Probability Theory

Instructor: Noam Berger

Lecture at TUM in Summer Semester 2014

July 8, 2014

Produced by me

Contents

1 Measure spaces

8.4

In the first few lectures we give, without proofs, background from Measure Theory which we will need in the Probability course, and set some of the basic definitions of Probability Theory.

Let Ω be a non-empty set.

Definition 1.1 $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, i.e. a collection $\mathcal{F}$ of subsets of Ω, is a *σ-algebra* on Ω, if

(i) $\mathcal{F} \neq \emptyset$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(iii) $A_n \in \mathcal{F},\ \forall n \in \mathbb{N} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$

Remark 1.2 Note that If $A_n \in \mathcal{F},\ \forall n \in \mathbb{N}$, then also $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Definition 1.3 If $\mathcal{F}$ is a σ-algebra on Ω, $(\Omega, \mathcal{F})$ is a *measurable space* and each $A \in \mathcal{F}$ is *measurable*.

Definition 1.4 Let $\mathcal{U} \subseteq 2^\Omega$. The *$\sigma$-algebra generated by $\mathcal{U}$* is defined to be

$$\sigma(\mathcal{U}) := \bigcap_{\substack{\mathcal{U} \subseteq \mathcal{F} \\ \mathcal{F}\ \sigma\text{-algebra}}} \mathcal{F}$$

Definition 1.5 Let (Ω, T) be a topological space. *Borel's σ-algebra w.r.t. the space* (Ω, T) is the σ-algebra $\sigma(T)$. It is usually denoted by $\mathcal{B}$.

We will mostly be interested in Borel's σ-algebra in the case of the Euclidean spaces $\mathbb{R}^d$.

Definition 1.6 Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be mesaurable spaces. The *product space* is defined as follows: The space is $\Omega = \Omega_1 \times \Omega_2$, and the σ-algebra is $\mathcal{F} = \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$.

Remark 1.7 Often notation is abused and the product σ-algebra is denoted $\mathcal{F}_1 \times \mathcal{F}_2$. Note that this is not the cartesian product of $\mathcal{F}_1$ and $\mathcal{F}_2$.

Definition 1.6 can be easily extended to any finite number of spaces. Note that we seem to have two natural ways of defining a σ-algebra on the space R^2 - using Borel directly, or multiplying the one-dimensional Borel σ-algebra with itself. However, both yield the same σ-algebra.

We now define the product σ-algebra of infinitely many spaces. Let $(\Omega_k, \mathcal{F}_k)_{k=1}^\infty$ be a collection of measurable spaces.

Definition 1.8 A *cylinder* is a set $A \subseteq \prod_{k=1}^\infty \Omega_k$ s.t. $A = \prod_{k=1}^\infty A_k$ such that

1. $A_k \in \mathcal{F}_k$ for every k, and
2. There exists k_0 s.t. $A_k = \Omega_k$ for all $k > k_0$.

Now, the *product σ-algebra* is defined to be the σ-algebra generated by the set of cylinders.

Remark 1.9 There is a slightly simpler way of defining the same object.

Definition 1.10 Let $\Omega, \mathcal{F}$ be a measurable space. A *measure* on $\Omega, \mathcal{F}$ is a function $\mu : \mathcal{F} \to [0, \infty]$ s.t:

1. $\mu(\emptyset) = 0$.
2. if $(A_k)_{k=1}^{\infty}$ are pairwise disjoint, then

$$\mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

A measure is called *σ-finite* if Ω is the union of countably many sets of finite measure. A measure P is called a *probability measure* if $P(\Omega) = 1$.

Example 1.11 Let Ω be finite, $\mathcal{F} = 2^{\Omega}$, and $P(A) = \frac{|A|}{|\Omega|}$. Then P is a probability measure.

Theorem 1.12 *There exist a unique measure λ on $(R, \mathcal{B})$ such that $\lambda([a, b]) = b - a$ for every $b > a$. λ is called* Lebesgue's measure.

Theorem 1.13 *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ σ-finite measure spaces. Then there exists a unique measure μ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ s.t. $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for every $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$.*

Theorem 1.14 *Let $(\Omega_k, \mathcal{F}_k, \mu_k)_{k=1}^{\infty}$ probability spaces. Then there exists a unique measure μ on $\prod_{k=1}^{\infty} \Omega_k, \prod_{k=1}^{\infty} \mathcal{F}_k$ such that $\mu(A) = \prod_{k=1}^{\infty} \mu_k(A_k)$ for every cylinder $A = \prod_{k=1}^{\infty} A_k$.*

11.4

Definition 1.15 Let $(\Omega, \mathcal{F})$ be a measurable space and let (X, T) be a topological space. $f : \Omega \to X$ is called a *measurable function* if $f^{-1}(A) \in \mathcal{F}$ for all $a \in T$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. An *event* is a measurable set. A *random variable* is a measurable function from Ω to $\mathbb{R}$. The σ-algebra generated by a random variable X is $\sigma(X) := \sigma(\{X^{-1}(A) : A \in T\})$.

1.1 integration

Let $(\Omega, \mathcal{F}, \mu)$be a σ-finite measure space. A measurable function $f : \Omega \to \mathbb{R}$ is called *simple* if it is non-negative and there is a partition $A_1, A_2, \ldots, A_n$ of Ω s.t. A_k is measurable for all k, and f is constant on each A_k. We define the *integral* of f to be

$$\int_\Omega f d\mu := \sum_{k=1}^{n} \mu(A_k) \cdot f|_{A_k}.$$

Exercise Show that this is well defined.
For a nonnegative measurable function f we define

$$\int_\Omega f d\mu = \sup \left\{ \int_\Omega g d\mu \ : \ g \leq f; g \text{ simple} \right\}.$$

Define

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) \leq 0 \end{cases} \quad ; \quad f^-(x) = \begin{cases} -f(x) & \text{if } f(x) \leq 0 \\ 0 & \text{if } f(x) \geq 0 \end{cases}.$$

Exercise Show that if f is measurable, then so are f^+ and f^-.

Definition 1.16 A measurable function f is said to be *inrtegrable* if $\int_\Omega f^+ d\mu < \infty$ and $\int_\Omega f^- d\mu < \infty$. In this case we define

$$\int_\Omega f d\mu := \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu. \tag{1.1}$$

We also define (infinite) integrals of non-integrable functions if the difference in (1.1) makes sense.
We say that a random variable X on a probability space $(\Omega, \mathcal{F}, P)$ has an expectation if it is integrable. In this case we write $E(X) := \int_\Omega X dP$. We say that X has a variance if X^2 has an expectation, and write

$$\text{var}(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2.$$

We say that X has a k-th moment if $E(|X|^k) < \infty$, and in this case the k-th moment of X is $E(X^k)$. For two variable X and Y, if they both have expectations and the variable XY has expectation too, then we say that they have a covariance, and define

$$\text{cov}(X, Y) := E(XY) - E(X)E(Y).$$

Exercise (1) show that if X and Y have variances, then they have a covariance. (2) show that this is not an "if and only if".
We now state without proof some theorems about convergence of integrals, which we will use in the course. For all of these theorems, $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space, typically a probability space. $(f_k : \Omega \to \mathbb{R})_k = 1^\infty$ are measurable.

Theorem 1.17 (Fatou's lemma)
If the functions f_k are non-negative, then

$$\int_\Omega \liminf f_k d\mu \leq \liminf \int_\Omega f_k d\mu$$

We now assume that there exists a function $f : \Omega \to \mathbb{R}$ such that

$$\mu\left(\left\{x \in \Omega \ : \ \lim_{k\to\infty} f_k(x) \neq f(x)\right\}\right) = 0.$$

Theorem 1.18 (Monotone convergence theorem)
If, in addition, f_k is a pointwise increasing sequence, and f_k are non-negative, then

$$\int_\Omega f d\mu = \lim \int_\Omega f_k d\mu$$

Theorem 1.19 (Dominated convergence theorem)
If there exists a non-negative integrable function g s.t. $|f_k(x)| \leq g(x)$ for every k and (almost) all x, then

$$\int_\Omega f d\mu = \lim \int_\Omega f_k d\mu$$

15.4

1.2 Calculus of probabilities

Let $(A_k)_{k=1}^{\infty}$ be a sequence of events. We define:

$$\liminf_{k\to\infty} A_k = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n \quad ; \quad \limsup_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n.$$

If $\liminf_{k\to\infty} A_k = \limsup_{k\to\infty} A_k$ we say that the sequence converges. We also say that the sequence converges if the equality is up to measure zero.

Example 1.20

1. If the sequence $(A_k)_{k=1}^{\infty}$ is increasing, then it converges to $\cup_{k=1^\infty} A_k$.

2. If the sequence $(A_k)_{k=1}^{\infty}$ is decreasing, then it converges to $\cap_{k=1^\infty} A_k$.

Theorem 1.21 (Continuity of Probability)
Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $(A_k)_{k=1}^{\infty}$ be a sequence of events. Then

$$P\left[\liminf_{k\to\infty} A_k\right] \leq \liminf_{k\to\infty} P[A_k], \ \text{and} \tag{1.2}$$

$$P\left[\limsup_{k\to\infty} A_k\right] \geq \limsup_{k\to\infty} P[A_k]. \tag{1.3}$$

As an immediate corollary we get the following:

Corollary 1.22 *If $(A_k)_{k=1}^{\infty}$ is a converging sequence of events, then*

$$P\left[\lim_{k\to\infty} A_k\right] = \lim_{k\to\infty} P[A_k].$$

Proof of Theorem 1.21: We start by showing this for increasing sequences. Let $(A_k)_{k=1}^\infty$ be increasing, and let $B_1 := A_1$ and $B_{k+1} := A_{k+1} - A_k$. Then the events $B_k, k = 1, \ldots$ are pairwise disjoint, and we have that $A_n = \cup_{k=1}^n B_k$ for every n, and in the limit $\lim_{k\to\infty} A_k = \cup_{k=1}^\infty B_k$. Therefore we get

$$P\left[\bigcup_{k=1}^\infty A_k\right] = P\left[\lim_{k\to\infty} A_k\right] = P\left[\bigcup_{k=1}^\infty B_k\right] = \sum_{k=1}^\infty P(B_k) = \lim_{n\to\infty}\sum_{k=1}^n P(B_k) = \lim_{n\to\infty} P(A_n).$$

Next we note that, by taking complements, the same holds for decreasing sequences. We now prove (1.2): For every k, Let

$$B_k = \bigcap_{n=k}^\infty A_n.$$

Then $B_k \subseteq A_k$ and therefore $P[B_k] \le P[A_k]$. Now note that (B_k) is an increasing sequence and that

$$\liminf_{k\to\infty} A_k = \lim_{k\to\infty} B_k.$$

Therefore,

$$P\left[\liminf_{k\to\infty} A_k\right] = P\left[\lim_{k\to\infty} B_k\right] = \lim_{k\to\infty} P[B_k] \le \liminf_{k\to\infty} P[A_k].$$

(1.3) is proven analogously. □

1.3 Distributions and independence

1.3.1 Independence

We say that two events A_1 and A_2 are *independent* if $P(A_1 \cap A_2) = P(A_1)P(A_2)$. We say that k events $A_1, \ldots, A_k$ are independent if for every nonempty subset $L \subseteq \{1, \ldots, k\}$,

$$P\left[\bigcap_{j\in L} A_j\right] = \prod_{j\in L} P[A_j].$$

Exercise Find three events A_1, A_2, A_3 s.t. every two are independent, but the three are not independent.

We say that a sequence of events $(A_k)_{k=1}^\infty$ is independent if every finite sub-collection is independent.

We can also define the very useful notion of independence of σ-algebras: Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{F}_1 \subseteq \mathcal{F}$ and $\mathcal{F}_2 \subseteq \mathcal{F}$ be σ-algebras. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent if A_1 and A_2 are independent for every $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$. equivalently, we can define independence of larger collections of σ-algebras - $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are independent if $A_1, \ldots, A_k$ are independent for all choices of $A_j \in \mathcal{F}_j, j = 1, \ldots, k$, and infinite collections of σ-algebras are independent if every finite sub-collection is independent.

