Probability Theory and Statistics

Lecture 3

Bence Csonka

Budapest University of Technology and Economics

csonkab@edu.bme.hu

September 15

Bayes reminder

Theorem (Simple Bayes' theorem)

Let $A, B \in \mathcal{F}$ *with* $\mathbb{P}(A) > 0$ *and* $\mathbb{P}(B) > 0$. *Then*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}.$$

Bayes reminder

Theorem (Simple Bayes' theorem)

Let $A, B \in \mathcal{F}$ *with* $\mathbb{P}(A) > 0$ *and* $\mathbb{P}(B) > 0$. *Then*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\overline{A})\mathbb{P}(\overline{A})}.$$

Theorem (Bayes' theorem)

Let $B, A_1, \ldots, A_n \in \mathcal{F}$ *with* $\mathbb{P}(B) > 0$ *and* $\mathbb{P}(A_i) > 0$ *for all* i, *and suppose* $A_1, \ldots, A_n$ *form a complete system. Then*

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(B|A_1)\mathbb{P}(A_1)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Example for Bayes theorem

After tossing a coin, two cases are possible:

- If the result is heads, then we roll a die once.
- If the result is tails, then we roll a die twice.

a) What is the probability that we roll only sixes?

b) What is the probability that, given we roll only sixes, the coin toss at the beginning was heads?

Examples of classical probability spaces — combinatorics, urn models

Classical probability spaces

Reminder: a classical probability space: Ω is finite, and every outcome/elementary event has the same probability.

Classical probability spaces

Reminder: a classical probability space: Ω is finite, and every outcome/elementary event has the same probability.

We will look at some classical examples today. These are the so-called *urn models*, where an urn (bag, etc.) contains balls of different types, and we draw some of them, either with or without replacement, and either with or without keeping track of order.

Classical probability spaces

Reminder: a classical probability space: Ω is finite, and every outcome/elementary event has the same probability.

We will look at some classical examples today. These are the so-called *urn models*, where an urn (bag, etc.) contains balls of different types, and we draw some of them, either with or without replacement, and either with or without keeping track of order.

The goal is to model a given random experiment by a suitable probability space. Choosing the right model is crucial, but there need not be a unique correct model.

Classical probability spaces

Reminder: a classical probability space: Ω is finite, and every outcome/elementary event has the same probability.

We will look at some classical examples today. These are the so-called *urn models*, where an urn (bag, etc.) contains balls of different types, and we draw some of them, either with or without replacement, and either with or without keeping track of order.

The goal is to model a given random experiment by a suitable probability space. Choosing the right model is crucial, but there need not be a unique correct model.

To determine the "number of favorable cases" and the "number of all cases", we first review the basic quantities of combinatorics, the so-called elementary counting results.

Variations (arrangements)

Definition (k-permutations (variations) of an n-element set)

Let k, n be natural numbers. An ordering of k distinct elements chosen from n distinct elements. This is defined for $0 \leq k \leq n$.
Their number: $V(n, k)$.

Definition (k-permutations with repetition)

A length-k sequence from n distinct elements, repetitions allowed. Here $k > n$ is also possible.
Their number: $V_{rep}(n, k)$.

Variations (arrangements)

Definition (k-permutations (variations) of an n-element set)

Let k, n be natural numbers. An ordering of k distinct elements chosen from n distinct elements. This is defined for $0 \leq k \leq n$.
Their number: $V(n, k)$.

Definition (k-permutations with repetition)

A length-k sequence from n distinct elements, repetitions allowed. Here $k > n$ is also possible.
Their number: $V_{rep}(n, k)$.

We use the usual factorial notation:

$$n! = n(n-1) \cdot \ldots \cdot 1.$$

Hence $0! = 1$ (the value of an empty product is always 1, while that of an empty sum is 0).

Variations

Theorem

$$V(n,k) = \frac{n!}{(n-k)!}.$$

(Proof)

Variations

Theorem

$$V(n,k) = \frac{n!}{(n-k)!}.$$

(Proof)

Theorem

$$V_{rep}(n,k) = n^k.$$

(Proof)

Variation without repetition (k-permutations)

Problem. From 10 distinct books, how many different ordered selections of 4 books can you place on a shelf?

