Probability
Spring 2008
Prof. A. Pisztora*

Chris Almost†

The textbook for this course is *Probability Theory and Examples* by Richard Durrett. Course notes by R.S. Varadhan are an additional reference, available from http://www.math.nyu.edu/faculty/varadhan/processes.html.

Contents

*pisztora@andrew.cmu.edu
†cdalmost@cmu.edu

1 Measure theoretic framework of probability

1.1 σ-fields

Let Ω be a non-empty set. We think of Ω as the collection of "outcomes" of an experiment.

1.1.1 Definition. A collection of subsets $\mathscr{F}$ of Ω is called a *σ-field* if

(i) $\varnothing, \Omega \in \mathscr{F}$;

(ii) $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}$; and

(iii) $A_1, A_2, \ldots \in \mathscr{F}$ implies $\bigcup_n A_n \in \mathscr{F}$.

By DeMorgan's law, if $A_1, A_2, \ldots \in \mathscr{F}$ then $\bigcup_n A_n \in \mathscr{F}$. Hence $\mathscr{F}$ contains the sets of the form $\bigcap_n \bigcup_{k \geq n} A_k$. This is the set of $\omega \in \Omega$ that are in infinitely many of the A_n, and will be denoted $\{A_n \text{ i.o.}\}$.

1.1.2 Examples.
(i) $\{\varnothing, \Omega\}$ is a σ-field.

(ii) The power set of Ω is a σ-field.

(iii) For any $A \subseteq \Omega$, $\{A, A^c, \Omega, \varnothing\}$ is a σ-field.

In using probability theory to model reality, σ-fields correspond to the collection of "observable events."

The (arbitrary) intersection of σ-fields is a σ-field, so given any collection $\mathscr{B}$ of subsets of Ω there is a smallest σ-field that contains $\mathscr{B}$. We denote this σ-field by $\sigma(\mathscr{B})$.

1.1.3 Examples.
(i) Let $(X, \mathscr{T})$ be a topological space. The σ-field $\sigma(\mathscr{T})$ generated by $\mathscr{T}$ is the *Borel σ-field*, and is denoted $\mathscr{B}(X)$.

(ii) Let $\mathscr{Z}$ be a countable partition of Ω. Then $\sigma(\mathscr{Z})$ has the particularly simple form $\{\bigcup_j Z_j \mid \{Z_j\} \subseteq \mathscr{Z}\}$.

1.1.4 Example. Consider the *time evolution* of a random system with *state space* $(S, \mathscr{S})$. The correct choice of Ω is S^I, where I is the collection of times under consideration, usually an interval or $\mathbb{N}$. For each $i \in I$, let $X_i \Omega \to S : \omega \mapsto \omega_i$ be the state of the system at time i. Clearly we would like all of the X_i to be measurable.

Suppose $I = \mathbb{N}$. Let $\mathscr{B}_n$ be the collection of sets of the form $\{X_n \in A\}$ for $A \in \mathscr{S}$, the collection of events observable at time $n \in \mathbb{N}$. $\mathscr{B}_n$ is a σ-field since $\mathscr{S}$ is a σ-field. Let $\mathscr{F}_n := \sigma(\mathscr{B}_1 \cup \cdots \cup \mathscr{B}_n)$, and let $\mathscr{F} := \sigma(\bigcup_{n=1}^{\infty} \mathscr{B}_n)$, the σ-field of "observable events." $(\Omega, \mathscr{F})$ is a *measurable space* and $\{\mathscr{F}_n\}$ is a *filtration* on this space.

Let $\mathscr{F}_n^* := \sigma(\bigcup_{k\geq n} \mathscr{B}_k)$, the σ-field of events observable at and after time n, and let $\mathscr{F}^* := \bigcap_{n=1}^{\infty} \mathscr{F}_n^*$, the *tail σ-field*. $\mathscr{F}^*$ contains events, for example, of the form $\{X_n \in A \text{ i.o}\}$ for a fixed $A \in \mathscr{S}$. Indeed,

$$\{X_n \in A \text{ i.o}\} = \bigcap_{n\geq 0} \bigcup_{k\geq n} \{X_k \in A\} = \bigcap_{n\geq m} \bigcup_{k\geq n} \{X_k \in A\} \in \mathscr{F}_m^*$$

for any $m \geq 1$, so $\{X_n \in A \text{ i.o}\} \in \mathscr{F}^*$.

1.2 Dynkin systems

1.2.1 Definition. A collection of subsets $\mathscr{D}$ of Ω is called a *Dynkin system* (or *λ-system*) if

(i) $\Omega \in \mathscr{D}$;

(ii) $A \in \mathscr{D}$ implies $A^c \in \mathscr{D}$; and

(iii) $A_1, A_2, \ldots \in \mathscr{D}$ with $A_i \cap A_j = \varnothing$ for all $i \neq j$ implies $\bigcup_i A_i \in \mathscr{D}$.

A non-empty collection of subsets Ω is called a *π-system* if it is closed under finite intersections.

Notice that if $A \subseteq B$ then $B \setminus A = (B^c \cup A)^c \in \mathscr{D}$.

1.2.2 Proposition. *If $\mathscr{D}$ is both a λ-system and a π-system then it is a σ-field.*

Proof: Indeed, in this case $A, B \in \mathscr{D}$ implies

$$A \cup B = (A \setminus A \cap B) \cup (A \cap B) \cup (B \setminus (A \cap B) \in \mathscr{D},$$

so $\mathscr{D}$ is closed under finite union. For arbitrary $A_1, A_2, \ldots \in \mathscr{D}$,

$$\bigcup_n A_n = \bigcup_n \left(\bigcup_{k\leq n} A_k \setminus \bigcup_{k<n} A_k \right) \in \mathscr{D}.$$

□

1.2.3 Theorem (Dynkin's π-λ Theorem). *Let $\mathscr{B}$ be a π-system. Then the smallest λ-system containing $\mathscr{B}$ is equal to $\sigma(\mathscr{B})$.*

Proof: Let $\mathscr{D}$ denote the smallest Dynkin system containing $\mathscr{B}$. $\mathscr{D} \subseteq \sigma(\mathscr{B})$ since every σ-field is a Dynkin system. We show that $\mathscr{D}$ is closed under finite intersection and conclude with the above Proposition. For $B \in \mathscr{B}$ consider the set

$$\{A \in \mathscr{D} \mid A \cap B \in \mathscr{D}\}.$$

This set is a Dynkin system (note $A^c \cap B = B \setminus (A \cap B)$) that contains $\mathscr{B}$ (since $\mathscr{B}$ is closed under finite intersection), so it is equal to $\mathscr{D}$.

Consider now, for any $D \in \mathscr{D}$, the set $\{A \in \mathscr{D} \mid A \cap D \in \mathscr{D}\}$. This set is a Dynkin system that contains $\mathscr{B}$ (by the first consideration) so it is equal to $\mathscr{D}$. Therefore $\mathscr{D}$ is closed under finite intersection. □

1.3 Probability measures

1.3.1 Definition. Let $(\Omega, \mathscr{F})$ be a measurable space. A *probability measure* is a map $\mathbb{P} : \mathscr{F} \to [0,1]$ such that

(i) $\mathbb{P}[\Omega] = 1$; and

(ii) if $A_1, A_2, \ldots \in \mathscr{F}$ with $A_i \cap A_j = \varnothing$ for all $i \neq j$ then $\mathbb{P}[\bigcup_i A_i] = \sum_i \mathbb{P}[A_i]$.

