Notes

Fabio Brau

February 6, 2023

Contents

Chapter 1

Markov Chains

In the section we will introduce the model of a classical Markov chain with discrete time for which each state belongs to some finite or countable set of possible state. In the next section we extend the definition to states with a continuous states or in general a continuous density function.

1.1 Basic Definitions

1.1.1 Discrete Probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. The class-function $\mathbb{P}$ is a finite measure over the sigma-algebra $\mathcal{F}$ such that $\mathbb{P}(\Omega) = 1$. A discrete random variable is represented by a measurable function $X : \Omega \to S$, where $S = \{s_1, \dots, s_n, \dots\}$ equipped with the discrete measure. The variable is associated to a discrete distribution λ.

Definition 1 (Discrete Distribution)**.** A sequence $\lambda = (\lambda_0, \dots, \lambda_n, \dots)$ is a *discrete distribution* if and only if $\sum_{i \in \mathbb{N}} \lambda_i = 1$.

We will say that a random variable X has distribution λ if and only if for each possible outcome $i \in \mathbb{N}$,

$$\mathbb{P}(X = i) := \mathbb{P}(\{w \in \Omega \,:\, X(w) = i\}) = \lambda_i. \tag{1.1}$$

1.1.2 Continuous Probability

As in the discrete case, let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A continuous d-dimensional random variable is defined as a measurable map $X : \Omega \to \mathbb{R}^d$, where $\mathbb{R}^d$ is equipped with the Lebesgue measure.

Remember that by definition, for each $A \subseteq \mathbb{R}^d$ measurable, we have that the probability that X takes values in the set A corresponds to the probability measure of the set $X^{-1}(A)$. In formulas,

$$\mathbb{P}(X \in A) := \mathbb{P}(\{w \in \Omega \,:\, X(w) \in A\}) \tag{1.2}$$

Definition 2 (Expectation)**.** Let X a random variable. The expectation $\mathbb{E}[X]$, when it exists , is defined by

$$\mathbb{E}\left[X\right] := \int_{\Omega} X(\omega)\, d\mathbb{P}(\omega). \tag{1.3}$$

As in the discrete case, is it possible to define the concept of independence even in the continuous case. In particular we can defined the independence of: Events, σ-algebras and random variables.

Definition 3 (Independence of Events)**.** Let $A, B \in \mathcal{F}$ two events, they are independent if and only if

$$\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(A\right)\mathbb{P}\left(B\right) \tag{1.4}$$

Definition 4 (Independence of σ-Algebras)**.** Two σ-algebras $\mathcal{G}_1, \mathcal{G}_2 \subseteq \mathcal{F}$ are independent if and only if for each two events A, B respectively in $\mathcal{G}_1, \mathcal{G}_2$, they are independent.

Before defining the independence of two random variables, let us remember that a random variable X generate a σ-algebra on Ω.

Remark 1. Let $X \in \mathbb{R}^d$ be a random variable, the class of set defined by

$$\sigma(X) := \left\{X^{-1}(A) \,:\, A \subseteq R^d \text{ measurable}\right\},$$

is a σ-algebra.

Definition 5 (Independence of Random Variables)**.** Two random variables X, Y are independent if and only if the corresponding sigma algebras are independent.

Observe that the last definition is equivalent to say that for each A, B measurable sets, then

$$\mathbb{P}\left((X,Y) \in A \times B\right) = \mathbb{P}\left(X \in A, Y \in B\right) = \mathbb{P}\left(X \in A\right)\mathbb{P}\left(Y \in B\right) \tag{1.5}$$

Similar, but not as easily as the discrete case, we can define the concepts of: Conditional Random Variable (a.k.a *Conditional Expectation*), and *Conditional probability.*

Proposition 1 (Conditional Expectation)**.** *Given a random variable X and given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists a unique random variable Z such that*

1. *Z is measurable in $\mathcal{G}$. In formulas $X \in \mathcal{G}$ or directly $\sigma(Z) \subseteq \mathcal{G}$.*
2. $\forall G \in \mathcal{G}, \quad \int_G X(\omega)\, d\mathbb{P}(\omega) = \int_G Y(\omega)\, d\mathbb{P}(\omega)$.

The random variable Z is named conditional expectation, and it is usually expressed as $\mathbb{E}\left[X \mid \mathcal{G}\right] := Z$.

Proof. The proof is a direct consequence of the Theorem of Radon-Nikodym. See Durrett for further details. □

Definition 6 (Conditional to a Variable)**.** Let X, Y two random variables measurable on $\mathcal{F}$. The *Conditional to a Variable* is defined as follows

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)], \tag{1.6}$$

where remember that $\sigma(Y)$ is the smallest σ-algebra for which Y is measurable.

