Bus 273:
Statistical Analysis For Business

Harald Schmidbauer

İSTANBUL BİLGİ ÜNİVERSİTESİ

About These Slides

- The present slides are not self-contained; they need to be explained and discussed. This will be done in the lectures.

- Even though being a "work in progress" and subject to revision, the slides constitute copyrighted material.
  If you want to reproduce or copy anything from the slides, please ask:

Harald Schmidbauer	**harald** at **hs-stat** dot **com**
Angi Rösch	**angi.r** at **t-online** dot **de**

- The slides were produced using LaTeX and R (the R project; www.R-project.org) on a GNU/Linux system.

- R files used for this course are available upon request.

Chapter 7:

Continuous Probability Distributions

7.1 Basics

Continuous random variables and continuous distributions.

- A random variable is called continuous if it can take on any value in a certain interval.

- The distribution of a continuous random variable is called a continuous distribution.

Examples:

- X = daily return on DAX
- X = waiting time of a customer at a call center until an incoming call is answered
- X = impurity of a chemical produced by a company
- ...

7.1 Basics

Computing probabilities.

- For a continuous random variable X,

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

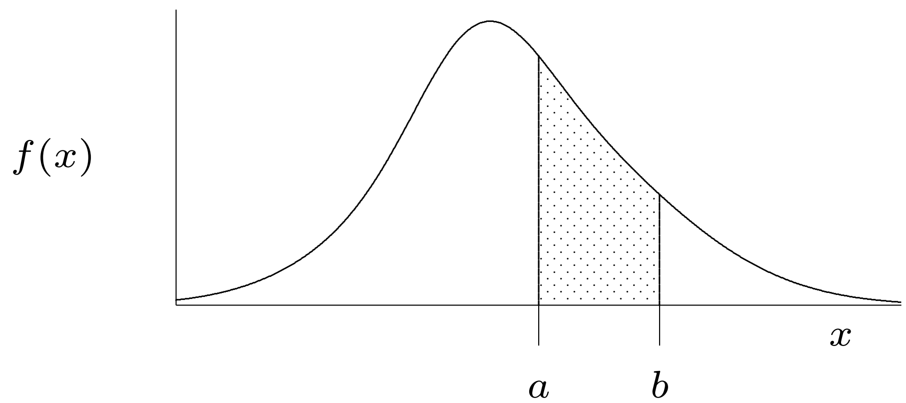
- For example, let X = body-height of a randomly selected person. Then, $P(X = 172) = P(X = 172.0000000) = 0$!

- Therefore, the concept of probability function doesn't work for a continuous random variable.

- A density is needed!

7.1 Basics

The density.



- Probabilities are areas below the density.
- For example,

$$P(a \le X \le b) = \int_a^b f(x)\mathrm{d}x$$

- A continuous probability distribution is given by these terms.

7.1 Basics

The distribution function.

- The distribution function of X is defined as

$$x \mapsto F(x) = P(X \leq x) = \int_{-\infty}^{x} f(\xi)\mathrm{d}\xi.$$

- With it, $P(a < X \leq b) = F(b) - F(a)$.

- As before, we can
  - display the distribution (here: the density),
  - compute location and variation measures.

7.1 Basics

The expectation: a location measure.

- For a continuous random variable X,

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\mathrm{d}x.$$

- Same principle as in Chapter 6, with
  - $f(x)\mathrm{d}x$ instead of $p_i = P(X = i)$,
  - $\int$ instead of $\sum$.

7.1 Basics

The variance: a variation measure.

- For a discrete random variable X,

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] = \mathrm{E}(X^2) - \mathrm{E}^2(X) \\ &= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \cdot f(x)\mathrm{d}x.\end{aligned}$$

- Same principle as in Chapter 6, with
  - $f(x)\mathrm{d}x$ instead of $p_i = P(X = i)$,
  - $\int$ instead of $\sum$.

7.2 The Normal Distribution

Definition.

- A random variable X with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma^2 > 0,$$

  is said to be normally distributed with parameters μ and σ^2.

- In symbols: $X \sim \mathrm{N}(\mu, \sigma^2)$

7.2 The Normal Distribution

This picture shows the density.