These two axioms imply many things about probability measures. For example,

(i) $\mathbb{P}[\varnothing] = 0$;

(ii) $\mathbb{P}[A] = 1 - \mathbb{P}[A^c]$ for all $A \in \mathscr{F}$;

(iii) $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$ for all $A, B \in \mathscr{F}$;

(iv) if $A_1, A_2, \ldots \in \mathscr{F}$ then $\mathbb{P}[\bigcup_n A_n] \leq \sum_n \mathbb{P}[A_n]$.

1.3.2 Example. If $\mathscr{X}$ is a countable partition of Ω...

1.3.3 Theorem (Monotone Convergence Theorem).
Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathscr{F})$.

(i) If $A_1 \subseteq A_2 \subseteq \cdots$ then $\mathbb{P}[\bigcup_n A_n] = \lim_{n\to\infty} \mathbb{P}[A_n]$; and

(ii) if $A_1 \supseteq A_2 \supseteq \cdots$ then $\mathbb{P}[\bigcap_n A_n] = \lim_{n\to\infty} \mathbb{P}[A_n]$.

1.3.4 Theorem (First Borel-Cantelli Lemma). *If $A_1, A_2, \ldots \in \mathscr{F}$ with $\sum_n \mathbb{P}[A_n] < \infty$ then $\mathbb{P}[\bigcap_n \bigcup_{k\geq n} A_k] = \mathbb{P}[\{A_n \text{ i.o.}\}] = 0$.*

PROOF:

$$\mathbb{P}\left[\bigcap_n \bigcup_{k\geq n} A_k\right] = \lim_{n\to\infty} \mathbb{P}\left[\bigcup_{k\geq n} A_k\right] \leq \liminf_{n\to\infty} \sum_{k\geq n} \mathbb{P}[A_k] = 0.$$

□

One must be careful about mixing limits and inequalities. It is not necessarily the case that $a_n \leq b_n$ for all n implies $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$, for the simple reason that either of those limits may not exist! However, one can conclude that both $\liminf_n a_n \leq \liminf_n b_n$ and $\limsup_n a_n \leq \limsup_n b_n$.

1.3.5 Theorem. *Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be two probability measures on a measurable space $(\Omega, \mathscr{F})$. If there is a π-system $\mathscr{B} \subseteq \mathscr{F}$ such that $\mathbb{P}_1$ and $\mathbb{P}_2$ agree on $\mathscr{B}$, then they agree on $\sigma(\mathscr{B})$.*

PROOF: Let $\mathscr{D} = \{A \in \mathscr{F} \mid \mathbb{P}_1[A] = \mathbb{P}_2[A]\}$. Then $\mathscr{D}$ contains $\mathscr{B}$ and is a Dynkin system. Indeed, if $A_1, A_2, \ldots \in \mathscr{D}$ are pairwise disjoint then

$$\mathbb{P}_1[\bigcup_n A_n] = \sum_n \mathbb{P}_1[A_n] = \sum_n \mathbb{P}_2[A_n] = \mathbb{P}_2[\bigcup_n A_n],$$

so $\bigcup_n A_n \in \mathscr{D}$. Whence $\mathscr{D}$ contains $\sigma(\mathscr{B})$ by Dynkin's π-λ Theorem. □

1.3.6 Example. Let S be a countable set, $\mathscr{S}$ be the power set of S, and $\Omega = S^{\mathbb{N}}$. Let $\mathscr{F} = \sigma\{\{X_k \in A\} \mid k \geq 0, A \in \mathscr{S}\}$. A *stochastic process* is any measure $\mathbb{P}$ on Ω. It is determined by its values on sets of the form $\{X_0 = s_0, X_1 = s_1, \ldots, X_k = s_k\}$ since the collection of sets of this form is closed under intersection and generates $\mathscr{F}$.

A sequence of *independent experiments* with values in S, where each experiment has outcome x with probability $\mu(x)$ (so μ is a probability measure on S), has the form

$$\mathbb{P}[X_0 = s_0, \ldots, X_k = s_k] = \mu(s_0) \cdots \mu(s_k).$$

A *Markov chain* on S with *initial distribution* μ (a probability measure on S) and *transition kernel* K is a probability measure of the form

$$\mathbb{P}[X_0 = s_0, \ldots, X_k = s_k] = \mu(s_0)K(s_0, s_1) \cdots K(s_{k-1}, s_k).$$

$K(s_j, s_k)$ is interpreted as the probability of jumping from state s_j to s_k in one time step. We must have $1 = \sum_{k \geq 1} K(s_j, s_k)$ for each $j \geq 1$. The existence of such a probability measure (given an initial distribution and transition kernel) follows from Carathéodory's extension theorem.

1.3.7 Theorem (Carathéodory). *Let $\mathscr{B}$ be an algebra on Ω and $\mathbb{P}$ be a normalized, σ-additive, set function on $\mathscr{B}$. Then there is a unique extension of $\mathbb{P}$ to $\sigma(\mathscr{B})$.*

The proof of this theorem is usually given in a course on measure theory. The uniqueness follows from the fact that $\mathscr{B}$ is closed under finite intersections. The following lemma comes in handy when actually applying Carathéodory's extension theorem.

1.3.8 Lemma. *Let $\mathscr{B}$ be an algebra on Ω and $\mathbb{P}$ an additive set function on $\mathscr{B}$. Then $\mathbb{P}$ is σ-additive on $\mathscr{B}$ if and only if "whenever $B_n \in \mathscr{B}$ are such that $B_n \searrow \varnothing$ then $\lim_{n\to\infty} \mathbb{P}(B_n) = 0$."*

1.4 Independence

1.4.1 Definition. A collection $\{\mathscr{B}_i\}_{i \in I}$ set-systems $\mathscr{B}_i \subseteq \mathscr{F}$ is *independent* if for every choice of $A_i \in \mathscr{B}_i$, the collection of events $\{A_i\}_{i\in I}$ is *independent*. That is to say, for every $J \subseteq I$ finite, $\mathbb{P}[\bigcap_{j\in J} A_j] = \prod_{j \in J} \mathbb{P}[A_j]$.

1.4.2 Theorem. *Let $\{\mathscr{B}_i\}_{i\in I}$ be a independent collection of set-systems that are closed under finite intersections. Then*

(i) $\{\sigma(\mathscr{B}_i)\}_{i\in I}$ is also independent; and

(ii) if $\{J_k\}_{k\in K}$ is a partition of I then $\{\sigma(\bigcup_{i\in J_k} \mathscr{B}_i)\}_{k\in K}$ is also independent.

