Erwartungswert

Naiv stellt man sich unter dem Erwartungswert einer Zufallsvariablen X Folgendes vor. Man führt das Experiment n-mal durch und bestimmt den Mittelwert (arithmetisches Mittel) der dabei für X erhaltenen Werte. Jetzt lässt man n gegen Unendlich gehen, und bestimmt den Grenzwert dieser Mittelwerte. Dieser Grenzwert ist der Erwartungswert $E(X)$ dieser Zufallsvariablen X. Damit ergibt sich auch die Vorstellung des Erwartungswerts, nämlich dass nach „vielen Versuchen“ der Mittelwert ungefähr der Erwartungswert ist.

Jetzt wollen wir den Erwartungswert axiomatisch definieren. Diese Definition entspricht der Definition des Integrals $\int X\,dP$ aus der Maßtheorie, das Lebesgue-Integral ist ein Spezialfall dieses maßtheoretischen Integrals. Man kann also sagen, dass $E(X) = \int X\,dP$ gilt, deshalb nennt man Zufallsvariable mit Erwartungswert auch integrierbar. Es ist die Definition des Erwartungswerts insofern ein bisschen einfacher als diejenige des allgemeinen maßtheoretischen Integrals, weil $P(\Omega) = 1$ gilt, während auf allgemeinen Maßräumen das Maß des ganzen Raumes auch den Wert ∞ ergeben kann (etwa beim Lebesgue-Maß). Zuerst wird der Erwartungswert für Zufallsvariable, die nur endlich viele Werte annehmen, definiert.

Definition 1. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die nur die endlich vielen Werte $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ annimmt. Dann definiert man $E(X) := \displaystyle\sum_{j=1}^{k} \alpha_j P(X = \alpha_j)$.

Proposition 1. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Es gäbe eine endliche Zerlegung $\mathcal{C}$ von Ω, sodass es für alle $C \in \mathcal{C}$ ein $a_C \in \mathbb{R}$ gibt, das $X(\omega) = a_C$ für alle $\omega \in C$ erfüllt. Dann gilt $E(X) = \displaystyle\sum_{C\in\mathcal{C}} a_C P(C)$.*

Beweis. Wegen der Voraussetzung nimmt die Zufallsvariable X nur endlich viele Werte $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ an. Für jedes $j \in \{1, 2, \ldots, k\}$ gibt es eine Teilmenge $\mathcal{C}_j \subseteq \mathcal{C}$ mit $a_C = \alpha_j$ für alle $C \in \mathcal{C}_j$. Deshalb erhält man, dass $\{X = \alpha_j\} = \bigcup_{C\in\mathcal{C}_j} C$ gilt. Nachdem $\mathcal{C}$ eine Zerlegung ist, sind ihre Elemente paarweise disjunkt, und daher ist $P(X = \alpha_j) = \sum_{C\in\mathcal{C}_j} P(C)$. Dadurch gilt auch $\sum_{C\in\mathcal{C}_j} \underbrace{a_C}_{=\alpha_j} P(C) = \alpha_j \underbrace{\displaystyle\sum_{C\in\mathcal{C}_j} P(C)}_{=P(X=\alpha_j)} = \alpha_j P(X = \alpha_j)$. Somit ist

$\sum_{C\in\mathcal{C}} a_C P(C) = \sum_{j=1}^{k} \underbrace{\sum_{C\in\mathcal{C}_j} a_C P(C)}_{=\alpha_j P(X=\alpha_j)} \underset{\text{Definition 1}}{=} E(X)$, womit das gewünschte Resultat bewiesen ist. □

Proposition 2. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die nur die endlich vielen Werte $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ annimmt, und $Y : \Omega \to \mathbb{R}$ sei eine Zufallsvariable, die nur die endlich vielen Werte $\{\beta_1, \beta_2, \ldots, \beta_s\}$ annimmt.*

(1) *Es gilt* $E(X+Y) = E(X) + E(Y)$.

(2) *Für* $\alpha \in \mathbb{R}$ *gilt* $E(\alpha X) = \alpha E(X)$.

(3) *Falls* $X \leq Y$ *erfüllt ist, dann gilt* $E(X) \leq E(Y)$.

Beweis. Für $j \in \{1, 2, \ldots, r\}$ und $k \in \{1, 2, \ldots, s\}$ setze

$$C_{j,k} := \{X = \alpha_j\} \cap \{Y = \beta_k\}\,.$$

Sei $\mathcal{C}$ die Zerlegung, die aus den nichtleeren $C_{j,k}$'s besteht. Dann gibt es für jedes $C \in \mathcal{C}$ ein $a_C \in \mathbb{R}$ und ein $b_C \in \mathbb{R}$, sodass $X(\omega) = a_C$ und $Y(\omega) = b_C$ für alle $\omega \in C$ gelten.

