Course 02402 Introduction to Statistics Lecture 3:

Continuous Distributions

Per Bruun Brockhoff

DTU Compute
Danish Technical University
2800 Lyngby – Denmark
e-mail: perbb@dtu.dk

Agenda

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables

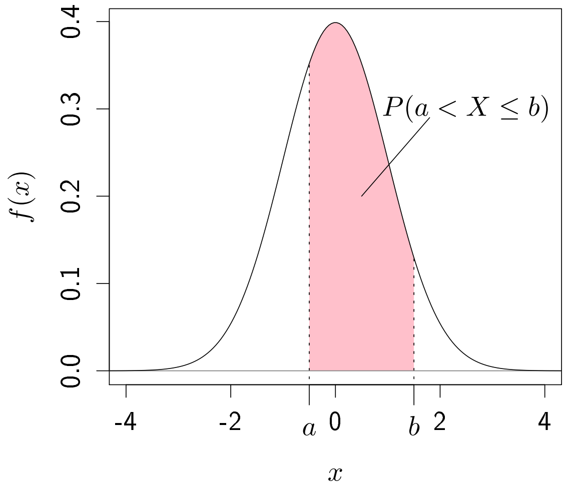
Oversigt

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables

The Density Function (pdf)

- The density function for a stochastic variable is denoted by $f(x)$
- $f(x)$ says something about the frequency of the outcome x for the stochastic variable X
- The density function for continuous variables does not correspond to the probability, that is $f(x) \neq P(X = x)$
- A nice plot of $f(x)$ is a histogram

The Density Function for Continuous Variables



The Density Function for Continuous Variables

The density function for a continuous variable is written as:

$$f(x)$$

The following is valid:

$$f(x) \geq 0 \quad \text{for all } x$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

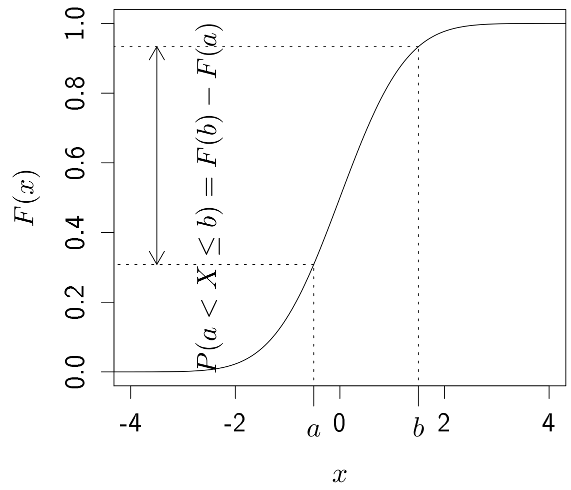
Distribution function or cumulative density function (cdf))

- The distribution function for a continuous stochastic variable is denoted by $F(x)$.
- The distribution function corresponds to the cumulative density function:

$$F(x) = P(X \leq x)$$

- $$F(x) = \int_{t=-\infty}^{x} f(t)dt$$
- A nice plot of $F(x)$ is the cumulative distribution plot
- $$f(x) = F'(x)$$

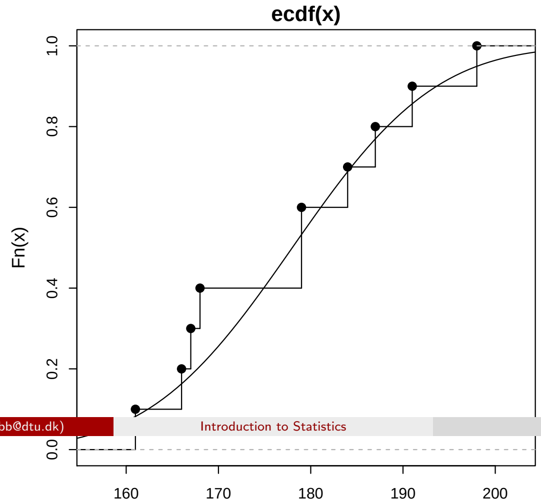
The distribution function(cdf))



