Course 02402 Introduction to Statistics

Lecture 3: Random variables and continuous distributions

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Overview

1. Continuous random variables and distributions
   - Density and distribution functions
   - Mean, variance, and covariance
2. Specific continuous distributions
   - The uniform distribution
   - The normal distribution
   - The log-normal distribution
   - The exponential distribution
3. Calculation rules for random variables

Overview

1. Continuous random variables and distributions
   - Density and distribution functions
   - Mean, variance, and covariance
2. Specific continuous distributions
   - The uniform distribution
   - The normal distribution
   - The log-normal distribution
   - The exponential distribution
3. Calculation rules for random variables

The density function, Definition 2.32

- The density function (probability density function, pdf) for a random variable is denoted by $f(x)$.

The density function, Definition 2.32

- The density function (probability density function, pdf) for a random variable is denoted by $f(x)$.
- The density function says something about the frequency of the outcome x for the random variable X.

The density function, Definition 2.32

- The density function (probability density function, pdf) for a random variable is denoted by $f(x)$.
- The density function says something about the frequency of the outcome x for the random variable X.
- The density function for a continuous random variable does *not* correspond directly to a probability. In fact, $P(X = x) = 0$ for all x.

The density function, Definition 2.32

- The density function (probability density function, pdf) for a random variable is denoted by $f(x)$.

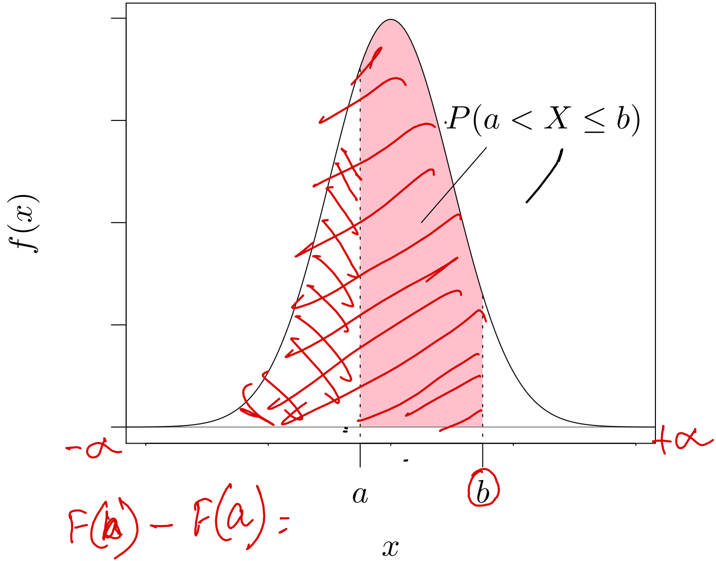
- The density function says something about the frequency of the outcome x for the random variable X.

- The density function for a continuous random variable does *not* correspond directly to a probability. In fact, $P(X = x) = 0$ for all x.

- The density function $f(x)$ for the distribution of a continuous random variable satisfies that

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

The density function



$F(b) - F(a) =$

The distribution function, Definition 2.33

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

The distribution function, Definition 2.33

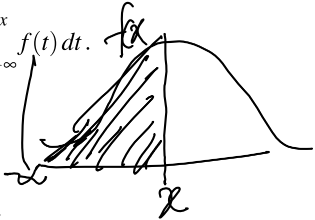
- The distribution function (cumulative density function, cdf) for a continuous random variable is denoted by $F(x)$.
- The distribution function is defined by

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

The distribution function, Definition 2.33

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

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



- Note that as a consequence of this definition,

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

The distribution function, Definition 2.33

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

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

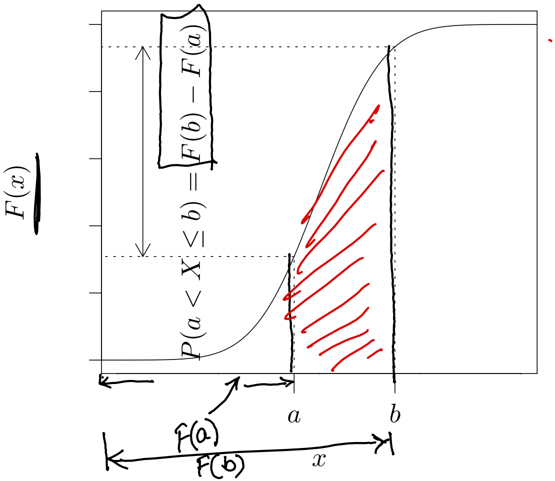
- Note that as a consequence of this definition,

