# Inhaltsverzeichnis

KAPITEL 2

# Grundbegriffe der Wahrscheinlichkeitstheorie

## 1. Problemstellungen der Wahrscheinlichkeitstheorie

Eine Nachrichtenquelle produziert eine Folge von 100 Zeichen auf folgende Art: Für jedes Zeichen wird gewürfelt. Fällt der Würfel auf 1, 2, oder 3, so wird das Zeichen $a$ produziert. Fällt der Würfel auf 4 oder 5, wird $b$ produziert. Fällt der Würfel auf 6, so wird $c$ produziert.

Wir fragen uns nun, wieviele $a$ wir erwarten dürfen. Zunächst würde man erwarten, dass etwa die Hälfte der Zeichen $a$ sind. Dennoch ist es möglich, dass der Würfel 100 mal hintereinander auf 4, 5, oder 6 fällt, und kein einziges Zeichen ein $a$ ist. "Möglich schon, aber sehr unwahrscheinlich", sagt die Wahrscheinlichkeitstheorie dazu.

Produziert man nun anstatt 100 Zeichen 10000 oder 100000 Zeichen, so scheint es plausibel, dass der Anteil an $a$ den erwarteten 50% schließlich ganz nahe kommt. Trotzdem ist es denkbar, dass auch unter 100000 Zeichen kein einziges $a$ vorkommt. In den "Gesetzen der großen Zahlen" erhalten wir mathematische Aussagen darüber.

## 2. Wahrscheinlichkeit

Wir wollen beschreiben, dass die drei Zeichen $a$, $b$, $c$ mit den "Wahrscheinlichkeiten" $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$ auftreten. Dazu definieren wir den Begriff *Wahrscheinlichkeitsraum.* Ein *Wahrscheinlichkeitsraum* besteht aus einer endlichen Menge, und einer Funktion, die jedem Element seine relative Häufigkeit zuordnet.

Definition 2.1 (Endliche Wahrscheinlichkeitsräume). Sei $\Omega$ eine endliche nichtleere Menge, und sei $P : \Omega \to [0, 1]$. Das Paar $(\Omega, P)$ ist ein *Wahrscheinlichkeitsraum*, falls

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

Wir können nun jede Teilmenge eines Wahrscheinlichkeitsraumes messen.

DEFINITION 2.2. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $A$ eine Teilmenge von $\Omega$. Dann definieren wir das *Maß von $A$*, $P(A)$, durch

$$P(A) := \sum_{a \in A} P(a).$$

In der Wahrscheinlichkeitstheorie beschreibt man Vorgänge, die vom Zufall abhängen, etwa vom Wurf einer Münze, von der Augenzahl, die ein Würfel zeigt, oder von der Kiste, in die eine Roulettekugel fällt. So hängt die Auswahl der Zeichen $a$, $b$, $c$ im Beispiel der vorigen Sektion davon ab, wie der Würfel gefallen ist. Es ist zufällig, welches der drei Zeichen $a$, $b$, $c$ die Nachrichtenquelle als nächstes produziert. Wenn aber der Würfel gefallen ist, so ist die Auswahl bestimmt. Für die mathematische Beschreibung der Auswahl eines Zeichens trennt man das Würfeln vom deterministischen Vorgang, der Auswahl des Zeichens aus der Augenzahl.

Das Würfeln kodiert man durch einen Wahrscheinlichkeitsraum $(\Omega, P)$ mit

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

und

$$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}.$$

Die Auswahl des Zeichens kodiert man durch eine Funktion von $\{1, 2, 3, 4, 5, 6\}$ nach $\{a, b, c\}$. Im genannten Beispiel verwenden wir die Funktion $X$ mit $X(1) = X(2) = X(3) = a$, $X(4) = X(5) = b$, $X(6) = c$. Funktionen, die auf der Trägermenge eines Wahrscheinlichkeitsraumes definiert sind, heißen *Zufallsvariablen.*

DEFINITION 2.3. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $M$ eine Menge. Eine $M$-wertige *Zufallsvariable* $X$ auf $(\Omega, P)$ ist eine Funktion $X : \Omega \to M$.

Wir interessieren uns nun dafür, wie häufig das Zeichen $b$ ausgewählt wird. Dazu messen wir, wie groß die Menge $\{\omega \mid X(\omega) = b\}$ ist, nennen das Maß dieser Menge *die Wahrscheinlichkeit für $X = b$*, und kürzen diese mit $P[X = b]$ ab. In unserem Fall erhalten wir

$$P[X = b] = P(\{\omega \mid X(\omega) = b\}) = P(\{4, 5\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

DEFINITION 2.4 (Wahrscheinlichkeit). Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, $M$ eine Menge, und $X : \Omega \to M$ eine Zufallsvariable. Sei $Z$ eine Teilmenge von $M$. Dann definieren wir *die Wahrscheinlichkeit, dass $X$ in $Z$ liegt*, abgekürzt $P[X \in Z]$, durch

$$P[X \in Z] := P(\{\omega \mid X(\omega) \in Z\}).$$

Wenn $Z$ einelementig ist und $Z = \{z\}$, dann schreiben wir auch $P[X = z]$ für $P[X \in Z]$.

