Stochastic processes

18 marzo 2024

Indice

Capitolo 1

Introduction

Important remark 1. Exam info:

- exam is oral: it's asked to discuss simple exercise. Exercises are simple version of exercies make in class. One is not requested to solve, only to discuss (discuss doesnt mean to be able to solve): the important information is the reaction to the exercise.
- after definitions at the exam add the interpretation

1.1 Conditional expectation

1.1.1 Definitions

Definition 1.1.1 (Conditional expectation)**.** Considered:

1. a probability space: $(\Omega, \mathscr{A}, \mathbb{P})$
2. a σ-field $\mathscr{G}$ contained in $\mathscr{A}$: $\mathscr{G} \subset \mathscr{A}$ (also called sub-σ-field $\mathscr{G}$)
3. a real random variable X, which has the mean: $\mathbb{E}\left[|X|\right] < +\infty$

A conditional expectation for X given $\mathscr{G}$ is *any* random variable $V = \mathbb{E}\left[X|\mathscr{G}\right]$ such that:

1. V has mean: $\mathbb{E}\left[|V|\right] < +\infty$
2. V is $\mathscr{G}$-measurable
3. $\mathbb{E}\left[1_A X\right] = \mathbb{E}\left[1_A V\right], \forall A \in \mathscr{G}$

Important remark 2 (Existence and uniqueness of conditional expectation)*.* It can be shown that:

- at least one V satisfying these 3 properties always *exists*,
- it is *unique* in the sense that if V_1 and V_2 are both conditional expectations (satisfying the properties), then are almost surely equal:

$$\mathbb{P}\left(V_1 \neq V_2\right) = 0$$

Important remark 3 ($\mathscr{G}$-measurability). Regarding the second requirement of the definition, what does mean that V is $\mathscr{G}$-measurable?
In general a real rv is a real function $X : \Omega \to \mathbb{R}$ which is measurable; measurable means that

$$X^{-1}(B) \in \mathscr{A}, \forall B \in \beta(\mathbb{R})$$
$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

Now V is $\mathscr{G}$ measurable means

$$V^{-1}(B) \in \mathscr{G}, \forall B \in \beta(\mathbb{R})$$

This is stronger condition rather than the previous because $\mathscr{G} \subset \mathscr{A}$. In general it could be that $X^{-1}(B) \in \mathscr{A}$ but $\notin \mathscr{G}$; thus the random variable would not be $\mathscr{G}$-measurable.

Important remark 4 (Interpretation of $\mathscr{G}$). The σ-field $\mathscr{G} \subset \mathscr{A}$ may be used to describe our *state of information*: $\forall A \in \mathscr{G}$, we already know wheter A is true or not.
So $\mathscr{G}$ is collection of event which we know if they are true or not; otherwise if $A \in \mathscr{A} \setminus \mathscr{G}$ we dont' know it.
In this interpretation *a general* X is $\mathscr{G}$-measurable means that the value of X is *known* under the information $\mathscr{G}$: if I have the information $\mathscr{G}$, I know the value taken by X.

Remark 1. Thus finally the requirement that $V = \mathbb{E}[X|\mathscr{G}]$ is g-measurable become easier to understand: it's something that is known under the information $\mathscr{G}$

Example 1.1.1. Suppose X is consumption of electricity in bologna tomorrow; this is a rv (since only tomorrow we will know). If X is $\mathscr{G}$-measurable, then if we have the information $\mathscr{G}$, we know the consumption. X being the consumption of electricity bologna tomorrow, $\mathscr{G}$ could be knowing that tomorro most people of bologna go to seaside (and not many electricity is used).
The interpretation of:

- $\mathbb{E}[X]$ is our best prediction of the random quantity X without knowing anything other information.
- $\mathbb{E}[X|\mathscr{G}]$ (so V) on the other hand is our best prediction of X under/having the information $\mathscr{G}$.

Important remark 5 (Interpretation of the third requirement). Under this interpretation, the third property in the definition of $V = \mathbb{E}[X|\mathscr{G}]$ becomes clearer: it's just the connection between X and V.
If V is our best prediction of X under $\mathscr{G}$, then they cannot be independent one of each other (they must be connected); this above is the condition connecting them.

Theorem 1.1.1 (Best prediction)**.** *If the random variable has second moment ($\mathbb{E}\left[X^2\right] < \infty$), the conditional expectation V of X given $\mathscr{G}$ is the best prediction of X, in the sense that minimizes the mean squared error:*

$$\mathbb{E}\left[(V-X)^2\right] = \min \mathbb{E}\left[(Z-X)^2\right]$$

being Z $\mathscr{G}$-measurable and $\mathbb{E}\left[Z^2\right] < +\infty$.

Remark 2 (Notation). If Y, V, Z are any random variables then we use the notation

$$\mathbb{E}\left[X|Y, V, Z, \ldots\right] = \mathbb{E}\left[X|\sigma(Y, V, Z, \ldots)\right]$$

with $\sigma(Y, V, Z, \ldots)$ the σ-field generated by Y, V, Z (least sigma field that make the random variables measurables).

Example 1.1.2. Suppose that X and Y are iid and $\mathbb{E}\left[|X|\right] < \infty$ (same for Y being iid). What can we say about $\mathbb{E}\left[X|X+Y\right]$?
eg if we know that $X + Y = 35$ what is our prediction for X? A reasonable answer is

$$\mathbb{E}\left[X|X+Y\right] = \frac{X+Y}{2}$$

That is, if we know the sum, a way to have the first component is to divide by 2, being X, Y iid.
In order to show that our guess $V = (X+Y)/2$ is actually the right answer, we have to show it satisfies the three properties of conditional expectation:

1. V do has the mean/expected value: X and Y have the mean so their means as well does;
2. V is $\mathscr{G}$-measurable: under the information $\mathscr{G}$ its know: if we know the value of $X+Y$ (eg say $X+Y = 35$) then we known the value of $V = 35/2$;
3. the third point is harder to show but it holds as well.

Important remark 6 (Modo di procedere). If we our interested in $\mathbb{E}\left[X|\mathscr{G}\right]$:

1. we set our conjecture is $\mathbb{E}\left[X|\mathscr{G}\right] = V$ (that can be anything);
2. we verify that the conjecture is true so that the three properties of conditional expectation are satisfied

1.1.2 Properties

Remark 3. We state the main properties of conditional expectation $V = \mathbb{E}\left[X|\mathscr{G}\right]$; the first three are straightforward, the remaining less obvious.

Important remark 7. We have

1. linearity: given X, Y random variables with mean ($\mathbb{E}\left[|X|\right] < \infty, \mathbb{E}\left[|Y|\right] < \infty$), and $a, b \in \mathbb{R}$ we have
$$\mathbb{E}\left[aX + bY|\mathscr{G}\right] = a\,\mathbb{E}\left[X|\mathscr{G}\right] + b\,\mathbb{E}\left[Y|\mathscr{G}\right] \tag{1.1}$$
2. positivity: if $X \geq 0$ (non negative), then $\mathbb{E}\left[X|\mathscr{G}\right] \geq 0$
3. exp.value of constant: $\mathbb{E}\left[c|\mathscr{G}\right] = c$, with $c \in \mathbb{R}$
4. if Z is $\mathscr{G}$-measurable, then
$$\mathbb{E}\left[XZ|\mathscr{G}\right] = Z \cdot \mathbb{E}\left[X|\mathscr{G}\right]$$

Interpretation: we want to predict the product of X and Z; all factors which are $\mathscr{G}$-measurable go out of the conditional expectation operator. Under the information $\mathscr{G}$, Z is known and is treated like a constant, and as constants go out of expectation operator; Z is not a constant, it's a random variable, but if it's $\mathscr{G}$-measurable, behaves as a constant.
As a consequence if X is $\mathscr{G}$-measurable the

$$\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X \cdot 1|\mathscr{G}] = X\,\mathbb{E}[1|\mathscr{G}] = X \cdot 1 = X$$

that is, if X is $\mathscr{G}$-measurable, its expected value under $\mathscr{G}$ is X itself.

5. suppose $X \perp\!\!\!\perp \mathscr{G}$ in the sense that

$$\mathbb{P}(A \cap (X \in B)) = \mathbb{P}(A) \cdot \mathbb{P}(X \in B), \quad \forall A \in \mathscr{G}, \forall B \in \beta(\mathbb{R})$$

then

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

This is reasonable from logical pov: if X is independent of $\mathscr{G}$, having or not having the info of $\mathscr{G}$ is the same in order to predict X (they are independent). And thus our prediction of X under $\mathscr{G}$ coincides with our prediction without any information.
An important special case where $X \perp\!\!\!\perp \mathscr{G}$ is when $\mathbb{P}(A) \in \{0,1\}, \forall A \in \mathscr{G}$. Infact $\forall A \in \mathscr{G}$ by assumption $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$ so:

- if $\mathbb{P}(A) = 0$ then

$$\mathbb{P}(A \cap (X \in B)) = 0 = 0 \cdot \mathbb{P}(X \in B) = \mathbb{P}(A) \cdot \mathbb{P}(X \in B)$$

- if $P(A) = 1$ then

$$\mathbb{P}(A \cap (X \in B)) = \mathbb{P}(X \in B) = 1 \cdot \mathbb{P}(X \in B) = \mathbb{P}(A) \cdot \mathbb{P}(X \in b)$$

Thus for example if $\mathscr{G} = \{\emptyset, \Omega\}$ then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$ (here $\mathscr{G}$ have no informative value)

NB: comunque lui in generale sembra indicare $\subset$ per intender $\subseteq$

6. if $\mathscr{G}_1 \subset \mathscr{G}_2 \subset \mathscr{A}$ (that is we have two sub-σ-field and we suppose they are contained, so under $\mathscr{G}_2$ we know more than $\mathscr{G}_1$) we have the so called *chain rule*:

$$\mathbb{E}[X|\mathscr{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1]$$

If $\mathscr{G}_1$ information is contained in $\mathscr{G}_2$ information then our prediction of X under the smaller information can be written in this way (we condition first on the bigger information and then under the smaller information).
For instance if we know that $\mathbb{E}[X|\mathscr{G}_2] = 0$ (eg for some reason we know that with more information our prediction of X is 0), then $\mathbb{E}[X|\mathscr{G}_1]$ (our prediction on lower information)

$$\mathbb{E}[X|\mathscr{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[0|\mathscr{G}_1] = 0$$

NB: ecco questa è la giustificazione per cui per lui $\subset = \subseteq$

As a consequence of the special case of $X \perp\!\!\!\perp \mathscr{G}$ at the previous point, by taking $\mathscr{G}_1 = \{\emptyset, \Omega\}$ and $\mathscr{G}_2 = \mathscr{G}_1$ (so we have two σ-fields with all events with probability 0 or 1). The chain rules yields:

$$\mathbb{E}[X|\mathscr{G}_1] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] \stackrel{(1)}{=} \mathbb{E}[X|\mathscr{G}_1] = \mathbb{E}[X]$$

where (1) . So $\mathbb{E}[X|\mathscr{G}_1] = \mathbb{E}[X]$.

TODO: not clear here?

Important remark 8 (Alternative explanation last point). We want to prove the equality

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

is true (i guess under $X \perp\!\!\!\perp \mathscr{G}$). To prove the equality holds we apply the chain rule: it we need two σ-field one contained in the other so we take $\mathscr{G}_2 = \mathscr{G}_1 = \{\emptyset, \Omega\}$

TODO: ma non basta notare che $X = \mathbb{E}[X|\mathscr{G}]$ sotto indipendenza?

$$\mathbb{E}[X] \stackrel{(1)}{=} \mathbb{E}[X|\mathscr{G}_1] \stackrel{(2)}{=} \mathbb{E}[E(X|\mathscr{G}_2)|\mathscr{G}_1] \stackrel{(3)}{=} \mathbb{E}[E(X|\mathscr{G}_2)]$$

where:

- (1) being $X \perp\!\!\!\perp \mathscr{G}_1$ for what we have seen in the previous point (in other words, all elements of $\mathscr{G}_1$ has prob 0 or 1 so not informative and we can condition on that)
- (2) applying the chain rule being $\mathscr{G}_1 \subset \mathscr{G}_2$
- (3) since again each element of $\mathscr{G}_1$ is prob 0 or 1 $\mathbb{E}[X|\mathscr{G}_2]$ is independent of $\mathscr{G}_1$ so we can remove the last conditioning

1.2 Conditional probability

Definition 1.2.1 (Conditional probability of an event given an event (refresher)). We remember that if $F, G \in \mathscr{A}$ and $\mathbb{P}(G) > 0$, then

$$\mathbb{P}(F|G) = \frac{\mathbb{P}(F \cap G)}{\mathbb{P}(G)}$$

Remark 4 (Indicator random variable (refresher)). We remember that

$$1_A = 1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$
$$\mathbb{E}[1_A] = \mathbb{P}(A)$$

Remark 5. Below we *define* the general conditional probability of an event given a *sigma*-field using the indicator function

Definition 1.2.2 (Conditional probability of an event given a σ-field). Given any event $F \in \mathscr{A}$ we *define* the conditional probability of F given $\mathscr{G}$ as:

$$\mathbb{P}(F|\mathscr{G}) := \mathbb{E}[1_F|\mathscr{G}]$$

Important remark 9 (Interpretation). The interpretation is the probability we give to the event F under the information $\mathscr{G}$.

Important remark 10 (Connection between the two definitions). What is the connection between $\mathbb{P}(F|G)$ and the more general $\mathbb{P}(F|\mathscr{G}) = \mathbb{E}[1_F|\mathscr{G}]$?
When we write $\mathbb{P}(F|G)$ we mean probability of F knowing that G is true. To connect the two definition let's take a field $\mathscr{G} = \{\emptyset, G, G^C, \Omega\}$ where the only unknown information provided is wheter G is happened or not (clearly the null and samplespace events are respectively false and true).
Now, it can be shown that, if a function $f : \Omega \to \mathbb{R}$ is $\mathscr{G}$-measurable when $\mathscr{G} = \{\emptyset, G, G^C, \Omega\}$, then it must be writable as linear combination of the indicators:

$$f = 1_G \cdot \alpha + 1_{G^c} \cdot \beta, \quad \alpha, \beta \in \mathbb{R}$$

Since $\mathbb{P}(F|\mathscr{G}) = \mathbb{E}[1_F|\mathscr{G}]$ must be $\mathscr{G}$-measurable (being a conditional expectation, by the second property of the definition), then it can be written as:

$$\mathbb{P}(F|\mathscr{G}) = 1_G \cdot \alpha + 1_{G^c} \cdot \beta$$

Which are α and β? To find them we use the third property of conditional expectation starting from the general case and substituting the indicator 1_F at the right moment

$$\begin{aligned}
\forall A \in \mathscr{G}, \mathbb{E}[1_A X] &= \mathbb{E}[1_A V] \\
\mathbb{E}[1_A X] &= \mathbb{E}[1_A \mathbb{E}[X|\mathscr{G}]] \\
\mathbb{E}[1_A 1_F] &= \mathbb{E}[1_A \mathbb{E}[1_F|\mathscr{G}]] \\
\mathbb{E}[1_A 1_F] &= \mathbb{E}[1_A \mathbb{P}(F|\mathscr{G})]
\end{aligned}$$

Now, by letting $A = G$

$$\begin{aligned}
\mathbb{E}[1_G 1_F] &= \mathbb{E}[1_G \mathbb{P}(F|\mathscr{G})] \\
\mathbb{E}[1_{F\cap G}] &= \mathbb{E}[1_G \mathbb{P}(F|\mathscr{G})] \\
\mathbb{P}(F \cap G) &= \mathbb{E}[1_G \mathbb{P}(F|\mathscr{G})]
\end{aligned}$$

By substituting $\mathbb{P}(F|\mathscr{G}) = 1_G \cdot \alpha + 1_{G^c} \cdot \beta$:

$$\mathbb{P}(F \cap G) = \mathbb{E}[1_G(\alpha 1_G + \beta 1_{G^c})] = \mathbb{E}[\alpha 1_G + \beta \cdot 0] = \alpha \mathbb{E}[1_G] = \alpha \mathbb{P}(G)$$

Hence:

$$\alpha = \frac{\mathbb{P}(F \cap G)}{\mathbb{P}(G)} = \mathbb{P}(F|G)$$

Similarly by repeating the same stuff with G^c instead of G we have that

$$\beta = \mathbb{P}(F|G^c)$$

So to sum up

$$\mathbb{P}(F|\mathscr{G}) = 1_G \mathbb{P}(F|G) + 1_{G^c} \mathbb{P}(F|G^c)$$

Remark 6 (Interpretation). In the last equation if G is true ($\omega \in G$) then $\mathbb{P}(F|G)$ is our prediction, otherwise it's $\mathbb{P}(F|G^c)$. So we arrived at the abstract object $\mathbb{P}(F|\mathscr{G})$ via calculations with elementary objects $\mathbb{P}(F|G)$ and $\mathbb{P}(F|G^c)$. Therefore the elementary definition is coherent with the general (it's a special case of the general above) in case we take $\mathscr{G} = \{\emptyset, \Omega, G, G^c\}$. We choose this special $\mathscr{G}$, because the only info provided is that G occurrs or not.

1.3 Stochastic processes

Definition 1.3.1 (Stochastic process)**.** Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and T any set. A stochastic process X is any collection of random variables indexed by T

$$X = \{X_t : t \in T\}$$

Remark 7. For each $t \in T$ we have a random variable; provided T can be any set is naturally interpreted as times (this is not mandatory, eg it could be a set of spaces as well).

Example 1.3.1. X_t could be the consumption of electricity in Bolo at time t.

Definition 1.3.2 (Discrete time process)**.** If T is finite or countable X is said to be a discrete time process.

Example 1.3.2. We have met discrete time processes: sequences of random variable are an example of discrete time processes. In case of sequences of random variable $T = \mathbb{N}$ (set of integer starting from 1 or 0).

Definition 1.3.3 (Continuous time process)**.** If T is an interval, X is called a continuous time process.

Example 1.3.3. Example of continuous time process later in the course.

Important remark 11. A stochastic process is actually a function of two variables. Infact, for fixed $t \in T$, X_t is a real random variable and thus X_t is a function of $\Omega \to \mathbb{R}$ (so it depends on an element $\omega \in \Omega$). So we can write

$$X_t(\omega) \quad \text{or} \quad X(t, \omega)$$

where the two variables are $t \in T$ and $\omega \in \Omega$.

Definition 1.3.4 (Path of a stochastic process X)**.** If we fix $\omega \in \Omega$, the deterministic function $p : t \to X(t, \omega)$ is called the path (or trajectory) of X.

Remark 8. How many path are there? we have a path for each point $\omega \in \Omega$ (and are depicted as a single time series)

Definition 1.3.5 (Filtration)**.** If $T \subset \mathbb{R}$, a filtration is a *nexted* collection of sub-σ-fields of $\mathscr{A}$, indexed by T. That is

- is a sequence $(\mathscr{F}_t : t \in T)$ where each $\mathscr{F}_t \subset \mathscr{A}$ is a sub-σ-field
- which is *nexted*, meaning $\mathscr{F}_s \subset \mathscr{F}_t$ if $s < t$

Example 1.3.4. For us the most important case, if $T = \{0, 1, 2, \ldots\}$ a filtration is $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{A}$

Important remark 12 (Interpretation)*.* $\mathscr{F}_0$ is our information at time 0, $\mathscr{F}_1$ is our informatin at time 1. With a filtration, we assume that as time goes on information increase so $\mathscr{F}_t \subset \mathscr{F}_{t+1}$ (tomorrow we have all the information of today plus maybe something else).

Definition 1.3.6 (Process adapted to a filtration)**.** A process X is said to be adapted to a filtration $\mathscr{F}_t$ if X_t is $\mathscr{F}_t$-measurable, $\forall t \in T$.

Remark 9 (Interpretation). In general, if a random variable X is $\mathscr{G}$-measurable, then under the information $\mathscr{G}$ we know the value taken by X.
In this case X is adapted to the information at time t, namely $\mathscr{F}_t$, X is $\mathscr{F}_t$-measurable, so $\mathscr{F}_t$ includes the value taken by X_t (and maybe other things). But certainly at time t we know the value taken by X.

Example 1.3.5 (σ-field generated by the process X until n)**.** Suppose the filtration $\mathscr{F}_n^* = \sigma(X_0, X_1, \ldots, X_n)$ be the σ-field generated by the process X until n (that is the least σ-field which makes $X_0, X_1, \ldots, X_n$ measurable).
Is the process X adapted to the filtration $\mathscr{F}_n^*$? Yes, by definition; if we are in $\mathscr{F}_n^*$ the process is necessarily adapted.

