Probability

Marco Biroli, Alessandro Pacco

September 7, 2020

Contents

Chapter 1

Probabilistic model

1.1 Definitions

Definition 1.1.1 (Universe)**.** We consider a random experiment, then the set of all possible outcomes of the experiment is denoted by Ω and is called the universe.

Definition 1.1.2 (Event)**.** An event E associated to the experiment is a set of results for which we can compute the probability.

Definition 1.1.3 (Collection)**.** The collection of all events is denoted by $\mathcal{F}$. Hence $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, the collection of all subsets of Ω.

Definition 1.1.4 (Disjoint Events)**.** Two events $A, B \in \mathcal{F}$ are disjoint or incompatible if they cannot occur simultaneously. In other words if $A \cap B = \emptyset$ (the null or impossible event).

Remark. *We require that the collection $\mathcal{F}$ of the events is an algebra of sets.*

Definition 1.1.5 (Algebra of Sets)**.** The collection $\mathcal{F}$ is called an algebra of sets if $\mathcal{F} \neq \emptyset$ (i.e. it is a non-empty collection of sets) and:

1. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (stability under complement)
2. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$ (stability under union)

Remark. *For the scope of this course we further require that $\mathcal{F}$ is stable under countable unions. In other words the second condition (2) above is replaced by (2'):*

$$(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$$

Definition 1.1.6 (σ-algebra)**.** A σ-algebra is an algebra of sets where the second condition is replaced by the stronger condition requiring stability under countable union.

1.2 Probability

Let us consider an event $E \in \mathcal{F}$. The probability $P(E)$ of E is the theoretical value for the proportion of experiments in which E occurs. Thus the probability is a function from $\mathcal{F}$ to $[0, 1]$ such that

1. $P(\Omega) = 1$.
2. $A, B \in \mathcal{F}, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.

In other words, P is an additive set function from $\mathcal{F}$ to $[0, 1]$.

Remark. *This definition however is not very well suited to infinite event sets. Then modern probability theory adds a condition to the above.*

Modern probability is built with the stronger condition (2') instead of (2):

$$(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}} \text{ s.t. } (\forall n, m \in \mathbb{N}, n \neq m \Rightarrow A_n \cap A_m = \emptyset) \Rightarrow P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$$

A priori, (2') does not collide with intuition. Moreover, this condition allows to prove much more interesting limit theorems. Conclusion:

Definition 1.2.1 (Probability Space)**.** A probability space is a triple $(\Omega, \mathcal{F}, P)$. Where:

- Ω is a set, the universe of all possible results
- $\mathcal{F}$ is a σ-algebra on Ω
- P is a probability function $\mathcal{F} \to [0,1]$ satisfying (1) and (2')

Remark. *The mathematical framework which defines probability theory actually comes from another mathematical framework called measure theory. This is why the elements of the σ-field are sometimes called the measurable sets and the probability function is sometimes called a probability measure.*

1.3 Finite Spaces

We consider the case where Ω is a finite set, we write $\Omega = \{x_1, \dots, x_n\}$. The natural σ-algebra on Ω is $\mathcal{P}(\Omega)$. It is the only σ-algebra which contains the singletons. Then let P be a probability on Ω and let us set $\forall i \in [\![1, n]\!], p_i = P(\{x_i\})$. Then the numbers p_i satisfy:

$$(\forall i \in [\![1, n]\!], 0 \leq p_i \leq 1) \wedge \sum_{i=1}^{n} p_i = 1 \tag{1.1}$$

Then for any $A \subset \Omega$ we have by additivity that:

$$P(A) = \sum_{x \in A} P(\{x\}) = \sum_{i : x_i \in A} p_i \tag{1.2}$$

Hence P is completely determined by the numbers p_i.

Remark. *Notice that conversely if we are given the numbers p_i satisfying equation (1.1) we can define a probability P on Ω by stating $P(\{x_i\}) = p_i$ and using equation (1.2). P will indeed be a probability measure.*

1.4 Countable Spaces

We suppose that Ω is countable and we set $\Omega := \{x_n, n \in \mathbb{N}\}$. The natural σ-field on Ω is again the power set of Ω, i.e. $\mathcal{P}(\Omega)$. Let P be a probability on Ω and let us set:

$$\forall n \in \mathbb{N}, p_n = P(\{x_n\}).$$

The sequence (p_n) satisfies:

$$\forall n,\ 0 \leq p_n \leq 1 \wedge \sum_{n \in \mathbb{N}} p_n = 1$$

If $A \in \Omega$, we have again:

$$P(A) = \sum_{x \in A} P(\{x\}) = \sum_{n \in \mathbb{N}, x_n \in A} p_n$$

the solution is similar to the finite case.

