Wahrscheinlichkeitstheorie und Statistik
SS2003
Peter Meyer-Nieberg

Mitschrift von
Sebastian Bitzer

16. Juli 2003

Missionsstatement bzw. Vorwort

Ich habe es mir zur Aufgabe gestellt, meine Aufzeichnungen aus der Vorlesung so gut wie mir möglich nach LaTeXzu überführen.

Dabei sind mehrere Probleme aufgetreten, die jedem Leser dieser Mitschrift bewusst sein sollten:

- ich kann etwas falsch abgeschrieben haben
- ich kann etwas falsches abgeschrieben haben
- die Struktur an der Tafel, stimmte nicht direkt mit der Struktur des richtigen Skriptes überein
- die Ausführungen von der Tafel reichen mitunter nicht aus um die Zusammenhänge zu verstehen
- ich könnte beim Aufschreiben etwas falsch interpretiert und damit falsch dargestellt haben
- ich stelle keinen Anspruch an Vollständigkeit

Natürlich habe ich versucht alles fehlerfrei ab- bzw. aufzuschreiben, da sich allerdings einige Dinge meinem Verständnis entziehen, kann ich keine Garantien auf Richtigkeit geben.

Da, wo es mir möglich war, habe ich versucht erklärende Kommentare zu geben bzw. den Punkt ausführlicher aufzuschreiben.

Als Begleitliteratur hat sich Krengel [**?**] nur bedingt als nützlich erwiesen auch wenn ich mir eher eine Vorlesung gewünscht hätte, die darauf aufbaute. Insgesamt wurde das Thema abstrakter und mit weniger Blick auf die Statistik als in [**?**] behandelt. Dafür sind sicherlich die Bücher von Bauer ([**?**] + [**?**]) besser geeignet. Auch die alte Ausgabe [**?**], die noch beide Bücher in einem enthält und in der Bibliothek als Lehrbuch zu erhalten ist, sollte als Begleitliteratur zu gebrauchen sein.

Inhaltsverzeichnis

1 Kombinatorik und diskrete Wahrscheinlichkeitsräume 3

2 Bedingte Wahrscheinlichkeit 11

2.1 Unabhängige Ereignisse 13

3 Zufallsvariablen auf diskreten Wahrscheinlichkeitsräumen 16

4 Maße und Wahrscheinlichkeitsmaße 20

4.1 Kontinuierliche Theorie 23

4.2 Unabhängigkeit von Ereignissen und Zufallsvariablen 29

5 Grenzwertsätze und Produkträume 32

6 Erwartungswert, Varianz und Verteilungen 36

7 Grenzwertaussagen 42

7.1 Zentraler Grenzwertsatz und Normalverteilung 44

8 Parameterschätzungen, grundlegende Statistik 49

8.1 Parameterschätzungen 51

8.2 Maximum-Likelihood-Schätzer 53

1 Kombinatorische Probleme und diskrete Wahrscheinlichkeitsräume

Bemerkung 1.1 *$A = a_1, \dots, a_n$*
A^r: r-fache kartesische Produkt aller r-Tupel $(a_{i_1}, \dots, a_{i_r}) 1 \leq i \leq n$
$\Rightarrow Anz(A^r) = |A^r| = n^r$ (klar, ... Induktion)

Folgerung 1.2 *A, B Mengen, $Anz(A) = r, Anz(B) = n$, dann existieren genau n^r Abbildungen $\varphi : A \to B$*

Beweis $(\varphi : A \to B) \overset{\psi}{\mapsto} (\varphi(a_1), \dots, \varphi(a_r)) \in B^r$
B^r hat n^r Elemente (siehe 1.1) und da ψ injektiv [1] + surjektiv [2]
$\Rightarrow$ gibt es auch n^r Elemente $(\varphi : A \to B)$

Bemerkung 1.3 *$A = \{a_1, \dots, a_n\} \quad (n \in \mathbb{N})$*
Sei $r \in \mathbb{N}$ mit $r \leq n$ setze $A^{(r)} = \{(b_1, \dots, b_r) \in A^r | b_1, \dots, b_r \in A$ paarweise verschieden$\}$

$$\Rightarrow Anz(A^{(r)}) = \frac{n!}{(n-r)!}$$

Beweis Induktion:
$r = 1$: ist klar $(\frac{n!}{(n-1)!} = n)$
$r \to r + 1$: fr den Fall $r + 1 \leq n$ (macht ja sonst keinen Sinn) Seien $(b_1, \dots, b_n) \in A^{(r)}$
Für die Wahl eines $b_{r+1} \in A$ mit $(b_1, \dots, b_r, b_{r+1}) \in A^{(r+1)}$ gibt es $(n - r)$ Möglichkeiten [3]
$\Rightarrow Anz(A^{(r+1)}) = Anz(A^{(r)}) \cdot (n - r) = \frac{n!}{(n-r)!} \cdot (n - r) = \frac{n!}{(n-(r+1))!}$ □

