Einführung in die Mathematische Stochastik

Bernhard Lorenz
(lorenzb@informatik.uni-muenchen.de)

Skript nach der Vorlesung von Dr. Pruscha
Ludwig-Maximilians-Universität München

Sommersemester '97
Letzte Änderung: 11. September 1997

Inhaltsverzeichnis

Kapitel 1

Diskrete Wahrscheinlichkeitsräume

1.1 Einführung und Beispiele

$$Stochastik = Wahrscheinlichkeitstheorie + Statistik$$

In der Stochastik werden Vorgänge untersucht, deren Ablauf vom Zufall (mit-) bestimmt wird. Sofern sie einer objektiven Beschreibung zugänglich sind, wollen wir diese *Zufallsexperimente* nennen.

1.1.1 Beispiele

1. Geschlecht (männlich / weiblich) eines Neugeborenen
2. Gewicht eines Neugeborenen
   Modell: Normalverteilung (mit gewichteten Parametern μ, σ)[1].
   Frage: Wie groß ist die Wahrscheinlichkeit, daß das Gewicht des Neugeborenen im Intervall $[3000, 3800]$ liegt?
3. 900-maliges Würfeln mit auszählen der gewürfelten 'Sechser'.
   Modell: Binomialverteilung mit Parametern $n = 900$, $p = \frac{1}{6}$ [2].
   Frage: Wie groß ist die Wahrscheinlichkeit, mindestens 150 mal '6' zu würfeln?
4. Zählen der - innerhalb einer Stunde - eingehenden Anrufe.
   Modell: Poissonverteilung (mit Parameter λ) [3].
   Frage: Wie groß ist die Wahrscheinlichkeit, daß innerhalb von 12 Stunden x oder mehr Anrufe eingehen?
5. Entnahme und Prüfung (defekt / in Ordnung) von 100 Stück aus einer Sendung von 10000 Glühbirnen.
6. Anzahl der Augen $(1, \ldots, 6)$ beim Werfen eines Würfels[4].

Die Wahrscheinlichkeitstheorie stellt für ein Zufallsexperiment unter Abzug der inhaltlichen Bedingungen (Abstraktion) ein mathematisches Modell auf. Konkrete Fragestellungen können dann auf der Ebene dieses Modells bearbeitet werden.

1.2 Ergebnisraum, Ereignisse, Wahrscheinlichkeit

Alle möglichen Ergebnisse (Werte), die bei der Durchführung eines Zufallsexperimentes auftreten können, fassen wir zum Ergebnisraum (Grundmenge) Ω zusammen. Jedes Element $\omega \in \Omega$ heißt *Ergebnis* oder *Realisation*.

[1] siehe 4.2 unten
[2] siehe 1.5
[3] siehe 1.2.5 unten
[4] Zu Beginn der Vorlesung werden häufig Beispiele aus dem Glücksspiel verwendet (einfach, anschaulich, historisch)

Beispiele aus 1.1.1:

1. $\Omega = \{m, w\}$ oder $\Omega = \{0, 1\}$, mögliches Ergebnis: $\omega = 0$
2. $\Omega = \{0, 1, \ldots, \infty\} \equiv \mathbb{R}_+$, mögliches Ergebnis: $\omega = 3475, 34[g]$
3. $\Omega = \{0, 1, \ldots, 900\}$, mögliches Ergebnis: $\omega = 157$
4. $\Omega = \{0, 1, \ldots\} \equiv \mathbb{Z}_+ \equiv \mathbb{N}_0$, mögliches Ergebnis: $\omega = 17(Anrufe)$

Ω wird vor der Durchführung des Zufallsexperimentes festgelegt. Je nach Fragestellung / Interessenlage sind unterschiedliche Festlegungen möglich.

zu Beispiel 3: Falls auch die maximale Länge von $'1er', '2er', \ldots$ -Sequenzen interessiert, so setzt man etwa $\Omega = \{1, 2, 3, 4, 5, 6\}^{900}$. Mögliches Ergebnis:

$$\omega = (\underbrace{'5', '1', '2', \ldots, '6', '1'}_{900-mal})$$

zu Beispiel 2: $\Omega = \mathbb{N}$ möglich, falls auf $100g$ gerundet wird.
In Kapitel 1 bis 3 werden nur Fälle mit höchstens abzählbarem (endlich oder abzählbar) Ereignisraum Ω betrachtet. (Ein $\Omega = \mathbb{R}_+$ wird erst ab Kapitel 4 zugelassen)
Häufig interessiert weniger das genaue Ergebnis ω eines Zufallsexperimentes, sondern mehr die Frage, ob die Realisation eine bestimmte Eigenschaft besitzt (= 'ob ein bestimmtes Ereignis aufgetreten ist')

zu Beispiel 3: Anzahl $'6er' \geq 200$ ('großes Glück').

zu Beispiel 2: $Gewicht \leq 2500$ ('Lebensgefahr').

zu Beispiel 4: $Anrufe \geq 100$ ('Kapazitätsüberschreitung').
Die *Ereignisse* eines Zufallsexperimentes sind die Teilmengen der Grundmenge Ω ('Ereignis $A \subset \Omega$ tritt ein' besagt also, daß $\omega \in A$ als Ereignis eingetreten ist.). Sind $A, B \subset \Omega$ Ereignisse, so bedeutet

$\overline{A} \equiv \Omega \setminus A$:	das Komplementärereignis ('A ist nicht eingetreten').
$A \cup B$:	Das Ereignis 'A und/oder B ist/sind eingetreten'.
$A \cap B$:	Das Ereignis 'A und B sind eingetreten'.
$\emptyset$:	Das unmögliche Ereignis.
Ω:	Das sichere Ereignis.
$\{\omega\}$:	Das Elementarereignis.
$A \cup B$:	Unvereinbare Ereignisse. (mit $A \cap B = \emptyset$)

Liegt eine Folge $A_1, A_2, \ldots$ von Ereignisse vor, so bezeichnet $\bigcup_{i=1}^{\infty} A_i$ das Ereignis: '*Mindestens ein* A_i ist eingetreten'. Sind die $A_1, A_2, \ldots$ paarweise disjunkt (d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$), so stellt $\bigcup_{i=1}^{\infty} A_i$ das Ereignis: '*Genau ein* A_i ist eingetreten' dar.
Die Menge aller Ereignisse ist die Potenzmenge $\mathfrak{P}(\Omega)$ von Ω.
Zur vollständigen Beschreibung eines Zufallsexperimentes werden den Ergebnissen $\omega \in \Omega$ 'Gewichte' (oder 'Wahrscheinlichkeiten') zugeordnet. Beim Werfen eines unverfälschten (symmetrischen) Würfels sind den Ergebnissen $'1', '2', \ldots, '6'$ jeweils das Gewicht $\frac{1}{6}$ zugeordnet, während ein (zugunsten der '6') verfälschter Würfel die Gewichte $\frac{1}{6} - \epsilon, \frac{1}{6}, \ldots, \frac{1}{6} + \epsilon$ bekommt.

1.2.1 Definitionen

a) Eine Abbildung $\mathbb{P}$ (Probability) von $\mathfrak{P}(\Omega)$ in $[0, 1]$ heißt Wahrscheinlichkeitsverteilung (oder Wahrscheinlichkeitsmaß), wenn sie die Eigenschaften erfüllt:

   (i) $\mathbb{P}(\Omega) = 1$ ('Normierung')

(ii) $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für jede Folge $A_1, A_2, \ldots$ von *paarweise disjukten* Ereignissen ('σ-Additivität')

b) Ein diskreter Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$, mit:

(i) Ω eine nichtleere, höchstens abzählbare Menge

(ii) $\mathfrak{P}(\Omega)$ ist die Menge aller Teilmengen

(iii) $\mathbb{P} : \mathfrak{P}(\Omega) \to [0, 1]$ ist eine Wahrscheinlichkeitsverteilung

Bemerkung zu a)(ii): Die Einschränkung 'paarweise disjunkt' ist *wesentlich*. Für die nicht paarweise disjunkten $A_i = \Omega$ gilt: $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = 1; \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$

Bezeichnungen (teilweise schon bekannt)

Ω	Ergebnisraum (bzw. Grundmenge)
$\omega \in \Omega$	Ergebnis (Realisation)
$\mathfrak{P}(\Omega)$	Ereignis algebra, $A \in \mathfrak{P}(\Omega)$, $[A \in \Omega]$ Ereignis
$\mathbb{P}(A)$	W.-keit von A, $\mathbb{P}$ W.-verteilung über Ω (genauer $\mathfrak{P}(\Omega)$)
$(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$	bildet ein mathematisches Modell für ein Zufallsexp.

1.2.2 Elementare Eigenschaften von $\mathbb{P}$

Allein aus den Axiomen 1.2.1 heraus leiten wir ab:

1) $\mathbb{P}(\emptyset) = 0$, denn a)(ii) liefert für alle $A_i = \emptyset : \mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset)$; die Ausnahme $\mathbb{P}(\emptyset) > 0$ liefert den Widerspruch.

2) $\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ für paarweise disjunkte $A_1, \ldots, A_n$ ('Additivität').
Insbesondere: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ für disjunkte A, B, *denn* setze in a)(ii): $A_{n+1} = A_{n+2} = \ldots = \emptyset$ und beachte 1.

3) $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$, denn a)(i): $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A}) \stackrel{2}{=} \mathbb{P}(A) + \mathbb{P}(\overline{A})$

4) $\mathbb{P}(A) \leq \mathbb{P}(B)$ für $A \subset B$ ('Monotonie'), denn $B = B \cap \overline{A} \cup A$ (disj.), also $\mathbb{P}(B) \stackrel{2}{=} \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A}) \geq \mathbb{P}(A)$

5) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$; $[B \setminus A \equiv B \cap \overline{A}]$

6) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
insbesondere: $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$;
allgemein: $\mathbb{P}(A_1, \ldots, A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$

Für die einelementigen (Elementar-) Ereignisse $\{\omega\}$ setzen wir: $p_\omega = \mathbb{P}(\{\omega\})$ und haben

1) $p_\omega \geq 0$

2) $\sum_{w \in \Omega} p_\omega \stackrel{\sigma-Add.}{=} \mathbb{P}(\Omega) = 1$

3) $\sum_{w \in \Omega} p_\omega \stackrel{\sigma-Add.}{=} \mathbb{P}(A)$ für jedes $A \in \Omega$

Wegen 3. ist die Wahrscheinlichkeitsverteilung $\mathbb{P}$ durch die Werte von p_ω eindeutig festgelegt.

1.2.3 Wahrscheinlichkeitsfunktion (Definition)

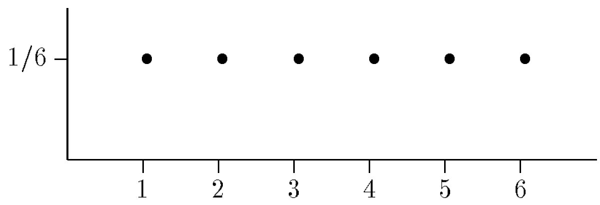
Eine Abb. $p : \Omega \to [0, 1]; \omega \mapsto P_\omega$ mit der Eigenschaft:

$$\sum_{\omega \in \Omega} p_\omega = 1 \quad \text{('Normierung')}$$

heißt Wahrscheinlichkeitsfunktion (oder diskrete W.-keitsdichte, oder Zähldichte).
Jede Wahrscheinlichkeitsfunktion p legt eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ eindeutig fest (wg. 3.: σ-Additivität von $\mathbb{P}$ folgt dann aus dem “Umordnugssatz” für absolut konvergente Reihen). Meistens gibt man eine Wahrscheinlichkeitsverteilung durch eine Wahrscheinlichkeitsfunktion an.

1.2.4 Gleichverteilung (Beispiel)

1 * Werfen mit einem symmetrischen Würfel. Auf $\Omega = \{1, 2, \ldots, 6\}$ definiert man die Wahrscheinlichleitsfunktion $p_1 = \ldots = p_6 = \frac{1}{6}$ ("Gleichverteilung" über $\{1, \ldots, 6\}$).



Ist $A = \{1, 3, 5\}$ das Ereignis "ungerade Augenzahl", so gilt $\mathbb{P}(A) = p_1 + p_3 + p_5 = \frac{3}{6} = \frac{1}{2}$.

1.2.5 Poissonverteilung (Beispiel)

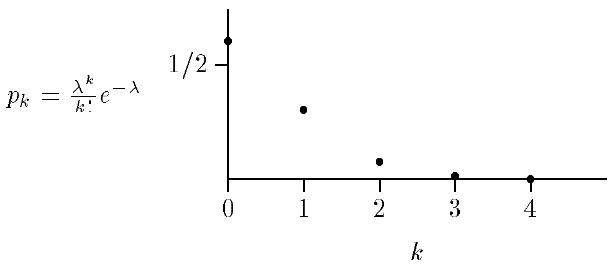
$\Omega = \{0, 1, 2, \ldots\} \equiv \mathbb{Z}_+$ (vgl. Beispiel 4 in 1.1.1). Für ein $\lambda > 0$ definiert man die Wahrscheinlichkeitsfunktion

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

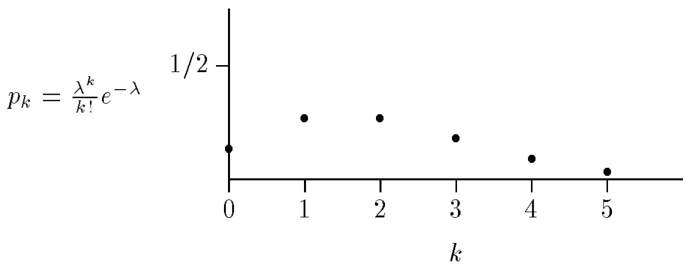
("Poissonverteilung mit Parameter λ", oder kurz P-λ-Verteilung).
Man prüft die Normierung nach:

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Poissonverteilung: $P(\lambda)$ für $\lambda = 0.5$



Poissonverteilung: $P(\lambda)$ für $\lambda = 2$



1.2.6 Einpunktverteilung (Beispiel)

Seien Ω und $w_0 \in \Omega$ beliebig. Dann definiert

$$p_{w_0} = 1, \; p_w = 0$$

(für alle $w \neq w_0$, "Einpunktverteilung") die in w_0 konzentrierte Wahrscheinlichkeitsverteilung ε_{w_0}. Es ist

$$\varepsilon_{w_0}(A) = \begin{cases} 0 & \text{für } w_0 \notin A \\ 1 & \text{für } w_0 \in A \end{cases}$$

1.2.7 Geometrische Verteilung (Beispiel)

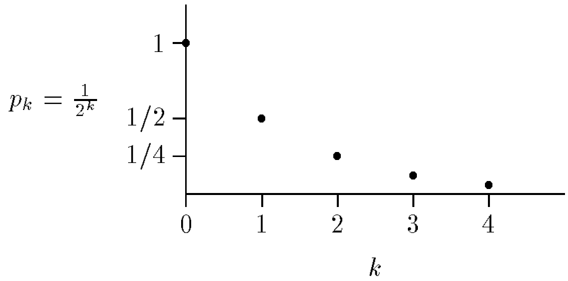
Eine symmetrische Münze wird sooft geworfen, bis *zum ersten Mal* Kopf auftritt. Der Ergebnisraum Ω besteht aus allen möglichen auftretenden Sequenzen:

$$w_1 = K,\ w_2 = ZK,\ w_3 = ZZK,\ \dots\ ,\ w_\infty = ZZZZ\dots$$

(Wir schreiben p_k anstatt p_{w_k}) Es ist $p_1 = \frac{1}{2}$. Heuristische[5] Überlegung zur Bestimmung der $p_k,\ k > 1$:
Wir betrachten $w_k = (\underbrace{Z\dots Z}_{k-1}K$ als Element der Menge $\Omega_k = \{Z, K\}^k$. Es ist[6]: $|\Omega_k| = 2^k$ Alle 2^k Elemente von Ω_k sind gleichwahrscheinlich. Es folgt:

$$p_k = \frac{1}{2^k},\ k \in \mathbb{N} \qquad \text{('Geometrische Verteilung')}$$

Geometrische Verteilung mit Parameter $\frac{1}{2}$:



Normierung:

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1$$

Man hat also $p_\infty = 0$ zu setzen.

1.2.8 Gleichverteilung (Definition)

Die Gleichverteilung über einer endlichen Grundmenge Ω wird bestimmt durch die Wahrscheinlichkeitsfunktion

$$p_\omega = \frac{1}{|\Omega|} \quad (\text{für alle } \omega \in \Omega) \tag{1.1}$$

Die Wahrscheinlichkeit für ein Ereignis $A \subset \Omega$ ist dann:

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega = \frac{|A|}{|\Omega|} \tag{1.2}$$

Die Gleichverteilung wird auch *Laplaceverteilung* genannt. Ein Zufallsexperiment, dem die Gleichverteilung zugrunde liegt, wird auch Laplaceexperiment genannt.
Interpretation der Formel (1.2):
Sei $A \subset \Omega$ gegeben. Ein Ereignis $\omega \in A$ heißt "günstig" für A. Jedes $\omega \in \Omega$ ist "möglich". $\frac{|A|}{|\Omega|}$ ist dann die Anzahl der günstigen durch die Anzahl der möglichen Ergebnisse.

[5] siehe 2.3 exakt
[6] $|\Omega|$ ist die Mächtigkeit der Menge Ω, bzw. die Anzahl der Elemente in der Menge Ω

1.3 Laplaceexperimente und Kombinatorik

Die im Folgenden betrachteten Mengen M seien stets nicht leer. Wir benutzen folgende Tatsachen:

1) Ist $|M| = n$, so existiert eine bijektive Abbildung $\varphi : M \to \{1, \dots, n\}$

2) Für endliche Mengen $M_1, \dots, M_r$ gilt $|M_1 \times \dots \times M_r| = |M_1| \cdot \dots \cdot |M_r|$. (Insbesondere: $|M^k| = |M|^k$)

   Hierbei ist zu beachten, daß $M_1 \times \dots \times M_r$ aus den (geordneten) r-Tupeln $(a_1, \dots, a_r)$ mit $a_1 \in M_1, \dots, a_r \in M_r$ besteht.

1.3.1 Geordnete Stichprobe (Definition)

a) Ein r-Tupel $(a_1, \dots, a_r)$ mit $a_i \in M$ heißt *geordnete* Stichprobe (aus M vom Umfang (od. Länge) r) **mit** Wiederholungen.

b) Ein r-Tupel $(a_1, \dots, a_r)$ mit $a_i \in M,\ a_i \neq a_j$ für $i \neq j$ heißt *geordnete* Stichprobe (aus M vom Umfang (od. Länge) r) **ohne** Wiederholungen.

1.3.2 Anzahl geordneter Stichproben (Satz)

Sei $|M| = n$.

a) Die Anzahl der geordneten Stichproben (aus M vom Umfang r) *mit* Wiederholungen ist

$$\tilde{P}_n^r = n^r$$

b) Die Anzahl der geordneten Stichproben (aus M vom Umfang r) *ohne* Wiederholungen ist

$$P_n^r = n \cdot (n-1) \cdot \dots \cdot (n-r+1) \equiv (n)_r$$

Beweis:

a) Es ist $\tilde{P}_n^r = \underbrace{|M \times \dots \times M|}_{r-mal} = |M|^r = n^r$

b) Heuristisch[7]: An die erste Stelle des r-Tupels $(a_1, \dots, a_r)$ können wir n verschiedene Elemente setzen. für die 2. Stelle bestehen dann $n-1$ Möglichkeiten, etc.(Formal: Krikeberg/Ziezold, S.10/11)

Bemerkung: Statt "geordnete" Stichproben sagt man auch: Stichproben in (unter Berücksichigung der) *Reihenfolge.* Im Spezialfall $r = n$ nennt man eine geordnete Stichprobe aus M *ohne* Wiederholung eine *Permutation* der Elemente von M:

1.3.3 Anzahl der Permutationen (Corollar)

Die Anzahl der Permutationen der Menge M, $|M| = n$, ist $P_n^n = n!$ $[0! = 1]$

1.3.4 Geordnete Stichprobe (Beispiele)

a) Toto: Für $r = 11$ Spiele sind $n = 3$ Ausgänge zu tippen: 1=Sieg Heim, 2=Sieg Gast, 3=Unentschieden.

   Es gibt $\tilde{P}_3^{11} = 177147$ Möglichkeiten, den Tippschein auszufüllen.

b) 3-Buchstabige Wörter (ohne Buchstabenwiederholung). Mit $n = 26$ Buchstaben beträgt die Anzahl $\tilde{P}_{26}^3 = 26 \cdot 25 \cdot 24 = 15600$

[7] Formal: siehe Krikeberg/Ziezold, S.10/11

1.3.5 Ungeordnete Stichprobe (Definition)

Eine Teilmenge $\{a_1, \ldots, a_r\} \subset M$ mit r (paarweise verschiedenen) Elementen heißt *ungeordnete* Stichprobe (oder Stichprobe ohne Berücksichtigung der Reihenfolge). aus M vom Umfang r *ohne* Wiederholung. Ihre Anzahl werde mit C_n^r bezeichnet ($n = |M|$, $0 \le r \le n$).

1.3.6 Binomialkoeffizient (Satz)

Für $n \in \mathbb{N}$ und $0 \le r \le n$ ist

$$C_n^r = \frac{n!}{(n-r)!r!} \equiv \binom{n}{r} \qquad \text{(“Binomialkoeffizient”)}$$

Beweis: Heuristisch[8]: Die Anzahl der geordneten Stichproben ist $P_n^r = (n)_r = n \cdot (n-1) \cdot \ldots \cdot (n-r+1)$. Jeweils $r!$ von ihnen unterscheiden sich nur durch die Reihenfolge. Das ergibt $C_n^r = \frac{(n)_r}{r!}$ ✓

Bemerkung: Es ist nützlich, $\binom{n}{r} = 0$ für $r < 0$ und $r > n$ zu setzen.

1.3.7 Lotto “6 aus 49” (Beispiel)

Die Anzahl der Möglichkeiten (Kombinationen), aus $n = 49$ Zahlen $r = 6$ auszuwählen beträgt:

$$C_6^{49} = \binom{49}{6} = 13983816$$

Da jede dieser Möglichkeiten gleich wahrscheinlich ist, beträgt die Wahrscheinlichkeit, mit 1 Tippreihe bei 1 Ziehung '6 Richtige' zu spielen:

$$W = \frac{1}{13983816}$$

1.3.8 Ungeordnete Stichprobe (Beispiel)

n-maliges Werfen einer symmetrischen Münze. Für $k = 0, \ldots, n$ bezeichne A_k das Ereignis, daß genau k-mal Kopf auftritt und $p_k = \mathbb{P}(A_k)$ die Wahrscheinlichkeit von A_k.
Zur Berechnung der p_k setze $\Omega = \{K, Z\}^n$, $A_k \subset \Omega$ die Menge aller n-Tupel ω, die genaue k-mal K enthält.
Es ist $|A_k| = |S|$, wobei S die Menge aller ungeordneten Stichproben $\{a_1, \ldots, a_k\}$ aus $\{1, \ldots, n\}$ vom Umfang k ohne Wiederholung ist. (In der Tat, $a_1, \ldots, a_k$ bezeichnen die Stellen (Positionen) im n-Tupel ω, an denen k steht. Bspw.: $n = 6$, $k = 3$, $\omega = (K, K, Z, K, Z, Z) \in A_3$, $\{a_1, a_2, a_3\} = \{1, 2, 4\} \in S$).
Also ist nach Satz 1.3.6 $A_k = C_n^k$, während $|\Omega| = 2^n$. Da wir alle $\omega \in \Omega$ als gleichwertig ansehen können (Laplaceexperiment, zur Gleichverteilung siehe 2.3), gilt:

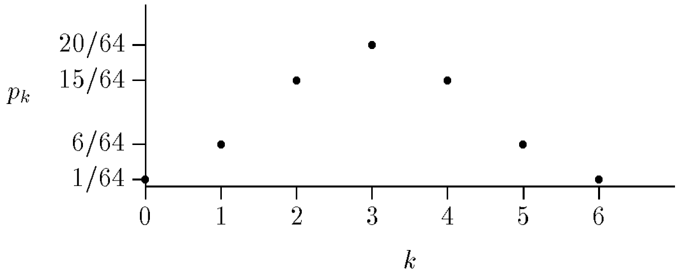
$$p_k = \frac{|A_k|}{|\Omega|} = \binom{n}{k} \cdot 2^{-n}, \; k = 0, \ldots, n$$

Die Zahlen p_k bilden eine Wahrscheinlichkeitsfunktion über $\{0, 1, \ldots, n\}$, denn

$$\sum_{k=0}^{n} p_k = 2^{-k} \sum_{n=0}^{k} \binom{n}{k} = 2^{-n} \cdot 2^n = 1 \qquad \text{(Binomischer Lehrsatz)}$$

[8] Formal: siehe Krikeberg/Ziezold, §II; Forster, Analysis 1

Sie definiert die *Binomialverteilung* mit Parameter[9] $\frac{1}{2}$, $n = 6$:



1.4 Urnenmodelle, Hypergeometrische Verteilung

Die kombinatorischen Begriffe aus 1.3 lassen sich auch über Urnenmodelle beschreiben.
In einer Urne befinden sich n Kugeln, die mit $1, \ldots, n$ numeriert sind. Eine geordnete Stichprobe $(a_1, \ldots, a_r)$ mit [ohne] Wiederholung ensteht durch r-maliges Ziehen mit [ohne] Zurücklegen der Kugeln in die Urne, wobei die Reihenfolge beachtet wird.

Überblick: Stichproben vom Umfang r aus der Menge $\{1, \ldots, n\}$

	mit Wiederholung	ohne Wiederholung
geordnet (in Reihenfolge)	$\tilde{P}_n^r = n^r$	$P_n^r = (n)_r$
ungeordnet (ohne Berücks. d. Reihenf.)	$\tilde{C}_n^r = \binom{n+r-1}{r}$ (Beweis: siehe Krengel, S. 9)	$C_n^r = \binom{n}{r}$

Mit Hilfe des Urnenmodells mit Kugeln von *zwei verschiedenen* Farben führt man eine weitere Wahrscheinlichkeitsverteilung ein.

Modell: In einer Urne mit $N = S + W$ Kugeln seien S schwarze und W weiße Kugeln. Es werden $n \leq N$ Kugeln *ohne* Zurücklegen gezogen. Bezeichne k die Anzahl der schwarzen unter den n gezogenen $0 \leq k \leq n)$, A_k das Ereignis, daß die Anzahl der schwarzen genau gleich k ist, und $h(k; n, N, S)$ die Wahrscheinlichkeit für das Ereignis A_k. Die Wahrscheinlichkeitsfunktion $h(k; n, N, S)$, $k = 0, \ldots, n$ definiert die sogenannte *hypergeometrische Verteilung* mit Parametern n, N, S auf $\Omega' = \{0, 1, \ldots, n\}$

1.4.1 Satz

Für $n \leq N$ und $S \leq N$ gilt:

$$h(k; n, N, S) = \frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}}; \quad k = 0, \ldots, n$$

Beweis: Sei Ω die Menge aller ungeordneten Stichproben $\{a_1, \ldots, a_n\}$ aus der Menge $\{1, \ldots, N\}$ ohne Wiederholung, $\mathbb{P}$ sei die Gleichverteilung auf Ω. Es ist $|\Omega| = C_N^n = \binom{N}{n}$. Die S schwarzen Kugeln seien mit den Nummern $1, \ldots, S$ versehen, die $N - S = W$ weißen Kugeln mit $S + 1, \ldots, N$.

Das Ereignis $A_k \subset \Omega$ besteht dann aus allen $\{a_1, \ldots, a_n\}$ mit genau k Elementen a_i aus $\{1, \ldots, S\}$.

Heuristisch[10]: Es gibt $\binom{S}{k}$ Möglichkeiten, k Kugeln aus S schwarzen [$\binom{N-S}{n-k}$ Möglichkeiten, $n - k$ Kugeln aus $N - S$ weißen] ohne Zurücklegen zu ziehen.

[9] Binomialverteilungen mit bel. Parameter p, $0 < p < 1$: siehe 2.3
[10] Formal: Krikeberg/Ziezold, S. 13

Jede Kombination einer der $\binom{S}{k}$ Möglichkeiten mit einer der $\binom{N-S}{n-k}$ Möglichkeiten ergibt genau ein Element aus A_k. Folglich $|A_k| = \binom{S}{k}\binom{N-S}{n-k}$ und $\mathbb{P}(A_n) = \frac{|A_k|}{|\Omega|}$ wie behauptet. ✓

Bemerkung: Im Spezialfall $n = k = S$ (d.h. Anzahl der Züge gleich Anzahl der schwarzen Kugeln; alle schwarzen Kugeln werden gezogen) ist

$$h = (S; S, N, S) = \frac{1}{\binom{N}{S}}$$

Für alle $n = k = S = 6$, $N = 49$ (Siehe Beispiel 1.3.7).

1.4.2 Qualitätskontrolle (Beispiel)

Aus einer Sendung von $N = 10000$ Glühbirnen, welche S defekte enthalten möge, werden $n = 100$ Stück (ohne Zurücklegen) entnommen und geprüft. Die Wahrscheinlichkeit, daß unter den 100 Proben genau k [höchstens k] defekt sind, ist

$$h(k; 100, 10000, S)$$

oder

$$\sum_{i=0}^{k} h(k; 100, 10000, S)$$

1.4.3 Skatspiel (Beispiel)

Wie groß ist die Wahrscheinlichkeit, daß beim Austeilen der $N = 32$ Karten der Spieler A unter seinen $n = 10$ Karten genau $k = 3$ [$k = 4$] der $S = 4$ Asse besitzt?

$$h(3; 10, 32, 4) = \frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} \approx 0,073 \approx \frac{1}{14}$$

$$\left[h(4; 10, 32, 4) = \frac{\binom{4}{4}\binom{28}{6}}{\binom{32}{10}} \approx 0,0058 \approx \frac{1}{172}\right]$$

1.5 Zufallsgrößen, Zufallsvariablen

Im Beweis zu Satz 1.4.1 traten zwei Wahrscheinlichkeitsräume auf:

1. Ω ist die Menge der ungeordneten Stichproben vom Umfang n aus $\{1, \ldots, N\}$ (ohne Wiederholung), $\mathbb{P}$ ist Gleichverteilung auf Ω
2. $\Omega' = \{0, 1, \ldots, n\}$, $\mathbb{P}'$ hypergeometrische Verteilung [Anzahl $k \in \Omega'$ schwarzer Kugeln hypergeometrisch verteilt]. Wir bezeichnen diese Anzahl mit X. X ist eine Abbildung von $\omega = \{a_1, \ldots, a_n\} \in \Omega : X\Omega \to \Omega'$, $\omega \to X(\omega) = |\omega \cap \{1, \ldots, S\}|$.