Example 1.23 Let $\Omega = [0,1]^2$, let $\mathcal{B}^{(2)}$ be the Borel σ-algebra for Ω, and let P be the two dimensional Lebesgue measure. Then $(\Omega, \mathcal{B}^{(2)}, P)$ is a probability space. Let $\mathcal{B}^{(1)}$ be the Borel σ-algebra of $[0,1]$. Let $\mathcal{F}_1 = \{A \times [0,1] \ : \ A \in \mathcal{B}^{(1)}\}$, and let $\mathcal{F}_2 = \{[0,1] \times A \ : \ A \in \mathcal{B}^{(1)}\}$. Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent σ-algebras.

We can now define independence of random variable: A collection of random variables is independent if the collection of induced σ-algebras is independent. We can also speak about independence of a σ-algebra from an event or of a random variable from a σ-algebra. 25.4

1.3.2 Distributions

Let $(\Omega, \mathcal{F}, P)$ be a Probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. The *distribution* of X is the Borel probability measure D_X on $\mathbb{R}$ induced by X. More precisely, for every Borel set $A \subseteq \mathbb{R}$, we take $D_X(A) = P\big(X^{-1}(A)\big)$.

Definition 1.24 The *distribution function* $F_X : \mathbb{R} \to \mathbb{R}$ of X is defined to be the function $F_X(A) = D_X\big((-\infty, a]\big) = P(X \le a)$.

Definition 1.25 Let $(\Omega, \mathcal{F}, P)$ be a probability space.

1. Let $X_1, \dots, X_k$ be random variables. The *joint distribution* of $X_1, \dots, X_k$ is the Borel probability measure $D_{X_1,\dots,X_k}$ on $\mathbb{R}^k$ defined by

$$D_{X_1,\dots,X_k}(A) = P\Big(\big\{\omega \in \Omega \ : \ (X_1(\omega), \dots, X_k(\omega)) \in A\big\}\Big).$$

2. Let $(X_k)_{k=1}^{\infty}$ be random variables. The *joint distribution* of $(X_k)_{k=1}^{\infty}$ is the Borel probability measure $D_{X_1,\dots,}$ on $\mathbb{R}^{\mathbb{N}}$ defined by

$$D_{X_1,\dots}(A) = P\Big(\big\{\omega \in \Omega \ : \ (X_1(\omega), X_2(\omega), \dots) \in A\big\}\Big).$$

1.4 Absolute continuity and the Radon-Nikodym's theorem

Today we will prove a measure theoretic theorem which is very useful in Probability Theory. Let $(\Omega, \mathcal{F})$ be a measurable space. Let μ and ν be measures on $(\Omega, \mathcal{F})$.

Definition 1.26

1. We say that ν is *absolutely continuous* with respect to μ if for every set $A \in \mathcal{F}$ such that $\mu(A) = 0$ we also have $\nu(A) = 0$. We denote this by $\nu \ll \mu$.

2. We say the ν and μ are *equivalent* if $\mu \ll \nu$ and $\nu \ll \mu$. We denote this by $\mu \sim \nu$.

3. We say that ν and μ are *singular* if there exists $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$. We denote this by $\mu \perp \nu$.

Example 1.27 Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space, and let $f : \Omega \to \mathbb{R}$ be measurable and non-negative. Define $\nu : \mathcal{F} \to \mathbb{R}$ by

$$\nu(A) = \int_A f d\mu := \int_\Omega f \mathbf{1}_A d\mu.$$

Then $\nu \ll \mu$.

Exercise Show that ν is a measure, and that it is absolutely continuous with respect to μ.

The next theorem will show that Exampel 1.27 is, in fact, the general case of absolute continuity.

Theorem 1.28 (Radon-Nikodym)
Let $(\Omega, \mathcal{F})$ be a measurable space, and let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. Assume, in addition, that $\nu \ll \mu$. Then there exists a measurable and non-negative $f : \Omega \to \mathbb{R}$ such that for every $A \in \mathcal{F}$,

$$\nu(A) = \int_A f d\mu. \tag{1.4}$$

Furthermore, f is unique up to measure zero. f is called the Radon-Nikodym derivative *of ν with respect to μ, and is denoted $\frac{d\nu}{d\mu}$.*

Exercise Find a counter example when the assumption of σ-finiteness is removed (it is sufficient when μ is not σ-finite).

The next definition and theorem (which we will not prove) diverge from our material. They are, however, important for one of the homework problems.

Definition 1.29 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We say that $(\Omega, \mathcal{F}, \mu)$ is *non-atomic* if for every $A \in \mathcal{F}$ with $\mu(A) > 0$ there exists $B \subseteq A$ in $\mathcal{F}$ such that $0 < \mu(B) < \mu(A)$.

Example 1.30

1. $(\mathbb{R}, \mathcal{B}, \lambda)$ is non-atomic.
2. Any finite probability space is atomic.

Theorem 1.31 *Let $(\Omega, \mathcal{F}, \mu)$ be non-atomic, and let $A \in \mathcal{F}$. Then for every $0 \leq \gamma \leq \mu(A)$ there exists $B \subseteq A$ in $\mathcal{F}$ such that $\mu(B) = \gamma$.*

1.4.1 Towards proving Theorem 1.28

We start with the following definition.

Definition 1.32 Let $(\Omega, \mathcal{F})$ be a measurable space. A *signed measure* Φ on $(\Omega, \mathcal{F})$ is a function $\Phi : \mathcal{F} \to \mathbb{R} \cup \{-\infty, +\infty\}$, such that:

1. $\Phi(\emptyset) = 0$.
2. If $(A_k)_{k=1}^{\infty}$ are disjoint measurable sets, then

$$\Phi\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \Phi(A_k),$$

   and the sum always converges (possibly to an infinite value).

Example 1.33 The difference of two measures, at least one of which finite, is a signed measure.

Definition 1.34 Let $(\Omega, \mathcal{F})$ be a measurable space, and let Φ be a signed measure on $(\Omega, \mathcal{F})$. Then we define

$$|\Phi|(A) := \sup\{|\Phi(B)| + |\Phi(A \setminus B)| : B \subseteq A\}.$$

Exercise Prove that $|\Phi|$ is a measure on $(\Omega, \mathcal{F})$, and that $|\Phi|(A) \geq |\Phi(A)|$ for every A.

Theorem 1.35 (Hahn's decomposition's theorem)
Let $(\Omega, \mathcal{F})$ be a measurable space, and let Φ be a signed measure on $(\Omega, \mathcal{F})$. Then there exist sets $A^+ \in \mathcal{F}$ and $A^- \in \mathcal{F}$ such that

1. $A^+ \cup A^- = \Omega$ *and* $A^+ \cap A^- = \emptyset$.
2. $\Phi(A) \geq 0$ *for every measurable* $A \subseteq A^+$. A^+ *is called* the positive set *of* Φ.
3. $\Phi(A) \leq 0$ *for every measurable* $A \subseteq A^-$. A^- *is called* the negative set *of* Φ.

A^+ and A^- are unique up to measure zero.

Proof: Let $S = \sup\{\Phi(A) : A \in \mathcal{F}\}$ and $I = \inf\{\Phi(A) : A \in \mathcal{F}\}$.
Exercise show that at most one of them is infinite.

Hint: show that otherwise the sum in definition 1.32 does not make sense.

29.4
Assume without loss of generality that $S < \infty$. then one can find a sequence of measurable sets $(A_k)_{k=1}^{\infty}$ such that for every k,

$$S - 2^{-k} \leq \Phi(A_k) \leq S.$$

Now let

$$A^+ := \limsup_{k \to \infty} A_k,$$

and let $A^- := \Omega \setminus A^+$.

Claim 1.36 $\Phi(A^+) = S$.

Now, if $A \subseteq A^+$ has negative measure, then $\Phi(A^+ \setminus A) > S$, in contradiction to the definition of S, and equivalently if $A \subseteq A^-$ has positive measure then $\Phi(A^+ \cap A) > S$, again in contradiction to the definition of S. □

Proof of Claim 1.36: For every k and every $A \subseteq A_k$ we have that $\Phi(A) \geq -2^{-k}$, and for every $A \subseteq A_k^c$ we have that $\Phi(A) \leq 2^k$. Therefore, for every k, j, we have

$$|\Phi|(A_k \setminus A_j) \leq 2^k + 2^j$$

and therefore

$$|\Phi|(A_k \triangle A_j) \leq 2(2^k + 2^j).$$

Now, let $B_k = \cup_{j=k}^{\infty} A_j$. Then

$$|\Phi|(B_k \setminus A_k) = \sum_{j=k+1}^{\infty} |\Phi|(A_j \setminus A_{j-1}) \leq \sum_{j=k+1}^{\infty} \left[2^{-j} + 2^{1-j}\right] \leq 2^{1-k}.$$

In particular, $\Phi(B_k) \geq S - 2^{2-k}$.
Now, B_k is a decreasing sequence, and

$$|\Phi|(B_k \setminus B_{k+1}) \leq 2^{3-k}.$$

Therefore,

$$|\Phi|(B_k \setminus A) = \sum_{j=k}^{\infty} |\Phi|(B_j \setminus B_{j-1}) \leq 2^{4-k}.$$

Therefore, for every k we have $\Phi(A) \geq \Phi(B_k) - |\Phi|(B_k \setminus A) \geq S - 2^{5-k}$, and therefore $\Phi(A) \geq S$. From the definition of S, we get now $\Phi(A) = S$.

□

Corollary 1.37 *Let A^+, A^- and B^+, B^- be two Hahn decompositions of the space $(\Omega, \mathcal{F}, \Phi)$. Then $|\Phi|(A^+ \triangle B^+) = |\Phi|(A^- \triangle B^-) = 0$.*

Proof: We show that $\Phi(A^+ \setminus B^+) = 0$. From symmetry this suffices. Let $C \subseteq A^+ \setminus B^+$. Then $\Phi(C) \geq 0$ because $C \subseteq A^+$, and $\Phi(C) \leq 0$ because $C \subseteq B^-$, and thus $\Phi(C) = 0$. From here we get that $|\Phi|(A^+ \setminus B^+) = 0$, as desired. □

We can now use the Hahn's theorem to prove Radon-Nikodym's theorem.

Proof of Theorem 1.28: We first assume that both μ and ν are finite. The extension to the σ-finite case is left as an (easy) exercise. For every $\alpha \geq 0$ rational, we define the signed measure $\Phi_\alpha := \alpha \cdot \mu - \nu$. Then, for every $\alpha \geq 0$ rational, we define A_α to be (a choice of) the positive set of Φ_α.

Claim 1.38 *Let $\alpha_1 < \alpha_2$. Then $\mu(A_{\alpha_1} \setminus A_{\alpha_2}) = \nu(A_{\alpha_1} \setminus A_{\alpha_2}) = 0$.*

Proof of Claim 1.38: Let $A = A_{\alpha_1} \setminus A_{\alpha_2}$. Then $\Phi_{\alpha_1}(A) \geq 0$ and $\Phi_{\alpha_2}(A) \leq 0$. However, by the definition of Φ_α, and since $\alpha_1 < \alpha_2$, we get $\Phi_{\alpha_2}(A) \geq \Phi_{\alpha_1}(A)$, and thus $\Phi_{\alpha_1}(A) = \Phi_{\alpha_2}(A) = 0$. Solving a linear equation we get $\mu(A) = \nu(A) = 0$. □

We now define the function $f : \Omega \to \mathbb{R}$ by

$$f(\omega) := \inf(\alpha : \omega \in A_\alpha) \leq \infty.$$

Exercise show that f is measurable, and that due to absolute continuity, $\mu(\{\omega : f(\omega) = \infty\}) = \nu(\{\omega : f(\omega) = \infty\}) = 0$.
We now need to show that (1.4) holds for every $A \in \mathcal{F}$. We first show that

$$\nu(A) \leq \int_A f d\mu.$$

To this end, let $\epsilon > 0$ be rational, and for every $n = 0, 1, \ldots$ let $A^{(n)} := \{\omega \in A : n\epsilon \leq f(\omega) < (n+1)\epsilon\}$. Define $\bar{f}(\omega) := (n+1)\epsilon$ on $A^{(n)}$. Then $\bar{f} < f + \epsilon$. Then, up to measure zero, $A_n \subseteq A_{(n+1)\epsilon}$. Therefore, $\nu(A_n) \leq (n+1)\epsilon\mu(A_n)$, and we get that

$$\nu(A) = \sum_{k=0}^{\infty} \nu(A_n) \leq \sum_{k=0}^{\infty} (n+1)\epsilon\mu(A_n) = \int_A \bar{f} d\mu \leq \int_A f d\mu + \epsilon\mu(A).$$

taking ϵ as small as we like proves the desired inequality. The opposite inequality follows similarly, and (1.4) is proved. □

Exercise Prove the uniqueness (up to μ-measure zero) of f.

