Probability Calculus 2021/2022
Lecture 1

1. Some Historical Remarks

Probability Calculus – in its present form – dates to the 1930s and it is thus a fairly new discipline of Mathematics. It would not be true to say, however, that probabilistic problems were strange do Mathematicians earlier. Probabilistic considerations may be traced back to the 17th century (some mathematical historians trace even further), and especially to Pascal and Fermat, whose correspondence pertaining to several probabilistic problems (among them, the paradox of Chevalier de Méré, and the problem of the division of stakes in an unfinished game) is now considered "classic". It would not be inappropriate to say that the foundations of probability theory lie in hazard games, as it is precisely the problems posed by those who "practiced" random games professionally and came upon properties they could not explain themselves that gave rise to the evolution of the discipline. Ad-hoc solutions to specific problems were provided throughout the centuries, but it is the work of Kolmogorov (published in 1933) which finally provided the foundations which permitted to formalize the problems posed by researchers and provide unambiguous solutions to many of them.

2. Basic combinatorics

In probability considerations, and especially in the classic scheme, it is often useful to refer to several combinatorial models. The most important are:

Variations with repetitions. The number of k-element sequences of elements of a given set A, if repetitions are allowed, is equal to $n \cdot n \cdot \ldots \cdot n = n^k$, where by n we denote the number of elements of A.

Variations without repetitions. The number of k-element sequences of elements of a given set A, if repetitions are not allowed, is equal to $n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = n!/(n-k)!$, where again by n we denote the number of elements of A. The formula makes sense for $k \leqslant n$; if $k > n$, the number is 0.

Permutations. The number of sequences of all elements of a given set A (consisting of n elements) is equal to $n!$. A special case of variations without repetitions.

Combinations. The number of k-element subsets of a given set A (consisting of n elements) is equal to $\binom{n}{k}$, where

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{if } 0 \leqslant k \leqslant n, \\ 0 & \text{otherwise} \end{cases}$$

3. Basic Definitions and Notation

If we wish to apply mathematical considerations to real-life problems, we need a notation to capture the outcomes of random experiments. For a given experiment, we will therefore need:

- a way to describe a single experiment outcome. We will usually denote a specific outcome by ω and call it an **elementary event**.
- a set of all possible (elementary) outcomes. We will denote this set of all possible ωs by Ω and call it a **sample space**.
- an **event** A, B, C, etc. – a subset of the sample space Ω, i.e. a set of possible outcomes which, for some reason or another, we wish to group together.

Throughout the lecture, we will assume the following notation:
If ω - is an elementary event and A an **event**, then
- if $\omega \in A$, we will say A occurred;
- if $\omega \notin A$, we will say A did not occur, or that the **complement of** A, which we define as $A' = \Omega \setminus A$, occurred.
- Ω is a certain event,

- $\emptyset$ is an impossible event,
- $A \cap B$ - both A and B occurred,
- $A \cap B = \emptyset$ - A and B are disjoint,
- $A \cup B$ - A or B occurred,
- $A \setminus B = A \cap B'$ - A occurred and B did not occurr
- $A \subseteq B$ - A implies B.

Let's see how this notation works in practice for some basic experiments.

(1) We flip a coin. There are two possible outcomes: heads and tails, so we may define $\Omega = \{H, T\}$. $|\Omega| = 2$.
(2) We roll a cubic die. Then, we may define $\Omega = \{1, 2, 3, 4, 5, 6\}$. $|\Omega| = 6$.
(3) We roll a pair of cubic dice. Then, if we take into account the results on both dice, $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}$. If, however, we look at the sum of points obtained, we have $\Omega = \{2, 3, \ldots, 12\}$. Note that this latter case is different than the previous cases in that the elementary events are not equally probable (for now, intuitively).
(4) We draw 13 cards from a deck of 52 cards. The outcome is a 13-element subset of the set of cards, thus Ω is a set of 13-element combinations and $|\Omega| = \binom{52}{13}$. In this case, we assume the order of the cards is not important.
(5) We toss a coin until we obtain heads. The outcome may be represented by the number of tosses, in which case we have $\Omega = \{1, 2, 3, \ldots\}$. This is an example of an infinite sample space.
(6) We throw a needle on a table (Buffon's needle). We measure the angle between the needle and a given edge of the table. The outcome may be represented by a number from the range $[0, 2\pi)$, and thus $\Omega = [0, 2\pi)$. This type of experiment is what we call a continuous experiment.

4. σ-algebras

We may not be interested in a particular event-outcome of a random experiment, but rather whether or not this element is a part of a given subset. For example, if we had a wish to determine who pays for the meal in a restaurant by flipping a coin, but we did not have a coin but had a die instead, we could establish that person A pays if the result is $\{1, 2, 3\}$, and person B pays if the result is $\{4, 5, 6\}$. Then, we are only interested in who pays for the meal, and not in the particular result of the roll (for example, we are interested whether A pays meaning the result was in the set $\{1, 2, 3\}$ and not whether the result was specifically 1 or 2 or 3).

In many simple cases, we may be interested in all possible results (the set of all possible results, i.e. the set of all subsets of Ω, is denoted 2^{Ω}). In other cases (for example in continuous experiments), it may simply be impossible to distinguish (or rather measure) all possible sets of outcomes. Given the sample space Ω, we will need to describe the set of events we wish to be able to measure. This set, if it is to behave reasonably, needs to be closed to sums, intersections and complements. If we allow infinite sums (and intersections), we come to the definition of a σ-algebra.

