Kapitel II
Diskrete Zufallsvariablen und ihre Kenngrößen

Wir werden in diesem Kapitel zeigen, daß sich das Ergebnis jedes diskreten ZE's als Realisierung (Wert) einer sogenannten *Zufallsvariable* beschreiben läßt und dessen Wahrscheinlichkeit dann vermöge der *Verteilung* dieser Zufallsvariablen bestimmt wird. Nach Einführung dieser Begriffe werden im weiteren Verlauf wichtige Kenngrößen wie den Erwartungswert und die Varianzdefinieren und ihre Eigenschaften untersuchen. Darüberhinaus erlauben Zufallsvariablen aber auch die formale Präzisierung des häufig gewollten Übergangs von einem gegebenen zu einem ”reduzierten” ZE, bei dem der ursprüngliche Ergebnisraum mittels einer Funktion in einen anderen (”kleineren”) Ergebnisraum überführt wird. Als typisches Beispiel nehme man beim Werfen von zwei Würfeln den Übergang von den einzelnen Augenzahlen (Ergebnisraum $= \{1, ..., 6\}^2$) zur Augensumme (Ergebnisraum $\{2, 3, ..., 11, 12\}$) [☞ 4.6].

4. Diskrete Zufallsvariablen und ihre Verteilungen

Zur Motivation der nachfolgenden Überlegungen betrachten wir zunächst ein Beispiel:

4.1. Beispiel. ***(n-facher Münzwurf)*** Wird eine faire Münze n-mal geworfen, so liegt bekanntlich ein Laplace-Experiment mit Ergebnisraum $\Omega = \{0, 1\}^n$ vor, sofern wir $\omega_k = 1$ bzw. 0 setzen, wenn im k-ten Wurf ”Zahl” bzw. ”Kopf” erscheint ($1 \le k \le n$). Nun ist man in der Regel nicht so sehr an den einzelnen Wurfergebnissen interessiert, sondern lediglich daran, wie oft die beiden Ergebnisse ”Zahl” und ”Kopf” aufgetreten sind. Betrachten wir die Häufigkeit von ”Zahl”, führt uns dies zu der Abbildung

$$S_n : \Omega \to \{0, ..., n\}, \quad (\omega_1, ..., \omega_n) \mapsto \sum_{k=1}^{n} \omega_k,$$

und weiter zu der Frage, mit welcher Wahrscheinlichkeit S_n die möglichen Werte $0, ..., n$ annimmt. Gesucht sind demnach die Wahrscheinlichkeiten der Ereignisse

$$E_k \stackrel{\text{def}}{=} \{(\omega_1, ..., \omega_n) : S_n(\omega_1, ..., \omega_n) = k\}, \quad k = 0, ..., n.$$

Da ein Laplace-Experiment zugrundeliegt, erhalten wir leicht

$$P(E_k) \;=\; \frac{|E_k|}{|\Omega|} \;=\; \frac{\binom{n}{k}}{2^n}$$

für jedes k. Für unsere Zwecke interessanter ist zu diesem Zeitpunkt jedoch die Beobachtung, daß vermöge der Abbildung S_n ein neues Zufallsexperiment *induziert* wird, und zwar mit Ergebnisraum $\hat{\Omega} = \{0, ..., n\}$ und Zähldichte $\hat{p}$, definiert durch

$$\hat{p}(k) \stackrel{\text{def}}{=} P(E_k) = \frac{\binom{n}{k}}{2^n} = \binom{n}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}.$$

Dabei folgt $\sum_{k=0}^n \hat{p}(k) = \sum_{k=0}^n P(E_k) = 1$ aus der Tatsache, daß die E_k, $0 \le k \le n$, die Menge Ω disjunkt zerlegen. Die zu $\hat{p}$ gehörende diskrete W-Verteilung $\hat{P}$ über $\hat{\Omega}$ ist im Fall $n \ge 2$ offenbar keine Laplace-Verteilung mehr, sondern eine Binomialverteilung mit Parametern n und $\frac{1}{2}$ (☞ Beispiel 1.8). Wir halten fest, daß sich $\hat{p}(k)$ ergibt, indem man feststellt, welche $\omega \in \Omega$ zu $S_n(\omega) = k$ führen, und dann deren Gesamtwahrscheinlichkeit im Ausgangsexperiment ausrechnet. Man bestimmt also die Wahrscheinlichkeit unter P für das *Urbild* der Menge $\{k\}$ unter S_n, definiert durch

$$S_n^{-1}(\{k\}) \stackrel{\text{def}}{=} \{\omega \in \Omega : S_n(\omega) = k\}.$$

Es wurde weiter oben mit E_k bezeichnet.

Will man für beliebiges $A \subset \hat{\Omega}$ die Wahrscheinlichkeit unter $\hat{P}$ berechnen, so folgt entsprechend

$$\hat{P}(A) = \sum_{k \in A} \hat{p}(k) = \sum_{k \in A} P(S_n^{-1}(\{k\})) = P(S_n^{-1}(A)),$$

wobei benutzt wird, daß die Urbilder $S_n^{-1}(\{k\})$, $k \in A$, p.d. sind und vereinigt das Urbild von A unter S_n ergeben. S_n^{-1} heißt die zu S_n gehörende *Urbildfunktion* und ist offenkundig definiert auf den Teilmengen des Wertebereichs von S_n. S_n^{-1} sollte somit nicht mit der Umkehrabbildung von S_n (die hier für $n \ge 2$ auch gar nicht existiert) verwechselt werden.

Um das anhand des vorherigen Beispiels skizzierte Vorgehen zu formalisieren, geben wir zunächst einige grundlegende Eigenschaften der Urbildfunktion.

4.2. Lemma. *Sei $X : \Omega \to \mathcal{X}$ eine Abbildung und $X^{-1} : \mathfrak{P}(\mathcal{X}) \to \mathfrak{P}(\Omega)$ die durch*

$$X^{-1}(A) \stackrel{\text{def}}{=} \{\omega \in \Omega : X(\omega) \in A\} \tag{4.1}$$

definierte Urbildfunktion. Dann gilt:

(a) $X^{-1}(A^c) = (X^{-1}(A))^c$ für alle $A \subset \mathcal{X}$.

(b) $X^{-1}(\cup_{i \in I} A_i) = \cup_{i \in I} X^{-1}(A_i)$ für beliebige (auch überabzählbare) Familien $(A_i)_{i \in I}$.

(c) $X^{-1}(\cap_{i \in I} A_i) = \cap_{i \in I} X^{-1}(A_i)$ für beliebige Familien $(A_i)_{i \in I}$.

*Aufgrund der Eigenschaften (a)-(c) nennt man X^{-1} **operationstreu**.*

(d) $X^{-1}(\sum_{i \in I} A_i) = \sum_{i \in I} X^{-1}(A_i)$ für beliebige Familien $(A_i)_{i \in I}$ p.d. Teilmengen von $\mathcal{X}$.

*(e) $X^{-1}(A_1) \subset X^{-1}(A_2)$ für alle $A_1 \subset A_2 \subset \mathcal{X}$ (X^{-1} ist also **ordnungstreu** bezüglich der partiellen Ordnung $\subset$ auf $\mathfrak{P}(\mathcal{X})$).*

Beweis: (a) ergibt sich vermöge

$$\omega \in X^{-1}(A^c) \;\Leftrightarrow\; X(\omega) \in A^c \;\Leftrightarrow\; X(\omega) \notin A \;\Leftrightarrow\; \omega \notin X^{-1}(A) \;\Leftrightarrow\; \omega \in (X^{-1}(A))^c.$$

Analog liefert

$$\begin{aligned} \omega \in X^{-1}(\cup_{i\in I} A_i) &\Leftrightarrow X(\omega) \in \cup_{i\in I} A_i \\ &\Leftrightarrow \exists i \in I : X(\omega) \in A_i \\ &\Leftrightarrow \exists i \in I : \omega \in X^{-1}(A_i) \\ &\Leftrightarrow \omega \in \cup_{i\in I} X^{-1}(A_i) \end{aligned}$$

Aussage (b). Den Beweis der übrigen Aussagen überlassen wir dem Leser. ◊

4.3. Satz. *Es sei (Ω, p) ein diskretes ZE mit zugehöriger diskreter W-Verteilung P und $X : \Omega \to \mathcal{X}$ eine Abbildung. Dann ist die durch*

$$P^X(A) \stackrel{\text{def}}{=} P(X^{-1}(A)), \quad A \subset \mathcal{X}, \tag{4.2}$$

definierte Abbildung $P^X : \mathfrak{P}(\mathcal{X}) \to [0,1]$ eine diskrete W-Verteilung über $\mathcal{X}$ mit der Zähldichte

$$p^X(x) \;=\; \sum_{\omega\in\Omega : X(\omega)=x} p(\omega), \quad x \in \mathcal{X}. \tag{4.3}$$

Beweis: Wegen $X^{-1}(\mathcal{X}) = \Omega$ gilt $P^X(\mathcal{X}) = P(\Omega) = 1$, d.h. P^X ist normiert. Gegeben eine beliebige Folge $(A_n)_{n\geq 1}$ p.d. Teilmengen von $\mathcal{X}$, ergibt sich außerdem vermöge Lemma 4.2(d) und der σ-Additivität von P

$$\begin{aligned} P^X\left(\sum_{n\geq 1} A_n\right) &= P\left(X^{-1}\left(\sum_{n\geq 1} A_n\right)\right) \\ &= P\left(\sum_{n\geq 1} X^{-1}(A_n)\right) \\ &= \sum_{n\geq 1} P(X^{-1}(A_n)) \\ &= \sum_{n\geq 1} P^X(A_n), \end{aligned}$$

also die σ-Additivität von P^X.

Sei Ω_0 der abzählbare Träger von P. Dann ist natürlich auch $\mathcal{X}_0 \stackrel{\text{def}}{=} X(\Omega_0)$ wieder abzählbar mit $X^{-1}(\mathcal{X}_0) \supset \Omega_0$. Es folgt

$$P^X(\mathcal{X}_0) \;=\; P(X^{-1}(\mathcal{X}_0)) \;\geq\; P(\Omega_0) \;=\; 1.$$

Nach Bemerkung 1.12 ist P^X damit tatsächlich eine *diskrete* W-Verteilung. Für ihre Zähldichte p^X ergibt sich (4.3) vermöge

$$p^X(x) \;=\; P^X(\{x\}) \;=\; P(X^{-1}(\{x\})) \;=\; P(\{\omega \in \Omega : X(\omega) = x\}) \;=\; \sum_{\omega\in\Omega : X(\omega)=x} p(\omega)$$

für alle $x \in \mathcal{X}$. $\diamondsuit$

Die vorherigen Überlegungen geben Anlaß zu folgenden Begriffsbildungen:

4.4. Definition. Eine Abbildung $X : \Omega \to \mathcal{X}$, definiert auf einer nichtleeren Menge Ω mit diskreter W-Verteilung P, heißt *diskrete Zufallsvariable*, im Fall $\mathcal{X} \subset \mathbb{R}$ auch *diskrete Zufallsgröße* und im Fall $\mathcal{X} \subset \mathbb{R}^d$, $d \geq 2$, auch *(d-dimensionaler) diskreter Zufallsvektor*. Die in Satz 4.3 eingeführte diskrete W-Verteilung P^X heißt die *durch X induzierte Verteilung* oder auch kurz *Verteilung von X (unter P)*.

