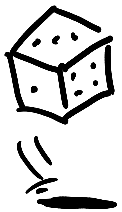
Grundlagen aus der Statistik

Bevor wir über ML reden können brauchen wir ein paar elementare Grundlagen & Begriffe aus der Statistik.

1.1) Zufallsexperimente & Zufallsvariablen

Prozesse, die in der realen Welt ablaufen lassen sich mit Hilfe des folgenden Konzepten beschreiben:

ZUFALLSEXPERIMENT	ELEMENTAREREIGNISSE	EREIGNISRAUM
 Würfeln	mögl. Ergebnisse $\omega_1=1, \omega_2=2, \dots, \omega_6=6$	$\Omega=\{\omega_i \mid i=1,2,\dots,6\}$ $\Omega=\{1,2,3,4,5,6\}$

Mit welcher Wahrscheinlichkeit beobachtet man ein bestimmtes ω_i ??

SATZ ⇒ Satz von Laplace

Sind alle ω_i $(i=1,\dots,n)$ gleich wahrscheinlich, dann gilt:

$$\boxed{P(\omega_i) = \frac{1}{|\Omega|}} \quad i=1,\dots,n \qquad |\Omega| = n \leftarrow \text{Mächtigkeit des Ereignisraumes } \Omega$$

Bsp.: Glücksrad

Experiment: "Rad drehen"

ω_i $(i=1,2,3)$

ω_1 = Sonne
ω_2 = Haus
ω_3 = Blume

$\Omega = \{\text{Sonne}, \text{Haus}, \text{Blume}\}$

$|\Omega| = 3$

$P(\omega_i) = \frac{1}{3}$ für $i = 1,2,3$

ABER ⇒ Wie beschreibt man komplexere Abläufe?

Würfelspiel: "Einsatz = 3 €, Gewinn: gewürfelte Augenzahl in €"

Beobachtung: Höhe des Gewinnes hängt vom Experiment "Würfeln" und $P(\omega_i)$ ab...

Def.: Zufallsvariable (ZV)

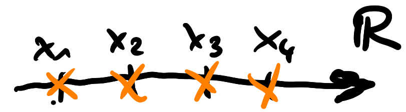
Eine **ZV X** ist eine Abbildung

$$\boxed{\begin{aligned} X: \Omega &\to \mathbb{R} \\ \omega_i &\mapsto X(\omega_i) = x_i \in \mathbb{R} \end{aligned}}$$

Jedem Elementarereignis ω_i wird ein konkreter Wert $x_i \in \mathbb{R}$ zugeordnet.
Die $x_i \in \mathbb{R}$ heißen **Realisierungen** der ZV X.

Wir unterscheiden 2 Szenarien:

- einzelne $x_i \in \mathbb{R}$ ⇨ DISKRETE ZV X



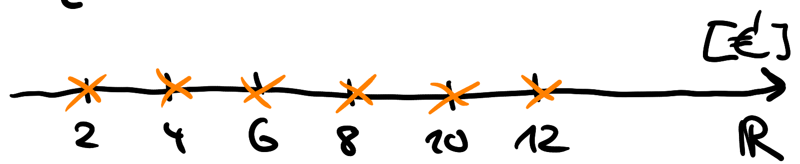
- kontinuierliche x_i ⇨ STETIGE ZV X

Bsp. 1 Würfelspiel (diskret)

$X \hat{=}$ "Augenzahl ×2 in €"

$\omega_1 = 1 \longmapsto X(\omega_1) = 2 \cdot 1 = 2€$

$\omega_2 = 2 \longmapsto X(\omega_2) = 2 \cdot 2 = 4€$

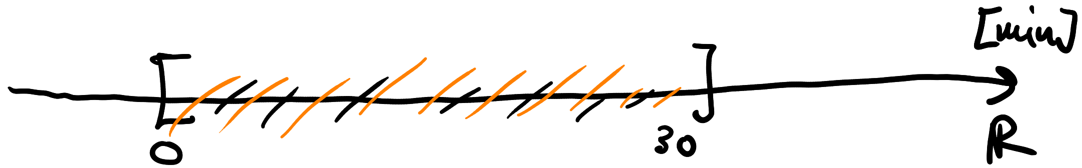


Bsp. 2 Wartezeit (stetig)

$X \hat{=}$ "Zeit, bis der nächste Bus kommt"

$\omega_1 = 0 \longmapsto X(\omega_1) = 0$ min

$\vdots$



FRAGE Wie bestimmt man die Wahrscheinlichkeiten der Realisierungen $x_i \in \mathbb{R}$?

