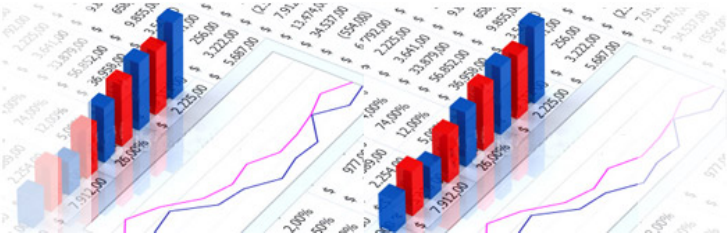
MTMS.01.099 Mathematical Statistics

Lecture 3. Continuous random variable

Tõnu Kollo



Fall 2017

Recall from the previous lecture

- A **random variable** is a variable whose possible values are numerical outcomes of a random phenomenon.
- If a rv can take on only a finite or countably infinite number of different values, the variable is said to be **discrete.**
- The **probability distribution** of rv describes how probabilities are distributed over the values of the rv.
- The **expected value** (mean) of any discrete rv is an average of possible outcomes, with each outcome weighted by its probability.
- **Standard deviation** is used as measure of dispersion. The square root of the variance is the standard deviation of rv.

Some dicrete distributions

Consider event A with probability $P(A) = p$.

Definition

A random variable X with possible values 0 and 1 has *Bernoulli distribution,* $X \sim Be(p)$ *if*

x_i	*0*	*1*
p_i	$1 - p$	p

Definition

A simple random variable X has *discrete uniform distribution, if*

$$P(X = x_i) = \frac{1}{n}, \quad i = 1, ..., n.$$

Binomial distribution

Definition

The random variable X that counts the number of successes k in the n trials is said to have a **binomial distribution** with parameters n and p, written $X \sim Bin(n, p)$ or $X \sim B(n, p)$.

The probability function of a binomial random variable X with parameters n and p is

$$p(k) = P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

for $k = 0, 1, \ldots, n$.
C_n^k counts the number of outcomes that include exactly k successes and $n - k$ failures.

Poisson distribution

A Poisson distributed random variable is often useful in estimating the number of occurrences over a specified interval of time or space.

Definition

If the random variable X has probability function

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where $k = 0, 1, \ldots$ and distribution parameter $\lambda > 0$ is a constant, then X is said to have a **Poisson distribution**, written $X \sim Po(\lambda)$ or $X \sim P(\lambda)$.

Continuous random variable

Situations that involve measuring some object often results in a continuous random variable.

Definition

Continuous random variable can take any value in some part of the real line(in uncountable set).

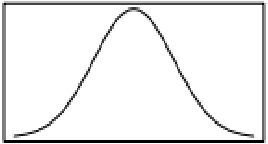
Definition

X is a **continuous random variable** if there is a function $f(x) \geq 0$ so that for any constants a and b, with $-\infty \leq a \leq b \leq \infty$,

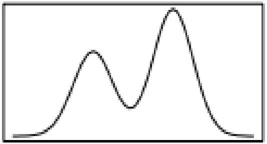
$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

The function $f(x)$ is the **probability density function** (pdf).

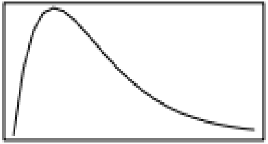
Common shapes of probability density functions



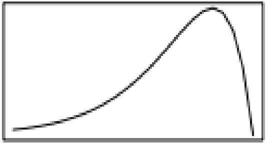
symmetric



bimodal



positively skewed



negatively skewed

Examples. Remarks

Examples of continuous random variables:

- the weight of a random person (a real number)
- randomly selected point inside a unit interval
- the waiting time until a taxi comes to the taxi-stop
- length of a telephone call
- ...

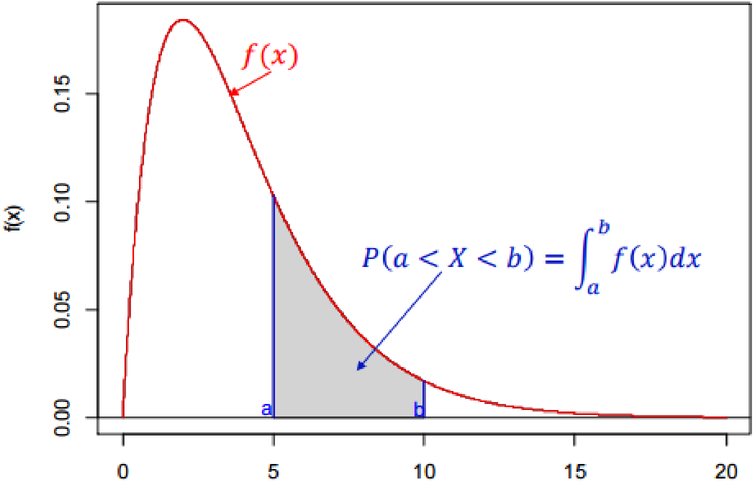
Remarks:

- A continuous random variable X has infinitely many possible values.
- All continuous probability models assign probability 0 to every individual outcome, e.g. $P(X = x) = 0$ for any x.
- Only intervals of values have positive probability.

