Lecture 4

CONTINUOUS RANDOM VARIABLES PROBABILITY DISTRIBUTION

Education

FPT UNIVERSITY

Department of Mathematics

Võ Văn Nam

Contents

1. Continuous random variables

Continuous random variables

A **continuous random variable** is a random variable whose possible values includes in an interval of real numbers.

Example.

1. The height of a student at FPT university can be any number between 150cm 190cm.
2. The weight of a newborn can be any number between 0.5kg - 4.5kg.

FPT Education
FPT UNIVERSITY

Probability density function (pdf)

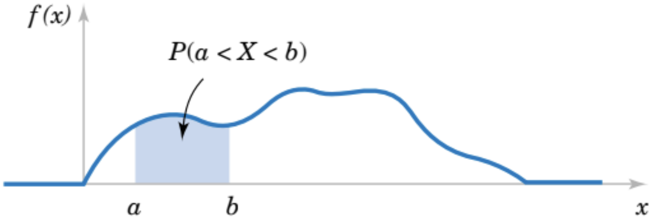
The **probability density function** (pdf) of a continuous random variable X is a function f such that:

- *Non-negativity* $f(x) \geq 0$ for all values of x.
- *Normalization*
$$\int_{-\infty}^{+\infty} f(x)\, dx = 1.$$
- *Probability*
$$P(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

Properties

If X is a continuous random variable, for any a and b,

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b).$$



For a continuous random variable, $P(X = c) = 0$ for any constant c.

Examples

Example 1. Suppose that the probability density function of a continuous random variable X is:

$$f(x) = e^{-(x-3)}, \quad \text{for } x \geq 3.$$

Determine

$$P(1 \leq X < 5),\ P(X < 8),\ P(X \geq 0).$$

Example 2. The probability density function of the length of a metal rod is

$$f(x) = cx^2,\ \text{for } 2 \leq x < 3.$$

a. What is the value of c?

b. Find

$$P(X < 2.5 \text{ or } X \geq 2.8).$$

Cumulative distribution function (cdf)

The **cumulative distribution function** of a continuous random variable X, denoted as $F(x)$, is given by

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

for $-\infty < x < +\infty$.

Remark. If X is continuous random variable with cumulative distribution function $F(x)$ then we can use

$$P(a < X < b) = F(b) - F(a).$$

FPT Education
FPT UNIVERSITY

Examples

Example 1. Suppose the cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0, & \text{if } x < 1 \\ 0.5x - 0.5, & \text{if } 1 \le x < 3 \\ 1, & \text{if } x \ge 3 \end{cases}$$

Find

$$P(X < 2.8),\ P(0 < X < 1.5).$$

Example 2. Suppose that the probability density function of a continuous random variable X is:

$$f(x) = e^{-(x-3)}, \quad \text{for } x \ge 3.$$

Find the cumulative distribution function of X.

2. Mean and Variance

Mean and Variance

Suppose X is a continuous random variable with probability density function $f(x)$.

1. The **mean** or **expected value** of X is defined by

$$\mu = E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$$

2. The **variance** of X, denoted as σ^2 or $V(X)$ is:

$$\sigma^2 = V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{+\infty} x^2 f(x)\, dx - \mu^2$$

3. The **standard deviation** of X is $\sigma = \sqrt{V(X)}$

FPT Education
FPT UNIVERSITY

Examples

Example 1. Assume that X is a continuous random variable with the following probability density function

$$f(x) = \frac{x^2}{18}, \text{ for } -3 < x < 3.$$

Determine the mean and variance of X.

Example 2. The cumulative distribution function of the random variable X is

$$F(x) = \begin{cases} 0, & \text{if } x < 1 \\ 0.5x - 0.5, & \text{if } 1 \le x < 3 \\ 1, & \text{if } x \ge 3 \end{cases}$$

Find the standard deviation of X.

3. Continuous Uniform Distribution

Continuous uniform distribution

pdf:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

Mean & Variance:

$$\mu = E(X) = \frac{b+a}{2}, \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

cdf:

$$F(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } a \le x < b \\ 1, & \text{if } x \ge b \end{cases}$$

Examples

Example 1. Suppose X has a continuous uniform distribution over the interval $[1; 10]$.

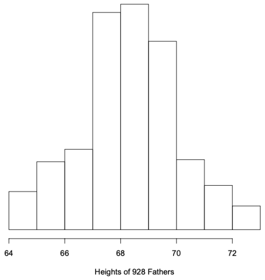
a. Determine the mean, variance and standard deviation of X.
b. Find $P(X < 6.5)$.
c. Determine the cumulative distribution function of X.