Answer. $$V(10,4) = \frac{10!}{(10-4)!} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040.$$

Variation with repetition (k-permutations with repetition)

Problem. How many different 5-digit codes can be formed using the digits 0–9 if digits may repeat?

Answer. $V_{\text{rep}}(10, 5) = 10^5 = 100{,}000.$

Urn model

Urn model related to variations: an urn contains n distinct balls, we draw k balls without replacement.
Possible outcomes: the possible orders of the k balls drawn.

One possible classical probability space for this experiment:
$\Omega = \{\text{the } k\text{-permutations of } n \text{ distinct balls}\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ (every subset is an event), $\mathbb{P}(\omega) = \dfrac{1}{V(n,k)} = \dfrac{(n-k)!}{n!}$ for all $\omega \in \Omega$.

Urn model 2

Urn model for variations with repetition: an urn contains n distinct balls, we draw one ball k times, each time returning it to the urn. Possible outcomes: the possible length-k sequences of draws.
A classical probability space for this experiment:
$\Omega = \{$the k-permutations with repetition of n distinct balls$\}$, $\mathcal{F} = \mathcal{P}(\Omega)$,
$\mathbb{P}(\omega) = \dfrac{1}{V_{\text{rep}}(n, k)} = \dfrac{1}{n^k}$ for all $\omega \in \Omega$.

This urn model is often called the Maxwell–Boltzmann urn model $\rightarrow$ statistical physics. (You don't need to memorize such names.)

Cartesian products of sets; choosing Ω

Previous example: $\Omega = \{$the k-permutations with repetition of n balls$\}$
$\longrightarrow$ how can we conveniently represent the elements of this set?

Cartesian products of sets; choosing Ω

Previous example: $\Omega = \{$the k-permutations with repetition of n balls$\}$
$\longrightarrow$ how can we conveniently represent the elements of this set?

For this (and many other things) we introduce the following notion/notation:

Definition

For sets $H_1, \ldots, H_n$, their Cartesian product *is the set of ordered n-tuples $(x_1, \ldots, x_n)$ with $x_1 \in H_1, \ldots, x_n \in H_n$. We denote it by $H_1 \times \ldots \times H_n$.*
If $H_1 = \ldots = H_n = H$ (for some set H), we also use the notation $H^n = \underbrace{H \times \ldots \times H}_{n \text{ factors}}$.

A familiar example: $\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ factors}}$, the set of ordered n-tuples of real numbers („n-high number columns”).

Cartesian product and choosing Ω

Previous example: $\Omega = \{$the k-permutations with repetition of n balls$\}$
$\longrightarrow$ how do we represent the elements of this set?

One possible representation of an outcome is: $(x_1, \ldots, x_k)$, where x_i is the element drawn i-th, an element of the set of balls in the urn.
Let H be the set of balls in the urn. Then one possible choice for Ω is:

$$\Omega = \underbrace{H \times H \times \ldots \times H}_{k \text{ factors}} = H^k.$$

(And $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(\omega) = 1/n^k$ for all $\omega \in \Omega$.)

E.g., rolling a die with 2 distinguishable dice (or rolling the same die twice in a row): $\Omega = \{1, 2, 3, 4, 5, 6\}^2 = [6]^2$. Typical elements: $(3, 4)$ means the first roll is three and the second is four; $(5, 5)$ means both rolls are five.

Cartesian product and variations without repetition

An earlier example:
$\Omega = \{$the k-permutations without repetition of n balls$\}$.

