Mitschrift zur Vorlesung

Einführung in die Stochastik

von Wolfgang Wefelmeyer

im Sommersemester 1997

Martin R. Elsner
$\mu_\rho \varepsilon$

30. August 1998

Inhaltsverzeichnis

1 Diskrete Experimente . . . 3
1.1 Zufallsvariablen . . . 4
1.2 Laplacesche Experimente . . . 4
2 Kombinatorik . . . 5
2.1 Permutationen . . . 5
2.2 Kombinationen . . . 5
2.3 Variationen . . . 6
3 Anwendungen . . . 6
4 Gekoppelte Experimente . . . 8
4.1 Zweistufige Experimente . . . 8
4.2 n-stufige Experimente . . . 9
5 Bedingte Wahrscheinlichkeit . . . 10
5.1 Rekonstruktion eines Wahrscheinlichkeitsmaßes aus bedingten Wahrscheinlichkeiten . . . 11
6 Unabhängigkeit . . . 12
7 Mittelwert und Erwartungswert (diskret) . . . 13
8 Die Kolmogorovschen Axiome und das Lebesgue-Borel-Maß . . . 13
9 Das Lebesgue-Integral . . . 14
9.1 Zufallsvariablen . . . 14
9.2 Lebesgue-Integral . . . 15
9.3 Induzierte Maße . . . 16
9.4 Integrale bezüglich induzierter Maße . . . 16
10 Produktmaße und unabhängige Zufallsvariablen . . . 17
11 Einige diskrete Verteilungen . . . 18
12 Stetige Wahrscheinlichkeitsmaße . . . 21
13 Verteilungsfunktion und Transformationssätze . . . 22
14 Gesetz der großen Zahl . . . 25
15 Erwartungstreue und konsistente Schätzer . . . 26
16 Zentraler Grenzwertsatz . . . 27

A Literatur **29**

1 Diskrete Experimente

(„Ein (Zufalls-)Experiment ist ein Ursachensystem, das ein vom Zufall beeinflusstes Ergebnis hervorbringt.“)

Beispiel 1.1 Werfen zweier (unterscheidbarer) Münzen
Ergebnis: Zahl oder Kopf für erste und zweite Münze:
$\Omega = \{zz, zk, kz, kk\}$ oder $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$ (Kopf=Gewinn=1)
Ereignis z.B.: $A = \{mind.\, einmal\, Kopf\} = \{zk, kz, kk\}$
$PA = P\{zk\} + P\{kz\} + P\{kk\} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$, weil $P\{\omega\} = \frac{1}{4}$ für $\omega \in \Omega$

Beispiel 1.2 Werfen zweier (*un*unterscheidbarer) Münzen
Ergebnis: Anzahl Kopf
Es ist $\Omega = \{0,1,2\}, \quad P\{0\} = p_0 = \frac{1}{4} = p_2, \quad P\{1\} = \frac{1}{2}$
$A = \{mind.\, einmal\, Kopf\} = \{1,2\}, \quad PA = P\{1\} + P\{2\} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

Definition 1.3 *Ein **diskretes Experiment** besteht aus einer endlichen oder abzählbaren Menge $\Omega = \{\omega_1, \omega_2, \ldots\}$ von Ergebnissen, denen Wahrscheinlichkeiten $P\{\omega_i\} = p_i$ zugeordnet sind. Dabei gilt $p_i \geq 0, \sum p_i = 1$ (insbesondere $p_i \leq 1$, und ω_i mit $p_i = 0$ können weggelassen werden).*

Definition 1.4 *Ein **Ereignis** ist eine Menge von Ergebnissen. Die Menge aller Ereignisse ist die **Potenzmenge** $\mathcal{P}(\Omega)$ aller Teilmengen von Ω.*
Die Wahrscheinlichkeit für eine Menge $A \subset \Omega$ ist $PA = \sum\limits_{\omega \in A} P\{\omega\} = \sum\limits_{\omega_i \in A} p_i$

Bemerkung 1.5 Folgende Mengenoperationen für $A, B \subset \Omega$ werden im weiteren benötigt:
Das Komplement $A^c = \{\omega : \omega \notin A\}$
Die Vereinigung $A \cup B = \{\omega : \omega \in A \, oder \, \omega \in B\}$
Der Durchschnitt $A \cap B = \{\omega : \omega \in A \, und \, \omega \in B\}$
Daraus abgeleitet:
Die Differenz $A - B = A \cap B^c$
Die symmetrische Differenz $A \triangle B = (A \cup B) - (A \cap B)$
A, B heißen disjunkt, wenn $A \cap B = \emptyset$. Für diesen Fall sei $A + B := A \cup B$.
Es gelten die *Regeln von de Morgan:* $(\bigcap A_i)^c = \bigcup A_i^c$

Bemerkung 1.6 Es gilt:
$P : \mathcal{P} \mapsto [0,1]$ mit $P\Omega = 1, \quad P(A+B) = PA + PB$ wenn A und B disjunkt.
Eine solche Mengenfunktion heißt **Wahrscheinlichkeitsmaß**.

Beweis: $P\Omega = \sum\limits_{\omega_i \in \Omega} P(\omega_i) = \sum p_i = 1$
$PA + PB = \sum\limits_{\omega_i \in A} p(\omega_i) + \sum\limits_{\omega_i \in B} p(\omega_i) = \sum\limits_{\omega_i \in A+B} p(\omega_i) = P(A+B)$ □

Beispiel 1.7 Wirkung eines Medikaments bei einer Person
Ergebnis: Zustand schlechter, unverändert oder besser; $\Omega = \{-1, 0, 1\}$

Beispiel 1.8 Klausur mit 60 Teilnehmern
1. Gleiche Punktzahl ausgeschlossen: $\Omega = \{(\omega_1, \ldots, \omega_{60}) : Permutation\ von\ 1, \ldots, 60\}$
2. Gleiche Punktzahl möglich:
$60 = \sum_{j=1}^{k} a_j$, $a_j \geq 1$, $k \geq 1$; Vektoren $(\underbrace{1 \ldots 1}_{a_1}, \underbrace{2 \ldots 2}_{a_2}, \ldots, \underbrace{k \ldots k}_{a_k})$
$\Omega = \{Permutationen\ dieses\ Vektors\}$

1.1 Zufallsvariablen

Beispiel 1.9 Zwei Münzen
$\Omega = \{(i,j) : i,j = 0,1\}, \quad P\{(i,j)\} = \frac{1}{4}$
$X(i,j) = i + j$ Anzahl Kopf
$P\{mind.\ einmal\ Kopf\} = P\{X \geq 1\} = P\{\omega : X(\omega) \geq 1\} = P\{(i,j) : i + j \geq 1\}$
$= P\{(0,1),(1,0),(1,1)\} = \frac{3}{4}$

Definition 1.10 *Eine Abbildung $X : \Omega \mapsto \mathbb{R}$ heißt **Zufallsvariable**. Die (induzierte) **Verteilung** von X ist gegeben durch*

$$P^X\{b\} = P\{X = b\} = P\{\omega : X(\omega) = b\}$$

Nach Bem. 1.6 ist dadurch ein Wahrscheinlichkeitsmaß auf dem Wertebereich $X(\Omega)$ von X gegeben. (Da Ω endlich oder abzählbar, ist auch $X(\Omega)$ höchstens abzählbar.)

Beispiel 1.11 Zwei Münzen
$\Omega = \{(i,j) : i,j = 0,1\}, \quad P\{(i,j)\} = \frac{1}{4}$
$X_1(i,j) = i$ erste Projektion
$X_2(i,j) = j$ zweite Projektion
a) $P\{erste\ Münze\ Kopf\} = P\{X_1 = 1\} = P\{(1,0),(1,1)\}$
b) $P\{mind.\ einmal\ Kopf\} = P\{X_1 + X_2 \geq 1\}$

Bemerkung 1.12 Die Verteilung einer Zufallsvariablen lässt sich durch die Urbildabbildung ausdrücken:

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

Insbesondere gilt: $X^{-1}\{b\} = \{\omega : X(\omega) = b\}$, also $P^X\{b\} = P(X^{-1}\{b\})$.

1.2 Laplacesche Experimente

Definition 1.13 *Sei Ω endlich. Die **Laplace-Verteilung** oder **Gleichverteilung** über Ω ist definiert durch $P\{\omega\} = \frac{1}{|\Omega|}$. Daraus ergibt sich*

$$PA = \frac{|A|}{|\Omega|} = \frac{Anzahl\ der\ günstigen\ Ergebnisse}{Anzahl\ der\ möglichen\ Ergebnisse}$$

2 Kombinatorik

2.1 Permutationen

Sei $k \in \{1, \ldots, n\}$. Die Anzahl der k-Tupel aus verschiedenen Zahlen $1, \ldots, n$ ist

$$(n)_k = n(n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Begründung:

1. Stelle des k-Tupels: n Möglichkeiten
2. " $n-1$ Möglichkeiten
$\vdots$
k. " $n-k+1$ Möglichkeiten

Äquivalent:

- Anzahl der Möglichkeiten, k unterscheidbare Kugeln auf n Fächer ohne Mehrfachbelegung zu verteilen (Begründung: Kugel $\Leftrightarrow$ Stelle im Tupel, Fach $\Leftrightarrow$ Zahl)
- Anzahl der injektiven Abbildungen $\{1, \ldots, k\} \mapsto \{1, \ldots, n\}$
- Spezialfall: Anzahl der Möglichkeiten, n Zahlen anzuordnen: $(n)_n = n!$

Beispiel 2.1 5 Gäste auf 10 Hotelzimmer: $(10)_5 = 30240$ Möglichkeiten

Beispiel 2.2 6 Kinder nebeneinander für Gruppenaufnahme: $6! = 720$ Möglichkeiten

2.2 Kombinationen

Sei $k \in \{0, \ldots, n\}$. Die Anzahl der k-elementigen Teilmengen in $\{1, \ldots, n\}$ ist

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \quad \textbf{(Binomialkoeffizient)}$$

Begründung:
Es existieren $(n)_k$ Tupel als Permutationen.
Für ein gegebenes k-Tupel existieren $k!$ Anordnungen $\Rightarrow \frac{(n)_k}{k!} = \binom{n}{k}$

Äquivalent:

- Anzahl der Möglichkeiten, k nicht unterscheidbare Kugeln ohne Mehrfachbelegung auf n Fächer zu verteilen (Begründung: Die belegten Fächer beschreiben die Teilmenge)

- Anzahl der Abbildungen $\{1,\ldots,n\} \mapsto \{0,1\}, \quad |f^{-1}\{1\}| = k$ (Begründung: Fach belegt $\Leftrightarrow$ Funktion=1)

Beispiel 2.3 6 aus 49: $\binom{49}{6} = 13983816$ Möglichkeiten

Beispiel 2.4 a Nullen und b Einsen in einer Reihe anordnen:
D.h. von $a+b$ Stellen a mit Nullen belegen: $\binom{a+b}{a} = \binom{a+b}{b}$