Das in Beweis zu 1.4.1 auftretende Ereignis A_k können wir so schreiben:

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

('Urbild der Menge $\{k\}$)

Durch die Definition:

$$\mathbb{P}'(\{k\}) = \mathbb{P}(\{\omega : X(\omega) = k\}) \equiv \mathbb{P}(X^{-1}\{k\}), \quad k \in \Omega \tag{1.3}$$

erhalten wir eine Wahrscheinlichkeitsfunktion auf Ω' (im Beispiel gerade die hypergeometrische Verteilung).
Für beliebige Ereignisse $A' \subset \Omega$ setzt man in Verallgemeinerung von (1.3)

$$\mathbb{P}'(A') = \mathbb{P}\{\omega : X(\omega) \in A'\} = \mathbb{P}(X^{-1}(A')) \tag{1.4}$$

1.5.1 Bild (Definition)

Sei $\mathbb{P}$ eine Wahrscheinlichkeitsfunktion über Ω. Ω' sei eine weitere (nicht-leere) höchstens abzählbare Menge.

a) Eine Abbildung $X : \Omega \to \Omega'$ heißt Zufallsgröße (oder zufällige Größe).

b) Die durch $\mathbb{P}'(A') = \mathbb{P}(X^{-1}(A'))$, $A' \subset \Omega'$, definierte Abbildung von $\mathfrak{P}(\Omega')$ in $[0,1]$ heißt das Bild von $\mathbb{P}$ unter X (vorläufige Bezeichnung!):

$$(\Omega, \mathfrak{P}(\Omega), \mathbb{P}) \underset{X^{-1}}{\overset{X}{\rightleftarrows}} (\Omega', \mathfrak{P}(\Omega'), \mathbb{P})$$

1.5.2 Diskreter Wahrscheinlichkeitsraum (Satz)

Mit dem in 1.5.1 b) definierten $\mathbb{P}'$ gilt: $(\Omega', \mathfrak{P}(\Omega'), \mathbb{P}')$ bildet einen diskreten Wahrscheinlichkeitsraum.

Beweis: Es ist zu zeigen, daß $\mathbb{P}'$ eine Wahrscheinlichkeitsverteilung über Ω' ist. $\mathbb{P}'(A') \in [0,1]$ (triv.)

i) Normierung: $\mathbb{P}'(\Omega') \overset{def.}{=} \mathbb{P}(X^{-1}(\Omega')) = \mathbb{P}(\Omega) = 1$

ii) σ-Additivität: Für eine Folge $A'_1, A'_2, \ldots$ vor paarweise disjunkten Mengen aus Ω' sind die Urbildmengen wieder paarweise disjunkt:

(1.5)
$$X^{-1}(A'_i) \cap X^{-1}(A'_j) = X^{-1}(\underbrace{A'_i \cap A'_j}_{=\emptyset \ f.\ i \neq j}) = \emptyset$$

Ferner gilt:

(1.6)
$$X^{-1}(\bigcup_i A'_i) = \bigcup_i X^{-1}(A'_i)$$

Dann folgt:

$$\mathbb{P}'(\bigcup_{i=1}^{\infty} A'_i) \overset{def.}{=} \mathbb{P}(X^{-1}(\bigcup_{i=1}^{\infty} A'_i) \overset{(1.6)}{=} \mathbb{P}(\bigcup_{i=1}^{\infty} X^{-1}(A'_i)) \overset{(1.5)}{=}$$
$$\overset{\mathbb{P}\ \sigma-Add.}{=} \sum_{i=1}^{\infty} \mathbb{P}(X^{-1}(A'_i)) \overset{def.}{=} \sum_{i=1}^{\infty} \mathbb{P}'(A'_i)$$

1.5.3 Wahrscheinlichkeitsverteilung (Definition)

Die durch $\mathbb{P}(A') = \mathbb{P}(X^{-1}(A'))$, $A' \in \Omega'$ definierte Wahrscheinlichkeitsverteilung $\mathbb{P}'$ auf Ω' heißt Wahrscheinlichkeitsverteilung von X und wird mit $\mathbb{P}_X$ bezeichnet. X heißt auf $\mathbb{P}'$-verteilt.

1.5.4 Bemerkungen und Bezeichnungen

1) $X^{-1}(A') \equiv \{\omega : X(\omega) \in A'\}$ schreibt man auch $\{X \in A'\}$ ("Ereignis 'X in A'"). Entsprechend schreibt man statt $\mathbb{P}_X(A') \overset{def.}{=} \mathbb{P}(X^{-1}(A'))$ auch $\mathbb{P}(X \in A')$ ("Wahrscheinlichkeit, daß X in A' ist").

2) Die zu $\mathbb{P}' \equiv \mathbb{P}_X$ gehörende Wahrscheinlichkeitsfunktion auf Ω' lautet $\omega \to \mathbb{P}_X(\{\omega'\}) = \mathbb{P}(X = \omega')$.

3) Zu jedem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ gibt es die Zufallsgröße $Id : \Omega \to \Omega$, $\omega \mapsto Id(\omega) = \omega$. In diesem Falle ist $\Omega' = \Omega$, $\mathbb{P}' = \mathbb{P}$.

4) Im Fall $\Omega' \subset \mathbb{R}$ [z.B. $\Omega = \mathbb{R}_+$ o.ä.] spricht man auch von einer Zufallsvariablen (statt Zufallsgröße). Beachte, daß diese eine schiefe (aber eingebürgerte) Bezeichnung ist: ω ist die *Variable*, $X : \Omega \to \mathbb{R}$ die *Funktion*. Im Fall $\Omega \subset \mathbb{R}^k$ spricht man auch von einem Zufalls*vektor*.

5) Da Zufallsvariablen reelwertige Funktionen sind, kann man mit ihnen in üblicherweise Rechnen:
$(X+Y)(\omega) = X(\omega) + Y(\omega)$,
$(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$,
$(e^X(\omega) = e^{X(\omega)}$,
$|X|(\omega) = |X(\omega)|$.

1.5.5 Beispiel

n-maliges Würfeln[11]. Auf $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$, $\Omega = \{1, \ldots, 6\}^n = \{(a_1, \ldots, a_n) : a_i \in \{1, \ldots, 6\}$, $\mathbb{P}$ Gleichverteilung auf Ω, definiere $X : \Omega \to \Omega' = \{0, \ldots, n\}$ durch $X(\omega) = |\{i \in \{1, \ldots, n\} : a_i = 6\}|$ ("Anzahl Sechser").

Behauptung: $\mathbb{P}_X(\{k\}) \equiv \mathbb{P}(X - k) = \binom{n}{k} (\frac{1}{6})^k (\frac{5}{6})^{1-k}$, $k = 0, \ldots, n$

Beweis: Siehe 2.3.7, sogar mit $p \in (0,1)$ statt nur $p = \frac{1}{6}$.

1.5.6 Binomialverteilung (Definition)

Die Wahrscheinlichkeitsverteilung auf $\{0, \ldots, n\}$, definiert durch

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

heißt *Binomialverteilung* mit Parametern n und p, kurz $B(n,p)$-Verteilung $(0 \leq p \leq 1)$. Die in 1.5.5 angegebene Zufallsvariable X ist also $B(n, \frac{1}{6})$-verteilt.

1.5.7 Indikatorvariablen (Beispiel)

Sei $A \subset \Omega$. Die Zufallsvariable $1_A : \Omega \to \{0,1\}$,

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

heißt *Indikatiorvariable* von A. Umgekehrt stellt jede $\{0,1\}$-wertige Zufallsvariable X eine Indikatorvariable 1_A dar, nämlich zu

$$A = \{\omega \in \Omega : X(\omega) = 1\} = \{X = 1\}$$

Sei eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf Ω gegeben und setze $p = \mathbb{P}(A)$. Dann ist die Verteilung von $X = 1_A$ gegeben durch $\mathbb{P}(X-1) = p$, $\mathbb{P}(X = 0) = 1-p$. X heißt Bernoulliverteilt mit Parameter p, kurz $B(1,p)$-verteilt.

Rechenregeln für Indikatorvariablen:
$1_{A \cap B} = 1_A \cdot 1_B$
$1_{A \cup B} = 1_A + 1_B - 1_{A \cap B}$
$1_{\overline{A}} = 1 - 1_A$, $(1 = 1_\Omega)$.
Die nun folgenden Erläuterungen beziehen sich auf den Fall von mehreren, aber auf dem gleichen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ definierten Zufallsvariablen.

[11] Vgl. Beispiel 1.3.8

1.5.8 Gemeinsame Verteilung (Definition)

Gegeben k Zufallsvariablen $X_1, \dots, X_k$, $X_i : \Omega \to \Omega'_i \subset \mathbb{R}$.
Sei $X = (X_1, \dots, X_k : \Omega \to \Omega' = \Omega'_1 \times \dots \times \Omega'_k \subset \mathbb{R}^k$ der aus ihnen gebildete Zufallsvektor. Dann heißt die Verteilung $\mathbb{P}_X = \mathbb{P}_{(X_1,\dots,X_k)}$ auch *gemeinsame* Verteilung der $X_1, \dots, X_k$; $\mathbb{P}_{X_i}$ heißt auch i-te Rand- oder Marginalverteilung.
Die Randverteilung $\mathbb{P}_{x_i}$ läßt sich aus der geminesamen Verteilung $\mathbb{P}_X$ wie folgt berechnen. Sei $A'_i \subset \Omega'_i$. Setze

$$A' = \Omega'_1 \times \dots \times \Omega'_{i-1} \times A'_i \times \Omega'_{i+1} \times \dots \times \Omega'_k$$

Dann ist

$$X^{-1}(A') = \{\omega : X_1(\omega) \in \Omega'_1, \dots, X_i(\omega) \in A'_i, \dots$$

$$\dots, X_k(\omega) \in \Omega'_k\} = \{\omega : X_i(\omega) \in A'_i\} = X_i^{-1}(A'_i) \tag{1.7}$$

$$\mathbb{P}_{X_i}(A'_i = \mathbb{P}(X_i^{-1}(A'_i)) \overset{(1.7)}{=} \mathbb{P}(X^{-1}(A')) = \mathbb{P}_X(A') \tag{1.8}$$

1.5.9 Beispiel

Werfen zweier (unterscheidbarer) Würfel. Setze $\Omega = \{1, \dots, 6\}^2$, $\omega = \{j, k\} \in \Omega$, $1 \leq j, k \leq 6$, und definiere die Zufallsvariablen $X_i : \Omega \to \Omega'_i = \{1, \dots, 6\}$, $(i = 1, 2)$ gemäß

$$\left.\begin{array}{l} X_1 = ((j,k)) = j \\ X_2 = ((j,k)) = k \end{array}\right\} \text{die Augenzahl des} \left\{\begin{array}{c} 1. \\ 2. \end{array}\right\} \text{Würfels.}$$

Hier ist $\Omega' = \Omega'_1 \times \Omega'_2 = \Omega$, und dir gemeinsame Verteilung der X_1, X_2 auf Ω' ist

$$\mathbb{P}_{(X_1,X_2)}(\{(j,k)\}) = \frac{1}{|\Omega'|} = \frac{1}{36} \quad \text{für alle } j, k = 1, \dots, 6.$$

Die Randverteilung von X_1 erhalten wir aus (1.8) wie folgt:

$$\mathbb{P}_{X_1}(\{j\}) \overset{(1.8)}{=} \mathbb{P}_{(X_1,X_2)}(\{j\} \times \{1, \dots, 6\}) \overset{disj.}{=} \sum_{k=1}^{6} \mathbb{P}_{(X_1,X_2)}(\{(j,k)\}) = \frac{1}{36} = \frac{1}{6}$$

Kapitel 2

Bedingte Wahrscheinlichkeit

2.1 Bedingte Wahrscheinlichkeit

Vorüberlegung: Gegeben sei ein endlicher Wahrscheinlichkeitsraum Ω, $\mathbb{P}$ Gleichverteilung über Ω. $A, B \subset \Omega$, $A \neq \emptyset$.
Wir nehmen an, daß das Ereignis A eintritt. Welche Definition ist sinnvoll für die Wahrscheinlichkeit von B, unter der Bedingung, daß Ereignis A eingetreten ist?

1. Wenn A eintritt, so kann B dann und nur dann eintreten, wenn $A \cap B$ eintritt.
2. Wir konzentrieren uns auf die Realisation $\omega \in A$ und betrachten sie als gleichwahrscheinlich.

Dann kann die Wahrscheinlichkeit von B unter der Bedingung A wie folgt definiert werden:

$$W \overset{1.,2.}{=} \frac{|A \cap B|}{|A|} = \frac{|A \cap B| \setminus |\Omega|}{|A| \setminus |\Omega|} \overset{Laplace}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Diese Überlegung liegt nahe:

2.1.1 Definitionen

Sei $\mathbb{P}$ Wahrscheinlichkeitsverteilung über Ω[1], $A \subset \Omega$ mit $\mathbb{P}(A) > 0$.

a) Die Abbildung $\mathbb{P}(\cdot|A) : \mathfrak{P}(\Omega) \to [0,1]$, die gemäß $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$, $B \subset \Omega$, definiert ist, heißt *bedingte Wahrscheinlichkeitsverteilung* über Ω unter (der Bedingung) A.

b) Die Zahl $\mathbb{P}(B|A)$ heißt *bedingte Wahrscheinlichkeit* von B unter (der Bedingung) A.

Bemerkungen

1. Man prüft leicht nach, daß $\mathbb{P}(\cdot|A)$ eine Wahrscheinlichkeitsverteilung auf Ω ist, i.S.v. Definitionen 1.2.1.
2. Es gilt $\mathbb{P}(A|A) = 1$ ("$\mathbb{P}(\cdot|A)$ ist auf A konzentriert") und $\mathbb{P}(B|A) = 0$ für $B \subset \overline{A}$ (denn:$\mathbb{P}(B|A) = \frac{\mathbb{P}\overbrace{(A \cup B)}^{=\emptyset}}{\mathbb{P}(A)} = 0$).

[1] $\Omega = (\Omega, \mathfrak{P}(\Omega), \mathbb{P})$, diskreter Wahrscheinlichkeitsraum

2.1.2 Beispiel

Weißer und schwarzer Würfel. Berechne die bedingte Wahrscheinlichkeit, mit dem schwarzen Würfel eine '6' zu würfeln (B), unter der Bedingung, daß die Summe der Augenzahlen '11' beträgt (A).
Setze $\Omega = \{1, \ldots, 6\}^2$, $|\Omega| = 36$, $\mathbb{P}$ Gleichverteilung auf Ω, dann ist $A = \{(5,6), (6,5)\}$ und $B = \{(i,6); i = \{1, \ldots, 6\}; A \cup B = \{5,6)\}$.
Es folgt:

$$\mathbb{P}(B|A) \overset{def.}{=} \frac{|A \cup B| \setminus |\Omega|}{|A| \setminus |\Omega|} = \frac{1}{2}$$

Formt man die Definitionsformeln in 2.1.1 um, zu

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) \tag{2.1}$$

so kann man (2.1) auch im Fall $\mathbb{P}(A) = 0$ einen Sinn geben. (Beide Seiten sind gleichzeitig 0: $\mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0$)
Wir betrachten eine Technik, eine Wahrscheinlichkeit auf Ω durch bedingte Wahrscheinlichkeiten zusammenzusetzen.

Definition $A_1, \ldots, A_m$ heißt eine Zerlegung von Ω, falls $\bigcup_{i=1}^m A_i = \Omega$, $A_i \cap A_j = \emptyset$ für alle $i \neq j$.

2.1.3 Satz

Sei $A_1, \ldots, A_m$ eine Zerlegung von Ω. Für jedes $i = 1, \ldots, m$ sei eineauf A_i konzentrierte Wahrscheinlichkeitsverteilung Q_{A_i} auf Ω gegeben (d.h. $Q_{A_i}(A_i) = 1$), sowie eine Zahl $p_i \in [0,1]$, mit $\sum_{i=1}^m p_i = 1$.

a) Dann existiert genau eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf Ω mit

   i) $\mathbb{P}(A_i) = p_i$

   ii) $\mathbb{P}(B|A_i) = Q_{A_i}(B)$, falls $p_i > 0$ für alle $i = 1, \ldots, m$.

b) Es gilt $\mathbb{P}(B) = \sum_{i=1}^m p_i \cdot Q_{A_i}(B)$ für jedes $B \subset \Omega$.

Beweis: Man definiert $\mathbb{P}$ gemäß Formel b) und rechnet nach[2], daß $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf Ω ist. Wie in Bemerkung zu 2.1.1 gilt für die paarweise disjunkten A_i

$$Q_{A_i}(A_i) = \begin{cases} 0 & \text{falls } i \neq j \\ 1 & \text{falls } i = j \end{cases} \tag{2.2}$$

Daraus folgt sofort i).
Für ein beliebiges $p_i > 0$ und $B \subset \Omega$ gilt

$$\mathbb{P}(B|A_i) \overset{def.}{=} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(A_i)} \overset{b),i)}{=} \frac{1}{p_i} \sum_{j=1}^m p_j Q_{A_j}(A_i \cap B) =$$

$$= \frac{p_i}{p_i} Q_{A_i}(A_i \cap B) + \overbrace{Q_{A_i}(\overline{A_i} \cap B)}^{=0} \overset{disj.}{=} Q_{A_i}(B)$$

also ii).
Zur Eindeutigkeit: Sei $\tilde{\mathbb{P}}$ eine (weitere) Wahrscheinlichkeitsverteilung auf Ω, die i) und ii) erfüllt. Dann gilt für $B \subset \Omega$ (wegen $B = \bigcup_{j=1}^m A_j \cup B$)

$$\tilde{\mathbb{P}}(B) = \tilde{\mathbb{P}}(\bigcup_{j=1}^m A_j \cap B) \overset{disj.}{=} \sum_{j=1}^m \tilde{\mathbb{P}}(A_j \cap B) \overset{1)}{=}$$

$$= \sum_{j=1}^m \tilde{\mathbb{P}}(A_j) \cdot \tilde{\mathbb{P}}(B|A_j) \overset{i),ii)}{=} \sum_{j=1}^m p_j Q_{A_j}(B) \overset{b)}{=} \mathbb{P}(B). \quad \checkmark$$

[2] siehe Übungsaufgabe 3

Bemerkung: Für jedes $B \subset \Omega$ gilt also

$$\mathbb{P}(B) = \sum_{j=1}^{m} \mathbb{P}(A_j)\mathbb{P}(B|A_j) \tag{2.3}$$

("Formel von der totalen Wahrscheinlichkeit")

2.1.4 Formel von Bayes (Satz)

Ist $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf Ω, $A_1, \ldots, A_m$ eine Zerlegung von Ω, so gilt für jedes $B \subset \Omega$ mit $\mathbb{P}(B) > 0$ und $i = 1, \ldots, m$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)}{\sum_{j=1}^{m} \mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}$$

("Formel von Bayes")

Beweis

$$\mathbb{P}(A_i|B) \overset{def.}{=} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \overset{(1),(2)}{=} \frac{\mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)}{\sum_{j=1}^{m} \mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)} \quad \checkmark$$

Bemerkung: Man beachte, daß auf der linken und rechten Seite in 2.1.4 "Argument und Bedingung" vertauscht auftreten. In einer außermathematischen Deutung spielen die A_j die Rolle von (verschiedenen) Ursachen für die Wirkung B.

2.1.5 Beispiel

Test auf eine Krankheit X. p*100% der Bevölkerung leide an der Krankheit X. Ein Test spreche bei k*100% der X-Kranken und bei g*100% der Gesunden positiv an (k = Sensitivität, $1 - g$ = Spezifität des Tests). Mit welcher Wahrscheinlichkeit w_1 [w_2] hat eine zufällig ausgewählte Person die Krankheit X, wenn der Test positiv [negativ] ausfällt?

Modell: Ω Bevölkerungsmenge, $A_1 \subset \Omega$ Teilmenge der X-Kranken, $A_2 = \Omega \setminus A_1$ Teilmenge der Gesunden, $B \subset \Omega$ Teilmenge der Test-Positiven.
$\mathbb{P}$ Gleichverteilung auf Ω: $\mathbb{P}(A_1) = p$, $\mathbb{P}(A_2) = 1 - p$, $\mathbb{P}(B|A_1) = k$, $\mathbb{P}(B|A_2) = g$.

Bayesformel:

$$w_1 = \mathbb{P}(A_1|B) = \frac{kp}{kp + g(1-p)};$$

$$w_2 = \mathbb{P}(A_1|\overline{B}) = \frac{(1-k)p}{(1-k)p + (1-g)(1-p)};$$

Zahlenbeispiel: $p = 0,01$, $k = 0,9$, $g = 0,2$. Dann rechne man:

$$w_1 = \frac{1}{23} \approx 0,044;$$

$$w_2 = \frac{1}{793} \approx 0,0013;$$

Bemerkung: Der folgende Satz verallgemeinert die Formel (2.1) auf n Faktoren zur sogenannten "Produktformel".

2.1.6 Produktformel (Satz)

Sei $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf Ω und seien $A_1, \ldots, A_n \subset \Omega$ mit $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt die sog. *"Produktformel"*:

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \ldots \cdot \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

Beweis: Die Faktoren auf der rechten Seite sind definiert wegen

$$\mathbb{P}(A_1) \geq \mathbb{P}(A_1 \cap A_2) \geq \ldots \geq \mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$$

Es gilt dann

$$\text{rechte Seite} \stackrel{def.}{=} \mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \cdot \ldots$$

$$\ldots \cdot \frac{\mathbb{P}(A_1 \cap \ldots \cap A_n)}{\mathbb{P}(A_1 \cap \ldots \cap A_{n-1})} = \mathbb{P}(A_1 \cap \ldots \cap A_n) \quad \checkmark$$

2.2 Unabhängige Ereignisse

Wir wollen zwei Ereignisse A, B unabhängig nennen, wenn die bedingte Wahrscheinlichkeit $\mathbb{P}(B|A)$ nicht von A abhängt: $\mathbb{P}(B|A) = \mathbb{P}(B)$. Wegen der lästigen Voraussetzung $\mathbb{P}(A) > 0$ und der fehlenden Symmetrie ziehen wir die folgende Definition vor.

2.2.1 Definition

Ist $\mathbb{P}$ Wahrscheinlichkeitsverteilung auf Ω und $A, B \subset \Omega$, dann heißen A und B stochastisch unabhängig, falls gilt:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \tag{2.4}$$

Bemerkung:

1. Ist $\mathbb{P}(A) = 0$ und B beliebig, dann sind A, B unabhängig (beachte, daß $\mathbb{P}(A \cap B) \leq \mathbb{P}(A)$).
2. Gleichung (2.4) ist im Fall $\mathbb{P}(A) > 0$ äquivalent mit

   $$\mathbb{P}(B|A) = \mathbb{P}(B) \tag{2.5}$$

   Definition (2.4) und Definition (2.5) zusammen mit Bemerkung 1 erweisen sich als äquivalent
3. Aus A, B unabängig folgt $\overline{A}, B$ unabhängig[3]

Erweiterung der Definition 2.2.1 auf n Ereignisse.

2.2.2 Definition

Sei $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf Ω und seien $A_1, \ldots, A_n \subset \Omega$. dann heißen $A_1, \ldots, A_n$ *(stochastisch) unabhängig*, falls für jede nichtleere Teilmenge $\{j_1, \ldots, j_k\} \subset \{1, \ldots, n\}$ gilt

$$\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}) = \mathbb{P}(A_{j_1}) \cdot \ldots \cdot \mathbb{P}(A_{j_k}) \tag{2.6}$$

[3] Übungsaufgabe

Bemerkung:

1. Ein Begriff, der in der Stochastik *keine* Rolle spielt, ist der der paarweisen Unabhängigkeit der $A_1, \ldots, A_n$:

   (2.7) $$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j), i \neq j$$

   Aus (2.7) folgt *nicht* notwendig die Unabhängigkeit der $A_1, \ldots, A_n$.

2. Folgende Eigenschaft der $A_1, \ldots, A_n$ erweist sich als äquivalent zu 2.2.2: Für jede echte, nichtleere Teilmenge $\{j_1, \ldots, j_k\} \subset \{1, \ldots, n\}$ mit $\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}) > 0$ und für jedes $i \in \{1, \ldots, n\} \setminus \{j_1, \ldots, j_k\}$ gilt

   (2.8) $$\mathbb{P}(A_i | A_{j_1} \cap \ldots \cap A_{j_k}) = \mathbb{P}(A_i)$$

2.2.3 Stochastische Unabhängigkeit (Satz)

Die Definitionen (2.6) und (2.8) der (stochastischen) Unabhängigkeit der $A_1, \ldots, A_n$ sind äquivalent.

Beweis: Aus (2.6)folgt im Fall $\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}) > 0$ sofort

$$\mathbb{P}(A_i | A_{j_1} \cap \ldots \cap A_{j_k}) \overset{def.}{=} \frac{\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k} \cap A_i)}{\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k})} \overset{(2.7)}{=} \mathbb{P}(A_i) \quad \checkmark$$

Gilt umgekehrt (2.8), so liefert unter der Voraussetzung

(2.9) $$\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_{k-1}}) > 0$$

die Produktformel

$$\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}) \overset{2.1.6}{=} \mathbb{P}(A_{j_1}) \cdot \mathbb{P}(A_{j_2} | A_{j_1}) \cdot \ldots \cdot \mathbb{P}(A_{j_k} | A_{j_1} \cap \ldots \cap A_{j_{k-1}}) \overset{(2.8)}{=}$$
$$= \mathbb{P}(A_{j_1}) \cdot \mathbb{P}(A_{j_2}) \cdot \ldots \cdot \mathbb{P}(A_{j_k}) \quad \text{d.h. (2.6)}$$

Gilt (2.9) nicht, so zeigt man, daß beide Seiten von (2.6) $\emptyset$ sind[4].

2.2.4 Beispiel

Weißer und schwarzer Würfel. Gegeben $A', B' \subset \{1, \ldots, 6\}$. Die Grundmenge $\Omega = \{1, \ldots, 6\}^2$ enthält die Teilmenge $A = A' \times \{1, \ldots, 6\}$ ($B = \{1, \ldots, 6\} \times B'$). Die Augenzahl des ersten (zweiten) Würfels liegt in A' (B'). $\mathbb{P}$ sei die Gleichverteilung auf Ω. Dann sind A, B unabhängige Ereignisse. In der Tat, wegen $A \cap B = A' \times B'$ ist

$$\mathbb{P}(A \cap B) = \mathbb{P}(A' \times B') \overset{Lapl.}{=} \frac{|A' \times B'|}{|\Omega|} = \frac{|A'| \cdot |B'|}{6 \cdot 6} =$$
$$= \frac{|A'| \cdot 6}{36} \cdot \frac{6 \cdot |B'|}{36} = \frac{|A|}{|\Omega|} \cdot \frac{|B|}{|\Omega|} \overset{Lapl.}{=} \mathbb{P}(A) \cdot \mathbb{P}(B)$$

2.3 Produktexperimente, unabh. Zufallsgrößen

Erweiterung des Begriffs "Unabhängigkeit" für Ereignisse auf

- Zufallsexperimente (Wahrscheinlichkeitsraum)
- Zufallsgrößen

Notation: *Für den Rest des Kapitels schreiben wir statt* $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ *nur noch:* $(\Omega, \mathbb{P})$.

[4] Krikeberg & Ziezold; S.48

Rückblick: Beispiel 2.2.4:
Seien $\begin{matrix}(\Omega_1,\mathbb{P}_1)\\(\Omega_2,\mathbb{P}_2)\end{matrix}$ Wahrscheinlichkeitsräume des $\begin{matrix}\text{“ersten Würfels”}\\\text{“zweiten Würfels”}\end{matrix}$, $\Omega_i = \{1,\dots,6\}$, $\mathbb{P}_i$ Gleichverteilung auf Ω_i, $(i = 1,2)$, $(\Omega,\mathbb{P})$ Wahrscheinlichkeitsraum “zweier Würfel”, $\Omega = \{1,\dots,6\}^2$, $\mathbb{P}$ Gleichverteilung auf Ω.
Es wurde gezeigt, daß

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \tag{2.10}$$

für *alle* $A_1 \subset \Omega_1$, $A_2 \subset \Omega_2$.
Dies gibt Anlaß zur folgenden Definition.

2.3.1 Produkt der Wahrscheinlichkeitsräume (Definition)

Gegeben sind (diskrete) Wahrscheinlichkeitsräume $(\Omega_1,\mathbb{P}_1)\dots(\Omega_n,\mathbb{P}_n)$. Der Wahrscheinlichkeitsraum $(\Omega,\mathbb{P})$ mit: $\Omega = \Omega_1 \times \dots \times \Omega_n$, $\mathbb{P}$ definiert auf $\mathfrak{P}(\Omega)$ mit

$$\mathbb{P}(A_1 \times \dots \times A_n) = \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_n) \tag{2.11}$$

für alle $A_1 \subset \Omega_1, \dots, A_n \subset \Omega_n$ heißt Produkt der Wahrscheinlichkeitsräume $(\Omega_i,\mathbb{P}_i)$, $i = 1,\dots,n$

2.3.2 Satz

Sind $(\Omega_1,\mathbb{P}_1),\dots,(\Omega_n,\mathbb{P}_n)$ Wahrscheinlichkeitsräume, dann existiert genau eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\Omega = \Omega_1 \times \dots \times \Omega_n$, welche (2.11) erfüllt.