Definition 1.39 Let X be a random variable, and let D_X be its distribution. We say that X *has a density* if $D_X \ll \lambda$. In this case, we say that the *density* of X is

$$f_X := \frac{dD_x}{d\lambda}.$$

2.5

2 Basic inequalities and types of convergence

2.1 Inequalities

We begin with the most basic and most useful inequalities in Probability Theory.

Theorem 2.1 (Cauchy-Schwarz inequality)
Let X, Y be variables with second moments. Then $E(XY)^2 \leq E(X^2)E(Y^2)$.

Proof: Assume without loss of generality that $E(X^2) = E(Y^2) = 1$. Define $Z = X - E(XY)Y$. Then $E(ZY) = E(XY) - E(XY)E(Y^2) = 0$. Therefore,

$$E(X^2)E(Y^2) = 1 = E(X^2) = E\big[(Z + E(XY)Y)^2\big]$$
$$= E(Z^2) + E(XY)^2E(Y^2) = E(XY)^2 + E(Z^2) \geq E(XY)^2.$$

□

Theorem 2.2 (Markov's inequality)
Let X be a non-negative random variable, and let $a > 0$. Then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof: Let Y be a random variable, defined as follows:

$$Y = \begin{cases} 0 & \text{if } X < a \\ a & \text{if } X \geq a \end{cases}.$$

Then $Y \leq X$, and therefore $E(Y) \leq E(X)$. We get

$$\frac{E(X)}{a} \geq \frac{E(Y)}{a} = \frac{a \cdot P(Y = a)}{a} = P(Y = a) = P(X \geq a).$$

□

We now prove a general inequality which we will use later.

Theorem 2.3 (Jensen's Inequality)
Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, and let X be a random variable s.t. X and $f(X)$ both have expectations. Then

$$E\left[f(X)\right] \geq f\left(E[X]\right).$$

Corollary 2.4 $\left[E[X]\right]^2 \leq E\left[X^2\right]$ *whenever well defined.*

Proof of Theorem 2.3: Since f is convex, for every x there exists an affine function g_x such that $g_x(x) = f(x)$ and $g_x(y) \leq f(y)$ for every y. We take $g = g_{E(X)}$. Since g is affine, we get $g\big(E[X]\big) = E\big[g(X)\big]$. Therefore,

$$E\left[f(X)\right] \geq E\left[g(X)\right] = g\left(E[X]\right) = f\left(E[X]\right).$$

□

We now prove two inequalities that relate to the second moment.

Theorem 2.5 (Chebishef I)
Let X be a random variable with variance. Then for every $\delta > 0$,

$$P\left[\big|X - E[x]\big| \geq \delta\right] \leq \frac{var[X]}{\delta^2}.$$

Proof: Let $Y = \big[X - E(X)\big]^2$. Then Y is non-negative, and $E(Y) = \text{var}(X)$. Then, by Markov's inequality,

$$P\left[\big|X - E[x]\big| \geq \delta\right] = P(Y \geq \delta^2) \leq \frac{E(Y)}{\delta^2} = \frac{\text{var}[X]}{\delta^2}.$$

□

Theorem 2.6 (Chebishef II)
Let X be a non-negative random variable with a second moment. Then

$$P[X > 0] \geq \frac{\left[E[X]\right]^2}{E\left[X^2\right]}.$$

Proof: Let $Y = \mathbf{1}_{X>0}$. Then $E[Y] = E\big[Y^2\big] = P[X > 0]$ and $X = XY$. Cauchy-Schwarz inequality tells us

$$\left[E[X]\right]^2 = \left[E[XY]\right]^2 \leq E\left[X^2\right] \cdot E\left[Y^2\right] = P[X > 0] \cdot E\left[X^2\right],$$

and we get the required inequality by dividing both sides by $E\big[X^2\big]$. □

2.2 The lemma of Borel-Cantelli

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $(A_k)_{k=1}^{\infty}$ be a sequence of events. The event $A := \limsup_{k\to\infty} A_k$ is the event that infinitely many of the events occur.

Theorem 2.7 (Borel-Cantelli's lemma) *1. If*

$$\sum_{k=1}^{\infty} P(A_k) < \infty,$$

then $P(A) = 0$.

2. *If the events are independent and*

$$\sum_{k=1}^{\infty} P(A_k) = \infty, \tag{2.1}$$

then $P(A) = 1$.

Exercise Find a sequence $(A_k)_{k=1}^{\infty}$ of events such that

$$\sum_{k=1}^{\infty} P(A_k) = \infty,$$

but $P(A) = 0$.

Proof of Theorem 2.7: 1: For every k,

$$A \subseteq \bigcap_{j=k}^{\infty} A_j,$$

and therefore

$$P(A) \le \sum_{j=k}^{\infty} P(A_k),$$

and as the RHS goes to zero as $k \to \infty$, we get that $P(A) = 0$.
2: Let $B_k = \cup_{j=k}^{\infty} A_j$. Then, using the continuity of probability (Theorem 1.21) as well as the continuity of the exponential function and the inequality $\log x \le x - 1$,

$$\begin{aligned} P\left[B_k^c\right] = P\left[\bigcap_{j=k}^{\infty} A_k^c\right] &= \prod_{j=k}^{\infty} P\left[A_k^c\right] = \exp\left(\sum_{j=k}^{\infty} \log P\left[A_k^c\right]\right) \\ &\le \exp\left(\sum_{j=k}^{\infty} \left(P\left[A_k^c\right] - 1\right)\right) = \exp\left(-\sum_{j=k}^{\infty} P\left[A_k\right]\right) = 0, \end{aligned}$$

where the last inequality follows from the divergence of the sum in (2.1). Therefore, $P\big[B_k\big] = 1$. Now, as $A = \cap_{k=1}^{\infty} B_k$, we get

$$P[A] = \lim_{k \to \infty} P[B_k] = 1.$$

□

6.5

2.3 Kolmogorov's zero-one law

In this section we show that certain events are trivial, namely that their probabilities are either zero or one. For the duration of this section we let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{A}_n)_{n=1}^{\infty}$ be a sequence of independent sub-σ-algebras of $\mathcal{F}$.
We let $\mathcal{F}_n = \sigma(\mathcal{A}_n, \mathcal{A}_{n+1}, \ldots)$ be the σ-algebra generated by $\mathcal{A}_k, k \ge n$ and let

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}_n.$$

$\mathcal{T}$ is called the *tail σ-algebra* of $(\mathcal{A}_n)$.

Theorem 2.8 *[Kolmogorov's zero-one law] For every event $A \in \mathcal{T}$, we have $P(A) \in \{0,1\}$.*

In order to prove Theorem 2.8, we need the following measure-theoretical lemma.

Lemma 2.9 *Let $A \in \mathcal{F}_1$. Then for every ϵ there exist n and $A_1 \in \sigma(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ such that $P(A \bigtriangleup A_1) < \epsilon$.*

Proof of Theorem 2.8: Let $A \in \mathcal{T}$. Then $A \in \mathcal{F}_1$. Fix ϵ. By Lemma 2.9, there exist n and $A_1 \in \sigma(\mathcal{A}_1, \ldots, \mathcal{A}_n)$ such that $P(A \bigtriangleup A_1) < \epsilon$. Note that $A \in \mathcal{F}_{n+1}$, and therefore A and A_1 are independent. Now,

$$P(A) - \epsilon \leq P(A \cap A_1) = P(A)P(A_1) \leq P^2(A) + \epsilon P(A) \leq P^2(A) + \epsilon$$

So $P(A) - P^2(A) < 2\epsilon$, and since this holds for all $\epsilon > 0$ we get that $P^2(A) \geq P(A)$ which implies $P(A) \in \{0,1\}$. □

Proof of Lemma 2.9: Let

$$\mathcal{G} = \{A \in \mathcal{F}_1 : \forall_\epsilon \exists_n \text{ and } _{A_1 \in \sigma(\mathcal{A}_1, \ldots, \mathcal{A}_n)} \; P(A \bigtriangleup A_1) < \epsilon\}$$

All we need to show is that $\mathcal{G}$ is a σ-algebra. If $A \in G$ then clearly $A^c \in \mathcal{G}$. if (A_n) are in $\mathcal{G}$ and $\epsilon > 0$, then there is m such that $P(\bigcap_{k=1}^{\infty} A_k) - P(\bigcap_{k=1}^{m} A_k) < \epsilon/2$, and for every $k \leq m$ we have n_k and B_k such that $B_k \in \mathcal{F}_{n_k}$ and $P(A_k \bigtriangleup B_k) < 2^{-k-1}\epsilon$. Take $n = \max\{n_1, \ldots, n_m\}$ and $B = B_1 \cup B_2 \cup \cdots \cup B_m$. Then $B \in \mathcal{F}_n$ and $P(A \bigtriangleup B) < \epsilon$. □

Example 2.10 Let $(X_n)_{n=1}^{\infty}$ be independent random variables. Then

$$P\left[\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k \text{ exists}\right] \in \{0,1\}.$$

2.4 Types of convergence

In this section we discuss various types of convergence of random variables, and the relations between those types of convergence.

2.4.1 Convergence in probability

Let $(X_k)_{k=1}^{\infty}$ be random variables, and let X be a random variable, all defined on the same probability space. We say that (X_k) *converges to X in probability*, and write

$$X_k \underset{\text{prob}}{\longrightarrow} X,$$

if for every $\delta > 0$,

$$\lim_{k\to\infty} P(|X_k - X| > \delta) = 0. \tag{2.2}$$

9.5

This is equivalent to

$$\forall_\epsilon \exists_N \forall_{n>N} P(|X_n - X| > \epsilon) < \epsilon. \tag{2.3}$$

Exercise Prove the equivalence of (2.2) and (2.3).

Example 2.11 (Weak version of the weak law of large numbers)
Let $(X_k)_{k=1}^\infty$ be i.i.d. variables with second moment, and let $E = E(X_1)$. Let $S_n = \frac{1}{n}\sum_{k=1}^n X_k$. Then

$$S_n \underset{\text{prob}}{\longrightarrow} E,$$

Proof: from Chebisheff.

2.4.2 Almost sure convergence

Let $(X_n)_{n=0}^\infty$ be random variables, and let X be a random variable. Then we say that (X_k) *converges to* X *almost surely*, and write

$$X_k \underset{\text{a.s.}}{\longrightarrow} X,$$

if

$$P\left(\left\{\omega \ : \ \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \tag{2.4}$$

This is equivalent to

$$\forall_\epsilon \exists_N P\left(\exists_{n>N}|X_n - X| > \epsilon\right) < \epsilon. \tag{2.5}$$

Exercise Prove the equivalence of (2.4) and (2.5).
Solution First, assume that (X_n) converges to X. Then for every ϵ, with probability 1 there exists N such that $|X_n - X| < \epsilon$ for every $n > N$, i.e.

$$P\left[\bigcap_{N=1}^{\infty} \exists_{n>N}|X_n - X| > \epsilon\right] = 0.$$

Continuity of probability now yields (2.5).
Now, assume that (2.5) holds. For every rational $\epsilon > 0$, and every $0 < \delta < \epsilon$, there exists N_δ such that $P\left(\exists_{n>N_\delta}|X_n - X| > \epsilon\right) < \delta$, and by continuity of probability,

$$P\left[\exists_N \forall_{n>N}|X_n - X| < \epsilon\right] = 1.$$

Since there are only $\aleph_0$ rationals, we get

$$P\left[\forall_{0<\epsilon\in\mathbb{Q}} \exists_N \forall_{n>N}|X_n - X| < \epsilon\right] = 1.$$

a.s. convergence follows.

