

Statistik für Informatiker, SS 2018

1.4 Erwartungswert, Varianz und Kovarianz

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/StatInfo18/>

14.5.2018 / 28.5.2018



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Der Erwartungswert ist eine wichtige Kenngröße der Verteilung einer reellwertigen Zufallsvariable X , er gibt eine Antwort auf die – etwas salopp formulierte – Frage „Wie groß ist X typischerweise?“

Sei X reelle ZV mit abzählbarem Wertebereich (auf einem W'raum (Ω, \mathcal{F}, P) definiert), d.h. es gibt eine abzählbare Menge $S = S_X \subset \mathbb{R}$ mit $P(X \in S) = 1$ und $\mathcal{L}_P(X)$ hat Gewichte $P(X = x)$, $x \in S$.

Sei X reelle ZV mit abzählbarem Wertebereich (auf einem W -raum (Ω, \mathcal{F}, P) definiert), d.h. es gibt eine abzählbare Menge $S = S_X \subset \mathbb{R}$ mit $P(X \in S) = 1$ und $\mathcal{L}_P(X)$ hat Gewichte $P(X = x)$, $x \in S$.

Definition 1.66

Der *Erwartungswert* von X ist definiert als

$$\mathbb{E}[X] := \sum_{x \in S_X} xP(X = x),$$

sofern die Reihe absolut konvergiert (d.h. sofern $\sum_{x \in S_X} |x|P(X = x) < \infty$ gilt, dann kann die Summation in beliebiger Reihenfolge erfolgen). Manchmal schreibt man auch $\mu_X := \mathbb{E}[X]$.

Man sagt dann „ X besitzt einen Erwartungswert“ und schreibt dies auch als $X \in \mathcal{L}^1$ (bzw. $X \in \mathcal{L}^1(P)$, wenn die zugrundeliegende W -keiten $P(\cdot)$ nicht aus dem Kontext klar sind).

Beispiel 1.67

1 A ein Ereignis, so ist

$$\mathbb{E}[I_A] = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) = P(A).$$

Beispiel 1.67

- ① A ein Ereignis, so ist

$$\mathbb{E}[I_A] = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) = P(A).$$

- ② W Augenzahl bei einem fairen Würfelwurf (W ist uniform auf $\{1, 2, 3, 4, 5, 6\}$), so ist

$$\begin{aligned}\mathbb{E}[W] &= \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 \\ &= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3,5\end{aligned}$$

Beispiel 1.67

- 1 A ein Ereignis, so ist

$$\mathbb{E}[I_A] = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A = 0) = P(A).$$

- 2 W Augenzahl bei einem fairen Würfelwurf (W ist uniform auf $\{1, 2, 3, 4, 5, 6\}$), so ist

$$\begin{aligned}\mathbb{E}[W] &= \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 \\ &= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = 3,5\end{aligned}$$

(allgemein: X uniform auf $\{1, 2, \dots, s\}$ mit $s \in \mathbb{N}$, so ist

$$\mathbb{E}[X] = \sum_{i=1}^s \frac{1}{s} \cdot i = \frac{1}{s} \frac{s(s+1)}{2} = \frac{s+1}{2}.)$$

Beispiel 1.67 (Fortsetzung)

③ X habe Werte in $S := \{2, 3, 4, \dots\} \cup \{-2, -3, -4, \dots\}$ mit Gewichten $P(X = n) = P(X = -n) = \frac{1}{2n(n-1)}$ für $n = 2, 3, \dots$

(es ist $\sum_{n=2}^{\infty} 2 \frac{1}{2n(n-1)} = \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = 1$, d.h. dies sind W'gewichte),
dann ist

$$\sum_{x \in S} |x| P(X = x) = \sum_{n=2}^{\infty} \frac{n}{n(n-1)} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

d.h. X besitzt keinen Erwartungswert.

Beispiel 1.67 (Fortsetzung)

- ③ Wenn man S durchnummerierte mit $x_{2i} = i + 1$,
 $x_{2i-1} = -i - 1$, $i \in \mathbb{N}$, so wäre

$$\sum_{j=1}^{\infty} x_j P(X = x_j) = \lim_{N \rightarrow \infty} \sum_{j=1}^N x_j P(X = x_j) = 0$$

(denn $\sum_{j=1}^{2N} x_j P(X = x_j) = 0$).

Beispiel 1.67 (Fortsetzung)

- 3 Wenn man andererseits S durchnummerierte mit $x_{3i} = -i - 1$, $x_{3i-2} = 2i$, $x_{3i-1} = 2i + 1$, $i \in \mathbb{N}$, so wäre

$$\sum_{j=1}^{\infty} x_j P(X = x_j) = \lim_{N \rightarrow \infty} \sum_{j=1}^N x_j P(X = x_j) = \frac{1}{2} \log(2) \neq 0$$

(denn

$$\sum_{j=1}^{3N} x_j P(X = x_j) = \sum_{i=2}^N \frac{-i}{2i(i-1)} + \sum_{i=2}^{2N} \frac{i}{2i(i-1)} = -\frac{1}{2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{2} \sum_{k=1}^{2N-1} \frac{1}{k} \sim -\frac{1}{2} \log(N) + \frac{1}{2} \log(2N) = \frac{1}{2} \log(2)).$$

Beispiel 1.67 (Fortsetzung)

- ③ Wenn man andererseits S durchnummerierte mit $x_{3j} = -j - 1$, $x_{3j-2} = 2j$, $x_{3j-1} = 2j + 1$, $j \in \mathbb{N}$, so wäre

$$\sum_{j=1}^{\infty} x_j P(X = x_j) = \lim_{N \rightarrow \infty} \sum_{j=1}^N x_j P(X = x_j) = \frac{1}{2} \log(2) \neq 0$$

(denn

$$\sum_{j=1}^{3N} x_j P(X = x_j) = \sum_{i=2}^N \frac{-i}{2i(i-1)} + \sum_{i=2}^{2N} \frac{i}{2i(i-1)} = -\frac{1}{2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{2} \sum_{k=1}^{2N-1} \frac{1}{k} \sim -\frac{1}{2} \log(N) + \frac{1}{2} \log(2N) = \frac{1}{2} \log(2)).$$

(Wir sehen hier ein Beispiel für die Tatsache aus der Analysis, dass der Wert einer bedingt konvergenten Reihe von der Summationsreihenfolge abhängt.)

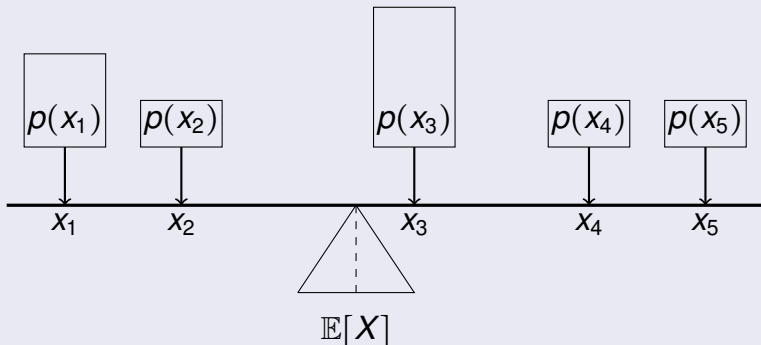
Bemerkung 1.68

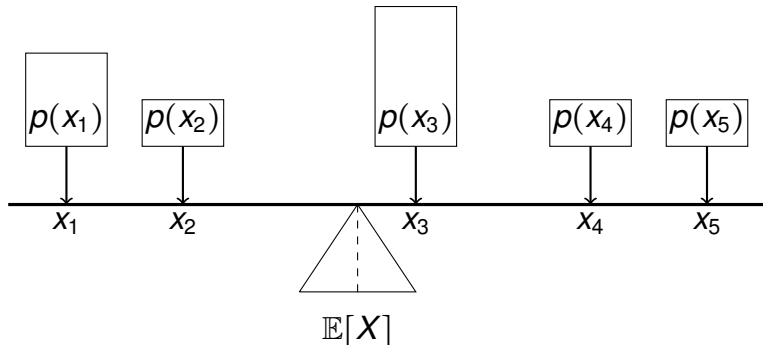
- 1 Eine beschränkte reellwertige ZV X
(d.h. es gibt eine Konstante $M < \infty$ mit
 $P(-M \leq X \leq M) = 1$)
besitzt stets einen Erwartungswert.

(Dies gilt insbesondere, wenn X nur endlich viele mögliche Werte hat.)

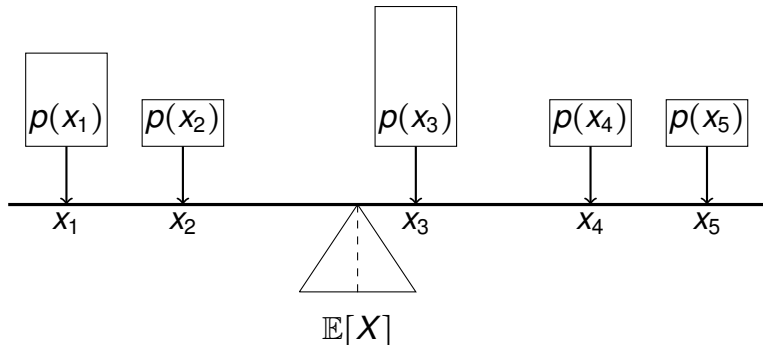
Bemerkung 1.68 (Fortsetzung)

- 2 Wenn X endlich viele mögliche Werte x_1, \dots, x_n (mit Gewichten $p(x_i) = P(X = x_i)$) hat, so kann man $\mathbb{E}[X]$ als den „Massenschwerpunkt“ interpretieren.