Definition 1.3.7 (Stopping time for a filtration)**.** A stopping time for the filtration $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \ldots \subset \mathscr{A}$ is any function $Tau : \Omega \to \{+\infty, 0, 1, \ldots\}$ such that the event $\{\tau = n\} \in \mathscr{F}_n$, $\forall n \geq 0$.

Important remark 13 (Interpretation). τ is a random time (look at the function) and is the first time “something” we are interested in happens (eg the time we marry for the first time is a stoppying time). It could be $\tau = +\infty$ if the event we are waiting for never happens.
There’s no process X involved in the definition, only a filtration.

Remark 10. Why we requested that the event $\{\tau = n\} \in \mathscr{F}_n \forall n \geq 0$?
Think of the following example: we go to the casino, an we start playing. At each time we can choose to play or to stop/exit: τ is time when we stop to play:

- if we could look in to the future, a good choice of τ would be to stop when our winning would reach the maximum; however this is impossible to know.
- thus a realistic strategy is $\tau = n$ (we stop playing at time n): that is decision to stop must depends only on information up to time n (can be any strategy based on info at time n).

$\{\tau = n\} \in F_n$ means literally:

- at time n you know wheter or not $\tau = n$;
- that the decision to stop playing at time n belongs to the information you have at time n/depends only on what we know up to n.

Example 1.3.6 (Stopping time important example)**.** Let:

- $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots \subset \mathscr{A}$ be a filtration
- $X_0, X_1, \ldots$ a sequence of random variable adapted to $(\mathscr{F}_n)$
- let $A \in \beta(R)$ be an event of interest, we’re interested in entering in the set A
- $\tau = \inf \{n \geq 0 : X_n \in A\}$ (τ is the first time the event A happen) with the convention $\inf \{\emptyset\} = +\infty$.

There are 2 possible situations:

- $X_n \notin A, \forall n \geq 0$: in this case the set $\{n \geq 0 : X_n \in A\}$ is empty, so $\tau(\emptyset) = +\infty$ by convention;
- $X_n \in A$ for some n: in this case τ is the first n such that $X_n \in A$

If X_n is price of actions at time n and $A = [1, +\infty)$ represent the decision to sell the action (when the price become greater or equal to 1), then

$$\tau = \text{first time since that the price is} \geq 1.$$

We need to set $\inf(\emptyset) = +\infty$ because it can be that never the price of the action becomes above 1.

Definition 1.3.8 (Finite dimensional distributions of a process X)**.** Given a process X, these are all the distributions of the n-variate random vector $(X_{t_1}, \dots, X_{t_n})$, $\forall n \geq 1$, $\forall t_1, \dots, t_n \in T$.

Important remark 14. What happens in real problems is that one first choose the finite dimensional distributions he/she wants and then looks for a process X having such finite dimensional distributions.

TODO: non ho capito se sono le marginali ordinate o marginali qualsiasi

Remark 11. This is not always in general possible.
For instance suppose that we want a process where the bivariate random variables are normal while the univariate random variables are binomial. That is where $(X_{t_1}, X_{t_2}) \sim N$, $\forall t_1, t_2 \in \mathbb{N}$ and X_t is binomial $\forall t$.
This is not possible since if the n-variate (eg bivariate) is normal all the marginals (eg univariate) are still normal. So here can't a marginal being binomial.

Remark 12. However there exists some theorems (the most important of which is the Kolmogorov consistency theorem) which give conditions on the finite dimensional distribution that guarantee the existence of a process X having such finite dimensional distributions.

Remark 13. So far we've talked about stochastic processes in general, next we start by looking at several stochastic processes, starting from a special/important type, the martingale.

Capitolo 2

Martingales

2.1 Introduction

Remark 14. The name martingale ("martingala" in italian) comes from betting: some types of bettings are called martingale, the names comes from there.

Definition 2.1.1 (Martingale)**.** Let $(\mathscr{F}_n)$ be a filtration and (X_n) a sequence of real random variable; then (X_n) is a martingale with respect to $(\mathscr{F}_n)$ if:

1. $\mathbb{E}[|X_n|] < +\infty$, $\forall n$ (all random variables have the mean)
2. (X_n) is adapted to $(\mathscr{F}_n)$
3. $\mathbb{E}[X_{n+1}|\mathscr{F}_n] = X_n$, $\forall n$ (most important/qualifying condition)

Remark 15 (Interpretation for third condition)*.* Suppose X_n is the price of an action at time n: if the process is a martingale our prediction of price tomorrow, given all we know today, is equal to the price today.
This is an assumption we make on the process: sometimes is reasonable, in other situations not.

Important remark 15 (Interpretation (Martingales as fair games))*.* Suppose you go to the casino and you play. Let X_n be the amount of money we have at time n; $X_{n+1} - X_n$ is the winning at time $n+1$ (could be negative as well).
If (X_n) is a martingale then the expectation

$$\mathbb{E}[X_{n+1} - X_n|\mathscr{F}_n] \overset{(1)}{=} \mathbb{E}[X_{n+1}|\mathscr{F}_n] - \mathbb{E}[X_n|\mathscr{F}_n] \overset{(2)}{=} X_n - X_n = 0$$

where:

- (1) because property of conditional expectation
- (2) because of the definition of martingale

So at every time the expected win of gambling, conditional to the information we have at that time, is 0; name of this gamble is "fair game" (gioco equo). Thus the martingale can be interpreted as a sequence of fair games.

Remark 16. If the process is a martingale is either a submartingale or supermartingale (but not necessary the contrary, a process could be a submartingale but not a martingale); below the weaker version of the notion martingale.

Definition 2.1.2 (Submartingale)**.** We have that $\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] \geq X_n$

Definition 2.1.3 (Supermartingale)**.** We have that $\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] \leq X_n$

Remark 17. Two remarks as propositions.

Proposition 2.1.1. *If (X_n) is a martingale with respect to some filtration $(\mathscr{F}_n)$, then (X_n) is also a martingale with respect to $(\mathscr{F}_n^*) = \sigma(X_0, X_1, \ldots, X_n)$ (σ-field generated by the process X until n).*

Dimostrazione. To prove it, we note that $\mathscr{F}_n^* \subset \mathscr{F}_n$: if X_n is a martingale with respect to F_n, then by the second condition it's adapted to F_n and thus $\mathscr{F}_n^* \subset \mathscr{F}_n$: that is at time n we have at least/surely the information provided by the process up to that time, plus maybe something else.
Therefore I can use the chain rule for conditional expectation toward the third condition of the definition:

$$\mathbb{E}\left[X_{n+1}|\mathscr{F}_n^*\right] \overset{(1)}{=} \mathbb{E}\left[\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right]|\mathscr{F}_n^*\right] \overset{(2)}{=} \mathbb{E}\left[X_n|\mathscr{F}_n^*\right] = X_n$$

where:

- in (1) we use the chain rule.
- in (2) we substitute $\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] = X_n$ by the martingale main property.
- in (3) given that X_n is adapted to $\mathscr{F}_n^*$ (so info of X_n is included in $\mathscr{F}_n^*$)

□

Remark 18. Then any time we know that X_n is a martingale with respect to any filtration we know it's also with respect to the particular filtration $(\mathscr{F}_n^*)$ derived by the process up to time n.
Eg if while reading a book the author does not specify which filtration a X_n martingale is adapted to, then we can suppose/assume the filtration is this one $(\mathscr{F}_n^*)$.
Suppose we know (X_n) is a martingale but no one specify which is the filtration; we are always allowed to take $\mathscr{F}_n^*$ as filtration.

Proposition 2.1.2. *If (X_n) is a martingale then all the rvs involved have the* same *mean:*

$$\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_0\right], \forall n$$

Dimostrazione. In fact

$$\mathbb{E}\left[X_{n+1}\right] = \mathbb{E}\left[\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right]\right] = \mathbb{E}\left[X_n\right]$$

So:

- for $n = 0$ we get $\mathbb{E}\left[X_1\right] = \mathbb{E}\left[X_0\right]$
- for $n = 1$ we get $\mathbb{E}\left[X_2\right] = \mathbb{E}\left[X_1\right] = \mathbb{E}\left[X_0\right]$
- thus by induction $\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_0\right]$, $\forall n$

□

2.2 Random walk

Remark 19. The most important random process and sometime example of martingale, is the random walk.

Definition 2.2.1 (Random walk). Any sequence (X_n) is a random walks if can be written as $X_n = X_0 + \sum_{i=1}^{n} Z_i$ where:

1. (Z_i) is iid sequence
2. the initial random variable X_0 is independent of (Z_i).

Remark 20 (Interpretation). In a sense, we are making a walk on the real axis, at every step we start from the previous position and modify it by an independent value.

Example 2.2.1 (Symmetric random walk on $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$). An important special case of random walk is defined with:

- X_0 in $\mathbb{Z}$ (take an integer at time 0 randomly)
- Z_i are such that

$$\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = \frac{1}{2}$$

At each step we go left or right of 1 with same probability.

Remark 21. A random walk is not necessarily a martingale: the thm below gives conditions under which this happens.

Theorem 2.2.1 (Randow walk and martingale). *Let $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n) = \sigma(X_0, Z_1, \dots Z_n)$; then the random walk (X_n) is a martingale if and only if:*

1. *X_0, Z_i have mean ($\mathbb{E}[|X_0|] < \infty$, $\mathbb{E}[|Z_1|] < \infty$) and*
2. *especially Z_i has mean 0 ($\mathbb{E}[Z_1] = 0$, as the others, being iid)*

Example 2.2.2. Is symmetric random walk on $\mathbb{Z}$ a martingale? Yes if X_0 has the mean, because Z_1 has mean and it's 0.

Remark 22 (Two typical exam question). Respectively:

1. what is a martingale?
2. what is a natural example of martingale? a random walk satysfying the conditions above.

2.3 Martingales and stopping time

Remark 23 (Connection martingale and stopping time). There is a nice connection between martingales and stopping time; some notation/definition before.

Important remark 16 (Notation). Given $a, b \in \mathbb{R}$, $a \wedge b$ means $\min(a, b)$.

Definition 2.3.1 (Stopped sequence)**.** Given a sequence of rvs (X_n) and a stopping time τ it's defined as:

$$(X_{n\wedge\tau}) = X_0, X_1, \dots, X_\tau, X_\tau, X_\tau, \dots$$

Remark 24. So the sequence is like the original up to τ and then all the following elements are X_τ.

Theorem 2.3.1 (Martingale and stopping time)**.** *If (X_n) is a martingale and τ is a stopping time, then the stopped sequence $(X_{n\wedge\tau})$ is still a martingale.*

Example 2.3.1 (Example of martingale: Gambler's ruin problem)**.** Me and Sarah are two player with each a certain amount of money: we roll a coin, if it comes tail I give a euro to Sarah, with head she gives an euro to me. The winner is the one which is able to ruin the other (making it go to 0). There are 3 outcomes at least in principle:

- I'm the winner
- Sarah is the winner
- the game never ends (eg we're very rich): actually it can be shown that this situation happens with probability 0.

For the math model of this process, let setup this as a symmetric random walk on integers (with 0 starting point): let X_n be my winning at time n, with $X_n = X_0 + \sum_{i=1}^{n} Z_i$, where $X_0 = 0$ (my winnning at time 0), Z_i are iid, and $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = -1) = \frac{1}{2}$.
Suppose I have $a > 0$ euros and Sarah $b > 0$ euros; in this case I win (and the games end) if $X_n = b$ (i ruin her) happens before $X_n = -a$ (i lost everything). So let's define the time when the game ends:

$$\tau = \inf\{n \geq 0 : X_n = b \text{ or } X_n = -a\}$$

X_τ are my winnings at the end of the game. Let's evaluate $\mathbb{E}[X_\tau]$: X_τ is discrete rv (so if it exists, the expected value is the sums of the value multiplied the probability of each one) where the only possible values of X_τ are b or $-a$:

$$\begin{aligned}
\mathbb{E}[X_\tau] &= b\,\mathbb{P}(X_\tau = b) + (-a)\,\mathbb{P}(X_\tau = -a) \\
&= b\,\mathbb{P}(X_\tau = b) + (-a)(1 - \mathbb{P}(X_\tau = b)) \\
&= b\,\mathbb{P}(X_\tau = b) - a + a\,\mathbb{P}(X_\tau = b)
\end{aligned}$$

Therefore

$$\mathbb{P}(X_\tau = b) = \mathbb{P}(\text{I win}) = \frac{\mathbb{E}[X_\tau] + a}{a + b}$$

Now we want to evaluate $\mathbb{E}[X_\tau]$: since the situation 3 (game never ends) has probability 0, the games ends with probability 1 so τ is finite almost surely ($\mathbb{P}(\tau < +\infty) = 1$). Thus being finite:

$$X_\tau = \lim_{n\to+\infty} X_{\tau\wedge n}$$

hence,

$$\begin{aligned}\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[\lim_{n\to+\infty} X_{\tau\wedge n}\right] &\overset{(1)}{=} \lim_{n\to+\infty} \mathbb{E}\left[X_{\tau\wedge n}\right] \\ &\overset{(2)}{=} \lim_{n\to+\infty} \mathbb{E}\left[X_{\tau\wedge 0}\right] = \lim_{n\to+\infty} \mathbb{E}\left[X_0\right] = \mathbb{E}\left[X_0\right] = 0\end{aligned}$$

where in:

- (1) lim goes out of expectation (we can interchange them here, take it as given);
- (2) since X_n is a martingale (being a symmetric random walk on the integers) then the stopped sequence $X_{\tau\wedge n}$ is a martingale as well; and being a martingale has constant mean value ($\mathbb{E}\left[X_{\tau\wedge n}\right] = \mathbb{E}\left[X_{\tau\wedge 0}\right]$)

So finally:

$$\mathbb{P}\left(\text{I win}\right) = \frac{\mathbb{E}\left[X_\tau\right] + a}{a+b} = \frac{a}{a+b}$$

2.4 Doob-Meyer theorem

Remark 25. Doob was american; Meyer french and thus is pronunced french-like with accent on the second e (not german-like on first one).

Remark 26. This next theorem provides a caraterization of submartingales. The sequence is a submartingale if it can be decomposed in the sum of two particular sequences.

Theorem 2.4.1 (Doob-Meyer theorem)**.** *(X_n) is a sub-martingale iff there exists two sequences $(A_n), (M_n)$ such that:*

- *$X_n = M_n + A_n$ where*
- *M_n is a martingale*
- *A_n is* increasing *(first two lines) and* predictable *(last) that is*

$$\begin{cases} A_0 = 0, \mathbb{E}\left[|A_n|\right] < +\infty, \forall n \\ A_0 \le A_1 \le A_2 \le \dots \\ A_n \textit{ is } \mathscr{F}_{n-1}\textit{-measurable } \forall n \end{cases}$$

Moreover the sequences (M_n) and (A_n) are also almost surely unique.

Remark 27. Predictability means that at time n our information $\mathscr{F}_n$ includes the value of A_{n+1} at the next time.

Dimostrazione. Let's suppose that (X_n) is a submartingale and define a sequence A_n such that $A_0 = 0$ and

$$A_n = \sum_{j=0}^{n-1} \left(\mathbb{E}\left[X_{j+1}|\mathscr{F}_j\right] - X_j\right)$$

We want to prove that A_n is predictable and increasing:

- (A_n) is predictable: in the definition of A_n the index j arrives at $n-1$ so it involves the variable up to X_{n-1} and the first term up to $\mathbb{E}\left[X_n|\mathscr{F}_{n-1}\right]$. Hence all the involved elements are measurable with respect to $\mathscr{F}_{n-1}$ thus A_n is measurable with respect to $\mathscr{F}_{n-1}$ (or equivalently sequence (A_n) is predictable).
- (A_n) is increasing: given that (X_n) is a submartingale (thus $\mathbb{E}\left[X_{j+1}|\mathscr{F}_j\right] \geq X_j, \forall j$) then:

  $$\mathbb{E}\left[X_{j+1}|\mathscr{F}_j\right] - X_j \geq 0, \forall j$$

  And therefore A_n is increasing being

  $$\begin{aligned} A_{n+1} &= \sum_{j=0}^{n} \left(\mathbb{E}\left[X_{j+1}|\mathscr{F}_j\right] - X_j\right) \\ &= \left(\sum_{j=0}^{n-1} \left(\mathbb{E}\left[X_{j+1}|\mathscr{F}_j\right] - X_j\right)\right) + \mathbb{E}\left[\left(X_{n+1}|\mathscr{F}_n\right)\right] - X_n \\ &= A_n + (\text{something} \geq 0) \end{aligned}$$

  and so $A_{n+1} \geq A_n$.

So A_n satisfies the conditions of the theorem.
Next define $M_n = X_n - A_n$; then obviously $X_n = M_n + A_n$.
It remains only to show that (M_n) is a martingale to end the proof. We have

$$\begin{aligned} \mathbb{E}\left[M_{n+1}|\mathscr{F}_n\right] &= \mathbb{E}\left[X_{n+1} - A_{n+1}|\mathscr{F}_n\right] \\ &= \mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] - \mathbb{E}\left[A_{n+1}|\mathscr{F}_n\right] \\ &\overset{(1)}{=} \mathbb{E}\left[X_{n+1|\mathscr{F}_n}\right] - A_{n+1} \\ &= \mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] - \left[A_n + \left(\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] - X_n\right)\right] \\ &= X_n - A_n \\ &= M_n \end{aligned}$$

where in (1) the $\mathscr{F}_n$-measurability of A_{n+1} was used.
On uniqueness we omit the demonstration. □

Remark 28. Before an example we need one more fact. Suppose we have the sequence (X_n) and a function f, and we are interested in $(f(X_n))$ especially if/when it's a submartingale. The following theorem provides conditions.

Theorem 2.4.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a real valued function such that $\mathbb{E}\left[|f(X_n)|\right] \leq \infty$ (the variable obtained applying the function has the mean). Now:*

1. *if (X_n) is a martingale and f is convex, OR ...*
2. *if (X_n) is a submartingale and f is increasing and convex ...*

... then the new sequence $(f(X_n))$ is a submartingale

Example 2.4.1. If (X_n) is a martingale and the rv have the second moments $\mathbb{E}\left[X_n^2\right] < \infty, \forall n$, then what can we say about the sequence (X_n^2)? We can say that:

- it's a submartingale by the first point of the last theorem since the square is a convex function;
- hence by Doob-Meyer thm, since X_n^2 is a submartingale we can write it as $X_n^2 = M_n + A_n$, with M_n a martingale and A_n increasing and predictable.

Example 2.4.2. Let $X_0 = 0$, $X_n = \sum_{i=1}^n Z_i$ where (Z_i) is iid, $\mathbb{E}[Z_1] = 0, \mathbb{E}\left[Z_1^2\right] = 1$. Is it a martingale being a random walk with Z_1 having mean 0. Now let:

- $M_n = X_n^2 - n$
- $\mathscr{F}_n = \sigma(Z_1, \ldots, Z_n)$

Then

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathscr{F}_n] &= \mathbb{E}\left[X_{n+1}^2 - (n+1)|\mathscr{F}_n\right] \\
&\overset{(1)}{=} \mathbb{E}\left[X_{n+1}^2|\mathscr{F}_n\right] - (n+1) \\
&= \mathbb{E}\left[(X_n + Z_{n+1})^2|\mathscr{F}_n\right] - (n+1) \\
&\overset{(2)}{=} \mathbb{E}\left[X_n^2 + Z_{n+1}^2 + 2X_n Z_{n+1}|\mathscr{F}_n\right] - (n+1) \\
&= X_n^2 + \mathbb{E}\left[Z_{n+1}^2|\mathscr{F}_n\right] + 2X_n\,\mathbb{E}[Z_{n+1}|\mathscr{F}_n] - (n+1) \\
&\overset{(3)}{=} X_n^2 + \mathbb{E}\left[Z_{n+1}^2\right] + 2X_n\,\mathbb{E}[Z_{n+1}] - (n+1) \\
&\overset{(4)}{=} X_n^2 - n \\
&= M_n
\end{aligned}$$

where in:

- (1) we take the constant $(n+1)$ out of expectation
- (2) used the $\mathscr{F}_n$-measurability of X_n^2
- (3) by independence of Z_{n+1} from $\mathscr{F}_n$ we drop conditioning
- (4), we have that $\mathbb{E}\left[Z_{n+1}^2\right]$ is the variance and is equal to 1 and $\mathbb{E}[Z_{n+1}] = 0$ by assumption

So we proved that (M_n) is a martingale. Therefore since $X_n^2 = M_n + n$, by the uniqueness in Doob-meyer thm one obtains $A_n = n$; this latter sequence (n) is both increasing (trivially: $1, 2, 3, \ldots$ is clearly increasing) and predictable (since is a constant).