$\rightsquigarrow$ unterscheide zwischen diskreten und stetigen ZV ...

1.2) W-Funktion & Verteilungsfunktion (DISKRETE ZV)

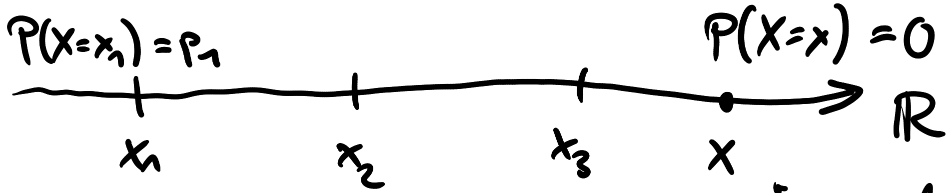
Def.: W-Funktion

Sei X eine diskrete ZV, $x \in \mathbb{R}$, $x_i \in \mathbb{R}$ $(i = 1, 2, \dots, n)$ Realisierungen von X

Dann heißt

$$f_X(x) = \begin{cases} P(X = x_i) = p_i & \text{für } x = x_i \quad (i = 1, 2, \dots n) \\ 0 & \text{sonst} \end{cases}$$

W-Funktion von X. Die W-Funktion ordnet jedem $x \in \mathbb{R}$ seine Wahrscheinlichkeit zu



wird nicht beobachtet!

Es gilt: $\boxed{\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} p_i = 1}$ und $\boxed{0 \leq f_X(x_i) = p_i \leq 1}$ Axiome von Kolmogorow!!

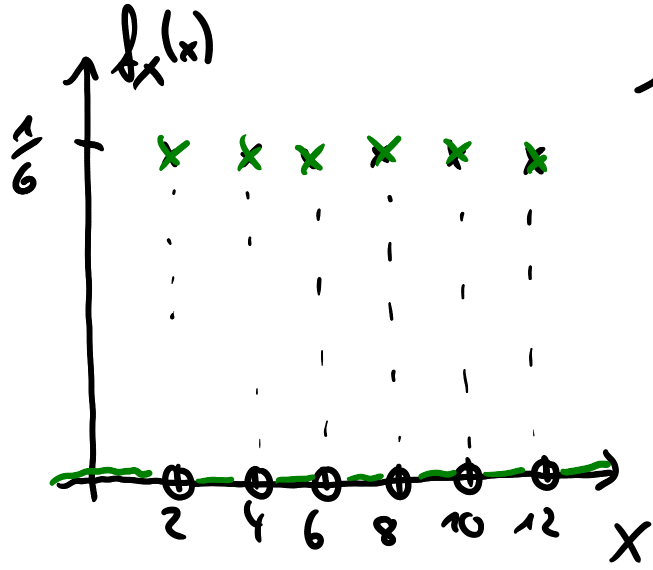
Die Funktion $\boxed{F_X(x) = P(X \le x) = \sum_{x_i \le x} f_X(x_i)}$ heißt Verteilungsfunktion von X.

Bsp. 1) Fortsetzung

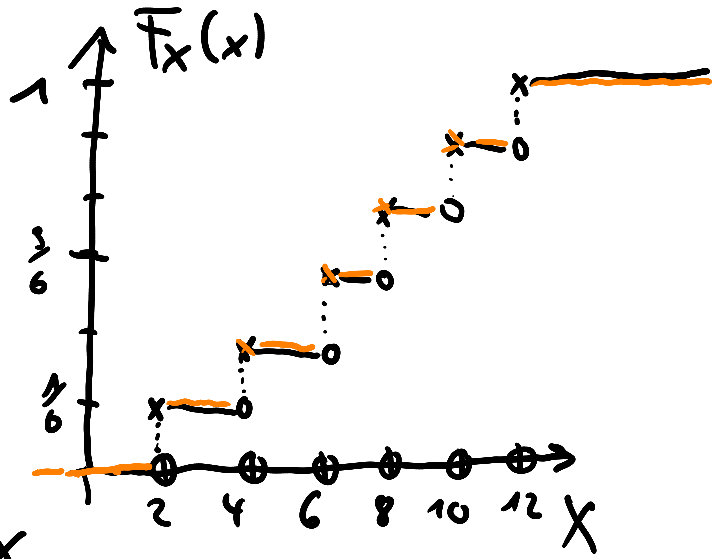
$X \hat{=}$ "Augenzahl × 2 in €"