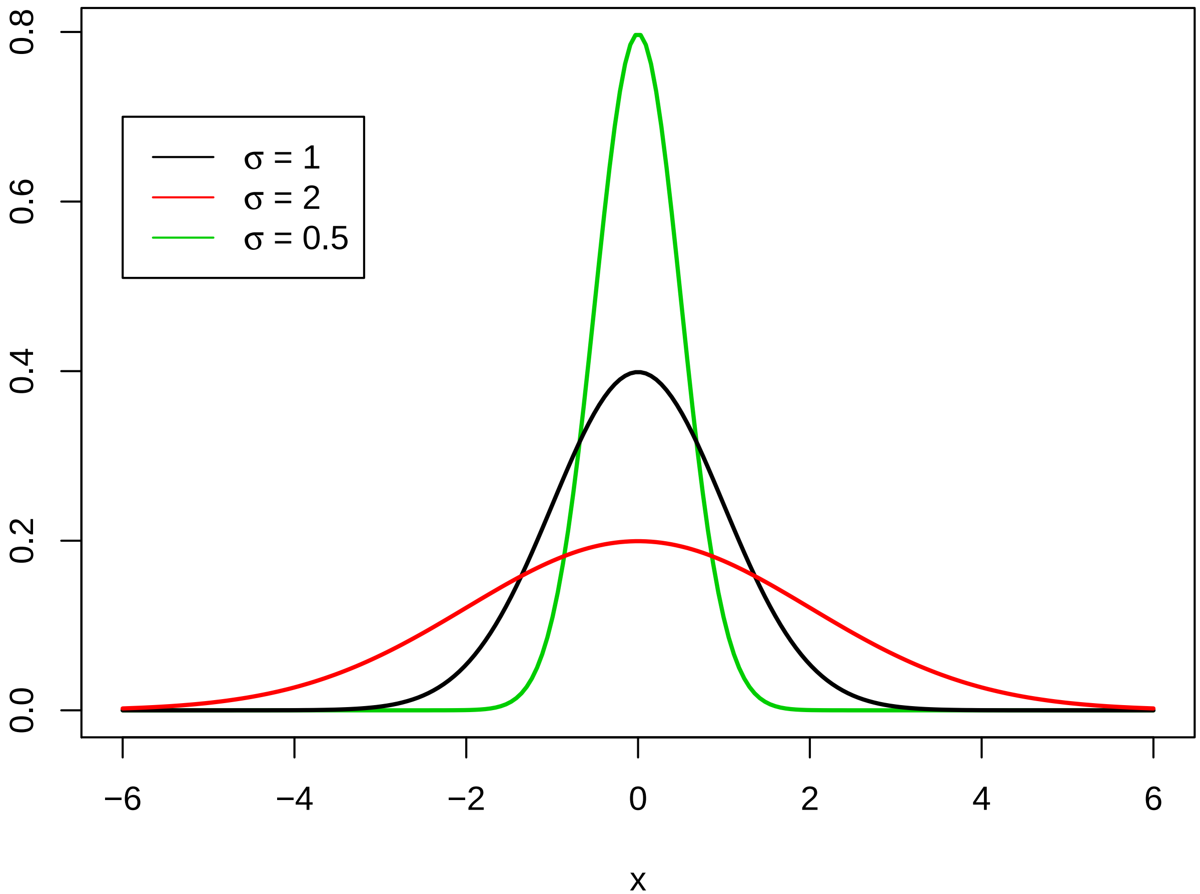
7.2 The Normal Distribution

Meaning of the parameters μ and σ^2.

- Analytical meaning of μ and σ^2:
  - The density has its maximum at $x = \mu$.
  - The inflection points are at $x = \mu \pm \sigma$.
- Statistical meaning of μ and σ^2: If $X \sim \mathsf{N}(\mu, \sigma^2)$, then
  - $\mathrm{E}(X) = \mu$,
  - $\mathrm{var}(X) = \sigma^2$.

7.2 The Normal Distribution

This picture shows the influence of σ^2.



7.2 The Normal Distribution

The standard normal distribution.