2.5 Martingale convergence

Remark 29. Next is an important thm on martingales (and submartingales as well): if X_n is a martingale and satisfies some defined conditions it converge to a limit as $n \to \infty$.

Important remark 17 (RV convergence (refresher)). Let (X_n) be any sequence of real random variable:

- $X_n \xrightarrow{a.s.} X$ means by definition that

$$\mathbb{P}\left(\omega \in \Omega : X(\omega) = \lim_n X_n(\omega)\right) = 1$$

- $X_n \xrightarrow{L_p} X \iff$ all the random variables have moment of order p ($\mathbb{E}[|X_n|^p] < \infty$, $\mathbb{E}\left[|X^P|\right] < \infty$) and $\mathbb{E}[|X_n - X|^p] \to 0$

Two remarks:

1. there's no connection between the two convergenges; they don't imply each other (different modes of convergence);
2. if $X_n \xrightarrow{L_p} X$ then $\mathbb{E}[X_n^p] \to \mathbb{E}[X^p]$

Theorem 2.5.1 (Martingale convergence thm)**.** *Let (X_n) be a submartingale:*

1. *if* $\sup_n \mathbb{E}[|X_n|] < \infty$*, then* $X_n \xrightarrow{a.s.} X$
2. *if (X_n) is* uniformly integrable*, then* $X_n \xrightarrow{as, L_1} X$
3. *if (X_n) is a martingale too and* $\sup_n \mathbb{E}\left[X_n^2\right] < \infty$ *then* $X_n \xrightarrow{as, L_2} X$

Remark 30. The thm holds not only for martingale but also for submartingale (if a thm is valid for submartingale it's better).
In the three points the conclusion is always the convergence of X_n, but the mode of convergence becomes better and better.

Remark 31 (On the first point)*.* Regarding the first point, when $\sup_n \mathbb{E}[|X_n|] < +\infty$ we not only say that all X_n have mean, but also the mean are upper bounded by a constant, that is there exists $c \in \mathbb{R}$ such that $E|X_n| \leq c$, $\forall n$.

Example 2.5.1. Suppose $\mathbb{E}[|X_n|] = n$, $\forall n$. Then each rv X_n has the mean, but the first condition is false since there's no constant that bounds $\mathbb{E}[|X_n|]$, so $\sup_n \mathbb{E}[|X_n|] = +\infty$. So in this case the condition of theorem fails.

Remark 32 (On the second point and uniform integrability)*.* Point 2 involvs *uniform integrability*: it's just a technical condition (don't attach an interpretation).

Definition 2.5.1 (Uniform integrability)**.** The sequence (X_n) is uniformly integrable if and only if

$$\lim_{c \to +\infty} \sup_n \mathbb{E}[|X_n| \, 1(|X_n| > c)] = 0$$

Remark 33. Uniform integrability implies that

$$\sup_n \mathbb{E}[|X_n|] < +\infty$$

(but the inverse is not true). So at point 2 we assume something more than point 1 (uniform integrability) but obtain something more (not only a.s. convergence but also L_1).

Remark 34. Exam question: make an example of martingal satisfying the sup condition (point 1 of the convergence thm) but is not uniformly integrable. Then one has to provide one of the next two examples.

Remark 35. In the following two examples we present a martingale (X_n) such that the $\sup_n \mathbb{E}\left[|(| X_n)\right] < \infty$ but (X_n) is not uniformly integrable.
Thanks to martingale convergence we can say it converge a.s., but does not converge in L_1.

Example 2.5.2. The $X_n = e^{-\frac{n}{2}+\sum_{i=1}^{n} Z_i}$ where $X_0 = 1$, (Z_i) is iid and $Z_1 \sim \mathrm{N}(0,1)$.
Now we prove the sequence is a martingale, it satifies the sup condition but it's not uniformly integrable:

1. the sequence is a martingale. Let $\mathscr{F}_n = \sigma(Z_1, \ldots, Z_n)$ then

$$\begin{aligned}
\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] &= \mathbb{E}\left[e^{-\frac{n+1}{2}} e^{\sum_{i=1}^{n+1} Z_i}|\mathscr{F}_n\right] \\
&= e^{-\frac{n+1}{2}} \cdot \mathbb{E}\left[e^{\sum_{i=1}^{n} Z_i} \cdot e^{Z_{n+1}}|\mathscr{F}_n\right] \\
&\overset{(1)}{=} e^{-\frac{n+1}{2}} e^{\sum_{i=1}^{n} Z_i}\, \mathbb{E}\left[e^{Z_{n+1}}|\mathscr{F}_n\right] \\
&\overset{(2)}{=} e^{-\frac{n+1}{2}+\sum_{i=1}^{n} Z_i}\, \mathbb{E}\left[e^{Z_{n+1}}\right]
\end{aligned}$$

   where

   - in (1) $e^{\sum_{i=1}^{n} Z_i}$ is measurable with respect to $\mathscr{F}_n$ so we take it out.
   - in (2) because independence of the conditioning $|\mathscr{F}_n$ so we can drop the conditioning

   Now we note that the final $\mathbb{E}\left[e^{Z_{n+1}}\right]$ is the moment generating function of $\mathrm{N}(0,1)$ evaluated at point 1; so we look into a book and see that this is $e^{1/2}$. So to continue

$$\mathbb{E}\left[X_{n+1}|\mathscr{F}_n\right] = e^{-\frac{n+1}{2}+\sum_{i=1}^{n} Z_i+\frac{1}{2}} = X_n$$

   and the sequence is a martingale.

2. the sup condition $\sup \mathbb{E}\left[|X_n|\right] < \infty$ is verified: looking at X_n,

$$\mathbb{E}\left[|X_n|\right] \overset{(1)}{=} \mathbb{E}\left[X_n\right] \overset{(2)}{=} \mathbb{E}\left[X_0\right] = \mathbb{E}\left[1\right] = 1$$

   where in:

   - (1) being involved an exponential, it's always positive and so the absolute value of the expected value is superfluous;
   - (2) because being a martingale expectation is constant

   Hence:

$$\sup_n \mathbb{E}\left[|X_n|\right] = 1 \leq +\infty$$

   Be careful that if X_n is a martingale then $\mathbb{E}\left[X_n\right] = \mathbb{E}\left[X_0\right]$, $\forall n$ but it's not true that $\mathbb{E}\left[|X_n|\right] = \mathbb{E}\left[|X_0|\right]$.

3. finally the sequence is not uniformly integrable. By the martingale convergence theorem one obtains $X_n \xrightarrow{a.s.} X$; but in this case it's easy to identify X, in fact:

$$X_n = e^{-\frac{n}{2}+\sum_{i=1}^n Z_i} = e^{n\left(\frac{\sum_{i=1}^n Z_i}{n}-\frac{1}{2}\right)}$$

What is the limit of $\sum_{i=1}^n Z_i/n$? For the strong law of large number we have:

$$\frac{\sum_{i=1}^n Z_i}{n} \xrightarrow{a.s.} \mathbb{E}\left[Z_1\right] = 0$$

Hence:

$$\left(\frac{\sum_{i=1}^n Z_i}{n} - \frac{1}{2}\right) \xrightarrow{a.s.} -\frac{1}{2}$$
$$n\left(\frac{\sum_{i=1}^n Z_i}{n} - \frac{1}{2}\right) \xrightarrow{a.s.} -\infty$$
$$e^{n\left(\frac{\sum_{i=1}^n Z_i}{n}-\frac{1}{2}\right)} \xrightarrow{a.s.} e^{-\infty} = 0$$

So in this example the limit X of sequence X_n is actually 0 (dirac).
Now suppose (X_n) is uniformly integrable: then by point 2 of convergence thm

$$X_n \xrightarrow{a.s., L_1} X$$

and thus

$$0 = \mathbb{E}\left[0\right] = \mathbb{E}\left[X\right] \overset{(1)}{=} \lim_n \mathbb{E}\left[X_n\right] = \lim_n \mathbb{E}\left[1\right] = 1$$

where in (1) if uniformly integrable we have L_1 convergence so the equality is provided by L_1 convergence.
So this is a contradiction (absurd): assuming its uniformly integrable we obtain an absurdity. Then it's not uniformly integrable

Example 2.5.3. One more example of the same type. Let

$$Y_0 = 1, Y_n = \sum_{i=1}^n Z_i, \quad Z_i \text{ is iid}$$
$$P(Z_1 = -1) = P(Z_1 = 1) = 1/2$$

It's a martingale since it's a symmetric random walk on the integers starting from 1 (instead of 0 but its the same). Now let

$$X_n = Y_{n\wedge\tau} = \begin{cases} Y_\tau & \text{if } n \geq \tau \\ Y_n & \text{if } n < \tau \end{cases}$$

where τ is the stopping time

$$\tau = \inf\left\{n \geq 0 : Y_n = 0\right\}$$

the behaviour of (X_n) is that it copies Y_n as long as this latter is positive and then when it become 0 repeats 0.
Now we prove the three point regarding X_n not Y_n:

1. (X_n) is a martingale as well, if we stop a martingale we have a martingale again;
2. the sup condition is verified: we have that X_n is positive ($X_n \geq 0$) with value $X_\tau = Y_\tau = 0$ and if $n < \tau \implies X_\tau = Y_n > 0$ (strictly $>$ because we start from 1). Since $X_n \geq 0$ we repeat as before:

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = \mathbb{E}[X_0] = \mathbb{E}[Y_{\tau\wedge 0}] = \mathbb{E}[Y_0] = \mathbb{E}[1] = 1$$

thus once again the sup condition is satisfied

$$\sup_n \mathbb{E}[|X_n|] = 1 < +\infty$$

3. finally to prove that is not uniformly integrable. Since τ is finite ($< +\infty$) almost surely:

$$X_n = Y_{n\wedge\tau} \xrightarrow{a.s.} Y_\tau = 0$$

Hence even in this example once again $X = 0$ and now we procede as before. If (X_n) is uniformly integrable we get

$$0 = \mathbb{E}[X] = \lim_n \mathbb{E}[X_n] = \mathbb{E}[X_0] = 1$$

absurd.

Remark 36. Before another example let's see another consequence of the martingale convergence thm

Example 2.5.4. Let Z be a real random variable with mean $\mathbb{E}[|Z|] < +\infty$. Define

$$X_n = \mathbb{E}[Z|\mathscr{F}_n]$$

X_n is our prediction of Z at time n: X_n is a martingale infact

$$\mathbb{E}[X_{n+1}|\mathscr{F}_n] = \mathbb{E}[\mathbb{E}[Z|\mathscr{F}_{n+1}]|\mathscr{F}_n] \overset{(2)}{=} \mathbb{E}[Z|\mathscr{F}_n] \overset{(3)}{=} X_n$$

where in

- (2) since $F_n \subset F_{n+1}$ using the chain rule
- (3) by definition

TODO: non mi è chiarissimo qui

So X_n is a martingale and it can be shown that X_n is uniformly integrable (take it as given).

By the convergence thm we can say that $X_n \xrightarrow{a.s.,L_1} X$; in this example we can also understand the form of X. X_n is our prediction of the information at time n. With $n \to +\infty$ it becomes $\mathbb{E}[Z|\mathscr{F}_\infty]$ that is the prediction of Z given the limit information (where $\mathscr{F}_\infty$ is the least σ-field which contains $\mathscr{F}_n$, $\forall n$).

Hence if Z is $\mathscr{F}_\infty$-measurable one obtains

$$X_n = \mathbb{E}[Z|\mathscr{F}_n] \xrightarrow{a.s.,L_1} \mathbb{E}[Z|\mathscr{F}_\infty] = Z$$

we discovered the hot water

Remark 37. As a last application of martingale convergence thm, let's prove the following result.

Proposition 2.5.2. *If (Z_n) is iid sequence with $\mathbb{E}[Z_1] = 0$, variance $\mathbb{E}\left[Z_1^2\right] = 1$ and τ is stopping time such that $\mathbb{E}[\tau] < +\infty$ then*

- $\mathbb{E}\left[\sum_{i=1}^{\tau} Z_i\right] = 0$
- $\mathbb{E}\left[\sum_{i=1}^{\tau} Z_i^2\right] = \mathbb{E}[\tau]$

Remark 38. Before proving that what is the meaning? Consider the following V sum of a *random number* of *random variables*:

$$\sum_{i=1}^{\tau} Z_i = V$$

suppose that $\tau = k$ (τ is constant dirac). Then

$$\mathbb{E}\left[\sum_{i=1}^{\tau} Z_i\right] = \mathbb{E}\left[\sum_{i=1}^{k} Z_i\right] = \sum_{i=1}^{k} \mathbb{E}[Z_i] = 0$$

and

$$\mathbb{E}\left[\sum_{i=1}^{\tau} Z_i^2\right] = \mathbb{E}\left[\sum_{i=1}^{k} Z_i^2\right] = \sum_{i=1}^{k} \mathbb{E}\left[Z_i^2\right] = k \cdot 1 = k$$

The thm generalize the mean and variance of the sum of standardized indipendent random variables when τ is not a constant but a random variable

Dimostrazione. Let's define the sequence Y_n

$$Y_0 = 0, \qquad Y_n = \left(\sum_{i=1}^{n} Z_i\right)^2 - n,$$

Then Y_n is a martingale (we prove this fact yesterday). Hence $X_n = Y_{n\wedge\tau}$ is still a martingale (it's a stopped martingale), hence

TODO: non sicurissimo qui, credo $\mathbb{E}[X_n]$

$$\mathbb{E}\left[\left(\sum_{i=1}^{n\wedge\tau} Z_i\right)^2\right] - \mathbb{E}[\tau \wedge n] = \mathbb{E}[Y_{n\wedge\tau}] = \mathbb{E}[X_n] = \mathbb{E}[X_0] = \mathbb{E}[Y_0] = 0$$

Hence by looking at first and second, we have that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n\wedge\tau} Z_i\right)^2\right] = \mathbb{E}[\tau \wedge n]$$

and thus

$$\sup_n \mathbb{E}\left[\left(\sum_{i=1}^{n\wedge\tau} Z_i\right)^2\right] = \sup_n \mathbb{E}[\tau \wedge n] \overset{(1)}{=} \mathbb{E}[\tau] \overset{(2)}{<} +\infty$$

where (1) since the last is monotone and the sup of the last expectation is $\mathbb{E}[\tau]$ and (2) by assumption.
Define now

$$L_0 = 0, \quad L_n = \sum_{i=1}^{n\wedge\tau} Z_i$$

Then (L_n) is a martingale and

$$\sup_n \mathbb{E}\left[Ln^2\right] = \sup_n \mathbb{E}\left[\left(\sum_{i=1}^{n\wedge\tau} Z_i\right)^2\right] < +\infty$$

So by martingale convergence theorem (3rd point) we have $L_n \xrightarrow{a.s.,L_2} L$. What is L? in this case, since

$$L_n = \sum_{i=1}^{n\wedge\tau} Z_i \xrightarrow{(1)} L = \sum_{i=1}^{\tau} Z_i$$

in (1), τ has finite mean value hence it's almost surely finite, so being finite with n that goes to $+\infty$ we keep τ.
And thus

$$\mathbb{E}\left[\left(\sum_{i=1}^{\tau} Z_i\right)^2\right] = \mathbb{E}\left[L^2\right] \overset{(1)}{=} \lim_n \mathbb{E}\left[L_n^2\right] = \lim_n \mathbb{E}\left[\left(\sum_{i=1}^{n\wedge\tau} Z_i\right)^2\right] = \mathbb{E}[\tau]$$

where (1) by L_2 convergence. And this proves the second fact of the theorem.
Then

$$\mathbb{E}\left[\sum_{i=1}^{\tau} Z_i\right] = \mathbb{E}[L] \overset{(1)}{=} \lim_n \mathbb{E}[L_n]$$

in (1) by L_1 convergence holds (since L_2 convergence) so

$$\lim_n \mathbb{E}[L_n] = \lim_n \mathbb{E}\left[\sum_{i=1}^{n\wedge\tau} Z_i\right] \overset{(1)}{=} \mathbb{E}[L_0] = 0$$

in 1 being a martingale $\square$

Capitolo 3

Markov chains

Remark 39. Markov chains are still discrete time processes.

3.1 Introduction

Definition 3.1.1. Let (X_n) be a sequence of random variables and $(\mathscr{F}_n)$ a filtration. Then X_n is a markov chain with respect to $(\mathscr{F}_n)$ if:

- (X_n) is adapted
- $\mathbb{P}(X_{n+1} \in A|\mathscr{F}_n) = \mathbb{P}(X_{n+1} \in A|X_n)$, $\forall n \geq 0$, $\forall A \in \beta(\mathbb{R})$

Remark 40 (Interpretation). The second condition capture the idea of the markov chain; the probability of something regarding the future given all the information up to now is equal to the probability of the future event conditionally only on the value taken today. Informally:

$$\mathbb{P}(\text{future}|\text{ present and the past}) = \mathbb{P}(\text{future}|\text{ present})$$

Again, this is an assumption reasonable for some phenomenon, not all.

Important remark 18. It follows from the definition that going forward in the future more than one step depends only on the value today:

$$\mathbb{P}(X_{n+1} \in A_1, \dots, X_{n+k} \in A_k|\mathscr{F}_n) = \mathbb{P}(X_{n+1} \in A_1, \dots, X_{n+k} \in A_k|X_n)$$

Remark 41. Spesso per n fissato la filtrazione che si ha in mente è $\mathscr{F}_n = \sigma(X_0, X_1, \dots, X_n)$, ossia si conosce il valore di tutte le variabili X sino a X_n $(X_0 = x_0, X_1 = x_1, \dots)$.

Important remark 19. In this course we speak about the most important type of markov chain (MC), the *homogeneous* Markov chains.

Definition 3.1.2 (Kernel of a Markov chain)**.** It's the probability of a future event given the actual value:

$$\alpha_n(x, A) = \mathbb{P}(X_{n+1} \in A|X_n = x), \quad \forall x \in \mathbb{R}, \forall A \in \beta(\mathbb{R})$$

Definition 3.1.3. A Markov chain is said to be homogeneous if its kernel is constant: $\alpha_n = \alpha_0$, $\forall n \geq 0$.

Remark 42 (Interpretation)*.* So the probability of starting on x and going into A does not depend on time, it's always the same.

Remark 43*.* Some remarks:

1. since $\alpha_n = \alpha_0 \ \forall n \geq 0$, the index is superfluous and we just write α instead of α_0
2. α is usually called the *kernel* of the markov chain (nucleo della catena);
3. in the rest of this course we will only deal with homogeneous Markov chains (if we forget to say homogeneous, this is implied)

Important remark 20*.* If the Markov chain is homogeneous the probability of the sequence depends only on α and the distribution of X_0. That is to write down the distribution of any X_n in the sequence we need only the kernel α and the starting point (so the probability distribution of X_0, being it a random variable).
Infact let's say $X_0 = x$ is the starting point (any):

1. i have $\mathbb{P}(X_1 \in A | X_0 = x)$: but this is $\alpha(x, A)$
2. next
$$\mathbb{P}(X_2 \in A | X_0 = x, X_1 = y) \overset{(1)}{=} \mathbb{P}(X_2 \in A | X_1 = y) = \alpha_1(y, A) = \alpha(y, A)$$
in (1) because the definition of markov chain.

Thus in practice if we have a phenomenon and we want to describe using a Markov chain we need the starting point (it's random so we need the probability distribution) and the kernel only.

