SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Measure-theoretic Probability

av

Jasper Lindblom

2021 - No K45

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Measure-theoretic Probability

Jasper Lindblom

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Sven Raum

2021

Measure-theoretic Probability

Jasper Lindblom

October 2021

Abstract

This thesis explores the deep connections of measure theory and probability theory. We introduce fundamental results of measure theory. We then use these results to rigorously develop probability theory. Special attention is given to independence and conditional expectation of random variables. The thesis aims to be self-contained to a high degree and most results are given detailed proofs.

Contents

1 Introduction

There are many motivations for letting measure theory provide the foundation for probability theory which the ardent student *almost surely* would discover on his or her own. But as a concrete example, conditional expectation is more easily treated. This is indispensable for the development of martingales. These are important in the theory of stochastic process and useful, for example, in financial modelling. Martingales are also interesting in their own right and can be used to prove some important theorems in real analysis. See, in particular, [5] and [7] for that kind of development.

However, we do without martingales in this thesis and instead focus on more fundamental concepts. In Section 2, we cover σ-algebras, measures, measurable functions, the Lebesgue integral and the main convergence theorems. In Section 3, we explore connections between real analysis and probability theory. We prove Weierstrass approximation theorem and Jensen's inequality in the process. In Section 4, we thoroughly develop the notion of independence between σ-algebras. The section ends with proofs of fundamental theorems, namely, Kolmogorov's 0-1 law and the two versions of Borel-Cantelli's lemma. Section 5 is largely devoted to proving useful properties of the conditional expectation with regards to a σ-algebra. In the concluding section suggestions for further reading are given.

The prerequisites are a basic understanding of probability theory and real analysis. Some familiarity with measure theory is helpful too but not necessary.

2 Measure Theory Preliminaries

For a quick probabilistic motivation of our first concept, consider an experiment consisting of flipping a coin until we get heads. Let A_n denote the event of getting heads after n throws. Then $A \coloneqq \cup_{n=1}^{\infty} A_{2n}$ is the event of getting heads after an *even* number of throws. From undergraduate theory, we would expect the probabilities $\mathbb{P}(A_n)$, $\mathbb{P}(A)$, $\mathbb{P}(A^c)$, and $\mathbb{P}(A \cup A^c)$ to all be well-defined. These well-defined sets are in a sense what makes up a σ-algebra.

Definition 2.1. Let S be a set. A collection $\mathcal{F}$ of subsets of S is a *σ-algebra* if

(i) $S \in \mathcal{F}$,

(ii) $A \in \mathcal{F} \implies S \setminus A \in \mathcal{F}$,

(iii) $A_1, A_2, \dots \in \mathcal{F} \implies \displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Since $\emptyset = S^c$, the σ-algebra contains the empty set. Furthermore, by De Morgan's laws, we have

$$A_1, A_2, \dots \in \mathcal{F} \implies \bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{F}.$$

The sets of $\mathcal{F}$ are called *measurable*. From the definition, it is evident that both $\{\emptyset, S\}$ and 2^S are σ-algebras on S. The second example shows that for any collection of sets, there exists a σ-algebra containing the collection.

It is also straightforward to verify that any intersection of σ-algebras on S is a σ-algebra. Thus the smallest σ-algebra containing a collection of sets $\mathcal{A}$ coincides with intersection of all σ-algebras containing $\mathcal{A}$. This σ-algebra is said to be *generated* by $\mathcal{A}$. It is denoted by $\sigma(\mathcal{A})$.

The *Borel* σ-algebra $\mathcal{B}(S)$ is the σ-algebra on a topological space S *generated* by the open sets of S. We often work with the Borel σ-algebra on the extended real line $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{-\infty, +\infty\}$ equipped with the topological basis consisting of all open intervals (a, b) along with $(a, \infty]$ and $[-\infty, b)$ where $a, b \in \mathbb{R}$. We will refer to this particular Borel set by $\mathcal{B}$.

Definition 2.2. Let $\mathcal{F}$ be a σ-algebra. A function $\mu : \mathcal{F} \to [0, \infty]$ is a *measure* if $\mu(\emptyset) = 0$ and

$$A_1, A_2, \dots \in \mathcal{F}, \text{ and } A_i \cap A_j = \emptyset \text{ for } i \neq j \implies \mu\left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Remark. The first condition is merely to avoid the trivial case when $\mu = \infty$. Otherwise, it is implied by the second condition. Also, using the notation above, if we let $A_n = \emptyset$ for $n \geq 3$, then $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$.

Consider the function $\mu : 2^S \to \mathbb{N}_0 \cup \{\infty\}$ which for $A \subset S$,

$$\mu(A) = \begin{cases} |A|, & \text{if } A \text{ finite}, \\ \infty, & \text{if } A \text{ infinite}, \end{cases}$$

where $|A|$ denotes the number of elements in A. This functions is a measure. It is called the *counting measure*. Another example of a measure is the Lebesgue measure which is defined on a σ-algebra larger than the Borel set of $\mathbb{R}$. It essentially maps intervals $[a, b)$ to $b - a$. Consult the literature, notably [1], [2] or [7], for elaboration on the Lebesgue measure.

Proposition 2.3. *If $A_1 \subset A_2 \subset \ldots \mathcal{F}$ and $A \coloneqq \lim_{n\to\infty} \cup_{n=1}^{\infty} A_n$, then for any measure μ we have*

$$\lim_{n\to\infty} \mu(A_n) = \mu(A).$$

Remark. Other notations for A are $\lim_{n\to\infty} A_n = A$ and $A_n \uparrow A$ as $n \to \infty$.

Proof. Let $B_n = \cup_{k=1}^{n} A_k \setminus \cup_{k=1}^{n-1} A_k$, then $B_i \cap B_j = \emptyset$ for $i \neq j$ and $A_n = \cup_{k=1}^{n} B_k$, thus

$$\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_n).$$

□

Definition 2.4. If $\mathcal{F}$ is a σ-algebra on S and μ is a measure on $\mathcal{F}$ then the triple $(S, \mathcal{F}, \mu)$ is a *measure space*. If we do not specify the measure, we say that $(S, \mathcal{F})$ is *measurable space*. If f is a function from one measurable space to another and the inverse image of every measurable set is also measurable, then f is a *measurable* function.

We will for the most part restrict ourselves to extended real-valued measurable functions. These will thus take the form of $f : (S, \mathcal{F}) \to (\overline{\mathbb{R}}, \mathcal{B})$ where $B \in \mathcal{B}$ implies $f^{-1}(B) \in \mathcal{F}$.

Lemma 2.5. *Suppose $f : X \to Y$ is a measurable function and $g : Y \to Z$ is continuous, where Y and Z are both topological spaces equipped with their Borel σ-algebra. Then the composition $g \circ f : X \to Z$ is measurable.*

Proof. Let B be open in Z, then $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B))$. By the definition of a continuous function, $g^{-1}(B)$ is open in Y. Since $g^{-1}(B)$ is a Borel set of Y, by the measurability of f, we have that $f^{-1}(g^{-1}(B))$ belongs to the σ-algebra of X. If B is a Borel set of Z which is not open, then it is still in the σ-algebra generated by open sets and is thus equal to a combination of countable intersections and unions of open sets. Now, the inverse image is closed w.r.t. those operations in the sense that for a countable sequence of Borel sets $(B_n)_{n=1}^{\infty}$, we have $\cup_{n=1}^{\infty} f^{-1}(B_n) = f^{-1}(\cup_{n=1}^{\infty} B_n)$ and likewise for intersections. Thus the initial argument holds for non-open B too. □

The last argument of the proof above yields the next lemma.

Lemma 2.6. *A function f on $(S,\mathcal{F})$ is measurable if and only if $f^{-1}((a,b)),\ f^{-1}((a,+\infty]),\ f^{-1}([-\infty,b))\in\mathcal{F}$ for all $a,b\in\mathbb{R}$.*

Lemma 2.7. *A function f on $(S,\mathcal{F})$ is measurable if and only if $f^{-1}([-\infty,b))\in\mathcal{F}$, for every $b\in\mathbb{R}$.*

Proof. For the "if"-part, fix $a,b\in\mathbb{R}$ such that $a<b$. Then

$$f^{-1}((a,+\infty]) = \bigcup_{n=1}^{\infty} f^{-1}([a+\frac{1}{n},+\infty]) = \bigcup_{n=1}^{\infty} f^{-1}([-\infty,a+\frac{1}{n}))^c,$$
$$f^{-1}((a,b)) = f^{-1}([-\infty,b))\cap f^{-1}((a,+\infty]).$$

Since a,b were arbitrarily chosen, this is sufficient. □

Theorem 2.8. *Let u and v be measurable functions on $(S,\mathcal{F})$, let $\Phi:\overline{\mathbb{R}}\times\overline{\mathbb{R}}\to\overline{\mathbb{R}}$ be continuous, and define $h(x):=\Phi(u(x),v(x))$ for $x\in S$. Then $h:S\to\overline{\mathbb{R}}$ is measurable.*

Remark. As usual, $\overline{\mathbb{R}}\times\overline{\mathbb{R}}$ is assigned the product topology. See [3] for more on topology. We let $\infty-\infty$ be undefined but $\infty\cdot 0:=0$. Thus this result implies the measurability of (well-defined) sums and products of measurable functions.

