

Kurzskript: Stochastik für Informatiker

Georg Neuhaus

März 2005

Inhaltsverzeichnis

1	Zufallsexperimente und Diskrete W-Räume	1
2	Einige Grundformeln der Kombinatorik	4
3	Einige Rechenregeln für W-Maße	5
4	Wichtige diskrete W-Verteilungen	5
5	Diskrete Zufallsvariable und ihre Verteilungen	9
6	Elementare bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit	11
6.1	Elementare bedingte Wahrscheinlichkeit	11
6.2	Einige Rechenregeln für bedingte Wahrscheinlichkeiten	13
6.3	(Stochastische) Unabhängigkeit von Ereignissen	14
6.4	(Stochastische) Unabhängigkeit von diskreten Zufallsvariablen	14
7	Erwartungswert und Varianz von diskreten Zufallsvariablen bzw. von ihren Verteilungen	17
8	Das schwache Gesetz der großen Zahlen	20
9	Verteilungskonvergenz und zentraler Grenzwertsatz	24
10	Schätzer, Konfidenzintervalle und Test: Die drei wichtigsten statistischen Verfahren am Beispiel von Binomialexperimenten	32
11	Allgemeine Maßräume und Zufallsexperimente über \mathbb{R}^k	34
11.1	W-Maße mit Riemann-Dichten	38
11.2	Erwartungswert und Varianz von Verteilungen mit R-Dichten	42
12	Allgemeine Zufallsvariable und ihre Verteilungen	42
13	Produkträume, Produktmaße, stochastische Unabhängigkeit und die Transformationsformel für R-Dichten	44
14	Koppelung von Experimenten	51
15	Monte Carlo Methode	53
16	Markov-Ketten – einige Grundbegriffe	55
16.1	Markov-Ketten im Gleichgewicht	58

17 Bediensysteme	62
17.1 Vorbemerkungen	62
17.2 Das Bedienmodell $M M 1 \infty$	63
17.3 Das $M M 1$ -Bediensystem im Gleichgewicht	68
17.4 Leistungsmaße im $M M 1$ -Bediensystem	70
18 Symbole, Abkürzungen, Verteilungen und Literaturliste	73
18.1 Symbole	73
18.2 Abkürzungen	74
18.3 Verteilungen und ihre Dichten	74
18.4 Literaturliste	75

1 Zufallsexperimente und Diskrete W-Räume

Der Begriff **Zufallsexperiment** steht für eine Situation, die ein vom Zufall beeinflusstes Ergebnis liefert. Die Menge der möglichen Ergebnisse ist Teilmenge des *Grundraumes* Ω . Bis auf die Tatsache, dass der Grundraum die möglichen Ergebnisse enthält, wird nichts vorausgesetzt, häufig wird er exakt gleich der Menge der möglichen Ergebnisse angesetzt. Die verschiedenen *Ereignisse*, d.h. Mengen von Ergebnissen, treten mit gewissen *Wahrscheinlichkeiten* auf.

Das Tripel (*Grundraum, Ereignisse, Wahrscheinlichkeiten*) bildet den sogenannten *Wahrscheinlichkeitsraum* (W-Raum).

1. $\Omega (\neq \emptyset)$: Der *Grundraum* oder *Stichprobenraum*, oder *Merkmalraum* in dem die möglichen Ausgänge eines Zufallsexperiments liegen.

Beispiel: Würfeln mit einem Würfel: $\Omega = \underline{6}$ (oder $\Omega = \mathbb{R}, \Omega = \mathbb{Z}, \dots$), Würfeln mit zwei Würfeln: $\Omega = \underline{6} \times \underline{6}$.

2. $\mathcal{P}(\Omega)$: Wenn nur abzählbar viele Versuchsausgänge möglich sind (und diesen Fall werden wir bis auf weiteres nur behandeln), können wir o.E. *alle* Teilmengen von Ω wählen. Jede solche Teilmenge repräsentiert ein *Ereignis*. Bei den später behandelten Experimenten mit überabzählbar vielen Versuchsausgängen, wie zum Beispiel \mathbb{R} , ist es dagegen unumgänglich, die Familie der betrachteten Ereignisse einzuschränken. Dies wird später zum Begriff der σ -Algebren \mathcal{A} führen und damit auch zu maßtheoretischen Überlegungen. Wir beginnen hier mit Zufallsexperimenten mit abzählbar vielen Versuchsausgängen. Dann treten keine maßtheoretischen Begriffe und Probleme auf.

Beispiel: Beim einzelnen Würfelwurf repräsentiert $A = \{2, 4, 6\}$ das Ereignis „eine gerade Zahl wird gewürfelt“. Beim doppelten Würfelwurf repräsentiert $A = \{(i, j) : i + j < 6\}$ das Ereignis „Augensumme ist kleiner als sechs“.

3. P : Ein *Wahrscheinlichkeitsmaß* (W-Maß), d.h. eine Abbildung $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ mit gewissen Eigenschaften (siehe Definition 1.1).

Beispiel: Idealer Würfel: $P\{i\} = \frac{1}{6}$ für $i \in \underline{6}$.

Neben dem Grundraum Ω , dem Ereignissystem $\mathcal{P}(\Omega)$ ist der dritte Bestandteil des Wahrscheinlichkeitsraumes die Bewertungsfunktion, die jedem Ereignis $A \in \mathcal{P}(\Omega)$ einen W-Wert zuordnet.

1.1 Definition. (*Diskreter W-Raum*) Gegeben sei eine Menge $\Omega (\neq \emptyset)$. Jede Abbildung $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ (P wie *Probability*) mit den Eigenschaften

$$P(A) \geq 0 \quad \forall A \in \mathcal{P}(\Omega) \quad (\text{Nichtnegativität}) \quad (1.1)$$

$$P(\Omega) = 1 \quad (\text{Normiertheit}) \quad (1.2)$$

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\sigma\text{-Additivität}) \quad (1.3)$$

$$P(\Omega_0) = 1 \quad \text{für eine abzählbare Teilmenge } \Omega_0 \subset \Omega \quad (\text{Diskretheit}), \quad (1.4)$$

wobei $\sum_{i=1}^{\infty} A_i := \bigcup_{i=1}^{\infty} A_i$ falls $A_i \cap A_j = \emptyset \quad \forall i \neq j$, ansonsten ist die Summe nicht definiert, heißt **diskretes Wahrscheinlichkeitsmaß** (diskretes W-Maß) auf $\mathcal{P}(\Omega)$. Das Tripel $(\Omega, \mathcal{P}(\Omega), P)$ heißt *diskreter Wahrscheinlichkeitsraum* (diskreter W-Raum) und man spricht von einem diskreten Zufallsexperiment. $PA := P(A)$ heißt die W. von A. A heißt ein fast unmögliches Ereignis, wenn $PA = 0$ und ein fast sicheres Ereignis, wenn $PA = 1$.

Da das Ereignissystem $\mathcal{P}(\Omega)$ hier feststeht, läßt man es meist weg und schreibt nur (Ω, P) .

Ist P ein diskretes W-Maß, so gilt: $P(\emptyset) = P(\sum_{i=1}^{\infty} \emptyset) \stackrel{(1.3)}{=} \sum_{i=1}^{\infty} P(\emptyset) \in \mathbb{R} \Rightarrow P(\emptyset) = 0$; man sagt: P ist **nulltreu**.

1.2 Bemerkung. Jedes W-Maß P ist auch endlich-additiv, d.h. (1.3) gilt auch für endlich viele, paarweise disjunkte Mengen $A_1, \dots, A_n \in \mathcal{P}(\Omega)$, denn: $P(A_1 + \dots + A_n) = P(A_1 + \dots + A_n + \emptyset + \dots) = P(A_1) + \dots + P(A_n) + P(\emptyset) + \dots = P(A_1) + \dots + P(A_n)$. \square

1.3 Beispiel. (Idealer Würfel) $\Omega = \underline{6}$. $P\{1\} = \dots = P\{6\}$ mit $P\{i\} = \frac{1}{6}$. Das reicht zur Festlegung von P , denn für $A \subset \Omega$ gilt $P(A) = P(\sum_{i \in A} \{i\}) = \sum_{i \in A} P\{i\} = \frac{1}{6}|A|$, also z.B. $P\{1, 2, 3\} = \frac{1}{6}3 = \frac{1}{2}$.

Ein diskreter W-Raum (Ω, P) beschreibt ein diskretes Zufallsexperiment. Bei Durchführung des Experiments erhält man irgendeinen Wert $\omega \in \Omega$ als Ergebnis. ω ist eine sogenannte *Realisierung* des Zufallsexperiments.

Der intuitive Wahrscheinlichkeitsbegriff basiert auf der Erfahrungstatsache, dass bei unabhängiger Wiederholung eines Zufallsexperiments sich die relative Häufigkeit des Eintretens eines Ereignisses A stabilisiert.

Sei $\omega_1, \omega_2, \dots$ eine Folge von unabhängigen Realisierungen eines Zufallsexperiments über Ω .

Die relative Häufigkeit des Eintretens von A in den ersten n Versuchen ist $H_n(A)$ mit

$$H_n(A) = \frac{\sum_{i=1}^n 1_A(\omega_i)}{n} \tag{1.5}$$

wobei

$$1_A(\omega) := \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$$

die *Indikatorfunktion* von A heißt.

Erfahrungstatsache: Bei unabhängiger Versuchswiederholung stabilisiert sich $H_n(A)$. $H_n(A)$ scheint gegen einen Grenzwert zu konvergieren, den man die Wahrscheinlichkeit $P(A)$ nennen möchte. (Kein mathematischer Satz!!) H_n hat die Eigenschaften (1.1) – (1.4) (Übungsaufgabe).

Frage: Warum fordert man σ -Additivität und nicht endliche Additivität?

Antwort: Das hat mathematische Gründe. Man möchte sichergehen, dass die W. einer unendlichen Teilmenge von Ω gleich der Summe der W.en der Punkte ist, aus denen sie sich

zusammensetzt. Additivität und σ -Additivität stehen in ähnlicher Relation zueinander, wie endliche und unendliche Zahlenfolgen in der Analysis.

Ist (Ω, P) ein diskreter W-raum, so gilt

$$P(A) = P\left(\sum_{\omega \in A \cap \Omega_0} \{\omega\}\right) = \sum_{\omega \in A \cap \Omega_0} P\{\omega\}. \quad (1.6)$$

Man nennt die durch

$$f(\omega) = P\{\omega\} \quad , \omega \in \Omega, \quad (1.7)$$

definierte Abbildung $f : \Omega \rightarrow \mathbb{R}$ die zu P gehörige **Zähldichte** .

Die (abzählbare!) Menge $T := \{\omega \in \Omega : f(\omega) > 0\}$ heißt der **Träger** von P . Offenbar kann in der Definition der diskreten W-Maße $\Omega_0 = T$ gesetzt werden. Man beweist leicht das

1.4 Lemma. *Sei $\Omega \neq \emptyset$. Eine Abbildung $f : \Omega \rightarrow \mathbb{R}_+$ ist genau dann eine Zähldichte (eines eindeutig bestimmten) diskreten W-Maßes auf $\mathcal{P}(\Omega)$, wenn gilt:*

$$\sum_{\omega \in \Omega} f(\omega) = 1, \quad (1.8)$$

wobei in der obigen evtl. überabzählbaren Summe alle verschwindenden Terme weggelassen werden, sodass eine abzählbare Summe übrigbleibt. Wir nennen jede Abbildung $f : \Omega \rightarrow \mathbb{R}_+$ mit (1.8) eine Zähldichte.

Man kann sich die Wahrscheinlichkeiten $f(\omega)$ als Massen auf den Punkten ω vorstellen. Die Gesamtmasse 1 ist also über den Raum Ω verteilt. W-Maße heißen deshalb auch **W-Verteilungen** .

Bei **endlichen Zufallsexperimenten**, d.h. diskreten Zufallsexperimenten mit endlichem Träger T , erscheint manchmal kein Element $\omega \in T$ bevorzugt zu sein. Man macht in solchen Fällen die **Gleichverteilungsannahme**, d.h. man nimmt an, dass alle $P\{\omega\}, \omega \in T$, denselben Wert haben. Dazu

1.5 Definition. *Ein Laplace-Experiment ist ein Paar (Ω, f) bestehend aus einer endlichen Menge Ω und der Zähldichte*

$$f(\omega) = 1/|\Omega| \quad \forall \omega \in \Omega. \quad (1.9)$$

Für jedes $A \subset \Omega$ gilt dann

$$\begin{aligned} P(A) &= \frac{|A|}{|\Omega|} \\ &= \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} \end{aligned} \quad (1.10)$$

Das zugehörige W-Maß heißt dann **Gleichverteilung** oder **Laplace-Verteilung** über Ω .

Der ideale Würfel wird z.B. durch die Gleichverteilung über $\Omega = \{1, \dots, 6\}$ beschrieben, d.h. alle Elemente von Ω erhalten dieselbe Masse $1/6$.

Laplace Experimente treten auf, wenn aus einer endlichen Grundgesamtheit Ω *zufällig* ein Element ausgewählt wird. Die Bestimmung der Anzahlen $|A|$ und $|\Omega|$ erfolgt häufig mit Hilfe kombinatorischer Überlegungen. Wir stellen im folgenden Abschnitt einige Grundformeln der Kombinatorik zusammen.

2 Einige Grundformeln der Kombinatorik

2.1 Formel. (Permutation mit Wiederholung)

Aus einer Menge $\{\omega_1, \dots, \omega_n\}$ von n Elementen kann man auf n^r verschiedene Arten geordnete Proben mit Wiederholung vom Umfang r entnehmen.

2.2 Formel. (Permutation ohne Wiederholung)

Aus einer Menge $\{\omega_1, \dots, \omega_n\}$ von n Elementen kann man auf

$$\frac{n!}{(n-r)!} = n(n-1) \cdots (n-r+1)$$

verschiedene Arten geordnete Proben ohne Wiederholung vom Umfang r entnehmen, $r \leq n$.

2.3 Formel. (Kombination ohne Wiederholung)

Aus einer Menge $\{\omega_1, \dots, \omega_n\}$ von n Elementen kann man auf

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

verschiedene Arten ungeordnete Proben ohne Wiederholung vom Umfang r entnehmen. Dabei $r \leq n$.

2.4 Formel. (Kombination mit Wiederholung)

Aus einer Menge $\{\omega_1, \dots, \omega_n\}$ von n Elementen kann man auf

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

verschiedene Arten ungeordnete Proben mit Wiederholung vom Umfang r entnehmen.

Die Berechnung von Fakultäten macht bei größeren Zahlen von n und r Schwierigkeiten. Als gute Approximation hat sich die Stirlingsche Formel erwiesen.

2.5 Formel. Stirlingsche Formel

Für alle $n \in \mathbb{N}$ gilt

$$n^n \cdot e^{-n} \sqrt{2\pi n} \cdot e^{1/(12n+1)} < n! < n^n \cdot e^{-n} \sqrt{2\pi n} \cdot e^{1/12n}. \quad (2.1)$$

Zum Beweis dieser Formel, die zur Näherung

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

benutzt wird, siehe Feller I.

3 Einige Rechenregeln für W-Maße

[BN]:1.6 Seien (Ω, P) ein diskreter W-Raum und $A, B, A_1, A_2, \dots \subset \Omega$. Dann gilt:

$$A \subset B \Rightarrow P(A) \leq P(B) \quad (\text{Isotonie}) \quad (3.1)$$

Es gilt

$$P(B \setminus A) = P(B) - P(A) \quad (\text{Subtraktivität}) \quad (3.2)$$

Beweis: $P(B) = P(A + B \setminus A) = P(A) + P(B \setminus A) \geq P(A)$. \square

Folgerung: Ist P ein W-Maß, so gilt

$$P(A^c) = 1 - P(A). \quad (3.3)$$

Ist $(A_i)_{i \geq 1}$ isoton, d.h. $A_1 \subset A_2 \subset \dots$, so gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n), \quad (3.4)$$

P ist also *stetig von unten*.

Ist $(A_i)_{i \geq 1}$ antiton, d.h. $A_1 \supset A_2 \supset \dots$, so folgt

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n), \quad (3.5)$$

P ist also *stetig von oben*.

Weiter gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \quad (\text{Sub-Sigma-Additivität}) \quad (3.6)$$

3.1 Bemerkung. Sei $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ nichtnegativ, nulltreu und endlich additiv. Dann ist die Stetigkeit von unten äquivalent zur σ -Additivität, äquivalent zur Stetigkeit von oben und sogar äquivalent zur \emptyset -Stetigkeit, d.h.

$$A_1 \supset A_2 \supset \dots \quad \text{mit} \quad \bigcap_{i=1}^{\infty} A_i = \emptyset, \quad \Rightarrow \quad \lim_{n \rightarrow \infty} P(A_n) = 0 \quad (3.7)$$

(P ist \emptyset -stetig).

4 Wichtige diskrete W-Verteilungen

4.1 Beispiel. [BN]:Sek.2 Urnenmodell: Ziehen mit Zurücklegen

In Fächern $1, \dots, n$ seien jeweils N Kugeln $1, \dots, N$ enthalten. Nacheinander werde jeweils eine Kugel aus jedem der n Fächer gezogen (äquivalent: n -maliges Ziehen aus einem Fach mit Zurücklegen).

$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \underline{N} \quad \forall i \in \underline{n}\}$ ist ein natürlicher Grundraum für das Experiment. Es ist $|\Omega| = N^n$. „Zufällige“ und „unabhängige“ Entnahme soll heißen, dass alle Folgen $\omega = (\omega_1, \dots, \omega_n)$ gleichberechtigt sind (Laplace-Experiment über Ω):

$$P\{\omega\} = \frac{1}{N^n} \quad \forall \omega \in \Omega.$$

4.2 Beispiel. Binomial (n, p) -Verteilung (kurz: $B(n, p)$ -Vtlg, [BN]:2.7)

Grundraum: $\Omega = \{0, \dots, n\}$

Zähldichte:

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n. \quad (4.1)$$

Wegen $\sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n$ gilt $\sum_{k=0}^n b(k; n, p) = 1$.

Wenn n gleichartige Versuche mit Versuchsausgängen Erfolg und Mißerfolg unabhängig voneinander durchgeführt werden und die Erfolgswahrscheinlichkeit beim Einzelversuch p ist, ist $b(k; n, p)$ die Wahrscheinlichkeit für „genau k Erfolge“.

Spezialfall: $n = 1$: Man nennt die $B(1, p)$ -Vtlg auch Bernoulli (p) -Vtlg (kurz: $B(p)$ -Vtlg).

$B(p)$ hat den Träger $\{0, 1\}$, wobei 1 die Wahrscheinlichkeit p und 0 die Wahrscheinlichkeit $q := (1-p)$ trägt.

$B(\frac{1}{2})$ beschreibt einen idealen Münzwurf.

Herleitung der Binomialverteilung:

Der Grundraum Ω ist derselbe wie in 4.1. Erfolg im i -ten Versuch soll bedeuten, dass für ein fe mit $K \in \underline{N}$ gilt: $\omega = (\omega_1, \dots, \omega_n)$ erfüllt $\omega_i \in \underline{K}$. Man denke sich die Kugeln mit Nummer $\leq K$ rot gefärbt, die übrigen schwarz gefärbt. Erfolg entspricht dann dem Ziehen einer roten Kugel. Es sei $A(k) := \{\omega \in \Omega : \text{für genau } k \text{ Versuche Erfolg}\}$, also

$$A(k) = \{\omega \in \Omega : \sum_{i=1}^n 1_{\underline{K}}(\omega_i) = k\},$$

und $A^0(k) := \{\omega \in \Omega : \text{Erfolg genau in den ersten } k \text{ Versuchen}\}$.

Es ist $|A^0(k)| = K^k (N-K)^{n-k}$.

Dasselbe Ergebnis folgt, wenn irgendwelche k Versuche statt der ersten k herausgegriffen werden.

Es gibt $\binom{n}{k}$ verschiedene Möglichkeiten k Versuche aus n Stück herauszugreifen. Es folgt

$$P(A(k)) = \binom{n}{k} \frac{K^k (N-K)^{n-k}}{N^n} = \binom{n}{k} p^k (1-p)^{n-k} =: b(k; n, p)$$

mit $p = \frac{K}{N}$ = Erfolgswahrscheinlichkeit im Einzelversuch. Die Eigenschaft

$$\sum_{k=0}^n b(k; n, p) = 1$$

folgt automatisch für die betrachteten $p = K/N$. Aus Stetigkeitsgründen folgt dann die Gleichheit für beliebige $p \in [0, 1]$.

4.3 Beispiel.

Poisson(λ)-Experiment, [BN]:2.8

Seien $T := \Omega := \mathbb{N}_0 = \{0, 1, \dots\}$, $\lambda \in (0, \infty)$. Dann ist $p(\cdot; \lambda)$ mit

$$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0, \quad (4.2)$$

eine Zähldichte über \mathbb{N}_0 . $\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda\right)$.

Kurz: $\mathcal{P}(\lambda)$ -Verteilung.

Die Poisson(λ)-Vtlg beschreibt z.B. die Anzahl der Zerfälle pro Zeiteinheit eines radioaktiven Präparates. Sie ist eine Approximation der Binomial(n, p)-Verteilung für große n und kleine p , s. Lemma 4.8.

4.4 Beispiel. Geometrische Verteilung Seien $T := \Omega := \mathbb{N}_0 = \{0, 1, \dots\}$, $p \in (0, 1)$. Dann ist wegen der geometrischen Reihe $1 + p + p^2 + \dots = (1 - p)^{-1}$ die Funktion $f(k) = (1 - p)p^k$ über Ω eine Zähldichte. Das zugehörige W -Maß heißt geometrische(p)-Verteilung, kurz: $Geo^0(p)$ -Verteilung. Die geometrische Verteilung beschreibt bei unabhängigen Bernoulli(p)-Experimenten die Anzahl der Versuche mit Misserfolg bis zum ersten Mal Erfolg eintritt.

4.5 Beispiel. n -maliges Ziehen aus einer Urne ohne Zurücklegen. Hypergeometrische Verteilung

$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \underline{N} \quad \forall i \in \underline{n} \wedge \omega_i \neq \omega_j \quad \forall i \neq j\}$, $|\Omega| = N(N - 1) \cdots (N - n + 1) =: (N)_n = \frac{N!}{(N-n)!}$.

Alle $\omega \in \Omega$ seien gleichwahrscheinlich: $P\{\omega\} = \frac{1}{|\Omega|}$ (Laplace-Experiment).

Wie oben sei $A(k) = \{\omega \in \Omega \mid \text{für genau } k \text{ Komponenten } \omega_i \text{ gilt } \omega_i \in \underline{K}\}$.

$A^0(k) := \{\omega \in \Omega \mid \text{Erfolg im Versuch } 1, \dots, k\}$.

$|A^0(k)| = [K(K - 1) \cdots (K - k + 1)][(N - K)(N - K - 1) \cdots (N - K - (n - k) + 1)] = \frac{K!}{(K-k)!} \frac{(N-K)!}{(N-K-(n-k))!}$.

Es gibt wiederum $\binom{n}{k}$ Möglichkeiten, Indizes (i_1, \dots, i_k) statt $(1, \dots, k)$ festzulegen. Folglich gilt

$$\begin{aligned} P(A(k)) &= \frac{\left(\frac{K!}{(K-k)!} \frac{(N-K)!}{(N-K-(n-k))!} \frac{n!}{k!(n-k)!}\right)}{\left(\frac{N!}{(N-n)!}\right)} \\ &= \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} =: h(k; N, K, n). \end{aligned} \quad (4.3)$$

Dabei sind die Nebenbedingungen

$$\max(0, n - N + K) \leq k \leq \min(K, n) \quad (4.4)$$

offenbar erfüllt. Wiederum folgt automatisch, dass \sum_k mit (4.4) $h(k; N, K, n) = 1$ gilt.

Ein W -Maß mit Träger $\{k : k \text{ erfüllt (4.4)}\}$ und mit Zähldichte $p(k) = h(k; N, K, n)$ heißt

Hypergeometrische (N, K, n) -Vtlg. (Kurz: $\mathcal{H}(N, K, n)$ -Vtlg, [BN]:2.9)

4.6 Beispiel. (zur Binomialvtlg.)

Bei einem Produktionsprozeß von elektrischen Sicherungen möchte man erreichen, dass höchstens 1% der Produktion defekt ist. Dazu kontrolliert man stündlich zehn Sicherungen. Falls nicht alle in Ordnung sind, wird der Prozeß abgebrochen und überprüft.

Frage: Wie groß ist die Wahrscheinlichkeit, dass nicht gestoppt wird, obwohl 2% der Sicherungen defekt sind?

Es liegt $P = B(n, p)$ mit $n = 10$ und $p = 0.02$ zugrunde. Gesucht ist $P\{0 \text{ defekte Stücke}\} = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n = 0.98^{10} = 0.82$.

4.7 Beispiel. (Zur hypergeometrischen Vtlg.)

Eine Lieferung vom Umfang $N = 1000$ darf laut Liefervertrag höchstens 2% Ausschluß haben.

Vorgeschlagenes Prüfverfahren: Die Lieferung wird zurückgewiesen, wenn eine Stichprobe vom Umfang $n = 10$ mindestens ein defektes Stück enthält.

Dann ist

$$\begin{aligned} P\{\text{Zurückweisen}\} &= 1 - P\{\underbrace{\text{Kein Zurückweisen}}_{\text{(Probe intakt)}}\} \\ &= 1 - h(0; N, K, n) \\ &= 1 - h(0; 1000, 20, 10) \\ &= 1 - \frac{\binom{K}{0} \binom{1000-K}{10}}{\binom{1000}{10}} \\ &= 1 - 0,82 = 18\% \text{ falls } K = 20. \end{aligned}$$

Bei diesem Test werden also 18% aller guten Ladungen zurückgewiesen.

4.8 Lemma. Ist n klein gegenüber N, K und $N - K$, so sind die Binomial- und die Hypergeometrische Verteilung nahezu gleich. Es gilt nämlich für eine Folge (p_n) mit $n \cdot p_n \rightarrow \lambda > 0$

$$\lim_{n \rightarrow \infty} \binom{n}{i} p_n^i (1-p_n)^{n-i} = \frac{\lambda^i}{i!} e^{-\lambda} \forall i \in \mathbb{N}_0. \quad (4.5)$$

Wir hatten als Erfahrungstatsache akzeptiert, dass bei unabhängiger Wiederholung eines Zufallsexperiments mit Erfolg (1) und Mißerfolg (0) die relative Häufigkeit der Erfolge gegen die Erfolgswahrscheinlichkeit p konvergiert, d.h. $\frac{\#Erfolge}{n} \approx p$ für großes n .

Dazu zeigen wir zum Abschluß dieses Paragraphen den Prototypen aller sogenannten *Gesetze der großen Zahlen*:

4.9 Satz. (Bernoullisches schwaches Gesetz der großen Zahlen, [BN]:S.99)

Sei P_n die $B(n, p)$ -Vtlg, dann gilt für alle $\varepsilon > 0$

$$P_n \left\{ k : \left| \frac{k}{n} - p \right| \geq \varepsilon \right\} \leq \frac{1}{n\varepsilon^2} p(1-p) \xrightarrow{n \rightarrow \infty} 0. \quad (4.6)$$

5 Diskrete Zufallsvariable und ihre Verteilungen

Seien $\Omega \neq \emptyset$ und $\mathcal{X} \neq \emptyset$ zwei Mengen und $X : \Omega \rightarrow \mathcal{X}$ eine Abbildung. (In [BN]:17 werden Abbildungen mit f, g statt wie hier mit X, Y bezeichnet.) (Abbildungen fassen i.a. Daten zusammen, sie verringern die Beobachtungstiefe durch Vergrößerung.)

5.1 Beispiel. Würfeln mit zwei idealen Würfeln

$\Omega = \underline{6} \times \underline{6}, P = \text{Laplace-Vtlg über } \Omega.$

$X(\omega) = i + j$ (Augensumme), $\Rightarrow \mathcal{X} = \{2, \dots, 12\}.$

$X : \Omega \rightarrow \mathcal{X}.$ Aus $(i + j)$ kann man i.a. nicht (i, j) zurückgewinnen.

Frage: Gibt es ein Zufallsexperiment, das direkt Augensumme beschreibt?

Augensumme = k wird in \mathcal{X} durch $\{k\}$ repräsentiert.

Augensumme = k wird in Ω durch $\{(i, j) : X(i, j) = k\}$ repräsentiert.

Ein neues Maß Q über \mathcal{X} muß deshalb vernünftigerweise die Gleichung

$$Q\{k\} = P\{\omega : X(\omega) = k\} \quad \forall k$$

erfüllen.

5.2 Definition. (Diskrete Zufallsvariable) Sei (Ω, P) ein diskreter W -Raum und \mathcal{X} eine nicht-leere Menge. Jede Abbildung

$$X : \Omega \rightarrow \mathcal{X} \tag{5.1}$$

heißt (diskrete) **Zufallsvariable (ZV)** auf (Ω, P) . Falls $\mathcal{X} \subset \mathbb{R}$, bzw. $\overline{\mathbb{R}}$, bzw. \mathbb{R}^n ist, spricht man von reellen, bzw. numerischen, bzw. n -dimensionalen Zufallsvariablen; um einen Bildraum \mathcal{X} allgemein anzugeben, spricht man oben von der \mathcal{X} -ZV X .

Für $B \subset \mathcal{X}$ repräsentieren die Mengen B und $A := \{\omega \in \Omega : X(\omega) \in B\}$ dasselbe Ereignis, nur in unterschiedlichen Räumen. Wir nennen Teilmengen $A \subset \Omega$, die wie oben definiert sind, *durch X beschreibbare Ereignisse*.

Schreibweise:

$$\{\omega \in \Omega : X(\omega) \in B\} =: \{X \in B\} =: X^{-1}(B)$$

und nennt diese Menge **Urbildmenge**. Es gilt offenbar

$$\omega \in X^{-1}(B) \Leftrightarrow X(\omega) \in B.$$

Achtung: X^{-1} ist keine Umkehrabbildung, sondern bildet Mengen in Mengen ab:

$$\begin{aligned} X^{-1} : \mathcal{P}(\mathcal{X}) &\rightarrow \mathcal{P}(\Omega) \\ B &\rightarrow X^{-1}(B) = \{X \in B\} \end{aligned}$$

Nur wenn X injektiv ist mit üblicher Inversen $\bar{X}: X(\Omega) \rightarrow \Omega$ gilt für $x \in X(\Omega) : \{\bar{X}^{-1}(x)\} = X^{-1}\{x\}$.

Die Mengenabbildung $X^{-1} : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{P}(\Omega)$ ist *operationstreu*, d.h. es gilt für $B, B_i \subset \mathcal{X}, i \in I =$ beliebige Indexmenge:

$$\begin{aligned} X^{-1}(\bigcap_{i \in I} B_i) &= \bigcap_{i \in I} X^{-1}(B_i), \\ X^{-1}(\bigcup_{i \in I} B_i) &= \bigcup_{i \in I} X^{-1}(B_i), \\ X^{-1}(B^c) &= (X^{-1}(B))^c, \\ B_1 \cap B_2 = \emptyset &\Rightarrow X^{-1}(B_1) \cap X^{-1}(B_2) = \emptyset, \\ X^{-1}(\emptyset) = \emptyset, \quad X^{-1}(\mathcal{X}) &= \Omega. \end{aligned} \tag{5.2}$$

Aus der Operationstreue folgt sofort, dass durch

$$P^X(B) = P(X^{-1}(B)) \quad \forall B \subset \mathcal{X} \tag{5.3}$$

ein W-Maß P^X auf \mathcal{X} definiert wird. Man nennt P^X das von X und P induzierte W-Maß oder die **Verteilung** von X unter P , kurz: $\mathcal{L}(X)$ oder $L(X)$, \mathcal{L} wie Law.

