Stochastische Unabhängigkeit und Gesetze großer Zahlen 2

Die Grundbegriffe der Wahrscheinlichkeitstheorie wie Wahrscheinlichkeitsraum, Zufallsvariable, Verteilung und Erwartungswert und Verteilungsfunktion werden auf maßtheoretischer Grundlage eingeführt und motiviert durch verschiedene Aufgabenstellungen. Grundlegende Themen dieses Kapitels sind der Begriff der stochastischen Unabhängigkeit, seine natürliche Verwendung zur Modellierung stochastischer Vorgänge und die fundamentalen Gesetze großer Zahlen, die einen Bezug zwischen Experimenten und den erklärenden Parametern des stochastischen Modells herstellen. Damit liefert die Wahrscheinlichkeitstheorie eine Brücke zwischen Empirie und Theorie, die wesentlich ihre Bedeutung für die Wissenschaften ausmacht.

2.1 Grundbegriffe der Wahrscheinlichkeitstheorie

In diesem einführenden Abschnitt werden zentrale Begriffe der Wahrscheinlichkeitstheorie eingeführt und Eigenschaften dieser Begriffe erläutert. Dies geschieht durch Spezialisierung und Interpretation von Begriffen und Aussagen der Maß- und Integrationstheorie. Einige Beispiele zeigen die Motivation dieser Begriffe und weiterführende Fragestellungen.

Definition 2.1.1 (Wahrscheinlichkeitsraum) *Sei $(\Omega, \mathcal{A})$ ein Messraum. Ein normiertes Maß P auf $(\Omega, \mathcal{A})$, d. h., es gilt: $P(\Omega) = 1$, heißt* ***Wahrscheinlichkeitsmaß*** *auf $(\Omega, \mathcal{A})$. $(\Omega, \mathcal{A}, P)$ heißt* ***Wahrscheinlichkeitsraum****.*

Interpretation Ein Experiment habe als Grundraum möglicher Elementarereignisse Ω. $\mathcal{A}$ sei das System der interessierenden Ereignisse. P beschreibt eine Gewichtsfunktion, nach der gewichtet als Ergebnis des Experiments ein zufälliges Ergebnis ω erhalten wird. Das stochastische Modell $(\Omega, \mathcal{A}, P)$ ist also ähnlich einem Urnenmodell, aus dem zufällig mit P gewichtet eine Kugel entnommen wird.

L. Rüschendorf, *Wahrscheinlichkeitstheorie*, Springer-Lehrbuch Masterclass,
DOI 10.1007/978-3-662-48937-6_2

Definition 2.1.2 (Zufallsvariable und Verteilung) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{A}')$ ein Messraum.*

a) *Eine messbare Abbildung $X : (\Omega, \mathcal{A}) \longmapsto (\Omega', \mathcal{A}')$ heißt* ***Zufallsvariable****. X heißt reelle Zufallsvariable, wenn $(\Omega', \mathcal{A}') = (\mathbb{R}^1, \mathcal{B}^1)$ ist.*
b) *Das Bildmaß P^X von P unter X heißt* ***Verteilung*** *von X.*

Interpretation Als Ergebnis eines Zufallsexperiments im Modell $(\Omega, \mathcal{A}, P)$ ergibt sich ein zufälliges Element $\omega \in \Omega$. Von Interesse ist aber nur der Wert $X(\omega)$ einer Funktion X. $X(\omega)$ ist dann auch eine zufällige Variable, deren Verteilung in $(\Omega', \mathcal{A}')$ durch $P^X =: P_X$ beschrieben wird.

Definition 2.1.3 (Erwartungswert und Varianz) *Sei X eine reelle Zufallsvariable (ZVe) auf $(\Omega, \mathcal{A}, P)$.*

a) *Für $X \in \overline{\mathcal{Z}}_+ \cup \mathcal{L}^1(P)$ heißt* $EX := \displaystyle\int X \, dP = \int x \, dP^X(x)$ ***Erwartungswert*** *von X.*
b) *Für $X \in \mathcal{L}^1(P)$ heißt* $\operatorname{Var}(X) := E(X - EX)^2$ ***Varianz*** *von X.*
Es gilt: $\operatorname{Var}(X) = EX^2 - (EX)^2$.