Wir überlegen uns nun, wie wir Vorgänge beschreiben, die von mehrmaligem Würfeln abhängen. Dazu betrachten wir folgendes Beispiel: Wie groß ist die Wahrscheinlichkeit, bei zweimaligem Würfeln die Augensumme 6 zu erhalten? Wir gehen davon aus, dass bei zweimaligem Würfeln jede Folge gleich wahrscheinlich ist: $(1,1)$ ist also gleich wahrscheinlich wie $(2,3)$, $(3,2)$ oder $(1,6)$. Daher definieren wir einen Wahrscheinlichkeitsraum $(\Omega, P)$ durch

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1,2,3,4,5,6\}\}, \\ P((\omega_1, \omega_2)) &= \tfrac{1}{36} \text{ für alle } \omega_1, \omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Die Zufallsvariable $X$ definieren wir durch

$$X((\omega_1, \omega_2)) := \omega_1 + \omega_2.$$

Wir suchen die Wahrscheinlichkeit für $X = 6$. Wir erhalten

$$\begin{aligned} P[X = 6] &= P(\{(\omega_1, \omega_2) \in \{1,2,3,4,5,6\}^2 \mid X((\omega_1, \omega_2)) = 6\}) \\ &= P(\{(\omega_1, \omega_2) \in \{1,2,3,4,5,6\}^2 \mid \omega_1 + \omega_2 = 6\}) \\ &= P(\{(1,5), (2,4), (3,3), (4,2), (5,1)\}) \\ &= \frac{5}{36}. \end{aligned}$$

Also ist die Augensumme mit Wahrscheinlichkeit $\frac{5}{36}$, also $13.888\ldots\%$, gleich 6.

Wir betrachten bei zweimaligem Würfeln folgende Zufallsvariablen:

(1) $X((\omega_1, \omega_2)) := \omega_1$,
(2) $Y((\omega_1, \omega_2)) := \omega_2$,
(3) $Z((\omega_1, \omega_2)) := \omega_1 + \omega_2$.

$X$ liefert also das Ergebnis des ersten Wurfes, $Y$ das Ergebnis des zweiten Wurfes, und $Z$ die Summe der Augenzahlen beider Würfe. Der Ausgang von $X$ und der Ausgang von $Y$ beeinflussen einander nicht; wenn ich auch weiß, dass der erste Wurf 5 ist, so bleiben für den zweiten Wurf doch alle Ausgänge gleich wahrscheinlich. Der Ausgang von $X$ liefert aber Einschränkungen für den Ausgang von $Z$. Ist der erste Wurf 5, so ist die Summe bestimmt nicht mehr 2 oder 12. Wenn Zufallsvariablen einander nicht beeinflussen, so nennt man sie *unabhängig*.

Definition 2.5. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $M, N$ Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. $X$ und $Y$ sind *unabhängig*, falls für alle $a \in M$ und $b \in N$ gilt:

$$P[X = a \; \& \; Y = b] = P[X = a] \cdot P[Y = b].$$

Dabei steht $P[X = a \; \& \; Y = b]$ für das Maß $P(\{\omega \in \Omega \mid X(\omega) = a \text{ und } Y(\omega) = b\})$.

Lemma 2.6. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $M, N$ Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. Äquivalent sind:*

(1) *$X$ und $Y$ sind unabhängig.*
(2) *Für alle Teilmengen $A$ von $M$ und für alle Teilmengen $B$ von $N$ gilt $P[X \in A \mathbin{\&} Y \in B] = P[X \in A] \cdot P[Y \in B]$.*

*Dabei steht $P[X \in A \mathbin{\&} Y \in B]$ für $P(\{\omega \in \Omega \mid X(\omega) \in A$ und $Y(\omega) \in B\})$.*

*Beweis:* (1)⇒(2):

$$
\begin{aligned}
P[X \in A \mathbin{\&} Y \in B] &= P(\{\omega \in \Omega \mid X(\omega) \in A \text{ und } Y(\omega) \in B\}) \\
&= \sum_{\substack{\omega \in \Omega \\ X(\omega) \in A \text{ und } Y(\omega) \in B}} P(\omega) \\
&= \sum_{a \in A} \sum_{b \in B} \sum_{\substack{\omega \in \Omega \\ X(\omega) = a \text{ und } Y(\omega) = b}} P(\omega) \\
&= \sum_{a \in A} \sum_{b \in B} P(\{\omega \in \Omega \mid X(\omega) = a \text{ und } Y(\omega) = b\}) \\
&= \sum_{a \in A} \sum_{b \in B} P[X = a \mathbin{\&} Y = b].
\end{aligned}
$$

