

Probability theory notes

Kotatsu

TOALDO COL CAZZO CHE DIVENTI ORDINARIO

Preface

This document stems from the fact that I just seem unable to pass the Probability Theory exam for the life of me. I regret with every ounce of my being the fact that I enrolled to the Stochastics and Data Science master degree a year ago. Since dear Professor Toaldo never really thrilled me with his insightful lectures about this delightful topic, I resorted to watch the old lectures by Professor Polito, who at least seems to know the subject and to be determined to explain it.

Unlike many among my esteemed colleagues I have NOT a background in mathematics so there will be a lot of repetitions and possibly mistakes. Do what you want with this information. YES I KNOW that there are the whiteboard registrations of his lectures but if I DECIDED TO DO THIS it was because I couldn't comprehend shit with only those notes.

I'll also try to compile the notes made by Professor Sacerdote, in the vain attempt to overcome the drowsiness that is congenitally entwined with every event that contemplates her uttering any words.

I take much pride in my custom environment and in my packages. If you don't like them I will be very sad. It is strongly recommended to play Metal Gear Solid, Metal Gear Solid 2 and Metal Gear Solid 3 before reading these notes to fully understand the subject treated.

Kotatsu

Contents

1	The basics of probability theory	1
1.1	Basics of probability	1
1.1.1	Random variables	2
1.1.2	Functions of random variables	7
1.1.3	Infinite product spaces	9
1.1.4	Stochastic processes	10
1.1.5	Example of random variables	10
1.2	Transition kernels	17
1.2.1	Products of kernels	22
1.3	Expectation	25
1.3.1	Properties of expectation	27
1.3.2	L^p spaces	30
1.3.3	Uniform integrability	32
1.4	σ -algebras and random variables	34
1.4.1	Filtrations	38
1.4.2	Independency for a finite class of sub- σ -algebras	39
1.4.3	Sums of independent random variables	40
1.4.4	Tail σ -algebra	41
1.5	Convergence of random variables and asymptotic behavior	44
1.5.1	Convergence of real sequences	44
1.5.2	Almost sure convergence	47
1.5.3	Borel-Cantelli Lemmas	48
1.5.4	Convergence in probability	52
1.5.5	Convergence in L^p spaces	55
1.5.6	Weak convergence and convergence in distribution	58
1.5.7	The law of large numbers	61
1.5.8	Central limit theorem	66
1.6	Conditional expectations	69
1.6.1	Conditional expectation conditional on arbitrary σ -algebras	71
1.6.2	Uniqueness of conditional expectation	72
1.6.3	Properties of conditional expectation	74
1.6.4	Conditioning as projection	76
1.6.5	Conditional expectations given random variables	77
1.6.6	Conditional probabilities and distributions	78
1.6.7	Disintegration of measures	80
1.6.8	Conditional independence	81
2	A first study in stochastic processes: Martingales	83
2.1	Revising stochastic processes and filtrations	83
2.1.1	Stopping times	85
2.1.2	Discrete and continuous stopping times	92
2.1.3	Conditioning at stopping times	93
2.2	Martingales	95
2.2.1	Properties of martingales	98
2.2.2	Create your own martingale	99
2.2.3	Doob's decomposition	101
2.2.4	Important stochastic processes and related martingales	102
2.2.5	Integration in discrete time	105
2.2.6	Stopped martingales and Doob's stopping theorem	107

2.2.7	Oscillations	109
2.2.8	Limit of sequences of random variables	115
3	Miscellanea: an integration (eheh)	119
3.1	Gaussian random variables	119
4	Exercises and questions	122
4.1	Exercises with Prof. Issoglio (2024/2025)	122
4.1.1	Exercise class 1	122
4.1.2	Exercise class 2	128
4.1.3	Exercise class 3	133
4.1.4	Exercise class 4	138
4.1.5	Exercise class 5	143
4.1.6	Exercise class 6	150
4.1.7	Exercise class 7	157
4.1.8	Exercise class 8	165
4.2	Questions for the oral examination	171
4.2.1	Transition kernels: definition, example, and their usage in the extension of measures to product spaces	171
4.2.2	Kolmogorov's 0-1 law: proof and an example of its usage	171
4.2.3	Almost sure convergence: definition, properties and characterization theorem	172
4.2.4	Borel Cantelli lemmas; proof of the divergence part	173
4.2.5	Borel Cantelli lemmas; proof of the convergence part, relations with a.s. convergence and examples of its application	175
4.2.6	Convergence in probability: definition, properties and theorem on the metric for convergence in probability	176
4.2.7	Uniform integrability: definitions and its consequences	177
4.2.8	Law of large numbers: proof of the related theorems	177
4.2.9	Central limit theorem (Lyapunov theorem) Proof of the theorem and of the corresponding lemma (Lindeberg Lemma)	179
4.2.10	Definition of conditional expectation and its main properties	182
4.2.11	Existence and uniqueness of conditional expectation (for L1 random variables)	183
4.2.12	Existence and uniqueness of conditional expectation (for L2 random variables, using orthogonal projection)	184
4.2.13	Filtration, adaptedness; example of non adapted process	185
4.2.14	Random times and Stopping Times: examples of random times that are not stopping times and examples of stopping times	186
4.2.15	Stopped filtration: definition and examples	186
4.2.16	Discrete stopping times associated to continuous stopping times	187
4.2.17	Conditional expectation at stopping time	188
4.2.18	Comparing stopping times: some interesting features	189
4.2.19	Functions of stopping times that are (or are not) stopping times	190
4.2.20	Foretold times and their relationship with stopping times; examples of foretold times.	191
4.2.21	Examples of Martingales: Markov Chains	191
4.2.22	Definition of martingale, sub-martingale and super-martingale and their properties	192
4.2.23	Definition of sub/super martingale and examples	193
4.2.24	Maximum, minimum of martingales	194
4.2.25	Predictable processes and Doob's decomposition	197
4.2.26	Doob's stopping theorem	199
4.2.27	Upcrossing inequality	200
4.2.28	Doob's decomposition	202
4.2.29	Stochastic integral in discrete time	202
4.2.30	Stochastic integral and its application	203
4.2.31	Impossibility to win against the system: related theorems and examples	205
4.2.32	Predictable and adapted processes: definition and examples	206
4.2.33	Poisson process and its martingale property	207
4.2.34	Stopped processes and their properties	208
4.2.35	Important inequalities for sub-martingales	209
4.2.36	Convergence theorems for sub-martingales	211
4.2.37	Uniform integrability and its consequences on convergence of martingales	212
4.2.38	Features of the sample paths of a submartingale (or martingale)	213

4.2.39	Uniform integrability and its role in convergence problems	213
4.2.40	Stopped martingales	214

Chapter 1

The basics of probability theory

1.1 Basics of probability

We start with the probability triplet: $(\Omega, \mathcal{H}, \mathbb{P})$. Here Ω is the set of sample space, \mathcal{H} is the σ -algebra built upon Ω and \mathbb{P} is the probability measure. Since \mathbb{P} is a measure, it will take values in \mathbb{R} .

We are interested in probability measure, which means:

- \mathbb{P} is a **finite measure** and $\mathbb{P}(\Omega) = 1$;
- $\omega \in \Omega$ will be called **outcomes**.

So consider the example of the roll of the die. If we roll it,

$$\Omega = \underbrace{\{1, 2, 3, 4, 5, 6\}}_{\text{outcomes}}$$

And if we consider the elements $A \in \mathcal{H}$ (which will be subsets of Ω) will be called **events**.

We want to quantify the possibility that the event A occurs: we want to measure, through \mathbb{P} , the set A : from a measure theory point of view, it's only sets in the σ -algebra.

The probability measure has the following properties:

- $\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0$
- **monotonicity of \mathbb{P} :** take 2 events $H, K \in \mathcal{H}$ such that $H \subset K$. Then $\mathbb{P}(H) \leq \mathbb{P}(K)$ ¹.
- **finite additivity:** take $H, K \in \mathcal{H}$ such that $H \cap K = \emptyset$. Then $\mathbb{P}(H \cup K) = \mathbb{P}(H) + \mathbb{P}(K)$;
- **countable additivity:** this requires that we consider collection of events. We denote them in this way:

$$(H_n)_{n \in \mathbb{N}} \subset \mathcal{H}$$

with $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ and $\mathbb{N}^* = \{1, 2, 3, 4, \dots\}$ such that they are disjoint pairwise (except identical pairs). Then

$$\mathbb{P}\left(\bigcup_n H_n\right) = \sum_n \mathbb{P}(H_n)$$

- **Boole inequality (sub-additivity):** if we have a collection $(H_n)_{n \in \mathbb{N}} \subset \mathcal{H}$ (not necessarily disjoint) then

$$\mathbb{P}\left(\bigcup_n H_n\right) \leq \sum_n \mathbb{P}(H_n)$$

- **sequential continuity:** consider the sequence $(H_n)_{n \in \mathbb{N}} \subset \mathcal{H}$ such that $H_n \nearrow H \in \mathcal{H}$ (H_n is an increasing sequence of numbers that has H as limit) then $\mathbb{P}(H_n) \nearrow \mathbb{P}(H)$. Moreover, if $(F_n)_{n \in \mathbb{N}} \subset \mathcal{H}$ such that $F_n \searrow F \in \mathcal{H}$ then $\mathbb{P}(F_n) \searrow \mathbb{P}(F)$. The second property is actually true because \mathbb{P} is finite (it is not true for infinite measures).

¹note that the notation is loose since we have proper subset on one side and leq on the other side. But this is not much of a problem, since i will kill myself very soon.

In measure theory we encounter the concept of **negligible sets**: these are sets of measure zero or non measurable sets included in measure zero sets. In probability theory, sets are **events**: so we have negligible events (events with probability 0 or non measurable events included in events with probability 0). Analogously, in measure theory a property which holds **almost everywhere** is allowed not to hold on negligible sets. In probability theory a property which holds **almost surely** is allowed not to hold on negligible events. We also have, in measure theory, *measurable functions* that in probability theory are **random variables**. Let's have a look back into what the absolute fuck a measurable function is. Also what is an integral? This course deals with distributions, measures and other hellish machinery that servers the sole purpose to confuse you.

Revise with Kotatsu!

Definition 1.1.1

Let (E, \mathcal{E}) and (F, \mathcal{F}) be measurable spaces. A mapping $f : E \mapsto F$ is said to be **measurable** relative to \mathcal{E} and \mathcal{F} if

$$f^{-1}(B) \in \mathcal{E} \quad \forall B \in \mathcal{F}.$$

There is an useful property to measurable functions. Take a function $f : E \mapsto F$. In order for f to be measurable relative to \mathcal{E} and \mathcal{F} it is necessary and sufficient that

$$f^{-1}(B) \in \mathcal{E} \quad \forall B \in \mathcal{F}_0$$

where \mathcal{F}_0 is a collection that generates \mathcal{F} , i.e. $\mathcal{F} = \sigma(\mathcal{F}_0)$.

1.1.1 Random variables

Consider a measurable space (E, \mathcal{E}) .

Definition 1.1.2

A mapping $X : \Omega \rightarrow E$ is called **random variable taking values in E** if X is measurable relative to \mathcal{H} and \mathcal{E} .

What does it mean²? The inverse image of the set A through X ($X^{-1}A$) with $A \in \mathcal{E}$ is actually the set of the ω s such that $X(\omega)$ arrives to A . So

$$X^{-1}A = \{\omega \in \Omega : X(\omega) \in A\} = \{X \in A\}$$

so that $X^{-1}A$ is an event for all A in \mathcal{E} .

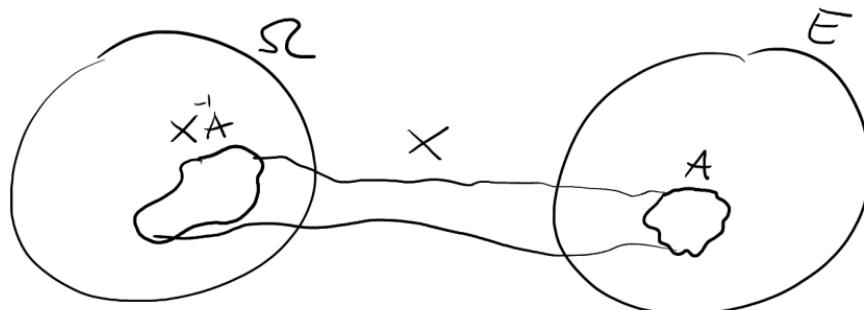


Figure 1.1: this is an early reminder of the fact that I will take my own life very soon.

So if $X^{-1}A$ is measurable by \mathbb{P} then it is in \mathcal{H} : otherwise it is not in \mathcal{H} . So

$$\mathbb{P}(X^{-1}A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

The message is that I am interested/able to evaluate \mathbb{P} over the set only if what I am evaluating is indeed an event (which means: it belongs to \mathcal{H} ³). If something is not in \mathcal{H} get it off my fucking face man and kill yourself

²who asked

³il lettore più arguto avrà notato che, a questo punto, il dio è ormai irrimediabilmente cane.

NOW⁴. This is the only restriction for a random variable. E can be whatever we need it to be: a graph, a tree, your mom being absolutely [REDACTED] by me. But most of the times, we have $E = \mathbb{R}$ or $E = \mathbb{R}^d$ with respectively $\mathcal{E} = \mathcal{B}^5(\mathbb{R}) = \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\mathbb{R}^d}$.

Remark

The simplest random variables are indicator functions of events. Example: take $H \in \mathcal{H}$. Define the function

$$\mathbb{1}_H : \Omega \rightarrow \mathbb{R}$$

$$\mathbb{1}_H(\omega) = \begin{cases} 0 & \omega \notin H \\ 1 & \omega \in H \end{cases}$$

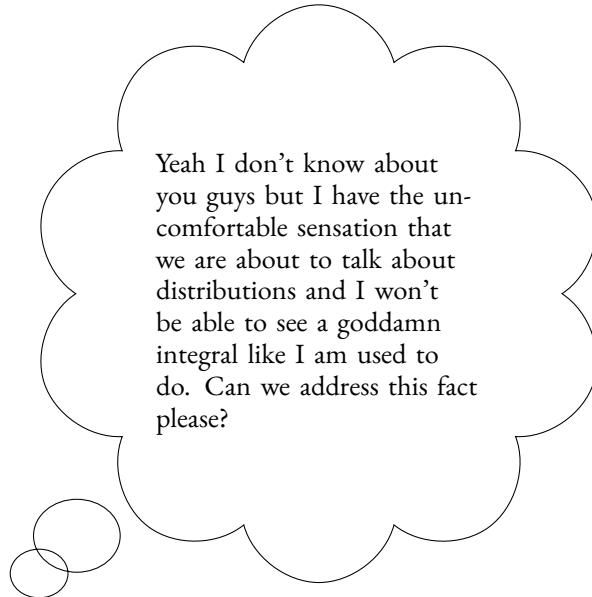
Remark

A random variable is said to be **simple** if it takes only finitely many values in \mathbb{R}^d .

Remark

A random variable is said to be **discrete** if it takes only countably many values.

We are now ready to define the concept of *distribution of a random variable*. But first...



Sure. Let's have a look back to Lebesgue integration.



⁴

⁵Borel σ -algebra. You don't know what a Borel σ -algebra is? https://en.wikipedia.org/wiki/Borel_set

Revise with Kotatsu!

Consider a measure space (E, \mathcal{E}, μ) . \mathcal{E} can be seen as the collection of all \mathcal{E} -measurable functions $f : E \mapsto \overline{\mathbb{R}}$ on E that can be denoted with an abuse of notation^a by $f \in \mathcal{E}$ and by $d \in \mathcal{E}_+$ if the functions are positive. Our aim is to define integrals of measurable functions with respect to the measure μ so that:

$$\mu f = \mu(f) = \int_E f(x) \mu(dx) = \int_E f d\mu$$

which is written as the product of μ and f . It is interesting to note, in the last part of the equation, that the integral reads something like: "integrate f over E with respect to the measure μ ". What is this measure?? This is the question. Turns out that the good old Riemann integral is just a particular case of the Lebesgue integral when a certain measure is chosen.

We consider them as the generalization of vectors and hence the scalar product becomes a sum, which transforms into an integral. We will define the Lebesgue integral in three steps:

1. Simple and positive functions:

Definition 1.1.3

The function f is called a **simple and positive function** if it can be written as

$$\sum_{i=1}^n a_i \mathbb{1}_{A_i}$$

where $A_i \in \mathcal{E}$ and $a_i \geq 0 \in \mathbb{R}$ for $i = 1, 2, \dots, n$.

Definition 1.1.4

For simple and positive functions, we define the Lebesgue integral as

$$\mu f := \sum_{i=1}^n a_i \mu(A_i)$$

2. Positive and measurable functions:

Theorem 1.1.1

Let $f \in \mathcal{E}_+$. Then there exists a sequence of simple and positive functions f_n such that $f_n \nearrow f$.

Thanks to this theorem, we can well pose the following definition"

Definition 1.1.5

Let $f \in \mathcal{E}_+$. We define

$$\mu f := \lim_n \mu f_n$$

where f_n is a sequence of simple and positive functions such that $f_n \nearrow f$.

3. Recall a general fact for real-valued functions.

Remark

Let f be a real-valued function. Then we can write

$$f = f^+ - f^-$$

With $f^+ := f \vee 0 = \max\{f, 0\}$, called **positive part** and $f^- := -(f \wedge 0) = -\min\{f, 0\}$, called **negative part**. Both of them are real and positive functions and f is measurable if and only if f^+ and f^- are real and positive functions.

We are now ready to define the Lebesgue integral for measurable functions in \mathbb{R} . The trick is to separate the positive and the negative part of the function, to treat them as the limit of sequence of simple functions and then lose ourselves in the bliss of measure theory.

Definition 1.1.6

Let $f \in \mathcal{E}$. We define

$$\mu f := \mu(f^+) - \mu(f^-)$$

Provided that at least one of the integrals is finite in order to be defined and not incur into indefinite forms like $+\infty$ or $-\infty$.

This definition can be easily converted if f is a complex function: we only have to remember that we can decompose any complex number in its real and imaginary part. Both of them will be measurable real functions.

$$f = \Re e^+ f - \Re e^- f + i(\Im m^+ f - \Im m^- f).$$

From now on we will use this notation for the Lebesgue integral on (E, \mathcal{E}) :

$$\mu f = \int_E f(x) \mu(dx) \quad \text{with } f \in \mathcal{E}$$

and if we choose $f = \mathbb{1}_B$ with $B \in \mathcal{E}$, then

$$\mu f = \mu \mathbb{1}_B = \int_E \mathbb{1}_B(x) \mu(dx) = \mathbb{1}_B \mu(dx) = \mu(B).$$

So this last equivalence helps us to understand one thing. Integrals are a device that needs a measure and a function to work. In the notation above, dx has the meaning of an infinitesimal amount of the variable x that is fed into the function f . Writing $\mu(dx)$ means measuring an infinitesimal amount of x using the measure μ .

In Riemann integration, dx represents an infinitesimal segment of the x -axis multiplied by the height of the function at x (which is, of course, $f(x)$) and summed (\int_a^b) with all the other infinitesimal segments of the x -axis over the interval $[a, b]$.

Here it's really the same thing with the difference that we multiply the height of the function $f(x)$ by calculating the "weight", or "measure" of a smaller and smaller part of the domain that "causes that function to be of that height", according to our method of measure of choice. We do this over the set E .

^aI am the only one being abused here.



The main difference between Riemann and Lebesgue integration is, in a certain way, *what* we are slicing. In the Riemann approach we basically do the following:

1. slice the x -axis in smaller and smaller slices;
2. compute $f(x)$;
3. sum all the cute little rectangles you got.

In the Lebesgue approach we basically *start by choosing different slices of the range of the function*, that is the co-domain. These little "slabs" of the range of the functions are nothing else but the "stepped" simple function version of our function:

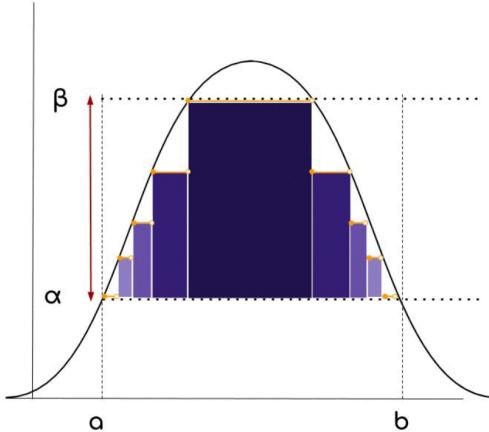


Figure 1.2: Imagine the steps getting smaller and smaller...

Since we are dealing with simple functions, we are effectively approaching the problem from the y -axis. This means that since we are choosing slices of height *first* our "slabs" may have different length when it comes to the x -axis. Anyway was it really SO DIFFICULT to explain? I don't think so. Fuck you mathematicians. So... will there ever be a measure and a set for which we will be able to circle back to our definition of Riemann integral? Hmm...

Definition 1.1.7

Distribution of a random variable. Let X be a random variable taking values in (E, \mathcal{E}) and let μ be the image of \mathbb{P} under X , that is,

$$\mu(A) = \mathbb{P}(X^{-1}A) = \mathbb{P}(X \in A) = \mathbb{P} \circ X^{-1}(A)^{\textcolor{violet}{a}}, A \in \mathcal{E}.$$

Then μ is a probability measure on (E, \mathcal{E}) and it is called **distribution of X** .

^ayou would know this if you knew fucking measure theory I guess

So we map, by means of X , sets belonging to \mathcal{E} into \mathcal{H} and then evaluates this sets by means of the measure \mathbb{P} . This is what we mean when we say that distributions are ultimately built with the probability measure and the random variable.

Distribution is itself a measure. To be exact it is a measure that we employ with a function (that in our case is a random variable) to form a Lebesgue integral just like we have seen in the revise box above. As we said, integrals are a machine that needs a function and a measure; in the case of probability theory these elements are respectively the **random variable** and the **probability distribution**.

Right now we can start to see the light at the end of the tunnel⁶ and start to have an intuition for all the ingredients to create this soup called "probability theory". Distributions are NOT cumulative density functions and neither they are probability density functions... They are something that transcends these "specialized" concepts and goes to the heart of how we evaluate (how we weigh; how we **measure**) a probability in a certain scenario.

Distributions are probability measures.

Remark

You should remember (LOL) that when we want to specify a measure on a σ -algebra, it's enough to do it on a π -system^a generating that σ algebra: by means of the monotone class theorem we are then able to extend the measure to the σ -algebra.

This means that to specify μ it is enough to specify it on a π -system which generates \mathcal{E} . For example, consider $E = \overline{\mathbb{R}}$, $\mathcal{E} = \mathcal{B}_{\overline{\mathbb{R}}}$. Consider the collection of sets $[-\infty, x]$, $x \in \mathbb{R}$ which is of course a π -system because it is closed under intersection. Moreover, this shit generates the Borel sigma algebra on $\overline{\mathbb{R}}$.

If we want to define a distribution, that is a measure, it is enough to define it on this π -system. Imagine

⁶This is only the first chapter.

that we apply our distribution measure to one set of this π -system

$$c(x)^b = \mu([-\infty, x]) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

by the monotone class theorem. So we have now specified the measure on the π -system. The part $\mathbb{P}(X \leq x)$ reminds us of the undergraduate times^c: it is a distribution function! This is what our professor did implicitly to avoid using measure theory^d.

^aa π -system is a simpler object than a σ -algebra: it is simply a collection of sets closed under intersection

^bbecause it is a function of x

^cI already wanted to kill myself at that time.

^dI have noticed that my life has not benefited in ANY form since I have been introduced to measure theory.

Revise with Kotatsu!

But what is the *monotone class Theorem*? First, we need the definition of *monotone class*:

Definition 1.1.8

A collection of functions \mathcal{M} is called **monotone class** provided that:

1. it includes the constant function 1;
2. taken f and $g \in \mathcal{M}_b$ (with \mathcal{M}_b being the subcollection of bounded functions in \mathcal{M}) and $a, b \in \mathbb{R}$, then $af + bg \in \mathcal{M}$;
3. if the sequence $(f_n)_n$ is contained in \mathcal{M}_+ (with \mathcal{M}_+ being the subcollection consisting of positive functions in \mathcal{M}) and $f_n \nearrow F$ then $f \in \mathcal{M}$.

Theorem 1.1.2

Monotone class Theorem:

Let \mathcal{M} be a monotone class of functions on E . Suppose, for some π -system \mathcal{C} generating \mathcal{E} , that $1_A \in \mathcal{C}$ for every $A \in \mathcal{C}$. Then \mathcal{M} includes all positive \mathcal{E} -measurable functions and all bounded \mathcal{E} -measurable functions.

So, turning back to the previous remark: in that case \mathcal{E} consists of the Borel σ -algebra on the extended real line ($\mathcal{B}_{\overline{\mathbb{R}}}$); our π -system is capable of generating the Borel σ -algebra (because every Borel set can be constructed with the combination $[-\infty, x]$ for all $x \in \mathbb{R}$ ⁷); we defined the measure μ on the π -system $[-\infty, x]$ for all $x \in \mathbb{R}$; the monotone class theorem states that if a class of sets (in this case, the class of sets where μ is well-defined) contains a π -system (✓) and is closed under monotone limits (i.e. is a monotone class), then it contains the σ -algebra generated by the π -system: this means that the class of sets where the distribution μ is well-defined will include the Borel σ -algebra $\mathcal{B}_{\overline{\mathbb{R}}}$. This is kinda cool, I'll have to admit. Unfortunately, I don't really care about this.

1.1.2 Functions of random variables

Consider X , a random variable taking values in (E, \mathcal{E}) and consider further a measurable space (F, \mathcal{F}) . Let $f : E \rightarrow F$ be a measurable function relative to \mathcal{E} and \mathcal{F} ⁸. This function should be measurable by means of \mathbb{P} , otherwise we couldn't do anything useful with it. Consider the composition

$$Y = f \circ X \quad \text{such that } Y(\omega) = f(X(\omega)) = f(X(\omega)), \quad \omega \in \Omega.$$

This composition is a random variable taking values in (F, \mathcal{F}) which comes from the fact that measurable functions of measurable functions are still measurable.

⁷I know, I know: the fuck is a Borel set? A Borel set is every set that can be formed by the countable union or countable intersection or complementation from any open or closed set. You see that every Borel set you can imagine can be constructed by $[-\infty, x]$.

⁸This basically means that this bitch won't do anything evil. The whole point of measure theory, σ algebras and all other shit is to ensure everything behaves.

Definition 1.1.9

Consider two random variables X, Y taking values in (E, \mathcal{E}) and (F, \mathcal{F}) respectively. Consider the pair

$$Z = (X, Y) : \Omega \rightarrow E \times F.$$

Why would we want to call it Z ? It's because, beside being a random vector, it is in turn a random variable:

$$Z(\omega) = (X(\omega), Y(\omega)).$$

Since $E \times F$ is a product space, we should attach it the product σ -algebra. So Z is a random variable taking values in $E \times F$.

Note that the product space $E \times F$ is endowed with the σ -algebra $\mathcal{E} \otimes \mathcal{F}$, that is the product σ -algebra generated by the collection of all possible rectangles between E and F . We frequently have to look to special cases like random vectors that must take values in measurable spaces for them to make sense. This measurable space is naturally generated by the product σ -algebra (but it may be generated by other σ -algebras⁹!).

Definition 1.1.10

We call **joint distribution** of X and Y the distribution of Z .

This is interesting, since we know that this variable has the specific structure of a random vector: we identify the distribution of this vector as the joint distribution of its two coordinates¹⁰.

Remark

The product σ -algebra $\mathcal{E} \otimes \mathcal{F}$ is generated by the π -system of measurable rectangles.

On the product space, it is enough to only specify it on this π -system.

Let denote with π the joint distribution of X, y . It is sufficient to specify

$$\pi(A \times B) = \mathbb{P}(X \in A, Y \in B) \quad \forall A \in \mathcal{E}, B \in \mathcal{F}.$$

We exploited the measurability of X and Y

Definition 1.1.11

Given the joint distribution π , consider sets $A \in \mathcal{E}, B \in \mathcal{F}$. Then we call **marginal distribution of X**

$$\mathbb{P}(X \in A) = \pi(A \times F) \quad \forall A \in \mathcal{E}$$

and we call **marginal distribution of Y**

$$\mathbb{P}(Y \in B) = \pi(E \times B) \quad \forall B \in \mathcal{F}.$$

We call it distribution because it is actually a measure! So we can call it with the notation of measure

$$\mu(A) = \mathbb{P}(X \in A) = \pi(A \times F) \quad \forall A \in \mathcal{E}$$

and

$$\nu(B) = \mathbb{P}(Y \in B) = \pi(E \times B) \quad \forall B \in \mathcal{F}.$$

This actually means that the second coordinate is fixed in being the whole space F . Think about integrating the second coordinate along the real line when doing marginal distributions... this is the same thing here. Now that we have joint and marginal distributions, what is the next step¹¹?

Definition 1.1.12

Let X, Y be random variables taking values in (E, \mathcal{E}) and (F, \mathcal{F}) respectively and let μ and ν be their respective distributions. Then X and Y are said to be **independent** if their joint distribution is the

⁹Repeatedly inflicting painful kicks on my gonads.

¹⁰it eludes me how anyone could find this interesting. We have to think about the whole vector as being distributed like its components separately

¹¹Abandoning myself in the sweet embrace of Death, methinks.

product measure formed by their marginals.

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \forall A \in \mathcal{E}, B \in \mathcal{F}.$$

This also means that

$$\pi = \mu \nu$$

Here the marginals do not interact with each other. This is true for two random variables but we need¹² something more general.

Definition 1.1.13

Let (X_1, X_2, \dots, X_n) be a finite collection of random variables. The collection is said to be an **independency** if the distribution of (X_1, X_2, \dots, X_n) is the product of $\mu_1, \mu_2, \dots, \mu_n$ where μ_i is the distribution of X_i , for $i = 1, \dots, n$.

Cinlar is stupid I wish him dead to be frank for this independency shit. Independency is not even an english word. What the fuck? Anyway, what about infinite collections?

Definition 1.1.14

Let $(X_n)_n$ be an infinite collection of random variables. It is said to be an **independency** if every finite sub-collection of it is an independency.

We now turn to stochastic processes¹³! But first...

1.1.3 Infinite product spaces

Let T be an arbitrary (countable or uncountable) set. We will think about this set as an "index" set. For each $t \in T$ consider the measurable (E_t, \mathcal{E}_t) . So we have a space for each index (plenty of measurable spaces hanging around). Consider a point x_t in E_t for each $t \in T$. The collection¹⁴ $(x_t)_{t \in T}$. If $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ then $(x_t)_{t \in T}$ is actually a function of T taking values on (E, \mathcal{E}) .

The set F of all possible functions $x = (x_t)_{t \in T}$ is called the **product space** $((E_t, \mathcal{E}_t))_{t \in T}$.

This is the natural generalization of what we do when we construct product spaces, albeit with a different notation. Usually F is denoted by $X_{t \in T} E_t$. But we know we also need a σ -algebra...

A **rectangle** in F is a subset of the form

$$\{x \in F : x_t \in A_t \forall t \in T\}$$

Where A_t differs from E_t for only a finite number of t . So I want to consider subsets of F (the space of functions) of the form above. I want only the functions x in F such that each coordinate belongs to A_t , a subset of E_t for each $t \in T$. It seems that we have a restriction on all the coordinates... But this may bring to problems when we have an uncountable number of coordinates and therefore an uncountable number of restrictions. But we can say that if $A_t = E_t$ (the whole space) we don't apply any restriction. So in this case X_t belongs to E_t so we can choose whatever X_t we like. So only a finite number of coordinates are restricted while the other infinite ones are free to vary¹⁵.

The σ -algebra generated by the collection of all measurable rectangles is denoted by

$$\bigotimes_{t \in T} \mathcal{E}_t.$$

This is the product σ -algebra in any infinite-dimensional space. So, the (natural) resulting measurable space in the end will be

$$\bigtimes_{t \in T} E_t, \bigotimes_{t \in T} \mathcal{E}_t.$$

¹²No.

¹³Please no.

¹⁴We could consider it a function of t but that wouldn't be exactly correct since each t has a different measurable space. We may have the same space but it's not true in general... I am thrilled to say the least.



¹⁵→ my honest reaction.

This is not in contrast with what we already know for finite product space, since these already have a finite number of restrictions. So this concept of rectangle, which can be a bit different from the one regarding the famous and well-tested geometrical shape¹⁶, is not restricted on all the coordinates (like the shape¹⁷) but only on a finite number of them.

We also have an alternative notation for this measurable space!

$$\bigotimes_{t \in T} (E_t, \mathcal{E}_t).$$

In the case that $(E_t, \mathcal{E}_t) = (E, \mathcal{E}) \forall t \in T$ the product space is denoted by

$$(E, \mathcal{E})^T$$

or

$$(E^T, \mathcal{E}^T)$$

These are not real powers but it's just notation... Anyway these are all different notations to indicate the infinite product space with the product σ -algebra built upon the π -system which is the collection of all possible rectangle defined in the way we saw above¹⁸.

1.1.4 Stochastic processes

Definition 1.1.15

Let (E, \mathcal{E}) be a measurable space and consider an index set T (as before, an arbitrary set countable or uncountable).

Let also X_t be a random variable taking values in (E, \mathcal{E}) . Then the collection of those random variables $(X_t)_{t \in T}$ is called a **stochastic process** with state space (E, \mathcal{E}) and parameter set T .

Note that there is no mention about time here. Just think about the index set, which indexes the stochastic process. If we interpret T as time then we have the most common interpretation of stochastic processes. But it could also be space (imagine \mathbb{R}^2) or your mom being █. Anyway the most natural interpretation is time.

Now take a $\omega \in \Omega$ and evaluate all these random variables on the same ω . What we get is

$$t \mapsto X_t(\omega)$$

which is a function from T to (E, \mathcal{E}) . So if we see it as a function of t for each ω we get a function which is an element of E^T . So what is a stochastic process, to sum it up? It's just a random variable taking values in the infinite product space E^T . That's why it is a problematic object: it's because mathematicians deserve to experience the sadness and evil they unleashed upon the world. Ever noticed how similar the words "measurable" and "miserable" are? I didn't think so. Yeah technically more structure helps us modeling real phenomena more accurately but who the FUCK cares.

1.1.5 Example of random variables

Consider some examples of simple random variables:

Example 1.1.1

Poisson random variables.

This random variable takes values in \mathbb{N} (it's a one dimensional random variable). We consider the power set of \mathbb{N} ¹⁷. We know that power sets are σ -algebras that we can use (but we could encounter some trouble with uncountable elements, for which we would need smaller σ -algebras¹⁸).

¹⁶Oh thank god someone finally said it. I was starting to get scared.

¹⁷I swear to god.

¹⁸NO I WON'T USE LABELS AND NUMBERED EQUATIONS.

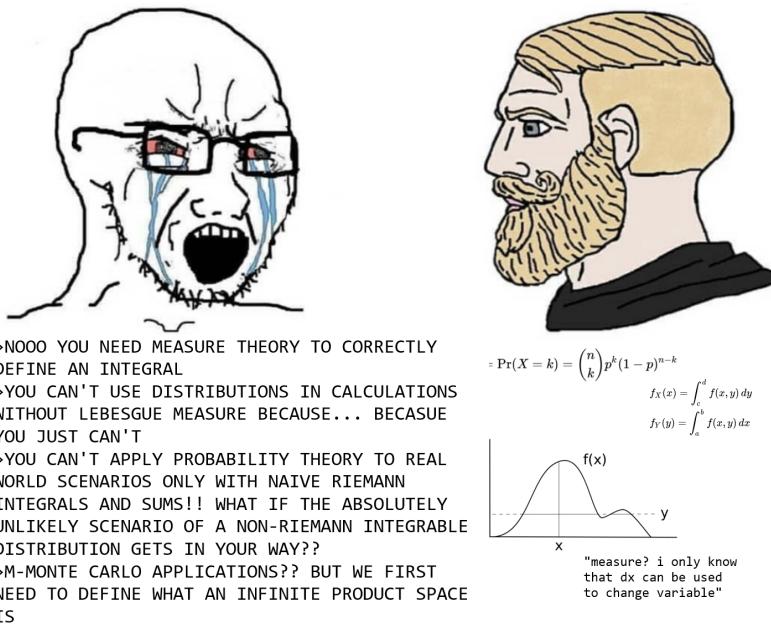


Figure 1.3: I'm sorry but here you're the soyjack and I'm the chad.

What is the distribution of this random variable?

$$\begin{aligned}
 \mu(A) &= \mathbb{P}(X \in A) = \sum_{n \in A} \mathbb{P}(X = n) \quad A \subset \mathbb{N} \\
 \text{with } \mathbb{P}(X = n) &= e^{-c} \frac{c^n}{n!}, \quad n \in \mathbb{N}, \quad c > 0.
 \end{aligned}$$

So imagine we have this kind of random variable. We consider a subset of the natural number and we want to evaluate the measure of this subset that we chose. We know that we define the random variable by defining the distribution. For each n we get a number $e^{-c} \frac{c^n}{n!}$. Another interesting implication is that

$$\sum_{n \in A} \mathbb{P}(X = n) = \sum_{n \in \mathbb{N}} \delta_n(A) \mathbb{P}(X = n)$$

where $\delta_n(A)$ is the **Dirac measure** sitting at n . So, n is a parameter and

$$\delta_n(A) = \begin{cases} 1 & n \in A \\ 0 & n \notin A \end{cases}.$$

The Dirac measure is similar to the indicator function (they behave basically in the same way) but the difference is that this one is a *measure* and the latter is a *function*. The Dirac measure has n as a parameter, while the indicator function has the set as a parameter ($\mathbb{1}_A(n)$).

^aSubset of all subsets of \mathbb{N} .

^bNo one really cares, not even Federico Polito.

Example 1.1.2

Exponential random variable

This random variable is again one-dimensional but this time this random variable is *absolutely continuous*. What does it mean? It actually means that the variable is absolutely continuous with respect to the Lebesgue measure^a. This is evident when we write down the distribution.

Consider a random variable taking values in \mathbb{R}_+ and further consider $\mathcal{B}_{\mathbb{R}_+}$. We have

$$\mu(dx) = \int_{\mathbb{R}_+} ce^{-cx} Leb(dx), \quad c > 0, x \in \mathbb{R}_+.$$

Ok no wait hold your fucking horses, cowboy. Why did we write dx instead of just x ? Also weren't densities, like, a fucking measure of some set in the form of $\mu(A)$? I like the fact that these densities resemble more closely the probability density function I was taught to work with during my sad Economics degree but there are many many things that creep me out.

We have an answer for this, but we need to do a bit of backtracking.

^atacci tua.

Revise with Kotatsu!

First of all: what does "absolutely continuous" even means?

Definition 1.1.16

let μ and ν be measures on a measurable space (E, \mathcal{E}) . Then, measure ν is said to be absolutely continuous with respect to measure μ if, for every set $A \in \mathcal{E}$,

$$\mu(A) = 0 \implies \nu(A) = 0.$$

Huh. That was pretty simple. Well, turns out we can exploit this fact to "switch" between different measures inside of integrals...

Theorem 1.1.3

Radon-Nikodyn Theorem. Suppose that measure μ is σ -finite and measure ν is absolutely continuous with respect to μ . Then there exists a positive \mathcal{E} -measurable function p such that

$$\int_E \nu(dx)f(x) = \int_E \mu(dx)p(x)f(x) \quad f \in \mathcal{E}_+.$$

If we use the alternative notation:

$$\int_E f d\nu = \int_E pf d\mu \quad f \in \mathcal{E}_+.$$

Moreover, p is unique up to equivalence: if the equation above holds for another $\hat{p} \in \mathcal{E}_+$ then $\hat{p}(x) = p(x)$ for μ -almost every^a $x \in \mathcal{E}_+$. This is an if and only if statement!

Also, p is called the **Radon-Nikodyn derivative** of ν with respect to μ :

$$\frac{\nu(dx)}{\mu(dx)} = \frac{d\nu}{d\mu} = p.$$

^aThis means that all the sets where this condition doesn't hold are negligible when weighted with measure μ .

With all this alternative notation this thing honestly feels like trying to understand the Metal Gear Solid plot, where identical characters named Snake keep cloning each other and being triple crossed by everyone until you finally understand that the storyline never made sense in the first place and that Hideo Kojima writes his games like a fucking fanfiction.

Now everything should make more sense¹⁹. If we know that a given random variable (say, the exponential random variable) has a distribution $\mu(dx)$ then we will be able to transform this in a distribution of the form Lebesgue measure $\cdot p(x)$ (we can lose f if f is constant). In this formulation dx stands for the Lebesgue measure and the second part of the equation (ce^{-cx}) is the $p(x)$, called **density function**. We can do this because we can see from the formula that this distribution, or measure, is indeed absolutely continuous with respect to the Lebesgue measure since we can express it in the form stated by the Radon-Nikodyn theorem. So $p(x) = ce^{-cx}$, $x \in \mathbb{R}_+$ is the density relative to μ . This should serve us as a demonstration that if we define the random variable we get the distribution/measure (remember! distributions are measures!) and vice versa.

¹⁹Enviable optimism.

It is interesting²⁰ to see that also discrete random variable turns out to be absolutely continuous... But not with respect to the Lebesgue measure. To exact, discrete random variables are absolutely continuous with respect to the *counting* measure. And here's why to do all this shit we need the Lebesgue integral: by changing the measure we are using to compute the integral, we can use just one object (the probability distribution) to treat both discrete (using a counting measure, which gives us the cumulative distribution function in the form of a sum) and continuous random variables (using the Lebesgue measure, which gives us the cumulative distribution function in the form of a Riemann integral.)

So you know what a Lebesgue measure is, right?

Of course not! Is that a bad thing?

You were adopted

Figure 1.4: Actual conversation happened between me and Professor Lods.

So we're due for a little refresh on what the hell a Lebesgue measure is. I'm sorry²¹ for our mathematician friends but I need this to be written loud and clear. This is from Professor Lods' Lecture notes from the pre-course in Measure Theory, with the hope that one day I'll be skilled like he is with \LaTeX typesetting.

Revise with Kotatsu!

Let's have a quick refresh about the Lebesgue measure over the measurable space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Let $S = \mathbb{R}$. First of all, what is an algebra?

Definition 1.1.17

A collection Σ_0 of subsets of S is called an algebra on S if:

- $S \in \Sigma_0$;
- if $A \in \Sigma_0$ then $A^c \in \Sigma_0$ where $A^c = S \setminus A$ is the complementary of A ;
- if $A, B \in \Sigma_0$ then $A \cup B \in \Sigma_0$.

We also need the concept of pre-measure, which is basically a measure but defined on an algebra (instead of a σ -algebra):

²⁰Debatable claim.

²¹Not really. I mean, I'm sorry for the fact that they *are* mathematicians but that's where my compassion starts and ends.

Definition 1.1.18

Let Σ_0 be an algebra on S (not necessarily a σ -algebra). A mapping $\ell : \Sigma_0 \mapsto [0, \infty]$ is said to be a **pre-measure** on Σ_0 if $\ell(\emptyset) = 0$ and for any pairwise disjoint $\{A_n\}_n \subset \Sigma_0$ with $\bigcup_n A_n \in \Sigma_0$ it holds:

$$\ell\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \ell A_n.$$

Moreover, a pre-measure ℓ is said to be σ -finite on Σ_0 if there exists a sequence $\{A_n\}_n \subset \Sigma_0$ with $\bigcup_n A_n \in \Sigma_0$ and $\ell(A_n) < \infty$ for any $n \in \mathbb{N}$.

We can immediately see that $\bigcup_n A_n \in \Sigma_0$ is an additional assumption: in σ -algebras this assumption is always met. So if Σ_0 is a σ -algebra any measure on Σ_0 is a pre-measure.

We also need one more thing: the **Caratheodory's extension Theorem**.

Theorem 1.1.4

Charatheodory's extension Theorem: Let S be a given set and let Σ_0 be an algebra on S and $\Sigma = \sigma(\Sigma_0)$. If $\ell : \Sigma_0 \mapsto [0, \infty]$ is a pre-measure on (S, Σ_0) then there exists a measure μ on (S, Σ) such that

$$\mu(A) = \ell(A) \quad \forall A \in \Sigma_0.$$

Moreover, if ℓ is a σ -finite pre-measure on Σ_0 , then such a measure μ on (S, Σ) is unique and σ -finite.

Apparently this is one of the principal results in measure theory since it allows to construct measures well-adapted to practical situations: once such measures are constructed, Caratheodory's theorem can go fuck itself off. But the most important question is: why do we care about these total nerds? Because we can now define

$$\mathcal{C}_0 = \{[a, b] : -\infty \leq a \leq b \leq \infty \in \mathbb{R}\}$$

and let

$$\Sigma_0 = \left\{ \bigcup_{j=1}^N I_j : I_j \in \mathcal{C}_0 \ \forall j, I_i \cap I_j = \emptyset \text{ if } i \neq j, N \in \mathbb{N} \right\}$$

We can prove without major difficulty that Σ_0 is an algebra on \mathbb{R} . Let's define a pre-measure on Σ_0 by setting:

- $\ell([a, b]) = b - a$ for any $b \geq a$;
- $\ell((-\infty, b)) = \ell((a, \infty)) = \ell(\mathbb{R}) = +\infty$;
- $\ell\left(\bigcup_{j=1}^N I_j\right) = \sum_{j=1}^N \ell(I_j)$ if $\{I_j\}_{j=1,\dots,N} \subset \mathcal{C}_0$ are pairwise disjoint.

It can be checked that this newly defined measure is σ -finite. Remember that $\sigma(\Sigma_0) = \sigma(\mathcal{C}_0) = \mathcal{B}_{\mathbb{R}}$, which is the Borel σ -algebra. Consider, additionally, the result of the Charatheodory's extension Theorem. By stitching all of these amenities together we get:

Theorem 1.1.5

There exists a unique measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that we denote λ (or m) and such that

$$\lambda([a, b]) = b - a \quad \forall a < b.$$

We call this measure the **Lebesgue measure** on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

So we just learnt how to FIND A
FUCKING INTERVAL ON THE
REAL LINE?

Look, I know that can be seen as useless formalism. And indeed it is! We just formalized the notion of *length*. If we take it to 2-dimensional spaces we end up with the notion of area, in 3-dimensional spaces is the notion of volume... and so on.

Huh. This makes sense. So this is the notion of volume when everything, including the real line, is a set. Kinda seems like the solution to a problem we ourselves created...



Remark

We can define in the same way the Lebesgue measure on (I, \mathcal{B}_I) for all $I \subset \mathbb{R}$.

Remark

The measure space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$ is σ -finite since $([-n, n])_n \nearrow \mathbb{R}$ but is not finite since

$$\lambda(\mathbb{R}) = \lim_n \lambda([-n, n]) = \lim_n 2n = \infty$$

Yeah, that mysterious measure I was talking about before to connect Riemann and Lebesgue integration was the Lebesgue measure. We often just write dx to express the Lebesgue measure (which is what we did on example 1.2 about the exponential random variable), but the meaning is always the same, with a striking similarity to the concept of Riemann integration: take the Lebesgue measure of a smaller and smaller element of our $\mathcal{B}_{\mathbb{R}}$ set, use it to "weight" (read: multiply) the value of the function for that element and then sum it all up together. What we end up with is basically a serie of simple functions that slice "horizontally" the co-domain of the function. Keep reducing the size and you end up with the Lebesgue integral. Of course, when the measure is the Lebesgue measure, the Riemann and the Lebesgue integral have a really similar interpretation.

Back to our topic: the exponential random variable is absolutely continuous with respect to the Lebesgue measure. It is interesting to see that also discrete random variables turn out to be absolutely continuous: the difference is that they are not absolutely continuous to the Lebesgue measure, but the *counting* measure. At the undergrad level we are used to say that a random variable is either discrete or absolutely continuous, buy this was ultimately a lie²².

²²Measure theory turns truths into lies. Truly a demonic machinery.

Example 1.1.3

Gamma distribution^a: Consider a random variable taking values in \mathbb{R}_+ and consider as a σ -algebra the Borel σ -algebra $\mathcal{B}_{\mathbb{R}_+}$. The distribution of the Gamma random variable is the following:

$$\mu(dx) = dx \frac{c^a x^{a-1} e^{-cx}}{\Gamma(a)}, \quad \begin{array}{l} a > 0, \\ \text{with } c > 0, \\ x \in \mathbb{R}_+ \end{array} .$$

Here $\Gamma(a)$ is the *Gamma function*:

$$\Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx.$$

The Gamma function is one of the most famous special functions that comes up almost everywhere. This definition of Gamma function is valid just for positive values and can be seen as a *Laplace transform^b* or as a *Mellin transform^c*. The first parameter a is called *shape parameter*; the parameter c is called *scale parameter*. This distribution is also continuous with respect to the Lebesgue measure. We have some special cases of the Gamma distribution but Federico Polito doesn't really care about. Just know that the χ^2 distribution is a special case of the Gamma random variable.

^aShe factorials on my γ 'till I β .

^b $\mathcal{L}\{f\}(s) = \int_0^{+\infty} f(t) e^{st} dt$ where s is a complex number $s = a + ib$.

^c $m[f; s] \equiv F(s) = \int_0^{+\infty} f(t) t^{s-1} dt$ where $s = a + ib$.

Example 1.1.4

This is a certified hood classic: **Gaussian distribution**.

Consider a random variable taking values in \mathbb{R} . Of course we consider $\mathcal{B}_{\mathbb{R}}$ and the distribution is (notice how also this one is absolutely continuous with respect to the Lebesgue measure):

$$\mu(dx) = dx \cdot \underbrace{\frac{1}{\sqrt{2\pi b}} e^{-\frac{(x-a)^2}{2b}}}_{p(x)}, \quad \begin{array}{l} a \in \mathbb{R}, \\ b > 0 \in \mathbb{R}, \\ x \in \mathbb{R}_+ \end{array} .$$

Of course, a is called the *mean* of the distribution and b is called the *variance*.

Example 1.1.5

This is a random variable that stems from two independent random variables having Gamma distribution. Consider γ_a (distribution of a Gamma random variable with parameters a and $c = 1$) and γ_b (distribution of a Gamma random variable with parameters b and $c = 1$). So, two gammas with different shape parameter.

Let $X \sim \gamma_a$ and $Y \sim \gamma_b$. Moreover, let them be independent. This is a random vector (X, Y) with two components... What is its distribution?

$$\pi(dx, dy) = \underbrace{\gamma_a(dx) \cdot \gamma_b(dy)}_{\text{because of independency}} = dx dy \frac{e^{-x} x^{a-1}}{\Gamma(a)} \cdot \frac{e^{-y} y^{b-1}}{\Gamma(b)}.$$

So it's easy to build joint distributions when the random variables are independent^a.

^aWell no shit. Even I can multiply two numbers

Example 1.1.6

Gaussian random variable with exponential variance.

Here the variance is random and is distributed exponentially^a. Consider a random variable X taking values in \mathbb{R}_+ and a random variable Y taking values in \mathbb{R} .

Here we are again in presence of a random vector. The distribution is the following:

$$\pi(dx, dy) = dx \, dy \cdot ce^{-cx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} \quad x \in \mathbb{R}_+, \quad y \in \mathbb{R}.$$

Remark

π in this case has a special form: it has the form

$$\pi(dx, dy) = \mu(dx)K(x, dy).$$

In particular, here $\mu(dx)$ is

$$dx \, dy \cdot ce^{-cx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}},$$

which is a docile exponential function, and $K(x, dy)$ is

$$dx \, dy \cdot ce^{-cx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}.$$

In this case $K(x, dy)$ cannot be the distribution of Y , because it has some x in it. So this distribution is not simply the product of marginal distribution. But let's take a closer look to the form $\mu(dx)K(x, dy)$. $\mu(dx)$ is certainly a measure, but what about $K(x, dy)$?

^abecause humans should never have the hubris to meddle with the horrific world of random necessities that the Gods have laid before us.

1.2 Transition kernels

Turns out that $K(x, dy)$, depending on x , is connected to the other measure $\mu(dx)$. The object $K(x, dy)$ is called **transition kernel**²³ and it's very important. Not that here $K(x, dy)$ has the form

$$B \mapsto K(x, B)$$

It should be seen as a set function. Why? Just look back to $\pi(dx, dy)$. It is in differential form, but if we integrate the joint distribution against one set in the product space we get π evaluated on that subset of the product function (it is a measure... measures are always set functions). So also $\mu(dx)K(x, dy)$ should be a set function. $\mu(dx)$ is of course a set function (it is a measure...²⁴) and so should be $K(x, dy)$. The presence of x in $K(x, dy)$ is what "links" the X coordinate with the Y coordinate of the random vector. The Gamma example was about two independent random variables: here it is impossible to obtain the product of measures. Transition kernels are incredibly important because they are ultimately connected with the *structure of dependency* between random variables: that's why this course is going to bust our ball into the oblivion about them²⁵.

²³Kernel? Colonel? I thought we were over with the Metal Gear Solid jokes.

²⁴All this passive-aggressiveness for what?

²⁵Professor Polito jokes that every year people complain about the abstractness of kernels and laughs about it. I'm happy that his sense of humor has been left untouched by my slightly scathing EDUMETER review of this course.



Figure 1.5: Yeah, no more Metal Gear Solid jokes after this one. I promise²⁶.

Remember in the undergraduate courses: when there was dependence we usually expressed it with the *conditional probability*.

Transition kernels let us manage random vectors that are not trivial and that have a dependence relationship with other random vectors.

Think about the previous example: the marginal distribution ν of Y has the form

$$\begin{aligned}\nu(B) &= \pi(\mathbb{R}_+ \times B) = \int_{\mathbb{R}_+} \mu(dx) K(x, B), \quad B \in \mathcal{B}_{\overline{\mathbb{R}}} \\ &= \int_{\mathbb{R}_+ \times B} \pi(dx, dy) = \int_{\mathbb{R}_+ \times B} dx dy \cdot ce^{-cx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} \\ &= \int_B dy n(y), \quad \text{where } n(y) = \int_0^\infty dx \cdot ce^{-cx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}}.\end{aligned}$$

So we have that the marginal distribution $\nu(B)$ is written as $\int_B dy n(y)$ and is therefore absolutely continuous with respect to the Lebesgue measure. We could actually solve this integral ($n(y)$ is a closed form called *two-sided exponent*). So we now have the marginal of Y and the marginal of X and we immediately realize that if we multiply the two densities we do not obtain the joint density (because they are dependent²⁷).

So we are now ready to define the concept of transition kernel.

Definition 1.2.1

Let (E, \mathcal{E}) and (F, \mathcal{F}) be measurable spaces. Let K be a mapping from $E \times F$ into $\overline{\mathbb{R}}_+$. Then K is called **transition kernel** from space (E, \mathcal{E}) into space (F, \mathcal{F}) if:

- the mapping $x \mapsto K(x, B)$ is \mathcal{E} -measurable $\forall B \in \mathcal{F}$;
- the mapping $B \mapsto K(x, B)$ (the second mapping of the kernel, the one regarding the set) is a measure $\forall x \in E$.

We can consider the transition kernel a hybrid object: if we look at it with respect to the first variable it is a *measurable function*, if we look at it with respect to the second variable it is a *measure*.

Example 1.2.1

Take ν , a finite measure on (F, \mathcal{F}) and take k , a positive function on $(E \times F)$ which is measurable with respect to $\mathcal{E} \otimes \mathcal{F}$, the product σ -algebra. Then, we integrate

$$\int_B \nu(dy) k(x, y) \quad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

We see how this object depends on x and on the choice of B (a function of x and B ...). It defines a

²⁶This is not, in fact, the last Metal Gear Solid joke.

²⁷OK I GET IT.

transition kernel

$$K(x, B) = \int_B \nu(dy) k(x, y) \quad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

from (E, \mathcal{E}) into $(F, |\mathcal{F}|)$.

Theorem 1.2.1

This theorem tells us what we can do with a kernel. Let K be a transition kernel from (E, \mathcal{E}) into (F, \mathcal{F}) . Then,

- ① we have

$$\int_F K(k, dy) f(y) \quad \text{with } x \in E, f \in \mathcal{F}_+.$$

This operation defines a function $Kf \in \mathcal{E}_+ \forall f \in \mathcal{F}_+$.

Note the notation!

The notation $f \in \mathcal{F}_+$ in Çinlar is either the σ -algebra \mathcal{F} or the set of functions measurable with respect to the σ -algebra \mathcal{F} . Another one of the many ways Çinlar chooses to sadden my day. But my revenge is on the way.

First of all we integrate the kernel with respect to the second variable (which basically means we use it as a measure on F). The integrand is F and since this function is a measure with respect to the second variable I can integrate \mathcal{F}_+ -measurable functions. Remember that $\mu f = \mu(f) = \int_E f(x) \mu(dx) = \int_E f d\mu$.

The integration over F "takes out" one part of the kernel from the equation (the measure part) so that we can write:

$$Kf(x) = \int_F K(k, dy) f(y);$$

- ② we want to use the kernel with respect to the first variable (obtaining a \mathcal{E} -measurable function), so we integrate

$$\int_E \mu(dx) K(x, B)$$

Remember that we must integrate the kernel with respect to a measure μ that is attached to the space (E, \mathcal{E}) . This operation (since we remove the "function" part of the kernel) defines a measure μK on (F, \mathcal{F}) for each measure μ on (E, \mathcal{E}) and we can write:

$$\mu K(B) = \int_E \mu(dx) K(x, B);$$

- ③ now we want to integrate everything: we consider the measure μK , take f and calculate its integral with respect to the measure μK

$$(\mu K)f.$$

With $f \in \mathcal{F}_+$. We can now link the function obtained in step 1 and the measure obtained in step 2:

$$(\mu K)f = \mu(Kf) = \int_E \mu(dx) \cdot \int_F K(x, dy) f(y)$$

for every choice of measure μ on (E, \mathcal{E}) and for every choice of $f \in \mathcal{F}_+$

Remember that here $(\mu K)f$ is shortened notation: μf means $\int_E f(x) \mu(dx)$. We are NOT applying the measure to the function!

Right... but if we choose $f \in \mathcal{E}$ (f is \mathcal{E} -measurable) as the indicator function $f = \mathbb{1}_B$, $B \in \mathcal{E}$ (the most simple example of \mathcal{E} -measurable function), this happens:

$$\begin{aligned} \mu \mathbb{1}_B &= \int_E \mathbb{1}_B(x) \mu(dx) \\ &= \int_B \mu(dx) = \mu(B) \end{aligned}$$

This is the reason behind this kind of notation that switches between measures and integrals. This is interesting and useful, if "interesting and useful" was slang for "boring and useless". So technically yeah, we *are* applying the measure to the function in the sense that we are weighing the function with the measure μ .

Proof

- First of all, remember that Kf resulting from the integral is a well defined function but that is not enough: we need a \mathcal{E} -measurable function of Kf . We need to proceed in a constructive way: we have to think about \mathcal{E} -measurable functions, in whose space there are different type of functions: indicators, simple function, positive functions, positive or negative function. We start from simple function and then extend this result to positive functions and then to a broader class.

First consider f to be a simple function with its canonical representation:

$$f = \sum_{i=1}^n b_i \mathbb{1}_{B_i}$$

for given weights b_i and given sets B_i . For this function we consider

$$Kf(x) = \sum_{i=1}^n b_i \underbrace{K(x, B_i)}_{\substack{\mathcal{E}\text{-measurable} \\ \text{with respect to } x}}.$$

So it is a linear combination of \mathcal{E} -measurable functions and therefore $Kf \in \mathcal{E}_+$ when f is simple.

We now want to extend the proof to the subclass of positive measurable functions. Take f positive. We know that we can approximate positive functions by means of simple functions, but reducing the "step" of the simple functions (discretizing the original function) by means of an auxiliary function called **dyadic function** (see below), thus producing a sequence of simple functions. So we have $f \in \mathcal{F}_+$ and $f_n = \vartheta_n \circ f$.

Lemma 1.2.1

Each ϑ_n is an increasing right-continuous simple function on $\overline{\mathbb{R}}_+$ and $\vartheta_n(r)$ increases to $r \forall r \in \overline{\mathbb{R}}_+$ as $n \rightarrow \infty$.

So as soon as n goes to infinity we get the original function. Moreover, remember the theorem that gives measurability of a function:

Revise with Kotatsu!

Theorem 1.2.2

A positive function on (E, \mathcal{E}) is \mathcal{E} -measurable if and only if it is the limit of an increasing sequence of positive simple functions.

Theorem 1.2.3

Let $(f_n)_n$ be a sequence of positive measurable functions in (S, Σ) with respect to a measure Σ such that $f_n \nearrow f$ almost everywhere on the set S , Then

$$\lim_n \int_S f_n d\mu = \int_S f d\mu$$

where the sequence $(\int_S f_n d\mu)_n$ is a nondecreasing sequence.

We consider, as we said before, the discretization

$$f_n = \vartheta_n \circ f.$$

What happens to $Kf(x)$? In this case, it is defined as:

$$Kf(x) = \lim_{n \rightarrow \infty} Kf_n(x) \quad \forall x$$

and this is true for the monotone convergence theorem for the measure $B \mapsto K(x, B)$. Then Kf is \mathcal{E} -measurable being the limit of the \mathcal{E} -measurable sequence of functions $(Kf_n)_n$.

2–3 Fix a measure μ on (E, \mathcal{E}) and define a functional

$$L : \mathcal{F}_+ \mapsto \overline{\mathbb{R}}_+$$

by setting

$$L(f) = \mu(Kf)$$

that is, we integrate Kf with respect to the measure μ . We note that

- i.) if $f = 0$ then $L(f) = 0$;
- ii.) if $f, g \in \mathcal{F}_+$ with $a, b \in \mathbb{R}_+$ then

$$\begin{aligned} L(af + bg) &= \mu(K(af + bg)) \\ &= a\mu(Kf) + b\mu(Kg) \\ &= aL(f) + bL(g). \end{aligned}$$

So the functional is a linear function.

iii.) if $(f_n)_n \subset \mathcal{F}_+$ and $f_n \nearrow f$ then $Kf_n \nearrow Kf$ and this is true by monotone convergence theorem with respect to integrals with respect to the measure $B \mapsto K(x, B)$.

iv.) note that

$$L(f_n) = \mu(Kf_n) \nearrow \mu(Kf)$$

again because of the monotone convergence theorem with respect to the measure μ .

So given that i.), ii.), iii.) and iv.) hold (and recurring to the theorem 4.21 of the Cinlar book^a) we have that there exists a measure ν on (F, \mathcal{F}) such that

$$L(f) = \nu f \quad \forall f \in \mathcal{F}_+,$$

Note that if we specifically take the function $f = \mathbb{1}_B$, $B \in \mathcal{F}$ we see that

$$\begin{aligned} \nu(B) &= \nu \mathbb{1}_B = L(\mathbb{1}_B) = \mu(K \mathbb{1}_B) = \mu \left(\int_B K(x, dy) \right) \\ &= \mu K(x, B) = \int_E K(x, B) \mu dx = \mu K(B). \end{aligned}$$

Then $\nu \equiv \mu K$, that is μK is a measure on (F, \mathcal{F}) and

$$(\mu K)f = \nu f = L(f) = \mu(Kf).$$

□

^ayeah not gonna check that.

Definition 1.2.2

The **dyadic function** is defined as:

$$\vartheta_n(r) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \mathbb{1}_{[\frac{k-1}{2^n}, \frac{k}{2^n})}(r) + n \mathbb{1}_{[n, +\infty)}(r), \quad r \in \overline{\mathbb{R}}_+$$

So we basically have two different sections in this function: after n the value of this function is equal to n , otherwise it is a step function in the sequence of interval from the start to n . To see the shape of this function we could see some examples.

Example 1.2.2

Take $n = 1$. We can calculate this dyadic function obtaining

$$\begin{aligned}\vartheta_1(r) &= 0, & r \in \left[0, \frac{1}{2}\right) \\ \vartheta_1(r) &= \frac{1}{2}, & r \in \left[\frac{1}{2}, 1\right) \\ \vartheta_1(r) &= 1, & r \geq 1.\end{aligned}$$

The function is right-continuous and a step function.

Example 1.2.3

$n = 1$. Try to do it by hand.

$$\begin{aligned}\vartheta_2(r) &= 0, & r \in \left[0, \frac{1}{4}\right) \\ \vartheta_2(r) &= \frac{1}{4}, & r \in \left[\frac{1}{4}, \frac{1}{2}\right) \\ \vartheta_2(r) &= \frac{1}{2}, & r \in \left[\frac{1}{2}, \frac{3}{4}\right) \\ \vartheta_2(r) &= \frac{3}{4}, & r \in \left[\frac{3}{4}, 1\right) \\ \vartheta_2(r) &= 1, & r \in \left[1, \frac{5}{4}\right) \\ \vartheta_2(r) &= \frac{5}{4}, & r \in \left[\frac{5}{4}, \frac{3}{2}\right) \\ \vartheta_2(r) &= \frac{3}{2}, & r \in \left[\frac{3}{2}, \frac{7}{4}\right) \\ \vartheta_2(r) &= \frac{7}{4}, & r \in \left[\frac{7}{4}, 2\right) \\ \vartheta_2(r) &= 2, & r \geq 2.\end{aligned}$$

1.2.1 Products of kernels

Let K be a kernel from (E, \mathcal{E}) into (F, \mathcal{F}) and let L be a kernel from (F, \mathcal{F}) into (G, \mathcal{G}) .

Definition 1.2.3

The product of K and L is the transition kernel from (E, \mathcal{E}) into (G, \mathcal{G}) such that $(KL)f = K(Lf)$ for $f \in \mathcal{G}_+$.

Let's now fill in some information for transition kernels.

Definition 1.2.4

A transition kernel from (E, \mathcal{E}) into (E, \mathcal{E}) is actually called **transition kernel on (E, \mathcal{E})** .

This is actually the most common transition kernel in probability²⁸, since random variables take values in the same space.

²⁸Oh right we are studying probability theory. Thanks for reminding!

Definition 1.2.5

A transition kernel on (E, \mathcal{E}) is called **Markov kernel** if

$$K(x, E) = 1 \quad \forall x \in E.$$

It is called **sub-Markov** if

$$K(x, E) \leq 1 \quad \forall x \in E.$$

So here we are, we summoned Markov for the first time in this course.



Figure 1.6: Andrej Andreevič Markov if he was cool.

Definition 1.2.6

Given a transition kernel on (E, \mathcal{E}) its **powers** are define recursively as follows:

$$\begin{aligned} K^0 &= I \\ K^1 &= K \\ &\vdots \\ K^n &= K \cdot K^{n-1} \end{aligned}$$

I is the identity kernel, i.e. $I(x, A) = \delta_x(A) = \mathbb{1}_A(x) \quad \forall x \in E, A \in \mathcal{E}$

Remark

$$If = f; \mu I = \mu; \mu If = \mu F; IK = KI = K$$

Definition 1.2.7

A transition Kernel K from (E, \mathcal{E}) into (F, \mathcal{F}) is said to be:

- **finite** if $K(x, F) < \infty$ for $\forall x \in E$;
- **bounded** if $x \mapsto K(x, F)$ is bounded;
- **σ -finite** if $B \mapsto K(x, B)$ is σ -finite for $\forall x \in E$;
- **σ -bounded** if it exists a measurable partition $(F_n)_n$ of F such that $x \mapsto K(x, F_n)$ is bounded for $\forall n$;
- **Σ -finite** if $K = \sum_{n=1}^{\infty} K_n$ for some sequence of finite kernels $(K_n)_n$.
- **Σ -bounded** if the K_n can be chose to be bounded.

Definition 1.2.8

If $K(x, F) = 1 \forall x \in E$ then the kernel K is said to be a **transition probability kernel**.

We now turn to extending measure to product spaces with respect to kernels²⁹. In order to formally solve this problem we need the following proposition.

Proposition 1.2.1

Let K be a Σ -finite kernel (the most general property we can think of) from (E, \mathcal{E}) into (F, \mathcal{F}) . We consider measurable functions with respect to the product space: for every positive function $f \in \mathcal{E} \otimes \mathcal{F}$ we have that:

$$Tf(x) = \int_F K(x, dy)f(x, y) \quad x \in E$$

And this object defines a function $Tf \in \mathcal{E}_+$. This is a similar operation to the previous theorem. Moreover the transformation $T : (\mathcal{E} \otimes \mathcal{F})_+ \mapsto \mathcal{E}_+$ is linear and continuous under increasing limits, that is:

- a) if we take $f, g \in (\mathcal{E} \otimes \mathcal{F})_+$ and $a, b \in \mathbb{R}_+$ we have

$$T(af + bf) = aTf + bTg;$$

- b) $Tf_n \nearrow Tf$ for \forall sequence $(f_n)_n \subset (\mathcal{E} \otimes \mathcal{F})_+$ with $f_n \nearrow f$.

So, similarly to the previous theorem we have constructed this operator Tf which operates on the function set $(\mathcal{E} \otimes \mathcal{F})_+$ giving us a positive \mathcal{E} -measurable function. So we can start from this function to build a method to construct measures on the product space $(\mathcal{E} \otimes \mathcal{F})_+$ with its related σ -algebra.

Theorem 1.2.4

Extension of measures on product spaces.

Let μ be a measure on the measurable space (E, \mathcal{E}) . Let K be a Σ -finite³⁰ transition kernel from space (E, \mathcal{E}) into (F, \mathcal{F}) . Then:

- ① if we take our function $f(x, y)$, integrate it against our kernel $K(x, dx)$ over F and then integrate again against measure μ over E , the operation

$$\pi f = \int_E \mu(dx) \int_F K(x, dy)f(x, y)$$

defines a measure π on $(E \times F, \mathcal{E} \otimes \mathcal{F})$;

- ② if μ is σ -finite and K is σ -bounded then π is σ -finite and it is the unique measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ satisfying

$$\pi(A \times B) = \int_A \mu(dx)K(x, B) \quad \forall A \in \mathcal{E}, B \in \mathcal{F}.$$

²⁹Erm... what the sigma?

So by means of kernels we are able to define measures on product spaces in this way³⁰.

114
BAND **EVERYTHING HAPPENS TO ME**
-MATT DENNIS/TOM ADAIR

³⁰I crave to be released from this prison of flesh.

Remark

When the kernel that we used to extend the measure has the form

$$K(x, B) = v(B)$$

for some Σ -finite measure v on (F, \mathcal{F}) , which means that it only depends on B and is therefore a measure, then we obtain the **product measure**

$$\pi = \mu v.$$

But what should we do when we have more than 2 spaces³¹? On finite product spaces we introduce in the same manner the product measure

$$\pi = \mu_1 \mu_2 \cdots \mu_n$$

where μ_i is Σ -finite on (E_i, \mathcal{E}_i) for $\forall i = 1, \dots, n$. So we can induce the presence of another measure from the product measure using the kernels, without assuming independence in the construction.

Example 1.2.4

Here we tackle $n = 3$. Take μ_1 on (E_1, \mathcal{E}_1) , the transition kernel K_2 from (E_1, \mathcal{E}_1) into (E_2, \mathcal{E}_2) and the transition kernel K_3 from $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ into (E_3, \mathcal{E}_3) . Not how we only defined the first measure and then only the transition kernels. We will use these fucking kernels to "move" from measure to measure and from space to space. We have

$$\pi f = \int_{E_1} \mu(dx_1) \int_{E_2} K_2(x, dx_2) \int_{E_3} K((x_1, x_2), dx_3) f(x_1, x_2, x_3)$$

with f positive and $(\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3)$ -measurable. So we defined a measure π which is different from the product measure we recalled earlier (actually that product measure is a special case of this measure) and π is a measure on the 3-dimensional product space. Writing π in differential form:

$$\begin{aligned}\pi(dx_1, dx_2, dx_3) &= \mu(dx_1) K_2(x, dx_2) K((x_1, x_2), dx_3) \\ &= \mu_1 K_2 K_3.\end{aligned}$$

This is for finite product spaces but we can also extend this to infinite product spaces³². What should we expect now?

1.3 Expectation

Well that was a cheap joke. We already know the meaning of expectation from our undergraduate courses... but here we will rock our world and learn some new interpretations. Let's start with measure theory³³. In probability the concept of expectation is strictly tied with the concept of integral: we could say they are almost the same thing.

Definition 1.3.1

Let X be a real valued ($\overline{\mathbb{R}}$) random variable on the probability space $(\Omega, \mathcal{H}, \mathbb{P})$. The **expectation** or **expected value** of X is

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

³¹I have an idea.

³²Why. Stop.

³³We have a great time ahead, I see.

Note the notation!

$$\mathbb{E}X = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) = \int_{\Omega} x d\mathbb{P}.$$

There is always the compact notation we use for integrals

$$\mathbb{E}X.$$

We are integrating the function over the space where it is defined, but since we are using Lebesgue integration we need to integrate with respect to a measure... which is the probability measure $\mathbb{P}(d\omega)$. This is a little bit different from the classic definition, but this is more correct.

Remark

The expectation of X exists if and only if the related integral exists. So the existence of the expectation is the existence of the integral.



Figure 1.7: And then he turned himself into a pickle, funniest shit I've ever seen.

Consider the random variable X , with its positive part X^+ and its negative part X^- . Moreover, remember that

$$X = X^+ - X^-$$

where both the positive part and the negative part are positive functions (remember?³⁴). If we apply the expectation to X , by the linearity of the integral operator we have that:

$$\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-.$$

Where, in particular:

$$\mathbb{E}X^+ = \int_{\Omega} X^+ \mathbb{P}(d\omega) = \begin{cases} < +\infty \\ +\infty \end{cases}$$

and

$$\mathbb{E}X^- = \int_{\Omega} X^- \mathbb{P}(d\omega) = \begin{cases} < +\infty \\ +\infty \end{cases}$$

The problem arises when both the expectation of the positive part and the expectation of the negative part are simultaneously infinite.

³⁴No. Happy?

Further, we say that $\mathbb{E}X$ exists finite when at the same time both expectations are finite:

$$\mathbb{E}X^+ < +\infty, \quad X^- < +\infty.$$

In this case

$$\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^- < \infty.$$

Finally, remember from undergraduate courses:

$$\mathbb{E}X^+ \mathbb{E}X^- = \mathbb{E}|X|$$

So if we check that the expectation of the absolute value is finite this will imply that the expectation of X is finite in both the negative and the positive part. A random variable with finite expectations is said to be **integrable**.