(1) Falls $C \in \mathcal{C}$, dann gilt $(X+Y)(\omega) = \underbrace{X(\omega)}_{=a_C} + \underbrace{Y(\omega)}_{=b_C} = a_C + b_C$ für alle $\omega \in C$. Deshalb gilt

$$\begin{aligned} E(X+Y) &\underset{\text{Proposition 1}}{=} \sum_{C\in\mathcal{C}} (a_C + b_C) P(C) = \\ &= \underbrace{\sum_{C\in\mathcal{C}} a_C P(C)}_{\underset{\text{Proposition 1}}{=} E(X)} + \underbrace{\sum_{C\in\mathcal{C}} b_C P(C)}_{\underset{\text{Proposition 1}}{=} E(Y)} = E(X) + E(Y)\,, \end{aligned}$$

womit (1) bewiesen ist.

(2) Es ist $(\alpha X)(\omega) = \alpha \underbrace{X(\omega)}_{=a_C} = \alpha a_C$ für alle $\omega \in C$, soferne $C \in \mathcal{C}$. Dadurch ist $E(\alpha X) \underset{\text{Proposition 1}}{=} \sum_{C\in\mathcal{C}} \alpha a_C P(C) = \alpha \underbrace{\sum_{C\in\mathcal{C}} a_C P(C)}_{\underset{\text{Proposition 1}}{=} E(X)} = \alpha E(X)$, wodurch auch (2) gezeigt ist.

(3) Wegen $X \leq Y$ gilt $a_C \leq b_C$ für jedes $C \in \mathcal{C}$. Somit ist

$$E(X) \underset{\text{Proposition 1}}{=} \sum_{C\in\mathcal{C}} \underbrace{a_C}_{\leq b_C} P(C) \leq \underbrace{\sum_{C\in\mathcal{C}} b_C P(C)}_{\underset{\text{Proposition 1}}{=} E(Y)} = E(Y)\,,$$

und auch (3) ist gezeigt. □

Wie üblich wird für $A \in \mathcal{A}$ die charakteristische Funktion $1_A : \Omega \to \mathbb{R}$ durch

$$1_A(\omega) := \begin{cases} 1\,, & \text{falls } \omega \in A, \\ 0\,, & \text{falls } \omega \in \Omega \setminus A, \end{cases}$$

definiert. Nimmt die Zufallsvariable X nur endlich viele Werte an, und ist $A \in \mathcal{A}$, so nimmt auch $1_A X$ nur endlich viele Werte an. Falls $X \geq 0$ gilt, und $A \in \mathcal{A}$ ist, dann gilt $X \geq 1_A X \geq 0$. Für $A, B \in \mathcal{A}$ mit $A \subseteq B$ und $X \geq 0$ gilt $X \geq 1_B X \geq 1_A X \geq 0$. Weiters ist $\sum_{C\in\mathcal{C}} 1_C = 1$, soferne $\mathcal{C}$ eine endliche Zerlegung von $(\Omega, \mathcal{A}, P)$ ist. Jetzt werden wir zeigen, dass $E(1_A X) = 0$ gilt, wenn X nur endlich viele Werte annimmt und $P(A) = 0$ ist.

Proposition 3. *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A \in \mathcal{A}$ mit $P(A) = 0$. Weiters sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die nur endlich viele Werte annimmt. Dann gilt $E(1_A X) = 0$.*

Beweis. Die Zufallsvariable X nimmt nur die Werte $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ an. Setze $C_0 := \Omega \setminus A$ und für $j \in \{1, 2, \ldots, k\}$ setze $C_j := A \cap \{X = \alpha_j\}$. Dann ist $\{C_0, C_1, C_2, \ldots, C_k\}$ eine endliche Zerlegung von $(\Omega, \mathcal{A}, P)$. Für $j \in \{1, 2, \ldots, k\}$ gilt $1_A X(\omega) = \alpha_j$ für alle $\omega \in C_j$ und es ist $0 \leq P(\underbrace{C_j}_{\subseteq A}) \leq P(A) = 0$, also $P(C_j) = 0$. Weiters gilt $1_A X(\omega) = 0$ für alle $\omega \in C_0$. Daher ist $E(1_A X) \underset{\text{Proposition 1}}{=} 0P(C_0) + \sum_{j=0}^{k} \alpha_j \underbrace{P(C_j)}_{=0} = 0$ und damit haben wir $E(1_A X) = 0$ gezeigt. □

Als nächsten Schritt wird der Erwartungswert für nichtnegative Zufallsvariable definiert. Dazu ist zunächst einmal auch der Wert $+\infty$ zugelassen. Um den Erwartungswert in diesem Fall zu definieren, benötigt man die folgenden Resultate.

Proposition 4. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die $X \geq 0$ erfüllt. Weiters seien $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsvariablen, die $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$, für jedes $n \in \mathbb{N}$ nimmt X_n nur endlich viele Werte an, $0 \leq Y_1 \leq Y_2 \leq Y_3 \leq \cdots$, für* *jedes $n \in \mathbb{N}$ nimmt Y_n nur endlich viele Werte an, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$ und $\lim_{n\to\infty} Y_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$ erfüllen. Dann gilt $\lim_{n\to\infty} E(X_n) = \sup_{n\in\mathbb{N}} E(X_n) = \lim_{n\to\infty} E(Y_n) = \sup_{n\in\mathbb{N}} E(Y_n)$, wobei für die Limiten und die Suprema auch der Wert $+\infty$ zugelassen ist.*

Beweis. Wegen Eigenschaft (3) aus Proposition 2 sind die Folgen $\big(E(X_n)\big)_{n\in\mathbb{N}}$ und $\big(E(Y_n)\big)_{n\in\mathbb{N}}$ monoton wachsend. Wird auch der Wert $+\infty$ als Grenzwert und Supremum zugelassen, so existieren die folgenden Grenzwerte und Suprema, und es gelten $\lim_{n\to\infty} E(X_n) = \sup_{n\in\mathbb{N}} E(X_n)$ und $\lim_{n\to\infty} E(Y_n) = \sup_{n\in\mathbb{N}} E(Y_n)$.

