

Stochastik für Informatiker

von

Steffen Dereich

Technische Universität Berlin

Version: 27.04.2009

Inhaltsverzeichnis

1	Einführung	1
1.1	Ereignisse und Wahrscheinlichkeiten	1
1.2	Der diskrete Wahrscheinlichkeitsraum	7
2	Bedingte Wahrscheinlichkeiten & Unabhängigkeit von Ereignissen	10
3	Diskrete Zufallsvariablen	17
3.1	Unabhängigkeit von Zufallsvariablen	19
3.2	Bedingte Verteilungen für Zufallsvariablen	22
3.3	Beispiele diskreter reeller Zufallsvariablen	23
4	Kenngrößen reeller Zufallsvariablen	26
4.1	Erwartungswert	26
4.2	Erwartungswerte wichtiger diskreter Zufallsvariablen	28
4.3	Die Varianz	31
4.4	Varianzen wichtiger reeller Zufallsvariablen	34
4.5	Die Kovarianz und Korrelation	35
5	Konvergenz von Zufallsvariablen	39
5.1	Die Chebychev Ungleichung	39
5.2	Das schwache Gesetz der großen Zahlen	41
5.3	Shannon's Quellkodierungssatz	42
5.4	Das starke Gesetz der Großen Zahlen	44
6	Markov-Ketten	49
6.1	Koppelung diskreter Zufallsvariablen	49
6.2	Homogene Markov-Ketten	51
6.3	Markov-Ketten mit endlichem Zustandsraum	56
6.4	Reversible Markov-Ketten	58
6.5	Austrittszeiten	59
7	Allgemeine Zufallsvariablen	63
7.1	Stetige Zufallsvariablen	65
	Konstruktion stetiger Zufallsvariablen	66
	Wichtige stetige Zufallsvariablen	66
7.2	Quantile und die Erzeugung von Zufallsvariablen mit dem Computer	68

7.3	Der Erwartungswert	70
7.4	Von diskreten zu allgemeinen Zufallsvariablen	73
8	Eine kleine Einführung in die Statistik	76
8.1	Die besondere Rolle der Normalverteilung	76
8.2	Die Grundidee der Statistik	77
	Konfidenzintervalle	78
	Testen von Hypothesen	78
8.3	Der t-Test	79
8.4	Approximatives Konfidenzintervall für Poisson-verteilte Zufallsvariablen	82

1 Einführung

Die Wahrscheinlichkeitstheorie befasst sich mit *Zufallsexperimenten*. Dies sind unter gleichen Bedingungen beliebig oft wiederholbare Experimente mit ungewissem, nicht vorhersagbarem Ausgang.

1.1 Ereignisse und Wahrscheinlichkeiten

Es folgt eine Auflistung der zu einem Zufallsexperiment gehörenden mathematischen Begriffe, jeweils mit einer Standardnotation.

<u>Elementarereignisse</u>	Menge aller möglichen Ausgänge des Experiments (endlich oder unendlich viele)
	Ω Menge der Elementarereignisse (Merkmalraum)
<u>Ereignisse</u>	setzen sich aus Elementarereignissen zusammen, sind also Teilmengen von Ω (aber: nicht jede Teilmenge von Ω muss als Ereignis angesehen werden)
	$A \subset \Omega$ Ereignis (weitere typische Bezeichnungen: B, C, D, \dots).

Beispiel 1.1. Würfelwurf

Um einen Würfelwurf w'theoretisch zu modellieren, wählen wir als Menge der *Elementarereignisse*

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

Alle Teilmengen von Ω sind nun Ereignisse. Z.Bsp. repräsentiert die Menge $A = \{2, 4, 6\}$ das Ereignis eine *gerade* Zahl zu würfeln oder die Menge $B = \{6\}$ eine sechs zu würfeln.

Wir bezeichnen meist mit $\omega \in \Omega$ einen möglichen Ausgang des Experiments. In Abhängigkeit von ω sagen wir

- das Ereignis A tritt ein, wenn $\omega \in A$, und
- das Ereignis A tritt nicht ein, wenn $\omega \notin A$.

Die Mengen \emptyset und Ω besitzen in der Wahrscheinlichkeitstheorie besondere Bezeichnungen:

\emptyset unmögliches Ereignis
 Ω sicheres Ereignis

Durch mengentheoretisches Verknüpfen von Ereignissen erhält man neue Ereignisse:

$A^c := \Omega \setminus A$	Komplementärereignis zu A (tritt genau dann ein, wenn Ereignis A nicht eintritt)
$A \cup B$	mindestens eins der Ereignisse A und B tritt ein (A oder B)
$A \cap B$	beide Ereignisse A und B treten ein (A und B)
$A \Delta B := (A \setminus B) \cup (B \setminus A)$	genau eines der Ereignisse A oder B tritt ein

Weiterhin kann man mengentheoretische Relationen gewissen w'theoretischen Aussagen zuordnen:

$A \cap B = \emptyset$ die Ereignisse A und B treten nicht gleichzeitig ein,
sie heißen *unvereinbar* oder *disjunkt*
 $A \subset B$ das Ereignis A impliziert B .

Wir nehmen nun an, dass wir das Zufallsexperiment beliebig oft wiederholen können, wobei der Ausgang jedes Experiments nicht von den anderen Experimenten beeinflusst wird. Wir bezeichnen mit $\omega_1, \omega_2, \dots \in \Omega$ die hierbei erhaltenen zufälligen Ausgänge des Experiments. Für ein gegebenes Ereignis $A \subset \Omega$, betrachten wir die relative Häufigkeit des Eintritts von A bei n -maliger Durchführung des Experiments:

$$h_n(A) := h_n(A; \omega_1, \dots, \omega_n) := \frac{1}{n} \sum_{i=1}^n 1_A(\omega_i) = \frac{\text{Zahl der günstigen Versuche}}{\text{Gesamtzahl der Versuche}}.$$

Hier bezeichnet $1_A(\cdot)$ die *Indikatorfunktion*

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \quad (\omega \in \Omega).$$

Die Funktion $h_n(A)$ hängt von den zufälligen Realisierungen $\omega_1, \dots, \omega_n$ ab. Die Erfahrung zeigt jedoch, dass der Wert $h_n(A)$ für große n sich um einen typischen Wert konzentriert. Diesen Wert bezeichnen wir mit $P(A)$ und nennen ihn *Wahrscheinlichkeit von A*. Insbesondere können wir aus diesem Zusammenhang einige Axiome, die eine sinnvolle *Wahrscheinlichkeitsverteilung* P erfüllen soll, ableiten:

$$\begin{aligned}
0 \leq h_n(A) \leq 1 & \implies 0 \leq P(A) \leq 1 \\
h_n(\emptyset) = 0, h_n(\Omega) = 1 & \implies P(\emptyset) = 0, P(\Omega) = 1 \\
h_n(A \cup B) = h_n(A) + h_n(B) & \implies P(A \cup B) = P(A) + P(B), \text{ falls die} \\
(A, B \text{ Ereignisse, } A \cap B = \emptyset) & \quad \text{Ereignisse } A \text{ und } B \text{ unvereinbar sind}
\end{aligned}$$

Mathematisch definieren wir ein Wahrscheinlichkeitsmaß P auf Ω wie folgt:

Definition 1.2. Eine Abbildung $P : \{\text{Menge der Ereignisse}\} \rightarrow [0, \infty)$ heißt *Wahrscheinlichkeitsverteilung* (kurz: *Verteilung*), falls

- (i) $P(\Omega) = 1$
- (ii) \forall Folgen von paarweise disjunkten Ereignissen $(A_n)_{n \in \mathbb{N}}$ (d.h. $A_i \cap A_j = \emptyset \quad \forall i \neq j$) gilt

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$$

(σ -Additivität)

Bemerkung 1.3. Um in der Praxis auftretende stochastische Vorgänge mathematisch analysieren zu können, assoziiert man in der *Modellbildung* (oder *Modellierung*) das Ausgangsproblem mit einem Merkmalraum und einer Wahrscheinlichkeitsverteilung. Die Modellbildung an sich ist kein mathematischer Vorgang und es kann verschiedene plausible Ansätze geben. Insbesondere hängen die Schlussfolgerungen (die sich aus der mathematischen Analyse des Modells ergeben) von der Wahl der Modellierung ab! Das Konzept der Wahrscheinlichkeit ist also nicht absolut in dem Sinne, dass in der Realität unvorhersehbaren Ereignissen vor ihrem Eintreten objektive W'keiten zugeordnet werden können.

Bemerkung 1.4. Ein Ereignis A ist eine Teilmenge von Ω . Manchmal ist es nicht möglich alle Teilmengen von Ω als Ereignis zuzulassen, da sonst kein geeignetes Wahrscheinlichkeitsmaß (zur Modellierung eines Experiments) definiert werden kann. Nichtsdestotrotz setzen wir voraus, dass die Menge der Ereignisse mächtig genug ist, um zu garantieren, dass alle obigen (endlichen und abzählbaren) mengentheoretischen Verknüpfungen von Ereignissen immer wieder ein Ereignis liefern und somit der entsprechenden Menge immer auch eine Wahrscheinlichkeit zugeordnet werden kann. Mathematisch fordert man, dass die Gesamtheit aller Ereignisse eine sogenannte σ -Algebra bildet (zur Vollständigkeit befindet sich die Definition einer σ -Algebra am Ende

dieses Kapitels). Für die von uns betrachteten Wahrscheinlichkeitsräume lässt sich in der Regel nur schwer eine nicht messbare Menge definieren. In den Betrachtungen in diesem Skript werden wir deshalb dieses Problem nicht weiter verfolgen.

Notation 1.5. Eine Besonderheit stellen Ereignisse A dar, die mit Wahrscheinlichkeit 1 eintreten. In diesem Fall nennen wir A ein *fast sicheres Ereignis* und sagen “Ereignis A tritt *fast sicher* ein”.

Wir fassen in dem folgenden Satz einige nützliche Rechenregeln für W 'keitsverteilungen zusammen:

Satz 1.6. Für Ereignisse A, B und eine Folge von Ereignissen (A_n) gilt:

$$(i) \quad P(\emptyset) = 0$$

$$(ii) \quad P(A^c) = 1 - P(A)$$

$$(iii) \quad A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$$

$$(iv) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$(v) \quad A \subset B \implies P(A) \leq P(B)$$

$$(vi) \quad A \subset \bigcup_n A_n \implies P(A) \leq \sum_{n \in \mathbb{N}} P(A_n)$$

Beweis. (i) Wähle $A_n := \emptyset$ für alle $n \in \mathbb{N}$. Da alle Mengen paarweise disjunkt sind, gilt

$$P(\emptyset) = P\left(\bigcup_n A_n\right) = \sum_{n \in \mathbb{N}} P(\emptyset).$$

Also ist $P(\emptyset) = 0$.

(iii) Wähle $A_1 := A$, $A_2 := B$ und $A_n := \emptyset$ ($n \geq 3$). Dann sind wieder alle Mengen der Folge (A_n) paarweise disjunkt und es gilt

$$P(A \cup B) = P\left(\bigcup_n A_n\right) = \sum_{n \in \mathbb{N}} P(A_n) = P(A) + P(B).$$

(v) Wähle $C = B \setminus A$. Falls $A \subset B$ gilt, ist B die disjunkte Vereinigung von A und C . Somit liefert Aussage (iii):

$$P(B) = P(A \cup C) = P(A) + P(C) \geq P(A).$$

(vi) Wähle $B_n = A_n \setminus (\bigcup_{m < n} A_m)$ ($n \in \mathbb{N}$). Dann sind die Mengen der Folge (B_n) paarweise disjunkt und es gilt $\bigcup_n A_n = \bigcup_n B_n$. Also folgt mit Hilfe von (v), dass

$$P(A) \leq P\left(\bigcup_n A_n\right) = P\left(\bigcup_n B_n\right) = \sum_{n \in \mathbb{N}} P(B_n) \leq \sum_{n \in \mathbb{N}} P(A_n).$$

□

Beispiel 1.7. *Idealer Würfelwurf*

Wir überlegen uns nun wie ein Wahrscheinlichkeitsmaß zur Modellierung eines Würfelwurfs aussehen sollte.

Für ein W'Maß P über $\Omega = \{1, \dots, 6\}$ gilt nun (Eigenschaften (i) und (iii) der Definition bzw. des Satzes)

$$1 = P(\Omega) = \sum_{i=1}^6 P(\{i\}).$$

Setzen wir nun voraus, dass bei einem Würfelwurf jeder Ausgang mit der gleichen W'keit eintritt, so muss

$$P(\{i\}) = \frac{1}{6}$$

für jedes $i \in \Omega$ gelten. Dies entspricht der Wahrscheinlichkeitsverteilung

$$P(A) = \frac{|A|}{6} \quad (A \subset \Omega),$$

wobei $|A|$ die Anzahl der Elemente von A bezeichnet.

Häufig sind die folgenden mengentheoretischen Rechenregeln von de Morgan in unseren Berechnungen nützlich:

Lemma 1.8. *Für zwei Mengen A und B gilt*

$$(A \cap B)^c = A^c \cup B^c \quad \text{und} \quad (A \cup B)^c = A^c \cap B^c.$$

Die Aussage gilt analog für Mengensysteme A_i ($i \in I$), wobei I eine beliebige Indexmenge bezeichnet:

$$\left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} A_i^c \quad \text{und} \quad \left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} A_i^c.$$

Manchmal werden wir die folgende Stetigkeitsaussage für Wahrscheinlichkeitsverteilungen benötigen:

Satz 1.9. *Es sei $(A_n)_{n \in \mathbb{N}}$ eine wachsende Folge von Ereignissen, d.h. $A_i \subset A_j$ für $i \leq j$. Dann gilt*

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Analog gilt für eine monoton fallende Folge $(B_n)_{n \in \mathbb{N}}$ von Ereignissen ($B_i \supset B_j$ für $i \leq j$):

$$P\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} P(B_n).$$

Beweis. Wir setzen $A_0 = \emptyset$ und wählen $C_n = A_n \setminus A_{n-1}$ ($n \in \mathbb{N}$). Nun sind alle Mengen C_n ($n \in \mathbb{N}$) paarweise disjunkt und es gilt $A_n = \bigcup_{m=1}^n C_m$ ($n \in \mathbb{N}$). Also folgt

$$P(A_n) = P\left(\bigcup_{m=1}^n C_m\right) = \sum_{m=1}^n P(C_m),$$

sodass

$$\begin{aligned} P\left(\bigcup_{m \in \mathbb{N}} A_m\right) &= P\left(\bigcup_{m \in \mathbb{N}} C_m\right) = \sum_{m=1}^{\infty} P(C_m) \\ &= \lim_{n \rightarrow \infty} \sum_{m=1}^n P(C_m) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Die zweite Aussage wird in einer Übungsaufgabe bewiesen. □

Bisher haben wir noch nicht festgelegt welche Teilmengen von Ω als Ereignisse angesehen werden. Eine sinnvolle Annahme ist, dass die Gesamtheit der Ereignisse, die wir mit \mathcal{F} bezeichnen wollen (eine Teilmenge der Potenzmenge von Ω), abgeschlossen ist bezüglich der oben genannten mengentheoretischen Verknüpfungen. Man kann zeigen, dass \mathcal{F} gerade dann abgeschlossen ist, falls es eine σ -Algebra ist:

Definition 1.10. Ein Mengensystem $\mathcal{F} \subset \mathfrak{P}(\Omega)$ (=Potenzmenge auf Ω) heißt σ -Algebra, falls

- (i) $\emptyset \in \mathcal{F}$,

$$(ii) \quad A \in \mathcal{F} \implies A^c \in \mathcal{F},$$

$$(iii) \quad (A_n)_{n \in \mathbb{N}} \text{ Folge in } \mathcal{F} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}.$$

In diesem Fall nennt man das Tupel (Ω, \mathcal{F}) *messbarer Raum*. Ist weiterhin P eine Wahrscheinlichkeitsverteilung auf dem messbaren Raum (Ω, \mathcal{F}) (d.h. in diesem Fall ist \mathcal{F} als Gesamtheit aller Ereignisse zu wählen), dann heißt (Ω, \mathcal{F}, P) *Wahrscheinlichkeitsraum*. Dieser Raum bildet die Basis für alle wahrscheinlichkeitstheoretischen Berechnungen.

1.2 Der diskrete Wahrscheinlichkeitsraum

Eine wichtige Klasse von Wahrscheinlichkeitsverteilungen bilden die *diskreten* Verteilungen, die wir in diesem Abschnitt einführen werden. Es bezeichne Ω eine beliebige Grundmenge. Wir lassen alle Teilmengen von Ω als Ereignisse zu, d.h. $\mathcal{F} = \mathfrak{P}(\Omega)$.

Definition 1.11. Eine Funktion $f : \Omega \rightarrow \mathbb{R}$ heißt *Folge von Wahrscheinlichkeitsgewichten* auf Ω , falls

$$(i) \quad \forall \omega \in \Omega : f(\omega) \geq 0,$$

$$(ii) \quad \sum_{\omega \in \Omega} f(\omega) = 1.$$

Satz 1.12. Falls $f : \Omega \rightarrow \mathbb{R}$ eine Folge von Wahrscheinlichkeitsgewichten ist, dann definiert die Abbildung

$$\begin{aligned} P : \mathfrak{P}(\Omega) &\rightarrow [0, 1] \\ A &\mapsto P(A) = \sum_{\omega \in A} f(\omega) \end{aligned}$$

eine Wahrscheinlichkeitsverteilung auf Ω . Wir nennen P die zu (Ω, f) gehörende *diskrete Wahrscheinlichkeitsverteilung*.

Bemerkung 1.13. Anschaulich gesprochen besagt der Satz, dass Zufallsexperimente mit endlichem oder abzählbarem Merkmalraum durch die Wahl einer *Folge von Wahrscheinlichkeitsgewichten* modelliert werden können!

Beweis.

$$(i) \quad \text{Da } f(\omega) \geq 0 \text{ } (\omega \in \Omega) \text{ gilt, folgt dass } P(A) = \sum_{\omega \in A} f(\omega) \geq 0 \text{ } (A \subset \Omega).$$

- (ii) Es gilt $P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = 1$ laut Voraussetzung.
 (iii) Für paarweise disjunkte Teilmengen $(A_n)_{n \in \mathbb{N}}$ von Ω gilt

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{\omega \in \bigcup_{n=1}^{\infty} A_n} f(\omega) \\ &= \sum_{n=1}^{\infty} \sum_{\omega \in A_n} f(\omega) = \sum_{n=1}^{\infty} P(A_n), \end{aligned}$$

da jedes Element $\omega \in A_i$ ($i \in \mathbb{N}$) in keiner anderen Menge A_j ($j \neq i$) liegt. \square

Eine wichtige diskrete Verteilung, ist die Verteilung, die man erhält, wenn man jedem Elementarereignis die gleiche Wahrscheinlichkeit zuordnet. Diese Verteilung haben wir bereits in Beispiel 1.7 kennengelernt.

Definition 1.14. Für eine gegebene endliche Menge Ω , nennen wir die Wahrscheinlichkeitsverteilung P gegeben durch

$$P(A) = \frac{|A|}{|\Omega|} \quad (A \subset \Omega)$$

Gleichverteilung (oder auch *Laplace-Verteilung*) auf Ω . In diesem Fall sind die zugehörigen W'keitsgewichte gegeben durch $f(\omega) = 1/|\Omega|$ ($\omega \in \Omega$).

Beispiel 1.15. *Lotto (Ziehung von 6 Zahlen aus einer Urne mit 49 Kugeln ohne Zurücklegen)*

Modellierung: $\Omega_1 = \{1, \dots, 49\}$

Merkmalraum:

$$\Omega = \{(\omega_1, \dots, \omega_6) \in \Omega_1^6 : \forall i, j \in \{1, \dots, 6\} \text{ mit } i \neq j \text{ gilt } \omega_i \neq \omega_j\}$$

Wahrscheinlichkeitsverteilung: P Gleichverteilung auf Ω .

Frage 1: Wie groß ist die W'keit, dass 6 vorher genannte Kugeln gezogen werden (Ereignis A)?

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6 \cdot 5 \cdot \dots \cdot 2 \cdot 1}{49 \cdot 48 \cdot \dots \cdot 44}$$

Frage 2: Wie groß ist die W'keit für 5 Richtige (Ereignis B)? Es gilt

$$|B| = \underbrace{6}_{\substack{\text{Möglichkeiten} \\ \text{5 richtige Kugeln} \\ \text{zu wählen}}} \cdot \underbrace{43}_{\substack{\text{Möglichkeiten} \\ \text{falsche Kugel} \\ \text{zu wählen}}} \cdot \underbrace{6!}_{\substack{\text{Möglichkeiten} \\ \text{die 6 Kugeln} \\ \text{anzuordnen}}}$$

und somit folgt, dass

$$P(B) = \frac{6 \cdot 43 \cdot 6!}{49 \cdot \dots \cdot 44}.$$

Satz 1.16. *Sei Ω ein endlicher Merkmalraum mit m Elementen. Dann ist die Anzahl der n elementigen Tupel aus Ω (Anzahl der Stichproben)*

- *ohne Berücksichtigung der Reihenfolge und mit paarweise verschiedenen Einträgen (Ziehen ohne Zurücklegen)*

$$\binom{m}{n} = \frac{m!}{n!(m-n)!}$$

- *ohne Berücksichtigung der Reihenfolge und mit beliebigen Einträgen (Ziehen mit Zurücklegen)*

$$\binom{m+n-1}{n}$$

- *mit Berücksichtigung der Reihenfolge und mit paarweise verschiedenen Einträgen (Ziehen ohne Zurücklegen)*

$$\frac{m!}{(m-n)!}$$

- *mit Berücksichtigung der Reihenfolge und mit beliebigen Einträgen (Ziehen mit Zurücklegen)*

$$m^n$$

Lernziele des 1. Kapitels

- Modellierung von Zufallsexperimenten durch Wahrscheinlichkeitsverteilungen (Wahrscheinlichkeitsräume)
- Rechenregeln für Wahrscheinlichkeitsverteilungen
- Diskrete Wahrscheinlichkeitsverteilungen, Wahrscheinlichkeitsgewichte
- Zählargumente (Mengen)

2 Bedingte Wahrscheinlichkeiten & Unabhängigkeit von Ereignissen

Eine typische Frage in der Stochastik lautet:

Gegeben Ereignis B ist eingetreten, wie groß ist dann die Wahrscheinlichkeit, dass auch A eintritt?

