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](http://www.math.nyu.edu/faculty/varadhan/processes.html).

Contents

1 Measure theoretic framework of probability 2
   1.1 \( \sigma \)-fields ............................................. 2
   1.2 Dynkin systems .............................................. 3
   1.3 Probability measures ......................................... 4
   1.4 Independence .................................................. 5
   1.5 Measurable maps .............................................. 7
   1.6 Distribution functions ....................................... 8
   1.7 Expectation ................................................... 8

2 Product spaces 12
   2.1 Kernels and Fubini's theorem ................................ 12
   2.2 Countable products .......................................... 13
   2.3 Applications .................................................. 14

3 Conditional Expectation 16
   3.1 Discrete-time martingales .................................... 16
   3.2 Gambling systems ............................................. 18

4 Weak Convergence 19
   4.1 Fourier transforms of probability measures .............. 19

Index 21
1 Measure theoretic framework of probability

1.1 \( \sigma \)-fields

Let \( \Omega \) be a non-empty set. We think of \( \Omega \) as the collection of “outcomes” of an experiment.

1.1.1 Definition. A collection of subsets \( \mathcal{F} \) of \( \Omega \) is called a \( \sigma \)-field if

(i) \( \emptyset, \Omega \in \mathcal{F} \);

(ii) \( A \in \mathcal{F} \) implies \( A^c \in \mathcal{F} \) and

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

By DeMorgan’s law, if \( A_1, A_2, \ldots \in \mathcal{F} \) then \( \bigcap_n A_n \in \mathcal{F} \). Hence \( \mathcal{F} \) 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) \( \{ \emptyset, \Omega \} \) is a \( \sigma \)-field.

(ii) The power set of \( \Omega \) is a \( \sigma \)-field.

(iii) For any \( A \subseteq \Omega \), \( \{ A, A^c, \Omega, \emptyset \} \) is a \( \sigma \)-field.

In using probability theory to model reality, \( \sigma \)-fields correspond to the collection of “observable events.”

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

1.1.3 Examples.

(i) Let \( (X, \mathcal{F}) \) be a topological space. The \( \sigma \)-field \( \sigma(\mathcal{F}) \) generated by \( \mathcal{F} \) is the
Borel \( \sigma \)-field, and is denoted \( \mathcal{B}(X) \).

(ii) Let \( \mathcal{Z} \) be a countable partition of \( \Omega \). Then \( \sigma(\mathcal{Z}) \) has the particularly simple form \( \{ \bigcup_j Z_j | \{ Z_j \} \subseteq \mathcal{Z} \} \).

1.1.4 Example. Consider the time evolution of a random system with state space \( (S, \mathcal{S}) \). The correct choice of \( \Omega \) is \( S^I \), where \( I \) is the collection of times under consideration, usually an interval or \( \mathbb{N} \). For each \( i \in I \), let \( X_i : \Omega \to S : \omega \to \omega_i \) 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 \( \mathcal{B}_n \) be the collection of sets of the form \( \{ X_n \in A \} \) for \( A \in \mathcal{S} \), the collection of events observable at time \( n \in \mathbb{N} \). \( \mathcal{B}_n \) is a \( \sigma \)-field since \( \mathcal{S} \) is a \( \sigma \)-field. Let \( \mathcal{B}_{\infty} := \sigma(\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_n) \), and let \( \mathcal{F} := \sigma(\bigcup_{n=1}^{\infty} \mathcal{B}_n) \), the \( \sigma \)-field of “observable events.” \( (\Omega, \mathcal{F}) \) is a measurable space and \( \{ \mathcal{F}_n \} \) is a filtration on this space.
Let $\mathcal{F}_n^* := \sigma(\bigcup_{k\geq n} \mathcal{F}_k)$, the $\sigma$-field of events observable at and after time $n$, and let $\mathcal{F}^* := \cap_{n=1}^{\infty} \mathcal{F}_n^*$, the tail $\sigma$-field. $\mathcal{F}^*$ contains events, for example, of the form $\{X_n \in A \text{ i.o.}\}$ for a fixed $A \in \mathcal{S}$. Indeed, 

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

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

### 1.2 Dynkin systems

**1.2.1 Definition.** A collection of subsets $\mathcal{D}$ of $\Omega$ is called a Dynkin system (or $\lambda$-system) if

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

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

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

A non-empty collection of subsets $\Omega$ is called a $\pi$-system if it is closed under finite intersections.

