Basic Probability Theory Concepts

Alex Isakson

March 2023

We start with some lose definitions of events.

Definition 0.1. A set Ω is called event space. A subset $A \subset \Omega$ is called event. The set of all subsets $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ is called power set of Ω.

We do not care what Ω is. Important is: It is a set and we are interested in its subsets. The power set is crazy and contains a lot of weird subsets that are very non-intiutive. So, we want to look at a smaller subset of the power set that contains 'normal' or intuitive subsets of Ω. We call this subset a σ-Algebra:

Definition 0.2. Let $\Omega \neq \emptyset$ (empty set). $\mathcal{A} \subset \mathcal{P}(\Omega)$ is called σ-Algebra over Ω if

(i) $\Omega \in \mathcal{A}$

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(iii) $A_i \in \mathcal{A}(i \in \mathbb{N}) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

So, a σ-Algebra is just a set of sets of Ω that follow the above system. In $\mathbb{R}$ the smallest σ-Algebra that contains all open sets is special. We call this set a borel σ-Algebra and denote it with $\mathcal{B}_{\mathbb{R}}$. This σ-Algebra is very important but it again has a similar idea: Take all sets in $\mathbb{R}$ that are not too crazy. Indeed nearly all sets of the real numbers you can think of are in $\mathcal{B}_{\mathbb{R}}$. It is not that easy to construct sets that are not part of it.
Why do we need σ-Algebras? The reason is because of measures. We want to define probability through measure theory and they do not work on crazy subsets as the power set. The definition of a measure makes it more clear:

Definition 0.3. Let $\mathcal{A}$ be a σ-Algebra over Ω. A measure is a function $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$ with

(i) $\mu(A) \geq 0$ for all $A \in \mathcal{A}$

(ii) $\mu(\emptyset) = 0$

(iii) For all countable sets $\{A_k\}_{k=1}^{\infty}$ of pairwise disjoints sets with $A_k \in \mathcal{A}$ it holds

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k).$$

A measure takes a set and gives it a value, with a couple of rules. We won't discuss many properties here, but heuristically it is very similar to measuring anything in the real world as well. The last rule says: if you measure a couple of things that don't interchange you can add them up and you get the same value as if you put everything together and measure it. Measures allow us to measure events. We can now define a probability measure:

Definition 0.4. Let $\mathcal{A}$ be a σ-Algebra over Ω. A probability measure is a function $\mathbb{P} : \mathcal{A} \to [0, 1]$ with

(i) $\mathbb{P}(\Omega) = 1$

(ii) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{A}$

(iii) For all countable sets $\{A_k\}_{k=1}^{\infty}$ of pairwise disjoints sets with $A_k \in \mathcal{A}$ it holds

$$\mathbb{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

The triple $(\Omega, \mathcal{A}, \mathbb{P})$ is called probability space.

A probability measure is a measure that only gives values in $[0, 1]$. Another helpful concept is the notion of independence.

Definition 0.5. The events $A, B \in \mathcal{A}$ are called stochasticly independent if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Now we can measure events, but we have a very limited understanding what an event is. To describe that, we use random variables. To define random variables, we need the concept of measurability.

Definition 0.6. Let Ω, Θ be sets and $\mathcal{A} \subset \mathcal{P}(\Omega), \mathcal{E} \subset \mathcal{P}(\Theta)$. A function $f : \Omega \to \Theta$ is called $(\mathcal{A}, \mathcal{E})$-measurable if

$$f^{-1}(E) = \{\omega \in \Omega : f(\omega) \in E\} \in \mathcal{A} \text{ for all } E \in \mathcal{E}.$$

Measurability just means that we can translate one event into something in the other set. It is usually no problem and nearly all functions are measurable with respect to certian subsets of the power sets. In practice, we don't care about it too much. Only in exotic cases, it might be checked. We can now define random variables.

Definition 0.7. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A $(\mathcal{A}, \mathcal{B}_{\mathbb{R}})$-measurable function $X : \Omega \to \mathbb{R}$ is called random variable.