The empirical cumulative distribution function - ecdf

Student height example from Chapter 1:

```
x <- c(168,161,167,179,184,166,198,187,191,179)
plot(ecdf(x), verticals = TRUE)
xp <- seq(0.9*min(x), 1.1*max(x), length.out = 100)
lines(xp, pnorm(xp, mean(x), sd(x)))
```



The Mean of a Continuous Random Variable

The Mean of a Continuous Random Variable

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

Compare with the discrete definition:

$$\mu = \sum_{i=1}^{\infty} x_i f(x_i)$$

The Variance of a Continuous Random Variable

The Variance of a Continuous Random Variable:

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

Compare with the discrete definition:

$$\sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i)$$

The Covariance of two random variables

The Covariance of two random variables:

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

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

Oversigt

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables

Specific Statistical Distributions

- A number of statistical distributions exist that can be used to describe and analyze different kind of problems

Now we consider continuous distributions

- The uniform distribution
- The normal distribution
- The log-normal distribution
- The Exponential distribution

The Uniform Distribution

Syntax:

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

Density function:

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

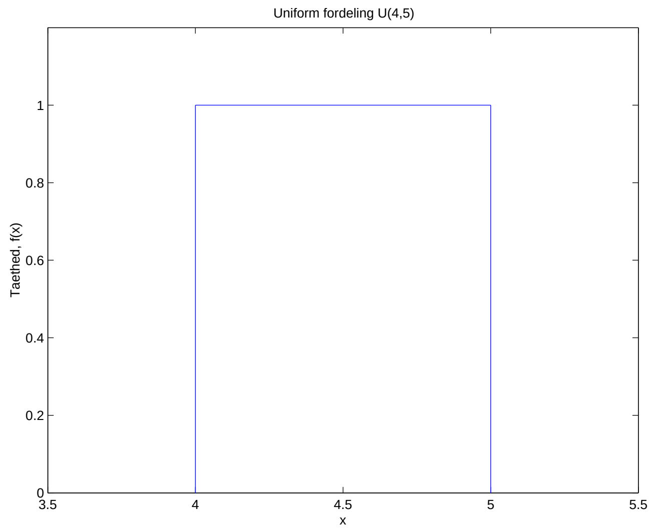
Mean:

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

Variance:

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

The Uniform distribution



Example 1

Students in a course arrive to 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 og 8.30?

Answer:

10/30=1/3

```
punif(30,0,30)-punif(20,0,30)
```

```
[1] 0.33
```


Example 1 - cont.

Question:

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

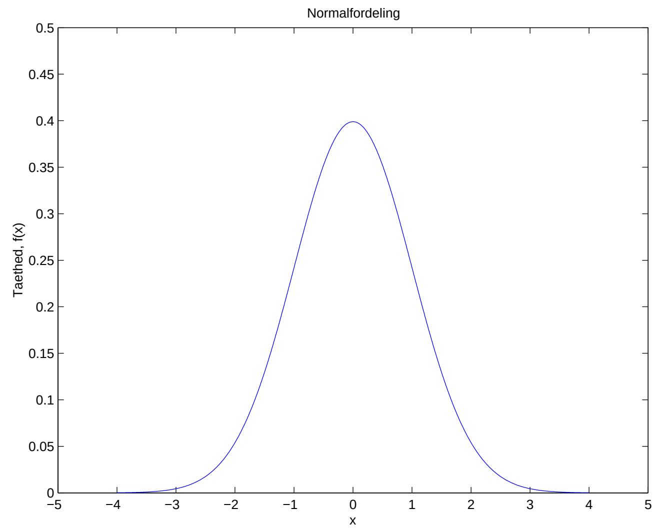
Answer:

0

```
1-punif(30,0,30)
```

```
[1] 0
```


The Normal Distribution



The Normal Distribution

Syntax:

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

Dnsity function:

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

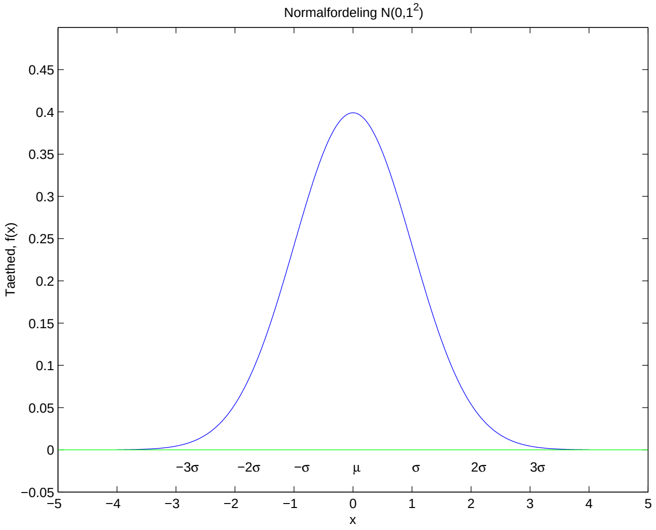
Mean:

$\mu = \mu$

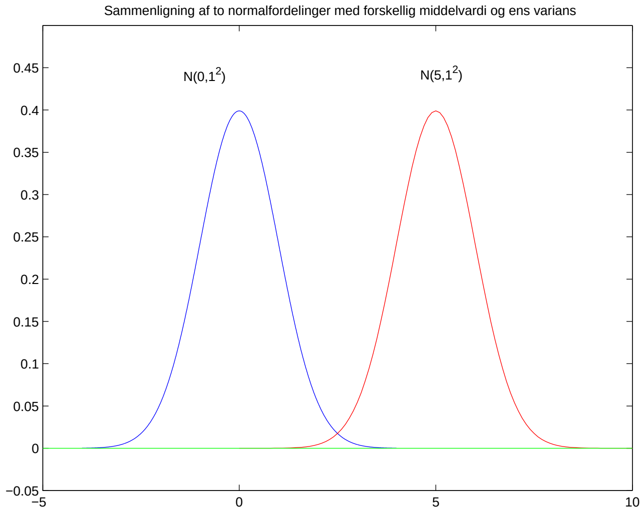
Variance:

$\sigma^2 = \sigma^2$

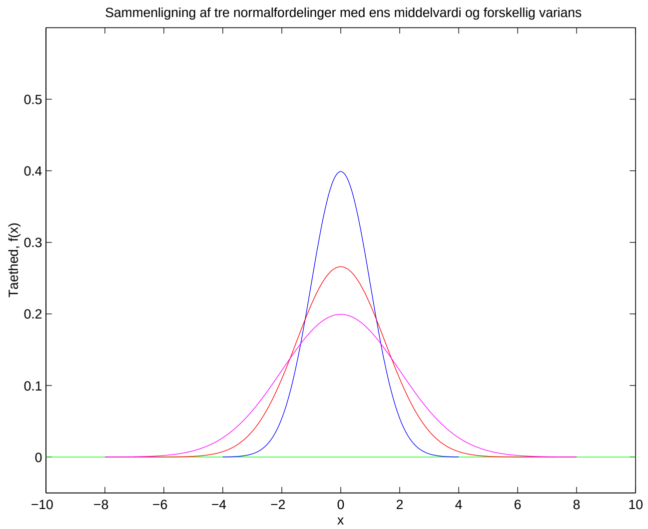
The Normal Distribution



The Normal Distribution



The Normal Distribution



The Normal Distribution

A standard normal distribution:

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

A normal distribution with mean 0 and variance 1.

Standardization:

An arbitrary normally distributed variable $X \sim N(\mu, \sigma^2)$ can be standardized by

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

Example 2

Measurement error:

A weight has a measurement error, Z, that can be described by a standard normal disrtribution, i.e.

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

that is, mean $\mu = 0$ and standard deviation $\sigma = 1$ gram.

