MDI Probability Theory.

Class 2: Concept of probability. Classical approach.

We have already covered two of three main probability basement's stones: these are **Sample space** Ω and **Event space** $\mathcal{F}$. Now we are left with one conception which completes the basics of Probability Theory, and gives us initial tools to describe random experiments and to go further.

Functions

For many many years mathematicians deal with sets. Moreover, they pay attention at how elements from one set interact with and relate to elements from the other set. Let us recall general definition of a function:

Def. Function F is a relation law between sets X (domain) and Y (codomain), such that for each element $x \in X$ there exists only one element $y \in Y$.

We also sometimes write it down with the help of *logic quantifiers*:
Function F is a relation law between sets X and Y, such that:

$$\forall x \in X \quad \exists!\, y \in Y.$$

Or just in a short arrow notation, which specifies domain and codomain of a function: $F : X \to Y$.

Probability concept

Now let's reveal why do we talk about functions anyway.
We talked about describing the random experiment, we can establish the sample space, the event space, but we are lacking the very important part. The overall motivation is to somehow get an information about *likeliness* or certainty of specific event.

The main breakthrough in Kolmogorov's general approach was to think about probability as of *function* which assigns an event to its certainty coefficient. And for operational simplicity this certainty coefficient is between 0 and 1, where 0 means "event is impossible", and "1" means "event will inevitably happen". Some more advanced books also call this relation a *measure*, which is more special type of function, and it sounds very in-place: we would like to *measure* event we are interested in and find out number between 0 and 1 as its "degree of likeliness".

So, third and last building block for modern probability is to interpret probability as a fuction which takes an event as an input, and returns real value as an output:

$$\mathcal{P} : \Omega \to [0, 1] \subset \mathbb{R}.$$

Example of notation: $P(A) = 0.6$, $P(B) = 0.1$, $P(A \cup B) = 0.2$.

Probability axioms

Not every function which can be defined on such domain and codomain will be probability function. We also need some important properties to be held.

1. (Nonnegativity) $P(A) > 0$ for every event $A \in \mathcal{F}$.
2. (Additivity) If A and B are two disjoint events ($A \cap B =$), then $P(A \cup B) = P(A) + P(B)$.
   This can be generalized further: If $A_1, A_2, \ldots, A_n$ are pairwise disjoint events ($A_i \cap B_j =,\ \forall i \neq j$), then $P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P\left(A_k\right)$.
3. (Normalization) The probability of event which is the whole sample space Ω: $P(\Omega) = 1$. By this we ensure that maximum value of our function is 1, as we required before.

These are 3 main basic probability axioms. There exist few more properties which can be naturally derived from the latter. Function $\mathcal{P} : \Omega \to [0, 1]$, for which the axioms above hold is called *Probability function* or *Probability measure*.

Addition to Axioms

We also need to fins out what happens if conditions for axiom #2 do not work. If A and B are two events such that $A \cap B \neq$, then we can not just sum up probabilities for $(A \cup B)$. For this case we use a special formula, which has its own name — *inclusion–exclusion principle*. Generally, it says that:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

But we use it in terms of probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

It is clear that if A and B are disjoint than this formula takes form as in Axiom # 2.

Probability of event (Discrete Probability Law)

In many cases we do not know value of probability function for each event beforehand. Moreover, this is our main goal while describing the random experiment -- to find out probability of the specific event. It turns out that, in order to do that, we need to know only some probabilities of simple outcomes. By using Probability axioms we can understand how to calculate probabilities of complex events, gathered from simple outcomes.

We have already seen that one simple outcome also can be an event we are interested in. In this case we treat outcome ω_i not as element of the Sample space, but rather set $\{\omega_i\}$ as an element of the Event space: $\{\omega_i\} \in \mathcal{F}$.

Suppose that we need to calculate probability of the event A, $A = \{\omega_1, \dots, \omega_n\}$, composed of n simple outcomes. But each of them also is an event, so the following is also true: $A = \{\{\omega_1\} \cup \{\omega_2\} \dots \cup \{\omega_n\}\}$. Because outcomes are the the smallest building blocks of Sample space Ω, then:

$$\forall i \neq j: \ \{\omega_i\} \cap \{\omega_j\} = \emptyset.$$

That means they are pairwise disjoint, and we can apply Additivity axiom:

$$P(A) = P\left(\{\{\omega_1\} \cup \{\omega_2\} \dots \cup \{\omega_n\}\}\right) = \sum_{k=1}^{n} P(\omega_i).$$

In a nutshell, to calculate probability of a complex event, you need to *sum up* probabilities of *every simple outcome* this event contains.

Classical approach

This may be the oldest intuition of the probability in history!

Main assumption: If the Sample space consists of finite number of elements (say, n), which are **equally likely** to happen (*i.e.* all single-element events have the same probability), then the probability of any event A is calculated as follows:

$$P(A) = \frac{\text{number of elements in } A}{n} = \frac{|A|}{|\Omega|}.$$