



University of St.Gallen



Methods: Statistics (4,120)

5. Continuous, Normal and Standard Normal Distribution

Spring 2022

Prof. Dr. Roland Füss

Swiss Institute of Banking and Finance (s/bf)



Contents

- I. Density Function of Continuous Random Variables**
- II. Measures of Central Tendency and Dispersion**
- III. Normal Distribution and Standard Normal Distribution**
- IV. Approximation Binomial Distribution**
- V. Exponential Distribution**

Learning Objectives

After this lecture, you know how:

- the **expected value** and the **variance** for continuous distributions can be calculated and interpreted.
- the normal distribution is applied in **business context** and why it is of central importance.
- a normal distribution is standardized and why the **standard normal distribution** is important.

Literature

Levine, D.M., K. A. Szabat, and D.F. Stephan. (2016). *Business Statistics: A First Course*, 7th ed. United States: Pearson, **Chapter 6.***

Stinerock, R. (2018). *Statistics with R*. United Kingdom: Sage. **Chapter 5.***

Shira, Joseph (2012). *Statistische Methoden der VWL und BWL*, 4th ed. Munich et al.: Pearson Studium, **Chapter 11**.

Weiers, R. M. (2011). *Introductory Business Statistics*, 7th ed., Canada: Thomson South-Western, **Chapter 7**.

*Mandatory literature

I. Density Function of Continuous Random Variables

In case of **continuous random** variables the **probability function** is replaced by the so-called density function. The **density function** $f(x)$ of a continuous random variable X is a continuous function for which holds:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 \quad \text{und} \quad f(x) \geq 0.$$

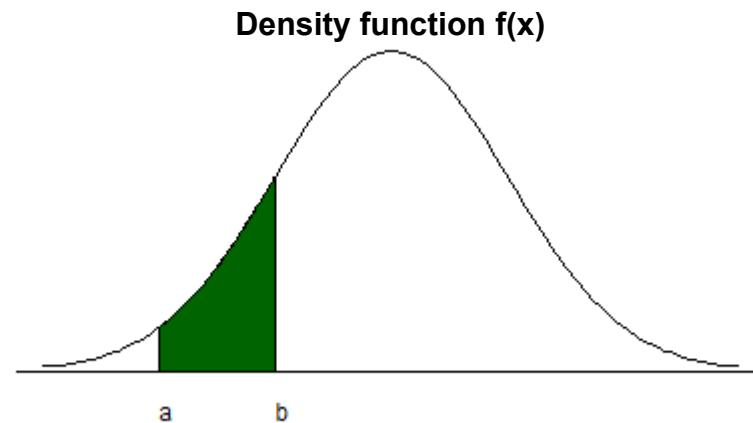
Thus, the density function has the following properties:

1. The density function cannot be **negative**.
2. The area between the density function and the x -axis **always equals 1**.

I. Density Function of Continuous Random Variables

Probability that X takes a value which lies in the **interval $[a, b]$** corresponds to the area below the density function within the boundaries a and b :

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f(x) \cdot dx \\ &= F(b) - F(a) \end{aligned}$$



$F(a) = P(X \leq a) = \int_{-\infty}^a f(x) dx$ is the **cumulative distribution function**.

I. Density Function of Continuous Random Variables

Note: The density of a continuous random variable X does **NOT** state the probability that the random variable **takes a certain value x !** This can also be seen from the fact that the density can take larger values than 1. In particular, for a single value x it holds:

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}$$

although $X = x$ is not an impossible event!

The area below the density function over a specific interval corresponds to the probability that the random variable **takes a value within this interval**. In this context, it is irrelevant if the boundaries a and b are included in the interval or not. Formally stated:

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

IV. Density Function (R-Example 1)

Open the file "L5-Example_1.R" in R-Studio and reproduce the R-Code.

```
# graphically represent the probability that x takes on a value in the interval [-2, -1]
# by coloring the area under the density function and the cumulative distribution function.

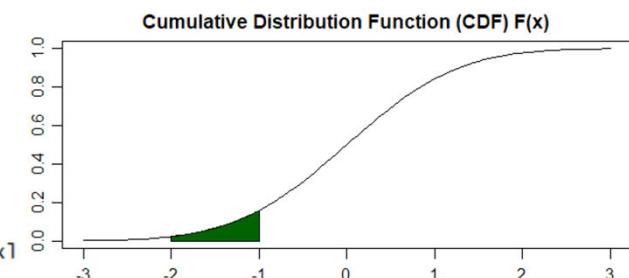
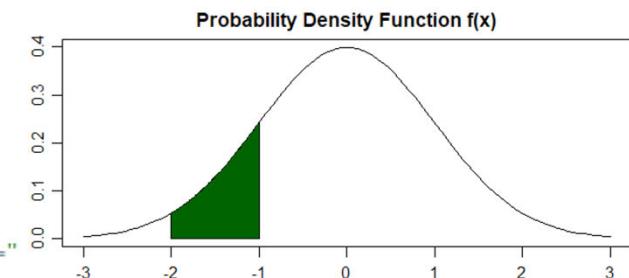
# for coloring an interval, the fields package is needed.
# install.packages('fields')
library(fields)

# probability density function
#-----
# Drawing the density function f(x) with mean value 0 and SDW 1.
op <- par(mfrow = c(2, 1), mgp = c(1.5, 0.8, 0), mar = .1+c(3,3,2,1))
curve(dnorm(x,0,1), xlim=c(-3,3),ylim=c(0,0.4), main="Probability Density Function f(x)",ylab="")

# draw the interval [-2, -1] in color:
from.z <- -2
to.z <- -1
s.x <- c(from.z, seq(from.z, to.z, 0.01), to.z)
s.y <- c(0, dnorm(seq(from.z, to.z, 0.01)), 0)
polygon(s.x,s.y, col="darkgreen")

# cumulative distribution function
#-----
# drawing the cumulative distribution function F(x) with mean value 0 and SDW 1.
curve(pnorm(x,0,1), xlim=c(-3,3), main="Cumulative Distribution Function (CDF) F(x)",ylab="",x1

# draw the interval [-2, -1] in color:
from.z <- -2
to.z <- -1
s.x <- c(from.z, seq(from.z, to.z, 0.01), to.z)
s.y <- c(0, pnorm(seq(from.z, to.z, 0.01)), 0)
polygon(s.x,s.y, col="darkgreen")
par(op)
```



II. Measures of Central Tendency and Dispersion

The characterization of density distributions of continuous random variables is done by the following measures:

Expected value:

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

Variance

$$Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot f(x) dx$$

Simplified:

$$Var(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx - [E(X)]^2 = E(X^2) - [E(X)]^2$$

III. Normal Distribution - $N(\mu, \sigma^2)$

The normal distribution is one of the **most important** continuous distributions, since many random variables occurring in practice are - at least approximately - normally distributed. A continuous random variable X with the density function:

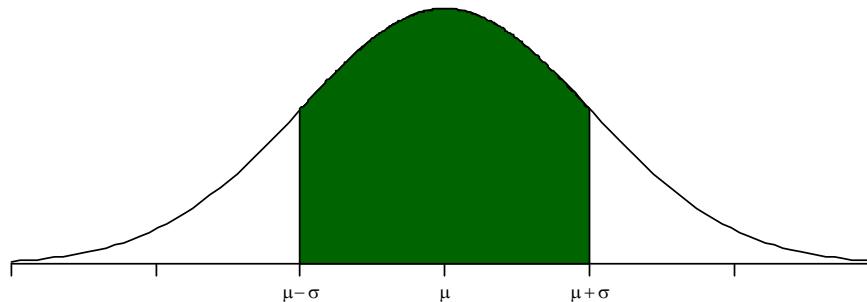
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

is normally distributed with the parameters μ and σ^2 . Many distributions (e.g., binomial distribution) can be **approximated** by the normal distribution.

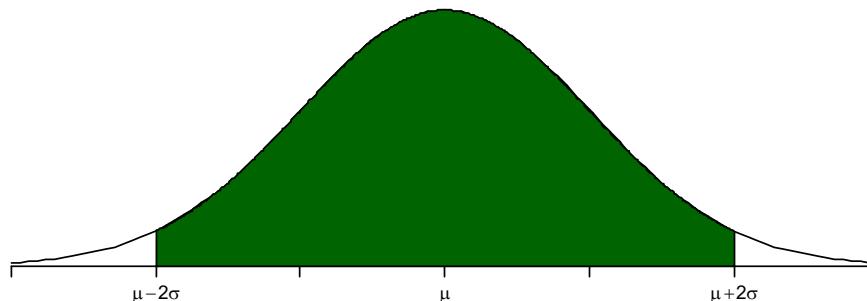
Sample means as well as proportions of repeated samples **tend** to be normally distributed (central limit theorem).