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

- It's particularly useful to note that

$$P(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(x)\,dx\,.$$

The distribution function

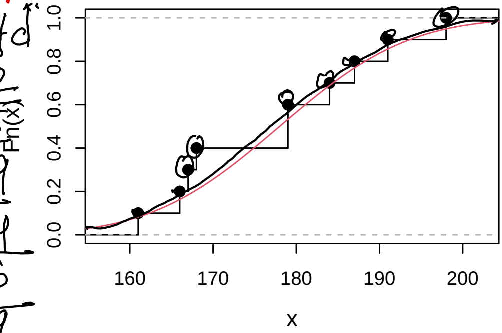


The empirical cumulative distribution function (ecdf)

```
# Empirical cdf for sample of height data from Chapter 1
x <- c(168, 161, 167, 179, 184, 166, 198, 187, 191, 179)
plot(ecdf(x), verticals = TRUE, main = "")

# 'True cdf' for normal distribution (with sample mean and variance)
xp <- seq(0.9*min(x), 1.1*max(x), length = 100)
lines(xp, pnorm(xp, mean(x), sd(x)), col = 2)
```

x (class)	freq f(x)	Cf
160 - 165	2	2
166 - 170	3	5
171 - 175	4	9
176 - 180	5	14
181 - 185	2	16
186 - 190	3	19



Histogram

Mean, continuous random variable, Definition 2.34

The mean/expected value of a continuous random variable:

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

Mean, continuous random variable, Definition 2.34

The mean/expected value of a continuous random variable:

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

Compare with the mean of a discrete random variable:

$$\mu = \sum_{\text{all } x} x f(x)$$

Variance, continuous random variable, Definition 2.34

The variance of a continuous random variable:

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

Variance, continuous random variable, Definition 2.34

The variance of a continuous random variable:

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

Compare with the variance of a discrete random variable:

$$\sigma^2 = \sum_{\text{all } x} (x-\mu)^2 f(x)$$

$\boxed{2.54}$ The mean and variance of a linear function of a random variable X

[Exam Question 2020dec 02402 VII.2 (21)]

a $E(aX+b) = aE(X)+b$

b $V(aX+b) = a^2V(X)$

V is the variance of the random variable.

Covariance, Definition 2.58

The covariance between two random variables:

Let X and Y be two random variables. Then, the covariance between X and Y is

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Relationship between covariance and independence:

If two random variables are *independent*, then their covariance is 0. *The reverse is not necessarily true!* i.e. if covariance is 0, then the value do not prove independence

. See Example 2.61 (P.89)

Remark 2.59 $\text{Cov}(X,X) = V(X)$

$\text{Cov}(X,Y) = \text{Cov}(Y,X)$

Variance of a Linear Combination of Two Random Variables

$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab\,\sigma_{X,Y}$

σ_{XY} is the covariance of X & Y

Overview

1. Continuous random variables and distributions
   - Density and distribution functions
   - Mean, variance, and covariance
2. Specific continuous distributions
   - The uniform distribution
   - The normal distribution
   - The log-normal distribution
   - The exponential distribution
3. Calculation rules for random variables

Specific continuous distributions

A number of statistical distributions exist (both continuous and discrete) that can be used to describe and analyze different types of problems.

Today, we'll take a closer look at the following continuous distributions:

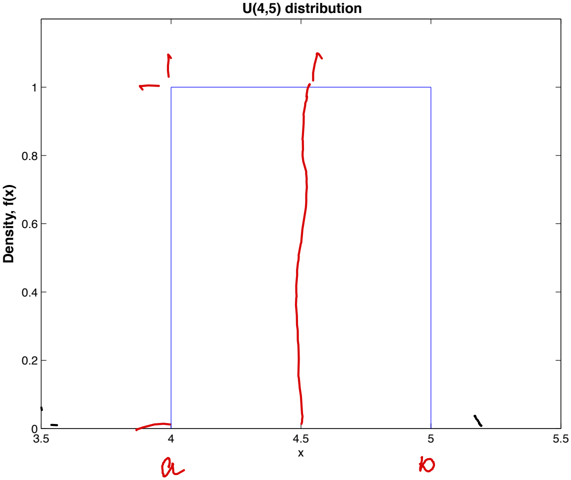
- The uniform distribution
- The normal distribution
- The log-normal distribution
- The exponential distribution

Continuous distributions in R

R	Distribution
norm	The normal distribution
unif	The uniform distribution
lnorm	The log-normal distribution
exp	The exponential distribution

- `d` Probability density function, $f(x)$.
- `p` Cumulative distribution function, $F(x)$.
- `q` Quantile function.
- `r` Random numbers from the distribution.

Density of a uniform distribution (example)



The uniform distribution, Def. 2.35 & Theo. 2.36

Syntax:

$X \sim U(\alpha, \beta)$

The uniform distribution, Def. 2.35 & Theo. 2.36

Syntax:

$X \sim U(\alpha, \beta)$

Density function:

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$

The uniform distribution, Def. 2.35 & Theo. 2.36

Syntax:

$X \sim U(\alpha, \beta)$

Density function:

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$

Mean:

$\mu = \frac{\alpha + \beta}{2}$

The uniform distribution, Def. 2.35 & Theo. 2.36

Syntax:

$X \sim U(\alpha, \beta)$

Density function:

0 30

$f(x) = \frac{1}{\beta - \alpha}$ for $\alpha \leq x \leq \beta$

$\frac{1}{30-0} = \frac{1}{30}$

Mean:

$\mu = \frac{\alpha+\beta}{2}$

Variance:

$\sigma^2 = \frac{1}{12}(\beta - \alpha)^2$

Example 1

Students attending a stats course arrive at a lecture between 8.00 and 8.30. It is assumed that the arival times can be described by a uniform distribution.