Example 3.1.1 (Modelling political behaviour (party choosen at election))**.** We suppose that the probability to choose the party Y at time depends at the first time we voted plus the kernel

$$\mathbb{P}(\text{I select } Y | \text{I selected } X \text{ in the previous election}) = a$$

$$\mathbb{P}(X_0) = \begin{cases} \text{prob I choose } X \text{ at the first election I participated} \\ \text{prob I choose } Y \text{ at the first election I participated} \\ \text{prob I choose } Z \text{ at the first election I participated} \end{cases}$$

Example 3.1.2. Any random walk is an homogeneous markov chain (it's not always a martingale, it depends; but it's always a markov chain).
Consider a random walk

$$X_n = X_0 + \sum_{i=1}^{n} Z_i, \quad Z_i \text{ iid}, \, X_0 \perp\!\!\!\perp (Z_i)$$

To prove it's a Markov chain we calculate

$$\begin{aligned} \mathbb{P}(X_{n+1} \in A | X_0 = x_0, \ldots, X_n = x_n) &= \mathbb{P}(X_n + Z_{n+1} \in A | X_0 = x_0, \ldots, X_n = x_n) \\ &\overset{(1)}{=} \mathbb{P}(x_n + Z_{n+1} \in A | X_0 = x_0, \ldots, X_n = x_n) \\ &\overset{(2)}{=} \mathbb{P}(x_n + Z_{n+1} \in A) \\ &\overset{(3)}{=} \mathbb{P}(x_n + Z_1 \in A) \end{aligned}$$

where in:

- (1) we replaced with x_n at left of conditioning since known
- (2) we noted Z_{n+1} are independent from $X_0, \dots, X_n$ so we drop conditioning
- (3) being identically distributed

Now suppose we want to evaluate the probability conditioning only on the last value; by repeating the same argument as above we have

$$\mathbb{P}\left(X_{n+1} \in A | X_n = x_n\right) = \mathbb{P}\left(x_n + Z_1 \in A\right)$$

So we proved that

$$\mathbb{P}\left(X_{n+1} \in A | X_0 = x_0, \dots, X_n = x_n\right) \overset{(1)}{=} \mathbb{P}\left(X_{n+1} \in A | X_n = x_n\right) = \mathbb{P}\left(x_n + Z_1 \in A\right)$$

where the (1) equality means this is a markov chain.
Furthermore the kernel does not depend on time making it an homogeneous markov chain

$$\alpha(x, A) = \mathbb{P}\left(X_1 \in A | X_0 = x\right) = \mathbb{P}\left(x + Z_1 \in A\right)$$

3.2 Discrete Markov chains

Important remark 21 (Discrete markov chain assumption). From now on, **we assume** that X_n is **discrete** $\forall n \geq 0$ (simplifying assumption). So all the random variables take value in a set which is finite or countable: $X_n \in S, \forall n \geq 0$ almost surely with S finite or countable.

Definition 3.2.1 (State space, states)**.** The set S is called the *state space* of the Markov chain; the elements of S are called *states*.

Remark 44 (Kernel notation). In this case, the notation become simpler. Instead of writing the kernel this way

$$\alpha(x, A) = \mathbb{P}\left(X_1 \in A | X_0 = x\right)$$

we can write

$$\alpha(x, y) = \mathbb{P}\left(X_1 = y | X_0 = x\right)$$

So α now is the probability of starting from x and arriving in y.

Example 3.2.1 (Vote example continued)**.** Here

$$S = \{\text{set of all party that can be selected}\} = \{x, y, z\}$$

The kernel is

$$\alpha(x, y) = \mathbb{P}\left(\text{I select } y \mid \text{at previous selected } x\right)$$

and the distribution at time X_0 being

$$\mathbb{P}\left(X_0 = x\right)$$
$$\mathbb{P}\left(X_0 = y\right)$$
$$\mathbb{P}\left(X_0 = z\right)$$

Example 3.2.2. Consider the symmetric random walk on $\mathbb{Z}$

$$\begin{aligned} &X_0 \in \mathbb{Z}, \text{ almost surely} \\ &X_n = X_0 + \sum_{i=1}^{n} Z_i, X_0 \perp\!\!\!\perp Z_i, Z_i \text{ is iid} \\ &\mathbb{P}(Z_1 = -1) = \mathbb{P}(Z_1 = 1) = \frac{1}{2} \end{aligned}$$

In this case $S = \mathbb{Z}$.
What is the kernel? Just write down the definition/formula to help with developing the kernel:

$$\alpha(x, y) = \mathbb{P}(X_1 = y | X_0 = x) = \begin{cases} \frac{1}{2} & \text{if } y = x + 1, \text{ we increment by 1} \\ \frac{1}{2} & \text{if } y = x - 1, \text{ we decrement by 1} \\ 0 & \text{otherwise} \end{cases}$$

Example 3.2.3 (Ehrenfest urn)**.** We have two boxes and a fixed number r balls. Some balls are in the first box, the remaining in the second. At each time n we select a ball at random (each with probability $1/r$) and the selected ball moves from the urn where it is to the other urn (if it belongs to box 1 it goes in box 2 and viceversa). This simple model is useful to describe termodynamical stuff/systems.
Let X_n be the number of balls in box 1 at time n. In this case the number of balls in box 1 at any time can be $0, 1, \ldots$, so $S = \{0, 1, \ldots, r\}$; the kernel

$$\begin{aligned} \alpha(x, y) &= \mathbb{P}(X_1 = y | X_0 = x) \\ &= \begin{cases} \frac{x}{r} & \text{if } y = x - 1 \text{ (we draw from this box, which decreases)} \\ 1 - \frac{x}{r} & \text{if } y = x + 1 \text{ (we draw from the other box, so increases)} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Definition 3.2.2 (Stopping time for state x)**.** Next, for each state $x \in S$ let

$$\begin{aligned} \tau_x &= \inf\{n > 0 : X_n = x\} \\ &= \begin{cases} +\infty & \text{if } X_n \neq x, \forall n \geq 0 \\ \text{first time } n \text{ such that } X_n = x & \text{otherwise} \end{cases} \end{aligned}$$

Definition 3.2.3 (Recurrent state)**.** The state $x \in S$ is called recurrent if

$$\mathbb{P}(\tau_x < +\infty | X_0 = x) = 1$$

Remark 45. A state is recurrent if we are certain to back if we start from it.

Proposition 3.2.1. *A state x is a recurrent state* $\iff \sum_n \mathbb{P}(X_n = x | X_0 = x) = +\infty$

Dimostrazione. Omitted. □

Definition 3.2.4 (Transient state)**.** State $x \in S$ is transient if it is not recurrent:

$$\mathbb{P}(\tau_x < +\infty | X_0 = x) < 1$$

Definition 3.2.5 (Null and positively recurrent)**.** If x is recurrent there are two possible situations:

1. x is null recurrent if

$$\mathbb{E}\left[\tau_x | X_0 = x\right] = +\infty$$

2. x is positively recurrent if

$$\mathbb{E}\left[\tau_x | X_0 = x\right] < +\infty$$

Remark 46. If the state is recurrent, i'm sure that sooner or later i'll be back to it, but the time this happens can be quite different and so we have different definitions regarding it:

- if null recurrent the expected time we come back is long
- instead if positively recurrent, the expected/mean time when this happens is (at least) finite

Example 3.2.4 (Funny example)**.** Both happy days and bad days are recurrent states (they come back sooner or later) but unfortunately bad days are positively recurrent while happy days are null recurrent.

Definition 3.2.6 (Irreducible Markov chain)**.** A Markov chain is irreducible if for any couple of states $\forall x, y \in S$, $\exists m, n$ such that

$$\mathbb{P}\left(X_m = y | X_0 = x\right) > 0 \text{ and } \mathbb{P}\left(X_n = x | X_0 = y\right) > 0$$

Remark 47 (Interpretation). If irreducible, whichever two states we choose, we go from the first to the second with positive probability (not necessary 1, just positive) and from the second to the first with positive probability again.

Proposition 3.2.2. *If a chain is irreducible all states are of the same type. They all are either transient, null recurrent or positive recurrent.*

Example 3.2.5. Both symmetric random walk and ehrenfest urn are irreducibile Markov chain.
In case of symmetric random walk this is obvious: supposing at a certain time we are in a point x and we want to arrive to a point y. Supposing $x < y$, if we want to go from x to y it's enough to observe 1 $(y - x)$ times (and this has a strictly positive probability); otherwis if we want to go from y to x it's enough to observe -1 exactly $(y - x)$ times (which has a positive probability).

Remark 48. Now we focus on the symmetric radom walk on **Z**, which types is it? Transient, null recurrent or positive recurrent

Proposition 3.2.3. *The symmetric random walk on* $\mathbb{Z}$ *is null recurrent.*

Dimostrazione. We know that all states are of the same type. Let's consider the state $x = 0$ and try to determine its type, by using proposition 3.2.1

$$\begin{aligned}\sum_n \mathbb{P}(X_n = 0|X_0 = 0) &\overset{(1)}{=} \sum_{n \text{ even}} \mathbb{P}(X_n = 0|X_0 = 0)\\ &= \sum_n \mathbb{P}(X_{2n} = 0|X_0 = 0)\\ &\overset{(2)}{=} \sum_n \mathbb{P}\left(X_0 + \sum_{i=1}^{2n} Z_i = 0|X_0 = 0\right)\\ &= \sum_n \mathbb{P}\left(\sum_{i=1}^{2n} Z_i = 0\right)\\ &= \sum_n \mathbb{P}\left(\sum_{i=1}^{2n} \frac{Z_i+1}{2} = n\right)\end{aligned}$$

where

TODO: rivedere ultimo passaggio algebricamente

- in (1) we can restrict indexes of the sum because if n is odd $\mathbb{P}(X_n = 0|X_0 = 0) = 0$ (at each step we can go up or low by 1, if the steps are odd we will never come back to 0);
- in (2) by substituting the random walk definition

Now what is the distribution of $\sum_{i=1}^{2n} \frac{Z_i+1}{2}$? Since Z_i can take only 0 or 1 and are iid, we're summing independent Bernoulli rvs so the sum is a binomial and especially $\sum_{i=1}^{2n} \frac{Z_i+1}{2} \sim \text{Bin}\left(2n, \frac{1}{2}\right)$. Thus to continue

$$\begin{aligned}\sum_n \mathbb{P}(X_n = 0|X_0 = x) &= \sum_n \mathbb{P}\left(\sum_{i=1}^{2n} \frac{Z_i+1}{2} = n\right)\\ &= \sum_n \mathbb{P}\left(\text{Bin}\left(2n, \frac{1}{n}\right) = n\right)\\ &= \sum_n \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\\ &\overset{(1)}{=} +\infty\end{aligned}$$

where in (1) we go to the friendly mathematician ad this sums up to $+\infty$ so $\sum_n \mathbb{P}(X_n = 0|X_0 = x) = +\infty$.
So we proved that 0 is a recurrent state and therefore all the states (being the same type being the chain irreducible) are recurrent.
Now we have to prove that 0 is null recurrent. Cconsider $\mathbb{E}[\tau_0|X_0 = 0]$ with $\tau_0 = \inf\{n > 0 : X_n = 0\}$ (so it's the first positive time we are in 0). We have two possible situations regarding the expected value:

$$\mathbb{E}[\tau_0|X_0 = 0] = \begin{cases} < \infty & \text{if true, 0 is positive recurrent} \\ = +\infty & \text{if true, 0 is null recurrent} \end{cases}$$

Toward a contradiction suppose that the expected time is finite ($\mathbb{E}[\tau_0|X_0 = 0] < \infty$), so 0 is positive recurrent. In this case we can say regardin the expected value of the process:

$$\mathbb{E}\left[X_{\tau_0}|X_0 = 0\right] \overset{(1)}{=} 0$$
$$\mathbb{E}\left[X_{\tau_0}^2|X_0 = 0\right] \overset{(1)}{=} \mathbb{E}\left[\tau_0|X_0 = 0\right]$$

where in (1) we used an application of the martingale convergence theorem: if you have the random walk and the Z_i are iid but of mean 0 and variance 1 and stopping time of finite mean, then the sum of 1 to tau of Z_i has mean zero and variance equal to the expectation of τ_0.
The latter expression is the variance of X_{τ_0}.
But $\tau_0 \geq 1$ so the variance of

$$\operatorname{Var}\left[X_{\tau_0}|X_0 = 0\right] = \mathbb{E}\left[X_{\tau_0}^2|X_0 = 0\right] \geq 1$$

and this is a contradition: since $X_{\tau_0} = 0$ by construction (being τ_0 the moment when we arrive in 0, so X_{τ_0} its a degenerate rv) and every degenerate rv has variance $= 0$.
So 0 is actually null recurrent because if it was positive recurrent (as assumed) we obtained a contradiction. □

3.3 Generalized symmetric random walk on $\mathbb{Z}^k$

Remark 49. Let's briefly generalize the symmetric random walk on $\mathbb{Z}^k$.

Definition 3.3.1 (Symmetric random walk on $\mathbb{Z}^k$)**.** Having $\mathbb{Z}^k = \underbrace{\mathbb{Z} \times \mathbb{Z} \times \ldots \times \mathbb{Z}}_{k \text{ times}} = \{(j_1, \ldots, j_k) : j_i \in \mathbb{Z}, \forall i\}$ we define

$$\mathbf{X}_n = \mathbf{X}_0 + \sum_{i=1}^{n} \mathbf{Z}_i,$$

where $\mathbf{X}_0 \perp\!\!\!\perp \mathbf{Z}_i$ and $(\mathbf{Z}_i)$ iid. Here $\mathbf{X}_0, \mathbf{Z}_i \in \mathbb{Z}^k$ so we're summing random vectors (but otherwise is the same).

Important remark 22. Looking at a single increment vector $\mathbf{Z}_i = \mathbf{Z}_1$ (in the space $\mathbb{Z}^k$) we have that its probability depends on k:

$$\mathbb{P}\left(\mathbf{Z}_1 = \mathbf{x}\right) = \frac{1}{2k}, \quad \forall \mathbf{x} \in \bigcup_{i=1}^{k} \left\{\mathbf{y} \in \mathbb{Z}^k : y_i \in \{-1, 1\}, y_j = 0, \forall j \neq i\right\}$$

The set on the right is composed by vectors $\mathbf{x}$ where one components (in turn) is either -1 or 1 and all the remainings are 0. So, for instance:

- if $k = 1$, $\mathbb{P}\left(Z_1 = 1\right) = \mathbb{P}\left(Z_1 = -1\right) = -\frac{1}{2}$ (that is in $\mathbb{Z}^1$ there are only two deltas available, $\{-1, +1\}$ each with probability 1/2, as seen before)
- if $k = 2$, $\mathbb{P}\left(\mathbf{Z}_1 = \mathbf{x}\right) = \frac{1}{4}$: in $\mathbb{Z}_2$ there are four admisible values/points/deltas are $\{(0, -1), (0, 1), (-1, 0), (1, 0)\}$ each with the same probability 1/4
- and so on ...

Remark 50 (Type of symmetric random walk on $\mathbb{Z}^k$)*.* This random walk is used often in application and gives origin to a strange phenomenon.
The symmetric random walk on $\mathbb{Z}^k$ is still **irreducible** (can be proved by exactly the same argument we done with $k = 1$), so all states are of the same type. It can be shown that:

- for $k = 1$ or $k = 2$ it is *null recurrent*;
- for $k \geq 3$ it's *transient*.

This is a curios fact: the type change drastically with the number of dimensions. If $k = 2$ and we start from a point eg, if $\mathbf{X}_0 = (i, j)^T$ then with probability 1 we will go back to this point; instead if $k \geq 3$ we start from a point but it could be that we never go back there (we don't prove it).

Remark 51*.* Exam question: can you give me an exameple of chain that is irreducible and transient? This is the only example in the course: it is the symmetric random walk on $\mathbb{Z}^k$ with $k \geq 3$.

NB: porta sta roba sopra dove piazzata meglio

Important remark 23*.* Some useful facts about states types which can be shown:

1. if state x is *recurrent* then

$$\mathbb{P}\left(X_n = x \text{ for infinitely many } n | X_0 = x\right) = 1$$

   That is, if we start from a recurrent state not only we go back to it one time almost surely, but we'll return back to it again and again almost surely.
   Idea: being in a Markov chain we forget about the past. We're sure that if we start from a point we will back to it; once this happens, you're in the state for the second time and working with a Markov chain you forget the past so you'll be back on the state a third time. An so on;

2. if state y is *transient* or *null recurrent*, then

$$\lim_{n \to +\infty} \mathbb{P}\left(X_n = y | X_0 = x\right) = 0, \forall x$$

   whichever starting point x probability to come back to it tends to go to 0 (it's not 0);

3. if S *is finite* (remember S can be finite or countable) then

   - there are no null recurrent state ($\nexists$)
   - exists at least one positive recurrent state.

   Eg in the case of symmetric rw on $\mathbb{Z}$ all states are null recurrent.

3.4 Stationarity

Definition 3.4.1 (Stationarity of a generic sequence)**.** An arbitrary sequence $(X_0, X_1, X_2, \dots)$ (not necessarily a Markov chain) is said to be stationary if the probability distribution doesn't change if we shift the sequence by 1, eg

$$(X_1, X_2, X_3, \dots) \sim (X_0, X_1, X_2, \dots)$$

Important remark 24 (Stationarity nd identically distributed processes)*.* If shifting by one doesn't change the distribution then even if we start from the shifted sequence. So by induction for a stationary sequence if we shift by k we still get the same distribution

$$(X_k, X_{k+1}, X_{k+2}, \dots) \sim (X_0, X_1, X_2, \dots)$$

It follows that since the two sequences above have the same distribution, the first element of the first sequence is distributed as the first element of the second sequence and so

$$X_0 \sim X_k, \quad \forall k$$

Hence if (X_n) is stationary it implies that is identically distributed.
The converse does not necessary hold: if a sequence is identically distributed does not need to be stationary.

Remark 52*.* Let's see now the concept of stationarity applied to a Markov chain.

Definition 3.4.2 (Stationary distribution for a Markov chain (X_n))**.** Let (X_n) be a Markov chain with state space S. Let π be a probability on S. Since S is finite or countable we have:

$$\pi(x) \geq 0, \forall x \in S$$
$$\sum_{x \in S} \pi(x) = 1$$

Then π is said to be a stationary distribution for (X_n) if

$$X_0 \sim \pi \implies X_1 \sim \pi$$

Remark 53*.* Recall that in order to describe the probability distribution of a Markov chain (any event concerning the sequence) we need the kernel α and the probability distribution of X_0. Now if the condition above holds, that is $X_0 \sim \pi \implies X_1 \sim \pi$, then π is said to be a stationary distribution for (X_n) because of the following theorem.

Theorem 3.4.1 (Stationary Markov chain)**.** *Let (X_n) be a Markov chain; if $X_0 \sim \pi$ and π is a stationary distribution ($X_0 \sim \pi \implies X_1 \sim \pi$), then the sequence $(X_0, X_1, \dots)$ is stationary.*

Remark 54*.* If π is a stationary distribution then, by definition

$$\mathbb{P}(X_0 = b) = \mathbb{P}(X_1 = b), \quad \forall b \in S$$

On the other hand

$$\begin{aligned}\mathbb{P}(X_1 = b) &= \sum_{a \in S} \mathbb{P}(X_0 = a) \cdot \mathbb{P}(X_1 = b | X_0 = a) \\ &= \sum_{a \in S} \pi(a) \cdot \alpha(a, b)\end{aligned}$$

Hence we have proved π is a stationary distribution if and only if $\iff$

$$\pi(b) = \sum_{a \in S} \pi(a)\alpha(a, b), \quad \forall b \in S$$

In practice to check whether a distribution π is stationary we have to check this latter equation.

Important remark 25. Stationary distribution are interesting objects: but given a Markov chain we're not sure a stationary distribution exists. And thus is useful a thm which explain in which situations a stationary distribution exists.