Nun verwenden wir die Unabhängigkeit von $X$ und $Y$ und erhalten

$$
\begin{aligned}
\sum_{a \in A} \sum_{b \in B} P[X = a \mathbin{\&} Y = b] &= \sum_{a \in A} \sum_{b \in B} P[X = a] \cdot P[Y = b] \\
&= (\sum_{a \in A} P[X = a]) \cdot (\sum_{b \in B} P[X = b]) \\
&= P[X \in A] \cdot P[X \in B].
\end{aligned}
$$

(2)⇒(1): Wir fixieren $a \in M$ und $b \in N$. Wir erhalten $P[X = a \mathbin{\&} Y = b] = P[X \in \{a\} \mathbin{\&} Y \in \{b\}]$. Wegen (2) ist der letzte Ausdruck gleich $P[X \in \{a\}] \cdot P[Y \in \{b\}] = P[X = a] \cdot P[X = b]$. □

Definition 2.7 (Bedingte Wahrscheinlichkeit). Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $M, N$ Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. Sei $A$ eine Teilmenge von $M$ und $B$ eine Teilmenge von $N$ mit $P[Y \in B] \neq 0$. Dann definieren wir

$$P[X \in A \mid Y \in B] := \frac{P[X \in A \mathbin{\&} Y \in B]}{P[Y \in B]},$$

und nennen die linke Seite *die Wahrscheinlichkeit, dass $X$ in $A$ ist, wenn wir schon wissen, dass $Y$ in $B$ ist.*

Wir schränken also unser Interesse auf jene Ereignisse ein, für die $Y$ in $B$ liegt, und messen, für welchen Anteil von diesen $X$ in $A$ liegt.

Wir überlegen uns dazu folgendes Beispiel: Wie groß ist die Wahrscheinlichkeit, dass bei zweimaligem Würfeln der erste Wurf ein Einser ist, wenn die Augensumme 4 ist?

Zur Lösung wählen wir den Wahrscheinlichkeitsraum

$$\begin{aligned} \Omega &= \{(\omega_1,\omega_2) \mid \omega_1,\omega_2 \in \{1,2,3,4,5,6\}\}, \\ P((\omega_1,\omega_2)) &= \tfrac{1}{36} \text{ für alle } \omega_1,\omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Wir definieren die Zufallsvariablen $X$ und $Y$ durch

$$\begin{aligned} X(\omega_1,\omega_2) &:= \omega_1 \text{ für alle } \omega_1,\omega_2 \in \{1,2,3,4,5,6\}, \\ Y(\omega_1,\omega_2) &:= \omega_1+\omega_2 \text{ für alle } \omega_1,\omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Gesucht ist dann $P[X = 1 \mid Y = 4]$. Wir berechnen

$$\begin{aligned} P[X = 1 \mid Y = 4] &= \frac{P[X = 1 \;\&\; Y = 4]}{P[Y = 4]} \\ &= \frac{P(\{\omega \in \Omega \mid X(\omega) = 1 \text{ und } Y(\omega) = 4\})}{P(\{\omega \in \Omega \mid Y(\omega) = 4\})} \\ &= \frac{P(\{(1,3)\})}{P(\{(1,3),(2,2),(3,1)\})} \\ &= \frac{\frac{1}{36}}{\frac{3}{36}} \\ &= \frac{1}{3}. \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass der erste Wurf ein Einser war, wenn wir schon wissen, dass die Augensumme 4 ist, ist also $\frac{1}{3}$.

Aus Neugier stellen wir auch folgende Frage: Wie groß ist die Wahrscheinlichkeit, dass die Augensumme 4 ist, wenn wir schon wissen, dass der erste Wurf ein Einser war. Wir berechnen dazu

$$\begin{aligned} P[Y = 4 \mid X = 1] &= \frac{P[Y = 4 \;\&\; X = 1]}{P[X = 1]} \\ &= \frac{P(\{(1,3)\})}{P(\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\})} \\ &= \frac{\frac{1}{36}}{\frac{6}{36}} \\ &= \frac{1}{6}. \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass die Augensumme 4 ist, wenn wir schon wissen, dass der erste Wurf ein Einser ist, ist also $\frac{1}{6}$. Wenn wir uns in dieses Experiment erst zu dem Zeitpunkt einschalten, an dem der erste Würfel auf Eins gefallen ist, müssen wir nur mehr beobachten, ob der zweite Würfel auf 3 fällt, damit die Augensumme 4 ist. Der zweite Würfel fällt mit Wahrscheinlichkeit $\frac{1}{6}$ auf 3.