Definition 1. *A family $\mathcal{F}$ of subsets of Ω is called a* σ-algebra*, if*

$$
\begin{aligned}
&(i) && \emptyset \in \mathcal{F}, \\
&(ii) && A \in \mathcal{F} \Rightarrow A' \in \mathcal{F}, \\
&(iii) && A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.
\end{aligned}
$$

5. Probability and a Probability Space

We all have an intuitive understanding of probability, and this intuitive understanding is that the probability resembles the fraction of favorable events (in a large sample). For example, if we toss a symmetric coin multiple times, we expect that, on average, in a long series of tosses, we will obtain more or less the same fraction of heads and tails. More generally, for a given sample space Ω and an event A, we expect that the proportion of event outcomes where A occurred should more or less resemble the probability that A will occur. Formally, if

$$\rho_n(A) = \frac{\text{number of occurrences of } A}{n},$$

then we expect that the limit of $\rho_n(A)$ as n approaches infinity should be equal to the probability of A. However, we do not have the tools yet to prove that such a limit exists, nor that it is equal to the probability we are looking for. We may, on the other hand, "reverse" the question and formally define probability as a function that satisfies the basic properties that the frequencies have. These most basic properties are the following:

$$\begin{aligned} (i) &\quad 0 \leqslant \rho_n(A) \leqslant 1, \\ (ii) &\quad \rho_n(\Omega) = 1, \\ (iii) &\quad A \cap B = \emptyset \Rightarrow \rho_n(A \cup B) = \rho_n(A) + \rho_n(B). \end{aligned}$$

We will modify the third property to fit the definition of the σ-algebra set of interesting events, which includes infinite sums, to obtain the following definition of a **probability measure** $\mathbb{P} : \mathcal{F} \to [0, 1]$:

$$\begin{aligned} (i) &\quad 0 \leqslant \mathbb{P}(A) \leqslant 1, \\ (ii) &\quad \mathbb{P}(\Omega) = 1, \\ (iii) &\quad \text{if } A_1, A_2, \ldots \in \mathcal{F} \text{ are pairwise disjoint, then } \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \end{aligned}$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space** or a **probability triple**. Note that all three elements of this triple are necessary to formally describe the random experiment we wish to formalize; a change in $\mathcal{F}$ results in the need to change $\mathbb{P}$, and for a given set of Ω and $\mathcal{F}$ we may define different probability measures. The choice of the model that best fits reality is sometimes not an easy one.

(1) A coin toss. $\Omega = \{H, T\}$. We can easily use the complete set of subsets for our σ-algebra: $\mathcal{F} = 2^{\Omega} = \{\{H\}, \{T\}, \Omega, \emptyset\}$. If the coin is symmetric, we may assume $\mathbb{P}(\{H\}) = \frac{1}{2}$, $\mathbb{P}(\{T\}) = \frac{1}{2}$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$. If we wish to account for an asymmetric coin, we may put $\mathbb{P}(\{H\}) = p$, $\mathbb{P}(\{T\}) = 1 - p$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$, for a given value of $p \in [0, 1] \setminus \frac{1}{2}$.
(2) A die roll. $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega}$, $\mathbb{P}(\omega) = 1/6$, and as can easily be shown, therefore $\mathbb{P}(A) = |A|/6$ for any subset A of Ω.
(3) More generally: **The classical probability scheme**. If Ω is a finite set, we may always put $\mathcal{F} = 2^{\Omega}$. If our experiment is such that *all elementary events are equally probable* (and thus the probability of obtaining a specific result is always $\frac{1}{n}$, where n is the number of elements of Ω), we have, for $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

(4) From a deck of cards we are randomly assigned a set of 13. We have seen above that Ω is the set of all 13-element combinations, and $|\Omega| = \binom{52}{13}$. Since, intuitively, all possible outcomes are equally probable, we may use the classical probability scheme. If, for example, we wish to determine the probability of obtaining four aces and four kings among the thirteen cards, we have that $|A| = \binom{44}{5}$, which gives us $\mathbb{P}(A) = \binom{44}{5} / \binom{52}{13}$.

The axiomatic definition of Probability introduced in this section fulfills a set of intuitive properties.

Theorem 1. *Let A, B, $A_1, A_2, \ldots \in \mathcal{F}$. Then*

(i) $\mathbb{P}(\emptyset) = 0$,

(ii) *If $A_1, A_2, \ldots, A_n$ are pairwise disjoint, then* $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.

(iii) $\mathbb{P}(A') = 1 - \mathbb{P}(A)$.

(iv) *If $A \subseteq B$, then* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.

(v) *If $A \subseteq B$, then* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

(vi) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

(vii) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Theorem 2 (Inclusion-exclusion principle)**.** *If $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then*

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = &\sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \ldots \\ &+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n).\end{aligned}$$

These properties not only facilitate the calculation of specific probabilities, but also permit us to define probability on the set of elementary events only – the probability of all other events is then determined unambiguously. For example,

(5) If $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$ is a countable (finite or infinite) set, we can define the probability function using a set of weights assigned to specific values of ωs. Formally, let $p_1, p_2, \ldots$ be a series of nonnegative numbers, such that their sum is equal to 1. We may then define $\mathcal{F} = 2^{\Omega}$, and $\mathbb{P}(\{\omega_i\}) = p_i$, $i = 1, 2, \ldots$. Then, for all $A \in \mathcal{F}$, we have

$$\mathbb{P}(A) = \sum_{i} 1_A(\omega_i) \cdot p_i,$$

where $1_A(x)$ is an **indicator function** of the set A, namely $1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$