Theorem 3.4.2. *A stationary distribution exists $\iff$ a positive recurrent state exists. Moreover if the Markov chain is irreducible and a stationary distribution exists, then it is unique.*

Corollary 3.4.3. *If S is finite, there's at least one positive recurrent state therefore there is certainly a stationary distribution*

Example 3.4.1 (Trivial but useful example)**.** Suppose S consist of two point

$$S = \{a, b\}, \quad \alpha(a, b) = \alpha(b, a) = 1$$

In this Markov chain we go from one point to the other with probability 1, and then we go back with probability 1.
Is the chain irreducible? yes. Furthermore S is finite so there's one and only one stationary distribution.
Who is the unique stationary distribution? In this example the unique stationary distribution could be $\pi(a) = \pi(b) = 1/2$ (it's a conjecture, we don't expect the two states have different probability). To verify it:

$$\sum_{x \in S} \pi(x) \cdot \alpha(x, a) = \pi(a) \underbrace{\alpha(a, a)}_{0} + \underbrace{\pi(b)}_{\frac{1}{2}} \underbrace{\alpha(b, a)}_{1} = \frac{1}{2} = \pi(a)$$

Similarly

$$\sum_{x \in S} \pi(x) \cdot \alpha(x, b) = \frac{1}{2} = \pi(b)$$

Example 3.4.2 (Ehrenfest urn, continued)**.** Let X_n be the number of balls in box 1 at time n, out of total r balls among the two boxes. We have that $S = \{0, 1, 2, \ldots, r\}$. The kernel is

$$\alpha(i, j) = \mathbb{P}(X_1 = j | X_0 = i) = \begin{cases} \frac{i}{r} & \text{if } j = i - 1 \text{ (draw from box 1)} \\ 1 - \frac{i}{r} & \text{if } j = i + 1 \text{ (draw from the other)} \\ 0 & \text{otherwise} \end{cases}$$

S is finite and the chain is irreducible; consequently there exists one (and it's unique) stationary distribution π. This time it's not easy/obvious which is it but it is:

$$\pi(j) = \frac{\binom{r}{j}}{2^r}$$

We now verify that it is actually the stationary distribution. Again π is a special distribution $\iff$

$$\pi(x) = \sum_{a \in S} \pi(a) \cdot \alpha(a, x), \quad \forall x \in S$$

In practice how can we write

$$\mathbb{P}(X_1 = b) = \sum_{a \in S} \pi(a) \cdot \alpha(a,b) = \sum_{a \in S} \mathbb{P}(X_0 = a)\, \mathbb{P}(X_1 = b | X_0 = a)$$

In case of ehrenfest urn we have:

$$\begin{aligned} \pi(b) &= \frac{\binom{r}{b}}{2^r} \\ \alpha(a,b) &= 0 \quad \text{if } a \notin \{b-1, b+1\} \end{aligned}$$

and thus we get

$$\begin{aligned} \sum_{a \in S} \pi(a)\alpha(a,b) &= \pi(b-1)\alpha(b-1,b) + \pi(b+1)\alpha(b+1,b) \\ &= \frac{\binom{r}{b-1}}{2^r}\left(1 - \frac{b-1}{r}\right) + \frac{\binom{r}{b+1}}{2^r}\frac{b+1}{r} \\ &= \frac{1}{2^r}\left[\frac{r!}{(b-1)!(r-b+1)!}\frac{r-b+1}{r} + \frac{r!}{(b+1)!(r-b-1)!}\frac{b+1}{r}\right] \\ &= \frac{1}{2^r}\left[\frac{(r-1)!}{(b-1)!(r-b)!} + \frac{(r-1)!}{(b)!(r-b-1)!}\right] \\ &= \frac{1}{2^r}\left[\frac{(r-1)!}{(b-1)!(r-b)(r-b-1)!} + \frac{(r-1)!}{(b)(b-1)!(r-b-1)!}\right] \\ &= \frac{1}{2^r}\frac{(r-1)!}{(b-1)!(r-b-1)!}\underbrace{\left[\frac{1}{r-b} + \frac{1}{b}\right]}_{\frac{r}{b(r-b)}} \\ &= \frac{1}{2^r}\frac{r!}{b!(r-b)!} \\ &= \frac{\binom{r}{b}}{2^r} \\ &= \pi(b) \end{aligned}$$

3.5 Ergodicity

Definition 3.5.1 (Ergodic Markov chain)**.** A Markov chain is said to be ergodic if there is a probability π on S such that:

$$\pi(y) = \lim_n \mathbb{P}(X_n = y | X_0 = x), \quad \forall x, y \in S$$

Important remark 26*.* Two remarks:

1. if the Markov chain is ergodic then can be shown that π is the *only stationary distribution*;

2. interpretation: if the MC is ergodic probability of arriving to the state y (somewhere) in the limit does not depend on the state from which we started (starting point).

Remark 55. Now we'll see some results: first a criterion that implies that the chain is actually ergodic. Before that a definition.

Definition 3.5.2 (Aperiodic state)**.** A state x is said to be aperiodic if the set of n such that

$$\{n : \mathbb{P}\left(X_n = x | X_0 = x\right) = 0\}$$

is finite.

(Paused) A: 00:29:31 / 01:30:31 (33

Important remark 27. If the chain is irreducible then either all states are aperiodic or all states are not aperiodic: that is all the states are of the same type even considered this kind of classification).

Example 3.5.1. The symmetrical random walk on $\mathbb{Z}$ is aperiodic or not? It's not, let's test it with state 0. If we start from 0 we can go back to 0 at odds times so the set of n such that:

$$\{n : \mathbb{P}\left(X_n = 0 | X = 0\right) = 0\} = \{n\colon\ n \text{ is odd}\}$$

which is not finite. So 0 is not aperiodic state: then being chain irreducible, all states are not aperiodic.

Theorem 3.5.1 (Sufficient condition for ergodicity)**.** *If a chain is irreducible, aperiodic and positive recurrent, then it is ergodic.*

NB: se mai riscrivi/riordina

Example 3.5.2. The symmetric RW on integers is not ergodic: the definition on ergodicity the probability π we should obtain in the lim is the unique stationary distribution, but the symmetric random walk on the integers does not have any stationary distribution because all states are null recurrent. And the existence of a stationary distribution is equivalent to the existence of a positive recurrent state. So being all states null recurrent there are no positive recurrent. Thus the symmetric rw does not admit any stationary distribution, and in particular is not ergodic, because if it was ergodic the probability π would be the unique distribution.

Capitolo 4

Brownian motions

Remark 56. So far martingales and Markov chains are discrete time process; brownian motions (next in line) are the first continuous stochastic processes we consider.

4.1 Introduction to continuous time process

Definition 4.1.1. A process X is said to be a continuous time process if the indexing set T is an interval on the real line.

Remark 57. For instance an important set is $T = [0, +\infty)$

Remark 58. Now three definition.

Definition 4.1.2. A filtration is now a collection of σ-fields $(\mathscr{F}_t : t \geq 0)$ such that $\mathscr{F}_s \subset \mathscr{F}_t \subset \mathscr{A}$, $\forall 0 \leq s < t$.

Remark 59. Interpretation is the same: $\mathscr{F}_t$ is our information at time t. In most situations is reasonable to assume that the information increases.

Definition 4.1.3 (Stopping time)**.** It's a function $\tau : \Omega \to [0, +\infty]$ such that the set $\{\tau \leq t\} \in \mathscr{F}_t$, $\forall t \geq 0$.

Remark 60. Interpretation is the same: time when something of interest happens (if $+\infty$ the event we're waitinig for doesn't happen)

Definition 4.1.4 (Martingale as cont. time process)**.** X is a martingale if

1. $\mathbb{E}\left[|X_t|\right] < +\infty$, $\forall t \geq 0$
2. X is adapted
3. $\mathbb{E}\left[X_t | \mathscr{F}_s\right] = X_s$, $\forall 0 \leq s < t$

Remark 61. That is the notion extend naturally/the same way. Now we're ready for the definition of interest

4.2 Brownian motion

Definition 4.2.1 (Brownian motion)**.** A brownian motion with respect to a filtration (F_t) is a continuous time process X such that:

1. X is adapted
2. X has *indipendent increments*
3. $X_0 = 0$ almost surely and $X_t \sim \mathrm{N}\left(0, \sigma^2 t\right)$, $\forall t > 0$, where $\sigma^2 > 0$ is a parameter

Remark 62. What does it mean indipendent increments?

Definition 4.2.2 (Process with independent increments)**.** A process X has independent increments if

$$(X_t - X_s) \perp\!\!\!\perp \mathscr{F}_s, \forall 0 \leq s < t$$

Remark 63. Interpretation: suppose to be at time s, then $X_t - X_s$ can be called the increment from s to t; the condition implies that the increment is indipendent of all the information we have up to time s.
At time s we can predict the increment ignoring all the stuff before now.

Remark 64 (History). During '800 Brown, a botanist, describe the motion of pollen in a fluid; 30 years after, Wiener describe a mathematical model for that motion. For this reason brownian motion can be called Wiener process as well.

Remark 65 (Wiener process). Wiener process is a triplet $(X_t, Y_t, Z_t) : t \geq 0$, where X, Y, Z are independent brownian motion (we need 3 coordinates to describe the motion in the space)

4.3 Paths of BM

Definition 4.3.1 (Path (refresher))**.** Recall that, for any process X the paths are functions $t \to X(t, \omega)$ for fixed $\omega \in \Omega$

Remark 66. Exam question: talk about the path of brownian motions (then we say the things below)

Important remark 28 (Paths of a brownian motion and their main properties). In case of brownian motion, the paths are functions from $[0, +\infty) \to \mathbb{R}$ with the following main properties:

1. the path are *continuous* ...
2. ... and *nowhere differentiable* (a function $f : I \to \mathbb{R}$ is said to be nowhere differentiable if, $\forall t \in I$, f is not differentiable at t) ...

... *up to equivalence.*

Remark 67. So it's something extremely erratic, hard to make a picture.

Remark 68. What does it means up to equivalence?

Definition 4.3.2 (Property satisfied up to equivalence)**.** A process X is said to satisfy some property up to equivalence if there is another process Z such that:

1. $\mathbb{P}(X_t \neq Z_t) = 0, \forall t$ (Z is "close" to X, in this sense)

2. Z satisfies the property

Remark 69 (Interpretation). Actually, it may be that X does *not* satisfy the property. However there's another process which satisfies it and is very close to X (in this case we can "throw away" X and use Z). So if we have a brownian motion we can say its path are continuous and nowhere differentiable tout court; but there's a process similar to it which satisfies the conditions.

4.4 Properties of BM

Definition 4.4.1 (Process with stationary increments)**.** In general a process X having *independent increments* is said to have *stationary increments* if the distribution of the increment $(X_t - X_s)$ is invariant under shifts of time which means that

$$(X_t - X_s) \sim (X_{t-s} - X_0), \forall 0 \leq s < t$$

Proposition 4.4.1 (Stationarity of increments of a BM)**.** *The BM has* stationary increments*; hence since $X_0 = 0$ almost surely one obtain*

$$X_t - X_s \sim X_{t-s} \sim \mathrm{N}\left(0, \sigma^2(t-s)\right)$$

Dimostrazione. Omitted. □

Proposition 4.4.2. *Let $Y_t = X_t^2 - \sigma^2 t$. If X is brownian motion then X and Y are both martingales.*

Dimostrazione. In this course when we want to prove that something is a martingale we prove only the third properties of their definition (taking the first two as given). So to prove

$$\mathbb{E}\left[X_t | \mathscr{F}_s\right] = X_s, \quad \forall 0 \leq s < t$$

fix s with $0 \leq s < t$ then

$$\begin{aligned}\mathbb{E}\left[X_t | \mathscr{F}_s\right] &= \mathbb{E}\left[X_s + (X_t - X_s) | \mathscr{F}_s\right] \overset{(1)}{=} X_s + \mathbb{E}\left[X_t - X_s | \mathscr{F}_s\right] \\ &\overset{(2)}{=} X_s + \mathbb{E}\left[X_t - X_s\right] = X_s + (\underbrace{\mathbb{E}\left[X_t\right] - \mathbb{E}\left[X_s\right]}_{=0} = X_s\end{aligned}$$

where in

- (1) since it's adapted

- (2) since process has independent increments $X_t - X_s \perp\!\!\!\perp \mathscr{F}_s$ so we drop conditioning

So we have proved that brownian motion X is a martingale.
Now we prove that even Y is a martingale: fix s such that $0 \leq s < t$ and

consider the expectation:

$$\begin{aligned}
\mathbb{E}\left[Y_t|\mathscr{F}_s\right] &= \mathbb{E}\left[X_t^2 - \underbrace{\sigma^2 t}_{\text{constant}} |\mathscr{F}_s\right] = -\sigma^2 t + \mathbb{E}\left[X_t^2|\mathscr{F}_s\right] \\
&= -\sigma^2 t + \mathbb{E}\left[(X_t + X_s - X_s)^2|\mathscr{F}_s\right] \\
&= -\sigma^2 t + \mathbb{E}\left[X_s^2 + (X_t - X_s)^2 + 2X_s(X_t - X_s)|\mathscr{F}_s\right] \\
&\overset{(1)}{=} -\sigma^2 t + X_s^2 + \mathbb{E}\left[(X_t - X_s)^2|\mathscr{F}_s\right] + 2X_s\,\mathbb{E}\left[X_t - X_s|\mathscr{F}_s\right] \\
&\overset{(2)}{=} -\sigma^2 t + X_s^2 + \underbrace{\mathbb{E}\left[(X_t - X_s)^2\right]}_{\sigma^2(t-s)} + 2X_s \underbrace{\mathbb{E}\left[X_t - X_s\right]}_{=0} \\
&\overset{(3)}{=} X_s^2 - \sigma^2 t + \sigma^2(t-s) \\
&= X_s^2 - \sigma^2 s \\
&= Y_t
\end{aligned}$$

where in:

- (1) i split the conditioning and take out X_s^2 because is measurable with respect to $\mathscr{F}_s$
- (2) because of independent increments I drop conditioning
- (3) $\mathbb{E}\left[(X_t - X_s)^2\right] = \sigma^2(t-s)$ since $X_t - X_s$ is normal with mean 0 and variance $\sigma^2(t-s)$ so the expected value $\mathbb{E}\left[(X_t - X_s)^2\right]$ is basically the variance of $X_t - X_s$

□

Proposition 4.4.3 (Covariance between terms of a BM)**.** *We have:*

$$\operatorname{Cov}(X_s, X_t) = \sigma^2 \cdot (s \wedge t)$$

Dimostrazione. Suppose $0 \le s \le t$: then

$$\begin{aligned}
\operatorname{Cov}(X_s, X_t) &= \mathbb{E}\left[X_s X_t\right] - \mathbb{E}\left[X_s\right]\mathbb{E}\left[X_t\right] \overset{(1)}{=} \mathbb{E}\left[X_s X_t\right] \\
&= \mathbb{E}\left[X_s(X_s + X_t - X_s)\right] \\
&\overset{(2)}{=} \underbrace{\mathbb{E}\left[X_s^2\right]}_{\sigma^2 s} + \underbrace{\mathbb{E}\left[X_s(X_t - X_s)\right]}_{0} \\
&= \sigma^2 s \\
&\overset{(3)}{=} \sigma^2(s \wedge t)
\end{aligned}$$

- where in (1) both $\mathbb{E}\left[X_s\right], \mathbb{E}\left[X_t\right]$ are null;
- in (2) since X_s has null expectation, the $\mathbb{E}\left[X_s^2\right]$ is the variance, and by definition of brownian motion is $\sigma^2 s$. On the other hand, the covariance between X_s and the increment is zero for independence increment (the covariance would actually be $\mathbb{E}\left[X_s(X_t - X_s)\right] - \mathbb{E}\left[X_s\right]\mathbb{E}\left[X_t - X_s\right]$ but the last part is null);
- in (3) we have supposed that $s \le t$ thus $\sigma^2 s = \sigma^2(s \wedge t)$.

□

4.5 Alternative definitions of BM

Remark 70. Above we've presented the official (most important/popular) definition of brownian motion: however there are many other equivalent definitions. We see three more of them.

Remark 71 (Canonical filtration of X). These 3 definitions refers to the filtration

$$\mathscr{F}_t^* = \sigma(X_s : s \leq t)$$

Such $\mathscr{F}_t^*$ is said to be the *canonical filtration* of X, and is the least filtration which makes X adapted. Suppose we work with this filtration: the only information we have at time t is the path of the process up to time t.

Remark 72. With the official definition given before we can choose any filtration $(\mathscr{F}_t)$ (so in a sense is more general) while with the following we're oblied with this filtration $(\mathscr{F}_t^*)$.

Definition 4.5.1 (Alternative definitions). Let X be any process such that $X_0 = 0$ almost surely. Then X is brownian motion if and only iff (any of the following equivalent situation occurs):

1. X is gaussian, $\mathbb{E}[X_t] = 0$, $\text{Cov}(X_s, X_t) = \sigma^2(s \wedge t)$, $\forall s, t$
2. X has stationary independent increments and continuous paths up to equivalence
3. X and Y are both martingale where $Y_t = X_t^2 - \sigma^2 t$ and X has continuous paths up to equivalence

Remark 73. Hence we can define BM in any of these above if we prefer.

Remark 74. Regarding the first, what does it mean that a process X is Gaussian?

Definition 4.5.2 (Gaussian process). A process is said to be gaussian if all finite dimension distribution are normal that is

$$(X_{t_1}, \ldots, X_{t_n}) \sim N, \quad \forall n \geq 1, \forall t_1, \ldots, t_n \in T$$

Remark 75. So it's gaussian if all the finite dimensional distribution are normal. Gaussian means not only $X \sim N$, $\forall t$: even bivariate and multivariate random vector should be normal.

Definition 4.5.3 (Standard brownian motion). A brownian motion process is said standard if $\sigma^2 = 1$.

Example 4.5.1. Let X be a standard brownian motion; define

$$\begin{aligned} &Y_0 = 0, \\ &Y_t = tX_{\frac{1}{t}}, \quad \forall t > 0 \end{aligned}$$

than Y is still a standard brownian motion.
To prove this fact we can use any of the alternative definition of BM: but in this example its convenient to use the first one To prove the various parts of the first definition:

- to prove it is gaussian: we can write Y as a linear transformation of normally distributed vector

$$\begin{bmatrix} Y_{t_1} \\ \dots \\ Y_{t_n} \end{bmatrix} = \begin{bmatrix} t_1 & 0 & \dots & 0 \\ 0 & t_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & t_n \end{bmatrix} \underbrace{\begin{bmatrix} X_{\frac{1}{t_1}} \\ X_{\frac{1}{t_2}} \\ \dots \\ X_{\frac{1}{t_n}} \end{bmatrix}}_{\text{normally distr.}}$$

  So being Y a linear transformation of a normally distributed vector, any linear transformation of normal is normal;

- regarding the expected value

$$\mathbb{E}[Y_t] = \mathbb{E}\left[tX_{\frac{1}{t}}\right] = t\underbrace{\mathbb{E}\left[X_{\frac{1}{t}}\right]}_{=0} \overset{(1)}{=} 0$$

  where in (1) the expected value is 0 because of assumptions (standard brownian motion).

- finally regarding the covariance

$$\begin{aligned} \operatorname{Cov}(Y_s Y_t) &= \mathbb{E}[Y_s Y_t] - \mathbb{E}[Y_s]\,\mathbb{E}[Y_t] = \mathbb{E}[Y_s Y_t] \\ &= \mathbb{E}\left[stX_{\frac{1}{s}}X_{\frac{1}{t}}\right] = st\,\mathbb{E}\left[X_{\frac{1}{s}}X_{\frac{1}{t}}\right] \overset{(2)}{=} st\left(\frac{1}{s} \wedge \frac{1}{t}\right) \\ &\overset{(3)}{=} \frac{st}{\max(s,t)} \overset{(4)}{=} \frac{\max(s,t)\min(s,t)}{\max(s,t)} = \min(s,t) \\ &= s \wedge t \end{aligned}$$

  where:

  - in (2) being a standard brownian motion, the covariance is the minimum between $\frac{1}{s}$ and $\frac{1}{t}$;
  - in (3) chossing the minimum between ratio is like choosing the maximum at the denominator;
  - in (4) we can rewrite the product of s and t as the product of the minimum times the maximum of them, whichever they are.

4.6 Brownian motion and strong law of large numbers

Remark 76. An important property of BM is that it satisfies the strong law of large number.

Theorem 4.6.1. *A brownian motion X satisifies the strong law of large numbers in the sense that*

$$\frac{X_t}{t} \xrightarrow{a.s.} 0, \quad \textit{for } t \to +\infty$$

Dimostrazione. We prove this fact in a special case only that is in the case where $tto+\infty$ along the integers $\mathbb{N}$; namely we consider $\frac{X_n}{n}$ instead of $\frac{X_t}{t}$.
We have that X_n/n can be written as

$$\frac{X_n}{n} \overset{(1)}{=} \frac{X_n - X_0}{n} \overset{(2)}{=} \frac{\sum_{i=1}^{n}(X_i - X_{i-1})}{n}$$

where:

- in (1) since $X_0 = 0$ in BM
- in (2) to see it, we have that

$$\sum_{i=1}^{n}(X_i - X_{i-1}) = X_1 - X_0 + X_2 - X_1 + \ldots + X_n - X_{n-1} = X_n - X_0$$

But since $X_i - X_{i-1} \sim \mathrm{N}\left(0, \sigma^2 \cdot 1\right)$ are iid (because BM stationary independent increment) then we can apply kolmogorov strong law of large number. Thus

$$\frac{\sum_{i=1}^{n}(X_i - X_{i-1})}{n} \xrightarrow{a.s.} \underbrace{\mathbb{E}\left[X_1 - X_0\right]}_{\text{common mean of } X_i - X_{i-1}} = 0$$

□

Let X be a standard brownian motion ($\sigma^2 = 1$). Fix a positive constant $a \in \mathbb{R}$, $a > 0$ and define a stopping time in this way

$$\tau = \inf\{t : X_t = a\}$$

τ is the first time such that the path of the brownian motion reaches the level a. Our gol is to determine the distribution function of τ.
Obviously if $t \leq 0$ then $\mathbb{P}(\tau \leq t) = 0$; so let $t > 0$. We can argue as follow; the probability can be written as:

$$\begin{aligned}
\mathbb{P}(\tau \leq t) &= \mathbb{P}(\tau \leq t, X_t = a) + \mathbb{P}(\tau \leq t, X_t < a) + \mathbb{P}(\tau \leq t, X_t > a) \\
&\overset{(2)}{=} 0 + \mathbb{P}(\tau \leq t, X_t < a) + \mathbb{P}(\tau \leq t, X_t > a) \\
&\overset{(3)}{=} 2\,\mathbb{P}(\tau \leq t, X_t > a)
\end{aligned}$$

where

- in (1) we split the event in three incompatible events for which we can sum the probability
- in (2) we have that

$$\mathbb{P}(\tau \leq t, X_t = a) \leq \mathbb{P}(X_t = a) = 0$$

  where the first inequality since the event $\{\tau \leq t, X_t = a\} \subseteq \{X_t = a\}$ and furthermore $\mathbb{P}(X_t = a) = 0$ (being X_t normal it's absolutely continuous and therefore the probability of a single point is null).