Ist T der Träger von P , so ist $X(T)$ der Träger von P^X . Die Z-Dichte von P^X nennt man auch einfach Z-Dichte von X .

P^X ist tatsächlich ein W-Maß über $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$, denn: $P^X(\mathcal{X}) = P(X^{-1}(\mathcal{X})) = 1, P^X(B) \geq 0, P^X(\sum B_i) = P(X^{-1}(\sum B_i)) = P(\sum X^{-1}(B_i)) = \sum P(X^{-1}(B_i)) = \sum P^X(B_i)$, Rest trivial.

□

Ist Q ein W-Maß mit $Q = P^X$, so sagt man: X ist gemäß Q verteilt, kurz: $X \sim Q$. Zu jedem diskreten W-Maß Q auf \mathcal{X} gibt es eine diskrete ZV X mit $X \sim Q$, nämlich $X = id_\Omega$, die Identität auf dem diskreten W-Raum $(\Omega, P) := (\mathcal{X}, Q)$. Dann gilt nämlich $P^X(B) = P\{\omega : X(\omega) \in B\} = P\{\omega : \omega \in B\} = Q(B)$.

Viele interessante zufällige Experimente verlaufen in einer endlichen Anzahl von i.a. zeitlichen Stufen. Ist \mathcal{X}_i der Grundraum der i -ten Stufe, $1 \leq i \leq n$, so ist $\mathcal{X} := \prod_{i=1}^n \mathcal{X}_i$ ein natürlicher Grundraum für das gesamte zufällige Experiment. Die Abbildung

$$\pi_i : \mathcal{X} \rightarrow \mathcal{X}_i \tag{5.4}$$

$$\pi_i(x) := x_i, \quad x = (x_1, \dots, x_n), \tag{5.5}$$

heißt **Projektion** von \mathcal{X} nach \mathcal{X}_i ; sie beschreibt innerhalb des Gesamtexperiments, das i -te Einzelexperiment. Offenbar gilt

$$\pi_i^{-1}(B_i) = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_{i-1} \times B_i \times \mathcal{X}_{i+1} \times \dots \times \mathcal{X}_n. \tag{5.6}$$

Ist P ein diskretes W-Maß auf einem Produktraum $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$, so nennt man die Verteilungen der Projektionen π_i die *eindimensionalen Randverteilungen* von P .

Betrachtet man auf einem allgemeinen diskreten Grundraum (Ω, P) mehrere Abbildungen $X_i :$

$\Omega \rightarrow \mathcal{X}_i, 1 \leq i \leq n$, so ist die Produktabbildung $X := (X_1, \dots, X_n)$ eine diskrete ZV mit Werten in $\mathcal{X} := \prod_{i=1}^n \mathcal{X}_i$. Offenbar gilt für die Projektionen π_i die Eigenschaft $\pi_i \circ X = X_i$. $\mathcal{L}(X) = P^X$ heißt **gemeinsame Verteilung** der X_i und die zugehörige Z-Dichte heißt **gemeinsame Z-Dichte** der X_i . Außerdem ist die i -te Randverteilung von $P^{(X_1, \dots, X_n)}$ gerade P^{X_i} :

$$(P^{(X_1, \dots, X_n)})^{\pi_i} = P^{X_i} \quad \forall i.$$

5.3 Satz. Ist f die gemeinsame Z-Dichte der \mathcal{X}_i -ZV $X_i, 1 \leq i \leq n$, so gilt für $B \subset \mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$

$$P\{X \in B\} = \sum_{x \in B \cap X(T)} f(x), \quad x = (x_1, \dots, x_n), \quad (5.7)$$

wobei T der Träger des diskreten Maßes P auf Ω ist. Bei $B = \prod_{i=1}^n B_i$ gilt

$$P\{X \in \prod_{i=1}^n B_i\} = \sum_{x_1 \in B_1 \cap X_1(T)} \sum_{x_2 \in B_2 \cap X_2(T)} \cdots \sum_{x_n \in B_n \cap X_n(T)} f(x_1, \dots, x_n). \quad (5.8)$$

Die Z-Dichte von X_i ergibt sich durch Aufsummieren aller von i verschiedenen Komponenten zu

$$x_i \rightarrow \sum_{x_1 \in X_1(T)} \cdots \sum_{x_{i-1} \in X_{i-1}(T)} \sum_{x_{i+1} \in X_{i+1}(T)} \cdots \sum_{x_n \in X_n(T)} f(x_1, \dots, x_n). \quad (5.9)$$

5.4 Lemma. Gegeben sei ein diskreter W -Raum (Ω, P) und Abbildungen X und Y gemäß

$$\Omega \xrightarrow{X} \mathcal{X} \xrightarrow{Y} \mathcal{Y}.$$

Dann ist $Y \circ X : \Omega \rightarrow \mathcal{Y}$ eine diskrete ZV, und es gilt

$$(P^X)^Y = P^{(Y \circ X)}, \quad (5.10)$$

was sich direkt aus der folgenden Gleichung ergibt.

$$X^{-1}(Y^{-1}(C)) = (Y \circ X)^{-1}(C) \quad \forall C \subset \mathcal{Y} \quad (5.11)$$

5.5 Beispiel. (Summenverteilung \mathbb{Z} -wertiger ZV) Seien X_1, X_2 zwei diskrete \mathbb{Z} -ZV mit gemeinsamer Z-Dichte f . Dann gilt

$$P\{X_1 + X_2 = k\} = \sum_i f(i, k - i) \quad \forall k \in \mathbb{Z} \quad (5.12)$$

6 Elementare bedingte Wahrscheinlichkeit und stochastische Unabhängigkeit

6.1 Elementare bedingte Wahrscheinlichkeit

In Anwendungen tritt häufig die Situation auf, dass die Wahrscheinlichkeit für A , unter der Bedingung, dass B eingetreten ist, angegeben werden soll.

Beispiel: Wie groß ist die Wahrscheinlichkeit, dass eine Glühbirne noch 100 Stunden brennt, wenn sie bereits 24 Stunden gebrannt hat?

Wie groß ist die Wahrscheinlichkeit, bei zwei unabhängigen Würfeln eine Augensumme > 7 zu würfeln, wenn der erste Wurf < 4 war?

Man kann sich an relativen Häufigkeiten orientieren: Wird ein Versuch n -mal unabhängig durchgeführt mit den Ergebnissen $\omega_1, \omega_2, \dots$, so gilt:

$$\begin{aligned} \text{rel. Häufigkeit von } A \text{ nach Eintreten von } B &= \frac{\text{Anzahl der } i \in \underline{n} \text{ mit } \omega_i \in A \cap B}{\text{Anzahl der } i \in \underline{n} \text{ mit } \omega_i \in B} \\ &= \frac{H_n(A \cap B)}{H_n(B)} \\ &\xrightarrow{\text{Erfahrung}} \frac{P(A \cap B)}{P(B)} \end{aligned}$$

6.1 Definition. [BN]:6.1 Sei (Ω, P) ein diskreter W-Raum. Dann ist für $A, B \subset \Omega$ mit $P(B) > 0$ die **bedingte Wahrscheinlichkeit** von A gegeben B , kurz $P(A|B)$ definiert durch

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0. \quad (6.1)$$

Falls $P(B) = 0$ ist, so ist auch $P(A|B) := 0$ (Konvention in Stochastik: $\frac{0}{0} = 0$).

6.2 Beispiel. [BN]:6.2 *Wurf mit zwei Würfeln:* $\Omega = \underline{6} \times \underline{6}, |\Omega| = 36, P\{\omega\} = \frac{1}{36}$.

$$\begin{aligned} A &:= \{(i, j) | i + j > 7\} \\ B &:= \{(i, j) | i < 4\} \\ A \cap B &= \{(2, 6), (3, 5), (3, 6)\} \\ B &= \{(1, 1), \dots, (1, 6), (2, 1), \dots, (2, 6), (3, 1), \dots, (3, 6)\} = \underline{3} \times \underline{6} \\ P(A|B) &= \frac{3/36}{18/36} = \frac{1}{6} \end{aligned}$$

6.3 Beispiel. *Zweimaliger Münzwurf: (Kopf=1, Zahl=0)* $\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}, |\Omega| = 4, P\{\omega\} = \frac{1}{4}$.

Frage: *Wie groß ist die Wahrscheinlichkeit für zweimal Kopf, wenn bereits der erste Wurf das Ergebnis Kopf hatte?*

$$\begin{aligned} A &= \{(1, 1)\} \\ B &= \{(1, 1), (1, 0)\} \\ P(A|B) &= \frac{1/4}{2/4} = \frac{1}{2} \end{aligned}$$

Frage: *Wie groß ist die Wahrscheinlichkeit für zweimal Kopf, wenn mindestens einmal Kopf geworfen wurde?*

$$\begin{aligned}
A &= \{(1, 1)\} \\
B &= \{(1, 1), (1, 0), (0, 1)\} \\
P(A|B) &= \frac{1/4}{3/4} = \frac{1}{3}
\end{aligned}$$

6.4 Lemma. [BN]:6.3 Falls $P(B) > 0$, so ist $(\Omega, P(\cdot|B))$ ein diskreter W-Raum mit $P(B^c) = 0$. (Wahrscheinlichkeit wird auf B eingeschränkt und normiert.)

6.2 Einige Rechenregeln für bedingte Wahrscheinlichkeiten

Sind B_1, B_2, \dots endlich oder abzählbar unendlich viele, paarweise disjunkte Teilmengen von Ω mit $\sum_i B_i = \Omega$, so gilt

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i) \quad (\text{Satz von der totalen W., [BN]:6.4}) \quad (6.2)$$

Seien $A_1, \dots, A_n \subset \Omega$. Dann gilt die **Multiplikationsregel**:

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}) \quad (6.3)$$

Seien B_1, B_2, \dots paarweise disjunkt $\subset \Omega$ mit $\sum_{i=1}^{\infty} B_i = \Omega$. Dann gilt für $A \subset \Omega$ mit $P(A) > 0$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}. \quad (\text{Bayes-Formel, [BN]:6.6}) \quad (6.4)$$

Beweis:

Zähler rechts = $P(A \cap B_k)$.

Nenner rechts = $P(A)$ dem Satz von der totalen Wahrscheinlichkeit. \square

Beispiele in der Vorlesung:

Die Paradoxa von Galton und von Simpson.

(6.3) ist besonders geeignet, wenn ein Experiment in mehreren Stufen abläuft. Dann kann man die rechte Seite von (6.3) heranziehen, um die linke Seite zu berechnen.

6.5 Beispiel. [BN]:6.8 Gegeben seien fünf Urnen $1, \dots, 5$ mit je zehn Kugeln.

Urne i enthält i rote und $10 - i$ schwarze Kugeln.

Experiment: Zunächst wähle eine Urne zufällig aus. Ziehe aus dieser Urne eine Kugel.

Frage:

1. Wie groß ist die Wahrscheinlichkeit für $A = \{ \text{Kugel ist rot} \}$?
2. Wie groß ist $P(B_k|A)$ mit $B_k = \{ \text{Kugel stammt aus Urne } k \}$?

Antwort:

$$1. P(A) = \sum_{i=1}^5 \underbrace{P(A|B_i)}_{\frac{1}{10}} \underbrace{P(B_i)}_{\frac{1}{5}} = \frac{5 \cdot 6/2}{50} = \frac{3}{10}.$$

$$2. P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)} = \frac{\frac{k}{10} \cdot \frac{1}{5}}{\frac{3}{10}} = \frac{k}{15}.$$

6.6 Beispiel. (Hübner, Bsp. 2.14) Bei der Übertragung von Daten komme nach dem Senden des Zeichens i , dargestellt durch das Ereignis B_i , mit W. $P(A_j|B_i)$ das Zeichen j an, dargestellt durch A_j . Aus dem Empfang von j und den als bekannt angenommenen Wahrscheinlichkeiten $P(B_i)$ und $P(A_j|B_i)$ ergibt sich mit der Bayes-Formel $P(B_j|A_j)$, also die W., daß bei Empfang von j auch tatsächlich j gesendet wurde.

6.3 (Stochastische) Unabhängigkeit von Ereignissen

6.7 Definition. [BN]:6.11 Seien (Ω, \mathcal{A}, P) ein W-Raum und $A, B \in \mathcal{A}$. Dann heißen A und B **stochastisch unabhängig**, wenn gilt:

$$P(B) = 0 \text{ oder } P(A|B) = P(A) \quad \text{falls} \quad P(B) > 0 \quad (6.5)$$

oder äquivalent

$$P(A) = 0 \text{ oder } P(B|A) = P(B) \quad \text{falls} \quad P(A) > 0 \quad (6.6)$$

oder äquivalent

$$P(A \cap B) = P(A)P(B). \quad (\text{symmetrisch in } A \text{ und } B) \quad (6.7)$$

Verallgemeinerung:

6.8 Definition. [BN]:6.12 Sei $I \neq \emptyset$ eine beliebige Indexmenge. Eine Familie $\{A_i : i \in I\}$ von Elementen aus \mathcal{A} heißt **stochastisch unabhängig**, wenn für jede endliche Teilmenge $\emptyset \neq J \subset I$ gilt:

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i). \quad (6.8)$$

6.4 (Stochastische) Unabhängigkeit von diskreten Zufallsvariablen

Sei (Ω, P) ein diskreter W-Raum und seien $X_i : \Omega \rightarrow \mathcal{X}_i, 1 \leq i \leq n$, ZV, deren Produktabbildung $X = (X_1, \dots, X_n)$ die Zähldichte $f(x_1, \dots, x_n)$ besitzt.

6.9 Definition. Seien X_1, \dots, X_n ZV auf einem diskreten W-Raum (Ω, P) und $f(x_1, \dots, x_n)$ Z-Dichte von $X = (X_1, \dots, X_n)$, sowie $f_i(x_i)$ Z-Dichte von X_i . Dann sind X_1, \dots, X_n genau dann **stochastisch unabhängig**, wenn gilt

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \quad \forall x_i \in \mathcal{X}_i, \quad (6.9)$$

wobei die Gleichheit nur für Trägerpunkte (x_1, \dots, x_n) von X nachzuprüfen ist, da sie sonst mit $0 = 0$ trivialerweise gilt.

Durch Summation über die x_i erhält man: Die ZV X_1, \dots, X_n sind genau dann stochastisch unabhängig, wenn gilt

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = \prod_{i=1}^n P\{X_i \in B_i\} \quad (6.10)$$

$$\forall B_i \subset \mathcal{X}_i, i = 1, \dots, n.$$

6.10 Bemerkung. Setzt man in (6.10) $B_i = \mathcal{X}_i$ für $i \in J$ mit einer Teilmenge J von $I := \{1, \dots, n\}$, so sieht man, dass jede Teilfamilie einer unabhängigen Familie von ZV'n wieder unabhängig ist. Dies motiviert die Definition der stochastischen Unabhängigkeit für Familien $(X_i : i \in I)$ mit beliebiger, nicht leerer Indexmenge I , nämlich dadurch, dass man für alle Familien $(X_i : i \in J)$ mit endlichen Teilmengen J von I die Unabhängigkeit fordert.

Die rechte Seite in (6.9) ist, wie man sofort nachrechnet Zähl-dichte eines Maßes auf $\prod_{i=1}^n \mathcal{X}_i$, das wegen der Produktform den Namen **Produktmaß** trägt. Wir definieren allgemein diskrete Produktmaße.

6.11 Definition. Seien (\mathcal{X}_i, P_i) , $1 \leq i \leq n$, diskrete W -Räume mit Zähl-dichten $f_i : \mathcal{X}_i \rightarrow \mathbb{R}_+$ und Trägern T_i . Dann ist die Funktion $f : \prod_{i=1}^n \mathcal{X}_i \rightarrow \mathbb{R}_+$ mit

$$f(x_1, \dots, x_n) := \prod_{i=1}^n f_i(x_i) \quad x_i \in T_i \forall i \quad (6.11)$$

und $f(x_1, \dots, x_n) := 0$ sonst, Zähl-dichte eines W -Maßes P auf $\mathcal{X} := \prod_{i=1}^n \mathcal{X}_i$ mit Träger $T := \prod_{i=1}^n T_i$. Das zugehörige W -Maß heißt **Produktmaß** der P_i und wird mit $\bigotimes_{i=1}^n P_i$ bezeichnet. Es gilt dann

$$\left(\bigotimes_{i=1}^n P_i\right)\left(\prod_{i=1}^n B_i\right) = \sum_{x_1 \in B_1 \cap T_1} \cdots \sum_{x_n \in B_n \cap T_n} \prod_{i=1}^n f_i(x_i) = \prod_{i=1}^n P_i(B_i). \quad (6.12)$$

Wir geben nun eine äquivalente Form der stochastischen Unabhängigkeit, die im allgemeinen maßtheoretischen Rahmen üblicherweise als Definition herangezogen werden wird.

6.12 Lemma. In der Situation von Definition 6.9 sind X_1, \dots, X_n genau dann stochastisch unabhängig, wenn

$$P^{(X_1, \dots, X_n)}\left(\prod_{i=1}^n B_i\right) = \left(\bigotimes_{i=1}^n P^{X_i}\right)\left(\prod_{i=1}^n B_i\right) \quad (6.13)$$

$$\forall B_i \subset \mathcal{X}_i, i = 1, \dots, n.$$

Betrachtet man in der Situation von Definition 6.11 den diskreten W -Raum $(\Omega, P) = \left(\prod_{i=1}^n \mathcal{X}_i, \bigotimes_{i=1}^n P_i\right)$ und definiert auf (Ω, P) ZV $X_i := \pi_i \forall i$, so haben diese die Verteilung P_i und sind stochastisch

unabhängig.

Wichtig: Man kann also zu vorgegebenen Randverteilungen einen W-Raum konstruieren und darauf stochastisch unabhängige ZV X_i mit $X_i \sim P_i$.

Anschaulich offensichtliche Rechenregeln für stochastisch unabhängige ZV sind in folgendem Satz zusammengestellt.

6.13 Satz. [BN]:21.9 Seien X_1, \dots, X_n stu. über einem diskreten W-Raum (Ω, P) mit Werten in \mathcal{X}_i und $I_1 + \dots + I_k = \{1, \dots, n\}$, $E_j := \bigtimes_{i \in I_j} \mathcal{X}_i$, $Y_j := (X_i : i \in I_j)$, $j \in \underline{k}$. Dann gilt:

1. Y_1, \dots, Y_k sind stu. (bzgl. P).
2. Seien $F_j, j \in \underline{k}$ weitere Räume und seien $f_j : E_j \rightarrow F_j$ gegeben, so sind auch die ZV $f_j(Y_j), 1 \leq j \leq k$ stu. (bzgl. P).

Beispielsweise gilt: Sind X, Y, Z stu. so auch X^2 und $e^Y + Z$.

Besonders wichtig sind Summenverteilungen von stu ZV. Aus (5.12) und (6.9) folgt

6.14 Satz. Seien X_1 und X_2 stu diskrete \mathbb{Z} -ZV mit Zähldichten f_1 bzw. f_2 . Dann hat $X_1 + X_2$ die Zähldichte

$$k \rightarrow \sum_{i \in \mathbb{Z}} f_1(i) f_2(k - i) = \sum_{i \in \mathbb{Z}} f_2(i) f_1(k - i). \quad (6.14)$$

Man bezeichnet diese Z-Dichte als **Faltung** $f_1 * f_2$ der Z-Dichten f_1 und f_2 . Das dazugehörige W-Maß heißt die **Faltung** der zu f_1 und f_2 gehörigen W-Maße.

6.15 Beispiel. Sind X_1, \dots, X_n stu mit $X_i \sim B(1, p)$, $p \in (0, 1)$, so gilt

$$\sum_{i=1}^n X_i \sim B(n, p). \quad (6.15)$$

Den Zusammenhang der stochastischen Unabhängigkeit von Mengen und ZV'n wird durch das folgende Lemma gegeben.

6.16 Lemma. Eine Familie $(A_i : 1 \leq i \leq n)$ von Ereignissen in einem diskreten W-Raum (Ω, P) ist genau dann stochastisch unabhängig, wenn die Familie $(1_{A_i} : 1 \leq i \leq n)$ stochastisch unabhängig ist.

Nach (6.10) ist diese Aussage äquivalent zu

$$P\left(\bigcap_{i=1}^n D_i\right) = \prod_{i=1}^n P(D_i) \text{ mit } D_i \in \{A_i, A_i^c\} \forall i. \quad (6.16)$$

7 Erwartungswert und Varianz von diskreten Zufallsvariablen bzw. von ihren Verteilungen

Betrachten wir folgendes Beispiel: Bei einem Glücksspiel mit r Ausgängen x_1, \dots, x_r (z.B. Würfel: $x_i = 1/6$, $1 \leq i \leq 6$,) werde beim Ausgang x_i eine Zahlung von x_i DM geleistet; p_i sei die W. für das Eintreten von x_i . Dann ist die Größe

$$x_1 p_1 + \dots + x_r p_r \quad (7.1)$$

intuitiv der *erwartete* oder *mittlere* Gewinn. Dies führt zu folgender

7.1 Definition. Sei X eine reelle ZV über einem diskreten W -Raum (Ω, P) und P habe den Träger T . Ist die Reihe

$$EX = E_P(X) = \sum_{x \in X(T)} x \cdot P\{X = x\} = \sum_{x \in X(T)} x \cdot P^X\{x\} \quad (7.2)$$

absolut konvergent, so heißt ihr Wert der **Erwartungswert** (kurz: *E-Wert*) von X und X heißt **integrabel**. Die letzte Gleichheit in (7.2) zeigt unter Beachtung, dass $X(T)$ der Träger von P^X ist, dass der Erwartungswert EX nur von der Verteilung P^X abhängt. Man kann deshalb auch vom Erwartungswert von P^X sprechen und genauso vom Erwartungswert EQ eines diskreten W -Maßes Q über \mathbb{R} mit Träger T_Q :

$$EQ = \sum_{x \in T_Q} x \cdot Q\{x\}.$$

Man sieht leicht unter Verwendung der Tatsache, dass absolut konvergente Reihen beliebig umgeordnet werden dürfen, die folgende Gleichheit.

7.2 Satz. (Transformationsformel) Mit den Bezeichnungen aus Definition 7.1 gilt Ist $X : \Omega \rightarrow \mathcal{X}$ eine diskrete ZV und $g : \mathcal{X} \rightarrow \mathbb{R}$ eine weitere Abbildung, so gilt, falls $g \circ X$ integrabel ist

$$Eg \circ X = \sum_{t \in g(X(T))} t \cdot P^{g \circ X}\{t\} = \sum_{x \in X(T)} g(x) \cdot P^X\{x\} \quad (7.3)$$

Die Menge der auf (Ω, P) definierten ZV, für die die Reihe (7.2) absolut konvergiert, wird mit $\mathcal{L}^1(P)$, oder wenn P klar ist, mit \mathcal{L}^1 bezeichnet. Auf diesem Raum mit der üblichen Addition und Multiplikation für reelle Abbildungen (nämlich punktweise) ist der Erwartungswert-Operator **linear**, d.h.

$$\begin{aligned} E(X_1 + X_2) &= EX_1 + EX_2 \\ E(aX) &= aEX \text{ für } a \in \mathbb{R} \end{aligned} \quad (7.4)$$

und **isoton**, d.h.

$$X_1 \leq X_2 \Rightarrow EX_1 \leq EX_2, \quad (7.5)$$

wobei $EX_1 = EX_2$ unter der Annahme $X_1 \leq X_2$ genau dann gilt, wenn **fast sicher** $X_1 = X_2$ gilt, d.h. bei $P\{X_1 = X_2\} = 1$.

7.3 Beispiel. a) Sei $X = 1_A$ die Indikatorfunktion von $A \subset \Omega$. Dann gilt

$$E1_A = PA \quad (7.6)$$

Insbesondere gilt mit $A = \Omega$, dass $E1 = 1$, also nach (7.4) auch $Ea = a$ für alle Konstanten a .

b) X sei eine Laplace verteilte ZV mit Träger $\{1, \dots, n\}$. Dann gilt

$$EX = \frac{n+1}{2} \quad (7.7)$$

c) Sei $X \sim B(n, p)$. Dann gilt

$$EX = np \quad (7.8)$$

d) Sei $X \sim \mathcal{P}(\lambda)$. Dann gilt

$$EX = \lambda \quad (7.9)$$

7.4 Bemerkung. Die Lage des E -Wertes kann man sich veranschaulichen, indem man sich einen gewichtslosen Stab vorstellt, bei dem an den Stellen x die Massen $P\{X = x\}$ angebracht sind. Der E -Wert ist dann diejenige Stelle, an der man den Stab unterstützen muß, damit er im Gleichgewicht ist.

Sei nun $\mathcal{L}^2(P)$ die Menge aller reellen ZV über dem diskreten W-Raum (Ω, P) mit Träger T von P , für die außerdem

$$E(X^2) = \sum_{x \in X(T)} x^2 P^X\{x\} \quad (7.10)$$

konvergiert, d.h. $X \in \mathcal{L}^2 \Leftrightarrow X^2 \in \mathcal{L}^1$. Die Zahl $E(X^2)$ heißt auch das **zweite Moment** von X . Aus $(x+y)^2 \leq 2(X^2+Y^2)$ folgt, dass \mathcal{L}^2 einen linearen Raum darstellt, und aus $|xy| \leq (x^2+y^2)/2$ ergibt sich

$$X, Y \in \mathcal{L}^2 \Rightarrow XY \in \mathcal{L}^1, \quad (7.11)$$

und aus $|x| \leq x^2 + 1$ folgt, dass $\mathcal{L}^2 \subset \mathcal{L}^1$. Insbesondere existiert EX für $X \in \mathcal{L}^2$. Man sieht, dass

$$\langle X, Y \rangle := E(X \cdot Y) \quad (7.12)$$

bilinear, und positiv semidefinit ist, d.h. $\langle X, X \rangle \geq 0$, und dass aus $\langle X, X \rangle = 0$ $X = 0$ fast sicher folgt, d.h. $P\{X = 0\} = 1$.

Damit ist $\langle \cdot, \cdot \rangle$ fast ein inneres Produkt. Der von den konstanten Funktionen aufgespannte Unterraum von \mathcal{L}^2 sei $[1]$ und der darauf senkrecht stehende Teilraum $[1]^\perp = \{X \in \mathcal{L}^2 : \langle X, 1 \rangle = EX = 0\}$. Man kann dann X in zwei orthogonale Anteile aufspalten

$$X = EX + (X - EX). \quad (7.13)$$

Dabei ist EX sozusagen der konstante Anteil von X und $X - EX$ ist die Schwankung um den konstanten Anteil. Wir definieren die Varianz von X als die erwartete quadratische Schwankung von X um seinen Mittelwert, also

7.5 Definition. Für eine ZV $X \in \mathcal{L}^2(P)$ heißt

$$\text{Var}X = E((X - EX)^2) = \sum_{x \in X(T)} (x - EX)^2 P^X\{x\} \quad (7.14)$$

die **Varianz** von X . Die Definition macht auch Sinn im Fall $X \in \mathcal{L}^1$, möglicherweise aber mit dem Wert ∞ .

Aus (7.13) folgt mit dem Satz des Pythagoras der sog. *Verschiebungssatz von Steiner*

$$E(X^2) = (EX)^2 + \text{Var}X \quad (7.15)$$

7.6 Beispiel. a) X besitze eine Laplace-Verteilung über $\{1, \dots, n\}$. Dann gilt

$$\text{Var}X = \frac{n^2 - 1}{12}. \quad (7.16)$$

b) Für $X \sim B(n, p)$ gilt

$$\text{Var}X = np(1 - p) \quad (7.17)$$

c) Für $X \sim \mathcal{P}(\lambda)$ gilt

$$\text{Var}X = \lambda \quad (7.18)$$

Nun einige Rechenregeln für die Varianz.

7.7 Lemma. Seien $X, Y \in \mathcal{L}^2(P)$. Dann gilt

1. $\text{Var}X \geq 0$
2. $\text{Var}X = 0 \Leftrightarrow X = \text{const.}P - f.s.$
(d.h. $\exists c \in \mathbb{R} : P\{X = c\} = 1$)
3. $\text{Var}(cX) = c^2 \text{Var}X \quad \forall c \in \mathbb{R}$
4. $\text{Var}(X + c) = \text{Var}X \quad \forall c \in \mathbb{R}$
5. $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2(E(XY) - EX \cdot EY).$

7.8 Definition. Es seien $X, Y \in \mathcal{L}^2$. Dann heißt

$$\text{Cov}(X, Y) := E(X - EX)(Y - EY) = EXY - EXEY \quad (7.19)$$

Kovarianz von X und Y (unter P).

In Verallgemeinerung von Lemma 7.7 gilt

7.9 Satz. Seien X_1, \dots, X_n reelle ZV auf dem diskreten W -Raum (Ω, P) , die zu $\mathcal{L}^2(P)$ gehören. Dann existiert $\text{Var}(\sum_{i=1}^n X_i)$, und es gilt

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}X_i + 2 \cdot \sum_{i,j:i < j}^n \text{Cov}(X_i, X_j). \quad (7.20)$$

Die $Cov(X, Y)$ läßt sich als *Maß für den linearen Zusammenhang* von X und Y interpretieren. Es gilt nämlich

7.10 Lemma. *Sei Y ZV auf einem diskreten W-Raum (Ω, P) mit $VarX > 0$ und $VarY > 0$. Mit dem sog. **Korrelationskoeffizienten***

$$\rho := \frac{Cov(X, Y)}{\sqrt{VarX}\sqrt{VarY}} \quad (7.21)$$

gilt:

a) Die Größen $a_0 = EX - \rho \cdot \frac{\sqrt{VarX}}{\sqrt{VarY}} EY$ und $b_0 = \rho \cdot \frac{\sqrt{VarX}}{\sqrt{VarY}}$ erfüllen

$$E((X - (a_0 - b_0 \cdot Y))^2) = \min\{E((X - (a + b \cdot Y))^2) : a, b \in \mathbb{R}\} \quad (7.22)$$

$$= VarX \cdot (1 - \rho^2). \quad (7.23)$$

b) Es gilt $\rho \in [-1, 1]$.

Wir werden jetzt sehen, dass stochastisch unabhängige ZV unkorreliert sind.

7.11 Satz. *Seien $X, Y \in \mathcal{L}^1(P)$ stochastisch unabhängige reelle ZV auf dem (diskreten) W-Raum (Ω, P) . Dann ist $XY \in \mathcal{L}^1(P)$ und es gilt*

$$E(XY) = EX \cdot EY \quad (7.24)$$

Aus Satz 7.9 ergibt sich nun sofort

7.12 Satz. *(Gleichung von Bienaymé) Sind X_1, \dots, X_n stochastisch unabhängige oder paarweise unkorrelierte ZV in $\mathcal{L}^2(P)$ auf einem diskreten W-Raum (Ω, P) . Dann gilt die Gleichung von Bienaymé*

$$Var(X_1 + \dots + X_n) = VarX_1 + \dots + VarX_n. \quad (7.25)$$

8 Das schwache Gesetz der großen Zahlen

Es soll versucht werden die Erfahrungstatsache *Stabilisierung der relativen Häufigkeiten* als Satz wiederzufinden. Das Ergebnis ist das schwache Gesetz der großen Zahlen (englisch: Weak law of large numbers (WLLN)). Anschließend wollen wir das WLLN zur Schätzung verschiedener Kenngrößen von W-Maßen benutzen. Im Folgenden ist (Ω, \mathbb{P}) stets ein diskreter W-Raum. Man benutzt gerne das Symbol \mathbb{P} für das W-Maß auf dem Hintergrundraum Ω , auf dem die ZVn leben, wenn es auf die spezielle Gestalt nicht ankommt. \mathbb{P} kann von Fall zu Fall seine Bedeutung ändern; im Englischen spricht man von einer *generic notation*.

