Probability and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Seminar 9

Inequalities, Laws of Large Numbers Central Limit Theorem

1. Chebyshev's Inequality

Problem 1

Let be $\mathbb{E}(\xi) = 1.6$, $\mathbb{D}(\xi) = 0.1$.
Provide a lower bound for the probability $\mathbb{P}(0.220 < \xi < 1.980)$.

Solution to Problem 1

We apply Chebyshev's inequality.

$$\mathbb{P}(\mathbb{E}(\xi) - \varepsilon < \xi < \mathbb{E}(\xi) + \varepsilon) > 1 - \frac{\mathbb{D}^2(\xi)}{\varepsilon^2}$$

First, we determine the accuracy.

$$\begin{aligned} 1.6 - \varepsilon_1 = 1.220 \qquad &\Longrightarrow \qquad \varepsilon_1 = 0.38, \\ 1.6 + \varepsilon_2 = 1.980 \qquad &\Longrightarrow \qquad \varepsilon_2 = 0.38, \end{aligned}$$

Thus, we obtain $\varepsilon = \min(\varepsilon_1, \varepsilon_2) = 0.38$. The desired lower bound is:

$$1 - \frac{\mathbb{D}^2(\xi)}{\varepsilon^2} = 1 - \frac{0.1^2}{0.38^2} = 0.9370.$$

2. Bernoulli's Law of Large Numbers

Problem 2

At least how many times must a fair coin be tossed so that the relative frequency of heads lies between 0.37 and 0.63 with probability at least 0.7?

Solution to Problem 2

We apply the Law of Large Numbers by Bernoulli. Since the coin is fair, $p = \frac{1}{2}$.

$$\left|\frac{k_n}{n} - p\right| < \varepsilon \quad \Longleftrightarrow \quad p - \varepsilon \leqslant \frac{k_n}{n} \leqslant p + \varepsilon,$$

from which we obtain that

$$\frac{1}{2} - \varepsilon_1 = 0.37 \qquad \Longrightarrow \qquad \varepsilon_1 = 0.13,$$
$$\frac{1}{2} + \varepsilon_2 = 0.63 \qquad \Longrightarrow \qquad \varepsilon_2 = 0.13,$$

thus we get that $\varepsilon = \min(\varepsilon_1, \varepsilon_2) = 0.13$. We need to solve the following inequality.

$$1 - \frac{p(1-p)}{n\varepsilon^2} = 1 - \frac{1}{4n \cdot 0.13^2} \geqslant 0.77$$

$n \geqslant \dfrac{1}{(1-0.77) \cdot 4 \cdot 0.13^2} = 64.3$, that is $n \geqslant 64$.

Problem 3

What is the minimum sample size we need to take if, under sampling with replacement, we want to estimate the defect rate with an accuracy of 0.11 (at most this deviation) and a confidence level of 0.83?

Solution to Problem 3

We apply the Law of Large Numbers by Bernoulli. $\varepsilon = 0.11$. The reliability is:

$$1 - \frac{1}{4n\varepsilon^2} \geqslant 0.83 \qquad \Longrightarrow \qquad n \geqslant \frac{1}{(1-0.83)\cdot 4 \cdot 0.11^2} = 121.5.$$

Thus we obtain $n \geqslant 122$.

3. De Moivre–Laplace Theorem

Problem 4

An urn contains 123 balls, of which 103 are red and 20 are white. We draw 600 balls with replacement. Provide an approximation of the probability that the number of red balls drawn falls within the interval $[478, 526]$.

Solution to Problem 4

Let ξ denote the number of red balls drawn. Then $\xi \sim B\left(n = 600, p = \frac{103}{123}\right)$. The probability we want to find

$$\mathbb{P}(\xi \in [478, 526]) = \sum_{k=478}^{526} \binom{600}{k} \left(\frac{103}{123}\right)^{k} \left(\frac{20}{123}\right)^{600-k}$$

in principle. However, this sum cannot be easily computed because it contains too many terms and involves multiplying very large numbers by very small numbers. Instead, we apply the De Moivre–Laplace theorem.

$$\mathbb{P}(478 \leqslant \xi \leqslant 526) \approx \Phi\left(\frac{526 - 600 \cdot \frac{103}{123} + \frac{1}{2}}{\sqrt{600 \cdot \frac{103}{123} \cdot \frac{20}{123}}}\right) - \Phi\left(\frac{478 - 600 \cdot \frac{103}{123} - \frac{1}{2}}{\sqrt{600 \cdot \frac{103}{123} \cdot \frac{20}{123}}}\right) =$$

$$= \Phi(2.6620) + \Phi(2.7592) - 1 = 0.996 + 0.9971 - 1 = 0.9932.$$

Problem 5

A target is shot at 200 times. Each shot hits with probability 0.4. Determine the range within which the number of hits will fall with 90% probability. Solve the problem using both the Law of Large Numbers and the De Moivre–Laplace Theorem.

