Continuous Distributions
Normal Distribution

Peter Büchel

HSLU W

SA: W04

Recap: Discrete Probability Distribution

- Random variable X: Assigns exactly one number to each random experiment
- Regard X as *function*
- Example: Random variable X assigns body height in cm to a randomly chosen person living in Switzerland
- Here: Body height is rounded to centimeters
- *Domain* of this random variable X: All people living in Switzerland
- Random variable X can only take following values (*range*)

$$W_X = \{0, 1, 2, \ldots, 500\}$$

- Range deliberately too large: All occurring values are included
- Set of values consists only of a finite numbers of integers
- Such a set is called *discrete*
- Important: Can choose *no* value between two values of range
- The quantity is “holey”
- This can be taken as a casual definition of *discreet*

- Random Variable X: Measures height of a randomly chosen person
- Now randomly (therefore random variable) choose a person
- Name of the person: *Tabea*
- Assume: Each name occurs only once, which is of course not the case
- Could have chosen AHV (social security) number, which is unique
- Tabea has height of 166 cm (rounded to cm)
- Formulation with random variable:

$$X(\text{Tabea}) = 166$$

- Choose another person: *Tadeo* with height of 176 cm

- Write:

$$X(\text{Tadeo}) = 176$$

- Can do this with every person living in Switzerland
- Expression:

$$X = 174$$

- *Event* of having chosen a person with a rounded height of 174 cm
- Speak of a *realization* $x = 174$ of X
- *Difference*: Capital and lower case letters:
  - $x = 174$ is a number
  - $X = 174$ is a set (people with rounded height 174 cm)

- Can assign probability to this event:

$$P(X = 174)$$

- Calculation: Divide number of people with rounded height of 174 cm by number of people living in Switzerland
- Example: $|X = 174| = 200\,000$, Population $|\Omega| = 8\,750\,000$:

$$P(X = 174) = \frac{|X = 174|}{|\Omega|} = \frac{200\,000}{8\,750\,000}$$

- Determine this way all probabilities for x in range of random variable:

$$P(X = x)$$

- In particular:

$$P(X = 500) = 0$$

- Because there no person with such a height
- Therefore it does not matter if range is way too large
- Can determine further probabilities
- So:

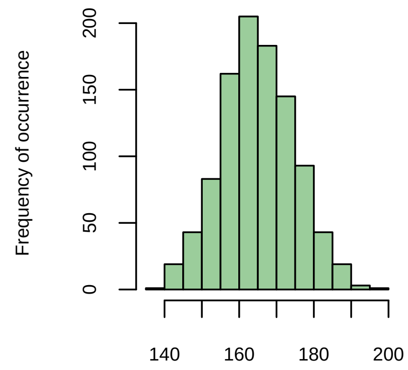
$$P(X \leq 170)$$

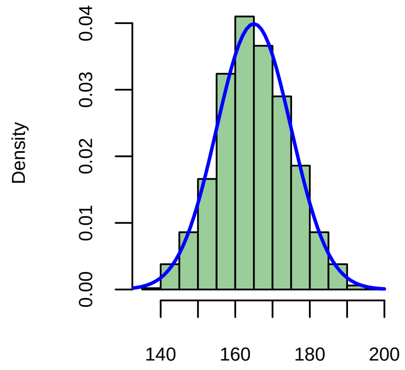
  - Probability that randomly chosen person has rounded height of 170 cm *or less*
- Important: Adding up all probabilities of distribution gives 1:

$$P(X = 0) + P(X = 1) + \ldots + P(X = 499) + P(X = 500) = 1$$

Example

- Choose randomly (see DoE) 1000 adult women
- Measure body height and create a histogram



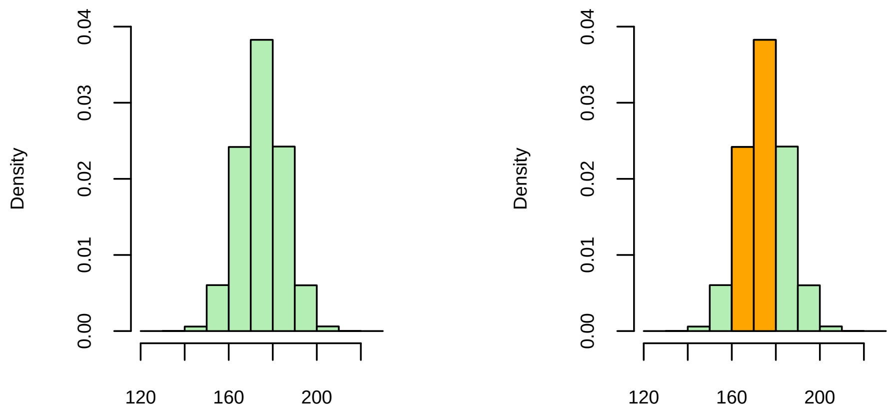


- Left: *Form* of histogram very typical: Occurs quite frequently
- In center bars high
- Flattens out further away from center
- Right: Try to draw a curve that follows histogram as closely as possible
- Apply density on vertical axis: Area of histogram 1
- Blue curve: *Curve of normal distribution*