Beispiel 2.5 Gegeben seien n_j Zeichen der Sorte j und k Sorten, also insgesamt $n_1+\ldots+n_k = n$ Zeichen
Dann ist die Anzahl der Anordnungen: $\binom{n}{n_1 \cdots n_k} = \frac{n!}{n_1! \cdots n_k!}$ (**Multinomialkoeffizient**)

Beweis: Für erstes Zeichen: $\binom{n}{n_1}$
Für zweites Zeichen: $\binom{n}{n_2}$
Für erstes und zweites Zeichen: $\binom{n}{n_1}\binom{n-n_1}{n_2} = \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}$ u.s.w. □

2.3 Variationen

Sei $k \in \mathbb{N}$. Die Anzahl der k-Tupel aus Zahlen $1,\ldots,n$ ist n^k

Äquivalent:

- k unterscheidbare Kugeln auf n Fächer verteilen
- Anzahl aller Abbildungen $\{1,\ldots,k\} \mapsto \{1,\ldots,n\}$

3 Anwendungen

Beispiel 3.1 Wie wahrscheinlich ist es, mit 4 Würfeln mind. eine 6 zu würfeln?
$\Omega = \{(i_1,\ldots,i_4) : i_k = 1,\ldots,6,\ k = 1,\ldots,4\}$
Trick: Komplement betrachten!
Möglich: 6^4, günstig für Komplement: 5^4
$\Rightarrow W = 1 - \frac{5^4}{6^4} \approx 0,52$
(bei 3 Würfeln: $1 - \frac{5^3}{6^3} = 1 - \frac{125}{216} < \frac{1}{2}$)

Beispiel 3.2 Olympialotterie 1971 (Glücksspirale)
Gewinnzahl wie folgt:
In einer Trommel sind je 7 Kugeln mit den Ziffern $0,\ldots,9$. 7 Kugeln werden nacheinander entnommen und zu einer Zahl angeordnet.
Allgemein: je m Kugeln mit $0,\ldots,9$; ziehen $k \leq m$.
Möglich: $(10m)_k$
Günstig für $(0,\ldots,0)$: $(m)_k$, $W = \frac{(m)_k}{(10m)_k} \approx 8,3 \cdot 10^{-10}$
Günstig für $(1,\ldots,k)$: m^k, $W = \frac{m^k}{(10m)_k} \approx 1,3 \cdot 10^{-7}$

Beispiel 3.3 Geburtstag mit s Personen
Wie wahrscheinlich ist es, dass mind. 1 Gast am gleichen Tag wie der Gastgeber Geburtstag hat?
Es sind $s-1$ Gäste anwesend, also $\Omega = \{(i_1,\ldots,i_{s-1}) : i_j \in \{1,\ldots,365\}\}$
Wenn der Geburtstag des Gastgebers i_s ist, dann sind $\{(i_1,\ldots,i_{s-1}) : i_j \in \{1,\ldots,365\}, i_j \neq i_s\}$ ungünstig. Also
$W = 1 - \frac{364^{s-1}}{365^{s-1}}$

$s-1$	5	10	15	20	40	253
W	0,01	0,03	0,04	0,05	0,10	0,50

Beispiel 3.4 Geburtstag mit s Personen
Wie wahrscheinlich ist es, dass mind. 2 am gleichen Tag Geburtstag haben?
Für $s \geq 366$ ist natürlich $W = 1$.
Möglich: $(i_1,\ldots,i_s)$ mit $i_j \in \{1,\ldots,365\}$
Ungünstig: alle i_j verschieden
$\Rightarrow W = 1 - \frac{(365)_s}{365^s}$

s	5	10	15	20	23	40	50	55
W	0,03	0,12	0,25	0,41	0,51	0,89	0,97	0,99

Satz 3.5 (Einschluss-Ausschluss-Formel von Poincaré-Sylvester)
Seien $A_1,\ldots,A_n \subset \Omega$ endliche Teilmengen. Es gilt

$$\begin{aligned} |A_1 \cup A_2| &= |A_1| + |A_2| - |A_1 \cap A_2| \\ |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| \\ &\quad + |A_1 \cap A_2 \cap A_3| \\ |\bigcup_{i=1}^{n} A_i| &= \sum_i |A_i| - \sum_{i \neq j} |A_i \cap A_j| + \ldots + (-1)^{n-1} |\bigcap_{i=1}^{n} A_i| \end{aligned}$$

Ist $|\bigcap_{j=1}^{m} A_{i_j}| = c(m)$, $i_j \neq i_k$ für $j \neq k$, dann gilt: $|\bigcup_{i=1}^{n} A_i| = \sum_{m=1}^{n} (-1)^{m-1} \binom{n}{m} c(m)$, denn es existieren $\binom{n}{m}$ Möglichkeiten, m verschiedene Indizes aus $1,\ldots,n$ zu wählen.

Beispiel 3.6 k Kugeln werden auf n Fächer verteilt, Mehrfachbelegung ist möglich. Wie wahrscheinlich ist es, dass mind. ein Fach leer bleibt?
Für $k < n$: $W = 1$
Für $i = 1,\ldots,n$ sei A_i die Menge der Zuordnungen, bei denen Fach i leer bleibt.
Die Anzahl der Zuordnungen, bei denen die Fächer $i_1,\ldots,i_m$ leer bleiben, ist dann $|A_{i_1} \cap \ldots \cap A_{i_m}| = (n-m)^k$.
Die Anzahl der Zuordnungen, bei denen mind. ein Fach leer bleibt, ist
$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{m=1}^{n} (-1)^{m-1} \binom{n}{m} (n-m)^k$.
Die Anzahl aller Zuordnungen ist n^k. Also
$W = \frac{1}{n^k} |\cup A_i| = \sum_{m=1}^{n} (-1)^{m-1} \binom{n}{m} \left(1 - \frac{m}{n}\right)^k$.

Beispiel 3.7 (Rencontre, matching, Koinzidenzen)
$1,\ldots,n$ werden permutiert. Wie wahrscheinlich ist es, dass mind. eine Zahl unverändert bleibt?
Für $i = 1\ldots,n$ sei A_i die Menge der Permutationen $(\nu_1,\ldots,\nu_n)$ mit $\nu_i = i$.
Es gilt:
$|A_{i_1} \cap \ldots \cap A_{i_n}| = (n-m)! = c(m)$
$|\bigcup A_i| = \sum_{m=1}^{n} (-1)^{m-1} \binom{n}{m} (n-m)! = n! \sum_{m=1}^{n} (-1)^{m-1} \frac{1}{m!} = n! \left(1 - \sum_{m=0}^{n} \frac{(-1)^m}{m!}\right)$
$W = 1 - \sum_{m=0}^{n} \frac{(-1)^m}{m!} \approx 1 - e^{-1} \approx 0,63$ (für $n = 4$ schon ca. 0,625)

Beispiel 3.8 Ein Los der Größe n enthalte k defekte Stücke. Mit welcher Wahrscheinlichkeit enthält eine Stichprobe vom Umfang n k defekte Stücke?
Anzahl der möglichen Stichproben: $\binom{N}{n}$
Die ersten Fächer seien defekt, danach folgen die nicht defekten.
k auf Fächer $1,\ldots,K$: $\binom{K}{k}$
$n-k$ auf Fächer $K+1,\ldots,N$: $\binom{N-K}{n-k}$
Anzahl der günstigen Ergebnisse: $\binom{K}{k}\binom{N-K}{n-k}$

$$W = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = H_{N,K,n}\{k\} \quad für \quad k = 0,\ldots,\min\{n,K\} \quad (= 0 \quad sonst)$$

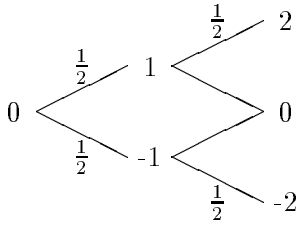
(Hypergeometrische Verteilung).
Kontrolle: $\sum_{k=0}^{\min\{n,K\}} \binom{K}{k}\binom{N-K}{n-k} = \binom{N}{n}$

Beispiel 3.9 Los $N = 1000$
Liefervertrag: höchstens 2% Ausschuss
Das Los gehe zurück, wenn eine Stichprobe mit $n = 10$ mindestens ein defektes Stück enthält.
Die Wahrscheinlichkeit, dass ein Los mit k defekten Stücken zurückgewiesen wird, ist
$H_{1000,k,10}\{1,\ldots,10\} = 1 - H_{1000,k,10}\{0\} = 1 - \frac{\binom{k}{0}\binom{1000-k}{10}}{\binom{1000}{10}} = 0,18 \quad für\, k = 20$

4 Gekoppelte Experimente

4.1 Zweistufige Experimente

Beispiel 4.1 Zwei Münzwürfe nacheinander
K,Z bedeuten 10 DM Gewinn,Verlust für Sie. Zeitlicher Verlauf Ihres Vermögens:



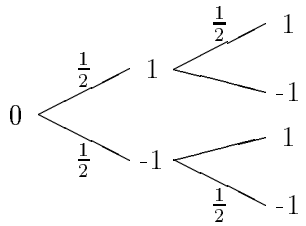
$p(1;2) = p_1(1) \cdot p_2(1;2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Definition 4.2 *Seien Ω_1, Ω_2 endliche Mengen, p_1 Wahrscheinlichkeiten auf Ω_1, p_2 Wahrscheinlichkeiten auf Ω_2. Das **gekoppelte Experiment** (zweistufiges Experiment) ist gegeben durch die Wahrscheinlichkeiten $p(\omega_1, \omega_2) = p_1(\omega_1)p_2(\omega_1;\omega_2)$*
Für das zugehörige Wahrscheinlichkeitsmaß P gilt also

$$P(A_1 \times A_2) = \sum_{(\omega_1,\omega_2)\in A_1\times A_2} p(\omega_1,\omega_2) = \sum_{\omega_1\in A_1} p_1(\omega_1) \sum_{\omega_2\in A_2} p_2(\omega_1;\omega_2) = \sum_{\omega_1\in A_1} p_1(\omega_1)P_2(\omega_1;A_2)$$

*Hängt p_2 nicht von ω_1 ab, heißt das Experiment **unabhängiges Produkt**, und es gilt $p(\omega_1,\omega_2) = p_1(\omega_1)p_2(\omega_2)$ und $P(A_1 \times A_2) = P_1A_1 \cdot P_2A_2$*

Beispiel 4.3 (Fortsetzung von 4.1) Ergebnis in jedem Schritt:



$p_1(1) = p_1(-1) = \frac{1}{2}$; $\quad p_2(\pm 1, 1) = p_2(\pm 1, -1) = \frac{1}{2} = p_2(1) = p_2(-1)$

4.2 n-stufige Experimente

Definition 4.4 *Seien $\Omega_1, \Omega_2, \ldots, \Omega_n$ endliche Mengen, $p_1, p_2(\omega_1;\cdot), \ldots, p_n(\omega_1, \ldots, \omega_{n-1};\cdot)$ Wahrscheinlichkeiten darauf. Das **gekoppelte Experiment (n-stufig)** ist gegeben durch*

$$p(\omega_1, \ldots, \omega_n) = p_1(\omega_1)p_2(\omega_1;\cdot) \cdot \ldots \cdot p_n(\omega_1, \ldots, \omega_n)$$

*Das Experiment heißt auch **stochastischer Prozess**. Ein Tupel $(\omega_1, \ldots, \omega_n)$ heißt **Pfad**. Für das zugehörige Wahrscheinlichkeitsmaß auf $\Omega_1 \times \ldots \times \Omega_n$ gilt:*

$$P(A_1 \times \ldots \times A_n) = \sum_{\omega_1\in A_1} p_1(\omega_1) \sum_{\omega_2\in A_2} p_2(\omega_1;\omega_2) \cdots \sum_{\omega_n\in A_n} p_n(\omega_1, \cdots, \omega_{n-1};\omega_n)$$

*Hängen die p_i nur von ω_{i-1} ab, so heißt der Prozess eine **Markov-Kette**. Hängen die p_i nicht von $\omega_1, \ldots, \omega_{i-1}$ ab, dann heißt das Experiment ein **unabhängiges Produkt** und es gilt*

$$P(A_1 \times \cdots \times A_n) = P_1A_1 \cdot P_2A_2 \cdot \ldots \cdot P_nA_n$$

Wir schreiben $P_1 \otimes \ldots \otimes P_n$.

