overline{f}(Tx),$$

by ergodicity, we have

$$(14) \qquad \underline{f}(x) = \text{constant and } \overline{f}(x) = \text{constant}$$

Fix $\varepsilon > 0$ and M.

Proof of the Birkhoff ET cont.

Proof (Cont.)

Let

$$n(x) := \inf\{n \geq 1 : S_n f(x) \leq n(\underline{f} + \varepsilon), \}$$

and

$$A := A_{M,\varepsilon} = \{x : n(x) \geq M\}.$$

By definition for every ε:

$$(15) \qquad \lim_{M\to\infty} \mu(A) = 0.$$

Proof of the Birkhoff ET cont.

Proof (Cont.)

We partition the numbers $\{0, 1, \ldots n-1\}$ into

- $\mathcal{I}_1 := \bigcup_{T^i x \notin A, i \leq n-M} \{i, i+1, \ldots, i+n(T^i x)\}$,
- $\mathcal{I}_2 := \{0 \leq i \leq n-M : T^i x \in A\}$,
- $\mathcal{I}_3 := \{n-M \leq i \leq n-1\}$.

Proof of the Birkhoff ET cont.

Proof (Cont.)

$$(16) \quad S_n f(x) = \sum_{i\in\mathcal{I}_1} f(T^i x) + \sum_{i\in\mathcal{I}_2} f(T^i x) + \sum_{i\in\mathcal{I}_3} f(T^i x).$$

Use the definition and divide both sides by n yields:

$$(17) \quad \frac{1}{n}\cdot S_n f(x) \leq \frac{\#\mathcal{I}_1}{n}\cdot(\underline{f}+\varepsilon) + \left(\frac{\#\mathcal{I}_2}{n} + \frac{M}{n}\right)\cdot\|f\|_\infty.$$

Observe that by T-invariance of μ, after integrating:

Proof of the Birkhoff ET cont.

Proof (Cont.)

- On the left hand side we get $\mu(A)$.
- On the right hand side observe that $\#\mathcal{I}_2 = \sum_{k=0}^{n-M} \mathbb{1}_A(T^k x)$. Hence

$$(18) \qquad \int \frac{\#\mathcal{I}_2}{n} d\mu = \frac{n-M}{n}\mu(A) \xrightarrow{n\to\infty} \mu(A).$$

(19)
$$\int \frac{\#\mathcal{I}_1}{n} d\mu = 1 - \frac{n-M}{n}\mu(A) - \frac{M}{n} \xrightarrow{n\to\infty} 1 - \mu(A).$$

Proof of the Birkhoff ET cont.

Proof (Cont.)

So, after taking the integral of both sides in (17) we get

(20)
$$\int f(x)d\mu(x) \leq \int \frac{\#\mathcal{I}_1}{n} d\mu \cdot (\underline{f}+\varepsilon) + \mu(A)\|f\|_\infty + \frac{M}{n}\|f\|_\infty.$$

First we let $n \to \infty$ and use (18) and (19) to get

$$(21) \quad \int f(x)d\mu(x) \leq (1-\mu(A))\cdot(\underline{f}+\varepsilon) + \mu(A)\|f\|_\infty.$$

Secondly, we let $M \to \infty$ (so $\mu(A) \to 0$ by (15)) finally, let $\varepsilon \downarrow 0$ to conclude that

Proof of the Birkhoff ET cont.

Proof (Cont.)

(22) $$\int f d\mu \leq \underline{f}.$$

Substituting f with $-f$ we obtain that

(23) $$\overline{f} \leq \int f d\mu.$$

Which implies that

$$\underline{f} = \overline{f} = \int f dm \text{ a.s.} \blacksquare$$

Rotation on the circle

Let $\Omega := S^1$ and $\mathcal{F}$ be the Borel σ-algebra and $\mathbb{P}$ be the Lebesgue measure. Further let $T : \Omega \to \Omega$ be the rotation with angle α. That is

$$Tx = x + \alpha \text{ mod } 1$$

Then T is measure preserving. (This is obvious.)

Theorem 4.13

Using the notation above, T is ergodic iff $\alpha \notin \mathbb{Q}$.

Let $f \in L^2$. Then f can be presented as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x} \text{ for } \mathbb{P}\text{-a.a. } x.$$

Rotation on the circle (Cont.)