$\sigma(x) := \sqrt{\operatorname{Var}(X)}$ *heißt* ***Streuung*** *von X.*
c) $EX^r = \int x^r \, dP^X(x)$ *heißt* ***Moment der Ordnung r*** *(falls existent).*

Interpretation Bei einem Spiel mit zufälligem Ausgang X ist EX der faire Einsatz. Ist P^X eine Masseverteilung, dann ist EX der Schwerpunkt. Die Varianz $\operatorname{Var}(X)$ ist die mittlere quadratische Schwankung um den Erwartungswert. EX und $\operatorname{Var}(X)$ sind zwei charakteristische Größen zur Beschreibung der Verteilung P^X von X.

Definition 2.1.4 (Kovarianz, Korrelationskoeffizient) *Seien X und Y reelle Zufallsvariable auf $(\Omega, \mathcal{A}, P)$.*

a) $\operatorname{Cov}(X, Y) := E(X - EX)(Y - EY)$ *heißt* ***Kovarianz*** *von X, Y (falls existent).*
b) $\varrho(X, Y) := \dfrac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$ *heißt* ***Korrelationskoeffizient*** *von X, Y (falls existent).*
c) *X, Y heißen* ***unkorreliert****, wenn* $\operatorname{Cov}(X, Y) = 0$.

Interpretation Die Kovarianz und der Korrelationskoeffizient beschreiben den Grad von linearer Abhängigkeit von X, Y. Diese Interpretation ergibt sich aus der folgenden Proposition.

Proposition 2.1.5 *Seien X, Y reelle Zufallsvariable.*

a) $\mathrm{Var}(X) = 0 \quad \Leftrightarrow \quad X = EX\,[P]$.
b) $|\varrho(X,Y)| \le 1$;
   $|\varrho(X,Y)| = 1 \quad \Leftrightarrow \quad \exists\, a, b \in \mathbb{R}^1 : Y = aX + b\,[P]$.
c) *Für $X, Y \in \mathcal{L}^2(P)$ gilt:*

$$E(X-a)^2 \ge E(X-EX)^2 = \mathrm{Var}(X), \quad \forall\, a \in \mathbb{R}^1.$$

d) *Für $X, Y \in \mathcal{L}^2(P)$ gilt:*

$$E(Y-(aX+b))^2 = \min_{a,b}$$

hat die Lösung

$$a^* := \varrho(X,Y)\frac{\sigma(X)}{\sigma(Y)}, \quad b^* := EY - a^*EX.$$

$g^*(x) = a^*x + b^*$ heißt **Regressionsgerade** von Y nach X.

Bemerkung 2.1.6 *Aussage b) in Proposition 2.1.5 ist eine Folgerung aus der Cauchy-Schwarz-Ungleichung. c) besagt, dass der Erwartungswert EX die beste Approximation von X durch eine Konstante im L^2-Sinne ist. Entsprechend ist nach d) $g^*(X)$, g^* die Regressionsgerade, die beste Vorhersage (Approximation) von Y durch eine lineare Funktion von X.* ⌟

2.2 Verteilungsfunktion und stochastische Unabhängigkeit

Verteilungsfunktionen erlauben eine einfache Beschreibung von Wahrscheinlichkeitsmaßen auf $\mathbb{R}^1$. Der Begriff der stochastischen Unabhängigkeit ist fundamental für die Modellbildung stochastischer Experimente. Wir behandeln die grundlegenden Rechenregeln für diesen Begriff.

Definition 2.2.1 (Verteilungsfunktion)
*Eine Abbildung $F : \mathbb{R} \longmapsto [0,1]$ heißt **Verteilungsfunktion** auf $\mathbb{R}^1$, wenn*

i) $F \uparrow$ *(d. h., F ist monton wachsend),*
ii) *F ist rechtsseitig stetig,*
iii) $\lim_{x\to\infty} F(x) = 1$ *und* $\lim_{x\to-\infty} F(x) = 0$, *$F$ ist normiert.*

Sei $M^1(\Omega, \mathcal{A})$ die Menge der Wahrscheinlichkeitsmaße auf (Ω, A) und sei $\mathcal{F} = \mathcal{F}^1$ die Menge der Verteilungsfunktionen auf $\mathbb{R}^1$.