$x_i \in \{2, 4, 6, 8, 10, 12\}$

$P(X = x_i) = \frac{1}{6}$



W-Funktion von X



Verteilungsfunktion von X

Bsp. 2) Wartezeit (Forts.)

$X =$ "Zeit bis der nächste Bus kommt"

$\omega_i \in [0; 30] \subset \mathbb{R}$

FRAGE ⇒ Welche Wahrscheinlichkeit hat

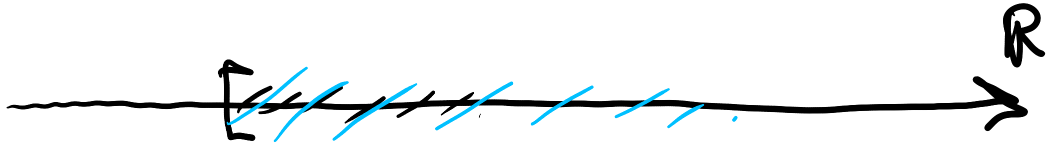
$x_i = 23$ min oder

$x_i = 12$ min 20 s. ??

offensichtlich nicht so einfach, denn ∞ viele x_i ...

1.3) W-Dichte & Verteilungsfunktion (STETIGE ZV)

Für stetige ZV X gilt: $x_i \in I \subseteq \mathbb{R}$



Die Realisierungen kommen aus einem Kontinuum, es gibt also ∞ viele davon!

$\Rightarrow$ man kann einem einzelnen $x_i \in I \subseteq \mathbb{R}$ keine sinnvolle Wahrscheinlichkeit zuordnen!

Die Idee von der W-Funktion klappt hier nicht...

Denn:

wäre $P(X=x_i) = p_i \in [0;1]$ für $x_i \in I \subseteq \mathbb{R}$

dann müßte gelten $\boxed{\sum_{i=1}^{\infty} p_i = 1}$

Wir brauchen eine andere Idee!

⇨ Def: W-Dichte

Sei X eine stetige ZV, $x \in \mathbb{R}$. Dann lässt sich X durch eine reelle Funktion mit den folgenden Eigenschaften beschreiben:

① $\boxed{f_X(x) \geq 0}$

② $\boxed{\int_{-\infty}^{\infty} f_X(x)\,dx = 1}$

$f_X(x)$ heißt <u>W-Dichte von X</u> und es gilt:

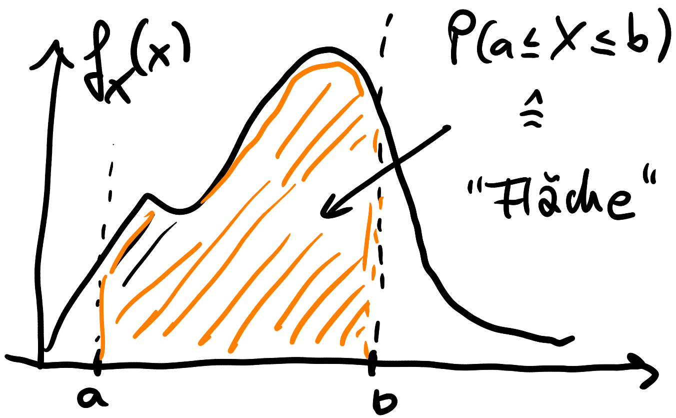
$$\boxed{P(a \leq X \leq b) = \int_a^b f_X(x)\,dx} \qquad a,b \in \mathbb{R}$$

Für die <u>Verteilungsfunktion von X</u> folgt:

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt}$$

und es gilt:

$$\boxed{P(a \leq X \leq b) = F_X(b) - F_X(a)}$$



Bemerkung: Die W-Dichte (W-Funktion) und die Verteilungsfunktion beschreiben die ZV X vollständig & gleichwertig

Bsp. 3) TK-Kost

Ein Kühllaster transportiert TK-Kost.
Die Temperatur darf nie über −18°C steigen,
die Aggregate haben aber eine Toleranz von 2°C. Es wird also auf −20°C gekühlt, damit eventuelle Abweichungen noch ok. sind.