- If $X \sim \mathsf{N}(0, 1)$, we say: X has a standard normal distribution.
- Let $X \sim \mathsf{N}(0, 1)$. Using the table (and using geometry):
  - $P(X < 0) = 0.5$
  - $P(X < 1) = 0.8413$
  - $P(X > 1) = 1 - 0.8413 = 0.1587$
  - $P(-1 < X < 1) = 1 - 2 \cdot (1 - 0.8413) = 0.6826$

  and:

$$\boxed{P(-1.96 < X < 1.96) = 0.95}$$

7.2 The Normal Distribution

The normal distribution — standardization.

- There is a table ONLY for the standard normal distribution.
- If $X \sim \mathsf{N}(\mu, \sigma^2)$ with any $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, we can standardize X:

$$Z := \frac{X - \mu}{\sigma} \sim \mathsf{N}(0, 1)$$

- For the new random variable Z, we can again use the standard normal distribution table.

7.2 The Normal Distribution

Example 1: IQ tests.

- Let $X =$ IQ of a randomly selected adult.
- IQ tests are designed such that $X \sim \mathrm{N}(100, 15^2)$.
- What is the probability that a randomly selected person's IQ is. . .
  - higher than 130?
  - less than 90?
  - between 90 and 110?

7.2 The Normal Distribution

Example 1: IQ tests.

- Now let

  $X =$ IQ of a randomly selected student of
  Bilgi University.

- Assuming $X \sim \mathrm{N}(105, 15^2)$:
  - What is the probability that a randomly selected student's IQ is higher than 150?
  - How many of the students of Bilgi University would we expect to have an IQ above 150?

7.2 The Normal Distribution

Example 2: Viscosity.

- Viscosity measurements from a batch chemical process are assumed to be a normal random variable with mean 15 and standard deviation 1.

- What is the probability that a batch has viscosity. . .
  - above 15?
  - above 15.5?
  - above 16?

7.2 The Normal Distribution

Example 3: Body-height.

- Assume that the body-height (in cm) of a randomly selected male person is a random variable $X \sim \mathrm{N}(178, \sigma^2)$.

- Which of the following could be a possible value of σ^2? —

  1 / 25 / 50 / 100 / 250 / 1000 ???