Anmerkung: Jedes r-Tupel $(b_1, \dots, b_r) \in A^{(r)}$ heißt auch eine r-Permutation der Elemente $\{a_1, \dots, a_n\}$

Folgerung 1.4 *i) $Anz(A) = r \in \mathbb{N}$, $Anz(B) = n \in \mathbb{N}$ mit $n \geq r$*
$\Rightarrow$ Es ex. genau $\frac{n!}{(n-r)!}$ injektive Abb. $\varphi : A \to B$

---

[1] ändert sich die Abbildung so ändert sich auch der Vektor in B^r - $\varphi(a_1)$ ist dann z.B. verschieden vom vorigen $\varphi(a_1)$

[2] man kann sich für jedes Element in B^r eine Abbildung vorstellen, die dieses Element erzeugt

[3] Forderung: $b_{r+1} \neq b_1, \dots, b_r$, r verschiedene Elemente aus A wurden schon genommen, bleiben nur noch $(n - r)$ übrig

ii) $Anz(A) = Anz(B) = n \in \mathbb{N}$
$\Rightarrow$ *Es ex. genau* $n!$ *bijektive Abb.* $\varphi : A \to B$

Beweis

i) $(\varphi : A \to B) \overset{\psi}{\mapsto} (\varphi(a_1), \ldots, \varphi(a_r)) \in B^{(r)}$ [4]
ψ ist injektiv + surjektiv

ii) $n = r$✓[5]
φ bijektiv $\iff$ φ injektiv $\iff$ φ surjektiv, da $Anz(A) = Anz(B) = n$

Satz 1.5 $Anz(A) = n \in \mathbb{N}$ *Sei* $r \in \mathbb{N}$ *mit* $r \leq n$.
Es ex. genau $\binom{n}{r}$ *Teilmengen* $B \subset A$ *mit* $Anz(B) = r$

Beweis $B \subset A$ mit $Anz(B) = r$
Es ex. genau $r!$ bijektive Abb. $\varphi : \{1, \ldots, r\} \to B$ (1.4 *ii)*)
Es ex. $\frac{n!}{(n-r)!}$ injektive Abb. $\varphi : \{1, \ldots, r\} \to A$ (1.4 *i)*)

Mit anderen Worten: Es gibt $\frac{n!}{(n-r)!}$ verschiedene Zuordnungen von Zahlen $1, \ldots, r$ zu r Elementen von A, die somit eine Menge $B \subset A$ bilden, für die gilt: $Anz(B) = r$. Allerdings enthält das auch Zuordnungen von $1, \ldots, r$ zu ein und der selben Menge B [6]. Nun, wie viele solche Mehrfachzuordnungen gibt es? Genau so viele wie verschiedene injektive Abb. zwischen gleich großen Mengen ($r!$). Wir müssen also für jede Menge B $r!$ Zuordnungen abziehen:

$$\Rightarrow \frac{n!}{(n-r)!} \cdot \frac{1}{r!} = \binom{n}{r} \quad \square$$

Bezeichnung: Auswahl einer k-elementigen Teilmenge aus einer n-elementigen Menge ($n > k$) $\equiv$ Entnahme einer ungeordneten Probe vom Umfang k aus einer n-elementigen Menge ohne Wiederholungen.

Satz 1.6 *A mit $Anz(A) = n \in \mathbb{N}$, $k \in \mathbb{N}$*
Man kann $\binom{n+k-1}{k}$ *unterschiedliche Proben vom Umfang k mit Wiederholungen aus A entnehmen.*

[4] da φ injektiv, sind alle $\varphi(a_k)$ verschieden und somit $(\varphi(a_1), \ldots, \varphi(a_r)) \in B^{(r)}$
[5] $\frac{n!}{(n-n)!} = n!$
[6] mal könnte z.B. die 1 auf ein Element von B zeigen und ein anderes mal könnte die 1 auf ein anderes Element von B zeigen ohne, dass sich B dabei ändert

Beweis oBdA sei $A = \{1, \dots, n\}$
Definiere auf A^r eine Äquivalenzrelation $(a_1, \dots, a_k) \sim (b_1, \dots, b_k)$
$\vec{a} \sim \vec{b} \iff$ Es ex. eine bijektive Abb. $\varrho : \{1, \dots, k\} \to \{1, \dots, k\}$ mit $b_j = a_{\varrho_j} \quad j = 1, \dots, k$
$\equiv$ Umordnung der Elemente (Permutationen)
$\Rightarrow$ oBdA ist erreichbar, dass $a_1 \leq a_2 \leq \dots \leq a_k$ (Schreibweise: $[a_1, \dots, a_k]$)

