

Stochastik für Studierende der Informatik

VON PETER PFAFFELHUBER

Version: 28. September 2013

Inhaltsverzeichnis

1	Grundlegendes	3
1.1	Häufige Modelle: Münzwurf, Würfeln, Urnen	3
1.2	Kombinatorik	4
1.3	Wahrscheinlichkeitsräume und Zufallsvariablen	5
1.4	Unabhängigkeit	10
2	Verteilungen und deren Eigenschaften	12
2.1	Grundlegendes	12
2.2	Laplace-Experimente	13
2.3	Die Binomial-Verteilung	13
2.4	Kenngrößen von Zufallsvariablen	14
3	Weitere wichtige Verteilungen	20
3.1	Die hypergeometrische Verteilung	20
3.2	Die Poisson-Verteilung und das Gesetz der kleinen Zahlen	22
3.3	Die geometrische und die Exponentialverteilung	23
3.4	Die Normalverteilung	25
3.5	Erzeugung von Zufallszahlen	26
4	Approximationssätze	27
4.1	Das schwache Gesetz der großen Zahlen	28
4.2	Der zentrale Grenzwertsatz und die Normalverteilung	29
5	Markov-Ketten	31
5.1	Bedingte Wahrscheinlichkeiten	31
5.2	Grundlegendes zu Markov-Ketten	33
5.3	Stationäre Verteilungen	38
6	Statistik	42
6.1	Grundlagen	42
6.2	Schätzprobleme	43
6.3	Testprobleme	48
7	Stochastische Optimierung	57
7.1	Simulated Annealing	58
7.2	Genetische Algorithmen	62

1 Grundlegendes

Die Stochastik beschäftigt sich mit zufälligen Ereignissen. Zunächst mag es erstaunen, dass man Regeln im Zufall finden kann. Auch wenn man nicht genau weiß, *welches* zufällige Ereignis passiert, kann man häufig angeben, *mit welcher Wahrscheinlichkeit* es passiert. Zentral ist hier der Begriff der *Wahrscheinlichkeitsverteilung*, die angibt, mit welchen (deterministischen) Wahrscheinlichkeiten zufällige Ereignisse passieren. Wir beschäftigen uns nicht damit, wie man diese Wahrscheinlichkeiten festlegt, sondern werden vor allem mit ihnen rechnen. Ein weiterer zentraler Begriff ist die *Zufallsvariable*, die ein zufälliges Element einer Menge beschreibt. In diesem ersten Kapitel werden wir neben der Einführung dieser Begriffe ein paar klassische Modelle kennen lernen, etwa das Werfen von Münzen. Dies wird bei der Behandlung von konkreten (Wahrscheinlichkeits-)Verteilungen sehr hilfreich sein.

1.1 Häufige Modelle: Münzwurf, Würfeln, Urnen

Bevor wir formal einführen, was Wahrscheinlichkeiten und Zufallsvariablen sind, behandeln wir häufige Beispiele.

Beispiel 1.1 (Münzwurf). Wenn man eine Münze wirft, zeigt sie bekanntlich *Kopf* oder *Zahl*. Bei einem p -Münzwurf ist die Wahrscheinlichkeit für *Kopf* gerade p . Damit ist die Wahrscheinlichkeit für *Zahl* gerade $1 - p$. Bei üblichen Münzen denkt man oft an den Fall $p = 1/2$, weil Kopf und Zahl mit derselben Wahrscheinlichkeit oben zu liegen kommen. Denkt man jedoch, dass man eine Reißzwecke wirft, die entweder mit der spitzen Seite oben oder unten zu liegen kommt, ist klar, dass auch $p \neq 1/2$ sein kann.

Wirft man zweimal mit derselben Münze, kann man nach der Wahrscheinlichkeit fragen, zweimal *Kopf* zu werfen. Dies ist p^2 .

Beispiel 1.2 (Würfeln). Eine bekannte Herausforderung ist es (etwa bei *Mensch ärgere Dich nicht*), beim Würfeln mit einem Würfel möglichst schnell eine 6 zu werfen. Dies ist ganz ähnlich wie beim $(1/6)$ -Münzwurf, schließlich gibt es nur zwei Möglichkeiten *6* oder *nicht 6*. Damit ist die Wahrscheinlichkeit, k mal keine 6 zu werfen, gerade $(5/6)^k$.

Wir berechnen noch die Wahrscheinlichkeit, mit zwei Würfeln ein *Mädchen* zu werfen, also gerade die Kombination 1,2. Diese ist

$$2 \cdot \frac{1}{6} \cdot \frac{1}{6}.$$

Denn: Entweder der erste Würfel zeigt eine 1 (Wahrscheinlichkeit $1/6$) und der zweite Würfel zeigt eine 2 (Wahrscheinlichkeit auch $1/6$), oder andersherum (deswegen der Faktor 2).

Beispiel 1.3 (Urne). In einer Urne befinden sich N Kugeln, und zwar K_1 Kugeln der Farbe 1, ..., K_n Kugeln der Farbe n . (Also ist $K_1 + \dots + K_n = N$.) Zieht man (mit verschlossenen Augen) aus der Urne, zieht man eine Kugel der Farbe i mit Wahrscheinlichkeit K_i/N . Zieht man anschließend nochmal aus der Urne (wobei die erste Kugel nicht zurückgelegt wurde), ist die Wahrscheinlichkeit nochmal eine Kugel der Farbe i zu ziehen $(K_i - 1)/(N - 1)$. Damit ist die Wahrscheinlichkeit, dass man zwei Kugeln derselben Farbe zu ziehen, gerade

$$\sum_{i=1}^n \frac{K_i}{N} \frac{K_i - 1}{N - 1}.$$

Beispiel 1.4 (Geburtstagsproblem). In einem Raum befinden sich 23 Personen. Wie groß ist die Wahrscheinlichkeit, dass es zwei Personen gibt die am selben Tag Geburtstag haben?

Um diese Wahrscheinlichkeit zu berechnen, ist es hilfreich, das Gegenteil *Alle Personen haben an unterschiedlichen Tagen Geburtstag* zu betrachten. Wir stellen die Personen (in Gedanken) in einer Reihe auf. Die Wahrscheinlichkeit dass die zweite Person an einem anderen Tag als die erste Person Geburtstag hat ist $364/365$. (Von Schaltjahren und dem 29.2. sehen wir einmal ab.) Weiter ist die Wahrscheinlichkeit, dass die dritte Person an einem anderen Tag als die Personen eins und zwei Geburtstag hat, dann $363/365$. Überlegen wir das weiter, so ist die Wahrscheinlichkeit, dass alle Personen an unterschiedlichen Tagen Geburtstag haben, gerade

$$\frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - 22}{365} \approx 0.493.$$

Damit ist die Wahrscheinlichkeit, dass es zwei Personen gibt, die am gleichen Tag Geburtstag haben, etwa $1 - 0.493 = 0.507$.

Beispiel 1.5 (Wartezeit bis zum nächsten Tor). Bei einem Fußballspiel nehmen wir an, dass in jeder Sekunde mit derselben Wahrscheinlichkeit $\lambda/3600$ ein Tor fällt (und zwar unabhängig von jeder anderen Sekunde). Hier ist λ die erwartete Zahl von Toren pro Stunde. Dann ist die Wahrscheinlichkeit, dass in t Stunden kein Tor fällt, gerade

$$\left(1 - \frac{\lambda}{3600}\right)^{3600t}.$$

Zerteilt man eine Stunde in noch kleinere Zeiteinheiten, behält jedoch die mittlere Anzahl von Toren pro Stunde bei, so beträgt die Wahrscheinlichkeit, dass in t Stunden noch kein Tor gefallen ist, gerade

$$\left(1 - \frac{\lambda}{n}\right)^{nt} \xrightarrow{n \rightarrow \infty} e^{-\lambda t}.$$

1.2 Kombinatorik

Bereits in den Beispielen 1.2 und 1.3 wird klar, dass man oftmals etwas abzählen muss, um Wahrscheinlichkeiten zu berechnen. Diese Kunst wird auch als Kombinatorik bezeichnet. Wir behandeln nun ein paar ausgewählte Fälle, eine Zusammenfassung findet man in Tabelle 6.1.

Lemma 1.6 (Kombinatorik ohne Wiederholungen). Sei $\underline{x} = (x_1, \dots, x_n)$ ein Vektor der Länge n , deren Einträge alle verschieden sind.

1. Es gibt $n! := n \cdot (n-1) \cdots 1$ verschiedene Vektoren der Länge n die dieselben Einträge wie \underline{x} haben.
2. Es gibt $n \cdot (n-1) \cdots (n-k+1)$ verschiedene Vektoren der Länge k , die aus den Einträgen von \underline{x} gewonnen werden können.
3. Es gibt $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ verschiedene, k -elementige Teilmengen von $\{x_1, \dots, x_n\}$.

Beweis. 1. Stellt man einen Vektor zusammen, so hat der erste Eintrag genau n Möglichkeiten. Für jede dieser Möglichkeiten hat der zweite Eintrag dann $n-1$ Möglichkeiten usw. Daraus ergibt sich die Formel.

2. folgt genau wie 1. nur dass man nach dem k -ten Eintrag abbrechen muss.

3. Geht man von 2. aus, so gibt es $\frac{n!}{(n-k)!}$ verschiedene Möglichkeiten, Vektoren der Länge k aus den Einträgen von \underline{x} zu bekommen. Von diesen Vektoren enthalten jeweils $k!$ dieselben Elemente. Will man also die Menge der k -elementigen Teilmengen zählen, so liefert jede dieser Teilmengen genau $k!$ Vektoren, also folgt das Ergebnis. \square

Lemma 1.7 (Kombinatorik mit Wiederholungen). *Gegeben seien n verschiedene Objekte, x_1, \dots, x_n .*

1. *Sind von den n Objekten genau r verschieden, und kommt das i -te Objekt genau k_i mal vor, $i = 1, \dots, r$, so gibt es genau $\frac{n!}{k_1! \dots k_r!}$ verschiedene Möglichkeiten, die Objekte in eine Reihenfolge zu bringen.*
2. *Zieht man aus den n Objekten genau k -mal mit Zurücklegen, so gibt es genau n^k mögliche Ausgänge, wenn man die Reihenfolge der gezogenen Objekte beachtet.*
3. *Zieht man aus den n Objekten genau k -mal mit Zurücklegen, so gibt es genau $\binom{n+k-1}{k}$ mögliche Ausgänge, wenn man die Reihenfolge der gezogenen Objekte nicht beachtet.*

Beweis. 1. Könnte man alle Objekte unterscheiden, so gäbe es $n!$ Möglichkeiten. Da man jedoch von dieser Anzahl jeweils $k_1!, \dots, k_r!$ für die möglichen Reihenfolgen der identischen Objekte zusammenfassen kann, folgt das Ergebnis.

2. ist klar.

3. Hier muss man etwas genauer nachdenken, wobei folgende Überlegung entscheidend ist: ist etwa $n = 5, k = 3$, so lässt sich jede gesuchte Möglichkeit durch einen Code der Art $\bullet\bullet**\bullet**$ darstellen. Dies bedeutet, dass vom ersten Objekt zwei Kopien gewählt wurden (deshalb $\bullet\bullet$), vom zweiten Objekt allerdings gar keines (deshalb ist zwischen den $*$'s kein \bullet), vom dritten eines, und vom vierten und fünften wiederum keines. Dieser Code ist also eine eindeutige Zuordnung der gesuchten Anordnungen auf die Menge der Vektoren der Länge $n + k - 1$ mit n Einträgen \bullet und $k - 1$ Einträgen $*$. (Man beachten, dass $*$ ja hier ein Trennzeichen ist, und man $n - 1$ Trennzeichen benötigt um n Felder zu trennen.) Die Anzahl dieser Vektoren ist gerade die gesuchte Anzahl, weil man nun die n -elementigen Teilmengen aller $n + k - 1$ Einträge sucht. \square

Bemerkung 1.8 (Permutation). Wir bezeichnen jede bijektive Abbildung einer Menge \mathcal{S} als Permutation von \mathcal{S} . Lemma 1.6.1 besagt, dass die Menge der Bijektionen von $\{1, \dots, n\}$ gerade $n!$ ist.

1.3 Wahrscheinlichkeitsräume und Zufallsvariablen

Definition 1.9 (Zufallsvariable). *Sei E eine Menge. Ein zufälliges Element von E heißt (E -wertige) Zufallsvariable.*

Bemerkung 1.10 (Zufälligkeit). 1. Eine Illustration einer Zufallsvariablen findet sich in Abbildung 1.1.

2. In obiger Definition ist zunächst unklar, was genau *zufällig* bedeuten soll. Dies wird nun durch den Begriff der *Verteilung* der Zufallsvariable behoben. Wir unterscheiden hierbei zwei Fälle: Entweder ist E höchstens abzählbar, oder Teilmenge der reellen Zahlen.

	Permutation	Variation	Kombination
ohne Wiederholung Beispiel	$n!$ Anordnung von n Zahlen, wobei jede Zahl nur einmal vorkommen darf	$\frac{n!}{(n-k)!}$ Ziehung von k Zahlen aus n Möglichen, wobei jede Zahl <i>nur einmal</i> vorkommen darf, <i>mit</i> Beachtung der Reihenfolge der Ziehung	$\binom{n}{k}$ Ziehung von k Zahlen aus n Möglichen, wobei jede Zahl <i>nur einmal</i> vorkommen darf, <i>ohne</i> Beachtung der Reihenfolge der Ziehung
mit Wiederholung Beispiel	$\frac{n!}{k_1! \cdot \dots \cdot k_r!}$ Anordnung von n Zahlen, die nicht alle verschieden sind	n^k Ziehung von k Zahlen aus n Möglichen, wobei jede Zahl <i>beliebig oft</i> vorkommen darf, <i>mit</i> Beachtung der Reihenfolge der Ziehung	$\binom{n+k-1}{k}$ Ziehung von k Zahlen aus n Möglichen, wobei jede Zahl <i>beliebig oft</i> vorkommen darf, <i>ohne</i> Beachtung der Reihenfolge der Ziehung

Tabelle 1.1: Kombinatorik beschäftigt sich mit dem Abzählen von Möglichkeiten. Generell unterscheidet man solche mit und ohne Wiederholungen, und mit (Permutationen und Variationen) und ohne (Kombinationen) Beachtung der Reihenfolge.

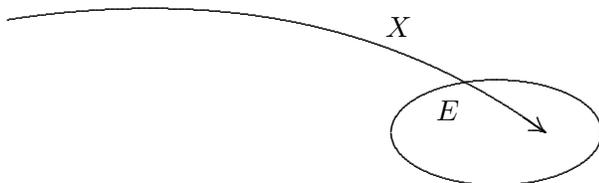


Abbildung 1.1: Eine Zufallsvariable ist ein zufälliges Element einer Menge.

Definition 1.11 (Verteilung einer Zufallsvariablen). Sei X eine E -wertige Zufallsvariable.

1. Angenommen, E ist höchstens abzählbar. Dann heißt X diskrete Zufallsvariable und die Funktion $x \mapsto \mathbf{P}[X = x]$ die Verteilung von X . Hier ist $\mathbf{P}[X = x]$ die Wahrscheinlichkeit, dass X den Wert x annimmt. Diese Abbildung heißt auch Zähldichte.
2. Angenommen, $E \subseteq \mathbb{R}$. Gilt für jedes Intervall $A \subseteq \mathbb{R}$, dass

$$\mathbf{P}[X \in A] = \int_A f(x) dx,$$

so heißt f Dichte der Verteilung von X und X eine stetige Zufallsvariable. Wir schreiben dann auch

$$\mathbf{P}[X \in dx] = f(x) dx.$$

3. Wir bezeichnen mit $\mathcal{P}(E)$ die Potenzmenge von E . Sei $\mathbf{P} : \mathcal{A} \subseteq \mathcal{P}(E) \rightarrow [0, 1]$. Gilt $\mathbf{P}[E] = 1$, $\mathbf{P}[A^c] = 1 - \mathbf{P}[A]$ und $\mathbf{P}[\bigcup_{n \in \mathbb{N}} A_n] = \sum_{n=1}^{\infty} \mathbf{P}[A_n]$ für paarweise disjunkte A_1, A_2, \dots (falls alle Werte definiert sind), so heißt \mathbf{P} eine Wahrscheinlichkeitsverteilung auf E . Sind \mathbf{P} und \mathbf{Q} Wahrscheinlichkeits-Verteilungen und ist X eine Zufallsvariable mit $\mathbf{P}[X \in A] = \mathbf{Q}[A]$, so schreiben wir auch $X \sim \mathbf{Q}$ oder $X \sim_{\mathbf{P}} \mathbf{Q}$.

Bemerkung 1.12. Offenbar definieren sowohl Zähldichten von diskreten Zufallsvariablen als auch Dichten von stetigen Zufallsvariablen Wahrscheinlichkeitsverteilungen.

Beispiel 1.13 (Münzwurf). Bei einem zweimaligen p -Münzwurf sei $X = (X_1, X_2)$ der zufällige, {Kopf, Zahl}-wertige Vektor, der den Ausgang des Wurfes beschreibt. Dann gilt etwa

$$\mathbf{P}[X = (\text{Kopf}, \text{Kopf})] = \mathbf{P}[X_1 = \text{Kopf}, X_2 = \text{Kopf}] = p^2.$$

Beispiel 1.14 (Wartezeit bis zum nächsten Tor). Sei X die Wartezeit bis zum nächsten Tor in Stunden wie in Beispiel 1.5. Ist $A = [a, b]$, so gilt also nach den Rechenregeln für Wahrscheinlichkeiten

$$\mathbf{P}[X \in A] = \mathbf{P}[X \geq a] - \mathbf{P}[X \geq b] = e^{-\lambda a} - e^{-\lambda b} = \int_a^b \lambda e^{-\lambda x} dx.$$

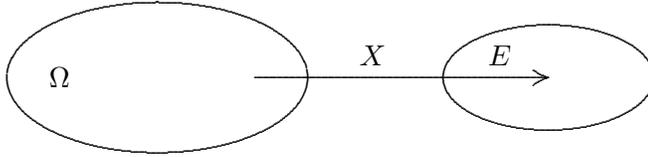


Abbildung 1.2: In der klassischen Modellierung ist eine Zufallsvariable eine Abbildung $X : \Omega \mapsto E$.

Also ist

$$f(x) := \lambda e^{-\lambda x}$$

die Dichte der Verteilung von X . Diese werden wir Exponentialverteilung zum Parameter λ nennen.

Bemerkung 1.15. In der Mathematik definiert man Zufallsvariablen und Verteilungen etwas anders. Während oben der Begriff der Zufallsvariable im Vordergrund steht, führt man in mathematischen Vorlesungen zur Stochastik zunächst Wahrscheinlichkeitsverteilungen ein. Für eine Menge Ω sind dies Abbildungen $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$, wobei $\mathcal{A} \subseteq \mathcal{P}[\Omega]$ eine σ -Algebra, also ein Mengensystem mit $\emptyset \in \mathcal{A}$, $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ ist. Für die Abbildung \mathbf{P} fordert man nun $\mathbf{P}[\emptyset] = 0$, $\mathbf{P}[A^c] = 1 - \mathbf{P}[A]$ und $\mathbf{P}[\bigcup_{n \in \mathbb{N}} A_n] = \sum_{n=1}^{\infty} \mathbf{P}[A_n]$ für paarweise disjunkte A_1, A_2, \dots . Hier sind nun automatisch alle Werte definiert, da \mathcal{A} eine σ -Algebra ist.

In dieser Form der Modellierung ist eine E -wertige Zufallsvariable eine Abbildung $X : \Omega \rightarrow E$; siehe auch Abbildung 1.2. Die Verteilung von X ist dann gegeben durch

$$\mathbf{P}[X \in A] = \mathbf{P}[X^{-1}(A)],$$

wobei im letzten Ausdruck $X^{-1}(A) \in \mathcal{A}$ gelten muss.

Wir werden jedoch (wie oben) den intuitiveren Begriff der Zufallsvariable in den Vordergrund stellen. Ganz egal wie man modelliert, wichtig ist, dass auf jeden Fall die Rechenregeln aus Definition 1.11.3 gelten.

Wir lernen nun noch eine wichtige Formel zum Rechnen mit Wahrscheinlichkeiten kennen.

Proposition 1.16 (Einschluss- Ausschluss-Formel). Sei X eine E -wertige Zufallsvariable und $A_1, \dots, A_n \subseteq E$. Dann gilt

$$\begin{aligned} & \mathbf{P}[X \in A_1 \cup \dots \cup A_n] \\ &= \sum_{i=1}^n \mathbf{P}[X \in A_i] - \sum_{1 \leq i < j \leq n} \mathbf{P}[X \in A_i \cap A_j] + \dots \pm \mathbf{P}[X \in A_1 \cap \dots \cap A_n]. \end{aligned}$$

Beweis. Zunächst sei $n = 2$. Klar ist, dass $\mathbf{P}[X \in A_2 \setminus (A_1 \cap A_2)] = \mathbf{P}[X \in A_2] - \mathbf{P}[X \in A_1 \cap A_2]$, da $A_1 \cap A_2$ und $A_2 \setminus (A_1 \cap A_2)$ disjunkt sind. Daraus folgt nun

$$\begin{aligned}\mathbf{P}[X \in A_1 \cup A_2] &= \mathbf{P}[X \in A_1] + \mathbf{P}[X \in A_2 \setminus (A_1 \cap A_2)] \\ &= \mathbf{P}[X \in A_1] + \mathbf{P}[X \in A_2] - \mathbf{P}[X \in A_1 \cap A_2].\end{aligned}$$

Wir beweisen die Aussage mit Hilfe von vollständiger Induktion. Es gilt wie im Fall $n = 2$

$$\begin{aligned}\mathbf{P}[X \in A_1 \cup \dots \cup A_{n+1}] &= \mathbf{P}[X \in A_1 \cup \dots \cup A_n] + \mathbf{P}[X \in A_{n+1}] \\ &\quad - \mathbf{P}[X \in (A_1 \cap A_{n+1}) \cup \dots \cup (A_n \cap A_{n+1})] \\ &= \sum_{i=1}^{n+1} \mathbf{P}[X \in A_i] - \sum_{1 \leq i < j \leq n} \mathbf{P}[X \in A_i \cap A_j] - \sum_{i=1}^n \mathbf{P}[X \in A_i \cap A_{n+1}] \\ &\quad \dots \pm \mathbf{P}[X \in A_1 \cap \dots \cap A_n] \pm \sum_{i=1}^n \mathbf{P}[X \in A_1 \cap \dots \cap A_{i-1} \cap A_{i+1} \cap \dots \cap A_{n+1}] \\ &\quad \mp \mathbf{P}[X \in A_1 \cap \dots \cap A_{n+1}].\end{aligned}$$

Damit ist die Aussage gezeigt. \square

Beispiel 1.17 (Fixpunkte in Permutationen). Als Anwendung von Proposition 1.16 berechnen wir die Wahrscheinlichkeit, dass es in einer Permutation der Zahlen $1, \dots, n$ keine Fixpunkte gibt. (Ein Fixpunkt einer Permutation σ ist ein i mit $\sigma(i) = i$.) Sei hierzu X ein rein zufälliges Element von \mathcal{S}_n ,

$$A_i := \{\sigma \in \mathcal{S}_n : \sigma(i) = i\}$$

und

$$Y := \sum_{i=1}^n 1_{X \in A_i}$$

die Anzahl der Fixpunkte von X . Klar ist, dass

$$\mathbf{P}[X \in A_i] = \frac{(n-1)!}{n!} = \frac{1}{n},$$

weil es gerade $(n-1)!$ Permutationen gibt, die i als Fixpunkt haben. Analog gilt für $1 \leq i_1 < \dots < i_k \leq n$

$$\mathbf{P}[X \in A_{i_1} \cap \dots \cap A_{i_k}] = \frac{(n-k)!}{n!} = \frac{1}{n \cdot \dots \cdot (n-k+1)}.$$

Weiter gilt

$$\begin{aligned}\mathbf{P}[Y > 0] &= \mathbf{P}[X \in A_1 \cup \dots \cup A_n] \\ &= n\mathbf{P}[X \in A_1] - \binom{n}{2}\mathbf{P}[A_1 \cap A_2] + \binom{n}{3}\mathbf{P}[X \in A_1 \cap A_2 \cap A_3] - \dots \\ &\quad \pm \binom{n}{n}\mathbf{P}[X \in A_1 \cap \dots \cap A_n] \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots \pm \frac{1}{n!}.\end{aligned}$$

Damit ist die Wahrscheinlichkeit, dass X keine Fixpunkte besitzt, gerade

$$\mathbf{P}[Y = 0] = 1 - \mathbf{P}[Y > 0] = \frac{1}{2!} - \frac{1}{3!} + \cdots \mp \frac{1}{n!}.$$

Wir bemerken, dass für große n

$$\mathbf{P}[Y = 0] = \sum_{k=2}^n \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} \xrightarrow{n \rightarrow \infty} e^{-1}.$$

1.4 Unabhängigkeit

Bereits in den Beispielen 1.1, 1.2, 1.4, 1.5 haben wir Wahrscheinlichkeiten miteinander multipliziert. Etwa hatten wir gesagt, dass die Wahrscheinlichkeit beim p -Münzwurf für das zweimalige Auftreten von *Kopf* gerade p^2 ist. Begründet haben wir dies mit der Unabhängigkeit der Münzwürfe. Um dieses Konzept einzuführen, benötigen wir zunächst die gemeinsame Verteilung von Zufallsvariablen.