Beispiel 4.5 Urne mit 2 weißen, 1 schwarzen Kugel
Zwei Kugeln werden ohne Zurücklegen gezogen. Wie wahrscheinlich ist es, zwei weiße zu ziehen?
$W = p_1(w_1)p_2(w_1;w_2) + p_1(w_2)p_2(w_2;w_1) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{3}$
Alternative 1:

$W = p_1(w)p_2(w,w) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$
Alternative 2:
$W = \frac{1}{\binom{3}{2}} = \frac{1}{3}$
Alternative 3:
$W = \frac{(2)_2}{(3)_2} = \frac{2!}{3 \cdot 2} = \frac{1}{3}$

Beispiel 4.6 Von 3 Spielen soll man 2 hintereinander gegen A oder B gewinnen. Sei B stärker als A; ist dann die Reihenfolge ABA oder BAB besser?
Seien p, q die Wahrscheinlichkeiten, gegen A, B zu gewinnen. Sie müssen $++-, -++$ oder $+++$ spielen.
$W_{ABA} = pq(1-p) + (1-p)qp + pqp = 2pq - pqp$
$W_{BAB} = 2pq - qpq$
Wegen $p > q$ ist $pqp > qpq$, also $W_{BAB} > W_{ABA}$

Beispiel 4.7 Zwei gleich starke Spieler verabreden, dass derjenige den Einsatz gewinnt, der zuerst 6 Spiele gewonnen hat. Beim Stand von 5:3 wird abgebrochen. Wie soll der Einsatz aufgeteilt werden?
Der zweite Spieler muss dreimal hintereinander gewinnen:
$W = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$, also Aufteilung 7:1.
Der erste Spieler muss $+$, $-+$ oder $--+$ spielen:
$W = \frac{1}{2} + \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{7}{8}$

5 Bedingte Wahrscheinlichkeit

Beispiel 5.1 Roulette: Zahlen $0, \ldots, 36$
0 grün, ungerade rot, gerade schwarz
a) Wahrscheinlichkeit für 7: $\frac{1}{37}$
b) Sei bekannt, dass rot eingetreten ist. Dann W. für 7: $\frac{1}{18}$
c) Sei bekannt, dass schwarz eingetreten ist. Dann W. für 7: 0
Formal:
$\Omega = \{0, \ldots, 36\}, \quad p(i) = \frac{1}{37}$
b) $A = \{rot\}, \quad B = \{7\}$
$P(B|A) = \frac{1}{18} = \frac{1/37}{18/37} = \frac{P(A \cap B)}{PA}$

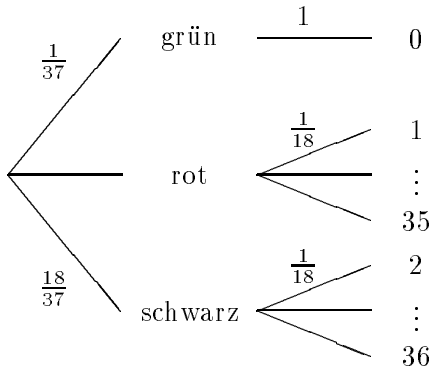
Definition 5.2 *Sei P ein Wahrscheinlichkeitsmaß auf Ω und $A, B \subset \Omega$ mit $PA > 0$. Die* ***bedingte Wahrscheinlichkeit*** *von B gegeben A ist*

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

Satz 5.3 *$P(\cdot|A)$ ist ein Wahrscheinlichkeitsmaß auf Ω.*

Beweis: 1. $P(B|A) = \frac{P(A \cap B)}{PA} \geq 0$
2. $P(\Omega|A) = \frac{PA}{PA} = 1$
3. $P(B + C|A) = \frac{P((B+C)|A)}{PA} = \frac{P(B \cap A + C \cap A)}{PA} = \frac{P(B \cap A)}{PA} + \frac{P(C \cap A)}{PA} = P(B|A) + P(C|A)$ □

Beispiel 5.4 Roulette als gekoppeltes Experiment



$p_2(rot; 7) = P(\{7\}|\{rot\})$

Beispiel 5.5 Ein roter und ein schwarzer Würfel werden geworfen.
a) Wahrscheinlichkeit für rote 6: $\frac{1}{6}$
b) Sei bekannt, dass die Augensumme 11 ist. Möglich: (5,6),(6,5); günstig: (6,5), also $W = \frac{1}{2}$
$P(B|A) = \frac{1/36}{2/36} = \frac{1}{2}$

5.1 Rekonstruktion eines Wahrscheinlichkeitsmaßes aus bedingten Wahrscheinlichkeiten

Satz 5.6 (Satz von der totalen Wahrscheinlichkeit) *Ist* $\Omega = \sum_{k=1}^{n} A_k$ *mit* $PA_k > 0$, *dann ist*

$$PB = \sum_{k=1}^{n} P(B|A_k)PA_k$$

Beweis: $PB = \sum P(B \cap A_k) = \sum \frac{P(B \cap A_k)}{PA_k} PA_k = \sum P(B|A_K)PA_k$ □

Satz 5.7 (Bayessche Regel)

$$P(A_m|B) = \frac{P(B|A_m)PA_m}{PB} = \frac{P(B|A_m)PA_m}{\sum_{k=1}^{n} P(B|A_k)PA_k}$$

Beweis: $P(A_m|B) = \frac{P(A_m \cap B)}{PB} = \frac{P(A_m \cap B)PA_m}{PA_m PB} = \frac{P(B|A_m)PA_m}{PB}$ und Satz 5.6. □

Beispiel 5.8 Diagnose:
Ein Test stuft eine Person mit 5% fälschlich als krank ein, mit 10% Wahrscheinlichkeit fälschlich als gesund.
Zustand: G, k; Diagnose: g, k

$\Omega = \{G, K\} \times \{g, k\}$
$P(k|G) = 0,05, \quad P(g|G) = 0,95; \qquad P(g|K) = 0,1, \quad P(k|K) = 0,9$
Der Anteil der Kranken in der Bevölkerung sei 2%.
$PK = 0,02 = P(K|g) + P(K|k), \quad PG = 0,98$
Wie groß ist die Wahrscheinlichkeit, dass ein als krank diagnostizierter tatsächlich krank ist?
$P(G|k) = \frac{P(k|G)PG}{P(k|G)PG+P(k|K)PK} = \frac{0,05 \cdot 0,98}{0,05 \cdot 0,98+0,9 \cdot 0,02} = 0,73$
Wie groß ist die Wahrscheinlichkeit, dass ein als gesund diagnostizierter tatsächlich krank ist?
$P(K|g) = \frac{P(g|K)PK}{P(g|K)PK+P(g|G)PG} = \frac{0,1 \cdot 0,02}{0,1 \cdot 0,02+0,95 \cdot 0,98} = 0,002$

Beispiel 5.9 Drei Niederlassungen einer Firma stellen Fernseher her, die mit den Wahrscheinlichkeiten q_1, q_2, q_3 funktionieren und mit Anteilen p_1, p_2, p_3 in ihrem Geschäft stehen; Sie kaufen einen.
a) Mit welcher Wahrscheinlichkeit funktioniert er?
Niederlassungen 1,2,3, Funktion $\pm$
Satz 5.6$\Rightarrow P(+) = P(+|1)P1 + P(+|2)P2 + P(+|3)P3 = q_1p_1 + q_2p_2 + q_3p_3$
b) Wenn er funktioniert, mit welcher Wahrscheinlichkeit stammt er von Niederlassung 1?
Mit Satz 5.7 und (a): $P(1|+) = \frac{P(+|1)P(1)}{P(+)} = \frac{q_1p_1}{q_1p_1+q_2p_2+q_3p_3}$

6 Unabhängigkeit

Beispiel 6.1 Ein roter und ein schwarzer Würfel werden geworfen.

1.a) W. für rote 6: $\frac{1}{6}$
b) W. für rote 6, gegeben Augensumme 11: $\frac{1/36}{2/36} = \frac{1}{2}$
c) W. für rote 6, gegeben Augensumme $\neq$11: $\frac{5/36}{34/36} = \frac{5}{34}$
2.b) W. für rote 6, gegeben Augensumme ungerade: $\frac{3}{18} = \frac{1}{6}$
c) W. für rote 6, gegeben Augensumme gerade: $\frac{3}{18} = \frac{1}{6}$

Formal:
$A = \{(i,j) : i + j \text{ gerade}\}, \quad B = \{(i,j) : i = 6\}$
$P(B|A) = \frac{P(B \cap A)}{PA} = PB$, ebenso $P(B|A^c) = PB$, also $P(A \cap B) = PA \cdot PB$

Satz 6.2 *Wenn $P(B|A) = P(B|A^c)$, dann $P(B \cap A) = PB \cdot PA$*

Beweis: $\frac{P(B \cap A)}{PA} = \frac{P(B \cap A^c)}{PA^c}$, also $PB = P(A \cap B) + P(A^c \cap B) = P(A \cap B)\left(1 + \frac{PA^c}{PA}\right) = \frac{P(A \cap B)}{PA}$
□

Definition 6.3 *Seien P ein Wahrscheinlichkeitsmaß und A, B zwei Ereignisse.*
*Dann heißen A, B **unabhängig**, wenn $P(A \cap B) = PA \cdot PB$*

Definition 6.4 *Ereignisse $A_1, \ldots, A_n$ heißen **unabhängig**, wenn $P(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} PA_{i_j}$ für alle verschiedenen Indizes.*

Bemerkung 6.5 Paarweise Unabhängigkeit reicht nicht, wie folgendes Beispiel zeigt: Zwei Münzen:
A: erste Münze Kopf
B: zweite Münze Kopf
C: genau eine Münze Kopf
Es ist $P(A \cap B) = \frac{1}{4} = PA \cdot PB$, ebenso $P(A \cap C), P(B \cap C)$; aber $P(A \cap B \cap C) = 0 \neq PA \cdot PB \cdot PC = \frac{1}{8}$

Bemerkung 6.6 A von A unabhängig oder A von A^c unabhängig gilt nur für $PA = 1$ oder $PA = 0$.