Setze $A_n^k = \{[a_1, \dots, a_n] | a_1 \leq \dots \leq a_n\}$, [7] $C = \{1, 2, \dots, n + k - 1\}$

$\phi : A_n^k \to \{B \in \mathcal{P}(C)$[8]$| Anz(B) = k\}$
$\phi([a_1, a_2, \dots, a_k]) = \{a_1, a_2 + 1, a_3 + 2, \dots, a_k + k - 1\} \subset C$
da ϕ injektiv + surjektiv und alle Elemente $\{a_1, a_2 + 1, a_3 + 2, \dots, a_k + k - 1\}$ unterschiedlich [9] kann man Satz 1.5 anwenden ($Anz(C) = n + k - 1$, $Anz(B) = k$):

$$\Rightarrow Anz(A_n^k) = \binom{n+k-1}{k} \quad \square$$

In A^k (ohne Reihenfolge) kommen diese Proben unterschiedlich oft vor (siehe 7)

Weitere Situation Menge Ω, $Anz(\Omega) = n \in \mathbb{N}$, $A \subset \Omega$ dann heit A Ereignis
$P : \mathcal{P}(\Omega) \to [0, 1]$, $P(\Omega) = 1$, $P(\emptyset) = 0$

$$\boxed{A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)}$$

$A = \{\omega\} \equiv$ Elementarereignis $P(\omega) = p_\omega = P(\{\omega\})$
$A \subset \omega \Rightarrow P(A) = \sum_{\omega \in A} p_\omega, \quad \sum_{\omega \in \Omega} p_\omega = 1$
$P(A) = 1 \equiv$ sicheres Ereignis
$P(A) = 0 \equiv$ unmögliches Ereignis

Bemerkung 1.7 $Anz(\Omega) \in \mathbb{N}, \quad A, B, A_1, \dots, A_n \in \mathcal{P}(\Omega)$
$\Rightarrow$

[7] A_n^k enthält somit alle geordneten Elemente aus A^k, z.B. $(1, 1, 2, 3)$ aber nicht $(1, 2, 3, 1)$, deren Anzahl der der unterschiedlichen Proben vom Umfang k mit Wiederholungen entspricht
[8] $\mathcal{P}(C)$ ist die Potenzmenge von C
[9] da wir wissen, dass $a_1 \leq a_2$, muss $a_1 < a_2 + 1$ sein

i) $P(A^c) = 1 - P(A) \quad A^c = \{\omega \in \Omega | \omega \notin A\}$

ii) $A \subset B \Rightarrow P(A) \leq P(B)$

iii) $P(A \setminus B) = P(A) - P(B \cap A)$, $P(A \cup B) + P(A \cap B) = P(A) + P(B)$

iv) $P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n)$

v) $A_i \cap A_j = \emptyset | i \neq j \Rightarrow P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$

vi) $P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} p_\omega$

Beweis *iii)*
$A = (A \cap B) \cup (A \setminus B)$, $(A \setminus B) \cap (A \cap B) = \emptyset$
$\Rightarrow P(A) = P(A \cap B) + P(A \setminus B)$
$P(A \cup B) = P(B \cup (A \setminus B)) = P(B) + P(A \setminus B) =$ [10] $P(A) + P(B) - P(A \cap B)$ $\square$

10.04.

Bemerkung 1.8

Laplace-Wahrscheinlichkeitsraum: eine endliche Menge, jedes Element dieser Menge ist gleich wahrscheinlich
$\Omega \quad Anz(\Omega) = n \quad A \subset \Omega$

$$P(A) = \frac{Anz(A)}{n} \equiv \text{Gleichverteilung}$$

Beispiel 1.9 *4 Würfel, man bestimme die Wahrscheinlichkeit, dass die erscheinenden Augenzahlen alle unterschiedlich sind (bei einem Wurf mit allen Würfeln)*

$\Omega_0 = \{1, 2, 3, 4, 5, 6\}$ Ergebnisse des Werfens eines Würfels
4 Würfel: $\Omega = \Omega_0^4$ =alle Tupel $\{(i_1, i_2, i_3, i_4) | i_1, \ldots, i_4 \in \{1, \ldots, 6\}\}$
$Anz(\Omega) = 6^4$
Alle Augenzahlen unterschiedlich (1.3) $Anz(\Omega_0^{(4)} = \frac{6!}{2!}$

$$p = \frac{6!}{2! 6^4} = \frac{5}{18}$$

[10] $P(A \setminus B) = P(A) - P(A \cap B)$

Beispiel 1.10 *Zahlenlotto 6 aus 49*

(1.5) $\Rightarrow \binom{49}{6}$ unterschiedliche Arten aus $\{1,\dots,49\}$ ungeordnete Proben vom Umfang 6 zu entnehmen

$$p_\omega = \frac{1}{\binom{49}{6}} \equiv \text{Wahrscheinlichkeit für}\,'6\,\text{Richtige}'$$