Remark

This is connected to the definition of expectation of a random variable that we already know. Consider the "change of variable" formula for Lebesgue integrals (see Çinlar, formula 5.3 page 30): consider $f \in \mathcal{E}$ and $h : (F, \mathcal{F}) \mapsto (E, \mathcal{E})$. We integrate

$$\int_F \nu(dx)f(h(x)) = \int_E \mu(dy)f(y)$$

where μ is a measure on (E, \mathcal{E}) and is the *image measure* of ν through the function h . For us:

- $h \equiv X$;
- $(F, \mathcal{F}) \equiv (\Omega, \mathcal{H})$;
- $\nu = \mathbb{P}$;
- μ is the distribution of X ;
- $(E, \mathcal{E}) \equiv (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

So the formula becomes

$$\int_{\Omega} \mathbb{P}(d\omega)f(X(\omega)) = \int_{\mathbb{R}} \mu(dx)f(x)$$

with f Borel-measurable. Now the last step is to choose a specific function f , for which we choose the *identity function* so that

$$\int_{\Omega} \mathbb{P}(d\omega)X(\omega) = \int_{\mathbb{R}} \mu(dx)x.$$

Here's the formula that we used to calculate the expectation in the undergraduate years! To be honest, that's what you'll ever really use in the sad case of you working in this field. If the distribution is absolutely continuous with respect to the lebesgue measure we get

$$\int_{\mathbb{R}} \mu(dx)x = \int_{\mathbb{R}} f_X(x) dx \cdot x$$

otherwise, if it is absolutely continuous with respect to a counting measure we get

$$\int_{\mathbb{R}} \mu(dx)x = \sum_{i=1}^{\infty} \mathbb{P}(X = x_i) \cdot x_i.$$

Note the notation!

Forget riemann integrals and sums, fucker, from now on you must learn to use

$$\int_{\mathbb{R}} \mu(dx)x.$$

1.3.1 Properties of expectation

- **Positivity:**

$$X \geq 0 \implies \mathbb{E}X \geq 0.$$

Remember that we are talking about random variables, so when we say $X \geq 0$ we actually mean " $X \geq 0$ almost surely with respect to \mathbb{P} ".³⁵

- **Monotonicity:**

$$X \geq Y \geq 0 \implies \mathbb{E}X \geq \mathbb{E}Y.$$

- **Linearity:**

$$X, Y \geq 0 \implies \mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

- **Insensitivity:**

$$X, Y \geq 0, X = Y \text{ almost surely} \implies \mathbb{E}X = \mathbb{E}Y.$$

- **Monotone convergence:** if we have, for $X_n \geq -0$,

$$(X_n)_n \nearrow X \implies \mathbb{E}X_n \nearrow \mathbb{E}X.$$

- **Fatou's Lemma:** for $X_n \geq 0$ we have

$$(X_n)_n \geq 0 \implies \mathbb{E}\liminf X_n \leq \liminf \mathbb{E}X_n.$$

- **Dominated convergence:** if $(X_n)_n$ is a sequence of random variables such that $\forall n |X_n| \leq Y$ and Y has finite expectation (it is integrable) and $\lim_{n \rightarrow \infty} X_n$ exists, then

$$\mathbb{E}\lim_n X_n = \lim_n \mathbb{E}X_n.$$

- **Bounded convergence:** if we have a sequence of random variables $(X_n)_n$ such that $|X_n| \leq b < \infty$ and $\lim_{n \rightarrow \infty} X_n$ exists, then

$$\mathbb{E}\lim_n X_n = \lim_n \mathbb{E}X_n.$$

Remark

X is positive (or non negative) and $\mathbb{E}X = 0$ if and only if $X = 0$ almost surely.

Remark

If we restrict $\mathbb{E}X$ and $\mathbb{E}Y$ on the subset H we get:

$$\text{if } \mathbb{E}X\mathbf{1}_H \geq \mathbb{E}Y\mathbf{1}_H \quad \forall H \implies X \geq Y \text{ a.s.}$$

Theorem 1.3.1

Let X be a random variable taking values in (E, \mathcal{E}) and be measurable relative to \mathcal{H} and \mathcal{E} . If μ is the distribution of X then

$$\mathbb{E}f \circ X = \mu f, \quad \forall f \in \mathcal{E}_+ \tag{★}$$

Conversely, if ★ holds for some measure μ on (E, \mathcal{E}) and $\forall f \in \mathcal{E}_+$, then μ is the distribution of X .

Proof

The proof should be simple. The first statement is basically the **change of variable** formula, rephrasing the theorem on integration with respect to image measures. The second converse statement requires more thought. If ★ holds $\forall f \in \mathcal{E}_+$, then it holds also for $f = \mathbf{1}_A$ for $A \in \mathcal{E}$ and we have that if we want to calculate the measure

$$\mu(A) = \mu \mathbf{1}_A = \mathbb{E} \mathbf{1}_A \circ X = \mathbb{P}(X \in A).$$

So we have identified this measure with the distribution of the random variable and hence μ is the distribution of X . \square

³⁵Because reality crumbles in front of the chaotic horror of randomness. A similar fate is reserved to my balls.

Example 1.3.1

- ① The **variance** of the random variable X is

$$\text{Var } X = \mathbb{E}(X - \mathbb{E}X)^2.$$

- ② Consider the expectation of an exponential transform of X :

$$\tilde{\mu}_r = \mathbb{E}e^{-rX} = \int_{\mathbb{R}_+} e^{-rx} \mu(dx), \quad r \in \mathbb{R}_+, x > 0.$$

This is known as the **Laplace transform** of distribution μ . It is connected to the moment generating function of a distribution.

- ③ We could be interested in the expectation of another exponential transform:

$$\hat{\mu}_r = \mathbb{E}e^{irX} = \int_{\mathbb{R}} \mu(dx) e^{irx}$$

and, exploiting the representation of complex numbers,

$$\hat{\mu}_r = \int_{\mathbb{R}} \mu(dx) \cos(rx) + i \int_{\mathbb{R}} \mu(dx) \sin(rx)$$

and this is called the **characteristic function** of μ .

- ④ Consider X , a random variable taking values in $\overline{\mathbb{N}}$. Consider

$$\mathbb{E}z^X = \sum_{n=1}^{\infty} z^n \cdot \mathbb{P}(X = n), \quad z \in [0, 1].$$

This is called **probability generating function of X** .

These are all special kinds of expectations.

Remark

The expectation $\mathbb{E}X$ is in some sense “optimal”: what the fuck? It is “optimal” because it is our best estimate of X . Imagine we are given the following integral:

$$f(a) = \int_{\Omega} (X(\omega) - a)^2 \mathbb{P}(d\omega).$$

Let us now derive the minimum value for the number a , that is the best value that cancels out the random variable X . Just take the derivative with respect to a :

$$\begin{aligned} \frac{d}{da} f(a) &= \int_{\Omega} \mathbb{P}(d\omega) \left(-2(X(\omega) - a) \right) = 0 \\ &= 2a \int_{\Omega} \mathbb{P}(d\omega) - 2 \int_{\Omega} \mathbb{P}(d\omega) X(\omega) = 0 \\ \implies a &= \underbrace{\int_{\Omega} \mathbb{P}(d\omega) X(\omega)}_{\text{definition of } \mathbb{E}X}. \end{aligned}$$

We must now recall some famous inequalities.

- **Markov inequality:** this is for positive, real-valued random variables. Let X be a positive and real-valued random variable. We may be interested in the probability that X exceeds some positive value b . We know that

$$\mathbb{P}(X > b) \leq \frac{1}{b} \mathbb{E}X \quad \forall b > 0.$$

- **Chebyshev inequality:** this is too for positive, real-valued random variables. Let X be a positive and

real-valued random variable. What we have is that

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq \frac{\text{Var } X}{\varepsilon^2}, \quad \varepsilon > 0.$$

- **Jensen inequality:** Let X be a real-valued random variable with finite expectation (integrable) and let f be a convex function on \mathbb{R} . Then

$$\mathbb{E}f(X) \geq f(\mathbb{E}X).$$

If the function is concave the inequality is the opposite.

Definition 1.3.2

Let X, Y be two real-valued random variables with finite variance. The **covariance** between X and Y is

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (try to prove it but it is easy even for me³⁶).

Remark

If we want to calculate the variance of two real-valued random variables then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

This can be proven very easily. Of course, if X and Y are an independency then $\text{Cov}(X, Y) = 0$ and $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Consider $\mathbb{E}|X| < +\infty$. We think of random variables with finite expectation in absolute value as “well-behaved”, without any crazy jumps or fluctuations. So it could be interesting to look for a way to treat all of these random variables with finite expectation in the same way.

Remark

A random variable (real-valued) random variable with finite expectations is called **integrable random variable**.

We are going back to functional analysis now. Brace yourselves.

1.3.2 L^p spaces

We start, as usual, with the $(\Omega, \mathcal{H}, \mathbb{P})$ probability space, with the real-valued random variable X and the parameter $p \in [1, \infty]$. Why doesn't p start from 0? It's dumbass mathematical reason we won't and shouldn't care about³⁷.

Define the so-called L^p -norm of X as:

$$\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}, \quad p < +\infty.$$

Remember that expectations are integrals. We can also define a *sup-norm*:

$$\|X\|_\infty = \inf(b \in \mathbb{R}_+ : |X| \leq b \text{ a.s.}).$$

So, taking a random variable X and fixing p we can calculate a number to attach to this random variable.

Example 1.3.2

Take $p = 1$. Our L^p -norm would then be

$$\|X\|_1 = \mathbb{E}|X|.$$

No shit. But as long as p is finite we can technically write the L^p norm of the random variable. The same goes for $p = \infty$, but with the different definition. We will call the *sup-norm* **essential supremum of X** . This

³⁶Proceeds *not* to do it.

³⁷ L in L^p stands for Lebesgue space. If $p < 0$ these spaces do not have the property of Lebesgue spaces so we don't care.

name has a meaning: imagine the random variable X “folded” by the absolute value. We are interested in the number b for which the probability mass is more or left all to the left of this number.

Note the notation!

$$\text{essup}(X) = \|X\|_\infty .$$

Remark

If $\|X\|_p = 0 \implies X = 0$ almost surely. Remember that the L^p -norm has an absolute value, so if the norm is 0 then the random variable must be 0 everywhere.

If we fix a number $c \geq 0$ we have

$$\|cX\|_p = c \|X\|_p .$$

If we have $1 \leq p \leq q \leq +\infty$ then

$$0 \leq \|X\|_p \leq \|X\|_q \leq +\infty .$$

Definition 1.3.3

The collection of real-valued random variables X with $\|X\|_p < +\infty$, $p \in [1, \infty]$ is called **L^P**.

So we are creating sub-collections of real-valued random variables such that they have a finite norm. The space L^p is a subset of the space of all random variables.

Remark

X is in L^p , $p \in [1, \infty)$ if and only if $|X|^p$ is integrable and X is in L^∞ if and only if X is almost surely bounded.

Remark

If $1 \leq p \leq q \leq +\infty$ then

$$L^q \subset L^p$$

So the L^p spaces are one inside the other, with L^∞ being the smallest of them all and L^1 being the biggest. Think about the meaning: L^1 is the space of all integrable random variables; L^2 is the space for which $|X|^2$ is integrable, but if $|X|^2$ is integrable so is $|X|$... and so on. We can thus prove certain properties for the whole class and this will extend to smaller classes.

The regularity of these functions is evident when we look at L^∞ : the functions that belong in this space are very regular because they are almost surely bounded and therefore their support doesn't explode towards infinity³⁸.

Remark

Some useful inequalities...

- **Minkowsky inequality:**

$$\|X + Y\|_r \leq \|X\|_p + \|Y\|_p .$$

- **Hölder inequality:**

$$\|XY\|_r \leq \|X\|_p \|Y\|_q$$

where $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ for $p, q, r \in [1, \infty)$.

- **Jensen inequality:** Consider:

- a convex domain D in \mathbb{R}^d ;
- a continuous and concave function $f : D \mapsto \mathbb{R}$;

³⁸Professor Polito said: “Good guys”. :3

- a sequence of random variables X_1, X_2, \dots, X_d integrable (in L^1 space) for which $(X_1, X_2, \dots, X_d) \in D$ almost surely.

Then

$$\mathbb{E}f(X_1, X_2, \dots, X_d) \leq f(\mathbb{E}(X_1, X_2, \dots, X_d)).$$

When we talk about integrability, we have this useful lemma:

Lemma 1.3.1

Let X be a real-valued random variable. X is integrable if and only if

$$\lim_{b \rightarrow \infty} \mathbb{E}|X| \mathbb{1}_{(|X| > b)} = 0.$$

This lemma states that this is an equivalent condition for integrability. The random variable $|X| \mathbb{1}_{(|X| > b)} = 0$ is equal to 0 until b and equal to the absolute value of X after b . So we are effectively truncating the random variable but maintaining the right tail. This means that if we integrate only the tails of the random variable we wouldn't get much information (we can say that the tail is *thin*).

Proof

Let's call Z_b the truncated random variable $|X| \mathbb{1}_{(|X| > b)} = 0$. Note that $Z_b \leq |X|$ and that $Z_b \xrightarrow[b \rightarrow \infty]{} 0$. Since X is integrable, we use the dominated convergence theorem:

$$\lim_{b \rightarrow \infty} \mathbb{E}|X| \mathbb{1}_{(|X| > b)} = \mathbb{E}\left[\lim_{b \rightarrow \infty} \mathbb{1}_{(|X| > b)}\right] = 0.$$

The first statement is proved. Conversely, if $\mathbb{E}Z_b \rightarrow 0$, we can choose b large enough such that

$$\mathbb{E}Z_b \leq 1.$$

Then

$$|X| \leq b + Z_b.$$

This is not immediate but it's true. Due to linearity of expectation, we know that

$$\mathbb{E}|X| \leq b + \mathbb{E}Z_b$$

but we know that $\mathbb{E}Z_b \leq 1$ so

$$\mathbb{E}|X| \leq b + 1 \leq \infty.$$

□

1.3.3 Uniform integrability

Definition 1.3.4

A collection of random variables K is said to be **uniformly integrable** if

$$k(b) = \sup_{X \in K} (\mathbb{E}|X| \mathbb{1}_{(|X| > b)})$$

goes to 0 as $b \rightarrow \infty$.

Note that it is a function of b . We consider the supremum of the collection to be “conservative”: if the supremum goes to 0 then I know the whole collection will.

Remark

- ① If K is finite and each $X \in K$ is integrable then K is uniformly integrable.
- ② If K is dominated by an integrable random variable Z then it is uniformly integrable.

- ③ Uniform integrability of a collection K implies the so-called L^1 -boundedness, which means

$$k \subset L^1 \quad \text{and} \quad k(0) = \sup_K \mathbb{E}|X| < \infty.$$

Note that $k(0)$ considers the whole random variable without truncation.

Proof

$$\mathbb{E}|x| \leq b + k(b) \quad \forall X \in K.$$

So, since now every random variable is in K and K is uniformly integrable, we use this property of K to choose a value for b such that $k(b) \leq 1$ and therefore it is finite. \square

- ④ We know that uniform integrability implies L^1 ... but the converse is not true. We can prove it by a counterexample.

Proof

Consider the probability space

$$\left((0, 1), \mathcal{B}_{(0,1)}, \underbrace{\lambda}_{\text{Lebesgue measure}} \right).$$

Normally the Lebesgue measure is infinite on the whole support, but if we restrict the measure on the unit interval then the lebesgue measure has maximum value 1 and so it is a probability measure.

Consider the collection

$$K = (X_n)_{n \geq 1} \quad \text{s.t. } X_n = \begin{cases} n, & \omega \leq \frac{1}{n} \\ 0, & \text{otherwise} \end{cases}$$

Note that

$$\forall n \quad \mathbb{E}X_n = 1$$

That is, K is L^1 -bounded.

But if we calculate $k(b)$ we realize it is equal to 1 for each b , so

$$\mathbb{E}X_n \mathbf{1}_{(X_n > b)} = \mathbb{E}X_n = 1 \quad \forall n > b$$

Therefore the collection K is *not* uniformly integral. \square

- ⑤ If K is L^p -bounded with $p > 1$ then it is uniformly integrable with $f(x) = x^p$.

To prove this we recur to the following proposition:

Proposition 1.3.1

Suppose it exists a positive borel function f on \mathbb{R}_+ such that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{x} = \infty$$

and write

$$c = \sup_{X \in K} \mathbb{E}f \circ |X|. \quad (\star)$$

If $c < \infty$ then K is uniformly integrable.

Proof

We are talking about the absolute value of X , so we don't lose generality if we talk about positive random variables. Let us assume, then, that $X \in \mathbb{R}_+$. Let us replace f with $f \vee 1$ if necessary and assume $f \geq 1$.

Let $g(x) = \frac{x}{f(x)}$ and note that if we are looking at the tail of the random variable we get

$$X \mathbf{1}_{(X_n > b)} = (f \circ X)(g \circ X) \mathbf{1}_{(X_n > b)}.$$

This is true because $(f \circ X)(g \circ X) = f(x) \frac{x}{f(x)}$.

Let us now bound this quantity:

$$\begin{aligned} X \mathbf{1}_{(X_n > b)} &= (f \circ X)(g \circ X) \mathbf{1}_{(X_n > b)} \\ &\leq (f \circ X) \sup_{x > b} g(x). \end{aligned}$$

If now $X \in \mathbb{R}$ (not necessarily positive) we can write this inequality as

$$|X| \mathbf{1}_{(|X| > b)} \leq (f \circ |X|) \sup_{x > b} g(x).$$

Take the expectation and the \sup_K , using equation \star (remember that $k(b) = \sup_{X \in K} (\mathbb{E}|X| \mathbf{1}_{(|X| > b)})$):

$$k(b) \leq c \cdot \sup_{x > b} g(x).$$

To study the behaviour in the limit for $b \rightarrow \infty$, we must study separately c (that is a finite number) and $\sup_{x > b} g(x)$ (that for $b \rightarrow \infty$, and therefore for $x \rightarrow \infty$, goes to 0 as $g(x) = \frac{x}{f(x)} = \frac{1}{\frac{f(x)}{x} \rightarrow 0}$). Hence, K is uniformly integrable. \square

Let's see one more³⁹ theorem about uniform integrability.

Theorem 1.3.2

The following are equivalent:

- ① K is uniformly integrable;
- ② $h(b) = \sup_K \int_b^{+\infty} dy \mathbb{P}(|X| > y) \xrightarrow[b \rightarrow \infty]{} 0$;
- ③ $\sup_K \mathbb{E} f \circ |X| < +\infty$ for some increasing convex function f on \mathbb{R}_+ such that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{x} = +\infty.$$

So we can consider all of these as alternative definition of uniform integrability and this is important because it is the extension of the concept of integrability for random variables. For example, we will treat only uniformly integrable martingales, which are good boys who's mommy good boy you are.

1.4 σ -algebras and random variables

There is a way to connect σ -algebras and random variables and we will see how linked the two objects are. Consider a random variable X taking values in (E, \mathcal{E}) . Then consider the collection of all inverse image of the set A through the random variable X :

$$\sigma X = X^{-1} \mathcal{E} = \{X^{-1} A : A \in \mathcal{E}\}$$

³⁹There will be consequences.

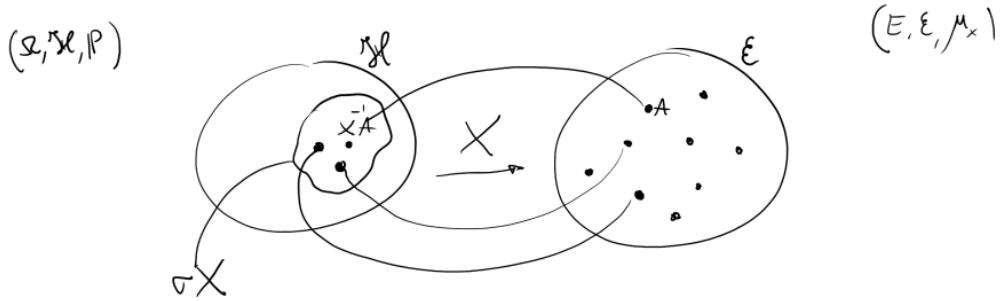


Figure 1.8: Remember that \mathcal{H} is made of the subsets of Ω and that \mathcal{E} is made of the subsets of E . Also, σX is a subset of \mathcal{H} .

σ -algebra

Remark

σX is the smallest σ -algebra \mathcal{G} on Ω such that X is measurable relative to \mathcal{G} and \mathcal{E} .

Take now T as an arbitrary index set (which can be countable or uncountable). Consider $\forall t \in T$, X_t being a random variable taking values in the measurable space (E_t, \mathcal{E}_t) . Then

$$\sigma \{X_t : t \in T\} = \sigma(X_t)_{t \in T} = \sigma \left(\bigcup_{t \in T} \sigma X_t \right).$$

This is just the union of the σ -algebras generated by the random variable X : we consider the σ -algebra of this union just to make sure that this union is, indeed, a σ -algebra.

Note the notation!

Remember that the σ -algebra generated by the collection $(X_t)_{t \in T}$ is noted as

$$\bigvee_{t \in T} \sigma X_t = \sigma \left(\bigcup_{t \in T} \sigma X_t \right)$$

Furthermore, it is the smallest σ -algebra \mathcal{G} on Ω such that $\forall t \in T$ the random variable X_t is measurable relative to \mathcal{G} and the corresponding \mathcal{E}_t .

Remark

Consider the collection $X = (X_t)_{t \in T}$ to be the random variable taking values in the product space

$$\bigotimes_{t \in T} (E_t, \mathcal{E}_t).$$

Define now $X(\omega)$ to be the point $(X_t(\omega))_{t \in T}$. The mapping

$$\omega \mapsto X_t(\omega)$$

is a random variable and is called **t -coordinate of X** .

⁴⁰This image will mean absolutely nothing to you unless you were terminally online in the summer of 2024. Otherwise it is hilarious.

Proposition 1.4.1

If $X = (X_t)_{t \in T}$ then

$$\sigma X = \sigma(X_t)_{t \in T}$$

So if we generate the σ -algebra from the collection as $\bigvee_{t \in T} \sigma X_t$ and we generate the σ -algebra from a collection seen a single random variable (as $\sigma X = X^{-1}\mathcal{E}$) we end up with the same object. Professor Polito think this is a nice thing. I am very keen on that man, to be honest. His enthusiasm is almost contagious.

Remember: when we define a random variable we actually do two things:

- we induce a distribution (that is a probability measure) to the arrival space:

$$(\Omega, \mathcal{H}, \mathbb{P}) \xrightarrow[X]{} (E, \mathcal{E}, \mu_x);$$

- we induce the existence of a special σ -algebra that links the σ -algebra \mathcal{H} and the sa \mathcal{E} (that is the σ -algebra σX).

We can define a new random variable V which is measurable with respect to the σ -algebra σX . Since we defined V as a random variable then it is automatically measurable with respect to \mathcal{H} and \mathcal{E} , but in particular when V is also measurable with respect to σX then it has a special representation. Before going through the theorem, let's revise once more a couple of measure theory concepts:

Revise with Kotatsu!

Definition 1.4.1

The collection of functions \mathcal{M} is called **monotone class** if:

- $1 \in \mathcal{M}$;
- let $f, g \in \mathcal{M}_b$ (bounded functions in \mathcal{M}) and let $a, b \in \mathbb{R}$. Then $af + bg \in \mathcal{M}$;
- $(f_n) \subset \mathcal{M}_+$ with $f_n \nearrow f$. Then $f \in \mathcal{M}$.

Theorem 1.4.1

Monotone class theorem for functions. Let \mathcal{M} be a monotone class of functions on E . Suppose that $1_A \in \mathcal{M}$ for each $A \in \mathcal{C}$ where \mathcal{C} is some π -system generating \mathcal{E} .

Then, \mathcal{M} includes all positive \mathcal{E} -measurable functions and all bounded \mathcal{E} -measurable functions.

We are going to use this theorem in the proof of the following theorem..

Theorem 1.4.2

“Theorem Δ ”:

Let X be random variable taking values in the measurable space (E, \mathcal{E}) . A mapping

$$V : \Omega \mapsto \overline{\mathbb{R}}$$

belongs to σX if and only if

$$V = f \circ X$$

for some deterministic function $f \in \mathcal{E}$.

The proof will be long as fuck.

Proof

(Sufficiency)

Consider a collection \mathcal{M} of all V of the form

$$V = f \circ X, \quad f \in \mathcal{E}.$$

We would like to prove that $\mathcal{M} \subset \sigma X$, that is to say: if we have a function in the form of V the it is σX -measurable. Measurable functions of measurable functions are still measurable: f is \mathcal{E} -measurable and X is a random variable: therefore X is \mathcal{H} -measurable but X is also measurable with respect to the σ -algebra generated by itself.

Hence

V is σX -measurable.

(Necessity)

We are planning to show that $\mathcal{M} \supset \sigma X$.

1. We show that \mathcal{M} is a monotone class of functions: we must check the three properties of the class;

(a) $1_E \in \mathcal{M}$ because $1_E = f \circ X$ with $f(x) = 1 \forall x \in E$.

(b) let $U, V \in \mathcal{M}_b$. Let $a, b \in \mathbb{R}$. Plainly

$$U = f \circ X, \quad V = g \circ X$$

for the right choice of $f, g \in \mathcal{E}$. Consider now the combination

$$aU + bV = h \circ X$$

with $h = af + bg$. Note that $h \in \mathcal{E}$. Hence

$$aU + bV \in \mathcal{M}.$$

H

- (c) consider $(V_n) \subset \mathcal{M}_+$ and $V_n \nearrow V$. For each $n, \exists f_n \in \mathcal{E}$ such that $V_n = f_n \circ X$. We now want to check that $V \in \mathcal{M}$.

Consider

$$f = \sup_n f_n \in \mathcal{E}$$

and

$$V = \sup_n V_n = \sup_n f_n(X) = f(X)$$

Hence $V \in \mathcal{M}$.

2. Consider $H \in \Omega$ such that it is in σX . If it is in σX then this means that H comes from the inverse image of a set A in \mathcal{E} :

$$H = X^{-1}A$$

for some $A \in \mathcal{E}$. So

$$1_H = 1_A \circ X \in \mathcal{M}.$$

So the class \mathcal{M} contains all the indicators of the events that are contained in σX . We need to check, for the monotone class theorem, what happens to the indicators of the elements of the π -system which in this case would be σX ... that is a σ -algebra but being a σ -algebra means also being a π -system.

We have proven that \mathcal{M} contains all the indicators in σX and applying the monotone class theorem for functions we have that \mathcal{M} contains all positive random variables that are σX -measurable.

3. Let $V \in \sigma X$ be arbitrary. Then of course $V^+ \in \sigma X$ and it is positive (the same can be said for V^-). Hence $V^+ = g \circ X$ for some $g \in \mathcal{E}$ and $V^- = h \circ X$ for some $h \in \mathcal{E}$. Then

$$V = V^+ - V^- = f \circ X$$

where

$$f(x) = \begin{cases} g(x) - h(x), & \text{if } g(x) \wedge h(x) = 0 \\ 0, & \text{otherwise} \end{cases}$$

and $f \in \mathcal{E}$.

This concludes $\sigma X \subset \mathcal{M}$. □

Corollary

For each $n \in \mathbb{N}^*$, let X_n be a random variable taking values in (E_n, \mathcal{E}_n) . A mapping

$$V : \Omega \mapsto \overline{\mathbb{R}}$$

is $\sigma(X_n)_{n \in \mathbb{N}^*}$ -measurable if and only if

$$V = f(X_1, X_2, X_3, \dots)$$

For some deterministic function $f \in \bigotimes_n \mathcal{E}_n$.

Remember that \mathbb{N}^* are the natural numbers without 0.

Remark

The above corollary can also be extended to uncountable collections. Should it? I don't know. You tell me.

Why did we destroyed our balls with this section about σ -algebras?

1.4.1 Filtrations

Let T be a subset of \mathbb{R} . Let \mathcal{F}_t be a sub- σ -algebra of \mathcal{H} $\forall t \in T$. The family $\mathcal{F} = (\mathcal{F}_t)_{t \in T}$ is called **filtration** if $\mathcal{F}_s \subset \mathcal{F}_t$ for every $s < t$. In figure 1.10 we can see how \mathcal{H} contains all events we are concerned with: it is built

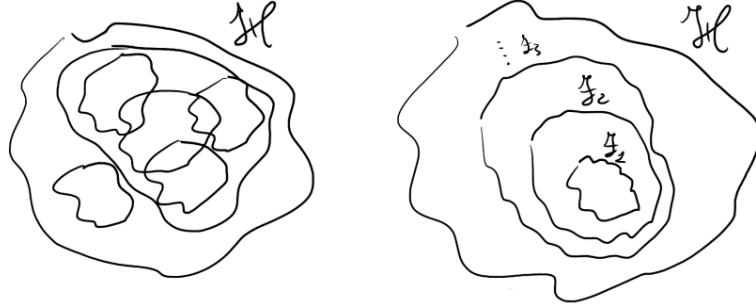


Figure 1.10: The first one is not a filtration, the second is for $T = \mathbb{N}^*$.

upon the *experience* of the event we are facing. Take, for example, \mathcal{F}_1 : it is much smaller than \mathcal{H} and if we say that a random variable is measurable with respect to \mathcal{F}_1 it means that we do not have much knowledge, but if we expand to \mathcal{F}_2 we are able to gain more knowledge about the random variable. If we interpret the index set T as time we may actually think about filtrations as our knowledge of the phenomenon as time passes.

Example 1.4.1

Consider a filtration generated by a stochastic process. Let $X = (X_t)_{t \in T}$. Of course each X_t takes value in the same state space. Define

$$\mathcal{F}_t = \sigma(X_s)_{\substack{s \leq t \\ s \in T}}.$$

Here we are generating a σ -algebra from all the random variables from the beginning of time until t . Consider now the family of σ -algebras

$$\mathcal{F} = (\mathcal{F}_t)_{t \in T} :$$

this is a filtration. This means that for this stochastic process, as time passes, we are gaining more knowledge about \mathcal{H} through the σ -algebras.

1.4.2 Independency for a finite class of sub- σ -algebras

We already know the concept of independence between random variables and the effect of independence with respect to the induced measure of probability. If we look at Çinlar , it actually introduces independence of random variables through the independence of σ -algebras. To do so we need to introduce the concept of **independency for a finite class of sub- σ -algebras**.

Definition 1.4.2

Consider the class \mathcal{H} as the collection of sub- σ -algebras \mathcal{F} :

$$\mathcal{H} = \{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n\} \quad \mathcal{F}_i = \text{sub-}\sigma\text{-algebra } \forall i = 1, \dots, n.$$

\mathcal{H} is said to be an **independency** if

$$\mathbb{E}[V_1 \cdot V_2 \cdot \dots \cdot V_n] = \mathbb{E}V_1 \cdot \mathbb{E}V_2 \cdot \dots \cdot \mathbb{E}V_n \quad ((\star))$$

for \forall random variable V_1, V_2, \dots, V_n belonging respectively to $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$.

But what does this definition mean? The definition is given in the terms of what happens to the random variables that are measurable with respect to the σ -algebras. If for each of them we get the factorization of the expectation, then our sub- σ -algebras are independency... and so, of course, are their algebras (all σ -algebras are algebras.)

This definition is valid only for a *finite* class of sub- σ -algebras, but we can extend it to infinite classes (Çinlar 's book extends it to countable collections of sub- σ -algebras and even to uncountable collections of sub- σ -algebras, so we shouldn't.). This definition, though, was just an intuition of the fact that independency of σ -algebras and independency of random variables are *not* unrelated, but are simply the same concept seen from two different points of view.

Proposition 1.4.2

Let $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ be sub- σ -algebras of \mathcal{H} , with $n \geq 2$. For each $i \leq n$, let C_i be a π -system generating \mathcal{F}_i . Then the collection $\{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n\}$ is an independency if and only if

$$\mathbb{P}(H_1 \cap H_2 \cap \dots \cap H_n) = \prod_{i=1}^n \mathbb{P}(H_i) \quad \forall H_i \in \overline{C}_i = C_i \cup \{\Omega\}, \quad i = 1, \dots, n.$$

We need to have clear that there is not an unique starting point in defining the concept of independency, so all of them are quite equivalent. We will build upon this fact and see other deep relations between these notions⁴¹.

Proposition 1.4.3

Every partition of an independency is an independency.

Take, for example, the collection $(\mathcal{F}_t)_{t \in T}$ of sub- σ -algebras and the partition $(T_i)_{i \in \mathbb{N}^*}$ of T : these two together form the sub-collection of sub- σ -algebras $\mathcal{F}_{T_i} = (\mathcal{F}_t)_{t \in T_i}$ with $i \in \mathbb{N}^*$.

Note the notation!

$$\mathcal{F}_{T_i} = \bigvee_{j \in T_i} \mathcal{F}_j.$$

We are interested in checking whether $(\mathcal{F}_{T_1}, \mathcal{F}_{T_2}, \dots)$ (remember: they are all σ -algebras!) are an independency... and they are.

Proposition 1.4.4

The notion of independence between random variables (taking values in general measurable spaces) is given in terms of independence of the σ -algebras generated by those random variables.

⁴¹Can't wait.

Example 1.4.2

Consider X_1, X_2 where X_1 takes values in (E_1, \mathcal{E}_1) and X_2 takes values in (E_2, \mathcal{E}_2) . If we are concerned about independence of the vector (X_1, X_2) we should look at the vector of the collection of generated σ -algebras $(\sigma X_1, \sigma X_2)$ and check whether it is an independency or not.

Proposition 1.4.5

Let X_1, X_2, \dots, X_n be random variables taking values respectively in (E_i, \mathcal{E}_i) for $i = 1, \dots, n$. They are independent if and only if

$$\mathbb{E}[f_1 \circ X_1, f_2 \circ X_2, \dots, f_n \circ X_n] = \mathbb{E}f_1 \circ X_1, \mathbb{E}f_2 \circ X_2, \dots, \mathbb{E}f_n \circ X_n$$

for every f_1, \dots, f_n positive and measurable with respect to $\mathcal{E}_1, \dots, \mathcal{E}_n$ respectively.

We need to prove this proposition. But before, just because we like it (like the little sluts we are), let's state an equivalent proposition.

Proposition 1.4.6

Let X_1, X_2, \dots, X_n be random variables taking values in (E_i, \mathcal{E}_i) for $i = 1, \dots, n$ respectively. Then they form an independency if and only if their joint distribution equals the product of their marginal distribution.

Well we already knew that... but thanks. Now we see more clearly how these two properties are basically the same thing.

Proof

We will prove the first proposition. We go back to formula of independence (\star) and show that it holds $\forall V_1, V_2, \dots, V_n$ positive and measurable with respect to

$$\sigma X_1, \sigma X_2, \dots, \sigma X_n$$

respectively. We know that (\star) holds because of Theorem A.

We will now prove the second proposition. From the first proposition we know that:

$$\begin{aligned} & \int_{E_1 \times \dots \times E_n} \underbrace{\pi(dx_1, \dots, dx_n)}_{\text{joint distribution of } (X_1, \dots, X_n)} f_1(x_1) \cdot \dots \cdot f_n(x_n) \\ &= \int_{E_1} \mu_1(dx_1) f_1(x_1) \cdot \int_{E_2} \mu_2(dx_2) f_2(x_2) \cdot \dots \cdot \int_{E_n} \mu_n(dx_n) f_n(x_n) \end{aligned}$$

Plainly, considering positivity of f_1, f_2, \dots, f_n , to have the above equality we obtain that

$$\pi = \mu_1 \times \mu_n.$$

□

You should convince yourself that the second proposition is just the first proposition rewritten. So, the independence of random variables is very important and right now we will see an application.

1.4.3 Sums of independent random variables

The simplest structure that we can construct with random variable is just adding them up. Let X, Y be independent real-valued random variables with distribution μ, ν respectively. Then the distribution $\mu \star \nu$ of $X + Y$ (i.e. the effect of $\mu \star \nu$ over f) is given by

$$(\mu \star \nu)f = \mathbb{E}f(X + Y) = \int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} \mu(dy) f(x + y).$$

This is called the **convolution** of the measures μ and ν .

So if we are asked to calculate the distribution it is interesting to choose a specific f to see why the distribution is defined in this way: let us choose $f = \mathbb{1}_{(-\infty, z]}(x + y)$ so that when we calculate the expectation we get

$$\begin{aligned}\mathbb{E}f(X + Y) &= \mathbb{E}\mathbb{1}_{(-\infty, z]}(X + Y) = \mathbb{P}(X + Y \leq z) \\ &= \int_{\mathbb{R}} \mu(dx) \int_{\mathbb{R}} \nu(dy) \mathbb{1}_{(-\infty, z]}(x + y) \\ &= \int_{\mathbb{R}} \nu(dy) \int_{\mathbb{R}} \mathbb{1}_{(-\infty, z-y]}(x) \mu(dx) \\ &= \int_{\mathbb{R}} \underbrace{F_X(z-y)}_{\text{distribution function of } X} \nu(dy)\end{aligned}$$

Which is what, supposedly, we saw in our undergraduate course⁴². From this last series of equations we can also have a look to a special case where X is absolutely continuous with respect to the Lebesgue measure with density $h(x)$:

$$\begin{aligned}\mathbb{P}(X + Y \leq z) &= \int_{\mathbb{R}} F_X(z-y) \nu(dy) = \int_{\mathbb{R}} \nu(dy) \int_{-\infty}^{z-y} h(t) dt \\ &\stackrel{w=t+y}{=} \int_{\mathbb{R}} \nu(dy) \int_{-\infty}^z h(w-y) dw \\ &= \int_{-\infty}^z \underbrace{\int_{\mathbb{R}} h(w-y) \nu(dy) dw}_{\text{density of } X+Y}.\end{aligned}$$

Further, if Y is also absolutely continuous with density g we get that the density of $X + Y$ becomes

$$\int_{\mathbb{R}} h(w-y) g(y) dy$$

Which is the convolution formula for calculating the density of $X + Y$. For a further example see ex. 2.1.3 of Cinlar page 43 (convolution of Gamma densities).

We can further consider a sequence of random variables $(X_n)_{n \in \mathbb{N}^*}$ and start adding them up:

$$S_n = X_1 + X_2 + X_3 + \dots + X_n$$

so we stop at a finite time. But then we have a new sequence, which is the sequence $(S_n)_{n \in \mathbb{N}^*}$ where S_n is called *partial sum*. The partial sum is in itself a random variable, made by little random variables added up and this is why S_n is also called **random walk**. The question is, as always,

“Should we care?”

and while the answer is often

“No”

this time it actually is

“Kinda”

which is probably the most astounding result of this whole course. This is one of the first kinds of real problems (albeit very simplistic) modeled through the use of random variables. To do this we need the notion of *tail σ -algebra*.

1.4.4 Tail σ -algebra

Did you think that we were about to dive into some interesting and useful practical applications of the The name alone suggests a concept of something away from us. Think of an experiment involving $n \geq 1$ trials, like the repeated roll of a die or toss of a coin. What matters is that we have a trial repeated infinite times and the result of all these trials (which are random variables) makes up our experiment, which is a random variable in itself.

The experiment is defined on $(\Omega, \mathcal{H}, \mathbb{P})$. Consider a sequence $(\mathcal{G}_n)_{n \in \mathbb{N}^*}$ of sub- σ -algebras of \mathcal{H} such that \mathcal{G}_n is the information on \mathcal{H} revealed by the n -th trial.

⁴²I didn't see anything remotely similar to this in my economics degree but what do I know. Economics is not real anyway and economists should kill themselves NOW.

Example 1.4.3

Imagine that the trials are given by the sequence of trials is given by

$$(X_n)_{n \in \mathbb{N}^*}$$

and the sub- σ -algebras are defined by

$$(\sigma X_n)_{n \in \mathbb{N}^*}$$

So the σ -algebras generated by each X_n are actually the *trajectories* of the random variables. Imagine we are sitting a time n . Consider now \mathcal{G}_m for $m > n$: I want to take the union of all these σ -algebras, but since I am not sure that we would end up with a σ -algebra, I take the σ -algebra of that union.

$$\tau_n = \sigma \left(\bigcup_{m > n} \mathcal{G}_m \right) = \bigvee_{m > n} \mathcal{G}_m.$$

This σ -algebra depends on n and it represents the *information about the future*, since we are now sitting at the time point n . Remember that when we talk about *revealed information* we are always talking about σ -algebras! This object is a bit strange... it is about the information that models the future. What the fuck? We now want to take the intersection of the various τ_n (which is surely a σ -algebra since we are intersecting):

$$\tau = \bigcap_n \tau_n.$$

This is now the *information about the remote future*: we are doing the *intersection*, not the *union*! This is not exactly what happens in the future, but the information revealed by all the infinite trials in the future. The σ -algebra τ is called **tail σ -algebra**. Remember that we always remain inside \mathcal{H} .

Example 1.4.4

Consider $(X_n)_{n \in \mathbb{N}^*}$ and $(\mathcal{G}_n)_{n \in \mathbb{N}^*} = (\sigma X_n)_{n \in \mathbb{N}^*}$. We have

$$S_n = \sum_{i=1}^n X_i.$$

① consider the event

$$\left\{ \omega : \lim_n S_n(\omega) \text{ exists.} \right\}$$

If we think about the tail σ -algebra, it is the intersection of the σ -algebras after n : it is composed by events whose occurrence is not influenced by the happenings in finite time! If we look at this event it consists of all the realization of the random walk until infinity: so if I were to remove the first n random variables in this sequence then the limit of this sequence wouldn't be affected: so this event *belongs* to τ .

② consider the set $B \in \mathcal{B}_{\mathbb{R}}$ and consider the event

$$\left\{ \omega : X_n(\omega) \in B \text{ i.o.} \right\}.$$

"i.o." means *infinitely often*: this means that

$$\left\{ \omega : X_n(\omega) \in B \text{ i.o.} \right\} = \left\{ \omega : \sum_n \mathbb{1}_B \circ X_n(\omega) = +\infty \right\}.$$

So we infinitely many $\mathbb{1}$'s in B . Again, it is clear that this event belongs in the tail σ -algebra because to test the divergence of the series $\sum_n \mathbb{1}_B \circ X_n(\omega)$ we need to consider the behavior in the limit. We don't care what happens in finite time, so if we remove the contribution of the first n random variables:

$$\left\{ \omega : X_n(\omega) \in B \text{ i.o.} \right\} \in \tau.$$

③ consider again $B \in \mathcal{B}_{\mathbb{R}}$ and consider

$$\left\{ \omega : S_n(\omega) \in B \text{ i.o.} \right\}.$$

This event does not belong in τ ! Why? Let's consider a special case of $B = \{0\}$ and therefore we consider the event

$$\{\omega : S_n(\omega) = 0 \text{ i.o.}\}$$

and we also specialize the “jumps”:

$$(X_n)_{n \in \mathbb{N}^*} \text{ is i.i.d. with } \begin{cases} \mathbb{P}(X_1 = 1) = p \\ \mathbb{P}(X_1 = -1) = 1 - p \end{cases}$$

with $p \in [0, 1]$. In this case $\{\omega : S_n(\omega) = 0 \text{ i.o.}\}$ is not a tail event. Let's realize the sequence of the “jumps” choosing an ω such that

$$X(\omega) = (+1, -1, +1, -1, +1, \dots)$$

Clearly $\omega \in \{\omega : S_n(\omega) = 0 \text{ i.o.}\}$:

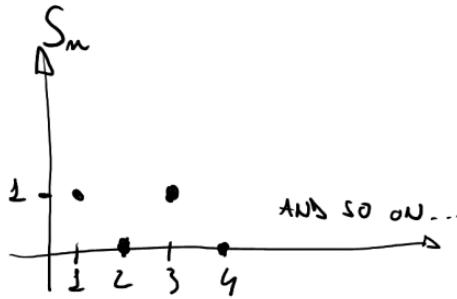


Figure 1.11: The random walk randomly walking

Now we choose a different $\tilde{\omega}$ such that:

$$X(\tilde{\omega}) = (+1, +1, +1, +1, -1, +1, -1, \dots)$$

So in this case we have a “ramp” before the oscillation: $\tilde{\omega} \notin \{\omega : S_n(\omega) = 0 \text{ i.o.}\}$! Try to make the same drawing: the walk will go up and the 0 will never be visited!

Observe that ω and $\tilde{\omega}$ agree from the fourth coordinate on: but this means that the first events are decisive when it comes to determine whether a certain ω belongs to $\{\omega : S_n(\omega) = 0 \text{ i.o.}\}$. This means that

$$\{\omega : S_n(\omega) = 0 \text{ i.o.}\} \notin \underbrace{\sigma(X_4, X_5, X_6, \dots)}_{\text{We need more information to decide!}}$$

So of course $\{\omega : S_n(\omega) = 0 \text{ i.o.}\} \notin \tau$. The nature of the event is changed by a finite number of occurrences...

④ let (X_1, X_2, \dots) be independent and $S_n = \sum_{i=1}^n X_i$. Then

$$\left\{ \omega : \frac{S_n(\omega)}{n} \text{ converges (exists finite)} \right\} \in \tau.$$

Consider

$$Z_1 = \liminf_{n \rightarrow +\infty} \frac{S_n}{n}, \quad Z_2 = \limsup_{n \rightarrow +\infty} \frac{S_n}{n}$$

so that our event becomes

$$\{\omega : Z_1(\omega) = Z_2(\omega)\}.$$

Rewrite

$$\frac{S_n}{n} = \underbrace{\frac{1}{n} \sum_{i=1}^m X_i}_{S_{n1}} + \underbrace{\frac{1}{n} \sum_{i=m+1}^n X_i}_{S_{n2}} \quad m \leq n$$

So that we “split” the partial sums in two sums before and after time m . We call these two partial sums S_{n1} and S_{n2} . Note that when $n \rightarrow \infty$ then $S_{n1} \rightarrow 0$ and therefore Z_1 and Z_2 do not depend upon the first n values of (X_1, X_2, \dots, X_m) . Therefore

$$\{Z_1 = Z_2\} \in \tau.$$

Theorem 1.4.3

Kolmogorov's 0-1 law

This is a theorem about independence. Let $\mathcal{G}_1, \mathcal{G}_2, \dots$ be independent. Then

$$\mathbb{P}(H) = \begin{cases} 0 \\ 1 \end{cases} \quad \forall H \in \tau.$$

So when the outcomes of a random variable are independent then the tail is made of almost sure or almost impossible events! The proof is short.

Proof

Remember that partition of independencies are independencies. Our independency is formed by the sub- σ -algebras $\mathcal{G}_1, \mathcal{G}_2, \dots$ and our partition is

$$(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n, \tau_n).$$

This is an independency $\forall n \in \mathbb{N}^*$. Since $\tau \subset \tau_n \forall n$, then $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n, \tau)$ is an independency for each n . But since we now have an independency made of finite elements for each n we have an independency of infinite components

$$(\tau, \mathcal{G}_1, \mathcal{G}_2, \dots)$$

by definition of independency for countably infinite sequences. But again, if we partition that last independency we still get an independency:

$$(\tau, \tau_0)$$

is still an independency. Now consider two σ -algebras $H \in \tau, G \in \tau_0$. Since they are independent, we have that $\mathbb{P}(H \cap G) = \mathbb{P}(H)\mathbb{P}(G)$. But since $\tau \subset \tau_0$ then we can choose $G = H$ and we get

$$\mathbb{P}(H) = \mathbb{P}(H \cap H) = \mathbb{P}(H)\mathbb{P}(H)$$

and the solution to this equation can only be 1 or 0. \square

We must remember that this is a special case where we have independence: this, in an experimental scenario, is modeled by independent trials. This result can be applied of concept of convergence. The idea is to build up instruments useful to the analysis of real random variable in their limits.

Oh, also: there is a fucking corollary of this piece of shit theorem. They didn't tell me and I was trying to do a fucking shit fucking god fucking dick in ass fucking kill myself exercise about tail σ -algebras and I couldn't for the fucking love of fucking god understand and I wanted to cry and I cried and this was because there is a fucking corollary of this horrible law that goes like this:

Corollary

If $\mathcal{G}_1, \mathcal{G}_2, \dots$ are independent, then every random variable X in the tail σ -algebra τ (which means, X is measurable with respect to τ) must be a constant $c \in \mathbb{R}$ \mathbb{P} – a.s..

1.5 Convergence of random variables and asymptotic behavior

1.5.1 Convergence of real sequences

Let $(x_n)_n$ be a sequence in $\mathbb{R}, n \in \mathbb{N}^*$. Then

$$\liminf_n x_n = \sup_m \inf_{n \geq m} x_n$$

and the

$$\limsup_n x_n = \inf_m \sup_{n \geq m} x_n.$$

Both are well defined numbers (and can possibly be infinity). We know that if

$$\liminf_n x_n = \limsup_n x_n$$

then the limit

$$\lim_n x_n$$

exists and the sequence is **convergent**.

Note the notation!

Let $\varepsilon \in \mathbb{R}_+$. We let i_ε be the indicator function of (ε, ∞) :

$$i_\varepsilon(x) = \mathbb{1}_{(\varepsilon, \infty)}(x) = \begin{cases} 1, & x > \varepsilon \\ 0, & x \leq \varepsilon. \end{cases}$$

We need two remarks about the convergence of x_n :

Remark

- Let $(x_n)_n$ be a sequence in \mathbb{R} . Then it converges to $x \in \mathbb{R}$ if and only if

$$|x_n - x| \xrightarrow{n} 0.$$

- Let $(x_n)_n$ be a sequence in \mathbb{R} . Then it converges to 0 if and only if

$$\forall \varepsilon > 0 \exists k \text{ s.t. } x_n \leq \varepsilon \quad \forall n \geq k.$$

This is the usual definition of the limit and it means that

$$x_n \xrightarrow{n} 0 \iff \sum_n i_\varepsilon < +\infty \quad \forall \varepsilon > 0.$$

This means that we have a finite number of x_n that are smaller than ε , so summing up the indicator function when we are in an interval larger than ε we will get less than infinity. This concept connects the convergence of real-valued positive sequences with the convergence to 0 of real-valued positive sequences. Further, the fact that

$$\sum_n i_\varepsilon(x_n) < +\infty \iff \limsup_n i_\varepsilon(x_n) = 0 \iff \lim_n i_\varepsilon(x_n) = 0$$

Proposition 1.5.1

Cauchy's criterion:

Let $(x_n)_n$ be a sequence in \mathbb{R} . Then $(x_n)_n$ converges if and only if

$$\lim_{m,n \rightarrow \infty} |x_n - x_m| = 0.$$

This criterion is interesting because it gives us a criterion to check whether the sequence converges without looking at the actual limit of the sequence...

Proposition 1.5.2

If there exists a positive sequence $(\varepsilon_n)_n$ such that

$$\sum_n \varepsilon_n < +\infty, \quad \sum_n i_{\varepsilon_n}(|x_{n+1} - x_n|) < +\infty$$

Then $(x_n)_n$ is convergent.

This is another way to prove convergence: we just need to prove that the two series described above are finite.

Definition 1.5.1

Given a sequence $(x_n)_n$, the sequence $(y_n)_n$ is said to be a **subsequence of $(x_n)_n$** if there exists an increasing sequence $(k_n)_n \in \mathbb{N}^*$ with $\lim_n k_n = \infty$ such that

$$y_n = x_{k_n} \quad \forall n.$$

So the subsequence is like a tool to extract elements from a sequence that are still a sequence!

Note the notation!

We will usually denote $(y_n)_n$ as $(x_n)_{n \in N}$ where $N = (k_n)_n$.

Remark

We say that $(x_n)_n$ converges **along N** to x if $(x_n)_{n \in N}$ converges to x . This basically means that we are extracting a subsequence from x and check if it converges.

Remark

$(x_n)_n$ converges to x if and only if every subsequence has the same limit x . The converse is also true!

Remark

Let $(x_n)_n$ be a bounded real sequence. Then it is always possible to extract from it a convergent subsequence.

So if we prove that $(x_n)_n$ is bounded we can always get a convergent subsequence...

Proposition 1.5.3

Selection principle:

If every subsequence that has a limit has the same limit value x , then the sequence that we are considering tends to the same limit x (that can be finite or infinite). If the sequence is bounded and every convergent subsequence has the same limit x , then the sequence converges to x .

We now consider a lemma about the behavior of \limsup and \liminf .

Lemma 1.5.1

Let $(x_n)_n$ be a sequence of positive real numbers and let $\bar{x}_n = \frac{\sum_{i=1}^n x_i}{n}$. Let $N = (n_k)_k$ be a subsequence of \mathbb{N} with

$$\lim_k \frac{n_{k+1}}{n_k} = r > 0.$$

If $(\bar{x}_n)_n$ converges along N to x then

$$\frac{x}{r} \leq \liminf_n \bar{x}_n \leq r \cdot x.$$

This is interesting because it gives us an idea of the fluctuation of the sequence of averages in terms of the behaviour of the subsequence N and the limit x . We will make use of this lemma soon...

Theorem 1.5.1

Helly's theorem:

For every sequence $(c_n)_n$ of distribution functions (this is not a real-valued sequence anymore) there exists a subsequence of distribution function $(b_n)_n$ and a limiting distribution function x such that

$$\lim_{n \rightarrow \infty} b_n(t) = c(t) \quad \forall t \text{ at which } c \text{ is continuous.}$$

Lemma 1.5.2

Kronecker's lemma:

Let $(x_n)_n$ a real-valued sequence. let $(a_n)_n$ be a strictly positive sequence increasing to ∞ . Finally, write

$$y_n = \frac{\sum_{k=1}^n x_k}{a_k}.$$

Then if (y_n) is convergent,

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k=1}^n x_k = 0.$$

We are now ready to talk about what we actually care about in probability theory.

1.5.2 Almost sure convergence

Definition 1.5.2

A real-valued sequence of random variables $(X_n)_n$ on $(\Omega, \mathcal{H}, \mathbb{P})$ is said to be **almost sure convergent** (a.s. convergent) is the numerical sequence

$$(X_n(\omega))_n$$

converges for almost all $\omega \in \Omega$.

It is said to converge to X if X is an almost sure real-valued random variable and

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

for almost all $\omega \in \Omega$.

Let us define

$$\Omega_0 = \left\{ \omega \in \Omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega) \right\}.$$

Of course this is an event and we note that $(X_n)_n$ is almost sure convergent if and only if Ω_0 is an almost sure event (i.e. $\mathbb{P}(\Omega_0) = 1$).

Theorem 1.5.2

Characterization of almost sure convergence:

A sequence of real valued random variables $(X_n)_n$ converges to X almost surely if and only if, for every $\varepsilon > 0$,

$$\sum_n i_\varepsilon \circ |X_n - X| < \infty \quad (\star\star)$$

almost surely.

Proof

Implication (\Rightarrow).

Start from the fact that

$$X_n \rightarrow X \text{ a.s.}$$

Write $Y_n = |X_n - X|$: for every $\omega \in \Omega$, by the analogous result for real sequences, we have that

$$\sum_n i_\varepsilon \circ Y_n(\omega) < \infty \quad \forall \varepsilon > 0$$

and therefore ($\star\star$) holds.

Then there is the converse implication (\Leftarrow).

Let ($\star\star$) hold and let us consider $(\varepsilon_k)_k$ be a strictly decreasing sequence converging to 0. Also, let

$$N_k = \sum_n i_{\varepsilon_k} \circ |X_n - X|.$$

Then

$$\mathbb{P}(N_k < \infty) = 1 \quad \forall k$$

and this is true because $(\star\star)$ holds! Now note that

$$\varepsilon_{k+1} < \varepsilon_k \implies i_{\varepsilon_{k+1}} \geq i_{\varepsilon_k} \implies N_{k+1} \geq N_k.$$

But now we have to pay attention to the nature of the events involved: consider the event

$$\{N_k < +\infty\} \quad \forall k.$$

The sequence $(\{N_k < +\infty\})_k$ decreases with k towards the intersection

$$\bigcap_k \{N_k < +\infty\} = \left\{ \omega \in \Omega : \sum_n i_\varepsilon \circ Y_n(\omega) \quad \forall \varepsilon > 0 \right\} \\ = \Omega_0.$$

Hence we want to evaluate

$$\mathbb{P}(\Omega_0) = \mathbb{P}\left(\lim_k \{N_k < +\infty\}\right) = \underbrace{\lim_k \mathbb{P}(N_k < +\infty)}_{\text{sequence of 1}}$$

Where the last equality is true because of sequential continuity. So $\forall \omega \in \Omega_0$ we have $X_n(\omega) \rightarrow X(\omega)$ which is the definition of almost surely convergence. \square

Let's study more what happens in the limits of these random walks. Remember that Kolmogorov's 0-1 law tells us something nice: if all the random variables are independent then events in the asymptotic future will either be certain or impossible. But we know nothing about situations in which random variables are dependent...

1.5.3 Borel-Cantelli Lemmas

Let $(H_n)_n$ be a sequence of events. Imagine evaluating the probability of all these events: we would end up with a sequence of real numbers. We may ask ourselves how the sum of the probability behaves.

Theorem 1.5.3

First Borel-Cantelli Lemma Let $(H_n)_n$ be a sequence of events. Then

$$\sum_n \mathbb{P}(H_n) < +\infty \implies \sum_n \mathbb{1}_{H_n} < +\infty \quad \text{a.s.}$$

Remember that the second sum is equivalent to

$$\mathbb{P}(H_n \text{ i.o.}) = 0 \quad \text{or} \quad \mathbb{P}(\limsup_n H_n) = 0.$$

If you want to know more see Cinlar page 101. Why do we need it to be almost surely? Well, the sum $\sum_n \mathbb{1}_{H_n}$ contains an indicator of events (which is nothing else but a Bernoulli random variable) so it is in itself a random variable. So we are asking yourself whether the sum converges *almost surely*. This theorem is important because if we can sum the probabilities then we can directly infer information about the presence of events! Let's see the proof⁴³.

Proof

Denote

$$N = \sum_n \mathbb{1}_{H_n}$$

⁴³Professor Polito says that we should be interested in the result as in the proof. I do not agree and wish to die.

So that

$$\sum_n \mathbb{P}(H_n) = \mathbb{E}N.$$

Our claim now becomes:

"If $\mathbb{E}N < +\infty$ then $N < \infty$ almost surely."

But this is obvious: if the expectation is finite then the random variable is finite almost surely (meaning that the support is finite). \square

That's it? That was cool and all. But let's see the first Borel-Cantelli lemma in action in the next proposition.

Proposition 1.5.4

Let

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty \quad \forall \varepsilon > 0.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

Here we are checking the convergence of the sum of a sequence of real numbers and we are given a result on random variables. This is pretty neat!

Proof

Consider

$$H_n = \{|X_n - X| > \varepsilon\}.$$

Then, we apply Borel-Cantelli Lemma and we calculate

$$\sum_n \mathbb{1}_{\{|X_n - X| > \varepsilon\}} < +\infty.$$

But now we are in the situation described by (**), so we know that

$$X_n \xrightarrow{\text{a.s.}} X$$

\square

So we can proof almost sure convergence by means of the application. So when we need to prove convergence always think about Borel-Cantelli lemmas!

Proposition 1.5.5

Suppose that there exists a sequence $(\varepsilon_n)_n$ decreasing to 0 such that

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < +\infty.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

So we don't need ε to be constant, but just to be decreasing to 0.

Proposition 1.5.6

Suppose that there exists a sequence of positive numbers $(\varepsilon_n)_n$ such that

$$\sum_n \varepsilon_n < +\infty, \quad \sum_n \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$$

Then X_n converges almost surely.

Let's state again the two Borel-Cantelli lemmas side by side (and with a slightly different formulation).

Theorem 1.5.4

Borel-Cantelli lemmas

a) Let $(B_n)_n$ be a sequence of Bernoulli random variables.

$$\sum_n \mathbb{E}B_n < +\infty \implies \sum_n B_n < +\infty \text{ a.s.}$$

b) If

$$\sum_n \mathbb{E}B_n = \infty \text{ and } (B_n)_n \text{ are pairwise independent}$$

then

$$\sum_n B_n = +\infty \text{ a.s.}$$

We will prove b), the so-called "divergence part".

Proof

Note the notation!

$$p_n = \mathbb{E}B_n, \quad a_n = \sum_{i=1}^n \mathbb{E}B_i = \sum_{i=1}^n p_i$$

and we get the partial sum and the limits

$$S_n = \sum_{i=1}^n B_i, \quad S = \lim_n S_n$$

First, we know that, due to pairwise independency in the hypothesis

$$\text{Var } S_n = \sum_{i=1}^n \text{Var } B_i$$

but we know that they are Bernoulli random variables, so we can write the variance explicitly:

$$\text{Var } S_n = \sum_{i=1}^n p_i(1-p_i) \leq \sum_{i=1}^n p_i = a_n$$

where the inequality is just arithmetical calculations.

Second, fix $b \in (0, +\infty)$. We know that $(a_n)_n$ increases to ∞ by hypothesis:

$$(a_n - \sqrt{ba_n})_n.$$

We are basically subtracting to a_n something that goes to ∞ as well, but slower than a_n . So the sequence $(a_n - \sqrt{ba_n})_n$ stays also increasing towards infinity. But if that is the case the event $\{S < +\infty\}$ is the limit of the increasing sequence of events

$$\{S < a_n - \sqrt{ba_n}\}.$$

Since $S_n \leq S \forall n$ consider further the following modified sequence of events:

$$\{S_n < a_n - \sqrt{ba_n}\} \supset \{S < a_n - \sqrt{ba_n}\}.$$

Next, we also have

$$\{S_n < a_n - \sqrt{ba_n}\} \subset \{|S_n - a_n| > \sqrt{ba_n}\}.$$

Now, let's switch to the probability point of view. Remember that the events are included one in the other so we can think about weak inequalities of probabilities.

$$\mathbb{P}(S < a_n - \sqrt{ba_n}) \leq \mathbb{P}(S_n < a_n - \sqrt{ba_n}) \leq \mathbb{P}(|S_n - a_n| > \sqrt{ba_n})$$

and we take the \limsup_n :

$$\limsup_n \mathbb{P}(S < a_n - \sqrt{ba_n}) \leq \limsup_n \mathbb{P}(|S_n - a_n| > \sqrt{ba_n})$$

But the left hand side of the inequality is

$$\begin{aligned}\limsup_n \mathbb{P}(S < a_n - \sqrt{ba_n}) &= \mathbb{P}\left(\lim_n \{\mathbb{P}(S < a_n - \sqrt{ba_n})\}\right) \\ &= \mathbb{P}(S < \infty).\end{aligned}$$

We get, thus,

$$\mathbb{P}(S < +\infty) \leq \limsup_n \mathbb{P}(|S_n - a_n| > \sqrt{ba_n}).$$

But now we can apply Chebyshev's inequality:

$$\mathbb{P}(S < +\infty) \leq \limsup_n \frac{\text{Var } S_n}{ba_n}$$

and we can now exploit the result we had reached before:

$$\mathbb{P}(S < +\infty) \leq \limsup_n \frac{\text{Var } S_n}{ba_n} \leq \limsup_n \frac{\sigma_n}{b\sigma_n} = \frac{1}{b}.$$

If we let $b \rightarrow \infty$ we should get the same result and immediately obtain

$$\mathbb{P}(S < +\infty) \leq \frac{1}{b} \xrightarrow[b \rightarrow \infty]{} 0$$

and hence

$$\mathbb{P}(S = +\infty) = 1.$$

□

Damn. I hate this. Let's see an example about coin toss!

Example 1.5.1

Imagine considering the probability of getting a head^a:

$$\mathbb{P}(H) \in (0, 1), \quad \Omega = \{H, T\}^\infty.$$

We are infinitely tossing coin and the probability of having head is always the same. Our probability space is thus the space of infinity sequences of H and T . Consider a chunk of tosses and define B_1 as the first subsequence of length k , B_2 as the second subsequence of length k and so on. So we get infinitely many chunks of length k .

Fix

$$s = (s_1, s_2, \dots, s_k) \in \{H, T\}^k.$$

Now consider the events

$$D_1 = \{B_1 = s\}$$

$$D_2 = \{B_2 = s\}$$

$$D_3 = \{B_3 = s\}$$

⋮

and suppose further that the sequence s has actually m heads and $k - m$ tails. Then

$$\mathbb{P}(D_i) = [\mathbb{P}(H)]^m \cdot [\mathbb{P}(T)]^{k-m}.$$

Imagine we want to calculate the sum:

$$\sum_i \mathbb{P}(D_i) = +\infty.$$

But now we can use the Borel-Cantelli lemma (point b)! So we get that

$$\sum_i \mathbb{1}_{D_i} = \infty$$

which means that

$$\mathbb{P}(D_1 \text{ i.o.}) = 1$$

or, in other words, for any given sequence of k outcomes we observe it infinitely often.

^a...so no head? Imma head out.

1.5.4 Convergence in probability

This definition is strictly related to almost sure convergence. In measure theory this is called *convergence in measure* (makes sense...).

Definition 1.5.3

Let $(X_n)_n$ be a sequence of real-valued random variables. Then $(X_n)_n$ converges to a further real-valued random variable **in probability** if

$$\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

So this means that the probability of X_n being at most ε away from X gets smaller and smaller as n gets bigger. It is quite different than the notion of almost sure convergence... and it depends heavily on the probability measure \mathbb{P} ! What we get now is a sequence of probabilities that we need to evaluate.

Example 1.5.2

Consider

$$\Omega = [0, 1], \mathcal{H} = \mathcal{B}_{[0,1]}, \mathbb{P} = \underline{\lambda} \text{ Lebesgue measure}$$

Consider the sequence of random variables

$$X_1, X_2, X_3, \dots$$

to be indicators of $(0, 1]$, $(0, \frac{1}{2}]$, $(\frac{1}{2}, 1]$, $(0, \frac{1}{3}]$, $(\frac{1}{3}, \frac{2}{3}]$, $(\frac{2}{3}, 1]$, ... respectively.
Then $\forall \varepsilon > 0, \varepsilon \in (0, 1)$, $\mathbb{P}(X_n > \varepsilon)$ forms the sequence

$$(1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \dots) \xrightarrow{n \rightarrow \infty} 0.$$

Hence

$$X_n \xrightarrow{\mathbb{P}} 0.$$

However,

$$X_n(\omega) = \begin{cases} 0 & \forall \omega \in \Omega \\ 1 & \end{cases}$$

and therefore we have that

$$\liminf_n X_n = 0, \limsup_n X_n = 1$$

and we do not have almost sure convergence... So

$$\Omega_0 = \left\{ \omega \in \Omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega) \right\}$$

is empty!

But how are these two different types of convergence related?

Theorem 1.5.5

Characterization theorem for convergence in probability.

- i) if $(X_n)_n$ converges to X almost surely, then it converges to X in probability;
- ii) if $(X_n)_n$ converges in probability to X , then it has a subsequence converging to the same random variable X almost surely;
- iii) if every subsequence of the main sequence has a further subsequence converging to X almost surely, then the main sequence converges to X in probability.

We will only prove point i) because it is the most useful.

Proof

Let $X_n > 0 \forall n$ and $X = 0$: this means replacing $|X_n - X|$ with X_n (as long as X_n is positive and converges to 0). Recall the definition of indicator $i_\varepsilon + \mathbb{1}_{(\varepsilon, \infty)}$ and define

$$p_n = p_n(\varepsilon) = \mathbb{E}[i_\varepsilon \circ X_n] = \mathbb{P}(X_n > \varepsilon).$$

Bernoulli r.v.

Here we switch between expectation and probability because of the property of expectations with indicator functions in this theorem. We know that $X_n \xrightarrow{\text{a.s.}} 0$. Fix $\varepsilon > 0$: now we have

$$\mathbb{1}_\varepsilon \circ X_n \xrightarrow{\text{a.s.}} 0.$$

Now take the expectation

$$p_n = \mathbb{E}i_\varepsilon \xrightarrow{\text{a.s.}} 0$$

Hence

$$\mathbb{P}(X_n > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

that is to say

$$X_n \xrightarrow{\mathbb{P}} 0$$

□

Proposition 1.5.7

Let $f : \mathbb{R} \mapsto \mathbb{R}$ be continuous. Then if $X_n \xrightarrow{\mathbb{P}} X$ in probability then

$$f(X_n) \xrightarrow{\mathbb{P}} f(x)$$

That is to say $f \circ N_n \xrightarrow{\mathbb{P}} f \circ X$.

Proof

We know that $X_n \xrightarrow{\mathbb{P}} X$. Let N be a subsequence of \mathbb{N}^* . Then $X_n \xrightarrow{\mathbb{P}} X$ along N . For the point ii) of the previous theorem there exists a subsequence N' along which the convergence takes place almost

surely:

$$X_n \xrightarrow{\text{a.s.}} X.$$

Then along N'

$$f \circ X_n \xrightarrow{\text{a.s.}} f \circ X$$

for the continuity of f : looking at the limit of f we exploit the continuity of f :

$$\lim f(X_n(\omega)) = f\left(\lim X_n(\omega)\right).$$

So saying that f is continuous is enough to prove the convergence of its composition with X .

Now by point *iii*) of the previous theorem we conclude that

$$f \circ X_n \xrightarrow{\mathbb{P}} F \circ N.$$

□

Remark

Convergence in probability is preserved under arithmetic operations.

So, for example if we have $X \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$ then $(X_n + Y_n)_n \xrightarrow{\mathbb{P}} X + Y$.

We will now introduce a metric for convergence in probability: since we are talking about *distance*, we may think⁴⁴ that this has something to do with *metric spaces*, where we measure the distance between different objects. Indeed there is a connection between measure spaces and metric spaces...

Let us introduce now a metric for convergence in probability. If we want to calculate a metric between random variables X and Y we can define the following metric:

$$d(X, Y) = \mathbb{E}(|X - Y| \wedge 1).$$

Remark

$$\textcircled{1} \quad d(X, Y) = 0 \iff X = Y \text{ a.s.}$$

$$\textcircled{2} \quad d(X, Y) + d(Y, Z) \geq d(X, Z).$$

d is a metric on the space of real-valued random variables if X and Y are identified as the same random variable if $X = Y$ almost surely.

Proposition 1.5.8

$$X_n \xrightarrow{\mathbb{P}} X \iff d(X_n, X) \xrightarrow{n \rightarrow \infty} 0.$$

Instead of proving this version, we will prove a slightly different version of the proposition, that is:

Theorem 1.5.6

$$X_n \xrightarrow{\mathbb{P}} X \iff \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n - X|}{1 + |X_n - X|} \right] = 0$$

Proof

Without losing generality we fix $X = 0$ a.s. We are concerned about the following claim:

$$X_n \xrightarrow{\mathbb{P}} 0 \iff \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] = 0.$$

⁴⁴A remark that reeks of overestimation.

(\implies) Suppose that $X_n \xrightarrow{\mathbb{P}} 0$. This means that

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Note that

$$\frac{|X_n|}{1 + |X_n|} \leq \frac{|X_n|}{1 + |X_n|} \mathbb{1}_{(|X_n| > \varepsilon)} + \varepsilon \mathbb{1}_{(|X_n| < \varepsilon)}$$

Because if $|X_n|$ is smaller than ε then also the fraction above is smaller than ε (duh). But we see that

$$\begin{aligned} \frac{|X_n|}{1 + |X_n|} &\leq \underbrace{\frac{|X_n|}{1 + |X_n|}}_{<1} \mathbb{1}_{(|X_n| > \varepsilon)} + \varepsilon \underbrace{\mathbb{1}_{(|X_n| < \varepsilon)}}_{\leq 1} \\ &\leq \mathbb{1}_{(|X_n| > \varepsilon)} + \varepsilon \end{aligned}$$

Then take the expectation:

$$\mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] \leq \underbrace{\mathbb{E} \mathbb{1}_{(|X_n| > \varepsilon)}}_{\mathbb{P}(|X_n| > \varepsilon)} + \varepsilon =$$

and the limit

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] \leq \varepsilon$$

Because we know that $\mathbb{P}(|X_n| > \varepsilon)$ goes to 0 as n gets larger. But since ε is chosen arbitrarily, then

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] = 0.$$

(\Leftarrow) Let

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] = 0.$$

We want to prove convergence in probability. Note that $f(x) = \frac{x}{1+x}$ is strictly increasing and therefore I can write the following inequality:

$$\frac{\varepsilon}{1 + \varepsilon} \mathbb{1}_{(|X_n| > \varepsilon)} \leq \frac{|X_n|}{1 + |X_n|} \leq \mathbb{1}_{(|X_n| > \varepsilon)} \leq \frac{|X_n|}{1 + |X_n|}.$$

Now take the expectation and the limit for $n \rightarrow \infty$:

$$[\frac{\varepsilon}{1 + \varepsilon} \lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon)] \leq \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{|X_n|}{1 + |X_n|} \right] = 0$$

where the last “=0” is by hypothesis. Since ε is chosen arbitrarily then

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \varepsilon) = 0$$

so we have proven the converse implication. \square

We said before that X_n converges in probability to X if the metric $d(X_n, X) \xrightarrow{n \rightarrow \infty} 0$. The last proof refers to the function $f(x) = \frac{|x|}{1+|x|}$ but the same holds for every bounded, non decreasing, continuous function g on \mathbb{R}_+ with $g(0) = 0, g(x) > 0 \forall x > 0$. For example, $g(x) = |x| \wedge 1$ (which is similar to the metric we established before).

1.5.5 Convergence in L^p spaces

Let's start with the proper definition.

Definition 1.5.4

A sequence $(X_n)_n$ of real-valued random variables is said to be convergent to the real-valued random variable X in L^p if, for each n , $X_n \in L^p$, $X \in L^p$ and

$$\lim_{n \rightarrow \infty} \mathbb{E}|X_n - X|^p = 0.$$

Revise with Kotatsu!

Recall the p -norm of X :

$$\|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}.$$

Then L^p is a normed vector space if we identify all the random variables that are almost surely equal as one single random variable: this means that the norm works up to an equivalence class (if A and B are equal a.s. then for us they are the same random variable as long as we work in L^p). In this case, the L^p -convergence is to be written as

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0.$$

Remark

What happens if we have two different sequences converging to the same limit in L^p ? If $(X_n)_n$ converges to X in L^p then X is unique (up to equivalence).

Proof

Consider

$$\begin{aligned} X_n &\xrightarrow{L^p} X \\ X_n &\xrightarrow{L^p} Y \end{aligned}$$

Revise with Kotatsu!

Recall the Minkowsky's inequality

$$\|X - Y\|_p \leq \|X - X_n\|_p + \|X_n - Y\|_p \xrightarrow{n \rightarrow 0} 0$$

But both $\|X - X_n\|_p$ and $\|X_n - Y\|_p$ tend to 0 as $n \rightarrow \infty$, we have that

$$X = Y \text{ a.s.}$$

□

So, since they are the same random variable in L^p then they differ only for elements of measure 0.

