Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

October 4, 2025

Week 5

Independence

Definition (Random Variable Independence)

Random variables $X_1, \ldots, X_n$ are independent if

$$\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2, \ldots, X_n \leq t_n) = \mathbb{P}(X_1 \leq t_1) \cdot \mathbb{P}(X_2 \leq t_2) \cdot \ldots \cdot \mathbb{P}(X_n \leq t_n)$$

holds for all real numbers $t_1, t_2, \ldots, t_n$.

Note: Functions of independent random variables will also be independent.

Normal distribution features

Let X and Y be independent random variables with normal distribution. Then

a for real numbers $c \neq 0$ and d, the random variable $cX + d$ has normal distribution with mean $c \cdot \mathbb{E}(X) + d$ and variance $c^2 \mathrm{Var}(X)$;

b for every real number t we have $(X - m)/\sigma$ has standard normal distribution, and

$$\mathbb{P}(X \leq t) = \mathbb{P}\left(\frac{X - m}{\sigma} \leq \frac{t - m}{\sigma}\right) = \Phi\left(\frac{t - m}{\sigma}\right).$$

c $X + Y$ has normal distribution, and $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, and $\mathrm{Var}(X + Y) = \mathrm{Var}(\mathrm{X}) + \mathrm{Var}(Y)$.

d If $X_1, X_2, \ldots, X_n$ are independent normal variables with mean m and variance σ^2, then their mean also has a normal distribution:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

Normal distribution features

Examples (1)

We install a program which consists of $n = 68$ components. The time of the installation of each file has mean $m = 10$ and $\sigma = 2$ (in seconds), and these are independent of each other.

a. Calculate the probability that the sum of the installation time is at most 12 minutes, supposing that the installation times have normal distribution.
b. Approximate the same probability without assuming normal distribution.
c. The next version consists of k files, with the same rules as above. We also know that the probability that the total time is less than 10 minutes is 95% percent. Determine the value of k.

Examples (2)

Let X and Y be independent random variables, both with an expected value of 0 and a standard deviation of 1. Let $W = X - Y$. Calculate the expected value and the standard deviation of W.

Law of large numbers

Random variables X, Y are identically distributed if their cumulative distribution function is the same: $P(X \leq t) = P(Y \leq t)$ holds for every t.

Theorem (Law of large numbers)

Let $X_1, X_2, \ldots$ be independent identically distributed random variables. Suppose furthermore that $m = \mathbb{E}(X_1) < \infty$. Then

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1) = m$$

holds with probability 1 as $n \to \infty$.

Central Limit Theorem

Theorem (Central limit theorem)

Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, for which $\mathbb{E}(X_1) = m$ and s.d.$(X_1) = \sigma < \infty$, that is, their variance is finite. Then for every real number t we have

$$P\left(\frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} \leq t\right) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty),$$

where Z has standard normal distribution, that is,

$$P(Z \leq t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Law of Large Numbers and CLT examples

Examples (1)

Suppose that $X_1, X_2, \dots,$ are independent random variables with Poisson distribution, and the variance of each X_j is equal to 3.

a Find the almost sure limit of $\frac{X_1+X_2+\ldots+X_n}{n}$.

b Find the limit of $\mathbb{P}\left(\frac{X_1+\ldots+X_n-3\cdot n}{\sqrt{n}} \leq 1\right)$ as n tends to infinity.

Examples (2)

Suppose that $X_1, X_2, \dots,$ are independent random variables with exponential distribution, and we also know that $\mathbb{P}(X_j \geq 1) = e^{-3}$.

a Find the almost sure limit of $\frac{X_1+X_2+\ldots+X_n}{n}$.

b Find the limit of $\mathbb{P}\left(\frac{X_1+\ldots+X_n-\frac{1}{3}\cdot n}{\sqrt{n}} \leq 1\right)$ as n tends to infinity.

Inequalities

Theorem (Markov-inequality)

Let $g : \mathbb{R} \to \mathbb{R}$ monotonically increasing, positive function, $X \geq 0$ a random variable, for which $EX < \infty$ and $\varepsilon > 0$ Then $P(X \geq \varepsilon) \leq \dfrac{E\left(g(X)\right)}{g(\varepsilon)}$. Especially, if $g(x) = x$, then $P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon}$

Theorem (Chebyshev-inequality)

Let X be an arbitrary random variable, for which $Var(X) < \infty$ and $\varepsilon > 0$. Then $P(|X - EX| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$.

Examples

Let X be a positive random variable with expected value $E(X) = 3$ and standard deviation $D(X) = 3$. Calculate the maximum probability that the variable takes a value of 13 or greater! What is the exact value of the probability, if we assume that the distribution is exponential?

Examples

1. Let $X \sim N(2, (\sqrt{5})^2)$ and $Y \sim N(5, 3^2)$ be independent random variables, and let $W = 3X - 2Y + 1$. Calculate: a) $E[W]$ and $D^2 W$ (Variance of W), and b) $P(W \leq 6)$. Given table value: $\Phi(1) = 0.8413$, where Φ is the Cumulative Distribution Function of the standard normal distribution $N(0, 1)$.
2. A scanned image has an average size of 600 KB with a standard deviation of 100 KB. What is the probability that 80 such images together occupy between 47 MB and 48 MB of storage space, assuming the image sizes are independent?
3. Suppose the mass of a bar of chocolate is normally distributed with an expected value of 100g and a standard deviation of 3g. What is the minimum number of chocolate bars we should pack into a box so that the average mass of the bars in the box is greater than 99.5g with a probability of at least 0.9, assuming the mass of each bar is independent?
4. An electric wire manufacturing company produces 40 m wires with a standard deviation of 0.2 m. What is the maximum probability that the length of the wire differs from the expected value of 40 m by at least 1 m?