A random variable allows us to translate an event (that is rather abstract) into a value that we can understand. Usually, we ignore what Ω is (and in most cases we just don't know) and work directly on $\mathbb{R}$ through the random variable X. Thus, it makes now sense to describe the probability measure through random variables.

Definition 0.8. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X a random variable. The induced probability by X is

$$\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)).$$

for $B \subset \mathbb{R}$. The function

$$F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \mathbb{P}(X \leq x)$$

is called cumulative distribution function.

There is one last concept that we need to be able to compute the probability. This concept is the measure integral. These things are quite complicated and I won't be able to explain them fully here, but they are key to understand Ito integrals that we will see later on. Instead, I will only mention them here and describe a method to calculate them.

Definition 0.9. A measure μ is absolutely continuous with respect to a measure ρ if for every ρ-measurable set A,

$$\rho(A) = 0 \Rightarrow \mu(A) = 0$$

We write $\mu \ll \rho$.

Note that $\mathbb{P}^X$ is a measure on Borel subsets of the real line. The key theorem to compute probabilities is the Radon-Nikodym theorem.

Theorem 0.10. If $\mathbb{P}^X \ll \rho$, then there exists a measurable function $f : \mathbb{R} \to [0, \infty)$ such that for every $B \in \mathcal{B}_{\mathbb{R}}$,

$$\mathbb{P}^X(B) = \int_B f d\rho.$$

We call f the probability density function.

Here, the integral $\int_B f d\rho$ is a measure integral and it is not clear how to compute that. There are specific rules and definitions how to handle this object, but we will skip this for now. Usually, one is only interested in ρ being either the Lebesgue measure λ (if Ω has uncountable infity many elements) or the counting measure μ (if Ω is countable). The distinction whether Ω is countable or not is also the difference between continuous (λ) or discrete (μ) random variables. Note that in nearly all cases, it holds that

$$\int_B f d\lambda = \int_B f(x)dx.$$

and

$$\int_B f d\mu = \sum_{x \in B} f(x).$$

In nearly all cases, $\mathbb{P}^X$ is absolute continuous with respect to either the Lebesgue or the counting measure. Mostly, the Lebesgue integral is the usual Riemann integral. If $\mathbb{P}^X$ is not absolute continuous with respect to either the Lebesgue or the counting measure, the density f does not exists and it is really hard to compute anything. We will not cover these cases here.

We now covered the most important definitions. There are still tons of things to define, explore and deduce, but at this point we know the base on what the probability definition is resting on and have tools to compute it. We finish this small crash course by defining some important values as the expectation, variance, covariance and how to handle multiple dimensions. In the following, we omit the X in $\mathbb{P}^X$ and just write $\mathbb{P}$.

Definition 0.11. Let X, Y be a random variables and f the density of X. We define

- Expectation: $\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)dx$ (if X is continuous) and $\mathbb{E}[X] = \sum_{i=0}^{\infty} x_i f(x_i)$ (if X is discrete),
- Variance: $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,
- Covariance: $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

For a function g, it also holds that

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x)dx$$

and

$$\mathbb{E}[g(X)] = \sum_{i=0}^{\infty} g(x_i) f(x_i)$$

for X continuous and discrete, respectively.

All the concepts that are presented here can be similarly defined on $\mathbb{R}^n$, $n \in \mathbb{N}$. Here, I only losely mention that in this case, if we observe two random variables (X, Y), we have to change the event space to $\Omega^X \times \Omega^Y$ and define all the above things in the same way. The only very important thing is that the density changes. Since we have two random variables the density becomes $f_{X,Y}(x, y)$ and the probability is computed via

$$\mathbb{P}((A, B)) = \int_A \int_B f_{X,Y}(x, y) dx dy,$$

but the marginal densities $f_X(x)$, $f_Y(y)$ still exist. It holds that

$$\int_{\mathbb{R}} f_{X,Y}(x, y) dx = f_Y(y)$$
$$\int_{\mathbb{R}} f_{X,Y}(x, y) dy = f_X(x).$$

If X and Y are independent it holds that $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.