Definition 1.18 (Gemeinsame Verteilung). Ist $\underline{X} = (X_1, \dots, X_n)$ mit Zielbereich $E_1 \times \dots \times E_n$, so heißt die Verteilung von \underline{X} auch gemeinsame Verteilung von X_1, \dots, X_n . Die Abbildung

$$A_i \mapsto \mathbf{P}[X_i \in A_i]$$

ist die i -te Rand- oder Marginalverteilung der gemeinsamen Verteilung von X_1, \dots, X_n .

Gilt für alle Intervalle $A_1, \dots, A_n \subseteq \mathbb{R}$, dass

$$\mathbf{P}[\underline{X} \in A_1 \times \dots \times A_n] = \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

so heißt f Dichte der gemeinsamen Verteilung von X_1, \dots, X_n . Wir schreiben dann auch

$$\mathbf{P}[\underline{X} \in dx_1 \cdots dx_n] = f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Definition 1.19 (Bildverteilung). Ist X eine E -wertige Zufallsvariable und

$$h : E \rightarrow E',$$

so ist $Y := h(X)$ eine E' -wertige Zufallsvariable und

$$\{Y = y\} = \{X \in h^{-1}(y)\},$$

also ist die Verteilung von Y gegeben durch

$$\mathbf{P}[Y \in B] = \mathbf{P}[X \in h^{-1}(B)].$$

Bemerkung 1.20 (Interpretation). Auch wenn die Definition abstrakt erscheint, sind Bilder von Zufallsvariablen häufig anzutreffen. Sei etwa (X_1, \dots, X_n) das Ergebnis eines n -fachen Würfelwurfs. Dann beschreibt die Zufallsvariable $Z_i := h(X_i)$ mit $h(x) := 1_{x=6}$, ob im i -ten Wurf eine 6 gefallen ist oder nicht. Weiter zählt die Zufallsvariable $\ell(Z_1, \dots, Z_n)$ mit $\ell(z_1, \dots, z_n) := z_1 + \dots + z_n$ die Anzahl der 6er.

Eine Illustration von Bildern von Zufallsvariablen findet sich in Abbildung 6.2.

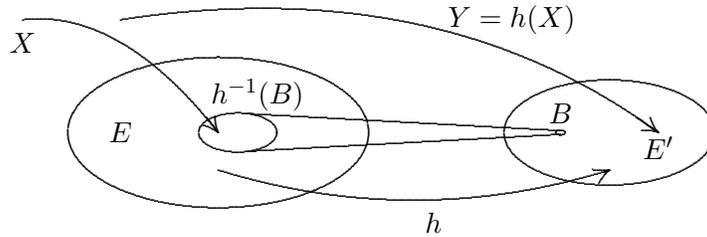


Abbildung 1.3: Das Bild einer Zufallsvariablen X unter einer Abbildung h ist wieder eine Zufallsvariable, nämlich $h(X)$.

Beispiel 1.21 (p -Münzwurf). Wir betrachten einen p -Münzwurf mit $q := 1 - p$. Also sei $X = (X_1, \dots, X_n)$ eine $\{0, 1\}^n$ -wertige Zufallsvariable mit

$$\mathbf{P}[X = (x_1, \dots, x_n)] = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}.$$

Sei $h(x_1, \dots, x_n) = x_1 + \dots + x_n$. Dann gilt $|h^{-1}(k)| = \binom{n}{k}$, also

$$\begin{aligned} \mathbf{P}(h(X) = k) &= \mathbf{P}(X \in h^{-1}(k)) \\ &= \sum_{(x_1, \dots, x_n) \in h^{-1}(k)} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}. \end{aligned}$$

Definition 1.22 (Unabhängigkeit). Seien X_1, \dots, X_n Zufallsvariable. Gilt

$$\mathbf{P}[X_1 \in A_1, \dots, X_n \in A_n] = \mathbf{P}[X_1 \in A_1] \cdots \mathbf{P}[X_n \in A_n] \quad (1.1)$$

für beliebige A_1, \dots, A_n , so heißt (X_1, \dots, X_n) unabhängig. Gilt (1.1) nur für Paare $i \neq j$, d.h.

$$\mathbf{P}[X_i \in A_i, X_j \in A_j] = \mathbf{P}[X_i \in A_i] \cdot \mathbf{P}[X_j \in A_j]$$

für beliebige A_i, A_j und $i \neq j$, so heißt (X_1, \dots, X_n) paarweise unabhängig.

Bemerkung 1.23 (Unabhängigkeit von diskreten und stetigen Zufallsvariablen).

Man überlegt sich einfach:

Sind X_1, \dots, X_n diskrete Zufallsvariable. Dann ist (X_1, \dots, X_n) genau dann unabhängig, wenn

$$\mathbf{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbf{P}[X_1 = x_1] \cdots \mathbf{P}[X_n = x_n]$$

für beliebige Wahl von x_1, \dots, x_n .

Weiter gilt:

Sind X_1, \dots, X_n stetige Zufallsvariablen, so dass die gemeinsame Verteilung eine Dichte f besitzt. Dann ist (X_1, \dots, X_n) genau dann unabhängig, wenn es Funktionen f_1, \dots, f_n gibt mit $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$. Dabei gilt

$$f_i(x_i) = \int f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Beispiel 1.24. 1. Beispiele für unabhängige Zufallsvariable haben wir bereits kennen gelernt. Beispielsweise ist ein p -Münzwurf (X_1, \dots, X_n) unabhängig.

2. Wir geben nun ein Beispiel für paarweise unabhängige, aber nicht unabhängige Zufallsvariablen an. Sei hierzu (X_1, X_2) ein $1/2$ -Münzwurf. Wir betrachten dann $\underline{Z} = (Z_1, Z_2, Z_3)$ mit $Z_1 = X_1, Z_2 = X_2, Z_3 = 1_{X_1=X_2}$. Die Zufallsgröße Z_3 gibt also gerade an, ob die beiden Ergebnisse der Münzwürfe identisch sind oder nicht. Es gilt $\mathbf{P}[Z_3 = 1] = \mathbf{P}[X_1 = X_2 = 0 \text{ oder } X_1 = X_2 = 1] = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$. Weiter ist

$$\mathbf{P}[Z_1 = 1, Z_3 = 1] = \mathbf{P}[X_1 = X_2 = 1] = \frac{1}{4} = \mathbf{P}[Z_1 = 1] \cdot \mathbf{P}[Z_3 = 1]$$

und analog für das Paar (Z_2, Z_3) . Damit ist Z paarweise unabhängig. Jedoch gilt

$$\mathbf{P}[Z_1 = 0, Z_2 = 1, Z_3 = 1] = 0 \neq \frac{1}{8} = \mathbf{P}[Z_1 = 0] \cdot \mathbf{P}[Z_2 = 1] \cdot \mathbf{P}[Z_3 = 1],$$

also ist \underline{Z} nicht unabhängig.

2 Verteilungen und deren Eigenschaften

2.1 Grundlegendes

Einen besonderen Stellenwert haben reellwertige Zufallsvariablen und deren Verteilungen. Für diese lässt sich die Verteilungsfunktion definieren.

Definition 2.1 (Verteilungsfunktion und Quantile). Sei $E \subseteq \mathbb{R}$ und X eine E -wertige Zufallsvariable. Dann heißt die Funktion

$$x \mapsto F_X(x) := \mathbf{P}[X \leq x]$$

Verteilungsfunktion von X . Sei weiter $\alpha \in (0, 1)$. Jedes $q_\alpha \in \mathbb{R}$ mit

$$\mathbf{P}[X < q_\alpha] \leq \alpha \leq \mathbf{P}[X \leq q_\alpha]$$

heißt α -Quantil von (der Verteilung von) X . Jedes 0.5 -Quantil heißt Median von (der Verteilung von) X .

Lemma 2.2 (Eigenschaften von Verteilungsfunktionen). Sei X eine (diskrete oder stetige) reellwertige Zufallsvariable. Dann gilt

$$\begin{aligned} \lim_{x \rightarrow \infty} F_X(x) &= 1, \\ \lim_{x \rightarrow -\infty} F_X(x) &= 0. \end{aligned}$$

Ist X eine stetige Zufallsvariable mit Dichte $f(x)dx$, so gilt

$$F_X(x) = \int_{-\infty}^x f(z)dz.$$

Inbesondere ist also $F' = f$ und die Verteilungsfunktion damit differenzierbar.

Beweis. Wir zeigen die ersten beiden Aussagen für diskrete Zufallsvariable. Sei $E \subseteq \mathbb{R}$ der diskrete Wertebereich von X . Dann gilt schließlich $\sum_{z \in E} \mathbf{P}[X = z] = 1$ und damit $\mathbf{P}[X \leq x] = \sum_{z \in E, z \leq x} \mathbf{P}[X = z] \xrightarrow{x \rightarrow \infty} 1$, da andernfalls die Gesamtsumme von $\mathbf{P}[X = z]$ nicht konvergiert. Die zweite Gleichung folgt analog.

Die Aussage für stetige Zufallsvariable ist klar, da $(-\infty, x]$ ein Intervall ist und damit

$$\mathbf{P}[X \leq x] = \mathbf{P}[X \in (-\infty, x]] = \int_{-\infty}^x f(z) dz$$

gilt. □

2.2 Laplace-Experimente

Man stelle sich einen Würfelwurf mit einem fairen Würfel vor. Hier sind alle möglichen Ausgänge gleichwahrscheinlich. In einem solchen Fall spricht man auch von einem Laplace-Experiment. Solche Experimente führen zu uniformen Verteilungen, die wir nun kennen lernen.

Definition 2.3 (Diskrete uniforme Verteilung). Sei E eine endliche Menge. Ist $x \mapsto \mathbf{P}[X = x]$ konstant, so heißt die Verteilung von X diskrete uniforme Verteilung. In diesem Fall gilt

$$\mathbf{P}[X \in A] = \frac{|A|}{|E|}.$$

Definition 2.4 (Stetige uniforme Verteilung). Sei $E = [a, b]$ ein Intervall und X eine E -wertige Zufallsvariable. Gilt für $a \leq c \leq d \leq b$

$$\mathbf{P}[X \in (c, d)] = \frac{d - c}{b - a},$$

so heißt die Verteilung von X stetige uniforme Verteilung. In diesem Fall hat X die Dichte

$$\frac{1}{b - a} 1_{[a, b]}(x) dx$$

und wir schreiben auch $X \sim U([a, b])$. Weiter können wir auch die Verteilungsfunktion von X berechnen. Es gilt

$$\mathbf{P}[X \leq x] = \begin{cases} 0, & x < a, \\ \frac{1}{b-a} \int_a^x dz = \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases}$$

2.3 Die Binomial-Verteilung

Denkt man an ein n -mal unabhängig wiederholt durchgeführtes Spiel, so ist es möglich, zwischen 0 und n -mal zu gewinnen. Die Anzahl der Gewinne ist eine Zufallsvariable, deren Verteilung wir nun als Binomialverteilung kennen lernen. Diese haben wir schon in Beispiel 1.21 kennen gelernt. Die Verteilungsgewichte und die Verteilungsfunktion der Binomialverteilung sind in Abbildung 2.1 dargestellt.

Definition 2.5 (Binomialverteilung). Sei $E = \{0, \dots, n\}$ und $0 \leq p \leq 1$. Ist X eine E -wertige Zufallsvariable mit

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad (2.1)$$

so heißt die Verteilung von X Binomialverteilung mit Parametern n und p und wir schreiben $X \sim B(n, p)$.

Bemerkung 2.6. Die $B(n, p)$ Verteilung gibt die Anzahl der Erfolge in einem n -mal unabhängig wiederholt ausgeführten Spiel dar, wobei in jedem Spiel die Gewinnwahrscheinlichkeit gerade p beträgt.

Denn: Jeder Spielausgang des wiederholten Spiels kann man mit einem $\{0, 1\}$ -wertigem Vektor der Länge n beschreiben. Ist etwa $n = 5$, so ist $(1, 1, 0, 0, 1)$ ein Gewinn in den ersten beiden und im letzten Spiel. Ein solcher Spielausgang hat gerade die Wahrscheinlichkeit $p \cdot p \cdot (1-p) \cdot (1-p) \cdot p = p^3(1-p)^2$. Da es genau $\binom{n}{k}$ Vektoren gibt, die aus genau k Ziffern 1 und $n-k$ Ziffern 0 bestehen, folgt (2.1).

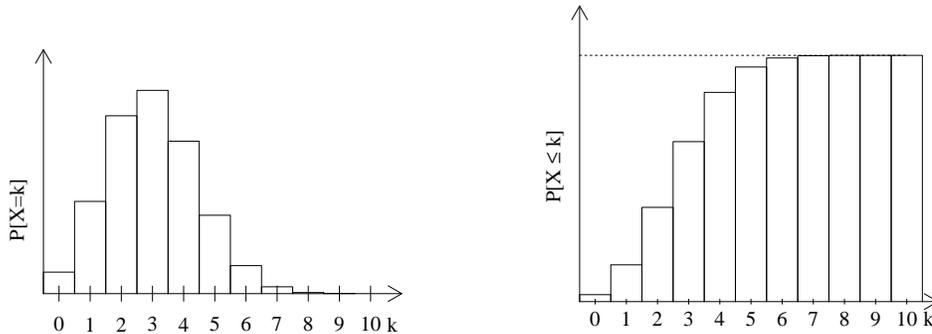


Abbildung 2.1: Verteilungsgewichte und Verteilungsfunktion von $B(10, 0.2)$.

2.4 Kenngrößen von Zufallsvariablen

Definition 2.7 (Erwartungswert). Sei X eine E -wertige Zufallsvariable und $h : E \rightarrow \mathbb{R}$.

1. Ist E diskret, so heißt

$$\mathbf{E}[h(X)] := \sum_{x \in E} h(x) \mathbf{P}[X = x]$$

Erwartungswert von $h(X)$, falls die Summe konvergiert.

2. Ist X stetig mit Dichte f , so ist

$$\mathbf{E}[h(X)] := \int h(x) f(x) dx$$

Erwartungswert von $h(X)$, falls das Integral existiert.

Ist speziell $E \subseteq \mathbb{R}$ und $h = \text{id}$, so heißt

$$\mathbf{E}[X] := \sum_{x \in E} x \mathbf{P}[X = x],$$

$$\mathbf{E}[X] := \int x f(x) dx$$

Erwartungswert von X , falls Summe und Integral existieren.

Beispiel 2.8 (Uniforme und Binomialverteilung). 1. Sei $X \sim B(n, p)$. Dann ist mit $q := 1 - p$

$$\begin{aligned} \mathbf{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-k} = np. \end{aligned} \tag{2.2}$$

2. Sei $X \sim U([a, b])$. Dann ist

$$\mathbf{E}[X] = \frac{1}{b-a} \int_a^b x dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{1}{2}(a+b).$$

Lemma 2.9 (Transformationssatz). Sei X eine E -wertige diskrete Zufallsvariable und $h : E \rightarrow \mathbb{R}$. Dann gilt

$$\mathbf{E}[h(X)] = \sum_{y \in h(E)} y \mathbf{P}[h(X) = y]$$

falls die Summe existiert.

Beweis. Es gilt nach Definition des Erwartungswertes

$$\begin{aligned} \mathbf{E}[h(X)] &= \sum_{x \in E} h(x) \mathbf{P}[X = x] = \sum_{y \in h(E)} \sum_{x \in h^{-1}(y)} h(x) \mathbf{P}[X = x] \\ &= \sum_{y \in h(E)} y \sum_{x \in h^{-1}(y)} \mathbf{P}[X = x] = \sum_{y \in h(E)} y \mathbf{P}[h(X) = y]. \end{aligned}$$

□

Lemma 2.10. Sei X eine \mathbb{Z}_+ -wertige Zufallsvariable. Dann gilt

$$\mathbf{E}[X] = \sum_{x=0}^{\infty} \mathbf{P}[X > x],$$

$$\mathbf{E}[X(X-1)] = 2 \cdot \sum_{x=0}^{\infty} x \cdot \mathbf{P}[X > x]$$

falls jeweils die Summe existiert.

Beweis. Für die erste Behauptung schreiben wir

$$\begin{aligned}\mathbf{E}[X] &= \sum_{x=0}^{\infty} x \mathbf{P}[X = x] = \sum_{x=1}^{\infty} \sum_{y=1}^x \mathbf{P}[X = x] = \sum_{y=1}^{\infty} \sum_{x=y}^{\infty} \mathbf{P}[X = x] \\ &= \sum_{y=1}^{\infty} \mathbf{P}[X \geq y] = \sum_{y=0}^{\infty} \mathbf{P}[X > y]\end{aligned}$$

wobei die Umordnung der Doppelsumme wegen der absoluten Konvergenz der Reihe möglich ist. Für die zweite Behauptung ist analog

$$\begin{aligned}\mathbf{E}[X(X-1)] &= 2 \sum_{x=1}^{\infty} \frac{x(x-1)}{2} \cdot \mathbf{P}[X = x] = 2 \sum_{x=1}^{\infty} \sum_{y=0}^{x-1} y \cdot \mathbf{P}[X = x] \\ &= 2 \sum_{y=0}^{\infty} \sum_{x=y+1}^{\infty} y \cdot \mathbf{P}[X = x] = 2 \sum_{y=0}^{\infty} y \cdot \mathbf{P}[X > y].\end{aligned}$$

□

Beispiel 2.11 (Diskrete uniforme Verteilung). Sei X uniform auf $\{k, \dots, \ell\}$ verteilt. Dann ist

$$\begin{aligned}\mathbf{E}[X] &= \sum_{x=k}^{\ell} x \frac{1}{\ell - k + 1} = \frac{1}{\ell - k + 1} \left(\sum_{x=0}^{\ell} x - \sum_{x=0}^{k-1} x \right) = \frac{\binom{\ell+1}{2} - \binom{k}{2}}{\ell - k + 1} \\ &= \frac{(\ell - k + 1)(\ell + k)}{2(\ell - k + 1)} = \frac{\ell + k}{2}.\end{aligned}$$

Genauso können wir schreiben

$$\begin{aligned}\mathbf{E}[X] &= \sum_{x=0}^{\ell} \mathbf{P}[X > x] = \sum_{x=0}^{k-1} 1 + \sum_{x=k}^{\ell} \frac{\ell - x}{\ell - k + 1} = k + \frac{1}{\ell - k + 1} \sum_{x=0}^{\ell-k} x \\ &= k + \frac{\binom{\ell-k+1}{2}}{\ell - k + 1} = k + \frac{\ell - k}{2} = \frac{\ell + k}{2}.\end{aligned}$$

Proposition 2.12 (Eigenschaften des Erwartungswertes). Seien X, Y (diskrete oder stetige) Zufallsvariablen.

1. Sei $a, b \in \mathbb{R}$. Dann gilt

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y],$$

falls alle Erwartungswerte existieren.

2. Gilt $X \leq Y$, so ist $\mathbf{E}[X] \leq \mathbf{E}[Y]$.

Beweis. Wir zeigen beide Aussagen nur im Fall von diskreten Zufallsvariablen. Es gilt

$$\begin{aligned}\mathbf{E}[aX + bY] &= \sum_{x,y \in E} (ax + by) \mathbf{P}[X = x, Y = y] \\ &= a \sum_{x,y \in E} x \mathbf{P}[X = x, Y = y] + b \sum_{x,y \in E} y \mathbf{P}[X = x, Y = y] \\ &= a \sum_{x \in E} \mathbf{P}[X = x] + b \sum_{y \in E} y \mathbf{P}[Y = y] = a\mathbf{E}[X] + b\mathbf{E}[Y].\end{aligned}$$

Für 2. genügt es zu zeigen, dass $\mathbf{E}[X] \geq 0$ für $X \geq 0$. (Dann ist nämlich $\mathbf{E}[Y] - \mathbf{E}[X] = \mathbf{E}[Y - X] \geq 0$.) Wir schreiben, falls $X \geq 0$

$$\mathbf{E}[X] = \sum_{x \in E} x \mathbf{P}[X = x] = \sum_{x \in E, x \geq 0} x \mathbf{P}[X = x] \geq \sum_{x \in E, x \geq 0} 0 \cdot \mathbf{P}[X = x] = 0.$$

□

Beispiel 2.13 (*p*-Münzwurf). Ist $X = (X_1, \dots, X_n)$ ein p -Münzwurf, so wissen wir bereits, dass $Y := X_1 + \dots + X_n \sim B(n, p)$. Also muss auch

$$\mathbf{E}[Y] = \mathbf{E}[X_1 + \dots + X_n] = \mathbf{E}[X_1] + \dots + \mathbf{E}[X_n] = n\mathbf{E}[X_1] = n\mathbf{P}[X_1 = 1] = np$$

in Übereinstimmung mit (2.2).

Definition 2.14 (Varianz). Sei X entweder eine diskrete, $E \subseteq \mathbb{R}$ -wertige Zufallsvariable, oder eine stetige Zufallsvariable und es existiere $\mu := \mathbf{E}[X]$. Dann ist

$$\sigma^2 := \mathbf{Var}[X] = \mathbf{E}[(X - \mu)^2]$$

die Varianz von X . In beiden Fällen heißt

$$\sigma := \sqrt{\mathbf{Var}[X]}$$

die Standardabweichung von X .

Bemerkung 2.15. Für diskrete Zufallsvariablen ist also

$$\mathbf{Var}[X] = \sum_{x \in E} (x - \mu)^2 \mathbf{P}[X = x],$$

für stetige mit Dichte $f(x)dx$ gerade

$$\mathbf{Var}[X] = \int (x - \mu)^2 f(x) dx$$

falls Summe und Integral existieren.

Proposition 2.16 (Rechenregeln für die Varianz). Sei X entweder eine diskrete, $E \subseteq \mathbb{R}$ -wertige Zufallsvariable oder eine stetige Zufallsvariable. Dann gilt

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \mathbf{E}[X(X - 1)] + \mathbf{E}[X] - \mathbf{E}[X]^2$$

falls alle Erwartungswerte existieren.

Beweis. Die zweite Gleichheit ist klar wegen der Linearität des Erwartungswertes. Sei $\mu := \mathbf{E}[X]$. Dann schreiben wir

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mu)^2] = \mathbf{E}[X^2 - 2\mu X + \mu^2] = \mathbf{E}[X^2] - 2\mu\mathbf{E}[X] + \mu^2 = \mathbf{E}[X^2] - \mu^2.$$

Damit ist alles gezeigt. □

Beispiel 2.17 (Uniforme und Binomialverteilung). 1. Sei $X \sim B(n, p)$. Dann ist mit $q := 1 - p$

$$\begin{aligned}\mathbf{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} q^{n-k} = n(n-1)p^2,\end{aligned}$$

also

$$\mathbf{Var}[X] = n(n-1)p^2 + np - (np)^2 = npq.$$

2. Sei $X \sim U([a, b])$. Dann ist

$$\begin{aligned}\mathbf{Var}[X] &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{(a+b)^2}{4} = \frac{1}{3} \frac{b^3 - a^3}{b-a} - \frac{(a+b)^2}{4} \\ &= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{(b-a)^2}{12}.\end{aligned}$$

Definition 2.18 (Covarianz und Korrelationskoeffizient). Seien X, Y diskrete oder stetige Zufallsvariablen, deren Erwartungswerte $\mu_X := \mathbf{E}[X], \mu_Y := \mathbf{E}[Y]$ sowie Varianzen $\sigma_X^2 := \mathbf{Var}[X], \sigma_Y^2 := \mathbf{Var}[Y]$ existieren. Dann existiert auch

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mu_X)(Y - \mu_Y)]$$

und heißt Covarianz von X und Y . Gilt $\mathbf{Cov}[X, Y] = 0$, so heißen X, Y unkorreliert.

Außerdem nennen wir

$$\mathbf{Kor}[X, Y] := \frac{\mathbf{Cov}[X, Y]}{\sigma_X \cdot \sigma_Y}$$

Korrelationskoeffizient von X und Y .

