Beachte, dass jede Indikatorvariable X_A eine Bernoulli-Variable mit der Erfolgswahrscheinlichkeit $\Pr(X_A = 1) = \Pr(A)$ ist. Somit kann man die Ereignisse als einen Spezialfall der Zufallsvariablen – nämlich als 0-1-wertige Zufallsvariablen – betrachten.

12.1 Erwartungswert und Varianz

Hat man eine Zufallsvariable $X : \Omega \to S$ mit dem Bildbereich $S = X(\Omega)$, so will man die Wahrscheinlichkeiten $\Pr(X \in R)$ für verschiedene Teilmengen $R \subseteq S$ bestimmen (oder zumindest abschätzen). Als Ausgangspunkt betrachtet man dazu zwei numerische Charakteristiken der Zufallsvariable X – ihren »Erwartungswert« und ihre »Varianz«.

Definition:
Der *Erwartungswert* $\mathrm{E}(X)$ von $X : \Omega \to S$ ist definiert durch

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega) \cdot \Pr(\omega)\,.$$

D. h. wir multiplizieren die Werte, die X annehmen kann, mit den entsprechenden Wahrscheinlichkeiten, und summieren die Terme auf. Der Erwartungswert ist also ein »verallgemeinerter Durchschnittswert«.

Beobachtet man, dass die Mengen $X^{-1}(a)$ mit $a \in S$ eine disjunkte Zerlegung des Wahrscheinlichkeitsraumes Ω bilden und

$$\Pr(X = a) = \sum_{\omega \in X^{-1}(a)} \Pr(\omega)$$

gilt, so erhält man eine äquivalente Definition von $\mathrm{E}(X)$:

$$\mathrm{E}(X) = \sum_{a \in S} a \cdot \Pr(X = a)\ .$$

Im Spezialfall, wenn der Wertebereich $S = \{a_1, \ldots, a_n\}$ endlich ist und X jeden Wert a_i mit gleicher Wahrscheinlichkeit $1/n$ annimmt, ist $\mathrm{E}(X)$ einfach das arithmetische Mittel

$$\mathrm{E}(X) = \frac{a_1 + \cdots + a_n}{n}.$$

Man kann den Erwartungswert auch rein mechanisch interpretieren. Wenn wir n Objekte mit Gewichten $p_i = \Pr(X = a_i)$ auf der x-Achse in der Positionen a_i $(i = 1, \ldots, n)$ ablegen, dann wird der Schwerpunkt genau an der Stelle $\mathrm{E}(X)$ sein (siehe Bild 12.1).

Falls die Zufallsvariable *unendlich* viele Werte $a_1, a_2, \ldots$ annehmen kann, dann ist der Erwartungswert als

$$\mathrm{E}(X) := \lim_{n \to \infty} \sum_{i=1}^{n} a_i \Pr(X = a_i) = \sum_{i=1}^{\infty} a_i \Pr(X = a_i)$$

definiert. Im Allgemeinen muss dieser Grenzwert nicht existieren. Ist aber (a_i) eine monoton fallende Nullfolge, dann existiert der Grenzwert nach dem Dirichlet-Kriterium (Satz 9.27), da die Partialsummen $\sum_{i=1}^{n} \Pr(X = a_i) \le 1$ beschränkt sind.

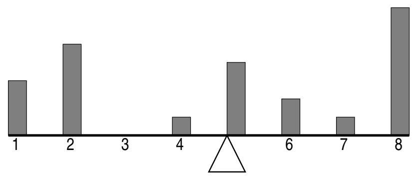

Bild 12.1: Erwartungswert als Schwerpunkt.

