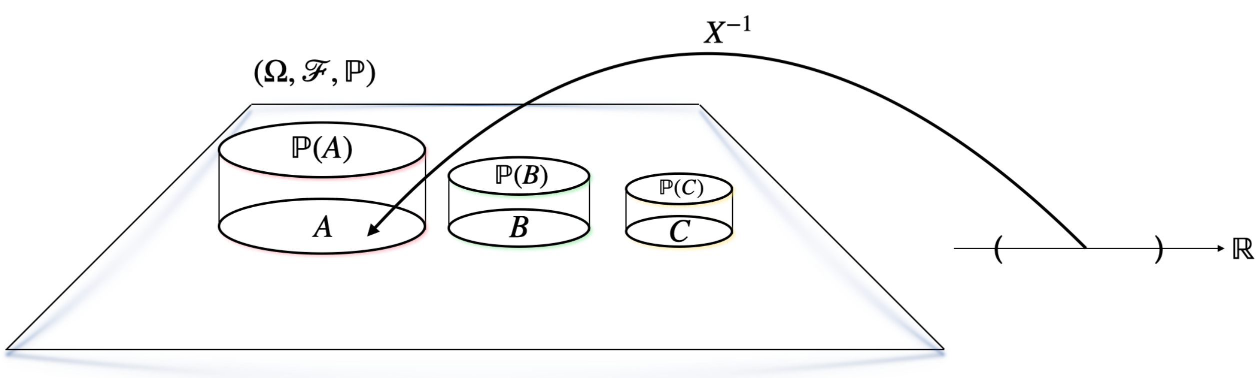
Probability

Jan Olucha Fuentes

September 9, 2024



This manuscript is a text that has evolved throughout my journey of learning about probability, it started out as a set of revision notes for my first course in probability, and was expanded as I continued taking more courses in the area. Its purpose is to provide a non-comprehensive, example-focused, light-hearted explanation of the most important concepts in the modern Mathematical Theory of Probability. Recommended knowledge for this text includes but is not limited to: elementary real analysis, elementary point-set topology, some experience about elementary probability. The reader is advised that the content in this text is not what one thinks that probability is during their first years of studies. A better name for this text would be Probability *Theory*, or *Measure Theoretic* Probability. As a note of motivation, I would like to explain that in my early stages of my mathematical education, I always despised elementary probability, as to me it lacked rigor or mathematical substance, this changed completely once I learnt the measure-theoretic foundations of probability, which provide an intuitive, yet powerful way to describe our ideas about probability. I wish the reader all the best.

Yours falsely,

J.O.F

PS: The examples and questions were obtained either from my lecturers at King's College London, from [Sto13], or self-curated.

Contents

1 Basic Constructions and Definitions

In this first chapter we address the fundamental questions: how can one formulate a mathematical theory of Probability? What does it mean for an event to have a probability, what even is an event? It turns out that our best bet at formalising a theory of Probability is to ground ourselves in the realm of measure theory, the mathematical theory of size, for probabilities could roughly be thought of as the sizes of possible events.

1.1 Sigma Algebras

There is a question we would like to explore in our process of formulating a theory of Probability, say: if X is a random variable (whatever that may be) taking values in the interval $[0,1]$, and A is any subset of $[0,1]$, does the quantity

$$\mathbb{P}(X \in A)$$

even make sense? As we will see, assigning probabilities is intimately related with assigning sizes, and if we focus our attention momentarily on the simplest case of the real line, we may wish to define the size, or *measure* of a set $[a,b]$ to be $b-a$, and then see from here how can we extend this notion of size to other sets. This size function is known as the Lebesgue measure, and as it turns out, provided that we assume the Axiom of Choice, there exists subsets of the real line (called the Vitali sets, if you are interested) whose Lebesgue measure does not exist. The approach we take to fix this issue is to restrict our attention to certain sets, which do not present a problem when talking about sizes. We don't want to dive too much into the technicalities, but from now on, we will only attempt to give sizes to sets that lie within a special family of sets, this special family comes in the shape of a σ-algebra.

Definition 1.1.1 Let Ω be a set. A σ-Algebra $\mathscr{F}$ is a subset of 2^{Ω} that contains Ω, is closed under countable unions and closed under complements. A set $A \in \mathscr{F}$ is called measurable.

The intuition behind this definition is that if (A_n) is a sequence of measurable sets, we would like to be able to assign a measure to the countable union. Similarly, we would like to be able to assign measures to the intersection, this can be accomplished by repeated application of De Morgan's Law. In the context of probabilities, as we will see, we also want to be able to assign a probability to $\Omega \setminus A$ whenever we are able to assign a probability to A.

In a very similar spirit to many other areas in Mathematics, we can generate a special object from a smaller object, e.g: the ideal generated by an element of a ring, the free group generated by a set of elements, etc. In this case we can do something similar:

Theorem 1.1.1 Let E be an element of 2^{Ω}. There exists a smallest σ-algebra (with respect to inclusion), denoted $\sigma(E)$ that contains E, we say E generates $\sigma(E)$ or that $\sigma(E)$ is generated by E.

Proof. The construction is rather simple

$$\sigma(E) = \bigcap_{E \subseteq \mathscr{F}} \mathscr{F}$$

with $\mathscr{F}$ being a σ-algebra □

Example 1 Suppose that our underlying set Ω is actually a topological space (Ω, τ). There is in some sense a *distinguished* sigma algebra which may be attached to this space, namely the sigma algebra generated by the open sets. This is referred to as the Borel sigma algebra, denoted $\mathfrak{B}(\Omega)$. In the case of $\mathbb{R}$, it can be shown that

$$\mathfrak{B}(\mathbb{R}) = \sigma(\{(a,b] \mid -\infty < a < b < \infty\})$$

1.1.1 Practice Questions

Question 1 Let E_1, E_2 be Sigma algebras in the same space Ω, is $E = E_1 \cap E_2$ a Sigma algebra?

Solution. Yes. First of all, $\Omega \in E_1$ and $\Omega \in E_2$ so $\Omega \in E$. Now take a countable collection of events $(A_n)_{n\in\mathbb{N}}$ in E. This collection belongs to both E_1 and E_2, which are Sigma algebras themselves, so the countable union of this collection belongs both in E_1 and E_2 which means it belongs in E. Let $A \in E$. Again, A belongs to both E_1 and E_2, so A^c belongs to both E_1 and E_2 so A^c belongs to A as well. □

Remark 1 (Due to [Sto13]) Is the union of two Sigma algebras a Sigma algebra?

Proof. No. Consider the space $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Then the sets $\mathscr{F}_1 = \{\varnothing, \{\omega_1\}, \{\omega_2, \omega_3\}\}$ and $\mathscr{F}_2 = \{\varnothing, \{\omega_2\}, \{\omega_1, \omega_3\}\}$ are both Sigma algebras but their union is not. Indeed $\{\omega_1\}, \{\omega_2\} \in \mathscr{F}_1 \cup \mathscr{F}_2$ yet $\{\omega_1, \omega_2\} \notin \mathscr{F}_1 \cup \mathscr{F}_2$. □

Question 2 Let $E = \{A \subseteq [0,\infty) \mid \text{ either } A \text{ or } A^c \text{ is countable }\}$. Prove E is a Sigma Algebra

Proof. Here it is implicit that $\Omega = [0,\infty)$ so $\Omega^c = \varnothing$ which is countable, so $\Omega \in E$. Take $A \in E$. Thus either A or A^c is countable. If A^c is countable we have shown $A^c \in E$. On the other hand if it is A that's countable, then A^c also belongs to E because $(A^c)^c = A$ which is countable. So E is closed under complements. Now take a countable collection (A_n) of events. Each A_n is either countable or co-countable. If all the A_n are countable then we note that a countable union of countable sets is countable, i.e: $\bigcup_n A_n \in E$. On the other hand, if among the A_n there is some co-countable set, say A_j, then notice that $\left(\bigcup_n A_n\right)^c = \bigcap A_n^c \subset A_j^c$, which is countable. Thus the complement of $\bigcup_n A_n$ is a subset of a countable set, hence countable. □

Question 3 Which of the following E are σ-algebras?

1. $\Omega = \{a,b,c,d\}$ with $E = \{\varnothing, \{a,b,c,d\}, \{a,b\}, \{c,d\}\}$
2. $\Omega = \{1,2,\cdots,10\}$ with $E = \{A \subseteq \Omega \mid \text{ cardinality is even}\}$

3. $\Omega = \mathbb{R}$ with $E = \{(a,b) \mid a < b\} \cup \{\emptyset\}$

Solution. The first one is easily checked to be a Sigma-Algebra. The second is **not** a Sigma Algebra. Let's check what fails: Ω has cardinality 10 so its even, let $A \in E$. The cardinality of A^c is $10 - |A|$ which is even. However, say for instance $A = \{1,2\}$ and $B = \{2,3\}$. Both of these sets are in E, but $A \cup B = \{1,2,3\}$ which has odd cardinality, so E is not stable under union. Number three is not a Sigma Algebra either, because it is not stable under unions as well. For example: $(1,2) \in E$ and $(3,4) \in E$ but the union of these two intervals is not an interval nor the empty set. $\square$

Question 4 Show that $\mathfrak{B}(\mathbb{R})$ contains the singletons.

Proof. Let $a \in \mathbb{R}$. Construct the family of open sets $A_n = \left(a - \frac{1}{n}, a + \frac{1}{n}\right)$. Open sets are the generators of $\mathfrak{B}(\mathbb{R})$, so (A_n) belongs to the Borel Algebra, which in addition is stable under countable intersections and $\bigcap_{n \in \mathbb{N}} A_n = \{a\}$. $\square$

Question 5 Let $\Omega = \mathbb{R}$. Is $E = \{\bigsqcup_{1 \leq i \leq n} (a_i, b_i]\}$ a ring? Is it a Sigma Algebra?

Solution. First notice that by setting $a = b$ then $E \ni \emptyset = (a,b]$. Let $A = (a,b] \in E$. Then $A^c = (-\infty, a] \cup (b, \infty]$. So E is stable under complements. Stability under union is built into the definition, so it is a ring. However, this definition does not allow for countable unions so it is not a Sigma Algebra. $\square$

Question 6 There exists a sigma-algebra $\mathscr{F} = \{\emptyset, A_1, \cdots, A_5, \Omega\}$ of exactly seven sets.

Solution. No. Since $\mathscr{F}$ must be closed under complements, for every $A_i \in \mathscr{F}$, A_i^c must also be in $\mathscr{F}$. No set satisfies $A = A^c$ so the fact that there is an odd number of sets in $\mathscr{F}$ means that at least there is one set whose complement is not in $\mathscr{F}$. $\square$

Question 7 Show sigma algebras are closed under countable intersections

Solution. Let (A_n) be a countable sequence of measurable sets in $\mathscr{F}$. Then for each A_n, $A_n^c \in \mathscr{F}$. Thus $\bigcup_{n=1}^{\infty} A_n^c \in \mathscr{F}$. Finally, we take the complement of this set and see that

$$\mathscr{F} \ni \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c = \bigcap_{n=1}^{\infty} A_n$$

$\square$

Question 8 Show that closed intervals $[a,b]$ for $-\infty < a < b < \infty$ are in $\mathfrak{B}(\mathbb{R})$ and that the Borel measure of these intervals is $b-a$.

Solution. The interval $\left(a-\frac{1}{n},b\right]$ belongs to $\mathfrak{B}(\mathbb{R})$ by definition. Since $\mathfrak{B}(\mathbb{R})$ is closed under countable intersections, then the set $\bigcap_n \left(a-\frac{1}{n},b\right] = [a,b] \in \mathfrak{B}(\mathbb{R})$. Observe that:

$$\left(a+\frac{1}{n},b\right] \subseteq [a,b] \subseteq \left(a-\frac{1}{n},b\right]$$

Thus by monotonicity of the measure:

$$\mu\left(\left(a+\frac{1}{n},b\right]\right) \leq \mu([a,b]) \leq \mu\left(\left(a-\frac{1}{n},b\right]\right)$$

This in turn says that:

$$b-a-\frac{1}{n} \leq \mu([a,b]) \leq b-a+\frac{1}{n}$$

By the sandwich theorem taking $n \to \infty$, we get the desired result. Alternatively, use Theorem 1.2.1. □

Question 9 Is the set of rational numbers $\mathbb{Q}$ Borel?

Solution. Yes. Let (r_n) be your favorite enumeration of the rational numbers. Then

$$\mathbb{Q} = \bigcup_{n=1}^{\infty} \{r_n\}$$

Observe that each $\{r_n\}$ is a singleton, and hence measurable (see Question 4 for a proof of this claim). □

1.2 Measures, Measure Spaces, Measurable Functions

We are now ready to formalise what we mean mathematically by size. A size function is called a measure.

Definition 1.2.1 Let Ω be a set and E a σ-algebra on Ω. A function $\mu : E \to [0,\infty]$ is a measure if

1. Given a sequence of disjoint measurable sets (A_n):

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$$

2. $\mu(\varnothing) = 0$

If $\mu(\Omega)$ is normalised, that is to say $\mu(\Omega) = 1$, then we refer to μ as a probability measure, and we talk about the triplet $(\Omega, \mathscr{E}, \mu)$ as a measure (or probability) space.

For notational convenience, whenever we are too lazy to indicate that a sequence of measurable sets is pairwise disjoint, we will write

$$\bigsqcup_n A_n$$

for their union instead of the usual union sign.

Example 2 The two canonical examples of measures are:

1. The Lebesgue measure λ on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. Which satisfies $\lambda([a,b]) = b - a$. The existence of this measure is a highly non-trivial fact, and its construction requires the use of something called Carethedory's Extension Theorem. Developing this would take too much valuable time away from our study of probability, so we delegate it to classic texts such as [Wil91].

2. The counting measure c on $(\mathbb{Z}, 2^{\mathbb{Z}})$. Which is defined by $c(S) = \#S$ for any $S \subseteq \mathbb{Z}$.

Theorem 1.2.1 (Properties of Measures) -

1. **Continuity:** Slogan - measure of the limit is limit of the measure
   (a) If (A_n) is an increasing sequence then $\lim \mu(A_n) = \mu\left(\bigcup A_n\right)$
   (b) If (A_n) is a decreasing sequence **of finite measure** ($\mu(A_N) < \infty$ for some N) then $\lim \mu(A_n) = \mu\left(\bigcap A_n\right)$

2. **Subadivitivity:** Slogan - triangle inequality

$$\mu\left(\bigcup_{n\geq 1} A_n\right) \leq \sum_{n\geq 1} \mu(A_n)$$

Proof of 1.a. The technique is to split each A_n into some disjoint sets that unite to A_n, thus being able to use additivity of μ. Define $B_n = A_n \setminus A_{n-1}$. Notice that

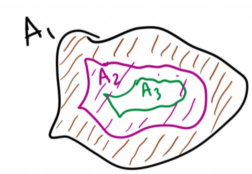
$$\mu(A_n) = \mu\left(\bigcup_{k=1}^{n} B_k\right) = \sum_{k=1}^{n} B_k \to \sum_{k=1}^{\infty} B_k = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right)$$

□

Proof of 1.b. We work by reducing to the case of increasing unions. Moreoever, notice that we are going to have to use the assumption of finite measure at some point, that is to say, we will need to take complements. This motivates the construction of the following increasing sequence:

$$B_n = A_1 \setminus A_n$$

Here's a graphical rendition of this



Now that we have an increasing sequence, we know that

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

On the one hand:

$$\mu(B_n) = \mu(A_1) - \mu(A_n)$$

On the other hand:

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_1 \cap A_n^c = A_1 \cap \left(\bigcap_{n=1}^{\infty} A_n\right)^c = A_1 \setminus \bigcap_{n=1}^{\infty} A_n$$

Now you can put this together. □

Proof of 2. We prove the simpler case $\mu(A\cup B) \le \mu(A) + \mu(B)$. The rest follows by a similar argument. Observe that one can write $A \cup B$ as $A \cup B \setminus A$. From Question 14 we know that since $B \setminus A \subseteq B$, then $\mu(B \setminus A) \le \mu(B)$. Combining this finishes the claim. □

Definition 1.2.2 (Independence) Two events $A, B \in \mathscr{F}$ are said to be independent if $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$. A collection of events (A_n) is said to be mutually independent if $\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$ for all valid index sequences $i_1, i_2, \cdots, i_k$. A collection of sigma algebras $\mathscr{F}_n \subseteq \mathscr{F}$ is said to be independent if any sequence (A_n) with $A_i \in \mathscr{F}_i$ of events is independent.

Remark 2 Note that if A and B are independent events, then A and B^c are independent events. Indeed:

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) \qquad \text{draw a diagram to convince yourself}$$

which in turnn is equal to $\mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B)$ by independence, hence

$$P(A \cap B^c) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(B^c)$$

Proposition 1.2.1 (Properties of measurable functions) Let $(\Omega, \mathscr{E}, \mu)$ be a measure space. The following properties hold:

1. If f, g are measurable functions into $\mathbb{R}$ with appropriate domains, then $f + g$, $f \cdot g$ and $f \circ g$ are measurable.
2. If (f_n) is a sequence of measurable functions, the limits $\liminf f_n, \limsup f_n$, if they exist, are measurable. Moreover, if the pointwise limit $\lim f_n$ exists, it is measurable.

Proof. We show each case separately.

1. (Stable under addition). The goal is to show that for any $a \in \mathbb{R}$, the set $(f+g)^{-1}(-\infty, a] \in E$. Or in other words, that the set $\{x \in \mathbb{R} \mid f(x) + g(x) \leq a\} \in E$. The key is to observe that since the rationals are dense in $\mathbb{R}$, one has:

$$\{x \in \mathbb{R} \mid f(x) + g(x) \leq a\} = \bigcup_{q \in \mathbb{Q}} \{x \in \mathbb{R} \mid f(x) \leq q\} \cap \{x \in \mathbb{R} \mid g(x) \leq a - q\} = \bigcup_{q \in \mathbb{Q}} f^{-1}(q) \cap g^{-1}(a-q)$$

   Which is a countable union of a collection of finite intersections of measurable sets, hence measurable.

2. (Stable under multiplication). Observe that

$$f \cdot g = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right]$$

$\square$

1.2.1 Random Variables

We now reach one of the central definitions of the theory, that of a random variable. Note that the key defining property is that we want to be able to assign probabilities to events of the form $\{X \in A\}$ for some measurable set A, it is therefore unsurprising that our definition of a random varibale is the following:

Definition 1.2.3 (Measurable functions, random variables) Let $(\Omega, \mathscr{E})$ and $(\Lambda, \mathscr{F})$ be measurable spaces and let $X : \Omega \to \Lambda$ be a function such that for any $E \in \mathscr{F}$, one has $X^{-1}(E) \in \mathscr{E}$. We then refer to X as a measurable function. If the measurable spaces are in particular probability spaces then we say X is a random variable.

One may introduce now a cheeky concepts, by which we generate σ-algebras that make functions measurable.

Definition 1.2.4 (σ-algebra generated by a function) Let Ω be a set, (Λ, F) be a measurable space, and $X:\Omega\to\Lambda$ be a function. The σ-algebra generated by X, $\sigma(X)$, is the smallest σ-algebra that makes X a measurable function. In other words

$$\sigma(X)=\sigma(\{X^{-1}(B):B\in F\})$$

Definition 1.2.5 (Independence of Random Variables) A collection (X_i) of random variables is independent the family of σ-algebras $(\sigma(X_i))$ is independent.

It is important to check that this new definition coincides with the elementary notion of independent random variables. Suppose that $X_1,\cdots,X_n$ are independent real-valued random variables. Then the σ-algebras $\sigma(X_i)$ are independent, and as such, given any sequence $X_{i_1},\cdots,X_{i_k}$, and Borel-measurable sets $B_i,\cdots,B_k$, we have that $(X_{i_j}^{-1}(B_j))_j$ is a sequence of events where the j^{th} entry belongs to $\sigma(X_{i_k})$ and as such

$$\mathbb{P}(X_{i_1}\in B_1,\cdots,X_{i_k}\in B_k)=\prod_j \mathbb{P}(X_{i_j}^{-1}\in B_j)$$

Thus we see that the usual notion of independence is recovered.

1.2.2 Distributions

Let $(\Omega,\mathscr{E},\mathbb{P})$ be a measure space and $X:\Omega\to\mathbb{R}$ be a measurable function. We can use this data to construct a new measure on $\mathbb{R}$. This new measure is defined as

$$\mathbb{P}_X:\mathfrak{B}(\mathbb{R})\to[0,1] \qquad A\mapsto\mathbb{P}\circ X^{-1}(A)$$

and gives rise to usual notions in probability, such as

Definition 1.2.6 With X as before, the **distribution** of X, $F_X:\mathbb{R}\to[0,1]$ is given by

$$F_X(x)=\mathbb{P}_X(-\infty,x]$$

Proposition 1.2.2 (Properties of the distribution function) Let $X:\Omega\to\mathbb{R}$ be a random variable. Then $F_X(x)$ is non-decreasing, right-continuous, $\lim_{x\to-\infty}F_X(x)=0$, and $\lim_{x\to\infty}F_X(x)=1$

Proof.