$1 \le j < 6$, gesucht: Wahrsch. p_j für 'j Richtige':
$\{\omega_1,\dots,\omega_6\} \ni j$-Zahlen $\to \binom{6}{j}$ [11]
$\Omega \setminus \{\omega_1,\dots,\omega_6\} \ni (6-j)$-Zahlen $\to \binom{43}{6-j}$ [12]

$$\Rightarrow p_j = \frac{\binom{6}{j}\binom{43}{6-j}}{\binom{49}{6}}$$

Bemerkung 1.11 $\Omega = \Omega_1 \cup \Omega_2\, mit\, \Omega_1 \cap \Omega_2 = \emptyset$
$Anz(\Omega_1) = n_1,\ Anz(\Omega_2) = n_2,\ n = n_1 + n_2 = Anz(\Omega)$
$m_1, m_2 \in \mathbb{N}_0 \quad m_i \le n_i \quad i = 1, 2$

i) *es ex. genau* $\binom{n_1}{m_1}\binom{n_2}{m_2}$ *Teilmengen* $A \in \Omega$ *mit* $Anz(A \cap \Omega_i) = m_i \quad i = 1, 2$

ii) $m = m_1 + m_2$, *Gleichverteilung* $\Rightarrow$ *Wahrscheinlichkeit der Auswahl einer Teilmengen* $A \in \Omega$, $Anz(A) = m$, $Anz(A \cap \Omega_i) = m_i$ *ist*

$$h(m_1, m_2, n_1, n_2) = \frac{\binom{n_1}{m_1}\binom{n_2}{m_2}}{\binom{n}{m}} \equiv \text{hypergeometrische Verteilung}$$

Begründung: Zahlenlotto

Beispiel 1.12 *Skat*

ges.: Wahrscheinlichkeit, dass Spieler I genau 3 Buben hat
$\Omega_1 \equiv$ Menge der Buben
$n_1 = 4 \quad m_1 = 3$
$n_2 = 28 \quad m_2 = 10$[13] $- 3 = 7$

$$p = \frac{\binom{4}{3}\binom{28}{7}}{\binom{32}{10}} = \frac{66}{899} = 0,0734$$

[11] wähle j-Zahlen aus den 6 Richtigen
[12] die restlichen der 6 gezogenen Zahlen kommen aus den 43 'Falschen'
[13] beim Skat bekommt jeder Spieler 10 Karten

Bemerkung 1.13 $n, n_1, \dots, n_r \in \mathbb{N} \quad r \in \mathbb{N}, r \geq 2 \quad Anz(\Omega) = n = n_1 + \dots + n_r$

Es ex. genau $\binom{n}{n_1, \dots, n_r} := \frac{n!}{n_1! \dots n_r!}$ *unterschiedliche Tupel* $(A_1, \dots, A_r)$ *von Teilmengen* $A_i \in \Omega$ *mit* $Anz(A_i) = n_i$, $A_i \cap A_j = \emptyset$ *für* $i \neq j$

$\rightarrow$ *Multinomialkoeffizienten*

Wir haben hier eine Menge Ω gegeben, die genau n Elemente hat. Des weiteren haben wir noch gegeben wie viele Teilmengen Ω hat (r) und wie viele Elemente jede Teilmenge haben soll ($n_1, \dots, n_r$).
Gesucht ist dann, wie viele Möglichkeiten es gibt, die verschiedenen Elemente von Ω auf die Teilmengen zu verteilen.
Zum Beispiel könnte Ω die Zahlen $1, \dots, 4$ enthalten und wir suchen die Anzahl der Möglichkeiten, wie man die 4 Zahlen auf eine Teilmenge mit 3 Elementen und eine Teilmenge mit 1 Element verteilen kann. Die Berechnung gibt an:

$$\frac{4!}{3! \cdot 1!} = 4$$

Das entspricht folgenden Verteilungen:

$A_1(n_1 = 3)$	$A_2(n_2 = 1)$
$1, 2, 3$	4
$2, 3, 4$	1
$3, 4, 1$	2
$4, 3, 2$	3

Dabei scheint die Reihenfolge innerhalb des Tupels von Teilmengen (A_1, A_2) festgelegt zu sein bzw. nicht zu interessieren.