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

**1.2.2 Proposition.** If $\mathcal{D}$ is both a $\lambda$-system and a $\pi$-system then it is a $\sigma$-field.

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

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

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

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

□

**1.2.3 Theorem (Dynkin’s $\pi$-$\lambda$ Theorem).** Let $\mathcal{B}$ be a $\pi$-system. Then the smallest $\lambda$-system containing $\mathcal{B}$ is equal to $\sigma(\mathcal{B})$.

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

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

This set is a Dynkin system (note $A^c \cap B = B \setminus (A \cap B)$) that contains $\mathcal{B}$ (since $\mathcal{B}$ is closed under finite intersection), so it is equal to $\mathcal{D}$. Consider now, for any $D \in \mathcal{D}$, the set $\{A \in \mathcal{D} \mid A \cap D \in \mathcal{D}\}$. This set is a Dynkin system that contains $\mathcal{B}$ (by the first consideration) so it is equal to $\mathcal{D}$. Therefore $\mathcal{D}$ is closed under finite intersection. □
1.3 Probability measures

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

(i) \(P(\Omega) = 1;\) and

(ii) if \(A_1, A_2, \ldots \in \mathcal{F}\) with \(A_i \cap A_j = \emptyset\) for all \(i \neq j\) then \(P(\bigcup A_i) = \sum P[A_i].\)

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

(i) \(P(\emptyset) = 0;\)

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

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

(iv) if \(A_1, A_2, \ldots \in \mathcal{F}\) then \(P(\bigcup^\infty A_n) \leq \sum P[A_n].\)

1.3.2 Example. If \(Z\) is a countable partition of \(\Omega\)...

1.3.3 Theorem (Monotone Convergence Theorem).

Let \(P\) be a probability measure on \((\Omega, \mathcal{F})\).

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

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

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

\textbf{Proof:}

\[
P\left[ \bigcap_n \bigcup_{k \geq n} A_k \right] = \lim_{n \to \infty} P\left[ \bigcup_{k \geq n} A_k \right] \leq \lim inf_{n \to \infty} \sum_{k \geq n} P[A_k] = 0. \quad \square
\]

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 \(\lim inf_n a_n \leq \lim inf_n b_n\) and \(\lim sup_n a_n \leq \lim sup_n b_n\).

1.3.5 Theorem. Let \(P_1\) and \(P_2\) be two probability measures on a measurable space \((\Omega, \mathcal{F})\). If there is a \(\pi\)-system \(\mathcal{B} \subseteq \mathcal{F}\) such that \(P_1\) and \(P_2\) agree on \(\mathcal{B}\), then they agree on \(\sigma(\mathcal{B})\).

\textbf{Proof:} Let \(\mathcal{D} = \{A \in \mathcal{F} \mid P_1[A] = P_2[A]\}\). Then \(\mathcal{D}\) contains \(\mathcal{B}\) and is a Dynkin system. Indeed, if \(A_1, A_2, \ldots \in \mathcal{D}\) are pairwise disjoint then

\[
P_1[\bigcup A_n] = \sum P_1[A_n] = \sum P_2[A_n] = P_2[\bigcup A_n],
\]

so \(\bigcup A_n \in \mathcal{D}\). Whence \(\mathcal{D}\) contains \(\sigma(\mathcal{B})\) by Dynkin’s \(\pi-\lambda\) Theorem. \(\square\)
1.3.6 Example. Let $S$ be a countable set, $\mathcal{S}$ be the power set of $S$, and $\Omega = S^\infty$. Let $\mathcal{F} = \sigma\{X_k \in A | k \geq 0, A \in \mathcal{S}\}$. A stochastic process is any measure $\mathbb{P}$ on $\Omega$. 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 $\mathcal{F}$.

A sequence of independent experiments with values in $S$, where each experiment has outcome $x$ with probability $\mu(x)$ (so $\mu$ 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 $\mu$ (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 $\mathcal{B}$ be an algebra on $\Omega$ and $\mathbb{P}$ be a normalized, $\sigma$-additive, set function on $\mathcal{B}$. Then there is a unique extension of $\mathbb{P}$ to $\sigma(\mathcal{B})$.

