Lecture 11 - Limit Theorems

Introduction

- By doing experiments we observe that the frequency of some event occurring tends to the probability of that event as the number of trials tends to infinity.
- The law of large numbers was formulated by Jacob Bernoulli as:

> 💡 "With sufficiently large number of trials, the frequency of a given random event will be arbitrarily close to the probability of this event."

- We often think about the law of large numbers as a way of defining the probabilities.

The frequency of events and Bernoulli scheme

- The occurrence/non-occurrence of a given event can be encoded by a Bernoulli random variable.
- Let n be the number of trials in which we observe if a given event A happened or not
- Define random variables $X_1, \ldots, X_n$ which concern consecutive trials as:

$$X_i(\omega) = \begin{cases} 1 & \text{if } \omega \in A \quad \text{"A occured"} \\ 0 & \text{if } \omega \notin A \quad \text{"A did not occur"} \end{cases}$$

- Those variables are independent and all have the Bernoulli distribution $B(p)$ with parameter $p = P(X_i = 1) = P(A)$
- We can define "frequency of the event A" as the number of successes: $S_n = X_1 + \ldots + X_n$ divided by the number of trials n: $\text{frequency} = \frac{S_n}{n}$.

- The law of large numbers states that as $n \to \infty$, $\frac{S_n}{n} \to p$, where p is the probability of the event A.

Bernoulli scheme

- Let $X_1, \ldots, X_n$ be independent random variables with the distribution $B(p)$.
- We know that the number of successes $S_n = X_1 + \ldots + X_n$ has the binomial distribution $B(n, p)$:

$$ES_n = np \quad D^2(S_n) = np(1-p)$$

- The distribution "grows", or stretches horizontally as $n \to \infty$.
- Let $\overline{X}_n = \frac{S_n}{n}$, the frequency of successes, then:

$$E\overline{X}_n = E\left(\frac{S_n}{n}\right) = \frac{ES_n}{n} = p$$
$$D^2(\overline{X}_n) = D^2\left(\frac{S_n}{n}\right) = \frac{D^2(S_n)}{n^2} = \frac{p(1-p)}{n}$$
$$D(\overline{X}_n) = \sqrt{\frac{p(1-p)}{n}}$$

- The distribution "shrinks" as $n \to \infty$.

Law of Large Numbers for Bernoulli distribution

- Remainder: Chebyshev's inequality:
  - For a random variable X with finite expectation and variance:

$$\forall \epsilon > 0 \quad P(|X - EX| \geq \epsilon) \leq \frac{D^2(X)}{\epsilon^2}$$

- If we apply this inequality to $\overline{X}_n$:

$$P(|\overline{X}_n - p| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$$

- By negating and adding one:

$$P(|\overline{X}_n - p| \leq \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2}$$

- And from this we arrive at the **Bernoulli's Law of Large Numbers**:
  - For any $\epsilon > 0$:

$$\lim_{n \to \infty} P(|\overline{X}_n - p| \leq \epsilon) = 1$$

  - Random variable $\overline{X}_n$ converges "in probability" to p.

Generalizing the Law of Large Numbers

- We have shown that the frequency of any given random event converges to its probability.
- We used the fact that with large probability the arithmetic mean of n Bernoulli random variables ("the frequency of successes") is close to its expected value (probability of success), but this property also holds for other random variables.
- We will generalize the LLN to:

> 💡 "For sufficiently large n, the arithmetic mean of n realizations of a random variable is arbitrarily close to its expectation"

- For example:
  - For large n, the mean of n rolls of a die is close to 3.5
  - For large n, the average win in n independent games is close to the expected value of win

- For large n, the average of n numbers drawn (independently) from a uniform distribution on $[0, 1]$ is close to 0.5

Sequence of independent random variables

- $X_1, \ldots, X_n$ — independent random variables with the same distribution, so also with the same expected value $EX_i = \mu$ and variance $D^2(X_i) = \sigma^2$.
- Define: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then:

$$E\overline{X}_n = \mu$$
$$D^2(\overline{X}_n) = \frac{\sigma^2}{n}$$

- Applying the Chebyshev's inequality:

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{D^2(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

- So as $n \to \infty$:

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq 0$$

Law of Large Numbers

Weak Law of Large Numbers (Khinchin)

- Let $X_1, \ldots$ be a sequence of independent random variables with the same distribution, expected value μ and variance $\sigma^2 < \infty$. Then for any $\epsilon > 0$:

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \leq \epsilon) = 1$$

- The Bernoulli's LLN is a special case of this theorem for random variables having Bernoulli distribution.

Convergence in probability

- Let $X_n = (X_1, X_2, \ldots)$ be a sequence of random variables.
- We say that sequence X_n converges to X in probability, if for any $\epsilon > 0$:

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

- We denote it as $X_n \xrightarrow{P} X$.
- In particular if X has a degenerate distribution, i.e. $P(X = c) = 1$, we denote $X_n \xrightarrow{P} c$.
- We can use this to rewrite the law of large numbers:

Weak Law of Large Numbers (Khinchin) (again)

- Let $X_1, \ldots$ be a sequence of independent random variables with the same distribution, expected value μ and variance $\sigma^2 < \infty$. Then:

$$\overline{X}_n \xrightarrow{P} \mu$$

Weak Law of Large Numbers (Chebyshev)

- Let $X_1, X_2, \ldots$ be a sequence of independent random variables with expected values $EX_i = \mu_i$ and variances $D^2(X_i) = \sigma_i^2$, jointly bounded by σ^2 (i.e., $\sigma_i^2 \leq \sigma^2$ for all i). Let $\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$. Then:

$$\overline{X}_n - \overline{\mu}_n \xrightarrow{P} 0$$

Convergence “with probability 1”

- Consider an event:

$$\lim_{n\to\infty} X_n = X$$

- This is a random event, since it is a statement about random variables.
- Consider an event:

$$\forall \omega \in \Omega \quad \lim_{n\to\infty} X_n(\omega) = X(\omega)$$

- This statement is called “sure” convergence: the convergence holds for all $\omega \in \Omega$.
- Usually it is too strong, since it must hold even for ω, for which $P(\{\omega\}) = 0$
- We say that a sequence X_n converges to a random variable X with probability 1 ("almost surely") if:

$$P(\lim_{n\to\infty} X_n = X) = 1$$

- We denote it by: $X_n \overset{\text{w. pr. 1}}{\to} X$
- Convergence with probability 1 implies convergence in probability (it is stronger):

$$X_n \overset{\text{w. pr. 1}}{\to} X \implies X_n \overset{P}{\to} X$$

Strong Law of Large Numbers

- We know that (from the proofs during lectures):

$$X_n \overset{P}{\to} X \not\Longrightarrow X_n \overset{\text{w. pr. 1}}{\to} X$$

- Despite that, the LLN can be strengthen to $X_n \overset{\text{w. pr. 1}}{\to} X$

Strong Law of Large Numbers (Khinchin)

- Let $X_1, X_2, \ldots$ be a sequence of independent random variables with the same distribution, expected value μ and variance $\sigma^2 < \infty$. Then:

$$\overline{X}_n \overset{\text{w. pr. 1}}{\to} \mu$$

Strong Law of Large Numbers (Chebyshev)

- Let $X_2, X_2, \ldots$ be a sequence of independent random variables with expected values $EX_i = \mu_i$ and variances $D^2(X_i) = \sigma_i^2$, jointly bounded by σ^2 (i.e., $\sigma_i^2 \leq \sigma^2$ for all i). Let $\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$. Then:

$$\overline{X}_n - \overline{\mu}_n \overset{\text{w. pr. 1}}{\rightarrow} 0$$