Lemma 2.19 (Rechenregeln für Covarianzen). Seien X, Y, Z (diskrete oder stetige) reellwertige Zufallsvariable mit endlicher Varianz. Dann gilt

$$\begin{aligned}\mathbf{Cov}[X, Y] &= \mathbf{Cov}[Y, X], \\ \mathbf{Var}[X] &= \mathbf{Cov}[X, X], \\ \mathbf{Cov}[X, Y] &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y], \\ \mathbf{Cov}[X, aY + bZ] &= a\mathbf{Cov}[X, Y] + b\mathbf{Cov}[X, Z].\end{aligned}$$

Beweis. Wir setzen $\mu_X := \mathbf{E}[X], \mu_Y := \mathbf{E}[Y]$. Die ersten beiden Gleichungen sind klar. Für die dritte Gleichung schreiben wir

$$\mathbf{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbf{E}[XY] - \mu_X \mathbf{E}[Y] - \mu_Y \mathbf{E}[X] + \mu_X \mu_Y = \mathbf{E}[XY] - \mu_X \mu_Y.$$

Die vierte folgt aus der Definition der Covarianz und der Linearität des Erwartungswertes. \square

Lemma 2.20 (Cauchy–Schwartz Ungleichung). Es gilt

$$\mathbf{E}[XY]^2 \leq \mathbf{E}[|XY|]^2 \leq \mathbf{Var}[X]\mathbf{Var}[Y].$$

Insbesondere ist

$$-1 \leq \mathbf{Kor}[X, Y] \leq 1.$$

Beweis. Die erste Ungleichung ist klar, da $XY \leq |XY|$. Für die zweite genügt es den Fall $\mathbf{E}[X^2] > 0$ zu betrachten. Andernfalls muss nämlich $\mathbf{P}[X = 0] = 1$ gelten und dann ist nichts zu zeigen. Für jedes $c \in \mathbb{R}$ gilt $0 \leq (-c|X| + |Y|)^2 = c^2X^2 - 2c|XY| + Y^2$. Insbesondere gilt für $c := \mathbf{E}[|XY|]/\mathbf{E}[X^2]$

$$\begin{aligned} 0 &\leq c^2\mathbf{E}[X^2] - 2c\mathbf{E}[|XY|] + \mathbf{E}[Y^2] = \mathbf{E}[|XY|^2]/\mathbf{E}[X^2] - 2\mathbf{E}[|XY|]^2/\mathbf{E}[X^2] + \mathbf{E}[Y^2] \\ &= \mathbf{E}[Y^2] - \mathbf{E}[|XY|^2]/\mathbf{E}[X^2] \end{aligned}$$

woraus die erste Behauptung direkt folgt. Die zweite folgt dann aus der Definition des Korrelationskoeffizienten. \square

Proposition 2.21 (Unabhängigkeit und Unkorreliertheit). *Sind X, Y (diskrete oder stetige) reellwertige, unabhängige Zufallsvariable, deren Varianz existiert. Dann sind X, Y auch unkorreliert, d.h. $\mathbf{Cov}[X, Y] = 0$.*

Beweis. Wir zeigen die Behauptung nur für diskrete Zufallsvariable. Es gilt

$$\begin{aligned} \mathbf{E}[XY] &= \sum_{x,y} xy\mathbf{P}[X = x, Y = y] = \sum_x x\mathbf{P}[X = x] \sum_y y\mathbf{P}[Y = y] \\ &= \mathbf{E}[X]\mathbf{E}[Y]. \end{aligned}$$

\square

Proposition 2.22 (Varianz einer Summe). *Seien X_1, \dots, X_n (diskrete oder stetige) reellwertige Zufallsvariable mit endlichen Varianzen. Dann gilt*

$$\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}[X_i] + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}[X_i, X_j].$$

Ist insbesondere (X_1, \dots, X_n) paarweise unkorreliert, so gilt

$$\mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}[X_i].$$

Beweis. Wir verwenden Lemma 2.19 und schreiben

$$\begin{aligned} \mathbf{Var}\left[\sum_{i=1}^n X_i\right] &= \mathbf{Cov}\left[\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right] = \sum_{i=1}^n \sum_{j=1}^n \mathbf{Cov}[X_i, X_j] \\ &= \sum_{i=1}^n \mathbf{Cov}[X_i, X_i] + \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{Cov}[X_i, X_j] + \sum_{i=1}^n \sum_{j=i+1}^n \mathbf{Cov}[X_i, X_j] \\ &= \sum_{i=1}^n \mathbf{Var}[X_i] + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}[X_i, X_j]. \end{aligned}$$

\square

Beispiel 2.23 (Runs in einem p -Münzwurf). Sei $\underline{X} = (X_1, \dots, X_n)$ ein p -Münzwurf. Ein *Run* ist ein maximaler Block aus 0ern oder 1ern, also enthält 1100010 genau vier Runs. Sei Y die Anzahl der Runs in \underline{X} . Dann gilt

$$Y = 1 + \sum_{i=2}^n 1_{E_i} \text{ mit } E_i = \{X_i \neq X_{i-1}\}.$$

Nun gilt

$$\begin{aligned} \mathbf{E}[1_{E_i}] &= \mathbf{P}[X_i \neq X_{i-1}] = 2pq, \\ \mathbf{Var}[1_{E_i}] &= \mathbf{E}[1_{E_i}] - \mathbf{E}[1_{E_i}]^2 = 2pq - (2pq)^2 = 2pq(1 - 2pq). \end{aligned}$$

Für $2 \leq i \leq n-1$ ist außerdem

$$\mathbf{E}[1_{E_i} \cdot 1_{E_{i+1}}] = \mathbf{P}[X_{i+1} \neq X_i \neq X_{i-1}] = pqp + qpq = pq.$$

Offenbar sind $1_{E_i}, 1_{E_j}$ für $|i - j| > 1$ unabhängig. Damit gilt

$$\begin{aligned} \mathbf{E}[Y] &= 1 + 2pq(n-1), \\ \mathbf{Var}[Y] &= \sum_{i=2}^n \mathbf{Var}[1_{E_i}] + 2 \sum_{i=2}^{n-1} \mathbf{Cov}[1_{E_i}, 1_{E_{i+1}}] \\ &= 2pq(1 - 2pq)(n-1) + pq(1 - 4pq)(n-2). \end{aligned}$$

3 Weitere wichtige Verteilungen

3.1 Die hypergeometrische Verteilung

Gegeben sei folgende Situation: In einer Urne befinden sich insgesamt g Kugeln, wovon w weiß sind. Wir ziehen n Kugeln blind und ohne Zurücklegen heraus. Sei nun

$X :=$ Anzahl der weißen Kugeln in der Stichprobe.

Wir fragen nun nach der Verteilung von X . Um die Wahrscheinlichkeit $\mathbf{P}[X = k]$ zu berechnen, stellen wir uns vor, dass die weißen Kugeln von 1 bis w und die anderen Kugeln von $w+1$ bis g nummeriert sind. Insgesamt ist jede mögliche Ziehung (wenn man die Kugeln nur nach Nummern unterscheidet) gleichwahrscheinlich und es gibt genau $\binom{g}{n}$ mögliche Ziehungen. (Wir unterscheiden ja die Reihenfolge nicht.) Wie viele dieser Möglichkeiten führen nun gerade dazu, dass man n weiße Kugeln zieht? Hierzu muss man gerade k der w weißen und $n-k$ der $g-w$ andersfarbigen Kugeln ziehen. Deshalb gibt es hierfür $\binom{w}{k} \binom{g-w}{n-k}$ Möglichkeiten. Weil alle Möglichkeiten gleich wahrscheinlich sind, gilt also

$$\mathbf{P}[X = k] = \frac{\binom{w}{k} \binom{g-w}{n-k}}{\binom{g}{n}}.$$

Definition 3.1 (Hypergeometrische Verteilung). Sei $n, g \in \mathbb{N}$ und $w \in \mathbb{Z}_+$ mit $n, w \leq g$. Eine $\{0, \dots, n\}$ -wertige Zufallsvariable X heißt hypergeometrisch verteilt mit n, g, w , wenn

$$\mathbf{P}[X = k] = \frac{\binom{w}{k} \binom{g-w}{n-k}}{\binom{g}{n}}, \quad k = 0, 1, \dots, n. \quad (3.1)$$

Wir schreiben dann auch $X \sim \text{Hyp}(n, g, w)$.

Proposition 3.2 (Erwartungswert und Varianz der hypergeometrischen Verteilung). Sei $X \sim \text{Hyp}(n, g, w)$. Dann gilt

$$\mathbf{E}[X] = np, \quad \mathbf{Var}[X] = npq \left(1 - \frac{n-1}{g-1}\right)$$

mit $p := \frac{w}{g}$, $q := 1 - p$.

Beweis. Wir stellen uns die oben beschriebene Urne mit den g Kugeln vor. Sei

$$Z_i := 1_{\{i\text{-te gezogene Kugel ist weiß}\}},$$

also

$$X = \sum_{i=1}^n Z_i.$$

Klar ist, dass $\mathbf{P}[Z_i = 1] = p$, $i = 1, \dots, n$ und damit

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[Z_i] = np.$$

Weiter gilt für $i \neq j$

$$\mathbf{Cov}[Z_i, Z_j] = \mathbf{P}[Z_i = Z_j = 1] - p^2 = p \left(\frac{w-1}{g-1} - \frac{w}{g} \right) = -p \frac{g-w}{g(g-1)} = -pq \frac{1}{g-1},$$

also nach Proposition 2.22

$$\begin{aligned} \mathbf{Var}[X] &= \sum_{i=1}^n \mathbf{Var}[Z_i] + 2 \sum_{1 \leq i < j \leq n} \mathbf{Cov}[Z_i, Z_j] \\ &= npq - n(n-1)pq \frac{1}{g-1} = npq \left(1 - \frac{n-1}{g-1}\right). \end{aligned}$$

□

Bemerkung 3.3 (Alternative Berechnung). Wir können Erwartungswert und Varianz von X alternativ auch direkt berechnen. Hierzu bemerken wir zunächst, dass

$$\sum_{k=1}^n \binom{w}{k} \binom{g-w}{n-k} = \binom{g}{w},$$

andernfalls würden die Wahrscheinlichkeiten in (3.1) nicht zu 1 addieren. Nun schreiben wir

$$\begin{aligned} \mathbf{E}[X] &= \sum_{k=0}^n k \frac{\binom{w}{k} \binom{g-w}{n-k}}{\binom{g}{n}} = w \sum_{k=1}^n \frac{\binom{w-1}{k-1} \binom{g-w}{n-k}}{\binom{g}{n}} = w \frac{\binom{g-1}{n-1}}{\binom{g}{n}} = n \frac{w}{g} = np, \\ \mathbf{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) \frac{\binom{w}{k} \binom{g-w}{n-k}}{\binom{g}{n}} \\ &= w(w-1) \sum_{k=2}^n \frac{\binom{w-2}{k-2} \binom{g-w}{n-k}}{\binom{g}{n}} = w(w-1) \frac{\binom{g-2}{n-2}}{\binom{g}{n}} = np \frac{(n-1)(w-1)}{g-1}, \end{aligned}$$

also

$$\begin{aligned}\mathbf{Var}[X] &= \mathbf{E}[X(X-1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 = np \frac{(n-1)(w-1)}{g-1} + np - (np)^2 \\ &= np \left(1 - \frac{nw}{g} + \frac{(n-1)(w-1)}{g-1} \right) = np \frac{g(g-1) + nw - (n+w-1)g}{g(g-1)} = npq \frac{g-n}{g-1}.\end{aligned}$$

Bemerkung 3.4 (Ziehen mit und ohne Zurücklegen). Eine Alternative zum Ziehen *ohne* Zurücklegen ist natürlich das Ziehen *mit* Zurücklegen. Führt man dies n -mal durch, und fragt sich bei g Kugeln in der Urne, von denen genau w weiß sind, nach der Anzahl Y der weißen Kugeln in der Stichprobe, so erhält man $Y \sim B(n, p)$ mit $p = \frac{w}{g}$. Damit ist also

$$\mathbf{E}[X] = \mathbf{E}[Y] = np,$$

sowie

$$\mathbf{Var}[X] = npq \left(1 - \frac{n-1}{g-1} \right), \quad \mathbf{Var}[Y] = npq.$$

Dies kann man etwa so interpretieren: Ist g groß, enthält die Urne also viele Kugeln, von denen etwa ein Anteil p weiß ist, und zieht man $n \ll g$ Kugeln ohne oder mit Zurücklegen heraus, so sind die Varianzen (und auch die Verteilungen) recht ähnlich.

Beispiel 3.5 (Lotto). Die Gewinnwahrscheinlichkeiten im Lotto kann man durch die hypergeometrische Verteilung beschreiben. Hier gibt es $g = 49$ Kugeln, von denen $w = 6$ die Gewinnzahlen sind. Füllt man einen Spielzettel aus, so kreuzt man $n = 6$ Zahlen an. Deshalb ist z.B. die Wahrscheinlichkeit, dass gerade $k = 4$ Zahlen richtig sind gerade

$$\frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} \approx 9.69 \cdot 10^{-4}.$$

3.2 Die Poisson-Verteilung und das Gesetz der kleinen Zahlen

Wir stellen uns (wie bei der Einführung der Binomialverteilung) vor, dass ein Zufallsexperiment unabhängig n -mal hintereinander ausgeführt wird, wobei in jedem Versuch mit Wahrscheinlichkeit p ein Erfolg zu verzeichnen ist. Wir betrachten nun den Fall großer n und kleiner p . Man denke hierbei etwa daran, dass viele Leute ein Glücksspiel (z.B. Lotto) spielen, das für jede Person nur eine sehr kleine Gewinnwahrscheinlichkeit hat. Nachwievor ist natürlich die Anzahl der Gewinner $X \sim B(n, p)$, also in Erwartung $\mathbf{E}[X] = np$. Es stellt sich nun heraus, dass man die Verteilung von X approximieren kann.

Proposition 3.6 (Gesetz der kleinen Zahlen). Sei X_n verteilt nach $B(n, p_n)$ für $n = 1, 2, \dots$, so dass

$$\mathbf{E}[X_n] = n \cdot p_n \xrightarrow{n \rightarrow \infty} \lambda > 0.$$

Dann gilt

$$\lim_{n \rightarrow \infty} \mathbf{P}[X_n = k] = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Beweis. Wir schreiben direkt

$$\begin{aligned}\mathbf{P}[X_n = k] &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{n^k} \cdot \frac{1}{k!} (np_n)^k \left(1 - \frac{np_n}{n} \right)^n (1-p_n)^{-k}.\end{aligned}$$

Für jedes k gilt nun $\frac{n(n-1)\cdots(n-k+1)}{n^k} \xrightarrow{n \rightarrow \infty} 1$, $(1 - np_n/n)^n \xrightarrow{n \rightarrow \infty} e^{-\lambda}$ und $(1 - p_n)^{-k} \xrightarrow{n \rightarrow \infty} 1$. Insgesamt ergibt sich damit die Behauptung. \square

Definition 3.7 (Poisson-Verteilung). Sei $\lambda \geq 0$ und X eine \mathbb{Z}_+ -wertige Zufallsvariable mit

$$\mathbf{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!},$$

so heißt die Verteilung von X Poisson-Verteilung zum Parameter λ und wir schreiben $X \sim \text{Poi}(\lambda)$.

Proposition 3.8 (Erwartungswert und Varianz der Poisson-Verteilung). Sei $\lambda \geq 0$ und $X \sim \text{Poi}(\lambda)$. Dann gilt

$$\mathbf{E}[X] = \mathbf{Var}[X] = \lambda.$$

Beweis. Wir schreiben, wegen $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda$

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda$$

und mit Proposition 2.22 folgt ebenso

$$\begin{aligned} \mathbf{Var}[X] &= \mathbf{E}[X(X-1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 \\ &= \left(\sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} \right) + \lambda - \lambda^2 \\ &= \left(\lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \right) + \lambda - \lambda^2 = \lambda. \end{aligned}$$

\square

Bemerkung 3.9 (Der Checkpot im Lotto). Um nun das Gesetz der kleinen Zahlen aus Proposition 3.6 anwenden zu können, betrachten wir das Lotto-Spiel 6 aus 49. Die Gewinnchancen für sechs Richtige sind hierbei wie in Beispiel 3.5

$$p := \frac{1}{\binom{49}{6}} = \frac{1}{13\,983\,816} \approx 7.15 \cdot 10^{-8}.$$

Geht man davon aus, dass an einer Ausspielung 10^7 Spiele abgegeben werden, gibt es (in Erwartung) etwa $\lambda := 0.715$ Sechser. Genauer können wir sagen, dass etwa die Anzahl X der gespielten Sechser Poisson-verteilt ist zum Parameter λ . Deshalb gilt z.B. für die Wahrscheinlichkeit, dass der Checkpot nicht geknackt wird

$$\mathbf{P}[X = 0] \approx e^{-\lambda} \approx 0.49.$$

3.3 Die geometrische und die Exponentialverteilung

Man betrachte einen (unendlichen) p -Münzwurf X_1, X_2, \dots . Wir wissen zwar schon, dass die Anzahl der Köpfe in den ersten n Würfeln gerade $B(n, p)$ verteilt ist, jedoch noch nicht, wie lange man auf das erste Mal Kopf warten muss. Die Verteilung dieser Wartezeit ist die geometrische Verteilung.

Definition 3.10 (Geometrische Verteilung). Sei $p \in (0, 1)$ und X eine \mathbb{N} -wertige Zufallsvariable mit

$$\mathbf{P}[X = k] = (1 - p)^{k-1}p, \quad k = 1, 2, \dots,$$

so heißt die Verteilung von X geometrische Verteilung zum Parameter p und wir schreiben $X \sim \text{geo}(p)$.

Proposition 3.11 (Eigenschaften der geometrischen Verteilung). Sei $X \sim \text{geo}(p)$. Dann ist

$$\begin{aligned} \mathbf{P}[X > k] &= (1 - p)^k, & k = 1, 2, \dots \\ \mathbf{E}[X] &= \frac{1}{p}, \\ \mathbf{V}[X] &= \frac{1 - p}{p^2}. \end{aligned}$$

Beweis. Zunächst ist

$$\mathbf{P}[X > k] = \sum_{x=k+1}^{\infty} (1 - p)^{x-1}p = p \sum_{x=k}^{\infty} (1 - p)^x = p(1 - p)^k \sum_{x=0}^{\infty} (1 - p)^x = (1 - p)^k.$$

Für den Erwartungswert und die Varianz verwenden wir Lemma 2.10 und schreiben

$$\begin{aligned} \mathbf{E}[X] &= \sum_{x=0}^{\infty} (1 - p)^x = \frac{1}{p}, \\ \mathbf{E}[X(X - 1)] &= 2 \sum_{x=0}^{\infty} x(1 - p)^x = 2 \frac{1 - p}{p} \sum_{x=0}^{\infty} x(1 - p)^{x-1}p = 2 \frac{1 - p}{p} \mathbf{E}[X] = 2 \frac{1 - p}{p^2}, \\ \mathbf{Var}[X] &= \mathbf{E}[X(X - 1)] + \mathbf{E}[X] - \mathbf{E}[X]^2 = 2 \frac{1 - p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1 - p}{p^2}. \end{aligned}$$

□

Ist der Erfolgsparameter p einer geometrisch verteilten Zufallsvariable klein, so muss man recht lange auf den ersten Erfolg warten. Dies lässt sich – übrigens ganz ähnlich wie in Beispiel 1.5 – zu einer Grenzwertaussage formalisieren.

Lemma 3.12 (Exponentialapproximation). Sei $X_n \sim \text{geo}(p_n)$ für $n = 1, 2, \dots$. Dann gilt für $x \geq 0$

$$\lim_{n \rightarrow \infty} \mathbf{P}[p_n X_n > x] = e^{-x}.$$

Beweis. Wir schreiben einfach

$$\mathbf{P}[p_n X_n > x] = \mathbf{P}[X_n > x/p_n] = (1 - p_n)^{x/p_n} \xrightarrow{n \rightarrow \infty} e^{-x}.$$

□

Definition 3.13 (Exponentialverteilung). Sei $\lambda \geq 0$ und X eine \mathbb{R}_+ -wertige Zufallsvariable mit Dichte

$$\mathbf{P}[X \in dx] = \lambda e^{-\lambda x} dx,$$

so heißt die Verteilung von X Exponentialverteilung zum Parameter λ und wir schreiben $X \sim \text{Exp}(\lambda)$.

Proposition 3.14 (Erwartungswert und Varianz der Exponentialverteilung). Für $\lambda > 0$ sei X nach $\text{Exp}(\lambda)$ verteilt. Dann gilt

$$\begin{aligned}\mathbf{E}[X] &= \frac{1}{\lambda}, \\ \mathbf{Var}[X] &= \frac{1}{\lambda^2}.\end{aligned}$$

Beweis. Mittels partieller Integration gilt

$$\begin{aligned}\mathbf{E}[X] &= \lambda \int_0^\infty x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}, \\ \mathbf{E}[X^2] &= \lambda \int_0^\infty x^2 e^{-\lambda x} dx = -x^2 e^{-\lambda x} \Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx = \frac{2}{\lambda^2}, \\ \mathbf{Var}[X] &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{1}{\lambda^2}.\end{aligned}$$

□

3.4 Die Normalverteilung

Die Normalverteilung – besonders bekannt durch die Gauss'sche Glockenkurve – spielt in statistischen Anwendungen eine große Rolle. Grund hierfür ist der zentrale Grenzwertsatz, den wir in Proposition 4.3 und Bemerkung 4.5 kennen lernen werden. Es folgt zunächst nur eine Definition.

Definition 3.15. Sei $\mu \in \mathbb{R}, \sigma^2 > 0$ und X eine reellwertige Zufallsvariable Z mit Dichte

$$\mathbf{P}[X \in dx] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \quad (3.2)$$

so heißt die Verteilung von X Normalverteilung mit Parametern μ und σ^2 und wir schreiben $X \sim N(\mu, \sigma^2)$.

Bemerkung 3.16. Aus der Analysis bekannt ist

$$\int \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}.$$

Damit zeigt man durch Substitution $z = (x - \mu)/\sqrt{\sigma^2}$ leicht, dass es sich bei (3.2) um eine Dichte handelt.

Proposition 3.17. Sei $X \sim N(\mu, \sigma^2)$. Dann ist

$$\frac{X - \mu}{\sqrt{\sigma^2}} \sim N(0, 1)$$

und

$$\begin{aligned}\mathbf{E}[X] &= \mu, \\ \mathbf{Var}[X] &= \sigma^2.\end{aligned}$$

Beweis. Für die erste Aussage schreiben wir mit $Z \sim N(0, 1)$

$$\begin{aligned} \mathbf{P}\left(\frac{X - \mu}{\sqrt{\sigma^2}} \leq x\right) &= \mathbf{P}\left(X \leq x\sqrt{\sigma^2} + \mu\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^{x\sqrt{\sigma^2} + \mu} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) dy \\ &= \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz = \mathbf{P}(Z \leq x). \end{aligned}$$

Zunächst ist mit partieller Integration

$$\begin{aligned} \int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2}\right) dx &= \exp\left(-\frac{x^2}{2}\right) \Big|_{-\infty}^{\infty} = 0, \\ \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx &= -x \exp\left(-\frac{x^2}{2}\right) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}, \end{aligned}$$

also $\mathbf{E}[Z] = 0$, $\mathbf{Var}[Z] = 1$. Daraus und aus der ersten Aussage folgt

$$\begin{aligned} \mathbf{E}[X] &= \sqrt{\sigma^2} \mathbf{E}\left[\frac{X - \mu}{\sqrt{\sigma^2}}\right] + \mu = \sqrt{\sigma^2} \mathbf{E}[Z] + \mu = \mu, \\ \mathbf{Var}[X] &= \sigma^2 \mathbf{Var}\left[\frac{X - \mu}{\sqrt{\sigma^2}}\right] = \sigma^2 \mathbf{Var}[Z] = \sigma^2. \end{aligned}$$

□

3.5 Erzeugung von Zufallszahlen

Will man ein stochastischen System numerisch untersuchen bieten sich mehrere Ansätze an. Als Beispiel stelle man sich die (zufällige) Tiefe eines Baumes im Quicksort Algorithmus vor. Die Verteilung dieser Tiefe ist schwer explizit zu bestimmen, jedoch kann man den Quicksort-Algorithmus (mit zufälliger Eingabesequenz) einfach simulieren, und damit auch die Verteilung der Tiefe des Baumes.

Um stochastische Systeme simulieren zu können, ist es notwendig Zufallsvariablen mit bestimmten Verteilungen mittels Zufallsgeneratoren zu ziehen. Während wir hier nicht über die Güte von Zufallsgeneratoren diskutieren wollen, stellen wir fest, dass diese meist uniform auf $[0, 1]$ verteilte Zufallszahlen liefern. Ausgehen von diesen kann man mittels des nächsten Satzes nach beliebigen Verteilungen verteilte Zufallszahlen generieren. Eine Illustration hierzu findet sich in Abbildung 3.1.