Remark

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X.$$

This is quite strong! Convergence in L^p is stronger! I wish the same could be said of my will to live.

Proof

We use Markov's inequality:

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \left(\frac{1}{\varepsilon}\right)^p \mathbb{E}|X_n - X|^p \xrightarrow{n \rightarrow 0} 0$$

□

But is the converse true? NO IT IS NOT IT IS NOT TRUE IT IS A LIE. Consider this counterexample:

Example 1.5.3

We show that the converse of what it is stated in the previous remark does NOT hold in general. Consider $\Omega = [0, 1]$, $\mathcal{H} = \mathcal{B}_{[0,1]}$, $\mathbb{P} = \lambda$. Consider $(X_n)_n$ a sequence of indicators of the intervals

$$(0, 1], \left(0, \frac{1}{2}\right], \left(\frac{1}{2}, 1\right], \left(0, \frac{1}{3}\right], \left(\frac{1}{3}, \frac{2}{3}\right], \left(\frac{2}{3}, 1\right], \dots$$

Note that we already proved that this sequences converges to 0 in probability. Note, moreover, that $X_n \xrightarrow{L^p} 0$ because $\mathbb{E}|X_n|$ forms the sequence

$$(1 + 0 \cdot p, 1 \cdot p + 0 \cdot q, \dots) = \left(1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \dots\right)$$

(remember that our \mathbb{P} is the Lebesgue measure, which is the length of the interval). This sequence goes to zero, clearly.

Now let's modify the sequence of random variables:

$$(\hat{X}_n)_n = (X_1, 2X_2, 2X_3, 3X_4, 3X_5, 3X_6, \dots)$$

and evaluate

$$\mathbb{P}(\hat{X}_n > \varepsilon) = \mathbb{E}|X_n|$$

because for indicators our probability coincides with the expectation. Now take the limit and see whether $(\hat{X}_n) \xrightarrow{\mathbb{P}} 0$. However, if we do the calculations we get

$$\mathbb{E}|\hat{X}_n| = 1 \quad \forall n.$$

Hence

$$\hat{X}_n \not\xrightarrow{L^1} 0$$

so there exists a p for which \hat{X}_n does not converge to 0.

Theorem 1.5.7

Characterization theorem for convergence in L^1 .

Let $(X_n)_n$ be a sequence of real-valued random variables then the following are equivalent:

- ① $(X_n)_n$ converges in L^1 ;
- ② $(X_n)_n$ converges in probability and is uniformly continuous;
- ③ $(X_n)_n$ is Cauchy for the L^1 -convergence:

$$\lim_{m,n \rightarrow \infty} \mathbb{E}|X_m - X_n| = 0.$$

Proposition 1.5.9

$$X_n \xrightarrow{L^1} X \implies \lim_{n \rightarrow \infty} \mathbb{E}X_n Y - \mathbb{E}XY$$

And this is true for every bounded random variable Y .

So it is not true in generale, but requires boundedness. Quite a strict need. See what I did there?

Remark

Setting $Y = 1$ a.s. we get that

$$X_n \xrightarrow{L^1} X \implies \lim_{n \rightarrow \infty} \mathbb{E}X_n = \mathbb{E}X.$$

1.5.6 Weak convergence and convergence in distribution

Definition 1.5.5

The sequence of probability measures $\mu_1, \mu_2, \mu_3, \dots$ on \mathbb{R} is said to **converge weakly** to μ if

$$\lim_{n \rightarrow \infty} \underbrace{\mu_n f}_{\text{integrals!!}} = \mu f \quad \forall f \in \mathbb{C}_b$$

where \mathbb{C}_b is the set of continuous and bounded functions from \mathbb{R} to \mathbb{R} .

So we are talking about convergence of probability measures and this is done through integrals. If the sequence of integrals converge towards a limiting integral then we say that the convergence happens for the *sequence of measures*: this shouldn't come as a surprise, since YOU should remember that measures are what defines integrals. But remember that random variables are ultimately characterized by their distributions (which are nothing else but probability measures!) so there must be some kind of connection between the distributions converging and their random variables doing the same thing...

Consider a sequence $(X_n)_n$ of random variables characterized by their sequence of distributions $(\mu_n)_n$.

Definition 1.5.6

The sequence $(X_n)_n$ is said to **converge in distribution** to a limiting random variable X with distribution μ if

$$(\mu_n)_n \xrightarrow{\text{weak}} \mu$$

that is to say if

$$\lim_{n \rightarrow \infty} \mathbb{E}f \circ X_n = \mathbb{E}f \circ X \quad \forall f \in \mathbb{C}_b.$$

Note the notation!

We usually denote convergence in distribution as

$$X_n \xrightarrow{d} X.$$

Remark

$$X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{d} X$$

So we see: convergence in distribution is the weakest of them all. It is related to the behavior in the limit of the probability distribution... and just having the same probability distribution is not saying much about a random variable. Imagine that we have $Y \sim N(0, 1)$ and $Z = -Y$. Z is still a normal random variable, but clearly $Y(\omega) = a \implies Z(\omega) = -a$ so they are very different! But there is a special case...

Remark

If $X_n \xrightarrow{d} X$ and $X = a$ a.s. (degenerate random variable¹). Then

$$X_n \xrightarrow{\mathbb{P}} X.$$

¹"Yeah I know some degenerates "random variables" around here."

What the actual fuck? Now convergence in distribution implies convergence in probability? I will end my own life RIGHT NOW!

Proof

Let us choose

$$f(x) = \frac{|x - a|}{1 - |x - a|}.$$

Note that $f \in C_b$. So we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \frac{|X_n - a|}{1 - |X_n - a|} = \mathbb{E} f \circ X = \mathbb{E} \frac{|a - a|}{1 - |a - a|} = 0.$$

Lol! Now

$$X_n \xrightarrow{\mathbb{P}} a.$$

□

Proposition 1.5.10

If μ and ν are probability measures on \mathbb{R} and

$$\mu f = \nu f \quad \forall f \in C_n \implies \mu = \nu.$$

Definition 1.5.7

Take a distribution function $C(x) = \mathbb{P}(X \leq x)$ and construct:

$$q(u) = \inf \{x \in \mathbb{R} : C(x) > u\}.$$

This is known as **quantile function**.

So the quantile functions finds the first point at which the distribution function exceeds the level u . This is just the compositional inverse of C , if we think about it.

Proposition 1.5.11

Convergence of distribution functions and quantile functions.

The following points are equivalent:

① $(\mu_n)_n \xrightarrow{\text{weak}} \mu$;

- ② take $C_n(x) = \mathbb{P}(X_n \leq x)$ (distribution function: since it depends on n we get a sequence of functions with respect to the index n). We have

$$C_n(x) \rightarrow C(x) = \mathbb{P}(X \leq x)$$

for each point of continuity x of the distribution function C ;

- ③ $q_n(u) \rightarrow q(u)$ for every point of continuity u of the quantile function q .

Theorem 1.5.8

Consider a sequence $(\mu_n)_n$ of probability measures such that

$$\mu_n \xrightarrow{\text{weak}} \mu,$$

where μ is a further probability measure. The above convergence is equivalent to the existence of a probability space

$$(\Omega', \mathcal{H}', \mathbb{P}')$$

and some random variables

$$Y_1, Y_2, \dots, Y$$

on it such that the distribution of Y_i is μ_i , $i \in \mathbb{N}$ and of Y is μ and we have that

$$Y_n \xrightarrow{\text{a.s.}} Y$$

on $(\Omega', \mathcal{H}', \mathbb{P}')$.

This theorem is really important. We start with a sequence of weakly convergent probability measures, so we already know that on our initial space $(\Omega, \mathcal{H}, \mathbb{P})$ there are some random variables X_1, X_2, \dots, X with those (convergent) measures. These random variables are still defined on $(\Omega, \mathcal{H}, \mathbb{P})$, but this theorem tells us that there exists *another space* $(\Omega', \mathcal{H}', \mathbb{P}')$ with other random variables Y_1, Y_2, \dots, Y that are actually convergent almost surely! So if we have convergence in distribution and we want convergence almost surely we can have it, but at the cost of working in a different probability space.

File: [mfw.jpg](#) (711 KB, 1170x1061)



Anonymous 08/07/24(Wed)17:24:38 No.16313511 ► [>>16313525](#)

>be me
>converging in distribution in my probability space $(\Omega, \mathcal{H}, \mathbb{P})$,
not a big deal
>meet chick in uni
>ask her out
>got rejected because she only dates almost surely converging
chads
>depression.jpg
>finally decide to get almost sure convergence to cope
>starting to realize that my sequences $(X_n)_n$ of random
variables converging to X seem different now
>mfw now I am in $(\Omega', \mathcal{H}', \mathbb{P}')$
>"anon... what the fuck is that probability space?"

>i am now the laughing stock of my uni

is it over for me bros?

Anonymous 08/07/24(Wed)17:34:13 No.16313525 ►

[>>16313511 \(OP\)](#)

absolutely brutal. As a 28 year old virgin who knows its over for me, this still hurts reading

Figure 1.12: Anon realizes he was deceived by probability theory⁴⁵.

Corollary

Skorokhod representation.

$(X_n) \xrightarrow{d} X$ if and only if there exist random variables Y_1, Y_2, \dots, Y on some probability space such that

$$Y_n \stackrel{d}{=} X_n \quad \forall n, Y \stackrel{d}{=} x \text{ and } Y_n \xrightarrow{\text{a.s.}} Y.$$

I could have stuffed another Metal Gear Solid joke here with the similarity between Skorokhod and Shagohod but I won't. Maybe. Anyway, the greentext in figure 1.12 is enough for now.

Proposition 1.5.12

If $X_n \xrightarrow{d} X$, the following are equivalent:

- ① $(X_n)_n$ is uniformly integrable;
- ② X_n and X are integrable and

$$\mathbb{E}|X_n| \rightarrow \mathbb{E}|X|.$$

But first let's introduce Fourier transform because... because fuck you. Let f_{n_n} be the sequence of Fourier transforms of $(\mu_n)_n$, $\forall n$:

$$f_n(r) = \int_{\mathbb{R}} e^{irx} \mu_n(dx), \quad r \in \mathbb{R}.$$

Remember that this is a function of r !

Theorem 1.5.9

Let $(\mu_n)_n$, our sequence of probability measures, be weakly convergent if and only if

$$\lim_n f_n(r) = f(r) \quad \forall r \in \mathbb{R}$$

and f is continuous at 0. Moreover, f is the transform of a probability measure μ on \mathbb{R} and μ is the weak limit of $(\mu_n)_n$.

With this notion we study weak convergence by means of the Fourier transform. Why would we want to do this? I am seriously at a loss for words. But dear Professor Polito comes to help! Look back at the equation for the Fourier transform of a measure μ :

$$f(r) = \int_{\mathbb{R}} e^{irx} \mu(dx), \quad r \in \mathbb{R}$$

But this is an integral, which is... an expected value! consider x as being the random variable X and write

$$f(r) = \mathbb{E}e^{irX}.$$

This is called the **characteristic function of X** or the **fourier transform of its measure**.

Corollary

The sequence $(X_n)_n \xrightarrow{d} X$ if

$$\lim_n \mathbb{E}e^{irX_n} = \mathbb{E}e^{irX} \quad r \in \mathbb{R}.$$

But up to now we didn't say anything about the limiting random variables... to do so we need to introduce one of the main results in asymptotic analysis: the law of large numbers.

1.5.7 The law of large numbers

“Large numbers” always mean “what happens in the limit”. Let's start with the basic theorem.

Theorem 1.5.10

Consider a sequence of real valued, pairwise independent random variables X_1, X_2, \dots with finite common expectation and variance

$$\mathbb{E}X_n = a, \quad \text{Var } X_n = b.$$

Then:

$$\textcircled{1} \quad \bar{x}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{L^2} a;$$

$$\textcircled{2} \quad \bar{x}_n \xrightarrow{\mathbb{P}} a;$$

⁴⁵Archived at <https://boards.4chan.org/sci/thread/16313511>

$$③ \bar{x}_n \xrightarrow{\text{a.s.}} a.$$

Huh. If you're anything like me this doesn't sound at all like the weak and strong laws of large numbers I've studied in the Economics degree. But again, what do I know. I'm sure we will find some surprising connections. Also we only need to prove the first and third point, because the second is implied by the third. Why does the theorem list it as the *third* point, though? What are you trying to do, *keeping me on the edge*?

Proof

① Define

$$S_n = n\bar{X}_n$$

and

$$\mathbb{E}S_n = na, \text{Var } S_n = nb$$

because of hypothesis of independence. Then

$$\mathbb{E}\bar{X}_n = a, \text{Var } \bar{X}_n = \frac{b}{n}.$$

So if I look at

$$\mathbb{E}|\bar{X}_n - a|^2 = \text{Var } \bar{X}_n \xrightarrow{n \rightarrow \infty} 0$$

because the variance is a constant divided by n . Hence

$$\bar{X}_n \xrightarrow{L^2} a$$

② This comes from Markov's inequality (which we already proved)...

③ Assume $X_n \geq 0$. This doesn't cause loss of generality because we can apply what we are going to prove to the positive and negative part of X_n (which are both positive!).

Define $N = (n_k)_{k \in \mathbb{N}^*}$ such that $n_k = k^2$. By Chebyshev's inequality we get

$$\mathbb{P}(|\bar{X}_n - a| > \varepsilon) \leq \frac{b}{\varepsilon^2 k^2}$$

since we are working on the subsequence N where the variance of the random variables \bar{X}_n is not $\frac{b}{n}$, but it is $\frac{b}{\text{value of subsequence}}$ which is k^2 . Now sum and multiply by ε^2 to get

$$\varepsilon^2 \sum_{n \in N} \mathbb{P}(|\bar{X}_n - a| > \varepsilon) \leq \sum_{k=1}^{\infty} \frac{b}{k^2} < \infty \quad \varepsilon > 0$$

since the latter is a geometric series. But this tells us that $\sum_{n \in N} \mathbb{P}(|\bar{X}_n - a| > \varepsilon)$ is a finite number and therefore... we can apply Borel-Cantelli lemma!

Revise with Kotatsu!

Borel Cantelli lemma:

$$\sum_n \mathbb{P}(H_n) < +\infty \implies \sum_n \mathbb{1}_{H_n} < +\infty \quad \text{a.s.}$$

and in particular if

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty \quad \forall \varepsilon > 0.$$

then

$$X_n \xrightarrow{\text{a.s.}} X.$$

So

$$\bar{X}_n \rightarrow a$$

along N . Let us call Ω_0 the almost sure set on which the convergence takes place and $\forall \omega \in \Omega_0$ we apply the following lemma (we have already proven)

Revise with Kotatsu!

If we have a sequence of positive numbers $(x_n)_n$ and consider the empirical sum $\bar{x}_n = \frac{\sum_{i=1}^n x_i}{n}$ and let $N = (n_k) \subset \mathbb{N}^*$ with $\lim_k \frac{n_{k+1}}{n_k} = r > 0$ (asymptotic linearity) and $(\bar{x}_n)_n$ converges along N to x then

$$\frac{x}{r} \leq \liminf \bar{x}_n \leq \limsup \bar{x}_n \leq rx.$$

Note that

$$\frac{(k+1)^2}{k^2} \xrightarrow{k \rightarrow \infty} 1$$

and hence $\forall \omega \in \Omega_0$ we have that

$$\lim_{n \rightarrow \infty} \bar{X}_n(\omega) = a \quad \forall \omega \in \Omega_0$$

with probability 1... which means, almost sure convergence!

□

This one was for sequence of random variables with finite variance and expectations... But what if both are infinite?

Proposition 1.5.13

Let $(X_n)_n$ be a sequence of positive independent and identically distributed (i.i.d.) random variables with

$$\mathbb{E}X_1 = +\infty.$$

Consider also a further random variable X distributed as X_1 (which means that they also have the same expectation). Then

$$\bar{X}_n \xrightarrow{\text{a.s.}} \infty.$$

This shouldn't be surprising, since we are given a sequence of positive random variables with infinite expectation, that means that there is a lot of the probability mass resides in the tails. This causes the expectation to diverge when doing the integral. To prove this fact we must rely on the previous result.

Proof

We know that this behaviour is caused by the tail of the distribution, so we are going to truncate the distribution.

Fix $b \in \mathbb{R}$ and let $Y_n = X_n \wedge b$. So X_n can have infinite support, while Y_n is definitely bounded with the support finishing at b (**truncated random variable**). Let also

$$\bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n}$$

and note that

$$\mathbb{E}Y_n = \mathbb{E}(X_n \wedge b) < +\infty.$$

By the previous theorem we know that

$$\bar{Y}_n \xrightarrow{\text{a.s.}} \mathbb{E}(X \wedge b).$$

The second result that we need is an ordering between X_n and Y_n . Which is the largest? For each n we have that $X_n \geq Y_n$ and this further implies that $\liminf_n \bar{X}_n \geq \lim_n \bar{Y}_n = \mathbb{E}(X \wedge b)$ almost surely. This must be true for any choice of b , even if $b \rightarrow \infty$ and thus

$$\mathbb{E}(X \wedge b) \xrightarrow{b \rightarrow \infty} \mathbb{E}X = +\infty$$

using the monotone convergence theorem. That is,

$$\liminf_n \bar{X}_n = +\infty \text{ a.s.}$$

Which basically means, since \bar{X}_n is a non decreasing sequence, that

$$\bar{X}_n \xrightarrow{\text{a.s.}} +\infty$$

□

There is a general theorem that puts together the last two theorems that we have seen so far:

Theorem 1.5.11

Law of large numbers:

Let $(X_n)_n$ be a sequence of pairwise independent random variables with the same distribution as X . If $\mathbb{E}X$ exists (infinite values are admitted!) then

$$\bar{X}_n \xrightarrow{n \rightarrow \infty} \mathbb{E}X.$$

This is an almost sure result and that's why we often call this theorem the **strong law of large numbers** (SLLN). Here are two useful inequalities:

- **Chebyshev's inequality:** here we are given a sequence of random variables X_n and build up the partial sums S_n . Here we must assume that $(X_n)_n$ are i.i.d. and such that $\mathbb{E}X_1 = 0$.

$$\varepsilon^2 \mathbb{P}(|S_n| > \varepsilon) \leq \text{Var } S_n = \underbrace{\mathbb{E}S_n^2}_{\text{second moment!}}$$

so the variance gives us an upper bound of the probability $\mathbb{P}(|S_n| > \varepsilon)$. If we don't have that $\mathbb{E}X_1 = 0$ then $\text{Var } S_n \leq \mathbb{E}S_n^2$ (the second moment is usually larger than the variance... still a good upper bound tho).

- **Kolmogorov's inequality:** let $(X_d)_n$ be a sequence of independent random variables such that $\mathbb{E}X_n = 0 \forall n$. Then $\forall a \in (0, \infty)$

$$a^2 \mathbb{P}\left(\max_{k \leq n} |S_k| > a\right) \leq \text{Var } S_n$$

We are interested in the times from 1 to n : the interesting thing is that the maximum value of the random walk is bounded by something that happens in time n .

We will prove this last inequality.

Proof

Fix $a > 0$, $n \geq 1$ and define

$$N(\omega) = \inf \{k \geq 1 : |S_k(\omega)| > a\} \quad \forall \omega.$$

So for that specific ω we consider the first time for which the random walk exceeds the value a . Of course this time index k depends on ω , so $N(\omega)$ is a random variable. Observe a couple of facts:

1) $N(\omega) = k \iff |S_k(\omega)| > a \text{ and } S_j(\omega) \leq a \quad \forall j < k$. Then $\mathbb{1}_{N=k}$ is a function of (X_1, X_2, \dots, X_n) because I need the knowledge of X up to k to know whether $N = k$.

2) For $k < n$ we define the following random variables:

$$U = S_k \mathbb{1}_{N=k} \quad \text{and} \quad V = S_n - S_k.$$

We know that U depends on the first k values of X_n , while V depends only on $X_{k+1}, X_{k+2}, \dots, X_n$. But this means that U and V are functions of independent random variables, so $U \perp\!\!\!\perp V$. Hence

$$\mathbb{E}UV = \mathbb{E}U \mathbb{E}V.$$

$=0$ because $\mathbb{E}X_n=0$

Therefore

$$\mathbb{E}S_k \mathbb{1}_{\{N=k\}}(S_n - S_k) = 0.$$

3) Consider S_n^2 :

$$S_n^2 = [S_k + (S_n - S_k)]^2 \geq S_k^2 + 2S_k(S_n - S_k)$$

The inequality is what I get if I neglect the negative term. So I am bounding what happens at time n with what happens at time k . Note that $|S_k|^2 > a^2$ when $\{N = k\}$ holds. Then

$$\mathbb{E}S_n^2 \mathbb{1}_{N=k} \leq \underbrace{a^2 \mathbb{E} \mathbb{1}_{\{N=k\}}}_{a^2 \mathbb{P}(N=k)} + \underbrace{2 \mathbb{E} S_k (S_n - S_k) \mathbb{1}_{\{N=k\}}}_{=0}$$

and summing both sides for $k \leq n$ we get

$$a^2 \mathbb{P}(N \leq n) \leq \mathbb{E}S_n^2 \mathbb{1}_{\{N \leq n\}} \leq \mathbb{E}S_n^2 = \underbrace{\text{Var } S_n}_{\text{because of independence}}$$

Hence

$$a^2 \mathbb{P}(N \leq n) \leq \text{Var } S_n$$

But we know that

$$\mathbb{P}(N \leq n) = \mathbb{P}(\max_{k \leq n} |S_k| > a)$$

So our proof is finished. \square

Proposition 1.5.14

Let $(X_n)_n$ be independent and with mean 0. If

$$\sum_n \text{Var } X_n < +\infty \implies \sum_n X_n < +\infty \text{ a.s.}$$

Proposition 1.5.15

Let $(X_n)_n$ be a bounded sequence of i.i.d. random variables. If

$$\sum_n (X_n - a_n) < +\infty \text{ a.s.}$$

for some sequence $(a_n)_n \subset \mathbb{R}$ then

$$\sum_n \text{Var } X_n < +\infty.$$

Theorem 1.5.12

Kolmogorov's 3-series theorem.

Let $(X_n)_n$ be independent random variables. We have

$$\sum_n X_n < +\infty \text{ a.s.}$$

if and only if the following 3 series are convergent:

1. $\sum_n \mathbb{P}(X_n \neq Y_n)$ where $Y_n = X_n \mathbb{1}_{\{|X_n| < b\}}$;
2. $\sum_n \mathbb{E}Y_n$;
3. $\sum_n \text{Var } Y_n$.

So: remember that expectation is a number, so the LLN tells us that \bar{X}_n (a random variable) tends to $\mathbb{E}X$ (a degenerate random variable): our variability gets lots as $n \rightarrow \infty$. This affects many characteristics of the random variable...

1.5.8 Central limit theorem

There are different versions of the Central limit theorem and we have probably already studied some of those: the most common in undergraduate courses is the **De Moivre-Laplace Central limit Theorem**.

Theorem 1.5.13

Let $(X_i)_i$ be a sequence of i.i.d. real-valued random variables with finite mean $\mathbb{E}X_1 = a$ and finite variance $\text{Var } X_1 = b$. Then

$$Z_n = \frac{S_n - na}{\sqrt{nb}} \xrightarrow{d} Z \sim N(0, 1)$$

The usual way to prove this theorem is relying on characteristic function. If $a = 0, b = 1$ then $Z_n = \frac{S_n}{\sqrt{n}}$ where Z_n is the sum of “small” independent random variables $X_{n,1} = \frac{X_1}{\sqrt{n}}, X_{n,2} = \frac{X_2}{\sqrt{n}}, \dots, X_{n,n} = \frac{X_n}{\sqrt{n}}$. All of them are “small” with respect to their original random variable X_n because we are dividing by \sqrt{n} : they get smaller and smaller as $n \rightarrow \infty$ but overall, for each n , Z_n has a stable mean 0 and variance 1. The general idea behind the CLT is exactly this: we divide all of these variables for the “right amount” so that the limit of our sequence of random variables stays stable and does not become a degenerate limit.

Definition 1.5.8

A **triangular array** (Professor Polito calls it “a nice object“) is an infinite matrix with a special structure:

$$[X_{nj}] = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \dots \\ X_{21} & X_{22} & X_{23} & \dots \\ X_{31} & X_{32} & X_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where the entries are real-valued random variables (so it is actually an infinite random metric). Furthermore, for each n there exists an integer k_n such that $X_{nj}, j > k_n$. The sequence of indices $(k_n)_n$ is increasing towards ∞ .

So the random variables are equal to 0 to the right hand side of the index k_n for each row. The first row has k_1 , the second row is composed by $k_2 > k_1$ (so there are more non-zero random variables than in row 1) and so on. So it’s not really a triangular matrix (there’s no rule on *how much* k_n is larger than k_{n-1}) but it is “basically” triangular.

Now define

$$Z_n = \sum_j X_{nj} \quad (j = 1, 2, \dots, k_n).$$

Remember that even if we are summing infinite random variables only a finite number of them are non-zero. Now consider the assumption that the random variables on each row are independent among them (there *may* be dependence with random variables in other rows.). With this structure in mind we can formalize the following formulation of the CLT:

Theorem 1.5.14

Lyapunov Central limit Theorem:

Suppose $\mathbb{E}X_{nj} = 0 \forall n, j$ and $\text{Var } Z_n = 1 \forall n$ and $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$. Then

$$Z_n \xrightarrow{d} Z \sim N(0, 1).$$

Here we have replaced the conditions on the random variables with the condition $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$ which allows us to use the structure of the triangular array. To prove this theorem we need the following lemma:

Lemma 1.5.3

Lindeberg’s lemma: Let (Y_1, Y_2, \dots, Y_k) be independent random variables with mean zero and let $S = \sum_{j=1}^k Y_j$. Let us further assume that $\text{Var } S = 1$. Let f be a function which can be differentiated 3

times and let f', f'', f''' be bounded and continuous and such that

$$|f'''| \leq c, \quad c \in \mathbb{R}_+.$$

Then for $Z \sim N(0, 1)$

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ Z| \leq c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

This means that if the hypotheses are met then the distance between S and a normal random variable is bounded so we can approximate S by means of a normal random variable. We will prove this lemma.

Proof

Let Z_1, \dots, Z_k be independent normal random variables with mean $\mathbb{E}Z_j = \mathbb{E}Y_j = 0$ for $j = 1, \dots, k$ and variance $\text{Var } Z_j = \text{Var } Y_j$ for $j = 1, \dots, k$. Then construct

$$T = \sum_{j=1}^k Z_j \sim N\left(0, \sum_{j=1}^k \text{Var } Z_j = \sum_{j=1}^k \text{Var } Y_j = 1\right).$$

So we know that T is distributed as Z (they are both $N(0, 1)$) so $T \stackrel{d}{=} Z$ and since we are using the expectation of Z we can replace $\mathbb{E}f \circ Z$ with $\mathbb{E}f \circ T$. So the Linderberg's lemma becomes

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ T| \leq c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

which we want to prove, to exploit the structure of T . Define now the random variables V_1, V_2, \dots, V_k as follows:

$$\begin{aligned} V_1 &\text{ s.t. } S = V_1 + Y_1 \\ V_2 &\text{ s.t. } V_1 + Z_1 = V_2 + Y_2 \\ &\vdots \\ V_j &\text{ s.t. } V_j + Z_j = V_{j+1} + Y_{j+1}, \quad 1 \leq j < k \\ &\vdots \\ V_k &\text{ s.t. } V_k + Z_k = T. \end{aligned}$$

Note that

$$\begin{aligned} V_1 &= Y_2 + Y_3 + \dots + Y_k \\ V_2 &= Z_1 + Y_3 + \dots + Y_k \\ V_3 &= Z_1 + Z_2 + Y_4 + \dots + Y_k \end{aligned}$$

so the Y get replaced by the Z one at the time in the V . We can now focus on the following expression:

$$\begin{aligned} f \circ S - f \circ T &= f(V_1 + Y_1) - f(V_k + Z_k) \\ &= f(V_1 + Y_1) + f(V_2 + Y_2) - \underbrace{f(V_2 + Y_2)}_{f(V_1 + Z_1)} + \underbrace{f(V_3 + Y_3) - f(V_2 + Y_2)}_{f(V_2 + Z_2)} + \\ &\quad + \dots + f(V_k + Z_k) \\ &= \sum_{j=1}^k f(V_j + Y_j) - \sum_{j=1}^k f(V_j + Z_j). \end{aligned}$$

Now take the expectation and the absolute value:

$$\begin{aligned} |\mathbb{E}f \circ S - \mathbb{E}f \circ T| &= \left| \sum_{j=1}^k \mathbb{E}f(V_j + Y_j) - \sum_{j=1}^k \mathbb{E}f(V_j + Z_j) \right| \\ &\leq \sum_{j=1}^k |\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \end{aligned}$$

and now we only need to prove that

$$|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \leq c\mathbb{E}|Y_j|^3.$$

Let's write the Taylor formula for this function:

$$f(v+x) = f(v) + f'(v)x + \frac{1}{2}f''(v)x^2 + R_2(v, x)$$

where

$$\begin{aligned} R_2(v, x) &= \frac{1}{2} \int_v^{v+x} (v+x-t)_2 f''(t) dt \\ &\leq \frac{1}{2} c \int_v^{v+x} (v+x-t)_2 dt = c \frac{x^3}{6} \end{aligned}$$

so that

$$|R_2(v, x)| \leq \frac{c}{6} |x|^3.$$

We now have

$$\begin{aligned} f(V_j + Y_j) &= f(V_j) + f'(V_j)Y_j + \frac{1}{2}f''(V_j)Y_j^2 + R_2(V_j, Y_j) \\ f(V_j + Z_j) &= f(V_j) + f'(V_j)Z_j + \frac{1}{2}f''(V_j)Z_j^2 + R_2(V_j, Z_j) \end{aligned}$$

and subtract side by side:

$$f(V_j + Y_j) - f(V_j + Z_j) = (Y_j - Z_j)f'(V_j) + \frac{1}{2}f''(V_j)(Y_j^2 - Z_j^2) + R_2(V_j, Y_j) - R_2(V_j, Z_j).$$

Now take the expectation

$$\begin{aligned} \mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j) &= \frac{1}{2} \mathbb{E}f''(V_j) (\underbrace{\mathbb{E}Y_j^2 - \mathbb{E}Z_j^2}_{=0 \text{ since they have same variance}}) + \mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)] \\ &= \mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)]. \end{aligned}$$

Now take the absolute value

$$\begin{aligned} |\mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)]| &\leq \mathbb{E}|[R_2(V_j, Y_j)]| + \mathbb{E}|[R_2(V_j, Z_j)]| \\ &\leq \frac{c}{6} (\mathbb{E}|Y_j|^3 + \mathbb{E}|Z_j|^3) \end{aligned}$$

so that

$$|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \leq \frac{c}{6} (\mathbb{E}|Y_j|^3 + \mathbb{E}|Z_j|^3).$$

Recall that $Z_j \sim N(0, b^2)$ where $b^2 = \mathbb{E}Y_j^2$. We know that

$$\mathbb{E}|Z_j|^3 = b^3 \sqrt{\frac{8}{\pi}} \leq 2b^3$$

and we also have

$$b = (\mathbb{E}Y_j^2)^{\frac{1}{2}} \leq (\mathbb{E}|Y_j|^3)^{\frac{1}{3}}$$

Because L^2 norm is less or equal than L^3 norm (revise inclusions in L^p -spaces for different values of p). But this last inequality is equivalent to

$$b^3 \leq \mathbb{E}|Y_j|^3$$

which leads to

$$\mathbb{E}|Z_j|^3 \leq 2b^3 \leq 2\mathbb{E}|Y_j|^3.$$

Finally we get

$$\begin{aligned} |\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| &\leq \frac{c}{6} (\mathbb{E}|Y_j|^3) \\ &= \frac{c}{6} 3\mathbb{E}|Y_3|^3 \\ &= \frac{c}{2}\mathbb{E}|Y_3|^3 \leq c\mathbb{E}|Y_3|^3. \end{aligned}$$

□

This was horrible, horrible. Truly an horrible experience and honestly useless proof. And we still have to prove Lyapunov's theorem.

Proof

Lyapunov's CLT. Recall

$$Z_n = \sum_j X_{nj} \quad Z \sim N(0, 1).$$

We are interested in evaluating the characteristic function.

$$\begin{aligned} e^{irZ_n} &= \cos rZ_n + i \sin rZ_n \\ e^{irZ} &= \cos rZ + i \sin rZ. \end{aligned}$$

Consider now

$$\begin{aligned} |\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}| &= |(\mathbb{E} \cos rZ_n - \mathbb{E} \cos rZ) + i(\mathbb{E} \sin rZ_n - \mathbb{E} \sin rZ)| \\ &\leq \|\mathbb{E} \cos rZ_n - \mathbb{E} \cos rZ\| + \underbrace{\|i\|}_1 \|\mathbb{E} \sin rZ_n - \mathbb{E} \sin rZ\|. \end{aligned}$$

By applying the above lemma we obtain

$$|\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}| \leq \sum_j |r|^3 \mathbb{E}|X_{nj}|^3 + \sum_j |r|^3 \mathbb{E}|X_{nj}|^3$$

and this is possible since both sine and cosine are differentiable three times and they are bounded by 1 (so our $c = 1$). Now, according to the hypotheses of Lyapunov's theorem, we need to take the limit considering the hypothesis that $\lim_{n \rightarrow \infty} \sum_j \mathbb{E}|X_{nj}|^3 = c$ and we obtain the claim. This theorem applies to all triangular arrays which include the one in the CLT. □

WHY.

1.6 Conditional expectations

Are you comfortable with the concept of expectation? Well brace yourself buddy because we are going to fuck it up and reveal that normal expectations are just a particular case of conditional expectations.

Consider our dear old probability space $(\Omega, \mathcal{H}, \mathbb{P})$, a $\overline{\mathbb{R}}$ -valued random variable X , a sub- σ -algebra $\mathcal{F} \in \mathcal{H}$ where \mathcal{F} is the space of \mathcal{F} -measurable random variables taking values in $\overline{\mathbb{R}}$. Remember that \mathcal{F} can be considered the body of information revealed by our random variable. Consider now the event H . Fix $\omega \in \Omega$ and pretend that our only knowledge of ω is that of the $\omega \in H$. First of all, remember that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \int_A \mathbb{1}_B d\mathbb{P}$$

Evaluate our best estimate for $X(\omega)$, which may very well be the average of X over H :

$$\frac{1}{\mathbb{P}(H)} \int_H X(\omega) \mathbb{P}(d\omega) = \frac{1}{\mathbb{P}(H)} \mathbb{E}X \mathbb{1}_H = \mathbb{E}_H X.$$

If $\mathbb{P}(H) = 0$ then we allow any value for $\mathbb{E}_H X$.

Note the notation!

We usually denote that quantity as $\mathbb{E}(X|H)$ but Çinlar denotes it as $\mathbb{E}_H X$.

Another interesting way to look at this question is the following: conditional probability basically gives us a new probability measure while leaving the rest of the “ingradients” of the probability triplet unchanged. And it makes quite sense: we change the way we weigh a certain event. So if we want, for instance, to condition our probability measure to the event B our probability space $(\Omega, \mathcal{H}, \mathbb{P})$ becomes $(\Omega, \mathcal{H}, \mathbb{P}(\cdot|B))$. This also means that our expectation $\mathbb{E}X = \int_{\Omega} X d\mathbb{P}$ becomes $\mathbb{E}_B X = \int_{\Omega} X d\mathbb{P}(\cdot|B)$.

Remark

This best estimate is “best” in the same sense as the expectation of X is the best estimate of X when we do not know anything about ω . The unconditional case is retrieved if $H = \Omega$.

Remark

The quantity $\mathbb{E}_H X$ is called **conditional expectation of X given the event H** .

Up to now, nothing new. Let’s go a step further. Now suppose that \mathcal{F} is generated by a measurable partition $(H_n)_n$ of Ω . Fix $\omega \in \Omega$ and consider the problem of building an estimate of the random variable $X(\omega)$ with only the body of information given by the partition/sub- σ -algebra \mathcal{F} : we could construct a “best estimate” of $X(\omega)$ for every event H_n :

$$\bar{X}(\omega) = \sum_n \mathbb{E}_{H_n} X \mathbf{1}_{H_n}(\omega). \quad (\text{•})$$

So for every specific ω only one term of the sum will be different from 0: if ω_0 belongs in H_0 then only $\mathbb{E}_{H_0} X$ in the sum will be non-zero and $\bar{X}(\omega_0) = \mathbb{E}_{H_0} X$. So in this sense we can summarize all the information available. So $\bar{X}(\omega)$ is a random variable that we will call **conditional expectation of X given \mathcal{F}** and we will write

$$\bar{X} = \mathbb{E}_{\mathcal{F}} X \quad \text{or} \quad \mathbb{E}[X|\mathcal{F}].$$

So expectation of a random variable is a *number*, while the conditional expectation with respect to a sub- σ -algebra is a *random variable*.

Example 1.6.1

Imagine we have $\Omega = (0, 1)$, $\mathcal{H} = \mathcal{B}_{(0,1)}$, $\mathbb{P} = \lambda$. Consider the partition of Ω given by $\mathcal{F} = \sigma(\{H_1, H_2, H_3\})$. The wiggly line is the realization of our process $X(\omega)$.

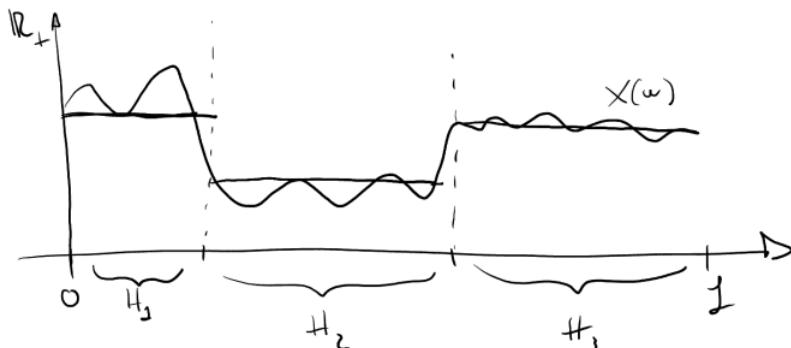


Figure 1.13: ♪ I think the apple is rotten right to the core

So according to our formula (•) in our first partition H_1 we have a certain expected value (which is an average⁴⁶: expectation of a random variable is a number...) which is different from the value in the second partition and in the third.

1.6.1 Conditional expectation conditional on arbitrary σ -algebras

Consider the following facts:

1. $\bar{X} \in \mathcal{F}$: it is measurable with respect to \mathcal{F} and this is clear from the definition \bullet of the random variable \bar{X} ;
2. we know that for each $V \in \mathcal{F}_+$ we have that

$$EVX = E\bar{X}$$

which is called **projection property**.

Proof

We start by letting $V = \mathbb{1}_{H_n}$ for some fixed n (and consider $\bar{X} = E_{H_n} X$). The statement becomes

$$E\mathbb{1}_{H_n} X = E\mathbb{1}_{H_n} \bar{X}.$$

The left hand side reads

$$E\mathbb{1}_{H_n} X = \int_{\Omega} \mathbb{1}_{H_n}(\omega) X(\omega) P(d\omega) = \int_{H_n} X(\omega) P(d\omega).$$

The right hand side reads

$$\begin{aligned} E \left[\mathbb{1}_{H_n} \cdot \frac{1}{P(H_n)} \int_{H_n} X(\omega) P(d\omega) \right] &= \frac{1}{P(H_n)} \int_{H_n} X(\omega) P(d\omega) P(H_n) \\ &= \int_{H_n} X(\omega) P(d\omega). \end{aligned}$$

Now consider a general random variable $V \in \mathcal{F}_+$. Of course we have (for construction)

$$V = \sum_n a_n \mathbb{1}_{H_n}$$

so V is a simple function. Here,

$$\begin{aligned} EV\bar{X} &= E \left[\sum_n a_n \mathbb{1}_{H_n} \sum_n (\mathbb{E}_{H_n} X \mathbb{1}_{H_n}) \right] \\ &= E \left[\sum_n a_n \mathbb{E}_{H_n} X \mathbb{1}_{H_n} \right] \\ &= \sum_n a_n E [\mathbb{1}_{H_n} \mathbb{E}_{H_n} X] \\ &= \sum_n a_n E [\mathbb{1}_{H_n} X] \quad \text{for the previous result} \\ &= E \left[\sum_n a_n \mathbb{1}_{H_n} X \right] \\ &= EVX. \end{aligned}$$

□

These two properties are the basis for the extended definition of conditional expectation.

Definition 1.6.1

Let \mathcal{F} be a sub- σ -algebra of \mathcal{H} . The conditional expectation $E_{\mathcal{F}} X$ of X given \mathcal{F} is defined in two steps:

- ① for $X \in \mathcal{H}_+$ (positive random variables) it is any random variable \bar{X} satisfying:

⁴⁶It's funny how this shit makes you really forget that the expected value is just the fucking average.

- (a) measurability ($\bar{X} \in \mathcal{F}_+$);
(b) projection property ($\mathbb{E}VX = \mathbb{E}V\bar{X} \quad \forall V \in \mathcal{F}_+$).

(b) for arbitrary $X \in \mathcal{H}$, if $\mathbb{E}X$ exists, we define

$$\mathbb{E}_{\mathcal{F}}X = \mathbb{E}_{\mathcal{F}}X^+ - \mathbb{E}_{\mathcal{F}}X^-.$$

Otherwise, if $\mathbb{E}X^+ = \mathbb{E}X^- = \infty$, then $\mathbb{E}_{\mathcal{F}}$ is left undefined.

Any random variable satisfying the first property is eligible as conditional expectation: that's why we call it a **version** of the conditional expectation of the random variable.

Revise with Kotatsu!

If Y and Z are positive and \mathcal{F} -measurable and if $\mathbb{E}VY \leq \mathbb{E}VZ$ for each $V \in \mathcal{F}_+$ then $Y \leq Z$ almost surely. Moreover, if $\mathbb{E}VY = \mathbb{E}VZ, \quad \forall V \in \mathcal{F}_+$ (or $\forall 1_H \in \mathcal{F}_+$), then $Y = Z$ a.s.

1.6.2 Uniqueness of conditional expectation

This is a simple matter. Let \bar{X} and $\bar{\bar{X}}$ be versions of $\mathbb{E}_{\mathcal{F}}X, X \geq 0$.

1. Both \bar{X} and $\bar{\bar{X}}$ are \mathcal{F}^+ -measurable;
2. $\mathbb{E}VX = \mathbb{E}V\bar{X} = \mathbb{E}V\bar{\bar{X}}$ for every $V \in \mathcal{F}_+$. Hence

$$\bar{X} = \bar{\bar{X}} \text{ a.s.}$$

Conversely, if $\mathbb{E}_{\mathcal{F}}X - \bar{X}$ and $\bar{\bar{X}} \in \mathcal{F}_+$ and $\bar{X} = \bar{\bar{X}}$ a.s. then $\bar{\bar{X}}$ satisfies the projection property $\mathbb{E}XV = \mathbb{E}\bar{\bar{X}}V$ (i.e. $\bar{\bar{X}}$ is a version of $\mathbb{E}_{\mathcal{F}}X$).

Remark

Tower rule: this is the projection property for $V = 1, X \in \mathcal{H}_+$. We have

$$\mathbb{E}X = \mathbb{E}\bar{X} = \mathbb{E}\mathbb{E}_{\mathcal{F}}X.$$

Moreover, if X is integrable then $\mathbb{E}_{\mathcal{F}}X$ is integrable.

Remark

If X is integrable we can rewrite the projection property as follows:

$$\mathbb{E}V(X - \bar{X}) = 0 \quad \forall V \in \mathcal{F}_b.$$

There is an interesting fact linked to this last representation. Let X be integrable: then $(X - \bar{X})$ is integrable. Define

$$\tilde{X} = X - \bar{X}$$

and write the following decomposition:

$$X = \bar{X} + \tilde{X}.$$

Here \bar{X} is determined by the information given by \mathcal{F} while \tilde{X} is orthogonal to \mathcal{F} , that is to say that $\mathbb{E}1_H\tilde{X} = 0, \forall H \in \mathcal{F}$. We call \bar{X} the **orthogonal projection of X on \mathcal{F}** .

He was forced to never mention again linear algebra in his job as a data scientist

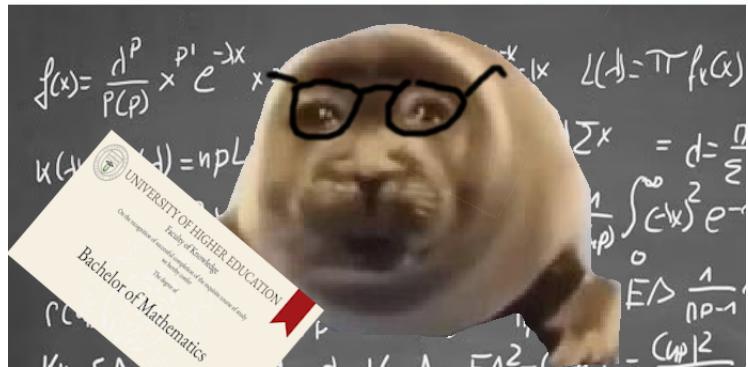


Figure 1.14: Absolutely honey you *need* all of this stuff to be a data scientist.

We are still to prove the existence of conditional expectation, but to do so we need some measure theory notions... Are you happy?

Revise with Kotatsu!

Let (E, \mathcal{E}, μ) be a measurable space and let g be \mathcal{E}_+ -measurable function. Define

$$\nu(A) = \mu(g \mathbb{1}_A) = \mathbb{1}_A g(x) \mu(dx) \quad \forall A \in \mathcal{E}.$$

It can be proved that ν is a measure on (E, \mathcal{E}) . So putting together (E, \mathcal{E}) with our new measure ν we can get a new measurable space.

Proposition 1.6.1

For every $f \in \mathcal{E}_+$ we have that (for such measures μ and ν)

$$\nu f = \mu(gf) = \int_E f(x) \nu(dx) = \int_E g(x)f(x) \mu(dx).$$

Definition 1.6.2

Consider a measurable space (E, \mathcal{E}) and two measures on it, say μ and ν (two general measures). Then ν is said to be **absolutely continuous** with respect to the other measure μ if

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \mathcal{E}.$$

Note that if $\nu(A) = \int_A g(x) \mu(dx)$ then ν is absolutely continuous with respect to μ , of course. This is a good way to construct an absolutely continuous measure!

Theorem 1.6.1

Radon-Nikodym theorem.

Suppose that μ is σ -finite and ν is absolutely continuous with respect to μ then there exists a positive \mathcal{E} -measurable function g such that

$$\int_E f(x)\nu(dx) = \int_E f(x)g(x)\mu(dx) \quad (\Delta)$$

Moreover, g is unique up to equivalence (that is, if (Δ) holds for another $\hat{g} \in \mathcal{E}_+$ then $g = \hat{g}$ μ -almost everywhere).

The function g is called **density** or **Radon-Nikodym derivative of measure ν with respect to measure μ** .

Theorem 1.6.2

Let $X \in \mathcal{H}$. Let \mathcal{F} be a sub- σ -algebra of \mathcal{H} . Then the conditional expectation $\mathbb{E}_{\mathcal{F}}X$ exists and it is unique up to equivalence.

Let's prove this theorem for \mathcal{H} -positive random variables.

Proof

$\forall H \in \mathcal{F}$ on the (reduced) measurable space (Ω, \mathcal{F}) . Now on the (reduced) measurable space we consider the restriction of \mathbb{P} on \mathcal{F} . Further consider

$$Q(H) = \int_H \mathbb{P}(d\omega)X(\omega)$$

Where \mathbb{P} is a probability measure and Q is a measure which is absolutely continuous with respect to \mathbb{P} . Remember that in this measurable space random variables are functions, so we can apply Radon-Nikodym theorem: this tells us that there exists a random variable $\bar{X} \in \mathcal{F}_+$ such that

$$\int_{\Omega} Q(d\omega)V(\omega) = \int_{\Omega} \mathbb{P}(d\omega)\bar{X}(\omega)V(\omega) \quad \forall V \in \mathcal{F}_+$$

So the random variable \bar{X} is a *version* of $\mathbb{E}_{\mathcal{F}}X$. □

To better understand this proof, set $V = \mathbb{1}_H, H \in \mathcal{F}$. We have

$$\begin{aligned} \int_{\Omega} Q(d\omega)\mathbb{1}_H(\omega) &= Q = \int_H X(\omega)\mathbb{P}(d\omega) \\ &= \int_H \bar{X}(\omega)\mathbb{P}(d\omega). \end{aligned}$$

1.6.3 Properties of conditional expectation

. We assume positivity and/or integrability of the random variables involved.

① monotonicity:

$$X \leq Y \implies \mathbb{E}_{\mathcal{F}}X \leq \mathbb{E}_{\mathcal{F}}Y;$$

② linearity:

$$\mathbb{E}_{\mathcal{F}}(aX + bY + c) = a\mathbb{E}_{\mathcal{F}}X + b\mathbb{E}_{\mathcal{F}}Y + c;$$

③ monotone convergence theorem:

$$(X_n)_n \text{ s.t. } X_n \geq 0 \forall n, X_n \nearrow X \implies \mathbb{E}_{\mathcal{F}}X_n \nearrow \mathbb{E}_{\mathcal{F}}X;$$

④ Fatou's lemma:

$$X \geq 0 \implies \mathbb{E}_{\mathcal{F}}\liminf X_n \leq \mathbb{E}_{\mathcal{F}}X_n;$$

⑤ **Dominated convergence theorem:**

$(X_n)_n$ a.s. $X_n \rightarrow X$, $|X_n| \leq Y$, Y integrable $\implies \mathbb{E}_{\mathcal{F}} X_n \rightarrow \mathbb{E}_{\mathcal{F}} X$;

⑥ **Jensen's inequality:**

f convex $\implies \mathbb{E}_{\mathcal{F}} f(x) \leq f(\mathbb{E}_{\mathcal{F}} X)$.

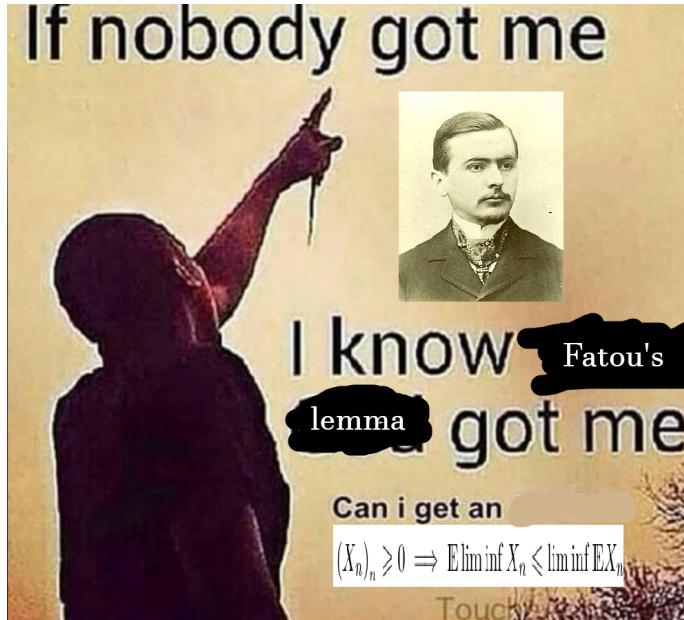


Figure 1.15: The name Fatou is hard as fuck, if you ask me. Apparently he teamed up with a dude with no nose, look it up.

Remember that these results hold only almost surely, not up to equivalence!

Theorem 1.6.3

Let \mathcal{F}, \mathcal{G} be two sub- σ -algebras of \mathcal{H} . Let W and X be \mathcal{H} -measurable random variables such that EX, EWX exist. Then

- a) $W \in \mathcal{F} \implies \mathbb{E}_{\mathcal{F}} WX = W\mathbb{E}_{\mathcal{F}} X$. This is called **conditional determinism**: we can basically treat W as a constant!
- b) $\mathcal{F} \subset \mathcal{G} \implies \mathbb{E}_{\mathcal{F}} \mathbb{E}_{\mathcal{G}} X = \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X$. So the smaller σ -algebra “wins” over the larger: this is called **repeated conditioning**.

Proof

- a) Let $X \in \mathcal{H}_+$, $W \in \mathcal{F}_+$. Then we already know that $\bar{X} = \mathbb{E}_{\mathcal{F}} \in \mathcal{F}_+$. Consider now

$$\mathbb{E} V(WX) = \mathbb{E}_{\mathcal{F}_+} (VW) \bar{X} = \mathbb{E}_{\mathcal{F}_+} (VW \bar{X}) = \mathbb{E} V(W\bar{X}) \quad \forall V \in \mathcal{F}_+$$

by projection property

and hence $\bar{W}\bar{X} = W\mathbb{E}_{\mathcal{F}} X$ is a version of $\mathbb{E}_{\mathcal{F}}(WX)$.

- b) Let $\mathcal{F} \subset \mathcal{G}$, $X \in \mathcal{H}_+$. By definition we know that $E = \mathbb{E}_{\mathcal{F}} X \in \mathcal{F}_+$ so $W \in \mathcal{G}_+$. But due to conditional determinism we have that $\mathbb{E}_{\mathcal{G}} W = W$. Hence

$$\mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X.$$

Now we have to prove that

$$\mathbb{E}_{\mathcal{F}} \mathbb{E}_{\mathcal{G}} X = \mathbb{E}_{\mathcal{F}} X.$$

Call $\mathbb{E}_{\mathcal{G}} X = Y$ and $\mathbb{E}_{\mathcal{F}} X = \bar{X}$, i.e.

$$\mathbb{E}_{\mathcal{F}} Y = \bar{X}.$$

Clearly $\bar{X} \in \mathcal{F}_+$ (\checkmark measurability) and we still have to prove projection property. By the definition of \bar{X} (as a version of $\mathbb{E}_{\mathcal{F}} X$) we have that

$$\mathbb{E}V\bar{X} = \mathbb{E}VX \quad \forall V \in \mathcal{F}_+.$$

Recall that $Y = \mathbb{E}_{\mathcal{G}} X \in \mathcal{G}_+$ and notice that V is also \mathcal{G} -measurable since $V \in \mathcal{F}_+ \subset \mathcal{G}_+$. For Y we have that $\mathbb{E}VY = \mathbb{E}VX$ for every $V \in \mathcal{G}_+$ and especially for those that are also \mathcal{F}_+ -measurable. Hence

$$\mathbb{E}V\bar{X} = \mathbb{E}VY$$

that is to say \bar{X} is also a version of the conditional expectation $\mathbb{E}_{\mathcal{F}} Y = \mathbb{E}_f [\mathbb{E}_{\mathcal{G}} X]$.

□

1.6.4 Conditioning as projection

Note the notation!

Consider the space L^2 . If we are talking about \mathcal{H} -measurable functions that belong to L^2 we write

$$L^2(\mathcal{H}) \ni X$$

And if we have a sub- σ -algebra \mathcal{F} of \mathcal{H} we will know that $L^2(\mathcal{F}) \subset L^2(\mathcal{H})$.

Theorem 1.6.4

$\forall X \in L^2(\mathcal{H})$ there exists a unique (up to equivalence) $\bar{X} \in L^2(\mathcal{H})$ such that

$$\mathbb{E}|X - \bar{X}|^2 = \inf_{Y \in L^2(\mathcal{F})} \mathbb{E}|X - Y|^2.$$

Furthermore, $X - \bar{X}$ is orthogonal to $L^2(\mathcal{F})$, i.e.

$$\mathbb{E}V(X - \bar{X}) = 0 \quad \forall V \in L^2(\mathcal{F})$$

Note that $L^2(\mathcal{H})$ is a complete Hilbert space in which the inner product of X and Y is given by $\mathbb{E}XY$. \bar{X} is the **orthogonal projection of the vector X** onto the subspace $L^2(\mathcal{F})$ and the decomposition

$$X = \bar{X} + \tilde{X}$$

holds.

Proof

Let's write the L^2 -norm of X calling it $\|X\|$:

$$\|X\| = \|X\|_2 = \sqrt{\mathbb{E}X^2}.$$

Fix $X \in L^2(\mathcal{H})$. Define

$$\delta = \inf_{Y \in L^2(\mathcal{F})} \|X - Y\|.$$

Let $(Y_n)_n \subset L^2(\mathcal{F})$ such that $\delta_n = \|X - Y_n\| \xrightarrow{n \rightarrow \infty} 0$. Let us prove that $(Y_n)_n$ is a Cauchy sequence for the $L^2(\mathcal{F})$ -convergence.

$$|Y_n - Y_m|^2 = 2|X - Y_m|^2 - 4|x - \underbrace{\frac{1}{2}(Y_n + Y_m)}_{\in L^2(\mathcal{F})}|^2.$$

Take the expectation on both sides:

$$\mathbb{E}|Y_n - Y_m|^2 \leq 2\delta_m^2 + 2\delta_n^2 - 4\delta^2.$$

Now we take the limit for n and m and what we get is

$$\lim_{m,n \rightarrow \infty} \mathbb{E}|Y_n - Y_m|^2 \leq 0.$$

Hence it is true that $(Y_n)_n$ is Cauchy and this means that there exists a $\bar{X} \in L^2(\mathcal{F})$ such that $|Y_n - \bar{X}| \xrightarrow{n \rightarrow \infty} 0$. Note that \bar{X} is unique up to equivalence (by definition of L^2 -norm). Note also

$$\bar{X} \in L^2(\mathcal{F}) \implies |X - \bar{X}| \geq \delta.$$

Now, by Minkowski's inequality we can write that

$$|X - \bar{X}| \leq \|X - Y_n\| + |Y_n - \bar{X}| \xrightarrow{n \rightarrow \infty} \delta + 0 = \delta.$$

We have thus

$$|X - \bar{X}| = \delta.$$

For $V \in L^2(\mathcal{F})$ and $a \in \mathbb{R}$, since $\mathbb{E}|X - \bar{X}|^2 = \delta$ then we have that

$$\begin{aligned} a^2 \mathbb{E}V^2 - 2a \mathbb{E}V(X - \bar{X}) + \delta^2 &= |aV - (X - \bar{X})|^2 \\ &= \|X - (\underbrace{aV + \bar{X}}_{\in L^2(\mathcal{F})})\|^2 = \delta^2 \end{aligned}$$

And therefore

$$a^2 \mathbb{E}V^2 - 2a \mathbb{E}V(X - \bar{X}) \leq 0 \quad \forall a \in \mathbb{R}$$

which is impossible unless

$$\mathbb{E}V(X - \bar{X}) = 0.$$

□

1.6.5 Conditional expectations given random variables

Consider Y , a random variable on a measurable space. We have σY that denotes the σ -algebra generated by Y . But σY is also the set of numerical random variables of the form $f \circ Y$ for a deterministic function f .

Definition 1.6.3

The conditional expectation of X given Y is

$$\mathbb{E}_{\sigma Y} X$$

If $(Y_t)_{t \in T}$ is a collection of random variables taking values in some measurable space. the conditional expectation of X given $(Y_t)_{t \in T}$ is

$$\mathbb{E}_{\mathcal{F}} X, \text{ where } \mathcal{F} = \sigma((Y_t)_{t \in T}).$$

Theorem 1.6.5

Let $X \in \mathcal{H}_+$. Let Y be a random variable taking values in the measurable space (E, \mathcal{E}) . Then every version of $\mathbb{E}_{\sigma Y} X$ has the form $f \circ Y$ for some $f \in \mathcal{E}_+$.

Conversely, $f \circ Y$ is a version of $\mathbb{E}_{\sigma Y} X$ if and only if

$$\mathbb{E}f \circ Y \cdot h \circ Y = \mathbb{E}X \cdot h \circ Y \quad \forall h \in \mathcal{E}_+.$$

Actually the last equality is just the projection property, if we think about it hard enough. The notation is once more problematic, though.

Note the notation!

Cinlar has a specific notation for conditional expectation, which is $\mathbb{E}_{\sigma Y} X$. He is stupid and autistic because the whole world uses $\mathbb{E}(X|Y)$. Professor Polito says that “Oh well but $\mathbb{E}_{\sigma Y} X$ is more precise b-because it’s more rigorous” well SHUT UP. SHUT UP NO ONE USES IT NO ONE EVEN KNOWS WHAT A FUCKING σ -ALGEBRA EVEN IS AND YOU KNOW WHY? BECAUSE IT IS USELESS BECAUSE ALL OF THIS SHIT IS USELESS IT IS A WASTE OF TIME AND IT IS NOT EVEN REMOTELY INTERESTING. LIKE I HAVE A FRIEND WHO STUDIES GEOLOGY THEY LITERALLY STUDY FUCKING ROCKS THEY STUDY ALL THE MINUTE CHEMICALS ON THE ROCKS YOU FIND ON THE GROUND THEY STUDY HOW MANY FUCKING DINOSAURS SHITTED AND PISSED AND CUMMED ON THAT FUCKING ROCK AFTER IT GOT EJECTED BY A STUPID VOLCANO A MILLION YEARS AGO AND IT IS MORE INTERESTING AND USEFUL THAN THIS LIKE IT IS LITERALLY MORE USEFUL KNOWLEDGE IN ANY REAL DATA SCIENCE JOB WHERE ANYONE LITERALLY SAYS $\mathbb{E}(X|Y)$ BECAUSE THAT’S WHAT YOU NEED TO KNOW YOU ONLY NEED TO KNOW THAT WE KNOW THAT Y HAPPENED WHEN WE EVALUATE THE PROBABILITY OF X YOU DON’T NEED TO KNOW ANYTHING ELSE I DON’T CARE AND NO ONE CARES.

The two notations are interchangeable but $\mathbb{E}_{\sigma Y} X$ is preferable.

1.6.6 Conditional probabilities and distributions

Definition 1.6.4

For every event $H \in \mathcal{H}$

$$\mathbb{P}_{\mathcal{F}}(H) = \mathbb{E}_{\mathcal{F}} \mathbf{1}_H$$

is called **conditional probability of H given \mathcal{F}** (with \mathcal{F} beinf a sub- σ -algebra of \mathcal{H}).

Lol what the fuck? Why are we defining probability through expectation? I don’t know but remember that conditional expectation is a full-fledged random variable!

Definition 1.6.5

Consider two events, G and H . The conditional probability of H given G is the number $\mathbb{P}_G(H) \in [0, 1]$ satisfying

$$\mathbb{P}(G \cup H) = \mathbb{P}(G)\mathbb{P}_G(H)$$

that is unique if $\mathbb{P}(G) > 0$.

So in this case is not a random variable anymore but only a number.

Let us consider the conditional probability $\mathbb{P}_{\mathcal{F}} H \forall H \in \mathcal{H}$ (we will remove parentheses from now on because you suck and you should die). Let $Q(H)$ be a version of $\mathbb{P}_{\mathcal{F}} H$ (remember... conditional probabilities are now expectations!). Assume $Q(\emptyset) = 0$ and $Q(\Omega) = 1$.

$Q(H)$ is a random variable and it is \mathcal{F} -measurable by definition. Let $Q_{\omega}(H)$ be the value at point $\omega \in \Omega$.

Note the notation!

If we called $Q(H) = Z$ then $Q_{\omega}(H)$ would be $Z(\omega)$.

So what is Q actually? It is a function $Q : (\omega, H) \mapsto Q_{\omega}(H)$ whith $\omega \in (\Omega, \mathcal{F})$ (a subset of (Ω, \mathcal{H})) and $H \in (\Omega, \mathcal{H})$. This is similar to a transition kernel... A transition kernel from (Ω, \mathcal{F}) into (Ω, \mathcal{H}) . But it is really? Do we care? Of course not, but we need to check because we need more reasons to end our life by our hand. For something to be a transition kernel it should verify:

1. the mapping $\omega \mapsto Q_{\omega}(H)$ should be \mathcal{F} -measurable (in this case it is ok for every $H \in \mathcal{H}$);
2. for every disjoint sequence $(H_n)_n \in \mathcal{H}$ it should hold $\forall \omega \in \Omega$ the following:

$$Q_{\omega}\left(\bigcup_n H_n\right) = \sum_n Q_{\omega}(H_n). \quad (\star)$$

This last requirement arises a problem: we could apply monotone convergence theorem for conditional expectations, but we know that those property hold only almost surely! I can already hear the reader say “oh but why should we care isn’t it supposed to be like the same of saying “equal to” and things like this and so on” but well, my dear idiotic little shit, when something is almost surely it is actually *almost* surely. It means that you’re *almost* guaranteed that nothing will fuck you in the ass unless something is small enough to creep in your asshole without you even noticing. Thank measure theory enthusiasts for this.



Figure 1.16: Snake has had enough and honestly so do I.

The fact is that we could have some ω 's with measure zero so technically there are some exception points for which it doesn't hold. These points depends on the choice of the sequence H_n . Let Ω_0 be the set for which (\star) holds:

$$\Omega_0 = \bigcap_h \Omega_h$$

where Ω_h is the almost surely set including all ω for the sequence $h = (H_n)_n$. But we may have uncountably many problems... Why are there so many problems with conditional expectation? The root of the problem is that conditional expectation is not an exactly well-defined object.

Definition 1.6.6

Let $Q(H)$ be a version of $\mathbb{P}_{\mathcal{F}}(H), \forall H \in \mathcal{H}$. Then $Q : (\omega, H) \mapsto Q_\omega(H)$ is said to be a **regular version** of the conditional probability $\mathbb{P}_{\mathcal{F}}$ provided that Q is a transition probability kernel from (Ω, \mathcal{F}) into (Ω, \mathcal{H}) .

Proposition 1.6.2

Let $\mathbb{P}_{\mathcal{F}}$ have a regular version Q . Then

$$QX : \omega \mapsto Q_\omega X = \int_{\Omega} Q_\omega(d\omega') X(\omega')$$

is a version of $\mathbb{E}_{\mathcal{F}} X \forall$ random variable X whose expectation exists.

So we are integrating over Ω our function X against our regular version of the conditional probability Q .

Theorem 1.6.6

If (Ω, \mathcal{H}) is a standard measurable space (which means that it is isomorphic to (F, \mathcal{B}_F) for some Borel subset F of \mathbb{R}). Then $\mathbb{P}_{\mathcal{F}}$ has a regular version.

So as long as we work in standard measurable spaces we do not need to worry. Now that we have assessed this fact we can define the concept of conditional distribution.

Definition 1.6.7

Consider a random variable Y taking values on (E, \mathcal{E}) and a sub- σ -algebra \mathcal{F} of \mathcal{H} . The **conditional distribution of Y given \mathcal{F}** is any transition probability kernel

$$L : (\omega, B) \mapsto L_{\omega}(B)$$

From (Ω, \mathcal{F}) into (E, \mathcal{E}) such that

$$\mathbb{P}_{\mathcal{F}}(Y \in B) = L(B), \quad B \in \mathcal{E}.$$

Here we already have our probability distribution of Y with our probability measure that Y belongs to B for every $B \in \mathcal{E}$. So the conditional transition is a transition probability kernel and moreover, if $\mathbb{P}_{\mathcal{F}}$ has a regular version Q then

$$L_{\omega}(B) = Q_{\omega}(\{Y \in B\}), \quad \omega \in \Omega, B \in \mathcal{E}$$

Defines a version L of the conditional distribution of Y given \mathcal{F} .

Note the notation!

This would be

$$\mathbb{P}(Y \in B | \mathcal{F}).$$

Theorem 1.6.7

If (E, \mathcal{E}) is a standard measurable space then there exists a version of the conditional distribution of Y given \mathcal{F} .

We will use this notion to perform a reverse operation that we did several lectures ago when we built measures on product spaces. Back then we talked about kernels in which the dependence structure is actually included in the kernel measure in itself.

1.6.7 Disintegration of measures

There may be a The Cure joke but my spirit is already too beaten down to think about one. We introduced measures on product spaces using Σ -finite transition kernels from $(D, \mathcal{D}) \rightarrow (E, \mathcal{E})$:

$$\pi f = \int_D \mu(dx) \int_E K(x, dy) \underbrace{f(x, y)}_{\in (\mathcal{D} \otimes \mathcal{E})_+}. \quad (\circ)$$

This integral defines a measure π on $(D \times E, \mathcal{D} \otimes \mathcal{E})$. This measure π is σ -finite and unique if μ is σ -finite and K is σ -bounded.

Let (D, \mathcal{D}) and (E, \mathcal{E}) be measurable spaces and let μ be a probability measure on (D, \mathcal{D}) . Let K be a probability transition kernel from (D, \mathcal{D}) into (E, \mathcal{E}) . Consider the measure π on the product space:

$$\pi f = \int_{D \times E} f(x, y) \pi(dx, dy).$$

Note also that π is the unique measure satisfying

$$\pi(A \times B) = \int_A \mu(dx) K(x, B), \quad \text{with } \begin{array}{l} A \in \mathcal{D} \\ B \in \mathcal{E} \\ x \in D \\ y \in E. \end{array}$$

We can also write the same thing in differential notation as

$$\pi(dx dy) = \underbrace{\mu(dx)}_{\text{joint distr.}} \underbrace{K(x, dy)}_{\substack{\text{distr. of } X \text{ conditional} \\ \text{probability}}} \cdot \underbrace{}_{\substack{\text{that } Y \in dy \text{ given that} \\ X = x}}.$$

Theorem 1.6.8

Let π be a probability measure on $(D \times E, \mathcal{D} \otimes \mathcal{E})$. Let (E, \mathcal{E}) be a standard space. Then there exists a probability measure μ on (D, \mathcal{D}) and a transition probability kernel K from (D, \mathcal{D}) into (E, \mathcal{E}) such that (o) holds.

Proof

Let $W = D \times E$, $\mathcal{W} = \mathcal{D} \otimes \mathcal{E}$, $\mathbb{P} = \pi$ so that we have our probability space (W, \mathcal{W}, π) . Now on this space we can define the random variables X and Y . Consider the notation $w \in W$: w has two components so that $w = (x, y)$. Define

$$\begin{aligned} X &\text{ s.t. } X(w) = x \text{ (it gives us the first coordinate)} \\ Y &\text{ s.t. } Y(w) = y \text{ (it gives us the second coordinate).} \end{aligned}$$

Further, let μ be the distribution of X , that is $\mu(A) = \pi(A \times E)$, $A \in \mathcal{D}$. Also note that Y takes values in (E, \mathcal{E}) which is standard. Since (E, \mathcal{E}) is standard it exists a regular version that we can call L of the conditional distribution of Y given the sub- σ -algebra $\mathcal{F} = \sigma X$. So this random variable is measurable with respect to \mathcal{F} that in this case is σX . But we know that if a random variable is measurable with respect to the σ -algebra generated by another random variable then we know that the first random variable is just a deterministic function of the second random variable (in this case X). Hence L is a function of X and

$$L_w(B) = K(X(w), B) \quad (\text{it doesn't depend on } Y \text{ at all!})$$

where K is a transition kernel from (D, \mathcal{D}) into (E, \mathcal{E}) . By the projection property for L we have:

$$\begin{aligned} \pi(A \times B) &= \mathbb{E} \mathbb{1}_A \circ X \cdot \mathbb{1}_B \circ Y = \mathbb{E} \mathbb{1}_A \circ X \cdot K(X, B) \\ &= \int_D \mathbb{1}_A \mu(dx) K(x, B). \end{aligned}$$

So for indicators (o) holds. □

Once we proved this for indicators we can go on for simple functions and positive functions in the usual way.

Theorem 1.6.9

If H has the representation (o) then the kernel L

$$L_w(B) = K(X(w), B) \quad w \in \mathbb{N}, B \in \mathcal{E}$$

is a **version of the conditional distribution** of Y given \mathcal{F} .

1.6.8 Conditional independence

We are now going to relate what we have studied so far with the idea of conditional probability.

Definition 1.6.8

Consider some sub- σ -algebras of \mathcal{H} :

$$\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$$

Then the sub- σ -algebras F_1, F_2, \dots, F_n are said to be **conditionally independent given \mathcal{F}** if

$$\mathbb{E}_{\mathcal{F}}V_1 \cdot V_2 \cdot \dots \cdot V_n = \mathbb{E}_{\mathcal{F}}V_1 \cdot \mathbb{E}_{\mathcal{F}}V_2 \cdot \dots \cdot \mathbb{E}_{\mathcal{F}}V_n$$

for every positive random variables V_1, V_2, \dots, V_n measurable with respect to $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ respectively.

Note that the usual notion of independence corresponds to the specific choice of trivial σ -algebra:

$$\mathcal{F} = \{\emptyset, \Omega\}.$$

It is over. It is finally over and I am a considerably worse person now. Martingales, anyway, are very linked to the notion of conditional probability... And so a lot of other stochastic processes. Whatever.

Chapter 2

A first study in stochastic processes: Martingales

Okay, I'll admit that this part is really hard for me to follow. I am already broken down by Professor Polito's lectures and right now... I only want to be happy again. Please note that the astonishing drop in the quality of the notes is not my fault. This shit should be easier to explain but they manage to mess this up anyway.

2.1 Revising stochastic processes and filtrations

Note the notation!

1. we will work in a fixed probability space $(\Omega, \mathcal{H}, \mathbb{P})$;
2. we will denote an index set \mathbb{T} (which is usually interpreted as time). It could be $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{R}$, for example;
3. $\{X_t : t \in \mathbb{T}\}$ is a random variable taking values in the product space

$$(\mathcal{F}, \mathcal{F}) = (E^{\mathbb{T}}, \mathcal{E}^{\mathbb{T}}).$$

We can think of this as an evolving stochastic process.

Let \mathbb{T} be a subset of \mathbb{R} . For each $t \in \mathbb{T}$ let \mathcal{F}_t be a sub- σ -algebra of \mathcal{H} . The family $\mathcal{F} = \{\mathcal{F}_t : t \in \mathbb{T}\}$ is called a **filtration** provided that

$$\mathcal{F}_s \in \mathcal{F}_t \quad \forall s < t.$$

Definition 2.1.1

A **stochastic process** $X = (X_t)_{t \in \mathbb{T}}$ generates a filtration that is named **natural filtration**.

As the time passes, I discover more and more information about \mathcal{H} . Each process defines a filtration by generating a “flow of information”.