PROPOSITION 2.8. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $M, N$ Mengen, seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen , und sei sei $A \subseteq M$, $B \subseteq N$. Wenn $X$ und $Y$ unabhängig sind und $P[Y \in B] \neq 0$, so gilt*

$$P[X \in A \mid Y \in B] = P[X \in A].$$

Selbst wenn wir wissen, dass $Y$ in $B$ liegt, hat das keinen Einfluss auf die Wahrscheinlichkeit dafür, dass $X$ in $A$ liegt.

*Beweis:*

$$P[X \in A \mid Y \in B] = \frac{P[X \in A \ \&\ Y \in B]}{P[Y \in B]}.$$

Nun verwenden wir, dass $X$ und $Y$ unabhängig sind, und erhalten

$$\begin{aligned} \frac{P[X \in A \ \&\ Y \in B]}{P[Y \in B]} &= \frac{P[X \in A] \cdot P[Y \in B]}{P[Y \in B]} \\ &= P[X \in A]. \end{aligned}$$

□

## 3. Erwartungswert und Varianz

Wenn $X$ eine Zufallsvariable vom Wahrscheinlichkeitsraum $\Omega$ nach $\mathbb{R}$ ist, dann bezeichnen wir mit $E(X)$ den *Durchschnitt* ihrer Ausgänge, und nennen ihn *Erwartungswert*.

DEFINITION 2.9. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann definieren wir den *Erwartungswert* $E(X)$ *von* $X$ durch

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Wie üblich bezeichnen wir $\{X(\omega) \mid \omega \in \Omega\}$ mit $X(\Omega)$.

LEMMA 2.10. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann gilt*

$$E(X) = \sum_{s \in X(\Omega)} s \cdot P[X = s]. \tag{2.1}$$

*Beweis:* Wir definieren $\delta(a,b) = 1$, falls $a = b$ und $\delta(a,b) = 0$, falls $a \neq b$. Wir formen nun die rechte Seite von (2.1) um und erhalten:

$$\begin{aligned}
\sum_{s \in X(\Omega)} s \cdot P[X = s] &= \sum_{s \in X(\Omega)} s \cdot P(\{\omega \in \Omega \mid X(\omega) = s\}) \\
&= \sum_{s \in X(\Omega)} s \cdot \sum_{\substack{\omega \in \Omega \\ X(\omega)=s}} P(\omega) \\
&= \sum_{s \in X(\Omega)} s \cdot \sum_{\omega \in \Omega} P(\omega) \cdot \delta(X(\omega), s) \\
&= \sum_{s \in X(\Omega)} \sum_{\omega \in \Omega} s \cdot P(\omega) \cdot \delta(X(\omega), s) \\
&= \sum_{\omega \in \Omega} \sum_{s \in X(\Omega)} s \cdot P(\omega) \cdot \delta(X(\omega), s).
\end{aligned}$$

In der inneren Summe im letzten Ausdruck ist nur der Summand mit $s = X(\omega)$ ungleich 0. Daher ist der letzte Ausdruck gleich

$$\sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \cdot 1,$$

und das ist genau der Erwartungswert $E(X)$. □

Eine wichtige Eigenschaft des Erwartungswertes ist seine Linearität.

SATZ 2.11. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $X, Y$ Zufallsvariablen von $\Omega \to \mathbb{R}$, und sei $t \in \mathbb{R}$. Dann gilt:*

(1) $E(X + Y) = E(X) + E(Y)$.
(2) $E(t \cdot X) = t \cdot E(X)$.

Dabei ist $t \cdot X$ die Funktion, die jedes $\omega$ auf $t \cdot X(\omega)$ abbildet.

*Beweis:* Die erste Eigenschaft beweisen wir durch

$$\begin{aligned}
E(X + Y) &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot P(\omega) \\
&= \sum_{\omega \in \Omega} (X(\omega) \cdot P(\omega) + Y(\omega) \cdot P(\omega)) \\
&= \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) + \sum_{\omega \in \Omega} Y(\omega) \cdot P(\omega) \\
&= E(X) + E(Y).
\end{aligned}$$

Die zweite Eigenschaft beweisen wir durch

$$\begin{aligned} E(t \cdot X) &= \sum_{\omega \in \Omega} (t \cdot X)(\omega) \cdot P(\omega) \\ &= \sum_{\omega \in \Omega} t \cdot X(\omega) \cdot P(\omega) \\ &= t \cdot \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \\ &= t \cdot E(X). \end{aligned}$$

□

Die folgende Eigenschaft unabhängiger Zufallsvariablen ist verblüffend:

LEMMA 2.12. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und seien $X, Y : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen. Dann gilt*

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

*Beweis:* Wir berechnen $E(X) \cdot E(Y)$. Wir verwenden dazu die Darstellung von $E(X)$ aus Lemma 2.10 und erhalten

$$\begin{aligned} E(X) \cdot E(Y) &= \Big( \sum_{r \in X(\Omega)} r \cdot P[X = r]\Big) \cdot \Big( \sum_{s \in Y(\Omega)} s \cdot P[Y = s]\Big) \\ &= \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r] \cdot P[Y = s]). \end{aligned}$$

