Probability and Mathematical Statistics

Miskolc, 2025.

Dr. Tamás Glavosits

Lecture 9

Inequalities, Laws of Large Numbers, Central Limit Theorem, Quantiles

1. Types of Convergence

Almost Surly or Strongly Convergence

Let (ξ_n) be a sequence of random variables and ξ be a random variable.

Definition

We say that ξ_n **converges to** ξ **almost surely** or **almost everywhere** or **strongly** if

$$\mathbb{P}\left(\lim_{n\to\infty} \xi_n = \xi\right) = \mathbb{P}\left(\left\{\omega \in \Omega | \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right\}\right) = 1$$

denoted by $\xi_n \stackrel{\text{a.s.}}{\to} \xi$

Convergence in probability

Definition

We say that ξ_n **converges to** ξ **in probability** if for every $\varepsilon > 0$

$$\mathbb{P}\left(\lim_{n\to\infty} |\xi_n - \xi| > 0\right) = \lim_{n\to\infty} \mathbb{P}\left(\{\omega \in \Omega \| \xi_n(\omega) - \xi(\omega)| < \varepsilon\}\right) = 1$$

denoted by $\xi_n \xrightarrow{\text{p}} \xi$

Weak Convergence

Definition

Let $\mathbb{F}_n$ denote the distribution function of ξ_n for every $n \in \mathbb{Z}_+$, and let $\mathbb{F}$ denote the distribution function of ξ.

We say that ξ_n converges **weakly** or **convergens in ditribution** to ξ if $\mathbb{F}_n(x) \to \mathbb{F}(x)$ at every point x where $\mathbb{F}$ is continuous, denoted by $\xi_n \xrightarrow{\mathcal{D}} \xi$.

Relationship between types of convergence

Remark

The almost surely convergence implies the convergence in probability.

2. Inequalities

Markov Inequality

Theorem

Let η be a nonnegative random variable such that there exists $\mathbb{E}(\eta)$, and let $c > 0$. Then

$$\mathbb{P}(\eta < c) \geqslant 1 - \frac{\mathbb{E}(\eta)}{c}.$$

Proof of Markov Inequality

Proof.

Since $\mathbb{P}(\eta < c) = 1 - \mathbb{P}(\eta \geqslant c)$, it is sufficient to prove that $\mathbb{P}(\eta \geqslant c) \leqslant \dfrac{\mathbb{E}(\eta)}{c}$.

a. In discrete cas:

$$\mathbb{E}(\eta) = \sum_i x_i \mathbb{P}(\xi = x_i) \geqslant \sum_{x_i \geqslant c} x_i \mathbb{P}(\eta = x_i) \geqslant \sum_{x_i \geqslant c} c\mathbb{P}(\eta = x_i) =$$

$$= c \sum_{x_i \geqslant c} \mathbb{P}(\eta = x_i) = c\mathbb{P}(\eta \geqslant c).$$

b. In absolutely continuous case:

$$\mathbb{E}(\eta) = \int_0^{+\infty} x f_\xi(x) dx \geqslant \int_c^{+\infty} x f_\xi(x) dx \geqslant \int_c^{+\infty} c f_\xi(x) dx =$$

$$= c \int_c^{+\infty} f_\xi(x) dx = c\mathbb{P}(\eta \geqslant c).$$

Chebyshev Inequality

Theorem

Let ξ *be a random variable such that there exists* $\mathbb{D}^2(\xi)$*, and let* $\varepsilon > 0$*. Then*

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| < \varepsilon) \geqslant 1 - \frac{\mathbb{D}^2(\xi)}{\varepsilon^2}.$$

Proof of Chebyshev Inequality

Proof.

This is a consequence of the Markov inequality. Let $\eta = (\xi - \mathbb{E}(\xi))^2$ and $c = \varepsilon^2$. Then η is a nonnegative random variable with an existing expected value. Since

$$(\xi - \mathbb{E}(\xi))^2 < c = \varepsilon^2 \quad \Longleftrightarrow \quad |\xi - \mathbb{E}(\xi)| < \varepsilon,$$

applying the Markov inequality, we obtain

$$\mathbb{P}\left(|\xi - \mathbb{E}(\xi)| < \varepsilon\right) \geqslant 1 - \frac{\mathbb{E}(\xi - \mathbb{E}(\xi))^2}{\varepsilon^2} = 1 - \frac{\mathbb{D}^2(\xi)}{\varepsilon^2}.$$

$\square$

Corollary

Let ξ be a random variable such that there exists $\mathbb{D}^2(\xi)$. Then the probability that ξ falls within a neighborhood of $\mathbb{E}(\xi)$ radius $k\mathbb{D}(\xi)$ is at least $1 - \frac{1}{k^2}$, that is,

$$\mathbb{P}\left(|\xi - \mathbb{E}(\xi)| < k\mathbb{D}(\xi)\right) \geqslant 1 - \frac{1}{k^2}.$$

Proof.