From Discrete to Continuous Probability Distribution

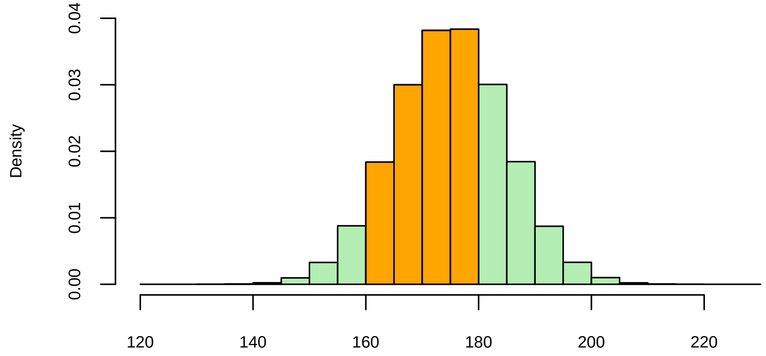
- Simulation of body height (in cm) of one million people
- Plot a histogram of these heights
- Assume that body height of each person is known as accurately as possible
- Histogram below (left) is normalised: Sum of area all bars is 1
- Start with a bar width of 10 cm

- Histogram:



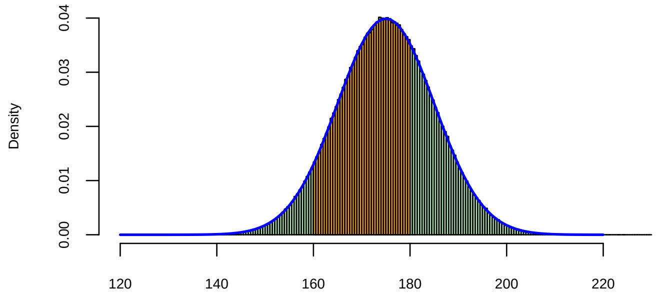
- Histogram (right): Two bars from 160 to 180 are coloured
- Because histogram normalised: Interpret *area* of two bars as *probability* that randomly chosen person of these 1 000 000 people has height between 160 cm and 180 cm
- Reasoning for this: Height of *any* of these people is contained in histogram
- Probability that *any* height of randomly chosen person is contained in histogram is 1
- That is area of sum of areas of all bars: 1
- Regard area of two bars as proportion of all people with height contained in these two bars
- This proportion: Corresponding probability

- Histogram with bar width of 5 cm:



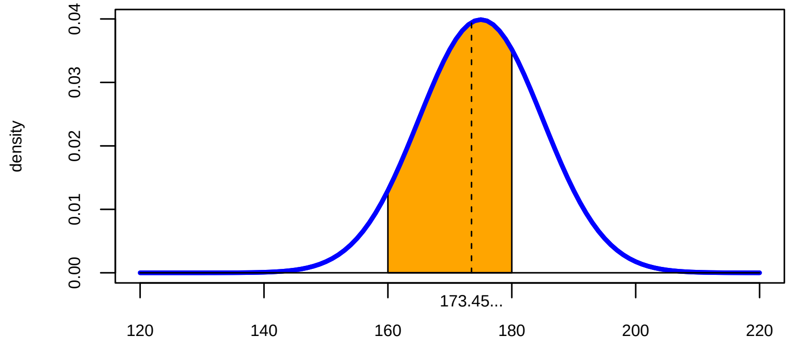
- Area of sum of all areas of bars is still 1
- Interpretation of the coloured area same as before
- Note: Area of the individual bars is less than area of individual bars in histogram before

- Histogram with bar width of 0.5 cm:



- The smaller bar width the more histogram follows a smooth curve

- Now final step
- Bar width tends to 0 (infinitely small):



- "Histogram" follows smooth curve
- Area under this curve is 1
- Coloured area is still probability, that a randomly chosen person has a height between 160 cm and 180 cm
- Area of individual "bars" is 0 (dashed line)
- Blue curve: *Probability density function*