1. Non-decreasing: Let $x_1\le x_2$. Then obviously $(-\infty,x_1]\subseteq(-\infty,x_2]$ and as such, $X^{-1}(-\infty,x_1]\subseteq X^{-1}(-\infty,x_2]$. The rest follows automatically.
2. Right-continuity: We see that

$$F_X(x+1/n)-F_X(x)=\mathbb{P}(X\in(x,x+1/n])=\mathbb{P}_X(x,x+1/n]$$

Since $\mathbb{P}_X$ is a measure, we can apply its (decreasing) continuity and observe that

$$\lim_{n\to\infty}\mathbb{P}_X(x,x+1/n]=\mathbb{P}_X\left(\bigcap_{n=1}^{\infty}(x,x+1/n]\right)=\mathbb{P}_X(\emptyset)=0$$

3. The limits: $\lim_{x\to\infty}F_X(x)=\mathbb{P}(X\in(-\infty,x))=\lim_{x\to\infty}\mathbb{P}_X(-\infty,x)$. Once again, using continuity of the measure, we see that this limit is equal to

$$\mathbb{P}_X\left(\bigcup_{x=1}^{\infty}(-\infty,x)\right)=\mathbb{P}_X(\mathbb{R})=1$$

in a similar manner,

$$\lim_{x\to-\infty}\mathbb{P}_X(-\infty,x)=\mathbb{P}_X\left(\bigcap_{x=1}^{\infty}(-\infty,-x])\right)=\mathbb{P}_X(\emptyset)=0$$

□

1.2.3 Practice Questions

Question 10 Show that in a probability space, if A is an event, then $\mathbb{P}(A^c)=1-\mathbb{P}(A)$. Moreoever, show that

$$\mathbb{P}(A_1\cup A_2)=\mathbb{P}(A_1)+\mathbb{P}(A_2)-\mathbb{P}(A_1\cap A_2)$$

Solution. We use σ-aditivity:

$$1=\mathbb{P}(\Omega)=\mathbb{P}(A\sqcup A^c)=\mathbb{P}(A)+\mathbb{P}(A^c)$$

Now we show the second statement, the strategy is the same, split $A_1\cup A_2$ into disjoint sets, and use aditivity of the measure:

$$\mathbb{P}(A_1\cup A_2)=\mathbb{P}((A_1\setminus A_2)\sqcup(A_1\cap A_2)\sqcup(A_2\setminus A_1))$$

Now we simply note that

$$\mathbb{P}(A)=\mathbb{P}((A\setminus B)\sqcup(A\cap B))\implies\mathbb{P}(A\setminus B)=\mathbb{P}(A)-\mathbb{P}(A\cap B)$$

which implementing into the equation above gives the desires result. □

Question 11 If $\mathbb{P}_1,\mathbb{P}_2$ are two probability measures on the same space, such that $\mathbb{P}_1,\mathbb{P}_2$ agree for some $A\in\mathscr{F}$ is it true that $\mathbb{P}_1(\Omega\setminus A)=\mathbb{P}_2(\Omega\setminus A)$? What about two general measures μ_1,μ_2 such that $\mu_1(\Omega)=\mu_2(\Omega)$

Solution. For the case of probability measures the answer is yes. Indeed:

$$\mathbb{P}_1(\Omega\setminus A)=\mathbb{P}_1(\Omega)-\mathbb{P}_1(A)=1-\mathbb{P}_1(A)=1-\mathbb{P}_2(A)=\mathbb{P}_2(\Omega\setminus A)$$

This does not hold for the case of general measures, indeed, take μ_1 to be the counting measure on $\Omega=\mathbb{Z}$, and $\mu_2=2\mu_1$, let $A=\mathbb{Z}\setminus\{0\}$. Then $\mu_1(A)=\mu_2(A)=\infty$ and $\mu_1(\Omega)=\mu_2(\Omega)=\infty$ but $\mu_1(\Omega\setminus A)=\mu_1(\{0\})=1$

but $\mu_2(\{0\}) = 2$. □

Question 12 Let $(\Omega, \mathscr{E}, \mathbb{P})$ be a probability space and $A, B \in \mathscr{E}$ be such that:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

Does it follow $A \cap B = \varnothing$.

Solution. No. Let $([0,1], [0,1] \cap \mathfrak{B}(\mathbb{R}), \mathbb{P})$ be the Lebesgue Measure space on $[0,1]$. Let $A = \mathbb{Z} \cap [0,1]$ and $B = \mathbb{Q} \cap [0,1]$. Then:

$$\mathbb{P}(A \cup B) = \mathbb{P}(\mathbb{Q} \cap [0,1]) = 0 = 0 + 0 = \mathbb{P}(\mathbb{Z} \cap [0,1]) + \mathbb{P}(\mathbb{Q} \cap [0,1])$$

However $A \cap B = \mathbb{Z} \cap [0,1]$ □

Question 13 In Question 2 we showed that

$$E = \{A \subseteq [0,\infty) \mid \text{ either } A \text{ or } A^c \text{ is countable } \}$$

was a sigma algebra. Show now that the function $\mathbb{P} : E \to \mathbb{R}$ given by

$$\mathbb{P}(A) = \begin{cases} 0 & A \text{ is countable} \\ 1 & A^c \text{ is countable} \end{cases}$$

Defines a probability measure on $([0,\infty), E)$

Solution. Naturally $\mathbb{P}(\varnothing) = 0$ for $\varnothing$ is countable. Similarly, $\mathbb{P}([0,\infty)) = 1$ for its complement, the empty set, is countable. Let (A_n) now be a collection of disjoint sets in E. Let us show additivity of the measure in the two cases.

1. If all (A_n) are countable, then it is clear that that their union is a countable set, and as such aditivity is satisfied.

2. Now suppose there are some sets in (A_n) that have countable complements. Observe that if two sets A and B have countable complements, then $(A \cap B)^c$ is countable, which means that $A \cap B$ is nonempty, thus they can't be disjoint. Therefore this case is reduced to the case where exactly one A_k has a countable complement. Now we that $\left(\bigcup_{n=1}^{\infty} A_n\right)^c \subseteq A_k^c$ which is countable, this means that the union of all A_n has a countable complement, and as such, aditivity is also satisfied.

□

Question 14 Let μ be a measure and suppose A, B are two measurable sets such that $A \subseteq B$. Show $\mu(A) \le \mu(B)$.

Proof. Write $B = A \cup B \setminus A$. Obviously $A \cap B \setminus A = \varnothing$ so

$$\mu(B) = \mu(A) + \mu(B \setminus A)$$

and thus the claim follows. $\square$

Question 15 Let $\{A_1, \cdots, A_k\}$ be a family of events. Does pairwise independence imply mutual independence?

Proof. No. Consider a box of four tickets labeled $112, 121, 211, 222$. Choose one ticket at random and let $A_1 = \{1 \text{ occurs in the first place}\}$, $A_2 = \{1 \text{ occurs in the second place}\}$, $A_3 = \{1 \text{ occurs in the third place}\}$. One observes that $\mathbb{P}(A_i \cap A_j)$ for $i \neq j$ is equal to $1/4$ which is also equal to $\mathbb{P}(A_i)\mathbb{P}(A_j)$. Thus having pairwise independence. However $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0 \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$. $\square$

Question 16 Does the relation $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ imply mutual independence?

Proof. No. Let $\Omega = \{1, \cdots, 8\}$ where each outcome has probability $1/8$. Let $A = \{1, 2, 3, 4\}$ and $B = C = \{1, 5, 6, 7\}$. Then $\mathbb{P}(A \cap B \cap C) = 1/8 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. However, B and C are obviously not independent. $\square$

Question 17 Does the function $\mu : \mathfrak{B}(\mathbb{R}) \to \mathbb{R}$

$$\mu(A) = \begin{cases} 0 & A \text{ is bounded} \\ 1 & A \text{ is unbounded} \end{cases}$$

define a measure?

Solution. Hell no man! Suppose it did. Then by σ-additivity:

$$1 = \mu(\mathbb{Z}) = \mu\left(\bigsqcup_{n \in \mathbb{Z}} \{n\}\right) = \sum_{n \in \mathbb{Z}} \mu(\{n\}) = 0$$

$\square$

Question 18 Let $f : \Omega \to \mathbb{R}$ be measurable. Show f^2 is measurable.

Proof. The general technique of these questions is as follows: We have to show that for any $\mathscr{E} \in \mathfrak{B}(\mathbb{R})$, one has $f^{-1}(\mathscr{E})$ is measurable in Ω. It is sufficient to show this claim holds on the generators of the Borel Sigma Algebra. In this case, it is most convenient to work with the generators of $\mathfrak{B}(\mathbb{R})$ of the form $(-\infty, a]$. Notice that f^2 is precisely $g \circ f$ where $g(x) = x^2$. We distinguish the following cases:

1. If $a < 0$. Then $g^{-1}(-\infty, a] = \varnothing$ so the preimage of our original function is equal to $f^{-1}(\varnothing)$. Recall f is measurable and the empty set is a measurable set so in this case the preimage of the generator under f^2 is measurable.

2. If $a \geq 0$. Then $g^{-1}(-\infty, a] = [-\sqrt{a}, \sqrt{a}]$. This is measurable in $\mathfrak{B}(\mathbb{R})$ so the preimage of this set under f is measurable in Ω. In this case, the preimage of the generator under f^2 is also measurable.

$\square$

Question 19 (Summer 2020 Question 1 (e)) Let $(\Omega, \mathscr{E}, \mu)$ be a measure space, and let $A \subseteq \Omega$ be any set. Let $E = \{\varnothing, A, A^c, \Omega\}$ be the Sigma Algebra on the space. Show that $f : \Omega \to \mathbb{R}$ is measurable if and only if it is constant on A and constant on A^c

Proof. First we suppose it is constant on A, say $f(\omega) = y_1$ for all $\omega \in A$ and constant on A^c, say $f(\omega) = y_2$ for all $\omega \in A^c$. Now we aim to show f is measurable by considering the preimage of generators $(-\infty, a]$ under f. Without loss of generality say $y_1 \leq y_2$. We will actually only show this with $y_1 < y_2$. The case for $y_1 = y_2$ is easier. Consider the following cases:

1. $a < y_1$. Then $f^{-1}(-\infty, a] = \varnothing$. Which is measurable

2. $y_1 \leq a < y_2$. Then $f^{-1}(-\infty, a] = A$. Which is measurable

3. $a \geq y_2$. Then $f^{-1}(-\infty, a] = \Omega$. Which is measurable

Now conversely, suppose f is a measurable function. Fix some $\omega_0 \in A$. We will show that for any $\omega \in A$, $f(\omega) = f(\omega_0) =: y_0$. Consider the preimage $f^{-1}(\{y_0\})$. By hypothesis, this preimage is either: $\varnothing, A, A^c$ or Ω. It cannot be the empty set, because ω_0 maps to y_0. We can assume that f takes at least two different values, otherwise it is constant in the entire domain and in particular constant in A and A^c. Thus the preimage $f^{-1}(y_0)$ cannot be Ω. It follows it must be exclusively A or A^c. It can't be A^c because there is an element of A, namely ω_0 that maps to y_0. Therefore $f^{-1}(y_0) = A$. In other words, f is constantly equal to y_0 on A. This argument repeats symmetrically to show that f is constant in A^c. $\square$

Question 20 (2019 Q1 (d)) Let $X, Y : \mathfrak{B}(\mathbb{R}) \to \mathbb{R}$ be random variables. Explain why if

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(X \leq y)$$

Then for all $A, B \in \mathfrak{B}(\mathbb{R})$ one has.

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

Attempt. Any set $A \in \mathfrak{B}(\mathbb{R})$ is generated from intervals $(-\infty, a]$ via unions, intersections and complements. As an example,

$$\mathbb{P}(X \in [a,b], Y \in [c,d]) = \mathbb{P}(X \leq b, Y \in [c,d]) - \mathbb{P}(X \leq a, Y \in [c,d])$$

and now repeat with $Y \in [c,d]$. $\square$

Question 21 (2019 Q1 (e)) Show that a random variable is independent of itself if and only if it is almost surely constant.

Proof. Let $A \in \mathfrak{B}(\mathbb{R})$. Clearly

$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A, X \in A).$$

X is independent of itself if and only if this product splits. I.e: If and only if $\mathbb{P}(X \in A) = \mathbb{P}(X \in A)^2$. Which holds if and only if $\mathbb{P}(X \in A) = 0$ or 1. Hence the X is almost surely constant. $\square$

Question 22 (Homework sheet 4) Suppose X and Y are two independent random variables such that $\mathbb{P}(X = Y) = 1$. Show that $X = Y = c$ for some constant c. Hint: study $\mathbb{P}(X \leq x, Y \leq x)$ in two different ways.

Solution. Observe that if $Y \leq x$ then either $X \leq x$ or $Y > X$. Hence

$$\mathbb{P}(Y \leq x) \leq \mathbb{P}(X \leq x) + \mathbb{P}(Y > X)$$

but by hypothesis, this last quantity is zero, therefore $\mathbb{P}(Y \leq x) = \mathbb{P}(X \leq x)$. By a symmetrical argument $\mathbb{P}(X \leq x) \leq \mathbb{P}(Y \leq x)$ and as such the two distributions are equal. Now we can finalise the argument by noting that:

$$\mathbb{P}(X \leq x, Y \leq x) = \mathbb{P}(X \leq x)$$

Here we have used simple conditional probability argument. Moreover,

$$\mathbb{P}(X \leq x, Y \leq x) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq x) = \mathbb{P}(X \leq x)^2$$

and as such

$$\mathbb{P}(X \leq x)^2 = \mathbb{P}(X \leq x)$$

So $\mathbb{P}(X \leq x)$ is either zero or one, which corresponds to X (and Y) being almost surely constant. $\square$

Question 23 (2019 Q1 (f)) Let ϵ_1, ϵ_2 be independent random variables with

$$\mathbb{P}(\epsilon_i = 1) = \mathbb{P}(\epsilon_i = -1) = 1/2 \qquad i = 1, 2$$

Are the random variables ϵ_1 and $\epsilon_1\epsilon_2$ independent?

Solution. Note that $\mathbb{P}(\epsilon_1\epsilon_2=1)=\mathbb{P}(\epsilon_1\epsilon_2=-1)=1/2$, so:

$$\mathbb{P}(\epsilon_1=1,\epsilon_1\epsilon_2=1)=\mathbb{P}(\epsilon_1=1,\epsilon_2=1)=\mathbb{P}(\epsilon_1=1)\mathbb{P}(\epsilon_2=1)=\frac{1}{4}=\mathbb{P}(\epsilon_1=1)\mathbb{P}(\epsilon_1\epsilon_2=1)$$

Similarly:

$$\mathbb{P}(\epsilon_1=1,\epsilon_1\epsilon_2=-1)=\mathbb{P}(\epsilon_1=1,\epsilon_2=-1)=\mathbb{P}(\epsilon_1=1)\mathbb{P}(\epsilon_2=-1)=\frac{1}{4}=\mathbb{P}(\epsilon_1=1)\mathbb{P}(\epsilon_1\epsilon_2=-1)$$

The other cases follow by symmetry. They are independent. □

Question 24 (Specimen 2024) Can every measurable function be expressed as a finite sum of indicator functions?

Solution. No. Take for example $f(x)=x$. Since f is continuous, it is measurable. A function that can be expressed as a finite sum of indicator functions takes finitely many values. □

Question 25 (Mock 2022) Let $(E,\mathscr{E},\mu_1)$ and $(G,\mathscr{G},\mu_2)$ be two measure spaces. If a function $f:E\to G$ is $\mathscr{E}-\mathscr{G}$ measurable and bijective, does it follow that $f^{-1}:G\to E$ is $\mathscr{G}-\mathscr{E}$ measurable?

Solution. I was going to write hell no, but its not actually an immediately obvious answer. Take $E=G=\mathbb{R}$ and $\mathscr{E}=\mathfrak{B}(\mathbb{R})$ and $\mathscr{G}=\{\varnothing,\mathbb{R}\}$. Then take $f(x)=x$. This is clearly $\mathscr{E}-\mathscr{G}$ measurable, but not the other way around. □

Question 26 (Mock 2022) Let $(E,\mathscr{E},\mu_1)$ and $(G,\mathscr{G},\mu_2)$ be two measure spaces and $f:E\to G$ be $\mathscr{E}-\mathscr{G}$ measurable. Let $\mathscr{E}'\subseteq\mathscr{E}$ and $\mathscr{G}'\subseteq\mathscr{G}$ be two sub-σ algebras. Does it follow:

1. f is $\mathscr{E}'-\mathscr{G}$ measurable?
2. f is $\mathscr{E}-\mathscr{G}'$ measurable?

Solution.

1. Hell no. Pick $\mathscr{E}=\mathscr{G}=\mathfrak{B}(\mathbb{R})$, $E=G=\mathbb{R}$, and $f(x)=x$. This one's obviously $\mathscr{E}-\mathscr{G}$ measurable. Now consider the sub-σ-algebra $\mathscr{E}'=\{\varnothing,E\}\subseteq\mathscr{E}$. Now f is clearly not $\mathscr{E}'-\mathscr{G}$ measurable.
2. Yes. Obvious.

□

Question 27 (Homework sheet 1) Show that if (A_n) is a sequence of events in a probability space, then

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty}(1 - \mathbb{P}(A_n))$$

Solution. This is a very easy direct computation:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} A_n^c\right)^c\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n^c\right) \geq 1 - \sum_{n=1}^{\infty} \mathbb{P}(A_n^c)$$

as required. □

Question 28 (Homework sheet 2) Show that if an event A has $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, then A is independent to any other event B.

Solution. We start with the case $\mathbb{P}(A) = 0$. Then we observe that since $A \cap B \subseteq A$, one has that $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0$. Therefore $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$ as required. (Observe the crucial requirement that we are in a probability space in this last step!). Now suppose that $\mathbb{P}(A) = 1$, then A^c is independent to B, and as such:

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B)$$

But one easily observes that $\mathbb{P}(A^c \cap B) = \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. The right hand side reads $(1 - \mathbb{P}(A))\mathbb{P}(B)$. Rearranging gives the required result. □

Question 29 (Homework sheet 2) Show that if A, B, C are independent events, then $A \cup B$ and C are independent.

Solution. Pretty straightforward once you see the idea. For convenience, let $\star = (A \cup B) \cap C$. Observe that $\star = (A \cap C) \cup (B \cap C)$. This is just some easy boolean algebra. Now we can see that

$$\mathbb{P}(\star) = \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C) - \mathbb{P}(A \cap B \cap C)$$

But since we are given all these events are independent, we may nicely rewrite this as:

$$\mathbb{P}(\star) = \mathbb{P}(C)[\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)]$$

But noticing that the thing inside the square brackets is nothing but $\mathbb{P}(A \cup B)$, we finish the claim. □

2 Convergence of Measurable Functions

Definition 2.0.1 (Modes of convergence of measurable functions) Let $(\Omega, \mathscr{E}, \mu)$ be a measure space, (f_n) be a sequence of measurable functions, and f a measurable function.

1. $f_n \to f$ **almost everywhere** if $\mu\big(\{x \in \Omega \mid f_n(x) \not\to f(x)\}\big) = 0$. In words: it converges pointwise everywhere except on a set of measure zero. In the context of a probability space we say **almost sure** convergence.

2. $f_n \to f$ **in measure** if for all $\epsilon > 0$, one has $\mu\big(\{x \in \Omega \mid |f_n(x) - f(x)| > \epsilon\}\big) \to 0$ as $n \to \infty$. In words: fixing any tolerance, the measure of the set of points that give an error greater than the tolerance converges to 0. In the context of a probability space we say convergence **in probability**

3. If we are talking about random variables, we say that $X_n \to X$ **in distribution** if the distribution function of X_n converges to the distribution function of X at each point of continuity.

Remark 3 (Common probability space? (Seen on 2021)) Note that for almost sure and convergence in probability we made reference to the random variables living on the same probability space, however, for the case of convergence in distribution, we only care about pointwise convergence of two distribution functions, therefore the underlying probability spaces could be distinct.

Theorem 2.0.1 We have the following convergence implications

$$(f_n \xrightarrow{\text{unif}} f) \longrightarrow (f_n \xrightarrow{\text{pt}} f) \longrightarrow (f_n \xrightarrow{\text{a.e}} f) \overset{\mu(\Omega)<\infty}{\dashrightarrow} (f_n \xrightarrow{\mu} f)$$

Proof. $(f_n \xrightarrow{pt} f) \implies (f_n \xrightarrow{a.e} f)$: The set of points x where $f_n(x) \not\to f(x)$ is precisely the empty set which has zero-measure.

$(f_n \xrightarrow{a.e} f) \implies (f_n \xrightarrow{\mu} f)$ if $\mu(\Omega) < \infty$: Fix an ϵ. The strategy is as follows. We will construct a decreasing sequence of sets, that in the limit capture the behaviour of almost everywhere convergence and the expression is made in terms of the measure. So define the set

$$A_N = \{x \in \Omega \mid \exists n \geq N \text{ with } |f_n(x) - f(x)| > \epsilon\}$$

Notice that if $x \in A_{N+1}$ then it also belongs to A_N, so we can consider the decreasing limit $\bigcap_{N=1}^{\infty} A_N$. Here you use the hypothesis of $\mu(\Omega)$ being finite, to guarantee that every A_N has finite measure so that you can use the fact that $\lim_{n\to\infty} \mu(A_n) = \mu\big(\bigcap_{N=1}^{\infty} A_N\big) = \mu\big(\{x \in \Omega \mid |f_n(x) - f(x)| > \epsilon \text{ infinitely often}\}\big)$ which is zero by assumption of almost everywhere convergence. $\square$

Example 3 (Breaking the opposite directions) Let us present some examples to break the reverse directions of these implications.