Wir betrachten wieder die Funktion der relativen Häufigkeiten

$$h_n(A) = h_n(A; \omega_1, \dots, \omega_n) = \frac{1}{n} \sum_{i=1}^n 1_A(\omega_i),$$

wobei $\omega_1, \dots, \omega_n$ Realisationen von n sich nicht beeinflussenden Experimenten sind. Für zwei Ereignisse A und B mit $h_n(B) > 0$, gibt gerade

$$\frac{h_n(A \cap B)}{h_n(B)} = \frac{\text{Anzahl der Ausgänge in } A \cap B}{\text{Anzahl der Ausgänge in } B}$$

die relative Häufigkeit von Ereignis A , wenn bereits B eingetreten ist, an. Dies motiviert die folgende Definition.

Definition 2.1. Für zwei Ereignisse A und B mit $P(B) > 0$ bezeichnen wir mit

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

die *bedingte Wahrscheinlichkeit von A gegeben B* . Aus technischen Gründen setzen wir außerdem $P(A|B) = 0$, falls $P(B) = 0$. (Diese Wahl ist willkürlich und 0 könnte durch jede beliebige andere Konstante ersetzt werden.)

Beispiel 2.2. *Würfelwurf mit zwei fairen Würfeln*

Modellierung: $\Omega = \{1, \dots, 6\}^2$ und P ist die Gleichverteilung auf Ω

Frage: Wie groß ist die W'keit, dass die gesamte Augensumme der Würfel gleich 7 ist, falls der erste Würfel 3 Augen zeigt?

Wir wählen $A = \{(i, j) \in \Omega : i + j = 7\}$ und $B = \{(3, i) : i \in \{1, \dots, 6\}\}$. Es gilt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(3, 4)\})}{P(\{(3, i) : i \in \{1, \dots, 6\}\})}.$$

Da

$$P(\{(3, 4)\}) = \frac{1}{36} \text{ und } P(\{(3, i) : i \in \{1, \dots, 6\}\}) = \frac{6}{36}$$

gilt, folgt

$$P(A|B) = \frac{1}{6}.$$

Also ist die W'keit, dass die Augensumme beider Würfel gleich 7 ist, gegeben dass der erste Würfel 3 Augen zeigt, gleich $1/6$.

Wir leiten nun einige elementare Formeln für bedingte Wahrscheinlichkeiten her.

Satz 2.3. *Es seien A, B Ereignisse. Falls $P(B) > 0$ ist, gilt:*

$$(i) \quad A \supset B \implies P(A|B) = 1$$

$$(ii) \quad A \cap B = \emptyset \implies P(A|B) = 0$$

$$(iii) \quad (A_n)_{n \in \mathbb{N}} \text{ paarweise disjunkte Ereignisse} \implies$$

$$P\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) = \sum_{n \in \mathbb{N}} P(A_n | B).$$

Beweis. (i)+(ii) folgen direkt aus der Definition.

(iii) Für paarweise disjunkte Ereignisse A_n sind auch die Ereignisse $A_n \cap B$ ($n \in \mathbb{N}$) paarweise disjunkt. Somit folgt

$$P\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) = \frac{P\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{P(B)} = \sum_{n \in \mathbb{N}} P(A_n | B).$$

□

Wir bezeichnen weiterhin mit B ein Ereignis mit $P(B) > 0$. Aus Aussage (i) des Satzes folgt insbesondere, dass $P(\Omega|B) = 1$. Also erfüllt die Abbildung Q gegeben durch $Q(A) := P(A|B)$ (A Ereignis), die Bedingungen

- $Q(\Omega) = 1$ und $Q(A) \geq 0$ (A Ereignis),
- $(A_n)_{n \in \mathbb{N}}$ paarweise disjunkte Folge von Ereignissen \implies

$$Q\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} Q(A_n).$$

Damit ist Q eine *Wahrscheinlichkeitsverteilung*, und es gelten alle Rechenregeln, die wir in den Sätzen 1.6 und 1.9 hergeleitet haben, auch für die bedingte Wahrscheinlichkeit $P(\cdot|B)$. Wir können also mit bedingten Wahrscheinlichkeiten wie gewohnt rechnen. Zum Beispiel gilt für zwei Ereignisse A_1 und A_2 :

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B).$$

Lemma 2.4. *Für zwei Ereignisse A und B gilt*

$$P(A \cap B) = P(A|B) \cdot P(B).$$

Weiterhin gilt für Ereignisse A , B und C :

$$P(A \cap B|C) = P(A|B \cap C) \cdot P(B|C).$$

Beweis. Die erste Aussage folgt direkt aus der Definition der bedingten W'keit. Die zweite Gleichung gilt, da

$$P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)}$$

und

$$P(A|B \cap C) P(B|C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} \cdot \frac{P(B \cap C)}{P(C)} = \frac{P(A \cap B \cap C)}{P(C)}.$$

□

Definition 2.5. Für eine Indexmenge I heißt ein Mengensystem $(B_i)_{i \in I}$ *Partition* von B , falls

- die Mengen B_i ($i \in I$) paarweise disjunkt sind und
- $B = \bigcup_{i \in I} B_i$ gilt.

Lemma 2.6. (Formel von der totalen Wahrscheinlichkeit)
Für zwei Ereignisse A und B gilt

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c).$$

Allgemeiner gilt für eine Partition B_1, \dots, B_n von Ω :

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i).$$

Beweis. Da die Ereignisse $A \cap B$ und $A \cap B^c$ disjunkt sind, gilt

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c) \\ &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c). \end{aligned}$$

Die zweite Aussage folgt analog aus der disjunkten Zerlegung $A = \bigcup_{i=1}^n (A \cap B_i)$. \square

Lemma 2.7. (Bayes-Umkehr-Formel)

Angenommen B_1, \dots, B_n ist eine Partition von Ω . Dann gilt für jedes $k = 1, \dots, n$ und jedes Ereignis A mit $P(A) > 0$:

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}.$$

Beweis. Das vorhergehende Lemma liefert

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

und es folgt

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)}.$$

\square

Bemerkung 2.8. Die Bayes-Umkehr-Formel ist für folgendes Problem nützlich: Es bezeichne A ein Ereignis und B_1, \dots, B_n eine Partition von Ω , d.h. eine Zerlegung des Merkmalraums in sich gegenseitig ausschließende Ereignisse. Angenommen wir kennen die Wahrscheinlichkeiten $P(B_i)$ und wissen wie sich das Eintreten des Ereignisses B_i auf das Eintreten von A , auswirkt, d.h. wir kennen $P(A|B_i)$. Dann liefert uns die Bayes-Umkehr-Formel, wie das Eintreten von A die Wahrscheinlichkeit, dass auch B_k eintritt, beeinflusst ($P(B_k|A)$). Wir können also mit Hilfe der Gleichung die umgekehrten noch nicht bekannten bedingten W'keiten berechnen.

Beispiel 2.9. Ein medizinischer Test soll herausfinden, ob eine Person von einem Virus infiziert ist oder nicht. Die Wahrscheinlichkeit, dass eine Person mit dem Virus infiziert ist, ist 0,5%. Weiterhin liefert der Test unabhängig

davon, ob die Person infiziert ist oder nicht, mit 95% W'keit das richtige Ergebnis und mit 5% W'keit das falsche Ergebnis.

Frage: Wie groß ist die W'keit, dass eine Person bei der der Test "positiv" ausfällt auch tatsächlich infiziert ist?

Wir benutzen die folgenden Ereignisse zur Modellierung:

- B_0 = der Patient ist gesund
- B_1 = der Patient ist mit dem Virus infiziert
- A = das Ergebnis des Tests ist "positiv"

Die obigen Aussagen lassen sich wie folgt übersetzen:

$$P(B_0) = 0,995, \quad P(B_1) = 0,005, \quad P(A|B_0) = 0,05, \quad P(A|B_1) = 0,95.$$

Nun liefert die Bayes-Umkehr-Formel:

$$P(B_1|A) = \frac{P(B_1) P(A|B_1)}{P(B_0) P(A|B_0) + P(B_1) P(A|B_1)} = 8,72 \dots \%$$

Für zwei gegebene Ereignisse A und B hängt häufig die Wahrscheinlichkeit, dass A eintritt nicht von dem Ereignis B ab, d.h. $P(A|B) = P(A)$. In diesem Fall gilt

$$P(A \cap B) = P(A) \cdot P(B).$$

Definition 2.10. Zwei Ereignisse A und B heißen (*stochastisch*) *unabhängig*, falls

$$P(A \cap B) = P(A) P(B).$$

Es bezeichne I eine beliebige Indexmenge. Wir nennen die Ereignisse A_i ($i \in I$) *unabhängig*, falls für alle endlichen Teilmengen $J \subset I$ die Gleichung

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

gilt.

Achtung: Um zu zeigen, dass gegebene Ereignisse A_1, \dots, A_n unabhängig sind, reicht es nicht die paarweise Unabhängigkeit, d.h. $P(A_i \cap A_j) = P(A_i) P(A_j)$ ($i \neq j$), zu überprüfen.

Beispiel 2.11. Wir betrachten nochmals Beispiel 2.2.

Frage: Sind die Ereignisse A (ges. Augensumme = 7) und B (erster Würfel zeigt 3 Augen) unabhängig?

Wir haben gesehen, dass $P(A|B) = \frac{1}{6}$. Da

$$P(A) = \frac{1}{36} |\{(i, j) : i + j = 7\}| = \frac{6}{36} = \frac{1}{6},$$

sind die Ereignisse A und B unabhängig.

Analog zur Definition der Unabhängigkeit kann man auch eine bedingte Unabhängigkeit einführen, d.h. gegeben ein Ereignis ist eingetreten sind die Ereignisse unabhängig:

Definition 2.12. Es bezeichne I eine Indexmenge und B ein Ereignis mit $P(B) > 0$. Die Ereignisse A_i ($i \in I$) heißen *unabhängig* gegeben B , falls für alle endlichen Teilmengen $J \subset I$ die Gleichung

$$P\left(\bigcap_{j \in J} A_j | B\right) = \prod_{j \in J} P(A_j | B)$$

gilt.

Eine Aussage, die wir später nochmals benötigen werden, ist das Lemma von Borel-Cantelli.

Lemma 2.13. *Es seien A_n ($n \in \mathbb{N}$) Ereignisse.*

i) *Falls $\sum_{n=1}^{\infty} P(A_n) < \infty$ ist, gilt*

$$P(\infty\text{-viele der Ereignisse } A_n \text{ treten ein}) = 0.$$

ii) *Falls $\sum_{n=1}^{\infty} P(A_n) = \infty$ ist und die Ereignisse A_n ($n \in \mathbb{N}$) unabhängig sind, gilt*

$$P(\infty\text{-viele der Ereignisse } A_n \text{ treten ein}) = 1.$$

Beweis. Übungsaufgabe!

Hinweis: Es bezeichne A das Ereignis, dass ∞ -viele Ereignisse A_n eintreten. Dann gilt

$$A = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

□

Lernziele des 2. Kapitels

- Bedingte Verteilungen
- Unabhängigkeit von Ereignissen
- Formel von der totalen Wahrscheinlichkeit
- Bayes-Umkehr-Formel

3 Diskrete Zufallsvariablen

Häufig interessieren wir uns für (reelle) Zufallsgrößen, die von einem Zufallsexperiment abhängen.

Beispiel 3.1. Wir wiederholen einen zufälligen Münzwurf n -mal und setzen jedes mal einen Euro auf “Kopf”. Falls wir gewinnen, erhalten wir zwei Euro zurück, andernfalls nichts. Ein wichtiger zufälliger Wert ist nun der Gewinn bzw. Verlust nach n Spielen.

Modellierung: $\Omega = \{K, Z\}^n$, P Gleichverteilung auf Ω

Wir interessieren uns für Eigenschaften des zufälligen Werts

$$\begin{aligned} S_n(\omega) &= \sum_{i=1}^n (2 \cdot 1_{\{K\}}(\omega_i) - 1) \\ &= \text{Anzahl “K” in } \omega - \text{Anzahl “Z” in } \omega. \end{aligned}$$

Eine Abbildung

$$X : \Omega \rightarrow X(\Omega)$$

mit abzählbarem Bild $X(\Omega)$, heißt *diskrete Zufallsvariable*. Wir werden vorerst ausschließlich mit diskreten Zufallsvariablen rechnen und bezeichnen jeweils mit Ω_X das diskrete Bild der Zufallsvariablen X .

Notation 3.2. Zur Rechnung mit Zufallsvariablen hat sich die folgende Schreib- und Sprechweise eingebürgert:

$$\begin{aligned} \{X \in A\} &= \{\omega \in \Omega : X(\omega) \in A\} && \text{“}X \text{ nimmt einen Wert aus } A \text{ an”} \\ \{X = x\} &= \{\omega \in \Omega : X(\omega) = x\} && \text{“}X \text{ nimmt den Wert } x \text{ an”} \\ \{X \leq x\} &= \{\omega \in \Omega : X(\omega) \leq x\} && \text{“}X \text{ nimmt einen Wert kleiner gleich } x \text{ an”}. \end{aligned}$$

Analog werden wir auch andere Gleichungen mit Zufallsvariablen in Ereignisse übersetzen.

Sei A eine Teilmenge von Ω_X . Wir interessieren uns nun für die Wahrscheinlichkeit der Menge $\{X \in A\}$, d.h. $P(\{X \in A\})$ oder kurz $P(X \in A)$. Damit diese wohldefiniert ist, müssen wir zusätzlich fordern, dass für eine diskrete Zufallsvariable X alle Mengen $\{X \in A\}$ ($A \subset \Omega_X$) Ereignisse sind. Dies stellt keine Einschränkung für unsere Berechnungen dar, da alle Zufallsvariablen, die wir aus natürlichen Fragestellungen ableiten werden, dies erfüllen werden.

Satz 3.3. Die Abbildung P_X gegeben durch $P_X(A) := P(X \in A)$ ($A \subset \Omega_X$) ist eine diskrete W'keitsverteilung über Ω_X mit Wahrscheinlichkeitsgewichten

$$f(x) := f_X(x) := P(X = x) \quad (x \in \Omega_X).$$

Definition 3.4. Wir nutzen die folgenden Bezeichnungen:

Verteilung der Zufallsvariable X : P_X

Wahrscheinlichkeitsgewichte von X : $f_X(x)$ ($x \in \Omega_X$)

Falls zwei diskrete Zufallsvariablen X und Y die gleichen W'keitsgewichte besitzen nennen wir X und Y *identisch verteilt*.

Bemerkung 3.5. Der Satz sagt insbesondere aus, dass alle Rechenregeln, die wir in den Sätzen 1.6 und 1.9 hergeleitet haben, auch auf diskrete Zufallsvariablen anwendbar sind, d.h. wir können mit Zufallsvariablen genauso rechnen, wie wir es bereits von dem Rechnen mit Verteilungen gewohnt sind. Zum Beispiel gilt für zwei Teilmengen A und B von Ω_X , dass

$$P(X \in A \cup B) = P(X \in A) + P(X \in B) - P(X \in A \cap B).$$

Bemerkung 3.6 (Zufallsvariablen in der Modellbildung (I)). Bisher wurden Zufallsexperimente durch einen Merkmalraum und ein Wahrscheinlichkeitsmaß modelliert. Im weiteren werden wir häufig zur Modellierung Zufallsvariablen verwenden, d.h. wir geben den Wertebereich Ω_X und die Wahrscheinlichkeitsgewichte f_X einer Zufallsvariable X an. Zum Beispiel können wir zur Modellierung eines Münzwurf eine Zufallsvariable X , die mit W'keit $\frac{1}{2}$ jeweils einen der Werte 'K' oder 'Z' annimmt, betrachten. Wir nennen nun nicht explizit den Raum Ω und die entsprechende Wahrscheinlichkeitsverteilung auf Ω .

Beweis von Satz 3.3. Es gilt

$$\begin{aligned} P_X(A) &= P(X \in A) = P(\{\omega : X(\omega) \in A\}) \\ &= P\left(\bigcup_{x \in A} \{X = x\}\right) = \sum_{x \in A} \underbrace{P(X = x)}_{=f(x)}. \end{aligned}$$

Insbesondere gilt

$$\sum_{x \in \Omega_X} f(x) = \sum_{x \in \Omega_X} P(X = x) = P(X \in \Omega_X) = 1.$$

Damit ist $(f(x))$ eine Folge von W'keitsgewichten und P_X die zugehörige diskrete Verteilung. \square

Beispiel 3.7. Wir betrachten nochmals Beispiel 3.1 mit $n = 2$ (zwei Münzwürfe). In diesem Fall ist $\Omega_{S_2} = \{-2, 0, 2\}$ und die entsprechenden W'keitsgewichte sind gegeben durch

$$f(k) = P(S_2 = k) = \begin{cases} P(\{KK\}) = 1/4 & k = 2 \\ P(\{KZ, ZK\}) = 1/2 & k = 0 \\ P(\{ZZ\}) = 1/4 & k = -2. \end{cases}$$

Lemma 3.8. Seien X_1, \dots, X_n beliebige diskrete Zufallsvariablen und φ eine Abbildung mit Definitionsbereich $\Omega_{X_1} \times \dots \times \Omega_{X_n}$. Dann ist

$$Z := \varphi(X_1, \dots, X_n)$$

eine diskrete Zufallsvariable mit W'keitsgewichten $(f_Z(z))$ gegeben durch

$$f_Z(z) = \sum_{x_1 \in \Omega_{X_1}} \dots \sum_{x_n \in \Omega_{X_n}} 1_{\{\varphi(x_1, \dots, x_n) = z\}} P(X_1 = x_1, \dots, X_n = x_n).$$

Bemerkung 3.9. Das Lemma garantiert, dass eine beliebige Verknüpfung von endlich vielen diskreten ZV'en wieder eine diskrete ZV ist. Z.Bsp. sind für zwei gegebene diskrete reelle ZV'en X und Y die Ausdrücke $X \cdot Y$ und $X + Y$ wieder diskrete Zufallsvariablen.

3.1 Unabhängigkeit von Zufallsvariablen

Das Konzept der Unabhängigkeit von Ereignissen lässt sich auch auf Zufallsvariablen übertragen.

Definition 3.10. Zwei diskrete Zufallsvariablen X und Y heißen *unabhängig*, falls

$$P(X = x, Y = y) = P(X = x) P(Y = y)$$

für alle $x \in \Omega_X$ und $y \in \Omega_Y$ gilt. Allgemeiner nennen wir für eine Indexmenge I die Zufallsvariablen X_i ($i \in I$) *unabhängig*, falls für alle endlichen Teilmengen $J \subset I$ und alle $(x_j)_{j \in J} \in \prod_{j \in J} \Omega_{X_j}$

$$P\left(\bigcap_{j \in J} \{X_j = x_j\}\right) = \prod_{j \in J} P(X_j = x_j)$$

gilt.

Die mathematische Definition formalisiert den intuitiven Begriff der Unabhängigkeit von Zufallsexperimenten.

Bemerkung 3.11. Wir betrachten nochmals das Münzwurfexperiment aus Beispiel 3.1. Für dieses Experiment kann man zeigen, dass die Zufallsvariablen

$$X_i(\omega) = 1_{\{K\}}(\omega_i) = \begin{cases} 1, & \text{falls der } i\text{-te Wurf "K" liefert} \\ 0, & \text{falls der } i\text{-te Wurf "Z" liefert} \end{cases} \quad (\omega \in \Omega)$$

für $i = 1, \dots, n$ unabhängig sind, d.h. unsere Anforderung, dass die Münzwürfe sich nicht gegenseitig beeinflussen, wird durch das gegebene Modell erfüllt. Wir verzichten an dieser Stelle auf einen Beweis der Unabhängigkeit.

Bemerkung 3.12 (Zufallsvariablen in der Modellbildung (II)). Wie bereits in Bemerkung 3.6 ausgeführt, kann die Modellierung von Zufallsexperimenten durch die Nennung von Zufallsvariablen und ihrer Verteilung (Verteilungsgewichte) erfolgen. Wird das Experiment durch mehrere Zufallsvariablen beschrieben, z.B. durch X_1, \dots, X_n , so beschreibt dieser Vorgang noch nicht die Abhängigkeitsstruktur der einzelnen Zufallsvariablen untereinander. Hierzu müsste man die Verteilung (W'keitsgewichte) von dem *Zufallsvektor* (X_1, \dots, X_n) angeben. Geht man in der Modellbildung davon aus, dass die Zufallsvariablen *unabhängig* sind, so reicht es die Verteilungen der einzelnen Zufallsvariablen anzugeben.

Wir können nun mit Hilfe von Lemma 3.8 die W'keitsgewichte der Summe zweier unabhängiger \mathbb{Z} -wertiger Zufallsvariable berechnen:

Lemma 3.13. *Es seien X und Y zwei unabhängige \mathbb{Z} -wertige ZV'en. Dann hat $Z = X + Y$ die W'keitsgewichte*

$$f_Z(z) = (f_X * f_Y)(z) \quad (z \in \mathbb{Z}),$$

wobei $*$ die Faltung zweier Funktionen bezeichnet, d.h.

$$(f_X * f_Y)(z) = \sum_{x \in \mathbb{Z}} f_X(x) \cdot f_Y(z - x).$$

Beweis. Wegen Lemma 3.8 und der Unabhängigkeit von X und Y gilt

$$\begin{aligned} f_Z(z) &= \sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{Z}} 1_{\{x+y=z\}} P(X = x, Y = y) \\ &= \sum_{x \in \mathbb{Z}} P(X = x, Y = z - x) = \sum_{x \in \mathbb{Z}} f_X(x) \cdot f_Y(z - x). \end{aligned}$$

□

Wir zeigen einige weitere hilfreiche Aussagen für unabhängige ZV'en.