Beispiel 12.4: Das »St. Petersburg-Paradoxon«

Sei X eine Zufallsvariable mit der Verteilung $\Pr\left(X = 2^k\right) = 1/2^k$ für alle $k = 1,2,\ldots$. Das ist eine legale Wahrscheinlichkeitsverteilung, da

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = \sum_{k=0}^{\infty} \frac{1}{2^k} - 1 = \frac{1}{1-(1/2)} - 1 = 1$$

gilt. Die Zufallsvariable X beschreibt zum Beispiel den Gewinn in dem folgenden Kasinospiel: Wir werfen eine faire 0-1 Münze bis erstmals eine Eins rauskommt; kommt die Eins in der k-ten Runde, so erhalten wir 2^k Euro ausgezahlt und das Spiel ist zu Ende. Der Gewinn richtet sich also nach der Anzahl der Münzwürfe insgesamt. War es nur einer, dann erhält der Spieler 2 Euro. Bei zwei Würfen (also Null, dann Eins) erhält er 4 Euro, bei drei Würfen 8 Euro, bei vier Würfen 16 Euro und bei jedem weiteren Wurf verdoppelt sich der Betrag.

Natürlich verlangt das Kasino vorher einen Teilnahmebetrag B und hofft, dass die Eins mit einer großen Wahrscheinlichkeit viel früher als in $k \leq \log_2 B$ Runden kommt; dann kassiert es die verbleibenden $B - 2^k$ Euro. Welchen Geldbetrag würde man für die Teilnahme an diesem Spiel bezahlen wollen?

Man kommt genau dann zum k-ten Wurf, wenn man man vorher $(k-1)$-mal 0 geworfen hat. Also ist die Wahrscheinlichkeit, dass das erste Mal beim k-ten Münzwurf 1 fällt, gleich $\left(\frac{1}{2}\right)^k = 2^{-k}$. Nach (9.2) mit $x = 1/2$ beträgt die erwartete Spieldauer nur

$$\sum_{k=1}^{\infty} k 2^{-k} = \sum_{k=1}^{\infty} k x^k = \frac{x}{(1-x)^2} = \frac{1/2}{(1-1/2)^2} = 2$$

Runden. Wieviel kann man im Durchschnitt erwarten zu gewinnen? Mit Wahrscheinlichkeit $1/2$ ist der Gewinn 2 Euro, mit Wahrscheinlichkeit $1/4$ ist er 4 Euro, mit Wahrscheinlichkeit $1/8$ ist er 8 Euro, usw. Der Erwartungswert ist daher

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} 2^{-k} \cdot 2^k = \sum_{k=1}^{\infty} 1 = \infty\,,$$

also unendlich! Sollte man die Entscheidung nach dem Erwartungswert treffen, könnte man daher jede beliebige Teilnahmegebühr akzeptieren. Dies widerspricht natürlich einer tatsächlichen Entscheidung, und scheint auch irrational zu sein, da man in der Regel nur einige Euro gewinnt. Dieses Paradoxon hat Daniel Bernoulli im Jahre 1738 entdeckt. Versuche, dieses Paradox aufzulösen, haben zu verschiedenen

Tabelle 12.1: Endliche Versionen des Kasinospiels.

Kasinokapital K	N	$\mathrm{E}(X)$	
100 €	6	7 €	Spiel unter Freunden
100 Millionen €	26	27 €	Spielkasino
100 Milliarden €	36	37 €	Haushalt eines (reichen) Landes

Theorien in der Ökonomie geführt. Hier betrachten wir die einfachste »Lösung«.

Das Unrealistische an dem Paradox ist, dass das Spiel *unendlich* lange laufen kann und die Gewinne unendlich hoch werden können. In der Praxis ist beides jedoch nicht möglich. Der Spieler kann nicht unendlich lange eine Münze werfen (klar) und das Kasino kann nicht unendlich hohe Gewinne ausgeben, da das Kapital K des Kasinos beschränkt ist. Daher kann das Kasino nur $N = \lfloor \log_2 K \rfloor$ Runden den Gewinn verdoppeln: Wird das Spiel länger als N Runden dauern, so wird jedenfalls nur $2^N = K$ € ausgezahlt. Der Erwartungswert eines solchen Spiels berechnet sich wie folgt:

$$\begin{aligned}\mathrm{E}(X) &= \sum_{k=1}^{N} 2^{-k} 2^k + 2^N \sum_{k=N+1}^{\infty} 2^{-k} = N + 2^N \Big(\sum_{k=1}^{\infty} 2^{-k} - \sum_{k=1}^{N} 2^{-k} \Big) \\ &= N + 2^N \Big(1 - \Big(1 - 2^{-N}\Big)\Big) = N + 1\,.\end{aligned}$$