Als Vorbereitung beweisen wir

8.1 Satz. *[BN]:8.1 und 19.10 Sei X eine reelle ZV aus $\mathcal{L}^1(\mathbb{P})$. Dann gilt die **Markov-Ungleichung***

$$\mathbb{P}\{|X| \geq \varepsilon\} \leq \frac{1}{\varepsilon} E(|X|) \quad \forall \varepsilon > 0. \quad (8.1)$$

Ist $X \in \mathcal{L}^2$, so gilt die **Tschebischev-Ungleichung**

$$\mathbb{P}\{|X - EX| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \text{Var}(X). \quad (8.2)$$

Beweis: Zu (8.1):

$$\begin{aligned} \mathbb{P}\{|X| \geq \varepsilon\} &= E(1_{\{|X| \geq \varepsilon\}}) \\ &= \frac{1}{\varepsilon} E(\varepsilon 1_{[\varepsilon, \infty)}(|X|)) \\ &\leq \frac{1}{\varepsilon} E|X| \quad \square \end{aligned}$$

Zu (8.2):

$$\begin{aligned} \mathbb{P}\{|X - EX| \geq \varepsilon\} &= \mathbb{P}\{|X - EX|^2 \geq \varepsilon^2\} \\ &\leq \frac{1}{\varepsilon^2} E(X - EX)^2 \\ &= \frac{1}{\varepsilon^2} \text{Var} X \quad \square \end{aligned}$$

8.2 Satz. (Schwaches Gesetz der großen Zahlen, [BN]:8.2)

Seien X_1, \dots, X_n i.i.d. (independent, identically distributed) oder paarweise unkorrelierte ZV aus \mathcal{L}^2 mit $EX_i = a \in \mathbb{R}$ und $\text{Var} X_i =: \sigma^2 < \infty$. Dann gilt für $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$

$$\mathbb{P}\{|\bar{X}_n - a| \leq \varepsilon\} \geq 1 - \frac{\sigma^2}{n\varepsilon^2} \quad \forall \varepsilon > 0. \quad (8.3)$$

Daraus folgt das **Schwache Gesetz der großen Zahlen**

$$\mathbb{P}\{|\bar{X}_n - a| \leq \varepsilon\} \rightarrow 1 \quad \forall \varepsilon > 0. \quad (8.4)$$

Beweis:

$$\begin{aligned} E\bar{X}_n &= E\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n EX_i = a \\ \text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var} \sum_{i=1}^n X_i \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \\ \mathbb{P}\{|\bar{X}_n - a| < \varepsilon\} &= 1 - \mathbb{P}\{|\bar{X}_n - a| \geq \varepsilon\} = 1 - \mathbb{P}\{|\bar{X}_n - E\bar{X}_n| \geq \varepsilon\} \\ &\geq 1 - \frac{1}{\varepsilon^2} \text{Var}(\bar{X}_n) \quad (\text{wg. (8.2)}) \\ &= 1 - \frac{1}{\varepsilon^2} \frac{\sigma^2}{n} \quad \square \end{aligned}$$

Das schwache Gesetz der großen Zahlen sagt aus, dass der Durchschnitt von ZVn, deren Erwartungswert in gewisser Weise stochastisch approximiert. Wenn der E-Wert a unbekannt ist, kann man in der Situation des WLLN das Stichprobenmittel \bar{X}_n als sog. *Schätzer* oder *Schätzfunktion* für a verwenden und weiß aufgrund des WLLN, dass der Schätzer für große n mit großer W. dicht bei a liegt. Um eine kurze Ausdrucksweise zu haben definieren wir nun

8.3 Definition. [BN]:8.5 X, X_1, X_2, \dots seien \mathbb{R}^k -wertige ZV über dem diskreten W-Raum (Ω, \mathbb{P}) . Falls gilt

$$\mathbb{P}\{|X_n - X| \geq \varepsilon\} \rightarrow 0 \quad \forall \varepsilon > 0, \quad (8.5)$$

so sagt man: $X_n, n \geq 1$ **konvergiert nach Wahrscheinlichkeit** gegen X (n.W. gegen X) oder kurz $X_n \rightarrow X$ n.W., $X_n \xrightarrow{\mathbb{P}} X, \mathbb{P} - \lim X_n = X$.

Bemerkung: Falls $X_n \rightarrow X$ n.W. und $X_n \rightarrow Y$ n.W., dann ist $X = Y$ [P].

Beweis:

$$\begin{aligned} \{|X - Y| > \varepsilon\} &\subset \{|X - X_n| > \frac{\varepsilon}{2}\} \cup \{|Y - X_n| > \frac{\varepsilon}{2}\} \\ \mathbb{P}\{|X - Y| > \varepsilon\} &\leq \underbrace{\mathbb{P}\{|X - X_n| > \frac{\varepsilon}{2}\}}_{\rightarrow 0} + \underbrace{\mathbb{P}\{|Y - X_n| > \frac{\varepsilon}{2}\}}_{\rightarrow 0} \end{aligned}$$

woraus auch $\mathbb{P}\{|X - Y| > \varepsilon\} = 0$ folgt. Damit ist auch

$$\mathbb{P}\{|X - Y| > 0\} = \lim_{n \rightarrow \infty} \left(\mathbb{P}\left\{|X - Y| \geq \frac{1}{n}\right\} \right) = 0$$

also $X = Y$ [P]. \square

Rechenregeln: [BN]:8.8 X, X_1, X_2, \dots und Y, Y_1, Y_2, \dots seien \mathbb{R}^k -wertige ZV über (Ω, \mathbb{P}) . $g : \mathbb{R}^k \rightarrow \mathbb{R}$ sei eine weitere Abbildung. Es gilt: (Übungsaufgabe)

$$g \text{ stetig in } a \in \mathbb{R}^k, X_n \rightarrow a \text{ n.W.} \Rightarrow g(X_n) \rightarrow g(a) \text{ n.W.} \quad (8.6)$$

$$g \text{ stetig auf } \mathbb{R}^k, X_n \rightarrow X \text{ n.W.} \Rightarrow g(X_n) \rightarrow g(X) \text{ n.W.} \quad (8.7)$$

$$X_n \rightarrow X \text{ n.W.}, Y_n \rightarrow Y \text{ n.W.} \Rightarrow X_n + Y_n \rightarrow X + Y \text{ n.W.} \quad (8.8)$$

$$X_n \rightarrow X \text{ n.W.}, Y_n \rightarrow Y \text{ n.W.} \Rightarrow X_n^T \cdot Y_n \rightarrow X^T \cdot Y \text{ n.W.} \quad (8.9)$$

$$\left. \begin{array}{l} X_n \rightarrow X \text{ n.W.}, Y_n \rightarrow Y \text{ n.W.}, \\ \mathbb{P}\{Y_n \neq 0\} = \mathbb{P}\{Y \neq 0\} = 1 \\ Y_n, Y \text{ reellwertige ZV} \end{array} \right\} \Rightarrow \frac{X_n}{Y_n} \rightarrow \frac{X}{Y} \text{ n.W.} \quad (8.10)$$

8.4 Beispiel. [BN]:8.9 X_1, \dots, X_n seien i.i.d. mit $EX_i = a$ und $\sigma^2 = \text{Var}X_i < \infty$.

\bar{X}_n schätzt a (denn $\bar{X}_n \rightarrow a$ n.W.). Gesucht ist ein Schätzer für σ^2 :

$Y_i := (X_i - a)^2, i \in \underline{n}$, sind i.i.d. mit $EW EY_i = E(X_i - a)^2 = \sigma^2$.

Falls $EY_i^2 = E(X_i - a)^4 < \infty$ (eigentlich unnötig) lässt sich das schwache Gesetz der großen Zahlen anwenden:

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n (X_i - a)^2 \rightarrow \sigma^2 \text{ n.W.}$$

Ersetzen von a durch \bar{X}_n und $\frac{1}{n}$ durch $\frac{1}{n-1}$ ergibt:

$$S_{XX,n} := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \text{ (Stichprobenvarianz)} \quad (8.11)$$

Es gilt:

$$ES_{XX,n} = \sigma^2 \quad (8.12)$$

($S_{XX,n}$ ist erwartungstreu) und

$$S_{XX,n} \rightarrow \sigma^2 \text{ n.W.} \quad (8.13)$$

Beweis:

Es gilt:

$$\begin{aligned} S_{XX,n} &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n ((X_i - a) + (a - \bar{X}_n))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i - a)^2 + \frac{1}{n-1} 2 \sum_{i=1}^n (X_i - a)(a - \bar{X}_n) + \frac{1}{n-1} \sum_{i=1}^n (a - \bar{X}_n)^2 \\ &= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n (X_i - a)^2 - (\bar{X}_n - a)^2 \right). \end{aligned}$$

Erwartungswertbildung auf beiden Seiten liefert (8.12):

$$\begin{aligned} (n-1)E(S_{XX,n}) &= n\sigma^2 - n\text{Var}\bar{X}_n = n\sigma^2 - n\frac{\sigma^2}{n} \\ &= (n-1)\sigma^2. \end{aligned}$$

zu (8.13):

$$S_{XX,n} = \underbrace{\frac{n}{n-1}}_{\rightarrow 1} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n (X_i - a)^2}_{\rightarrow \sigma^2 \text{ n.W.}} - \underbrace{2(\bar{X}_n - a)^2}_{\rightarrow 0 \text{ n.W.}} \right) \rightarrow \sigma^2 \text{ n.W.} \quad \square$$

8.5 Beispiel. [BN]:8.10 $(X_1, Y_1), (X_2, Y_2), \dots$ seien i.i.d. aus \mathbb{R}^2 , $X_i, Y_i \in \mathcal{L}^2$. Dann wird

$$S_{XY,n} := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n) \quad (8.14)$$

die **Stichprobenkovarianz** genannt. Außerdem wird als **Stichprobenkorrelationskoeffizient** definiert:

$$R_{XY,n} := \frac{S_{XY,n}}{\sqrt{S_{XX,n} \cdot S_{YY,n}}}. \quad (8.15)$$

Es gilt mit ähnlichem Beweis wie für $S_{XX,n}$:

$$ES_{XY,n} = \text{Cov}(X, Y) \text{ (Erwartungstreue)} \quad (8.16)$$

$$S_{XY,n} \rightarrow \text{Cov}(X, Y) \text{ n.W.} \quad (8.17)$$

Falls $0 < \text{Var} X < \infty$ und $0 < \text{Var} Y < \infty$ so gilt

$$R_{XY,n} \rightarrow \rho(X, Y) \text{ n.W.} \quad (8.18)$$

9 Verteilungskonvergenz und zentraler Grenzwertsatz

[BN]:9 Das schwache Gesetz der großen Zahlen, das wir oben behandelt haben, ist das einfachste Beispiel aus einer ganzen Familie von Gesetzen der großen Zahlen, auf die wir hier jedoch nicht näher eingehen können. Daneben gibt es eine zweite, in Theorie und Praxis äußerst wichtige Familie von wahrscheinlichkeitstheoretischen Resultaten, die *zentralen Grenzwertsätze*, von denen hier der *Satz von Lindeberg* gezeigt werden soll. Als Vorbereitung stellen wir zunächst den Begriff der Verteilungskonvergenz für reelle ZV W -Maße über einem diskreten W -Raum, sowie einige Rechenregeln bereit.

Der gängige Beweis für zentrale Grenzwertsätze verwendet charakteristische Funktionen (Fourier-Transformationen), die uns hier nicht zur Verfügung stehen. E. Bolthausen gab mir den Hinweis auf den hier dargestellten Beweis nach der Steinschen Methode, die ohne weitergehende Hilfsmittel auskommt, und R.-D.Reiss stellte mir seinen Beweis für die Erweiterung vom i.i.d. Fall auf den Lindeberg-Fall zur Verfügung.

Stein, C. (1970): Proc. VIth Berkeley Symp. Math. Stat. Prob. II, 583 – 602.

Barbour, A.D. & Hall, T. (1984): Austr. J. Statist. 26, 8 – 115.

Sei X eine reelle ZV auf einem diskreten W -Raum (Ω, \mathbb{P}) und $P := \mathbb{P}^X$ sei das induzierte diskrete Maß über \mathbb{R} . Durch die Gleichung

$$\mathbb{P}\{X \leq x\} = P(-\infty, x] = F(x), \quad x \in \mathbb{R}, \quad (9.1)$$

wird eine Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ definiert mit den Eigenschaften

$$F(x) \leq F(y) \quad \forall x \leq y, \text{ (monoton steigend)} \quad (9.2)$$

$$F(x+h) \rightarrow F(x) \text{ falls } h \downarrow 0, \text{ (stetig von rechts)} \quad (9.3)$$

$$\begin{cases} F(\infty) := \lim_{n \rightarrow \infty} F(-n) = 0 \\ F(\infty) := \lim_{n \rightarrow \infty} F(n) = 1 \end{cases} \quad \text{(Normierung)}. \quad (9.4)$$

F heißt die **Verteilungsfunktion** (kurz: VF) von X oder von $P = \mathbb{P}^X$. Jede Funktion F mit (9.2)–(9.4) nennt man Verteilungsfunktion.

Wir werden später sehen (ohne allerdings in die dazu nötige Maßtheorie einzusteigen), dass zu jeder VF F ein W -Maß über \mathbb{R} existiert, allerdings nicht mehr auf der gesamten Potenzmenge $\mathcal{P}(\mathbb{R})$ sondern nur auf einer Teilfamilie, der sog. Borelschen σ -Algebra \mathbb{B} , die aber alle in der Praxis vorkommenden Mengen enthält. Zum jetzigen Zeitpunkt ist der Begriff VF nur ein Abkürzung für die obigen Eigenschaften.

9.1 Lemma. [BN]:14.10 Sei F die VF von X . Dann lässt sich aus F das diskrete W -Maß $P = \mathbb{P}^X$ über \mathbb{R} zurückgewinnen. Es gilt nämlich für alle $x, y \in \mathbb{R}$ mit $x \leq y$:

$$P(-\infty, x) = \lim_{n \rightarrow \infty} F\left(x - \frac{1}{n}\right) =: F(x_-) \quad (9.5)$$

$$P\{x\} = F(x) - F(x_-) \quad (9.6)$$

$$P\{x\} = 0 \Leftrightarrow F \text{ stetig in } x \quad (9.7)$$

$$P(x, y] = 0 \Leftrightarrow F(x) = F(y) \quad (9.8)$$

9.2 Beispiel. Die VF des diskreten Maßes P auf \mathbb{R} mit Träger $T \subset \mathbb{R}$ und Zähldichte $f(x) = P\{x\}$ ergibt sich aus $P(B) = \sum_{t \in T \cap B} f(t) = \sum_{t \in T} f(t) 1_B(t)$, $B \in \mathcal{P}(\mathbb{R})$, zu

$$F(x) = P(-\infty, x] = \sum_{t \in T, t \leq x} f(t) = \sum_{t \in T} f(t) 1_{(-\infty, x]}(t) \quad (9.9)$$

9.3 Beispiel. (empirische VF, [BN]:17.16)

Seien X_1, \dots, X_n reellwertige (im Moment noch diskrete, später beliebige) (ZV auf (Ω, P) , die z.B. n -maliges Messen einer physikalischen Größe beschreiben sollen. Für ein festes $\omega \in \Omega$ sei $P_{n, \omega}$ die Laplace-Vtlg über den Realisierungen $X_1(\omega) =: x_1, \dots, X_n(\omega) =: x_n$.

Es gilt also für die zu $P_{n, \omega}$ gehörige VF $\hat{F}_n = \hat{F}_{n, \omega}$

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(x_i). \quad (9.10)$$

Bei festem x ist $\hat{F}_n(x)$ eine (diskrete) ZV.

\hat{F}_n heißt empirische Verteilungsfunktion zu X_1, \dots, X_n . Sind X_1, \dots, X_n stochastisch unabhängig, alle mit derselben Verteilungsfunktion $F(x) := P\{X_1 \leq x\}$. Nach dem WLLN konvergiert \hat{F}_n für $n \rightarrow \infty$ nach W . gegen F .

9.4 Definition. (Verteilungskonvergenz über \mathbb{R})

F und F_n , $n \in \mathbb{N}$, seien VFn über \mathbb{R} . Gilt für die Stetigkeitsmenge $C(F) := \{y \in \mathbb{R} : F \text{ stetig im Punkt } y\}$ die Aussage

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \forall x \in C(F), \quad (9.11)$$

so heißt die Folge $(F_n, n \geq 1)$ verteilungskonvergent gegen F , kurz:

$$F_n \xrightarrow{\mathcal{L}} F.$$

Eine Folge $(X_n, n \geq 1)$ von reellen ZVn heißt verteilungskonvergent gegen F , wenn die VFn F_n von \mathbb{P}^{X_n} nach Verteilung gegen F konvergieren. Man schreibt dann auch $\mathcal{L}(X_n) \xrightarrow{\mathcal{L}} F$. Kurz: $X_n \xrightarrow{\mathcal{L}} F$.

Das \mathcal{L} steht für den englischsprachigen Ausdruck „Law“, der in diesem Zusammenhang „Verteilung“ bedeutet. Die Verteilungskonvergenz einer Folge von ZV X_n gegen eine VF F ist nach obiger Definition lediglich eine *Eigenschaft der zugehörigen induzierten Verteilungen* bzw. der entsprechenden Verteilungsfunktionen. Daher ist es völlig irrelevant, ob die ZVn auf unterschiedlichen W-Räumen (Ω_n, \mathbb{P}_n) definiert sind.

Liegt Verteilungskonvergenz $X_n \xrightarrow{\mathcal{L}} F$ vor, so kann (für große n) die Wahrscheinlichkeit $\mathbb{P}^{X_n}(a, b]$ durch $F(b) - F(a)$ approximiert werden, wenn $a, b \in C(F)$ gilt.

Der Begriff der Verteilungskonvergenz ist schwächer als der Begriff der Konvergenz nach Wahrscheinlichkeit.

9.5 Lemma. (*Konvergenz n.W. impliziert Verteilungskonvergenz*)

$X_n, n \in \mathbb{N}$, und X seien reelle ZV über einem diskreten W-Raum (Ω, \mathbb{P}) mit $X_n \rightarrow X$ n.W. für $\rightarrow \infty$. Sind F_n bzw. F die zugehörigen VFn, so gilt $F_n \xrightarrow{\mathcal{L}} F$.

Beweis: Zu $x \in C(F)$ und $\varepsilon > 0$ gibt es ein $y > 0$ mit den Eigenschaften

$$F(x) - \varepsilon \leq F(x - y) = \mathbb{P}\{X \leq x - y\} \quad \text{und} \quad F(x + y) \leq F(x) + \varepsilon.$$

Nun gilt aber für alle n die Inklusion

$$\{X \leq x - y\} \subset \{X_n \leq x\} \cup \{|X_n - X| > y\},$$

denn aus $X_n > x$ und $|X_n - X| \leq y$ folgt $X = (X - X_n) + X_n > x - y$. Hiermit folgt die Abschätzung

$$F(x) \leq F(x - y) + \varepsilon \leq F_n(x) + y_n + \varepsilon \quad \text{mit} \quad y_n := \mathbb{P}\{|X_n - X| > y\} \rightarrow 0.$$

Entsprechend gilt für alle n die Inklusion

$$\{X_n \leq x\} \subset \{X \leq x + y\} \cup \{|X_n - X| > y\},$$

was die Abschätzung

$$F_n(x) \leq F(x + y) + y_n \leq F(x) + \varepsilon + y_n$$

impliziert. Insgesamt folgt $|F_n(x) - F(x)| \leq \varepsilon + y_n$ mit $y_n = y_n(x, \varepsilon) \rightarrow 0$ und $\varepsilon > 0$ beliebig, also $F_n(x) \rightarrow F(x)$. \square

9.6 Bemerkung. (*Verteilungskonvergenz impliziert i.a. nicht Konvergenz n.W.*) Mit obigen Bezeichnungen gilt: Aus $F_n \xrightarrow{\mathcal{L}} F$ folgt i.a. nicht $X_n \rightarrow X$ n.W. Denn z.B. für $X \sim B(1, 1/2)$ und $X_n := X$, falls n gerade, bzw. $X_n := 1 - X$, falls n ungerade, gilt $\mathcal{L}(X_n) = \mathcal{L}(X) \forall n$, also auch $F_n \xrightarrow{\mathcal{L}} F$, jedoch $\mathbb{P}\{|X_n - X| \geq \varepsilon\} = 1 \not\rightarrow 0$ für alle ungeraden $n \in \mathbb{N}$ und $0 < \varepsilon < 1$. \square

Allerdings sind Verteilungskonvergenz gegen eine Einpunktverteilung und Konvergenz n.W. äquivalent, wenn alle Zufallsvariablen X_n auf demselben W-Raum definiert sind:

9.7 Lemma. (Konvergenz n.W. gegen Konstanten und Verteilungskonvergenz)

Gegeben sei eine Folge von reellen ZVn X_n , $n \in \mathbb{N}$, über dem diskreten W -Raum (Ω, \mathbb{P}) . Für $a \in \mathbb{R}$ (aufgefasst als Abbildung auf Ω) bezeichne δ_a die Einpunktverteilung über \mathbb{R} (Dirac-Maß) mit Träger $\{a\}$ und zugehöriger VF F_a . Dann gilt:

$$X_n \rightarrow a \text{ n.W.} \quad \Leftrightarrow \quad X_n \xrightarrow{\mathcal{L}} F_a. \quad (9.12)$$

Beweis: Es gilt die nachfolgende Kette von Äquivalenzen:

$$\begin{aligned} X_n &\xrightarrow{\text{n.W.}} a \\ \Leftrightarrow \quad \forall \varepsilon > 0 \text{ gilt } &\lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - a| \leq \varepsilon\} = 1 \\ \Leftrightarrow \quad \forall \varepsilon > 0 \text{ gilt } &\lim_{n \rightarrow \infty} \mathbb{P}\{X_n \leq a + \varepsilon\} = 1, \quad \lim_{n \rightarrow \infty} \mathbb{P}\{X_n \geq a - \varepsilon\} = 1 \\ \Leftrightarrow \quad \lim_{n \rightarrow \infty} F_n(x) = 1 \quad \forall x > a, &\quad \lim_{n \rightarrow \infty} F_n(x) = 0 \quad \forall x < a \\ \Leftrightarrow \quad \lim_{n \rightarrow \infty} F_n(x) = 1_{[a, \infty)}(x) = F(x) &\quad \forall x \in C(F) = \mathbb{R} \setminus \{a\}. \quad \square \end{aligned}$$

Wir werden nun ein (hinreichendes) Kriterium für die Verteilungskonvergenz von X_n gegen F beweisen, das einen relativ einfachen Beweis des zentralen Grenzwertsatzes ermöglicht.

9.8 Satz. (Hinreichendes Kriterium für Verteilungskonvergenz)

Seien $X_n, n \in \mathbb{N}$, diskrete reelle ZV mit zugehörigen VFn F_n und F eine VF auf \mathbb{R} , die differenzierbar ist mit Ableitung F'

$$\lim_{n \rightarrow \infty} Ef(X_n) = \int_{-\infty}^{\infty} f(x) \cdot F'(x) dx \quad \forall f \in \mathcal{F} \quad (9.13)$$

mit $\mathcal{F} := \{f : \mathbb{R} \rightarrow \mathbb{R} : f \text{ stetig, } \lim_{x \rightarrow \pm\infty} f(x) \text{ existiert}\}$; offenbar sind alle $f \in \mathcal{F}$ gleichmäßig stetig und beschränkt.

Dann ist die Folge X_n verteilungskonvergent gegen F .

Es sei hier angemerkt, dass die Differenzierbarkeit von F hier nur deshalb benötigt wird, weil uns bis jetzt das allgemeine Maßintegral noch nicht zur Verfügung steht.

Beweis: Sei $g \in \mathcal{F}$ definiert durch

$$g(x) = 1_{(-\infty, 0]}(x) + (1 - x) \cdot 1_{(0, 1)}(x), \quad x \in \mathbb{R}.$$

Mit Hilfe von g definieren wir für jedes $k \in \mathbb{N}$ die Funktion $f_k(x) = g(kx)$, $x \in \mathbb{R}$, die offenbar in \mathcal{F} liegt. Wegen (9.13) gilt dann für alle $x \in \mathbb{R}$ und alle $k \in \mathbb{N}$ einerseits

$$\begin{aligned} \limsup_{n \rightarrow \infty} F_n(x) &= \limsup_{n \rightarrow \infty} E1_{(-\infty, x]}(X_n) \\ &\leq \limsup_{n \rightarrow \infty} Ef_k(X_n - x) = \int f_k(y - x) F'(y) dy \\ &\leq \int 1_{(-\infty, x + \frac{1}{k}]}(y) F'(y) dy = F(x + \frac{1}{k}), \end{aligned}$$

und andererseits

$$\begin{aligned}
 \liminf_{n \rightarrow \infty} F_n(x) &= \liminf_{n \rightarrow \infty} E 1_{(-\infty, x]}(X_n) \\
 &\geq \liminf_{n \rightarrow \infty} E f_k(X_n - x + \frac{1}{k}) \\
 &= \int f_k(y - x + \frac{1}{k}) F^{\vee}(y) dy \\
 &\geq \int 1_{(-\infty, x - \frac{1}{k}]}(y) F^{\vee}(y) dy = F(x - \frac{1}{k}).
 \end{aligned}$$

Ist x ein Stetigkeitspunkt von F , d.h. $x \in C(F)$,

so folgt zusätzlich $F(x + \frac{1}{k}) \rightarrow F(x)$ und $F(x - \frac{1}{k}) \rightarrow F(x)$ für $k \rightarrow \infty$, also $\lim_{n \rightarrow \infty} F_n(x) = F(x)$. \square

Mit Satz 9.8 haben wir ein handliches hinreichendes Kriterium für Verteilungskonvergenz bewiesen. Tatsächlich ist die in Satz 9.8 angegebene Bedingung sogar *äquivalent* zur Verteilungskonvergenz von F_n gegen F .

Die Grenzverteilungsfunktion F im folgenden **zentralen Grenzwertsatz** (englisch: Central Limit Theorem (**CLT**)) ist eine der wichtigsten Funktionen der Wahrscheinlichkeitstheorie, die sogenannte **Normalverteilungsfunktion**, die wir wie allgemein üblich mit Φ statt F bezeichnen und die folgendermaßen definiert ist

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy \tag{9.14}$$

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, y \in \mathbb{R}. \tag{9.15}$$

Die Funktion φ beschreibt die **Gauss'sche Glockenkurve**. Man kann zeigen das Φ eine differenzierbare VF ist mit Ableitung φ .

Alle ZV seien im folgenden diskret. Wir beweisen das CLT unter folgender

9.9 Lindeberg-Bedingung. Für $n \in \mathbb{N}$ gelte: $X_{1,n}, \dots, X_{n,n}$ sind unabhängig mit

$$EX_{i,n} = 0, \quad i = 1, \dots, n,$$

und

$$\sigma_{i,n}^2 := EX_{i,n}^2 \in (0, \infty).$$

Mit $\tau_n^2 := \sum_{i=1}^n \sigma_{i,n}^2$ gelte:

$$\forall \varepsilon > 0 : \lim_{n \in \mathbb{N}} \sum_{i=1}^n E \left(\left(\frac{X_{i,n}}{\tau_n} \right)^2 1_{\left\{ \frac{|X_{i,n}|}{\tau_n} \geq \varepsilon \right\}} \right) = 0.$$

Den obigen Summenausdruck, über den der Grenzwert $\lim_{n \in \mathbb{N}}$ gebildet wird, heißt *Lindeberg Ausdruck*

9.10 Satz. (Zentraler Grenzwertsatz) Unter der Lindeberg-Bed. 9.9 gilt:

$$\forall t \in \mathbb{R} : P \left\{ \frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \leq t \right\} \xrightarrow{n \rightarrow \infty} \Phi(t).$$

9.11 Lemma. Sei h eine differenzierbare Funktion, sodass h' gleichmäßig stetig und beschränkt ist. Unter der Lindeberg-Bed.9.9 gilt für

$$s(x) := h'(x) - xh(x)$$

folgendes:

$$E \left(s \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) \rightarrow 0, \quad n \rightarrow \infty. \quad (9.16)$$

9.12 Bemerkung. Es gilt $\int s(x)\varphi(x) dx = 0$.

Beweis zu Lemma 9.11: Setze

$$Z_{i,n} = \sum_{j=1, j \neq i}^n X_{j,n}.$$

Damit ergibt sich

$$\begin{aligned} & E \left(s \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) \\ &= E \left(h' \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) - \sum_{i=1}^n E \left(\frac{X_{i,n}}{\tau_n} h \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) \end{aligned}$$

N.R.: Mittels einer Taylor-Entwicklung erhält man für $\vartheta = \vartheta(X_{i,1}, \dots, X_{i,n})$:

$$\begin{aligned} & h \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) = h \left(\frac{Z_{i,n}}{\tau_n} + \frac{X_{i,n}}{\tau_n} \right) \\ &= h \left(\frac{Z_{i,n}}{\tau_n} \right) + h' \left(\frac{Z_{i,n}}{\tau_n} + \vartheta \frac{X_{i,n}}{\tau_n} \right) \frac{X_{i,n}}{\tau_n} \\ &= h \left(\frac{Z_{i,n}}{\tau_n} \right) + h' \left(\frac{Z_{i,n}}{\tau_n} \right) \frac{X_{i,n}}{\tau_n} \\ &\quad + \left(h' \left(\frac{Z_{i,n}}{\tau_n} + \vartheta \frac{X_{i,n}}{\tau_n} \right) - h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \frac{X_{i,n}}{\tau_n} \\ &= E \left(h' \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) - \sum_{i=1}^n \frac{\sigma_{i,n}^2}{\tau_n^2} E \left(h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \\ &\quad - \sum_{i=1}^n E \left(\left(h' \left(\frac{Z_{i,n}}{\tau_n} + \vartheta \frac{X_{i,n}}{\tau_n} \right) - h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \left(\frac{X_{i,n}}{\tau_n} \right)^2 \right) \\ &=: A_n - B_n, \end{aligned}$$

da $Z_{i,n}$ und $X_{i,n}$ unabhängig sind und $EX_{i,n} = 0$ gilt.

Wir zeigen: $B_n \rightarrow 0$, $n \rightarrow \infty$:

Sei $\varepsilon > 0$, sodaß für alle x, y , mit $|y| \leq \varepsilon$

$$|h'(x+y) - h'(x)| \leq \delta$$

und $b = \sup_{x \in \mathbb{R}} |h'(x)|$.