Da wir es in diesem Text weitestgehend nur mit diskreten W-Verteilungen P zu tun haben, verzichten wir für diskrete Zufallsvariablen häufig auf das Attribut "diskret". Zur Illustration betrachten wir einige weitere Beispiele:

4.5. Beispiel. ***(Indikatorfunktionen)*** Sei P eine diskrete W-Verteilung über Ω und $E \subset \Omega$. Interessieren wir uns nur dafür, ob das Ereignis E eintritt oder nicht, so wird dies durch die Zufallsgröße

$$\mathbf{1}_E(\omega) \;\stackrel{\text{def}}{=}\; \begin{cases} 1, & \text{falls } \omega \in E \\ 0, & \text{falls } \omega \notin E \end{cases} \tag{4.4}$$

mit Wertebereich $\{0, 1\}$ beschrieben. $\mathbf{1}_E$ heißt *Indikatorfunktion von E*. Für ihre Verteilung $P^{\mathbf{1}_E}$ ergibt sich

$$P^{\mathbf{1}_E}(A) \;=\; P(\mathbf{1}_E^{-1}(A)) \;=\; \begin{cases} P(\Omega) = 1, & \text{falls } A = \{0,1\} \\ P(E), & \text{falls } A = \{1\} \\ P(E^c) = 1 - P(E), & \text{falls } A = \{0\} \\ 0, & \text{falls } A = \emptyset \end{cases}. \tag{4.5}$$

$P^{\mathbf{1}_E}$ ist also eine Bernoulli-Verteilung mit Parameter $\theta = P(E)$ (☞ Beispiel 1.6).

4.6. Beispiel. ***(Augensumme bei zweimaligem Würfeln)*** Wird ein fairer Würfel zweimal geworfen, beschrieben durch ein Laplace-Experiment mit Ergebnisraum $\Omega = \{1, ..., 6\}^2$, so gibt die Zufallsgröße $S : \Omega \to \{2, 3, ..., 11, 12\}$,

$$S(\omega_1, \omega_2) \;=\; \omega_1 + \omega_2$$

k	$S^{-1}(\{k\})$	$P^S(\{k\})$
2	$\{(1,1)\}$	$\frac{1}{36}$
3	$\{(1,2),(2,1)\}$	$\frac{2}{36}$
4	$\{(1,3),(2,2),(3,1)\}$	$\frac{3}{36}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$	$\frac{4}{36}$
6	$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$	$\frac{5}{36}$
7	$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$	$\frac{6}{36}$
8	$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$	$\frac{5}{36}$
9	$\{(3,6),(4,5),(5,4),(6,3)\}$	$\frac{4}{36}$
10	$\{(4,6),(5,5),(6,4)\}$	$\frac{3}{36}$
11	$\{(5,6),(6,5)\}$	$\frac{2}{36}$
12	$\{(6,6)\}$	$\frac{1}{36}$

Die Verteilung der Augensumme bei zweimaligem Würfeln

die Augensumme der beiden Würfe an. Zur Bestimmung ihrer Verteilung benötigen wir die Urbilder $S^{-1}(\{k\})$ für $k = 2, ...12$. Sie finden sich mit den zugehörigen Wahrscheinlichkeiten in der umseitigen Tabelle zusammengefaßt.

4.7. Beispiel. ***(Hypergeometrische Verteilung)*** Aus einer Urne mit r roten und s schwarzen Kugeln werden n Kugeln ohne Zurücklegen gezogen. Wir fragen nach der Verteilung der Zufallsgröße X, die die Anzahl der gezogenen roten Kugeln angibt. Wir denken uns die Kugeln mit $1, ..., r+s$ durchnumeriert, wobei die ersten r rot und die übrigen schwarz seien. Dann liegt ein Laplace-Experiment mit Ergebnisraum $\mathfrak{P}_n(\{1, ..., r+s\})$ der Mächtigkeit $\binom{r+s}{n}$ vor, und es gilt

$$X(\{\omega_1, ..., \omega_n\}) \ = \ |\{\omega_1, ..., \omega_n\} \cap \{1, ..., r\}|.$$

$X^{-1}(\{k\})$ besteht offenbar aus den n-elementigen Teilmengen von $\{1, ..., r+s\}$ mit genau k Elementen aus $\{1, ..., r\}$ und $n-k$ Elementen aus $\{r+1, ..., r+s\}$, vorausgesetzt $k \le \min(r, n)$ und $n-k \le \min(s, n)$, d.h. $\max(0, n-s) \le k \le \min(r, n)$. Es folgt $|X^{-1}(\{k\})| = \binom{r}{k}\binom{s}{n-k}$, denn es gibt $\binom{r}{k}$ Möglichkeiten k rote aus den r vorhandenen auszuwählen und entsprechend $\binom{s}{n-k}$ Möglichkeiten, die übrigen $n-k$ Kugel unter den schwarzen auszuwählen. Wir erhalten demnach

$$(4.6) \qquad P^X(\{k\}) \ = \ \begin{cases} \dfrac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, & \text{falls } \max(0, n-s) \le k \le \min(r, n) \\ 0, & \text{sonst} \end{cases} .$$

Diese Verteilung heißt *hypergeometrische Verteilung mit Parametern $r+s$, r und n*, kurz *$H(r+s, r, n)$-Verteilung.*

Ersetzt man die $r+s$ Kugeln der Urne durch N Produktionsgüter, unter denen M defekt und $N-M$ in Ordnung sind, so beschreibt die $H(N, M, n)$-Verteilung die in der Qualitätskontrolle wichtige Situation, daß bei einer Überprüfung von n zufällig herausgegriffenen

Stücken eine bestimmte Anzahl k defekt ist.

4.8. Beispiel. ***(Multinomialverteilung)*** Als letztes wollen wir ein wichtiges Beispiel eines mehrdimensionalen diskreten Zufallsvektors vorstellen. In 3.6 hatten wir einige Teilchen-Fächer-Modelle im Rahmen der statistischen Physik studiert. Greifen wir beispielhaft noch einmal die Maxwell-Boltzmann-Statistik heraus. Dann können sich beliebig viele der k Teilchen in jeder der n Zellen (Fächer) aufhalten, und es ist jeder Mikrozustand $(\omega_1, ..., \omega_k) \in \Omega = \{1, ..., n\}^k$ gleichwahrscheinlich. Unter der Annahme, daß m Energieniveaus $\mathfrak{E}_1, ..., \mathfrak{E}_m$ existieren, wobei n_j Zellen das Niveau $\mathfrak{E}_j$ besitzen ($n_1 + ... + n_m = n$), hatten wir eingesehen, daß für den Zufallsvektor $(N_1, ..., N_m)$ der Besetzungszahlen zu den Energieniveaus gilt:

$$\begin{aligned} P^{(N_1,...,N_m)}(\{(k_1,...,k_m)\}) &= P((N_1,...,N_m)^{-1}(\{(k_1,...,k_m)\})) \\ &= P(\omega \in \Omega : N_1(\omega) = k_1, ..., N_m(\omega) = k_m) \\ &= \frac{k!}{k_1! \cdot ... \cdot k_m!} \left(\frac{n_1}{n}\right)^{k_1} \cdot ... \cdot \left(\frac{n_m}{n}\right)^{k_m} \end{aligned} \tag{4.7}$$

für alle $(k_1, ..., k_m) \in \mathbb{N}_0^m$ mit $k_1 + ... + k_m = k$. Es handelt sich hierbei um eine *Multinomialverteilung mit Parametern k und $\frac{n_1}{n}, ..., \frac{n_m}{n}$*, kurz *$M(k, \frac{n_1}{n}, ..., \frac{n_m}{n})$-Verteilung*. Sie stellt eine Verallgemeinerung der Binomialverteilung dar, die im Fall $m = 2$ auftritt, wenn man statt (N_1, N_2) nur N_1 (oder N_2) betrachtet, was wegen $N_2 = k - N_1$ ausreicht.

4.9. Bemerkungen. (a) Man kann *jede* diskrete W-Verteilung als induzierte Verteilung einer diskreten Zufallsvariablen auffassen. Bezeichnet nämlich $X = \mathrm{id}_\Omega : \Omega \to \Omega$ die Identität auf Ω, so gilt offensichtlich

$$P^X(A) = P(X^{-1}(A)) = P(A)$$

für alle $A \subset \Omega$ gilt, d.h. $P^X = P$.

(b) Aus $P^{X_1} = P^{X_2}$ folgt i.a. nicht $X_1 = X_2$, wie das folgende einfache Beispiel lehrt: Sei P die Laplace-Verteilung über $\Omega = \{0, 1\}$ und $X_1 = 1 - X_2 = \mathrm{id}_\Omega$. Dann haben X_1 und X_2 offenkundig dieselben Verteilungen, ohne gleich zu sein.

4.10. Beispiel. [☞ 4.6] In Beispiel 4.6 hatten wir die Verteilung der Augensumme S beim zweimaligen Werfen eines fairen Würfels bestimmt. Sei weiter

$$D(\omega_1, \omega_2) \stackrel{\text{def}}{=} \omega_1 - \omega_2$$

die Augendifferenz und $X \stackrel{\text{def}}{=} (S, D)$. Dann liefert X eine Bijektion zwischen Ω und

$$X(\Omega) = \left\{ \begin{array}{c} (2,0), (3,-1), (4,-2), (5,-3), (6,-4), (7,-5) \\ (3,1), (4,0), (5,-1), (6,-2), (7,-3), (8,-4) \\ \vdots \\ (7,5), (8,4), (9,3), (10,2), (11,1), (12,0) \end{array} \right\},$$

wobei $X^{-1}(s,d) = (\frac{s+d}{2}, \frac{s-d}{2})$ (hier ist natürlich die Umkehrabbildung gemeint). Folglich besitzt X eine Laplace-Verteilung über $X(\Omega)$.

Man kann nun die Verteilungen von S und D auch sehr leicht aus der von X bestimmen, und zwar vermöge

$$P^S(\{s\}) \;=\; P^X(\{(s,k) : (s,k) \in X(\Omega)\}) \;=\; \sum_{(s,k)\in X(\Omega)} P^X(\{(s,k)\})$$

sowie

$$P^D(\{d\}) \;=\; P^X(\{(k,d) : (k,d) \in X(\Omega)\}) \;=\; \sum_{(k,d)\in X(\Omega)} P^X(\{(k,d)\}).$$

Man summiert demnach unter P^X alle Wahrscheinlichkeiten dafür auf, als erste Komponente s bzw. als zweite Komponente d zu erhalten. Das Ergebnis für S kennen wir bereits aus 4.6, das für D sieht genauso aus, wenn man die Werte $2, ..., 12$ jeweils durch $-5, ..., 5$ ersetzt.

Die vorherige Überlegung gibt Anlaß zu einer Definition:

4.11. Definition. Sei $X = (X_1, ..., X_n)$ ein diskreter Zufallsvektor mit der Verteilung P^X und $\{i_1, ..., i_k\} \subset \{1, ..., n\}$ mit $1 \le i_1 < ... < i_k \le n$. Dann heißt $P^{(X_{i_1},...,X_{i_k})}$ eine *k-dimensionale Randverteilung (Marginalverteilung) von* P^X oder auch X.

4.12. Satz. *Gegeben eine k-dimensionale Randverteilung $P^{(X_{i_1},...,X_{i_k})}$ eines diskreten Zufallsvektors $X = (X_1, ..., X_n)$, gilt für alle $A \subset \mathbb{R}^k$*

$$P^{(X_{i_1},...,X_{i_k})}(A) \;=\; P^X(A \boxtimes_I \mathbb{R}^{n-k}), \tag{4.8}$$

wobei $I \stackrel{\text{def}}{=} \{i_1, ..., i_k\}$ und $A \boxtimes_I \mathbb{R}^{n-k} \stackrel{\text{def}}{=} \{(x_1, ..., x_n) \in \mathbb{R}^n : (x_{i_1}, ..., x_{i_k}) \in A\}$.