1.5 Continuous Spaces

We consider here the more complicated situation where Ω is continuous. If we take the simplest example of $\Omega = \mathbb{R}$ then the intuitive σ-field being the power set turns out to be too complicated to be useful. We will instead consider a simpler σ-field on $\mathbb{R}$. We will start with closed intervals $[a, b]$, $a < b \in \mathbb{R}$. We consider the smallest σ-field on $\mathbb{R}$ which contains these intervals. It can be proved that it exists and is well-defined. This σ-field is called the Borel field and denoted by $\mathcal{B}(\mathbb{R})$. What does it contain: closed intervals, open intervals, "semi-open" intervals, all possible countable unions of closed intervals ecc.. It can be proved that $\mathcal{P}(\mathbb{R}) \neq \mathcal{B}(\mathbb{R})$.

1.6 Random Variables

A random variable is known only after the realization of the experiment, hence it is random. So a random variable X is a map from Ω to $\mathbb{R}$. We want to compute the probability that X belongs to some interval of $\mathbb{R}$, that is why we ask it to be "measurable" (to be defined below).

Definition 1.6.1 (Random Variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable X on $(\Omega, \mathcal{F}, P)$ is map from Ω to $\mathcal{F}$ which is measurable:

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

Remark. *This definition is equivalent to: for all interval I of $\mathbb{R}$ we have that $X^{-1}(I) \in \mathcal{F}$. The point is that, for any interval, $X^{-1}(I)$ is an event, and its probability is well defined.*

Notation. *The event $X^{-1}(I)$ is denoted by $\{X \in I\}$ or even simply $X \in I$. Thus $P(X^{-1}(I)) = P(X \in I)$. Secondly, random variables are denoted by capital letters typically X, Y, U, V and their possible values are denoted by the corresponding lowercase letters x, y, u, v.*

1.7 The law of a random variable

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be a random variable defined on Ω.

Definition 1.7.1 (Law of a random variable)**.** The law of X is the probability measure P_X on $\mathbb{R}$ defined by:

$$\forall B \in \mathcal{B}(\mathbb{R}), P_X(B) = P(X \in B) = P(X^{-1}(B))$$

Proof. Let us check that P_X is indeed a probability measure. We have that:

$$P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1.$$

Furthermore let $(B_n)_{n\in\mathbb{N}} \in \mathcal{B}(\mathbb{R})^{\mathbb{N}}$ be a disjoint sequence of Borel sets. Then:

$$P_X\left(\bigcup_{n\in\mathbb{N}} B_n\right) = P\left(X \in \bigcup_{n\in\mathbb{N}} B_n\right) = P\left(\bigcup_{n\in\mathbb{N}} \{X \in B_n\}\right) = \sum_{n\in\mathbb{N}} P(X \in B_n) = \sum_{n\in\mathbb{N}} P_X(B_n)$$

$\square$

Let us insist on the fact that the law P_X of X is a probability measure on $\mathbb{R}$, and this whatever the set Ω is.

Notation. *The law P_X of X is sometimes called the distribution of X. We furthermore say that two variables X, Y have the same law if $P_X = P_Y$. The object of primary interest for a random variable is its law. Indeed, we want to compute the probabilities of events associated to X, and this is done with the help of its law.*

"The law is fundamental"

1.8 Probability measures on $\mathbb{R}$

We start with $\mathbb{R}$ and the Borel σ-field $\mathcal{B}(\mathbb{R})$. Let f be a non-negative function $\mathbb{R} \to \mathbb{R}$, which is integrable and such that $\int_{\mathbb{R}} f(x)dx = 1$. We define next $\forall A \in \mathcal{B}(\mathbb{R})$,

$$P(A) = \int_A f(x)dx.$$

This formula defines a probability measure on $\mathbb{R}$, called the probability measure with density function f. The good definition of integral is the Lebesgue integral, which we will use all along the course.

Definition 1.8.1 (Law)**.** Let f be a non-negative function $\mathbb{R} \to \mathbb{R}^+$ which is integrable and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$. We define next:

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad P(A) = \int_A f(x)\mathrm{d}x$$

This formula defines a probability measure on $\mathbb{R}$, called the probability measure with density function f.