7 Mittelwert und Erwartungswert (diskret)

Beispiel 7.1 Kirmes
Sie gewinnen bei 2 von 2000 Losen DM 1000,-, sonst einen Trostpreis von DM 1,-. Lohnt es sich, DM 2,- für ein Los zu bezahlen?
Erwarteter Gewinn: $\frac{2}{2000} \cdot 1000 + \frac{1998}{2000} \cdot 1 = 1,999$

Definition 7.2 *Sei P ein diskretes Wahrscheinlichkeitsmaß auf $\mathbb{R}$ mit $P(A_i) = p_i$.*
*Der **Mittelwert** von P ist $\sum p_i a_i$*

Definition 7.3 *Sei X eine diskrete Zufallsvariable.*
*Der **Erwartungswert** von X ist der Mittelwert von P^X, $\sum P(X = a_i) a_i$*

Beispiel 7.4 In Beispiel 7.1:
$P\{1000\} = \frac{2}{2000} = P(X = 1000), \quad P\{1\} = \frac{1998}{2000} = P(X = 1)$

Beispiel 7.5 Sei P die Gleichverteilung auf $a_1, \ldots, a_n \in \mathbb{R}$. Dann ist der Mittelwert $\frac{1}{n} \sum_{i=1}^{n} a_i$

Beispiel 7.6 Erwartete Augenzahl bei einmaligem Würfeln: 3,5

8 Die Kolmogorovschen Axiome und das Lebesgue-Borel-Maß

Kolmogorovsche Axiome: (1933)
Ein *Wahrscheinlichkeitsraum* (Zufallsexperiment) $(\Omega, \mathcal{F}, P)$ besteht aus
- einem Grundraum $\Omega \neq \emptyset$
- einer σ-Algebra $\mathcal{F}$
- einem Wahrscheinlichkeitsmaß P auf $\mathcal{F}$

Definition 8.1 *Ein Mengensystem $\mathcal{F}$ heißt **σ-Algebra**, wenn*

1. $\Omega \in \mathcal{F}$

2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

3. $A_i \in \mathcal{F}, i \in I\!N \Rightarrow \bigcup A_i \in \mathcal{F}$

Bemerkung 8.2 Es heißen $\omega \in \Omega$ Ergebnis und $A \in \mathcal{F}$ Ereignis.

Definition 8.3 *Eine Funktion $P : \Omega \mapsto \overline{I\!R} = [-\infty, +\infty]$ heißt **Wahrscheinlichkeitsmaß** auf $\mathcal{F}$, wenn*

1. $PA \geq 0, A \in \mathcal{F}$

2. $P\Omega = 1$

3. $A_i \in \mathcal{F}, i \in I\!N$, *paarweise disjunkt* $\Rightarrow P\left(\sum A_i\right) = \sum PA_i$ *(σ-Additivität)*

*Ist nur Bedingung 2 nicht erfüllt, so sprechen wir von einem **Maß**.*

Lemma 8.4 *Sei $\mathcal{S} \subset \mathcal{P}(\Omega)$ ein Mengensystem. Dann existiert eine kleinste σ-Algebra, die $\mathcal{S}$ enthält. Sie heißt die von $\mathcal{S}$ erzeugte σ-Algebra $\sigma(\mathcal{S})$*

Bemerkung 8.5 $<r, s>$ bezeichne im Weiteren ein Intervall mit den Grenzen r und s.

Definition 8.6 *Die **Borel-Algebra** $\mathcal{B}^m$ ist die kleinste σ-Algebra über $I\!R^m$, die alle beschränkten Quader $\times<r_i, s_i>$ enthält.*

Bemerkung 8.7 Statt der Quader reichen auch die abgeschlossenen unteren Quadranten $(-\infty, r_1] \times \ldots \times (-\infty, r_m]$.

Definition 8.8 *Für einen beschränkten Quader definieren wir*

$$\lambda_m(\times<r_i, s_i>) = \prod(s_i - r_i)$$

Satz 8.9 *Sei $\mathcal{R}$ die Menge der beschränkten Quader.*
*Die eben definierte Mengenfunktion $\lambda_m : \mathcal{R} \mapsto [0, \infty)$ lässt sich eindeutig zu einem Maß auf $\mathcal{B}^m$ fortsetzen. Es heißt **Lebesgue-Borel-Maß**.*

9 Das Lebesgue-Integral

9.1 Zufallsvariablen

Definition 9.1 *Seien $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ Mengen mit σ-Algebren. Eine Funktion $X : \Omega \mapsto \Omega'$ heißt **messbar**, wenn für $A' \in \mathcal{F}'$ gilt:*

$$\{X \in A'\} (= \{\omega : X(\omega) \in A'\}) = X^{-1}A' \in \mathcal{F}$$

Satz 9.2 *Ist $\mathcal{F}' = \sigma(\mathcal{S})$ und $X^{-1}S \in \mathcal{F}$, $S \in \mathcal{S}$, dann ist X messbar.*

Beweis: $T = \{A' \subset \Omega' : X^{-1}A' \in \mathcal{F}\}$ ist σ-Algebra und $\mathcal{S} \subset T$, also $\sigma(\mathcal{S}) \subset T$ □

Bemerkung 9.3 Messbare Funktionen $X : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^m, \mathcal{B}^m)$ heißen **Zufallsvektoren**, für m=1 **Zufallsvariablen**.

Satz 9.4 *Die Menge der Zufallsvariablen ist abgeschlossen unter Linearkombinationen, abzählbaren Suprema/Infima und Limites (soweit nicht $\pm\infty$).*

Beweis: Wir betrachten Urbilder offener Halbstrahlen.
1. $a \neq 0$: $\{aX < r\} = \begin{cases} \{X < r/a\} & a > 0 \\ \{X > r/a\} & a < 0 \end{cases}$
$a = 0$: $\{aX < r\} = \begin{cases} \Omega & r > 0 \\ \emptyset & r < 0 \end{cases}$
2. $\{X + Y < r\} = \bigcup_{u \in \mathbb{Q}} \{X < u\} \cap \{Y < r - u\}$
$\supset$: $X + Y < u + r - u = r$
$\subset$: $X < r - Y \Leftrightarrow \exists u \in \mathbb{Q}$: $X < u < r - Y$
3. $\{\inf X_n < r\} = \bigcup_n \{X_n < r\}$
4. $\sup X_n = -\inf(-X_n)$
5. $\lim X_n = \inf_n \sup_{k \geq n} X_k$ □

Definition 9.5 $1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$ *heißt* ***Indikatorfunktion*** *von A.*

Definition 9.6 *Eine Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ heißt (messbare)* ***Treppenfunktion****, wenn sie nur endlich viele Werte animmt:*

$$X = \sum_{i=1}^{n} a_i 1_{A_i}$$

Satz 9.7 *Sei $X : \Omega \mapsto [0, \infty)$ Zufallsvariable.*
Dann existiert eine messbare Treppenfunktion X_n mit $X_n \uparrow X$.

Beweis: $X_n(\omega) = \begin{cases} (k-1)2^{-n} & (k-1)2^{-n} \leq X(\omega) \leq k2^{-n} \\ n & X(\omega) \geq n \end{cases}$ □

9.2 Lebesgue-Integral

Gegeben seien im Folgenden ein Lebesgue-Maß λ_1 auf $\mathcal{B}$ und das Intervall (a, b).
Für $f = 1_{(a,b)}$ gilt $\int f(x)\, dx = b - a$ bzw. allgemeiner $\int 1_A(\omega)\mu(d\omega) = \mu A$.
Im Weiteren sei $X^+ = X1_{(X \geq 0)}$, $X^- = -X1_{(X \leq 0)}$, sodass $X = X^+ - X^-$.

Definition 9.8 *Gegeben seien $(\Omega, \mathcal{F})$ und ein Maß μ*

1. *Für messbare Indikatorfunktionen sei*

$$\int 1_A d\mu = \mu A$$

2. *Für messbare Treppenfunktionen sei*

$$\int \sum a_i 1_{A_i} d\mu = \sum a_i \mu A_i$$

3. *Für eine nichtnegative Zufallsvariable X mit $X_n \uparrow X$, X_n messbare Treppenfunktion, sei*

$$\int X \, d\mu = \lim \int X_n d\mu$$

4. *Eine Zufallsvariable X heißt* ***Lebesgue-integrierbar****, wenn*
$\int X^+ d\mu < \infty$ oder $\int X^- d\mu < \infty$. Dann ist

$$\int X \, d\mu = \int X^+ d\mu - \int X^- d\mu$$

Satz 9.9 (Satz von der monotonen Konvergenz (Levi))
Sei $0 \leq X_n \uparrow X$, X_n messbar. Dann gilt $\int X_n d\mu \to \int X d\mu$.

Satz 9.10 (Satz von der dominierten Konvergenz (Lebesgue))
Seien $X_n \to X$, X_n messbar, $-Y \leq X_n \leq Y$, Y messbar und endlich integrierbar.
Dann gilt $\int X_n d\mu \to \int X \, d\mu$.

9.3 Induzierte Maße

Satz 9.11 *Gegeben $(\Omega, \mathcal{F})$, $\mu|\mathcal{F}$, $(\Omega', \mathcal{F}')$, $X : \Omega \mapsto \Omega'$ messbar.*
Dann definiert $\mu^X A' = \mu(X^{-1}A')$, $A' \in \mathcal{F}'$ ein Maß μ^X auf $\mathcal{F}'$, das durch X ***induzierte Maß****.*

Beweis: 1. $\mu^X A' = \mu(X^{-1}A') \geq 0$
2. Sind $A'_n \in \mathcal{F}$, $n \in \mathbb{N}$, paarweise disjunkt, so auch $X^{-1}A'_n$, $n \in \mathbb{N}$.
$\mu^X(\sum A'_n) = \mu(X^{-1} \sum A'_n) = \mu \sum X^{-1} A'_n = \sum \mu X^{-1} A'_n = \sum \mu^X A'_n$ □

Satz 9.12 *Gegeben $(\Omega'', \mathcal{F}'')$, $Y : \Omega' \mapsto \Omega''$ messbar. Dann ist $(\mu^X)^Y = \mu^{Y \circ X}$*

Beweis: Für $C \in \mathcal{F}$:
$(\mu^X)^Y C = \mu^X(Y^{-1}C) = \mu X^{-1}(Y^{-1}C) = \mu(Y \circ X)^{-1} C = \mu^{Y \circ X} C$ □

9.4 Integrale bezüglich induzierter Maße

Gegeben seien im Folgenden $(\Omega, \mathcal{F}, \mu) \overset{X}{\mapsto} (\Omega', \mathcal{F}') \overset{Y}{\mapsto} (\Omega'', \mathcal{F}'') = (\mathbb{R}, \mathcal{B})$