Beweis: Es gilt für alle $A \subset \mathbb{R}^k$

$$\begin{aligned}
P^{(X_{i_1},...,X_{i_k})}(A) &= P((X_{i_1}, ..., X_{i_k})^{-1}(A)) \\
&= P(\{\omega \in \Omega : (X_{i_1}(\omega), ..., X_{i_k}(\omega)) \in A) \\
&= P(\{\omega \in \Omega : (X_{i_1}(\omega), ..., X_{i_k}(\omega)) \in A, X_j(\omega) \in \mathbb{R} \text{ für } j \notin \{i_1, ..., i_k\}) \\
&= P(\{\omega \in \Omega : (X_1(\omega), ..., X_n(\omega)) \in A \boxtimes_I \mathbb{R}^{n-k}) \\
&= P((X_1, ..., X_n)^{-1}(A \boxtimes_I \mathbb{R}^{n-k}) \\
&= P^X(A \boxtimes_I \mathbb{R}^{n-k}).
\end{aligned}$$
◊

Der Satz zeigt insbesondere, daß die Verteilung von X alle eindimensionalen Randverteilungen P^{X_i} eindeutig festlegt. Wie das folgende einfache Beispiel belegt, ist die Umkehrung jedoch falsch:

4.13. Beispiel. Sei P die Laplace-Verteilung über $\Omega = \{0,1\}^2$ und $(X_1, X_2) = \mathrm{id}_\Omega$. Dann gilt

$$P^{(X_1,X_2)}(\{(0,0)\}) \;=\; P^{(X_1,X_2)}(\{(0,1)\}) \;=\; P^{(X_1,X_2)}(\{(1,0)\}) \;=\; P^{(X_1,X_2)}(\{(1,1)\}) \;=\; \frac{1}{4}$$

und folglich für $i = 1, 2$

$$P^{X_i}(\{0\}) \;=\; P^{X_i}(\{1\}) \;=\; \frac{1}{2}.$$

X_1 und X_2 sind also jeweils Laplace-verteilt über $\Omega = \{0,1\}$. Setzen wir nun $Y_1 = Y_2 \overset{\text{def}}{=} X_1$, so gilt zwar $P^{X_1} = P^{X_2} = P^{Y_1} = P^{Y_2}$, aber aus

$$P^{(Y_1,Y_2)}(\{(0,0)\}) \;=\; P^{X_1}(\{0\}) \;=\; \frac{1}{2} \quad \text{und} \quad P^{(Y_1,Y_2)}(\{(1,1)\}) \;=\; P^{X_1}(\{1\}) \;=\; \frac{1}{2}$$

folgt sofort

$$P^{(X_1,X_2)} \;\neq\; P^{(Y_1,Y_2)}.$$

Die folgende Überlegung zeigt, daß es sogar überabzählbar viele verschiedene diskrete W-Verteilungen auf $\{0,1\}^2$ gibt, deren eindimensionalen Randverteilungen der Laplace-Verteilung auf $\{0,1\}$ entsprechen. Wir identifizieren dazu eine W-Verteilung auf $\{0,1\}^2$ mit der 2×2-Matrix, deren (i,j)-te Komponente die Wahrscheinlichkeit für das Ergebnis (i,j) angibt. Für jedes $\theta \in [0,1]$ sei dann $P_\theta = P_\theta^{(X_1,X_2)}$ die durch die Matrix

$$\begin{pmatrix} \frac{\theta}{2} & \frac{1-\theta}{2} \\ \frac{1-\theta}{2} & \frac{\theta}{2} \end{pmatrix}$$

gegebene W-Verteilung auf $\{0,1\}^2$. Da die Zeilen- und Spaltensummen der Matrix gerade die Wahrscheinlichkeiten für die Ergebnisse 0 und 1 unter den beiden eindimensionalen Randverteilungen von $P_\theta^{(X_1,X_2)}$ angeben und da diese hier stets $\frac{1}{2}$ ergeben, folgt die behauptete Aussage, denn die $P_\theta^{(X_1,X_2)}$ sind offenkundig alle verschieden.

Am Ende dieses Abschnitt führen wir im Rahmen einer Tabelle noch folgende in der Stochastik üblichen Bezeichnungsweisen ein:

A	Alternative Schreibweise für $X^{-1}(A)$
beliebig	$\{X \in A\}$
B^c	$\{X \notin B\}$
$\{x\}, \{x\}^c$	$\{X = x\}, \{X \neq x\}$
$[a,b], (a,b)$	$\{a \leq X \leq b\}, \{a < X < b\}$
$(a,b], [a,b)$	$\{a < X \leq b\}, \{a \leq X < b\}$
$(-\infty, x], (-\infty, x)$	$\{X \leq x\}, \{X < x\}$
$[x, \infty), (x, \infty)$	$\{X \geq x\}, \{X > x\}$

Wir benutzen ebenfalls die Schreibweisen

$$\{X_1 \in A_1, ..., X_n \in A_n\} \quad \text{oder} \quad \{X_i \in A_i, 1 \le i \le n\}$$

anstelle von

$$(X_1, ..., X_n)^{-1}(A_1 \times ... \times A_n)$$

und verzichten bei Wahrscheinlichkeiten auf die Verwendung von Mengenklammern, schreiben also

$$P(X \in A),\ P(X = x),\ \text{etc.}$$

für

$$P(\{X \in A\}),\ P(\{X = x\}),\ \text{etc.}$$

Dies entspricht gängiger Praxis in der Stochastik-Literatur.

5. Die Verteilungsfunktion

Gegeben eine diskrete W-Verteilung P über einer Menge Ω sowie eine Zufallsgröße $X : \Omega \to \mathbb{R}$, werden wir in diesem Abschnitt zeigen, daß die Verteilung P^X, bekanntlich definiert auf der Potenzmenge von $\mathbb{R}$, auch durch eine monoton wachsende Funktion $F^X : \mathbb{R} \to [0,1]$ beschrieben werden kann, die sich durch Auswertung von P^X auf den Intervallen $(-\infty, x]$, $x \in \mathbb{R}$, ergibt.

5.1. Definition. Sei X eine diskrete Zufallsgröße mit Zähldichte p^X. Die durch

$$F^X(x) \overset{\text{def}}{=} P^X((-\infty, x]) = P(X \le x) = \sum_{\mathcal{X}_0 \ni y \le x} p^X(y) \tag{5.1}$$

definierte Funktion $F^X : \mathbb{R} \to [0,1]$ heißt *Verteilungsfunktion von* X oder P^X.

Die wesentlichen Eigenschaften einer Verteilungsfunktion fassen wir in einem Lemma zusammen:

5.2. Lemma. *Die Verteilungsfunktion F^X einer diskreten Zufallsgröße X besitzt folgende Eigenschaften:*

(a) F^X ist monoton wachsend, d.h. $x_1 \le x_2$ impliziert $F^X(x_1) \le F^X(x_2)$.

(b) F^X ist rechtsseitig stetig, d.h. $\lim_{y \downarrow x} F^X(y) = F^X(x)$ für alle $x \in \mathbb{R}$.

(c) $\lim_{x \to -\infty} F^X(x) = 0$ und $\lim_{x \to \infty} F^X(x) = 1$.

(d) F^X besitzt linksseitige Limiten, d.h. $F^X(x-) \overset{\text{def}}{=} \lim_{y \uparrow x} F^X(y)$ existiert für alle $x \in \mathbb{R}$. Genauer gilt $F^X(x-) = P(X < x)$.

(e) F^X ist eine Sprungfunktion, deren Sprungstellen den Elementen des Trägers $\mathcal{X}_0$ von P^X entsprechen, und sie ist außerdem konstant auf jedem Intervall $I \ne \emptyset$ mit $I \cap \mathcal{X}_0 = \emptyset$.

Beweis: (a) ist aufgrund der Definition von F^X klar. Zum Nachweis von (b) sei $(x_n)_{n\geq 1}$ eine monoton fallende Folge mit Limes x. Dann folgt aus der Stetigkeit von oben von P^X (Satz 2.3(c))

$$\lim_{n\to\infty} F^X(x_n) \;=\; \lim_{n\to\infty} P^X((-\infty,x_n]) \;=\; P^X\left(\bigcap_{n\geq 1}(-\infty,x_n]\right) \;=\; P^X((-\infty,x]) \;=\; F^X(x).$$

Im Fall $x_n \downarrow -\infty$ erhalten wir analog

$$\lim_{n\to\infty} F^X(x_n) \;=\; \lim_{n\to\infty} P^X((-\infty,x_n]) \;=\; P^X\left(\bigcap_{n\geq 1}(-\infty,x_n]\right) \;=\; P^X(\emptyset) \;=\; 0$$

sowie im Fall $x_n \uparrow \infty$ vermöge der Stetigkeit von unten von P^X (Satz 2.3(b))

$$\lim_{n\to\infty} F^X(x_n) \;=\; \lim_{n\to\infty} P^X((-\infty,x_n]) \;=\; P^X\left(\bigcup_{n\geq 1}(-\infty,x_n]\right) \;=\; P^X(\mathbb{R}) \;=\; 1.$$

Damit ist auch (c) gezeigt. Die Existenz von $F^X(x-)$ ist klar aufgrund der Monotonie von F^X. Ferner gilt im Fall $x_n \uparrow x$ unter erneuter Verwendung der Stetigkeit von unten von P^X

$$\lim_{n\to\infty} F^X(x_n) \;=\; \lim_{n\to\infty} P^X((-\infty,x_n]) \;=\; P^X\left(\bigcup_{n\geq 1}(-\infty,x_n]\right) \;=\; P^X((-\infty,x)) \;=\; P(X<x),$$

was den Beweis von (d) komplettiert. Der Leser beachte für spätere Zwecke, daß wir bisher nirgendwo verwendet haben, daß die Zufallsgröße diskret ist.

Kommen wir schließlich zum Nachweis von (e). Wir zeigen zuerst, daß F^X genau dann unstetig in x ist, falls $x \in \mathcal{X}_0$ gilt. Betrachte Folgen $(x_n)_{n\geq 1}$ und $(x_n')_{n\geq 1}$ mit $x_n \uparrow x$ bzw. $x_n' \downarrow x$, falls $n \to \infty$. Dann gilt $(x_n, x_n'] \downarrow \{x\}$, und wir folgern die Behauptung aus

$$\begin{aligned} p^X(x) \;&=\; \lim_{n\to\infty} P^X((x_n,x_n']) \\ &=\; \lim_{n\to\infty}\Big(F^X(x_n') - F^X(x_n)\Big) \;=\; F^X(x) - F^X(x-), \end{aligned} \tag{5.2}$$

wobei erneut die Stetigkeit von P^X verwendet wurde. Bezeichnet schließlich I ein nichtleeeres Intervall mit $I \cap \mathcal{X}_0 = \emptyset$, das o.B.d.A. nicht nur aus einem Punkt besteht, so folgt für alle $x,y \in I$ mit $x<y$

$$F^X(y) - F^X(x) \;=\; P(x < X \leq y) \;\leq\; P(X \in I) \;=\; \sum_{z \in I\cap\mathcal{X}_0} p^X(z) \;=\; 0,$$

d.h. $F^X(x) = F^X(y)$. ♢

5.3. Korollar. *P^X ist durch F^X eindeutig bestimmt.*

Beweis: Da P^X durch seine Zähldichte p^X festgelegt wird, ergibt sich die Behauptung unmittelbar aus (5.2) ♢

5.4. Bemerkung. Sofern der Träger $\mathcal{X}_0$ von P^X keinen Häufungspunkt besitzt, lassen sich seine Elemente offenbar aufsteigend in der Form $(\ldots <) x_{-m} < \ldots < x_0 \leq 0 < x_1 < x_2 < \ldots < x_n (< \ldots)$ anordnen, und wir stellen fest, daß F^X eine Treppenfunktion mit Sprungstellen x_i der Höhe $p^X(x_i)$ ist. Die Pünktchen in Klammern sollen andeuten, daß natürlich sowohl unendlich viele negative als auch unendlich viele positive Trägerpunkte vorliegen können.

5.6. Beispiele. Die folgenden Graphiken zeigen die Zähldichten und Verteilungsfunktionen einiger bereits bekannter Verteilungen:

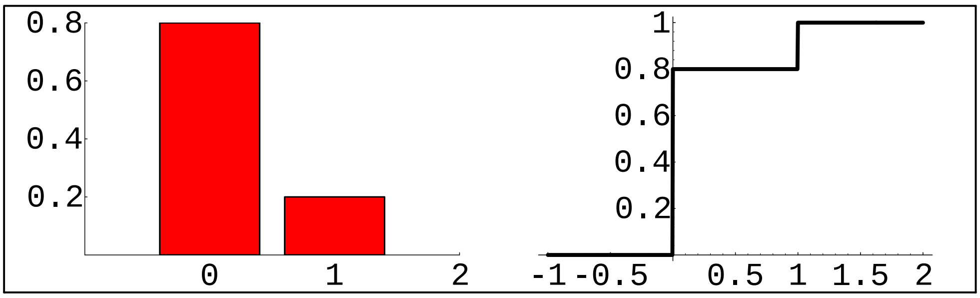

BILD 5.1. Die Bernoulli-Verteilung mit Parameter 0.2.