Es sei $s < \lim_{n\to\infty} E(X_n)$. Dann gibt es ein $N \in \mathbb{N}$ mit $E(X_N) > s$. Die Zufallsvariable X_N nimmt die Werte $\{\alpha_1, \alpha_2, \dots, \alpha_r\}$ an, dabei ist $\alpha_j \geq 0$ für alle $j \in \{1, 2, \dots, r\}$. Setze $\delta := \frac{1}{3}(E(X_N) - s)$. Dann ist $\delta > 0$ und es gilt $E(X_N) - 2\delta > s$. Für $j \in \{1, 2, \dots, r\}$ setze $A_j := \{X_N = \alpha_j\}$.

Fixiere ein $j \in \{1, 2, \dots, r\}$. Falls $\omega \in A_j$ ist, dann gilt $\alpha_j - \delta < \alpha_j \leq X(\omega)$. Definiere $A_{j,n} := \{\omega \in A_j : Y_n(\omega) > \alpha_j - \delta\}$. Weil $(Y_n)_{n\in\mathbb{N}}$ monoton wachsend ist, gilt $A_{j,n} \subseteq A_{j,n+1}$ für alle $n \in \mathbb{N}$. Nachdem $\lim_{n\to\infty} Y_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$, ist $\bigcup_{n=1}^{\infty} A_{j,n} = A_j$. Wegen der Stetigkeitseigenschaft gilt $\lim_{n\to\infty} P(A_{j,n}) = P(A_j)$ und daher $\lim_{n\to\infty}(\alpha_j - \delta)P(A_{j,n}) = (\alpha_j - \delta)P(A_j)$. Wenn $P(A_j) > 0$ ist, dann ist $(\alpha_j - \delta)P(A_j) > (\alpha_j - 2\delta)P(A_j)$, und im Fall $P(A_j) = 0$ ist $P(A_{j,n}) = 0$ für alle n. Deshalb gibt es ein $N_j \in \mathbb{N}$ mit $(\alpha_j - \delta)P(A_{j,n}) \geq (\alpha_j - 2\delta)P(A_j)$ für alle $n \geq N_j$. Ist $n \geq N_j$, so ist $1_{A_{j,n}} Y_n \geq (\alpha_j - \delta)1_{A_{j,n}}$, und damit

$$\begin{aligned} E(1_{A_j} Y_n) &\geq E(1_{A_{j,n}} Y_n) \geq E\big((\alpha_j - \delta)1_{A_{j,n}}\big) \underset{\text{(2) aus Proposition 2}}{=} \\ &= (\alpha_j - \delta)\underbrace{E(1_{A_{j,n}})}_{=P(A_{j,n})} = \underbrace{(\alpha_j - \delta)P(A_{j,n})}_{\geq(\alpha_j - 2\delta)P(A_j)} \geq (\alpha_j - 2\delta)P(A_j)\ . \end{aligned}$$

Somit wurde gezeigt, dass es zu jedem $j \in \{1, 2, \dots, r\}$ ein $N_j \in \mathbb{N}$ gibt, sodass $E(1_{A_j} Y_n) \geq (\alpha_j - 2\delta)P(A_j)$ für alle $n \geq N_j$ gilt.

Betrachte ein $n \in \mathbb{N}$ mit $n \geq \max\{N_1, N_2, \dots, N_r\}$. Weil $\{A_1, A_2, \dots, A_r\}$ eine endliche Zerlegung ist, gilt $\sum_{j=1}^{r} 1_{A_j} = 1$ Dann ist

$$E(Y_n) = E(Y_n \underbrace{1}_{=\sum_{j=1}^{r} 1_{A_j}}) = E\left(\sum_{j=1}^{r} 1_{A_j} Y_n\right) \underset{\text{(1) aus Proposition 2}}{=}$$

$$= \sum_{j=1}^{r} \underbrace{E(1_{A_j} Y_n)}_{\geq(\alpha_j - 2\delta)P(A_j)} \geq \underbrace{\sum_{j=1}^{r} \alpha_j P(A_j)}_{\underset{\text{Definition 1}}{=} E(X_N)} - 2\delta \underbrace{\sum_{j=1}^{r} P(A_j)}_{=1} = E(X_N) - 2\delta > s\ .$$