Dann folgt:

$$\begin{aligned} |B_n| &= \left| \sum_{i=1}^n E \left(\left(h' \left(\frac{Z_{i,n}}{\tau_n} + \vartheta \frac{X_{i,n}}{\tau_n} \right) - h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \left(\frac{X_{i,n}}{\tau_n} \right)^2 \right) \right| \\ &\leq \left| \sum_{i=1}^n E \left((\dots) \left(\frac{X_{i,n}}{\tau_n} \right)^2 1_{\left\{ \left| \frac{X_{i,n}}{\tau_n} \right| < \varepsilon \right\}} \right) \right. \\ &\quad \left. + \sum_{i=1}^n E \left((\dots) \left(\frac{X_{i,n}}{\tau_n} \right)^2 1_{\left\{ \left| \frac{X_{i,n}}{\tau_n} \right| \geq \varepsilon \right\}} \right) \right| \\ &\leq \delta \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 + 2b \sum_{i=1}^n E \left(\left(\frac{X_{i,n}}{\tau_n} \right)^2 1_{\left\{ \left| \frac{X_{i,n}}{\tau_n} \right| \geq \varepsilon \right\}} \right) \end{aligned}$$

Aus der Lindeberg-Bedingung 9.9 folgt:

$$\overline{\lim}_{n \in \mathbb{N}} |B_n| \leq \delta.$$

Da $\delta > 0$ beliebig ist, folgt dann $B_n \xrightarrow[n \rightarrow \infty]{} 0$.

Wir zeigen $A_n \rightarrow 0$, $n \rightarrow \infty$:

$$\begin{aligned} |A_n| &= \left| E \left(h' \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) - \sum_{i=1}^n \frac{\sigma_{i,n}^2}{\tau_n^2} E \left(h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \right| \\ &= \left| \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 E \left(h' \left(\frac{Z_{i,n}}{\tau_n} + \frac{X_{i,n}}{\tau_n} \right) - h' \left(\frac{Z_{i,n}}{\tau_n} \right) \right) \right| \\ &\leq \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 E \left(\left| \cdot \right| 1_{\left\{ \left| \frac{X_{i,n}}{\tau_n} \right| < \varepsilon \right\}} \right) \\ &\quad + \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 E \left(\left| \cdot \right| 1_{\left\{ \left| \frac{X_{i,n}}{\tau_n} \right| \geq \varepsilon \right\}} \right) \\ &\leq \delta + 2b \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 P \left\{ \left| \frac{X_{i,n}}{\tau_n} \right| \geq \varepsilon \right\} \\ &\leq \delta + 2b \sum_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2 \frac{\sigma_{i,n}^2}{\varepsilon^2 \tau_n^2} \\ &\leq \delta + 2 \frac{b}{\varepsilon^2} \max_{i=1}^n \left(\frac{\sigma_{i,n}}{\tau_n} \right)^2, \end{aligned}$$

wobei die vorletzte Ungleichung aus der Markovschen Ungleichung folgt.

Aus dem nachfolgenden Lemma 9.13 folgt:

$$\overline{\lim}_{n \in \mathbb{N}} |A_n| \leq \delta$$

und damit $A_n \xrightarrow[n \rightarrow \infty]{} 0$.

9.13 Lemma. Die Lindeberg-Bedingung impliziert:

$$\max_{i=1}^n \frac{\sigma_{i,n}}{\tau_n} \xrightarrow[n \rightarrow \infty]{} 0.$$

Beweis des zentralen Grenzwertsatzes 9.10:

Es ist zu zeigen: $P \left\{ \frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \leq t \right\} \rightarrow \Phi(t), n \rightarrow \infty$.

Nach Satz 9.8 ist nur zu zeigen:

$$\forall f \in \mathcal{F} : E \left(f \left(\frac{1}{\tau_n} \sum_{i=1}^n X_{i,n} \right) \right) \xrightarrow[n \rightarrow \infty]{} \int f(x) \varphi(x) dx$$

o.E. ist $\int f(x) \varphi(x) dx = 0$, sonst betrachte man: $f - \int f(x) \varphi(x) dx$.

Wir zeigen: Es existiert eine differenzierbare Funktion h , für die h' gleichmäßig stetig und beschränkt ist, sodaß

$$f(x) = h'(x) - xh(x) \tag{9.17}$$

gilt. Aus Lemma 9.11 folgt damit die Behauptung.

Zu (9.17): Mit $\varphi = \Phi'$ setze

$$h(x) = \int_{-\infty}^x f(y) \varphi(y) dy / \varphi(x)$$

Wegen $\boxed{\varphi'(x) = -x\varphi(x)}$ gilt:

$$\begin{aligned} h'(x) &= \frac{f(x)\varphi(x)}{\varphi(x)} - \frac{\int_{-\infty}^x f(y)\varphi(y) dy \varphi'(x)}{\varphi^2(x)} \\ &= f(x) + xh(x). \end{aligned}$$

Die gleichmäßige Stetigkeit und Beschränktheit von h' folgt aus:

$$\begin{aligned} \lim_{x \rightarrow \pm\infty} xh(x) &= \lim_{x \rightarrow \pm\infty} \frac{\int_{-\infty}^x f(y)\varphi(y) dy}{\varphi(x)/x} \\ &= \lim_{x \rightarrow \pm\infty} (f(x)\varphi(x)) / \left(\frac{\varphi'(x)}{x} - \frac{\varphi(x)}{x^2} \right) = \lim_{x \rightarrow \pm\infty} \left(-\frac{f(x)}{1+x^{-2}} \right) \\ &= -\lim_{x \rightarrow \pm\infty} f(x) \text{ ex. } \in \mathbb{R}, \end{aligned}$$

wobei die zweite Gleichung wegen der Regel von de l'Hospital folgt.

9.14 Lemma. Es gelte

$$\forall n \in \mathbb{N} \forall i = 1, \dots, n: \quad \mathcal{L} \left(\frac{X_{i,n}}{\sigma_{i,n}} \right) = \mathcal{L} \left(\frac{X_{1,1}}{\sigma_{1,1}} \right)$$

Dann gilt:

$$\max_{i=1}^n \frac{\sigma_{i,n}}{\tau_n} \xrightarrow{n \rightarrow \infty} 0 \quad \Leftrightarrow \quad \text{Lindeberg-Bed. 9.9}$$

9.15 Satz. (Lindeberg-Lévy) Seien X_1, X_2, X_3, \dots i.i.d. mit $EX_1 =: \mu \in \mathbb{R}$ und $E((X_1 - \mu)^2) =: \sigma^2 \in (0, \infty)$. Dann gilt:

$$\forall t \in \mathbb{R}: \quad P \left\{ \frac{n^{1/2}}{\sqrt{\text{Var}(X_1)}} \left(\frac{1}{n} \sum_{i=1}^n X_i - EX_1 \right) \leq t \right\} \xrightarrow{n \rightarrow \infty} \Phi(t).$$

10 Schätzer, Konfidenzintervalle und Test: Die drei wichtigsten statistischen Verfahren am Beispiel von Binomialexperimenten

Seien in diesem Kapitel X_1, \dots, X_n i.i.d. $X_i \sim B(1, p) \forall i$, wobei der Parameter $p \in (0, 1)$ als unbekannt angesehen wird.

Schätzer für p

Ein vernünftiger Schätzer für p ist

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

denn er ist **erwartungstreu**, d.h. $E\bar{X}_n = p$, wenn p der wahre Parameter ist, und **konsistent**, d.h. $\bar{X}_n \rightarrow p$ n.W.

Ein erwartungstreuer und konsistenter Schätzer für die Varianz $p(1-p)$ einer $B(1, p)$ -Verteilung ist

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Konfidenzintervalle für p

Hier betrachten wir nur sog. approximative Konfidenzintervalle. Eine leichte Folgerung aus dem zentralen Grenzwertsatz ist die Aussage

$$\mathbb{P} \left\{ -t \leq \sqrt{n} \frac{\bar{X}_n - p}{\hat{\sigma}_n} \leq t \right\} \rightarrow \int_{-t}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =: I(t),$$

bei $t > 0$ oder äquivalent

$$\mathbb{P}\left\{\bar{X}_n - \frac{t}{\sqrt{n}}\hat{\sigma}_n \leq p \leq \bar{X}_n + \frac{t}{\sqrt{n}}\hat{\sigma}_n\right\} \rightarrow I(t).$$

Ist nun für $\alpha \in (0, 1)$ die Zahl u_α das sog. α -Fraktile der Standard Normalverteilung, d.h. der Wert u_α für den die VF Φ der Standard Normalverteilung

$$\Phi(u_\alpha) = 1 - \alpha,$$

ergibt, so gilt $I(u_{\alpha/2}) = 1 - \alpha$. Folglich enthält das sog. Konfidenzintervall

$$\left[\bar{X}_n - \frac{u_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n \leq p \leq \bar{X}_n + \frac{u_{\alpha/2}}{\sqrt{n}}\hat{\sigma}_n\right]$$

den unbekannt Parameter p mit approximativer W. $1 - \alpha$. Man nennt $1 - \alpha$ das **Konfidenzniveau** des Konfidenzintervalls. Übliche Werte für $1 - \alpha$ sind 95% oder 99%.

Tests für Hypothesen über p

Aufgrund der Beobachtungen X_1, \dots, X_n soll zwischen zwei sich gegenseitig ausschließenden Aussagen über p entschieden werden, z.B. für einen vorgegebenen Wert $p_0 \in (0, 1)$:

$$\text{Hypothese } H_0 : p \leq p_0 \quad \text{Alternative } K : p > p_0.$$

Man trifft die Entscheidung zwischen H_0 und K mit Hilfe eines Tests zum Niveau α , d.h. eines Verfahrens, bei dem die Wahrscheinlichkeit einer Entscheidung für K , wenn H_0 richtig ist (sog. **Fehlerwahrscheinlichkeit 1. Art**,) den vorgegebenen Wert α nicht überschreitet. Übliche Werte für α sind 0.05 oder 0.01.

Ein vernünftiges Testverfahren ist mit $S_n = X_1 + \dots + X_n$ (Anzahl der Erfolge) offenbar

$$\begin{aligned} \text{Entscheidung für } K, \text{ falls } & S_n > c_n(\alpha) \\ \text{Entscheidung für } H_0, \text{ falls } & S_n \leq c_n(\alpha), \end{aligned}$$

wobei $c_n(\alpha)$ aus den beiden Bedingungen

$$\mathbb{P}_{p_0}\{S_n > c_n(\alpha)\} \leq \alpha \text{ und } \mathbb{P}_{p_0}\{S_n \leq c_n(\alpha)\} > \alpha$$

bestimmt wird (Tabellen, Programme). Man nennt $c_n(\alpha)$ das α -Fraktile der Verteilung $\mathbb{P}_{p_0}^{S_n}$. Die Eigenschaften des Tests kann man an der sog. **Gütefunktion**

$$\beta_n(p) = \mathbb{P}_p\{S_n > c_n(\alpha)\}$$

ablesen, die die W. sich für K zu entscheiden angibt, wenn p der wahre Parameter ist. Für $p \leq p_0$, also für H_0 , ist $\beta_n(p)$ die Fehlerw. 1. Art, während für $p > p_0$, also für K der Wert $1 - \beta_n(p)$ die sog. Fehlerw. 2. Art ist, d.h. die W., sich fälschlich für H_0 zu entscheiden.

Der Test ist unsymmetrisch konstruiert, nämlich so, dass die Fehlerw. 1. Art unter Kontrolle ist.

Die Fehlerw. 2. Art ist dann nur noch über den Stichprobenumfang n zu kontrollieren: Je größer n , umso kleiner wird die Fehlerw. 2. Art.

Wenn bei der Durchführung des Zufallsexperiments der beobachtete Wert von S_n den kritischen Wert $c_n(\alpha)$ überschreitet, kann man relativ sicher sein, dass K wirklich vorliegt, da die W., sich fälschlich für K zu entscheiden α nicht übersteigt. Anders im Fall, dass der beobachtete Wert von S_n den kritischen Wert $c_n(\alpha)$ nicht überschreitet. Dann kann man nur feststellen, dass aufgrund des Datenmaterials eine Ablehnung von H_0 nicht möglich ist. Eine genauere Diskussion erfolgt in der Vorlesung.

11 Allgemeine Maßräume und Zufallsexperimente über \mathbb{R}^k

Bemerkung: Das schwache Gesetz der großen Zahlen und der zentrale Grenzwertsatz wurden hier nur für diskrete ZV bewiesen. Die Beweise haben wir jedoch so durchgeführt, dass sie ohne Änderung auch im allgemeinen Rahmen, den wir jetzt schaffen werden, ihre Gültigkeit behalten. Es wurden nämlich beim Beweis nur Linearität und Isotonie des Erwartungswert, sowie die Tatsache, dass stochastisch unabhängige ZV unkorreliert sind benutzt. Beim zentralen Grenzwertsatz für i.i.d. ZV wurde darüberhinaus benutzt, dass für eine quadratintegrierbare ZV Z stets $E Z^2 1\{Z^2 \geq t\} \rightarrow 0$, falls $t \rightarrow \infty$, was für diskrete ZV klar ist und für allgemeine ZV aus dem sog. Satz von der monotonen Konvergenz folgt, s.[BN]:Satz 18.6.

Der zentrale Grenzwertsatz besagt, dass sich unter gewissen Bedingungen die Verteilung von $Y_n := \sqrt{n}(\bar{X}_n - EX_1)/\sqrt{Var X_1}$ für $n \rightarrow \infty$ stabilisiert in dem Sinne, dass die W. $\mathbb{P}^{Y_n}(a, b]$ für $n \rightarrow \infty$ sich immer mehr dem Wert $F(b) - F(a)$ mit $F = \Phi$ annähert. Es stellt sich deshalb die Frage, ob man ein Limes W-Maß P auf \mathbb{R} definieren kann, das die Gleichung

$$P((a, b]) = F(b) - F(a) \quad \forall a \leq b, \quad (11.1)$$

erfüllt. Ein solches P kann bei $F = \Phi$ kein diskretes W-Maß sein, da wegen der Stetigkeit von Φ notwendig $P\{a\} = 0 \quad \forall a \in \mathbb{R}$ gelten müßte. Es muss also eine neue Art von W-Maß sein. Will man nun ein W-Maß über $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ mit den Eigenschaften (1.1)- (1.3) definieren, so treten mathematische Schwierigkeiten auf, die nicht überwindbar sind. Diese liegen daran, dass die Potenzmenge *zu viele* Mengen enthält, darunter auch pathologische Mengen, die in der Praxis uninteressant sind.

Wenn man unsere früheren Überlegungen mit diskreten W-Maßen P anschaut, so sieht man, dass von den Potenzmengen $\mathcal{P}(\Omega)$ nur verwendet wurde, dass die *interessierenden Mengen dazugehören* (dies waren die Einpunktmengen, die aus Trägerpunkte des W-Maßes gebildet waren) und dass man bestimmte Mengenoperationen (abzählbare Vereinigungen und Durchschnitte, Komplemente u.s.w.) ausführen konnte, und $P(A)$ dann auch für die so entstandene Menge A definiert war.

Ein eleganter Ausweg aus den obigen Schwierigkeiten besteht nun darin, nur noch Definitions-

bereiche für W-Maße zu betrachten, die abgeschlossen gegenüber den eben genannten Mengenoperationen sind, sog. σ -Algebren, und zwar jeweils die kleinste σ -Algebra, die die momentan interessierenden Mengen enthält. In der obigen Situation mit \mathbb{R} als Grundraum und den interessierenden Mengen $(a, b]$, würde man also als Definitionsbereich eines W-Maßes P mit (11.1) die kleinste, σ -Algebra über \mathbb{R} wählen, die alle Intervalle $(a, b]$ enthält. Dies ist die sogenannte Borelsche σ -Algebra, s.unten. Wir kommen also in natürlicher Weise zu

11.1 Definition. [BN]:13.2 Gegeben sei ein Stichprobenraum $\Omega (\neq \emptyset)$. Dann heißt jedes System $\mathcal{A} \subset \mathcal{P}(\Omega)$ (=Potenzmenge von Ω) mit

$$\Omega \in \mathcal{A} \tag{11.2}$$

$$A^c \in \mathcal{A} \text{ falls } A \in \mathcal{A} \tag{11.3}$$

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \text{ falls } A_i \in \mathcal{A} \quad \forall i \in \mathbb{N} \tag{11.4}$$

eine σ -Algebra über Ω .

Das Paar (Ω, \mathcal{A}) heißt **Meßraum**. Ein Element $A \in \mathcal{A}$ heißt **meßbare Menge**.

Gilt (11.4) nur für endlich viele $A_i \in \mathcal{A}$, so spricht man von einer *Algebra* über Ω . („ σ -“ bedeutet soviel wie „abzählbar unendlich“)

(11.3) besagt, daß \mathcal{A} abgeschlossen gegenüber Komplementbildung ist: besitzt A einen W-Wert, so auch A^c .

(11.4) besagt, daß \mathcal{A} abgeschlossen gegenüber abzählbaren Vereinigungen ist: Haben die Ereignisse A_1, A_2, \dots W-Werte bekommen, so auch das Ereignis $A =$ „mindestens ein A_i tritt ein.“ = $\bigcup_{i=1}^{\infty} A_i$.

σ -Algebren werden die Definitionsbereiche von W-Maßen sein.

Bemerkung: [BN]:S. 158 Ist \mathcal{A} eine σ -Algebra, dann ist \mathcal{A} auch Algebra, denn aus (11.2) und (11.3) folgt $\emptyset = \Omega^c \in \mathcal{A}$ und somit ist für $A_1, \dots, A_n \in \mathcal{A}$ nach (11.4) auch $A_1 \cup \dots \cup A_n \cup \emptyset \cup \dots \in \mathcal{A}$, also $A_1 \cup \dots \cup A_n \in \mathcal{A}$.

11.2 Beispiel. (σ -Algebren),[BN]:13.2

1. $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$ ist die größte σ -Algebra über Ω .
2. $\{\emptyset, \Omega\}$ ist die kleinste σ -Algebra über Ω .
3. Ist \mathcal{A} eine Algebra, die nur aus endlich vielen Elementen (= Mengen) besteht, so ist \mathcal{A} schon eine σ -Algebra.
4. Falls $|\Omega| = \infty$ ist, so ist $\mathcal{A} = \{A \subset \Omega : A \text{ oder } A^c \text{ endlich}\}$ eine Algebra aber keine σ -Algebra.

5. Sind $\mathcal{A}_i, i \in I$, beliebig viele σ -Algebren über Ω , so ist das Mengensystem $\mathcal{A} := \{A \subset \Omega : A \in \mathcal{A}_i \ \forall i \in I\}$ wieder eine σ -Algebra. Faßt man jede Menge $A \subset \Omega$ als Element der Potenzmenge $\mathcal{P}(\Omega)$ auf, so ist gerade $\mathcal{A} = \bigcap_{i \in I} \mathcal{A}_i$ (Durchschnittsbildung in $\mathcal{P}(\Omega)$).

Beweis: Zu (11.2): $\Omega \in \mathcal{A}$, da $\Omega \in \mathcal{A}_i \ \forall i$.

Zu (11.3): Sei $A \in \mathcal{A}$, d.h. $A \in \mathcal{A}_i \ \forall i \Rightarrow A^c \in \mathcal{A}_i \ \forall i \Rightarrow A^c \in \mathcal{A}$.

Zu (11.4): Seien $A_k \in \mathcal{A}, k \in \mathbb{N}$, d.h. $A_k \in \mathcal{A}_i \ \forall i \in I \ \forall k \in \mathbb{N} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}_i \ \forall i$, d.h. $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$. \square

Wie oben diskutiert hat man bei vielen Problemstellungen eine Familie von Mengen, die auf jeden Fall einen W -Wert bekommen sollen. Dann gibt es immer eine kleinste σ -Algebra, die die vorgegebenen Mengen enthält:

\mathcal{E} sei ein gegebenes System von Mengen $\subset \mathcal{P}(\Omega)$.

Dann heißt die σ -Algebra

$$\sigma_{\Omega}(\mathcal{E}) := \bigcap_{\mathcal{A} \supset \mathcal{E}} \mathcal{A} \quad (11.5)$$

die von \mathcal{E} über Ω erzeugte σ -Algebra, [BN]:(13.5).

\mathcal{E} heißt Erzeuger von $\sigma_{\Omega}(\mathcal{E})$. Jede σ -Algebra hat verschiedene Erzeuger.

(Ist \mathcal{A} σ -Algebra, so erzeugt \mathcal{A} sich selbst; wenn $A \in \mathcal{A}$ ist, so erzeugt $\mathcal{A} \setminus \{A\} =: \mathcal{E}$ ebenfalls \mathcal{A} .)

6. Ist $\Omega (\neq \emptyset)$ eine abzählbare Menge, so erzeugt $\mathcal{E} = \{\{\omega\} : \omega \in \Omega\}$ (d.h. das System der Einpunktmengen in Ω) die Potenzmenge $\mathcal{P}(\Omega)$, da für $A \subset \Omega$ gilt $A = \bigcup_{\omega \in A} \{\omega\} \in \sigma_{\Omega}(\mathcal{E})$. Dies ist, wie wir wissen, der Grund dafür, daß man bei abzählbaren Grundräumen ohne Maßtheorie auskommt: Jede Teilmenge A von Ω bekommt einen W -Wert (nämlich die Summe der W -Werte aller $\omega \in A$, siehe Definition 11.3).

7. σ -Algebren sind wie erwähnt deshalb nötig, damit man Zufallsexperimente über komplizierteren Grundräumen beschreiben kann, z.B. in Zusammenhang mit dem zentralen Grenzwertsatz:

$$\Omega = \mathbb{R}$$

$\mathcal{E} = \mathcal{G}_1 := \{(a, b] : a, b \in \mathbb{R}, a \leq b\}$ (mindestens die Intervalle $(a, b]$ sollen einen W -Wert bekommen).

Dann heißt $\mathbb{B} := \sigma_{\mathbb{R}}(\mathcal{G}_1)$ die Borel- σ -Algebra über \mathbb{R} (die kleinste σ -Algebra über \mathbb{R} , die die Intervalle $(a, b]$ enthält).

8. Ganz analog definiert man die Borelsche σ -Algebra \mathbb{B}^k über $\mathbb{R}^k, k \geq 1$ gemäß

$$\mathbb{B}^k = \sigma_{\mathbb{R}^k}(\mathcal{G}_k)$$

mit

$$\mathcal{G}_k := \{(a, b] : a, b \in \mathbb{R}^k, a \leq b\},$$

wobei $a = (a_1, \dots, a_k), b = (b_1, \dots, b_k), (a, b] = \bigtimes_{i=1}^k (a_i, b_i]$ und $a \leq b \Leftrightarrow a_i \leq b_i \forall i$.

\mathbb{B}^k wird praktisch immer verwendet, wenn $\Omega = \mathbb{R}^k$ ist.

Man kann zeigen: $\mathbb{B}^k \neq \mathcal{P}(\mathbb{R}^k)$.

\mathbb{B}^k enthält alle offenen und abgeschlossenen Teilmengen in \mathbb{R}^k , z.B. auch die Einpunktmengen, aber auch alle endlichen und abzählbar unendlichen Mengen.

Vorsicht: Man kann \mathbb{B}^k nicht konstruktiv durch Vereinigungs-, Durchschnitts- und Komplexbildung aus den Rechtecken $(a, b]$ aufbauen.

Analog zur Definition im diskreten Fall definieren wir nun allgemeine Wahrscheinlichkeitsräume.

11.3 Definition. (Kolmogorowsche Axiome (1933), [BN]:13.4)

Gegeben sei eine σ -Algebra \mathcal{A} über $\Omega (\neq \emptyset)$. Jede Abbildung $P : \mathcal{A} \rightarrow \mathbb{R}$ (P wie Probability) mit den Eigenschaften

$$P(A) \geq 0 \quad \forall A \in \mathcal{A} \quad (\text{Nichtnegativität}) \quad (11.6)$$

$$P(\Omega) = 1 \quad (\text{Normiertheit}) \quad (11.7)$$

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\sigma\text{-Additivität}) \quad (11.8)$$

wobei $\sum_{i=1}^{\infty} A_i := \bigcup_{i=1}^{\infty} A_i$ falls $A_i \cap A_j = \emptyset \quad \forall i \neq j$, ansonsten ist die Summe nicht definiert, heißt **Wahrscheinlichkeitsmaß** (W -Maß) auf \mathcal{A} .

Das Tripel (Ω, \mathcal{A}, P) heißt ein **Wahrscheinlichkeitsraum** (W -Raum).

Falls P den Wert ∞ annehmen darf und die Bedingungen (11.6), (11.8) und als Ersatz für (11.7) die Bedingung $P(\emptyset) = 0$ (Nulltreue) erfüllt, heißt P ein **Maß** und (Ω, \mathcal{A}, P) ein **Maßraum**. Wir werden

Statt P schreibt man in diesem Fall meist μ und entsprechend für den Maßraum $(\Omega, \mathcal{A}, \mu)$.

Ist P ein W -Maß, so gilt wie im diskreten Fall: $P(\emptyset) = P(\sum_{i=1}^{\infty} \emptyset) \stackrel{(11.8)}{=} \sum_{i=1}^{\infty} P(\emptyset) \in \mathbb{R} \Rightarrow P(\emptyset) = 0 \Rightarrow P$ ist ein Maß.

Übrigens gilt für jedes Maß μ : Die Bedingung $\mu(\emptyset) = 0$ ist äquivalent zu $\exists A \in \mathcal{A}$ mit $\mu(A) < \infty$. (Beweis ist trivial).

Maße μ mit Wert ∞ treten z.B. bei der Längenmessung in \mathbb{R} auf (das Intervall $(-\infty, a]$ hat die Länge ∞). Die in Kapitel 3 hergeleiteten Rechenregeln gelten mit identischen Beweisen für allgemeine (W -Räume) (Ω, \mathcal{A}, P) . Da für allgemeine Maße μ auch der Wert ∞ vorkommen kann, gilt die Stetigkeit von oben für allgemeine Maße nur dann, wenn in (3.5) mit $P = \mu$ für mindestens ein A_i gilt $P(A_i) < \infty$.

Zu Beginn dieses Kapitels hatten wir uns gefragt, ob es für eine Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$ ein W -Maß P über (\mathbb{R}, \mathbb{B}) gibt mit der Eigenschaft

$$P(a, b] = F(b) - F(a) \quad \forall a \leq b.$$

Die Antwort darauf liefert der Korrespondenzsatz.

11.4 Satz. (Korrespondenzsatz, [BN]:14.9)

Zu jeder VF F über \mathbb{R} existiert genau ein W -Maß P über (\mathbb{R}, \mathbb{B}) mit

$$P(a, b] = F(b) - F(a) \quad \forall a \leq b. \quad (11.9)$$

Umgekehrt gibt es zu jedem W -Maß P genau eine VF F mit (11.9), für die dann notwendig gilt:

$$P(-\infty, x] = F(x) \quad \forall x \in \mathbb{R}. \quad (11.10)$$

Letzteres gilt wegen $((-\infty, x] = \bigcup_{n=1}^{\infty} \underbrace{(-n, x]}_{\in \mathbb{B}}$, also $P(-\infty, x] = \lim_{n \rightarrow \infty} P(-n, x] = F(x) - F(-n) = F(x)$ (Stetigkeit von oben)

Folgerung: Wie schon bei diskreten Maßen lassen sich auch gemäß Lemma 9.1 alle Eigenschaften eines W -Maßes P über (\mathbb{R}, \mathbb{B}) auch durch seine VF ausdrücken.

Dabei gehören offene Intervalle $(-\infty, x)$ zu \mathbb{B} , denn:

$$(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - \frac{1}{n}] \in \mathbb{B} \Rightarrow \{x\} = (-\infty, x] \setminus (-\infty, x) = (-\infty, x] \cap (-\infty, x)^c \in \mathbb{B}.$$

Ein W -Maß P über (\mathbb{R}, \mathbb{B}) heißt *stetig*, wenn die zugehörige VF F stetig ist.

11.1 W -Maße mit Riemann-Dichten

11.5 Definition. [BN]:15.1,15.2 Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt eine **Riemann-Dichte** (kurz: *R-Dichte*) über \mathbb{R} , wenn sie (uneigentlich) Riemann-integrierbar ist und

$$f(x) \geq 0 \quad \forall x \in \mathbb{R} \quad \text{und} \quad \int_{-\infty}^{\infty} f(x) dx = 1 \quad (11.11)$$

gilt. Die Funktion

$$F(x) = \int_{-\infty}^x f(y) dy \quad (11.12)$$

ist eine stetige VF und für das zugehörige W -Maß P über (\mathbb{R}, \mathbb{B}) gilt

$$P(a, b] = \int_a^b f(y) dy \quad \forall a \leq b. \quad (11.13)$$

Ganz entsprechend kann man *R-Dichten* f über \mathbb{R}^k und zugehörige W -Maße P über $(\mathbb{R}^k, \mathbb{B}^k)$ definieren. Man fasse dazu z.B. in (11.12) das Integral als k -faches *R-Integral*

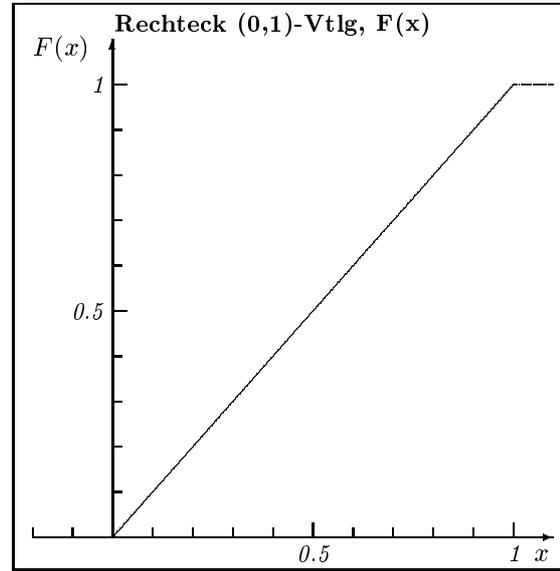
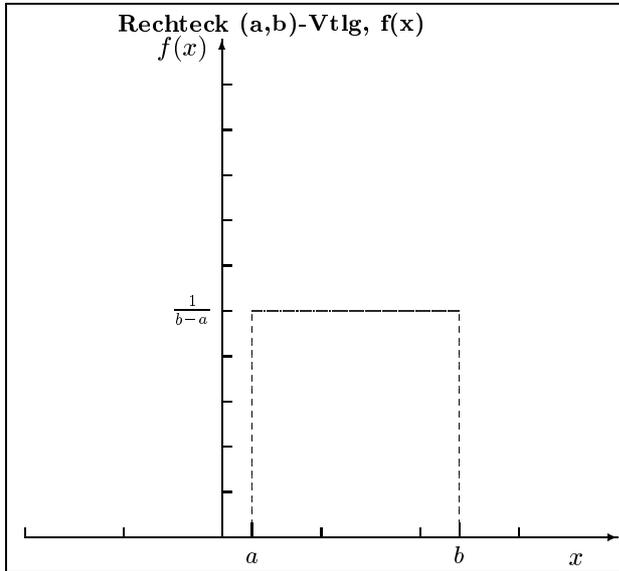
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_k) dx_1 \cdots dx_k = 1$$

auf und entsprechend in (11.13).

Definition 11.5 und (11.12) lassen sich sinngemäß auch auf $f : (0, \infty) \rightarrow \mathbb{R}$ oder allgemein $f : (a, b) \rightarrow \mathbb{R}$ übertragen, falls f auf (a, b) uneigentlich R-integrierbar ist. Entsprechend in \mathbb{R}^k .

11.6 Beispiel. [BN]:15.3 Rechteck(a, b)-Vtlg (kurz: $\mathcal{R}(a, b)$ -Vtlg)

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x), x \in \mathbb{R}. \quad (11.14)$$



Spezialfall: $a = 0, b = 1$. In diesem Fall ist $F(x) = \int_{-\infty}^x 1_{(0,1)}(y) dy$ (siehe Bild).