$X \hat{=}$ "Abweichungen vom Soll in °C" mit W-Dichte $f_X(x) = \begin{cases} \frac{1}{12}(8-x^3) & x \in [0;2] \\ 0 & \text{sonst} \end{cases}$

a) Ist $f_X(x)$ tatsächlich eine zulässige W-Dichte?

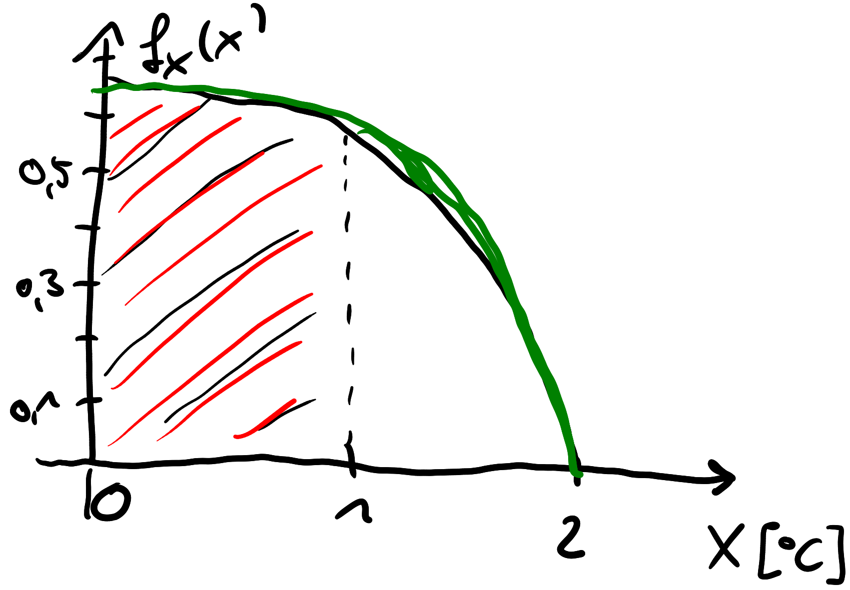
b) Wie lautet die Verteilungsfunktion $F_X(x)$?

c) Wie wahrscheinlich sind Abweichungen von 0 bis 1°C?

Lösung:

a) zu prüfen ist ① $f_X(x) \geq 0 \quad (x \in \mathbb{R})$

② $\int_{-\infty}^{\infty} f_X(x) = 1$



① $\frac{1}{12} \geq 0$ und $x^3 \in [0;8]$ für $x \in [0;2]$ $\Rightarrow$ $8-x^3 \geq 0$ ✓

② $\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{2} \frac{1}{12}(8-x^3)\,dx + \int_{2}^{\infty} 0\,dx = \frac{1}{12}\; 8x - \frac{1}{4}x^4 \Big|_0^2 =$

$= \frac{1}{12}[16-4] = \frac{12}{12} = 1$ ✓

Aus ① + ② folgt: $f_X(x)$ ist eine zulässige W-Dichte!

b) $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} \frac{1}{12}(8-t^3)\,dt = \frac{1}{12}\left[8t - \frac{1}{4}t^4 \Big|_0^x\right] =$

$= \frac{1}{12} 8x - \frac{1}{48}x^4 = \frac{2}{3}x - \frac{1}{48}x^4$ ← Verteilungsfunktion von X

c) $P(0 \leq X \leq 1) = \int_{0}^{1} f_X(x)\,dx = \frac{1}{12}\int_{0}^{1}(8-x^3)\,dx = \frac{1}{12}\left[8x - \frac{1}{4}x^4 \Big|_0^1\right] =$

$= \frac{1}{12}\left(8 - \frac{1}{4}\right) = \frac{31}{48} = 0{,}6458 \mathrel{\hat{=}} 64{,}58\,\%$

Die Wahrscheinlichkeit beträgt 64,58 %.