Other examples: the dirac mass. Any convex combination ($\frac{1}{2}(P_1 + P_2)(A) = \frac{1}{2}(P_1(A) + P_2(A))$) is still a probability measure. There exist probability measure that do not belong to this catalog, but this is another story.

1.9 Expectation

We say that the random variable X has an expectation or that it is integrable if:

$$\int_{\mathbb{R}} |x| \mathrm{d}P_X(x) < +\infty$$

Then the expectation is defined as:

$$E(X) = \int_{\mathbb{R}} x \mathrm{d}P_X(x) = \int_{\Omega} X \mathrm{d}P = \int_{\omega \in \Omega} X(\omega) \mathrm{d}P(\omega) = \int_{\mathbb{R}} \mathrm{Id}_{\mathbb{R}} \mathrm{d}P_X$$

From this formula we see that the expectation is completely dependent on the law of the random variable. "Expectations depend on the laws".

Chapter 2

Coin tossing games

2.1 Model

We take a fair coin and we consider the experiment which consists in throwing n times the coin. If -1 denotes tail and +1 denotes head, then the result of the experiment is a sequence of length n of signs -1,+1, that is an element of

$$\Omega = \{-1, +1\}^n.$$

The σ-algebra is already defined. By symmetry, all the sequences of signs have the same probability, thus we set:

$$\forall w = (w_1, \ldots, w_n) \in \Omega,\ P(w) = \frac{1}{|\Omega|} = \frac{1}{2^n}.$$

Let X_k be the result of the k-th throw. Then X_k is a random variable, defined by:

$$X_k : w = (w_1, \ldots, w_n) \in \Omega \to X_k(w) = w_k$$

2.2 Graphical representation

To the sequence $X_1, \ldots, X_n$, we associate the partial sums: $S_0 = 0, S_1 = X_1, \ldots, S_n = X_1 + \ldots X_n$. THe sequence $S_0, \ldots, S_n$ contains exactly the same information as the initial sequence $X_1, \ldots X_n$. We can represent the result of the experiment by a poligonal line, the line which joins successively the points

$$(0, S_0), \ldots, (n, S_n)$$

disegno, with slope $= \pm 1$. Such as polygonal line, associated to a sequence of signs will be called a path.

2.3 Interpretation of this model

There are more interpretations:

1. Coin tossing game. P and V play the following game: P throws a coin and V tries to guess the result. If V guesses correctly, then P gives 1 euro to V, otherwise V gives 1 euro to P. Here S_n represents the algebraic gain of P after n turns.

2. Random walk. A drunkard perfomrs a random walk on $\mathbb{Z}$ with the following mechanism:

   - at time 0 he starts at 0
   - at time 1, he tosses a coin. If the result is heads he goes to the right, if it is tail, he goes to the left. Picture
   - he reiterates this procedure from this new position.

   with this interpretatoin, S_n represents the position of the drunkard after n steps.

2.4 Distribution or law of S_n

Proposition 2.4.1. *The law of S_n is the probability distribution on $\{-n, \ldots, n\}$ given by*

$$\forall k \in \{-n, \ldots, n\} \;\; P(S_n = k) = \frac{1}{2^n} C_n^{\frac{n+k}{2}}$$

Proof. By graphical representation,

$$P(S_n = k) = \frac{1}{2^n} \cdot |\{\text{Paths from (0,0) to (n,k)}\}|$$

let us consider a path and let us denote by α the number of ascending steps (+1) in the path, and by β the number of descending steps. We must have

$$\begin{cases} \alpha + \beta = n \\ \alpha - \beta = k \end{cases}$$

then $\alpha = \frac{n+k}{2}$. To count the number of paths, I count the number of possible choices for the ascending steps. There are $C_n^\alpha = (n\alpha)$. Convention: $C_n^x = 0$ if $x \notin \mathbb{Z}$ $x < 0, x > n$.