Beweis:

i) Eindeutigkeit:
Jede Wahrscheinlichkeitsverteilung $\mathbb{P}$, welche (2.11) erfüllt, besitzt ein und dieselbe Wahrscheinlichkeitsfunktion $q(\omega)$, $\omega = (\omega_1,\dots,\omega_n) \in \Omega$, nämlich

$$q(\omega) \overset{def.}{=} \mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega_1\} \times \dots \times \{\omega_n\}) \overset{(2.11)}{=} \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\}) \tag{2.12}$$

ii) Existenz:
Man definiere die Abbildung $q : \Omega \to [0,1]$ gemäß (2.12), d.h. $q(\omega) = \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \mathbb{P}_n(\{\omega_n\})$.
q ist normiert (!), also eine Wahrscheinlichkeitsfunktion auf Ω, die eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf Ω definiert. $\mathbb{P}$ erfüllt (2.11), denn

$$\mathbb{P} = (A_1 \times \dots \times A_n) = \sum_{\omega_1 \in A_1,\dots,\omega_n \in A_n} q((\omega_1,\dots,\omega_n)) \overset{(2.12)}{=}$$

$$= \sum_{\omega_1 \in A_1} \mathbb{P}_1(\{\omega_1\}) \cdot \dots \cdot \sum_{\omega_n \in A_n} \mathbb{P}_n(\{\omega_n\}) = \mathbb{P}_1(A_1) \cdot \dots \cdot \mathbb{P}_n(A_n) \quad \checkmark$$

2.3.3 Produktverteilung (Definition)

Die Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\Omega_1 \times \dots \times \Omega_n = \Omega$, welche (2.11) erfüllt, heißt *Produktverteilung*, genauer Produkt $\mathbb{P} = \mathbb{P}_1 \times \dots \times \mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$, der $\mathbb{P}_1,\dots,\mathbb{P}_n$.

Notation: $\otimes$ und $\times$ hier gleichbedeutend.

Bemerkung

1. Die einzelnen Faktoren $\mathbb{P}_1, \dots, \mathbb{P}_n$ der Produktverteilung $\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$ (auch Randverteilungen genannt) lassen sich (wie folgt, aus $\mathbb{P}$) zurückgewinnen:
Ist nämlich $A_i \in \Omega_i$, $(i \in \{1, \dots, n\})$, so setzt man $A = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n$ und hat

$$\mathbb{P}(A) \overset{(2.11)}{=} 1 \cdot \ldots \cdot 1 \cdot \mathbb{P}_i(A_i) \cdot 1 \cdot \ldots \cdot 1 = \mathbb{P}_i(A_i) \tag{2.13}$$

2. Spezialfall: $\Omega_1, \dots, \Omega_n$ endlich
$\mathbb{P} = \bigotimes_{i=1}^n \mathbb{P}_i$ ist genau dann Gleichverteilung auf $\Omega = \Omega_1 \times \ldots \times \Omega_n$, wenn jedes $\mathbb{P}_i$ $(i = 1, \dots, n)$ Gleichverteilung auf Ω_i ist.
In der Tat:

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega_1| \cdot \ldots \cdot |\Omega_n|} \quad \text{für alle} \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega$$

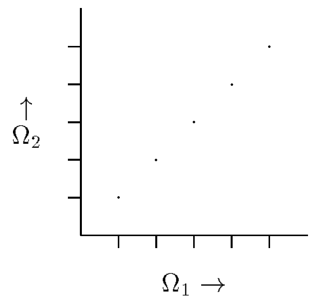
$$(2.11) \Uparrow \quad \Downarrow (2.13)$$

$$\mathbb{P}_i(\{\omega_i\}) = \frac{1}{|\Omega_i|} \quad \text{für alle} \quad i = 1, \dots, n, \quad \omega_i \in \Omega$$

3. Ist $\Omega_1 = \ldots \Omega_n$, $\mathbb{P}_1 = \ldots = \mathbb{P}_n$, so bildet Ω_1^n, $\bigotimes_{i=1}^n \mathbb{P}_i$ ein Modell für die n-fache *unabhängige* Wiederholung des Zufallsexperimentes $(\Omega_1, \mathbb{P}_1)$. (sog. "Produktexperiment")

4. Es gibt viele Möglichkeiten, eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\Omega_1 \times \ldots \times \Omega_n$ *anders* zu definieren, als durch $\bigotimes_{i=1}^n \mathbb{P}_i$, so daß immernoch (2.13) erfüllt ist.
Beispiel:
$\Omega_1 = \ldots \Omega_n$, $\mathbb{P}_1 = \ldots = \mathbb{P}_n$; definierte Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf Ω_1^n durch die Wahrscheinlichkeitsfunktion

$$\mathbb{P}((\omega_1, \dots, \omega_n)) = \begin{cases} \mathbb{P}_1(\{\omega_1\}) & \text{falls } \omega_1 = \ldots = \omega_n \\ 0 & \text{sonst} \end{cases} \tag{2.14}$$



$\mathbb{P}$ erfüllt (2.13), jedoch nicht (2.13) ($\Rightarrow$ ist also *nicht* die Produktwahrscheinlichkeit).
In der Tat, für ω_1 mit $0 < \mathbb{P}_1(\{\omega_1\}) < 1$ gilt

$$\mathbb{P}(\underbrace{\{\omega_1\} \times \ldots \times \{\omega_1\}}_{\text{n-mal}}) = \mathbb{P}(\{(\omega_1, \dots, \omega_1)\}) =$$

$$\overset{(2.14)}{=} \mathbb{P}_1(\{\omega_1\}) \neq [\mathbb{P}_1(\{\omega_1\})]^n$$

Der Wahrscheinlichkeitsraum $(\Omega_1^n, \mathbb{P})$, $\mathbb{P}$ wie in (2.14), steht für das Zufallsexperiment: "n identische Wiederholungen", das extreme Gegenteil von "n unabhängige Wiederholungen".

2.3.4 Defintion

Auf dem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ seien Zufallsgrößen $X_1, \ldots, X_n$ definiert, $X_i : \Omega \to \Omega'$, $i = 1, \ldots, n$.
$X_1, \ldots, X_n$ heißen (stochastisch) unabhängig, falls

$$\mathbb{P}(X_1^{-1}(B_1) \cap \ldots \cap X_n^{-1}(B_n)) = \mathbb{P}(X_1^{-1}(B_1)) \cdot \ldots \cdot \mathbb{P}(X_n^{-1}(B_n)) \tag{2.15}$$

für alle $B_i \subset \Omega_i$, $i = 1, \ldots, n$.
$X_1, \ldots, X_n$

Bemerkungen

1. Für unabh. $X_1, \ldots, X_n$ sind die Ereignisse $X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n)$ unabhängig (i.S.v. Definition 2.2.2).
   In der Tat, ist $J = \{j_1, \ldots, j_k\} \subset \{1, \ldots, n\}$ nicht leer, dann lautet (2.15) mit $B_j = \Omega'_j$ für $j \notin J$ wegen $X_j^{-1}(\Omega'_j) = \Omega$:
   $$\mathbb{P}(X_{j_1}^{-1}(B_{j_1}) \cap \ldots \cap X_{j_k}^{-1}(B_{j_k})) = \mathbb{P}(X_{j_1}^{-1}(B_{j_1})) \cdot \ldots \cdot \mathbb{P}(X_{j_k}^{-1}(B_{j_k}))$$

2. Gleichung (2.15) läßt sich auch auf folgende Weisen schreiben:
   - $\mathbb{P}((X_1, \ldots, X_n) \in B_1 \times \ldots \times B_n) = \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_n \in B_n)$ [5]
   - $\mathbb{P}_{(X_1, \ldots, X_n)}(B_1 \times \ldots \times B_n) = \mathbb{P}_{X_1}(B_1) \cdot \ldots \cdot \mathbb{P}_{X_n}(B_n)$ jeweils für alle $B_i \subset \Omega'_i$, $i = 1, \ldots, n$

Die letzte Gleichung führt unter Berücksichtigung von Definition 2.3.3 zu folgendem Satz.

2.3.5 Satz

Die Zufallsgrößen $X_1, \ldots, X_n$, $X_i : \Omega \to \Omega'_i$, $i = 1, \ldots, n$, sind genau dann unabhängig, falls

$$\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{X_1} \times \ldots \times \mathbb{P}_{X_n} \tag{2.16}$$

(In Worten: falls die gemeinsame Verteilung der $X_1, \ldots, X_n$ gleich dem Produkt der Randverteilungen ist.)

2.3.6 Satz

Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $\varphi_1, \ldots, \varphi_n$ Funktionen von $\mathbb{R}$ in $\mathbb{R}$, dann sind die Zufallsvariablen $\varphi_1 \circ X_1, \ldots, \varphi_n \circ X_n$ ebenfalls unabhängig.

Beweis: Wegen $(\varphi_i \circ X_i)^{-1}(B_i) = X_i^{-1}(\varphi_i^{-1}(B_i))$, $B_i \subset \mathbb{R}$ gilt

$$\begin{aligned}
&\mathbb{P}((\varphi_1 \circ X_1)^{-1}(B_1) \cap \ldots \cap (\varphi_n \circ X_n)^{-1}(B_n)) = \\
&= \mathbb{P}(X_1^{-1}(\varphi_1^{-1}(B_1)) \cap \ldots \cap X_n^{-1}(\varphi_n^{-1}(B_n))) = \\
&\overset{Unabh.,(2.15)}{=} \mathbb{P}(X_1^{-1}(\varphi_1^{-1}(B_1))) \cdot \ldots \cdot \mathbb{P}(X_n^{-1}(\varphi_n^{-1}(B_n))) = \\
&= \mathbb{P}((\varphi_1 \circ X_1)^{-1}(B_1)) \cdot \ldots \cdot \mathbb{P}((\varphi_n \circ X_n)^{-1}(B_n)) \quad \checkmark
\end{aligned}$$

Sind also die beiden Zufallsvariablen X, Y unabhängig, dann auch die Zufallsvariablen $|X|, Y^2$ und auch die Zufallsvariablen $e^X, sin(Y)$, **nicht** aber die Zufallsvariablen $X + Y, X - Y$.

[5] Ausführlich: $\mathbb{P}(\{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in \ldots\})$

2.3.7 Binomialverteilung

Sei $(\Omega_1, \mathbb{P}_1)$ Wahrscheinlichkeitsraum. Wir fixieren ein Ereignis $A \subset \Omega_1$ und setzen $p = \mathbb{P}(A)$ ("Erfolgswahrscheinlichkeit").
Mit $b(k; n, p)$ bezeichnen wir die Wahrscheinlichkeit, daß bei n-maliger unabhängiger Wiederholung des Zufallsexperimentes $(\Omega_1, \mathbb{P}_1)$ genau k-mal das Ereignis A auftritt. Behauptung:

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}; \quad k = 0, \dots, n \tag{2.17}$$

Beweis: Auf dem Produktraum $(\Omega, \mathbb{P})$, $\Omega = \Omega_1^n$, $\mathbb{P} = \underbrace{\mathbb{P}_1 \times \ldots \times \mathbb{P}_1}_{n-mal}$, definieren wir die Zufallsvariablen $X_1, \dots, X_n$ gemäß

$$X_i((\omega_1, \dots, \omega_n)) = \begin{cases} 1 & \text{falls } \omega_i \in A \\ 0 & \text{sonst} \end{cases}$$

("Ereignis tritt bei i-ter Wiederholung ein/nicht ein").
X_i hat die sog. Bernoulliverteilung: $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$.
Die Zufallsvariablen $X_1, \dots, X_n$ sind unabhängig (Sätze 2.3.5, 2.3.6, (!)). Setze

$$X = X_1 + \ldots + X_n$$

("Häufigkeit des Auftretens von A")

dann ist $\mathbb{P}(X = k) = b(k; n, p)$.

Zur Berechnung von $b(k; n, p)$:
Für jede der $\binom{n}{k}$ Zerlegungen $\{1, \dots, n\} = \{i_1, \dots, i_k\} \cup \{j_1, \dots, j_{n-k}\}$ (disj.) gilt:

$$\mathbb{P}(X_{i_1} = 1, \dots, X_{i_k} = 1, X_{j_1} = 0, \dots, X_{j_{n-k}} = 0) =$$

$$\stackrel{unabh.}{=} \mathbb{P}(X_{i_1} = 1) \cdot \ldots \cdot \mathbb{P}(X_{j_{n-k}} = 0) = p^k (1-p)^{n-k}$$

also: $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ ✓

2.3.8 Binomialverteilung (Defintion)

Die durch (2.17) auf $\{0, \dots, n\}$ def. Wahrscheinlichkeitsverteilung heißt *Binomialverteilung* mir Parametern n und p, kurz $B(n,p)$-Verteilung, $n \in \mathbb{N}$, $0 \leq p \leq 1$ (vgl. Beispiele 1.3.8, 1.5.5).
Die $B(1,p)$-Verteilung heißt auch *Bernoulliverteilung*, das oben beschriebene Produktexperiment heißt auch *Bernouliiexperiment*. Man sagt: Die (oben definierte) Zufallsvariable X ist $B(n,p)$-verteilt.
Für unabhängige Zufallsvariablen $X_1, \dots, X_n$ läßt sich die Verteilung der $\mathbb{P}_X$ der Summe $X = X_1 + \ldots + X_n$ explizit auf den Randverteilungen berechnen.

2.3.9 Satz

Sind $X_1, \dots, X_n$ *unabhängige* Zufallsvariablen auf $(\Omega, \mathbb{P})$, so gilt für die Verteilung $\mathbb{P}_X$ von $X = X_1 + \ldots + X_n$:

$$\mathbb{P}_X(\{x\}) = \sum_{\substack{(X_1, \dots, X_n) \in X_1(\Omega) \times \dots \times X_n(\Omega) \\ X_1 + \dots + X_n = X}} \mathbb{P}_{X_1}(\{x_1\}) \cdot \ldots \cdot \mathbb{P}_{X_n}(\{x_n\}) \tag{2.18}$$

für alle $x \in X(\Omega)$.

Beweis: Zerlegung von $\{X = x\}$ in paarweise disj. Mengen $A_{(X_1, \dots, X_n)} = \langle X_1 = x_1, \dots, X_n = x_n \rangle$, $X_1 + \ldots + X_n = X$. Im Fall $n = 2$ lautet (2.18)für $x \in (X_1 + X_1)(\Omega)$

$$\mathbb{P}_{X_1 + X_2}(\{x\}) = \sum_{x_1 \in X_1(\Omega)} \mathbb{P}_{X_1}(\{x_1\}) \cdot \mathbb{P}_{X_2}(\{x - x_1\})$$

2.3.10 Faltung (Definition)

Die in (2.18) angegebene Verteilung von $X = X_1 + \ldots + X_n$ heißt *Faltung* der Verteilungen $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_n}$.

Notation: $\mathbb{P}_X = \mathbb{P}_{X_1} * \ldots * \mathbb{P}_{X_n}$

Beispiel $B(n,p)$-Verteilung ist die Faltung von n $B(1,p)$-Verteilungen:

$$B(n,p) = \underbrace{B(1,p) * \ldots * B(1,p)}_{\text{n-mal}}$$

Kapitel 3

Momente

3.1 Erwartungswert

Vorbemerkung: Zwei Personen A und B vereinbaren ein Würfelspiel:

- Ausgang '1': A zahlt an B 5DM
- Ausgang '2',...,'6': B zahlt an A 1DM

die Gewinnerwartung für beide Spieler beträgt hier Null: Erwarteter Gewinn von B:

$$5 \cdot \underbrace{\frac{1}{6}}_{\mathbb{P}(\{1\})} - 1 \cdot \underbrace{\frac{1}{6}}_{\mathbb{P}(\{2\})} - \ldots - 1 \cdot \underbrace{\frac{1}{6}}_{\mathbb{P}(\{2\})} = 0$$

("faires Spiel")

3.1.1 Erwartungswert (Definiton)

Sei X eine auf dem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ definierte Zufallsvariable, dann heißt

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}_X(\{x\}) \tag{3.1}$$

Erwartungswert von X, vorausgesetzt, daß $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty$
Falls $\sum_x |x| \cdot \mathbb{P}(X = x) < \infty$, dann ist die Reihe in (3.1) absolut konvergent (Die Fälle $\mathbb{E}(x) = \pm\infty$ und $\mathbb{E}(x) = \infty - \infty$ werden also ausgeschlossen).

3.1.2 Satz

Bezeichnet $p_\omega = \mathbb{P}(\{\omega\})$ die Wahrscheinlichkeitsfunktion von $\mathbb{P}$ auf Ω, so gilt für eine auf $(\Omega, \mathbb{P})$ definierte Zufallsvariable X:

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) p_\omega \tag{3.2}$$

(sofern $\sum |X(\omega)| p_\omega < \infty$)

Beweis: Es ist

$$\sum_{\omega \in \Omega} X(\omega) p_\omega = \sum_{x \in X(\Omega)} \sum_{\omega : X(\omega) = x} X(\omega) p_\omega =$$
$$= \sum_{x \in X(\Omega)} x \sum_{\omega : X(\omega) = x} p_\omega = \sum_{x \in X(\Omega)} x \mathbb{P}_X(\{\omega\}) = (3.1) \quad \checkmark$$

[Großer Umordnungssatz für konvergente Reihen]

Bemerkung

1. Im Fall der absoluten Konvergenz der Reihen in (3.1) oder (3.2) sagt man:
   - X besitzt einen Erwartungswert, oder
   - der Erwartungswert von X existiert.

   Dies ist bei endlichen Ω immer der Fall.
2. Schreibweise: Statt $\mathbb{E}(X)$ auch $\mathbb{E}X$, oder auch $\mathbb{E}_{\mathbb{P}}X$.
3. X besitzt genau dann einen Erwartungswert, falls $\mathbb{E}|X| < \infty$;
   $[\mathbb{E}|X| \overset{(3.2)}{=} \sum_\omega |X(\omega)|p_\omega]$
4. Aus (3.2) folgt: $\mathbb{E}(1) = 1$, $\mathbb{E}(X) \leq \mathbb{E}(Y)$ falls $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$.

3.1.3 Beispiel

Ist $X(\omega) = \{x_1, \dots, x_m\}$, und ist $\mathbb{P}_X$ die Gleichverteilung auf $X(\Omega)$, so lautet der Erwartungswert

$$\mathbb{E}(X) \overset{(3.1)}{=} \frac{1}{m}\sum_{i=1}^{m} x_i \quad (\text{“arithmetisches Mittel”})$$

[Der Erwartungswert der Augenzahl beim Würfeln ist $3,5$]

3.1.4 Beispiel

Für die Indikatorvariable $X = 1_A$ eines Ereignisses $A \subset \Omega$ gilt:

$$\mathbb{E}1_A \overset{(3.1)}{=} 0 \cdot \mathbb{P}(1_A = 0) + 1 \cdot \mathbb{P}(1_A = 1) = \mathbb{P}(A)$$

Eine direkte Folge der Darstellung (3.2) ist folgender Satz.

3.1.5 Satz

Besitzen die Zufallsvariablen X_1, X_2 Erwartungswerte, so auch die Zufallsvariablen aX_1, bX_2, $a, b \in \mathbb{R}$, und es gilt:

$$\mathbb{E}(aX_1 + bX_2 = a\mathbb{E}(X_1) + b\mathbb{E}(X_2) \quad (\text{“Linearität”})$$

3.1.6 Beispiel

Für ein $B(n,p)$-verteiltes X gilt:

$$\mathbb{E}(X) = np$$

Beweismöglichkeiten

1. $$\mathbb{E}(X) \overset{(3.1)}{=} \sum_{k=k}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \dots \overset{Bin.Lehrsatz}{=} n \cdot p \quad \checkmark$$

2. Nach 2.3.7 ist $X = X_1 + \ldots + X_n$ mit $\mathbb{E}X_i \overset{3.1.4}{=} p$. Dann ist

   $$\mathbb{E}(X) \overset{3.1.5}{=} \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = n \cdot p \quad (\text{wg. Linearität}) \quad \checkmark$$

Der Erwartungswert von Funktionen $\varphi \circ X$ einer Zufallsgröße berechnet sich wie folgt.

3.1.7 Satz

Ist $X : \Omega \to \Omega'$ Zufallsgröße auf $(\Omega, \mathbb{P})$ und $\varphi : \Omega' \to \mathbb{R}$ eine Abbildung, dann gilt

$$\mathbb{E}(\varphi \circ X) = \sum_{x \in X(\Omega)} \varphi(x) \mathbb{P}_X(\{x\}) \tag{3.3}$$

[Falls die Reihe absolut konvergiert].

Beweis:

$$\mathbb{E}_{\mathbb{P}}(\varphi \circ X) \overset{(3.2)}{=} \sum_{\omega \in \Omega} \underbrace{\varphi(X(\omega))}_{=(\varphi \circ X)(\omega)} p_\omega \overset{Zerl. v. \Omega}{=} \sum_{x \in X(\omega)} \sum_{\omega : X(\omega) = x} \varphi(X(\omega)) p_\omega =$$

$$= \sum_{x \in X(\omega)} \varphi(x) \sum_{\omega : X(\omega) = x} p_\omega = \sum_{x \in X(\omega)} \varphi(x) \mathbb{P}_X(\{x\})$$

[Großer Umordnungssatz für absolut konvergente Reihen].
Gemäß (3.3) gilt z.B.:

$$\mathbb{E}(x^2) = \sum_x x^2 \mathbb{P}_X(\{x\}),$$

$$\mathbb{E}(e^x) = \sum_x e^x \mathbb{P}_X(\{x\}),$$

usw.

Bemerkung: Nach (3.2) kann man die rechte Seite von (3.3) auch $\mathbb{E}_{\mathbb{P}_X}(\varphi)$ schreiben.
In der Tat, $\mathbb{P}_X(\{x\})$ ist Wahrscheinlichkeitsfunktion auf Ω' und φ wird als Zufallsvariable auf $(\Omega', \mathbb{P}_X)$ aufgefaßt:

$$\mathbb{E}_{\mathbb{P}}(\varphi \circ X) = \mathbb{E}_{\mathbb{P}_X}(\varphi) \tag{3.4}$$

3.1.8 Satz

Die Zufallsvariablen X, Y mögen *unabhängig* sein und Erwartungswerte besitzen. Dann besitzt auch $X \cdot Y$ einen Erwartungswert und es gilt

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Beweis:

i) Zur Existenz von $\mathbb{E}(X \cdot Y)$:

$$\mathbb{E}(|X \cdot Y|) = \sum_{\omega \in \Omega} |X(\omega) \cdot Y(\omega)| p_\omega =$$

$$= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} \sum_{\substack{\omega : X(\Omega) = x, \\ Y(\Omega) = y}} |X(\omega) \cdot Y(\omega)| p_\omega =$$

$$= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} |x \cdot y| \mathbb{P}(X = x, Y = y) =$$

$$\overset{unabh.}{=} \sum_{x \in X(\Omega)} |x| \cdot \mathbb{P}(X = x) \sum_{y \in Y(\Omega)} |y| \cdot \mathbb{P}(Y = y) =$$

$$= \mathbb{E}|X| \cdot \mathbb{E}|Y| \overset{Vorauss.}{<} \infty \quad \checkmark$$

[Reihe links oben also absolut konvergent].

ii) Die gleiche Rechnung wie in i) ohne Betragsstriche liefert:

$$\mathbb{E}(X \cdot Y) = \sum_{\omega \in \Omega} X(\omega) \cdot Y(\omega) p_\omega = \ldots$$

$$\ldots = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) \sum_{y \in Y(\Omega)} y \cdot \mathbb{P}(Y = y) =$$

$$= \mathbb{E}(X) \cdot \mathbb{E}(Y) \quad \checkmark$$

Es wurden angewandt: Doppelreihensatz und Umordnungssatz für Reihen mit nicht negativen Gliedern (in i)) und für absolut konvergente Reihen (in ii)).

Bemerkung:

1. Allgemein gilt für unabhängige Zufallsvariablen $X_1, \ldots, X_n$ mit existierenden Erwartungswerten:
$$\mathbb{E}(X_1 \cdot \ldots \cdot X_n) = \mathbb{E}(X_1) \cdot \ldots \cdot \mathbb{E}(X_n)$$

2. Bei fehlender Unabhängigkeit folgt aus $\mathbb{P}|X| < \infty$, $\mathbb{P}|Y| < \infty$ **nicht** notwendigerweise[1] $\mathbb{P}|X \cdot Y| < \infty$.

3.1.9 Bedingter Erwartungswert (Definition)

Sei $(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum, X Zufallsvariable mit existierendem Erwartungswert und $A \subset \Omega$ mit $\mathbb{P}(A) > 0$. Dann heißt

$$\mathbb{E}(X|A) = \mathbb{E}_{\mathbb{P}(\cdot|A)}(X) \tag{3.5}$$

der *bedingte Erwartungswert* von X unter (der Bedingung) A. [$\mathbb{P}(\cdot|A)$ ist die bedingte Wahrscheinlichkeitsverteilung auf Ω unter A, vgl. 2.1.1]

Formeln zur Berechnung von $\mathbb{E}(X|A)$

$$\mathbb{E}(X|A) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}|A) = \sum_{\omega \in \Omega} X(\omega) \frac{\mathbb{P}(\{\omega\} \cap A)}{\mathbb{P}(A)} =$$

$$= \underline{\frac{1}{\mathbb{P}(A)} \sum_{\omega \in A} X(\omega) p_\omega} = \frac{1}{\mathbb{P}(A)} \sum_{\omega \in \Omega} 1_A(\omega) X(\omega) p_\omega = \underline{\frac{1}{\mathbb{P}(A)} \mathbb{E}(1_a X)}$$

Bemerkung:

1. Da $\sum_\Omega X(\omega) p_\omega$ absolut konvergiert, so auch $\sum_A X(\omega) p_\omega$; also existiert $\mathbb{E}(X|A)$.
2. Spezialfälle: $\mathbb{E}(1_A|A) = 1$, $\mathbb{E}(X|\Omega) = \mathbb{E}(X)$
3. Der hier eingeführte Begriff des bedingten Erwartungswertes spielt eine untergeordnete Rolle. In der allgemeinen Wahrscheinlichkeitstheorie definiert man bedingte Erwartungswerte der Art $\mathbb{E}(X|Y)$, Y Zufallsvariable, welche von großer Wichtigkeit sind (jedoch hier in der Einführung *nicht* gebraucht werden).

3.2 Varianz, Kovarianz

Der Erwartungswert $\mathbb{E}(X)$dient als Maß der zentralen Lage der Verteilung $\mathbb{P}_X$ von X. Wir definieren nun ein Streuungs- (Schwankungs-) Maß der Verteilung $\mathbb{P}_X$.

[1] vgl. 3.2.6 unten

3.2.1 Varianz, Standardabweichung (Definition)

Ist X Zufallsvariable auf $(\Omega, \mathbb{P})$ und ist $\mathbb{E}(X^2) < \infty$, so heißt $Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}((X - \mathbb{E}(X))^2)$ *Varianz* von X, und $\sigma(X) = \sqrt{Var(X)}$ *Standardabweichung* von X. Zur Existenz der Erwartungswerte siehe 3.2.2, i) unten.

Sprechweise: Wir sagen im folgenden $X = 0\mathbb{P}$ fast überall ($\mathbb{P}$-f.n.), falls $X(\omega) = 0 \forall \omega \in \Omega$ mit $p_\omega > 0$ (äquivalent mit$\mathbb{P}(X = 0) = 1$).

3.2.2 Satz

i) Falls $\mathbb{E}(X^2) < \infty$, so auch $\mathbb{E}(X) < \infty$, $\mathbb{E}(X - a)^2 < \infty$, für $a \in \mathbb{R}$.

ii) Es gilt $\mathbb{E}(X^2) = 0$ und $Var(X) = 0$ genau dann, wenn $X = 0$ und X =const., $\mathbb{P}$ fast überall.

Beweis:

i) Wegen $|x| \leq 1 + x^2$ gilt

$$\sum_X |x| \mathbb{P}_X(\{x\}) \leq 1 + \underbrace{\sum_X x^2 \mathbb{P}_X(\{x\})}_{=\mathbb{E}(X^2)} < \infty$$

Wegen $(X - a)^2 \leq 2x^2 + 2a^2$ (!) gilt

$$\sum_X (X - a)^2 \mathbb{P}_X(\{x\}) \leq 2 \underbrace{\sum_X x^2 \mathbb{P}_X(\{x\})}_{=\mathbb{E}(X^2)} + 2a^2 < \infty$$

ii) Die Darstellung $\mathbb{E}(X^2) = \sum_\omega X^2(\omega) p_\omega = 0$, bzw. Formel (3.7) unten für $Var(X)$.

3.2.3 Formeln zur Varianz von X

(3.6) $$Var(X) = \sum_{x \in X(\Omega)} (x - \mathbb{E}(X)^2) \mathbb{P}_x(\{x\}) \quad [3.1]$$

(3.7) $$Var(X) = \sum_{\omega \in \Omega} (X(\omega) - \mathbb{E}(X)^2) p_\omega \quad [3.2]$$

(3.8) $$Var(X) = \mathbb{E}(X)^2 - (\mathbb{E}(X))^2 \quad [^2]$$

denn:

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

[allgemein: $\mathbb{E}(X - a)^2 = Var(X) + (\mathbb{E}(X) - a)^2$, $a = 0$ liefert (3.8).]

(3.9) $$Var(aX + b) = a^2 Var(X) \quad \text{für } a, b \in \mathbb{R}, \sigma(aX + b) = (a(\sigma(X)))$$

Bemerkung:

1. Denkt man sich $\mathbb{P}_X\{X\}$ als eine Massenverteilung (Gesamtmasse=1), so entspricht $\mathbb{E}(X)$ dem *Schwerpunkt* und $Var(X)$ dem *Trägheitsmoment.*

2. Ist X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$, $\sigma(X) > 0$, so gilt für die sogenannte *Standardisierte* von X, d.i. $X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}$, $\mathbb{E}(X^*) = 0$, $Var(X^*) = 1 = \sigma(X^*)$

3.2.4 Beispiel

$X : \Omega \to \Omega' = \{1, \ldots, n\}$, $\mathbb{P}_X$ Gleichverteilung auf Ω'. Dann ist

$$\mathbb{E}(X) \overset{3.1.3}{=} \frac{1}{2}(n+1)$$

$$\mathbb{E}(X^2) = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{1}{6}(n+1)(2n+1)$$

$$Var(X) \overset{(3.8)}{=} \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \frac{1}{12}(n^2-1)$$

Wir bilden nun Maße für den Zusammenhang *zweier* Zufallsvariablen X und Y.

3.2.5 Kovarianz (Definition)

Für Zufallsvariable X und Y auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$ definiert

$$\text{a)} \quad Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

die *Kovarianz* von X und Y[3] und

$$\text{b)} \quad \rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X) \cdot \sigma(Y)}$$

(sofern $\sigma(X) \cdot \sigma(Y) > 0$) den *Korrelationskoeffizienten* von X und Y.

c) X und Y heißen *unkorreliert*, falls $Cov(X,Y) = 0$

Zum Studium der Größen Cov und ρ benötigen wir den folgenden Satz.