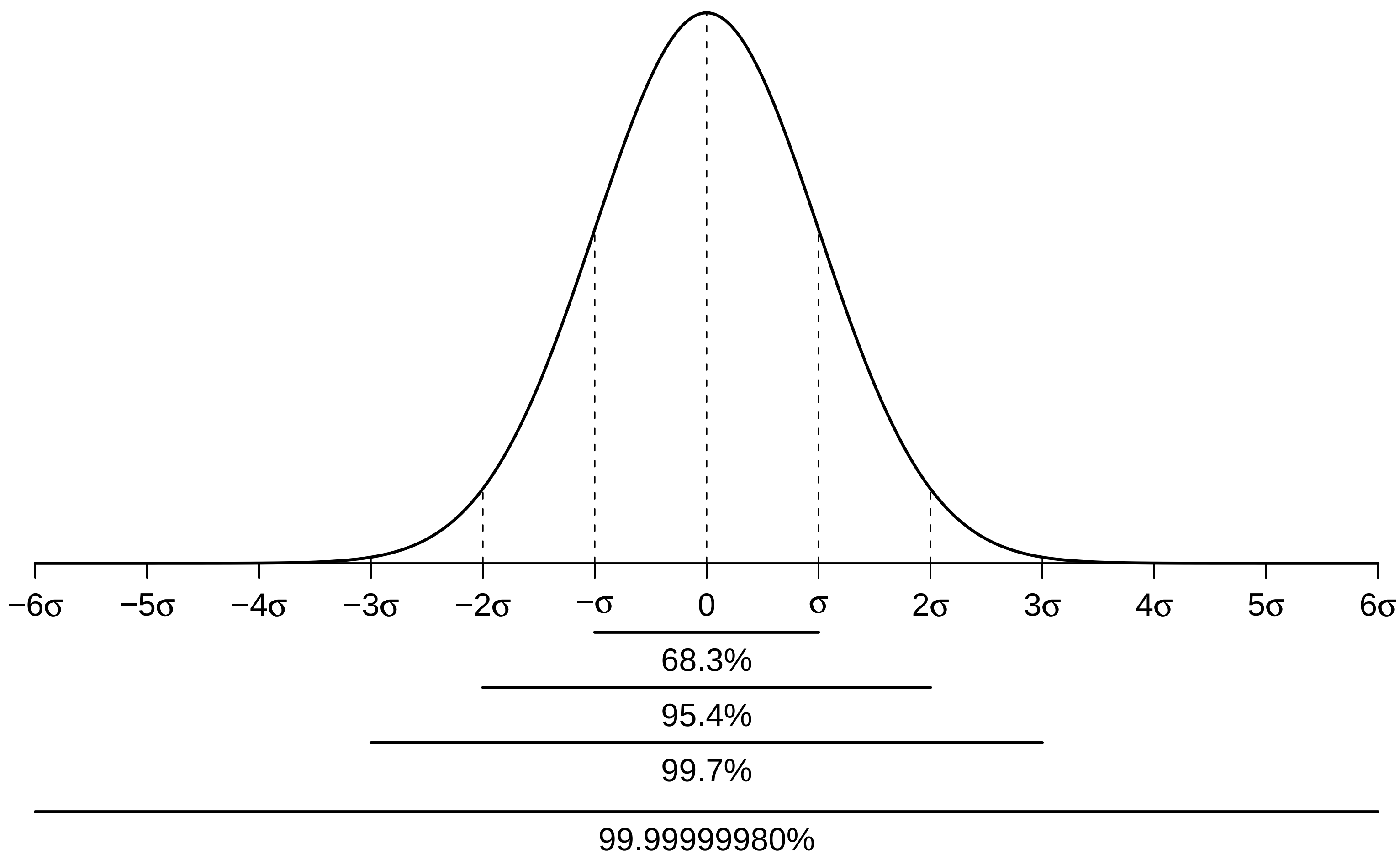
7.2 The Normal Distribution

The sigma rules again.

- If $X \sim \mathsf{N}(\mu, \sigma^2)$:
  - $P(\mu - 1 \cdot \sigma \leq X \leq \mu + 1 \cdot \sigma) \approx 0.68$
  - $P(\mu - 2 \cdot \sigma \leq X \leq \mu + 2 \cdot \sigma) \approx 0.95$
  - $P(\mu - 3 \cdot \sigma \leq X \leq \mu + 3 \cdot \sigma) \approx 0.997$
  - $P(\mu - 6 \cdot \sigma \leq X \leq \mu + 6 \cdot \sigma) \approx 0.9999999980$
- The last rule — the six-σ rule — gave rise to the six-sigma concept in quality management (Motorola, mid-1980s).

7.2 The Normal Distribution

The sigma rules again.



7.2 The Normal Distribution

Further important properties of the normal distribution.

- Let $X \sim \mathsf{N}(\mu, \sigma^2)$. Then, for $a, b \in \mathbb{R}$:

$$aX + b \sim \mathsf{N}(a\mu + b, a^2\sigma^2).$$

- Let $X_1 \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$; X_1, X_2 independent. Then:

$$X_1 + X_2 \sim \mathsf{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

7.2 The Normal Distribution

Further important properties of the normal distribution.

Consequences are:

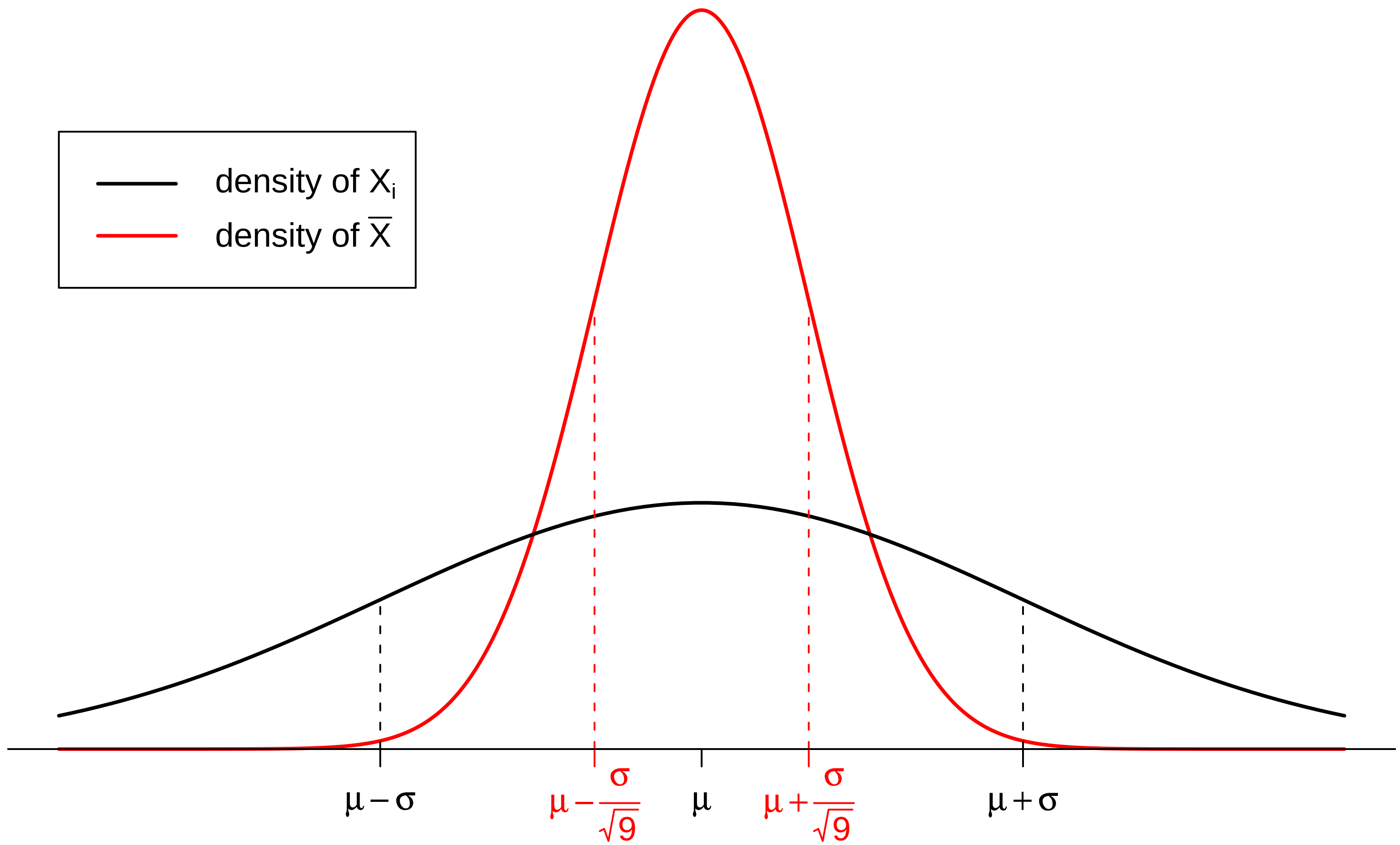
- Let $X_1, \ldots, X_n \sim \mathsf{N}(\mu, \sigma^2)$ and independent. Then:

$$\sum_{i=1}^{n} X_i \;\sim\; \mathsf{N}(n\mu, n\sigma^2),$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\sim\; \mathsf{N}(\mu, \frac{\sigma^2}{n})$$

7.2 The Normal Distribution

Densities of X_i and $\bar{X}$ (with $n = 9$).



7.2 The Normal Distribution

Example 4: Airline passenger weight.

- An airline assumes that the weight (in kg) of a passenger, including carry-on baggage weight, is a random variable

$$X \sim \mathrm{N}(84, 400).$$

- An airplane with 10 seats has a capacity of 1000 kg.
- Compute the probability that this limit is exceeded, i.e. that 10 passengers weigh more than 1000 kg.
- The normality assumption is not really needed. (CLT!)

7.2 The Normal Distribution

Example 5: Average IQ.

- Süleyman Hoca is offering an elective course at Bilgi University.
- Seven students have registered.
- What is the probability that the average IQ of these students is above 115?
  - Assume again that student IQ is normally distributed with mean 105 and standard deviation 15.
  - Also assume that the IQs of the students in this course are independent. (Is this assumption plausible?)

7.3 The Lognormal Distribution

Derivation of the lognormal distribution.