Satz 2.2.2 (Korrespondenzsatz)

1. *Sei* $P \in M^1(\mathbb{R}^1, \mathcal{B}^1)$, *dann ist* $F_P(x) := P((-\infty, x])$ *für* $x \in \mathbb{R}$ *eine Verteilungsfunktion.*
2. *Sei* $F \in \mathcal{F}$, *dann existiert genau ein Wahrscheinlichkeitsmaß* $P \in M^1(\mathbb{R}^1, \mathcal{B}^1)$, *so dass* $F = F_P$.
3. *Die Abbildung* $\begin{array}{ccc} M^1(\mathbb{R}^1, \mathcal{B}^1) & \longrightarrow & \mathcal{F} \\ P & \longmapsto & F_P \end{array}$ *ist bijektiv.*

Beweis

1) i) Ist $x \le y$, dann folgt $F_P(x) = P((-\infty, x]) \le P((-\infty, y]) = F_P(y)$.

ii) Sei $x_n \downarrow x$, dann folgt nach dem Stetigkeitssatz für Maße:

$$F_P(x_n) = P((-\infty, x_n]) \downarrow P((-\infty, x]).$$

iii) Sei $x_n \uparrow \infty$, dann folgt:

$$F_P(x_n) = P((-\infty, x_n]) \uparrow P(\mathbb{R}^1) = 1;$$

für $x_n \downarrow -\infty$ folgt

$$F_P(x_n) = P((-\infty, x_n]) \downarrow P(\varnothing) = 0.$$

2) **Existenz.** Definiere für ein Intervall $I = (a, b]$ mit $a < b$,

$$P((a, b]) := F(b) - F(a). \tag{2.1}$$

Sei $A := \sum_{i=1}^n (a_i, b_i] \in \mathcal{F}^1$ eine Vereinigung paarweise disjunkter Intervalle, d. h. eine eindimensionale Figur, und definiere

$$P(A) := \sum_{i=1}^n (F(b_i) - F(a_i)).$$

Wie bei der Konstruktion des Lebesgue-Maßes ist P wohldefiniert auf $\mathcal{F}^1$, endlich additiv und ein Prämaß (kompakt approximierbar). Außerdem ist P σ-endlich auf $\mathcal{F}^1$. Nach dem Maßerweiterungssatz existiert daher eine Fortsetzung von P auf $\mathcal{B}^1 = \sigma(\mathcal{F}^1)$, und es ist $F = F_P$ nach (2.1).

Eindeutigkeit. Da $\mathcal{J}^1 = \{(-\infty, x];\ x \in \mathbb{R}\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{B}^1$ ist und $(-\infty, n] \uparrow \mathbb{R}^1$ mit $P((-\infty, n]) \le 1$, folgt aus dem Eindeutigkeitssatz die Eindeutigkeit der Maßfortsetzung.

3) folgt aus 1) und 2). □

Die Eigenschaften einer Verteilungsfunktion lassen sich in der Regel einfach nachprüfen, und daher erhalten wir nach dem Korrespondenzsatz eindeutige zugeordnete Wahrscheinlichkeitsmaße auf $\mathbb{R}^1$. Eine wichtige Klasse von Beispielen von Verteilungsfunktionen wird über Lebesgue-Dichten definiert.

