Alma Mater Studiorum - Universitàà di Bologna

Departement of Statistical Science
Curriculum Data Science

Stochastic Processes

Presented by:
Alberto Trashaj

Appello I
Anno Accademico 2022-2023

Abstract

Dear reader, thank you for your interest in my notes on stochastic processes. I have created these notes with care and attention to detail, attending every lecture and spending many hours outside of class to ensure their accuracy and clarity. While I do not wish to sell these notes, I would greatly appreciate any donation you may be willing to offer in exchange for their use. As a full-time student, attending these classes is a significant investment of time and resources for me, and your support would be greatly appreciated. Thank you for your consideration. If you have found these notes helpful, and would like to support my efforts in creating them, you can send a donation via PayPal to **trashajalberto@gmail.com**. Your contribution will go a long way in helping me continue to create and share high-quality educational materials. Thank you again for your support.

Contents

Chapter 1

Introduction

1.1 Definition

The first concept that we need to present in this course is the definition of stochastic process:

Definition 1.1.1 (Stochastic process)**.** A stochastic process is any collection of random variables

So, quite simple, no?
Let's now briefly recall what is a random variables since it will be useful to understanding what we are talking about:

A random variable is a measurable map defined as follows:
$X : \Omega \longrightarrow S$
where we just recall that the triplet $(\Omega, \mathcal{A}, P)$ is the well known probability space and $(S, \mathcal{B})$ is the measurable space.
In particular, S is a set and B is a σ-field of subset of S.
It can be possible that you are not familiar with this concepts since some notion of the probability theory are needed, so if you have already problems with understanding what are we talking about pause the reading and try to comprehend the concept so far since they will be very useful for the rest of the course.

```
Disclaimer: in the repository
https://github.com/alberello00/stochastic_processes
you will find some notes of the course "Advanced probability"
which can be useful
to understanding some concepts.
```

So far we just briefly recalled the two main ingredients for defining a random variable: another ingredient that has to be mentioned when defining what is a stochastic process is the set T.
The set T can be any set, but the most natural and intuitive set that we can think of is the $Time$ set. But, later on we will define better what T actually is.
To give a formal definition we can repeat that a stochastic process X is any collection of random variables namely:

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

where $X_t : \Omega \longrightarrow S$ is a random variable $\forall t \in T$

Another parentheses should be opened here: in fact, if one remember what it actually is a random variable can point out that a random variable has to be a measurable map, i.e. the random variables that are in our stochastic processes has to be measurable too:

$$X_t^{-1}(B) \in \mathcal{A} \; \forall B \in \mathcal{B}$$

where

$$X_t^{-1}(B) = \{\omega \in \Omega : X_t(\omega) \in B\}$$

1.1.1 Remarks on S and T

- S is said to be the **state space** of the process: of course, the most important case is when the state space is equal to the real line, namely

$$S = R \text{ and } B \text{ is the Borel } \sigma\text{-field on } R$$

- T is an arbitrary set, which is usually called the **indexing set** of the **parameter space** of the process: a natural interpretation of this set can be thought as the *time* set.
- If T is a finite or countable set, X is said to be a discrete time process: for example, a sequence of random variables.
- If T is an interval of the real line $\mathcal{R}$, X is said to be a **continuous time process**

1.2 A process is a bivariate function

An important way to think of what is a stochastic process is that it is nothing else than a function of two variables:

$$X : \Omega \times T \longrightarrow S$$

So, we can write

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

So, since a stochastic process can be thought as a function of two variables let's see what happens if we fix one of the two variables and let one be freely tho change.

- Let fix $t \in T$:
  Then, X_t is still a random variable, namely a measurable function on Ω.
  Hence, we have

$$X_t(\omega) \; \forall \omega \in \Omega$$

- Let fix $\omega \in \Omega$:
  Then, we obtain a function of t, namely:

$$t \longrightarrow X_t(\omega)$$

  This function is usually called the **path** or **trajectory** of the process.
  For instance, if we choose the state process as $S = R$ and $T = [0, \infty)$: the path is a function that maps from $[0, \infty)$ into $\mathcal{R}$

1.2.1 Another way to see what is a stochastic process

A process can be always regarded as a random function: it is sufficient to think of X as the map $\omega \longrightarrow$ to the path associated to ω, namely:

- $\omega \longrightarrow X(\omega, *)$

1.3 Equality of the processes

One can wonder when two processes can be regarded as "equal", "similar" or, in a more general sense when two processes can be compared and see if they are in some sense indistinguishable.
In general, we can say that there are three level of "equality" of two processes:

1. $X \sim Y \Rightarrow (X_{t_1}, X_{t_2}, ..., X_{t_n}) \sim (Y_{t_1}, Y_{t_2}, ..., Y_{t_n})$
2. X is equal to $Y \longleftrightarrow P(X_t \neq Y_t) = 0 \ \forall \ t \in T$
3. X is indistinguishable from $Y \longleftrightarrow X_t(\omega) = Y_t(\omega) \ \forall \ t \in T$ and $\forall \ \omega \in A$ where $A \in \mathcal{A}$ is such that $P(A) = 1$

The natural question that can pop out in our mind is if this three type of equality are in someway related:

We can state the following implications:
$$3 \Rightarrow 2 \Rightarrow 1$$
$$1 \nRightarrow 2 \nRightarrow 3$$

1.3.1 Equivalent doesn't imply indistinguishable

Usually to say that A doesn't imply B, it is sufficient to prove that A is true while B is false.
So, let's find an example where we have two processes that satisfies the second type of equality but not the third one.

Example 1.3.1. Let V be any random variable such that $V \geq 0$ (positive-definite) and $P(V = v) = 0 \ \forall \ v \geq 0$. For instance, we can take the absolute value of a standard Gaussian, so $V = |Z|$ where $Z \sim N(0, 1)$. Now, let's define two random process:

- $X(t, \omega) = 0 \ \forall \ t \in [0, +\infty) \ , \ \forall \ \omega \in \Omega$
- $$Y(t, \omega) = \begin{cases} 1, & \text{if } t = V(\omega). \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

Then, it's clear that X and Y are not indistinguishable since for $t = V(\omega)$ we get

$$Y(\omega, t) = 1 \text{ and } X(\omega, t) = 0, \text{ so } X(\omega, t) \neq Y(\omega, t)$$

However, with a close analysis, we can say that

$P(X_t \neq Y_t) = P(Y_t \neq 0)$ since $X_t = 0$ as assumption, and also $Y_t \neq 0$ when $t = V(\omega)$, therefore the $P(Y_t \neq 0) = P(V(\omega) = t)$ that we know it's equal to 0 since the probability of a point is 0 by the assumption that $P(V = v) = 0 \ \forall \ v \geq 0$

So, in conclusion we showed that X_t and Y_t are equivalent processes but they are not indistinguishable

1.4 Stopping times

Let take a set $T = 0, 1, 2, ...$ and the usual sample space $(\Omega, \mathcal{A}, \mathcal{P})$; then we define a $filtration$ as an increasing sequence of sub-σ fields of $\mathcal{A}$, namely

$$\mathcal{F}_, \subset \mathcal{F}_\infty \subset \mathcal{F}_\in \subset ... \subset \mathcal{A}.$$

Given this ingredients we can give a formal definition of the stopping time

Definition 1.4.1. A stopping time is a map

$$T : \Omega \Rightarrow \{+\infty, 0, 1, 2, ...\}$$

such that $\{T = n\} \in \mathcal{F}_\backslash \ \forall n \geq 0$.

In general, a σ-field $\mathcal{G} \subset \mathcal{A}$ may be used to describe the so called state of information. By now, it's not so clear what are we referring to when we talk about the "state of information" of $\mathcal{G}$, so let's make some examples:

Example 1.4.1.

- $\mathcal{G} = \{\varnothing, \Omega\}$ contains null information since, by definition, we always know that the $\varnothing$ is always $False$ and the Ω is always $True$.
- $\mathcal{G} = \{\varnothing, \Omega \, A, A^c\}$ in this set the only information we have is whether or not A is $True$.

Now, I think that another example should clarify better what we mean by "state of the information":

Example 1.4.2. Let's take $\Omega = \{1, 2, 3, 4, 5, 6\}$ and define a σ-algebra $\mathcal{F} = \{\$1,2,34,5\varnothing, \{\Omega\}\}$: now, we can see that, for example, the subset $\{1, 2\}$ is not contained, so our probabilistic model can't make conclusion in this last set, so the "information" is contained only in the set $\mathcal{F}$. Therefore, the only "probabilistic" conclusion we can make are only in the σ-field we constructed before; there will no exists random variable that maps the subset $\{1, 2\}$ on the interval $[0, 1]$.

