PTS Vorlesung – Kap 2

2 Zufallsvariablen (Random Variables)

Wirf 2 Würfel: Summe $X = w_1 + w_2$

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

ist eine Zufallsvariable.

Idee: Wir interessieren uns nicht für beliebige Ereignisse, sondern nur für solche, die mittels X ausgedrückt werden können

Etwa: Gewicht $\geq$ 100 kg

Höhe $<$ 1,60 m

Zurück zu den Würfeln:

$P(X = 2) = \frac{1}{36}$

$P(X = 3) = \frac{2}{36}$

$P(X = 4) = \frac{3}{36}$

$P(X = 5) = \frac{4}{36}$

$P(X = 6) = \frac{5}{36}$

$P(X = 7) = \frac{6}{36}$

$P(X = 8) = \frac{5}{36}$

$P(X = 9) = \frac{4}{36}$

$P(X = 10) = \frac{3}{36}$

$P(X = 11) = \frac{2}{36}$

$P(X = 12) = \frac{1}{36}$

Probe (Check):

$$\sum_{i=2}^{12} P(X = i) = 1$$

Ereignisse in Termini von X:
(in terms of)

$P(5 \leq X \leq 9) = \frac{24}{36} = \frac{2}{3}$

$P(X = 1) = 0$

Die ZV (= Zufallsvariable) X ist

- diskret, wenn $P(X = x_i) > 0$ nur für endlich viele $(x_1, \dots, x_n)$ oder abzählbar unendlich viele $(x_1, \dots, x_n, \dots)$ Werte gilt
- stetig, wenn sie ein Kontinuum von Werten annehmen können (später zu präzisieren!)

<u>Definition 24</u>: Die (kumulative) Verteilungsfunktion von X ist

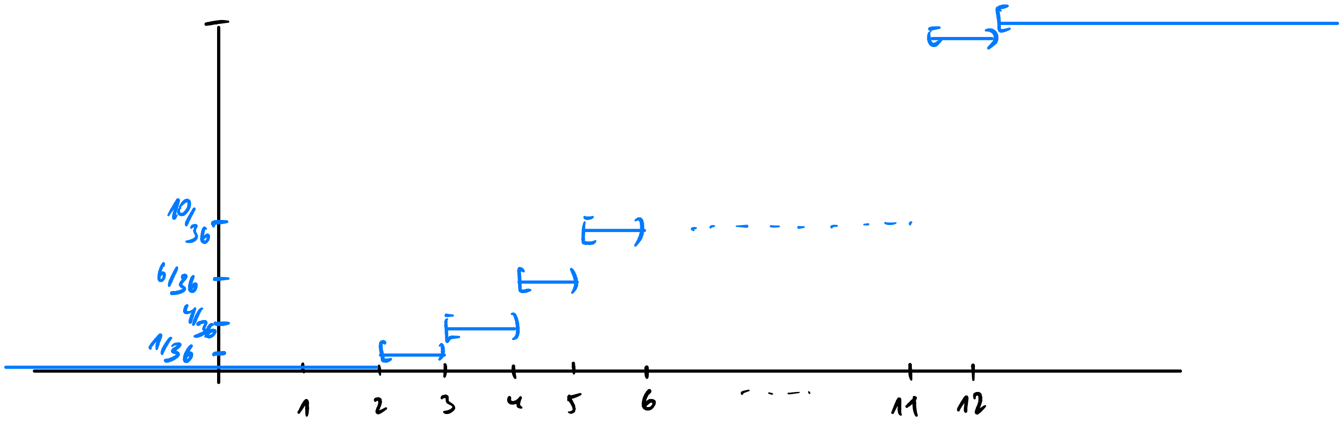
$$F: \mathbb{R} \to [0,1]$$

$$F(x) = P(X \leq x)$$

"$X \sim F$" bedeutet "F ist die Verteilungsfunktion von X"

(cumulative distribution function, cdf)

Verteilung für $X = W_1 + W_2$



$F(2) = P(X \leq 2) = \frac{1}{36}$

$F(3) = P(X \leq 3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$

$F(1{,}9) = P(X \leq 1.9) = 0$

<u>F beantwortet alle W.keitsfragen über X</u>

Etwa $\quad P(a < X \leq b) = ?$

Ereignis "$X \leq b$" $\rightsquigarrow [X \leq b] = [X \leq a] \cup [a < X \leq b]$

$[a < X \leq b] = [X \leq b] \setminus [X \leq a]$

$([X \leq a] \subseteq [X \leq b])$

$\Rightarrow \quad P[a < X \leq b] = P[X \leq b] - P[X \leq a]$

Beispiel: Wartezeit bei einem Prozess ohne Gedächtnis

Wir warten, bis

- ein Atom zerfällt
- ein Kunde anruft
- ein Computer ausfällt.

Sei X diese Wartezeit. Bei diesen Vorgängen gilt (im Prinzip):

Wenn wir schon s Minuten gewartet haben,
ist die W.keit noch einmal t Minuten zu warten,
gleich der W.keit, von Anfang an t Minuten zu warten.

Formal:

$$P[X > s+t \mid X > s] = P[X > t]$$

für alle $s, t \geq 0$

Prozess ohne Gedächtnis: Verteilungsfunktion

$$P[X > s+t \mid X > s] = P[X > t] \quad , \quad s,t \geq 0$$

bedeutet

$$\frac{P[X > s+t, X > s]}{P[X > s]} = P[X > t]$$

$$\Longleftrightarrow \quad P[X > s+t] = P[X > s] \cdot P[X > t]$$

Sei $G(s) := P[X > s]$. Dann gilt

- $G(0) = 1$; $G(s) = 1$ f.a. $s < 0$
- $G(s+t) = G(s) \cdot G(t)$
- $G(s) = P[X > s] = 1 - P[X \leq s] = 1 - F_X(s)$
- $\lim_{s \to \infty} G(s) = 0$

Prozess ohne Gedächtnis: Verteilungsfunktion (Forts.)

Welche Funktionen $G: \mathbb{R} \to \mathbb{R}$ erfüllen

- $G(s+t) = G(s) \cdot G(t)$
- $\lim_{s \to \infty} G(s) = 0$
- $G(s) = 1$ f.a. $s \le 0$?

Antwort:

$$G(s) = \begin{cases} 1, & s < 0 \\ A^s, & s \geq 0 \end{cases} \quad \text{für ein } A \text{ mit } 0 < A < 1$$

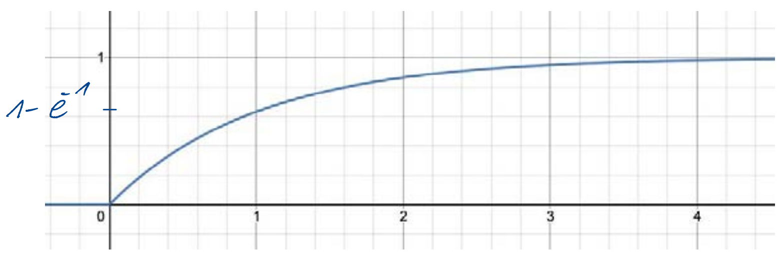
Also: $G(s) = e^{(\log A) \cdot s}$.

Erinnerung: $0 < A < 1 \Rightarrow \log A < 0$. Sei $\lambda := -\log A$.

Dann: $G(s) = e^{-\lambda \cdot s} \Rightarrow F_X(s) = 1 - G(s) = 1 - e^{-\lambda s}$ für ein $\lambda > 0$

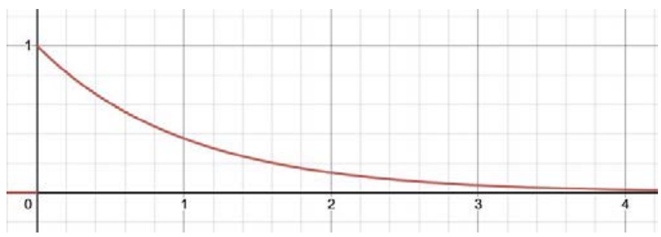
Beispiel 25a: Sei $X \sim F$ mit

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$



$F(x) = P[X \le x]$

$P[X \le 1] = F(1)$
$= 1 - e^{-1} \approx 0{,}632$



$F'(x) = e^{-x}$
Dichte der Exponentialverteilung

chain rule

$$F(x) = 1 - e^{-\lambda x}$$

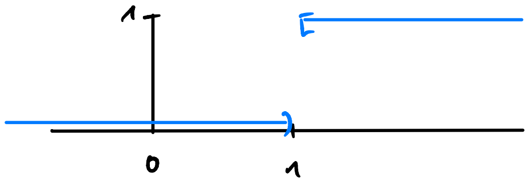
$$F'(x) = -e^{-\lambda x} \cdot (-\lambda) = \lambda \cdot e^{-\lambda x}$$

Eine Verteilungsfunktion F hat folgende Charakteristika:

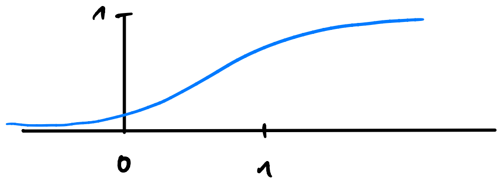
- $0 \leq F(x) \leq 1$ (denn $F(x) = P[X \leq x]$ ist eine W-keit)
- F ist monoton wachsend (monotonically increasing)
  (denn $x \leq y \Rightarrow [X \leq x] \subseteq [X \leq y]$
  $\Rightarrow F(x) = P[X \leq x] \leq P[X \leq y] \leq F(y)$)
- $\lim\limits_{x \to -\infty} F(x) = 0$ (denn $\lim\limits_{x \to -\infty} P[X \leq x] = P(\emptyset)$)
- $\lim\limits_{x \to \infty} F(x) = 1$ (denn $\lim\limits_{x \to \infty} P[X \leq x] = P(\Omega) = 1$)
- $\lim\limits_{\substack{x \to +x_0 \\ x > x_0}} F(x) = F(x_0)$