Continuous Measurement Data

- In many applications: *Measurement data*
- Measurement data: Can take (theoretically) any value in certain range
- Example: Body height (in cm) of humans can take *each* value in *interval* $[0, 500]$
- Example:

$$145.325\,986\,54\ldots$$

- Prerequisite: Measurements as accurate as possible

Definitions

- *Range* W_X of a random variable: Set of all values X can take
- Random variable X *continuous*, if its range W_X is continuous
- Continuous: "Continuous" and not "holey", like set $\{1, 2, 3\}$
- Generally: Range is interval (part of number line)
- Important continuous value range:

$$W_X = \mathbb{R},\ \mathbb{R}^+ \quad \text{or} \quad [0, 1]$$

- Last case: Numbers 0 and 1 *and* all numbers in between

Point Probability 0

- Probability distribution of *discrete* random variable: Probability $P(X = x)$ for all possible x in range
- Example: $P(X = 174) = \frac{200\,000}{8\,750\,000}$ in Slide 6
- But for continuous random variable X for all $x \in W_X$:

$$P(X = x) = 0$$

- Conclusion: Probability distribution of X can not be described by $P(X = x)$

Example: Height

- Measuring height of people
- Probability of *exactly* a height of $168.254\,680\,895\,434\ldots$ cm: 0

$$P(X = 168.254\,680\,895\,434\ldots) = 0$$

- Dashed line in Slide 16: Area 0
- But possible: Probability that measurement lies in certain range (interval)

- Example: Between 174 and 175 cm:

$$P(174 < X \leq 175)$$

- This probability is no longer 0
- Because
$$P(X = 174) = P(X = 175) = 0$$
- It follows:

$$P(174 < X \leq 175) = P(174 \leq X \leq 175) = P(174 < X < 175)$$

- New concept: *Probability density*

Probability Density

- Probability densities are almost arbitrary under restrictions mentioned below
- Generalisation of height example Slide 16

Properties Probability Density Function

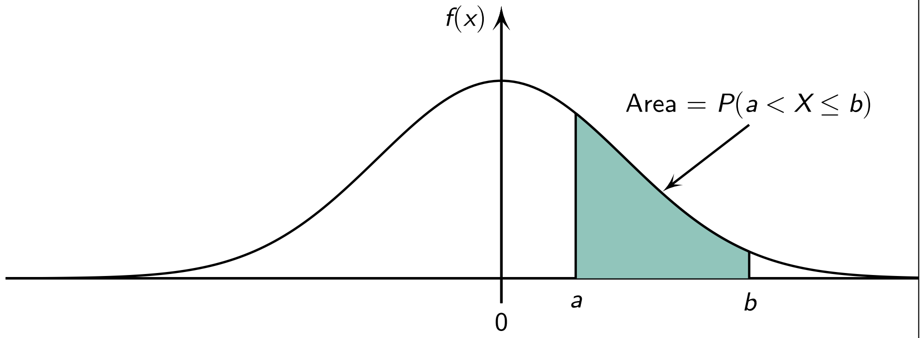
Probability density function $f(x)$ has following properties:

- Function is not negative:

$$f(x) \geq 0$$

  - This means that curve lies on or above x axis
- Probability
$$P(a < X \leq b)$$
  - Corresponds to area between a and b under $f(x)$
- Total area under curve is 1
  - This is probability that *any* value is measured
- Important: Values of $f(x)$ are *not* probabilities, only areas are

- Sketch:

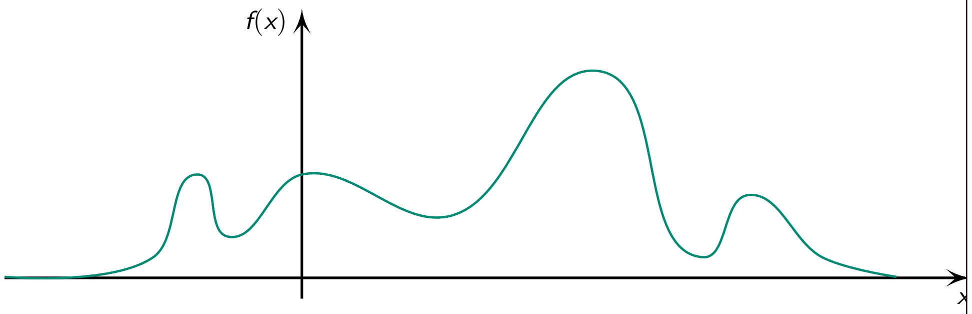


- Important: Relationship between probability and areas:

For continuous probability distributions, probabilities correspond to areas under density function

Remark

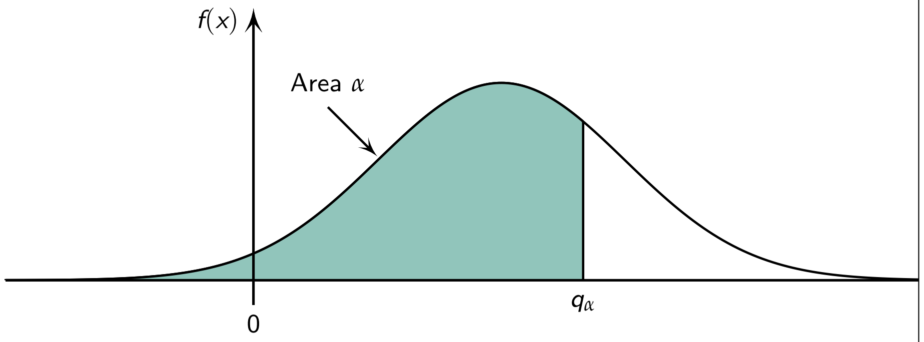
- Probability density functions do not have to look „nice":



- Here only "nice" functions as normal distribution and related t distribution

Quantiles

- For continuous distributions, the α quantile q_α is value where area (probability) under density function from $-\infty$ to q_α is just α
- 50 % quantile is called *median*
- Sketch:



Example: Body Height

- For $\alpha = 0.75$, corresponding quantile is

$$q_\alpha = 182.5$$

- Means: 75 % of heights are less than or equal to 182.5 cm

Normal (Gaussian) Distribution: $X \sim \mathcal{N}(\mu, \sigma^2)$

- Definition "must" have been seen once
- *Range*:

$$W = (-\infty, \infty)$$

- *Density function*:

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

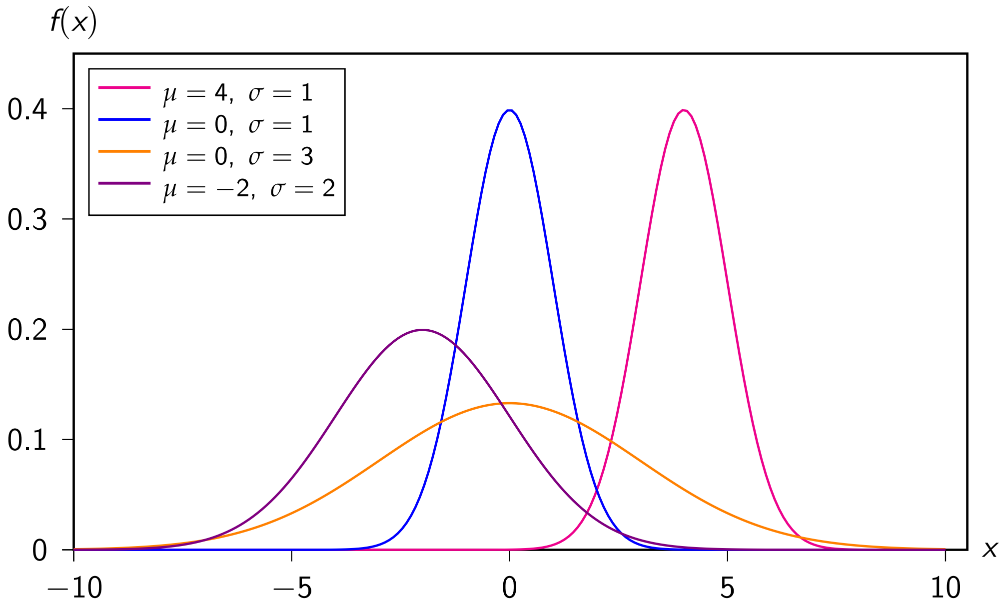
- *Expected value*:

$$\mathrm{E}(X) = \mu$$

- *Variance*:

$$\mathrm{Var}(X) = \sigma^2$$

Normal distribution: Illustration Graphs



Properties of Normal Distribution Density Function

- Parameter μ: Shifting curve horizontally from origin:
  - To right: If μ positive
  - To left: If μ negative
  - μ: Where peak of curve is
- Parameter σ: *Shape* of curve:
  - Narrow and high around μ: If σ small (close to 0)
  - Wide and low around μ: If σ large