3.2.6 Satz

Für Zufallsvariablen X, Y mit $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$ gilt:

i) $\mathbb{E}|X \cdot Y| < \infty$

ii) Cauchy-Schwarzsche-Ungleichung:

$$[\mathbb{E}(X \cdot Y)]^2 \le \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2) \tag{3.10}$$

iii) Das Gleichheitszeichen in (3.10) gilt genau dann, wenn es Zahlen $a, b \in \mathbb{R}$ gibt, mit $a^2 + b^2 > 0$ und

$$a \cdot X(\omega) + b \cdot X(\omega) = 0 \quad (\mathbb{P} \text{ fast überall lin. abh.}) \tag{3.11}$$

Beweis:

1. Aus der Ungleichung $2|x \cdot y| \le x^2 + y^2$ folgt $\mathbb{E}(|X \cdot Y|) < \infty$, d.i. i)
2. Sei $\mathbb{E}(X^2) = 0$. Dann ist nach 3.2.2 i) $X(\omega) = 0$ für alle ω mit $p_\omega > 0$ und es gilt $\mathbb{E}(X \cdot Y) = 0$ und das Gleichheitszeichen in (3.10). Ferner ist (3.10) erfüllt (mit $a = 1$, $b = 0$).
3. Sei $\mathbb{E}(X^2) > 0$. Zunächst gilt für ein beliebiges $c \in \mathbb{R}$:

$$(*) \quad 0 \le \mathbb{E}(cX - Y)^2 \overset{lin.v.\mathbb{E}}{=} c^2\mathbb{E}(X^2) - 2c\mathbb{E}(X \cdot Y) + \mathbb{E}(Y^2)$$

Einsetzen von $c = \frac{\mathbb{E}(X \cdot Y)}{\mathbb{E}(X^2)}$ liefert:

$$0 \le \frac{[\mathbb{E}(X \cdot Y)]^2}{\mathbb{E}(X^2)} - 2\frac{[\mathbb{E}(X \cdot Y)]^2}{\mathbb{E}(X^2)} + \mathbb{E}(Y^2)$$

[3] Wegen der Existenz des Erwartungswertes $\mathbb{E}[\ ]$, siehe Folg. 1 unten.

oder

$$0 \leq \mathbb{E}(Y^2) - \frac{[\mathbb{E}(X \cdot Y)]^2}{\mathbb{E}(X^2)} \quad \text{, d.h. (3.10).}$$

Hierin gilt das Gleichheitszeichen genau dann, wenn es in (*) gilt, d.h. wenn gilt:

$$(**) \quad \mathbb{E}(cX - Y)^2 = 0$$

Aus (**) folgt über 3.2.2 ii) die Gleichung (3.11) mit $a = c$, $b = -1$. Umgekehrt folgt auch aus (3.11) die Gleichung (**) (!). ✓

Folgerungen aus 3.2.6:

1. Aus

$$[(+) \quad \mathbb{E}(X^2) < \infty,\ \mathbb{E}(Y^2) < \infty$$

folgt also die Existenz des Erwartungswertes von $X \cdot Y$: $\mathbb{E}(X \cdot Y) < \infty$.
Im Fall der *Unabhängigkeit* der X, Y benötigen wir in 3.1.8 anstatt (+) nur $\mathbb{E}|X| < \infty, \mathbb{E}|Y| < \infty$.
Aus (+) folgt die Existenz der Kovarianz ($Cov(X,Y)$). In der Tat, in der Formel

$$(++) \quad Cov(X,Y) = \mathbb{E}[X \cdot Y - Y \cdot \mathbb{E}(X) - X \cdot \mathbb{E}(Y) + \mathbb{E}(X) \cdot \mathbb{E}(Y)]$$

existieren gemäß 3.2.2 i) und 3.2.6 i) sämtliche Erwartungswerte.

2. Formel (++) vereinfacht sich zur "Verschiebungsformel":

$$Cov(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{3.12}$$

3. Setzt man in (3.10) $X - \mathbb{E}(X)$, $Y - \mathbb{E}(Y)$ ein (anstatt X, Y), so erhält man

$$[Cov(X,Y)]^2 \leq \sigma^2(x) \cdot \sigma^2(Y) \tag{3.13}$$

d.h. es gilt $-1 \leq \rho(X,Y) \leq 1$

4. $|\rho(X,Y)| = 1$ genau dann, wenn

$$Y(\omega) - \mathbb{E}(Y) = c(X(\omega) - \mathbb{E}(X))$$

für alle ω mit $\rho_\omega > 0$.
Interpretation: $\rho(X,Y)$ ist ein Maß für den linearen Zusammenhang von X und Y.

3.2.7 Eigenschaften der Varianz und Kovarianz

Sei $\mathbb{E}(X^2) < \infty$, $\mathbb{E}(Y^2) < \infty$ vorausgesetzt.

a)

$$Cov(aX + b, a'Y + b') = a \cdot a' Cov(X,Y)$$

Insbesondere:

$$Cov(X^*, Y^*) = \rho(X^*, Y^*) = \rho(X,Y)$$

für $X^* = (X - \mathbb{E}(X))/\sigma(X)$, $Y^* = (Y - \mathbb{E}(Y))/\sigma(Y)$.

b) $Cov(X,Y) = Cov(Y,X)$; $Cov(X,X) = Var(X)$

c)

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i + \sum_{i \neq j} Cov(X_i, X_j)$$

Insbesondere gilt für paarweise unkorrelierte (d.h. $Cov(X_i, X_j) = 0$, für $i \neq j$) $X_1, \ldots, X_n$ die *"Formel von Bienaymé"*:

$$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i)$$

d) X, Y unabhängig $\Rightarrow$ X, Y unkorreliert.

Beweis zu c): Wegen der Eigenschaft (3.9) der Varianz und a) können wir annehmen, daß $\mathbb{E}(X_i) = 0$. Dann gilt:

$$Var(X_1 + \ldots + X_n) = \mathbb{E}(X_1 + \ldots + X_n)^2 = \mathbb{E}[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i \cdot X_j] =$$

$$= \sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i \cdot X_j)$$

ad d): X, Y unabhängig $\Rightarrow \mathbb{E}(X \cdot Y) \overset{3.1.8}{=} \mathbb{E}(X) \cdot \mathbb{E}(Y) \Rightarrow Cov(X, Y) = 0$ (Verschiebungformel).

3.3 Erzeugende Funktionen

Erzeugende Funktionen nützen bei der Berechnung von:

- Momenten
- Faltungen
- Grenzwerten von Wahrscheinlichkeiten

3.3.1 Erzeugende Funktion (Definition)

Ist X eine Zufallsvariable auf $(\Omega, \mathbb{P})$ mit Werten in $\mathbb{Z}_+$, so heißt:

$$G_X(s) = G_{\mathbb{P}_X}(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}_X\{k\} \quad , 0 \leq s \leq 1 \tag{3.14}$$

die *erzeugende Funktion* von X [von $\mathbb{P}_X$].
Wegen $\sum_{k=0}^{\infty} \mathbb{P}\{k\} = 1$ stellt G_X eine Potenzreihe dar, mit Konvergenzradius ≥ 1. (3.14) ist also wohldefiniert, $G_X(s)$ ist beliebig oft differenzierbar in $[0, 1[$.

Bemerkung:

1. Wegen $\frac{d^n}{ds^n} G_X|_{s=0} = n! \mathbb{P}_X\{n\}$, $n \in \mathbb{Z}_+$, so daß die Zuordnung $\mathbb{P}_X \to G_X(s)$ injektiv ist.
2. Man beachte auch die folgende Schreibweise: $G_x(s) = \mathbb{E}s^x$ (vgl. Satz 3.1.7).
3. Für unabhängige Zufallsvariablen $X_1, \ldots, X_n$ mit Werten in $\mathbb{Z}_+$ gilt:

$$G_{X_1, \ldots, X_n}(s) = \mathbb{E}s^{X_1} \cdot \ldots \cdot s^{X_n} \overset{3.1.8}{=} \mathbb{E}s^{X_1} \cdot \ldots \cdot \mathbb{E}s^{X_n} = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s)$$

3.3.2 Satz

Sei X eine Zufallsvariable mit Werten in $\mathbb{Z}_+$.

a) Der (linksseitige) Grenzwert $G_X'(1-) = \lim_{s \uparrow 1} G'(s)$ existiert genau dann, wenn $\mathbb{E}(X)$ existiert. In diesem Fall ist $\mathbb{E}(X)$=G'(1-)

b) Es existiere $\mathbb{E}(X)$. $G''(1-)$ existiert genau dann, wenn $\mathbb{E}(X^2)$ existiert. In diesem Fall ist $Var(X) = G_X''(1-) + G_X'(1-) - (G_X'(1-))^2$.

Beweis:

a) Zunächst gilt für $s \in [0,1[$:

$$G'_X(s) = \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}_X\{k\}$$

i) Sei $\mathbb{E}(X) < \infty$.

$$G'_X(s) \leq \sum_{k=1}^{\infty} k \mathbb{P}_X\{k\} = \mathbb{E}(X) < \infty$$

Also auch

$$\lim_{s \uparrow 1} G'_X(s) \equiv G'_X(1-) \leq \mathbb{E}(X)$$

ii) Sei $G'_X(1-) < \infty$. Für $s \in [0,1[$ gilt:

$$\sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}_X\{k\} \leq G'_X(s) \leq \lim_{n \uparrow 1} G'_X(n) = G'_X(1-) < \infty$$

Also auch beliebige $s \uparrow 1$:

$$\sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}_X\{k\} \leq G'_X(1-)$$

und bei $n \uparrow \infty$:

$$\mathbb{E}(X) \leq G'_X(1-) < \infty$$

Aus i) und ii) ferner:

$$G'_X(1-) \leq \mathbb{E}(X) \leq G'_X(1-)$$

b) **Vorbemerkung:** $\mathbb{E}(X^2) < \infty$ genau dann, wenn $\mathbb{E}(\underbrace{X(X-1)}_{\geq 0}) < \infty$ (vgl. 3.2.2, i))

Ausgehend von $G''_X(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} \mathbb{P}_X\{k\}$, $s \in [0,1[$, zeigt man wie in a) die gleichzeitige Existenz von $\mathbb{E}(X(X-1))$ und $G''_X(1-)$. In welchem Fall dann $\mathbb{E}(X(X-1)) = G''_X(1-)$ ist.
die Verschiebungsformel schließlich liefert $Var(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X^2)$ ✓

Beispiel: (Binomialverteilung, B(n,p))
Sei X $B(n,p)$-verteilt, so rechnet man mit $q = 1-p$

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} (p \cdot s)^k q^{n-k} \overset{Bin.LS}{=} (q+ps)^n \quad , s \in \mathbb{R}$$

Ableitung an der Stelle 1 liefert, in Übereinstimmung mit 3.1.6:

$$G'_X(1) = n(q+ps)^{n-1} \cdot p|_{s=1} = np = \mathbb{E}(X)$$

$$G''_X(1) = n(n-1)(q+ps)^{n-2} \cdot p^2|_{s=1} = n(n-1)p^2$$

$$\rightsquigarrow Var(X) = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

Seien X_1, X_2 unabhängige $B(n_1,p)$- bzw. $B(n_2,p)$-verteilte Zufallsvariablen. Aus Bemerkung 3 folgt:

$$G_{X_1+X_2}(s) = G_{X_1}(s) \cdot G_{X_2}(s) = (q+ps)^{n_1+n_2}$$

d.h. $\mathbb{P}_{X_1+X_2}$ ist die $B(n_1+n_2,p)$-Verteilung (mit Bemerkung 1), kurz:

$$B(n_1,p) * B(n_2,p) = B(n_1+n_2,p)$$

3.3.3 Poissonverteilung (Beispiel)

Ist X $P(\lambda)$-verteilt, so ist

$$G_X(s) = \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \cdot e^{-\lambda} = e^{\lambda s} e^{-\lambda} = e^{\lambda(s-1)} \quad , e \in \mathbb{R}.$$

Ableitungen an der Stelle 1:

$$G'_X(1) = \lambda e^{\lambda(s-1)}|_{s=1} = \lambda$$
$$G''_X(1) = \lambda^2 e^{\lambda(s-1)}|_{s=1} = \lambda^2$$
$$Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

(Erwartungswert und Varianz jew. gleich λ)
Sind X_1 und X_2 unabhängige, $P(\lambda_1)$- bzw. $P(\lambda_2)$-verteilte Zufallsvariablen, so gilt:

$$G_{X_1+X_2}(s) \overset{Bem.3}{=} G_{X_1}(s) \cdot G_{X_2}(s) = e^{(\lambda_1+\lambda_2)(s-1)}$$

d.h. $P(\lambda_1) * P(\lambda_2) = P(\lambda_1 + \lambda_2)$, mit Bemerkung 1.

3.3.4 Negative Binomialverteilung (Definition)

[$(NB(n,p))$, $n \in \mathbb{N}$, $0 < p \leq 1$, setze $q = 1-p$]
Eine Wahrscheinlichkeitsverteilung auf $\mathbb{Z}$ mit der Wahrscheinlichkeitsfunktion

$$p_k = \binom{n+k-1}{k} \cdot p^n q^k \quad , k = 0, 1, \ldots$$

heißt *negative Binomialverteilung*. Man kann auch schreiben:

$$p_k = \binom{-n}{k} (-1)^k p^n q^k$$

wobei $\binom{r}{k} = \frac{(r)_k}{k!}$, $r \in \mathbb{R}$.
Im Spezialfall $n = 1$ spricht man von einer *geometrischen* Verteilung: $P_k = p \cdot q^k$, $k = 0, 1, \ldots$ (vgl. 1.2.7).
Zählt X die Anzahl der 'Mißerfolge' 0 bis zum Auftreten des n-ten 'Erfolges' 1 (unabhängige Wiederholung), so ist X $NB(n,p)$-verteilt, $p = \mathbb{P}\{1\} > 0$.
Man rechnet $G_X(s) = \left(\frac{p}{1-qs}\right)^n$, $0 \leq s \leq \frac{1}{q}$ (Binomische Reihe[4])

$$\mathbb{E}(X) = \underbrace{n\frac{q}{p}}_{\equiv \mu} \ , \quad Var(X) = \underbrace{n\frac{q}{p^2}}_{\equiv \sigma^2} \ (!)$$

[$\sigma^2 = \frac{\mu}{p} \overset{0<p<1}{>} \mu$, 'overdispension']
Es gilt:

$$\underbrace{NB(1,p) * \ldots * NB(1,p)}_{n-mal} = NB(n,p)$$

Grenzwerte von Wahrscheinlichkeitsverteilungen:

[4] siehe Forster I, S.182

Notation: Ist $(p_0, p_1, \ldots)$ eine Wahrscheinlichkeitsfunktion auf $\mathbb{Z}_+$, ($p_k \geq 0$, $\sum_{k=0}^\infty p_k = 1$), so bezeichnet $G(s) = \sum_{k=0}^\infty s_k p_k$, $0 \leq s \leq 1$, ihre erzeugende Funktion.
Liegt eine Folge

(3.15) $$(p_0^{(n)}, p_1^{(n)}, \ldots) \quad , n \geq 1$$

von Wahrscheinlichkeitsfunktionen auf $\mathbb{Z}_+$ vor, so konvergieren die Wahrscheinlichkeiten $p_k^{(n)}$ bei $n \to \infty$ genau dann, wenn die Folge der zugehörenden erzeugenden Funktionen

(3.16) $$G^{(n)}(s) = \sum_{k=0}^\infty s^k p_k^{(n)} \ , \quad 0 \leq s \leq 1,\ n \geq 1$$

konvergiert. Genauer:

3.3.5 "Stetigkeitssatz"

Gegeben sei eine Folge (3.15) von Wahrscheinlichkeitsfunktionen aus $\mathbb{Z}_+$, mit Folge (3.16) der zugehörenden erzeugenden Funktionen. Dann existieren die Limiten

(3.17) $$a_k = \lim_{n\to\infty} p_k^{(n)}$$

für alle $k = 0, 1, \ldots$ genau dann, wenn der Limes

(3.18) $$A(s) = \lim_{n\to\infty} G^{(n)}(s)$$

für alle $s \in (0, 1)$ (offenes Intervall) existiert. In diesem Fall ist

(3.19) $$A(s) = \sum_{k=0}^\infty s^k a_k$$

Bemerkung: Aus (3.17) folgt $a_k \geq 0$ und $\sum_{k=0}^\infty a_k \leq 1$ ($a_0, a_1, \ldots$ bilden nicht notwendigerweise eine Wahrscheinlichkeitsfunktion).

Beweis:

i) Wir nehmen (3.17) an und definieren $A(s)$ durch (3.19). Wegen $|p_k^{(n)} - G_k| \leq 1$ gilt für $s \in (0, 1)$:

$$|G^{(n)}(s) - A(s)| \leq \sum_{k=0}^{r} |p_k^{(n)} - a_k| s^k + \underbrace{\sum_{k=r+1}^\infty s^k}_{\frac{s^r}{(1-s)}}$$

Zu $\epsilon > 0$ wähle r so groß, daß $\frac{s^r}{(1-s)} < \epsilon$. Dann $\lim_{n\to\infty} |G^{(n)}(s) - A(s)| \leq \epsilon$.

ii) Gelte nun (3.18). Wir zeigen (3.17) sukzessive für $k = 0, 1, \ldots$.
Zunächst $k = 0$:

$$p_0^{(n)} \leq G^{(n)}(s) \leq p_0^{(n)} + \sum_{k=1}^\infty 1 s^k = p_0^{(n)} + \frac{s}{1-s} \ , \quad s \in (0, 1)$$

Für jeden Häufungspunkt $\overline{p_0}$ der beschränkten Folge $p_0^{(n)}$, $n \geq 1$, gilt demnach:

(3.20) $$A(s) - \frac{s}{1-s} \leq \overline{p_0} \leq A(s) \ , \quad \text{für alle } s \in (0, 1)$$

Wegen der Monotonie von $A(s)$ existiert also $A(0) = \lim_{s\to 0} A(s)$, so daß (3.20) bei $s \downarrow 0$ liefert:

$$(*) \quad \overline{p_0} = A(0) = \lim_n p_0^{(n)}$$

Zu $k = 1$: Aus (3.18) folgt:

$$\frac{G^{(n)}(s) - p_0^{(n)}}{s} \overset{n\to\infty}{\Longrightarrow} \frac{A(s) - A(0)}{s}$$

links: Potenzreihe mit $p_1^{(n)}$ als Anfangsglied, rechts: Differenzquotient, der gegen $A'(0)$ bei $s \downarrow 0$ konvergiert (!).
Analog zu k=0:

$$\overline{p_1} = A'(0) = \lim_n p_1^{(n)} \quad (\dots) \quad \checkmark$$

3.3.6 Beispiel

(Poissonverteilung als Grenzwert der negativen Binomialverteilung)
Nach 3.3.4 lautet die erzeugende Funktion der $NB(n,p)$ gleich $\left(\frac{1-q}{1-sq}\right)^n$. Nun gehe $n \to \infty$, aber so, daß $p = p_n \to 1$, genauer:
Es gibt ein $\lambda > 0$ mit (*) $nq_n \to \lambda$, ($q_n = 1 - p_n$). Die erzeugende Funktion der $NB(n,p)$ lautet dann, mit $\lambda_n = n \cdot q_n \to \lambda$:

$$G_s^{(n)} = \left(\frac{1 - \frac{\lambda_n}{n}}{1 - \frac{\lambda_n s}{n}}\right)^n \overset{(**), n\to\infty}{\longrightarrow} \frac{e^{-\lambda}}{e^{\lambda s}} = e^{\lambda(s-1)} \equiv G(s)\ , \quad s \in \mathbb{R}$$

$G(s)$ ist gemäß 3.3.3 erzeugende Funktion der $P(\lambda)$-Verteilung. Also folgt aus dem Stetigkeitssatz:

$$\binom{n+k-1}{k} p_n^n q_n^k \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}\ , \quad \text{unter } (*),\ k = 0, 1, \dots \tag{3.21}$$

Argument zu (**): Sei $\epsilon, \delta > 0$: Für $n \geq n_0 = n_0(\epsilon, \delta)$ gilt:

$$(1 - \frac{\lambda_n}{n})^n \begin{Bmatrix} \leq (1 - \frac{\lambda - \epsilon}{n})^n \leq e^{-(\lambda-\epsilon)} + \delta \\ \geq (1 - \frac{\lambda + \epsilon}{n})^n \geq e^{-(\lambda-\epsilon)} - \delta \end{Bmatrix} \rightsquigarrow (1 - \frac{\lambda_n}{n}) \overset{\substack{n\to\infty \\ \lambda_1 \to \lambda}}{\longrightarrow} e^{-\lambda}$$

Interpretation: Für große n ist k die Anzahl der auftretenden Ereignisse E ('Mißerfolge' in 3.3.4) in einer *langen* Beobachtungsperiode, wobei E eine sehr kleine Auftrittswahrscheinlichkeit $q = \frac{\lambda}{n}$ hat.

Bemerkung: Die Anwendung von Satz 3.3.5 ist auf alle Fälle beschränkt, in denen die 'Grenzverteilung' (in 3.3.6 die Poissonverteilung) die diskrete Wahrscheinlichkeitsverteilung ist. Der wichtigste Fall ist dabei gerade der mit einer 'stetigen Grenzverteilung' (vgl. 3.4.2, 5.2 unten), so daß wir dort einen anderen 'Stetigkeitssatz' benötigen werden.

3.4 Approximation der Binomialverteilung

Für große n sind Wahrscheinlichkeiten, die nicht mit der $B(n,p)$-Verteilung verknüpft sind, größenordnungsmäßig schlecht zu erfassen und umständlich zu berechnen. Wir wollen sie durch handlichere Größen approximieren. Aber auch theoretisches Interesse an solch einer Approximation (3.4.2).

Vorüberlegungen zu 3.4.2: Sei $X^{(n)}$ $B(n,p)$-verteilt, $0 < p < 1$. Wegen $\mathbb{E}X^{(n)} = np$ und $Var(X^{(n)}) = np(1-p)$ läuft die Verteilung für wachsendes n einerseits nach rechts, andererseits 'verläuft' sie auch in die Breite. Wir bilden deshalb die Standardisierte von $X^{(n)}$, d.h. $X_n = \frac{X^{(n)} - np}{\sqrt{np(1-p)}}$.
Die Verteilung von X_n liegt 'glockenförmig' um 0, allerdings in 'diskreter Form'. Um die ideale 'Glockenform' analytisch zu beschreiben, führen wir die Funktion $\varphi(x)$ ein:

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

Wir nennen φ die Dichte der Standard-Normalverteilung[5].

3.4.1 Hilfssatz

Es gilt:

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

Beweis:

$$\left(\int_{-\infty}^{\infty} e^{\frac{-t^2}{2}} dt\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} e^{\frac{-y^2}{2}} dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(x^2+y^2)}{2}} dxdy =$$

$$= \int_{-\infty}^{2\pi}\int_{-\infty}^{\infty} e^{\frac{-r^2}{2}} rdrd\vartheta = 2\pi \underbrace{\int_0^{\infty} e^{\frac{-r^2}{2}} rdr}_{=1} = 2\pi \quad \checkmark$$

Polarkoordinatentransformation:

$$x = r \cdot \cos\vartheta\ ,\ \ y = r \cdot \sin\vartheta$$

$$y^2 + x^2 = r^2$$

$$\left|\det\begin{pmatrix} \cos\vartheta & -r\cdot\sin\vartheta \\ \sin\vartheta & r\cdot\cos\vartheta \end{pmatrix}\right| = r(\sin^2\vartheta + \cos^2\vartheta) = r$$

zum Beweis: Ferner definieren wir die sogenannte Verteilungsfunktion ϕ der Standardnormalverteilung[6]:

$$\phi(x) = \int_{-\infty} \infty\varphi(t)dt\ ,\ \ x \in \mathbb{R}$$

3.4.2 Satz (DeMoivre-Laplace)

Ist $X^{(n)}$ $B(n,p)$-verteilt, $0 < p < 1$, so gilt mit der Standardisierten $X_n^k = \frac{X^{(n)}-np}{\sqrt{np(1-p)}}$:

$$\mathbb{P}(a \le X_n^* \le b) \overset{n\to\infty}{\longrightarrow} \int_a^b \varphi(x)dx = \phi(b) - \phi(a)$$

Beweis: Der Satz ist ein Spezialfall des sogenannten zentralen Grenzwertsatzes[7]. ✓

Bemerkung:

1. Insbesondere gilt: $\mathbb{P}(X_n^* \le b) \overset{n\to\infty}{\longrightarrow} \phi(b)$, $\mathbb{P}(X_n^* \ge a) \overset{n\to\infty}{\longrightarrow} 1 - \phi(b) = \phi(-a)$.
2. Zur approx. Berechnung der Wahrscheinlichkeit von $k \le X^{(n)} \le l$ geht man wie folgt vor: Bilde $a = \frac{k-np}{\sqrt{np(1-p)}}$, $b = \frac{l-np}{\sqrt{np(1-p)}}$; $[a \equiv a_n,\ b \equiv b_n]$, dann

$$\mathbb{P}(\underbrace{k \le X^{(n)} \le l}_{E_1}) = \mathbb{P}(\underbrace{a \le X_n^* \le b}_{E_2}) \overset{3.4.2}{\approx} \phi(b) - \phi(a)$$

   mit E_1, E_2 gleiche Ereignisse, und wenn n groß genug ist. [Faustregel: $np(1-p) \ge 10$]
3. Für kleine n ist noch eine sogenannte 'Stetigkeitskorrektur' nützlich:

$$a = \frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}}\ ,\ \ b = \frac{l - np + \frac{1}{2}}{\sqrt{np(1-p)}}$$

[5] vgl. 4.2.9 unten
[6] vgl. 4.2.9
[7] siehe 5.2 unten

3.4.3 Beispiel

Ein Medikament heile einen Patienten mit der Wahrscheinlichkeit $p = 0,8$. Wie groß ist die Wahrscheinlichkeit, daß unter $n = 1000$ Patienten (denen das Medikament verabreicht wird) mindestens $k = 780$ Patienten geheilt werden?
Man rechnet:

$$a = \frac{k - np}{\sqrt{np(1-p)}} = \frac{780 - 800}{\sqrt{160}} = -1.581$$

$$W = \mathbb{P}(X^{(n)} \geq 780) \approx 1 - \phi(-1.581) = \phi(1.581) = 0.943$$

(Mit dem Stetigkeitssatz hätte man erhalten: $a = -1.621$, $\phi(-a) = 0.947$.)
Lassen wir p nicht konstant, sondern geht $p = p_n$ bei wachsendem n gegen 0, so kann die Binomialverteilung durch die Poissonverteilung approximiert werden. Setze $b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$.

3.4.4 Poissonscher Grenzwertsatz

Ist p_n, $n \geq 1$, eine Folge $\in (0, 1)$ mit

$$(*) \quad \lim_{n \to \infty} n p_n = \lambda > 0$$

so gilt

$$\lim_{n \to \infty} b(k; n, p_n) = \frac{\lambda^k}{k!} e^{-\lambda} \ , \quad k \in \mathbb{Z}$$

Beweis: Übungsaufgabe 23 (Stetigkeitssatz anwenden!)

3.4.5 Beispiel

2% der Bevölkerung habe eine bestimmte Krankheit. Wie groß ist die Wahrscheinlichkeit, daß unter 100 Personen (zufällig herausgegriffen) mindestens 3 diese Krankheit haben?

a) X binomialverteilt, $B(100, 0.02)$-Verteilung
   $W = 0.323314$, $(\mathbb{P}(X \geq 3))$

b) X poissonverteilt, $P(2)$-Verteilung
   $W = \mathbb{P}(X \geq 3) = 0.323324$

[Letzter Grenzwertsatz dieser Art: Übungsaufgabe 22]

Kapitel 4

Allgemeine Wahrscheinlichkeitstheorie

4.1 Wahrscheinlichkeitsräume

Bisher beschränkten wir uns auf ein abzählbares Ω (Vermeidung technischer Schwierigkeiten). Es gibt jedoch Zufallsexperimente, für welche ein überabzählbares Ω angemessen ist.

1. Messung einer physikalischen Größe mit großer Genauigkeit. ($\Omega = \mathbb{R}$)
2. Exakter Zeitpunkt des Eintretens eines Erdbebenstoßes oder eines Telephonanrufs. ($\Omega = \mathbb{R}_+$)
3. Idealisiertes Roulette. ($\Omega = [0, 2\pi[$)
4. Pseudo-Zufallszahlen. ($\Omega = [0, 1]$)

In Beispiel 4. verlangen wir intuitiv von einer Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $[0,1]$: $\mathbb{P}([a,b]) = b - a$, $0 \leq a \leq b \leq 1$, inbesondere $(*)$ $\mathbb{P}(\{\omega\}) = 0$.
Das mathematische Problem besteht nun darin, daß es keine Abbildung $\mathbb{P} : \mathfrak{P}([0,1]) \to [0,1]$ gibt, die normiert und σ-additiv ist und (*) erfüllt.
Ausweg aus dem Dilemma: Statt auf ganz $\mathfrak{P}(\Omega)$, definiert man $\mathbb{P}$ nur nich aus ein Teilsystem von Teilmengen $A \subset \Omega$, d.h. auf einem Teilsystem $\mathfrak{A} \subset \mathfrak{P}(\Omega)$. $\mathfrak{A}$ sollte so beschaffen sein, daß die üblichen Mengenoperationen $(\cup, \cap \ldots)$ nicht aus $\mathfrak{A}$ herausführen.

4.1.1 σ-Algebra (Definition)

Ist Ω eine beliebige, nichtleere Menge, so heißt ein Mengensystem $\mathfrak{A} \subset \mathfrak{P}(\Omega)$ eine σ-Algebra über Ω, wenn gilt:

a) $\Omega \in \mathfrak{A}$

b) $A \in \mathfrak{A} \Rightarrow \overline{A} \in \mathfrak{A}$

c) $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}$

Bemerkungen:

1. Es ist $\emptyset = \overline{\Omega} \in \mathfrak{A}$ (wg. a), b))
2. $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}} \in \mathfrak{A}$ (wg. b), c))
3. In c) bzw. 2. können wir auch $\bigcup_{i=1}^{\infty}$ bzw. $\bigcap_{i=1}^{\infty}$ einsetzen. Setze $A_{n+1} = A_{n+2} = \ldots = \emptyset$ bzw. $A_{n+1} = A_{n+2} = \ldots = \Omega$.
4. $\mathfrak{P}(\Omega)$ ist eine σ-Algebra ('größte'), $\{\emptyset, \Omega\}$ ist eine σ-Algebra ('kleinste').