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

1.3.8 Lemma. Let $\mathcal{B}$ be an algebra on $\Omega$ and $\mathbb{P}$ an additive set function on $\mathcal{B}$. Then $\mathbb{P}$ is $\sigma$-additive on $\mathcal{B}$ if and only if “whenever $B_n \in \mathcal{B}$ are such that $B_n \searrow \emptyset$ then $\lim_{n \to \infty} \mathbb{P}(B_n) = 0.”

1.4 Independence

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

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

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

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

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


does not tell us what happens at what happens inside any box of finite size. By Kolmogorov’s 0-1 Law, the monkey will type your favourite book infinitely many times.

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. \( P[A_i] > 0 \), 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 \( (\mathcal{F}_n)_{n \in \mathbb{N}} \) be a countable collection of independent \( \sigma \)-fields. Define

\[
\mathcal{F}^* := \bigcap_{n \geq 1} \sigma(\bigcup_{k \geq n} \mathcal{F}_k),
\]

the tail \( \sigma \)-field. Then \( P[A] = 0 \) or 1 for all \( A \in \mathcal{F}^* \).

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

\[
\]

Indeed, \( \{\mathcal{F}_1, \ldots, \mathcal{F}_n, \sigma(\bigcup_{k \geq n} \mathcal{F}_k)\} \) are independent and \( \mathcal{F}^* \subseteq \sigma(\bigcup_{k \geq n} \mathcal{F}_k) \), so \( \{\mathcal{F}_1, \ldots, \mathcal{F}_n, \mathcal{F}^*\} \) are independent. It follows that \( \{\mathcal{F}^*, \mathcal{F}_1, \mathcal{F}_2, \ldots\} \) are independent, so \( \mathcal{F}^* \) and \( \mathcal{F}_\infty \) are independent.

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

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

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

\[
P[X_{n_1}, \ldots, X_{n_k} = 1, X_{y_1}, \ldots, X_{y_s} = 0] = p^k(1-p)^s.
\]

Then \( (\mathcal{F}_k)_{k \geq 0} \) is an independent family of \( \sigma \)-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 \( A \in \mathcal{F}^* = \bigcap_{n \geq 0} \sigma(\bigcup_{k \geq n} \mathcal{F}_k) \) since it doesn’t matter what happens inside any box of finite size. By Kolmogorov’s 0-1 Law, \( P[A] = 0 \) 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 $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces and let $X : \Omega' \rightarrow \Omega$ be a function. Define

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

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

1.5.2 Exercises.

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

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

(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 $X : \Omega \rightarrow \mathbb{R}$, we say that $X$ is $\mathcal{F}$-measurable if $X$ is $\mathcal{F}/\mathcal{B}(\mathbb{R})$-measurable.

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

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

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

$$\{ T < c \} = \bigcup_{n: \pi_n = 1} \left\{ X_n = 0 \} \cap \bigcap_{k < n} \{ X_k = c_k \} \right\} \in \mathcal{F},$$

so $T$ is measurable. If we define $P[X_0 = x_0, \ldots, X_k = x_k] = 2^{-k}$ then it turns out that $P[T < c] = c$, 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 $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, $X : \Omega' \rightarrow \Omega$ a measurable function, and $P'$ a probability measure on $(\Omega', \mathcal{F}')$. Then $P_X[A] := P'[X \in A]$ is a probability measure on $(\Omega, \mathcal{F})$, called the image measure of $P'$ under $X$, or the distribution of $X$ with respect to $P'$, or the law of $X$, and is written $P' \circ X^{-1}$.
1.6 Distribution functions

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

1.6.2 Proposition. The probability measure \( \mu \) is uniquely determined by \( F_\mu \), and \( F_\mu \) 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) \( F_\mu \) is increasing; and

(iii) \( F_\mu \) 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 \( \mu_F((-\infty, x]) := \mu \). It can be shown that \( \mu_F \) is a probability measure on \( \mathbb{R} \), and \( F_\mu = F \). Indeed, set \( G(y) := \inf \{ x \mid F(x) > y \} \), the unique right continuous “inverse” of \( F \). Then

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

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. \( X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) (where \( \mathcal{B}(\mathbb{R}) \) is generated by the collection \( \{ [-\infty, c] \mid c \in \mathbb{R} \} \) ) is called a random variable if it is \( \mathcal{F} / \mathcal{B}(\mathbb{R}) \)-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, \( \sum_{i=1}^{n} c_i 1_{A_i} \). 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.

\[
X_n := \sum_{k=0}^{n^2 - 1} \frac{k}{2^n} 1_{k2^n \leq X < (k+1)2^n} + n 1_{n \leq X}.
\]

\[ \square \]

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

Proof (Measure-theoretic induction):
(i) If $X = 1_A$ for some $A \in \sigma(T)$ then $A = \{ T \in B \}$ for some $B \in \mathcal{F}$, so take $X' = 1_B$.