Example 2.1.1

Imagine a car competition composed by two different races. Consider the events

$$\begin{aligned}\omega &= \{\text{car A wins}\} \\ \ell &= \{\text{car A loses}\}.\end{aligned}$$

Consider 2 races: our Ω will be

$$\Omega = \{\omega\omega, \ell\omega, \omega\ell, \ell\ell\}.$$

Before the starting of the race, at time 0, we have no information about what is going to happen so

our filtration will be

$$\mathcal{F}_0 = \left\{ \underbrace{\emptyset}_{\varnothing}, \underbrace{\{\omega\omega, \ell\omega, \omega\ell, \ell\ell\}}_{\Omega} \right\}$$

which is the trivial σ -algebra. Remember that in this case the empty set \emptyset represents impossible events (e.g. all cars win all races). So the random variable X_0 measurable with respect to \mathcal{F}_0 maps all the outcomes of the race to the same value. The idea is that we cannot ask ourselves “what is the probability of the car winning the first and the second race?” because we aren’t able to measure these events in \mathcal{H} . Now consider the situation after the first race: I am now able to “discern” the events regarding the first race:

$$\mathcal{F}_1 = \{\{0\}, \{\omega\ell, \omega\omega\}, \{\ell\ell, \ell\omega\}, \{\omega\omega, \omega\ell, \ell\omega, \ell\ell\}\}.$$

Now I am able to know what happened after the first race and this gives me a richer sub- σ -algebra, with $\mathcal{F}_0 \subset \mathcal{F}_1$. I am now able to have a random variable X_1 that is measurable with respect to \mathcal{F}_1 . For instance, imagine defining the following random variable:

$$X_1 = \begin{cases} 1 & \text{if the car wins the first race} \\ 0 & \text{otherwise.} \end{cases}$$

But, for instance, the random variable “the car wins the second race” is still not measurable with respect to \mathcal{F}_1 . Consider now the filtration after the second race:

$$\mathcal{F}_2 = \{\{0\}, \{\omega\omega\}, \{\ell\ell\}, \{\omega\ell\}, \{\ell\omega\}, \{\omega\ell, \omega\omega\}, \{\ell\ell, \ell\omega\}, \{\omega\omega, \omega\ell, \ell\omega, \ell\ell\}\}.$$

Of course $\mathcal{F}_2 \supset \mathcal{F}_1 \supset \mathcal{F}_0$. We could define a random variable X_3 of the following type:

$$X_3 = \begin{cases} 0 \\ 1 \\ 2 \end{cases}$$

which counts the numbers of victories of our car. X_3 is \mathcal{F}_2 -measurable but not \mathcal{F}_1 -measurable. Now \mathcal{F} is a filtration:

$$\mathcal{F} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}.$$

Example 2.1.2

What if the same experiment is observed by different people? In other ways, what if we have different learning from the same experiment? Imagine that we have two filtrations \mathcal{F} and \mathcal{G} . We say that \mathcal{F} is finer than \mathcal{G} if

$$\mathcal{F}_t \supset \mathcal{G}_t \quad \forall t \in \mathbb{T}.$$

Definition 2.1.2

Given $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$, a stochastic process $\{X_t\}_{t \in \mathbb{T}}$ is **adapted** to \mathcal{F} if X_t is \mathcal{F}_t -measurable for each $t \in \mathbb{T}$ and \mathcal{E} -measurable.

We already mentioned that every process generates its *natural filtration* $\sigma(X_t, t \in \mathbb{T})$, for which the process is clearly adapted. But we may want to consider another filtration $\mathcal{F} = (\mathcal{F}_t)$ such that $\{X_t\}$ is adapted to \mathcal{F} because all the information on X that I care about are contained in this filtration. What does this mean and what is the relationship between the two? Which one of the two is finer considering that our process is adapted to both? The answer is

$$\mathcal{F} \supset \sigma(X_t) \quad (\mathcal{F} \text{ is finer than the natural filtration})$$

because $\sigma(X_t)$ is always the smallest possible σ -algebra that a process can generate (the natural filtration is the *minimum* filtration).

Another way we could see the question of adaptation is:

“Can X be determined using only information available in \mathcal{F} ? ”

If yes, then X is adapted to \mathcal{F} .

Example 2.1.3

This is called “the secretary problem”: in this case we must start from the filtration and then understand the problem. Here I have N candidates for a position; a candidate disregarded after the interview is lost. The interviewer wants to hire exactly 1 candidate and each candidate has different abilities and the interviewer knows only the relative ability of those already interviewed so far. Our goal is to maximize the probability of hiring the best one. We have three questions:

1. what is Ω ?
2. what is the filtration \mathcal{F} for this experiment?
3. what process should we use?

In this case $\Omega = N!$ permutations of the ranking of the candidates (the order in which they show up) and the filtration is the information earned from interview up to time t (that is the ranking of the candidates up to time t). But what is the process that I should use? Consider the sequence

$$V_1, V_2, \dots, \{V_i\}_{i \geq 1}$$

with $V_i = 1$ if and only if the best candidate is the i -th candidate and $V_i = 0$ otherwise. Could this process $\{V_i\}_{i \geq 1}$ be used? No, because V is not adapted to \mathcal{F} ... because to understand if i -th candidate is the best we need to compare it to the other candidates, including the ones that didn't show up yet! But then how can we get an adapted process? Let us consider the expectation

$$U_n = \mathbb{E}[V_n | \mathcal{F}_n]$$

What do we know about the measurability of U_n ? We know that it is for sure \mathcal{F}_n -measurable. This trick gives us a simple way to build an adapted process. So now we will have: $U_n = 0$ if the candidate is not the best up to n and $U_n = 1$ otherwise. More specifically, we will have

$$\begin{aligned} U_n &= 1 \cdot \text{probability that the best candidate is among the first } n + 0 \cdot \text{probability that the best candidate is not among the first } n \\ &= 1 \cdot \frac{n}{N} + 0 \cdot \frac{N-n}{N} \\ &= \frac{n}{N} \end{aligned}$$

This is a quantity that I can measure and it is therefore adapted.

2.1.1 Stopping times

Consider now the same problem under a different point of view: we want to establish the “best time” when to stop discarding candidates. This is a random time and it is called a **stopping time**. This is nothing else than a time that we choose basing ourselves on elements or events that may or may not happen: this is why they’re a random variable. The stopping times are a random variable that can cause problems, since stopping times can be infinite: imagine that our stopping time is “whenever a volcano erupts” but the stupid fucking volcano decides to go into quiescence forever. This is why beside our index set \mathbb{T} we need to introduce $\bar{\mathbb{T}} = \mathbb{T} \cup \{\infty\}$.

Definition 2.1.3

$$T : \Omega \mapsto \bar{\mathbb{T}}$$

is called **stopping time** of \mathcal{F} if

$$\{T \leq t\} \in \mathcal{F}_t \quad \forall t \in \mathbb{T}.$$

Remember that stopping times have to be \mathcal{F} -measurable: for example “when will the fucking volcano erupt for the last time” is *not* a stopping time because it requires information that we don’t have in t !
Stopping time have an alternative definition:

statisticians Hate Him!

See how he made a process adapted in an afternoon with one simple trick



Figure 2.1: You conditioned in the wrong hood, buddy.

Definition 2.1.4

T is a stopping time if the process

$$Z_t = \mathbb{1}_{\{T \leq t\}} \quad t \in \mathbb{T}$$

is adapted to F .

From this definition it follows that when $\mathbb{T} = \mathbb{N}$ then this is equivalent to having

$$\frac{\dot{Z}_n}{Z_n - Z_{n-1}} = \mathbb{1}_{\{T=1\}} \quad \forall n \in \mathbb{N}$$

that needs being adapted to \mathcal{F}_n .

Not all random times are stopping time: think about an experiment in which we produce a signal when a physical event happens. Let us call Z the process indicator and T our stopping time:

$$Z_t \omega = \begin{cases} 0 & t < T(\omega) \\ 1 & t \geq T(\omega). \end{cases}$$

Then if the information flow allows us to recognize the occurrence of the event then T is a stopping time.

Example 2.1.4

Some stopping times:

- ① The first time that $X(\omega) \in H \in \Omega$;

$$T(\omega) = \begin{cases} \inf \{n \in \mathbb{N} : X_n(\omega) \in H\} \\ +\infty \end{cases} \quad \text{if } X_n(\omega) \notin H \forall n$$

So

$$\{T \leq n\} = \bigcup_{k=0}^n \{X_k \in H\}.$$

② Consider i.i.d. random variables X_1, X_2, \dots . Consider the probabilities

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

and the random walk

$$S_n = \sum_1^n X_i.$$

Let's define

$$T_1 = \begin{cases} \min \{n < 50 : S_n = 3\} \\ 50 \end{cases} \quad \text{otherwise.}$$

This is a stopping time because I can write $\{T_1 \leq n\}$ as

$$\bigcup_{k=1}^n \underbrace{\{S_k = 3\}}_{\mathcal{F}_k\text{-measurable}} \quad n < 50.$$

Moreover for $n = 50$ we have $T_1 \in \mathcal{F}_{50}$.

③ Starting from the previously defined random walk, consider the quantity

$$M_n = \min(S_1, \dots, S_n)$$

And the random time

$$T_2 = \min \{n : S_n \geq M_m + 2\}$$

is a stopping time. On the contrary,

$$T_3 = \begin{cases} \max \{n < 50 : S_n = 7\} & \text{if not empty} \\ 50 & \text{otherwise} \end{cases}$$

is not a stopping time. Why? Because I have to wait until $n = 50$ to answer the question.

Consider the random times on \mathbb{R}_+

$$0 < T_1 < T_2 < \dots$$

With $\lim_{n \rightarrow \infty} T_n = +\infty$. Define the process $\{N_t\}$ as

$$N_t := \sum \mathbf{1}_{[0,t]}(T_n).$$

This is called **counting process**. It is a basic count of the number of events happened up to time n .

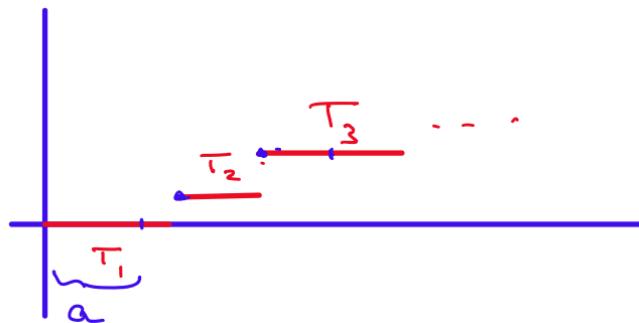


Figure 2.2: Yeah I'm counting on ending my life very soon.

N_t is increasing, right continuous and increases by unitary jumps. Moreover, $N_0 = 0, N_t < \infty$ for $t \in \mathbb{R}_+$. Of course $\lim_{t \rightarrow \infty} N_t = \infty$. Counting processes generate their natural filtration \mathcal{F} .

Some problems require to stop the observation at a stopping time (because we don't care anymore¹)... So

¹Assuming we ever did.

we don't actually need the whole knowledge of the complete filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$. The problem is that as we said before the stopping time is a random time. So what we need is the *information known up to time T* \mathcal{F}_T , basically the σ -field that is the filtration at time T .

Remark

We said that \mathcal{F}_t does not include $t = \infty$... but T may diverge to ∞ ! We need to enlarge \mathcal{F} to include $t = \infty$.

Think, for instance the stopping time “first time of overflowing of a dam”. It is something that may very well never happen and is therefore called “Vajont stopping time²”.

Definition 2.1.5

Let $\bar{\mathbb{T}} = \mathbb{T} \cup \{+\infty\}$. We define

$$\mathcal{F}_\infty = \lim_{t \rightarrow \infty} \mathcal{F}_t = \bigvee_{\substack{t \in \mathbb{T} \\ \sigma(\bigcup \mathcal{F}_t)}} \mathcal{F}_+ .$$

Definition 2.1.6

$\mathcal{F}_{\bar{\mathbb{T}}}$ is a filtration on $\bar{\mathbb{T}}$ and T is a stopping time on it if and only if it is a stopping time on $\mathcal{F}_{\mathbb{T}}$.

Remark

Every stochastic process on $(\mathcal{F}_t)_{t \in \mathbb{T}}$ can be extended to $\bar{\mathbb{T}}$ by appending any random variable X_∞ to it (with $X_\infty \in \mathcal{F}_\infty$).

So adding this random variable doesn't really add any information but only ensures that nerds don't whine about completeness. It is *technically* information that we will never user, since \mathcal{F}_∞ does not contain any further information with respect to \mathcal{F} . We denote with \mathcal{F} also the extended filtration, with the ease of mind given by the fact that a divergent stopping time won't make us commit harakiri.

Definition 2.1.7

Let \mathcal{F} be a filtration on $\bar{\mathbb{T}}$ and let T be a stopping time on \mathcal{F} . We call **past until T** the σ -algebra $\mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{H}$ such that:

$$\mathcal{F}_T = \left\{ H \in \mathcal{H} : H \cap \{T < t\} \in \mathcal{F}_t, \forall t \in \bar{\mathbb{T}} \right\} .$$

This means that \mathcal{F}_T , which represents the evolution within T , contains all the events H such that they are before the stopping time in all the filtrations. This is selecting the events H that happen within the time T .

Remark

If we fix a $T \equiv t$ then $\mathcal{F}_T \equiv \mathcal{F}_t$ (it's the normal filtration at time t).

If T is a stopping time of \mathcal{F} then $\{T \leq r\}$ belongs to \mathcal{F}_T , $\forall r \geq 0$:

$$\{T \leq r\} \cap \{T \leq t\} = \{T \leq \min(T, r)\} \in \mathcal{F}_t$$

So T is \mathcal{F}_T -measurable.

Remark

\mathcal{F}_t can be read as the collection of all \mathcal{F}_t -measurable random variables V : the value of $V(\omega)$ can be told by the time $T(\omega)$: we can read the value of $V(\omega)$ before the ringing of an alarm.

²Okay I'm sorry I'm sorry I am a terrible person yes sir yes yes I concur it was an extremely inappropriate joke yes yes sir yes you have my word yes no no never again no you don't have to worry no no.

Theorem 2.1.1

A random variable V belongs to \mathcal{F}_T if and only if

$$V \mathbb{1}_{T \leq t} \in \mathcal{F}_t, \quad \forall t \in \overline{\mathbb{T}}. \quad (i)$$

If $\overline{\mathbb{T}} \equiv \overline{\mathbb{N}}$ the condition becomes

$$V \mathbb{1}_{T=n} \in \mathcal{F}_n \quad \forall n \in \overline{\mathbb{N}}.$$

Proof

Note that

$$X_t = V \mathbb{1}_{T \leq t}$$

is a stochastic process. For $\forall r \in \mathbb{R}_+$ and $t \in \mathbb{T}$

$$\{V > r\} \cap \{T \leq t\} = \{X_t > r\}.$$

But now we know that (i) is just a consequence of the definition of \mathcal{F}_t :

$$\{V > r\} \in \mathcal{F}_T \quad \forall r \underset{\text{def. of } \mathcal{F}_T}{\iff} \{X_t > r\} \in \mathcal{F}_t \quad \forall r, \forall t \in \overline{\mathbb{T}}$$

The special case in which $\overline{\mathbb{T}} \equiv \overline{\mathbb{N}}$ means that

$$V \mathbb{1}_{\{T=n\}} = \begin{cases} X_n - X_{n-1} & \text{if } n \in \mathbb{N} \\ X_\infty - \sum_{n \in \mathbb{N}} (X_n - X_{n-1}) & \text{if } n = \infty \end{cases}$$

so that

$$V \in \mathcal{F}_T \iff V \mathbb{1}_{T=n} \in \mathcal{F}_n$$

□

But what if there is a problem that involves more than 1 stopping time? We cannot use two different “pasts” because that would cause problems with the measurability of random variables.

Example 2.1.5

Imagine a factory. The production is blocked when

- a) the temperature of the room is above a certain threshold;
- b) the machinery has not been cleaned for more than 12 hours.

Alternatively, imagine that the sales of shares happens:

- a) when the price is above a fixed value;
- b) when the increase of the price is smaller than a fixed value.

Both of this cases require the use of 2 different stopping times S and T . We may think to evaluate two typical functions of S and T of interest:

$$\underbrace{\min\{S, T\}}_{S \wedge T} \quad \underbrace{\max\{S, T\}}_{S \vee T}$$

Example 2.1.6

Truncated stopping time: let T be a stopping time (for example, the time at which we sell certain shares) and that we want a finite horizon for this decision. In this case the quantity of interest is

$$S = T \wedge n = \min\{T, n\}$$

where n could be some sort of time horizon.

Imagine that two cyclists participate to a race. Their children will have their snack when both the parents will arrive to the finish line. How long will the children wait for their snack? We can think about the following stopping times:

$$\begin{aligned} T &: \text{time employed by the first cyclist} \\ S &: \text{time employed by the second cyclist} \\ U &: \max\{S, T\}. \end{aligned}$$

The waiting time for the children will be U .

This is cool and all, but are we sure that $S \wedge T$ and $S \vee T$ are stopping times?

Remark

Yes.

Why?

Remark

Fuck you.

Aw hell nah. Now even the color boxes are against me.

You fucking prick

You will NEVER pass this exam.

Well this is not a surprise.

You fucking prick

Not if you keep jerking around.

Proof

We know that $\{S \leq t\}$ and $\{T \leq t\}$ belong to \mathcal{F}_t for every t , since they are stopping times. Consider now $\{\min\{(S, T) \leq t\}\}$ and $\{\max\{(S, T) \leq t\}\}$. We have

$$\{\min\{(S, T) \leq t\}\} = \underbrace{\{S \leq t\}}_{\in \mathcal{F}_t} \cup \underbrace{\{T \leq t\}}_{\in \mathcal{F}_t} \underbrace{\quad}_{\in \mathcal{F}_t}$$

and

$$\{\max\{(S, T) \leq t\}\} = \underbrace{\{S \leq t\}}_{\in \mathcal{F}_t} \cap \underbrace{\{T \leq t\}}_{\in \mathcal{F}_t} \underbrace{\quad}_{\in \mathcal{F}_t}.$$

□

There is a further theorem:

Theorem 2.1.2

① if $S \leq T \implies \mathcal{F}_S \subset \mathcal{F}_T$;

② $\mathcal{F}_{\min\{S, T\}} = \mathcal{F}_S \cap \mathcal{F}_T$;

③ if $V \in \mathcal{F}$, the following processes are in $\mathcal{F}_{\min\{S,T\}}$:

$$(a) \quad V\mathbb{1}_{S \leq T}, \quad (b) \quad V\mathbb{1}_{S=T}, \quad (c) \quad V\mathbb{1}_{S < T}.$$

Proof

1. Imagine $\{S \leq T\}$. Now consider the event $\{S \leq n\}$. Since S happens before T then we have that $\{T \leq n\} \subset \{S \leq n\}$. We know that if $A \in \mathcal{F}_S$ then

$$A \cap \{T \leq n\} = A \cap \underbrace{\{S \leq n\}}_{\in \mathcal{F}_n} \cap \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n}$$

which means that $A \cap \{T \leq n\} \in \mathcal{F}_n$ and by definition of \mathcal{F}_T this corresponds to saying

$$\mathcal{F}_S \subset \mathcal{F}_T.$$

2. (a) Let us first prove that $\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cup \mathcal{F}_T$. Note that $\min\{S, T\}$ is dominated by \mathcal{F}_S and \mathcal{F}_T , which means that $\mathcal{F}_{\min\{S,T\}}$ is included both in \mathcal{F}_S and \mathcal{F}_T . By point 1 we have

$$\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cap \mathcal{F}_T.$$

- (b) It remains to prove that $\mathcal{F}_{\min\{S,T\}} \supset \mathcal{F}_S \cup \mathcal{F}_T$. Let $H \in \mathcal{F}_S \cap \mathcal{F}_T$. This means that

$$H \cap \{S \leq T\} \in \mathcal{F}_{\min\{S,T\}}$$

by part 1 of this theorem, since $H \in \mathcal{F}_S$. But we also know that

$$H \cap \{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}$$

by part 3 of this theorem since also $H \in \mathcal{F}_T$. But this means that

$$H = \{H \cap \{S \leq T\}\} \cup \{H \cap \{T \leq S\}\}$$

belongs to $\mathcal{F}_{\min\{S,T\}}$ and so

$$\mathcal{F}_{\{S \cap T\}} \subset \mathcal{F}_{\min\{S,T\}}.$$

3. (a) $V\mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}$. To prove this we use the following theorem:

Theorem 2.1.3

Let T be a stopping time of \mathcal{F} . Then

$$\mathcal{F}_T = \{X_T : X \in \mathcal{F}\}$$

with X in \mathcal{F} .

This basically means that \mathcal{F}_t becomes the values X_T of X in \mathcal{F} at time T . Let us consider $X_t = V\mathbb{1}_{\{S \leq t\}}$ with $t = \min\{S, T\}$. We know that $\min\{S, T\}$ is a stopping time of $\mathcal{F}_{\min\{S,T\}}$ and by the theorem above

$$X_{\min\{S,T\}} \in \mathcal{F}_{\min\{S,T\}}$$

or, alternatively,

$$\mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}.$$

- (b) Let $V = 2 \implies \{S \leq T\} \in \mathcal{F}_{\min\{S,T\}}$ and by symmetry $\{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}$. This means that

$$\{S = T\} = \{S \leq T\} \cap \{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}.$$

Furthermore

$$\{S < T\} = \{S \leq T\} \setminus \{S = T\}$$

Which implies membership of $\mathcal{F}_{\min\{S,T\}}$.

- (c) Consider the fact already proven that $V\mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}$. Multiplying by $\mathbb{1}_{\{S=T\}}$ or $\mathbb{1}_{\{S < T\}}$ does not alter the membership in $\mathcal{F}_{\min\{S,T\}}$.

□

Consider now a machinery what when it becomes too hot (above a certain threshold) an alarm rings and after 3 seconds the machinery stops. Here we deal with 2 times:

$$\begin{aligned} T : & \text{ time of the alarm (stopping time)} \\ t = 3 : & \text{ deterministic time (stopping time as well)} \end{aligned}$$

and the time of our interest is the time S when the machinery stops:

$$S = T + t$$

Remark

At time T we know S : We know that when the alarm rings within three seconds the machinery will stop, so S is actually \mathcal{F}_T -measurable. We say that S is **foretold by T** .

Definition 2.1.8

Let

S be a stopping time

T be a random time $T > S$ but whose value can be told by time S .

We say that T is **foretold by S** (and it is a stopping time).

Example 2.1.7

Consider S and $T = 2S$. Surprise surprise, T is a foretold time.

Did we need an example for this?

Example 2.1.8

Yes.

Yeah this meta-joke has been running long enough.

2.1.2 Discrete and continuous stopping times

Definition 2.1.9

Let

$$d_n(t) = \begin{cases} \frac{k+1}{2^n} & \text{if } \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \\ +\infty & \text{if } t = \infty. \end{cases}$$

We got a function

$$d_n : \overline{\mathbb{R}}_+ \mapsto \overline{\mathbb{R}}_+.$$

Consider the case $n = 1$ (remember that $k = 0, 1, 2, \dots$):

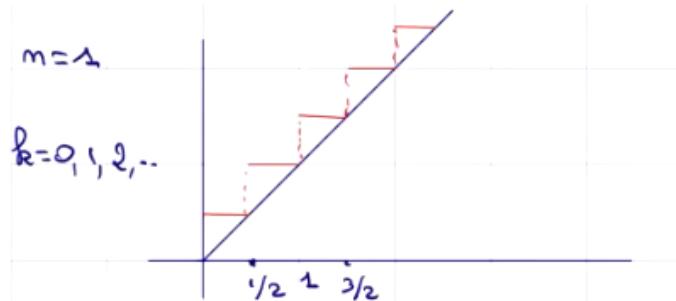


Figure 2.3: These are screenshots taken from the video feed. I couldn't bother. Sorry.

But after $n = 2$ we get the same function with smaller "steps":

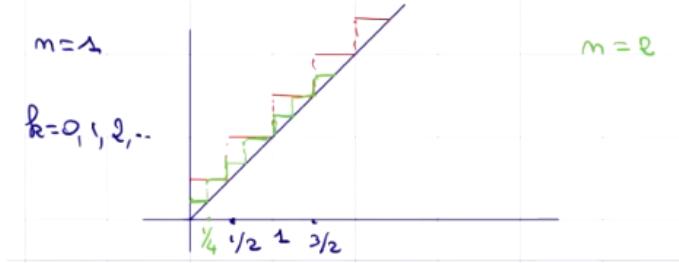


Figure 2.4: The aesthetic problem of musical execution is reduced by Enrico Fubini to a problem of musical improvisation: I honestly think that this perspective is limiting both in its conceptions of “execution” and “improvisation” even if I agree with the intent of blurring the line between the two concepts unjustly separated in our Western modern musical. That is, a modern perspective: the reader will surely remember that Bach was famous first and foremost for his improvisational abilities and that the musical form of the fugue was often used to display pyrotechnical capabilities in that regard.

Observe that $d_1 \geq d_2 \geq d_3 \geq \dots$. What are the other properties of this function?

1. it is a step function;
2. it is right-continuous;
3. $d_n(t) > t$ (the diagonal line);
4. $\lim_{n \rightarrow \infty} d_n(t) = t$.

Now we apply this simple function to our stopping times so that we get the following proposition.

Proposition 2.1.1

Let \mathcal{F} be a filtration on $\overline{\mathbb{R}}_+$ and let T be a stopping time. Let

$$T_n = d_n \circ T.$$

Then $\{T_n\}_n$ is a sequence of **discrete stopping times** of \mathcal{F} decreasing to T .

Remark

Of course T_n is a foretold time by T , because if I know T I also know the value of T_n .

Proof

Fix n :

- T_n is a measurable function of $T \implies T_n \in \mathcal{F}_n$;
- $d_n(t) \geq t \quad \forall t < \infty$ and $d_n(\infty) = \infty \implies d_n(T) = T_n \geq T$;
- T_n is a foretold time by T ;
- T_n is a stopping time;
- T_n is discrete.

But $d_n(t) \searrow t$ as $n \rightarrow \infty \implies d_n(T) = T_n \searrow T$. We have thus switched from a continuous stopping time to a sequence of discrete stopping times. \square

An important thing to note is that during our process of discretization we never break the property of the stopping times... and this is possible because we are using foretold times!

2.1.3 Conditioning at stopping times

Imagine that we have our information about X that stops at the stopping time T : this means knowing \mathcal{F}_T . Consider the classic conditioning situation

$$\mathbb{E}(X|\mathcal{F}_T).$$

We already said how this is our best estimate of X using our information available (that is \mathcal{F}_T).

Note the notation!

We will write

$$\mathbb{E}_T \equiv \mathbb{E}_{\mathcal{F}_T} \equiv \mathbb{E}(\cdot | \mathcal{F}_T).$$

Remark

Can I use T fixed as a deterministic time t and thus get $\mathbb{E}(X|\mathcal{F}_t)$? Is this a special case or a different think altogether? Well yes, because t is a stopping time being \mathcal{F}_t -measurable! $\mathbb{E}(X|\mathcal{F}_t)$ is a special case of $\mathbb{E}(X|\mathcal{F}_T)$.

This becomes clear when we see the properties of stopped expectation. For $\forall X, Y, Z$ being positive random variables and for $\forall S, T$ stopping times of \mathcal{F} the following properties hold:

- **defining property:**

$$\mathbb{E}_T X = Y \iff V \in \mathcal{F}_T \text{ and } \mathbb{E} V X = \mathbb{E} V Y;$$

- **unconditioning:**

$$\mathbb{E}\mathbb{E}_T X = \mathbb{E} X;$$

- **repeated conditioning:**

$$\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_{\min\{S, T\}} X = \mathbb{E}_{S \wedge T} X;$$

- **conditional determinism:**

$$\mathbb{E}_T(X + YZ) = X + Y\mathbb{E}_T(Z)$$

if $X, Y \in \mathcal{F}_T$.

We need to prove the third property.

Proof

Notice that if $S \leq T \implies \mathcal{F}_S \subset \mathcal{F}_T$. This means that

$$\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_S (= \mathbb{E}_{S \wedge T})$$

(the “poorer” σ -algebra wins!). Moreover notice that if S and T are arbitrary we can apply the same result to $S \wedge T$ since $S \wedge T \leq T$ and

$$\mathbb{E}_{S \wedge T} \mathbb{E}_T X = \mathbb{E}_{S \wedge T} \quad (\bullet)$$

Remember that $Y = \mathbb{E}_T X$. Let's now try to write the statement of our proof in a different way:

$$\mathbb{E}_S Y = \mathbb{E}_{S \wedge T} X.$$

But using our property (\bullet) we get

$$\begin{aligned} \mathbb{E}_S Y &= \mathbb{E}_{S \wedge T} \mathbb{E}_T X \\ &= \mathbb{E}_{S \wedge T} Y. \end{aligned}$$

Let's prove this, by steps.

1. note that $\mathcal{F}_{S \wedge T} \subset \mathcal{F}_S$. This means that $\mathbb{E}_{S \wedge T} Y$ is in \mathcal{F}_S and therefore $\mathbb{E}_{S \wedge T} Y$ has the measure property and it is a candidate to be $\mathbb{E}_S Y$. So we should check whether this candidate satisfies the defining property of stopped expectation.
2. Check the defining property

$$\mathbb{E} V Y = \mathbb{E} V \mathbb{E}_{S \wedge T} Y$$

for each positive V in \mathcal{F}_S . We start by fixing such V , positive and in \mathcal{F}_S . We proved that

$$V \mathbf{1}_{\{S \leq T\}} \in \mathcal{F}_{S \wedge T}$$

and the defining property for $\mathbb{E}_{S \wedge T}$ gives

$$\mathbb{E} V \mathbf{1}_{\{S \leq T\}} Y = \mathbb{E} V \mathbf{1}_{\{S \leq T\}} \mathbb{E}_{S \wedge T} Y \quad (\bullet\bullet)$$

Now we observe that $Y \in \mathcal{F}_T$ by definition. Also notice that $Y \mathbb{1}_{\{T < S\}} \in \mathcal{F}_{S \wedge T}$. Apply now this fact to the defining property and obtain

$$\text{EV}Y \mathbb{1}_{\{T < S\}} = \text{EV} \mathbb{E}_{S \wedge T} Y \mathbb{1}_{\{T < S\}}$$

$\mathcal{F}_{S \wedge T}$ -meas.

and by conditional determinism we get

$$\begin{aligned}\text{EV}Y \mathbb{1}_{\{T < S\}} &= \text{EV} \mathbb{1}_{\{T < S\}} \\ &= \text{EV} \mathbb{1}_{\{T < S\}} \mathbb{E}_{S \wedge T} Y.\end{aligned}$$

By putting together this result with (••) we get

$$\text{EV}Y = \text{EV} \mathbb{E}_{S \wedge T} Y.$$

□

2.2 Martingales

When we study stochastic processes, the typical way of introducing classes of stochastic processes for which we have methods to study is determining the finite dimensional distributions. If we have them, thanks to the Kolmogorov Consistency theorem, we can give a complete description of the process but this is boring and for fucking losers. So we introduce classes of stochastic processes for which we can write the distribution.

Imagine, from a different view point, another approach: instead of starting from the distribution of a stochastic process we start from a property; then as a consequence of that property we infer the behavior towards infinity. This is the philosophy behind the class of **martingales**.

Example 2.2.1

Consider the process $\{X_n\}$ that describes the fortune of a player. This player wins with probability p and loses with probability $1 - p = q$. Consider the fortune at time n $\{S_n\}_{n \geq 0}$, which is

$$S_n = \sum_{i=0}^n X_i.$$

Is the game fair? Imagine also that the X_i 's are the value that I get by losing or selling options or stock. We know that the game is fair if $p = q = \frac{1}{2}$. But this could also be written as a property in the following way:

$$\mathbb{E} \left[S_{n+1} \mid \mathcal{F}_n \right] = S_n.$$

So I don't have any "advantages" or "bias".

Proof

$$\begin{aligned}\mathbb{E} [S_{n+1} | \mathcal{F}_n] &= \mathbb{E} [S_n + X_{n+1} | \mathcal{F}_n] \\ &= S_n + \mathbb{E} [X_{n+1} | \mathcal{F}_n] \\ &= S_n + \underbrace{\mathbb{E} [X_{n+1}]}_{=0 \text{ if } p=q=\frac{1}{2}} \\ &= S_n\end{aligned}$$

□

We will not define martingales through the event/filtration that allows the process Y to remain constant:

$$\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = Y_n.$$

Then we should ask some questions about this σ -algebra. We know that $\{\mathcal{F}_n\}$ can be increasing, constant or decreasing but Y must remain constant in mean. Look at it through the tower property.

$$\mathbb{E}Y_n = \mathbb{E}[\mathbb{E}[Y_{n+1}|\mathcal{F}_n]] = \mathbb{E}Y_{n+1}.$$

Definition 2.2.1

A real-valued process

$$X = (X_t)_{t \in \mathbb{T}}$$

is called a **\mathcal{F} -martingale** if:

1. it is adapted to \mathcal{F} ;
2. it is integrable for each $t \in \mathbb{T}$;
3. $\mathbb{E}(X_t - X_s|\mathcal{F}_s) = 0 \quad \forall s < t$.

If $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \geq 0 \quad \forall s < t$ then the process is called **\mathcal{F} -submartingale** and if $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \leq 0 \quad \forall s < t$ it is called **\mathcal{F} -supermartingale**.

These three properties are what we need to check to establish whether a process has martingale properties. The first two are really just regularity conditions, while the third is the “real” defining property of a martingale. There is another definition for the third property:

Definition 2.2.2

$$3'. \mathbb{E}(X_t|\mathcal{F}_s) = X_s.$$

In the first definition we were looking at the increment, now we are looking at the mean of a further point t in time for the process X and discovering that, conditioning it to information available in time s , it doesn’t change from the process in time s .

If $\mathbb{E}X_t^+ < \infty$ then the sequence $\{X_t\}$ (that can be a sub-martingale) converges to a limit.

Remark

A martingale is also a sub-martingale and a super-martingale.

We will use sub-martingales in our demonstrations but it really does not change anything because if $\{X_t\}$ is a \mathcal{F} -sub-martingale then $\{-X_t\}$ is a \mathcal{F} -super-martingale.

Remark

The number of terms in $\{X_t\}$ can be bounded or unbounded: even a process with 4 terms can be a martingale.

There are consequence to the way we defined martingales. Consider the times $s < t < u$ and the \mathcal{F} -martingale $\{X_t\}$. Then

$$\mathbb{E}_s(X_u - X_t)$$

means knowing the evolution of the process at time s and looking for the increment of the process between t and u .

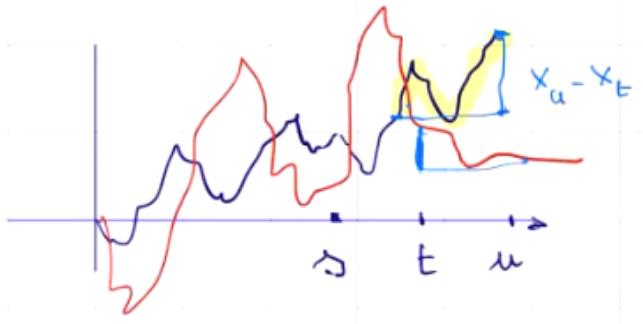


Figure 2.5: It should not

We can write

$$\begin{aligned}\mathbb{E}_s(X_u - X_t) &= \mathbb{E}_s\left(\mathbb{E}_t(X_u - X_t)\right) \\ &= 0 \text{ bc it is a martingale} \\ &= \mathbb{E}_s 0 = 0.\end{aligned}$$

This means that given the information up to time s the estimation of any future increment will be null! In the special case $\mathbb{T} = \mathbb{N}$ then this property becomes

$$\mathbb{E}_n(X_{n+1} - X_n) = 0.$$

Proposition 2.2.1

If $\{X_i\}$ is an \mathcal{F}_n -martingale, it holds

$$\begin{aligned}\mathbb{E}X_i &= \mathbb{E}X_{i+1} && \text{if it is a martingale} \\ \mathbb{E}X_{i+1} &> \mathbb{E}X_i && \text{if it is a sub-martingale.}\end{aligned}$$

Proof

Consider the expectation

$$\mathbb{E}(X_{i+1}|\mathcal{F}_i) \geq X_i$$

being a sub-martingale. Take the expectation

$$\underbrace{\mathbb{E}[\mathbb{E}(X_{i+1}|\mathcal{F}_i)]}_{\mathbb{E}X_{i+1}} \geq \mathbb{E}X_i.$$

We can now define the process $\{Y_i\}_{i \in \mathbb{N}}$ as

$$Y_{i+1} = X_{i+1} - X_i.$$

We know that $\{X_i\}$ is a \mathcal{F} -martingale. Note that

$$\sum_{j=1}^i Y_j = X_1 - X_0 + X_2 - X_1 + \dots + X_i - X_{i-1}$$

so that we can write

$$X_i = c + \sum_{j=1}^i Y_j.$$

We know that

$$\mathbb{E}[X_{i+1}|\mathcal{F}] = X_i$$

But we also know that

$$\mathbb{E}[Y_{i+1}|\mathcal{F}] = 0.$$

□

The quantity $\{\mathbb{E}Y_{i+1}|\mathcal{F}_i\}$ is called **martingale difference**.

Example 2.2.2

Imagine we have a sequence $\{Y_i\}$ of independent random variables with $\mathbb{E}Y_i = 0$. Consider \mathcal{F}_1 , the σ -algebra generated by $\{Y_j, 0 \leq j \leq i\}$. Call $S_i = c + \sum^i Y_j$. $\{S_i\}$ is an \mathcal{F} -martingale.

2.2.1 Properties of martingales

1. Consider X and Y being \mathcal{F} -sub-martingales and $a, b \in \mathbb{R}^+$. Then

$$aX + bY$$

is a \mathcal{F} -sub-martingale.

2. Consider X, Y being two \mathcal{F} -sub-martingales. Then

$$\max\{X, Y\}$$

is a \mathcal{F} -sub-martingale.

Proof

Consider $\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}]$. Remember that we have to prove:

- measurability property;
- integrability property;
- martingale property.

Often times measurability and integrability are implied in the form of the function. Since we are using the maximum between X and Y we know that

$$\begin{aligned}\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] &\geq \mathbb{E}[X_n|\mathcal{F}_{n-1}] \\ \mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] &\geq \mathbb{E}[Y_n|\mathcal{F}_{n-1}].\end{aligned}$$

But since X and Y are martingales we know that $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$ and $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] \geq Y_{n-1}$; hence we get that

$$\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] \geq \max\{X_{n-1}, Y_{n-1}\}$$

so the martingale property is satisfied. \square

3. Consider two \mathcal{F} -super-martingales X_n, Y_n . Then

$$\min\{X_n, Y_n\}$$

is a \mathcal{F} -super-martingale.

Proof

Consider $\max\{-X_n, -Y_n\}$. We know that $-X_n$ is a sub-martingale and so is $-Y_n$. So as proved above $\max\{-X_n, -Y_n\}$ is a sub-martingale. But $\max\{-X_n, -Y_n\} = -\min\{X_n, Y_n\}$ so

$$-\min\{X_n, Y_n\} \text{ is a sub-martingale} \iff \min\{X_n, Y_n\} \text{ is a super-martingale.}$$

\square

4. Consider a function f that is convex on \mathbb{R} . If X is a \mathcal{F} -martingale and $f \circ X$ is integrable then $f \circ X$ is a \mathcal{F} -sub-martingale.

Proof

We have $s < t$ and we want to study $\mathbb{E}[f(X_t)|\mathcal{F}_s]$. By Jensen's inequality we have

$$\begin{aligned}\mathbb{E}[f(X_t)|\mathcal{F}_s] &\geq f\left[\underbrace{\mathbb{E}(X_t|\mathcal{F}_s)}_{X_s}\right] \\ &= f(X_s).\end{aligned}$$

□

Remark

Some examples of positive functions of X are $X^+, X^-, |X|$. If X is martingale, these are sub-martingales. $|X|^p$ is a sub-martingale if X is a martingale and $\mathbb{E}|X|^p < \infty$.

5. if f is convex and increasing and X is a \mathcal{F} -sub-martingale with $f(X_t)$ integrable $\forall t$ then $f(X_t)$ is again a \mathcal{F} -sub-martingale.

2.2.2 Create your own martingale

1. Consider an increasing family of σ -algebras $\{\mathcal{F}_i\}$ and a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ which is integrable and adapted to \mathcal{F} . Define

$$X_i = \mathbb{E}[X|\mathcal{F}_i].$$

Then the process $\{X_i\}_{i \in \mathbb{N}}$ is called **Doob martingale** and plays an important role in many processes. As i increases, so does the information about our random variable while retaining the underlying process X . We have now to prove that $\{X_i\}$ is a \mathcal{F} -martingale.

Proof

We need to check that

$$\begin{aligned}\mathbb{E}[X_{i+1}|\mathcal{F}_i] &= \mathbb{E}[\mathbb{E}[X|\mathcal{F}_{i+1}]|\mathcal{F}_i] \\ &= \mathbb{E}[X|\mathcal{F}_i]\end{aligned}$$

and this is because of the properties of conditional expectation since $\mathcal{F}_{i+1} \supset \mathcal{F}_i$. But $\mathbb{E}[X|\mathcal{F}_i] = X_i$ by definition. □

The same result holds when \mathbb{T} is continuous.

2. Consider sums of independent random variables $\{X_i\}$ with $\mathbb{E}[X_i] = 0$. We have

$$S_0 = 0 \quad S_n = S_0 + X_1 + X_2 + \dots + X_n \quad n \geq 1.$$

We define $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ as the filtration generated by $S = (S_n)_{n \in \mathbb{N}}$. Consider the following facts:

- (a) \mathcal{F}_n is adapted to S ;
- (b) Each S_n is integrable and $\mathbb{E}S_n = 0$;
- (c) $\mathbb{E}_n(S_{n+1} - S_n) = \mathbb{E}_n(X_{n+1}) = \mathbb{E}X_{n+1} = 0$ (because of independence).

So S_n is a martingale. What happens, though, if $\mathbb{E}X_i = 1$ for each i ? We would get a sub-martingale (because we would have an increasing process).

3. Consider the product of random variables. Consider R_1, R_2, \dots being independent random variables with

$$\mathbb{E}R_i = 1 \quad \text{Var } R_i < \infty$$

that don't necessarily need to be identically distributed. Let $M_0 = 1$ and $M_n = M_0 \cdot R_1 \cdot R_2 \cdot \dots \cdot R_n$ and let \mathcal{F} be the filtration generated by $M = \{M_n\}_{n \in \mathbb{N}}$. Observe that

- (a) M is adapted to \mathcal{F} by definition;

- (b) $\forall M_n$ is integrable: consider $\|M_n\| \leq \|M_{n-1}\| \|R_n\|$ since $M_n = M_{n-1}R_n$. But since R_n has finite variance then $\|M_{n-1}\| \|R_n\| < \infty$. So since it is bounded it is integrable;
(c) consider the expectation

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}[M_n R_{n+1}|\mathcal{F}_n] \\ &= M_n \mathbb{E}[R_{n+1}|\mathcal{F}_n] \quad (\text{because } M_n \text{ is } \mathcal{F}_n\text{-meas.}) \\ &= M_n \mathbb{E}[R_{n+1}] \\ &= M_n.\end{aligned}$$

Example 2.2.3

This is considered a good model for finance: M_n is the price of a share at time n and R_{n+1} is the return at time $(n+1)$ per euro invested at time n . \mathcal{F}_n is the information that is available to the whole market.

4. Consider functions of sums. Consider $\{X_n\}$, a sequence of i.i.d. random variables with

$$\mathbb{E}X_i = 0 \quad \text{Var } X_i = \sigma^2 < \infty \quad \forall i.$$

Consider

$$S_n = \sum_{i=1}^n X_i$$

and define

$$M_n = S_n^2 - n\sigma^2.$$

M_n is a \mathcal{F} -martingale. This is useful to study for example the occurrence of the time in which something happens for the first time. Consider:

- (a) M_n is integrable ($\text{Var } X_i = \sigma^2$);
(b) M_n is adapted by definition;
(c) consider the expectation

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[\underbrace{(S_n + X_{n+1})^2}_{S_{n+1}^2} - (n+1)\sigma^2|\mathcal{F}_n\right] \\ &= \mathbb{E}\left[S_n^2 + N_{n+1}^2 + 2S_n X_{n+1} - (n+1)\sigma^2|\mathcal{F}_n\right] \\ &= S_n^2 + 2S_n \underbrace{\mathbb{E}(X_{n+1}|\mathcal{F}_n)}_{=0} + \underbrace{\mathbb{E}[X_{n+1}^2|\mathcal{F}_n]}_{\mathbb{E}[X_{n+1}^2]=\sigma^2} - (n+1)\sigma^2 \\ &= S_n^2 + \cancel{\sigma^2} - n\sigma^2 - \cancel{\sigma^2} = M_n.\end{aligned}$$

Example 2.2.4

Consider $(X_n)_{n \geq 1}$ being i.i.d. random variables with common probability density function $g(x)$. Let $f(x)$ be another probability density function with the property

$$f(x) = 0 \quad \text{for } \frac{x}{g(x)} = 0.$$

Define

$$\begin{aligned}L_n &= \prod_{i=1}^n \frac{f(X_i)}{g(X_i)} \\ L_0 &= 1.\end{aligned}$$

likelihood ratio

Likelihood ratio is a \mathcal{F} -martingale where \mathcal{F} is the filtration generated by $\{X_n\}$. Contemplate:

- (a) it is integrable by definition;
(b) it is adapted by definition;

(c) consider the expectation:

$$\begin{aligned}
 \mathbb{E}\left[L_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[L_n \frac{f(X_{n+1})}{g(X_{n+1})|\mathcal{F}_n}\right] \\
 &= L_n \mathbb{E}\left[\frac{f(X_{n+1})}{g(X_{n+1})}|\mathcal{F}_n\right] \quad (\text{because } L_n \text{ is } \mathcal{F}_n\text{-meas.}) \\
 &= L_n \mathbb{E}\left[\frac{f(X_{n+1})}{g(X_{n+1})}\right] \\
 &= L_n \underbrace{\int_{\mathbb{R}} \frac{f(x)}{g(x)} g(x) dx}_{=1} = L_n.
 \end{aligned}$$

This really is exhausting. Like it literally drains the life from me

2.2.3 Doob's decomposition

We may want to understand whether there is any relationship between a stochastic process and a martingale. In this part our set \mathbb{T} will be \mathbb{N} .

Up to now we said that a process is adapted to \mathcal{F} if X_n is \mathcal{F}_n -measurable.

Definition 2.2.3

The process $F = (F_n)_{n \geq 1}$ is **\mathcal{F} -predictable** if $F_0 \in \mathcal{F}_0$ and $\mathcal{F}_{n+1} \in \mathcal{F}_n$, for $\forall n \in \mathbb{N}$.

This means that the available information up to n is “enough” to have a bet in the period $n + 1$. Some predictable processes are:

- any deterministic processes;
- consider two stopping times S, T of \mathcal{F} and let $S \leq T$. Consider the random variable V in \mathcal{F}_n . Then

$$(1) \quad V \mathbf{1}_{(S,T]} \quad (2) \quad V \mathbf{1}_{(S,\infty]} \quad (3) \quad \mathbf{1}_{(S,T]} \quad (4) \quad \mathbf{1}_{(0,T]}$$

are predictable processes.

Proof

(2) Consider $F = V \mathbf{1}_{(S,\infty]}$, so that $F_n = \mathbf{1}_{(S,\infty]}(n)$. Consider

$$\begin{aligned}
 F_{n+1} &= V \mathbf{1}_{(S,\infty]}(n+1) \\
 &= V \mathbf{1}_{\{S < n+1\}} \\
 &= V \mathbf{1}_{\{S \leq n\}} \in \mathcal{F}_n
 \end{aligned}$$

so the process is \mathcal{F} -measurable and predictable.

(1) Since we know by hypothesis that $S \leq T$ then $V \in \mathcal{F}_S \subset \mathcal{F}_T$. This means that $V \in \mathcal{F}_T$. Hence, of a consequence,

$$V \mathbf{1}_{(T,\infty]} - V \mathbf{1}_{(S,\infty]} = V \mathbf{1}_{(S,T]}$$

is predictable.

(3) Take $V = 1$.

(4) Take $T = \infty$, $V = 1$. $\mathbf{1}_{(S,\infty]}$ is predictable. But then

$$\mathbf{1}_{[0,S]} = \mathbf{1} - \mathbf{1}_{(S,\infty]}$$

so it is predictable.

□

Theorem 2.2.1

Doob's decomposition.

X is a stochastic process which is adapted to \mathcal{F} and integrable. Then

1. it can be decomposed as

$$X_n = X_0 + M_n + A_n, \quad n \in \mathbb{N}$$

where:

- M_n is a \mathcal{F} -martingale with $M_0 = 0$;
- A_n is a predictable process with $A_0 = 0$.

2. The decomposition is unique up to equivalence.
3. If X_n is a sub-martingale then $\{A_n\}_{n \geq 0}$ increasing, while if X_n is a super-martingale then $\{A_n\}_{n \geq 0}$ is decreasing.

Proof

Put $A_0 = M_0 = 0$. Define M and A through their increments:

$$\begin{aligned} A_{n+1} - A_n &= \underbrace{\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n]}_{\mathcal{F}_n\text{-meas.}} \\ M_{n+1} - M_n &= (X_{n+1} - X_n) - (A_{n+1} - A_n). \end{aligned}$$

If we look at these quantities we see that A is predictable and M is martingale. Imagine now there is another decomposition: let

$$X = X_0 + M' + A'$$

be another decomposition. We must have

$$X_0 + M' + A' = X_0 + M + A \iff A - A' = M - M' = B.$$

Now B is a process and it is predictable and martingale (because it is the difference between two martingales). Since B is predictable *and* a martingale we have

$$\begin{aligned} B_{n+1} - B_n &= \underbrace{\mathbb{E}[B_{n+1} - B_n | \mathcal{F}_n]}_{\text{predictable}} \\ &= 0 \\ &\stackrel{\text{martingale}}{\implies} B_{n+1} = B_n = B_0 \text{ a.s., } A = A' \text{ a.s., } M = M' \text{ a.s.} \end{aligned}$$

If X is a sub-martingale then we have

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \geq 0$$

and this means that we have $A_{n+1} \geq A_n$ is increasing. □

Consider the increments on X :

$$X_{n+1} - X_n = \underbrace{A_{n+1} - A_n}_{\text{prediction term}} + \underbrace{M_{n+1} - M_n}_{\text{innovation}}.$$

Engineers, much better people than us, often use this terms.

2.2.4 Important stochastic processes and related martingales

A quick note before starting:

Markov chains

Consider a stochastic process $\{X_n\}_{n \geq 1}$. All trials are on the same probability space $(E_n, \mathcal{E}_n) = (E, \mathcal{E})$. Define a **Markov kernel** on (E, \mathcal{E})

$$(E, \mathcal{E}) \mapsto (E, \mathcal{E})$$

such that

$$K_n(x_0, x_1, \dots, x_n; A) = \mathbb{P}(x_n, A) \quad \forall n \in \mathbb{N}, \forall (x_0, x_1, \dots, x_n) \in E, A \in \mathcal{E}.$$

So the probability that governs the evolution is dependent *only* on the last state and nothing else. If $x_{n+1} = j, x_n = i$ then $\mathbb{P}(i, j) = p_{ij}$. The process $X = \{X_n\}_{n \geq 0}$ is said to be a **Markov chain** over $(\Omega, \mathcal{H}, \mathbb{P})$ with state space (E, \mathcal{E}) , initial distribution μ and transition kernel P .

$$P = (p_{ij})_{i,j}$$

Now consider a Markov chain with transition probability matrix P such that

$$P \cdot \underbrace{f}_{\text{eigenvector}} = \underbrace{\lambda}_{\text{eigenvalue}} \cdot f.$$

This means that we have the following system of equations:

$$\begin{cases} \sum_j p_{ij} f(j) = \lambda f(i) & i = 1 \\ \vdots & i = 2 \\ \vdots & \ddots \end{cases}$$

We can clearly write these expressions in form of expectation:

$$\mathbb{E}[f(X_{n+1})|X_n = i] = \lambda f(i).$$

We can also write it without fixing i :

$$\mathbb{E}[f(X_{n+1})|X_n] = \lambda f(X_n). \quad (\bullet)$$

Look at the last equation: it is not too different from the martingale property. Consider the ratio $\frac{f(X_n)}{\lambda^n}$ and the filtration generated by $X \sigma(X_0, X_1, \dots, X_n)$. Thanks to the relation given by the equation (\bullet) it is a martingale with respect to $\sigma(X_0, X_1, \dots, X_n)$.

Proof

Consider the expectation

$$\begin{aligned} \mathbb{E}\left[\frac{f(X_{n+1})}{\lambda^{n+1}} \middle| X_0, \dots, X_n\right] &= \frac{1}{\lambda^n} \frac{1}{\lambda} \underbrace{\mathbb{E}[f(X_{n+1}|X_n)]}_{\lambda f(X_n)} \\ &= \frac{1}{\lambda^n} \frac{1}{\lambda} \lambda f(X_n) \\ &= \frac{1}{\lambda^n} f(X_n) \\ &= Y_n \end{aligned}$$

where the first equality is thanks to Markov property. This is a general property for Markov chains. \square

Example 2.2.5

Branching processes. We study the life of a generation of individual.

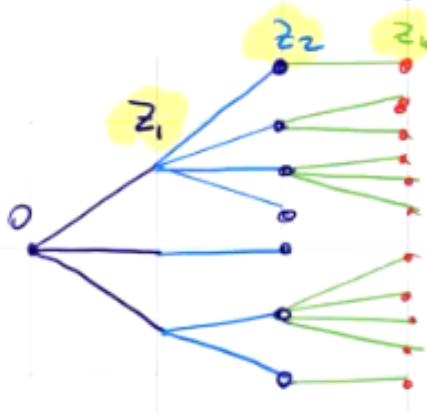


Figure 2.6: I don't even know anymore.

The offsprings have a distribution $\{p_k\}$. The mean number of offsprings per individual is $m = \sum k p_k$. Consider the i.i.d random variables $\{Z^{(n)}(i), n \geq 0\}$ describing the number of children of the i -th individual of generation n . Define

$$Z_{n+1} = \begin{cases} Z^{(n)}(1) + Z^{(n)}(2) + \dots + Z^{(n)}(Z_n) & \text{if } Z_n \neq 0 \\ 0 & \text{if } Z_n = 0 \end{cases}$$

We immediately realize that this is a Markov process because every generation's number of children can be entirely derived from the previous generation's law of reproduction. I don't need to know whose child the individual is. So Z_n (number of individual of the n -th generation) gives rise to a Markov chain.

$$\sum_{j=0}^{\infty} p_{ij} j = m \cdot i$$

Is the equation that tells us the complete number of individuals. Now consider $Pf = mf$ with $f(j) = j$. So we have $Pj = mi$ where m plays the role of λ . The process $\{\frac{Z_n}{m^n}, \sigma(Z^{(0)}, \dots, Z^{(n)})\}$ is a martingale.

Poisson martingale

Consider \mathbb{R}_+ as our index set. \mathcal{F} is our filtration and we consider the counting process $N = (N_t)_{t \geq 0}$ (it counts the number of events up to time t). It has unit jumps and any path starts from 0 ($N_0(\omega) = 0$), is increasing and right continuous.

Definition 2.2.4

The counting process N is said to be a **Poisson process** with rate λ with respect to \mathcal{F} if it is adapted to \mathcal{F} and

$$\mathbb{E}[f(N_{t+s} - N_s) | \mathcal{F}_s] = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} f(k) \quad \forall s, t \in \mathbb{R}_+, \forall \text{positive } f \mapsto \mathbb{N}.$$

Theorem 2.2.2

Let N be a counting process. It is a Poisson process with rate λ with respect to \mathcal{F} if and only if:

$$M_t = N_t - \lambda t$$

is a \mathcal{F} -martingale.

We only prove that M_t is a martingale if N_t is Poisson.

Proof

We know that

$$\begin{aligned}
 \mathbb{E}[M_t | \mathcal{F}_s] &= \mathbb{E}[M_t - M_s + M_s | \mathcal{F}_s] \\
 &= \mathbb{E}[M_t - M_s | \mathcal{F}_s] + M_s \\
 &= \mathbb{E}[M_t - M_s] + M_s \\
 &= \mathbb{E}[N_t - N_s + \lambda t + \lambda s] + M_s \\
 &= \underbrace{\mathbb{E}[N_t - N_s]}_{\lambda(t-s)} - \lambda(t-s) + M_s \\
 &= M_s
 \end{aligned}$$

□

We may use the Doob's decomposition theorem to get $N_t = M_t + \lambda t$ where λt is our predictable process. So the same result can be achieved from two different viewpoints.

2.2.5 Integration in discrete time

Let us consider two real-valued processes $M = (M_n)_n$ and $F = (F_n)_n$ and let us define

$$X_n = F_0 M_0 + (M_1 - M_0) F_1 + \dots + (M_n - M_{n-1}) F_n.$$

We say that $\{X_n\}$ is the integral of F with respect to M and we write

$$X_n = \int F \, dM$$

where dM is a random signed measure. Remember the Lebesgue-Stieltjes integral? Me neither, but as long as M has bounded variation this is a Lebesgue-Stieltjes integral.

Yeah no wait hold your fucking horses.

What's up now?

First of all how is that shit an integral?? And what the fuck is a Lebesgue-Stieltjes integral?

I should have NEVER let you enroll in this master.

Professor Lods, I will exact my revenge on Università di Torino very soon.

Okay okay. So a little explanation is due since I actually never saw a Stieltjes integral. Too busy doing things, y'know. Things like, for example, sex. Heh. Yeah. Sex.

Revise with Kotatsu!

Oookay so very shortly: a Riemann-Stieltjes integral is a generalization of our dear old Riemann integral. It is notated as

$$\int_a^b f(x)dg(x).$$

Actually this is not too different from the concept of the Lebesgue integral. Basically it's about partitioning the interval (a, b) and then summing for every partition (that is, for each i) each value of $f(x_i)$ multiplied for $[g(x_i) - g(x_{i-1})]$. Just take the limit for the partition to be infinitely small and voilà, le intégral c'est servie (I don't know French). El integral està servido (I know a little bit of Spanish). Omnibus integral paratum est ut facile intellegatur (I actually know Latin really well. Useless skills are my strength).

Anyway what's important is: *we take the function $f(x)$ and use it to calculate how much the variation of $g(x)$ impacts on the total sum.* And if $g(x) = x$, well... We have our cute little Riemann integral!

Yeah after this smug declaration of sterile knowledge we are maybe ready to better understand how the discrete time integral is, well, an integral. Our function $g(x)$ is M so we consider the variation of M ; since we are working in discrete time we will never reach a partition so fine it can be called continuous. So it's like... a *diet* Riemann-Stieltjes integral.

Theorem 2.2.3

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 2.2.6

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n-1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^n (M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k , but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

- We have M being a martingale and $F_0, F_1, \dots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \dots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to \mathcal{F} .

2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b . This implies

$$|X_n| < b(|M_0| + |M_1 + M_0| + \dots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

We know that using a policy that it is predictable it is impossible to beat the system. Finance bros try to overcome this possibility using stopping times. If my policy is not only based on a predictable process but I add the randomness of the time in which I decide to sell or buy can I break the curse of martingales and make shareholders want to suck my dick? Well, it depends.

2.2.6 Stopped martingales and Doob's stopping theorem

Definition 2.2.5

Define $M = (M_n)_{n \in \mathbb{N}}$ as a process and let T be a random time with values on $\overline{\mathbb{N}}$. The process

$$X_n(\omega) = M_{n \wedge T}(\omega) = \begin{cases} M_n(\omega) & n < T(\omega) \\ M_T(\omega) & n > T(\omega) \end{cases}$$

(where $n \wedge T$ is a truncated random time) is called **M stopped at T** .

As a consequence X is exactly the discrete time integral if $F = \mathbb{1}_{[0,T]}$:

$$X_n = \underbrace{M_0 F_0}_0 + (M_1 + M_0) \mathbb{1}_{[0,T]}(1) + \dots + (M_n - M_{n-1}) \mathbb{1}_{[0,T]}(n).$$

The indicators only select the current time interval. If this is the case we can observe that $F_{[0,T]}$ is bounded, positive and predictable. Hence if M is a martingale the theorem applies with this special choice of M and we can write the result as a different theorem:

Theorem 2.2.4

Let T be a stopping time and let X be the process M stopped at T . If M is a martingale then so is X (the same holds for sub-martingales and super-martingales).

So we cannot determine a policy based on stopping times that can change the nature of our martingale. In the remote case in which you are interested in this you can read Williams - Introduction to martingales. A further theorem about this is **Doob's stopping theorem**. For a martingale we know

$$\mathbb{E}[X_t - X_s | \mathcal{F}_s] = 0.$$

The question is: is this true also if s and t are substituted by stopping times $S, T, S \leq T$?

Theorem 2.2.5

Let M be adapted to \mathcal{F} . Then the following are equivalent:

- ① M is a submartingale;
- ② for every bounded stopping time $S \leq T$ the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0;$$

③ for each pair of bounded stopping times the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S] \geq 0.$$

Remark

If M is a martingale the theorem can be read in a different way:

$$\mathbb{E}M_T = \mathbb{E}M_S = \mathbb{E}M_0.$$

Previously, $\mathbb{E}M_n = \mathbb{E}M_{n-1} = \mathbb{E}M_0$.

Proof

To prove the theorem we need to show that from condition 1 follows 2 from which follows 3 from which follows 1.

1 → 2 by hypothesis M is a sub-martingale and our thesis is that if $S(\omega) < T(\omega) < n$ (because we asked for bounded times) then:

- (a) M_S and M_T are integrable;
- (b) $\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0$.

We know that S and T are bounded by n . Let V be a positive bounded random variable and define $F = V\mathbf{1}_{(S,T]}$ and use it in the discrete time integral:

$$X_n = \underbrace{M_0 F_0}_{X_0} + \underbrace{(M_1 - M_0)}_{V\mathbf{1}_{\{1 \in (S,T)\}}} \underbrace{F_1}_{+ \dots + (M_n - M_{n-1})F_n}$$

So we have

$$X_n - X_0 = V(M_T - M_S)$$

So X_n is a sub-martingale. Take $V = 1$ and $S = 0$: now we have

$$\begin{array}{ccc} X_n & - & X_0 \\ \downarrow & & \downarrow \\ \text{int.} & & \text{int.} \end{array} = M_T \implies M_T \text{ is integrable.}$$

Now take $V = 1$ and $T = n$ so that we get

$$\begin{array}{ccc} X_n & - & X_0 \\ \downarrow & & \downarrow \\ \text{int.} & & \text{int.} \end{array} = M_n - M_S \implies M_S \text{ is integrable.}$$

We recall that $V \in \mathcal{F}_S$ and we use the defining property for $\mathbb{E}(\cdot | \mathcal{F}_S)$. So we can write

$$\begin{aligned} \mathbb{E}\mathbb{E}(M_T - M_S | \mathcal{F}_S) &= \mathbb{E}V(M_T - M_S) \\ &\stackrel{\text{def. prop.}}{=} \mathbb{E}[X_n - X_0] \\ &\stackrel{\text{discr. time int.}}{=} \mathbb{E}[X_n - X_0] \\ &\geq 0 \end{aligned}$$

proved above

and this is true $\forall V > 0, V < b, V \in \mathcal{F}_s$. Hence

$$\mathbb{E}(M_T - M_S | \mathcal{F}_S) \geq 0$$

So 1 → 2.

2 → 3 We can use the tower rule. Take the expectation of point 2:

$$\mathbb{E}[\mathbb{E}(M_T - M_S | \mathcal{F}_S)] = \mathbb{E}[M_T - M_S] \geq 0.$$

$3 \rightarrow 1$ Let 3 hold, so that $\mathbb{E}[M_T - M_S] \geq 0$. Choose $T = n$ and $S = 0$. Then M_n is integrable. Move to adaptness: this holds by hypothesis. Move to the martingale inequality:

$$\mathbb{E}[M_n - M_m | \mathcal{F}_m] \geq 0.$$

Note that this is equivalent to prove

$$\mathbb{E}[\mathbf{1}_H \mathbb{E}[M_n - M_m | \mathcal{F}_m]] \geq 0 \quad H \in \mathcal{F}_m, 0 \leq m \leq n.$$

Fix H, m, n and define

$$S(\omega) = m \quad T(\omega) = n \mathbf{1}_H(\omega) + m \mathbf{1}_{\{\Omega \setminus H\}}(\omega)$$

The indicators are non-zero in complementary instances. Notice that:

- (a) S is a fixed time so it is a stopping time;
- (b) $S \leq T \leq n$ by definition of S and T because the indicators are non-zero in complementary instances;
- (c) $T \geq S$ is a foretold time by $S = m$;
- (d) $H \in \mathcal{F}_S$ by definition.

So we can write $M_T - M_S = \mathbf{1}_H(M_n - M_m)$ where $M_T - M_S \geq 0$ by hypothesis. This means that we have

$$\underbrace{\mathbb{E}[\mathbf{1}_H \mathbb{E}(M_n - M_m | \mathcal{F}_m)]}_{\mathbb{E}[M_n - M_m | \mathcal{F}_m] \geq 0} \geq 0$$

□

2.2.7 Oscillations

We will talk about oscillations in a strip (a, b) during a time interval. Consider an adapted process M and two real numbers $a < b \in \mathbb{R}$. We want to set a level a and a level b . First, by convention, fix $T_0 = -1$. Then define:

$$\begin{aligned} S_k &:= \inf\{n > T_{k-1} : M_n \leq a\} \\ T_l &:= \inf\{n > S_k : M_n \geq b\}. \end{aligned}$$

So it is the first time in which the process crosses the threshold a or b after having crossed the other threshold. We call S_k **downcrossing time** while we call T_k **upcrossing time**.

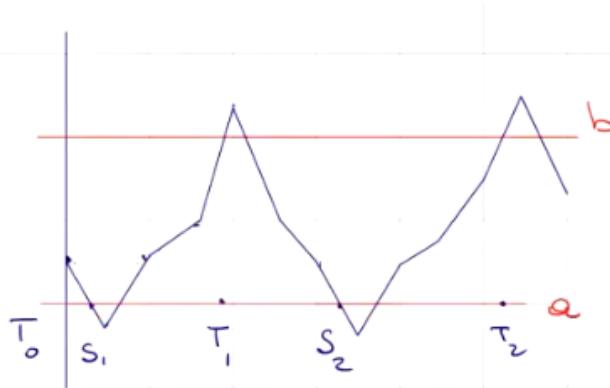


Figure 2.7: Up and downs. Just like my consumption of cocaine and my mood respectively on an average tuesday evening.

Now consider the sequence $(T_0, S_1, T_1, S_2, T_2, \dots)$. It is an increasing sequence of stopping times. If we want to focus on the number of oscillations we may count how many times we go from an upcrossing to a downcrossing time (or vice-versa). Say we want to count the upcrossing. We can define a quantity

$$U_n(a, b) = \sum_{k=1}^{\infty} \mathbf{1}_{(0, n]}(T_k)$$

which of course is a counting process.

Example 2.2.7

We are still with our financebros. Suppose that (M_n) is the price of an asset at time n . We want to buy when the price is below a at time S_i and sell when it is above b at time T_i . In $(0, n]$ we have $U_n(a, b)$ cycles of buying and selling so our strategy could consist in holding a number F_n of shares during period $(m - 1, m]$. This means introducing

$$F_m = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}(n)$$

with $F_0 = 0$. We can thus trace the evolution of our capital with the discrete-time integral:

$$X = \int F \, dM$$

And the profit during $(0, n]$ will be $X - X_0$.

The profit is at least

$$(b - a)U_n(a, b).$$

Proposition 2.2.2

If M is a sub-martingale with respect to its natural filtration then

$$(b - a)\mathbb{E}U_n(a, b) \leq \mathbb{E}[(M_n - a)^+ - (M_0 - a)^+].$$

We wanted to find a bound for our profit, but our profit is a stochastic quantity: so it's only natural to think about the expectation to give a bound to the number of expected up/downcrossing.

Observe that the number of upcrossings does not depend on the value of T_0 that we fix.

Proof

Choose $a = 0$. Consider hence the process $(M - a)^+$ that is sub-martingale (if M is a sub-martingale).

Take $M \geq 0$ and let

$$F_n = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}(n)$$

and consider

$$X = \int F \, dM$$

like in our example. We know that F is predictable since by definition $F_{k+1} \in \mathcal{F}_k$. Consider the expectation of the increment

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] = \mathbb{E}[F_{k+1}(M_{k+1} - M_k) | \mathcal{F}_k].$$

But since F_{k+1} is predictable we can take it out the expectation:

$$F_{k+1}\mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

But since F_{k+1} is an indicator we know it is ≤ 1 :

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] \leq \mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

Now take the expectation of both sides:

$$\mathbb{E}[X_{k+1} - X_k] \leq \mathbb{E}[M_{k+1} - M_k].$$

If we sum these inequalities over k we get:

$$\mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0].$$

So we now get that

$$bU_n(a, n) \leq \mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0]$$

but given that $a = 0$ we get that

$$bU_n(0, b) \leq \mathbb{E}[M_n - M_0].$$

Clearly we have to take the positive part. \square

This characterizes our martingale and its boundedness. The number of oscillations of a sub-martingale is bounded! The next question is: can we say anything about the behaviour of maximum/minimum of a martingale or sub-martingale?

Remark

Consider a sequence $\{X_n\}$ of independent random variables with $\mathbb{E}X_n = 0$, $S_n = \sum X_i$. We proved that

$$a^2 \mathbb{P}(\max_{k \leq n} |S_k| > a) \leq \text{Var } S_n$$

and we called this **Kolmogorov's inequality**.

What we are doing here is considering a random walk whose jumps have 0 mean. We wonder whether it is above the level a as seen in figure 4.7. About this, we know that $\{S_n\}$ is a \mathcal{F} -martingale but in proving the

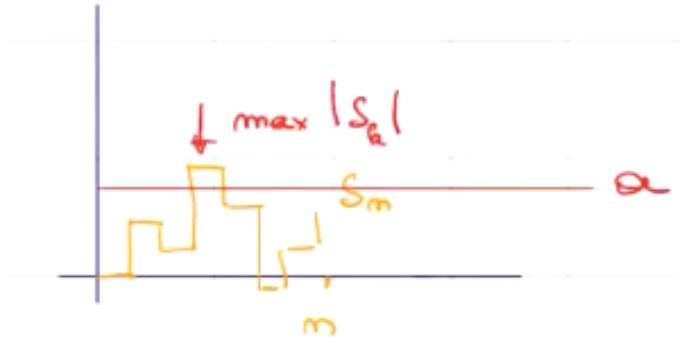


Figure 2.8: The maximum of the random walk.

Kolmogorov's inequality we never talked about the martingale property! We could improve this inequality using the **Doob's martingale inequality**. The problem is that we can prove very general inequalities that hold true for any random variables but specifying more characteristic we can obtain stricter bounds. In this framework let's define

$$\begin{aligned} M_n^* &= \max_{k \leq n} M_k \\ m_n^* &= \min_{k \leq n} M_k \end{aligned}$$

as current maximum and current minimum of M .

Theorem 2.2.6

Take M as a sub-martingale. For $b > 0$ it holds:

1. $b\mathbb{P}(M_n^* \geq b) \leq \mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq b\}}] \leq \mathbb{E}[M_n^+];$
2. $b\mathbb{P}(m_n^* \leq -b) \leq -\mathbb{E}M_0 + \mathbb{E}[M_n \mathbf{1}_{\{m_n^* \leq -b\}}] \leq \mathbb{E}M_n^+ - \mathbb{E}M_0.$

So we can further bound the result looking into the property of M_n .

Example 2.2.8

Prof. Sacerdote fucked up so we now need to define the brownian motion (or Weiner process).

Definition 2.2.6

A real-valued stochastic process $B = (B_t)_{t \geq 0}$ is called **brownian motion** if:

1. the index set is \mathbb{R}^+ ;
2. $B_0(\omega) = 0$ for almost all ω ;
3. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent for $\forall 0 = t_0 < t_1 < \dots < t_n < \infty$;
4. $B_t - B_s \sim B_{t+b} - B_{s+b}$ for every $0 \leq s < t < \infty \quad \forall n > -s$;
5. $B_t - B_s \sim N(0, t - s)$;
6. $t \mapsto B_t(\omega)$ are continuous for every ω .

The brownian motion is in itself a martingale, but there is a class of martingales strictly related to it:

$$M_t = \exp \left\{ \lambda B_t - \frac{1}{2} \lambda^2 t \right\}, \quad t \in \mathbb{R}^+.$$

Now we can get to the actual example: if B is a brownian motion, we have

$$\mathbb{P} \left(\sup_{0 \leq t \leq \theta} B_t \geq b \right) \leq \exp \left\{ -\frac{b^2}{s\theta} \right\}.$$

The sample paths of brownian motions are extremely irregular. We are asking with which probability the max of our process will be over b at time θ .

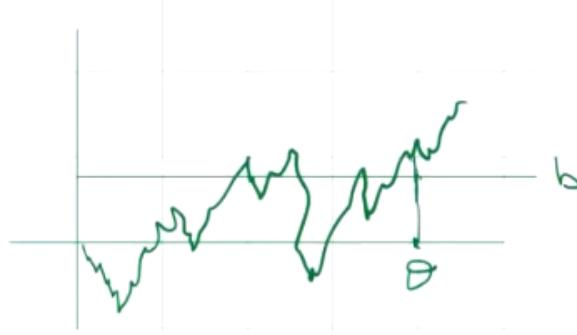


Figure 2.9: Check out Autechre. A staple in electronic experimental music.

I am basically asking if the maximum attained is above or below b . How can we use the Doob's inequality?

Proof

Consider

$$\mathbb{P}(\sup B_t \geq b) = \mathbb{P}(\sup e^{\lambda B_t} \geq e^{\lambda b})$$

and using Doob's inequality

$$\begin{aligned} \mathbb{P}(\sup B_t \geq b) &= \mathbb{P}(\sup e^{\lambda B_t} \geq e^{\lambda b}) \\ &\leq \underbrace{\mathbb{E}\left[e^{\lambda B_\theta - \frac{\lambda^2}{2}\theta}\right]}_{\text{martingale}} \\ &= e^{\lambda b} e^{-\frac{\lambda^2}{2}\theta} \\ &= e^{-\lambda b + \frac{\lambda^2}{2}\theta} \quad \forall \lambda > 0 \end{aligned}$$

□

Now we have to prove Doob's inequality.