11.7 Beispiel. [BN]:15.4 Exponential (α)-Vtlg, Weibull (α, β)-Vtlg

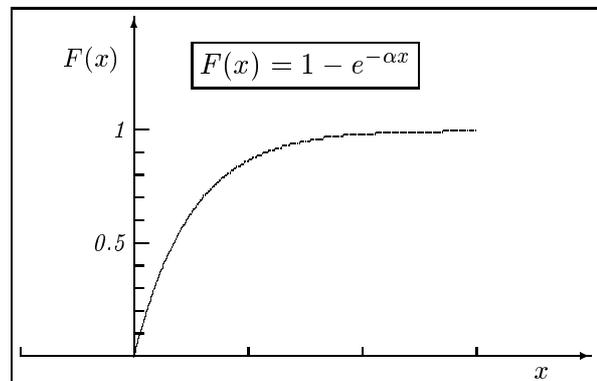
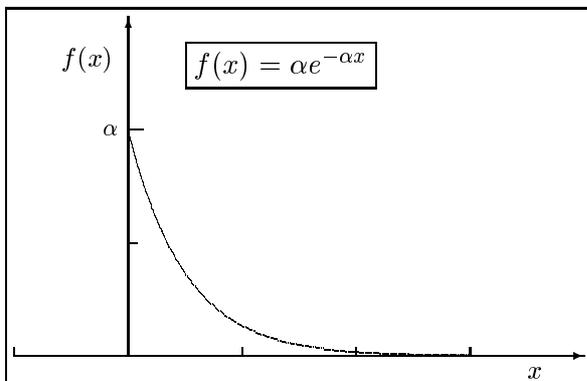
Sei $0 < \alpha < \infty$.

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0 \\ 0 & , x \leq 0 \end{cases} \quad (11.15)$$

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \alpha e^{-\alpha x} dx = -e^{-\alpha x} \Big|_0^{\infty} = 0 - (-1) = 1.$$

Folglich ist f eine Dichte. Die zugehörige VF ist

$$F(x) = \int_{-\infty}^x f(y) dy = -e^{-\alpha y} \Big|_0^x 1_{(0,\infty)}(x) = \begin{cases} 1 - e^{-\alpha x} & \text{falls } x > 0 \\ 0 & \text{falls } x \leq 0. \end{cases} \quad (11.16)$$



Die Exponential (α)-Verteilung, kurz $\mathcal{E}(\alpha)$ -Vtlg, beschreibt die Lebensdauer vtlg von Stoffen oder Produkten, die keinem Alterungsprozeß unterliegen, z.B. radioaktive Atome, Glühlampen (approximativ), Transistoren.

Eigenschaft: Die Ausfallrate $r(x)$ zum Zeitpunkt x ist konstant, dabei ist die Ausfallrate in x definiert gemäß:

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \frac{P(x, x + \varepsilon]}{P(x, \infty)} &= \lim_{\varepsilon \downarrow 0} \frac{F(x + \varepsilon) - F(x)}{\varepsilon} \frac{1}{1 - F(x)} \\ &= \frac{f(x)}{1 - F(x)} =: r(x) \end{aligned}$$

$r(x)$ für $\mathcal{E}(\alpha)$ -Vtlg:

$$r(x) = \frac{\alpha e^{-\alpha x}}{1 - (1 - e^{-\alpha x})} = \alpha = \text{const}$$

Weibull (α, β)-Vtlg:

Für $0 < \alpha < \infty$ und $0 < \beta < \infty$ ist

$$f(x) = \begin{cases} \alpha \beta x^{\beta-1} e^{-\alpha x^\beta}, & x > 0 \\ 0 & , x \leq 0 \end{cases} \quad (11.17)$$

die Dichte der Weibull (α, β)-Vtlg. Damit ist

$$F(x) = \int_{-\infty}^x f(y) dy = \begin{cases} 1 - e^{(-\alpha x^\beta)}, & x > 0 \\ 0 & , x \leq 0. \end{cases} \quad (11.18)$$

Ausfallrate $r(x) = \frac{\alpha \beta x^{\beta-1} e^{-\alpha x^\beta}}{e^{-\alpha x^\beta}} = \alpha \beta x^{\beta-1}$.

$\beta > 1$ Ausfallrate steigend (Alterung)

$\beta = 1$ Ausfallrate konstant (keine Alterung, Exponentialvtlg)

$\beta < 1$ Ausfallrate fallend (Kinderkrankheiten)

11.8 Beispiel. [BN]:15.5 Normal (μ, σ^2)-Vtlg, $\mu \in \mathbb{R}, \sigma^2 > 0$ Die schon in (9.15) definierte Gaußsche Glockenkurve:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}, \quad (11.19)$$

ist eine R-Dichte. Dazu die folgenden Überlegungen.

Sei $\Psi(z) := (2\pi)^{-\frac{1}{2}} \int_0^z e^{-\frac{x^2}{2}} dx, z \geq 0$. Dann gilt

1.

$$\frac{1}{2} \left(1 - e^{-\frac{z^2}{2}}\right)^{\frac{1}{2}} \leq \Psi(z) \leq \frac{1}{2} \left(1 - e^{-z^2}\right)^{\frac{1}{2}}$$

und somit für die VF

$$\Phi(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$$

der $\mathcal{N}(0, 1)$ -Verteilung die Abschätzung

2.

$$\frac{1}{2} + \frac{1}{2} \left(1 - e^{-\frac{z^2}{2}}\right)^{\frac{1}{2}} \leq \Phi(z) \leq \frac{1}{2} + \frac{1}{2} \left(1 - e^{-z^2}\right)^{\frac{1}{2}}, z \geq 0.$$

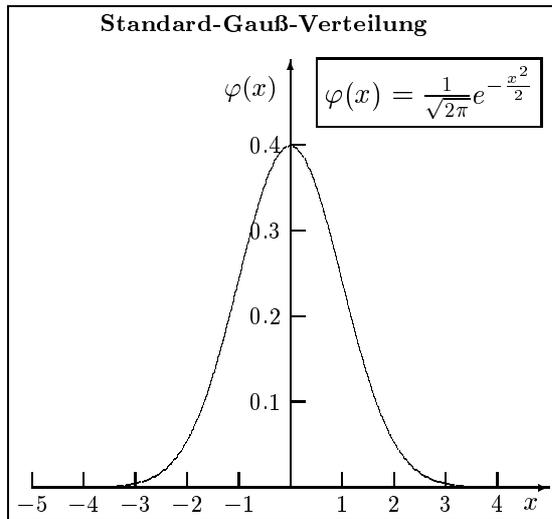
Aus 1. folgt $\int_0^\infty \varphi(x) dx = \frac{1}{2}$. Aus Symmetriegründen gilt auch $\int_{-\infty}^0 \varphi(x) dx = \frac{1}{2}$, also $\int_{-\infty}^\infty \varphi(x) dx = 1$.

Der Beweis ist eine Übungsaufgabe, hier ein Hinweis:

Es ist

$$(\Psi(z))^2 = \int_0^z \int_0^z \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dx dy.$$

Man führe Polarkoordinaten (r, α) , $r > 0, 0 \leq \alpha < 2\pi$ ein, wobei $x = r \sin \alpha, y = r \cos \alpha$, also $x^2 + y^2 = r^2$ und $dx dy = -r dr d\alpha$, und vergleiche das Integral über $Q = [0, z]^2$ mit dem größten Viertelkreis um $(0,0)$, der innerhalb Q liegt, bzw. dem kleinsten Viertelkreis, der Q umfaßt.



Fläche unter der Gauß-Kurve:

$$\int_{-1}^1 \varphi(x) dx = 0.683$$

$$\int_{-2}^2 \varphi(x) dx = 0.954$$

$$\int_{-3}^3 \varphi(x) dx = 0.997$$

Das zu φ gehörige W-Maß heißt $\mathcal{N}(0, 1)$ -Vtlg, *Standard-Normalverteilung* oder auch *Standard-Gauß-Verteilung*.

Mit φ ist auch

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma > 0, \quad (11.20)$$

eine Dichte, denn:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} \varphi\left(\underbrace{\frac{x - \mu}{\sigma}}_{=: y, dy = \frac{1}{\sigma} dx}\right) dx = \int_{-\infty}^{\infty} \varphi(y) dy = 1.$$

Die zugehörige Vtlg heißt $\mathcal{N}(\mu, \sigma^2)$ -Vtlg, wobei man μ auch einen *Lageparameter* und σ einen *Skalenparameter* nennt (bei $x = \mu$ liegt der Extrempunkt der Kurve). Sobald wir den Erwartungswert und die Varianz von diskreten ZVen auf allgemeine reelle ZVe verallgemeinert ha-

ben, werden wir zeigen können, dass μ der Erwartungswert und σ^2 die Varianz einer $\mathcal{N}(\mu, \sigma^2)$ -Verteilung ist.

Normalverteilungen treten, wie wir schon im Zentralen Grenzwertsatz von Lindeberg gesehen auf, wenn viele kleine Störungen einen wahren Meßwert überlagern, z.B. bei physikalischen Messungen.

11.2 Erwartungswert und Varianz von Verteilungen mit R-Dichten

11.9 Definition. Sei P ein W-Maß über \mathbb{R} mit R-Dichte f . Falls $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$, so heißt

$$E(P) = \int_{-\infty}^{\infty} xf(x)dx \quad (11.21)$$

Erwartungswert von P und

$$\text{Var}(P) = \int_{-\infty}^{\infty} (x - E(P))^2 |f(x)dx \quad (11.22)$$

heißt **Varianz** von P , falls das Integral endlich ist.

Im Vorgriff auf später definierte reellwertige ZV X mit Verteilung P^X , die eine R-Dichte f besitze, definieren wir dann wie im diskreten Fall $EX = E(P^X)$ und $\text{Var}X = \text{Var}(P^X)$. Allgemeiner, wird dann für eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ derart, dass $g(x)f(x)$ R-integrierbar ist der E-Wert von $g(X)$ definiert gemäß

$$Eg(X) = \int_{-\infty}^{\infty} g(x)f(x)dx. \quad (11.23)$$

Im Anhang A.2 des Buches von Hübner finden sich z.B. eine Reihe von wichtigen Verteilungen mit ihren R-Dichten, E-Werten und Varianzen.

12 Allgemeine Zufallsvariable und ihre Verteilungen

In Kapitel 5 wurden Abbildungen $X : \Omega \rightarrow \mathcal{X}$ betrachtet und aufgrund der Tatsache, dass für $B \subset \mathcal{X}$ die Mengen

$$\{\omega \in \Omega : X(\omega) \in B\} \text{ und } B$$

dasselbe Ereignis repräsentieren (nur in verschiedenen Räumen), induzierte W-Maße eingeführt, s.(5.3). Da wir nun auch im Bildraum eine σ -Algebra betrachten wollen, werden wir im allgemeinen Rahmen in natürlicher Weise auf messbare Abbildungen im folgenden Sinne geführt.

12.1 Definition. [BN]:17.1 Seien (Ω, \mathcal{A}) und $(\mathcal{X}, \mathcal{B})$ zwei Messräume.

Eine Abbildung $X : \Omega \rightarrow \mathcal{X}$ heißt dann **\mathcal{A} - \mathcal{B} -messbar**, falls

$$\{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} \in \mathcal{A} \quad \forall B \in \mathcal{B}. \quad (12.1)$$

erfüllt ist.

In Analogie zum diskreten Fall können wir nun allgemein induzierte Maße definieren.

12.2 Definition. [BN]:17.2 Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\mathcal{X}, \mathcal{B})$ ein Meßraum und $X : \Omega \rightarrow \mathcal{X}$ eine \mathcal{A} – \mathcal{B} -meßbare Abbildung. Dann wird durch

$$\mu^X(B) = \mu(X^{-1}(B)), \quad B \in \mathcal{B}, \quad (12.2)$$

ein Maß μ^X über $(\mathcal{X}, \mathcal{B})$ definiert. μ^X heißt das von X und μ über $(\mathcal{X}, \mathcal{B})$ **induzierte Maß**. Ist $\mu = P$ ein W -Maß, so auch P^X . P^X wird auch **Verteilung** von X unter P genannt (Bez.: $\mathcal{L}_P(X)$ oder $\mathcal{L}(X)$, \mathcal{L} wie Law).

Ist Q ein W -Maß mit $Q = P^X$, so sagt man: X ist gemäß Q verteilt, kurz: $X \sim Q$. X heißt in diesem Zusammenhang auch **Zufallsvariable** (ZV). (Gegenüber dem diskreten Fall ist nur die Meßbarkeit von X hinzugekommen.)

Es bietet sich nun an, den Begriff der *diskreten* ZV ein wenig allgemeiner zu fassen als bisher: Falls $\{x\} \in \mathcal{B} \forall x \in \mathcal{X}$ und P^X diskret ist in dem Sinne, dass eine abzählbare Teilmenge \mathcal{X}_0 von \mathcal{X} existiert mit $P^X(\mathcal{X}_0) = 1$, so heißt X **diskrete ZV**.

Diskrete ZV im früheren Sinn bleiben nach dieser Definition immer noch solche. Darüberhinaus sind nun alle messbaren ZV, die nur abzählbar viele Werte annehmen, diskret. Früher musste man vom Hintergrundraum $(\Omega, \mathcal{P}(\Omega), P)$ dazu ebenfalls die Diskretheit verlangen.

Dass P^X tatsächlich ein W -Maß über $(\mathcal{X}, \mathcal{B})$ ist oder so zumindest so aufgefasst werden kann, sieht man wie im diskreten Fall.

Sprechweise wie früher: Eine ZV X auf einem W -Raum (Ω, \mathcal{A}, P) , $X : \Omega \rightarrow \mathcal{X}$ beschreibt ein Zufallsexperiment $(\mathcal{X}, \mathcal{B}, Q)$ wenn $P^X = Q$ ist, d.h. wenn die Vtlg. von X gleich Q ist. Kurz: $X \sim Q$ und $\mathcal{L}(X) = Q$.

Wie im diskreten Fall kann jedes Zufallsexperiment $(\mathcal{X}, \mathcal{B}, Q)$ durch eine ZV X beschrieben werden. Man wähle analog wie früher $\Omega = \mathcal{X}$, $\mathcal{A} = \mathcal{B}$, $X = id_\Omega$, die Identität auf Ω , und $P = Q$. Als Beispiel zur Erinnerung:

Ein ideales Würfelexperiment wird durch eine reelle ZV X mit $X \sim$ Laplace-Vtlg. über $\{1, \dots, 6\}$ beschrieben.

Praktisch alle von uns betrachteten ZV'n und Abbildungen sind reell oder haben Werte in \mathbb{R}^k und sind messbar. Letzteres folgt aus den beiden folgenden Lemmata.

12.3 Lemma. [BN]:17.10 Gegeben seien meßbare Abbildungen X und Y gemäß

$$(\Omega, \mathcal{A}) \xrightarrow{X} (\mathcal{X}, \mathcal{B}) \xrightarrow{Y} (\mathcal{Y}, \mathcal{C}).$$

$Y \circ X : \Omega \rightarrow \mathcal{Y}$ ist dann \mathcal{A} – \mathcal{C} -meßbar.

12.4 Lemma. [BN]:17.12, 17.13

a) Ist (Ω, \mathcal{A}) ein Meßraum, so ist $f : \Omega \rightarrow \mathbb{R}$ genau dann \mathcal{A} - \mathbb{B} -meßbar, wenn gilt

$$\{f > a\} = \{\omega \in \Omega : f(\omega) > a\} = f^{-1}((a, \infty)) \in \mathcal{A} \quad \forall a \in \mathbb{R}. \quad (12.3)$$

Dieselbe Aussage gilt mit $\{f \geq a\}$, bzw. $\{f < a\}$ und $\{f \leq a\}$ anstelle von $\{f > a\}$ in (12.3).

Da $\{f > a\}$ offen und eine Teilmenge aus \mathbb{B} ist, ist insbesondere jede stetige Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ auch \mathbb{B} - \mathbb{B} -meßbar.

b) Ist (Ω, \mathcal{A}) ein Meßraum, so ist $(f_1, \dots, f_k) = f : \Omega \rightarrow \mathbb{R}^k$ genau dann \mathcal{A} - \mathbb{B}^k -meßbar, wenn jede Komponente \mathcal{A} - \mathbb{B} -meßbar ist.

Als Folgerung erkennen wir z.B., dass Summen und Produkte reeller ZV reelle ZV sind.

13 Produkträume, Produktmaße, stochastische Unabhängigkeit und die Transformationsformel für R-Dichten

Wie schon bei der Behandlung diskreter Modelle bemerkt, verlaufen viele Experimente in einer endlichen Anzahl von Stufen (s. Kapitel 5). Die dortigen Begriffe Projektion, Randverteilung, gemeinsame Verteilung bleiben im allgemeinen Rahmen praktisch unverändert, nur durch Messbarkeitsforderungen ergänzt: Ist \mathcal{X}_i der Grundraum der i -ten Stufe mit σ -Algebra \mathcal{B}_i , $1 \leq i \leq n$, so ist $\mathcal{X} := \prod_{i=1}^n \mathcal{X}_i$ ein natürlicher Grundraum für das Gesamtexperiment den man üblicherweise mit der sog. Produkt- σ -Algebra $\mathcal{B} := \bigotimes_{i=1}^n \mathcal{B}_i$ versieht, die durch das Mengensystem

$$B_1 \times \dots \times B_n; \quad B_i \in \mathcal{B}_i \forall i$$

erzeugt wird. Dann sind die Projektionen $\pi_i : \mathcal{X} \rightarrow \mathcal{X}_i$ automatisch $\mathcal{B} - \mathcal{B}_i$ -messbar.

Ist P ein W-Ma auf dem Produktraum $(\mathcal{X}, \mathcal{B})$, so nennt man wie im diskreten Fall die Verteilung der Projektion π_i die i -te Randverteilung von P .

Betrachtet man auf einem allgemeinen W-Raum (Ω, \mathcal{A}, P) mehrere Abbildungen $\mathcal{X}_i : \Omega \rightarrow \mathcal{X}_i, 1 \leq i \leq n$, so ist die Produktabbildung $X = (X_1, \dots, X_n)$ mit Werten in $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$ eine $\mathcal{A} - \mathcal{B}$ -messbare Abbildung. $\mathcal{L}(X) = P^X$ heißt *gemeinsame Verteilung* der X_i , wobei wie im diskreten Fall gilt:

$$(P^{(X_1, \dots, X_n)})^{\pi_i} = P^{X_i} \forall i.$$

Bemerkung: Bei $\mathcal{X}_i = \mathbb{R}, \mathcal{B}_i = \mathbb{B} \forall i$ gilt für die Produkt- σ -Algebra

$$\mathcal{B} := \bigotimes_{i=1}^n \mathcal{B}_i = \mathbb{B}^n.$$

Wir definieren allgemein Produktmae.

13.1 Definition. (Produktmaß) Seien $(\mathcal{X}_i, \mathcal{B}_i, P_i)$, $1 \leq i \leq n$, Dann gibt es auf $(\prod_{i=1}^n \mathcal{X}_i, \prod_{i=1}^n \mathcal{B}_i)$ ein eindeutig bestimmtes W-Maß $P := \bigotimes_{i=1}^n P_i$, das **Produktmaß** der P_i , mit der Eigenschaft

$$\bigotimes_{i=1}^n P_i \left(\prod_{i=1}^n B_i \right) = \prod_{i=1}^n P_i(B_i). \quad (13.1)$$

Zur Existenz des Produktmaßes s. [BN]: Kapitel IV

In Analogie zum diskreten Fall definieren wir nun

13.2 Definition. In der Situation zu Beginn dieses Kapitels heißen X_1, \dots, X_n genau dann *stochastisch unabhängig*, wenn

$$P^{(X_1, \dots, X_n)} \left(\prod_{i=1}^n B_i \right) = \left(\bigotimes_{i=1}^n P^{X_i} \right) \left(\prod_{i=1}^n B_i \right) \quad (13.2)$$

$$\forall B_i \in \mathcal{B}_i, i = 1, \dots, n.$$

Offenbar ist Gleichung (13.2), also die stochastische Unabhängigkeit der X_i äquivalent zu

$$P\{X_1 \in B_1, \dots, X_n \in B_n\} = \prod_{i=1}^n P\{X_i \in B_i\} \quad (13.3)$$

$$\forall B_i \in \mathcal{B}_i, i = 1, \dots, n.$$

Wie im diskreten Fall sieht man: Definiert man auf dem W-Raum $(\Omega, \mathcal{A}, P) = (\prod_{i=1}^n \mathcal{X}_i, \prod_{i=1}^n \mathcal{B}_i, \bigotimes_{i=1}^n P_i)$

ZV $X_i := \pi_i \forall i$, so haben diese X_i die Verteilung P_i und sind stochastisch unabhängig. Den Raum (Ω, \mathcal{A}, P) nennt man oft Hintergrundraum. Seine spezielle Struktur spielt dabei meist keine Rolle. Man kann also zu vorgegebenen Randverteilungen, wie im diskreten Fall einen Hintergrund-W-Raum konstruieren und darauf stochastisch unabhängige ZV X_i mit $X_i \sim P_i$.

Man kann zeigen, dass es auch für ∞ viele W-Räume $(\mathcal{X}_i, \mathcal{B}_i, P_i)$, $i \geq 1$ ein Produktmaß $\bigotimes_{i=1}^{\infty} P_i$ gibt, sowie stochastisch unabhängige ZV $X_i, i \geq 1$ mit $X_i \sim P_i$, s. [BN]:. Es gilt bei Vorliegen von R-Dichten

13.3 Lemma. [BN]:22.6 Seien X_1, \dots, X_n reelle, st.u. ZV mit R-Dichten f_1, \dots, f_n von P^{X_1}, \dots, P^{X_n} .

Dann ist $\prod_{i=1}^n f_i(x_i)$ eine R-Dichte der gemeinsamen Verteilung $P^{(X_1, \dots, X_n)}$.

Beweis: Für jedes n-dimensionale Intervall $(a, b] = \prod_{i=1}^n (a_i, b_i]$ gilt für $X = (X_1, \dots, X_n)$

$$\begin{aligned} P^X(a, b] &= P^{X_1}(a_1, b_1] \cdots P^{X_n}(a_n, b_n] \\ &= \int_{a_1}^{b_1} f_1(x_1) dx_1 \cdots \int_{a_n}^{b_n} f_n(x_n) dx_n \\ &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_1(x_1) \cdots f_n(x_n) dx_1 \cdots dx_n, \end{aligned}$$

d.h. die gemeinsame Verteilung P^X hat die R-Dichte

$$(x_1, \dots, x_n) \rightarrow f_1(x_1) \cdots f_n(x_n).$$

Dies ist natürlich auch hinreichend für die stochastische Unabhängigkeit der ZV'n X_1, \dots, X_n .

□

Für die Verteilungsfunktion von $X_1 + X_2$, d.h. von $P^{(X_1+X_2)}$ ergibt sich somit

$$\begin{aligned} P^{(X_1+X_2)}(-\infty, t] &= P^{(X_1+X_2)}\{(x_1, x_2) : x_1 + x_2 \leq t\} \\ &= \int \int_{x_1+x_2 \leq t} f_1(x_1) f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{t-x_1} f_1(x_1) f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \left(\int_{-\infty}^{t-x_1} f_2(x_2) dx_2 \right) dx_1 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \left(\int_{-\infty}^t f_2(y-x_1) dy \right) dx_1 \\ &= \int_{-\infty}^t \int_{-\infty}^{\infty} f_1(x_1) f_2(y-x_1) dx_1 dy. \end{aligned}$$

Hieraus folgt, dass die Verteilung von $X_1 + X_2$ die Dichte

$$y \rightarrow h(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y-x) dx \quad (13.4)$$

besitzt. Die Verteilung der Summe $X_1 + \dots + X_n$ bezeichnen wir hier wie im diskreten Fall als *Faltung* der Verteilungen P^{X_i} . Das in (13.4) auftretende Integral heißt *Faltungsformel für R-Dichten* von Verteilungen unabhängiger ZV'r.

13.4 Beispiel. [BN]:22.9 Seien X, Y st.u., $X \sim \mathcal{N}(a, \sigma^2), Y \sim \mathcal{N}(b, \tau^2)$, so gilt $X + Y \sim \mathcal{N}(a+b, \sigma^2 + \tau^2)$.

Beweis: Die Dichte von $\mathcal{N}(0, 1)$ ist bekanntlich $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

1. Sei zunächst $a = b = 0, \sigma^2 = 1, \tau^2 > 0$

$$\begin{aligned} g(z) &= \int_{-\infty}^{\infty} \varphi(z-y) \frac{1}{\tau} \varphi\left(\frac{y}{\tau}\right) dy \\ &= \frac{1}{2\pi\tau} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-y)^2 - \frac{1}{2}\frac{y^2}{\tau^2}} dy \\ &\stackrel{c:=1+\frac{1}{\tau^2}}{=} \frac{1}{2\pi\tau} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-\frac{z}{c})^2 c} dy e^{-\frac{1}{2}z^2(1-\frac{1}{c})} \\ &= \frac{1}{\sqrt{2\pi}\tau} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{c}}}}_{=1} e^{-\frac{1}{2}(y-\frac{z}{c})^2 c} dy e^{-\frac{1}{2}z^2(1-\frac{1}{c})} \sqrt{\frac{1}{c}} \\ &= \frac{1}{\sqrt{2\pi(1+\tau^2)}} e^{-\frac{1}{2}z^2 \frac{1}{1+\tau^2}}. \end{aligned}$$

Dies ist die Dichte einer $\mathcal{N}(0, 1 + \tau^2)$ -Verteilung.

$$2. Z \sim \mathcal{N}(a, \sigma^2) \Rightarrow tZ + b \sim \mathcal{N}(ta + b, t^2\sigma^2)$$

Beweis: Sei $t > 0$. Dann gilt

$$\begin{aligned} P\{tZ + b \leq z\} &= P\left\{Z \leq \frac{z-b}{t}\right\} \\ &= \int_{-\infty}^{\frac{z-b}{t}} \varphi\left(\frac{x-a}{\sigma}\right) \frac{1}{\sigma} dx \\ &= \int_{-\infty}^z \frac{1}{\tau} \varphi\left(\frac{y-\mu}{\tau}\right) dy, \end{aligned}$$

wobei im letzten Schritt $\tau^2 := t^2\sigma^2$ und $\mu := ta + b$ gesetzt wurden. Eine ähnliche Rechnung gilt für $t < 0$.

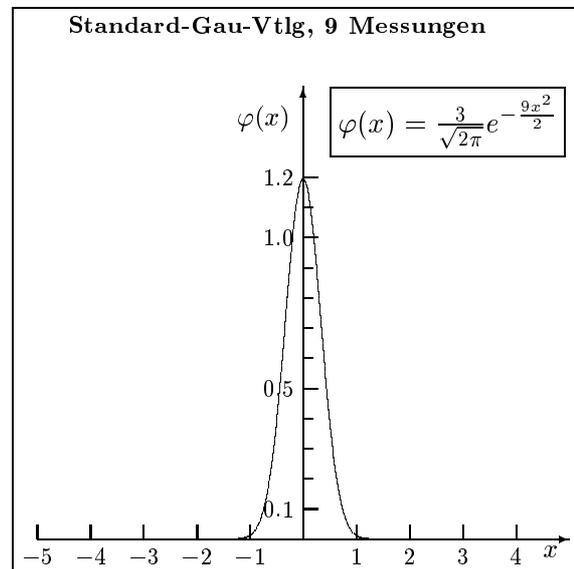
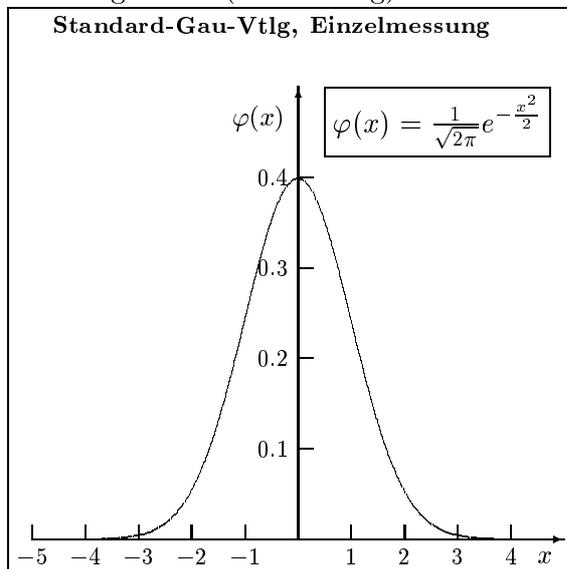
3. Allgemeiner Fall:

$$X + Y = \underbrace{\left(\underbrace{\frac{X-a}{\sigma}}_{\sim \mathcal{N}(0,1)} + \underbrace{\frac{Y-b}{\sigma}}_{\sim \mathcal{N}(0, \frac{\tau^2}{\sigma^2})}\right)}_{\sim \mathcal{N}(0, 1 + \frac{\tau^2}{\sigma^2})} \sigma + (a+b) \sim \mathcal{N}(a+b, \sigma^2 + \tau^2) \quad \square$$

Folgerung: Sind X_1, \dots, X_n st.u. mit $X_i \sim \mathcal{N}(a, \sigma^2)$, dann ist

$$\frac{X_1 + \dots + X_n}{n} =: \bar{X}_n \sim \mathcal{N}\left(a, \frac{1}{n}\sigma^2\right),$$

das heißt also, da die Werte der Durchschnittsverteilung dichter am Mittelwert a liegen, als die Einzelergebnisse (s. Zeichnung).



Eines der wichtigsten Hilfsmittel zur Berechnung von R-Dichten ist die Transformationsformel für R-Dichten.

13.5 Satz. [BN]:23.3 (Transformationsatz für R-Dichten)

Gegeben sei ein W -Maß P über $(\mathbb{R}^k, \mathbb{B}^k)$ mit R -Dichte f und $f = 0$ auf G^c , G offen. Es sei $G := \bigcup_{i \in I} G_i$, G_i offen, I abzählbar. Weiter seien $g_i = (g_{i1}, \dots, g_{ik}) : G_i \rightarrow \mathbb{R}^k$ stetig differenzierbar, injektiv mit

$$\Delta_{g_i} := \det \begin{pmatrix} \frac{\partial g_{i1}}{\partial x_1} & \dots & \frac{\partial g_{i1}}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial g_{ik}}{\partial x_1} & \dots & \frac{\partial g_{ik}}{\partial x_k} \end{pmatrix} \neq 0$$

auf G_i . Dann hat die induzierte Verteilung P^g für $g := \sum_{i \in I} g_i 1_{G_i}$, eine R -Dichte h gemäß

$$h(y) = \sum_{i \in I} \frac{f(g_i^{-1}(y))}{|\Delta_{g_i}(g_i^{-1}(y))|} 1_{g_i(G_i)}(y) \quad (13.5)$$

für $y \in \mathbb{R}^k$.

Beweis: siehe [BN].

13.6 Beispiel. [BN]:23.6 $X = (X_1, \dots, X_n)^T$ sei \mathbb{R}^k -wertig mit R -Dichte f . A sei eine invertierbare $k \times k$ -Matrix, $a \in \mathbb{R}^k$. Dann besitzt $Y = AX + a$ die R -Dichte

$$h(y) = f(A^{-1}(y - a)) \frac{1}{|\det A|} \quad (13.6)$$

für $y \in \mathbb{R}^k$.

