Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

September 27, 2025

Week 4

Recap from last week

1. We organize a party, and we know in advance that the number of participants is eight with probability $1/4$, nine with probability $1/3$ and ten otherwise. Calculate the expected number and the variance of the number of participants.
2. Flipping a coin (let p be the probability of heads), let X denote the length of the first sequence of identical outcomes. (For example, if the sequence is HHT..., then $X = 2$.) Determine the distribution of X.
3. Suppose that 3 probability practice sessions have 15, 20, and 25 students, respectively. What is the expected size of a randomly selected student's group?
4. Suppose that we have 10 servers in a system. On each day, each of them breaks down with probability 0.01, independently of each other. Let Z be the number of servers (among these 10) which break down tomorrow (that is, on a given day). Calculate the probability $\mathbb{P}(Z = 2)$, and the expectation and variance of Z.
5. Suppose that the number of downloads of a webpage within an hour has Poisson distribution, and the probability that there are 0 downloads is $1/e^2$. Suppose furthermore that the number of downloads independent for disjoint time intervals. a) What is the variance of the number of downloads within an hour? b) Given that the number of downloads within an hour is at most 1, what is the probability that there are 0 downloads within this hour?

Absolute Continuous distribution

Definition (Absolute Continuous Random Variable)

A random variable X is absolutely continuous if there exists a function $f(x)$ such that the cumulative distribution function (CDF) $F(x)$ can be expressed as:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

In this case, $f(x)$ is called the **probability density function** (PDF).

Note: For the PDF f to exist, it is necessary (but not sufficient) that the CDF F is continuous (i.e., $P(X = x) = 0$ for all x).

Absolute Continuous distribution

Theorem

Let X have an absolutely continuous distribution. Then:

- $f(x) = F'(x)$
- $f(x) \geq 0$ *for all* x
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $P(X = x) = 0$ *for all* x
- $P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a)$

Absolute Continuous distribution

Definition (Expected Value)

Let X be an absolutely continuous random variable with probability density function $f(x)$. The expected value of X is:

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$$

This integral is said to exist if it is absolutely convergent.

Definition (Standard Deviation)

The standard deviation of X is given by

$$s.d.(X) = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}.$$

where

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx.$$

Examples

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

1. What is the value of C?
2. Find $P\{X > 1\}$.
3. Determine the expectation of X.
4. Determine the standard deviation of X.

Uniform distribution

Definition (Uniform distribution)

CDF: $F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x \leq b \\ 1 & \text{if } b < x \end{cases}$, while PDF: $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x \leq b \\ 0 & \text{otherwise} \end{cases}$

Examples

Assume that a computer's time of failure is between 0 and 10 years and can be described by a uniform distribution. Determine the distribution function of the phenomenon.

1. Calculate $\mathbb{P}(X \leq 9)$, that is, the probability that the delivery is before 9.
2. Calculate the probability $\mathbb{P}(8.5 < X < 9)$.
3. Calculate the probability $\mathbb{P}(X > 9.75)$.
4. Draw the curve $\mathbb{P}(X \leq t)$ (the cumulative distribution function of X) as a function of t, where $t \in \mathbb{R}$ is a real number.

Exponential distribution

Definition

A random variable X has exponential distribution with parameter λ is its density function is given by
$f(x) = \begin{cases} 0, & x \leq 0; \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$, while CDF: $F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

Examples

Suppose that the response time of a server (in seconds) has exponential distribution with parameter $\lambda = 2$.

1. Determine the cumulative distribution function and the density function of the response time.
2. What is the probability that the response time is more than 0.5 seconds?
3. the response time is at least 1 second?
4. given that the response time is at least 0.5 second. what is probability that it is at least 2.5 seconds?
5. What is the probability that the response time is between 1 and 2 seconds?
6. For which t is it true that the probability that the response time is at most t is equal to $1/2$?
7. Find the expectation and standard deviation of X.
8. Generate a sample of 1000 random variables from the exponential distribution with parameter $\lambda = 2$. Make a histogram, and find the mean, standard deviation and median of the sample. Compare the results with the values calculated above.

Normal Distribution

Definition

A random variable has normal distribution with mean m and variance σ^2 if its density function is defined by $f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$. This is denoted by $X \sim N(m, \sigma^2)$

A random variable Z has standard normal distribution if it has normal distribution with mean $m = 0$ and variance $\sigma = 1$, that is, $Z \sim N(0, 1)$. Its density function is given by $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$.

That is, for every $a < b$ we have

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\, dx = \int_a^b \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

The cumulative distribution function of the standard normal distribution is denoted by Φ:

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

We can use that $\Phi(-x) = 1 - \Phi(x)$, *and* $\Phi(0.5) = 0.6915$, $\Phi(1) = 0.8413$, $\Phi(2) = 0.9772$, $\Phi(3) = 0.9987$.

Examples

Suppose that the temperature in Budapest, on October 31 at midnight has normal distribution X with mean 1 and variance 4 (that is, $X \sim N(1, 4)$), in Celsius degree).