Satz 9.13 $\int Y \, d\mu^X = \int Y \circ X \, d\mu$

Beweis: („algebraische Induktion")
0. $Y = 1_B$: $\int 1_B d\mu^X = \mu^X B = \mu(X^{-1}B) = \int 1_{X^{-1}B} d\mu = \int 1_B \circ X \, d\mu$
1. $Y = \sum b_i 1_{B_i}$ mit 0. und Linearität.
2. $Y \geq 0$. Wähle Treppenfunktion $Y_n \uparrow Y$. Dann auch $Y_n \circ X \uparrow Y \circ X$
$\int Y \, d\mu^X \overset{Levi}{=} \lim \int Y_n d\mu^X \overset{1.}{=} \lim \int Y_n \circ X \, d\mu \overset{Levi}{=} \int Y \circ X \, d\mu$
3. Y bel., $Y = Y^+ - Y^-$. Dann $(Y \circ X)^\pm = Y^\pm \circ X$. Jetzt 2. anwenden. □

Definition 9.14 *Gegeben $(\Omega, \mathcal{F}, P)$. Der **Mittelwert** von P ist $\int xP(dx)$ (falls existiert). Der **Erwartungswert** einer Zufallsvariablen X ist der Mittelwert von P^X.*

$$EX = \int x dP^X(dx) \overset{\substack{Satz\ 9.13 \\ mit\ Y(X)=X}}{=} \int X \, dP$$

Bemerkung 9.15 Diskrete Verteilung: Für $P\{a_i\} = p_i$ ist $\int xP(dx) = \sum a_i p_i$

10 Produktmaße und unabhängige Zufallsvariablen

Definition 10.1 *Gegeben $(\Omega_i, \mathcal{F}_i)$, $i \in I\!N$.*
*Die von den messbaren Zylindern $A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \cdots$ mit $A_i \in \mathcal{F}_i$ erzeugte σ-Algebra heißt **Produkt-σ-Algebra** $\bigotimes_{i=1}^{\infty} \mathcal{F}_i$.*

Satz 10.2 *Sind $(\Omega_i, \mathcal{F}_i, P_i)$, $i \in I\!N$, Wahrscheinlichkeitsräume, so lässt sich die Mengenfunktion $P(A_1 \times \ldots \times A_n \times \Omega_{n+1} \times \ldots) = \prod_{i=1}^n P_i A_i$ eindeutig zu einem Wahrscheinlichkeitsmaß auf $\bigotimes \mathcal{F}_i$ fortsetzen. Es wird $\bigotimes P_i$ geschrieben und heißt **Produktmaß**.*

Definition 10.3

1. *Gegeben $(\Omega, \mathcal{F}, P)$. Zwei Ereignisse $A, B \in \mathcal{F}$ heißen **unabhängig**, wenn $P(A \cap B) = PA \cdot PB$.*
2. *Zwei Mengensysteme $\mathcal{G}, \mathcal{H} \in \mathcal{F}$ heißen **unabhängig**, wenn A, B unabhängig für alle $A \in \mathcal{G}$, $B \in \mathcal{H}$.*
3. *Gegeben $(\Omega, \mathcal{F}, P)$, $(\Omega', \mathcal{F}')$, $(\Omega'', \mathcal{F}'')$, $X : \Omega \mapsto \Omega'$, $Y : \Omega \mapsto \Omega''$ messbar. Dann heißen X, Y **unabhängig**, wenn $X^{-1}A'$, $Y^{-1}B''$ unabhängig für alle $A' \in \mathcal{F}'$, $B'' \in \mathcal{F}''$.*

Bemerkung 10.4 Folgende Aussagen sind äquivalent:

1. X, Y unabhängig: $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$ $\forall A \in \mathcal{F}', B \in \mathcal{F}''$
2. $P^{(X,Y)}(A \times B) = P^X A \cdot P^Y B$ $\forall A \in \mathcal{F}', B \in \mathcal{F}''$
3. $P^{(X,Y)} = P^X \otimes P^Y$, also $P^{(X,Y)}$ Produktmaß

Bemerkung 10.5 Für mehr als zwei Mengen: $A_i,\ i \in \mathbb{N}$, sind unabhängig, wenn $P(A_{i_1} \cap \ldots \cap A_{i_m}) = \prod_{j=1}^{m} PA_{i_j}$ für alle verschiedenen $i_1, \ldots, i_m$.

Bemerkung 10.6 Seien Ω_1, Ω_2 endliche Mengen.
Dann ist die Gleichverteilung auf $\Omega_1 \times \Omega_2$ das unabhängige Produkt der Gleichverteilungen auf Ω_1 und Ω_2.

Beweis: $\frac{|A_1 \times A_2|}{|\Omega_1 \times \Omega_2|} = \frac{|A_1| \cdot |A_2|}{|\Omega_1| \cdot \Omega_2|} = \frac{|A_1|}{|\Omega_1|} \cdot \frac{|A_2|}{|\Omega_2|}$ □

11 Einige diskrete Verteilungen

Satz 11.1 *Gegeben $(\Omega, \mathcal{F})$, $\mathcal{F}$ enthält Einpunktmengen, $\omega_i \in \Omega$ und $p_i \geq 0,\ i \in I\!N,\ \sum p_i = 1$. Dann definiert $PA = \sum_{\omega_i \in A} p_i,\ A \in \mathcal{F}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$, es heißt **diskret**.*

Beweis: 1. $PA \geq 0$
2. $P\Omega = \sum_{\omega_i \in \Omega} p_i = 1$
3. $A_i \in \mathcal{F},\ i \in \mathbb{N}$, paarweise disjunkt.
Dann $P(\sum A_j) = \sum_{\omega_i \in \Sigma A_j} p_i = \sum_j \sum_{\omega_i \in A_j} p_i = \sum_j PA_j$ □

Bemerkung 11.2 Ein diskretes Wahrscheinlichkeitsmaß P auf einem Produktraum $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ ist ein unabhängiges Produkt genau dann, wenn $P(\{\omega_1, \omega_2\}) = P_1\{\omega_1\}P_2\{\omega_2\}$. Dabei sind $P_1,\ P_2$ die Randverteilungen von P: $P_1(A_1) = P(A_1 \times \Omega_2),\ P_2(A_2) = P(\Omega_1 \times A_2)$

Beweis: $\Rightarrow$: klar.
$\Leftarrow$: $P(A_1 \times A_2) = \sum_{(\omega_1,\omega_2) \in A_1 \times A_2} P\{(\omega_1, \omega_2)\} = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} P_1\{\omega_1\}P\{\omega_2\}$
$= \sum_{\omega_1 \in A_1} P_1\{\omega_1\} \sum_{\omega_2 \in A_2} P_2\{\omega_2\} = P_1A_1 \cdot P_2A_2$ □

Definition 11.3 *Gegeben $(\Omega, \mathcal{F}, P)$. Eine Zufallsvariable X heißt **diskret**, wenn $P^X|\mathcal{B}$ diskret ist.*

Beispiel 11.4 (Kapitel 1.2) X heißt **Laplace-verteilt** oder **gleichverteilt** auf $a_1, \ldots, a_n$, wenn $P(X = a_i) = \frac{1}{n},\ i = 1, \ldots, n$
Ohne Zufallsvariable:
$P|\mathcal{B}$ heißt gleichverteilt auf $a_1, \ldots, a_n$, wenn $P\{a_i\} = \frac{1}{n},\ i = 1, \ldots, n$

Beispiel 11.5 (Kapitel 3) X heißt **hypergeometrisch verteilt** $(X \sim H_{N,K,n})$, wenn $P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},\ k = 0, \ldots, \min\{n, K\}$

Definition 11.6 *X heißt **Bernoulli-verteilt** $(X \sim B_{1,p})$, wenn*

$$P(X = 1) = p,\ P(X = 0) = 1 - p$$

Beispiel 11.7 Sei P ein Wahrscheinlichkeitsmaß auf Ω und $A \in \mathcal{F}$ ein Ereignis.
Dann ist $X = 1_A$ Bernoulli-verteilt mit $p = PA$:
$P(X = 1) = P\{\omega : 1_A(\omega) = 1\} = P\{\omega : \omega \in A\} = PA$

Definition 11.8 *Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen.*
Die Verteilung von $X_1 + \ldots + X_n$ heißt ***Faltung****.*

Satz 11.9 *Sind X, Y unabhängig mit Werten in $\mathbb{Z}$, so gilt*

$$P(X + Y = i) = \sum_k P(X = i - k)P(Y = k)$$

Beweis: $\{X + Y = i\} = \bigcup_k \{X = i - k, Y = k\}$, also $P(X + Y = i) = \sum_k P(X = i - k, Y = k) = \sum_k P(X = i - k)P(Y = k)$ □

Beispiel 11.10 Seien X, Y unabhängig und auf $1, \ldots, 6$ Laplace-verteilt.
$P(X + Y = i) = \sum_k P(X = i - k)P(Y = k) = \frac{1}{36}|\{k : 1 \le k \le 6,\ 1 \le i - k \le 6\}|$
$= \frac{1}{36}|\{k : max\{1, i - 6\} \le k \le \min\{6, i - 1\}|$

Definition 11.11 *X heißt* ***binomialverteilt*** *($X \sim B_{n,p}$), wenn*

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},\ k = 0, \ldots, n$$

Satz 11.12 *Sind $X_1, \ldots, X_n$ unabhängig und $B_{1,p}$-verteilt, dann ist $X_1 + \ldots + X_n$ $B_{n,p}$-verteilt.*

Beweis: δ_i stehe für 0 oder 1.
$P(X_1 = \delta_1, \ldots, X_n = \delta_n) = \prod P(X_i = \delta_i) = p^{\Sigma\delta_i}(1 - p)^{n - \Sigma\delta_i}$
Für $\sum \delta_i = k$ gibt es $\binom{n}{k}$ Möglichkeiten, also
$P(X_1 + \ldots + X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ □

Beispiel 11.13 Sie spielen n-mal ein Spiel, das Sie mit Wahrscheinlichkeit p gewinnen. Wie wahrscheinlich ist es, dass sie genau einmal gewinnen?
$P(X_1 + \ldots + X_n = 1) = np(1 - p)^{n-1} \approx np$, wenn p klein ist.