Es ist auch möglich die Verteilungsfunktion zu benutzen:

$P(0 \leq X \leq 1) = F_X(1) - F_X(0) = \left(\frac{2}{3} - \frac{1}{48}\right) - 0 = 0{,}6458$

⇨ Beide Lösungen sind absolut gleichwertig! ☺

1.4) Erwartungswert & Varianz

ZV lassen sich über ihre W-Dichte (W-Funktion) bzw. die Verteilungsfunktion vollständig beschreiben. Sie haben aber auch charakteristische Merkmale, die schon sehr viel über sie aussagen.

⇨ ERWARTUNGSWERT (Lagemaß)

VARIANZ (Streuungsmaß)

⇨ Def.: Erwartungswert

Sei X eine ZV mit Realisierungen $x_i \in \mathbb{R}$.
Dann ist der Erwartungswert $E(X)$ definiert als:

DISKRETE ZV ⇨ $$E(X) = \sum_{i=1}^{n} x_i f_X(x_i)$$

← Mittelwert der Realisierungen gewichtet mit ihren Wahrscheinlichkeiten

STETIGE ZV ⇨ $$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

← $E(X)$ hat die Einheit der ZV
z.B. €, $\frac{m}{s}$, $, ...

Rechenregeln:

X, Y ZV

$a, b, c \in \mathbb{R}$

$$\boxed{\begin{array}{ll} E(a) = a & E(X+Y) = E(X) + E(Y) \\ E(a+X) = a + E(X) & E(aX + bY + c) = a\,E(X) + b\,E(Y) + c \\ E(a \cdot X) = a \cdot E(X) & \end{array}}$$

Bsp. 3) Fortsetzung

$f_X(x) = \frac{1}{12}(8 - x^3)$

$x \in [0; 2]$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = \int_0^2 x \cdot \frac{1}{12}(8 - x^3)\, dx = \frac{1}{12} \int_0^2 8x - x^4\, dx =$$

$$= \frac{1}{12}\left[8 \cdot \frac{1}{2}x^2 - \frac{1}{5}x^5 \Big|_0^2 \right] = \frac{4}{5} = 0{,}8$$

$\Rightarrow$ Es ist erwartungsgemäß mit einer Abweichung von 0,8 °C zu rechnen.

Merke Der Erwartungswert ist also derjenige Wert, den man "den wahrscheinlichsten" beobachtet.

In der deskriptiven Statistik: Lagemaß $\rightarrow$ Mittelwert, $\rightarrow$ Median $\leftarrow$ interessant auch die Streuung $\Rightarrow$ Standardabweichung

Def.: Varianz

Die Varianz einer ZV X entspricht der erwarteten quadratischen Abweichung der möglichen Werte vom Erwartungswert.

DISKRETE ZV $\Rightarrow$ $\boxed{V(X) = \sum_{i=1}^{k} (x_i - E(X))^2 \cdot f_X(x)}$ $\leftarrow$ einheitenlos!

STETIGE ZV $\Rightarrow$ $\boxed{V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f_X(x)\, dx}$

$\sigma_X = \sqrt{V(X)}$ heißt Standardabweichung von X (Einheit von X)

Für zwei unabhängige ZV X, Y ($P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$) gelten die folgenden Rechenregeln:

$a \in \mathbb{R}$

X, Y unabhängig

$$\boxed{\begin{array}{ll} V(a) = 0 & V(X+Y) = V(X) + V(Y) \\ V(a+X) = V(X) & \\ V(a \cdot X) = a^2 V(X) & \end{array}}$$

Bsp. 3) Fortsetzung

$f_X(x) = \frac{1}{12}(8-x^3)$

$E(X) = \frac{4}{5}$

$$V(X) = \int_0^2 \left(x-\frac{4}{5}\right)^2 \cdot \frac{1}{12}(8-x^3)\,dx = \dots = 0{,}2489$$

$$\Rightarrow \sigma_X = \sqrt{V(X)} = \sqrt{0{,}2489} = 0{,}4989 \ [°C]$$

Die Beschreibung von Abläufen, die von Zufall geprägt sind wird also mit Hilfe von ZV gemacht.