This follows from the Chebyshev inequality by choosing $\varepsilon = k\mathbb{D}(\xi)$. For example: the 3σ rule ($\sigma = \mathbb{D}(\xi)$):

$$\mathbb{P}\left(|\xi - \mathbb{E}(\xi)| < 3\sigma\right) \geqslant 1 - \frac{1}{9} = \frac{8}{9}.$$

$\square$

3. Laws of Large Numbers

Weak Laws of Large Numbers

The weak laws of large numbers assert **convergence in probability**.

Theorem (Chebyshev's Law of Large Numbers)

Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with an existing variance, m denotes their common expected value, and $\varepsilon > 0$. Then

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\xi_1 + \cdots + \xi_n}{n} - m\right| < \varepsilon\right) = 1,$$

(that is, the expected value can be approximated by the sample mean).

Proof of Weak Laws of Large Numbers

Proof.

$$\mathbb{E}\left(\frac{\xi_1+\cdots+\xi_n}{n}\right)=m,$$

$$\mathbb{D}^2\left(\frac{\xi_1+\cdots+\xi_n}{n}\right)=\frac{1}{n^2}n\mathbb{D}^2(\xi_1)=\frac{\mathbb{D}^2(\xi_1)}{n}.$$

Thus, from Chebyshev's inequality, we obtain that

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{\xi_1+\cdots+\xi_n}{n}-m\right|<\varepsilon\right)\geqslant\lim_{n\to\infty}\left(1-\frac{\frac{\mathbb{D}^2(\xi_1)}{n}}{\varepsilon^2}\right)=1.$$

$\square$

Bernoulli's Law of Large Numbers

Theorem

Consider a sequence of independent trials and let us observe the number of occurrences of an event with probability p. Let $\frac{k_n}{n}$ denote the relative frequency of the event, i.e., the event A occurs k_n times in n trials. Then

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

Clearly, $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{k_n}{n} - p\right| \leqslant \varepsilon\right) = 1$. *(That is, the relative frequency of the event approaches the theoretical probability in the limit.) Hence, the relative frequencies converge in probability to the probability p.*

Proof of Bernoulli's Law of Large Numbers

Proof.

$k_n \sim \mathrm{B}(n, p)$. Thus,

$$\mathbb{E}\left(\frac{k_n}{n}\right) = \frac{1}{n}\mathbb{E}(k_n) = \frac{1}{n}np = p,$$
$$\mathbb{D}^2\left(\frac{k_n}{n}\right) = \frac{1}{n^2}\mathbb{D}^2(k_n) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n},$$

Applying the Chebyshev inequality, we obtain

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

$\square$

Proof of Bernoulli's Law of Large Numbers, Continued

Proof.

If the probability p is unknown, then

$$\mathbb{P}\left(\left|\frac{k_n}{n} - p\right| \leqslant \varepsilon\right) \geqslant 1 - \frac{1}{4n\varepsilon^2},$$

since with completing the square we obtain that

$$-p(1-p) = p^2 - p = \left(p - \frac{1}{2}\right)^2 - \frac{1}{4} \geqslant -\frac{1}{4}.$$

$\square$

Accuracy and Reliability

Definition

The number ε is called the **accuracy**, and the right-hand side is called the **reliability**.

Jacob Bernoulli (1654–1705), Swiss mathematician.

Strong Laws of Large Numbers

The strong laws of large numbers assert **almost sure convergence** or convergence with **probability 1**.
This theorem is due to Nasrollah Etemadi, professor emeritus at the University of Chicago, who is still living.

Theorem

Let $\xi_1, \xi_2, \ldots$ be identically distributed, pairwise independent random variables with $m = \mathbb{E}(\xi_i) < +\infty$. Then $\frac{\xi_1 + \cdots + \xi_n}{n}$ converges to m with probability 1, that is

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\xi_1 + \cdots + \xi_n}{n} = m\right) = \mathbb{P}\left(\omega \in \Omega | \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \xi_i(\omega)}{n} = m\right) = 1$$

4. Central Limit Theorem

Central Limit Theorem

The central limit theorem expresses **convergence in distribution**. It plays a central role in probability theory and mathematical statistics.

Theorem

Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with finite variance ($m = \mathbb{E}(\xi_i)$, $\sigma = \mathbb{D}(\xi_i)$), and let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$.

$$\lim_{n \to \infty} \left(\frac{S_n - nm}{\sigma \sqrt{n}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \, (= \varphi(x)) \quad (x \in \mathbb{R})$$

History of the Theorem

The history of the theorem:

- The first version is attributed to the French mathematician Abraham de Moivre (1733)
- Pierre-Simon Laplace (1812) rediscovered the theory
- Aleksandr Mikhailovich Lyapunov (1901) was the first to rigorously prove the theorem

Interestingly, a proof of the theorem also appears in Alan Turing's doctoral dissertation. The name "central limit theorem" comes from a 1920 paper by György Pólya.

5. Moivre-Laplace Theorem

Moivre-Laplace Theorem

Since the binomial distribution is the sum of independent Bernoulli distributions, the central limit theorem can be applied.

Theorem (Moivre-Laplace Theorem)

Let $\xi \sim \mathrm{B}(n, p)$, $0 \leqslant a < b \leqslant n$, *then*

$$\mathbb{P}\left(a \leqslant \xi \leqslant b\right) \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{np(1-p)}}\right).$$

End of Lecture 9