III. Normal Distribution - $N(\mu, \sigma^2)$

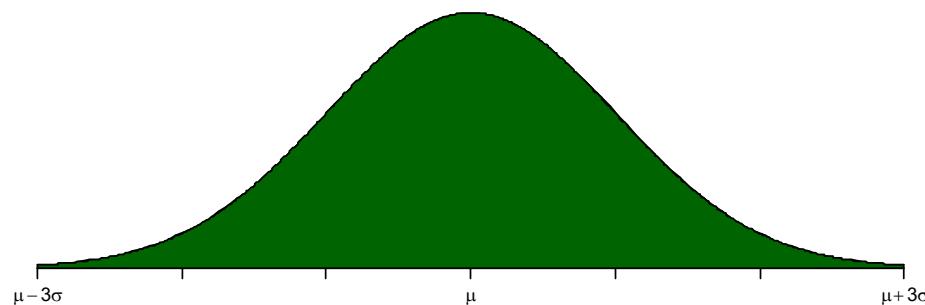
68.3% of observations within 1 standard deviation around mean



95.5% of observations within 2 standard deviation around mean



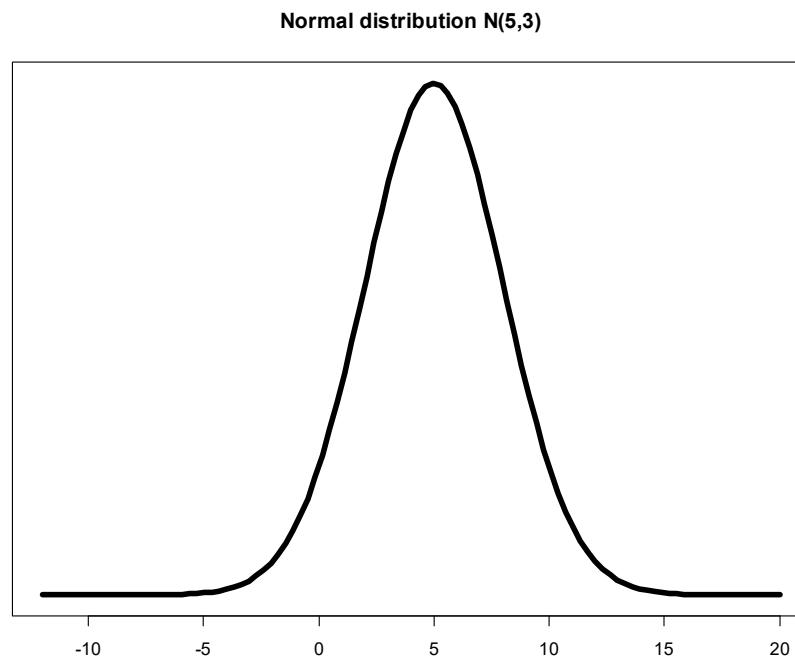
99.7% of observations within 3 standard deviation around mean



III. Normal Distribution (Example)

Normal distribution with expected value $\mu = 5$ and standard deviation of $\sigma = 3$:

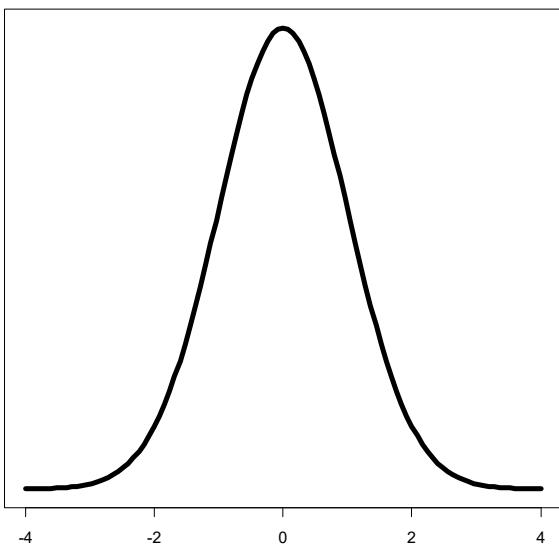
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



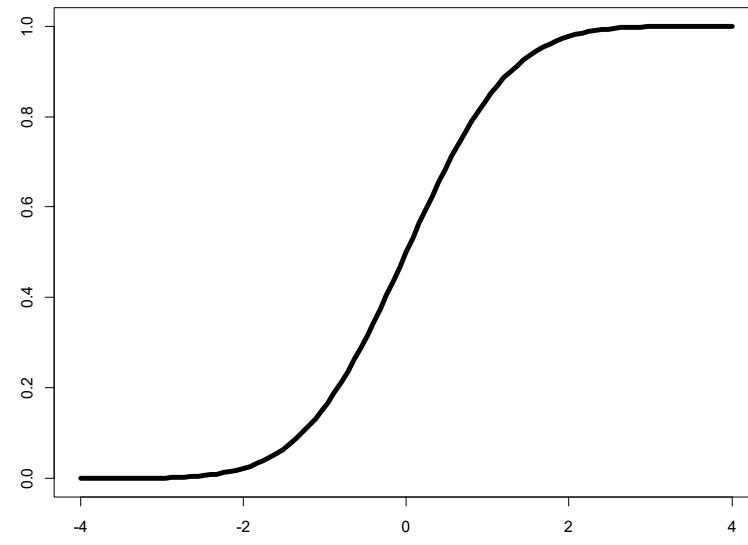
III. Standard Normal Distribution - N(0,1)