Proof. Define $f(x):=(u(x),v(x))$, then $h(x)=(\Phi\circ f)(x)$. Thus by Lemma 2.5 it suffices to prove the measurability of f. For an open rectangle $I=I_1\times I_2\in\overline{\mathbb{R}}\times\overline{\mathbb{R}}$, we have $f^{-1}(I)=u^{-1}(I_1)\cap v^{-1}(I_2)\in\mathcal{F}$. We are now done since any open subset of $\overline{\mathbb{R}}\times\overline{\mathbb{R}}$ can be written as the countable union of open rectangles. □

Proposition 2.9. *Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions, then $\sup\limits_{n\ge1} f_n,\ \inf\limits_{n\ge1} f_n,\ \limsup\limits_{n\to\infty} f_n$ and $\liminf\limits_{n\to\infty} f_n$ are measurable.*

Proof. Denote $\inf\limits_{n\ge1} f_n$ by f. Note that for every $b\in\mathbb{R},\ f(x)<b$ implies $f_n(x)<b$ for some n. By using the previous lemma, it is sufficient to note that

$$f^{-1}([-\infty,b)) = \bigcup_{n=1}^{\infty} f_n^{-1}([-\infty,b)).$$

Consequently,

$$\sup_{n\ge1} f_n = -\inf_{n\ge1}(-f_n),$$
$$\limsup_{n\to\infty} f_n = \inf_{m\ge1}\left(\sup_{n\ge m} f_n\right),$$
$$\liminf_{n\to\infty} f_n = \sup_{m\ge1}\left(\inf_{n\ge m} f_n\right),$$

all follow. □

Remark. In particular, for a measurable function f we have that $f^+(x) := \max\{f(x), 0\}$ and $f^-(x) := \max\{-f(x), 0\}$ are measurable.

The *indicator function* of a subset $E \subset S$ is defined as

$$I_E(x) = \begin{cases} 1, & \text{if } x \in E, \\ 0, & \text{if } x \notin E. \end{cases}$$

Proposition 2.10. *The indicator function I_E is measurable if and only if E is measurable.*

Proof. We have

$$I_E^{-1}((a,b)) = \begin{cases} \emptyset, & \text{if } a \geq 1 \text{ or } b \leq 0, \\ E^c, & \text{if } a < 0 < b \leq 1, \\ E, & \text{if } 0 \leq a < 1 < b, \\ S, & \text{if } a < 0 < 1 < b. \end{cases}$$

Hence the function is measurable if and only if these four sets are measurable. Since $\sigma(E) = \{\emptyset, E, E^c, S\}$, we are done. □

A *simple* function is a real function with finite range. An indicator function is trivially a simple function, but $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ is not, because the range is an uncountable set. In general, a simple function s can be written as

$$s(x) = \sum_{i=1}^{n} \alpha_i I_{A_i}(x),$$

where each α_i is distinct, and $\{A_1, \ldots A_n\}$ is a partition of the domain. It is evident that s is measurable only if each A_i is measurable.

The next proposition gives a good characterization of the relationship between simple functions and measurable functions.

Proposition 2.11. *Let $f : S \to [0, \infty]$ be a measurable function. Then there exists a nondecreasing sequence of nonnegative measurable simple functions, $(s_n)_{n=1}^{\infty}$, that converges pointwise to f.*

Proof. Let

$$s_n(x) := \sum_{i=0}^{n2^n} i2^{-n} I_{A_{n,i}}(x),$$

where $A_{n,i} := \{x : i2^{-n} \leq f(x) < (i+1)2^{-n}\} = A_{n+1,2i} \cup A_{n+1,2i+1}$. □

Definition 2.12. Let E be a measurable subset and

$$s(x) = \sum_{i=1}^{n} \alpha_i I_{A_i}(x),$$

a nonnegative measurable simple function. Then we define

$$\int_E s\, d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E),$$

where $0 \cdot \infty := 0$ as noted before. If f a nonnegative measurable function, then the *Lebesgue integral* of f over E μ of f is defined as

$$\int_E f\, d\mu := \sup \int_E s\, d\mu.$$

where supremum is taken over all measurable simple functions $0 \leq s \leq f$.

If f is allowed to be negative and $\int_S |f|\, d\mu < \infty$, then it is *integrable*, or μ*-integrable.* Its integral is defined as

$$\int_E f\, d\mu := \int_E f^+\, d\mu - \int_E f^-\, d\mu.$$

We denote it by $f \in \mathcal{L}^1(S, \mathcal{F}, \mu)$, or just $f \in \mathcal{L}^1$. More generally, $f \in \mathcal{L}^p$ for $p \in [0, \infty)$ if and only if

$$\|f\|_p := \left(\int_S |f|^p\, d\mu \right)^p < \infty.$$

Here are some elementary properties of the Lebesgue integral. We omit a proof.

Proposition 2.13. *If $f, g, h \in \mathcal{L}^1(S, \mathcal{F}, \mu)$ with $f \leq h$ and $a, b \in \mathbb{R}$, then $af + bg \in \mathcal{L}^1$ and*

(a) $\displaystyle\int_S (af + bg)\, d\mu = a \int_S f\, d\mu + b \int_S g\, d\mu,$

(b) $\displaystyle\int_S f\, d\mu \leq \int_S h\, d\mu,$

(c) $\displaystyle\left| \int_S f\, d\mu \right| \leq \int_S |f|\, d\mu.$

We have now arrived to one of our main results in this section.

Theorem 2.14 (Monotone convergence theorem)**.** *Suppose $(f_n)_{n=1}^{\infty}$ is an non-decreasing sequence of nonnegative measurable functions on S which converges pointwise to a function f, then f is measurable and*

$$\lim_{n\to\infty} \int_S f_n\, d\mu = \int_S f\, d\mu.$$

Proof. Since $f = \sup\limits_{n\geq 1} f_n$, it follows that f is measurable. Thus the integral makes sense and we have

$$\int_S f_n\, d\mu \leq \int_S f\, d\mu, \quad \forall n. \tag{1}$$

The limit of the left side of (1) exists because of monotonicity. If the limit is infinite the equality is trivially true. We can thus assume that

$$\lim_{n\to\infty} \int_S f_n = \alpha \in \mathbb{R}.$$

Now, (1) implies, by taking the limit of the left side,

$$\alpha \le \int_S f \, d\mu. \tag{2}$$

For the reverse inequality, take any measurable simple function $0 \le s \le f$, and any $c \in (0,1)$. Consider the set $E_n = \{x \in S : f_n(x) \ge cs(x)\}$. Clearly $E_n \subset E_{n+1}$, hence $\bigcup_{n=1}^{N} E_n = E_N$. Furthermore, $E_n \uparrow S$ as $n \to \infty$. The latter assertion needs a clarification. Take $x \in S$. If $s(x) = 0$, then $x \in E_1$. If $s(x) > 0$, then $f(x) - cs(x) = \varepsilon > 0$. By pointwise convergence $f(x) - f_n(x) < \varepsilon$ for $n \ge N$ for some N. This implies $x \in E_N$. Further,

$$\int_S f_n \, d\mu \ge \int_{E_n} f_n \, d\mu \ge c \int_{E_n} s \, d\mu.$$

This implies

$$\alpha \ge c \int_{E_n} s \, d\mu,$$

and by letting $n \to \infty$, we get

$$\alpha \ge c \int_S s \, d\mu.$$

Since the inequality holds for any $c \in (0,1)$, it must also hold for $c = 1$, resulting in

$$\alpha \ge \int_S s \, d\mu.$$

Finally, since s is arbitrary we can invoke the definition of the Lebesgue integral, and thus

$$\alpha \ge \int_S f \, d\mu,$$

which coupled with (2) yields the desired equality. □

Remark. The condition $f_1 \ge 0$ can be replaced by $\int_S f_1^- < \infty$, i.e. finite integral of the negative part of f_1. This is seen by applying the theorem on $f_n + f_1^-$, which *is* nonnegative, and then subtract.

Theorem 2.15 (Fatou's lemma)**.** *Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of nonnegative measurable functions, then*

$$\int_S \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int_S f_n \, d\mu.$$

Proof. We have $\liminf_{n\to\infty} f_n(x) = \sup_{m\to\infty}\left(\inf_{n\geq m} f_n(x)\right)$, for every $x \in S$, that is, $\inf_{n\geq m} f_n(x) \uparrow \liminf_{n\to\infty} f_n(x)$ as $m \to \infty$. Now,

$$\int_S \inf_{n\geq m} f_n \, d\mu \leq \int_S f_k \, d\mu$$

holds for any $k \geq m$. Thus

$$\int_S \inf_{n\geq m} f_n \, d\mu \leq \liminf_{n\to\infty} \int_S f_n \, d\mu$$

holds for any m. We are therefore allowed to take the limit of the left side, which is given by the Monotone convergence theorem. □

Corollary 2.16 (Reverse Fatou's lemma)**.** *Suppose* $(f_n)_{n=1}^{\infty}$ *is a sequence of nonnegative measurable functions, and there exists an integrable function* g *such that* $f_n \leq g$ *for all* n*. Then*

$$\int_S \limsup_{n\to\infty} f_n \, d\mu \geq \limsup_{n\to\infty} \int_S f_n \, d\mu.$$

Proof. By Fatou's lemma we have

$$\begin{aligned}\int_S \liminf_{n\to\infty}(g - f_n) \, d\mu &= \int_S g \, d\mu - \int_S \limsup_{n\to\infty} f_n \, d\mu \\ &\leq \liminf_{n\to\infty} \int_S (g - f_n) \, d\mu = \int_S g \, d\mu - \limsup_{n\to\infty} \int_S f_n \, d\mu,\end{aligned}$$

and since $\int_S g \, d\mu$ is finite we can subtract it from both sides. □

Remark. There are set versions of Fatou's lemma as well. For a sequence $(E_n)_{n=1}^{\infty}$ of measurable sets, we define

$$\liminf_{n\to\infty} E_n \coloneqq \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n,$$

$$\limsup_{n\to\infty} E_n \coloneqq \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n.$$

