PROBABILITY AND STATISTICS, Lesson 0.

1. Find the probability p_k that in a group of k people there are at least two celebrating their birthdays on the same day! **Birthday paradox**: surprisingly, $p_{23} \approx 0.5$ and $p_{55} \approx 0.99$.

2. **Birthday holidays:** k workers are employed in a factory. Each working day, each worker manifactures a chair. When any of them has a birthday, there is a holiday (the factory is closed, they do not work at all). Under these conditions, how many workers have to be employed, if they want to maximize the number of yearly produced chairs in the factory? Equivalent problem: A worker's legal code specifies a holiday any day during which at least one worker in a certain factory has a birthday. All other days are working days. How many workers (k) must the factory employ so that the expected number of working man-days is maximized during the year?

3. **Monte Carlo integration:** $\int_0^1 t(x)\,dx = ?$ where $t(x) = e^{x^2+\sin\sqrt{x}}$. Let $X_1, \dots, X_n$ be i.i.d. $\mathcal{U}[0,1]$ random numbers and $Y_i := t(X_i)$. Then

$$\frac{Y_1 + \cdots + Y_n}{n} \to \mathbb{E}(Y) = \int_0^1 t(x)\,dx \qquad \text{as } n \to \infty$$

   almost surely, by the **Laws of Large Numbers**.

4. **Hypothesis testing (Believe or not?)** *Parametric inference:* let $x_1, \dots, x_n$ be the realization of an i.i.d. sample from a Gaussian population with unknown mean μ and known variance σ_0^2. E.g., we measure the weight of n packages of sugar powder (in kg). Let $n = 25$, $\sigma_0 = 0.05$, $\bar{x} = 0.98$. We want to decide about

$$H_0 : \mu = 1(kg) \quad \text{versus} \quad H_0 : \mu \neq 1(kg).$$

   Give the decision with p-values 0.05 and 0.01. (The two-sided critical value belonging to $p = 0.05$ is 1.96 and to $p = 0.01$ is 2.58.)

   For large n ($n \geq 30$), in case of continuous i.i.d. data, $\bar{X}$ is nearly Gaussian (**Central Limit Theorem**).

5. *Nonparametric inference:* decide whether the following *die is fair or not.* We cast it $n = 1200$ times and the frequencies of the sides are: $\nu_1 = 184$, $\nu_2 = 212$, $\nu_3 = 190$, $\nu_4 = 208$, $\nu_5 = 212$, $\nu_6 = 194$. The chi-square statistic is:

$$\chi^2 = \sum_{i=1}^{6} \frac{(\nu_i - 200)^2}{200} = \frac{16^2 + 12^2 + 10^2 + 8^2 + 12^2 + 6^2}{200} = 3.72.$$

   Decide, if $n = 12000$ and $\nu_1 = 1840$, $\nu_2 = 2120$, $\nu_3 = 1900$, $\nu_4 = 2080$, $\nu_5 = 2120$, $\nu_6 = 1940$.

   Decide, if $n = 1200$ and $\nu_1 = 184$, $\nu_2 = 212$, $\nu_3 = 160$, $\nu_4 = 238$, $\nu_5 = 212$, $\nu_6 = 194$.

   (The critical value belonging to $p = 0.05$ is 11.07 and to $p = 0.01$ is 15.09 with $df = 5$.)