Dadurch ist $\lim_{n\to\infty} E(Y_n) = \sup_{n\in\mathbb{N}} E(Y_n) \geq s$, und deswegen erhalten wir $\lim_{n\to\infty} E(Y_n) = \sup_{n\in\mathbb{N}} E(Y_n) \geq \lim_{n\to\infty} E(X_n) = \sup_{n\in\mathbb{N}} E(X_n)$. Vertauscht man die Rollen von $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ so ergibt sich daraus auch $\lim_{n\to\infty} E(X_n) = \sup_{n\in\mathbb{N}} E(X_n) \geq \lim_{n\to\infty} E(Y_n) = \sup_{n\in\mathbb{N}} E(Y_n)$. Damit ist die Gleichheit gezeigt und das Resultat bewiesen. □

Proposition 5. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die $X \geq 0$ erfüllt. Für $n \in \mathbb{N}$ definiere $X_n : \Omega \to \mathbb{R}$ durch*

$$X_n(\omega) := \begin{cases} \frac{k}{2^n}\,, & \textit{falls } \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n} \textit{ für ein } k \in \{0, 1, 2, \ldots, 2^{2n} - 1\}, \\ 2^n\,, & \textit{falls } X(\omega) \geq 2^n. \end{cases}$$

Dann ist X_n für jedes $n \in \mathbb{N}$ eine Zufallsvariable, die nur endlich viele Werte annimmt. Weiters gilt $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$ und für jedes $\omega \in \Omega$ gilt $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$.

Beweis. Offensichtlich nimmt die Funktion X_n nur die endlich vielen Werte $\{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n\}$ an, und es gilt $\{X_n = \frac{k}{2^n}\} = \{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\} \in \mathcal{A}$, falls $k < 2^{2n}$, und $\{X_n = \frac{k}{2^n} = 2^n\} = \{X \geq 2^n\} \in \mathcal{A}$, falls $k = 2^{2n}$. Damit ist X_n eine Zufallsvariable, die nur endlich viele Werte annimmt, und $X_n \geq 0$ erfüllt. Sei $\omega \in \Omega$. Betrachte ein beliebiges $n \in \mathbb{N}$. Wenn $X(\omega) \geq 2^{n+1}$ gilt, dann ist $X_n(\omega) = 2^n \leq 2^{n+1} = X_{n+1}(\omega)$. Ist $X(\omega) \geq 2^{n+1}$, so gibt es ein $k \in \{0, 1, \ldots, 2^{2(n+1)} - 1\}$ mit $\frac{k}{2^{n+1}} \leq X(\omega) < \frac{k+1}{2^{n+1}}$. Soferne dieses $k \geq 2^{2n+1}$ ist, ist $X(\omega) \geq 2^n$ und $X_n(\omega) = 2^n = \frac{2^{2n+1}}{2^{n+1}} \leq \frac{k}{2^{n+1}} = X_{n+1}(\omega)$. Andernfalls gibt es ein $p \in\in \{0, 1, \ldots, 2^{2n} - 1\}$ mit $k = 2p$ oder $k = 2p + 1$. Dadurch ist $2p \leq k < k+1 \leq 2p+2 = 2(p+1)$ und somit $\frac{p}{2^n} \leq \frac{k}{2^{n+1}} \leq X(\omega) < \frac{k+1}{2^{n+1}} \leq \frac{p+1}{2^n}$. Deshalb ist $X_n(\omega) = \frac{p}{2^n} \leq \frac{k}{2^{n+1}} = X_{n+1}(\omega)$.

Jetzt sei $\varepsilon > 0$. Wähle $N \in \mathbb{N}$ so, dass $X(\omega) < 2^N$ und $\frac{1}{2^N} < \varepsilon$. Sei $n \geq N$. Dann gibt es ein $k \in \{0, 1, 2, \ldots, 2^{2n} - 1\}$ mit $\frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}$. Deswegen ist $X_n(\omega) = \frac{k}{2^n}$ und daher $|\underbrace{X_n(\omega)}_{=\frac{k}{2^n}} - X(\omega)| = \underbrace{X(\omega)}_{<\frac{k+1}{2^n}} - \frac{k}{2^n} < \frac{1}{2^n} < \varepsilon$. Nachdem $\omega \in \Omega$ beliebig war, gilt $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$ und $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$. □

Definition 2. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die $X \geq 0$ erfüllt. Dann definiert man $E(X)$ durch $E(X) := \lim\limits_{n\to\infty} E(X_n)$, wobei der Wert $+\infty$ für $E(X)$ zugelassen ist, und $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen ist, die $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$, für jedes $n \in \mathbb{N}$ nimmt X_n nur endlich viele Werte an, und $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$ erfüllen.

Diese Definition bedarf einer Rechtfertigung. Zunächst folgt aus Proposition 5, dass es eine Folge $(X_n)_{n\in\mathbb{N}}$ mit den in der Definition geforderten Eigenschaften gibt. Aus Proposition 4 folgt, dass diese Definition nicht von der konkreten Wahl der Folge abhängt. Jetzt zeigen wir, dass für nichtnegative Zufallsvariable mit nur endlich vielen Werten Definition 1 und Definition 2 denselben Wert ergeben.