Theorem 3.18 (Simulationslemma). *Sei X eine (diskrete oder stetige) reellwertige Zufallsvariable mit Verteilungsfunktion F . Wir definieren die Pseudoinverse von F für $x \in [0, 1]$ durch*

$$F^{-1}(x) := \inf\{q : F(q) \geq x\}.$$

Dann ist $F^{-1}(U)$ genauso verteilt wie X .

Beweis. Wir verwenden, dass die Verteilungsfunktion F die Verteilung der Zufallsvariable X eindeutig bestimmt. Da F nicht notwendigerweise injektiv ist, muss F^{-1} nicht die Umkehrfunktion von F sein. Es gilt jedoch wegen der Konstruktion $F^{-1}(q) \leq x$ genau dann, wenn $q \leq F(x)$. Daraus folgt

$$\mathbf{P}[F^{-1}(U) \leq q] = \mathbf{P}[U \leq F(q)] = F(q).$$

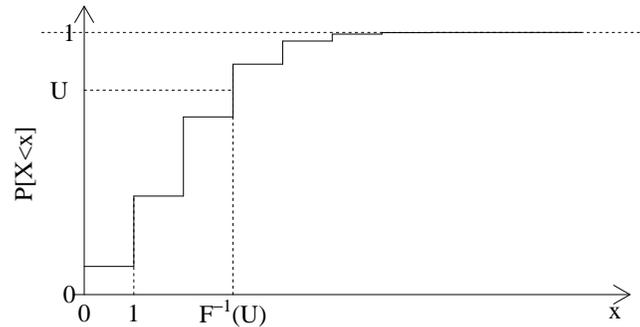


Abbildung 3.1: Illustration zum Simulationslemma, Theorem 3.18.

Das bedeutet, dass $F^{-1}(U)$ die Verteilungsfunktion F hat, woraus die Aussage folgt. \square

Bemerkung 3.19 (Anwendung). Angenommen, von einer Verteilung ist die Verteilungsfunktion F (und damit die Pseudoinverse F^{-1}) bekannt. Wir wollen unabhängige Zufallsvariablen X_1, X_2, \dots nach dieser Verteilung erzeugen. Verwenden dürfen wir hierzu jedoch nur die vom Computer bereit gestellten uniform verteilten Zufallsvariablen U_1, U_2, \dots . Das Simulationslemma besagt, dass in diesem Fall $F^{-1}(U_1), F^{-1}(U_2), \dots$ genau die gewünschte Verteilung besitzen.

Bemerkung 3.20 (Alternativen). Während das Verfahren aus Theorem 3.18 allgemeingültig ist, gibt es für spezielle Verteilungen schnellere Verfahren, Zufallszahlen zu erzeugen. Wir erwähnen hier (ohne Beweis) nur eine, nämlich die Erzeugung normalverteilter Zufallszahlen. Hierzu seien U_1, U_2 zwei unabhängige Zufallszahlen. Setzt man nun

$$X := \cos(2\pi U_1) \sqrt{-2 \log(U_2)},$$

so ist $X \sim N(0, 1)$.

4 Approximationssätze

Wir beschäftigen uns nun mit der Verteilung der Summe unabhängiger und identisch verteilter Zufallsvariablen. Dies ist eine Situation, die in der Praxis häufig vorkommt, etwa wenn man die Anzahl der Gewinne in einem Spiel zählt. Für unabhängige, identisch verteilte Zufallsvariable X_1, X_2, \dots sei nun $Y_n := X_1 + \dots + X_n$. Selbst bei bekannter Verteilung der X_i ist meistens die Verteilung von Y nicht bekannt. Allerdings gilt (falls Erwartungswert bzw. Varianz von X_1 existieren)

$$\begin{aligned} \mathbf{E}[Y_n] &= \mathbf{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{E}[X_i] = n\mathbf{E}[X_1], \\ \mathbf{Var}[Y_n] &= \mathbf{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbf{Var}[X_i] = n\mathbf{Var}[X_1]. \end{aligned} \tag{4.1}$$

Geht man also davon aus, dass die typische Schwankung der Zufallsvariable Y_n in etwa ihrer Standardabweichung entspricht, sieht man das \sqrt{n} -Gesetz: Y_n streut ihre Werte typischerweise in einem Bereich der Größenordnung \sqrt{n} .

4.1 Das schwache Gesetz der großen Zahlen

Das schwache Gesetz der großen Zahlen ist eine Konvergenzaussage für den Mittelwert von unabhängigen, identisch verteilten Zufallsgrößen. Als Beispiel betrachte man die relative Häufigkeit $\frac{1}{n} \sum_{i=1}^n X_i$ der Köpfe in einem p -Münzwurf X_1, X_2, \dots . Intuitiv klar ist, dass

$$\frac{1}{n} \sum_{i=1}^n X_i \approx p$$

gelten sollte. (Die relative Häufigkeit der Köpfe entspricht also in etwa der Wahrscheinlichkeit Kopf zu werfen.) In welchem Sinne diese Konvergenz zu verstehen ist, werden wir hier zwar nicht beleuchten können, das schwache Gesetz der großen Zahlen formalisiert diese Aussage jedoch. Eine Illustration findet sich in Abbildung 4.1. Zunächst benötigen wir zwei wichtige Ungleichungen.

Proposition 4.1 (Markov- und Tschebychev-Ungleichung). *Sei X eine \mathbb{R}_+ -wertige Zufallsvariable. Dann gilt für alle $\varepsilon > 0$ die Markov-Ungleichung*

$$\mathbf{P}[X \geq \varepsilon] \leq \frac{1}{\varepsilon} \mathbf{E}[X].$$

Existiert $\mu := \mathbf{E}[X]$, so gilt außerdem für $\varepsilon > 0$ die Tschebychev-Ungleichung

$$\mathbf{P}[|X - \mu| \geq \varepsilon] \leq \frac{\mathbf{Var}[X]}{\varepsilon^2}.$$

Beweis. Zum Beweis der Markov-Ungleichung bemerken wir $X \geq \varepsilon \cdot 1_{X \geq \varepsilon}$. Damit ist wegen der Monotonie des Erwartungswertes

$$\varepsilon \cdot \mathbf{P}[X \geq \varepsilon] = \mathbf{E}[\varepsilon \cdot 1_{X \geq \varepsilon}] \leq \mathbf{E}[X].$$

Die Tschebychev-Ungleichung folgt nun aus der Markov-Ungleichung, wenn man die Zufallsvariable $(X - \mu)^2$ betrachtet. \square

Theorem 4.2 (Schwaches Gesetz großer Zahlen). *Seien X_1, X_2, \dots reellwertige, unabhängige, identisch verteilte Zufallsvariable mit endlichem Erwartungswert $\mathbf{E}[X_1] = \mu$ und endlicher Varianz. Dann gilt für alle $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbf{P}\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right] = 0.$$

Beweis. Der Beweis erfolgt mittels der Chebyshev-Ungleichung und (4.1), denn

$$\mathbf{P}\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right] \leq \frac{\mathbf{Var}\left[\frac{1}{n}(X_1 + \dots + X_n)\right]}{\varepsilon^2} = \frac{\mathbf{Var}[X_1]}{n\varepsilon} \xrightarrow{n \rightarrow \infty} 0.$$

\square

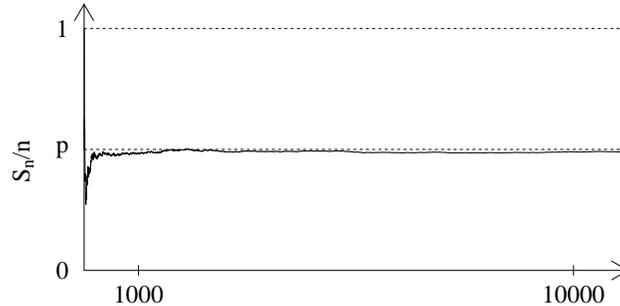


Abbildung 4.1: Illustration zum Gesetz der großen Zahlen, Theorem 4.2.

4.2 Der zentrale Grenzwertsatz und die Normalverteilung

Während das Gesetz der großen Zahlen eine Aussage über die Konvergenz des Mittelwertes von unabhängigen, identisch verteilten Zufallsvariablen X_1, X_2, \dots gegen deren Erwartungswert trifft, beschäftigt sich der zentrale Grenzwertsatz mit den Schwankungen in dieser Konvergenz. Da $\mathbf{Var}[\frac{1}{n}(X_1 + \dots + X_n)] = \mathbf{Var}X_1/n$, eine typische Schwankung also von der Größenordnung $1/\sqrt{n}$ ist, werden wir hier

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i - \mathbf{E}[X_i]}{\sqrt{n \mathbf{Var}[X_1]}}$$

betrachten. Obwohl es möglich ist, ein allgemeines Resultat zu beweisen, beschränken wir uns auf den Fall von $B(1, p)$ -verteilten Zufallsgrößen X_1, X_2, \dots und bemerken, dass $X_1 + \dots + X_n \sim B(n, p)$. Ohne Beweis werden wir außerdem die Stirling Formel zur Approximation von Fakultäten,

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

verwenden. Hierbei bedeutet \approx , dass

$$\frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} \xrightarrow{n \rightarrow \infty} 1.$$

Eine Illustration des Satzes findet sich in Abbildung 4.2.

Proposition 4.3 (Satz von deMoivre-Laplace). *Sei Z eine $N(0, 1)$ verteilte Zufallsvariable und X_1, X_2, \dots eine Folge binomialverteilter Zufallsvariable mit $\mathbf{Var}[X_n] \xrightarrow{n \rightarrow \infty} \infty$. Dann ist für $-\infty \leq c < d \leq \infty$*

$$\mathbf{P}\left[c \leq \frac{X_n - \mathbf{E}[X_n]}{\sqrt{\mathbf{Var}[X_n]}} \leq d\right] \xrightarrow{n \rightarrow \infty} \mathbf{P}[c \leq Z \leq d].$$

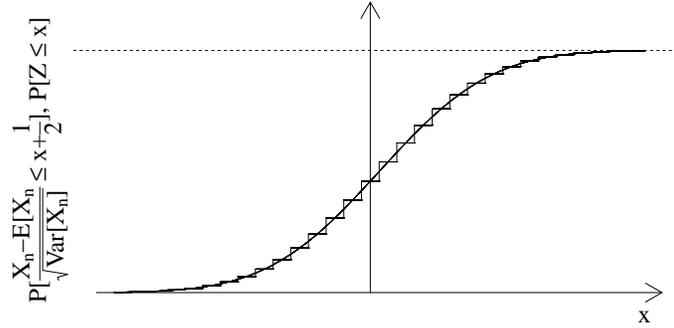


Abbildung 4.2: Illustration zum Satz von deMoivre-Laplace, Proposition 4.3. Gezeigt sind jeweils die Verteilungsfunktionen von $N(0,1)$ (die glatte Kurve) und eine Transformation von $B(100,0.5)$.

Beweisskizze. Mit der Stirling-Formel und $\eta(t) = t \ln \frac{t}{p} + (1-t) \ln \frac{1-t}{q}$ ist

$$\begin{aligned} \mathbf{P}[X = k] &= \binom{n}{k} p^k q^{n-k} \approx \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \\ &= \frac{1}{\sqrt{2\pi n \frac{k}{n} \frac{n-k}{n}}} \exp\left(-n\eta\left(\frac{k}{n}\right)\right) \\ &\approx \frac{1}{\sqrt{2\pi n \frac{k}{n} \frac{n-k}{n}}} \exp\left(-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2\right), \end{aligned}$$

da $\eta(p) = \eta'(p) = 0, \eta''(p) = \frac{1}{pq}$ mit einer Taylor-Entwicklung von η um p . Damit gilt für $\mu_n = np, \sigma_n^2 = npnq$ und $z_{x,n} = \frac{x-\mu_n}{\sqrt{\sigma_n^2}}$

$$\begin{aligned} \mathbf{P}\left(c \leq \frac{X_n - \mathbf{E}[X_n]}{\sqrt{\mathbf{Var}[X_n]}} \leq d\right) &= \sum_{k:c \leq z_{k,n} \leq d} \mathbf{P}[X_n = k] \\ &\approx \sum_{k:c \leq z_{k,n} \leq d} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{z_{k,n}^2}{2}\right) \approx \frac{1}{\sqrt{2\pi}} \int_c^d e^{-\frac{z^2}{2}} dz, \end{aligned}$$

wobei wir die letzte Summe als Riemannsumme interpretiert haben. □

Bemerkung 4.4 (Anwendung). Sei X eine $B(n, p)$ verteilte Zufallsgröße mit n groß, npq groß, sowie $c, d \in \mathbb{Z}$. Dann gilt

$$\begin{aligned} \mathbf{P}(c \leq X \leq d) &\approx \sum_{k=c}^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_{k,n}^2}{2}\right) (z_{k+\frac{1}{2},n} - z_{k-\frac{1}{2},n}) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{z_{c-1/2,n}}^{z_{d+1/2,n}} \exp\left(-\frac{z^2}{2}\right) dz = \Phi\left(\frac{d+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{c-\frac{1}{2}-np}{\sqrt{npq}}\right). \end{aligned}$$

Man beachte hierbei jeweils die Verschiebung um $\frac{1}{2}$ in der Binomialverteilung. Mit ihr erreicht man eine bessere Approximationsgüte. Diese Korrektur wurde etwa auch in Abbildung 4.2 angewandt.

Bemerkung 4.5 (Zentraler Grenzwertsatz). Allgemeiner als der Satz von deMoivre-Laplace gilt auch der Zentraler Grenzwertsatz:

Seien X_1, X_2, \dots reellwertige, unabhängige, identisch verteilte Zufallsvariablen mit endlichem Erwartungswert $\mathbf{E}[X_1] = \mu$ und Varianz $\mathbf{Var}[X_1] = \sigma^2$. Dann gilt für

$$Y_n^* := \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

und alle $-\infty \leq c < d \leq \infty$

$$\lim_{n \rightarrow \infty} \mathbf{P}[c \leq Y_n^* \leq d] = \mathbf{P}[c \leq Z \leq d],$$

wobei Z eine nach $N(0, 1)$ verteilte Zufallsvariable ist.

5 Markov-Ketten

Wir betreten nun den Bereich der stochastischen Prozesse. Ein stochastischer Prozess ist hierbei eine Familie $(X_t)_{t \in I}$ von Zufallsvariablen mit Indexmenge $I \subseteq \mathbb{R}$. Wir werden im Folgenden die Index-Menge $I = \mathbb{Z}_+$ betrachten. Markov-Ketten zeichnen sich dadurch aus, dass X_{t+1} nur von X_t abhängt (und nicht von den Zufallsvariablen X_0, \dots, X_{t-1}). Diese Abhängigkeit wird durch bedingte Wahrscheinlichkeiten beschrieben.

5.1 Bedingte Wahrscheinlichkeiten

Definition 5.1 (Bedingte Wahrscheinlichkeit). Seien X_1 und X_2 reellwertige Zufallsvariable mit Zielbereichen E_1 und E_2 .

1. Die bedingte Wahrscheinlichkeit von $\{X_2 \in A_2\}$, gegeben $\{X_1 \in A_1\}$, ist durch

$$\mathbf{P}[X_2 \in A_2 | X_1 \in A_1] := \frac{\mathbf{P}[X_1 \in A_1, X_2 \in A_2]}{\mathbf{P}[X_1 \in A_1]}$$

gegeben. Ist $\mathbf{P}[X_1 \in A_1] = 0$, so definieren wir die rechte Seite als 0.

2. Die Abbildung $A_2 \mapsto \mathbf{P}[X_2 \in A_2 | X_1 \in A_1]$ heißt die bedingte Verteilung von X_2 , gegeben $\{X_1 \in A_1\}$.
3. Ist X_1 diskret, so ist die bedingte Verteilung von X_2 gegeben X_1 die Abbildung

$$\mathbf{P}[X_2 \in A_2 | X_1] : \begin{cases} E_1 & \rightarrow [0; 1] \\ a_1 & \mapsto \mathbf{P}[X_2 \in A_2 | X_1 = a_1]. \end{cases}$$

Lemma 5.2 (Einfache Eigenschaften bedingter Wahrscheinlichkeiten). Seien X_1 und X_2 Zufallsvariable mit Zielbereichen E_1 und E_2 und $A_1 \subseteq E_1, A_2 \subseteq E_2$.

1. Es gilt

$$\mathbf{P}[X_2 \in A_2 | X_1 \in A_1] = \sum_{a_2 \in A_2} \mathbf{P}[X_2 = a_2 | X_1 \in A_1].$$

2. Die Zufallsvariablen X_1 und X_2 sind genau dann unabhängig, wenn

$$\mathbf{P}[X_2 \in A_2 | X_1 \in A_1] = \mathbf{P}[X_2 \in A_2]$$

für alle A_1, A_2 gilt. Ist X_1 diskret, so ist dies genau dann der Fall, wenn

$$\mathbf{P}[X_2 \in A_2 | X_1] = \mathbf{P}[X_2 \in A_2]$$

für alle A_2 gilt.

Beweis. Beide Eigenschaften folgen aus der Definition der bedingten Wahrscheinlichkeit. \square

Beispiel 5.3 (Gedächtnislosigkeit der geometrischen Verteilung). Sei $X \sim \text{geo}(p)$. Dann gilt

$$\mathbf{P}[X > i + j | X > i] = \mathbf{P}[X > j].$$

Dies interpretiert man so: Man führt unabhängige Experimente mit Erfolgswahrscheinlichkeit p hintereinander durch. Wenn i Versuche erfolglos verliefen, ist die Wahrscheinlichkeit für mindestens weitere j erfolglose Versuche genauso groß wie am Anfang.

Denn: Mit $q = 1 - p$ schreiben wir

$$\begin{aligned} \mathbf{P}[X > i + j | X > i] &= \frac{\mathbf{P}[X > i + j, X > i]}{\mathbf{P}[X > i]} = \frac{q^{i+j}}{q^i} = q^j \\ &= \mathbf{P}[X > j]. \end{aligned}$$

Eine analoge Aussage gilt übrigens auch für Exponentialverteilungen.

Theorem 5.4 (Formel für die totalen Wahrscheinlichkeit und Bayes'sche Formel). Seien X_1, X_2 diskrete Zufallsvariable mit Zielbereichen E_1 und E_2 . Dann gilt die Formel von der totalen Wahrscheinlichkeit

$$\mathbf{P}[X_2 \in A_2] = \sum_{a_1} \mathbf{P}[X_2 \in A_2 | X_1 = a_1] \cdot \mathbf{P}[X_1 = a_1].$$

für $A_2 \subseteq E_2$. Weiter gilt die Bayes'sche Formel für $A_1 \subseteq E_1$ mit $\mathbf{P}[X_1 \in A_1] > 0$

$$\mathbf{P}[X_1 \in A_1 | X_2 \in A_2] = \frac{\mathbf{P}[X_2 \in A_2 | X_1 \in A_1] \cdot \mathbf{P}[X_1 \in A_1]}{\sum_{a_1} \mathbf{P}[X_2 \in A_2 | X_1 = a_1] \cdot \mathbf{P}[X_1 = a_1]}.$$

Beweis. Die Formel von der totalen Wahrscheinlichkeit ergibt sich durch Einsetzen der Definition von $\mathbf{P}[X_2 \in A_2 | X_1 = a_1]$ in

$$\mathbf{P}[X_2 \in A_2] = \sum_{a_1} \mathbf{P}[X_2 \in A_2, X_1 = a_1].$$

In der Bayes'schen Formel ist der Zähler der rechten Seite gleich $\mathbf{P}[X_1 \in A_1, X_2 \in A_2]$ nach der Definition der bedingten Wahrscheinlichkeit, und der Nenner ist

$$\sum_{a_1} \mathbf{P}[X_2 \in A_2 | X_1 = a_1] \cdot \mathbf{P}[X_1 = a_1] = \sum_{a_1} \mathbf{P}[X_2 \in A_2, X_1 = a_1] = \mathbf{P}[X_2 \in A_2].$$

\square

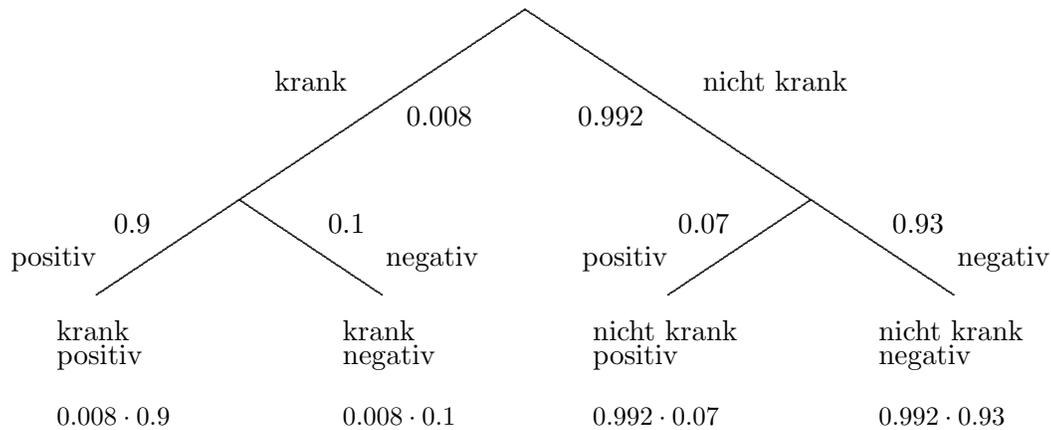


Abbildung 5.1: Ein Wahrscheinlichkeitsbaum für das Beispiel 5.5.

Beispiel 5.5 (Reihenuntersuchungen). In einer Reihenuntersuchung werden Personen auf eine bestimmte Krankheit getestet. Dabei kann es fälschlicherweise vorkommen, dass gesunde Personen durch den Test als krank eingestuft werden, oder dass kranke Personen nicht als solche erkannt werden.

In einer Population sind insgesamt 0.8% der Personen erkrankt. Eine kranke Person wird in 90% der Fälle positiv getestet (der Test fällt also positiv aus), eine gesunde Person mit 7%.

Wie groß ist die Wahrscheinlichkeit, dass eine positiv getestete Person wirklich krank ist?

Um dies zu beantworten, setzen wir $X = 1$ wenn die Person erkrankt ist (sonst $X = 0$) und $Y = 1$ wenn die Person positiv getestet wird (sonst $Y = 0$). Die genannten Angaben übersetzen wir zu

$$\mathbf{P}[X = 1] = 0.008, \quad \mathbf{P}[Y = 1|X = 1] = 0.9, \quad \mathbf{P}[Y = 1|X = 0] = 0.07.$$

Mit der Formel von Bayes ergibt sich für die Wahrscheinlichkeit, dass eine positiv getestete Person wirklich erkrankt ist

$$\begin{aligned} \mathbf{P}[X = 1|Y = 1] &= \frac{\mathbf{P}[Y = 1|X = 1] \cdot \mathbf{P}[X = 1]}{\mathbf{P}[Y = 1|X = 1] \cdot \mathbf{P}[X = 1] + \mathbf{P}[Y = 1|X = 0] \cdot \mathbf{P}[X = 0]} \\ &= \frac{0.9 \cdot 0.008}{0.9 \cdot 0.008 + 0.07 \cdot 0.992} \approx 0.0939 \end{aligned}$$

Ein positiver Test ist also nicht unbedingt ein sicheres Zeichen dafür, ob eine Person erkrankt ist.

Bemerkung 5.6 (Wahrscheinlichkeitsbäume). Bedingte Wahrscheinlichkeiten kann man durch Wahrscheinlichkeitsbäume darstellen. Siehe Abbildung 5.1.

5.2 Grundlegendes zu Markov-Ketten

Eine besondere Form der Abhängigkeit von Zufallsvariablen tritt in Markov-Ketten zu Tage. Hier werden der Reihe nach Zufallsvariablen X_0, X_1, \dots realisiert, und zwar so, dass X_{t+1} nur von X_t abhängt. Man sagt auch, X_{t+1} ist unabhängig von X_1, \dots, X_{t-1} , wenn X_t bekannt ist. Diese Form der Abhängigkeit wird häufig zur stochastischen Modellierung verwendet.

Beispielsweise könnte X_0, X_1, \dots der Preis einer Aktie an Tagen $0, 1, \dots$ sein. Die Markov-Eigenschaft besagt in diesem Fall, dass die Verteilung der Kursänderungen am Tag $t + 1$ nur davon abhängt, wie der Kurs X_t am Tag t war.

Wir beginnen mit der Definition von Markov-Ketten. Wir werden vor allem den Fall von homogenen Markov-Ketten behandeln. Bei solchen gibt es eine sich zeitlich nicht ändernde stochastische Übergangsvorschrift, wie die Verteilung des Zustandes X_{t+1} ist, wenn der Zustand X_t der Kette zur Zeit t bekannt ist. Diese Vorschrift wird mit Hilfe einer Matrix zusammengefasst, der Übergangsmatrix.