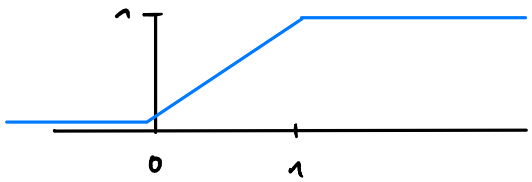
Jede Funktion mit diesen Eigenschaften ist eine Verteilungsfunktion

Mögliche Formen von Verteilungsfunktion



diskrete Verteilung mit 1 als einzigem möglichen Wert





2.1 Arten von Zufallsvariablen

Sei $\mathfrak{X}$ diskret. Dann ist

$$p : \mathbb{R} \longrightarrow [0,1] \qquad p(x) = P[\mathfrak{X} = x]$$

die Wahrscheinlichkeitsfunktion von $\mathfrak{X}$

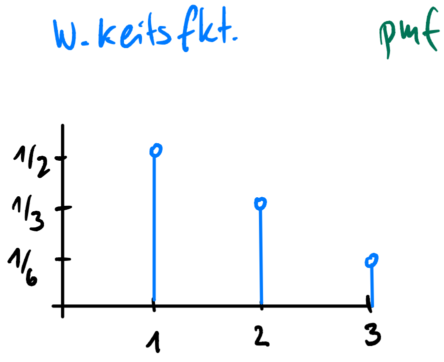
(engl. probability mass function, pmf)

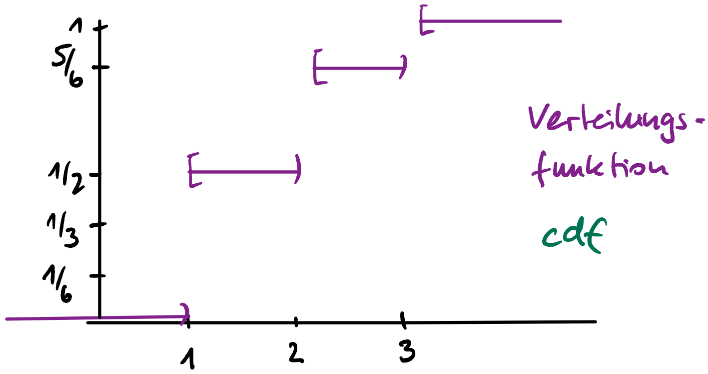
Seien $x_1, \dots, x_n, \dots$ die möglichen Werte von $\mathfrak{X}$. Dann ist

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Beispiel 26: X habe die Werte $\{1, 2, 3\}$

mit $p(2) = \frac{1}{3}$, $p(3) = \frac{1}{6}$ $\Rightarrow$ $p(1) = \frac{3}{6} = \frac{1}{2}$





$F(x) = P[X \leq x]$

F ist konstant auf jedem Intervall $[x_i, x_{i+1})$

$\Rightarrow$ F ist eine Treppenfunktion (step function)

Definition 27: X ist stetig (continuous) wenn es eine Funktion f gibt,

$f: \mathbb{R} \to \mathbb{R}$, $f \geq 0$ so dass

$$P[X \in B] = \int_B f(x)\, dx$$

für alle „vernünftigen" Mengen $B \subseteq \mathbb{R}$

(d.h. im Wesentlichen, B ist eine Vereinigung von Intervallen,
also $B = \bigcup_{i=1}^{\infty} [a_i, b_i]$)

Wir nennen f eine W.keitsdichte oder Dichte
(probability density function, pdf)

Bemerkung:

- $\int_{-\infty}^{\infty} f(x)\, dx = P[X \in (-\infty, \infty)] = 1$
- $P[a \leq X \leq b] = \int_a^b f(x)\, dx$

Quiz: Welche Konstante?

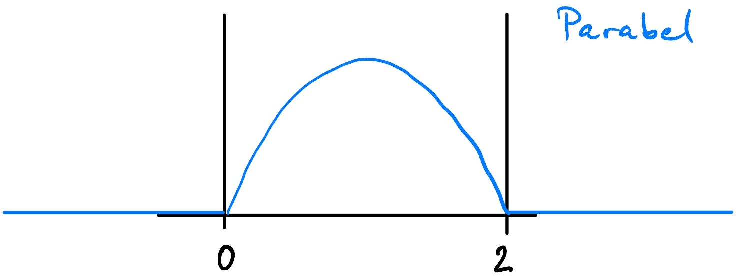
Die Funktion f sei definiert als

$$f(x) = c\,(2x - x^2) \quad \text{für} \quad 0 < x < 2 \quad \text{und} \quad 0 \ \text{sonst.}$$

Wie muss c gewählt werden, damit f eine Dichtefunktion ist?

Wie sieht f aus?

$$f(x) = c\,(2x - x^2)$$
$$= c \cdot x \cdot (2 - x)$$



Zusammenhang zwischen Verteilung und Dichte:

$$F(x) = P[X \le] = \int_{-\infty}^{x} f(y)\,dy$$

$$\Longleftrightarrow \quad F'(x) = f(x) \quad \text{und} \quad \lim_{x \to -\infty} F(x) = 0$$

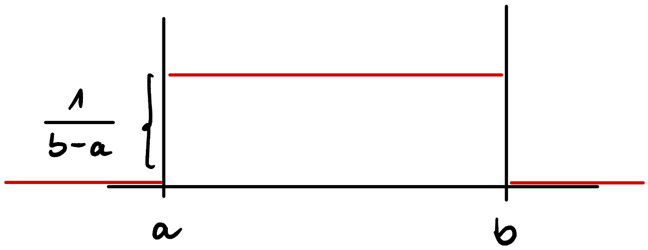
Wir hatten

$$f(x) = \begin{cases} e^{-x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

Dann

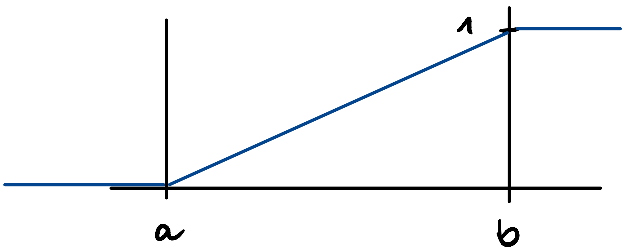
$$F(x) = \begin{cases} \int_{-\infty}^{x} 0\,dy = 0 & \text{wenn } x < 0 \\ \int_{-\infty}^{0} 0\,dy + \int_{0}^{x} e^{-y}\,dy & \\ = \left[-e^{-y}\right]_0^x = -e^{-x} - (-1) & \text{wenn } x \ge 0 \\ = 1 - e^{-x} & \end{cases}$$

Beispiel: Stetige Gleichverteilung



Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & , \; a \le x \le b \\ 0 & , \; \text{sonst} \end{cases}$$



Verteilung

$$F(x) = \begin{cases} 0 & , \; x < a \\ \frac{x-a}{b-a} & , \; a \le x \le b \\ 1 & , \; x > b \end{cases}$$

Die Verteilung ist eine stetige Funktion, die Dichte nicht (unstetig in a und b)

Quiz: Welche Konstante?

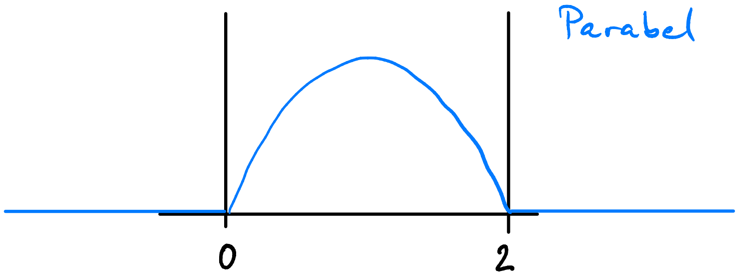
Die Funktion f sei definiert als

$$f(x) = c\,(2x - x^2) \quad \text{für} \quad 0 < x < 2 \quad \text{und} \quad 0 \text{ sonst.}$$

Wie muss c gewählt werden, damit f eine Dichtefunktion ist?

Wie sieht f aus?

$$f(x) = c\,(2x - x^2)$$
$$= c \cdot x \cdot (2 - x)$$



Parabel

Berechne c:

Wir wollen, dass $\int_{-\infty}^{\infty} f(x)\, dx = 1$

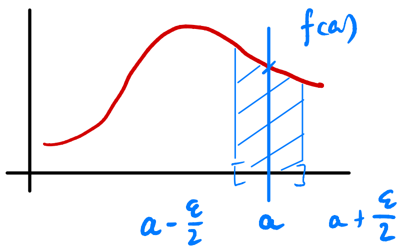
Hier $1 = \int_{-\infty}^{\infty} f(x)\, dx = \int_0^2 f(x)\, dx = \int_0^2 c\,(2x - x^2)\, dx$

$$= c\left[2 \cdot \frac{x^2}{2} - \frac{x^3}{3}\right]_0^2 = c\left(2 \cdot \frac{2^2}{2} - \frac{2^3}{3} - 0 + 0\right)$$

$$= c\left(4 - \frac{8}{3}\right) = c\,\frac{12-8}{3} = c\,\frac{4}{3} \implies c = \frac{3}{4}$$

Dichte: Intuition

Sei f eine stetige Dichte



$$P\left[a - \frac{\varepsilon}{2} \le X \le a + \frac{\varepsilon}{2}\right] \approx \varepsilon \cdot f(a)$$

Die W.keit, dass X in ein Intervall um a fällt ist proportional zu $f(a)$.

