02403 Introduction to Mathematical Statistics

Lecture 2: Random variables and continuous distributions

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Agenda

1. Introduction
2. Continuous Distributions
   - Density and Distribution Functions
   - Mean, variance, and covariance
3. Some concrete continuous distributions
4. Calculation rules for mean and variance
5. The normal distribution revisitet
6. Multivariate distributions
7. Covariance and corrrelation

Overview

Discrete and continuous random variables

- We distinguish between *discrete* and *continuous* random variables.
- Discrete:
  - Number of people in this room who wear glasses.
  - Number of planes departing from CPH within the next hour.
- Continuous:
  - Wind speed measurement.
  - Transport time to DTU.

Random variable

Before the experiment is carried out, we have a random variable

$$Y \text{ (or } Y_1, \ldots, Y_n)$$

indicated with capital letters.

After the experiment is carried out, we have a *realization* or *observation*

$$y \text{ (or } y_1, \ldots, y_n)$$

indicated with lowercase letters.

Overview

Continuous Random Variables

Random variable Y.

The sample space S is now continuous.

Examples:

- Height of students
- Measurement of wind speed
- Time to cycle to DTU
- Measurement of blood sugar in patients
- ...

The density function, Definition 2.32

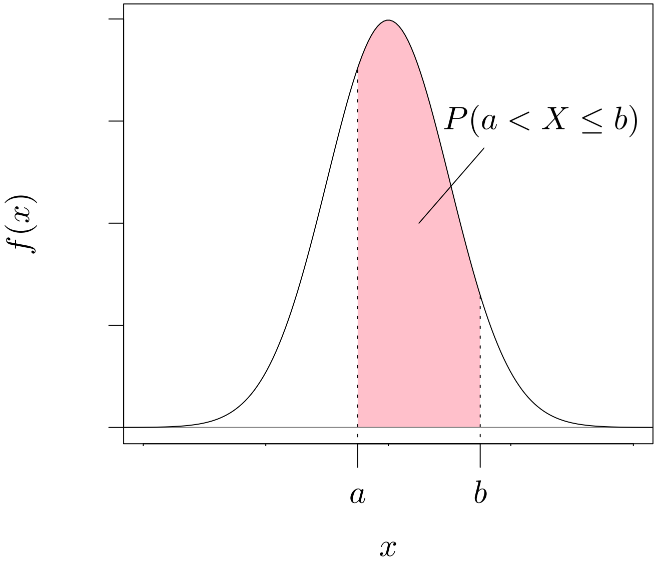
- The probability density function (pdf) for a random variable is denoted by $f(y)$.
- The density function says something about the frequency of the outcome y for the random variable Y.
- The probability that Y takes a value in the interval $[a;b]$ is given by the integral (the area under the curve):

$$P(a < Y \leq b) = \int_a^b f(y)\, dy\,.$$

- *No* direct probability for pdf. In fact, $P(Y = y) = 0$ for all y.
- The density function $f(y)$ for the distribution of a continuous random variable satisfies that

$$f(y) \geq 0 \text{ for all } y \quad \text{and} \quad \int_{-\infty}^{\infty} f(y)\, dy = 1\,.$$

The density function (Continuous)



The distribution function, Definition 2.33

- The distribution function (cumulative density function, cdf) for a continuous random variable is denoted by $F(y)$.
- The distribution function is defined by