Auf einer Balkenwaage (deren Balken Eigengewicht 0 habe) liege an der Position x_i das Gewicht $p(x_i)$, damit der Balken in Ruhelage ist, muss man in an der Stelle $\sum_{i=1}^n x_i p(x_i) = \mathbb{E}[X]$ unterstützen



Auf einer Balkenwaage (deren Balken Eigengewicht 0 habe) liege an der Position x_i das Gewicht $p(x_i)$, damit der Balken in Ruhelage ist, muss man in an der Stelle $\sum_{i=1}^n x_i p(x_i) = \mathbb{E}[X]$ unterstützen, denn dann ist das Gesamtdrehmoment (proportional zu)

$$\sum_{i=1}^n p(x_i)(x_i - \mathbb{E}[X]) = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

Bemerkung 1.68 (Fortsetzung)

- ③ Der Erwartungswert von X muss nicht notwendigerweise ein möglicher Wert von X sein ($P(X = \mathbb{E}[X]) = 0$ ist durchaus möglich, siehe Bsp. 1.67),
daher kann man die Interpretation von $\mathbb{E}[X]$ als „typischer Wert von X “ i.A. nicht wörtlich nehmen.

Bemerkung 1.68 (Fortsetzung)

- ③ Es gilt aber: Sind X_1, X_2, \dots unabhängig mit derselben Verteilung wie X , so konvergiert

$$M_n := \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} \mathbb{E}[X] = \sum_x xP(X = x)$$

(in geeignetem Sinn), dies ist die Aussage des *Gesetzes der großen Zahlen*, das wir später sehen werden.

Bemerkung 1.68 (Fortsetzung)

- ③ Es gilt aber: Sind X_1, X_2, \dots unabhängig mit derselben Verteilung wie X , so konvergiert

$$M_n := \frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{n \rightarrow \infty} \mathbb{E}[X] = \sum_x xP(X = x)$$

(in geeignetem Sinn), dies ist die Aussage des *Gesetzes der großen Zahlen*, das wir später sehen werden.

Es ist nämlich

$$M_n = \sum_x x \cdot \frac{\#\{i \leq n : X_i = x\}}{n}$$

und $\#\{i \leq n : X_i = x\}/n \xrightarrow{n \rightarrow \infty} P(X = x)$.

Illustration: X_1, X_2, \dots uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie

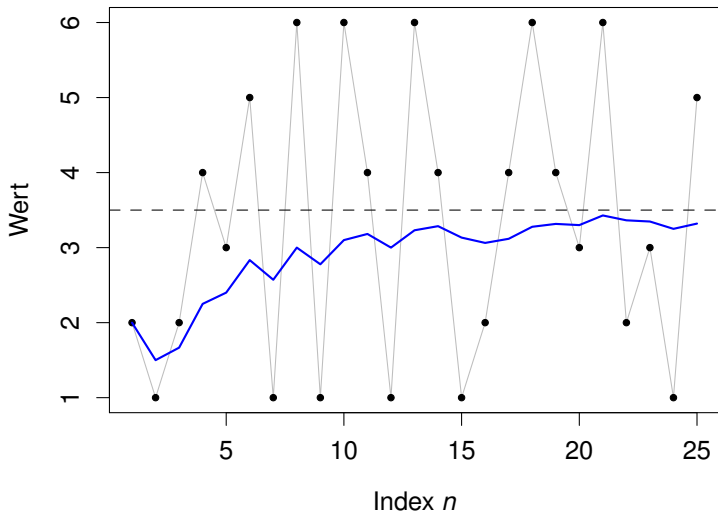


Illustration: X_1, X_2, \dots uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie

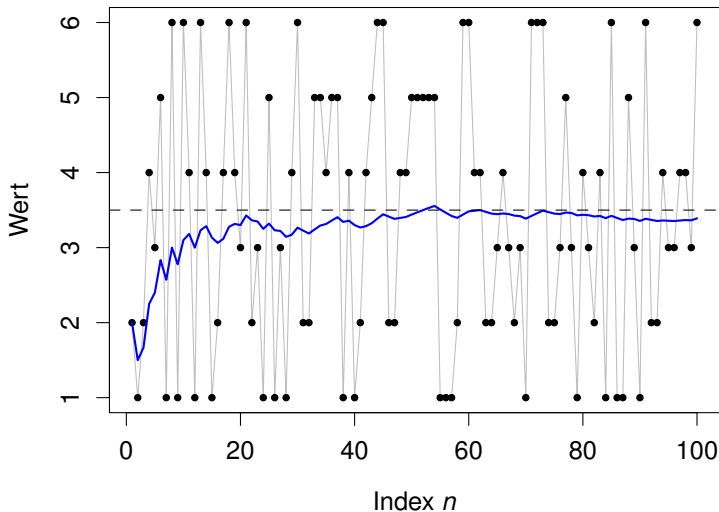


Illustration: X_1, X_2, \dots uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie

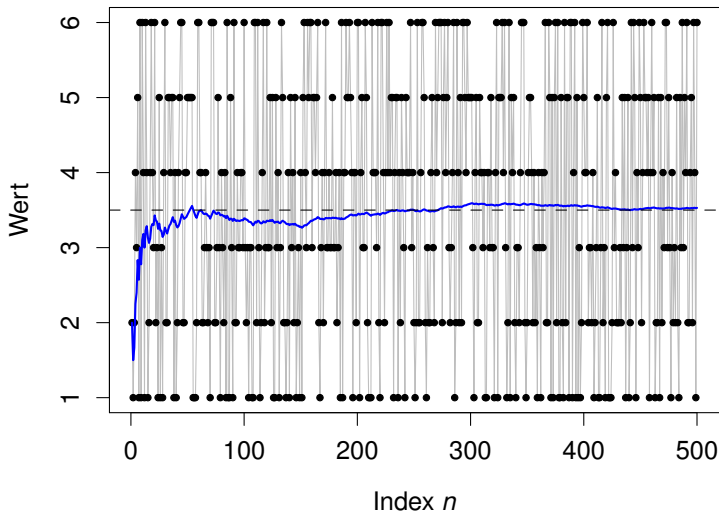
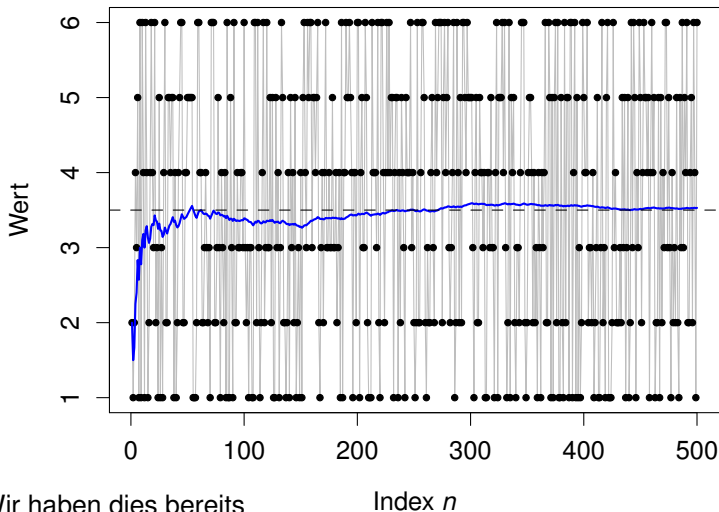


Illustration: X_1, X_2, \dots uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie



Wir haben dies bereits
in Kap. 0 („eine Monte-Carlo-Methode zur Integration“) verwendet.

Bemerkung 1.68 (Fortsetzung)

- 4 Man kann $\mathbb{E}[X]$ als den erforderlichen Einsatz in einem „fairen Spiel“ interpretieren, bei dem man eine zufällige Auszahlung X erhält.

Bemerkung 1.68 (Fortsetzung)

- ④ Man kann $\mathbb{E}[X]$ als den erforderlichen Einsatz in einem „fairen Spiel“ interpretieren, bei dem man eine zufällige Auszahlung X erhält.
- ⑤ Der Erwartungswert ist eine Eigenschaft der Verteilung: $\mathcal{L}(X) = \mathcal{L}(Y)$ impliziert $\mathbb{E}[X] = \mathbb{E}[Y]$.
(Klar, da dann $P(X = x) = P(Y = x)$ für alle x gilt.)

Beispiel 1.69

1 Sei $X \sim \text{Bin}_{n,p}$, $n \in \mathbb{N}$, $p \in [0, 1]$:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^n kP(X = k) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \text{Bin}_{n-1,p}(\{0, 1, \dots, n-1\}) \\ &= np\end{aligned}$$

Beispiel 1.69 (Fortsetzung)

2 Sei $X \sim \text{Geom}_p$, $p \in [0, 1]$:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} np(1-p)^n = p \sum_{n=1}^{\infty} \sum_{k=1}^n 1 \cdot (1-p)^n \\ &= p \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} (1-p)^n = p \sum_{k=1}^{\infty} (1-p)^k \sum_{j=0}^{\infty} (1-p)^j \\ &= \sum_{k=1}^{\infty} (1-p)^k = \frac{1}{p} - 1\end{aligned}$$

Beispiel 1.69 (Fortsetzung)

2 Sei $X \sim \text{Geom}_p$, $p \in [0, 1]$:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} np(1-p)^n = p \sum_{n=1}^{\infty} \sum_{k=1}^n 1 \cdot (1-p)^n \\ &= p \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} (1-p)^n = p \sum_{k=1}^{\infty} (1-p)^k \sum_{j=0}^{\infty} (1-p)^j \\ &= \sum_{k=1}^{\infty} (1-p)^k = \frac{1}{p} - 1\end{aligned}$$

3 Sei $X \sim \text{Poi}_\alpha$, $\alpha > 0$:

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} ne^{-\alpha} \frac{\alpha^n}{n!} = \alpha \sum_{n=1}^{\infty} e^{-\alpha} \frac{\alpha^{n-1}}{(n-1)!} = \alpha$$

Satz 1.70 (Rechenregeln für Erwartungswerte)

Seien $X, Y, X_1, X_2, \dots, Y_1, Y_2, \dots \in \mathcal{L}^1(P)$.

1. (Linearität) Für $a, b \in \mathbb{R}$ gilt $aX + bY \in \mathcal{L}^1(P)$ und

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Satz 1.70 (Rechenregeln für Erwartungswerte)

Seien $X, Y, X_1, X_2, \dots, Y_1, Y_2, \dots \in \mathcal{L}^1(P)$.