2.4 Der Erwartungswert

Würfeln: Sei X die gewürfelte Zahl. Im Durchschnitt sehen wir

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

Münzwurf, n-mal: # Kopf im Durchschnitt:

$$0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$ bei einem Wurf

$n = 2$

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 0 + \frac{1}{2} + \frac{1}{2} = 1$$

Im allgemeinen

$$\frac{n}{2} = n \cdot \frac{1}{2}$$

Definition: Sei X eine diskrete ZV, mit Werten $x_1, \dots, x_n, \dots$

Dann ist

$$E[X] := \sum_{i=1}^{n} x_i \cdot P[X = x_i], \text{ wenn } X \text{ } n \text{ Werte hat}$$

$$E[X] := \sum_{i=1}^{\infty} x_i \cdot P[X = x_i], \text{ wenn } X \text{ } \infty \text{ viele Werte hat}$$

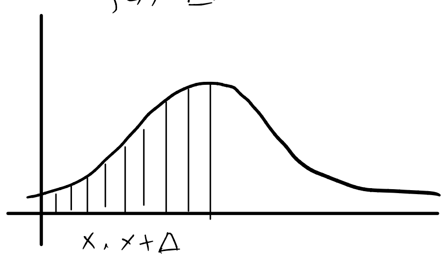
$E[X]$ ist der Erwartungswert von X (expected value, expectation) (und die Summe existiert)

Sei X stetig mit Dichte f. Dann ist

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

der Erwartungswert von X (falls das Integral existiert)

$f(x) \cdot \Delta \approx P[x \le X \le x + \Delta]$



Erwartungswert der Exponentialverteilung

Sei X verteilt nach

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0 \end{cases} \quad , \lambda > 0$$

Exponential-verteilung

Was ist $E[X]$? Die Dichte von X ist

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Dann ist

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} x \cdot e^{-\lambda x} dx = ?$$

Erwartungswert der Exponentialverteilung (Forts.)

$$\lambda \int_0^{\infty} \underbrace{x}_{g} \cdot \underbrace{e^{-\lambda x}}_{h'} dx = \lambda \left[\underbrace{x}_{g} \cdot \left(\underbrace{-\frac{1}{\lambda} e^{-\lambda x}}_{h} \right) \right]_0^{\infty} - \lambda \int_0^{\infty} \underbrace{1}_{g'} \cdot \left(\underbrace{-\frac{1}{\lambda} e^{-\lambda x}}_{h} \right) dx$$

partielle Integration

$$= \lambda (0 - 0) + \lambda \int_0^{\infty} \frac{1}{\lambda} e^{-\lambda x} dx = \int_0^{\infty} e^{-\lambda x} dx$$

$$= \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} = \frac{1}{\lambda} \left(0 - (-1) \right) = \frac{1}{\lambda} = E[X]$$

Intuition: λ ist eine Frequenz oder Rate, z.B.

- $\lambda = 5$ atomare Zerfälle pro Minute,
- $\lambda = 10$ Anrufe pro Stunde.

Dann warten wir im Schnitt $1/5$ Minute, $1/10$ Stunde bis zum nächsten Zerfall oder Anruf.

2.2 Gemeinsame Verteilungen

Wir betrachten zwei ZVen zusammen

- Größe und Gewicht von Personen
- Messfehler in x- und y-Richtung
- Niederschlag (precipitation) und Temperatur

und untersuchen W.keiten der Art

$$P[X = x,\ Y = y]$$

$$P[a < X \leq b,\ c < Y \leq d]$$

Beispiel 29: 9 Batterien: 2 neu, 3 teilweise geladen, 4 leer

wähle zufällig 3 der 9 Batterien

X = # neue Batterien $\qquad X \in \{0,1,2\}$

Y = # teilw. geladene Batterien $\qquad Y \in \{0,1,2,3\}$

Sei $p(x,y) = P[X = x,\ Y = y]$ für gemeinsame W.keitsfunktion von X und Y

$$p(0,0) = \frac{\binom{4}{3}}{\binom{9}{3}} = \frac{\binom{4}{1}}{\frac{{}^{3}\cancel{9}\cdot\cancel{8}\cdot 7^{4}}{\cancel{3}\cdot\cancel{2}\cdot 1}} = \frac{4}{84}$$

$$p(0,1) = \frac{\binom{3}{1}\binom{4}{2}}{84} = \frac{3\cdot 6}{84} = \frac{18}{84}$$

$$p(0,2) = \frac{\binom{3}{2}\binom{4}{1}}{84} = \frac{3\cdot 4}{84} = \frac{12}{84}$$

$$p(0,3) = \frac{\binom{3}{3}}{84} = \frac{1}{84}$$

$$p(1,0) = \frac{\binom{2}{1}\binom{3}{0}\binom{4}{2}}{84} = \frac{2 \cdot 6}{84} = \frac{12}{8}$$

$$p(1,1) = \frac{\binom{2}{1}\binom{3}{1}\binom{4}{1}}{84} = \frac{2 \cdot 3 \cdot 4}{84} = \frac{24}{84}$$

$$p(1,2) = \frac{\binom{2}{1}\binom{3}{2}\binom{4}{0}}{84} = \frac{2 \cdot 3}{84} = \frac{6}{84}$$

$$p(2,0) = \frac{\binom{2}{2}\binom{3}{0}\binom{4}{1}}{84} = \frac{1 \cdot 1 \cdot 4}{84} = \frac{4}{84}$$

$$p(2,1) = \frac{\binom{2}{2}\binom{3}{1}\binom{4}{0}}{84} = \frac{3}{84}$$

Wir haben die gemeinsame W.keitsfunktion von X und Y berechnet.

Zusammenfassung in Tabelle (als Vielfache (multiples) von $\frac{1}{84}$)

gemeinsame W.keitsfkt.

$X \backslash Y$	0	1	2	3	Summe
0	4	18	12	1	35
1	12	24	6	0	42
2	4	3	0	0	7
Summe	20	45	18	1	84

Wkeit, dass $X = i$, d.h. $P[X = i]$

marginale W.keitsfkt.

W.keit, dass $Y = j$, d.h. $P[Y = j]$

Seien $\mathfrak{X}, \mathcal{Y}$ diskret mit gemeinsamer W.keitsfunktion

$$p(x,y) = P[\mathfrak{X}=x, \mathcal{Y}=y]$$

Die gemeinsame Verteilung von $\mathfrak{X}$ und $\mathcal{Y}$ ist

$$F : \mathbb{R} \times \mathbb{R} \longrightarrow [0,1]$$

$$F(x,y) = P[\mathfrak{X} \leq x, \mathcal{Y} \leq y]$$

Die marginale Verteilung / Randverteilung von $\mathfrak{X}$ ist

$$F_{\mathfrak{X}}(x) = P[\mathfrak{X} \leq x] = P[\mathfrak{X} \leq x, \mathcal{Y} < \infty]$$

$$= \lim_{y \to \infty} F(x,y)$$

Wie erhalten wir die marginale W.keitsfunktion von $\mathfrak{X}$ aus der gemeinsamen W.keitsfunktion

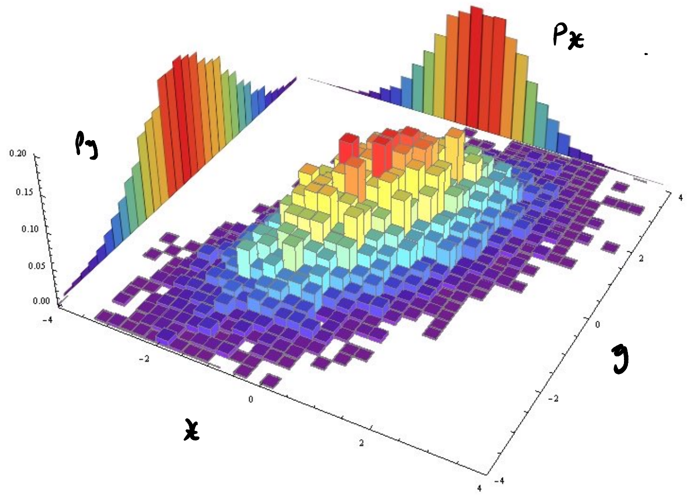
$$p(x,y) = P[\mathfrak{X}=x, \mathcal{Y}=y] \ ?$$

$$p_{\mathfrak{X}}(x) = P[\mathfrak{X}=x] = \sum_{j=1}^{n} P[\mathfrak{X}=x, \mathcal{Y}=y_j]$$

$$= \sum_{j=1}^{n} p(x, y_j)$$

Summiere über alle möglichen Werte von y

$$p_{\mathcal{Y}}(y) = P[\mathcal{Y}=y] = \sum_{i=1}^{m} p(x_i, y)$$

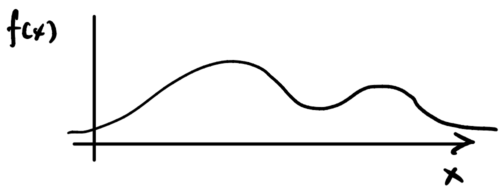


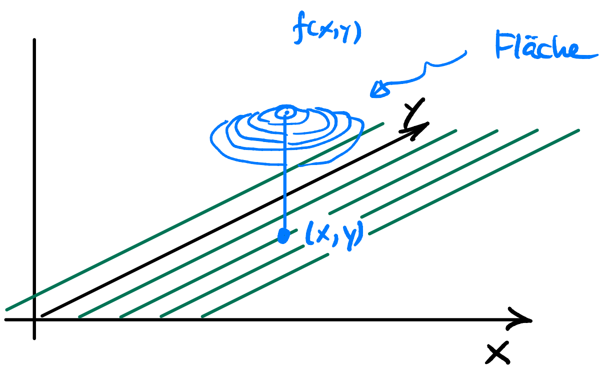
Stetige gemeinsame Verteilungen

Wie modellieren?