Hence,

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

Density function of a continuous random variable



(Cumulative) Distribution Function (cdf)

The distribution function of a continuous rv is defined as for discrete rv-s:

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]).$$

Let $F'(x)$ be the derivative of $F_X(x)$ with respect to x.

Theorem 3.

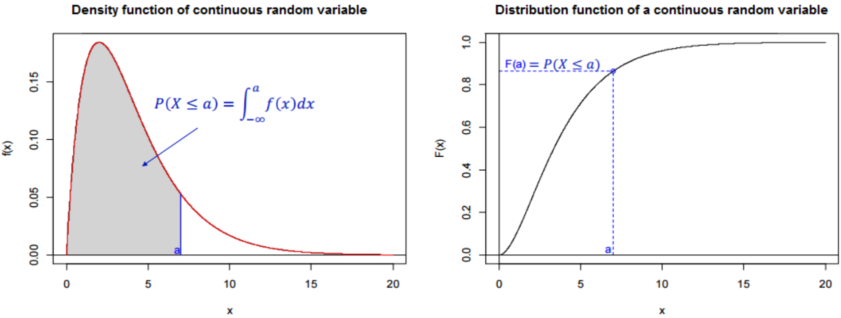
At any point x of continuity of $f_X(x)$:

$$F'_X(x) = f_X(x).$$

Hence,

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

Density function and distribution function of a continuous random variable



Some properties

Properties:

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- For two numbers a and b, if $a < b$, then $F(b) \geq F(a)$
- (Theorem 2) If $a \leq b$, then

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

- $F_X(b) - F_X(a) = \int_a^b f_X(t)dt$
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$ (the total area under the density function is 1)

Expectation of continuous random variables

Definition

The expectation of the continuous random variable X with pdf $f_X(x)$ is given by

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

- Intuitively, this comes from the discrete case by replacing $\sum$ with $\int$ and $p(x_i)$ with $f(x)dx$.

Also, if $Y = g(X)$ we have

$$E(Y) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx.$$

Example 1.1

Example

The density function of X is

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \end{cases}$$

Find the expected value of X.

Example 1.1

Example

The density function of X is

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \end{cases}$$

Find the expected value of X.

Solution:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \Big|_a^b = \frac{a+b}{2}.$$

Variance of continuous random variables

Recall that the variance of a random variable is defined as

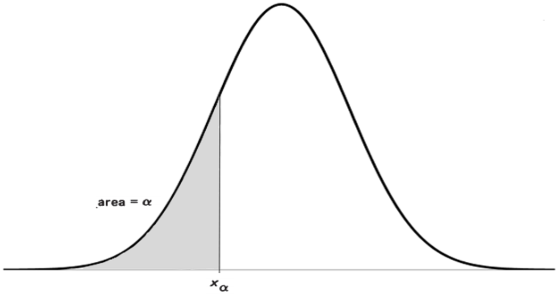
$$Var(X) = E(X - E(X))^2.$$

Using the definition of continuous random variable, we obtain the formula for the variance of continuous rv:

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

NB! The basic properties of mean and variance are the same as for discrete random variables!

Quantiles



If x_α is determined such that the area under the density function to the left of x_α is equal to a given quantity α, $0 < \alpha < 1$, one obtains the so-called α-**quantile** of the distribution (see the figure above).

Quantiles

The α-quantile can also be defined using the distribution function.

Definition

α-quantile of a continuous rv X is the number, denoted by x_α, with the property

$$F(x_\alpha) = P(X \leq x_\alpha) = \alpha,$$

where F is the distribution function of X.

Thus, $x_{0.95}$ separates the top 5% of the probability mass of X from the rest.

Quantiles (2)

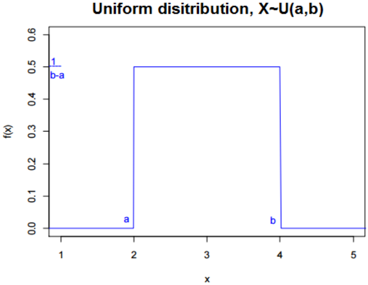
- $F(x_{0.5}) = P(X \leq x_{0.5}) = 0.5$ – $x_{0.5}$ is the median.
- $F(x_{0.75}) = P(X \leq x_{0.75}) = 0.75$ – $x_{0.75}$ is also called the **upper quartile.**
- $F(x_{0.25}) = P(X \leq x_{0.25}) = 0.25$ – $x_{0.25}$ is also called the **lower quartile.**
- For any given α, x_α can be found by solving the equation $F(x_\alpha) = \alpha$, for x_α.