Beweis $r = 1$✓ [14]

$r \rightarrow r+1: \quad n_0, n_1, \dots, n_r, n = n_0 + \underbrace{n_1, \dots, n_r}_{m} \Rightarrow m < n$

$A_0 \subset \Omega$ fest, $Anz(A_0) = n_0$ fest

$\Omega_0 = \Omega \setminus A_0 \quad Anz(\Omega_0) = m = n_1 + \dots + n_r$

Es ex. $\binom{m}{n_1, \dots, n_r}$ unterschiedliche Tupel $(A_1, \dots, A_r)$, $Anz(A_j) = n_j$ mit $j = 1, \dots, r$, $A_i \cap A_j = \emptyset$ $i \neq j$, $A_j \subset \Omega_0$

Insgesamt sind es dann:

$$\binom{n}{n_0}\binom{m}{n_1, \dots, n_r} = \frac{n!}{n_0! \cancel{(n - n_0)!}} \cdot \frac{\cancel{m!}}{n_1! \dots n_r!} = \binom{n}{n_0, n_1, \dots, n_r} \quad \square$$

[14] $\frac{n!}{n_1!} = 1$

Beispiel 1.14

Wie lange muss gewürfelt werden, damit sicher einmal 6 erscheint?
Unexakt: Es gibt keine richtige Antwort auf diese Frage mit $k \in \mathbb{N}$. Gibt es sie für $k \to \infty$?
Eine halbwegs richtige Antwort (approximativ) erfordert $\Omega = \mathbb{N}$

$P : \mathcal{P}(\mathbb{N}) \to [0,1]$, $P(\mathbb{N}) = 1$, $P(\emptyset) = 0$

Zusatzforderung: $p_n = P(\{n\}) \in [0,1]$ $\boxed{\sum_{n=1}^{\infty} p_n = 1}$

$A \cap B = \emptyset \Rightarrow p(A \cup B) = P(A) + P(B)$
Es gelten dann auch die üblichen Eigenschaften wie in 1.7.
(A_n) disjunkt $P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$, da die absolute Konvergenz gegeben ist (indifferentes Ergebnis gegenüber Umordnung)

Bemerkung 1.15 *Bernoulli-Experiment:*
$\Omega_0 = \{a, b\}$, $0 < p < 1$
$p = P(\{a\})$ *Wahrscheinlichkeit des Erfolges*
$q = 1 - p = P(\{b\})$ *Wahrscheinlichkeit des Misserfolges*
n*-mal unabhängig wiederholt:* $\Omega = \Omega_0^n$

Folgerung 1.16

Frage: Wie hoch ist die Wahrscheinlichkeit, dass man bei n Wiederholungen j-mal $(0 \leq j \leq n)$ Erfolg hat?
Anzahl der a in $(c_1, \ldots, c_n) \in \Omega = j$, Anzahl der $b = n - j$
Es gibt also $\binom{n}{j}$ verschiedene Möglichkeiten an welcher Stelle im Vektor $(c_1, \ldots, c_n)$ j a vorkommen.

$$\Rightarrow P(j\text{−mal Erfolg}) = \binom{n}{j} p^j q^{n-j} = b(n, j, p) \equiv \underline{\text{Binomialverteilung}}$$

$\sum_{j=0}^{n} b(n, j, p) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} = (p + q)^n = 1^n = 1$

Frage: Wie wahrscheinlich ist es, dass genau bei der j-ten Wiederholung der erste Erfolg eintritt? $(\equiv P(A))$

$$\underbrace{(b, \ldots, b, a)}_{j} \rightsquigarrow P(A) = q^{j-1} p,\ j \in \mathbb{N} \quad \equiv \underline{\text{geometrischeVerteilung}}$$

$\sum_{j=1}^{\infty} q^{j-1} p = p \sum_{r=0}^{\infty} q^r = p \cdot \frac{1}{1-q} = 1$

$$\sum_{j=1}^{n} q^{j-1} p = \frac{1-q^n}{1-q} \cdot p$$

$\equiv$ Wahrscheinlichkeit, dass bis zum n-ten Wurf mindestens einmal p gekommen ist

14.04.

Bemerkung 1.17 *negative Binomialverteilung $\equiv$ Pascal-Verteilung*
$\Omega_0 = \{a, b\}$, $p = P(\{a\}) = Erfolgswahrscheinlichkeit$
$r, k \in \mathbb{N} \quad r + k = n\, Wiederholungen : \Omega^n$
$f(k, r, p)$ ist die Wahrscheinlichkeit, dass dem r-ten Erfolg genau k-Misserfolge vorausgehen

$$\Rightarrow f(k, r, p) = \binom{k+r-1}{k} p^r q^k$$

Situation: $(n = r+k)$-te Stelle ist a, sonst gibt es noch $r+k-1$ Stellen, in denen k-mal b vorkommt und $(r-1)$-mal a. Alle Möglichkeiten aus $r+k-1$ Stellen k zu ziehen:

$$\binom{r+k-1}{k} = \binom{r+k-1}{r-1} \Rightarrow \text{Wert} : \binom{r+k-1}{k} p^r q^k$$