Further,

$$f(T(x)) = \sum_{n=-\infty}^{\infty} \left(c_n e^{2\pi i n \alpha}\right) e^{2\pi i n x} \quad n \in \mathbb{Z}.$$

T is ergodic iff $f(x) = f(Tx)$ for a.a. x. That is

(24) $$c_n = c_n e^{2\pi i n \alpha} \quad n \in \mathbb{Z}.$$

- If α irrational then (24) is equivalent to $c_n = 0$ for all $n \neq 0$ which means that f is constant (a.s.) which means that T is ergodic.
- If $\alpha \in \mathbb{Q}$ then we can find $n \in \mathbb{Z} \setminus \{0\}$ such that $1 = e^{2\pi i n \alpha} = e^{-2\pi i n \alpha}$.

Ration on the circle (Cont.)

For such an n consider the function

$$f(x) = e^{2\pi i n \alpha x} + e^{-2\pi i n \alpha x} = \cos(2\pi n x).$$

This is not constant but $f(x) = f(Tx)$ so, in this case T is not ergodic.

Definitions

Definition 4.14

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $T : \Omega \to \Omega$ be a measurable map.

- The set $A \in \mathcal{F}$ is called invariant if $T^{-1}A = A$.
- $\mathcal{I} := \{A \in \mathcal{F} : T^{-1}A = A\}$ the sub σ-algebra of invariant sets.
- We say that T is measure preserving if $\mathbb{P}(A) = \mathbb{P}(T^{-1}A)$ holds for all $A \in \mathcal{F}$.
- In this case $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is called measure-preserving dynamical system.

See Homework # 44.

Notation

Let $(S, \mathcal{B})$ be a Polish space endowed with its Borel σ-algebra. Given a sequence $\{\xi_k\}_{k=0}^{\infty}$ of $(S, \mathcal{B})$-valued r.v. Define the corresponding canonical process $\{X_n\}$ as we did on slides 12-14 and the corresponding canonical dynamical systems $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ (see Definition 3.3). on the space also defined on slides 12-14.

Fact 4.15

X is stationary iff $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ is measure preserving.

Ergodicity

Definition 4.16

We say that the stochastic process in Definition 46 is ergodic if $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ is ergodic.

Remark 4.17

If the process $\{\xi_k\}$ is stationary and ergodic with mean μ then it follows from Birkhoff Ergodic Therem that

$$\lim_{n \to \infty} \frac{\xi_1 + \cdots + \xi_n}{n} = \mu.$$

Relation between Ergodicity and iid

Let $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ and $X = (X_i)$ be as in Definition 46.

Fact 4.18

If $(X_n)_{n=0}^{\infty}$ is iid then $(X_n)_{n=0}^{\infty}$ is ergodic.

Proof

Let $A \in \mathcal{I}$ that is $\sigma^{-1}(A) = A$. Then for every n:

$$A = \sigma^{-n}(A) = \{\omega : \sigma^n(\omega) \in A\} \in \sigma(X_n, X_{n+1}, \dots).$$

That is $A \in \mathcal{T}$, where $\mathcal{T}$ is the tail σ-algebra. Hence $\mathcal{I} \subset \mathcal{T}$. Since by Kolmogorov $0 - 1$ law $\mathcal{T}$ is trivial, so $\mathcal{I}$ is trivial. This implies that $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ is ergodic. So, by the definition, $\{X_n\}$ is ergodic.

The organization of the rest of the Section

- Definition 4.16 defines ergodicity for a general stochastic process $\{\xi_k\}$. However, we have alraedy defined ergodicity for Markov chain in the Course "Stochastic processes". Below we answer the question: Are these two definitions equivalent for Markov chains?
- However, before doing that we recall some theorems proved in the Course "Stochastic processes".
- However, even before that we give a little bit more general definitions than the ones we learned in the previous courses.

1 A number of boring definitions
2 Kolmogorov Extension Theorem
3 Stationary processes
4 Birkhoff pointwise Ergodic Theorem
- Invariant and ergodic measures
- Weak $*$-topology
- existence of invariant measure
- Birkhoff ET Theorem
5 Revisit Markov chains
6 Martingal CHT
- Recall: CLT for iid with finite mean and variance
- CLT for martingales with stationary L^2 differences
- CLT for Markov chains

Notation used in the subsection

Here we follow Klenke's book [15, Chapter 17].