1. (Linearität) Für $a, b \in \mathbb{R}$ gilt $aX + bY \in \mathcal{L}^1(P)$ und

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2. (Monotonie) Wenn $X \geq Y$ (es genügt $P(X \geq Y) = 1$), so gilt $\mathbb{E}[X] \geq \mathbb{E}[Y]$; insbesondere gilt $\mathbb{E}[X] \geq 0$ für $X \geq 0$.

Satz 1.70 (Rechenregeln für Erwartungswerte)

Seien $X, Y, X_1, X_2, \dots, Y_1, Y_2, \dots \in \mathcal{L}^1(P)$.

1. (Linearität) Für $a, b \in \mathbb{R}$ gilt $aX + bY \in \mathcal{L}^1(P)$ und

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2. (Monotonie) Wenn $X \geq Y$ (es genügt $P(X \geq Y) = 1$), so gilt $\mathbb{E}[X] \geq \mathbb{E}[Y]$; insbesondere gilt $\mathbb{E}[X] \geq 0$ für $X \geq 0$.

3. $P(X \geq 0) = 1$ und $\mathbb{E}[X] = 0 \Rightarrow P(X = 0) = 1$.

Satz 1.70 (Rechenregeln für Erwartungswerte)

Seien $X, Y, X_1, X_2, \dots, Y_1, Y_2, \dots \in \mathcal{L}^1(P)$.

1. (Linearität) Für $a, b \in \mathbb{R}$ gilt $aX + bY \in \mathcal{L}^1(P)$ und

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

2. (Monotonie) Wenn $X \geq Y$ (es genügt $P(X \geq Y) = 1$), so gilt $\mathbb{E}[X] \geq \mathbb{E}[Y]$; insbesondere gilt $\mathbb{E}[X] \geq 0$ für $X \geq 0$.

3. $P(X \geq 0) = 1$ und $\mathbb{E}[X] = 0 \Rightarrow P(X = 0) = 1$.

4. (Faktorisierung für unabhängige Produkte) Wenn X und Y unabhängig sind, so ist $XY \in \mathcal{L}^1(P)$ und

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y].$$

Beweis.

1. Beachte, dass $aX + bY$ ebenfalls diskret ist, der Wertebereich $\{ax + by : x \in S_X, y \in S_Y\}$ ist abzählbar. Es ist

$$\begin{aligned} \sum_z |z| P(aX + bY = z) &= \sum_{x,y} |ax + by| \underbrace{P(X = x, Y = y)}_{\leq |a||x| + |b||y|} \\ &\leq |a| \sum_x |x| P(X = x) + |b| \sum_y |y| P(Y = y) < \infty, \end{aligned}$$

d.h. $aX + bY \in \mathcal{L}^1(P)$. Analog ist

$$\begin{aligned} \mathbb{E}[aX + bY] &= \sum_{x,y} (ax + by) P(X = x, Y = y) \\ &= a \sum_{x,y} x P(X = x, Y = y) + b \sum_{x,y} y P(X = x, Y = y) \\ &= a\mathbb{E}[X] + b\mathbb{E}[Y]. \end{aligned}$$

Beweis.

2.

$$\begin{aligned}\mathbb{E}[X] &= \sum_x xP(X = x) = \sum_{x,y} x \underbrace{P(X = x, Y = y)}_{=0 \text{ falls } y > x} \\ &\geq \sum_{x,y} yP(X = x, Y = y) \\ &= \sum_y yP(Y = y) = \mathbb{E}[Y]\end{aligned}$$

Beweis.

2.

$$\begin{aligned}\mathbb{E}[X] &= \sum_x xP(X = x) = \sum_{x,y} x \underbrace{P(X = x, Y = y)}_{=0 \text{ falls } y > x} \\ &\geq \sum_{x,y} yP(X = x, Y = y) \\ &= \sum_y yP(Y = y) = \mathbb{E}[Y]\end{aligned}$$

3. Sei $X \geq 0$. $\mathbb{E}[X] = \sum_{x(\geq 0)} xP(X = x)$ wäre > 0 , wenn $P(X = x) > 0$ für ein $x > 0$ gälte. □

Beweis.

4. Beachte, dass XY wiederum diskret ist. Weiter ist

$$\begin{aligned}\sum_z |z| P(XY = z) &= \sum_{x,y \neq 0} |xy| \underbrace{P(X = x, Y = y)}_{=P(X=x)P(Y=y)} \\ &= \sum_{x \neq 0} |x| P(X = x) \cdot \sum_{y \neq 0} |y| P(Y = y) = \mathbb{E}[|X|] \mathbb{E}[|Y|],\end{aligned}$$

d.h. $XY \in \mathcal{L}^1(P)$. Analog folgt

$$\begin{aligned}\mathbb{E}[XY] &= \sum_z z P(XY = z) = \sum_{x,y \neq 0} xy P(X = x, Y = y) \\ &= \sum_{x \neq 0} x P(X = x) \cdot \sum_{y \neq 0} y P(Y = y) \\ &= \mathbb{E}[X] \mathbb{E}[Y].\end{aligned}$$



Beobachtung 1.71 (Erwartungswerte für Kompositionen)

X (diskrete) reelle ZV, $g : \mathbb{R} \rightarrow \mathbb{R}$, $Y := g(X)$.

Dann besitzt Y einen Erwartungswert g.d.w.

$\sum_x |g(x)|P(X = x) < \infty$ und in diesem Fall ist

$$\mathbb{E}[Y] = \sum_x g(x)P(X = x).$$

Beobachtung 1.71 (Erwartungswerte für Kompositionen)

X (diskrete) reelle ZV, $g : \mathbb{R} \rightarrow \mathbb{R}$, $Y := g(X)$.

Dann besitzt Y einen Erwartungswert g.d.w.

$\sum_x |g(x)|P(X = x) < \infty$ und in diesem Fall ist

$$\mathbb{E}[Y] = \sum_x g(x)P(X = x).$$

(Schreibe $\sum_y yP(Y = y) = \sum_y \sum_{x:g(x)=y} g(x)P(X = x) = \sum_x g(x)P(X = x)$.)

Beispiel 1.72

1. Seien X_1, \dots, X_n u.i.v., $\sim \text{Ber}_p$, so ist
 $X := X_1 + \dots + X_n \sim \text{Bin}_{n,p}$ und

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = np.$$

(Wir hatten den Erwartungswert einer binomialverteilten ZV bereits in Bsp. 1.69, 2. bestimmt, hier kommen wir allerdings ohne explizite Rechnung aus.)

Beispiel 1.72 (Fortsetzung)

2. Sei $X \sim \text{Hyp}_{s,w,k}$ hypergeometrisch verteilt,

$$(P(X = \ell) = \frac{\binom{s}{\ell} \binom{w}{k-\ell}}{\binom{s+w}{k}}, \text{ dies war Bsp. 1.17})$$

Denken wir an eine Urne mit s schwarzen und w weißen Kugeln, aus der k mal ohne Zurücklegen gezogen wird, so ist

$$X \stackrel{d}{=} \mathbf{1}_{A_1} + \dots + \mathbf{1}_{A_k} \quad \text{mit } A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$$

und $P(A_1) = P(A_2) = \dots = P(A_k) = \frac{s}{s+w}$, also

$$\mathbb{E}[X] = k \cdot \frac{s}{s+w}.$$

Definition 1.73

Sei X reellwertige ZV mit Dichte f_X , dann besitzt X einen Erwartungswert (auch $X \in \mathcal{L}^1$ geschrieben), wenn gilt $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$ und man setzt dann

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) dx.$$

Beispiel 1.74

1. $X \sim \mathcal{N}_{0,1}$ hat $\mathbb{E}[X] = 0$, denn aus der Symmetrie der Dichte folgt

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx = 0\end{aligned}$$

Beispiel 1.74

1. $X \sim \mathcal{N}_{0,1}$ hat $\mathbb{E}[X] = 0$, denn aus der Symmetrie der Dichte folgt

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx = 0\end{aligned}$$

(strenggenommen muss man auch prüfen, dass
 $(2\pi)^{-1/2} \int_{-\infty}^{\infty} |x| e^{-x^2/2} dx = (2/\pi)^{1/2} \int_0^{\infty} x e^{-x^2/2} dx =$
 $(2/\pi)^{1/2} [-e^{-x^2/2}]_0^{\infty} = (\pi/2)^{-1/2} < \infty$)

Beispiel 1.74

1. $X \sim \mathcal{N}_{0,1}$ hat $\mathbb{E}[X] = 0$, denn aus der Symmetrie der Dichte folgt

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx = 0\end{aligned}$$

(strenggenommen muss man auch prüfen, dass $(2\pi)^{-1/2} \int_{-\infty}^{\infty} |x| e^{-x^2/2} dx = (2/\pi)^{1/2} \int_0^{\infty} x e^{-x^2/2} dx = (2/\pi)^{1/2} [-e^{-x^2/2}]_0^{\infty} = (\pi/2)^{-1/2} < \infty$)

Somit gilt auch: $Y \sim \mathcal{N}_{\mu, \sigma^2}$ hat $\mathbb{E}[Y] = \mu$, denn $\sigma X + \mu \stackrel{d}{=} Y$ nach Bsp. 1.32.