Proposition 6. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die nur endlich viele Werte annimmt und $X \geq 0$ erfüllt. Weiters sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$, für jedes $n \in \mathbb{N}$ nimmt X_n nur endlich viele Werte an, und $\lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$ erfüllen. Dann ist $\lim\limits_{n\to\infty} E(X_n) = E(X)$.*

Beweis. Definiere $Y_n := X$ für $n \in \mathbb{N}$. Dann ist $0 \leq Y_1 \leq Y_2 \leq Y_3 \leq \cdots$, für jedes $n \in \mathbb{N}$ nimmt Y_n nur endlich viele Werte an, und $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$. Nach Proposition 4 ist $\lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} \underbrace{E(Y_n)}_{=E(X)} = E(X)$, wodurch das Resultat gezeigt ist. □

Wie üblich wird im folgenden Resultat die Konvention $x + (+\infty) = (+\infty) + x = +\infty$ für alle $x \in \mathbb{R}$ und $(+\infty) + (+\infty) = +\infty$. Außerdem soll wie üblich $x < +\infty$ für alle $x \in \mathbb{R}$ gelten.

Proposition 7. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, und seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ Zufallsvariable mit $X \geq 0$ und $Y \geq 0$.*

(1) *Es gilt* $E(X + Y) = E(X) + E(Y)$.

(2) *Für $\alpha \in \mathbb{R}$, $\alpha \geq 0$ gilt* $E(\alpha X) = \alpha E(X)$.

(3) *Es ist* $E(X) \geq 0$.

(4) *Wenn $X \leq Y$ gilt, dann ist* $E(X) \leq E(Y)$.

(5) *Falls $A \in \mathcal{A}$ und $P(A) = 0$, dann gilt* $E(1_A X) = 0$

Beweis. Wegen Proposition 5 gibt es Folgen $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ von Zufallsvariablen, sodass X_n und Y_n für jedes $n \in \mathbb{N}$ nur endlich viele Werte annehmen, $0 \leq X_1 \leq X_2 \leq X_3 \leq \cdots$, $0 \leq Y_1 \leq Y_2 \leq Y_3 \leq \cdots$, und $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ und $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$ für alle $\omega \in \Omega$ gelten.

(1) Auch $X_n + Y_n$ nimmt für jedes $n \in \mathbb{N}$ nur endlich viele Werte an, es ist $0 \leq X_1 + Y_1 \leq X_2 + Y_2 \leq X_3 + Y_3 \leq \cdots$, und für jedes $\omega \in \Omega$ gilt $\lim_{n\to\infty}(X_n + Y_n)(\omega) = \lim_{n\to\infty} X_n(\omega) + \lim_{n\to\infty} Y_n(\omega) = X(\omega) + Y(\omega) =$

$(X+Y)(\omega)$. Deshalb ist $E(X+Y) = \lim_{n\to\infty} \underbrace{E(X_n+Y_n)}_{\underset{(1)\text{ aus Proposition 2}}{=} E(X_n)+E(Y_n)} = \lim_{n\to\infty} E(X_n) + \lim_{n\to\infty} E(Y_n) = E(X) + E(Y)$.

(2) Weil $\alpha \geq 0$ ist, nimmt αX_n für alle $n \in \mathbb{N}$ nur endlich viele Werte an, wir haben $0 \leq \alpha X_1 \leq \alpha X_2 \leq \alpha X_3 \leq \cdots$, und für jedes $\omega \in \Omega$ gilt $\lim_{n\to\infty}(\alpha X_n)(\omega) = \alpha \lim_{n\to\infty} X_n(\omega) = \alpha X(\omega) = (\alpha X)(\omega)$. Man erhält $E(\alpha X) = \lim_{n\to\infty} \underbrace{E(\alpha X_n)}_{\underset{(2)\text{ aus Proposition 2}}{=} \alpha E(X_n)} = \alpha \lim_{n\to\infty} E(X_n) = \alpha E(X)$.

(3) Aus (3) in Proposition 2 folgt, dass $E(X_n) \geq E(0) = 0$ für alle $n \in \mathbb{N}$. Daher ist $E(X) = \lim_{n\to\infty} \underbrace{E(X_n)}_{\geq 0} \geq 0$.

(4) Setze $Z := Y - X$. Dann ist Z eine Zufallsvariable und wegen $X \leq Y$ ist $Z \geq 0$. Man erhält wegen (3), dass $E(Z) \geq 0$ gilt. Da $X+Z = X+Y-X = Y$, ergibt sich aus (1), dass $E(Y) = E(X+Z) = E(X) + \underbrace{E(Z)}_{\geq 0} \geq E(X)$.

(5) Es nimmt auch $1_A X_n$ für alle $n \in \mathbb{N}$ nur endlich viele Werte an, wir haben $0 \leq 1_A X_1 \leq 1_A X_2 \leq 1_A X_3 \leq \cdots$, und für jedes $\omega \in \Omega$ gilt $\lim_{n\to\infty}(1_A X_n)(\omega) = (1_A X)(\omega)$. Somit ergibt sich

$$E(1_A X) = \lim_{n\to\infty} \underbrace{E(1_A X_n)}_{\underset{\text{Proposition 3}}{=} 0} = 0\ ,$$

womit das gewünschte Resultat gezeigt ist. □

Damit können wir jetzt den Erwartungswert allgemein definieren. Zur Erinnerung erwähnen wir, dass $X = X^+ - X^-$ und $|X| = X^+ + X^-$ gelten.