- If $Y \sim \mathsf{N}(\mu, \sigma^2)$, then $X := \mathrm{e}^Y$ is said to have a lognormal distribution with parameters μ and σ^2.

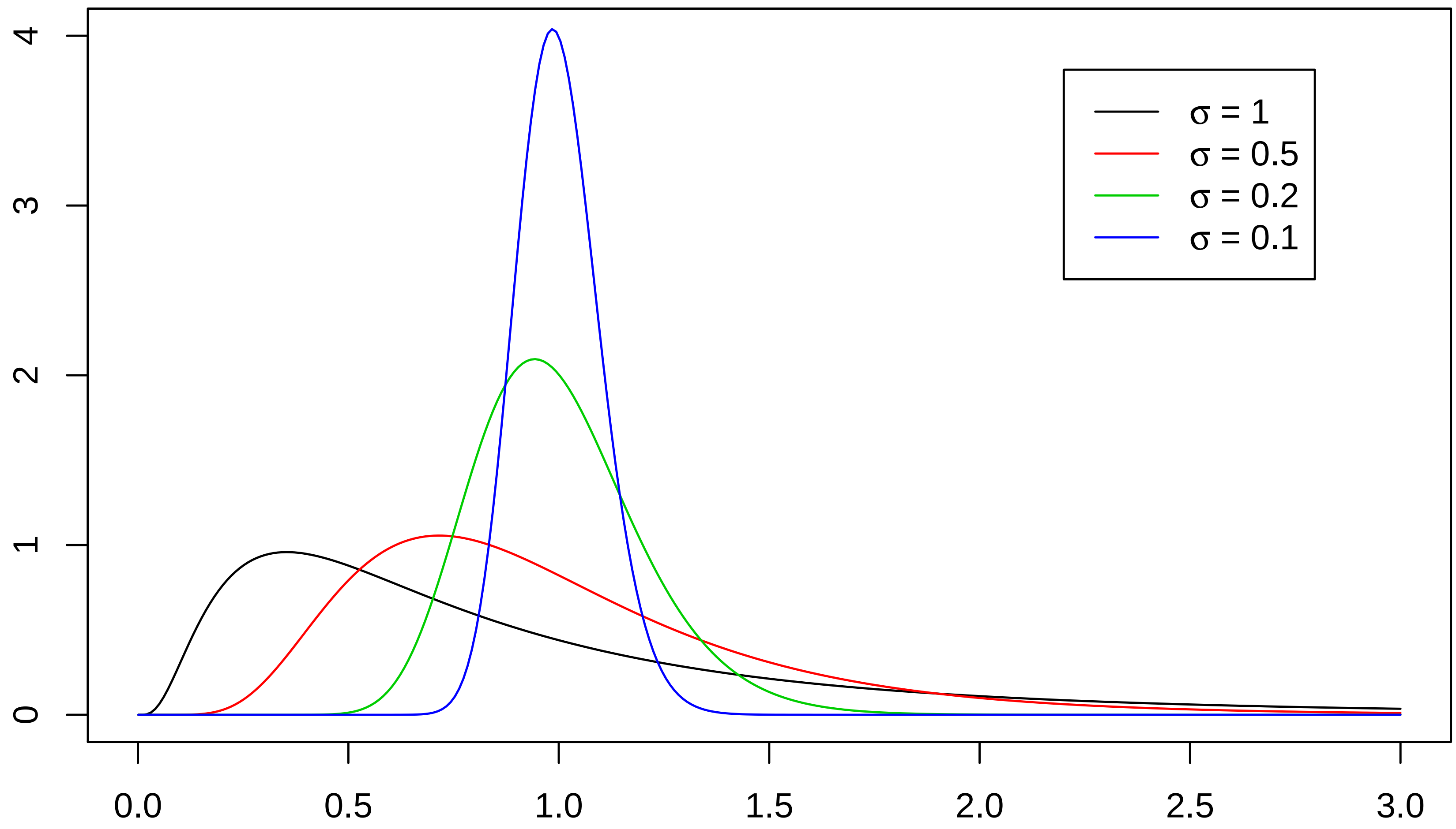
- Its density is:

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \, \mathrm{e}^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0, \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

- In symbols: $X \sim \mathsf{LN}(\mu, \sigma^2)$

7.3 The Lognormal Distribution

This picture shows the density. (Each expectation equals 1.)