Beispiel 1.74 (Fortsetzung)

2. $X \sim \text{Exp}_\lambda$ hat $\mathbb{E}[X] = 1/\lambda$, denn

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x \lambda e^{-\lambda x} dx = \left[x(-e^{-\lambda x}) \right]_0^\infty - \int_0^\infty 1 \cdot (-e^{-\lambda x}) dx \\ &= \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} dx = \frac{1}{\lambda}\end{aligned}$$

Beispiel 1.74 (Fortsetzung)

2. $X \sim \text{Exp}_\lambda$ hat $\mathbb{E}[X] = 1/\lambda$, denn

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x \lambda e^{-\lambda x} dx = \left[x(-e^{-\lambda x}) \right]_0^\infty - \int_0^\infty 1 \cdot (-e^{-\lambda x}) dx \\ &= \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} dx = \frac{1}{\lambda}\end{aligned}$$

3. Die Cauchy-Verteilung mit Dichte $\frac{1}{\pi} \frac{1}{1+x^2}$ besitzt keinen Erwartungswert:

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{|x|}{1+x^2} dx = 2 \int_0^{\infty} \frac{1}{\pi} \frac{x}{1+x^2} dx = \left[\frac{1}{\pi} \log(1+x^2) \right]_0^{\infty} = \infty.$$

Bericht 1.75

1. Man kann prinzipiell den Fall mit Dichte aus dem diskreten Fall herleiten: X habe Dichte f_X , so nimmt $X_{(n)} = \frac{1}{n} \lfloor nX \rfloor$ den Wert $\frac{k}{n}$, $k \in \mathbb{N}$ an mit

$$P\left(X_{(n)} = \frac{k}{n}\right) = \int_{k/n}^{(k+1)/n} f_X(x) dx,$$

also ist (sofern die Reihe absolut konvergiert, was man analog überprüft)

$$\begin{aligned} \mathbb{E}[X_{(n)}] &= \sum_{k \in \mathbb{Z}} \frac{k}{n} \int_{k/n}^{(k+1)/n} f_X(x) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{n} \lfloor nx \rfloor f_X(x) dx \xrightarrow{n \rightarrow \infty} \int_{-\infty}^{\infty} x f_X(x) dx \end{aligned}$$

Bericht 1.75

2. (Analogon zu Beob. 1.71 im Fall mit Dichte)

Sei $X = (X_1, \dots, X_d)$ \mathbb{R}^d -wertig mit Dichte $f_X : \mathbb{R}^d \rightarrow [0, \infty]$,
 $g : \mathbb{R}^d \rightarrow \mathbb{R}$, $Y := g(X)$. Dann gilt $Y \in \mathcal{L}^1$ g.d.w.

$$\int_{\mathbb{R}^d} |g(x_1, \dots, x_d)| f_X(x_1, \dots, x_d) dx_1 \dots dx_d < \infty$$

und in diesem Fall

$$\mathbb{E}[Y] = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) f_X(x_1, \dots, x_d) dx_1 \dots dx_d < \infty.$$

Bericht 1.75

2. (Analogon zu Beob. 1.71 im Fall mit Dichte)

Sei $X = (X_1, \dots, X_d)$ \mathbb{R}^d -wertig mit Dichte $f_X : \mathbb{R}^d \rightarrow [0, \infty]$, $g : \mathbb{R}^d \rightarrow \mathbb{R}$, $Y := g(X)$. Dann gilt $Y \in \mathcal{L}^1$ g.d.w.

$$\int_{\mathbb{R}^d} |g(x_1, \dots, x_d)| f_X(x_1, \dots, x_d) dx_1 \dots dx_d < \infty$$

und in diesem Fall

$$\mathbb{E}[Y] = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) f_X(x_1, \dots, x_d) dx_1 \dots dx_d < \infty.$$

3. Die Rechenregeln aus Satz 1.70 gelten auch im Fall mit Dichte.

(Siehe z.B. das Buch von Georgii)

Für eine reellwertige Zufallsvariable X heißt $\mathbb{E}[X^2]$ das *2. Moment von X* (allgemein heißt $\mathbb{E}[X^p]$ das p -te Moment).

Man sagt, dass X ein 2. Moment besitzt, wenn $\mathbb{E}[X^2] < \infty$ gilt und schreibt dies auch als $X \in \mathcal{L}^2$ (bzw. $X \in \mathcal{L}^2(P)$, wenn die zugrundeliegende Wahrscheinlichkeiten $P(\cdot)$ nicht aus dem Kontext klar sind).

Definition 1.76

Für $X, Y \in \mathcal{L}^2$ heißt

- 1 $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ die *Varianz* von X
(manchmal schreibt man auch $\sigma_X^2 := \text{Var}[X]$)

Definition 1.76

Für $X, Y \in \mathcal{L}^2$ heißt

- 1 $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ die *Varianz* von X

(manchmal schreibt man auch $\sigma_X^2 := \text{Var}[X]$),

$\sqrt{\text{Var}[X]}$ die *Standardabweichung* (oder *Streuung*) von X

(manchmal auch $\sigma_X = \sqrt{\sigma_X^2}$ geschrieben),

Definition 1.76

Für $X, Y \in \mathcal{L}^2$ heißt

- ① $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ die *Varianz* von X

(manchmal schreibt man auch $\sigma_X^2 := \text{Var}[X]$),

$\sqrt{\text{Var}[X]}$ die *Standardabweichung* (oder *Streuung*) von X

(manchmal auch $\sigma_X = \sqrt{\sigma_X^2}$ geschrieben),

- ② $\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
 $= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

die *Kovarianz* von X und Y .

Definition 1.76

Für $X, Y \in \mathcal{L}^2$ heißt

- ① $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ die *Varianz* von X

(manchmal schreibt man auch $\sigma_X^2 := \text{Var}[X]$),

$\sqrt{\text{Var}[X]}$ die *Standardabweichung* (oder *Streuung*) von X

(manchmal auch $\sigma_X = \sqrt{\sigma_X^2}$ geschrieben),

- ② $\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$
 $= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

die *Kovarianz* von X und Y .

X und Y heißen *unkorreliert*, wenn $\text{Cov}[X, Y] = 0$.

Die Standardabweichung σ_X ist – neben dem Erwartungswert $\mathbb{E}[X]$ – eine weitere wichtige Kenngröße der Verteilung einer Zufallsvariable X , sie gibt eine Antwort auf die – etwas salopp formulierte – Frage „Wie sehr weicht X typischerweise von $\mathbb{E}[X]$ ab?“

Die Standardabweichung σ_X ist – neben dem Erwartungswert $\mathbb{E}[X]$ – eine weitere wichtige Kenngröße der Verteilung einer Zufallsvariable X , sie gibt eine Antwort auf die – etwas salopp formulierte – Frage „Wie sehr weicht X typischerweise von $\mathbb{E}[X]$ ab?“

Einen Hinweis dazu gibt die Chebyshev-Ungleichung \rightsquigarrow (nächste Folie)

Satz 1.77

Sei X reelle ZV, $f : [0, \infty) \rightarrow [0, \infty)$ monoton wachsend.

1. Für $a > 0$ mit $f(a) > 0$ gilt

$$P(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}[f(|X|)] \quad (\text{Markov}^1\text{-Ungleichung}). \quad (1)$$

2. Für $X \in \mathcal{L}^2$ gilt

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2} \quad (\text{Chebyshev}^2\text{-Ungleichung}). \quad (2)$$

¹Andrei Andrejewich Markov, 1856–1922.

²Pafnuty Lvovich Chebyshev, 1821–1894.

$$1. P(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}[f(|X|)]$$

$$2. P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}$$

$$1. P(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}[f(|X|)]$$

$$2. P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}$$

Beweis.

1. Sei $Y := f(a)I_{\{|X| \geq a\}}$, so ist $Y \leq f(|X|)$ und

$$\mathbb{E}[Y] = f(a)P(|X| \geq a) \leq \mathbb{E}[f(|X|)]$$

nach Satz 1.70, 2.

$$1. P(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}[f(|X|)]$$

$$2. P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}$$

Beweis.

1. Sei $Y := f(a)I_{\{|X| \geq a\}}$, so ist $Y \leq f(|X|)$ und

$$\mathbb{E}[Y] = f(a)P(|X| \geq a) \leq \mathbb{E}[f(|X|)]$$

nach Satz 1.70, 2.

2. Wende 1. an auf $\tilde{X} := X - \mathbb{E}[X]$ und $f(a) = a^2$. □

$$1. P(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}[f(|X|)]$$

$$2. P(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}$$

Beweis.

1. Sei $Y := f(a)I_{\{|X| \geq a\}}$, so ist $Y \leq f(|X|)$ und

$$\mathbb{E}[Y] = f(a)P(|X| \geq a) \leq \mathbb{E}[f(|X|)]$$

nach Satz 1.70, 2.

2. Wende 1. an auf $\tilde{X} := X - \mathbb{E}[X]$ und $f(a) = a^2$. □

Insbesondere (wähle $a = b\sigma_X$ in (2)):

Die W'keit $P(|X - \mathbb{E}[X]| \geq b\sigma_X)$, dass X von $\mathbb{E}[X]$ um mehr als das b -fache von σ_X abweicht, ist höchstens $1/b^2$.

Beobachtung 1.78

- 1 Wegen $|XY| \leq X^2 + Y^2$ ist die Kovarianz wohldefiniert.
Es gilt (offensichtlich)

$$\text{Cov}[X, Y] = \text{Cov}[Y, X].$$

Beobachtung 1.78

- 1 Wegen $|XY| \leq X^2 + Y^2$ ist die Kovarianz wohldefiniert.
Es gilt (offensichtlich)

$$\text{Cov}[X, Y] = \text{Cov}[Y, X].$$

- 2 Es ist

$$\begin{aligned} & \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[YX] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

(und analog für $\text{Var}[X] = \text{Cov}[X, X]$).

Beobachtung 1.78 (Fortsetzung)

- ③ $\text{Var}[X] = 0 \iff P(X = \mathbb{E}[X]) = 1$
(„ \Leftarrow “ ist klar, für „ \Rightarrow “ wende Satz 1.70, 3. an auf die ZV $(X - \mathbb{E}[X])^2$)

Beobachtung 1.78 (Fortsetzung)

- ③ $\text{Var}[X] = 0 \iff P(X = \mathbb{E}[X]) = 1$
(„ \Leftarrow “ ist klar, für „ \Rightarrow “ wende Satz 1.70, 3. an auf die ZV $(X - \mathbb{E}[X])^2$)

- ④ $\text{Var}[X]$ ist eine Eigenschaft der Verteilung von X ,
 $\text{Cov}[X, Y]$ ist eine Eigenschaft der gemeinsamen Verteilung von X und Y .

Beispiel 1.79

1. $X \sim \text{Ber}_p$, $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$.

Beispiel 1.79

1. $X \sim \text{Ber}_p$, $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$.