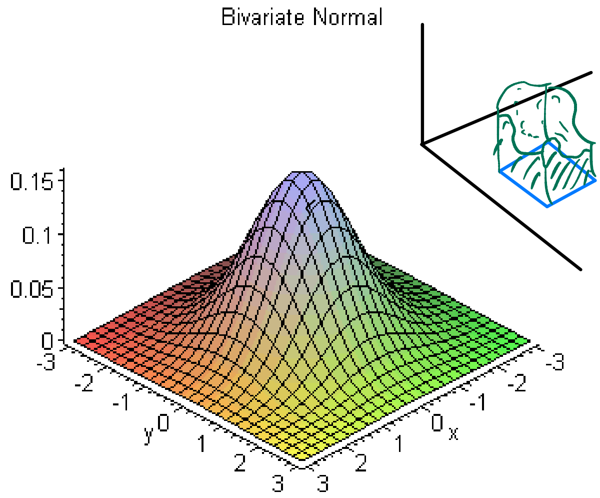
diskreter Fall: gemeinsame Wkeitsfunktion $p(x,y)$

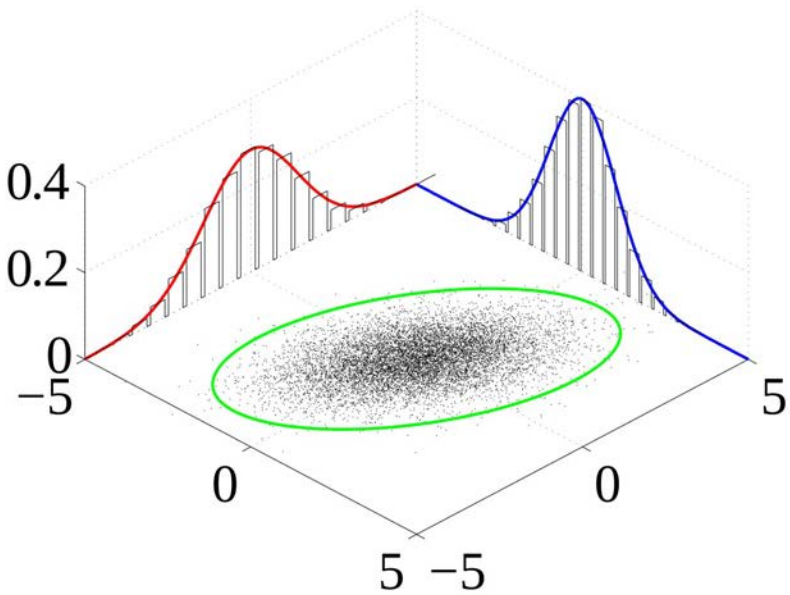
stetiger Fall:





Integral = Volumen unter der Fläche


Bivariate Normal
0.15
0.1
0.05
0
-3
-2
-1
y 0
1
2
3
3
2
1
0 x
-1
-2
-3


0.4
0.2
0
−5
0
5
−5
0
5

Gemeinsame Dichte stetiger ZVen

Seien X, Y stetig. Die gemeinsame Dichte ist ein Funktion

$$f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R},$$

so dass für „vernünftige" $C \subseteq \mathbb{R} \times \mathbb{R}$ gilt

$$P[(X,Y) \in C] = \iint_{(x,y) \in C} f(x,y) \, d(x,y)$$

Volumen unter der Fläche von f über C

Damit dies funktioniert, brauchen wir

- $f(x,y) \geq 0$
- $\iint_{\mathbb{R} \times \mathbb{R}} f(x,y) \, d(x,y) = 1$

Reihenfolge beim Integrieren: Satz von Fubini

Für alle „vernünftigen" Funktionen $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ und alle „vernünftigen" Mengen $A, B \subseteq \mathbb{R}$ gilt

Iterierte Integrale

$$\iint_{A \times B} f(x,y) \, d(x,y) = \int_A \underbrace{\left(\int_B f(x,y) \, dy \right)}_{} dx$$

Raum über $A \times B$ und unter f

- Für festes x, integriere bzgl. y
- $\leadsto$ Funktion von x
- Integriere bzgl. x

$$= \int_B \underbrace{\left(\int_A f(x,y) \, dx \right)}_{} dy$$

- Für festes y, integriere bzgl. x
- $\leadsto$ Funktion von y
- Integriere bzgl. y

D.h. die Reihenfolge ist egal.

Analogie: Vertauschung der Reihenfolge beim Summieren

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$$

Es ist

$$\sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} = (a_{11} + a_{12}) + (a_{21} + a_{22})$$

Summe über Zeilen

$$\sum_{j=1}^{2} \sum_{i=1}^{2} a_{ij} = (a_{11} + a_{21}) + (a_{12} + a_{22})$$

Summe über Spalten

Beide Summen sind gleich!

Analog $\int_A \int_B f(x,y)\, dy\, dx = \int_B \int_A f(x,y)\, dx\, dy$

Beispiel: Sei die gemeinsame Dichte von X, Y

$$f(x,y) = \begin{cases} c \cdot (x+2y), & 0 \leq x,y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Was ist c?

$$\iint_{\mathbb{R}\times\mathbb{R}} f(x,y)\, d(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy\, dx = \int_0^1 \left(\int_0^1 c(x+2y)\, dx \right) dy$$

$$= c \int_0^1 \left[\frac{x^2}{2} + 2yx \right]_0^1 dy = c \int_0^1 \frac{1}{2} + 2y\, dy$$

$$= c \left[\frac{1}{2} y + y^2 \right]_0^1 = c \left(\frac{1}{2} + 1 \right) = c \cdot \frac{3}{2}$$

Also: $c = \frac{2}{3}$

Andere Integrationsreihenfolge

$$\int_0^1 \left(\int_0^1 c(x+2y)\, dy \right) dx = c \int_0^1 \left[xy + y^2 \right]_0^1 dx$$

$$= c \int_0^1 x + 1 \, dx = \left[\frac{x^2}{2} + x \right]_0^1$$

$$= c\left(\frac{1}{2} + 1\right) = c \cdot \frac{3}{2}$$

Beispiel: Sei die gemeinsame Dichte von $\mathcal{X}, \mathcal{Y}$

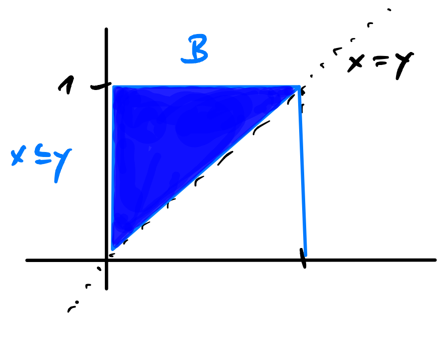
$$f(x,y) = \begin{cases} \frac{2}{3} \cdot (x+2y), & 0 \le x, y \le 1 \\ 0 & \text{sonst} \end{cases}$$

Was ist $P[\mathcal{X} \le \mathcal{Y}]$?

Wir wissen: $f(x,y) = 0$ falls $x \ge 1$ oder $y \ge 1$
und falls $x \le 0$ oder $y \le 0$.

Also ist

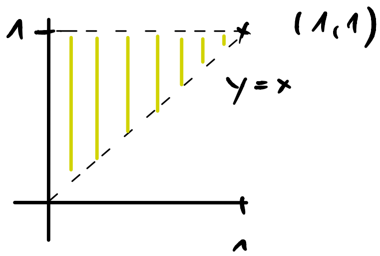
$$P[\mathcal{X} \le \mathcal{Y}] = \int_B f(x,y)\, d(x,y)$$



mit $B = \{ (x,y) \mid 0 \le x \le 1, 0 \le y \le 1, x \le y \}$

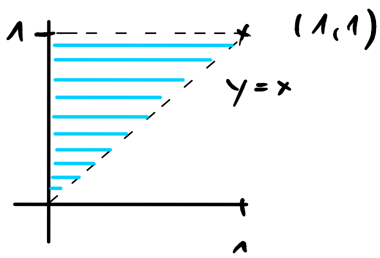
Wie berechnet man $P[X \le Y] = \int_B f(x,y)\, d(x,y)$?