$$F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(t)\,dt\,.$$

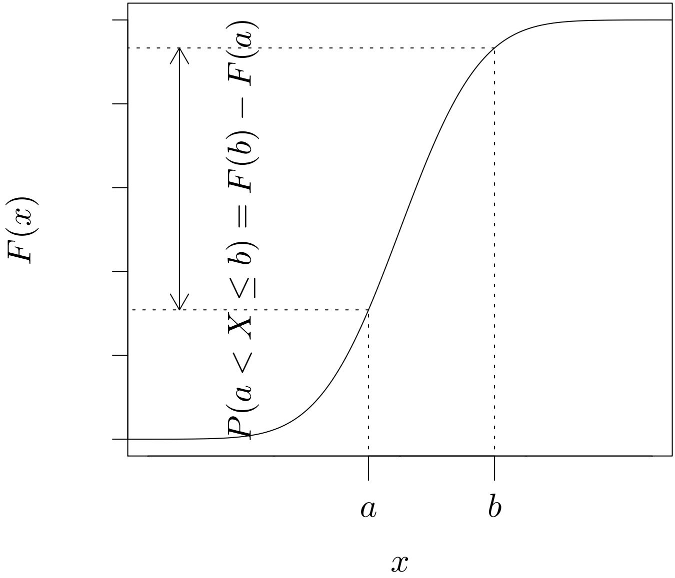
- Note that as a consequence of this definition,

$$f(y) = F^{'}(y)\,.$$

- It's particularly useful to note that

$$P(a < Y \leq b) = \int_{a}^{b} f(y)\,dy \; = F(b) - F(a).$$

Continuous distribution function



Mean, continuous random variable, Definition 2.34

The mean/expected value and variance of a continuous random variable:

$$\mu = \mathrm{E}[Y] = \int_{-\infty}^{\infty} y\, f(y)\, dy$$
$$\sigma^2 = \mathrm{E}[(Y-\mu)^2] = \int_{-\infty}^{\infty} (y-\mu)^2\, f(y)\, dy$$

Continuous and discrete distributions: some facts

Discrete	Continuous
$f(y) = P(Y = y)$	$f(y) \neq P(Y = y) = 0$
$F(y) = \sum_{-\infty}^{y} f(i)$	$F(y) = \int_{-\infty}^{y} f(t)dt$
$P(Y \leq y) \neq P(Y < y)$	$P(Y \leq y) = P(Y < y)$
$P(y_1 < Y \leq y_2) = \sum_{y_1+1}^{y_2} f(i)$	$P(y_1 < Y \leq y_2) = \int_{y_1}^{y_2} f(t)dt$
$P(y_1 < Y \leq y_2) = F(y_2) - F(y_1)$	$P(y_1 < Y \leq y_2) = F(y_2) - F(y_1)$
$E[Y] = \sum_{-\infty}^{\infty} y f(y)$	$E[Y] = \int_{-\infty}^{\infty} y f(y)$
$V[Y] = \sum_{-\infty}^{\infty} (y - E[Y])^2 f(y)$	$V[Y] = \int_{-\infty}^{\infty} (y - E[Y])^2 f(y) dy$

Overview

Continuous distributions used in this course

We will consider the continuous distributions

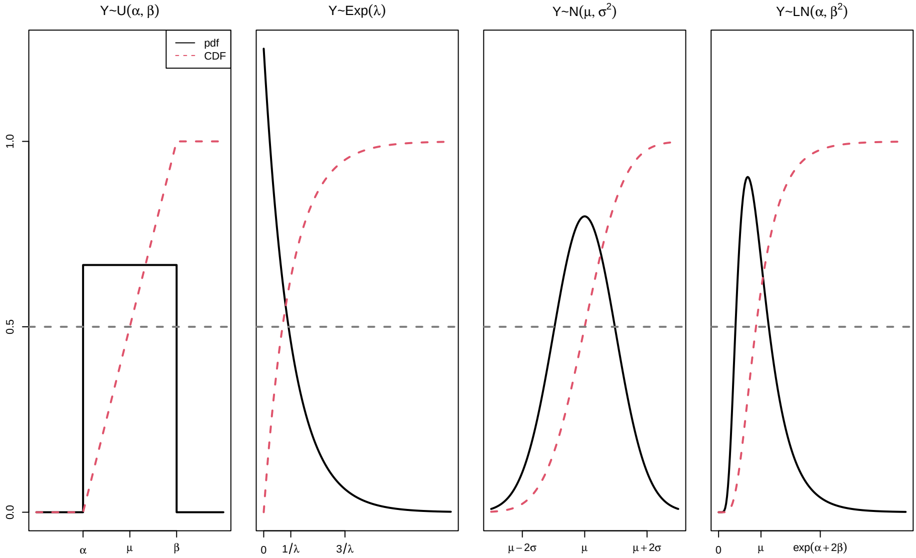
- The uniform distribution $U(\alpha, \beta)$
- The normal distribution $N(\mu, \sigma^2)$
- The log-normal distribution $LN(\alpha, \beta^2)$
- The exponential distribution $Exp(\lambda)$