(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) $E[1_A] := P[A]$;

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

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


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 $(\Omega, \mathcal{F}, P)$. Then
$$E[X] := \int_{\Omega} X dP = \int_{\Omega} X(\omega) P(d\omega).$$

$L^1(\Omega, \mathcal{F}, P)$ denotes the collection of integrable r.v.’s, i.e. those $X$ for which $E[X^+] < \infty$ and $E[X^-] < \infty$. 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 $X_0 \in L^1(\Omega, \mathcal{F}, P)$ and $X_0 \leq X_1 \leq \cdots$ $P$-a.s. then
$$E[\lim_{n \to \infty} X_n] = \lim_{n \to \infty} E[X_n].$$

This is also called the Beppo-Levi Theorem. That $X_0 \in L^1$ is necessary. For a counterexample, take $\Omega = (0, 1]$ and $X_0 := -\frac{1}{x}$ and $X_n := X_0 1_{(0,1/n]}$. Then $X_n \not\nearrow 0$ but $E[X_n] = -\infty$ 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
$$-\infty < E[\liminf_{n \to \infty} X_n] \leq \liminf_{n \to \infty} E[X_n].$$
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,
\[
-\infty \leq \mathbb{E}[Y] \leq \mathbb{E}\left(\liminf_{n \to \infty} X_n\right) = \lim_{n \to \infty} \mathbb{E}\left(\inf_{k \geq n} X_n\right) \leq \limsup_n \mathbb{E}[X_n]. \quad \square
\]

**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,
\[
\mathbb{E}[X] = \mathbb{E}[\limsup_n X_n] \geq \limsup_n \mathbb{E}[X_n] \geq \liminf_n \mathbb{E}[X_n] = \mathbb{E}[X].
\]
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 \( \varphi : \mathbb{R} \to \mathbb{R} \) be a non-negative measurable function. Let \( A \in \mathcal{B}(\mathbb{R}) \) 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 \( X \in \mathcal{L}^1 \) 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 \( \sigma \)-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
\[
\mathbb{E}[u(X)] \geq \mathbb{E}[\ell(X)] = \ell(\mathbb{E}[X]) = u(\mathbb{E}[X]). \quad \square
\]
1.7.11 Example. For \( p \geq 1 \), let \( \mathcal{L}^p \) 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

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

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

1.7.12 Theorem. Let \( T : (\Omega', \mathcal{F}') \rightarrow (\Omega, \mathcal{F}) \) be measurable and \( \mathbb{P}' \) be a measure on \( (\Omega', \mathcal{F}') \) and \( \mathbb{P} \) the induced measure on \( (\Omega, \mathcal{F}) \). Then for any non-negative r.v. \( X \) on \( (\Omega, \mathcal{F}) \),

\[
\mathbb{E}[X] = \mathbb{E}'[X \circ T].
\]

Proof: By measure-theoretic induction. \( \square \)

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 \( (X, Y) : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^2 \),

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

\( \mu \) is the induced measure on \( \mathbb{R}^2 \), the joint distribution of \( X \) and \( Y \).

\[
\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]
\]

\[
= \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]
\]

\[
= \int_{\mathbb{R}^2} xy \mu(dx, dy) - \int_{\mathbb{R}} x \mathbb{P}[X \in dx] \int_{\mathbb{R}} y \mathbb{P}[Y \in dy]
\]

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

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

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

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

1.7.17 Theorem. Random variables \(X\) and \(Y\) are independent if and only if
\[
\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \mathbb{E}[g(Y)] \quad \text{for all } f, g \geq 0 \text{ measurable.}
\]
In this case, if \(X, Y \in L^1\) then \(XY \in L^1\).

2 Product spaces

2.1 Kernels and Fubini’s theorem

Let \((S_1, \mathcal{S}_1)\) be measurable spaces and \(\Omega = S_1 \times S_2\) and \(X_i : \Omega \rightarrow S_i\) denote projection on the \(i\)th coordinate.