For $\mu = 0$ and $\sigma^2 = 1$, the standard normal distribution is obtained:

Standard normal distribution



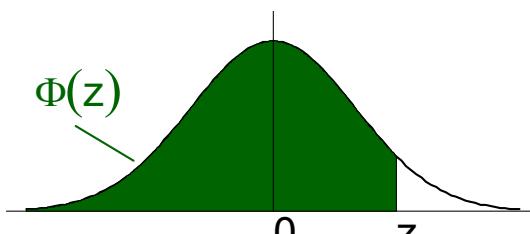
Standard normal CDF (F(x))



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

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

III. Standard Normal Distribution - N(0,1)



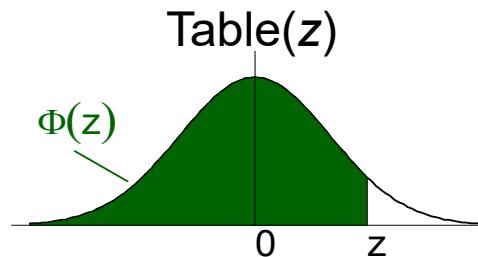
$$\int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

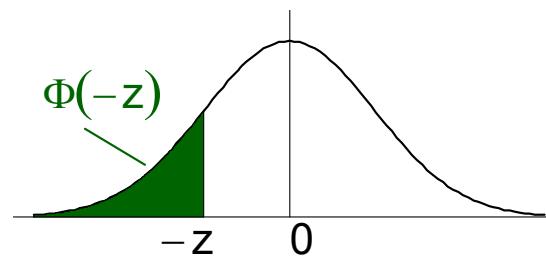
III. Standard Normal Distribution - $N(0,1)$

Using distribution tables:

- $P(Z \leq z) = \Phi(z) = \text{Table}(z)$



- $P(Z \leq -z) = \Phi(-z) = 1 - \text{Table}(z)$



- "Table(z)" represents values listed in the table on the previous slide.

III. Standard Normal Distribution - N(0,1) (Examples)

1. By means of the previously mentioned "tables" for the standard normal distribution, we are searching the following probability:

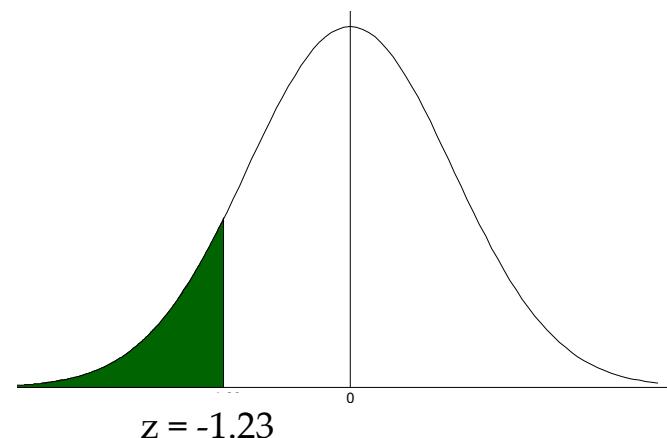
$$P(1.00 \leq z \leq 1.32) ; z \sim N(0; 1)$$

The probability results from the subtraction of the corresponding areas:

$$\rightarrow 0.9066 - 0.8413 = 0.0653$$

2. Search the value $z = -1.23$. What probability does it reflect?

$$\Phi(-1.23) = 1 - 0.8907 = 0.1093$$



III. Calculation of Probabilities

For the standardized normal distribution (short: $\Phi(z)$) there are "tables" which can be used to obtain "probability-areas" of standard normally distributed random variables. Each normally distributed random variable X with the parameters μ and σ^2 can be transformed into a standard normally distributed random variable Z (**standardization**):

$$X \sim N(\mu, \sigma^2) \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

The distribution function F of the random variable X is expressed by the distribution function Φ of the standard normal distribution:

$$F(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

III. Calculation of Probabilities

For the determination of the probability that a $N(\mu, \sigma^2)$ -distributed random variable X takes a value between a and b , we make use of the “**standardization**”:

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) = F(b) - F(a) \\ &= \Phi(z_b) - \Phi(z_a) \end{aligned}$$

with:

$$z_b = \frac{b - \mu}{\sigma}; \quad z_a = \frac{a - \mu}{\sigma}$$

III. Normal Distribution (Example)

Work process: Let the average (normally distributed) duration of a certain operation be 12.1 minutes with a standard deviation of 2 minutes. What is the probability that a worker needs:

- longer than 14.1 minutes,
- less than 8.1 minutes,
- between 10.1 and 16.1 minutes?