2. $X \sim \text{Poi}_\alpha$,

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) e^{-\alpha} \frac{\alpha^k}{k!} = \alpha^2 \sum_{k=2}^{\infty} e^{-\alpha} \frac{\alpha^{k-2}}{(k-2)!} = \alpha^2,$$

Beispiel 1.79

1. $X \sim \text{Ber}_p$, $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p)$.

2. $X \sim \text{Poi}_\alpha$,

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) e^{-\alpha} \frac{\alpha^k}{k!} = \alpha^2 \sum_{k=2}^{\infty} e^{-\alpha} \frac{\alpha^{k-2}}{(k-2)!} = \alpha^2,$$

also

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\ &= \alpha^2 + \alpha - \alpha^2 = \alpha \end{aligned}$$

Beispiel 1.79 (Fortsetzung)

3. $X \sim \text{Bin}_{n,p}$,

$$\begin{aligned}\mathbb{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} \\ &= n(n-1)p^2\end{aligned}$$

also

$$\begin{aligned}\text{Var}[X] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\ &= n(n-1)p^2 + np - (np)^2 = -np^2 + np \\ &= np(1-p)\end{aligned}$$

Beispiel 1.79 (Fortsetzung)

4. $X \sim \text{Geom}_p$, $p \in [0, 1]$ (wir hatten gesehen, dass $\mathbb{E}[X] = (1 - p)/p$, siehe Bsp. 1.69, 2)

Es ist

$$\begin{aligned} \mathbb{E}[X(X-1)] &= \sum_{n=0}^{\infty} n(n-1)p(1-p)^n \\ &= p(1-p)^2 \sum_{n=2}^{\infty} n(n-1)p(1-p)^{n-2} = 2 \frac{(1-p)^2}{p^2} \end{aligned}$$

(verwende, dass $f(t) := \sum_{n=0}^{\infty} t^n = \frac{1}{1-t}$ (für $|t| < 1$) erfüllt $\frac{d^2}{dt^2} f(t) = \frac{2}{(1-t)^3} = \sum_{n=2}^{\infty} n(n-1)t^{n-2}$), somit

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X](1 - \mathbb{E}[X]) \\ &= 2 \frac{(1-p)^2}{p^2} - \frac{1-p}{p} \cdot \frac{2p-1}{p} = \frac{1-p}{p^2}. \end{aligned}$$

Beispiel 1.79 (Fortsetzung)

5. $X \sim \mathcal{N}_{\mu, \sigma^2}$ hat $\text{Var}[X] = \sigma^2$ (wir hatten in Bsp. 1.74 bereits gesehen, dass $\mathbb{E}[X] = \mu$):

$$\begin{aligned}
 \text{Var}[X] &= \mathbb{E}[(X - \mu)^2] = \int_{\mathbb{R}} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx \\
 &= \int_{\mathbb{R}} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{z^2 e^{-z^2/2}}_{=z\left(-\frac{d}{dz} e^{-z^2/2}\right)} dz \\
 &= \frac{\sigma^2}{\sqrt{2\pi}} \left(\left[z \left(-\frac{d}{dz} e^{-z^2/2} \right) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-z^2/2} dz \right) \\
 &= \frac{\sigma^2}{\sqrt{2\pi}} (0 + \sqrt{2\pi}) = \sigma^2
 \end{aligned}$$

(Wir haben Bericht 1.75, 2. verwendet, dann im Integral $z = (x - \mu/\sigma)$ substituiert und partiell integriert.)

Satz 1.80 (Rechenregeln für Varianz und Kovarianz)

Seien $X, Y, X_1, X_2, \dots, X_n \in \mathcal{L}^2$, $a, b, c, d \in \mathbb{R}$.

1. $aX + b, cY + d \in \mathcal{L}^2$ und

$$\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y],$$

insbesondere $\text{Var}[aX + b] = a^2\text{Var}[X]$ (die Kovarianz ist eine Bilinearform, die Varianz ein quadratisches Funktional).

Satz 1.80 (Rechenregeln für Varianz und Kovarianz)

Seien $X, Y, X_1, X_2, \dots, X_n \in \mathcal{L}^2$, $a, b, c, d \in \mathbb{R}$.

1. $aX + b, cY + d \in \mathcal{L}^2$ und

$$\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y],$$

insbesondere $\text{Var}[aX + b] = a^2\text{Var}[X]$ (die Kovarianz ist eine Bilinearform, die Varianz ein quadratisches Funktional).

$$2. \text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i] + \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \text{Cov}[X_i, X_j],$$

insbesondere gilt für paarweise unkorrelierte X_1, \dots, X_n

$$\text{also } \text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i].$$

Satz 1.80 (Rechenregeln für Varianz und Kovarianz, 2)

3. Sind X und Y unabhängig, so gilt $\text{Cov}[X, Y] = 0$.

³nach Augustin-Louis Cauchy (1789–1857)
und Hermann Amandus Schwarz (1843–1921)

Satz 1.80 (Rechenregeln für Varianz und Kovarianz, 2)

3. Sind X und Y unabhängig, so gilt $\text{Cov}[X, Y] = 0$.

4. Es gilt

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \sqrt{\text{Var}[Y]}$$

(die Cauchy-Schwarz-Ungleichung³)

³nach Augustin-Louis Cauchy (1789–1857)
und Hermann Amandus Schwarz (1843–1921)

Beweis.

1. Es ist

$$\text{Cov}[aX + b, cY + d] = \text{Cov}[aX, cY]$$

(denn $\mathbb{E}[aX + b] = \mathbb{E}[aX] + b$ und $\mathbb{E}[cY + d] = \mathbb{E}[cY] + d$)

$$= \mathbb{E}[aX cY] - \mathbb{E}[aX] \mathbb{E}[cY]$$

$$= ac(\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y])$$

$$= ac \text{Cov}[X, Y].$$



Beweis.

2. Dies folgt etwa per Induktion über n aus 1., oder direkt folgendermaßen:

Sei o.E. $\mathbb{E}[X_1] = \dots = \mathbb{E}[X_n] = 0$ (sonst ziehe jeweils die Erwartungswerte ab, verwende 1.), dann ist

$$\begin{aligned}\operatorname{Var}\left[\sum_{i=1}^n X_i\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] = \sum_{i,j=1}^n \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^n \operatorname{Var}[X_i] + \sum_{i \neq j} \operatorname{Cov}[X_i, X_j]\end{aligned}$$



Beweis.

3. Klar: für X und Y unabhängig ist $\mathbb{E}[X Y] = \mathbb{E}[X] \mathbb{E}[Y]$
nach Satz 1.70, 4, also

$$\text{Cov}[X, Y] = \mathbb{E}[X Y] - \mathbb{E}[X] \mathbb{E}[Y] = 0$$

Beweis.

3. Klar: für X und Y unabhängig ist $\mathbb{E}[X Y] = \mathbb{E}[X] \mathbb{E}[Y]$ nach Satz 1.70, 4, also

$$\text{Cov}[X, Y] = \mathbb{E}[X Y] - \mathbb{E}[X] \mathbb{E}[Y] = 0$$

4. Falls $\text{Var}[Y] = 0$, so ist die Ungleichung

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \sqrt{\text{Var}[Y]} \text{ (als } 0 \leq 0 \text{) erfüllt}$$

(denn dann ist $P(Y - \mathbb{E}[Y] = 0) = 1$ nach Beob. 1.78, 3. und somit auch $\text{Cov}[X, Y] = 0$).

Beweis.

3. Klar: für X und Y unabhängig ist $\mathbb{E}[X Y] = \mathbb{E}[X] \mathbb{E}[Y]$ nach Satz 1.70, 4, also

$$\text{Cov}[X, Y] = \mathbb{E}[X Y] - \mathbb{E}[X] \mathbb{E}[Y] = 0$$

4. Falls $\text{Var}[Y] = 0$, so ist die Ungleichung

$$|\text{Cov}[X, Y]| \leq \sqrt{\text{Var}[X]} \sqrt{\text{Var}[Y]} \text{ (als } 0 \leq 0 \text{) erfüllt}$$

(denn dann ist $P(Y - \mathbb{E}[Y] = 0) = 1$ nach Beob. 1.78, 3. und somit auch $\text{Cov}[X, Y] = 0$).

Falls $\text{Var}[Y] > 0$, setze $\alpha := -\frac{\text{Cov}[X, Y]}{\text{Var}[Y]}$, es ist

$$0 \leq \text{Var}[X + \alpha Y] \text{Var}[Y]$$

$$\stackrel{!}{=} (\text{Var}[X] + 2\alpha \text{Cov}[X, Y] + \alpha^2 \text{Var}[Y]) \text{Var}[Y]$$

$$= \text{Var}[X] \text{Var}[Y] - (\text{Cov}[X, Y])^2.$$



Bemerkung 1.81

Es gilt Gleichheit in der Cauchy-Schwarz-Ungleichung g.d.w.

es gibt $a, b, c \in \mathbb{R}$ (mit $a \neq 0$ oder $b \neq 0$), so dass

$$P(aX + bY + c = 0) = 1.$$

In diesem Fall heißen X und Y *perfekt korreliert*.

Bemerkung 1.81

Es gilt Gleichheit in der Cauchy-Schwarz-Ungleichung g.d.w.

es gibt $a, b, c \in \mathbb{R}$ (mit $a \neq 0$ oder $b \neq 0$), so dass

$$P(aX + bY + c = 0) = 1.$$

In diesem Fall heißen X und Y *perfekt korreliert*.

(Denn wir sehen aus dem Beweis, dass Gleichheit genau dann eintritt, wenn $\text{Var}[Y] = 0$ oder $\text{Var}[X + \alpha Y] = 0$.)

Beispiel 1.82

- ① $X \sim \text{Bin}_{n,p}$, schreibe $X = Y_1 + \dots + Y_n$ mit Y_i u.i.v. $\sim \text{Ber}_p$, so ist (mit Satz 1.80, 2.)

$$\text{Var}[X] = \sum_{i=1}^n \text{Var}[Y_i] = n\text{Var}[Y_1] = np(1-p)$$

(vgl. auch Bsp. 1.79, 3.).