Während das Kapital $K = 2^N$ des Kasinos exponentiell erhöht wird, steigt der erwartete Gewinn nur linear. Man müsste also ein enorm großes Kapital des Kasinos annehmen, um auf hohe Gewinnerwartungen zu kommen (siehe Tabelle 12.1). Würde also ein Kasino mehr als 30 € als Teilnahmebetrag verlangen, dann sollten wir am besten ein anderes Kasino aufsuchen.

Die allerwichtigste Eigenschaft des Erwartungswertes überhaupt ist seine *Linearität*. Diese Eigenschaft ist sehr robust: Sie gilt für *beliebige* (nicht nur für unabhängige) Zufallsvariablen!

Satz 12.5: Linearität des Erwartungswertes

Seien X, Y Zufallsvariablen und a, b beliebige reelle Zahlen. Dann gilt

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y)\,.$$

Da X und Y *beliebige* Zufallsvariablen sind, kann man diese Eigenschaft für mehrere Zufallsvariablen $X_1, \dots, X_n$ erweitern:

$$\mathrm{E}(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1\,\mathrm{E}(X_1) + a_2\,\mathrm{E}(X_2) + \cdots + a_n\,\mathrm{E}(X_n)\,.$$

Beweis:

$$\begin{aligned}\mathrm{E}(aX+bY) &= \sum_{\omega\in\Omega}(aX(\omega)+bY(\omega))\Pr(\omega)\\ &= a\sum_{\omega\in\Omega}X(\omega)\Pr(\omega)+b\sum_{\omega\in\Omega}Y(\omega)\Pr(\omega) = a\,\mathrm{E}(X)+b\,\mathrm{E}(Y)\,.\end{aligned}$$

□

Ist $f(x)$ eine *nicht* lineare Funktion, so gilt $\mathrm{E}(f(X)) = f(\mathrm{E}(X))$ im Allgemeinen nicht! Das zu »behaupten« ist ein sehr häufiger Fehler. Ist zum Beispiel $f(x) = x^2$ und X eine Indikatorvariable mit $\Pr(X=1) = 1/2$, dann haben wir $\mathrm{E}(f(X)) = \mathrm{E}(X) = 1/2$ und $f(\mathrm{E}(X)) = (1/2)^2 = 1/4$.

Beispiel 12.6: Zufällige Teilmengen

Sei N eine endliche Menge mit $|N| = n$ Elementen. Wir wollen eine zufällige Teilmenge $S \subseteq N$ erzeugen, zu der jedes Element $x \in N$ mit Wahrscheinlichkeit p gehört. Dazu nehmen wir eine Münze, bei der die Wahrscheinlichkeit für den Ausgang »Wappen« gleich p ist. Wir werfen für jedes potenzielle Element $x \in N$ diese Münze und nehmen das Element x in die Menge S auf, wenn das Ergebnis »Wappen« ist. Somit gilt $\Pr(x \in S) = p$ für jedes $x \in N$. Die zufällige Wahl der Menge $S \subseteq N$ entspricht also der n-maligen Wiederholung eines Bernoulli-Experiments mit der Erfolgswahrscheinlichkeit p und $|S|$ ist dann genau die Anzahl der Erfolge. Um die erwartete Größe der Menge S zu bestimmen, sei I_x die Indikatorvariable für das Ereignis »$x \in S$«. Nach der Linearität des Erwartungswertes gilt dann

$$\mathrm{E}(|S|) = \mathrm{E}\Big(\sum_{x\in N} I_x\Big) = \sum_{x\in N}\mathrm{E}(I_x) = \sum_{x\in N}\Pr(x\in S) = pn\,.$$