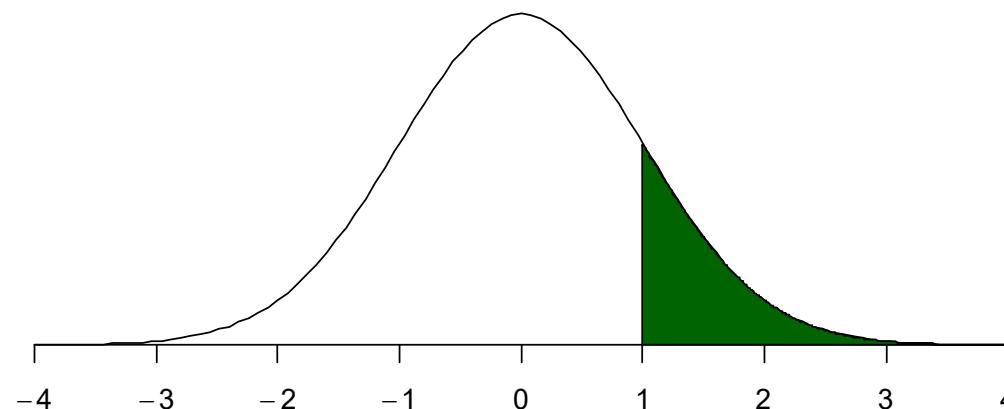
III. Normal Distribution (Example)

Given: $\mu = 12.1$ minutes, $\sigma = 2.0$ minutes

Wanted: Probability of «time > 14.1 minutes»?

$$z = \frac{x - \mu}{\sigma} = \frac{14.1 - 12.1}{2.0} = 1.00$$

$$\begin{aligned}\Leftrightarrow P(x > 14.1) &= P(z > 1.00) \\ &= 1 - 0.8413 = \mathbf{0.1587}\end{aligned}$$



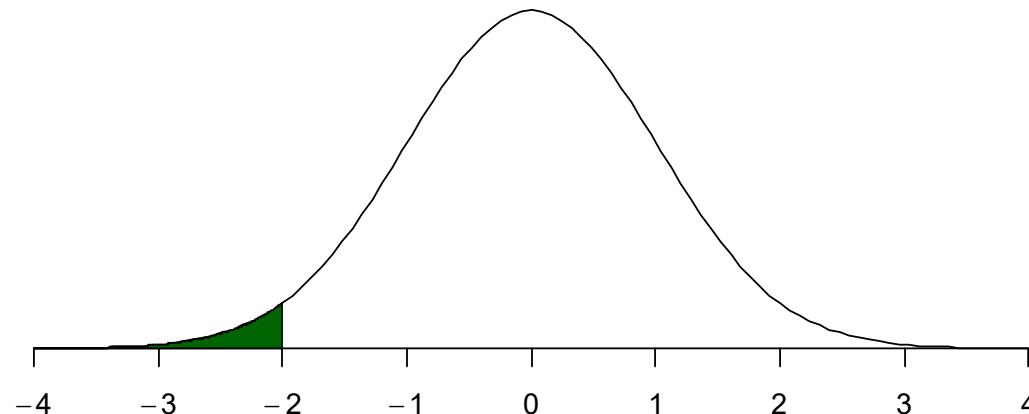
III. Normal Distribution (Example)

Given: $\mu = 12.1$ minutes, $\sigma = 2.0$ minutes

Wanted: probability of «time < 8.1 minutes»?

$$z = \frac{x - \mu}{\sigma} = \frac{8.1 - 12.1}{2.0} = -2.00$$

$$\begin{aligned}\Leftrightarrow P(x < 8.1) &= P(z < -2.00) \\ &= 1 - 0.9772 = \mathbf{0.0228}\end{aligned}$$



