

```
In [1]: import numpy as np
import pandas as pd

import matplotlib.pyplot as plt

%matplotlib inline

In [2]: from scipy.special import factorial
```

Seminar 1

Seminar outlook

- I introduce myself
- You introduce yourselves
- Administrative announcements
- Counting rules and naive definition of probability

Let's get to know each other

- My name is Nikolai Stulov
- I myself got my BSc from MIPT
- Got my MSc from HSE and Skoltech
- 6 years of experience as Data Scientist and ML researcher
- 4th time teaching Probability at MSAI

I will be happy to hear short introductions from you! (but it's OK if you don't want to)

Course outlook

- Lectures will be distributed several days in advance (perhaps Monday), it's best to watch the lecture before the webinar
- The webinars are mostly delivered in Jupyter notebooks, notebooks will be distributed via GitHub
- You can get a maximum of 10 points for the course. Graded activities: homework, project, exam.
- Homework will be weekly with 2 weeks time to solve. Each homework includes regular and bonus problems. Homeworks are distributed in Telegram, collected with Google Forms.
- Fully solved regular problems in all homeworks will give you 7 points for the

course. Bonus problems will give you additional 2 points for the course. If AI assistants are used, it must be explicitly stated and prompt must be provided.

- A project is a take-home literature review + coding assignment. Project will give you additional 2 points for the course, but is not compulsory for anyone.
- Exam is not compulsory if you get more than 3 points with homeworks. If you get 3 or less, you must take exam. Exam can give you a maximum of 3 points.

Counting rules

Problem 1 (entrance exam)

Russian car plate consists of three letters and three digits. Any digits are permitted, but the only permitted letters are the ones that have English-lookalikes. How many car plates are possible in one region?

Solution 1

How many letters are there?

A, B, C, E, H, K, M, O, P, T, X, Y - total 12 letters. We choose 3 digits from them. Do we sample with or without replacement?

Using sampling with replacement, because there are no restrictions on repetitions of letters:

- We choose three of ten digits: 10^3
- We choose three of twelve letters: 12^3

Since the choice of the digits and the letters is independent, the total number of plates is therefore $10^3 \cdot 12^3 = 1728000$.

Problem 2 (entrance exam)

How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1?

Solution 2

We independently choose each digit. Do we sample with or without replacement?

Using sampling with replacement, because there are no restrictions on repetitions of numbers:

- We choose the first digit from reduced set of 8 digits: 8
- We choose the rest 6 digits: 10^6

The total number of phone numbers is therefore $8 \cdot 10^6$.

$$n \cdot (n - 1) \cdot \dots \cdot (n - k + 1)$$

$$n = k \Rightarrow n \cdot (n - 1) \cdot \dots \cdot (n - n + 1) = n!$$

Problem 3

How many paths are there from the point (0,0) to the point (110,111) in the plane such that each step either consists of going one unit up or one unit to the right?

Solution 3

We will encode a path as a sequence of letters U (for up step) and R (for right step), like $URURURUR \dots UURUR$. How many R s and U s will be in the complete sequence?

The sequence must consist of 110 R s and 111 U s, because we need to get from 0 to 110 horizontally by only moving right and from 0 to 111 vertically by only moving up.

Let's use the factorial rule: the number of shuffles of this UR sequence is $(110 + 111)! = 221!$. Is it correct?

It is not correct, because we do not care about individual permutations of R s and U s, but we counted these permutations as different. We need to **adjust for overcounting**.

We need to get rid of permutations that we counted multiple times. In order to do that, we divide by the number of such permutations, and this gives the correct answer:

$$\frac{221!}{110!111!}$$

```
In [3]: factorial(221) / (factorial(110) * factorial(111))
```

```

/var/folders/33/j0cl7y453td68qb96j7bqcj4cf41kc/T/ipykernel_87094/300022860
4.py:1: RuntimeWarning: overflow encountered in double_scalars
  factorial(221) / (factorial(110) * factorial(111))
/var/folders/33/j0cl7y453td68qb96j7bqcj4cf41kc/T/ipykernel_87094/300022860
4.py:1: RuntimeWarning: invalid value encountered in double_scalars
  factorial(221) / (factorial(110) * factorial(111))

```

Out[3]: nan

Why didn't we overcount previously?

Because we didn't use the number of shuffles formula $n!$, that assumes that objects are distinguishable.

Naive definition

Problem 4 (entrance exam)

A child is playing with cubes with letters A, A, C, E, H, I, K, M, M, S, T, T. What is the probability that a random ordering of the cubes in one line will form the word MATHEMATICS?

Solution 4

A, A, C, E, H, I, K, M, M, S, T, T - total 12 letters.

Let's count the number of favorable cases: 2 ways to get an M, 2 ways to get an A, 2 ways to get a T, one way to get H and E, one left way to get an M, etc all ones.

Multiplying, we get $2 \cdot 2 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \dots = 2^3$.

Let's count the total number of cases: it is $12!$. Does this mean our answer is

$$\frac{2^3}{12!}$$

No, it does not. We overcounted the total number of cases, because we counted twice the cases that differ in the positions of the same letters. We must get rid of the overcounting:

$$\frac{2^3}{\frac{12!}{2!2!2!}} = \frac{2^6}{12!}$$

In [4]: `2 ** 6 / factorial(12)`

Out [4]: 1.3361124472235584e-07

Problem 5 (entrance exam)

A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

$$\frac{neg}{all} = \frac{all - fav}{all} = 1 - \frac{fav}{all}$$

Solution 5

We will compute the probability of the complement.

- All cases: There are 6^6 possible configurations for which robbery occurred where.
- Favorable cases: There are $6!$ configurations where each district had exactly 1 of the 6.

So the probability of the complement of the desired event is $6!/6^6$.

Finally, the probability of some district having more than 1 robbery is $1 - 6!/6^6$.

```
In [5]: 1 - factorial(6) / (6 ** 6)
```

Out [5]: 0.9845679012345679