Solution to Probelm 5

We solve the problem in two ways.

- **Using the Bernoulli Law of Large Numbers**, we obtain that

$$\mathbb{P}\left(\left|\frac{k_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant 1 - \frac{p(1-p)}{\varepsilon^2 n}.$$

Since $p = 0.4$ and $n = 200$, we have that

$$1 - \frac{0.4 \cdot 0.6}{\varepsilon^2 200} = 0.9$$

from the inequality we obtain that

$$\sqrt{\frac{0.4 \cdot 0.6}{(1-0.9) \cdot 200}} = 0.1095.$$

We have that $-\varepsilon \leqslant \frac{k_n}{n} - p \leqslant \varepsilon$, whence we obtain that

$$\underbrace{(-\varepsilon + p) \cdot n}_{58} \leqslant k_n \leqslant \underbrace{(\varepsilon + p) \cdot n}_{102}$$ it will fall within the interval.

- Using the **De Moivre–Laplace theorem**:
  We know that number of hits denoted by k_200 is a random variable distributed $\mathcal{B}(n = 200, p = 0.4)$ so
  $\mathbb{E}(k_{200}) = np = 200 \cdot 0.4 = 80$,
  $\mathbb{D}(k_n) = \sqrt{np(1-p)} = \sqrt{200 \cdot 0.4 \cdot 0.6} = \sqrt{48}$. We consider a symmetric interval. Let $a = 80 - r$, $b = 80 + r$.
  By the De Moivre–Laplace theorem we obtain that

$$\mathbb{P}(a \leqslant k_{200} \leqslant b) \sim \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right) =$$

$$= \Phi\left(\frac{r}{\sqrt{48}}\right) - \Phi\left(\frac{-r}{\sqrt{48}}\right) = 2\Phi\left(\frac{r}{\sqrt{48}}\right) - 1 > 0.9.$$

  Thus, we obtain $\Phi\left(\frac{r}{\sqrt{48}}\right) > \frac{0.9+1}{2} = 0.95$. Looking up the value of $\frac{r}{\sqrt{48}}$ we have that $\frac{r}{\sqrt{48}} \geqslant 1.645$ whence $r \geqslant 11.4$.
  Therefore, $a = 80 - r = 68$, $b = 80 + r = 92$, that is,
  $68 < k_n < 92$.

4. Central Limit Theorem

Problem 6

In certain stores, the number of customers appearing between 11 a.m. and 12 p.m. is a Poisson-distributed random variable with expected value $\lambda = 30$. Considering 100 such stores, what is the probability that the total number of customers appearing between 11 a.m. and 12 p.m. falls between 3000 and 3100?

Solution to Problem 6

Let $\xi_1, \xi_2, \ldots, \xi_{100}$ be independent Poisson-distributed random variables with expected value $\lambda = 30$. Let $S_{100} = \xi_1 + \xi_2 + \cdots + \xi_{100}$. Then the common mean is $m = 30$, and the common variance is $\sigma^2 = 30$.
Based on the Central Limit Theorem, we obtain that $\eta := \frac{S_n - nm}{\sqrt{n}\sigma} \sim \mathcal{N}(0,1)$, whence we obtain that

$$\mathbb{P}\left(3000 \leqslant \sum_{k=1}^{100} \xi_k \leqslant 3100\right) =$$
$$= \mathbb{P}\left(\frac{3000 \cdot -100 \cdot 30}{\sqrt{100} \cdot \sqrt{30}} \leqslant \eta \leqslant \frac{3100 \cdot -100 \cdot 30}{\sqrt{100} \cdot \sqrt{30}}\right) =$$
$$= \Phi(1.83) - \Phi(0) = 0.964 - 0.5 = 0.464.$$

End of Seminar 9