Observation: if H is the set of balls in the urn, then $\Omega = H^k$ is still a possible choice, since every outcome can be written as $(x_1, \ldots, x_k)$ (x_i: the element drawn in the i-th step).
However, for $n > k > 1$ this will not be a classical probability space, because elements of the form $(x_1, x_1, x_3, \ldots, x_k)$ are *not* possible outcomes. Here

$$\mathbb{P}\big((x_1, \ldots, x_k)\big) = \begin{cases} \dfrac{1}{V(n,k)}, & \text{if } x_1, \ldots, x_k \text{ are pairwise distinct,} \\ 0, & \text{otherwise.} \end{cases}$$

Definition (A permutation of n elements)

An ordering of n distinct elements.

:

Permutations 1

Definition (A permutation of n elements)

An ordering of n distinct elements.

:

Theorem

The number of permutations of n elements is $n!$.

(Proof: this is the special case $k = n$ of variations.)

Permutations 1

Definition (A permutation of n elements)

An ordering of n distinct elements.

:

Theorem

The number of permutations of n elements is n!.

(Proof: this is the special case $k = n$ of variations.)

Accordingly, the related urn model is a special case of the one for variations: an urn contains n distinct balls; we repeatedly draw one ball without replacement until all are drawn; outcomes are the orders of the drawn balls.

Permutations 2

An equivalent, more common definition:

Definition

The permutations (of order n) are the bijections of the set $[n]$ onto itself.

Permutations 2

An equivalent, more common definition:

Definition

The permutations (of order n) are the bijections of the set $[n]$ onto itself.

Permutations are usually denoted by Greek letters (π, σ, etc.).

A common notation is, e.g., $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$: this is the permutation $\pi\colon [3] \to [3]$ with $\pi_1 := \pi(1) = 3$, $\pi_2 = 2$ and $\pi_3 = 1$.
Sometimes this is abbreviated as the permutation 3 2 1, though this notation is somewhat outdated.

A frequent misunderstanding: a permutation itself is a function $[n] \to [n]$, not the associated rook placement.

Permutations with repetition

Let $n, l, k_1, \ldots, k_l$ be natural numbers with $n = k_1 + \ldots + k_l$.

Definition

An n-element permutation with repetition (with parameters $k_1, \ldots, k_l$) is an arrangement of n elements where k_1 elements are identical, k_2 elements are different from the first group but identical among themselves, and so on, finally k_l elements are different from all previous ones but identical among themselves.

Theorem

Their number is

$$\frac{n!}{k_1! \ldots k_l!}.$$

(Proof.)

Urn model for permutations with repetition

An urn contains $n = k_1 + \ldots + k_l$ balls, of which k_1 are identical, k_2 are different from those but identical among themselves, and so on, finally k_l are different from all previous ones but identical among themselves.
We draw one ball at a time without replacement until they run out; outcomes are the orders in which the balls are drawn.

A classical probability space for this urn model:

$$\Omega = \{\text{the permutations with repetition of } n \text{ balls with parameters } k_1, \ldots, k_l\},$$
$$\mathcal{F} = \mathcal{P}(\Omega).$$

$\mathbb{P}(\omega) = \dfrac{k_1! \ldots k_l!}{n!}$ for all $\omega \in \Omega$.

Combinations without repetition

Let n, k be natural numbers with $0 \leq k \leq n$.

Definition

A k-combination without repetition from n elements is a choice of k elements from n, where order does not matter. Their number: $C(n, k)$.

Theorem

$$C(n, k) = \frac{n!}{k!\,(n-k)!}.$$

Combinations without repetition

Let n, k be natural numbers with $0 \leq k \leq n$.

Definition

A k-combination without repetition from n elements is a choice of k elements from n, where order does not matter. Their number: $C(n,k)$.

Theorem

$$C(n,k) = \frac{n!}{k!\,(n-k)!}.$$

(Proof — via the connection to variations without repetition.)
Standard notation:

$$\underbrace{\binom{n}{k}}_{\text{binomial coefficient}} := \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k}.$$

Properties of binomial

Properties (also known from high school/other university courses?):

- $\binom{n}{k} = \binom{n}{n-k}$.
- $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = \binom{n}{n-1} = n$.
- $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. (See Pascal's triangle.)
- Binomial theorem: see analysis. Simple case:
  $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ for positive integer n and $a, b > 0$.
  Important special case: $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Urn model for combinations without repetition

We have n distinct balls in an urn; we draw k of them without replacement; order does not matter.