Example 2. Suppose X has a continuous uniform distribution over $[5; 15]$. What is the mean and variance of $Y = 8X$?

4. Normal distribution

The normal curve

Many data have histograms that look bell-shaped, e.g. heights, weights, IQ scores:



The data follow the normal curve. But remember that some data have histograms that look quite different, e.g. incomes, house prices.

Normal Distribution

Definition

The normal distribution, also known as the Gaussian distribution or bell curve, is a continuous probability distribution that is symmetric around its mean, which is also its median and mode. The shape of the normal distribution is characterized by its bell-shaped curve.

The probability density function (PDF) of a normal distribution is given by the formula:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where:

- μ is the mean of the distribution.
- σ is the standard deviation.

FPT Education
FPT UNIVERSITY

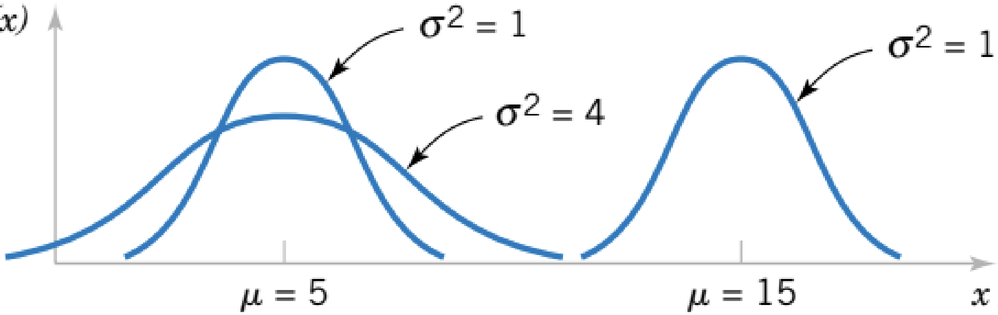


Figure: Normal probability density functions for selected values of the parameters μ and σ^2.

Properties of the normal distribution

1. The normal distribution is often denoted as $\mathcal{N}(\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance.

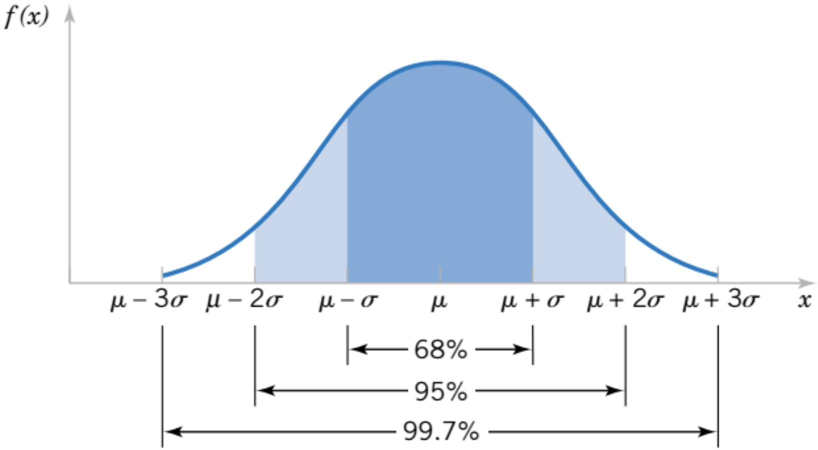
2. It is symmetric around the mean.

3. About 68% of the data falls within one standard deviation of the mean $(\mu \pm \sigma)$.

   About 95% falls within two standard deviations $(\mu \pm 2\sigma)$.

   About 99.7% falls within three standard deviations $(\mu \pm 3\sigma)$.

The empirical rule



Standard normal distribution

Definition

The standard normal distribution, often denoted as Z, is a special case of the normal distribution where the mean (μ) is 0 and the standard deviation (σ) is 1.

The probability density function (PDF) of the standard normal distribution is given by:

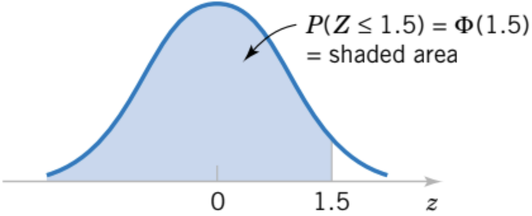
$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

The cumulative distribution function (CDF) of the standard normal distribution, denoted as $\Phi(z)$ or $P(Z \leq z)$, represents the probability that a standard normal random variable is less than or equal to z.