III. Normal Distribution (Example)

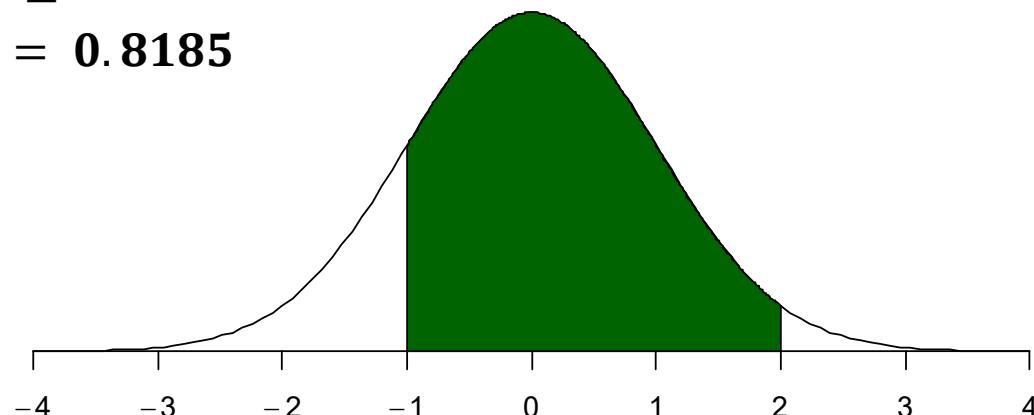
Given: $\mu = 12.1$ minutes, $\sigma = 2.0$ minutes

Wanted: probability of « $10.1 < \text{time} < 16.1$ minutes»?

$$z_a = \frac{x - \mu}{\sigma} = \frac{10.1 - 12.1}{2.0} = -1.00$$

$$z_b = \frac{x - \mu}{\sigma} = \frac{16.1 - 12.1}{2.0} = 2.00$$

$$\begin{aligned}\Leftrightarrow P(10.1 < x < 16.1) &= \\ &= P(-1.00 < z < 2.00) \\ &= 0.9772 - (1 - 0.8413) = \mathbf{0.8185}\end{aligned}$$



IV. Normal Distribution (R-Example 2)

Open the file "L5-Example_2.R" in R-Studio and reproduce the R-Code.

```
# example: IQ test
#-----
# a random experiment consists of having a person take an IQ test.
# the random variable x is given by the score on the test. The points are
# standardized with expected value 100 and standard deviation of 15.

# a) What is the probability P(85<=x<=115)?
# The limits 85 and 115 fall exactly on one standard deviation.
# By the expected value of 100 (15 above and below each).
# The probability is therefore about 68% (chebychev).
z1<-(85-100)/15
z2<-(115-100)/15
pnorm(z2)-pnorm(z1)

# b) what is the lowest possible IQ score that a person can have to
# still be in the top 1% of all IQ scores?
qnorm(0.99,mean=100,sd=15) # qx(0.99) quantile-function

# c) Find the z values 0.025, 0.01, and 0.005
# (these will be important for estimations and hypothesis testing)
qnorm(c(0.025,0.01,0.005),lower.tail=FALSE) # !!!!!!
qnorm(c(0.975,0.99,0.995)) # same result

# example: z-values
#-----
# Given is a normally distributed random variable z with N(mean=0,sd=1):
# a) P(z>2.64)
pnorm(2.64,lower.tail=F)

# b) P(0<=z<=0.87)
pnorm(0.87)-1/2
```

IV. Approximation Binomial Distribution

An important characteristic of the normal distribution is that it can be used to approximate other distributions. For instance, the binomial distribution

$$f(x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

approaches the form of a normal distribution with growing n . A good approximation results, when the conditions

$$n\pi \geq 5 \text{ and } n(1 - \pi) \geq 5$$

are fulfilled.