Wir können B auf zwei Arten durchlaufen:



- zuerst, für jedes x zwischen 0 und 1, entlang der y-Achse von x bis 1; dann entlang der x-Achse von 0 bis 1; dies entspricht dem Integral

$$\int_0^1 \int_x^1 f(x,y)\, dy\, dx$$



- zuerst, für jedes y zwischen 0 und 1, entlang der x-Achse von 0 bis y; dann entlang der y-Achse von 0 bis 1; dies entspricht dem Integral

$$\int_0^1 \int_0^y f(x,y)\, dx\, dy$$

Wir rechnen:

$$\int_B f(x,y)\, dx\, dy = \int_0^1 \int_x^1 f(x,y)\, dy\, dx$$

$$= \int_0^1 \int_x^1 \frac{2}{3}(x+2y)\, dy\, dx = \frac{2}{3}\int_0^1 \left[xy + 2\frac{y^2}{2} \right]_x^1 dx$$

$$= \frac{2}{3}\int_0^1 \left(x + 2\frac{1^2}{2} - x^2 - x^2 \right) dx = \frac{2}{3}\int_0^1 x + 1 - 2x^2\, dx$$

$$= \frac{2}{3}\left[\frac{x^2}{2} + x - 2\frac{x^3}{3} \right]_0^1$$

$$= \frac{2}{3}\left(\frac{1}{2} + 1 - \frac{2}{3} \right) = \frac{\cancel{2}^{\,1}}{3} \cdot \frac{3+6-4}{\cancel{6}\,3} = \frac{5}{9}$$

Zweite Variante:

$$\int_B f(x,y)\,dx\,dy = \int_0^1 \int_0^y f(x,y)\,dx\,dy$$

$$= \int_0^1 \int_0^y \frac{2}{3}(x+2y)\,dx\,dy = \frac{2}{3}\int_0^1 \left[\frac{x^2}{2} + 2yx\right]_0^y dy$$

$$= \frac{2}{3}\int_0^1 \left(\frac{y^2}{2} + 2y^2 - \frac{0^2}{2} - 2y0\right) dy$$

$$= \frac{2}{3}\left[\frac{y^3}{3\cdot 2} + 2\frac{y^3}{3}\right]_0^1$$

$$= \frac{2}{3}\left(\frac{1}{6} + \frac{2}{3}\right) = \frac{\cancel{2}^{1}}{3} \cdot \frac{1+4}{\cancel{6}\,3} = \frac{5}{9}$$

Marginale Dichten

Sei f die gemeinsame Dichte von X und Y.
Dann ist die Randverteilung (marginal distribution) von X

$$F_X(x) = P[X \le x] = P[X \le x, Y < \infty]$$

$$= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx$$

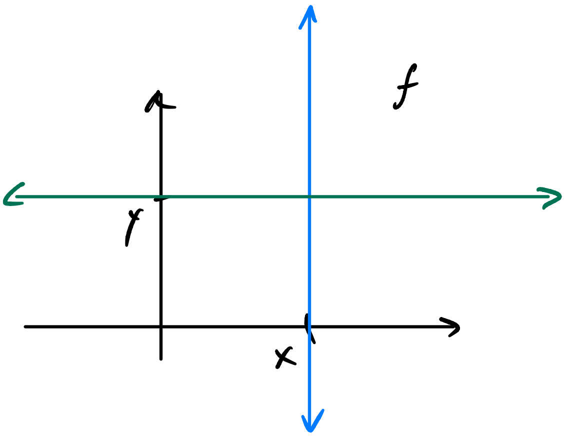
Die marginale Dichte (marginal density) von X ist die Ableitung von F_X, also

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx$$

$$= \int_{-\infty}^{\infty} f(x,y)\,dy$$

Marginale Dichten: Intuition

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$



$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

2.3 Unabhängige Zufallsvariablen

Erinnerung: 2 Ereignisse E, F sind unabhängig wenn

$$P(E F) = P(E) \cdot P(F)$$

Idee: Wissen über E gibt uns kein Wissen über F und umgekehrt.

Seien X, Y ZVen. Dann sind X, Y unabhängig wenn für alle $A, B \subseteq \mathbb{R}$ gilt

"$X \in A$" und "$Y \in B$" sind unabhängig

Oder:

alle Ereignisse die in Termini von X beschrieben werden können

"$5 < X < 9$" "$X > 2$" "$X < 0$ oder $X > 2$"

sind unabhängig von Ereignissen, die in Termini von Y beschrieben werden können

"$Y \geq 4$"

Satz: $\mathcal{X}, \mathcal{Y}$ sind unabhängig genau dann, wenn

$$P[\mathcal{X} \le a, \mathcal{Y} \le b] = P[\mathcal{X} \le a] \cdot P[\mathcal{Y} \le b]$$

für alle $a, b \in \mathbb{R}$.

Beweis: $\mathcal{E} = [\mathcal{X} \le x]$ ist ein Ereignis in Termini von $\mathcal{X}$,
$\mathcal{F} = [\mathcal{Y} \le y]$ ist ein Ereignis in Termini von $\mathcal{Y}$.
Daher sind $\mathcal{E}$ und $\mathcal{F}$ unabhängig und

$$P[\mathcal{E}\mathcal{F}] = P[\mathcal{E}] \cdot P[\mathcal{F}]$$

Folgerung 1: Seien $\mathcal{X}, \mathcal{Y}$ diskret, mit gemeinsamer W.keitsfunktion p.
Dann sind $\mathcal{X}$ und $\mathcal{Y}$ unabhängig gdw

$$p(x_i, y_j) = p_{\mathcal{X}}(x_i) \cdot p_{\mathcal{Y}}(y_j) \qquad \text{f.a. } x_i, y_j$$

wobei $p_{\mathcal{X}}, p_{\mathcal{Y}}$ die marginalen W-keitsfunktionen von $\mathcal{X}$ und $\mathcal{Y}$ sind.

Folgerung 2: Seien $\mathcal{X}, \mathcal{Y}$ stetig verteilt, mit gemeinsamer Dichte f.
Dann sind $\mathcal{X}$ und $\mathcal{Y}$ unabhängig gdw

$$f(x,y) = f_{\mathcal{X}}(x) \cdot f_{\mathcal{Y}}(y) \qquad \text{f.a. } x, y \in \mathbb{R},$$

wobei $f_{\mathcal{X}}, f_{\mathcal{Y}}$ die marginalen Dichten von $\mathcal{X}$ und $\mathcal{Y}$ sind.

Summen unabhängiger Zufallsvariablen

Seien X, Y unabhängig. Was können wir über die Verteilung von $Z = X + Y$ sagen?

Wir unterscheiden zwischen

diskreten ZVen

stetig verteilten ZVen

Summen diskreter unabhängiger Zufallsvariablen

Seien X, Y diskret:

- X habe die möglichen Werte $\{0, 1, \ldots, m\}$
- Y habe die möglichen Werte $\{0, 1, \ldots, n\}$

Die möglichen Werte von $Z = X + Y$ sind dann $\{0, 1, \ldots, m+n\}$

Was ist $P[Z = k]$, $0 \leq k \leq m+n$?

$$
\begin{aligned}
P[Z=k] &= P[X+Y=k] \\
&= P[X=0, Y=k] + P[X=1, Y=k-1] + \ldots + P[X=k, Y=0] \\
&= p_X(0) \cdot p_Y(k) + p_X(1) \cdot p_Y(k-1) + \ldots + p_X(k) \cdot p_Y(0) \\
&= \sum_{\ell=0}^{k} p_X(\ell) \cdot p_Y(k-\ell)
\end{aligned}
$$

$$P(Z=k) = \sum_{\ell=0}^{k} p_X(\ell) \cdot p_Y(k-\ell)$$

$f \sim X$

$g \sim Y$

$h \sim Z = X+Y$

$$h(z) = \int_{-\infty}^{\infty} f(x) \cdot g(z-x) \, dx$$

Beispiel: X, Y seien unabh. mit möglichen Werten $0, 1$, wobei

$$P[X=1] = P[Y=1] = p$$

und $$P[X=0] = P[Y=0] = (1-p).$$

Etwa: Münzwurf

Sei $Z = X+Y$. Dann ist

$$P[Z=0] = p_X(0) \cdot p_Y(0) = (1-p)^2$$

$$P[Z=1] = p_X(0) \cdot p_Y(1) + p_X(1) \cdot p_Y(0) = 2 \cdot (1-p)p$$

$$P[Z=2] = p_X(1) \cdot p_Y(1) = p^2$$

Summen stetig verteilter unabhängiger Zufallsvariablen

Seien X, Y unabhängig mit Dichten f und g, sei $Z = X + Y$.
Was ist die Dichte von Z?