Ist nun eine Teilmenge $T \subseteq N$ gegeben, was kann man über die erwartete Größe des Schnitts $S \cap T$ sagen? Diese Frage ist wegen der Linearität des Erwartungswertes wiederum leicht zu beantworten:

$$\mathrm{E}(|S\cap T|) = \mathrm{E}\Big(\sum_{x\in T} I_x\Big) = \sum_{x\in T}\mathrm{E}(I_x) = \sum_{x\in T}\Pr(x\in S) = p|T|\,.$$

Die Linearität des Erwartungswertes kann man *nicht* ohne weiteres auf unendlich vielen Zufallsvariablen $X_1, X_2, \ldots$ erweitern. Dazu muss die Reihe $\sum_{i=0}^{\infty}\mathrm{E}(|X_i|)$ konvergieren. Es gilt nämlich:

Satz 12.7: Unendliche Linearität des Erwartungswertes

Seien $X_0, X_1, \ldots$ Zufallsvariablen. Konvergiert die Reihe $\sum_{i=0}^{\infty}\mathrm{E}(|X_i|)$, so gilt

$$\mathrm{E}(X_0 + X_1 + \cdots) = \mathrm{E}(X_0) + \mathrm{E}(X_1) + \cdots\,.$$

Beispiel 12.8: Kasino

Wir spielen in einem Kasino ein Spiel mit Gewinnwahrscheinlichkeit $p = 1/2$. Wir werfen zum Beispiel eine faire 0-1 Münze und wir gewinnen, falls 1 kommt. Wir können einen beliebigen Betrag einsetzen. Geht das Spiel zu unseren Gunsten aus, erhalten wir den Einsatz zurück und zusätzlich denselben Betrag aus der Bank. Endet das Spiel ungünstig, verfällt unser Einsatz.

Wir betrachten die folgende Strategie: In jedem Schritt *verdoppeln* wir unseren Einsatz bis erstmals 1 kommt; dann hören wir auf. Wir wollen den erwarteten Gewinn dieser Strategie bestimmen. Sei K unser erster Einsatz und sei X_i das im i-ten Schritt gewonnene Kapital. Dann ist $Y = \sum_{i=0}^{\infty} X_i$ das (am Ende des Spiels) von uns gewonnene Kapital.

Da in jedem Schritt die Gewinnchance $p = 1/2$ ist, werden wir im i-ten Schritt mit gleicher Wahrscheinlichkeit entweder $K \cdot 2^{i-1}$ Euro gewinnen oder denselben Betrag verlieren, d. h. der Gewinn im i-ten Schritt ist entweder positiv ($X_i = +K2^{i-1}$) oder negativ ($X_i = -K2^{i-1}$). Deshalb ist der erwartete Gewinn $\mathrm{E}(X_i) = 0$ für alle $i = 1,2,\ldots$ gleich Null und man könnte daraus »schließen«, dass wir keinen Gewinn erwarten sollten:

$$\mathrm{E}(Y) = \mathrm{E}\Big(\sum_{i=1}^{\infty} X_i\Big) \overset{(*)}{=} \sum_{i=1}^{\infty} \mathrm{E}(X_i) = \sum_{i=1}^{\infty} 0 = 0\,.$$

Aber die Gewinnwahrscheinlichkeit ist in jedem Schritt positiv, also muss die Münze mit Sicherheit irgendwann auf 1 landen. D. h. wir sollten mit Wahrscheinlichkeit 1 mindestens K Euro gewinnen. Was war dann hier falsch? Unsere Argumentation, dass $\mathrm{E}(X_i) = 0$ für alle i gilt, war richtig. Der Fehler steckt aber in der »Gleichung« $(*)$, da die Reihe $\sum_{i=1}^{\infty} \mathrm{E}(|X_i|)$ nicht konvergent ist: Es gilt $|X_i| = K \cdot 2^{i-1}$ mit Wahrscheinlichkeit 2^{-i} und deshalb gilt auch:

$$\sum_{i=1}^{\infty} \mathrm{E}(|X_i|) = \sum_{i=1}^{\infty} = K \cdot 2^{i-1} \cdot 2^{-i} = \frac{1}{2} \sum_{i=1}^{\infty} K = \infty\,.$$