Nun verwenden wir die Unabhängigkeit der Zufallsvariablen $X$ und $Y$.

$$\sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r] \cdot P[Y = s] = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r \ \& \ Y = s].$$

Wir verwenden wieder die Funktion $\delta$ mit $\delta(a, b) = 1$, falls $a = b$ und $\delta(a, b) = 0$, falls $a \neq b$. Wir fassen für jedes $t \in \mathbb{R}$ alle Summanden zusammen, für die $r \cdot s = t$ ist. So erhalten wir

$$\begin{aligned} &\sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r \ \& \ Y = s] \\ &\qquad = \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \ \& \ Y = s] \cdot \delta(r \cdot s, t). \end{aligned}$$

Wir zeigen nun als nächstes:

$$\text{Für alle } t \in \mathbb{R}: \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \ \& \ Y = s] \cdot \delta(r \cdot s, t) = P[X \cdot Y = t]. \tag{2.2}$$

Um das zu begründen, beweisen wir zunächst:

(2.3) Für alle $t \in \mathbb{R}$ gilt:

$$\bigcup_{r \in X(\Omega)} \bigcup_{s \in Y(\Omega)} \{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\} = \{\omega \in \Omega \mid X(\omega) \cdot Y(\omega) = t\}.$$

Wir fixieren $t \in \mathbb{R}$. Die Inklusion $\subseteq$ ist offensichtlich. Für die Inklusion $\supseteq$ fixieren wir $\omega$ in $\Omega$ so, dass $X(\omega) \cdot Y(\omega) = t$. Das Element $\omega$ kommt dann in jener Menge der Vereinigung auf der linken Seite von (2.3) vor, für die $r := X(\omega)$ und $s := Y(\omega)$ ist. Das beweist (2.3). Da die Mengen, die auf der linken Seite von (2.3) vereinigt werden, paarweise disjunkt sind, gilt

$$\begin{aligned} & P\left( \bigcup_{r \in X(\Omega)} \bigcup_{s \in Y(\Omega)} \{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\} \right) \\ & \quad = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P(\{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\}) \\ & \qquad = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P(\{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s\}) \cdot \delta(r \cdot s, t). \end{aligned}$$

Das Maß der Menge auf der linken Seite von (2.3) steht also auf der linken Seite von (2.2), und das Maß der Menge auf der rechten Seite von (2.3) steht auf der rechten Seite von (2.2). Das beweist (2.2). Wir verwenden nun (2.2), um weiterzurechnen, und erhalten

$$\begin{aligned} \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \,\&\, Y = s] \cdot \delta(r \cdot s, t) &= \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot P[X \cdot Y = t] \\ &= E(X \cdot Y). \end{aligned}$$

□

Ein Maß für die Schwankung einer Zufallsvariablen um ihren Erwartungswert ist die *Varianz*. Für einen Wahrscheinlichkeitsraum $(\Omega, P)$ und eine reelle Zahl $a$ bezeichnen wir mit $\overline{a}$ die Funktion, die durch

$$\begin{array}{cccc} \overline{a} : & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & a \end{array}$$

gegeben ist; $\overline{a}$ ist also die konstante Funktion mit Funktionswert $a$.

Definition 2.13 (Varianz). Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann ist die *Varianz* von $X$, abgekürzt $V(X)$, definiert durch

$$V(X) := E\left( \left( X - \overline{E(X)} \right)^2 \right).$$

Die Varianz ist also der Erwartungswert der Zufallsvariablen $Y$, die durch $Y(\omega) = (X(\omega) - E(X))^2$ definiert ist.

Satz 2.14. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, sei $n \in \mathbb{N}$, und seien $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen, die paarweise voneinander unabhängig sind. Dann gilt*

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n).$$

*Beweis:*

$$\begin{aligned}
V\left(X_1 + X_2 + \cdots + X_n\right) &= E\left(\left(\left(\sum_{i=1}^n X_i\right) - \overline{E\left(\sum_{i=1}^n X_i\right)}\right)^2\right) \\
&= E\left(\left(\left(\sum_{i=1}^n X_i\right) - \overline{\sum_{i=1}^n E\left(X_i\right)}\right)^2\right) \\
&= E\left(\left(\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \overline{E\left(X_i\right)}\right)^2\right) \\
&= E\left(\left(\sum_{i=1}^n \left(X_i - \overline{E\left(X_i\right)}\right)\right)^2\right) \\
&= E\left(\sum_{i=1}^n \sum_{j=1}^n \left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right).
\end{aligned}$$

Wir trennen nun die Summanden, für die $i = j$ ist, von denen, für die $i \neq j$. Wir erhalten

$$\begin{aligned}
&\sum_{i=1}^n \sum_{j=1}^n E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) \\
&= \sum_{i=1}^n E\left(\left(X_i - \overline{E\left(X_i\right)}\right)^2\right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right).
\end{aligned} \tag{2.4}$$