- in (3) it can be shown that

$$\mathbb{P}(\tau \leq t, X_t < a) = \mathbb{P}(\tau \leq t, X_t > a)$$

Now, since X has continuous path (up to equivalence), with $X_0 = 0$ and $X_t > a$, there must exist a point $s \in [0, t)$, such that $X_s = a$: hence $\tau \leq s < t$ (τ is the first time the path of the brownian motion reach a). In other terms the event $\{X_t > a\} \subset \{\tau \leq t\}$ and so

$$\begin{aligned}\mathbb{P}(\tau \leq t) &= 2\,\mathbb{P}(\tau \leq t, X_t > a) = 2\,\mathbb{P}(X_t > a) \overset{(1)}{=} 2\,\mathbb{P}\left(\frac{X_t}{\sqrt{t}} > \frac{a}{\sqrt{t}}\right)\\ &= 2\,\mathbb{P}\left(\mathrm{N}(0,1) > \frac{a}{\sqrt{t}}\right) = 2\left[1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right]\end{aligned}$$

where in (1) we merely divide both terms by $\sqrt{t}$ and in (2) we note that being X_t normal with mean 0 and variance t), then $X_t/\sqrt{t}$ is distributed as standard normal.
To sum up the distribution is the following:

$$\mathbb{P}(\tau \leq t) = 2\left[1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right]$$

So far $a > 0$ was a positive constant. Suppose now that $a < 0$; in t his case

$$\tau = \inf\{t : X_t = a\} = \inf\{t : -X_t = -a\} \overset{(1)}{=} \inf\{t : -X_t = |a|\}$$

where in (1) since if $a < 0$, $-a$ is $|a|$. Here

1. $-X$ is still a standard brownian motion (from the definition, is still adapted, take value 0 at 0, has independent increments, and $-X_t$ is normal with mean 0 and variance t). So similarly to the fact that if $Z \sim \mathrm{N}(0,1)$ then $-Z \sim \mathrm{N}(0,1)$ here both X and $-X$ are standard brownian motion.

2. with $|a|$ positive

we can repeat the same argument as before, and we end up with an equation valid for all cases

$$\mathbb{P}(\tau \leq t) = 2\left[1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)\right], \quad \forall t > 0, \forall a \neq 0$$

Here if $a > 0$ $|a|$ doesn't change, but if $a < 0$ absolute value is needed.

Example 4.6.1. Let us evaluate the probability of τ being finite (that is the probability that sooner or later the path of the brownian motion reach the level a):

$$\mathbb{P}(\tau < +\infty) = \lim_{t \to +\infty} \mathbb{P}(\tau \leq t) = \lim_{t \to +\infty} 2\left[1 - \Phi\left(\frac{\alpha}{\sqrt{t}}\right)\right] = 2\left[1 - \Phi(0)\right] = 1$$

So almost surely the brownian motion reach the value a (sooner or later). If X describes the price of an action, we fix a level $a \neq 0$ and decide that when it will become a then we sell. By this theorem this will happen, maybe it will take gozilion of years, but will happen

Remark 77. Every time we know that something will happen almost surely, we are interested how much time will be needed.

Let us evaluate $\mathbb{E}[\tau]$. Infact $\mathbb{E}[\tau]$ provides a rough estimate of a time required to reach a. τ is a random variable so the time to reach is random. Being continuous, the density of τ is the first derivative of the distribution function

$$\begin{aligned}\frac{\partial}{\partial t}\mathbb{P}(\tau \le t) = \frac{\partial}{\partial t}2\left[1-\Phi\left(\frac{|a|}{\sqrt{t}}\right)\right] &= -2\Phi'\left(\frac{|a|}{\sqrt{t}}\right)\cdot|a|\left(-\frac{1}{2}t^{-3/2}\right)\\ &\overset{(1)}{=} \frac{|a|}{t^{3/2}}\cdot\frac{e^{-\frac{1}{2}\frac{a^2}{t}}}{\sqrt{2\pi}}\end{aligned}$$

where in (1), $\Phi'\left(\frac{|a|}{\sqrt{t}}\right)$ is merely the density of a standard normal evaluated in $|a|/\sqrt{t}$.
Hence being the latter the density, to evaluate the expected value

$$\mathbb{E}[\tau] = \int_0^{+\infty} t\cdot f(t)\,\mathrm{d}t = \int_0^{+\infty} t\cdot\frac{|a|}{t^{3/2}}\cdot\frac{e^{-\frac{1}{2}\frac{a^2}{t}}}{\sqrt{2\pi}} = \frac{|a|}{\sqrt{2\pi}}\int_0^{+\infty}\frac{t}{t^{3/2}}e^{-\frac{a^2}{2}\frac{1}{t}}\,\mathrm{d}t \overset{(1)}{=} +\infty$$

where in (1) if $t \to +\infty$ the term $e^{-\frac{a^2}{2}\frac{1}{t}} \to 1$ while $\int_0^{+\infty}\frac{1}{\sqrt{t}}\,\mathrm{d}t = +\infty$. So to sum up:

$$\begin{aligned}&\mathbb{P}(\tau < +\infty) = 1\\ &\mathbb{E}[\tau] = +\infty\end{aligned}$$

The time to reach a is very large. we will go in a but the time to arrive there is large. Using the language of Markov chain state we could say that a is null recurrent state.

Remark 78. A last fact on brownian motion.

Proposition 4.6.2. *Fix $t > 0$ and we subdivide the interval $[0, t]$ in this way*

$$0 = t_0^n < t_1^n < t_2^n < \ldots < t_{k_n}^n = t$$

if X is standard brownian motion and

$$\max_j(t_j^n - t_{j-1}^n) \to 0, \quad \text{as } n \to \infty$$

then

$$\sum_{j=1}^{k_n}\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2 \xrightarrow{L_2} t$$

so the convergence is to the length of the interval

Remark 79. If X is a standard brownian motion and the length of the subinterval goes to 0 as n goes to infty, we can say that if we take the sum of the increments to the power 2, then this goes to t in L_2.
For instance, let $t = 1$ and $t_0^n = 0$ and $t_j^n = \frac{j}{n}$ for $j = 1, \ldots, n$; that is for each n I split the unit interval into n subintervals of equal length. Then

$$\max_j\left(t_j^n - t_{j-1}^n\right) = \max_j\frac{1}{n} = \frac{1}{n} \to 0$$

here we can say that

$$\sum_{j=1}^{n}\left(X_{\frac{j}{n}}-X_{\frac{j-1}{n}}\right)^2 \xrightarrow{L_2} 1$$

where 1 is the length of the interval

Remark 80. Da spostare in probabilità la roba di sotto, se mancante

Proposition 4.6.3. *In general* $Y_n \xrightarrow{L_2} c$ *with* $c \in \mathbb{R}$, *if and only if both* $\mathbb{E}[Y_n] \to c$ *and* $\operatorname{Var}[Y_n] \to 0$.

Dimostrazione. Infact

$$\begin{aligned}
\mathbb{E}\left[(Y_n-c)^2\right] &\overset{(1)}{=} \mathbb{E}\left[\left((Y_n-\mathbb{E}[Y_n])+(\mathbb{E}[Y_n]-c)\right)^2\right]\\
&= \underbrace{\mathbb{E}\left[(Y_n-\mathbb{E}[Y_n])^2\right]}_{\operatorname{Var}[Yn]}+\underbrace{\mathbb{E}\left[(\mathbb{E}[Y_n]-c)^2\right]}_{(\mathbb{E}[Y_n]-c)^2}+\mathbb{E}\left[2(Y_n-\mathbb{E}[Y_n])(\mathbb{E}[Y_n]-c)\right]\\
&= \operatorname{Var}[Y_n]+(\mathbb{E}[Y_n]-c)^2+2(\mathbb{E}[Y_n]-c)\underbrace{\mathbb{E}[Y_n-\mathbb{E}[Y_n]]}_{=0}\\
&= \operatorname{Var}[Y_n]+(\mathbb{E}[Y_n]-c)^2
\end{aligned}$$

Therefore $\mathbb{E}\left[(Y_n-c)^2\right]\to 0 \iff \mathbb{E}[Y_n]\to c$ and $\operatorname{Var}[Y_n]\to 0$.

□

Dimostrazione. Now to prove the theorem, we let

$$Y_n=\sum_{j=1}^{k_n}\left(X_{t_j^n}-X_{t_{j-1}^n}\right)^2$$

Then it suffices to prove that $\mathbb{E}[Y_n]\to t$ and $\operatorname{Var}[Y_n]\to 0$. So:

- regarding the expected value

$$\begin{aligned}
\mathbb{E}[Y_n] &= \sum_{j=1}^{k_n}\mathbb{E}\left[\left(X_{t_j^n}-X_{t_{j-1}^n}\right)^2\right] \overset{(1)}{=} \sum_{j=1}^{n}\left(t_j^n-t_{j-1}^n\right)\\
&= t_1^n-t_0^n+t_2^n-t_1^n+\ldots t_{k_n}^n-t_{k_{n-1}}^n\\
&= t_{k_n}^n-t_0^n=t-0\\
&= t
\end{aligned}$$

  where in (1) since X is a standard brownian motion then $X_t-X_s\sim \mathrm{N}(0,t-s)$, and since its the sum of variance of the increments (which are independent) its the sum of the variances.
  So being even equal to t is goes to t as well

- regarding variance

$$\operatorname{Var}[Y_n] = \operatorname{Var}\left[\sum_j \left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right] \overset{(1)}{=} \sum_j \operatorname{Var}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^2\right]$$

$$\overset{(2)}{\leq} \sum_j \mathbb{E}\left[\left(X_{t_j^n} - X_{t_{j-1}^n}\right)^4\right] \overset{(3)}{=} \sum_j \mathbb{E}\left[\left(\frac{X_{t_j^n} - X_{t_{j-1}^n}}{\sqrt{t_j^n - t_{j-1}^n}}\right)^4 (t_j^n - t_{j-1}^n)^2\right]$$

$$\overset{(4)}{=} \sum_j \left(t_j^n - t_{j-1}^n\right)^2 \underbrace{\mathbb{E}\left[\mathrm{N}(0,1)^4\right]}_{\text{it's 3 btw}}$$

$$\overset{(5)}{\leq} \mathbb{E}\left[\mathrm{N}(0,1)^4\right] \cdot \max_j \left(t_j^n - t_{j-1}^n\right) \cdot \underbrace{\sum_j \left(t_j^n - t_{j-1}^n\right)}_{=t}$$

$$= t \cdot \mathbb{E}\left[\mathrm{N}(0,1)^4\right] \max_j \left(t_j^n - t_{j-1}^n\right)$$

where in:

– (1) by independence of the increments (and the variance of the sum of independent random variables has no covariance)

– (2) since that in general for any rv Y we have that $\operatorname{Var}[Y] \leq \mathbb{E}\left[Y^2\right]$

– (3) we divided and multiplied the difference of the two variables for their standard deviation in order to standardize

– (4) we have that $\frac{X_{t_j^n} - X_{t_{j-1}^n}}{\sqrt{t_j^n - t_{j-1}^n}}$ is standard normal because is the difference of the variable with his mean (0) and divided by the standard deviation

– (5) to construct the upper bound of the sum of squares square we choosed the maximum value multiplied for the remaining

So looking at the final equation the term $\max_j \left(t_j^n - t_{j-1}^n\right)$ by assumption goes to 0 as $n \to +\infty$. Therefore $\operatorname{Var}[Y_n] \to 0$ as $n \to +\infty$.

$\square$

Capitolo 5

Poisson (and Levy) processes

5.1 Introduction

Definition 5.1.1 (Poisson process)**.** A process $N = \{N_t : t \geq 0\}$ indexed by non negative reals is a Poisson process if:

1. N is adapted (to some filtration $\mathscr{F}_t$);
2. N has independent increments;
3. $N_0 = 0$ almost surely and $N_t \sim \text{Pois}\,(\lambda t)$, $\forall t > 0$ where $\lambda > 0$ is a parameter

Remark 81. Some remarks:

1. N is quite similar to brownian motion. The only difference is that $N_t \sim \text{Pois}\,(\lambda \cdot t)$, $\forall t > 0$ while in case of brownian motion X we have $X_t \sim \text{N}\left(0, \sigma^2 t\right)$, $\forall t > 0$.
   The two characterization are exactly the same up to the property of the paths; then in case of BM paths are continuous (up to equivalence), while in case of Poisson process paths are not continuous tout court but satisfies the long list of properties.
2. the analogy between the two processes can be highlighted/stressed in another way.
   Let $\mathscr{F}_t^* = \sigma(X_s : s \leq t)$ (that is the canonical filtration of the process X, is the least/smaller filtration which makes the process X adapted). Then among the four charaterization of BM, the third states that X is a brownian motion iff X has independent and stationary increments, $X_0 = 0$ almost surely and X has continuous paths up to equivalence. For the poisson process holds a characterization which is similar.

Proposition 5.1.1 (Process characterization)**.** *X is a poisson process $\iff$:*

- *X has stationary and independent increments;*
- *$X_0 = 0$ almost surely (thus, so far the same condition of BM);*
- *the paths of X are (up to equivalence) increasing, right continuous, piecewise constant with unitary jumps and $X_t \xrightarrow{a.s.} +\infty$ as $t \to +\infty$.*

Important remark 29 (Shape of paths). To remind all the properties think it's actually a counting process: usually in application one uses Poisson process when we're counting random arrivals (eg at the bus stop you remain for an hour and count the number of cars that passes, or counting the visit to a website in a time interval).
Thus, the form of the path of Poisson process is a starway starting from 0 with value 0; the process remains 0 until the first arrival. Then is right continuous with unitary increases that occurs at time where an arrival occurs. Jumps are unitary because we assume there are no events that occurs in the same moment.

Remark 82 (Stationary increments and deltas distribution). It follows from the previous characterization that a Poisson process has not only independent increment (by the definition) but also has *stationary increments*. Hence

$$N_t - N_s \sim N_{t-s} - N_0$$

(by definition of stationary increments) but in case of poisson $N_0 = 0$ so

$$N_t - N_s \sim N_{t-s} - N_0 = N_{t-s} \sim \text{Pois}\,(\lambda(t-s))$$

5.2 Stopping times and counting process

Remark 83. In order to better understand that a Poisson process is actually a counting process, let's introduce the stopping times

Definition 5.2.1. In the Poisson process the stopping times are defined as

$$\begin{aligned}
&\tau_0 = 0\\
&\tau_1 = \inf\{t : N_t = 1\}\\
&\dots\\
&\tau_k = \inf\{t : N_t = k\}
\end{aligned}$$

τ_1 is the time of first arrival, τ_k is first time where we have k arrival. So, on the time line we have that $0 < \tau_1 < \tau_2 < \dots$.

Proposition 5.2.1. *If N is a poisson process the sequence $(\tau_n - \tau_{n-1} : n \geq 1)$ is iid with common distribution* $\text{Exp}\,(\lambda)$

Dimostrazione. We dont' prove time deltas are independent o identically distributed (take it as given) but we show that the common distribution of $\tau_n - \tau_{n-1}$ is actually $\text{Exp}\,(()\lambda)$.
We start from the complement to 1 of the distribution function:

$$\mathbb{P}(\tau_1 - \tau_0 > t) \overset{(1)}{=} \mathbb{P}(\tau_1 > t) \overset{(2)}{=} \mathbb{P}(N_t = 0) \overset{(3)}{=} e^{-\lambda t} \quad \forall t \geq 0$$

where in

- (1) since $\tau_0 = 0$ by construction
- (2) the event is that at time t we have no arrival (since the first occurs at τ_1 which is following t)

- (3) is the probability that a poisson with λt parameter to take value 0. Remember that if $X \sim \text{Pois}(\lambda)$ then $\mathbb{P}(X = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$ $\forall x \in \{0, 1, \ldots\}$, so $\mathbb{P}(X = 0) = e^{-\lambda}$

So with $e^{-\lambda t}$ we have the complement to 1 of the distribution function of an exponential (which is $1-e^{-\lambda t}$). Therefore since we started from the complement to 1 of the distribution function of the differences between times we can say that

$$\tau_1 - \tau_0 \sim \text{Exp}(\lambda)$$

□

Remark 84. So in case of a Poisson process, the times between two arrivals are independents and have same distribution which is an exponential with parameter λ.
The times τ are *not* independent (they're increasing, can't be indepepndent) but the deltas $(\tau_n - \tau_{n-1})$ are so $(\tau_1 - 0)$ is independent from $(\tau_2 - \tau_1)$ etc.

Definition 5.2.2 (Poisson process as counting process)**.** Using the stopping times, the process N can be written as follows: the event $\{N_t = j\}$ (at time t we have j arrivals), can be written as $\{\tau_j \leq t \leq \tau_{j+1}\}$ (where τ_j is the time of the j-th arrival). Thus we can rewrite

$$N_t = j \iff t \in [\tau_j, \tau_{j+1}], \quad \forall j = 0, 1, 2, \ldots$$

so that N_t can be regarded as

$$N_t = \{\text{number of arrivals in the interval } [0, t]\}$$

Example 5.2.1. Given $0 < s < t$, we want to evaluate the probability $P(N_s = j|N_t = n)$ where j is in $0 \leq j \leq n$

$$\begin{aligned}
\mathbb{P}(N_s = j|N_t = n) &\overset{(1)}{=} \frac{\mathbb{P}(N_s = j, N_t = n)}{\mathbb{P}(N_t = n)} \\
&\overset{(2)}{=} \frac{\mathbb{P}(N_s = j, N_t - N_s = n - j)}{\mathbb{P}(N_t = n)} \\
&\overset{(3)}{=} \frac{\mathbb{P}(N_s = j) \cdot \mathbb{P}(N_t - N_s = n - j)}{\mathbb{P}(N_t = n)} \\
&\overset{(4)}{=} \frac{\frac{e^{-\lambda s}(\lambda s)^j}{j!} \frac{e^{-\lambda(t-s)}(\lambda(t-s))^{n-j}}{(n-j)!}}{\frac{e^{-\lambda t}(\lambda t)^n}{n!}} \\
&= \binom{n}{j} \frac{s^j \cdot (t-s)^{n-j}}{t^n} \\
&= \binom{n}{j} \left(\frac{s}{t}\right)^j \left(1 - \frac{s}{t}\right)^{n-j}
\end{aligned}$$

where the last equation is a binomial with parameter $\frac{s}{t}$ and in

- (1) we apply the definition being $\mathbb{P}(N_t = n) > 0$;
- (2) we rewrite the second event;

- (3) by independent increments the two random variables at the numerator are independent;
- (4) all are poisson with different parameters due to time indexes.

Remark 85 (Interpretation). We have $N_t = n$ meaning we have n arrivals up to t, while $N_s = j$ we have j arrivals up to s (with $s < t$ and $j \leq n$). We proved that if N is a poisson process then:

$$\mathbb{P}\left(\text{we had j arrivals at } s|\text{at the end we had } n \text{ arrivals up to } t\right) = \text{binomial with parameter } \frac{s}{t}$$

In practice this means that suppose we know that $N_t = n$; then conditionally on this information we can say that the arrival times are iid and uniformly distributed on $(0, t)$.

Example 5.2.2. Suppose the value at time t is $n = 1$; we have two possibilities regarding the value at time s

$$\begin{aligned}
\mathbb{P}\left(N_s = 1|N_t = 1\right) = \mathbb{P}\left(\tau_1 \leq s|N_t = 1\right) &= \binom{1}{1}\left(\frac{s}{t}\right)^1\left(1 - \frac{s}{t}\right)^0 = 1 \cdot \frac{s}{t} \cdot 1 \\
&= \frac{s}{t} \\
\mathbb{P}\left(N_s = 0|N_t = 1\right) = \mathbb{P}\left(\tau_1 > s|N_t = 1\right) &= \binom{1}{0}\left(\frac{s}{t}\right)^0\left(1 - \frac{s}{t}\right)^1 = 1 \cdot 1 \cdot \left(1 - \frac{s}{t}\right) \\
&= 1 - \frac{s}{t}
\end{aligned}$$

So it basically depends on the ratio between, that is it's uniform, and precisely $\text{Unif}\left(0, t\right)$.
I know that there is 1 arrival up to t so it happened in interval $(0, t)$: so conditional distribution $\tau_1|N_t = 1 \sim \text{Unif}\left(0, t\right)$.
On the other hand, however, note that unconditional distribution $\tau_1 \sim \text{Exp}\left(\lambda\right)$.
Same interpretation holds if we have n arrivals from 0 to t, they are conditionally $\text{Unif}\left(0, t\right)$.