The following two examples aim at providing an idea of the definition of conditional expectation. The conditional expectation can be thought as the best estimate of X under the hypothesis that certain events happen. We focus on two opposite cases, in the first can be though as the case of *Conditioning by providing all the information* the latter *Conditioning without adding information.*

Remark 2 (No-Information)**.** Let us assume that X is a random variable measurable on $\mathcal{G} \subseteq \mathcal{F}$. This means that finding the probability that X takes certain values in B by knowing that a set of possible events $\mathcal{G}$ happen. If the event $\{X \in B\}$ is in $\mathcal{G}$, than we are not providing a further information. In formulas

$$X \in \mathcal{G} \quad \Rightarrow \quad \mathbb{E}[X \mid \mathcal{G}] = X. \tag{1.7}$$

Remark 3 (All-Information)**.** Let us assume that X is independent of a σ-algebra $\mathcal{G}$. Then the conditional expectation of X on $\mathcal{G}$ is given by

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]. \tag{1.8}$$

Proof. Let $A \in \mathcal{G}$. Since X is independent of $\mathcal{G}$, then, by definition, for each $A \in \mathcal{G}$, X is independent of the characterization function over A. This implies that

$$E[X\, I_A] = E[X]\, E[I_A] \tag{1.9}$$

from which it is easy to deduce the thesis. $\square$

Definition 7 (Density Distribution)**.** A random variable X admits a *density distribution* if there exists a function $f_X : \mathbb{R}^d \to \mathbb{R}_+$ such that, for any open set $A \subseteq \mathbb{R}^d$ the following equality holds

$$\mathbb{P}(X \in A) := \mathbb{P}(\{w \in \Omega \,:\, X(w) \in A\}) = \int_A f_X(x)\, dx \tag{1.10}$$

Observe that not all the random variables admit a density function, a typical examples is given by the random variable $X \equiv a$, that is equal to the constant a. In this case, for each open set $A \subseteq \mathbb{R}$ we have that

$$\mathbb{P}(X \in A) = \begin{cases} 1, & \text{if } a \in A \\ 0, & \text{otherwise} \end{cases} \tag{1.11}$$

Given two random variables $X \in \mathbb{R}^m, Y \in \mathbb{R}^n$ we can defined the joint random variable as follows $(X, Y) \in \mathbb{R}^{m+n}$.

1.2 Countable States Markov Chains

Let $(X_t)_{t\in\mathbb{N}}$ a sequence of random variables. Let us assume that each instant $t \in \mathbb{N}$ the variable X_t takes values in a countable state space S.

Definition 8 (Markov Property)**.** The sequence $(X_t)_{t\in\mathbb{N}}$ satisfies the Markov property if for each time t and for each states $s, s_0, \ldots, s_n \in S$

$$\mathbb{P}(X_{n+1} = s \,|\, X_0 = s_0, \ldots, X_n = s_n) = \mathbb{P}(X_{n+1} = s \,|\, X_n = s_n). \tag{1.12}$$

That is, the state assumed at a certain instant t only depends on the previous state and not on the whole history.

Observe that, since we are assuming that S is finite, then we are assuming that there exists an enumeration $S = \{s_1, \ldots, s_n, \ldots\}$. Hence, for the sake of simplicity, and without loss of generality, we can assume that $S = \mathbb{N}$, from which $X_t \in \mathbb{N}$ for each t.

Based on the latter assumption over S, a *Markov Chain* with discrete time and countable state set is represented by a tuple (λ, P) as stated in the following definition.

Definition 9 (Transaction Matrix)**.** A transaction matrix $P = (p_{ij})$ is matrix with infinite entries such that

$$\forall i \in \mathbb{N}, \quad \sum_{j\in\mathbb{N}} p_{ij} = 1. \tag{1.13}$$

In other words, a matrix P is a transaction matrix if every row $P_{i:}$ is a discrete distribution.

Definition 10 (Markov Chain)**.** Let λ be a discrete distribution, and let P be a transaction matrix. A sequence of random variable $(x_i)_{i\in\mathbb{N}}$ is a Markov chain with initial distribution λ and transaction matrix P

- $\lambda_i = \mathbb{P}(X_0 = i)$;
- $p_{ij} = \mathbb{P}(X_n = j \,|\, X_{n-1} = i)$ for each state n.
- X_{n+1} is independent from $X_0, \ldots, X_{n-1}$.

1.3 Markov Chain with Continuous State Space

Let assume $(X_t)_{t\in\mathbb{N}}$ is a sequence of continuous random variables, and that the initial variable X_0 has a continuous distribution $q(x)$.

1.3.1 Distribution Notation

From know on we will make use of a useful distributional notation. Let X be a continuous random variable from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the continuous state space $(E, \mathcal{E})$, e.g., $E = \mathbb{R}^n$ equipped with the Lebesgue measure and $\mathcal{E}$ the Borel σ-algebra of measurable sets.

With the symbol $p(X)$ is the density function of X, such that $p(X) : E \to \mathbb{R}$ and is such that

$$\mathbb{P}\left(X \in \mathbb{A}\right) = \int_A p(X)\, dX \int_A p(X)(x) \tag{1.14}$$

Remember this means that

$$\forall A \subseteq \mathbb{R}, \quad \mathbb{P}\left(X \in A\right) = \int_A p(x)\, dx \tag{1.15}$$

Definition 11 (Markov Chain)**.** Let p a distribution and let $q(x, y)$ a join distribution. The process (X_t) is a Markov chain with continuous state space if

- $X_0 \sim p$
- $(X_{t+1} \,|\, X_t = x) \sim p(x, \cdot)$
- X_{t+1} is independent from $X_0, \ldots, X_{t-1}$.