5. Ist $\mathfrak{F} \subset \mathfrak{P}(\Omega)$ eine vorgegebenes Mengensystem, so existiert unter den σ-Algebraen, die $\mathfrak{F}$ umfassen, eine kleinste σ-Algebra (!). Wir nennen sie die von $\mathfrak{F}$ erzeugte σ-Algebra $\sigma(\mathfrak{F})$. $\mathfrak{F}$ heißt dann Erzeugendensystem von $\sigma(\mathfrak{F})$.

6. Ein Paar $(\Omega, \mathfrak{A})$, $\mathfrak{A}$ σ-Algebra über Ω, heißt ein meßbarer Raum.

4.1.2 Definition

Ein Triplet $(\Omega, \mathfrak{A}, \mathbb{P})$ heißt (allgemeiner) Wahrscheinlichkeitsraum, falls

a) Ω nichtleere Menge

b) $\mathfrak{A}$ σ-Algebra über Ω

c) $\mathbb{P}\colon \mathfrak{A} \to [0, 1]$ mit

   (i) $\mathbb{P}(\Omega) = 1$

   (ii) $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty}(A_i)$ für *paarweise disjunkte* $A_1, A_2, \ldots \in \mathfrak{A}$

Bemerkungen:

1. $\mathbb{P}$ heißt Wahrscheinlichkeitsverteilung auf $(\Omega, \mathfrak{A})$. Auch die übrigen Bezeichnungen von 1.2.1 und die Folgerungen aus 1.2.2 sind weiterhin gültig, wenn man $\mathfrak{P}(\Omega)$ durch die σ-Algebra $\mathfrak{A}$ ersetzt ($A \in \mathfrak{A}$ statt $A \in \Omega$ (oder $A \in \mathfrak{P}(\Omega)$)).

2. Der diskrete Wahrscheinlichkeitsraum 1.2.1 ergibt sich als Spezialfall von 4.1.2:
   Ω abzählbar, $\mathfrak{A} = \mathfrak{P}(\Omega) \overset{(!)}{=} \sigma\{\{\omega\}\omega \in \Omega\}$

Der für das folgende wichtigste Fall eines meßbaren Raumes $(\Omega, \mathfrak{A})$ wird wie folgt konstruiert.

4.1.3 Borelsche σ-Algebra

Konstruktion der Borelschen σ-Algebra über $\Omega = \mathbb{R}^k$, $k \geq 1$.
Das Mengensystem $\mathfrak{F}^k \subset \mathfrak{P}(\overline{\mathbb{R}^k})$ bestehe aus allen k-dimensionalen Intervallen. Für $a = (a_1, \ldots, a_k)$, $b = (b_1, \ldots, b_k) \in \overline{\mathbb{R}^k}$ (d.h. $a_i = -\infty$ und $b_i = +\infty$ sind zugelassen) mit $a < b$ (d.h. $a_i < b_i$ für $i = 1, \ldots, k$) definiert man das k-dimensionale Intervall $(a, b] = \bigotimes_{i=1}^{k}(a_i, b_i] = \{x \in \mathbb{R}^k : a_i < x_i < b_i$, für $i = 1, \ldots, k\}$.[1]
Man führt das Mengensystem $\mathfrak{F}^k = \{(a, b], a < b\}$ ein (beachte $\mathbb{R}^k \in \mathfrak{F}^k$).
Gemäß Bemerkung 4., 4.1.1, sei $\mathfrak{B}^k = \sigma(\mathfrak{F}^k)$ die kleinste σ-Algebra, die alle k-dimensionalen Intervalle auf $\mathfrak{F}^k$ enthält. $\mathfrak{B}$ heißt *σ-Algebra der Borelschen Mengen* oder kurz *Borelsche σ-Algebra*.

4.1.4 Satz

Satz aus der Topologie / Maßtheorie:

a) Die σ-Algebra $\mathfrak{B}^k$ der Borelschen Mengen enthält alle offenen und alle abgeschlossenen Mengen des $\mathbb{R}^k$.

b) $\mathfrak{B}$ wird auch erzeugt von jedem der drei folgenden Mengensystemen:

   - das System der offenen Intervalle (a, b) des $\mathbb{R}^k$
   - das System der abgeschlossenen Intervalle $[a, b]$ des $\mathbb{R}^k$
   - das System der links abgeschlossenen und rechts offenen Menge $[a, b)$ des $\mathbb{R}^k$

c) Es gibt nicht-Borelsche Mengen des $\mathbb{R}^k$, d.h. es ist $\mathfrak{B}^k \not\subseteq \mathfrak{P}(\mathbb{R}^k)$.

Zur Festlegung einer Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\mathbb{R}^k$ braucht man (zum Glück!) nicht alle $\mathbb{P}(B)$, $B \in \mathfrak{B}^k$ auf allen Intervallen. Es gilt nämlich folgender Satz.

[1] $(a_i < x_i < \infty$, falls $b = \infty)$

4.1.5 Fortsetzungssatz von Caratheodory

Sei $\tilde{\mathbb{P}} : \mathfrak{J}^k \to [0,1]$ eine Abbildung, so daß gilt:

i) $\tilde{\mathbb{P}}(\mathbb{R}^k = 1$

ii) $\mathbb{P}(\bigcup_{i=1}^{\infty} I_i) = \sum_{i=1}^{\infty} \mathbb{P}(I_i)$ für paarweise disjunkte $I_1, I_2, \ldots \in \mathfrak{J}^k$ mit $\bigcup_{i=1}^{\infty} I_i \in \mathfrak{J}^k$

Dann existiert genau eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\mathfrak{B}^k$, so daß $\mathbb{P}|\mathfrak{J}^k = \tilde{\mathbb{P}}$ (d.h. $\mathbb{P}(I) = \tilde{\mathbb{P}}(I)$ für alle $I \in \mathfrak{J}^k$). [$\mathbb{P}$ heißt Fortsetzung von $\tilde{\mathbb{P}}$ auf ganz $\mathfrak{B}^k$.]

Bemerkung: Öfter ist nun eine Teilmenge von $\mathbb{R}^k$ als Ergebnisraum Ω von Interesse (z.B.: $\Omega = [0,1]^k$). Dann werden alle Größen auf $\Omega \subset \mathbb{R}^k$ eingeschränkt: $\Omega \cap \mathfrak{J}^k = \{\Omega \cap I,\ I \in \mathfrak{J}^k\}$ statt $\mathfrak{J}^k$; $\Omega \cap \mathfrak{B}^k = \{\Omega \cap \mathfrak{B}^k,\ B \in \mathfrak{B}^k\}$ statt $\mathfrak{B}^k$ ('Borelsche Mengen in Ω'); $\mathbb{P}|\Omega \cap \mathfrak{B}^k$ statt $\mathbb{P}$ ('Restriktionen von $\mathbb{P}$ auf $\Omega \cap \mathfrak{B}^k$') $(\Omega,\ \Omega \cap \mathfrak{B}^k,\ \mathbb{P}|\Omega \cap \mathfrak{B}^k)$ bildet einen Wahrscheinlichkeitsraum (!).

4.1.6 Idealisiertes Roulette (Beispiel)

$\Omega = [0, 2\pi)$, $\Omega \cap \mathfrak{J}^1 = \{(a,b], 0 \le a < b < 2\pi\}$.
Durch $\tilde{\mathbb{P}}(a,b] = \frac{b-a}{2\pi}$ wird auf $\Omega \cap \mathfrak{J}^1$ eine Abbildung in $[0,1]$ definiert, welche Eigenschaften (i) und (ii) aus 4.1.5 erfüllt.
$\tilde{\mathbb{P}}$ legt eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\Omega \cap \mathfrak{B}^1$ fest ('Gleichverteilung auf $[0, 2\pi)$').

4.1.7 Beispiel

Zeitpunkt des Auftretens eines Ereignisses $\Omega = [0, \infty)$; durch $\tilde{\mathbb{P}}(a,b] = e^{-\lambda a} e^{-\lambda b}$, $0 \le a \le b < \infty$ ($\lambda > 0$ fest) wird eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $[0,\infty) \cap \mathfrak{B}^1 \equiv \mathfrak{B}^1_+$ festgelegt ('Exponentialverteilung mit Parameter λ').

Bemerkung: Zukünftig schreiben wir statt $\tilde{\mathbb{P}}$ ebenfalls $\mathbb{P}$.

4.1.8 Unabhängigkeit

Die Unabhängigkeit von Ereignissen $A_1, \ldots, A_n \in \mathfrak{A}$ in einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$ definiert man wie in 2.2.2 durch die Eigenschaft: $\mathbb{P}(A_{j_1} \cap \ldots \cap A_{j_k}) = \mathbb{P}(A_{j_1}) \cdot \ldots \cdot \mathbb{P}(A_{j_k})$; für alle $\emptyset \ne \{j_1, \ldots, j_k\} \subset \{1, \ldots, n\}$. Sind $\mathbb{P}_1, \ldots, \mathbb{P}_k$ Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathfrak{B}^1)$, so heißt die Wahrscheinlichkeitsverteilung (wie in 2.3.3) $\mathbb{P}$ auf $(\mathbb{R}^k, \mathfrak{B}^k_j)$ *Produkt* der $\mathbb{P}_1, \ldots, \mathbb{P}_k$, kurz $\mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_k$, falls $\mathbb{P}(B_1 \times \ldots \times B_k) = \mathbb{P}_1(B_1) \cdot \ldots \cdot \mathbb{P}_k(B_k)$; für alle $B_1, \ldots, B_k \in \mathfrak{B}^1$.

Bemerkung: Der Begriff des Produktes von (allgemeinen) Wahrscheinlichkeitsräumen $(\Omega_i, \mathfrak{A}_i, \mathbb{P}_i)$, $i = 1, \ldots, n$ verlangt den Begriff der Produkt-σ-Algebra $\mathfrak{A} = \mathfrak{A}_1 \times \ldots \times \mathfrak{A}_k$. Wir beschränken uns auf den Spezialfall: $\mathfrak{A}_i = \mathfrak{B}^1$, $\mathfrak{A} = \mathfrak{B}^k$, für den wir diesen Begriff nicht benötigen.

4.1.9 (Elementare) bedinget Wahrscheinlichkeit

Der Begriff $\mathbb{P}(B|A)$. falls $\mathbb{P}(A) > 0$, der (elementaren) bedingten Wahrscheinlichkeit, und die Formel von der totalen Wahrscheinlichkeit, Bayessche Formel, Produkt gelten auch für Wahrscheinlichkeitsräume $(\Omega, \mathfrak{A}, \mathbb{P})$, falls alle auftretenden Ereignisse $A, B, A_1, \ldots, A_m$ aus $\mathfrak{A}$ genommen werden. Der allgemeine Begriff der bedingten Wahrscheinlichkeitsverteilung und des bedingten Erwartungswertes wird hier nicht gebracht[2].

[2] vgl. 3.1.9, Bemerkung 3.

4.2 Verteilungsfunktion, Dichte

Zunächst Beschränkung auf der Wahrscheinlichkeitsraum $(\mathbb{R}, \mathfrak{B}^1, \mathbb{P})$. Zur Festlegung von Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathfrak{B}^1)$ (bzw. auf $(\mathbb{R} \cap \Omega, \mathfrak{B}^1 \cap \Omega)$) reicht es aus, wegen $(a, b] = (-\infty, b] \setminus (-\infty, a]$ und $\mathbb{P}(a, b] = \mathbb{P}(-\infty, b] - \mathbb{P}(-\infty, a]$, alleine die Funktion $F(t) = \mathbb{P}(-\infty, t]$, $t \in \mathbb{R}$, zu betrachten.

4.2.1 (kumulative) Verteilungsfunktion (Definition)

Sei $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}, \mathfrak{B}^1)$. Dann heißt die Funktion $F_{\mathbb{P}} : \mathbb{R} \to [0, 1]$, $F_{\mathbb{P}}(t) = \mathbb{P}(-\infty, t]$, $t \in \mathbb{R}$, *(kumulative) Verteilungsfunktion* von $\mathbb{P}$.
[Im Folgenden sei $F(t+) = \lim_{s \downarrow t} F(s)$, $F(t-) = \lim_{s \uparrow t} F(s)$, (falls existiert)]

4.2.2 Satz

Sei $F(t) \equiv F_{\mathbb{P}}(t)$, $t \in \mathbb{R}$, Verteilungfunktion von $\mathbb{P}$. Dann gilt:

i) $F(t)$ ist (nicht notw. *streng*) monoton wachsend, $0 \leq F(t) \leq 1$.

ii) $F(t+) = F(t)$ ("rechtsseitig Stetig")

iii) $F(t-) = F(t) - \mathbb{P}(\{t\})$

iv) $\lim_{t \to -\infty} F(t) = 0$, $\lim_{t \to \infty} F(t) = 1$

Bemerkung: die Limiten in b), c), d) existieren wegen a).

Beweis:

i) Monotonieeigenschaften von $\mathbb{P}$, siehe 1.2.2, 4.

ii) Sei $t_n \downarrow t$ $(t_n > t)$. Zerlege $(*)$ $(t, t_n] \overset{disj.}{=} \bigcup_{i=n}^{\infty} (t_{i+1}, t_i]$. Dann ist

$$F(t+) = \lim_{n \to \infty} F(t_n) = \lim_{n \to \infty} [F(t) + \mathbb{P}(t, t_n]] \overset{(*)}{=} \lim_{n \to \infty} [F(t) + \sum_{i=n}^{\infty} \mathbb{P}(t_{i+1}, t_i]] = F(t) + 0$$

da die Reihe $\sum_{i=1}^{\infty} \mathbb{P}(t_{i+1}, t_i] = \mathbb{P}(t, t_1] < \infty$ konvergiert. ✓

iii) Sei $t_n \uparrow t$ $(t_n < t)$. Zerlege $(*)$ $(t_1, t_n] \overset{disj.}{=} \bigcup_{i=1}^{n-1} (t_i, t_{i+1}]$. Dann ist

$$F(t-) = \lim_{n \to \infty} F(t_n) = \lim_{n \to \infty} [F(t_1) + \mathbb{P}(t_1, t_n]] \overset{(*)}{=} \lim_{n \to \infty} [F(t_1) + \sum_{i=1}^{n-1} \mathbb{P}(t_i, t_{i+1}]] =$$

$$F(t_1) + \mathbb{P}(\bigcup_{i=1}^{\infty} (t_i, t_{i+1}]) = F(t_1) + \mathbb{P}(t_1, t) = \mathbb{P}(-\infty, t) = F(t) - \mathbb{P}(\{t\}) \quad ✓$$

iv) Analog zu b) $(t_n = -n)$, bzw. c) $(t_n = n)$. ✓

Notation: Im Folgenden bezeichne $< a, b >$ für $-\infty \leq a < b \leq \infty$ eines der Intervalle $[a, b]$, (a, b), $[a, b)$, $(a, b]$. Wobei im Fall $a = -\infty$ nur $(-\infty, b >$ und im Fall $b = \infty$ nur $< a, \infty)$ zugelassen wird.

4.2.3 Formeln für $\mathbb{P} < a, b >$

Sei F Verteilungsfunktion von $\mathbb{P}$; Alles Folgerungen aus 4.2.1, 4.2.2.

- $\mathbb{P}(a, b] = F(b) - F(a)$, insbesondere $\mathbb{P}(-\infty, b] = F(b)$.
- $\mathbb{P}(a, b) = F(b-) - F(a)$, insbesondere $\mathbb{P}(a, \infty) = 1 - F(a)$.
- $\mathbb{P}[a, b] = F(b) - F(a-)$.
- $\mathbb{P}[a, b) = F(b-) - F(a-)$.

Bemerkungen:

1. Falls F bei a stetig und F auf dem Intervall $[a, b]$ konstant ist, so ist $\mathbb{P}[a, b] = 0$.
2. Zusammen mit 4.1.5 folgt, daß $\mathbb{P}$ durch Vorgabe einer Verteilungsfunktion (d.i. eine Funktion $F(t)$, $t \in \mathbb{R}$, mit den Eigenschaften a), b), d) aus 4.2.2) eindeutig festgelegt wird, wenn man setzt $\mathbb{P}(a, b] = F(b) - F(a)$.[3]

Im Fall der Exponetialvert. aus 4.1.7, bei der $F(t) = \left\{ \begin{array}{ll} 0 & , t < 0 \\ 1 - e^{-\lambda t} & , t \geq 0 \end{array} \right\}$ ist, stellt man fest, daß $F' = f$ bzw. $F(t) = \int_{-\infty}^{t} f(x)dx$, mit $f(x) = \left\{ \begin{array}{ll} 0 & , x < 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{array} \right\}$.

4.2.4 Wahrscheinlichkeitsdichte (Definition)

Sei $F_{\mathbb{P}}(t)$, $t \in \mathbb{R}$, Verteilungsfunktion von $\mathbb{P}$. Existiert dann eine 'meßbare' Funktion $f_{\mathbb{P}} : \mathbb{R} \to [0, \infty]$ mit $F_{\mathbb{P}}(t) = \int_{-\infty}^{\infty} f_{\mathbb{P}}(*)dx$ für alle $t \in \mathbb{R}$, so heißt $f_{\mathbb{P}}(x)$, $x \in \mathbb{R}$, *Wahrscheinlichkeitsdichte* oder kurz *Dichte* von $\mathbb{P}$.

Bemerkungen:

1. Das Intergral $\int_{-\infty}^{\infty}$ läßt sich als uneigentliches Riemann-Integral oder als Lebesgue-Integral auffassen. Wegen des Begriffs "meßbar" siehe 4.3.1 unten.
2. Ist die stetige Verteilungsfunktion F auf $\mathbb{R} \setminus D$ (D leer oder endlich) stetig diefferenzierbar, so besitzt F die Dichte $f(x) = F'(x)$, $x \in \mathbb{R} \setminus D$, [$f(x)$ auf D beliebig festgelegt].
3. Besitzt F eine Dichte, so ist $F(t)$, $t \in \mathbb{R}$ stetig [d.h. $F(t) = F(t+) = F(t-1)$] und 4.2.3 liefert für alle 4 Intervalltypen die Formel $\mathbb{P} < a, b >= F(b) - F(a)$.

4.2.5 Satz

Besitzt eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B})$ eine Dichte $f = f_{\mathbb{P}}$, so gilt:

$$\mathbb{P} < a, b >= \int_a^b f(x)dx \tag{4.1}$$

Inbesondere gilt:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{4.2}$$

Beweis: folgt direkt aus Bemerkung 3. oben.

Bemerkungen:

1. Wir können also eine Dichte f durch die Eigenschaft $f : \mathbb{R} \to [0, \infty)$, f integrierbar mit $\int_{-\infty}^{\infty} f(x)dx = 1$ festlegen.
2. Durch Vorgabe einer Dichte f ist eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B})$ eindeutig festgelegt[4].
3. Der Begriff der Dichte spielt im Fall $\Omega = \mathbb{R}$ die gleiche Rolle wie der Begriff der Wahrscheinlichkeitsfunktion im Falle eines abzählbaren Ω (nur: eine Dichte braucht nicht notwendigerweise zu existieren!).

[3] Eigenschaften i), ii) aus 4.1.5 erfüllt (!)

[4] vermöge (4.1) und Beweis 2. in 4.2.3

4.2.6 Gleichverteilung (Beispiel)

Gleichverteilung auf dem Intervall $\Omega = [A, B] \subset \mathbb{R}$, $A < B$ (vgl. 4.1.6)

$$\text{Dichte:} \quad f(x) = \begin{cases} 0 & , x \notin [A, B] \\ \frac{1}{B-A} & , x \in [A, B] \end{cases}$$

$$\text{Verteilungsfunktion:} \quad F(x) = \begin{cases} 0 & , x < A \\ \frac{x-A}{B-A} & , A \leq x \leq B \\ 1 & , x > B \end{cases}$$

4.2.7 Exponentialverteilung (Beispiel)

Exponentialverteilung mit Parametern $\lambda > 0$ (vgl. 4.1.7)

$$\text{Dichte:} \quad f(x) = \begin{cases} 0 & , x > 0 \\ \lambda e^{-\lambda x} & , x \geq 0 \end{cases}$$

$$\text{Verteilungsfunktion:} \quad F(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-\lambda x} & , x \geq 0 \end{cases}$$

Verwendung: Wartezeit (bis zum Eintreten eines Ereignisses).

4.2.8 Diskrete Verteilung (Beispiel)

Diskrete Verteilung auf $\{x_1, x_2, \ldots\} \subset \mathbb{R}$ (oder $\{x_1, \ldots, x_n\} \subset \mathbb{R}$) mit vorgegebener Wahrscheinlichkeitsfunktion $\mathbb{P}\{x_i\}$. Setze für $A \in \mathfrak{B}^1$:

$$\mathbb{P}(A) = \sum_{i: x_i \leq x} \mathbb{P}(\{x_i\})$$

$\mathbb{P}$ bildet ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathfrak{B}^1)$ (!), mit Verteilungsfunktion:

$$F(x) = \sum_{i: x_i \leq x} \mathbb{P}(\{x_i\})$$

Es existiert jedoch keine Dichte!

4.2.9 Normalverteilung (Beispiel)

Normalverteilung mit Parametern μ und σ^2, $\mu \in \mathbb{R}$, $\sigma^2 > 0$:

$$\text{Dichte:} \quad \varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\text{Verteilungsfunktion:} \quad \phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(t) dt$$

Abkürzung: $N(\mu, \sigma^2)$

Verwendung: Symmetrisch um einen 'wahren' Wert μ streuende Meßgröße.

Spezialfall: $N(0, 1)$ 'Standard-Normalverteilung' (vgl. 3.4), man schreibt $\varphi = \varphi_{0,1}$, $\phi = \phi_{0,1}$.

Umrechnung: $\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$, $\phi_{\mu,\sigma^2}(x) = \phi(\frac{x-\mu}{\sigma})$ (Substitutionsregel).
Aus dieser Beziehung folgt:

$$\int_{-\infty}^{\infty} \varphi_{\mu,\sigma^2}(x)dx = \phi_{\mu,\sigma^2}(\infty) = \phi(\infty) = \int_{-\infty}^{\infty} \varphi(x)dx \overset{3.4.1}{=} 1$$

so daß φ_{μ,σ^2} eine Dichte ist.
Das Konzept der Dichte läßt sich auch im Fall $\Omega = \mathbb{R}^k$ verwirklichen. Eine Dichte im $\mathbb{R}^k$ ist eine nicht negative (aber meßbare) Funktion $f(x) = f(x_1, \dots, x_k)$, $x \in \mathbb{R}^k$ mit (Integrierbarkeit vorausgesetzt):

$$\int \underset{\mathbb{R}^k}{\dots} \int f(x_1, \dots, x_k)dx_1 \dots dx_k = 1 \tag{4.3}$$

Wobei $\int \underset{\mathbb{R}^k}{\dots} \int = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{k-mal}$.
Für ein $B \in \mathfrak{B}^k$ definiert man

$$\int \underset{B}{\dots} \int f(x_1, \dots, x_k)dx_1 \dots dx_k = \int \underset{\mathbb{R}^k}{\dots} \int f(x)dx_1 \dots dx_k =$$

$$= \int \underset{\mathbb{R}^k}{\dots} \int 1_B(x)f(x)dx_1 \dots dx_k$$

Wir benötigen den folgenden Satz aus der Integrationstheorie.

4.2.10 Satz

Ist $f \geq 0$ eine integrierbare Funktion auf dem $\mathbb{R}^k$, so wird durch $B \mapsto \int \underset{B}{\dots} \int f(x)dx_1 \dots dx_k$ $(B \in \mathfrak{B}^k)$ eine σ-additive Abbildung von $\mathfrak{B}^k$ in $[0, \infty)$ definiert. [d.h. für paarweise disjunkte $B_1, \dots, B_k \in \mathfrak{B}^k$ gilt:

$$\int \underset{\cup B_i}{\dots} \int f(x)dx_1 \dots dx_k = \sum_{i=1}^{\infty} \int \underset{B_i}{\dots} \int f(x)dx_1 \dots dx_k$$

Beweis: über den Satz von der Monotonen Konvergenz aus Analysis III.

4.2.11 Satz

Sei $f(x)$, $x \in \mathbb{R}^k$ eine Dichte und $(a, b]$ ein n-dimensionales Intervall (vgl. 4.1) $(a, b] = \bigotimes_{i=1}^k (a_i, b_i]$.

a) Setzt man

$$\mathbb{P}(a, b] = \int \underset{(a,b]}{\dots} \int f(x)dx_1 \dots dx_k \tag{4.4}$$

so wird eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\mathbb{R}^k, \mathfrak{B}^k)$ eindeutig festgelegt. (Anstelle von $(a, b]$ läßt sich in (4.4) auch jeder andere Intervalltyp $< a, b >= \bigotimes_{i=1}^k < a_i, b_i >$ einsetzen.)

b) Für die Wahrscheinlichkeitsverteilung $\mathbb{P}$ aus a) gilt, allgemeiner als (4.4):

$$\mathbb{P}(B) = \int \underset{B}{\dots} \int f(x)dx_1 \dots dx_k \quad (B \in \mathfrak{B}^k) \tag{4.5}$$

Beweis:

a) Durch (4.4) wird eine Abbildung $\mathbb{P} : F^k \to [0,1]$ definiert, die wegen (4.3) normiert ist und aufgrund von Satz 4.2.11 σ-additiv auf F^k. Vermöge Satz 4.1.5 hat sie eine eindeutige Fortsetzung auf $\mathfrak{B}^k$.

b) folgt dann aus Satz 4.2.10 und der Eindeutigkeitsaussage von a)

4.2.12 Beispiel

k-dimensionale Normaleverteilung mit Parameter $\mu \in \mathbb{R}^k$ und $\sum$ (Symmetrische $k \times K$-Matrix, positiv definit), kurz $N(\mu, \sum)$-Verteilung.

(4.6) Dichte: $$f(x) = \frac{1}{\sqrt{(s\pi)^k \det(\sum)}} e^{-\frac{1}{2}(x-\mu)^T \cdot \sum^{-1} \cdot (x-\mu)}$$

$x \in \mathbb{R}^k$, Abkürzung: $(N_k(\mu, \sum))$.

Spezialfall: $N_k(0, I_k)$ (k-dimensionale Standard-Normalverteilung).
Im Fall $\mu = 0 \in \mathbb{R}^k$ und $\sum = I_k$ (k-dimensionale Einheitsmatrix) reduziert sich (4.6) auf:

(4.7) $$f(x) = \frac{1}{\sqrt{(s\pi)^k}} e^{-\frac{1}{2}(x_1^2 + \dots x_k^2)} = \prod_{i=1}^{k} \frac{1 e^{-\frac{1}{2}x}}{\sqrt{(s\pi)}}$$

$x = (x_1, \dots, x_k) \in \mathbb{R}^k$.
Aus 3.4.1 folgt, daß $f(x)$ in (4.7) die Normierungseigenschaft (4.3) hat.
In 4.3.9 werden wir (4.3) auch für $f(x)$ in (4.6) nachweisen können.

4.3 Zufallsvariablen, Zufallsvektoren

In 1.5 hatten wir jede Abbildung: $\Omega \to \Omega'$ als Zufallsgröße bezeichnet: $(\Omega, \mathfrak{A}, \mathbb{P}) \xrightarrow{x} (\Omega', \mathfrak{A}', \mathbb{P}_X)$. Jetzt müssen wir sicherstellen, daß die Urbilder $X^{-1}(A')$, $A' \in \mathfrak{A}'$ auch Element von $\mathfrak{A}$ sind.

4.3.1 Zufallsgröße (Definition)

a) Sind $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$ meßbare Räume, so heißt eine Abbildung $X : \Omega \to \Omega'$ *Zufallsgröße* (auf $(\Omega, \mathfrak{A})$, mit Werten in Ω'), falls:

   (4.8) $$X^{-1}(A') \in \mathfrak{A} \forall A' \in \mathfrak{A}'$$

b) Ist $X : \Omega \to \Omega'$ Zufallsgröße und $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung auf $(\Omega, \mathfrak{A})$, so heißt $\mathbb{P}_X \mathfrak{A}' \to [0,1]$ mit

   (4.9) $$\mathbb{P}_X(A') = \mathbb{P}(X^{-1}(A')) \ , \quad A' \in \mathfrak{A}'$$

   Verteilung von X.

Bemerkungen:

1. Genau wie in 1.5.2 zeigt man, daß $\mathbb{P}_X$ eine Wahrscheinlichkeitsverteilung auf $(\Omega', \mathfrak{A}')$ ist.
2. In der Maßtheorie nennt man eine Abbildung X mit der Eigenschaft (4.8) *meßbar* bezüglich $\mathfrak{A}, \mathfrak{A}'$. (Eine meßbare Funktion $f : \mathbb{R}^k \to \mathbb{R}$ ist also meßbar bezüglich $\mathfrak{B}^k, \mathfrak{B}^1$
3. Im Fall $\Omega' = \mathbb{R}^k$, $\mathfrak{A}' = \mathfrak{B}^k$ spricht man von einem *k-dimensionalen Zufallsvektor*, im Fall $k = 1$ von einer *Zufallsvariablen.*
4. Es gibt *nichtmeßbare* Funktionen $f : \mathbb{R} \to \mathbb{R}$. Ist nämlich $C \subset \mathbb{R}^1$ nicht borelsch[5], so ist $f = 1_C$ nicht meßbar.

Zum Nachweis von (4.8) brauchen wir nur Mengen $A' \in F'$, F' Erzeugendensystem von $\mathfrak{A}'$.