Lemma 3.14. *Für zwei unabhängige Zufallsvariablen X und Y sind die Ereignisse*

$$\{X \in A\} \quad \text{und} \quad \{Y \in B\}$$

für alle $A \subset \Omega_X, B \subset \Omega_Y$ unabhängig.

Beweis. Es gilt

$$P(X \in A, Y \in B) = P\left(\bigcup_{\substack{(x,y): x \in A, \\ y \in B}} \{X = x, Y = y\}\right).$$

Die Vereinigung in der Formel ist eine abzählbare Vereinigung disjunkter Mengen, sodass

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{\substack{(x,y): x \in A, \\ y \in B}} P(X = x, Y = y) \\ &= \sum_{x \in A} \sum_{y \in B} P(X = x) \cdot P(Y = y) \\ &= \left(\sum_{x \in A} P(X = x)\right) \left(\sum_{y \in B} P(Y = y)\right) \\ &= P(X \in A) \cdot P(Y \in B). \end{aligned}$$

□

Lemma 3.15. *Es seien X und Y unabhängige ZV'en und $\varphi : \Omega_X \rightarrow \varphi(\Omega_X)$ und $\psi : \Omega_Y \rightarrow \psi(\Omega_Y)$ zwei beliebige Abbildungen.*

Dann sind auch die Zufallsvariablen $\varphi(X)$ und $\psi(Y)$ unabhängig.

Beweis. Setze $Z_1 = \varphi(X)$ und $Z_2 = \psi(Y)$. Nun gilt für $z_1 \in \Omega_{Z_1}$ und $z_2 \in \Omega_{Z_2}$ mit Hilfe von Lemma 3.14:

$$\begin{aligned} P(Z_1 = z_1, Z_2 = z_2) &= P(X \in \varphi^{-1}(\{z_1\}), Y \in \psi^{-1}(\{z_2\})) \\ &= P(X \in \varphi^{-1}(\{z_1\})) \cdot P(Y \in \psi^{-1}(\{z_2\})) \\ &= P(Z_1 = z_1) \cdot P(Z_2 = z_2). \end{aligned}$$

□

3.2 Bedingte Verteilungen für Zufallsvariablen

In Abschnitt 2 haben wir bedingte Wahrscheinlichkeiten eingeführt. Wir erweitern dieses Konzept nun auf diskrete Zufallsvariablen.

Satz 3.16. Für ein Ereignis B mit $P(B) > 0$ und eine diskrete Zufallsvariable X definiert

$$P_{X|B}(A) := P(X \in A|B) \quad (A \subset \Omega_X)$$

eine diskrete Wahrscheinlichkeitsverteilung auf Ω_X mit Wahrscheinlichkeitsgewichten

$$f_{X|B}(x) := P(X = x|B) \quad (x \in \Omega_X).$$

Beweis. Wir bezeichnen mit Q die bedingte Wahrscheinlichkeit

$$Q(A) = P(A|B) \quad (A \text{ Ereignis}).$$

Nun ist wegen Satz 3.3

$$P_{X|B}(A) = Q(X \in A) \quad (A \subset \Omega_X)$$

eine diskrete Wahrscheinlichkeitsverteilung mit W'keitsgewichten

$$Q(X = x) = P(X = x|B) = f_{X|B}(x) \quad (x \in \Omega_X)$$

□

Definition 3.17. Für ein Ereignis B mit $P(B) > 0$ und eine Zufallsvariable X nennen wir

$$P_{X|B}(A) := P(X \in A|B) \quad (A \subset \Omega_X)$$

bedingte (Wahrscheinlichkeits-)Verteilung von X gegeben B . Wir benutzen $P_{X|B}$ und $f_{X|B}$ als Standardnotationen für die bedingte Verteilung und deren Wahrscheinlichkeitsgewichte.

3.3 Beispiele diskreter reeller Zufallsvariablen

In diesem Abschnitt führen wir einige wichtige Zufallsvariablen ein.

Bernoulli Zufallsvariable.

Eine Zufallsvariable X heißt Bernoulli Variable (mit Erfolgswahrscheinlichkeit p), falls

- $\Omega_X = \{0, 1\}$
- $P(X = 1) = p$ und $P(X = 0) = 1 - p$

Wir werten den Ausgang $X = 1$ als Erfolg.

Binomialverteilung.

Experiment: Ein Computernetz besteht aus einem Zentralrechner und n Terminals. In einem gegebenen Zeitintervall, greift ein Terminal auf den Zentralrechner mit W'keit p zu, wobei die Zugriffe der Terminals als unabhängig angesehen werden.

Frage: Was ist die Verteilung der Anzahl der Zugriffe auf den Server ?

Modellierung:

Es seien X_1, \dots, X_n unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit p . Wir interessieren uns nun für die Anzahl der Erfolge nach n -maliger Durchführung des Experiments:

$$S_n = \sum_{i=1}^n X_i.$$

Es ist $\Omega_{S_n} = \{0, 1, \dots, n\}$ und wir erhalten für $k \in \Omega_{S_n}$:

$$P(S_n = k) = P(X_1 + \dots + X_n = k) = P\left(\bigcup_{\substack{(x_1, \dots, x_n): \\ x_1 + \dots + x_n = k}} \{X_1 = x_1, \dots, X_n = x_n\}\right).$$

Die Vereinigung ist eine endliche disjunkte Vereinigung, also folgt

$$\begin{aligned} P(S_n = k) &= \sum_{\substack{(x_1, \dots, x_n) \in \{0, 1\}^n: \\ x_1 + \dots + x_n = k}} P(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{\substack{(x_1, \dots, x_n) \in \{0, 1\}^n: \\ x_1 + \dots + x_n = k}} p^k (1 - p)^{n-k}. \end{aligned}$$

Es verbleibt die Anzahl der Summanden zu berechnen. Es bezeichne $\Sigma = \{(z_1, \dots, z_n) \in \{0, 1\}^n : z_1 + \dots + z_n = k\}$ die Indexmenge der Summation. Für einen Eintrag $\mathbf{x} = (x_1, \dots, x_n) \in \Sigma$ betrachten wir die Menge der Erfolge:

$$F_{\mathbf{x}} = \{i \in \{1, \dots, n\} : x_i = 1\}.$$

Die Menge $F_{\mathbf{x}}$ enthält genau k Elemente. Umgekehrt findet man für jede k -elementige Menge $F \subset \{1, \dots, n\}$ genau einen Eintrag $\mathbf{x} \in \Sigma$ mit $\mathbf{x} = (1_F(i))_{i=1, \dots, n}$. Also ist die Anzahl $|\Sigma|$ gerade gleich der Anzahl der k -elementigen Teilmengen von $\{1, \dots, n\}$ und wir folgern:

$$|\Sigma| = \binom{n}{k} = \frac{n \dots (n - k + 1)}{k!}.$$

Wir erhalten, dass

$$f_{S_n}(k) := P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Die Verteilung von S_n nennen wir *Binomialverteilung* oder kurz $b(n, p)$ -Verteilung. Die Zufallsvariable S_n heißt binomial-verteilte Zufallsvariable (oder $b(n, p)$ -verteilte ZV).

Geometrische Verteilung.

Experiment: Wir warten auf das erste Eintreten eines Ereignisses (z.Bsp. beim Würfeln mit zwei Würfeln, der erste Pasch), wobei wir voraussetzen, dass die Ausgänge sich nicht gegenseitig beeinflussen und jeweils mit der gleichen W'keit ein Erfolg eintritt.

Modellierung:

Es seien X_1, X_2, \dots unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit $p \in (0, 1]$, und es bezeichne Y die Zufallsvariable gegeben durch

$$Y = \min\{i \in \mathbb{N} : X_i = 1\},$$

d.h. die Wartezeit auf den ersten erfolgreichen Ausgang. Nun gilt $\Omega_Y = \mathbb{N}$ und

$$P(Y = k) = P(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) = (1 - p)^{k-1} p \quad (k \in \mathbb{N}).$$

Die Verteilung von Y heißt *geometrische Verteilung* zum Parameter p , und Y wird geometrisch verteilte Zufallsvariable genannt.

Übungsaufgabe

- Man zeige, dass für einen Parameter $\lambda > 0$, die Folge $(e^{-\lambda} \frac{\lambda^k}{k!})_{k \in \mathbb{N}_0}$ eine Folge von Wahrscheinlichkeitsgewichten ist (*Poisson Verteilung* mit Parameter λ ; kurz $\text{Poiss}(\lambda)$).
- Seien $\lambda_1, \lambda_2 > 0$ und X und Y unabhängige Poisson-verteilte Zufallsvariablen mit Parametern λ_1 und λ_2 . Wie ist nun $X + Y$ verteilt? Hierzu sind die Wahrscheinlichkeitsgewichte von $X + Y$ zu berechnen.
- Man bestimme die bedingte Verteilung von X gegeben $X + Y = n$ für ein $n \in \mathbb{N}_0$, d.h. $P_{X|X+Y=n}$.

Lernziele des 3. Kapitels

- Diskrete Zufallsvariablen
- Modellierung mittels Zufallsvariablen
- Unabhängige Zufallsvariablen
- Verteilungen von Summen unabhängiger Zufallsvariablen (Faltung)
- Spezielle Verteilungen

4 Kenngrößen reeller Zufallsvariablen

4.1 Erwartungswert

In diesem Abschnitt führen wir die erste wichtige Kenngröße für reelle Zufallsvariablen ein.

Angenommen x_1, \dots, x_n sind die Ausgänge (Realisierungen) einer Zufallsvariable X von n sich nicht beeinflussenden Experimenten. Wir betrachten das Mittel

$$m(n; x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i = \sum_{x \in \Omega_X} x \underbrace{\frac{1}{n} \sum_{i=1}^n 1_{\{x_i=x\}}}_{=\text{relative Häufigkeit von } x \text{ in } (x_1, \dots, x_n)}$$

Die Erfahrung zeigt, dass die relative Häufigkeit von x für große n typischerweise nahe bei $f_X(x)$ liegt, d.h. wir erwarten, dass der zufällige Wert $m(n; x_1, \dots, x_n)$ nahe bei

$$\sum_{x \in \Omega_X} x P(X = x)$$

liegt.

Definition 4.1. Wir nennen eine diskrete reelle Zufallsvariable X *integrierbar*, falls

$$\sum_{x \in \Omega_X} |x| f_X(x) < \infty. \quad (1)$$

Für eine integrierbare Zufallsvariable X heißt

$$E(X) = \sum_{x \in \Omega_X} x f_X(x) = \sum_{x \in \Omega_X} x P(X = x) \quad (2)$$

Erwartungswert von X . Die Bedingung (1) garantiert die absolute Summierbarkeit der Folge, sodass die Reihe gegen einen eindeutigen endlichen Wert konvergiert.

Bemerge: Der Erwartungswert von X hängt nur von der Verteilung von X (W'keitsgewichten) ab. Insbesondere haben identisch verteilte Zufallsvariablen die gleichen Erwartungswerte.

Falls die Integrierbarkeitsbedingung (1) nicht erfüllt ist, aber die Summe über die negativen oder positiven Summanden in (2) endlich ist, ist der Erwartungswert dennoch wohldefiniert. In diesem Fall erhält man den Erwartungswert ∞ bzw. $-\infty$. Insbesondere ist der Erwartungswert für jede Zufallsvariable, die ausschließlich nichtnegative Werte annimmt immer wohldefiniert.

Beispiel 4.2. Für zwei Münzwürfe (siehe Beispiele 3.1 und 3.7) erhalten wir

$$E(S_2) = 2 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} - 2 \cdot \frac{1}{4} = 0.$$

Zum Rechnen mit Erwartungswerten sind folgende Regeln hilfreich:

Satz 4.3. *Es seien X und Y Zufallsvariablen, die jeweils integrierbar oder nichtnegativ sind.*

(i) *Der Erwartungswert der Summe $X + Y$ ist wohldefiniert und es gilt*

$$E(X + Y) = E(X) + E(Y).$$

(ii) *Für $a \in \mathbb{R}$ gilt*

$$a E(X) = E(a X),$$

wobei wir hier die Konvention $0 \cdot \infty = 0$ annehmen.

(iii) *Falls X eine positive Zufallsvariable ist, gilt*

$$E(X) \geq 0.$$

(iv) *Für die konstante Zufallsvariable $1 = 1_\Omega$ gilt*

$$E(1) = 1.$$

Wir werden nur Teil (i) beweisen. Hierzu nutzen wir den folgenden Satz.

Satz 4.4. *Es seien X_1, \dots, X_n beliebige diskrete Zufallsvariablen und $\varphi : \Omega_{X_1} \times \dots \times \Omega_{X_n} \rightarrow \mathbb{R}$ eine beliebige Abbildung. Es gilt*

$$E\varphi(X_1, \dots, X_n) = \sum_{x_1 \in \Omega_{X_1}} \dots \sum_{x_n \in \Omega_{X_n}} \varphi(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n),$$

wobei der Erwartungswert wohldefiniert ist, wenn die Summe über die negativen Einträge endlich ist (also z.Bsp. wenn φ nur positive Werte annimmt oder $\varphi(X_1, \dots, X_n)$ integrierbar ist).

Beweis. Setze $Z = \varphi(X_1, \dots, X_n)$. Mit Hilfe von Lemma 3.8 erhalten wir

$$\begin{aligned} E(Z) &= \sum_{z \in \Omega_Z} z f_Z(z) \\ &= \sum_{z \in \Omega_Z} z \left(\sum_{x_1 \in \Omega_{X_1}} \cdots \sum_{x_n \in \Omega_{X_n}} 1_{\{\varphi(x_1, \dots, x_n) = z\}} P(X_1 = x_1, \dots, X_n = x_n) \right). \end{aligned}$$

Vertauschen der Summationsreihenfolge liefert also

$$\begin{aligned} E(Z) &= \sum_{x_1 \in \Omega_{X_1}} \cdots \sum_{x_n \in \Omega_{X_n}} P(X_1 = x_1, \dots, X_n = x_n) \left(\sum_{z \in \Omega_Z} 1_{\{\varphi(x_1, \dots, x_n) = z\}} z \right) \\ &= \sum_{x_1 \in \Omega_{X_1}} \cdots \sum_{x_n \in \Omega_{X_n}} \varphi(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n). \end{aligned}$$

□

Beweis von Satz 4.3 (i). Es gilt

$$\begin{aligned} E(X + Y) &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} (x + y) P(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} x P(X = x, Y = y) + \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} y P(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} x \sum_{y \in \Omega_Y} P(X = x, Y = y) + \sum_{y \in \Omega_Y} y \sum_{x \in \Omega_X} P(X = x, Y = y) \\ &= \sum_{x \in \Omega_X} x P(X = x) + \sum_{y \in \Omega_Y} y P(Y = y) = E(X) + E(Y). \end{aligned}$$

□

4.2 Erwartungswerte wichtiger diskreter Zufallsvariablen

In diesem Abschnitt berechnen wir die Erwartungswerte der in Abschnitt 3.3 eingeführten Zufallsvariablen. Weiterhin werden wir eine neue Verteilung einführen.

Bernoulli Zufallsvariable.

Es bezeichne X eine Bernoulli Variable mit Erfolgswahrscheinlichkeit p . Dann gilt

$$E(X) = p \cdot 1 + (1 - p) \cdot 0 = p.$$

Binomialverteilung.

Es seien X_1, \dots, X_n unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit p . Dann ist $S_n = \sum_{i=1}^n X_i$ eine $b(n, p)$ -verteilte Zufallsvariable und es gilt

$$E(S_n) = E(X_1) + \dots + E(X_n) = np.$$

Geometrische Verteilung.

Es sei Y eine geometrisch-verteilte Zufallsvariable zum Parameter $p \in (0, 1]$, d.h.

$$P(Y = k) = (1 - p)^{k-1} p \quad (k \in \mathbb{N}).$$

Nun ist der Erwartungswert von Y gegeben durch

$$E(Y) = p \sum_{k=1}^{\infty} k (1 - p)^{k-1}.$$

Wir benutzen folgenden Rechentrick zur Berechnung der Reihe: Wir differenzieren die Potenzreihe

$$g(q) := \sum_{k=0}^{\infty} q^k = \frac{1}{1 - q} \quad (|q| < 1)$$

und erhalten

$$g'(q) = \sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1 - q)^2}.$$

Nun kann man $q = 1 - p$ wählen und erhält

$$E(Y) = p \frac{1}{p^2} = \frac{1}{p}.$$

Poisson Verteilung (Poisson Approximation).

Experiment: Wir betrachten wieder das Netzwerk mit einem Zentralrechner und n Terminals. Angenommen die Anzahl der Terminals ist sehr groß und die Zugriffswahrscheinlichkeit sehr klein (Z.Bsp. $n = 1000$, $p = 0,002$).

Frage: Gibt es eine einfachere Verteilung als die Binomialverteilung um dieses Problem “gut” zu modellieren ?

Alternative Frage: Wie soll man die Verteilung modellieren, wenn wir den Erwartungswert der Zugriffe kennen, die Anzahl der Terminals aber unbekannt ist?

Wir bezeichnen mit $\lambda > 0$ die erwartete Anzahl an Zugriffen und lassen n (die Anzahl der Terminals) gegen ∞ laufen. Damit der Erwartungswert konstant λ bleibt, müssen wir die Erfolgswahrscheinlichkeit $p_n = \lambda/n$ wählen. Die entsprechende $b(n, p_n)$ -verteilte Zufallsgröße bezeichnen wir mit S_n . Nun gilt für ein $k \in \mathbb{N}_0$:

$$\begin{aligned} P(S_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{1}{k!} n \dots (n - k + 1) \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \underbrace{\frac{n}{n} \dots \frac{n - k + 1}{n}}_{\rightarrow 1 \text{ für } n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

Weiterhin gilt

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \exp\left\{\underbrace{\frac{n-k}{n}}_{\rightarrow 1} \ln\left(\underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}}\right)\right\} \rightarrow e^{-\lambda}$$

und wir folgern, dass

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda} =: f(k) \quad (k \in \mathbb{N}_0).$$

Die Folge $(f(k))_{k \in \mathbb{N}_0}$ ist eine Folge von Wahrscheinlichkeitsgewichten, da

$$\sum_{k=0}^{\infty} f(k) = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k}_{=e^{\lambda}} = 1.$$

Die zugehörige Verteilung nennen wir *Poisson-Verteilung* zum Parameter $\lambda > 0$.

In unserem Beispiel ist $n = 1000$ und $p = 0,002$, also $\lambda = n \cdot p = 2$. Dann können wir die Binomialverteilung durch die Poissonverteilung mit Parameter 2 approximieren. Die folgende Tabelle enthält einen Vergleich der beiden Verteilungen, wobei X eine Poisson-verteilte ZV zum Parameter 2 bezeichnet:

k	$P(S_n = k)$	$P(X = k)$
0	0.135065	0.135335
1	0.27067	0.270671
2	0.270942	0.270671
3	0.180628	0.180447
4	0.0902234	0.0902235
5	0.036017	0.0360894
6	0.0119696	0.0120298

Beispiel 4.5 (ALOHA Protokoll). Ein Computernetz bestehe aus einem Kanal auf dem jeder Teilnehmer in jedem Zeitfenster ein Signal senden kann. Senden mehrere Teilnehmer gleichzeitig werden die Nachrichten verstümmelt und müssen später erneut gesendet werden. Beim ALOHA Protokoll emuliert jeder Teilnehmer pro Zeitfenster unabhängig von den anderen eine Bernoulli Zufallsvariable. Liefert diese Zufallsvariable einen Erfolg sendet er seine Nachricht, andernfalls wartet er weiter ab.

Problem: Das Netz bestehe aus n Teilnehmern (wobei n groß ist), die alle eine Nachricht senden wollen. Wie muß die Erfolgswahrscheinlichkeit $p > 0$ der Bernoulli Zufallsvariablen gewählt werden, damit es mit maximaler Wahrscheinlichkeit eine erfolgreiche Übertragung einer Nachricht gibt?

Antwort: Die Erfolgswahrscheinlichkeit ist gerade $P(S_n = 1)$, wobei S_n eine $b(n, p)$ -verteilte Zufallsvariable ist. Da n groß ist können wir diese Wahrscheinlichkeit durch eine Poisson-verteilte ZV S zum Parameter $\lambda = np$ approximieren:

$$P(\text{erfolgreiche Übertragung}) = P(S = 1) = \lambda e^{-\lambda} =: g(\lambda).$$

Die Funktion g nimmt ihr globales Maximum für $\lambda = 1$ an, d.h. die Teilnehmer sollten $p = 1/n$ wählen. In diesem Fall ist die Wahrscheinlichkeit einer erfolgreichen Übertragung $P(S = 1) = e^{-1} \approx 36,8\%$.

4.3 Die Varianz

In diesem Abschnitt werden wir eine weitere Kenngröße für Zufallsvariablen einführen: die Varianz. Wir betrachten folgendes Beispiel.

Beispiel 4.6. Für ein gegebenes $t > 0$ bezeichne X eine Zufallsvariable mit

$$P(X = \pm t) = 1/2.$$

Die Zufallsvariable hat Erwartungswert $E(X) = 0$. Für große t sagt dies wenig über die Zufallsvariable aus, da die Realisierungen weit weg vom Erwartungswert liegen. Wir benötigen ein Maß für die Streuung bzw. die typische Abweichung von dem Mittelwert.