We omit a proof, but we nonetheless have that $\liminf_{n\to\infty} E_n$ and $\limsup_{n\to\infty} E_n$ are measurable. Further,

$$\mu(\liminf_{n\to\infty} E_n) \leq \liminf_{n\to\infty} \mu(E_n),$$

and if $\mu(E_1) < \infty$, then

$$\mu(\limsup_{n\to\infty} E_n) \geq \limsup_{n\to\infty} \mu(E_n).$$

Theorem 2.17 (Dominated convergence theorem). *Suppose that $g, (f_n)_{n=1}^{\infty} \in \mathcal{L}^1$, $|f_n| \leq g$ for all n and f_n converges pointwise to function f. Then*

$$\lim_{n\to\infty} \int_S |f_n - f| \, d\mu = 0.$$

Proof. We first note that $f = \limsup_{n\to\infty} f_n$ and is thus measurable. Since $0 \leq |f_n - f| \leq 2g$, we can apply the previous corollary,

$$\limsup_{n\to\infty} \int_S |f_n - f| \, d\mu \leq \int_S \limsup_{n\to\infty} |f_n - f| \, d\mu = 0 \leq \liminf_{n\to\infty} \int_S |f_n - f| \, d\mu.$$

□

Remark. The conclusion implies $\lim_{n\to\infty} \int_S f_n \, d\mu = \int_S f \, d\mu$ since $|\int_S f_n \, d\mu - \int_S f \, d\mu| \leq \int_S |f_n - f| \, d\mu$ by Proposition 2.13c.

A property p is said to hold *almost everywhere*, abbreviated a.e., if $\mu(\{x \in S : p \text{ does not hold}\}) = 0$. Since the integral over subsets with measure 0 always equals 0, we may replace the condition of pointwise convergence in the Monotone convergence theorem and the Dominated convergence theorem with pointwise convergence a.e. This extra flexibility will be useful later when dealing with conditional expectation.

3 Probability Theory

We begin by introducing the basic measure-theoretic definitions in probability theory. A *probability space* is a measure space, $(\Omega, \mathcal{F}, \mathbb{P})$, where the elements of Ω, are called *outcomes*, usually denoted by ω, the elements of $\mathcal{F}$ are called *events*, and $\mathbb{P}(\Omega) = 1$. A *random variable* X is a measurable function, with the notational convention that $\mathbb{P}(X \in B) \coloneqq \mathbb{P}(\{\omega : X(\omega) \in B\})$. The *expected value* of an integrable random variable is defined as $\mathrm{E}[X] \coloneqq \int_\Omega X \, d\mathbb{P}$.

A property is said to hold *almost surely*, abbreviated a.s., if said property has probability 1 of occurring.

Fubini's theorem, see [5] for a concise proof, gives a nontrivial representation of the expected value. Let μ denote the Lebesgue measure, then we have

$$X^+(\omega) = \int_0^{X^+(\omega)} d\mu(x) = \int_0^\infty I_{\{X^+(\omega) > x\}} \, d\mu(x),$$

hence

$$\begin{aligned}
\mathrm{E}[X^+] &= \int_\Omega \left(\int_0^\infty I_{\{X^+(\omega)>x\}} \, d\mu(x) \right) d\mathbb{P}(\omega) \\
&= \int_0^\infty \left(\int_\Omega I_{\{X^+(\omega)>x\}} \, d\mathbb{P}(\omega) \right) d\mu(x) = \int_0^\infty \mathbb{P}(X > x) d\mu(x).
\end{aligned}$$

Similarly,

$$X^-(\omega) = \int_{-X^-(\omega)}^0 d\mu(x) = \int_{-\infty}^0 I_{\{X^-(\omega) \geq -x\}} \, d\mu(x) = \int_{-\infty}^0 I_{\{-X^-(\omega) \leq x\}} \, d\mu(x),$$

hence

$$\begin{aligned}
\mathrm{E}[X^-] &= \int_\Omega \left(\int_{-\infty}^0 I_{\{-X^-(\omega) \leq x\}} \, d\mu(x) \right) d\mathbb{P}(\omega) \\
&= \int_{-\infty}^0 \left(\int_\Omega I_{\{-X^-(\omega) \leq x\}} \, d\mathbb{P}(\omega) \right) d\mu(x) = \int_{-\infty}^0 \mathbb{P}(X \leq x) d\mu(x).
\end{aligned}$$

In total, given that X is integrable,

$$\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-] = \int_0^\infty \mathbb{P}(X > x) d\mu(x) - \int_{-\infty}^0 \mathbb{P}(X \leq x) d\mu(x).$$

The constructive version of the proof of the next theorem shows that probability theory is helpful for mathematical analysis.

Theorem 3.1 (Weierstrass approximation theorem)**.** *For any continuous real-valued function, f, on an interval $[a, b] \subset \mathbb{R}$, there exists a sequence of polynomials which converges uniformly to f on that interval.*

Proof. Without loss of generality, we let the interval be $[0,1]$. Indeed, if the theorem holds for this specific case, and f satisfies the original conditions, then g, defined as $g(x) := f(a+(b-a)x)$ on $[0,1]$, satisfies the new conditions.

Let $X_n \sim \mathrm{Bin}(n,p)$ be binomially distributed, that is,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Then

$$\mathrm{E}[f(n^{-1}X_n)] = \sum_{k=0}^{n} \binom{n}{k} f(n^{-1}k) p^k (1-p)^{n-k},$$

which is the polynomial we will use for approximation with p as independent variable. As we know from the undergraduate theory, $\mathrm{E}[n^{-1}X_n] = p$ and $\mathrm{Var}(n^{-1}X_n) = n^{-1}p(1-p)$. Hence by Chebyshev's inequality,

$$\mathbb{P}(|n^{-1}X_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} < \frac{1}{n\varepsilon^2}.$$

Since f is continuous and $[0,1]$ is compact, f is bounded, that is, $|f(x)| \leq K$ for every $x \in [0,1]$ and some $K \in \mathbb{R}$. Furthermore, f is uniformly continuous. Thus for a fixed $\varepsilon > 0$, there exists some δ such that $|f(x) - f(y)| < \frac{\varepsilon}{2}$ if $|x - y| < \delta$.

Using that δ, we let $A_{n,\delta} := \{\omega : |n^{-1}X_n(\omega) - p| < \delta\}$ and $h(n,p) := |f(n^{-1}X_n) - f(p)|$. We then have

$$\begin{aligned}
|\mathrm{E}[f(n^{-1}X_n)] - f(p)| &\leq \mathrm{E}[|f(n^{-1}X_n) - f(p)|] \\
&= \mathrm{E}[h(n,p)] \\
&= \mathrm{E}[h(n,p)|A_{n,\delta}]\mathbb{P}(A_{n,\delta}) + \mathrm{E}[h(n,p)|A_{n,\delta}^c]\mathbb{P}(A_{n,\delta}^c) \\
&\leq \frac{\varepsilon}{2}\mathbb{P}(|n^{-1}X_n - p| < \delta) + 2K\mathbb{P}(|n^{-1}X_n - p| \geq \delta) \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

whenever $n > \frac{4K}{\varepsilon\delta^2}$. The choices of δ and n are independent of p. Uniform convergence has thereby been proved. □

A real function φ is *convex* on an open interval (a,b) if for every $x, y \in (a,b)$ and $\lambda \in (0,1)$ we have

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y).$$

By reorganizing, this is seen to be equivalent to

$$\frac{\varphi(t) - \varphi(s)}{t - s} \leq \frac{\varphi(u) - \varphi(t)}{u - t},$$

where $a < s < t < u < b$. Indeed, this yields

$$\varphi(t) \leq \frac{t-s}{u-s}\varphi(u) + \frac{u-t}{u-s}\varphi(s),$$

and if $\lambda = \frac{t-s}{u-s}$, then $1 - \lambda = \frac{u-t}{u-s}$ and $t = \lambda u + (1-\lambda)s$. We now show that convexity implies continuity. Let $a < r < s < t < u < b$, then

$$\frac{\varphi(s) - \varphi(r)}{s - r} \leq \frac{\varphi(t) - \varphi(s)}{t - s} \leq \frac{\varphi(u) - \varphi(t)}{u - t}.$$

Thus the middle fraction seen as a function of $s, t \in (r, u)$ is bounded. As the denominator becomes arbitrarily close to zero, so must also the numerator.