[5] siehe 4.1.4, c)

4.3.2 Satz

Seien $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$ meßbare Räume, $F' \subset \mathfrak{A}'$ sei Erzeugendensystem von $\mathfrak{A}'$ (d.h. $\sigma(F') = \mathfrak{A}'$). Die Abbildung $X : \Omega \to \Omega'$ ist genau dann Zufallsgröße (i.e. meßbar), wenn

$$X^{-1}(A') \in \mathfrak{A}\ ,\ \ \forall A' \in F' \tag{4.10}$$

Beweis: Aus 1) folgt 3) (triv.). Sei nun 3) erfüllt. Setze

$$\varphi' = \{A' \in \mathfrak{A}' : X^{-1}(A') \in \mathfrak{A}\}$$

man zeigt, da" φ' eine σ-Algebra ist (!). Aus $F' \overset{(4.10)}{\subset} \varphi' \subset \mathfrak{A}'$ folgt:

$$\underbrace{\underbrace{\sigma(F')}_{F=Erz.Sys.}}_{=\mathfrak{A}'} \subset \underbrace{\sigma(\varphi')}_{=\varphi'} \subset \underbrace{\sigma(\mathfrak{A}')}_{=\mathfrak{A}'} \implies \varphi' = \mathfrak{A}'$$

4.3.3 Korollar

Sei $(\Omega, \mathfrak{A})$ ein meßbarer Raum. $X : \Omega \to \mathbb{R}$ ist Zufallsvariable genau dann, wenn:

$$\{X < b\} \equiv X^{-1}(-\infty, b] \in \mathfrak{A} \forall b \in \mathbb{R}$$

[äquivalent: $\geq, >, \leq$ statt $<$] Insbesondere ist jede stetige (stückweise stetige) Abbildung $X : \mathbb{R}^k \to \mathbb{R}$ Zufallsvariable auf $(\mathbb{R}^k, \mathfrak{B}^k)$.

Beweis: (Nur für "$<$")
Setze $F' = \{(-\infty, b] : b \in \mathbb{R}\}$. Man zeigt, daß $\sigma(F') = \mathfrak{B}^1$ (!), so daß Satz 4.3.2 anwendbar ist.
Für ein stetiges $X : \mathbb{R}^k \to \mathbb{R}$ ist $X^{-1}(-\infty, b)$ offene Menge, ist in $\mathbb{R}^k$, also aus $\mathfrak{B}^k$. ✓

4.3.4 Satz

Sei $X = (X_1, \dots, X_k)$ eine Abbildung: $\Omega \to \mathbb{R}^k$, und $(\Omega, \mathfrak{A})$ meßbarer Raum. Dann ist X ein Zufallsvektor genau dann, wenn jedes X_i eine Zufallsvariable ist $(i = 1, \dots, k)$.

Beweis: Wie in 1.5.9 gilt:

$$X_i^{-1}(a, b] = X^{-1}(\mathbb{R} \times \ldots \times \mathbb{R} \underbrace{(a, b]}_{i-te\ Stelle} \times \mathbb{R} \times \ldots \times \mathbb{R})$$

woraus die Behauptung folgt.

4.3.5 Satz

Sind $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$, $(\Omega'', \mathfrak{A}'')$ meßbare Räume und $X : \Omega \to \Omega'$, $Y : \Omega' \to \Omega''$ Zufallsgrößen, so ist auch $Y \circ X : \Omega \to \Omega''$ eine Zufallsgröße (Beweis klar).

4.3.6 Sprechweise

Die in 4.2 eingeführten Notationen "F ist Verteilungfunktion von $\mathbb{P}$" (in 4.2.1) und "f ist Dichte von $\mathbb{P}$" (in 4.2.10) wird auch die Verteilung $\mathbb{P}_X$ von X angewandt:
Man sagt dann "F ist Verteilungsfunktion von X" (d.h. $F_X(x) = \mathbb{P}(X \in x) = \mathbb{P}_X(-\infty, x]$ für eine Zufallsvariable von X) und "f_x ist Dichte von X" (aber X hat Dichte f).

Beispiel: ist die Zufallsvariable X eine Wartezeit und $\mathbb{P}_X$ eine Exponentialverteilung (mit $\lambda > 0$, so hat X nach 4.2.7 die

$$\text{Verteilungsfunktion:} \quad F_X(x) = \left\{ \begin{array}{cc} 0 & x < 0 \\ 1 - e^{-\lambda x} & <\geq 0 \end{array} \right\}$$

$$\text{bzw. Dichte:} \quad f_X(x) = \left\{ \begin{array}{cc} 0 & x < 0 \\ \lambda e^{-\lambda x} & <\geq 0 \end{array} \right\}$$

Hat der Zufallsvektor $X = (X_1, \ldots, X_k)$ die Dichte $f(x) = f_X(x)$, $x \in \mathbb{R}^k$, so gilt gemäß 4.2.11 für ein k-dimensionales Intervall $(a, b] = \bigotimes_{i=1}^k (a_i, b_i]$:

$$\mathbb{P}_X(a, b] = \int_{a_k}^{b_k} \ldots \int_{a_1}^{b_1} f(x_1, \ldots, x_k) dx_1 \ldots dx_k \tag{4.11}$$

Hat X die Dichte $f(x)$, $x \in \mathbb{R}^k$, so hat die Komponente X_i die Randdichte:

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k) dx_1 \ldots dx_{i-1} dx_{i+1} dx_k \tag{4.12}$$

Derfolgende Satz gibt die Dichte von $\varphi \circ X$ an, wenn die Dichte von X gegeben ist.

4.3.7 Transformationssatz für Dichten

Der k-dimensionale Zufallsvektor X besitzt die Dichte $f(x)$, $x \in \mathbb{R}^k$, wobei für eine offene Menge $U \subset \mathbb{R}^k$ gilt: $f(x) = 0$ für $X \notin U$. Sei $\varphi : U \to V$, $V \subset \mathbb{R}^k$, eine bijektive Abbildung mit φ, φ^{-1} stetig differenzierbar.
Dann hat der k-dimensionale Zufallsvektor $Y = \varphi \circ X$ eine Dichte, und es gilt:

$$\begin{array}{ll} g(y) = 0 & y \notin V \\ g(y) = f(\varphi^{-1}(y)) \cdot |\det(\frac{d\varphi^{-1}}{dy}(y))| & y \in V \end{array} \tag{4.13}$$

wobei $(\frac{d\varphi^{-1}}{dy}(y)) = (\frac{\partial \varphi_i^{-1}}{\partial y_j}(y)\ ,\ \ i, j = 1, \ldots, k)$ die $k \times k$ Funktionsmatrix von φ^{-1} ist.

Bemerkung: Zur Festlegung der Verteilung (und damit der Dichte) von $Y = \varphi \circ X$ genügt es, φ allein auf U festzulegen. Sind nämlich φ und $\tilde{\varphi} : \mathbb{R}^k \to \mathbb{R}^k$ mit $\varphi|U = \tilde{\varphi}|U$, so gilt $\mathbb{P}_{\tilde{\varphi} \circ X} = \mathbb{P}_{\varphi \circ X}$. In der Tat, sei $B \in \mathfrak{B}^k$, dann

$$\mathbb{P}_{\tilde{\varphi} \circ X}(B) = \mathbb{P}(X \in \tilde{\varphi}^{-1}(B)) \overset{(*)}{=} \mathbb{P}(X \in \tilde{\varphi}^{-1}(B) \cap U) =$$

$$= \mathbb{P}(X \in \varphi^{-1}(B) \cap U) = \mathbb{P}(X \in \varphi^{-1}(B)) = \mathbb{P}_{\varphi \circ X}(B)$$

Dabei gilt (*) wegen

$$\mathbb{P}(X \in \overline{U}) \overset{4.2.11}{=} \underset{\overline{U}}{\int \ldots \int} f(x) dx = 0$$

da $f(x) = 0$ für $x \notin U$.

Beweis: Sei $A \in \mathfrak{B}^k$ offen, dann gilt wegen (*) $[\mathbb{P}(X \in \overline{U}) = 0]$:

$$\mathbb{P}_{\varphi \circ X}(A) = \mathbb{P}(X \in \varphi^{-1}(A)) \overset{(*)}{=} \mathbb{P}(X \in \varphi^{-1}(A) \cap U) =$$

$$= \underset{\varphi^{-1}(A) \cup U}{\int \ldots \int} f(x) dx \overset{U = \varphi^{-1}(V)}{=} \underset{\varphi^{-1}(A \cap V)}{\int \ldots \int} f(x) dx =$$

$$\overset{Transf.satz}{=} \underset{(A\cap V)}{\int\ldots\int} f(\varphi^{-1}(x))\cdot|\det\frac{d\varphi^{-1}}{dy}(y)|dy \equiv \underset{A}{\int\ldots\int} g(y)dy$$

wobei wir den Transformationssatz für Integrale angewandt haben[6]. Speziell gilt für offenes $A = \bigotimes_{i=1}^k (a_i, b_i)$:

$$\mathbb{P}_{\varphi\circ X}(A) = \int_{a_k}^{b_k}\ldots\int_{a_1}^{b_1} g(y)dy$$

d.h. $g(y)$ ist Dichte von $\varphi\circ X$ nach Satz 4.2.11 a). ✓

4.3.8 Corollar

Besitzt der k-dimensionale Zufallsvektor X die Dichte $f(x)$, $x\in\mathbb{R}^k$, so lautet die Dichte $g(y)$, $y\in\mathbb{R}^k$ von $Y = A\cdot X + b$, (A invertierbare $k\times k$-Matrix, $b\in\mathbb{R}^k$):

$$g(y) = \frac{1}{|\det A|} f(A^{-1}(y-b)) \ , \quad y\in\mathbb{R}^k$$

Beweis: $\varphi(X) = A\cdot X + b$ ist auf $U = \mathbb{R}^k$ bijektiv, mit $\varphi^{-1}(y) = A^{-1}(y-b)$ und $\det(\frac{d\varphi^{-1}(y)}{dy} = \det(A^{-1}) = \frac{1}{\det(A)}$. ✓

4.3.9 Beispiel

k-dimensionale Normalverteilung (vgl. 4.2.12).

1. Ist X $N_k(0, I_k)$-verteilt (d.h. $f(x) = (\frac{1}{\sqrt{2\pi}})^k\cdot e^{-\frac{1}{2}x^Tx}$, $x = \begin{bmatrix} x_1\\ \vdots \\ x_k\end{bmatrix}$), so besitzt $Y = A\cdot X+\mu$ (A invertierbare $k\times k$-Matrix, $\mu\in\mathbb{R}^k$) die Dichte

$$g(y) = \frac{1}{|\det(A)|}(\frac{1}{\sqrt{2\pi}})^k e^{-\frac{1}{2}(y-\mu)^T(A^{-1})^T\cdot A^{-1}(y-\mu)} =$$

$$= \frac{1}{\sqrt{(2\pi)^k\det(\sum)}} e^{-\frac{1}{2}(y-\mu)^T\cdot\sum^{-1}\cdot(y-\mu)}$$

   mit $\sum = A\cdot A^T$. (Dann ist $\sum^{-1} = (A\cdot A^T)^{-1} = (A^{-1})\cdot A^{-1}$, $\sum$ symmetrisch, positiv definit, $\det(\sum) = (\det(A))^2$).
   Y ist also $N_k(\mu,\sum)$-verteilt.

2. Ist umgekehrt Y $N_k(\mu,\sum)$-verteilt ($\sum$ symm. pos. def.), so ist $X = (\sum^{-\frac{1}{2}})^T(y-\mu)$ $N_k(o, I_k)$-verteilt.
   Dabei ist $\sum^{-\frac{1}{2}}$ eine invertierbare $k\times k$-Matrix mit $\sum^{-1} = (\sum^{-\frac{1}{2}})\cdot(\sum^{-\frac{1}{2}})^T$ $[(\sum^{-\frac{1}{2}})^T = A$ aus 1.].

Bemerkung: Für symmetrische, positiv definite B existieren verschiedene 'Wurzeln' $B^{1/2}$ von B mit (+) $B = B^{1/2}\cdot(B^{1/2})^T$ (oben mit $B = \sum^{-1}$, $B^{1/2} = A^{-1}$):

1. symmetrische Wurzel, $B^{1/2}$ symmetrisch, positiv definit
2. Cholesky Wurzel, $B^{1/2}$ obere Dreiecksmatrix (Numerische Mathematik)

In jedem Fall ist $\det(B^{1/2}) = \sqrt{\det B}$ und (+).

[6] Forster III, S.120 (dort ϕ anstatt φ^{-1})

4.4 Unabhängige Zufallsvariablen

Analog zu 2.3.4 definiert man

4.4.1 Definition

a) Die auf $(\Omega, \mathfrak{A}, \mathbb{P})$ definierten Zufallsvariablen $X_1, \ldots, X_n$ heißen *unabhängig*, falls für alle $B_1, \ldots, B_n \in \mathfrak{B}^1$ gilt:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_n \in B_n) \tag{4.14}$$

b) Abzählbar viele Zufallsvariablen $X_1, X_2, \ldots$ heißen *unabhängig*, falls je endlich viele $X_{i_1}, \ldots, X_{i_n}$ unabhängig sind.

4.4.2 Satz

Die auf $(\Omega, \mathfrak{A}, \mathbb{P})$ definierten Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig genau dann, wenn:

$$\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{(X_1)} \times \mathbb{P}_{(X_2)} \times \ldots \times \mathbb{P}_{(X_n)}$$

Beweis:

- Gelte $\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{(X_1)} \times \ldots \times \mathbb{P}_{(X_n)}$:
  Seien $B_1, \ldots, B_n \in \mathfrak{B}^1$, dann gilt:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}_{(X_1, \ldots, X_n)}(B_1 \times \ldots \times B_n) =$$
$$= \mathbb{P}_{X_1} \times \ldots \times \mathbb{P}_{X_n}(B_1 \times \ldots \times B_n) = \mathbb{P}_{X_1}(B_1) \cdot \ldots \cdot \mathbb{P}_{X_n}(B_n) =$$
$$= \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_n \in B_n)$$

- Gelte $X_1, \ldots, X_n$ unabhängig:

$$\mathbb{P}_{(X_1, \ldots, X_n)}(B_1 \times \ldots \times B_n) = \mathbb{P}(X_1 \subset B_1) \cdot \ldots \cdot \mathbb{P}(X_n \subset B_n) =$$
$$= \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_n \in B_n) = \mathbb{P}_{X_1}(B_1) \cdot \ldots \cdot \mathbb{P}_{X_n}(B_n)$$

Bemerkung: Da die σ-Algebra $\mathfrak{B}^n$ von dem System der intervalle $(a, b] = \bigotimes_1^n (a_i, b_i]$ erzeugt werden, genügte es, statt (4.14) für alle $a, b \in \mathbb{R}^n$, $a < b$ zu finden:

$$\begin{aligned} &\mathbb{P}(a_1 \subset X_1 \leq b_1, \ldots, a_n \subset X_n \leq b_n) = \\ &= \mathbb{P}(a_1 \subset X_1 \leq b_1) \cdot \ldots \cdot \mathbb{P}(a_n \subset X_n \leq b_n) \end{aligned} \tag{4.15}$$

Auch die Intervalltypen $[), [], ()$ können anstelle von $(]$ verwendet werden.

4.4.3 Satz

Die Zufallsvariablen $X_1, \ldots, X_n$ mögen die Dichten $f_1, \ldots, f_n$ besitzen.
Dann gilt:

$$X_1, \ldots, X_n \text{ unabhängig}$$
$$\Updownarrow$$
$$(X_1, \ldots, X_n) \text{ hat Dichte } f(x_1, \ldots, x_n) = f(x_1) \cdot \ldots \cdot f(x_n)$$

Beweis:

"$\Rightarrow$" Sind $X_1, \dots, X_n$ unabhängig, dann folgt:

$$\mathbb{P}_{(X_1,\dots,X_n)}((a_1, b_1] \times \dots \times (a_n, b_n]) = \mathbb{P}_{X_1}(a_1, b_1] \cdot \dots \cdot \mathbb{P}_{X_n}(a_n, b_n] =$$
$$= \int_{a_1}^{b_1} f_1(x_1)dx_1 \cdot \dots \cdot \int_{a_n}^{b_n} f_n(x_n)dx_n =$$
$$= \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_1(x_1) \cdot \dots \cdot f_n(x_n)dx_1 \cdot \dots \cdot dx_n$$
$$\Rightarrow (X_1, \dots, X_n) \text{ hat Dichte } f(x_1, \dots, x_n)\ ,\ (x_1, \dots, x_n) \in \mathbb{R}^n.$$

"$\Leftarrow$" Analog.

4.4.4 Beispiel

$X = (X_1, \dots, X_n)$ ist $\mathfrak{N}(0, I_k)$-verteilt genau dann, wenn die $X_1, \dots, X_n$ unabhängig und $\mathfrak{N}(0,1)$-verteilt sind.

Beweis: Sind $X_1, \dots, X_n$ unabhängig mit den Dichten $f_{x_i}(x) = f_i(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, so hat $(X_1, \dots, X_n)$ gemäß 4.4.3 die Dichte

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f_i(x_i) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)} = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x^T \cdot x)}$$

mit $x = (x_1, \dots, x_n)$. Umgekehrt folgt: X hat die Dichte

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x^T \cdot x)} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x_i^2)}$$

Man stellt fest, durch Integration $\int_{-\infty}^{\infty}$ über die Komponenten $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, daß $f_i(x_i)$ die Dichte von X_i sein muß, so daß die Unabhängigkeit und $\mathfrak{N}(0,1)$-Verteilung der X_i folgt.

Faltungsformel für unabhängige X_1, X_2:

$$\mathbb{P}_{x_1+X_2}(\{x\}) = \sum_k \mathbb{P}_{X_1}(\{x_1\}) \cdot \mathbb{P}_{X_2}(\{x - x_1\})$$

4.4.5 Satz

Sind X_1, X_2 unabhängig Zufallsvariablen mit Dichten f_1, f_2, dann besitzt die Zufallsvariable $X_1 + X_2$ die Dichte

$$f(x) = \int_{-\infty}^{\infty} f_1(x_1) f_2(x - x_1) dx \tag{4.16}$$

Beweis: Für die Verteilungsfunktion $F(y) : \mathbb{P}(X_1 + X_2 \leq y)$ weisen wir $F(y) \int_{-\infty}^{y} f(x)dx$ nach. Es ist

$$\mathbb{P}(X_1 + X_2 \leq y) = \mathbb{P}_{(X_1, X_2)}(\{(x_1, x_2) : x_1 + x_2 \leq y\}) =$$
$$= \underset{\{(x_1, x_2) : x_1 + x_2 \leq y\}}{\int \dots \int} f_1(x_1) \cdot f_2(x_2) dx_2 dx_1 =$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y - x_1} f_1(x_1) \cdot f_2(x_2) dx_2 dx_1 =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_1(x_1) \cdot f_2(x_2 - x_1) dx_2 dx_1 =$$
$$= \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(x_2 - x) dx_1 dx_2 =$$
$$= \int_{-\infty}^{y} \underbrace{\int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(x - x_1) dx_1 \, dx_2}_{=Dichte} =$$

4.4.6 Definition

Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen, so heißt die Wahrscheinlichkeitsverteilung

$$\mathbb{P}_{X_1 + \ldots + X_n} \equiv \mathbb{P}_{X_1} * \ldots * \mathbb{P}_{X_n}$$

Faltung von $\mathbb{P}_{X_1 + \ldots + X_n}$ (mit "*" ist Faltungssysmbol).

4.4.7 Unabhängige Wartezeiten (Beispiel)

Sei S_n die Wartezeit zwischen dem $n-1$-ten und n-ten Ereignis. Die Zufallsvariable $T_n = S_1 + \ldots + S_n$ stellt die Wartezeit des n-ten Ereignisses dar. Unter den Voraussetzungen:

1. Die Zufallsvariablen $S_1, S_2, \ldots$ sind unabhängig.
2. Jedes S_i ist exponentialverteilt mit Parameter λ ("$\varepsilon(\lambda)$-verteilt").

wollen wir die Dichte der Zufallsvariable T_n berechnen. Es gilt (ÜA):

$$f_n(x) = \frac{\lambda^n \cdot x^{n-1}}{(n-1)!} \cdot e^{-\lambda x} \ , \quad x \geq 0 \ (f_n(x) = 0 \text{ f. } x < 0)$$

Die Wahrscheinlichkeitsverteilung mit der Dichte f_n heißt *Gamma*verteilung mit Parametern n und λ, kurz $\Gamma(n, \lambda)$ (I_n heißt dann $\Gamma(n, \lambda)$-verteilt).

$$\Gamma(n, \lambda) = \underbrace{\varepsilon(\lambda) * \ldots * \varepsilon(\lambda)}_{n-mal}$$

Zerlegt man einen Satz von Zufallsvariablen in disjunkte Gruppen und setzt auf die Gruppen Funktionen an, so erhalten wir unabhängig Zufallsvariablen.

4.4.8 Satz

$X_1, \ldots, X_n$ seien unabhängige Zufallsvariablen, für $m \leq n$ sei $\{1, \ldots, n\} = I_1 \cup \ldots \cup I_m$ eine Zerlegung der Indexmenge und φ_j Zufallsvariable auf $(\mathbb{R}^{k_j}, \mathfrak{B}^{k_j})$, $k_j = |I_j|$, $j = 1, \ldots, m$ $(\sum_{j=1}^{m} k_j = n)$. Bezeichet Y_j den k_j-dimensionalen Zufallsvektor $(X_i, \, i \in I)$, dann sind

$$\varphi_1 \cdot Y_1, \ldots, \varphi_m \cdot Y_m$$

unabhängige Zufallsvariablen.

Beweis: Ohne Einschränkung sei

$$I_1 = \{1, \ldots, k_1\} \ , \quad I_2 = \{k_1 + 1, \ldots, k_1 + k_2\} \ , \quad \ldots$$

i) Zunächst zeigen wir, daß die m Zufallsvektoren $Y_1, \ldots, Y_m$ unabhängig sind, im Sinne von

$$(*) \quad \mathbb{P}_{(Y_1, \ldots, Y_m)}(C_1 \times \ldots \times C_m) = \mathbb{P}_{Y_1}(C_1) \cdot \ldots \cdot \mathbb{P}_{Y_m}(C_m)$$

für alle $C_j \in \mathfrak{B}^{k_j}$, $j = 1, \ldots, m$.
Für die speziellen C_j der Form $C_j = B_1^{(j)} \times \ldots \times B_{k_j}^{(j)}$, $B_l^{(j)} \in \mathfrak{B}^1$ gilt wegen $(Y_1, \ldots, Y_m) = (X_1, \ldots, Y_n)$, $(C_1, \ldots, C_m) = (B_1^{(j)}, \ldots, B_{k_m}^{(j)})$:

$$\mathbb{P}_{(Y_1,\ldots,Y_m)}(C_1 \times \ldots \times C_m) = \mathbb{P}_{(X_1,\ldots,X_n)}(B_1^{(1)} \times \ldots \times B_{k_m}^{(m)}) =$$

$$\overset{Unabh.d.X_i}{=} \mathbb{P}_{X_1}(B_1^{(1)}) \cdot \ldots \cdot \mathbb{P}_{X_n}(B_{k_m}^{(m)}) =$$

$$= \bigotimes_{i \in I_1} \mathbb{P}_{X_i}(B_1^{(1)} \times \ldots \times B_{k_1}^{(1)}) \cdot \ldots \cdot \bigotimes_{i \in I_m} \mathbb{P}_{X_i}(B_1^{(m)} \times \ldots \times B_{k_m}^{(m)}) =$$

$$= \mathbb{P}_{Y_1}(C_1) \cdot \ldots \cdot \mathbb{P}_{Y_m}(C_m)$$

Nach Satz 4.1.5 gilt dann (*) auch für **alle** $C_j \in \mathfrak{B}^{k_j}$.

ii) Nun wird die Unabhängigkeit der $\varphi_1 \circ Y_1, \ldots, \varphi_m \circ Y_m$ gezeigt. Analog zu Satz 2.3.6 gilt:

$$\mathbb{P}_{(\varphi_1 \circ Y_1,\ldots,\varphi_m \circ Y_m)}(B_1 \times \ldots \times B_m) \overset{\text{(siehe 2.3.6)}}{=} \ldots = \ldots \underset{(*)}{=}$$

$$= \mathbb{P}_{\varphi_1 \circ Y_1}(B_1) \cdot \ldots \cdot \mathbb{P}_{\varphi_m \circ Y_m}(B_m) \quad \checkmark$$

4.5 Momente von Zufallsvariablen

Wir führen den Begriff des Erwartungswert $\mathbb{E}(X)$ einer Zufallsvariable X ein, indem wir uns an den entsprechenden Begriff für den diskreten Fall[7] durch eine Approximation von X (durch eine Folge diskreter Zufallsvariablen $X_{(n)}$) anhängen.

4.5.1 Definition

Für eine beliebige Zufallsvariable X auf $(\Omega, \mathfrak{A})$ definiert man für jedes $n \in \mathbb{N}$ die Zufallsvariable (n-te Approximierte):

$$X_{(n)}(\omega) = \sum_{k=-\infty}^{\infty} \frac{k}{n} 1_{A_{k,n}}(\omega) \;,$$

$$A_{k,n} = \{\omega : \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}\}$$

$$\text{d.h.} \quad \begin{cases} X_n(\omega)\frac{k}{n} & \text{für } \omega \in A_{k,n} \\ 0 & \text{sonst} \end{cases}$$

Es ist $A_{k,n} \subset \mathfrak{A}$, so daß $X_{(n)}$ eine Zufallsvariable ist, und zwar mit höchstens abzählbar vielen Werten ($\pm\frac{k}{n}$, $k \in \mathbb{Z}$).
Gemäß 3.1.1 setzen wir für die diskrete Zufallsvariable $X_{(n)}$:

$$\mathbb{E}(X_{(n)}) = \sum_{k=-\infty}^{\infty} \mathbb{P}(A_{k,n}) \tag{4.17}$$

[mit $\mathbb{P}(A_{k,n}) = \mathbb{P}_{X_{(n)}}\{\frac{k}{n}\}$], sofern

$$\sum_{k=-\infty}^{\infty} \frac{|k|}{n} \mathbb{P}(A_{k,n}) \equiv \mathbb{E}(|X_{(n)}|) < \infty \;.$$

[7] vgl. 3.1.1

4.5.2 Eigenschaften von $X_{(n)}, \mathbb{E}(X_{(n)})$

a) $X_{(n)} \leq X \leq X_{(n)} + \frac{1}{n}$, insbesondere $|X - X_{(n)}| \leq \frac{1}{n}$

b) $|X_{(n)} - X_{(m)}| \leq \frac{1}{n} + \frac{1}{m}$, denn $|X_{(n)} - X_{(m)}| \leq |X_{(n)} - X| + |X - X_{(m)}|$ und a)

c) $\mathbb{E}|X_{(n)} - X_{(m)}| \leq \frac{1}{n} + \frac{1}{m}$, aus b) und Eigenschaften von $\mathbb{E}$

d) Existiert $\mathbb{E}(X_{(n)})$ für $n \in \mathbb{N}$, so existiert auch $\mathbb{E}(X_{(m)})$ für alle $m \geq n$, denn

$$\mathbb{E}(X_{(m)}) \leq \underbrace{\mathbb{E}|X_{(m)} - X_{(n)}|}_{\frac{1}{n}+\frac{1}{m}} + \underbrace{\mathbb{E}(X_{(m)})}_{<\infty} < \infty$$

e) Existiert $\mathbb{E}(X_{(n)})$ für (mindestens) ein $n \in \mathbb{N}$, so bildet $\mathbb{E}(X_{(n)})$, $n \geq n_0$, eine Cauchy-Folge, denn:

$$|\mathbb{E}(X_{(n)}) - \mathbb{E}(X_{(m)})| \stackrel{linear.}{=} |\mathbb{E}(X_{(n)} - X_{(m)})| \stackrel{Monot.}{\leq}$$
$$\leq \mathbb{E}|X_{(n)} - X_{(m)}| \stackrel{c)}{\leq} \frac{1}{n} + \frac{1}{m} \to 0\ , \quad n, m \to \infty$$

4.5.3 Definition

Falls für (mindestens) ein $n \in \mathbb{N}$ der Erwartungswert $\mathbb{E}(X_{(n)})$ der n-ten Approximierten für X existiert, so setzt man $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_{(n)})$ (Existenz nach e) gesichert) und man sagt: $\mathbb{E}(X)$ existiert oder X besitzt einen Erwartungswert. Man schreibt auch: $\mathbb{E}(X) = \int_a X(\omega)\mathbb{P}(d\omega)$.

Bemerkung: Dieser "$\mathbb{P}$-Integral von X" ist vom Typ "Lebesgue-Stieltjes" (Intervall-einteilung auf der y-Achse), im Unterschied zum Riemann-Integral (Einteilung auf der x-Achse).

4.5.4 Eigenschaften von $\mathbb{E}(X)$

a) $\mathbb{E}(X)$ existiert genau dann, wenn $\mathbb{E}|X|$ existiert (d.h. $\mathbb{E}|X| < \infty$)

b) Ist $X(\Omega)$ abzählbar, so ist (in Übereinstimmung mit 3.1.1) $\mathbb{E}(X) = \sum_{x\in X(\Omega)} x \cdot \mathbb{P}_X\{x\}$, falls diese Reihe absolut konvergiert.

Beweis:

a) Mehrfache Anwendung von 4.5.2 a) liefert $|X_{(n)}| \leq |X| + |X - X_{(n)}| \leq |X| + \frac{1}{n} \leq |X|_{(n)} + \frac{2}{n}$ und $|X|_{(n)} \leq \ldots \leq |X_{(n)}| + \frac{1}{n}$, woraus a) folgt. ✓

b) Setze $I_{k,n} = (\frac{k}{n}, \frac{k+1}{n}]$. Wegen $\mathbb{P}(X_{(n)} = \frac{k}{n}) = \sum_{x\in I_{k,n}} \mathbb{P}(X = x)$ ist:

$$(*)\left\{\begin{array}{c} \mathbb{E}(X_{(n)}) = \sum_{k=-\infty}^{\infty} \frac{k}{n} \sum_{x\in I_{k,n}} \mathbb{P}(X = x) \leq \\ \leq \sum_{k=-\infty}^{\infty} \sum_{x\in I_{k,n}} x\mathbb{P}(X = x) = \\ = \sum_{x\in X(\Omega)} x\mathbb{P}(X = x) \leq \\ \leq \sum_{k=-\infty}^{\infty} \frac{k+1}{n} \sum_{x\in I_{k,n}} \mathbb{P}(X = x) = \mathbb{E}(X_{(n)}) + \frac{1}{n} \end{array}\right.$$

Falls die Reihe $\sum x\mathbb{P}(X = x)$ absolut konvergiert, so wegen $\mathbb{E}(X_{(n)}) \leq \sum |x|\mathbb{P}(X = x) + \frac{1}{n} < \infty$ (ähnliche Abschätzung wie bei (*)) auch die Reihe $\mathbb{E}(X_{(n)})$, so daß $\lim_{n\to\infty}$ in (*) die Behauptung liefert. ✓

Im speziellen Fall, daß X eine Dichte besitzt, berechnet sich $\mathbb{E}(X)$ wie folgt.