7.3 The Lognormal Distribution

Moments of the lognormal distribution.

- Let $X \sim \mathsf{LN}(\mu, \sigma^2)$. Then,

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{e}^{\mu+\sigma^2/2}, \\ \mathrm{var}(X) &= \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2} - 1). \end{aligned}$$

- What is the median of X?

7.3 The Lognormal Distribution

A financial application.

- Define stock prices:

  V_0 : price today (known), V_t : price at time $t > 0$ (unknown)

- Some celebrated models (e.g., the Black-Scholes model for option pricing) assume that

$$V_t = V_0 \cdot X, \quad \text{where} \quad X \sim \mathsf{LN}(\mu, \sigma^2)$$

- We shall see a justification of this assumption in Chapter 8!

- Under the lognormality assumption: $\mathrm{E}(V_t|V_0) = V_0 \cdot \mathrm{e}^{\mu + \frac{\sigma^2}{2}}$. (If $\mu = -\sigma^2/2$, the process (V_t) will be a martingale.)

7.3 The Lognormal Distribution

Example: Stock prices.

- Consider a certain stock, and define stock prices

  V_0 (today), $\quad V_1$ (one year ahead of now).

  Assume: $V_1 = X \cdot V_0$, where $X \sim \text{LN}(0, 0.16)$, $V_0 =$ € 100.

- With this model: What is. . .

  – the probability that the stock price is at least € 130
  – the expected stock price

  after one year?

7.4 The Exponential Distribution

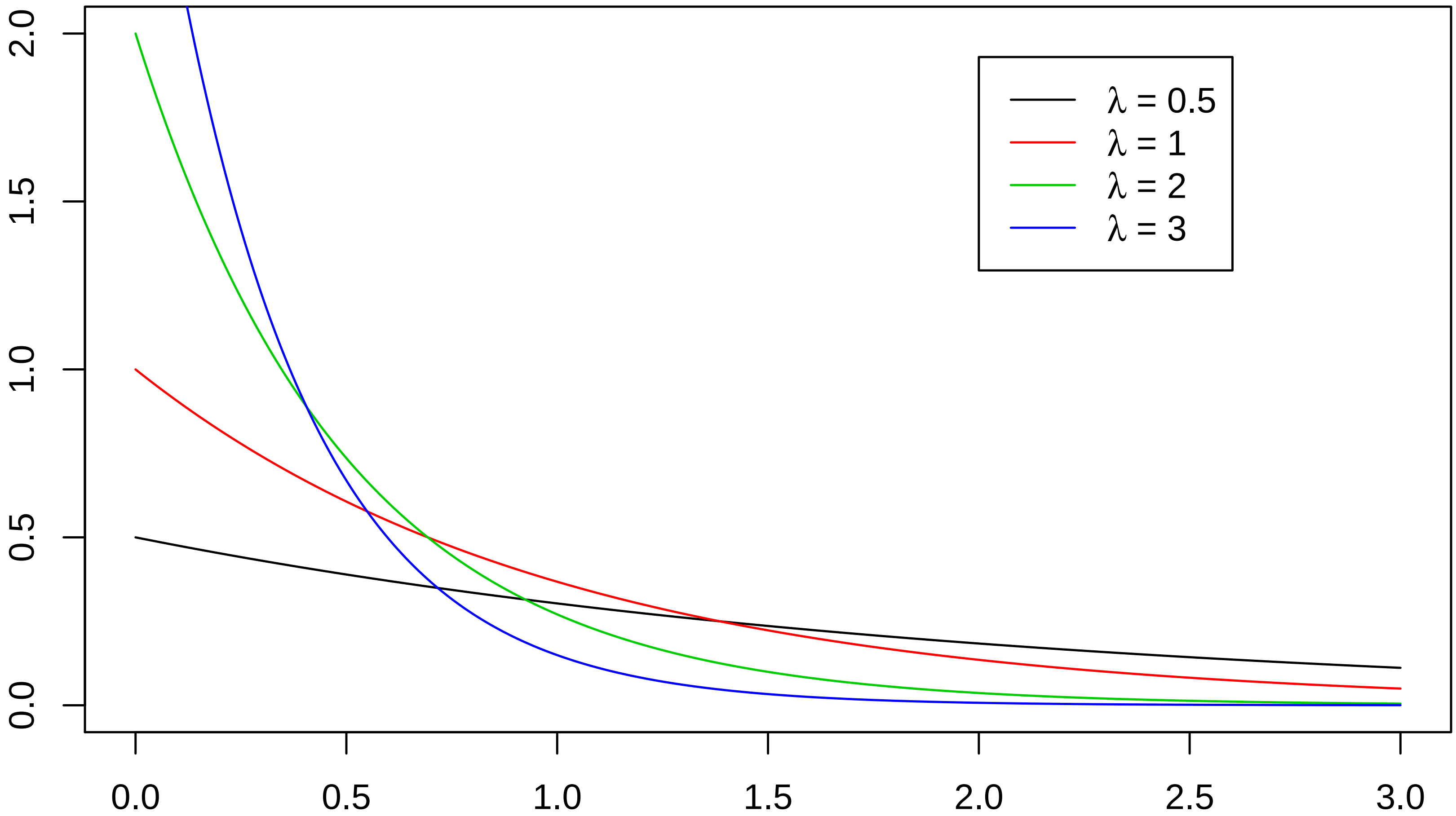
Definition, some properties.