Expected value: $\mu = n\pi$

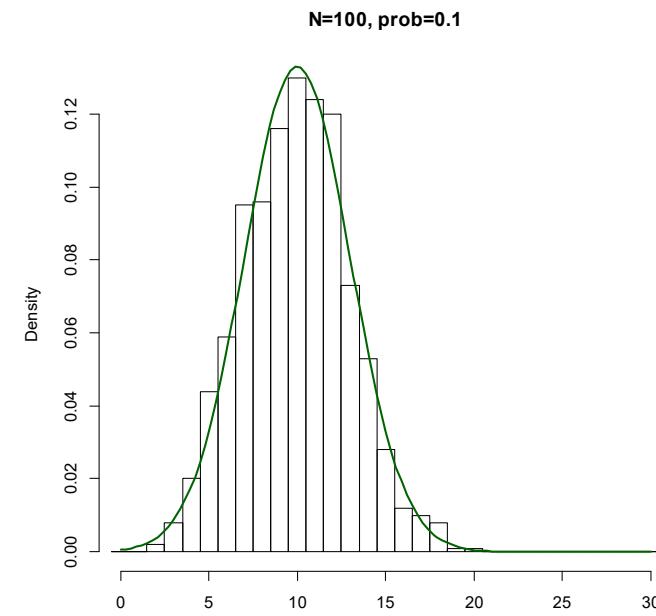
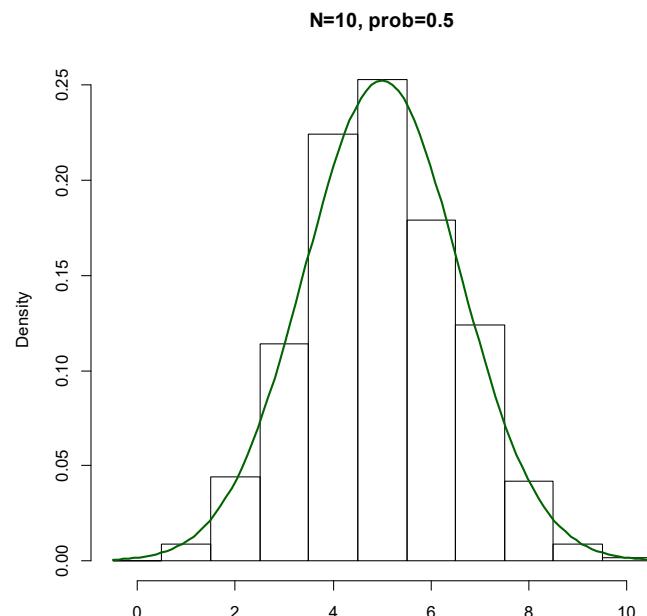
Variance: $\sigma^2 = n\pi(1 - \pi)$

IV. Approximation Binomial Distribution

Binomially distributed random variable r : $r \sim B(n, p)$

Expected value $\mu = np$, variance $\sigma^2 = np(1-p)$

For $n \rightarrow \infty$: binomial distribution converges to the normal distribution with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$



IV. Approximation Binomial Distribution (Example)

Exam: 40 students take an exam. The probability to pass the exam is 60%. What is the probability that at least 30 students pass the exam?

Option 1: Binomial distribution

$$p = p(30) + p(31) + \dots + p(40)$$

$$p(x) = \binom{40}{x} \cdot 0.6^x \cdot 0.4^{40-x}$$

→ extensive calculation yields: $p = 3.52\%$

IV. Approximation Binomial Distribution (Example)

Option 2: Approximation with normal distribution!

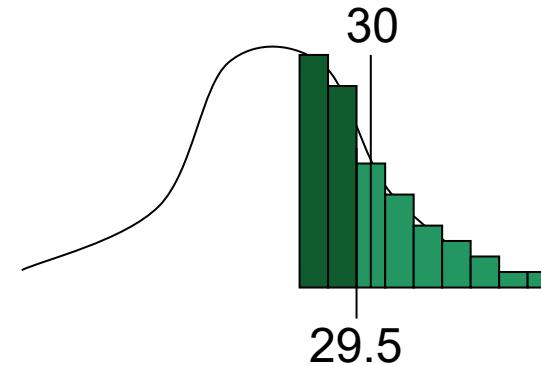
$$(n\pi = 24 \geq 5, n(1 - \pi) = 16 \geq 5)$$

$$X \sim N(\mu = n\pi, \sigma^2 = n\pi(1 - \pi)), n = 40, \pi = 0.6$$

$$X \sim N(\mu = 24, \sigma^2 = 9.6)$$

$$z = \frac{29.5 - 24}{\sqrt{9.6}} = 1.78$$

→ from the normal distribution table we get: $p = 3.75\%$.



Since the binomial distribution is discrete, the approximation is improved by the so-called **continuity correction**:

$$x \rightarrow [x - 0.5]; [x + 0.5]$$

V. Exponential Distribution

The exponential distribution describes the continuous **random variable x = the amount of time, space, or distance between occurrences of these rare events** (e.g., x is the time between successive arrivals). The cumulative distribution and density function are:

$$F(x) = 1 - e^{-\lambda x} \quad f(x) = \lambda e^{-\lambda x} \quad \text{für } \lambda > 0, x \geq 0$$