Question:

What is the probability that a randomly selected student arrives between 8.20 and 8.30?

Example 1

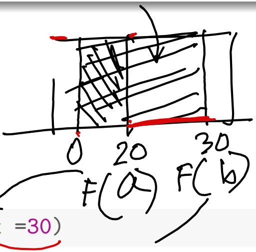
Students attending a stats course arrive at a lecture between 8.00 and 8.30. It is assumed that the arival times can be described by a uniform distribution.

Question:

What is the probability that a randomly selected student arrives between 8.20 and 8.30?

Answer:

$10/30 = 1/3$



Let $X \sim U(0, 30)$ represent arrival time. Then:
$P(20 \leq X \leq 30) = P(X \leq 30) - P(X \leq 20) = 1 - 2/3 = 1/3$

```
punif(q=30, min=0, max=30) - punif(q=20, min=0, max =30)
```

```
[1] 0.33
```

Example 1 (continued)

Question:

What is the probability that a randomly selected student arrives after 8.30?

Example 1 (continued)

Question:

What is the probability that a randomly selected student arrives after 8.30?

Answer:

0

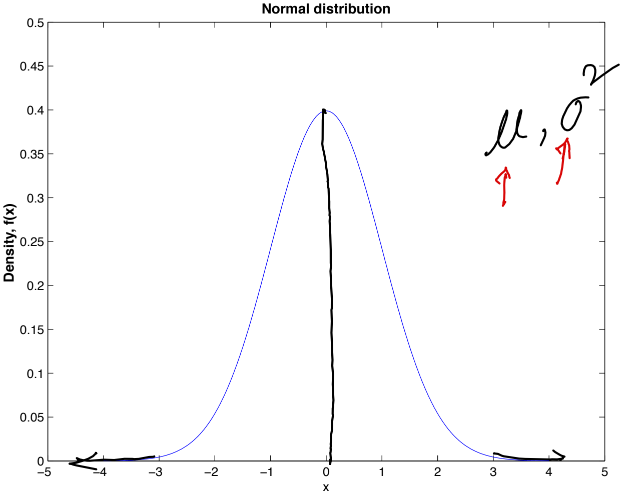
Let $X \sim U(0, 30)$ represent arrival time. Then:
$P(X > 30) = 1 - P(X \leq 30) = 1 - 1 = 0$

```
1 - punif(q=30, min=0, max=30)
```

[1] 0

Kahoot! 1-2

Density of a normal distribution (example)



The normal distribution, Def. 2.37 & Theo. 2.38

Syntax:

$X \sim N(\mu, \sigma^2)$

The normal distribution, Def. 2.37 & Theo. 2.38

Syntax:

$X \sim N(\mu, \sigma^2)$

Density function:

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

The normal distribution, Def. 2.37 & Theo. 2.38

Syntax:

$X \sim N(\mu, \sigma^2)$

Density function:

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

Mean:

$\mu = \mu$

The normal distribution, Def. 2.37 & Theo. 2.38

Syntax:

$X \sim N(\mu, \sigma^2)$

Density function:

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

Mean:

$\mu = \mu$

Variance:

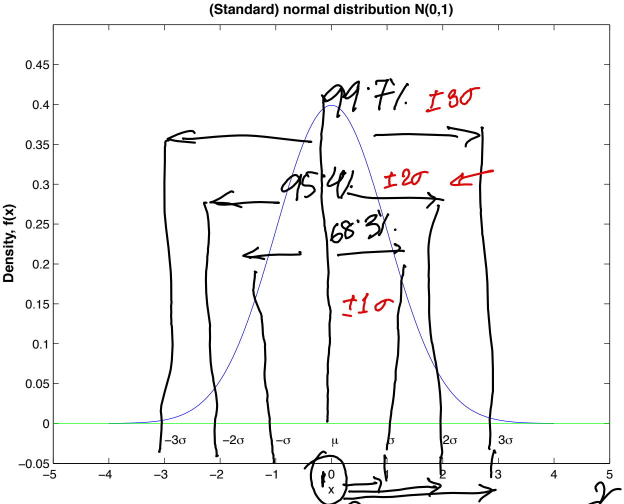
$\sigma^2 = \sigma^2$

Density of a standard normal distribution

$N(0, 1^2)$

See 2.43

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

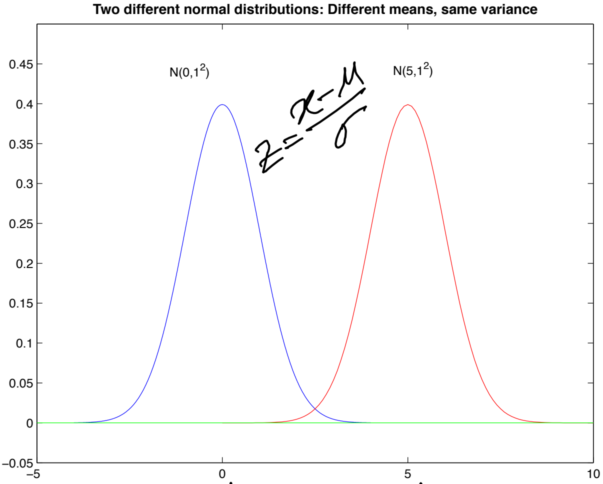


$z = 0$

$\mu = 0$

$\sigma^2 = 1$

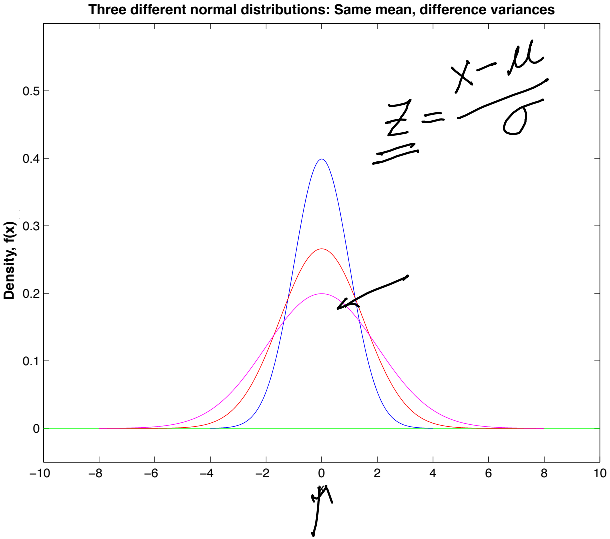
Density of two normal distributions (example)



$z = \frac{x-\mu}{\sigma}$

μ μ

Density of three normal distributions (example)



The standard normal distribution

The standard normal distribution:

$$Z \sim N(\underset{\mu}{0}, \underset{\sigma^2}{1^2})$$

The normal distribution with mean 0 and variance 1.

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 $X \sim N(\mu, \sigma^2)$ can be *standardized* by

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

$\bar{x}$.

s

Example 2

Measurement error:

A scale has a measurement error, Z, that can be described by the standard normal distribution, i.e.

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

That is, the mean measurement error is $\mu = 0$ with standard deviation $\sigma = 1$ gram. The scale is used to measure the weight of a product.

Example 2

Measurement error:

A scale has a measurement error, Z, that can be described by the standard normal distribution, i.e.

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

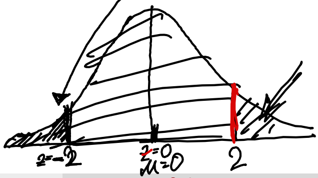
That is, the mean measurement error is $\mu = 0$ with standard deviation $\sigma = 1$ gram. The scale is used to measure the weight of a product.

Question a):

What is the probability that the scale yields a measurement which is at least 2 grams smaller than the true weight of the product?

1 - pnorm(2)

pnorm(-2)



Example 2

Measurement error:

A scale has a measurement error, Z, that can be described by the standard normal distribution, i.e.

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

That is, the mean measurement error is $\mu = 0$ with standard deviation $\sigma = 1$ gram. The scale is used to measure the weight of a product.

Question a):

What is the probability that the scale yields a measurement which is at least 2 grams smaller than the true weight of the product?

Answer:

$P(Z \leq -2) = 0.02275$

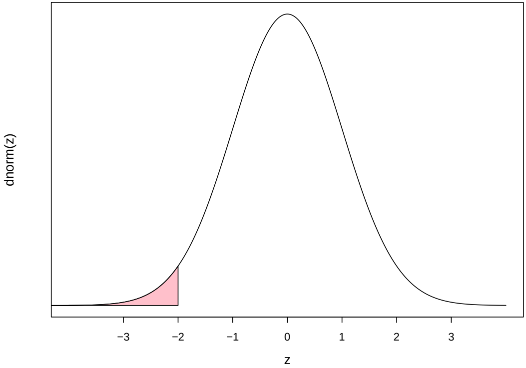
```
pnorm(-2); pnorm(q=-2, mean =0, sd=1)
```

Example 2

Answer:

```
pnorm(-2)
```

[1] 0.023



Example 2

Question b):

What is the probability that the scale yields a measurement which is at least 2 grams larger than the true weight of the product?

Example 2

Question b):

What is the probability that the scale yields a measurement which is at least 2 grams larger than the true weight of the product?

Answer:

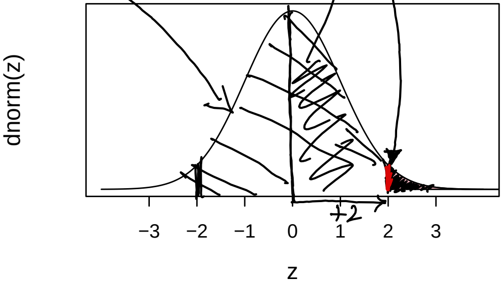
$P(Z \geq 2) = 0.02275$

```
1 - pnorm(2)
```

95.4% /2

+2

0.5



Example 2

Question c):

What is the probability that the scale is off by at most ± 1 *gram?*

Example 2

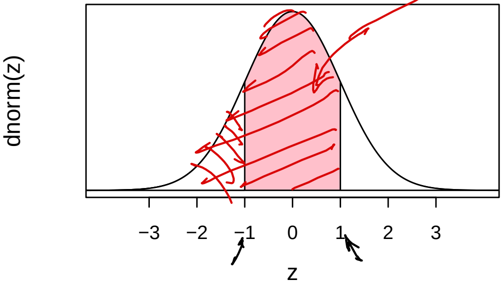
Question c):

What is the probability that the scale is off by at most ± 1 gram?