Each sutiable for different situations.

Continuous distributions: overview

Distribution	pdf	μ	σ^2	Typical application
$Y \sim U(\alpha,\beta)$	$\frac{1}{\beta-\alpha}$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$	Constant density in the interval (α,β), zero outside.
$Y \sim Exp(\lambda)$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Time between arrivals when mean time equal λ.
$Y \sim N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	Distribution of measurement errors.
$Y \sim LN(\alpha,\beta^2)$	$\frac{1}{x\beta\sqrt{2\pi}} e^{-\frac{(\ln(x)-\alpha)^2}{2\beta^2}}$	$e^{\alpha+\beta^2/2}$	$\mu^2(e^{\beta^2}-1)$	Distribution of concentrations.

Density functions



As we did with the discrete distributions, will we use `Scipy.stats` for the continuous distributions (see documentation online).

General 'methods' for different distributions are:

`scipy.stats`	`.uniform/.norm/.lognorm/.expon`
`.rvs`	'random variates' (simulate random numbers)
`.pdf`	'probability density function' (pdf/density function)
`.cdf`	'cumulative distribution function' (distribution function)
`.ppf`	'percent point function' (inverse cdf / quantile function)
`.mean /.var /.std`	'mean'/'variance'/'standard deviation'

Example

You know that within each hour one bus is leaving. You arrive at the bus stop at 08:00.

- What is the probability that you will have to wait for more the 15 minutes?
- What is the probability that you have to wait for excactly 15 minutes?
- What is the probability that you will have to wait more than 1 hour?

Relations between distribution

Some exact ralations between distributions is

- If $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$ and independent, then $Y_1 + Y_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- If the time between events is Exponential distributed, then the number of events in a fixed time interval follow a Poisson distribution.
- If $Y \sim N(\mu, \sigma^2)$ then $e^Y \sim LN(\mu, \sigma^2)$
- If Y have distribution function $F(y)$ and $U \sim U(0,1)$ then $X = F^{-1}(U)$ have distribution function $F(x)$ (i.e. same as Y)

Some approximate relations between distributions

- If n is large and $Y \sim B(p,n)$ then Y will approximately follow a $N(np, np(1-p))$-distribution.
- If Y_i are independent and have mean μ and variance σ^2, then, for n large, the random variable $\bar{Y} = \frac{1}{n}\sum Y_i$ approximately follow a $N\left(\mu, \frac{\sigma^2}{n}\right)$-distribution. [1]

[1] If $Y_i \sim N\left(\mu, \sigma^2\right)$ then the relation is exact.

Example

Two students are counting the number of cars passing by on different stretches of road. They assume that the number of cars passing by in specific time intervals follow Poisson distributions. On the first road (road 1) they assume that the expected number of cars passing by is $\lambda_1 = 10/\text{hour}$, while on the second road (road 2) they assume that the expected number of cars passing by is $\lambda_2 = 15/\text{hour}$. [2]
Now they define two random variables:

- Y_1: number of cars passing by on road 1 in 15 minutes
- Y_2: number of cars passing by on road 2 in 10 minutes.

You can assume that Y_1 and Y_2 are independent.