- Let $(S, \mathcal{B})$ be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$
- Let $\mathcal{I} \subset \mathbb{R}$ be the index set. We assume that $0 \in \mathcal{I}$ and $\mathcal{I}$ is closed for addition.
- Let $X = (X_t)_{t \in \mathcal{I}}$ be $(S, \mathcal{B})$-valued stochastic process.
- We write $\mathbb{F} := (\mathcal{F}_t)_{t \in \mathcal{I}}$ for the filtration defined by X. That is $\mathbb{F}$ is an increasing sequence of σ-algebras with $\mathcal{F}_s \subset \mathcal{F}_t$ and $X_s \in \mathcal{F}_s$ for all $s \leq t$.

Markov process: the definition

Definition

We say that $X = (X_t)_{t \in \mathcal{I}}$ is a time-homogeneous Markov process with distributions $(\mathbb{P}_x)_{x \in S}$ on the space $(\Omega, \mathcal{A})$ if

- $\forall x \in S$, X is a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P}_x)$, with $\mathbb{P}_x (X_0 = x) = 1$.
- The map

$$\kappa : S \times \mathcal{B}^{\mathcal{I}} \to [0,1], \qquad \kappa(x, A) := \mathbb{P}_x (X \in A)$$

is a stochastic kernel. (See slide # 54.)

Markov process: the definition (Cont.)

Definition (Cont.)

- X has the time-homogeneous Markov Property (MP). This means that

(25) $\forall B \in \mathcal{B}, \ \forall x \in S, \ s, t \in \mathcal{I},$

$$\mathbb{P}_x (X_{t+s} \in B | \mathcal{F}_s) = \kappa_t(X_s, B),$$

where for $t \in \mathcal{I}$, the map $\kappa_t : S \times \mathcal{B} \to [0,1]$ is the so called transition kernel defined by:

(26) $\forall x \in E, \ B \in \mathcal{B} \quad \kappa_t(x, B) :=$

$$\kappa \left(x, \left\{\mathbf{y} \in S^{\mathcal{I}} : y(t) \in B\right\}\right) = \mathbb{P}_x (X_t \in B),$$

Stochastic kernel

where κ is a stochastic kernel defined below. Consider the measurable spaces $(S, \mathcal{B})$ and $(S^{\mathcal{I}}, \mathcal{B}^{\mathcal{I}})$. The map $\kappa : S \times \mathcal{B}^{\mathcal{I}} \to [0,1]$

$$\kappa(x, A) := \mathbb{P}_x (X \in A), \quad \forall A \in \mathcal{B}^{\mathcal{I}}$$

is a stochastic kernel or Markov kernel. It satisfies

- for every fixed Borel set $A \in \mathcal{B}^{\mathcal{I}}$ the map, $x \mapsto \kappa(x, A)$ is $\mathcal{B}$-measurable.
- for every fixed $x \in S$, the map $A \mapsto \kappa(x, A)$ is a Borel probability measure on the measurable space $(S^{\mathcal{I}}, \mathcal{B}^{\mathcal{I}})$.

Markov process: the definition (Cont.)

Definition (Cont.)

- The family

$$(\kappa_t(x, B), t \in \mathcal{I}, x \in S, B \in \mathcal{B})$$

is the family of transition probabilities of X.
- If S is countable then X is a discrete Markov process.
- We write $\mathbb{E}_x$ for the expectation w.r.t. $\mathbb{P}_x$.
- If $\mathcal{I} = \mathbb{N} = \{0, 1, 2, \dots\}$ then X is a Markov chain. In this case κ_n is the family of n-steps probabilities.

Recall: Markov chains

In the course Stochastic Processes in File B we called a Markov chain ergodic if it is

Definition 5.1 (Ergodic Markov Chain)

- irreducible,
- every state is aperiodic,
- every state is positive recurrent,

If a state is both aperiodic and positive recurrent then the state was called ergodic. In the course Stochastic Processes in File B we proved the following two theorems:

Recall: Markov chains (Cont.)

Theorem 5.2

For an ergodic Markov Chain with transition probability matrix $P = p(i,j)$ there exists a unique stationary distribution π

(27) $$\lim_{n\to\infty} p^n(j,i) = \pi(i) = \frac{1}{\mathbb{E}_i[T_i]},$$

where
- $p^n(j,i)$ is the probability that starting from j we are in i in n steps,
- $T_i \geq 1$ is the first time we get to i and
- $\mathbb{E}_j$: the expectation conditional on starting from j.