Beispiel: Wartezeiten

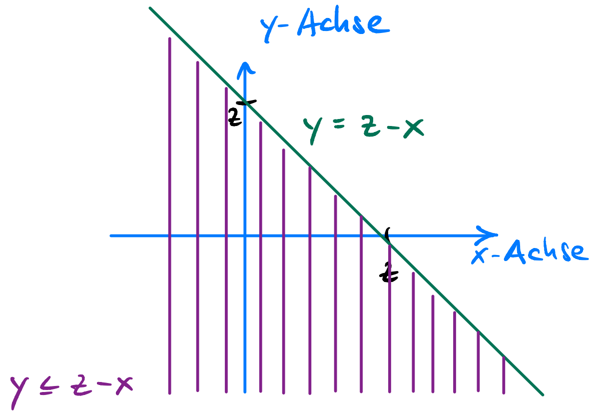
Die Verteilungsfunktion von Z ist

$$F_Z(z) = P[Z \le z] = P[X + Y \le z]$$

$F_Z(z)$ ist das Integral der gemeinsamen Dichte $f(x) \cdot g(y)$ über dem Gebiet

$$C = \{(x,y) \in \mathbb{R}^2 \mid x + y \le z\}$$
$$= \{(x,y) \in \mathbb{R}^2 \mid y \le z - x\}$$

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x) \cdot g(y) \, dy \, dx$$



Die Dichte der Summen-ZV:

Was ist $\frac{d}{dz} F_Z(z)$?

$$\frac{d}{dz} F_Z(z) = \frac{d}{dz} \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x) \cdot g(y) \, dy \, dx$$

Vertauschung von Ableitung und Integral

$$= \int_{-\infty}^{\infty} \frac{d}{dz} \int_{-\infty}^{z-x} f(x) \cdot g(y) \, dy \, dx$$

$$= \int_{-\infty}^{\infty} f(x) \frac{d}{dz} \int_{-\infty}^{z-x} g(y) \, dy \, dx$$

$$= \int_{-\infty}^{\infty} f(x) \frac{d}{dz} G(z-x) \, dx$$

mit $G(y) = \int_{-\infty}^{y} g(v) \, dv$, Verteilungsfunktion von Y

$$= \boxed{\int_{-\infty}^{\infty} f(x) \, g(z-x) \, dx}$$

$$= (f * g)(z)$$

Faltung (convolution) von f und g

Satz: Sind $\mathcal{X}, \mathcal{Y}$ unabhängige, stetig verteilte ZVen, mit Dichten f, g, dann hat die Summe $\mathcal{X}+\mathcal{Y}$ als Dichte die Faltung $f*g$,

$$\boxed{(f*g)(z) = \int_{-\infty}^{\infty} f(x)\, g(z-x)\, dx}$$

2.5 Eigenschaften des Erwartungswerts

Sei $\mathcal{X}$ ZV, $g: \mathbb{R} \to \mathbb{R}$ Funktion. Dann ist auch $g(\mathcal{X})$ eine ZV

Beispiel 37: Sei W die Punktezahl eines Würfels, $g(x) = x^2$

$\Rightarrow$ $g(W) = W^2$ ist das Quadrat der Punktezahl

W^2 ist eine neue ZV

Werte	W.keiten
1	1/6
4	1/6
9	1/6
16	1/6
25	1/6
36	1/6

$E[W^2]$

$$= 1 \cdot 1/6 + 4 \cdot 1/6 + 9 \cdot 1/6 + 16 \cdot 1/6 + 25 \cdot 1/6 + 36 \cdot 1/6$$

$$= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}$$

$$= \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} = \sum_{i=1}^{6} g(i) \cdot p_{\mathcal{X}}(i)$$

Beispiel: Würfel mit Zahlen $-3, -2, -1, 1, 2, 3$

X sei die gewürfelte Zahl $\Rightarrow E[X] = 0$

Sei $Z := X^2$. Was ist $E[Z]$?

W.keitsfkt. von Z

Werte	W.keiten
1	$1/6 + 1/6 = 1/3$
4	$1/6 + 1/6 = 1/3$
9	$1/6 + 1/6 = 1/3$

1) Bestimme die W.keitsfkt. von Z und berechne $E[Z]$:

$$E[Z] = 1 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 9 \cdot \frac{1}{3} = \frac{14}{3}$$

2) Berechne das gewichtete Mittel der $g(x_i)$:

$$E[X^2] = (-3)^2 \cdot \frac{1}{6} + (-2)^2 \cdot \frac{1}{6} + (-1)^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 1^2 \cdot \frac{1}{6}$$

Wir sehen:

$$E[g(X)] = \sum_i g(x_i) \cdot p_X(x_i)$$

Satz 39 ("Law of the Unconscious Statistician", LOTUS)

Seien X ZV, $g: \mathbb{R} \to \mathbb{R}$

- $E[g(X)] = \sum_i g(x_i)\, p_X(x_i)$, X diskret
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$, X stetig verteilt, f Dichte von X

Anwendungen des LOTUS-Theorems (1)

X ZV, $a, b \in \mathbb{R}$. $g(x) = ax + b$

Erwartungswert unter skalarer Multiplikation und Verschiebung

$$E[aX + b] \overset{?}{=} a\,E[X] + b$$

$$E[g(X)] = \int_{-\infty}^{\infty} (ax+b) \cdot f(x)\, dx = \int_{-\infty}^{\infty} ax \cdot f(x) + b \cdot f(x)\, dx$$

$$= a \int_{-\infty}^{\infty} x \cdot f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx$$

$$= a\,E[X] + b \cdot 1$$

$$= a\,E[X] + b$$

Für diskrete ZVen ist die Rechnung analog

Anwendungen des LOTUS-Theorems (2)

Seien X, Y ZVen. Dann ist

Erwartungswert der Summe

$$E[X + Y] = E[X] + E[Y]$$

LOTUS gilt auch für 2-dimensionale W.keitsfkt.en und Dichten:

Sei $f(x,y)$ die gemeinsame Dichte von X und Y, $g: \mathbb{R}^2 \longrightarrow \mathbb{R}$.

Dann ist

$$E[g(X,Y)] = \iint_{\mathbb{R} \times \mathbb{R}} g(x,y) \cdot f(x,y)\, d(x,y)$$

Das Summengesetz folgt mit $g(x,y) = x + y$

$E[g(X,Y)]$ $g(x,y)$

$$E[X+Y] \overset{\text{LOTUS}}{=} \iint_{\mathbb{R}\times\mathbb{R}} (x+y)\cdot f(x,y)\, d(x,y)$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} x\cdot f(x,y)\, dx\, dy + \int_{\mathbb{R}}\int_{\mathbb{R}} y\cdot f(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} x \underbrace{\left(\int_{\mathbb{R}} f(x,y)\, dy\right)}_{f_X(x)} dx + \int_{\mathbb{R}} y \underbrace{\left(\int_{\mathbb{R}} f(x,y)\, dx\right)}_{f_Y(y)} dy$$

$$= \int_{\mathbb{R}} x\cdot f_X(x)\, dx + \int_{\mathbb{R}} y\cdot f_Y(y)\, dy$$

$$= E[X] + E[Y]$$

Verallgemeinerung:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Beispiel: Summe zweier Würfel W_1, W_2: $E[W_i] = \frac{7}{2}$, $i=1,2$

$$\Rightarrow E[W_1+W_2] = E[W_1] + E[W_2] = \frac{7}{2} + \frac{7}{2} = 7$$

Beispiel 42: n-faches Münzwurf, $E[\#\text{Kopf}] = ?$

Sei $P[K] = p$, $P[Z] = (1-p)$.

Sei $X_i = 1$ gdw der i-te Wurf ist Kopf

Dann ist $E[X_i] = 1\cdot p + 0\cdot(1-p) = p$.

n Würfe: $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = n\cdot p$

Erwartungswert von Produkten unabhängiger ZVen

Bemerkung: Seien X, Y unabhängig $\Rightarrow$ $E[X \cdot Y] = E[X] \cdot E[Y]$

Beweis: Wir zeigen die Behauptung für stetig verteilte ZVen. Für diskrete ZVen ist das Argument analog.

Seien $X \sim f$, $Y \sim g$.

$$E[X \cdot Y] \overset{\text{LOTUS}}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot y \cdot f(x) \cdot g(y) \, dy \, dx$$

$$= \int_{\mathbb{R}} x \cdot f(x) \cdot \left(\int_{\mathbb{R}} y \cdot g(y) \, dy \right) dx$$

„(.)" ist eine Konstante. Man kann sie aus dem Integral herausziehen.

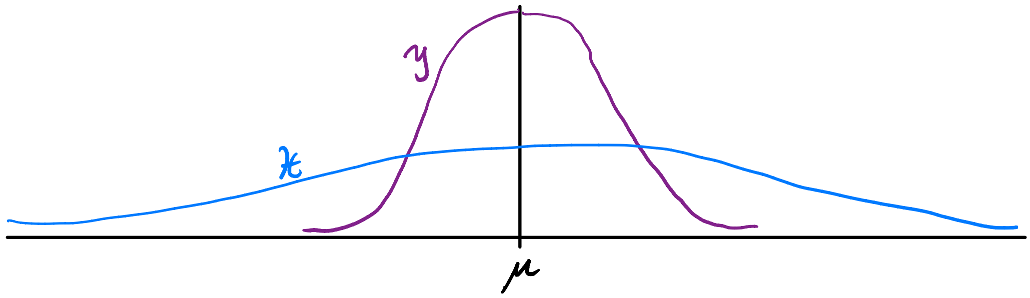
$$= \int_{\mathbb{R}} x \cdot f(x) \, dx \cdot \left(\int_{\mathbb{R}} y \cdot g(y) \, dy \right)$$

$$= E[X] \cdot E[Y]$$

2.6 Varianz

Beispiel: Seien X, Y ZVen,

$\mu = E[X] = E[Y]$ der Mittelwert von X, Y



Die Werte von X sind weiter um μ gestreut als die Werte von Y.

Wir wollen den mittleren Unterschied zwischen X zu μ messen.