1. A sequence of random variables that converges **in distribution but not in probability**: Consider the probability space $([0,1], \mathfrak{B}(\mathbb{R}), m)$ where m is the Lebesgue measure, and define $X_n : [0,1] \to \mathbb{R}$ by:
$$X_{2n}(x) = x \qquad X_{2n-1}(x) = 1 - x$$
We check that they converge in distribution to the uniform distribution. Indeed,
$$m(X_{2n} \leq x) = m([0,x]) = x \qquad m(X_{2n-1} \leq x) = m(1-x, 1) = x$$
thus for all n, X_n has the uniform distribution, in particular it converges to the uniform distribution. Now we check that there is no convergence in probability. Indeed, let $\epsilon > 0$ be given and let X be the random variable $X(x) = x$ then:
$$m(\{x \mid |X_n - X| > \epsilon\}) = \begin{cases} 0 & n \text{ is even} \\ \text{some constant } c(\epsilon) > 0 & \text{otherwise} \end{cases}$$
one sees that this quantity does not converge to zero.

2. A sequence of random variables that converge **in probability but not almost surely**: Define the sequence of IIDs
$$X_n \sim Bern(1/n)$$
This sequence converges in probability to the zero random variable. Indeed, $\mathbb{P}(|X_n| > \epsilon) = 1/n \to 0$. However, let A_N be the event $\{|X_N| > \epsilon\}$. These events are independent and $\sum_{N=1}^{\infty} \mathbb{P}(A_N) = \sum_{N=1}^{\infty} 1/N = \infty$ so by BC2, $\mathbb{P}(\{A_N \text{ infinitely often }\}) = 1$. So X_n does not converge almost surely to zero.

3. A sequence of measurable functions that converges almost everywhere but not pointwise: $\exp(-nx)$. This converges to the zero function everywhere except when $x = 0$. Take a suitable measure space to make this work now.

4. The example above also works in $(0,1)$ to have an example of a sequence of functions that converge pointwise but not uniformly.

Remark 4 (On the importance of $\mu(E) < \infty$ for a.e $\implies$ μ convergences) Observe that the Theorem needed $\mu(E) < \infty$ for almost everywhere convergence to imply convergence in measure. Indeed, consider the sequence of functions $f_n(x) = \mathbf{1}_{[n,2n]}(x)$ on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), m)$ where m is the Lebesgue measure. Then

1. Obviously f_n converges pointwise to zero so in particular we have $f_n \to f \equiv 0$ a.e.

2. However, let $0<\epsilon<1$ be given, then

$$m\big(\{x \mid |f_n(x)-f|>\epsilon\}\big)=m([n,2n])=n$$

thus showing that we don't have convergence in measure.

Question 30 (2023) Prove that if $f=h$ a.e and $h=g$ a.e, then $f=g$ a.e

Proof. Let $x\in\{\omega\in\Omega \mid f(\omega)\neq g(\omega)\}=:X_1$. Then either:

1. $f(x)=h(x)$ and $g(x)\neq h(x)$
2. $f(x)\neq h(x)$ and $g(x)=h(x)$
3. $f(x)\neq h(x)$ and $g(x)\neq h(x)$

Thus: $X_1\subseteq\Big(\{\omega\in\Omega \mid f(\omega)=h(\omega)\}\cap\{\omega\in\Omega \mid h(\omega)\neq g(\omega)\}\Big)\cup$
$\Big(\{\omega\in\Omega \mid f(\omega)\neq h(\omega)\}\cap\{\omega\in\Omega \mid h(\omega)=g(\omega)\}\Big)\cup\Big(\{\omega\in\Omega \mid f(\omega)\neq h(\omega)\}\cap\{\omega\in\Omega \mid h(\omega)\neq g(\omega)\}\Big)$
By countable sub-additivity and monotonicity, the measure of X_1 is less than or equal to the sum of the measures of the three sets on the right-hand side. Each of these sets is in turn a subset of $\{\omega\in\Omega \mid h(\omega)\neq g(\omega)\}$, $\{\omega\in\Omega \mid f(\omega)\neq h(\omega)\}$ and $\{\omega\in\Omega \mid h(\omega)\neq g(\omega)\}$ respectively, which by assumption all have measure 0. Once again, by monotonicity, it follows that the overall sum is less than or equal to zero. Measures being non-negative means that $\mu(X_1)=0$. This explanation can be made much shorter and still be valid. □

Question 31 (Spin-off 2023) Let $(\Omega,\mathscr{F},\mu)$ be $(\mathbb{Z},\mathscr{P}(\mathbb{Z}),\mu_{\text{counting}})$. Give sufficient and necessary conditions for two functions $f,g:\Omega\to\mathbb{R}$ to be equal almost everywhere.

Proof. f,g are equal almost everywhere if and only if they differ on a set of measure zero. The only set of measure zero with respect to the counting measure is the empty set. Thus it is sufficient and necessary for f and g to take the same values on the integers. □

Question 32 (2023) Give a condition on two measurable sets A and B so that the condition holds if and only if $\mathbf{1}_A=\mathbf{1}_B$ almost everywhere.

Proof. One must have $\mu(A\setminus B)=\mu(B\setminus A)=0$. □

Theorem 2.0.2 Convergence $X_n\to X$ in probability implies convergence in distribution

Proof. We recall the following important Lemma: For random variables X, Y, $a \in \mathbb{R}, \epsilon > 0$

$$\{Y \leq a\} \subseteq \{X \leq a + \epsilon\} \cup \{|X - Y| > \epsilon\}$$

With this, observe the following two facts:

1. $\mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$
2. $\mathbb{P}(X_n \leq x - \epsilon) \leq \mathbb{P}(X \leq x) + \mathbb{P}(|X_n - X| > \epsilon)$

Combine them to get:

$$\mathbb{P}(X_n \leq x - \epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

If x is a point of continuity of $F_X(x)$, then we take $\epsilon \to 0$ and as $n \to \infty$:

$$\lim_{n \to \infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x) \leq \lim_{n \to \infty} \mathbb{P}(X_n \leq x)$$

□

Theorem 2.0.3 The converse to Theorem 2.0.2 holds when X is a constant random variable.

Proof. By assumption:

$$\mathbb{P}(X_n \leq x) \to \mathbf{1}_{\{x \geq c\}}(x)$$

Let $\epsilon > 0$ be given. It is quite clear that

$$\mathbb{P}(\{x : |X_n - c| > \epsilon\}) = \mathbb{P}(\{x : X_n < c - \epsilon\} \cup \{x : X_n > c + \epsilon\})$$

By subaditivity, we then see that

$$\mathbb{P}(\{x : |X_n - c| > \epsilon\}) \leq \mathbb{P}(\{x : X_n \leq c - \epsilon\}) + 1 - \mathbb{P}(\{x : X_n \leq c + \epsilon\})$$

As $n \to \infty$, this latter quantity is precisely

$$\mathbf{1}_{\{x \geq c\}}(c - \epsilon) + 1 - \mathbf{1}_{\{x \geq c\}}(c + \epsilon) = 0 + 1 - 1 = 0$$

as required. □

Question 33 (Specimen 2024) Let U_n be Uniformly distributed in $\{1, \cdots, n\}$. Show that $\frac{1}{n} U_n$ converges in distribution to $U \sim \text{Unif}([0,1])$

Proof. Note:

$$\mathbb{P}\left(\frac{1}{n}U_n \le x\right) = \mathbb{P}(U_n \le xn) = \frac{\lfloor xn \rfloor}{n} \to x$$

□

Question 34 (Specimen 2024) With U_n as in the previous question, does $\frac{1}{n^\alpha}U_n$ converge in distribution?:

1. $\alpha < 1$
2. $\alpha > 1$

Solution.

1. If $\alpha < 1$ then: $\mathbb{P}(U_n \le n^\alpha x) = \frac{\lfloor n^\alpha x \rfloor}{n} \to 0$ as $n \to \infty$. Regardless of x. This cannot be the case since we require distribution functions to approach 1 as $x \to \infty$
2. Otherwise, as $n \to \infty$, $\mathbb{P}(U_n \le n^\alpha x)$ will be 1 because $n^\alpha x$ becomes eventually greater than n and $\mathbb{P}(U_n \le n) = 1$. This holds for any $x \in [0,1]$ so if $\alpha > 1$ the random variable converges to the zero random variable in distribution.

□

Question 35 (2022 B3 + Mock 2022 + Homework shet 4) Determine whether the following sequences of functions $\mathbb{R} \to \mathbb{R}$ converge: almost everywhere, in measure. Find their limits if they exist with respect to each mode of convergence.

1. $f_n(x) = \exp(-x^2/2n^2)$
2. $f_n(x) = \exp(-n^2x^2/2)$
3. $f_n(x) = \frac{1}{\sqrt{n}}\exp(-x^2/n)$
4. $f_n(x) = \mathbf{1}_{[-\log(n+1), \log(n+1)]}$
5. $f_n(x) = \frac{\mathbf{1}_{[-n^2,n^2](x)}}{\log(n+1)}$
6. X_n is a sequence of random variables such that $X_n \sim \text{Unif}[1, 1+1/n]$
7. $f_n(x) = n\mathbf{1}_{[1/n, 2/n]}$
8. $f_n(x) = \mathbf{1}_{\mathbb{R}\setminus[n,n+1]}(x)$
9. $f_n(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 1/n & x \notin \mathbb{Q} \end{cases}$

Solution.

1. The first function converges pointwise to the function $f(x) \equiv 1$. Therefore it converges a.e. However, it does not converge in measure. Indeed, fixing any ϵ and regardless of how large n is, for values of $|x|$ large enough, the distance between 1 and $f_n(x)$ will be larger than ϵ. So the measure of the set of points where the discrepancy is larger than ϵ is always infinite in measure.

2. The second function converges obviously a.e to zero. And given any $\epsilon > 0$, the only point that is always further away than ϵ from zero is precisely the point $x = 0$. Thus we have convergence in measure.

3. Observe that the numerator converges pointwise to 1, so the whole function converges pointwise to 0. Moreover, since the numerator is bounded, the denominator squishes it down towards zero and as such eventually it will lie below the ϵ threshold away from zero.

4. This function converges pointwise to the constant function $f(x) \equiv 1$ just like the first case. But the convergence in measure does not happen because the set of points that disagree more than ϵ with 1 is precisely $(-\infty, -log(n+1)) \cup (\log(n+1), \infty)$ which has measure infinite.

5. It is easy to see that this function converges pointwise to the zero function, hence almost surely. Now let us verify whether it converges in measure. Indeed, given any $\epsilon > 0$ the set of points where f_n is larger than ϵ has initially positive measure, but as $n \to \infty$, f_n falls below the threshold and as such, no points lie above the ϵ threshold.

6. By hypothesis, $\mathbb{P}(X_n \in [1, 1+1/n]) = 1$ or in other words, $\mathbb{P}(|X_n - 1| > 1/n) = 0$. Now we claim that $X_n \to 1$ almost surely. We shall employ the Borel-Cantelli Lemma. Let $\epsilon > 0$ be given then by our observation, for n large enough $\mathbb{P}(|X_n - 1| > \epsilon) = 0$, therefore, the sum

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - 1| > \epsilon)$$

is finite and as such $X_n \to X$ almost surely.

7. We see that given any $x \in \mathbb{R}$, $f_n(x) \to 0$, so we have pointwise convergence and in particular, almost everywhere convergence. We also see that the set of points of disagreement between f_n and 0 has measure $1/n$, hence we have convergence in measure.

8. Here we have pointwise convergence to 1, indeed, this sequence of functions is just a constant line at $y = 1$ with a dip to zero between $[n, n+1]$. However, the points of disagreement between 1 and f_n has constant length of 1 therefore we do not have convergence in measure.

9. We see that $f_n \to \mathbf{1}_{\mathbb{Q}}$ pointwise, so in fact $f_n \to 0$ almost everywhere, for $\mathbb{Q}$ has measure zero. By a similar argument, we see that the set of points of disagreement between $f_n(x)$ and 0 is precisely the rationals so it also converges in measure.

□

Question 36 (2021 A1) Show that if $f_n : E \to \mathbb{R}$ is a sequence of measurable functions whose a.e limit is f, then f is measurable.

Proof. We recall that a function $f : E \to \mathbb{R}$ is measurable if and only if $f^{-1}(a, \infty)$ is a measurable set for all $a \in \mathbb{R}$. Now we aim to show that $\limsup_{n\to\infty} f_n$ is measurable. First we show that $\sup_n f_n(x)$ is a measurable function. Recall that the meaning of $\sup_n f_n(x)$ is:

$$\sup_n f_n(x) = \sup\{f_n(x) \mid n \in \mathbb{N}\}$$

and as such, the set:

$$\{x \mid \sup_n f_n(x) > c\} = \{x \mid \sup\{f_n(x) \mid n \in \mathbb{N}\} > c\} = \bigcup_{n=1}^{\infty} \{x \mid f_n(x) > c\}$$

this latter set is measurable since each f_n is measurable. Similarly one shows that $\inf_n f_n$ is measurable. Now we see that

$$\limsup f_n = \inf_n \sup_{m \geq n} f_n$$

is measurable as required. □

Question 37 (2019 3 (c)) With X_n as in the previous question, give necessary and sufficient conditions on $\alpha \in \mathbb{R}$ such that $X^n/n^\alpha \to 0$ in probability.

Solution. $X^n/n^\alpha \to 0$ in Probability if given any $\epsilon > 0$, we have that:

$$P = \mathbb{P}\left(\left|\frac{X_n}{n^\alpha}\right| > \epsilon\right) \to 0$$

Noting that $X_n \geq 0$ and rearranging inside the probability sign gives:

$$P = 1 - \mathbb{P}(X_n \leq n^\alpha \cdot \epsilon) = 1 - \int_0^{n^\alpha \cdot \epsilon} \frac{\mathrm{d}x}{(x+1)^2} = 1 - \left[-\frac{1}{x+1}\right]_0^{n^\alpha \cdot \epsilon} = \frac{1}{n^\alpha \cdot \epsilon + 1}$$

This last quantity converges to 0 as $n \to \infty$ if and only if $\alpha > 0$. □

Question 38 (Homework sheet 7) Show that if $X_n \to X$ in distribution, then $X_n^2 \to X^2$ in distribution.

Solution. Observe that

$$\mathbb{P}(X_n^2 \leq x) = \mathbb{P}(X_n \leq \sqrt{x}) - \mathbb{P}(X_n \leq -\sqrt{x})$$

Naturally we take $x \geq 0$ as X_n^2 cannot take negative values. Thus if F_X is discontinuous as $\pm\sqrt{x}$, then F_{X^2} is discontinuous at x. Suppose then that F_{X^2} is continuous at x, we may write

$$\mathbb{P}(X_n^2 \leq x) \to \mathbb{P}(X \leq \sqrt{x}) - \mathbb{P}(X \leq -\sqrt{x}) = \mathbb{P}(X^2 \leq x)$$

□

Question 39 (Homework sheet 4) Let (U_n) be a family of IID Uniform distributions on $[0,1]$. Define $M_n = \max\{U_1, \cdots, U_n\}$. Show that

1. $M_n \to 1$ in probability.
2. $n(1-M_n)$ converges in distribution. Identify the limit.

Solution. To show the first part, let $\epsilon > 0$ be given. We are interested in studying

$$\mathbb{P}(|M_n - 1| > \epsilon)$$

By drawing a diagram or something, convince yourself that this probability is the same as the probability of each U_i being less than or equal to $1-\epsilon$. That is to say:

$$\mathbb{P}(|M_n - 1| > \epsilon) = \prod_{i=1}^{n} \mathbb{P}(U_i \leq 1-\epsilon) = (1-\epsilon)^n \to 0$$

Here we have used the fact that (U_i) are IIDs in order to introduce the product.

For the second task, don't ask me how I thought of this, we consider the probability $\mathbb{P}(n(1-M_n) \geq x)$ instead of $\leq x$. Rearranging, this is equal to

$$\mathbb{P}(M_n \leq) = \prod_{i=1}^{n} \mathbb{P}\left(U_i \leq 1 - \frac{x}{n}\right) = \left(1 - \frac{x}{n}\right)^n \to \exp(-x)$$

thus showing that $n(1-M_n)$ converges in distribution to an exponential random variable with parameter 1. □

3 The Borel–Cantelli Lemmas

Now we have perhaps our first interesting result in this measure-theoretic framework of probability, and this result has to do with how can we determine the long-run behavior of a sequence of events.

Definition 3.0.1 Let (A_n) be a sequence of events. We say that (A_n) **holds**:

1. **Infinitely Often**:

$$\{A_n \text{ holds infinitely often}\} := \{x : x \in A_n \text{ for infinitely many } n\} := \limsup A_n$$

2. **Eventually:**

$$\{A_n \text{ holds eventually}\} := \{x : x \in A_n \text{ for all } n \text{ large enough}\} := \liminf A_n$$

Proposition 3.0.1 Let (A_n) be a sequence of events. Then we may express **events**:

1. **Infinitely Often**:
$$\{A_n \text{ holds infinitely often}\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$
2. **Eventually:**
$$\{A_n \text{ holds eventually}\} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

Theorem 3.0.1 (Borel-Cantelli Lemma 1) Let (A_n) be a sequence of events. If $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\{A_n \text{ i.o }\}) = 0$

Proof. Slogan - If the sum converges, then the summands must eventually go to zero. Now stare at the definition of infinitely often and observe that if $\left(\bigcup_{n=k}^{\infty} A_n\right)$ holds for all $k > 0$, we can write for any N:

$$\mathbb{P}(\{A_n \text{ i.o}\}) \leq \mathbb{P}\left(\bigcup_{k=N}^{\infty} A_n\right) \leq \sum_{k=N}^{\infty} \mathbb{P}(A_k)$$

Now letting $N \to \infty$ we get that $\mathbb{P}(\{A_n \text{ i.o}\}) \to 0$ $\square$

Theorem 3.0.2 (BC2) Let (A_n) be a sequence of **independent** events such that $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$. Then $\mathbb{P}(\{A_n \text{ i.o }\}) = 1$

3.1 Application of Borel-Canteli

The BC Lemmas are very useful to prove or disprove that sequences of random variables converge or not almost surely.

Proposition 3.1.1 Let (X_n) be a sequence of random variables and X another random variable on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Suppose that

$$\forall \epsilon > 0 \quad \sum_{n=1}^{\infty} \mathbb{P}(\{|X_n - X| > \epsilon\}) < \infty$$

Then $X_n \to X$ almost surely

Proof. We have that for any $\epsilon > 0$, the event $\{|X_n - X| > \epsilon$ infinitely often $\}$ has probability zero. That is to say, for any $\epsilon > 0$, with probability 1, there exists an N such that whenever $n > N, |X_n - X| < \epsilon$. $\square$

Proposition 3.1.2 Suppose (X_n) is a sequence of IID random variables such that for any $\epsilon > 0$, the following sum diverges:

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) = \infty$$

Then $X_n \not\to X$ almost surely.

Proof. By BC2, given any $\epsilon > 0$, the event $\{|X_n - X| > \epsilon\}$ holds infinitely often so in particular, $\mathbb{P}(X_n \to X) \neq 0$ $\square$

3.2 Practice Questions

Question 40 Let (A_n) be a sequence of independent events such that: $\mathbb{P}(A_n) = \frac{1}{(n+1)^2}$. Moreover, let (B_n) be a sequence of events defined by $B_n = \{$ exactly one of $A_1, \cdots, A_n$ holds$\}$

1. Find $\mathbb{P}(\{A_n \text{ i.o}\})$
2. Show $\mathbb{P}(B_n) \geq \frac{1}{3}\mathbb{P}(A_1^c \cap A_2^c \cap \cdots \cap A_n^c)$
3. Establish whether $\sum_{n=1}^{\infty} \mathbb{P}(B_n)$ is finite or infinite.
4. Decide whether $\mathbb{P}(\{B_n \text{ i.o}\})$ is equal to $0, 1$ or something in between.

Proof. -

1. We use Borel-Cantelli.
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$
therefore $\mathbb{P}(\{A_n \text{ i.o}\}) = 0$.

2. Observe that
$$B_n = \{A_1 \cap A_2^c \cap \cdots \cap A_n^c\} \cup \cdots$$
therefore
$$\mathbb{P}(B_n) \geq \mathbb{P}(\{A_1 \cap A_2^c \cap \cdots \cap A_n^c\})$$
since the events are assumed to be independent, we see
$$\mathbb{P}(B_n) \geq \mathbb{P}(A_1)\mathbb{P}(\{A_2^c \cap \cdots \cap A_n^c\}) = \frac{1}{4}\mathbb{P}(\{A_2^c \cap \cdots \cap A_n^c\})$$
Observe that
$$\mathbb{P}(\{A_1^c \cap A_2^c \cap \cdots \cap A_n^c\}) = (1 - \mathbb{P}(A_1))\mathbb{P}(\{A_2^c \cap \cdots \cap A_n^c\}) = \frac{3}{4}\mathbb{P}(A_2^c \cap \cdots \cap A_n^c\})$$
combining these two last lines gives the desired result.