Example with R: Distribution of IQ

- Example: IQ tests follow normal distribution with mean 100 and standard deviation 15
- Constructed in such a way
- X: IQ of a randomly chosen person
- X normally distributed with $\mu = 100$ and $\sigma = 15$
- Write:

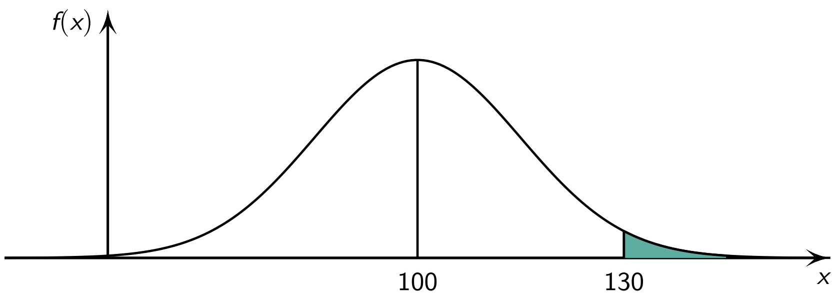
$$X \sim \mathcal{N}(100, 15^2)$$

- How large is probability that randomly chosen person has an IQ of more than 130, i.e. is considered highly gifted?
- Sought:

$$P(X > 130)$$

where $X \sim \mathcal{N}(100, 15^2)$

- Sketch:



- Calculation of $P(X > 130)$ with R command `pnorm(...)`
- Command calculates probability:

$$P(X \leq 130)$$

- Note direction of inequality sign!
- Calculation:

```
pnorm(q = 130, mean = 100, sd = 15)

[1] 0.9772499
```

- Calculates area (probability) from $-\infty$ to `q = 130` under normal distribution curve with $\mu = 100$ and $\sigma = 15$
- But this is *not* asked for probability $P(X > 130)$

- But since total area under curve is 1:

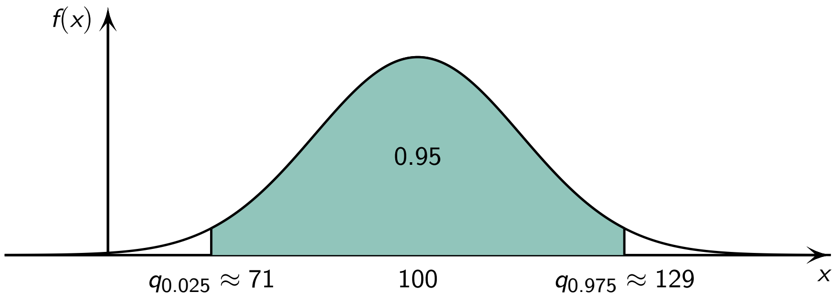
$$P(X > 130) = 1 - P(X \leq 130)$$

- Thus:

```
1 - pnorm(q = 130, mean = 100, sd = 15)

[1] 0.02275013
```

- About 2 % of population is highly gifted

- Which symmetrical interval contains 95 % of IQ's around mean $\mu = 100$?
- Again: Represent this probability as area:



- Green Area in center of Figure: 95 % of total area
- Small white areas on left and right: 0.025 each

- Probabilities given: Look for corresponding values (quantiles)
- Determine *quantiles* $q_{0.025}$ and $q_{0.975}$
- With R:

```
qnorm(p = 0.025, mean = 100, sd = 15)
[1] 70.60054

qnorm(p = 0.975, mean = 100, sd = 15)
[1] 129.3995
```

- Or shorter:

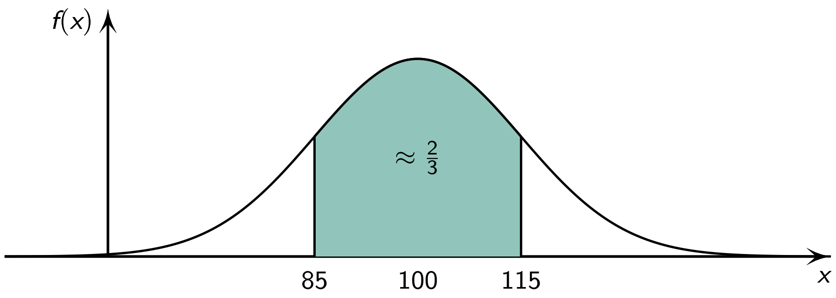
```
qnorm(p = c(0.025, 0.975), mean = 100, sd = 15)

[1]  70.60054 129.39946
```

- 95 % of people have an IQ between about 70 and 130
- Corresponds to distance of ≈ 2 standard deviations from mean 100