Definition 4.7. Für eine integrierbare Zufallsvariable X heißt

$$\text{var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

Varianz von X . Wie der Erwartungswert hängt die Varianz nur von der Verteilung der Zufallsvariable ab.

Die Varianz hat nicht dieselbe Einheit wie die Zufallsvariable. Deshalb betrachtet man häufig auch die *Standardabweichung* von X :

$$S(X) = \sqrt{\text{var}(X)}.$$

Zur Berechnung der Varianz können wir Satz 4.4 (i) anwenden:

Lemma 4.8. Für eine diskrete integrierbare ZV X gilt

$$\text{var}(X) = \sum_{x \in \Omega_X} (x - E(X))^2 f_X(x).$$

Definition 4.9. Für $n \in \mathbb{N}$ nennen wir

$$E(X^n)$$

das *n -te Moment von X* . Für ungerade n ist dies möglicherweise nicht wohldefiniert. Wir sagen X hat endliches n -tes Moment, falls

$$E(|X|^n) < \infty.$$

Lemma 4.10. Eine Zufallsvariable mit endlichem 2-ten Moment ist integrierbar und besitzt endliche Varianz.

Beweis. Übungsaufgabe

□

Lemma 4.11. Für zwei unabhängige reelle Zufallsvariablen X und Y mit endlichen 2-ten Momenten gilt

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

Beweis. Die Abschätzung $E(|XY|) \leq E(X^2 + Y^2) < \infty$ impliziert, dass $X \cdot Y$ integrierbar ist. Es folgt

$$\begin{aligned} E(XY) &= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} xy \underbrace{P(X=x, Y=y)}_{=P(X=x)P(Y=y)} \\ &= \sum_{x \in \Omega_X} x P(X=x) \sum_{y \in \Omega_Y} y P(Y=y) = E(X) E(Y). \end{aligned}$$

□

Definition 4.12. Wir nennen zwei Zufallsvariablen *unkorreliert*, falls

$$E(XY) = E(X) E(Y).$$

Wir haben gesehen, dass die Unabhängigkeit von zwei ZV'en die Unkorreliertheit impliziert. Wie das folgende Beispiel zeigt gilt die Umkehrung jedoch nicht!

Beispiel 4.13. Sei $\Omega = \{\omega_1, \omega_2, \omega_3\}$ und P die Gleichverteilung auf Ω . Wir betrachten die Zufallsvariablen

$$X(\omega) = \begin{cases} -1 & \omega = \omega_1 \\ 1 & \omega = \omega_2 \\ 0 & \omega = \omega_3 \end{cases}, \quad Y(\omega) = \begin{cases} -1 & \omega \in \{\omega_1, \omega_2\} \\ 2 & \omega = \omega_3 \end{cases}.$$

Dann ist $E(X) = E(Y) = E(XY) = 0$ und die Zufallsvariablen X und Y sind unkorreliert. Die Zufallsvariablen sind aber nicht unabhängig, da die Ereignisse $\{X = 0\}$ und $\{Z = 2\}$ nicht unabhängig sind.

Wir fassen nun einige elementare Eigenschaften der Varianz im folgenden Satz zusammen.

Satz 4.14. Es seien X, Y unkorrelierte ZV'en mit endlichen 2-ten Momenten. Dann gilt

$$(i) \operatorname{var}(aX + b) = a^2 \operatorname{var}(X) \quad \text{und} \quad S(aX + b) = |a| S(X)$$

$$(ii) \operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y)$$

Beweis. (i) Es gilt:

$$\begin{aligned} \operatorname{var}(aX + b) &= E((aX + b - E(aX + b))^2) = E((a(X - E(X)))^2) \\ &= a^2 E((X - E(X))^2) = a^2 \operatorname{var}(X). \end{aligned}$$

(ii) Es gilt

$$\begin{aligned} \operatorname{var}(X + Y) &= E((X + Y - E(X + Y))^2) \\ &= E((X - E(X))^2 + 2(X - E(X))(Y - E(Y)) + (Y - E(Y))^2) \\ &= \operatorname{var}(X) + 2(E(XY) - E(X)E(Y)) + \operatorname{var}(Y). \end{aligned}$$

Nun impliziert die Unkorreliertheit von X und Y , dass

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y).$$

□

Lemma 4.15. *Es bezeichne X eine Zufallsvariable mit endlichen 2-ten Momenten. Es gilt:*

$$\operatorname{var}(X) = 0 \iff X \text{ ist fast sicher konstant.}$$

4.4 Varianzen wichtiger reeller Zufallsvariablen

Bernoulli Zufallsvariable Sei X eine Bernoulli Zufallsvariable mit Erfolgswahrscheinlichkeit p . Es gilt

$$\operatorname{var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p).$$

Binomialverteilung Seien X_1, \dots, X_n unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit p . Dann ist

$$S_n = \sum_{i=1}^n X_i$$

$b(n, p)$ -verteilt und es gilt

$$\text{var}(S_n) = \text{var}(X_1) + \cdots + \text{var}(X_n) = n p (1 - p).$$

Poisson Verteilung Es bezeichne Y eine Poisson-verteilte Zufallsvariable zum Parameter λ . Nun gilt

$$\begin{aligned} E(Y^2) &= \sum_{k=0}^{\infty} k^2 \frac{1}{k!} e^{-\lambda} \lambda^k \\ &= \sum_{k=0}^{\infty} k(k-1) \frac{1}{k!} e^{-\lambda} \lambda^k + \sum_{k=0}^{\infty} k \frac{1}{k!} e^{-\lambda} \lambda^k \\ &= e^{-\lambda} \left(\underbrace{\lambda^2 \sum_{k=2}^{\infty} \frac{1}{(k-2)!} \lambda^{k-2}}_{=e^{\lambda}} + \lambda \underbrace{\sum_{k=1}^{\infty} \frac{1}{(k-1)!} \lambda^{k-1}}_{=e^{\lambda}} \right), \end{aligned}$$

sodass

$$\text{var}(Y) = E(Y^2) - (E(Y))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Geometrische Verteilung

Es bezeichnen X_1, X_2, \dots unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit p . Nun ist $Y = \min\{i \in \mathbb{N} : X_i = 1\}$ eine geometrisch verteilte ZV für die

$$E(Y^2) = \sum_{k=0}^{\infty} k^2 p (1-p)^{k-1}.$$

Die Berechnung der Reihe erfolgt ähnlich, wie die Berechnung des Erwartungswerts und liefert

$$\text{var}(Y) = \frac{1-p}{p^2}.$$

Sie wird im Tutorium durchgeführt.

4.5 Die Kovarianz und Korrelation

Ein Maß für die gegenseitige Beeinflussung von Zufallsvariablen stellen die Kovarianz und die Korrelation dar.

Definition 4.16. Für zwei ZV'en X und Y mit endlichen 2-ten Momenten nennen wir

$$\text{cov}(X, Y) = E((X - E(X))(Y - E(Y)))$$

die *Kovarianz von X und Y* . Falls die Varianzen von X und Y strikt positiv sind, definieren wir

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}$$

und nennen den Wert *Korrelation zwischen X und Y* .

Die Kovarianz erfüllt die folgenden Rechenregeln:

Lemma 4.17. *Es seien X, Y und Z reelle Zufallsvariablen mit endlichen 2-ten Momenten und a, b beliebige Konstanten.*

- (i) $\text{var}(X) = \text{cov}(X, X)$
- (ii) $\text{cov}(X, Y) = \text{cov}(Y, X)$ (*Symmetrie*)
- (iii) $\text{cov}(aX + bY, Z) = a \text{cov}(X, Z) + b \text{cov}(Y, Z)$ (*Linearität*)

Das folgende Lemma verknüpft den Begriff der Unkorreliertheit mit der Korrelation:

Lemma 4.18. *Es seien X und Y zwei Zufallsvariablen mit endlichen 2-ten Momenten. Es gilt:*

$$X \text{ und } Y \text{ sind unkorreliert} \iff \text{cov}(X, Y) = 0.$$

Beweis. Übungsaufgabe. □

Im weiteren Verlauf bezeichnen wir mit X und Y zwei Zufallsvariablen mit endlichen 2-ten Momenten. Wir wollen nun Y durch eine Zufallsvariable

$$\tilde{Y} = aX + b$$

möglichst gut approximieren, wobei a und b geeignet zu wählende Konstanten sind.

Problem: Finde $a, b \in \mathbb{R}$ für die der *quadratische Fehler*

$$E((Y - (aX + b))^2)$$

minimal ist.

Technische Annahme: X und Y haben strikt positive Varianz. (Andernfalls ist eine der Zufallsvariablen fast sicher konstant und das Problem trivial.)

Wir bezeichnen mit $Z = Y - \tilde{Y}$ den (zufälligen) Approximationsfehler. Für ihn gilt

$$E(Z^2) = \text{var}(Z) + (E(Z))^2.$$

Die Varianz von Z hängt nicht von der Wahl des Parameters b ab. D.h. für ein gegebenes $a \in \mathbb{R}$ ist dasjenige $b \in \mathbb{R}$ optimal für das

$$E(Z) = E(Y) - E(\tilde{Y}) = E(Y) - aE(X) - b$$

gleich 0 ist; also $b = E(Y) - aE(X)$. Für diese Wahl folgt $E(Z^2) = \text{var}(Z)$. Es verbleibt $a \in \mathbb{R}$ zu bestimmen für das die Varianz $\text{var}(Z)$ minimal ist. Nun ist

$$\begin{aligned} \text{var}(Z) &= E(((Y - E(Y)) - a(X - E(X)))^2) \\ &= \text{var}(Y) - 2a \text{cov}(X, Y) + a^2 \text{var}(X) \end{aligned}$$

ein Polynom in a und wir erhalten mit Hilfe einer quadratischen Ergänzung:

$$\begin{aligned} \text{var}(Z) &= a^2 \text{var}(X) - 2a \text{cov}(X, Y) + \frac{\text{cov}(X, Y)^2}{\text{var}(X)} - \frac{\text{cov}(X, Y)^2}{\text{var}(X)} + \text{var}(Y). \\ &= \underbrace{a^2 \text{var}(X) - 2a \text{cov}(X, Y) + \frac{\text{cov}(X, Y)^2}{\text{var}(X)}}_{= \left(a \sqrt{\text{var}(X)} - \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}}\right)^2} + \text{var}(Y). \end{aligned}$$

Der Ausdruck wird minimal, wenn der quadratische Term gleich 0 ist, d.h. für

$$a = \frac{\text{cov}(X, Y)}{\text{var}(X)} = \rho(X, Y) \frac{S(Y)}{S(X)}.$$

Der optimale quadratische Approximationsfehler ist damit

$$\text{var}(Y) - \frac{\text{cov}(X, Y)^2}{\text{var}(X)} = \text{var}(Y) (1 - \rho^2(X, Y)).$$

Satz 4.19. *Es seien X und Y zwei Zufallsvariablen mit endlichen 2-ten Momenten und strikt positiven Varianzen. Die beste lineare Approximation $\tilde{Y} = aX + b$ für Y erhält man für*

$$a = \frac{\text{cov}(X, Y)}{\text{var}(X)} = \rho(X, Y) \frac{S(Y)}{S(X)} \quad \text{und} \quad b = E(Y) - aE(X).$$

Für diese Approximation ist der mittlere quadratische Fehler gerade

$$E((Y - \tilde{Y})^2) = \text{var}(Y) - \frac{\text{cov}(X, Y)^2}{\text{var}(X)} = \text{var}(Y) (1 - \rho^2(X, Y)).$$

Bemerkung 4.20. Für zwei Zufallsvariablen kann man generell die Frage stellen inwiefern die Kenntnis der einen Information über die andere liefert. In dem Satz zeigen wir, dass für lineare Approximationen gerade der Betrag der Korrelation anzeigt wie gut dies möglich ist.

Korollar 4.21. *Für zwei reelle Zufallsvariablen X und Y mit endlichen 2-ten Momenten und strikt positiven Varianzen, gilt*

$$|\rho(X, Y)| \leq 1.$$

Insbesondere gilt die Gleichheit genau dann, wenn es Konstanten a und b gibt, sodass $Y = aX + b$ fast sicher, d.h. $P(Y = aX + b) = 1$.

Man kann die Aussage des vorhergehenden Korollars alternativ wie folgt ausdrücken:

Korollar 4.22. (Cauchy-Schwarz-Ungleichung) *Für zwei Zufallsvariablen X und Y mit endlichen zweiten Momenten gilt:*

$$|\text{cov}(X, Y)| \leq \sqrt{\text{var}(X)} \sqrt{\text{var}(Y)}.$$

Lernziele des 4. Kapitels

- Varianz, Standardabweichung
- Kovarianz, Korrelation
- Unkorreliertheit
- Lineare Approximation von einer Zufallsvariablen durch eine andere
- Poisson-Approximation

5 Konvergenz von Zufallsvariablen

In diesem Kapitel werden wir erstmals Konvergenzbegriffe für Zufallsvariablen einführen. Weiterhin werden Techniken zum Beweisen von Konvergenzaussagen hergeleitet.

5.1 Die Chebychev Ungleichung

In dem vorhergehenden Kapitel haben wir die Einführung der Varianz als ein Maß für die Streuung motiviert. Wir werden in diesem Kapitel diesen Zusammenhang konkretisieren.

Lemma 5.1. (Chebychev Ungleichung) *Für eine Zufallsvariable X mit endlichem 2-ten Moment und $t > 0$ gilt*

$$P(|X - E(X)| \geq t) \leq \frac{\text{var}(X)}{t^2} = \left(\frac{S(X)}{t} \right)^2.$$

Beweis. Wir betrachten die Zufallsvariable $Y = X - E(X)$. Die Ungleichung $Y^2 \geq 1_{\{|Y| \geq t\}} t^2$ impliziert

$$\text{var}(X) = E(Y^2) \geq E(1_{\{|Y| \geq t\}} t^2) = t^2 P(|Y| \geq t) = t^2 P(|X - E(X)| \geq t).$$

□

Lemma 5.1 schätzt die W'keit von weit entfernten Realisierungen gegen einen Erwartungswert (die Varianz) ab. Indem wir die quadratische Funktion durch die Exponentialfunktion ersetzen, erhalten wir auf analogem Weg eine ähnliche Abschätzung.

Lemma 5.2. *Es sei X eine Zufallsvariable und $a > 0$. Angenommen*

$$E \exp(aX) < \infty. \tag{3}$$

Dann gilt für alle $t \geq 0$:

$$P(X \geq t) \leq \frac{E(e^{aX})}{e^{at}}.$$

Beweis. Übungsaufgabe

□

Bemerkung 5.3. Die Wahrscheinlichkeit von Realisationen fern des Erwartungswerts oder des Ursprungs kann man gegen bestimmte Erwartungswerte abschätzen (z.Bsp. die Varianz oder (3)). Falls der Erwartungswert endlich ist, impliziert dies, dass die Wahrscheinlichkeit großer Abweichungen von einer gewissen Ordnung gegen 0 konvergiert.

Beispiel 5.4. Wir betrachten eine Maschine, die in Massenproduktion Bauteile herstellt. Wir nehmen an, dass jedes Bauteil unabhängig von den anderen Bauteilen mit W'keit p fehlerhaft ist und mit W'keit $1 - p$ fehlerfrei ist. Angenommen wir kennen p noch nicht und wollen Aussagen über p treffen.

Modellierung: X_1, X_2, \dots unabhängige Bernoulli-verteilte ZV'en mit Erfolgswahrscheinlichkeit p .

Gesucht: p !

Antwort: Wir betrachten den Mittelwert $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ($n \in \mathbb{N}$). Die Zufallsvariable \bar{X}_n besitzt den Erwartungswert

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

und die Varianz

$$\text{var}(\bar{X}_n) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n} \underbrace{\text{var}(X_1)}_{=p(1-p)}.$$

Also gilt für $\varepsilon > 0$

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{(2\sqrt{n\varepsilon})^2},$$

d.h. für sehr große Stichproben (sehr große $n \in \mathbb{N}$) liegt der zufällige Wert \bar{X}_n sehr nahe bei der Wahrscheinlichkeit p . Insbesondere gilt

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - p| \geq \varepsilon) = 0.$$

In einem gewissen Sinn konvergieren die Zufallsvariablen \bar{X}_n gegen p .

Definition 5.5. Eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen *konvergiert in Wahrscheinlichkeit* gegen $a \in \mathbb{R}$, falls für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - a| > \varepsilon) = 0.$$

Wir schreiben in diesem Fall

$$\lim_{n \rightarrow \infty} X_n = a, \quad \text{in Wahrscheinlichkeit.}$$

Wir werden später noch weitere Konvergenzbegriffe kennenlernen.

5.2 Das schwache Gesetz der großen Zahlen

Wir haben in Beispiel 5.4 ein sogenanntes “Gesetz der großen Zahlen” das erste Mal kennengelernt. Der folgende Satz liefert ein ähnliches Ergebnis in einem allgemeineren Fall.

Satz 5.6. (Schwaches Gesetz der großen Zahlen) *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von paarweise unkorrelierten Zufallsvariablen mit identischem Erwartungswert $a = E(X_1)$ und*

$$\sup_{n \in \mathbb{N}} \text{var}(X_n) < \infty.$$

Dann gilt für $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ($n \in \mathbb{N}$) und alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - a| > \varepsilon) = 0,$$

d.h. (\bar{X}_n) konvergiert in Wahrscheinlichkeit gegen a .

Beweis von Satz 5.6. Wir betrachten den Erwartungswert und die Varianz von \bar{X}_n . Es gilt

$$E(\bar{X}_n) = \frac{1}{n} (\underbrace{E(X_1)}_{=a} + \cdots + \underbrace{E(X_n)}_{=a}) = a$$

und

$$\begin{aligned} \text{var}(\bar{X}_n) &= \text{cov}\left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) \leq \frac{C}{n}, \end{aligned}$$

wobei $C := \sup_{m \in \mathbb{N}} \text{var}(X_m)$. Also gilt für alle $\varepsilon > 0$ wegen der Chebychev Ungleichung

$$P(|\bar{X}_n - a| \geq \varepsilon) \leq \frac{\text{var}(\bar{X}_n)}{\varepsilon^2} \leq \frac{C}{n \varepsilon^2} \longrightarrow 0, \quad n \rightarrow \infty.$$

□

Das schwache Gesetz der großen Zahlen lässt sich analog auch unter schwächeren Bedingungen an die Kovarianz beweisen:

Satz 5.7. (Schwachtes Gesetz der großen Zahlen) *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit identischem Erwartungswert $a = E(X_1)$ und*

$$\text{cov}(X_n, X_m) \leq C \frac{1}{(1 + |n - m|)^\alpha} \quad (n, m \in \mathbb{N})$$

für geeignete Konstanten $C > 0$ und $\alpha \in (0, 1)$. Dann gilt für $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ($n \in \mathbb{N}$) und alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - a| > \varepsilon) = 0.$$

Beweis. Übungsaufgabe. □

5.3 Shannon's Quellkodierungssatz

Als eine Anwendung der Grenzwertsätze werden wir nun Shannon's Quellkodierungssatz herleiten.

Experiment: Ein Quellmedium liefert unabhängig identisch verteilte Zeichen X_1, X_2, \dots aus einer endlichen Menge $\Sigma := \Omega_{X_1}$ (*Alphabet*). Wir bezeichnen mit $f(x) = f_{X_1}(x)$ die entsprechenden Wahrscheinlichkeitsgewichte.

Frage: Wieviel Bit benötigt man zur Speicherung der ersten n Zeichen ?

1. Antwort: Ein Signal der Länge n liegt in der endlichen Menge

$$\Sigma^n = \underbrace{\Sigma \times \dots \times \Sigma}_{n\text{-mal}}.$$

Da wir mit m -Bit langen 0-1-Zeichenketten gerade 2^m verschiedene Werte indexieren können, reicht es also $m \in \mathbb{N}$ so zu wählen, dass

$$2^m \geq |\Sigma|^n$$

oder äquivalent

$$m \geq n \cdot \log_2 |\Sigma|.$$

D.h. wenn wir m so wählen, können wir sicherstellen, dass jedem Element aus Σ^n ein eindeutiger 0-1-String der Länge m zugeordnet werden kann.

Die Antwort ist in gewisser Weise nicht zufriedenstellend, da möglicherweise manche Werte in Σ^n häufiger und manche Werte seltener angenommen

werden, wir aber für alle die gleiche Stringlänge zur Kodierung verwenden. Um das Verfahren zu verbessern sollte man für “typische” Eingangssignale kürzere Stringlängen wählen und dafür längere Stringlängen für untypische Eingangssignale in Kauf nehmen.