3. One has
$$\sum_{n=1}^{\infty}\mathbb{P}(B_n)\geq\frac{1}{3}\sum_{n=1}^{\infty}\mathbb{P}(A_1^c\cap\cdots\cap A_n^c)=\frac{1}{3}\sum_{n=1}^{\infty}\prod_{i=1}^{n}\left(1-\frac{1}{(i+1)^2}\right)=$$
$$=\frac{1}{3}\sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{i(i+2)}{(i+1)^2}=\frac{1}{3}\sum_{n=1}^{\infty}\frac{n+2}{2(n+1)}=\infty.$$

4. We notice that $\mathbb{P}(\{B_n \text{ i.o}\})=\mathbb{P}(A_i \text{ exactly once } i\geq 1)$. On the one hand,
$$\mathbb{P}(A_i \text{ exactly once } i\geq 1)=1-\mathbb{P}(A_1^c\cap A_2^c\cdots)-\mathbb{P}(A_i \text{ at least twice })\leq 1-\mathbb{P}(A_1^c\cap A_2^c\cdots)=1-\frac{1}{2}$$
On the other hand
$$\mathbb{P}(A_i \text{ exactly once } i\geq 1)=\mathbb{P}(\{A_1\cap A_2^c\cap\cdots\}\cup\cdots)\geq\mathbb{P}(\{A_1\cap A_2^c\cap\cdots\})=\frac{1}{6}\quad\square$$

Question 41 Consider a sequence of independent experiments, where the n^{th} experiment has a probability of $n^{-\alpha}$ where $0<\alpha<1$ of succeeding.

1. What is the probability that infinitely many successes will be observed?
2. Fixing a $k\geq 2$, what is the probability that k successes in a row will be observed infinitely often?

Solution. The first question is an immediate consequence of the Borel Cantelli Lemmas (and so is the second one, but requires slightly more thought). Denote by A_n the event that the n^{th} experiment is a success.

1. Observe that
$$\sum_{n=1}^{\infty}\mathbb{P}(A_n)=\sum_{n=1}^{\infty}n^{-\alpha}$$
Since α is too small, this decay is too slow and so the sum diverges. Since the events are assumed to be independent, it follows that $\mathbb{P}(\{A_n \text{ i.o }\})=1$.

2. Now we have to put some more thought. Observe that we are interested in the probability of the following event B_n occurring infinitely often:
$$B_n=\{A_n,A_{n+1},\cdots,A_{n+k-1} \text{ all succeed}\}$$
We also make the observation that since all A_i are independent, we can bound above and below the probability of B_n occurring. Since
$$\mathbb{P}(B_n)=\prod_{i=0}^{k-1}(n+i)^{-\alpha}$$
we deduce that
$$(n+k-1)^{-k\alpha}<\mathbb{P}(B_n)<n^{-k\alpha}$$

the important thing is that the asymptotic behaviour is the same both sides. Therefore from this we deduce that the sum to infinity of $\mathbb{P}(B_n)$ will converge if and only if $k\alpha > 1$. That is to say:

(a) If $\alpha > 1/k$ then the probability of observing k successes in a row infinitely often is 1.

(b) Otherwise its zero.

□

Question 42 (2019 Q3 (e-f)) Let (X_n) be IID random variables with density function $f_X(x) = \frac{1}{(x+1)^2}\mathbf{1}_{\{x\geq 0\}}$. Show that

$$M_n := \min_{1\leq i\leq n} X_n \to 0$$

almost surely. Then show:

$$nM_n \to M$$

In distribution, where M is a random variable you need to identify.

Solution. As before, the strategy is to show that for any $\epsilon > 0$. One has:

$$\sum_{n=1}^{\infty} \mathbb{P}(|M_n| > \epsilon) < \infty$$

Observe that $\mathbb{P}(M_n > \epsilon) = \mathbb{P}(X_1 > \epsilon, X_2 > \epsilon, \cdots, X_n > \epsilon) = [\mathbb{P}(X_1 > \epsilon)]^n$. This last equality is since each X_i is independent. We compute that:

$$\mathbb{P}(X_1 > \epsilon) = 1 - \int_0^{\epsilon} \frac{\mathrm{d}x}{(1+x^2)} = \frac{1}{1+\epsilon}$$

Since $\epsilon > 0$ and therefore $(1+\epsilon) > 1$ it follows that the following sum converges:

$$\sum_{n=1}^{\infty} \mathbb{P}(|M_n| > \epsilon) = \sum_{n=1}^{\infty} \frac{1}{(1+\epsilon)^n} < \infty$$

Finishing the proof of the first part. To solve the second part we need to identify the pointwise limit of the function $F_{nM_n}(x) = \mathbb{P}(nM_n \leq x) = \mathbb{P}(M_n \leq x/n) = 1 - \mathbb{P}(M_n \geq x/n) = 1 - [\mathbb{P}(X_1 \geq x/n)]^n = 1 - \frac{1}{(1+x/n)^n} \to 1 - e^{-x}$. Thus we see that nM_n converges in distribution to an exponential random variable of parameter $\lambda = 1$ □

Question 43 (Summer 2021 Question A2 (c) (iii)) Let $X_n \to X$ almost surely. Suppose $\mathbb{E}[X_n] = 0$ for all $n \geq 1$. Does it follow $\mathbb{E}[X] = 0$?

Solution. The answer is no. Consider the sequence of random variables X_n defined by

$$X_n = \begin{cases} -1 \text{ with probability } \frac{n^2-1}{n^2} \\ n^2 - 1 \text{ with probability } \frac{1}{n^2} \end{cases}$$

We check that $\mathbb{E}[X_n] = -1 \cdot \frac{n^2-1}{n^2} + (n^2-1) \cdot \frac{1}{n^2} = 0$. We check that $X_n \to -1$ almost surely. Indeed, we see that $\sum_{n=1}^{\infty} \mathbb{P}(|X_n + 1| > \epsilon) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$ so by the Borel Cantelli Lemma, we see that $X_n \to -1$ almost surely. However, the expectation of the constant random variable -1 is not 0. □

Question 44 (Summer 2021 Question A2 (c) (iv)) Let $X_n \to X$ almost surely, with $|X_n| \leq 1$ and $\mathbb{E}[X_n] = 0$ for all $n \geq 1$. Does it follow that $\mathbb{E}[X] = 0$?

Solution. The answer is yes. We can apply the Dominated Convergence Theorem with dominating function 1. Note that $\int_\Omega 1 d\mathbb{P} = 1$ so the function 1 is integral, so the Theorem applies. It follows that we can interchange the integrals with the limits and we deduce that $\mathbb{E}[X] = 0$ □

4 Tail Events

We are interested in looking at the tail events, that is to say, events that depend on the asymptotic behavior of a sequence of random variables.

Definition 4.0.1 (Tail σ-algebra) Let (X_n) be a sequence of random variables, define $\tau_n = \sigma(X_n, X_{n+1}, \cdots)$, i.e, the sigma algebra generated by X_n, and so on. (For a refresher on what this means please see the Analysis Intermezzo). Then we define the **tail σ-algebra** to be

$$\tau = \bigcap_n \tau_n$$

Remark 5 (Intuitive Explanation of Tail σ-algebras) Suppose (X_n) is a sequence of random variables. For convenience, we can think of the sequence as some process indexed by a time variable n. We can explain the meaning of τ_T in very simple terms: it is simply the events that can be determined purely by knowing the value that the process takes for values of time larger than T. That is to say, whatever happened before T has no impact on the probability of the event occurring. The events that therefore lie in τ, are those that are determined by what occurs in the **far distant future**, that is to say, if you alter the values of (X_n) for any finite number of time units, the event will remain unaffected.

Proposition 4.0.1 (Examples of tail events) Let (X_n) be a sequence of real valued random variables. The events

$$\{x : \lim X_n(x) \text{ exists}\},$$

$$\left\{x : \sum X_n(x) \text{ converges}\right\},$$

$$\left\{x : \lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} X_i(x) \text{ exists}\right\}$$

all belong to τ.

Proof. 1. We note that

$$\{x : \lim X_n \text{ exists}\} = \{x : \limsup X_n < \infty\} \cap \{x : \liminf X_n > -\infty\} \cap g^{-1}(\{0\})$$

where $g(x) = \limsup f_n(x) - \liminf f_n(x)$. The task thus is reduced to showing that $\limsup$ and $\liminf$ are τ-measurable. To do so, we simply note that for any given $p \in \mathbb{N}$:

$$\liminf X_n = \liminf X_{n+p}$$

It is now easy to see that each function in the sequence $(X_{n+p})_n$ is $\sigma(X_p, X_{p+1}, \cdots)$ measurable. Invoking the result from Measure Theory that if (f_n) is a sequence of measurable functions, then so is $\liminf$ and $\limsup$ of f_n, we finish the claim.

2. Follows immediately from 1, using $S_n = \sum_{i=1}^{n} X_i$.

3. Follows immediately from previous work.

$\square$

Theorem 4.0.1 (Kolmogorov $1-0$ law) Let (X_n) be a sequence of independent random variables. Let τ be their tail sigma algebra. Then, given any $A \in \tau$, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Consequently, if Ξ is a τ-measurable random variable, then Ξ is almost surely constant. That is to say, $\mathbb{P}(\Xi = c) = 1$ for some c.

Proof. The strategy is to show that τ is independent of itself as a sigma algebra. It will then follow that any event will be independent of itself, and the desired result will follow.

Define $\varphi_n = \sigma(X_1, \cdots, X_n)$. We have that φ_n is generated by the π-system of events of the form

$$\{X_1 \le x_1, \cdots X_n \le x_n\}$$

whereas τ_n is generated by events of the form

$$\{X_n \le x_{n+1}, \cdots X_{n+k} \le x_{n+k}\}$$

where k is some natural number. We use the fact that if the generating π-systems are independent, then the σ-algebras are independent. Hence, since each X_n is independent, it follows that φ_n and τ_n are independent. The fact that $\tau \subseteq \tau_n$ for each n automatically implies that φ_n and τ are independent. We now claim that $\varphi_\infty := \sigma(X_n : n \in \mathbb{N})$ is independent of τ. We note that the π-system $\bigcup_n \varphi_n$ generates φ_∞. Moreover, $\bigcup_n \varphi_n$ and τ are independent as π-systems due to a preceding argument: indeed, if $A \in \bigcup_n \varphi_n$, then $A \in \varphi_k$ for some k but we know that φ_k and τ are independent. Thus $\sigma\left(\bigcup_n \varphi_n\right) = \varphi_\infty$ and τ are independent as σ-algebras. We now finalise the claim by noting that $\tau \subseteq \sigma\left(\bigcup_n \varphi_n\right) = \varphi_\infty$. Therefore τ is independent of itself. $\square$

Remark 6 In light of Proposition 4.0.1, we see that for example, if (X_n) is a sequence of independent random variables:

1. $\mathbb{P}(\lim X_n$ exists) is zero or one.
2. $\mathbb{P}\left(\sum X_n \text{ converges}\right)$ is zero one.

And so on.

5 Integration

5.1 Integration of simple functions

Definition 5.1.1 Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. A function $f : \Omega \to \mathbb{R}$ is said to be simple if f takes finitely many values, or in other words:

$$f(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_n}(\omega)$$

Where each $A_n \in \mathscr{F}$ and $a_i \geq 0$. Then the integral of f with respect to μ is defined to be

$$\int_\Omega f \,\mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(A_n)$$

It is also common notation to write $\mu(f)$ for the integral.

Remark 7 Note that by definition, simple functions are measurable.

5.2 Integration of measurable functions

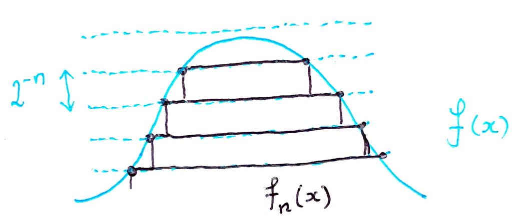
Definition 5.2.1 Let $f : \Omega \to \mathbb{R}$ be non-negative measurable function. The integral of f is defined as

$$\int_\Omega f \,\mathrm{d}\mu = \sup \left\{ \int_\Omega g \,\mathrm{d}\mu \,\middle|\, g \text{ simple, and } g \leq f \right\}$$

Remark 8 This definition is not nonsensical because every non-negative measurable function can be **monotonically**, uniformly approximated by simple functions. For example, for a given measurable $f : \Omega \to \mathbb{R}$, we have

$$f_n(\omega) = 2^{-n} \lfloor 2^n f \rfloor$$

A result that justifies that this approximation of functions yields the correct integral is the following:



Theorem 5.2.1 (Monotone Convergence Theorem) Let (f_n) be a monotone sequence of measurable functions, whose almost-everywhere pointwise is f. This means:

1. $0 \leq f_i(x) \leq f_j(x) \qquad \forall i \leq j$
2. $f_n(x) \to f(x) \qquad$ for $\mu-$almost all $x \in \Omega$

Then:

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu$$

Example 4 (Monotonicity is crucial) The sequence

$$f_n(x) = n\mathbf{1}_{[0,1/n]}$$

has μ-almost everywhere pointwise limit $f = 0$, yet this convergence is not monotone, for example $f_1(3/4) \geq f_2(3/4)$

Definition 5.2.2 A measurable function $f : \Omega \to \mathbb{R}$ is integrable if:

$$\int_\Omega f^+ \, \mathrm{d}\mu < \infty \quad \& \quad \int_\Omega f^- \, \mathrm{d}\mu < \infty$$

where $f^+ = \max(f(x), 0)$ and $f^-(x) = -\min(f(x), 0)$. We then define

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu$$

A whole new world opens now

Question 45 (Homework sheet 7) Show that if f is a measurable function, then f^+ (and hence f^-) are measurable.

Proof. We shall check whether $(f^+)^{-1}(-\infty, a]$ is Borel. If $a < 0$ then there is nothing to show, since f^+ does not output negative values. Otherwise

$$(f^+)^{-1}(-\infty, a] = (f^+)^{-1}[0, a] = f^{-1}[0, a]$$

and this latter set is obviously measurable. To show measurability of f^- we do an identical argument. □

Theorem 5.2.2 (Properties of the integral) Let f, g be integrable functions and $\alpha, \beta \geq 0$ be constants. Then

1. The integral is linear:

$$\int_\Omega \alpha f + \beta g \, \mathrm{d}\mu = \alpha \int_\Omega f \, \mathrm{d}\mu + \beta \int_\Omega g \, \mathrm{d}\mu$$

2. The integral is "almost-everywhere" inequality preserving: if $f \le g$ almost everywhere, then

$$\int_\Omega f \, d\mu \le \int_\Omega g \, d\mu$$

3. The integral is "non-degenerate", i.e:

$$\int_\Omega f \, d\mu = 0 \text{ if and only if } f = 0 \quad a.e$$

A whole new world opens up now, because this last Theorem tells us that integrable functions actually form a linear space, the L^1 space. We have a natural generalisation

Definition 5.2.3 Let $(\Omega, \mathscr{E}, \mu)$ be a measure space. Then we define $\mathscr{L}^p(\Omega, \mathscr{E}, \mu)$ as

$$\mathscr{L}^p(\Omega, \mathscr{E}, \mu) = \{f \text{ measurable } : \mu(|f|^p) < \infty\}$$

This is a linear space, but it turns out that it (almost) is something much better. We have the following result

Theorem 5.2.3 (Minkowski's Inequality) Let $f, g \in \mathscr{L}^p(\Omega, \mathscr{E}, \mu)$, then

$$\mu(|f+g|^p)^{1/p} \le \mu(|f|^p)^{1/p} + \mu(|g|^p)^{1/p}$$

This looks awfully lot like a triangle-type inequality, and so we may wonder whether $\mathscr{L}^p(\Omega, \mathscr{E}, \mu)$ is actually a normed space, with the norm of a function f given by $\mu(|f|^p)^{1/p}$. After all, we also have that $\mu(|\alpha f|^p)^{1/p} = |\alpha|\mu(|f|^p)^{1/p}$, but the problem we have is that non-degeneracy fails, indeed, it is easy to see that there will be functions for which the integral is zero, yet the function is not the zero function. However, in view of Theorem 5.2.2, we have that actually, if the integral is zero, then the function must be zero almost everywhere, so there turns out to be an easy fix to turn $\mathscr{L}^p(\Omega, \mathscr{E}, \mu)$ into a normed space. The solution is to take a quotient.

Definition 5.2.4 Let $(\Omega, \mathscr{E}, \mu)$ be a measure space, we define the linear space

$$L^p(\Omega, \mathscr{E}, \mu) = \mathscr{L}^p(\Omega, \mathscr{E}, \mu)/\sim$$

where $f \sim g$ if $f = g$ almost everywhere.

Although this space is no longer a function space, but rather a space of equivalence classes, it is easy to see that this forms a linear space under appropriate notions of addition and scalar multiplication. Moreover, we now can define a norm on this space, where $\|f\| = \mu(|f|)^{1/p}$. What's even better, is that this space enjoys some great metric properties, most importantly

Theorem 5.2.4 The space $L^p(\Omega, \mathscr{E}, \mu)$ is complete with respect to its norm.

In other words, $L^p(\Omega, \mathscr{E}, \mu)$ is a Banach space.

5.2.1 Practice Questions

Question 46 True or False: a simple function integrates to zero if and only if it is equal to zero almost everywhere. What if instead of simple, the function is integrable?

Solution. True. Indeed, since simple functions are non-negative by definition, it must be that if $f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_n}$ and some a_i is not zero, then it must be that $\mu(A_i)$ is zero. Thus we conclude that f is almost everywhere zero. For the second part the answer is False, indeed: the function $f = \mathbf{1}_{[0,1]} - \mathbf{1}_{[-1,0]}$ is integrable and integrates to zero.

□

Question 47 Is every measurable function integrable? If f is measurable and integrable, is $|f|$ integrable? If (f_n) is a sequence of measurable integrable functions, and $f = \lim_{n\to\infty} f_n$ a.e, is f integrable?

Solution. -

1. No. We know continuous functions are measurable, yet the simplest continuous function $f : \mathbb{R} \to \mathbb{R}$, the constant function $f \equiv a > 0$ is not integrable on the whole of $\mathbb{R}$.
2. Yes. If f is integrable then both its negative and positive parts integrate to a finite number so the sum of two finite numbers is finite.
3. No. Consider the sequence of functions: $f_n = \mathbf{1}_{[-n,n]}$ living in $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$ where λ is the Lebesgue measure. Clearly, f_n is measurable and integrable. Its limit is simply the constant function 1. Is 1 Integrable? Hell no

□

Remark 9 Can a function be Riemann Integrable but not Lebesgue integrable? Yes! Consider the function $f(x) = \frac{\sin(x)}{x}$. The limit

$$\lim_{N\to\infty} \int_{2\pi}^{K} f(x)\,\mathrm{d}x < \infty$$

in the Riemann sense, but f^+ and f^- have infinite Lebesgue integral so f is not Lebesgue integrable.

Question 48 (2022) Are the following true or false?

1. If $f : E \to \mathbb{R}$ is integrable then f^2 is integrable.
2. If $f : E \to \mathbb{R}$ is measurable and f^2 is integrable, then f is integrable.
3. if $f : [0,1] \to \mathbb{R}$ is measurable where $[0,1]$ is equipped with the Borel measure, and f^2 is integrable, then f is integrable.

Solution. -

1. False. Take $E=(0,1)$ and $f(x)=\frac{1}{\sqrt{x}}$. We know that $\int_0^1 f(x)<\infty$ so f is integrable. However, $\int_0^1 \frac{1}{x}\,\mathrm{d}x$ is not finite.

2. False. Take $E=(1,\infty)$ and $f(x)=1/x$. Clearly f is continuous hence measurable, and $f^2(x)$ is integrable. Yet f is not integrable.

3. We refer to the Cauchy-Schwarz inequality for random variables:

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X]\mathbb{E}[Y]$$

Observe that in this setting, f corresponds to a random variable. We can take $X=f$ and $Y=1$. Observe that due to the finiteness of $[0,1]$, $\mathbb{E}[Y]=1$. Therefore

$$\mathbb{E}[f]^2=\left(\int_{[0,1]} f\,\mathrm{d}x\right)^2 \leq \int_{[0,1]} f^2\,\mathrm{d}x<\infty$$

□

5.3 Convergence of Integrals

We have already seen one tool to study convergence of integrals, namely the Monotone Convergence Theorem, there is another useful tool.