Proof

We introduce 2 stopping times:

$$\begin{aligned} T &= \inf\{n \geq 0 : M_n \geq b\} \\ S &= \inf\{n \geq 0 : M_n \leq -b\}. \end{aligned}$$

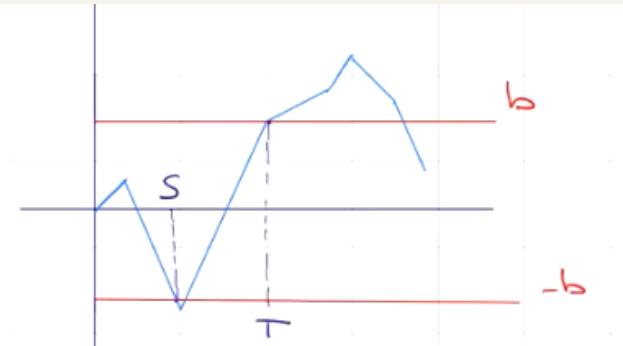


Figure 2.10: Also Aphex Twin. Really, Autechre and Aphex Twin are probably the best example of the boom of experimentation in popular electronic music between 1995 and 2005.

Consider the current maximum and minimum above the level b :

$$\{M_n^* \geq b\} = \{T \leq n\} \quad \{m_n^* < -b\} = \{S \leq n\}.$$

Fix b and n . When $\{T \leq n\}$ we have

$$M_{T \wedge n} = M_T \geq b.$$

Multiply by $\mathbf{1}_{\{T \leq n\}}$:

$$b \mathbf{1}_{\{T \leq n\}} \leq M_{T \wedge n} \mathbf{1}_{\{T \leq n\}}.$$

Using Doob's stopping theorem we know that

$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0$$

and using $S = T \wedge n$ and $T = n$ we obtain

$$\mathbb{E}[M_n | \mathcal{F}_{T \wedge n}] > M_{T \wedge n}.$$

So summing it all up:

$$\begin{aligned} b \mathbb{1}_{\{T \leq n\}} &\leq M_{T \wedge m} \mathbb{1}_{\{T \leq n\}} \\ &\leq \mathbb{1}_{\{T \leq n\}} \mathbb{E}[M_n | \mathcal{F}_{T \wedge n}] \\ &= \mathbb{E}[M_n \mathbb{1}_{\{T \leq n\}} | \mathcal{F}_{T \wedge n}]. \end{aligned}$$

Now take the expectation:

$$\begin{aligned} b \mathbb{P}(T \leq n) &= b \mathbb{P}(M_n^* \geq b) \\ &\leq \underbrace{\mathbb{E}[M_n \mathbb{1}_{\{T \leq n\}}]}_{\mathbb{E}[M_n \mathbb{1}_{\{M_n^* \geq b\}}]} \\ &\leq \mathbb{E}(M_n^*). \end{aligned}$$

For the minimum we work on $\{S \leq n\}$ and we get

$$\begin{aligned} M_{S \wedge n} &= M_S \mathbb{1}_{\{S \leq n\}} + M_S \mathbb{1}_{\{S > n\}} \\ &\leq -b \mathbb{1}_{\{S \leq n\}} + M_n \mathbb{1}_{\{S > n\}}. \end{aligned}$$

Now take the expectation:

$$\mathbb{E} M_{S \wedge n} \leq -b \mathbb{P}(m_n^* \leq -b) + \mathbb{E}[M_n \mathbb{1}_{\{S > n\}}].$$

So that we get

$$\mathbb{P}(m_n^* \leq -b) \leq -\mathbb{E} M_{S \wedge n} + \mathbb{E}[M_n \mathbb{1}_{\{S > n\}}].$$

Now use Doob's stopping theorem with $T = 0$ and $S = S \wedge n$ so that $\mathbb{E} M_0 \leq \mathbb{E} M_{\{S \wedge n\}}$. This gets us our result:

$$\begin{aligned} b \mathbb{P}(m_n^* - b) &\leq -\mathbb{E} M_0 + \mathbb{E}[M_n \mathbb{1}_{\{m_n^*\}}] \\ &\leq \mathbb{E} M_n^* - \mathbb{E} M_0. \end{aligned}$$

□

Remark

If M is a martingale then $|M|^p$ is a sub-martingale for $p \geq 1$. If $M_n \in L^p \forall n$ we can apply Doob's inequality.

Corollary

If M is martingale in L^p for some $p \geq 1$ then for $b > 0$ we have that

$$b^p \mathbb{P}(\max_{k \leq n} |M_k| > b) \leq \mathbb{E}|M_n|^p.$$

There are also other bounds:

- $b \mathbb{P}(\max_{k \leq n} |M_k| > b) \leq 2\mathbb{E} M_n^+ - 3M_0$;
- **Doob's norm inequality**: if M is a martingale in L^p , $p \geq 1$ and q is the exponent conjugate to p ($\frac{1}{p} + \frac{1}{q} = 1$) then
$$\mathbb{E} \max_{k \leq n} |M_k|^p \leq q^p \mathbb{E}|M_n|^p.$$
- Consider L^2 -bounded martingales characterized by final coordinate X with $\text{Var } X = \sigma^2$ (that is I am fixing the variance of the last value I consider). We want to assess the variability of this process.

Theorem 2.2.7

Dubin & Schwartz 1998: it holds

1. $\mathbb{E} [\max_{0 \leq T \leq t} M_T] \leq \sigma$;
2. $\mathbb{E} [\max_{0 \leq T \leq t} |M_T|] \leq \sigma\sqrt{2}$.

Moreover, there exist suitable martingales for which this bound is attained and is strict.

2.2.8 Limit of sequences of random variables

Consider a sequence of random variables $\{X_n\}$ with independence between the random variable. Then we know that if $\lim X_n$ exists then the limit is a constant. Why?

Remark

If the $\lim_n X_n$ exists than it belongs to the tail σ -algebra τ . Since the random variables are independent then we can apply the 0-1 Kolmogorov law, so the probability that the random variable is infinity is either 0 or 1 but if this is the case then it must be a constant.

This is about the behavior in the limits of a sequence of random variables with independence between $\{X_n\}$. Then if the limit exists it is a constant, since $\lim_n X_n$ belongs to the tail σ -algebra of the sequence. Independence between random variables tells us that we can apply Kolmogorov's 0-1 law to say that $\mathbb{P}(\lim_n X_n = \infty)$ is either 0 or 1... but if we somehow know that the limit exists then it can't be ∞ so its probability of being ∞ is 0 and therefore it is a constant almost surely! This is cool because it tells us that the limit is either infinite or finite. If $\{X_n\}$ is a submartingale consider the limit

$$\lim_{n \rightarrow \infty} X_n = X_\infty$$

What the fuck? What is this object? Consider the case in which $T = \mathbb{N}$. We know that the expected value $\{\mathbb{E}X_n\}$ is an increasing sequence if the process is a sub-martingale. We have our first theorem about convergence:

Theorem 2.2.8

Let X be a sub-martingale. If (and note that is a sufficient condition)

$$\sup_n \mathbb{E}X_n^+ < \infty$$

Then

1. $\{X_n\}$ converges a.s.;
2. $\{X_n\}$ converges to an integrable random variable.

Before proving think about it.

1. The condition of this theorem requires that X^+ is in L^1 and is bounded. If X is a sub-martingale then also X^+ is a sub-martingale... But this means that $\mathbb{E}X_n^+$ is an increasing sequence. Our condition gives a bound to the increase of this sequence! But this means that I am trying to avoid the divergence of the expected value (which means: trying to control the rate of our martingale).
2. Now think about a negative sub-martingale: this would verify the condition so there is no problem about it.
3. Now think about a positive super-martingale: it is only a negative sub-martingale with another sign so it converges.
4. Consider now a positive or bounded martingale or a negative martingale. All of these cases converge.
5. $\sup \mathbb{E}X_n^+ < \infty \iff \sup \mathbb{E}|X_n| < \infty$. Let $\{X_n\}$ be a sub-martingale such that

$$\mathbb{E}X_n > \mathbb{E}X_0 \quad \forall n.$$

But now

$$\mathbb{E}|X_n| = \mathbb{E}X_n^+ - \mathbb{E}X_n^- = \mathbb{E}X_n^+ + \mathbb{E}_n^+ - \mathbb{E}X_n$$

So

$$\begin{aligned}\mathbb{E}X_n^+ &\leq \mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \\ &\leq 2 \underbrace{\mathbb{E}X_n^+}_{<\infty} - \mathbb{E}X_0 \leq \infty\end{aligned}$$

So $\mathbb{E}|X_n| \leq \infty$.

Proof

We prove the theorem by contradiction. Pick an outcome ω and suppose that $\{X_n(\omega)\}$ is a numerical sequence that has not a limit. But if it doesn't have a limit, then

$$\exists \inf \lim \neq \sup \lim \quad \inf \lim < \sup \lim.$$

So there exist at least 2 rationals $a < b$ such that

$$\inf \lim < a < b < \sup \lim.$$

The sequence $\{X_n(\omega)\}$ crosses (a, b) ∞ many times. Now take the union over rational a and b , $a < b$ of the sets

$$\{U(a, b) = \infty\}$$

with $U(a, b) = \lim_{n \rightarrow \infty} U_n(a, b)$. Our aim is now to show that $U(a, b) \leq \infty$ almost surely to get a contradiction.

Fix a, b . We know that $U_n(a, b)$ is increasing with n . Now consider

$$\begin{aligned}(b-a)\mathbb{E}U(a, b) &= (b-a)\mathbb{E}\lim U_n(a, b) \\ &= (b-a)\lim \mathbb{E}U_n(a, b) \\ &\stackrel{\text{monotone conv.}}{\leq} \sup \mathbb{E}(X_n - a)^+ \\ &\stackrel{\text{upcross inequalities}}{\leq} \sup \mathbb{E}X_n^+ + |a|.\end{aligned}$$

So this means that $\mathbb{E}U(a, b) < \infty$. But this is a contradiction, so it exists a limit $X_n = X_\infty$ a.s.. Now consider the second part of the theorem:

$$\begin{aligned}\mathbb{E}|X_\infty| &= \mathbb{E}\liminf |X_n| \\ &\stackrel{\text{Fatou's lemma}}{\leq} \liminf \mathbb{E}|X_n| \\ &\leq 2 \sup \mathbb{E}X_n^+ - \mathbb{E}X_0 \leq \infty\end{aligned}$$

so the limit is integrable. □

Here everything is related to integrability conditions but this is a sufficient condition. What if we want a sufficient and necessary condition? Well, consider $\sup_n \mathbb{E}X_n^+ < \infty$. We want to control the tail uniformly not only on the tail but on the whole process. We recall that we already introduced a strong condition for uniform integrability. We will need:

1. a collection \mathcal{K} of real random variables is said to be uniformly integrable if

$$k(b) = \sup_{X \in \mathcal{K}} \mathbb{E}|X| \mathbf{1}_{\{X>b\}} \xrightarrow[b \rightarrow \infty]{} 0.$$

2. If \mathcal{K} is dominated by an integrable random variable Z then it is uniformly integrable.

3. uniform integrability implies L^1 -boundedness but not the converse.

4. If \mathcal{K} is L^p -bounded for some $p > 1$ then it is uniformly integrable.

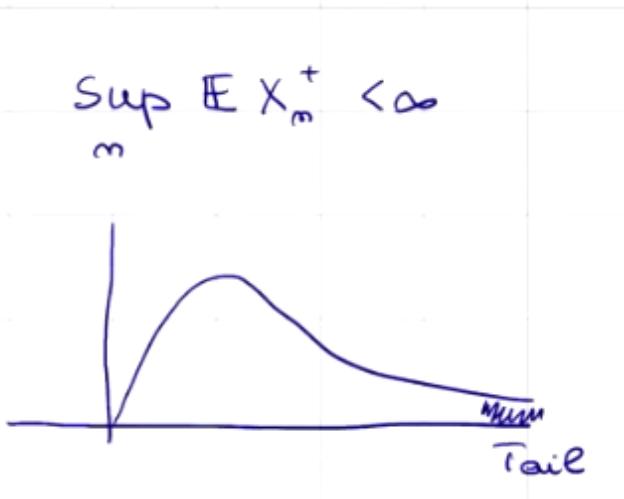


Figure 2.11: It's astonishing how much a bad teacher can have you hate a subject.

Lemma 2.2.1

Let Z be an integrable random variable. Then

$$\mathcal{K} = \{X : X = \mathbb{E}(Z|\mathcal{G})\}$$

for some sub- σ -algebra \mathcal{G} of \mathcal{H} is uniformly integrable.

Proposition 2.2.3

Let Z be an integrable random variable. Define

$$X_t = \mathbb{E}(Z|\mathcal{F}_t) \quad t \in \mathbb{T}.$$

This means that $\{X_t\}$ is a uniformly integrable \mathcal{F} -martingale.

Theorem 2.2.9

Let $\{X_n\}$ be a sequence of real-valued random variables. The following are equivalent:

1. it converges in L^1 ;
2. it converges in probability and it is uniformly integrable.

We can now prove the theorem about the convergence of sub-martingales.

Theorem 2.2.10

Let X be a sub-martingale. We have that X converges almost surely and in L^1 if and only if it is uniformly integrable. Moreover, if it is so, setting

$$X_\infty = \lim X_n$$

extends X to a sub-martingale

$$\bar{X} = (X_n)_{n \in \overline{\mathbb{N}}}.$$

We only prove the first part of the theorem.

Proof

Necessity. If X converges in L^1 by the theorem above it is uniformly integrable.

Sufficiency. If X is uniformly integral then it is L^1 -bounded for the property above. So our previous

theorem holds and the martingale converges almost surely with X_∞ integrable. Furthermore, for the property above, it also converges in L^1 . \square

Is there a way to get a uniformly integrable sub-martingale?

Theorem 2.2.11

A process $M = (M_n)_{n \in \mathbb{N}}$ is a uniformly integrable martingale if and only if for some integrable random variable Z

$$M_n = \mathbb{E}[Z|\mathcal{F}_n] \quad n \in \mathbb{N}. \quad (\bullet)$$

If so it converges almost surely and in L^1 to the integrable random variable

$$M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty].$$

Corollary

For every integrable random variable Z we have

$$\mathbb{E}(Z|\mathcal{F}_n) \xrightarrow{\text{a.s.}} \mathbb{E}(Z|\mathcal{F}_\infty).$$

Theorem 2.2.12

Let Z be an integrable random variable and let

$$M_n = \mathbb{E}(Z|\mathcal{F}_n)_{n \in \overline{\mathbb{N}}}.$$

For every stopping time T define

$$M_T = \mathbb{E}[Z|\mathcal{F}_T]$$

and for arbitrary stopping times S and T we get

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_{S \wedge T}.$$

This lets us rethink Doob's theorem.

Theorem 2.2.13

If S and T are arbitrary stopping times such that $S \leq T$ then

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S$$

for an uniformly integrable martingale.

The dominated convergence theorem requires adaptness.

Theorem 2.2.14

Hunt's dominated convergence theorem.

Let $\{X_n\}$ be dominated by an integrable random variable and suppose that exists

$$X_\infty = \lim X_n$$

So $(\mathbb{E}_n X_n)_n$ converges to $\mathbb{E}X_\infty$ almost surely and in L^1 .

To prove we should use Borel-cantelli somehow.

Chapter 3

Miscellanea: an integration (eheh)

3.1 Gaussian random variables

Definition 3.1.1

A **standard Gaussian random variable** (of one dimension) is a random variable Z with density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}.$$

Definition 3.1.2

A **Gaussian random variable** with parameters μ and σ^2 is a random variable X with density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Proposition 3.1.1

We have $\mathbb{E}X = \mu$, $\text{Var } X = \sigma^2$ and in particular $\mathbb{E}Z = 0$, $\text{Var } Z = 1$.

Proposition 3.1.2

Let $Z \sim \mathcal{N}(0, 1)$ and $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$. Then

$$X := \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2).$$

Proposition 3.1.3

The characteristic function of X is

$$\mathbb{E}[e^{i\lambda X}] = e^{i\lambda\mu} \cdot e^{-\frac{1}{2}\sigma^2\lambda^2}$$

if and only if we have $X \sim \mathcal{N}(\mu, \sigma^2)$.

Let's generalize this to n -dimensional random variables.

Definition 3.1.3

Let $X = (X_1, \dots, X_n)$ be a n -dimensional vector of random variables. We say that X is Gaussian if for every vector $r \in \mathbb{R}^n$ we have that $r \cdot X = \sum_{i=1}^n r_i X_i$ is Gaussian.

Proposition 3.1.4

Let $Z = (Z_1, \dots, Z_k)$ be k -dimensional vector of independent random variables $Z_i \sim \mathcal{N}(0, 1)$. Let $\mu \in \mathbb{R}^n$ and A be a $(n \times k)$ -dimensional matrix $A = (a_{ij})_{i,j}$. We set

$$\begin{matrix} i=1, \dots, n \\ j=1, \dots, k \end{matrix}$$

$$\underset{(n \times 1)}{X} := \underset{(n \times 1)}{\mu} + \underset{(n \times k)(k \times 1)}{A} \underset{(k \times 1)}{Z}.$$

In particular,

$$X_i := \mu_i + \sum_{j=1}^k a_{ij} Z_j.$$

In this case, X is Gaussian.

The n -dimensional Gaussian random variables have the following properties:

- the characteristic function of X is given by

$$\mathbb{E}[e^{ir \cdot X}] = \mathbb{E}\left[e^{i\sum_{i=1}^n r_i X_i}\right] = e^{ia(r) - \frac{1}{2}b(r)}, \quad r \in \mathbb{R}^n$$

where

$$\begin{aligned} a(r) &:= \mathbb{E}[r \cdot X] \\ b(r) &:= \text{Var}(r \cdot X). \end{aligned}$$

Yeah, why not. Why not making my life even more miserable.

- Setting $\mu_i := \mathbb{E}X_i$ and $\Gamma_{ij} = \text{Cov}(X_i, X_j)$, then

$$a(r) = r \cdot \mu \quad \text{and} \quad b(r) = \underset{(n \times 1)}{r} \cdot \underset{(n \times n)(n \times 1)}{\Gamma} \underset{(n \times 1)}{r}^\top.$$

What the fuck does this mean? The Gaussian vector X is determined by its mean μ and covariance matrix Γ , because they determine the characteristic function of X . We write

$$X \sim \mathcal{N}(\mu, \Gamma)$$

where X is a n -dimensional vector (and also μ).

Proposition 3.1.5

Any Gaussian random variable X can be written as $X = \mu + AZ$ with $\mu \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times k}$ and $Z = (Z_1, \dots, Z_k)$ with $Z_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$

Let's now talk about independence. Since any Gaussian vector X can be written as $X = \mu + AZ$ where the $Z_i \perp\!\!\!\perp Z_j, i \neq j, i, j = 1, \dots, k$, what can we say about the independence of $X_i, X_j, i \neq j, i, j = 1, \dots, n$? To see this, calculate $\text{Cov}(X_i, X_j)$:

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] \\ &= \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathbb{E}\left[(\mu_i + \sum_{l=1}^k a_{il}z_l - \mu_i)(\mu_j + \sum_{m=1}^k a_{jm}z_m - \mu_j)\right] \\ &= \mathbb{E}\left[\left(\sum_{l=1}^k a_{il}z_l\right)\left(\sum_{m=1}^k a_{jm}z_m\right)\right] \\ &= \sum_{l=1}^k \sum_{m=1}^k a_{il}a_{jm} \mathbb{E}[Z_l Z_m] \\ &= \sum_{l=1}^k a_{il}a_{jl} \underbrace{\mathbb{E}[Z_l^2]}_{=1 \text{ since } Z_l \sim \mathcal{N}(0, 1)} = (\mathbf{A}\mathbf{A}^\top)_{ij}. \end{aligned}$$

Since it is necessary (but not sufficient) that $\text{Cov}(X_i, X_j) = 0$ to have $X_i \perp\!\!\!\perp X_j$ then X_i is not independent from X_j in general.

Note the notation!

We proved that $\text{Cov}(X_i, X_j) = (\mathbf{A}\mathbf{A}^\top)_{ij}$, hence with the notation above we get

$$\underset{n \times n}{\boldsymbol{\Gamma}} = \underset{n \times k}{\mathbf{A}} \underset{k \times n}{\mathbf{A}^\top}.$$

Proposition 3.1.6

If $\boldsymbol{\Gamma}$ is invertible (i.e. $\text{rank}(\boldsymbol{\Gamma}) = n$) then $X = \mu + \mathbf{A}\mathbf{Z}$ has density given by

$$f_X(x) = \frac{1}{(2\pi)^{\frac{k}{2}}} \cdot \frac{1}{\sqrt{\det(\boldsymbol{\Gamma})}} \cdot e^{-\frac{1}{2} \underset{(n \times 1)}{(x-\mu)} \cdot \underset{(n \times n)}{\boldsymbol{\Gamma}^{-1}} \cdot \underset{(n \times 1)}{(x-\mu)}}$$

Note that if $\boldsymbol{\Gamma}$ is singular (i.e. $\text{rank}(\boldsymbol{\Gamma}) = r < n$) then the distribution of X is not absolutely continuous with respect to the Lebesgue Measure (and so X has no density). Indeed, the distribution lives on a hyperplane of \mathbb{R}^n and at least one component of X is a linear combination of the others¹.

¹What the fuck? This is absolutely heinous.

Chapter 4

Exercises and questions

4.1 Exercises with Prof. Issoglio (2024/2025)

Here I gathered the exercises seen with Prof. Issoglio. Each class begins with a small recap on the topics that the exercises will be based on.

4.1.1 Exercise class 1

Revise with Kotatsu!

→ Measurable and measure spaces:

$$(E, \mathcal{E}, \nu)$$

set σ -algebra measure

An example is given by discrete spaces (E is finite and countable) or the real spaces ($E = \mathbb{R}$), ($E = \mathbb{R}^n$).

→ Measurable functions form (E, \mathcal{E}) to (F, \mathcal{F}) . It is a function $f : E \mapsto F$ such that $\forall B \in \mathcal{F}$ we have $f^{-1}(B) \in \mathcal{E}$. This is the least possible “regularity” we can ask for to still be able to do some analysis (it is much less restrictive than continuity).

Property: if f_n are measurable, also $\liminf_n f_n$, $\limsup_n f_n$, $f_1 + f_2$, $f_1 \cdot f_2$, λf_1 are measurable.

→ Probability spaces: special case of a measure space where $\nu(E) = 1$. Given a real random variable $X : \Omega \mapsto \mathbb{R}$ we can consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}_X)$ where \mathcal{L}_X is the law of X given by

$$\mathcal{L}_X := \mathbb{P}(X^{-1}(A)), \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

We have the mapping

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}_X).$$

→ σ -algebras generated by random variables: given $X : \Omega \mapsto \mathbb{R}$ we define

$$\sigma(X) := \{X^{-1}(A) \text{ for some } A \in \mathcal{B}(\mathbb{R})\}.$$

This is the *smallest* σ -algebra such that X is a random variable on $(\Omega, \sigma(X))$. By properties of p-systems and d-systems it is enough to consider a set of elements of \mathbb{R} that generates $\mathcal{B}(\mathbb{R})$ (for example: all open sets) when checking measurability related properties. This means that

$$\begin{aligned} \sigma(X) &= \{X^{-1}(A) \text{ for some } A \in \mathcal{B}(\mathbb{R})\} \\ &= \{X^{-1}(A) \text{ for some } A \text{ open}\} \end{aligned}$$

Exercise 1

Consider the experiment where we throw a coin infinitely many times, so that

$$\Omega \subset \{H, T\}^{\mathbb{N}} = (\Omega \times \{0, 1\})^{\mathbb{N}}$$

In particular, the event $\omega \in C$ is of the form $\omega = (\omega_1, \omega_2, \dots)$ with $\omega_i \in \{0, 1\}$. Let

$$C = \{ \omega \in \Omega : \omega_k \in \omega, \omega \in \{0, 1\}^{\mathbb{N}} \}$$

and

$$\mathcal{F} := \sigma(C).$$

We want to check whether \mathcal{F} is a good modeling choice for our experiment. For example, if we set

$$A = \left\{ \omega : \frac{\#\{k \leq n : \omega_k = 1\}}{n} \xrightarrow[n \rightarrow \infty]{} \frac{1}{2} \right\}$$

do we have $A \in \mathcal{F}$?

Let us introduce a random variable (which is a map) $X_n : \Omega \mapsto \mathbb{R}$ given by

$$X_n = \begin{cases} 1 & \text{if } \omega_n = 1 \\ 0 & \text{if } \omega_n = 0. \end{cases}$$

Now, X_n is \mathcal{F} -measurable (i.e. it is a random variable on (Ω, \mathcal{F})) because for $\forall B \in \mathcal{B}(\mathbb{R})$ we have

$$X_n^{-1}(B) = \{ \omega \in \Omega : X_n(\omega) \in B \}$$

and this set actually belongs to \mathcal{F} . Why? Because it basically is

$$\{ \omega \in \Omega : X_n(\omega) = B \} = \begin{cases} \emptyset & \text{if } 0, 1 \notin B \\ \Omega & \text{if } 0, 1 \in B \\ \{ \underbrace{\omega \in \Omega : X_n(\omega) = 1}_{\text{all the } \omega \text{ with } \omega_n = 1 \implies \in C} \} & \text{if } 0 \notin B, 1 \in B \\ \{ \underbrace{\omega \in \Omega : X_n(\omega) = 0}_{\text{all the } \omega \text{ with } \omega_n = 0 \implies \in C} \} & \text{if } 0 \in B, 1 \notin B \end{cases}$$

So the process is measurable with respect to \mathcal{F} : the σ -algebra generated by C has all the ω necessary to fully explain the process. Remember that $B \subset \Omega$! Now let $S_n = \sum_{i=1}^n X_i$. Then our set A becomes

$$A = \left\{ \omega : \frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} \frac{1}{2} \right\}.$$

We know that S_n is a sum of \mathcal{F} -measurable functions and therefore it is \mathcal{F} -measurable. We can write A as

$$\begin{aligned} A &= \left\{ \omega \in \Omega : \frac{S_n}{n} \rightarrow \frac{1}{2} \right\} \\ &= \left\{ \omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_n}{n} = \limsup_{n \rightarrow \infty} \frac{S_n}{n} = \frac{1}{2} \right\} \\ &= \left\{ \underbrace{\omega \in \Omega : \liminf_{n \rightarrow \infty} \frac{S_n}{n} = \frac{1}{2}}_{\in \mathcal{F}} \right\} \cap \left\{ \underbrace{\omega \in \Omega : \limsup_{n \rightarrow \infty} \frac{S_n}{n} = \frac{1}{2}}_{\in \mathcal{F}} \right\} \end{aligned}$$

since S_n is \mathcal{F} -measurable, so are its \limsup and \liminf . By doing so we went from the limit, which is not a nice thing to handle¹ when it comes to measurability, to \liminf and \limsup that are more manageable objects.

¹Unlike a gun pointed to my head by myself.

Exercise 2

Let $\Omega \subset \{0, 1\}^n$. For a given $n \in \mathbb{N}$ let

$$\mathcal{A}_n = \{\{\omega = (\omega_1, \omega_2, \dots) \in \Omega : (\omega_1, \dots, \omega_n) \in B\} \text{ for some } B \subset \{0, 1\}^n\}.$$

- ① Show that \mathcal{A}_n is a σ -algebra.
- ② Show that $\bigcup_{n \in \mathbb{N}} \mathcal{A}_n$ is not a σ -algebra.

- ① If we show that \mathcal{A}_n is finite, then it is enough to prove that it is an algebra, since the closure under *union, intersection and complementation* will automatically follow from this.

We easily see that \mathcal{A}_n is finite because n is fixed so $\{0, 1\}^n$ is finite and since the varying set in this case is $B \subset \{0, 1\}^n$ then also \mathcal{A}_n must be finite. For example we would get

$$\mathcal{A}_1 = \{\{\omega \in \Omega : \omega_1 \in \{0\}\}, \{\omega \in \Omega : \omega_1 \in \{1\}\}, \{\omega \in \Omega : \omega_1 \in \{0, 1\}\}, \{\omega \in \Omega : \omega_1 \in \emptyset\}\}.$$

We can also check that this is an algebra:

- (a) it is obvious that $\emptyset \in \mathcal{A}_n$;
- (b) let $A \in \mathcal{A}_n$. This means that $\exists B \subset \{0, 1\}^n$ such that

$$A = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}$$

and so

$$\begin{aligned} A^c &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \notin B\} \\ &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B^c\} \in \mathcal{A}_n \end{aligned}$$

since $B^c \subset \{0, 1\}^n$. So \mathcal{A}_n is closed under intersection.

- (c) Let $A_1, A_2 \in \mathcal{A}_n$. Then

$$\begin{aligned} A_1 \cup A_2 &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B_1\} \cup \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B_2\} \\ &= \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B_1 \cup B_2\} \in \mathcal{A}_n \end{aligned}$$

because of course $B_1 \cup B_2 \subset \{0, 1\}^n$. So \mathcal{A}_n is closed under union.

So we conclude that \mathcal{A}_n is an algebra and since it is finite it is also a σ -algebra.

- ② To prove that $\bigcup_n \mathcal{A}_n$ is *not* a σ -algebra it is sufficient to find an element $A \in \sigma(\bigcup_n \mathcal{A}_n)$ such that $A \notin \bigcup_n \mathcal{A}_n$. This is because a σ -algebra must be closed under countable unions (beside countable intersections and complementation) and if we find an element in the σ -algebra generated by the union of the sets \mathcal{A}_n that is not in the union we would get that the sets are not closed under union²!

We notice that for each n , \mathcal{A}_n does not include the set formed by the single element $\mathbf{0} = (0, 0, \dots)$ because any element $A \in \mathcal{A}_n$ contains infinitely many ω s. Thus $\{\mathbf{0}\} \notin \bigcup_n \mathcal{A}_n$. But we can write $\{\mathbf{0}\} = \bigcap_n A_n$ where

$$A_n = \{\omega \in \Omega : (\omega_1, \dots, \omega_n) = (0, \dots, 0)\}.$$

Clearly $A_n \in \mathcal{A}_n \subset \bigcup_m \mathcal{A}_m$, so $\{\mathbf{0}\} \in \sigma(\bigcup_n \mathcal{A}_n)$ since it is expressed as the countable intersection of $\bigcup_n \mathcal{A}_n$.

Exercise 3

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let \mathfrak{F} be defined by

$$\mathfrak{F} = \{A \in \mathcal{A} : \mathbb{P}(A) \in (0, 1)\}.$$

- ① Show that \mathfrak{F} is a σ -algebra.
- ② Find examples such that
 - (a) $\mathfrak{F} = \{\emptyset, \Omega\}$;
 - (b) \mathfrak{F} has infinitely many elements.

²The only one doing an union here is me.

- ①
- We show that $\emptyset \in \mathcal{F}$. Notice that $\mathbb{P}(\Omega) = 1$ and so $\Omega \in \mathcal{F}$. notice that \mathcal{F} is closed under complementation because if $F \in \mathcal{F}$ then either $\mathbb{P}(F) = 0$ or $\mathbb{P}(F) = 1$. Hence $\mathbb{P}(F^c) = 1 - \mathbb{P}(F) \in \{0, 1\}$. Thus $\Omega^c \in \mathcal{F} \implies \emptyset \in \mathcal{F}$.
 - We show that if $F \in \mathcal{F}$ then $F^c \in \mathcal{F}$... but we already did that.
 - We show that if $F_n \in \mathcal{F}$ then $\bigcup_n F_n \in \mathcal{F}$. We consider two cases:
 - if $\mathbb{P}(F_n) = 0 \forall n$ then $\mathbb{P}(\bigcup_n F_n) \leq \sum_n \mathbb{P}(F_n) = 0$ and therefore

$$\bigcup_n F_n \in \mathcal{F};$$

• if there exists at least one \bar{n} such that $\mathbb{P}(F_{\bar{n}}) = 1$ then using that $F_{\bar{n}} \subseteq \bigcup_n F_n$ we get that

$$1 = \mathbb{P}(F_{\bar{n}}) \leq \mathbb{P}\left(\bigcup_n F_n\right) \implies \bigcup_n F_n \in \mathcal{F}.$$

- ②
- Take any discrete finite Ω such that each element $\omega \in \Omega$ has positive probability and strictly less than 1 (for example, a uniform probability space). In this case there are no elements $A \in \mathcal{A}$ such that $\mathbb{P}(A) \in \{0, 1\}$ apart from \emptyset and Ω for which probabilities are respectively 0 and 1 so that $\mathcal{F} = \{\emptyset, \Omega\}$.
 - Here we need an infinite set Ω . Let us choose $\Omega = [0, 1]$ with $\mathcal{A} = \mathcal{B}([0, 1])$ and $\mathbb{P} = \lambda$. In this case

$$\mathbb{P}(\{x\}) = 0 \quad \forall x \in [0, 1]$$

and therefore \mathcal{F} is finite (even uncountable).

Exercise 4

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space such that $\{\omega\} \in \mathcal{A} \forall \omega \in \Omega$. Show that

$$A := \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > 0\}$$

is countable.

Let us set $A_n = \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{n}\}$ for $n \geq 1$. For example:

$$\begin{aligned} A_1 &= \{\omega \in \Omega : \mathbb{P}(\{\omega\}) > 1\} = \emptyset \\ A_2 &= \left\{ \omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{2} \right\} \\ A_3 &= \left\{ \omega \in \Omega : \mathbb{P}(\{\omega\}) > \frac{1}{3} \right\} \\ &\vdots \end{aligned}$$

Since $\mathbb{P}(\Omega) = 1$ we have $\#A_1 = 1, \#A_2 < 2, \#A_3 < 3$ so in general

$$\#A_n \leq n.$$

Notice moreover that $A_1 \subset A_2 \subset A_3 \dots$ and so

$$A = \bigcup_{n \in \mathbb{N}} A_n.$$

Thus A is the countable union of finite sets, hence it is countable.

Exercise 5

Let $(\Omega, \mathcal{F}, \mu)$ be a given measure space. Let \mathcal{N} be the set of null sets ($\mu(A) = 0$, negligible sets) of Ω and let $\bar{\mathcal{F}} = \sigma(\mathcal{F} \cup \mathcal{N})$.

- a) Prove that if $B \in \bar{\mathcal{F}}$ we have $B = A \cup N$ for $A \in \mathcal{F}$ and $N \in \mathcal{N}$.

- b) Letting

$$\bar{\mu}(A \cup N) := \mu(A) \quad \forall A \in \mathcal{F}, N \in \mathcal{N} \quad (\heartsuit)$$

Prove that:

$$(\text{a}) \quad \bar{\mu}(A) = \mu(A), \quad \forall A \in \mathcal{F};$$

(b) $\bar{\mu}$ is well-defined and $\bar{\mu}$ defines a measure as required in $\bar{\mathcal{F}}$;

(c) $(\Omega, \bar{\mathcal{F}}, \bar{\mu})$ is complete.

- a) Setting

$$F = \{A \cup N : A \in \mathcal{F}, N \in \mathcal{N}\}$$

we see that

$$\mathcal{F} \cup \mathcal{N} = F \subset \bar{\mathcal{F}} := \sigma(\mathcal{F} \cup \mathcal{N}). \quad (\otimes)$$

If we prove that F is a σ -algebra then we have $\sigma(F) = F$ hence by \otimes we have $F = \bar{\mathcal{F}}$. We know that

- $\emptyset \in F$ because $\emptyset = \emptyset \cup \emptyset$ and $\emptyset \in \mathcal{F}, \emptyset \in \mathcal{N}$ (\checkmark);
- if $B \in F$ then $B^c \in F$. Indeed $B = A \cup N$ for some $A \in \mathcal{F}, N \in \mathcal{N}$ so

$$B^c = (A \cup N)^c = A^c \cap N^c$$

by DeMorgan's law. Now we observe that since for $N \in \mathcal{N}$ there exists $M \supseteq N$ such that $M \in \mathcal{F}, \mu(M) = 0$ we can write

$$N^c = M^c \cup (M \setminus N)$$

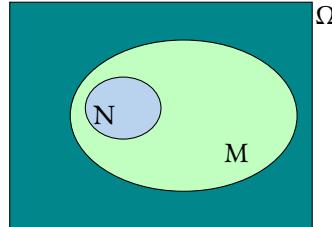


Figure 4.1: Of course, it is never too late for some sickz tickz pickz

Then from DeMorgan Law we get that

$$\begin{aligned} B^c &= A^c \cap (M^c \cup (M \setminus N)) \\ &= \underbrace{A^c}_{\in \mathcal{F}} \cap \underbrace{M^c}_{\in \mathcal{F}} \cup \underbrace{A^c \cap (M \setminus N)}_{\in \mathcal{F}} \subseteq M \setminus N \subset M \text{ and } \mu(M) = 0 \\ &\implies A^c \cap (M \setminus N) \in \mathcal{N} \\ &\implies B^c \in F. \end{aligned}$$

- Let $\{B_n\}_n$ be a countable family, $B_n \in F$. We prove that

$$\bigcup_n B_n \in F.$$

By hypothesis $\exists A_n \in \mathcal{F}, N_n \in \mathcal{N}$ such that $B_n = A_n \cup N_n$ and therefore

$$\begin{aligned} \bigcup_n B_n &= \bigcup_n (A_n \cup N_n) \\ &= \underbrace{\bigcup_n A_n}_{\in \mathcal{F}} \cup \underbrace{\bigcup_n N_n}_{\subseteq \bigcup_n M_n \text{ with } M_n \in \mathcal{F}, \mu(M_n) = 0}. \end{aligned}$$

Moreover,

$$\mu\left(\bigcup_n M_n\right) \leq \sum_n \mu(M_n) = 0 \implies \bigcup_n N_n \in \mathcal{N} \implies \bigcup_n B_n \in \mathcal{F}.$$

(b) Recall that $\forall A, N$ we let $\bar{\mu}(A \cup N) = \mu(A)$.

b1) Let $A \in \mathcal{F}$. Then $A = A \cap \emptyset$ with $\emptyset \in \mathcal{N}$. Then

$$\bar{\mu}(A) = \bar{\mu}(A \cap \emptyset) = \mu(A).$$

b2) We have to show that

- i. the DeMorgan equation is well-posed, that is

$$A \cup N = A' \cup N' \implies \bar{\mu}(A \cup N) = \bar{\mu}(A' \cup N');$$

ii. $\bar{\mu}$ is a measure on $(\Omega, \bar{\mathcal{F}})$;

iii. $\bar{\mu}$ is the only measure on $(\Omega, \bar{\mathcal{F}})$ such that the DeMorgan equation holds.

We need to prove each of these points.

- i. We notice that

$$A \subset A \cup N = A' \cup N' \subset A' \cup M'$$

where $M' \supset N'$, $M' \in \mathcal{F}$ and $\mu(M') = 0$. Moreover, $A \in \mathcal{F}$ and $A' \cup M' \in \mathcal{F}$. This implies

$$\mu(A) \leq \mu(A' \cup M') \leq \mu(A') + \mu(M') = \mu(A').$$

Analogously we get

$$\mu(A') \leq \mu(A) \implies \mu(A) = \mu(A').$$

Thus

$$\bar{\mu}(A \cup N) := \mu(A) = \mu(A') =: \bar{\mu}(A' \cup N').$$

ii. To show that $\bar{\mu}$ is a measure on $(\Omega, \bar{\mathcal{F}})$ we check:

- $\bar{\mu}(\emptyset) = \bar{\mu}(\emptyset \cup \emptyset) = \mu(\emptyset) = 0$;
- let $\{B_n\}_n \subset \bar{\mathcal{F}}$ with B_n pairwise disjoint. Then, since $B_n = \underbrace{A_n}_{\in \mathcal{F}} \cup \underbrace{N_n}_{\in \mathcal{N}}$

$$\begin{aligned} \bar{\mu}\left(\bigcup_n B_n\right) &= \bar{\mu}\left(\bigcup_n (A_n \cup N_n)\right) \\ &= \bar{\mu}\left(\left(\bigcup_n \underbrace{A_n}_{\in \mathcal{F}}\right) \cup \left(\bigcup_n \underbrace{N_n}_{\in \mathcal{N}}\right)\right) \\ &= \bar{\mu}\left(\bigcup_n A_n\right) \quad \text{by DeMorgan eqn.} \\ &= \sum_n \mu(A_n) \quad \text{bc they are pair. disj.} \\ &= \sum_n \bar{\mu}(A_n \cup N_n) \\ &= \sum_n \bar{\mu}(B_n). \end{aligned}$$

iii. To show that $\bar{\mu}$ is unique we suppose that $\exists \nu$, a measure on $(\Omega, \bar{\mathcal{F}})$ such that

$$\nu(A \cup N) = \mu(A), \quad \forall A \in \mathcal{F}. \tag{*}$$

Then from DeMorgan and * we get

$$\nu(A \cup N) = \bar{\mu}(A \cup N), \quad \forall A \in \mathcal{F}, \forall N \in \mathcal{N}.$$

Since μ and ν coincide on all $F = A \cup N \in \mathcal{F} = \bar{\mathcal{F}}$ they are the same measure.

b3) To show that the space is complete we must show that every negligible set of \mathcal{F} is actually measurable ($\in \bar{\mathcal{F}}$) and has measure 0. Let N be such that $N \subseteq M$, $M \in \bar{\mathcal{F}}$ with $\mu(M) = 0$. Since $N = \underbrace{N}_{\in \mathcal{F}} \cup \underbrace{\emptyset}_{\in \mathcal{F}}$ it means $N \in \bar{\mathcal{F}}$. Moreover

$$\bar{\mu}(N) = \bar{\mu}(N \cup \emptyset) = \mu(\emptyset) = 0.$$

Wow, this was useless.

4.1.2 Exercise class 2

Revise with Kotatsu!

→ Independence of random variables:

Definition 4.1.1

Given two random variables X_1 and X_2 we say that they are **independent** if $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

Two σ -algebras are independent if $\forall A_1 \in \sigma(X_1)$ and $A_2 \in \sigma(X_2)$ we have

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2).$$

An equivalent definition of independence for real-valued random variables is that

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1) \cdot \mathbb{P}(X_2 \leq x_2) \quad \forall x_1, x_2 \in \mathbb{R}.$$

To see the equivalence note that $\{(-\infty, x], x \in \mathbb{R}\}$ is a p-system for the Borel σ -algebra. Setting $C_i = \{\omega : X_i \in (-\infty, x], x \in \mathbb{R}\}$ we have $\sigma(C_i) = \sigma(X_i)$ for $i = 1, 2$. The equivalence follows.

→ Uniform integrability for a family of random variables:

Definition 4.1.2

K is uniformly integrable if

$$\lim_{b \rightarrow \infty} \sup_{X \in K} \mathbb{E}[|X| \mathbb{1}_{\{|X| > b\}}] = 0.$$

Theorem 4.1.1

K is uniformly integrable if and only if it is *L^1 – bounded* and $\forall \varepsilon > 0 \exists \delta > 0$ such that for \forall event H we have

$$\mathbb{P}(H) \leq \delta \implies \sup_{X \in K} \mathbb{E}[|X| \mathbb{1}_H] \leq \varepsilon.$$

This is known as the “ $\varepsilon - \delta$ characterization” and it means that on every small set the integrals of the X are uniformly small. The “tails” of these random variables behave uniformly well, especially on small-probability sets.

→ Transition kernels:

Definition 4.1.3

Let (E, \mathcal{E}) and (F, \mathcal{F}) be measurable spaces. We say that

$$K : E \times \mathcal{F} \mapsto \overline{\mathbb{R}}_+$$

is a transition kernel from (E, \mathcal{E}) to (F, \mathcal{F}) if

- a) the map $X \mapsto K(x, B)$ (from E to $\overline{\mathbb{R}}_+$) is \mathcal{E} -measurable for $\forall B \in \mathcal{F}$;
- b) the map $B \mapsto K(x, B)$ (from \mathcal{F} to $\overline{\mathbb{R}}_+$) is a measure on (F, \mathcal{F}) for every $X \in E$.

Exercise 1

Choose 2 numbers uniformly between 0 and 1, denoted by X_1 and X_2 . Consider both the case with replacement and without replacement. Are X_1 and X_2 independent?

First we have to specify the probability space and view X_1 and X_2 as random variables, so the concept of independence makes sense.

Let

$$\Omega = [0, 1] \times [0, 1] \quad \mathcal{F} = \mathcal{B}([0, 1] \times [0, 1]) \quad \mathbb{P} = \lambda$$

with $(\omega \in \Omega \rightarrow \omega = (\omega_1, \omega_2))$ and

$$\begin{aligned} X_1 : \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X_1(\omega) := \omega_1 \\ X_2 : \omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X_2(\omega) := \omega. \end{aligned}$$

X_i are random variables because they are measurable. Indeed, for $A \in \mathcal{B}(\mathbb{R})$ (e.g. for $i = 1$) we have that

$$\begin{aligned} X_1^{-1}(A) &= \{\omega \in \Omega : X_1(\omega) \in A\} \\ &= \{\omega \in \Omega : \omega_1 \in A\} \\ &= \{\omega \in \Omega : \omega_1 \in A, \omega_1 \in [0, 1]\} \in \mathcal{B}([0, 1] \times [0, 1]) \end{aligned}$$

since both A and $[0, 1]$ belong in $\mathcal{B}([0, 1])$. The same can be said for $i = 1$. To check the dependence (or independence) we compute the joint distribution of (X_1, X_2) :

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \begin{cases} 1 & \text{if } x_1, x_2 > 1 \\ x_1 x_2 & \text{if } 0 \leq x_1, x_2 \leq 1 \\ x_1 & \text{if } 0 \leq x_1 \leq 1, x_2 > 1 \\ x_2 & \text{if } 0 \leq x_2 \leq 1, x_1 > 1 \\ 0 & \text{else.} \end{cases}$$

The situation can be summarized in the following drawings:

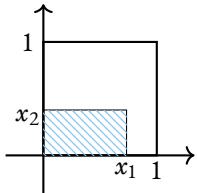


Figure 4.2: Case with $0 \leq x_1, x_2 \leq 1$

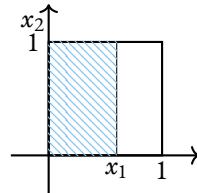


Figure 4.3: Case with $0 \leq x_1 \leq 1, x_2 > 1$

Calculate the marginal distribution as well:

$$\mathbb{P}(X_i \leq x_1) = \begin{cases} 1 & \text{if } x_i > 1 \\ x_i & \text{if } - \leq x_i \leq 1 \\ 0 & \text{else} \end{cases} \quad \text{for } i = 1, 2.$$

So we can basically see that

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1) \cdot \mathbb{P}(X_2 \leq x_2)$$

and therefore we have that X_1 and X_2 are independent.

In this case we must change the probability space to account for this difference. Let

$$\Omega = \{\omega \in [0, 1] \times [0, 1] \text{ s.t. } \omega_1 \neq \omega_2\} \quad \mathcal{F} = \mathcal{B}([0, 1] \times [0, 1], x_1, x_2) \quad \mathbb{P} = \lambda$$

The random variables are the same as before ($X_i(\omega)$, $i = 1, 2$) and so $\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$ is as above because $\mathbb{P}(x_1 = x_2) = 0$ since we are using fucking Lebesgue measure as probability measure. So also in this case X_1 and X_2 are independent. What the fuck?? And why can't I seem to be able to have centered captions anymore?

Exercise 2

Let c be a cumulative distribution function, that is

$$c: [0, 1] \rightarrow \mathbb{R} \text{ increasing, right-continuous}$$

Let q be the corresponding quantile function, that is

$$\begin{aligned} q: (0, 1) &\rightarrow \overline{\mathbb{R}} \\ u \mapsto q(u) &= \inf\{x \in \mathbb{R} : c(x) > u\}. \end{aligned}$$

Let $\mu = \lambda_{(0,1)} \circ q^{-1}$.

- (a) Show that μ is a probability measure on $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$.
- (b) Show that $\mu((-\infty, x]) = c(x) \quad \forall x \in \overline{\mathbb{R}}$.
- (c) Show that for any cumulative distribution function on $\overline{\mathbb{R}}$ there exists a unique probability measure μ on $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ and vice versa.

Set

$$\lambda = \lambda_{(0,1)}.$$

- (a) First we show that μ is a measure since it is the composition of a Borel function (q^{-1} , since q is Borel) with a measure. To show it is a probability measure we calculate

$$\begin{aligned} \mu(\overline{\mathbb{R}}) &= \lambda(q^{-1}(\overline{\mathbb{R}})) \\ &= \lambda(\{x \in [0, 1] : q(x) \in \overline{\mathbb{R}}\}) \\ &= \lambda(0, 1) = 1. \end{aligned}$$

- (b) Consider the following:

$$\begin{aligned} \mu((\infty, x]) &= \lambda(q^{-1}((-\infty, x])) \\ &= \lambda(\{u \in [0, 1] : q(u) \in (-\infty, x]\}) \\ &= 1 - \underbrace{\lambda(\{u \in [0, 1] : q(u) \notin (-\infty, x]\})}_{\lambda \text{ is a probability measure!}} \\ &= 1 - \lambda(\{u \in [0, 1] : q(u) > x\}) \\ &= 1 - \underbrace{\lambda(\{u \in [0, 1] : c(x) \leq u\})}_{\text{notice that } u \in (0, 1) : q(u) > x \iff u \in [0, 1] : c(x) \leq u} \\ &= 1 - \lambda([c(x), 1)) \\ &= 1 - (1 - c(x)) \\ &= c(x). \end{aligned}$$

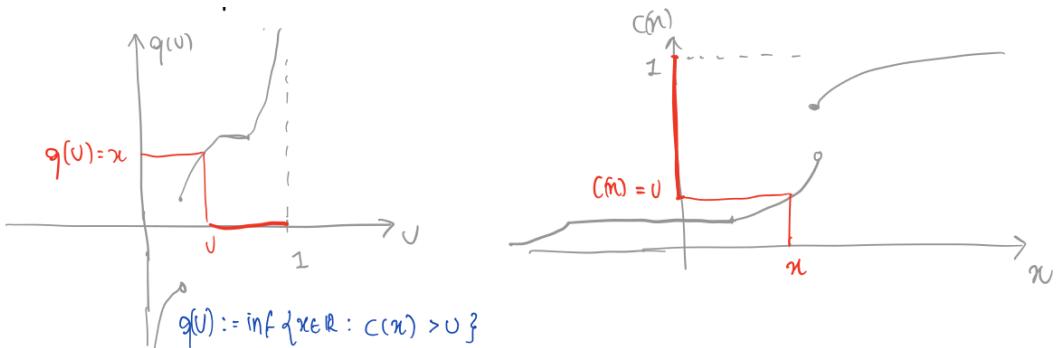


Figure 4.4: Ok but tikz are hard and long... ; that is WHY i use them.

- (c) We construct μ as in the previous two points. That μ is unique thanks to the theorem on uniqueness of measures (since μ is defined on the p-system $(-\infty, x]$). Vice versa, given the probability measure μ on

$(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ we define

$$\begin{aligned} c : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto c(x) = \mu([\infty, x]). \end{aligned}$$

We notice that c is a cumulative distribution function. Indeed:

- c is increasing (it is non-decreasing but whatever):

$$\forall x_1 \leq x_2 \quad c(x_1) = \mu[\infty, x_1] \leq \mu[\infty, x_2] = c(x_2)$$

since $[\infty, x_1] \subseteq [\infty, x_2]$.

- c is right continuous: let $x \in \mathbb{R}$ be fixed and let $A_n = [-\infty, x + \frac{1}{n}]$, $\forall n \geq 1$. Clearly $A_n \in \mathcal{B}(\mathbb{R})$ and $A_n \searrow A := [-\infty, x]$. Then by properties of measures we know that

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

and since $\mu(A_n) = c(x + \frac{1}{n})$ and $\mu(A) = c(x)$ we conclude

$$\lim_{n \rightarrow \infty} c\left(x + \frac{1}{n}\right) = c(x).$$

Exercise 3

Let $X_1, X_2, \dots, X_n \in L^1$. Show that $\{X_1, X_2, \dots, X_n\}$ is uniformly integrable.

We know that a random variable is uniformly integrable if $\lim_{b \rightarrow \infty} \sup_{X \in K} \mathbb{E}[|X| \mathbb{1}_{\{|X|>b\}}] = 0$. So, since we are operating over a finite set, we can swap lim and max to conclude that this is equivalent to showing

$$\max_{i=1, \dots, n} \lim_{b \rightarrow \infty} \mathbb{E}[|X_i| \mathbb{1}_{\{X_i>b\}}] = 0 \tag{*}$$

Let's fix i . We have

$$\lim_{b \rightarrow \infty} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i|>b\}}] = \mathbb{E}\left[\underbrace{\lim_{b \rightarrow \infty} |X_i| \mathbb{1}_{\{|X_i|>b\}}}_{\text{this is 0 because } |X_i| \in L^1}\right] = 0. \tag{**}$$

Here we can swap expectation and limit thanks to the dominated convergence theorem, since $|X_i|$ (weakly) dominates $|X_i| \mathbb{1}_{\{|X_i|>b\}}$. Moreover, $\lim_{b \rightarrow \infty} |X_i| \mathbb{1}_{\{|X_i|>b\}}$ is 0 because being in L^1 means having "low" tails. So ** holds for all i and therefore * holds.

Exercise 4

Let $\{X_i\}$ be a family of uniformly integrable variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $X \in L^1$. Show that the family $\{X_i - X\}$ is uniformly integrable.

Let $K := \{X_i - X\}$. We use the $\varepsilon - \delta$ characterization for uniform integrability. So we should check that:

1. $\sup_{Y \in K} \mathbb{E}[Y] < \infty$;
2. $\forall \varepsilon \exists \delta : \text{if } F \in \mathcal{F} \text{ with } \mathbb{P}(F) \leq \delta \text{ then } \sup_{Y \in K} \mathbb{E}[|Y| \mathbb{1}_F] \leq \varepsilon$.

1. We have

$$\begin{aligned} \sup_{Y \in K} \mathbb{E}|Y| &= \sup_i \mathbb{E}|X_i - X| \\ &\leq \sup_i (\mathbb{E}|X_i| + \mathbb{E}|X|) \\ &\leq \underbrace{\mathbb{E}|X|}_{<\infty \text{ bc U.I.}} + \sup_i \underbrace{\mathbb{E}|X_i|}_{<\infty \text{ bc } \in L^1} < \infty \end{aligned}$$

The third step is simple triangle inequality...

2. Since $\{X_i\}$ is uniformly integrable, for any ε_1 there exists a δ_1 such that if $\mathbb{P}(F) < \delta_1$ then $\sup_i \mathbb{E} [|X_i| \mathbf{1}_F] \leq \varepsilon_1$. On the other hand, the fact that X is in L^1 means that $\{X\}$ is uniformly integrable because $\lim_{b \rightarrow \infty} \mathbb{E} [|X| \mathbf{1}_{\{|X| > b\}}]$, so for each ε_2 we can find a δ_2 such that if $\mathbb{P}(F) \leq \delta_2$ then $\sup_i \mathbb{E} [|X_i| \mathbf{1}_F] \leq \varepsilon_2$. Then, for any $\varepsilon > 0$, choose $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$. We need to find our δ so we say $\delta = \min\{\delta_1, \delta_2\}$. In this way, for $F \in \mathcal{F}$ such that $\mathbb{P}(F) \leq \delta$ (which means that is also $\leq \delta_1$ and $\leq \delta_2$) we have

$$\begin{aligned}\sup_{Y \in K} \mathbb{E} [|Y| \mathbf{1}_F] &= \sup_i \mathbb{E} [|X_i - X| \mathbf{1}_F] \\ &\leq \sup_i \mathbb{E} [|X_i| \mathbf{1}_F] + \mathbb{E} [|X| \mathbf{1}_F] \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Exercise 5

Let X and Y be random variables with values in $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ respectively. Let $Z = (X, Y)$ and

$$\mathbb{P}_Z(dx, dy) = \lambda \frac{e^{-\lambda x}}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dx dy \quad x \in \mathbb{R}_+, y \in \mathbb{R}$$

be the joint density function.

- ① Find the marginal distribution \mathbb{P}_X of X .
- ② Find the transition kernel K from $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mathbb{P}_Z = \mathbb{P}_X \times K \quad (\mathbb{P}_Z(dx, dy) = \mathbb{P}_X(dx) \cdot K(x, dy)).$$

Prove that K is a Markov transition kernel.

- ③ Find the marginal distribution $\mathbb{P}_Y \circ f_X$ in \mathbb{R} ; use that

$$\int_0^\infty \frac{1}{\sqrt{x}} \cdot e^{-\lambda x} \cdot \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \geq 0, \theta > 0$$

- ④ Determine whether X and Y are independent.

- ① Let's find $\mathbb{P}_X(A)$ for each $A \in \mathcal{B}(\mathbb{R}_+)$.

$$\begin{aligned}\mathbb{P}_X(A) &= \mathbb{P}_Z(A \times \mathbb{R}) = \int_{A \times \mathbb{R}} \mathbb{P}_Z(dx, dy) \\ &= \int_A \int_{\mathbb{R}} \lambda \frac{e^{-\lambda x}}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dx dy \\ &= \int_A \lambda e^{-\lambda x} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dy}_{\text{density of } N(0, x)} dx \\ &= \int_A \lambda e^{-\lambda x} \cdot 1 \cdot dx.\end{aligned}$$

So $\mathbb{P}_X(dx) = \lambda e^{-\lambda x} \sim \text{Exp}(\lambda)$.

- ② Since we can now write

$$\mathbb{P}_Z(dx, dy) = \mathbb{P}_X(dx) \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dy$$

then we see that

$$K(x, dy) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dy$$

or, more formally, for $B \in \mathcal{B}(\mathbb{R})$ we have

$$K(x, B) = \int_B \frac{1}{\sqrt{2\pi x}} e^{-\frac{y^2}{2x}} dy.$$

Let's now check that $K(x, B)$ is a Markov transition kernel:

- it is a transition kernel because
 - for $B \in \mathcal{B}(\mathbb{R})$ fixed, $K(\cdot, B)$ is positive and measurable (it is even continuous);
 - for $x \in \mathbb{R}_+$ fixed, $K(x, \cdot)$ is a measure on $\mathcal{B}(\mathbb{R})$: in particular by inspection we see that it is a Gaussian measure with mean 0 and variance x .
- It is a Markov kernel because $K(x, \mathbb{R}) = 1 \forall x$ since $K(x, \cdot)$ is a probability measure $\forall x \in \mathbb{R}$.

③ Let's find $\mathbb{P}_Y(B)$ for every $B \in \mathcal{B}(\mathbb{R})$.

$$\begin{aligned}\mathbb{P}_B(B) &= \mathbb{P}_Z(\mathbb{R}_+ \times B) = \int_{\mathbb{R}_+ \times B} \lambda e^{-\lambda x} \frac{e^{\frac{y^2}{2x}}}{\sqrt{2\pi x}} dx dy \\ &= \int_B \left(\int_{\mathbb{R}_+} \lambda \frac{e^{-\lambda x}}{\sqrt{2\pi x}} e^{\frac{-y^2}{2x}} dx \right) dy\end{aligned}$$

Now use the hint with $\gamma = \lambda$, $\beta = \frac{y^2}{2}$ to get

$$\begin{aligned}\frac{1}{\sqrt{2\pi}} \lambda \int_{\mathbb{R}_+} e^{-\lambda x} \frac{1}{\sqrt{x}} e^{\frac{y^2}{2x}} dx &= \frac{\lambda}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{\lambda}} \cdot e^{-2\sqrt{\lambda \cdot \frac{y^2}{2}}} \\ &= \sqrt{\frac{\lambda}{2}} e^{-\sqrt{2\lambda} \cdot |y|}.\end{aligned}$$

So our function becomes

$$\int_B \frac{1}{2} \sqrt{2\lambda} e^{-\sqrt{2\lambda} |y|} dy.$$

Which is the probability density function of a two-sided exponential random variable. Thus $\mathbb{P}_Y(dy) = \frac{1}{2} \sqrt{2\lambda} e^{-\sqrt{2\lambda} |y|} dy$.

④ We observe that $\mathbb{P}_X \cdot \mathbb{P}_Y \neq \mathbb{P}_Z$ so X and Y are not independent.

4.1.3 Exercise class 3

Exercise 1

Let X, Y be independent random variables $\sim N(0, 1)$. Compute the joint distribution of $(X + Y, X - Y)$.

The vector is Gaussian, since it is a linear combination of Gaussian random variables.

$$\begin{aligned}W &:= (X + Y, X - Y) = (W_1, W_2) \\ Z &:= (X, Y).\end{aligned}$$

We write W in terms of Z :

$$W = \mu + \mathbf{A}Z \quad \text{for some } \mu = (\mu_1, \mu_2) \text{ and } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Clearly $\mu = (0, 0)$ because

$$\begin{aligned}\mathbb{E}W_1 &= \mathbb{E}X + Y = \mathbb{E}X + \mathbb{E}Y = 0 \\ \mathbb{E}W_2 &= \mathbb{E}X - Y = \mathbb{E}X - \mathbb{E}Y = 0\end{aligned}$$

So now we have

$$\begin{aligned}W_1 &= X + Y = a_{11}X + a_{12}Y \\ W_2 &= X - Y = a_{21}X + a_{22}Y \implies \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\end{aligned}$$

Let's calculate the covariance matrix Γ of W , given by

$$\Gamma = \mathbf{A} \mathbf{A}^\top = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Since $\det(\Gamma) = 4 \neq 0$ then Γ has full rank so the random variable W has density given by

$$f_W(W_1, W_2) = \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{\sqrt{\det(\Gamma)}} e^{-\frac{1}{2}(W_1 W_2) \Gamma^{-1} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}}.$$

Compute Γ^{-1} :

$$\Gamma^{-1} = \frac{1}{\det(\Gamma)} \begin{bmatrix} 2 & -0 \\ -0 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

And now plug the results inside the density formula:

$$\begin{aligned} f_W(W_1, W_2) &= \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{\sqrt{\det(\Gamma)}} e^{-\frac{1}{2}(W_1 W_2) \Gamma^{-1} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}} \\ &= \frac{1}{(2\pi)^{\frac{2}{2}}} \cdot \frac{1}{2} e^{-\frac{1}{2}(W_1 W_2) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}} \\ &= \frac{1}{4\pi} e^{-\frac{1}{2}(\frac{1}{2}W_1^2 + \frac{1}{2}W_2^2)}. \end{aligned}$$

So this is the density of the random variable $W = (X + Y, X - Y)$.

Remark

Since $\Gamma_{12} = \Gamma_{21} = 0$ we have $W_1 \perp\!\!\!\perp W_2$. We can see this also because the joint distribution factorizes in the product of the marginal distributions.

$$f_W(W_1, W_2) = \underbrace{\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{1}{4}W_1^2}}_{\sim n(0,2)} \cdot \underbrace{\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{1}{4}W_2^2}}_{\sim n(0,2)}$$

Exercise 2

Let X, Y be independent real random variables with distributions μ_X and μ_Y respectively. Show that the distribution of $Z := X + Y$ denoted μ_Z is given by

$$\mu_Z(B) = \int_{\mathbb{R}} \mu_X(B - x) \mu_Y(dx) \quad \forall B \in \mathcal{B}(\mathbb{R})$$

where

$$\{B - x\} := \{y - x \in B \mid \text{for all } y \in B\}.$$

Note that since $X + Y \sim Z + Y$ we also have

$$\mu_Z(B) = \int_{\mathbb{R}} \mu_X(B - x) \mu_X(dx)$$

Take $B \in \mathcal{B}(\mathbb{R})$. We have

$$\begin{aligned} \mu_Z(B) &= \mathbb{P}(Z \in B) \\ &= \mathbb{P}(X + Y \in B) \\ &= \mathbb{E} [\mathbb{1}_{\{X+Y \in B\}}]. \end{aligned}$$

We know that $\mu_{(X,Y)} = \mu_X \cdot \mu_Y$ since $X \perp\!\!\!\perp Y$, so we have that

$$\begin{aligned} \mu_Z(B) &= \mathbb{P}(X + Y \in B) \\ &= \mathbb{E} [\mathbb{1}_{\{X+Y \in B\}}] \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X+Y \in B\}} \mu_{(X,Y)}(dx, dy) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X+Y \in B\}} \mu_X(dx) \mu_Y(dy) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(x + y) \mu_X(dx) \mu_Y(dy). \end{aligned} \tag{*}$$

Now consider

$$\mathbb{1}_B(x+y) = \mathbb{1}_{\{B-x\}}(y) = \begin{cases} 1 & \text{if } x+y \in B \iff y \in \{B-x\} \\ 0 & \text{if } x+y \notin B \iff y \notin \{B-x\} \end{cases}.$$

Now consider *:

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_B(x+y) \mu_X(dx) \mu_Y(dy) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{B-x\}}(y) \mu_X(dx) \mu_Y(dy) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \underbrace{\mathbb{1}_{\{B-x\}}(y) \mu_Y(dy)}_{\int_{\{B-x\}} \mu_Y(dy) = \mu_Y(B-x)} \right) \mu_X(dx) \\ &= \int_{\mathbb{R}} \mu_Y(B-x) \mu_X(dx) \end{aligned}$$

Exercise 3

Let X, Y be independent random variables with densities f_X, f_Y . Show that $Z := X + Y$ also has a density function given by

$$f_Z(z) := \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx, \quad z \in \mathbb{R}.$$

Take $B \in \mathcal{B}(\mathbb{R})$. The law of Z, μ_Z , is such that

$$\mu_Z(B) = \mathbb{P}(Z \in B)$$

and we want to show that indeed

$$\mathbb{P}(Z \in B) = \int_B f_Z(z) dz.$$

Using what we found in exercise 2 we have that

$$\begin{aligned} \mu_Z(B) &= \int_{\mathbb{R}} \mu_Y(B-x) \mu_X(dx) \\ &= \int_{\mathbb{R}} \int_{B-x} f_Y(y) dy f_X(x) dx. \end{aligned}$$

If $y \in \{B-x\}$ it means that $y = t - x$ for some $t \in B$. So we can make a change of variable $y = t - x$:

$$\int_{\{B-x\}} f_Y(y) dy = \int_B f_Y(t-x) dt$$

which implies

$$\begin{aligned} \mu_Z(B) &= \int_{\mathbb{R}} \int_B f_Y(y)(t-x) dt f_X(x) dx \\ &= \int_B \left(\int_{\mathbb{R}} f_Y(t-x) f_X(x) dx \right) dt \\ &= \int_B f_Z(t) dt. \end{aligned}$$

Exercise 4

Let (X_n) be a sequence of independent random variables. Show that if $\lim_{n \rightarrow \infty} X_n = X$ P.a.s, then $Y = c$ P.a.s. for some $c \in \mathbb{R}$.

Start from this set:

$$\begin{aligned} A &= \left\{ \omega : \lim_n X_n(\omega) \text{ exists} \right\} \\ &= \left\{ \omega : \liminf_n X_n(\omega) = \limsup_n X_n(\omega) \right\}. \end{aligned}$$

We know that $\mathbb{P}(A) = 1$ because we know that X_n always converges and we want to show that $A \in \tau$ because this (together with independence) would allow us to apply Kolmogorov's 0-1 law's corollary that tells us that if an event is part of a tail σ -algebra then it is a constant³. Let's show that $\limsup_n X_n$ is τ -measurable.

$$\limsup_n X_n = \inf_n \sup_{\substack{m \geq n \\ \text{under}}} X_m$$

is measurable with respect to $\bigvee_{m \geq n} \sigma(X_m) = \tau_n$

Thus $\inf_n \sup_{m \geq n} X_m$ is measurable with respect to

$$\bigcap_{n \geq 1} \bigvee_{m \geq n} \sigma(X_m) = \bigcap_{n \geq 1} \tau_n = \tau$$

so $\limsup_n X_n$ is τ -measurable. Similarly for \liminf we get that is is τ -measurable so $A \in \tau$. We recall that X is such that

$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega) \quad \forall \omega \in A \in \tau.$$

Introduce

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

We introduced this new random variable because we are technically working with the *limit* of X_n (which is a sequence) and measurability of limits is not a beautiful thing to handle. By definition $X = Y \mathbb{P} - \text{a.s.}$ (because $\mathbb{P}(A^c) = 0$) and Y is τ -measurable because $A \in \tau$. This implies that X is τ -measurable. By corollary of the Kolmogorov's 0-1 law, $X = c \mathbb{P} - \text{a.s.}$ for some $c \in \mathbb{R}$.

Exercise 5

I.e. (X_n) be a sequence of i.i.d. random variables with $X_n \sim \mathcal{E}_{\lambda}$, ($\lambda > 0$) with density

$$f(x) = \begin{cases} e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else.} \end{cases}$$

Determine whether

$$\frac{1}{n} \max\{X_1, X_2, \dots, X_n\} \xrightarrow{\mathbb{P}} 0$$

We guess yes because why not. Let $\varepsilon > 0$. Take the set

$$B_n := \{\omega : X_n(\omega) \geq n\varepsilon\}.$$

This gives us

$$\begin{aligned} \mathbb{P}(B_n) &= \mathbb{E}[1_{B_n}] \\ &= \int_{n\varepsilon}^{+\infty} e^{-x} dx = e^{-n\varepsilon}. \end{aligned}$$

Thus $\sum_n \mathbb{P}(B_n) = \sum_n e^{-n\varepsilon} < \infty$. By Borel-Cantelli 1 we have

$$\mathbb{P}(\{B_n \text{ i.o.}\}) = 0.$$

Take the complement:

$$\mathbb{P}(\{B_n^c \text{ f.o.}\}) = 1.$$

For every $\omega \in \{B_n^c \text{ f.o.}\}$ there exists $m(\omega, \varepsilon)$ such that $\forall n > m(\omega, \varepsilon)$ then $\underbrace{X_n(\omega)}_{B_n^c} < n\varepsilon$. Set

$$M(\omega, \varepsilon) := \max\{X_1(\omega), X_2(\omega), \dots, X_{m(\omega, \varepsilon)}(\omega)\}$$

so that for $\forall n$ we have

$$\max\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} \leq M + n\varepsilon$$

so

$$\frac{1}{n} \max\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} \leq \frac{M(\omega, \varepsilon)}{n} + \varepsilon.$$

³If you have already read the theory then you know how this thing caused me a mental breakdown.

Check the limit:

$$\begin{aligned}\limsup_n \frac{1}{n} \max\{X_1(\omega), \dots, X_n(\omega)\} &\leq \limsup_n \frac{M(\omega, \varepsilon)}{n} + \varepsilon \\ &= \lim_n \frac{M(\omega, \varepsilon)}{n} + \varepsilon = \varepsilon.\end{aligned}$$

Now we take the limit as $\varepsilon \rightarrow 0$:

$$\limsup_n \frac{1}{n} \max\{X_1(\omega), \dots, X_n(\omega)\} \leq \lim_{\varepsilon \rightarrow 0} \varepsilon = 0$$

Thus, using the fact that X_i are positive, we conclude

$$\lim_{n \rightarrow \infty} \frac{1}{n} \max\{X_1(\omega), \dots, X_n(\omega)\} = 0.$$

Since this holds for $\forall \omega \in \{B_n^x \text{ ev.}\}$ and $\mathbb{P}(\{B_n^c \text{ ev.}\}) = 1$ we have

$$\frac{1}{n} \max\{X_1, \dots, X_n\} \rightarrow 0 \quad \text{a.s.}$$

Exercise 6

Let (X_i) be independent random variables non-negative such that

$$\mathbb{P}(X_i > \delta) \geq \varepsilon, \quad (1)$$

for some fixed $\delta, \varepsilon > 0$. Determine whether $S_n = \sum_{i=1}^n X_i$ converges a.s. as $n \rightarrow \infty$.

Guess: S_n diverges to $+\infty$ because the random variables X_i are such that $\{X_i > \delta\}$ has positive probability. We claim

$$\limsup_n \{X_i \geq \delta\} \subset \{\lim_{n \rightarrow \infty} S_n = \infty\} \quad (\heartsuit)$$

Let's prove it.

$$\begin{aligned}\omega \in \limsup_i \{X_i \geq \delta\} &= \bigcap_{i \geq 1} \bigcup_{m \geq i} \{X_m \geq \delta\} \\ &= \{X_i \geq \delta \text{ i.o.}\}.\end{aligned}$$

This means that \exists a subsequence i_k such that $X_{i_k}(\omega) \geq \delta$. Summing over k we get

$$\sum_{k=1}^{\infty} X_{i_k}(\omega) \geq \sum_{k=1}^{\infty} X_{i_k}(\omega) = \infty.$$

So taking the limit we obtain

$$\begin{aligned}\lim_{n \rightarrow \infty} S_n^{(\omega)} &= \lim_{n \rightarrow \infty} \sum_{i=1}^n X_i(\omega) \\ &= \sum_{i=1}^{\infty} X_i(\omega) \\ &\geq \sum_{k=1}^{\infty} X_{i_k}(\omega) = \infty \implies \lim_{n \rightarrow \infty} S_n(\omega) = \infty\end{aligned}$$

hence \heartsuit^4 holds. We now apply Borel-Cantelli 2.

⁴Yes, I have ONLY NOW discovered dingbats. My future endeavours will have more of them, do not worry. Even if they look like shit with this font, on god.

Revise with Kotatsu!

Proposition 4.1.1

"Divergence". Let $(A_n)_n$ be a sequence of pairwise independent events.

$$\text{If } \sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty \text{ then } \sum_{n=1}^{\infty} \mathbb{1}_{A_n} = \infty \text{ a.s.}$$

$\underbrace{\phantom{\sum_{n=1}^{\infty}}}_{=\mathbb{P}(\{A_n \text{ i.o.}\})}$

An equivalent formulation of Borel-Cantelli 2 is:

Proposition 4.1.2

Take a sequence of Bernoulli random variables B_n defined as

$$B_n = \begin{cases} 1 & \text{on } A_n \\ 0 & \text{on } A_n^c \end{cases} \quad (\text{or, more simply, } B_n = \mathbb{1}_{A_n})$$

so that B_n is a Bernoulli random variable and we have

$$\mathbb{E}B_n = \mathbb{P}(A_n).$$

The B_n are pairwise independent.

$$\text{If } \sum_{n=1}^{\infty} \mathbb{E}B_n = \infty \text{ then } \sum_{n=1}^{\infty} B_n = \infty \text{ a.s.}$$

Remember that $\limsup_n \{X_n = 1\}$ is the same as saying that $\{X_n = 1 \text{ i.o.}\}$.

In our case we have $B_i = \{X_i \geq \delta\}$. We know that $\mathbb{P}(B_i) \geq \epsilon$ by assumption, so

$$\sum_{i=1}^{\infty} \mathbb{P}(B_i) = \infty.$$

Moreover, the B_i are independent. Hence we can apply BC-2 and say that

$$\mathbb{P}(\{X_i \geq \delta\} \text{ i.o.}) = 1 \iff \mathbb{P}(\limsup_i \{X_i \geq \delta\}) = 1.$$

By  we have

$$1 = \mathbb{P}(\limsup_n \{X_i \geq \delta\}) \leq \mathbb{P}(\lim_n S_n = \infty)$$

hence

$$\mathbb{P}(\lim_n S_n = \infty) = 1$$

which means that we have almost sure divergence of the sum.