T should be regarded as the first time when something we are interested actually happens: so, when we say that $T = n$ we are implicitly saying that at the time n something happened for the first time.
Remember when we defined the set where T takes values? Yes, at first can seems strange to include the $+\infty$ but now it should be more clear that saying $T = +\infty$ means that something does not happen.

Example 1.4.3. (X_n) is a sequence of real random variable and $A \in \mathcal{B}(\mathcal{R})$; $T = inf\{n : X_n \in A\}$, so the first time n such that $X_n \in A$.

Example 1.4.4 (Casinò)**.** Suppose you and your friends are playing at a roulette table in a casino, and you want to come up with a strategy to maximize your winnings. You decide that you will keep playing until a certain point in time, at which you will stop and walk away with your winnings. The catch is that you don't know ahead of time when this stopping time will occur - it could happen after 10 minutes of playing, or it could happen after several hours.
However, you have a secret weapon: a fortune teller who has told you that the stopping time will occur at the 10th round of the game. In other words, you will keep playing for 10 rounds, and then stop.
In this example, the stopping time is the $10th$ round of the game. This is a stopping time because it is determined by information that is available at or before the $10th$ round of the game. Specifically, the information that you have received from the fortune teller is available to you before the $10th$ round, and therefore the event $T = 10$ (where T is the stopping time) belongs to the sigma-algebra $\mathcal{F}_{10}$.

1.5 Finite dimensional distributions

Definition 1.5.1. Let X be a process indexed by T $\forall n \geq 1$ $\forall t_1, ..., t_n \in T$ we have a n-dimensional random variable $(X_{t_1}, X_{t_2}, ..., X_{t_n})$: the distributions of $(X_{t_1}, X_{t_2}, ..., X_{t_n})$ for all $n \geq 1$ and $\forall t_1, ..., t_n \in T$ are called finite dimensional distribution.

Example 1.5.1. Suppose that $T = \{1, 2\}$ and $X = \{X_1, X_2\}$: $X_1 \sim$ Binomial and $X_2 \sim$ Poisson. Then, for obvious reasons, $(X_1, X_2) \sim N$ it's not possible and such process fail to exist.

Usually, in application we choose the finite dimensional distributions and we look for a process having such a finite dimensional distribution: the problem here arise since this process may not exist.
However there are some theorem usually called *consistency* theorems which provides conditions on the finite dimensional distributions under which the processes (with such finite dimensional distribution) exists.

1.6 Conditional expectations

In order to give a proper definition of conditional expectations, three ingredients are necessary:

1. A sample space, $(\Omega, \mathcal{A}, \mathcal{P})$
2. A sub-σ-field of $\mathcal{A}$: $\mathcal{G} \subset \mathcal{A}$
3. A real random variable X such that $E|X| < +\infty$.

Now, let's define a conditional expectation:

Definition 1.6.1. A conditional expectation of X given $\mathcal{G}$ is any real random variable $V : \Omega \Rightarrow \mathcal{R}$ such that:

1. $E(|V|) < +\infty$
2. V is $\mathcal{G}$-measurable
3. $E(\mathbb{1}_{\mathbb{A}} \cdot X) = E(\mathbb{1}_{\mathbb{A}} \cdot V)\ \forall A \in \mathcal{G}$

Where, on the third point we have,

$$X \cdot \mathbb{1}_{\mathbb{A}} = \begin{cases} \text{X, on } A \\ 0, \text{on } A^c. \end{cases} \tag{1.2}$$

In particular, since $\Omega \in \mathcal{G}$, we obtain:

$$E(X) = E(X \cdot \mathbb{1}) = E(V \cdot \mathbb{1}) = E(V)$$

We need to make a second remark on the second condition stated in the definition: a random variable V is said to be $\mathcal{G}$-measurable if

$$V^{-1}(B) \in \mathcal{G}\ \forall B \in \mathcal{BR}$$

In other terms, V is a random variable not only in the "big" probability space $(\Omega, \mathcal{G}, \mathcal{P})$ but also in the "small" probability space $(\Omega, \mathcal{G}, \mathcal{P})$.
For instance, take the set $\mathcal{G} = \{\varnothing, \Omega\}$ (null information), then if X is $\mathcal{G}$-measurable, it must be that

$$X^{-1}(B) = \Omega \text{ or } X^{-1}(B) = \varnothing\ \forall B \in \mathcal{B}(\mathcal{R})$$

Hence, with the set taken the only $\mathcal{G}$-measurable random variables are the constants. At the opposite extreme, if $\mathcal{G} = \mathcal{A}$, then every random variable is G-measurable. However, the interpretation of $\mathcal{G}$-measurability is under the information $\mathcal{G}$, so a $\mathcal{G}$-measurable random variable reduces to a constant: suppose we have a random variable X that is $\mathcal{G}$-measurable for some sub-σ-field $\mathcal{G}$. Intuitively, this means that X behaves like a constant, because its value can be determined from the information in $\mathcal{G}$: let's see with an example why this is true.