Theorem 3.2 (Jensen's inequality)**.** *Suppose $\varphi : \mathcal{O} \to \mathbb{R}$ is convex on an open interval $\mathcal{O} \subset \mathbb{R}$, and X is an integrable random variable such that $X(\omega) \in \mathcal{O}$ for all $\omega \in \Omega$. If $\varphi(X) \in \mathcal{L}^1$, then*

$$\varphi(\mathrm{E}[X]) \leq \mathrm{E}[\varphi(X)].$$

Proof. For any $t \in \mathcal{O}$, let $(s_n)_{n=1}^{\infty}$ and $(u_n)_{n=1}^{\infty}$ be sequences in $\mathcal{O}$ such that $s_n < t < u_n$ for every $n \in \mathbb{N}$, and $\lim_{n\to\infty} s_n = \lim_{n\to\infty} u_n = t$. By convexity and continuity,

$$(D_-\varphi)(t) \coloneqq \lim_{n\to\infty} \frac{\varphi(t) - \varphi(s_n)}{t - s_n}, \text{ and } (D_+\varphi)(t) \coloneqq \lim_{n\to\infty} \frac{\varphi(u_n) - \varphi(t)}{u_n - t}$$

both exist, and $(D_-\varphi)(t) \leq (D_+\varphi)(t)$ for every $t \in \mathcal{O}$. Let $x, t \in \mathcal{O}$, then, if $x > t$,

$$(D_-\varphi)(t) \leq (D_+\varphi)(t) \leq \frac{\varphi(x) - \varphi(t)}{x - t}$$

and so

$$(D_-\varphi)(t)(x - t) \leq (D_+\varphi)(t)(x - t) \leq \varphi(x) - \varphi(t).$$

If $x < t$,

$$(D_+\varphi)(t) \geq (D_-\varphi)(t) \geq \frac{\varphi(x) - \varphi(t)}{x - t},$$

and since the denominator is negative,

$$(D_+\varphi)(t)(x - t) \leq (D_-\varphi)(t)(x - t) \leq \varphi(x) - \varphi(t).$$

Hence for any $k \in [(D_-\varphi)(t), (D_+\varphi)(t)]$ and any $x, t \in \mathcal{O}$,

$$\varphi(x) \geq k(x - t) + \varphi(t).$$

Thus we have almost surely,

$$\varphi(X) \geq k(X - \mathrm{E}[X]) + \varphi(\mathrm{E}[X]),$$

and by taking expectations,

$$\mathrm{E}[\varphi(X)] \geq \mathrm{E}[k(X - \mathrm{E}[X]) + \varphi(\mathrm{E}[X])] = \varphi(\mathrm{E}[X]). \qquad \square$$

Remark. Using the notation of the proof above, we have by continuity

$$\varphi(x) = \sup_{t\in\mathcal{O}}[(D_+\varphi)(t)(x-t) + \varphi(t)] = \sup_n (a_n x + b_n) \text{ for all } x \in \mathcal{O},$$

where $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ are sequences in $\mathbb{R}$. This will be needed later for the proof of the conditional expectation version of Jensen's inequality.

We also have a few other important inequalities. For simple and effective proofs via Young's inequality of the next two theorems, see [7].

Theorem 3.3 (Hölder's inequality)**.** *Suppose $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then for any two measurable functions f, g on $(S, \mathcal{F}, \mu)$, we have*

$$\int_S |fg|\, d\mu \leq \left(\int_S |f|^p\, d\mu\right)^{\frac{1}{p}} \left(\int_S |g|^q\, d\mu\right)^{\frac{1}{q}},$$

or more concisely put,

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Theorem 3.4 (Minkowski's inequality)**.** *If the conditions of the previous theorem hold and $f, g \in \mathcal{L}^p$, then*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Remark. Minkowski's inequality is the triangle inequality in $\mathcal{L}^p$-spaces. As a side note, $\|\cdot\|_p$ defines a norm of equivalence classes of $\mathcal{L}^p$ where f and g are equivalent if and only if $f = g$ almost everywhere. Furthermore, these normed vector spaces are complete, i.e. they are Banach spaces. For $p = 2$ the norm defines an inner product and is thus a Hilbert space.

Proposition 3.5 (Lyapunov's inequality)**.** *Let $0 < p < r < \infty$, then $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ whenever $X \in \mathcal{L}^r(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore,*

$$\mathrm{E}[|X|^p]^{\frac{1}{p}} \leq \mathrm{E}[|X|^r]^{\frac{1}{r}}.$$

Proof. We give two proofs. Since $r/p > 1$, the function $\varphi(x) = x^{\frac{r}{p}}$ has a nondecreasing derivative on $[0, \infty)$ and is thus convex. Since $\min\{|X|, n\}^p \in \mathcal{L}^1$ for every $n \in \mathbb{N}$, we can can apply Jensen's inequality,

$$\begin{aligned}\mathrm{E}[\min\{|X|, n\}^p]^{\frac{r}{p}} &= \varphi(\mathrm{E}[\min\{|X|, n\}^p]) \\ &\leq \mathrm{E}[\varphi(\min\{|X|, n\}^p)] = \mathrm{E}[\min\{|X|, n\}^r] \leq \mathrm{E}[|X|^r].\end{aligned}$$

Applying the Monotone convergence theorem completes the proof.

Alternatively, let $q := (1 - \frac{p}{r})^{-1}$, then $\frac{r}{p}, q > 1$ and $\frac{p}{r} + \frac{1}{q} = 1$. If $Y \in \mathcal{L}^{\frac{r}{p}}, Z \in \mathcal{L}^q$, then by Hölder's inequality,

$$\mathrm{E}[|YZ|] \leq \mathrm{E}[|Y|^{\frac{r}{p}}]^{\frac{p}{r}} \mathrm{E}[|Z|^q]^{\frac{1}{q}}.$$

Set $Y = |X|^p$ and $Z = I_\Omega$ to yield the desired inequality. $\square$

4 Independence

We now attempt to put independence of random variables on a rigorous footing. For this we introduce the concepts of π-systems and d-systems. These will help us define independence more generally and then recover familiar results from undergraduate theory.

Definition 4.1. A nonempty collection $\mathcal{I}$ of subsets of S is a *π-system* if it is closed with respect to finite intersection.

Definition 4.2. A collection $\mathcal{D}$ of subsets of S is a *d-system*, or *Dynkin system*, if

(i) $S \in \mathcal{D}$,

(ii) $A, B \in \mathcal{D}$ and $B \subset A \implies A \setminus B \in \mathcal{D}$,

(iii) $A_1 \subset A_2 \subset \cdots \in \mathcal{D}, \ \lim_{n\to\infty} A_n = A \implies A \in \mathcal{D}$.

Lemma 4.3. *A collection $\mathcal{F}$ is a σ-algebra if and only if $\mathcal{F}$ is both a π-system and a d-system.*

Proof. The "only if"-part is clear. Now suppose $\mathcal{F}$ is π-system and a d-system. The first two conditions of our definition of a σ-algebra clearly hold. Furthermore, $\mathcal{F}$ being closed with respect to finite intersection and with respect to complements implies it being closed with respect to finite unions by De Morgan's laws. Let $A_1, A_2 \cdots \in \mathcal{F}$ and $B_n := \bigcup_{i=1}^n A_i$, then $B_1 \subset B_2 \subset \cdots \in \mathcal{F}$, hence by the third property of d-systems, $\bigcup_{i=1}^\infty A_i = \lim_{n\to\infty} B_n \in \mathcal{F}$. □

Theorem 4.4 (Dynkin's lemma)**.** *Let $d(\mathcal{I})$ be the d-system generated by all the sets belonging to the π-system $\mathcal{I}$. Then $d(\mathcal{I}) = \sigma(\mathcal{I})$.*

Proof. Clearly $d(\mathcal{I}) \subset \sigma(\mathcal{I})$. Thus by the previous lemma it suffices to show that $d(\mathcal{I})$ is a π-system. Let $\mathcal{D}_1 = \{A \in d(\mathcal{I}) : A \cap B \in d(\mathcal{I}), \ \forall B \in \mathcal{I}\}$, then $\mathcal{I} \subset \mathcal{D}_1 \subset d(\mathcal{I})$. Also, $\mathcal{D}_1$ is a d-system.

Indeed, take any $B \in \mathcal{I}$, then $S \cap B = B \in d(\mathcal{I})$, thus $S \in \mathcal{D}_1$. Also, if $A_1, A_2 \in \mathcal{D}_1$ and $A_2 \subset A_1$, then $A_1 \cap B, A_2 \cap B \in d(\mathcal{I})$ and so $(A_1 \setminus A_2) \cap B = (A_1 \cap B) \setminus (A_2 \cap B) \in d(\mathcal{I})$. For the third and last property, let $A_1 \subset A_2 \subset \cdots \in \mathcal{D}_1$, then $A_1 \cap B \subset A_2 \cap B \subset \cdots \in d(\mathcal{I})$. Thus $(\lim_{n\to\infty} A_n) \cap B = \lim_{n\to\infty} A_n \cap B \in d(\mathcal{I})$. This establishes that $\mathcal{D}_1 = d(\mathcal{I})$.

Now, let $\mathcal{D}_2 = \{A \in d(\mathcal{I}) : A \cap B \in d(\mathcal{I}), \ \forall B \in d(\mathcal{I})\}$, then by considering the preceding result we have $\mathcal{I} \subset \mathcal{D}_2$. We show that $\mathcal{D}_2$ is a d-system too. It follows in almost identical fashion as previously.

Take any $B \in d(\mathcal{I})$, then $S \cap B = B \in d(\mathcal{I})$, thus $S \in \mathcal{D}_2$. If $A_1, A_2 \in \mathcal{D}_2$ and $A_2 \subset A_1$, then $A_1 \cap B, A_2 \cap B \in d(\mathcal{I})$ and so $(A_1 \setminus A_2) \cap B = (A_1 \cap B) \setminus (A_2 \cap B) \in d(\mathcal{I})$. Finally, let $A_1 \subset A_2 \subset \cdots \in \mathcal{D}_2$, then $A_1 \cap B \subset A_2 \cap B \subset \cdots \in d(\mathcal{I})$. Thus $(\lim_{n\to\infty} A_n) \cap B = \lim_{n\to\infty} A_n \cap B \in d(\mathcal{I})$. This establishes that $\mathcal{D}_2 = d(\mathcal{I})$. And it is now easily seen that $\mathcal{D}_2$, and so $d(\mathcal{I})$, are π-systems as were to be proved. □

Theorem 4.5. *Suppose μ_1 and μ_2 are two measures on $(S, \mathcal{F})$ such that $\mu_1(S) = \mu_2(S) < \infty$ and $\sigma(\mathcal{I}) = \mathcal{F}$ for some π-system $\mathcal{I}$. If $\mu_1(I) = \mu_2(I)$ for every $I \in \mathcal{I}$, then $\mu_1 = \mu_2$.*

Proof. Let $\mathcal{D} = \{F \in \mathcal{F} : \mu_1(F) = \mu_2(F)\}$, then $\mathcal{I} \subset \mathcal{D} \subset \mathcal{F}$. It suffices to show that $\mathcal{D}$ is a d-system since by the previous theorem we then have $\mathcal{F} = d(\mathcal{I}) \subset \mathcal{D}$.