We now measure the weight of a single piece

Question a):

What is the probability that the weight measures at least 2 grams too little?

Answer:

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

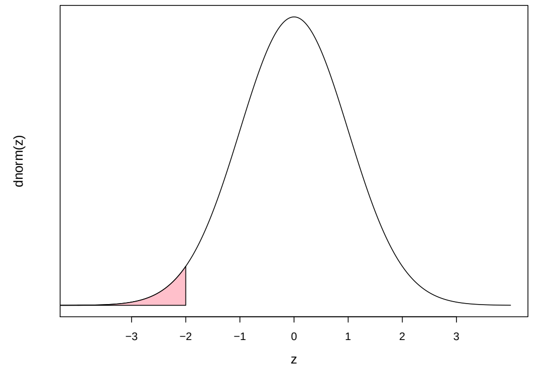
```
pnorm(-2)
```


Example 2

Answer:

```
pnorm(-2)
```

[1] 0.023



Example 2

Question b):

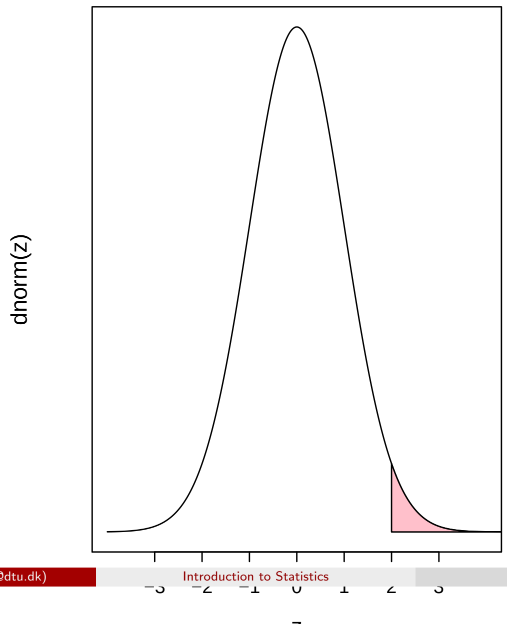
What is the proba ... nuch?

Answer:

$P(Z \geq 2) = 0.022$

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

[1] 0.023



Example 2

Question c):

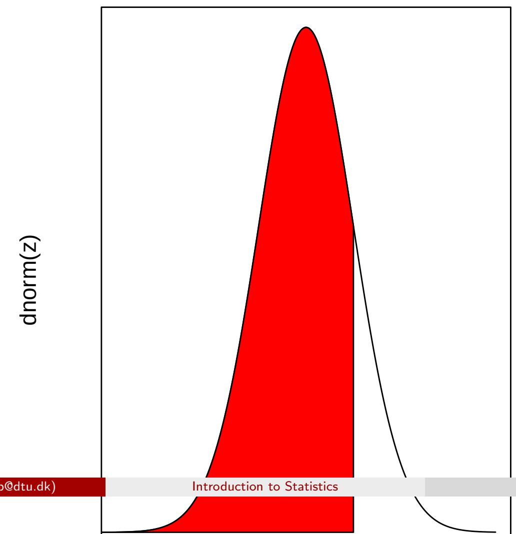
What is the probability that the weight measures at most ± 1 gram wrong?

Answer:

$P(|Z| \leq 1) = P(-$

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

[1] 0.68



Example 2

Question c):

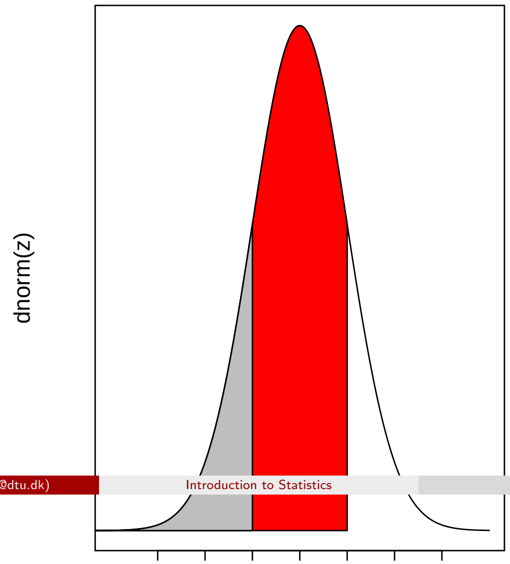
What is the probability that the weight measures at most ± 1 gram wrong?