Wir fixieren nun $i, j \in \{1, 2, \ldots, n\}$ mit $i \neq j$ und betrachten den Ausdruck

$$E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right). \tag{2.5}$$

Wir rechnen

$$\begin{aligned}&E\left(\left(X_i - \overline{E(X_i)}\right) \cdot \left(X_j - \overline{E(X_j)}\right)\right)\\&\qquad = E\left(X_i \cdot X_j - \overline{E(X_i)} \cdot X_j - X_i \cdot \overline{E(X_j)} + \overline{E(X_i)} \cdot \overline{E(X_j)}\right).\end{aligned}$$

Wir verwenden die Linearität des Erwartungswertes; außerdem verwenden wir, dass $\overline{E(X_i)}$ und $\overline{E(X_j)}$ konstante Funktionen sind und somit $E\left(\overline{E(X_i)} \cdot X_j\right) = E(E(X_i) \cdot X_j) = E(X_i) \cdot E(X_j)$ gilt. Außerdem verwenden wir, dass der Erwartungswert einer konstanten Funktion gleich ihrem (einzigen) Funktionswert ist.

$$\begin{aligned}&E\left(X_i \cdot X_j - \overline{E(X_i)} \cdot X_j - X_i \cdot \overline{E(X_j)} + \overline{E(X_i)} \cdot \overline{E(X_j)}\right)\\&\quad = E(X_i \cdot X_j) - E(X_i) \cdot E(X_j) - E(X_j) \cdot E(X_i) + E(X_i) \cdot E(X_j)\\&\quad = E(X_i \cdot X_j) - E(X_i) \cdot E(X_j).\end{aligned}$$

Da $X_i$ und $X_j$ unabhängig sind, liefert Lemma 2.12, dass der letzte Ausdruck 0 ist. Somit ist (2.5) gleich 0. Wenn wir nun in (2.4) weiterrechnen, erhalten wir

$$\begin{aligned}&\sum_{i=1}^{n} E\left(\left(X_i - \overline{E(X_i)}\right)^2\right) + \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} E\left(\left(X_i - \overline{E(X_i)}\right) \cdot \left(X_j - \overline{E(X_j)}\right)\right)\\&\qquad = \sum_{i=1}^{n} E\left(\left(X_i - \overline{E(X_i)}\right)^2\right) = \sum_{i=1}^{n} V(X_i).\end{aligned}$$

□

## 4. Das schwache Gesetz der großen Zahlen

Zur Beschreibung von Zufallsexperimenten, bei denen mehrmals hintereinander gewürfelt wird, eignen sich *Produkträume*. Wir kürzen Vektoren mit fettgedruckten Buchstaben ab, also zum Beispiel $(\omega_1, \omega_2, \ldots, \omega_n)$ mit $\boldsymbol{\omega}$.

Definition 2.15 (Produktraum). Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum. Der *n-fache Produktraum* von $(\Omega, P)$, abgekürzt $(\Omega, P)^n$, ist das Paar $\left(\Omega^n, P^{(n)}\right)$, das durch

$$\begin{aligned}\Omega^n &= \{\boldsymbol{\omega} \mid \omega_1, \omega_2, \ldots, \omega_n \in \Omega\},\\ P^{(n)}(\boldsymbol{\omega}) &= P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_n)\end{aligned}$$

gegeben ist.

Tatsächlich ist auch im Produktraum die Summe der Wahrscheinlichkeiten wieder 1.

PROPOSITION 2.16. *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, und sei $n \in \mathbb{N}$. Dann gilt*

$$\sum_{\boldsymbol{\omega} \in \Omega^n} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_n) = 1. \tag{2.6}$$

*Beweis:* Wir beweisen die Gleichheit (2.6) mit Induktion nach $n$. Für $n = 1$ folgt (2.6) unmittelbar aus der Tatsache, dass sich in $\Omega$ die Maße der Elemente zu 1 addieren. Sei nun $n \in \mathbb{N}$, $n \geq 2$. Es gilt

$$\begin{aligned}
&\sum_{\boldsymbol{\omega} \in \Omega^n} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_n) \\
&\qquad = \sum_{\omega_n \in \Omega} \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}) \cdot P(\omega_n) \\
&\qquad = \sum_{\omega_n \in \Omega} P(\omega_n) \cdot \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}).
\end{aligned}$$

Wir verwenden die Induktionsvoraussetzung und erhalten

$$\sum_{\omega_n \in \Omega} P(\omega_n) \cdot \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}) = \sum_{\omega_n \in \Omega} P(\omega_n) \cdot 1.$$

Da $\Omega$ ein Wahrscheinlichkeitksraum ist, ist die Summe der Maße der Elemente in $\Omega$ gleich 1. $\square$