1.4.3 Theorem (Second Borel-Cantelli Lemma). *If $A_1, A_2, \ldots \in \mathscr{F}$ are independent and $\sum_n \mathbb{P}[A_n] = \infty$ then $\mathbb{P}[\bigcap_n \bigcup_{k\geq n} A_k] = \mathbb{P}[\{A_n \text{ i.o.}\}] = 1$.*

PROOF: It suffices to show that $\mathbb{P}[\bigcup_{k\geq n} A_k] = 1$ for all $n \geq 1$, or equivalently that $\mathbb{P}[\bigcap_{k\geq n} A^c] = 0$, for all $n \geq 1$. But $(A_k^c)_{k\geq 1}$ is independent, so

$$\mathbb{P}\left[\bigcap_{k\geq n} A_k^c\right] = \lim_{m\to\infty} \mathbb{P}\left[\bigcap_{k=1}^m A_k^c\right]$$
$$= \lim_{m\to\infty} \prod_{k=1}^m (1 - \mathbb{P}[A_k]) \leq \liminf_m \exp\left(-\sum_{k=1}^m \mathbb{P}[A_k]\right) = 0$$

using the inequality $1 - x \leq e^{-x}$. □

1.4.4 Example. Let A_i be the event that an immortal, randomly typing monkey types your favourite book in the i^{th} one million characters that he types. $\mathbb{P}[A_i] > 0$, they have the same probability, and we may assume that they are independent. Therefore, by the second Borel-Cantelli lemma, the monkey will type your favourite book infinitely many times.

1.4.5 Theorem (Kolmogorov's 0-1 Law). *Let $(\mathscr{F}_n)_{n\in\mathbb{N}}$ be a countable collection of independent σ-fields. Define*

$$\mathscr{F}^* := \bigcap_{n\geq 1} \sigma\left(\bigcup_{k\geq n} \mathscr{F}_k\right),$$

the tail σ-field. Then $\mathbb{P}[A] = 0$ or 1 for all $A \in \mathscr{F}^*$.

PROOF: Set $\mathscr{F}_\infty = \sigma(\bigcup_{n\geq 1} \mathscr{F}_n) \supseteq \mathscr{F}^*$. We will show that $\mathscr{F}^*$ and $\mathscr{F}_\infty$ are independent, so that for any $A \in \mathscr{F}^*$, $A \in \mathscr{F}_\infty$ and

$$\mathbb{P}[A] = \mathbb{P}[A \cap A] = \mathbb{P}[A]^2.$$

Indeed, $\{\mathscr{F}_1, \dots, \mathscr{F}_n, \sigma(\bigcup_{k>n} \mathscr{F}_k)\}$ are independent and $\mathscr{F}^* \subseteq \sigma(\bigcup_{k>n} \mathscr{F}_k)$, so $\{\mathscr{F}_1, \dots, \mathscr{F}_n, \mathscr{F}^*\}$ are independent. It follows that $\{\mathscr{F}^*, \mathscr{F}_1, \mathscr{F}_2, \dots\}$ are independent, so $\mathscr{F}^*$ and $\mathscr{F}_\infty$ are independent. □

1.4.6 Example (Site percolation). Let $\Omega = \{0,1\}^{\mathbb{Z}^d}$ and let

$$\mathscr{F}_k = \sigma\{X_z \mid \|z\|_\infty = k, z \in \mathbb{Z}^d\}.$$

Define $\mathbb{P}$ on $(\Omega, \sigma\{X_z \mid z \in \mathbb{Z}^d\})$ by

$$\mathbb{P}[X_{z_1}, \dots, X_{z_k} = 1, X_{y_1}, \dots, X_{y_n} = 0] = p^k(1-p)^n.$$

Then $(\mathscr{F}_k)_{k\geq 0}$ is an independent family of σ-fields. Define a set $C \subseteq \mathbb{Z}^d$ to be *connected* if it is connected in the sense of the game of Go. An *open cluster* is a connected set of cells with value 1. Let A be the event that there is an open cluster of infinite size. Then $A \in \mathscr{F}^* = \bigcap_{n\geq 0} \sigma(\bigcup_{k\geq n} \mathscr{F}_k)$ since it doesn't matter what happens inside any box of finite size. By Kolmogorov's 0-1 Law, $\mathbb{P}[A] = 0$ or 1, so there is either always an infinite cluster or there is never one. This critical probability p_c must be strictly between 0 and 1 for $d \geq 2$. Unfortunately, the Law does not tell us what happens at $p = p_c$, and in fact little is known.

1.5 Measurable maps

1.5.1 Definition. Let $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{F}')$ be measurable spaces and let $X : \Omega' \to \Omega$ be a function. Define

$$\sigma(X) = \{\{X \in A\} \mid A \in \mathscr{F}\} = \{X^{-1}(A) \mid A \in \mathscr{F}\} = X^{-1}(\mathscr{F}),$$

the σ-field generated by X. The map X is said to be *$\mathscr{F}'/\mathscr{F}$-measurable* if $\sigma(X) \subseteq \mathscr{F}'$, i.e. if $\{X \in A\} \in \mathscr{F}'$ for all $A \in \mathscr{F}$.

1.5.2 Exercises.

(i) Check that $\sigma(X)$ truly is a σ-field.

(ii) Show that it is sufficient to check that $\{X \in B\} \in \mathscr{F}$ for all $B \in \mathscr{B} \subseteq \mathscr{F}$, where $\mathscr{B}$ generates $\mathscr{F}$. (Hint: $\{A \subseteq \Omega \mid \{X \in A\} \in \mathscr{F}'\}$ is a σ-field containing $\mathscr{B}$.)

(iii) Prove the composition of measurable maps is measurable.

1.5.3 Examples.

(i) A continuous map between topological spaces is measurable with respect to the corresponding Borel σ-fields. For $X : \Omega \to \mathbb{R}$, we say that X is $\mathscr{F}$-measurable if X is $\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable.

(ii) If $\mathscr{F} = \sigma(Z)$, where Z is a countable partition of Ω, then a measurable map is constant on each atom.

1.5.4 Lemma. *$\Phi : \Omega' \to \Omega$ is always $\sigma(\Phi)/\mathscr{F}$-measurable. If $X : \Omega' \to \mathbb{R}$ is $\sigma(\Phi)$-measurable then there is an $\mathscr{F}$-measurable map $\varphi : \Omega \to \mathbb{R}$ such that $X = \varphi \circ \Phi$.*

1.5.5 Example. Let $\Omega = \{0,1\}^{\mathbb{N}}$, and as usual let X_n be the n^{th} coordinate map, $\mathscr{F}_n := \sigma(X_n)$, and $\mathscr{F} := \sigma(\bigcup_{n\geq 0} \mathscr{F}_n)$. Let $T := \sum_{n\geq 0} 2^{-n} X_n$, so that $T : \Omega \to [0,1]$, and let $\Omega_0 = \{\{X_n = 1\} \text{ i.o.}\}$, so that $T|_{\Omega_0}$ is a bijection with $(0,1]$. We have for $c \in (0,1]$,

$$\{T < c\} = \bigcup_{n:\bar{c}_n = 1} \left\{ \{X_n = 0\} \cap \bigcap_{k<n} \{X_k = \bar{c}_k\} \right\} \in \mathscr{F},$$

so T is measurable. If we define $\mathbb{P}[X_0 = x_0, \dots, X_k = x_k] = 2^{-k}$ then it turns out that $\mathbb{P}[T < c] = c$, so T has a uniform distribution on $[0,1]$, and the law of T is Lebesgue measure on $[0,1]$.

1.5.6 Definition. Let $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{F}')$ be measurable spaces, $X : \Omega' \to \Omega$ a measurable function, and $\mathbb{P}'$ a probability measure on $(\Omega', \mathscr{F}')$. Then $\mathbb{P}_X[A] := \mathbb{P}'[X \in A]$ is a probability measure on $(\Omega, \mathscr{F})$, called the *image measure* of $\mathbb{P}'$ under X, or the *distribution* of X with respect to $\mathbb{P}'$, or the *law* of X, and is written $\mathbb{P}' \circ X^{-1}$.