Recall: Markov chains (Cont.)

Theorem 5.3

If
- ξ_k is an ergodic Markov chain with (countable) state space S and
- stationary distribution π and
- $f : S \to \mathbb{R}$ is a function satisfying: $\int |f(i)| d\pi(i) = \sum_i |f(i)|\pi(i) < \infty$

Then almost surely:

(28) $$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(\xi_k) = \sum_{x\in S} f(x)\pi(x) = \int_S f(x)d\pi(x).$$

Theorem 5.3 vs. Birkhoff ET (Cont.)

Let us define $\widetilde{f} : \Omega \to \mathbb{R}$,

$$\widetilde{f}(\omega) := f(\omega_1)$$

Then related to the left hand side (28):

$$\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) = \frac{1}{n}\sum_{k=0}^{n-1} \widetilde{f}\left(\sigma^k\omega\right).$$

Related to the right hand side (28):

$$\int_S f(x)d\pi(x) = \sum_{x\in S} f(x)\pi(x) = \sum_{\omega_1 = x} \widetilde{f}(\omega)\mathbb{P}(\omega_1 = x)$$
$$= \sum_{x\in S}\int_{\omega_1 = x} \widetilde{f}(\omega)d\mathbb{P}(\omega) = \int_\Omega \widetilde{f}(\omega)d\mathbb{P}(\omega).$$

Concluion

This indicates that.

Theorem 5.4

For every Markov chain which is ergodic in sense of Definition 5.1 the corresponding cannonical dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ is also ergodic.

The proof can be found in [15, Example 20.17] So, the definition of ergodicity we gave above is in coherence with that of we gave in Definition 5.1.

Be careful! In the Theorem above the opposite implication is NOT true!

1. A number of boring definitions
2. Kolmogorov Extension Theorem
3. Stationary processes
4. Birkhoff pointwise Ergodic Theorem
   - Invariant and ergodic measures
   - Weak $*$-topology
   - existence of invariant measure
   - Birkhoff ET Theorem
5. Revisit Markov chains
6. Martingal CHT
   - Recall: CLT for iid with finite mean and variance
   - CLT for martingales with stationary L^2 differences
   - CLT for Markov chains

Recall: Central Limit Theorem as you learned in the course Probability I

Theorem 6.1 (Central Limit Theorem (CLT))

Let $X_1, X_2, \dots$ be i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\mathrm{var}(X_i) = \sigma^2 \in (0,\infty)$. If $S_n = X_1 + \dots + X_n$ then

(29) $$\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} \Rightarrow \kappa,$$

where κ is the standard normal distribution and "$\Rightarrow$" means the convergence in distribution.

Recall: Central Limit Theorem as you learned in the course Probability I (cont.)

That is

(30) $$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} < a\right) = \Phi(a),$$

where Φ is the distribution function of the standard normal distribution.

The following theorem is the martingale Central Limit Theorem (CLT). Although we do not have time to prove it (the proof is available in[24, Chapter 6.5]) we will study how to apply it.

CLT for Martingales

Theorem 6.2

Let $\xi_1, \xi_2, \dots\dots$ be an ergodic stationary sequence of random variables with $\mathbb{E}[\xi_n] = 0$. We assume that there exist a square integrable martingale $(M_n)_{n=0}^{\infty}$ such that $M_n - M_{n-1} = \xi_n$, $n = 1, 2, \dots$. Then the CLT holds for $\xi_1, \xi_2, \dots$. That is

(31) $$\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \mathbb{E}\left[\xi_1^2\right]$ and "$\Rightarrow$" means the convergence in distribution.

CHT for Martingales

Corollary 6.3

Let X_n be a stationary stochastic process with $\mathbb{E}[X_n]=0$. Assume that $\exists \xi_n, \eta_n$ such that:

- (A1) *ξ_n is a square integrable martingale difference with $\mathbb{E}[\xi_n]=0$, $\mathrm{Var}(\xi_n)=\sigma^2$ (like in Theorem 6.2).*
- (A2) *η_n is small: $\mathbb{E}\left[\left(\sum_{j=1}^{n}\eta_j\right)^2\right]=\mathfrak{o}(n)$ (little o n).*
- (A3) *$X_n=\xi_{n+1}+\eta_{n+1}$ for all n.*

Then $\frac{X_1+\cdots+X_n}{\sqrt{n}} \Rightarrow \mathcal{N}(0,\sigma^2)$, where "$\Rightarrow$" means the convergence in distribution.