2.1.1 Definition. A stochastic kernel from \(S_1\) to \(S_2\) is a map \(K : S_1 \times \mathcal{S}_2 \rightarrow [0, 1]\) 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 \(\mathcal{S}_1\)-measurable for all \(A \in \mathcal{S}_2\).

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 \(T : S_1 \rightarrow S_2\) 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 \(\sigma\)-field, and take \(K(x, \{y\}) = q_{xy}\) to be a (countable) matrix with \(q_{xy} \geq 0\) 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 \(\sigma\)-field, and \(K(x, \cdot) = \mathcal{N}(0, \beta x^2)\). 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 \((S_1, \mathcal{S}_1)\) and a stochastic kernel \(K\) from \(S_1\) to \(S_2\), we would like to construct a probability measure \(\mathbb{P}\) on the product space such that
\[
\mathbb{P}[X_1 \in A_1] = \mathbb{P}_1[A_1] \quad \text{and} \quad \mathbb{P}[X_2 \in A_2 \mid X_1 = x_1] = K(x_1, A_2)
\]
2.1.3 Theorem (Fubini). Let $\Omega := S_1 \times S_2$ and $\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. Given a probability measure $P_1$ on $(S_1, \mathcal{F}_1)$ and a stochastic kernel $K$ from $S_1$ to $S_2$, there is a unique probability measure $P$ on $\Omega$ such that for all $f \in m^+ F$,

$$\int_{\Omega} f \, dP := \int_{S_1} \left( \int_{S_2} f(x_1, x_2) K(x_1, dx_2) \right) \, P_1(dx_1).$$

In particular,

(i) $P[A_1 \times A_2] = \int_{A_1} P_1(dx_1) K(x_1, A_2)$.

(ii) $P[A] = \int_{S_1} P_1(dx_1) K(x_1, A_{x_1})$ 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 $P := P_1 \times K$.

**Proof:** Make notes on the proof of this. □

2.1.4 Corollary. If $K(x, \cdot) = P_2[\cdot]$ has no coupling then $P = P_1 \times P_2$ and we recover the classical theorem of Fubini.

Remark. Fubini’s theorem is true for $\sigma$-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.$X$,

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

$$= \int_{\Omega} P(d\omega) \int_0^\infty 1_{[0, X(\omega)]}(s) \, ds$$

$$= \int_0^\infty ds \left( \int_{\Omega} 1_{(s, \infty]}(X(\omega)) \, P(d\omega) \right)$$

$$= \int_0^\infty P[X > s] \, ds$$

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

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

2.2 Countable products

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

$$(S^n, \mathcal{F}^n) := (S_0 \times \cdots \times S_n, \mathcal{F}_0 \otimes \cdots \otimes \mathcal{F}_n).$$
Let \( \mu_0 \) 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 \( f \in m^+ S^n \),

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

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

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

PROOF: Since the \( \mathcal{F}_n \)'s are increasing, \( \mathcal{A} := \bigcup_{n \geq 0} \mathcal{F}_n \) is an algebra. Define \( \mu \) on \( \mathcal{A} \) by

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

for \( A^n \in \mathcal{A} \). Now \( \mu \) is well-defined since if \( A \in \mathcal{F}_n \cap \mathcal{F}_{n-1} \) and \( A = A^n \times S_{n+1} \times \cdots = A^{n-1} \times S_{n} \times \cdots \) then

\[
\mu(A) = \mu^n(A^n) = \mu^n(A^{n-1} \times S_n)
\]

\[
= \int_{S^{n-1}} \mu^{n-1}(d\hat{x}) \int_{S_n} K(\hat{x}, dx_n) 1_{A^n}(\hat{x}) 1_{S^n}(x_n)
\]

\[
= \mu^{n-1}(A^{n-1}),
\]

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

Suppose that \( A_n \in \mathcal{A} \) are such that \( A_n \searrow \emptyset \). Without loss of generality, suppose that \( A_n \in \mathcal{A}_n \) 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. \( \square \)

2.3 Applications

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

PROOF: Recall that the variance of \( X_i \), variance is \( \mathbb{E}[(X_i - \mu)^2] = \sigma^2 \).

\[
\mathbb{E}[(\bar{X}_n - \mu)^2] = \mathbb{E} \left[ \frac{1}{n^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right]
\]

by independence

\[
= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \to 0
\]
so \( X_n \to 0 \) in \( \mathcal{L}^2 \). That the convergence is also in probability follows from the next lemma.