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \, dx$$

Computing percentiles for normal data

The standard normal distribution is crucial in statistics and probability theory.
Many statistical methods and tables are based on the properties of the standard normal distribution.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Examples

Example 1. Find the standard normal-curve area between $z = 1$ and $z = 0.4$?

Hint. Use Appendix Table III or NORM.S.DIST(a;1) in Excel to find $P(z < a)$.

Example 2. Assume that z- scores are normally distributed with a mean of 0 and a standard deviation of 1. If $P(z < a) = 0.1487$, find a.

Hint. Use Appendix Table III or NORM.S.INV(p) in Excel to find a such that $P(z < a) = p$.

Standardizing Normal Distribution

Theorem

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma}$ is a standard normal random variable $\mathcal{N}(0, 1)$.

Example 1. A machine pours beer into 16 oz bottles. Experience has shown that the number of ounces poured is normally distributed, with a standard deviation of 1.3 ounces. *Find the probability that the amount of beer the machine will pour into the next bottle will be more than 16.25 ounces.*

Example 2. The tread life of a particular brand of tire is a random variable best described by a normal distribution with a mean of 65,000 miles and a standard deviation of 1,500 miles. *What warranty should the company use if they want 95% of the tires to outlast the warranty?*

5. Exponential distribution

Exponential distribution

Definition

The random variable X that equals the distance between two consecutive events with mean number of events $\lambda > 0$ per unit interval is an exponential random variable with parameter λ. The probability density function of X is:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

Theorem

If a random variable X has exponential distribution with parameter λ, then

$$\mu = E(X) = \frac{1}{\lambda}, \quad \sigma^2 = V(X) = \frac{1}{\lambda^2}.$$

Examples

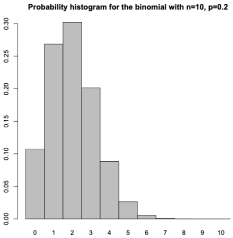
Example 1. The time between customer arrivals at a furniture store has an approximate exponential distribution with a mean of 9 minutes.
If a customer just arrived, find the probability that the next customer will not arrive for at least 15 minutes.

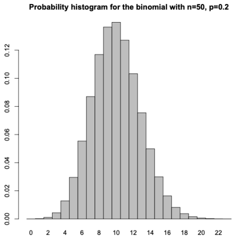
Example 2. The time between patients arriving at an outpatient clinic follows an exponential distribution at a rate of 15 patients per hour.
What is the probability that a randomly chosen arrival interval will not exceed 6 minutes?

6. Normal approximations

Normal approximation to the Binomial Distribution

As the number of experiments n gets larger, the probability histogram of the binomial distribution looks more and more similar to the normal curve:





In fact, we can approximate binomial probabilities using **normal approximation**: to standardize, subtract off np and then divide by $\sqrt{np(1-p)}$.

FPT UNIVERSITY

Normal approximation to the Binomial Distribution

Theorem

If X has a Binomial distribution $B(n, p)$, then random variable

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately a standard random variable $\mathcal{N}(0, 1)$.

We have

$$P(X \le x) \approx P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right), \quad P(X \ge x) \approx P\left(Z \ge \frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Remark. The approximation is good for $np > 5$ and $n(1-p) > 5$.

Example

The manufacturing of semiconductor chips produces 3% defective chips. Assume the chips are independent and that a lot contains 800 chips. Approximate the probability that more than 30 chips are defective.

Hint. $X \sim B(n = 800; p = 0.03)$ use Normal approximation.

Remark. Can use BINOM.DIST in Excel to find actual value.

Normal Approximation to the Poisson Distribution

Theorem

If X has Poisson distribution $P(\lambda)$, then random variable

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard random variable $\mathcal{N}(0, 1)$.

We have

$$P(X \leq x) \approx P\left(Z \leq \frac{x + 0.5 - \lambda}{\sqrt{\lambda}}\right), \quad P(X \geq x) \approx P\left(Z \geq \frac{x - 0.5 - \lambda}{\sqrt{\lambda}}\right)$$

Remark. The approximation is good for $\lambda > 5$.

Example

The number of customers that arrive at a fast-food business during a one-hour period is known to be Poisson distributed, with a mean equal to 9.6. What is the probability that more than 10 customers will arrive in a one-hour period?

Hint. $X \sim P(9.6)$ use Normal approximation.

Remark. Can use POISSON.DIST in Excel to find actual value.

FPT Education
FPT UNIVERSITY

Thank you!

namvv14@fe.edu.vn