

Statistik für Informatiker, SS 2021

**1 Grundlagen aus der
Wahrscheinlichkeitstheorie**

1.4 Erwartungswert, Varianz und Kovarianz

Matthias Birkner

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?“

Diskreter Fall

Sei X reelle ZV mit abzählbarem Wertebereich (auf einem W'raum $(\Omega, \mathcal{F}, P)$ 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 $P(X = x)$, $x \in S$.

Diskreter Fall

Sei X reelle ZV mit abzählbarem Wertebereich (auf einem W'raum $(\Omega, \mathcal{F}, P)$ 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

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

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

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

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 $\frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$

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

d.h. X besitzt keinen Erwartungswert.

*zähle S_X auf
als: $-2, 2, -3, 3, -4, 4, \dots$*

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

$$\underbrace{(-2)P(X=-2) + 2P(X=2)}_{=0} + \underbrace{(-3)P(X=-3) + 3P(X=3)}_{=0} + \dots$$

Aufzählung: $2, 3, -2, 4, 5, -3, 6, 7, -4, \dots$

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

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

$$\begin{aligned} \text{(denn } \sum_{j=1}^{3N} x_j P(X = x_j) &= \sum_{i=2}^N \frac{-i}{2i(i-1)} + \sum_{i=2}^{2N} \frac{i}{2i(i-1)} = \\ &= -\frac{1}{2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{2} \sum_{k=1}^{2N-1} \frac{1}{k} \sim \underline{\underline{-\frac{1}{2} \log(N)}} + \underline{\underline{\frac{1}{2} \log(2N)}} = \underline{\underline{\frac{1}{2} \log(2)}}). \end{aligned}$$

$= \log 2 + \log N$

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

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

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

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

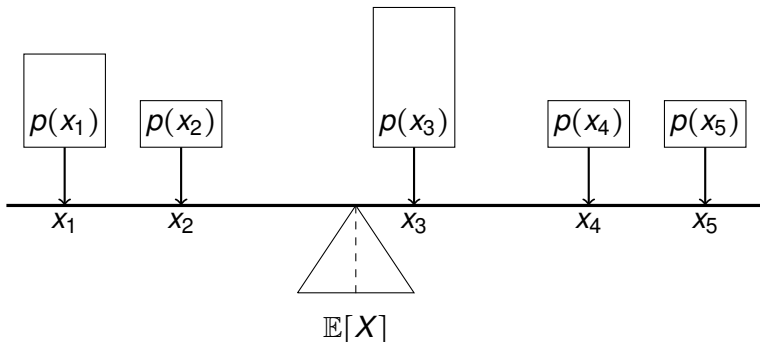
Bemerkung 1.68

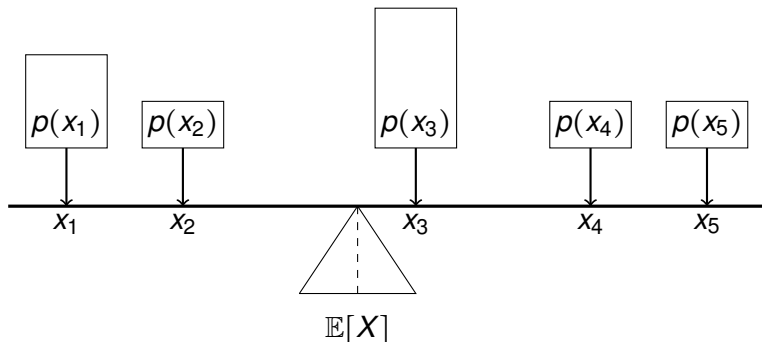
1. Eine beschränkte reellwertige ZV X
(d.h. es gibt eine Konstante $M < \infty$ mit $P(-M \leq X \leq M) = 1$)
besitzt stets einen Erwartungswert.