Answer:

$P(|Z| \leq 1) = P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1) = 0.683$ → 68.3%

```
pnorm(1) - pnorm(-1)
```



Example 3

Income distribution:

It is assumed that the annual salary distribution of elementary school teachers can be described using a normal distribution with mean $\mu = 290$ (in DKK thousand) and standard deviation $\sigma = 4$ (DKK thousand).

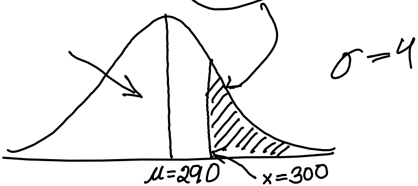
Example 3

Income distribution:

It is assumed that the annual salary distribution of elementary school teachers can be described using a normal distribution with mean $\mu = 290$ (in DKK thousand) and standard deviation $\sigma = 4$ (DKK thousand).

Question a):

What is the probability that a randomly selected teacher earns more than DKK 300.000?



Example 3

Question a):

What is the probability that a randomly selected teacher earns more than DKK 300.000?

Example 3

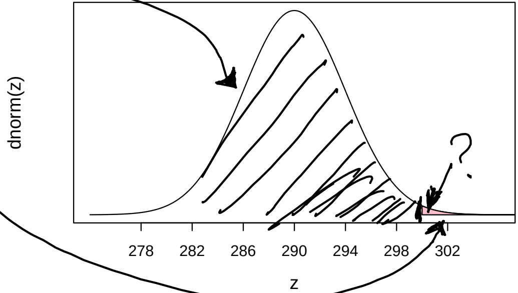
Question a):

What is the probability that a randomly selected teacher earns more than DKK 300.000?

Answer:

```r
1 - pnorm(300, m = 290, s = 4)
```

[1] 0.0062



Example 4

(Same income distribution):

It is assumed that the annual salary distribution of elementary school teachers can be described using a normal distribution with mean $\mu = 290$ (DKK thousand) and standard deviation $\sigma = 4$ (DKK thousand).

Example 4

(Same income distribution):

It is assumed that the annual salary distribution of elementary school teachers can be described using a normal distribution with mean $\mu = 290$ (DKK thousand) and standard deviation $\sigma = 4$ (DKK thousand).

"Opposite question"

Give a salary interval (symmetric around the mean) which covers 95% of all teachers' salary.

Example 4

(Same income distribution):

It is assumed that the annual salary distribution of elementary school teachers can be described using a normal distribution with mean $\mu = 290$ (DKK thousand) and standard deviation $\sigma = 4$ (DKK thousand).

"Opposite question"

5%

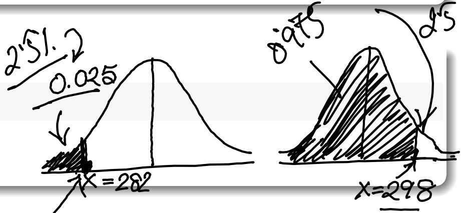
Give a salary interval (symmetric around the mean) which covers 95% of all teachers' salary.

Quantile ~ the specific values

Answer:

```
qnorm(c(0.025, 0.975), m = 290, s = 4)
```

```
[1] 282 298
```

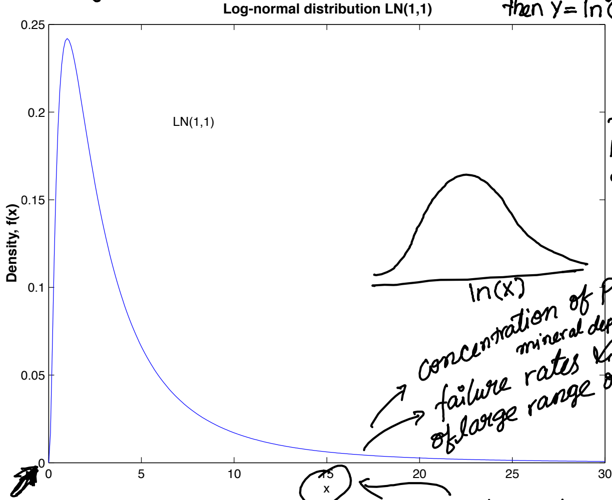


Kahoot! 3-5

The log-normal distribution

A lognormal distribution is a cont. prob. distribution of a random variable in which logarithm is normally distributed. Thus, if the random variable X has a lognormal distribution, then $Y = \ln(x)$ has a normal distribution.

Likewise, if Y has a normal distri, then $X = \exp(Y)$ has a lognormal distribution.



concentration of pollutants in environment / mineral deposit size etc. :

failure rates & other phenomena of large range of data

A variable that is known as never taking on negative values is normally assigned a lognormal distribution rather than a normal distribution.