Definition 5.7 (Stochastischer Prozess und Markov-Ketten). 1. Seien I und E Mengen. Ein (E -wertiger) stochastischer Prozess (mit Indexmenge I) ist eine Familie von Zufallsvariablen $\mathcal{X} = (X_t)_{t \in I}$ mit Wertebereich E . Die Menge E heißt auch Zustandsraum von \mathcal{X} und I seine Indexmenge.

2. Sei $I = \{0, 1, 2, \dots\}$, E abzählbar und $\mathcal{X} = (X_t)_{t \in I}$ ein E -wertiger stochastischer Prozess. Falls

$$\mathbf{P}[X_{t+1} = i \mid X_0, \dots, X_t] = \mathbf{P}[X_{t+1} = i \mid X_t] \quad (5.1)$$

für alle $i \in E$, so heißt \mathcal{X} eine Markov-Kette. Sie heißt endlich, falls E endlich ist.

3. Sei $\mathcal{X} = (X_t)_{t \in I}$ eine E -wertige Markov-Kette. Existiert eine Matrix $P = (P_{ij})_{i,j \in E}$ mit

$$P_{ij} := \mathbf{P}[X_{t+1} = j \mid X_t = i]$$

(unabhängig von t), so heißt \mathcal{X} zeitlich homogen und P heißt Übergangsmatrix von \mathcal{X} .

Bemerkung 5.8 (Interpretation). 1. Die Eigenschaft (5.1) bedeutet in Worten: die zukünftige Entwicklung von \mathcal{X} nach t hängt von X_1, \dots, X_t nur durch den aktuellen Zustand X_t ab.

2. Sei P die Übergangsmatrix einer homogenen Markov-Kette \mathcal{X} mit Zustandsraum E . Dann gilt

$$\begin{aligned} 0 \leq P_{ij} \leq 1, & \quad i, j \in E, \\ \sum_{j \in E} P_{ij} = 1, & \quad i \in E. \end{aligned} \quad (5.2)$$

Die erste Eigenschaft ist klar, da die Einträge in P Wahrscheinlichkeiten sind. Außerdem ist

$$1 = \mathbf{P}[X_{t+1} \in E \mid X_t = i] = \sum_{j \in E} \mathbf{P}[X_{t+1} = j \mid X_t = i] = \sum_{j \in E} P_{ij}.$$

Matrizen P mit den Eigenschaften (5.2) heißen *stochastische Matrizen*.

3. Sei \mathcal{X} eine homogene Markov-Kette mit Übergangsmatrix P . Definiere einen gewichteten, gerichteten Graphen (E, K, W) wie folgt: die Menge der Knoten ist E , die Menge der (gerichteten) Kanten ist $K := \{(i, j) : P_{ij} > 0\}$. Das Gewicht der Kante (ij) ist $w_{(ij)} := P_{ij}$ und $W = (w_{(ij)})_{(ij) \in K}$. Der Graph (E, K, W) heißt *Übergangsgraph* von \mathcal{X} .

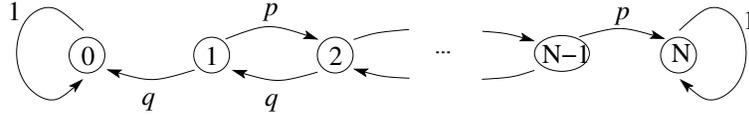


Abbildung 5.3: Übergangsgraph der Markov-Kette aus Beispiel 5.10.

Wir bezeichnen mit

$$p_n := \mathbf{P}[X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n]$$

die Wahrscheinlichkeit, dass Spieler 1 gewinnt. Wir werden zeigen, dass

$$p_n = \begin{cases} \frac{n}{N}, & p = q = \frac{1}{2}, \\ \frac{1 - (q/p)^n}{1 - (q/p)^N}, & \text{sonst.} \end{cases} \quad (5.4)$$

Zunächst gilt für $n = 1, \dots, N-1$

$$\begin{aligned} p_n &= q \cdot \mathbf{P}[X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n, X_1 = n-1] + p \cdot \mathbf{P}[X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n, X_1 = n+1] \\ &= q \cdot \mathbf{P}[X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n-1] + p \cdot \mathbf{P}[X_t \xrightarrow{t \rightarrow \infty} N \mid X_0 = n+1] \\ &= q \cdot p_{n-1} + p \cdot p_{n+1}, \end{aligned}$$

also mit $\Delta p_n := p_n - p_{n-1}$

$$q\Delta p_n = p\Delta p_{n+1}.$$

Im Fall $p = q = \frac{1}{2}$ folgt mit $\sum_{m=1}^N \Delta p_m = p_N - p_0 = 1$ daraus bereits

$$p_n = \frac{\sum_{m=1}^n \Delta p_m}{\sum_{m=1}^N \Delta p_m} = \frac{n\Delta p_1}{N\Delta p_1} = \frac{n}{N}.$$

Im Fall $p \neq q$ setzen wir $u := \frac{q}{p}$ und berechnen iterativ

$$\Delta p_n = u\Delta p_{n-1} = u^2\Delta p_{n-2} \cdots = u^{n-1}\Delta p_1 = u^{n-1}p_1.$$

Weiter ist

$$1 = \sum_{m=1}^N \Delta p_m = p_1 \sum_{m=0}^{N-1} u^m = p_1 \frac{1 - u^N}{1 - u}.$$

Also

$$p_n = \sum_{m=1}^n \Delta p_m = p_1 \sum_{m=1}^n u^{m-1} = \frac{1-u}{1-u^N} \frac{1-u^n}{1-u} = \frac{1-u^n}{1-u^N}$$

und die Behauptung ist gezeigt. \square

Beispiel 5.11 (Ehrenfest'sche Urne). Betrachte folgendes Urnenmodell: In einer Urne gibt es zwei durch eine Trennwand getrennte Kammern. Insgesamt liegen n Kugeln in den beiden Kammern. Wir ziehen eine Kugel rein zufällig aus der Urne und legen sie anschließend in die andere Kammer zurück. In Abbildung 5.4 findet sich eine Illustration.

Wir betrachten den $\{0, \dots, n\}$ -wertigen stochastischen Prozess $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$, wobei X_t die Anzahl der Kugeln in der linken Kammer nach dem t -Schritt darstellt. Dann ist \mathcal{X} eine homogene Markov-Kette, denn der Ausgang des Schrittes $t+1$ hängt nur von X_t ab. Die Übergangsmatrix von \mathcal{X} ist gegeben durch

$$P_{ij} = \begin{cases} \frac{i}{n}, & j = i - 1, \\ \frac{n-i}{n}, & j = i + 1, \\ 0, & \text{sonst.} \end{cases} \quad (5.5)$$

□

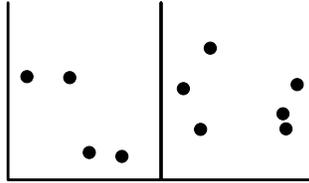


Abbildung 5.4: Die Ehrenfest'sche Urne mit $n = 10$ Kugeln aus Beispiel 5.11. Im nächsten Schritt wird mit Wahrscheinlichkeit $\frac{4}{10}$ eine Kugel von der linken Kammer in die rechte und mit Wahrscheinlichkeit $\frac{6}{10}$ von der rechten in die linke Kammer gelegt.

Wir kommen nun zu einem wichtigen Zusammenhang zwischen der Verteilung einer Markov-Ketten zum Zeitpunkt t und Potenzen der Übergangsmatrix.

Theorem 5.12. Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine homogene Markov-Kette mit Übergangsmatrix P und $\mu^{(t)} = (\mu^{(t)}(i))_{i \in E}$,

$$\mu^{(t)}(i) = \mathbf{P}[X_t = i]$$

für $t = 0, 1, 2, \dots$. Dann gilt

$$\mu^{(t)} = \mu^{(0)} P^t, \quad (5.6)$$

für $t = 0, 1, 2, \dots$, wobei die rechte Seite als Multiplikation des Zeilenvektors $\mu^{(0)}$ und der Matrix $P^t = \underbrace{P \cdots P}_{t \text{ mal}}$ zu verstehen ist.

Beweis. Der Beweis geht mittels Induktion über t . Für $t = 0$ ist die Aussage klar. Ist (5.6) für t gezeigt, so gilt

$$\begin{aligned} \mu^{(t+1)}(j) &= \mathbf{P}[X_{t+1} = j] = \sum_{i \in E} \mathbf{P}[X_{t+1} = j \mid X_t = i] \cdot \mathbf{P}[X_t = i] \\ &= \sum_{i \in E} \mu^{(t)}(i) \cdot P_{ij} = \sum_{i \in E} (\mu^{(0)} P^t)_i \cdot P_{ij} = (\mu^{(0)} P^{t+1})_j. \end{aligned} \quad \square$$

Beispiel 5.13 (Irrfahrt im Dreieck). Betrachte die Markov-Kette \mathcal{X} aus Beispiel 5.9 und Übergangsmatrix P aus (5.3). Sei $X_0 = 1$, die Markov-Kette startet also in 1, d.h. $\mu^{(0)} = (1, 0, 0)$. Weiter berechnen wir

$$P = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 2pq & q^2 & p^2 \\ p^2 & 2pq & q^2 \\ q^2 & p^2 & 2pq \end{pmatrix}$$

und damit

$$\mu^{(2)} = (2pq, q^2, p^2).$$

Also ist etwa $\mathbf{P}[X_2 = 1] = 2pq$. Dies ist klar, weil $X_2 = 1$ genau dann, wenn die ersten beiden Schritte im Übergangsgraphen aus Abbildung 5.2 einmal mit und einmal entgegen dem Uhrzeigersinn gingen. Es ist $X_2 = 3$ genau dann, wenn zwei Schritte im Uhrzeigersinn realisiert wurden, was Wahrscheinlichkeit p^2 hat. \square

5.3 Stationäre Verteilungen

Wir beginnen nun, uns für Markov-Ketten zu interessieren, die schon lange gelaufen sind, also für X_t mit großem t . Solche Markov-Ketten können (mehr dazu im nächsten Abschnitt) so beschaffen sein, dass sich die Verteilung nicht mehr viel ändert. Verteilungen auf dem Zustandsraum E , die invariant sind gegenüber Schritten der Markov-Kette heißen stationär.

Definition 5.14 (Stationäre Verteilung). Sei \mathcal{X} eine homogene Markov-Kette mit Übergangsmatrix P und π eine Verteilung auf E , gegeben durch einen Vektor $\pi = (\pi_i)_{i \in E}$.

1. Gilt

$$\pi P = \pi,$$

so heißt π stationäre Verteilung von \mathcal{X} .

2. Gilt

$$\pi_i P_{ij} = \pi_j P_{ji}$$

für alle i, j , so heißt π reversible Verteilung.

Bemerkung 5.15 (Interpretation von reversiblen Verteilungen). Sei π reversibel, $\mathbf{P}[X_t = i] = \pi_i$ und $i_t, \dots, i_{t+s} \in E$. Dann gilt

$$\begin{aligned} \mathbf{P}[X_t = i_t, \dots, X_{t+s} = i_{t+s}] &= \pi_{i_t} P_{i_t, i_{t+1}} \cdots P_{i_{t+s-1}, i_{t+s}} \\ &= \pi_{i_{t+1}} P_{i_{t+1}, i_{t+2}} \cdots P_{i_{t+s-1}, i_{t+s}} \cdot P_{i_{t+1}, i_t} = \cdots \\ &= \pi_{i_{t+s}} P_{i_{t+s}, i_{t+s-1}} \cdots P_{i_{t+1}, i_t} \\ &= \mathbf{P}[X_t = i_{t+s}, \dots, X_{t+s} = i_t]. \end{aligned}$$

Das bedeutet, dass die Wahrscheinlichkeit, die Zustände i_t, \dots, i_{t+s} zu durchlaufen, dieselbe ist wie die, die Zustände in umgekehrter Reihenfolge (reversibel) zu durchlaufen.

Lemma 5.16 (Einfache Eigenschaften stationärer Verteilungen). Sei $\mathcal{X} = (X_t)_{t \in I}$ eine E -wertige Markov-Kette mit Übergangsmatrix P und π eine Verteilung auf E .

1. Sei π stationär und $\mathbf{P}[X_t = i] = \pi_i$. Dann gilt, dass

$$\mathbf{P}[X_{t+1} = j] = \mathbf{P}[X_t = j].$$

Das bedeutet, dass X_t und X_{t+1} (und damit auch X_{t+2}, X_{t+3}, \dots) identisch verteilt sind.

2. Ist π reversibel, dann ist π auch stationär.

Beweis. 1. Wir verwenden Theorem 6.9 und schreiben

$$\begin{aligned} \mathbf{P}[X_{t+1} = j] &= \sum_{i \in E} \mathbf{P}[X_t = i] \cdot \mathbf{P}[X_{t+1} = j \mid X_t = i] = \sum_{i \in E} \pi_i P_{ij} = (\pi P)_j = \pi_j \\ &= \mathbf{P}[X_t = j]. \end{aligned}$$

2. Ist π reversibel, so gilt

$$(\pi P)_j = \sum_{i \in E} \pi_i P_{ij} = \sum_{i \in E} \pi_j P_{ji} = \pi_j,$$

da P Zeilensumme 1 hat. □

Beispiel 5.17 (Irrfahrt im Dreieck). Betrachte die Irrfahrt auf dem Dreieck \mathcal{X} aus Beispiel 5.9 mit Übergangsmatrix P aus (5.3). Für $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ ist

$$\pi P = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix} = \pi.$$

Damit ist die Gleichverteilung auf $\{1, 2, 3\}$ stationäre Verteilung von \mathcal{X} . Allerdings ist π nicht reversibel (außer für $p = \frac{1}{2}$), denn es gilt etwa

$$\pi_1 P_{12} = \frac{1}{3}p \neq \frac{1}{3}q = \pi_2 P_{21}.$$

Letztere Beobachtung ist nicht erstaunlich: Sei etwa $p > \frac{1}{2}$. Dann erwarten wir, dass \mathcal{X} das Dreieck öfter im Uhrzeigersinn durchläuft als umgekehrt. Wäre nun π reversibel, würde daraus folgen (siehe Bemerkung 5.15), dass die Wahrscheinlichkeit für einen Durchlauf des Dreiecks gegen und im Uhrzeigersinn dieselbe ist. □

Beispiel 5.18 (Ruinproblem). Im Ruinproblem aus Beispiel 5.10 gibt es zwar stationäre Verteilungen, diese sind jedoch nicht sonderlich interessant. Jede Verteilung $\pi = (p', 0, \dots, 0, q')$ mit $p' \in (0, 1)$ und $q' = 1 - p'$ ist stationäre (aber nicht reversible) Verteilung, wie man leicht nachrechnet. Insbesondere sind $(1, 0, \dots, 0)$ und $(0, \dots, 0, 1)$ stationär. Das bedeutet, dass die Zustände $X_t = 0$ und $X_t = n$ von der Markov-Kette nicht mehr verlassen werden. Man sagt auch, 0 und n sind *Fallen* für \mathcal{X} . □

Beispiel 5.19 (Ehrenfest'sche Urne). Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ die Anzahl der Kugeln in der linken Kammer einer Ehrenfest'schen Urne, wie in Beispiel 5.11. Hier ist die Binomialverteilung $\pi = B(n, \frac{1}{2})$ reversible Verteilung für \mathcal{X} . Es gilt nämlich mit (5.5) für $i = 1, \dots, n$

$$\pi_i P_{i,i-1} = \binom{n}{i} \frac{1}{2^n} \frac{i}{n} = \binom{n}{i-1} \frac{1}{2^n} \frac{n-i+1}{n} = \pi_{i-1} P_{i-1,i}.$$

Die Tatsache, dass $B(n, p)$ stationär ist, interpretiert man am besten so: nach langer Zeit ist jede der Kugeln oft von der linken in die rechte Kammer und umgelegt worden, so dass jede Kugel mit Wahrscheinlichkeit $\frac{1}{2}$ jeweils in der linken und rechten Kammer liegt. Diese Verteilung ändert sich nicht mehr, weil in jedem Schritt nur eine der n Kugeln nochmals in die andere Kammer gelegt wird. \square

Wir geben nun noch – ohne Beweis – den wichtigen Markov-Ketten-Konvergenzsatz an. Dieser stellt Bedingungen auf, wann eine Markov-Kette eine eindeutige stationäre Verteilung besitzt. Konvergenz der Markov-Kette $\mathcal{X} = (X_t)_{t \in I}$ bedeutet dabei, dass $\mathbf{P}[X_t = i] \xrightarrow{t \rightarrow \infty} \pi(i)$ für diese stationäre Verteilung π gilt, unabhängig von der Verteilung von X_0 . Um den Satz zu formulieren, benötigen wir noch eine Definition.

Definition 5.20 (Irreduzibilität, Aperiodizität). Sei \mathcal{X} eine E -wertige homogene Markov-Kette.

1. Der Zustand $i \in E$ kommuniziert mit $j \in E$, falls es ein t gibt mit

$$\mathbf{P}[X_t = j \mid X_0 = i] > 0.$$

In diesem Fall schreiben wir $i \rightarrow j$. Falls $i \rightarrow j$ und $j \rightarrow i$, schreiben wir $i \leftrightarrow j$.

2. Falls $i \leftrightarrow j$ für alle $i, j \in E$, so heißt \mathcal{X} irreduzibel. Andernfalls heißt \mathcal{X} reduzibel.
3. Ein Zustand $i \in E$ heißt aperiodisch, falls

$$d(i) := \text{ggT}\{t : \mathbf{P}[X_t = i \mid X_0 = i] > 0\} = 1.$$

Andernfalls heißt i periodisch mit Periode $d(i)$.

4. Falls alle $i \in E$ aperiodisch sind, so heißt \mathcal{X} aperiodisch.

Bemerkung 5.21 (Irreduzibilität, Aperiodizität und die Übergangsmatrix).

1. Die Begriffe der Irreduzibilität und Aperiodizität lassen sich auch mittels der Übergangsmatrix P der Markov-Kette \mathcal{X} erklären. Es gilt etwa $i \rightarrow j$ genau dann, wenn es ein t gibt mit $(P^t)_{ij} > 0$. Außerdem ist $d(i) = \text{ggT}\{t : P_{ii}^t > 0\}$.
2. Es gibt einen einfachen Zusammenhang zwischen dem Begriff der kommunizierenden Zustände und dem Übergangsgraphen: ein Zustand i kommuniziert mit einem Zustand j , wenn man einen Pfad $i \rightarrow j$ im Übergangsgraphen der Markov-Kette findet. Das bedeutet, dass es eine endliche Folge $(i, i_1), (i_1, i_2), \dots, (i_{n-1}, i_n), (i_n, j)$ im Übergangsgraphen gibt, also der Zustand j von i aus durch eine endliche Folge von Kanten erreicht werden kann.
3. Sei \mathcal{X} eine reduzible Markov-Kette. Der Begriff der Reduzibilität erklärt sich so, dass sich die Markov-Kette \mathcal{X} auf einen Zustandsraum $E' \subseteq E$ reduzieren lässt. Das bedeutet, dass aus $X_0 \in E'$ folgt, dass $X_t \in E'$ für alle $t = 1, 2, \dots$

Beispiel 5.22 (Irrfahrt auf dem Dreieck). Sei \mathcal{X} die Irrfahrt auf dem Dreieck aus Beispiel 5.9. In Beispiel 5.13 haben wir ausgerechnet, dass $P^2 > 0$ ist. Weiter ist auch $P^3 > 0$, was bereits zeigt, dass \mathcal{X} sowohl irreduzibel als auch aperiodisch ist. \square

Beispiel 5.23 (Ruinproblem). Für die Markov-Kette \mathcal{X} aus dem Ruinproblem, Beispiel 5.10, gilt sicher, dass $0 \not\rightarrow i$ für $i = 1, \dots, N$, und $N \not\rightarrow i$ für $i = 0, \dots, N - 1$. Mit anderen Worten: Ist Spieler 1 pleite (d.h. $X_t = 0$ für ein t), so wird er nach den Spielregeln nie wieder Geld bekommen, d.h. alle Zustände $1, \dots, n$ sind für ihn unerreichbar. Das bedeutet also, \mathcal{X} ist reduzibel. \square

Beispiel 5.24 (Ehrenfest'sche Urne). Die Markov-Kette \mathcal{X} , die die Anzahl der Kugeln in der linken Kammer einer Ehrenfest'schen Urne beschreibt, ist irreduzibel. Betrachtet man nämlich Zustände i, j , etwa mit $i > j$, so genügt es ja, genau $i - j$ Kugeln nacheinander von der linken in die rechte Kammer zu legen. Allerdings ist die Markov-Kette periodisch. Sei etwa $X_0 = 2i$ gerade, dann folgt, dass X_1 ungerade (entweder $2i - 1$ oder $2i + 1$) ist. Damit kann nur zu geraden Zeiten t der Zustand $X_t = 2i$ eintreten und es ist $\text{ggT}\{t : \mathbf{P}(X_t = 2i | X_0 = 2i)\} = 2$. \square

Nun kommen wir also zum Markov-Ketten-Konvergenzsatz.

Theorem 5.25 (Markov-Ketten-Konvergenzsatz). Sei $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ eine endliche, homogene, irreduzible und aperiodische Markov-Kette. Dann hat \mathcal{X} genau eine stationäre Verteilung π und es gilt für alle $i \in E$

$$\mathbf{P}[X_t = i] \xrightarrow{t \rightarrow \infty} \pi(i). \quad (5.7)$$

Beweis. Der Beweis findet sich in allen gängigen Lehrbüchern, in denen Markov-Ketten behandelt werden. \square

Bemerkung 5.26. Vor allem ist an dem Resultat bemerkenswert, dass (5.7) nicht von der Verteilung von X_0 abhängt. Das bedeutet, dass die Markov-Kette nach langer Zeit *vergisst*, in welchem Zustand X_0 sie zur Zeit 0 gestartet ist.

Beispiel 5.27 (Irrfahrt auf dem Dreieck). Von unseren drei Beispielen erfüllt (nur) die Irrfahrt auf dem Dreieck \mathcal{X} aus Beispiel 5.9 die Voraussetzungen des Satzes. Nach Beispiel 5.22 ist \mathcal{X} irreduzibel und aperiodisch. Nach Beispiel 5.17 und obigem Satz ist die Gleichverteilung auf $\{1, 2, 3\}$ die einzige stationäre Verteilung für \mathcal{X} . Außerdem gilt

$$\mathbf{P}[X_t = i] \xrightarrow{t \rightarrow \infty} \frac{1}{3}.$$

\square

Beispiel 5.28 (Ruinproblem). Für die Markov-Kette \mathcal{X} aus dem Ruinproblem, Beispiel 5.10, gibt es stationäre Verteilungen π , nämlich solche mit $\pi(0) + \pi(N) = 1$. (Bei solchen ist einer der beiden Spielern pleite.) Offenbar gilt auch

$$\lim_{t \rightarrow \infty} \mathbf{P}[X_t = N | X_0 = n] = p_n$$

mit p_n aus (5.4). Das bedeutet, dass die (Verteilung der) Markov-Kette konvergiert, die Grenzverteilung aber vom Startwert (hier $X_0 = n$) abhängt. Dies ist im Einklang mit Theorem 5.25, da \mathcal{X} nicht irreduzibel ist.

Beispiel 5.29 (Ehrenfest'sche Urne). Die Markov-Kette \mathcal{X} , die die Anzahl der Kugeln in der linken Kammer einer Ehrenfest'schen Urne beschreibt, ist periodisch, und damit ist Theorem 5.25 nicht anwendbar. Ist etwa $X_0 = 2n$ gerade, so ist X_{2t} auch gerade, also $\mathbf{P}[X_{2t} \text{ gerade} | X_0 = 2n] = 1$, aber $\mathbf{P}[X_{2t} \text{ gerade} | X_0 = 2n + 1] = 0$. Insbesondere existiert der Grenzwert von $\mathbf{P}[X_t = k | X_0 = n]$ nicht.

6 Statistik

Es ist nicht übertrieben zu behaupten, dass in der heutigen Welt immer mehr *Daten* jeglicher Art erhoben werden. Diese zu ordnen und aus Daten Schlussfolgerungen zu ziehen ist Aufgabe der Statistik.

Man teilt dabei diese Aufgaben in zwei Gebiete auf. Die *deskriptive Statistik* dient rein der Beschreibung der Daten, etwa durch geeignete Wahl von Statistiken, die die Daten zusammenfassen. Anders ist dies bei der hier behandelten *schließenden* oder *induktiven Statistik*. Die Aufgabe ist hier, mit Hilfe von stochastischen Modellen Aussagen darüber zu treffen, welchen Annahmen den Daten zugrunde liegen könnten.

6.1 Grundlagen

Wir beginnen mit einem Beispiel.

Beispiel 6.1 (Erfolgswahrscheinlichkeit beim Münzwurf).