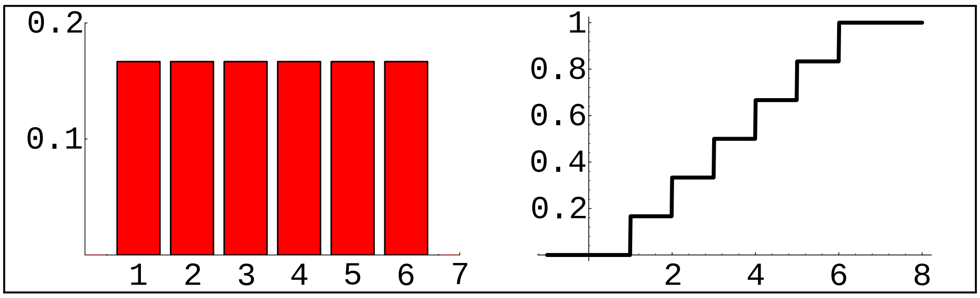

BILD 5.2. Die Laplace-Verteilung auf $\{1, 2, 3, 4, 5, 6\}$.

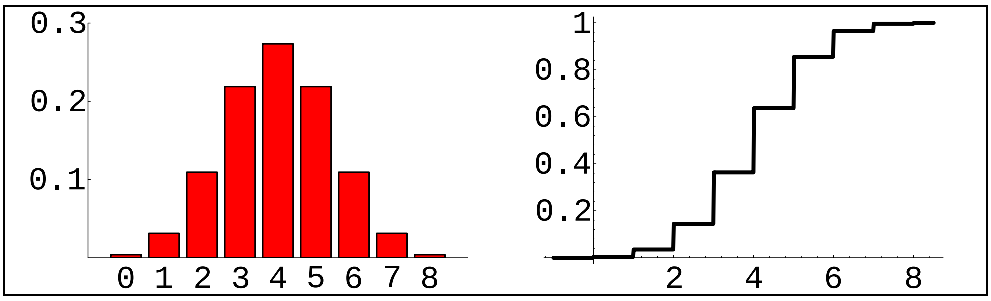

BILD 5.3. Die Binomialverteilung mit Parametern 8 und 0.5.

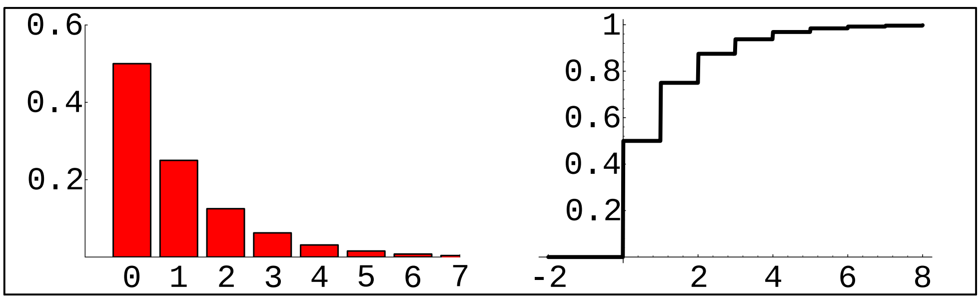

BILD 5.4. Die geometrische Verteilung mit Parameter 0.5.

6. Der Erwartungswert

Während Zähldichte und Verteilungsfunktion die Verteilung einer Zufallsgröße vollständig charakterisieren, trifft dies für andere, nicht minder wichtige Kenngrößen, zu denen wir als nächstes kommen, nicht mehr zu. Betrachten wir zur Motivation wieder ein einfaches Beispiel:

6.1. Beispiel. Stellen Sie sich vor, man böte Ihnen folgendes Spiel an: Sie zahlen vor Spielbeginn 1 DM Spieleinsatz und erhalten $g+1$ DM zurück, falls der anschließende Wurf von zwei fairen Würfeln einen Pasch liefert. Wie hoch sollte g sein, damit Sie bereit wären, sich auf dieses Spiel einzulassen? Da es 36 gleichwahrscheinliche Spielausgänge gibt, von denen 6 für Sie günstig sind, werden Sie sich vermutlich am *durchschnittlichen Gewinn* orientieren, die sich dadurch ergibt, daß der jedem Ergebnis (ω_1, ω_2) zugeordnete Gewinn

$$X(\omega_1,\omega_2) \stackrel{\text{def}}{=} \begin{cases} g, & \text{falls } \omega_1 = \omega_2 \\ -1, & \text{sonst} \end{cases}$$

mit seiner Eintrittswahrscheinlichkeit $\frac{1}{36}$ multipliziert wird. Mit anderen Worten, es wird das *gewichtete Mittel*

$$g\,P(X=g) \;+\; (-1)P(X=-1) \;=\; g\cdot\frac{6}{36} \;+\; (-1)\cdot\frac{30}{36} \;=\; \frac{g-5}{6}$$

gebildet. Falls $g \geq 6$, ist das Spiel also für sie vorteilhaft.

Das Beispiel gibt Anlaß zu folgender Definition:

6.2. Definition. Sei X eine diskrete Zufallsgröße mit der Zähldichte p^X, für die

(6.1) $$\sum_{x\in\mathbb{R}} |x| p^X(x) \;<\; \infty$$

gilt. Dann ist auch

(6.2) $$E(X) \stackrel{\text{def}}{=} \sum_{x\in\mathbb{R}} x p^X(x)$$

endlich und heißt *Erwartungswert* oder auch *Mittelwert von* X (unter P).

6.3. Bemerkungen. (a) Statt $E(X)$ schreibt man oft kürzer einfach EX. Man spricht ferner manchmal auch vom Erwartungswert von P^X, weil EX offenbar von der Zufallsgröße X nur über deren Verteilung P^X abhängt.

(b) Die in (6.1) und (6.2) auftretenden Summen machen Sinn, weil $p^X(x)$ nur für abzählbare viele $x \in \mathbb{R}$ von 0 verschieden ist. Bezeichnet $\mathcal{X}_0$ wieder den Träger von P^X, erhalten wir demnach

(6.3) $$EX \;=\; \sum_{x\in\mathcal{X}_0} x p^X(x) \;=\; \sum_{x\in\mathcal{X}_0} x P(X=x).$$

(c) *Nicht jede Zufallsgröße besitzt einen Erwartungswert*, wie das folgende Beispiel lehrt: Sei $\mathcal{X}_0 = \mathbb{Z}\backslash\{0\}$ und $p^X(n) = p^X(-n) = \frac{1}{2n(n+1)}$ für $n \in \mathbb{N}$. Dann ist Bedingung (6.1) wegen der Divergenz der harmonischen Reihe verletzt und EX existiert nicht.

(d) Eine "physikalische" Interpretation des Erwartungswerts ergibt sich aus der Interpretation der $p^X(x) = P^X(\{x\})$ als "Massenbelegungen": Wir nehmen dazu an, daß der Träger $\mathcal{X}_0$ von P^X endlich ist, also

$$\mathcal{X}_0 \;=\; \{x_1, ..., x_n\}$$

für ein $n \in \mathbb{N}$. In diesem Fall kann man $\mathcal{X}_0$ als Gesamtheit von Massenpunkten mit positiver Massenbelegung auffassen. Denkt man sich diese Punkte auf einer gewichtslosen Achse, der reellen Zahlengeraden, angeordnet, so erhält man den *Schwerpunkt (Massenmittelpunkt)* a des so entstehenden Körpers von der Masse 1 aus der Gleichgewichtsbedingung

$$\sum_{k=1}^{n} (x_k - a)p^X(x_k) \;=\; 0.$$

Es folgt

$$\sum_{k=1}^{n} x_k p^X(x_k) \;=\; a \sum_{k=1}^{n} p^X(x_k) \;=\; a,$$

und folglich $a = EX$.

6.4. Beispiele. (a) ***(Laplace-Verteilung)*** Ist X Laplace-verteilt auf $\{1, ..., n\}$ für ein $n \in \mathbb{N}$, so gilt

$$EX \;=\; \frac{1}{n}\sum_{k=1}^{n} k \;=\; \frac{n(n+1)}{2n} \;=\; \frac{n+1}{2}. \tag{6.4}$$

Dies liefert insbesondere $EX = 3.5$ im Fall $n = 6$ (einfacher Würfelwurf).

(b) ***(Binomialverteilung)*** Ist X $B(n, \theta)$-verteilt für ein $n \in \mathbb{N}$ und $\theta \in (0, 1)$, d.h. $P(X = k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$ für $k \in \mathcal{X}_0 = \{0, ..., n\}$ (☞ 1.8), folgt

$$\begin{aligned} EX \;&=\; \sum_{k=0}^{n} k\binom{n}{k}\theta^k(1-\theta)^{n-k} \\ &=\; \sum_{k=1}^{n} k\,\frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k} \\ &=\; n\theta\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}\theta^{k-1}(1-\theta)^{n-k} \\ &=\; n\theta\underbrace{\sum_{k=0}^{n-1}\binom{n-1}{k}\theta^k(1-\theta)^{n-1-k}}_{=1} \;=\; n\theta. \end{aligned} \tag{6.5}$$

(c) ***(Geometrische Verteilung)*** Ist X geometrisch verteilt mit Parameter $\theta \in (0,1)$, d.h. $P(X=k) = \theta(1-\theta)^k$ für $k \in \mathcal{X}_0 = \mathbb{N}_0$, ergibt sich

$$
\begin{aligned}
EX &= \theta \sum_{k\geq 0} k(1-\theta)^k \\
&= \theta(1-\theta) \sum_{k\geq 1} k(1-\theta)^{k-1} \\
&= \theta(1-\theta) \left(\sum_{k\geq 0} x^k \right)'_{\big|x=1-\theta} \\
&= \theta(1-\theta) \left(\frac{1}{(1-x)^2} \right)_{\big|x=1-\theta} \;=\; \frac{1-\theta}{\theta}.
\end{aligned} \tag{6.6}
$$

6.5. Bemerkung. Wählt man in 6.4(b) $\theta = \frac{n+1}{2n}$, folgt $EX = \frac{n+1}{2}$, d.h. derselbe Erwartungswert wie für eine Laplace-verteilte Zufallsgröße auf $\{1,...,n\}$. *Weder Zufallsgrößen noch ihre Verteilung sind demnach durch ihren Erwartungswert eindeutig bestimmt.*

Die folgende Transformationsformel zeigt, daß sich der Erwartungswert auch unter Benutzung der auf dem Definitionsbereich von X zugrundeliegenden W-Verteilung P berechnen läßt, sofern P als diskret oder regulär stetig vorausgesetzt wird.

6.6. Satz und Definition. *Sei P eine diskrete W-Verteilung über Ω mit Zähldichte p und Träger Ω_0 sowie $X : \Omega \to \mathbb{R}$ eine Zufallsgröße. Genau dann existiert der Erwartungswert von X, wenn*

$$
\sum_{\omega\in\Omega} |X(\omega)| p(\omega) \;=\; \sum_{\omega\in\Omega_0} |X(\omega)| p(\omega) \;<\; \infty \tag{6.7}
$$

*In diesem Fall bezeichnet man X als **(P-)integrierbar**, und es gilt weiter*

$$
EX \;=\; \sum_{\omega\in\Omega} X(\omega) P(\{\omega\}) \;=\; \sum_{\omega\in\Omega_0} X(\omega) P(\{\omega\}) \tag{6.8}
$$

6.7. Bemerkungen. (a) Die Bezeichnung "P-integrierbar" stammt aus der Integrationstheorie. Dort wird nämlich der Erwartungswert von X bezüglich einer beliebigen, d.h. nicht notwendig diskreten W-Verteilung P über Ω durch $\int X\,dP$, das Integral von X bezüglich P definiert, sofern dieses Integral existiert. Selbstverständlich ist diese Definition mit den unsrigen konsistent.