Beweis: Satz 13.5 mit $|I| = 1$, $G = \mathbb{R}^k$, $g : G \rightarrow \mathbb{R}^k$, $g(x) = Ax + a$, $g(G) = \mathbb{R}^k$, $\Delta_g(x) = \det A \neq 0$, $g^{-1}(y) = A^{-1}(y - a)$. \square

Man erhält also erneut die im Beweis von Beispiel 13.4 hergeleitete Aussage:

$$X \sim \mathcal{N}(0, 1) \Rightarrow \sigma X + a \sim \mathcal{N}(a, \sigma^2), \sigma \neq 0.$$

13.7 Beispiel. [BN]:23.7 (k -dimensionale Normalverteilung) Seien X_1, \dots, X_k st.u., $X_i \sim \mathcal{N}(0, 1) \quad \forall i$, dann ist eine R -Dichte von $(X_1, \dots, X_k)^T$ gegeben durch

$$f(x) = \frac{1}{\sqrt{2^k \pi^k}} e^{-\frac{1}{2} x^T x}.$$

$g(x) = Ax + a$ wie in (13.6) liefert eine Dichte h von $Y = AX + a$ gemäß

$$\begin{aligned} h(y) &= \frac{1}{\sqrt{2^k \pi^k}} e^{-\frac{1}{2} (A^{-1}(y-a))^T (A^{-1}(y-a))} \frac{1}{|\det A|} \\ &= \frac{1}{\sqrt{2^k \pi^k}} e^{-\frac{1}{2} (y-a)^T (A^{-1})^T A^{-1} (y-a)} \frac{1}{|\det A|} \\ &= \frac{1}{\sqrt{2^k \pi^k}} e^{-\frac{1}{2} (y-a)^T (AA^T)^{-1} (y-a)} \frac{1}{\sqrt{\det(AA^T)}} \\ &= \frac{1}{\sqrt{2^k \pi^k \det S}} e^{-\frac{1}{2} (y-a)^T S^{-1} (y-a)} \end{aligned} \quad (13.7)$$

mit $y \in \mathbb{R}$, $S := AA^T$.

Eine Verteilung über $(\mathbb{R}^k, \mathbb{B}^k)$ mit R -Dichte (13.7) heißt Normalverteilung $\mathcal{N}(a, S)$ mit Mittelwertvektor a und Kovarianzmatrix S .

S ist positiv definit, d.h. $x^T S x > 0 \quad \forall x \neq 0$, denn:

$$x^T S x = x^T A A^T x = (A^T x)^T A^T x = \|A^T x\|^2 > 0$$

da $A^T x \neq 0$.

S ist symmetrisch, $S^T = S$, denn $(A A^T)^T = A A^T$.

In der linearen Algebra wird gezeigt: Jede positiv definite symmetrische Matrix S lässt sich in der Form $S = A A^T$ mit geeignetem A , A invertierbar, schreiben. Da a beliebig ist, wird durch (13.7) genau dann eine Dichte einer $\mathcal{N}(a, S)$ -Verteilung definiert, wenn $a \in \mathbb{R}^k$ und S positiv definit ist.

13.8 Korollar. [BN]:23.8 Ist Y eine \mathbb{R}^k -wertige ZV mit $Y \sim \mathcal{N}(a, S)$ und $b \in \mathbb{R}^k$ sowie B eine invertierbare $k \times k$ -Matrix, so folgt

$$B Y + b = \mathcal{N}(B a + b, B S B^T) \quad (13.8)$$

Beweis: Mit $X = (X_1, \dots, X_k)^T, X_1, \dots, X_k$ st.u., $X_i \sim \mathcal{N}(0, 1)$, gilt

$$\begin{aligned} Y &\sim A X + a, \text{ also } B Y + b &\sim B(A X + a) + b \\ &= (B A) X + B a + b \\ &\sim \mathcal{N}(B a + b, (B A)(B A)^T) \\ &= \mathcal{N}(B a + b, B A A^T B^T) \\ &= \mathcal{N}(B a + b, B S B^T). \quad \square \end{aligned}$$

13.9 Lemma. [BN]:23.9 Sei Y eine \mathbb{R}^k -wertige ZV mit $Y \sim \mathcal{N}(a, S)$ und $c := (c_1, \dots, c_k)^T \neq 0$. Dann gilt

$$c^T Y = c_1 Y_1 + \dots + c_k Y_k \sim \mathcal{N}(c^T a, c^T S c) \quad (13.9)$$

und insbesondere für $c = (0, \dots, 0, 1, 0, \dots, 0)^T$, wobei die 1 an der i -ten Stelle steht,

$$Y_i \sim \mathcal{N}(a_i, S_{ii}). \quad (13.10)$$

Beweis:

$$\begin{aligned}
c^T Y &\sim c^T (AX + a) \\
&= (c^T A)X + c^T a \\
&= \sum_{j=1}^k \left(\sum_{i=1}^k c_i a_{ij} \right) X_j + c^T a \\
&\sim \mathcal{N} \left(c^T a, \sum_{j=1}^k \left(\sum_{i=1}^k c_i a_{ij} \right)^2 \right) \\
&= \mathcal{N} \left(c^T a, \sum_{i=1}^k \sum_{l=1}^k c_i c_l \left(\sum_{j=1}^k a_{ij} a_{lj} \right) \right) \\
&= \mathcal{N} \left(c^T a, \sum_{i=1}^k \sum_{l=1}^k c_i c_l (AA^T)_{il} \right) \\
&= \mathcal{N}(c^T a, c^T S c). \quad \square
\end{aligned}$$

13.10 Beispiel. [BN]:23.4 Sei $X \sim \mathcal{N}(0, 1)$. Was ist $\mathcal{L}(X^2)$?

$$\begin{aligned}
\varphi(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \\
g(x) &= x^2 \quad ; \quad g \text{ ist nicht eindeutig} \\
G_1 &:= (-\infty, 0) & g_1(x) &:= x^2 & g_1 : G_1 &\rightarrow g_1(G_1) = (0, \infty) \\
G_2 &:= (0, \infty) & g_2(x) &:= x^2 & g_2 : G_2 &\rightarrow g_2(G_2) = (0, \infty) \\
G &:= G_1 \cup G_2 \\
P &= \mathcal{N}(0, 1) \Rightarrow P(G^c) = P(\{0\}) = 0
\end{aligned}$$

Mit φ ist auch die undefinierte Funktion $f := \varphi|_{\{0\}^c}$ eine R-Dichte von P .

Gesucht ist P^g : $\Delta_{g_1}(x) = 2x = \Delta_{g_2}(x)$ und $g_1^{-1}(y) = -\sqrt{y}$, $g_2^{-1}(y) = \sqrt{y}$. Damit folgt als R-Dichte von P^g :

$$\begin{aligned}
h(y) &= f(g_1^{-1}(y)) \frac{1}{|\Delta_{g_1}(g_1^{-1}(y))|} 1_{(0, \infty)}(y) + f(g_2^{-1}(y)) \frac{1}{|\Delta_{g_2}(g_2^{-1}(y))|} 1_{(0, \infty)}(y) \\
&= \left(\varphi(-\sqrt{y}) \frac{1}{2\sqrt{y}} + \varphi(\sqrt{y}) \frac{1}{2\sqrt{y}} \right) 1_{(0, \infty)}(y) \\
&= \frac{\varphi(\sqrt{y})}{\sqrt{y}} 1_{(0, \infty)}(y) \\
&= \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} 1_{(0, \infty)}(y)
\end{aligned}$$

Das heißt

$$h(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} 1_{(0, \infty)}(y) \tag{13.11}$$

ist λ - und auch R-Dichte von X^2 . $\mathcal{L}(X^2)$ heißt χ_1^2 -Verteilung.

Sind X_1, \dots, X_n i.i.d.¹ $\sim \mathcal{N}(0, 1)$, so heißt

$$\mathcal{L}(X_1^2 + \dots + X_n^2) \quad (13.12)$$

die χ_n^2 -Verteilung oder auch χ^2 -Verteilung mit n Freiheitsgraden.

14 Koppelung von Experimenten

Wir betrachten in diesem Kapitel W-Maße auf Produkträumen $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_n$, wobei wir der Einfachheit annehmen, dass die \mathcal{X}_i abzählbar sind. Ist P ein W-Maß auf \mathcal{X} , so gilt für festes $x = (x_1, \dots, x_n) \in \mathcal{X}$ mit $P\{x\} > 0$ gemäß der Multiplikationsregel (6.3) mit Mengen $A_i = \{y = (y_1, \dots, y_n) \in \mathcal{X} : y_i = x_i\}$,

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}),$$

d.h. mit der Zähldichte $f(x) = P\{x\}$ von P und den sog. **Übergangszähldichten**

$$f_i^{i-1}(x_1, \dots, x_{i-1}; x_i) := P(A_i|A_1 \cap \dots \cap A_{i-1}), i \geq 2$$

sowie der **Startzähldichte** $f_1(x_1) := P(A_1)$ gilt

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot f_2^1(x_1; x_2) \cdot \dots \cdot f_n^{n-1}(x_1, \dots, x_{n-1}; x_n). \quad (14.1)$$

Man sieht sofort, dass die Übergangsdichten $f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$ tatsächlich bei festen Werten der (x_1, \dots, x_{i-1}) Zähldichten von W-Maßen auf \mathcal{X}_i sind.

Umgekehrt gilt, wenn für alle $i \geq 2$ und alle (x_1, \dots, x_{i-1}) solche Übergangsdichten *vorgegeben* sind, sowie eine Start-Zähldichte $f_1(x_1)$, dass durch (14.1) eine Zähldichte auf dem Produktraum \mathcal{X} und damit ein Maß P definiert wird. Dies ist eine der gängigsten Methoden zur Konstruktion von Verteilungen auf Produkträumen.

Ist $X = (X_1, \dots, X_n)$ eine ZV mit Werten im Produktraum \mathcal{X} (auf einem beliebigen Hintergrundraum), deren Verteilung das eben konstruierte W-Maß P ist, $\mathbb{P}^X = P$, so gilt in offensichtlicher Notation

$$\mathbb{P}(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) = f_i^{i-1}(x_1, \dots, x_{i-1}; x_i), i \geq 2,$$

und $\mathbb{P}\{X_1 = x_1\} = f_1(x_1)$. Wie schon früher diskutiert, ist als Hintergrundraum (Ω, \mathbb{P}) z.B. der Raum (\mathcal{X}, P) selbst wählbar mit ZV'n $X_i = \pi_i$, den Projektionen auf die i -te Komponente.

Hat man auf die oben dargestellte Weise das W-Maß $P_n := P$ auf $\mathcal{X}^{(n)} := \prod_{i=1}^n \mathcal{X}_i$ konstruiert mit

¹independent, identically distributed = stochastisch unabhängig mit gleicher Verteilung

Zähldichte $f_n := f$ und ist $f_{n+1}^n(x_1, \dots, x_n; x_{n+1})$ eine weitere Übergangszähldichte auf einem abzählbaren Raum \mathcal{X}_{n+1} so läßt sich auf $\mathcal{X}^{(n+1)} = \prod_{i=1}^{n+1} \mathcal{X}_i$ gemäß

$$f_{n+1}(x_1, \dots, x_n, x_{n+1}) = f_n(x_1, \dots, x_n) f_{n+1}^n(x_1, \dots, x_n; x_{n+1})$$

wie in (14.1) eine Zähldichte f_{n+1} auf $\mathcal{X}^{(n+1)}$ definieren mit zugehörigem W-Maß P_{n+1} .

Dieses Verfahren kann beliebig oft wiederholt werden. Nach dem Satz von Ionescu-Tulcea, s. [BN]:40.4, kann man sogar zeigen, dass dies unendlich oft durchführbar ist, d.h. bei ∞ vielen abzählbaren Räumen \mathcal{X}_i und gegebenen Übergangszähldichten f_i^{i-1} , $i \geq 2$, sowie einer Start-Zähldichte f_1 , gibt es auf dem Produktraum $\mathcal{X}^{(\infty)} = \prod_{i=1}^{\infty} \mathcal{X}_i$ ein W-Maß P_{∞} derart, dass die endlichen Randverteilungen

$$P_{\infty}(B^{(n)} \times \prod_{i=n+1}^{\infty} \mathcal{X}_i), \quad B^{(n)} \subset \mathcal{X}^{(n)}$$

genau die W-Maße P_n sind $\forall n$. In ZV'n ausgedrückt bedeutet dies, dass es ZV'n X_1, X_2, \dots gibt auf einem Hintergrundraum derart, dass $\forall n$ die Verteilung von (X_1, \dots, X_n) gerade P_n ist.

Wie oben kann der Hintergrundraum als der unendliche Produktraum gewählt werden, sowie die ZV'n als die Projektionen.

Wir merken einen schon bekannten Sonderfall an: *Stochastische Unabhängigkeit*.

Wenn die Übergangszähldichten $f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$ nicht von den Bedingungen x_1, \dots, x_{i-1} abhängen, also ganz gewöhnliche Zähldichten f_i sind:

$$f_i(x_i) := f_i^{i-1}(x_1, \dots, x_{i-1}; x_i),$$

so vereinfacht sich (14.1) zu

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n) \quad \forall x = (x_1, \dots, x_n).$$

Das bedeutet, dass das Maß $P_n = P$ auf $\mathcal{X}^{(n)}$ ein Produktmaß ist und die ZV'n X_1, \dots, X_n stochastisch unabhängig sind mit vorgegebenen Zähldichten f_1, \dots, f_n .

Dies gilt auch im Fall $n = \infty$. Folglich gibt es dann zu ∞ vielen Zähldichten f_1, f_2, \dots auch ∞ viele stochastisch unabhängige ZV X_1, X_2, \dots mit diesen Z-Dichten.

Ein weiterer in der Praxis sehr wichtiger Sonderfall ist der Fall der sog. Markov-Koppelung, bei dem die Übergangszähldichten nicht von allen bis dahin beobachteten Werten abhängen, sondern nur von der unmittelbaren Vergangenheit.

14.1 Definition. *Hängen bei einem mehrstufigen Experiment mit obiger Notation die Übergangszähldichten $f_i^{i-1}(x_1, \dots, x_{i-1}; x_i)$ nur von x_{i-1} ab, Schreibweise:*

$$f_i^{i-1}(x_{i-1}; x_i) \quad \forall i \geq 2, \tag{14.2}$$

so spricht man von einer **Markov-Kette**. Auch die zugehörigen ZV'n X_1, X_2, \dots werden Markov-Kette genannt. Meist sind die Räume \mathcal{X}_i für alle i gleich, z.B. gleich I . Wir sprechen dann von einer Markov-Kette mit **Zustandsraum** I . Für die gesamte Markov Kette schreiben wir (X_n) .

15 Monte Carlo Methode

Sei $g : [0, 1] \rightarrow \mathbb{R}$ stetig. Gesucht ist $\int_0^1 g(x) dx$.

Falls $X \sim \mathcal{R}(0, 1)$, so ist $Eg(X) = \int_0^1 g(x) dx$.

$$\text{Var } g(X) \leq Eg^2(X) = \int_0^1 g^2(x) dx \leq \sup_x |g(x)|^2 =: \|g\|^2.$$

Falls X_1, X_2, \dots i.i.d., $X_i \sim \mathcal{R}(0, 1)$ so gilt

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n g(X_i) - \int_0^1 g(x) dx \right| < \varepsilon \right\} \geq 1 - \frac{\text{Var } g(X)}{n\varepsilon} \quad (15.1)$$

Sind x_1, \dots, x_n beobachtete Werte von X_1, \dots, X_n , so nennt man diese Werte *Zufallszahlen* und verwendet $\frac{1}{n} \sum_{i=1}^n g(x_i)$ als Näherung für $\int_0^1 g(x) dx$.

Das sogenannte Anwendungspostulat besagt, daß Ereignisse großer Wahrscheinlichkeit in der Praxis mit Sicherheit eintreten. Ist die rechte Seite in (15.1) „groß“, wobei der Anwender vorgibt, was er als „groß“, also nahe bei 1 gelegen, ansieht, so tritt das Ereignis $\left| \frac{1}{n} \sum g(x_i) - \int_0^1 g(x) dx \right| < \varepsilon$ in der Praxis mit Sicherheit ein.

In der Realität ist X_i nicht exakt verteilt nach $\mathcal{R}(0, 1)$, sondern z.B. $X_i \sim$ Laplace über

$$\left\{ \frac{0}{m}, \dots, \frac{m-1}{m} \right\} =: I_m.$$

Damit ist für eine Funktion g mit integrierbarer Ableitung g' die Differenz $|Eg(X_i) - \int_0^1 g(x) dx| \leq \frac{1}{m} \int_0^1 |g'(x)| dx$ sehr klein, falls m sehr groß ist.

Statt (15.1) erhält man

$$\frac{1}{m^n} \left| \left\{ (x_1, \dots, x_n) \in I_m^n : \left| \frac{1}{n} \sum_{i=1}^n g(x_i) - Eg(X) \right| < \varepsilon \right\} \right| \geq 1 - \frac{\|g\|^2}{n\varepsilon^2}.$$

Ist z.B. $1 - \frac{\|g\|^2}{n\varepsilon^2} = 0.99$, so bedeutet dies: Für 99% aller Zahlenfolgen $x_1, \dots, x_n \in I_m$ liegt $\frac{1}{n} \sum g(x_i)$ in $Eg(X) \pm \varepsilon$. Deshalb ist für „regellose“ Zahlenfolgen x_1, \dots, x_n die Summe $\frac{1}{n} \sum g(x_i)$ eine Annäherung von $\int_0^1 g(x) dx$.

Pseudozufallszahlen: Beispiel: Kongruenzgeneratoren

Man wähle $m \in \mathbb{Z}$ (den Modul), $a \in \mathbb{Z}$ (den Faktor) und $r \in \mathbb{Z}$ (das Inkrement), wobei $0 \leq a, r, y_0 < m$ gelten soll. Nun berechnet man rekursiv eine Folge y_1, y_2, \dots gemäß

$$y_{i+1} := (ay_i + r) \bmod m$$

mit $0 \leq y_{i+1} < m$. Die Folge $x_i := \frac{y_i}{m}, i = 1, 2, \dots$, wird dann als Folge von Realisierungen von i.i.d. $\mathcal{R}(0, 1)$ -verteilten ZV X_1, X_2, \dots angesehen.

Kongruenzgeneratoren sind stets periodisch mit maximaler Periodenlänge m . Auf Binärcomputern wählt man meist $m = 2^k$ als Modul, wobei k die maximale Wortlänge des Rechners ist.

Durch geeignete Wahl von a und r kann man erreichen, daß bei beliebigem y_0 die Periodenlänge gleich m wird, d.h. es werden alle Zahlen x_i aus $\{\frac{0}{m}, \dots, \frac{m-1}{m}\}$ im Verlauf der Periode genau einmal erzeugt.

15.1 Bemerkung. (Erzeugung von reellen ZV mit beliebiger VF F aus $\mathcal{R}(0, 1)$ -verteilten)

Sei für eine VF $F : \mathbb{R} \rightarrow \mathbb{R}, F^{-1} : (0, 1) \rightarrow \mathbb{R}$ definiert gemäß

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}, 0 < u < 1 \quad (15.2)$$

die **Pseudoinverse** von F (=übliche Inverse, falls F invertierbar). [BN]:S.299 Dann folgt: Falls $U \sim \mathcal{R}(0, 1)$, so gilt $F^{-1}(U) \sim F$.

Beweis: F^{-1} ist monoton steigend, also meßbar. Folgende Implikationen gelten:

$$\begin{aligned} F^{-1}(u) &\leq x \\ \Rightarrow F(F^{-1}(u)) &\leq F(x) \\ &\Rightarrow u \leq F(x) \\ \Rightarrow F^{-1}(u) &\leq x \end{aligned}$$

Dabei wurde verwendet, daß $F(F^{-1}(u)) \geq u$, da F rechtsseitig stetig ist (wähle $x_n \downarrow F^{-1}(u)$ mit $F(x_n) \geq u$). Es folgt also

$$F^{-1}(u) \leq x \Leftrightarrow u \leq F(x) \quad \forall x \in \mathbb{R} \quad (15.3)$$

$\Rightarrow P\{F^{-1}(U) \leq x\} = P\{U \leq F(x)\} = F(x)$ für alle $x \in \mathbb{R}$. \square

Es werde z.B. eine ZV X mit der Exponential(α)-Verteilung gesucht, $\alpha > 0$.

$$\begin{aligned} f(x) &= \begin{cases} \alpha e^{-\alpha x} & \text{falls } x > 0 \\ 0 & \text{falls } x \leq 0 \end{cases} \\ F(x) &= \begin{cases} 1 - e^{-\alpha x} & \text{falls } x > 0 \\ 0 & \text{falls } x \leq 0 \end{cases} \end{aligned}$$

Sei $0 < u < 1$:

$u = F(x) = 1 - e^{-\alpha x} \Rightarrow -\frac{1}{\alpha} \ln(1 - u) = F^{-1}(u) = x \Rightarrow -\frac{1}{\alpha} \ln(1 - U) \sim \text{Exponential}(\alpha)$.

Entsprechend: $\Phi^{-1}(U) \sim \mathcal{N}(0, 1)$ und $\sigma\Phi^{-1}(U) + a \sim \mathcal{N}(a, \sigma^2)$.

16 Markov-Ketten – einige Grundbegriffe

Die nun folgenden Teile dieses Kurzschrifts sind im wesentlichen dem Buch von G.Hübner entlehnt. Wir wiederholen die Definition einer Markov-Kette, wobei wir jetzt aus traditionellen Gründen mit dem Zeitpunkt $n = 0$ statt $n = 1$ zu zählen beginnen

16.1 Definition. Eine Markov-Kette (MK) ist die Folge der Beobachtungen (bzw. der Koordinaten-Variablen) X_0, X_1, X_2, \dots in einem unendlich-stufigen Versuch mit Markov-Koppelung und abzählbarer Zustandsmenge I . Die Zufallsvariablen $X_n: \Omega \rightarrow I$ beschreiben also den Zustand eines Systems zu den Zeitpunkten $n = 0, 1, 2, \dots$.

Bemerkung: Der Begriff „Kette“ beschreibt sehr anschaulich die Markov-Eigenschaft, weil dabei jedes Kettenglied immer nur am unmittelbar vorangehenden hängt.

16.2 Definition. Eine Markov-Kette (X_n) heißt **homogen**, falls die Übergangswahrscheinlichkeiten (ÜZ-Dichten) $f_n^{n-1}(i; j) = P(X_n = j | X_{n-1} = i)$ für alle Zeitpunkte gleich sind. In diesem Fall schreibt man $p_{ij} := f_n^{n-1}(i; j)$.

Die Matrix $(p_{ij}) := (p_{ij}, i, j \in I)$ heißt Übergangsmatrix, kurz \ddot{U} -Matrix. Sie kann auch (abzählbar) unendlich viele Zeilen und Spalten besitzen. Die Zeilensumme ist stets $= 1$. (Solche Matrizen nennt man „stochastische Matrizen“.)

Wir werden uns im folgenden auf homogene Markov-Ketten (HMK) beschränken. Wir beginnen mit einem einfachen Beispiel.

16.3 Beispiel. Die Anzahl der belegten Telefonleitungen soll durch eine HMK modelliert werden. Zur Vereinfachung nehmen wir an, dass nur zwei Leitungen vorhanden sind. Es gibt also die Zustände $i = 0, 1, 2$. (Die Zeiteinheit sei z.B. 1 Minute.)

Für die Übergangsmatrix (p_{ij}) nehmen wir die folgenden Werte an:

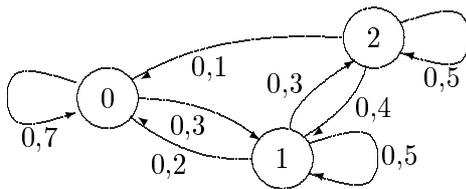
$$(p_{ij}) = \begin{pmatrix} 0,7 & 0,3 & 0,0 \\ 0,2 & 0,5 & 0,3 \\ 0,1 & 0,4 & 0,5 \end{pmatrix}$$

Der Wert $p_{12} = 0,3$ (z.B.) ist dann die (bedingte) Wahrscheinlichkeit, dass zur Zeit n zwei Leitungen belegt sind, falls zur Zeit $n - 1$ eine Leitung belegt war.

Das Übergangsverhalten einer HMK kann man oft noch übersichtlicher darstellen durch den sogenannten Übergangs-Graph:

16.4 Definition. Ein **Übergangs-Graph** (kurz \ddot{U} -Graph) einer HMK besteht aus allen möglichen Zuständen als Ecken des und den mit positiver Wahrscheinlichkeit möglichen Übergängen als gerichtete Kanten (Pfeile). An der Kante von i nach j wird jeweils der Wert p_{ij} notiert.

Für das obige Beispiel 16.3 sieht also der Ü-Graph wie folgt aus. Man beachte auch diejenigen Pfeile, die zum selben Zustand zurückkehren.



Zur Beschreibung des Ablaufs einer (homogenen) Markov-Kette benötigt man neben der Ü-Matrix oder dem Ü-Graph noch einen (festen) Startpunkt $i_0 \in I$ oder – für den Fall eines zufälligen Startpunkts – eine sogenannte **Startverteilung**, nämlich eine Z-Dichte ($P(X_0 = i)$, $i \in I$). Dann ist die Wahrscheinlichkeit für jede endliche Zustandsfolge festgelegt durch

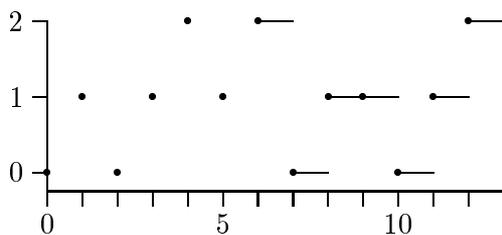
$$P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0)p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \quad (16.1)$$

Bemerkung. Ein fester Startpunkt $i_0 \in I$ wird ebenfalls durch eine Startverteilung beschrieben. Es ist dann $P(X_0 = i_0) = 1$ und $P(X_0 = i) = 0$ sonst.

16.5 Definition. Ein einzelner Verlauf einer Markov-Kette für einen festen Wert ω aus dem Hintergrundraum, also $(X_0(\omega), X_1(\omega), \dots)$ heißt ein **Pfad** der Markov-Kette.

Für das Beispiel 16.3 könnte ein Pfad etwa wie unten dargestellt aussehen. Aus optischen Gründen wird der Pfad oft so gezeichnet, dass die Zustände auch zwischen den Zeitpunkten $0, 1, 2, 3, \dots$ erhalten bleiben (hier in der rechten Hälfte dargestellt).

Man kann sich dazu vorstellen, dass der Prozess in stetiger Zeit $t \in [0, \infty)$ abläuft, dass aber Zustandsänderungen nur in ganzzahligen Zeitpunkten stattfinden.



Meist interessiert bei einem Prozess nicht die Wahrscheinlichkeit für eine (lückenlose) Zustandsfolge wie in (16.1), sondern die Wahrscheinlichkeit $P(X_n = i_n)$ für einen festen Zeitpunkt n .

Diesen Wert kann man im Prinzip direkt aus (16.1) durch Bildung der n -ten Randverteilung berechnen, d.h. man muss (16.1) über alle $i_0, i_1, \dots, i_{n-1} \in I$ aufsummieren. Dieser Weg ist aber sehr aufwendig und für große Werte von n technisch kaum durchführbar. Deshalb geht man besser induktiv vor.

Für $n = 1$ erhält man auf diesem Weg direkt (mit $i_0 = i, i_1 = j$)

$$P(X_1 = j) = \sum_{i \in I} P(X_0 = i, X_1 = j) = \sum_{i \in I} P(X_0 = i) \cdot p_{ij}$$

Dies geht natürlich ebenso, wenn 0 und 1 durch $n - 1$ und n ersetzt werden:

$$P(X_n = j) = \sum_{i \in I} P(X_{n-1} = i, X_n = j) = \sum_{i \in I} P(X_{n-1} = i) \cdot p_{ij} \quad (16.2)$$

Damit erhält man die Verteilung von X_n in $n - 1$ gleichartigen Schritten, die der Multiplikation eines Vektors mit einer Matrix entsprechen.

Wir notieren die beiden wichtigsten Formeln (16.1) und (16.2) nochmals gemeinsam:

16.6 Bemerkung. Für eine homogene Markov-Kette mit \ddot{U} -Matrix $(p_{ij})_{i,j \in I}$ und Startverteilung $(P(X_0 = i), i \in I)$ gilt

$$P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0) p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \quad (16.1)$$

$$P(X_n = j) = \sum_{i \in I} P(X_{n-1} = i) \cdot p_{ij} \quad (16.2)$$

Wir stellen nun noch einige ergänzende Begriffe und Eigenschaften zusammen, die wir im folgenden nur zum Teil benötigen.

16.7 Definition. (und Folgerung) (n -Schritt Übergangsmatrix) Für eine HMK (X_n) ist die Matrix $(p_{ij}^{(n)})$ mit $p_{ij}^{(n)} := P(X_{m+n} = j | X_m = i)$ ($i, j \in I$) unabhängig von m und heißt n -Schritt-Übergangsmatrix. Analog zu (16.2) kann man zeigen, dass gilt:

$$p_{ij}^{(n)} = \sum_{k \in I} p_{ik}^{(n-1)} p_{kj}, \quad i, j \in I, \quad \text{mit } p_{ik}^{(1)} = p_{ik}. \quad (16.3)$$

Mit den Abkürzungen $\mathbb{P} := (p_{ij})$, $\mathbb{P}^{(n)} := (p_{ij}^{(n)})$ folgt hieraus $\mathbb{P}^{(n)} = \mathbb{P}^{(n-1)} \cdot \mathbb{P}$ (mit Matrix-Multiplikation) und durch Induktion $\mathbb{P}^{(n)} = \mathbb{P} \cdot \mathbb{P} \cdots \mathbb{P} = \mathbb{P}^n$. Die n -Schritt-Übergangsmatrix ist also die n -te Potenz der \ddot{U} -Matrix (p_{ij}) . Die Folgerung

$$p_{ij}^{(n+m)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}^{(m)}, \quad (i, j) \in I, \quad n, m \in \{1, 2, \dots\} \quad (16.4)$$

ist auch als Chapman-Kolmogorov-Gleichung bekannt.

16.8 Definition. (Zerlegung in Klassen, irreduzibel) Die Zustandsmenge I einer HMK wird in (disjunkte) **Klassen** zerlegt durch die Vorschrift, dass zwei Zustände i und j zur selben Klasse gehören sollen, wenn $i = j$ oder wenn der Zustand j ausgehend von i in endlich vielen Schritten mit positiver Wahrscheinlichkeit erreicht werden kann ($i \rightsquigarrow j$) und umgekehrt i von j aus „erreichbar“ ist ($j \rightsquigarrow i$). Jeder Zustand $i \in I$ gehört zu genau einer Klasse k . Eine HMK heißt **irreduzibel**, falls alle Zustände zur selben Klasse gehören.

Die einfachsten Beispiele reduzibler (= zerlegbarer) HMK besitzen zwei Zustände, z.B. $I = \{1, 2\}$, und \ddot{U} -Matrizen der Form $\begin{pmatrix} 1-a & a \\ 0 & 1 \end{pmatrix}$ mit $0 \leq a \leq 1$. Es gibt zwei einelementige Klassen $\{1\}$ und $\{2\}$, weil man vom Zustand 2 nicht zum Zustand 1 kommt.

16.9 Definition. (Periode, aperiodisch) *Eine Klasse K heißt **periodisch** mit **Periode** d , wenn es d (≥ 2) disjunkte Teilmengen in K gibt, die der Reihe nach in d Schritten durchlaufen werden. Eine HMK heißt **aperiodisch**, wenn es keine periodische Klasse gibt. Das einfachste Beispiel einer periodischen HMK besteht (wieder) aus zwei Zuständen und besitzt eine \ddot{U} -Matrix (p_{ij}) der Form $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Der Prozess springt immer zwischen den beiden Zuständen hin und her, hat also Periode 2.*

16.1 Markov-Ketten im Gleichgewicht

In der Praxis ist die Bestimmung des mittelfristigen Verhaltens einer Markov-Kette wegen des Rechenaufwands und der großen Datenmengen meist nicht möglich oder nicht sinnvoll. Es müssten ja für jeden Zeitpunkt n und jeden Zustand j die Wahrscheinlichkeiten $P(X_n = j)$ berechnet und ausgewertet werden. Dagegen liefert das einfacher zu bestimmende Langzeitverhalten oft ausreichende Information.