Bevor wir statistische Konzepte einführen, betrachten wir folgendes Beispiel: eine Münze wird 100 mal geworfen. Dabei ist die Wahrscheinlichkeit für *Kopf* (was wir im Folgenden als Erfolg werten wollen) noch unbekannt. Von den 100 Würfeln sind 59 ein Erfolg.

Unsere statistischen Überlegungen gehen nun von der Vorstellung aus, dass die 100 Münzwürfe die Realisierung einer Zufallsvariable $X = (X_1, \dots, X_{100})$ sind, wobei X_1, \dots, X_{100} unabhängig und identisch verteilt sind mit

$$X_i = \begin{cases} 1, & \text{falls der } i\text{-te Wurf Kopf zeigt,} \\ 0, & \text{sonst.} \end{cases}$$

und es gilt

$$\mathbf{P}[X_i = 1] = p.$$

Jetzt ist $X_1 + \dots + X_n$ die Gesamtzahl der Erfolge. Als Summe von n unabhängigen Bernoulli-verteilten Zufallsvariablen ist diese Summe $B(n = 100, p)$ -verteilt. Wichtig ist, dass zwar $n = 100$ bereits fest steht (schließlich wissen wir ja, dass wir 100 mal die Münze geworfen haben), nicht jedoch p . In dieser Situation gibt es zwei *statistische Probleme*.

- *Schätzproblem*: Wir können versuchen, den Erfolgsparameter p zu schätzen. Wir werden hierzu aus den Daten (59 Erfolge aus 100 Versuchen) einen Wert \hat{p} ableiten (*Punktschätzer*). Alternativ kann man auch ein Intervall $[a, b]$ angeben, in dem der wahre Parameter p mit hoher Wahrscheinlichkeit liegt (*Intervallschätzer*). Hierauf werden wir allerdings nicht näher eingehen.
- *Testproblem*: Stellen wir uns vor, der Werfer der Münze behauptet, dass die Münze fair ist, also $p = \frac{1}{2}$ gilt. Dieser Meinung können wir skeptisch gegenüber stehen, da ja sogar 59 aus 100 Würfeln ein Erfolg waren. Wir können versuchen, die Hypothese $p = \frac{1}{2}$ zu testen. Das bedeutet, dass wir untersuchen, wie gut die Hypothese mit den Daten in Einklang steht.

Wie im Beispiel gesehen besteht ein statistisches Experiment aus einer (oder mehreren) Zufallsvariablen (etwa die Anzahl der Erfolge) und verschiedenen (möglichen) Verteilungen der Zufallsvariable; hier $B(n = 100, p)$ für variierendes p . Dies führt zur Definition des statistischen Modells. Siehe auch Abbildung 6.1.

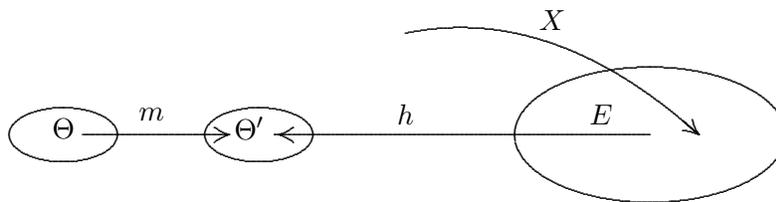


Abbildung 6.1: Veranschaulichung von statistischen Modellen

Definition 6.2 (Statistisches Modell). Seien E, Θ, Θ' Mengen. Ein statistisches Modell ist ein Paar $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$, wobei X eine Zufallsvariable mit Zielbereich E ist, bei deren Verteilung noch ein Parameter $\vartheta \in \Theta$ frei, also unbestimmt, ist. Das bedeutet, dass es eine Funktion $\vartheta \mapsto \rho_\vartheta$ gibt mit¹

$$\mathbf{P}_\vartheta[X \in da] = \rho_\vartheta(a)da.$$

Die Menge Θ heißt Parameterraum, die Menge E Beobachtungsraum. Jedes Zufallsvariable $h(X)$ mit $h : E \rightarrow \Theta'$ heißt Statistik.

Beispiel 6.3 (Erfolgswahrscheinlichkeit beim Münzwurf). In Beispiel 6.4 ist einfach X die Anzahl der Erfolge und $\mathbf{P}_p := B(n = 100, p)$ (wobei wir ϑ durch p ersetzt haben).

6.2 Schätzprobleme

Ein Schätzproblem ist das Folgende: Gegeben sei eine Funktion $m : \Theta \rightarrow \Theta'$ (was meist die Identität und $\Theta = \Theta'$ ist). Bestimme eine Funktion $h : E \rightarrow \Theta'$, so dass $h(X)$ den Wert von $m(\vartheta)$ (möglichst gut) schätzt. Interessant hierbei ist, was es denn überhaupt bedeutet, guter Schätzer zu sein.

Beispiel 6.4 (Erfolgswahrscheinlichkeit beim Münzwurf). In der Situation aus Beispiel 6.1 heißt der freie Parameter p anstatt ϑ . Es ist $E = \{0, \dots, 100\}$ und X die Anzahl der Erfolge in $n = 100$ Versuche. Damit ist $\mathbf{P}_p = B(n = 100, p)$ mit dem freien Parameter $p \in \Theta = [0, 1]$. Es ist mit $\Theta' = \Theta$

$$h : \begin{cases} E & \rightarrow \Theta, \\ x & \mapsto \frac{1}{100}x. \end{cases}$$

und $\hat{p} = h(X)$ ist ein Schätzer für p . (Im Folgenden werden wir einen Schätzer für einen Parameter \bullet meistens mit $\hat{\bullet}$ bezeichnen.) Für unser Datenbeispiel ist also

$$\hat{p} = \frac{1}{n}(X_1 + \dots + X_n) = \frac{59}{100} = 0.59.$$

¹Wir wollen im Folgenden die dauernde Unterscheidung zwischen diskreten Zufallsvariablen und Zufallsvariablen mit Dichten durch die Notation $\mathbf{P}[X \in da]$ vermeiden. Ist der Wertebereich E von X diskret und $a \in E$, ist damit

$$\mathbf{P}[X \in da] := \mathbf{P}[X = a]$$

gemeint. Hat X die Dichte $f(a)da$, ist

$$\mathbf{P}[X \in da] := f(a)da.$$

Da \hat{p} von den Daten abhängt, die wir uns als Realisierung von einer Zufallsvariable gedacht haben, ist \hat{p} also auch eine Zufallsvariable. Warum ist der Schätzer \hat{p} gut? Nehmen wir an, wir wüssten den wahren Parameter p . Dann leistet \hat{p} zumindest im Mittel das gewünschte: (Wir schreiben hier und im Folgenden $\mathbf{P}_p[\cdot]$ und $\mathbf{E}_p[\cdot]$, wenn wir Wahrscheinlichkeiten und Erwartungswerte unter der Hypothese ausrechnen wollen, dass p der wahre Parameter ist.)

$$\mathbf{E}_p[\hat{p}] = \frac{1}{n} \mathbf{E}_p[X_1 + \cdots + X_n] = p.$$

Wir sagen auch, der Schätzer \hat{p} ist erwartungstreu (oder unverzerrt oder unbiased).

Eine weitere wünschenswerte Eigenschaft eines Schätzers ist, dass er immer besser wird, je größer die zu Grunde liegende Datenmenge ist. Eine große Datengrundlage bedeutet in unserem Fall, dass die Münze oft geworfen wurde, also n groß ist. Aus dem schwachen Gesetz großer Zahlen wissen wir, dass

$$\mathbf{P}_p[|\hat{p} - p| \geq \varepsilon] = \mathbf{P}_p\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mathbf{E}_p[X_1]\right| \geq \varepsilon\right] \xrightarrow{n \rightarrow \infty} 0$$

für alle $\varepsilon > 0$. Die Eigenschaft, dass \hat{p} mit hoher Wahrscheinlichkeit immer näher am wahren Wert p liegt, wenn mehr Daten zur Verfügung stehen, nennen wir Konsistenz.

Definition 6.5 (Punktschätzer, unverzerrte und konsistente Schätzer).

1. Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $m : \Theta \rightarrow \Theta'$. Jede Statistik $\hat{m} := \hat{m}(X)$ mit $\hat{m} : E \rightarrow \Theta'$ heißt (Punkt-)Schätzer für $m(\vartheta)$.

Der Schätzer \hat{m} heißt unverzerrt (erwartungstreu, unbiased), falls

$$\mathbf{E}_\vartheta[\hat{m}] = m(\vartheta)$$

für alle $\vartheta \in \Theta$.

2. Sei $(X^n, (\mathbf{P}_\vartheta^n)_{\vartheta \in \Theta})_{n=1,2,\dots}$ eine Folge statistischer Modelle mit derselben Parametermenge Θ und $\hat{m}_1(X^1), \hat{m}_2(X^2), \dots$ eine Folge von Schätzern für $m(\vartheta)$. Die Folge $\hat{m}_1(X^1), \hat{m}_2(X^2), \dots$ heißt konsistent, falls

$$\mathbf{P}_\vartheta^n[|\hat{m}_n(X^n) - m(\vartheta)| \geq \varepsilon] \xrightarrow{n \rightarrow \infty} 0$$

für alle $\varepsilon > 0, \vartheta \in \Theta$.

3. Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell. Die Abbildung

$$L : \begin{cases} E \times \Theta & \rightarrow [0, 1] \\ (a, \vartheta) & \mapsto \mathbf{P}_\vartheta[X \in da] \end{cases}$$

heißt Likelihood-Funktion. Für eine Abbildung $h : E \mapsto \Theta$ mit

$$L(a, h(a)) = \max_{\vartheta \in \Theta} L(a, \vartheta)$$

heißt $\hat{\vartheta}_{ML} = h(X)$ Maximum-Likelihood-Schätzer von ϑ .

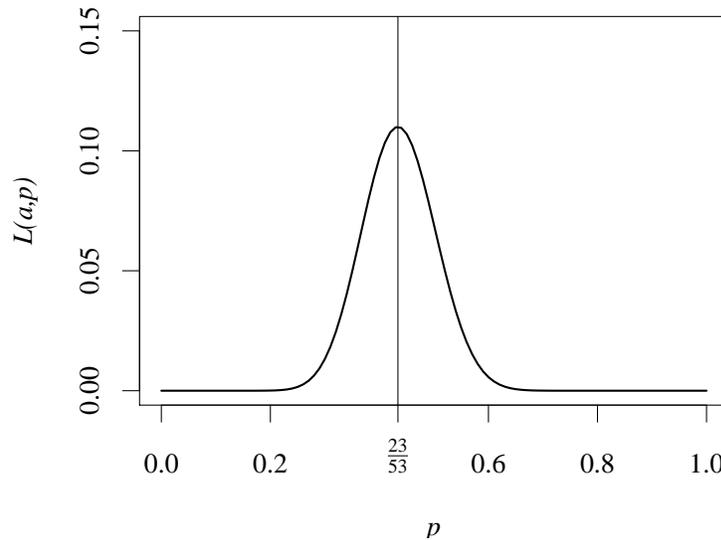


Abbildung 6.2: Die Likelihood-Funktion beim Münzwurf für $X = 23$ aus Beispiel 6.7.

Bemerkung 6.6 (Interpretation von Maximum-Likelihood-Schätzern). Sei X diskret und $X = a$. Da die Vorstellung die ist, dass die erhobenen Daten Ergebnis eines Zufallsexperiments (d.h. die Realisierung einer Zufallsvariable X) sind, sagt man auch, dass die Daten $X = a$ sind. Ein Maximum-Likelihood-Schätzer ist also ein Parameter ϑ , unter dem die Wahrscheinlichkeit, die Daten $X = a$ zu beobachten – das ist $\mathbf{P}_\vartheta[X = a]$ – maximal ist.

Beispiel 6.7 (Erfolgswahrscheinlichkeit beim Münzwurf). Betrachten wir also wieder das Beispiel der Schätzung der Erfolgswahrscheinlichkeit beim Münzwurf. Wir hatten $X = 59$ Erfolge bei $n = 100$ Erfolgen verzeichnet. Unter \mathbf{P}_p ist X nach $B(n = 100, p)$ verteilt, also ist

$$L(X, p) = \binom{n}{X} p^X (1-p)^{n-X}.$$

die Likelihood-Funktion; siehe auch Abbildung 6.2. Um diese für gegebenes X zu maximieren, berechnen wir den Wert p , für den $\log L(X, p)$ maximal ist. Die Bestimmung des Maximums der log-Likelihood-Funktion $\log L(a, p)$ genügt, da \log eine streng monotone Funktion ist. Wir berechnen

$$\frac{\partial \log L(X, p)}{\partial p} = \frac{\partial}{\partial p} \left(\log \binom{n}{X} + X \log p + (n - X) \log(1 - p) \right) = \frac{X}{p} - \frac{n - X}{1 - p}.$$

Am Maximum \hat{p}_{ML} muss $\frac{\partial \log L(X, p)}{\partial p} = 0$ sein, also ist

$$(1 - \hat{p}_{ML})X = \hat{p}_{ML}(n - X), \quad \hat{p}_{ML} = \frac{X}{n}$$

ein Maximum-Likelihood-Schätzer für p . Da es nur ein einziges Maximum der Likelihood gibt, ist dies auch der einzige Maximum-Likelihood-Schätzer.

Soeben haben wir gesehen, dass der Maximum-Likelihood-Schätzer für eine Summe von Zufallsvariablen (was der Anzahl an Erfolgen im Münzwurf entspricht) gerade durch den Mittelwert gegeben ist. Die Verwendung des Mittelwertes für eine solche Schätzung ist generell eine gute Idee, wie wir nun zeigen werden.

Definition 6.8 (Mittelwert und empirische Varianz). Sei $\underline{X} = (X_1, \dots, X_n)$ ein Vektor von Zufallsvariablen. Dann heißt

$$\bar{X} := \frac{1}{n}(X_1 + \dots + X_n)$$

Mittelwert von \underline{X} und

$$s^2(\underline{X}) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

empirische Varianz von \underline{X} .

Theorem 6.9 (Mittelwert und empirische Varianz als Schätzer).

Sei $(\underline{X} = (X_1, \dots, X_n), (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell, so dass X_1, \dots, X_n unter allen \mathbf{P}_ϑ identisch verteilt sind und $\mu_\vartheta := \mathbf{E}_\vartheta[X_1]$ existiert.

1. Es ist \bar{X} ein unverzerrter Schätzer für μ_ϑ . Sind außerdem X_1, \dots, X_n unter allen \mathbf{P}_ϑ unabhängig mit endlichem $\mathbf{V}_\vartheta[X_1]$, so ist \bar{X} auch konsistent.
2. Sei $n \geq 2$ und $\underline{X} = (X_1, \dots, X_n)$ so, dass X_1, \dots, X_n unter allen \mathbf{P}_ϑ paarweise unkorreliert und identisch verteilt sind mit $\sigma_\vartheta^2 := \mathbf{V}_\vartheta[X_1] < \infty$. Dann ist die empirische Varianz $s^2(\underline{X})$ ein unverzerrter Schätzer für σ_ϑ^2 .

Beweis. 1. Zunächst ist

$$\mathbf{E}_\vartheta[\bar{X}] = \frac{1}{n}(\mathbf{E}_\vartheta[X_1] + \dots + \mathbf{E}_\vartheta[X_n]) = \mathbf{E}_\vartheta[X_1] = \mu_\vartheta,$$

was bereits die Unverzerrtheit von \bar{X} als Schätzer von μ_ϑ zeigt. Für die Konsistenz berechnen wir für $\varepsilon > 0$

$$\mathbf{P}_\vartheta[|\bar{X} - \mu_\vartheta| \geq \varepsilon] = \mathbf{P}_\vartheta\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbf{E}_\vartheta[X_1]\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

nach dem schwachen Gesetz der großen Zahlen.

2. Wir schreiben zunächst

$$\mathbf{E}_\vartheta[s^2(\underline{X})] = \frac{1}{n-1} \sum_{i=1}^n \mathbf{E}_\vartheta[X_i^2 - 2X_i\bar{X} + \bar{X}^2] = \frac{n}{n-1} \mathbf{E}_\vartheta[X_1^2 - 2X_1\bar{X} + \bar{X}^2].$$

Nun ist

$$\begin{aligned} \mathbf{E}_\vartheta[X_1^2] &= \mu_\vartheta^2 + \sigma_\vartheta^2, \\ \mathbf{E}_\vartheta[X_1\bar{X}] &= \mu_\vartheta^2 + \frac{1}{n}\sigma_\vartheta^2, \\ \mathbf{E}_\vartheta[\bar{X}^2] &= \mathbf{E}_\vartheta[X_1\bar{X}], \end{aligned}$$

also

$$\mathbf{E}_\vartheta[s^2(\underline{X})] = \frac{n}{n-1} \mathbf{E}_\vartheta[X_1^2 - X_1\bar{X}] = \sigma_\vartheta^2,$$

was die Unverzerrtheit bereits zeigt. □

Beispiel 6.10 (Erfolgswahrscheinlichkeit beim Münzwurf). Betrachten wir noch einmal das einführende Beispiel der Schätzung des Erfolgsparameters in einem Münzwurf in n Versuchen. Wir wählen dazu die Notation aus den Beispielen 6.1 und 6.4. Der Schätzer $\hat{p} = h(X)$ ist unverzerrt und konsistent, wie wir soeben gezeigt haben.

Ein weiterer Schätzer für p wäre $\hat{p}' = X_1$. Das bedeutet, dass \hat{p}' nur die beiden Werte, 0 und 1 annehmen kann, und genau dann 1 ist, wenn der erste Wurf einen Erfolg zeigt. Der Schätzer \hat{p}' ist ebenfalls erwartungstreu, denn

$$\mathbf{E}_p[\hat{p}'] = \mathbf{E}_p[X_1] = p.$$

Allerdings ist \hat{p}' nicht konsistent, da

$$\mathbf{P}_p(|\hat{p}' - p| > \varepsilon) = 1,$$

falls $\varepsilon < \min(p, 1 - p)$, unabhängig von n .

Beispiel 6.11 (Maximum-Likelihood-Schätzer bei Normalverteilungen).

Wir betrachten den Fall einer unabhängigen, normalverteilten Stichprobe. Sei also $(\underline{X}, (\mathbf{P}_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+})$ so, dass $\underline{X} = (X_1, \dots, X_n)$ und X_1, \dots, X_n unter $\mathbf{P}_{(\mu, \sigma^2)}$ unabhängig und identisch nach $N(\mu, \sigma^2)$ verteilt sind.

Wir berechnen nun die Maximum-Likelihood-Schätzer für μ und σ^2 . Genau wie im letzten Beispiel berechnen wir zunächst die log-Likelihood-Funktion

$$\begin{aligned} \log L((X_1, \dots, X_n), (\mu, \sigma^2)) &= \log \left(\frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left(- \sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^2} \right) \right) \\ &= -n \log \sigma - \sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^2} + C, \end{aligned}$$

wobei C weder von μ noch von σ abhängt. Ableiten nach μ und σ ergibt

$$\begin{aligned} \frac{\partial \log L((X_1, \dots, X_n), (\mu, \sigma^2))}{\partial \mu} &= \sum_{i=1}^n \frac{X_i - \mu}{\sigma^2}, \\ \frac{\partial \log L((X_1, \dots, X_n), (\mu, \sigma^2))}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^3}. \end{aligned}$$

Für die Maximum-Likelihood-Schätzer $\hat{\mu}_{ML}$ und $\hat{\sigma}_{ML}^2$ gilt notwendigerweise

$$\begin{aligned} \sum_{i=1}^n (X_i - \hat{\mu}_{ML}) &= 0, \\ \frac{n}{\hat{\sigma}_{ML}^2} - \sum_{i=1}^n \frac{(X_i - \hat{\mu}_{ML})^2}{\hat{\sigma}_{ML}^3} &= 0. \end{aligned}$$

Die Maximum-Likelihood-Schätzer sind also gegeben durch

$$\begin{aligned} \hat{\mu}_{ML} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \\ \hat{\sigma}_{ML}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} s^2(\underline{X}). \end{aligned}$$

Insbesondere sehen wir, dass \bar{X} nicht nur erwartungstreu und konsistent (siehe Theorem 6.9) ist, sondern auch ein Maximum-Likelihood-Schätzer für μ . Allerdings ist der Maximum-Likelihood-Schätzer für σ^2 nicht erwartungstreu, wie man aus Theorem 6.9 abliest. Immerhin ist $\hat{\sigma}_{ML}^2$ für große n annähernd erwartungstreu, da $\hat{\sigma}_{ML}^2 - s^2(\underline{X}) \xrightarrow{n \rightarrow \infty} 0$.

6.3 Testprobleme

Ein Testproblem ist das Folgende: Sei $\Theta_0 \subset \Theta$. Kann die Hypothese $\vartheta \in \Theta_0$ auf Grundlage der Daten verworfen werden?

Beispiel 6.12 (Erfolgswahrscheinlichkeit beim Münzwurf). Nehmen wir an, der Werfer der Münze behauptet, sie sei fair, also $p = \frac{1}{2}$. Es ist also $\Theta_0 = \{\frac{1}{2}\}$. Können wir diese Hypothese aufgrund der Daten verwerfen? Zunächst stellen wir fest, dass wir prinzipiell zwei Arten von Fehlern mit unserer Entscheidung machen können. Wenn wir die Hypothese verwerfen, könnte sie doch wahr sein, und wenn wir die Hypothese nicht verwerfen, könnte sie doch richtig sein.

Da wir nicht grundlos dem Werfer der Münze widersprechen wollen, wollen wir die Wahrscheinlichkeit, dass wir die Hypothese ablehnen (wir dem Werfer der Münze widersprechen), obwohl sie wahr ist (die Hypothese des Werfers richtig ist), kontrollieren. Das bedeutet, dass

$$\mathbf{P}_{p=1/2}(\text{Hypothese verwerfen}) \leq \alpha$$

für ein anfangs gewähltes $\alpha \in (0, 1)$ sein soll. Klar ist, dass damit die Hypothese umso seltener abgelehnt werden kann, je kleiner α ist. Nun kommen wir zu der Regel, mit der wir die Hypothese ablehnen wollen. In unserem Beispiel haben wir für die Hypothese $p = \frac{1}{2}$ zu viele (59 von 100) Erfolge. Wir würden die Hypothese ablehnen wollen, wenn

$$\mathbf{P}_{p=1/2}(\text{Daten mindestens so extrem wie tatsächliche Daten}) \leq \alpha. \quad (6.1)$$

Wir wählen $\alpha = 5\%$. Für $X \sim B(n = 100, p = \frac{1}{2})$ verteilt, erwarten wir 50 Erfolge. Um die Wahrscheinlichkeit einer Abweichung, die größer ist als die der Daten zu berechnen, betrachten wir eine nach $N(0, 1)$ verteilte Zufallsvariable Z und berechnen

$$\begin{aligned} & \mathbf{P}_{p=1/2}(|X_1 + \dots + X_n - 50| \geq 9) \\ &= 1 - \mathbf{P}_{p=1/2}\left(-\frac{9}{\sqrt{np(1-p)}} < \frac{Y_n - np}{\sqrt{np(1-p)}} < \frac{9}{\sqrt{np(1-p)}}\right) \\ &\approx 1 - \mathbf{P}_{p=1/2}(-1.8 \leq Z \leq 0.1.8) \approx 7.19\% \end{aligned}$$

Da dieser Wert größer als $\alpha = 5\%$ ist, kann die Hypothese nicht verworfen werden, siehe (6.1).

Wir beginnen mit der Einführung wichtiger Begriffe wie Teststatistik, Nullhypothese, Alternative, Ablehnungsbereich, Signifikanzniveau und p -Wert.

Definition 6.13 (Statistischer Test). Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell, E der Zielbereich von X , und $\Theta_0, \Theta_A \subseteq \Theta$ disjunkt mit $\Theta_0 \cup \Theta_A = \Theta$.

1. Ein Paar (Y, C) mit $Y = t(X)$ für $t : E \rightarrow E'$ und $C \subseteq E'$ heißt statistischer Test von

$$H_0 : \vartheta \in \Theta_0 \text{ gegen } H_A : \vartheta \in \Theta_A.$$

Hier heißt Y Teststatistik, C kritischer oder Ablehnungsbereich des Tests, H_0 heißt Nullhypothese und H_A heißt Alternativhypothese. Man sagt, der Test (Y, C) hat (Signifikanz-)Niveau $\alpha \in [0, 1]$, falls

$$\sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}(Y \in C) \leq \alpha.$$

Falls $Y \in C$, sagt man, dass H_0 abgelehnt (und damit H_A angenommen) ist. Falls $Y \notin C$, sagt man, dass H_0 nicht abgelehnt ist (und H_A abgelehnt ist).

Sei nun (Y, C) ein Test der Nullhypothese H_0 gegen H_A .