(b) Da der Erwartungswert einer Zufallsgröße nur von deren Verteilung abhängt, können wir zum Beweis von Aussagen über den Erwartungswert einer diskreten Zufallsgröße $X : \Omega \to \mathbb{R}$ mit Verteilung $Q = P^X$ o.B.d.A. voraussetzen, daß Ω abzählbar ist, indem wir Ω als Träger von Q, $P = Q$ und X als Identität auf Ω wählen (☞ auch 4.9(a)).

Beweis von Satz 6.6: Existiert EX, folgt aus der absoluten Konvergenz der Reihe $\sum_{x\in\mathbb{R}} xP(X=x)$

$$\begin{aligned} EX &= \sum_{x\in\mathbb{R}} xP(X=x) \\ &= \sum_{x\in X(\Omega_0)} xP(X=x) \\ &= \sum_{x\in X(\Omega_0)} x \sum_{\omega\in\Omega_0 : X(\omega)=x} p(\omega) \\ &= \sum_{x\in X(\Omega_0)} \sum_{\omega\in\Omega_0 : X(\omega)=x} xp(\omega) \\ &= \sum_{\omega\in\Omega_0} X(\omega)p(\omega) \end{aligned}$$

und per analoger Rechnung mit $|x|$ anstelle von x

$$\sum_{x\in\mathbb{R}} |x|P(X=x) = \sum_{\omega\in\Omega_0} |X(\omega)|p(\omega).$$

Damit ist eine Richtung der Behauptung bewiesen. Für die andere braucht man aber unter Annahme von $\sum_{\omega\in\Omega_0} |X(\omega)|p(\omega) < \infty$ die obige Rechnung nur rückwärts zu lesen. $\diamondsuit$

Da Satz 6.6 insbesondere $E|X| = \sum_{\omega\in\Omega} |X(\omega)|p(\omega)$ bzw. $= \int_\Omega |X(\omega)|f(\omega)\,d\omega$ impliziert, läßt sich das dort angegebene notwendige und hinreichende Kriterium (6.7) für die Existenz von EX nun auch in der Form "$E|X| < \infty$" schreiben. Deshalb merke:

EX existiert genau dann, wenn dies für $E|X|$ der Fall ist.

oder mit anderen Worten:

Genau dann ist X integrierbar, wenn dies für $|X|$ der Fall ist.

Eine weitere wichtige Transformationsformel gilt für den Erwartungswert zusammengesetzter Zufallsgrößen der Form $G\circ X$.

6.8. Transformationssatz. *Sei $X:\Omega\to\mathcal{X}$ eine diskrete Zufallsvariable mit Zähldichte p^X. Sei ferner $G:\mathcal{X}\to\mathbb{R}$ eine Funktion, so daß $EG\circ X$ existiert. Dann gilt*

(6.9) $$EG\circ X = \sum_{x\in\mathcal{X}} G(x)p^X(x) = \sum_{x\in\mathcal{X}_0} G(x)p^X(x),$$

wobei $\mathcal{X}_0$ den Träger von P^X bezeichnet.

Beweis: Sei Ω_0 der Träger von P. Sofern $EG \circ X$ existiert, folgt nun vermöge Satz 6.6

$$\begin{aligned} EG \circ X &= \sum_{y \in \mathbb{R}} yP(G \circ X = y) \\ &= \sum_{\omega \in \Omega_0} G \circ X(\omega) P(\{\omega\}) \\ &= \sum_{x \in \mathcal{X}} \sum_{\omega \in \Omega_0 : X(\omega) = x} G(x) P(\{\omega\}) \\ &= \sum_{x \in \mathcal{X}} G(x) p^X(x). \end{aligned}$$ ◇

Zwei wichtige Eigenschaften, nämlich die *Linearität* und die *Monotonie des Erwartungswertes*, gibt der nächste Satz.

6.9. Satz. *Es seien X, Y zwei integrierbare diskrete Zufallsgrößen. Dann gilt:*

(a) Für alle $a, b \in \mathbb{R}$ ist $aX + bY$ integrierbar und $E(aX + bY) = aEX + bEY$.

(b) $X \geq Y$ impliziert $EX \geq EY$; insbesondere folgt $EX \geq 0$ aus $X \geq 0$.

Beweis: Sei wieder Ω_0 der Träger von P.

(a) Aus

$$\begin{aligned} &\sum_{\omega \in \Omega_0} |aX(\omega) + bY(\omega)| P(\{\omega\}) \\ &\qquad \leq \sum_{\omega \in \Omega_0} (|a||X(\omega)| + |b||Y(\omega)|) P(\{\omega\}) \\ &\qquad = |a| \sum_{\omega \in \Omega_0} |X(\omega)| P(\{\omega\}) + |b| \sum_{\omega \in \Omega_0} |Y(\omega)| P(\{\omega\}) < \infty \end{aligned}$$

für alle $a, b \in \mathbb{R}$ folgt gemäß Satz 6.6 die Existenz von $E(aX + bY)$, und es ergibt sich

$$\begin{aligned} &\sum_{\omega \in \Omega_0} (aX(\omega) + bY(\omega)) P(\{\omega\}) \\ &\qquad = a \sum_{\omega \in \Omega_0} X(\omega) P(\{\omega\}) + b \sum_{\omega \in \Omega_0} Y(\omega) P(\{\omega\}) \\ &\qquad = aEX + bEY. \end{aligned}$$

(b) Hier liefert (6.8) im Fall $X \geq Y$

$$EX = \sum_{\omega \in \Omega_0} X(\omega) P(\{\omega\}) \geq \sum_{\omega \in \Omega_0} Y(\omega) P(\{\omega\}) = EY,$$

also die Behauptung. ◇

Abschließend geben wir dem Leser einige hinreichende Kriterien für die Integrierbarkeit einer Zufallsgröße an die Hand:

6.10. Satz. *Hinreichend für die Integrierbarkeit einer diskreten Zufallsgröße X sind:*

(a) *X ist beschränkt.*

(b) *$|X| \leq Y$ für eine integrierbare Zufallsgröße $Y \geq 0$.*

(c) *$|X|^p$ ist integrierbar für ein $p \geq 1$.*

Beweis: Es sei wiederum o.B.d.A. Ω_0 der Träger von P.

(a) Aus $|X| \leq c$ folgt

$$\sum_{\omega\in\Omega_0} |X(\omega)|P(\{\omega\}) \ \leq \ c \sum_{\omega\in\Omega_0} P(\{\omega\}) \ = \ c \ < \ \infty$$

und somit die Existenz von $E|X|$ und EX.

(b) Analog liefert $|X| \leq Y$

$$\sum_{\omega\in\Omega_0} |X(\omega)|P(\{\omega\}) \ \leq \ \sum_{\omega\in\Omega_0} Y(\omega)P(\{\omega\}) \ = \ EY \ < \ \infty$$

und folglich wiederum die Existenz von $E|X|$ und EX.

(c) Hier genügt der Hinweis auf (b), denn $|X| \leq |X|^p + 1$ für alle $p \geq 1$. ♢

7. Varianz, Kovarianz und Korrelation

Wir zeigen als nächstes, daß der Erwartungswert noch eine andere Charakterisierung als die des Schwerpunkts einer Verteilung besitzt. Eine Zufallsgröße X heiße *quadratisch integrierbar*, falls X^2 integrierbar ist, also EX^2 existiert.

7.1. Lemma. *Sei X eine quadratisch integrierbare Zufallsgröße. Dann existiert auch $G(a) \stackrel{\text{def}}{=} E(X-a)^2$ für alle $a \in \mathbb{R}$, und die Funktion G besitzt in $a = EX$ ein Minimum, d.h.*

(7.1)
$$E(X - EX)^2 \ = \ \min_{a\in\mathbb{R}} E(X-a)^2.$$

Beweis: Da $(X-a)^2 \leq X^2 + 2|a||X| + a^2$ und mit EX^2 auch $E|X|$ existiert (☞ Satz 6.10(c)), erhalten wir vermöge Satz 6.10(b) die Existenz von $E(X-a)^2$ für alle $a \in \mathbb{R}$.

Zum Beweis von (7.1) bemerken wir zuerst, daß

$$E(X - EX) \ = \ EX - E(EX) \ = \ 0$$

aufgrund der Linearität des Erwartungswerts (Satz 6.9(a)) gilt. Damit folgt für alle $a \in \mathbb{R}$

$$\begin{aligned} E(X-a)^2 \ &= \ E(X - EX + EX - a)^2 \\ &= \ E(X-EX)^2 + 2E\big((X-EX)(EX-a)\big) + E(EX-a)^2 \\ &= \ E(X-EX)^2 + 2(EX-a)\underbrace{E(X-EX)}_{=0} + \underbrace{(EX-a)^2}_{\geq 0} \\ &\geq \ E(X-EX)^2, \end{aligned}$$

also die Behauptung. $\diamondsuit$

Der Erwartungswert einer Zufallsgröße X läßt sich also auch als derjenige Wert charakterisieren, der die mittlere quadratische Abweichung von X minimiert. Wir definieren:

7.2. Definition. Gegeben eine quadratisch integrierbare Zufallsgröße X, heißt

$$\mathrm{Var}(X) \stackrel{\text{def}}{=} E(X - EX)^2 = \sum_{x \in \mathbb{R}} (x - EX)^2 p^X(x). \tag{7.2}$$

die *Varianz von X* (unter P). Ihre Wurzel $\sqrt{\mathrm{Var}(X)}$ wird als *Standardabweichung von X* bezeichnet.

Statt $\mathrm{Var}(X)$ schreiben wir im folgenden einfach $\mathrm{Var}X$. Die folgende Darstellung der Varianz wird häufig anstelle von (7.2) zur expliziten Berechnung verwendet.

7.3. Lemma. *Gegeben eine quadratisch integrierbare Zufallsgröße X, gilt*

$$VarX = EX^2 - (EX)^2. \tag{7.3}$$

Beweis: Mittels Ausquadrieren und der zuvor gezeigten Linearität des Erwartungswertes (Satz 6.9(a)) erhalten wir sofort die Behauptung:

$$\begin{aligned} \mathrm{Var}X &= E\Big(X^2 - 2XEX + (EX)^2\Big) \\ &= EX^2 - 2(EX)(EX) + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$ $\diamondsuit$

7.4. Beispiele. (a) ***(Laplace-Verteilung)*** Ist X Laplace-verteilt auf $\{1, ..., n\}$, gilt unter Benutzung der Summationsformel $\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$

$$EX^2 = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{6}.$$

Damit folgt weiter (☞ (6.4))

$$\mathrm{Var}X = EX^2 - (EX)^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n-1)(n+1)}{12}. \tag{7.4}$$

(b) ***(Binomialverteilung)*** Für eine $B(n, \theta)$-verteilte Zufallsgröße ergibt sich unter Benutzung von $EX = n\theta$ (☞ (6.5))

$$\begin{aligned} EX^2 &= \sum_{k=1}^n k^2 \binom{n}{k} \theta^k (1-\theta)^{n-k} \\ &= \sum_{k=1}^n \big(k(k-1) + k\big) \binom{n}{k} \theta^k (1-\theta)^{n-k} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k}\theta^k(1-\theta)^{n-k} \; + \; n\theta \\
&= n(n-1)\theta^2 \sum_{k=2}^{n} \binom{n-2}{k-2}\theta^{k-2}(1-\theta)^{n-k} \; + \; n\theta \\
&= n(n-1)\theta^2 \underbrace{\sum_{k=0}^{n-2} \binom{n-2}{k}\theta^{k}(1-\theta)^{n-2-k}}_{=1} + n\theta \\
&= (n\theta)^2 + n\theta(1-\theta)
\end{aligned}
$$

und damit

$$\text{(7.5)} \qquad \mathrm{Var}X \;=\; (n\theta)^2 + n\theta(1-\theta) - (n\theta)^2 \;=\; n\theta(1-\theta).$$

(c) ***(Geometrische Verteilung)*** Ist X geometrisch verteilt mit Parameter $\theta > 0$, gilt unter Benutzung von $\sum_{k\geq 2} k(k-1)x^{k-2} \frac{d^2}{dx^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3}$ und $EX = \frac{1-\theta}{\theta}$ (☞ (6.6))

$$
\begin{aligned}
EX^2 &= \sum_{k\geq 1} k^2\theta(1-\theta)^k \\
&= \sum_{k\geq 2} \big(k(k-1)+k\big)\theta(1-\theta)^k \\
&= \theta(1-\theta)^2 \sum_{k\geq 2} k(k-1)(1-\theta)^{k-2} \; + \; \sum_{k\geq 1} k\theta(1-\theta)^k \\
&= \theta(1-\theta)^2 \frac{2}{\theta^3} \; + \; \frac{1-\theta}{\theta} \\
&= 2\left(\frac{1-\theta}{\theta}\right)^2 \; + \; \frac{1-\theta}{\theta}.
\end{aligned}
$$

Für die Varianz liefert dies

$$\text{(7.6)} \qquad \mathrm{Var}X \;=\; \left(\frac{1-\theta}{\theta}\right)^2 \; + \; \frac{1-\theta}{\theta} \;=\; \frac{1-\theta}{\theta^2}.$$

Kommen wir als nächstes zu einer Reihe von Rechenregeln für die Varianz.