Definition 3. Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Man sagt X ist integrierbar, falls $E(X^+) < +\infty$ und $E(X^-) < +\infty$. Für eine integriebare Zufallsvariable X wird der Erwartungswert $E(X)$ durch $E(X) := E(X^+) - E(X^-)$ definiert.

Definition 4. Die Familie aller integrierbaren Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ bezeichnet man mit $L^1(\Omega, \mathcal{A}, P)$ (oder kurz L^1, soferne es keine Missverständnisse geben kann).

Proposition 8. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann ist X genau dann integrierbar, wenn $|X|$ integrierbar ist, und es gilt $\big|E(X)\big| \leq E\big(|X|\big)$.*

Beweis. Zuerst sei X integrierbar. Wegen $E(X^+) < +\infty$ und $E(X^-) < +\infty$ ist nach (1) aus Proposition 7 $E(|X|) = E(X^+ + X^-) = E(X^+) + E(X^-) < +\infty$. Damit ist $|X|$ integrierbar.

Jetzt sei $|X|$ integrierbar. Weil $0 \leq X^+ \leq |X|$ und $0 \leq X^- \leq |X|$ folgt wegen (4) in Proposition 7, dass $E(X^+) \leq E(|X|) < +\infty$ und $E(X^-) \leq E(|X|) < +\infty$. Deswegen ist X integrierbar.

Es ist $|E(X)| = |E(X^+) - E(X^-)| \leq E(X^+) + E(X^-) = E(X^+ + X^-) = E(|X|)$. Wir haben damit auch die „Dreiecksungleichung“ für den Erwartungswert gezeigt. □

Proposition 9. *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, und es sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, die nur endlich viele Werte annimmt. Dann ist X integrierbar und der in Definition 2 definierte Erwartungswert stimmt mit den in Definition 3 definierten Erwartungswert überein.*

Beweis. Setze $A_1 := \{X \geq 0\}$ und $A_2 := \Omega \setminus A_1$. Dann ist $\{A_1, A_2\}$ eine endliche Zerlegung und es gelten $X^+ = 1_{A_1} X$ und $X^- = -1_{A_2} X$. Als endliche Summen sind $E(X^+) < +\infty$ und $E(X^-) < +\infty$, wodurch X integrierbar ist. Somit ist $E(X^+) - E(X^-) = E(1_{A_1}X) - \underbrace{E(-1_{A_2}X)}_{\underset{\text{(2) aus Proposition 2}}{=} -E(1_{A_2}X)} = E(1_{A_1}X) + E(1_{A_2}X) \underset{\text{(1) aus Proposition 2}}{=} E(\underbrace{(1_{A_1} + 1_{A_2})}_{=1} X) = E(X)$, womit gezeigt ist, dass beide Definitionen übereinstimmen. □

Proposition 10. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ seien integrierbare Zufallsvariable.*

(1) *Dann ist $X + Y$ integrierbar und es gilt* $E(X + Y) = E(X) + E(Y)$.

(2) *Für $\alpha \in \mathbb{R}$ ist αX integrierbar und es gilt* $E(\alpha X) = \alpha E(X)$.

(3) *Wenn $X \leq Y$ gilt, dann ist* $E(X) \leq E(Y)$.

(4) *Falls $A \in \mathcal{A}$ und $P(A) = 0$, dann gilt* $E(1_A X) = 0$

Beweis. (1) Setze $A_1 := \{X \geq 0\} \cap \{Y \geq 0\}$, $A_2 := \{X < 0\} \cap \{Y < 0\}$, $A_3 := \{X \geq 0\} \cap \{Y < 0\} \cap \{X + Y \geq 0\}$, $A_4 := \{X \geq 0\} \cap \{Y < 0\} \cap \{X + Y < 0\}$, $A_5 := \{X < 0\} \cap \{Y \geq 0\} \cap \{X + Y \geq 0\}$, $A_6 := \{X < 0\} \cap \{Y \geq 0\} \cap \{X + Y < 0\}$. Dann ist $\{A_1, A_2, A_3, A_4, A_5, A_6\}$ eine endliche Zerlegung. Auf A_1 ist $X^+ = X$, $Y^+ = Y$, $(X+Y)^+ = X+Y = X^+ + Y^+$, und $X^- = Y^- = (X+Y)^- = 0$, also $1_{A_1}(X+Y)^+ = 1_{A_1}(X^+ + Y^+)$ und $1_{A_1}(X+Y)^- = 1_{A_1}(X^- + Y^-)$. Daher ist wegen (1) aus Proposition 7 $E(1_{A_1}(X+Y)^+) = E(1_{A_1}X^+ + 1_{A_1}Y^+) = E(1_{A_1}X^+) + E(1_{A_1}Y^+)$ und $E(1_{A_1}(X+Y)^-) =$