2. Die Hypothese $H : \vartheta \in \Theta_H$ (die entweder H_0 oder H_A sein kann) heißt einfach wenn $\Theta_H = \{\vartheta^*\}$ für ein $\vartheta^* \in \Theta$. Andernfalls heißt H zusammengesetzt.
3. Ist $\Theta = (\underline{\vartheta}, \bar{\vartheta})$ ein Intervall (wobei $\underline{\vartheta} = -\infty$ und $\bar{\vartheta} = \infty$ zugelassen sind und die Intervalle auch abgeschlossen sein können), so heißt der Test (Y, C) einseitig, falls $\Theta_H = (\underline{\vartheta}, \vartheta^*)$ oder $\Theta_H = (\vartheta^*, \bar{\vartheta})$. Falls $\Theta_H = (\vartheta^+, \vartheta^*)$ mit $\underline{\vartheta} < \vartheta^+ \leq \vartheta^* < \bar{\vartheta}$, so heißt der Test (Y, C) zweiseitig.
4. Der Test (Y, C) heißt unverfälscht, falls

$$\mathbf{P}_{\vartheta_0}(Y \in C) \leq \mathbf{P}_{\vartheta_A}(Y \in C)$$

für alle $\vartheta_0 \in \Theta_0, \vartheta_A \in \Theta_A$ gilt.

Bemerkung 6.14 (Interpretation und Fehler eines Tests).

1. Einen statistischen Test hat man sich am besten so vorzustellen (siehe auch das nächste Beispiel): die Daten sind gegeben durch die Zufallsvariable X . Diese Daten fasst man durch die meist reellwertige Funktion t zusammen zur Teststatistik $Y = t(X)$. Die Daten können entweder nach \mathbf{P}_{ϑ} mit $\vartheta \in \Theta_0$ (d.h. die Nullhypothese ist richtig) oder mit $\vartheta \in \Theta_A$ (d.h. die Alternativhypothese ist richtig) verteilt sein. Ziel ist es, die Nullhypothese genau dann (anhand der Daten X) abzulehnen, wenn H_A richtig ist. Der Ablehnungsbereich C ist so gewählt, dass H_0 genau dann abgelehnt wird, wenn $Y \in C$. Dabei können zwei verschiedene Arten von Fehler auftreten; siehe auch Tabelle 6.1.

	H_0 abgelehnt	H_0 nicht abgelehnt
H_0 richtig	Fehler erster Art	richtige Entscheidung
H_0 falsch	richtige Entscheidung	Fehler zweiter Art

Tabelle 6.1: Die möglichen Fehler eines statistischen Tests.

Gehen wir zunächst davon aus, dass $\vartheta \in \Theta_0$. Hat der Test ein Niveau α , so wissen wir, dass $\mathbf{P}_{\vartheta}(Y \in C) \leq \alpha$. Da H_0 genau dann abgelehnt wird, wenn $Y \in C$, wissen wir also, dass die Nullhypothese höchstens mit Wahrscheinlichkeit α abgelehnt wird, wenn sie zutrifft. Damit hat man also die Wahrscheinlichkeit, die Nullhypothese abzulehnen, falls sie zutrifft, durch α beschränkt. Falls $Y \in C$, aber $\vartheta \in \Theta_0$, die Nullhypothese also irrtümlicherweise verworfen wird, sprechen wir über einen *Fehler erster Art* (dessen Wahrscheinlichkeit durch α kontrolliert wird).

Geht man davon aus, dass $\vartheta \in \Theta_A$, liegt eine Fehlentscheidung genau dann vor, wenn $Y \notin C$, die Nullhypothese also nicht abgelehnt wird. In diesem Fall sprechen wir von einem *Fehler zweiter Art*. Das Niveau des Tests liefert keinen Anhaltspunkt dafür, mit welcher Wahrscheinlichkeit ein solcher Fehler auftritt.

2. Auf den ersten Blick besteht eine scheinbare Symmetrie zwischen H_0 und H_A . Schließlich lehnen wir H_0 genau dann ab (und nehmen H_A an), wenn $Y \in C$ und wir lehnen H_0 nicht ab (und lehnen damit H_A ab) wenn $Y \notin C$. Allerdings wird diese Symmetrie durch das Niveau des Tests gebrochen. Weiß man, dass (Y, C) ein Test zum Niveau α ist, bedeutet das, dass die Wahrscheinlichkeit, die Nullhypothese H_0 abzulehnen, obwohl sie wahr ist, höchstens α ist. Mit anderen Worten ist die Wahrscheinlichkeit für einen Fehler erster Art höchstens α . Allerdings hat man keine Kontrolle über den Fehler zweiter Art.

Wegen dieser Asymmetrie ist in der Praxis die Nullhypothese genau so zu wählen, dass eine Ablehnung der Nullhypothese möglichst sicher auf die Richtigkeit der Alternativhypothese zurückzuführen ist. Wir betrachten das Beispiel der Münzwürfe aus Beispiel 6.1. Bevor wir die Daten des Experimentes erhoben haben, haben wir die Vorstellung, dass der Würfel gezinkt sein könnte. Außerdem legen wir ein Signifikanzniveau α fest (was in der Praxis oft $\alpha = 5\%$ ist). Um unsere Vorstellung über die Münze zu überprüfen, testen wir

$$\begin{aligned} H_0 &: \text{die Münze ist nicht gezinkt, } p = \frac{1}{2} \\ &\text{gegen} \\ H_A &: \text{die Münze ist gezinkt, } p \neq \frac{1}{2}. \end{aligned}$$

Kommt es nämlich jetzt zu einer Ablehnung von H_0 , so wissen wir, dass dies mit Wahrscheinlichkeit höchstens α dann passiert, wenn H_0 wahr ist, die Münze also nicht gezinkt ist. Damit können wir uns relativ sicher sein, dass die Ablehnung der Nullhypothese darauf zurückzuführen ist, dass H_A zutrifft. Damit ist unsere Vorstellung, dass die Münze gezinkt ist, bei Ablehnung der Nullhypothese höchstwahrscheinlich bestätigt.

3. Die Forderung von unverfälschten Tests ist klar zu verstehen: Da wir H_0 dann ablehnen, wenn $Y \in C$, soll zumindest die Wahrscheinlichkeit, dass H_0 abgelehnt wird, unter \mathbf{P}_{ϑ_A} , $\vartheta_A \in \Theta_A$ größer sein als für \mathbf{P}_{ϑ_0} , $\vartheta_0 \in \Theta_0$.

Bemerkung 6.15 (p -Werte und alternative Definition eines Tests).

1. Sei $Y = y$, d.h. dass die Teststatistik Y , angewendet auf die echten Daten, ergibt y . Dann heißt der Wert

$$p_y := \sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}(Y \text{ mindestens so extrem wie } y)$$

p -Wert des Tests für $Y = y$. Dabei hängt die Bedeutung davon, was 'extrem' heißt davon ab, was genau die Alternative ist. (Dies ist oftmals in konkreten Beispielen einfach zu verstehen, siehe etwa den Binomialtest, Abbildung 6.3.) Immer gilt jedoch $p_y \leq p_{y'}$, falls y mindestens so extrem wie y' ist. Es ist wichtig zu beachten, dass es dadurch einen engen Zusammenhang zwischen dem Niveau α des Tests und dem p -Wert gibt. Ist nämlich (Y, C) ein Test zum Niveau α und

$$C = \{y : y \text{ mindestens so extrem wie } y_0\}$$

für ein y_0 , so wird H_0 genau dann abgelehnt, wenn

$$\alpha \geq \sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}(Y \text{ mindestens so extrem wie } y_0) = p_{y_0}.$$

Ist $Y = y$ und gilt $p_y \leq p_{y_0}$, so wird H_0 also abgelehnt. Es genügt also, für einen Test zum Niveau α und $Y = y$ den Wert p_y zu bestimmen. Ist $p_y \leq \alpha$, so wird H_0 abgelehnt. Dieses Vorgehen wird bei vielen Statistik-Programmen angewendet, bei denen ausschließlich p -Werte ausgegeben werden. Dabei muss man meist angeben, was genau die Alternative ist (einseitig oder zweiseitig), damit das Programm weiß, in welche Richtungen Abweichungen als extrem zu betrachten sind.

Wir formalisieren nun noch das Eingangsbeispiel 6.1, was uns zum Binomialtest führt.

Proposition 6.16 (Binomialtest). Sei $\alpha \in [0, 1]$, $n \in \mathbb{N}$ und $(X, (\mathbf{P}_p)_{p \in [0,1]})$ ein statistisches Modell, so dass X unter \mathbf{P}_p nach $B(n, p)$ verteilt ist.

(a) Ist $\Theta_0 = p^*$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{0, \dots, k\} \cup \{l, \dots, n\})$ ein unverfälschter Test zum Niveau α , falls

$$\mathbf{P}_{p^*}(X \leq k) \leq \alpha/2, \quad \mathbf{P}_{p^*}(X \geq l) \leq \alpha/2.$$

(b) Ist $\Theta_0 = [0, p^*]$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{k, \dots, n\})$ ein unverfälschter Test zum Niveau α , falls

$$\mathbf{P}_{p^*}(X \geq k) \leq \alpha.$$

(c) Ist $\Theta_0 = [p^*, 1]$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{0, \dots, k\})$ ein unverfälschter Test zum Niveau α , falls

$$\mathbf{P}_{p^*}(X \leq k) \leq \alpha.$$

Beweis. Wir beweisen nur (c), die anderen beiden Aussagen folgen analog. Klar ist, dass der Test unverfälscht ist. Es ist außerdem

$$\sup_{p \in \Theta_0} \mathbf{P}_p(X \in \{0, \dots, k\}) = \mathbf{P}_{p^*}(X \leq k) \leq \alpha$$

nach Voraussetzung. Also folgt bereits die Aussage. \square

Beispiel 6.17 (Binomialtest). Sei $\alpha = 5\%$, $n = 100$ und $(X, (\mathbf{P}_p)_{p \in [0,1]})$ wie in der Proposition. Wir wollen nun

$$H_0 : p = 1/2 \text{ gegen } H_A : p \neq 1/2$$

testen, wenn wir in 100 Versuchen 59 Erfolge erzielt haben. Nach Proposition 6.16 ist der kritische Bereich von der Form $\{0, \dots, k\} \cup \{l, \dots, 100\}$. Es ist

$$\mathbf{P}_{p=1/2}(X \leq 40) + \mathbf{P}_{p=1/2}(X \geq 60) \approx 5.69\%.$$

Da $40 < 59 < 60$, liegt 59 nicht im Ablehnungsbereich von H_0 . Damit kann die Nullhypothese aufgrund der Daten ($X = 59$) nicht abgelehnt werden. Auf dasselbe Ergebnis kommt man mit Hilfe des p -Wertes. Es ist

$$\mathbf{P}_{p=1/2}(X \text{ mindestens so extrem wie } 59) = \mathbf{P}_{p=1/2}(X \leq 41) + \mathbf{P}_{p=1/2}(X \geq 59) \approx 8.86\%.$$

Da dieser Wert größer als $\alpha = 5\%$ ist, kann man die Nullhypothese nicht ablehnen.

Der Binomialtest

überprüft, ob bestimmte Erfolgswahrscheinlichkeiten einer Binomialverteilung angenommen werden.

Statistisches Modell X unter \mathbf{P}_p nach $B(n, p)$ verteilt

Hypothesen (a) $H_0 : p \in \{p^*\}$ gegen $H_A : p \notin \{p^*\}$
 (b) $H_0 : p \in [0, p^*]$ gegen $H_A : p \in (p^*, 1]$
 (c) $H_0 : p \in [p^*, 1]$ gegen $H_A : p \in [0, p^*)$

Teststatistik X unter \mathbf{P}_p verteilt nach $B(n, p)$

Ablehnungsbereich (a) $\{0, \dots, k, l, \dots, n\}$ mit $\mathbf{P}_{p^*}(X \leq k), \mathbf{P}_{p^*}(X \geq l) \leq \alpha/2$,
 (b) $\{l, \dots, n\}$ mit $\mathbf{P}_{p^*}(X \geq l) \leq \alpha$
 (c) $\{0, \dots, k\}$ mit $\mathbf{P}_{p^*}(X \leq k) \leq \alpha$,

p -Wert, falls $X = x$ (a) $\mathbf{P}_{p^*}(X \leq x \wedge (2p^*n - x)) + \mathbf{P}_{p^*}(X \geq x \vee (2p^*n - x))$
 (b) $\mathbf{P}_{p^*}(X \geq x)$
 (c) $\mathbf{P}_{p^*}(X \leq x)$

Abbildung 6.3: Schematische Darstellung des Binomialtests aus Proposition 6.16

Wie behandeln nun noch einen Test der auf normalverteilten Daten beruht. Um die Verteilung der Teststatistik anzugeben, benötigen wir hierfür eine Definition.

Definition 6.18 (Student- t -Verteilung). Sei $n > 1$ und $\underline{X} = (X_1, \dots, X_n)$ unabhängig und nach $N(0, 1)$ verteilt, sowie \bar{X} der Mittelwert und $s^2(\underline{X})$ die empirische Varianz. Dann heißt die Verteilung von

$$T := \frac{\bar{X}}{\sqrt{s^2(\underline{X})/n}}$$

(Student-) t -Verteilung mit $n - 1$ Freiheitsgraden. Diese bezeichnen wir auch mit $t(n)$.

Bemerkung 6.19 (Transformation von $N(\mu, \sigma^2)$ -verteilten Zufallsvariablen). Sei $n > 1$ und $\underline{X} = (X_1, \dots, X_n)$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt. Man überlegt sich leicht, dass dann Y_1, \dots, Y_n mit $Y_i := \frac{X_i - \mu}{\sqrt{\sigma^2}} \sim N(0, 1)$. Deswegen gilt in dieser Situation, dass

$$\frac{\bar{Y}}{\sqrt{s^2(\underline{Y})}} = \frac{\bar{X} - \mu}{\sqrt{s^2(\underline{X})}} \sim t(n - 1).$$

Proposition 6.20 (Einfacher t -Test).

Sei $\alpha \in [0, 1]$, $\mu^* \in \mathbb{R}$ und $(X = (X_1, \dots, X_n), (\mathbf{P}_{\mu, \sigma^2})_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+})$ ein statistisches Modell, so

dass X_1, \dots, X_n unter $\mathbf{P}_{\mu, \sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt sind. Weiter sei

$$T := \frac{\bar{X} - \mu^*}{\sqrt{s^2(X)/n}}$$

und $t_{n,p}$ für $p \in [0, 1]$ das p -Quantil von $t(n)$.

- (a) Ist $\Theta_0 = \{\mu^*\} \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, (-\infty, t_{n-1, \alpha/2}) \cup (t_{n-1, 1-\alpha/2}, \infty))$ ein unverfälschter Test zum Niveau α .
- (b) Ist $\Theta_0 = (-\infty, \mu^*] \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, [t_{n-1, 1-\alpha}, \infty))$ ein unverfälschter Test zum Niveau α .
- (c) Ist $\Theta_0 = [\mu^*, \infty) \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, (-\infty, t_{n-1, \alpha}])$ ein unverfälschter Test zum Niveau α .

Beweis. Wieder beweisen wir nur (c), da die anderen beiden Aussagen analog folgen. Klar ist, dass der Test unverfälscht ist. Nach Definition der t -Verteilung folgt, dass T unter $\mathbf{P}_{\mu^*, \sigma^2}$ nach $t(n-1)$ verteilt ist. Damit gilt

$$\sup_{\mu \geq \mu^*} \mathbf{P}_{\mu, \sigma^2}(T \leq t_{n-1, \alpha}) = \mathbf{P}_{\mu^*, \sigma^2}(T \leq t_{n-1, \alpha}) = \alpha,$$

woraus die Behauptung sofort folgt. \square

Beispiel 6.21 (Normalverteilte Schlafdauern). Wir untersuchen ein Beispiel aus der Medizin. Ein Medikament wird daraufhin untersucht, ob es den Schlaf von Probanden verlängert. Dazu wird jeweils die Schlafdauerdifferenz bei zehn Patienten notiert. Man erhält

$$1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4.$$

Diese Beobachtungen betrachten wir als Realisierungen von Zufallsvariablen X_1, \dots, X_{10} , die nach $N(\mu, \sigma^2)$ verteilt sind, wobei weder μ noch σ^2 bekannt sind. Wir berechnen

$$\bar{X} = 2.33, \quad s^2(X) = 4.01.$$

Damit ist

$$T = 2.33 / \sqrt{4.01/10} \approx 3.68.$$

Testet man also

$$H_0 : \mathbf{E}_{\mu, \sigma^2}[\bar{Y} - \bar{X}] = 0 \text{ gegen } \mathbf{E}_{\mu, \sigma^2}[\bar{Y} - \bar{X}] \neq 0,$$

auf dem Niveau 5%, so ist dessen Ablehnungsbereich $C = (-\infty, -2.262) \cup (2.262, \infty)$. Da $3.68 \in C$, kann man H_0 auf dem Niveau 5% verwerfen. Das bedeutet, dass vermutlich eine Veränderung der Schlafdauer durch Einnahme des Medikamentes stattfand.

Bemerkung 6.22 (Kontingenztafeln und Unabhängigkeitstests). Wir kommen nun zum letzten hier behandelten Test, Fisher's exaktem Test. Gegeben seien Items mit zwei Merkmalen. (Beispielsweise Pflanzenblätter, die sowohl eine Farbe als auch eine bestimmte Form haben.) Solche Daten lassen sich in einer Kontingenztafel zusammenfassen; siehe unten. Falls beide Merkmale genau zwei verschiedene Möglichkeiten haben, lässt sich mit Hilfe des exakten Tests von Fisher bestimmen, ob beide Ausprägungen unabhängig sind.

Der einfache t -Test

überprüft, ob der Erwartungswert einer Normalverteilung (oder irgendeiner Verteilung bei approximativ unendlich großer Samplegröße) gleich einer vorgegebenen Größe μ^* ist, wenn die Varianz unbekannt ist.

Statistisches Modell	$X = (X_1, \dots, X_n)$ und X_1, \dots, X_n unter $\mathbf{P}_{\mu, \sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt
Hypothesen	(a) $H_0 : \mu \in \{\mu^*\}$ gegen $H_A : \mu \notin \{\mu^*\}$ (b) $H_0 : \mu \in (-\infty, \mu^*]$ gegen $H_A : \mu \in (\mu^*, \infty)$ (c) $H_0 : \mu \in [\mu^*, \infty)$ gegen $H_A : \mu \in (-\infty, \mu^*)$
Teststatistik	$T = \frac{\bar{X} - \mu^*}{\sqrt{s^2(X)/n}}$ unter $\mathbf{P}_{\mu^*, \sigma^2}$ verteilt nach $t(n-1)$
Ablehnungsbereich	(a) $(-\infty, t_{n-1, \alpha/2}) \cup (t_{n-1, 1-\alpha/2}, \infty)$, (b) $(t_{n-1, 1-\alpha}, \infty)$ (c) $(-\infty, t_{n-1, \alpha})$
p -Wert, falls $T = t$	(a) $2(1 - \mathbf{P}[T \leq t])$, wenn T nach $t(n-1)$ verteilt ist (b) $\mathbf{P}[T \geq t]$ (c) $\mathbf{P}[T \leq t]$

Abbildung 6.4: Schematische Darstellung des einfachen t -Tests aus Proposition 6.20

	Merkmal 2, Möglichkeit 1	Merkmal 2 Möglichkeit 2	Σ
Merkmal 1, Möglichkeit 1	S_{11}	S_{12}	$S_{1\bullet}$
Merkmal 1, Möglichkeit 2	S_{21}	S_{22}	$S_{2\bullet}$
Σ	$S_{\bullet 1}$	$S_{\bullet 2}$	S

Proposition 6.23 (Fisher's exakter Test). Sei $\alpha \in [0, 1]$, \mathcal{I}, \mathcal{J} endliche Mengen mit $|\mathcal{I}| = |\mathcal{J}| = 2$ und

$$\Theta = \{\underline{p} = (p_{ij})_{i \in \mathcal{I}, j \in \mathcal{J}} \text{ Verteilungsgewichte einer Verteilung auf } \mathcal{I} \times \mathcal{J}\}.$$

Weiter sei $((X, Y) = ((X_1, Y_1), \dots, (X_n, Y_n)), (\mathbf{P}_{\underline{p}})_{\underline{p} \in \Theta})$ ein statistisches Modell, so dass $(X_1, Y_1), \dots, (X_n, Y_n)$ unter $\mathbf{P}_{\underline{p}}$ unabhängig und nach \underline{p} verteilt sind (d.h. $\mathbf{P}(X_k = i, Y_k = j) = p_{ij}$). Setze für $i \in \mathcal{I}, j \in \mathcal{J}$

$$\begin{aligned} S_{ij} &:= |\{k : X_k = i, Y_k = j\}|, \\ S_{i\bullet} &:= |\{k : X_k = i\}| = \sum_{j \in \mathcal{J}} S_{ij}, \\ S_{\bullet j} &:= |\{k : Y_k = j\}| = \sum_{i \in \mathcal{I}} S_{ij} \end{aligned}$$

und $p_{i\bullet} := \sum_{j \in \mathcal{J}} p_{ij}, p_{\bullet j} := \sum_{i \in \mathcal{I}} p_{ij}$ für $\underline{p} \in \Theta$. Sei

$$\Theta_0 = \{\underline{p} \in \Theta : p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für } i \in \mathcal{I}, j \in \mathcal{J}\},$$

$\Theta_A = \Theta \setminus \Theta_0$. Dann ist (S_{11}, C) ein unverfälschter Test zum Niveau α , falls

$$H(S_{1\bullet}, S, S_{\bullet 1})(C) \leq \alpha. \quad (6.2)$$

Beweis. Gegeben sind Daten von S Objekten. Wir gehen davon aus, dass $S_{1\bullet}, S_{2\bullet}, S_{\bullet 1}$ und $S_{\bullet 2}$ bekannt sind. In diesem Fall gilt es, die S Objekte so auf die Kontingenztabelle zu verteilen, dass die Randeinträge stimmen. Wir stellen uns vor, die $S_{\bullet 1}$ Elemente mit Merkmal 2, Möglichkeit 1 seien weiß, die anderen $S_{\bullet 2}$ schwarz. Aus diesen wählen wir $S_{1\bullet}$ aus. Sind die Merkmale wirklich unabhängig, so erhalten wir eine $H(S_{1\bullet}, S, S_{\bullet 1})$ verteilte Zufallsgröße als Eintrag von S_{11} . Ist diese entweder zu groß oder zu klein, können wir die Unabhängigkeit verwerfen. \square

Bemerkung 6.24. In (6.2) besteht eine vermeintliche Unsymmetrie zwischen dem ersten und zweiten Merkmal, da die Parameter der hypergeometrischen Verteilung $S_{1\bullet}$ die Anzahl der gezogenen und $S_{\bullet 1}$ die Anzahl der weißen Kugeln bedeutet. Allerdings gilt

$$\begin{aligned} H(S_{1\bullet}, S, S_{\bullet 1})(k) &= \frac{\binom{S_{\bullet 1}}{k} \binom{S_{\bullet 2}}{S_{1\bullet} - k}}{\binom{S}{S_{1\bullet}}} = \frac{S_{\bullet 1}! S_{\bullet 2}! S_{1\bullet}! S_{2\bullet}!}{S! k! (S_{\bullet 1} - k)! (S_{1\bullet} - k)! (S - S_{\bullet 1} - S_{1\bullet} + k)!} \\ &= H(S_{\bullet 1}, S, S_{1\bullet})(k), \end{aligned}$$

was diese vermeintliche Unsymmetrie erklärt.

Fisher's exakter Test

überprüft, ob zwei Merkmale, die in jeweils zwei möglichen Ausprägungen vorliegen, stochastisch unabhängig sind.

Statistisches Modell $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig, identisch verteilt,

$$S_{ij} := |\{k : X_k = i, Y_k = j\}|,$$

$$S_{i\bullet} := |\{k : X_k = i\}| = \sum_{j \in \mathcal{J}} S_{ij},$$

$$S_{\bullet j} := |\{k : Y_k = j\}| = \sum_{i \in \mathcal{I}} S_{ij}.$$

Hypothesen $H_0 : \mathbf{P}(X_k = i, Y_k = j) = \mathbf{P}(X_k = i) \cdot \mathbf{P}(Y_k = j)$ für alle i, j
 $H_A : \mathbf{P}(X_k = i, Y_k = j) \neq \mathbf{P}(X_k = i)\mathbf{P}(Y_k = j)$ für
 mindestens ein Paar i, j

Teststatistik S_{11} ist – gegeben $S_{1\bullet}, S_{2\bullet}, S_{\bullet 1}$ und $S_{\bullet 2}$ – nach
 $H(S_{1\bullet}, S, S_{\bullet 1})$ verteilt.