Satz 5.8. (Shannon’s Quellkodierungssatz)

Es bezeichne

$$H(X_1) = - \sum_{x \in \Sigma} f(x) \log_2 f(x)$$

die Entropie von X_1 . Für jedes $\varepsilon > 0$ gibt es Mengen $\mathcal{T}_n \subset \Sigma^n$ ($n \in \mathbb{N}$) (typische Mengen) mit

$$|\mathcal{T}_n| \leq 2^{(H(X_1)+\varepsilon)n} \quad \text{und} \quad \lim_{n \rightarrow \infty} P((X_1, \dots, X_n) \in \mathcal{T}_n) = 1.$$

Beweis. Wir betrachten als typische Mengen:

$$\mathcal{T}_n = \{(x_1, \dots, x_n) \in \Sigma^n : \frac{1}{n} \sum_{i=1}^n \log_2 f(x_i) \geq -(H(X_1) + \varepsilon)\} \quad (n \in \mathbb{N}).$$

Indem wir den schwachen Grenzwertsatz auf die Zufallsvariablen $Z_n := \log_2 f(X_n)$ ($n \in \mathbb{N}$) anwenden erhalten wir:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \log_2 f(X_i) + H(X_1)\right| \leq \varepsilon\right) = 1.$$

Somit ist

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{n} \sum_{i=1}^n \log_2 f(X_i) \geq -(H(X_1) + \varepsilon)\right) = 1 \quad (4)$$

und es gilt

$$\lim_{n \rightarrow \infty} P((X_1, \dots, X_n) \in \mathcal{T}_n) = 1.$$

Es verbleibt die Größe der typischen Menge \mathcal{T}_n abzuschätzen. Es gilt für ein beliebiges Tupel $(x_1, \dots, x_n) \in \mathcal{T}_n$ ($n \in \mathbb{N}$):

$$\sum_{i=1}^n \log_2 f(x_i) \geq -(H(X_1) + \varepsilon)n$$

und somit

$$\log_2 \left(\prod_{i=1}^n f(x_i) \right) \geq -(H(X_1) + \varepsilon)n.$$

Da $\prod_{i=1}^n f(x_i) = P(X_1 = x_1, \dots, X_n = x_n)$ ist, folgern wir, dass für alle Einträge $(x_1, \dots, x_n) \in \mathcal{T}_n$:

$$P(X_1 = x_1, \dots, X_n = x_n) \geq 2^{-(H(X_1)+\varepsilon)n}.$$

Andererseits gilt

$$\begin{aligned} 1 &\geq \sum_{(x_1, \dots, x_n) \in \mathcal{T}_n} \underbrace{P(X_1 = x_1, \dots, X_n = x_n)}_{\geq 2^{-(H(X_1)+\varepsilon)n}} \\ &\geq |\mathcal{T}_n| 2^{-(H(X_1)+\varepsilon)n} \end{aligned}$$

und damit ist

$$|\mathcal{T}_n| \leq 2^{(H(X_1)+\varepsilon)n}.$$

□

Bemerkung 5.9. Der Quellkodierungssatz sagt, dass typischerweise eine Realisierung von (X_1, \dots, X_n) in dem Codebuch \mathcal{T}_n liegt. Die Menge \mathcal{T}_n enthält weniger als $2^{(H(X_1)+\varepsilon)n}$ Elemente, und es reichen ca. $H(X_1) + \varepsilon$ Bits pro Symbol um das komplette Codebuch zu indexieren. Wir haben in dem Beweis nicht benutzt, dass das Alphabet des Eingangssignals Σ endlich ist. Die Aussage gilt analog für abzählbar unendliche Mengen Σ , falls die Entropie des Eingangssignals $H(X_1)$ endlich ist.

5.4 Das starke Gesetz der Großen Zahlen

Wir definieren einen weiteren Grenzwertbegriff.

Definition 5.10. Eine Folge von Zufallsvariablen (X_n) *konvergiert fast sicher* gegen einen Wert $a \in \mathbb{R}$, falls das Ereignis

$$\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = a\}$$

fast sicher eintritt, d.h. W'keit 1 hat. Wir schreiben in diesem Fall:

$$\lim_{n \rightarrow \infty} X_n = a, \quad \text{fast sicher.}$$

Lemma 5.11. *Es seien X_n ($n \in \mathbb{N}$) Zufallsvariablen. Die folgenden Aussagen sind äquivalent:*

- $(X_n)_{n \in \mathbb{N}}$ konvergiert fast sicher gegen a
- Für alle $\varepsilon > 0$ treten fast sicher nur endlich viele der Ereignisse

$$\{|X_n - a| > \varepsilon\} \quad (n \in \mathbb{N})$$

ein, d.h. für alle $\varepsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{m \geq n} \{|X_m - a| > \varepsilon\}\right) = 0.$$

Die fast sichere Konvergenz impliziert die Konvergenz in Wahrscheinlichkeit:

Lemma 5.12. *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen und $a \in \mathbb{R}$. Es gilt*

$$\begin{aligned} (X_n)_{n \in \mathbb{N}} \text{ konvergiert fast sicher gegen } a \\ \implies (X_n)_{n \in \mathbb{N}} \text{ konvergiert in Wahrscheinlichkeit gegen } a. \end{aligned}$$

Die Umkehrung der Aussage des Lemmas stimmt nicht wie das folgende Beispiel zeigt.

Beispiel 5.13. Wir betrachten folgendes Zufallsexperiment: Es seien X_1, X_2, \dots unabhängige Bernoulli Zufallsvariablen mit Erfolgswahrscheinlichkeit $1/2$ (z.Bsp. Münzwürfe). Nun definieren wir induktiv Zeiten $(t_n)_{n \in \mathbb{N}}$ durch die Vorschriften

- $t_1 = 0$ und
- für $n \geq 2$ ist t_n die größte natürliche Zahl kleiner gleich $t_{n-1} + \log_2 n$,

und betrachten die Ereignisse

$$\begin{aligned} A_n : \text{kein Erfolg in den Versuchen } t_n + 1, \dots, t_{n+1} \\ (\text{d.h. } X_i = 0 \text{ für } t_n + 1 \leq i \leq t_{n+1}) \end{aligned}$$

Wir bezeichnen mit Z_n die Zufallsvariable $Z_n(\omega) = 1_{A_n}(\omega)$ ($n \in \mathbb{N}$). Nun kann man mit Hilfe des Lemmas von Borel-Cantelli zeigen, dass Z_n stochastisch, aber nicht fast sicher, gegen 0 konvergiert.

Satz 5.14. (Starkes Gesetz der großen Zahlen) *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von paarweise unkorrelierten Zufallsvariablen mit identischen Erwartungswerten und*

$$\sup_{n \in \mathbb{N}} \text{var}(X_n) < \infty.$$

Dann gilt für $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ($n \in \mathbb{N}$)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E(X_1), \quad \text{fast sicher.}$$

Beweis. Wir zeigen die Aussage nur für Zufallsvariablen X_n ($n \in \mathbb{N}$) mit Erwartungswert 0. Für andere Erwartungswerte kann man dann die Aussage auf $\tilde{X}_n = X_n - a$ ($n \in \mathbb{N}$) anwenden und erhält direkt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i - a = 0 \quad \text{fast sicher,}$$

was äquivalent zur Aussage des Theorems ist.

Wir setzen $I = \{m^2 : m \in \mathbb{N}\}$ und bezeichnen mit $\pi(n)$ ($n \in \mathbb{N}$) die größte Zahl in I kleiner gleich n . Der Beweis basiert auf der Zerlegung

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^{\pi(n)} X_i + \frac{1}{n} \sum_{i=\pi(n)+1}^n X_i.$$

Es gilt

$$|\bar{X}_n| \leq \left| \frac{1}{\pi(n)} \sum_{i=1}^{\pi(n)} X_i \right| + \left| \frac{1}{\pi(n)} \sum_{i=\pi(n)+1}^n X_i \right|.$$

Um zu zeigen, dass \bar{X}_n gegen 0 konvergiert reicht es zu zeigen, dass für beliebiges $\varepsilon > 0$ beide Terme fast sicher für alle bis auf endlich viele n kleiner als $\varepsilon/2$ sind. Damit ist dann auch die Summe fast sicher für alle bis auf endlich viele $n \in \mathbb{N}$ kleiner als ε .

Wir benutzen die Chebychev Ungleichung um die Wahrscheinlichkeiten abzuschätzen. Für ein $m \in I$ gilt

$$\begin{aligned} P\left(\underbrace{\left| \frac{1}{m} \sum_{i=1}^m X_i \right| \geq \varepsilon/2}_{=: A_m}\right) &\leq \frac{4}{\varepsilon^2} E\left(\left(\frac{1}{m} \sum_{i=1}^m X_i\right)^2\right) = \frac{4}{\varepsilon^2 m^2} E\left(\left(\sum_{i=1}^m X_i\right)\left(\sum_{j=1}^m X_j\right)\right) \\ &= \frac{4}{\varepsilon^2 m^2} E\left(\sum_{i=1}^m \sum_{j=1}^m X_i X_j\right) \end{aligned}$$

und die paarweise Unkorreliertheit der Zufallsvariablen X_n ($n \in \mathbb{N}$) impliziert, dass

$$P(A_m) \leq \frac{4}{\varepsilon^2 m^2} \sum_{i=1}^m \underbrace{E(X_i^2)}_{\leq C} \leq \frac{4C}{\varepsilon^2 m}.$$

Da

$$\sum_{m \in I} P(A_m) \leq \frac{4C}{\varepsilon^2} \sum_{m \in I} \frac{1}{m} = \frac{4C}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

liefert das Lemma von Borel-Cantelli, dass fast sicher nur endlich viele der Ereignisse A_m ($m \in I$) eintreten.

Nun betrachten wir den zweiten Term:

$$P\left(\underbrace{\left|\frac{1}{\pi(n)} \sum_{i=\pi(n)+1}^n X_i\right| \geq \varepsilon/2}_{=: B_n}\right) \leq \frac{4}{\varepsilon^2} E\left(\left(\frac{1}{\pi(n)} \sum_{i=\pi(n)+1}^n X_i\right)^2\right)$$

Daraus folgt, dass

$$\begin{aligned} \sum_{n=1}^{\infty} P(B_n) &\leq \frac{4}{\varepsilon^2} C \sum_{n=1}^{\infty} \frac{1}{\pi(n)^2} (n - \pi(n)) \\ &= \frac{4}{\varepsilon^2} C \sum_{m=1}^{\infty} \sum_{k=0}^{(m+1)^2 - m^2} \frac{1}{m^4} k. \end{aligned}$$

Die zweite Summe läuft über $2m + 2 (\leq 4m)$ Summanden und jeder der Summanden ist kleiner gleich $(2m + 1)/m^4 (\leq 3/m^3)$. Also folgt

$$\sum_{n=1}^{\infty} P(B_n) \leq \frac{48}{\varepsilon^2} C \sum_{m=1}^{\infty} \frac{1}{m^2} < \infty$$

und es treten fast sicher nur endlich viele der Ereignisse B_n ($n \in \mathbb{N}$) ein. \square

Bemerkung 5.15. Durch eine Verfeinerung des Arguments kann man das starke Gesetz der großen Zahlen auch unter den Voraussetzungen von Satz 5.7 beweisen.

Lernziele des 5. Kapitels

- Chebychev Ungleichung
- Schwaches und starkes Gesetz der großen Zahlen
- Shannon's Quellkodierungssatz, Entropie von Zufallsvariablen

6 Markov-Ketten

6.1 Koppelung diskreter Zufallsvariablen

Bisher haben wir häufig angenommen, dass X_1, X_2, \dots (bzw. X_0, X_1, \dots) unabhängige diskrete Zufallsvariablen mit gewissen Verteilungen sind. Wir haben mit diesen Variablen gerechnet, ohne uns zu überlegen wie ein geeigneter Wahrscheinlichkeitsraum, d.h. ein Merkmalraum Ω ein System von messbaren Mengen \mathcal{F} und eine Wahrscheinlichkeitsverteilung P , aussehen müsste (und ob überhaupt ein geeigneter Wahrscheinlichkeitsraum existiert). In diesem Abschnitt wollen wir eine Technik vorstellen mit der wir allgemeine (nicht unabhängige) diskrete Zufallsvariablen X_0, X_1, \dots, X_n stochastisch beschreiben werden.

Zufallsexperiment: Am Anfang des Experiments befinden sich 2 Personen in einem Raum. Die eine Person ist Anhänger der Partei A die andere nicht. Wir lassen nun weitere Leute einzeln in den Raum. Jede Person, die den Raum betritt, sucht sich zufällig eine Person aus und übernimmt deren Meinung. Wie kann man nun die zufälligen Meinungen der nacheinander eintretenden Personen modellieren ?

Um entsprechende Zufallsvariablen einzuführen kann man die sogenannte *Koppelung* zwischen den einzelnen Zufallsvariablen betrachten.

Betrachten wir erst beliebige diskrete Zufallsvariablen X_0, X_1, \dots, X_n . Wir wissen bereits, dass die W'keit, dass der Vektor (X_0, \dots, X_n) den Wert (x_0, \dots, x_n) annimmt, wie folgt mit bedingten Wahrscheinlichkeiten dargestellt werden kann:

$$\begin{aligned} P(X_0 = x_0, \dots, X_n = x_n) \\ &= P(X_0 = x_0) P(X_1 = x_1, \dots, X_n = x_n | X_0 = x_0) \\ &\quad \vdots \\ &= P(X_0 = x_0) P(X_1 = x_1 | X_0 = x_0) P(X_2 = x_2 | X_0 = x_0, X_1 = x_1) \dots \\ &\quad \dots P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \end{aligned} \tag{5}$$

Wir bezeichnen die bedingten W'keiten in (5) mit p_0 bis p_n , d.h. für $x_0 \in \Omega_{X_1}$ ist

$$p_0(x_0) = P(X_0 = x_0),$$

und für $m \geq 1$ und $(x_0, \dots, x_m) \in \Omega_{X_0} \times \dots \times \Omega_{X_m}$ setzen wir

$$p_m(x_0, \dots, x_{m-1}; x_m) = P(X_m = x_m | X_0 = x_0, \dots, X_{m-1} = x_{m-1}).$$

Falls wir hierbei auf ein Ereignis mit W'keit 0 bedingen, spielt es keine Rolle wie wir die entsprechende Funktion p_m definieren. Die Funktion p_0 beschreibt die *Startverteilung* und die Funktionen p_m ($m = 1, \dots, n$) die *Koppelung* der einzelnen Zufallsvariablen X_m mit den Zufallsvariablen X_0, \dots, X_{m-1} . Analog zu Gleichung (5) erhalten wir für alle $m = 0, \dots, n$ und x_0, \dots, x_m :

$$P(X_0 = x_0, \dots, X_m = x_m) = p_0(x_0)p_1(x_0; x_1) \dots p_m(x_0, \dots, x_{m-1}; x_m). \quad (6)$$

Frage: Zu welcher Startverteilung und Koppelung (d.h. zu welchen Funktionen p_1, \dots, p_n) existieren Zufallsvariablen X_0, \dots, X_n (und ein geeigneter W'keitsraum), sodass die Gleichung (6) für alle m und alle x_0, \dots, x_m gilt? Angenommen das Ereignis $\{X_0 = x_0, \dots, X_{m-1} = x_{m-1}\}$ tritt mit strikt positiver W'keit ein, dann definiert $(p_m(x_0, \dots, x_{m-1}; x_m))_{x_m \in \Omega_{X_m}}$ eine Folge von W'keitsgewichten. Es gilt also

- (i) $\forall x_m \in \Omega_{X_m}: p_m(x_0, \dots, x_{m-1}; x_m) \geq 0$ und
- (ii) $\sum_{x_m \in \Omega_{X_m}} p_m(x_0, \dots, x_{m-1}; x_m) = 1.$

Nun kann man zeigen, dass diese beiden Bedingungen auch hinreichend für die Existenz geeigneter Zufallsvariablen sind:

Satz 6.1. *Es bezeichne I die Indexmenge $I = \{0, \dots, n\}$ oder $I = \mathbb{N}_0$. Weiterhin seien Ω_m ($m \in I$) beliebige abzählbare Mengen und p_m ($m \in I$) eine Familie von Funktionen*

$$p_m : \prod_{i=0}^{m-1} \Omega_i \times \Omega_m \rightarrow [0, 1]$$

mit der Eigenschaft, dass $\forall m \in I$ und $\forall (x_1, \dots, x_{m-1}) \in \Omega_1 \times \dots \times \Omega_{m-1}$

$$\sum_{x_m \in \Omega_m} p_m(x_1, \dots, x_{m-1}; x_m) = 1.$$

Dann existieren Zufallsvariablen X_m ($m \in I$) (auf einem geeigneten W'keitsraum), für die die Gleichung (6) für alle $m \in I$ und alle x_0, \dots, x_m gilt. Insbesondere gilt

$$P(X_m = x_m | X_0 = x_0, \dots, X_{m-1} = x_{m-1}) = p_m(x_1, \dots, x_{m-1}; x_m)$$

immer wenn die bedingte W'keit wohldefiniert ist.

Bemerkung 6.2. Das Theorem besagt, dass wir, falls die Startverteilung und die Koppelungen die Eigenschaften (i) und (ii) erfüllen, ohne weiteres annehmen können, dass entsprechende Zufallsvariablen existieren, die gerade diese bedingten W'keiten besitzen.

Wir kommen zurück zum Zufallsexperiment vom Anfang dieses Abschnitts.

Lösung: Wir wollen die Meinung der $m + 1$ -ten Person durch eine Zufallsvariable X_m mit Werten in $\Omega_m = \{0, 1\}$ beschreiben, wobei $\{X_m = 1\}$ das Ereignis bezeichnet, dass die $m + 1$ -te eintretende Person ein Anhänger der Partei wird. Wir können das Experiment wie folgt beschreiben: Für $m \in \mathbb{N}_0$ und $x_0, \dots, x_{m-1} \in \{0, 1\}$ setzen wir

$$p_m(x_0, \dots, x_{m-1}; 1) = \frac{\sum_{i=0}^{m-1} x_i + 1}{m + 2} = \text{rel. Anzahl der Sympatisanten}$$

und

$$p_m(x_0, \dots, x_{m-1}; 0) = 1 - p_m(x_0, \dots, x_{m-1}; 1).$$

Natürlich erfüllt p_m alle Voraussetzungen von Satz 6.1. Also existieren Zufallsvariablen X_1, X_2, \dots , die dieses Experiment beschreiben.

6.2 Homogene Markov-Ketten

Problem: (*Callcenter*) Ein Callcenter ist über 3 Leitungen erreichbar. Wir möchten die Anzahl der belegten Leitungen zu jedem Zeitpunkt $n \in \mathbb{N}_0$ durch eine $\{0, 1, 2, 3\}$ -wertige Zufallsvariable modellieren.

In diesem Fall beschreiben die Zufallsvariablen X_n ($n \in \mathbb{N}_0$) einen zufälligen zeitlich ablaufenden Prozess.

Definition 6.3. Es seien

- T eine abzählbare Indexmenge (meist $T = \mathbb{N}_0 = \{0, 1, 2, \dots\}$) (*Zeitbereich*),

- S eine weitere abzählbare Menge (*Zustandsraum*),
- $X_t : \Omega \rightarrow S$ ($t \in T$) Zufallsvariablen (*Zustand zum Zeitpunkt t*).

Dann nennen wir $(X_t)_{t \in T}$ einen (*diskreten*) *stochastischen Prozess* mit Zustandsraum S . Ein endliches oder unendliches Tupel mit Einträgen im Zustandsraum S nennen wir *Pfad*.

Im Beispiel des Callcenters ist der *Zustandsraum* S gerade $\{0, 1, 2, 3\}$. Um die Auslastung mit Zufallsvariablen zu modellieren, müssen wir nun noch die Verteilung der Anfangsauslastung

$$p_0(x_0) = P(X_0 = x_0)$$

und die Koppelungen

$$p_m(x_0, \dots, x_{m-1}; x_m) = P(X_m = x_m | X_1 = x_1, \dots, X_{m-1} = x_{m-1})$$

beschreiben. Häufig hängt der Wert der m -ten Koppelung nicht von dem kompletten vorhergehenden Pfad (x_0, \dots, x_{m-1}) sondern nur vom letzten Zustand x_{m-1} ab.

Definition 6.4. Wir nennen einen stochastischen Prozess $(X_n)_{n \in \mathbb{N}_0}$ eine *Markov-Kette*, falls für alle endlichen Pfade (x_0, \dots, x_n) mit $P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) > 0$ die Gleichung

$$P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

gilt. Die Verteilung einer Markov-Kette wird also durch

- einen Zustandsraum S ,
- die Startverteilung $p_0 : S \rightarrow [0, 1]$ (die wir im Folgenden mit μ bezeichnen werden) und
- die Koppelungen $p_n : S \times S \rightarrow [0, 1]$, $(x_{n-1}, x_n) \mapsto p_n(x_{n-1}, x_n)$ ($n \in \mathbb{N}$)

beschrieben.

Falls die Koppelung für jeden Zeitpunkt $n \in \mathbb{N}$ durch dieselbe Funktion $p : S \times S \rightarrow [0, 1]$ beschrieben werden kann (die Koppelung also nicht von der Zeit abhängt), nennen wir $(X_n)_{n \in \mathbb{N}_0}$ eine *homogene Markov-Kette*. Diese wird beschrieben durch

- einen Zustandsraum S ,
- die Startverteilung $\mu : S \rightarrow [0, 1]$ und
- die Koppelung $p : S \times S \rightarrow [0, 1]$, $(x, y) \mapsto p(x, y)$.

Wir werden im Folgenden ausschließlich homogene Markov-Ketten betrachten, wobei wir die Notationen $(X_n)_{n \in \mathbb{N}_0}$ und p für die Markov-Kette und die Koppelung beibehalten werden. Häufig möchte man das Verhalten einer Markov-Kette für unterschiedliche Startverteilungen μ untersuchen. Hierzu bezeichnen wir mit P^μ ein W'keitsmaß unter dem der Prozess $(X_n)_{n \in \mathbb{N}_0}$ gerade eine homogene Markov-Kette mit Startverteilung μ ist. Falls wir in einem deterministischen Zustand $x \in S$ starten (d.h. $\mu(y) = 1_{\{x\}}(y)$ ($y \in S$)), schreiben wir auch P^x . Analog schreiben wir für die entsprechenden Erwartungswerte E^μ und E^x .

Beispiel 6.5. (*Servicecenter*)

Wir modellieren die Auslastung eines Servicecenters durch eine homogene Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit Zustandsraum $S = \mathbb{N}_0$. Wir wählen als Koppelung gerade

$$p(x, y) = \begin{cases} 0,4 & \text{falls } y = x + 1 \\ 0,6 & \text{falls } y = x - 1 \text{ oder } y = x = 0 \\ 0 & \text{sonst.} \end{cases}$$

Frage: Wie groß ist die W'keit, dass sich zum Zeitpunkt 2 genau 1 Kunde im Servicecenter befindet, falls das Servicecenter zum Zeitpunkt 0 leer ist (also $\mu(x) = 1_{\{0\}}(x)$ ist) ?

$$P^0(X_2 = 1) = P^0(X_1 = 0, X_2 = 1) = \mu(0) p(0, 0) p(0, 1) = 0,6 \cdot 0,4.$$

In dem Beispiel gibt es nur einen Pfad der Länge 2 von 0 nach 1 entlang dem die Markov-Kette mit positiver W'keit läuft. Wir führen folgende Notation für homogene Markov-Ketten ein.