Bemerkung 1.18 *Poisson-Verteilung*

$$e^{-\alpha} \frac{\alpha^k}{k!}, \quad \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} = 1$$

$\Rightarrow$ Approximation der Binomialverteilung bei sehr kleinem p für $\alpha = p \cdot n$
Rechtfertigung:
Binomialverteilung $\binom{n}{k} p^k q^{n-k}$, für sehr kleines p ist q fast 1, kürze $(n-k)!$ aus Binomialkoeffizient: $\frac{n(n-1)...(n-k+1)}{1 \cdot 2 ... k} p^k$

der endliche Raum der Binomialverteilung wird durch den unendlichen Raum der Poissson-Verteilung ersetzt $\Rightarrow$ besser berechenbar

2 Bedingte Wahrscheinlichkeit

$\Omega \neq \emptyset$, abzählbar, $\sum_{\omega \in \Omega} p_\omega = 1$, $p_\omega \geq 0$, $P(A) = \sum_{\omega \in A} p_\omega$

Definition 2.1 *Bedingte Wahrscheinlichkeit*
$B \in \mathcal{P}(\Omega)$, $P(B) > 0$, $A \in \mathcal{P}(\Omega)$
Die bedingte Wahrscheinlichkeit des Eintreffens von A unter der Hypothese B:

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

Beispiel 2.2

2 Schränke (gleiches Aussehen) haben jeweils 4 Schubladen
Schrank I : in jeder Schublade ist eine Kugel
Schrank II: in einer Schublade ist eine Kugel, in den anderen sind keine
Nun wird eine Schublade geöffnet, darin liegt eine Kugel. Wie groß ist die Wahrscheinlichkeit, dass diese Schublade in Schrank I war?

B ist das Ereignis, dass eine Schublade mit einer Kugel geöffnet wurde
A ist das Ereignis, dass eine Schublade aus Schrank I geöffnet wurde
LaPlace-Raum:

$$P(B) = \frac{5}{8},\ P(A) = \frac{1}{2},\ P(A \cap B) = \frac{4}{8} \qquad \Rightarrow P(A|B) = \frac{4}{5}$$

Satz 2.3

i) $B \in \mathcal{P}(\Omega)$, $P(B) > 0$
$\mathcal{P}(\Omega) \ni A \mapsto P(A|B)$ *ist eine Wahrscheinlichkeit*

ii) $B_1, \ldots, B_r \in \mathcal{P}(\Omega)$ *paarweise disjunkt (unvereinbar)*
$\Omega = B_1 \cup \cdots \cup B_r$ *totale Wahrscheinlichkeit*

$$P(A) = \sum_{k=1}^{r} P(A|B_k) P(B_k)$$

Festetzung: $P(A|B_k) = 0$ *falls* $P(B_k) = 0$

iii) ***Formel von Bayes***
$A \in \mathcal{P}(\Omega)$, $P(A) > 0$
Voraussetzung: ii)

$$P(B_i|A) = \frac{P(B_i) P(A|B_i)}{\sum_{k=1}^{r} P(A|B_k) P(B_k)}$$

Beweis

i) dies ist leicht nachzuprüfen und ist nur Schreibarbeit

ii) $A = A \cap \Omega = A \cap (B_1 \cup \cdots \cup B_r) = (A \cap B_1) \cup \cdots \cup (A \cap B_r)$
$\Rightarrow P(A) = \sum_{k=1}^{r} P(A \cap B_k)$
$= \sum_{k=1}^{r} P(A|B_k)P(B_k)$

iii) $$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)} \overset{ii)}{=} \frac{P(A|B_i)P(B_i)}{\sum_{k=1}^{r} P(A|B_k)P(B_k)}$$

Beispiel 2.4 *Skat*

Spieler I hat zwei Buben.
Gesucht ist die Wahrscheinlichkeit, dass die anderen Spieler jeweils einen Buben haben.
Ereignisse:
$A_1 \equiv$ I hat genau zwei Buben
$A_2 \equiv$ II hat genau einen Buben
$A_3 \equiv$ III hat genau einen Buben

$$P(A_1) = \frac{\binom{4}{2}\binom{28}{8}}{\binom{32}{10}} \quad P(A_2|A_1) = \frac{\binom{2}{1}\binom{20}{9}}{\binom{22}{10}} \quad P(A_3|A_1 \cap A_2) = \frac{\binom{1}{1}\binom{11}{9}}{\binom{12}{10}}$$
$$P(A_2 \cap A_3|A_1) = P(A_3|A_1 \cap A_2) \cdot P(A_2|A_1) = 0,4329$$

17.04.

