Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

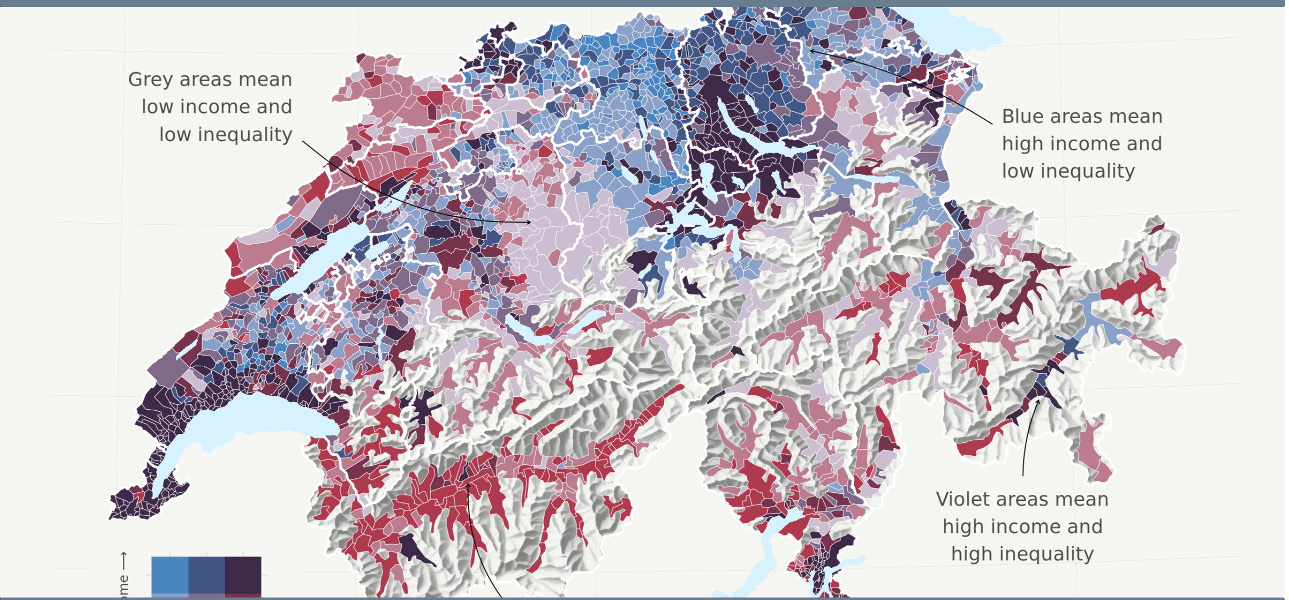


figure source

Statistics for WING

Vidushi Bigler, FS 2020, Classes W1 & W2

▶ Technik und Informatik

- Motivation

- Random variables
- Probability distribution
- Cummulative distribution function
- Expectation value & Variance

- Independent & dependent random variables

- Diagnostic tests

- Motivation

- **Random variable**
- **Probability distribution**
- **Cumulative distribution function**
- **Expectation value & variance**

Real important

- Independent & dependent random variables

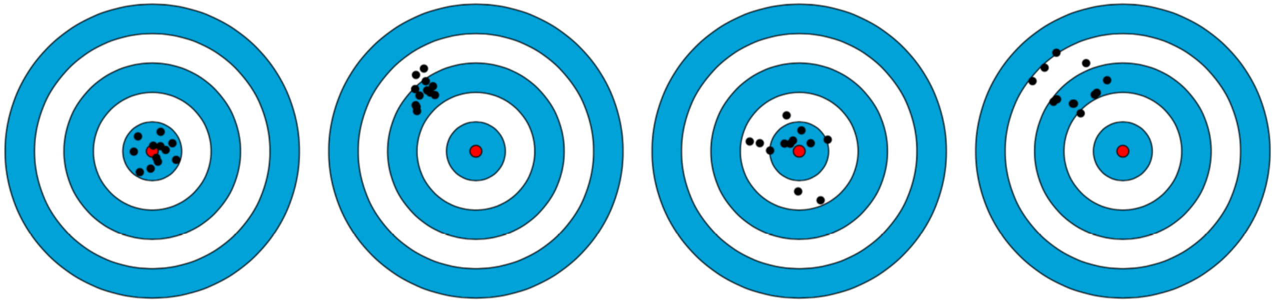
- Diagnostic tests

Motivation

Grundlegendes Problem in der Statistik

In diesem Abschnitt werden die Begriffe **Präzision**, **Richtigkeit** und **Genauigkeit** genauer erläutert. Präzision und Richtigkeit können illustrativ anhand einer Zielscheibe wie in Abbildung 1.1 dargestellt einfach erklärt werden. Dabei ist es das Ziel den roten Punkt in der Mitte der Scheibe zu treffen. Die schwarzen Punkte stellen die Schusslöcher dar, welche entstanden sind, als vier Bogenschützen ihr Glück versucht haben.

- Bogenschütze (a) ist sowohl präzise als auch richtig. Er ist präzise, weil alle Schusslöcher nahe beieinander liegen. Hohe Präzision bedeutet insbesondere hohe Wiederholbarkeit; im konkreten Fall trifft der Bogenschütze wiederholt den innersten Kreis. Die Schüsse zeigen zudem eine hohe Richtigkeit, weil sie alle um den roten Punkt streuen.
- Bogenschütze (b) ist zwar präzise aber nicht richtig. Seine Wiederholbarkeit ist sehr hoch, da die Streuung der Schusslöcher klein ist. Aber die Schusslöcher sind nicht um den roten Punkt herum verteilt.
- Bogenschütze (c) ist nicht präzise aber richtig. Die Schusslöcher streuen über einen grösseren Bereich und folglich ist die Präzision niedrig, bzw. die Wiederholbarkeit schlecht. Die Schusslöcher sind jedoch um den roten Punkt herum verteilt und daher ist die Richtigkeit gegeben.
- Bogenschütze (d) ist weder präzise noch richtig.



(a) präzise und richtig, gute Genauigkeit

(b) präzise, aber nicht richtig, schlechte Genauigkeit

(c) richtig, aber nicht präzise, schlechte Genauigkeit

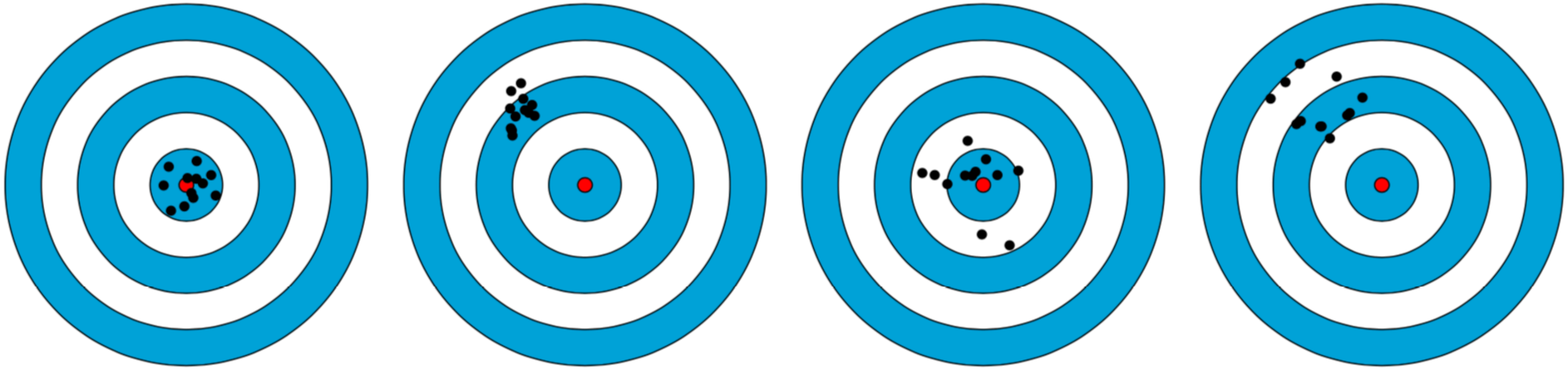
(d) weder präzise noch richtig, schlechte Genauigkeit

Abbildung 1.1.: Illustration der Präzision, Richtigkeit und Genauigkeit

Grundlegendes Problem in der Statistik

Wir wissen meistens nicht wo dieser rote Punkt liegt!

- wir müssen den Bias (Offset) in den Griff bekommen
- und wir müssen die Streuung quantifizieren können