Beispiel 1.82 (Fortsetzung)

- ② $X \sim \text{Hyp}_{s,w,n}$, stelle dar als $X = Y_1 + \dots + Y_n$ mit $Y_i = \mathbf{1}_{A_i}$,
 $A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$
(bei n -fachem Ziehen ohne Zurücklegen aus einer Urne
mit s schwarzen und w weißen Kugeln)

Beispiel 1.82 (Fortsetzung)

② $X \sim \text{Hyp}_{s,w,n}$, stelle dar als $X = Y_1 + \dots + Y_n$ mit $Y_i = \mathbf{1}_{A_i}$,
 $A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$

(bei n -fachem Ziehen ohne Zurücklegen aus einer Urne mit s schwarzen und w weißen Kugeln)

Erinnerung: Für $y_1, \dots, y_n \in \{0, 1\}$ mit $y_1 + \dots + y_n = k$ gilt

$$P(Y_1 = y_1, \dots, Y_n = y_n) = \frac{s(s-1)(s-2)\cdots(s-k+1) \cdot w(w-1)\cdots(w-n+k+1)}{(s+w)(s+w-1)(s+w-2)\cdots(s+w-n+1)},$$

was nicht von der Reihenfolge abhängt – die Y_i sind „austauschbar“.

Beispiel 1.82 (Fortsetzung)

- ② Es ist $\mathbb{E}[Y_i] = P(A_i) = P(A_1) = \frac{s}{s+w} =: p$,
 $\text{Var}[Y_i] = p(1-p)$; für $i \neq j$ ist

$$\mathbb{E}[Y_i Y_j] = \mathbb{E}[Y_1 Y_2] = P(A_1 \cap A_2) = \frac{s}{s+w} \frac{s-1}{s+w-1},$$

$$\begin{aligned} \text{Cov}[Y_i, Y_j] &= \mathbb{E}[Y_i Y_j] - \mathbb{E}[Y_i] \mathbb{E}[Y_j] \\ &= \frac{s}{s+w} \frac{s-1}{s+w-1} - \left(\frac{s}{s+w} \right)^2 \\ &= \frac{s}{s+w} \underbrace{\left(\frac{s-1}{s+w-1} - \frac{s}{s+w} \right)}_{=-\frac{w}{s+w} \frac{1}{s+w-1}} \\ &= -p(1-p) \frac{1}{s+w-1}, \end{aligned}$$

Beispiel 1.82 (Fortsetzung)

② $X = Y_1 + \dots + Y_n \sim \text{Hyp}_{s,w,n}$
mit $Y_i = \mathbf{1}_{A_i}$, $A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$

Beispiel 1.82 (Fortsetzung)

② $X = Y_1 + \dots + Y_n \sim \text{Hyp}_{s,w,n}$
mit $Y_i = \mathbf{1}_{A_i}$, $A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$

Also (mit $p = s/(s + w)$)

$$\begin{aligned}\text{Var}[X] &= \sum_{i=1}^n \text{Var}[Y_i] + \sum_{i \neq j}^n \text{Cov}[Y_i, Y_j] \\ &= np(1-p) - n(n-1) \left(-p(1-p) \frac{1}{s+w-1} \right) \\ &= np(1-p) \left(1 - \frac{n-1}{s+w-1} \right)\end{aligned}$$

Beispiel 1.82 (Fortsetzung)

- ② $X = Y_1 + \dots + Y_n \sim \text{Hyp}_{s,w,n}$
mit $Y_i = \mathbf{1}_{A_i}$, $A_i = \{i\text{-te gezogene Kugel ist schwarz}\}$

Also (mit $p = s/(s+w)$)

$$\begin{aligned}\text{Var}[X] &= \sum_{i=1}^n \text{Var}[Y_i] + \sum_{i \neq j}^n \text{Cov}[Y_i, Y_j] \\ &= np(1-p) - n(n-1) \left(-p(1-p) \frac{1}{s+w-1} \right) \\ &= np(1-p) \left(1 - \frac{n-1}{s+w-1} \right)\end{aligned}$$

Wir sehen: Die Varianz ist kleiner im Fall ohne Zurücklegen als im Fall mit Zurücklegen – insbes. ist sie natürlich = 0 im Fall $n = s + w$.

Beispiel 1.82 (Fortsetzung)

- ③ Z reelle ZV mit $\mathbb{E}[|Z|^3] < \infty$ und *symmetrischer* Verteilung, d.h. es gilt $P(Z > z) = P(Z < -z)$ für alle $z \geq 0$ (z.B. $Z \sim \mathcal{N}_{0,1}$), setze

$$Y := Z^2$$

Beispiel 1.82 (Fortsetzung)

- ③ Z reelle ZV mit $\mathbb{E}[|Z|^3] < \infty$ und *symmetrischer* Verteilung, d.h. es gilt $P(Z > z) = P(Z < -z)$ für alle $z \geq 0$ (z.B. $Z \sim \mathcal{N}_{0,1}$), setze

$$Y := Z^2,$$

dann gilt

$$\begin{aligned}\text{Cov}[Y, Z] &= \mathbb{E}[Z^2 Z] - \mathbb{E}[Z^2]\mathbb{E}[Z] \\ &= \mathbb{E}[Z^3] - \mathbb{E}[Z^2]\mathbb{E}[Z] = 0 - \mathbb{E}[Z^2] \cdot 0 = 0.\end{aligned}$$

Beispiel 1.82 (Fortsetzung)

- ③ Z reelle ZV mit $\mathbb{E}[|Z|^3] < \infty$ und *symmetrischer* Verteilung, d.h. es gilt $P(Z > z) = P(Z < -z)$ für alle $z \geq 0$ (z.B. $Z \sim \mathcal{N}_{0,1}$), setze

$$Y := Z^2,$$

dann gilt

$$\begin{aligned}\text{Cov}[Y, Z] &= \mathbb{E}[Z^2 Z] - \mathbb{E}[Z^2]\mathbb{E}[Z] \\ &= \mathbb{E}[Z^3] - \mathbb{E}[Z^2]\mathbb{E}[Z] = 0 - \mathbb{E}[Z^2] \cdot 0 = 0.\end{aligned}$$

Z und Y sind also unkorreliert, aber i.A. *nicht* unabhängig.

Definition 1.83

Seien $X, Y \in \mathcal{L}^2$.

$$\kappa_{X,Y} := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \text{Var}[Y]}} \in [-1, 1]$$

heißt *Korrelationskoeffizient* von X und Y (manche Autoren schreiben auch $\rho_{X,Y}$).

(Die Cauchy-Schwarz-Ungleichung (Satz 1.80, 4.) zeigt, dass $|\kappa_{X,Y}| \leq 1$.)

Beobachtung 1.84 (Interpretation des Korrelationskoeffizienten via „beste lineare Vorhersage“)

Es ist

$$\begin{aligned} & \min_{\beta_0, \beta_1 \in \mathbb{R}} \mathbb{E}[(Y - \beta_1 X - \beta_0)^2] \\ &= (1 - \kappa_{X,Y}^2) \min_{\beta_0 \in \mathbb{R}} \mathbb{E}[(Y - \beta_0)^2] \quad \left(= (1 - \kappa_{X,Y}^2) \text{Var}[Y] \right), \end{aligned}$$

Denn

$$\begin{aligned}\mathbb{E}[(Y - \beta_1 X - \beta_0)^2] &= \text{Var}[Y - \beta_1 X - \beta_0] + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \text{Var}[Y] - 2\beta_1 \text{Cov}[X, Y] + \beta_1^2 \text{Var}[X] + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \sigma_Y^2 - 2\beta_1 \sigma_X \sigma_Y \kappa_{X,Y} + \beta_1^2 \sigma_X^2 + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \sigma_Y^2 (1 - \kappa_{X,Y}^2) + \sigma_X^2 \left(\beta_1 - \frac{\sigma_Y}{\sigma_X} \kappa_{X,Y} \right)^2 + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2\end{aligned}$$

Denn

$$\begin{aligned}\mathbb{E}[(Y - \beta_1 X - \beta_0)^2] &= \text{Var}[Y - \beta_1 X - \beta_0] + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \text{Var}[Y] - 2\beta_1 \text{Cov}[X, Y] + \beta_1^2 \text{Var}[X] + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \sigma_Y^2 - 2\beta_1 \sigma_X \sigma_Y \kappa_{X,Y} + \beta_1^2 \sigma_X^2 + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2 \\ &= \sigma_Y^2 (1 - \kappa_{X,Y}^2) + \sigma_X^2 \left(\beta_1 - \frac{\sigma_Y}{\sigma_X} \kappa_{X,Y} \right)^2 + (\mathbb{E}[Y] - \beta_1 \mathbb{E}[X] - \beta_0)^2\end{aligned}$$

was offensichtlich minimal wird für die Wahl

$$\beta_1 = \beta_1^* := \frac{\sigma_Y}{\sigma_X} \kappa_{X,Y}, \quad \beta_0 = \beta_0^* := \mathbb{E}[Y] - \beta_1^* \mathbb{E}[X]$$

und dann den Wert $(1 - \kappa_{X,Y}^2) \sigma_Y^2$ hat.

Für den Zusatz beachte analog:

$$\mathbb{E}[(Y - \beta)^2] = \mathbb{E}[Y^2] - 2\beta\mathbb{E}[Y] + \beta^2 = \text{Var}[Y] + (\beta - \mathbb{E}[Y])^2$$

ist minimal für die Wahl $\beta = \mathbb{E}[Y]$.

Für den Zusatz beachte analog:

$$\mathbb{E}[(Y - \beta_0)^2] = \mathbb{E}[Y^2] - 2\beta\mathbb{E}[Y] + \beta^2 = \text{Var}[Y] + (\beta - \mathbb{E}[Y])^2$$

ist minimal für die Wahl $\beta = \mathbb{E}[Y]$.

Im Sinne einer möglichst kleinen quadratischen Abweichung ist $\mathbb{E}[Y]$ die beste konstante „Vorhersage“ von Y . Man kann demnach um einen Faktor $(1 - \kappa_{X,Y}^2)$ besser vorhersagen, wenn man stattdessen eine affin-lineare Funktion von X verwenden darf.