- What percentage of population is within *one* standard deviation of mean?
- Sought for probability:

$$P(85 \leq X \leq 115)$$

- Again, this probability is represented as area:



- With **R**:

```
pnorm(q = 115, mean = 100, sd = 15) - pnorm(q = 85, mean = 100, sd = 15)

[1] 0.6826895
```

- This means: About $\frac{2}{3}$ of population have an IQ between 85 and 115

Normal Distribution: Properties

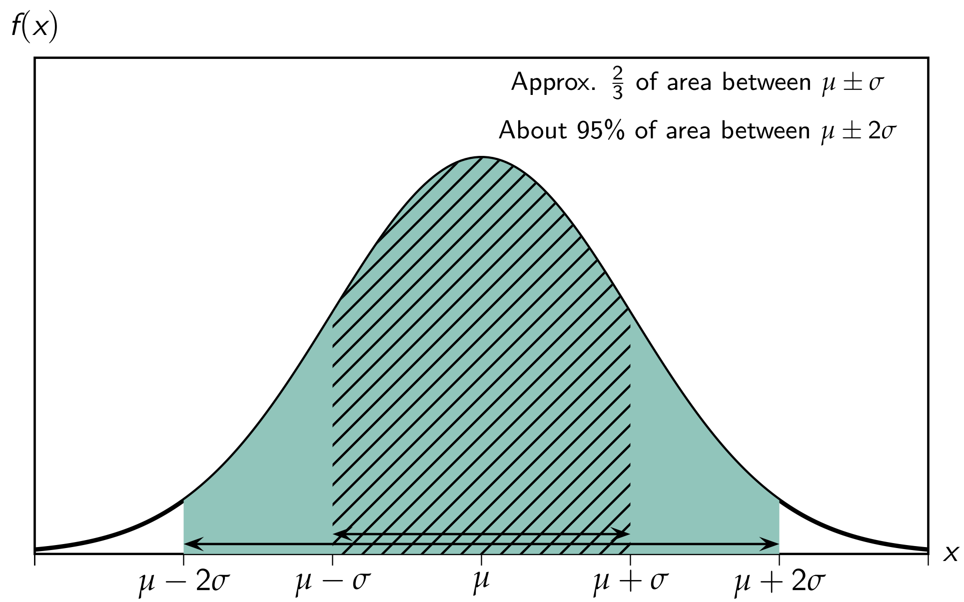
- Last result from Example applies to *all* normal distribution $\mathcal{N}(\mu, \sigma^2)$
- Probability that an observation deviates from expected value by at most one standard deviation is about $\frac{2}{3}$:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx \frac{2}{3}$$

- Normal distribution: Concrete statement for spread as "deviation" from expected value
- Deviation from expected value by at most two standard deviations:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$

Normal Distribution: Properties



Functions of Several Random Variables

- So far: Considered distribution of *one* random variable (RV)
- But: Usually *same* quantity is measured several times
- Example: Measure weight several times
- In general: Observations $x_1, x_2, \ldots, x_n$ realisations of random variables:
$$X_1, \ldots, X_n$$
- X_i: ith repetition of random experiment

Example

- Want to check whether beer cans *really* have content 500 ml
- Buy 20 cans: Measure contents, take average of observations
- One can can deviate from 500 ml slightly, average should deviate less
- Observations (concrete values):
$$x_1 = 501.35, \quad x_2 = 499.95, \quad \ldots, \quad x_{20} = 498.67$$
- Realisations of RV:
$$X_1,\ X_2,\ \ \ldots,\ X_{20}$$
- Assumption: 20 RV with *same* probability distribution
- Interested in: *Average* of these observations and distribution of corresponding RV

Sum and Average

- Given RV:
$$X_1, \ldots, X_n$$
- *Sum*:
$$S_n = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$$
- *Arithmetic mean*:
$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n} S_n$$

Key Figures of S_n and $\overline{X}_n$

- Assumption:
$$X_1, \ldots, X_n \ \text{ i.i.d.}$$
- First “i”: *independent*
- “i.d”: *identically distributed*
- Second “i” in i.i.d.: X_i same distribution with same key figures:
$$\mathrm{E}(X_i) = \mu \quad \text{and} \quad \mathrm{Var}(X_i) = \sigma_X^2$$
- Looking for: Expected value and variance of:
  - Sum S_n:
$$S_n = X_1 + X_2 + \ldots + X_n$$
  - Average $\overline{X}_n$:
$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Graphical Example