(a) präzise und richtig, gute Genauigkeit

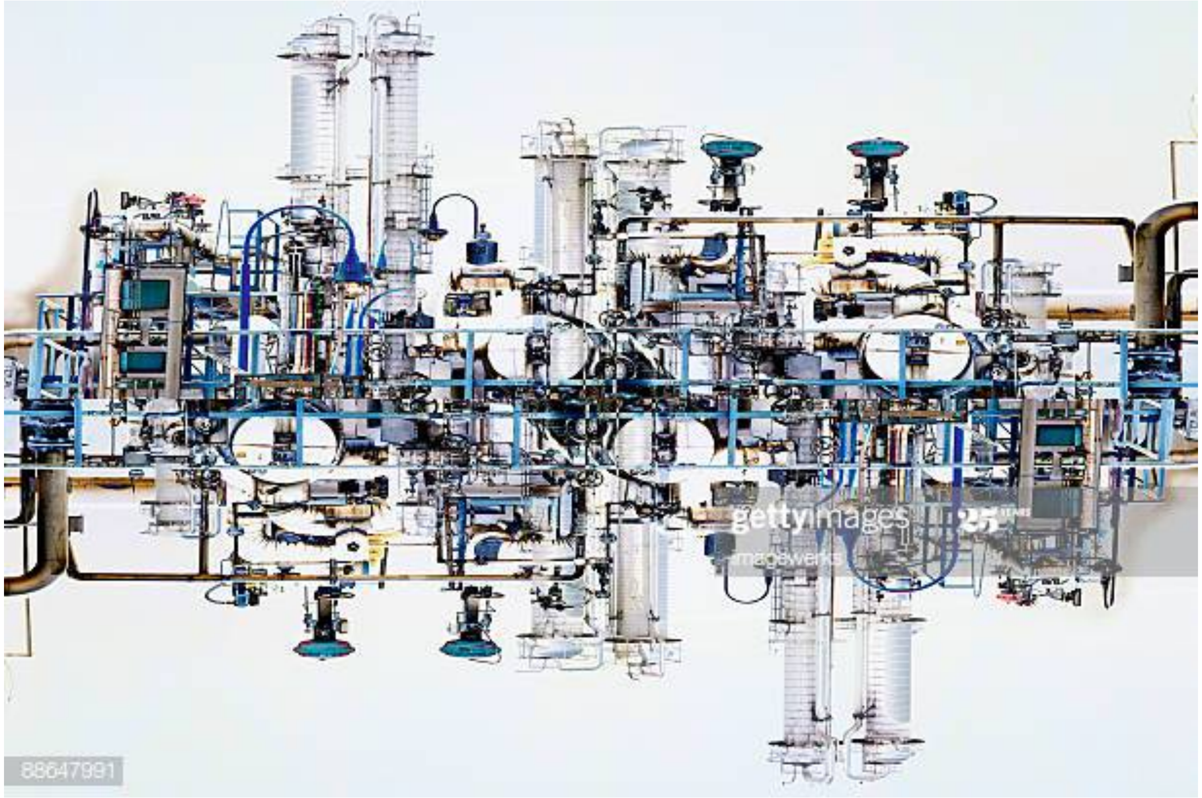
(b) präzise, aber nicht richtig, schlechte Genauigkeit

(c) richtig, aber nicht präzise, schlechte Genauigkeit

(d) weder präzise noch richtig, schlechte Genauigkeit

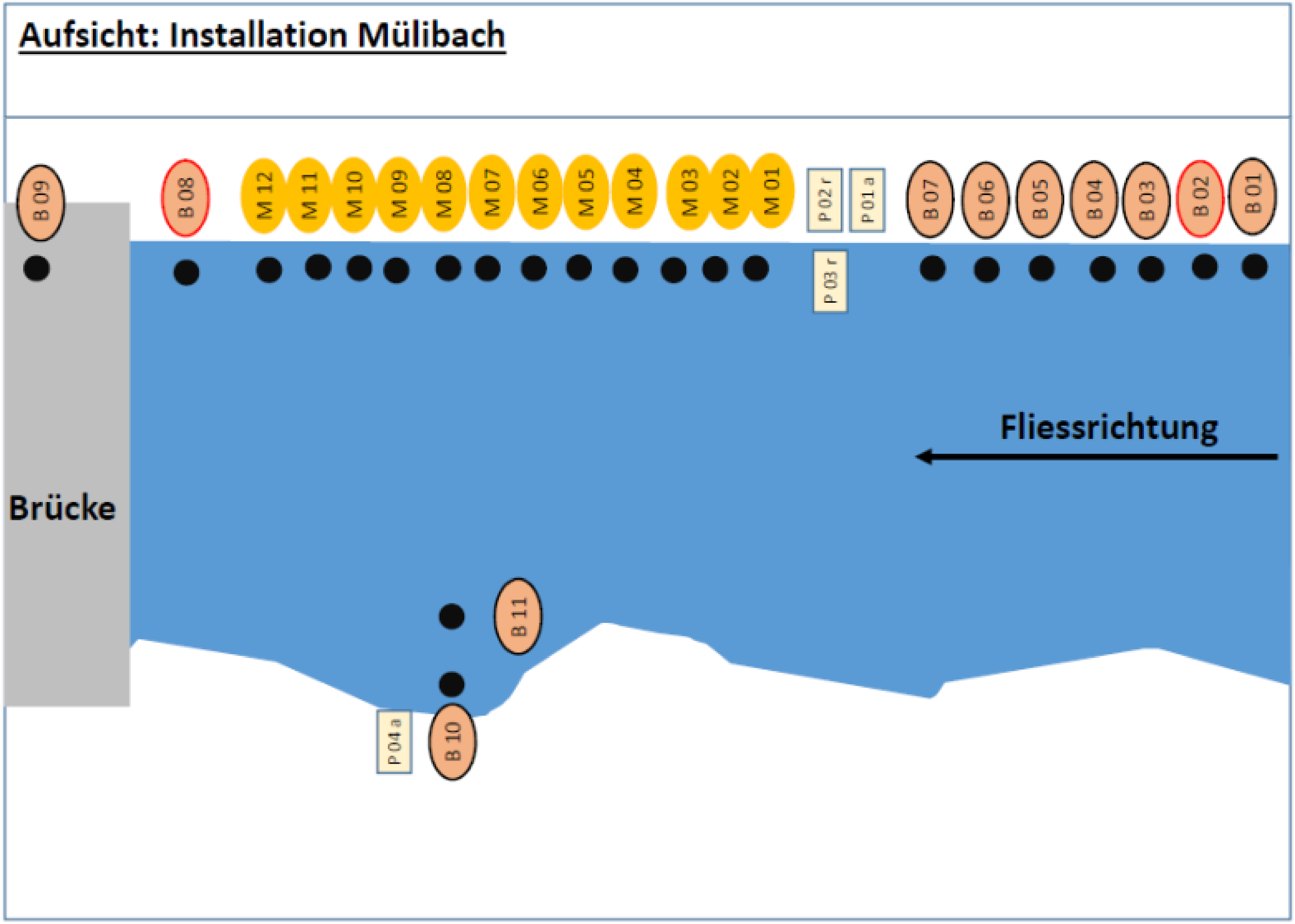
Abbildung 1.1.: Illustration der Präzision, Richtigkeit und Genauigkeit