$\Omega = \{$the k-combinations without repetition of n balls$\}$, $\mathcal{F} = \mathcal{P}(\Omega)$,
$\mathbb{P}(\omega) = \dfrac{1}{C(n,k)} = \dfrac{1}{\binom{n}{k}} = \dfrac{k!\,(n-k)!}{n!}$ for all $\omega \in \Omega$.

Example: 5-out-of-90 lottery: $n = 90$, $k = 5$.
A given ticket wins the jackpot with probability
$\dfrac{1}{\binom{90}{5}} = 2.27535075214449 \cdot 10^{-8}$.

When does order matter? (without repetition)

A common modeling question: does order matter?

Often this is clear, but not always. In some cases, multiple correct models exist. If a problem can be solved both by variations without repetition and by combinations, then both methods yield the same result.

Example:
Grandma Magda fills in 20 lottery tickets at random (on each ticket she chooses 5 numbers from 1 to 90). Determine the following probability: Any two tickets are different.

Combinations with repetition

Definition

Let n, k be natural numbers. A k-combination with repetition from n types is a selection of k elements from n types, repetitions allowed. Their number: $C_{rep}(n, k)$. (Here $k > n$ is possible.)

Example: choosing k pastries from n pastry types.

Combinations with repetition

Definition

Let n, k be natural numbers. A k-combination with repetition from n types is a selection of k elements from n types, repetitions allowed. Their number: $C_{rep}(n, k)$. (Here $k > n$ is possible.)

Example: choosing k pastries from n pastry types.

Encoding combinations with repetition by 0–1 strings.
If $a_1, \ldots, a_n \geq 0$ and $a_1 + a_2 + \ldots + a_n = k$, then the corresponding combination with repetition ("we take a_1 pieces of type 1, ..., a_n pieces of type n") is encoded as

$$\underbrace{1\ldots1}_{a_1 \text{ times}}0\underbrace{1\ldots1}_{a_2 \text{ times}}0\ldots0\underbrace{1\ldots1}_{a_n \text{ times}}.$$

E.g., $12 = 6 + 0 + 0 + 3 + 2 + 1$ is encoded by 11111100011101101.
Such a code always has $n - 1$ zeros and k ones. Hence these are 0–1 strings of length $n + k - 1$ with k ones. It follows that:

Proof

Encoding combinations with repetition by 0–1 strings.
If $a_1, \ldots, a_n \geq 0$ and $a_1 + a_2 + \ldots + a_n = k$, then the corresponding combination with repetition ("we take a_1 pieces of type 1, ..., a_n pieces of type n") is encoded as

$$\underbrace{1\ldots1}_{a_1 \text{ times}} 0 \underbrace{1\ldots1}_{a_2 \text{ times}} 0\ldots0 \underbrace{1\ldots1}_{a_n \text{ times}}.$$

E.g., $12 = 6 + 0 + 0 + 3 + 2 + 1$ is encoded by 11111100011101101.

Such a code always has $n - 1$ zeros and k ones. Hence these are 0–1 strings of length $n + k - 1$ with k ones. It follows that:

Theorem

$C_{rep}(n, k) = \binom{n+k-1}{k}$.

(Proof)

Random Variables

So far: Random experiments $\rightarrow$ outcomes, events, $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

Instead of working with the abstract probability space, it is often simpler to work with real-valued functions.

Definition

A (one-dimensional) random variable is a function $X : \Omega \rightarrow \mathbb{R}$ that assigns to each outcome $\omega \in \Omega$ a real number $X(\omega) \in \mathbb{R}$.

Random Variables

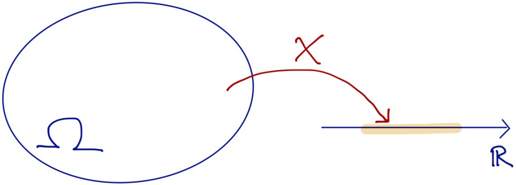
So far: Random experiments $\rightarrow$ outcomes, events, $(\Omega, \mathcal{F}, \mathbb{P})$ probability space.