$E(1_{A_1}X^- + 1_{A_1}Y^-) = E(1_{A_1}X^-) + E(1_{A_1}Y^-)$. Wir erhalten auf A_2, dass $X^+ = Y^+ = Y = (X+Y)^+ = 0$, und $X^- = -X$, $Y^- = -Y$ und $(X+Y)^- = -(X+Y) = X^- + Y^-$, also $1_{A_2}(X+Y)^+ = 1_{A_2}(X^+ + Y^+)$ und $1_{A_2}(X + Y)^- = 1_{A_2}(X^- + Y^-)$. Nach (1) aus Proposition 7 ergibt sich $E(1_{A_2}(X + Y)^+) = E(1_{A_2}X^+ + 1_{A_2}Y^+) = E(1_{A_2}X^+) + E(1_{A_2}Y^+)$ und $E(1_{A_2}(X+Y)^-) = E(1_{A_2}X^- + 1_{A_2}Y^-) = E(1_{A_2}X^-) + E(1_{A_2}Y^-)$. Man erhält auf A_3, dass $X^+ = X$, $Y^- = -Y$, $(X+Y)^+ = X+Y = X^+ - Y^-$, und $X^- = Y^+ = (X+Y)^- = 0$, also $1_{A_3}(X + Y)^+ = 1_{A_3}(X^+ - Y^-)$. Wegen (1) aus Proposition 7 ist $E(1_{A_3}X^+) = E(1_{A_3}(X + Y)^+ + 1_{A_3}Y^-) = E(1_{A_3}(X + Y)^+) + E(1_{A_3}Y^-)$, also $E(1_{A_3}(X + Y)^+) = E(1_{A_3}X^+) - E(1_{A_3}Y^-)$, und $E(1_{A_3}(X + Y)^-) = 0 = E(1_{A_3}X^-) - E(1_{A_3}Y^+)$. Auf A_4 ist $X^+ = X$, $Y^- = -Y$, $(X + Y)^- = -(X + Y) = -X^+ + Y^-$, und $X^- = Y^+ = (X + Y)^+ = 0$, und deshalb $1_{A_4}(X + Y)^- = 1_{A_4}(-X^+ + Y^-)$. Aus (1) in Proposition 7 erhalten wir $E(1_{A_4}Y^-) = E(1_{A_4}(X + Y)^- + 1_{A_4}X^+) = E(1_{A_4}(X + Y)^-) + E(1_{A_4}X^+)$, also $E(1_{A_4}(X + Y)^-) = -E(1_{A_4}X^+) + E(1_{A_4}Y^-)$, und $E(1_{A_4}(X + Y)^+) = 0 = -E(1_{A_4}X^-) + E(1_{A_4}Y^+)$. Wir erhalten auf A_5, dass $X^- = -X$, $Y^+ = Y$, $(X + Y)^+ = (X + Y) = -X^- + Y^+$, und $X^+ = Y^- = (X + Y)^- = 0$, also $1_{A_5}(X + Y)^+ = 1_{A_5}(-X^- + Y^+)$. Wegen (1) aus Proposition 7 ist $E(1_{A_5}Y^+) = E(1_{A_5}(X + Y)^+ + 1_{A_5}X^-) = E(1_{A_5}(X + Y)^+) + E(1_{A_5}X^-)$, also $E(1_{A_5}(X + Y)^+) = -E(1_{A_5}X^-) + E(1_{A_5}Y^+)$, und $E(1_{A_5}(X + Y)^-) = 0 = -E(1_{A_5}X^+) + E(1_{A_5}Y^-)$. Schließlich ergibt sich auf A_5, dass $X^- = -X$, $Y^+ = Y$, $(X+Y)^- = -(X+Y) = X^- - Y^+$, und $X^+ = Y^- = (X+Y)^- = 0$, also $1_{A_6}(X + Y)^- = 1_{A_6}(X^- - Y^+)$. Deshalb erhalten wir wegen (1) aus Proposition 7, dass $E(1_{A_6}X^-) = E(1_{A_6}(X + Y)^- + 1_{A_6}Y^+) = E(1_{A_6}(X + Y)^-) + E(1_{A_6}Y^+)$, somit $E(1_{A_6}(X + Y)^-) = E(1_{A_6}X^-) - E(1_{A_6}Y^+)$, und $E(1_{A_6}(X + Y)^+) = 0 = E(1_{A_6}X^+) - E(1_{A_6}Y^-)$. Deshalb ist $E(X + Y) = E\big((X + Y)^+\big) - E\big((X + Y)^-\big) = E\big(1_{A_1}(X + Y)^+\big) + E\big(1_{A_2}(X + Y)^+\big) + E\big(1_{A_3}(X+Y)^+\big) + E\big(1_{A_4}(X+Y)^+\big) + E\big(1_{A_5}(X+Y)^+\big) + E\big(1_{A_6}(X+Y)^+\big) - E\big(1_{A_1}(X+Y)^-\big) - E\big(1_{A_2}(X+Y)^-\big) - E\big(1_{A_3}(X+Y)^-\big) - E\big(1_{A_4}(X+Y)^-\big) - E\big(1_{A_5}(X+Y)^-\big) - E\big(1_{A_6}(X+Y)^-\big) = E(1_{A_1}X^+) + E(1_{A_1}Y^+) + E(1_{A_2}X^+) + E(1_{A_2}Y^+) + E(1_{A_3}X^+) - E(1_{A_3}Y^-) - E(1_{A_4}X^-) + E(1_{A_4}Y^+) - E(1_{A_5}X^-) + E(1_{A_5}Y^+) + E(1_{A_6}X^+) - E(1_{A_6}Y^-) - E(1_{A_1}X^-) - E(1_{A_1}Y^-) - E(1_{A_2}X^-) - E(1_{A_2}Y^-) - E(1_{A_3}X^-) + E(1_{A_3}Y^+) + E(1_{A_4}X^+) - E(1_{A_4}Y^-) + E(1_{A_5}X^+) - E(1_{A_5}Y^-) - E(1_{A_6}X^-) + E(1_{A_6}Y^+) = E(1_{A_1}X^+) + E(1_{A_2}X^+) + E(1_{A_3}X^+) + E(1_{A_4}X^+) + E(1_{A_5}X^+) + E(1_{A_6}X^+) - E(1_{A_1}X^-) - E(1_{A_2}X^-) - E(1_{A_3}X^-) - E(1_{A_4}X^-) - E(1_{A_5}X^-) - E(1_{A_6}X^-) + E(1_{A_1}Y^+) + E(1_{A_2}Y^+) + E(1_{A_3}Y^+) + E(1_{A_4}Y^+) + E(1_{A_5}Y^+) + E(1_{A_6}Y^+) - E(1_{A_1}Y^-) - E(1_{A_2}Y^-) - E(1_{A_3}Y^-) - E(1_{A_4}Y^-) - E(1_{A_5}Y^-) - E(1_{A_6}Y^-) = E(X^+) - E(X^-) + E(Y^+) - E(Y^-) = E(X) + E(Y)$.

