Seminar 7

*In all exercises where it is required to estimate a probability, you **must also estimate the accuracy** of the approximation.*

Exercise 1 [M]

There are 1000 students studying at the institute. There are 105 seats in the cafeteria. Each student goes to the cafeteria during the long break with a probability of 0.1. Estimate the probability that on a normal school day:

    a. the cafeteria will be filled to no more than two-thirds capacity;
    b. there will not be enough seats for everyone.

> 💡 Solution
>
> Let X be the number of students who come to the cafeteria. $X \sim \text{Binomial}(n = 1000, p = 0.1)$. We use the normal approximation (CLT): $\mu = np$, $\sigma^2 = np(1-p)$, $X \approx X' \mathcal{N}(\mu, \sigma)$.
>
>   a. We are looking for
>
> $$\mathbb{P}(X \leq 2/3 \cdot 105) \approx \mathbb{P}(X' \leq 70.5) \approx \mathbb{P}\left(\frac{X' - \mu}{\sigma} \leq (70.5 - 100)/\sqrt{90}\right) \approx 0.0009368$$
>
>   b. We are looking for
>
> $$\mathbb{P}(X \geq 105) \approx \mathbb{P}\left(\frac{X' - \mu}{\sigma} \geq (105 - 100)/\sqrt{90}\right) \approx 0.2810$$
>
> **Accuracy**: For the binomial to normal approximation, once we correct with the "half-integer" value (as above), the error for the probability of intervals is uniformly bounded by $1/\sqrt{1 + 12np(1-p)}$. In this case this gives 0.030415, which does not catch the accuracy of the approximation in reality.

Exercise 2 [M]

On average, the number of people infected with the flu virus during an epidemic is 25%, that is, in an organization of 600 people, one can assume an average of 150 infected. Specify the value a such that with probability 0.9 the number of infected people in this organization will be from 150-a to 150+a. The resulting interval is called a *confidence interval*.

Exercise 3 [H]

On average, the number of people infected with the flu virus during an epidemic is 20%, but in a specific case, by polling N residents, one can find out the relative frequency of infections. What is the smallest number N of people that should be polled in order to assert with probability 0.95 that the recorded infection frequency differs from the assumed value of 0.2 by less than 0.05? How can the situation be explained if the observed frequency at such an N turns out to be, for example, equal to 0.1?

Exercise 4

In the production of identical parts, a defect occurs with probability $p = 10^{-4}$. Estimate the probability that in a batch of 500 parts there is more than one defective part.

Exercise 5

In the production of identical parts at a Chinese factory, a defect occurs with probability $p = 3 \cdot 10^{-4}$; in the production of the same parts at an Indian factory, a defect occurs with probability $p = 2 \cdot 10^{-4}$. A shipment from China of 100 parts was combined with a shipment from India of 200 parts. Estimate the probability that in the resulting shipment of 300 parts there is not a single defective one.

Exercise 6

A call center receives calls on average once every five minutes. Estimate the probability that in 10 minutes there will be exactly three calls.

Exercise 7

Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables for which the Law of Large Numbers (LLN) holds:

$$S_n := \sum_{i=1}^{n} \xi_i, \qquad \lim_n \mathbb{P}\left(\left| \frac{S_n - \mathbb{E}S_n}{n} \right| \geq \varepsilon \right) = 0 \qquad \forall \varepsilon > 0.$$

Is it true that the LLN must also hold for the sequence $|\xi_1|, |\xi_2|, \ldots |$?

Exercise 8* [Bernstein's Theorem]

Let random variables $\xi_1, \xi_2, \ldots$ have uniformly bounded variances, and $\mathrm{Cov}(\xi_i, \xi_j) \to 0$ uniformly as $|i - j| \to \infty$. Prove that this sequence satisfies the LLN.

Exercise 9

Show that the independence condition cannot be dropped from the law of large numbers.

Exercise 10

Show that the independence condition cannot be dropped from the central limit theorem.