Intuition: Seien

$x = (x_1, \dots, x_n) \in \mathbb{R}^n$ ein beliebiger Vektor,

$\mu = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$ der Mittelwert der x_i,

$m = (\mu, \dots, \mu) \in \mathbb{R}^n$ ein Vektor, dessen Komponenten alle $= \mu$ sind

Der Abstand zwischen x und m in $\mathbb{R}^n$ ist

$$\|x - m\| = \sqrt{(x_1-\mu)^2 + \dots + (x_n - \mu)^2} = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2}$$

Das Quadrat des Abstands ist dann

$$\|x - m\|^2 = (x_1-\mu)^2 + \dots + (x_n - \mu)^2 = \sum_{i=1}^{n} (x_i - \mu)^2$$

Für eine ZV X ist die Varianz (variance) von X

$$Var(X) = E[(X-\mu)^2]$$

und die Standardabweichung (standard deviation)

$$\sigma = \sqrt{Var(X)} \qquad \Rightarrow \quad Var(X) = \sigma^2$$

Was kann man sich darunter vorstellen?

Sei X eine diskrete ZV mit Werten $x_1, \dots, x_n$, W.keiten $p(x_i) = p_i$; Mittelwert μ
Dann ist

$$Var(X) = E[(X-\mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$$

$$\sigma = \sqrt{Var(X)} = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i}$$

Ähnliche Ausdrücke erhält man auch für stetig verteilte ZVen

Im Vergleich:

$$\|x-m\|^2 = \boxed{\sum_{i=1}^{n} (x_i-\mu)^2} \qquad \|x-m\| = \boxed{\sqrt{\sum_{i=1}^{n} (x_i-\mu)^2}}$$

$$Var(X) = \boxed{\sum_{i=1}^{n} (x_i-\mu)^2 \cdot p_i} \qquad \sigma = \boxed{\sqrt{\sum_{i=1}^{n} (x_i-\mu)^2 \cdot p_i}}$$

Varianz und Standardabweichung messen den gewichteten Abstand zwischen

- der Verteilung von X und
- der konstanten Verteilung μ.

Der Erwartungswert $\mu = E[X]$ ist gerade die Konstante, für die dieser Abstand minimal ist.

Variance of a continuous X with density f:

$$Var(X) = \sum_{i=1}^{n} (x_i-\mu)^2 \, p(x_i) \qquad \text{discrete case}$$

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \, f(x) \, dx \qquad \text{cont. case}$$

Wie kann man die Varianz berechnen?

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2X\mu + \mu^2] \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2\mu\cdot\mu + \mu^2 \\
&= E[X^2] - \mu^2 \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

Beispiel: Würfelzahl W

Wir wissen schon: $E[W] = \frac{7}{2}$ $\quad E[W^2] = \frac{91}{6}$ (Bsp 37)

Also $Var(W) = E[W^2] - E[W]^2$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182 - 147}{12} = \frac{35}{12}$$

Eigenschaften der Varianz

- $Var(X + b) = Var(X)$
- $Var(aX) = a^2 Var(X)$
- $Var(X + Y) \overset{?}{=} Var(X) + Var(Y)$ falls X, Y unabhängig

Beweis später

Wir rechnen (mit $\mu = E[X]$):

$$
\begin{aligned}
Var(aX+b) &= E[Y^2] - E[Y]^2 \\
&= E[(aX+b)^2] - E[aX+b]^2 \\
&= E[a^2X^2 + 2abX + b^2] - (a\,E[X] + b)^2 \\
&= a^2 E[X^2] + 2ab\,E[X] + b^2 - a^2 E[X]^2 - 2ab\,E[X] - b^2 \\
&= a^2 E[X^2] + \cancel{2ab\mu} + \cancel{b^2} - a^2\mu^2 - \cancel{2ab\mu} - \cancel{b^2} \\
&= a^2 (E[X^2] - \mu^2) = a^2\,Var(X)
\end{aligned}
$$

Folgerung über die Standardabweichung σ

$$\boxed{\sigma_{aX} = a\,\sigma_X}$$

Wir rechnen:

$$\sigma_{aX} = \sqrt{Var(aX)} = \sqrt{a^2 Var(X)} = a\sqrt{Var(X)} = a\,\sigma_X$$

2.7 Kovarianz

Wir haben gesehen

$$Var(X+X) = Var(2X) = 4\,Var(X)$$
$$\neq Var(X) + Var(X)$$

Was macht den Unterschied aus?

Definition 45: Seien X, Y ZVen mit gemeinsamer Verteilung, $\mu_X = E[X]$, $\mu_Y = E[Y]$.
Die Kovarianz von X und Y ist

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Idee: X und Y werden um ihre Mittelwerte zentriert, dann werden ihre Abweichungen vom Mittelwert verglichen.

Kovarianz und Varianz

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$
$$= E[(X-\mu)\cdot(X-\mu)]$$
$$= \mathrm{Cov}(X, X)$$

Die Varianz ist ihre eigene Kovarianz.

Eigenschaften der Kovarianz

$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$

Wie kann man die Kovarianz (einfach) berechnen?

$$\mathrm{Cov}(X,Y) = E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y)$$
$$= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y$$
$$= E[XY] - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y$$
$$= E[XY] - E[X]\,E[Y]$$

Beobachtung: X, Y unabhängig $\Rightarrow E[XY] = E[X]\cdot E[Y]$

$\Rightarrow \mathrm{Cov}(X,Y) = 0$

Die Umkehrung gilt nicht!

Intuition: $\mathrm{Cov}(X,Y)$ misst, wie stark X, Y voneinander abhängen.

Kovarianz und Addition

„Die Kovarianz ist bilinear"

Satz 46: $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

Wir rechnen

$$Cov(X_1 + X_2, Y) = E[(X_1 + X_2) \cdot Y] - E[X_1 + X_2] \cdot E[Y]$$

$$= E[X_1 Y] + E[X_2 \cdot Y] - E[X_1] \cdot E[Y] - E[X_2] \cdot E[Y]$$

$$= E[X_1 Y] - E[X_1] \cdot E[Y] + E[X_2 \cdot Y] - E[X_2] \cdot E[Y]$$

$$= Cov(X_1, Y) + Cov(X_2, Y)$$

Theorem 47: $Cov\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j Cov(X_i, Y_j)$

Varianz der Summe versus Summe der Varianzen

$$Var\left(\sum_i X_i\right) = Cov\left(\sum_i X_i, \sum_i X_i\right)$$

$$= \sum_i \sum_j Cov(X_i, X_j)$$

$$= \sum_i \left(\sum_{j \neq i} Cov(X_i, X_j) + Cov(X_i, X_i) \right)$$

$$= \sum_i \sum_{j \neq i} Cov(X_i, X_j) + \sum_i Cov(X_i, X_i)$$

$$= \sum_i Var(X_i) + \sum_i \sum_{j \neq i} Cov(X_i, X_j)$$

Für $n = 2$:

Dies ist wie $(x+y)^2 = x^2 + y^2 + 2xy$

$$Var(X + Y) = Var(X) + Var(Y) + 2\, Cov(X, Y)$$

Bedeutung der Kovarianz

$Cov(\mathcal{X}, \mathcal{Y})$ ist

- >0 : $\mathcal{X}, \mathcal{Y}$ haben meistens ähnliche Werte
- <0 : $\mathcal{X}, \mathcal{Y}$ haben meistens entgegengesetzte Werte
- ≈ 0 : $\mathcal{X}, \mathcal{Y}$ haben wenig miteinander zu tun

Korrelation

Idee: Normalisiere die Kovarianz auf Werte zwischen -1 und 1, d.h., definiere

$$\mathcal{X}_0 := \frac{\mathcal{X}}{\sigma_\mathcal{X}} = \frac{\mathcal{X}}{\sqrt{Var(\mathcal{X})}} \qquad \mathcal{Y}_0 := \frac{\mathcal{Y}}{\sigma_\mathcal{Y}} = \frac{\mathcal{Y}}{\sqrt{Var(\mathcal{Y})}}$$

Also $Var(\mathcal{X}_0) = Var(\mathcal{Y}_0) = 1$.

Dann ist

$$Corr(\mathcal{X}, \mathcal{Y}) := Cov(\mathcal{X}_0, \mathcal{Y}_0) = \frac{Cov(\mathcal{X}, \mathcal{Y})}{\sigma_\mathcal{X} \cdot \sigma_\mathcal{Y}} = \rho_{\mathcal{X},\mathcal{Y}}$$

der Pearsonsche Korrelationskoeffizient (nach Karl Pearson)

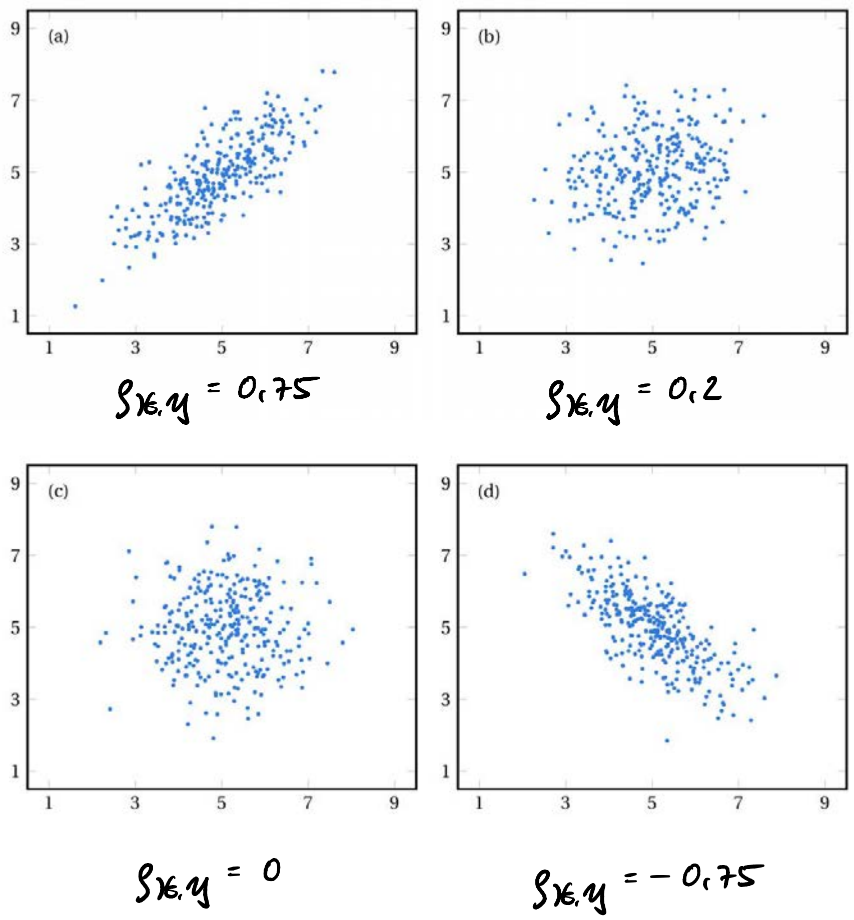


Figure 9: Random variables $\mathscr{X}$ and $\mathscr{Y}$ with correlations (a) 0.75; (b) 0.2; (c) 0; and (d) -0.75.