Proof (Cont.)

Proof of the Corollary (Cont.)

Let $H_\varepsilon^n := \left\{ \frac{\left|\sum_{i=1}^{n}\eta_{i+1}\right|}{\sqrt{n}} \geq \varepsilon \right\}$. Observe that by Markov inequality:

(32) $$h_\varepsilon^n := \mathbb{P}(H_\varepsilon^n) \leq \frac{1}{\varepsilon^2}\mathbb{E}\left[\frac{\left(\sum_{i=1}^{n}\eta_{i+1}\right)^2}{n}\right]$$

So, by Assumption (A2) we have

(33) $$\lim_{n\to\infty} h_\varepsilon^n = 0.$$

Proof (Cont.)

Proof of the Corollary (Cont.)

(34)
$$p_n := \mathbb{P}\left(\frac{\sum_{i=1}^{n}X_i}{\sqrt{n}} < a\right) = \mathbb{P}\left(\frac{\sum_{i=1}^{n}\xi_{i+1}}{\sqrt{n}} < a - \frac{\sum_{i=1}^{n}\eta_{i+1}}{\sqrt{n}}; H_\varepsilon^n\right)$$
$$+\,\mathbb{P}\left(\frac{\sum_{i=1}^{n}\xi_{i+1}}{\sqrt{n}} < a - \frac{\sum_{i=1}^{n}\eta_{i+1}}{\sqrt{n}}; (H_\varepsilon^n)^c\right)$$

Proof (Cont.)

Proof of the Corollary (Cont.)

From here we obtain that

$$\mathbb{P}\left(\frac{\sum_{i=1}^{n}\xi_{i+1}}{\sqrt{n}} < a-\varepsilon\right) - h_n^\varepsilon \leq p_n$$
$$\leq \mathbb{P}\left(\frac{\sum_{i=1}^{n}\xi_{i+1}}{\sqrt{n}} < a+\varepsilon\right) + h_n^\varepsilon$$

Proof (Cont.)

Proof of the Corollary (Cont.)

Now we let $n\to\infty$ and then $\varepsilon \downarrow 0$ to get by Theorem 6.2:

(35) $$\frac{\sum_{i=1}^{n}X_i}{\sqrt{n}} \Rightarrow \mathcal{N}(0,\sigma^2).\blacksquare$$

Find a representation like in (A3)

Given a stationary process $X=\{X_n\}$ with zero mean. That is $\mathbb{E}[X_n]=0$ for all n. Our next assumption is:

- (A4) The following rv. exist and square integrable:

(36) $$Z_n := \sum_{\ell=n}^{\infty}\mathbb{E}[X_\ell|\mathcal{F}_n]$$

Claim 1

- (a) $M_n := Z_n + \sum_{i\leq n-1} X_i$ is a martingale.
- (b) $\mathbb{E}\left[Z_{n+1}^2\right]=\mathfrak{o}(n)$.

Find a representation like in (A3) (Cont.)

Proof of part (a) of the Claim

$M_n \in \mathcal{F}_n$ by definition. $\mathbb{E}[M_{n+1}|\mathcal{F}_n]=M_n$ by the tower property and the definition. $M_n\in L^1$ because $Z_n\in L^1$ and $\sum_{i\leq n-1}X_i \in L^1$.

Proof of part (b) of the Claim

This is immediate from the fact that $\{X_n\}$ is stationary so $\{Z_n\}$ is stationary, so $\mathbb{E}\left[Z_{n+1}^2\right]=\mathbb{E}\left[Z_1^2\right]<\infty$.

Find a representation like in (A3) (Cont.)

Assume that assumption (A4) holds. Let

(37) $$\xi_{n+1} := Z_{n+1} - \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = Z_{n+1}-Z_n+X_n$$

and

$$\eta_{n+1} := Z_n - Z_{n+1}.$$

Then

$$X_n = \xi_{n+1}+\eta_{n+1}$$

and $\xi_n = M_{n+1}-M_n$, and $\xi_n\in L^2$, further

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\eta_i\right)^2\right] = \mathbb{E}\left[(Z_1-Z_{n+1})^2\right] = \mathfrak{o}(n).$$

Find a representation like in (A3) (Cont.)