Definition 2.2.3 *Eine messbare Abbildung $f : (\mathbb{R}^1, \mathcal{B}^1) \longrightarrow (\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+)$ heißt* ***Lebesgue-Dichte*** *auf $\mathbb{R}^1$, wenn sie normiert ist, d. h. wenn $\int f \, d\lambda^1 = 1$.*

Proposition 2.2.4 *Sei f eine Lebesgue-Dichte auf $\mathbb{R}^1$ und definiere $F(x) := \int_{(-\infty,x]} f \, d\lambda_1$, dann gilt:*

1) *$F \in \mathcal{F}$, F ist eine Verteilungsfunktion und das zugehörige Wahrscheinlichkeitsmaß $P = f\lambda^1$ ist das Maß mit der Dichte f bzgl. λ^1.*
2) *F ist stetig und P ist nicht-atomar (stetig), d. h., für alle $x \in \mathbb{R}^1$ ist $P(\{x\}) = 0$.*
3) *Ist f stetig in x_0, dann ist F in x_0 differenzierbar und es gilt: $F'(x_0) = f(x_0)$.*

Beweis

1) Aus der Definition und den Eigenschaften des Integrals folgt für alle x

$$P((-\infty, x]) = F(x) = \int_{(-\infty,x]} f \, d\lambda^1 = f\lambda^1((-\infty, x]).$$

Somit ist $P = f\lambda^1$ auf $\mathcal{J}^1$, und aus dem Eindeutigkeitssatz folgt: $P = f\lambda^1$, d. h.

$$P(A) = \int_A f \, d\lambda^1 \quad \text{für alle } A \in \mathcal{B}^1.$$

2) Es ist $P = f\lambda^1 \ll \lambda^1$.
Da $\lambda^1(\{x\}) = 0$, $\forall\, x \in \mathbb{R}^1$, folgt $P(\{x\}) = 0$, d. h., P ist stetig.
Behauptung: F ist stetig $\Leftrightarrow$ P ist stetig
„$\Rightarrow$“: $P(\{x\}) = F(x) - F(x-) = 0$, also folgt „$\Rightarrow$“
„$\Leftarrow$“: F ist rechtsseitig stetig nach dem Stetigkeitssatz

$$\lim_{h \downarrow 0} (F(x+h) - F(x)) = \lim_{h \downarrow 0} P((x, x+h]) = 0.$$

F ist auch linksseitig stetig, da wieder nach dem Stetigkeitssatz

$$\lim_{h \downarrow 0} \big(\underbrace{F(x) - F(x-h)}_{= P((x-h,x])} \big) = P(\{x\}) = 0.$$

3) Sei f stetig in x_0, dann existiert eine Umgebung $U = (x_0 - h, x_0 + h)$ von x_0 und $k \in \mathbb{R}^1$ mit $|f(y)| \le k$ für $y \in U$. Daraus folgt: für alle $x, y \in U$ mit $x < x_0 < y$ gilt

$$\frac{F(y) - F(x_0)}{y - x_0} = \frac{1}{y - x_0} \int_{(x_0,y]} f \, d\lambda^1 \xrightarrow[y \downarrow x_o]{} f(x_0)$$

und ebenso

$$\frac{F(x_0) - F(x)}{x_0 - x} \xrightarrow[x \uparrow x_0]{} f(x_0).$$

Falls f stetig ist, folgt die Behauptung auch direkt aus dem Hauptsatz der Differential- und Integralrechnung, d. h., $F(x) = \int_{-\infty}^{x} f(y)\,dy$ ist differenzierbar in x_0 und $F'(x_0) = f(x_0)$. □

Bemerkung 2.2.5

a) *Ist $P = f\lambda^1$, dann gilt für $F = F_P$, dass F' λ^1 fast sicher existiert und $F' = f[\lambda^1]$. Dies ist der Inhalt des **Lebesgue'schen Differentiationssatzes**.*

b) *Ist F stetig, so folgt nicht, dass das zugehörige Maß P Lebesgue-stetig ist. Es existiert $P \in M^1(\mathbb{R}^1, \mathcal{B}^1)$ mit $P \perp \lambda^1$, d. h., P und λ^1 haben disjunkte Träger so dass F_P stetig ist.*

c) ***Maße mit Dichten.** Ist μ ein Maß auf $(\mathbb{R}^1, \mathcal{B}^1)$ und $f \in \overline{\mathcal{Z}}_+ \cap \mathcal{L}^1(\mu)$ mit $\int f\,d\mu = 1$, dann definiert $F(x) = \int_{(-\infty,x]} f\,d\mu$ eine Verteilungsfunktion, $F \in \mathcal{F}^1$. Das zugehörige Wahrscheinlichkeitsmaß P ist das Maß mit Dichte f bzgl. μ, d. h. $P = f\mu$. Der Beweis ist analog zu dem von Proposition 2.2.4.*
Insbesondere durch die Wahl von $\mu = \lambda_A$ als abzählendes Maß einer abzählbaren Menge $A \subset \mathbb{R}^1$ erhält man die kanonischen diskreten Wahrscheinlichkeitsmaße. ⌟