4.5.5 Satz

Besitzt die Zufallsvariable X eine Dichte $f(x)$, $x \in \mathbb{R}$, so ist

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx\ , \quad \text{sofern} \int_{-\infty}^{\infty} |x|f(x)dx < \infty$$

Beweis: Wegen $\mathbb{P}(X_{(n)} = \frac{k}{n}) = \mathbb{P}(\frac{k}{n} \leq X \leq \frac{k+1}{n}) = \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx$ ist:

$$(*) \quad \mathbb{E}(X_{(n)}) = \sum_{k=-\infty}^{\infty} \frac{k}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx \leq \sum_{k=-\infty}^{\infty} \int_{\frac{k}{n}}^{\frac{k+1}{n}} xf(x)dx =$$

$$\overset{\sigma-Add.d.Int.}{=} \int_{-\infty}^{\infty} xf(x)dx \leq \sum_{k=-\infty}^{\infty} \frac{k+1}{n} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)dx = \mathbb{E}(X_{(n)}) + \frac{1}{n}$$

(Ähnlich der Überlegung zur absoluten Konvergenz $\mathbb{E}(X_{(n)})$) $\lim_{n\to\infty}$ in (*) liefert die Behauptung. ✓

Allgemeiner gilt der folgende Satz (k-dimensionaler Zufallsvektor X, Komposition $\varphi \circ X$).

4.5.6 Satz

Besitzt der k-dimensionale Zufallsvektor X die Dichte $f(x)$, $x \in \mathbb{R}$, und ist φ eine (meßbare) Funktion von $\mathbb{R}^k \to \mathbb{R}$ o gilt:

$$\mathbb{E}(\varphi \circ X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x)f(x)dx_1 \dots dx_k$$

sofern $\int \dots \int |\varphi| f \, dx < \infty$.

Beweis: Gemäß 4.3.5 ist $\varphi \circ X$ Zufallsvariable. Ähnlich wie oben gilt:

$$\mathbb{E}(\varphi \circ X)_{(n)} = \dots \leq \dots = \underset{\mathbb{R}^k}{\int \dots \int} \varphi(x)f(x)dx \leq \dots = \mathbb{E}(\varphi \circ X)_{(n)} + \frac{1}{n} \quad ✓$$

Wie in 3.1 haben wir die Linearität und Monotonie des Erwartungswertes.

4.5.7 Satz

Sind X und Y Zufallsvariablen mit Erwartungswerten $\mathbb{E}(X)$ und $\mathbb{E}(Y)$, so gilt:

a) $\mathbb{E}(aX + bY)$ existiert, und $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ für alle $a, b \in \mathbb{R}$. ('Linearität')

b) $\mathbb{E}(X) \leq \mathbb{E}(Y)$, falls $X \leq Y$. ('Monotonie')

c) $\mathbb{E}(1) = 1$. ('Normiertheit')

Beweis: folgt aus den entsprechenden Eigenschaften für diskrete Zufallsvariablen. ✓

Für die Existenz des Erwartungswertes ist das sogenannte *Majorantenkriterium* nützlich.

4.5.8 Satz

Sind X, Y Zufallsvariablen mit $|X| \leq Y$ und $\mathbb{E}(Y)$ existiert (d.h. $\mathbb{E}(Y) < \infty$), so existiert auch $\mathbb{E}(X)$ (und es ist $\mathbb{E}(x) \leq \mathbb{E}(Y)$ nach 4.5.7 b)).

Beweis: Für die approximierten (diskreten) Zufallsvariablen $|X|_{(n)}$ und $Y_{(n)}$ gilt $|X|_{(n)} \leq Y_{(n)}$ und deshalb:

$$\mathbb{E}(|X|_{(n)}) \leq \mathbb{E}(Y_{(n)}) < \infty$$

(letzteres für $n \geq n_0$ nach Voraussetzung). Also existiert auch $\mathbb{E}|X|$ und - nach 4.5.4 a) - auch $\mathbb{E}(X)$.
In Verallgemeinerung von Satz 3.1.8 gilt der folgende Satz.

4.5.9 Satz

Existieren für die unabhängigen Zufallsvariablen X und Y die Erwartungswerte $\mathbb{E}(X)$ und $\mathbb{E}(y)$, so existiert auch der Erwartungswert für $X \cdot Y$, und es gilt:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{4.18}$$

Beweis:

i) Man kann die Approximation $X_{(n)}$, $Y_{(n)}$ in der Form $X_{(n)} = \varphi(X)$, $Y_{(n)} = \varphi(Y)$ schreiben, mit eine geeigneten (meßbaren) Funktion $\varphi \equiv \varphi_n$.[8]
Nach Satz 4.4.8 sind dann auch $X_{(n)}$, $Y_{(n)}$ unabhängige Zufallsvariablen, und es gilt nach 3.1.8 $X_{(n)} \cdot Y_{(n)}$ besitzt einen Erwartungswert und es gilt:

$$(*) \quad \mathbb{E}(X_{(n)} \cdot Y_{(n)}) = \mathbb{E}(X_{(n)}) \cdot \mathbb{E}(Y_{(n)})$$

ii) Wir haben die Ungleichung:

$$|(X \cdot y)_{(n)} - X_{(n)} \cdot Y_{(n)}| \leq$$
$$\leq |(X \cdot Y)_{(n)} - |X \cdot Y| + |X \cdot Y| - |X_{(n)} \cdot Y| + |X_{(n)} \cdot Y| - X_{(n)} \cdot Y_{(n)}| \leq$$
$$\leq \frac{1}{n} + \frac{1}{n}|Y| + \frac{1}{n}|X_{(n)}| \leq \frac{1}{n}(2 + |X| + |Y|) \equiv \frac{1}{n}Z$$

Daraus folgt (!):

- $\mathbb{E}(X \cdot Y)_{(n)}$ existiert, also auch $\mathbb{E}(X \cdot Y)$
- $\mathbb{E}(X_{(n)} \cdot Y_{(n)}) - \mathbb{E}(X \cdot y)_{(n)} \overset{n\to\infty}{\longrightarrow} 0$ so daß $\lim_{n\to\infty}$ in (*) die Gleichung (4.18) liefert. ✓

Für das nun Folgende ('höhere Momente') wird wiederholt folgende Ungleichung benützt:

$$|a \pm b|^m \leq C_m(|a|^m + |b|^m) \tag{4.19}$$

für alle $a, b \in \mathbb{R}$, $m \in \mathbb{N}$, mit $C_m = 2^{m-1}$.
Dieser Ungleichung folgt aus der *Jensenschen Ungleichung* in der Form ($r, m \in \mathbb{N}$, $a_i \geq 0$):

$$\frac{1}{r^m}(a_1 + \ldots + a_r)^m \leq \frac{1}{r}(a_1^m + \ldots + a_r^m)$$

(in (4.19) ist $r = 2$ gesetzt).

4.5.10 Definitionen

Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

a) Für $m \in \mathbb{N}$ bezeichne $\mathfrak{L}_m \equiv \mathfrak{L}_m(\mathbb{P})$ die Menge aller Zufallsvariable auf $(\Omega, \mathfrak{A})$ mit $\mathbb{E}|X|^m < \infty$.
Für $X \in \mathfrak{L}_m$ heißt $\mathbb{E}|X|^m$ das *absolute* m-te Moment ($\mathbb{E}(X^m)$ das m-te).

b) Für $X \in \mathfrak{L}_m$ führt man noch ein: das m-te zentrierte Moment $\mathbb{E}((X - \mathbb{E}X)^m)$ und das absolute m-te zentrierte Moment $\mathbb{E}(|X - \mathbb{E}X|^m)$.[9]

c) Speziell für $X \in \mathfrak{L}_2$ heißt $Var(X) = \mathbb{E}(X_{\mathbb{E}}X)^2$ Varianz von X und $\sigma(X) = \sqrt{Var(X)}$ Standardabweichung von X. Wie in 3.2.3 haben wir $Var(aX + b) = a^2 Var(X)$ und $Var(X) = \mathbb{E}(X^2 - (\mathbb{E}X)^2)$.
Ferner:

- $Var(X) = 0$ genau dann, wenn $\mathbb{P}(X = \text{const.}) = 1$ ['X =const., $\mathbb{P}$ fast überall[10]']
- $\mathbb{E}X^2 = 0$ genau dann, wenn $\mathbb{P}(X = 0) = 1$ ['$X = 0$, $\mathbb{P}$ f.ü.']

---

[8] siehe 4.5.1 oben

[9] Existenz von $\mathbb{E}X$ durch 4.5.12, unten, gesichert; $\mathbb{E}|X - \mathbb{E}X|^m \overset{4.19}{=} C_m(\mathbb{E}|X|^m + |\mathbb{E}X|^m) < \infty$ für ein $X \in \mathfrak{L}_m$

[10] f.ü. (fast überall) = f.s. (fast sicher) = a.s. (almost sure)

4.5.11 Beispiel

1. X gleichverteilt auf $[a, b]$, $a < b$. Dann ist $X^* = \frac{X-a}{b-a}$ gleichverteilt auf $[0,1]$ und

$$\mathbb{E}X^* \overset{4.5.5}{=} \int_0^1 x \cdot 1 \cdot dx = \frac{1}{2} \overset{(X=a+(b-a)X^*)}{\leadsto} \mathbb{E}X = a \cdot \frac{1}{2}(b-a) = \frac{1}{2}(a+b)$$

$$\mathbb{E}(X^*) \overset{4.5.6}{=} \int_0^1 x \cdot 1 \cdot dx = \frac{1}{3}$$

$$Var(X^*) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

also $Var(X) = \frac{1}{12}(b-a)^2$.

2. X exponentialverteilt mit Parameter $\lambda > 0$.

$$\mathbb{E}X = \int_0^\infty x \cdot \lambda \cdot e^{-\lambda x} \cdot dx \overset{s=\lambda x}{=} \frac{1}{\lambda} \underbrace{\int_0^\infty s \cdot e^{-s} \cdot ds}_{1} \overset{part.Int.}{=} \frac{1}{\lambda}$$

$$\mathbb{E}X^2 = \int_0^\infty x^2 \cdot \lambda \cdot e^{-\lambda x} \cdot dx \overset{s=\lambda x}{=} \frac{1}{\lambda^2} \underbrace{\int_0^\infty s^2 \cdot e^{-s} \cdot ds}_{\lambda} \overset{part.Int.}{=} \frac{2}{\lambda^2}$$

$$Var(X^*) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

3. Normalverteilung $N(\mu, \sigma^2)$
Ist X $N(\mu,\sigma^2)$-verteilt, dann ist $X^* = \frac{X-\mu}{\sigma}$ $N(0,1)$-verteilt, gemäß 4.3.9. Es gilt:

$$\mathbb{E}X^* = \int_{-\infty}^{\infty} |x > e^{-\frac{1}{2}x^2} dx = 0$$

wegen $\varphi(x) = \varphi(-x)$ und wegen $\int_{-\infty}^{\infty} |x > e^{-\frac{1}{2}x^2} dx < \infty$.
Ferner:

$$Var(X^*) = \mathbb{E}(X^*)^2 = \int_{-\infty}^{\infty} x \cdot x\varphi(x)dx \overset{part.Int.}{=} \int_{-\infty}^{\infty} \varphi(x)dx \overset{3.4.1}{=} 1$$

Es folgt für $X = \mu + \sigma X^*$: $\mathbb{E}X = \mu$, $Var(X) = \sigma^2$.
Die $N(\mu,\sigma^2)$-Verteilung kann also als Normalverteilung mit Erwartungswert μ und Varianz σ^2 charakterisiert werden.

Den Anschluß an die Lineare Algebra / Funktionalanalysis liefert folgender Satz.

4.5.12 Satz

Seien $(\Omega, \mathfrak{A}, \mathbb{P})$ und $m \in \mathbb{N}$ vorgegeben.

a) $\mathfrak{L}_m$ ist ein linearer Raum.

b) $\mathfrak{L}_n \subset \mathfrak{L}_m$ für alle $n \geq m$. D.h. aus $\mathbb{E}|X|^n < \infty$ für ein $n \in \mathbb{N}$ folgt $\mathbb{E}|X|^m < \infty$ für $m \leq n$, insbesondere ist $\mathbb{E}|X| < \infty$.

Beweis:

a) Majorantenkriterium 4.5.8 und Ungleichung (4.19) nach 4.5.9 liefern für $a, b \in \mathbb{R}$:

$$|aX + bY|^m \overset{4.19}{\leq} C_m(|a|^m|X|^m + |b|^m|Y|^m)$$

b) Sei $\mathbb{E}|X|^n < \infty$. Dann gilt für $m \leq n$ wegen $|X|^m \leq 1 \cdot 1_{\{|X|\leq 1\}} + |X| 1_{\{|X|\geq 1\}} \leq 1 + |X|^n$ auch $\mathbb{E}|X|^m < \infty$. ✓

Wichtig sind die folgenden "stochastischen Ungleichungen".

4.5.13 Ungleichungen

Markov-Ungleichung: Ist $X \in \mathfrak{L}_m$ für ein $m \in \mathbb{N}$, so gilt für jedes $\varepsilon < 0$:

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|^m}{\varepsilon^m} \tag{4.20}$$

Tschebyschoff-Ungleichung: Insbesondere für $X \in \mathfrak{L}_2$:

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \tag{4.21}$$

Beweis: Wiederholte Anwendung der Monotonieeigenschaften von $\mathbb{E}$:

$$\mathbb{E}|X|^m \geq \mathbb{E}(|X|^m 1_{\{|X|\geq 1\}}) \geq \varepsilon^m \mathbb{E}(1_{\{|X|\geq 1\}}) = \varepsilon \mathbb{P}(|X| \geq \varepsilon)$$

setzt man in (4.20) speziell $X - \mathbb{E}X$ statt X ein, sowie $m = 2$, so erhält man (4.21). ✓
Genauso wie in 3.2.6 beweist man die Cauchy-Schwarzsche-Ungleichung.

4.5.14 Satz

Für Zufallsvariablen $X, Y \in \mathfrak{L}_2$ gilt $X \cdot Y \in \mathfrak{L}_1$ und $[\mathbb{E}(X \cdot Y)]^2 \leq \mathbb{E}(X)^2 \cdot \mathbb{E}(Y)^2$. Das "="-Zeichen gilt genau dann, wenn $aX + bY = 0$, $\mathbb{P}$ f.ü. für $a, b, a^2 + b^2 > 0$

4.5.15 Bemerkungen zum $\mathfrak{L}_2$

Im linearen Raum $\mathfrak{L}_2$ können wir ein 'Fast-Skalarprodukt' einführen:
Für $X, Y \in \mathfrak{L}_2$ setze $< X, Y >= \mathbb{E}(X, Y)$. $< \cdot, \cdot >$ ist dann eine bilineare, symmetrische, positiv semidefinite ($< X, X > \geq 0$) Form. Aus $< X, X >= 0$ folgt aber nur $X = 0$ f.ü. (und nicht $X = 0$).

4.5.16 Definitionen

Sind $X, Y \in \mathfrak{L}_2$, dann heißen

a) $Cov(X, Y) = \mathbb{E}((X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)) = \mathbb{E}(X \cdot y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$ die *Kovarianz* von X und Y.

b) X, Y *unkorreliert*, falls $Cov(X, Y) = 0$.

c) $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma(X)\cdot\sigma(Y)}$ *Korrelation* (oder *Korrelationskoeffizient*) von X und Y, sofern $\sigma(X) \cdot \sigma(Y) > 0$.

Die Folgerungen 2. und 3. nach 3.2.6[11] gelten weiterhin, sowie die Eigenschaften 3.2.7 von Var und Cov[12]
Im Hinblick auf 4.5.15: X, Y unkorreliert, falls $X - \mathbb{E}X \perp Y - \mathbb{E}Y$ (bezüglich $< \cdot, \cdot >$).[13]

4.5.17 Beispiel

Momente der k-dimensionalen Normalverteilung[14].

Ist $X = (X_1, \ldots, X_k)$ $N_k(\mu, \Sigma)$-verteilt, $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} \in \mathbb{R}^k$, $\Sigma = (\sigma_{i,j})$ symmetrisch, positiv definite $k \times k$-Matrix.

Behauptung: $\mathbb{E}X_i = \mu_i$, $Cov(X_i, X_j) = \sigma_{i,j}$

[11] Insbesondere: $-1 \leq \rho(X, Y) = 1$, $|\rho(X, Y)| = 1$ genau dann, wenn $X - \mathbb{E}X = c(Y - \mathbb{E}Y)\mathbb{P}$ f.ü. ($c \neq 0$).

[12] Insbesondere: X, Y unabhängig $\Rightarrow$ X, Y unkorreliert, $Var(X_1 + \ldots + X_n) = \sum_{i=1}^n Var(X_i)$, falls $X_1, \ldots, X_n$ paarweise unkorreliert.

[13] "$\perp$" = orthogonal

[14] vgl. 4.2.12

Beweis: (Ü.A. Blatt 10)

Bemerkung: Die Parameter μ, Σ der $N_k(\mu, \Sigma)$-Verteilung bilden also den Erwartungswert-Vektor bzw. die Matrix der Kovarianzen (Cov-Matrix) des $N_k(\mu, \Sigma)$-verteilten Zufallsvektors X.

4.6 Charakteristische Funktion

Für diskrete Zufallsvariablen X mit Werten $\mathbb{Z}_+$ erwies sich die erzeugende Funktion $G(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}_X\{k\}$ als nützlich, und zwar bei der Berechnung von: Momenten, Faltungen und Grenzverteilungen.
Eine vergleichbare Funktion hat die charakteristische Funktion in der allgememinen Wahrscheinlichkeitstheorie, in der X einen beliebige Zufallsvariable ist. Anstelle des Erwartungswertes s^X (der nicht notwendigerweise existiert) bildet man den Erwartungswert der komplexwertigen Variablen "e^{iX}"

4.6.1 Erinnerung: Komplexe Zahlen

Für eine komplexe Zahl $z = a+bi$, $a = Re(z) \in \mathbb{R}$, $b = Im(z) \in \mathbb{R}$ setze man: $|z| = \sqrt{a^2+b^2} = \sqrt{z \cdot \overline{z}}$.
Es ist $z = r \cdot e^{i\varphi}$ mit $r = |z|$, $e^{i\varphi} = \cos\varphi + i \sin\varphi$.
Es gilt $|z \cdot w| = |z| \cdot |w|$.

4.6.2 Definition

Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

a) Sind z_1, z_2 Zufallsvariable auf $(\Omega, \mathfrak{A})$, ($z_i : \Omega \to \mathbb{R}$) so bildet $z = z_1 + iz_2$ eine komplexwertige Zufallsgröße auf $(\Omega, \mathfrak{A})$, ($z : \Omega \to \mathbb{C}$).

b) Existieren $\mathbb{E}(z_1), \mathbb{E}(z_2)$, so heißt die komplexe Zahl $\mathbb{E}(z) := \mathbb{E}(z_1) + i\mathbb{E}(z_2)$ Erwartungswert von z.

Hilfssatz:

a) Sind $z, \overline{z}$ komplexe Zufallsgrößen und existieren $\mathbb{E}(z), \mathbb{E}(\overline{z})$, so gilt:

$$\mathbb{E}(z + \overline{z}) = \mathbb{E}(z) + \mathbb{E}(\overline{z})$$

$$\mathbb{E}(v \cdot z) = v \cdot \mathbb{E}(z) \ , \quad v \in \mathbb{C}$$

b) $|\mathbb{E}(z)| \leq \mathbb{E}(|z|) < \infty$, (Ü.A. 38a)

4.6.3 Charakteristische Funktion (Definition)

Ist X eine Zufallsvariable auf $(\Omega, \mathfrak{A})$, so heißt die komplexwertige Funktion $\varphi_X : \mathbb{R} \to \mathbb{C}$:

$$\varphi_X(t) \equiv \varphi(t) = (e^{itx}) = \mathbb{E}(\cos(tx)) + i\mathbb{E}(\sin(tx))$$

charakteristische Funktion von X.

Bemerkungen:

1. Aus $e^{itx} = \cos(tx) + i\sin(tx)$ folgt wegen $|\cos(tx)| \leq 1$, $|\sin(tx)| \leq 1$ die Existenz von $\mathbb{E}(\cos(tx))$ und $\mathbb{E}(\sin(tx))$, also von (e^{itx}).

2. Beispiele für charakteristische Funktionen:

   - $\varphi_X(t) = 1$
   - $\varphi_X(t) = \cos(t)$
   - $\varphi_X(t) = e^{-\frac{t^2}{2}}$

- $\varphi_X(t) = e^{e^{it}-1}$
- $\varphi_X(t) = \frac{1}{1+it}$
- $\varphi_X(t) = \frac{1}{it}(e^{it}-1)$, $(\varphi_X(0) = 1)$
- $\varphi_X(t) = e^{-|t|}$
- $\varphi_X(t) = (1-|t|)1_{[-1,1]}(t)$

Keine charakteristischen Funktionen sind:

- $\varphi(t) = \sin(t)$
- $\varphi(t) = 1 - t^2$
- $\varphi(t) = 1_{[-1,1]}(t)$
- $\varphi(t) = e^{-|t|^2}$

3. Wegen $|e^{itx}| = 1$ gilt $|\varphi(t)| = |\mathbb{E}(e^{itx})| \leq \mathbb{E}(|e^{itx}|) = \mathbb{E}(1) = 1$, $\varphi(0) = \mathbb{E}(1) = 1$.

4. φ_X ist gleichmäßig stetig. (ohne Beweis)

5. $\varphi_{aX+b}(t) = \mathbb{E}(e^{it(aX+b)}) = e^{itb} \cdot \mathbb{E}(e^{itaX}) = e^{itb} \cdot \varphi_X(ta)$, $a, b \in \mathbb{R}$.

6. Ist X eine Zufallsvariable mit Werten in $\mathbb{Z}_+$, so ist

$$\varphi_X(t) = (e^{itx}) = \mathbb{E}(\cos(tx)) + i\mathbb{E}(\sin(tx)) =$$

$$= \sum_{k=0}^{\infty} \cos(tk) \cdot \mathbb{P}(X = k) + i \sum_{k=0}^{\infty} \sin(tk) \cdot \mathbb{P}(X = k) =$$

$$= \sum_{k=0}^{\infty} (e^{it})^k \cdot \mathbb{P}(\{k\})$$

(vgl. mit $g_x(s) = \mathbb{E}(s^x) = \sum_{k=0}^{\infty} s^k \cdot \mathbb{P}_x(\{k\})$, $s \in [0,1]$) Also (!) lautet die charakteristische Funktion von X:

X $B(n,p)$**-verteilt:** $\varphi_X(t) = (1-p+pe^{it})^n$, $t \in \mathbb{R}$

X $P(\lambda)$**-verteilt:** $\varphi_X(t) = e^{\lambda e^{it}-1}$, $t \in \mathbb{R}$

4.6.4 Beispiele

a) X exponentialverteilt mit Parameter $\lambda > 0$:

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}\,, \quad t \in \mathbb{R}$$

b) Xsei $N(0,1)$-verteilt:

$$\varphi_X(t) = (e^{itx}) = \mathbb{E}(\cos(tx)) + i\mathbb{E}(\sin(tx)) =$$

$$= \int_{-\infty}^{\infty} (\cos(tx)) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + i \int_{-\infty}^{\infty} (\sin(tx)) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$\Longrightarrow \varphi_X'(t) \overset{(!)}{=} \int_{-\infty}^{\infty} \underbrace{(\sin(tx))}_{n'} \underbrace{\frac{1}{\sqrt{2\pi}}(-x)e^{-\frac{x^2}{2}}}_{v'} dx =$$

$$= \left[(\sin(tx)) \frac{1}{\sqrt{2\pi}}(-x)e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} =$$

$$= \int_{-\infty}^{\infty} t \cdot \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = -t \cdot \varphi_X(t)$$

$$\frac{d}{dt}(\varphi_X(t) \cdot e^{-\frac{t^2}{2}}) = \varphi'_X(t) \cdot e^{-\frac{t^2}{2}} + \varphi_X(t) \cdot t \cdot e^{-\frac{t^2}{2}} =$$

$$= e^{-\frac{t^2}{2}} [\underbrace{\varphi'_X(t) + t \cdot \varphi_X(t)}_{=0}] = 0$$

$$\Longrightarrow \varphi_X(t) \cdot e^{-\frac{t^2}{2}} = c = \text{const.}$$

$$\Longrightarrow \left. \begin{array}{l} \varphi_X(t) = c \cdot e^{-\frac{t^2}{2}} \\ \varphi_X(0) = 1 \Rightarrow c = 1 \end{array} \right\} \Rightarrow \varphi_X(t) = e^{-\frac{t^2}{2}}$$

c) X sei $N(0,1)$-verteilt $\Rightarrow \varphi_X(t) = ?$
$X = \sigma\Gamma + \mu$, Γ $N(0,1)$-verteilt.
$\Rightarrow \varphi_X(t) \cdot e^{it\mu} \cdot \varphi_Y(\sigma t) = e^{it\mu} \cdot e^{-\frac{1}{2}\sigma^2 t^2}$.

4.6.5 Eindeutigkeitssatz

Seien X, Y Zufallsvariablen. Dann gilt:

$$\varphi_X = \varphi_y \Longleftrightarrow \mathbb{P}_X = \mathbb{P}_Y$$

Beweis: Gänßler / Stute: Wahrscheinlichkeitstheorie, S....

4.6.6 Faltungssatz

Sind X, Y unabhängige Zufallsvariablen, so gilt: $\varphi_{X+Y} = \varphi_X + \varphi_Y$

Hilfssatz: Seiein X, Y unabhängige Zufallsvariablen, $f = f_1 + f_2$, $g = g_1 + g_2$ komplexwertige Funktionen, so gilt, falls $\mathbb{E}(f(x))$, $\mathbb{E}(g(y))$ existieren:

$$\mathbb{E}(f(x) \cdot g(y)) = \mathbb{E}(f(x)) \cdot \mathbb{E}(g(y))$$

Beweis hierzu: (Ü.A.)

Beweis:

$$\varphi_{x+y}(t) = \mathbb{E}(e^{it(x+y)}) = \mathbb{E}(e^{itx} \cdot e^{ity} = \mathbb{E}(e^{itx}) \cdot \mathbb{E}(e^{ity}) = \varphi_x + \varphi_x$$

Beispiel: Es gilt $N(\mu_1, \sigma_1^2) \times N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Beweis: Sei X_1 $N(\mu_1, \sigma_1^2)$-verteilt und X_2 $N(\mu_2, \sigma_2^2)$-verteilt, mit X_1, X_2 unabhängig.

$$\varphi_{x_1+x_2}(t) = \varphi_{x_1}(t) \cdot \varphi_{x_2}(t) = e^{it\mu_1} \cdot e^{-\frac{1}{2}\sigma_1^2 t^2} \cdot e^{it\mu_2} \cdot e^{-\frac{1}{2}\sigma_2^2 t^2} =$$

$$= e^{it(\mu_1+\mu_2)} \cdot e^{-\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2} \stackrel{4.6.5}{\Rightarrow} \text{Beh.}$$

4.6.7 Satz (Berechnung von Momenten)

Für die Zufallsvariable X existieren $\mathbb{E}(X^m)$ für ein $m \in \mathbb{N}$. Dann ist die charakterisierende Funktion φ_x m-mal stetig differenzierbar mit:

$$\varphi_x^{(m)}(0) = i^m \mathbb{E}(X^m)$$

(für m gerade gilt auch die Umkehrung).

Kapitel 5

Grenzwertsätze

5.1 Gesetz der großen Zahlen

Vorbemerkung: Gesetze der großen Zahlen haben die Konvergenz von $\frac{1}{n}((X_1 - \mu_1) + \ldots + (X_n - \mu_n))$ gegen 0 zum Inhalt, wenn $X_1, X_2, \ldots$ eine Folge von Zufallsvariablen ist, und $\mu_i = \mathbb{E}(X_i)$.

Beispiel: Sind $X_1, X_2, \ldots$ unabhängige, $B(n,p)$-verteilte Zufallsvariablen, so vermutet man eine Konvergenz von $\frac{1}{n}\sum_{i=1}^n X_i$ ('relative Häufigkeit') gegen p ('Auftrittswahrscheinlichkeit'). Dabei müssen Konvergenzbegriffe der Stochastik eingeführt werden (In 5.1 und 5.2 werden 3 verschiedene Konvergenzen einführen).

5.1.1 Definitionen

a) Wie sagen, daß eine Folge $Y_1, Y_2, \ldots$ von Zufallsvariablen (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$) stochastisch gegen eine Zufallsvariable Y konvergiert, falls $\forall \varepsilon < 0$:

$$\mathbb{P}(|Y_n - Y| \geq \mathbb{E}) \to 0 \ , \quad (n \to \infty) \tag{5.1}$$

gilt. Man schreibt dafür $Y_n \xrightarrow{\mathbb{P}} Y$.

b) Wir sagen, daß eine Folge $X_1, X_2, \ldots$ von Zufallsvariablen (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$) mit existierendem Erwartungswert $\mu_i = \mathbb{E}(X_i)$ das schwache Gesetz der großen Zahlen erfüllt, falls eine Folge

$$Y_n = \frac{1}{n}[(X_1 - \mu_1) + \ldots + (X_n - \mu_n)] n = 1.2\ldots$$

von Zufallsvariablen stochastisch gegen 0 konvergiert.

$$Y_n \xrightarrow{\mathbb{P}} 0$$

5.1.2 Schwaches Gesetz der großen Zahlen (Satz)

Sind X_1, X_2 paarweise unkorrelierte Zufallsvariablen (auf Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$) mit $X_i \in \mathfrak{L}_2$ und mit $\frac{1}{n^2}\sum_{i=1}^n Var(X_i) \to 0$, $(n \to \infty)$, so erfüllt diese Folge $X_1, X_2, \ldots$ das schwache Gesetz der großen Zahlen.