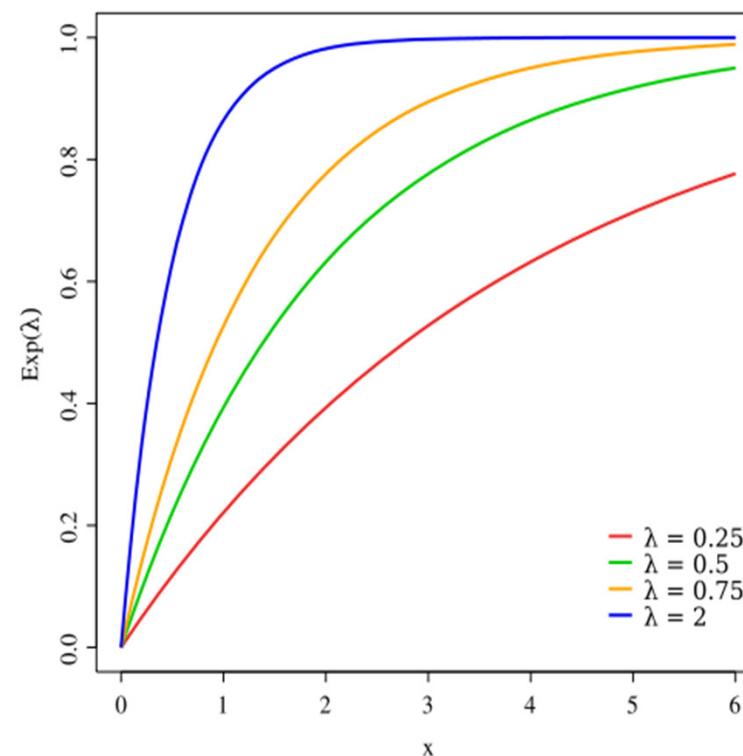
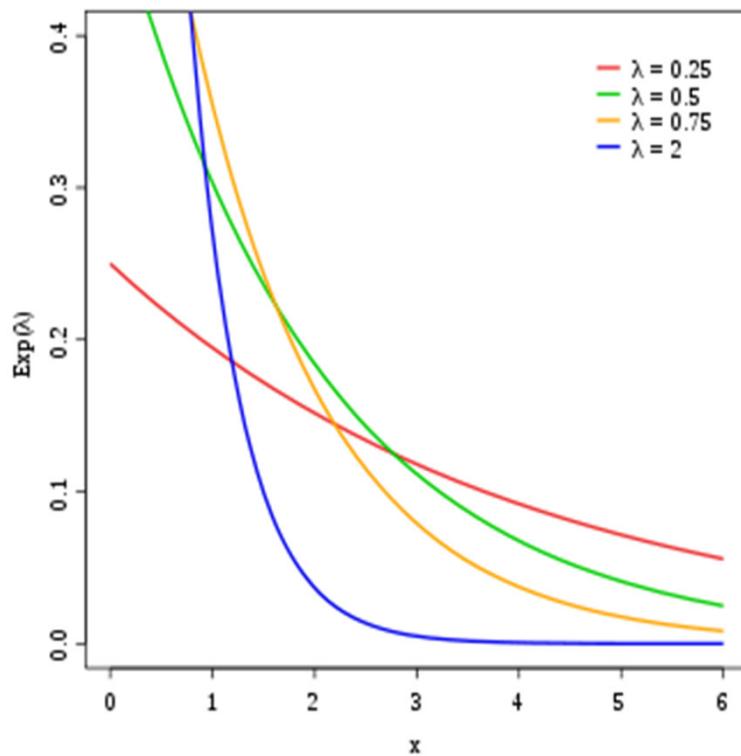
Mean and variance of the exponential distribution are:

$$\text{Expected value} = \frac{1}{\lambda}$$

$$\text{Variance} = \frac{1}{\lambda^2}$$

V. Exponential Distribution

Density and distribution function for different values of the parameter λ :



V. Exponential distribution (example)

Bank counter: Assume the time span (in minutes) between the arrival of two bank customers at the counter is exponentially distributed with $\lambda = 0.4$. What is the probability that more than 2 minutes pass between the arrival of two customers?

$$\begin{aligned}P(X > 2) &= 1 - P(X \leq 2) \\&= 1 - (1 - e^{-0.4 \cdot 2}) \\&= e^{-0.8} = 0.4493\end{aligned}$$

V. Exponential Distribution (R-Example)

Open the file "L5-Example_3.R" in R-Studio and reproduce the R-Code.

```
# example: call center
#-----
# the time between two calls in a call center is exponentially distributed
# with lambda = 3, so the mean is 1/3 and the variance is 1/3^2.

# What is the probability that the time span is less than 2 minutes?
pexp(2,1/3)

# What is the probability that the time span is less than 5 minutes?
pexp(5,1/3)

# What is the probability that the time span is more than 3 minutes?
1-pexp(3,1/3)

# What is the probability that the time span is between 2 and 5 minutes?
pexp(5,1/3) - pexp(2, 1/3)
```