Um den erwarteten Gewinn $\mathrm{E}(Y)$ doch zu bestimmen, schauen wir das Problem genauer an. Unser Wahrscheinlichkeitsraum Ω besteht aus allen 0-1 Vektoren der Form

$$\omega = 0^{k-1}1 = \overbrace{0\cdots0}^{k-1 \text{ mal}}\ 1 \quad (k-1 \text{ Nullen gefolgt von einer Eins})\,.$$

Jeder solche Vektor entspricht einem möglichen Verlauf des Spiels: Eine Eins erst im k-ten Schritt. Bezeichnet nun X_i das im i-ten Schritt gewonnene Kapital, so gilt $X_i(0^{k-1}1) = K \cdot 2^k$ für $i = k$, und $X_i(0^{k-1}1) = -K \cdot 2^i$ für $i < k$; für $i > k$ können wir o. B. d. A. $X_i(0^{k-1}1) = 0$ setzen (das Spiel war bereits früher beendet). Daher ist auf *jedem* Elementarereignis $\omega = 0^{k-1}1$ der Wert von $Y = X_1 + X_2 + \cdots$ gleich

$$Y(\omega) = K \cdot 2^k - K \cdot \sum_{i=1}^{k-1} 2^i = 2K \qquad \text{geometrische Reihe}$$

und somit muss auch der Erwartungswert von Y gleich $2K$ sein.

Wenn die Zufallsvariable X nur *natürliche* Zahlen als Werte annimmt, gibt es eine alternative (und oft geeignetere) Art und Weise den Erwartungswert $\mathrm{E}(X)$ zu bestimmen.

Satz 12.9: Erwartungswert diskreter Zufallsvariablen

Ist $X : \Omega \to \mathbb{N}$ eine Zufallsvariable mit endlichem Erwarungswert, so gilt

$$\mathrm{E}(X) = \sum_{k=0}^{\infty} \Pr(X > k) \; .$$

Beweis:

Da X nur ganze Zahlen 0,1,2,... als Werte annimmt, gilt

$$\Pr(X > k) = \Pr(X = k+1) + \Pr(X = k+2) + \Pr(X = k+3) + \cdots$$

und deshalb gilt auch

$$\begin{aligned}
\sum_{k=0}^{\infty} \Pr(X > k) &= \Pr(X > 0) + \Pr(X > 1) + \Pr(X > 2) + \Pr(X > 3) + \cdots \\
&= \underbrace{\Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) + cldots}_{\Pr(X>0)} \\
&\qquad + \underbrace{\Pr(X = 2) + \Pr(X = 3) + \cdots}_{\Pr(X>1)} \\
&\qquad\qquad + \underbrace{\Pr(X = 3) + \cdots}_{\Pr(X>2)} \\
&= \Pr(X = 1) + 2 \cdot \Pr(X = 2) + 3 \cdot \Pr(X = 3) + \cdots \\
&= \sum_{k=1}^{\infty} k \cdot \Pr(X = k) = \mathrm{E}(X) \, .
\end{aligned}$$

□

Beispiel 12.10:

Wir haben ein Kommunikationsnetz, in dem viele Pakete verschickt werden sollen. Angenommen der Versand eines Pakets kann sich nur mit Wahrscheinlichkeit $1/k$ um k oder mehr Sekunden verzögern. Das klingt gut: Es ist nur 1% Chance, dass der Versand eines Pakets um 100 oder mehr Sekunden verzögert wird. Aber wenn wir die Situation genauer betrachten, ist das Netz gar nicht so gut. Tatsächlich ist die erwartete Verzögerung eines Pakets unendlich! Sei X die Verzögerung eines Pakets. Dann gilt nach Satz 12.9

$$\mathrm{E}(X) = \sum_{k=0}^{\infty} \Pr(X > k) \geq \sum_{k=0}^{\infty} \frac{1}{k+1} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty \qquad \text{harmonische Reihe} \, .$$

ENTWURF