Wir bauen eine riesige Maschinerie auf

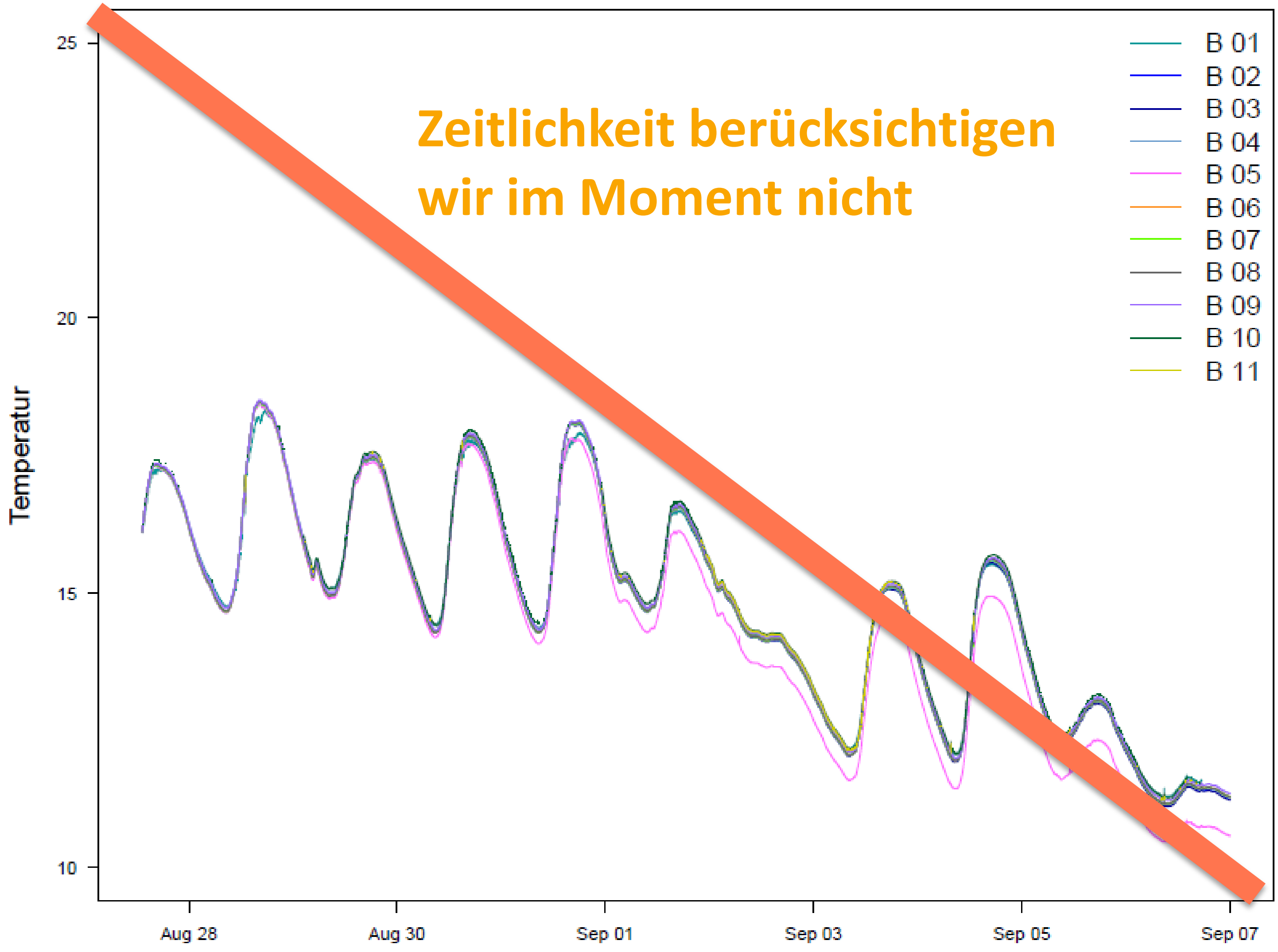


Credit: Getty Images/Imagewerks Japan

Beispiel I: Feldstudie Mülibach Burgdorf

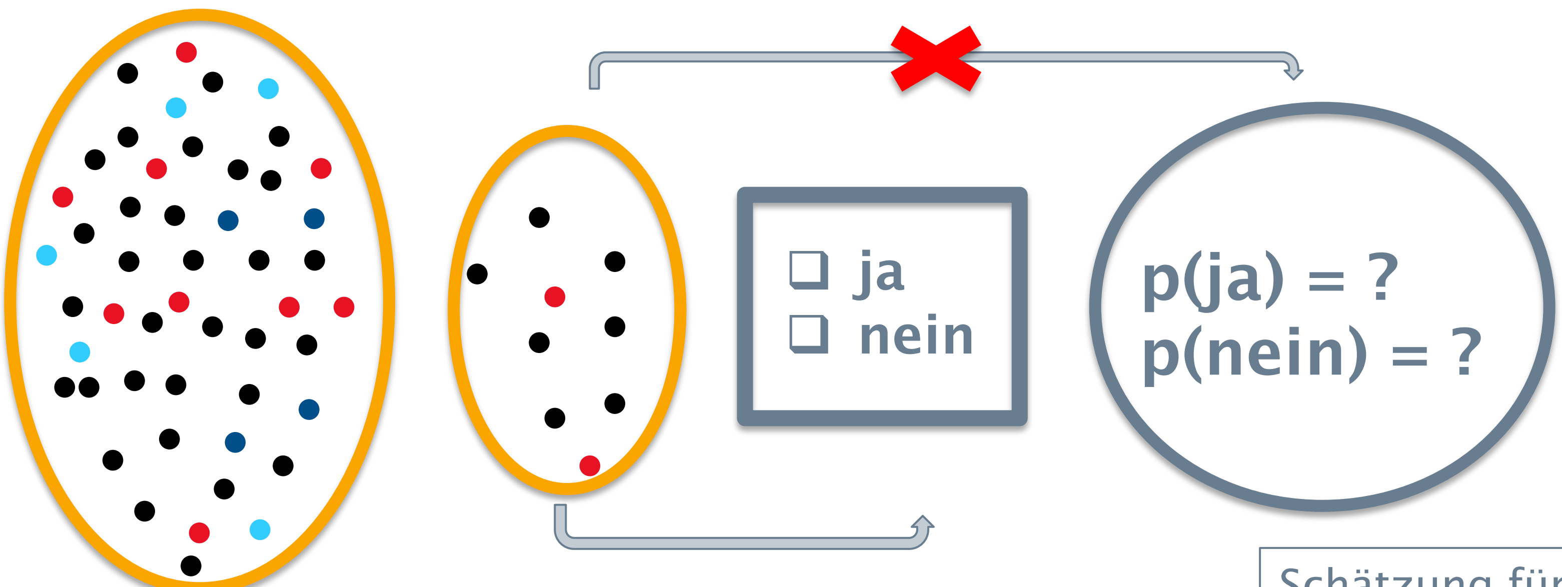


Beispiel I: Ergebnisse Feldstudie



Beispiel II: die Wiederwahl von Trump

Angenommen Sie machen eine Meinungsumfrage. Sie wählen zufälligerweise Menschen aus der Menge der Amerikaner und Amerikanerinnen aus und befragen sie, ob sie Trump wählen werden oder nicht.



Menge aller amerikanischer Wahlberechtigten

Ausgewählte Personen (unsere Stichprobe)

Zufallsvariable X codiert mit entweder 0 oder 1

Schätzung für die wahre Wahrscheinlichkeiten

Beispiel III: faire Münze

Guter Startpunkt, weil wir in diesem Fall alles mögliche exakt ausrechnen können!

Without further ado

Beispiel III: faire Münze

Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Frage: Welche Werte nimmt X an? Mit welchen Wahrscheinlichkeiten?

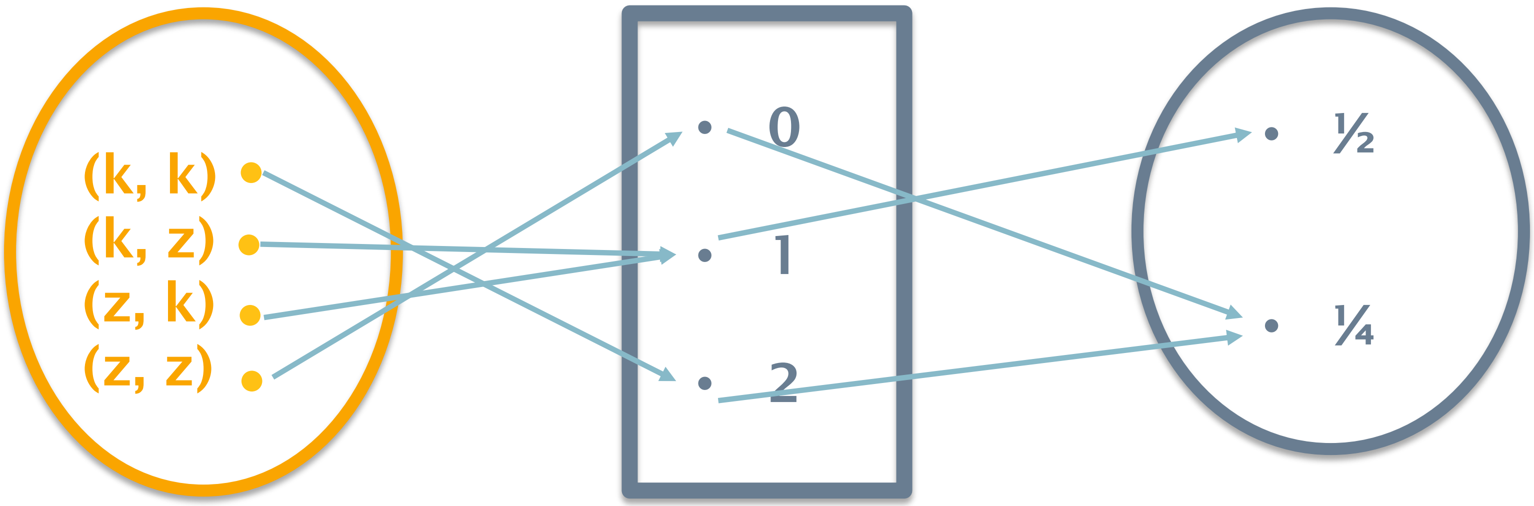
k: Kopf, z: Zahl
Das Erbegnisraum $\Omega = \{(k, k), (k, z), (z, k), (z, z)\}$

X kann die Werte 0, 1, 2 annehmen mit den Wahrscheinlichkeiten

$$\left.\begin{aligned} &P(X = 0) = \tfrac{1}{4} \\ &P(X = 1) = \tfrac{1}{2} \\ &P(X = 2) = \tfrac{1}{4} \end{aligned}\right\} \textbf{Wahrscheinlichkeitsverteilung}$$

Beispiel II: faire Münze

$X: \Omega \longrightarrow \mathbb{R}$



Gesamt Ω

Zufallsvariable X

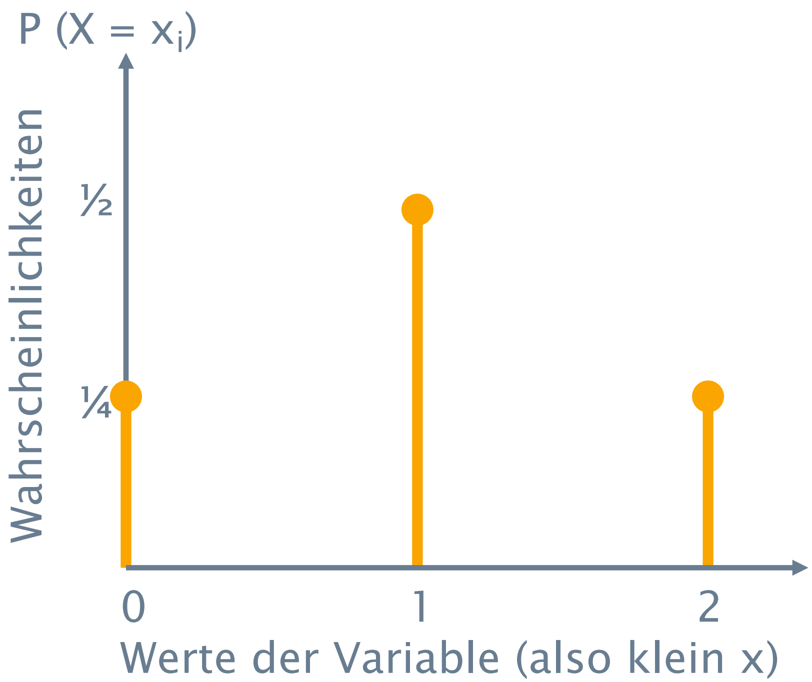
Exakte Wahrscheinlichkeiten

Beispiel III: faire Münze

Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Frage: Zeichnen Sie die zugehörige Wahrscheinlichtsverteilung als Graphen.

$X = x_i$	0	1	2
$P(X = x_i)$	¼	½	¼



Beispiel III: faire Münze

Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Fragen: Was ist die Wahrscheinlichkeit, dass Sie:

1. **genau ein Mal Kopf erhalten?**
2. **höchstens einmal Kopf erhalten?**
3. **mehr als einmal Kopf erhalten?**
4. **weniger als einmal Kopf erhalten?**
5. **... usw. usf.**

Cumulative distribution function

Beispiel III: faire Münze

Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Frage: Berechnen Sie die zugehörige cumulative Verteilungsfunktion.

Definition: $\boxed{F(x) = P(X \leq x)}$

Schritt 1: Summe (diskrete Zufallsvariable)

$X = x_i$	0	1	2
$P(X = x_i)$	¼	½	¼
am einfachsten Summe berechnen	¼	¼ + ½ = ¾	¼ + ½ + ¼ = 1

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} P(X = k)$$

Beispiel III: faire Münze

Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Frage: Berechnen Sie die zugehörige cumulative Verteilungsfunktion.