4.1.4 Exercise class 4

Revise with Kotatsu!

Hi bitch. Remember all the types of convergence? Me neither!

 **Convergence in L^p :**

$$X_n \xrightarrow{L^p} X$$

if $X_n, X \in L^p$ and

$$\mathbb{E}|X_n - X|^p \rightarrow 0 \text{ as } n \rightarrow \infty.$$

A necessary (but not sufficient...) condition for convergence in L^p is that

$$\mathbb{E}|X_n|^p \rightarrow \mathbb{E}|X|^p.$$

Convergence in probability:

$$X_n \xrightarrow{\mathbb{P}} X$$

if $\forall \varepsilon > 0$ we have

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Almost sure convergence:

$$X_n \xrightarrow{\text{a.s.}} X$$

if $\exists \Omega' \subset \Omega$ with $\mathbb{P}(\Omega') = 1$ such that

$$X_n(\omega) \rightarrow X(\omega) \quad \forall \omega \in \Omega'.$$

Convergence in distribution:

$$X_n \xrightarrow{d} X$$

if $F_n(x) \rightarrow F(x)$ for any point x such that $F(x)$ is continuous. Here F_n and F denote the cumulative distribution functions of X_n and X .

Link between different convergence modes:

$$\begin{array}{c} X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X \iff X_n \xrightarrow{\text{a.s.}} X \\ \Downarrow \\ X_n \xrightarrow{d} X. \end{array}$$

Properties:

- if $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$ then $X = Y$ a.s.;
- if $X_n \xrightarrow{L^p} X$ and $\lim_n X_n$ exists a.s., then $X_n \xrightarrow{\text{a.s.}} X$.

Exercise 1

Let $\{X_n\}$ be a sequence of independent random variables with $X_n = \text{Det}\left(\frac{1}{n}\right)$. Determine whether X_n converges a.s. or in L^p for some $p > 1$.

I will now stop⁵ with the stupid LATEX symbols.

L^p convergence: note that

$$\mathbb{E}|X_n|^p = \mathbb{E}X_n^p = 1^p \cdot \frac{1}{n} + 0^p \cdot \left(1 - \frac{1}{n}\right) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0, \quad \forall p > 0.$$

Hence the candidate limit is $X = 0$. Let's check

$$\mathbb{E}|X_n - 0|^p = \mathbb{E}|X_n|^p = \frac{1}{n} \rightarrow 0$$

so

$$X_n \xrightarrow{L^p} X, \quad \forall p \geq 1.$$

Almost sure convergence: we note that

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty$$

⁵This is a lie.

hence by BC-2 we have that

$$\mathbb{P}(\limsup_n \{X_n = 1\}) = 1.$$

This implies that we cannot have $X_n \xrightarrow{n} 0$.

Exercise 2

Let $(X_n)_n$ be a sequence of i.i.d. random variables with $X_1 \sim U(0, 1)$. Let $Y_n = \min\{X_1, X_2, \dots, X_n\}$. Determine whether Y_n converges a.s. and in L^p for some $p \geq 1$.

We must figure out the distribution of Y_n . for $x \in (0, 1)$ we have:

$$\begin{aligned}\mathbb{P}(Y_n > x) &= \mathbb{P}(X_1 > x, X_2 > x, \dots, X_n > x) \\ &= \mathbb{P}(X_1 > x) \cdot \mathbb{P}(X_2 > x), \dots, \mathbb{P}(X_n > x) \\ &= (1-x)^n.\end{aligned}$$

Therefore,

$$\mathbb{P}(Y_n \leq x) = 1 - (1-x)^n = \int_0^x f_{Y_n}(y) dy$$

so taking the derivative we have:

$$f_{Y_n}(y) = \begin{cases} n(1-y)^{n-1} & \text{for } y \in (0, 1) \\ 0 & \text{else.} \end{cases}$$

Let's check if $\mathbb{E}|Y_n|^p$ converges for some p . We start with $p = 1$.

Revise with Kotatsu!

Remember integration by parts? Lmao.

$$\int f g' = f g - \int f' g$$

$$\begin{aligned}\mathbb{E}|Y_n| &= \mathbb{E}Y_n \\ &= \int_0^1 \underbrace{y}_f \underbrace{n(1-y)^{n-1}}_{g'} dy \\ &= y(-(1-y)^n) \Big|_0^1 - \int_0^1 -(1-y)^n dy \\ &= 0 + -\frac{(1-y)^{n+1}}{n+1} \Big|_0^1 \\ &= 0 + \frac{1}{1+n}\end{aligned}$$

Hence $\mathbb{E}|Y_n| \xrightarrow{n \rightarrow \infty} 0$, so the candidate Y such that $\mathbb{E}|Y_n - Y| \rightarrow 0$ is 0. This implies $\mathbb{E}|Y_n - 0| = \mathbb{E}|Y_n| \rightarrow 0$

so that we have $Y_n \xrightarrow{L^1} 0$. What about convergence in L^p though? First check a.s. convergence.

By definition of Y_n we have $Y_{n+1}(\omega) \leq Y_n(\omega)$ for almost all ω ; this means that Y_n is monotone \mathbb{P} -almost surely. This implies that Y_n has a limit \mathbb{P} -almost surely, say Y (moreover, $Y_n(\omega) \geq 0$ hence $Y(\omega) \geq 0$ \mathbb{P} -a.s.). By uniqueness of the limit⁶ we have that $Y = 0$. Hence we conclude $Y_n \xrightarrow{\text{a.s.}} 0$.

To check for convergence in L^p , if we can apply the dominated convergence theorem then we have:

$$\lim_{n \rightarrow \infty} \mathbb{E}|Y_n - Y|^p = \mathbb{E} \left[\lim_{n \rightarrow \infty} |Y_n|^p \right] = 0.$$

Here dominated convergence works because $0 \leq X_k \leq 1$ hence $|Y_n| \leq 1$ and $|Y_n|^p \leq 1$ where 1 is integrable. Moreover, the a.s. limit $Y = 0$ is also integrable.

⁶ $Y_n \xrightarrow{L^1} 0 \implies Y_n \xrightarrow{\mathbb{P}} 0; Y_b \xrightarrow{\text{a.s.}} Y \implies Y = 0 \mathbb{P} - \text{a.s.}$

Exercise 3

Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \lambda_{[0,1]})$. Let

$$X_n(\omega) = n \mathbb{I}_{[\frac{1}{n}, \frac{1}{n-1}]}(\omega).$$

Determine whether $\sum X_n$ converges a.s. or in L^p for some $p \geq 1$.

Let's calculate $\mathbb{E}|X_n|^p$ for some $p \geq 1$.

$$\mathbb{E}|X_n|^p = n^p \cdot \mathbb{P}\left(\left[0, \frac{1}{n}\right]\right) + 0^p \cdot \mathbb{P}\left(\left[\frac{1}{n}, 1\right]\right) = n^p \cdot \frac{1}{n} = n^{p-1}$$

which diverges for $n \rightarrow \infty$ if $p > 1$. Thus there can be no L^p -convergence for $p > 1$.

Let's check a.s. convergence. For any $\omega > 0 \exists N = N(\omega)$ s.t. $\forall n > N$ then $X_n(\omega) = 0$ (since the interval $[0, \frac{1}{n}]$ shrinks). Hence $X_n(\omega) \rightarrow 0 \forall \omega \mathbb{P}$ -a.s. given that $\mathbb{P}([0, 1]) = 1$.

A It does not converge for $\omega = 0$ but $\{\omega = 0\}$ has Lebesgue measure zero.

Back to L^p -convergence with $p = 1$: if $p = 1$ then $\mathbb{E}|X_n| = 1$ but by uniqueness of the limit we should have $X = 0$ and $\mathbb{E}|X| = 0 \neq 1$ so we do not have convergence in L^1 .

Exercise 4

Let $(X_n)_n$ be a sequence of independent random variables such that

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

Determine whether X_n converges to 1 L^p or almost surely.

$$\mathbb{E}|X_n| = \mathbb{E}|X_n| = n^2 \cdot \frac{1}{n^2} + 0 = 1.$$

Let us check whether $X_n \xrightarrow{L^p} 1$.

$$\begin{aligned} \mathbb{E}|X_n - 1| &= |n^2 - 1| \cdot \frac{1}{n^2} + |0 - 1| \left(1 - \frac{1}{n^2}\right) \\ &= \left|\frac{n^2 - 1}{n^2}\right| + \left|\frac{n^2 - 1}{n^2}\right| \\ &= 2 \left|\frac{n^2 - 1}{n^2}\right| \rightarrow 2 \neq 0 \end{aligned}$$

so there is no L^1 convergence. Since $L^p \implies L^1$ $p \geq 1$ there is no L^p convergence either. What about a.s. convergence?

We cannot argue directly as in Exercise 3, but do so through Borel-Cantelli: we know that $\forall \varepsilon > 0, \mathbb{P}(|X_n| \geq \varepsilon) = \frac{1}{n^2}$, hence

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq \varepsilon) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Thus by Borel-Cantelli 1 we have

$$\mathbb{P}\left(\limsup_n \{|X_n| \geq \varepsilon\}\right) = 0$$

and, taking the complement,

$$\mathbb{P}\left(\liminf_n \{|X_n| < \varepsilon\}\right) = 1.$$

Thus, $\forall \varepsilon > 0 \exists N = N(\varepsilon)$ such that $|X_n(\omega)| < \varepsilon$ for \mathbb{P} -almost all ω and this implies $X_n \xrightarrow{\text{a.s.}} 0$.

Exercise 5

Let $(X_n)_n$ be a sequence of independent random variables with $X_n \sim \text{Exp}(\frac{1}{n})$. Determine if $\sum X_n = \min(X_1, \dots, X_n)$ converges in probability.

Revise with Kotatsu!

The exponential distribution has the following cumulative distribution function:

$$\mathbb{P}(X_n \geq x) = \int_0^x \frac{1}{k} e^{-\frac{z}{k}} dz = -e^{-\frac{z}{k}} \Big|_0^x = 1 - e^{-\frac{x}{k}}$$

Let's figure out the distribution of Y_n . We have

$$\begin{aligned}\mathbb{P}(Y_n > y) &= \mathbb{P}(X_1 > y, X_2 > y, \dots, X_n > y) \\ &= \prod_{i=1}^n \mathbb{P}(X_i > y) \\ &= \prod_{i=1}^n e^{-\frac{y}{i}} \\ &= e^{-y \cdot \sum_{i=1}^n \frac{1}{i}}\end{aligned}$$

hence

$$Y_n \sim \text{Exp}\left(\sum_{i=1}^n \frac{1}{i}\right).$$

At the limit we would have $f_Y(y) = \infty \cdot e^{-\infty \cdot y}$ but we guess that $e^{-\infty}$ is stronger than ∞ and the limit is zero. Thus we have, for $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - 0| > \varepsilon) = \lim_{n \rightarrow \infty} e^{\varepsilon \cdot \sum_{i=1}^n \frac{1}{i}} = 0$$

So by definition $Y_n \xrightarrow{\mathbb{P}} 0$.

Exercise 6

Let $(X_n)_n$ be a sequence of i.i.d. random variables with $X_n \sim \text{Exp}(1)$. Let

$$Y_n = \begin{cases} \frac{1}{n} & \text{if } X_n < \log n \\ 1 & \text{else} \end{cases}$$

Determine, for varying λ , the convergence in probability of $\ln(Y_n)$ to some $\nu \geq 0$.

First, we calculate the distribution of Y_n :

$$\mathbb{P}(Y_n \geq y) = \begin{cases} 1 & \text{if } Y \geq 1 \\ \mathbb{P}(X_n < \log n) & \text{if } \frac{1}{n} \leq y < 1 \\ 0 & \text{if } y < \frac{1}{n} \end{cases}$$

where

$$\mathbb{P}(X_n < \log n) = 1 - e^{\lambda \cdot \log n} = 1 - e^{\log(n^{-\lambda})} = 1 - n^{-\lambda}.$$

From this, it looks like $\ln(Y_n) \rightarrow 1$ as $n \rightarrow \infty$. The complement of the cumulative distribution function, which we are interested in to compute the limit, is

$$\begin{aligned}\mathbb{P}(Y_n > y) &= 1 - \mathbb{P}(Y_n \leq y) \\ &= \begin{cases} 0 & \text{if } Y \geq 1 \\ n^{-\lambda} & \text{if } \frac{1}{n} \leq y < 1 \\ 1 & \text{if } y < \frac{1}{n} \end{cases}\end{aligned}$$

and it looks like Figure 4.5.

- Convergence in \mathbb{P} : we have that $\forall \varepsilon > 0$ (since $Y_n \geq 0$),

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - 0| > \varepsilon) = \lim_{n \rightarrow \infty} \mathbb{P}(Y_n > \varepsilon) = 0$$

which implies

$$Y_n \xrightarrow{\mathbb{P}} 0, \quad \forall \lambda > 0.$$

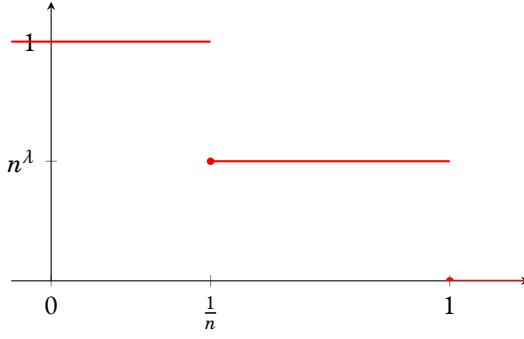


Figure 4.5: It's really just like a cumulative distribution “backwards”...

- Almost sure convergence: We now tackle almost sure convergence and we try to apply Borel-Cantelli. Let $\varepsilon < 1$ and consider

$$\sum_{i=1}^{\infty} \mathbb{P}(Y_n > \varepsilon) = \sum_{n=1}^{N^*} 1 + \sum_{n=N^*+1}^{\infty} n^{-\lambda} \leq N^* + \sum_{i=1}^{\infty} n^{-\lambda} < \infty \quad \text{if } \lambda > 1$$

where $N^* = \max \{n : \varepsilon < \frac{1}{n}\}$. Thus, by BC1 (given that $\lambda > 1$),

$$\mathbb{P} \left(\limsup_k \{|Y_n| > \varepsilon\} \right) = 0 \iff \mathbb{P} \left(\limsup_k \{|Y_n| \leq \varepsilon\} \right) = 1$$

i.e., $\forall \varepsilon \exists N = N(\varepsilon)$ such that $|Y_n(\omega)| \leq \varepsilon$ \mathbb{P} -almost all ω . This means that

$$Y_n \xrightarrow{\text{a.s.}} 0 \quad \text{if } \lambda > 1.$$

On the other hand, if $\lambda \leq 1$ we instead have

$$\sum_{n=1}^{\infty} \mathbb{P}(Y_n > \varepsilon) = \sum_{n=1}^{N^*} 1 + \sum_{n=N^*+1}^{\infty} n^{-\lambda} \leq N^* + \sum_{i=1}^{\infty} n^{-\lambda} = \infty \quad \text{if } \lambda \leq 1.$$

Thus, by BC2, we get (since $X_n \perp \!\!\! \perp Y_n$)

$$\mathbb{P} \left(\limsup_n \{Y_n > \varepsilon\} \right) = 1$$

So the sequence Y_n cannot converge almost surely to 0. By uniqueness of the limit, this implies that $Y_n \not\xrightarrow{\text{a.s.}} Y$ to any Y .

- Convergence in L^p : let's check the necessary condition

$$\mathbb{E}|Y_n|^p \rightarrow \mathbb{E}|Y|^p = 0.$$

Again, by uniqueness of the limit, we must have $Y = 0$ \mathbb{P} -a.s. Considering that $X_n \sim \text{Exp}(\lambda)$, we have

$$\begin{aligned} \mathbb{E}|Y_n|^p &= \mathbb{E}Y_n^p \\ &= \frac{1}{n^p} \mathbb{P}(X_n < \log n) + 1 \cdot \mathbb{P}(X_n \geq \log n) \\ &= \frac{1}{n^p} \left(1 - e^{\lambda \log n} \right) + e^{-\lambda \log n} \\ &= \frac{1}{n^p} \left(1 - n^{-\lambda} \right) + n^{-\lambda} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \forall \lambda, \forall \lambda > 0. \end{aligned}$$

Hence,

$$\mathbb{E}|Y_n - 0|^p = \mathbb{E}|Y_n|^p \rightarrow 0$$

that is

$$Y_n \xrightarrow{L^p} 0 \quad \forall p \geq 1, \forall \lambda > 0.$$

4.1.5 Exercise class 5

Revise with Kotatsu!

- ◆ **Strong Law Of Large Numbers (SLLN):** if X_n are pairwise independent and identically distributed as X . If $\mathbb{E}X$ exists ($\pm\infty$ is allowed) then

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mu.$$

- ◆ **Weak Law Of Large Numbers (WLLN):**

- (a) if X_n are pairwise independent and identically distributed as X with $\mathbb{E}X = \mu < \infty$ and $\text{Var } X = \sigma^2 < \infty$ then

$$\bar{X}_n \rightarrow \mu \quad \text{in } L^2, \text{ in } \mathbb{P} \text{ and a.s.}$$

- (b) If X_n are uncorrelated and $\sum_n \text{Var} \left(\frac{X_n}{b_n} \right) < \infty$ for some (b_n) strictly positive and increasing to ∞ then

$$\frac{\sum_{i=1}^n X_i - \mathbb{E}(\sum_{i=1}^n X_i)}{b_n} \rightarrow 0 \quad \text{in } L^2 \text{ and in } \mathbb{P}.$$

If the random variables are independent, the convergence holds almost surely.

- ◆ **Central Limit Theorem (CLT)** (easy version, whatever this may mean): if X_n are i.i.d. random variables with $\mathbb{E}X_i = \mu < \infty$, $\text{Var } X = \sigma^2 < \infty$ then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} N(0, 1).$$

- ◆ **Weak Convergence** (for probability measures): a sequence $(\mu_n)_n$ of probability measures converges weakly to μ if for any f bounded and continuous function we have

$$\int f \, d\mu_n \rightarrow \int f \, d\mu.$$

In this case we write $\mu_n \xrightarrow{\text{weak}} \mu$.

Proposition 4.1.3

$$X_n \xrightarrow{\text{d}} X \iff \mu_n \xrightarrow{\text{weak}} \mu$$

where μ_n is the law of X_n and μ is the law of X .

Proposition 4.1.4

1. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{d}} X;$
2. $X_n \xrightarrow{\text{d}} X$ and $X(\omega) = x$ a.s. $\implies X_n \xrightarrow{\mathbb{P}} X$.

Exercise 1

Let (Y_n) be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}Y > \mu > 0$ and $\mathbb{E}|Y| < \infty$. Let $K \sim \text{Bin}(p)$, $n \in \{0, 1\}$ be a random variable independent of $\{Y_n\}_{n \geq 1}$ and let $Z_n = Y_n^{K_n}$ with $Z_{n0} = \frac{1}{n} \sum_{i=1}^n Z_{ni}$.

- A) Determine the convergence of Z_n in the a.s. sense.

- B) Compute $\mathbb{P}(A)$ where $A = \{\omega : \bar{Z}_n \rightarrow \mu\}$.

1. Note that $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n YX_n = Y\bar{X}_n$. Moreover $\mathbb{E}Z_n = \mathbb{E}Y\mathbb{E}X_n = \mu\mathbb{E}Y < \infty$, but Z_n are not independent: hence we cannot apply SLLN... However, we can apply it to \bar{X}_n . Let's guess the limit,

which is $Y \mathbb{E} X$ (because Y does not average). We have:

$$|\bar{Z}_n - Y \mathbb{E} X| = \left| Y \sum_{i=1}^n X_i - Y \mathbb{E} X \right| \leq |Y| \left| \sum_{i=1}^n X_i \mathbb{E} X \right| \leq |\bar{X}_n - \mathbb{E} X|.$$

Hence $\forall \varepsilon > 0$ and for any $n \geq 1$:

$$\left\{ |\bar{Z}_n - Y \mathbb{E} X_n| \leq \varepsilon \right\} \supset \left\{ |\bar{X}_n - \mathbb{E} X_n| \leq \varepsilon \right\}$$

which means

$$\bigcap_{n=k}^{\infty} \left\{ |\bar{Z}_n - Y \mathbb{E} X| \leq \varepsilon \right\} \subset \bigcap_{n=k}^{\infty} \left\{ |\bar{X}_n - \mathbb{E} X_n| \leq \varepsilon \right\}, \quad \forall k$$

and also

$$\begin{aligned} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ |\bar{Z}_n - Y \mathbb{E} X| \leq \varepsilon \right\} &\subset \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{ |\bar{X}_n - \mathbb{E} X_n| \leq \varepsilon \right\}, \quad \forall k \\ &\Updownarrow \\ \liminf_n \left\{ |\bar{Z}_n - Y \mathbb{E} X| \leq \varepsilon \right\} &\supset \liminf_n \left\{ |\bar{X}_n - \mathbb{E} X_n| \leq \varepsilon \right\} \\ &\Downarrow \\ \mathbb{P} \left(\left\{ |\bar{Z}_n - Y \mathbb{E} X| \leq \varepsilon \right\} \right) &\geq \mathbb{P} \left(\left\{ |\bar{X}_n - \mathbb{E} X_n| \leq \varepsilon \right\} \right) \\ &\quad \xrightarrow{\text{--> 1 by SLLN: } \bar{X}_n \rightarrow \mathbb{E} X \text{ a.s.}} \\ &\Downarrow \\ \mathbb{P} \left(\liminf_n \left\{ |\bar{Z}_n - Y \mathbb{E} X| \leq \varepsilon \right\} \right) &= 1 \end{aligned}$$

i.e.

$$\bar{Z}_n \xrightarrow{\text{a.s.}} Y \mathbb{E} X.$$

Remark

There is a shorter solution: we know that $X_n \rightarrow X$ a.s. and $Y_n \rightarrow Y$ a.s. and this implies that

$$X_n Y_n \xrightarrow{\text{a.s.}} XY$$

in the special case where $Y_n \equiv Y$.

2. We know that

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(Y \mathbb{E} X = \mu) \\ &= \mathbb{P}(Y \mathbb{E} X = \mathbb{E} X) \\ &= \mathbb{P}(Y = 1) = p \in (0, 1) \end{aligned}$$

since $\mathbb{E} X \neq 0$.

Exercise 2

Monte Carlo integration. Let $f: [0, 1] \rightarrow \mathbb{R}$ be a Borel-measurable function with $f \in L^1([0, 1])$. Let $(U_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables with $U_i \sim U[0, 1]$. Let

$$I_n = \frac{1}{n} (f(U_1) + \dots + f(U_n)).$$

Determine the convergence a.s. of I_n .

Since we know that $U_i \perp \!\!\! \perp$ and that f is Borel-measurable, then $f(U_i)$ are random variables and independent ones. Note that, since U_i are absolutely continuous random variables,

$$\begin{aligned} \mathbb{E}|f(U_i)| &= \int_0^1 |f(u)| \cdot f_{U_i}(u) du \\ &= \int_0^1 |f(u)| \cdot \underset{\substack{\square \\ \text{p.d.f. of } U}}{1} du < \infty \end{aligned}$$

since $f \in L^1$. Setting $Y_i = f(U_i)$ we have $I_n = \bar{Y}_n$. By SLLN we have

$$\bar{Y}_n \rightarrow \mathbb{E}Y = \int_0^1 f(u) du.$$

Exercise 3

Let (X_n) be a sequence of i.i.d. random variables with $X_1 \sim U(0, 1)$. Let $K_n = \min(X_1, \dots, X_n)$ and $Z_n = n \cdot K_n$. Determine the convergence in distribution of Z_n .

Let $x \in (0, n)$. We have

$$\begin{aligned}\mathbb{P}(Z_n > x) &= \mathbb{P}(nX_1 > x, nX_2 > x, \dots, nX_n > x) \\ &= \mathbb{P}\left(X_1 > \frac{x}{n}\right) \cdot \dots \cdot \mathbb{P}\left(X_n > \frac{x}{n}\right) \\ &= \left(1 - \frac{x}{n}\right) \cdot \dots \cdot \left(1 - \frac{x}{n}\right) \\ &= \left(1 - \frac{x}{n}\right)^n.\end{aligned}$$

So

$$\mathbb{P}(Z_n \leq x) = 1 - \left(1 - \frac{x}{n}\right)^n$$

and for $x \geq n$ we have $\mathbb{P}(Z_n \geq x) = 1$, for $x \leq 0$ we have $\mathbb{P}(Z_n \geq x) = 0$.

Revise with Kotatsu!

Remember that

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

Thus

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = \begin{cases} 1 - e^{-x} & \text{for } 0 < x < \infty \\ 0 & \text{for } x \geq 0. \end{cases}$$

It follows that

$$Z_n \xrightarrow{d} Z$$

where $Z \sim \text{Exp}(1)$.

Exercise 4

Let (X_n) be a sequence of independent random variables such that $X_n \sim U(0, 2^n)$. Let $Y_n = (X_n)^{\frac{1}{n}}$. Determine the convergence of Y_n in distribution, in probability and a.s.

"We want to do."

ⓘ **A.s. convergence:** if this holds, we also have convergence in probability and in distribution. The distribution of Y_n is

$$\begin{aligned}\mathbb{P}(Y_n \leq y) &= \mathbb{P}\left(X_n^{\frac{1}{n}} \leq y\right) = \begin{cases} 0 & \text{if } y < 0 \\ \mathbb{P}(X_n \leq y^n) & \text{if } y \geq 0. \end{cases} \\ &= \begin{cases} 0 & \text{if } y < 0 \\ y^n \cdot 2^{-n} & \text{if } 0 \leq y^n \leq 2^n \iff 0 \leq y \leq 2 \\ 1 & \text{if } y^n > 2^n \iff y > 2. \end{cases}\end{aligned}$$

Revise with Kotatsu!

Do you really need to remember that for uniform distributions the probability density function

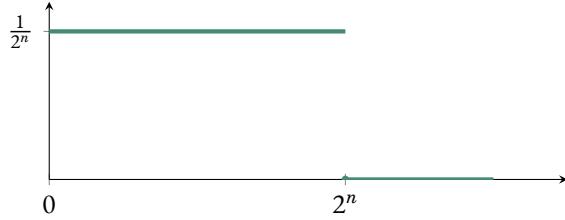


Figure 4.6: Remember how a $U(0, 2^n)$ uniform distribution is made?

is

$$f(x) = \frac{1}{a-b}?$$

Since $y^n 2^{-n} \rightarrow 0$ when $y < 2$ and $y^n 2^{-n} \rightarrow 1$ when $y = 2$ we have:

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n \leq y) = \begin{cases} 0 & \text{if } y < 2 \\ 1 & \text{if } y \geq 2. \end{cases}$$

This means that we have convergence in distribution to the degenerate random variable $Y = 2$. Let's check a.s. convergence with BC: for $\varepsilon > 0$ and $\varepsilon < 2$ (because for $\varepsilon \geq 2$ the probability is simply 0) we have

$$\begin{aligned} \mathbb{P}(|Y_n - 2| > \varepsilon) &= \mathbb{P}(-Y_n + 2 > \varepsilon) = \mathbb{P}(Y_n < 2 - \varepsilon) \\ &\quad \text{since } Y_n = X_n^{\frac{1}{n}} \text{ and } Y_n \in [0, 2^n] \\ &= \left(\frac{2 - \varepsilon}{2}\right)^n \end{aligned}$$

hence, since we know that $0 > \varepsilon < 2$ and therefore $\frac{2-\varepsilon}{2} < 1$

$$\sum_{i=1}^{\infty} \mathbb{P}(|Y_i - 2| > \varepsilon) = \sum_{i=1}^{\infty} \left(\frac{2 - \varepsilon}{2}\right)^n < +\infty.$$

Thus, by BC-1 we have

$$\mathbb{P}(\limsup_n \{|Y_n - 2| > \varepsilon\}) = 0 \iff \mathbb{P}(\liminf_n \{|Y_n - 2| \geq \varepsilon\}) = 1$$

which implies a.s. convergence of Y_n to 2, since $\varepsilon > 0$ was arbitrary.

Exercise 5

Let $(X_n)_n$ be a sequence of random variables independent and such that

$$\mathbb{P}(X_1 = x) = \begin{cases} \frac{1}{2} & \text{if } x \in (-1, 1) \\ 0 & \text{else} \end{cases}$$

and

$$\mathbb{P}(X_n = x) = \begin{cases} \frac{1}{2} & \text{if } x \in (-n, n) \\ \frac{1}{2} \left(1 - \frac{1}{n}\right) & \text{if } x \in (-1, 1) \\ 0 & \text{else.} \end{cases}$$

i.e.

$$X_n = \begin{cases} \frac{1}{n} & \text{if } x \in (-1, 1) \\ 0 & \text{else.} \end{cases}$$

Determine the convergence in distribution of $(Y_n)_n$.

Let's check whether we can apply CLT (ez version). We know that the X_n are independent with $\mathbb{E}X_n = 0$ and

$$\text{Var } X_n = \mathbb{E}X_n^2 = 2 \cdot n^2 \cdot \frac{1}{2n^2} + 2 \cdot 1 \cdot \frac{1}{2} \left(1 - \frac{1}{n^2}\right) = 1 + \left(1 - \frac{1}{n^2}\right)$$

so $\text{Var } X_n$ depends on n and we can't apply CLT! Let's proceed differently. Consider

$$\phi(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

and notice that

$$Y_n = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^n \phi(X_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \phi(X_i)). \quad (\ddot{\lambda})$$

We study the convergence of the two parts separately. Note that

$$\begin{aligned} \mathbb{P}(X_i \neq \phi(X_i)) &= \mathbb{P}(X_i \neq \pm 1) \\ &= \mathbb{P}(X_i = i \cup X_i = -i) \quad \text{for } i > 1 \\ &= \mathbb{P}(X_i = i) + \mathbb{P}(X_i = -i) \\ &= \frac{1}{2i^2} + \frac{1}{2i^2} = \frac{1}{i^2}. \end{aligned}$$

Hence we have that

$$\sum_{i=1}^{\infty} \mathbb{P}(X_i \neq \phi(X_i)) = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

and by BC-1 we can take the complement $\mathbb{P}(\liminf_i \{X_i = \phi(X_i)\}) = 1$ and thus $\mathbb{P}\text{-a.a.}\omega, \exists N(\omega)$ such that $\forall i \geq N(\omega), X_i(\Omega) = \phi(X_i(\omega))$ which implies that for a set Ω' of measure 1 we have $\sum_{i=1}^{\infty} X_i - \phi(X_i) < \infty$. Then the second term in $\ddot{\lambda}$ converges to 0 \mathbb{P} - a.s. (and in distribution) because of the factor $\frac{1}{\sqrt{n}}$ in front of the finite series. For the other terms in $\ddot{\lambda}$ we have

$$\begin{aligned} \phi(X_i) &\in \{1, -1\} \\ \mathbb{P}(\phi(X_i) = 1) &= \mathbb{P}(X_i \geq 0) = \frac{1}{2} \quad \text{i.e. } \phi(X_i) \sim U(\{-1, 1\}) \\ \mathbb{P}(\phi(X_i) = -1) &= \mathbb{P}(X_i < 0) = \frac{1}{2} \end{aligned}$$

so $\phi(X_i)$ are identically distributed random variables with mean 0 and variance

$$\text{Var } \phi(X_i) = \mathbb{E}\phi(X_i)^2 = 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1.$$

They are independent because X_i are independent and thus, by CLT (ez version) we have

$$\underbrace{\frac{\frac{1}{n} \sum_{i=1}^n \phi(X_i) - 0}{\frac{1}{\sqrt{n}}}}_{\xrightarrow{d} N(0,1)} = \frac{\sqrt{n}}{n} \sum_{i=1}^n \phi(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(X_i)$$

hence $Y \xrightarrow{d} N(0, 1) + 0 = N(0, 1)$ because of the following fact:

Remark

if $X_n \xrightarrow{d} X, Z_n \xrightarrow{d} c, c \in \mathbb{R}$, then $(X_n, Z_n) \xrightarrow{d} (X, c)$.

From this fact we know that $X_n + Z_n \xrightarrow{d} X + c$ and we apply this with

$$X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(X_i) \xrightarrow{d} N(0, 1) \quad \text{and} \quad Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \phi(X_i)) \xrightarrow{d} 0.$$

Exercise 6

Prove that if $X_n \xrightarrow{P} X$ then there is a subsequence $(X_{n_k})_k$ s.t. that $X_{n_k} \xrightarrow{s} X$ (\xrightarrow{s} converges along the subsequence).

Let $k \in \mathbb{N}$ and select n_k such that $n_k > n_{k-1}$ and

$$\mathbb{P} \left(|X_{n_k} - X| > \frac{1}{k} \right) \leq \frac{1}{k^2}$$

which is always possible since $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0 \forall \varepsilon > 0$ thanks to the fact that we know about the convergence. Notice, moreover, that $n_k \rightarrow \infty$ as $k \rightarrow \infty$. We have

$$\sum_{k=1}^{\infty} \mathbb{P}\left(|X_{n_k} - X| > \frac{1}{k}\right) \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

so by BC-1 we have

$$\mathbb{P}\left(\limsup_k \left\{|X_{n_k} - X| > \frac{1}{k}\right\}\right) = 0 \iff \mathbb{P}\left(\liminf_k \left\{|X_{n_k} - X| \leq \frac{1}{k}\right\}\right) = 1$$

which means that $\forall \omega \in \{\liminf_k |X_{n_k} - X| \leq \frac{1}{k}\} = \Omega'$ (which is the set of the ω where the event $|X_{n_k} - X| \leq \frac{1}{k}$ is true as $k \rightarrow \infty$) we have

$$|X_{n_k}(\omega) - X(\omega)| \leq \frac{1}{k}$$

i.e.

$$\lim_{k \rightarrow \infty} |X_{n_k}(\omega) - X(\omega)| \leq 0 \implies \lim_{k \rightarrow \infty} |X_{n_k} - X(\omega)| = 0$$

and since $\mathbb{P}(\Omega') = 1$ we have a.s. convergence along the subsequence n_k .

Exercise 7

Let $(X_i)_i$ be a sequence of i.i.d. random variables (real-valued).

(i) Show that for any $f: \mathbb{R} \rightarrow \mathbb{R}^+$ Borel-measurable and bounded we have

$$\frac{1}{n} \sum_{i=1}^n f(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[f(X_1)].$$

(ii) For $A \in \mathcal{B}(\mathbb{R})$ and $\omega \in \Omega$ let $F_n(A)$ be the random variable

$$F_n(A) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_A(X_i).$$

Given $A \in \mathcal{B}(\mathbb{R})$ determine whether $F_n(A)$ converges a.s.

- (i) Let $Y_i := f(X_i)$. $(Y_i)_i$ are random variables because f is Borel-measurable and $Y_i \perp\!\!\!\perp$ because $X_i \perp\!\!\!\perp$. Moreover Y_i are identically distributed because

$$\begin{aligned} \mathbb{P}(Y_i \leq y) &= \mathbb{P}(f(X_i) \leq y) \\ &= \mathbb{P}(f(X_i) \in (-\infty, y]) \\ &= \mathbb{P}(X_i \in f^{-1}(-\infty, y]) \\ &= \mathbb{P}(X_1 \in f^{-1}(-\infty, y]) \\ &= \mathbb{P}(f(X_1) \in (-\infty, y]) \\ &= \mathbb{P}(Y_1 \leq y) \quad \forall i \geq 1 \end{aligned}$$

Moreover, since $\text{Var } Y_i \leq c \|f\|_\infty =: \sigma^2$ (and it is independent of i) then

$$\sum_{i=1}^k \text{Var}\left(\frac{Y_i}{i}\right) \leq \sigma^2 \cdot \sum_{i=1}^n \frac{1}{i^2} < \infty$$

So by the second point of the WLLN we have

$$\begin{aligned} \frac{\sum_{i=1}^n Y_i - \mathbb{E}(\sum_{i=1}^n Y_i)}{n} &\xrightarrow{\text{a.s.}} 0 \\ \downarrow \\ \frac{1}{n} \sum_{i=1}^n f(X_i) - \frac{1}{n} \mathbb{E}[n \cdot Y_i] &= \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X_1)] \\ \Downarrow \\ \frac{1}{n} \sum_{i=1}^n f(X_i) &\xrightarrow{\text{a.s.}} \mathbb{E}[f(X_1)]. \end{aligned}$$

(ii) By (i) with $f = \mathbb{1}_A$ we have

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_A(X_i) \xrightarrow{\text{a.s.}} \mathbb{E}[\mathbb{1}_A(X_1)] = \mathbb{P}(X_1 \in A).$$

Remark

Notice that (i) and (ii) provide a Monte-Carlo method to estimate probabilities and expectations.

4.1.6 Exercise class 6

Revise with Kotatsu!

Definition 4.1.4

Conditional expectation: let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space and $\mathcal{F} \subset \mathcal{H}$. Let X be \mathcal{H} -measurable with $\mathbb{E}|X| < \infty$. The conditional expectation of X given \mathcal{F} , denoted by \bar{X} or $\mathbb{E}(X|\mathcal{F})$ is a random variable such that:

- \bar{X} is \mathcal{F} -measurable (**measurability**);
- $\mathbb{E}[\mathbb{1}_F \bar{X}] = \mathbb{E}[\mathbb{1}_F X]$, $\forall F \in \mathcal{F}$ (**projection**).

This actually defines the conditional expectation of positive random variable first and then uses $X = X^+ - X^-$.

Here are the **main properties of conditional expectation**. Let $X, Y \in L^1(\Omega, \mathcal{H}, \mathbb{P})$, $\mathcal{G} \subset \mathcal{H}$, $\mathcal{F} \subset \mathcal{H}$. Then:

- (a) if Y is a version of $\mathbb{E}(X|\mathcal{G})$ then $\mathbb{E}Y = \mathbb{E}X$;
 - (b) if X is \mathcal{G} -measurable then $\mathbb{E}(X|\mathcal{G}) = X$, \mathbb{P} – a.s.;
 - (c) $\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$, \mathbb{P} – a.s.;
 - (d) if $X \geq 0$ then $\mathbb{E}[X|\mathcal{G}] \geq 0$
- (e) **tower property:** if $\mathcal{F} \subset \mathcal{G}$ then

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{F}] = \mathbb{E}[\mathbb{E}(X|\mathcal{F})|\mathcal{G}] = \mathbb{E}(X|\mathcal{F});$$

- (f) if Y is bounded and \mathcal{G} -measurable then

$$\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G});$$

- (g) if \mathcal{F} is independent of $\sigma(\mathcal{G}, \sigma(X))$ then

$$\mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{F})) = \mathbb{E}(X|\mathcal{G}).$$

Remark

Special case with $\mathcal{G} = \{\emptyset, \Omega\}$: if $X \perp \mathcal{F}$ then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}X$.

Conditioning on another random variable:

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y)).$$

Remark

The expectation $\mathbb{E}(X)$ can be viewed as a condition expectation given the trivial σ -algebra $\mathcal{H}_0 = \{\emptyset, \Omega\}$ i.e.

$$\mathbb{E}(X) = \mathbb{E}(X|\mathcal{H}_0).$$

This is useful when applying the tower property, because it gives us

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|\mathcal{G})).$$

Theorem 4.1.2

Suppose X and Y are random variables on $(\Omega, \mathcal{H}, \mathbb{P})$ with values in (D, \mathcal{D}) , (E, \mathcal{E}) respectively. Suppose that the joint probability distribution has the form

$$\pi(dx, dy) = \mu(dx)k(x, dy)$$

with some probability kernel form (D, \mathcal{D}) to (E, \mathcal{E}) , i.e. for every f which is $(\mathcal{D} \times \mathcal{E})$ -measurable we have

$$\int_{D \times E} f(x, y) \pi(dx, dy) = \int_D \left(\int_E f(x, y) k(x, dy) \right) \mu(dx).$$

Then the kernel L defined by

$$L_\omega(B) = k(X(\omega), B), \quad \forall \omega \in \Omega, B \in \mathcal{E}$$

is a version of the conditional distribution of Y given X and for every positive f which is $(\mathcal{D} \times \mathcal{E})$ -measurable it holds

$$\mathbb{E}[f(X, Y)|X] = \int_E f(X, y) k(X, dy).$$

This last equation is also known as **freezing lemma**. It means that I can effectively calculate the conditional expectation given X as if X was a constant (and not a random variable). Cool! Was it so hard to put it in this way?

Exercise 1

I.e. $(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow ([0, 1], \mathcal{B}([0, 1]), \lambda_{[0,1]}).$ I.e.

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in [0, \frac{1}{3}] \\ 2 & \text{if } \omega \in [\frac{1}{3}, \frac{2}{3}] \\ 3 & \text{if } \omega \in [\frac{2}{3}, 1]. \end{cases}$$

I.e. $\bar{Y}(\omega) = 2\omega^2$. Determine a version of $\mathbb{E}(\bar{Y}|X)$.

Note that $\sigma(X)$ is generated by three disjoint intervals:

$$I_1 := [0, \frac{1}{3}], \quad I_2 := [\frac{1}{3}, \frac{2}{3}], \quad I_3 := [\frac{2}{3}, 1].$$

We know that $\bar{Y} := \mathbb{E}(Y|X)$ must be $\sigma(X)$ -measurable (by property of measurability) so it must be constant on each interval I_i (since the σ -algebras of constants are constants so any function measurable by a constant must be a constant):

$$(\forall A \in \mathcal{B}(\mathbb{R}), \quad \bar{Y}^{-1}(A) \in \sigma(X) = \sigma(I_1, I_2, I_3)).$$

Let's denote by a_i the values of \bar{Y} on I_i , that is

$$\bar{Y}(\omega) = \mathbb{E}(Y|X) = \begin{cases} a_1 & \text{on } I_1 \\ a_2 & \text{on } I_2 \\ a_3 & \text{on } I_3. \end{cases}$$

It must be that, for the projection property,

$$\mathbb{E}(\mathbb{1}_{I_i} \bar{Y}) = \mathbb{E}(\mathbb{1}_{I_i} Y) \quad \forall I_i \in \sigma(X),$$

that is

$$\int_{I_i} \bar{Y}(\omega) d\mathbb{P}(\omega) = \int_{I_i} Y(\omega) d\mathbb{P}(\omega), \quad \forall i = 1, 2, 3. \quad (\textcircled{2})$$

The left hand side of $\textcircled{3}$ gives

$$\int_{I_i} \bar{Y}(\omega) d\mathbb{P}(\omega) = \int_{I_i} a_i d\mathbb{P} = a_i \mathbb{P}(I_i) = a_i \cdot \frac{1}{3}.$$

The right hand side of $\textcircled{3}$ gives

$$\begin{aligned} \int_{I_i} Y(\omega) d\mathbb{P}(\omega) &= \int_{I_i} 2\omega^2 d\mathbb{P}(\omega) \\ &= \begin{cases} \frac{2}{3}\omega^3 \Big|_0^{\frac{1}{3}} = \frac{2}{81} & \text{if } i = 1 \\ \frac{2}{3}\omega^3 \Big|_{\frac{1}{3}}^{\frac{2}{3}} = \frac{14}{81} & \text{if } i = 2 \\ \frac{2}{3}\omega^3 \Big|_{\frac{2}{3}}^1 = \frac{38}{81} & \text{if } i = 3. \end{cases} \end{aligned}$$

Thus we must have

$$\begin{aligned} \frac{1}{3}a_1 &= \frac{2}{81}, & \frac{1}{3}a_2 &= \frac{14}{81}, & \frac{1}{3}a_3 &= \frac{38}{81} \\ &\Downarrow \\ a_1 &= \frac{2}{27}, & a_2 &= \frac{14}{27}, & a_3 &= \frac{38}{27} \end{aligned}$$

Exercise 2

Special case of freezing lemma. Let X, Y be discrete random variables with values in E and F respectively and let $\rho_{X,Y}(x,y)$ denote their joint discrete density. Let $g : E \rightarrow \mathbb{R}$ be such that $g(X) \in L^1$. Show that

$$\mathbb{E}[g(X)|Y] = \varphi(Y)$$

where

$$\varphi(y) = \sum_x g(x) \rho_{X|Y}(x|y)$$

and

$$\rho_{X|Y}(x|y) = \begin{cases} \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)} & \text{if } \rho_Y(y) > 0 \\ 0 & \text{else.} \end{cases}$$

According to the definition we must show:

[i] $\sigma(Y)$ -measurability ($g(X) \in L^1$ by assumption);

[ii] projection.

Let's tackle this one at the time.

[i] For this property it is enough to show that $y \mapsto \varphi(y)$ is measurable. We have $\varphi(y) = \sum_x g(x) \rho_{X|Y}(x|y)$ where $\rho_{X|Y}(x|y)$ depends on $\rho_{X,Y}(x,y)$ and $\rho_Y(y)$ which are both measurable because X and Y are random variables.

[ii] $\forall A \in \sigma(Y)$ we should check whether $\mathbb{E}[\mathbb{1}_A \varphi(Y)] = \mathbb{E}[\mathbb{1}_A g(X)]$. The set $A \in \sigma(Y)$ must be of the form $A = \{\omega : Y(\omega) \in B\}$ for some $B \in \mathcal{F}$. Thus $\mathbb{1}_A(\omega) = \mathbb{1}_B(Y)$. We have

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A \varphi(Y)] &= \int_{\Omega} \mathbb{1}_A(\omega) \varphi(Y(\omega)) \mathbb{P}(d\omega) \\ &= \int_{\Omega} \mathbb{1}_B(Y(\omega)) \mathbb{P}(d\omega) \\ &= \sum_{y \in F} \mathbb{1}_B(y) \varphi(y) \rho_Y(y)^{\textcircled{7}} \\ &= \sum_{y \in B} \sum_{x \in E} g(x) \rho_{X|Y}(x|y) \rho_Y(y) \\ &= \sum_{y \in B} \sum_{x \in E} g(x) \rho_{X,Y}(x,y). \end{aligned} \tag{D}$$

On the other hand,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A g(Y)] &= \int_{\Omega} \mathbb{1}_A(\omega) g(X(\omega)) \mathbb{P}(d\omega) \\ &= \int_{\Omega} \mathbb{1}_B(Y(\omega)) g(X(\omega)) \mathbb{P}(d\omega) \\ &= \sum_{(x,y) \in E \times F} \mathbb{1}_B(y) g(x) \rho_{X,Y}(x,y) \\ &= \sum_{X \in E} \sum_{y \in B} g(x) \rho_{X,Y}(x,y)\end{aligned}$$

which is the same as in D .

Exercise 3

Let $X \sim \text{Geom}(1-p)$, and $Y \sim \text{Geom}(1-q)$ be independent and let $Z := \min(X, Y)$. Determine a version of the conditional expectation $\mathbb{E}(X|Z)$.

Here we mean that $\mathbb{P}(X = x) = (1-p)p^x$, $x \in \mathbb{N} \cup \{0\}$. This is also known as “modified Geometric”. In this case $\mathbb{E}X = \frac{1}{1-p} - 1$. We want to use the result from exercise 2 and hence we compute the joint distribution of (X, Z) :

$$\begin{aligned}\rho_{X,Z}(x, z) &= \mathbb{P}(X = x, Z = z) \\ &= \mathbb{P}(X = x, \min(X, Y) = z) \\ &= \begin{cases} 0 & \text{if } z > x \\ \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geq x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\ &= \begin{cases} \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geq x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\ &= \begin{cases} \mathbb{P}(X = x, Y = z) & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ \mathbb{P}(X = x, Y \geq x) & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\ &= \begin{cases} (1-p)p^x(1-q)^z & \text{if } z < x, \quad z, x \in \mathbb{N} \cup \{0\} \\ (1-p)p^x \sum_{k=1}^{\infty} (1-q)^k q^k & \text{if } z = x, \quad z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else.} \end{cases} \quad (\text{f})\end{aligned}$$

Revise with Kotatsu!

Remember the geometric series:

$$\begin{aligned}\sum_{k=0}^{\infty} q^k &= \frac{1}{1-q} \\ \sum_{k=0}^N q^k &= \frac{1 - q^{N+1}}{1-q}.\end{aligned}$$

⁷In the sum only the terms with $\rho_Y(y) > 0$ give a contribution.

In this case we have

$$\begin{aligned}
\mathbb{P}(Y \geq x) &= \sum_{k=x}^{\infty} q^k \\
&= - \sum_{k=0}^{x-1} (1-q)q^k + \sum_{k=0}^{\infty} (1-q)q^k \\
&= -(1-q) \cdot \frac{1-q^x}{1-q} + (1-q) \frac{1}{1-q} \\
&= 1 - (1-q^x) = q^x.
\end{aligned} \tag{8}$$

We compute now the marginal distribution of Z . Let $z \geq 0, z \in \mathbb{N}$.

$$\begin{aligned}
\rho_Z(z) &= \mathbb{P}(Z = z) \\
&= \mathbb{P}(\min(X, Y) = z) \\
&= \mathbb{P}(\min(X, Y) \geq z) - \mathbb{P}(\min(X, Y) \geq z+1) \\
&= \mathbb{P}(X \geq z)\mathbb{P}(Y \geq z) - \mathbb{P}(X \geq z+1)\mathbb{P}(Y \geq z+1) \\
&= p^z q^z - p^{z+1} q^{z+1} \quad \text{due to } \textcolor{violet}{\cancel{8}} \\
&= p^z q^z (1-pq)
\end{aligned} \tag{9}$$

hence $Z \sim \text{Geom}(1-pq)$. We can thus define the conditional probability $\rho_{X|Z}$ as

$$\begin{aligned}
\rho_{X|Z} &= \begin{cases} \frac{\rho_{X,Z}(x,z)}{\rho_Z(z)} & \text{if } \rho_Z(z) > 0 \\ 0 & \text{if } \rho_Z(z) = 0 \end{cases} \\
\text{due to } \textcolor{violet}{\cancel{8}} \text{ and } \textcolor{violet}{\cancel{9}} &= \begin{cases} \frac{(1-p)p^x(1-q)q^z}{p^z q^z (1-pq)} & \text{if } z < x, z, x \in \mathbb{N} \cup \{0\} \\ \frac{(1-p)p^x q^z}{p^z q^z (1-pq)} & \text{if } z = x, z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \frac{(1-p)(1-q)p^{x-z}}{1-pq} & \text{if } z < x, z, x \in \mathbb{N} \cup \{0\} \\ \frac{(1-p)}{1-pq} & \text{if } z = x, z, x \in \mathbb{N} \cup \{0\} \\ 0 & \text{else.} \end{cases} \tag{10}
\end{aligned}$$

Using $\textcolor{violet}{\cancel{10}}$ and the result of exercise 2 with $g(x) = x$ we define

$$\begin{aligned}
\varphi(z) &= \sum_{x=0}^{\infty} x \rho_{X|Z}(x|z) \\
&= \sum_{x=0}^{\infty} (x - z + z) \rho_{X|Z}(x|z) \\
&= z \sum_{x=0}^{\infty} \rho_{X|Z}(x|z) + \sum_{x=0}^{\infty} (x - z) \rho_{X|Z}(x|z) \\
&= z + \sum_{x=z}^{\infty} (x - z) \rho_{X|Z}(x|z) \\
&= z + 0 + \sum_{x=z+1}^{\infty} (x - z) \cdot \frac{(1-p)(1-q)p^{x-z}}{1-pq} \\
&= z + \frac{(1-p)(1-q)}{1-pq} \sum_{x=z+1}^{\infty} \underbrace{(x - z)}_{=y} p^{x-z} \\
&= z + \frac{(1-p)(1-q)}{1-pq} \sum_{y=1}^{\infty} y p^y \\
&= z + \frac{1-q}{1-pq} \sum_{y=1}^{\infty} y p^y (1-p) \\
&\quad \text{EV where } V \sim \text{Geom}(1-p) \text{ so } \mathbb{E}V = \frac{1}{1-p} - 1 \\
&= z + \frac{1-q}{1-pq} \cdot \frac{p}{1-p}.
\end{aligned}$$

So a version of the conditional expectation $\mathbb{E}(X|Z)$ is

$$\mathbb{E}(X|Z) = Z + \frac{(1-q)p}{(1-pq)(1-p)}.$$

Exercise 4

Let $(Z_n)_{n \geq 1}$ be independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite mean. Let $X_0 := a \in \mathbb{R}$ and $Z_n = a + Z_1 + Z_2 + \dots + Z_n, \forall n \geq 1$. Find a version of $\mathbb{E}(X_n|\mathcal{F}_n), \forall n \geq 1$, where $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$.

Idk this had no solution.

Exercise 5

Let $(Y_n)_{n \geq 1}$ be a sequence of positive i.i.d. random variables in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Let N be a random variable with values in \mathbb{N} and belonging to $L^1(\Omega, \mathcal{F}, \mathbb{P})$, $N \wedge (X_n)_n$. Let

$$Y(\omega) = Y_1(\omega) - Z_1(\omega) + \dots + (-1)^{N(\omega)} Y_{N(\omega)}(\omega)$$

Find a version of $\mathbb{E}(Y|N)$.

Take $A \in \sigma(N)$ arbitrary. Note that the events $\{N = n\}_{n \in \mathbb{N}}$ form a partition of Ω so

$$Y = \sum_{n=1}^{\infty} \mathbb{1}_{\{N=n\}} = \sum_{n=1}^{\infty} \sum_{k=1}^n X_k \mathbb{1}_{\{N=n\}}.$$

Hence

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A Y] &= \mathbb{E}\left[\mathbb{1}_A \sum_{n=1}^{\infty} \sum_{k=1}^n X_k \mathbb{1}_{\{N=n\}}\right] \\ \text{monotone convergence} &= \sum_{n=1}^{\infty} \mathbb{E}\left[\mathbb{1}_A \sum_{k=1}^n X_k \mathbb{1}_{\{N=n\}}\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}\left[\sum_{k=1}^n X_k\right] \cdot \mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{N=n\}}] \\ &= \sum_{n=1}^{\infty} n \cdot \mathbb{E}X_1 \cdot \mathbb{E}[\mathbb{1}_A \mathbb{1}_{\{N=n\}}] \\ &= \mathbb{E}X_1 \cdot \mathbb{E}\left[\sum_{n=1}^{\infty} n \cdot \mathbb{1}_{\{N=n\}} \cdot \mathbb{1}_A\right] \\ &= \mathbb{E}X_1 \cdot \mathbb{E}[N \mathbb{1}_A] \\ &= \mathbb{E}[\mathbb{E}[X_1] \cdot N \mathbb{1}_A]. \end{aligned} \quad (\blacksquare)$$

Hence, since $\mathbb{E}(X_1) \cdot N$ is $\mathcal{F}(N)$ -measurable, we have that the second point of definition and expectation is satisfied for

$$\mathbb{E}(Y|N) = \mathbb{E}(X_1) \cdot N$$

which is then a version of the conditional expectation. Notice that we should also check that $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ which is true by similar computations as in with $A = \Omega$ so

$$\mathbb{E}[|Y|] = \mathbb{E}(Y) = \mathbb{E}(\mathbb{E}X_1 \cdot N) = \mathbb{E}X_1 \cdot \mathbb{E}N < \infty$$

since $X_1, N \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Exercise 6

Let $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ be such that $\mathbb{E}(Y|X) = \mathbb{E}X$ a.s. and \mathbb{P} . Show that $\text{Cov}(X, Y) = 0$ and that the converse is not true.

We know that

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY|X)) = \mathbb{E}(X\mathbb{E}(Y|X)) = \underbrace{\mathbb{E}(X\mathbb{E}Y)}_{\text{by assumption}} = \mathbb{E}X\mathbb{E}Y.$$

Thus $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = 0$. Assume now $\text{Cov}(X, Y) = 0$. To show that $\mathbb{E}(Y|X) \neq \mathbb{E}Y$ in general, we find a counterexample. Consider

$$U \sim U(\{(0,1), (-1,-1), (1,-1)\}) \quad U = (X, Y).$$

Then $X \sim U(\{0, -1, 1\})$ which means $\mathbb{E}X = 0$ and

$$XY = \begin{cases} 0 & \text{with prob. } \frac{1}{3} \\ 1 & \text{with prob. } \frac{1}{3} \\ -1 & \text{with prob. } \frac{1}{3}. \end{cases}$$

Hence $XY \sim U(\{0, -1, 1\})$ which means $\mathbb{E}(XY) = 0$. We then have $\text{Cov}(XY) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0 - 0 = 0$. However, $\mathbb{E}(Y|X=0) = 1$ but

$$Y = \begin{cases} 1 & \text{with prob. } \frac{1}{3} \\ -1 & \text{with prob. } \frac{2}{3} \end{cases} \implies \mathbb{E}Y = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3} \neq 1.$$

Exercise 7

Let X and Y be real-valued independent random variables. Let F_{X+Y} be the distribution function of $X+Y$ and F_Y the distribution function of Y . Show that $F_Y(t-X)$ is a version of the conditional expectation

$$\mathbb{E}(\mathbb{1}_{\{X+Y \leq t\}}|X)$$

Deduce that if $X \sim U(0,1)$, then $\mathbb{E}_{X,Y}[Y|t] = \int_{-\infty}^t F_Y(t-u) du$.

We want to apply the freezing lemma:

Theorem 4.1.3

Let (X, Y) be \mathbb{R} -valued random variables such that their joint distribution is

$$\pi(dx, dy) = \mu(dx) k(x, dy). \quad \underbrace{k(x, dy)}_{\text{prob. kernel}}$$

Then \forall positive measurable f we have that $\int_{\mathbb{R}} f(X, Y)k(X, dy)$ is a version of conditional expectation $\mathbb{E}[f(X, Y)|X]$.

We want to apply this shit to (X, Y) and $f(X, Y) = \mathbb{1}_{\{X+Y \leq t\}}$ for some $t \in \mathbb{R}$ given. We have by independence $\pi(dx, dy) = \mu_X(dx)\mu_Y(dy)$ and $\mathbb{1}_{\{X+Y \leq t\}} \geq 0$ and measurable. Thus a version of the conditional expectation $\mathbb{E}(\mathbb{1}_{\{X+Y \leq t\}}|X)$ is given by

$$\int_{\mathbb{R}} \mathbb{1}_{\{X+Y \leq t\}}(y)\mu_Y(dy).$$

Let's work out the integral (with X replaced by x):

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{1}_{\{x+y \leq t\}}(y)\mu_Y(dy) &= \int_{\mathbb{R}} \mathbb{1}_{\{y \geq t-x\}}(y)\mu_Y(dy) \\ &= \int_{-\infty}^{t-x} \mu_Y(dy) \\ &= F_Y(t-x). \end{aligned}$$

Hence a version of the conditional expectation is given by $F_Y(t-x)$. In the special case when $X \sim U(0,1)$ we

get $\mu(dx) = dx$ on $(0,1)$. Then

$$\begin{aligned} F_{X+Y}(t) &= \mathbb{P}(X + Y \leq t) \\ &= \mathbb{E}[\mathbb{1}_{\{X+Y \leq t\}}] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{X+Y \leq t\}}|X]] \\ &= \mathbb{E}[F_Y(t - X)] \\ &= \int_{\mathbb{R}} F_Y(t - x) \mu_X(dx) \\ &= \int_{\mathbb{R}} F_Y(t - x) dx. \end{aligned}$$

4.1.7 Exercise class 7

Revise with Kotatsu!

Definition 4.1.5

Sub/Super Martingale: a real-valued stochastic process $(X_t)_{t \in T}$ is a \mathcal{F} -submartingale/supermartingale/martingale if:

1. X is \mathcal{F} -adapted;
2. X_t is integrable $\forall t \in T$, i.e. $\mathbb{E}|X_t| < \infty$;
3. $\mathbb{E}[X_t|\mathcal{F}_s] \geq / \leq / = X_s, \forall s \geq t, t, s \in T$.

For the discrete time processes we have $T \equiv \mathbb{N}$.

Definition 4.1.6

Stopping time: a random time $T : \Omega \rightarrow \mathbb{R}^+$ is a stopping time for a filtration $(\mathcal{F}_t)_{t \in T}$ if $\{T \leq t\} \in \mathcal{F}_t \forall t \in T$.

In the discrete case we have $T : \Omega \rightarrow \mathbb{N} \cup \infty$ s.t. $\{T = n\} \in \mathcal{F}_n \forall n \geq 0$.

Theorem 4.1.4

Doob's stopping theorem. Let M be a \mathcal{F} -adapted process. The following are equivalent:

- M is a \mathcal{F} -submartingale;
- $\forall T, S$ stopping times such that $S \geq T$ and S, T bounded, M_S and M_T are integrable and $\mathbb{E}(M_T - M_S|\mathcal{F}_S) \geq 0$;
- $\forall T, S$ stopping times such that $S \geq T$ and S, T bounded, M_S and M_T are integrable and $\mathbb{E}(M_T - M_S) \geq 0$.

We can replace the submartingale with the **martingale** by switching \geq with $=$. Same for **supermartingale** (switching with \leq).

Theorem 4.1.5

Variation/extension of Doob's stopping theorem. Let M be a \mathcal{F} -martingale and T a stopping time. Then M_T is integrable and $\mathbb{E}M_T = \mathbb{E}M_0$ if and only if one of the following condition holds:

1. T is bounded;
2. M is bounded and $T < \infty \mathbb{P} - \text{a.s.}$;
3. $\mathbb{E}T < \infty$ and M has bounded increments;
4. M is uniformly integrable.

Exercise 1

Let $(\Omega, \mathcal{A}, \mathbb{P}) = ((0, 1), \mathcal{B}([0, 1]), \mathbb{P}_{[0, 1]})$. Let

$$C_n := \left[\left[\frac{k}{2^n}, \frac{k+1}{2^n} \right], k = 0, 1, \dots, 2^n - 1 \right]$$

and $\mathcal{G}_n = \sigma(C_n)$. Define

$$X_n(\omega) = \begin{cases} 2^{-n} & 0 < \omega \leq 2^{-n} \\ 0 & \text{else} \end{cases}$$

Determine whether $(\mathcal{G}_n)_n$ is a filtration and if so whether X is a \mathcal{G} -martingale.

Each \mathcal{G}_n is a σ -algebra by definition. We must prove that $\mathcal{F}_n \subseteq \mathcal{G}_{n+1}$. We have:

$$\begin{aligned} C_n &\ni \left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] = \left(\frac{2k}{2^{n+1}}, \frac{2(k+1)}{2^{n+1}} \right] \\ &= \underbrace{\left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}} \right]}_{\in C_{n+1}} \cup \underbrace{\left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}} \right]}_{\in C_{n+1}}. \end{aligned}$$

$\in \sigma(C_{n+1}) = \mathcal{G}_{n+1}$

Since it holds $\forall k = 0, \dots, 2^n - 1$ we have $C_n \subset \mathcal{G}_{n+1}$ as wanted. Let's check whether X is a \mathcal{G} -martingale.

- “Is X \mathcal{G} -adapted?” We need to check that $\forall B \in \mathcal{B}(\mathbb{R})$, $X_n^{-1}(B) \in \mathcal{G}_n$. This is like asking whether we have $\sigma(X_n) \subseteq \mathcal{G}_n$.

$$\begin{aligned} \sigma(X_n) &= \{X_n^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} \\ &\equiv \{\emptyset, \Omega, (0, 2^n], (2^n, 1]\} \subset \mathcal{G}_n. \end{aligned}$$



- “Is X integrable?” We know that

$$\begin{aligned} \mathbb{E}|X_n| &= \int_{\Omega} |X_n(\omega)| d\mathbb{P}(\omega) \\ &= \int_0^1 |X_n(\omega)| d\omega \\ &= \int_0^{2^{-n}} 2^n d\omega + \int_{2^{-n}}^1 0 d\omega = 1 < \infty. \end{aligned}$$



- “Does X have martingale property?” It is enough to show for n and $n+1$ that $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = X_n$. We have to show that

$$\mathbb{E}(X_n \mathbf{1}_A) = \mathbb{E}(X_{n+1} \mathbf{1}_A) \quad \forall A \in \mathcal{G}_n.$$

Since $\mathcal{G}_n = \sigma(C_n)$ we have that $A = \bigcup_k A_k$ where A_k have the form $A_k = \left(\frac{k}{2^n}, \frac{k+1}{2^n} \right]$ and are thus disjoint. Thus

$$\begin{aligned} \mathbb{E}(X_{n+1} \mathbf{1}_A) &= \mathbb{E}(X_{n+1} \mathbf{1}_{\bigcup_k A_k}) \\ &= \sum_k \mathbb{E}(X_{n+1} \mathbf{1}_{A_k}) \end{aligned}$$

and similarly

$$\mathbb{E}(X_n \mathbf{1}_A) = \sum_k \mathbb{E}(X_n \mathbf{1}_{A_k}).$$

We then just have to check that

$$\mathbb{E}(X_n \mathbf{1}_{A_k}) = \mathbb{E}(X_{n+1} \mathbf{1}_{A_k}) \quad \forall A_k \in C_n \tag{*}$$

(which means for $k = 0, 1, \dots, 2^n - 1$). Notice that if $k \geq 1$ then $\left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] \cap \left(0, \frac{1}{2^n} \right] = \emptyset$ and $\left(\frac{k}{2^n}, \frac{k+1}{2^n} \right] \cap \left(0, \frac{1}{2^{n+1}} \right] = \emptyset$. Thus $X_n \mathbf{1}_{A_k} = 0$ and $X_{n+1} \mathbf{1}_{A_k} = 0$, so $*$ holds.
If $k = 0$ then

$$X_n \mathbf{1}_{A_0} = 2^n \cdot \mathbf{1}_{(0, 2^{-n})} \cdot \mathbf{1}_{(0, 2^{-n})} = 2^n \cdot \mathbf{1}_{(0, 2^{-n})} = X_n$$

and

$$X_{n+1} \mathbb{1}_{A_0} = 2^{n+1} \mathbb{1}_{(0,2^{-n-1})} \mathbb{1}_{(0,2^{-n})} = 2^{n+1} \mathbb{1}_{(0,2^{-n-1})} = X_{n+1}.$$

Moreover, $\mathbb{E}X_n = \mathbb{E}N_{n+1} = 1$. Hence $\#$ holds also for $k = 0$. ✓

Exercise 2

Let X be a supermartingale, with $X_n \geq 0 \quad \forall n$. Show that

$$\mathbb{E}(X_{n+k} \mathbb{1}_{\{X_n=0\}}) = 0 \quad \text{a.s.}$$

By the supermartingale property we know that $\mathbb{E}(X_{n+k} | \mathcal{G}_n) \leq X_n$ so for any $A \in \mathcal{G}_n$ we have

$$\mathbb{E}(X_{n+k} \mathbb{1}_A) = \mathbb{E}[(\mathbb{E}X_{n+k} | \mathcal{G}_n) \mathbb{1}_A] \leq \mathbb{E}(X_n \mathbb{1}_A). \quad (\dagger\dagger)$$

We can use tower property because if a random variable is \mathcal{G} -adapted and positive then it is \mathcal{G} -measurable. Moreover, since $X_n \geq 0 \forall n$ and also $\mathbb{1}_A \geq 0$ we have

$$0 \leq \mathbb{E}(X_{n+k} \mathbb{1}_A). \quad (\dagger\dagger\dagger)$$

Putting $\dagger\dagger$ and $\dagger\dagger\dagger$ together we get

$$0 \leq \mathbb{E}(X_{n+k} \mathbb{1}_A) \leq \mathbb{E}(X_n \mathbb{1}_A)$$

and choosing $A = \{X_n = 0\} \in \mathcal{G}_n$ we get

$$0 \leq \mathbb{E}(X_{n+k} \mathbb{1}_{\{X_n=0\}}) \leq \mathbb{E}(X_n \mathbb{1}_{\{X_n=0\}}) = 0$$

this is always 0...

which implies $X_{n+k} \mathbb{1}_{\{X_n=0\}} = 0$ a.s. since $X_{n+k} \geq 0$.

Exercise 3

Let $(X_j)_n$ be a sequence of random variables in $L^2(\Omega)$ such that, $\forall n \geq 0$, $\mathbb{E}(X_{n+1} | \mathcal{G}_n) \leq 0$ a.s., where \mathcal{G} is the natural filtration generated by X_j . Let

$$S_n = \sum_{j=0}^n X_j \quad \text{and} \quad Z_n = S_n - \sum_{j=0}^n \mathbb{E}(X_{j+1}^2 | \mathcal{G}_j).$$

Check whether $(Z_n)_n$ is a \mathcal{G} -martingale.

We need to check the three properties.

1. Check that $Z_n \in m\mathcal{G}_n$ (= it is measurable with respect to \mathcal{G}_n): this is true because $X_j \in m\mathcal{G}_n \forall j \leq n$, thus $S_n \in m\mathcal{G}_n$ and also $S_n^2 \in m\mathcal{G}_n$. Finally $\mathbb{E}(X_{j+1}^2 | \mathcal{G}_j)$ is \mathcal{G}_j -measurable, so its sum up to $n-1$ is \mathcal{G}_{n-1} -measurable.
2. Check integrability:

$$\begin{aligned} \mathbb{E}|Z_n| &\leq \mathbb{E}|S_n|^2 + \sum_{j=0}^{n-1} \mathbb{E}\left(\left|\mathbb{E}(X_{j+1}^2 | \mathcal{G}_j)\right|\right) \\ &\leq \mathbb{E}\left(\left|\sum_{j=0}^n X_j\right|^2\right) + \sum_{j=0}^{n-1} \mathbb{E}(X_{j+1}^2) \\ &\leq n \sum_{j=0}^n \mathbb{E}|X_j|^2 + \sum_{j=0}^{n-1} \mathbb{E}|X_{j+1}|^2 < \infty \quad \text{since } X_j \in L^2(\Omega). \end{aligned}$$

3. Check martingale property:

$$\begin{aligned}
Z_{n+1} &= S_{n+1}^2 - \sum_{j=0}^n \mathbb{E}(X_{j+1}^2 | \mathcal{G}_n) \\
&= S_n^2 + 2X_{n+1}S_n + X_{n+1}^2 - \sum_{j=0}^{n-1} \mathbb{E}(X_{j+1}^2 | \mathcal{G}_n) - \mathbb{E}(X_{n+1}^2 | \mathcal{G}_n) \\
&= Z_n + 2X_{n+1}S_n + X_{n+1}^2 - \mathbb{E}(X_{n+1}^2 | \mathcal{G}_n).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}(Z_{n+1} | \mathcal{G}_n) &= \mathbb{E}(Z_n | \mathcal{G}_n) + 2\mathbb{E}(X_{n+1}S_n | \mathcal{G}_n) + \cancel{\mathbb{E}(X_{n+1}^2 | \mathcal{G}_n)} - \cancel{\mathbb{E}(\mathbb{E}(X_{n+1}^2 | \mathcal{G}_n) | \mathcal{G}_n)} \\
&= Z_n + 2S_n \underbrace{\mathbb{E}(X_{n+1} | \mathcal{G}_n)}_{=0 \text{ by assumption}} \\
&= Z_n.
\end{aligned}$$

Exercise 4

Let $(X_j)_{j \geq 1}$ be a sequence of i.i.d. random variables s.t. $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. Let $S_n = \sum_{j=1}^n X_j$, $T_1 \geq 1$ and $S_{T_1} \geq 0$. Let

$$\begin{aligned}
T_1 &= \min\{n \geq 1 : S_n \geq 1\} \\
T_2 &= \min\{n \geq 1 : S_n \geq 2\}.
\end{aligned}$$

1. Determine whether $(S_n)_{n \geq 0}$ is a \mathcal{G} -martingale, where $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$.
2. Determine whether T_1 and/or T_2 are stopping times for \mathcal{G} .
3. Compute $\mathbb{E}S_{T_1}$ and $\mathbb{E}S_{T_2}$.

1. We must check the three conditions:

- (a) *measurability*: $X_j \in m\mathcal{G}_n, \forall j \leq n$, hence $S_n \in m\mathcal{G}_n$; ✓
- (b) *integrability*:

$$\mathbb{E}|S_n| = \mathbb{E}\left|\sum_{j=1}^n X_j\right| \leq \mathbb{E}\sum_{j=1}^n |X_j| = n\mathbb{E}|X_j| = n < \infty.$$
✓

- (c) *martingale property*:

$$\begin{aligned}
\mathbb{E}(S_{n+1} | \mathcal{G}_n) &= \mathbb{E}\left(\sum_{j=1}^{n+1} X_j | \mathcal{G}_n\right) \\
&= \mathbb{E}(X_{n+1} + S_n | \mathcal{G}_n) \\
&= \mathbb{E}(X_{n+1} | \mathcal{G}_n) + \mathbb{E}(S_n | \mathcal{G}_n) \\
&= \mathbb{E}(X_{n+1}) + S_n \quad \text{since } X_{n+1} \perp \mathcal{G}_n \text{ and } S_n \in m\mathcal{G}_n \\
&= 0 + S_n.
\end{aligned}$$
✓

2. We have clearly $T_i : \Omega \rightarrow \mathbb{N}$ and they are measurable (hence random variables) because S_n, S_{n-1} are random variables. Let's check the condition $\{T_i = n\} \in \mathcal{G}_n, \forall n \geq 0$. Remember that this is equivalent to check $\{T_i \leq n\} \in \mathcal{G}_n, \forall n \geq 0$.

$$\begin{aligned}
\{T_1 \leq n\} &= \left\{ \omega : \inf\{m \geq 1 : \underbrace{S_m - S_{m-1}}_{X_m} = 1\} \leq n \right\} \\
&= \left\{ \omega : \min\{m \geq 1 : X_m = 1\} \leq n \right\} \\
&= \left\{ \omega : X_m(\omega) = 1 \text{ for some } m \leq n \right\} \\
&= \bigcup_{m=1}^n \underbrace{\{X_m(\omega) = 1\}}_{\in \mathcal{G}_n, \forall m \leq n} \in \mathcal{G}_n.
\end{aligned}$$

So T_1 is a stopping time. Now check $\{T_2 \leq n\}$.

$$\begin{aligned}\{T_2 \leq n\} &= \left\{ \omega : S_m(\omega) = 1 \text{ for some } m \leq n \right\} \\ &= \bigcup_{m=1}^n \{S_m = 1\} \\ &= \bigcup_{m=1}^n \left\{ \sum_{j=1}^n X_j = 1 \right\} \in \mathcal{G}_n. \\ &\in \mathcal{G}_n, \forall m \leq n \text{ because } X_j \in m\mathcal{G}_n \text{ for } j \leq m \leq n.\end{aligned}$$

So T_2 is a \mathcal{G} -stopping time.