1.6 Distribution functions

1.6.1 Definition. If μ is a probability measure on $\mathbb{R}$ then $F_\mu(x) := \mu((-\infty, x])$ is the *distribution function* (or *cumulative distribution function*) of μ.

1.6.2 Proposition. *The probability measure μ is uniquely determined by F_μ, and F_μ has the following properties.*

(i) $F_\mu : \mathbb{R} \to [0,1]$, *and* $\lim_{n\to-\infty} F_\mu(x) = 0$ *and* $\lim_{n\to\infty} F_\mu(x) = 1$*;*

(ii) F_μ *is increasing; and*

(iii) F_μ *is right continuous.*

1.6.3 Definition. A function $F : \mathbb{R} \to [0,1]$ that satisfies all of the properties of 1.6.2 is called a *distribution function*.

In this case define $\mu_F((-\infty, x]) := F(x)$. It can be shown that μ_F is a probability measure on $\mathbb{R}$, and $F_{\mu_F} = F$. Indeed, set $G(y) := \inf\{x \mid F(x) > y\}$, the unique right continuous "inverse" of F. Then

$$\{G \le x\} = \begin{cases} [0, F(x)) & \text{if } F \text{ is constant after } c \\ [0, F(x)] & \text{if } F \text{ is increasing after } c \end{cases}$$

so G is a measurable function $[0,1] \to \mathbb{R}$, and $\mu_F = \text{meas} \circ G^{-1}$.

1.7 Expectation

1.7.1 Definition. $X : (\Omega, \mathscr{F}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ (where $\mathscr{B}(\overline{\mathbb{R}})$ is generated by the collection $\{[-\infty, c] \mid c \in \mathbb{R}\}$) is called a *random variable* if it is $\mathscr{F}/\mathscr{B}(\overline{\mathbb{R}})$-measurable.

It is useful to note that the collection of r.v.'s is closed under "countable operations" such as sup, lim sup, and lim.

1.7.2 Definition. A *step function* (or *elementary function*) is a (finite) linear combination of indicator functions of measurable sets, $\sum_{i=1}^n c_i \mathbf{1}_{A_i}$. We may assume that the A_i's are disjoint and the c_i's are distinct.

1.7.3 Theorem. *If X is a non-negative r.v. then there is an increasing sequence of step functions X_n such that $X = \lim_{n\to\infty} X_n$.*

PROOF: Take X_n to be the following and check the conclusion of the theorem.

$$X_n := \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{k2^n \le X < (k+1)2^n} + n\mathbf{1}_{n \le X}.$$

□

1.7.4 Theorem (Lifting). *Let $T : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{F}')$ be measurable and let X be a r.v. that is also $\sigma(T)$-measurable. Then there is a r.v. X' on Ω' such that $X = X' \circ T$.*

PROOF (MEASURE-THEORETIC INDUCTION):

(i) If $X = \mathbf{1}_A$ for some $A \in \sigma(T)$ then $A = \{T \in B\}$ for some $B \in \mathscr{F}$, so take $X' = \mathbf{1}_B$.

(ii) If X is a step function then take X' to be the corresponding step function assembled from the bits obtained in Step 1.

(iii) If X is a non-negative r.v. then take step functions $X_n \nearrow X$. Use the X_n''s from Step 2 and check that $X' := \lim_{n\to\infty} X_n'$ works. Indeed,

$$X' \circ T = \lim_{n\to\infty} X_n' \circ T = \lim_{n\to\infty} X_n = X.$$

(iv) For a general r.v. write $X = X^+ - X^-$ and use the result from Step 3. □

1.7.5 Definition. The definition of *expectation* is by measure-theoretic induction.

(i) $\mathbb{E}[\mathbf{1}_A] := \mathbb{P}[A]$;

(ii) $\mathbb{E}[\sum_i c_i \mathbf{1}_{A_i}] := \sum_i c_i \mathbb{E}[\mathbf{1}_{A_i}]$;

(iii) For $X \geq 0$ with step functions $X_n \nearrow X$, $\mathbb{E}[X] := \lim_{n\to\infty} \mathbb{E}[X_n]$;

(iv) For any r.v. X, $\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

This definition requires checking. Namely, that it doesn't depend on the decomposition of a simple function and that it doesn't depend on the approximating sequence. It can be checked that expectation is linear, monotone, and the expectation of a r.v. that is zero a.s. is zero.

Let X be a r.v. on $(\Omega, \mathscr{F}, \mathbb{P})$. Then

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

$\mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ denotes the collection of *integrable* r.v.'s, i.e. those X for which $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$. An r.v. for which at most one (but not both) of those inequalities fails is said to be *semi-integrable*.

1.7.6 Theorem (Monotone Convergence Theorem).
If $X_0 \in \mathscr{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ and $X_0 \leq X_1 \leq \cdots$ $\mathbb{P}$-a.s. then

$$\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n].$$

This is also called the Beppo-Levi Theorem. That $X_0 \in \mathscr{L}^1$ is necessary. For a counterexample, take $\Omega = (0,1]$ and $X_0 := -\frac{1}{x}$ and $X_n := X_0 \mathbf{1}_{(0,\frac{1}{n}]}$. Then $X_n \nearrow 0$ but $\mathbb{E}[X_n] = -\infty$ for all n.

1.7.7 Theorem (Fatou's Lemma).
Suppose that $\{X_n\}_{n\geq 0}$ is bounded below by an integrable function. Then

$$-\infty < \mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n].$$

Similarly, if $\{X_n\}_{n\geq 0}$ is bounded above by an integrable function then

$$\infty > \mathbb{E}[\limsup_n X_n] \geq \limsup_n \mathbb{E}[X_n].$$

PROOF: Applying the monotone convergence theorem and the fact that expectation is monotone,

$$-\infty \leq \mathbb{E}[Y] \leq \mathbb{E}[\liminf_n X_n] = \lim_{n\to\infty} \mathbb{E}[\inf_{k\geq n} X_n] \leq \liminf_n \mathbb{E}[X_n]. \qquad \square$$

1.7.8 Theorem (Lebesgue Dominated Convergence Theorem).
Suppose that $|X_n|$ is dominated by an integrable function and $X := \lim_{n\to\infty} X_n$ $\mathbb{P}$-a.s. Then $\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$ and $\mathbb{E}[|X - X_n|] \to 0$ as $n \to \infty$.

PROOF: First, $|X_n| \to |X|$ a.s., so $|X|$ is dominated by an integrable function and thus is integrable. Applying Fatou's Lemma,

$$\begin{aligned}\mathbb{E}[X] = \mathbb{E}[\limsup_n X_n] &\geq \limsup_n \mathbb{E}[X_n] \\ &\geq \liminf_n \mathbb{E}[X_n] \geq \mathbb{E}[\liminf_n X_n] = \mathbb{E}[X]\end{aligned}$$

Further, applying the result to $|X - X_n|$ gives that $\mathbb{E}[|X - X_n|] \to 0$. $\square$

Remark. In measure theory $0 \cdot \infty = 0$.