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

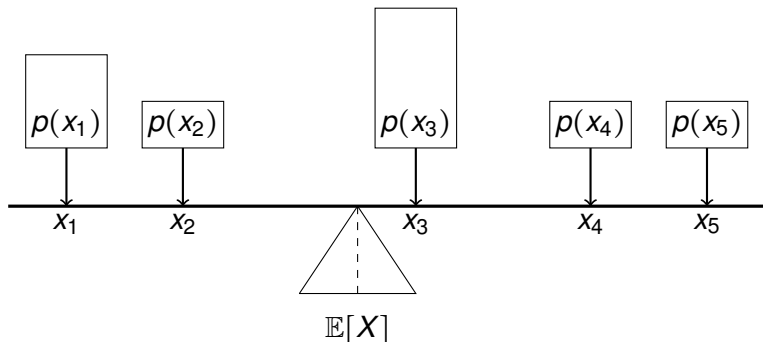
Bemerkung 1.68 (Fortsetzung)

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

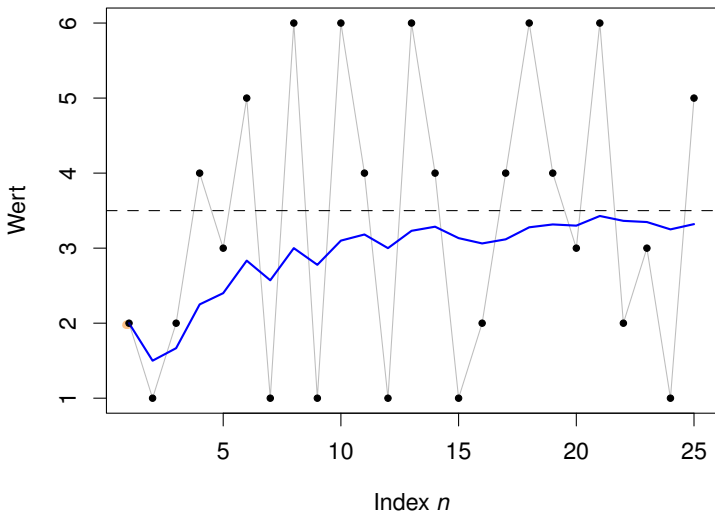


Illustration: $X_1, X_2, \dots$ uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie

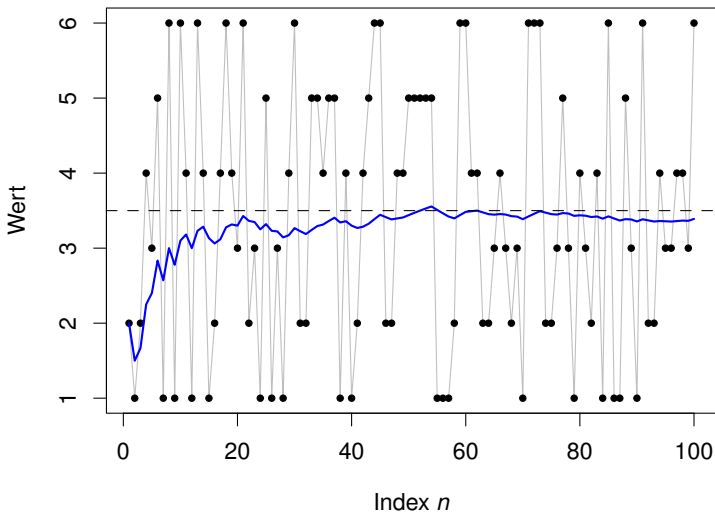


Illustration: $X_1, X_2, \dots$ uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie

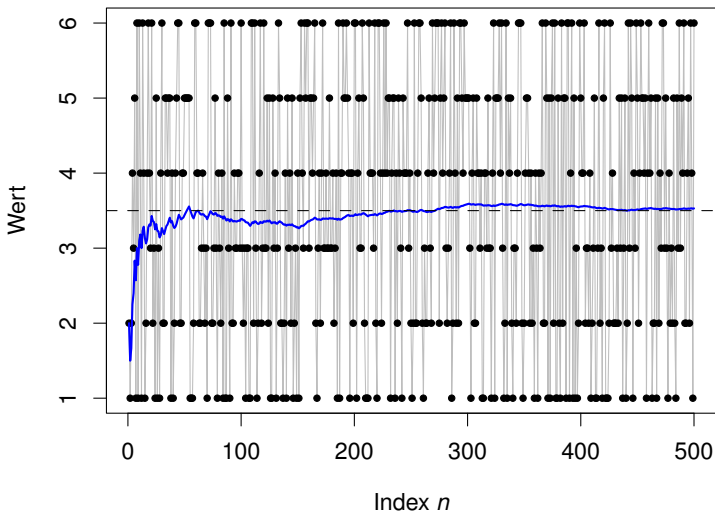
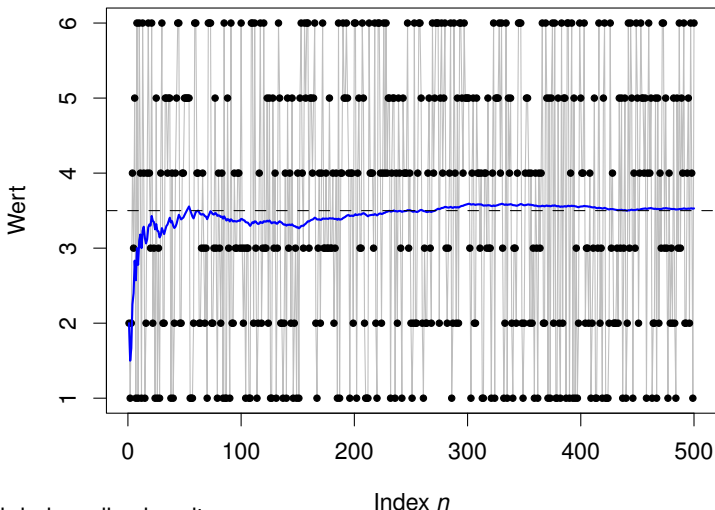


Illustration: $X_1, X_2, \dots$ uniform auf $\{1, 2, 3, 4, 5, 6\}$,
 X_n sind jeweils die schwarzen Punkte, M_n die blaue Linie



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

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

Handwritten notes: $\frac{n \cdot (n-1)!}{k \cdot (k-1)! \cdot ((n-1)-(k-1))!}$ (above the binomial coefficient)

Beispiel 1.69 (Fortsetzung)

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

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

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

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

Handwritten notes:
 - A bracket above the sum from $n=1$ to ∞ indicates $\alpha \cdot \alpha^{n-1} = 1$.
 - A bracket below the sum from $n=1$ to ∞ indicates $n \cdot (n-1)! = (n-1)!$.

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

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

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.

Linearität

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

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

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. Monotonie : Sei $X \geq Y$

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

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.

X, Y unabh.

← Wertebere. $\{x \cdot y : x \in S_X, y \in S_Y\}$

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

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

□

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

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


 $= p \cdot 1 + (1-p) \cdot 0 = p$

(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

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

$$\text{und } P(A_1) = P(A_2) = \dots = P(A_k) = \frac{s}{s+w}, \text{ also } \mathbb{E}[X] = k \cdot \frac{s}{s+w}.$$

Der Fall mit Dichte

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

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

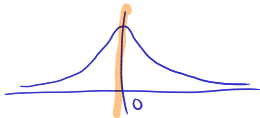
Beispiel 1.74

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

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

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-x^2/2} dx = 0$$



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.

/
Mittelwertparameter
(Erwartungswert)

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)

Varianz und Kovarianz

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).

Varianz und Kovarianz

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 $\sigma_X^2 := \text{Var}[X]$)

Varianz und Kovarianz

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),

Varianz und Kovarianz

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 .

Varianz und Kovarianz

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 $\leadsto$
(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

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

²Andrei 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 $1/b^2$.