Answer:

$P(|Z| \leq 1) = P(-$

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

[1] 0.68



Example 3

Indkomstfordeling:

It is assumed that among a group af elementary school teachers, the salary distribution can be described as a normal distribution with mean $\mu = 280.000$ and standard deviation $\sigma = 10.000$..

Question a):

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

Answer:

$P(X > 300) = P(Z > \frac{300-280}{10}) = P(Z > 2) = 0.023$

$$X \sim N(300, 10^2) \Rightarrow Z = \frac{X - 280}{10} \sim N(0, 1^2)$$

Example 3

Question a):

What is the proba ... *an 300.000?*

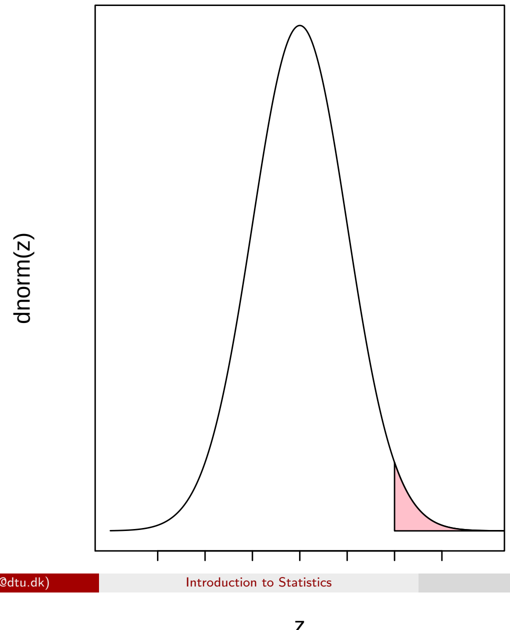
Answer:

```
1-pnorm(300, m
```

[1] 0.023

```
1-pnorm((300-28
```

[1] 0.023



Example 4

A more narrow distribution:

It is assumed that among a group af elementary school teachers, the salary distribution can be described as a normal distribution with mean $\mu = 290.000$ and standard deviation $\sigma = 4.000$.

Question a):

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

Example 4

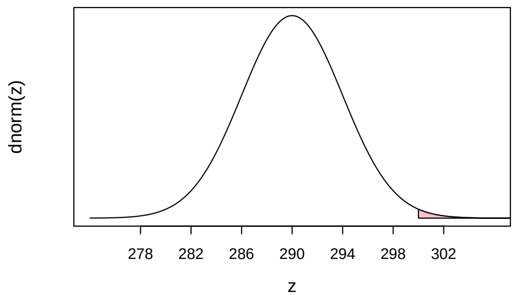
Question a):

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

Answer:

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

[1] 0.0062



Example 5

Same income distribution

It is assumed that among a group af elementary school teachers, the salary distribution can be described as a normal distribution with mean $\mu = 290.000$ and standard deviation $\sigma = 4.000$.

"Opposite question"

Give the salary interval that covers 95% of all teachers' salary

Answer:

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

[1] 282 298

The Log-Normal distribution

Syntax:

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

Density function:

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

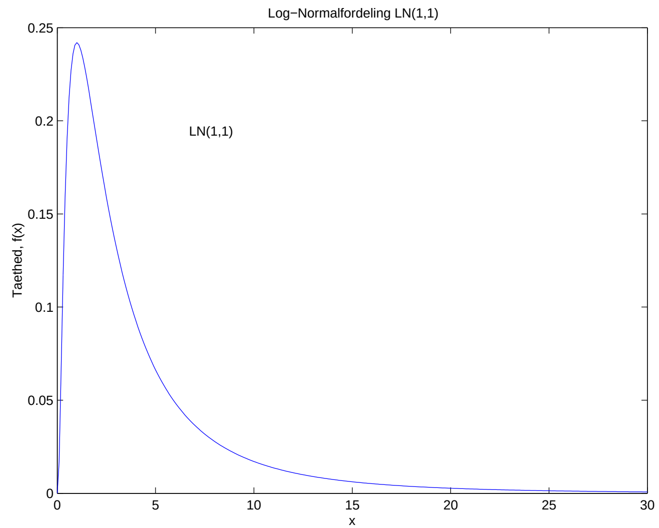
Mean:

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

Variance:

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

The Log-Normal distribution



The Log-Normal distribution

Log-normal and Normal distributions:

A log-normally distributed variable $Y \sim LN(\alpha, \beta)$ can be transformed into a standard normally distributed variable X by:

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

dvs.

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

Continuous distributions in R

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

- `d` $(f(x))$ probability density function.
- `p` $(F(x))$cumulative distribution function.
- `q` Quantile in distribution.
- `r` Random numbers from distribution

Oversigt

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables

The Exponential Distribution

Density function

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

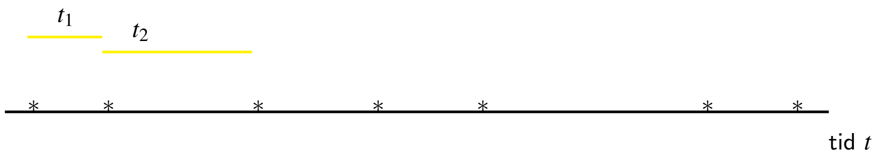
The Exponential Distribution

- 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
- Mean $\mu = \beta$
- Variance $\sigma^2 = \beta^2$

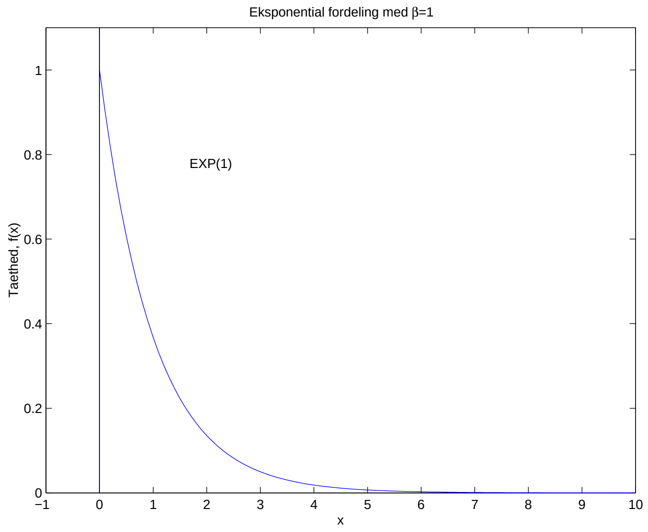
Connection between the Exponetial- and Poisson Distribution

Poisson: Discrete events pr./ unit

Exponential: Continuous distance between events



The Exponential Distribution



Example 6

Qeuing model - poisson proces

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

Question:

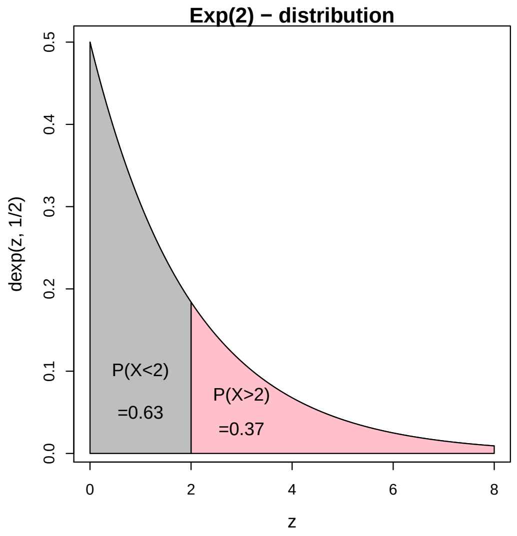
One customer is just arrived. What is the probability that no other costumers will arrive in the next period of 2 minutes?