By the definitions of μ_1, μ_2, we have $S \in \mathcal{D}$. If $F_1, F_2 \in \mathcal{D}$ and $F_2 \subset F_1$, then $\mu_1(F_1 \setminus F_2) = \mu_1(F_1) - \mu_1(F_2) = \mu_2(F_1) - \mu_2(F_2) = \mu_2(F_1 \setminus F_2)$, and so $F_1 \setminus F_2 \in \mathcal{D}$. Lastly, suppose $F_1 \subset F_2 \subset \dots \in \mathcal{D}$ and $F_n \uparrow F$ as $n \to \infty$, then by Proposition 2.3 we have $\mu_1(F) = \lim_{n\to\infty} \mu_1(F_n) = \lim_{n\to\infty} \mu_2(F_n) = \mu_2(F)$, and so $F \in \mathcal{D}$. Thus $\mathcal{D}$ is a d-system by the previous lemma and we are done. □

A subset $\mathcal{G}$ of a σ-algebra $\mathcal{F}$ on S is a *sub-σ-algebra* if $\mathcal{G}$ is itself a σ-algebra on S. We now introduce the main concept of this section.

Definition 4.6. Sub-σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \dots \subset \mathcal{F}$ are *independent* if whenever $G_i \in \mathcal{G}_i$, $i \in \mathbb{N}$, and $i_1, \dots i_n$ are distinct, we have

$$\mathbb{P}(G_{i_1} \cap \dots \cap G_{i_n}) = \prod_{k=1}^{n} \mathbb{P}(G_{i_k}).$$

Random variables $X_1, X_2, \dots$ are independent if $\sigma(X_1), \sigma(X_2), \dots$ are independent, where $\sigma(X)$ is the smallest σ-algebra making X measurable. Events $E_1, E_2, \dots$ are independent if their corresponding indicator functions $I_{E_1}, I_{E_2}, \dots$ are independent.

Remark. As we will see with conditional expectation, $\sigma(X)$ can informally be seen as the information of the random variable X. This motivates this general definition of independence.

Proposition 4.7. *Suppose $\mathcal{G}, \mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$ and for two π-systems $\mathcal{I}, \mathcal{J}$ we have $\sigma(\mathcal{I}) = \mathcal{G}$, $\sigma(\mathcal{J}) = \mathcal{H}$. Then $\mathcal{G}, \mathcal{H}$ are independent if and only if $\mathcal{I}, \mathcal{J}$ are independent, that is, $\mathbb{P}(I \cap J) = \mathbb{P}(I)\mathbb{P}(J)$ for every $I \in \mathcal{I}, J \in \mathcal{J}$.*

Proof. The "only if"-part is clear. Suppose $\mathcal{I}, \mathcal{J}$ are independent. Fix $I \in \mathcal{I}$ and define two measures μ_1, μ_2 on $\mathcal{H}$ by $H \mapsto \mathbb{P}(I \cap H)$ and $H \mapsto \mathbb{P}(I)\mathbb{P}(H)$, respectively. These are measures. Indeed if $H_1, H_2, \dots \in \mathcal{H}$ and are pairwise disjoint, then

$$\begin{aligned}\mu_1\Big(\bigcup_{n=1}^{\infty} H_n\Big) &= \mathbb{P}\Big(I \cap \bigcup_{n=1}^{\infty} H_n\Big) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} (I \cap H_n)\Big) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(I \cap H_n) = \sum_{n=1}^{\infty} \mu_1(H_n),\end{aligned}$$

also

$$\mu_2(\bigcup_{n=1}^{\infty} H_n) = \mathbb{P}(I)\mathbb{P}(\bigcup_{n=1}^{\infty} H_n) = \mathbb{P}(I)\sum_{n=1}^{\infty}\mathbb{P}(H_n)$$
$$= \sum_{n=1}^{\infty}\mathbb{P}(I)\mathbb{P}(H_n) = \sum_{n=1}^{\infty}\mu_2(H_n).$$

Moreover, μ_1, μ_2 agree on $\mathcal{J}$, thus by the previous theorem they agree on $\sigma(\mathcal{J}) = \mathcal{H}$. Now fix $H \in \mathcal{H}$ and define new measures μ_3, μ_4 on $\mathcal{G}$ by $G \mapsto \mathbb{P}(H \cap G)$ and $G \mapsto \mathbb{P}(H)\mathbb{P}(G)$, respectively. By comparison with the previously defined measures, we see that μ_3, μ_4 agree on $\mathcal{I}$, and therefore also on $\sigma(\mathcal{I}) = \mathcal{H}$. Thus $\mathbb{P}(G \cap H) = \mathbb{P}(G)\mathbb{P}(H)$ for any pair of $G \in \mathcal{G}, H \in \mathcal{H}$. □

Corollary 4.8. *Suppose X and Y are two (finite) random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then X and Y are independent if and only if $\mathbb{P}(X \leq x,\ Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ for every $x, y \in \mathbb{R}$.*

Proof. It suffices to note that the collection of sets $(-\infty, x]$, where $x \in \mathbb{R}$, constitute a π-system. □

Proposition 4.9. *If $X, Y \in \mathcal{L}^1$ and are independent, then $XY \in \mathcal{L}^1$ and $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.*

Proof. By linearity, we only need to treat the case when $X, Y \geq 0$. Suppose X, Y are simple, or more precisely, $X = \sum_{i=1}^{n} a_i I_{A_i},\ Y = \sum_{i=1}^{m} b_i I_{B_i}$ then

$$\mathrm{E}[XY] = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{P}(A_i \cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{P}(A_i)\mathbb{P}(B_j)$$
$$= \sum_{i=1}^{n} a_i \mathbb{P}(A_i) \sum_{j=1}^{m} b_j \mathbb{P}(B_j) = \mathrm{E}[X]\mathrm{E}[Y].$$

Suppose now that X, Y are not simple. By Proposition 2.11 there are sequences $(X_n)_{n=1}^{\infty}, (Y_n)_{n=1}^{\infty}$ of nonnegative simple random variables such that $X_n \uparrow X$ and $Y_n \uparrow Y$. From their construction X_n and Y_n are easily seen to be independent. By the Monotone convergence theorem,

$$\mathrm{E}[XY] = \lim_{n\to\infty} \mathrm{E}[X_n Y_n] = \lim_{n\to\infty} \mathrm{E}[X_n]\mathrm{E}[Y_n] = \mathrm{E}[X]\mathrm{E}[Y].$$ □

Our excursion into the intricacies of independence yields an impressive theorem, namely Kolmogorov's 0-1 theorem. But we need to first introduce the concept of a *tail σ-algebra*.

Definition 4.10. Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables, and define $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \dots)$ and $\mathcal{T} := \bigcap_{n=1}^{\infty} \mathcal{T}_n$, where $\sigma(X_{n+1}, X_{n+2}, \dots)$ is the smallest σ-algebra making $X_{n+1}, X_{n+2}, \dots$ all measurable. Then $\mathcal{T}$ is the *tail σ-algebra*.

Proposition 4.11. *The following events belong to $\mathcal{T}$:*

$$F_1 := \{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\},$$
$$F_2 := \{\omega : \sum_{n=1}^{\infty} X_n(\omega) \text{ converges}\},$$
$$F_3 := \{\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) \text{ exists}\}.$$

Proof. For any $n \in \mathbb{N}$ we have

$$\limsup_{k\to\infty} X_k = \inf_{m\geq 0}\left(\sup_{k\geq m} X_k\right) = \inf_{m\geq n+1}\left(\sup_{k\geq m} X_k\right),$$

thus $\mathcal{T}_n$ makes $\limsup_{k\to\infty} X_k$ measurable. A similar argument shows that the same property holds for $\liminf_{k\to\infty} X_k$. Now,

$$F_1 = \{\limsup_{k\to\infty} X_k < \infty\} \cap \{\liminf_{k\to\infty} X_k > -\infty\} \cap \{\limsup_{k\to\infty} X_k - \liminf_{k\to\infty} X_k = 0\},$$

thus $F_1 \in \mathcal{T}_n$. Since n was arbitrary this holds for every n, hence $F_1 \in \bigcap_{n=1}^{\infty} \mathcal{T}_n = \mathcal{T}$. For the second case, let $n \in \mathbb{N}$ be fixed. Define $S_m := \sum_{k=n+1}^{m} X_k$ for $m > n$, then $\mathcal{T}_n$ makes S_m measurable for $m > n$, thus

$$F_2 = \{\omega : \sum_{k=1}^{\infty} X_k(\omega) \text{ converges}\} = \{\omega : \lim_{m\to\infty} S_m(\omega) \text{ exists}\} \in \mathcal{T}_n,$$

by the previous case. Since n was arbitrary, F_2 belongs to $\mathcal{T}$. Lastly, let n be fixed again. Define $S'_m := \frac{1}{m}\sum_{k=n+1}^{n+1+m} X_k$ for $m > n$, then $\mathcal{T}_n$ makes S'_m measurable for $m > n$, thus once again

$$F_3 = \{\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k(\omega) \text{ exists}\} = \{\omega : \lim_{m\to\infty} S'_m(\omega) \text{ exists}\} \in \mathcal{T}_n,$$

hence $F_3 \in \mathcal{T}$. □

Theorem 4.12 (Kolmogorov's 0-1 law)**.** *Let $\mathcal{T}$ be the tail σ-algebra corresponding to the sequence $(X_n)_{n=1}^{\infty}$ of independent random variables. Then $\mathbb{P}(F)$ equals 0 or 1 for every $F \in \mathcal{T}$. Moreover, if f is a $\mathcal{T}$-measurable function, then there exists $c \in \overline{\mathbb{R}}$ such that $\mathbb{P}(f = c) = 1$, that is, f is a.s. constant.*

Proof. Let $\mathcal{X}_n := \sigma(X_1, \ldots, X_n)$ and $\mathcal{T}_n$ be defined as before. Consider the two π-systems given by sets of the form $\{\omega : X_i(\omega) \leq x_i \text{ for } 1 \leq i \leq n\}$ and sets of the form $\{\omega : X_i(\omega) \leq x_i \text{ for } n+1 \leq i \leq n+r\}$ where $x_i \in \mathbb{R}$ and $r \in \mathbb{N}$. These π-systems generate $\mathcal{X}_n$ and $\mathcal{T}_n$, respectively, thus $\mathcal{X}_n$ and $\mathcal{T}_n$ are independent. Since $\mathcal{T} \subset \mathcal{T}_n$ for every n, we have that $\mathcal{X}_n$ and $\mathcal{T}$ are independent for every n. This in turn implies that the π-system $\mathcal{X}_\infty := \bigcup_{n=1}^{\infty} \mathcal{X}_n$ is independent of $\mathcal{T}$.