1. What is the probability that the temperature at midnight on the given day is below 0° C?
2. What is the probability that the temperature is between $-1^\circ C$ and $3^\circ C$?
3. Generate a random independent sample of size 1000 from the normal distribution with mean 1 and variance 4. Make a histogram, calculate the mean and the standard deviation, and determine the proportion of the elements between -1 and 3.

Notable Absolutely Continuous Distributions

Name (parameters)	Range	CDF (F)	PDF (f)	EX	D^2X
Uniform $U[a,b]$	$[a,b]$	$\begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x \leq b \\ 1 & \text{if } b < x \end{cases}$	$\begin{cases} \frac{1}{b-a} & \text{if } a < x \leq b \\ 0 & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential $\mathrm{Exp}(\lambda)$	$(0,\infty)$	$\begin{cases} 1-e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Standard Normal $N(0,1)$	$(-\infty,\infty)$ $x \in \mathbb{R}$	$\Phi(x)$ (from tables)	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	0	1
Normal $N(\mu,\sigma^2)$	$(-\infty,\infty)$ $x \in \mathbb{R}$	reducible to $\Phi(x)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Gamma $\Gamma(\alpha,\lambda)$	$(0,\infty)$	no closed-form elementary formula	$\begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

Examples

1. We have ordered an item online, and we know that the delivery is between 8 : 00 and 10 : 00, at a time which is uniformly distributed on the interval $[8, 10]$: the probability that it is between a and b is proportional to $b - a$, if $8 \leq a \leq b \leq 10$. Let X be the time of the delivery (it is a random element of $[8, 10]$, hence this is a random variable).
   a Calculate $\mathbb{P}(X \leq 9)$, that is, the probability that the delivery is before 9.
   b Calculate the probability $\mathbb{P}(8.5 < X < 9)$, that is, the probability is between 8 : 30 and 9 : 00.
   c Calculate the probability $\mathbb{P}(X > 9.75)$.
   d Draw the curve $\mathbb{P}(X \leq t)$ (the cumulative distribution function of X) as a function of t, where $t \in \mathbb{R}$ is a real number.
   e Does there exist a function $f : \mathbb{R} \to \mathbb{R}_+$, such that $\mathbb{P}(a < X < b) = \int_a^b f(x)\, dx$ is satisfied for all real numbers $a < b$?
2. Suppose that the volume of mineral water in a bottle has normal distribution, with standard deviation $0,01$ (in litres). What is the mean of the distribution if the probability that the volume is less than $0,5$ litres is 2%?
3. Let Y be a random variable where $0 < Y < 3$. Its cumulative distribution function on this interval is $F(x) = cx^3$. What are the values of c and $P(-1 < Y < 1)$?
4. Let X be an absolutely continuous random variable on the interval $[0, c]$, with the probability density function:

$$f(x) = \begin{cases} \frac{1}{9}x^2, & \text{if } 0 \leq x < c \\ 0, & \text{if } x < 0 \text{ or } x \geq c \end{cases}$$

   Determine c and the cumulative distribution function of X.

Examples

1. Let the probability density function (PDF) of X be $f(x) = \frac{c}{x^4}$ if $x > 1$, and 0 otherwise.
   a What is the value of c?
   b What is the Expected Value $E(X)$?

2. Experience shows that the distance a certain type of scooter travels until its first breakdown is an exponentially distributed random variable. This distance is on average 6000 km. What is the probability that a randomly selected scooter:
   (a) breaks down after traveling less than 4000 km?
   (b) breaks down after traveling more than 6500 km?
   (c) breaks down after traveling more than 4000 km but less than 6000 km?
   (d) What is the maximum distance traveled until the first breakdown by the earliest 20% of scooters to fail?

3. Let X be the daily milk yield (in liters), modeled by a Normal distribution: $X \sim N(\mu, \sigma^2)$. Mean (Expected Value): $\mu = 22.1$ liters, Standard Deviation: $\sigma = 1.5$ liters.
   a Probability that the yield is between 23 and 25 liters?
   b robability that the yield is between $\mu - \sigma$ and $\mu + \sigma$

   Given Standard Normal CDF values: $\Phi(0.6) = 0.7257$, $\Phi(1.93) = 0.9732$, $\Phi(1) = 0.8413$.

4. How long should we set the warranty period (W) if we want at most 10% of our products to require repair within the warranty period, given that the lifetime of the device (X) can be approximated by a Normal distribution with an expected value (mean) of $\mu = 10$ years and a standard deviation of $\sigma = 2$ years.

Examples

1. Let X be the Intelligence Quotient (IQ), modeled by a Normal distribution: $X \sim N(\mu, \sigma^2)$.
   - Mean (Expected Value): $\mu = 110$
   - Standard Deviation: $\sigma = 10$

   We need to find the probability $P(X > 120)$. Given Standard Normal CDF value: $\Phi(1) = 0.8413$.
2. Let X be uniformly distributed on the interval $[1, 4]$. Calculate the expected value of $(X - 1)^2$, denoted as $E[(X - 1)^2]$.