Definition 6.6. Es bezeichne (x_0, \dots, x_n) einen endlichen Pfad. Wir nennen (x_0, \dots, x_n) einen *guten Pfad* von x_0 nach x_n , falls für alle $i = 1, \dots, n$

$$p(x_{i-1}, x_i) > 0$$

ist.

Allgemein kann man die Aufenthaltswahrscheinlichkeit für einen Zeitpunkt mit Hilfe der folgenden Regel induktiv herleiten:

Satz 6.7. Für $n \in \mathbb{N}$ und $x \in S$ gilt

$$P(X_n = x) = \sum_{y \in S} P(X_{n-1} = y) p(y, x).$$

Definition 6.8. Wir sagen eine Markov-Kette ist im *Gleichgewicht*, falls für beliebige $x \in S$ die W'keiten $P(X_n = x)$ nicht von der Zeit $n \in \mathbb{N}_0$ abhängen.

Satz 6.9. Eine Markov Kette ist genau dann im Gleichgewicht, falls die Startverteilung μ eine Lösung der Gleichung

$$\mu(x) = \sum_{y \in S} \mu(y) p(y, x) \quad (x \in S)$$

ist.

Definition 6.10. Eine Abbildung $\pi : S \rightarrow \mathbb{R}$, die die *Gleichgewichtsgleichung*

$$\pi(x) = \sum_{y \in S} \pi(y) p(y, x) \quad (x \in S) \tag{7}$$

löst, heißt *invariant* unter der Koppelung p . Falls $(\pi(x))_{x \in S}$ zusätzlich eine Folge von W'keitsgewichten ist, nennen wir π eine *invariante Verteilung der Markov-Kette*.

Beispiel 6.11. (*Invariante Verteilung: Servicecenter*)

Angenommen $(\pi(s))_{s \in \mathbb{N}_0}$ löst die Gleichung (7) für die Übergangswahrscheinlichkeiten p wie in Beispiel 6.5. Dann liefert die Gleichung (7) für $x = 0$

$$\pi(0) = \frac{3}{5}(\pi(1) + \pi(0)),$$

sodass

$$\pi(1) = \frac{2}{3}\pi(0).$$

Nun betrachten wir die Gleichgewichtsgleichung für $x = 1$:

$$\frac{2}{3}\pi(0) = \pi(1) = \frac{3}{5}\pi(2) + \frac{2}{5}\pi(0)$$

Damit folgt

$$\pi(2) = \frac{4}{9}\pi(0).$$

Man kann nun analog $\pi(3), \pi(4), \dots$ berechnen. Wenn man sich die entsprechenden Werte betrachtet, kommt man schnell zu der Vermutung, dass für alle $n \in \mathbb{N}_0$

$$\pi(n) = \left(\frac{2}{3}\right)^n \pi(0)$$

gilt. Die Gleichung lässt sich leicht mittel vollständiger Induktion für alle $n \in \mathbb{N}_0$ beweisen. Damit $(\pi(n))_{n \in \mathbb{N}_0}$ eine Folge von Wahrscheinlichkeitsgewichten definiert muss noch zusätzlich gelten, dass

$$1 = \sum_{n=1}^{\infty} \pi(n) = \pi(0) \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = 3 \pi(0)$$

also $\pi(0) = 1/3$ ist.

Frage: Angenommen wir starten die Markov-Kette deterministisch in einem Zustand $x_0 \in S$ und warten eine lange Zeit. Wie verhält sich die Aufenthaltswahrscheinlichkeit $P^{x_0}(X_n = x)$ für einen Zustand x für große n ? Hat der Startwert einen Einfluss auf diese W'keit?

Zur Beantwortung der Frage benötigen wir noch zwei weitere Definitionen.

Definition 6.12. Wir nennen eine Markov-Kette

- *irreduzibel*, falls es für alle Zustände x und y einen guten Pfad von x nach y gibt.
- *aperiodisch*, falls der größte gemeinsame Teiler der Menge

$$J = \{n \in \mathbb{N} : \exists \text{ einen guten Pfad } (x_0, \dots, x_n) \text{ der Länge } n \text{ mit } x_0 = x_n\}$$

gleich 1 ist.

Satz 6.13. *Es sei*

- $(X_n)_{n \in \mathbb{N}}$ *eine homogene irreduzible aperiodische Markov-Kette und*
- $(\pi(x))_{x \in S}$ *eine invariante Verteilung.*

Dann gilt für eine beliebige Startverteilung μ (oder auch einen beliebigen deterministischen Startwert $x_0 \in S$)

$$\lim_{n \rightarrow \infty} P(X_n = x) = \pi(x) \quad (x \in S).$$

Insbesondere existiert maximal eine invariante Verteilung.

Bemerkung 6.14. Der Satz setzt voraus, dass zumindest eine invariante Verteilung existiert. Es kann aber durchaus passieren, dass die Markov-Kette keine invariante Verteilung besitzt. In diesem Fall gilt

$$\lim_{n \rightarrow \infty} P(X_n = x) = 0 \quad (x \in S).$$

Beispiel 6.15. Für das Beispiel des Servicecenters besagt der Satz, dass unabhängig davon wie wir das System starten (wie wir μ wählen), die Gleichung

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{1}{3} \left(\frac{2}{3}\right)^k \quad (k \in \mathbb{N}_0)$$

gilt.

6.3 Markov-Ketten mit endlichem Zustandsraum

Wir betrachten nun eine allgemeine homogene Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit endlichem Zustandsraum $S = \{s_1, \dots, s_m\}$. Es bezeichne wieder

- μ die Startverteilung und
- $p : S \times S \rightarrow [0, 1]$ die Koppelung

von $(X_n)_{n \in \mathbb{N}_0}$.

Um mit der Markov-Kette besser rechnen zu können führen wir die folgende Vektor-Matrix-Schreibweise ein.

Notation 6.16. Wir assoziieren eine beliebige Abbildung $\pi : S \rightarrow \mathbb{R}$ mit dem Vektor

$$\bar{\pi} = (\pi(s_1), \dots, \pi(s_m)).$$

Weiterhin setzen wir

$$p_{i,j} := p(s_i, s_j) \quad (i, j = 1, \dots, m)$$

und bezeichnen mit \mathbf{P} die $m \times m$ -Matrix (*Übergangsmatrix*) mit Einträgen p_{ij} , d.h.

$$\mathbf{P} = (p_{i,j})_{i,j \in \{1, \dots, m\}} = \begin{pmatrix} p_{1,1} & \dots & p_{1,m} \\ \vdots & & \vdots \\ p_{m,1} & \dots & p_{m,m} \end{pmatrix}.$$

Nun ist die Vektorschreibweise für die Wahrscheinlichkeitsgewichte von X_n gerade

$$\bar{p}^{(n)} := (P(X_n = s_1), \dots, P(X_n = s_m)).$$

Mit Hilfe der neuen Notationen können wir die rekursive Formel aus Theorem 6.7 mittels einer Matrixmultiplikation darstellen:

Satz 6.17. *Es gilt für $n \in \mathbb{N}$*

$$\bar{p}^{(n)} = \bar{p}^{(n-1)} \mathbf{P}.$$

Die Gleichung impliziert direkt, dass

$$\bar{p}^{(n)} = \bar{\mu} \mathbf{P}^n,$$

und für einen deterministischen Startwert s_i erhält man

$$P^{s_i}(X_n = s_j) = (\mathbf{P}^n)_{i,j}.$$

Satz 6.18. *Es sei $\pi : S \rightarrow \mathbb{R}$. Dann ist π genau dann invariant unter p , falls $\bar{\pi}$ die Gleichung*

$$\bar{\pi} \mathbf{P} = \bar{\pi} \tag{8}$$

löst. Die Gleichung (8) ist äquivalent zu

$$\bar{\pi}(\mathbf{P} - \mathbb{I}_m) = 0,$$

wobei \mathbb{I}_m die m -dimensionale Einheitsmatrix bezeichnet.

Im Fall eines unendlichen Zustandsraums ist es möglich, dass keine invariante Verteilung existiert. Hingegen gilt für Markov-Ketten mit endlichem Zustandsraum:

Satz 6.19. *Eine Markov-Kette mit endlichem Zustandsraum besitzt mindestens eine invariante Verteilung.*

Bemerkung 6.20. Zur Berechnung der invarianten Verteilung kann man wie folgt vorgehen:

- (i) Zuerst berechnet man den Kern von $(\mathbf{P} - \mathbb{I}_m)^*$. Ein Vektor $\bar{\pi}$ ist genau dann invariant unter der Koppelung p , falls $\bar{\pi}^*$ im Kern von $(\mathbf{P} - \mathbb{I}_m)^*$ liegt.
- (ii) Berechnung eines invarianten Vektors $\bar{\pi}$ mit

$$\sum_{i=1}^m \bar{\pi}_i = 1 \quad \text{und} \quad \bar{\pi}_i \geq 0 \quad (i = 1, \dots, m).$$

Man kann überprüfen, dass die Markov-Kette sehr schnell gegen die invariante Verteilung konvergiert und somit sehr schnell den Startwert “vergisst”. Eine Schlussfolgerung aus dieser Eigenschaft ist das folgende Theorem:

Satz 6.21. *Es bezeichne $(X_n)_{n \in \mathbb{N}_0}$ eine homogene, irreduzible, aperiodische Markov-Kette mit endlichem Zustandsraum und invarianter Verteilung $(\pi(x))_{x \in S}$. Für eine beliebige Funktion $f : S \rightarrow \mathbb{R}$ und eine beliebige Startverteilung μ gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = E^\pi(f(X_0)) = \sum_{x \in S} \pi(x) f(x) \quad \text{fast sicher.}$$

6.4 Reversible Markov-Ketten

Eine besondere Klasse von Markov-Ketten stellen die reversiblen Markov-Ketten dar. Wir werden sie im Folgenden einführen, wobei wir weiterhin annehmen, dass der Zustandsraum $S = \{s_1, \dots, s_m\}$ endlich ist.

Es bezeichne $(C_{i,j})_{i,j \in \{1, \dots, m\}}$ eine symmetrische Matrix mit nicht-negativen Einträgen. Nun betrachten wir eine Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit Übergangsmatrix $\mathbf{P} = (p_{i,j})_{i,j \in \{1, \dots, m\}}$ gegeben durch

$$p_{i,j} := \frac{C_{i,j}}{C_j} \quad (i, j \in \{1, \dots, m\}),$$

wobei $C_i := \sum_{j=1}^m C_{i,j}$.

Für diese Markov-Kette lässt sich leicht eine invariante Verteilung angeben.

Satz 6.22. *Die Gewichte*

$$\pi(s_i) = \frac{C_i}{\sum_{j=1}^m C_j} \quad (i = 1, \dots, m)$$

definieren eine invariante Verteilung für die Markov-Kette.

Eine Markov-Kette, die wie oben definiert wurde, heißt reversible Markov-Kette. Sie verdankt diesen Namen der folgenden Eigenschaft.

Satz 6.23. *Für eine invariante Verteilung π (es kann durchaus eine andere als die im letzten Theorem hergeleitete sein), gilt für alle $n \in \mathbb{N}$ und $x_0, \dots, x_n \in S$*

$$P^\pi(X_0 = x_0, \dots, X_n = x_n) = P^\pi(X_0 = x_n, \dots, X_n = x_0).$$

Dies bedeutet, dass wir die Zufallsvektoren (X_0, \dots, X_n) und (X_n, \dots, X_0) statistisch nicht unterscheiden können.

6.5 Austrittszeiten

Wir betrachten weiterhin eine allgemeine homogene Markov-Kette mit endlichem Zustandsraum $S = \{s_1, \dots, s_m\}$ und starten die Markov-Kette deterministisch in dem Zustand $x = s_i$.

Fragestellung: Was ist die erwartete Wartezeit bis der erste Zustand in $A \subset S$ eintritt? Alternativ kann man auch fragen wie häufig ein Zustand $y = s_j \in A^c$ im Mittel besucht wird, bevor ein Zustand in A eintritt.

Modellierung: Wir bezeichnen mit T die zufällige Zeit zu der die Markov-Kette, das erste Mal einen Wert in A annimmt, d.h.

$$T = \min\{n \in \mathbb{N}_0 : X_n \in A\}.$$

Weiterhin betrachten wir die Zufallsvariable N_y , die die Häufigkeit der Besuche in y vor der Zeit T zählt:

$$N_y = \sum_{n=0}^{T-1} 1_{\{X_n=y\}}.$$

Lösung des Problems: Um das Problem zu lösen betrachten wir einen leicht modifizierten Prozess (\tilde{X}_n) . Der Prozess (\tilde{X}_n) folgt (X_n) bis zur zufälligen

Zeit T und bleibt dann auf dem ersten Zustand in A stehen:

$$\tilde{X}_n = \begin{cases} X_n & \text{falls } n \leq T \\ X_T & \text{falls } n \geq T. \end{cases}$$

Der so erhaltene Prozess ist wieder eine Markov-Kette. Die entsprechende Koppelung ist gerade gegeben durch:

$$\tilde{p}(z_1, z_2) = \begin{cases} p(z_1, z_2) & \text{falls } z_1 \notin A \\ 1_{\{z_1=z_2\}} & \text{falls } z_1 \in A. \end{cases}$$

Gesucht: $E^x(N_y)$!

Es gilt

$$E^x(N_y) = E^x\left(\sum_{n=0}^{T-1} 1_{\{X_n=y\}}\right) = E^x\left(\sum_{n=0}^{\infty} 1_{\{\tilde{X}_n=y\}}\right) = \sum_{n=0}^{\infty} P^x(\tilde{X}_n = y).$$

Da $x = s_i$ und $y = s_j$ erhalten wir, dass

$$P^x(\tilde{X}_n = y) = (\tilde{\mathbf{P}}^n)_{i,j}.$$

Also finden wir $E^x(N_y)$ indem wir

$$\sum_{n=0}^{\infty} (\tilde{\mathbf{P}}^n)_{i,j}$$

berechnen. Wir nehmen nun an, dass die Menge A gerade durch $A = \{s_{k+1}, \dots, s_m\}$ gegeben ist. (Möglicherweise müssen wir hierzu die Zustände neu durchnummerieren.) Dann ist

$$\tilde{\mathbf{P}} = \begin{pmatrix} p_{1,1} & \dots & p_{1,k} & p_{1,(k+1)} & \dots & p_{1,m} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{k,1} & \dots & p_{k,k} & p_{k,(k+1)} & \dots & p_{k,m} \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & \vdots & & \ddots & \\ 0 & \dots & 0 & 0 & & 1 \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{P}} & \mathbf{Q} \\ 0 & \mathbb{I}_{m-k} \end{pmatrix}$$

Multiplikation von $\tilde{\mathbf{P}}$ mit sich selbst liefert nun

$$\tilde{\mathbf{P}}^2 = \begin{pmatrix} \bar{\mathbf{P}}^2 & \bar{\mathbf{P}}\mathbf{Q} + \mathbf{Q} \\ 0 & \mathbb{I}_{m-k} \end{pmatrix}.$$

Allgemein erhält man

$$\tilde{\mathbf{P}}^n = \begin{pmatrix} \bar{\mathbf{P}}^n & * \\ 0 & \mathbb{I}_{m-k} \end{pmatrix},$$

wobei an der Stelle $*$ eine für uns uninteressante Matrix steht. Es gilt also

$$(\tilde{\mathbf{P}}^n)_{i,j} = (\bar{\mathbf{P}}^n)_{i,j}.$$

Wenn wir annehmen, dass von jedem Startpunkt ein guter Pfad nach A existiert, konvergiert die Reihe

$$\sum_{n=0}^{\infty} \bar{\mathbf{P}}^n = (\mathbb{I}_k - \bar{\mathbf{P}})^{-1}.$$

Insbesondere gilt

$$E^x(N_y) = (\mathbb{I}_k - \bar{\mathbf{P}})_{i,j}^{-1}.$$

Wir fassen unsere Ergebnisse im folgenden Theorem zusammen.

Satz 6.24. *Es seien $A = \{s_{k+1}, \dots, s_m\}$ und $x = s_i$ und $y = s_j$ zwei Zustände in A^c . Angenommen es existiert von jedem Zustand in A^c ein guter Pfad nach A . Dann gilt für die Anzahl der Besuche N_y von y vor dem Erreichen von A :*

$$E^x(N_y) = (\mathbb{I}_k - \bar{\mathbf{P}})_{i,j}^{-1},$$

wobei

$$\bar{\mathbf{P}} = (p_{l,m})_{l,m \in \{1, \dots, k\}}.$$

Aus dem Satz können wir nun die Antwort auf die ursprüngliche Fragestellung ableiten. Wir behalten die obigen Notationen bei.

Korollar 6.25. *Für eine in $x = s_i \in A^c$ gestartete Markov-Kette gilt*

$$E^x(T) = \sum_{j=1}^k E^x(N_{s_j}) = \sum_{j=1}^k (\mathbb{I}_k - \bar{\mathbf{P}})_{i,j}^{-1}.$$

Beweis. Es gilt

$$T = \sum_{n=0}^{\infty} 1_{A^c}(\tilde{X}_n) = \sum_{n=0}^{\infty} \sum_{j=1}^k 1_{\{\tilde{X}_n = s_j\}} = \sum_{j=1}^k N_{s_j},$$

wobei wir im letzten Schritt die Summationsreihenfolge vertauscht haben. Nun folgt das Resultat direkt mit Hilfe von Satz 6.24. \square

Frage: Wie groß ist die W'keit $P^x(X_T = y)$ für zwei Elemente $y \in A$ und $x \in A^c$?

Satz 6.26. Für $x = s_i \in \{1, \dots, s_k\} = A^c$ und $y = s_j \in A$ gilt

$$P^x(X_T = y) = \sum_{l=1}^k (\mathbb{I} - \bar{\mathbf{P}})_{i,l}^{-1} p_{l,j}.$$

Beweis. Man erhält direkt:

$$\begin{aligned} P^x(X_T = y) &= \sum_{n=1}^{\infty} P^x(T = n, \tilde{X}_n = y) = \sum_{n=1}^{\infty} P^x(\tilde{X}_{n-1} \in A^c, \tilde{X}_n = y) \\ &= \sum_{n=1}^{\infty} \sum_{l=1}^k P^x(\tilde{X}_{n-1} = s_l, \tilde{X}_n = y) \\ &= \sum_{l=1}^k \sum_{n=1}^{\infty} \underbrace{P^x(\tilde{X}_{n-1} = s_l)}_{=(\tilde{\mathbf{P}}^{n-1})_{i,l}} p_{l,j} = \sum_{l=1}^k (\mathbb{I} - \bar{\mathbf{P}})_{i,l}^{-1} p_{l,j}. \end{aligned}$$

\square

Lernziele des 6. Kapitels

- Modellierung stochastischer Prozesse (Abläufe) in diskreter Zeit
- Homogene Markov-Ketten
- Invariante Verteilung
- Langzeitverhalten von Markov-Ketten
- Reversible Markov-Ketten
- Ausrittszeiten

7 Allgemeine Zufallsvariablen

Experiment: Wir lassen eine Stecknadel auf eine flache Oberfläche fallen und bezeichnen mit X den zufälligen Winkel zwischen der Nadel und der Nord-Süd Richtung. Dieser Winkel ist ein zufälliger Wert zwischen 0 und 2π ohne Präferenz für bestimmte Werte. Zur Modellierung benötigen wir eine *Gleichverteilung* auf dem Intervall $[0, 2\pi)$. Da das Intervall $[0, 2\pi)$ nicht abzählbar ist, lässt sich das Experiment nicht mit einer diskreten Zufallsvariable modellieren.

Problem: Wir benötigen ein Konzept um allgemeine reelle Zufallsgrößen zu modellieren.

Eine reelle Zufallsvariable soll wieder als Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

modelliert werden. Weiterhin behalten wir die Mengenschreibweisen $\{X \in A\}$, $\{X \leq x\}$ und $\{X = x\}$ bei.

Achtung: Beim Rechnen mit allgemeinen Zufallsvariablen ist es typischerweise nicht möglich jeder Menge $A \subset \mathbb{R}$ eine Wahrscheinlichkeit $P(X \in A)$ zuzuordnen, da nicht alle Mengen $\{X \in A\}$ Ereignisse sind.

Wir müssen uns auf eine kleinere Familie von *messbaren* Mengen beschränken und eine weitere technische Forderung an die Abbildung X stellen.

Definition 7.1. (i) Wir bezeichnen mit $\mathcal{B} \subset \mathfrak{P}(\mathbb{R})$ die kleinste σ -Algebra (siehe Definition 1.10), die alle Mengen $(-\infty, t]$ ($t \in \mathbb{R}$) enthält.

(ii) Wir nennen eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ *reelle Zufallsvariable*, falls die Menge $\{X \in A\}$ für alle $A \in \mathcal{B}$ ein Ereignis ist, d.h. falls für jede Menge $A \in \mathcal{B}$ die W'keit $P(X \in A)$ wohldefiniert ist.

Alle von uns aus natürlichen Fragestellungen abgeleiteten Zufallsvariablen können in diesem Sinne als Zufallsvariablen über einem geeigneten (für uns uninteressanten) W'keitsraum modelliert werden. Die Einschränkung, dass nun nicht mehr allen Mengen $A \subset \mathbb{R}$ eine W'keit zugeordnet werden kann, bedeutet für uns keine Einschränkung, da die Familie \mathcal{B} sehr reichhaltig ist: Die Familie \mathcal{B} enthält z. Bsp.