3. Let's calculate $\mathbb{E}S_{T_2}$ first. We have $S_{T_2} = 1$ a.s., hence $\mathbb{E}(S_{T_2}) = 1$. Notice that $\mathbb{E}(S_0) = 0 \neq \mathbb{E}(S_{T_2})$, so the variation/extension of the Doob's theorem can't apply. This means that T_2 must be unbounded... Let's calculate $\mathbb{E}S_{T_1}$ now. Notice that T_1 is unbounded hence Doob's theorem does not apply. Let's check whether S_{T_1} is integrable. Since $\{\{T_1 = n\}, n \in \mathbb{N}\}$ is a partition of Ω we have:

$$\begin{aligned}\mathbb{E}|S_{T_1}| &= \sum_{n=1}^{\infty} \mathbb{E}(|S_n| \mathbf{1}_{\{T_1=n\}}) \\ &= \sum_{n=1}^{\infty} \mathbb{E}\left[\left|\sum_{j=1}^n X_j\right| \mathbf{1}_{\{T_1=n\}}\right] \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^n \mathbb{E}[|X_j| \mathbf{1}_{\{T_1=n\}}] \\ &= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} \mathbb{E}[|X_j| \mathbf{1}_{\{T_1=n\}}] \\ &= \sum_{j=1}^{\infty} \mathbb{E}[|X_j| \mathbf{1}_{\{T_1 \geq j\}}] \\ &= \mathbb{E}|X_1| \sum_{j=1}^{\infty} \mathbb{P}(\{T_1 \geq j\}) \\ &= \mathbb{E}|X_1| \cdot \mathbb{E}(T_1).\end{aligned}$$

The last equation comes from the fact that X_j are i.i.d., that $X_j \in m\mathcal{G}_j$ and that $\{T_1 \geq j\} = \{T_1 \leq j-1\}^c \in \mathcal{G}_{j-1}$, so we know that $X_j \perp \{T_1 \geq j\}$ (for $X \geq 0$, $\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geq x) dx = \sum_{k=0}^\infty \mathbb{P}(X \geq k)$ ⁸). We have $\mathbb{P}(T_1 = k) = \frac{1}{2} \cdot \frac{1}{2^{k-1}} = \frac{1}{2^k}$, hence

$$\mathbb{E}T_1 = \sum_{k=1}^{\infty} k \frac{1}{2^k} < \infty \implies S_{T_1} \in L^1(\Omega).$$

Doing the same computations without absolute value we get

$$\mathbb{E}(S_{T_1}) = \mathbb{E}(X_1)\mathbb{E}(T_1) = 0 \cdot \mathbb{E}(T_1) = 0.$$

before that we put \leq due to $|\sum X_j| \leq \sum |X_j|$

There is an alternative (shorter) solution to compute $\mathbb{E}S_{T_2}$, using the extended Doob's Theorem since $T_2 < \infty \mathbb{P} - \text{a.s.}$ Given that $T_2 \sim \text{Geom}\left(\frac{1}{2}\right)$ hence

$$\begin{cases} \mathbb{E}T_2 < \infty \\ T_2 \geq 0 \end{cases} \implies T_2 < +\infty \mathbb{P} - \text{a.s.}$$

Moreover, the increments of the martingale S are

$$S_n - S_{n-1} = X_n \sim \text{Be}\left(\frac{1}{2}\right)$$

hence $|X_n(\omega)| \leq 1 \forall \omega \in \Omega$. This tells us that S_{T_2} is integrable and $\mathbb{E}(S_{T_2}) = \mathbb{E}(S_0) = 0$.

⁸What the fuck?

Exercise 5

Let X be a random variable with values in \mathbb{N} such that $\mathbb{P}(X \geq n) = p^n$ for some $p \in (0, 1)$. Let $Y_n = p^{-n} \mathbb{1}_{\{X \geq n\}}$, $n \geq 0$, and $\mathcal{G}_n = \sigma(Y_1, \dots, Y_n)$.

1. Determine if \mathcal{G}_n is a martingale.
2. Determine if $T := \min\{n \geq 0 : Y_n = 0\}$ is a stopping time for \mathcal{G}_n .
3. Show that $T \in \mathcal{H}^{\mathbb{P}}(\Omega)$.
4. Compute $\mathbb{E}[T]$.

1. (a) Measurability is obvious by definition of \mathcal{G}_n .

- (b) To check for integrability:

$$\mathbb{E}|Y_n| = \mathbb{E}(Y_n) = \mathbb{E}(p^{-n} \mathbb{1}_{\{X \geq n\}}) = p^{-n} \mathbb{P}(X \geq n) = p^{-n} p^n = 1.$$

- (c) Martingale property: $\mathbb{E}(Y_{n+1} | \mathcal{G}_n) = Y_n$. We must show that, $\forall G \in \mathcal{G}$,

$$\mathbb{E}(\mathbb{1}_G Y_{n+1}) = \mathbb{E}(\mathbb{1}_G Y_n). \quad (\clubsuit)$$

Since we know that $\mathcal{G}_n = \sigma(Y_1, \dots, Y_n)$, then

$$\begin{aligned} Y_n = 0 &\iff X < n \\ Y_n = p^{-n} &\iff X \geq n. \end{aligned}$$

So G is of the form

$$G = \{Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n\}.$$

If $Y_n \neq 0$ then we have $y_1, y_2, \dots, y_{n-1} \neq 0$ because $X \geq n > n-1 > \dots > 1$. If $y_n = 0 \iff X < n$ then

$$X > n + 1 \iff Y_{n+1} = 0. \quad (\spadesuit)$$

This also means=

$$\mathbb{E}(\mathbb{1}_G Y_n) = \begin{cases} 0 & \text{if } y_n = 0 \\ p^{-n} & \text{if } y_n \neq 0 \end{cases} \quad (\clubsuit\clubsuit)$$

$$\mathbb{E}(\mathbb{1}_G Y_{n+1}) = \begin{cases} 0 & \text{if } y_n = 0 \text{ (because of } \spadesuit) \\ ? & \text{if } y_n \neq 0. \end{cases} \quad (\clubsuit\clubsuit\clubsuit)$$

There are two possible values for Y_{n+1} :

$$\begin{aligned} Y_{n+1} = 0 &\iff X < n + 1 \\ Y_{n+1} = p^{-n-1} &\iff X \geq n + 1. \end{aligned}$$

If $y_n \neq 0$ we know that $X \geq n$. So we get

$$\begin{aligned} \mathbb{1}_G Y_{n+1} = 0 &\iff X < n + 1 \text{ given that } X \geq n \\ \mathbb{1}_G Y_{n+1} = p^{-n-1} &\iff X \geq n + 1 \text{ given that } X \geq n \end{aligned}$$

that is

$$\begin{aligned} \mathbb{P}(\mathbb{1}_G Y_{n+1} = 0) &= \mathbb{P}(X < n + 1 | X \geq n) \\ &= \frac{\mathbb{P}(X < n + 1, X \geq n)}{\mathbb{P}(X \geq n)} \\ &= \frac{\mathbb{P}(X = n) - \mathbb{P}(X \geq n + 1)}{\mathbb{P}(X \geq n)} \\ &= \frac{p^n - p^{n+1}}{p^n} \\ &= 1 - p \end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}(\mathbb{1}_G Y_{n+1} = p^{n-1}) &= \mathbb{P}(X \geq n+1 | X \geq n) \\
&= \frac{\mathbb{P}(X \geq n+1, X \geq n)}{\mathbb{P}(X \geq n)} \\
&= \frac{\mathbb{P}(X \geq n+1)}{\mathbb{P}(X \geq n)} \\
&= \frac{p^{n+1}}{p^n} \\
&= p.
\end{aligned}$$

Thus if $y_n \neq 0$ we get

$$\begin{aligned}
\mathbb{E}(\mathbb{1}_G Y_{n+1}) &= 0 \cdot \mathbb{P}(\mathbb{1}_G Y_{n+1} = 0) + p^{-n-1} \mathbb{P}(\mathbb{1}_G Y_{n+1} = p^{-n-1}) \\
&= 0 + p^{-n-1} \cdot p = p^{-n}
\end{aligned}$$

and plugging this into purple flower we get purple flower . Thus purple flower holds and $(Y_n)_n$ is a martingale with respect to \mathcal{G} .

2. We have

$$\begin{aligned}
T &= \min\{n \geq 0 : Y_n = 0\} \\
&= \min\{n \geq 0 : \mathbb{1}_{\{X \geq n\}} = 0\} \\
&= X + 1.
\end{aligned}$$

Note that the indicator $\mathbb{1}_{\{X \geq n\}}$ is 1 if $X \geq n$ and when $X < n$ it switches to 0. $X \geq n$ means $X = n - 1 \iff n = X + 1$. Thus the conditions for stopping time reduce to

$$\{T \leq n\} = \{X + 1 \leq n\} \in \mathcal{G}_n$$

and

$$\{X + 1 \leq n\}^n = \{X + 1 > n\} = \{X > n + 1\} = \underbrace{\{X \leq n\}}_{\text{since } Y_n = p^{-n} \mathbb{1}_{\{X \geq n\}}} \in \mathcal{G}_n$$

so $\{T \leq n\} \in \mathcal{G}_n$, hence it is a stopping time.

3. Consider

$$\begin{aligned}
\mathbb{E}X &= \sum_{n \geq 1} \mathbb{P}(X \geq n) \\
&= \sum_{n \geq 1} p^n \\
&= -1 + \sum_{n \geq 0} p^n \\
&= -1 + \frac{1}{1-p}
\end{aligned} \tag{88}$$

We know that

$$\begin{aligned}
\mathbb{E}|T| &= \mathbb{E}|X + 1| \leq 1 + \mathbb{E}|X| = 1 + \mathbb{E}X \\
&= 1 + \sum_{n=0}^{\infty} n \cdot \mathbb{P}(X = n) \\
&= 1 + \sum_{n=0}^{\infty} n(\mathbb{P}(X \geq n) - \mathbb{P}(X \geq n + 1)) \\
&= 1 + \sum_{n=0}^{\infty} (p^n - p^{n+1}).
\end{aligned} \tag{89}$$

Compute the members one by one.

$$\sum_{n=0}^{\infty} p^n = \frac{1}{1-p} \quad \text{so} \quad \sum_{n=0}^{\infty} np^{n-1} = \frac{1}{(1-p)^2} \quad \text{for } p < 1$$

so

$$\sum_{n=0}^{\infty} np^n = p \sum_{n=0}^{\infty} np^{n-1} = \frac{p}{(1-p)^2}$$

and

$$\sum_{n=0}^{\infty} np^{n+1} = p^2 \sum_{n=0}^{\infty} np^{n-1} = \frac{p^2}{(1-p)^2}.$$

This gives us

$$\begin{aligned} &= 1 + \frac{p}{(1-p)^2} - \frac{p^2}{(1-p)^2} \\ &= \frac{(1-p)^2 + p - p^2}{(1-p)^2} \\ &= \frac{1 + p^2 - 2p + p - p^2}{(1-p)^2} \\ &= \frac{1-p}{(1-p)^2} \\ &= \frac{1}{1-p} < \infty. \end{aligned}$$

Remark

To compute $\mathbb{E}(X)$ we can also use the formula for positive random variables:

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \quad \text{or} \quad \mathbb{E}(X) = \int_0^{\infty} \mathbb{P}(X \geq x) dx$$

and get the same result $\mathbb{E}(X) = \frac{1}{1-p} - 1$. Take

$$A := \{(x, \omega) : 0 \leq x \leq X(\omega)\}.$$

Then

$$\begin{aligned} \int_{\Omega} \int_0^{\infty} \mathbb{1}_A(x, \omega) dx d\mathbb{P} &= \int_0^{\infty} \int_{\omega} \mathbb{1}_A(x, \omega) d\mathbb{P} dx \\ &= \int_0^{\infty} \mathbb{P}(X \geq x) dx. \end{aligned}$$

Exercise 6

Let $X \in L^1(\Omega)$ and \mathcal{G}_n be a filtration. Let $Y_n := \mathbb{E}(X|\mathcal{G}_n)$.

1. Show that $(Y_n)_{n \in \mathbb{N}}$ is a \mathcal{F} -martingale.
2. Show that for any \mathcal{G} -stopping time T such that $\mathbb{P}(T < \infty) > 0$ one has $Y_T \in L^1(\Omega)$.

1. (a) Measurability: obvious by definition. ✓
- (b) Integrability: obvious by definition. ✓
- (c) Martingale property:

$$\begin{aligned} \mathbb{E}[Y_{n+1}|\mathcal{G}_n] &= \mathbb{E}[\mathbb{E}(X|\mathcal{G}_{n+1})|\mathcal{G}_n] \\ &= \mathbb{E}[X|\mathcal{G}_n] = Y_n. \end{aligned}$$

2.

$$\begin{aligned}
\mathbb{E}|Y_T| &= \mathbb{E}\left|\sum_{n=0}^{\infty} Y_n \mathbf{1}_{\{T=n\}}\right| \\
&\leq \sum_{n=0}^{\infty} \mathbb{E}(\mathbb{E}(|X| \mid \mathcal{G}_n) \mathbf{1}_{\{T=n\}}) \\
\mathbf{1}_{\{T=n\}} \in m\mathcal{G}_n \longrightarrow &= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{E}(|X| \mathbf{1}_{\{T=n\}} \mid \mathcal{G}_n)] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[|X| \mathbf{1}_{\{T=n\}}] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty} |X| \mathbf{1}_{\{T=n\}}\right] \\
&= \mathbb{E}|X| < +\infty.
\end{aligned}$$
✓

Exercise 7

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent random variables with $\mathbb{E}|X_n| > 1$, $\forall n \in \mathbb{N}$. Let $Y_n = \prod_{i=1}^n X_i$ and $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$. Determine whether Y_n is a \mathcal{G} -martingale.

(a) Measurability: follows from the measurability of X_i and the fact that Y_n is a product of X_i with $i \leq n$. ✓

(b) Integrability:

$$\mathbb{E}|Y_n| = \mathbb{E}\left|\prod_{i=1}^n X_i\right| = \prod_{i=1}^n \mathbb{E}|X_i| = 1 < \infty.$$
✓

(c) Martingale probability:

$$\begin{aligned}
\mathbb{E}(Y_{n+1} \mid \mathcal{G}_n) &= \mathbb{E}\left(\prod_{i=1}^{n+1} X_i \mid \mathcal{G}_n\right) \\
&= \mathbb{E}(X_{n+1} \cdot Y_n \mid \mathcal{G}_n) \\
Y_n \in m\mathcal{G}_n \longrightarrow &= Y_n \mathbb{E}(X_{n+1} \mid \mathcal{G}_n) \\
X_{n+1} \perp \mathcal{G}_n \longrightarrow &= Y_n \mathbb{E}(X_{n+1}) = Y_n.
\end{aligned}$$
✓

4.1.8 Exercise class 8

Revise with Kotatsu!

Theorem 4.1.6

Martingale convergence theorem. Let X be a submartingale. Suppose that $(X_n)_{n \in \mathbb{N}}$ is bounded in L^1 , uniformly in n (that is, $\sup_n \mathbb{E}|X_n| < \infty$). Then X_n converges a.s. to an integrable random variable.

Remark

$$\sup_n \mathbb{E}|X_n| < \infty \iff \sup_n \mathbb{E}X_n^+ < \infty.$$

Definition 4.1.7

A sequence of random variables $(X_n)_{n \in \mathbb{N}}$ is said to be **uniformly integrable** if

$$\lim_{b \rightarrow \infty} \sup_n \mathbb{E}[|X_n| \mathbf{1}_{\{|X_n| > b\}}] = 0.$$

If a random variable is uniformly integrable then it is L^1 -bounded. If it is L^p -bounded and $p > 1$ then it is uniformly integrable.

Theorem 4.1.7

Martingale convergence theorem (equivalence). Let X be a submartingale. Then X converges a.s. and in L^1 to an integrable random variable if and only if it is uniformly integrable.

Exercise 1

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables with $\mathbb{E}X_n = 0$, $\mathbb{E}X_n^2 = 1$. Let $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. Determine the a.s. convergence of S_n . Decide whether S_n is uniformly L^1 -bounded.

If $\lim_n S_n$ exists and it is finite a.s. it should be absolutely summable so it must be that $|X_i| \rightarrow 0$ a.s. as $i \rightarrow \infty$. This however is impossible because X_i are i.i.d.. So S_n does not converge a.s.. If $\sup_n \mathbb{E}|S_n| < \infty$ then by the martingale convergence theorem (if S_n is a martingale) we would have S_n converging a.s., but this is not the case and thus S_n is not L^1 -bounded.

(a) Measurability with respect to natural filtration. ✓

(b) Integrability:

$$\mathbb{E}|S_n| \leq \sum_{i=1}^n \mathbb{E}|X_i| = n \cdot c < \infty.$$
✓

(c) Martingale property:

$$\begin{aligned} \mathbb{E}[S_{n+1} | \mathcal{F}_n] &= \mathbb{E}\left[\sum_{i=1}^n X_i + X_{n+1}\right] \\ &= \mathbb{E}(S_n | \mathcal{F}_n) + \mathbb{E}(X_{n+1} | \mathcal{F}_n) \\ &= S_n + \mathbb{E}(X_{n+1}) = S_n, \end{aligned}$$
✓

Exercise 2

Let $(\xi_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables such that $\mathbb{E}\xi_i = 0$ $\forall i \in \mathbb{N}$ and $\mathbb{E}|\xi_i| = \frac{1}{2^n}$, $\forall i \in \mathbb{N}$. Let $X_n = \sum_{i=1}^n \xi_i$, $n \geq 1$. Determine the a.s. convergence of X_n . Say if X_n is uniformly L^1 -bounded.

$(X_n)_n$ is a martingale (same as the exercise 1, we used only the fact that ξ_i are independent and $\mathbb{E}\xi_i = 0$). Thus we could apply martingale convergence theorem:

$$\sup_n \mathbb{E}|X_n| = \sup_n \mathbb{E}\left|\sum_{i=1}^n \xi_i\right| \leq \sup_n \sum_{i=1}^n \mathbb{E}|\xi_i| \leq \sup_n \sum_{i=1}^n \frac{1}{2^n} = 1.$$

Hence $X_n \rightarrow X$ a.s. for some $X \in L^1$ (since $(X_n)_n$ is L^1 -bounded).

Exercise 3

Let $(Y_n)_{n \in \mathbb{N}}$ be a sequence of i.i.d. random variables with distribution $Y_n \sim \text{I}((1, -1))$. Let $X_n = \sum_{i=1}^n Y_i$. Determine the a.s. convergence of X_n . Say if X_n is uniformly L^1 -bounded.

$$\mathbb{E}\frac{Y_i}{i} = \frac{1}{i} \left(1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2}\right) = 0.$$

Thus X_n is a martingale with respect to its natural filtration, because it is a sum of independent random

variables with mean 0 (see exercise 1). Let's check whether $(X_n)_n$ is L^1 -bounded.

$$\begin{aligned}\sup_n \mathbb{E}|X_n| &= \sup_n \mathbb{E} \left| \sum_{i=1}^n \frac{Y_i}{i} \right| \\ &\leq \sup_n \sum_{i=1}^n \frac{1}{i} \mathbb{E}|Y_i| \\ &= \sup_n \sum_{i=1}^n \frac{1}{i} \left(1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \right) \\ &= \sum_{i=1}^n \frac{1}{i} = +\infty.\end{aligned}$$

From this alone we cannot conclude anything. Let's see if $(X_n)_n$ is L^2 -bounded. This would imply L^1 -boundedness, hence convergence a.s. to a L^1 random variable.

$$\begin{aligned}\mathbb{E}|X_n|^2 &= \mathbb{E} \left(\sum_{i=1}^n \frac{Y_i}{i} \right)^2 \\ &= \mathbb{E} \left(\sum_{i=1}^n \frac{Y_i^2}{i^2} + \sum_{\substack{i,j=1 \\ i \neq j}} \frac{Y_i Y_j}{i j} \right) \\ &= \sum_{i=1}^n \frac{1}{i^2} \mathbb{E} Y_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}} \frac{1}{i j} \underbrace{\mathbb{E}(Y_i Y_j)}_{\substack{=0 \\ =0}} \\ &= \sum_{i=1}^n \frac{1}{i^2} \cdot 1 + 0.\end{aligned}$$

So

$$\sup_n \mathbb{E}|X_n|^2 = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

hence $(X_n)_n$ is L^2 -bounded \Rightarrow it is uniformly integrable \Rightarrow converges a.s. to a random variable in L^1 and it converges also in L^1 .

Exercise 4

Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables defined recursively as follows:

$$X_1 \sim \text{U}(0, 1)$$

$$X_2 \sim \text{U}(0, X_1)$$

$$\vdots$$

$$X_n \sim \text{U}(0, X_{n-1}).$$

Let $\mathcal{G}_m = \sigma(X_1, X_2, \dots, X_m)$.

1. Compute $\mathbb{P}[X_n]$.

2. Determine \mathcal{G}_n such that $X_n \in \mathcal{G}_n$, X_n is a \mathcal{G} -measurable

3. Determine the convergence of X_n

- We formalise the procedure to construct X_n and write the conditional p.d.f.:

$$\begin{aligned}f_{X_1}(x) &= \mathbb{1}_{[0,1]}(x) = \frac{1}{x_0} \mathbb{1}_{[0,x_0]}(x) \quad \text{with } x_0 = 1 \\ f_{X_2|X_1=x_1}(x) &= \frac{1}{x_1} \mathbb{1}_{[0,x_1]}(x) \\ &\vdots \\ f_{X_n|X_{n-1}=x_{n-1}, \dots, X_1=x_1}(x) &= f_{X_n|X_{n-1}=x_{n-1}}(x) = \frac{1}{x_{n-1}} \mathbb{1}_{[0,x_{n-1}]}(x).\end{aligned}$$

Thus we have

$$\mathbb{E}(X_n | \mathcal{G}_{n-1}) = \mathbb{E}(X_n | X_{n-1})$$

and

$$\mathbb{E}(X_n | X_{n-1} = x_{n-1}) = \int x f_{X_n | X_{n-1}=x_{n-1}}(x) dx = \frac{1}{x_{n-1}} \int_0^{x_{n-1}} x dx = \frac{1}{x_{n-1}} \cdot \frac{x_{n-1}^2}{2} = \frac{x_{n-1}}{2}.$$

So $\mathbb{E}(X_n | X_{n-1}) = \frac{x_{n-1}}{2}$ by the freezing lemma. This implies that $\frac{x_{n-1}}{2}$ is a version of the conditional expectation $\mathbb{E}(X_n | \mathcal{G}_{n-1})$. This helps because by the tower property

$$\begin{aligned} \mathbb{E}(X_n) &= \mathbb{E}(\mathbb{E}(X_n | \mathcal{G}_{n-1})) \\ &= \mathbb{E}\left(\frac{X_{n-1}}{2}\right) \\ &= \frac{1}{2} \mathbb{E}(\mathbb{E}(X_{n-1} | \mathcal{G}_{n-2})) \\ &= \frac{1}{2} \mathbb{E}\left(\frac{X_{n-2}}{2}\right) \\ &\quad \vdots \\ &\text{by recurrence } \rightarrow \quad \vdots \\ &= \frac{1}{2^{2-1}} \mathbb{E}X_1 = \frac{1}{2^{n-1}} \cdot \frac{1}{2} = \frac{1}{2^n}. \end{aligned}$$

2. Given any $(C_n)_n$ we have

$$\mathbb{E}(Y_n | \mathcal{G}_{n-1}2) = C_n \mathbb{E}(X_n | \mathcal{G}_{n-1}) = \frac{C_n}{2} X_{n-1} \stackrel{?}{=} Y_{n-1}.$$

For $Y_{n-1} = C_{n-1}X_n$ we must have $C_{n-1} = \frac{C_n}{2}$ i.e. $C_n = 2C_{n-1}$ defined recursively with C_1 initial condition to get

$$C_n = 2C_{n-1} = 2^2 C_{n-2} = 2^{n-1} C_1.$$

3. Notice that if $C_1 > 0$ then $Y_n \geq 0$, if $C_1 < 0$ then $Y_n \leq 0$. Let's consider the case $C_1 > 0$ (the other case is analogous with a minus sign).

$$\mathbb{E}|Y_n| = \mathbb{E}(Y_n) = \mathbb{E}[C_n X_n] = 2^{n-1} C_1 \mathbb{E}X_n = 2^{n-1} C_1 \cdot \frac{1}{2} = \frac{C_1}{2}$$

and this implies

$$\sup_n \mathbb{E}|Y_n| = \frac{C_1}{2} < \infty \quad (\text{independent of } n.)$$

Hence by the martingale convergence theorem we have $Y_n \rightarrow Y_\infty$ a.s. wth $Y_\infty \in L^1$.

Exercise 5

Consider the following betting game. A player has random initial capital $X_0 \in L$ and $X_0 > 0$ a.s.. She places a bet of $\frac{X_0}{2}$ which gives X_1 if she wins and 0 if she loses so that her total capital after one bet is $X_1 = X_0 + \frac{X_0}{2}$ (win) or $X_1 = \frac{X_0}{2}$ (loss). The probability of losing is $\frac{1}{2}$ and s is the probability of winning. The player plays repeatedly (one bet is independent of the previous one) and each time she bets $\frac{X_n}{2}$ on her current capital. We denote by X_n her capital after n bets.

- Determine if $(X_n)_n$ is a martingale, where $\mathcal{G}_n = \sigma(X_0, \dots, X_n)$
- Determine if $X_n \rightarrow X$ a.s. and specify X
- Determine if $X_n \rightarrow X$ in L^1 .

1. (a) Measurability is obvious.

(b) Integrability:

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_n \leq \mathbb{E}\left(X_{n-1} + \frac{X_{n-1}}{2}\right) \leq \dots \text{const} \cdot \mathbb{E}(X_0) \leq \infty.$$

- (c) Martingale property: we have to show $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = X_n$ for every $n \in \mathbb{N}$. Notice that $\mathbb{E}(X_{n+1}|\mathcal{G}_n) = \mathbb{E}(X_{n+1}|X_n)$. We want to apply the freezing lemma so we calculate

$$\mathbb{E}(X_{n+1}|X_n = x_n) = \frac{1}{2} \left(\frac{x_n}{2} + \frac{3}{2}x_n \right) = x_n$$

thus by the freezing lemma we have

$$\mathbb{E}(X_{n+1}|X_n) = X_n.$$

2.

$$\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_0 < \infty \implies \sup_n \mathbb{E}|X_n| = \mathbb{E}X_0$$

hence $X_n \rightarrow X$ a.s. for some $X \in L^1$. In order to determine X we make the following calculations"

$$X_{n+1} = \begin{cases} \frac{X_n}{2} \\ \frac{X_n}{2} + X_n \end{cases} \implies |X_{n+1} - X_n| = \begin{cases} \left| \frac{X_n}{2} - X_n \right| \\ \left| \frac{X_n}{2} + X_n - X_n \right| \end{cases} = \frac{|X_n|}{2} \\ \Downarrow \\ 2|X_{n+1} - X_n| = X_n$$

Taking the a.s. limit on both sides we get $2|X - X| = X$ which implies $X = 0$.

3. Since $\mathbb{E}[|X_n - 0|] = \mathbb{E}|X_n| = \mathbb{E}X_n \neq 0$ since $X_0 > 0$ a.s. this implies that X_n does not converge in 0 in L^1 .

Exercise 6

Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. random variables with distribution $X_i \sim \text{Be}\left(\frac{1}{2}\right)$ and let $S_n = \sum_{i=1}^n X_i$. Fix $p \in (0, 1)$ and $q := 1 - p$. Define $M_n = (2p)^{S_n} (2q)^{n-S_n}$.

- Determine whether (M_n) is a martingale with respect to $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$.
- Determine the a.s. convergence of M_n to M and determine M .

1. (a) Measurability holds because M_n is a continuous function of X_1, \dots, X_n .
(b) Integrability:

$$\begin{aligned} \mathbb{E}|M_n| = \mathbb{E}M_n &= \mathbb{E}[(2p)^{S_n} (2q)^{n-S_n}] \\ &= \mathbb{E}\left[(2p)^{\sum_{i=1}^n X_i} (2q)^{\sum_{i=1}^n (1-X_i)}\right] \\ &= \mathbb{E}\left[\prod_{i=1}^n (2p)^{X_i} (2q)^{1-X_i}\right] \\ &= \prod_{i=1}^n \mathbb{E}[(2p)^{X_i} (2q)^{1-X_i}] \\ &= \left[\frac{1}{2}(2p)^0 \cdot (2q)^{1-0} + \frac{1}{2}(2p)^1 (2q)^{1-1}\right]^n \\ &= \left[\frac{1}{2}2q + \frac{1}{2}2p\right]^n \\ &= (p+q)^n \\ &= 1^n = 1 < \infty. \end{aligned}$$

- (c) Martingale property: setting $Z_i = (2p)^{X_i} (2q)^{1-X_i}$ we have $M_n = \prod_{i=1}^n Z_i$ and $Z_i \perp \mathcal{L}$. Then

$$\begin{aligned} \mathbb{E}(M_{n+1}|\mathcal{G}_n) &= \mathbb{E}(Z_{n+1}M_n|\mathcal{G}_n) \\ M_n \in m\mathcal{G}_n \rightarrow &= M_n \mathbb{E}(Z_{n+1}|\mathcal{G}_n) \\ Z_{n+1} \perp \mathcal{G}_n \rightarrow &= M_n \mathbb{E}_{(Z_{n+1})} \\ \mathbb{E}Z_{n+1} = 1 \rightarrow &= M_n. \end{aligned}$$

2.

$$\mathbb{E}|M_n| = \mathbb{E}M_n = \mathbb{E} \prod_{i=1}^n Z_i = \prod_{i=1}^n \mathbb{E}Z_i = 1$$

thus $\sup_n \mathbb{E}|M_n| = 1$. By martingale convergence theorem $M_n \rightarrow M$ a.s. for some $M \in L^1$. To determine M we must distinguish for different values of p .

- If $p = \frac{1}{2}$ then $M_n \equiv 1$ and so $M = 1$.
- If $p \neq \frac{1}{2}$ we note that

$$\begin{aligned} M_n &= (2p)^{S_n} (2q)^{n-S_n} \\ &= \exp \{ \log [(2p)^{S_n} (2q)^{n-S_n}] \} \\ &= \exp \{ \log(2p)^{S_n} + \log(2q)^{n-S_n} \} \\ &= \exp \{ S_n \log(2p) + (n - S_n) \log(2q) \} \\ &= \exp \left\{ n \left[\frac{S_n}{n} \log(2p) + \left(1 - \frac{S_n}{n}\right) \log(2q) \right] \right\}. \end{aligned}$$

By SLLN we have $\frac{S_n}{n} \rightarrow \mathbb{E}X_1$ a.s. with $\mathbb{E}X_1 = \frac{1}{2}$. Thus setting

$$Y_n = \frac{S_n}{n} \log(2p) + \left(1 - \frac{S_n}{n}\right) \log(2q)$$

we have

$$Y_n \xrightarrow{\text{a.s.}} \frac{1}{2} \log(2p) + \frac{1}{2} \log(2q) = \frac{1}{2} \log(4pq)$$

where

$$4pq = 4p(1-p) < 1 \implies \frac{1}{2} \log(4pq) < 0.$$

We also have $M_n = \mathbb{E}(n \cdot Y_n)$ and by limit of composite functions (on a set of probability 1) we get

$$\lim_{n \rightarrow \infty} M_n = e^{-\infty} = 0 \quad \text{a.s.}$$

which means $M = 0$.

Exercise 7

Let (\mathcal{F}_n) be a filtration and $(X_n)_{n \in \mathbb{N}}, (Y_n)_{n \in \mathbb{N}}$ be two \mathcal{F} -martingales. Let $T = \inf\{n \geq 1 : X_n = Y_n\}$

1. Determine if T is a \mathcal{F} -stopping time.
2. Letting $Z_n = (X_n - Y_n) \mathbf{1}_{\{T \geq n\}}$ determine if $(Z_n)_{n \in \mathbb{N}}$ is a \mathcal{F} -martingale

1. We must check whether $\{T \leq n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$, or equivalently if $\{T = n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

$$\{T = n\} = \{\underbrace{X_1 \neq Y_1}_{\in \mathcal{F}_1 \subset \mathcal{F}_n}, \underbrace{X_2 \neq Y_2}_{\in \mathcal{F}_2 \subset \mathcal{F}_n}, \dots, \underbrace{X_{n-1} \neq Y_{n-1}}_{\in \mathcal{F}_{n-1} \subset \mathcal{F}_n}, \underbrace{X_n = Y_n}_{\in \mathcal{F}_n}\} \in \mathcal{F}_n.$$

Remember that X and Y are \mathcal{F} -martingales and hence measurable.

2. • Measurability:

$$\{T \geq n\} = \{T > n-1\} = \{T \leq n-1\}^c$$

and by the previous point we know that $\{T \leq n-1\} \in \mathcal{F}_{n-1}$ so $\{T \geq n\} \in m\mathcal{F}_{n-1}$. Moreover $X_n, Y_n \in m\mathcal{F}_n$ have $Z_n \in m\mathcal{F}_n$.

- Integrability:

$$\mathbb{E}[(X_n - Y_n) \mathbf{1}_{\{T \geq n\}}] \leq \mathbb{E}|X_n| + \mathbb{E}|Y_n| < \infty.$$

- Martingale property:

$$\begin{aligned} \mathbb{E}[(X_n - Y_n) \mathbf{1}_{\{T \geq n\}} | \mathcal{G}_{n-1}] &= \mathbf{1}_{\{T \geq n\}} \mathbb{E}[X_n - Y_n | \mathcal{G}_{n-1}] \\ &= \mathbf{1}_{\{T \geq n\}} (X_{n-1} - Y_{n-1}) \\ &= (\mathbf{1}_{\{T \geq n-1\}} - \mathbf{1}_{T=n-1}) (X_{n-1} - Y_{n-1}) \\ &= \mathbf{1}_{\{T \geq n-1\}} (X_{n-1} - Y_{n-1}) - \underbrace{\mathbf{1}_{\{T=n-1\}} (X_{n-1} - Y_{n-1})}_{=0 \text{ because if } T = n-1 \text{ it means } X_{n-1} = Y_{n-1}} \\ &= \mathbf{1}_{\{T \geq n-1\}} (X_{n-1} - Y_{n-1}). \end{aligned}$$

4.2 Questions for the oral examination

Here I will try and answer to the questions for the anal examination that were on the moodle page for the academic year 2022-2023. I hope that they will be valid for the years to come as well.

4.2.1 Transition kernels: definition, example, and their usage in the extension of measures to product spaces

First of all, the definition:

Definition 4.2.1

Let (E, \mathcal{E}) and (F, \mathcal{F}) be measurable spaces. Let K be a mapping from $E \times \mathcal{F}$ into $\overline{\mathbb{R}}_+$. Then K is called **transition kernel** from space (E, \mathcal{E}) into space (F, \mathcal{F}) if:

- the mapping $x \mapsto K(x, B)$ is \mathcal{E} -measurable $\forall B \in \mathcal{F}$;
- the mapping $B \mapsto K(x, B)$ (the second mapping of the kernel, the one regarding the set) is a measure $\forall x \in E$.

Then the example:

Example 4.2.1

Take v , a finite measure on (F, \mathcal{F}) and take k , a positive function on $(E \times F)$ which is measurable with respect to $\mathcal{E} \otimes \mathcal{F}$, the product σ -algebra. Then, we integrate

$$\int_B v(dy)k(x, y) \quad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

We see how this object depends on x and on the choice of B (a function of x and B ...). It defines a transition kernel

$$K(x, B) = \int_B v(dy)k(x, y) \quad \begin{matrix} B \in \mathcal{F} \\ x \in E \end{matrix}$$

from (E, \mathcal{E}) into (F, \mathcal{F}) .

Now the extensions of measures on product spaces;

Theorem 4.2.1

Extension of measures on product spaces.

Let μ be a measure on the measurable space (E, \mathcal{E}) . Let K be a Σ -finite^a transition kernel from space (E, \mathcal{E}) into (F, \mathcal{F}) . Then:

- ① if we take our function $f(x, y)$, integrate it against our kernel $K(x, dx)$ over F and then integrate again against measure μ over E , the operation

$$\pi f = \int_E \mu(dx) \int_F K(x, dy) f(x, y)$$

defines a measure π on $(E \times F, \mathcal{E} \otimes \mathcal{F})$;

- ② if μ is σ -finite and K is σ -bounded then π is σ -finite and it is the unique measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ satisfying

$$\pi(A \times B) = \int_A \mu(dx) K(x, B) \quad \forall A \in \mathcal{E}, B \in \mathcal{F}.$$

^aErm... what the sigma?

4.2.2 Kolmogorov's 0-1 law: proof and an example of its usage

First, the statements of the theorem.

Theorem 4.2.2

Kolmogorov's 0-1 law

This is a theorem about independence. Let $\mathcal{G}_1, \mathcal{G}_2, \dots$ be independent. Then

$$\mathbb{P}(H) = \begin{cases} 0 & \forall H \in \tau. \\ 1 & \end{cases}$$

and here's the short proof:

Proof

Start from your independency $(\mathcal{G}_1, \mathcal{G}_2, \dots)$, then put the last \mathcal{G} 's in a partition (which is still an independency) $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n, \tau_n)$. Now consider that the tail σ -algebra is a subset of τ_n and this means that $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n, \tau)$ is still an independency... But we also know that this is true for every n up to ∞ and so we can extend the independency to a collection of countably many partitions, since $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n, \tau)_n$ us a countably infinite sequence of independencies. So

$$(\tau, \mathcal{G}_1, \mathcal{G}_2, \dots)$$

is an independency. Do another partition (which still gives us an independency) and get the independency

$$(\tau, \tau_0)$$

but then remember that τ_0 is an *union* of the σ -algebras from 1 to ∞ while τ is an *intersection* of all τ_n including τ_0 , so we have that $\tau \subset \tau_0$.

Consider now $\mathbb{P}(H \cap G)$ where $H \in \tau$ and $G \in \tau_0$. Due do independency we have that

$$\mathbb{P}(H \cap G) = \mathbb{P}(H)\mathbb{P}(G)$$

but since all elements of τ are also elements of τ_0 we can choose $G \equiv H$ and our equation becomes

$$\mathbb{P}(H \cap G) = \mathbb{P}(H \cap H) = \mathbb{P}(H) = \mathbb{P}(H)^2$$

and the solution to this can only be 0 or 1. \square

We still need an example of the application... that would be, for example, the behavior in the limits of a sequence of random variables with independence between $\{X_n\}$. Then if the limit exists it is a constant, since $\lim_n X_n$ belongs to the tail σ -algebra of the sequence. Independence between random variables tells us that we can apply Kolmogorov's 0-1 law to say that $\mathbb{P}(\lim_n X_n = \infty)$ is either 0 or 1... but if we somehow know that the limit exists then it can't be ∞ so its probability of being ∞ is 0 and therefore it is a constant almost surely! This is cool because it tells us that the limit is either infinite or finite. This is useful for sub-martingales.

4.2.3 Almost sure convergence: definition, properties and characterization theorem

Start with the definition:

Definition 4.2.2

A real-valued sequence of random variables $(X_n)_n$ on $(\Omega, \mathcal{H}, \mathbb{P})$ is said to be **almost sure convergent** (a.s. convergent) is the numerical sequence

$$(X_n(\omega))_n$$

converges for almost all $\omega \in \Omega$.

It is said to converge to X if X is an almost sure real-valued random variable and

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

for almost all $\omega \in \Omega$.

The characterization is

Theorem 4.2.3

Characterization of almost sure convergence:

A sequence of real valued random variables $(X_n)_n$ converges to X almost surely if and only if, for every $\varepsilon > 0$,

$$\sum_n i_\varepsilon \circ |X_n - X| < \infty$$

almost surely.

Where i_ε is the indicator function of (ε, ∞) :

$$i_\varepsilon(x) = \mathbb{1}_{(\varepsilon, \infty)}(x) = \begin{cases} 1, & x > \varepsilon \\ 0, & x \leq \varepsilon. \end{cases}$$

This basically means that in every interval from a certain point ε to ∞ the sum of the differences of X_n and X must not diverge. Here are some properties:

- Comparison with Other Types of Convergence:

- If $\{X_n\}$ converges almost surely to X , then $\{X_n\}$ also converges to X in probability.
- Almost sure convergence implies convergence in distribution, but not necessarily vice versa.

Closure Under Linear Operations:

- **Addition:** If $X_n \rightarrow X$ a.s. and $Y_n \rightarrow Y$ a.s., then $X_n + Y_n \rightarrow X + Y$ a.s.
- **Multiplication:** If $X_n \rightarrow X$ a.s. and $Y_n \rightarrow Y$ a.s., then $X_n Y_n \rightarrow XY$ a.s. (provided X and Y are bounded or measurable in a compatible way).

- Countable Additivity:

- **Countable Unions:** If $\{X_n\}$ converges almost surely to X , then for any countable collection of events $\{A_i\}$ with $\Pr(A_i) \rightarrow 0$, $\Pr(\limsup_{i \rightarrow \infty} A_i) = 0$.

- Uniform Integrability:

- **Expectation Convergence:** If $\{X_n\}$ converges almost surely to X and $\{X_n\}$ is uniformly integrable, then $X_n \rightarrow X$ in L^1 , meaning $\mathbb{E}[|X_n - X|] \rightarrow 0$.

- Continuity from Below and Above:

- **From Below:** If $X_n \leq X_{n+1}$ a.s. for all n and $X_n \rightarrow X$ a.s., then $X_n \leq X$ a.s.
- **From Above:** If $X_n \geq X_{n+1}$ a.s. for all n and $X_n \rightarrow X$ a.s., then $X_n \geq X$ a.s.

- Interchange of Limits:

- **Expectation Limit:** If $X_n \rightarrow X$ a.s. and $\{X_n\}$ is uniformly integrable, then $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

- Fatou's Lemma and Monotone Convergence Theorem:

- **Fatou's Lemma:** If $X_n \geq 0$ almost surely and $X_n \rightarrow X$ a.s., then $\mathbb{E}[X] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n]$.
- **Monotone Convergence Theorem (for non-negative random variables):** If $X_n \rightarrow X$ a.s. and $X_n \geq 0$ a.s., then $\mathbb{E}[X] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n]$.

4.2.4 Borel Cantelli lemmas; proof of the divergence part

Enunciate the two lemmas side by side (and with a slightly different formulation).

Theorem 4.2.4

Borel-Cantelli lemmas

- a) Let $(B_n)_n$ be a sequence of Bernoulli random variables.

$$\sum_n \mathbb{E}B_n < +\infty \implies \sum_n B_n < +\infty \text{ a.s.}$$

b) If

$$\sum_n \mathbb{E}B_n = \infty \text{ and } (B_n)_n \text{ are pairwise independent}$$

then

$$\sum_n B_n = +\infty \text{ a.s.}$$

Now let's prove the divergent part.

Note the notation!

$$p_n = \mathbb{E}B_n, \quad a_n = \sum_{i=1}^n \mathbb{E}B_i = \sum_{i=1}^n p_i$$

and we get the partial sum and the limits

$$S_n = \sum_{i=1}^n B_i, \quad S = \lim_n S_n$$

Proof

Ok, this is long. We know that, since the random variables are pairwise independent,

$$\text{Var } S_n = \sum_{i=1}^n \text{Var } B_n = \sum_{i=1}^n p_n(1-p_n) \leq \sum_{i=1}^n p_n = a_n.$$

Then fix $b \in (0, \infty)$. Since we know that $(a_n)_n$ is increasing towards ∞ then the same can be said of

$$(a_n - \sqrt{ba_n})_n$$

since we are subtracting to a_n a quantity that goes towards ∞ more slowly. So the sequence still goes towards infinity and this means that the event $S < \infty$ is basically the limit of the increasing sequence of events

$$\{S < a_n - \sqrt{ba_n}\}_n.$$

Since $S_n \leq S$ we have that

$$\{S < a_n - \sqrt{ba_n}\} \subset \{S_n < a_n - \sqrt{ba_n}\}$$

and we also have that

$$\{S_n < a_n - \sqrt{ba_n}\} \subset \{|S_n - a_n| > \sqrt{ba_n}\}$$

Since we are dealing with inclusion of events we can look at this from the point of view of probability measures:

$$\mathbb{P}(S < a_n - \sqrt{ba_n}) \leq \mathbb{P}(S_n < a_n - \sqrt{ba_n}) \leq \mathbb{P}(|S_n - a_n| > \sqrt{ba_n}).$$

We take the \limsup_n for two of these probabilities:

$$\limsup_n (\mathbb{P}(S < a_n - \sqrt{ba_n})) \leq \limsup_n (\mathbb{P}(|S_n - a_n| > \sqrt{ba_n}))$$

but the first term of this inequality becomes

$$\begin{aligned} \limsup_n (\mathbb{P}(S < a_n - \sqrt{ba_n})) &= \mathbb{P}(\lim_n (\mathbb{P}(S < a_n - \sqrt{ba_n}))) \\ &= \mathbb{P}(S < \infty) \end{aligned}$$

since the quantity $\{S < a_n - \sqrt{ba_n}\}_n$ goes towards infinity. Now we can rewrite the previous inequality and use Chebyshev's inequality:

$$\begin{aligned} \mathbb{P}(S < \infty) &\leq \limsup_n \left(\mathbb{P}(|S_n - a_n| > \sqrt{ba_n}) \right) \\ &\leq \limsup_n \left(\frac{\text{Var } S_n}{ba_n} \right) \\ &\leq \limsup_n \left(\frac{a_n}{ba_n} \right) = \frac{1}{b}. \end{aligned}$$

Since b is arbitrary we can let it go towards infinity and thus obtain that

$$\mathbb{P}(S < \infty) \leq 0$$

But this means that the probability of the complementary event must be 1:

$$\mathbb{P}(S = \infty) = 1.$$

□

4.2.5 Borel-Cantelli lemmas; proof of the convergence part, relations with a.s. convergence and examples of its application

Start by enunciating the theorem.

Theorem 4.2.5

First Borel-Cantelli Lemma Let $(H_n)_n$ be a sequence of events. Then

$$\sum_n \mathbb{P}(H_n) < +\infty \implies \sum_n \mathbb{1}_{H_n} < +\infty \quad \text{a.s.}$$

The proof is no shit:

Proof

Denote $N = \sum \mathbb{1}_{H_n}$ so that $\sum_n \mathbb{P}(H_n) = \sum_n \mathbb{E} \mathbb{1}_{H_n} = \mathbb{E} \sum \mathbb{1}_{H_n} = \mathbb{E}N$. This means that the new claim is

“If $\mathbb{E}N < \infty$ then $N < \infty$ ”

But this is true because if the expectation is finite so is its random variable. Of course almost surely... □

Now the implications for a.s. convergence:

Proposition 4.2.1

Let

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty \quad \forall \varepsilon > 0.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

Proposition 4.2.2

Suppose that there exists a sequence $(\varepsilon_n)_n$ decreasing to 0 such that

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < +\infty.$$

Then

$$X_n \xrightarrow{\text{a.s.}} X.$$

So we don't need ε to be constant, but just to be decreasing to 0.

Proposition 4.2.3

Suppose that there exists a sequence of positive numbers $(\varepsilon_n)_n$ such that

$$\sum_n \varepsilon_n < +\infty, \quad \sum_n \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) < +\infty$$

Then X_n converges almost surely.

4.2.6 Convergence in probability: definition, properties and theorem on the metric for convergence in probability

Let's start with the definition:

Definition 4.2.3

Let $(X_n)_n$ be a sequence of real-valued random variables. Then $(X_n)_n$ converges to a further real-valued random variable **in probability** if

$$\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

I am not sure about properties, why can't you state things clearly? Fuck off.

Theorem 4.2.6

Characterization theorem for convergence in probability.

- i) if $(X_n)_n$ converges to X almost surely, then it converges to X in probability;
- ii) if $(X_n)_n$ converges in probability to X , then it has a subsequence converging to the same random variable X almost surely;
- iii) if every subsequence of the main sequence has a further subsequence converging to X almost surely, then the main sequence converges to X in probability.

Remark

Convergence in probability is preserved under arithmetic operations.

So, for example if we have $X \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$ then $(X_n + Y_n)_n \xrightarrow{\mathbb{P}} X + Y$.

We will now introduce a metric for convergence in probability: since we are talking about *distance*, we may think⁹ that this has something to do with *metric spaces*, where we measure the distance between different objects. Indeed there is a connection between measure spaces and metric spaces...

Let us introduce now a metric for convergence in probability. If we want to calculate a metric between random variables X and Y we can define the following metric:

$$d(X, Y) = \mathbb{E}(|X - Y| \wedge 1).$$

Remark

$$\textcircled{1} \quad d(X, Y) = 0 \iff X = Y \text{ a.s.}$$

$$\textcircled{2} \quad d(X, Y) + d(Y, Z) \geq d(X, Z).$$

d is a metric on the space of real-valued random variables if X and Y are identified as the same random variable if $X = Y$ almost surely.

⁹A remark that reeks of overestimation.

Proposition 4.2.4

$$X_n \xrightarrow{\mathbb{P}} X \iff d(X_n, X) \xrightarrow{n \rightarrow \infty} 0.$$

4.2.7 Uniform integrability: definitions and its consequences

Start with the definition

Definition 4.2.4

A collection of random variables K is said to be **uniformly integrable** if

$$k(b) = \sup_{X \in K} (\mathbb{E}|X| \mathbb{1}_{(|X| > b)})$$

goes to 0 as $b \rightarrow \infty$.

What about the properties?

- ① If K is finite and each $X \in K$ is integrable then K is uniformly integrable.
- ② If K is dominated by an integrable random variable Z then it is uniformly integrable.
- ③ Uniform integrability of a collection K implies the so-called L^1 -boundedness, which means

$$k \subset L^1 \quad \text{and} \quad k(0) = \sup_K \mathbb{E}|X| < \infty.$$

Note that $k(0)$ considers the whole random variable without truncation.

- ④ We know that uniform integrability implies L^1 ... but the converse is not true. We can prove it by a counterexample.
 - ⑤ If K is L^p -bounded with $p > 1$ then it is uniformly integrable with $f(x) = x^p$.
- To prove this we recur to the following proposition:

Proposition 4.2.5

Suppose it exists a positive borel function f on \mathbb{R}_+ such that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{x} = \infty$$

and write

$$c = \sup_{X \in K} \mathbb{E} f \circ |X|. \quad (\star)$$

If $c < \infty$ then K is uniformly integrable.

and one last theorem

Theorem 4.2.7

The following are equivalent:

- ① K is uniformly integrable;
- ② $h(b) = \sup_K \int_b^{+\infty} dy \mathbb{P}(|X| > y) \xrightarrow{b \rightarrow \infty} 0$;
- ③ $\sup_K \mathbb{E} f \circ |X| < +\infty$ for some increasing convex function f on \mathbb{R}_+ such that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{x} = +\infty.$$

4.2.8 Law of large numbers: proof of the related theorems

What the hell? This is the definition

Theorem 4.2.8

Law of large numbers:

Let $(X_n)_n$ be a sequence of pairwise independent random variables with the same distribution as X . If $\mathbb{E}X$ exists (infinite values are admitted!) then

$$\bar{X}_n \xrightarrow{n} \mathbb{E}X.$$

To do this we need two theorems with their related proof.

Theorem 4.2.9

Consider a sequence of real valued, pairwise independent random variables X_1, X_2, \dots with finite common expectation and variance

$$\mathbb{E}X_n = a, \quad \text{Var } X_n = b.$$

Then:

$$\textcircled{1} \quad \bar{x}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{L^2} a;$$

$$\textcircled{2} \quad \bar{x}_n \xrightarrow{\mathbb{P}} a;$$

$$\textcircled{3} \quad \bar{x}_n \xrightarrow{\text{a.s.}} a.$$

Proof

1. Define

$$S_n = n\bar{X}_n.$$

Now we have, due to independency,

$$\mathbb{E}S_n = na, \quad \text{Var } S_n = nb$$

and, consequently,

$$\mathbb{E}\bar{X}_n = a, \quad \text{Var } \bar{X}_n = \frac{b}{n}.$$

To prove L^2 -convergence, consider

$$\mathbb{E}|\bar{X}_n - a|^2 = \text{Var } \bar{X}_n \xrightarrow[b \rightarrow \infty]{} 0$$

$$\text{so } \bar{X}_n \xrightarrow{L^2} a.$$

2. This one comes from Markov's inequality.

3. Assume $X_n \geq 0$ (without loss of generality as always since we can just divide negative and positive part). Consider a subsequence $N = (n_k)_{k \in \mathbb{N}^*}$ where $n_k = k^2$. Apply Chebyshev's inequality

$$\mathbb{P}(|\bar{X}_n - a| > \varepsilon) \leq \frac{b}{\varepsilon^2 k^2}$$

and by summing everything

$$\varepsilon^2 \sum_{n \in N} \mathbb{P}(|\bar{X}_n - a| > \varepsilon) \leq \sum_{k=1}^{\infty} \frac{b}{k^2} \leq \infty$$

so we can apply borel-cantelli to $\mathbb{P}(|\bar{X}_n - a| > \varepsilon)$ since it is a finite quantity. This means that

$$\bar{X}_n \xrightarrow{\text{a.s.}} a$$

along N . Call Ω_0 the almost sure set on which the convergence take place. Remember that if we have $N = (n_k) \in \mathbb{N}^*$ and $\lim_k \frac{n_{k+1}}{n_k} = r > 0$ and the sequence $(X_n)_n$ converges along N to a then

$$\frac{a}{r} \leq \liminf \bar{X}_n \leq \limsup \bar{X}_n \leq ra.$$

So note that

$$\frac{(k+1)^2}{k^2} \xrightarrow[k \rightarrow \infty]{} 1$$

and hence $\forall \omega \in \Omega_0$ we have that

$$\lim_{n \rightarrow \infty} \bar{X}_n(\omega) = a$$

with probability 1 which means almost sure convergence. \square

Then there is the second statement:

Proposition 4.2.6

Let $(X_n)_n$ be a sequence of positive independent and identically distributed (i.i.d.) random variables with

$$\mathbb{E}X_1 = +\infty.$$

Consider also a further random variable X distributed as X_1 (which means that they also have the same expectation). Then

$$\bar{X}_n \xrightarrow{\text{a.s.}} \infty.$$

Proof

Start by defining $Y_n = X_n \wedge b$ for some $b \in \mathbb{R}$. Then we have $\bar{Y}_n = \frac{\sum_{i=1}^n Y_i}{n}$. Since the random variable is truncated we have

$$\mathbb{E}Y_n = \mathbb{E}[X_n \wedge b] < \infty$$

but by the previous result we know that

$$Y_n \xrightarrow{\text{a.s.}} \mathbb{E}[X_n \wedge b].$$

Now consider the fact that $X_n \geq Y_n$ for every n , so we expect

$$\liminf_n X_n \geq \lim_n \bar{Y}_n = \mathbb{E}[X_n \wedge b]$$

for every b , even when b tends to ∞ (which causes $\mathbb{E}[X_n \wedge b]$ to become $\mathbb{E}X_n$). But this means

$$\liminf_n \bar{X}_n = \infty \text{ a.s.}$$

and since \bar{X}_n is a non-decreasing sequence

$$\bar{X}_n \xrightarrow{\text{a.s.}} \infty.$$

\square

4.2.9 Central limit theorem (Lyapunov theorem) Proof of the theorem and of the corresponding lemma (Lindeberg Lemma)

Theorem 4.2.10

Lyapunov Central limit Theorem:

Suppose $\mathbb{E}X_{nj} = 0 \forall n, j$ and $\text{Var } Z_n = 1 \forall n$ and $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$. Then

$$Z_n \xrightarrow{d} Z \sim N(0, 1).$$

Here we have replaced the conditions on the random variables with the condition $\lim_n \sum_n \mathbb{E}|X_{nj}|^3 = 0$ which allows us to use the structure of the triangular array. To prove this theorem we need the following lemma:

Lemma 4.2.1

Lindeberg's lemma: Let (Y_1, Y_2, \dots, Y_k) be independent random variables with mean zero and let $S = \sum_{j=1}^k Y_j$. Let us further assume that $\text{Var } S = 1$. Let f be a function which can be differentiated 3 times and let f', f'', f''' be bounded and continuous and such that

$$|f'''| \leq c, \quad c \in \mathbb{R}_+.$$

Then for $Z \sim N(0, 1)$

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ Z| \leq c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

Proof

Let Z_1, \dots, Z_k be independent normal random variables with mean $\mathbb{E}Z_j = \mathbb{E}Y_j = 0$ for $j = 1, \dots, k$ and variance $\text{Var } Z_j = \text{Var } Y_j$ for $j = 1, \dots, k$. Then construct

$$T = \sum_{j=1}^k Z_j \sim N\left(0, \sum_{j=1}^k \text{Var } Z_j = \sum_{j=1}^k \text{Var } Y_j = 1\right).$$

So we know that T is distributed as Z (they are both $N(0, 1)$) so $T \xrightarrow{d} Z$ and since we are using the expectation of Z we can replace $\mathbb{E}f \circ Z$ with $\mathbb{E}f \circ T$. So the Lindeberg's lemma becomes

$$|\mathbb{E}f \circ S - \mathbb{E}f \circ T| \leq c \sum_{j=1}^k \mathbb{E}|Y_j|^3$$

which we want to prove, to exploit the structure of T . Define now the random variables V_1, V_2, \dots, V_k as follows:

$$\begin{aligned} V_1 &\text{ s.t. } S = V_1 + Y_1 \\ V_2 &\text{ s.t. } V_1 + Z_1 = V_2 + Y_2 \\ &\vdots \\ V_j &\text{ s.t. } V_j + Z_j = V_{j+1} + Y_{j+1}, \quad 1 \leq j < k \\ &\vdots \\ V_k &\text{ s.t. } V_k + Z_k = T. \end{aligned}$$

Note that

$$\begin{aligned} V_1 &= Y_2 + Y_3 + \dots + Y_k \\ V_2 &= Z_1 + Y_3 + \dots + Y_k \\ V_3 &= Z_1 + Z_2 + Y_4 + \dots + Y_k \end{aligned}$$

so the Y get replaced by the Z one at the time in the V . We can now focus on the following expression:

$$\begin{aligned} f \circ S - f \circ T &= f(V_1 + Y_1) - f(V_k + Z_k) \\ &= f(V_1 + Y_1) + f(V_2 + Y_2) - \underbrace{f(V_2 + Y_2)}_{f(V_1 + Z_1)} + f(V_3 + Y_3) - \underbrace{f(V_3 + Y_3)}_{f(V_2 + Z_2)} + \\ &\quad + \dots + f(V_k + Z_k) \\ &= \sum_{j=1}^k f(V_j + Y_j) - \sum_{j=1}^k f(V_j + Z_j). \end{aligned}$$

Now take the expectation and the absolute value:

$$\begin{aligned} |\mathbb{E}f \circ s - \mathbb{E}f \circ t| &= \left| \sum_{j=1}^k \mathbb{E}f(V_j + Y_j) - \sum_{j=1}^k \mathbb{E}f(V_j + Z_j) \right| \\ &\leq \sum_{j=1}^k |\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \end{aligned}$$

and now we only need to prove that

$$|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \leq c\mathbb{E}|Y_j|^3.$$

Let's write the Taylor formula for this function:

$$f(v+x) = f(v) + f'(v)x + \frac{1}{2}f''(v)x^2 + R_2(v, x)$$

where

$$\begin{aligned} R_2(v, x) &= \frac{1}{2} \int_v^{v+x} (v+x-t)_2 f''(t) dt \\ &\leq \frac{1}{2} c \int_v^{v+x} (v+x-t)_2 dt = c \frac{x^3}{6} \end{aligned}$$

so that

$$|R_2(v, x)| \leq \frac{c}{6} |x|^3.$$

We now have

$$\begin{aligned} f(V_j + Y_j) &= f(V_j) + f'(V_j)Y_j + \frac{1}{2}f''(V_j)Y_j^2 + R_2(V_j, Y_j) \\ f(V_j + Z_j) &= f(V_j) + f'(V_j)Z_j + \frac{1}{2}f''(V_j)Z_j^2 + R_2(V_j, Z_j) \end{aligned}$$

and subtract side by side:

$$f(V_j + Y_j) - f(V_j + Z_j) = (Y_j - Z_j)f'(V_j) + \frac{1}{2}f''(V_j)(Y_j^2 - Z_j^2) + R_2(V_j, Y_j) - R_2(V_j, Z_j).$$

Now take the expectation

$$\begin{aligned} \mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j) &= \frac{1}{2}\mathbb{E}f''(V_j)\underbrace{(\mathbb{E}Y_j^2 - \mathbb{E}Z_j^2)}_{=0 \text{ since they have same variance}} + \mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)] \\ &= \mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)]. \end{aligned}$$

Now take the absolute value

$$\begin{aligned} |\mathbb{E}[R_2(V_j, Y_j) - R_2(V_j, Z_j)]| &\leq \mathbb{E}|[R_2(V_j, Y_j)]| + \mathbb{E}|[R_2(V_j, Z_j)]| \\ &\leq \frac{c}{6} (\mathbb{E}|Y_j|^3 + \mathbb{E}|Z_j|^3) \end{aligned}$$

so that

$$|\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| \leq \frac{c}{6} (\mathbb{E}|Y_j|^3 + \mathbb{E}|Z_j|^3).$$

Recall that $Z_j \sim N(0, b^2)$ where $b^2 = \mathbb{E}Y_j^2$. We know that

$$\mathbb{E}|Z_j|^3 = b^3 \sqrt{\frac{8}{\pi}} \leq 2b^3$$

and we also have

$$b = (\mathbb{E}Y_j^2)^{\frac{1}{2}} \leq (\mathbb{E}|Y_j|^3)^{\frac{1}{3}}$$

Because L^2 norm is less or equal than L^3 norm (revise inclusions in L^p -spaces for different values of p). But this last inequality is equivalent to

$$b^3 \leq \mathbb{E}|Y_j|^3$$

which leads to

$$\mathbb{E}|Z_j|^3 \leq 2b^3 \leq 2\mathbb{E}|Y_j|^3.$$

Finally we get

$$\begin{aligned} |\mathbb{E}f(V_j + Y_j) - \mathbb{E}f(V_j + Z_j)| &\leq \frac{c}{6} (\mathbb{E}|Y_j|^3) \\ &= \frac{c}{6} 3\mathbb{E}|Y_3|^3 \\ &= \frac{c}{2} \mathbb{E}|Y_3|^3 \leq c\mathbb{E}|Y_3|^3. \end{aligned}$$

□

This was horrible, horrible. Truly an horrible experience and honestly useless proof. And we still have to prove Lyapunov's theorem.

Proof

Lyapunov's CLT. Recall

$$Z_n = \sum_j X_{nj} \quad Z \sim N(0, 1).$$

We are interested in evaluating the characteristic function.

$$\begin{aligned} e^{irZ_n} &= \cos rZ_n + i \sin rZ_n \\ e^{irZ} &= \cos rZ + i \sin rZ. \end{aligned}$$

Consider now

$$\begin{aligned} |\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}| &= |(\mathbb{E}\cos rZ_n - \mathbb{E}\cos rZ) + i(\mathbb{E}\sin rZ_n - \mathbb{E}\sin rZ)| \\ &\leq \|\mathbb{E}\cos rZ_n - \mathbb{E}\cos rZ\| + \|i\| \|\mathbb{E}\sin rZ_n - \mathbb{E}\sin rZ\|. \end{aligned}$$

By applying the above lemma we obtain

$$|\mathbb{E}e^{irZ_n} - \mathbb{E}e^{irZ}| \leq \sum_j |r|^3 \mathbb{E}|X_{nj}|^3 + \sum_j |r|^3 \mathbb{E}|X_{nj}|^3$$

and this is possible since both sine and cosine are differentiable three times and they are bounded by 1 (so our $c = 1$). Now, according to the hypotheses of Lyapunov's theorem, we need to take the limit considering the hypothesis that $\lim_{n \rightarrow \infty} \sum_j \mathbb{E}|X_{nj}|^3 = c$ and we obtain the claim. This theorem applies to all triangular arrays which include the one in the CLT. □

4.2.10 Definition of conditional expectation and its main properties

First of all, remember that

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \int_A \mathbb{1}_B d\mathbb{P}$$

Evaluate our best estimate for $X(\omega)$, which may very well be the average of X over H :

$$\frac{1}{\mathbb{P}(H)} \int_H X(\omega) \mathbb{P}(d\omega) = \frac{1}{\mathbb{P}(H)} \mathbb{E}X \mathbb{1}_H = \mathbb{E}_H X.$$

If $\mathbb{P}(H) = 0$ then we allow any value for $\mathbb{E}_H X$.

Note the notation!

We usually denote that quantity as $\mathbb{E}(X|H)$ but Çinlar denotes it as $\mathbb{E}_H X$.

Remark

The quantity $\mathbb{E}_H X$ is called **conditional expectation of X given the event H** .

Consider the following facts:

1. $\bar{X} \in \mathcal{F}$: it is measurable with respect to \mathcal{F} and this is clear from the definition \mathbb{J} of the random variable \bar{X} ;
2. we know that for each $V \in \mathcal{F}_+$ we have that

$$\mathbb{E}VX = \mathbb{E}\bar{X}$$

which is called **projection property**.

Definition 4.2.5

Let \mathcal{F} be a sub- σ -algebra of \mathcal{H} . The conditional expectation $\mathbb{E}_{\mathcal{F}} X$ of X given \mathcal{F} is defined in two steps:

- ① for $X \in \mathcal{H}_+$ (positive random variables) it is any random variable \bar{X} satisfying:
 - (a) measurability ($\bar{X} \in \mathcal{F}_+$);
 - (b) projection property ($\mathbb{E}VX = \mathbb{E}\bar{X} \quad \forall V \in \mathcal{F}_+$).
- ② for arbitrary $X \in \mathcal{H}$, if $\mathbb{E}X$ exists, we define

$$\mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F}} X^+ - \mathbb{E}_{\mathcal{F}} X^-.$$

Otherwise, if $\mathbb{E}X^+ = \mathbb{E}X^- = \infty$, then $\mathbb{E}_{\mathcal{F}}$ is left undefined.

more properties:

① monotonicity:

$$X \leq Y \implies \mathbb{E}_{\mathcal{F}} X \leq \mathbb{E}_{\mathcal{F}} Y;$$

② linearity:

$$\mathbb{E}_{\mathcal{F}}(aX + bY + c) = a\mathbb{E}_{\mathcal{F}} X + b\mathbb{E}_{\mathcal{F}} Y + c;$$

③ monotone convergence theorem:

$$(X_n)_n \text{ s.t. } X_n \geq 0 \forall n, X_n \nearrow X \implies \mathbb{E}_{\mathcal{F}} X_n \nearrow \mathbb{E}_{\mathcal{F}} X;$$

④ Fatou's lemma:

$$X \geq 0 \implies \mathbb{E}_{\mathcal{F}} \liminf X_n \leq \mathbb{E}_{\mathcal{F}} X_n;$$

⑤ Dominated convergence theorem:

$$(X_n)_n \text{ a.s. } X_n \rightarrow X, |X_n| \leq Y, Y \text{ integrable} \implies \mathbb{E}_{\mathcal{F}} X_n \rightarrow \mathbb{E}_{\mathcal{F}} X;$$

⑥ Jensen's inequality:

$$f \text{ convex} \implies \mathbb{E}_{\mathcal{F}} f(x) \leq f(\mathbb{E}_{\mathcal{F}} X).$$

4.2.11 Existence and uniqueness of conditional expectation (for L1 random variables)

Theorem 4.2.11

Let $X \in \mathcal{H}$. Let \mathcal{F} be a sub- σ -algebra of \mathcal{H} . Then the conditional expectation $\mathbb{E}_{\mathcal{F}} X$ exists and it is unique up to equivalence.

Proof

$\forall H \in \mathcal{F}$ on the measurable space (Ω, \mathcal{F}) consider the restriction of \mathbb{P} on \mathcal{F} . Consider now

$$Q(H) = \int_H \mathbb{P}(d\omega) X(\omega)$$

where \mathbb{P} is a probability measure and Q is a measure which is absolutely continuous with respect to \mathbb{P} . In this measurable space random variables are functions, so we can apply Radon-Nikodym theorem.

So it exists a random variable $\bar{X} \in \mathcal{F}_+$ such that

$$\int_{\Omega} Q(d\omega) V(\omega) = \int_{\Omega} \mathbb{P}(d\omega) \bar{X}(\omega) V(\omega) \quad \forall V \in \mathcal{F}_+$$

so the projection property is satisfied and \bar{X} is a version of $\mathbb{E}_{\mathcal{F}} X$.

About the uniqueness: Let \bar{X} and $\bar{\bar{X}}$ be versions of $\mathbb{E}_{\mathcal{F}} X$, $X \geq 0$.

1. Both \bar{X} and $\bar{\bar{X}}$ are \mathcal{F}^+ -measurable;
2. $\mathbb{E}VX = \mathbb{E}V\bar{X} = \mathbb{E}V\bar{\bar{X}}$ for every $V \in \mathcal{F}_+$. Hence

$$\bar{X} = \bar{\bar{X}} \text{ a.s.}$$

Conversely, if $\mathbb{E}_{\mathcal{F}} X = \bar{X}$ and $\bar{X} \in \mathcal{F}_+$ and $\bar{X} = \bar{\bar{X}}$ a.s. then $\bar{\bar{X}}$ satisfies the projection property $\mathbb{E}XV = \mathbb{E}\bar{\bar{X}}V$ (i.e. $\bar{\bar{X}}$ is a version of $\mathbb{E}_{\mathcal{F}} X$).

□

4.2.12 Existence and uniqueness of conditional expectation (for L2 random variables, using orthogonal projection)

Theorem 4.2.12

$\forall X \in L^2(\mathcal{H})$ there exists a unique (up to equivalence) $\bar{X} \in L^2(\mathcal{H})$ such that

$$\mathbb{E}|X - \bar{X}|^2 = \inf_{Y \in L^2(\mathcal{F})} \mathbb{E}|X - Y|^2.$$

Furthermore, $X - \bar{X}$ is orthogonal to $L^2(\mathcal{F})$, i.e.

$$\mathbb{E}V(X - \bar{X}) = 0 \quad \forall V \in L^2(\mathcal{F})$$

Note that $L^2(\mathcal{H})$ is a complete Hilbert space in which the inner product of X and Y is given by $\mathbb{E}XY$. \bar{X} is the **orthogonal projection of the vector X onto the subspace $L^2(\mathcal{F})$** and the decomposition

$$X = \bar{X} + \tilde{X}$$

holds.

Proof

Let's write the L^2 -norm of X calling it $\|X\|$:

$$\|X\| = \|X\|_2 = \sqrt{\mathbb{E}X^2}.$$

Fix $X \in L^2(\mathcal{H})$. Define

$$\delta = \inf_{Y \in L^2(\mathcal{F})} \|X - Y\|.$$

Let $(Y_n)_n \subset L^2(\mathcal{F})$ such that $\delta_n = \|X - Y_n\| \xrightarrow{x \rightarrow \infty} 0$. Let us prove that $(Y_n)_n$ is a Cauchy sequence

for the $L^2(\mathcal{F})$ -convergence.

$$|Y_n - Y_m|^2 = 2|X - Y_m|^2 - 4|x - \underbrace{\frac{1}{2}(Y_n + Y_m)}_{\in L^2(\mathcal{F})}|^2.$$

Take the expectation on both sides:

$$\mathbb{E}|Y_n - Y_m|^2 \leq 2\delta_m^2 + 2\delta_n^2 - 4\delta^2.$$

Now we take the limit for n and m and what we get is

$$\lim_{m,n \rightarrow \infty} \mathbb{E}|Y_n - Y_m|^2 \leq 0.$$

Hence it is true that $(Y_n)_n$ is Cauchy and this means that there exists a $\bar{X} \in L^2(\mathcal{F})$ such that $|Y_n - \bar{X}| \xrightarrow{n \rightarrow \infty} 0$. Note that \bar{X} is unique up to equivalence (by definition of L^2 -norm). Note also

$$\bar{X} \in L^2(\mathcal{F}) \implies |X - \bar{X}| \geq \delta.$$

Now, by Minkowski's inequality we can write that

$$|X - \bar{X}| \leq \|X - Y_n\| + |Y_n - \bar{X}| \xrightarrow{n \rightarrow \infty} \delta + 0 = \delta.$$

We have thus

$$|X - \bar{X}| = \delta.$$

For $V \in L^2(\mathcal{F})$ and $a \in \mathbb{R}$, since $\mathbb{E}|X - \bar{X}|^2 = \delta$ then we have that

$$\begin{aligned} a^2 \mathbb{E}V^2 - 2a \mathbb{E}V(X - \bar{X}) + \delta^2 &= |aV - (X - \bar{X})|^2 \\ &= \|X - (\underbrace{aV + \bar{X}}_{\in L^2(\mathcal{F})})\|^2 = \delta^2 \end{aligned}$$

And therefore

$$a^2 \mathbb{E}V^2 - 2a \mathbb{E}V(X - \bar{X}) \leq 0 \quad \forall a \in \mathbb{R}$$

which is impossible unless

$$\mathbb{E}V(X - \bar{X}) = 0.$$

□

4.2.13 Filtration, adaptedness; example of non adapted process

Let T be a subset of \mathbb{R} . Let \mathcal{F}_t be a sub- σ -algebra of \mathcal{H} $\forall t \in T$. The family $\mathcal{F} = (\mathcal{F}_t)_{t \in T}$ is called **filtration** if $\mathcal{F}_s \subset \mathcal{F}_t$ for every $s < t$. Take, for example, \mathcal{F}_1 : it is much smaller than \mathcal{H} and if we say that a random variable is measurable with respect to \mathcal{F}_1 it means that we do not have much knowledge, but if we expand to \mathcal{F}_2 we are able to gain more knowledge about the random variable. If we interpret the index set T as time we may actually think about filtrations as our knowledge of the phenomenon as time passes.

Definition 4.2.6

Given $\mathcal{F} = (\mathcal{F}_t)_{t \in T}$, a stochastic process $\{X_t\}_{t \in T}$ is **adapted** to \mathcal{F} if X_t is \mathcal{F}_t -measurable for each $t \in T$ and \mathcal{E} -measurable.