Ablehnungsbereich $S_{11} \in C := \{0, \dots, k, \ell, \dots, S_{1\bullet} \wedge S_{\bullet 1}\}$ falls $H(S_{1\bullet}, S, S_{\bullet 1})(C) \leq \alpha$.

Abbildung 6.5: Schematische Darstellung von Fisher's exaktem Test

Beispiel 6.25 (Mendel's Experimente). Wir betrachten nochmal die beiden Merkmale der Blütenfarbe und Blattmuster einer Pflanze. Beispielsweise könnte ein Züchter von Pflanzen zählen, wieviele seiner Pflanzen rot und weiß blühen, und wieviele glatte und gesprenkelte Blätter haben. Wir gehen davon aus, dass (X_i, Y_i) die Farbe und die Blättermusterung der i -ten Pflanze ist. Ziel ist es herauszufinden, ob X_i und Y_i stochastisch unabhängig sind, d.h. ob etwa

$$\mathbf{P}(X_i = \text{rot}, Y_i = \text{glatt}) = \mathbf{P}(X_i = \text{rot}) \cdot \mathbf{P}(Y_i = \text{glatt})$$

(und ähnliche Gleichungen für die anderen Farben/Blattmusterungen) gilt. Bei einem Versuch, in dem man 400 Blumen als Stichprobe untersucht hat, erhält man folgende Daten, die wir in einer Kontingenztabelle zusammen fassen:

	glatt	gesprenkelt	Σ
weiß	41	74	115
rot	96	189	285
Σ	137	263	400

Wir bezeichnen diese Anzahlen mit $S_{\text{weiß, glatt}}, \dots, S_{\text{rot, gesprenkelt}}$. Die Marginalien bezeichnen wir mit $S_{\text{weiß}\bullet}, \dots, S_{\text{rot}\bullet}$ und $S_{\bullet \text{glatt}}, S_{\bullet \text{gesprenkelt}}$. Kann aufgrund dieser Daten die Hypothese der Unabhängigkeit auf einem Signifikanzniveau von 5% verworfen werden? Wären die beiden

Merkmale unabhängig, so wäre $\mathbf{E}[S_{11}] = 400 \cdot \frac{137}{400} \cdot \frac{115}{400} \approx 39.39$. Beobachtet wurden aber $S_{11} = 41$.

Zunächst stellen wir fest, dass

$$H(115, 400, 137)\{0, \dots, 30\} \approx 0.018 < 0.032 = H(115, 400, 137)\{0, \dots, 31\}$$

und

$$H(115, 400, 137)\{49, \dots, 115\} \approx 0.018 < 0.030 = H(115, 400, 137)\{48, \dots, 115\}.$$

Damit ist

$$C = \{0, \dots, 30, 49, \dots, 115\}$$

der Ablehnungsbereich für Fisher's exakten Test. Damit kann die Nullhypothese einer unabhängigen Ausprägung der Blütenfarbe und Blattmusterung nicht abgelehnt werden.

7 Stochastische Optimierung

Stochastische Optimierung meint das Optimieren mit stochastischen Methoden.² Das abstrakte Problem ist dabei wie folgt: Gegeben sei eine endliche Menge E und eine Funktion $f : E \rightarrow \mathbb{R}$. Finde s_* (oder s^*) mit

$$f(s_*) = \min_{s \in E} f(s) \quad (\text{oder } f(s^*) = \max_{s \in E} f(s)).$$

Minimierungs- und Maximierungsprobleme kann man dabei durch den Übergang zur Funktion $-f$ ineinander umwandeln.

Beispiel 7.1 (Traveling Salesman Problem (TSP)). Ein Handlungsreisender will m Städte besuchen und dann nach Hause zurückkehren. Der Abstand der Städte ist dabei durch eine (symmetrische) $m \times m$ -Matrix D gegeben. Bei seiner Rundreise will er seine zurückgelegte Wegstrecke minimieren. Hier ist

$$E = \mathcal{S}_m$$

die Menge aller Permutationen von $\{1, \dots, m\}$ und

$$f(\xi) = \sum_{i=1}^{m-1} D_{\xi_i, \xi_{i+1}} + D_{\xi_m, \xi_1}. \quad (7.1)$$

Legt man einen Wert d fest und fragt, ob es eine Permutation gibt mit $f(\xi) \leq d$, so handelt es sich um das TSP-Entscheidungsproblem. Dieses ist NP-vollständig: Bis heute sind keine deterministischen Algorithmen bekannt, die das TSP-Entscheidungsproblem in einer Zeit, die polynomial ist in m , entscheiden können. Ein Ansatz, der hier weiterführen kann, ist es, stochastische Algorithmen zu untersuchen, die zumindest solche Wege finden, deren Länge kurz ist. Dabei kann man nur hoffen, mit stochastischen Methoden eine möglichst gute Lösung zu erhalten, jedoch nicht hoffen, mit stochastischen Methoden einen Beweis für die Optimalität einer Lösung zu finden.

Zwei Methoden der stochastischen Optimierung, die man auf obige Probleme anwenden kann, ist *Simulated Annealing* und *genetische Algorithmen*.

²Oft wird unter *stochastischer Optimierung* auch das Optimieren unter unsicheren (stochastischen) Daten verstanden. Das wollen wir hier nicht behandeln.

7.1 Simulated Annealing

Der Begriff des *Annealing* kommt aus der Metallverarbeitung und beschreibt das *Auskühlen* von Metallen. Dort sieht man sich dem Problem gegenüber, ein Metall so auszukühlen, dass lokal keine Spannungen entstehen, die das Material instabil werden lassen. Mit anderen Worten will man die Energie des Metalls durch das Ausglühen minimieren.

Die Idee des *Simulated Annealing* ist die folgende: Starte eine Markoff-Kette mit Zustandsraum E und Übergangsmatrix P_1 und einer stationären Verteilung, die großes Gewicht auf Zustände $s \in E$ mit kleinem $f(s)$ legt. Warte eine Zeit N_1 , bis diese annähernd im Gleichgewicht ist. Starte nun in diesem Zustand eine neue Markoff-Kette mit Übergangsmatrix P_2 , die noch größeres Gewicht auf $s \in E$ mit kleinem $f(s)$ legt. Warte wieder eine Zeit N_2 bis diese Markoff-Kette im Gleichgewicht ist. Setzt man dieses Vorgehen fort, so besteht die Hoffnung, dass sich die Markoff-Kette am Ende in einem Zustand mit minimalem $f(s)$ befindet. Formal wird dies durch eine zeitinhomogene Markoff-Kette mit Übergangsmatrix

$$P^{(n)} = \begin{cases} P_1, & 0 < n \leq N_1 \\ P_2, & N_1 < n \leq N_1 + N_2 \\ \dots & \end{cases}$$

beschrieben.

Zunächst wollen wir die *Boltzmann-Verteilung* für eine Funktion $f : E \rightarrow \mathbb{R}$ definieren, die obige Forderung, hohes Gewicht auf Zustände s mit kleinem $f(s)$ zu legen, erfüllt. Im Anschluss definieren wir eine Markov-Kette, die die Boltzmann-Verteilung als stationäre Verteilung hat.

Definition 7.2 (Boltzmann-Verteilung). Sei $f : E \rightarrow \mathbb{R}$ nach unten beschränkt und $T > 0$. Dann ist die Boltzmann-Verteilung $\pi_{f,T}$ gegeben durch

$$\pi_{f,T}(s) = \frac{1}{Z_{f,T}} \exp\left(-\frac{f(s)}{T}\right),$$

$$Z_{f,T} = \sum_{s \in E} \exp\left(-\frac{f(s)}{T}\right).$$

Der Parameter T heißt Temperatur.

Je kleiner T als Parameter in der Boltzmann-Verteilung, desto größer sind die relativen Unterschiede für Zustände $s, t \in E$ mit $f(s) \neq f(t)$. Je größer $f(t)$, desto kleiner ist das Gewicht $\pi_{f,T}(t)$. Noch genauer gilt das folgende Resultat.

Lemma 7.3. Sei $T > 0$, E eine endliche Menge und $f : E \rightarrow \mathbb{R}$ beliebig. Weiter sei

$$E_* := \{s \in E : f(s) = \min_{t \in E} f(t)\}.$$

Dann gilt

$$\lim_{T \rightarrow 0} \pi_{f,T}(E_*) = 1.$$

Beweis. Sei $E_* = \{s_1, \dots, s_n\}$, $a = f(s)$ und $b = \min_{t \in E \setminus E_*} f(t)$. Es gilt

$$\lim_{T \rightarrow 0} \exp\left(\frac{a-b}{T}\right) = 0$$

und damit

$$\begin{aligned}\pi_{f,T}(E_*) &= \frac{|E_*| \exp\left(-\frac{a}{T}\right)}{\sum_{t \in E} \exp\left(-\frac{f(t)}{T}\right)} = \frac{|E_*| \exp\left(-\frac{a}{T}\right)}{|E_*| \exp\left(-\frac{a}{T}\right) + \sum_{t \in E \setminus E_*} \exp\left(-\frac{f(t)}{T}\right)} \\ &\geq \frac{|E_*| \exp\left(-\frac{a}{T}\right)}{|E_*| \exp\left(-\frac{a}{T}\right) + (|E| - |E_*|) \exp\left(-\frac{b}{T}\right)} = \frac{1}{1 + \left(\frac{|E|}{|E_*|} - 1\right) \exp\left(\frac{a-b}{T}\right)} \xrightarrow{T \rightarrow 0} 1.\end{aligned}$$

□

Bemerkung 7.4 (Herleitung der Metropolis-Kette). Für eine Verteilung π auf E benötigen wir nun noch eine Markov-Kette $(X_t)_{t=0,1,\dots}$ mit stationärer Verteilung π . Diese wird durch die Metropolis-Kette geliefert, für die π sogar reversibel ist (siehe Definition 5.14). Diese erhält man wie folgt:

- Wähle eine Übergangsmatrix Q einer irreduziblen Markov-Kette.
- Wähle die Übergangsmatrix P der gesuchten Markov-Kette so, dass

$$\pi_i P_{ij} = \min(\pi_i Q_{ij}, \pi_j, Q_{ji})$$

für $i \neq j$. Dann ist wegen der Symmetrie der rechten Seite π reversible Verteilung für die Markov-Kette mit Übergangsmatrix P .

- Auflösen der letzten Gleichung liefert

$$P_{ij} = \min\left(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right) Q_{ij}.$$

Dies ist die gesuchte Übergangsmatrix für

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}.$$

Da nämlich $\sum_{j \neq i} Q_{ij} \leq 1$, muss auch $\sum_{j \neq i} P_{ij} \leq 1$ gelten.

Definition 7.5 (Metropolis-Kette). Sei Q die Übergangsmatrix einer irreduziblen Markov-Kette mit Zustandsraum E und π eine Verteilung auf E . Dann ist die Metropolis-Kette $\mathcal{X} = (X_t)_{t=0,1,\dots}$ die Markov-Kette mit Übergangsmatrix P mit

$$P_{ij} = \min\left(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right) Q_{ij}$$

und

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}.$$

Beispiel 7.6 (Traveling Salesman Problem). Wir wollen nun für das TSP eine Metropolis-Kette konstruieren, die die Boltzmann-Verteilung $\pi_{f,T}$ zur Funktion f aus Beispiel 7.1 als stationäre Verteilung hat. Zur Definition der Metropolis-Kette benötigen wir zunächst einen

Graphen mit der Punktmenge E . Wir müssen also angeben, welche Permutationen ξ und η benachbart sein sollen. Dazu definieren wir für $i < j$

$$\sigma_{i,j}(\xi) = (\xi_1, \dots, \xi_{i-1}, \xi_j, \xi_{j-1}, \dots, \xi_{i+1}, \xi_i, \xi_{j+1}, \dots, \xi_m).$$

Also ist $\sigma_{i,j}(\xi)$ die Permutation, die entsteht, indem man in ξ den Abschnitt ξ_i, \dots, ξ_j in umgekehrter Reihenfolge durchläuft. Damit definieren wir

$$\xi \sim \eta : \iff \exists i < j : \eta = \sigma_{i,j}(\xi).$$

Damit hat jede Permutation genau $\binom{m}{2}$ Nachbarn. Wir betrachten die Übergangsmatrix Q mittels

$$Q_{\xi,\eta} = \begin{cases} \frac{1}{\binom{m}{2}}, & \xi \sim \eta \\ 0, & \xi \not\sim \eta. \end{cases} \quad (7.2)$$

Die Markov-Kette mit dieser Übergangsmatrix ist irreduzibel, weil jede Permutation als Komposition von Transpositionen dargestellt werden kann. Für die dazugehörigen Boltzmann-Verteilung entsteht die Metropolis-Kette nun durch Definition 7.5 aus

$$P_{\xi,\eta} = \begin{cases} \frac{1}{\binom{m}{2}} \min \left(\exp \left(\frac{f(\xi) - f(\eta)}{T} \right), 1 \right), & \xi \sim \eta \\ 0, & \xi \not\sim \eta, \xi \neq \eta \\ 1 - \sum_{\xi' \sim \xi} \frac{1}{\binom{m}{2}} \min \left(\exp \left(\frac{f(\xi) - f(\xi')}{T} \right), 1 \right), & \xi = \eta. \end{cases}$$

Ist die Markoff-Kette im Zustand $X_t = \xi$, so verfährt sie also folgendermaßen:

- Wähle zuerst nach der Gleichverteilung zwei Positionen $1 \leq i < j \leq m$.
- Ist $f(\sigma_{i,j}(\xi)) \leq f(\xi)$, so ist $X_{t+1} = \sigma_{i,j}(\xi)$. Ist jedoch $f(\sigma_{i,j}(\xi)) > f(\xi)$, so ist

$$X_{t+1} = \begin{cases} \sigma_{i,j}(\xi) & \text{mit Wahrscheinlichkeit } \exp \left(\frac{f(\xi) - f(\eta)}{T} \right), \\ \xi & \text{mit Wahrscheinlichkeit } 1 - \exp \left(\frac{f(\xi) - f(\eta)}{T} \right). \end{cases}$$

Es ist wichtig zu betonen, dass wir zur Berechnung der Übergangswahrscheinlichkeiten der Metropolis-Kette die Normalisierungskonstanten $Z_{f,T}$ nicht berechnen mussten. Das liegt an der allgemeinen Eigenschaft der Metropolis-Ketten, dass zu deren Berechnung nur Quotienten der Werte der stationären Verteilung bekannt sein müssen.

Die Markov-Kette \mathcal{X} ist irreduzibel und aperiodisch, da $P_{\xi,\xi} > 0$ für manche ξ (falls f nicht konstant ist). Durch Theorem 5.25 und da $\pi_{f,T}$ reversibel für P ist, wissen wir, dass $\mathbf{P}[X_t = i] \xrightarrow{t \rightarrow \infty} \pi_{f,T}(i)$ gelten muss. Schön wäre es nun auszunutzen, dass für E_* wir in Lemma 7.3

$$\mathbf{P}[X_t \in E_*] \xrightarrow{t \rightarrow \infty} \pi_{f,T}(E_*) \xrightarrow{T \rightarrow 0} 1.$$

Allerdings sind hier zwei Grenzübergänge zu betrachten. Dies wird durch eine Wahl von T_1, T_2, \dots und N_1, N_2, \dots durchgeführt: Ziel ist es immer, die Temperatur zu erniedrigen, um dem Minimum näher zu kommen. Passiert dies jedoch zu schnell, so läuft man Gefahr, nur

ein lokales Minimum von f zu finden. Ein Theorem von Geman und Geman besagt: Gilt für die Temperatur $T^{(t)}$ im t -ten Schritt

$$T^{(t)} \geq \frac{|E|(\max_{s \in E} f(s) - \min_{s \in E} f(s))}{\log t},$$

so ist

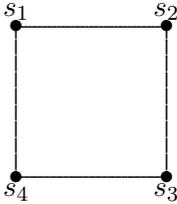
$$\mathbf{P}[X_t \in E_*] \xrightarrow{t \rightarrow \infty} 1.$$

Leider ist das Annealing-Schema, das durch dieses Theorem gegeben wird, sehr langsam. (Beispielsweise ist schon $|E|$ sehr groß.) In der Praxis wird man mehrere Annealing-Schemas ausprobieren, um sagen zu können, welches in akzeptabler Zeit gute Resultate liefert.

Beispiel 7.7 (Zu schnelles Abkühlen). Um ein einfaches Beispiel anzugeben, in dem ein zu schnelles Abkühlen der Markoff-Kette zu einem falschen Ergebnis führt, betrachten wir $E = \{s_1, s_2, s_3, s_4\}$. Weiter haben wir

$$f(s_1) = 1, \quad f(s_2) = 2, \quad f(s_3) = 0, \quad f(s_4) = 2.$$

Um eine Metropolis-Kette zu definieren, müssen wir angeben, welche Punkte benachbart sind. Dies ist durch nachfolgende Grafik veranschaulicht (jeweils mit Übergangswahrscheinlichkeiten $\frac{1}{2}$):



Damit ergibt sich die Übergangsmatrix der Metropolis-Kette

$$P = \begin{pmatrix} 1 - e^{-1/T} & \frac{1}{2}e^{-1/T} & 0 & \frac{1}{2}e^{-1/T} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2}e^{-2/T} & 1 - e^{-2/T} & \frac{1}{2}e^{-2/T} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Wir starten die Markoff-Kette in $X_0 = s_1$ und verwenden ein Annealing-Schema $T^{(1)}, T^{(2)}, \dots$. Sie wird sicher dann nicht in das globale Minimum s_3 konvergieren, wenn sie immer im Zustand s_1 bleibt. Es gilt

$$\mathbf{P}[X_0 = X_1 = \dots = s_1] = \lim_{t \rightarrow \infty} \prod_{i=1}^t (1 - e^{-1/T^{(i)}}) = \prod_{i=1}^{\infty} (1 - e^{-1/T^{(i)}}).$$

Damit gilt nach einem Satz der Analysis

$$\mathbf{P}[X_0 = X_1 = \dots = s_1] > 0 \quad \text{genau dann wenn} \quad \sum_{i=1}^{\infty} e^{-1/T^{(i)}} < \infty.$$

Damit besteht beispielsweise für $T^{(i)} = \frac{1}{i}$ die Möglichkeit, dass die Markoff-Kette immer im Zustand s_1 bleibt und somit nicht in das globale Minimum s_3 konvergiert.

Allgemein kann man sagen, dass ein solches Verhalten auf zwei Dinge zurückzuführen ist: ein zu schnelles Annealing-Schema und ein Punkt s , der zwar ein lokales, aber kein globales Minimum von f ist.

7.2 Genetische Algorithmen

Bei genetischen Algorithmen handelt es sich um Algorithmen, die durch eine ähnliche Dynamik wie sie bei der Fortpflanzung von Individuen beobachtet wird, versuchen, eine Zielfunktion zu maximieren. Diese Zielfunktion heißt meist *Fitnessfunktion*.

Bemerkung 7.8 (Vererbung). In der Biologie geht man davon aus, dass alle Individuen durch ihr genetisches Material, also ihre Chromosomen, bestimmt werden. Chromosomen bestehen aus DNA, also aus der aneinandergereihten Aminosäuren. Mathematisch können Chromosomen als endliche Folge von Buchstaben aus einem Alphabet modelliert werden. Jeder solchen Folge wird der Wert einer Fitness-Funktion zugeordnet. Genetische Algorithmen modellieren nun die Fortpflanzung dieses genetischen Materials. Pflanzen sich Organismen fort, so liegt von beiden Eltern je ein Chromosom vor. Beispielsweise liegen zwei Chromosomen mit der Codierung

$$\begin{array}{cccccccc} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ \text{und} & & & & & & & \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}$$

vor; jedes Chromosom besteht also aus acht Aminosäuren, die wir auch Bits nennen wollen. (Dabei haben wir angenommen, dass es nur genau zwei verschiedene Aminosäuren gibt, was eine Vereinfachung darstellt.) Nun findet *Mutation*, *Rekombination* und *Vererbung mit Selektion* statt.

- *Mutation*: Es besteht die Möglichkeit, dass sich einzelne Bits zufällig ändern. Passiert die beispielsweise an der vierten Stelle des Chromosoms, so ergibt sich

$$1 \ 0 \ 0 \ \mathbf{1} \ 1 \ 1 \ 10$$

- *Rekombination* (oder engl. *Crossover*): Das Chromosom des Nachkommens besteht aus Stücken beider Elternteile. Beispielsweise werden die ersten drei und die letzten beiden Bits des ersten und die restlichen Bits des zweiten Elternteils genommen. Dies ergibt das Chromosom

$$\begin{array}{ccc|ccc|cc} \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & 0 & 1 & \mathbf{1} & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{0} & \mathbf{1} & \mathbf{1} & 1 & 1 \end{array} \longrightarrow 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0$$

Die Punkte des Chromosoms, an denen die Rekombination stattfindet, heißen Crossover-Punkte.

- *Reproduktion mit Selektion*: Es werden zwei Individuen i und j der Population gezogen, das Individuum i bekommt einen Nachkommen, und j stirbt. Die Wahrscheinlichkeit, dass sich i fortpflanzen darf ist gegeben durch proportional zur Fitness von Individuum i .

Die Evolutionstheorie sagt voraus, dass sich Organismen, die besser an ihre Umwelt angepasst sind oder auf andere Art und Weise *fitter* als andere, sich öfter fortpflanzen werden, was durch den letzten Schritt realisiert ist.

Bemerkung 7.9 (In silico Evolution). Ziel ist es, das Maximum einer Funktion $f : E \rightarrow \mathbb{R}$ zu finden. Hierzu sein eine Population der Größe N gegeben. Individuum i hat einen Typ $\alpha(i) \in E$ im Typenraum E . Die Fitness von Typ α ist nun gegeben durch $f(\alpha)$. Da sich gute Typen eher durchsetzen als schlechte, betrachtet man eine Markov-Kette $\mathcal{X} = (X_t)_{t=0,1,\dots}$ mit $X_t = (\alpha_t(1), \dots, \alpha_t(N))$, wobei $\alpha_t(i)$ der Typ von Individuum i zur Zeit t ist. Man geht nun folgendermaßen vor, um das gesuchte Maximum zu finden: Mit bestimmten Wahrscheinlichkeiten findet in jedem Schritt eine der folgenden drei Mechanismen statt.

- *Mutation:* Wähle ein Individuum i und setze $\alpha_{t+1}(i) = \beta$ mit Wahrscheinlichkeit $Q_{\alpha_t(i),\beta}$ für eine stochastische Matrix Q .
- *Rekombination:* Wähle zwei Individuen i, j aus und setze $\alpha_{t+1}(i) = \beta$ mit Wahrscheinlichkeit $Q'_{\alpha_t(i),\alpha_t(j);\beta}$ für eine stochastische Matrix Q' .
- *Reproduktion und Selektion:* Wähle ein Individuum j rein zufällig aus. Weiter wähle Individuum i mit Wahrscheinlichkeit $\frac{f(i)}{\sum_{j \in N} f(j)}$ aus. Setze nun $\alpha_{t+1}(j) = \alpha_t(i)$. (Das bedeutet, dass Individuum i stirbt und durch einen Nachkommen von i ersetzt wird.)

Dieses Schema führt hoffentlich dazu, dass sich die Population aus immer fitteren Individuen zusammensetzt.

Beispiel 7.10 (Traveling Salesman Problem (TSP)). Wir wollen nun einen genetischen Algorithmus angeben, mit dem man eventuell das TSP lösen kann. Wir müssen hierzu definieren, was genau ein Individuum ist und welche Mutations- Rekombinationsmechanismen zugelassen sind und welche Fitness-Funktion wir verwenden.

Der Typenraum ist hier der Raum $E = \mathcal{S}_m$ aller Permutationen der Länge m . Die Fitness von $\xi \in E$ (die es ja zu maximieren gilt) ist z.B. gegeben durch

$$g(\xi) = \frac{1}{f(\xi)}$$

mit f aus (7.1).

- *Mutation:* Eine Mutation besteht aus der zufälligen Vertauschung zweier besuchter Orte, d.h. Q ist gegeben durch (7.2).
- *Rekombination:* Sei $i \in \{1, \dots, m\}$ rein zufällig gewählt. Um ξ, η an der Stelle i zu rekombinieren, gehen wir folgendermaßen vor: Bis Ort i werden die Orte in der Reihenfolge von i besucht. Im Anschluss werden die Orte, die noch nicht besucht worden, in der Reihenfolge besucht, wie sie durch η vorgegeben sind. Sei etwa $m = 5$, $\xi = (3, 5, 2, 1, 4)$, $\eta = (5, 3, 1, 4, 2)$, so rekombinieren die beiden an Ort 5 zu $(3, 5, 1, 4, 2)$.
- *Reproduktion mit Selektion:* Dies folgt wieder dem allgemeinen Schema aus Bemerkung 7.9.