7.5. Satz. *Gegeben eine quadratisch integrierbare Zufallsgröße X, gilt:*

(a) Var$X = 0$ genau dann, wenn X P-f.s. konstant ist, d.h., wenn ein $c \in \mathbb{R}$ existiert mit $P(X = c) = 1$.

(b) Var$(aX + b) = a^2$Var$X$ für alle $a, b \in \mathbb{R}$.

Beweis: (a) Es bezeichne $\mathcal{X}_0$ wie üblich den Träger von P^X. Dann folgt aus

$$\mathrm{Var}X \;=\; \sum_{x\in\mathcal{X}_0} (x - EX)^2 P(X = x) \;=\; 0$$

wegen der Nichtnegativität der Summanden offenbar $(x - EX)^2 P(X = x) = 0$ für alle $x \in \mathcal{X}_0$ und auch die Umkehrung. Dies ist aber wiederum äquivalent zu $P(X = x) = 0$ für alle $x \neq EX$ und somit zu $P(X = EX) = 1$. X ist also mit Wahrscheinlichkeit 1 (P-f.s.) konstant.

(b) Für beliebige $a, b \in \mathbb{R}$ erhalten wir mittels Ausquadrieren und den Rechenregeln für Erwartungswerte

$$\begin{aligned} \mathrm{Var}(aX + b) &= E\big(aX + b - E(aX + b)\big)^2 = E\big(a(X - EX)\big)^2 \\ &= E\big(a^2(X - EX)^2\big) = a^2 E(X - EX)^2 = a^2 \mathrm{Var}X, \end{aligned}$$

also die behauptete Identität. ♢

Aussage (b) des vorherigen Satzes können wir auch in folgenden Merkregeln festhalten:

Die Varianz einer Zufallsgröße ist
— invariant unter Verschiebungen ($\mathrm{Var}(X + b) = \mathrm{Var}X$),
— invariant unter Spiegelung am Nullpunkt ($\mathrm{Var}(-X) = \mathrm{Var}X$),
— quadratisch unter Skalentransformationen ($\mathrm{Var}(aX) = a^2\mathrm{Var}X$).

Wählt man speziell $a = 2$, ergibt sich $\mathrm{Var}(X + X) = \mathrm{Var}(2X) = 4\mathrm{Var}X$. Im Gegensatz zum Erwartungswert ist die Varianz demnach i.a. nicht additiv. Es gilt vielmehr:

7.6. Lemma. *Gegeben zwei quadratisch integrierbare Zufallsgrößen X, Y, ist auch $X+Y$ quadratisch integrierbar, und es gilt*

$$\mathit{Var}(X + Y) = \mathit{Var}X + \mathit{Var}Y + 2\Big(E(XY) - EXEY\Big). \tag{7.7}$$

Beweis: Wir bemerken als erstes, daß $E|XY|$ wegen $|XY| \leq (X^2 + Y^2)/2$ existiert. Wegen $(X + Y)^2 = X^2 + Y^2 + 2XY$ liefert dies die Existenz von $E(X + Y)^2$ und damit die von $\mathrm{Var}(X + Y)$. Weiter erhalten wir nun

$$\begin{aligned} \mathrm{Var}(X + Y) &= E\big(X^2 + Y^2 + 2XY\big) - \big(EX + EY\big)^2 \\ &= EX^2 + EY^2 + 2E(XY) - (EX)^2 - (EY)^2 - 2EXEY \\ &= \mathrm{Var}X + \mathrm{Var}Y + 2\big(E(XY) - EXEY\big). \end{aligned}$$ ♢

Anstelle einer linearen Beziehung zwischen $\mathrm{Var}(X + Y)$ und $\mathrm{Var}X, \mathrm{Var}Y$ erhalten wir somit den zusätzlichen "Korrekturterm" $E(XY) - EXEY$, der einen eigenen Namen trägt:

7.7. Definition. Gegeben zwei quadratisch integrierbare Zufallsgrößen X, Y, heißt

$$\mathrm{Cov}(X, Y) \stackrel{\text{def}}{=} E(XY) - EXEY$$

die *Kovarianz von X und Y* (unter P).

Einfaches Nachrechnen zeigt, daß auch

$$(7.8) \qquad \mathrm{Cov}(X,Y) \;=\; E\big((X-EX)(Y-EY)\big)$$

gilt. Als direkte Verallgemeinerung von Lemma 7.6, die mittels einer Induktion gezeigt werden kann, notieren wir:

7.8. Satz. *Gegeben $n \geq 2$ quadratisch integrierbare Zufallsgrößen $X_1,...,X_n$, ist auch deren Summe $X_1+...+X_n$ quadratisch integrierbar, und es gilt*

$$(7.9) \qquad Var\left(\sum_{k=1}^{n} X_k\right) \;=\; \sum_{k=1}^{n} VarX_k \;+\; 2\sum_{1\leq i<j\leq n} Cov(X_i,X_j).$$

Wir wollen als nächstes zeigen, daß die Kovarianz als Maßzahl für die "lineare Abhängigkeit" zwischen zwei Zufallsgrößen interpretierbar ist. Seien dazu X,Y Zufallsgrößen mit existierenden positiven Varianzen. Wir hatten zuvor eingesehen, daß EX den mittleren quadratischen Abstand zu X unter allen Konstanten minimiert. Wir stellen nun die Frage, welche lineare Funktion $a+bY$, $a,b\in\mathbb{R}$ diesen Abstand zu X, d.h. $E(X-a-bY)^2$, minimiert. Wir geben die Lösung in einem Lemma:

7.9. Lemma. *X und Y seien quadratisch integrierbare Zufallsgrößen mit positiven Varianzen. Ferner sei*

$$\rho \stackrel{\text{def}}{=} \frac{Cov(X,Y)}{(VarX)^{1/2}(VarY)^{1/2}}.$$

(a) Für $a_0 = EX - \rho\frac{(VarX)^{1/2}}{(VarY)^{1/2}}EY$ und $b_0 = \rho\frac{(VarX)^{1/2}}{(VarY)^{1/2}}$ gilt

$$E(X-a_0-b_0Y)^2 \;=\; \min\{E(X-a-bY)^2 : a,b\in\mathbb{R}\}.$$

(b) Es gilt $\rho\in[-1,1]$, d.h.

$$(7.10) \qquad |Cov(X,Y)| \;\leq\; (VarX)^{1/2}(VarY)^{1/2},$$

genannt ***Cauchy-Schwarzsche Ungleichung****.*

Beweis: (a) Wir nehmen zunächst $EX = EY = 0$ und $\mathrm{Var}X = \mathrm{Var}Y = 1$ an. Zufallsgrößen mit Erwartungswert 0 und Varianz 1 heißen *standardisiert.* Dann ergibt sich für $a,b\in\mathbb{R}$

$$\begin{aligned} E(X-a-bY)^2 &= EX^2+a^2+b^2EY^2-2aEX-2bE(XY)+2abEY \\ &= 1+a^2+b^2-2bE(XY). \end{aligned}$$

Das Minimum wird demnach für $a_0 = 0$ und $b_0 = E(XY)$ angenommen.

Nicht standardisierte Zufallsgrößen mit positiver Varianz lassen sich durch eine lineare Transformation in standardisierte Zufallsgrößen überführen. Wir setzen zu diesem Zweck

$$X^* \stackrel{\text{def}}{=} \frac{X-\mu}{\sigma} \quad \text{und} \quad Y^* \stackrel{\text{def}}{=} \frac{Y-\nu}{\tau},$$

wobei $\mu = EX, \nu = EY, \sigma^2 = \mathrm{Var}X$ und $\tau^2 = \mathrm{Var}Y$. Wegen

$$\begin{aligned} E(X-a-bY)^2 &= E\Big(\sigma X^* + \mu - a - b(\tau Y^* + \nu)\Big)^2 \\ &= \sigma^2 E\left(X^* - \left(\frac{a-\mu+b\nu}{\sigma}\right) - \frac{b\tau}{\sigma}Y^*\right)^2 \end{aligned}$$

wird $E(X-a-bY)^2$ nach dem vorher Gezeigten genau dann minimal, wenn

$$\frac{a-\mu+b\nu}{\sigma} = 0 \quad \text{und} \quad \frac{b\tau}{\sigma} = EX^*Y^*.$$

Mit $\rho = \frac{Cov(X,Y)}{\sigma\tau} = EX^*Y^*$ folgt hieraus für die gesuchten Werte a_0, b_0

$$a_0 = \mu - \rho\nu\frac{\sigma}{\tau} \quad \text{und} \quad b_0 = \rho\frac{\sigma}{\tau},$$

also

$$\begin{aligned} 0 &\le \min\{E(X-a-bY)^2 : a,b \in \mathbb{R}\} = E(X-a_0-b_0Y)^2 \\ &= E\left(X - \mu - \rho\frac{\sigma}{\tau}(Y-\nu)\right)^2 \\ &= \mathrm{Var}X + \rho^2\frac{\sigma^2}{\tau^2}\mathrm{Var}Y - 2\rho\frac{\sigma}{\tau}Cov(X,Y) \\ &= \sigma^2(1-\rho^2). \end{aligned}$$

(b) Aus der letzten Ungleichung ersieht man, daß $\rho^2 \le 1$ und damit $\rho \in [-1,1]$ gilt. ♢

Die zuvor eingeführte Zahl ρ hat einen besonderen Namen:

7.10. Definition. Gegeben quadratisch integrierbare Zufallsgrößen X, Y, heißt

$$\rho(X,Y) \stackrel{\text{def}}{=} \frac{Cov(X,Y)}{(\mathrm{Var}X)^{1/2}(\mathrm{Var}Y)^{1/2}}$$

der *Korrelationskoeffizient* oder auch *die Korrelation von X und Y*.