1.7.9 Theorem (Chebyshev's inequality).
Let $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be a non-negative measurable function. Let $A \in \mathscr{B}(\overline{\mathbb{R}})$ and set $C_A = \inf_{a\in A} \varphi(a)$. Then for any r.v. X,

$$C_A \mathbb{P}[X \in A] \leq \mathbb{E}[\varphi(X); X \in A] \leq \mathbb{E}[\varphi(X)].$$

1.7.10 Theorem (Jensen's inequality). *Let $X \in \mathscr{L}^1$ and u be convex.*

(i) $\mathbb{E}[u(X)] \geq u(\mathbb{E}[X])$.

(ii) If u is strictly convex and X is not constant a.s. then $\mathbb{E}[u(X)] > u(\mathbb{E}[X])$.

Warning: This statement of Jensen's inequality holds only for probability measures (and not for general measures, or even σ-finite measures).

PROOF: Since u is convex it has a support line $\ell(x) = ax + b$ over $\mathbb{E}[X]$, i.e. such that $u(x) \geq \ell(x)$ for all $x \in \mathbb{R}$ and $u(\mathbb{E}[X]) = \ell(\mathbb{E}[X])$. Then

$$\mathbb{E}[u(X)] \geq \mathbb{E}[\ell(X)] = \ell(\mathbb{E}[X]) = u(\mathbb{E}[X]). \qquad \square$$

1.7.11 Example. For $p \geq 1$, let $\mathscr{L}^p$ be the collection of r.v.'s X such that $\mathbb{E}[|X|^p] < \infty$. Then for any $1 \leq p' \leq p$ we have

$$\infty > \mathbb{E}[|X|^p] = \mathbb{E}[|X|^{p' \frac{p}{p'}}] \geq \mathbb{E}[|X|^{p'}]^{\frac{p}{p'}},$$

so $\mathscr{L}^p \subseteq \mathscr{L}^{p'}$. For probability measures the L^p spaces are decreasing, and the embeddings are contractions.

1.7.12 Theorem. *Let $T : (\Omega', \mathscr{F}') \to (\Omega, \mathscr{F})$ be measurable and $\mathbb{P}'$ be a measure on $(\Omega', \mathscr{F}')$ and $\mathbb{P}$ the induced measure on $(\Omega, \mathscr{F})$. Then for any non-negative r.v. X on $(\Omega, \mathscr{F})$, $\mathbb{E}[X] = \mathbb{E}'[X \circ T]$.*

PROOF: By measure-theoretic induction. ☐

1.7.13 Corollary. *If X is a r.v. such that $\mathbb{P}[X \in \mathbb{R}] = 1$ then the distribution of $\mu = \mathbb{P} \circ X^{-1}$ is concentrated on $\mathbb{R}$ and for each measurable function $\varphi \geq 0$,*

$$\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)\mu(dx)$$

if the expectation on the left is defined.

1.7.14 Examples.

(i) $\mathbb{E}[X^+] = \int_0^\infty x\mu(dx)$ and $\mathbb{E}[X^-] = \int_{-\infty}^0 x\mu(dx)$.

(ii) $\mathbb{E}[X] = \int_{-\infty}^\infty x\mu(dx)$ whenever one of the above expectations is defined.

(iii) $\mathbb{E}[|X|^k] = \int_{-\infty}^\infty |x|^k\mu(dx)$, the k^{th} *moment* of X.

(iv) For a pair of r.v.'s $(X, Y) : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}^2$,

$$\mu((-\infty, a] \times (-\infty, b]) := \mathbb{P}[(X, Y) \in (-\infty, a] \times (-\infty, b]] = \mathbb{P}[X \leq a, Y \leq b]$$

μ is the induced measure on $\mathbb{R}^2$, the *joint distribution* of X and Y.

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\ &= \int_{\mathbb{R}^2} xy\,\mu(dx, dy) - \int_{\mathbb{R}} x\,\mathbb{P}[X \in dx] \int_{\mathbb{R}} y\,\mathbb{P}[Y \in dy]\end{aligned}$$

(v) $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

1.7.15 Theorem.

(i) For $p \leq q$, $\|X\|_p \leq \|X\|_q$, which implies that $\mathscr{L}^p \supseteq \mathscr{L}^q$.

(ii) Hölder's inequality: $\|XY\|_1 \leq \|X\|_p \|Y\|_q$ when $\frac{1}{p} + \frac{1}{q} = 1$ and for all $X \in \mathscr{L}^p$, $Y \in \mathscr{L}^q$.

(iii) *Minkowski's inequality:* $\|X + X'\|_p \le \|X\|_p + \|X'\|_p$.

1.7.16 Theorem (Riesz-Fischer). $(L^p, \|\cdot\|_p)$ *is complete for all* $1 \le p \le \infty$.

1.7.17 Theorem. *Random variables* X *and* Y *are independent if and only if*

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)] \text{ for all } f, g \ge 0 \text{ measurable.}$$

In this case, if $X, Y \in \mathscr{L}^1$ *then* $XY \in \mathscr{L}^1$

2 Product spaces

2.1 Kernels and Fubini's theorem

Let $(S_i, \mathscr{S}_i)$ be measurable spaces and $\Omega = S_1 \times S_2$ and $X_i : \Omega \to S_i$ denote projection on the i^{th} coordinate.

2.1.1 Definition. A *stochastic kernel* from S_1 to S_2 is a map $K : S_1 \times \mathscr{S}_2 \to [0,1]$ such that

(i) $K(x, \cdot)$ is a probability measure on S_2 for all $x \in S_1$; and

(ii) $K(\cdot, A)$ is $\mathscr{S}_1$-measurable for all $A \in \mathscr{S}_2$.

The interpretation of a stochastic kernel is that $K(x, A)$ is the probability of a transition from x (the current location) into the set A.

2.1.2 Examples.

(i) $K(x, \cdot) = \mathbb{P}_2[\cdot]$ for all $x \in S_1$ is a kernel with no coupling; all transitions are independent.

(ii) $K(x, \cdot) = \delta_{T(x)}$ for $T : S_1 \to S_2$ measurable is a kernel with deterministic coupling; the only transition from x_1 is to $x_2 = T(x_1)$.

(iii) Countable Markov chains. Take $S_1 = S_2 = S$ countable with the power set σ-field, and take $K(x, \{y\}) = q_{xy}$ to be a (countable) matrix with $q_{xy} \ge 0$ for all x, y and $\sum_y q_{xy} = 1$ for all x. Then $K(x, A) = \sum_{y \in A} q_{xy}$.

(iv) Take $S_1 = S_2 = \mathbb{R}$ with the Borel σ-field, and $K(x, \cdot) = \mathscr{N}(0, \beta x^2)$. Is there a $\beta > 0$ such that the system converges to zero? (What do we mean by "converges" in this case?)

Given a probability measure $\mathbb{P}_1$ on $(S_1, \mathscr{S}_1)$ and a stochasitic kernel K from S_1 to S_2, we would like to construct a probability measure $\mathbb{P}$ on the product space such that

$$\mathbb{P}[X_1 \in A_1] = \mathbb{P}_1[A_1] \quad \text{and} \quad \text{“}\mathbb{P}[X_2 \in A_2 \mid X_1 = x_1] = K(x_1, A_2)\text{”}$$

2.1.3 Theorem (Fubini). *Let $\Omega := S_1 \times S_2$ and $\mathscr{F} := \mathscr{S}_1 \otimes \mathscr{S}_2 := \sigma(\mathscr{S}_1 \times \mathscr{S}_2)$. Given a probability measure $\mathbb{P}_1$ on $(S_1, \mathscr{S}_1)$ and a stochastic kernel K from S_1 to S_2, there is a unique probability measure $\mathbb{P}$ on Ω such that for all $f \in m^+\mathscr{F}$,*

$$\int_\Omega f \, d\,\mathbb{P} := \int_{S_1} \left(\int_{S_2} f(x_1, x_2) K(x_1, dx_2) \right) \mathbb{P}_1(dx_1).$$

In particular,

(i) $\mathbb{P}[A_1 \times A_2] = \int_{A_1} \mathbb{P}_1(dx_1) K(x_1, A_2)$.