The log-normal distribution, Def. 2.46 & Theo. 2.47

Syntax: $N(\mu, \sigma^2)$

$X \sim LN(\alpha, \beta^2)$ (with $\beta > 0$)

The log-normal distribution, Def. 2.46 & Theo. 2.47

Syntax:

$X \sim LN(\alpha, \beta^2)$ (with $\beta > 0$)

Density function:

$$f(x) = \begin{cases} \frac{1}{\beta\sqrt{2\pi}} x^{-1} e^{-(\ln(x)-\alpha)^2/2\beta^2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

The log-normal distribution, Def. 2.46 & Theo. 2.47

Syntax:

$X \sim LN(\alpha, \beta^2)$ (with $\beta > 0$)

Density function:

$$f(x) = \begin{cases} \frac{1}{\beta\sqrt{2\pi}} x^{-1} e^{-(\ln(x)-\alpha)^2/2\beta^2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Mean:

$\mu = e^{\alpha + \beta^2/2}$

The log-normal distribution, Def. 2.46 & Theo. 2.47

Syntax:

$X \sim LN(\alpha, \beta^2)$ (with $\beta > 0$)

Density function:

$$f(x) = \begin{cases} \frac{1}{\beta\sqrt{2\pi}} x^{-1} e^{-(\ln(x)-\alpha)^2/2\beta^2} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Mean:

$\mu = e^{\alpha+\beta^2/2}$

Example: $\boxed{\text{See Exam Question}}$ 2020Dec_02402_VII.2 (18)

2019dec_02402_XI.1 (26) & XI.2 (27)

Variance:

$\sigma^2 = e^{2\alpha+\beta^2}(e^{\beta^2} - 1)$

The log-normal distribution

Log-normal and normal distributions:

A log-normal distributed variable $Y \sim LN(\alpha, \beta^2)$ can be transformed into a normal distributed variable:

$$\underline{X = \ln(Y)}$$

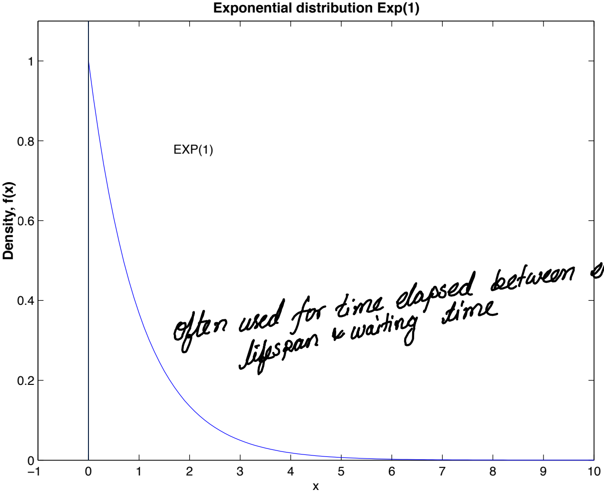
is normal distributed with mean α and variance β^2, i.e. $X \sim N(\alpha, \beta^2)$.

$$Z = \frac{\ln(Y) - \alpha}{\beta}$$

is standard normal distributed, i.e. $Z \sim N(0, 1)$.

See Exam question 2018 Dec_02402_Question VII.1 (16)

The exponential distribution



often used for time elapsed between events:
lifespan & waiting time

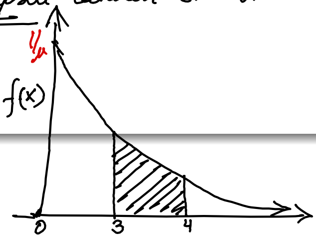
The exponential distribution, Def. 2.48 & Theo. 2.49

Syntax:

Often used to model the time elapsed between events.

$$X \sim \text{Exp}(\lambda)$$

with $\lambda > 0$.



Density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Mean:

$$\mu = \frac{1}{\lambda}$$

$$\lambda = \frac{1}{\mu}$$

Variance:

$$\sigma^2 = \frac{1}{\lambda^2}$$

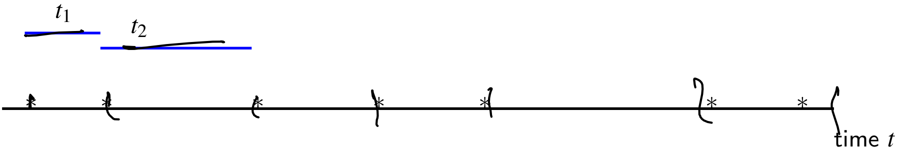
The exponential distribution

- The exponential distribution is a special case of the gamma distribution.
- The exponential distribution is used to describe lifespan and waiting times.
- The exponential distribution can be used to describe (waiting) time between Poisson events.

Connection between the exponential and Poisson distributions

Poisson: Discrete events per unit

Exponential: Continuous distance between events



Example 5

Queuing model – Poisson process

The time between customer arrivals at a post office is exponentially distributed with mean $\mu = 2$ minutes.

Example 5

Queuing model – Poisson process

The time between customer arrivals at a post office is exponentially distributed with mean $\mu = 2$ minutes.

Question:

One customer has just arrived. What is the probability that no other customers will arrive during the next 2 minutes?