- What is the probability that the time between two cars passing by is greater than 2 minutes on road 2?

[2] 2024 June

Example

Let $Y_i \sim LN(\mu, \sigma^2)$, $i = \{1, \ldots, n\}$ be independent random variables. [3]

- What is the probability $P(Y_1 Y_2 > k)$?

[3] 2022 June

Overview

Rules for stochastic variables (*both discrete and continuous*)

Let Y be a random variable, while a and b are constants, then

$$\mathrm{E}[aY+b] = a\mathrm{E}[Y]+b$$
$$\mathrm{V}[aY+b] = a^2\mathrm{V}[Y]$$

Let $Y_1,\ldots,Y_n$ be random variables, then

$$\mathrm{E}\left[\sum_{i}^{n} a_i Y_i\right] = \sum_{i=1}^{n} a_i \mathrm{E}[Y_i]$$

If $Y_1,\ldots,Y_n$ are *independent*, then

$$\mathrm{V}\left[\sum_{i=1}^{n} a_i Y_i\right] = \sum_{i=1}^{n} a_i^2 \mathrm{V}[Y_i]$$

Example 6

Planning for Airline

The individual weight of passengers on a flight, Y_i, is assumed to be normally distributed as $Y_i \sim N(70, 10^2)$.

A plane that can take 55 passengers can be loaded with a maximum of 4000 kg (only the passengers' weight is considered here as the load).

Question:

Calculate the probability that the plane will be overloaded.

Example

- Let $Y_1, \ldots, Y_n$ be independent identically distributed (iid.) random variables, what is $E(\overline{Y})$ and $Var(\overline{Y})$?
- If we additionally assume that Y_i are normally distributed which distribution will $\overline{Y}$ then follow?

Example

Let $Y_i \sim LN(\mu, \sigma^2)$, $i = \{1, \dots, n\}$ be independent random variables. [4]

- The quantity

$$Q = \left(\prod_{i=1}^{n} Y_i \right)^{1/n}$$

is also called the geometric mean. What is the mean and variance of Q?

[4] 2022 June

Overview

The standard normal distribution

The standard normal distribution:

$$Z \sim N(0, 1^2)$$

The normal distribution with mean 0 and variance 1.

Standardization:

An arbitrary normal distributed variable $Y \sim N(\mu, \sigma^2)$ can be *standardized* by

$$Z = \frac{Y - \mu}{\sigma}$$

Example

Measurement Error:

A given scale has a measurement error (measured in grams), Z, which can be described by a standard normal distribution, $Z \sim N(0, 1^2)$.

Question:

a) What is the probability that the scale gives a result that is at least 2 grams less than the true weight of the product?
b) What is the probability that the scale gives a result that is at least 2 grams more than the true weight of the product?
c) What is the probability that the scale has a deviation of at most ± 1 *gram?*
d) Find d *such that* $P(-d < Z < d) = 0.95$
e) With $Y \sim N(\mu, \sigma^2)$*, find* d *such that* $P(\mu - d < Y < \mu + d) = 0.95$

Visual Studio Code

Overview

Multivariate probability density functions

A multivariate probability density function for the random variable $\boldsymbol{Y} \in \mathbb{R}^n$, is a function from $\mathbb{R}^n$ into $\mathbb{R}_0$,

$$f(\boldsymbol{y}) = f(y_1, y_2, ..., y_n) \geq 0$$

such that

$$\int f(\boldsymbol{y}) d\boldsymbol{y} = \int \int \cdots \int f(y_1, y_2, ..., y_n) dy_1 dy_2 \cdots dy_n = 1,$$

further the marginal distribution for Y_i is given by

$$f_{Y_i}(y_i) = \int \int \cdots \int f(y_1, y_2, ..., y_n) dy_1 \cdots dy_{i-1} dy_{i+1} \cdots dy_n.$$

If a random variable $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T]^T$ have the joint density $f_{\boldsymbol{Y}}(\boldsymbol{y})$, then the marginal density of $\boldsymbol{Y}_1$ is

$$f_{\mathbf{Y}_1}(\boldsymbol{y}_1) = \int f_{\boldsymbol{Y}}(\boldsymbol{y}) d\boldsymbol{y}_2.$$

Example

Assume that the joint distribution of $[Y_1, Y_2]^T$ is

$$f(y_1, y_2) = k y_1 y_2; \quad [Y_1, Y_2]^T \in [0,1]^2,$$

and zero otherwise.