2.4.3 Convergence in L^p

Let $1 \leq p < \infty$. Let $(X_n)_{n=0}^{\infty}$ be random variables, and let X be a random variable. Then we say that (X_k) *converges to X in L^p*, and write

$$X_k \underset{L^p}{\longrightarrow} X,$$

If

$$\lim_{n\to\infty} E\left[|X_n - X|^p\right] = 0.$$

Exercise Show that if $p > q$, then convergence in L^p yields convergence in L^q.

Remark 2.12 We will mostly be interested in convergence in L^1 and in L^2.

Remark 2.13 In some books it is required that the variables (X_n) are in L^p, and in others not. We do not make such a requirement, but in almost all interesting cases (X_n) are indeed in L^p.

2.4.4 Convergence in distribution

Let $(X_n)_{n=0}^{\infty}$ be random variables, and let X be a random variable. Then we say that (X_k) *converges to X in distribution*, and write

$$X_k \underset{dist}{\longrightarrow} X,$$

If for every continuous, bounded f with compact support,

$$\lim_{n\to\infty} E[f(X_n)] = E[f(X)].$$

exercise Prove that this is equivalent to the following condition: For ever x such that F_X is continuous in x,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

2.5 Relations between various types of convergence

In this section we will draw a complete diagram of relations between types of convergence.

Theorem 2.14 *Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables, and let X be a random variable.*

1. *If*

$$X_k \underset{a.s.}{\longrightarrow} X,$$

then

$$X_k \underset{prob}{\longrightarrow} X.$$

2. *If*

$$X_n \xrightarrow[prob]{} X,$$

then there exists a strictly increasing sequence $(n_k)_{k=1}^{\infty}$ *such that*

$$X_{n_k} \xrightarrow[a.s.]{} X.$$

Proof: 1. If a sequence (X_n) converges almost surely to a random variable X, then by (2.5), for every ϵ there exist N such that w.p. $1-\epsilon$ for every $n > N$ we have $|X - X_n| < \epsilon$. In particular, $P[|X_n - X| > \epsilon] < \epsilon$ for every $n > N$, so (2.3) is satisfied.

13.5

2. We assume that the sequence (X_n) converges in probability to a random variable X. Then by (2.3) we can define a subsequence as follows: n_1 is chosen so that $P[|X_{n_1} - X| > 2^{-1}] < 2^{-1}$. Then inductively n_k is chosen to be a number larger than n_{k-1} such that $P[|X_{n_k} - X| > 2^{-k}] < 2^{-k}$. such choice is possible by (2.3). Now choose $\epsilon > 0$. Then there exists K such that $\epsilon > 2^{-K}$. Then

$$P\left[\exists_{k>K}|X_{n_k} - X| > \epsilon\right] \leq \sum_{k=K+1}^{\infty} P\left[|X_{n_k} - X| > \epsilon\right]$$

$$\leq \sum_{k=K+1}^{\infty} P\left[|X_{n_k} - X| > 2^{-k}\right] \leq \sum_{k=K+1}^{\infty} 2^{-k} = 2^{-K} < \epsilon.$$

Therefore, by (2.5), the convergence is a.s. □

Theorem 2.15 *If* (X_n) *converges to* X *in* L^p*, then* (X_n) *converges to* X *in probability.*

Proof: Exercise, use Markov's inequality. □

Exercise Find examples showing that there is no implication between a.s. convergence and convergence in L^p.

Theorem 2.16 1. *If* (X_n) *converges to* X *in probability, then* (X_n) *converges to* X *in distribution.*

2. *If* (X_n) *converges to* X *in distribution and* X *is an almost sure constant, then* (X_n) *converges to* X *in probability.*

Proof: 1. Let f be a bounded, continuous function with bounded support. Then f is uniformly continuous. Let M be a bound for $|f|$. Fix ϵ, and let $\delta < \epsilon$ be such that if $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$. Let N be such that for all $n > N$ we have

$$P[|X_n - X| > \delta] < \delta.$$

Then for $n > N$, let A be the event $A = \{|X - X_n| > \delta\}$

$$E(f(X)) - E(f(X_n)) = \big(E[f(X)\mathbf{1}_A] - E[f(X_n)\mathbf{1}_A]\big) + \big(E[f(X)\mathbf{1}_{A^c}] - E[f(X_n)\mathbf{1}_{A^c}]\big).$$

Now,

$$\big(E[f(X)\mathbf{1}_{A^c}] - E[f(X_n)\mathbf{1}_{A^c}]\big) \leq 2MP(A^c) < 2M\epsilon,$$

and

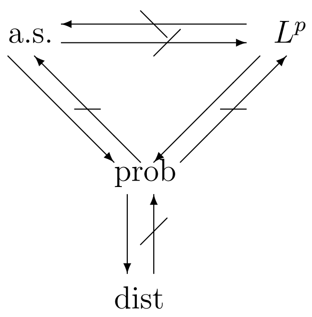
$$\big(E[f(X)\mathbf{1}_{A}] - E[f(X_n)\mathbf{1}_{A}]\big) \leq \epsilon.$$

Therefore

$$|E(f(X)) - E(f(X_n))| < (2M+1)\epsilon.$$

2. Exercise. □

So we get the following diagram.



16.5

3 The law of large numbers

In this section we discuss the following theorem.

Theorem 3.1 *Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables, and assume $E(|X_1|) < \infty$. Let $E := E(X_1)$. Let*

$$S_n = \sum_{k=1}^{n} X_k.$$

Then

$$\frac{S_n}{n} \underset{a.s.}{\longrightarrow} E.$$

Conversely, we have the following theorem.

Theorem 3.2 *Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables, and let*

$$S_n = \sum_{k=1}^{n} X_k.$$

Assume in addition that there exists a random variable X s.t.

$$\frac{S_n}{n} \underset{a.s.}{\longrightarrow} X.$$

Then $E(|X_1|) < \infty$, and $X \equiv E(X_1)$ (i.e. $P(X = E(X_1)) = 1$).

We will prove Theorem 3.1 in full when we study martingales (this is in the future...), but at the moment we prove an important special case.

Proof od Theorem 3.1 in the case $\mathrm{var}(X_1) < \infty$: We assume that $\mathrm{var}(X_1) < \infty$. Write $\sigma^2 = \mathrm{var}(X_1)$. First we prove that

$$\frac{S_{n^2}}{n^2} \underset{\text{a.s.}}{\longrightarrow} E.$$

Indeed,

$$\mathrm{var}(S_{n^2}) = \sum_{k=1}^{n^2} \mathrm{var}(X_k) = n^2\sigma^2,$$

and thus

$$\mathrm{var}\left[\frac{S_{n^2}}{n^2}\right] = \frac{1}{n^4}\mathrm{var}(S_{n^2}) = \frac{\sigma^2}{n^2}.$$

At the same time,

$$E\left[\frac{S_{n^2}}{n^2}\right] = E.$$

Fix $\epsilon > 0$ rational. By Chebisheff's inequality,

$$P\left[\left|\frac{S_{n^2}}{n^2} - E\right| \geq \epsilon\right] \leq \frac{\mathrm{var}[S_{n^2}/n^2]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n^2},$$

and thus by Borel-Cantelli with probability 1 there exists $N_\epsilon < \infty$ such that for every $n > N_\epsilon$,

$$\left|\frac{S_{n^2}}{n^2} - E\right| < \epsilon \tag{3.1}$$

Therefore w.p. 1 (3.1) holds for every rational $\epsilon > 0$, and therefore

$$\frac{S_{n^2}}{n^2} \xrightarrow[\text{a.s.}]{} E.$$

We now need to show that the non-squares do not do too much damage. Again, fix $\epsilon > 0$ (rational). For every n, let

$$U_n = \max_{n^2 < k \leq (n+1)^2} \left|\frac{S_k}{k} - \frac{S_{n^2}}{n^2}\right|.$$

We want to estimate $P(U_n \geq \epsilon)$.
We first estimate

$$P\left[\left|\frac{S_k}{k} - \frac{S_{n^2}}{n^2}\right| > \epsilon\right]$$

for a given $n^2 < k \leq (n+1)^2$. First we write

$$J_k := \frac{S_k}{k} - \frac{S_{n^2}}{n^2} = \frac{1}{k}\sum_{j=n^2}^{k} X_j - \left(\frac{1}{n^2} - \frac{1}{k}\right)\sum_{j=1}^{n^2} X_j$$

Thus

$$E(J_k) = \left(\frac{k-n^2}{k} - n^2\left[\frac{1}{n^2} - \frac{1}{k}\right]\right) \cdot E = 0.$$

We also need to calculate the variance.

$$\mathrm{var}(J_k) = \frac{k-n^2}{k^2}\sigma^2 + \left(\frac{k-n^2}{kn^2}\right)^2 \cdot n^2\sigma^2 \leq \left(\frac{3n}{n^4} + \frac{9n^2}{n^6}\right) \cdot \sigma^2 \leq \frac{4\sigma^2}{n^3}$$

for all n large enough and $n^2 < k \leq (n+1)^2$.
Thus, by Chebisheff,

$$P[|J_k| > \epsilon] \leq P[|J_k - E(J_k)| > \epsilon/2] \leq \frac{4\mathrm{var} J_k}{\epsilon^2} \leq \frac{C_1}{n^3},$$

and by a union bound,

$$P(U_n \geq \epsilon) \leq \left[(n+1)^2 - n\right] \cdot \frac{C_1}{n^3} \leq \frac{C_2}{n^2},$$

and by Borel-Cantelli w.p.1 $U_n < \epsilon$ for all n large enough.
The theorem follows.

□

Proof of Theorem 3.2: It is sufficient to prove that $E(|X_1|) < \infty$ due to Theorem 3.1. For contradiction, assume $E(|X_1|) = \infty$, and we keep the convergence assumption. Then

$$\sum_{k=1}^{\infty} P(|X_1| \geq k) = (-1) + \sum_{k=0}^{\infty} P(|X_1| \geq k) \geq (-1) + E(|X_1|) = \infty.$$

Therefore

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{k=1}^{\infty} P(|X_1| \geq k) = \infty,$$

and by Borel-Cantelli there a.s. exists a subsequence $(X_{n_k})_{k=1}^{\infty}$ such that $|X_{n_k}| \geq n_k$. For such n_k,

$$\frac{S_{n_k}}{n_k} - \frac{S_{n_k-1}}{n_k-1} = \frac{X_{n_k}}{n_k} + \left(\frac{n_k-1}{n_k} - 1\right)\frac{S_{n_k-1}}{n_k-1} = \frac{X_{n_k}}{n_k} - \frac{1}{n_k}\frac{S_{n_k-1}}{n_k-1}.$$

Due to Cauchy's criterion,

$$\lim_{k\to\infty}\left(\frac{X_{n_k}}{n_k} - \frac{1}{n_k}\frac{S_{n_k-1}}{n_k-1}\right) = 0. \tag{3.2}$$

On the other hand, due to convergence,

$$\lim_{k\to\infty}\frac{1}{n_k}\frac{S_{n_k-1}}{n_k-1} = 0, \tag{3.3}$$

and due to the choice of the sequence (n_k),

$$\liminf_{k\to\infty}\left|\frac{X_{n_k}}{n_k}\right| \geq 1. \tag{3.4}$$

(3.3) and (3.4) contradict (3.2).

$\square$

Exercise: Is the following true: Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables, and let

$$S_n = \sum_{k=1}^{n} X_k.$$

Assume in addition that there exists a random variable X s.t.

$$\frac{S_n}{n} \underset{\text{prob}}{\longrightarrow} X.$$

Then $E(|X_1|) < \infty$, and $X \equiv E(X_1)$ (i.e. $P(X = E(X_1)) = 1$)?