As an example consider “the secretary problem”: in this case we must start from the filtration and then understand the problem. Here I have N candidates for a position; a candidate disregarded after the interview is lost. The interviewer wants to hire exactly 1 candidate and each candidate has different abilities and the interviewer knows only the relative ability of those already interviewed so far. Our goal is to maximizing the probability of hiring the best one. We have three questions:

1. what is Ω ?

2. what is the filtration \mathcal{F} for this experiment?

3. what process should we use?

In this case $\Omega = N!$ permutations of the ranking of the candidates (the order in which they show up) and the filtration is the information earned from interview up to time t (that is the ranking of the candidates up to time t). But what is the process that I should use? Consider the sequence

$$V_1, V_2, \dots \quad \{V_i\}_{i \geq 1}$$

with $V_i = 1$ if and only if the best candidate is the i -th candidate and $V_i = 0$ otherwise. Could this process $\{V_i\}_{i \geq 1}$ be used? No, because V is not adapted to \mathcal{F} ... because to understand if i -th candidate is the best we need to compare it to the other candidates, including the ones that didn't show up yet! But then how can we get an adapted process? Let us consider the expectation

$$U_n = \mathbb{E}[V_n | \mathcal{F}_n]$$

What do we know about the measurability of U_n ? We know that it is for sure \mathcal{F}_n -measurable. This trick gives us a simple way to build an adapted process. So now we will have: $U_n = 0$ if the candidate is not the best up to n and $U_n = 1$ otherwise. More specifically, we will have

$$\begin{aligned} U_n &= 1 \cdot \text{probability that the best candidate is among the first } n + 0 \cdot \text{probability that the best candidate is not among the first } n \\ &= 1 \cdot \frac{n}{N} + 0 \cdot \frac{N-n}{N} \\ &= \frac{n}{N} \end{aligned}$$

This is a quantity that I can measure and it is therefore adapted.

4.2.14 Random times and Stopping Times: examples of random times that are not stopping times and examples of stopping times

Definition 4.2.7

$$T : \Omega \mapsto \overline{\mathbb{T}}$$

is called **stopping time** of \mathcal{F} if

$$\{T \leq t\} \in \mathcal{F}_t \quad \forall t \in \mathbb{T}.$$

Stopping time have an alternative definition:

Definition 4.2.8

T is a stopping time if the process

$$Z_t = \mathbb{1}_{\{T \leq t\}} \quad t \in \mathbb{T}$$

is adapted to \mathcal{F} .

Starting from the previously defined random walk, consider the quantity

$$M_n = \min(S_1, \dots, S_n)$$

And the random time

$$T_2 = \min\{n : S_n \geq M_m + 2\}$$

is a stopping time. On the contrary,

$$T_3 = \begin{cases} \max\{n < 50 : S_n = 7\} & \text{if not empty} \\ 50 & \text{otherwise} \end{cases}$$

is not a stopping time. Why? Because I have to wait until $n = 50$ to answer the question.

4.2.15 Stopped filtration: definition and examples

Definition 4.2.9

Let \mathcal{F} be a filtration on $\overline{\mathbb{T}}$ and let T be a stopping time on \mathcal{F} . We call **past until T** the σ -algebra $\mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{H}$ such that:

$$\mathcal{F}_T = \left\{ H \in \mathcal{H} : H \cap \{T < t\} \in \mathcal{F}_t, \forall t \in \overline{\mathbb{T}} \right\}.$$

This means that \mathcal{F}_T , which represents the evolution within T , contains all the events H such that they are before the stopping time in all the filtrations. This is selecting the events H that happen within the time T .

Remark

If we fix a $T \equiv t$ then $\mathcal{F}_T \equiv \mathcal{F}_t$ (it's the normal filtration at time t).

If T is a stopping time of \mathcal{F} then $\{T \leq r\}$ belongs to \mathcal{F}_T , $\forall r \geq 0$:

$$\{T \leq r\} \cap \{T \leq t\} = \{T \leq \min(T, r)\} \in \mathcal{F}_t$$

So T is \mathcal{F}_T -measurable.

Remark

\mathcal{F}_t can be read as the collection of all \mathcal{F}_t -measurable random variables V : the value of $V(\omega)$ can be told by the time $T(\omega)$: we can read the value of $V(\omega)$ before the ringing of an alarm.

Imagine a factory. The production is blocked when

- a) the temperature of the room is above a certain threshold;
- b) the machinery has not been cleaned for more than 12 hours.

Alternatively, imagine that the sales of shares happens:

- a) when the price is above a fixed value;
- b) when the increase of the price is smaller than a fixed value.

4.2.16 Discrete stopping times associated to continuous stopping times

Definition 4.2.10

Let

$$d_n(t) = \begin{cases} \frac{k+1}{2^n} & \text{if } \frac{k}{2^n} \leq t < \frac{k+1}{2^n} \\ +\infty & \text{if } t = \infty. \end{cases}$$

We got a function

$$d_n : \overline{\mathbb{R}}_+ \mapsto \overline{\mathbb{R}}_+.$$

Observe that $d_1 \geq d_2 \geq d_3 \geq \dots$. What are the other properties of this function?

1. it is a step function;
2. it is right-continuous;
3. $d_n(t) > t$ (the diagonal line);
4. $\lim_{n \rightarrow \infty} d_n(t) = t$.

Now we apply this simple function to our stopping times so that we get the following proposition.

Proposition 4.2.7

Let \mathcal{F} be a filtration on $\overline{\mathbb{R}}_+$ and let T be a stopping time. Let

$$T_n = d_n \circ T.$$

Then $\{T_n\}_n$ is a sequence of **discrete stopping times** of \mathcal{F} decreasing to T .

Of course T_n is a foretold time by T , because if I know T I also know the value of T_n .

Proof

First of all, fix n . We know that:

- T_n is a measurable function of T so it is for sure \mathcal{F}_n -measurable;
- $d_n(t) \geq t$ for every $t < \infty$ and $d_n(\infty) = \infty$, so if we apply d_n to T we get $d_n(T) = T_n \geq T$;
- we know that T_n is a foretold time by T ;
- we know that T_n is a stopping time as well;
- we know that T_n is discrete.

But we also know that $d_n(t) \searrow t$ as $n \rightarrow \infty$ and this means that $d_n(T) = T_n \searrow T$. \square

We have thus switched from a continuous stopping time to a sequence of discrete stopping time.

4.2.17 Conditional expectation at stopping time

Can I use T fixed as a deterministic time t and thus get $\mathbb{E}(X|\mathcal{F}_t)$? Is this a special case or a different thing altogether? Well yes, because t is a stopping time being \mathcal{F}_t -measurable! $\mathbb{E}(X|\mathcal{F}_t)$ is a special case of $\mathbb{E}(X|\mathcal{F}_T)$. This becomes clear when we see the properties of stopped expectation. For $\forall X, Y, Z$ being positive random variables and for $\forall S, T$ stopping times of \mathcal{F} the following properties hold:

- **defining property:**

$$\mathbb{E}_T X = Y \iff Y \in \mathcal{F}_T \text{ and } \mathbb{E} Y X = \mathbb{E} Y;$$

- **unconditioning:**

$$\mathbb{E}\mathbb{E}_T X = \mathbb{E} X;$$

- **repeated conditioning:**

$$\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_{\min\{S,T\}} X = \mathbb{E}_{S \wedge T} X;$$

- **conditional determinism:**

$$\mathbb{E}_T(X + YZ) = X + Y\mathbb{E}_T(Z)$$

if $X, Y \in \mathcal{F}_T$.

We need to prove the third property.

Proof

Notice that if $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$ and since the poorer σ -algebra always wins we have

$$\mathbb{E}_S \mathbb{E}_T X = \mathbb{E}_S X = \mathbb{E}_{S \wedge T} X.$$

But if S and T are arbitrary we can apply the same result to $S \wedge T$ since for sure $S \wedge T \leq T$:

$$\mathbb{E}_{S \wedge T} \mathbb{E}_T X = \mathbb{E}_{S \wedge T} X. \quad (\star)$$

Remember that $Y = \mathbb{E}_T X$. We can now write the statement of the proof in a different way:

$$\mathbb{E}_S Y = \mathbb{E}_{S \wedge T} X.$$

and using property \star we get

$$\begin{aligned} \mathbb{E}_S Y &= \mathbb{E}_{S \wedge T} X \\ &= \mathbb{E}_{S \wedge T} X. \end{aligned}$$

This is what we need to prove now. Let's go through the usual two steps.

1. Note that $\mathcal{F}_{S \wedge T} \subset \mathcal{F}_S$. This means that $\mathbb{E}_{S \wedge T} Y$ is in \mathcal{F}_S and therefore $\mathbb{E}_{S \wedge T} Y$ has the measure property and it is a candidate to be a version of $\mathbb{E}_S Y$. We must now check whether this candidate satisfies the defining property of stopped expectation.

2. Check the defining property:

$$\mathbb{E}VY = \mathbb{E}V\mathbb{E}_{S \wedge T} Y$$

for each positive $V \in \mathcal{F}_S$. We start by fixing such $V \in \mathcal{F}_S$ positive. We proved that

$$V1_{(S \leq T)} Y = \mathbb{E}V1_{(S \leq T)} \mathbb{E}_{S \wedge T} Y. \quad (\textcircled{S})$$

Now we observe that $Y \in \mathcal{F}_T$ by definition. Notice that $Y1_{(T < S)} \in \mathcal{F}_{S \wedge T}$. We can apply this fact to the defining property and obtain

$$\mathbb{E}VY1_{(T < S)} = \mathbb{E}V\mathbb{E}_{S \wedge T} Y \underbrace{1_{(T < S)}}_{\mathcal{F}_{S \wedge T}\text{-measurable}}$$

and by conditional determinism we get

$$\begin{aligned} \mathbb{E}VY1_{(T < S)} &= \mathbb{E}V1_{(T < S)} \\ &= \mathbb{E}V1_{(T < S)} \mathbb{E}_{S \wedge T} Y. \end{aligned}$$

By putting this together with \textcircled{S} , we get

$$\mathbb{E}VY = \mathbb{E}V\mathbb{E}_{S \wedge T} Y.$$

□

4.2.18 Comparing stopping times: some interesting features

Uuuuh I'm not really sure what she meant by this. Of course god forbid a fucking SDS students has their life made easy once. I think she means this:

Theorem 4.2.13

- ① if $S \leq T \implies \mathcal{F}_S \subset \mathcal{F}_T$;
- ② $\mathcal{F}_{\min\{S,T\}} = \mathcal{F}_S \cap \mathcal{F}_T$;
- ③ if $V \in \mathcal{F}$, the following processes are in $\mathcal{F}_{\min\{S,T\}}$:

$$\begin{array}{ccc} (a) & (b) & (c) \\ V1_{S \leq T}, & V1_{S=T}, & V1_{S > T}. \end{array}$$

Proof

- Imagine $\{S \leq T\}$. Now consider the event $\{S \leq n\}$. Since S happens before T then we have that $\{T \leq n\} \subset \{S \leq n\}$. We know that if $A \in \mathcal{F}_S$ then

$$A \cap \{T \leq n\} = A \cap \underbrace{\{S \leq n\}}_{\in \mathcal{F}_n} \cap \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n}$$

which means that $A \cap \{T \leq n\} \in \mathcal{F}_n$ and by definition of \mathcal{F}_T this corresponds to saying

$$\mathcal{F}_S \subset \mathcal{F}_T.$$

- (a) Let us first prove that $\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cup \mathcal{F}_T$. Note that $\min\{S, T\}$ is dominated by \mathcal{F}_S and \mathcal{F}_T , which means that $\mathcal{F}_{\min\{S,T\}}$ is included both in \mathcal{F}_S and \mathcal{F}_T . By point 1 we have

$$\mathcal{F}_{\min\{S,T\}} \subset \mathcal{F}_S \cap \mathcal{F}_T.$$

- (b) It remains to prove that $\mathcal{F}_{\min\{S,T\}} \supseteq \mathcal{F}_S \cup \mathcal{F}_T$. Let $H \in \mathcal{F}_S \cap \mathcal{F}_T$. This means that

$$H \cap \{S \leq T\} \in \mathcal{F}_{\min\{S,T\}}$$

by part 3 of this theorem, since $H \in \mathcal{F}_S$. But we also know that

$$H \cap \{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}$$

by part 3 of this theorem since also $H \in \mathcal{F}$. But this means that

$$H = \{H \cap \{S \leq T\}\} \cup \{H \cap \{S \leq T\}\}$$

belongs to $\mathcal{F}_{\min\{S,T\}}$ and so

$$\mathcal{F}_{\{S \cap T\}} \subset \mathcal{F}_{\min\{S,T\}}.$$

3. (a) $V \mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}$. To prove this we use the following theorem:

Theorem 4.2.14

Let T be a stopping time of \mathcal{F} . Then

$$\mathcal{F}_T = \{X_T : X \in \mathcal{F}\}$$

with X in \mathcal{F} .

This basically means that \mathcal{F}_T becomes the values X_T of X in \mathcal{F} at time T . Let us consider $X_T = V \mathbb{1}_{\{S \leq t\}}$ with $t = \min\{S, T\}$. We know that $\min\{S, T\}$ is a stopping time of $\mathcal{F}_{\min\{S,T\}}$ and by the theorem above

$$X_{\min\{S,T\}} \in \mathcal{F}_{\min\{S,T\}}$$

or, alternatively,

$$\mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}.$$

- (b) Let $V = 2 \implies \{S \leq T\} \in \mathcal{F}_{\min\{S,T\}}$ and by symmetry $\{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}$. This means that

$$\{S = T\} = \{S \leq T\} \cap \{T \leq S\} \in \mathcal{F}_{\min\{S,T\}}.$$

Furthermore

$$\{S < T\} = \{S \leq T\} \setminus \{S = T\}$$

Which implies membership of $\mathcal{F}_{\min\{S,T\}}$.

- (c) Consider the fact already proven that $V \mathbb{1}_{\{S \leq T\}} \in \mathcal{F}_{\min\{S,T\}}$. Multiplying by $\mathbb{1}_{\{S=T\}}$ or $\mathbb{1}_{\{S < T\}}$ does not alter the membership in $\mathcal{F}_{\min\{S,T\}}$. □

4.2.19 Functions of stopping times that are (or are not) stopping times

Same, I think she means this: are we sure that $S \wedge T$ and $S \vee T$ are stopping times?

Proof

We know that $\{S \leq t\}$ and $\{T \leq t\}$ belong to \mathcal{F}_t for every t , since they are stopping times. Consider now $\{\min\{(S, T) \leq t\}\}$ and $\{\max\{(S, T) \leq t\}\}$. We have

$$\{\min\{(S, T) \leq t\}\} = \underbrace{\{S \leq t\}}_{\in \mathcal{F}_t} \cup \underbrace{\{T \leq t\}}_{\in \mathcal{F}_t} \underbrace{\quad}_{\in \mathcal{F}_t}$$

and

$$\{\max\{(S, T) \leq t\}\} = \underbrace{\{S \leq t\}}_{\in \mathcal{F}_t} \cap \underbrace{\{T \leq t\}}_{\in \mathcal{F}_t} \underbrace{\quad}_{\in \mathcal{F}_t}$$

4.2.20 Foretold times and their relationship with stopping times; examples of foretold times.

Definition 4.2.11

Let

S be a stopping time

T be a random time $T > S$ but whose value can be told by time S .

We say that T is **foretold by S** (and it is a stopping time).

Example 4.2.2

Consider S and $T = 2S$. Surprise surprise, T is a foretold time.

Remark

At time T we know S : We know that when the alarm rings within three seconds the machinery will stop, so S is actually \mathcal{F}_T -measurable. We say that **S is foretold by T** .

4.2.21 Examples of Martingales: Markov Chains

Consider a stochastic process $\{X_n\}_{n \geq 1}$. All the possible outcomes are on the same probability space so $(E_n, \mathcal{E}_n) = (E, \mathcal{E})$. Define a Markov kernel on (E, \mathcal{E}) such that $(E, \mathcal{E}) \mapsto (E, \mathcal{E})$ and such that

$$K_n(x_0, x_1, \dots, x_n; A) = \mathbb{P}(x_n, A) \quad \forall n \in \mathbb{N}, \forall (x_0, x_1, \dots, x_n) \in E, A \in \mathcal{E}.$$

The probability that governs the evolution of the process only depends on the last state. If $x_{n+1} = j$ and $x_n = i$ then $\mathbb{P}(i, j) = p_{ij}$ and the process $X = \{X_n\}_{n \geq 0}$ is said to be a Markov chain over $(\Omega, \mathcal{H}, \mathbb{P})$ with state spaces (E, \mathcal{E}) , initial distribution μ and transition kernel P where

$$P = (p_{ij})_{i,j}$$

Now consider a Markov chain with transition probability matrix P such that

$$P \cdot \underbrace{f}_{\text{eigenvector}} = \underbrace{\lambda}_{\text{eigenvalue}} \cdot f.$$

This means that we have the following system of equations:

$$\begin{cases} \sum_j p_{ij}f(j) = \lambda f(i) & i = 1 \\ \vdots & i = 2 \\ \vdots & \ddots \end{cases}$$

We can clearly write these expressions in form of expectation:

$$\mathbb{E}[f(X_{n+1})|X_n = i] = \lambda f(i).$$

We can also write it without fixing i :

$$\mathbb{E}[f(X_{n+1})|X_n] = \lambda f(X_n). \quad (\bullet)$$

Look at the last equation: it is not too different from the martingale property. Consider the ratio $\frac{f(X_n)}{\lambda^n}$ and the filtration generated by X $\sigma(X_0, X_1, \dots, X_n)$. Thanks to the relation given by the equation (\bullet) it is a martingale with respect to $\sigma(X_0, X_1, \dots, X_n)$.

Proof

Consider the expectation

$$\begin{aligned}\mathbb{E} \left[\frac{f(X_{n+1})}{\lambda^{n+1}} \middle| X_0, \dots, X_n \right] &= \frac{1}{\lambda^n} \frac{1}{\lambda} \underbrace{\mathbb{E} [f(X_{n+1}|X_n)]}_{\lambda f(X_n)} \\ &= \frac{1}{\lambda^n} \frac{1}{\lambda} \chi f(X_n) \\ &= \frac{1}{\lambda^n} f(X_n) \\ &= Y_n\end{aligned}$$

where the first equality is thanks to Markov property. This is a general property for Markov chains.
An example of Markov process is the Branching process. \square

4.2.22 Definition of martingale, sub-martingale and super-martingale and their properties

Definition 4.2.12

A real-valued process

$$X = (X_t)_{t \in \mathbb{T}}$$

is called a **\mathcal{F} -martingale** if:

1. it is adapted to \mathcal{F} ;
2. it is integrable for each $t \in \mathbb{T}$;
3. $\mathbb{E}(X_t - X_s | \mathcal{F}_s) = 0 \quad \forall s < t$.

If $\mathbb{E}(X_t - X_s | \mathcal{F}_s) \geq 0 \quad \forall s < t$ then the process is called **\mathcal{F} -submartingale** and if $\mathbb{E}(X_t - X_s | \mathcal{F}_s) \leq 0 \quad \forall s < t$ it is called **\mathcal{F} -supermartingale**.

Now the properties:

1. Consider X and Y being \mathcal{F} -sub-martingales and $a, b \in \mathbb{R}^+$. Then

$$aX + bY$$

is a \mathcal{F} -sub-martingale.

2. Consider X, Y being two \mathcal{F} -sub-martingales. Then

$$\max\{X, Y\}$$

is a \mathcal{F} -sub-martingale.

Proof

Consider $\mathbb{E}[\max\{X_n, Y_n\} | \mathcal{F}_{n-1}]$. Remember that we have to prove:

- measurability property;
- integrability property;
- martingale property.

Often times measurability and integrability are implied in the form of the function. Since we are using the maximum between X and Y we know that

$$\begin{aligned}\mathbb{E}[\max\{X_n, Y_n | \mathcal{F}_{n-1}\}] &\geq \mathbb{E}[X_n | \mathcal{F}_{n-1}] \\ \mathbb{E}[\max\{X_n, Y_n | \mathcal{F}_{n-1}\}] &\geq \mathbb{E}[Y_n | \mathcal{F}_{n-1}].\end{aligned}$$

But since X and Y are martingales we know that $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq X_{n-1}$ and $\mathbb{E}[Y_n|\mathcal{F}_{n-1}] \geq Y_{n-1}$; hence we get that

$$\mathbb{E}[\max\{X_n, Y_n|\mathcal{F}_{n-1}\}] \geq \max\{X_{n-1}, Y_{n-1}\}$$

so the martingale property is satisfied. \square

3. Consider two \mathcal{F} -super-martingales X_n, Y_n . Then

$$\min\{X_n, Y_n\}$$

is a \mathcal{F} -super-martingale.

Proof

Consider $\max\{-X_n, -Y_n\}$. We know that $-X_n$ is a sub-martingale and so is $-Y_n$. So as proved above $\max\{-X_n, -Y_n\}$ is a sub-martingale. But $\max\{-X_n, -Y_n\} = -\min\{X_n, Y_n\}$ so

$$-\min\{X_n, Y_n\} \text{ is a sub-martingale} \iff \min\{X_n, Y_n\} \text{ is a super-martingale.}$$

\square

4. Consider a function f that is convex on \mathbb{R} . If X is a \mathcal{F} -martingale and $f \circ X$ is integrable then $f \circ X$ is a \mathcal{F} -sub-martingale.

Proof

We have $s < t$ and we want to study $\mathbb{E}[f(X_t)|\mathcal{F}_s]$. By Jensen's inequality we have

$$\begin{aligned} \mathbb{E}[f(X_t)|\mathcal{F}_s] &\geq f\left[\underbrace{\mathbb{E}(X_t|\mathcal{F}_s)}_{X_s}\right] \\ &= f(X_s). \end{aligned}$$

\square

Remark

Some examples of positive functions of X are $X^+, X^-, |X|$. If X is martingale, these are sub-martingales. $|X|^p$ is a sub-martingale if X is a martingale and $\mathbb{E}|X|^p < \infty$.

5. if f is convex and increasing and X is a \mathcal{F} -sub-martingale with $f(X_t)$ integrable $\forall t$ then $f(X_t)$ is again a \mathcal{F} -sub-martingale.

4.2.23 Definition of sub/super martingale and examples

Definition 4.2.13

A real-valued process

$$X = (X_t)_{t \in \mathbb{T}}$$

is called a **\mathcal{F} -martingale** if:

1. it is adapted to \mathcal{F} ;
2. it is integrable for each $t \in \mathbb{T}$;
3. $\mathbb{E}(X_t - X_s|\mathcal{F}_s) = 0 \quad \forall s < t$.

If $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \geq 0 \quad \forall s < t$ then the process is called **\mathcal{F} -submartingale** and if $\mathbb{E}(X_t - X_s|\mathcal{F}_s) \leq 0 \quad \forall s < t$ it is called **\mathcal{F} -supermartingale**.

1. Consider X and Y being \mathcal{F} -sub-martingales and $a, b \in \mathbb{R}^+$. Then

$$aX + bY$$

is a \mathcal{F} -sub-martingale.

2. Consider X, Y being two \mathcal{F} -sub-martingales. Then

$$\max\{X, Y\}$$

is a \mathcal{F} -sub-martingale.

3. Consider two \mathcal{F} -super-martingales X_n, Y_n . Then

$$\min\{X_n, Y_n\}$$

is a \mathcal{F} -super-martingale.

4. Consider a function f that is convex on \mathbb{R} . If X is a \mathcal{F} -martingale and $f \circ X$ is integrable then $f \circ X$ is a \mathcal{F} -sub-martingale.

Remark

Some examples of positive functions of X are $X^+, X^-, |X|$. If X is martingale, these are sub-martingales. $|X|^p$ is a sub-martingale if X is a martingale and $\mathbb{E}|X|^p < \infty$.

5. if f is convex and increasing and X is a \mathcal{F} -sub-martingale with $f(X_t)$ integrable $\forall t$ then $f(X_t)$ is again a \mathcal{F} -sub-martingale.

4.2.24 Maximum, minimum of martingales

Not so sure what it is being asked here.

Example 4.2.3

We are still with our financebros. Suppose that (M_n) is the price of an asset at time n . We want to buy when the price is below a at time S_i and sell when it is above b at time T_i . In $(0, n]$ we have $U_n(a, b)$ cycles of buying and selling so our strategy could consists in holding a number F_n of shares during period $(m - 1, m]$. This means introducing

$$F_m = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}$$

with $F_0 = 0$. We can thus trace the evolution of our capital with the discrete-time integral:

$$X = \int F \, dM$$

And the profit during $(0, n]$ will be $X - X_0$.

The profit is at least

$$(b - a)U_n(a, b).$$

Proposition 4.2.8

If M is a sub-martingale with respect to its natural filtration then

$$(b - a)\mathbb{E}U_n(a, b) \leq \mathbb{E}[(M_n - a)^+ - (M_0 - a)^+].$$

We wanted to find a bound for our profit, but our profit is a stochastic quantity: so it's only natural to think about the expectation to give a bound to the number of expected up/downcrossing. Observe that the number of upcrossings does not depend on the value of T_0 that we fix.

Proof

Choose $a = 0$. Consider hence the process $(M - a)^+$ that is sub-martingale (if M is a sub-martingale). Take $M \geq 0$ and let

$$F_n = \sum_{k=1}^{\infty} \mathbb{1}_{(S_k, T_k]}(n)$$

and consider

$$X = \int F \, dM$$

like in our example. We know that F is predictable since by definition $F_{k+1} \in \mathcal{F}_k$. Consider the expectation of the increment

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] = \mathbb{E}[F_{k+1}(M_{k+1} - M_k) | \mathcal{F}_k].$$

But since F_{k+1} is predictable we can take it out the expectation:

$$F_{k+1} \mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

But since F_{k+1} is an indicator we know it is ≤ 1 :

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] \leq \mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

Now take the expectation of both sides:

$$\mathbb{E}[X_{k+1} - X_k] \leq \mathbb{E}[M_{k+1} - M_k].$$

If we sum these inequalities over k we get:

$$\mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0].$$

So we now get that

$$bU_n(a, n) \leq \mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0]$$

but given that $a = 0$ we get that

$$bU_n(0, b) \leq \mathbb{E}[M_n - M_0].$$

Clearly we have to take the positive part. □

This characterizes our martingale and its boundedness. The number of oscillations of a sub-martingale is bounded! The next question is: can we say anything about the behaviour of maximum/minimum of a martingale or sub-martingale?

Remark

Consider a sequence $\{X_n\}$ of independent random variables with $\mathbb{E}X_n = 0$, $S_n = \sum X_i$. We proved that

$$a^2 \mathbb{P}(\max_{k \leq n} |S_k| > a) \leq \text{Var } S_n$$

and we called this **Kolmogorov's inequality**.

What we are doing here is considering a random walk whose jumps have 0 mean. We wonder whether it is above the level a as seen in figure 4.7. About this, we know that $\{S_n\}$ is a \mathcal{F} -martingale but in proving the Kolmogorov's inequality we never talked about the martingale property! We could improve this inequality using the **Doob's martingale inequality**. The problem is that we can prove very general inequalities that hold true for any random variables but specifying more characteristic we can obtain stricter bounds. In this framework let's define

$$\begin{aligned} M_n^* &= \max_{k \leq n} M_k \\ m_n^* &= \min_{k \leq n} M_k \end{aligned}$$

as current maximum and current minimum of M .

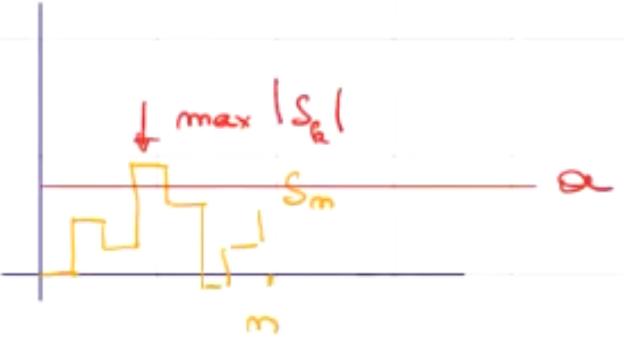


Figure 4.7: The maximum of the random walk.

Theorem 4.2.15

Take M as a sub-martingale. For $b > 0$ it holds:

1. $b\mathbb{P}(M_n^* \geq b) \leq \mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq b\}}] \leq \mathbb{E}[M_n^*];$
2. $b\mathbb{P}(M_n^* \leq -b) \leq -\mathbb{E}M_0 + \mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq b\}}] \leq \mathbb{E}M_n^* - \mathbb{E}M_0.$

So we can further bound the result looking into the property of M_n .

Example 4.2.4

Now need to define the brownian motion (or Weiner process).

Definition 4.2.14

A real-valued stochastic process $B = (B_t)_{t \geq 0}$ is called **brownian motion** if:

1. the index set is \mathbb{R}^+ ;
2. $B_0(\omega) = 0$ for almost all ω ;
3. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent for $\forall 0 = t_0 < t_1 < \dots < t_n < \infty$;
4. $B_t - B_s \sim B_{t+b} - B_{s+b}$ for every $0 \leq s < t < \infty \quad \forall n > -s$;
5. $B_t - B_s \sim N(0, t - s)$;
6. $t \mapsto B_t(\omega)$ are continuous for every ω .

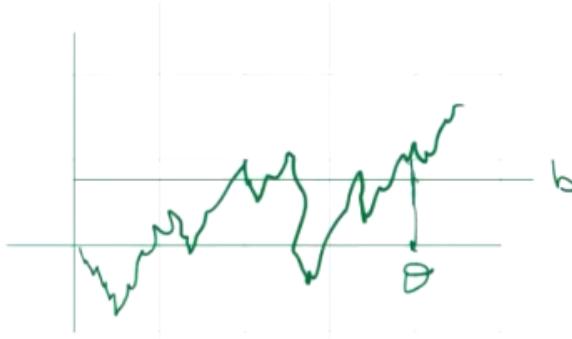
The brownian motion is in itself a martingale, but there is a class of martingales strictly related to it:

$$M_t = \exp \left\{ \lambda B_t - \frac{1}{2} \lambda^2 t \right\}, \quad t \in \mathbb{R}^+.$$

Now we can get to the actual example: if B is a brownian motion, we have

$$\mathbb{P}(\sup_{0 \leq t \leq \theta} B_t \geq b) \leq \exp \left\{ -\frac{b^2}{s\theta} \right\}.$$

The sample paths of brownian motions are extremely irregular. We are asking with which probability the max of our process will be over b at time θ .



I am basically asking if the maximum attained is above or below b . How can we use the Doob's inequality?

Proof

Consider

$$\mathbb{P}(\sup B_t \geq b) = \mathbb{P}(\sup e^{\lambda B_t} \geq e^{\lambda b})$$

and using Doob's inequality

$$\begin{aligned} \mathbb{P}(\sup B_t \geq b) &= \mathbb{P}(\sup e^{\lambda B_t} \geq e^{\lambda b}) \\ &\leq \underbrace{\mathbb{E}\left[e^{\lambda B_\theta - \frac{\lambda^2}{2}\theta}\right]}_{\text{martingale}} \\ &\leq e^{\lambda b} e^{-\frac{\lambda^2}{2}\theta} \\ &= e^{-\lambda b + \frac{\lambda^2}{2}\theta} \quad \forall \lambda > 0 \end{aligned}$$

□

4.2.25 Predictable processes and Doob's decomposition

Definition 4.2.15

The process $F = (F_n)_{n \geq 1}$ is **\mathcal{F} -predictable** if $F_0 \in \mathcal{F}_0$ and $\mathcal{F}_{n+1} \in \mathcal{F}_n$, for $\forall n \in \mathbb{N}$.

This means that the available information up to n is “enough” to have a bet in the period $n+1$. Some predictable processes are:

- any deterministic processes;
- consider two stopping times S, T of \mathcal{F} and let $S \leq T$. Consider the random variable V in \mathcal{F}_n . Then

$$V\mathbf{1}_{(S,T]} \quad V\mathbf{1}_{(S,\infty]} \quad \mathbf{1}_{(S,T]} \quad \mathbf{1}_{(0,T]}$$

are predictable processes.

Proof

(2) Consider $F = V\mathbf{1}_{(S,\infty]}$, so that $F_n = \mathbf{1}_{(S,\infty]}(n)$. Consider

$$\begin{aligned} F_{n+1} &= V\mathbf{1}_{(S,\infty]}(n+1) \\ &= V\mathbf{1}_{\{S < n+1\}} \\ &= V\mathbf{1}_{\{S \leq n\}} \in \mathcal{F}_n \end{aligned}$$

so the process is \mathcal{F} -measurable and predictable.

(1) Since we know by hypothesis that $S \leq T$ then $V \in \mathcal{F}_S \subset \mathcal{F}_T$. This means that $V \in \mathcal{F}_T$. Hence, of a consequence,

$$V\mathbb{1}_{(T,\infty]} - V\mathbb{1}_{(S,\infty]} = V\mathbb{1}_{(S,T]}$$

is predictable.

(3) Take $V = 1$.

(4) Take $T = \infty, V = 1$. $\mathbb{1}_{(S,\infty]}$ is predictable. But then

$$\mathbb{1}_{[0,S]} = \mathbb{1} - \mathbb{1}_{(S,\infty]}$$

so it is predictable. \square

Theorem 4.2.16

Doob's decomposition.

X is a stochastic process which is adapted to \mathcal{F} and integrable. Then

1. it can be decomposed as

$$X_n = X_0 + M_n + A_n, \quad n \in \mathbb{N}$$

where:

- M_n is a \mathcal{F} -martingale with $M_0 = 0$;
- A_n is a predictable process with $A_0 = 0$.

2. The decomposition is unique up to equivalence.

3. If X_n is a sub-martingale then $\{A_n\}_{n \geq 0}$ increasing, while if X_n is a super-martingale then $\{A_n\}_{n \geq 0}$ is decreasing.

Proof

Put $A_0 = M_0 = 0$. Define M and A through their increments:

$$\begin{aligned} A_{n+1} - A_n &= \underbrace{\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n]}_{\mathcal{F}_n\text{-meas.}} \\ M_{n+1} - M_n &= (X_{n+1} - X_n) - (A_{n+1} - A_n). \end{aligned}$$

If we look at these quantities we see that A is predictable and M is martingale. Imagine now there is another decomposition: let

$$X = X_0 + M' + A'$$

be another decomposition. We must have

$$\cancel{X_0} + M' + A' = \cancel{X_0} + M + A \iff A - A' = M - M' = B.$$

Now B is a process and it is predictable and martingale (because it is the difference between two martingales). Since B is predictable and a martingale we have

$$\begin{aligned} B_{n+1} - B_n &= \underbrace{\mathbb{E}[B_{n+1} - B_n | \mathcal{F}_n]}_{\text{predictable}} \\ &= 0 \\ &\stackrel{\text{martingale}}{\implies} B_{n+1} = B_n = B_0 \text{ a.s., } A = A' \text{ a.s., } M = M' \text{ a.s.} \end{aligned}$$

If X is a sub-martingale then we have

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \geq 0$$

and this means that we have $A_{n+1} \geq A_n$ is increasing. \square

4.2.26 Doob's stopping theorem

For a martingale we know

$$\mathbb{E}[X_t - X_s | \mathcal{F}_s] = 0.$$

The question is: is this true also if s and t are substituted by stopping times $S, T, S \leq T$?

Theorem 4.2.17

Let M be adapted to \mathcal{F} . Then the following are equivalent:

- ① M is a submartingale;
- ② for every bounded stopping time $S \leq T$ the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0;$$

- ③ for each pair of bounded stopping times the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S] \geq 0.$$

Remark

If M is a martingale the theorem can be read in a different way:

$$\mathbb{E}M_T = \mathbb{E}M_S = \mathbb{E}M_0.$$

Previously, $\mathbb{E}M_n = \mathbb{E}M_{n-1} = \mathbb{E}M_0$.

Proof

To prove the theorem we need to show that from condition 1 follows 2 from which follows 3 from which follows 1.

1 → 2 by hypothesis M is a sub-martingale and our thesis is that if $S(\omega) < T(\omega) < n$ (because we asked for bounded times) then:

- (a) M_S and M_T are integrable;
- (b) $\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0$.

We know that S and T are bounded by n . Let V be a positive bounded random variable and define $F = V\mathbb{1}_{(S,T]}$ and use it in the discrete time integral:

$$X_n = \underbrace{M_0 F_0}_{X_0} + \underbrace{(M_1 - M_0)}_{V\mathbb{1}_{\{1 \in (S,T]\}}} \underbrace{F_1}_{+ \dots + (M_n - M_{n-1})F_n}$$

So we have

$$X_n - X_0 = V(M_T - M_S)$$

So X_n is a sub-martingale. Take $V = 1$ and $S = 0$: now we have

$$\begin{array}{ccc} X_n & - & X_0 \\ \downarrow & & \downarrow \\ \text{int.} & & \text{int.} \end{array} = M_T \implies M_T \text{ is integrable.}$$

Now take $V = 1$ and $T = n$ so that we get

$$\begin{array}{ccc} X_n & - & X_0 \\ \downarrow & & \downarrow \\ \text{int.} & & \text{int.} \end{array} = M_n - M_S \implies M_S \text{ is integrable.}$$

We recall that $V \in \mathcal{F}_S$ and we use the defining property for $\mathbb{E}(\cdot | \mathcal{F}_S)$. So we can write

$$\begin{aligned}\mathbb{E}\mathbb{E}(M_T - M_S | \mathcal{F}_S) &= \mathbb{E}V(M_T - M_S) \\ &\stackrel{\text{def. prop.}}{=} \mathbb{E}[X_n - X_0] \\ &\stackrel{\text{discr. time int.}}{\geq 0} \\ &\stackrel{\text{proved above}}{\geq 0}\end{aligned}$$

and this is true $\forall V > 0, V < b, V \in \mathcal{F}_S$. Hence

$$\mathbb{E}(M_T - M_S | \mathcal{F}_S) \geq 0$$

So 1 \rightarrow 2.

2 \rightarrow 3 We can use the tower rule. Take the expectation of point 2:

$$\mathbb{E}[\mathbb{E}(M_T - M_S | \mathcal{F}_S)] = \mathbb{E}[M_T - M_S] \geq 0.$$

3 \rightarrow 1 Let 3 hold, so that $\mathbb{E}[M_T - M_S] \geq 0$. Choose $T = n$ and $S = 0$. Then M_n is integrable. Move to adaptness: this holds by hypothesis. Move to the martingale inequality:

$$\mathbb{E}[M_n - M_m | \mathcal{F}_m] \geq 0.$$

Note that this is equivalent to prove

$$\mathbb{E}1_H \mathbb{E}[M_n - M_m | \mathcal{F}_m] \geq 0 \quad H \in \mathcal{F}_m, 0 \leq m \leq n.$$

Fix H, m, n and define

$$S(\omega) = m \quad T(\omega) = n1_H(\omega) + m1_{\{\Omega \setminus H\}}(\omega)$$

The indicators are non-zero in complementary instances. Notice that:

- (a) S is a fixed time so it is a stopping time;
- (b) $S \leq T \leq n$ by definition of S and T because the indicators are non-zero in complementary instances;
- (c) $T \geq S$ is a foretold time by $S = m$;
- (d) $H \in \mathcal{F}_S$ by definition.

So we can write $M_T - M_S = 1_H(M_n - M_m)$ where $M_T - M_S \geq 0$ by hypothesis. This means that we have

$$\underbrace{\mathbb{E}[1_H \mathbb{E}(M_n - M_m | \mathcal{F}_m)]}_{\mathbb{E}[M_n - M_m | \mathcal{F}_m] \geq 0} \geq 0$$

□

4.2.27 Upcrossing inequality

I believe that what is being asked is

Proposition 4.2.9

If M is a sub-martingale with respect to its natural filtration then

$$(b - a)\mathbb{E}U_n(a, b) \leq \mathbb{E}[(M_n - a)^+ - (M_0 - a)^+].$$

We wanted to find a bound for our profit, but our profit is a stochastic quantity: so it's only natural to think about the expectation to give a bound to the number of expected up/downcrossing. Observe that the number of upcrossings does not depend on the value of T_0 that we fix.

Proof

Choose $a = 0$. Consider hence the process $(M - a)^+$ that is sub-martingale (if M is a sub-martingale). Take $M \geq 0$ and let

$$F_n = \mathbb{1}_{(S_k, T_k]}(n)$$

and consider

$$X = \int F dM$$

like in our example. We know that F is predictable since by definition $F_{k+1} \in \mathcal{F}_k$. Consider the expectation of the increment

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] = \mathbb{E}[F_{k+1}(M_{k+1} - M_k) | \mathcal{F}_k].$$

But since F_{k+1} is predictable we can take it out the expectation:

$$F_{k+1} \mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

But since F_{k+1} is an indicator we know it is ≤ 1 :

$$\mathbb{E}[X_{k+1} - X_k | \mathcal{F}_k] \leq \mathbb{E}[M_{k+1} - M_k | \mathcal{F}_k].$$

Now take the expectation of both sides:

$$\mathbb{E}[X_{k+1} - X_k] \leq \mathbb{E}[M_{k+1} - M_k].$$

If we sum these inequalities over k we get:

$$\mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0].$$

So we now get that

$$bU_n(a, n) \leq \mathbb{E}[X_n - X_0] \leq \mathbb{E}[M_n - M_0]$$

but given that $a = 0$ we get that

$$bU_n(0, b) \leq \mathbb{E}[M_n - M_0].$$

Clearly we have to take the positive part. □

But I also found this

Proposition 4.2.10

Suppose that $\{X = X_t : t \in [0, \infty)\}$ satisfies the basic assumptions with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \in [0, \infty)\}$ and let $a, b \in \mathbb{R}$ with $a < b$. Let $U_t = u_t(a, b, X)$ the random number of upcrossings of $[a, b]$ by X up to time $t \in [0, \infty)$.

① if X is a super-martingale relative to \mathcal{F} then

$$\mathbb{E}(U_t) \leq \frac{1}{b-a} \mathbb{E}[(X_t - a)^-] \leq \frac{1}{b-a} [\mathbb{E}(X_t^-) + |a|] \leq \frac{1}{b-a} [\mathbb{E}(|X_t|) + |a|];$$

② if X is a sub-martingale relative to \mathcal{F} then

$$\mathbb{E}(U_t) \leq \frac{1}{b-a} \mathbb{E}[(X_t - a)^+] \leq \frac{1}{b-a} [\mathbb{E}(X_t^+) + |a|] \leq \frac{1}{b-a} [\mathbb{E}(|X_t|) + |a|];$$

4.2.28 Doob's decomposition

Theorem 4.2.18

Doob's decomposition.

X is a stochastic process which is adapted to \mathcal{F} and integrable. Then

- it can be decomposed as

$$X_n = X_0 + M_n + A_n, \quad n \in \mathbb{N}$$

where:

- M_n is a \mathcal{F} -martingale with $M_0 = 0$;
- A_n is a predictable process with $A_0 = 0$.

- The decomposition is unique up to equivalence.
- If X_n is a sub-martingale then $\{A_n\}_{n \geq 0}$ increasing, while if X_n is a super-martingale then $\{A_n\}_{n \geq 0}$ is decreasing.

Proof

Put $A_0 = M_0 = 0$. Define M and A through their increments:

$$\begin{aligned} A_{n+1} - A_n &= \mathbb{E} \left[X_{n+1} - X_n | \mathcal{F}_n \right] \\ M_{n+1} - M_n &= (X_{n+1} - X_n) - (A_{n+1} - A_n). \end{aligned}$$

\mathcal{F}_n -meas.

If we look at these quantities we see that A is predictable and M is martingale. Imagine now there is another decomposition: let

$$X = X_0 + M' + A'$$

be another decomposition. We must have

$$X_0 + M' + A' = X_0 + M + A \iff A - A' = M - M' = B.$$

Now B is a process and it is predictable and martingale (because it is the difference between two martingales). Since B is predictable and a martingale we have

$$\begin{aligned} B_{n+1} - B_n &= \mathbb{E}[B_{n+1} - B_n | \mathcal{F}_n] \\ &= 0 \\ \implies B_{n+1} &= B_n = B_0 \text{ a.s., } A = A' \text{ a.s., } M = M' \text{ a.s.} \end{aligned}$$

If X is a sub-martingale then we have

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}_n] \geq 0$$

and this means that we have $A_{n+1} \geq A_n$ is increasing. □

4.2.29 Stochastic integral in discrete time

Let us consider two real-valued processes $M = (M_n)_n$ and $F = (F_n)_n$ and let us define

$$X_n = F_0 M_0 + (M_1 - M_0) F_1 + \dots + (M_n - M_{n-1}) F_n.$$

We say that $\{X_n\}$ is the integral of F with respect to M and we write

$$X_n = \int F \, dM$$

where dM is a random signed measure. Remember the Lebesgue-Stieltjes integral? Me neither, but as long as M has bounded variation this is a Lebesgue-Stieltjes integral. So a little explanation is due since I actually

never saw a Stieltjes integral.

Theorem 4.2.19

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 4.2.5

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n-1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^n (M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k , but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \dots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \dots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to \mathcal{F} .
2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b . This implies

$$|X_n| < b(|M_0| + |M_1 - M_0| + \dots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

4.2.30 Stochastic integral and its application

Let us consider two real-valued processes $M = (M_n)_n$ and $F = (F_n)_n$ and let us define

$$X_n = F_0 M_0 + (M_1 - M_0)F_1 + \dots + (M_n - M_{n-1})F_n.$$

We say that $\{X_n\}$ is the integral of F with respect to M and we write

$$X_n = \int F \, dM$$

where dM is a random signed measure. Remember the Lebesgue-Stieltjes integral? Me neither, but as long as M has bounded variation this is a Lebesgue-Stieltjes integral. So a little explanation is due since I actually never saw a Stieltjes integral.

Theorem 4.2.20

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 4.2.6

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n-1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^n (M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k , but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \dots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \dots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to \mathcal{F} .
2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b . This implies

$$|X_n| < b(|M_0| + |M_1 + M_0| + \dots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$\underbrace{F_n \mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

and for the applications:

Definition 4.2.16

Define $M = (M_n)_{n \in \mathbb{N}}$ as a process and let T be a random time with values on $\overline{\mathbb{N}}$. The process

$$X_n(\omega) = M_{n \wedge T}(\omega) = \begin{cases} M_n(\omega) & n < T(\omega) \\ M_T(\omega) & n > T(\omega) \end{cases}$$

(where $n \wedge T$ is a truncated random time) is called **M stopped at T** .

As a consequence X is exactly the discrete time integral if $F = \mathbb{1}_{[0,T]}$:

$$X_n = \underbrace{M_0 F_0}_0 + (M_1 - M_0) \mathbb{1}_{[0,T]}(1) + \dots + (M_n - M_{n-1}) \mathbb{1}_{[0,T]}(n).$$

The indicators only select the current time interval. If this is the case we can observe that $F_{[0,T]}$ is bounded, positive and predictable. Hence if M is a martingale the theorem applies with this special choice of M and we can write the result as a different theorem:

Theorem 4.2.21

Let T be a stopping time and let X be the process M stopped at T . If M is a martingale then so is X (the same holds for sub-martingales and super-martingales).

So we cannot determine a policy based on stopping times that can change the nature of our martingale. In the remote case in which you are interested in this you can read Williams - Introduction to martingales.

4.2.31 Impossibility to win against the system: related theorems and examples

Theorem 4.2.22

Consider F bounded and predictable. Then if M is a martingale then X is a martingale; If M is a sub(super)-martingale then X is a sub(super)-martingale.

This means that... **we can't beat the system!**

I can beat something else though.

Example 4.2.7

Consider M_n as the price of a share at time n and F_n as the number of shares owned during $(n-1, n]$. Our profit will be

$$(M_n - M_{n-1})F_n$$

and our total profit X_n gained during $(0, n]$ will be:

$$X_n = X_0 + \underbrace{\sum_{k=1}^n (M_k - M_{k-1})F_k}_{\text{discrete time integral}}$$

F_n is based on the knowledge in $n-1$ so it is predictable. The process M_n should be a martingale (otherwise if it is a sub/super-martingale everyone/no one will buy). So the total profit will also be a martingale! We can only choose our buying politics F_k , but there is no way to select a politics that will change a martingale in a super-martingale or sub-martingale.

Clearly this works in mean!

Proof

1. We have M being a martingale and $F_0, F_1, \dots, F_n \in \mathcal{F}_n$ as well as $M_0, M_1, \dots, M_n \in \mathcal{F}_n$. Therefore $X_n \in \mathcal{F}_n$ and X is adapted to \mathcal{F} .

2. We need to check whether the discrete time integral is a martingale. We know by hypothesis that F is bounded, so $F < b$ for some b . This implies

$$|X_n| < b(|M_0| + |M_1 - M_0| + \dots + |M_n - M_{n-1}|)$$

Since M is a martingale and it is integrable, we get that X_n is bounded and integrable.

3. Consider

$$\mathbb{E}[X_{n+1} - X_n | \mathcal{F}] = \mathbb{E}[F_{n+1}(M_{n+1} - M_n) | \mathcal{F}_n]$$

since all the terms cancel out and only the last ones survive. But $F_{n+1} \in \mathcal{F}_n$ so we can take it out of the expectation:

$$F_n \underbrace{\mathbb{E}[M_{n+1} - M_n | \mathcal{F}_n]}_{=0} = 0.$$

□

We know that using a policy that it is predictable it is impossible to beat the system. Finance bros try to overcome this possibility using stopping times. If my policy is not only based on a predictable process but I add the randomness of the time in which I decide to sell or buy can I break the curse of martingales and make shareholders want to suck my dick? Well, it depends.

4.2.32 Predictable and adapted processes: definition and examples

Definition 4.2.17

The process $F = (F_n)_{n \geq 1}$ is **\mathcal{F} -predictable** if $F_0 \in \mathcal{F}_0$ and $\mathcal{F}_{n+1} \in \mathcal{F}_n$, for $\forall n \in \mathbb{N}$.

This means that the available information up to n is “enough” to have a bet in the period $n + 1$. Some predictable processes are:

- any deterministic processes;
- consider two stopping times S, T of \mathcal{F} and let $S \leq T$. Consider the random variable V in \mathcal{F}_n . Then

$$\begin{array}{cccc} (1) & (2) & (3) & (4) \\ V\mathbf{1}_{(S,T]} & V\mathbf{1}_{(S,\infty]} & \mathbf{1}_{(S,T]} & \mathbf{1}_{(0,T]} \end{array}$$

are predictable processes.

Proof

(2) Consider $F = V\mathbf{1}_{(S,\infty]}$, so that $F_n = \mathbf{1}_{(S,\infty]}(n)$. Consider

$$\begin{aligned} F_{n+1} &= V\mathbf{1}_{(S,\infty]}(n+1) \\ &= V\mathbf{1}_{\{S < n+1\}} \\ &= V\mathbf{1}_{\{S \leq n\}} \in \mathcal{F}_n \end{aligned}$$

so the process is \mathcal{F} -measurable and predictable.

(1) Since we know by hypothesis that $S \leq T$ then $V \in \mathcal{F}_S \subset \mathcal{F}_T$. This means that $V \in \mathcal{F}_T$. Hence, of a consequence,

$$V\mathbf{1}_{(T,\infty]} - V\mathbf{1}_{(S,\infty]} = V\mathbf{1}_{(S,T]}$$

is predictable.

(3) Take $V = 1$.

(4) Take $T = \infty$, $V = 1$. $\mathbb{1}_{(S, \infty]}$ is predictable. But then

$$\mathbb{1}_{[0, S]} = \mathbb{1} - \mathbb{1}_{(S, \infty]}$$

so it is predictable. \square

Example of adapted process:

Example 4.2.8

This is called “the secretary problem”: in this case we must start from the filtration and then understand the problem. Here I have N candidates for a position; a candidate disregarded after the interview is lost. The interviewer wants to hire exactly 1 candidate and each candidate has different abilities and the interviewer knows only the relative ability of those already interviewed so far. Our goal is to maximize the probability of hiring the best one. We have three questions:

1. what is Ω ?
2. what is the filtration \mathcal{F} for this experiment?
3. what process should we use?

In this case $\Omega = N!$ permutations of the ranking of the candidates (the order in which they show up) and the filtration is the information earned from interview up to time t (that is the ranking of the candidates up to time t). But what is the process that I should use? Consider the sequence

$$V_1, V_2, \dots, \{V_i\}_{i \geq 1}$$

with $V_i = 1$ if and only if the best candidate is the i -th candidate and $V_i = 0$ otherwise. Could this process $\{V_i\}_{i \geq 1}$ be used? No, because V is not adapted to \mathcal{F} ... because to understand if i -th candidate is the best we need to compare it to the other candidates, including the ones that didn't show up yet! But then how can we get an adapted process? Let us consider the expectation

$$U_n = \mathbb{E}[V_n | \mathcal{F}_n]$$

What do we know about the measurability of U_n ? We know that it is for sure \mathcal{F}_n -measurable. This trick gives us a simple way to build an adapted process. So now we will have: $U_n = 0$ if the candidate is not the best up to n and $U_n = 1$ otherwise. More specifically, we will have

$$\begin{aligned} U_n &= 1 \cdot \text{probability that the best candidate is among the first } n + 0 \cdot \text{probability that the best candidate is not among the first } n \\ &= 1 \cdot \frac{n}{N} + 0 \cdot \frac{N-n}{N} \\ &= \frac{n}{N} \end{aligned}$$

This is a quantity that I can measure and it is therefore adapted.

4.2.33 Poisson process and its martingale property

Consider \mathbb{R}_+ as our index set. \mathcal{F} is our filtration and we consider the counting process $N = (N_t)_{t \geq 0}$: this counts the number of events up to time t , it has unit jumps and any path starts from 0 so that $N_0(\omega) = 0$, it is increasing and it is right continuous.

Definition 4.2.18

The counting process N is said to be a **Poisson process** with rate λ with respect to \mathcal{F} if it is adapted to \mathcal{F} and

$$\mathbb{E}[f(N_{t+s} - N_s) | \mathcal{F}_s] = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} f(k) \quad \forall s, t \in \mathbb{R}_+, \forall \text{positive } f \mapsto \mathbb{N}.$$

Theorem 4.2.23

Let N be a counting process. It is a Poisson process with rate λ with respect to \mathcal{F} if and only if:

$$M_t = N_t - \lambda t$$

is a \mathcal{F} -martingale.

We only prove that M_t is a martingale if N_t is Poisson.

Proof

We know that

$$\begin{aligned} \mathbb{E}[M_t | \mathcal{F}_s] &= \mathbb{E}[M_t - M_s + M_s | \mathcal{F}_s] \\ &= \mathbb{E}[M_t - M_s | \mathcal{F}_s] + M_s \\ &= \mathbb{E}[M_t - M_s] + M_s \\ &= \mathbb{E}[N_t - N_s + \lambda t + \lambda s] + M_s \\ &= \underbrace{\mathbb{E}[N_t - N_s]}_{\lambda(t-s)} - \lambda(t-s) + M_s \\ &= M_s \end{aligned}$$

□

4.2.34 Stopped processes and their properties

Example 4.2.9

Some stopping times:

- ① The first time that $X(\omega) \in H \in \Omega$;

$$T(\omega) = \begin{cases} \inf \{n \in \mathbb{N} : X_n(\omega) \in H\} \\ +\infty \end{cases} \quad \text{if } X_n(\omega) \notin H \forall n$$

So

$$\{T \leq n\} = \bigcup_{k=0}^n \{X_k \in H\}.$$

- ② Consider i.i.d. random variables X_1, X_2, \dots . Consider the probabilities

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$$

and the random walk

$$S_n = \sum_{i=1}^n X_i.$$

Let's define

$$T_1 = \begin{cases} \min \{n < 50 : S_n = 3\} \\ 50 \end{cases} \quad \text{otherwise.}$$

This is a stopping time because I can write $\{T_1 \leq n\}$ as

$$\bigcup_{k=1}^n \underbrace{\{S_k = 3\}}_{\mathcal{F}_k\text{-measurable}} \quad n < 50.$$

Moreover for $n = 50$ we have $T_1 \in \mathcal{F}_{50}$.

③ Starting from the previously defined random walk, consider the quantity

$$M_n = \min(S_1, \dots, S_n)$$

And the random time

$$T_2 = \min\{n : S_n \geq M_m + 2\}$$

is a stopping time. On the contrary,

$$T_3 = \begin{cases} \max\{n < 50 : S_n = 7\} & \text{if not empty} \\ 50 & \text{otherwise} \end{cases}$$

is not a stopping time. Why? Because I have to wait until $n = 50$ to answer the question.

Consider the random times on \mathbb{R}_+

$$0 < T_1 < T_2 < \dots$$

With $\lim_{n \rightarrow \infty} T_n = +\infty$. Define the process $\{N_t\}$ as

$$N_t := \sum \mathbf{1}_{[0,t]}(T_n).$$

This is called **counting process**. It is a basic count of the number of events happened up to time n . N_t is increasing, right continuous and increases by unitary jumps. Moreover, $N_0 = 0$, $N_t < \infty$ for $t \in \mathbb{R}_+$. Of course $\lim_{t \rightarrow \infty} N_t = \infty$. Counting processes generate their natural filtration \mathcal{F} .

Some problems require to stop the observation at a stopping time (because we don't care anymore¹⁰)... So we don't actually need the whole knowledge of the complete filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$. The problem is that as we said before the stopping time is a random time. So what we need is the *information known up to time T* \mathcal{F}_T , basically the σ -field that is the filtration at time T .

Example 4.2.10

Truncated stopping time: let T be a stopping time (for example, the time at which we sell certain shares) and that we want a finite horizon for this decision. In this case the quantity of interest is

$$S = T \wedge n = \min\{T, n\}$$

where n could be some sort of time horizon.

Imagine that two cyclists participate to a race. Their children will have their snack when both the parents will arrive to the finish line. How long will the children wait for their snack? We can think about the following stopping times:

$$\begin{aligned} T &: \text{time employed by the first cyclist} \\ S &: \text{time employed by the second cyclist} \\ U &: \max\{S, T\}. \end{aligned}$$

The waiting time for the children will be U .

4.2.35 Important inequalities for sub-martingales

The problem is that we can prove very general inequalities that hold true for any random variables but specifying more characteristic we can obtain stricter bounds. In this framework let's define

$$\begin{aligned} M_n^* &= \max_{k \leq n} M_k \\ m_n^* &= \min_{k \leq n} M_k \end{aligned}$$

as current maximum and current minimum of M .

¹⁰Assuming we ever did.

Theorem 4.2.24

Take M as a sub-martingale. For $b > 0$ it holds:

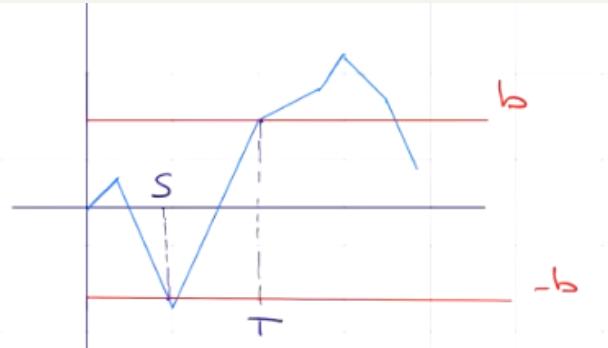
1. $b\mathbb{P}(M_n^* \geq b) \leq \mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq b\}}] \leq \mathbb{E}[M_n^*];$
2. $b\mathbb{P}(m_n^* \leq -b) \leq -EM_0 + \mathbb{E}[M_n \mathbf{1}_{\{m_n^* \leq -b\}}] \leq EM_n^+ - EM_0.$

So we can further bound the result looking into the property of M_n .

Proof

We introduce 2 stopping times:

$$\begin{aligned} T &= \inf\{n \geq 0 : M_n \geq b\} \\ S &= \inf\{n \geq 0 : M_n \leq -b\}. \end{aligned}$$



Consider the current maximum and minimum above the level b :

$$\{M_n^* \geq b\} = \{T \leq n\} \quad \{m_n^* < -b\} = \{S \leq n\}.$$

Fix b and n . When $\{T \leq n\}$ we have

$$M_{T \wedge n} = M_T \geq b.$$

Multiply by $\mathbf{1}_{\{T \leq n\}}$:

$$b\mathbf{1}_{\{T \leq n\}} \leq M_{T \wedge n}\mathbf{1}_{\{T \leq n\}}.$$

Using Doob's stopping theorem we know that

$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0$$

and using $S = T \wedge n$ and $T = n$ we obtain

$$\mathbb{E}[M_n | \mathcal{F}_{T \wedge n}] > M_{T \wedge n}.$$

So summing it all up:

$$\begin{aligned} b\mathbf{1}_{\{T \leq n\}} &\leq M_{T \wedge n}\mathbf{1}_{\{T \leq n\}} \\ &\leq \mathbf{1}_{\{T \leq n\}} \mathbb{E}[M_n | \mathcal{F}_{T \wedge n}] \\ &= \mathbb{E}[M_n \mathbf{1}_{\{T \leq n\}} | \mathcal{F}_{T \wedge n}]. \end{aligned}$$

Now take the expectation:

$$\begin{aligned} b\mathbb{P}(T \leq n) &= b\mathbb{P}(M_n^* \geq b) \\ &\leq \underbrace{\mathbb{E}[M_n \mathbf{1}_{\{T \leq n\}}]}_{\mathbb{E}[M_n \mathbf{1}_{\{M_n^* \geq b\}}]} \\ &\leq \mathbb{E}(M_n^+). \end{aligned}$$

For the minimum we work on $\{S \leq n\}$ and we get

$$\begin{aligned} M_{S \wedge n} &= M_S \mathbb{1}_{\{S \leq n\}} + M_S \mathbb{1}_{\{S > n\}} \\ &\leq -b \mathbb{1}_{\{S \leq n\}} + M_n \mathbb{1}_{S > n}. \end{aligned}$$

Now take the expectation:

$$\mathbb{E}M_{S \wedge n} \leq -b \mathbb{P}(m_n^* \leq -b) + \mathbb{E}[M_n \mathbb{1}_{\{S > n\}}].$$

So that we get

$$\mathbb{P}(m_n^* \leq -b) \leq -\mathbb{E}M_{S \wedge n} + \mathbb{E}[M_n \mathbb{1}_{\{S > n\}}].$$

Now use Doob's stopping theorem with $T = 0$ and $S = S \wedge n$ so that $\mathbb{E}M_0 \leq \mathbb{E}M_{\{S \wedge n\}}$. This gets us our result:

$$\begin{aligned} b \mathbb{P}(m_n^* \leq -b) &\leq -\mathbb{E}M_0 + \mathbb{E}[M_n \mathbb{1}_{\{m_n^*\}}] \\ &\leq \mathbb{E}M_n^+ - \mathbb{E}M_0. \end{aligned}$$

□

4.2.36 Convergence theorems for sub-martingales

Theorem 4.2.25

Let X be a sub-martingale. If (and note that is a sufficient condition)

$$\sup_n \mathbb{E}X_n^+ < \infty$$

Then

1. $\{X_n\}$ converges a.s.;
2. $\{X_n\}$ converges to an integrable random variable.

Proof

We prove the theorem by contradiction. Pick an outcome ω and suppose that $\{X_n(\omega)\}$ is a numerical sequence that has not a limit. But if it doesn't have a limit, then

$$\exists \inf \lim \neq \sup \lim \quad \inf \lim < \sup \lim.$$

So there exist at least 2 rationals $a < b$ such that

$$\inf \lim < a < b < \sup \lim.$$

The sequence $\{X_n(\omega)\}$ crosses (a, b) ∞ many times. Now take the union over rational a and b , $a < b$ of the sets

$$\{U(a, b) = \infty\}$$

with $U(a, b) = \lim_{n \rightarrow \infty} U_n(a, b)$. Our aim is now to show that $U(a, b) \leq \infty$ almost surely to get a contradiction.

Fix a, b . We know that $U_n(a, b)$ is increasing with n . Now consider

$$\begin{aligned} (b - a) \mathbb{E}U(a, b) &= (b - a) \mathbb{E} \lim U_n(a, b) \\ &= (b - a) \lim \mathbb{E}U_n(a, b) \\ &\stackrel{\text{monotone conv.}}{=} \sup_{\text{upcross inequalities}} \mathbb{E}(X_n - a)^+ \\ &\leq \sup \mathbb{E}X_n^+ + |a|. \end{aligned}$$

So this means that $\mathbb{E}U(a, b) < \infty$. But this is a contradiction, so it exists a limit $X_n = X_\infty$ a.s..

Now consider the second part of the theorem:

$$\begin{aligned}\mathbb{E}|X_\infty| &= \mathbb{E} \liminf |X_n| \\ &\leq \liminf \mathbb{E}|X_n| \\ &\leq 2 \sup \mathbb{E} X_n^+ - \mathbb{E} X_0 < \infty\end{aligned}$$

so the limit is integrable. □

4.2.37 Uniform integrability and its consequences on convergence of martingales

We will need:

1. a collection \mathcal{K} of real random variables is said to be uniformly integrable if

$$k(b) = \sup_{X \in \mathcal{K}} \mathbb{E}|X| \mathbf{1}_{\{X>b\}} \xrightarrow[b \rightarrow \infty]{} 0.$$

2. If \mathcal{K} is dominated by an integrable random variable Z then it is uniformly integrable.
3. uniform integrability implies L^1 -boundedness but not the converse.
4. If \mathcal{K} is L^p -bounded for some $p > 1$ then it is uniformly integrable.

Lemma 4.2.2

Let Z be an integrable random variable. Then

$$\mathcal{K} = \{X : X = \mathbb{E}(Z|\mathcal{G})\}$$

for some sub- σ -algebra \mathcal{G} of \mathcal{H} is uniformly integrable.

Proposition 4.2.11

Let Z be an integrable random variable. Define

$$X_t = \mathbb{E}(Z|\mathcal{F}_t) \quad t \in \mathbb{T}.$$

This means that $\{X_t\}$ is a uniformly integrable \mathcal{F} -martingale.

Theorem 4.2.26

Let $\{X_n\}$ be a sequence of real-valued random variables. The following are equivalent:

1. it converges in L^1 ;
2. it converges in probability and it is uniformly integrable.

We can now prove the theorem about the convergence of sub-martingales.

Theorem 4.2.27

Let X be a sub-martingale. We have that X converges almost surely and in L^1 if and only if it is uniformly integrable. Moreover, if it is so, setting

$$X_\infty = \lim X_n$$

extends X to a sub-martingale

$$\bar{X} = (X_n)_{n \in \overline{\mathbb{N}}}.$$

We only prove the first part of the theorem.

Proof

Necessity. If X converges in L^1 by the theorem above it is uniformly integrable.

Sufficiency. If X is uniformly integral then it is L^1 -bounded for the property above. So our previous theorem holds and the martingale converges almost surely with X_∞ integrable. Furthermore, for the property above, it also converges in L^1 . \square

4.2.38 Features of the sample paths of a submartingale (or martingale)

Remark

If M is a martingale then $|M|^p$ is a sub-martingale for $p \geq 1$. If $M_n \in L^p \forall n$ we can apply Doob's inequality.

Corollary

If M is martingale in L^p for some $p \geq 1$ then for $b > 0$ we have that

$$b^p \mathbb{P}(\max_{k \leq n} |M_k| > b) \leq \mathbb{E}|M_n|^p.$$

There are also other bounds:

- $b \mathbb{P}(\max_{k \leq n} |M_k| > b) \leq 2\mathbb{E}M_n^+ - 3M_0$;
- **Doob's norm inequality:** if M is a martingale in L^p , $p \geq 1$ and q is the exponent conjugate to p ($\frac{1}{p} + \frac{1}{q} = 1$) then

$$\mathbb{E} \max_{k \leq n} |M_k|^p \leq q^p \mathbb{E}|M_n|^p.$$

- Consider L^2 -bounded martingales characterized by final coordinate X with $\text{Var } X = \sigma^2$ (that is I am fixing the variance of the last value I consider). We want to assess the variability of this process.

Theorem 4.2.28

Dubin & Schwartz 1998: it holds

1. $\mathbb{E} [\max_{0 \leq T \leq t} M_T] \leq \sigma$;
2. $\mathbb{E} [\max_{0 \leq T \leq t} |M_T|] \leq \sigma\sqrt{2}$.

Moreover, there exist suitable martingales for which this bound is attained and is strict.

4.2.39 Uniform integrability and its role in convergence problems

Theorem 4.2.29

A process $M = (M_n)_{n \in \mathbb{N}}$ is a uniformly integrable martingale if and only if for some integrable random variable Z

$$M_n = \mathbb{E}[Z|\mathcal{F}_n] \quad n \in \mathbb{N}. \quad (\bullet)$$

If so it converges almost surely and in L^1 to the integrable random variable

$$M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty].$$

Corollary

For every integrable random variable Z we have

$$\mathbb{E}(Z|\mathcal{F}_n) \xrightarrow{\text{a.s.}} \mathbb{E}(Z|\mathcal{F}_\infty).$$

Theorem 4.2.30

Let Z be an integrable random variable and let

$$M_n = \mathbb{E}(Z|\mathcal{F}_n)_{n \in \overline{\mathbb{N}}}.$$

For every stopping time T define

$$M_T = \mathbb{E}[Z|\mathcal{F}_T]$$

and for arbitrary stopping times S and T we get

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_{S \wedge T}.$$

This lets us rethink Doob's theorem.

Theorem 4.2.31

If S and T are arbitrary stopping times such that $S \leq T$ then

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S$$

for an uniformly integrable martingale.

The dominated convergence theorem requires adaptness.

Theorem 4.2.32

Hunt's dominated convergence theorem.

Let $\{X_n\}$ be dominated by an integrable random variable and suppose that exists

$$X_\infty = \lim X_n$$

So $(\mathbb{E}_n X_n)_n$ converges to $\mathbb{E}X_\infty$ almost surely and in L^1 .

4.2.40 Stopped martingales

Definition 4.2.19

Define $M = (M_n)_{n \in \mathbb{N}}$ as a process and let T be a random time with values on $\overline{\mathbb{N}}$. The process

$$X_n(\omega) = M_{n \wedge T}(\omega) = \begin{cases} M_n(\omega) & n < T(\omega) \\ M_T(\omega) & n > T(\omega) \end{cases}$$

(where $n \wedge T$ is a truncated random time) is called **M stopped at T** .

As a consequence X is exactly the discrete time integral if $F = \mathbb{1}_{[0,T]}$:

$$X_n = \underbrace{M_0 F_0}_0 + (M_1 + M_0) \mathbb{1}_{[0,T]}(1) + \dots + (M_n - M_{n-1}) \mathbb{1}_{[0,T]}(n).$$

The indicators only select the current time interval. If this is the case we can observe that $F_{[0,T]}$ is bounded, positive and predictable. Hence if M is a martingale the theorem applies with this special choice of M and we can write the result as a different theorem:

Theorem 4.2.33

Let T be a stopping time and let X be the process M stopped at T . If M is a martingale then so is X (the same holds for sub-martingales and super-martingales).

So we cannot determine a policy based on stopping times that can change the nature of our martingale. In the remote case in which you are interested in this you can read Williams - Introduction to martingales. A further theorem about this is **Doob's stopping theorem**. For a martingale we know

$$\mathbb{E}[X_t - X_s | \mathcal{F}_s] = 0.$$

The question is: is this true also if s and t are substituted by stopping times $S, T, S \leq T$?

Theorem 4.2.34

Let M be adapted to \mathcal{F} . Then the following are equivalent:

- ① M is a submartingale;
- ② for every bounded stopping time $S \leq T$ the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S | \mathcal{F}_S] \geq 0;$$

- ③ for each pair of bounded stopping times the random variables M_S and M_T are integrable and

$$\mathbb{E}[M_T - M_S] \geq 0.$$

Remark

If M is a martingale the theorem can be read in a different way:

$$\mathbb{E}M_T = \mathbb{E}M_S = \mathbb{E}M_0.$$

Previously, $\mathbb{E}M_n = \mathbb{E}M_{n-1} = \mathbb{E}M_0$.

That's all, folks

That's it. I really hated this last part. What really makes me angry is that this shit could even be remotely interesting if it wasn't all a slop of notions and theorems. Professor Sacerdote really takes the joy out of everything.

I'm sorry for the mistakes, there might be many. But still, this should be better than looking at the notes by Polito or Sacerdote. At least I hope so.

Anonymous 08/07/24(Wed)23:03:03 No. 16314090 ►

Guys, if you understood anything the OP said I'm afraid you'll end up like him

Figure 4.8: Is it over for me anon?

List of Figures

1.1	References to suicide are to be taken seriously	2
1.2	Imagine the steps getting smaller and smaller...	6
1.3	I'm sorry but here you're the soyjack and I'm the chad.	11
1.4	Actual conversation happened between me and Professor Lods.	13
1.5	Colonel Campbell embraces her true self	18
1.6	Andrej Andreevič Markov if he was cool.	23
1.7	And then he turned himself into a pickle, funniest shit I've ever seen.	26
1.8	Professor Polito draws circles and arrows	35
1.9	I'm so Julia	35
1.10	The first one is not a filtration, the second is for $T = \mathbb{N}^*$	38
1.11	The random walk randomly walking	43
1.12	Anon studies probability theory	60
1.13	Brat summer never ends	70
1.14	Mathematicians when	73
1.15	The name Fatou is hard as fuck, if you ask me. Apparently he teamed up with a dude with no nose, look it up.	75
1.16	Snake has had enough and honestly so do I.	79
2.1	You conditioned in the wrong hood, buddy.	86
2.2	Yeah I'm counting on ending my life very soon.	87
2.3	These are screenshots taken from the video feed. I couldn't bother. Sorry.	92
2.4	The problem of improvisation	93
2.5	It should not	97
2.6	I don't even know anymore.	104
2.7	Up and downs. Just like my consumption of cocaine and my mood respectively on an average tuesday evening.	109
2.8	The maximum of the random walk.	111
2.9	Check out Autechre. A staple in electronic experimental music.	112
2.10	Also Aphex Twin. Really, Autechre and Aphex Twin are probably the best example of the boom of experimentation in popular electronic music between 1995 and 2005.	113
2.11	It's astonishing how much a bad teacher can have you hate a subject.	117
4.1	Of course, it is never too late for some sickz tickz pickz	126
4.2	Case with $0 \leq x_1, x_2 \leq 1$	129
4.3	Case with $0 \leq x_1 \leq 1, x_2 > 1$	129
4.4	Whatever. I hate exercises	130
4.5	It's really just like a cumulative distribution "backwards"	143
4.6	Remember how a $U(0, 2^n)$ uniform distribution is made?	147
4.7	The maximum of the random walk.	196
4.8	The one and only truth	216