This shows that

Theorem 6.4

If $X = (X_n)$ is a stationary process with zero mean and we assume (A4) on slide # 70. Then the CLT holds for (X_n) with $\sigma^2 = \mathrm{Var}(X_n)$:

$$\text{(38)} \qquad \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \sigma^2),$$

where "$\Rightarrow$" means the convergence in distribution.

Application for Markov Chains

Given a stationary ergodic Markov chain X with finite or countable state space S. The probability transition matrix is $(p(x,y))_{x,y\in S}$ and the unique stationary distribution is π. On $L^1(S,\pi)$ we define the operator P by

$$(Pf)(i) := \sum_{j\in S} p(i,j) f(j).$$

Remeber that on slide # 10 in File A we have already introduced this operator. Let $f : S \to \mathbb{R}$, $f \in L^2(\pi)$, satisfying

$$\text{(39)} \qquad \mathbb{E}_\pi [f] := \sum_{i\in S} f(i)\pi(i) = 0$$

Application for Markov Chains (Cont.)

We define

$$Y_i := f(X_i).$$

We assume

(A5) $\exists \mathbf{U} : S \to \mathbb{R}$, $\mathbf{U} \in L^2(\pi)$ such that

(40)

$$((I-P)\cdot \mathbf{U})(i) = f(i), \quad \text{for all } i \in S.$$

Then like in (36):

$$\text{(41)} \qquad Z_n = \sum_{j=0}^{\infty} \mathbb{E}\left[f(X_{n+j})|X_n\right] = \sum_{j=0}^{\infty} (P^j \cdot \mathbf{f})(X_n),$$

Application for Markov Chains (Cont.)

where we consider f as a vector $\mathbf{f}$ indexed by the elements of S and $(P^j \cdot \mathbf{f})(X_n)$ is the $X_n \in S$-th component of the vector $(P^j \cdot \mathbf{f})$. Observe that

$$\lim_{n\to\infty} P^n =: \Pi$$

where Π is a matrix which for which all the rows are equal to π. Then

$$\text{(42)} \qquad \sum_{k=0}^{\infty} P^k (I-P) = I - \lim_{n\to\infty} P^{n+1} = I - \Pi.$$

Apply this, (40) and (41) to obtain

Application for Markov Chains (Cont.)

$$Z_n = U(X_n) - \mathbb{E}_\pi [\mathbf{U}]$$

It follows from (37) that

$$\xi_{n+1} = Z_{n+1} - Z_n + f(X_n) = U(X_{n+1}) - U(X_n) + f(X_n).$$

is an L^2 martingale difference. Hence, by Corollary 6.3 we have:

Application for Markov Chains (Cont.)

Theorem 6.5 (Markov chain CLT)

Let X_n be a stationary and ergodic Markov chain with stationary distribution π. Let $f : S \to \mathbb{R}$, $f \in L^2(\pi)$, and $\mathbb{E}_\pi [f] = 0$. Further we assume the assumption (A5) on slide #75 holds . Then

$$\text{(43)} \qquad \frac{f(X_1) + \cdots + f(X_n)}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \mathbb{E}_\pi \left[\xi_1^2\right] = \mathbb{E}_\pi \left[(U(X_1) - U(X_0) + f(X_0))^2\right].$$

We compute all of these on a simple example:

Example

Let $S := \{1,2\}$ and the transition probability matrix

$$P := \begin{pmatrix} p & q \\ q & p \end{pmatrix}, \quad p + q = 1, \ p, q \geq 0.$$

Then $\pi = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$. Let $f \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then $\mathbb{E}_\pi [f] = 0$ and $I - P = \begin{pmatrix} q & -q \\ -q & q \end{pmatrix}$

$$\text{(44)} \qquad (I-P)\cdot \mathbf{U} = \begin{pmatrix} q & -q \\ -q & q \end{pmatrix} \cdot \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Example (Cont.