- A continuous random variable X is said to have an exponential distribution with parameter $\lambda > 0$ if it has the density
$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \quad x > 0.$$
- In symbols: $X \sim \text{EXPO}(\lambda)$.
- Expectation and variance are:

$$\mathrm{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{var}(X) = \frac{1}{\lambda^2}.$$

7.4 The Exponential Distribution

This picture shows the density.



7.4 The Exponential Distribution

Example: Interarrival times of customers in a copy-shop.

Interarrival times. . .

. . . before 2 p.m.:

(#)		
(11)	**0**⋆	00001122334
(6)	**0**•	678889
(6)	**1**⋆	033344
(2)	**1**•	57
(2)	**2**⋆	44
(1)	**2**•	7
(1)	**3**⋆	3
	3•	
	4⋆	
(1)	**4**•	5
(1)	**5**⋆	2
	5•	
(31)		

. . . after 2 p.m.:

(#)		
(24)	**0**⋆	000011111112222223334444
(12)	**0**•	566667778899
(4)	**1**⋆	0113
(1)	**1**•	8
(3)	**2**⋆	002
	2•	
	3⋆	
(1)	**3**•	5
	4⋆	
	4•	
	5⋆	
	5•	
(45)		

1|0=10 minutes

7.4 The Exponential Distribution

Forgetfulness.

- "Forgetfulness" or "memorylessness" of the distribution of a random variable X:

$$P(X \geq t+s | X \geq s) = P(X \geq t) \quad \text{for all } s,t > 0$$

- The following two statements are equivalent:

  A: The distribution of the continuous random variable X is forgetful.

  B: The random variable X is exponentially distributed.

7.4 The Exponential Distribution

The Poisson process.

- Suppose events happen according to the following assumptions:
  - The number of events occurring in $[s, s+t]$ has a Poisson distribution with parameter λt.
  - The numbers of events occurring in non-overlapping intervals are independent.

- Let $N_t = \#$ events in $[0, t]$.

- The process $(N_t)_{t \geq 0}$ is called Poisson process with intensity λ.

7.4 The Exponential Distribution

Interarrival times.

- Consider a Poisson process with intensity λ.
- Let

$$X = \text{length of the time interval between two successive events.}$$

- Then, $X \sim \text{EXPO}(\lambda)$.

7.4 The Exponential Distribution

Example: Occurrence of power failures.

- Power failures occur in a certain area of Istanbul according to a Poisson process with intensity $\lambda = 0.7$ per week.

- What is the probability that there is...
  - a power failure during the next 12 hours?
  - no power failure during the next 3 days?

- Hint: If $X \sim \text{EXPO}(\nu)$, then the distribution function of X is

$$x \mapsto P(X \leq x) = \int_0^x \nu \mathrm{e}^{-\nu\xi} \mathrm{d}\xi = 1 - \mathrm{e}^{-\nu x}.$$