Instead of working with the abstract probability space, it is often simpler to work with real-valued functions.

Definition

A (one-dimensional) random variable is a function $X : \Omega \rightarrow \mathbb{R}$ that assigns to each outcome $\omega \in \Omega$ a real number $X(\omega) \in \mathbb{R}$.

X has to satisfy certain conditions $\rightarrow$ precise definition: coming soon.



Example for random variables

Example: modeling a coin toss, $\Omega = \{H, T\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. $X(H) := 1$, $X(T) := 0$. Then X is a random variable, namely the number of heads obtained in the experiment.

Random Variables

When there are only finitely many possible outcomes, one could often regard any function $X\colon \Omega \to \mathbb{R}$ as a random variable without difficulty. However, this is not what we do, because in more complicated cases of Ω it is important that simple subsets of $\mathbb{R}$ (such as intervals) have *preimages* in Ω that are events (that is, elements of the σ-algebra $\mathcal{F}$).
This motivates the following definitions.

Random Variables

When there are only finitely many possible outcomes, one could often regard any function $X \colon \Omega \to \mathbb{R}$ as a random variable without difficulty. However, this is not what we do, because in more complicated cases of Ω it is important that simple subsets of $\mathbb{R}$ (such as intervals) have *preimages* in Ω that are events (that is, elements of the σ-algebra $\mathcal{F}$).
This motivates the following definitions.

Definition

Let $X \colon \Omega \to \mathbb{R}$ be a function. For $x \in \mathbb{R}$, define

$$\{X < x\} \overset{def.}{=} \{\omega \in \Omega \mid X(\omega) < x\},$$

that is, the set of outcomes $\omega \in \Omega$ for which $X(\omega) < x$. These sets are called the level sets of X.

Random Variables

$$\{X < x\} = X^{-1}((-\infty, x)) \subseteq \Omega.$$

Similarly, we can introduce the following notations (where $x < y$ are real numbers):

$$\begin{aligned}
\{X \leq x\} &= \{\omega \in \Omega \mid X(\omega) \leq x\} = X^{-1}((-\infty, x]), \\
\{X > x\} &= \{\omega \in \Omega \mid X(\omega) > x\} = X^{-1}((x, \infty)), \\
\{X = x\} &= \{\omega \in \Omega \mid X(\omega) = x\} = X^{-1}(\{x\}), \\
\{x < X \leq y\} &= \{\omega \in \Omega \mid x < X(\omega) \leq y\} = X^{-1}((x, y]) \quad \text{etc.}
\end{aligned}$$

In fact, for any set $A \subseteq \mathbb{R}$,

$$\{X \in A\} = \{\omega \in \Omega \mid X(\omega) \in A\} = X^{-1}(A).$$

Using the notation $\{X < x\}$, we can now finally give the definition of a random variable.

Random Variables

Definition

The function $X : \Omega \to \mathbb{R}$ *is called a random variable if for every* $x \in \mathbb{R}$ *we have* $\{X < x\} \in \mathcal{F}$.

Definition

The function $X\colon \Omega \to \mathbb{R}$ *is called a random variable if for every* $x \in \mathbb{R}$ *we have* $\{X < x\} \in \mathcal{F}$.

In other words, $X\colon \Omega \to \mathbb{R}$ is a random variable if each of its level sets is an event.

If X is a random variable, then for any $x < y$, the sets $\{X \leq x\}, \{X \geq x\}, \{X > x\}, \{X = x\}, \{x \leq X \leq y\}, \{x < X < y\}, \{x \leq X < y\}, \{x < X \leq y\}$ are also events.
Countable unions of such sets are also events.
(This follows from the properties of a σ-algebra, though we will not prove it here. Some cases are easy, e.g. $\{X \geq x\} = \overline{\{X < x\}}$.)