Example 5

Queuing model – Poisson process

The time between customer arrivals at a post office is exponentially distributed with mean $\mu = 2$ minutes.

Question:

One customer has just arrived. What is the probability that no other customers will arrive during the next 2 minutes?

Or the next customer will arrive after 2 minutes or later?

$\lambda = \frac{1}{\mu} = \frac{1}{2}$

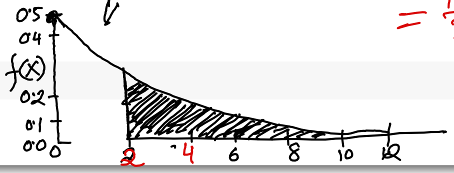
Answer:

$X \sim \text{Exp}(1/2)$ represents waiting time until next customer.

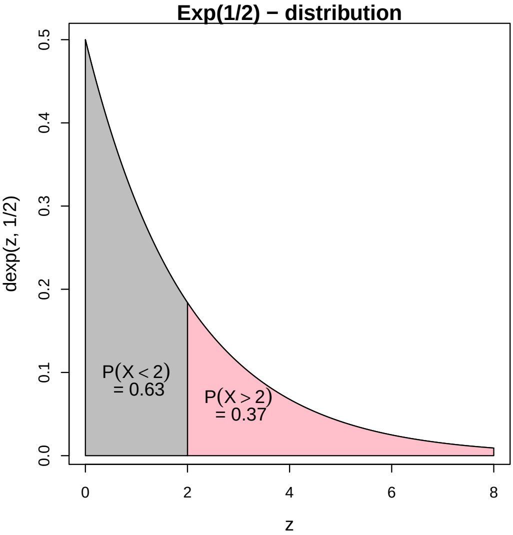
$P(X > 2) = 1 - P(X \leq 2)$

```
1 - pexp(2, rate = 1/2)
```

[1] 0.37



Example 5


Exp(1/2) − distribution
dexp(z, 1/2)
z
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
P(X < 2)
= 0.63
P(X > 2)
= 0.37

Example 6

Question:

One customer has just arrived. Use the Poisson distribution to calculate the probability that no other costumers will arrive during the next two minutes.

Example 6

Question:

One customer has just arrived. Use the Poisson distribution to calculate the probability that no other costumers will arrive during the next two minutes.

Answer:

$\lambda_{2min} = 1$, $P(X = 0) = \frac{e^{-1}}{1!} 1^0 = e^{-1}$

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$f(x)$ = the probability of x occurances in an interval

λ = expected value or mean number of occurances in an interval

$e = 2.71828$

```r
dpois(0,1)
```

lambda

x

[1] 0.37

```r
exp(-1)
```

[1] 0.37

Kahoot! 6-7

Overview

Calculation rules for random variables

These rules work for both continuous and discrete random variables!

X is a random variable, a and b are constants.

Calculation rules for random variables

These rules work for both continuous and discrete random variables!

X is a random variable, a and b are constants.

Mean rule:

$$\mathsf{E}(aX+b) = a\mathsf{E}(X)+b$$

Calculation rules for random variables

These rules work for both continuous and discrete random variables!

X is a random variable, a and b are constants.

Mean rule:

$$\mathsf{E}(aX+b) = a\mathsf{E}(X)+b$$

Variance rule:

$$\mathsf{Var}(aX+b) = a^2\mathsf{Var}(X)$$

Example 7

X is a random variable with mean 4 and variance 6.

Question:

Calculate the mean and variance of $Y = -3X + 2$

Example 7

X is a random variable with mean 4 and variance 6.

Question:

Calculate the mean and variance of $Y = -3X + 2$

Answer:

$$\mathsf{E}(Y) = -3\mathsf{E}(X) + 2 = -3 \cdot 4 + 2 = -10$$

$$\mathsf{Var}(Y) = (-3)^2 \mathsf{Var}(X) = 9 \cdot 6 = 54$$

Calculation rules for random variables

$X_1, \ldots, X_n$ are *independent* random variables.

Calculation rules for random variables

$X_1, \ldots, X_n$ are *independent* random variables.

Mean rule:

$$\mathsf{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1\mathsf{E}(X_1) + a_2\mathsf{E}(X_2) + \cdots + a_n\mathsf{E}(X_n)$$

Calculation rules for random variables

$X_1,\ldots,X_n$ are *independent* random variables.

Mean rule:

$$\mathsf{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1\mathsf{E}(X_1) + a_2\mathsf{E}(X_2) + \cdots + a_n\mathsf{E}(X_n)$$

Variance rule:

$$\mathsf{Var}(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1^2\mathsf{Var}(X_1) + \cdots + a_n^2\mathsf{Var}(X_n)$$

Example 8

Airline Planning

The weight of each passenger on a flight is assumed to be normal distributed $X \sim N(70, 10^2)$.

A plane, which can take 55 passengers, may not have a load exceeding 4000 kg (only the weight of the passengers is considered load).

Example 8

Airline Planning

The weight of each passenger on a flight is assumed to be normal distributed $X \sim N(70, 10^2)$.

A plane, which can take 55 passengers, may not have a load exceeding 4000 kg (only the weight of the passengers is considered load).