Allerdings: Die ZV X selbst ist ein abstraktes Konstrukt

In der Realität gilt:

$$\boxed{\begin{array}{l} \text{ZV } X \\ X: \Omega \to \mathbb{R} \\ \quad \omega_i \mapsto X(\omega_i) = x_i \in \mathbb{R} \end{array}}$$

Realisierungen der ZV

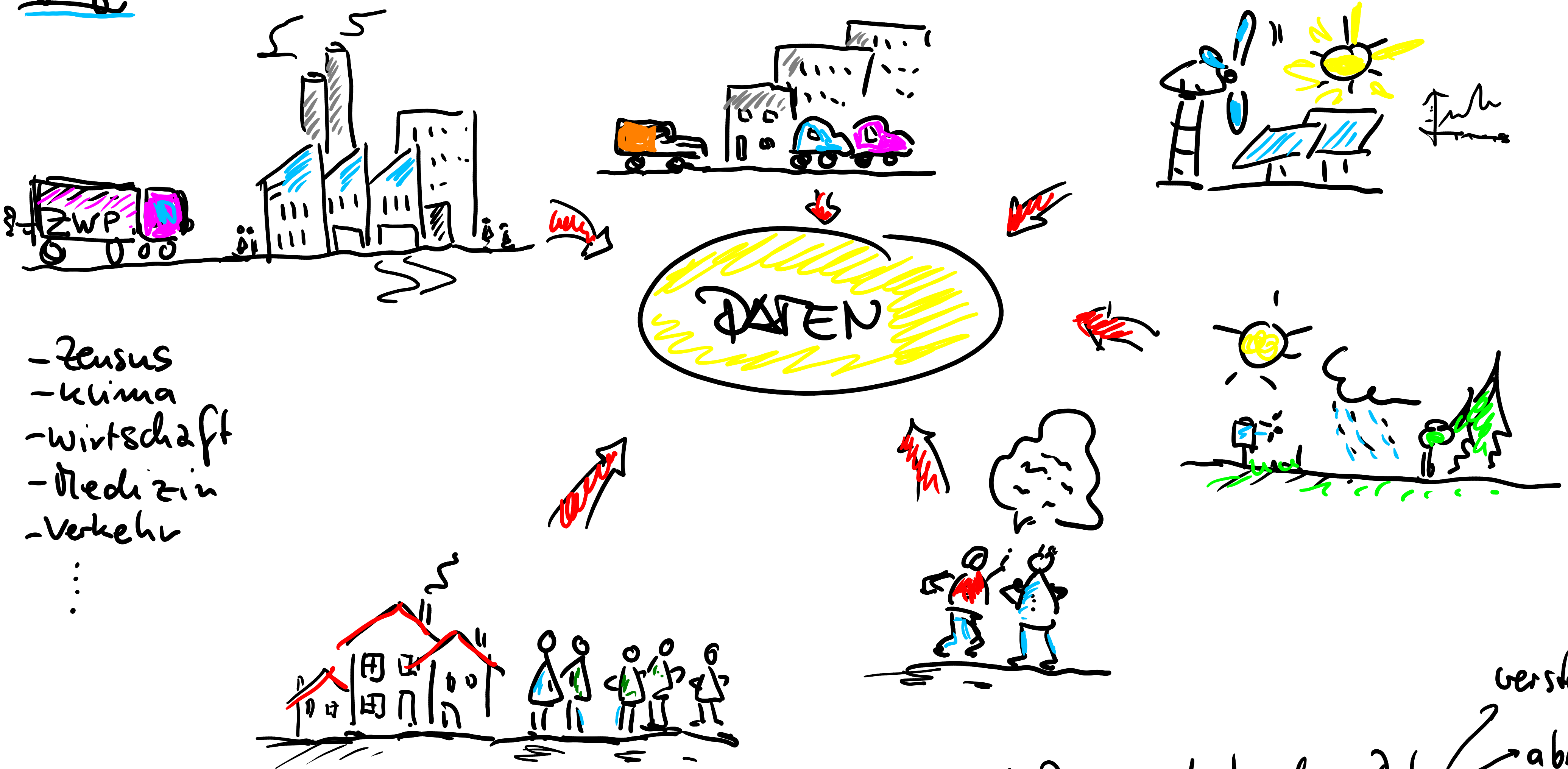
wir „sehen“ in der echten Welt immer nur die REALISIERUNGEN $x_i \in \mathbb{R}$

$\boxed{\text{DATEN}}$

2. Induktive Statistik – "Lernen aus Daten"

Bisher: theoretischer Unterbau & Konzepte

Jetzt: wir sehen \ beobachten \ messen ⇒ Daten! und was weiter??