- Find k
- Find $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$

Overview

Second order moment representation

If a random vector $\boldsymbol{Y} \in \mathbb{R}^n$ have the probability density function $f_{\boldsymbol{Y}}$ then the mean and variance of Y_i is

$$E[Y_i] = \mu_i = \int y_i f_{Y_i}(y_i) dy_i$$
$$V[Y_i] = \sigma_{ii} = \int (y_i - \mu_i)^2 f_{Y_i}(y_i) dy_i,$$

and the covariances between Y_i and Y_j is

$$Cov[Y_i, Y_j] = \sigma_{ij} = \int (y_i - \mu_i)(y_j - \mu_j) f_{Y_i,Y_j}(y_i, y_j) dy_i dy_j.$$

Further the mean value vector of a random vector $\boldsymbol{Y} = [Y_1, \ldots, Y_n]^T$ is defined by

$$\boldsymbol{\mu} = E[\boldsymbol{Y}] = \begin{bmatrix} E[Y_1] \\ \vdots \\ E[Y_n] \end{bmatrix}$$

and the variance-covariance matrix is

$$\boldsymbol{\Sigma} = V[\boldsymbol{Y}],$$

where the elements of $\boldsymbol{\Sigma}$ are $\boldsymbol{\Sigma}_{ij} = Cov[Y_i, Y_j]$. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is referred to as the second order moment representation.

Example

Assume that the joint distribution of $[Y_1, Y_2]^T$ is

$$f(y_1, y_2) = k y_1 y_2; \quad [Y_1, Y_2]^T \in [0,1]^2,$$

and zero otherwise.

- Find $E[Y_1]$, $E[Y_2]$, $Cov[Y_1, Y_2]$

Calculation rules for random variable

Apply to both continuous and discrete random variable!

Let Y_1 and Y_2 be random variables with $Z_1 = a_0 + a_1 Y_1 + a_2 Y_2$ og $Z_2 = b_0 + b_1 Y_1 + b_2 Y_2$ then

$$\begin{aligned}
E(Z_1) =& a_0 + a_1 E(Y_1) + a_2 E(Y_2) \\
Cov(Z_1, Z_2) =& a_1 b_1 V(Y_1) + a_2 b_2 V(Y_2) + (a_1 b_2 + a_2 b_1) Cov(Y_1, Y_2) \\
V(Z_1) =& a_1^2 V(Y_1) + a_2^2 V(Y_2) + 2 a_1 a_2 Cov(Y_1, Y_2)
\end{aligned}$$

Let $Y_1, \ldots, Y_n$ be random variables, then:

$$\begin{aligned}
Z =& a_1 Y_1 + a_2 Y_2 + .. + a_n Y_n \\
Var(Z) =& a_1^2 Var(Y_1) + .. + a_n^2 Var(Y_n) + 2 a_1 a_2 Cov(Y_1, Y_2) + ... + \\
& 2 a_1 a_n Cov(Y_1, Y_n) + 2 a_2 a_3 Cov(Y_2, Y_3) + ... + \\
& 2 a_{n-1} a_n Cov(Y_{n-1}, Y_n)
\end{aligned}$$