Definition 11.14 *X heißt* ***geometrisch verteilt*** *($X \sim G_p$), wenn*

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \ldots$$

Definition 11.15 *Sind $X_1, X_2, \ldots$ unabhängig und Bernoulli-verteilt, dann heißt*

$$W_1 = \inf\{k : X_k = 1\} \quad mit\ \inf \emptyset = \infty$$

die ***Wartezeit bis zum ersten Erfolg****.*
Induktiv definiert man: $W_n = \inf\{k > W_{n-1} : X_k = 1\}$

Satz 11.16 *Sind $X_1, X_2, \ldots$ unabhängig und $B_{1,p}$-verteilt, dann ist W_1 G_p-verteilt.*

Beweis: $P(W_1 > k) = P(X_1 = \ldots = X_k = 0) = (1-p)^k$
$P(W_1 = k) = P((W_1 > k-1) \setminus (W_1 > k)) = P(W_1 > k-1) - P(W_1 > k) = (1-p)^{k-1} - (1-p)^k = p(1-p)^{k-1}$ □

Beispiel 11.17 Münzwurf: $p = \frac{1}{2}$, $P(W_1 = k) = 2^{-k}$

Definition 11.18 *X heißt **negativ binomialverteilt** ($X \sim B^-_{n,p}$), wenn*

$$P(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k = n, n+1, \ldots$$

Satz 11.19 *Sind $X_1, X_2, \ldots$ unabhängig und $B_{1,p}$-verteilt, dann ist W_n $B^-_{n,p}$-verteilt.*

Beweis: Setze $Y_1 = W_1,\ Y_2 = W_2 - W_1, \ldots, Y_n = W_n - W_{n-1}$.
Dann ist Y_i die Wartezeit vom $(i-1)$-ten zum i-ten Erfolg.
Es gilt $W_n = \sum_{i=1}^n Y_i$, Y_i unabhängig und G_p-verteilt.
$P(Y_1 = k_1, \ldots, Y_n = k_n) = \prod_{i=1}^n P(Y_i = k_i) = \prod_{i=1}^n p(1-p)^{k_i - 1} = p^n (1-p)^{\sum_{i=1}^n k_i - n}$
$P(W_n = k) = P(\sum_{i=1}^n Y_i = k) = \sum_{\sum_{i=1}^n k_i = k} p^n (1-p)^{k-n} = \binom{k-1}{n-1} p^n (1-p)^{k-n}$ □

Bemerkung 11.20
1. Es gilt $B^-_{1,p} = G_p$.
2. Der früheste Zeitpunkt für n Erfolge ist n.

Definition 11.21 *X heißt **Poisson-verteilt** ($X \sim P_\lambda,\ \lambda > 0$), wenn*

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Die Poisson-Verteilung approximiert die Binomialverteilung, wenn n groß und p klein ist:

Satz 11.22 $B_{n,\frac{\lambda}{n}}\{k\} \to P_\lambda\{k\}, \quad n \to \infty$

Beweis: $B_{n,\frac{\lambda}{n}}\{k\} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \underbrace{\frac{n!}{(n-k)!n^k}}_{\to 1} \frac{\lambda^k}{k!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \to e^{-\lambda} \frac{\lambda^k}{k!}$ □

Beispiel 11.23 Ein Bäcker rechnet pro Brötchen im Schnitt 10 Rosinen. Wie wahrscheinlich ist es, dass Ihr Brötchen keine Rosinen enthält?
Gegeben: n Rosinen, $\frac{n}{10}$ Brötchen
Die W., dass eine geg. Rosine in Ihr Brötchen gerät, ist $\frac{10}{n}$. Also
$W = B_{n,\frac{10}{n}}\{0\} \to P_{10}\{0\} = e^{-10}$

Bemerkung 11.24 Der Erwartungswert einer diskreten Zufallsvariablen X mit $P(X = a_i) = p_i$ ist $EX = \sum a_i p_i$ (siehe Kapitel 7).
Somit folgt für die bekannten Verteilungen:
Bernoulli-Verteilung $X \sim B_{1,p}$: $EX = p$
Binomial-Verteilung $X \sim B_{n,p}$: $EX = np$
Poisson-Verteilung $X \sim P_\lambda$: $EX = \lambda$
(Nach Satz 11.22 gilt $B_{n,\frac{\lambda}{n}}\{k\} \to P_\lambda\{k\}$ und somit gilt für EX $n\frac{\lambda}{n} = \lambda \to \lambda$.)

Beispiel 11.25 Alternative zu Bsp. 11.23:
Erwartete Anzahl der Rosinen ist 10, also:
$W \approx P_\lambda\{0\} = e^{-10}$

Beispiel 11.26 Eine Buchseite enthält im Schnitt 4 Druckfehler. Wie groß ist der Anteil der Seiten mit mind. 2 Druckfehlern?
$P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-4}(1+4) = 1 - 5e^{-4}$

12 Stetige Wahrscheinlichkeitsmaße

Satz 12.1 *Sei $p : \mathbb{R}^m \mapsto [0,\infty)$ messbar mit $\int p(x)\,dx = 1$.*
Dann ist $P(A) = \int_A p(x)\,dx = \int 1_A(x)p(x)\,dx$, $A \in \mathcal{B}^m$, ein Wahrscheinlichkeitsmaß auf $\mathcal{B}^m$;
*es heißt **stetig mit Dichte p**.*

Beweis: 1. $P(A) \geq 0$ weil $p \geq 0$
2. $P(\mathbb{R}^m) = \int\limits_{\mathbb{R}^m} p(x)\,dx = 1$
3. $A_n \in \mathcal{B}^m$, $n \in \mathbb{N}$, disjunkt:
$P(\sum A_n) = \int_{\Sigma A_n} p(x)\,dx = \int 1_{\Sigma A_n}(x)p(x)\,dx = \int \sum 1_{A_n}(x)p(x)\,dx$
$= \int \lim_{N\mapsto\infty} \sum_{n=1}^N 1_{A_n}(x)p(x)\,dx \overset{Levi}{=} \lim_{N\mapsto\infty} \int \sum_{n=1}^N 1_{A_n}p(x)\,dx = \lim_{n\mapsto\infty} \sum_{n=1}^N \int 1_{A_n}(x)p(x)\,dx$
$= \sum_{n=1}^\infty \int 1_{A_n}(x)p(x)\,dx = \sum_{n=1}^\infty P(A_n)$ □

Lemma 12.2 *Ist $P|\mathcal{B}^{m_1+m_2}$ ein Wahrscheinlichkeitsmaß mit Dichte p, so ist P ein unabhängiges Produkt genau dann, wenn $p(x,y) = p_1(x)p_2(y)$.*

Beweis: (nur für $\Leftarrow$)
$P(A \times B) = \int\limits_{A\times B} p(x,y)\,d(x,y) = \int\int 1_A(x)1_B(y)p_1(x)p_2(y)\,dx\,dy = \int\limits_A p_1(x)\,dx \int\limits_B p_2(y)\,dy =$
$P_1(A)P_2(B)$ □

Beispiel 12.3 Sei $A \in \mathcal{B}^m$ mit $0 < \lambda(A) < \infty$.
Die Gleichverteilung U_A über A hat die Dichte

$$p(x) = \frac{1}{\lambda_m(A)} 1_A(x)$$

Definition 12.4 *Die **Exponentialverteilung** E_a hat die Dichte*

$$p(x) = \begin{cases} \frac{1}{a} e^{-x/a} & x > 0 \\ 0 & sonst \end{cases}$$

Satz 12.5 $E_a((t+c,\infty)|(t,\infty)) = E_a(c,\infty) \quad \forall t, c$

Beweis: $\frac{E_a(t+c,\infty)}{E_a(t,\infty)} = \frac{e^{\frac{-(t+c)}{a}}}{e^{-\frac{t}{a}}} = e^{-\frac{c}{a}} = E_a(c,\infty)$ □

13 Verteilungsfunktion und Transformationssätze

Definition 13.1 *Sei $P|\mathcal{B}$ ein Wahrscheinlichkeitsmaß. Dann heißt $F(r) = P(-\infty, r]$, $r \in \mathbb{R}$, **Verteilungsfunktion** von P.*

Satz 13.2 *Sei F eine Verteilungsfunktion. Dann gilt:*

1. *F ist nichtfallend.*
2. $\lim_{r \to -\infty} F(r) = 0, \quad \lim_{r \to \infty} F(r) = 1$
3. *F ist rechtsstetig: $F(r_n) \to F(r)$ für $r_n \downarrow r$*

Beweis: 1. $(-\infty, a] \subset (-\infty, b]$, $a < b$, also $F(a) = P(-\infty, a] \leq P(-\infty, b] = F(b)$
2. $r_n \to -\infty$: $(-\infty, r_n] \to \emptyset$. Mit Übung 34 folgt $F(r_n) = P(-\infty, r_n] \to 0$
$r_n \to \infty$: $(-\infty, r_n] \to \mathbb{R}$. Mit Übung 34 folgt $F(r_n) = P(-\infty, r_n] \to 1$
3. $r_n \downarrow r$: $(-\infty, r_n] \to (-\infty, r]$. Mit Übung 34 folgt $F(r_n) \to F(r)$. □

Satz 13.3 *Zu jeder Funktion F mit den Eigenschaften 1.-3. aus Satz 13.2 existiert genau ein Wahrscheinlichkeitsmaß P mit Verteilungsfunktion F.*

Satz 13.4 *Ist F differenzierbar, dann ist $f = F'$ Dichte von P.*

Beweis: $F(b) - F(a) = \int_a^b F'(x)\,dx$, insbesondere $P(a,b] = \int_a^b F'(x)\,dx$. Daraus folgt $PA = \int_A F'(x)\,dx$ □

Bemerkung 13.5 Hat P die Dichte p, so gilt $F(r) = \int_{-\infty}^{r} p(x)\,dx$

Satz 13.6 (Transformationssatz) *X habe die Verteilungsfunktion F, $T : \mathbb{R} \mapsto \mathbb{R}$ sei stetig und streng monoton steigend und $T(\mathbb{R}) = (a,b)$, $-\infty \leq a < b \leq \infty$.*
Dann hat $T \circ X$ (oder $P^{T \circ X}$) die Verteilungsfunktion

$$y \mapsto \begin{cases} 0 & y \leq a \\ F(T^{-1}(y)) & a < y < b \\ 1 & y \geq b \end{cases}$$

Beweis: $P^{T\circ X}(-\infty,r] = P^X(T \le r) = P^X(-\infty, T^{-1}(r)] = F(T^{-1}(r)), \quad a < r < b$ □

Satz 13.7 *X habe die Dichte p, $T : \mathbb{R} \mapsto \mathbb{R}$ sei stetig differenzierbar und $T' \neq 0$. Dann hat $T \circ X$ die Dichte*

$$y \mapsto \begin{cases} \frac{p(T^{-1}(y))}{|T'(T^{-1}(y))|} & y \in [a,b] \\ 0 & sonst \end{cases}$$

Beweis: Sei o.E. T steigend. Mit Satz 13.6 ist $\frac{d}{dy}F(T^{-1}(y)) = \frac{1}{T'(T^{-1}(y))}p(T^{-1}(y))$ für $y \in [a,b]$ □

Beispiel 13.8 Hat X die Dichte p, dann hat $aX + b$, $a \neq 0$, die Dichte $y \mapsto \frac{1}{|a|}p\left(\frac{y-b}{a}\right)$.

Definition 13.9 *Die eindimensionale **Normalverteilung** N_{μ,σ^2}, $\mu \in \mathbb{R}$, $\sigma^2 > 0$, hat die Dichte*

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

*Die **Standard-Normalverteilung** $N_{0,1}$ hat die Dichte $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.*

Beispiel 13.10 Hat X die Verteilung $N_{0,1}$, dann hat $\sigma X + \mu$ die Verteilung N_{μ,σ^2} für $\sigma > 0$ (folgt aus Bsp. 13.8).