Beispiel 2.2.6

a) ***Diskrete Maße***
   1. ***Bernoulli-Verteilung:*** $\mathcal{B}(1,\vartheta)$, $\vartheta \in [0,1]$
      $A = \{0,1\}$, Dichte $f(1) = \vartheta$, $f(0) = 1 - \vartheta$.
   2. ***Binomialverteilung:*** $\mathcal{B}(n,\vartheta)$
      $A = \{0,1,\dots,n\}$, Dichte $f_\vartheta(k) = \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$.
   3. ***Poisson-Verteilung:*** $\mathcal{P}(\lambda)$, $0 < \lambda < \infty$
      $A = \mathbb{N}_0$, Dichte $f_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}$.
   4. ***Geometrische Verteilung:*** $\mathcal{G}(\vartheta)$, $0 < \vartheta \le 1$
      $A = \mathbb{N}$, Dichte $f_\vartheta(n) = (1-\vartheta)^{n-1}\vartheta$.

b) ***Stetige Maße***
   Für $\mu = \lambda^1$ erhält man die stetigen Standardverteilungen wie z. B.
   1. ***Gleichverteilung:*** *$U_A = U(A)$ auf $A \subset \mathbb{R}^1$ mit $0 < \lambda^1(A) < \infty$*
      Dichte $f = \frac{1}{\lambda^1(A)}\mathbb{1}_A$,
   2. ***Normalverteilung:*** $N(\mu,\sigma^2)$, $\mu \in \mathbb{R}^1$, $\sigma^2 > 0$
      Dichte $\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$,
   3. ***Exponentialverteilung:*** $\mathcal{E}(\lambda)$, $\lambda > 0$
      Dichte $f_\lambda(x) = \lambda e^{-\lambda x}, \quad x \ge 0$. ◇

Beispiel 2.2.7 (Normalverteilung) *Sei $\varphi(x) := \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}^1$, die Dichte der **Standard-Normalverteilung** $P = \varphi\lambda^1 = N(0,1)$. Ist X eine normalverteilte Zufalls-*

variable auf $(\Omega, \mathcal{A}, P)$, *d. h.* $P^X = N(0,1)$, *dann gilt:*

$$EX = \int X\,dP = \int x\,dP^X(x) = \int x\varphi(x)\,dx = 0$$
$$\textit{und} \quad \operatorname{Var}(X) = EX^2 - (EX)^2 = \int x^2\varphi(x)\,dx = 1.$$

Für $X \sim N(\mu, \sigma^2)$ *gilt analog* $EX = \mu$, $\operatorname{Var} X = \sigma^2$. *Die allgemeinen Momente* α_k *für* $X \sim N(0,1)$ *erhält man mit Induktion und partieller Integration. Es ist*

$$\alpha_{2k+1} = 0, \quad \alpha_{2k} = EX^{2k} = (2k-1)(2k-3)\cdots 1,$$

speziell $\alpha_2 = 1, \alpha_4 = 3$ *und* $\alpha_6 = 15$. ◇

Der Begriff der stochastischen Unabhängigkeit wird zunächst für zwei Ereignisse A, B eingeführt. Dazu ist es nützlich, an die elementare bedingte Wahrscheinlichkeit zu erinnern.