20.5

4 Characteristic functions and the central limit theorem

4.1 Characteristic functions

Let X be a random variable, and let $t \in \mathbb{R}$. Then the (complex) random variable e^{itX} is bounded by 1 and therefore has an expectation. Therefore, for every random variable X, we can define its *characteristic function* $\phi_X : \mathbb{R} \to \mathbb{C}$ by $\phi_X(t) = E[e^{itX}]$.
We start by discussing the basic properties of the characteristic function.

Lemma 4.1 *Let X be a random variable, and let $\phi_X : \mathbb{R} \to \mathbb{C}$ be its characteristic function.*

1. *ϕ_X is determined only by the distribution of X.*
2. *$\phi_X(0) = 1$.*
3. *$|\phi_X(t)| \le 1$ for every $t \in \mathbb{R}$.*
4. *ϕ_X is continuous in $\mathbb{R}$.*
5. *If X has an expectation, then ϕ_X is everywhere differentiable, and $\phi'_X(0) = iE(X)$. Further, $|\phi'_X(t)| \le E(|X|)$ for every t.*
6. *If X has a variance, then ϕ_X is everywhere twice differentiable, and $\phi''_X(0) = -E(X^2)$. Further, $|\phi''_X(t)| \le E(X^2)$ for every t.*
7. *If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$ for every t.*

Exercise: Show that in parts 5 and 6, the function is, in fact, continuously (twice) differentiable.

Proof of Lemma 4.1: 1. and 2. are obvious. 3. This is clear because e^{itX} is bounded by 1.
4. Fix ϵ and let M be such that $P[|X| \ge M] < \epsilon/2$. Let t_1, t_1 be such that $|t_1 - t_2| < \frac{\epsilon}{4M}$. Then,

$$\begin{aligned} |\phi_X(t_1) - \phi_X(t_1)| &\le \left|E\left[\left(e^{it_1X} - e^{it_2X}\right) \cdot \mathbf{1}_{X<M}\right]\right| + \left|E\left[\left(e^{it_1X} - e^{it_2X}\right) \cdot \mathbf{1}_{X\ge M}\right]\right| \\ &\le M|t_1 - t_2| + 2P[|X| > M] < \epsilon. \end{aligned}$$

5. For every t and every $h > 0$,

$$\left|\frac{e^{i(t+h)X} - e^{itX}}{h}\right| \le |X|.$$

As $|X|$ is integrable, by the dominated convergence theorem,

$$\begin{aligned} E[Xie^{itX}] &= E\left[\lim_{h\to 0}\frac{e^{i(t+h)X}-e^{itX}}{h}\right] = \lim_{h\to 0} E\left[\frac{e^{i(t+h)X}-e^{itX}}{h}\right] \\ &= \lim_{h\to 0}\frac{1}{h}\left(E\left[e^{i(t+h)X}\right]-E\left[e^{itX}\right]\right) = \frac{d\phi_X(t)}{dt}. \end{aligned}$$

23.5 Jan Nagel

6. Same proof as that of 5.

7. From independence,

$$\phi_{X+Y}(t) = E[e^{it(X+Y)}] = E[e^{itX}e^{itY}] = E[e^{itX}]E[e^{itY}] = \phi_X(t)\phi_Y(t).$$

□

Example 4.2 Let $X \sim \mathcal{N}(0,1)$. We will calculate $\phi_X(t)$ for every $t \in \mathbb{R}$. Let f be the density function of X, i.e.

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Then,

$$\begin{aligned} \phi_X(t) = E[e^{itX}] &= \int_{-\infty}^{\infty} e^{itx} f(x)dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(itx - x^2/2)dx \\ = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(\frac{1}{2}(t^2+2itx-x^2-t^2)\right)dx &= \frac{e^{-t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(\frac{-1}{2}(x^2-2itx-t^2)\right)dx \\ = \frac{e^{-t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(\frac{(x-it)^2}{2}\right)dx = e^{-t^2/2}, \end{aligned}$$

where the last inequality (i.e. the calculation of the integral) follows from Cauchy's theorem.

Exercise: Calculate the characteristic functions of Bernoulli, exponential and binomial distributions.

4.2 The central limit theorem

Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables, and assume $E(X_1) = \mu$, $\text{var}(X_1) = \sigma^2 < \infty$. We want to understand the behavior of the variable

$$S_n = \sum_{k=1}^{n} X_k$$

for n large. In particular, we want to establish some sort of convergence to a non trivial variable. $E(S_n) = n\mu$, so in order to get convergence we need to subtract $n\mu$. Once done this, the variance of $S_n - n\mu$ is $n\sigma^2$, so we may want to divide by $\sigma\sqrt{n} = \sqrt{n\sigma^2}$.
Our purpose now is to prove the following theorem.

Theorem 4.3 (Central limit theorem)
Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables, and assume $E(X_1) = \mu$, $var(X_1) = \sigma^2 < \infty$. Then,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \underset{dist}{\longrightarrow} \mathcal{N}(0,1).$$

Before proving the theorem, we may note that w.l.o.g. we may assume $\mu = 0, \sigma^2 = 1$ and then we get the more esthetic form

$$\frac{S_n}{\sqrt{n}} \underset{\text{dist}}{\longrightarrow} \mathcal{N}(0,1).$$

We start with a lemma that gives the intuitive explanation the CLT.

Lemma 4.4 *Let $(X_n)_{n=1}^{\infty}$ be i.i.d. random variables such that $E(X_1) = 0$ and $var(X_1) = 1$. Let $X \sim \mathcal{N}(0,1)$. Let*

$$S_n = \sum_{k=1}^{n} X_k,$$

and let $U_n = S_n/\sqrt{n}$. Then for every $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \phi_{U_n}(t) = \phi_X(t).$$

Proof: Fix t. Then $\phi_{S_n}(t) = [\phi_{X_1}(t)]^n$, and

$$\phi_{U_n}(t) = E[e^{iU_nt}] = E[e^{iS_nt/\sqrt{n}}] = \phi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \left[\phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

$\phi_{X_1}(0) = 1$ and ϕ_{X_1} is continuous, and therefore for all n large enough,

$$\log\left[\phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right]$$

is well defined. Thus, for n large enough,

$$\log(\phi_{U_n}(t)) = n \log \phi_{X_1}(t/\sqrt{n}).$$

Note that by Taylor's theorem,

$$\log \phi_{X_1}(t/\sqrt{n}) = \log \phi_{X_1}(0) + (\log \phi_{X_1})'(0) \cdot \frac{t}{\sqrt{n}} + (\log \phi_{X_1})''(0) \cdot \frac{t^2}{2n} + R_n,$$

where $\lim_{n\to\infty} nR_n = 0$. Also, note that $\log \phi_{X_1}(0) = 0$, $(\log \phi_{X_1})'(0) = 0, (\log \phi_{X_1})''(0) = -1$. Therefore,

$$\lim_{n\to\infty} \log(\phi_{U_n}(t)) = \lim_{n\to\infty} n \log \phi_{X_1}(t/\sqrt{n}) = \lim_{n\to\infty} n \cdot \frac{-t^2}{2n} + nR_n = \frac{-t^2}{2},$$

and therefore

$$\lim_{n\to\infty} \phi_{U_n}(t) = e^{\frac{-t^2}{2}}.$$

□

We now show that the convergence of the characteristic functions indeed guarantees convergence of the distributions. We do it in a few steps. The first step seems to be the converse of the desired statement.

Lemma 4.5 *If*

$$X_n \underset{dist}{\longrightarrow} X,$$

then for every $t \in \mathbb{R}$*,*

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t).$$

Proof: First we define for every $M > 0$ the cut-off function $g_M : \mathbb{R} \to \mathbb{R}$ by

$$g_M(x) = \begin{cases} 1 & |x| \le M \\ M + 1 - |x| & M \le |x| \le M + 1 \\ 0 & |x| \ge M + 1 \end{cases}.$$

Let $f : \mathbb{R} \to \mathbb{C}$ be defined by $f(x) = e^{its}$, and let $f_M(x) = f(x)g_M(x)$. Then for every M,

$$\lim_{n\to\infty} E(f_M(X_n)) = E(f_M(X)).$$

For every ϵ, there exists M such that $P(|X| > M) < \epsilon$, and $P(|X| > M) < \epsilon$ for every n. For this M and each n, we have that $\big|E[f_M(X_n) - E[f(X_n)]]\big| < \epsilon$, and equivalently, $\big|E[f_M(X) - E[f(X)]]\big| < \epsilon$. The lemma follows. □

27.5

The next lemma shows that the characteristic function determines the distribution.

Lemma 4.6 *Assume* X_1 *and* X_2 *are two random variables with the same characteristic function* ϕ*. Then* X_1 *and* X_2 *have the same distribution.*

Proof: Let P_1 be the distribution of X_1, and P_2 that of X_2. It suffices to show that for every interval $[a, b]$, where neither a nor b are atoms of P_1 or P_2, we have $P_1([a, b]) = P_2([a, b])$. By translation and multiplication by a constant, we may assume that the interval $[a, b]$ is in fact the interval $[-1, 1]$.
We now claim:

$$P_1([-1,1]) = P_2([-1,1]) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{it} - e^{-it}}{it} \phi(t)dt.$$

Indeed, let

$$I(T) = \int_{-T}^{T} \frac{e^{it} - e^{-it}}{it} \phi(t)dt = \int_{-T}^{T} \int \frac{e^{it} - e^{-it}}{it} e^{itx} dP(x)dt.$$

Note that the integrand is bounded by 2, and both measures are finite, and therefore we may use Fubini and say

$$I(T) = \int_{-T}^{T} \int \frac{e^{it} - e^{-it}}{it} e^{itx} dP(x) dt = \int \int_{-T}^{T} \frac{e^{it} - e^{-it}}{it} e^{itx} dt dP(x)$$
$$= \int \left[\int_{-T}^{T} \frac{\sin((x+1)t)}{t} dt - \int_{-T}^{T} \frac{\sin((x-1)t)}{t} dt \right] dP(x).$$

We introducing the notation $R(T, \theta) = \int_{-T}^{T} \frac{\sin(\theta t)}{t} dt$ and $S(T) = \int_0^T \frac{\sin t}{t} dt$, we get

$$I(T) = \int \left(R(T, x+1) - R(T, x-1)\right) dP(x). \tag{4.1}$$

Changing the integration variables tells us that

$$R(T, \theta) = 2(\operatorname{sgn}\theta) S(T|\theta|).$$

It is well known that $\lim_{T\to\infty} S(T) = \pi/2$ (and even if not, it only changes the constant in front of the integral), and therefore, for every x,

$$\lim_{T\to\infty} R(T, x+1) - R(T, x-1) = \begin{cases} 0 & |x| > 1 \\ \pi & |x| = 1 \\ 2\pi & |x| < 1 \end{cases}.$$

$S(T)$ is bounded, and therefore R is also bounded, and therefore by applying the bounded convergence theorem we get the desired result.

□

We now define the notion of tightness for random variables. Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. We say that (X_n) is *tight* if for every ϵ there exists M such that for every n,

$$P(|X_n| > M) < \epsilon.$$

30.5

Lemma 4.7 *Let $(X_n)_{n=1}^{\infty}$ be a tight sequence of random variables. Then there exist a subsequence $(X_{n_k})_{k=1}^{\infty}$ which converges in distribution.*

To prove Lemma 4.7, we need to use the Riesz representation theorem.

Theorem 4.8 (Riesz representation)
Let C be the space of continuous functions with bounded support from $\mathbb{R}$ to $\mathbb{R}$. Then every bounded linear functional $\psi : C \to \mathbb{R}$ can be represented as the integral with respect to a signed measure.