(2) Im Fall $\alpha \geq 0$ ist $(\alpha X)^+ = \alpha X^+$ und $(\alpha X)^- = \alpha X^-$. Daher folgt wegen (2) aus Proposition 7, dass $E((\alpha X)^+) = E(\alpha X^+) = \alpha E(X^+)$ und

$E((\alpha X)^-) = E(\alpha X^-) = \alpha E(X^-)$. Daraus ergibt sich $E(\alpha X) = E((\alpha X)^+) - E((\alpha X)^-) = \alpha(E(X^+) - E(X^-)) = \alpha E(X)$. Für $\alpha < 0$ ist $(\alpha X)^+ = (-\alpha)X^-$ und $(\alpha X)^- = (-\alpha)X^+$. Wegen (2) aus Proposition 7 erhalten wir $E((\alpha X)^+) = E((-\alpha)X^-) = -\alpha E(X^-)$ und $E((\alpha X)^-) = E((-\alpha)X^+) = -\alpha E(X^+)$, und deshalb $E(\alpha X) = E((\alpha X)^+) - E((\alpha X)^-) = -\alpha(E(X^-) - E(X^+)) = \alpha E(X)$.

(3) Da $X \leq Y$ ist, ist $Y - X \geq 0$ und wegen (3) aus Proposition 7 erhält man, dass $E(Y - X) \geq 0$. Unter Verwendung von (1) und (2) ergibt sich daraus, dass $0 \leq E(Y - X) = E(Y + (-1)X) = E(Y) + E((-1)X) = E(Y) + (-1)E(X) = E(Y) - E(X)$, und deshalb $E(X) \leq E(Y)$.

(4) Hier ist $(1_A X)^+ = 1_A X^+$ und $(1_A X)^- = 1_A X^-$. Wegen (5) aus Proposition 7 ist $E(1_A X^+) = 0$ und $E(1_A X^-) = 0$. Deswegen erhält man $E(1_A X) = E((1_A X)^+) - E((1_A X)^-) = E(1_A X^+) - E(1_A X^-) = 0$. □

Proposition 11. *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.*

(1) *Falls $Y : \Omega \to \mathbb{R}$ eine integrierbare Zufallsvariable ist, und $X = Y$ fast überall gilt (also $P(X \neq Y) = 0$ ist), dann ist X integrierbar und es gilt $E(X) = E(Y)$.*

(2) *Es gäbe ein $\alpha \in \mathbb{R}$ mit $X = \alpha$ fast überall (also $P(X = \alpha) = 1$). Dann ist X integrierbar und es gilt $E(X) = \alpha$.*

Beweis. (1) Setze $A_1 := \{X \neq Y\}$ und $A_2 := \Omega \setminus A_1$. Dann ist $\{A_1, A_2\}$ eine endliche Zerlegung und $P(A_1) = 0$. Es ist

$$E(X) = E(1_{A_1}X + 1_{A_2}X) \underset{\text{(1) aus Proposition 10}}{=} \underbrace{E(1_{A_1}X)}_{\underset{\text{(4) aus Proposition 10}}{=} 0} + E(\underbrace{1_{A_2}X}_{=1_{A_2}Y}) =$$

$$= \underbrace{0}_{\underset{\text{(4) aus Proposition 10}}{=} E(1_{A_1}Y)} + E(1_{A_2}Y) \underset{\text{(1) aus Proposition 10}}{=} E(1_{A_1}Y + 1_{A_2}Y).$$

(2) Nach Proposition 9 ist $E(\alpha) = \alpha P(\Omega) = \alpha$. Wegen (1) ist X integrierbar und $E(X) = \alpha$. □