Question:

Calculate the probability that the plain is overloaded

Example 8

Airline Planning

The weight of each passenger on a flight is assumed to be normal distributed $X \sim N(70, 10^2)$.

A plane, which can take 55 passengers, may not have a load exceeding 4000 kg (only the weight of the passengers is considered load).

Question:

Calculate the probability that the plain is overloaded

What is Y = Total passenger weight?

Example 8

Airline Planning

The weight of each passenger on a flight is assumed to be normal distributed $X \sim N(70, 10^2)$.

A plane, which can take 55 passengers, may not have a load exceeding 4000 kg (only the weight of the passengers is considered load).

Question:

Calculate the probability that the plain is overloaded ne

We need the mean & variance of the plane's load.

What is Y = Total passenger weight?

What is Y?

Definitely NOT: $Y = 55 \cdot X$ 70 one passenger's weight? 70.

Example 8

What is Y = Total passenger weight?

$Y = \sum_{i=1}^{55} X_i$, where $X_i \sim N(70, 10^2)$ (and assumed to be independent)

Example 8

$E(X_1) + E(X_2) + \cdots + E(X_{55})$ — 70, 70, 70

What is Y = Total passenger weight?

$Y = \sum_{i=1}^{55} X_i$, where $X_i \sim N(70, 10^2)$ (and assumed to be independent)

Mean and variance of Y:

addition of all passengers' expected values & applying normal distribution approximate.

$$\mathsf{E}(Y) = \sum_{i=1}^{55} \mathsf{E}(X_i) = \sum_{i=1}^{55} 70 = 55 \cdot 70 = 3850$$

$E(X) = 70$
$V(X) = 10^2 = 100$

$$\mathsf{Var}(Y) = \sum_{i=1}^{55} \mathsf{Var}(X_i) = \sum_{i=1}^{55} 100 = 55 \cdot 100 = 5500$$

$Var(X_1) + Var(X_2) + \cdots Var(X_{55})$

$100 + 100 + 100$

Be Careful! it is not theorem 2.54 & formula (2-72)

NOT

Example 8

What is Y = Total passenger weight?

$Y = \sum_{i=1}^{55} X_i$, where $X_i \sim N(70, 10^2)$ (and assumed to be independent)

Mean and variance of Y:

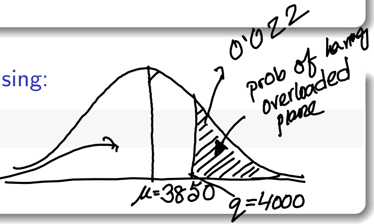
$$\mathsf{E}(Y) = \sum_{i=1}^{55} \mathsf{E}(X_i) = \sum_{i=1}^{55} 70 = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = \sum_{i=1}^{55} \mathsf{Var}(X_i) = \sum_{i=1}^{55} 100 = 55 \cdot 100 = 5500$$

Y is normal distributed, so we may find $P(Y > 4000)$ using:

```
1-pnorm(4000, mean = 3850, sd = sqrt(5500))
```

[1] 0.022



Example 8 - WRONG ANALYSIS

What is Y?

Definitely NOT: $Y = 55 \cdot X$

Example 8 - WRONG ANALYSIS

What is Y?

Definitely NOT: $Y = 55 \cdot X$

Mean and variance of WRONG Y:

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Example 8 - WRONG ANALYSIS

What is Y?

Definitely NOT: $Y = 55 \cdot X$

Mean and variance of WRONG Y:

a^2

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$
$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Wrong Y is also normal distributed. Finding $P(Y > 4000)$ using WRONG Y:

```
1 - pnorm(4000, mean = 3850, sd = 550)
```

[1] 0.39

Example 8 - WRONG ANALYSIS

Assume price of bikes vary and independent Price follow normal dist N(70, 10^2). We sell 55 bikes each month. What is the probability that revenue exceeds 4000 euros?
(Be careful! Applying the theorem 2.54 will be wrong analysis)

What is Y?

Definitely NOT: $Y = 55 \cdot X$

Mean and variance of WRONG Y:

$$\mathsf{E}(Y) = 55 \cdot 70 = 3850$$

$$\mathsf{Var}(Y) = 55^2 \mathsf{Var}(X) = 55^2 \cdot 100 = 550^2$$

Wrong Y is also normal distributed. Finding $P(Y > 4000)$ using WRONG Y:

```
1 - pnorm(4000, mean = 3850, sd = 550)
```

[1] 0.39

In this case, the price of each bike sold do not vary but in case of weight of passengers, they vary!

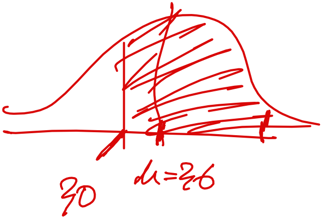
See typical example 2.55 (P.85)

Consequence of wrong calculation:

A LOT of wasted money for the airline company!!!

Overview

1. Continuous random variables and distributions
   - Density and distribution functions
   - Mean, variance, and covariance
2. Specific continuous distributions
   - The uniform distribution
   - The normal distribution
   - The log-normal distribution
   - The exponential distribution



3. Calculation rules for random variables