Now, $\mathcal{T} \subset \sigma(\mathcal{X}_\infty)$, hence $\mathcal{T}$ is independent of itself. Thus for every $F \in \mathcal{T}$, we have $\mathbb{P}(F) = \mathbb{P}(F \cap F) = \mathbb{P}(F)\mathbb{P}(F)$, hence $\mathbb{P}(F)$ equals either 0 or 1.

For the second assertion, let f be $\mathcal{T}$-measurable. Since $\{\omega : f(\omega) \leq \alpha\} \in \mathcal{T}$ for every $\alpha \in \overline{\mathbb{R}}$, we have $\mathbb{P}(f \leq \alpha) = 0$ or 1. Let $c = \inf\{\alpha \in \overline{\mathbb{R}} : \mathbb{P}(f \leq \alpha) = 1\}$, then $\mathbb{P}(f = c) = 1$, that is, $f = c$ a.s. $\square$

We end this section with two well-known theorems

Theorem 4.13 (First Borel-Cantelli lemma)**.** *Let* $(E_n)_{n=1}^\infty$ *be a sequence of events. If* $\sum_{n=1}^\infty \mathbb{P}(E_n) < \infty$*, then* $\mathbb{P}(\limsup\limits_{n\to\infty} E_n) = 0$.

Proof. For $\varepsilon > 0$, there exists some $N \in \mathbb{N}$ such that $\sum_{n=N}^\infty \mathbb{P}(E_n) < \varepsilon$. By definition

$$\limsup_{n\to\infty} E_n = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty E_n \subset \bigcup_{n=N}^\infty E_n,$$

thus

$$\mathbb{P}(\limsup_{n\to\infty} E_n) \leq \mathbb{P}(\bigcup_{n=N}^\infty E_n) \leq \sum_{n=N}^\infty \mathbb{P}(E_n) < \varepsilon.$$

Since ε is arbitrary, the proof follows. $\square$

Theorem 4.14 (Second Borel-Cantelli lemma)**.** *Let* $(E_n)_{n=1}^\infty$ *be a sequence of independent events. If* $\sum_{n=1}^\infty \mathbb{P}(E_n) = \infty$*, then* $\mathbb{P}(\limsup\limits_{n\to\infty} E_n) = 1$.

Remark. Since $\limsup\limits_{n\to\infty} E_n$ is in the tail σ-algebra and by assuming independence, Kolmogorov's 0-1 applies here. However, it will not be needed for the proof.

Proof. First note that

$$(\limsup_{n\to\infty} E_n)^c = \left(\bigcap_{m=1}^\infty \bigcup_{n=m}^\infty E_n\right)^c = \bigcup_{m=1}^\infty \bigcap_{n=m}^\infty E_n^c.$$

Hence it is enough to show $\mathbb{P}(\cap_{n=m}^\infty E_n^c) = 0$ for every $m \in \mathbb{N}$ since then

$$\mathbb{P}(\limsup_{n\to\infty} E_n)^c) \leq \sum_{m=1}^\infty \mathbb{P}(\bigcap_{n=m}^\infty E_n^c) = 0.$$

By independence, we have

$$\mathbb{P}\left(\bigcap_{n=m}^N E_n^c\right) = \prod_{n=m}^N \mathbb{P}(E_n^c),$$

for every $N \geq m$. Hence by first taking the limit of the left and then the right side, we get

$$\mathbb{P}\left(\bigcap_{n=m}^{\infty} E_n^c\right) \leq \prod_{n=m}^{\infty} \mathbb{P}(E_n^c).$$

and, reversing the order,

$$\mathbb{P}\left(\bigcap_{n=m}^{\infty} E_n^c\right) \geq \prod_{n=m}^{\infty} \mathbb{P}(E_n^c).$$

Using that $1 - x \leq e^{-x}$ for $x \in [0, \infty)$, we finally get

$$\mathbb{P}\left(\bigcap_{n=m}^{\infty} E_n^c\right) = \prod_{n=m}^{\infty} \mathbb{P}(E_n^c) = \prod_{n=m}^{\infty} (1 - \mathbb{P}(E_n)) \leq \exp\left\{-\sum_{n=m}^{\infty} \mathbb{P}(E_n)\right\} = 0. \quad \square$$

5 Conditional Expectation

We first need a few auxiliary results before defining the the expected value conditioned on a σ-algebra.

Lemma 5.1. *Let $f \in \mathcal{L}^1(S, \mathcal{F}, \mu)$, if for every $F \in \mathcal{F}$ we have $\int_F f\, d\mu = 0$, then $f = 0$ a.e.*

Proof. For any $\varepsilon > 0$, we have

$$0 \leq \varepsilon\mu(\{x : f(x) \geq \varepsilon\}) = \int_{\{x: f(x) \geq \varepsilon\}} \varepsilon\, d\mu \leq \int_{\{x: f(x) \geq \varepsilon\}} f\, d\mu = 0,$$

thus $\mu(\{x : f(x) \geq \varepsilon\}) = 0$. A similar argument shows that $\mu(\{x : f(x) \leq -\varepsilon\}) = 0$. Together we have

$$\mu(\{x : f(x) \neq 0\}) = \mu(\bigcup_{n=1}^{\infty} \{x : f(x) \geq \frac{1}{n}\} \cup \{x : f(x) \leq -\frac{1}{n}\}) = 0. \qquad \square$$

Remark. The result and proof are similar when $\int_F f\, d\mu \geq 0$.

Lemma 5.2. *Let f be a nonnegative measurable function on $(S, \mathcal{F}, \mu)$, then the function $v : \mathcal{F} \to [0, \infty]$ defined by $v(F) = \int_F f\, d\mu$ is a measure.*

Proof. Let $F_1, F_2, \dots \in \mathcal{F}$ and $F_i \cap F_j = \emptyset$ for $i \neq j$. Define $E_n := \bigcup_{k=1}^n F_k$ and $E := \lim_{n\to\infty} E_n$. Hence $0 \leq I_{E_n} f \leq I_{E_{n+1}} f$ and $I_{E_n} f(x) \to I_E f(x)$ as $n \to \infty$, for every $x \in S$. Using the Monotone convergence theorem, we have

$$\begin{aligned} v(\bigcup_{n=1}^{\infty} F_n) &= v(E) = \int_E f\, d\mu = \int_S I_E f\, d\mu = \lim_{n\to\infty} \int_S I_{E_n} f\, d\mu \\ &= \lim_{n\to\infty} \int_{\bigcup_{k=1}^n F_k} f\, d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} \int_{F_k} f\, d\mu = \sum_{k=1}^{\infty} v(F_k). \qquad \square \end{aligned}$$

Let μ be a measure on $(S, \mathcal{F})$. If $\mu(S) < \infty$ then μ is *finite*. If there exists a sequence $E_1 \subset E_2 \subset \dots \in \mathcal{F}$ such that $\mu(E_n) < \infty$ for all $n \in \mathbb{N}$ and $E_n \uparrow S$ as $n \to \infty$, then μ is *σ-finite*. Thus all finite measures, including probability measures, are also σ-finite. The Lebesgue measure on $\mathbb{R}$ is an example of a σ-finite measure which is not finite.

Let $\mathcal{N}_\mu$ denote the collection of zero sets of μ, that is, $\mathcal{N}_\mu = \{E \in \mathcal{F} : \mu(E) = 0\}$. Let ν be another measure, then ν is *absolutely continuous* with respect to μ, if $\mathcal{N}_\mu \subset \mathcal{N}_\nu$. It is denoted by $\nu \ll \mu$.

We can now state an important result in measure theory. We refer the reader to [7] for a constructive proof. Let $\mathcal{M}^+(\mathcal{F})$ denote the set nonnegative $\mathcal{F}$-measurable functions.

Theorem 5.3 (Radon-Nikodým)**.** *Suppose μ and ν are σ-finite measures on $(S, \mathcal{F})$. Then the following are equivalent:*

(i) $\nu \ll \mu$,

(ii) $\nu(E) = \int_E f\, d\mu$, *for some $f \in \mathcal{M}^+(\mathcal{F})$ and all $E \in \mathcal{F}$.*

Remark. By Lemma 5.2 the second condition of the theorem makes sense. The function f is referred to as the *Radon-Nikodým derivative* and is denoted by $\frac{d\nu}{d\mu}$.