Beispiel 2.5

Urne I : 4 rote, 2 schwarze Kugeln
Urne II: 8 rote, 12 schwarze Kugeln
Eine Urne werde zufällig ausgewählt und aus dieser dann eine Kugel gezogen.
Frage: Wie groß ist die Wahrscheinlichkeit, dass diese Kugel rot ist?
Falsch wäre eine reine Betrachtung der farbigen Kugeln: 12 rote, 14 schwarze
$\cancel{\Rightarrow p = \frac{12}{26}}$
Richtig: B_i $(i = 1, 2)$ sei das Ereignis, dass Urne i ausgewählt wurde, $P(B_1) = P(B_2) = \frac{1}{2}$
A sei das Ereignis, dass die Kugel rot ist.

$$\Rightarrow P(A|B_1) = \frac{2}{3} \quad P(A|B_2) = \frac{2}{5} \qquad \Rightarrow P(A) = \frac{2}{3} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2} = \frac{8}{15}$$

2.1 Unabhängige Ereignisse

Vorbetrachtung:
2 Experimente, diese sollen nacheinander ablaufen, keine gegenseitige Beeinflussung
1. Experiment: Ω_1, P_1, $\mathcal{P}(\Omega_1)$, $\omega_1 \in \Omega_1$
2. Experiment: Ω_2, P_2, $\mathcal{P}(\Omega_2)$, $\omega_2 \in \Omega_2$
$\Omega_1 \times \Omega_2$ Produktraum

$$p_{\omega_1,\omega_2} = p_{\omega_1} \cdot p_{\omega_2}$$

$A_1 \in \Omega_1,\ A_2 \in \Omega_2$

$$P(A_1 \times A_2) = \sum_{\omega_1,\omega_2 \in A_1 \times A_2} p_{\omega_1,\omega_2}$$

$B_1 = A_1 \times \Omega_2 \quad \Rightarrow$ [15] $P(B_1) = P_1(A_1)$
$B_2 = \Omega_1 \times A_2 \quad \Rightarrow P(B_2) = P_2(A_2)$

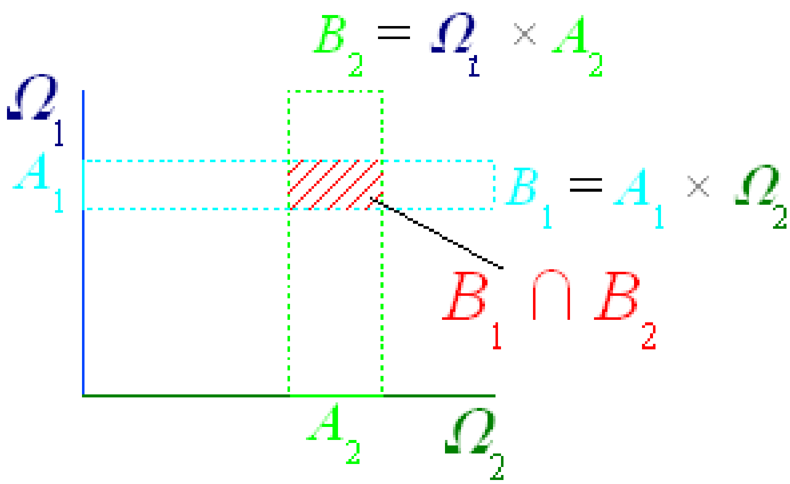

Abbildung 1: Übersicht Produkträume

$$\begin{aligned} P(B_1 \cap B_2) &= \textstyle\sum_{(\omega_1,\omega_2) \in B_1 \cap B_2} p_{\omega_1,\omega_2} = \sum_{\substack{\omega_1 \in A_1 \\ \omega_2 \in A_2}} p_{\omega_1} p_{\omega_2} \\ &= \textstyle\sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} p_{\omega_1} p_{\omega_2} = P_1(A_1) \cdot P_2(A_2) \\ &= P(B_1) \cdot P(B_2) \end{aligned}$$

[15] $P(A_1 \times \Omega_2) = \sum_{\omega_1,\omega_2 \in A_1 \times \Omega_2} p_{\omega_1,\omega_2} = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in \Omega_2} p_{\omega_1} \cdot p_{\omega_2} = \sum_{\omega_1 \in A_1} p_{\omega_1} \cdot \sum_{\omega_2 \in \Omega_2} p_{\omega_2} = \sum_{\omega_1 \in A_1} p_{\omega_1} \cdot 1$

Definition 2.6 *Unabhängigkeit von Ereignissen*
Ω, $\mathcal{P}(\Omega)$, P *diskreter Wahrscheinlichkeitsraum*
$A, B \subset \mathcal{P}(\Omega)$ *unabhängig* $\iff P(A \cap B) = P(A) \cdot P(B)$