Answer:

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

```
[1] 0.37
```

Example 6



Example 6

```
z=seq(0,8,by=0.01)

plot(z,dexp(z, 1/2),type = "l", main = "Exp(2) - distribution")

polygon(c(2, seq(2, 8, by = 0.01), 8, 2),
        c(0, dexp(seq(2, 8, by = 0.01), 1/2), 0, 0),
        col = "pink")

text(3,0.07,"P(X>2)")
text(3,0.03,"=0.37")

polygon(c(2, seq(2, 0, by = -0.01), 0, 2),
        c(0, dexp(seq(2, 0, by =- 0.01), 1/2), 0, 0),
        col = "grey")

text(1,0.1,"P(X<2)")
text(1,0.05,"=0.63")
```


Example 7

Question:

One customer is just arrived. Using the Poisson distribution, calculate the probability that no other costumers will arrive in the next period of 2

Answer:

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

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

[1] 0.37

```
exp(-1)
```

[1] 0.37

Example 8

Other time periods:

The time between customer arrivals at a post office is exponentially distributed with mean $\mu = 2$ minutes. Now consider a period of 10 minutes

Question:

Using the Poisson distribution, calculate the probability that no other costumers will arrive in this period

Answer:

$\lambda_{10min} = 5,\ P(X=0) = \frac{e^{-5}}{1!}5^0 = e^{-5}$

```
dpois(0,5)
```

[1] 0.0067

Oversigt

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables

Calculation Rules for Random Variables

(Holds for AS WELL continuous as discrete variables)

X is a random variable

. We assume that a and b are constants.Then we have:

Mean rule:

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

Variance rule:

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

Example 9

X is a random variable

. A random variable X has mean 4 and variance 6.

Question:

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

Answer:

$$\mathrm{E}(Y)=-3\mathrm{E}(X)+2=-3\cdot 4+2=-10$$
$$\mathrm{Var}(Y)=(-3)^2\mathrm{Var}(X)=9\cdot 6=54$$

Calculation Rules for Random Variables

$X_1,\ldots,X_n$ are random variables

Then (when independent)

Mean rule:

$$E(a_1X_1+a_2X_2+..+a_nX_n)$$
$$=a_1E(X_1)+a_2E(X_2)+..+a_nE(X_n)$$

Variance rule::

$$Var(a_1X_1+a_2X_2+..+a_nX_n)$$
$$=a_1^2Var(X_1)+..+a_n^2Var(X_n)$$

Example 10

Airline Planning

The weight of the passengers on a flight is assumed Normal distributed $X \sim N(70, 10^2)$.

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

Question:

Calculate the probability that the plain is overloaded

What is Y=Total passenger weight?

What is Y?

Definitely NOT: $Y = 55 \cdot X$!!!!!!

Example 10

What is Y=Total passenger weight?

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

Mean and variance of Y:

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

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

We use a normal distribution for Y:

```
1-pnorm(4000, m = 3850, s = sqrt(5500))
```

[1] 0.022

Example 10 - WRONG ANALYSIS

What is Y?

Definitely NOT: $Y = 55 \cdot X$!!!!!!

Mean and variance of Y:

$$E(Y) = 55 \cdot 70 = 3850$$

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

We use a normal distribution for Y:

```
1-pnorm(4000, m = 3850, s = 550)
```

[1] 0.39

Consequence of wrong calculation:

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

Agenda

1. Continuous random variables and distributions
   - The Density Function
   - Distribution function
   - The Mean of a Continuous Random Variable
   - The Variance of a Continuous Random Variable
   - The Covariance of two random variables
2. Specific Statistical Distributions
   - The Uniform Distribution
   - The Normal Distribution
   - The Log-Normal distribution
3. The Exponential Distribution
4. Calculation Rules for Random Variables