The solution of (44) is $\mathbf{U} = \begin{pmatrix} \frac{1}{q} + a \\ a \end{pmatrix}$ for every a, w.l.g. we may choose $a = 0$ (see the next formula).

$$\begin{aligned}
\sigma^2 &= \mathbb{E}_\pi \left[(\mathbf{U}(X_1) - \mathbf{U}(X_0) + f(X_0))^2\right] \\
&= \frac{1}{2}\cdot p \cdot \left(\frac{1}{q} - \frac{1}{q} + 1\right)^2 + \frac{1}{2}\cdot q \cdot \left(0 - \frac{1}{q} + 1\right)^2 \\
&\quad + \frac{1}{2}\cdot q \cdot \left(\frac{1}{q} - 0 - 1\right)^2 + \frac{1}{2}\cdot p \cdot (0 - 0 - 1)^2 \\
&= \frac{p}{q}.
\end{aligned}$$

Example (Cont.)

Conclusion: Let

$$S_n := (\#\text{visit to } 1) - (\#\text{visit to } -1).$$

Then

$$\frac{S_n}{\sqrt{n}} \Rightarrow \mathcal{N}(0, \sigma^2).$$

- $p \downarrow 0$: $\sigma \downarrow 0$ (determinisztikus)
- $p = \frac{1}{2}$: $\sigma = 1$ the digits $1, -1$ appear with $\frac{1}{2} - \frac{1}{2}$ probability
- $p \uparrow 1$: sign change happens rarely, $\sigma \to \infty$.

[1] BalÃązs MÃąrton, TÃşth BÃąlint
ValÃşszÃŋnÅśsÃľgszÃąmÃŋtÃąs 1. jegyzet matematikusoknak Ãľs fizikusoknak
BÃąlÃązs MÃąrton Honlapja, 2012. Az internettes vÃąltozatÃľrt kattintson ide.

[2] P. Billingsley
Convergence of probability measures
Wiley, 1968

[3] B. Driver
Analysis tools with examples
Lecturenotes, 2012. Click here.

[4] R. Durrett
Essentials of Stochastic Processes, Second edition
Springer, 2012. Click here

[5] R. Durrett
Probability: Theory with examples, 4th edition
Cambridge University Press, 2010.

[6] R. Durrett
Probability: Theory and Examples
Click here

[7] D.H. Fremlin
Measure Theory Volume I
Click here

[8] D.H. Fremlin
Measure Theory Volume II
Click here

[9] O. van Gaans
Probability measures on metric spaces
Click here

[10] I.I. Gihman, A.V. Szkorohod
BevezetÃľs a sztochasztikus folyamatok elmÃľletÃľbe
MÅśszaki KÃűnyvkiadÃş1975, Budapest, 1985

[11] S. Karlin, H.M. Taylor
A first course in stochastic processes
Academic Press, New York, 1975

[12] S. Karlin, H.M. Taylor
Sztochasztikus Folyamatok
Gondolat, Budapest, 1985

[13] S. Karlin, H.M. Taylor
A second course in stochastic processes
, Academic Press, 1981

[14] A.F. Karr
Probability
Springer-Verlag, 1993

[15] A. Klenke
Probability Theory, A comprehensive course, 2nd ed.
Springer, 2013.

[16] G. Lawler
Intoduction to Stochastic Processes
Chapmann & Hall 1995.

[17] D.A. Levin, Y. Peres,E.L. Wilmer
Markov chains and mixing times
American Mathematical Society, 2009.

[18] Major PÃľter
Folytonos idejÅś Markov lÃąncok http://www.renyi.hu/~major/debrecen/debrecen2008a/markov3.html

[19] P. Mattila Geometry of sets and measure in Euclidean spaces. Cambridge, 1995.

[20] RÃľnyi AlfrÃľd
ValÃşszÃŋnÅśsÃľgszÃąmÃŋtÃąs, (negyedik kiadÃąs)
TankÃűnyvkiadÃş Budapest, 1981.

[21] S.I. Resnik
A Probability Path
Birkhäuser, 2005

[22] S. Ross
A First Course in Probability, 6th ed.
Prentice Hall, 2002

[23] TÃşth BÃąlint *Sztochasztikus folyamatok jegyzet*
TÃşth BÃąlint JegyzetÃľrt kattintson ide

[24] S.R.S: Varadhan
Probability Theory
Cambridge 2005

[25] P. Walters
An Introduction to Ergodic Theory
Springer, 1982.

[26] D. Williams
Probability with Martingales
Lecture Notes, Courant Institute 2000