2.5 Equalization or return to 0

We say that there is an equalization or return to 0 at the time n if $S_n = 0$. Since n and S_n have the same parity, this occurs only at even times, and

$$P(S_{2n} = 0) = \frac{1}{2^{2n}} C_{2n}^n = \frac{1}{2^{2n}} \frac{(2n)!}{(n!)^2}.$$

We use Stirling formula:

$$n! = (\frac{n}{2}\Big)^n \sqrt{2\pi n}\Big(1 + \frac{1}{12n} + \frac{1}{288n^2} + o(1/n^2)$$

We get

$$P(S_{2n} = 0) \sim_{n \to +\infty} \frac{1}{\sqrt{\pi n}}$$

This gives an excellent approximation, even for small values of n. For example:

$$P(S_{10} = 0)\text{true: 0,2461, approx: 0,2523}$$

2.6 The lamplighter walk

We consider an infinite street, equipped with laterns, one every meter, and a lamplighter, which lights the lanterns. He starts at 0, he lights the lantern, then he throws a coin to decide whether he goes left or right, and he goes on this way. The positioin at time n of the lamplighter is S_n. Picture. The process $(S_n)_{n \in \mathbb{N}}$ is the symmetric random walk on $\mathbb{Z}$. What is the probability that the lamplighter comes back to the initial lantern? Notice that we prefer unions rather than existance symbols:

$$P(\exists n \in \mathbb{N}^* |\; S_n = 0) = P(\cup_{n \geq 1} \{S_{2n} = 0\})$$

here we are a bit stuck cause we cannot use the σ-additivity, since the events are not disjoint. Hence we "disjoint the union", in a standard procedure:

$$\begin{aligned} &= P(\{S_2 = 0\} \cup (\{S_4 = 0\} \backslash \{S_2 = 0\}) \cup \ldots \cup (\{S_{2n} = 0\} \backslash (\{S_2 = 0\} \cup \ldots \cup \{S_{2n-2} = 0\}))....) \\ &= P(\cup_{n \geq 1}(\{S_{2n} = 0\} \backslash (\{S_2 = 0\} \cup \ldots \cup \{S_{2n-2} = 0\}))) \\ &= P(\cup_{n \geq 1}\{S_2 \neq 0, \ldots, S_{2n-2} \neq 0, S_{2n} = 0\}) \end{aligned}$$

These events are disjoint, so, by σ- additivity,

$$= \sum_{n \geq 1} P(S_2 \neq 0, \ldots S_{2n-2} \neq 0, S_{2n} = 0)$$

but

$$P(S_2 \neq 0, \ldots, S_{2n-2} \neq 0, S_{2n} = 0) = P(S_1 > 0, \ldots S_{2n-1} > 0, S_{2n} = 0) + P(S_1 < 0, \ldots S_{2n-1} < 0, S_{2n} = 0) = 2P(S_1 > 0, \ldots S$$

Picture. We then compute

$$= \frac{2}{2^{2n}}(|\text{paths from (1,1) to (2n-1,1)}| - |\text{paths from (1,1) to (2n-1,1) which touch the x-axis}|)$$

2.7 The reflection principle

(William Feller Intro to proba theory and applications) Let $A = (a, \alpha)$ and $B = (b, \beta)$ be two points with $0 < a < b$, $\alpha, \beta > 0$. Picture. To find the number of paths from A to B which touch the axis, we make a reflection of A with respcet to the x-axis, call it $A' = (a, -\alpha)$. Hence the number of paths from A to B which touch the axis is equal to the paths from $A' = (a, -\alpha)$ to B(without constraint).

Proof. Let $s = (s_a = \alpha, s_{a+1}, \ldots, s_b)$ be a path from A to B which touches the axis. Let t be the first time it touches,

$$t = \min\{i \geq a : s_i = 0\}$$

Let $T = (t, 0)$. To the path s, we associate the path $\phi(s)$ obtained from s by taking the reflection of the position A with respect to the axis. I claim that ϕ is a one to one map from the set of paths from A to B which touch the axis onto the set of paths from A' to B. In fact, $\phi^2 = Id$, whih we can infer by making again the symmetry.

2.8 The ballot theorem

Let $x, n > 0$. THe number of paths from (0,0) to (n, x), picture, which don't touch 0 after time 0 is equal to

$$\frac{x}{n} C_n^{\frac{n+x}{2}}$$

.