Proof of Lemma 4.7: Let F be a countable dense collection of functions in C. Let $f_1, f_2, \ldots$ be an ordering of F. We take a subsequence $(X_n^{(1)})$ of (X_n) such that $E(f_1(X_n^{(1)}))$ converges, then a subsequence $(X_n^{(2)})$ of $(X_n^{(1)})$ such that $E(f_2(X_n^{(2)}))$ converges, and so on. At the end we take the diagonal subsequence $X_{n_k} = X_k^{(k)}$. Let $\phi : C \to \mathbb{R}$ be the functional defined by $\phi(f) = \lim_{k\to\infty} E(X_{n_k}(f))$ on F, and elsewhere by continuity. ϕ is bounded, and therefore is the integral w.r.t. a signed measure μ. ϕ is non-negative, and therefore μ is non-negative. All we ned to prove is that μ has total measure one. to this end, we use tightness. Fix ϵ, and let M be such that $P(|X_n| > M) < \epsilon$ for every n. Let

$$f = \begin{cases} 1 & |x| \leq M \\ 0 & |x| \geq M+1 \\ M+1-|x| & M \leq |x| \leq M+1 \end{cases}.$$

Then $\mu(f) \geq 1-\epsilon$, and thus $\mu(\mathbb{R}) \geq 1-\epsilon$, and thus $\mu(\mathbb{R}) \geq 1$. To see that $\mu(\mathbb{R}) \leq 1$ is an easy exercise. Therefore, X_{n_k} converge in distribution to any variable with distribution μ. □

Lemma 4.9 *Let (X_n) be random variables, and assume that ϕ_{X_n} converges pointwise to ϕ, where $\phi(0) = 1$ and ϕ is continuous at zero. Then the sequence X_n is tight.*

Proof: Fix ϵ. There exists δ s.t. for every n,

$$\frac{1}{2\delta}\int_{-\delta}^{\delta} \phi_{X_n}(t)dt > 1 - \epsilon/2.$$

Now fix (large) M.

$$\frac{1}{2\delta}\int_{-\delta}^{\delta} \phi_{X_n}(t)dt \leq P(|X_n| < M) + \frac{1}{2\delta}\max_{m>M}\left|\int_{-\delta}^{\delta} e^{imt}dt\right|$$
$$\leq P(|X_n| < M) + \frac{1}{2\delta}\max_{m>M}\frac{e^{im\delta} - e^{-im\delta}}{m} = P(|X_n| < M) + \frac{1}{\delta M}.$$

Therefore, for M such that $1/\delta M < \epsilon/2$, we get that $P(|X_n| > M) < \epsilon$.

□

3.6

Proof of Theorem 4.3: By Lemma 4.4, $\phi_{X_n}(t) \longrightarrow e^{-t^2/2}$ which is a continuous function. Therefore, by Lemma 4.9, the sequence (X_n) is tight, and therefore by Lemma 4.7 every subsequence of (X_n) has a convergent subsubsequence. By Lemma 4.5 the characteristic function of the limit of the subsubsequence is $e^{-t^2/2}$, and thus by Lemma 4.6 the limit is $\mathcal{N}(0,1)$. The theorem follows. □

5 Conditional expectation and martingales

5.1 Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X be a random variable on it. The expectation of the random variable X is its average value. We may define its average conditioned on an event A. This is done by integrating with respect to the probability measure conditioned on A. If we now partition Ω into events $A_1, \ldots, A_k$ (partition means that the A_i-s are disjoint and their union in Ω), we can calculate the expectation conditioned on each of those events, and define a new variable Y as follows: If A_i occurs, then Y takes the value $E(X|A_i)$.
It is easy to see that the variable Y is the unique variable satisfying the following properties:

1. Y is measurable w.r.t. the σ-algebra $\mathcal{G} := \sigma(A_1, \ldots, A_k)$.

2. For every (bounded) random variable Z which is measurable w.r.t. $\mathcal{G}$, we have $E(ZY) = E(ZX)$.

This formulation allows us to generalize the definition of conditional expectation to more general σ-algebras.

Definition 5.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space, let X be a random variable which has an expectation, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω. We say that a random variable Y is the *conditional expectation of* X *w.r.t.* $\mathcal{G}$, and denote $Y = E(X|G)$ if:

1. Y is measurable w.r.t. $\mathcal{G}$.

2. For every bounded random variable Z which is measurable w.r.t. $\mathcal{G}$, we have $E(ZY) = E(ZX)$.

It is not immediately obvious that the conditional expectation exists, or that it is unique. We will give an example where this is a priori not obvious, and then prove both existence and uniqueness of the conditional expectation.

Example 5.2 Let $\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}$, $P = \lambda$. We take $X(\omega) = \sin \omega$. We then let Y_k be the $k - th$ digit in the decimal expansion of ω, and take $\mathcal{G} = \sigma(Y_2, Y_4, Y_6, \ldots)$.

Lemma 5.3 (Existence)
The conditional expectation always exists.

Proof: Let Q be the following measure defined on $(\Omega, \mathcal{F})$:

$$Q(A) = \int_A X dP.$$

Note that P and Q are also defined on $(\Omega, \mathcal{G})$ because $\mathcal{G} \subseteq \mathcal{F}$. Now note that $Q \ll P$. Indeed, if $P(A) = 0$ then $Q(A) = 0$. Therefore there exists a Radon-Nykodim derivative

of Q w.r.t. P on the measurable space $(\Omega, \mathcal{G})$. We call this derivative Y, and claim that Y is the conditional expectation. We verify the conditions in the definition:
1. This is obvious as Y was defined on the space $(\Omega, \mathcal{G})$.
2. Z is measurable in the space $(\Omega, \mathcal{G})$. Therefore,

$$\int ZdQ = \int ZYdP = E(ZY).$$

Z is also measurable in the space $(\Omega, \mathcal{F})$, and thus

$$\int ZdQ = \int ZXdP = E(ZX).$$

and we get $E(ZY) = E(ZX)$. □

6.6

Lemma 5.4 (Uniqueness)
Assume that Y_1 is the conditional expectation of X w.r.t. $\mathcal{G}$, and that at the same time Y_2 is the conditional expectation of X w.r.t. $\mathcal{G}$. Then $P(Y_1 = Y_2) = 1$.

Proof: Assume $P(Y_1 \neq Y_2) > 0$. Without loss of generality, there exists $\epsilon > 0$ s.t. $P(Y_1 - Y_2 > \epsilon) > \epsilon$. Let $Z = \mathbf{1}_{Y_1 - Y_2 > \epsilon}$. Then Z is bounded and is measurable w.r.t. $\mathcal{G}$. Now, by the choice of Z, we have $E(ZY_1) - E(ZY_2) \geq \epsilon^2 > 0$, in contradiction to the assumption that $E(ZY_1) = E(ZX) = E(ZY_2)$. □

Claim 5.5 *Let $Y = E(X|\mathcal{G})$. Let Z be $\mathcal{G}$-measurable. If ZX has an expectation, then ZY has an expectation and $E(ZX) = E(ZY)$.*

Proof: First assume that X and Z are both non-negative. Fix M, and let $Z_M = Z \cdot \mathbf{1}_{Z<M}$. Z_M is bounded and $\mathcal{G}$-measurable, so $E(XZ_M) = E(YZ_M)$. The sequences $(XZ_M)_{M=1}^{\infty}$ and $(YZ_M)_{M=1}^{\infty}$ are monotone increasing and converge point wise to XZ and respectively YZ, so by the monotone convergence theorem,

$$E(XZ) = \lim_{M\to\infty} E(XZ_M) = \lim_{M\to\infty} E(YZ_M) = E(YZ).$$

If Z is nonnegative but X is not necessarily non-negative, let $Y_+ = E(X^+|\mathcal{G})$ and $Y_- = E(X^-|\mathcal{G})$. Note that $Y = Y_+ - Y_-$. Let $\bar{Y} = Y_+ - Y^+ = Y_- - Y^-$. Note that by Jensen's inequality $\bar{Y}$ is non-negative. By the assumptions, $E(ZX^+) = E(ZY_+) < \infty$ and $E(ZX^-) = E(ZY_-) < \infty$. Note that $\bar{Y} \leq Y_+$ and thus $0 \leq \bar{Y}Z \leq Y_+Z$ and thus $E(Z\bar{Y}) < \infty$. Therefore, $E(ZY^+) = E(Z\bar{Y}) + E(ZY_+) = E(Z\bar{Y}) + E(ZX^+)$ and equivalently $E(ZY^-) = E(Z\bar{Y}) + E(ZY_-) = E(Z\bar{Y}) + E(ZX^-)$. In particular $E(ZY^+), E(ZY^-) < \infty$ and

$$E(YZ) = E(ZY^+) - E(ZY^-) = E(ZX^+) - E(Z\bar{Y}) - E(ZX^-) + E(Z\bar{Y}) = E(ZX).$$

If both X and Z are not necessarily non-negative, note that

$$E\big((YZ)^+\big) = E(Y^+Z^+) + E(Y^-Z^-) \leq E(X^+Z^+) + E(X^-Z^-) \leq E(|XZ|) < \infty,$$

and similarly for $E\big((YZ)^-\big)$, so that ZY has an expectation. Note also that

$$E\big(|X|(Z^+)\big) \le E\big(|XZ|\big) < \infty, \tag{5.1}$$

and then

$$E(YZ) = E\big(Y(Z^+)\big) - E\big(Y(Z^-)\big) = E\big(X(Z^+)\big) - E\big(X(Z^-)\big) = E(XZ).$$

□

Notation: For a random variable Y, we define $E(X|Y) := E(X|\sigma(Y))$. for a (finite or infinite) collection of random variables $Y_1, Y_2, \ldots$, we write $E(X|Y_1, Y_2, \ldots) := E(X|\sigma(Y_1, Y_2, \ldots))$.
We now collect some useful facts regarding the conditional expectation.

Lemma 5.6 *1. Let $\mathcal{G} = \{\Omega, \emptyset\}$ be the trivial σ-algebra. Then $E(X|\mathcal{G}) = E(X)$ a.s.*

2. Let $\mathcal{G}$ be independent of X. Then $E(X|\mathcal{G}) = E(X)$ a.s.

3. Let $\mathcal{G}_2 \subseteq \mathcal{G}_1 \subseteq \mathcal{F}$. Then $E\big(E(X|\mathcal{G}_1)|\mathcal{G}_2\big) = E(X|\mathcal{G}_2)$.

4. Jensen's inequality holds for conditional expectations, namely if f is convex then $E(f(X)|\mathcal{G}) \ge f(E(X|\mathcal{G}))$ a.s. when everything is well-defined.

5. Define

$$var(X|\mathcal{G}) := E\left[\big(X - E(X|\mathcal{G})\big)^2 \Big| \mathcal{G}\right] = E(X^2|\mathcal{G}) - \left[E(X|\mathcal{G})\right]^2.$$

If $var(X) < \infty$, then $var(X) = E(var(X|\mathcal{G})) + var(E(X|\mathcal{G}))$.

6. If W is measurable w.r.t. $\mathcal{G}$ and $E(|XW|) < \infty$, then $E(WX|\mathcal{G}) = WE(X|\mathcal{G})$ a.s.

Exercise Find a variable X and σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ such that $E\big(E(X|\mathcal{G}_1)|\mathcal{G}_2\big) \neq E\big(E(X|\mathcal{G}_2)|\mathcal{G}_1\big)$.
Exercise Prove that $E(E(X|\mathcal{G})) = E(X)$.