Beispiel: 10-mal unabhängig Würfeln

Sei X_i das i-te Ergebnis

$$\mathrm{Var}\left(\sum_{i=1}^{10} X_i\right) \overset{\text{Unabhängigkeit der } X_i}{=} \sum_{i=1}^{10} \mathrm{Var}(X_i)$$

$$= \sum_{i=1}^{10} \frac{35}{12} = 10 \cdot \frac{35}{12}$$

Was ist mit der Standardabweichung?

- Var ist um den Faktor 10 gewachsen
- σ wächst um den Faktor $\sqrt{10}$

2.8 Das schwache Gesetz der großen Zahlen

Wir leiten zunächst zwei Ungleichungen her.

Sei $X \geq 0$ und f die Dichte von X. Sei $a > 0$. Dann:

$$E[X] = \int_0^\infty x\, f(x)\, dx \geq \int_a^\infty x\, f(x)\, dx \geq \int_a^\infty a\, f(x)\, dx$$

$$= a \int_a^\infty f(x)\, dx = a\, P[X \geq a]$$

Argument funktioniert auch für diskrete ZVen

Also: $P[X \geq a] \geq \frac{E[X]}{a}$

Theorem 50 (Markow-Ungleichung): Sei X eine ZV mit $X \geq 0$ und $a > 0$.
Dann gilt

$$P[X \geq a] \leq \frac{E[X]}{a}$$

(Markov's Inequality)

Wir wenden die Markow-Ungleichung an auf

$$Y := (X - \mu)^2 \qquad a = k^2$$

Annahme: $Var(X) = \sigma^2 < \infty \Rightarrow E[Y] = Var(X) = \sigma^2$. Dann:

$$P[\,|X-\mu| \geq k\,] = P[(X-\mu)^2 \geq k^2] = P[Y \geq k^2]$$
$$\leq \frac{E[Y]}{k^2} = \frac{\sigma^2}{k^2}$$

Korollar 51 (Tschebyscheff-Ungleichung). Sei X eine ZV mit Mittelwert μ und Varianz σ^2. Dann gilt für jedes $k > 0$

$$P[\,|X-\mu| \geq k\,] \leq \frac{\sigma^2}{k^2}$$

Beispiel 52. Eine Person habe eine Wochenarbeitszeit X mit $\mu = 40$ Stunden.

1) Wie wahrscheinlich ist es, dass diese Person mehr als 60 Stunden pro Woche arbeitet?

Wende an: $P[X \geq a] \leq \frac{E[X]}{a}$

$$P[X \geq 60] \leq \frac{40}{60} = \frac{2}{3}$$

Beispiel 52. Eine Person habe eine Wochenarbeitszeit X mit $\mu = 40$ Stunden.

2) Wenn $Var(X) = 16$ ist, wie wahrscheinlich ist es dann, dass die Person zwischen 32 und 48 Stunden arbeitet?

Wende an: $P[\,|X - \mu| \geq k\,] \leq \frac{\sigma^2}{k^2}$

$$P[\,|X - 40| \geq 8\,] \leq \frac{\sigma^2}{k^2} = \frac{16}{8^2} = \frac{2}{8} = \frac{1}{4}$$

$$\Rightarrow P[\,|X - 40| \leq 8\,] = 1 - P[\,|X - 40| \geq 8\,]$$

$$\geq 1 - \frac{1}{4} = \frac{3}{4}$$

Beispiel: Kleine Schulen

Erziehungswissenschaftler fanden heraus, dass es unter den Schulen, die bei der Bewertung des Unterrichtserfolgs am besten abschneiden, immer verhältnismäßig mehr kleine Schulen gibt, als es kleine Schulen unter allen Schulen gibt. (Siehe die Statistik aus North Carolina.)

Die Gates-Stiftung beschloss Anfang der 2000er Jahre, massiv in die Einrichtung kleiner Schulen zu investieren (z.B. durch Aufteilung größerer Schulen in kleinere).

War das eine gute Entscheidung?

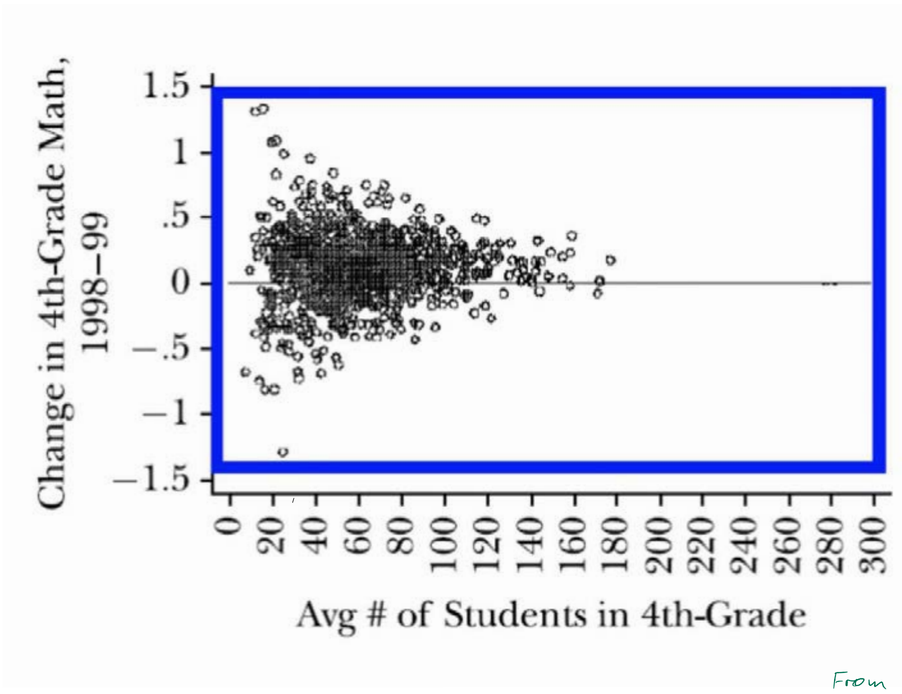
Die Geschichte stammt aus dem Buch "Thinking Fast and Slow" von Daniel Kahnemann

Erfolge kleiner Schulen in North Carolina

School Size	Percentage Ever "Top 25" 1997–2000
Smallest decile	27.7%
2nd	11.8
3rd	8.2
4th	3.6
5th	2.4
6th	3.6
7th	4.8
8th	7.1
9th	0
Largest decile	1.2
Total	7.0

Aus
Alex Tabarrok
"The Small Schools Myth",
September 2, 2010
https://marginalrevolution.com

Verteilung der Erfolge in Bezug zur Schülerzahl



From
Alex Tabarrok
"The Small Schools Myth",
September 2, 2010
https://marginalrevolution.com

Was geschieht, wenn wir ein Experiment viele Male durchführen und Durchschnitte der Ergebnisse ausrechnen?

Sei X eine ZV, seien $X_1, \dots, X_n$ ZVen welche
1) dieselbe Verteilung haben wie X
2) unabhängig sind.

Man redet dann von unabhängig und identisch verteilten (uiv) ZVen
(independent and identically distributed (iid) RVs)

Sei
- $\sigma^2 = Var(X) = Var(X_i)$
- $\mu = E[X] = E[X_i]$
- $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$

Es ist also $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. Dann gilt:

- $E[\overline{X}_n] = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \cdot n \cdot \mu = \mu$
- $Var(\overline{X}_n) = Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right)$

Unabhängigkeit der X_i

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Wir wenden die Tschebyscheff-Ungleichung auf $\overline{X}_n$ mit $k = \varepsilon$ an:

$$P\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right] = P[|\overline{X} - \mu| > \varepsilon]$$

$$\leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{1}{n} \cdot \frac{\sigma^2}{\varepsilon^2} \longrightarrow 0 \quad (n \to \infty)$$

Theorem (Schwaches Gesetz der großen Zahlen)

Seien $X_1, \dots, X_n, \dots$ u.i.v. ZVen mit Mittelwert μ und $Var(X_i) < \infty$

Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n \to \infty} P\left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \varepsilon\right] = 0$$

D.h. die Wahrscheinlichkeit von ε-Ausreißern (ε-outliers) geht gegen 0.