Theorem 5.4 (Definition and existence of conditional expectation.)**.** *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ be a random variable and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ which satisfies*

$$\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P},$$

for every $G \in \mathcal{G}$. This random variable Y is a version of the conditional expectation *of X given $\mathcal{G}$. It is denoted by* $\mathrm{E}[X|\mathcal{G}]$.

Proof. We assume that $X \geq 0$. Let $\mathbb{P}_0$ be the restriction of $\mathbb{P}$ to $\mathcal{G}$. Let ν be the measure on $(\Omega, \mathcal{G})$ for which $G \mapsto \int_G X \, d\mathbb{P}$. Since X is integrable, ν is finite. We also have $\nu \ll \mathbb{P}_0$, hence by the Radon-Nikodým theorem,

$$\int_G X \, d\mathbb{P} = \nu(G) = \int_G \frac{d\nu}{d\mathbb{P}_0} \, d\mathbb{P}_0 = \int_G \frac{d\nu}{d\mathbb{P}_0} \, d\mathbb{P}.$$

Since $\frac{d\nu}{d\mathbb{P}_0}$ is $\mathcal{G}$-measurable, it is a version of the conditional expectation. The general case follows by decomposing, $X = X^+ - X^-$, and linearity. □

Proposition 5.5 (Uniqueness of conditional expectation)**.** *If Y' is another random variable with the properties of Y in Definition 5.4, then $Y' = Y$ a.s.*

Proof. This is an immediate consequence of Lemma 5.1. □

Remark. It is usually more convenient to refer to Y or $\mathrm{E}[X|\mathcal{G}]$ as *the* conditional expectation of X given $\mathcal{G}$, even though most results involving it only hold a.s.

Proposition 5.6. *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ where $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. If Y satisfies*

$$\int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \tag{1}$$

for every $G \in \mathcal{I}$, where $\mathcal{I}$ is a π-system which contains Ω and generates $\mathcal{G}$, then $Y = \mathrm{E}[X|\mathcal{G}]$ a.s.

Proof. We will show that the collection $\mathcal{D}$ of sets satisfying (1) form a d-system. We have $\Omega \in \mathcal{D}$ by the definition of $\mathcal{I}$. If $A, B \in \mathcal{D}$ and $B \subset A$, then

$$\int_{A \setminus B} Y \, d\mathbb{P} = \int_A Y \, d\mathbb{P} - \int_B Y \, d\mathbb{P} = \int_A X \, d\mathbb{P} - \int_B X \, d\mathbb{P} = \int_{A \setminus B} X \, d\mathbb{P},$$

thus $A \setminus B \in \mathcal{D}$. If $A_1 \subset A_2 \subset \cdots \in \mathcal{D}$, $\lim_{n\to\infty} A_n = A$, then for any $\varepsilon > 0$ there exists some N such that $n \geq N$ implies $\mathbb{P}(A \setminus A_n) < \varepsilon$. Thus

$$\int_A Y \, d\mathbb{P} - \int_{A_n} Y \, d\mathbb{P} < \varepsilon \mathrm{E}[|Y|] < \infty$$

since ε was arbitrary, $\lim_{n\to\infty} \int_{A_n} Y \, d\mathbb{P} = \int_A Y \, d\mathbb{P}$. The same argument holds for X too. Hence

$$\int_A Y \, d\mathbb{P} = \lim_{n\to\infty} \int_{A_n} Y \, d\mathbb{P} = \lim_{n\to\infty} \int_{A_n} X \, d\mathbb{P} = \int_A X \, d\mathbb{P},$$

thus $A \in \mathcal{D}$. Dynkin's lemma, Theorem 4.4, completes the proof. $\square$

Proposition 5.7. *Consider $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ as a Hilbert space with norm $(X, Y) \mapsto E[XY]$. Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then the orthogonal projection of X onto $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is given by* $\mathrm{E}[X|\mathcal{G}]$.

Proof. The orthogonal projection Z satisfies $\mathrm{E}[(X-Z)Y] = 0$ for every bounded $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. In particular, for $Y = I_G$, where $G \in \mathcal{G}$, we have

$$\int_\Omega (X - Z) I_G \, d\mathbb{P} = \int_G (X - Z) \, d\mathbb{P} = 0.$$

Put another way, for every $G \in \mathcal{G}$, we have

$$\int_G Z \, d\mathbb{P} = \int_G X \, d\mathbb{P},$$

which is the definition of $\mathrm{E}[X|\mathcal{G}]$. $\square$

We can find a useful expression for $\mathrm{E}[X|Y]$ given that the joint density $f_{X,Y}$ is known. We first need a lemma.

Lemma 5.8. *Let X, Y be random variables. Then X is $\sigma(Y)$-measurable if and only if $X = F(Y)$ for some measurable function $F : \mathbb{R} \to \mathbb{R}$.*

Proof. The "if" part is clear. For the "only-if" part, if $X = I_A$ for some $A \in \sigma(Y)$, then $X = I_{F^{-1}(A)} \circ Y$. Similarly, if $X = aI_A + bI_B$ for $a, b \in \mathbb{R}$ and $A, B \in \sigma(Y)$, then $(aI_{F^{-1}(A)} + bI_{F^{-1}(B)}) \circ Y$. Thus it holds if X is simple. If X is not simple, then assume $X \geq 0$ and take an increasing sequence of simple random variables $(X_n)_{n=1}^\infty$ given by Proposition 2.11. Let $X_n := F_n \circ Y$, and $F := \liminf_{n\to\infty} F_n$, then

$$F \circ Y = (\liminf_{n\to\infty} F_n) \circ Y = \lim_{n\to\infty} F_n \circ Y = \lim_{n\to\infty} X_n = X.$$

The general case follows by decomposing: $X = X^+ - X^-$. $\square$

Now, by Lemma 5.8, we let $F(Y) := \mathrm{E}[X|Y]$, then for any $G \in \sigma(Y)$,

$$\int_{Y\in G} X \, d\mathbb{P} = \int_G \left(\int_{\mathbb{R}} x f_{X,Y}(x, y) \, dx \right) dy.$$

Meanwhile,

$$\int_{Y\in G} X \, d\mathbb{P} = \int_{Y\in G} F(Y) \, d\mathbb{P} = \int_G \left(F(y) \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx \right) dy.$$

Combining these, we have almost surely

$$F(Y)\int_{\mathbb{R}} f_{X,Y}(x,Y)\,dx = \int_{\mathbb{R}} x f_{X,Y}(x,Y)\,dx,$$

or

$$\mathrm{E}[X|Y] = \frac{\int_{\mathbb{R}} x f_{X,Y}(x,Y)\,dx}{\int_{\mathbb{R}} f_{X,Y}(x,Y)\,dx}.$$

The expectation conditioned on another random variable $\mathrm{E}[X|Z]$ is defined as $\mathrm{E}[X|\sigma(Z)]$. The expectation conditioned on an event is not as straightforward because we want it to be constant. Therefore it is defined as $\mathrm{E}[X|A] \coloneqq \frac{1}{\mathbb{P}(A)}\int_A X\,d\mathbb{P}$, where $\mathbb{P}(A) \neq 0$.

Next we will prove some interesting properties of the conditional expectation. We assume for the remaining part of this text that $X, Y, (X_n)_{n=1}^{\infty}, (Y_n)_{n=1}^{\infty} \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}, \mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$. All the equalities, inequalities and convergences hold almost surely.

We begin with linearity.

Proposition 5.9. $\mathrm{E}[aX + bY|\mathcal{G}] = a\mathrm{E}[X|\mathcal{G}] + b\mathrm{E}[Y|\mathcal{G}]$, *for* $a, b \in \mathbb{R}$.

Proof. For any $G \in \mathcal{G}$, we have

$$\begin{aligned}\int_G \mathrm{E}[aX + bY|\mathcal{G}]\,d\mathbb{P} = \int_G (aX + bY)\,d\mathbb{P} = a\int_G X\,d\mathbb{P} + b\int_G Y\,d\mathbb{P} \\ = a\int_G \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} + b\int_G \mathrm{E}[Y|\mathcal{G}]\,d\mathbb{P},\end{aligned}$$

which conforms with Definition 5.4. □

Proposition 5.10. $\mathrm{E}[\mathrm{E}[X|\mathcal{G}]] = \mathrm{E}[X]$.

Proof. $\mathrm{E}[\mathrm{E}[X|\mathcal{G}]] = \int_\Omega \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} = \int_\Omega X\,d\mathbb{P} = \mathrm{E}[X]$. □

Next we prove the conditional expectation versions of the main convergence theorems in Section 1; the Monotone convergence theorem, Fatou's lemma and the Dominated convergence theorem. One reason the proofs are different are that we demand integrability for the conditional expectation to be well-defined.

Proposition 5.11.

(a) $X_n \uparrow X \implies \mathrm{E}[X_n|\mathcal{G}] \uparrow \mathrm{E}[X|\mathcal{G}]$.

(b) $X_n \downarrow X \implies \mathrm{E}[X_n|\mathcal{G}] \downarrow \mathrm{E}[X|\mathcal{G}]$.

(c) $|X_n| \le Y \implies \mathrm{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \le \liminf_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]$.

(d) $|X_n| \le Y \implies \mathrm{E}[\limsup_{n\to\infty} X_n|\mathcal{G}] \ge \limsup_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]$.

(e) $|X_n| \le Y$ *and* $X_n \to X$ *a.s.* $\implies \mathrm{E}[X_n|\mathcal{G}] \to \mathrm{E}[X|\mathcal{G}]$.