Uniform distribution

Definition

A continuous random variable X has a uniform distribution in interval $[a, b]$, if the pdf is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{elsewhere.} \end{cases}$$

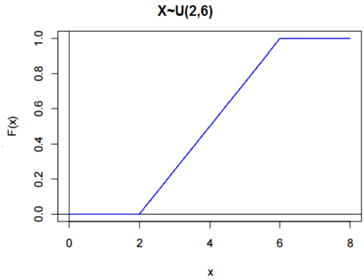


Uniform distribution

The distribution function of uniform distribution is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b, \\ 1 & \text{if } x > b. \end{cases}$$

The mean and variance of a random variable X having the uniform distribution: $E(X) = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$.



Exponential distribution

Definition

A continuous rv X has an **exponential distribution** with parameter $\lambda > 0$ if the pdf is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

It is written as $X \sim Exp(\lambda)$. The distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

The mean and variance of a random variable X having the exponential distribution: $E(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

Exponential distribution

Example

Let X be a waiting time in minutes till the first e-mail. Server receives e-mails in average in every 15 seconds (in $\frac{1}{4}$ minute), eg $\lambda = 4$ per minute.

What is the probability that the server will not receive any e-mail within 30 seconds (0.5 min)?

Exponential distribution

Example

Let X be a waiting time in minutes till the first e-mail. Server receives e-mails in average in every 15 seconds (in $\frac{1}{4}$ minute), eg $\lambda = 4$ per minute.

What is the probability that the server will not receive any e-mail within 30 seconds (0.5 min)?

We have to find $P(X \geq 0.5)$.

Exponential distribution

Example

Let X be a waiting time in minutes till the first e-mail. Server receives e-mails in average in every 15 seconds (in $\frac{1}{4}$ minute), eg $\lambda = 4$ per minute.

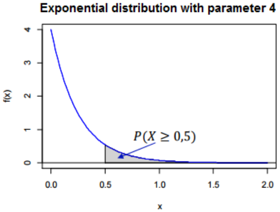
What is the probability that the server will not receive any e-mail within 30 seconds (0.5 min)?

We have to find $P(X \geq 0.5)$.

$$P(X \geq 0{,}5) = 1 - P(X < 0{,}5) = 1 - P(X \leq 0{,}5) =$$

$$= 1 - \int_0^{0{,}5} f(x)dx = 1 - \int_0^{0{,}5} 4e^{-4x}dx \approx$$

$$\approx 1 - 0{,}86 = 0{,}14.$$



Normal distribution

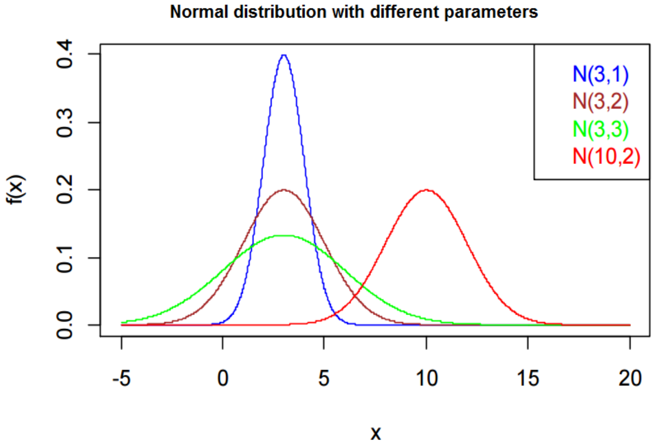
Definition

A continuous rv X is said to have a **normal distribution** with parameters μ and σ, where $-\infty < \mu < \infty$ and $\sigma > 0$, if the pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ -\infty < x < \infty.$$

It is written as $X \sim N(\mu, \sigma^2)$.
The mean and variance of a random variable X having the normal distribution: $E(X) = \mu$ and $Var(X) = \sigma^2$.

Normal distribution with different parameters



Standard normal distribution

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called a **standard normal distribution.** The random variable is usually denoted by Z. The pdf of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty < z < \infty.$$

The distribution function of Z is denoted by Φ. Thus

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(x) dx.$$

Recall that $\Phi(-z) = 1 - \Phi(z)$.

Theorem

If $X \sim N(\mu, \sigma^2)$ and $Y = \frac{X-\mu}{\sigma}$, then $Y \sim N(0, 1)$.