Proof: 1. Let $Y = E(X|\mathcal{G})$. Y is a constant, and it has to satisfy $Y = E(Y \cdot 1) = E(X \cdot 1) = E(X)$.
2. Let $Y = E(X|\mathcal{G})$. Then Y and X are independent and, by parts (1) and (3), $E(X) = E(Y)$. Now, using Claim 5.5, $\mathrm{var}(Y) = E(Y^2) - E(Y)^2 = E(YX) - E(Y)^2 = E(Y)E(X) - E(Y)^2 = E(Y)^2 - E(Y)^2 = 0$, so Y is an almost sure constant, and as $E(Y) = E(X)$ we get $Y = E(X)$ a.s.
3. Let $Y_1 = E(X|\mathcal{G}_1)$ and $Y_2 = E(X|\mathcal{G}_2)$. Let Z be measurable w.r.t. $\mathcal{G}_2$. Note that Z is also measurable w.r.t. $\mathcal{G}_1$ because $\mathcal{G}_2 \subseteq \mathcal{G}_1$. Therefore,

$$E(ZY_2) = E(ZX) = E(ZY_1),$$

so $Y_2 = E(Y_1|\mathcal{G}_2)$.
4. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a Borel measurable function s.t. (a) for every x, the fiber $g(x, \cdot)$ is

a linear function, (b) $g(x,y) \le f(y)$ for every x and y, and (c) $g(x,x) = f(x)$ for every x.
Exercise: Prove the existence of such g.
Now, we may take $Y = g(E(X|\mathcal{G}), X)$. Then $Y \le f(X)$. Now, a.s,

$$f(E(X|\mathcal{G})) = E(Y|\mathcal{G}) \le E(f(X)|\mathcal{G}).$$

13.6
5. Assume w.l.o.g. that $E(X) = 0$. Let $Y = E(X|\mathcal{G})$ and $W = X - Y$. Then $E(XY) = E(Y^2)$ and therefore $E(WY) = 0$. Therefore, using Claim 5.5,

$$\begin{aligned} \mathrm{var}(x) = E(X^2) = E(Y^2) + E(W^2) &= \mathrm{var}(E(X|\mathcal{G})) + E\left[\left(X - E(X|\mathcal{G})\right)^2\right] \\ &= E(\mathrm{var}(X|\mathcal{G})) + \mathrm{var}(E(X|\mathcal{G})). \end{aligned}$$

6. Let $Y = E(X|\mathcal{G})$ and Z bounded and $\mathcal{G}$ measurable. Then, with the help of Claim 5.5, $E(ZWY) = E(ZWX)$ and therefore $E(WX|\mathcal{G}) = WE(X|\mathcal{G})$ a.s.

□

5.2 Filtrations

Definition 5.7 Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *filtration* is a sequence $(\mathcal{G}_n)_{n=1}^{\infty}$ of σ-algebras on Ω such that

1. For every n, $\mathcal{G}_{n+1}$ is a *refinement* of $\mathcal{G}_n$, i.e. $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$.
2. For every n, $\mathcal{G}_n \subseteq \mathcal{F}$.

We may also speak about finite filtrations, namely finite sequences of σ-algebras.
Examples:

1. Let $\Omega = [0,1]^{\aleph_0}$ with product σ-algebra and measure. We take $\mathcal{G}_n = \mathcal{B} \times \mathcal{B} \times \cdots \times \mathcal{B} \times \{\emptyset, [0,1]\} \times \{\emptyset, [0,1]\} \times \cdots$.
2. Let (X_n) be a sequence of random variables, then take $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$.

We can also define the notion of an inverse filtration which we will use later.

Definition 5.8 Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *inverse filtration* is a sequence $(\mathcal{G}_n)_{n=1}^{\infty}$ of σ-algebras on Ω such that

1. For every n, $\mathcal{G}_n$ is a *refinement* of $\mathcal{G}_{n+1}$, i.e. $\mathcal{G}_{n+1} \subseteq \mathcal{G}_n$.
2. For every n, $\mathcal{G}_n \subseteq \mathcal{F}$.

Example: Let (X_n) be a sequence of random variables, then take $\mathcal{G}_n = \sigma(X_n, X_{n+1}, \ldots)$.

5.3 Martingales

Definition 5.9 Let (X_n) be a sequence of random variables, and let $(\mathcal{G}_n)$ be a filtration. We say that (X_n) is a martingale with respect to $(\mathcal{G}_n)$ if

1. $E(|X_n|) < \infty$ for every n.
2. X_n is measurable with respect to $\mathcal{G}_n$ for every n.
3. $X_n = E(X_{n+1}|\mathcal{G}_n)$ for every n.

Definition 5.10 A sequence (X_n) of random variables is called a martingale if there exists a filtration $(\mathcal{G}_n)$ such that (X_n) is a martingale w.r.t. $(\mathcal{G}_n)$.

Examlpes

1. Simple random walk.
2. conditional expectations of same variable w.r.t. a filtration.

Equivalently, we can define the notion of an inverse martingale.

Definition 5.11 Let (X_n) be a sequence of random variables, and let $(\mathcal{G}_n)$ be an inverse filtration. We say that (X_n) is an inverse martingale with respect to $(\mathcal{G}_n)$ if

1. $E(|X_n|) < \infty$ for every n.
2. X_n is measurable with respect to $\mathcal{G}_n$ for every n.
3. $X_{n+1} = E(X_n|\mathcal{G}_{n+1})$ for every n.

Definition 5.12 A sequence (X_n) of random variables is called an inverse martingale if there exists an inverse filtration $(\mathcal{G}_n)$ such that (X_n) is an inverse martingale w.r.t. $(\mathcal{G}_n)$.

Examlpe Let $\{X_n\}$ be i.i.d. and assume that $E(|X_1|) < \infty$. Let

$$Z_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Then $\{Z_n\}$ is an inverse martingale.
17.6

Lemma 5.13 *1. A sequence (X_n) of random variables is a martingale if and only if it is a martingale with respect to its natural filtration $\mathcal{K}_n = \sigma(X_1, \ldots, X_n)$.*

2. Let (X_n) be a martingale w.r.t. a filtration $(\mathcal{G}_n)$. For every n, m, $E(X_n) = E(X_m)$. Also, if $m > n$ then $E(X_n) = E(X_m|\mathcal{G}_n)$.

3. If (X_n) and (Y_n) are both martingales w.r.t. $(\mathcal{G}_n)$, then (X_n+Y_n) is also a martingale w.r.t. $(\mathcal{G}_n)$.

Exercise:

1. State and prove an equivalent of Lemma 5.13 for inverse martingales.

2. Find two martingales whose sum is not a martingale.

3. Find an example of a martingale (X_n) w.r.t. a filtration $(\mathcal{G}_n)$ such that there is no variable X s.t. $X_n = E(X|\mathcal{G}_n)$ for all n.

Proof of Lemma 5.13: 1. Let (X_n) be a martingale w.r.t. $(\mathcal{G}_n)$. Clearly, $E(|X_n|) < \infty$ for every n, and it is also obvious that X_n is measurable with respect to $\mathcal{K}_n$. We need to show that $X_n = E(X_{n+1}|\mathcal{G}_n)$. Note that $\mathcal{K}_n \subseteq \mathcal{G}_n$, and therefore for every Z which is measurable w.r.t. $\mathcal{K}_n$, it is also measurable w.r.t. $\mathcal{G}_n$. Thus for all such Z, $E(ZX_{n+1}) = E(ZX_n)$ and $X_n = E(X_{n+1}|\mathcal{K}_n)$.
2. This follows from Lemma 5.6.
3. This follows from the linearity of the conditional expectation. □

20.6 Tal Orenshtein

Definition 5.14 Let (X_n) be a sequence of random variables, and let $(\mathcal{G}_n)$ be a filtration. We say that (X_n) is a sub-martingale (resp super-martingale) with respect to $(\mathcal{G}_n)$ if

1. $E(|X_n|) < \infty$ for every n.

2. X_n is measurable with respect to $\mathcal{G}_n$ for every n.

3. $X_n \le E(X_{n+1}|\mathcal{G}_n)$ (resp. $X_n \ge E(X_{n+1}|\mathcal{G}_n)$) for every n.

Examples

1. Random walk with a drift.

2. Let (X_n) be a martingale, and let f be convex. Then $(f(X_n))$ is a sub-martingale.

3. Let (X_n) be a martingale, and let f be concave. Then $(f(X_n))$ is a super-martingale.

4. Let (X_n) be a martingale. $(|X_n|)$, (X_n^+) and (X_n^-) are sub-martingales.

5.4 Stopping times and the Optional stopping theorem

Definition 5.15 Let $(\mathcal{G}_n)_{n=1}^{\infty}$ be a filtration. A positive integer variable $T \le \infty$ is called a *stopping time* if for every n, the event $\{T \le n\}$ belongs to $\mathcal{G}_n$.

Examples:

1. Let (X_n) be a simple random walk, and let $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Take $T = \inf\{n : X_n = 9\}$. Then T is a stopping time.

2. Let (X_n) be a simple random walk, and let $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Take $T = \sup\{n < 100 : X_n = 0\}$. Then T is not a stopping time.

3. Let (X_n) be a simple random walk, and let $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Take $T = \inf\big\{n < 100 : X_n = \max\{X_k : k = 1, \ldots, 99\}\big\}$. Then T is not a stopping time.

4. Let (X_n) be a simple random walk, and let $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$. Take $T = \inf\big\{n \geq 100 : X_n = \max\{X_k : k = 1, \ldots, 99\}\big\}$. Then T is a stopping time.

Exercise: Prove the assertions above.

We now discuss stopped martingales. We start with a useful lemma.

Lemma 5.16 *Let $(\mathcal{G}_n)$ be a filtration, let (X_n) be a martingale w.r.t. $(\mathcal{G}_n)$, and let T be a stopping time w.r.t. $(\mathcal{G}_n)$. Let*

$$Y_n = X_{n \wedge T} = \begin{cases} X_n & n \leq T \\ X_T & n \geq T \end{cases}.$$

The Y_n is a martingale.

Exercise Find an example of a martingale (X_n) and a random time T such that $X_{n\wedge T}$ is not a martingale.

<u>Proof of Lemma 5.16:</u>

$$E(Y_{n+1}|\mathcal{G}_n) = \mathbf{1}_{T\leq n} E(Y_{n+1}|\mathcal{G}_n) + \mathbf{1}_{T>n} E(Y_{n+1}|\mathcal{G}_n)$$

On the event $\{T \leq n\}$, we have $Y_{n+1} = X_T = Y_n$, and thus

$$\mathbf{1}_{T\leq n} E(Y_{n+1}|\mathcal{G}_n) = \mathbf{1}_{T\leq n} Y_n.$$

On the event $\{T > n\}$, we have $Y_n = X_n$ and $Y_{n+1} = X_{n+1}$, and thus

$$\mathbf{1}_{T>n} E(Y_{n+1}|\mathcal{G}_n) = \mathbf{1}_{T>n} Y_n.$$

The lemma follows. □

From here we get an important theorem known as the optional sampling theorem.

Theorem 5.17 (Optional Sampling Theorem)
Let $(\mathcal{G}_n)$ be a filtration, let (X_n) be a martingale w.r.t. $(\mathcal{G}_n)$, and let T be a stopping time w.r.t. $(\mathcal{G}_n)$. Then $E(X_T) = E(X_1)$ if at least one of the following conditions hold:

1. *T is bounded (i.e. there exists M s.t. $P(T < M) = 1$).*

2. *(X_n) is bounded (i.e. $\exists_M P(\forall_n |X_n| < M) = 1$) and T is finite.*

3. *$E\big[|X_{n+1} - X_n|\big|\mathcal{G}_n\big]$ is bounded and $E(T) < \infty$.*

Proof: 1. Since $X_T = Y_M$ and $X_1 = Y_1$, and (Y_n) is a martingale, we get $E(X_T) = E(Y_M) = E(Y_1) = E(X_1)$.
2. Let N be a large number, and let $S = \min(T, N)$. Then S is a bounded stopping time, so $E(X_S) = E(X_1)$. Now, $|X_S - X_T| < 2M$, and $P(X_S \neq X_T) = P(T > N)$. Therefore $\left|E(X_T) - E(X_1)\right| < 2MP(T > N)$ and as $\lim_{N\to\infty} P(T > N) = 0$ we get $E(X_T) = E(X_1)$.
3. Again, let $S = \min(T, N)$. Then $E(X_S) = E(X_1)$, and

$$E(|X_T - X_S|) \leq E\left(\sum_{k=N}^{T-1} |X_{k+1} - X_k|\right) = \sum_{k=n}^{\infty} E(|X_{k+1} - X_k| \mathbf{1}_{T>k})$$

$$= \sum_{k=n}^{\infty} E\left(\mathbf{1}_{T>k} E\left[|X_{k+1} - X_k| \middle| \mathcal{G}_k\right]\right) \leq M \sum_{k=N}^{\infty} P(T > k),$$

and as the last sum converges to 0 as $N \to \infty$ we get that $E(X_T) = E(X_1)$. □

We now see several applications of the Optional Sampling Theorem. Let $(Z_n)_{n=1}^{\infty}$ be i.i.d. with distribution $P(Z_1 = 1) = P(Z_1 = -1) = 1/2$, and let

$$X_n = \sum_{k=1}^{n} Z_k.$$

The sequence (X_n) is called Simple Random Walk (or SRW).