Definition 2.2.8 *Sei* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum und* $A, B \in \mathcal{A}$, *dann heißt*

$$P(A \mid B) := \begin{cases} \frac{P(A\cap B)}{P(B)}, & P(B) > 0, \\ P(A), & \textit{sonst}, \end{cases}$$

die ***elementare bedingte Wahrscheinlichkeit****.*

Definition 2.2.9 (Stochastische Unabhängigkeit) *Sei* $(\Omega, \mathcal{A}, P)$ *ein Wahrscheinlichkeitsraum und* $A, B, C \in \mathcal{A}$.

a) A, B heißen **stochastisch unabhängig**

$$\begin{aligned} &\Leftrightarrow \; P(A \cap B) = P(A)P(B) \\ &\Leftrightarrow \; P(A \mid B) = P(A). \end{aligned}$$

b) *Die Ereignisse* A, B, C *heißen stochastisch unabhängig, wenn sie paarweise unabhängig sind und zusätzlich*

$$P(A \cap B \cap C) = P(A)P(B)P(C) \tag{2.2}$$

gilt.

Bemerkung 2.2.10

a) *Es gilt:* A, B *sind stochastisch unabhängig*

$$\Leftrightarrow \quad \mathbb{1}_A, \mathbb{1}_B \text{ sind unkorreliert,}$$

d. h. $\operatorname{Cov}(\mathbb{1}_A, \mathbb{1}_B) = 0$.

b) *Bedingung* (2.2) *ist äquivalent zu*

$$P(A \cap B \mid C) = P(A \cap B) = P(A)P(B)$$

oder äquivalent zu

$$P(B \cap C \mid A) = P(B \cap C) \quad \text{oder auch zu} \quad P(A \cap C \mid B) = P(A \cap C). \quad \lrcorner$$

Für allgemeinere Mengensysteme wird der Unabhängigkeitsbegriff durch eine Produkteigenschaft definiert.

Definition 2.2.11

a) *Ein System* $(A_i)_{i \in I} \subset \mathcal{A}$ *heißt stochastisch unabhängig, wenn*

$$\forall\, J \subset I \text{ endlich} : P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

b) *Seien* $\mathcal{E}_i \subset \mathcal{A}$ *für* $i \in I$. $(\mathcal{E}_i)_{i \in I}$ *heißt* ***stochastisch unabhängig***, *wenn*

$$\forall\, A_i \in \mathcal{E}_i, \forall\, i \in I \text{ gilt} : (A_i)_{i \in I} \text{ ist stochastisch unabhängig.}$$

c) *Seien* $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$, $i \in I$, *Zufallsvariablen, dann heißt* $(X_i)_{i \in I}$ ***stochastisch unabhängig***, *wenn* $(\sigma(X_i))_{i \in I}$ *stochastisch unabhängig ist.*

Bemerkung 2.2.12 $(\mathcal{E}_i)_{i \in I}$ *ist stochastisch unabhängig genau dann, wenn für alle* $J \subset I$ *endlich* $(\mathcal{E}_j)_{j \in J}$ *stochastisch unabhängig ist. Die stochastische Unabhängigkeit ist also eine Eigenschaft endlicher Teilsysteme.* $\lrcorner$

Beispiel 2.2.13 Geometrische Verteilung. Die Wartezeit W auf den ersten Erfolg einer unabhängigen Bernoulli-Folge (X_n), $X_n \sim \mathcal{B}(1, \vartheta)$ mit Erfolgswahrscheinlichkeit ϑ (pro Versuch) ist geometrisch verteilt: $W \sim \mathcal{G}(\theta)$, d. h., es gilt

$$f_W(n) := P(W = n) = (1 - \theta)^{n-1}\theta, \quad n \in \mathbb{N}, \text{ und es gilt} : EW = \frac{1}{\theta}.$$

Da $W = \inf\{n \in \mathbb{N} : X_n = 1\}$, gilt wegen der Unabhängigkeit der (X_n):

$$\begin{aligned} P(W = n) &= P(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdots P(X_{n-1} = 0) P(X_n = 1) = (1 - \vartheta)^{n-1}\vartheta. \end{aligned}$$

Die Formel $EW = \frac{1}{\vartheta}$ folgt elementar.