Für den Zusatz beachte analog:

$$\mathbb{E}[(Y - \beta_0)^2] = \mathbb{E}[Y^2] - 2\beta\mathbb{E}[Y] + \beta^2 = \text{Var}[Y] + (\beta - \mathbb{E}[Y])^2$$

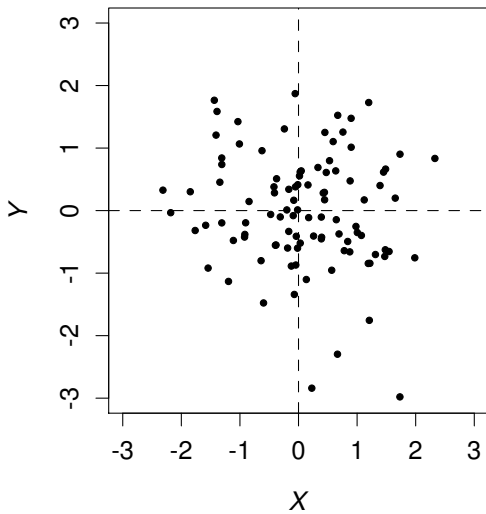
ist minimal für die Wahl $\beta = \mathbb{E}[Y]$.

Im Sinne einer möglichst kleinen quadratischen Abweichung ist $\mathbb{E}[Y]$ die beste konstante „Vorhersage“ von Y . Man kann demnach um einen Faktor $(1 - \kappa_{X,Y}^2)$ besser vorhersagen, wenn man stattdessen eine affin-lineare Funktion von X verwenden darf.

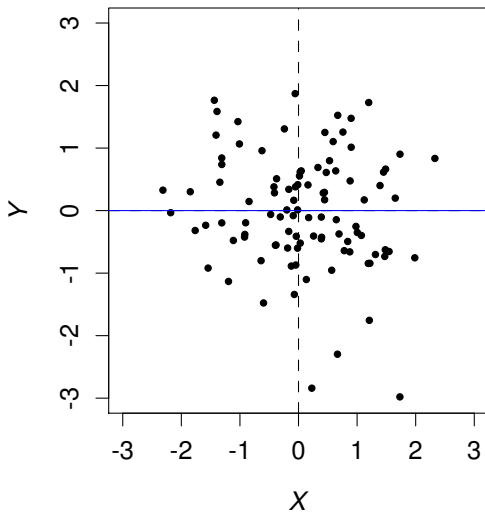
Die folgenden Scatterplots zeigen jeweils 100 simulierte Paare (X, Y) , wobei $\sigma_X = \sigma_Y = 1$ und $\kappa_{X,Y}$ den angegebenen Wert hat.

(Blau eingezeichnet ist die „Vorhersagegerade“
 $x \mapsto \beta_1^* x + \beta_0^*$.)