Die letzte Ungleichung im Beweis von Lemma 7.9(a) zeigt, inwiefern man den Korrelationskoeffizienten – und damit auch die Kovarianz – als Maß für den *linearen "Zusammenhang" von X und Y* ansehen kann: Die mittlere quadratische Abweichung ist bei optimaler linearer Approximation von X durch $a_0 + b_0Y$ umso kleiner, je größer $|\rho|$ ist. Der Extremfall $|\rho| = 1$ liegt genau dann vor, wenn

$$E(X-a_0-b_0Y)^2 = \mathrm{Var}X(1-\rho^2) = 0,$$

d.h., wenn

$$P(\{\omega \in \Omega : X(\omega) = a_0 + b_0 Y(\omega)\}) \;=\; 1$$

gilt. Da außerdem die mittlere Abweichung bei optimaler linearer Approximation

$$\begin{aligned} E(X - a_0 - b_0 Y) \;&=\; EX - a_0 - b_0 EY \\ &=\; EX - EX + \rho\left(\frac{\mathrm{Var}X}{\mathrm{Var}Y}\right)^{1/2} EY - \rho\left(\frac{\mathrm{Var}X}{\mathrm{Var}Y}\right)^{1/2} EY \;=\; 0, \end{aligned}$$

erhält man vermöge Lemma 7.3 noch die Gleichheit

$$\mathrm{Var}(X - a_0 - b_0 Y) \;=\; E(X - a_0 - b_0 Y)^2 \;=\; (1 - \rho^2)\mathrm{Var}X.$$

7.11. Definition. Man bezeichnet zwei Zufallsgrößen X, Y als *positiv korreliert, unkorreliert* bzw. *negativ korreliert*, wenn ihre Kovarianz $\mathrm{Cov}(X, Y) >, =$ bzw. < 0 ist.

Die nachfolgenden Beispiele sollen belegen, daß Korrelationskoeffizient und Kovarianz tatsächlich nur über den linearen und nicht über den allgemeinen funktionalen Zusammenhang von zwei Zufallsgrößen Auskunft geben.

7.12. Beispiele. (a) Sei X eine auf der Menge $\{-2, -1, 1, 2\}$ Laplace-verteilte Zufallsgröße und $Y = X^2$. Da X beschränkt ist, existieren EX^2 und EY^2 (Satz 6.10(a)), und sowohl X als auch Y besitzen nach Satz 7.5(a) positive Varianz. Ferner gilt $EX = 0$, woraus weiter

$$\mathrm{Cov}(X, Y) \;=\; E(XY) \;=\; EX^3 \;=\; \frac{1}{4}(-8 - 1 + 1 + 8) \;=\; 0$$

und damit $\rho(X, Y) = 0$ folgt.

(b) Seien W_1 und W_2 die Augenzahl des ersten bzw. zweiten Wurfs bei zweimaligem Werfen eines fairen Würfels. Beide Zufallsgrößen besitzen also eine Laplace-Verteilung auf $\{1, 2, 3, 4, 5, 6\}$. Augensumme und Augendifferenz werden dann durch die Zufallsgrößen $X = W_1 + W_2$ bzw. $Y = W_1 - W_2$ gegeben. Da X, Y wiederum beschränkt sind, existieren EX^2 und EY^2, und es folgt $\mathrm{Var}X > 0$ sowie $\mathrm{Var}Y > 0$ aus Satz 7.5(a). Da W_1 und W_2 dieselbe Verteilung besitzen, erhalten wir ferner

$$EY \;=\; E(W_1 - W_2) \;=\; EW_1 - EW_2 \;=\; 0,$$

und dann

$$\mathrm{Cov}(X, Y) \;=\; E(XY) \;=\; E\Big((W_1 + W_2)(W_1 - W_2)\Big) \;=\; EW_1^2 - EW_2^2 \;=\; 0$$

sowie $\rho(X, Y) = 0$.

In beiden Beispielen sind X und Y also unkorreliert, obwohl zwischen ihnen ein funktionaler Zusammenhang besteht; im ersten ist dieser wegen $Y = X^2$ offensichtlich, im zweiten dagegen von schwächerer Form; hier gilt nur, daß X und Y entweder beide gerade oder ungerade sind.

Wir beschließen diesen Abschnitt mit einigen Eigenschaften der Kovarianz, die der Leser leicht selbst nachprüft:

7.13. Lemma. *Gegeben zwei quadratisch integrierbare Zufallsgrößen X, Y, gilt:*
(a) $Cov(X, X) = VarX$.
(b) $Cov(X, Y) = Cov(Y, X)$.
(c) $Cov(aX + b, cY + d) = acCov(X, Y)$ *für alle* $a, b, c, d \in \mathbb{R}$.

8. Ungleichungen für Erwartungswerte

Als erstes notieren wir eine elementare Ungleichung zwischen EX und $E|X|$.

8.1. Lemma. *Gegeben eine integrierbare Zufallsgröße X, gilt*

$$|EX| \;\le\; E|X|. \tag{8.1}$$

Beweis: Aus $\pm X \le |X|$ folgt $\pm EX \le E|X|$ und damit die Behauptung. $\diamondsuit$

In Lemma 7.9 hatten wir die Cauchy-Schwarzsche Ungleichung gezeigt, die im folgenden Satz noch etwas allgemeiner formuliert wird:

8.2. Satz. (Cauchy-Schwarzsche Ungleichung) *Gegeben zwei quadratisch integrierbare Zufallsgrößen X, Y, gilt*

$$|EXY| \;\le\; (EX^2)^{1/2}(EY^2)^{1/2}, \tag{8.2}$$

wobei Gleichheit genau dann vorliegt, wenn $P(X = aY) = 1$ oder $P(Y = aX) = 1$ für ein $a \in \mathbb{R}$.

Die in Lemma 7.9 gezeigte Ungleichung (7.10) ergibt sich, wenn man in (8.2) X, Y durch $X - EX, Y - EY$ ersetzt. Entsprechend folgt mit $|X|, |Y|$ anstelle von X, Y

$$E|XY| \;\le\; (EX^2)^{1/2}(EY^2)^{1/2}. \tag{8.3}$$

Beweis: Die (Un-)Gleichung ist im Fall $EX^2 = 0$ oder $EY^2 = 0$ trivial, weil dann $P(X = 0) = 1$ bzw. $P(Y = 0) = 1$ folgt und somit auch $E(XY) = 0$.

Seien nun EX^2, EY^2 beide positiv. Definieren wir $U = X/(EX^2)^{1/2}$, $V = Y/(EY^2)^{1/2}$ und benutzen die Ungleichung $\pm xy \le \frac{x^2+y^2}{2}$, so ergibt sich

$$\frac{\pm E(XY)}{(EX^2)^{1/2}(EY^2)^{1/2}} \;=\; \pm E(UV) \;\le\; \frac{EU^2 + EV^2}{2} \;=\; 1, \tag{8.4}$$

d.h. (8.2) nach Multiplikation beider Seiten mit $(EX^2)^{1/2}(EY^2)^{1/2}$.

Weiterhin $EX_1^2 > 0$ und $EY^2 > 0$ voraussetzend, liegt in (8.2) offenbar genau dann Gleichheit vor, wenn dies in (8.4) für $-E(UV)$ oder $E(UV)$ der Fall ist, wenn also

$$EU^2 + EV^2 \pm 2EUV \;=\; E(U \pm V)^2 \;=\; 0.$$

Dies ist aber äquivalent zu

$$1 \;=\; P(U = \mp V) \;=\; P\left(X = \mp\left(\frac{EX^2}{EY^2}\right)^{1/2} Y\right). \qquad \diamondsuit$$

Eine weitere äußerst nützliche Ungleichung betrifft Erwartungswerte der Form $E\varphi(X)$ für konvexe Funktionen φ. Eine Funktion $\varphi : I \to \mathbb{R}$, I ein beschränktes oder unbeschränktes Intervall, heißt bekanntlich konvex, wenn

$$\varphi(\lambda x + (1-\lambda)y) \;\leq\; \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

für alle $x, y \in I$ und $\lambda \in [0,1]$. Für unsere Zwecke genügt es im folgenden, nur offene Intervalle I zu betrachten. Bevor wir die angekündigte Ungleichung formulieren, fassen wir die wichtigsten Eigenschaften konvexer Funktionen in einem Lemma zusammen:

8.3. Lemma. *Sei I ein offenes Intervall und $\varphi : I \to \mathbb{R}$ eine konvexe Funktion. Dann gilt:*

(a) φ ist stetig.
(b) φ ist rechts- und linksseitig differenzierbar mit endlichen Ableitungen φ'_+ bzw. φ'_-.
(c) $\varphi(x) \geq \varphi(a) + \varphi'_+(a)(x-a)$ für alle $a, x \in I$.

Beweis: ☞ z.B. [1], Theorem (4.43) auf S. 199f. $\diamondsuit$

8.4. Satz. (Jensensche Ungleichung) *Es sei $I \subset \mathbb{R}$ ein offenes Intervall, $\varphi : I \to \mathbb{R}$ eine konvexe Funktion und X eine Zufallsgröße mit $P(X \in I) = 1$. Ferner existiere sowohl EX als auch $E\varphi(X)$. Dann gilt $EX \in I$ und*

$$\varphi(EX) \;\leq\; E\varphi(X). \tag{8.5}$$

Beweis: Sei $I = (a,b)$ für $a < b$, $a \in [-\infty, \infty)$ und $b \in (-\infty, \infty]$. Existiert EX, so impliziert $P(X > a) = 1$ auch $EX > a$ und analog $P(X < b) = 1$ auch $EX < b$. Damit gilt $EX \in I$ unter der Voraussetzung $P(X \in I) = P(a < X < b) = 1$.

Für den Nachweis der Ungleichung (8.5) benutzen wir $\varphi(x) \geq \varphi(y) + \varphi'_+(y)(x-y)$ für alle $x, y \in I$ (Lemma 8.3(c)). Es folgt

$$\begin{aligned} E\varphi(X) \;&\geq\; E\Big(\varphi(EX) + \varphi'_+(EX)(X - EX)\Big) \\ &=\; \varphi(EX) \;+\; \varphi'_+(EX)\,E(X - EX) \;=\; \varphi(EX), \end{aligned}$$

denn $E(X - EX) = 0$. ◊

Eine entsprechende Ungleichung mit umgekehrtem Ungleichheitszeichen gilt natürlich für konkave Funktionen φ, denn $-\varphi$ ist in diesem Fall konvex. Wir fassen zum Abschluß einige wichtige Spezialfälle in einem Korollar zusammen:

8.5. Korollar. *Gegeben eine integrierbare Zufallsgröße X, gilt jede der folgenden Ungleichungen, sofern der Erwartungswert auf der rechten Seite existiert:*

$$(E|X|)^r \leq E|X|^r \qquad (r \geq 1), \tag{8.6}$$

$$E|X|^r \leq (E|X|)^r \qquad (0 < r \leq 1), \tag{8.7}$$

$$\left(\frac{1}{E|X|}\right)^r \leq E\left(\frac{1}{|X|^r}\right) \qquad (r > 0), \tag{8.8}$$

$$e^{aEX} \leq Ee^{aX} \qquad (a \in \mathbb{R}). \tag{8.9}$$

Beachte, daß die für (8.8) notwendige Existenz von $E\left(\frac{1}{|X|^r}\right)$ insbesondere $P(X \neq 0) = 1$ und damit $E|X| > 0$ erzwingt.

9. Das schwache Gesetz der großen Zahlen

Wir kommen als nächstes zu zwei weiteren Ungleichungen, die die anschaulichen Interpretationen von Erwartungswert und Varianz als "Schwerpunkt" bzw. Maß für die "Streuung um den Schwerpunkt" untermauern. Mit ihrer Hilfe können wir anschließend auf sehr einfache Weise ein erstes wichtiges Konvergenzresultat beweisen, das unter dem Titel dieses Abschnitts bekannt ist.

Wir nennen eine Zufallsgröße X *r-fach integrierbar* ($r > 0$), wenn $|X|^r$ integrierbar ist, wenn also $E|X|^r$ existiert.