Theorem 5.3.1 (Dominated Convergence Theorem) Let (f_n) be a sequence of measurable functions with almost-everywhere pointwise limit f. Moreover, suppose there exists a measurable function g such that $|f_n|\leq g$ almost everywhere for all n, and g is integrable. Then for all n, f_n and f are integrable and moreover

$$\int_\Omega f_n\,\mathrm{d}\mu \to \int_\Omega f\,\mathrm{d}\mu$$

5.3.1 Practice Questions

Example 5 (The case where Ω has finite measure) Suppose (f_n) and f are as in Theorem 5.3.1 and moreover, suppose that f_n is uniformly bounded, i.e: there exists a constant $C>0$ such that for all n, $|f_n|\leq C$ almost-everywhere. Noticing that

$$\int_\Omega C\,\mathrm{d}\mu = C\mu(\Omega)<\infty$$

gives that

$$\int_\Omega f_n\,\mathrm{d}\mu \to \int_\Omega f\,\mathrm{d}\mu$$

Example 6 (Example of DCT usage) Evaluate the limit $n \to \infty$:

$$\int_0^\infty \exp(-nx)\,\mathrm{d}x$$

Solution. This is an easy application of the DCT: Observe that $0 < \exp(-nx) < \exp(-x)$ for all n, and moreover,

$$\int_0^\infty \exp(-x)\,\mathrm{d}x = 1 < \infty$$

It follows that

$$\int_0^\infty \exp(-nx)\,\mathrm{d}x \to 0$$

□

Example 7 (Another example of DCT) Evaluate the limit $n \to \infty$:

$$\int_0^\pi \sin\left(x + \frac{\cos(x)^n}{n}\right)\mathrm{d}x$$

Solution. This is a clear use of the remark made in Example 5. Since $f_n(x) = \sin\left(x + \frac{\cos(x)^n}{n}\right)$ is uniformly bounded by your favourite number greater than or equal to one, the DCT applies and as such

$$\int_0^\pi \sin\left(x + \frac{\cos(x)^n}{n}\right)\mathrm{d}x \to \int_0^\pi \sin(x)\,\mathrm{d}x = 2$$

□

Question 49 (2022 B3) Determine whether the limit

$$\lim_{a\to 0} \int_a^1 \frac{\sin(x)}{x^{3/2}}\,\mathrm{d}x$$

exists, and determine whether $\sin(x)/x^{3/2}$ is integrable on $[0,1]$

Solution. We translate this integral into the language of the DCT using the simple observation that

$$\lim_{a\to 0} \int_a^1 f(x)\,\mathrm{d}x = \lim_{N\to\infty} \int_0^1 f(x)\mathbf{1}_{[1/N,1]}\,\mathrm{d}x$$

We thus are looking at a sequence of functions

$$f_n(x) = \frac{\sin(x)}{x^{3/2}}\mathbf{1}_{[1/n,1]}(x)$$

It is easy to see that $|f_n(x)| \leq \frac{1}{\sqrt{x}}$ for all $x \in [0,1]$ and this latter function is integrable on $[0,1]$. It follows that the limit of $f_n(x)$, namely $\sin(x)/x^{3/2}$ is integrable, and of course the limit of the integral also exists. $\square$

Question 50 (2021 B3) Compute the $n \to \infty$ limits of the following Lebesgue integrals:

1. $$\int_0^\infty \frac{\sin(\exp(nx))}{n + x^{3/2}} \mathrm{d}x$$
2. $$\int_0^2 \frac{\sin(n^2 x)}{x} \mathrm{d}x$$
3. $$\int_0^1 \mathbf{1}_{C_n} \mathrm{d}x$$
   where C_n is a decreasing sequence of subsets $C_n \subseteq [0,1]$ converging to a set $C = \bigcap_n C_n$ of measure zero.

Proof.
1. Use DCT. We see that the integrand is a sequence f_n of measurable functions and moreover
   $$|f_n| \leq \frac{1}{1 + x^{3/2}}$$
   which is an integrable function on $[0,\infty)$. Therefore the integral equals zero.
2. Make a substitution and see that the integral is equal to
   $$\int_0^{2n^2} \frac{\sin(x)}{x} \mathrm{d}x$$
   we know this integral is not well defined in the Lebesgue sense.
3. The integral is precisely equal to $\mu(C_n)$. By continuity of the measure we see that $\lim_{n\to\infty} \mu(C_n) = \mu\left(\bigcap C_n\right) = \mu(C) = 0$.

$\square$

Question 51 (Mock 2022) Compute the following integral, where f is some measurable and continuous function $[0,\pi] \to \mathbb{R}$
$$I = \lim_{n\to\infty} \int_0^\pi f(\sin(x)^n) \mathrm{d}x$$

Solution. Since f is continuous on $[0,\pi]$, which is a compact interval, it is then bounded so we may apply the DCT. Since f is continuous we pass the limit inside the function and easily see that
$$I = \pi f(0)$$

$\square$

Question 52 (Homework sheet 7) Compute (if it exists) the limit of the following integrals.

1. $$\int_{\mathbb{R}} \frac{\sin(\exp(x))}{1+nx^2}\,\mathrm{d}x$$
2. $$\int_0^1 \frac{n\sin(x)}{1+n^2\sqrt{x}}\,\mathrm{d}x$$
3. $$\int_0^1 \frac{\sin(nx)}{x}\,\mathrm{d}x$$

Solution.

1. DCT with $\frac{1}{1+x^2}$, which is integrable. Limit is zero.
2. Obviously uniformly bounded by some large enough number. DCT applies. Limit is zero.
3. Limit doesn't exist, indeed, making a substitution $nx = y$, the integral becomes
$$\int_0^n \frac{\sin(y)}{y}\,\mathrm{d}y$$
which does not converge in the sense of Lebesgue. (But it does in the sense of Riemann)

$\square$

Question 53 (Q2 (d) 2019) Does there exist a function $g(x)$ with $\int_0^\infty |g(x)|\,\mathrm{d}x < \infty$ such that

$$\frac{e^{-at}-e^{-a^2t}}{t} \leq |g(t)|$$

For all $t > 0, a > 1$?

Solution. No. Suppose there was such a function. Then by the Dominated Convergence Theorem, letting $f_a(t) = \frac{e^{-at}-e^{-a^2t}}{t}$, one would have $\int_0^\infty f_a\,\mathrm{d}x \to \int_0^\infty f$ where f is the $a.e$ limit of f_a. Clearly f_a converges pointwise to 0 for $t > 0$ so in particular, it converges almost everywhere to the zero function. However, in the previous part we have computed the integral $\int_0^\infty f_a\,\mathrm{d}x$, which is equal to $\log(a)$. Since $a > 1$ it follows $\log(a) \neq 0$, contradiction, since according to the DCT we would expect the integral $\int_0^\infty f_a$ to converge to 0. $\square$

Question 54 (2019 Q2 (e)) Compute the following limits:

$$\lim_{n\to\infty}\int_0^1 \frac{1-e^{-nx}\sin(nx)}{1+x^2}\,\mathrm{d}x, \qquad \lim_{n\to\infty}\int_0^1 n\log\Big(1+\frac{x}{n}\Big)\,\mathrm{d}x$$

Solution. Let us compute the first limit. Notice that we are performing an integration over a space of finite measure (the interval $[0,1]$) and the integrand is uniformly bounded. Indeed: let $f_n = \frac{1-e^{-nx}\sin(nx)}{1+x^2}$. Then for any n, we have for example $f_n < 2$. So by the DCT, we have that:

$$\int_0^1 f_n\,\mathrm{d}x \to \int_0^1 f\,\mathrm{d}x$$

Where f is the $a.e$ limit of f. Observe that $f_n \to \frac{1}{1+x^2}$ pointwise. Now we conclude that:

$$\int_0^1 f_n\,\mathrm{d}x \to \int_0^1 \frac{1}{1+x^2}\,\mathrm{d}x = \arctan(1) - \arctan(0) = \frac{\pi}{4}.$$

A similar argument can be used for the second problem. First note that

$$f_n = n\log\Big(1+\frac{x}{n}\Big) = \log\Big[\Big(1+\frac{x}{n}\Big)^n\Big]$$

In particular, in the limit $n\to\infty$ one has:

$$f_n \to \log(e^x) = x$$

Returning to the integral, we can apply once again the DCT because the interval has a finite measure and the function is uniformly bounded. Therefore:

$$\int_0^1 f_n\,\mathrm{d}x \to \int_0^1 x\,\mathrm{d}x = 1/2.$$

□

5.4 Product measure spaces, Fubini's Theorem

Recall the following definition:

Definition 5.4.1 (σ-algebra generated by a function) Let Ω be a set, and $(A, \mathscr{A})$ be a measurable space. Let $f:\Omega\to A$ be any function. Then the σ-algebra generated by f, denoted $\sigma(f)$, is the smallest σ-algebra, $\mathscr{F}$ that makes f an $\mathscr{F}-\mathscr{A}$ measurable function. That is to say, $\sigma(f)$ is the σ-algebra generated by $f^{-1}(S)$ for all $S\in\mathscr{A}$.

Definition 5.4.2 (Product σ-algebra) Let $(A,\mathscr{A})$ and $(B,\mathscr{B})$ be two measurable spaces. Consider the projection maps $\pi_A : A\times B \to A$ and $\pi_B : A\times B\to B$ defined in the obvious ways. We define the product sigma algebra, denoted by $\mathscr{A} \otimes \mathscr{B}$ (nothing to do with the tensor product) is the σ-algebra generated by the projection maps, i.e:

$$\mathscr{A} \otimes \mathscr{B} = \sigma(\pi_A, \pi_B)$$

Remark 10 Since given any measurable set $S \in \mathscr{A}$, we have that $\pi_A^{-1}(S) = S \times B$, and something identical holds for π_B, we have that $\mathscr{A} \otimes \mathscr{B}$ contains all sets of the form $S_1 \times S_2$, where $S_1 \in \mathscr{A}$ and $S_2 \in \mathscr{B}$. For the case of a countable product of σ-algebras, it can be shown that Cartesian products like that of $S_1 \times S_2$ actually generate the product of algebras, so we may use either characterisation, depending on what's more useful. For uncountable products however, we use the definition given above.

The goal is to construct a measure on $\mathscr{A} \otimes \mathscr{B}$, and we do so by explicit construction.

Lemma 5.4.1 Let $(A, \mathscr{A}, \mu_A)$ and $(B, \mathscr{B}, \mu_B)$ be measure spaces. Let $\mathscr{E} = \mathscr{A} \otimes \mathscr{B}$ and $f = f(x_1, x_2)$ be $\mathscr{E}$-measurable. Then the function

$$x_1 \mapsto \int_B f(x_1, x_2) \mu_B(dx_2)$$

is $\mathscr{A}$-measurable.

This Lemma ensures that the following result makes sense.

Theorem 5.4.1 Let $(A, \mathscr{A}, \mu_A)$ and $(B, \mathscr{B}, \mu_B)$ be measure spaces. Let $\mathscr{E} = \mathscr{A} \otimes \mathscr{B}$. Then there exists a unique measure μ which is sometimes denoted as $\mu_A \otimes \mu_B$ on $\mathscr{A} \otimes \mathscr{B}$ such that

$$\mu(S_1 \times S_2) = \mu_A(S_1) \mu_B(S_2)$$

Proof. The measure is constructed as

$$\mu(S) = \int_A \left(\int_B \mathbf{1}_S(x_1, x_2) \mu_B(dx_2) \right) \mu_A(dx_1) \quad S \in \mathscr{A} \otimes \mathscr{B}$$

□

Theorem 5.4.2 (Fubini's / Tonelli's Theorem) Let f be non-negative $\mathscr{A} \otimes \mathscr{B}$-measurable. Then

$$\int_{A \times B} f \, d(\mu_A \otimes \mu_B) = \int_A \left(\int_B f \, \mathrm{d}\mu_B \right) \mathrm{d}\mu_A = \int_B \left(\int_A f \, \mathrm{d}\mu_A \right) \mathrm{d}\mu_B$$

5.4.1 Practice Questions

Question 55 (Q2 (b) 2019) Let $0 < a < b$ and

$$f(x,y) = \begin{cases} e^{-xy} & \text{if } x \in (0,\infty), y \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

Show f is integrable.

Solution. By Fubini's Theorem:

$$\int_{\mathbb{R}^2} f(x,y)\mathrm{d}\mu = \int_a^b \int_0^\infty e^{-xy}\,\mathrm{d}x\,\mathrm{d}y = \log(y)\Big|_{y=a}^{y=b} < \infty$$

□

Question 56 (Q2 (c) 2019) By using Fubini's Theorem or otherwise evaluate:

$$\int_0^\infty \frac{e^{-at} - e^{-bt}}{t}\,\mathrm{d}t$$

Solution. Looking at the previous question, we see that:

$$\int_{\mathbb{R}^2} f\,\mathrm{d}\mu = \log(b) - \log(a)$$

We could have swapped the order of integration using Fubini's Theorem and get that:

$$\log(b/a) = \int_0^\infty \int_a^b e^{-xy}\,\mathrm{d}y\,\mathrm{d}x = \int_0^\infty \frac{e^{-at} - e^{-bt}}{t}\,\mathrm{d}t$$

□

5.5 Expectation

In this chapter, we shall concentrate on spaces of measurable functions whose moments have finite expectation, it turns out that the geometry of one of these spaces, $L^2(\Omega, \mathscr{F}, \mathbb{P})$ has deep insights into probabilistic results.

Definition 5.5.1 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The **expectation** of X is defined to be

$$\mathbb{E}[X] = \int_\Omega X\,\mathrm{d}\mathbb{P}$$

If we take $X \in L^2(\Omega, \mathscr{E}, \mathbb{P})$, we may define the **variance**,

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Note that the variance corresponds to the square of the *distance* between the vectors X and $\mathbb{E}[X]$ in $\mathscr{L}^2$, this is a great geometric insight into probability because in this measure-theoretic formulation of probability, the variance (or rather the standard deviation) corresponds to how far away X and $\mathbb{E}[X]$ are as vectors in L^2, this is truly fascinating if you ask me, because it links a concept from elementary probability with geometry, and it goes to show how powerful this theory of probability is.

Proposition 5.5.1 (Properties of expectations and variances) Let $X, Y \in L^2(\Omega, \mathscr{E}, \mathbb{P})$ Then the expectation is:

1. Linear:
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

2. Almost surely monotonic: if $X \leq Y$ a.s then:
$$\mathbb{E}[X] \leq \mathbb{E}[Y]$$

   The variance satisfies:

   (a) (Variance is square of norm) Quadratic scaling:
$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$$

   (b) (Expectation is linear) Translational invariance:
$$\mathrm{Var}(X + b) = \mathrm{Var}(X)$$

   (c) (Pythagoras' Theorem) If X and Y are independent:
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Notice in this last statement of the Proposition, I have labeled this result as Pythagoras' Theorem. This is because L^2 is actually even better than a Banach space, it is actually a Hilbert space, since it is equipped with an inner product
$$\langle f, g \rangle = \mu(fg)$$
Which enables the measurement of "angles". It turns out that the angle between two random variables is their correlation. Hence the interpretation of two random variables being orthogonal is that they are uncorrelated, so technically, the result could be relaxed to just say uncorrelated instead of independent. Obviously, independent random variables are uncorrelated.

Question 57 Is it true that if $X_n \to X$ in distribution and $\mathbb{E}[X_n] \to \mathbb{E}[X]$, then $\mathbb{E}[X_n^2] \to \mathbb{E}[X]$?

Solution. False. Indeed, define the random variable

$$X_n = \begin{cases} \sqrt{n} & \text{with } \mathbb{P} = 1/n \\ 0 & \text{with } \mathbb{P} = 1 - 1/n \end{cases}$$

We verify that given any $\epsilon > 0$,

$$\mathbb{P}(X_n > \epsilon) = 1/n \to 0$$

and as such $X_n \to 0$ in distribution. Moreover,

$$\mathbb{E}[X_n] = \frac{\sqrt{n}}{n} \to 0$$

but $\mathbb{E}[X_n^2] = 1$ □

Example 8 (Two examples of expectations) Here are two well known examples of expectations, phrased in this better language:

1. Dirac δ measure: given a measure space $(E, \mathscr{E})$ and $x_0 \in E$, we define the *point-mass* measure at x_0

$$\delta_{x_0}(A) = \begin{cases} 1 & x_0 \in A \\ 0 & x_0 \notin A \end{cases}, \quad A \in \mathscr{E}$$

   Now given a function $f : \mathbb{R} \to \mathbb{R}$, a random variable $X = c$ a.s, and the Dirac measure δ_c we see that

$$\mathbb{E}[f(X)] \overset{def}{=} \int_{\mathbb{R}} f(X) d\delta_c = \int_{\mathbb{R}} f(c) \mathbf{1}_{\{c\}} d\delta_c = f(c)\delta_c(\{c\}) = f(c)$$

2. Discrete random variable: given X taking values in $\{1, 2, \cdots\}$ then almost surely

$$X = \lim_{N \to \infty} \sum_{n=1}^{N} n \mathbf{1}_{\{X=n\}}$$

   since this is a monotone limit, we may apply MCT and see that

$$\mathbb{E}[X] = \lim_{N \to \infty} \mathbb{E}\left[\sum_{n=1}^{N} n \mathbf{1}_{\{X=n\}}\right] = \lim_{N \to \infty} \sum_{n=1}^{N} n \mathbb{E}[\mathbf{1}_{\{X=n\}}]$$

   this latter quantity is precisely

$$\sum_{n=1}^{\infty} n \mathbb{P}(X = n)$$

Let's take a second to appreciate this last example. Once again, we see how this incredibly technological formulation of probability, through the use of measure theory and Lebesgue integration, recovers familiar concepts such as the expectation of a discrete random variable, whilst opening the doors to a myriad of more complicated insights that elementary probability could have not given us.

Question 58 (Homework sheet 7) Let (X_n) be a collection of non-negative random variables. By using the MCT, show that

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n]$$

Then show that if X is an $\mathbb{N}$-valued random variable, then

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$$

Hence, show that if $\mathbb{E}[X] = \infty$, then with probability one, $X_n \geq n$ infinitely often.

Solution. We define the sequence of measurable functions

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

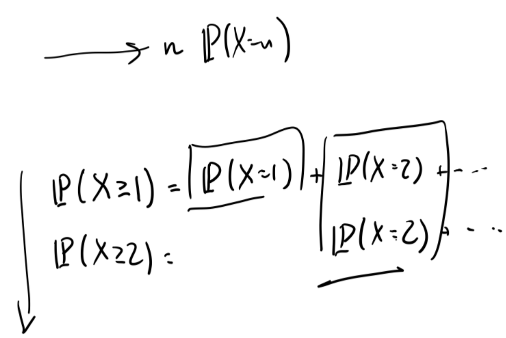
Given that X_n is non-negative, this sequence is monotone increasing, so by the Monotone Convergence Theorem:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \mathbb{E}\left[\lim_{N\to\infty} \sum_{n=1}^{N} X_n\right] = \lim_{N\to\infty} \mathbb{E}\left[\sum_{n=1}^{N} X_n\right] = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{E}[X_n] = \sum_{n=1}^{\infty} \mathbb{E}[X_n]$$

Now we can focus on the next part. Indeed, if X is $\mathbb{N}$-valued, we can write $X = \sum_{n=1}^{\infty} n\mathbf{1}_{\{n\}}$. Hence using the first part,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{n=1}^{\infty} n\mathbf{1}_n\right] = \sum_{n=1}^{\infty} n\mathbb{E}[\mathbf{1}_n] = \sum_{n=1}^{\infty} n\mathbb{P}(X = n)$$

I now claim that this last quantity is precisely equal to $\sum_{n=1}^{\infty} \mathbb{P}(X \geq n)$ its really just a trivial combinatoric argument, but here is a pictorial proof, sorry for the messy drawing.



Now to solve the last part, we simply use the Borel-Cantelli Lemma. $\square$

Definition 5.5.2 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable, and $\mathbb{P}_X$ be the measure induced by X on $(\mathbb{R},\mathfrak{B}(\mathbb{R}))$. If there exists a function f such that

$$\mathbb{P}_X(A)=\int_{\mathbb{R}} f_X \mathbf{1}_A \,\mathrm{d}x$$

then we refer to f_X as the **probability density** of X.

Remark 11 (Unicity of probability densities) Suppose f_X and $\tilde{f}_X$ are two functions satisfying the defining property of a density function, then define the set $A=\{x\in X \mid f_X(x)\neq \tilde{f}_X(x)\}$. By hypothesis:

$$\int_{\mathbb{R}}(f_X-\tilde{f}_X)\mathbf{1}_A \,\mathrm{d}x=0$$

Therefore by the non-negativity of $f_X,\tilde{f}_X$, it follows:

$$(f_X-\tilde{f}_X)\mathbf{1}_A=0 \quad \mu-\text{almost-everywhere}$$

but by assumption, on A, the two functions differ so it follows that $\mu(A)=0$.

Remark 12 The density depends not only on the random variable but also on the underlying probability space. Indeed, if $\mathrm{d}x$ is the Lebesgue measure on $\mathbb{R}$, suppose the zero random variable $X=0$ had a density f_X. Then

$$1=\mathbb{P}_X(\{0\})=\int_{\mathbb{R}} f_X\,\mathrm{d}x=\int_{\{0\}} f_X\,\mathrm{d}x=f_X(0)\,\mathrm{d}x(\{0\})=0$$

However, if we take μ to be the counting measure, then it does! and $f_X(x)=\begin{cases}1 & x=0\\ 0 & \text{otherwise}\end{cases}$

Question 59 (2022 Mock) Let X and Y be two random variables with densities f_X and f_Y, is it true that the random variable $X+Y$ has density f_X+f_Y?

Solution. False. Take the probability space to be $(\mathbb{R},\mathfrak{B}(\mathbb{R}),m)$ where m is the Lebesgue measure and take $Y=-X$ where X is any random variable with density f_X. Then $X+Y$ is the zero random variable but as seen above, in this probability space, the zero random variable has no density. □

Question 60 (2022 Mock) Let $X:\Omega\to\mathbb{R}$ be a random variable with density f_X. Is it true that $Y=X+c$ has density $f_X(x-c)$?

Proof. Yes! Indeed:

$$\mathbb{P}(Y \in (-\infty, a]) = \mathbb{P}(X \in (-\infty, a-c]) = \int_{-\infty}^{a-c} f_X(x)\mathrm{d}x = \int_{-\infty}^{a} f_X(x-c)\mathrm{d}x$$

as required. □

Question 61 (2022 Mock) If X has density f_X, what is the density of the random variable $Y = \exp(X)$?