Wir überlegen uns, wie groß die Wahrscheinlichkeit ist, bei dreimaligem Würfeln zumindest einen Sechser zu würfeln. Wir setzen

$$\begin{aligned}
\Omega &= \{1, 2, 3, 4, 5, 6\} \\
\Omega^3 &= \{(1, 1, 1), (1, 1, 2), \ldots, (6, 6, 6)\}, \\
P^{(3)}((\omega_1, \omega_2, \omega_3)) &= \left(\tfrac{1}{6}\right)^3 = \tfrac{1}{216}.
\end{aligned}$$

Die Zufallsvariable $X$, die uns interessiert, definieren wir so:

$$X((\omega_1, \omega_2, \omega_3)) := \text{ die Anzahl der Sechser in } (\omega_1, \omega_2, \omega_3).$$

Dann suchen wir also

$$P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) \geq 1\}\right).$$

Da die Summe der Maße aller Elemente von $\Omega^3$ gleich 1 ist, kann man diese Wahrscheinlichkeit auch durch den Ausdruck

$$1 - P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) = 0\}\right)$$

berechnen. Klarerweise ist $X(\omega)$ genau dann 0, wenn $\omega$ in $\{1,2,3,4,5\}^3$ liegt. Es gilt also

$$\begin{aligned} 1 - P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) = 0\}\right) &= 1 - P^{(3)}\left(\{1,2,3,4,5\}^3\right) \\ &= 1 - 5^3 \cdot \frac{1}{216} \\ &\approx 0.4213 \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass mindestens ein Sechser gewürfelt wird, ist also ungefähr 42%.

Wir beschäftigen uns jetzt mit der Frage, wie es bei längerem Würfeln mit der Anzahl der Sechser aussieht. Man sollte ja erwarten, dass etwa ein Sechstel der Würfe Sechser werden. Ein mathematische Aussage darüber finden wir im "schwachen Gesetz der großen Zahlen". Zuvor brauchen wir aber noch ein harmloses Lemma.

LEMMA 2.17 (Čebyšëv, Tschebyschow, Tschebyscheff). *Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, und sei $\varepsilon \in \mathbb{R}$ mit $\varepsilon > 0$. Dann gilt*

$$P(\{\omega \mid |X(\omega)| \geq \varepsilon\}) \leq \frac{E(X^2)}{\varepsilon^2}.$$

Wir berechnen $E(X^2)$.

$$\begin{aligned} E\left(X^2\right) &= \sum_{\omega \in \Omega} P(\omega)\,(X(\omega))^2 \\ &= \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega)\,(X(\omega))^2 + \sum_{\substack{\omega \in \Omega \\ |X(\omega)| < \varepsilon}} P(\omega)\,(X(\omega))^2 . \end{aligned}$$

Nun wissen wir, dass die Summanden in der ersten Summe mindestens die Größe $P(\omega) \cdot \varepsilon^2$ haben, und dass die Summanden der zweiten Summe positiv sind. Insgesamt erhalten wir

$$\begin{aligned} \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega)\,(X(\omega))^2 + \sum_{\substack{\omega \in \Omega \\ |X(\omega)| < \varepsilon}} P(\omega)\,(X(\omega))^2 &\geq \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega)\,\varepsilon^2 + 0 \\ &= \varepsilon^2 \cdot \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega) = \varepsilon^2 \cdot P(\{\omega \mid |X(\omega)| \geq \varepsilon\}) . \end{aligned}$$

Daher gilt

$$\frac{E(X^2)}{\varepsilon^2} \geq P(\{\omega \mid |X(\omega)| \geq \varepsilon\}) .$$

$\square$

SATZ 2.18 (Schwaches Gesetz der großen Zahlen). *Sei* $(\Omega, P)$ *ein endlicher Wahrscheinlichkeitsraum, sei* $X : \Omega \to \mathbb{R}$ *eine Zufallsvariable, und sei* $\varepsilon > 0$. *Sei* $E_n$ *gegeben durch*

$$E_n := P^{(n)}\left(\{\boldsymbol{\omega} \mid \left| \left(\frac{1}{n}\sum_{i=1}^{n} X(\omega_i)\right) - E(X) \right| \geq \varepsilon\}\right).$$

*Dann gilt*

$$E_n \leq \frac{V(X)}{\varepsilon^2 \cdot n}.$$

Die Wahrscheinlichkeit, dass bei $n$ Würfen zwischen 16% und 17% Sechser fallen, konvergiert also gegen 1.