5.3 Compensated and spatial poisson process

Remark 86. Now two remarks regarding the poisson process

Important remark 30. The poisson process is not a martingale. Infact

$$\mathbb{E}\left[N_t\right] = \lambda t \neq 0 = \lambda 0 = \mathbb{E}\left[N_0\right]$$

If the poisson process would be a martingale the expectation would be constant in time. However is easy to change the process to make it a martingale (being the mean not constant we only subtract it to make it constant)

Proposition 5.3.1 (Compensated Poisson process is a martingale)**.** *Let N_t be a poisson process, the process $\widetilde{N}_t$ defined by subtracting from the original process the expected value $\mathbb{E}\left[N_t\right]$:*

$$\widetilde{N}_t = N_t - \lambda t$$

is called compensated poisson process *and is a martingale.*

Dimostrazione. Given $0 \leq s < t$

$$\begin{aligned}
\mathbb{E}\left[\widetilde{N}_t|\mathscr{F}_s\right] &= \mathbb{E}\left[N_t - \lambda t|\mathscr{F}_s\right] = -\lambda t + \mathbb{E}\left[N_s + N_t - N_s|\mathscr{F}_s\right] \\
&\overset{(1)}{=} -\lambda t + N_s + \mathbb{E}\left[N_t - N_s|\mathscr{F}_s\right] \overset{(2)}{=} -\lambda t + N_s + \mathbb{E}\left[N_t - N_s\right] \\
&\overset{(3)}{=} -\lambda t + N_s + \lambda(t-s) = N_s - \lambda s \\
&= \widetilde{N}_s
\end{aligned}$$

where in

- (1) being N_s being $\mathscr{F}_s$ measurable we can take it out of expectation
- (2) having independent increments we can drop the conditioning
- (3) $N_t - N_s \sim \text{Pois}\left(\lambda(t-s)\right)$

□

Remark 87. The next remark is more important; it's a modification of the Poisson process which is useful in applications.

Definition 5.3.1 (Spatial Poisson process)**.** Let be $H \subset \mathbb{R}^n$ be a bounded subset; let $\mathscr{U}$ be a class of borel subset of H and consider a process

$$N = \{N(A) : A \in \mathscr{U}\}$$

N is said to be a spatial Poisson process if satisfies two properties:

1. if $A_1, \ldots, A_k \in \mathscr{U}$ are pairwise disjoint, then the corrisponding random variables $N(A_1), \ldots, N(A_k)$ are independent
2. $\forall A \in \mathscr{U}$, $N(A) \sim \text{Pois}\left(\lambda \cdot m(A)\right)$ where m is lebesgue measure. Eg in dimension $n = 1$ the meaning of lebesgue measure is the length, in dimension $n = 2$ the area, in dimension 3 the volume and so on; if we are working in dim 2 the random variable depends on a lambda and something connected to the area of A

Example 5.3.1. Associamo ad ogni elemento A di $\mathscr{U}$ una variabile aleatoria: for instance $H = ITALY$, $\mathscr{U} = \{\text{italians region}\}$, and

$$N(A) = \{\text{number of cases of some disease in a certain time interval in region A}\}$$

Suppose we want to use a poisson process in the italian region example. so we're making the two assumptions above. are these reasonable? no they're not reasonable, especially the first: the number of cases in emilia is not independent of the number of disease in tuscany.

5.4 Levy processes

Definition 5.4.1 (Levy process)**.** $X = \{X_t : t \geq 0\}$ is a levey process with respect fo a filtration $(\mathscr{F}_t)$, if

1. X is adapted to $(\mathscr{F}_t)$

2. X has stationary indipendent increments
3. $X_0 = 0$ almost surely
4. $X_s \xrightarrow{p} X_t$ as $s \to t$

Remark 88. Both brownian motion and poisson process are the two most important cases of Levy process. The first three condition are evident; we want to prove the fourth holds for both the processes

Dimostrazione. The fourth property for BM is obviously true because the paths of BM are continuous up to equivalence; if $s \to t$, $X_s \to X_t$.
Similarly, suppose that X is a poisson process, what does $X_s \xrightarrow{p} X_t$ mean? The definition is

$$\lim_{s\to t} \mathbb{P}\left(|X_s - X_t| > \varepsilon\right) = 0, \forall \varepsilon > 0$$

Now if X is poisson process the distribution of $X_s - X_t$ is Poisson and precisely

$$\mathbb{P}\left(|X_s - X_t| > \varepsilon\right) = \mathbb{P}\left(X_{|t-s|} > \varepsilon\right) \overset{(1)}{=} 1 - \mathbb{P}\left(X_{|t-s|} = 0\right) \overset{(2)}{=} 1 - e^{-\lambda|t-s|}$$

where:

- in (1) suppose that $0 < \varepsilon < 1$; the only value which does not satisfies $X_{|t-s|} > \varepsilon$ is 0;
- in (2) we substitute the probability that a poisson be 0.

Thus if $s \to t$ then int $\mathbb{P}\left(|X_s - X_t| > \varepsilon\right) \to 1 - e^0 = 0$, and this proves the fouth condition of the levy process for a poisson one. □

Remark 89. Another Levy process is the following.

Definition 5.4.2 (Compound poisson process)**.**

$$X_t = 1(N_t > 0) \sum_{j=1}^{N_t} Z_j, \quad \forall t \geq 0$$

where (Z_j) is iid, (Z_j) independent of N and N is a poisson process.

Important remark 31 (Interpretation). What is the idea of the process:

- we start with a poisson process N and we have the arrival times of a poisson process $0 < \tau_1 < \tau_2 < \dots$ (with τ_1 time of first arrival);
- in the poisson process jumps are unitary, here are iid random variables Z_j (of any kind): so the graph with the path which start at 0, at τ_1 will goes to Z_1, at τ_2 will goes to $Z_1 + Z_2$ (which differently from poisson is not needed to be $Z_1 + 1$, can be lower than Z_1 if Z_2 is negative) and so on.

Capitolo 6

Further topics

6.1 Stationary and exchangeable sequences

Remark 90. Here **we go back to discrete time process**; the following are important sequences of random variables. The goal is to discuss stationary sequences and exchangeable sequences (de Finetti).

Remark 91. Stationary is important/common assumption for time time series (statistican call time series sequences of random variables).

Definition 6.1.1 (Stationary sequences)**.** Let $(X_0, X_1, X_2, \ldots)$ be a sequence of real random variables. The sequence (X_n) is stationary if and only if the distribution doesn't change if we shift it by one, eg $(X_1, X_2, X_3, \ldots) \sim (X_0, X_1, X_2, \ldots)$

Important remark 32 (Shift by k and identically distributed sequence)*.* By induction one obtains that

$$(X_k X_{k+1}, X_{k+2}, \ldots) \sim (X_0, X_1, X_2, \ldots)$$

so we can shift the sequence by $k \geq 1$.

Remark 92. Exchangeable sequences were introduced by Bruno de Finetti; it's a nice assumption for a sequence, applied in many fields, especially bayesian statistics.

Definition 6.1.2 (Exchangeable sequences)**.** A sequence (X_n) is exchangeble by definition iff $(X_{\pi_0}, X_{\pi_1}, \ldots, X_{\pi_n}) \sim (X_0, X_1, \ldots, X_n)$ $\forall n \geq 1$, $\forall$ permutation $(\pi_0, \pi_1, \ldots, \pi_n)$ of $(0, 1, \ldots, n)$.

Example 6.1.1. So:

- if $n = 1$ it must be that $(X_0, X_1) \sim (X_1, X_0)$
- if $n = 2$ the permutation are $3! = 6$ and must be:

$$(X_0, X_1, X_2) \sim (X_0, X_2, X_1) \sim (X_2, X_0, X_1) \sim (X_1, X_0, X_2) \sim (X_1, X_2, X_0) \sim (X_2, X_1, X_0)$$

6.2 Relationship between notions of dependence

Important remark 33. For a sequence of random variable $(X_0, X_1, X_2, \ldots)$ we have many notions of dependence:

- iid
- exchangeable
- stationary
- identically distributed

The implication is downward in the list above, that is:

iid sequence $\Longrightarrow$ exchangeable seq. $\Longrightarrow$ stationary seq. $\Longrightarrow$ identically distributed seq.

6.2.1 IID and exchangeable sequences

Proposition 6.2.1. *IID sequences are exchangeable, but the converse does not hold necessary.*

Dimostrazione. Let's prove iid implies exchangeability. If (X_n) is iid:

$$\begin{aligned}
\mathbb{P}\left(X_{\pi_0} \in A_0, X_{\pi_1} \in A_1, \ldots, X_{\pi_n} \in A_n\right) &\overset{(1)}{=} \prod_{j=0}^{n} \mathbb{P}\left(X_{\pi_j} \in A_j\right) \\
&\overset{(2)}{=} \prod_{j=0}^{n} \mathbb{P}\left(X_j \in A_j\right) \\
&\overset{(3)}{=} \mathbb{P}\left(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n\right)
\end{aligned}$$

where in

1. (1) due to being independent
2. (2) to being identically distributed
3. (3) using independence again (reversed this time)

Therefore we conclude that

$$(X_{\pi_0}, X_{\pi_1}, \ldots, X_{\pi_n}) \sim (X_0, X_1, \ldots, X_n)$$

and this is the definition of exchangeability. □

Example 6.2.1 (Counterexample exchangeability $\not\Longrightarrow$ iid)**.** Now we give a counterexample with an exchangeable sequence which is not iid.
Fix a non degenerate random variable V and define $X_n = V$, forall $n \geq 0$. So the sequence is $(V, V, V, \ldots)$:

- the sequence is trivially exchangeable because if we make any permutation nothing happens

$$(X_{\pi_0}, X_{\pi_1}, \ldots, X_{\pi_n}) = (V, V, \ldots, V)$$

- however it is not true that is iid because the variable is the same and all the values are equal (più dipendente di così non si può)

6.2.2 Exchangeable and stationary sequences

Proposition 6.2.2. *Exchangeable implies stationary but the converse does not hold.*

Dimostrazione. We prove exchangeability implies stationarity. If (X_n) is exchangeable we can take:

$$(\pi_0, \pi_1, \ldots, \pi_n) = (1, 2, \ldots, n, 0)$$

getting that

$$(X_1, X_2, \ldots, X_n, X_0) \sim (X_0, X_1, \ldots, X_{n-1}, X_n)$$

In particular

$$(X_1, X_2, \ldots, X_n) \sim (X_0, X_1, \ldots, X_{n-1}), \quad \forall n \geq 1$$

and this implies

$$(X_1, X_2, X_3, \ldots) \sim (X_0, X_1, X_2, \ldots)$$

so exchangeable $\implies$ stationary. An example of a stationary sequences which is not exchangeable to conclude the proof follows $\square$

Example 6.2.2 (Counterexample stationary $\not\Longrightarrow$ exchangeable)**.** Fix a symmetric random variable V ($V \sim -V$, eg normal with mean 0) with $\mathbb{P}(V = 0) < 1$ (non degenerate on the centre). Given V we let:

$$X_n = \begin{cases} V & \text{if } n \text{ is even} \\ -V & \text{if } n \text{ is odd} \end{cases}$$

We have to prove that this sequence is stationary but not exchangeable

- to prove it is not exchangeable

$$\mathbb{P}(X_0 = X_2) = \mathbb{P}(V = V) = 1$$
$$\mathbb{P}(X_0 = X_1) = \mathbb{P}(V = -V) \overset{(1)}{=} \mathbb{P}(V = 0) < 1$$

  where in (1) $V = -V$ is possible only if $V = 0$. Hence $(X_0, X_1) \nsim (X_0, X_2)$ and thus (X_n) is not exchangeable.

- however (X_n) is stationary. Infact, at time 0 we choose the value of V (eg $V = 18$), this is the only random moment; then we go on in a deterministic way (oscillating $18, -18, 18 \ldots$). After we choose V there is nothing random. So considering both the original and the shifted distribution we have for each $ngeq1$

$$(X_0 X_1, \ldots, X_n) = (V, -V, V, -V, \ldots, (-1)^n V)$$
$$(X_1, X_2, \ldots, X_{n+1}) = (-V, V, -V, V, \ldots (-1)^{n+1} V)$$

  these two sequences have the same distribution because $V = -V$ is symmetrical. hence

$$(X_0, X_1, \ldots, X_n) \sim (X_1, X_2, \ldots X_{n+1}), \forall n$$

  thus (X_n) is stationary

6.2.3 Stationary and identically distributed sequences

Proposition 6.2.3. *Stationary implies identically distribution but the converse does not hold.*

Dimostrazione. Supposing that (X_n) is stationary, $(X_1, X_2, \ldots) \sim (X_0, X_1, \ldots)$; by induction one obtains that

$$(X_k X_{k+1}, X_{k+2}, \ldots) \sim (X_0, X_1, X_2, \ldots)$$

In particular $X_k \sim X_0, \forall k \geq 1$, so the sequence is identically distributed. □

Important remark 34. In the special case where (X_n) is a Markov chain, the two condition are equivalent: a MC is stationary $\iff$ it's identically distributed.

Example 6.2.3 (Counterexample identically distributed $\not\Longrightarrow$ stationary)**.** Take a sequence (Z_i) iid $\mathrm{N}(0,1)$. we define

$$X_0 \sim \mathrm{N}(0,1), X_0 \perp\!\!\!\perp (Z_i)$$
$$X_n = \frac{\sum_{i=1}^n Z_i}{\sqrt{n}}, \forall n \geq 1$$

Now:

- the sequence is clearly identically distributed: X_0 is normal while $\sum_{i=1}^n Z_i$ is again normal $\sim \mathrm{N}(0,n)$ (variance is the sum because independent n of variance 1). Hence $X_n = \sum_{i=1}^n (Z_i)/\sqrt{n} \sim \mathrm{N}(0,1)$ so the sequence (X_n) is identically distributed (each random variable in the sequence has the same distribution, which is a $\mathrm{N}(0,1)$).
- we want to see that it's not stationary; let's evaluate the covariance

$$\mathrm{Cov}(X_1, X_2) \stackrel{(1)}{=} \mathbb{E}[X_1 \cdot X_2] = \mathbb{E}\left[Z_1 \frac{Z_1 + Z_2}{\sqrt{2}}\right] = \frac{1}{\sqrt{2}} \mathbb{E}[Z_1(Z_1 + Z_2)]$$
$$= \frac{\mathbb{E}\left[Z_1^2\right] + \mathbb{E}[Z_1 Z_2]}{\sqrt{2}} \stackrel{(2)}{=} \frac{1+0}{\sqrt{2}}$$

  where in (1) we avoid $-\mathbb{E}[X_1]\mathbb{E}[X_2]$ (because both terms are null) and in (2) the covariance between $Z_1 Z_2$ is 0 because of their independence. Hence X_1 is not independent of X_2: the covariance is not null (if they were independent the covariance would be 0) while $X_0 \perp\!\!\!\perp X_1$ by assumption/definition. Thus,

$$(X_0, X_1) \stackrel{(1)}{\nsim} (X_1, X_2) \implies (X_n) \text{ is not stationary}$$

  in (1) because the first (X_0, X_1) are independent while (X_1, X_2) are not idependent.
  If X_n was stationary we could say by definition $(X_1, X_2, X_3, \ldots) \sim (X_0, X_1, X_2, \ldots)$; in particular the first two elements of the first sequence should have the same distribution $(X_1, X_2) \sim (X_0, X_1)$; in this case they have not the same distribution and thus we can conclude it's not stationary.

6.3 Exchangeable sequences, de Finetti representation

Remark 93. The following theorem is an important characterization of exchangeable sequences.

Theorem 6.3.1 (de Finetti representation theorem). *A sequence* $(X_0, X_1, \ldots)$ *of real rv is exchangeable if and only if:*

$$\begin{aligned}&\mathbb{P}\left(X_0 \in A_0, \ldots, X_n \in A_n\right) = \mathbb{E}_\nu\left[\nu(A_0) \cdot \ldots \cdot \nu(A_n))\right], \\ &\forall n \geq 0, \forall A_0, \ldots, A_n \in \beta(\mathbb{R})\end{aligned}$$

where ν *denotes a probability on* $\beta(\mathbb{R})$ *and* $\mathbb{E}_\nu$ *the expectation with respect to* ν*.*

Important remark 35 (Interpretation). Supposing we want to evaluate a general probability of the sequence $\mathbb{P}\left(X_0 \in A_0, \ldots, X_n \in A_n\right)$. If (X_n) was iid (it's not necessary iid but suppose for the moment it is) with common distribution ν, then this probability would be:

$$\begin{aligned}\mathbb{P}\left(X_0 \in A_0, \ldots, X_n \in A_n\right) &\overset{(1)}{=} \mathbb{P}\left(X_0 \in A_0\right) \cdot \ldots \cdot \mathbb{P}\left(X_n \in A_n\right) \\ &\overset{(2)}{=} \nu(A_0) \cdot \ldots \cdot \nu(A_n)\end{aligned}$$

with (1) because of independence and (2) because of identical distribution. The final expression is exactly the stuff within parentheses in the theorem definition. So the theorem states that if the sequence is exchageable, every time i want to evaluate a general probability of an event regarding the process, we start by evaluating it under the assumption of it being iid; then we calculate the mean $\mathbb{E}_\nu$ with respect to all the possible probability distributions ν.

Example 6.3.1. Suppose we have an urn with black and white balls, drawing with replacement, $\theta in (0,1)$ be the proportion of white ball in the urn. Let X_n be the indicator

$$X_n = \begin{cases} 1 & \text{white ball at trial } n \\ 0 & \text{otherwise} \end{cases}$$

We're making drawings with replacement so, supposing we want to evaluate the general event $\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right)$, considered that $x_1, x_2, \ldots x_n \in \{0,1\}$ and the sequence is iid due to replacement:

$$\begin{aligned}\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n\right) &= \mathbb{P}\left(X_1 = x_1\right) \cdot \ldots \cdot \mathbb{P}\left(X_n = x_n\right) \\ &= \theta^{\text{n. of white balls}} \cdot (1-\theta)^{\text{n. of black balls}} \\ &= \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}\end{aligned}$$

So far exchangeability doesn't have a role.
However suppose that θ is unknown. We need it to estimate our probability:

- we could estimate θ, considered as unknown constant/parameter, by maximum likelihood in a frequentist approach
- on the other way going bayesian we treat θ like a random variable (don't knowing it) so having a distribution object of interest.

In the second case it seems reasonable to assume it's exchangeable, thus we can apply de Finetti's theorem. Assuming there's a distribution of θ (in this case the ν of the theorem statement is θ) and we take the expected value to evaluate the probability calculated only under iid hypothesis so:

TODO: non chiaro ché qui non usa il fatt iid implica scambiabil assume

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{E}_\theta\left[\theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n x_i}\right]$$

where $\mathbb{E}_\theta$ means the expectation with respect to θ.
For instance if θ has absolutely continuous distribution with density f (which must be limited to $[0,1]$ being θ a probability), then

$$\begin{aligned}\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) &= \mathbb{E}_\theta\left[\theta^{\sum_{i=1}^n x_i}(1-\theta)^{\sum_{i=1}^n x_i}\right]\\ &= \int_0^1 \left(\theta^{\sum_{i=1}^n x_i}\cdot(1-\theta)^{\sum_{i=1}^n x_i}\right)\cdot f(\theta)\,\mathrm{d}\theta\end{aligned}$$

In this case distribution of θ, $f(\theta)$, is the so called prior.
So assuming the sequence is exchangeable and by assigning a proper prior distribution of θ I can evaluate any event regarding the sequence.

Remark 94. Two more examples of exchangeable sequences.