Beispiel 13.11 Sei X nichtnegativ mit Dichte p. Dann hat X^a für $a \neq 0$ die Dichte

$$y \mapsto \frac{1}{|a|}y^{\frac{1-a}{a}}p(y^{1/a})$$

Beweis: Satz 13.7 mit $T(X) = X^a$; $T^{-1}(y) = y^{1/a}$
$T'(X) = aX^{a-1}$, $T'(T^{-1}(y)) = ay^{\frac{a-1}{a}}$ □

Definition 13.12 *Die **Gamma-Verteilung** $\Gamma_{a,b}|\mathcal{B}$, $a,b > 0$, hat die Dichte*

$$p(x) = \frac{1}{a^b\Gamma(b)}x^{b-1}e^{-x/a}, \quad x > 0$$

(Für $b = 1$ ist dies gerade E_a; wenn $X \sim \Gamma_{1,b}$, dann $aX \sim \Gamma_{a,b}$)

Beispiel 13.13 Ist X N_{0,σ^2}-verteilt, dann ist X^2 $\Gamma_{2\sigma^2,1/2}$-verteilt.

Beweis: $T(x) = x^2$ ist nicht injektiv; aber N_{0,σ^2} ist symmetrisch um 0. Also reicht es, zweimal $(0,\infty)$ zu betrachten.

Beispiel 13.11 für $a = 2$, $p(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$
$\Rightarrow \frac{2}{|2|}y^{\frac{1-2}{2}}e^{\frac{-y}{2\sigma^2}}\frac{1}{\sqrt{2\pi}\sigma} = \frac{1}{(2\sigma^2)^{1/2}\sqrt{\pi}}y^{1/2-1}e^{\frac{-y}{2\sigma^2}}$ und $\sqrt{\pi} = \Gamma(\frac{1}{2})$ □

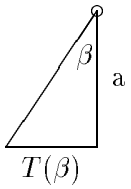
Definition 13.14 *Die* ***Cauchy-Verteilung*** $C_a|\mathcal{B}$, $a > 0$, *hat die Dichte*

$$y \mapsto \frac{1}{\pi a}\frac{1}{1+(y/a)^2}$$

Bemerkung 13.15 Der Mittelwert existiert nicht; es genügt, $a = 1$ zu betrachten:

$$EX = \underbrace{\int_0^\infty xp(x)dx}_{1/\pi \int_0^\infty \frac{x}{1+x^2}dx=\infty} + \underbrace{\int_{-\infty}^0 xp(x)dx}_{-\infty}$$

Beispiel 13.16 Die Helligkeitsverteilung unter einer Straßenlaterne
a sei die Höhe der Laterne, β der Winkel zwischen einem Lichtstrahl und der Senkrechten, $T(\beta)$ der Abstand des vom Strahl getroffenen Punktes zur Laterne.



β ist gleichverteilt über $(-\frac{\pi}{2}, \frac{\pi}{2})$; $\tan\beta = \frac{T(\beta)}{a}$, $\quad T(\beta) = a\tan\beta$
$P(T \le X) = P\{\beta : a\tan\beta \le x\} = P(\beta \le \arctan\frac{x}{a}) = \frac{1}{\pi}(\arctan\frac{x}{a} - \frac{\pi}{2})$
Mit Satz 13.4 ist die Dichte $p(x) = \frac{1}{\pi a}\frac{1}{1+(x/a)^2}$

Satz 13.17 *Hat (X,Y) die Dichte p, dann hat $X+Y$ die Dichte*

$$x \mapsto \int p(X-Y,Y)\,dY$$

Sind X, Y unabhängig mit Dichten p_1, p_2, dann hat $X+Y$ die Dichte

$$x \mapsto \int p_1(X-Y)p_2(Y)\,dY$$

(vgl. Satz 11.9 diskret)

Beweis: Wegen Satz 13.3 genügt es, Halbstrahlen zu betrachten:
$P(X+Y \le r) = \int\limits_{X+Y\le r} p(X,Y)d(X,Y) = \int(\int\limits_{X\le r-Y} p(X,Y)dX)dY = \int(\int_{X\le r} p(X-Y,Y)dX)dY =$
$\int_{X\le r}(\int p(X-Y,Y)dY)dX = \int_{-\infty}^{r}(\int p(X-Y,Y)dY)dX$ □

Beispiel 13.18 Sind X, Y unabhängig und $\Gamma_{a,b}$- bzw. $\Gamma_{a,c}$-verteilt, dann ist $X+Y$ $\Gamma_{a,b+c}$-verteilt.

Beweis: Die Dichte von $X+Y$ ist nach Satz 13.17
$\int p_{a,b}(x-y)p_{a,c}(y)dy = \frac{1}{a^{b+c}}\frac{1}{\Gamma(b)\Gamma(c)}\int_0^x (x-y)^{b-1}e^{-\frac{x-y}{a}}y^{c-1}e^{-y/a}dy$
und mit der Transformation $y \mapsto xy$
$= e^{-x/a}\frac{1}{a^{b+c}}\frac{1}{\Gamma(b)\Gamma(c)}x^{b+c-1}\underbrace{\int_0^1 (1-y)^{b-1}y^{c-1}dy}_{\frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)}} = \frac{1}{a^{b+c}}\frac{1}{\Gamma(b+c)}e^{-x/a}x^{b+c-1}$ für $x > 0$, $\quad 0$ sonst. □

14 Gesetz der großen Zahl

Im Weiteren sei $(\Omega, \mathcal{F}, P)$ gegeben.

Bemerkung 14.1 Der Erwartungswert ist linear: $E(cX) = cE(X)$, $E(X+Y) = EX + EY$

Definition 14.2 *Sei X eine Zufallsvariable mit endlichem Erwartungswert EX. Dann heißt*

$$VarX = E(X - EX)^2 = E(X^2) - (EX)^2$$

*die **Varianz** von X.*

Bemerkung 14.3 Es gilt $Var(cX) = c^2 VarX$

Beweis: $Var(cX) = E(cX - E(cX))^2 = c^2 E(X - EX)^2 = c^2 VarX$ □

Bemerkung 14.4 $(VarX)^{1/2}$ ist ein Maß für die Streuung von X (bzw. von P^X).

Beispiel 14.5 $X \sim E_a$:
$EX = \int_0^\infty x \frac{1}{a} e^{-x/a} dx = a \int_0^\infty x e^{-x} dx = a \left([-xe^{-x}]_0^\infty - \int_0^\infty (-e^{-x}) dx \right) = a$
$EX^2 = \int_0^\infty x^2 \frac{1}{a} e^{-x/a} dx = a^2 \int_0^\infty x^2 e^{-x} dx = a^2 \left([-x^2 e^{-x}]_0^\infty - 2 \int_0^\infty x(-e^{-x}) dx \right) = 2a^2$,
also $VarX = E(X^2) - (EX)^2 = 2a^2 - a^2 = a^2$

Definition 14.6 *Seien X, Y Zufallsvariablen mit endlicher Varianz. Dann heißt*

$$Cov(X,Y) = E(X - EX)(Y - EY)$$

***Kovarianz** von X und Y.*

Bemerkung 14.7 $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n VarX_i + 2 \sum_{i<j} Cov(X_i, X_j)$

Beweis: O.B.d.A. sei $EX_i = 0$ (sonst setze $X_i' := X_i - EX_i$): $Var(\sum X_i) = E(\sum X_i)^2 = E(\sum X_i^2 + \sum_{i \neq j} X_i X_j) = \sum EX_i^2 + \sum_{i \neq j} E(X_i X_j)$ □

Satz 14.8 *Sind X, Y unabhängig und X, Y, XY integrierbar, so gilt*

$$EXY = EX \cdot EY$$

Beweis: $EXY = \int xy P^{(X,Y)}(d(x,y)) = \int xy P^X(dx) P^Y(dy) = \int x P^X(dx) \cdot \int y P^Y(dy) = EX \cdot EY$ □

Korollar 14.9 *Sind $X_1, \ldots, X_2$ unabhängig mit endlichen Varianzen, so gilt:*

$$Var(\sum X_i) = \sum Var(X_i)$$

Beweis: Wegen Bem. 14.7 und Satz 14.8 genügt es, für $i \neq j$ zu zeigen: $Cov(X_i, X_j) = E(X_i - EX_i)(X_j - EX_j) = E(X_i - EX_i) \cdot E(X_j - EX_j) = 0$ □

Satz 14.10 (Tschebyschewsche Ungleichung)

$$P(|X - EX| \geq c) \leq \frac{VarX}{c^2}$$

Beweis: $1_{(|X-EX|\geq c)} \leq \frac{1}{c^2}(X - EX)^2$ □

Satz 14.11 (Gesetz der großen Zahl) *Sind $X_1, \ldots, X_n$ unabhängig und identisch verteilt (mit endlicher Varianz), so gilt:*

$$P\left(\left|\frac{1}{n}\sum X_i - EX_1\right| \geq \varepsilon\right) \leq \frac{VarX_1}{n\varepsilon^2}$$

Beweis: Satz 14.10 mit $X := \frac{1}{n}\sum X_i$; es ist $EX = \frac{1}{n}\sum EX_i = EX_1$
$VarX = \frac{1}{n^2}\sum VarX_i = \frac{VarX_1}{n}$ □

15 Erwartungstreue und konsistente Schätzer

$\overline{X}_n$ bezeichne im Folgenden das arithmetische Mittel von $X_1, \ldots, X_n$.