(ii) $\mathbb{P}[A] = \int_{S_1} \mathbb{P}_1(dx_1) K(x_1, A_{x_1})$ *where* $A_{x_1} := \{x_2 \in S_2 \mid (x_1, x_2) \in A\}$ *is the* x_1*-section of* A*;*

We denote $\mathbb{P} := \mathbb{P}_1 \times K$.

PROOF: Make notes on the proof of this. □

2.1.4 Corollary. *If* $K(x, \cdot) = \mathbb{P}_2[\cdot]$ *has no coupling then* $\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2$ *and we recover the classical theorem of Fubini.*

Remark. Fubini's theorem is true for σ-finite measures, such as Lebesgue measure on an infinite interval, but in this case the integrand needs to be integrable.

2.1.5 Example. For a non-negative r.v. X,

$$\begin{aligned}
\mathbb{E}[X] &= \int_\Omega X(\omega) \mathbb{P}[d\omega] \\
&= \int_\Omega \mathbb{P}[d\omega] \int_0^\infty \mathbf{1}_{[0, X(\omega))}(s) \, ds \\
&= \int_0^\infty ds \left(\int_\Omega \mathbf{1}_{(s,\infty]}(X(\omega)) \mathbb{P}[d\omega] \right) \\
&= \int_0^\infty \mathbb{P}[X > s] \, ds
\end{aligned}$$

and in fact $\int f \, d\mu = \int_0^\infty \mu(f > c) \, dc$ for any finite measure μ, and this implies the inequality

$$\int |f| \, d\mu \geq \sup_c c\mu(|f| > c).$$

2.2 Countable products

Let $\{(S_n, \mathscr{S}_n)\}_{n \geq 0}$ be a countable collection of measurable spaces and let

$$(S^n, \mathscr{S}^n) := (S_0 \times \cdots \times S_n, \mathscr{S}_0 \otimes \cdots \otimes \mathscr{S}_n)$$

Let μ_0 be a probability measure on S_0 and for $n \geq 1$, let K_n be a stochastic kernel from S^{n-1} to S_n (note the indices). Set $\mu^0 := \mu_0$ and inductively define $\mu^n := \mu^{n-1} \times K_n$, a measure on S^n, so that for $f \in m^+\mathscr{S}^n$,

$$\int_{S^n} f \, d\mu^n = \int_{S_0} \mu_0(dx_0) \int_{S_1} K_1(x_0, dx_1) \cdots \int_{S_n} K_n(x_0 x_1 \cdots x_{n-1}, dx_n) f(x).$$

Finally, set $\Omega := S_0 \times S_1 \times \cdots$, let $X_n(\omega) := \omega_n$ be the projections, and let $\mathscr{F}_n := \sigma(X_0, \dots, X_n)$ and $\mathscr{F} := \sigma(X_0, X_1, \dots) =: \sigma(\bigcup_{n\geq 0} \mathscr{F}_n)$. We would like a probability measure μ on $(\Omega, \mathscr{F})$ such that $\mu \circ X_{\{0,\dots,n\}}^{-1} = \mu^n$ for all $n \geq 0$.

2.2.1 Theorem (Ionescu-Tulcea). *Given the setup above, there is a unique probability measure μ on $(\Omega, \mathscr{F})$ such that $\mu \circ X_{\{0,\dots,n\}}^{-1} = \mu^n$ for all $n \geq 0$.*

Proof: Since the $\mathscr{F}_n$'s are increasing, $\mathscr{A} := \bigcup_{n\geq 0} \mathscr{F}_n$ is an algebra. Define μ on $\mathscr{A}$ by

$$\mu(A^n \times S_{n+1} \times S_{n+2} \times \cdots) := \mu^n(A^n)$$

for $A^n \in \mathscr{S}^n$. Now μ is well-defined since if $A \in \mathscr{F}_n \cap \mathscr{F}_{n-1}$ and $A = A^n \times S_{n+1} \times \cdots = A^{n-1} \times S_n \times \cdots$ then

$$\begin{aligned}
\mu(A) = \mu^n(A^n) &= \mu^n(A^{n-1} \times S_n) \\
&= \int_{S^{n-1}} \mu^{n-1}(d\hat{x}) \int_{S_n} K(\hat{x}, dx_n) \mathbf{1}_{A^n}(\hat{x}) \mathbf{1}_{S^n}(x_n) \\
&= \mu^{n-1}(A^{n-1}),
\end{aligned}$$

and μ is additive on $\mathscr{A}$ by monotonicity. To apply the Carathéodory extension theorem we must prove that μ is σ-additive on $\mathscr{A}$, and then we may conclude that there is a unique extension of μ to a probability measure on $\mathscr{F} = \sigma(\mathscr{A})$ satisfying the conclusion of the theorem.

Suppose that $A_n \in \mathscr{A}$ are such that $A_n \searrow \varnothing$. Without loss of generality, suppose that $A_n \in \mathscr{A}_n$ for all $n \geq 1$ (see course notes), and write $A_n = A^n \times S_{n+1} \times \cdots$.

Fill in the details from the course notes. □

2.3 Applications

2.3.1 Theorem (Weak Law of Large Numbers).
Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s (in fact uncorrelated is enough) with finite variance σ^2 and mean μ. Then $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ converges to μ in $\mathscr{L}^2$ and in probability.

Proof: Recall that the variance of X_i variance is $\mathbb{E}[(X_i - m)^2] = \sigma^2$.

$$\begin{aligned}
\mathbb{E}[(\overline{X}_n - m)^2] &= \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n (X_i - m)^2\right] && \text{by independence} \\
&= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0
\end{aligned}$$

so $\overline{X}_n \to 0$ in $\mathscr{L}^2$. That the convergence is also in probability follows from the next lemma. □

2.3.2 Lemma. *For any $p > 0$, convergence in $\mathscr{L}^p$ implies convergence in probability.*

PROOF: Suppose that $X_n \to X$ in $\mathscr{L}^p$. By subtraction it suffices to consider $X_n \to 0$. Let $\varepsilon > 0$. By Chebyshev's inequality, with $\varphi(x) = |x|^p$, $A = \{|x| \geq \varepsilon\}$, and $C_A = \varepsilon^p$, we get $\mathbb{P}[|X_n| \geq \varepsilon] \leq \varepsilon^{-p}\,\mathbb{E}[|X_n|^p] \to 0$ as $n \to \infty$. □

2.3.3 Theorem (Strong Law of Large Numbers).
Let $X_1, X_2, \dots$ be a sequence of pairwise independent identically distributed r.v.'s with $\mu := \mathbb{E}X_1 < \infty$. Then $\overline{X}_n \to \mu$, $\mathbb{P}$-a.s.

PROOF:

(i) Show WLOG $X_1 \geq 0$.

(ii) Apply truncation.

(iii) Estimate variance.