2.3.2 Lemma. For any \( p > 0 \), convergence in \( \mathcal{L}^p \) implies convergence in probability.

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

2.3.3 Theorem (Strong Law of Large Numbers).

Let \( X_1, X_2, \ldots \) be a sequence of pairwise independent identically distributed r.v.’s with \( \mu := \mathbb{E}X_1 < \infty \). Then \( \bar{X}_n \to \mu, \mathbb{P}\text{-a.s.} \)

**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 \( X_1, X_2, \ldots \) be a sequence of i.i.d. r.v.’s with \( \mathbb{E}[X_1^+] = +\infty \) and \( \mathbb{E}[X_1^-] < \infty \). Then \( \bar{X}_n \to +\infty, \mathbb{P}\text{-a.s.} \)

2.3.6 Example (Renewals). Let \( X_1, X_2, \ldots \) be i.i.d. with \( 0 < X_i < \infty, \mathbb{P}\text{-a.s.} \), 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 \( \frac{1}{N_t} \to \frac{1}{\mu}, \mathbb{P}\text{-a.s.} \). Indeed, by the SLIN, \( \frac{1}{n} T_n \to \mu, \mathbb{P}\text{-a.s.} \). 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 \( t \to \infty \),

\[
\frac{\mu}{N_t} \leq \frac{T_{N_t}}{N_t} < \frac{T_{N_t+1}}{N_t} \to \mu, \mathbb{P}\text{-a.s. as } t \to \infty.
\]
2.3.7 Example (Monte-Carlo Integration). How do we approximate an integral of the form
\[
\int_{[0,1]} \cdots \int_{[0,1]} \varphi(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]
when \( \varphi \) may be complicated to deal with? Assume that \( X_1, \ldots, X_n \) are i.i.d. uniform on \([0,1]\). Then \( \vec{X} \) has the distribution of \( \lambda_n \) on \([0,1]^n\), so
\[
E[\varphi(\vec{X})] = \int_{[0,1]^n} \varphi(x_1, \ldots, x_n) \lambda_n(dx).
\]
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 \( \mathcal{F}_0 \subseteq \mathcal{F} \) and \( X \in m\mathcal{F} \). An \( \tilde{X} \in m\mathcal{F}_0 \) such that
\[
E[X 1_{A_0}] = E[X \tilde{X} 1_{A_0}]
\]
for all \( A_0 \in \mathcal{F}_0 \) is (a version of) the conditional expectation of \( X \) given \( \mathcal{F}_0 \), and is denoted \( E[X | \mathcal{F}_0] \).

3.0.9 Theorem. Let \( \mathcal{F}_0 \subseteq \mathcal{F} \) and \( X \in m\mathcal{F} \). If \( X \in L^1 \) or \( X \geq 0 \) then \( E[X | \mathcal{F}_0] \) exists and is unique.

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

3.1 Discrete-time martingales

Skipped the definitions of filtrations, martingales.

3.1.1 Definition. \( \mathcal{H} \subseteq L^1 \) is said to be uniformly integrable (or u.i.) if
\[
\lim_{c \to \infty} \sup_{X \in \mathcal{H}} \int_{\{|X| > c\}} |X| \, d\mathbb{P} = 0
\]

3.1.2 Examples.
(i) If \( X \in L^1 \) 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 \( 1_{|X| \leq c} |X| \to |X| \) as \( c \to \infty \).

(ii) If there is \( Y \in L^1 \) such that \( |X| \leq Y \) for all \( X \in \mathcal{H} \) then \( \mathcal{H} \) is u.i. To show this we need another characterization of uniform integrability.
3.1.3 Lemma. If there is \( g : \mathbb{R} \rightarrow \mathbb{R} \) with \( \frac{g(x)}{x} \rightarrow \infty \) as \( x \rightarrow \infty \) such that
\[
\sup_{\mathcal{H}} \int g(|X|) \, d\mathbb{P} < \infty
\]
then \( \mathcal{H} \) is u.i.

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

3.1.4 Theorem. \( f_n \rightarrow f \) in \( L^1 \) if and only if the set \( \{f_n\} \) is u.i. and \( f_n \rightarrow f \) in probability.

In the particular case where \( f_n \rightarrow f \) pointwise (a stronger form of convergence then in probability) and there is \( g \in L^1 \) with \( |f_n| \leq g \) 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 \( L^1 \).