Matrix calculation rules for random variable

Let $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ be random vectors with

$$V\begin{bmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12}\\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22}\end{bmatrix},$$

then

Theorem

Let the variance-covariance matrix of $[\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T]^T$ be as above and let $\boldsymbol{b}$ be a vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of appropriate dimensions, then

$$\begin{aligned} E[\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{b}] &= \boldsymbol{A}E[\boldsymbol{Y}_1] + \boldsymbol{b} \\ Cov[\boldsymbol{A}\boldsymbol{Y}_1, \boldsymbol{B}\boldsymbol{Y}_2] &= \boldsymbol{A}Cov[\boldsymbol{Y}_1, \boldsymbol{Y}_2]\boldsymbol{B}^T = \boldsymbol{A}\boldsymbol{\Sigma}^{12}\boldsymbol{B}^T \end{aligned}$$

and as a special case

$$V[\boldsymbol{A}\boldsymbol{Y}_1] = Cov[\boldsymbol{A}\boldsymbol{Y}_1, \boldsymbol{A}\boldsymbol{Y}_1] = \boldsymbol{A}\boldsymbol{\Sigma}^{11}\boldsymbol{A}^T.$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be such that $\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{B}\boldsymbol{Y}_2$ can be formed, then

$$V[\boldsymbol{A}\boldsymbol{Y}_1 + \boldsymbol{B}\boldsymbol{Y}_2] = \boldsymbol{A}\boldsymbol{\Sigma}^{11}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{\Sigma}^{22}\boldsymbol{B}^T + \boldsymbol{A}\boldsymbol{\Sigma}^{12}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Sigma}^{21}\boldsymbol{A}^T.$$

Example

Let $Y_1 \sim N(2,3)$ and $Y_2 \sim N(0,1)$, with $Cov(Y_1,Y_2) = 0$, what is $\mathsf{E}(Z)$, $\mathsf{V}(Z)$ and $\mathsf{Cov}(Y_1,Z)$ when $Z = Y_1 + 2Y_2$?

Independence

Definition

Let $f_{\boldsymbol{Y}}$ be the joint distribution of the random vector $\boldsymbol{Y} = [\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T]^T$, then $\boldsymbol{Y}_1$, and $\boldsymbol{Y}_2$ are independent if

$$f_Y(\boldsymbol{y}) = f_{Y_1}(\boldsymbol{y}_1) f_{Y_2}(\boldsymbol{y}_2).$$

If $\boldsymbol{Y}_1$, and $\boldsymbol{Y}_2$ are independent then

$$Cov(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \boldsymbol{0},$$

the converse statement does not apply.

Example

Assume that the joint distribution of $[Y_1, Y_2]^T$ is

$$f(y_1, y_2) = ky_1y_2; \quad [Y_1, Y_2]^T \in [0,1]^2,$$

and zero otherwise.

- Show that Y_1 and Y_2 are independent.

Covariance and correlation

The correlation coefficient between Y_i and Y_j is defined as

$$\rho_{ij} = \frac{Cov[Y_i, Y_j]}{\sqrt{V[Y_i]V[Y_j]}}$$

The covariance matrix is often decomposed into

$$\boldsymbol{\Sigma} = \boldsymbol{\sigma} \boldsymbol{R} \boldsymbol{\sigma}$$

with $\boldsymbol{\sigma}_{ii} = \sqrt{\boldsymbol{\Sigma}_{ii}}$, $\boldsymbol{\sigma}_{ij} = 0$ for $i \neq j$ and

$$\boldsymbol{R}_{ij} = \rho_{ij} = \frac{Cov[Y_i, Y_j]}{\sqrt{V[Y_i]V[Y_j]}} = \frac{\boldsymbol{\Sigma}_{ij}}{\sqrt{\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj}}}$$

Example

Let $Y_1 \sim N(2,3)$ and $Y_2 \sim N(0,1)$, with $Cov(Y_1,Y_2) = 0$ define $Z = Y_1 + 2Y_2$. For the random vector $[Y_1, Z]$ write down $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}$ and $\boldsymbol{R}$.

Summary of today