Beobachtet man den Ablauf realer Prozesse über längere Zeit, dann stellt man häufig fest, dass immer wieder dieselben Zustände auftreten und dass jeder dieser Zustände eine ziemlich konstante – vom Zustand abhängige – relative Häufigkeit besitzt.

16.10 Beispiel. *Das eben beschriebene Verhalten kann man z.B. an der Länge einer Warteschlange beobachten (die wir später genauer modellieren und untersuchen werden), S.126 im Buch von Hübner.*

Man sagt, dass ein Prozess mit einem solchen Verhalten (ohne Berücksichtigung der Anlaufphase) „im Gleichgewicht“ ist.

Genaugenommen schließt die nachfolgende Definition für „Markov-Ketten im Gleichgewicht“ gewisse periodische Effekte aus, die bei der obigen verbalen Charakterisierung durch „konstante relative Häufigkeiten“ noch möglich sind.

Solche Effekte treten aber bei den hier vorgesehenen Anwendungen nicht auf. (Gegebenenfalls können sie durch zeitliche Mittelung über volle Perioden beseitigt werden.)

16.11 Definition. (Markov-Kette im Gleichgewicht) *(a) Eine homogene Markov-Kette (X_n) ist im **Gleichgewicht**, wenn für alle Zustände $i \in I$ die Wahrscheinlichkeiten $P(X_n = i)$ unabhängig vom Zeitpunkt n sind. Man setzt dann $\pi_i := P(X_n = i)$ und bezeichnet die Z -Dichte $(\pi_i, i \in I)$ als **Gleichgewichtsverteilung** (GGV) der HMK (X_n) .*

(b) Allgemein heißt $(\pi_i, i \in I)$ eine GGV (zur \ddot{U} -Matrix (p_{ij})), wenn eine aus (π_i) als Startverteilung und der \ddot{U} -Matrix (p_{ij}) konstruierte HMK (X_n^i) im Gleichgewicht ist.

Bemerkungen:

1. Es hat sich hier eine laxe Sprechweise durchgesetzt. Genauer wäre $(\pi_i, i \in I)$ „die Z-Dichte einer Gleichgewichtsverteilung“.
2. Statt „Prozess (oder HMK) im Gleichgewicht“ sagt man auch „stationäre(r) Prozess (oder HMK)“. Entsprechend heißt $(\pi_i, i \in I)$ auch „stationäre Verteilung“.

Es ist jetzt zu klären, ob es zu einer gegebenen \ddot{U} -Matrix (p_{ij}) eine GGV (π_i) gibt und wie man diese bestimmen kann. Die Frage, ob sich bei einer vorgegebenen Startverteilung langfristig ein Gleichgewicht einstellen wird, werden wir etwas später ansprechen.

16.12 Satz. Berechnung der Gleichgewichts-Verteilung (a) Die HMK (X_0, X_1, \dots) mit \ddot{U} -Matrix $(p_{ij}, i, j \in I)$ sei im Gleichgewicht, d.h. es gelte $P(X_n = i) = \pi_i$ für alle $n = 0, 1, 2, \dots$ und $i \in I$.

Wegen (16.2) $P(X_n = j) = \sum_{i \in I} P(X_{n-1} = i) p_{ij}$ gelten dann für die Werte $\pi_i, i \in I$, die folgenden beiden **Gleichgewichtsbedingungen**

$$\pi_j = \sum_{i \in I} \pi_i p_{ij} \quad \text{für alle } j \in I, \tag{G}$$

$$\pi_j \geq 0 \quad \text{für alle } j \in I \quad \text{und} \quad \sum_{j \in I} \pi_j = 1 \tag{N}$$

(Die Bezeichnungen (G) und (N) bedeuten hier „Gleichgewicht“ und „Normierung“.)

(b) Gilt (G) und (N) für $(\pi_j, j \in I)$, dann ist die HMK mit Startverteilung $(\pi_j, j \in I)$ und \ddot{U} -Matrix $(p_{ij}, i, j \in I)$ im Gleichgewicht.

(c) Zu der \ddot{U} -Matrix $(p_{ij}, i, j \in I)$ gibt es genau dann keine/genau eine/mehrere Gleichgewichtsverteilung(en), wenn die Bedingungen (G) und (N) keine/genau eine/mehrere Lösung(en) besitzen.

Der Beweis von (a) ist einfach, wenn man (16.2) mit (G) vergleicht. Ebenso folgt (b) mit (16.2) induktiv, und (c) folgt aus (a) und (b). \square

16.13 Beispiel. Für das Beispiel 16.3 aus dem vorangehenden Abschnitt mit $I = \{0, 1, 2\}$ und der \ddot{U} -Matrix

$$(p_{ij}) = \begin{pmatrix} 0,7 & 0,3 & 0 \\ 0,2 & 0,5 & 0,3 \\ 0,1 & 0,4 & 0,5 \end{pmatrix}$$

lauten die Gleichgewichtsbedingungen:

$$(G_0) \quad \pi_0 = 0,7\pi_0 + 0,2\pi_1 + 0,1\pi_2$$

$$(G_1) \quad \pi_1 = 0,3\pi_0 + 0,5\pi_1 + 0,4\pi_2 \quad \text{und} \quad (N) \quad \pi_0 + \pi_1 + \pi_2 = 1$$

$$(G_2) \quad \pi_2 = 0,3\pi_1 + 0,5\pi_2$$

(Man beachte, dass spaltenweise, nicht zeilenweise, addiert werden muss.)

Als eindeutige Lösung erhält man daraus

$$\pi_0 = \frac{13}{37} \approx 0,35, \quad \pi_1 = \frac{15}{37} \approx 0,41, \quad \pi_2 = \frac{9}{37} \approx 0,24$$

Im Gleichgewicht sind also mit der Wahrscheinlichkeit $\pi_2 = 24\%$ beide Leitungen belegt.

Bei großer Zustandsmenge I und einem günstig strukturierten \ddot{U} -Graph ist zur Lösung des Gleichungssystems (G) oft das folgende „Schnittprinzip“ hilfreich:

16.14 Satz. (Schnittprinzip) (a) Zerlegt man die Zustandsmenge I einer HMK in Teilmengen K und K^c , dann folgt aus der Gleichgewichtsbedingung (G)

$$\sum_{i \in K} \sum_{j \in K^c} \pi_i p_{ij} = \sum_{i \in K} \sum_{j \in K^c} \pi_j p_{ji}, \quad (16.5)$$

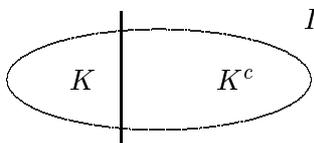
d.h. die Wahrscheinlichkeit eines \ddot{U} bergangs von K nach K^c ist im Gleichgewicht ebenso groß wie die eines \ddot{U} bergangs von K^c nach K . (Andernfalls würden sich die Wahrscheinlichkeiten verschieben.)

(b) Zerteilt ein Schnitt zwischen den Zuständen i und j den \ddot{U} -Graph in zwei (disjunkte) Teile, dann folgt aus (G) die (sogenannte) lokale Gleichgewichtsbedingung

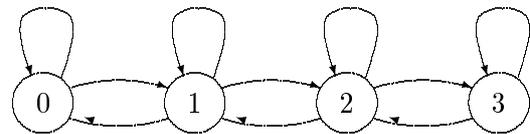
$$\pi_i p_{ij} = \pi_j p_{ji}. \quad (L_{ij})$$

(c) Ist der \ddot{U} -Graph einfach zusammenhängend, d.h. zerlegt ihn jeder Schnitt zwischen zwei verbundenen Zuständen in zwei Teile, dann ist (G) äquivalent zu

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{für alle } i, j \in I. \quad (L)$$



Der formale Beweis zu (a) ergibt sich, wenn man (G) nach Vertauschen der Indizes i und j über $i \in K$ summiert, die linke Seite zu $\sum_{i \in K} \pi_i = \sum_{i \in K} \sum_{j \in I} \pi_i p_{ij}$ erweitert und $\sum_{j \in I} \dots$ in $\sum_{j \in K} \dots + \sum_{j \in K^c} \dots$ aufteilt. Die Äquivalenz in (c) ergibt sich durch Summation von (L) über $i \in I$.



Bemerkung: Das Schnittprinzip ist besonders nützlich bei linear angeordneten \ddot{U} -Graphen, die im folgenden Kapitel über Bediensysteme häufig vorkommen.

Die Frage, ob eine HMK sich langfristig wie eine HMK im Gleichgewicht verhalten wird, werden wir nur unter gewissen zusätzlichen Voraussetzungen beantworten. Diese Voraussetzungen werden aber bei der Anwendung auf Bedienungsmodelle erfüllt sein.

16.15 Satz. (Grenzwertsatz für homogene Markov-Ketten) *Ist die HMK(X_n) mit \ddot{U} -Matrix $(p_{ij}, i, j \in I)$ irreduzibel und aperiodisch, dann konvergiert (für alle $i \in I$) $P(X_n = i)$ unabhängig von der Startverteilung gegen einen Wert π_i mit $0 \leq \pi_i \leq 1$. Dabei sind*

- entweder alle $\pi_i > 0$, und $(\pi_i, i \in I)$ ist die einzige GGK zu (p_{ij}) ,
- oder es sind alle $\pi_i = 0$, und es gibt keine GGK zu (p_{ij}) .

Der zweite Fall kommt nur bei unendlicher Zustandsmenge I vor. Ohne die Voraussetzung „aperiodisch“ konvergiert (i.a.) nur das langfristige Mittel. (Für einen Beweis dieses Satzes vgl. z.B. Kohlas, J. Stochastische Methoden des Operations Research, Satz 3.9)

Wir wissen nun, dass im irreduziblen und aperiodischen Fall langfristig ein Gleichgewicht erreicht wird, falls überhaupt eine GGK existiert (zumindest also bei endlicher Zustandsmenge I), und dass diese GGK dann eindeutig ist.

Dass wir auch unendliche Zustandsmengen I zulassen – und damit Modelle, in denen keine GGK existiert – liegt daran, dass das Gleichungssystem (G), (??) in Bedienungsmodellen ohne Warteraum-Beschränkung im allgemeinen leichter auszuwerten ist und dann Rückschlüsse auf endliche Modelle möglich sind.

So könnte etwa in einem unbeschränkten Bedienungssystem unter Überlast die Länge der Warteschlange gegen ∞ gehen. Dann existiert keine GGK. In der Realität treten aber durch Beschränkung des Warteraums Überlauf- oder Blockierungseffekte auf oder ein Teil der Kunden wird freiwillig das System verlassen.

Bemerkung: Das hier von der Betrachtung ausgeschlossene langfristige Verhalten von reduzierbaren oder periodischen HMK kann an den einfachen Modellen mit zwei Zuständen aus Definition ?? und Definition ?? andeutungsweise demonstriert werden.

1. Für eine reduzierbare HMK mit $(p_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ist jede der unendlich vielen möglichen Startverteilungen auch GGK (es ist stets $X_1 = X_2 = \dots$). Die langfristigen relativen Häufigkeiten sind aber stets $= 0$ oder $= 1$.

2. Eine HMK mit $(p_{ij}) = \begin{pmatrix} 0,5 & 0,5 \\ 0 & 1 \end{pmatrix}$ ist ebenfalls reduzierbar. Die einzige GGK ist $(\pi_1, \pi_2) = (0, 1)$, also nicht strikt positiv. Diese GGK wird sich bei jeder Startverteilung langfristig einstellen.

3. Für eine periodische HMK mit $(p_{ij}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ist nur $(\pi_1, \pi_2) = (1/2, 1/2)$ eine GGK. Für andere Startverteilungen ändert sich $P(X_n = i)$ mit der Periode 2. In diesem Fall konvergiert $P(X_n = i)$ nicht.

17 Bediensysteme

17.1 Vorbemerkungen

In einem Bediensystem, auch Bedienstation genannt, kommen Kunden oder Aufträge in einer zeitlichen Abfolge einzeln oder in Gruppen an, werden in einer festgelegten oder zufälligen Reihenfolge bedient bzw. bearbeitet und verlassen nach einer (i.a.) zufälligen Bedienzeit das System.

Werden mehrere Bedienstationen miteinander verbunden, und zwar dadurch, dass die abgehenden Kunden bzw. Aufträge eines Teils dieser Stationen direkt anderen Stationen als Ankünfte zugeleitet werden, dann spricht man von Bedien-Netzen oder Bedien-Netzwerken. Wir werden uns aber vorerst nur mit einzelnen Bediensystemen befassen.

Bediensysteme und Bediennetze waren bisher schon gelegentlich als Beispiele vorgekommen. Hier sollen sie nun systematisch behandelt werden.

In der Regel werden in der Literatur und in Lehrbüchern Bedien-Vorgänge als stochastische Prozesse mit stetigem Zeitparameter modelliert. Hierdurch ergeben sich (zumindest für die einfachen Probleme) zwar übersichtliche und leicht zu interpretierende Ergebnisse, aber zur Lösung dieser Probleme sind Hilfsmittel erforderlich, die im Rahmen dieser Vorlesung nicht bereitgestellt werden können.

Andererseits erhält man bei Modellierung mit diskretem Zeitparameter kompliziertere Ergebnisse, die außerdem nur schwer zu den oben erwähnten einfachen Ergebnissen in Beziehung gesetzt werden können.

Dieses Dilemma soll hier dadurch gelöst werden, dass wir zwar mit diskretem Zeitparameter arbeiten, aber die Länge der Zeittakte als hinreichend klein annehmen. Damit können wir an entsprechenden Stellen Modell-Vereinfachungen vornehmen, die numerisch vernachlässigbar sind, aber zu den einfachen Ergebnissen des stetigen Falles führen.

Dadurch ist gleichzeitig sichergestellt, dass die hier erarbeiteten Ergebnisse mit den in der Praxis angewandten „Formeln“ (die auf der stetigen Theorie beruhen) übereinstimmen. Dieser zunächst überraschende Effekt kommt daher, dass die ursprünglichen diskreten Modelle exakt in die stetigen Modelle übergehen, wenn man die Taktlänge gegen 0 gehen lässt.

Zum Abschluss dieser Vorbemerkungen soll auf die Klassifikation und Bezeichnungsweise für Bedienmodelle eingegangen werden.

Bediensysteme werden unter sehr unterschiedlichen Annahmen über den Ankunftsprozess und den Bedienvorgang modelliert. Um die wesentlichen Eigenschaften eines Bedienmodells auf einen Blick zu erkennen, wird allgemein eine einheitliche Kurzbezeichnung benutzt.

Klassifikation z.B. $M|M|1|\infty$

Die Kurzbezeichnung für Bedienmodelle setzt sich (meist) aus 5 Elementen zusammen in der

Form

$$A|B|s|c|R$$

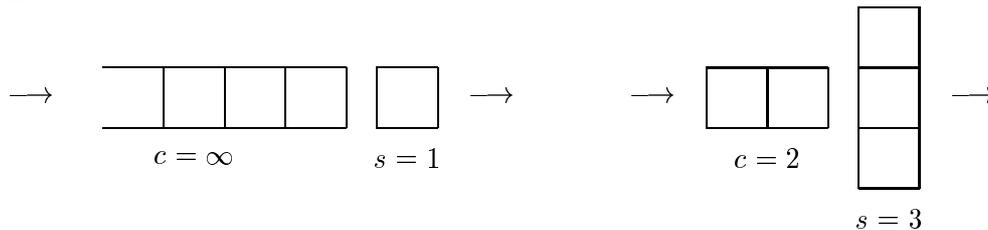
- Dabei steht
- A** für die Art des **A**nkunftsprozesses,
 - B** für die Art des **B**edienvorganges,
 - s** für die Anzahl der Bediener (**s**erver),
 - c** für die Größe des Warteraums (**c**apacity),
 - R** für die **R**eihenfolge der Bedienung.

Für **A** und **B** stehen z.B. die Buchstaben **M** (für **M**arkov-Eigenschaft), **D** (für **d**eterministisch) oder **G** (für **g**enerelle Annahmen).

Die Werte **s** und **c** sind ganze Zahlen ≥ 1 bzw. ≥ 0 oder ∞ , weil aus formalen Gründen auch unendlich viele Bediener und unendlich viele Warteplätze zugelassen werden.

Wenn die Bedienung in der Reihenfolge der Ankunft erfolgt, steht für die Bedienregel **FCFS** (**f**irst **c**ome **f**irst **s**erved), andere Regeln sind z.B. **LCFS** (**l**ast **c**ome **f**irst **s**erved, der letzte angekommene Kunde wird zuerst bedient) und **SIRO** (**s**ervice **i**n **r**andom **o**rd(er), Bedienung in zufälliger Reihenfolge). Der Standardfall, FCFS, wird meist nicht notiert, ebenso entfällt meist $c = \infty$. Bei $s = \infty$ entfällt die Angabe von c immer, weil kein Warteraum benötigt wird.

Gelegentlich werden Bediensysteme symbolisch dargestellt, etwa entsprechend den folgenden Beispielen:



Solche Symbole werden insbesondere bei der grafischen Darstellung eines Bediennetzes benutzt.

17.2 Das Bedienmodell $M|M|1|\infty$

Es mag überraschen, dass wir mit dem Bediensystem $M|M|1|\infty$ (kurz $M|M|1$) anfangen, einem System mit unendlichem Warteraum, wo doch in der Realität ein Warteraum stets endlich ist. Aber ein solches System lässt sich tatsächlich einfacher modellieren, weil die Berücksichtigung von Kunden, die ein volles System vorfinden, zusätzliche Überlegungen erfordert. Die endlichen Modelle lassen sich dann durch Modifikation des $M|M|1$ -Modells erzeugen.

Wir werden zuerst den Ankunftsprozess und den Bedienvorgang getrennt modellieren und dann daraus das eigentliche Bedienmodell zusammensetzen.

Um möglichst konkret zu sein, werden wir alle Überlegungen am Beispiel des Zugriffs auf einen Rechner veranschaulichen.

17.1 Beispiel. *Auf einen Zentralrechner werde von vielen Arbeitsplatz-Rechnern zugegriffen. Wir nehmen an, dass in jeder Minute durchschnittlich 10 Zugriffe erfolgen und dass die Bear-*

beitung der Aufträge im Mittel jeweils 5 Sekunden erfordert. Man möchte z.B. wissen, wie lange die Beantwortung der Aufträge im Mittel dauert (einschließlich Wartezeit).

Um ein Modell mit diskreter Zeit zu erhalten, betrachten wir einen getakteten Zeitablauf:

Zeittakt Taktlänge h

Die Zeit sei in „Takte“ der einheitlichen Länge h eingeteilt. Der Ankunfts- und Bedienungsprozess wird also zu den Zeitpunkten $0, h, 2h, 3h, \dots$ beobachtet. In unserem Beispiel wählen wir für die Taktlänge $h = 1/10$ Sekunde.

Wir modellieren nun den Ankunftsprozess. Zur Vereinfachung werden wir nur von ankommenden „Kunden“ sprechen, auch wenn je nach Anwendung „Aufträge“ oder „Datenpakete“ ankommen.

Ankunftsrate λ

Wir gehen davon aus, dass in gleichen Zeiträumen *im Mittel* gleich viele Kunden ankommen. Die mittlere Zahl von Ankünften je Zeiteinheit heißt dann die *Ankunftsrate*, die mit λ bezeichnet wird. In einem Zeitintervall der Länge l kommen also im Mittel $\lambda \cdot l$ Kunden an.

In unserem Beispiel ist demnach die Ankunftsrate $\lambda = 10$ pro Minute = $1/6$ pro Sekunde. In einem Takt ist dann die mittlere Zahl der Ankünfte $\lambda \cdot h = 1/60$.

Bernoulli-Ankunftsprozess

Für den Ankunftsprozess nehmen wir außerdem an, dass in jedem Takt höchstens *ein* Kunde ankommt. Für jeden einzelnen Takt liegt dann ein Bernoulli-Versuch vor, bei dem „1“ bzw. „Erfolg“ einer Ankunft entspricht. Diese Annahme ist immer dann realistisch, wenn die Takte kurz genug sind und die Kunden unabhängig voneinander ankommen, also keine Gruppenankünfte vorliegen.

Ist die Ankunfts-Wahrscheinlichkeit für diesen Bernoulli-Versuch gleich p_a (a für Ankunft), dann ist die erwartete (= mittlere) Zahl der Ankünfte in *einem* Takt ebenfalls p_a . Nach der obigen Definition der Ankunftsrate λ gilt also $p_a = \lambda h$, im Beispiel $p_a = 1/60$.

Es muss also die Taktlänge h so klein gewählt werden, dass zumindest $\lambda h \leq 1$ gilt. Wegen der Vernachlässigung von zwei und mehr Ankünften in einem Takt muss zudem „ λh klein gegen 1“ gelten, z.B. $\lambda h \leq 0,02$.

Wenn wir außerdem annehmen, dass eine Ankunft in jedem Takt die gleiche Wahrscheinlichkeit $p_a = \lambda h$ besitzt und von den Ankünften in allen anderen Takten unabhängig ist, dann können wir die Ankünfte in n Takten als *n -fachen Bernoulli(λh)-Versuch* modellieren. Für den gesamten Ankunftsprozess ergibt sich entsprechend eine unbeschränkte Folge von Bernoulli- (λh) -Versuchen, auch *Bernoulli(λh)-Prozess* genannt.

Eigenschaften des Ankunftsprozesses

17.2 Korollar. Für den durch einen Bernoulli(λh)-Prozess modellierten Ankunftsprozess gilt:

- (a) Die Ankunftsanzahlen in disjunkten Zeitintervallen sind stochastisch unabhängig.
- (b) Die Zahl der Ankünfte in n Takten ist Binomial($n, \lambda h$)-verteilt mit Erwartungswert $n\lambda h$.
- (c) Die Zahl der Takte bis zur nächsten Ankunft bzw. zwischen zwei Ankünften ist $\text{Geo}^+(\lambda h)$ -verteilt mit Erwartungswert $1/(\lambda h)$. Die reale mittlere Wartezeit auf eine Ankunft ist also $h \cdot 1/(\lambda h) = 1/\lambda$ (in Beispiel 17.1 also 60 Takte = 6 Sekunden). Entsprechend ergibt sich die (reale) Streuung als $h \cdot \sqrt{1 - \lambda h}/(\lambda h) = \sqrt{1 - \lambda h}/\lambda \approx 1/\lambda$ (im Beispiel also 5,95 Sekunden).

Die Modellierung des Bedienprozesses beginnen wir ebenfalls mit einer „Rate“, der Bedienrate.

Bedienrate μ

Wir setzen analog zum Ankunftsprozess auch hier voraus, dass die mittlere Anzahl μ der Bedienungen je Zeiteinheit konstant ist (solange immer Kunden da sind, die bedient werden). Die Größe μ heißt entsprechend *Bedienrate*.

17.3 Bemerkung. Ist statt der Bedienrate μ die mittlere Bedienzeit, z.B. b , gegeben (wie im Beispiel), dann ist die Bedienrate $\mu = 1/b$. So ergibt sich im Beispiel aus der mittleren Bedienzeit $b = 5$ Sekunden (= 1/12 Minute) die Bedienrate $\mu = 12$ pro Minute. Auf einen Takt (= $\frac{1}{600}$ Minute) bezogen sind dies im Mittel $\mu h = \frac{12}{600} = \frac{1}{50}$ Bedienungen. Dieser Zusammenhang ist für eine feste Bedienzeit b sofort klar. Dass dies auch für stochastische Bedienzeiten gilt, wird in der Erneuerungstheorie gezeigt (für unabhängige und identisch verteilte Bedienzeiten).

Bernoulli-Bedien-Prozess

Den Bedienprozess werden wir ebenso wie den Ankunftsprozess mit Hilfe eines Bernoulli-Prozesses modellieren. Es liege also ein Bernoulli(p_b)-Prozess vor (b für Bedienung), der vom Bernoulli-Ankunftsprozess stochastisch unabhängig sei. Nun interpretieren wir jeden „Erfolg“ dieses Prozesses als das Ende einer Bedienung, sofern ein Kunde da ist. Damit erhält man die Bedienzeiten als Wartezeiten bis zum nächsten Bedienende.

Den Parameter p_b erhält man wie beim Ankunftsprozess daraus, dass bei einem Bernoulli(p)-Versuch die Erfolgswahrscheinlichkeit p und die erwartete Anzahl der Erfolge übereinstimmen. Also ist hier $p_b = \mu h$, im Beispiel = $\frac{1}{50}$ (s.o).

Auch hier muss wie bei $p_a = \lambda h$ darauf geachtet werden, dass „ μh klein gegen 1“ gilt, z.B. $\mu h \leq 0,02$.

Wie in Folgerung 17.2(c) ergibt sich daraus:

17.4 Korollar. Die Bedienzeiten B_k sind – als Wartezeiten in einem Bernoulli(μh)-Prozess – nach Takten gezählt $\text{Geo}^+(\mu h)$ -verteilt und besitzen den Erwartungswert

$$EB_k = \frac{1}{\mu h} \text{ (in Takten) bzw. } h \cdot \frac{1}{\mu h} = \frac{1}{\mu} \text{ (in Zeiteinheiten).}$$

Die Streuung ist entsprechend

$$\text{Str}B_k = \frac{\sqrt{1 - \mu h}}{\mu h} \text{ (in Takt)} \text{ bzw. } \frac{\sqrt{1 - \mu h}}{\mu} \approx \frac{1}{\mu} \text{ (in Zeiteinheiten)}.$$

Auch die Restbedienzeit von einem beliebigen Zeitpunkt an hat dieselbe $\text{Geo}^+(\mu h)$ -Verteilung wie die Gesamtbedienzeit. Dies folgt direkt aus der Konstruktion über den Bernoulli-Prozess und hängt mit der Gedächtnislosigkeit der geometrischen Verteilung zusammen.

17.5 Bemerkung. Die geometrisch verteilten Bedienzeiten, die sich aus dem Bernoulli-Prozess ergeben, kommen in der Praxis selten vor. Deshalb ist der Bernoulli-Ansatz hier weniger realistisch als bei der Modellierung des Ankunftsprozesses. Aber es gibt doch gute Gründe für diesen Ansatz:

(1) Man erhält ein einfaches Gesamtmodell mit Markov-Eigenschaft, das sich gut auswerten lässt.

(2) Man kann diesen Ansatz später so modifizieren, dass jede andere Bedienzeitverteilung approximiert wird.

Aus der Überlagerung des Bernoulli(λh)-Ankunftsprozesses und des Bernoulli(μh)-Bedienprozesses erhält man nun das Gesamtmodell, wenn man als Zustand des Systems die daraus resultierende Zahl der Kunden im System angibt, und zwar zu den Zeitpunkten $0, h, 2h, \dots$. Diese Zeitpunkte werden mit $n = 0, 1, 2, \dots$ nummeriert.

Dabei ändert sich die Kundenzahl in einem Takt

- um $+1$, wenn ein Kunde ankommt, kein Kunde geht,
- um -1 , wenn kein Kunde ankommt, ein Kunde geht,
- nicht, wenn kein Kunde ankommt, kein Kunde geht,

oder wenn ein Kunde ankommt, ein Kunde geht.

Für den hierdurch definierten Prozess gilt die Markov-Eigenschaft, weil der Zustand nach einem Takt nur vom Zustand vor diesem Takt und den beiden Bernoulli-Versuchen während des Taktes abhängt, wobei diese Bernoulli-Versuche von allen vorausgehenden (und gegenseitig) stochastisch unabhängig sind.

Modell 1

17.6 Korollar. Wir erhalten also als Modell 1 eine homogene Markov-Kette X_0, X_1, X_2, \dots mit dem Zustandsraum $I = \{0, 1, 2, \dots\}$, einer beliebigen Startverteilung $(p_0(i), i \in I)$ und der folgenden Übergangsmatrix $(p_{ij}, i, j \in I)$ (zur Abkürzung steht $1 - \dots$ für $(1 - \lambda h)(1 - \mu h) + \lambda h \mu h = 1 - \lambda h - \mu h + 2\lambda h \mu h$):

$$(p_{ij}) = \begin{pmatrix} 1 - \lambda h & \lambda h & 0 & 0 & \dots \\ (1 - \lambda h)\mu h & 1 - \dots & \lambda h(1 - \mu h) & 0 & \dots \\ 0 & (1 - \lambda h)\mu h & 1 - \dots & \lambda h(1 - \mu h) & \dots \\ 0 & 0 & (1 - \lambda h)\mu h & 1 - \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (17.1)$$

Als Zahlenwerte erhalten wir in unserem Beispiel:

$$(p_{ij}) = \begin{pmatrix} 1 - \frac{1}{60} & \frac{1}{60} & 0 & \dots \\ \frac{1}{50} - \frac{1}{3000} & 1 - \frac{1}{60} - \frac{1}{50} + \frac{2}{3000} & \frac{1}{60} - \frac{1}{3000} & \ddots \\ \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (17.2)$$

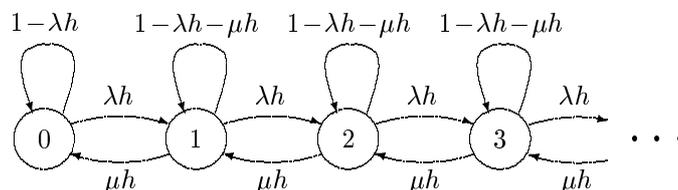
Weil die Übergangswahrscheinlichkeiten in den Formeln und den Zahlenwerten relativ unübersichtlich sind, werden wir im folgenden die Terme $\lambda h \mu h$ (im Beispiel = $\frac{1}{3000}$) vernachlässigen. Dies ist deshalb sinnvoll, weil wir ohnehin die Taktlänge h als so kurz vorausgesetzt hatten, dass „ λh und μh klein gegen 1“ gilt (z.B. $\leq 0,02$). (Dadurch ist zugleich sichergestellt, dass die sich ergebenden Diagonalelemente $1 - \lambda h - \mu h$ nicht-negativ sind.) Damit erhält man eine vereinfachte homogene Markov-Kette, die wir *Modell 2* nennen.

Modell 2

17.7 Korollar. Modell 2 unterscheidet sich von Modell 1 nur durch die vereinfachte Übergangsmatrix (p_{ij}) :

$$(p_{ij}) = \begin{pmatrix} 1 - \lambda h & \lambda h & 0 & 0 & \dots \\ \mu h & 1 - \lambda h - \mu h & \lambda h & 0 & \dots \\ 0 & \mu h & 1 - \lambda h - \mu h & \lambda h & \dots \\ 0 & 0 & \mu h & 1 - \lambda h - \mu h & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (17.3)$$

Der zugehörige Übergangs-Graph hat folgende Form (vgl. Def. 16.4):



Damit ist die Modellierung des $M|M|1|\infty$ -Bediensystems abgeschlossen. Im nächsten Abschnitt werden wir das Verhalten von Modell 2 im Gleichgewicht untersuchen.