- Rolling fair die
- X: RV for rolled number of eyes
- Expected value:

$$E(X) = \mu = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

- Variance:

$$\begin{aligned}\text{Var}(X) &= \frac{1}{6}\left((1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\right)\\ &= 2.92\end{aligned}$$

- **R**:

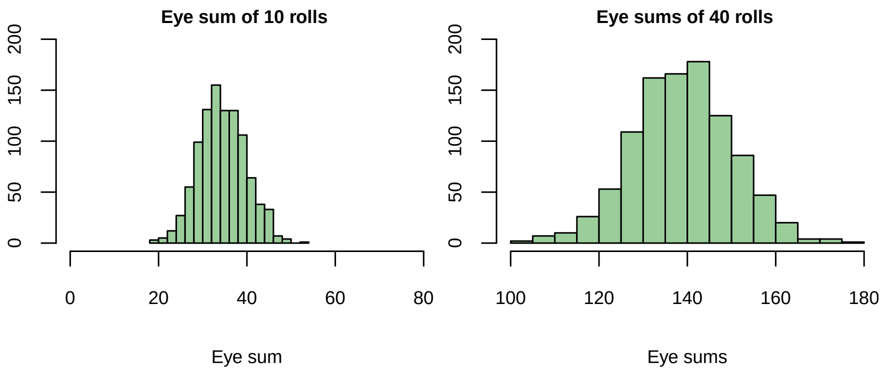
```
x <- c(1, 2, 3, 4, 5, 6)

ave <- mean(x)
ave
[1] 3.5

var <- mean((x - ave)^2)
var
[1] 2.916667
```


- Rolling die 10 times
- RV:

$$X_1, X_2, \ldots, X_{10} \quad \text{i.i.d.}$$

- X_i: Score in ith roll
- Expected value and variance: Values of RV X_i above
- Note down eye sum s_{10} of these 10 rolls
- Repeat 1000 times: Histogram of all occurring eye totals
- Same with 40 rolls
- Simulation with **R**

Histograms



Findings

- Eye sum histogram shifts to right with more rolls
- Panel left: Largest frequency at about 35, so

$$10 \cdot 3.5 = 10 \cdot \mu$$

- $\mu = 3.5$: Expected value for *one* roll
- Panel right: Largest frequency at about 140

$$40 \cdot 3.5 = 40 \cdot \mu$$

- Conjecture:

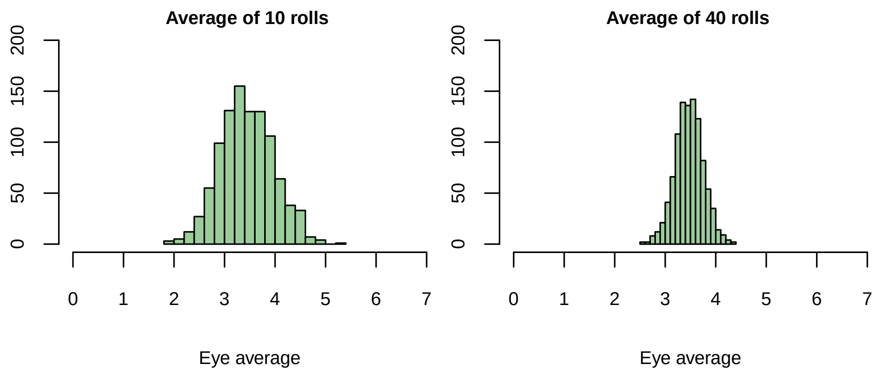
$$E(S_n) = n\mu$$

- Conjecture true (without proof)

- Variance/standard deviation *increases* with increasing number of rolls
- Law (without proof):

$$\mathrm{Var}(S_n) = n\,\mathrm{Var}(X), \qquad \sigma_{S_n} = \sqrt{n}\sigma_X$$

- Consider $\overline{X}_n$ of these 1000 rolls
- Histograms:



Findings

- Both histograms: Largest frequency at 3.5, so μ
- Conjecture (true, but without proof):

$$\mathrm{E}(\overline{X}_n) = \mu$$

- Variance/standard deviation *decreases* with increasing number of rolls
- Law (without proof):

$$\mathrm{Var}(\overline{X}_n) = \frac{\mathrm{Var}(X)}{n}, \qquad \sigma_{\overline{X}_n} = \frac{\sigma_X}{\sqrt{n}}$$

- Above observations are hold generally

In General

- Assumption:

$$X_1, \ldots, X_n \text{ i.i.d.}$$

- It follows:

Key figures of S_n

$$\mathrm{E}(S_n) = n\mu$$
$$\mathrm{Var}(S_n) = n\,\mathrm{Var}(X_i)$$
$$\sigma(S_n) = \sqrt{n}\sigma_X$$

Key figures of $\overline{X}_n$

$$\mathrm{E}(\overline{X}_n) = \mu$$
$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma_X^2}{n}$$
$$\sigma(\overline{X}_n) = \frac{\sigma_X}{\sqrt{n}}$$

Remarks

- Standard deviation of $\overline{X}_n$: Called *standard error* of arithmetic mean
- Standard deviation of sum: Grows with increasing n, but slower than the number of observations n
- I.e.: Larger spread for growing n
- Expected value of $\overline{X}_n$: Equal to that of a single RV X_i, but *spread decreases with increasing n*

Standard error

Standard deviation of arithmetic mean (*standard error*) is *not* proportional to $1/n$, but decreases with factor $1/\sqrt{n}$:

$$\sigma_{\overline{X}_n} = \frac{1}{\sqrt{n}}\sigma_X$$

To halve *standard error*, need *four times* as many observations

This is also called $\sqrt{n}$*-law*

Central Limit Theorem (CLT)

- Known: Key figures of S_n and $\overline{X}_n$
- Unknown: Distribution of S_n and $\overline{X}_n$
- Fair die: X_i equally distributed:

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Obviously: Not normally distributed
- *How* is S_n and $\overline{X}_n$ distributed?
- Will see: Both approximately *normally distributed*
- This is statement of *central limit theorem* (no proof)
- Simulation of statement: Lecture notes Example **??**

Central Limit Theorem

- X_i's i.i.d. (not necessarily normally distributed), then:

Central Limit Theorem

$X_1, \ldots, X_n$ i.i.d. with any distribution with expected value μ and variance σ^2, then (without proof):

$$S_n \approx \mathcal{N}(n\mu, n\sigma_X^2)$$

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right)$$

- Approximation is generally better for larger n
- Approximation better the closer original distribution of X_i is to the normal distribution $\mathcal{N}(\mu, \sigma_X^2)$

Example

- PWD (public works departement) has stored enough road salt to cope with a total snowfall of 80 cm per year
- Daily average snowfall is 1.5 cm with a standard deviation of 0.3 cm
- What is probability that stored salt is sufficient for next 50 days?

Solution

- X_i: RV for fallen snow during ith day
- Assumption i.i.d.: Justified?
  - Not really
  - If it snows today, it is more likely to snow tomorrow as well than on a sunny day
  - But let's assume, the RVs are i.i.d.
- It follows $\mu = 1.5$ and $\sigma_X = 0.3$
- Snow amount (total) S_{50} of the next 50 days
- Should not exceed 80

- It applies approximately:

$$S_{50} \sim \mathcal{N}\left(50 \cdot \mu, 50 \cdot \sigma_X^2\right) = \mathcal{N}(75, 4.5)$$

- Sought:

$$P(S_n \leq 80) = 0.991$$

```
pnorm(q = 80, mean= 50*1.5, sd = sqrt(50)*0.3)

[1] 0.9907889
```

- Probability of about 99 % means: Once in a century stored salt is not sufficient

Example

- Lifetime of a given electrical part is on average 100 hours with standard deviation of 20 hours
- Test 16 such parts
- How large is probability that the sample mean
  - is under 104 hours or
  - is between 98 and 104 hours?

Solution

- X_i: Random variable for lifetime of the part i
- It follows $\mu = 100$ and $\sigma_X = 20$
- Assumption i.i.d.: Justified?
- Considering average lifetime $\overline{X}_{16}$
- Approximately distributed as:

$$\overline{X}_{16} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(100, \frac{20^2}{16}\right) = \mathcal{N}(100, 25)$$

- Sought:

$$P(\overline{X}_{16} \leq 104) = 0.788$$

```
pnorm(q = 104, mean = 100, sd = 20 / sqrt(16))

[1] 0.7881446
```

- Sought:

$$P(98 \leq \overline{X}_{16} \leq 104) = 0.444$$

```
pnorm(q=104, mean=100, sd=20 / sqrt(16) ) - pnorm(98, 100, 20/sqrt(16))

[1] 0.4435663
```

- Probability that lifespan of electrical part is between 98 and 104 hours is almost 0.44

Remark: $\mathcal{N}(\mu, \sigma^2)$ or $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

- $\mathcal{N}(\mu, \sigma^2)$ for *one* observation
- $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ for *mean* of n observations