Proof. For (a), by the remark to Lemma 5.1, we have $\mathrm{E}[X_n|\mathcal{G}] \le \mathrm{E}[X_{n+1}|\mathcal{G}]$ for $n \ge 1$. By taking limits, we get $\lim\limits_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}] \le \mathrm{E}[X|\mathcal{G}]$. For the reverse inequality, take any $G \in \mathcal{G}$, then

$$\int_G \lim_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]\, d\mathbb{P} \ge \int_G \mathrm{E}[X_n|\mathcal{G}]\, d\mathbb{P},$$

for all $n \ge 1$. By the remark to the Monotone convergence theorem and the fact that $0 \le \mathrm{E}[X_1^-] \le \mathrm{E}[|X_1|] < \infty$, we have for any $G \in \mathcal{G}$,

$$\begin{aligned}\int_G \lim_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}]\, d\mathbb{P} &\ge \lim_{n\to\infty} \int_G \mathrm{E}[X_n|\mathcal{G}]\, d\mathbb{P} \\ &= \lim_{n\to\infty} \int_G X_n\, d\mathbb{P} = \int_G X\, d\mathbb{P} = \int_G \mathrm{E}[X|\mathcal{G}]\, d\mathbb{P}.\end{aligned}$$

Applying the remark to Lemma 5.1 again gives $\lim\limits_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}] \ge \mathrm{E}[X|\mathcal{G}]$ a.s.

For (b), apply (a) to $-X_n$. For (c), we have that $\inf\limits_{m\ge n} X_m \uparrow \liminf\limits_{n\to\infty} X_n$ as $n \to \infty$. Because of the added condition, the random variables $\liminf\limits_{n\to\infty} X_n$ and $\inf\limits_{m\ge n} X_m$ are all integrable for all $n \ge 1$. Thus their conditional expectations are well-defined. We have by (a),

$$\liminf_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}] = \lim_{n\to\infty} \mathrm{E}[\inf_{m\ge n} X_m|\mathcal{G}] = \liminf_{n\to\infty} \mathrm{E}[\inf_{m\ge n} X_m|\mathcal{G}] \le \liminf_{n\to\infty} \mathrm{E}[X_n|\mathcal{G}],$$

where the last inequality is given by the fact that $\inf\limits_{m\ge n} X_m \le X_n$.

For (d), it suffices to note that $\limsup\limits_{n\to\infty} X_n = -\liminf\limits_{n\to\infty} -X_n$. The last result (e) follows by combining (c) and (d). □

We now state the conditional expectation version of Jensen's inequality.

Proposition 5.12. *If $\varphi : \mathbb{R} \to \mathbb{R}$ is convex and $\varphi(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then $\varphi(\mathrm{E}[X|\mathcal{G}]) \le \mathrm{E}[\varphi(X)|\mathcal{G}]$.*

Proof. From the remark to Jensen's inequality there are sequences $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ such that $\varphi(x) = \sup_n(a_n x + b_n)$ for every $x \in \mathbb{R}$. Thus $a_n X + b_n \le \varphi(X)$ and so $a_n \mathrm{E}[X|\mathcal{G}] + b_n \le \mathrm{E}[\varphi(X)|\mathcal{G}]$ a.s. for every $n \in \mathbb{N}$. As demonstrated before, the countable union of sets with measure 0 has measure 0, hence we may neglect the instance were the inequality fails to hold. Consequently,

$$\sup_n (a_n \mathrm{E}[X|\mathcal{G}] + b_n) = \varphi(\mathrm{E}[X|\mathcal{G}]) \le \mathrm{E}[\varphi(X)|\mathcal{G}]$$

holds almost surely. □

Proposition 5.13. *If Y is $\mathcal{G}$-measurable and $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, then $\mathrm{E}[XY|\mathcal{G}] = Y\mathrm{E}[X|\mathcal{G}]$.*

Proof. If $Y = I_A$, then $A \in \mathcal{G}$, thus, for any $G \in \mathcal{G}$,

$$\begin{aligned}\int_G \mathrm{E}[XI_A|\mathcal{G}]\,d\mathbb{P} &= \int_G XI_A\,d\mathbb{P} = \int_{G\cap A} X\,d\mathbb{P} \\ &= \int_{G\cap A} \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} = \int_G I_A \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P}.\end{aligned}$$

By linearity, the conclusion holds if Y is simple, and we only need to consider the case when $X, Y \geq 0$. For simple nondecreasing $X_n \uparrow X, Y_n \uparrow Y$ as $n \to \infty$, and any $G \in \mathcal{G}$, we have

$$\begin{aligned}\int_G \mathrm{E}[XY|\mathcal{G}]\,d\mathbb{P} &= \int_G XY\,d\mathbb{P} = \lim_{n\to\infty} \int_G X_n Y_n\,d\mathbb{P} \\ &= \lim_{n\to\infty} \int_G Y_n \mathrm{E}[X_n|\mathcal{G}]\,d\mathbb{P} = \int_G Y \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P}.\end{aligned}$$ □

Proposition 5.14. *If $\mathcal{H} \subset \mathcal{G}$, then $\mathrm{E}[\mathrm{E}[X|\mathcal{G}]|\mathcal{H}] = \mathrm{E}[\mathrm{E}[X|\mathcal{H}]|\mathcal{G}] = \mathrm{E}[X|\mathcal{H}]$.*

Proof. For any $H \in \mathcal{H}$, we have $H \in \mathcal{G}$, and the conclusion follows immediately from Definition 5.4, as

$$\begin{aligned}\int_H \mathrm{E}[\mathrm{E}[X|\mathcal{G}]|\mathcal{H}]\,d\mathbb{P} &= \int_H \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} = \int_H X\,d\mathbb{P} \\ &= \int_H \mathrm{E}[X|\mathcal{H}] = \int_H \mathrm{E}[\mathrm{E}[X|\mathcal{H}]|\mathcal{G}]\,d\mathbb{P}.\end{aligned}$$ □

Proposition 5.15. *If $\mathcal{H}$ and $\sigma(\sigma(X), \mathcal{G})$ are independent, then* $\mathrm{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathrm{E}[X|\mathcal{G}]$.

Proof. By linearity, we can assume $X \geq 0$. Note that sets of the form $G \cap H$ where $G \in \mathcal{G}, H \in \mathcal{H}$ form a π-system $\mathcal{I}$ such that $\sigma(\mathcal{I}) = \sigma(\mathcal{G}, \mathcal{H})$. Now, for any $G \in \mathcal{G}, H \in \mathcal{H}$, we have

$$\begin{aligned}\int_{G\cap H} \mathrm{E}[X|\sigma(\mathcal{G}, \mathcal{H})]\,d\mathbb{P} &= \int_{G\cap H} X\,d\mathbb{P} = \mathrm{E}[XI_G I_H] \\ &= \mathrm{E}[XI_G]\mathrm{E}[I_H] = \mathbb{P}(H) \int_G X\,d\mathbb{P}.\end{aligned}$$

Similarly,

$$\begin{aligned}\int_{G\cap H} \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} &= \int_\Omega \mathrm{E}[X|\mathcal{G}] I_G I_H\,d\mathbb{P} = \mathrm{E}[\mathrm{E}[X|\mathcal{G}] I_G I_H] = \mathrm{E}[\mathrm{E}[X|\mathcal{G}] I_G]\mathrm{E}[I_H] \\ &= \mathbb{P}(H) \int_G \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P} = \mathbb{P}(H) \int_G X\,d\mathbb{P}.\end{aligned}$$

Thus the absolutely continuous measures defined by $F \mapsto \int_F \mathrm{E}[X|\sigma(\mathcal{G}, \mathcal{H})]\,d\mathbb{P}$ and $F \mapsto \int_F \mathrm{E}[X|\mathcal{G}]\,d\mathbb{P}$ agree on $\mathcal{I}$, and so on $\sigma(\mathcal{G}, \mathcal{H})$ too. □

Corollary 5.16. *If $\sigma(X)$ and $\mathcal{H}$ are independent, then $\mathrm{E}[X|\mathcal{H}] = \mathrm{E}[X]$.*

Proof. Set $\mathcal{G} \coloneqq \{\emptyset, \Omega\}$. Then $\sigma(\sigma(X), \mathcal{G})$ and $\mathcal{H}$ are independent. By Proposition 5.15,

$$\mathrm{E}[X|\mathcal{G}] = \mathrm{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathrm{E}[X|\mathcal{H}].$$

Evidently, $\mathrm{E}[X]$ satisfies the conditions of Definition 5.4. $\square$

6 Suggestions for Further Reading

For a similar introduction to measure theory, see chapter 1 in [1]. For a quick but still rigorous introduction to martingales, see [5]. As an opposite, [8] develops more concepts but slower. Most suitable for self-studying are [6], [7] and [9]. The first lacks rigor but has many problems with solutions making it good exercise. The second is highly rigorous. It develops probability theory while real analysis is the main focus. A complete solution manual can be found online. The third book strikes a balance between the other two. It reads like a graduate course in probability theory. Half of the given problems have solutions online.

Bibliography

[1] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 3rd ed., 1987.

[2] A. Friedman. *Foundations of Modern Analysis*. Holt, Rinehart and Winston, 1970.

[3] J. Munkres. *Topology*. Prentice Hall, 2nd ed., 2000.

[4] R. Ash. *Probability and Measure Theory*. Academic Press, 2nd ed., 2000.

[5] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[6] Z. Brzezniak and T. Zastawniak. *Basic Stochastic Processes*. Springer, 1998.

[7] R. Schilling. *Measures, Integrals and Martingales*. Cambridge University Press, 2nd ed., 2017.

[8] A. Gut. *Probability: A Graduate Course*. Springer, 2nd ed., 2013.

[9] J. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific Publishing Co., 2nd ed., 2006.