- Zensus
- Klima
- Wirtschaft
- Medizin
- Verkehr

⋮

... wir wollen die Prozesse hinter den Daten verstehen / abbilden

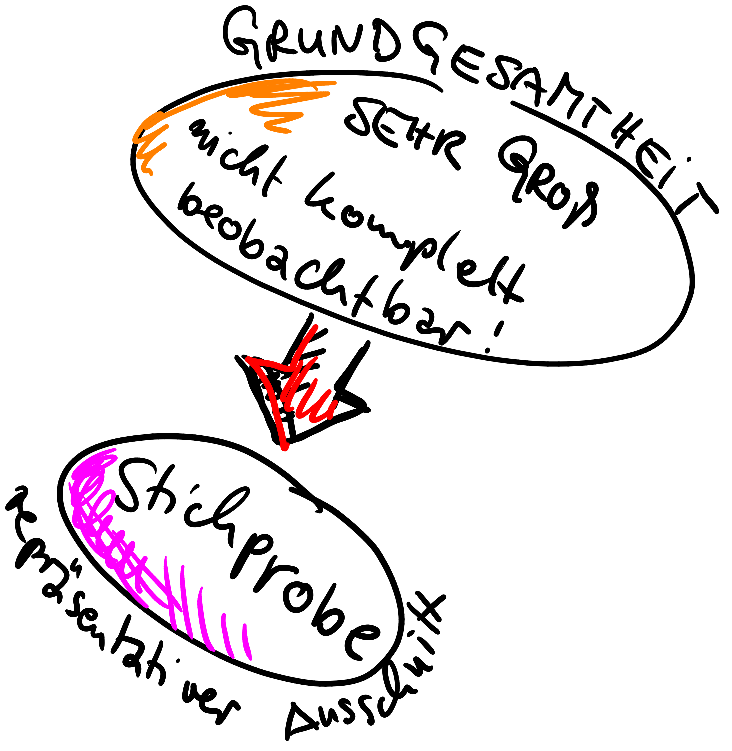
KONZEPT: hinter jeder Beobachtung steckt eine ZV X mit einer beschreibenden Verteilungsfunktion.
Die beobachteten Daten entsprechen (den) Realisierungen dieser ZV X.

ZIEL ⇒ Ziehe aus den Beobachtungen Rückschlüsse auf die zugrundeliegende ZV:
→ Verteilungsfkt., Parameter?
→ $E(X) = ?$
→ $V(X) = ?$

⇒ bei mehrdimensionalen Größen $X = (X_1, X_2)$ oder 2 ZV X, Y geht es darum, Zusammenhangsstrukturen (mehrdim. Verteilungsfunktion) aufzudecken und zu modellieren.

2.1) Stichproben & Schätzer

Allgemein: Elemente einer (sehr großen) Grundgesamtheit haben eine bestimmte Eigenschaft



Bsp.: Personen Ü18 in Deutschland

Führerschein? → ja (1)
→ nein (0)

- repräsentativen Ausschnitt wählen
- Daten erheben
- Daten auswerten

} Rückschlüsse ziehen über ZV $X \hat{=}$ "Führerschein?"

Def.: Stichprobe

Eine n-dim ZV $X=(X_1,\dots,X_n)$ heißt <u>Stichprobe</u>, wenn die ZV X_i $(i=1,\dots,n)$ unabhängig und identisch verteilt gemäß der Grundgesamtheit sind.

Die <u>Stichprobenziehung</u> liefert eine <u>Realisierung</u> $x=(x_1,\dots,x_n)$ von X, wobei x_i die Realisierungen der X_i $(i=1,\dots,n)$ sind.

Die Dimension n der Stichprobe heißt <u>Stichprobenlänge</u>.

$$\left.\begin{array}{ll} E(X)=\mu & \text{Erwartungswert} \\ V(X)=\sigma^2 & \text{Varianz} \end{array}\right\} \text{in der GRUNDGESAMTHEIT}$$

<u>unbekannt</u>

müssen aus den Daten (Realisierungen $x=(x_1,\dots,x_n)$) geschätzt werden.