9.1. Satz. (Markov-Ungleichung) *Sei $X : \Omega \to \mathbb{R}$ eine r-fach integrierbare Zufallsgröße. Dann gilt für alle $c > 0$*

$$P(|X| \geq c) \leq \frac{E|X|^r}{c^r}. \tag{9.1}$$

Beweis: Wir betrachten weiterhin nur den diskreten Fall. Dann gilt vermöge des Transformationssatzes 6.8

$$\begin{aligned} E|X|^r &= \sum_{x \in \mathbb{R}} |x|^r P(X = x) \\ &= \sum_{x:|x| \geq c} |x|^r P(X = x) + \sum_{x:|x|<c} |x|^r P(X = x) \\ &\geq c^r \sum_{x:|x| \geq c} P(X = x) \\ &= c^r P(|X| \geq c). \end{aligned}$$

◊

Ersetzt man in (9.1) X durch die Zufallsgröße $X - EX$ und wählt $r = 2$, so ergibt sich speziell:

9.2. Korollar. (Tschebyschev-Ungleichung) *Gegeben eine quadratisch integrierbare Zufallsgröße mit Erwartungswert μ und Varianz σ^2, gilt für alle $c > 0$*

$$P(|X - \mu| \geq c) \;\leq\; \frac{\sigma^2}{c^2}. \tag{9.2}$$

Zur Motivation des angekündigten Konvergenzresultats betrachten wir zuerst ein Beispiel:

9.3. Beispiel. ***(Wiederholtes Werfen eines fairen Würfels)*** Das n-fache Werfen eines fairen Würfels wird bekanntlich durch ein Laplace-Experiment über dem Ergebnisraum $\Omega = \{1, 2, 3, 4, 5, 6\}^n$ beschrieben. Für die einzelnen Wurfergebnisse X_i,

$$X_i(\omega_1, ..., \omega_n) \;\stackrel{\text{def}}{=}\; \omega_i,$$

gilt dann

$$P(X_i = k) \;=\; \frac{1}{6}$$

für $k = 1, ..., 6$, d.h. jedes X_i ist Laplace-verteilt über $\{1, 2, 3, 4, 5, 6\}$ und hat folglich den Erwartungswert $EX_i = 3.5$ (☞ (6.4)) und die Varianz $\mathrm{Var}X_i = \frac{35}{12}$ (☞ (7.4)). Ferner gilt für $1 \leq i \neq j \leq n$

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) \;&=\; \frac{1}{36}\Big(1 \cdot (1 + ... + 6) + ... + 6 \cdot (1 + ... + 6)\Big) - (3.5)^2 \\ &=\; \frac{1}{36}\left(\frac{6 \cdot 7}{2}\right)^2 - (3.5)^2 \;=\; 0. \end{aligned}$$

Die X_i sind also außerdem paarweise unkorreliert.

Die Zufallsgröße $\overline{X}_n \stackrel{\text{def}}{=} n^{-1}(X_1 + ... + X_n)$ beschreibt offenkundig das *Durchschnittsergebnis* der n Würfe. Für großes n wird man – auch ohne etwas von Stochastik zu verstehen – erwarten, daß $\overline{X}_n \approx 3.5$. Dahinter verbirgt sich die Vorstellung, daß sich deutlich nach unten und oben abweichende Wurfergebnisse im Mittel kompensieren. Mit anderen Worten, man erwartet mit hoher Wahrscheinlichkeit ein Durchschnittsergebnis nahe bei 3.5, dem Erwartungswert bei einem Wurf. Allerdings bedarf es für die Bestätigung einer derartigen Aussage zunächst der Präzisierung dessen, was wir unter der Erwartung mit hoher Wahrscheinlichkeit verstehen.

9.4. Definition. $X, X_1, X_2, ...$ seien Zufallsgrößen. Die Folge $(X_n)_{n\geq 1}$ *konvergiert nach Wahrscheinlichkeit* oder *P-stochastisch* gegen X, kurz $X_n \stackrel{P}{\to} X$, wenn

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) \;=\; 0 \tag{9.3}$$

für jedes $\varepsilon > 0$ gilt.

9.3. (Fortsetzung) Wir wollen nun zeigen, daß $\overline{X}_n$ nach Wahrscheinlichkeit gegen die konstante Zufallsgröße $X = 3.5$ konvergiert. Zu diesem Zweck müssen wir $P(|\overline{X}_n - 3.5| \geq \varepsilon)$ abschätzen und bedienen uns dabei der Tschebyschev-Ungleichung. Da die X_i paarweise unkorreliert sind, gilt nach Satz 7.8

$$\mathrm{Var}\overline{X}_n \;=\; \frac{1}{n^2}\mathrm{Var}(X_1 + ... + X_n) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}X_i \;=\; \frac{\mathrm{Var}X_1}{n},$$

wobei für die letzte Gleichung benutzt wurde, daß die X_i alle dieselbe Verteilung besitzen. Mittels der Tschebyschev-Ungleichung erhalten wir nunmehr

$$P(|\overline{X}_n - 3.5| \geq \varepsilon) \;\leq\; \frac{\mathrm{Var}\overline{X}_n}{\varepsilon^2} \;=\; \frac{\mathrm{Var}X_1}{n\varepsilon^2} \;\to\; 0 \quad (n \to \infty)$$

für jedes $\varepsilon > 0$ und somit eine Bestätigung unserer Intuition.

Der nachfolgende Satz zeigt, daß das arithmetische Mittel von Zufallsgrößen schon unter weit schwächeren Voraussetzungen nach Wahrscheinlichkeit konvergiert.

9.5. Satz. (Schwaches Gesetz der großen Zahlen) *Seien $X_n, n \geq 1$, quadratisch integrierbare Zufallsgrößen mit Erwartungswerten μ_n. Sei ferner $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ und $\overline{\mu}_n = n^{-1}\sum_{i=1}^{n}\mu_i$ für $n \geq 1$. Es gelte die*

Markov-Bedingung:
$$\lim_{n\to\infty}\frac{1}{n^2}Var\left(\sum_{i=1}^{n} X_i\right) \;=\; 0.$$

Dann konvergiert $\overline{X}_n - \overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i)$ für $n \to \infty$ nach Wahrscheinlichkeit gegen 0, kurz $\overline{X}_n - \overline{\mu}_n \xrightarrow{P} 0$.

Beweis: Mittels der Tschebyschev-Ungleichung ergibt sich wegen $Var(\overline{X}_n - \overline{\mu}_n) = Var\overline{X}_n = n^{-2}\mathrm{Var}(\sum_{i=1}^{n} X_i)$

$$P(|\overline{X}_n - \overline{\mu}_n| \geq \varepsilon) \;\leq\; \frac{\mathrm{Var}(\sum_{i=1}^{n} X_i)}{\varepsilon^2 n^2}$$

für jedes $\varepsilon > 0$ und somit die Behauptung, wenn die Markov-Bedingung erfüllt ist. $\diamondsuit$

9.6. Beispiel. ***(Relative Häufigkeiten)*** Gegeben die Situation von Beispiel 9.3, definieren wir nun für festes $k \in \{1, 2, 3, 4, 5, 6\}$ und $i = 1, ..., n$

$$Y_i(\omega_1, ..., \omega_n) \;=\; \mathbf{1}_{\{k\}}(\omega_i) \;=\; \begin{cases} 1, & \text{falls } \omega_i = k \\ 0, & \text{sonst} \end{cases} \;=\; \mathbf{1}_{\{k\}} \circ X_i(\omega_1, ..., \omega_n).$$

Dann gilt

$$EY_i \;=\; P(X_i = k) \;=\; \frac{1}{6} \;\stackrel{\text{def}}{=}\; \mu$$

und wegen $Y_i = Y_i^2$

$$\mathrm{Var} Y_i \;=\; EY_i^2 - (EY_i)^2 \;=\; EY_i - (EY_i)^2 \;=\; \frac{5}{36} \stackrel{\text{def}}{=} \sigma^2$$

für jedes i. Ferner erhalten wir für $1 \leq i \neq j \leq n$

$$Cov(Y_i, Y_j) \;=\; E(Y_i Y_j) - EY_i EY_j \;=\; P(X_i = X_j = k) - \frac{1}{36} \;=\; 0,$$

d.h., die Y_i sind paarweise unkorreliert und

$$\frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) \;=\; \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var} Y_i \;=\; \frac{\sigma^2}{n}.$$

Die $Y_1, ..., Y_n$ genügen also bei Grenzübergang $n \to \infty$ der Markov-Bedingung. Nach dem schwachen Gesetz der großen Zahlen folgt demnach für $\overline{Y}_n$, die *relative Häufigkeit* der Augenzahl k in n Würfen, und jedes $\varepsilon > 0$

$$P(|\overline{Y}_n - \mu| \geq \varepsilon) \;=\; P(|\overline{Y}_n - \tfrac{1}{6}| \geq \varepsilon) \;\to\; 0,$$

falls $n \to \infty$. Wir haben damit ein intuitiv erwartetes Ergebnis erhalten: Bei gegen ∞ strebender Anzahl von Würfen konvergiert die Wahrscheinlichkeit dafür, daß die relative Häufigkeit einer beliebigen Zahl "nahezu" $\frac{1}{6}$ beträgt, gegen 1.

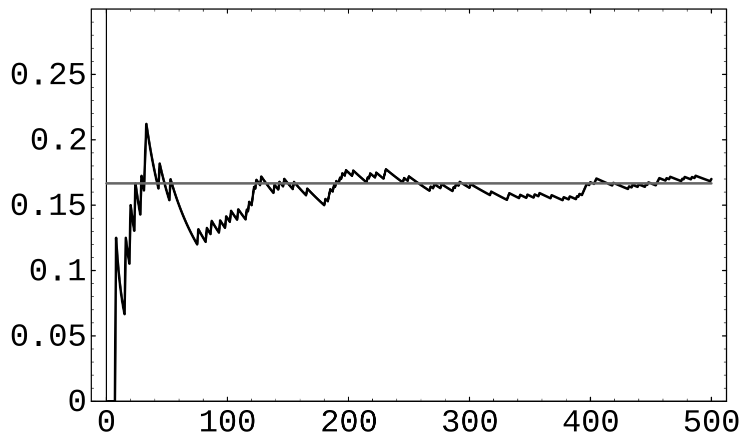

BILD 9.1. Relative Häufigkeit der Sechs bei einer Simulation von 500 Würfelwürfen.

9.7. Bemerkung. Es stellt sich angesichts des obigen Beispiels die Frage, ob nicht $\overline{Y}_n(\omega)$ für jedes ω gegen $\frac{1}{6}$ strebt (☞ Bild 9.1). Die Antwort, der wir im Rahmen dieser Vorlesung nicht weiter nachgehen werden, lautet: "Beinah". Es gilt nämlich

$$\lim_{n \to \infty} P(\{\omega : \overline{Y}_n(\omega) \to \tfrac{1}{6}\}) \;=\; 1.$$

Die Konvergenz gilt also bis auf ein Ausnahmeereignis, das Wahrscheinlichkeit 0 hat. Man spricht in diesem Fall von *P-fast sicherer Konvergenz.* Ein Problem, das hierbei jedoch völlig ausgeklammert und durch unsere Notation kaschiert wurde, ist die Frage nach einem mathematischen Modell für die Beschreibung unendlich vieler Würfelwürfe. Dies läßt sich in der Tat nur mittels eines überabzählbaren Ergebnisraums Ω, etwa $\Omega = \{1, 2, 3, 4, 5, 6\}^{\mathbb{N}}$, bewerkstelligen, auf dem die Y_i definiert sind. Da uns hier die Möglichkeiten fehlen, dieses maßtheoretische Problem zu behandeln (der interessierte Leser sei auf den Kurs "Wahrscheinlichkeitstheorie" verwiesen), werden wir die Existenz von unendlichen Folgen von Zufallsgrößen im folgenden einfach unterstellen. Wir weisen allerdings darauf hin, daß unsere Definitionen der $Y_1, ..., Y_n$ in 9.3 für jedes $n \geq 1$ auf einem Ergebnisraum vorgenommen wurden, der selbst von n abhing, nämlich $\Omega = \{1, 2, 3, 4, 5, 6\}^n$. Damit hing aber auch die Laplace-Verteilung P über Ω von n ab und hätte streng genommen mit n indiziert werden müssen.

Literatur

SCHMITZ, N. *Stochastik für Lehramtskandidaten.* Lit Verlag, Münster (1997).

[1] STROMBERG, K. *An Introduction to Classical Real Analysis.* Wadsworth, Belmont, Kalifornien (1981).