Beispiel 15.1 Ein Glücksspiel habe die Gewinnwahrscheinlichkeit p, p sei nicht bekannt; wir spielen n-mal.
$X_1, \ldots, X_n$ sind dann unabhängig und $B_{1,p}$-verteilt.
Ein **Schätzer** für p ist der Anteil der gewonnenen Spiele:

$$T_n = \frac{1}{n}|\{i : X_i = 1\}| = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad EX_1 = p$$

Je größer n ist, umso genauer ist T_n.
T_n ist **konsistent** für p (Satz 14.11):

$$P(|T_n - p| \geq \varepsilon) \leq \frac{VarX_1}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \to 0$$

wegen $VarX_1 = (0-p)^2(1-p) + (1-p)^2 p = p(1-p)$.
T_n ist auch **erwartungstreu** für p:

$$ET_n = E\left(\frac{1}{n}\sum X_i\right) = EX_1 = p$$

Beispiel 15.2 N Taxis seien durchnumeriert; wir wollen N schätzen und schreiben n Taxi-Nummern auf. Diese können durch n unabhängige, auf $1,\ldots,N$ gleichverteilte Zufallsvariablen $X_1,\ldots,X_n$ beschrieben werden.
1. Es gilt $EX_1 = \frac{N+1}{2}$, $N = 2EX_1 - 1$
Weil $\overline{X}_n = \frac{1}{n}\sum X_i$ für EX_1 konsistent und erwartungstreu ist, ist $T_n = 2\overline{X}_n - 1$ konsistent und erwartungstreu für N.
2. $T'_n = \max\{X_1,\ldots,X_n\}$ ist wegen $T'_n \leq N$ nicht erwartungstreu, aber konsistent:
$P(|T'_n - N| \geq 1) = P(T'_n < N) = P(X_1 < N,\ldots,X_n < N) = (P(X_1 < N))^n = \left(\frac{N-1}{N}\right)^n$
3. Verbesserung:
Seien $x_1,\ldots,x_n$, $x_i \leq x_{i+1}$, die beobachteten Taxinummern. Schätze die Lücke zwischen x_i und x_{i+1} mit $\frac{T'_n}{n}$, also $T''_n = \frac{n+1}{n}T'_n$

Satz 15.3

1. *Sind $X_1,\ldots,X_n$ unabhängig und identisch verteilt mit endlichem Erwartungswert, so ist $\overline{X}_n$ erwartungstreu für EX_1.*

2. *Ist die Varianz endlich, so ist $\overline{X}_n$ konsistent für EX_1.*

Beweis: 1. Bem. 14.1, 2. Satz 14.11 □

Beispiel 15.4 Seien $X_1,\ldots,X_n$ unabhängig und E_a-verteilt, a sei unbekannt. Wir suchen einen Schätzer für die Varianz:
Nach Beispiel 14.5 ist $VarE_a = a^2$, nach Satz 15.3 ist $\overline{X}_n$ konsistent für a.
Also ist $\overline{X}_n^2$ konsistent für $a^2 = VarE_a$
$P(|\overline{X}_n^2 - a^2| \geq \varepsilon) = P(|\overline{X}_n - a||\overline{X}_n + a| \geq \varepsilon) \leq P(|\overline{X}_n - a||\overline{X}_n + a| \geq \varepsilon, |\overline{X}_n + a| \leq 3a)$
$+P(|\overline{X}_n + a| > 3a) \leq P(|\overline{X}_n - a| \geq \frac{\varepsilon}{3a}) + P(|\overline{X}_n + a| > 3a) \to 0$

Allgemein gilt

Bemerkung 15.5 Ist T_n konsistent für a und f stetig, so ist $f(T_n)$ konsistent für $f(a)$.

Beispiel 15.6 Seien $X_1,\ldots,X_n$ unabhängig und identisch verteilt mit Mittelwert μ und Varianz σ^2. Wir suchen einen Schätzer für σ^2:
1. Sei μ bekannt. Es gilt $\sigma^2 = VarX_1 = E(X_1 - \mu)^2$. Also ist $\frac{1}{n}\sum(X_i - \mu)^2$ konsistent und erwartungstreu für σ^2.
2. Sei μ unbekannt. Dann Schätzer $\frac{1}{n}\sum(X_i - \overline{X}_n)^2$:
$T_n = \frac{1}{n}\sum(X_i - \mu)^2 - 2(\overline{X}_n - \mu)\frac{1}{n}\sum(X_i - \mu) + (\overline{X}_n - \mu)^2$. Jetzt $P(|\overline{X}_n - \mu| \geq \varepsilon) \to 0$ anwenden.

16 Zentraler Grenzwertsatz

Wenn $X_1,\ldots,X_n$ unabhängig und identisch verteilt sind mit $VarX_1 < \infty$, so gilt nach dem Gesetz der großen Zahl $\frac{1}{n}\sum_{i=1}^n X_i \to EX_1$. Ausserdem gilt $Var(\frac{1}{n}\sum X_i) = \frac{1}{n}VarX_1$. Also ist $E(n^{-1/2}\sum(X_i - EX_1))^2 = Var(n^{-1/2}\sum X_i) = VarX_1$.

Satz 16.1 (Zentraler Grenzwertsatz) *Sind $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit Mittelwert $\mu = EX_1$ und Varianz $\sigma^2 = E(X_1 - EX_1)^2$, so gilt für beliebige a, b mit $a < b$*

$$P\left(a < n^{-1/2}\sum_{i=1}^{n}(X_i - \mu) < b\right) \to N_{0,\sigma^2}(a,b) = \frac{1}{\sqrt{2\pi}\sigma}\int_a^b e^{-\frac{x^2}{2\sigma^2}}dx$$

Beispiel 16.2 (Fortsetzung zu 15.6)
Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit Mittelwert μ und Varianz σ^2, und $E(X_1 - \mu)^4$ sei endlich.
1. Sei μ bekannt. Dann gilt für $T_n = \frac{1}{n}\sum(X_i - \mu)^2$:
$n^{1/2}(T_n - \sigma^2) = n^{-1/2}\sum((X_i - \mu)^2 - \sigma^2)$ ist asymptotisch normal mit Varianz
$E(((X_1 - \mu)^2 - \sigma^2)^2) = E(X_1 - \mu)^4 - \sigma^4$
2. Sei μ unbekannt. Dann gilt für $T_n' = \frac{1}{n}\sum(X_i - \overline{X}_n)^2$:
$n^{1/2}(T_n' - \sigma^2) = n^{-1/2}\sum((X_i - \overline{X}_n)^2 - \sigma^2) = n^{-1/2}\sum((X_i - \mu)^2 - \sigma^2)$
$+2 \underbrace{n^{1/2}(\overline{X}_n - \mu)}_{\substack{in\ Wahrscheinlichkeit \\ beschränkt}} \underbrace{\frac{1}{n}\sum(X_i - \mu)}_{\to 0} + \underbrace{(\overline{X}_n - \mu)n^{-1/2}(\overline{X}_n - \mu)}_{\to 0},$
gleiche asymptotische Verteilung wie T_n.

Beispiel 16.3 (Messfehler)
Seien $X_1, , \ldots, X_n$ Messungen für eine unbekannte physikalische Konstante μ. Die Messfehler $X_i - \mu$ seien unabhängig und identisch verteilt mit Mittelwert 0 und bekannter Varianz $\sigma^2 < \infty$.
Dann ist $\overline{X}_n$ konsistenter Schätzer für μ.
In welchem Bereich liegt der Schätzfehler $\overline{X}_n - \mu$ mit Wahrscheinlichkeit 0,95?
$n^{1/2}\frac{\overline{X}_n - \mu}{\sigma}$ ist ungefähr $N_{0,1}$-verteilt (wegen $Var\left(n^{1/2}\frac{\overline{X}_n - \mu}{\sigma}\right) = 1$)
Bekannt ist: $N_{0,\sigma}(-1.96, 1.96) \approx 0,95$
$P(|\overline{X}_n - \mu|) < 1,96n^{-1/2}\sigma) = P\left(\left|n^{1/2}\frac{\overline{X}_n - \mu}{\sigma}\right| < 1,96\right) \approx 0.95$
z.B. $\sigma = 1,\ n = 100:\quad 1,96n^{-1/2}\sigma = 0,196$

Beispiel 16.4 (Rundungsfehler)
Zahlen $R_1, \ldots, R_n$ werden auf ganze Zahlen gerundet: $R_i = Z_i + S_i, \quad Z_i \in \mathbb{Z},\ S_i \in [-\frac{1}{2}, \frac{1}{2})$
$\sum R_i$ hat Rundungsfehler $\sum S_i$
Ann.: S_i unabhängig und gleichverteilt auf $[-\frac{1}{2}, \frac{1}{2})$
Es gilt $ES_1 = 0,\ VarS_1 = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}$
$P(|n^{-1/2}\sum S_i| < t) \to N_{0,\frac{1}{12}}(-t, t)$
z.B. $n = 100$: Rundungsfehler $< 5,7$ mit Wahrscheinlichkeit 0,95.

Beispiel 16.5 Normalapproximation der Binomialverteilung
Ist X verteilt nach $B_{n,p}$, so ist $n^{-1/2}(X - np)$ ungefähr $N_{0,p(1-p)}$-verteilt:
$X \sim X_1 + \ldots + X_n,\ X_i$ unabhängig $B_{1,p}$-verteilt (Satz 11.12).
Es gilt $EX_1 = p,\ VarX_1 = p(1-p)$.
Jetzt Satz 16.1 auf $n^{-1/2}(X - np) \sim n^{-1/2}\sum(X_i - p)$ anwenden.

Anhang A

Literatur

[1] Pfanzagl: *Elementare Wahrscheinlichkeitsrechnung*, De Gruyter 1991

[2] Mathar R. Pfeifer D.: *Stochastik für Informatiker*, Teubner 1990

[3] Krengel U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg 1991

[4] Behnen K. Neuhaus G.: *Grundkurs Stochastik*, Teubner 1987

Index

σ-Algebra, 13

Bayessche Regel, 11
Borel, 14
Borel-Algebra, 14

de Morgan, 3
Dichte, 21
diskrete Zufallsvariable, 18
diskretes Wahrscheinlichkeitsmaß, 18

Einschluss-Ausschluss-Formel, 7
Ereignis, 3, 14
Ergebnis, 14
Erwartungswert, 13, 17
Experiment
    diskretes, 3
    Zufalls-, 13

Faltung, 19

Gesetz der großen Zahl, 26

Indikatorfunktion, 15
induziertes Maß, 16

Kolmogorovsche Axiome, 13
Kovarianz, 25

Lebesgue, 14, 16
Lebesgue-integrierbar, 16
Levi, 16

Maß, 14
    induziertes, 16
    Lebesgue-Borel-, 14
    Produkt-, 17
    Wahrscheinlichkeits-, 3, 14
Markov-Kette, 9

messbar, 14
Mittelwert, 13, 17

Pfad, 9
Poincaré, 7
Potenzmenge, 3
Produktmaß, 17

Satz
    Transformations-, 22
    von der dominierten Konvergenz, 16
    von der monotonen Konvergenz, 16
Schätzer, 26
    erwartungstreuer, 26
    konsistenter, 26
stetiges Wahrscheinlichkeitmaß, 21
stochastischer Prozess, 9
Sylvester, 7

Treppenfunktion, 15
Tschebyschew, 26
    -Ungleichung, 26

unabhängig
    -e Ereignisse, 12, 17
    -e Mengensysteme, 17
    -e Wahrscheinlichkeitsmaße, 12
    -e Zufallsvariablen, 17
    -es Produkt, 9

Varianz, 25
Verteilung, 4
    Bernoulli-, 18
    Binomial-, 19
    Cauchy-, 24
    Exponential-, 22
    Gamma-, 23
    geometrische, 19

Gleich-, 4, 18
Hypergeometrische, 8
hypergeometrische, 18
Laplace-, 4, 18
negative Binomial-, 20
Normal-, 23
Poisson-, 20
Standard-Normal-, 23
Verteilungsfunktion, 22

Wahrscheinlichkeit
bedingte, 10
Wahrscheinlichkeitsmaß, 3
diskretes, 18
stetiges, 21
unabhängige ∼e, 12
Wahrscheinlichkeitsraum, 13
Wartezeit, 19

Zentraler Grenzwertsatz, 28
Zufallsvariable, 4, 15
unabhängige ∼n, 17
Zufallsvektor, 15