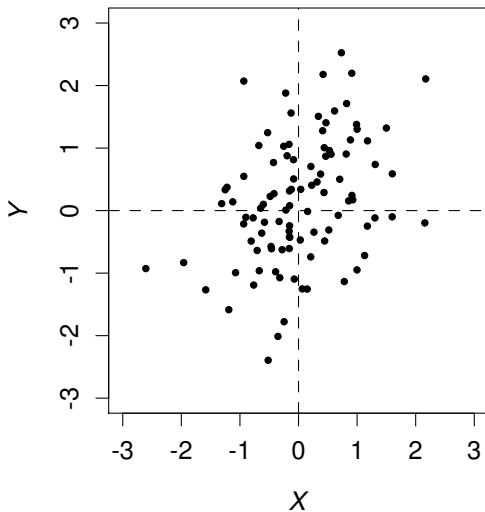
$$\kappa_{X,Y} = 0$$



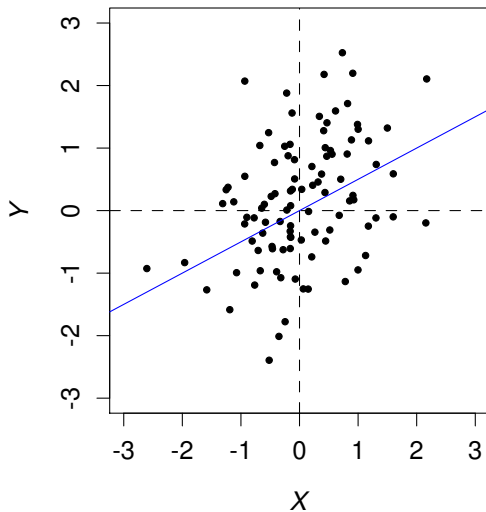
$$\kappa_{X,Y} = 0$$



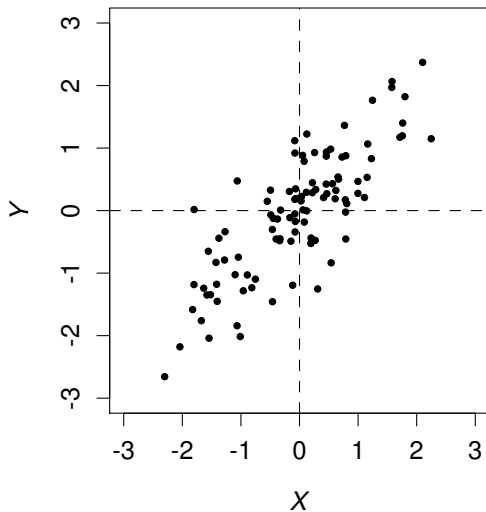
$$\kappa_{X,Y} = 0.5$$



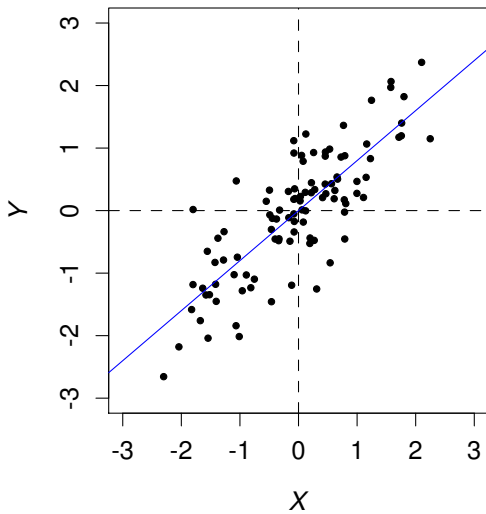
$$\kappa_{X,Y} = 0.5$$



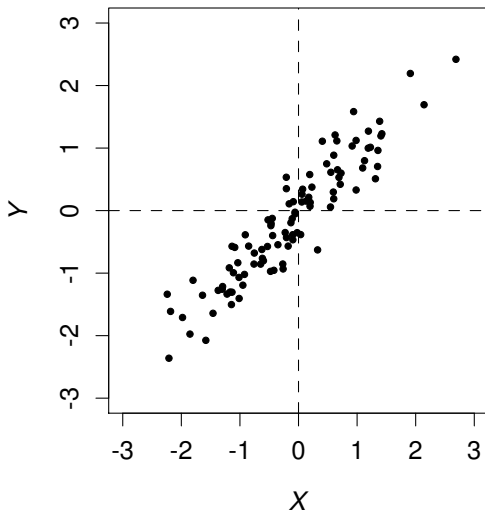
$$\kappa_{X,Y} = 0.8$$



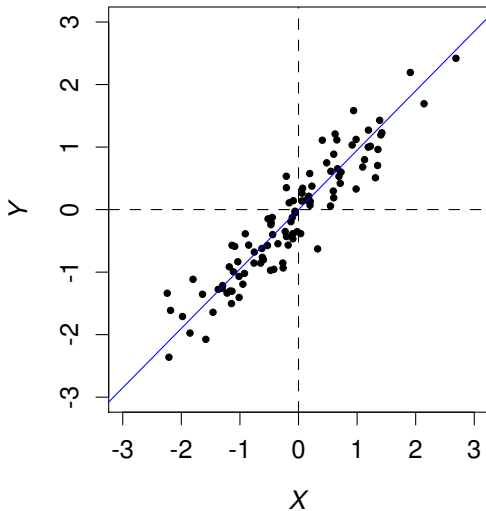
$$\kappa_{X,Y} = 0.8$$



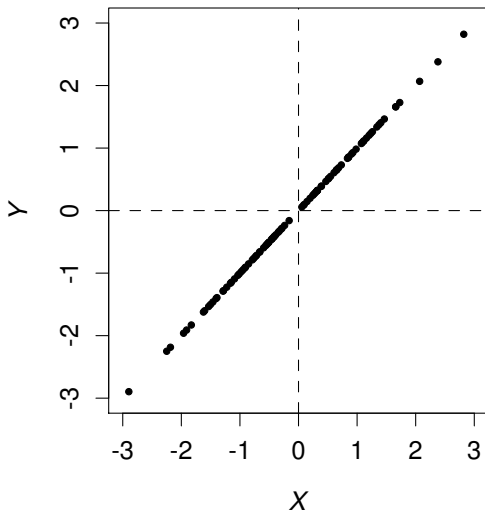
$$\kappa_{X,Y} = 0.95$$



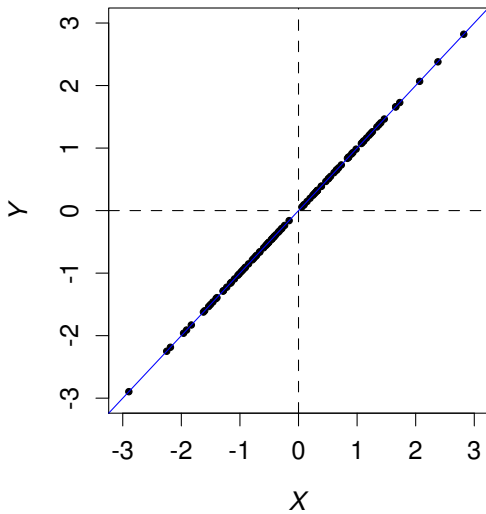
$$\kappa_{X,Y} = 0.95$$



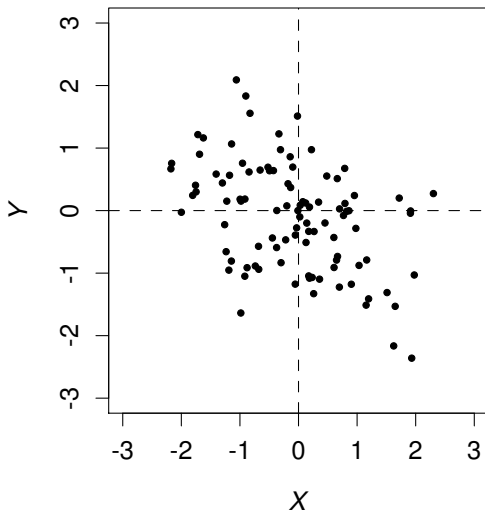
$$\kappa_{X,Y} = 1$$



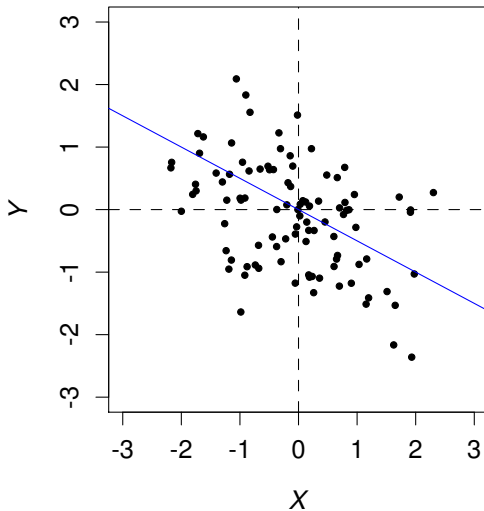
$$\kappa_{X,Y} = 1$$



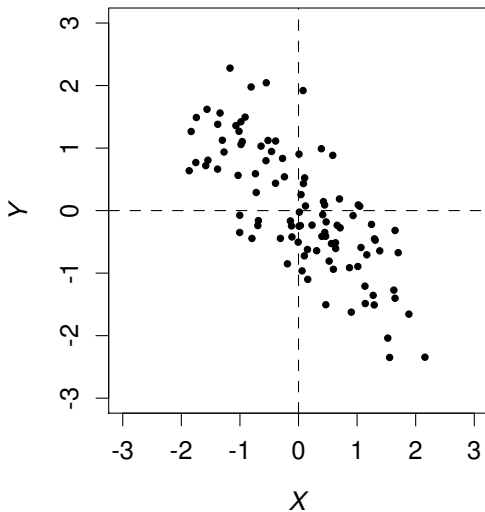
$$\kappa_{X,Y} = -0.5$$



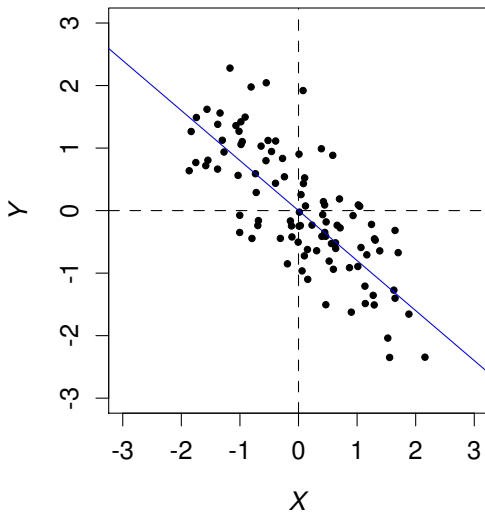
$$\kappa_{X,Y} = -0.5$$



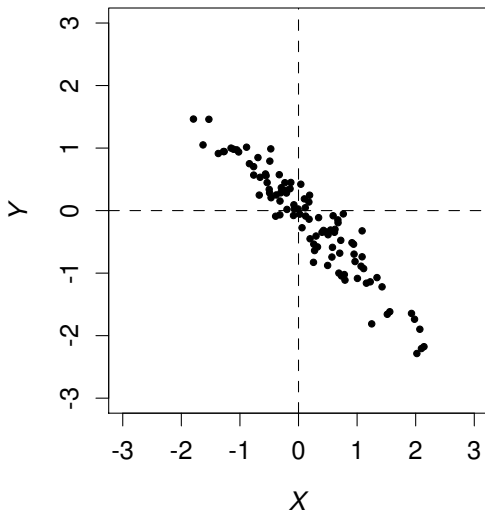
$$\kappa_{X,Y} = -0.8$$

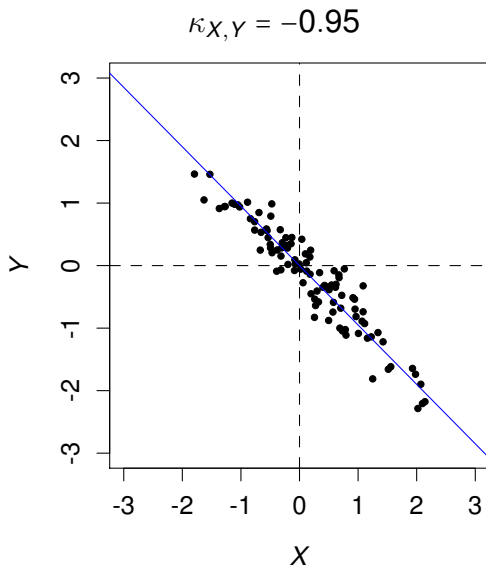


$$\kappa_{X,Y} = -0.8$$

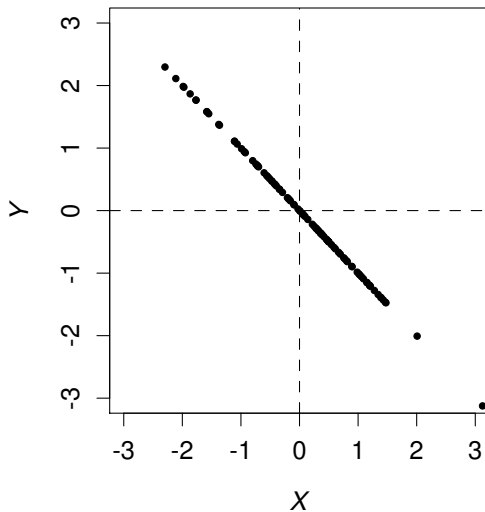


$$\kappa_{X,Y} = -0.95$$

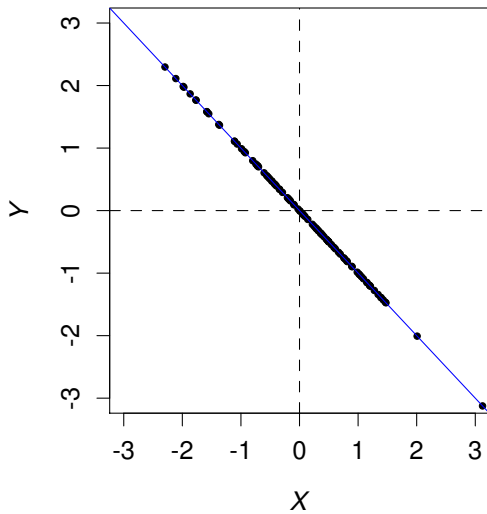




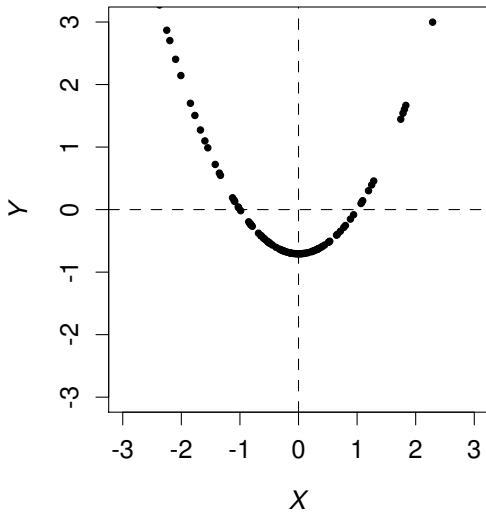
$$\kappa_{X,Y} = -1$$



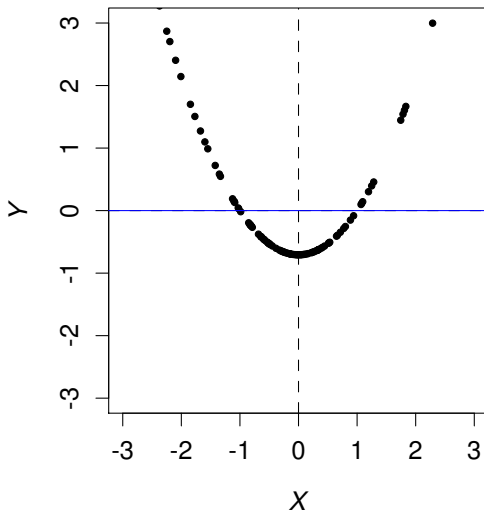
$$\kappa_{X,Y} = -1$$



$$\kappa_{X,Y} = 0$$



$$\kappa_{X,Y} = 0$$



Demnach (vgl. auch Bem. 1.81)

$|\kappa_{X,Y}| = 1 \quad \Leftrightarrow$ perfekter linearer Zusammenhang
zwischen X und Y

$\kappa_{X,Y} = 1 \quad \Leftrightarrow$ perfekter linearer Zusammenhang zwischen
 X und Y mit positivem Koeffizienten
(X größer als $\mathbb{E}[X]$ \Leftrightarrow Y größer als $\mathbb{E}[Y]$)

$\kappa_{X,Y} = -1 \quad \Leftrightarrow$ perfekter linearer Zusammenhang zwischen
 X und Y mit negativem Koeffizienten
(X größer als $\mathbb{E}[X]$ \Leftrightarrow Y kleiner als $\mathbb{E}[Y]$)

Nicht-lineare Zusammenhänge erfasst der
Korrelationskoeffizient möglicherweise nicht korrekt
(oder gar nicht), vgl. Bsp. 1.82, 3.