$X = x_i$	0	1	2
$P(X = x_i)$	¼	½	¼
$F(x_i)$	¼	¼ + ½ = ¾	¼ + ½ + ¼ = 1

Schritt 2: reellwertige Funktion angeben

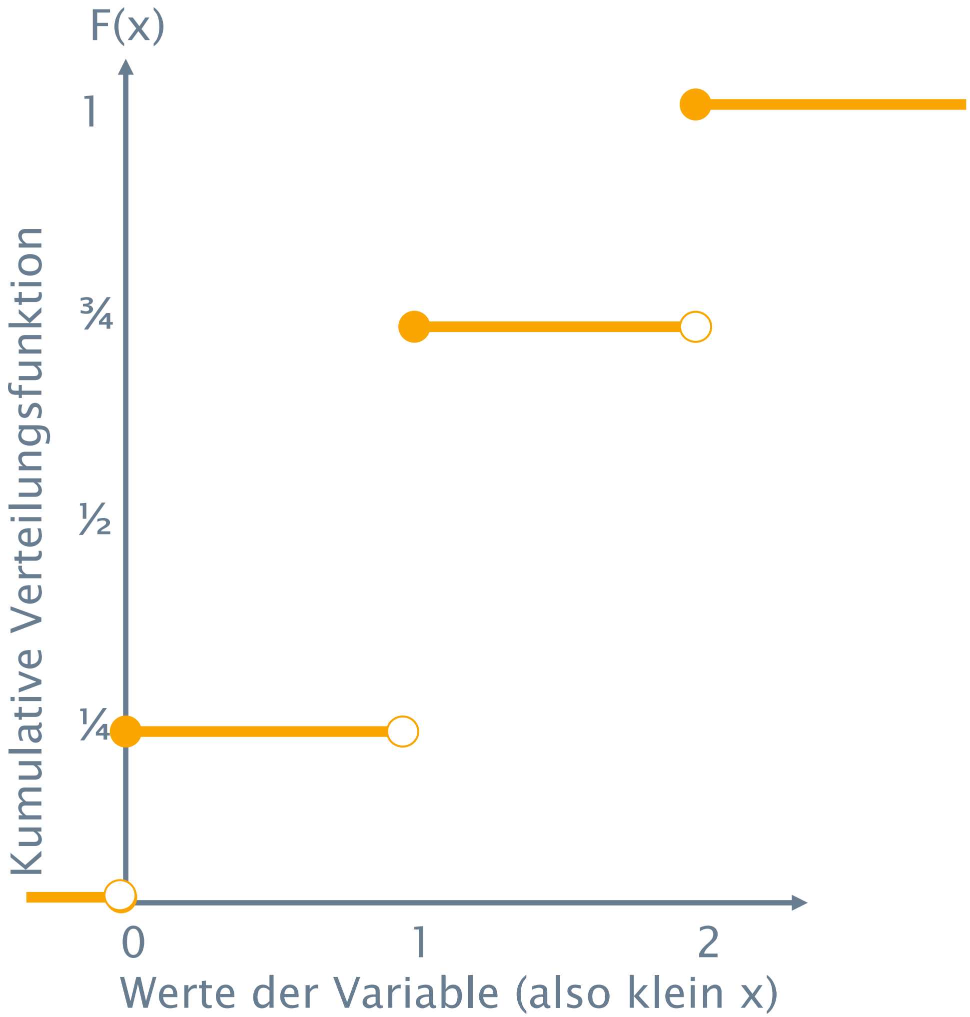
$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{4} & \text{falls } 0 \leq x < 1 \\ \frac{3}{4} & \text{falls } 1 \leq x < 2 \\ 1 & \text{sonst} \end{cases}$$

Beispiel III: faire Münze

Frage: Zeichnen Sie die zugehörige cumulative Verteilungsfunktion.

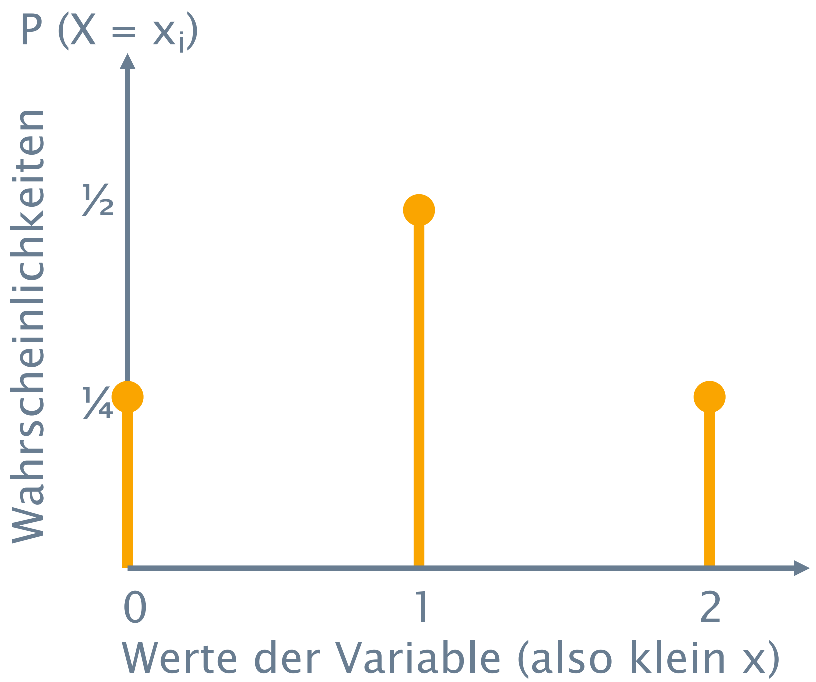
$$F(x) = P(X \leq x) = \sum_{k=0}^{2} P(X = k)$$

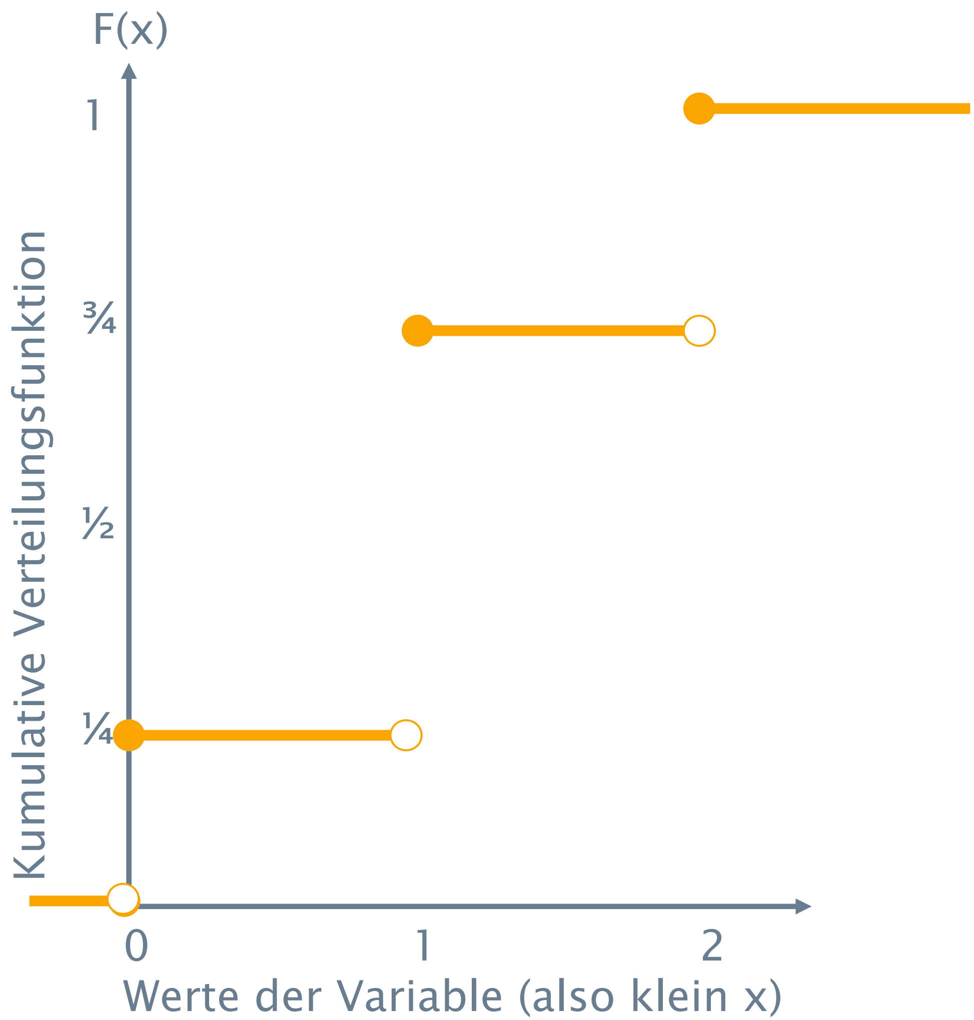
$$F(x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{4} & \text{falls } 0 \leq x < 1 \\ \frac{3}{4} & \text{falls } 1 \leq x < 2 \\ 1 & \text{sonst} \end{cases}$$