(iv) Prove convergence along a subsequence.

(v) Fill the gap between the subsequence and the full sequence. □

2.3.4 Lemma. *Almost sure convergence implies convergence in probability.*

PROOF: Fill this in. □

2.3.5 Theorem (Strong Law of Large Number for Semi-integrable). *Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s with $\mathbb{E}[X_1^+] = +\infty$ and $\mathbb{E}[X_1^-] < \infty$. Then $\overline{X}_n \to +\infty$, $\mathbb{P}$-a.s.*

2.3.6 Example (Renewals). Let $X_1, X_2, \dots$ be i.i.d. with $0 < X_i < \infty$, $\mathbb{P}$-a.s., and let $T_n := X_1 + \cdots + X_n$. Think of X_i as the lifetime of a light bulb and T_n as the time of the replacement of the n^{th} light bulb. Let

$$N_t := \sup\{n \geq 1 \mid T_n \leq t\},$$

the number of replacements up to time t. We claim that if $\mathbb{E}X_1 = \mu \leq \infty$ then $\frac{1}{t}N_t \to \frac{1}{\mu}$, $\mathbb{P}$-a.s. Indeed, by the SLLN, $\frac{1}{n}T_n \to \mu$, $\mathbb{P}$-a.s. Note that since $X_i < \infty$, $N_t \to \infty$ as $t \to \infty$. By definition of N_t, $T_{N_t} \leq t < T_{N_t+1}$, so dividing both sides by N_t and taking $t \to \infty$,

$$\mu \leftarrow \frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t+1}\frac{N_t+1}{N_t} \to \mu, \ \mathbb{P}\text{-a.s. as } t \to \infty.$$

2.3.7 Example (Monte-Carlo Integration). How do we approximate an integral of the form

$$\int_{[0,1]} \cdots \int_{[0,1]} \varphi(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

when φ may be complicated to deal with? Assume that $X_1, \dots, X_n$ are i.i.d. uniform on $[0,1]$. Then $\vec{X}$ has the distribution of λ_n on $[0,1]^n$, so

$$\mathbb{E}[\varphi(\vec{X})] = \int_{[0,1]^n} \varphi(x_1, \dots, x_n)\, \lambda_n(dx).$$

By the SLLN we can approximate the integral by computing a large number of realizations and taking the average.

3 Conditional Expectation

3.0.8 Definition. Let $\mathscr{F}_0 \subseteq \mathscr{F}$ and $X \in m\mathscr{F}$. An $X_0 \in m\mathscr{F}$ such that $\mathbb{E}[X\mathbf{1}_{A_0}] = \mathbb{E}[X_0\mathbf{1}_{A_0}]$ for all $A_0 \in \mathscr{F}_0$ is (a version of) the *conditional expectation* of X given $\mathscr{F}_0$, and is denoted $\mathbb{E}[X \mid \mathscr{F}_0]$.

3.0.9 Theorem. *Let $\mathscr{F}_0 \subseteq \mathscr{F}$ and $X \in m\mathscr{F}$. If $X \in \mathscr{L}^1$ or $X \geq 0$ then $\mathbb{E}[X \mid \mathscr{F}_0]$ exists and is unique.*

PROOF: If $X \in \mathscr{L}^1$ then it is a difference of non-negative random variables, so we may assume that $X \geq 0$. Let $Q[A_0] = \mathbb{E}[X\mathbf{1}_{A_0}]$ for $A_0 \in \mathscr{F}_0$, a σ-finite measure on $\mathscr{F}_0$ that is absolutely continuous with respect to $\mathbb{P}|_{\mathscr{F}_0}$. By the Radon-Nikodym theorem there is $X_0 \in m^+\mathscr{F}_0$ such that $Q[A_0] = \mathbb{E}[X_0\mathbf{1}_{A_0}]$. Uniqueness is trivial. □

3.1 Discrete-time martingales

Skipped the definitions of filtrations, martingales.

3.1.1 Definition. $\mathscr{H} \subseteq \mathscr{L}^1$ is said to be *uniformly integrable* (or *u.i.*) if

$$\lim_{c\to\infty} \sup_{\mathscr{H}} \int_{\{|X|>c\}} |X|\, d\,\mathbb{P} = 0$$

3.1.2 Examples.

(i) If $X \in \mathscr{L}^1$ then $\{X\}$ is u.i., since

$$\int_{\{|X|>c\}} |X|\, d\,\mathbb{P} = \int |X|\, d\,\mathbb{P} - \int_{\{|X|\leq c\}} |X|\, d\,\mathbb{P} \to 0 \text{ as } c \to \infty$$

by the monotone convergence theorem, since $\mathbf{1}_{|X|\leq c}|X| \nearrow |X|$ as $c \to \infty$.

(ii) If there is $Y \in \mathscr{L}^1$ such that $|X| \leq Y$ for all $X \in \mathscr{H}$ then $\mathscr{H}$ is u.i. To show this we need another characterization of uniform integrability.

3.1.3 Lemma. *If there is $g : \mathbb{R} \to \mathbb{R}$ with $\frac{g(x)}{x} \to \infty$ as $x \to \infty$ such that*

$$\sup_{\mathscr{H}} \int g(|X|)\, d\,\mathbb{P} < \infty$$

then $\mathscr{H}$ is u.i.

In particular, if $\mathscr{H}$ is bounded in $\mathscr{L}^p$ for any $p > 1$ then it is u.i. If $\mathscr{H}$ is bounded in $\mathscr{L}^1$ then it is not necessarily u.i.

3.1.4 Theorem. *$f_n \to f$ in $\mathscr{L}^1$ if and only if the set $\{f_n\}$ is u.i. and $f_n \to f$ in probability.*

In the particular case where $f_n \to f$ pointwise (a stronger form of convergence then in probability) and there is $g \in \mathscr{L}^1$ with $|f_n| \le g$ for all n (which implies in particular that $\{f_n\}$ is u.i.) this theorem is seen to imply the Lebesgue dominated convergence theorem. It is better theorem because it characterizes convergence in $\mathscr{L}^1$.

3.1.5 Example (Gambler's Ruin). Let $X_n = x + S_n$, where S_n is a simple random walk with probability p of increase. Let $a < x < b$ and

$$T := \inf\{n \ge 0 \mid X_n \notin (a, b)\}.$$

By the Borel-Cantelli Lemma $\mathbb{P}[T < \infty] = 1$, and in fact it can be shown that $\mathbb{E}[T] < \infty$. Indeed, let A_n be the event that X_n leaves (a, b) at a time in the set

$$\{n(b-a), n(b-a)+1, \dots, (n+1)(b-a)-1\}.$$

Then $\mathbb{P}[A_n] \ge p^{b-a} > 0$, the latter being the probability of the path consisting of $(b-a)$ up-steps. It follows that $\sum_{n\ge 0} \mathbb{P}[A_n] = \infty$, and since these events are independent, $\mathbb{P}[A_n \text{ i.o.}] = 1$ by the second Borel-Cantelli Lemma. In particular

$$\mathbb{P}[T < \infty] = \mathbb{P}[\bigcup_{n\ge 0} A_n] \ge \mathbb{P}[A_n \text{ i.o.}] = 1$$

The probability of ruin is $r(x) := \mathbb{P}[X_T = a]$. When $p = \frac{1}{2}$ X itself is a martingale, so

$$x = \mathbb{E}[X_0] = \mathbb{E}[X_{T\wedge n}] \to \mathbb{E}[X_T] = b(1 - r(x)) + ar(x),$$

by dominated convergence, since X_T is bounded. Whence in the case of a symmetric random walk $r(x) = \frac{b-x}{b-a}$.