17.8 Bemerkung. Wenn zu der homogenen Markov-Kette von Modell 2 der Ankunfts- und der Bedienprozess rekonstruiert wird unter der Annahme, dass nicht gleichzeitig eine Ankunft und ein Abgang vorkommt, dann erhält man wie bei dem zuvor modellierten System einen Bernoulli(λh)-Ankunftsprozess und einen Bernoulli(μh)-Bedienprozess, wie man direkt aus den Übergangswahrscheinlichkeiten sieht. Aber die beiden Bernoulli-Prozesse sind nicht mehr stochastisch unabhängig.

Behält man dagegen die (sonst in der Bedienungstheorie übliche) Unabhängigkeit von Ankunfts- und Bedienprozess bei, dann muss die Ankunfts- und Bedienrate bei den Zuständen $i \geq 1$

geändert werden (näherungsweise λh in $\lambda h(1+\mu h)$ und μh in $\mu h(1+\lambda h)$), also zustandsabhängig sein. Beide Probleme „verschwinden“ aber, wenn die Taktlänge h gegen 0 geht.

Die Approximation bei Modell 2 wird exakt, wenn man statt der konstanten Taktlängen h zufällige (st.u.) $\text{Exp}(1/h)$ -verteilte Taktlängen benutzt. Aus den geometrischen Verteilungen in Folg. 17.2 und 17.4 werden dann $\text{Exp}(\lambda)$ - bzw. $\text{Exp}(\mu)$ -Verteilungen. Dieser Ansatz soll aber hier nicht weiter ausgeführt werden.

17.3 Das $M|M|1$ -Bediensystem im Gleichgewicht

Es ist sinnvoll, das langfristige Verhalten von Markov-Ketten – und damit auch von Bediensystemen – anhand der Gleichgewichtsverteilungen zu untersuchen. Deshalb werden wir jetzt die Gleichgewichtsverteilung $(\pi_i, i \in I)$ (falls sie existiert) für das in Abschnitt 17.2 entwickelte Modell 2 des $M|M|1$ -Bediensystems bestimmen. Dazu müssen die Gleichgewichtsbedingungen

$$(G) \quad \pi_j = \sum_{i \in I} \pi_i p_{ij}, \quad j \in I \quad \text{und} \quad (N) \quad \pi_j \geq 0, \quad j \in I, \quad \sum_j \pi_j = 1$$

gelöst werden.

Durch die spezielle Struktur des \ddot{U} -Graphs von Modell 2 (vgl. Folgerung 17.7) vereinfacht sich aber nach dem Schnittprinzip die Bedingung (G) zur äquivalenten Bedingung (L) $\pi_i p_{ij} = \pi_j p_{ji}$ ($i, j \in I$), hier also

$$\pi_j \lambda h = \pi_{j+1} \mu h, \quad j = 0, 1, 2, \dots, \quad (17.4)$$

wobei der Faktor h entfallen kann. Daraus folgt durch Induktion die äquivalente Darstellung

$$\pi_{j+1} = \frac{\lambda}{\mu} \pi_j = \left(\frac{\lambda}{\mu}\right)^{j+1} \pi_0, \quad j = 0, 1, 2, \dots \quad (17.5)$$

Die Bedingung (N) ergibt $\pi_0 \geq 0$ und

$$1 = \pi_0 + \pi_1 + \pi_2 + \dots = \pi_0 \left(1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \dots\right). \quad (17.6)$$

Für $\lambda \geq \mu$ geht die Summe gegen ∞ . Es gibt also keine Lösung π_0 von (17.6) und damit keine Lösung für das Gleichungssystem (G), (N).

Für $\lambda < \mu$ ist die Summe in (17.6) konvergent gegen $(1 - \lambda/\mu)^{-1}$. Damit existiert genau eine Lösung von 17.6, nämlich $\pi_0 = 1 - \lambda/\mu$, also hat das System (G), (N) die eindeutige Lösung

$$\pi_j = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^j, \quad j = 0, 1, 2, \dots \quad (17.7)$$

Dies ist die Z-Dichte der $\text{Geo}^0(1 - \lambda/\mu)$ -Verteilung.

In Beispiel 17.1 ist $\lambda/\mu = 5/6$, also gilt dort $\pi_j = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^j, j = 0, 1, 2, \dots$

Man sieht, dass die für $\lambda < \mu$ eindeutig existierende Gleichgewichtsverteilung von Modell 2 unabhängig von der Taktlänge h ist. Da dieses Modell für kleine Werte von h konzipiert war, ist es nicht überraschend, dass sich dieselbe Verteilung auch für stetige Modelle ergibt. Wir formulieren deshalb den folgenden Satz für beide Fälle gemeinsam. Für $h \rightarrow 0$ geht diese Verteilung in (17.7) über.)

$M|M|1$ im Gleichgewicht

17.9 Satz. In einem $M|M|1$ -Bediensystem, modelliert durch Modell 2 oder ein Modell mit stetiger Zeit, gilt:

- (a) Für $\lambda < \mu$ besitzen die Gleichgewichtsbedingungen (G), (N) genau eine Lösung, und im Gleichgewicht ist die Zahl X_n der Kunden im System $\text{Geo}^0(1-\lambda/\mu)$ -verteilt (mit der Z -Dichte (17.7)).
- (b) Für $\lambda \geq \mu$ gibt es keine Gleichgewichtsverteilung.

17.10 Bemerkung. Die Bedingung „ $\lambda < \mu$ “ bedeutet, dass im Mittel weniger Kunden ankommen als bedient werden können. Also „stabilisiert sich“ das System. Umgekehrt bedeutet „ $\lambda > \mu$ “, dass die ankommenden Kunden vom Bediener langfristig nicht bewältigt werden können. Die Kundenzahl geht im Laufe der Zeit gegen ∞ . Dass auch bei „ $\lambda = \mu$ “ kein Gleichgewicht existiert, liegt daran, dass dann keine freie Bedienungskapazitäten vorhanden sind, um eine entstandene Warteschlange abzubauen.

Die Eindeutigkeit der Gleichgewichtsverteilung folgt auch aus Satz 16.15, weil die Markov-Kette aus Modell 2 irreduzibel ist (jeder Zustand ist von jedem anderen erreichbar). Außerdem folgt aus $1 - \lambda h > 1 - \lambda h - \mu h \geq 0$ auch die Eigenschaft „aperiodisch“ und damit die Annäherung der Markov-Kette bei beliebiger Startverteilung an das Gleichgewicht. (Wenn keine Übergänge $i \rightarrow i$ möglich wären, hätte man Periode 2.)

17.11 Bemerkung. Bei einem Bediensystem im Gleichgewicht bedeutet $\pi_i = P(X_n = i)$, dass man bei einer Inspektion zu einem (beliebigen) Zeitpunkt n das System mit Wahrscheinlichkeit π_i im Zustand $X_n = i$ vorfindet, sofern man keine Information über den Prozess besitzt. Es ist klar, dass bei Information über eine Ankunft im Takt n $P(X_n = 0 | X_n = X_{n-1} + 1) = 0 \neq \pi_0$ ist.

Es gibt aber Beispiele von Informationen, die das Gleichgewicht nicht verfälschen:

17.12 Satz. Für ein $M|M|1$ -System im Gleichgewicht gilt:

(a) Ein ankommender Kunde trifft das System (vor seiner Ankunft) im Gleichgewicht an, d.h. $P(X_n = i | X_{n+1} = X_n + 1) = \pi_i$ ($i \in I$).

(b) Ein abgehender Kunde hinterlässt das System im Gleichgewicht, d.h.

$$P(X_n = i | X_n = X_{n-1} - 1) = \pi_i \quad (i \in I).$$

Die zweite Aussage ist besonders überraschend, da man intuitiv annehmen würde, dass nach einem Abgang weniger Kunden „als normal“ im System sind.

Zum Beweis muss man die bedingten Wahrscheinlichkeiten nachrechnen. Für Teil (a) gilt $P(1 \text{ Ankunft in Takt } n) = \lambda h$.

Daraus folgt $P(X_{n+1} = X_n + 1) = \lambda h$. Daraus folgt

$$P(X_n = i | X_{n+1} = X_n + 1) = P(X_n = i, X_{n+1} = i + 1) / P(1 \text{ Ankunft}) = \pi_i \cdot \lambda h / \lambda h$$

$= \pi_i$ ($i \in I$). Den Beweis von Teil (b) führen wir hier nicht durch, s. Hübner.

17.4 Leistungsmaße im $M|M|1$ -Bediensystem

Im letzten Abschnitt hatten wir untersucht, unter welchen Voraussetzungen das $M|M|1$ -Bediensystem langfristig im Gleichgewicht sein wird und welche Verteilung dann die Zahl der anwesenden Kunden annimmt. Ausgehend davon sollen nun auch andere Eigenschaften des Systems im Gleichgewicht untersucht werden.

Eigenschaften eines (Bedien-)Systems, meist im Gleichgewicht, die durch wenige reellwertige Größen ausgedrückt werden können, werden **Leistungsmaße** des Systems genannt. Dazu gehören z.B. die mittlere Zeit, die ein Kunde im System verbringt, oder der durchschnittliche Zeiteanteil, in dem der Bediener untätig ist, weil keine Kunden da sind.

Wir werden diese Größen hier zwar für allgemeine Bediensysteme einführen, aber zunächst nur für das $M|M|1$ -System diskutieren und auswerten. Im nächsten Abschnitt werden wir dann die Betrachtungen auch auf andere Bediensysteme ausdehnen.

Wir beginnen mit einer Leistungsgröße, die u.a. aussagt, ob es überhaupt ein Gleichgewicht gibt.

17.13 Definition. Auslastung *Die Auslastung ρ wird allgemein definiert als*

$$\rho := \frac{\text{(ankommende) Verkehrsdichte}}{\text{maximale Bedienkapazität}}.$$

Dabei ist die Verkehrsdichte die je Zeiteinheit ankommende Anforderung von (erwarteter) Bedienzeit, die maximale Bedienkapazität ist die Anzahl der zur Verfügung stehenden Bediener. Im $M|M|1$ -Fall ergibt dies

$$\rho = \frac{\lambda \cdot 1/\mu}{1} = \frac{\lambda}{\mu} = \frac{\text{Ankunftsrate}}{\text{Bedienrate}}.$$

Im allgemeinen Fall ist diese Darstellung nicht brauchbar, wenn es Kunden mit unterschiedlichen Bedienzeitverteilungen gibt.

Man sieht leicht ein, dass auch für beliebige Bediensysteme nur für $\rho < 1$ ein Gleichgewicht existieren kann (wenn kein „Überlauf“ möglich ist), weil für $\rho \geq 1$ die Bedienkapazität nicht zur Bewältigung des Ankunftsstroms ausreicht.

17.14 Bemerkung. *Die Auslastung ρ gibt im Gleichgewichtsfall gleichzeitig den mittleren Anteil der beschäftigten Bediener an. Bei einem Bediener ist dies auch die Wahrscheinlichkeit, dass der Bediener zu einem festen Zeitpunkt beschäftigt ist, also gilt $\rho = 1 - P(\text{System leer}) = 1 - \pi_0$, nicht nur für $M|M|1$ -Systeme.*

Viele der im folgenden genannten Leistungsmaße hängen nur über ρ von λ und μ ab. Deshalb werden die entsprechenden Formeln auch mit Hilfe von ρ formuliert. Dies erleichtert oft die Berechnung, zumal ρ eine dimensionslose Größe ist. Insbesondere ist für ein $M|M|1$ -System die $Geo^0(1 - \lambda/\mu)$ -Gleichgewichtsverteilung nur von ρ abhängig.

Es liegt nahe, den Erwartungswert und die Streuung der Gleichgewichtsverteilung als Leistungsmaß heranzuziehen.

Für ein Bediensystem im Gleichgewicht ergibt sich der Erwartungswert EX_n der Zahl der Kunden im System, kurz die *mittlere Kundenzahl*, ebenso wie die *Streuung der Kundenzahl*, $\text{Str}X_n$, direkt aus der Gleichgewichtsverteilung. (Im Gleichgewicht hängt die Verteilung von X_n und damit EX_n und $\text{Str}X_n$ nicht von n ab.)

Für das $M|M|1$ -System gilt also wegen der $\text{Geo}^0(1 - \lambda/\mu)$ -Verteilung von X_n

$$EX_n = \frac{\lambda/\mu}{1 - \lambda/\mu} = \frac{\rho}{1 - \rho}, \quad \text{Str}X_n = \frac{\sqrt{\lambda/\mu}}{1 - \lambda/\mu} = \frac{\sqrt{\rho}}{1 - \rho}. \quad (17.8)$$

In Beispiel 17.1 mit $\rho = \lambda/\mu = 5/6$ gilt also

$$EX_n = \frac{5/6}{1 - 5/6} = 5, \quad \text{Str}X_n = \frac{\sqrt{5/6}}{1 - 5/6} = \sqrt{30} \approx 5,5.$$

Der Erwartungswert und die Streuung reichen meist für die Beurteilung der in Anspruch genommenen Warteplätze nicht aus. Deshalb spielen die Quantile (vgl. Abschnitt ??) bei dieser Frage eine wichtige Rolle.

Das α -Quantil u_α der Gleichgewichtsverteilung gibt eine Schranke für die Kundenzahl an, die zu einem festen Zeitpunkt mit Wahrscheinlichkeit α nicht überschritten wird. Für $\alpha = 0,9$ z.B. bedeutet dies gleichzeitig, dass langfristig in 90% der Zeitpunkte die Kundenzahl X_n nicht größer als das 90%-Quantil $u_{0,9}$ (oder $u_{90\%}$) sein wird.

In der folgenden auszugsweisen Tabelle sind die Z-Dichte und die Verteilungsfunktion der Gleichgewichtsverteilung $\text{Geo}^0(1 - \rho)$ mit $\rho = 5/6$ angegeben. Daraus kann man insbesondere die Quantile $u_{90\%} = 12$ und $u_{99\%} = 25$ ablesen.

k	0	1	2	...	11	12	...	24	25
$P(X_n = k)$	0,167	0,139	0,116	...	0,022	0,019	...	0,0021	0,0017
$P(X_n \leq k)$	0,167	0,306	0,421	...	0,888	0,907	...	0,9895	0,9913

Bei 12 bzw. 25 Plätzen im System wird man also in 90% bzw. in 99% der Zeit keinen Überlauf bzw. keine Blockierung erwarten.

An die Bestimmung der Leistungsmaße EX_n , $\text{Str}X_n$ und $u_{90\%}/u_{99\%}$ schließt sich die Frage an, wie stark diese Werte auf Änderungen von ρ reagieren.

Wenn sich die Auslastung ρ dem Wert 1 nähert, steigt im Gleichgewicht die Zahl der Kunden im System stark an. Wenn in unserem Beispiel die Ankunftsrate λ von 10 auf 11 (und damit ρ von $5/6$ auf $11/12$) steigt, also um 10%, dann steigen die Werte EX_n von 5 auf 11, $\text{Str}X_n$ von 5,5 auf 11,5, $u_{90\%}$ von 12 auf 26 und $u_{99\%}$ von 25 auf 52, also jeweils auf mehr als das Doppelte. Die **mittlere Bedienzeit** EW_b gehört auch zu den Leistungsmaßen, obwohl sie meist schon von Anfang an vorgegeben ist oder direkt aus der Bedienrate folgt (vgl. Abschnitt 17.2).

Im $M|M|1$ -Modell ist die Bedienzeit W_b (vgl. Folgerung 17.4) in Takten $\text{Geo}^+(\mu h)$ -verteilt mit Erwartungswert

$$EW_b = 1/\mu h \quad (\text{Takte}) \quad \text{bzw.} \quad EW_b = 1/\mu \quad (\text{Zeiteinheiten}).$$

Entsprechend gilt für die *Streuung der Bedienzeit*

$$\text{Str}W_b = \frac{\sqrt{1 - \mu h}}{\mu h} \text{ (in Takten) bzw. } = \frac{\sqrt{1 - \mu h}}{\mu} \approx \frac{1}{\mu} \text{ (in Zeiteinheiten).}$$

Im Beispiel war $b = EW_b = 5$ Sekunden vorgegeben. Mit $\mu h = 1/50$ folgt für die Streuung $\text{Str}W_b = \sqrt{0,98}/\mu = 0,99 b \approx b = 5$ Sekunden.

17.15 Definition. (Mittlere Verweilzeit EW)

Die Verweilzeit, auch „Antwortzeit“ oder auch „Durchlaufzeit“, ist die (zufällige) Zeit, die ein Kunde nach seiner Ankunft in der Warteschlange und in Bedienung verbringt. Weil wir das System im Gleichgewicht untersuchen, ist die Verteilung der Verweilzeit für jeden Kunden gleich (bei gleichartigen Kunden). Man betrachtet deshalb die Verweilzeit W eines beliebig herausgegriffenen Kunden.

Als Leistungsmaß wird die *mittlere Verweilzeit* EW angegeben, evtl. auch die Streuung $\text{Str}W$. Die Verteilung der Verweilzeit ist zwar kein Leistungsmaß im strengen Sinn, aber z.B. die daraus abgeleiteten Quantile. Allerdings ist die Bestimmung der Verweilzeitverteilung für viele Bediensysteme ziemlich schwierig.

17.16 Korollar. (Verteilung der Verweilzeit) *In einem $M|M|1$ -Bediensystem im Gleichgewicht gilt:*

- (a) Die Verweilzeit W der Kunden (in Takten gerechnet) ist $\text{Geo}^+((\mu - \lambda)h)$ -verteilt.
 (b) Die mittlere Verweilzeit und die Streuung sind

$$EW = \frac{1}{(\mu - \lambda)h} \text{ (Takte) bzw. } \frac{1}{\mu - \lambda} \text{ (Zeiteinheiten),}$$

$$\text{Str}W = \frac{\sqrt{1 - (\mu - \lambda)h}}{(\mu - \lambda)h} \text{ (Takte) bzw. } \frac{\sqrt{1 - (\mu - \lambda)h}}{\mu - \lambda} \approx \frac{1}{\mu - \lambda} \text{ (Zeiteinheiten).}$$

Im Beispiel 17.1 ist $\mu = 12$ und $\lambda = 10$, also $\mu - \lambda = 2$ (je Minute). Damit erhält man $EW = 1/2$ Minute und $\text{Str}W \approx 1/2$ Minute.

Für das 90%- bzw. 99%-Quantil der Verweilzeit erhält man aus $P(W > kh) = (1 - (\mu - \lambda)h)^k \leq 1 - \alpha$ und $1 - (\mu - \lambda)h = 299/300$ die Werte $u_{90\%} = 1,15$ Min. und $u_{99\%} = 2,3$ Min. Diese Werte sind im Vergleich zur mittleren Bedienzeit von 5 Sekunden bemerkenswert groß.

Der Weg zur Bestimmung der mittleren Verweilzeit EW war schon hier sehr mühsam. Für andere Bediensysteme ist eine entsprechende Herleitung äußerst schwierig oder unmöglich. Es gibt aber trotzdem eine einfache und ziemlich allgemein gültige Formel zur Berechnung von EW aus der mittleren Kundenzahl EX_n .

17.17 Satz. (Formel von Little) *Sofern in einem beliebigen Bediensystem die langfristigen Durchschnitte konvergieren, gilt für die mittlere Verweildauer EW , die mittlere Kundenzahl EX_n und die mittlere Ankunftsrate $\bar{\lambda}$*

$$EW = EX_n / \bar{\lambda}.$$

Denn für die (zufällige) Zahl Y_t der Ankünfte bis t und für die gesamte bis t im System verbrachte Kundenzeit $G_t = \sum_{n=0}^{t-1} X_n$ gilt, dass die mittlere Zeit je Kunde, $W_t = G_t/Y_t$, gegen EW , die mittlere Kundenzahl G_t/t gegen EX_n und die mittlere Zahl der Ankünfte Y_t/t gegen $\bar{\lambda}$ konvergiert. Daraus folgt $EW = EX_n/\bar{\lambda}$.

Die Formel von Little kann man sich für ein $M|M|1$ -Bediensystem im Gleichgewicht mit $FCFS$ -Bedienregel leichter veranschaulichen, wenn man bedenkt, dass ein abgehender Kunde im Mittel EX_n Kunden zurücklässt (vgl. Satz 17.12(b)). Dies sind aber genau die Kunden, die während seiner Verweilzeit W mit der Rate λ angekommen sind (wegen $FCFS$). Also gilt (im Mittel) $EX_n = EW \cdot \lambda$.

18 Symbole, Abkürzungen, Verteilungen und Literaturliste

18.1 Symbole

\underline{n}	$\{1, 2, \dots, n\}$
\mathbb{N}_0	$\{0, 1, 2, \dots\}$
\mathbb{N}	$\{1, 2, 3, \dots\}$
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$
\mathbb{B}	Borel σ -Algebra
\log	natürlicher Logarithmus
\xrightarrow{P}	Konvergenz nach Wahrscheinlichkeit
$P - \lim$	Konvergenz nach Wahrscheinlichkeit
$B(p)$	Bernoulliverteilung
$B(n, p)$	Binomialverteilung
χ_n^2	χ^2 -Verteilung mit n Freiheitsgraden
$\mathcal{E}(\alpha)$	Exponentialverteilung
$\mathcal{H}(N, K, n)$	Hypergeometrische Verteilung
$\mathcal{N}(0, 1)$	Standard-Normalverteilung
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung
Φ	Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung
$\mathcal{P}(\lambda)$	Poisson-Verteilung
$\mathcal{R}(a, b)$	Rechteck-Verteilung
t_n	Student'sche t -Verteilung

18.2 Abkürzungen

<i>f.s.</i>	fast sicher
i.i.d.	independent and identically distributed
n.W.	nach Wahrscheinlichkeit
R-Dichte	Riemann-Dichte
st.u.	stochastisch unabhängig
ÜW	Übergangswahrscheinlichkeit
VF	Verteilungsfunktion
ZV	Zufallsvariable

18.3 Verteilungen und ihre Dichten

Bernoulliverteilung	$b(k; p) = p^k(1-p)^{1-k}, \quad k = 0, 1$
Binomialverteilung	$b(k; n, p) = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, \dots, n$
Cauchy(α, β)-verteilung	$f(x) = \frac{1}{\pi\beta} \frac{1}{1 + \left(\frac{x-\alpha}{\beta}\right)^2}, \quad x \in \mathbb{R}$
χ_n^2 -Verteilung	$h_n(x) = \frac{2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2})} 1_{(0, \infty)}(x), \quad x \in \mathbb{R}$
Exponentialverteilung	$f(x) = 1_{(0, \infty)}(x) \alpha e^{-\alpha x}, \quad \alpha \in (0, \infty), x \in \mathbb{R}$
Gamma(α, β)-verteilung	$f_{\alpha, \beta}(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} 1_{(0, \infty)}(x), \quad x \in \mathbb{R}$
Gammafunktion	$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \quad \alpha > 0$
Hypergeometrische Verteilung	$h(k; N, K, n) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad k = 0, \dots, n, n < N, K < N,$ $\max(0, n - N + K) \leq k \leq \min(K, n)$
Standard-Normalverteilung	$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}$
Normalverteilung	$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \quad \mu \in \mathbb{R}, \sigma^2 > 0$
Poisson-Verteilung	$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}, \lambda \in (0, \infty)$
Rechteck-Verteilung	$f(x) = 1_{[a, b]}(x) \frac{1}{b-a}, \quad a < b, a, b \in \mathbb{R}$
Student'sche t -Verteilung	$f_n(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}$
Weibull(α, β)-Verteilung	$f(x) = 1_{(0, \infty)}(x) \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} \quad \alpha, \beta \in (0, \infty)$

18.4 Literaturliste

- H.Bauer Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie
de Gruyter, Berlin 1978, 3. Auflage
- H.Bauer Maß- und Integrationstheorie
de Gruyter, Berlin 1990
- H.Bauer Wahrscheinlichkeitstheorie
de Gruyter, Berlin 1991, 4.Auflage
- K.Behnen, G.Neuhaus Grundkurs Stochastik
PD-Verlag, Heidenau, 4.Auflage
- P.J.Bickel, K.A.Doksum Mathematical Statistics
Holden-Day, San Francisco, Düsseldorf 1977
- Y.S. Chow, H.Teicher Probability Theory
Springer, Heidelberg 1978
- H.Dinges, H.Rost Prinzipien der Stochastik
Teubner Studienbücher 1982
- P.Gänssler, W.Stute Wahrscheinlichkeitstheorie
Springer, Heidelberg 1977
- K.Hinderer Grundbegriffe der Wahrscheinlichkeitstheorie
Springer, Heidelberg 1972
- G. Hübner Stochastik
Vieweg, Braunschweig, 4. Auflage 2003
- M.Loeve Probability Theory I
Springer, Heidelberg 1977
- J.Pfanzagl Elementare Wahrscheinlichkeitsrechnung
de Gruyter 1988
- N.Schmitz Vorlesungen in Wahrscheinlichkeitstheorie
Teubner Studienbücher 1995

Index

- diskretes W-Maß, 2
- endliches Zufallsexperiment, 3

- Algebra, 35
- Alternative, 33
- Ankunftsrate, 64
- Anwendungspostulat, 53
- Ausfallrate, 40
- Auslastung, 70

- Bayes-Formel, 13
- Bedienrate, 65
- Bernoulli Prozess, 64
- Bernoulli(p)-Verteilung, 6
- Bernoullisches schwaches Gesetz der großen Zahlen, 8
- Binomial(n, p)-Verteilung, 6
- Bolthausen, E., 24
- Borel- σ -Algebra, 36

- Chapman-Kolmogorov, 57
- CLT, 28

- Dichte
 - Riemann, 38
- Dirac-Maß, 27
- diskrete Zufallsvariable, 43
- Diskreter W-Raum, 1
- diskretes Wahrscheinlichkeitsmaß, 2

- E-Wert, 17
- Einpunktverteilung, 27
- empirische Verteilungsfunktion, 25
- endlich-additiv, 2
- Ereignis, 1
- Erwartungstreue, 23
- Erwartungswert, 17, 42
- Erzeuger, 36
- Experiment
 - Poisson(λ), 7
- Exponential(α)-Verteilung, 39

- Faltungsformel
 - für R-Dichten, 46
- fast sicher, 2
- fast unmöglich, 2
- Fehlerwahrscheinlichkeit, 33
- Formel von Little, 72

- Gauss'sche Glockenkurve, 28,40
- $Geo^0(p)$ -Verteilung, 7
- Gesetz der großen Zahlen
 - schwaches, 21
- Gesetz der großen Zahlen
 - schwaches, 21
- Gesetze der großen Zahlen
 - schwaches, 8
- Gleichgewicht, 58
- Gleichgewichtsbedingung
 - lokale, 60
- Gleichgewichtsbedingungen, 59
- Gleichverteilung, 3
- Glockenkurve, 40
- Grenzwertsatz
 - Zentraler , 32
- Grundraum, 1
- Gütefunktion, 33

- Homogene Markov-Kette, 55
- Hypergeometrische(N, K, n)-Verteilung, 7
- Hypothese, 33

- i.i.d., 21
- Indikatorfunktion, 2
- induziertes Maß, 43
- integrabel, 17
- Ionescu-Tulcea

Satz von, 52
 Isotonie, 5

 k -dimensionale Normalverteilung, 48
 Klasse, 57
 Kolmogorovsche Axiome, 37
 Konfidenzniveau, 33
 Kongruenzgeneratoren, 53
 Konvergenz
 nach Wahrsch., 26
 nach Wahrscheinlichkeit, 22
 Verteilungs-, 25
 aus Konvergenz n.W., 26
 hinr. Kriterium, 27
 Korrelationskoeffizient, 20
 Korrespondenzsatz, 38
 Kundenzahl, 66
 mittlere, 71

 Lageparameter, 41
 Laplace-Experiment, 3
 Laplace-Verteilung, 3
 Leistungsmaß, 70
 Lindeberg-Bedingung, 28
 Lindeberg-Lévy
 Satz von, 32

 Markov-Kette, 53, 55
 irreduzible, 57
 Markov-Ungleichung, 20
 Maß, 37
 induziertes, 43
 Maßraum, 37
 Modell 1, 66
 Merkmalraum, 1
 messbar, 42
 messbare Menge, 35
 Messraum, 35
 Mittlere Bedienzeit, 71
 Modell 2, 67

 Monte-Carlo-Methode, 53
 Multiplikationsregel, 13

 Normal (μ, σ^2) -Verteilung, 40
 Normalverteilung
 k -dimensionale, 48
 Normalverteilungsfunktion, 28
 \emptyset -Stetigkeit, 5
 Nulltreue, 37

 Operationstreue, 10

 Periode, 58
 Pfad, 56
 Poisson(λ)-Experiment, 7
 Poisson(λ)-Verteilung, 7
 Potenzmenge, 35
 Pseudoinverse, 54
 Pseudozufallszahlen, 53

 R-Dichten
 Transformationssatz, 47
 Randverteilung, 10
 Realisierung, 2
 Rechteck(a, b)-Verteilung, 39
 Reiss, R.-D., 24
 relative Häufigkeit, 2
 Rest-Bedienzeit, 66
 Riemann-Dichte, 38

 Satz
 Korrespondenzsatz, 38
 von der totalen Wahrscheinlichkeit, 13
 Schätzer, 21
 Schätzfunktion, 21
 schwaches Gesetz der großen Zahlen, 8, 21
 schwaches Gesetz der großen Zahlen, 21
 Sensitivität bei $\rho \rightarrow 1$, 71
 σ -Additivität, 1,37
 σ -Algebra, 35
 Skalenparameter, 41

Standard-Gauss-Verteilung, 41
 Standard-Normalverteilung, 41
 Startverteilung, 56
 Steinsche Methode, 24
 Stetigkeit
 von oben, 5
 von unten, 5
 Stichprobenkorrelationskoeffizient, 23
 Stichprobenkovarianz, 23
 Stichprobenraum, 1
 Stichprobenvarianz, 22
 stochastisch unabhängig, 14
 Sub-sigma-Additivität, 5
 Subtraktivität, 5

 Takt, 64
 Test, 33
 Träger, 3
 Transformationsformel, 17
 Transformationssatz
 für R-Dichten, 47
 Tschebyschev-Ungleichung, 21

 Übergangsgraph, 55
 Übergangsmatrix
 n-Schritt, 57
 Übergangszahldichte, 51
 Übergangsmatrix, 55
 Unabhängigkeit
 stochastische, 14
 von Ereignissen, 14, 16
 Urbildmenge, 9
 Urnenmodell, 5

 Varianz, 19, 42
 Verschiebungssatz von Steiner, 19
 Verteilung, 10, 43
 $Geo^0(p)$, 7
 Bernoulli(p), 6
 Binomial(n, p), 6
 Exponential (α), 39
 gemeinsame, 11
 Hypergeometrische (N, K, n), 7
 Normal (μ, σ^2), 40
 Poisson(λ), 7
 Rechteck(a, b), 39
 Standard-Normal, 41
 stationäre, 59
 Weibull (α, β), 39, 40
 Verteilungsfunktion, 24
 empirische, 25
 Verteilungskonvergenz, 25
 Verweilzeit
 Verteilung, 72

 W-Maß, 1, 37
 W-Verteilung, 3
 Wahrscheinlichkeit
 bedingte, 12
 Satz von der totalen, 13
 Wahrscheinlichkeitsmaß, 1, 37
 Weibull (α, β)-Verteilung, 39, 40

 Z-Dichte
 gemeinsame, 11
 Zähldichte, 3, 10
 Zentraler Grenzwertsatz, 28
 Zentraler Grenzwertsatz von Lindeberg-Lévy,
 32
 Ziehen
 mit Zurücklegen, 6
 ohne Zurücklegen, 7
 Zufallsexperiment, 1
 Zufallsvariable, 9, 43
 diskrete, 43
 Zufallszahlen, 53
 Zustandsraum, 53