Example 5.18 Fix some a, b positive and integer. Let $T = \min(n : X_n \in \{-a, b\})$. Then

$$P(X_T = a) = \frac{b}{a+b}.$$

24.6

Example 5.19 Fix some a positive and integer. Let $T = \min(n : X_n \in \{-a, a\})$. Then

$$E(T) = a^2.$$

Example 5.20 Let $T = \min(n : X_n = 1)$. Then $E(T) = \infty$.

The same holds for sub and super martingales. We state the theorem for sub martingales, and the proof is left as an exercise.

Theorem 5.21 *Let $(\mathcal{G}_n)$ be a filtration, let (X_n) be a sub-martingale w.r.t. $(\mathcal{G}_n)$, and let T be a stopping time w.r.t. $(\mathcal{G}_n)$. Then $E(X_T) \geq E(X_1)$ if at least one of the following conditions hold:*

1. *T is bounded (i.e. there exists M s.t. $P(T < M) = 1$).*
2. *(X_n) is bounded (i.e. $\exists_M P(\forall_n |X_n| < M) = 1$) and T is finite.*
3. *$E\left[|X_{n+1} - X_n| \middle| \mathcal{G}_n\right]$ is bounded and $E(T) < \infty$.*

Exercise: Prove Theorem 5.21

5.5 Convergence Theorems

We start with a simple theorem, which nevertheless contains the main idea of the general convergence theorem, and will slowly work our way to more and more general theorems.

Theorem 5.22 *Let* (X_n) *be a positive martingale. Then* $\lim_{n\to\infty} X_n$ *exists a.s.*

Theorem 5.22 is a special case of the following theorem:

Theorem 5.23 *Let* (X_n) *be a positive sub-martingale, and assume*

$$\sup\{E(X_n) : n = 1, 2, \ldots\} < \infty.$$

Then $\lim_{n\to\infty} X_n$ *exists a.s.*

Proof: Let

$$S = \sup\{E(X_n) : n = 1, 2, \ldots\}.$$

We first show that $P(\sup_n X_n < \infty) = 1$. To this end we will show that

$$\lim_{k\to\infty} P(\sup_n X_n > k) = 0.$$

Fix k, and let $T = \inf\{n : X_n > k\} \leq \infty$. Let $j > i$ be natural numbers. Then $E(X_j|T = i) > k$ because (X_n) is a sub-martingale. The events $\big(\{T = i\}\big)_{i=1,\ldots}$ are disjoint, and therefore $E(X_j|T < j) > k$. Using Markov's inequality,

$$P(T < j) < \frac{E(X_j)}{k} < \frac{S}{k}.$$

Since this holds for every t we get $P(T < \infty) < \frac{S}{k}$, and a result $P(\sup_{X_n} < \infty) = 1$.
27.6
We now proceed to prove the convergence. Let $a < b$ be rational. We define a sequence of stopping times T_k and S_k as follows:

$$\begin{aligned} T_1 &= \inf\{n : X_n < a\} \leq \infty; \\ S_k &= \inf\{n > T_k : X_n > b\} \leq \infty; \\ T_{k+1} &= \inf\{n > S_k : X_n < a\} \leq \infty. \end{aligned}$$

We let $C_{a,b} = \max\{k : S_k < \infty\}$, and call $C_{a,b}$ the *number of up-crossing from* a *to* b.
We define a new sub-martingale (Y_n) as follows: $Y_1 = X_1$. For every n, if there exists k s.t. $T_k \leq n < S_k$, then $Y_{n+1} = Y_n + X_{n+1} - X_n$ and else $Y_{n+1} = Y_n$. We make a few observations. The first is that $E(Y_n) \leq E(X_n)$ for every n. Indeed, let A_n be the event that there exists k s.t. $T_k \leq n < S_k$. Then $A_n \in \mathcal{G}_n$, and thus $E(1_{A_n}(X_{n+1} - X_n)|\mathcal{G}_n) = 1_{A_n} E(X_{n+1} - X_n|\mathcal{G}_n)$ and thus $E(Y_{n+1} - Y_n) \leq E(X_{n+1} - X_n)$ and by induction $E(Y_n) \leq E(X_n)$.
The second observation is that whenever $S_k < \infty$, we have that $Y_{S_k} \geq k(b - a)$. Thus, if we write $C_{a,b}^{(n)} = \max\{k : S_k \leq n\}$ (with the understanding that $\max\emptyset = 0$) for the

number of upcrossings up to time n, then we have

$$E(Y_n|C_{a,b}^{(n)} \geq k) = E(Y_n|S_k \leq n) = \sum_{j=1}^{n} E(Y_n|S_k = j)P(S_k = j|S_k \leq n)$$

$$\geq \sum_{j=1}^{n} E(Y_j|S_k = j)P(S_k = j|S_k \leq n) \geq \sum_{j=1}^{n} k(b-a)P(S_k = j|S_k \leq n) = k(b-a).$$

Therefore, $P(C_{a,b}^{(n)} \geq k) \leq (b-a)E(Y_n)/k \leq (b-a)E(X_n)/k \leq (b-a)S/k$, and as $C_{a,b} = \lim_{n\to\infty} C_{a,b}^{(n)}$ we get that $P(C_{a,b} \geq k) \leq (b-a)S/k$ for every k and thus $C_{a,b} < \infty$ with probability 1.
Therefore, a.s, for every rational a and b the number of up-crossing is finite, and therefore (X_n) converges almost surely. □

We can now state and prove the main theorem in this section.

Theorem 5.24 (The Martingale Convergence Theorem)
Let $(X_n)_{n=1}^{\infty}$ and assume that $\sup_n E(|X_n|) < \infty$. Then the sequence (X_n) converges almost surely.

Proof: By Theorem 5.23, both (X_n^+) and (X_n^-) converge, and therefore (X_n) converges. □

The same proof idea is useful in proving the inverse martingale convergence theorem.

Theorem 5.25 (Inverse Martingale Convergence Theorem)
Let $(X_n)_{n=1}^{\infty}$ be an inverse martingale. Then (X_n) converges almost surely.

Proof: Fix $a < b$ rational, and let $C_{a,b}(k)$ be the number of up crossings up to time k. The number of up crossings in $X_1, \ldots, X_k$ is the same as the number of down crossings in $X_k, X_{k-1}, \ldots X_1$. Since $X_k, X_{k-1}, \ldots X_1$ is a martingale, we get $E(C_{a,b}(k)) \leq E(|X_1 - X_k|)/(b-a) \leq 2E(|X_1|)$, and therefore, and $C_{a,b}(k)$ increases with k, with probability 1 we have that $\lim_{k\to\infty} C_{a,b}(k) < \infty$. Therefore, with probability 1,

$$\liminf_{k\to\infty} X_k = \limsup_{k\to\infty} X_k$$

and all we need to show is that the limit is not infinite. This is left as an exercise. □

1.7

5.6 Uniform integrability and convergence in L^1

Definition 5.26 Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. We say that the sequence (X_n) is *uniformly integrable* if for every $\epsilon > 0$ there exists M such that $E(|X_n| \cdot \mathbf{1}_{|X_n|>M}) < \epsilon$ for every n.

Exercises

1. If $\sup E(X_n^2) < \infty$ then (X_n) is uniformly integrable.

2. "Double or nothing": Let (Y_n) be i.i.d. variables with $P(Y_1 = 0) = P(Y_1 = 2) = 0.5$. Let $X_1 = 1$, and $X_{n+1} = X_n Y_n$ for all n. Then (X_n) is not uniformly integrable.

Lemma 5.27 *Let X be a random variable s.t. $E(|X|) < \infty$, and let $(\mathcal{F}_n)_{n=1}^{\infty}$ be σ-algebras. Let $X_n = E(X|\mathcal{F}_n)$. Then the sequence (X_n) is uniformly integrable.*

Proof: We first note that for every ϵ there exists δ such that for every A, if $P(A) < \delta$ then $E(|X| \cdot \mathbf{1}_A) < \epsilon$. Indeed, otherwise we can find a sequence of events A_n with $P(A_n) < 2^{-n}$ and $E(|X| \cdot \mathbf{1}_{A_n}) \geq \epsilon$. We then get the contradiction using the dominated convergence theorem, when we note that $|X| \cdot \mathbf{1}_{A_n} \leq |X|$ and that $|X| \cdot \mathbf{1}_{A_n}$ goes a.s. to zero.
Now, since $\{|X_n| > M\} \in \mathcal{F}_n$, we get $E(|X_n| \cdot \mathbf{1}_{|X_n|>M}) = E(|X| \cdot \mathbf{1}_{|X_n|>M})$, and since $P(|X_n| > M) \leq E(|X|)/M$, for M large enough $P(|X_n| > M) < \delta$ uniformly in n and we are done. □

Corollary 5.28 *Every Inverse martingale is uniformly integrable.*

Theorem 5.29 *Let $(X_n)_n^{\infty}$ be a uniformly integrable sequence of random variables, and assume that (X_n) converges almost surely to X. Then (X_n) converges to X in L^1.*

Proof: Unorthodoxly, we write $x \wedge M$ for $\text{sgn}(x) \cdot \min(|x|, M)$. Note that for fixed positive M, $x \wedge M$ is continuous in x. Fix $\epsilon > 0$. Fix M s.t. $E(|X_n| \cdot \mathbf{1}_{|X_n|>M}) < \epsilon$ for every n. Then $X_n \wedge M$ converges to $X \wedge M$ almost surely, and since they are bounded, also in L^1. Now, $E(|X_n - X_n \wedge M|) < \epsilon$, and by Fatou's lemma also $E(|X - X \wedge M|) < \epsilon$, so for every n large enough, $E(|X_n - X|) < 2\epsilon$. □

Example 5.30 Let $(X_n)_{n=1}^{\infty}$ be i.i.d. s.t. $E(|X_1|) < \infty$ and let

$$U_n = \frac{1}{n} \sum_{k=1}^{n} X_n.$$

Then (U_n) is an inverse martingale.

Proof of the law of large numbers: We just saw that $\lim_{n\to\infty} U_n$ exists, and by Kolmogorov's zero-one law the limit is an almost sure constant. We need to show that the limit is indeed $E(X_1)$. We first prove that U_n converge in L^1 to its a.s. limit. This follows from the fact that the sequence U_n, as any inverse martingale, is uniformly integrable. Now, as the expectation in continuous in L^1 we get that $U_n \to E(X_1)$.

□

Theorem 5.31 *Let $(X_n)_{n=1}^{\infty}$ be a martingale w.r.t a filtration $(\mathcal{G}_n)_{n=1}^{\infty}$. Then (X_n) is uniformly integrable if and only if there exists a variable X such that $X_n = E(X|\mathcal{G}_n)$ for every X.*

Proof: If such X exists, then (X_n) is uniformly integrable by Lemma 5.27. Conversely, assume that (X_n) is uniformly integrable. Let M be s.t. $E(|X_n|\mathbf{1}_{|\mathbf{X_n}|} > \mathbf{M}) < 1$ for every n. Then $E(|X_n|) < M + 1$ for every n, and thus (X_n) converges a.s. to some variable X. By uniform integrability, the convergence is also in L^1 (Theorem 5.29). Fix n and let $Y_n = E(X|\mathcal{G}_n)$. We want to show that $Y_n = X_n$ a.s. Fix $\epsilon > 0$. Then for m large enough, $E(|X - X_m|) < \epsilon$, and by Jensen inequality

$$E(|Y_n - X_n|) = E\big(\big|E(X|\mathcal{G}_n) - E(X_m|\mathcal{G}_n)\big|\big) = E\big(\big|E(X - X_m|\mathcal{G}_n)\big|\big) \leq E(|X - X_m|) < \epsilon$$

and we are done since ϵ is arbitrary. □