When $p \ne \frac{1}{2}$, let $h(z) = (\frac{1-p}{p})^z$, and notice that $h(X_n)$ is a martingale. Indeed,

$$\begin{aligned}\mathbb{E}[h(X_{n+1}) \mid \mathscr{F}_n] &= \mathbb{E}[(\tfrac{1-p}{p})^{X_n + Y_{n+1}} \mid \mathscr{F}_n] \\ &= (\tfrac{1-p}{p})^{X_n}\big((\tfrac{1-p}{p})^1 p + (\tfrac{1-p}{p})^{-1}(1-p)\big) = h(X_n)\end{aligned}$$

By optional stopping and dominated convergence,

$$h(x) = \mathbb{E}[X_0] = \mathbb{E}[h(X_T)] = h(b)(1 - r(x)) + h(a)r(x)$$

so in non-symmetric case

$$r(x) = \frac{1 - (\frac{p}{1-p})^{b-x}}{1 - (\frac{p}{1-p})^{b-a}}.$$

Note that if $p < \frac{1}{2}$ then $r(x) \geq 1 - (\frac{p}{1-p})^{b-x}$, and this lower bound does not depend on a. In the particular case of casino roulette, the probability of winning is $\frac{18}{37}$.

3.2 Gambling systems

Let $(V_n, n \geq 1)$ be a *previsible process*, i.e. a process for which V_n is $\mathscr{F}_{n-1}$-measurable for all $n \geq 1$. Let $(X_n, n \geq 0)$ be a martingale. Then the *discrete stochastic integral* $V \cdot X$, defined by

$$(V \cdot X)_n := \sum_{k=1}^{n} V_k(X_k - X_{k-1})$$

for $n \geq 1$ and $(V \cdot X)_0 = 0$, is a martingale.

A common and illustrative way to interpret V is as a *gambling system*. If $(Y_n, n \geq 1)$ is a sequence of i.i.d. Bernoulli random variables with parameter $p = 2$, representing the sequence of outcomes of a fair game, then $X_n := \sum_{i=1}^{n} Y_i$ is a martingale, and $(V \cdot X)_n = \sum_{i=1}^{n} V_n Y_n$ is the cumulative winnings up to and including time n.

3.2.1 Example. How long do you have to wait for the occurrence of a fixed binary text $[a_1 \cdots a_N]$ in a random binary sequence? Let $(Y_n, n \geq 0)$ be a sequence of independent Bernoulli r.v.'s with $p = \frac{1}{2}$, and set

$$T := \inf\{n \geq 1 \mid y_{n-N+1} = a_1, \dots, Y_n = a_N\}.$$

By the second Borel-Cantelli Lemma $\mathbb{P}[T < \infty] = 1$ (Divide the infinite text into blocks of text of length N. There is a non-zero probability that a block is the desired text, and the blocks are independent, so a block of text is the desired string infinitely often. Estimate to show that $T \in \mathscr{L}^1$.) What is $\mathbb{E}[T]$?

Consider the following gambling system...

3.2.2 Example (Branching process). Let $\{Y_{n,k} \mid n, k \geq 1\}$ be an array of i.i.d. r.v.'s with distribution μ (not deterministic), with $\mathbb{E}[Y_{1,1}] = m < \infty$. Let $X_0 := 1$ and

$$X_n := Y_{n,1} + \cdots + Y_{n,X_n}.$$

X represents the size of the total population at the n^{th} generation, where $Y_{n,k}$ represents the number of children of the k^{th} individual, and parent die every generation. Let $\mathscr{A}_n = \sigma(Y_{\ell,k}, 1 \leq \ell \leq n, k \geq 1)$ and $M_n = \frac{X_n}{m^n}$. Then M_n is a martingale for this filtration, bounded in $\mathscr{L}^1$.

Let $T = \min\{n \geq 0 \mid X_n = 0\}$ be the time of extinction.

3.2.3 Example (Microeconomics). Let $g : \mathbb{R} \to \mathbb{R}$ be a given (payoff) function. Let $X_n = x + Y_1 + \cdots + Y_n$ be a random walk on $\{a < n < b\}$ and let $S := \min\{n \geq 0 \mid X_n \in \{a, b\}\}$ be the first hitting time of X to the boundary. We wish to find a stopping time T for which $\mathbb{E}[g(X_T^S)]$ is maximal.

For the solution, let h the concave envelope of g (the smallest concave function at least g pointwise). Then $h(X_n^S)$ is a super-martingale, so

$$h(x) = \mathbb{E}^x[h(X_0^S)] \geq \mathbb{E}^x[h(X_T^S)] \geq \mathbb{E}^x[g(X_T^S)]$$

for any stopping time T. If $h(x) = g(x)$ then the solution is to take $T := 0$, and if $h(x) > g(x)$ then we take $T := \min\{n \geq 0 \mid X_n \in \{g = h\}\}$. Then $h(X^T)$ is a martingale since h is linear on this range, so

$$h(x) = \mathbb{E}^x[h(X_0)] = \mathbb{E}^x[h(X_T)] = \mathbb{E}^x[g(X_T)]$$

is best possible.

4 Weak Convergence

4.1 Fourier transforms of probability measures

Let $\mu \in \mathscr{M}_1(\mathbb{R}^d)$. There is a r.v. X such that the law of μ is X, and we will make use of this fact heavily, but it is possible to develop this theory for $\mu \in \mathscr{M}_f(\mathbb{R}^d)$.

4.1.1 Definition. For $u \in \mathbb{R}^d$,

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{iu\cdot x} \mu(dx) = \mathbb{E}[e^{iu\cdot X}],$$

the *Fourier transform* of μ.

Notice that $|\mathbb{E}[e^{iu\cdot X}]| \leq \mathbb{E}[|e^{iu\cdot X}|] = 1$ for all u, and in fact the Fourier transform is continuous.

4.1.2 Examples.

(i) If $f \in L^1(\mathbb{R}^d)$ then $\hat{f}(u) := \int_{\mathbb{R}^d} e^{iu\cdot x} f(x) dx$. Further, $|\hat{f}(u)| \leq \|f\|_1$ for all $u \in \mathbb{R}^d$, so $\hat{f} \in C_b(\mathbb{R}^d)$.

(ii) Let $\varphi_\varepsilon(x) := \frac{1}{\varepsilon\sqrt{2\pi}} e^{-\frac{x^2}{2\varepsilon^2}}$, the density of a $\mathscr{N}(0, \varepsilon^2)$ distribution with respect to Lebesgue measure. Then $\hat{\varphi}_\varepsilon(u) = e^{\frac{1}{2}\varepsilon^2 u^2}$.

4.1.3 Theorem. *Let $f \in C_c(\mathbb{R})$. Then*

$$\lim_{\varepsilon \downarrow 0} \sup_{x \in \mathbb{R}} |f * \varphi_\varepsilon(x) - f(x)| = 0.$$

This says that $f * \varphi_\varepsilon \to f$ in $(C_0, \|\cdot\|_\infty)$ as $\varepsilon \to 0$. Unfortunately $f * \varphi_\varepsilon \notin C_c$.

Index