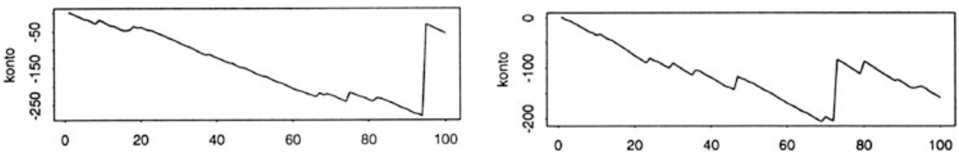

Abb. 2.1 Spielfolge, Einsatz 6 €

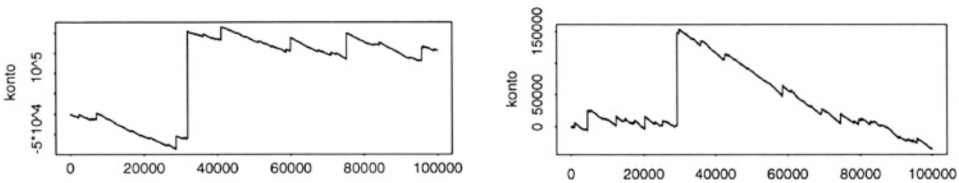

Abb. 2.2 Spielfolge, Einsatz 10 €

Als Anwendung ergibt sich das auch historisch interessante **Petersburger Paradoxon.** Eine faire Münze wird so lange geworfen, bis zum ersten Mal die Seite mit dem Wappen erscheint. Die Wartezeit X ist nach a) dann geometrisch verteilt, $X \sim \mathcal{G}(\frac{1}{2})$. Als Gewinn wird einem Spieler dieses Spiels

$$Y = 2^{X-1}$$

ausgezahlt. Die Frage ist nun: Was ist ein fairer Einsatz für dieses Spiel?

Es ist $P(X = n) = 2^{-n}$, $n \geq 1$ und nach a) ist $EX = 2$. Im Mittel dauert es nur zwei Würfe bis zum ersten Wappen. Die Wahrscheinlichkeit für lange Zahl-Serien nimmt schnell ab, z. B. ist $P(X \geq 11) = \frac{1}{1024}$. Es gilt aber

$$\begin{aligned} EY = E2^{X-1} &= \sum_{n=1}^{\infty} 2^{n-1} P(X = n) \\ &= \sum_{n=1}^{\infty} \frac{1}{2} = \infty. \end{aligned}$$

Dieses scheinbar paradoxe Resultat: im Mittel sind Serien kurz, aber der faire Einsatz $EY = \infty$, erklärt sich mit Hilfe von Simulationen. Bei einem Einsatz von 6 € und $n = 100$ Spielen tritt typischerweise (aber nicht immer) ein negativer Verlauf ein (vgl. Abb. 2.1). Ist n jedoch sehr groß, z. B. $n = 10.000$, dann gibt es typischerweise (aber nicht immer) eine lange Serie und einen beträchtlichen Gewinn (vgl. Abb. 2.2 für $n = 10.000$, Einsatz 10 €). $\diamond$

Die folgenden Rechenregeln über „Vergrößern“ und „Zusammenfassen“ stochastisch unabhängiger Systeme sind nützlich.

Satz 2.2.14 („Vergrößern“ und „Zusammenfassen“) *Sei $(\mathcal{E}_i)_{i\in I} \subset \mathcal{A}$ ein stochastisch unabhängiges System, dann gilt:*

1) $(\mathcal{D}(\mathcal{E}_i))_{i\in I}$ *ist stochastisch unabhängig.*
2) *Sind $\mathcal{E}_i$ $\cap$-stabil, $i \in I$, dann ist $(\sigma(\mathcal{E}_i))_{i\in I}$ stochastisch unabhängig.*
3) *Seien $\mathcal{E}_i$ $\cap$-stabil, $i \in I$, sei $I = \bigcup_{j\in K} I_j$ eine disjunkte Zerlegung von I und sei $\mathcal{A}_{I_i} := \sigma(\bigcup_{j\in I_i} \mathcal{E}_j)$. Dann sind die Systeme $(\mathcal{A}_{I_i})_{i\in K}$ stochastisch unabhängig.*