Proof. We can perform this via direct method or using some technology which gives a shortcut, now we restrict ourselves to the direct method. Suppose firstly that $x > 0$:

$$\mathbb{P}(Y \in (-\infty, x]) = \mathbb{P}(\exp(X) \in (-\infty, x]) = \mathbb{P}(X \in (-\infty, \log(x)] = \int_{-\infty}^{\log(x)} f_X(x)\mathrm{d}x$$

now we make a substitution $x \to \log(x)$ and we reach the conclusion that this last integral is precisely

$$\int_{-\infty}^{x} \frac{f_X(\log(x))}{x}\mathrm{d}x$$

Clearly if $x \leq 0$ then $\mathbb{P}(Y \in (-\infty, x]) = 0$ so we reach the conclusion that

$$\mathbb{P}(Y \in (-\infty, x]) = \int_{-\infty}^{x} \mathbf{1}_{>0} \frac{f_X(\log(x))}{x}\mathrm{d}x$$

for all x. □

Question 62 (2022 Mock) Suppose X is a random variable such that $\mathbb{P}(X \in \mathbb{Q}) > 0$. Does X have a density with respect to the Lebesgue measure?

Proof. Hell no. Indeed:

$$0 < \mathbb{P}(x \in \mathbb{Q}) = \int_{\mathbb{Q}} f_X(x)\mathrm{d}\mu = 0 \qquad \text{since } \mu(\mathbb{Q}) = 0$$

where μ is the Lebesgue measure. □

Question 63 (Homework sheet 8) Let $R \sim \text{Exp}(\lambda)$ and let A be the random variable denoting the area of a circle of radius R. Find the density of A.

Solution. We begin by finding the distribution of A:

$$\mathbb{P}(A \leq a) = \mathbb{P}(\pi R^2 \leq a) = \mathbb{P}\left(R \leq \sqrt{\frac{a}{\pi}}\right)$$

in this last equality we have used the fact that R, being exponentially distributed, does not take negative values. This last quantity is obviously equal to

$$1-\exp\left(-\lambda\sqrt{\frac{a}{\pi}}\right)$$

and by differentiating this quantity we reach the conclusion of $f_A(a)$. □

Definition 5.5.3 (Change of measure) Let $(\Omega,\mathscr{F},\mu)$ be a measure space and $f:\Omega\to\mathbb{R}$ be a non-negative measurable function. The **change of measure** given by f is the measure $\nu:\mathscr{F}\to\mathbb{R}$ given by

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

Proposition 5.5.2 The measure induced by a non-negative measurable function is indeed a measure. Moreover, if g is an integrable function, then

$$\int_\Omega g\,\mathrm{d}\nu=\int_\Omega fg\,\mathrm{d}\mu$$

Proposition 5.5.3 Let Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and $X:\Omega\to\mathbb{R}$ be a random variable. Then

$$\mathbb{E}[X]\overset{\text{def}}{=}\int_\Omega X\,\mathrm{d}\mathbb{P}=\int_\mathbb{R} x\,\mathrm{d}\mathbb{P}_X=\int_\mathbb{R} x f_X\,\mathrm{d}x$$

where the second equality holds if f_X exists.

Question 64 (2022, pending verification) True/False: If there exists some $c\in\mathbb{R}$ so that $\mathbb{P}_X(\{c\})>0$, then X does not have a density function.

Proof. Suppose it has a density function $f_X(x)$. Then

$$0<\mathbb{P}_X(\{c\})=\int_{\{c\}}\mathrm{d}\mathbb{P}_X=\int_{\{c\}} f_X(x)\,\mathrm{d}x=0$$

contradiction.

□

Question 65 (2022, pending verification) If X has a density function, then so does $|X|$.

Solution. Yes. Let $f_X(x)$ be the density function of X. We must find a function $f_{|X|}(x)$ such that

$$\mathbb{P}(|X| \in (-\infty, x)) = \int_{-\infty}^{x} f_{|X|}(y)\,\mathrm{d}y$$

Observe that $\mathbb{P}(|X| \in (-\infty, x)) = \mathbb{P}(X \in (-x, x))) = \int_{\mathbb{R}} f_X(y)\mathbf{1}_{[-x,x]}\,\mathrm{d}y = \int_{-\infty}^{x} f_X(y)\mathbf{1}_{[-x,x]}\,\mathrm{d}y$. Thus we have our density function. $\square$

Remark 13 If $F_X(x)$ is the distribution function of a random variable and $F_X(x)$ is continuously differentiable for all but finitely many points, then $F_X'(x)$ is the distribution of X. This follows by the FTC.

Example 9 (Exponential distribution) A random variable X is said to be of exponential distribution with parameter λ if

$$\mathbb{P}(X \geq x) = \exp(-\lambda x)$$

From the above remark it follows that $\lambda \exp(-\lambda x)$ is the pdf of X.

5.6 Inequalities concerning expectations

Theorem 5.6.1 (Markov's Inequality) Let X be a non-negative random variable. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Note that if X has infinite expectation this still holds vacuously.

Key of proof. Consider the expectation of $Y = a\mathbf{1}_{\{X \geq a\}}$ □

Theorem 5.6.2 (Chebyshev's Inequality) Let X be a random variable with finite mean and variance. Then for all $k > 0$.

$$\mathbb{P}(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Key of proof. Apply Markov with $Y = (X - \mu)^2$ □

Theorem 5.6.3 (Chernoff bounds) Let X be a random variable with finite mean and variance and $t \geq 0$ be such that $\mathbb{E}[\exp(tX)] < \infty$. Then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[\exp(tX)]}{\exp(ta)}$$

Theorem 5.6.4 (Jensen's inequality) Let X be an integrable random variable and $g : \mathbb{R} \to \mathbb{R}$ a convex function, then:

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

Theorem 5.6.5 (Cauchy-Schwarz) Let X, Y be two random variables, then

$$(\mathbb{E}[|XY|])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

5.6.1 Some questions

Question 66 (Homework sheet 8) Show that for a random variable X and $p \in (0, \infty)$, we have that

$$\mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}[|X|^p]}{x^p}$$

Solution. This follows from a simple application of Markov's inequality:

$$\mathbb{P}(|X| \geq x) = \mathbb{P}(|X|^p \geq x^p)$$

□

Question 67 (Homework sheet 8) Let $X \geq 0$ almost surely. Show that for some range of z,

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[z^X]}{z^x}$$

Solution. We apply once again Markov's inequality but now we will apply it to the variable z^X. The thing to be careful of, is that in order to preserve the order of the inequalities, we must have $z \geq 1$. □

Question 68 (Homework sheet 8) Let $X_n \sim \text{Bin}(n, 1/3)$. By using Markov's Inequality, Chebyshev's Inequality and the inequality obtained in Question 67, obtain upper bounds on $\mathbb{P}\big(X_n \geq \frac{2n}{3}\big)$.

Solution.

1. Markov:
$$\mathbb{P}\left(X_n \geq \frac{2n}{3}\right) \leq \frac{n/3}{2n/3} = 1/2$$
So this is not a great inequality because its always the same. But that's to be expected because Markov is a very "dumb" inequality.

2. Chebyshev: We rather implement "the proof" of Chebyshev's inequality to obtain this bound:
$$\mathbb{P}(X_n \geq 2n/3) = \mathbb{P}(X_n - \mu \geq n/3) = \mathbb{P}((X_n - \mu)^2 \geq n^2/9) \leq \frac{\text{Var}[X_n]}{n^2/9} = \frac{2}{n}$$
Much better.

3. PGF: We must first compute
$$\mathbb{E}[z^{X_n}] = \sum_{k=0}^{n} z^k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k} = (zp + 1 - p)^n = \left(\frac{2}{3} + \frac{z}{3}\right)^n$$
Hence
$$\mathbb{P}(X \leq x) \leq \left(\frac{2+z}{3z^{2/3}}\right)^n$$
Now one could sweat a bit to find a nice value of z that minimises the thing inside the bracket. In fact when $z = 4$, the thing inside the bracket is $2^{-1/3}$ so we see that this inequality is super good. As its usual in Mathematics, the more you have to think the better the result it.

□

Question 69 (Homework sheet 7) We say $X_n \to X$ in L_1 if $\mathbb{E}[|X_n - X|] \to 0$. Show that convergence in L_1 implies convergence in probability.

Solution. We use Markov's inequality. For a fixed $\epsilon > 0$:

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\epsilon} \to 0$$

□

5.7 Change of variables

Sometimes we are given some random variable X with its density, and we wish to study the density of $g(X)$ for some measurable function g. This can be done in two different ways. The direct way and the sledgehammer approach.

Example 10 (Direct way) Consider $X \sim \text{Exp}(\lambda)$ and set $Y = cX$. Compute the distribution function of Y using the direct way:

$$\mathbb{P}(Y \leq x) = \mathbb{P}(cX \leq x) = \mathbb{P}(X \leq x/c) = 1 - \exp(-\lambda x/c)$$

Sometimes this is too hard to do, so we resort to some extra technology:

Proposition 5.7.1 (Beefed up change of variables) Let X be a continuous random variable with density $f_X(x)$ and $g: \mathbb{R} \to \mathbb{R}$ a measurable function **such that**

1. g is strictly monotonic, thus ensuring the existence of g^{-1} as per the inverse function theorem.
2. g^{-1} is differentiable everywhere.

Then the density of Y is given by

$$f_Y(y) = f_X(g^{-1}(x))\left|\frac{d}{dy} g^{-1}(y)\right|$$

Key of proof. Consider the distribution function:

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Now use the chain rule. □

Example 11 (Example 10 revisited) We now perform the same task using our proposition. Here $g(x) = cx$, which satisfies the conditions. Clearly $g^{-1}(x) = \frac{1}{c}x$ and you can easily verify that the same result as obtained in Example 10 holds.

6 Multivariate Probability

Definition 6.0.1 Let (X,Y) be an $\mathbb{R}^2$-valued random variable (which we refer to as a random vector). The **joint distribution** function is defined as

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y)$$

The **joint density** is (if it exists) the function $f_{X,Y} : \mathbb{R}^2 \to [0,\infty)$ that satisfies

$$F_{X,Y}(x,y) = \iint_{-\infty,-\infty}^{x,y} f_{X,Y}(u,v) du dv$$

Remark 14 By the Fundamental Theorem of Calculus, it follows

$$f_{X,Y} = \partial_x \partial_y F_{X,Y}$$

provided that $F_{X,Y}$ is differentiable with respect to x and y.

Definition 6.0.2 (Marginal density) Let (X,Y) be a random vector with density $f_{X,Y}$. The marginal density $f_X(x)$ is defined as:

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y) \mathrm{d}y$$

Remark 15 This allows us to recover the expectations of each individual random variable, indeed, since given a measurable $g : \mathbb{R}^2 \to \mathbb{R}$ we have that

$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}^2} g(X,Y) f_{X,Y} \, \mathrm{d}x \, \mathrm{d}y$$

we can choose the projection map $\pi_X(x,y) = x$, which is clearly measurable, and as such

$$\mathbb{E}[X] = \mathbb{E}[\pi_X(X,Y)] = \int_{\mathbb{R}} X \left(\int_{\mathbb{R}} f_{X,Y} \, \mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}} X f_X(x) \mathrm{d}x$$

Example 12 (Sheet 9) Let (X,Y) be a random vector uniformly distributed in the triangle

$$T = \{y \geq x \mid 0 \leq x, y \leq 1\}$$

Compute $\mathbb{E}[X]$.

Proof. We begin by computing the density of (X,Y). Given that T has area of $1/2$, then $f_{X,Y}(x,y) = 2\mathbf{1}_T$. To compute $\mathbb{E}[X]$, we could do two things, compute the integral

$$\int\int_{\mathbb{R}^2} x f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

Or alternatively, compute the marginal density

$$f_X(x) = \int_{-\infty}^{\infty} 2\mathbf{1}_T(x,y)\,\mathrm{d}y = 2(1-x)\mathbf{1}_{[0,1]}$$

Hence $\mathbb{E}[X] = \int_{\mathbb{R}} 2x(1-x)\mathbf{1}_{[0,1]}\,\mathrm{d}x$ □

6.1 Covariance and dependence

Definition 6.1.1 (Covariance) Let (X, Y) be a random vector. Define $g(x, y) = (x - \mathbb{E}[X])(y - \mathbb{E}[Y])$. Then $\mathbb{E}[g(X, Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ is the covariance of (X, Y), denoted $\mathrm{Cov}(X, Y)$

Remark 16 (Geometric interpretation in L^2) Suppose that $X, Y \in L^2(\Omega, \mathscr{E}, \mathbb{P})$ are both zero-mean random variables. Then $\mathrm{Cov}(X, Y) = \langle X, Y \rangle$. This has a deep geometric interpretation as hinted in the chapter on Expectations. Since inner product measures angles, we see that the interpretation of two random variables being orthogonal in L^2 corresponds to them having zero covariance. Something even cooler is actually true, the cosine of the angle between the two random variables corresponds to their correlation.

Definition 6.1.2 (Correlation) Let X and Y be two random variables, then

$$\mathrm{Corr}(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Remark 17 Correlation is scale invariant and by the Cauchy-Schwarz inequality, $\mathrm{Corr}(X, Y) \in [-1, 1]$, which checks out with the fact that it corresponds to the cosine of an angle.

Remark 18 (Independence of random variables and covariance) By definition, X and Y are independent if and only if $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ so by taking suitable derivatives, we see that X and Y are independent if and only if $f_{X,Y} = f_X(x)f_Y(y)$, that is to say, $f_{X,Y}$ splits as a product of two function, one that depends strictly on X, one that depends strictly on Y. A consequence to this observation is that if X and Y are independent, $\mathbb{E}[XY]$ will split as an integral into $\mathbb{E}[X]\mathbb{E}[Y]$ which implies that $\mathrm{Cov}(X, Y) = 0$. The converse is however not true.

Example 13 (Zero covariance yet not independent) Consider the random variable

$$X = \begin{cases} 1 & \mathbb{P} = 1/2 \\ -1 & \mathbb{P} = 1/2 \end{cases}$$

Now construct the random variable Y, as follows

$$Y = \begin{cases} 0 & \text{if } X = 1 \\ \begin{cases} 1 & \mathbb{P} = 1/2 \\ -1 & \mathbb{P} = -1/2 \end{cases} & \text{if } X = -1 \end{cases}$$

Obviously these random variables are not independent. Yet it is easy to see that $\mathbb{E}[XY] = \mathbb{E}[X] = \mathbb{E}[Y] = 0$.

Proposition 6.1.1 Given a collection of random variables $X_1, \cdots, X_n$,

$$\mathrm{Var}\left(\sum_i X_i\right) = \sum_{i,j} \mathrm{Cov}(X_i, X_j)$$

Proposition 6.1.2 (Covariance is bilinear) The Covariance is bilinear.

7 Transformations of Multivariate Random Variables

7.1 Convolutions, simple case of transformations.

Proposition 7.1.1 Suppose X and Y are independent Random Variables with densities f_X and f_Y. Then the density of $Z = X + Y$, f_Z is given by

$$f_Z(z) = (f_X * f_Y)(z)$$

Proof. We start by considering the distribution function

$$F_Z(z) = \mathbb{P}(X + Y \le z) = \int_{\{x+y \le z\}} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y =$$

$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_X(x) f_Y(y)\,\mathrm{d}y\,\mathrm{d}x$$

substitute $y = \omega - x$ and get

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_X(x) f_Y(\omega - x) d\omega\,\mathrm{d}x =$$

$$= \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_X(x) f_Y(\omega - x)\,\mathrm{d}x\,d\omega = \int_{-\infty}^{z} g(\omega) d\omega$$

where $g(\omega) = (f_X * f_Y)(\omega)$. Now differentiating finishes the claim.

□

Example 14 Let X and Y both be distributed according to $Exp(\lambda)$ and consider $Z = X + Y$, then noting that $f_X(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\{x \ge 0\}}$, it also follows that $f_Y(z - y) = \lambda \exp(-\lambda(z - y)) \mathbf{1}_{\{y \le z\}}$

$$f_Z(z) = \lambda^2 \int_0^z \exp(-\lambda x) \exp(-\lambda(z - x))\,\mathrm{d}x = \lambda^2 z \exp(-\lambda z)$$

Remark 19 This is a particular case of a more general phenomenon, that of the family of distributions called the Gamma distribution, denoted by $\Gamma(n,\lambda)$, where $n = 1, 2, \cdots$. Its density is

$$x^{n-1}\exp(-\lambda x)\frac{\lambda^n}{(n-1)!}\mathbf{1}_{\{x\geq 0\}}$$

and generally speaking if $X_1, \cdots, X_n$ are I.I.D exponentials with parameter λ, it follows that

$$\sum_{i=1}^{n} X_i \sim \Gamma(n,\lambda)$$

7.2 General case of transformations

Proposition 7.2.1 Let (X, Y) be a random variable with joint density $f_{XY}(x, y)\mathbf{1}_D$ and

$$\phi : D \subseteq \mathbb{R}^2 \to C \subseteq \mathbb{R}^2$$

be an injective map. We define the random variable

$$(U, V) = \phi(X, Y) = (u(X, Y), v(X, Y))$$

If ϕ^{-1} has continuous partial derivatives almost everywhere on the codomain, then the density f_{UV} of (U, V) is given by

$$f_{UV}(u, v) = f_{XY}(\phi^{-1}(u, v))|J(\phi^{-1})|\mathbf{1}_C$$

Where $J(\phi^{-1})$ is the Jacobean of ϕ^{-1}. Generally, if $\varphi(x, y) = (u(x, y), v(x, y))$, then

$$J(\varphi) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

7.2.1 Example computations

Example 15 Let X and Y be two I.I.D $Exp(\lambda)$ random variables. Define the random variables

$$Z = X + Y \qquad Q = \frac{X}{X+Y}$$

and compute $f_{Z,Q}$.

Solution. We begin by noting that our map is given by

$$\phi : (X, Y) \mapsto \left(X + Y, \frac{X}{X+Y}\right)$$

this map is injective, and we may write the inverse:

$$\phi^{-1} : (Z, Q) \mapsto (ZQ, Z(1-Q))$$

which has continuous partial derivatives, so we may apply the proposition above. First we construct the Jacobean $J(\phi^{-1})$:

$$J(\phi^{-1}) = \begin{pmatrix} Q & Z \\ 1-Q & -Z \end{pmatrix}$$

and its determinant is nothing but

$$|J(\phi^{-1})| = |-QZ + QZ - Z| = Z$$

Since $f_{XY}(x, y) = f_X(x) f_Y(y)$ given that they are I.I.D, we see that

$$f_{XY}(\phi^{-1}(u, v)) = f_X(uv) f_X(u - uv) = \lambda^2 \exp(-\lambda u)$$

and as such $f_{ZQ}(z, q) = \lambda^2 z \exp(-\lambda z) \mathbf{1}_{\{z \geq 0\}}$ notice that incidentally, Q is uniformly distributed as that accounts for a 1 in the product above, and Z has the shape of $\Gamma(2, \lambda)$ as computed before via the convolution. □

Example 16 Let X and Y be two I.I.D standard Normal distributions. By writing $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$ where now (R, Θ) is another random vector, study the joint distribution of (R, Θ).

Solution. We proceed as standard. First we notice that $f_{XY} = \frac{1}{2\pi} \exp\left(-\frac{(x^2+y^2)}{2}\right)$. We have some $\phi : (X, Y) \mapsto (R, \Theta)$. The inverse of this map is very easy to write. Indeed:

$$\phi^{-1} : (R, \Theta) \mapsto (R\cos(\Theta), R\sin(\Theta))$$

We construct the Jacobean

$$J(\phi^{-1}) = \begin{pmatrix} \cos(\Theta) & -R\sin(\Theta) \\ \sin(\Theta) & R\cos(\Theta) \end{pmatrix}$$

The determinant is very clearly R and so applying the Proposition, we see that

$$f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} r^2\right) r$$

What we conclude for this, is that $f_\Theta(\theta) = 1/2\pi$ and $f_R(r) = \exp\left(-\frac{1}{2} r^2\right) r$. Meaning that the random vector (X, Y) is radially symmetric □

A good observation is that

$$|J(\phi)| = \frac{1}{|J(\phi^{-1})|}$$

this might ease some computations.

Question 70 (Homework sheet 9) Let $T \subseteq \mathbb{R}^2$ be the triangle with vertices $(0,0),(0,1),(1,1)$. Let $(X,Y) \sim$ Unif(T). Let $S = X+Y$ and $U = Y-X$. Find $f_{SU}(s,u)$. Are S and U independent? Find $\mathbb{E}[S]$ and $\mathbb{E}[U]$.

Solution. Clearly $f_{XY} \equiv 2$. The map ϕ is as follows $\phi(X,Y) = (X+Y, Y-X)$ so $J(\phi) = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ Therefore $|J^{-1}(\phi)| = \frac{1}{|J(\phi)|} = 1/2$. Note that $\phi(T) = T'$ where T' is the triangle with vertices $\phi(0,0) = (0,0)$, $\phi(0,1) = (1,1)$ and $\phi(1,1) = (2,0)$. This gives $f_{S,U} = \mathbf{1}_{T'}$, which is consistent with the fact that T' has area one, therefore the integral of the density over the whole T' is one as should be expected.

The density does not split as a product, so they are not independent.

To find the expectation, we note that $f_{S,U}$ is symmetric about $S = 1$ so $\mathbb{E}[S] = 1$. To compute $\mathbb{E}[U]$ we perform the integral

$$\mathbb{E}[U] = \iint_{T'} u du ds = 2\int_0^s \int_0^1 u ds du = \frac{1}{3}$$

□

Question 71 (Homework sheet 9) Let X, Y be independent random variables. Define $S = X+Y$. Compute the joint density of (X,S) and by finding the marginal distribution of S, verify the convolution result.