- alle Intervalle,

- alle offenen Mengen,
- alle abgeschlossenen Mengen und auch
- alle Mengen, die sich mithilfe abzählbar vieler mengentheoretischer Verknüpfungen konstruieren lassen.

Für all diese Mengen sind die Wahrscheinlichkeiten $P(X \in A)$ wohldefiniert.

Wir haben in Definition 7.1 Anforderungen an eine Modellierung eines reellen zufälligen Werts gestellt. Um den Ausgang eines Zufallsexperiments nun durch eine reelle Zufallsvariable beschreiben zu können, benötigen wir Kriterien, die die Existenz geeigneter Zufallsvariablen sicherstellen. Weiterhin müssen wir klären wie wir die Statistik (also die Verteilung) einer Zufallsvariable eindeutig beschreiben können. Wir gehen hierzu wie folgt vor: Zuerst werden Eigenschaften von reellen Zufallsvariablen hergeleitet. Danach geben wir Voraussetzungen an, die die Existenz geeigneter Zufallsvariablen sicherstellen.

Definition 7.2. Die Funktion

$$F(t) := F_X(t) := P(X \leq t) \quad (t \in \mathbb{R})$$

heißt *Verteilungsfunktion der (reellen) Zufallsvariable X* . Aus der Definition folgt direkt, dass

$$P(X \in (s, t]) = F(t) - F(s) \quad (t \geq s).$$

Wir nennen zwei Zufallsvariablen X und Y *identisch verteilt*, falls sie die gleiche Verteilungsfunktion besitzen.

Für die uns bereits bekannten diskreten Zufallsvariablen kann man die Verteilungsfunktion mit Hilfe der Wahrscheinlichkeitsgewichte darstellen:

Lemma 7.3. Für eine reelle diskrete Zufallsvariable X gilt

$$F_X(t) = P(X \leq t) = \sum_{\substack{x \in \Omega_X: \\ x \leq t}} f_X(x) \quad (t \in \mathbb{R}).$$

Beispiel 7.4. Sei X eine Bernoulli Zufallsvariable mit Erfolgswahrscheinlichkeit $p \in [0, 1]$. Dann ist

$$F_X(t) = \begin{cases} 0 & \text{falls } t < 0, \\ 1 - p & \text{falls } t \in [0, 1), \\ 1 & \text{falls } t \geq 1. \end{cases}$$

Alle Verteilungsfunktionen besitzen die folgenden Eigenschaften:

Lemma 7.5. (i) $F : \mathbb{R} \rightarrow \mathbb{R}$ ist monoton wachsend,

(ii) $\lim_{t \rightarrow -\infty} F(t) = 0$ und $\lim_{t \rightarrow \infty} F(t) = 1$,

(iii) F ist rechtsseitig stetig, d.h. $\forall t \in \mathbb{R} : \lim_{\varepsilon \downarrow 0} F(t + \varepsilon) - F(t) = 0$.

Das Lemma fasst 3 Eigenschaften von Verteilungsfunktionen zusammen. Andererseits kann man zeigen, dass diese drei Eigenschaften die Existenz einer Zufallsvariablen mit Verteilungsfunktion F sicherstellen:

Satz 7.6. Es sei F eine Funktion die den Eigenschaften (i), (ii) und (iii) aus Lemma 7.5 genügt. Dann existiert eine Zufallsvariable X (und ein geeigneter W'keitsraum), sodass F gerade die Verteilungsfunktion von X ist.

7.1 Stetige Zufallsvariablen

Eine wichtige Klasse von reellen Zufallsvariablen bilden die stetigen Zufallsvariablen.

Definition 7.7. Wir nennen X eine reelle stetige Zufallsvariable, falls es eine nicht-negative Riemann integrierbare Funktion $f = f_X$ gibt, sodass

$$F_X(t) = \int_{-\infty}^t f(x) dx.$$

Dann ist

$$P(X \in (a, b]) = \int_a^b f(x) dx.$$

Wir nennen f (Wahrscheinlichkeits-)Dichte von X .

Bemerkung 7.8. Es sei X eine stetige Zufallsvariable. Wir wollen für ein $b \in \mathbb{R}$ die W'keit $P(X = b)$ berechnen. Da $\{b\} = \bigcap_{n=1}^{\infty} (b - \frac{1}{n}, b]$ ist, liefert Satz 1.9, dass

$$P(X = b) = P\left(X \in \bigcap_{n=1}^{\infty} (b - \frac{1}{n}, b]\right) = \lim_{n \rightarrow \infty} \int_{b - \frac{1}{n}}^b f(x) dx = 0.$$

Also ist die W'keit, dass eine stetige Zufallsvariable X einen bestimmten vorgegebene Wert b annimmt, gleich 0.

Konstruktion stetiger Zufallsvariablen

Damit wir für eine gegebene nicht-negative Riemann-integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ eine Zufallsvariable X mit Dichte f betrachten können, muss wegen Satz 7.6 die Funktion

$$F(t) = \int_{-\infty}^t f(x) dx$$

die Eigenschaften (i), (ii) und (iii) aus Lemma 7.5 erfüllen. Eigenschaften (i) und (iii) und der erste Teil von (ii) sind trivialerweise erfüllt. Also muss man nur noch überprüfen, ob

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Wichtige stetige Zufallsvariablen

Gleichverteilte Zufallsvariable auf dem Intervall $[a, b]$ ($a < b$)

Eine stetige Zufallsvariable X mit Dichte

$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}(x) \quad (x \in \mathbb{R}).$$

heißt gleichverteilte Zufallsvariable auf $[a, b]$. Sie besitzt die Verteilungsfunktion

$$F_X(t) = \begin{cases} 0 & \text{falls } t \leq a, \\ \frac{t-a}{b-a} & \text{falls } t \in [a, b], \\ 1 & \text{falls } t \geq b. \end{cases}$$

Experiment: Eine gleichverteilte Zufallsvariable auf $[0, 2\pi]$ kann zum Beispiel zur Modellierung des Stecknadel-experiments vom Anfang des Kapitels

verwendet werden.

Exponentialverteilung zum Parameter $\lambda > 0$

Eine Zufallsvariable X mit Dichte

$$f_X(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x > 0 \\ 0 & \text{falls } x \leq 0 \end{cases}$$

heißt exponentialverteilte Zufallsvariable zum Parameter λ . Ihre Verteilungsfunktion ist

$$\begin{aligned} F_X(t) &= 1_{(0,\infty)}(t) \int_0^t \lambda e^{-\lambda x} dx = 1_{(0,\infty)}(t) (1 - e^{-\lambda t}) \\ &= \begin{cases} 1 - e^{-\lambda t} & \text{falls } t \geq 0 \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

Für eine exponentialverteilte Zufallsvariable X gilt für $t > s \geq 0$ und $a > 0$

$$P(X \in [s, t]) = P(X \in [a + s, a + t] | X \geq a).$$

Wir sagen die Exponentialverteilung ist *gedächtnislos*. Umgekehrt ist jede gedächtnislose Zufallsvariable mit Werten in $[0, \infty)$ exponentialverteilt.

Experiment: Angenommen die W'keit, dass ein elektrisches Bauteil in einem zukünftigen Zeitintervall kaputt geht, hängt nicht von der bisherigen Lebensdauer ab. In diesem Fall ist die zufällige Zeit an der das Bauteil zerstört wird, exponentialverteilt. In der Informatik werden häufig die Zeitintervalle zwischen aufeinanderfolgenden Ereignissen (zum Beispiel die Zeitspanne zwischen zwei aufeinanderfolgenden Anfragen an einen Server) exponentialverteilt gewählt.

Normalverteilung $\mathcal{N}(a, \sigma^2)$

Die Normalverteilung wird durch zwei Parameter beschrieben: den *Erwartungswert* a und die *Varianz* $\sigma^2 > 0$. (Wir werden die Begriffe Erwartungswert und Varianz später noch für Zufallsvariablen mit Dichte einführen.) Eine Zufallsvariable X mit Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/(2\sigma^2)} \quad (x \in \mathbb{R})$$

heißt *normalverteilte Zufallsvariable mit Parametern μ und σ^2* . Wenn $\mu = 0$ und $\sigma^2 = 1$ sind, nennen wir die Zufallsvariable X *standardnormalverteilte Zufallsvariable*. Die Dichte der Normalverteilung wird häufig auch Gauß'sche Glockenkurve genannt. Sie war auf dem 10 DM-Schein abgebildet. Die Verteilungsfunktion einer normalverteilten Zufallsvariable kann nicht explizit angegeben werden. Stattdessen gibt es Tabellen in denen man die W'keiten aufgelistet findet. Weiterhin können die meisten Mathematikprogramme die Werte der Verteilungsfunktion berechnen. Normalverteilte Zufallsvariablen besitzen die folgende Eigenschaft:

Lemma 7.9. *Sei X eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable. Dann ist für $\alpha \in \mathbb{R} \setminus \{0\}$ und $\beta \in \mathbb{R}$ die Zufallsvariable*

$$\alpha X + \beta$$

gerade $\mathcal{N}(\alpha\mu + \beta, (\alpha\sigma)^2)$ -verteilt.

Das Lemma besagt insbesondere, dass sich mithilfe einer standardnormalverteilten Zufallsvariable X eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable generieren lässt. In der Tat ist

$$\sigma X + \mu$$

wegen des Lemmas $\mathcal{N}(\mu, \sigma^2)$ -verteilt.

Experiment: Eine Maschine stellt Metallstifte her. Die Stifte haben im Mittel eine Länge von 10cm und eine Standardabweichung von 0,1cm. Wenn man annimmt, dass die Abweichung von den 10cm durch mehrere kleine sich nicht beeinflussende Effekte verursacht wird, ist eine normalverteilte Zufallsvariable gut zum Modellieren der Stiftlänge geeignet.

Achtung: Jede normalverteilte Zufallsvariable nimmt mit positiver W'keit negative Werte an. Die W'keit für einen negativen Wert ist aber in unserem Beispiel so gering, dass unsere Modellierung dennoch sinnvolle Ergebnisse liefert.

7.2 Quantile und die Erzeugung von Zufallsvariablen mit dem Computer

Häufig bieten Programmiersprachen oder Mathematikprogramme die Möglichkeit gleichverteilte (Pseudo-)Zufallsvariablen auf dem Intervall $(0, 1)$ zu generieren. Mit deren Hilfe lassen sich auch andere Zufallsvariablen generieren. Dazu führen wir den Begriff des Quantils ein:

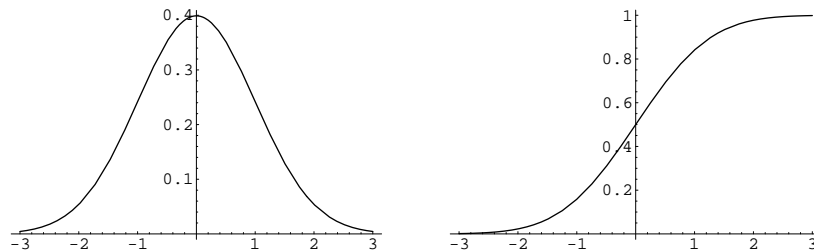


Abbildung 1: Dichte und Verteilungsfunktion der Standardnormalverteilung

Definition 7.10. Sei X eine Zufallsvariable mit Verteilungsfunktion F . Ein Wert $t \in \mathbb{R}$ mit

$$F(t) \geq \alpha \text{ und } F(t - \varepsilon) < \alpha \quad \forall \varepsilon > 0 \quad (9)$$

heißt α -Quantil der Verteilung von X (oder der Verteilungsfunktion F).

Bemerkung 7.11. Falls die Verteilungsfunktion F stetig ist, ist die Bedingung (9) äquivalent zur Bedingung

$$F(t) = \alpha.$$

Wenn die Funktion weiterhin injektiv ist, folgt, dass das α -Quantil t gerade gleich $F^{-1}(\alpha)$ ist. Für allgemeine Verteilungsfunktionen liefert

$$F^{\leftarrow}(\alpha) = \inf\{t \in \mathbb{R} : F(t) \geq \alpha\}. \quad (10)$$

ein α -Quantil von F . In der Tat gilt für $t = F^{\leftarrow}(\alpha)$

- $F(t) \geq \alpha$, da F rechtsseitig stetig ist und
- $F(t - \varepsilon) < \alpha$, da t der kleinste Wert mit $F(t) \geq \alpha$.

Satz 7.12. Sei U eine auf dem Intervall $(0, 1)$ gleichverteilte ZV und sei F eine beliebige Funktion, die die Eigenschaften (i), (ii) und (iii) erfüllt. Dann besitzt die Zufallsvariable

$$X := F^{\leftarrow}(U)$$

die Verteilungsfunktion F .

Beweis. Übungsaufgabe □

Bemerkung 7.13. Satz 7.12 liefert einen Ansatzpunkt zur Simulation von reellen Zufallsvariablen. Die meisten Softwarepakete stellen Routinen zur Generierung von auf $(0, 1)$ gleichverteilten Zufallszahlen bereit. Setzt man diese in F^{\leftarrow} ein, so erhält man nun Zufallszahlen, die zu einer beliebig wählbaren Verteilungsfunktion gehören. Selbst braucht man also nur F^{\leftarrow} zu berechnen.

7.3 Der Erwartungswert

Mit Hilfe der Verteilungsfunktion können wir allgemeinen reellen Zufallsvariablen einen Erwartungswert zuordnen:

Definition 7.14. Für eine Zufallsvariable X definieren wir

$$E(X) = \int_0^\infty (1 - F_X(t)) dt - \int_0^\infty F_X(-t) dt \quad (11)$$

und nennen $E(X)$ *Erwartungswert der Zufallsvariable X* .

Achtung: Es kann passieren, dass beide Integrale gleich ∞ sind. In diesem Fall können wir X keinen Erwartungswert zuordnen.

Wir nennen die Zufallsvariable X *integrierbar*, falls beide Integrale einen endlichen Wert liefern.

Für diskrete Zufallsvariablen X entspricht diese Definition gerade dem Erwartungswert, den wir in Abschnitt 4 kennengelernt haben.

Wir können mit dem allgemeinen Erwartungswert genauso rechnen, wie wir es bereits gewohnt sind:

Satz 7.15. *Es seien X und Y Zufallsvariablen, für die jeweils der zweite Term in (11) endlich ist.*

(i) *Der Erwartungswert der Summe $X + Y$ ist wohldefiniert und es gilt*

$$E(X + Y) = E(X) + E(Y).$$

(ii) *Für $a \in \mathbb{R}$ gilt*

$$a E(X) = E(a X),$$

wobei wir hier die Konvention $0 \cdot \infty = 0$ annehmen.

(iii) Falls X eine positive Zufallsvariable ist, gilt

$$E(X) \geq 0.$$

(iv) Für die konstante Zufallsvariable $1 = 1_\Omega$ gilt

$$E(1) = 1.$$

Beweis. Der Beweis von Eigenschaft (i) ist nicht ganz einfach. Deshalb beweisen wir nur die restlichen drei Eigenschaften.

(ii): Wir zeigen die Eigenschaft nur für den Fall $a > 0$. Die Zufallsvariable aX besitzt die Verteilungsfunktion

$$P(aX \leq t) = P(X \leq t/a) = F_X(t/a),$$

sodass eine Substitution

$$\begin{aligned} E(aX) &= \int_0^\infty (1 - F_X(t/a)) dt - \int_0^\infty F_X(-t/a) dt \\ &= \int_0^\infty a(1 - F_X(s)) ds - \int_0^\infty aF_X(-s) ds = aE(X) \end{aligned}$$

liefert.

(iii): Falls X nur positive Werte annimmt, gilt für $t < 0$

$$F_X(-t) = P(X \leq t) = 0$$

und es folgt

$$E(X) = \int_0^\infty (1 - F_X(t)) dt \geq 0.$$

(iv): Die Zufallsvariable, die konstant 1 ist, besitzt als Verteilungsfunktion $F_1(t) = 1_{[1, \infty)}$. Dies impliziert

$$E(1) = \int_0^\infty (1 - F_1(t)) dt = 1.$$

□

Wir werden nun die Erwartungswerte der im vorhergehenden Kapitel eingeführten stetigen Verteilungen berechnen. Hierzu ist die folgende einfachere Darstellung des Erwartungswerts für stetige Zufallsvariablen nützlich.

Satz 7.16. *Es sei X eine Zufallsvariable mit Dichte f . Es gilt*

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Insbesondere ist der Erwartungswert genau dann wohldefiniert, falls wir dem Integral einen eindeutigen Wert zuordnen können.

Beweis. Übungsaufgabe □

Gleichverteilung auf $[a, b]$

Für eine auf $[a, b]$ gleichverteilte ZV X gilt

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

Exponentialverteilung

Für eine exponentialverteilte Zufallsvariable X zum Parameter λ gilt:

$$\begin{aligned} E(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= [-xe^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}. \end{aligned}$$

Normalverteilung

Es bezeichne X eine $\mathcal{N}(a, \sigma^2)$ -verteilte Zufallsvariable. Eine Substitution $y = x - a$ liefert

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} x e^{-(x-a)^2/(2\sigma^2)} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} (y+a) e^{-y^2/(2\sigma^2)} dy. \end{aligned}$$

Da $y \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/(2\sigma^2)} dy$ die Dichte einer Normalverteilung ist, folgt, dass das Integral über diesen Term gleich 1 ist. Also ist

$$E(X) = a + \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} y e^{-y^2/(2\sigma^2)} dy}_{\text{ungerade Funktion}}$$

Da die Funktion im Integral ungerade (d.h. $g(-x) = -g(x)$) ist und wir das Integral über ein symmetrisches Intervall auswerten, ist dieses Integral gleich 0. Es folgt

$$E(X) = a.$$

7.4 Von diskreten zu allgemeinen Zufallsvariablen

Die Begriffe Varianz, Kovarianz, und das Moment einer Zufallsvariable sind über Erwartungswerte definiert. D.h. man kann diese Begriffe genau wie wir es für diskrete Zufallsvariablen getan haben, auch für allgemeine Zufallsvariablen verwenden. Weiterhin kann man den intuitiven Begriff der Unabhängigkeit von Zufallsvariablen formal wie folgt definieren:

Definition 7.17. Es bezeichne I eine Indexmenge. Die reellen Zufallsvariablen X_i ($i \in I$) heißen unabhängig, falls für alle endlichen Teilmengen $J \subset I$ und alle Mengen $A_j \in \mathcal{B}$ ($j \in J$)

$$P\left(\bigcap_{j \in J} \{X_j \in A_j\}\right) = \prod_{j \in J} P(X_j \in A_j)$$

gilt.

Für diskrete Zufallsvariablen mit endlichen zweiten Momenten impliziert die Unabhängigkeit die Unkorreliertheit. Dies gilt analog für allgemeine Zufallsvariablen:

Satz 7.18. *Es seien X, Y unabhängige Zufallsvariablen mit endlichen 2-ten Momenten. Dann gilt:*

$$X, Y \text{ unabhängig} \implies X, Y \text{ unkorreliert.}$$

D.h. die wichtigsten Rechenregeln, die wir in den Kapiteln 4 und 5 genutzt haben, gelten auch für allgemeine Zufallsvariablen. Entsprechend sind auch die meisten Schlussfolgerungen aus diesen Kapiteln für allgemeine Zufallsvariablen gültig. Um nachvollziehbar zu machen welche der vorhergehenden Aussagen auch für allgemeine Zufallsvariablen gelten, wurde folgende Konvention beachtet:

Wenn in einer Aussage bzw. einer Definition nicht explizit gefordert ist, dass die Zufallsvariablen diskret sind, gilt diese Aussage analog für allgemeine Zufallsvariablen!

Insbesondere gilt das Gesetze der großen Zahlen auch für allgemeine Zufallsvariablen.

Wir berechnen im Rest dieses Kapitels noch die Varianzen der uns bekannten stetigen Zufallsvariablen. Hierzu benötigen wir die folgende Formel:

Satz 7.19. Es sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion und X eine stetige Zufallsvariable. Falls $Z := \varphi(X)$ eine Zufallsvariable ist, gilt

$$E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx.$$

Insbesondere ist der Erwartungswert wohldefiniert, wenn eins der Integrale

$$\int_{-\infty}^{\infty} \max(\varphi(x), 0) f_X(x) dx \quad \text{oder} \quad \int_{-\infty}^{\infty} \min(\varphi(x), 0) f_X(x) dx$$

endlich ist.

Gleichverteilung auf $[a, b]$

Für eine auf $[a, b]$ gleichverteilte ZV X gilt

$$\text{var}(X) = E((X - E(X))^2) = \int_a^b \frac{1}{b-a} \left(x - \frac{a+b}{2}\right)^2 dx$$

und die Substitution $z = x - \frac{a+b}{2}$ liefert

$$\text{var}(X) = \frac{1}{b-a} \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} z^2 dz = \frac{1}{b-a} \frac{2}{3} \left(\frac{b-a}{2}\right)^3 = \frac{1}{12} (b-a)^2.$$

Exponentialverteilung

Für eine exponentialverteilte Zufallsvariable X zum Parameter λ gilt

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{\infty} + \underbrace{\int_0^{\infty} 2x e^{-\lambda x} dx}_{=\frac{2}{\lambda} E(X)} \end{aligned}$$

und es folgt

$$\text{var}(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Normalverteilung

Wir berechnen die Varianz einer standardnormalverteilten Zufallsvariable X . Partielle Integration liefert:

$$\begin{aligned} \text{var}(X) &= E(X^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx \\ &= \underbrace{\frac{1}{\sqrt{2\pi}} [-x e^{-x^2/2}]_{-\infty}^{\infty}}_{=0} + \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx}_{=1} = 1. \end{aligned}$$

Lernziele des 7. Kapitels

- Modellierung mit allgemeinen reelle Zufallsvariablen
- Erwartungswert, Varianz
- Stetige Zufallsvariablen
- Exponentialverteilung , Gleichverteilung, Normalverteilung

8 Eine kleine Einführung in die Statistik

8.1 Die besondere Rolle der Normalverteilung

Die Relevanz der Normalverteilung in der Statistik beruht auf dem folgenden Satz.