Example 6.3.2 (An exchageable sequence)**.** Let (Y_n) be an iid sequence and $Z \perp\!\!\!\perp (Y_n)$

$$X_n = Z + Y_n, \forall n \geq 0$$

(X_n) is not iid because all the X_n shares a common factor Z so can't be independent. However X_n is exchangeable because we can write:

$$\begin{aligned}\mathbb{P}(X_0 \in A_0, \ldots, X_n \in A_n) &\overset{(0)}{=} \mathbb{E}_z\left[\mathbb{P}(X_0 \in A_0, \ldots, X_n \in A_n | Z = z)\right]\\ &= \mathbb{E}_z\left[\mathbb{P}(Z + Y_0 \in A_0, \ldots, Z + Y_n \in A_n | Z = z)\right]\\ &= \mathbb{E}_z\left[\mathbb{P}(z + Y_0 \in A_0, \ldots, z + Y_n \in A_n | Z = z)\right]\\ &\overset{(1)}{=} \mathbb{E}_z\left[\mathbb{P}(z + Y_0 \in A_0, \ldots, z + Y_n \in A_n)\right]\\ &\overset{(2)}{=} \mathbb{E}_z\left[\mathbb{P}(z + Y_0 \in A_0)\cdot\ldots\cdot\mathbb{P}(z + Y_n \in A_n)\right]\\ &\overset{(3)}{=} \mathbb{E}_z\left[\mathbb{P}(z + Y_0 \in A_0)\cdot\ldots\cdot\mathbb{P}(z + Y_0 \in A_n)\right]\end{aligned}$$

where in

- (0) in order to prove the sequence is exchangeable, in a sense in the first step we need to regard Z as a constant; Z is not a constant, it's a rv, but if we condition on Z it becomes a constant. In probability every time we want to treat something as fixed for some reason we can condition on its values. Here we treat like a constant because in the case the sequence $z + Y_n$ becomes iid;
- (1) since $Z \perp\!\!\!\perp Y$ i drop conditioning;
- (2) the the sequence $(z + Y_n)$ is iid for every fixed value of z, and thus probability of intersection is product of marginals

- (3) being iid.

Thus (X_n) is exchangeable by de Finetti theorem (ci si poteva fermare alla penultima per vederlo direi).
Idea of the example: Y_n is iid (and thus exchangeable); Z is independent of Y_n; if I sum them exchangeability is preserved but not independence because all the sums have in common Z. The sequence $(Z + Y_n)$ is not iid, while $(z + Y_n)$ is iid for each fixed z. at the last passage it remains to make the expectation with respect to Z but this depends on the distribution of Z. however, whatever distribution of Z, we have exchangeable because $z + Y_0, z + Y_1, \dots$ is iid and thus the probability of our interest depends on Z but the probability is invariant under permutation. And the conclusion is that X_n is exchangeable

Example 6.3.3 (Polya urn)**.** Instead of drawing with or witout replacement another extraction scheme is dued to Polya.
We have an urn with $a > 0$ white and $b > 0$ black ball: I draw a ball, look at the color, put it back in the urn and add add k more ball of the same color. Let X_n be the indicator:

$$X_n = \begin{cases} 1 & \text{white ball at trial } n \\ 0 & \text{black ball at trial } n \end{cases}$$

Clearly we have that

$$\mathbb{P}\left(X_0 = 1\right) = \frac{a}{a+b}, \quad \mathbb{P}\left(X_0 = 0\right) = \frac{b}{a+b}$$

Now the conditional probability at second extraction is:

$$\mathbb{P}\left(X_1 = 1 | X_0 = x\right) = \frac{a + kx}{a+b+k}$$

where:

- at the denominator the total number of balls at second extraction: $a+b+k$
- at the numerator the number of white depends on first extraction. If $X_0 = x = 1$, we have $a + k$, while if $x = 0$, a, so we can summarize by using x.

In general we have that

$$\mathbb{P}\left(X_{n+1} = 1 | X_0 = x_0, \dots, X_n = x_n\right) = \frac{a + k\sum_{i=0}^{n} x_i}{a+b+k(n+1)}$$

We see that thanks this conditional probability, the sequence is exchangeable.
Now suppose we want to evaluate

$$\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) = \mathbb{P}\left(X_0 = x_0, \dots, X_{n-1} = x_{n-1}\right) \cdot \mathbb{P}\left(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}\right)$$

$$\begin{aligned}\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) = \mathbb{P}\left(X_0 = x_0, \dots, X_{n-2} = x_{n-2}\right) \cdot \mathbb{P}\left(X_{n-1} = x_{n-1} | X_0 = x_0, \dots, X_{n-2} = x_{n-2}\right) \cdot \\ \cdot \mathbb{P}\left(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}\right)\end{aligned}$$

A simple example

$$\begin{aligned}\mathbb{P}\left(X_0 = 0, X_1 = 1, X_2 = 1\right) &= \mathbb{P}\left(X_0 = 0, X_1 = 1\right) \cdot \mathbb{P}\left(X_2 = 1 | X_0 = 0, X_1 = 1\right) \\ &= \mathbb{P}\left(X_0 = 0\right) \cdot \mathbb{P}\left(X_1 = 1 | X_0 = 0\right) \cdot \mathbb{P}\left(X_2 = 1 | X_0 = 0, X_1 = 1\right) \\ &= \frac{b}{a+b} \cdot \frac{a}{a+b+k} \cdot \frac{a+k\cdot 1}{a+b+2k}\end{aligned}$$

depends only on a, b and j (number of white), not on x_0, x_1.
In general suppose $X_n = 1$ and $X_{n-1} = 0$ we can substitute the conditional probabilities using the formula developed above (for $n+1$)

$$\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) = \mathbb{P}\left(X_0 = x_0, \dots, X_{n-2} = x_{n-2}\right) \cdot \frac{b + k\sum_0^{n-2}(1-x_i)}{a+b+k(n-1)} \cdot \frac{a + k\sum_0^{n-1} x_i}{a+b+kn}$$

We could go this way long for all the steps, but if set $j = \sum_0^n x_i$ as the sum of white balls in the first $n+1$ extractions, in general one obtains:

$$\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) = \frac{[a+k(j-1)]\,[a+k(j-2)] \cdot \ldots \cdot a \cdot [b+k(n-j)]\,[b+k(n-j-1)] \cdot \ldots \cdot}{[a+b+k\cdot n] \cdot [a+b+k\cdot(n-1)] \cdot \ldots \cdot [a+b]}$$

So $\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right)$ depends on $X_0, \dots, X_n$ only through the sum $j = \sum_0^n x_i$, that is the single values $x_0 \dots x_n$ are not in the final formula (the formula depends on a, b, k, n too but these are given).
Thus if I make a permutation the sum remains the same and nothing happens.
Hence since the previous probability depends only on the sum j we get

$$\mathbb{P}\left(X_{\pi_0} = x_0, \dots, X_{\pi_n} = x_n\right) = \mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right)$$

$\forall$ permutation $(\pi_0, \dots, \pi_n)$ of $(0, \dots, n)$. Hence the sequence (X_n) is exchangeable.

Being exchangeable, by de Finetti's thm we can write the probability of interest as product of iid events, and taking the expected value for all the possible value of the unknown parameter:

$$\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) = \mathbb{E}_\theta\left[\theta^{\sum_0^n x_i} \cdot (1-\theta)^{\sum_0^n (1-x_i)}\right]$$

In the special case of Polya urn we can also find the distribution of θ. Before that we need a results

Proposition 6.3.2 (Convergence of exchangeable sequence mean)**.** *If (X_n) is any exchangeable sequence such that $\mathbb{E}\left[|X_0|\right] < +\infty$ (having the mean), letting*

$$\overline{X}_n = \frac{1}{n}\sum_0^{n-1} X_i$$

then the sample mean converges to a random variable V, $\overline{X} \xrightarrow{a.s.} V$.

Dimostrazione. By de Finetti thm we have that

$$\mathbb{P}\left(\overline{X}_n \text{ converges}\right) = \mathbb{E}_\nu\left(\mathbb{P}_\nu(\overline{X}_n \text{ converges})\right)$$

where $\mathbb{P}_\nu$ denotes the probability distribution of an iid sequence with common distribution ν. So the probability that the sample mean converges is equal to the expectation (with respect to ν) of the probability of convergence, under assumption of iid sequence with common distribution ν.
But in the hypothesis the sequence is iid and the mean exists, the probability of convergence is 1 by strong law of large numbers of Kolmogorov so

$$\mathbb{P}\left(\overline{X}_n \text{ converges}\right) = \mathbb{E}_\nu\left(\mathbb{P}_\nu(\overline{X}_n \text{ converges})\right) = E_\nu(1) = 1$$

This proves that if the sequence (X_n) is exchangeable and the mean exists ($\mathbb{E}[|X_0|] < +\infty$) then the sample mean $\overline{X} \xrightarrow{a.s.} V$. □

Se la successione fosse iid con media finita allora per la legge forte di kolmogorov avremmo convergenza almost sure e la probabilità di convergenza sarebbe 1.
The notation given in the definetti thm

$$\mathbb{P}(\text{something}) = \mathbb{E}_\nu\left\{\mathbb{P}_\nu(\text{something})\right\}$$

means that the probability of some events, if the sequence is exchangeable, is binded to the probability $\mathbb{P}_\nu(\text{something})$ (which is the probability that something under the assumption that (X_n) is iid with distribution ν) via expectation on all possible probability function ν.
Quindi iid non l'abbiamo ma grazie al teorema possiamo calcolare la probabilità sotto exchangeable (ipotizzando la sequenza fosse iid).
Nel nostro esempio something è l'evento che la media campionaria converga

Remark 95. So as a consequence of de Finetti theorem we've proved strong law of large number holds for exchangeable sequences as well (not only IID as proved by Kolmogorov:

- if the sequence is IID the sample mean converges to the common mean by Kolmogorov

- in the more general exchangeable case, the sample mean converges to V which is a generic non degenerate random variable (we don't know what is it, we know it exists).

Example 6.3.4 (Polya urn continued)**.** Let us go back to Polya sequences/urn; by what already proved $\overline{X} \xrightarrow{a.s.} V$. Moreover, given an integer $m \geq 1$, convergence is also in L_m (converge in L_m perché la successione è limitata, le medie campionarie sono tutte ≤ 1, e posso maggiorare di una costante il valore assoluto; questo basta per dire che la convergenza a.s. implica la convergenza in LM (cosa che normalmente non avviene)..

Hence:

$$\mathbb{E}\left[V^m\right] = \lim_n \mathbb{E}\left[\overline{X}_n^m\right] = \lim_n \mathbb{E}\left[\left(\frac{\sum_{i=0}^{n-1} X_i}{n}\right)^m\right] = \lim_n \frac{1}{n^m}\mathbb{E}\left[\left(\sum_0^{n-1} X_i\right)^m\right]$$

$$= \lim_n \frac{1}{n^m} \sum_{j_1=0}^{n-1}\sum_{j_2=0}^{n-1}\dots\sum_{j_m=0}^{n-1} \mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right]$$

$$\overset{(1)}{=} \lim_n \frac{1}{n^m}\left[\underbrace{\sum_{j_1=0}^{n-1}\sum_{j_2=0}^{n-1}\dots\sum_{j_m=0}^{n-1}}_{j_1,\dots,j_m \text{ distinct}} \mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right] + \underbrace{\sum_{j_1=0}^{n-1}\sum_{j_2=0}^{n-1}\dots\sum_{j_m=0}^{n-1}}_{j_1,\dots,j_m \text{ not distinct}} \mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right]\right]$$

$$\overset{(2)}{=} \lim_n \frac{1}{n^m}\left[n(n-1)\dots(n-m+1)\,\mathbb{E}\left[X_0X_1,\dots X_{n-1}\right] + \underbrace{\sum_{j_1=0}^{n-1}\sum_{j_2=0}^{n-1}\dots\sum_{j_m=0}^{n-1}}_{j_1,\dots,j_m \text{ not distinct}} \mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right]\right]$$

where in:

- (1) we split the sum in two parts where in the first all the indexes are all distinct, while in the second there are repetitions
- in (2) where $j_1,\dots,j_m$ are all distinct under exchangeability the distribution is invariant under permutation so

  $$\mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right] = \mathbb{E}\left[X_0\cdot\ldots\cdot X_{n-1}\right]$$

  and we have $n(n-1)\cdot\ldots\cdot(n-m+1)$ of these expected values with different indexes (n choices as first index, ..., $(n-m+1)$ as the m-th index)

Now we have that the two terms goes like

$$\frac{n(n-1)\cdot\ldots\cdot(n-m+1)}{n^m} \to 1 \quad \text{as } n\to\infty$$

$$\frac{\sum_{j_1}\sum_{j_2}\dots\sum_{j_m}\mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_m}\right]}{n^m} \to 0$$

Regarding the second factor we don't know the value of the expectation $\mathbb{E}\left[X_{j_1}\cdot\ldots\cdot X_{j_n}\right]$ but surely is ≤ 1 (since each X can be only 0 or 1) so we can upper bound it to one and just use the cardinality of the double sum:

$$\frac{\overbrace{\sum_{j_1}\dots\sum_{j_m}}^{\text{not all distinct}} \mathbb{E}\left[X_{j_1}\dots X_{j_m}\right]}{n^m} \overset{(1)}{\leq} \frac{n^m - n(n-1)\dots(n-m+1)}{n^m} = 1 - \frac{n(n-1)\dots(n-m+1)}{n^m} \to 1-1$$

where in (1) being $n(n-1)\cdot\ldots\cdot(n-m+1)$ the number of distict touples and n^m all the touples, $n^m - n(n-1)\cdot\ldots\cdot(n-m+1)$ are the touples with at least

a repeated index.
Hence

$$\begin{aligned}
\mathbb{E}\left[V^m\right] &= \mathbb{E}\left[X_0 X_1 \dots X_{n-1}\right] \\
&\overset{(1)}{=} \mathbb{P}\left(X_0 = X_1 = \ldots = X_{n-1} = 1\right) \\
&\overset{(2)}{=} \frac{[a+k(m-1)][a+k(m-2)] \cdot \ldots \cdot a}{[a+b+k(m-1)] \cdot \ldots \cdot [a+b]} \\
&\overset{(3)}{=} \frac{(\frac{a}{k}+(m-1))(\frac{a}{k}+(m-2)) \cdot \ldots \cdot \frac{a}{k}}{(\frac{a+b}{k}+(m-1))(\frac{a+b}{k}+(m-2)) \cdot \ldots \cdot \frac{a+b}{k}}
\end{aligned}$$

where in:

- (1) since X_i are indicators the product of indicators is still an indicator (all $=1$), and the expected value of indicator is the probability of its event;
- (2) we apply the formula derived before;
- (3) we rewrited by dividing by k.

The last equation is the moment of order m of a beta random variable with parameters $\frac{a}{k}$ and $\frac{b}{k}$.
Now, in general $\mathbb{E}\left[X^m\right] = \mathbb{E}\left[Y^m\right], \forall m \not\Longrightarrow X \sim Y$. However if furthermore X (or Y) has finite moment generating function the implication holds.
Here since $0 \leq V \leq 1$, then V has finite moment generating function: if $a \leq V \leq b$ we have that

$$e^{at} \leq \mathbb{E}\left[e^{tV}\right] \leq e^{bt}$$

and the moment generating function $\mathbb{E}\left[e^{tV}\right]$ is finite.
Therefore having V the same moments of a beta and finite moment generating function, then V is beta distributed $V \sim \text{Beta}\left(\frac{a}{k}, \frac{b}{k}\right)$.
Hence we can say that $\theta = V$ is beta distributed and finally

$$\begin{aligned}
\mathbb{P}\left(X_0 = x_0, \dots, X_n = x_n\right) &= \mathbb{E}_\theta\left[\theta^{\sum_0^n x_i}(1-\theta)^{\sum_0^n (1-x_i)}\right] \\
&= \int_0^1 \theta^{\sum_0^n x_i}(1-\theta)^{\sum_0^n (1-x_i)} \cdot f(\theta)\, \mathrm{d}\theta
\end{aligned}$$

where f is a density of a beta distribution with parameter $V \sim \text{Beta}\left(\frac{a}{k}, \frac{b}{k}\right)$.

6.4 Renewal processes (processi di rinnovo)

Remark 96. It's a generalization of Poisson process.

Definition 6.4.1 (Renewal process)**.** Let $Z_1, Z_2, \dots$ be a sequence of random variables such that

- (Z_n) is independent
- $Z_2 \sim Z_n \forall n \geq 2$ (the sequence starting from the second variable on is iid)
- $Z_n > 0$ almost surely $\forall n \geq 1$

Now let $\tau_0 = 0, \tau_n = \sum_{i=1}^{n} Z_i$. Since the random variable are positive we have $0 = \tau_0 < \tau_1 < \tau_2 < \ldots$. which can be represented on the time line. Interpretation is the same of the Poisson process: τ_1 is time of first arrival/renewal, τ_2 of the second and son on.
Now let's define

$$N_t = \max\{n \geq 0 : \tau_n \leq t\}, \forall t \in [0, +\infty)$$

While (Z_n) is a discrete sequence N_t is a process in continuous time, and can be regarded as the number of arrivals (or renewals, in this process language) in the interval $[0, t]$.
So we have the time line with the ticks $0, \tau_1, \tau_2, \ldots$; regarding the value of the process

- if t is in the interval $[0, \tau_1)$ the value of $N_t = 0$;
- if $t \in [\tau_1, \tau_2)$ then $N_t = 1$;
- ...and so on, it follows easily that $N_t = n \iff t \in [\tau_n, \tau_{n+1})$ or equivalently $t \in [\tau_{N_t}, \tau_{N_t+1})$.

The process N_t is said to be a renewal process.

Important remark 36 (The first two distribution). Regarding the first two random variables Z_1, Z_2, both the following possibilities are allowed by the general definition above:

$$Z_1 \sim Z_2$$
$$Z_1 \nsim Z_2$$

Why Z_2 is not forced to have the same distribution of Z_1 in the general framework? Z_1 is the meaning time of first arrival, while Z_2 is the time between the first and the second: reason is that in real processes/applications time of first arrival doesn't have to be the same distribution of the time between the first and second arrival (eg first and second girl/boyfriend).
The Poisson process is a special case that is obtained assuming further that $Z_1 \sim Z_2 \sim \mathrm{Exp}(\lambda)$ (while in the general case above $Z_1 \sim Z_2$ is not required).

Theorem 6.4.1. *If $\mathbb{E}[Z_2] < +\infty$, then*

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{\mathbb{E}[Z_2]}, \quad \text{for } t \to \infty$$

Dimostrazione. In fact by definition if $N_t = n$ with $t \in [\tau_{N_t}, \tau_{N_t+1})$ that is $\tau_{N_t} \leq t < \tau_{N_t+1}$. Hence

$$\frac{N_t}{\tau_{N_t+1}} < \frac{N_t}{t} \leq \frac{N_t}{\tau_{N_t}}$$

Now looking separately at the two margin for $t \to \infty$:

- we have that

$$\frac{N_t}{\tau_{N_t}} = \frac{1}{\frac{\tau_{N_t}}{N_t}} \overset{(1)}{=} \frac{1}{\frac{\sum_{i=1}^{N_t} Z_i}{N_t}} \overset{(2)}{=} \frac{1}{\mathbb{E}[Z_2]}$$

in

- (1) by definition of τs
- (2) rv starting from the second are iid so we use kolmogorov and the denominator goes to the common mean which is the expectation of Z_2 as $t \to +\infty$ (index of the sum $N_t \xrightarrow{a.s.} +\infty$ as $t \to +\infty$)

- similarly

$$\frac{N_t}{\tau_{N_t+1}} \overset{(1)}{=} \frac{N_t}{N_t+1} \cdot \frac{1}{\frac{\tau_{N_t+1}}{N_t+1}} \overset{(2)}{\to} 1 \cdot \frac{1}{\mathbb{E}\left[Z_2\right]}$$

  where in

  - (1) we merely multiplied and divided by $N_t + 1$ at denominator
  - (2) since $N_t \to +\infty$ as $t \to +\infty$, then $\frac{N_t}{N_t+1} \xrightarrow{a.s.} 1$ as $t \to \infty$; for the other term the same as before

Thus being $\frac{N_t}{t}$ bounded by terms which goes to $1/\mathbb{E}\left[Z_2\right]$ as $t \to +\infty$, it behaves the same way. □

Remark 97. This theorem applies to any renewal process. In the following the application to the Poisson process.

Example 6.4.1. Applying this thm to the Poisson process one obtains that:

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{\mathbb{E}\left[Z_2\right]} \overset{(1)}{=} \lambda$$

where in (1) we have that if $Z_2 \sim \operatorname{Exp}(\lambda)$, then has mean $1/\lambda$.
So in the special case of the Poisson process N_t/t goes almost surely to λ. This fact helps to understand the meaning of parameter λ of Poisson process; being

$$\frac{N_t}{t} = \frac{\text{number of arrivals in } [0,t]}{t}$$

λ is a sort of asymptotic "intensity" of the process as $t \to +\infty$ (number of arrivals over legth of the interval).