Solution. We have that the map ϕ given by

$$\phi : (X,Y) \mapsto (X, X+Y)$$

has Jacobean

$$J(\phi) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Hence by the observation in the box above, we have that $|J(\phi^{-1})| = 1$. As such

$$f_{X,S}(x,s) = f_{X,Y}(x, s-x) = f_X(x) f_Y(s-x)$$

now we find the marginal.

$$f_S(s) = \int_{\mathbb{R}} f_{X,S}(x,s) \mathrm{d}x = \int_{\mathbb{R}} f_X(x) f_Y(s-x) \mathrm{d}x = (f_X * f_Y)(s)$$

as required. □

Question 72 (Homework sheet 9) Let X and Y be two random variables with joint density f_{XY}. Find the joint density of $Z = Y/X$ and the marginal of Z.

Solution. The map is given by

$$\phi : (X, Y) \mapsto (X, X/Y)$$

so the Jacobean is $1/x$, so $|J(\phi^{-1})| = |x|$. Therefore

$$f_{X,Z}(x,z) = |x| f_{X,Y}(x, xz)$$

You can find the marginal. □

Question 73 (Homework sheet 9 + 2020) Let $X \sim Exp(\lambda)$ and $Y \sim Exp(\mu)$ be independent. Determine the distribution of $\min(X, Y)$. Define $Y_n = \min_{1 \le i \le n} X_i$. What is the distribution of Y_n? Determine if Y_n converges in probability or almost surely, and if it does, determine which random variable it converges to. Also determine necessary and sufficient conditions on $\alpha \in \mathbb{R}$ so that Y_n / n^α converges to zero in probability and determine the distribution of nY_n.

Solution. Determining the distribution of $\min(X, Y)$ is a pretty standard computation:

$$\mathbb{P}(\min(X, Y) \ge x) = \mathbb{P}(X \ge x, Y \ge x) = \mathbb{P}(X \ge x)\mathbb{P}(Y \ge x)$$

where in the last equality we have used independence. Now, inspecting the shape of the exponential distribution, we see that this last product is also the distribution of the exponential distribution with parameters $\mu + \lambda$. Therefore $\min(X, Y) \sim Exp(\mu + \lambda)$. Inductively, now one argues that $Y_n \sim Exp(\lambda n)$. And by inspecting the distribution function, it seems reasonable to guess that as $n \to \infty$, this should converge in some sense to the zero random variable. Let us inspect firs:

1. Convergence in probability: Let $\epsilon > 0$ be given, then $\mathbb{P}(Y_n > \epsilon) = \exp(\lambda n \epsilon) \to 0$ as $n \to \infty$. Thus we have convergence in probability.
2. Almost sure convergence. We employ the Borel-Cantelli lemma. Let $\epsilon > 0$ be given, then

$$\sum_{n=1}^{\infty} \mathbb{P}(Y_n > \epsilon) = \sum_{n=1}^{\infty} \exp(-\lambda n \epsilon) < \infty$$

   and as such we have almost sure convergence.
3. If we want Y_n / n^α to converge to zero in probability, we want that for any given $\epsilon > 0$, the probability

$$\mathbb{P}(Y_n > \epsilon n^\alpha)$$

   goes to zero. This quantity is precisely $\exp(-\lambda n^{\alpha+1} \epsilon)$, which provided $\alpha > -1$, does indeed converge to zero.
4. To conclude, we note that if $Y_n \sim Exp(\lambda n)$, then $nY_n \sim Exp(\lambda)$. No need for further calculations.

□

8 Conditional probability

Definition 8.0.1 (Conditional distribution of X given A) Let X be a random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We define the random variable X subject to an event $A \in \mathscr{F}$ with $\mathbb{P}(A) > 0$ by having the distribution function

$$\mathbb{P}(X \in C \mid A) = \frac{\mathbb{P}(\{X \in C\} \cap A)}{\mathbb{P}(A)}$$

Remark 20 Provided that $\mathbb{P}(A) > 0$, the definition above satisfies the definitions of a distribution and we can define

$$F_{X|A}(x) = \mathbb{P}(X \leq x \mid A)$$

Example 17 (Memory-less property of the exponential distribution) Let $X \sim Exp(\lambda)$ and $A = \{X \geq a\}$. Then

$$(X \mid A) \sim a + Exp(\lambda)$$

Indeed:

$$\mathbb{P}(X \geq t + a \mid X \geq a) = \frac{\mathbb{P}(X \geq t + a)}{\mathbb{P}(X \geq a)} = \exp(-\lambda(t + a))/\exp(-\lambda a) = \exp(-\lambda t)$$

We now take a look at the concept of conditional expectation, which will be central in our later discussion of Martingale Theory. Conditional expectation should, as we will see later, be seen as our best approximation to a given random variable, provided that we restrict our information available about said random variable.

Definition 8.0.2 (Expectation conditioned to an event) By observing Definition 8.0.1 it is natural to define for an event $A \in \mathscr{F}$ of non-zero probability, the expectation of X given A to be

$$\mathbb{E}[X \mid A] = \frac{1}{\mathbb{P}(A)} \mathbb{E}[X \mathbf{1}_A]$$

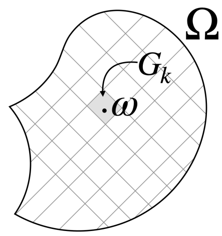
Definition 8.0.3 (Conditional expectation) Let (G_n) be a collection of pairwise disjoint events with the property that

$$\Omega = \bigcup_n G_n$$

Define $\mathscr{G} = \sigma(\{G_n : n \geq 0\})$. The random variable $\mathbb{E}[X \mid \mathscr{G}]$ is defined to be

$$\mathbb{E}[X \mid \mathscr{G}] = \sum_{n=1}^{\infty} \mathbb{E}[X \mid G_n] \mathbf{1}_{G_n}$$

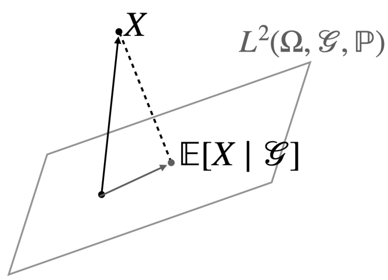
The way to think about this is as follows: our collection G_n partitions our sample space Ω, and the interpretation of the conditional expectation of X given $\mathscr{G}$ is that the only information about the chosen sample point $\omega \in \Omega$ is which "box" it lands in, and as such, we can say that our best guess for the value of X given that we only know in which "box" we are, is precisely the average of X in that box.



$\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X \mid G_k]$

We would now like to understand this random variable a bit more profoundly. First we make a geometric statement, which should be fairly natural to the reader who has understood the definition of conditional expectation.

Proposition 8.0.1 Let $L^2(\Omega, \mathscr{G}, \mathbb{P})$ be the subspace of $L^2(\Omega, \mathscr{F}, \mathbb{P})$ of $\mathscr{G}$-measurable functions. Then if $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$, then $\mathbb{E}[X \mid \mathscr{G}]$ is the orthogonal projection of X onto $L^2(\Omega, \mathscr{G}, \mathbb{P})$.



This formalises the intuition that the conditional expectation is the best guess you can do given the information encoded in $\mathscr{G}$.

Proof of Proposition 8.0.1. It suffices to show that $\mathrm{dist}(X, L^2(\mathscr{G}))$ is minimised by $\mathbb{E}[X \mid \mathscr{G}]$. The key here is that since $\mathscr{G} = \sigma(\{G_n : n \geq 0\})$, we have that any $\mathscr{G}$-measurable function W is constant on the G_k's, indeed: [JUSTIFY]

Hence,

$$W = \sum_{n=1}^{\infty} a_n \mathbf{1}_{G_n}$$

Thus,

$$\mathbb{E}[(X-W)^2] = \mathbb{E}\left[\left(\sum_{n=1}^{\infty}(X-a_n)\mathbf{1}_{G_n}\right)^2\right]$$
$$=$$

□

8.1 Conditioning on an event of zero probability

Sometimes we have a random vector (X, Y) and make an observation on X, and still want to obtain some information on Y. Naturally, if X is continuous, and X was observed to take the value $X = x$, this event has probability zero. However, we can still make a meaningful definition:

Definition 8.1.1 Let (X, Y) be a random vector with density $f_{X,Y}$ the conditional density of $(Y \mid X = x)$ is given by

$$f_{Y|X=x} = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Remark 21 (Comment on independence) It is easy to verify that if X and Y are independent, then

$$f_{Y|X=x} = f_Y$$

This gives a sensible way of defining conditional expectation of one random variable subject to another random variable.

Definition 8.1.2 The expectation of a random variable Y given that X takes value x is given by:

$$\mathbb{E}[Y \mid X = x] = \int_{\mathbb{R}} y f_{Y|X=x}(y) \mathrm{d}y =: \psi(x)$$

We may then try to make sense of $\mathbb{E}[Y \mid X]$ by defining it as the random variable $\psi(X)$.

Example 18 (Question sheet 10) Let $X \sim \Gamma(2, \lambda)$ and let Y conditioned on $X = x$ take the distribution $Unif(0, x)$. Find:

1. The joint density $f_{X,Y}$.
2. The marginal density of Y.
3. The conditional density of X given $Y = y$.
4. $\mathbb{E}[X \mid Y = y]$

Comment on the distribution $(X \mid Y = y)$ and the joint distribution of Y and $X - Y$

Solution.
1. To find the joint density we see that $f_{Y|X=x} = \mathbf{1}_{[0,x]}(y)$ and $f_X(x) = \lambda^2 x \exp(-\lambda x)\mathbf{1}_{\{\geq 0\}}(x)$. Putting this together we see that

$$f_{X,Y}(x, y) = \lambda^2 \exp(-\lambda x)\mathbf{1}_{[0,x]}(y).$$

2. From this we may compute the marginal of Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}x = \int_y^{\infty} \lambda^2 \exp(-\lambda x) \mathrm{d}x = \lambda \exp(-\lambda y)\mathbf{1}_{\{\geq 0\}}(y)$$

3. Now we have that $f_{X|Y=y} = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \lambda \exp(-\lambda(x - y))\mathbf{1}_{[0,x]}(y)$

4. A quick computation gives

$$\mathbb{E}[X \mid Y=y] = \lambda \exp(\lambda y) \int_{\mathbb{R}} x \exp(-\lambda x) \mathbf{1}_{[0,x]}(y) \mathrm{d}x = \frac{1}{\lambda} + y$$

Observe that given $Y=y$, $\mathbb{P}(X \le x)$ is given by some function $\Phi(x) = \int_{-\infty}^{x} f_{X|Y=y}(x)\mathrm{d}x$. Therefore $\mathbb{P}(X-y \le x) = \Phi(x+y)$ which is precisely the distribution function of an exponential distribution, in other words, subject to $Y=y$, $X-Y \sim Exp(\lambda)$ and since this distribution is independent of what value Y takes, it follows that $X-Y$ and Y are independent Exponential distributions with parameter λ, and as such the joint PDF splits as a product. □

Proposition 8.1.1 (Tower Law)

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$$

Key ideas: Use the law of total probability inside an integral. □

Remark 22 The expectation of a random variable conditioned to another random variable taking a value, i.e:

$$\mathbb{E}[X \mid Y=y]$$

is a linear operator in its own right!

9 Gaussian Random Variables

Definition 9.0.1 A **Gaussian Random Variable** X is a Random Variable distributed according to a **Normal Distribution**, $X \sim N(\mu,\sigma^2)$. That is to say, X is a Gaussian if its density is of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Proposition 9.0.1 (Scaling of the Normal Distribution) If $X \sim N(0,1)$, then $Z = \sigma X + \mu \sim N(\mu,\sigma^2)$

Idea of proof. The idea of the proof is to use moment generating functions, recall that

$$m_X(t) := \mathbb{E}[\exp(tX)]$$

So in particular, for the case of $X \sim N(0,1)$ we have

$$m_X(t) = \exp\left(\frac{t^2}{2}\right)$$

and generally, if $X \sim N(\mu, \sigma^2)$, then

$$m_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$$

And it is known that if two random variables X and Y with the same domain have that for every t in some neighborhood of 0, $(-\epsilon, \epsilon)$ the two MGFs agree, then $X \stackrel{d}{=} Y$. □

Proposition 9.0.2 **Independent** Normal distributions are closed under linear combinations.

Proof. Using the MGF of $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, one can check that

$$m_{X_1+X_2} = \exp\left(t(\mu_1 + \mu_2) + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2} \right)$$

hence $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ □

Remark 23 The converse is generally not true!

This motivates the next section

9.1 Gaussian Random Vectors

Definition 9.1.1 A random vector (X, Y) is Gaussian if: **for all** $b, c \in \mathbb{R}$:

$$bX + cY \text{ is a one-dimensional Gaussian}$$

Remark 24 We always have that for any two random variables X, Y:

$$\mathbb{E}[bX + cY] = c\mathbb{E}[X] + c\mathbb{E}[Y]$$

and

$$\mathrm{Var}[bX + cY] = b^2 \mathrm{Var}[X] + c^2 \mathrm{Var}[Y] + 2bc\,\mathrm{Cov}[X, Y]$$

so one may rewrite this in terms of Linear Algebra to see that:

$$\mathbb{E}[bX + cY] = (b\ c)\begin{pmatrix} \mathbb{E}[X] \\ \mathbb{E}[Y] \end{pmatrix}$$

and

$$\mathrm{Var}[bX + cY] = (b\ c)\begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}[X, Y] \\ \mathrm{Cov}[Y, X] & \mathrm{Var}[Y] \end{pmatrix}(b\ c)^T$$

We can make our definition slightly more general now:

Definition 9.1.2 (Gaussian vector) A random vector $X = (X_1, \cdots, X_n) \in \mathbb{R}^n$ is a **Gaussian vector** if **given any** $u \in \mathbb{R}^n$

$$\langle u, X \rangle$$

is a **one-dimensional** Gaussian vector. Given a random vector $X \in \mathbb{R}^n$, we define the **expectation** $\mu = \mathbb{E}[X]$ to be coordinate-wise, i.e:

$$\mu = (\mathbb{E}[X_1], \cdots, \mathbb{E}[X_n]) \in \mathbb{R}^n$$

and the **covariance matrix** V to be as

$$V = (\mathrm{Cov}[X_i, X_j])_{i,j} = \mathbb{E}[\langle X - \mu, X - \mu \rangle] \in \mathrm{Mat}_{n \times n}(\mathbb{R})$$

Proposition 9.1.1 (Projection of a Gaussian vector) If X is a Gaussian vector with mean μ and covariance matrix $V = \mathrm{Cov}(X)$, then given any $u \in \mathbb{R}^n$

$$\langle u, X \rangle \sim N\left(\langle u, \mu \rangle, \langle u, Vu \rangle\right)$$

Proof as per Homework sheet 10. We first compute the expectation:

$$\mathbb{E}[\langle u, X \rangle] = \sum_{i=1}^{n} u_i \mathbb{E}[X_i] = \sum_{i=1}^{n} u_i \mu_i = \langle u, \mu \rangle$$

Similarly:

$$\mathrm{Var}\left[\sum_{i=1}^{n} u_i X_i\right] = \sum_{i,j} \mathrm{Cov}[u_i X_i, u_j X_j] == \sum_i u_i \sum_j \mathrm{Cov}(X_i, X_j) u_j = \sum_i u_i (\mathrm{Cov}(X) u)_i = \langle u, \mathrm{Cov}(X) u \rangle$$

□

Question 74 (2022 Mock) Let (X_n) be a sequence of Gaussian vectors, with mean $\mu_n = \left(1 + \frac{1}{n}, 2 - \frac{1}{n}\right)$ and covariance matrix

$$\mathrm{Cov}(X_n) = \begin{pmatrix} 2 & 1 + \frac{1}{n^2} \\ 1 + \frac{1}{n^2} & 3 \end{pmatrix}$$

If $X_n = (Y_{n,1}, Y_{n,2})$, what is the distribution of $2Y_{n,1} - 2Y_{n,2}$?

Solution. We note that $2Y_{n,1} - 2Y_{n,2} = \langle u, X_n \rangle$ where $u = (2, -2)$, so we know that $2Y_{n,1} - 2Y_{n,2}$ is a one-dimensional Gaussian and as such we only need to finds its expectation and variance to conclude the problem. By using linearity of expectation we easily see that

$$\mathbb{E}[2Y_{n,1} - 2Y_{n,2}] = -2 + \frac{4}{n}$$

Now to find the variance, we note that

$$\mathrm{Var}[2Y_{n,1} - 2Y_{n,2}] = \mathrm{Cov}[2Y_{n,1} - 2Y_{n,2}, 2Y_{n,1} - 2Y_{n,2}]$$

and now by bilinearity this is equal to

$$4\,\mathrm{Var}[Y_{n,1}] - 8\,\mathrm{Cov}[Y_{n,1}, Y_{n,2}] + 4\,\mathrm{Var}[Y_{n,2}]$$

Now use the covariance matrix to fill in the details. □

Proposition 9.1.2 (Properties of Linear Combinations of Gaussian vectors) Let X be a random Gaussian vector, $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ and $b \in \mathbb{R}^n$. It is easy to see that $AX + b$ is a Gaussian random vector. Moreover:

1. $\mathbb{E}[AX + b] = A\mathbb{E}[X] + b$
2. $\mathrm{Cov}(AX) = A\,\mathrm{Cov}(X)A^T$, where $\mathrm{Cov}(X)$ is the covariance matrix V.

Proof. I use Einstein's summation notation, tough shit if you don't understand it.

1. $$(\mathbb{E}[AX + b])_i = \mathbb{E}[A_{ij}X^j + b_i] = A_{ij}\mathbb{E}[X^j] + b_i = A_{ij}\mathbb{E}[X]^j + b_i = (A\mathbb{E}[X] + b)_i$$

2. $$\mathrm{Cov}(AX) = \mathbb{E}[(AX - A\mu)(AX - A\mu)^T] = \mathbb{E}[A(X-\mu)(X-\mu)^T A^T] = A\mathbb{E}[(X-\mu)(X-\mu^T)]A^T = A\,\mathrm{Cov}(X)A^T$$

□

Question 75 (2023 B4) Show that if A is an $n \times n$ matrix and X is a Gaussian vector, then AX is a Gaussian vector.

Proof. Easy, let $u \in \mathbb{R}^n$ be any given vector, let us compute

$$\langle u, AX\rangle = \left\langle A^T u, X\right\rangle$$

Clearly $A^T u$ is just another vector in $\mathbb{R}^n$. X being Gaussian means that this latter vector is also a one dimensional Gaussian as required. □

9.2 MGFs in $\mathbb{R}^n$

Definition 9.2.1 Let X be a random vector in $\mathbb{R}^n$, the MGF of X is given by

$$m_X(u) = \mathbb{E}[\exp(\langle u, X\rangle)]$$

for some $u \in \mathbb{R}^n$

Remark 25 The corresponding result about MGFs in the case of random vectors is that if two random vectors X and Y have that for some neighborhgood $[-\epsilon, \epsilon]^n$, their MGFs are both finite and agree, then the two random vectors are equal in distribution.

In some situations, including the Gaussians, the fast decay will ensure the MGF converges for any $u \in \mathbb{R}^n$

Proposition 9.2.1 (MGF of Gaussian Random Vector) This is the analogous result to Proposition 9.0.2. We have that if X is a Gaussian random vector, then

$$m_X(u) = \exp\left(\langle u, \mu\rangle + \frac{1}{2}\langle u, Vu\rangle\right)$$

Proof. Immediate consequence of Proposition 9.1.1 □

9.3 Bivariate Gaussians

Proposition 9.3.1 Let (X, Y) be a random Gaussian vector. Then X and Y are idendependent if and only if $\mathrm{Cov}(X, Y) = 0$.

Remark 26 This does **not** hold in general for other random variables!

Key idea of proof. The key is that it is **generally true**, that two random variables X and Y are independent if and only if the MGF of (X, Y) splits as a product of a function of X_1 and a function of X_2. Thus one can check that if $\mathrm{Cov}(X, Y) = 0$, then the corresponding MGF splits. □

Proposition 9.3.2 Let (X, Y) be a Gaussian random vector. Then there exists a scalar $a \in \mathbb{R}$ and a Gaussian random variable Z such that:

1. $Y = aX + Z$
2. (X, Z) are independent.

Key ideas.

1. Take $Z = Y - aX$. Then obviously Z is Gaussian.
2. Observe that (X, Z) is a Gaussian vector.
3. To force independence of X and Z, observe that

$$\mathrm{Cov}[X, Z] = \mathrm{Cov}[X, Y - aX]$$

now use bilinearity of Covariance and set a so that the left hand side vanishes. □

9.4 Standard Gaussians

Observe that we have waffled quite a lot about Gaussian vectors but actually we haven't really proved that any Gaussian vector exists at all! In this subsection we will aim to combine the previous results to try to write any given Gaussian vector as some combination of the **Standard Gaussian Vector**, much like the result of Proposition 9.0.1.

Definition 9.4.1 The Standard Gaussian vector in $\mathbb{R}^n$ is the random vector $X=(X_1,\cdots,X_n)$ where each $X_i \sim N(0,1)$. Its density is gien by

$$f_X(u)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_i^2\right)=\frac{1}{(2\pi)^{n/2}}\exp(-\frac{1}{2}||x||^2)$$

Remark 27 In view of Proposition 9.0.2, since each X_i is independent, it follows that X as above is indeed a Gaussian random vector.