Example 1.6.1. Fix $A \in \mathcal{A}$ and define

- $\mathcal{G} = \{\varnothing, \Omega, A, A^c\}$
- $X = 3\mathbb{1}_{\mathbb{A}} - 2\mathbb{1}_{\mathbb{A}^c}$

Then, X is $\mathcal{G}$-measurable:

$$X^{-1}(B) = \begin{cases} \Omega, \text{if } -2, 3 \in \mathcal{B} \\ \varnothing, \text{if } -2, 3 \notin \mathcal{B} \\ A, \text{if } 3 \in \mathcal{B} \text{ and } -2 \notin \mathcal{B} \\ A^c, \text{if } 3 \notin \mathcal{B} \text{ and } -2 \in \mathcal{B} \end{cases} \tag{1.3}$$

But, under the information of $\mathcal{G}$, we know whether A is $True$ or $False$, thus X becomes a constant.

Theorem 1.6.1 (Theorem)**.** A conditional expectation V always exists, and it is almost surely unique, namely, if V_1 and V_2 are both conditional expectations, then $P(V_1 \neq V_2) = 0$ or, equivalently, $P(V_1 = V_2) = 1$

From now on, to denote a conditional expectation of X given $\mathcal{G}$, we adopt the notation $E(X|\mathcal{G})$, namely $V = E(X|\mathcal{G})$.
The interpretation is that the expected value is our prediction of X under the information $\mathcal{G}$: in the special case where $V = E(X)$, the natural interpretation is that $E(X)$ is our prediction of Xwithout any specific information.
Now, should be more clear why $E(X|\mathcal{G})$ should be $\mathcal{G}$-measurable: in fact, our prediction of X under the information $\mathcal{G}$ should be something which only depends on the information $\mathcal{G}$.

1.6.1 Best predictor?

In general, to make a prediction, we are not only interested in finding a good predictor, but it would be better to find the **best** predictor: don't worry, we already find it! Yes, $E(X|\mathcal{G})$ is the best prediction we can make in the following sense:

- Suppose $E(X^2) < +\infty$: then,

$$E\{(X - (E(X|\mathcal{G}))^2\} = \min E\{(X - Z)^2\}$$

We can think of Z as a predictor of X under the information $\mathcal{G}$

1.6.2 Properties

There are 7 useful properties of the conditional expectation:

1. Linearity: $E(aX + bY|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G})$
2. Positivity: if $X \geq 0 \rightarrow E(X|\mathcal{G}) \geq 0$
3. Constant: $E(constant|\mathcal{G}) = constant$
4. Measurability: if Z is $\mathcal{G}$-measurable, then $E(X \cdot Z|\mathcal{G}) = Z \cdot E(X|\mathcal{G})$
   It follows that, if X is $\mathcal{G}$-measurable, then

$$E(X|\mathcal{G}) = E(X \cdot 1|\mathcal{G}) = X \cdot E(1|\mathcal{G}) = X \cdot 1 = X$$

5. Independence: X is said to be independent from $\mathcal{G}$, by definition, if

$$P(A, X \in B) = P(A) \cdot P(X \in B) \forall A \in \mathcal{G} \forall B \in \mathcal{B}(\mathcal{R})$$

   If X is independent from $\mathcal{G}$, then $E(X|\mathcal{G}) = E(X)$

6. Chain rule: if $\mathcal{G}_\infty \subset \mathcal{G}_\in \subset \mathcal{A}$, then $E(X|\mathcal{G}_\infty) = E(E(X|\mathcal{G}_\in)|\mathcal{G}_\infty)$.
   For instance if we know that $E(X|\mathcal{G}_\in) = 0$, then
   $E(X|\mathcal{G}_\infty) = E(E(X|\mathcal{G}_\in)|\mathcal{G}_\infty) = E(0|\mathcal{G}_\infty) = 0$

7. Conditional probability: if $A \in \mathcal{A}$, we can define

$$P(A|\mathcal{G}) = E(\mathbb{1}_\mathbb{A}|\mathcal{G})$$