Proof.

|paths from (0,0) to (n,x), no touch| = |paths from (1,1) to (n,x), no touch| = |(1,1) to (n,x)|−|(1,1) to (n,x) touch| $=_{reflect}$

for the first term we have

$$\begin{cases} \alpha + \beta = n - 1 \\ \alpha - \beta = x - 1 \end{cases} \Rightarrow \alpha = \frac{n+x}{2} - 1$$

and for the second term we have

$$\begin{cases} \alpha - \beta = n - 1 \\ \alpha - \beta = x + 1 \end{cases} \Rightarrow \alpha = \frac{n+x}{2}$$

Finally we get

$$= C_{n-1}^{\frac{n+x}{2}-1} - C_{n-1}^{\frac{n+x}{2}} = C_{n-1}^{\frac{n+x}{2}-1} - C_{n-1}^{\frac{n-x}{2}-1} = \frac{n+x}{2n} C_n^{\frac{n+x}{2}} - \frac{n-x}{2n} C_n^{\frac{n+x}{2}} = \frac{x}{n} C_n^{\frac{n+x}{2}}$$

where we applied the two formulas $C_n^k = C_n^{n-k}$ and $\frac{k}{n} C_n^k = C_{n-1}^{k-1}$. Application: In an election, candidate P scores p votes, reps Q scores q votes, $p > q$. The probability that the winning candidate is always ahead during the reading of the votes is

$$\frac{p-q}{p+q}$$

2.9 End of the computation

THanks to the Ballot thereom we then get:

$$P(S_2 \neq 0, \ldots, S_{2n-2} \neq 0, S_{2n} = 0) = \frac{2}{2^{2n}} |\text{paths from (0,0) to (2n-1,1) which stay positive}| = \frac{2}{2^{2n}} \frac{1}{2n-1} C_{2n-1}^n = \frac{2}{2^{2n}} \frac{1}{2n-1}$$

. where at the end we applyed the two notorious formulae stated above. Thus

$$P(S_2 \neq 0, \ldots, S_{2n-2} \neq 0, S_{2n} = 0) = \frac{1}{2n-1} P(S_{2n} = 0)$$

Hence

$$P(\text{lamplighter returns to 0}) = \sum_{n \geq 1} \frac{1}{2n-1} P(S_{2n} = 0) = \sum_{n \geq 1} \frac{1}{2n-1} \frac{1}{2^{2n}} \frac{(2n)!}{(n!)^2}$$

Retry:

$$\frac{1}{2n-1} P(S_{2n} = 0) = \frac{1 - 2n + 2n}{2n-1} P(S_{2n} = 0) = \frac{2n}{2n-1} P(S_{2n} = 0) - P(S_{2n} = 0)$$

but

$$\frac{2n}{2n-1} P(S_{2n} = 0) = \frac{2n}{2n-1} \frac{1}{2^{2n}} C_{2n}^n = \frac{2n}{2n-1} \ldots. = \frac{2n}{2n-1} \frac{1}{2^{2n}} 2 \frac{2n-1}{n} C_{2n-2}^{n-1} = P(S_{2n-2} = 0) = \sum_{n \geq 1} \frac{1}{2n-1} P(S_{2n} = 0) = \sum_{n \geq}$$

2.10 Fundamental lemma

we have obtained

$$P(S_2 \neq 0, \dots S_{2n-2} \neq 0, S_{2n} = 0) = P(S_{2n-2} = 0) - P(S_{2n} = 0)$$

yet

$$P(S_2 \neq 0, \dots, S_{2n-2} \neq 0, S_{2n} = 0) = P(S_2 \neq 0, \dots, S_{2n-2} \neq 0) - P(S_2 \neq 0, \dots S_{2n} \neq 0)$$

so for $n \geq 1$

$$P(S_2 \neq 0, \dots, S_{2n-2} \neq 0) - P(S_2 \neq 0, \dots, S_{2n} \neq 0) = P(S_{2n-2} = 0) - P(S_{2n} = 0)$$

moreover $P(S_2 \neq 0) = 1/2 = P(S_2 = 0)$ Fundamental Lemma:

$$P(S_2 \neq 0, \dots, S_{2n} \neq 0) = P(S_{2n} = 0)$$

2.11 Last tie

Consider a coin tossing game of length $2n$ and the time of the last tie before $2n$.

$$T = \max\{k \leq 2n : S_k = 0\}$$

Intuition tells us that the winning player should change frequently during the game. What is the distribution of T? So we image that T should be close to $2n$. But this is completely false. In fact, the law of T is symmetric with respect to n:

$$P(T < n) = P(T > n)$$

hence,

$$P(T \leq n) > \frac{1}{2}$$

Proposition 2.11.1. *(Arcsinuns law for T)*

$$\forall k \in \{0, \dots, n\} : P(T = k) = P(S_{2k} = 0)P(S_{2n-2k} = 0)$$

Example: coin tossing 1 toss each second for 1 year day and night. Proba =1/10 T occurs in 2 first days. 1/20 2,25 day. 1/100 2h 15 min.