To achieve this goal of ours, we resort to some Linear Alebra, recall the following thing from Year 1:

Lemma 9.4.1 Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ be a positive semi-definite matrix, that is to say, for any $u \in \mathbb{R}^n$, $\langle u, Au\rangle \geq 0$, then there exists some matrix $B \in \mathrm{Mat}_{n\times n}$ such that $A = BB$.

Then one shows the following:

Lemma 9.4.2 The covariance matrix is positive semi-definite.

Now we are ready.

Theorem 9.4.1 (Existence of Gaussian Random Vectors) Let $\mu \in \mathbb{R}^n$ be given and V be a positive semi-definite matrix, then there exists a Gaussian Random Vector X such that $\mathbb{E}[X]=\mu$ and $\mathrm{Cov}(X)=V$. Moreover, if $\det(V)\neq 0$, the density of X is given by

$$f_X(x)=\frac{1}{(2\pi)^{n/2}\sqrt{\det(V)}}\exp\left(-\frac{1}{2}\left\langle x-\mu, V^{-1}(x-\mu)\right\rangle\right) \qquad x\in\mathbb{R}^n$$

Main idea of proof. Construct X as $X=V^{1/2}Z+\mu$, where Z is the Standard Gaussian and $V^{1/2}$ is the "square root" of V. Now show the rest. □

10 Random Walks

Definition 10.0.1 Let $(X_n)_{n\geq 1}$ be a sequence of IID random variables. The sequence of random variables $(S_n)_{n\geq 0}$ where $S_0 = 0$ and otherwise $S_n = \sum_{i=1}^{N} X_i$ is called a random walk.

Example 19 (Simple random walk) Let $X \sim 2 \cdot Ber(1/2) - 1$ and construct the random walk with the sequence (X_n) where each $X_n = X$. For example, let n be even. If we wish to calculate $\mathbb{P}(S_n = 0)$, the best strategy is enumeration:

$$\mathbb{P}(S_n = 0) = \frac{\#\{\text{paths from } (0,0) \text{ to } (n,0)\}}{\#\{\text{total paths}\}} = \frac{1}{2^n}\binom{n}{n/2}$$

Example 20 (Biased simple random walk) Similarly to before, we can take $X \sim 2 \cdot Ber(p) - 1$ and construct the random walk in the same way we did. Now we may wonder what happens to (S_n) as $n \to \infty$ in view of **the drift**, $\mathbb{E}[X]$.

10.1 Limiting Results

Theorem 10.1.1 (Strong Law of Large Numbers) Let $(S_n)_{n\geq 0}$ be a random walk with $\mathbb{E}(S_1) = \mu < \infty$. Then we have almost surely $S_n/n \to \mu$

Question 76 (2019 3 (b)) Let $X_1, X_2, \cdots$ be IID random variables each with density $f(x) = \frac{1}{(x+1)^2}$ for $x \geq 0$. Does the SLLN apply to the sample average $\frac{1}{n}\sum_{i=1}^{n} X_i$?

Solution. It does not. Indeed:

$$\mathbb{E}[X_1] = \int_0^\infty \frac{x}{(x+1)^2}\,\mathrm{d}x = \infty$$

□

Theorem 10.1.2 (Weak Law of Large Numbers) With (S_n) as before, we have that $S_n/N \to \mu$ in probability

Remark 28 Since almost sure convergence implies convergence in probability, it is no surprise that one of these theorems is called Strong and the other one is called Weak.

As another comment, it can be shown that convergence of a sequence of random variables to a constant in probability is equivalent to it converging in distribution, thus we may phrase the Weak Law in either way, it doesn't really matter.

Question 77 (Homework sheet 11) Define the following sequence of independent random variables

$$X_i = \begin{cases} 1 & \mathbb{P} = 1/2 \\ 0 & \mathbb{P} = 1/2 - n^{-\alpha}/2 \\ -n^{\alpha} & \mathbb{P} = n^{-\alpha}/2 \end{cases}$$

Define the random walk $S_n = X_1 + \cdots X_n$

1. By adapting the proof of WLLN, (that is, using Chebyshev's inequality) show that $S_n/n \xrightarrow{\mathbb{P}} 0$ when $\alpha < 1$
2. Show that SLLN does not hold for $\alpha > 1$, yet $\mathbb{P}(S_n/n \text{ converges}) = 1$. Explain.
3. Show that when $\alpha = 1$, $\mathbb{P}(S_n/n \text{ converges}) = 0$.

Solution. 1. Recall Chebyshev's inequality:

$$\mathbb{P}(|X - \mu| > k\sigma) \leq 1/k^2$$

Before doing anything, let us compute μ. That's not hard, because $\mathbb{E}[X_i]$ is easily verified to be zero. Therefore $\mathbb{E}[S_n] = 0$. Our setup is that given $\epsilon > 0$, we can bound:

$$\mathbb{P}(|S_n| \geq \epsilon n) = \mathbb{P}\left(|S_n| \geq \frac{\epsilon n \sigma}{\sigma}\right) \leq \frac{\mathrm{Var}(S_n)}{\epsilon^2 n^2}$$

So our goal is to show that $\mathrm{Var}(S_n) = o(n^2)$. Let us compute it then:

$$\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Observe that for any n,

$$\mathrm{Var}(X_n) = \mathbb{E}[X_n^2] = \frac{n^{\alpha} + 1}{2} = O(n^{\alpha})$$

Hence

$$\frac{1}{n^2}\mathrm{Var}(S_n) = O\left(\frac{n^{\alpha+1}}{n^2}\right) = O(n^{\alpha-1})$$

When $\alpha < 1$, it follows that the desired quantity converges to zero, and as such we have convergence in probability.

2. Observe that $\mathbb{P}\{X_n = -n^{\alpha} \text{ infinitely often}\}$ is zero due to a simple consequence of the Borel Cantelli Lemma. This means that with probability one, X_n eventually takes values 1 and 0 with equal probability and as such the proportion converges to $1/2$ almost surely. This violates the SLLN which prescribes the limit to be $\mathbb{E}[S_n] = 0$.

3. To show this one, we employ an **approximation argument**, its good to have seen this argument

at least once: By similar reasoning as before, now that $\alpha = 1$, with probability one, X_n will take the value $-n^\alpha$ infinitely often. When this happens, $S_n - S_{n-1} = -n^\alpha$ hence

$$\left|\frac{S_n}{n} - \frac{S_{n-1}}{n-1}\right| \approx 1$$

For S_n/n to converge it is necessary that successive terms become closer as $n \to \infty$, yet we see that infinitely often, a term will be approximately a distance of 1 away from the next term.

□

Theorem 10.1.3 (Central Limit Theorem) With S_n as before, and now assuming finite variance of the increments, i.e: $\text{Var}(X_1) < \infty$ one has that

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1)$$

Remark 29 (Interpretations of CLT) From the CLT we gather that:

1. S_n becomes concentrated around $n\mu$.
2. The fluctuations about $n\mu$ are of order $\sqrt{n}$.
3. The exact distribution becomes approximately normal.

10.2 Using the Limiting Results

10.2.1 Strong and Weak Laws

Example 21 (Generalised coin tossing) Let $X \sim Ber(p)$ and construct the random walk S_n via X. It is not hard to see that $S_n \sim Bin(n,p)$ and the quantity S_n/n represents the proportion of successes among the first n trials. Observe that $\mathbb{E}[X] = p$. Let's interpret S_n/n as $n \to \infty$ in view of the laws of large numbers.

1. The WLLN tells us that the probability of finding the proportion of successes to be away from p becomes arbitrarily small as $n \to \infty$. In other words, the proportion of successes is very likely to be close to p when n is large.
2. The SLLN tells us that in fact, with probability 1, the proportion of successes will converge to the constant p as n gets large. This is what one may take as an intuitive definition of probability!

Knowing the proportion of course tells us information about S_n as well. Knowing that almost surely

$S_n/n \to \mu = \mathbb{E}[X]$ tells us that almost surely:

$$S_n \to \begin{cases} +\infty & \text{if } p > 0 \\ -\infty & \text{if } p < 0 \end{cases}$$

this follows by a very simple analytic argument.

Example 22 (Poisson Process) Let $X \sim Exp(\lambda)$. We know X represents the time taken for an arrival to occur. We may build a random walk S_n by using X and then S_n/n represents the total time taken for the n^{th} arrival to occur in proportion to n. It is not a surprise then that the SLLN guarantees that with probability one,

$$S_n/n \to \mathbb{E}[X] = 1/\lambda$$

because S_n/n as $n \to \infty$ symbolises the average waiting time. Instead of this, let us try to understand how many arrivals occured during a given timespan.

$$N(t) := \#\{\text{number of arrivals during} t \text{ time units}\} = \max\{n : S_n < t\}$$

To convince yourself of this last line, notice that S_n measures the total time taken for the n^{th} arrival to occur, so $\max\{n : S_n < t\}$ measures how many arrivals occured before the total time surpassed t.

Proposition 10.2.1

$$\frac{N(t)}{t} \xrightarrow{a.s} \lambda$$

Proof. We begin by noting the two following facts

$$N(S_n) = n \qquad S_{N(t)} \leq t < S_{N(t)} + 1$$

Moreover, since S_n/n converges almost surely to $1/\lambda$ as $n \to \infty$ and $N(t) \to \infty$ as $t \to \infty$, then $N(t)/S_{N(t)} \to \lambda$ almost surely as $t \to \infty$. Now we sandwich the desired quantity:

$$\frac{N(t)}{S_{N(t)+1}} \leq \frac{N(t)}{t} \leq \frac{N(t)}{S_{N(t)}}$$

equivalently

$$\frac{N(t)+1}{N(t)+1} \frac{N(t)}{S_{N(t)+1}} \leq \frac{N(t)}{t} \leq \frac{N(t)}{S_{N(t)}}$$

Now observing that almost surely

$$\frac{N(t)}{N(t)+1} \to 1$$

as $t \to \infty$ finishes the claim. □

Observe how in this argument, by sandwiching the quantities, we have converted SLLN from discrete to continuous times.

Question 78 (Homework Sheet 11) Let $(X_1, X_2, \cdots)$ be a collection of independent random variables such that $(X_2, \cdots)$ are also identically distributed. Given that $\mathbb{E}[X_2] = \mu < \infty$, show that

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s} \mu$$

Proof. Observe that since a single random variable cannot take the value ∞, it follows that almost surely $X_1/n \to 0$ as $n \to \infty$. Now observe that

$$\frac{\sum_{i=2}^{n} X_i}{n} = \frac{n+1}{n} \frac{1}{n+1} \sum_{i=2}^{n} X_i \xrightarrow{a.s} \mu$$

combine this and fuck yourself. □

Question 79 (Homework Sheet 11) Let $X_1, \cdots, X_n$ be Poisson Random variables with parameter one. What is the distribution of $X_1 + \cdots + X_n$? By writing

$$\exp(-n)\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right)$$

as a probability, use the CLT to show this quantity converges to $1/2$ as $n \to \infty$.

Proof. We know that the MGF of the sum of two independent Poisson random variables with parameters λ_1 and λ_2 is

$$m_{X_1+X_2}(t) = \mathbb{E}[\exp([X_1 + X_2]t)] = \mathbb{E}[\exp(X_1 t)]\mathbb{E}[\exp(X_2 t)] = \exp(-\lambda_1(1-t))\exp(-\lambda_2(1-t))$$

which is the MGF of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Therefore by induction we see that $S_n = X_1 + \cdots X_n \sim Poisson(n)$. We see that the quantity in question is precisely

$$\mathbb{P}(S_n = 0) + \mathbb{P}(S_1 = 1) + \cdots + \mathbb{P}(S_n = n) = \mathbb{P}(S_n \leq n)$$

We note that $\mathbb{E}[X_1] = \mathrm{Var}[X_1] = 1$, and as such:

$$\mathbb{P}(S_n \leq n) = \mathbb{P}\left(\frac{S_n - n}{\sqrt{n}} \leq 0\right) \to \mathbb{P}(Z \leq 0) = 1/2$$

where $Z \sim N(0,1)$. □

Question 80 (Homework sheet 11) What is the distribution of the sum of n independent Bernoulli random variables with parameter p? Let $0 \le \alpha < \beta \le 1$. With this inspiration, determine

$$\lim_{n\to\infty} \sum_{r\in\mathbb{N}:\alpha n\le r\le n\beta} \binom{n}{r} p^r(1-p)^{n-r}$$

Solution. We know that the sum of Bernoullis as given is distributed according to $Binomial(n,p)$. Coincidentally, the sum in question is precisely: $\lim_{n\to\infty}\mathbb{P}(\alpha n \le B \le n\beta)$ where $B \sim Bernoulli(n,p)$. We know that $\mathbb{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$. Therefore we may rewrite our sum as

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{\alpha n - np}{n\sqrt{p(1-p)}} \le \frac{B-np}{n\sqrt{p(1-p)}} \le \frac{\beta n - np}{n\sqrt{p(1-p)}}\right) = \Phi\left(\frac{\beta - p}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{\alpha - p}{\sqrt{p(1-p)}}\right)$$

□

Question 81 (Homework sheet 10) Let $(X_i) \sim N(0,1)$ be a family of IID normals, and define the random walk $S_n = \sum_{i=1}^n X_i$. Prove that the random vector

$$S = (S_1, \cdots, S_n)$$

is a Gaussian random vector.

Proof. Let $u \in \mathbb{R}^n$ be given, then

$$\langle u, S\rangle = \sum_{i=1}^n u_i \left(\sum_{j=1}^i X_j\right) = \sum_{j=1}^n \left(\sum_{i=j}^n u_i\right) X_j$$

which is a linear combination of IID normals. Hence another normal distribution. Recall that **it is absolutely crucial** that the normals are assumed to be independent from the start, because we know that if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then we can guarantee that $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Otherwise we cannot guarantee that this sum is also a normal. □

Question 82 (Homework sheet 10) Let $X \sim N(0,1)$ and define

$$Y = \begin{cases} X & |X| < 100 \\ -X & |X| \ge 100 \end{cases}$$

Show that Y is normally distributed. Is (X, Y) a Gaussian random vector?

Solution. We shall study the distribution of Y, that is to say $\mathbb{P}(Y \le y)$. And we shall do so via the Law of Total Probability:

$$\mathbb{P}(Y \le y) = \mathbb{P}(Y \le y \mid |X| < 100)\mathbb{P}(|X| < 100) + \mathbb{P}(Y \le y \mid |X| \ge 100)\mathbb{P}(|X| \ge 100)$$

Call $k_1 = \mathbb{P}(|X| < 100)$ and $k_2 = \mathbb{P}(|X| \geq 100)$ for convenience, and we may rewrite the above as:

$$\mathbb{P}(Y \leq y) = k_1 \mathbb{P}(X \leq y) + k_2 \mathbb{P}(X \geq -y)$$

Since X has mean zero, it is symmetric about zero and as such $\mathbb{P}(X \geq -y) = \mathbb{P}(X \leq y)$ (Sketch the distribution to convince yourself of this). And as such:

$$\mathbb{P}(Y \leq y) = (k_1 + k_2)\mathbb{P}(X \leq y) = \mathbb{P}(X \leq y)$$

so in fact $Y \sim N(0,1)$.

Finally we note that (X, Y) is not a Gaussian vector. Indeed, if it were, we could take the linear combination $X + Y$ which should be a Gaussian, in which case $\mathbb{P}(X + Y = 0)$ should be zero because the probability of a Gaussian taking any single value is always zero, however, we see that in this case, that probability is the same as the probability of $|X| \geq 100$, which is most definitely not zero. $\square$

10.3 Exam Questions

Question 83 (2023 B4) Let $W \sim N(0,1)$ and define, conditional on $W = w$, the IID random variables $(Z_1, \cdots)$ distributed according to $N(\omega, \sigma^2)$. Consider the random walk $R_n = \sum_{i=1}^n Z_i$.

1. Show that (W, R_n) is a Gaussian random vector.
2. Find $\mathbb{P}(\{R_n \geq 0$ for infinintely many $n\})$

Proof. First of all, let's make the following "moral observation" about what this question is saying:

$$\mathbb{P}(\{Z_i \leq x\} \mid \{W = \omega\}) = \mathbb{P}(Y \leq x) \qquad Y \sim N(\omega, \sigma^2)$$

From this, we see that

$$\mathbb{P}(\{Z_i - W \leq x\} \mid \{W = \omega\}) = \mathbb{P}(Y' \leq x) \qquad Y' \sim N(0, \sigma^2)$$

This is of course just a "moral observation" because we cannot condition on an event of probability zero. Regardless, since the distribution of the random variable $(Z_i - W \mid \{W = w\})$ is irrespective of ω, we gather that the density of $(Z_i - W \mid \{W = \omega\})$, $f_{Z_i - W|W=\omega}$ has no ω dependence. Now using the fact that

$$f_{Z_i - W|W=\omega} = \frac{f_{Z_i - W, W}(x, \omega)}{f_W(\omega)}$$

we see that in fact $f_{Z_i - W, W}(x, \omega)$ splits as a product of two functions depending on each parameter and from this we gather that $Z_i - W$ is independent of W. Thus we may write $Z_i = X_i + W$ where $X_i \sim N(0, \sigma^2)$ and X_i and W are independent. From this we easily conclude that (W, R_n) is a Gaussian vector.

Now we show the second part. From the Law of total probability we see that

$$\mathbb{P}(\{R_n \geq 0 \text{ i.o}\}) = \mathbb{P}(\{R_n \geq 0 \text{ i.o}\} \mid \{W \geq 0\})\mathbb{P}(\{W \geq 0\}) + \mathbb{P}(\{R_n \geq 0 \text{ i.o}\} \mid \{W \leq 0\})\mathbb{P}(\{W \leq 0\})$$

Observe that $R_n \geq 0$ if and only if $R_n/n \geq 0$, and given that $R_n/n \to \omega$, $R_n \geq 0$ if and only if, in the almost sure limit, $R_n/n \geq 0$, or in other words, if $\omega \geq 0$. Noticing that since W is symmetric, $\mathbb{P}(W \geq 0) = \mathbb{P}(W \leq 0) = 1/2$, we can rewrite

$$\mathbb{P}(\{R_n \geq 0 \text{ i.o}\}) = \frac{1}{2}\mathbb{P}(\{R_n \geq 0 \text{ i.o}\} \mid \{W \geq 0\}) + \frac{1}{2}\mathbb{P}(\{R_n \geq 0 \text{ i.o}\} \mid \{W \leq 0\})$$

The first of these probabilities is one, and the second is zero. □

10.4 Moment Generating Functions in limiting processses

We have seen before that if two Random Variables have MGFs that agree on a neighborhood, then the random variables are equal in distribution. Now we are going to see an analogous result in terms of sequences. The following Theorem will explain why if a sequence of random variables have MGFs converging to the MGF of some limiting random variable, then we actually have convergence in distribution of this sequence to said random variable.

Theorem 10.4.1 (Lévy's continuity Theorem) Let (X_n) be a sequence of Random Variables and X be a random variable with MGFs $m_{X_N}(t)$ and $m_X(t)$ respectively. Suppose $m_X(t) < \infty$ for all t in some neighborhood $(-\epsilon, \epsilon)$ of zero. Then if

$$m_{X_N}(t) \to m_X(t) \qquad \text{for all} \, t \in (-\epsilon, \epsilon)$$

we have convergence in distribution $X_n \to X$.

Question 84 (2024 Specimen) Let (Z_n) be a sequence of random variables defined by $Z_n \sim N(\mu_n, \sigma_n^2)$. Where μ_n is a sequence converging to μ and σ_n^2 converges to σ^2. Show that $Z_n \xrightarrow{d} Z \sim N(\mu, \sigma_n^2)$.

Proof. We employ Lévy's continuity Theorem. To do so, we must show that the Moment Generating Functions of (X_n) exhibit convergence towards that of X.

$$m_{X_n}(t) = \exp\left(\mu_n t - \frac{\sigma_n^2 t^2}{2}\right)$$

Clearly $m_{X_n}(t)$ is a continuous function in t so we may pass the limit $n \to \infty$ inside it, and as such we see that

$$m_{X_n}(t) \to m_X(t)$$

which is in turn defined for all t due to the fast decay of the exponential. The convergence in distribution follows immediately. □

Question 85 (2024 Specimen Spinoff) With (Z_n) defined as in the previous question and assumed to be independent, define the random walk $R_n = \sum_{i=1}^n Z_n$? An inexperienced probability student would think that since $Z_n \to Z \sim N(\mu, \sigma^2)$ in distribution, then the Central Limit Theorem, says that

$$\frac{R_n - n\mu}{\sqrt{n}\sigma} \to N(0,1)$$

Is this true?
Hint: take $\sigma_n \equiv 1$ and you may use the fact that $(n+1)^\alpha - n^\alpha \to 0$ as $n \to \infty$ for all $0 < \alpha < 1$.

Proof. It's false, indeed, define $Z_n \sim N(\mu_n, \sigma_n^2)$ where $\mu_n = (n+1)^\alpha - n^\alpha$ and $\sigma_n \equiv 1$. Then since the Z_i are assumed to be independent:

$$\frac{R_n}{\sqrt{n}} \sim N\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mu_i, 1\right)$$

but with our choice of μ_n:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (i+1)^\alpha - i^\alpha = \frac{(n+1)^\alpha}{\sqrt{n}}$$

pick $1/2 < \alpha < 1$. Then this quantity goes to $+\infty$ as n gets large. $\square$

11 Martingales

References

[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[Sto13] Jordan M Stoyanov. *Counterexamples in Probability Third Edition*. Dover Publications, 2013.