Satz 8.1. (Zentraler Grenzwertsatz) *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Erwartungswert $a \in \mathbb{R}$ und endlicher Varianz $\sigma^2 > 0$. Wir betrachten*

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - a}{\sigma} \quad (n \in \mathbb{N}).$$

Es gilt

$$\lim_{n \rightarrow \infty} \sup_{-\infty \leq a \leq b \leq \infty} \left| P(S_n \in [a, b]) - \underbrace{\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx}_{=\Phi(b) - \Phi(a)} \right| = 0,$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Wir betrachten das folgende Problem zur Veranschaulichung des Zentralen Grenzwertsatzes.

Problem: Eine Maschine stellt 100 Metallstäbe her. Die Länge der einzelnen Stäbe wird als eine Folge unabhängiger identisch verteilter Zufallsvariablen X_1, \dots, X_{100} modelliert. Angenommen wir kennen die Standardabweichung $\sigma := \sqrt{\text{var}(X_1)}$, aber nicht den Erwartungswert $a := E(X_1)$ der einzelnen ZV'en. Wie kann man aufgrund der Beobachtung von X_1, \dots, X_{100} Aussagen über den Parameter a treffen?

Antwort: Wir haben bereits gesehen, dass der Mittelwert

$$\bar{X}_{100} := \frac{1}{100} \sum_{i=1}^{100} X_i$$

typischerweise nahe an dem Erwartungswert a liegt. Um nun die Abweichung dieses Wertes von a zu kontrollieren nutzen wir, dass nach dem Grenzwertsatz

$$S_{100} := \frac{1}{\sqrt{100}} \sum_{n=1}^{100} \frac{X_n - a}{\sigma} = \frac{1}{10\sigma} (100 \bar{X}_{100} - 100a) = \frac{10}{\sigma} (\bar{X}_{100} - a) \quad (12)$$

ähnlich wie eine standardnormalverteilte Zufallsvariable verteilt ist. Aus Berechnung (12) folgt

$$\frac{\sigma}{10}S_{100} = \bar{X}_{100} - a$$

und

$$P(|\bar{X}_{100} - a| \leq \varepsilon) = P\left(\left|\frac{\sigma}{10}S_{100}\right| \leq \varepsilon\right) = P\left(-\frac{10}{\sigma}\varepsilon \leq S_{100} \leq \frac{10}{\sigma}\varepsilon\right).$$

Aufgrund des zentralen Grenzwertsatzes gilt also

$$P(|\bar{X}_{100} - a| \leq \varepsilon) \approx \Phi\left(\frac{10}{\sigma}\varepsilon\right) - \Phi\left(-\frac{10}{\sigma}\varepsilon\right) = 2\Phi\left(\frac{10}{\sigma}\varepsilon\right) - 1.$$

Dies liefert zum Beispiel für $\sigma = 1$ und $\varepsilon = 0.2$:

$$P(|\bar{X}_{100} - a| \leq 0.2) \approx 2\Phi(2) - 1 = 0.9545\dots$$

Der Wert \bar{X}_{100} liegt also mit großer W'keit nahe bei dem wahren Parameter a .

8.2 Die Grundidee der Statistik

Die Statistik befasst sich mit der Entwicklung von Methoden, die es erlauben mithilfe von zufallsgesteuerten Beobachtungen, die zugrundeliegenden stochastischen Gesetzmäßigkeiten zu erkunden. Wir gehen hierbei davon aus, dass X_1, X_2, \dots unabhängige Zufallsvariablen mit identischer Verteilung sind. Nun wollen wir mithilfe der Beobachtung der Zufallsvariablen X_1, \dots, X_n Aussagen über die zugrundeliegende Verteilung treffen. Wir haben bereits in Abschnitt 8.1 ein statistisches Problem behandelt (die Approximation des unbekannten Erwartungswerts a).

Um allgemeine statistische Probleme zu behandeln, legt man im ersten Schritt das sogenannte *statistische Modell* fest. Dies beschreibt aus welcher Klasse von Verteilungen die wahre (dem Beobachter unbekannte) Verteilung von X_1 (und den anderen ZV'en) stammt. Diese Klasse besteht im vorhergehenden Beispiel gerade aus allen Verteilungen mit vorgegebener Varianz σ^2 . Im vorhergehenden Beispiel haben wir den *wahren* Erwartungswert a durch die Zufallsvariable \bar{X}_{100} approximiert. Die Zufallsvariable \bar{X}_{100} wird *Schätzer* des wahren Parameters a genannt.

Konfidenzintervalle

Im Beispiel von Kapitel 8.1, haben wir berechnet, dass für $\sigma = 1$ unabhängig von dem wahren Parameter a

$$P(\underbrace{[\bar{X}_{100} - 0.2, \bar{X}_{100} + 0.2]}_{=:C} \ni a) \approx 0.9545 \geq 1 - \underbrace{0.05}_{=: \alpha}$$

gilt. D.h. der wahre Parameter a liegt mit großer W'keit in dem zufälligen Intervall C . Wir nennen C ein *approximatives Konfidenzintervall* für den Parameter a zum *Konfidenzniveau* $\alpha = 0.05$.

Allgemein nennt man ein zufälliges Intervall $C = C(X_1, \dots, X_n)$ *Konfidenzintervall* für den wahren Parameter ϑ zum Konfidenzniveau α , falls für alle im statistischen Modell enthaltenen zugrundeliegenden Verteilungen P

$$P(C \ni \vartheta) \geq 1 - \alpha$$

gilt.

Testen von Hypothesen

Nun wollen wir mithilfe eines statistischen Tests entscheiden, ob in dem statistischen Beispielmmodell von Kapitel 8.1 der wahre Parameter a größer gleich oder kleiner als 10 ist. Wir bezeichnen mit

- H_0 : alle Verteilungen des statistischen Modells mit wahren Parameter $a = E(X_1) \geq 10$ (*Hypothese*) und mit
- H_1 : alle Verteilungen des statistischen Modells mit wahren Parameter $a = E(X_1) < 10$ (*Alternative*).

Aufgrund der Beobachtung von X_1, \dots, X_{100} soll nun entschieden werden, ob die zugrundeliegende Verteilung P in der Menge H_0 oder der Menge H_1 liegt. Falls wir uns für die Menge H_0 entscheiden, sagt man, dass man die *Hypothese akzeptiert*. Andernfalls, spricht man vom *Verwerfen der Hypothese*.

Bei dieser Entscheidung können wir zwei Fehler machen:

- Die Hypothese wird verworfen, obwohl sie wahr ist (Fehler 1. Art).
- Die Hypothese wird akzeptiert, obwohl sie falsch ist (Fehler 2. Art).

Häufig ist es unvermeidbar Fehler bei der Entscheidung zu machen. Angenommen in dem Beispielmmodell liegt der wahre Parameter sehr nahe bei 10. Dann ist es nicht möglich mithilfe der Beobachtung X_1, \dots, X_{100} eine klare Entscheidung zu fällen. In diesem Fall verlangen wir von dem Test, dass der Fehler 1. Art klein ist. Im Zweifel entscheiden wir uns also für die Hypothese und gegen die Alternative. Betrachten wir folgenden Test für das Beispielmmodell: Entscheidung

- für die Hypothese, falls $\bar{X}_{100} > 10 - 0.02 = 9.98$ und
- gegen die Hypothese, falls $\bar{X}_{100} \leq 9.98$ ist.

Dann gilt für jede zugrundeliegende Verteilung aus H_0 :

$$\begin{aligned} P(\text{Hypothese wird verworfen}) &= P(\bar{X}_{100} \leq 9.98) \\ &\leq P(\bar{X}_{100} - a \leq 0.02) \approx 0.02275\% \leq 2.5\% \end{aligned}$$

D.h. uns unterläuft ein Fehler 1. Art mit *approximativer* W'keit kleiner als 2.5% (unabhängig von der wahren zugrundeliegenden Statistik). Wir haben hier einen ersten *approximativen Hypothesentest* kennengelernt.

Allgemein teilt man beim Testen von Hypothesen das statistische Modell in zwei Bereiche: H_0 und H_1 . Wir sagen nun, dass ein Test der Hypothese H_0 gegen die Alternative H_1 *Signifikanzniveau* (oder kurz *Niveau*) $\alpha > 0$ hat, falls für alle Verteilungen P aus H_0 , die Abschätzung

$$P(\text{Hypothese wird verworfen}) \leq \alpha$$

gilt. D.h. unabhängig vom zugrundeliegenden Modell machen wir einen Fehler 1. Art mit W'keit kleiner gleich α .

8.3 Der t-Test

Wir führen in diesem Abschnitt einen in der Praxis häufig angewandten Test ein.

Statistisches Modell: X_1 (und die weiteren Zufallsvariablen X_2, \dots) sind normalverteilt zu den unbekannten Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$.

Ziel: Herleitung eines Konfidenzintervalls für den wahren Parameter a zum Niveau α auf der Basis von $n \geq 2$ Beobachtungen.

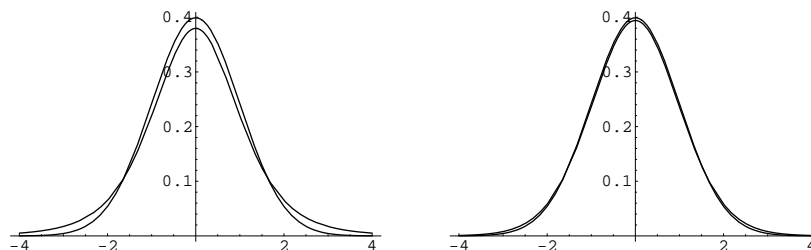


Abbildung 2: Vergleich der Dichten der Standardnormalverteilung mit der t_5 (bzw. t_{20})-Verteilung)

Wir konstruieren ein Konfidenzintervall für μ mithilfe einer Statistik (Zufallsvariable), deren Verteilung nicht von der Wahl des zugrundeliegenden Modells (von der Wahl von P) abhängt. Wir nutzen weiterhin die Notation

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Satz 8.2. Sei μ der wahre Erwartungswert des zugrundeliegenden Modells und setze

$$\hat{s}_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2 \text{ und } T = \sqrt{n} \frac{\bar{X}_n - \mu}{\hat{s}_n}.$$

Dann ist die Zufallsvariable T Student-verteilt zum Parameter $n-1$ (kurz: t_{n-1} -verteilt).

Bemerkung 8.3. Die t_n -Verteilung ist eine stetige Verteilung mit symmetrischer Dichte. Ihre Verteilungsfunktion sowie deren Quantile können von den meisten Mathematikprogrammen berechnet werden. Weiterhin kann man die Quantile der t_n -Verteilung in Tabellen nachschlagen. Die t_n -Verteilung ähnelt für große n der Standardnormalverteilung (siehe Abbildung).

Der Satz besagt insbesondere, dass die Verteilung von T für alle Verteilungen des statistischen Modells gleich ist. Es bezeichne nun $\alpha > 0$ das Niveau des angestrebten Konfidenzintervalls für den wahren Parameter μ . Wir wählen τ_α als $(1 - \alpha/2)$ -Quantil der t_{n-1} -Verteilung. Wegen der Symmetrie, ist $-\tau_\alpha$

ein $\alpha/2$ -Quantil und es folgt unabhängig von der wahren zugrundeliegenden Verteilung:

$$P(T \in [-\tau_\alpha, \tau_\alpha]) = 1 - \alpha.$$

Wir übersetzen die Aussage $|T| \leq \tau_\alpha$ nun in eine Aussage bzgl. des Abstands von \bar{X}_n nach μ :

$$|T| \leq \tau_\alpha \iff \sqrt{n} \frac{|\bar{X}_n - \mu|}{\hat{s}_n} \leq \tau_\alpha \iff |\bar{X}_n - \mu| \leq \frac{\tau_\alpha}{\sqrt{n}} \hat{s}_n.$$

Also gilt allgemein:

$$P([\bar{X}_n - \frac{\tau_\alpha}{\sqrt{n}} \hat{s}_n, \bar{X}_n + \frac{\tau_\alpha}{\sqrt{n}} \hat{s}_n] \ni \mu) = 1 - \alpha.$$

Satz 8.4. Sei $\alpha \in (0, 1)$, $n \geq 2$ und τ_α das $(1 - \alpha/2)$ -Quantil der t_{n-1} -Verteilung. Für das Gauß'sche statistische Modell dieses Kapitels ist

$$[\bar{X}_n - \frac{\tau_\alpha}{\sqrt{n}} \hat{s}_n, \bar{X}_n + \frac{\tau_\alpha}{\sqrt{n}} \hat{s}_n]$$

ein Konfidenzintervall für μ zum Niveau α .

Analog können wir einen statistischen Test zum Niveau α angeben, der die

- Hypothese, dass $\mu \leq \mu_0$ für ein festgelegten Wert $\mu_0 \in \mathbb{R}$ ist, gegen die
- Alternative, dass $\mu > \mu_0$ ist,

testet. Wir nutzen wieder, dass

$$T = \sqrt{n} \frac{\bar{X}_n - \mu}{\hat{s}_n}$$

unabhängig von der zugrundeliegenden Verteilung t_{n-1} -verteilt ist. Es bezeichne nun q_α das $1 - \alpha$ -Quantil der t_{n-1} -Verteilung, d.h.

$$P(T \geq q_\alpha) = \alpha.$$

Wir formen wieder das Ereignis äquivalent um:

$$T \geq q_\alpha \iff \sqrt{n} \frac{\bar{X}_n - \mu}{\hat{s}_n} \geq q_\alpha \iff \bar{X}_n \geq \mu + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n.$$

Also folgt

$$\alpha = P(T \geq q_\alpha) = P(\bar{X}_n \geq \mu + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n).$$

Falls der wahre Parameter μ kleiner gleich μ_0 ist (die Hypothese also wahr ist), erhalten wir

$$P(\bar{X}_n \geq \mu_0 + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n) \leq P(\bar{X}_n \geq \mu + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n) = \alpha.$$

Also hat der Test, der

- die Hypothese akzeptiert, falls $\bar{X}_n < \mu_0 + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n$ ist, und
- die Hypothese verwirft, falls $\bar{X}_n \geq \mu_0 + \frac{q_\alpha}{\sqrt{n}} \hat{s}_n$ ist,

Signifikanzniveau α .

8.4 Approximatives Konfidenzintervall für Poisson-verteilte Zufallsvariablen

Wir betrachten nun das folgende

Statistische Modell: X_1, X_2, \dots sind unabhängig Poisson-verteilte Zufallsvariablen zum unbekannten Parameter $\lambda > 0$.

Die Zufallsvariablen können zum Beispiel die Anzahl der Zugriffe auf einen Server in aufeinanderfolgenden Zeitintervallen beschreiben. Aufgrund der ersten n Beobachtungen wollen wir nun den wahren Parameter λ schätzen. Natürlich ist auch in diesem Fall \bar{X}_n ein geeigneter Schätzer für den Erwartungswert und damit auch für den Parameter λ . Wir konstruieren nun ein approximatives Konfidenzintervall zum Parameter $\alpha \in (0, 1)$.

Der Erwartungswert und die Varianz der Zufallsvariablen sind gleich λ , und der zentrale Grenzwertsatz sagt aus, dass für große $n \in \mathbb{N}$

$$S_n := \sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \tag{13}$$

sich statistisch ähnlich wie eine standardnormalverteilte ZV verhält. Wir bezeichnen mit τ_α das $1 - \alpha/2$ -Quantil der Standardnormalverteilung. Dann gilt

$$P\left(-\tau_\alpha \leq \sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq \tau_\alpha\right) \approx 1 - \alpha.$$

Als nächstes übersetzen wir die Bedingung $\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq \tau_\alpha$ an eine anschaulichere Bedingung an λ . Hierbei benutzen wir die Notationen $z = \tau_\alpha / \sqrt{n}$ und

$\mu = \sqrt{\lambda}$:

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq \tau_\alpha \iff \bar{X}_n - \lambda \leq z\sqrt{\lambda} \iff \mu^2 + z\mu - \bar{X}_n \geq 0.$$

Die Parabel $\mu \mapsto \mu^2 - z\mu - \bar{X}_n$ ist nach oben geöffnet und hat genau eine positive Nullstelle (da der Funktionswert an der Stelle 0 negativ ist). Die positive Nullstelle ist gerade

$$\mu_0 := -\frac{z}{2} + \sqrt{\frac{z^2}{4} + \bar{X}_n}.$$

Also folgt

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq \tau_\alpha \iff \lambda \geq \mu_0^2 = \left(\sqrt{\bar{X}_n + \frac{z^2}{4}} - \frac{z}{2} \right)^2.$$

Andererseits gilt:

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \geq -\tau_\alpha \iff \bar{X}_n - \lambda \geq -z\sqrt{\lambda} \iff \mu^2 - z\mu - \bar{X}_n \leq 0$$

Die Parabel hat als einzige positive Nullstelle

$$\mu_1 = \frac{z}{2} + \sqrt{\frac{z^2}{4} + \bar{X}_n}$$

und wir erhalten:

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \geq -\tau_\alpha \iff \lambda \leq \mu_1^2 = \left(\sqrt{\bar{X}_n + \frac{z^2}{4}} + \frac{z}{2} \right)^2.$$

Also gilt für

$$C := \left[\left(\sqrt{\bar{X}_n + \frac{\tau_\alpha^2}{4n}} - \frac{\tau_\alpha}{2\sqrt{n}} \right)^2, \left(\sqrt{\bar{X}_n + \frac{\tau_\alpha^2}{4n}} + \frac{\tau_\alpha}{2\sqrt{n}} \right)^2 \right]$$

die Abschätzung

$$P(C \ni \lambda) \approx 1 - \alpha$$

und C ist ein approximatives Konfidenzintervall für λ zum Niveau α .

Bemerkung 8.5. Um das approximative Konfidenzintervall zu berechnen, haben wir die echte Verteilung der Zufallsvariable S_n aus (13) durch die Standardnormalverteilung approximiert. Die Güte der Approximation hängt nur von dem Wert $n\lambda$ ab. Wie gut die Approximation ist, zeigt die Abbildung 3. Die Abbildung vergleicht die wahren Quantile der Zufallsvariable S_n mit denen der Standardnormalverteilung für den Fall $n\lambda = 500$.

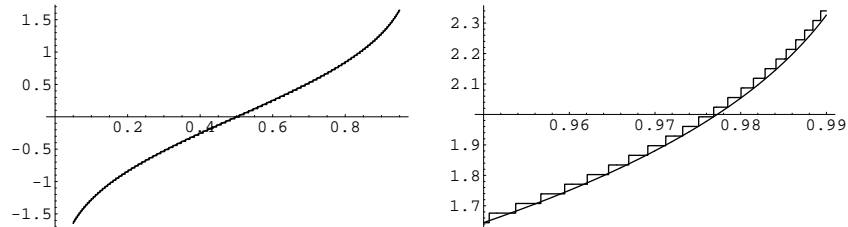


Abbildung 3: Quantile der Standardnormalverteilung und von S_n im Fall $n\lambda = 500$

Beispiel 8.6. Es soll die Intensität der Zugriffe auf einen Server geschätzt werden. Die Anzahl der Zugriffe werden hierbei als unabhängige Poisson-verteilte Zufallsvariablen zum unbekannten Parameter $\lambda > 0$ modelliert. Es soll ein Konfidenzintervall für den Parameter λ zum Niveau $\alpha = 0.05$ bestimmt werden.

Angenommen wir beobachten den Server über 100 Zeitintervalle und erhalten als Realisierung von \bar{X}_{100} gerade den Wert 7.4. Um ein Konfidenzintervall anzugeben müssen wir gerade

$$n = 100 \text{ und } \tau_{0.05} = 1.95996$$

in die obige Darstellung einsetzen und erhalten das Konfidenzintervall

$$[6.88569, 7.95272]$$

für λ . Zum Vergleich: Hätten wir nach 1000 Messungen als Mittelwert \bar{X}_{1000} die Zahl 7.4 realisiert, ergäbe dies als Konfidenzintervall

$$[7.23331, 7.57053]. \quad (14)$$

Bemerkung 8.7. Wenn n sehr groß ist, unterscheidet sich das obige Konfidenzintervall nur wenig von dem Intervall

$$\left[\bar{X}_n - \sqrt{\bar{X}_n} \frac{\tau_\alpha}{\sqrt{n}}, \bar{X}_n + \sqrt{\bar{X}_n} \frac{\tau_\alpha}{\sqrt{n}} \right].$$

Hier verwendet man als Schätzer für die Standardabweichung $\sqrt{\bar{X}_n}$. Wenn man z.Bsp. nach 1000 Zeitintervallen den Wert 7.4 für \bar{X}_{1000} erhält, ist das entsprechende Intervall gerade

$$[7.2314, 7.5686].$$

Dies unterscheidet sich nur wenig von dem oben berechneten Konfidenzintervall (14). Weiterhin kann man das Konfidenzintervall auch mit dem Konfidenzintervall aus Satz 8.4 vergleichen und wird typischerweise für große Stichproben kaum Unterschiede feststellen.

Bemerkung 8.8. Die Erfahrung zeigt, dass das Konfidenzintervall aus Satz 8.4 auch für viele andere nicht normalverteilte Zufallsvariablen gute Ergebnisse liefert und die meisten mathematischen Softwarepakete berechnen Konfidenzintervalle auf dieser Basis.

Lernziele des 8. Kapitels

- Schätzung von Parametern
- Konfidenzintervalle
- Testen von Hypothesen