Beweis
1) Sei o. B. d. A. $|I| < \infty$ und für $i_0 \in I$ sei

$$\mathcal{D}_{i_0} := \{E \in \mathcal{A};\ \ ((\mathcal{E}_i)_{i\neq i_0}, \{E\}) \text{ stochastisch unabhängig}\}.$$

   Nach Voraussetzung gilt: $\mathcal{D}_{i_0} \supset \mathcal{E}_{i_0}$. Daraus folgt: $\mathcal{D}_{i_0} \supset \mathcal{D}(\mathcal{E}_{i_0})$.
   Also sind $((\mathcal{E}_i)_{i\neq i_0}, \mathcal{D}(\mathcal{E}_{i_0}))$ stochastisch unabhängig. Induktiv ergibt sich: $(\mathcal{D}(\mathcal{E}_i))_{i\in I}$ ist stochastisch unabhängig.
2) Die Behauptung folgt aus 1), da $\mathcal{D}(\mathcal{E}_i) = \sigma(\mathcal{E}_i)$.
3) Sei $\mathcal{A}_i := \sigma(\mathcal{E}_i)$, $i \in I$. Diese sind nach 2) stochastisch unabhängig.
   Für $j \in K$ definiere $\mathcal{B}_j := \{\bigcap_{i\in J} C_i;\ \ J \subset I_j \text{ endlich}, C_i \in \mathcal{A}_i, i \in J\}$.
   Es folgt $\mathcal{B}_j$ ist $\cap$-stabil, $(\mathcal{B}_j)_{j\in K}$ sind stochastisch unabhängig und $\mathcal{A}_{I_j} = \sigma(\mathcal{B}_j)$.
   Nach 2) folgt $(\mathcal{A}_{I_j})_{j\in K}$ sind stochastisch unabhängig. □

Korollar 2.2.15 *Seien $X_i : (\Omega, \mathcal{A}) \longmapsto (\Omega_i, \mathcal{A}_i)$ Zufallsvariable und $\mathcal{E}_i$ $\cap$-stabile Erzeuger von $\mathcal{A}_i$ mit $\Omega_i \in \mathcal{E}_i$, $1 \le i \le n$. Dann gilt:*

$$\begin{aligned} &(X_i)_{1\le i\le n} \textit{ sind stochastisch unabhängig} \\ \Leftrightarrow\ & P\left(\bigcap_{i=1}^{n}\{X_i \in E_i\}\right) = \prod_{i=1}^{n} P(\{X_i \in E_i\}), \quad \forall\, E_i \in \mathcal{E}_i. \end{aligned}$$

Beweis „$\Rightarrow$“: ist klar.
„$\Leftarrow$“: $\mathcal{B}_i := X_i^{-1}(\mathcal{E}_i)$ sind $\cap$-stabile Erzeuger von $\sigma(X_i)$ und $\Omega \in \mathcal{B}_i$. Daraus folgt: $(\mathcal{B}_i)_{1\le i\le n}$ sind stochastisch unabhängig. (Wähle einige $E_i = \Omega_i$.)
Daher gilt: $(\mathcal{D}(\mathcal{B}_i)) = (\sigma(\mathcal{B}_i)) = (\sigma(X_i))$ sind stochastisch unabhängig. □

Funktionen stochastisch unabhängiger Zufallsvariablen sind stochastisch unabhängig.

Lemma 2.2.16 *Seien $X_i : (\Omega, \mathcal{A}) \longrightarrow (\Omega_i, \mathcal{A}_i)$ Zufallsvariable und $f_i : (\Omega_i, \mathcal{A}_i) \longrightarrow (\Omega_i', \mathcal{A}_i')$, $i \in I$. Sind die $(X_i)_{i\in I}$ stochastisch unabhängig, dann sind auch $(f_i \circ X_i)_{i\in I}$ stochastisch unabhängig.*

http://www.springer.com/978-3-662-48936-9

Wahrscheinlichkeitstheorie
Rüschendorf, L.
2016, VIII, 469 S. 48 Abb., 3 Abb. in Farbe., Softcover
ISBN: 978-3-662-48936-9