*Beweis:* Sei $n \in \mathbb{N}$. Wir definieren auf dem Produktraum $(\Omega, P)^n$ die Zufallsvariablen

$$\begin{array}{rcl}
X_1(\boldsymbol{\omega}) & := & X(\omega_1) \\
X_2(\boldsymbol{\omega}) & := & X(\omega_2) \\
& \vdots & \\
X_i(\boldsymbol{\omega}) & := & X(\omega_i) \text{ für } i \in \{1, 2, \ldots, n\} \\
& \vdots & \\
X_n(\boldsymbol{\omega}) & := & X(\omega_n)
\end{array}$$

und

$$Y(\boldsymbol{\omega}) := \left(\frac{1}{n}\sum_{i=1}^{n} X(\omega_i)\right) - E(X).$$

Wir beobachten zunächst, dass für alle $i$ die Gleichheiten $E(X_i) = E(X)$ und $V(X_i) = V(X)$ gelten. Außerdem sind $X_i$ und $X_j$ für $i \neq j$ unabhängig. Wir berechnen nun den Erwartungswert von $Y$. Es gilt

$$\begin{aligned}
E(Y) &= E\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \overline{E(X)}\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} E(X_i)\right) - E\left(\overline{E(X)}\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} E(X)\right) - E(X) \\
&= \frac{1}{n} \cdot n \cdot E(X) - E(X) \\
&= E(X) - E(X) \\
&= 0.
\end{aligned}$$

Als nächstes berechnen wir den Erwartungswert von $Y^2$. Es gilt

$$\begin{aligned}
E\left(Y^2\right) &= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}(\boldsymbol{\omega}) \cdot (Y(\boldsymbol{\omega}))^2 \\
&= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}(\boldsymbol{\omega}) \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i(\omega) - E(X))\right)^2 \\
&= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}(\boldsymbol{\omega}) \cdot \frac{1}{n^2} \cdot \left(\left(\sum_{i=1}^{n} X_i(\omega)\right) - \left(\sum_{i=1}^{n} E(X_i)\right)\right)^2 \\
&= \frac{1}{n^2} \cdot \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}(\boldsymbol{\omega}) \cdot \left(\left(\sum_{i=1}^{n} X_i(\omega)\right) - \left(\sum_{i=1}^{n} E(X_i)\right)\right)^2 \\
&= \frac{1}{n^2} \cdot E\left(\left(\sum_{i=1}^{n} X_i - \overline{E\left(\sum_{i=1}^{n} X_i\right)}\right)^2\right) \\
&= \frac{1}{n^2} \cdot V\left(\sum_{i=1}^{n} X_i\right).
\end{aligned}$$

Für $i \neq j$ sind die Zufallsvariablen $X_i$ und $X_j$ unabhängig. Satz 2.14 ergibt also

$$\begin{aligned}
\frac{1}{n^2} \cdot V\left(\sum_{i=1}^{n} X_i\right) &= \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} V(X) \\
&= \frac{1}{n^2} \cdot n \cdot V(X) \\
&= \frac{1}{n} V(X).
\end{aligned}$$

Aus dem Lemma von Čebyšëv, Lemma 2.17, wissen wir, dass

$$P^{(n)}(\{\boldsymbol{\omega} \mid |Y(\boldsymbol{\omega})| \geq \varepsilon\}) \leq \frac{E(Y^2)}{\varepsilon^2}.$$

Es gilt also

$$P^{(n)}\left(\{\boldsymbol{\omega} \mid \left|\left(\frac{1}{n}\sum_{i=1}^{n} X(\omega_i)\right) - E(X)\right| \geq \varepsilon\}\right) \leq \frac{\frac{1}{n}V(X)}{\varepsilon^2},$$

und somit

$$E_n \leq \frac{V(X)}{\varepsilon^2 \cdot n}.$$

## 5. Konstruktion von Zufallsvariablen

Wir konstruieren nun aus der Zufallsvariable $X$, die die Auswahl eines Zeichens auf Basis eines Wurfes mit einem Würfel beschreibt, eine Zufallsvariable $X^n$, die die Auswahl von $n$ Zeichen durch $n$-maliges Würfeln beschreibt.

Definition 2.19. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, sei $A$ eine Menge, sei $n \in \mathbb{N}$ und sei $X : \Omega \to A$ eine Zufallsvariable. Mit $X^n$ bezeichnen wir die Zufallsvariable, die durch

$$\begin{array}{rcrcl} X^n & : & \Omega^n & \longrightarrow & A^n \\ & & \boldsymbol{\omega} & \longmapsto & (X(\omega_1), X(\omega_2), \ldots, X(\omega_n)) \end{array}$$

definiert ist.

Die folgende Konstruktion brauchen wir, um die Ergebnisse zweier Zufallsvariablen gemeinsam zu untersuchen.

Definition 2.20. Sei $(\Omega, P)$ ein Wahrscheinlichkeitsraum, seien $M, N$ Mengen, und seien $X : \Omega \to M$, $Y : \Omega \to N$ Zufallsvariablen. Dann definieren wir die Zufallsvariable $X \otimes Y$ durch

$$\begin{array}{rcrcl} X \otimes Y & : & \Omega & \longrightarrow & M \times N \\ & & \omega & \longmapsto & (X(\omega), Y(\omega)). \end{array}$$