Beispiel III: faire Münze

Beide zusammen





Beispiel III: faire Münze

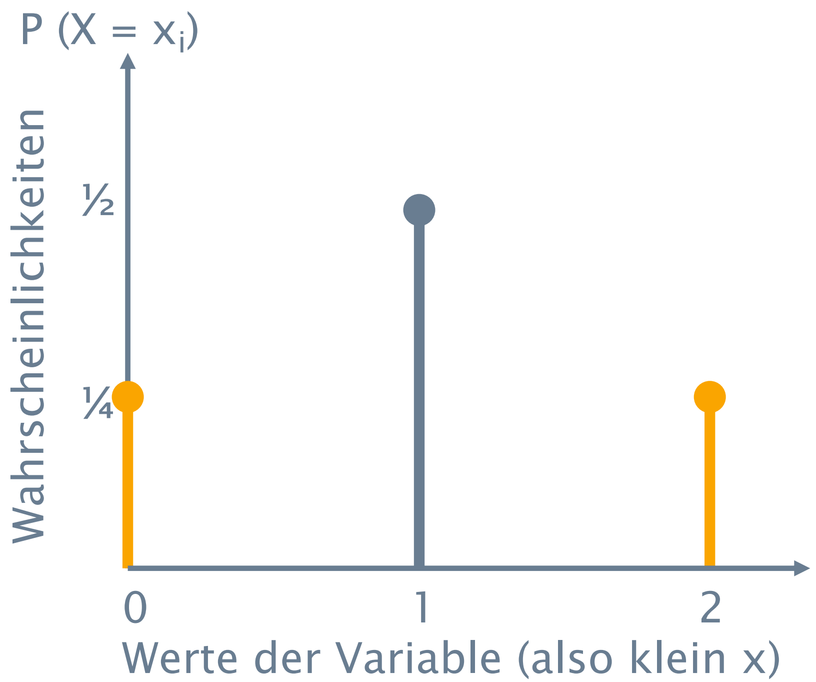
Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

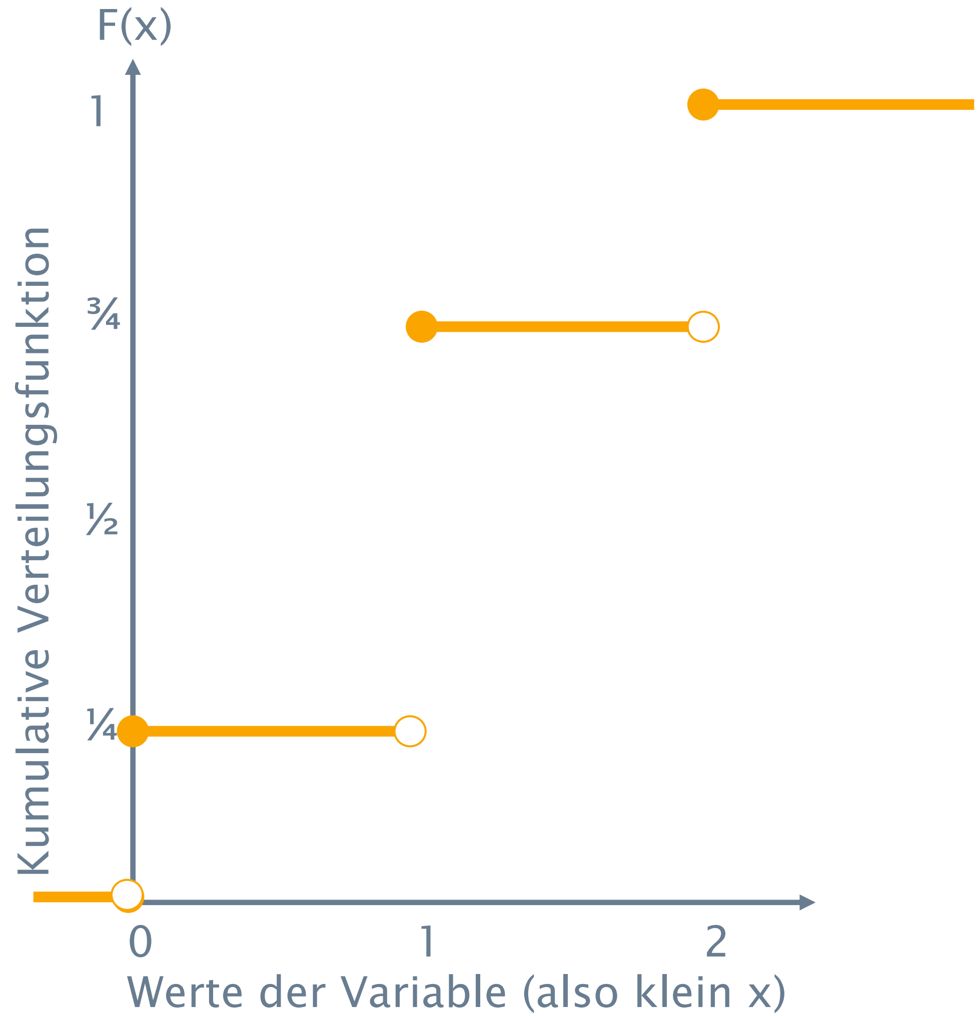
Fragen: Was ist die Wahrscheinlichkeit, dass Sie:

1. **genau ein Mal Kopf erhalten?**
2. **höchstens einmal Kopf erhalten?**
3. **mehr als einmal Kopf erhalten?**
4. **weniger als einmal Kopf erhalten?**
5. **... usw. usf.**

Beispiel III: faire Münze

1. **P(X = 1) = ½**





Beispiel III: faire Münze

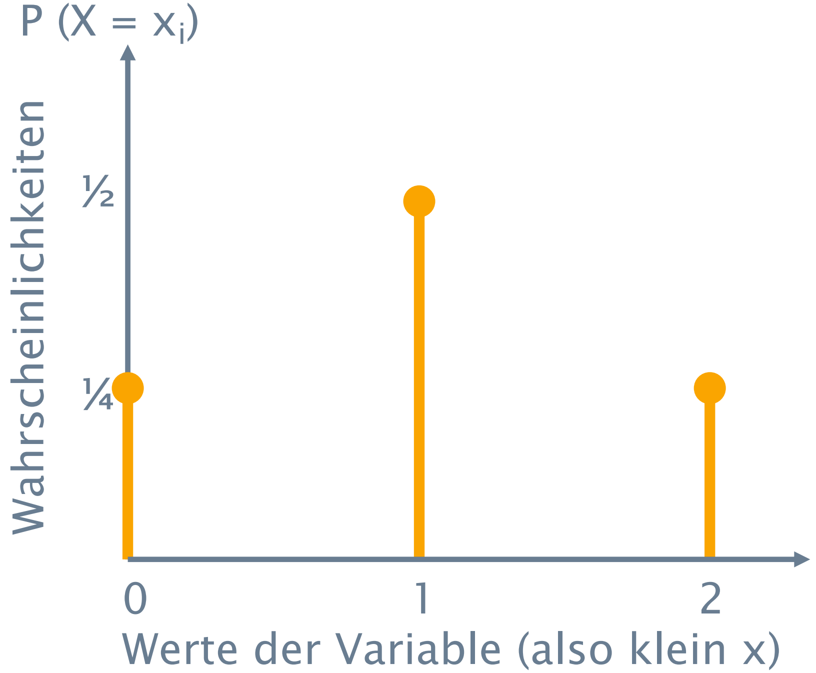
Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

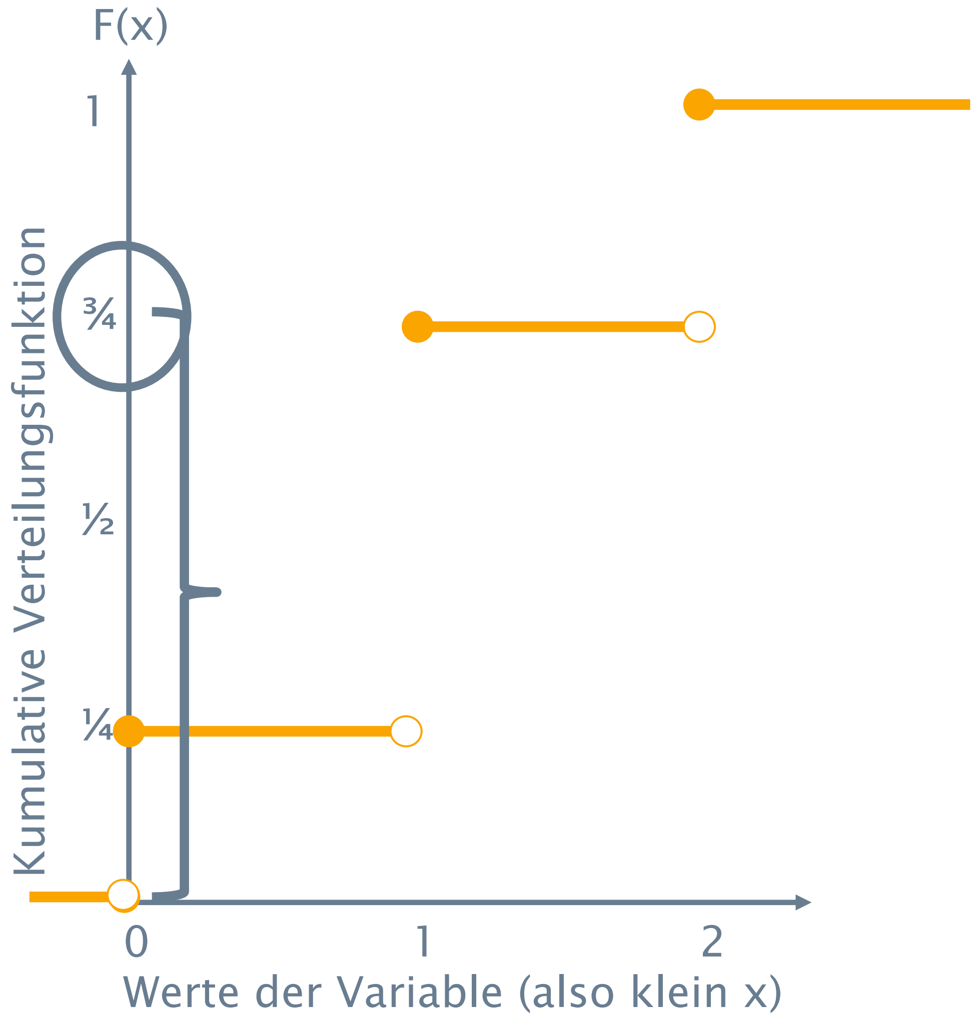
Fragen: Was ist die Wahrscheinlichkeit, dass Sie:

1. **genau ein Mal Kopf erhalten?**
2. **höchstens einmal Kopf erhalten?**
3. **mehr als einmal Kopf erhalten?**
4. **weniger als einmal Kopf erhalten?**
5. **... usw. usf.**

Beispiel III: faire Münze

2. **$F(1) = P(X \leq 1) = ¾$**





Beispiel III: faire Münze

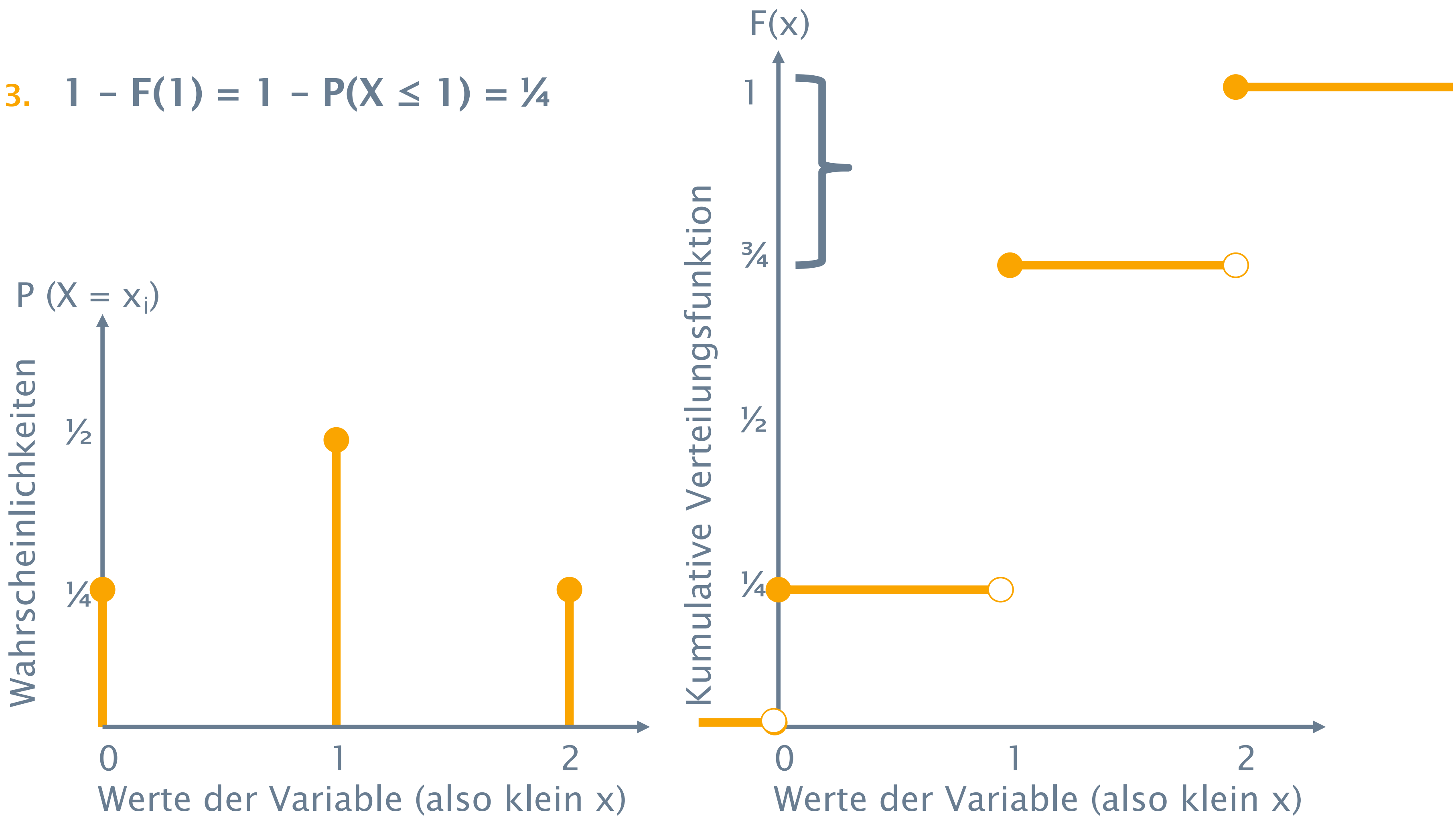
Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

Fragen: Was ist die Wahrscheinlichkeit, dass Sie:

1. **genau ein Mal Kopf erhalten?**
2. **höchstens einmal Kopf erhalten?**
3. **mehr als einmal Kopf erhalten?**
4. **weniger als einmal Kopf erhalten?**
5. **... usw. usf.**

Beispiel III: faire Münze

3. **$1 - F(1) = 1 - P(X \leq 1) = ¼$**



Beispiel III: faire Münze

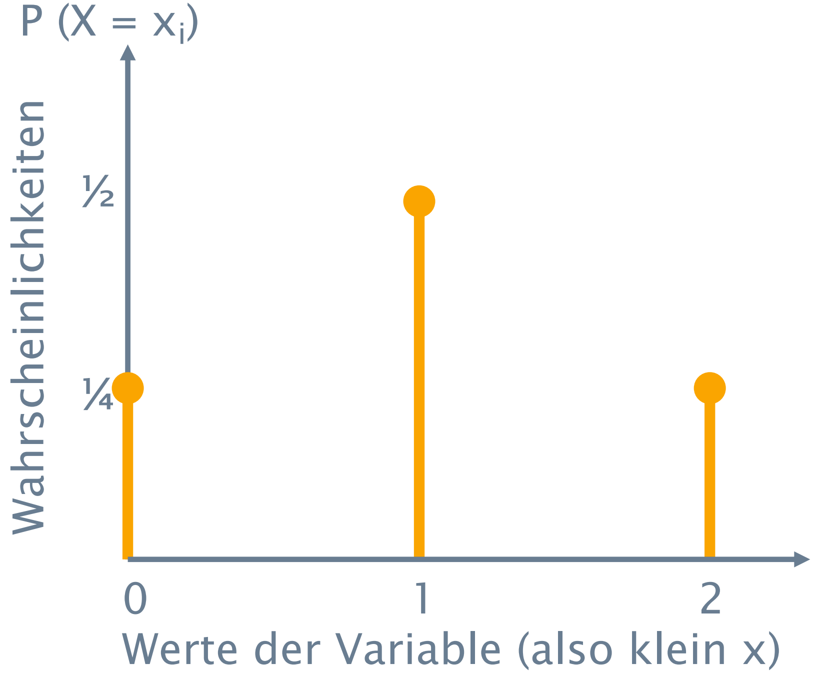
Angenommen Sie werfen eine faire Münze 2 mal nach einander.
Die Zufallsvariable X bezeichnet die Anzahl Kopf.

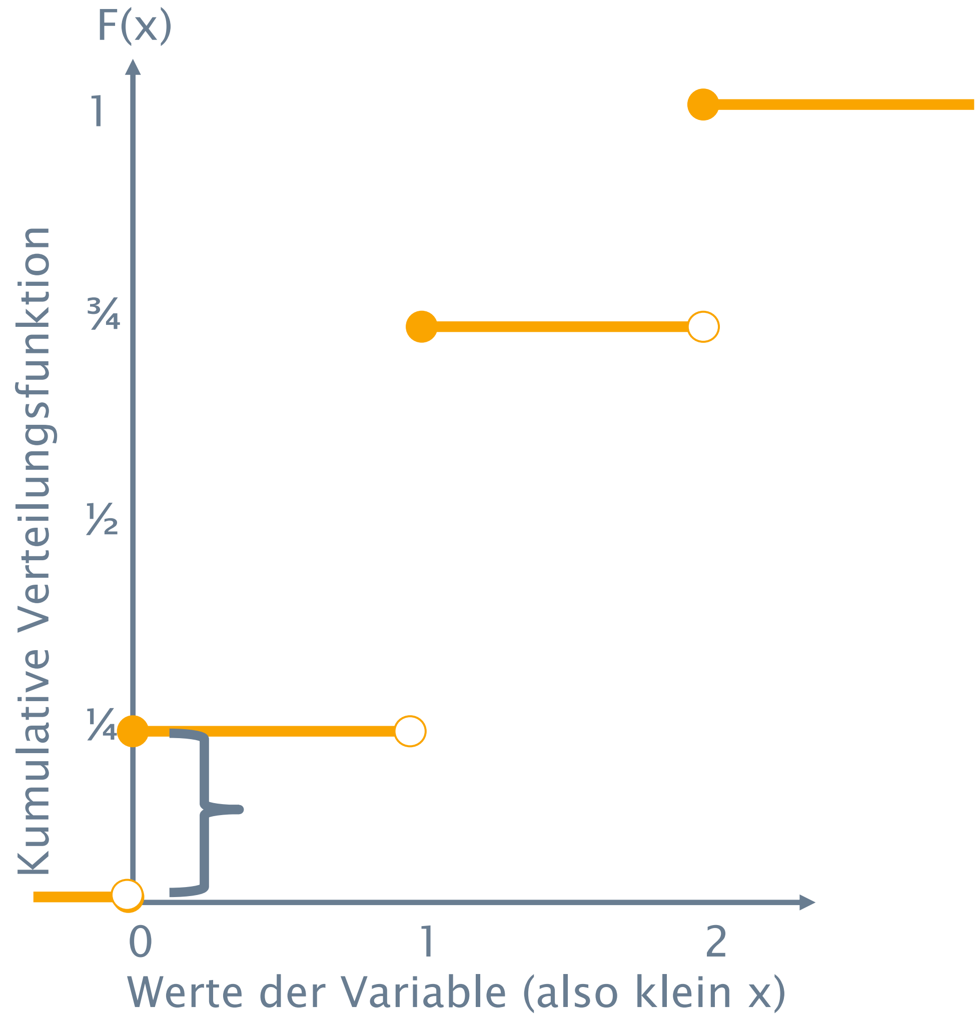
Fragen: Was ist die Wahrscheinlichkeit, dass Sie:

1. **genau ein Mal Kopf erhalten?**
2. **höchstens einmal Kopf erhalten?**
3. **mehr als einmal Kopf erhalten?**
4. **weniger als einmal Kopf erhalten?**
5. **... usw. usf.**

Beispiel III: faire Münze

4. **F(0) = P(X < 1) = ¼**



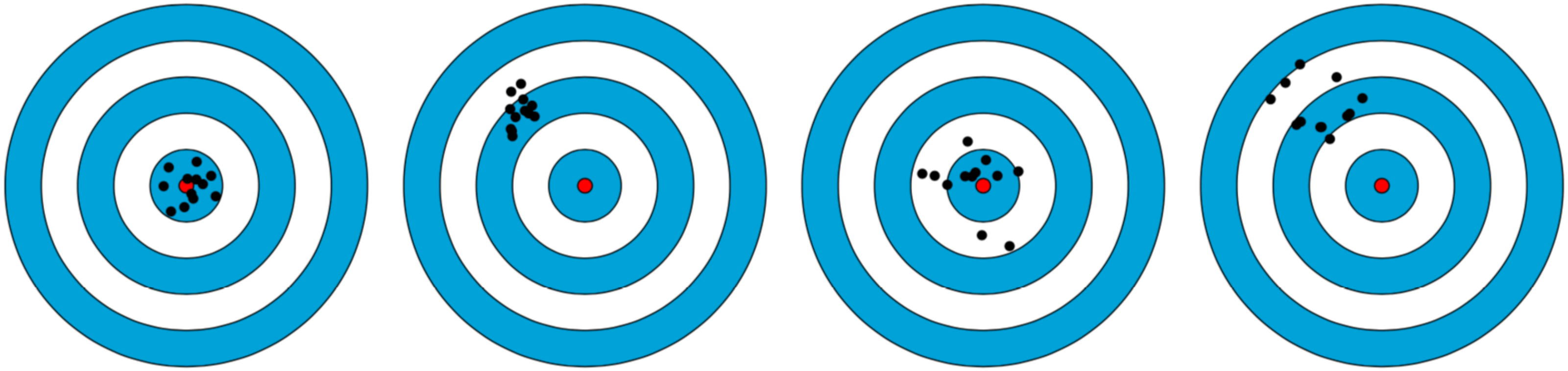


Expectation value, variance & standard deviation

Grundlegendes Problem in der Statistik

Wir möchten

- den roten Punkt in der Mitte bestimmen,
- und die Streuung quantifizieren.



(a) präzise und richtig, gute Genauigkeit

(b) präzise, aber nicht richtig, schlechte Genauigkeit

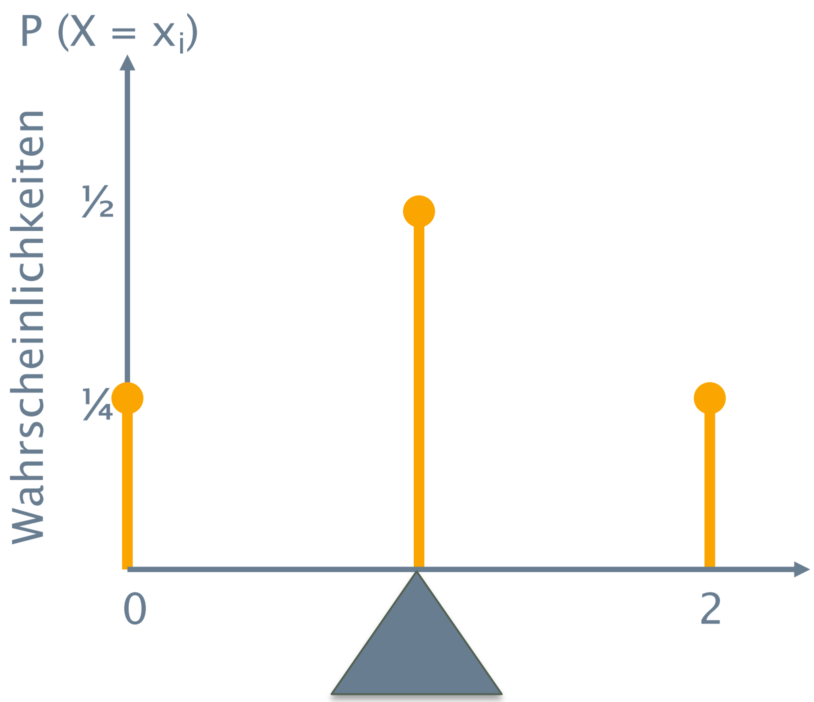
(c) richtig, aber nicht präzise, schlechte Genauigkeit

(d) weder präzise noch richtig, schlechte Genauigkeit

Abbildung 1.1.: Illustration der Präzision, Richtigkeit und Genauigkeit

Erwartungswert

Der Erwartungswert berechnet sich als nach Wahrscheinlichkeit gewichtetes Mittel der Werte, die die Zufallsvariable annimmt. Der Erwartungswert einer Zufallsvariablen kann als Schwerpunkt der Wahrscheinlichkeitsmasse betrachtet werden.



$$E(X) = \sum_{k=0}^{2} k \bullet P(X = k)$$
$$= 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4}$$
$$= 1$$

Varianz

Die Varianz ist ein Mass für die Streuung der Wahrscheinlichketsverteilung um ihren Schwerpunkt. Sie kann physikalisch als Trägheitsmoment interpretiert werden.

$$
\begin{aligned}
V(X) &= \sum_{k=0}^{2} (k - E(X))^2 \cdot P(X = k) \\
&= (0-1)^2 \cdot \tfrac{1}{4} + (1-1)^2 \cdot \tfrac{1}{2} + (2-1)^2 \cdot \tfrac{1}{4} \\
&= \tfrac{1}{4} + \tfrac{1}{4} \\
&= \underline{\tfrac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
V(X) &= \left[\sum_{k=0}^{2} k^2 \cdot P(X = k)\right] - (E(X))^2 \\
&= 0^2 \cdot \tfrac{1}{4} + 1^2 \cdot \tfrac{1}{2} + 2^2 \cdot \tfrac{1}{4} - 1^2 \\
&= \underline{\tfrac{1}{2}}
\end{aligned}
$$

Standardabweichung

Varianz hat leider nicht die gleiche Einheit wie der Erwartungswert, die Standardabweichung schon! Die Standardabweichung σ ist das wichtigste Mass für die Streuung.