Beweis: Für die Zufallsvariablen $Y_n = \frac{1}{n}\sum(X_i, \mu_i)$ gilt $\mathbb{E}Y_n = 0$ und $Var(Y_n) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) \to 0$, $(n \to \infty)$, Tschebyschow Ungleichung[1] liefert:

$$\mathbb{P}(|Y_n = \underbrace{\mathbb{E}(Y_n)}_{=0}| \geq \varepsilon) \leq \frac{Var(Y_n)}{\varepsilon^2} \to \ , \quad (n \to \infty)$$

[1] siehe 4.5.13

5.1.3 Korollar

Sind X_1 und X_2 unabhängige Zufallsvariablen aus $\mathfrak{L}_2$ mit gleichmäßig beschränkten Varianzen (d.h. $Var(X_i) \leq M < \infty \; \forall i = 1, 2, \ldots$), dann erfüllt diese Folge das schwache Gesetz der großen Zahlen.

5.1.4 Beispiel

Ist $X^n = X_1 + \ldots + X_n$ $B(n,p)$-verteilt ($X_1, X_2, \ldots$ unabhängig $B(1,p)$-verteilt), so gilt:

$$\frac{1}{n} X^n \xrightarrow{\mathbb{P}} p'$$

Umgangssprachlich: die relativen Häufigkeiten des Ereignisses '1' konvergiert stochastisch gegen ω.

Bemerkung: Die *stochastische Konvergenz* stellt einen relativ schwachen Konvergenzbegriff dar. So braucht für kein $\omega \in \Omega$ gewöhnliche Konvergenz $Y_n(\omega) \to Y(\omega)$, $(n \to \infty)$, stattzufinden, wie das folgende Beispiel zeigt.

5.1.5 Beispiel

Sei $(\Omega, \mathfrak{A}, \mathbb{P}) = ([0,1], [0,1] \cap \mathfrak{B}^1$, Gleichverteilung). Man definiere die Folge $Y_n = 1_{A_n}$, $n \geq 1$, durch:

$$A_n = \{\omega \leftarrow [0,1] : \text{ es exisitiert ein } m \in \mathbb{N} \text{ mit } \omega + m \leftarrow [a_{n-1}, a_n]\}$$

wobei $a_0 = 0$ und $a_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$, $(\omega \in [a_{n-1}, a_n] mod 1)$.
Es gilt

1. $Y_n \xrightarrow{\mathbb{P}} 0$, denn für $\varepsilon \in (0,1)$ ist $\mathbb{P}(|Y_n - 0| \geq \varepsilon) = \mathbb{P}(Y_n = 1) = \mathbb{P}(A_n) = \frac{1}{n} \to 0$
2. Die Folge $Y_n(\omega)$ konvergiert für kein $\omega \in [0,1]$, wegen der Konvergenz der harmonischen Reihe.

Der Konvergenzbegriff $Y_n(\omega) = Y(\omega) \; \forall \omega \in \Omega$ ist für die Stochastik unbrauchbar. So ist für $Y_n = \frac{1}{n} X^{(n)}$, $X^{(n)}$ $B(n,p)$-verteilt:

$$Y_n(\omega) \text{ nicht konvergent}$$

für viele ω.
Wir nehmen die Sprechweise wieder auf: Eine Aussage gilt '$\mathbb{P}$ *fast überall*' oder '$\mathbb{P}$ *fast sicher*' (synonym), wenn die Menge A aller ω für die die Aussage richtig ist, die Wahrscheinlichkeit 1 hat: $\mathbb{P}(A) = 1$.

5.1.6 Definition

a) Eine Folge $Y_1, Y_2, \ldots$ von Zufallsvariablen (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$) konvergiert fast sicher gegen die Zufallsvariable Y, falls:

$$\mathbb{P}\{\omega : \lim_{n \to \infty} Y_n(\omega) = Y(\omega)\} = 1 \tag{5.2}$$

Man schreibt kürzer: $\mathbb{P}(\lim_n Y_n = Y) = 1$ bzw. $Y_n \to Y$ $\mathbb{P}$ fast sicher.

b) Man sagt, daß eine Folge $X_1, X_2, \ldots$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, \mathbb{P})$ mit existierenden Erwartungswerten $\mu_i \equiv \mathbb{E}(X_i)$ das starke Gesetz der großen Zahlen erfüllt, falls die Folge $Y_n = \frac{1}{n}[(X_1 - \mu_1) + \ldots + (X_n - \mu_n)]$, $n = 1, 2, \ldots$, $\mathbb{P}$-f.s. gegen 0 konvergiert: $Y_n \to 0$ $\mathbb{P}$-f.s.

Bemerkung: Aus $Y_n \to Y$ $\mathbb{P}$-f.s. folgt $Y_n \xrightarrow{\mathbb{P}} Y$ (ohne Beweis), Beispiel 5.1.5 zeigt, daß die Umkehrung nicht gilt (ein vereinfachtes Beispiel dazu folgt in 5.1.8).
Das wichtigste Hilfsmittel zum Beweis eines starken Gesetzes der großen Zahlen ist das folgende Lemma von Borel-Cantelli, das auch sonst wichtig ist.

5.1.7 Lemma (von Borel-Cantelli)

Sei $(\Omega, \mathfrak{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \ldots$ eine Folge von Ereignissen aus $\mathfrak{A}$. Sei A^* das Ereignis, daß unendlich viele der A's eintreten:

$$A^* = \{\omega \in \Omega : \ \omega \in A_i; \ \text{ für unendlich viele } i \in \mathbb{N}\}$$

a) Gilt $\sum_{i=1}^{\infty} \mathbb{P}(A_i) \leq \infty$, dann ist $\mathbb{P}(A^*) = 0$

b) Sind die $A_1, A_2, \ldots$ unabhängig und ist $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$, dann ist $\mathbb{P}(A^*) = 1$

Beweis:

a) Es ist $\omega \in A^*$ genau dann, wenn es $\forall n \in \mathbb{N}$ ein $i \geq n$ gibt, $\omega \in A_i$. D.h.:

$$A^* = \bigcap_{n=1}^{\infty} \bigcup_{i \geq n} A_i$$

Da $A^* \subset \bigcup_{1 \geq n} A_i$ für jedes $n \in \mathbb{N}$ ist, gilt:

$$\mathbb{P}(A^*) = \mathbb{P}(\bigcup_{1 \geq n} A_i) \overset{(!)}{\leq} \sum_{i \geq n} \mathbb{P}(A_i) \to 0$$

für $n \to \infty$.

b) Wir benutzen die Ungleichung $1 - x \leq e^{-x}$, $\forall x \in \mathbb{R}$, und die Unabhängigkeit der $\overline{A_1}, \overline{A_2}, \ldots$. Es gilt für alle n und $N \geq n$:

$$\mathbb{P}(\bigcap_{i=1}^{\infty} \overline{A_i}) \leq \mathbb{P}(\bigcap_{i=n}^{N} \overline{A_i}) \overset{unabh.}{=} \prod_{i=n}^{N} (1 - \underbrace{\mathbb{P}(A_i)}_{x} \leq \prod_{i=n}^{N} e^{-\mathbb{P}(A_i)} =$$

$$= \exp(-\sum_{i=n}^{N} \mathbb{P}(A_i)) \to 0$$

für $N \to \infty$, wegen der Divergenz der Reihe. Also $\mathbb{P}(\overline{A_i}) = 0$ für jedes n:

$$\mathbb{P}(\overline{A_i^*}) = \mathbb{P}(\bigcup_{n=1}^{\infty} \bigcap_{1 \geq n} \overline{A_i}) \leq \sum_{n=1}^{\infty} \mathbb{P}(\bigcap \overline{A_i}) = 0$$

d.h. $\mathbb{P}(A^*) = 1$.

5.1.8 Bemerkungen zu 5.1.7

1. Teil b) rechtfertigt den populären Ausspruch: "Ein Ereignis, das (mit positiver Wahrscheinlichkeit) eintreten kann, tritt mit ($\mathbb{P}$-) Sicherheit einmal ein (sogar beliebig oft), wenn nur genügend (unabhängige) Versuche durchgeführt werden".

2. Mit Teil b) läßt sich ein weiteres Beispiel einer Folge Y_n, $n \geq 1$ angeben, die stochastisch konvergiert, aber nicht fast sicher. Seien $Y_1, Y_2, \ldots$ unabhängige $B(1, \frac{1}{n})$-verteilte Zufallsvariablen. Dann gilt (wie in 5.1.5) $Y_n \xrightarrow{\mathbb{P}} 0$, denn für ein $0 < \varepsilon < 1$ ist $\mathbb{P}(|Y_n| > \varepsilon) = \mathbb{P}(Y_n = 1) = \frac{1}{n} \to 0$, $(n \to \infty)$.

3. Andererseits konvergiert die Folge für $\mathbb{P}$ fast alle ω[2] **nicht**! Denn wegen $\sum_n \mathbb{P}(\underbrace{Y_n = 1}_{A_n}) =$ $\sum_n \frac{1}{n} \to \infty$ folgt

$$\mathbb{P}(\limsup Y_n = 1) = mp(A^*) \overset{b)}{=} 1$$

[2] äquivalent zu 'fast sicher'

und wegen $\sum_n mp(\underbrace{Y_n = 0}_{B_n}) = \sum_n (1 - \frac{1}{n}) = \infty$ folgt

$$\mathbb{P}(\liminf Y_n = 0) = mp(B^*) \overset{b)}{=} 1$$

5.1.9 Starkes Gesetz der großen Zahlen (Satz)

Bilden $X_1, X_2, \ldots$ eine Folge paarweise unkorrelierter Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$, aus $\mathfrak{L}_2$ mit beschränkter Varianz (d.h. $Var(X_i) \leq M < \infty$ für alle i), so erfüllt die Folge das *starke Gesetz der großen Zahlen.*

Beweis: Definiere $Y_n = \frac{1}{n}\sum_{i=1}^n X_i'$, $X_i' = X_i - \mathbb{E}(X_i)$.

i) Wir zeigen zunächst, daß $Y_{n^2} = \frac{1}{n^2}\sum_{i=1}^{n^2} X_i' \to 0$ $\mathbb{P}$-f.s.
Gemäß 3.2.7 c) Bienaymé ist:

$$Var(Y_{n^2}) = \frac{1}{n^4}\sum_{i=1}^{n^2} Var(X_i) \leq \frac{1}{n^2}M$$

so daß nach Tschebyschoff (siehe 4.5.13) für alle $\varepsilon > 0$ und für die Menge $A_k^{(\varepsilon)} = \{\omega : |Y_{n^2}(\omega)| \geq \varepsilon\}$ gilt:

$$\mathbb{P}(A_k^{(\varepsilon)}) \leq \frac{1}{\varepsilon^2}Var(Y_{n^2}) \leq \frac{M}{n^2\varepsilon^2}$$

sowie

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n^{(\varepsilon)}) < \infty$$

Borel-Cantelli-Lemma Teil a) liefert für $A^{*(\varepsilon)} = \{\omega : |Y_{n^2}(\omega)| \geq \varepsilon$, für ∞ viele $n\}$:

$$\mathbb{P}(A^{*(\varepsilon)}) = 0$$

Es folgt:

$$\mathbb{P}(\bigcup_{k=1}^{\infty} A^{*\frac{1}{k}}) \overset{!}{\leq} \sum_{k=1}^{\infty} \mathbb{P}(A^{*\frac{1}{k}}) = 0 \text{ bzw. } \mathbb{P}(\bigcap_{k=1}^{\infty} A^{*\frac{1}{k}}) = 1$$

denn für $\omega \in \bigcap_{k=1}^{\infty} A^{*\frac{1}{k}}$ gilt $|Y_{n^2}(\omega)| \geq \frac{1}{k}$ nur für endlich viele n (für alle k) d.h. für $\mathbb{P}$ fast sicher (für alle ω) gilt: $\forall\, \varepsilon > 0\ \exists\, m_0 = m_0(\omega, \varepsilon)$, so daß:

(5.3) $$|Y_{n^2}(\omega)| \leq \varepsilon \quad \forall\, n^2 \geq m_0$$

ii) Für beliebige $m \in \mathbb{N}$ sei $n = n(m)$ diejenige natürliche Zahl, für welche $n^2 \leq m < (n+1)^2$.
Mit analogen Methoden wie in i) zeigt man für die Menge

$$B^{*(\varepsilon)} = \{\omega : |\underbrace{\frac{m}{n^2}Y_m(\omega)}_{\frac{1}{n^2}\sum_1^m X_i'} - \underbrace{Y_{n^2}(\omega)}_{\frac{1}{n^2}\sum_1^{n^2} X_i}| \geq \varepsilon \text{ für } \infty \text{ viele } m\},$$

daß

$$\mathbb{P}(\bigcap_{n=1}^{\infty} B^{*(\varepsilon)}) = 1$$

[3] Folglich gilt für $\mathbb{P}$ fast sicher: $\forall\, \varepsilon > 0\ \exists\, m_0 \equiv m_0(\omega, \varepsilon)$ mit

(5.4) $$|\frac{m}{n^2}Y_m(\omega) - Y_{n^2}(\omega)| \leq \varepsilon \quad \text{für alle } m \leq m_0$$

[3] vgl. Krengel, S.148 f.

iii) Gleichungen (5.3) und (5.4) liefern für $\mathbb{P}$ fast sicher: $\forall\, \varepsilon > 0\ \exists\, m_0 \equiv m_0(\omega, \varepsilon)$ mit

$$|Y_m(\omega)| \leq \underbrace{\frac{m}{n^2}}_{\geq 1} |Y_m(\omega)| \overset{ungl.}{\leq} |\frac{m}{n^2} Y_m(\omega) - Y_{n^2}(\omega)| + |Y_{n^2}(\omega)| \overset{(5.3),(5.4)}{\leq} 2\varepsilon$$

für alle $m \geq m_0$. Das heißt aber $Y_n \to 0$ $\mathbb{P}$ fast sicher.

Bemerkung: Entsprechend der starken Aussage benötigt Satz 5.1.9 auch eine stärkere Voraussetzung, als Satz 5.1.2.

5.1.10 Beispiel

Ist $X^{(n)}$ $B(n,p)$-verteilt, so gilt $\frac{1}{n} X^{(n)} \to p$ $\mathbb{P}$ fast sicher. Hierdurch wird die Aussage von 5.1.4 verbessert. Dieses Ergebnis bestätigt die Brauchbarkeit unseres Wahrscheinlichkeitstheoretischen Konzeptes. Es präzisiert die Intuition, daß sich für große n annähert.
$\frac{1}{n}(X^n)$ beobachte relative Häufigkeit eines Ereignisses an p (axiomatisch eingeführte Wahrscheinlichkeit der Ereignisse).

5.2 Zentrale Grenzwertsätze

In diesem Abschnitt Verallgemeinerung (und Beweis) des Grenzwertsatzes 3.4.2 von 'de Moivre-Laplace' auf Summen unabhängiger, identisch verteilter Zufallsvariablen (anstatt nur unabhängiger Bernoullivariablen). Der folgende Beweis von 5.2.8 benutzt einen 'Stetigkeitssatz' für charakteristische Funktionen und einen dritten Konvergenzbegriff ('Verteilungskonvergenz').

5.2.1 Definition

Seien $X_1, X2, \ldots$ Zufallsvariablen aus $\mathfrak{L}_2$. Man sagt, daß diese Folge den zentralen Grenzwertsatz erfüllt, falls für die Standardisierten der Partialsummen $S_n = X_1 + \ldots + X_n$ mit

$$S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \quad (\equiv\text{standardisieren}) \tag{5.5}$$

gilt:

$$\lim_{n\to\infty} \mathbb{P}(a < S_n \leq b) \to \phi(b) - \phi(a) \quad \forall\, a < b;\ a, b \in \mathbb{R}$$

Dabei ist $\phi(x)$, $x \in \mathbb{R}$, die Verteilungfunktion der $N(0,1)$-Verteilung[4]. Es reicht, $\lim_{n\to\infty} \mathbb{P}(S_n^* \leq x) = \phi(x)\ \forall\, x \in \mathbb{R}$ zu zeigen.

Bemerkungen:

1. Die Gültigkeit des zentralen Grenzwertsatzes eröffnet die Möglichkeit, unter Umständen nicht (oder nur schwer) berechenbare Wahrscheinlichkeiten $\mathbb{P}(a < S^* \leq b)$ durch die Werte der $N(0,1)$-Verteilung zu approximieren.

2. Sind $X_1, X_2, \ldots$ unabhängig, mit identischen Erwartungswerten $\mu = \mathbb{E}(X_i)$ und identischen Varianzen $\sigma^2 = Var(X_i)$, so wird aus (5.5):

$$S_n^z = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)}_{X_i^*\ standardis.} = \sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \tag{5.6}$$

$$\neq \overline{X_m} = \frac{1}{n} \sum_{i=1} n X_i$$

[4] vgl. 3.4

3. Um einen zentralen Grenzwertsatz zu beweisen, müssen wir zeigen:

   (5.7) $$F_n^*(x) \to \phi(x), \quad \forall\, x \in \mathbb{R},\, n \to \infty$$

   wenn $F_n^*(x)$ die Verteilungfunktion von $S_n^*(x)$ ist.
   Diese Aussage (5.7) stellt einen dritten Konvergenzbegriff dar (Verteilungskonvergenz).

Allgemein wird folgendes definiert.

5.2.2 Verteilungskonvergenz (Definition)

Eine Folge Y_n, $n \geq 1$ von Zufallsvariablen heißt *Verteilungskonvergent* gegen die Zufallsvariable Y_0, falls bei $n \to \infty$:

(5.8) $$F_n(x) \to F_0(x) \quad \forall\ \in \mathfrak{C}(F_0)$$

dabei bezeichnet F_n und F_0 die Verteilungsfunktion von Y_n unf Y_0 und $\mathfrak{C}(F_0) \subset \mathbb{R}$ die Menge alles Stetigkeitsstellen von F_0. Man schreibt für (5.8) kurz:

$$Y_n \xrightarrow{\mathfrak{D}} Y$$

[oder auch $Y_n \xrightarrow{\mathfrak{V}} Y$], wobei '$\mathfrak{D}$' hier 'Distribution' bedeutet.

Bemerkungen:

1. Der Begriff der Verteilungskonvergenz verlangt nicht, daß alle Y_n, Y_0 auf demselben Wahrscheinlichkeitsraum definiert sind.

2. Für stetige Verteilungsfunktionen F_0, wie z.B. ϕ ist $\mathfrak{C}(F_0) = \mathbb{R}$ (vgl. (5.7). Die Forderung:

   (5.9) $$F_n(x) \to F_0(x) \quad \forall\ \in \mathbb{R}$$

   erweist sich als zu restriktiv. So gilt im folgenden Beispiel (5.8), jedoch nicht (5.9):
   Y_0, Y_n seien "entartete" Zufallsvariablen mit $\mathbb{P}(Y_n = \frac{1}{n}) = 1$, $\mathbb{P}(Y_0 \to 0) = 1$. Für

   $$F_n(x) = 1_{[\frac{1}{n},\infty[}(x)$$

   $$F_0(x) = 1_{[0,\infty)}(x)$$

   gilt:

   $$\mathfrak{C}(F_0) = \mathbb{R} \setminus \{0\} \quad \text{und}$$

   $$\lim F_n(x) = \left\{ \begin{array}{ll} 1 & x > 0 \\ 0 & x < 0 \end{array} \right\} = F_0(x)$$

   Bei $x = 0$ gilt: $0 = \lim_n F_n(x) \neq F_0(0) = 1$

3. Der nächste Satz zeigt, daß aus stochastischer Konvergenz die Verteilungskonvergenz folgt. Zusammen mit 5.1.6: $Y_n \to Y_0$ $\mathbb{P}$ fast sicher $\Longrightarrow Y_n \xrightarrow{\mathbb{P}} Y_0 \to Y_n \xrightarrow{\mathfrak{D}} Y_0$

5.2.3 Satz

Sind Y_n, $n \geq 1$, und Y_0 Zufallsvariablen (auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathbb{P})$), mit $Y_n \xrightarrow{\mathbb{P}} Y_0$, so gilt $Y_n \xrightarrow{\mathfrak{D}} Y_0$.

Beweis: Sei $x \in \mathbb{R}$ und $\varepsilon > 0$ beliebig. Dann folgt au der Alternative "$Y_0 - Y_n Y \varepsilon$" die Inklusion:

$$\{\omega : Y_n(\omega) \leq x\} \subset \{\omega : Y_0(\omega) \leq x + \varepsilon\} \cap \{\omega : Y_0(\omega) - Y_n(\omega) > \varepsilon\}$$

und damit

$$\mathbb{P}(Y_n \leq x) \leq \mathbb{P}(Y_0 \leq x + \varepsilon) + \mathbb{P}(Y_0 - Y_n > \varepsilon)$$

Wegen $Y_n \to Y_0$ konvergiert der zweite Summand gegen 0, so daß

$$\lim_{n\to\infty} \sup \mathbb{P}(Y_n \leq x) \leq \mathbb{P}(Y_0 \leq x + \varepsilon) \equiv F_0(x + \varepsilon)$$

Analog: $\lim_{n\to\infty} \inf \mathbb{P}(Y_n \leq x) \geq F_0(x - \varepsilon)$.
Ist also $x \in \mathfrak{C}(F_0)$, so folgt mit $F_n(x) = \mathbb{P}(Y_n \leq x)$:

$$\limsup_n F_n(x) \leq F_0(x) \leq \liminf_n F_n(x) \text{ , d.i.}$$

$$\lim F_n(x) = F_0(x)$$

Die Umkehrung ist nicht richtig!

5.2.4 Beispiele

Sei Y_0 $B(1, \frac{1}{2})$-verteilt und $Y_n = 1 - Y_0$ für alle $n \geq 1$. Dann ist jedes Y_n wieder $B(1, \frac{1}{2})$-verteilt und damit $Y_n \xrightarrow{\mathfrak{D}} Y_0$ (sogar $Y_n \overset{\mathfrak{D}}{=} Y_0$).
Y_n, $n \geq 1$ konvergiert aber nicht stochastisch gegen Y_0, denn für $\varepsilon \in (0, 1)$ ist

$$\mathbb{P}(|Y_n - Y_0| > \varepsilon) \overset{Y_n = 1 - Y_0}{=} \mathbb{P}(|1 - lY_0| > \varepsilon) = 1 \quad \forall\, n \geq 1$$

In 3.3.5 haben wir einen Stetigkeitssatz für diskrete Wahrscheinlichkeitsverteilungen kennengelernt. Es ist der Limes einer Folge von Wahrscheinlichkeitsfunktionen, d.h.

$$a_n = \lim_{n\to\infty} \rho_k^{(n)} \quad k = 0, 1, \ldots \tag{5.10}$$

genau dann, wenn der Limes der Folge der zugehörenden erzeugenden Funktion existiert. Zunächst stellen wir fest, daß die Aussage (5.10) eine Verteilungskonvergenz bedeutet.

5.2.5 Lemma von Schiffé

Sind Y_n, $n \geq 1$, und Y_0 $\mathbb{Z}_+$-wertige Zufallsvariablen und setzt man $\rho_k^{(n)} = \mathbb{P}(Y_n = k)$, $k \in \mathbb{Z}_+$, $n = 1, 2, \ldots$, so gilt $\rho_k^{(0)} = \lim_n \rho_k^{(n)}$ genau dann, wenn

$$\mathbb{P}(Y_0 \in \mathfrak{A}) = \lim_{n\to\infty} \mathbb{P}(Y_n \in \mathfrak{A})$$

in allen $A \in \mathbb{Z}$ [5].

Bemerkung: Setzt man $A = (-\infty, x]$, so hat man $Y_n \xrightarrow{\mathfrak{D}} Y_0$, $n \to \infty$.
In der allgemeinen Wahrscheinlichkeitstheorie wird der Stetigkeitssatz mit Hilfe der charakteristischen Funktionen $\varphi(n) = \mathbb{E}(e^{itY_n})$, $t \in \mathbb{R}$ formuliert[6].

5.2.6 Stetigkeitssatz

Seien Y_n, $n \geq 1$, eine Folge von Zufallsvariablen und φ_n die Folge der zugehörenden charakteristischen Funktionen.
Y_n ist verteilungskonvergent gegen eine Zufallsvariable Y_0 genau dann, wenn φ_n gegen eine Funktion φ_0 konvergiert, die an der Stelle 0 stetig ist. φ_0 ist dann charakteristische Funktion von Y_0: $\varphi(0) = \mathbb{E}(e^{itY_0})$, $t \in \mathbb{R}$

[5] Beweis siehe Schmitz '96, S.228
[6] vgl. 4.6.3

Beweis: Siehe Schmitz '96, S.244

Kurzfassung: $Y_n \xrightarrow{\mathfrak{D}} Y_0 \Longleftrightarrow \varphi(t) = \varphi_0(t),\ \forall\ t \in \mathbb{R}$. Die Stetigkeit von φ_0 bei 0 garantiert erst, daß φ_0 wieder charakteristische Funktion eine Zufallsvariablen ist[7].
Im folgenden Beispiel ist das nicht der Fall.

5.2.7 Beispiel

Y_n sei gleichverteilt auf $(-n, n)$. Dann gilt:

$$\varphi_n(t) = \left\{ \begin{array}{cc} \frac{\sin(nt)}{nt} & t \neq 0 \\ 1 & t = 0 \end{array} \right\}$$

und

$$\lim \varphi_n(t) = \left\{ \begin{array}{cc} 0 & t \neq 0 \\ 1 & t = 0 \end{array} \right\}$$

mit bei 0 unstetigen Grenzfunktionen. Für die Verteilungfunktion $F_n(x)$ von Y_0 gilt:

$$\lim F_n(x) = \lim_{n\to\infty} \left\{ \begin{array}{cc} 0 & x < -n \\ \frac{n+x}{2n} & x \in (-n, n) \\ 1 & x > n \end{array} \right\} = \frac{1}{2}$$

was **keine** Verteilungsfunktion darstellt. Es gibt *kein* Y_0 mit $Y_n \xrightarrow{\mathfrak{D}} Y_0$.
Statt $Y_n \xrightarrow{\mathfrak{D}} Y_0$, Y_0 $N(0,1)$-verteilt, schreibt man auch 'gemischt':

$$Y_n \xrightarrow{\mathfrak{D}} N(0,1)$$

Nun zeigen wir, daß die standardisierten Partialsummen S_n^* (nehmen jetzt die Rolle von Y_n ein) verteilungskonvergent gegen die $N(0,1)$-Verteilung sind.

5.2.8 Zentraler Grenzwertsatz von Lindberg-Lexy (Satz)

Gegeben sei eine Folge $X_1, X_2, \ldots$ von unabhängigen, indentisch verteilten[8] Zufallsvariablen aus $\mathfrak{L}_2$ ($\mu \equiv \mathbb{E}(X)$, $\sigma^2 \equiv Var(X_i) > 0$). Dann gilt für die Folge

$$S_n^* = \frac{(X_1 + \ldots + X_n)n\mu}{\sqrt{n}\sigma}$$

der standardisierten Partialsummen von X_n, $n \geq 1$, die Verteilungskonvergenz

$$S_n^* \xrightarrow{\mathfrak{D}} N(0,1) \quad (n \to \infty)$$

Beweis:

i) Ist $\varphi(t)$ die charakteristische Funktion von $X_i - \mu$ (für alle i dieselbe), so lautet die charakteristische Funktion:

$$\varphi_{S_n^*} = \varphi_n^* = \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^{n} X_i - \mu$$

$$\varphi_n^*(t) \overset{4.6.3-3}{=} \varphi_{\sum_{i=1}^n X_i - \mu}\left(\frac{t}{\sqrt{n}\sigma}\right) \overset{4.6.6}{=} \prod_{i=1}^{n} \varphi\left(\frac{t}{\sqrt{n}\sigma}\right) = \left(\varphi\left(\frac{t}{\sqrt{n}\sigma}\right)\right)^n$$

[7] siehe 4.6.3, 5.
[8] strenge Bedingung

ii) Taylorentwicklung von $\varphi(t)$ an der Stelle $t = 0$:

$$\varphi(t) = 1 + \varphi'(0) \cdot t + \frac{1}{2}\varphi''(0) \cdot t^2 + r_2(t)$$

mit $\frac{r_2(t)}{t^2} \to 0$ bei $t \to \infty$.
Nach 4.6.7 ist

$$\varphi'(0) = i \cdot \mathbb{E}(X_i - \mu) = 0$$
$$(*) \quad \varphi''(0) = -\mathbb{E}(X_i - \mu)^2 = -\sigma^2,$$

so daß

$$\varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + r_2(t)$$

iii) Das φ_n^* aus Teil i) lautet mit Formel (*):

$$\varphi_n^*(t) = [1 - \frac{1}{2}\frac{t^2}{n} + r_2(\frac{t}{\sqrt{n}\sigma})]^n = (1 - \frac{t^2}{2n}(1 + a(t))^n$$

mit

$$a(t) = \frac{r_2(\frac{t}{\sqrt{n}\sigma})}{\frac{t^2}{2n}} \longrightarrow 0 \quad \text{für } t \to \infty$$

Es folgt mit einem ε-Argument wie in 3.3.6:

$$\varphi_n^*(t) \overset{n\to\infty}{\longrightarrow} e^{\frac{t^2}{2}} \quad \forall\, t \in \mathbb{R}$$

Nach 4.6.4 ist die charakteristische Funktion der $N(0,1)$-Verteilung, so daß der Stetigkeitssatz 5.2.6 zusammen mit dem Eindeutigkeitssatz 4.6.5 die Behauptung liefern.

Bemerkungen:

1. Im Spezialfall unabhängiger, $N(\mu, \sigma^2)$-verteilter X_i ist gemäß Beispiel 4.6.6 jeder S_n^* $N(0,1)$-verteilt, so daß hier sogar Gleichheit $F_{S_n} = \phi$ für jedes n gilt.

2. Im zentralen Grenzwertsatz kann die unabhängig-Voraussetzung nicht ersatzlos gestrichen werden. Als Gegenbeispiel wähle man identische $X_1 = X_2 = \dots$.

3. **Anwendungsbeispiel:** Gewinnung von $N(0,1)$-verteilten Zufallsvariablen aus $U[0,1]$-verteilten Zufallsvariablen.
   Sind $X_1, X_2, \dots$ unabhängig und $[0,1]$ gleichverteilt, so ist wegen $\mu = \frac{1}{2}$, $\sigma^2 = \frac{1}{12}$ (vgl. 4.5.11, 1.):
   $$\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$
   approximiert $N(0,1)$-verteilt ($S_n = X_1 + \ldots + X_n$).
   Für $n = 48$ ist $\frac{S_n - 24}{2}$ angenähert $N(0,1)$-verteilt.