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
\[
T := \inf\{n \geq 0 \mid X_n \notin (a, b)\}.
\]
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
\[
\{n(b - a), n(b - a) + 1, \ldots, (n + 1)(b - a) - 1\}.
\]
Then \( \mathbb{P}[A_n] \geq p^{b-a} > 0 \), the latter being the probability of the path consisting of \((b - a)\) up-steps. It follows that \( \sum_{n \geq 0} \mathbb{P}[A_n] = \infty \), and since these events are independent, \( \mathbb{P}[A_n \text{ i.o.}] = 1 \) by the second Borel-Cantelli Lemma. In particular
\[
\mathbb{P}[T < \infty] = \mathbb{P}\left[\bigcup_{n \geq 0} A_n \right] = \mathbb{P}[A_n \text{ i.o.}] = 1
\]
The probability of ruin is \( r(x) := \mathbb{P}[X_T = a] \). When \( p = \frac{1}{2} \) \( X \) itself is a martingale, so
\[
x = \mathbb{E}[X_0] = \mathbb{E}[X_{T \wedge n}] \rightarrow \mathbb{E}[X_T] = b(1 - r(x)) + ar(x),
\]
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 \( p \neq \frac{1}{2} \), let \( h(z) = (\frac{1-p}{p})^z \), and notice that \( h(X_n) \) is a martingale. Indeed,
\[
\mathbb{E}[h(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[((\frac{1-p}{p})^{X_n}Y_{n+1}) | \mathcal{F}_n] = (\frac{1-p}{p})^{X_n}((\frac{1-p}{p})^1p + (\frac{1-p}{p})^{-1}(1-p)) = h(X_n)
\]
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
\[ r(x) = \frac{1 - (\frac{p}{1-p})^{b-x}}{1 - (\frac{p}{1-p})^{b-x} - a}. \]
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 \(\mathcal{F}_{n-1}\)-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 \((V \cdot X)_n = \sum_{i=1}^{n} V_i Y_i\) 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
\[
T := \inf\{n \geq 1 \mid Y_{n-N+1} = a_1, \ldots, Y_n = a_N\}.
\]
By the second Borel-Cantelli Lemma \(P[T < \infty] = 1\) (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 \(T \in \mathcal{L}^1\).) 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 \(\mu\) (not deterministic), with \(\mathbb{E}[Y_{1,1}] = m < \infty\). Let \(X_0 := 1\) and
\[
X_n := Y_{n,1} + \cdots + Y_{n,N}.
\]
\(X\) 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 \(\sigma\) be \(\sigma(Y_{\ell,k}, 1 \leq \ell \leq n, k \geq 1)\) and \(M_n = \frac{X_n}{m^n}\). Then \(M_n\) is a martingale for this filtration, bounded in \(\mathcal{L}^1\).

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 $h(X^T_S)$ is a super-martingale, so

$$h(x) = \mathbb{E}^x[h(X^T_0)] \geq \mathbb{E}^x[h(X^T_T)] \geq \mathbb{E}^x[g(X^T_T)]$$

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

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 $\mu$.

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 $\varphi_\epsilon(x) := \frac{1}{\epsilon \sqrt{2\pi}} e^{-\frac{x^2}{2\epsilon^2}}$, the density of a $\mathcal{N}(0, \epsilon^2)$ distribution with respect to Lebesgue measure. Then $\varphi_\epsilon(u) = e^{i\epsilon u^2}$.

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

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

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


Index

λ-system, 3
$\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, 9
$k^{th}$ moment, 11
π-system, 3
σ-field, 2
Borel σ-field, 2
conditional expectation, 16
cumulative distribution function, 8
discrete stochastic integral, 18
distribution, 7
distribution function, 8
Dynkin system, 3
elementary function, 8
expectation, 9
filtration, 2
Fourier transform, 19
gambling system, 18
image measure, 7
independent, 5
independent experiments, 5
initial distribution, 5
integrable, 9
joint distribution, 11
law, 7
Markov chain, 5
measurable, 7
measurable space, 2
previsible process, 18
probability measure, 4
random variable, 8
semi-integrable, 9
state space, 2
step function, 8

stochastic kernel, 12
stochastic process, 5
tail σ-field, 3, 6
time evolution, 2
transition kernel, 5
u.i., 16
uniformly integrable, 16