$$\sigma = \sqrt{V(X)}$$

Für unser Beispiel erhalten wir

$\sigma = \sqrt{0.5} \approx 0.707$

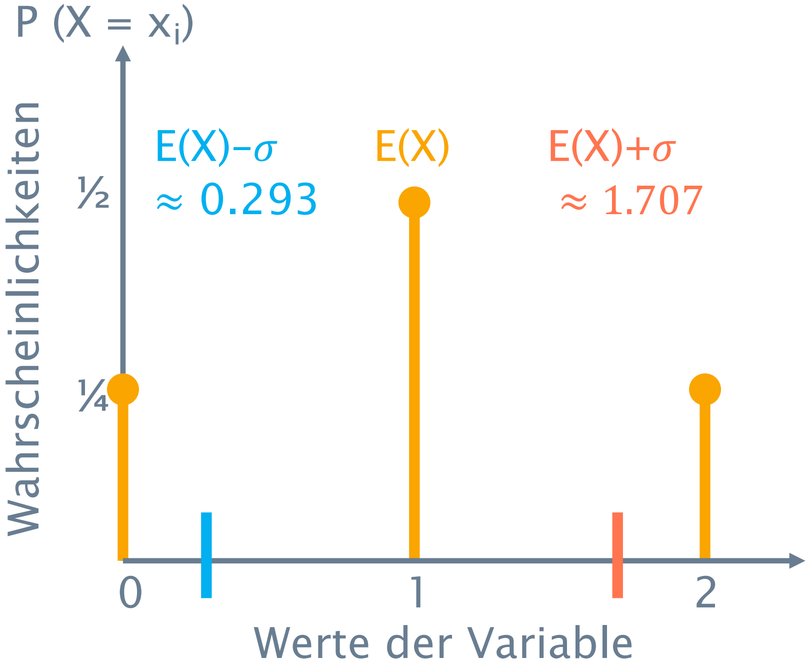
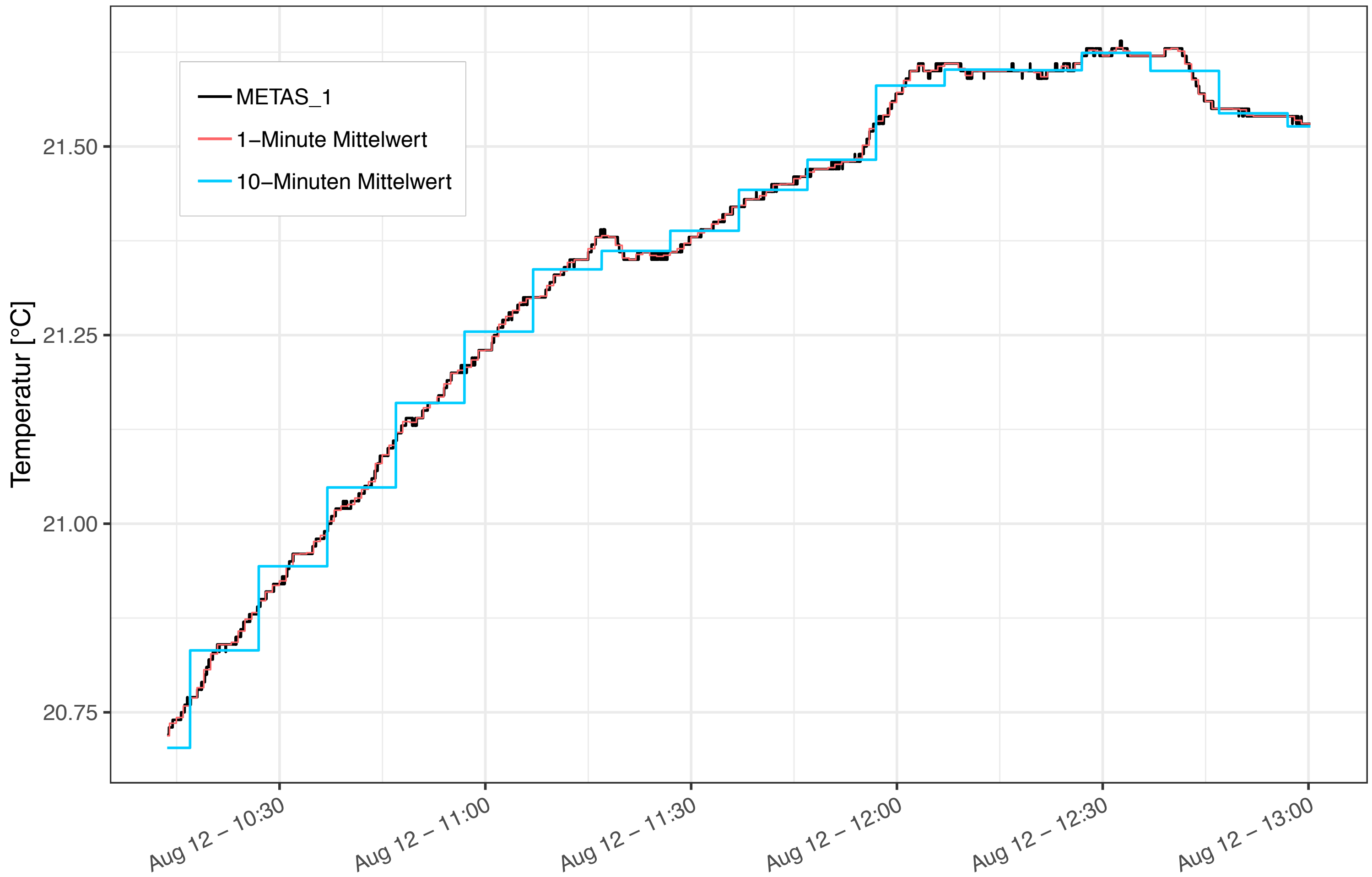


Illustration: Schwerpunkt und Trägheitsmoment

[Quelle]

[Quelle]

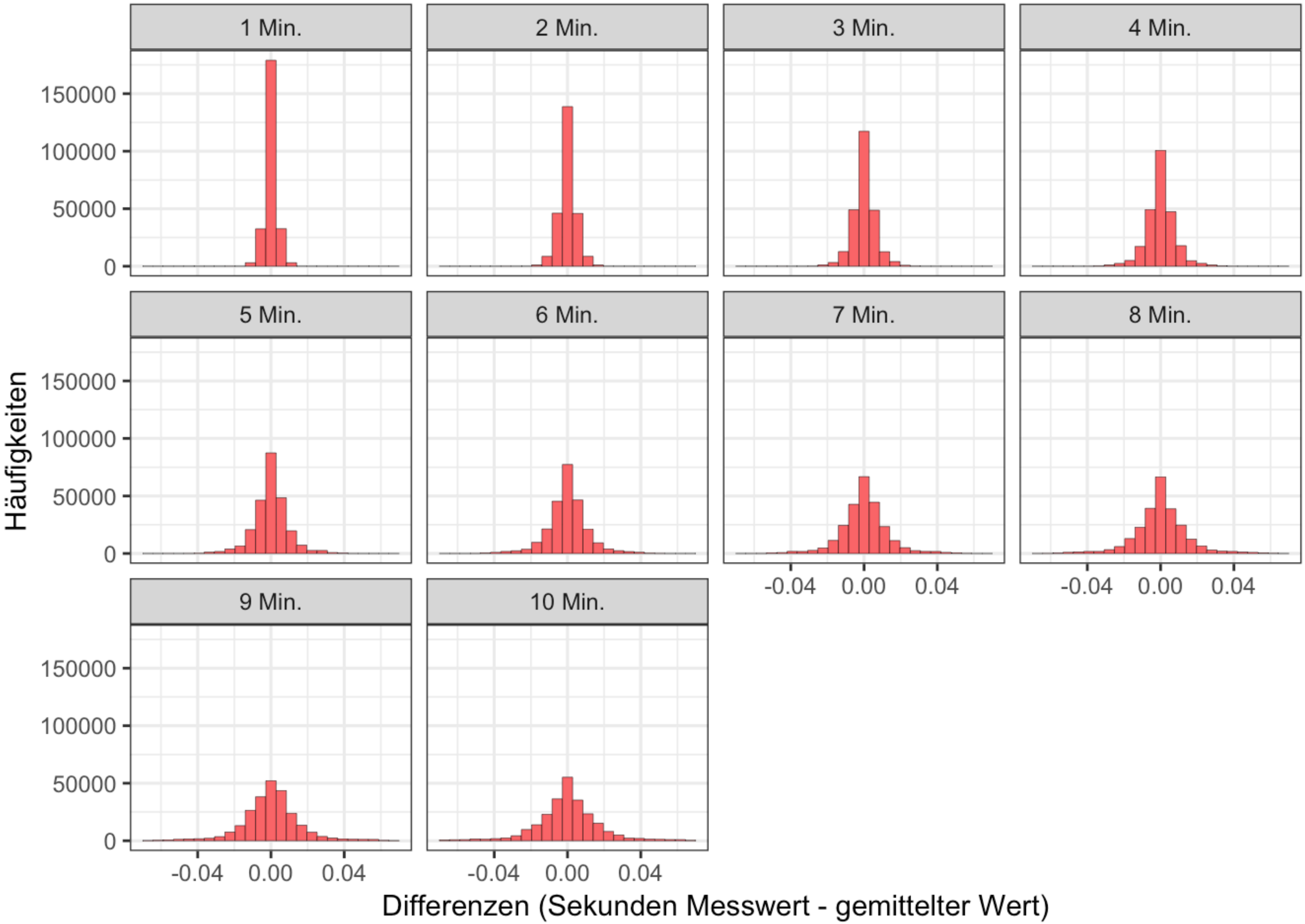
Beispiel: Abtastrate bei Temperaturmessungen



Beispiel: Abtastrate bei Temperaturmessungen

Messwert T_{sek}	18	19	20	19	19.5	18	$\cdots$	$\cdots$	$\cdots$
Mittelwert $\overline{T}$		19			19.5			$\cdots$	
Differenz $T_{sek} - \overline{T}$	-1	0	1	-0.5	0	-1.5	$\cdots$	$\cdots$	$\cdots$

Tabelle 3.3.: Schematische Darstellung der Mittelwertbildung bzw. der Differenzbildung. Hier werden 3 Messwerte gemittelt und die Differenz zum wahren Messwert wird berechnet.


1 Min.
2 Min.
3 Min.
4 Min.
5 Min.
6 Min.
7 Min.
8 Min.
9 Min.
10 Min.
150000
100000
50000
0
-0.04
0.00
0.04
Häufigkeiten
Differenzen (Sekunden Messwert - gemittelter Wert)

Beispiel: Abtastrate bei Temperaturmessungen

s bezeichnet die empirische Standardabweichung

Auflösung	ca.68.3% der Differenzen liegen im Intervall $\pm s$	ca.95% der Differenzen liegen im Intervall $\pm 2 \cdot s$	ca.99.7% der Differenzen liegen im Intervall $\pm 3 \cdot s$
1 Min.	± 0.003	± 0.006	± 0.010
2 Min.	± 0.005	± 0.009	± 0.014
3 Min.	± 0.006	± 0.012	± 0.018
4 Min.	± 0.008	± 0.015	± 0.023
5 Min.	± 0.009	± 0.018	± 0.028
6 Min.	± 0.011	± 0.022	± 0.033
7 Min.	± 0.013	± 0.025	± 0.038
8 Min.	± 0.014	± 0.028	± 0.043
9 Min.	± 0.016	± 0.032	± 0.047
10 Min.	± 0.018	± 0.035	± 0.053

Tabelle 3.4.: Übersicht der Genauigkeit in Abhängigkeit der zeitlichen Auflösung für die Metas-Sonde

Dependant & independent random variables

Abhängigkeit und Unabhängigkeit von Zufallsvariablen

Angenommen X und Y seinen 2 Zufallsvariablen.

Der Erwartungswert der Summe dieser Variablen ist sehr einfach zu berechnen

$$E(X + Y) = E(X) + E(Y)$$

Es ist egal, ob X und Y abhängig oder unabhängig voneinander.

Falls X und Y unabhängig sind, gilt auch für die Varianz

$$V(X + Y) = V(X) + V(Y)$$

Falls X und Y abhängig sind, gilt diese Formel nicht mehr. Wir erhalten einen zusätzlichen Term, 2 * Kovarianz, den wir in 1 bis 2 Wochen genauer kennenlernen werden.

Exercises 4, 5, 6 & 7 from the series 5

Exercise 3 from the exercise series 5

read pages 26 – 30 of the script to be found under week 2 on moodle

3 c) requires a deep understanding

Set theory
I forgot to mention that

- $\wedge$ is an “and”

- $\vee$ is an “or”

Thank you for your time and attention.