Bsp: Führerschein

Grundgesamtheit:	Stichprobe der Länge	Zahlung
alle Ü-18	$n = 1000$	$x = (1,0,\dots,1,1,0)$
	repräsentativ: X_i unabh.	Realisierung/Daten
	gleiche Frage: X_i id. verteilt	

FRAGE Wie schätzt man den Erwartungswert $E(X)$, also die Wahrscheinlichkeit für eine "ja"-Antwort bei einer beliebigen Person?

... was würden Sie mit der Realisierung $x = (1,0,\dots,1,1,0)$ anfangen??

Def.: Schätzfunktion & Schätzung

Sei Θ ein beliebiger unbekannter Parameter der Verteilung $F_\Theta(x)$ und $X=(X_1,\dots,X_n)$ eine Stichprobe mit Realisierung $x=(x_1,\dots,x_n)\in\mathbb{R}^n$. Dann heißt

$\boxed{T_\Theta(X)=T_\Theta(X_1,\dots,X_n)}$ **Schätzfunktion** von Θ (ZV)

$\boxed{\hat{\Theta}=T_\Theta(x_1,\dots,x_n)=t_\Theta(x_1,\dots,x_n)}$ **Schätzung** von Θ ($\hat{\Theta}\in\mathbb{R}$)

wichtige Eigenschaft: **ERWARTUNGSTREUE**

$\boxed{E(T_\Theta(X))=\Theta}$ für ∞ viele Wiederholungen der Schätzung konvergiert der Mittelwert der $\hat{\Theta}$ gegen das "wahre" Θ !!

... Schätzer sind oft so konstruiert, daß sie erwartungstreu sind. Falls nicht gibt es einen BIAS

$$\boxed{\text{bias}(T_\Theta(X)) = E(T_\Theta(X)) - \Theta}$$ ← Maß für die Verzerrung

Bsp.: Schätzung für Erwartungswert & Varianz

① Erwartungswert

$X = (X_1, \ldots, X_n)$ $\qquad E(X_i) = \mu \quad (i = 1, \ldots, n)$

$x = (x_1, \ldots, x_n)$ $\qquad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\boxed{T_\mu(X) = \frac{1}{n} \sum_{i=1}^{n} X_i} \rightsquigarrow \boxed{\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

Schätzer für $E(X)$ $\qquad$ Schätzung von μ

BIAS?

$E(T_\mu(X)) =$

② Varianz

$$\boxed{T_{\sigma^2}(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} \rightsquigarrow \boxed{\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

! wg. Erwartungstreue nötig!!

Bemerkung: Selbst wenn die Schätzfunktion erwartungstreu ist ($\Rightarrow$ bias $= 0$), gibt es immer noch einen Fehler beim Schätzen. Der hängt von der Varianz der Schätzfunktion ab!!

Def.: MSE – Mean Squared Error

falls $T_\Theta(x)$ erwartungstreu

$$\boxed{MSE(T_\Theta(X)) = E\left((T_\Theta(X) - \Theta)^2\right) = Var(T_\Theta(X)) + \underbrace{bias(T_\Theta(X))^2}_{}}$$

$bias = 0$

Bemerkung:

$$\boxed{\sigma_{T_\theta}(x) = \sqrt{Var(T_\theta(x))}}$$

heißt **Standardfehler der Schätzung**

Bsp. 4) Service - Center

i	1	2	3	4	5
# Anrufe	20	27	35	53	42

$X = (X_1, \dots, X_5) \quad n = 5$

$x = (20, 27, 35, 53, 42)$

$E(X) = \mu = ?$

$Var(X) = \sigma^2 = ?$

$\hat{\mu} = \frac{1}{5}(20 + 27 + \dots + 42) = 35{,}4$ (Anrufe)

$\hat{\sigma}^2 = \frac{1}{4}\left((20 - 35{,}4)^2 + \dots + (42 - 35{,}4)^2\right) = 165{,}3$

$\hat{\sigma} = 12{,}86$ (Standardfehler)

FRAGE

Welche Möglichkeiten hat man zusätzlich zum Standardfehler der Schätzung um die Unsicherheit der Schätzung zu bewerten?

⇒ Konfidenzintervalle

⇒ Hypothesentests

MEHR IN DER STATISTIK-VORESUNG