Folgerung 2.7
$P(B) > 0$
$A, B \subset \mathcal{P}(\Omega)$ *unabhängig* $\iff P(A|B) = P(A)$

Beweis
"$\Rightarrow$" $P(A|B) = \frac{P(A\cap B)}{P(B)} \overset{2.6}{=} \frac{P(A)\cdot P(B)}{P(B)} = P(A)$
"$\Leftarrow$" $P(A|B) = \frac{P(A\cap B)}{P(B)} \overset{Vor.}{\Leftrightarrow} P(A) = \frac{P(A\cap B)}{P(B)} \Leftrightarrow P(A) \cdot P(B) = P(A \cap B)$ $\square$

Definition 2.8
$\emptyset \neq \mathcal{A}, \mathcal{B} \subset \mathcal{P}(\Omega)$ *unabhängig* $\iff A \in \mathcal{A},\ B \in \mathcal{B} \Rightarrow P(A \cap B) = P(A) \cdot P(B)$

Beispiel: $\Omega = \Omega_1 \times \Omega_2 \qquad \mathcal{A} = A_1 \times \Omega_2 | A_1 \subset \Omega_1 \qquad \mathcal{B} = \Omega_1 \times B_2 | B_2 \subset \Omega_2$

Beispiel 2.9 *Würfel, zweimal gewürfelt*

1. $A \equiv$ Augenzahl des ersten Wurfes ist gerade
   $B \equiv$ Augenzahl des zweiten Wurfes $= 1$
   $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{6}$
   $P(A \cap B) = P((2,1), (4,1), (6,1)) = \frac{3}{36} = \frac{1}{12} = P(A) \cdot P(B)$

2. $A \equiv$ Summe der Augenzahlen beider Würfel ist gerade
   $B \equiv$ Augenzahl des zweiten Wurfes ist gerade
   $P(A)$[16] $= \frac{1}{2} = P(B) \qquad P(A \cap B)$[17] $= \frac{1}{4} = P(A) \cdot P(B)$

3. $A \equiv$ Summe der Augenzahlen beider Würfel ist gerade und ≤ 8
   $B \equiv$ zweiter Wurf $= 4$ oder 6
   $P(A) = \frac{14}{36}$, $P(B) = \frac{1}{3}$
   $P(A \cap B) = \frac{1}{12} \Rightarrow$ nicht unabhängig

Bemerkung 2.10
$\mathcal{A} \subset \mathcal{P}(\Omega)$ *System unabhängiger Ereignisse*
$\equiv A, B \in \mathcal{A},\ A \neq B \Rightarrow P(A \cap B) = P(A)P(B)$
exakter: $A_1, \ldots, A_n \in \mathcal{A} \Rightarrow P(\bigcap A_j) = \prod P(A_j)$

[16] die Augenzahlen der zwei Würfe müssen entweder beide gerade oder ungerade sein, damit die Summe gerade ist
[17] in der Hälfte aller Durchgänge ist die erste Augenzahl gerade, wiederum bei der Hälfte davon ist die zweite Augenzahl gerade

i) $\mathcal{A}_0 = \mathcal{A} \cup \{\emptyset, \Omega\}$ *ist ein System unabhängiger Ereignisse*

ii) $\mathcal{A} = \{A_i | i \in J\} \quad B_i \in \{A_i, A_i^c, \emptyset, \Omega\} \quad P(A^c) = 1 - P(A)$
$\Rightarrow \mathcal{B} = \{B_i | i \in J\}$ *ist ein System unabhängiger Ereignisse*

24.04.

3 Zufallsvariablen auf diskreten Wahrscheinlichkeitsräumen

$\Omega \neq \emptyset,\ \mathcal{P}(\Omega),\ \Omega \ni \omega \mapsto p_\omega \in [0,1],\ \sum_{\omega\in\Omega} p_\omega = 1,\ P(A) = \sum_{\omega\in A} p_\omega$

Dann ist eine reelle Zufallsvariable eine Funktion $X : \Omega \to \mathbb{R}$,
eine komplexe Zufallsvariable eine Funktion $X : \Omega \to \mathbb{C}$
und ein (reeller) Zufallsvektor eine Funktion $X : \Omega \to \mathbb{R}^q$

Definition 3.1 ***Erwartungswert*** *einer Zufallsvariablen*

$$E(|X|) := \sum_{\omega\in\Omega} |X(\omega)| \cdot p_\omega < \infty \quad \Rightarrow E(X) = \sum_{\omega\in\Omega} X(\omega) \cdot p_\omega$$ [18]

Man beachte: Der Erwartungswert ist damit wohldefiniert, da aus $E(|X|) < \infty$ die absolute Konvergenz von $E(|X|)$ folgt und somit auch $E(X)$ konvergiert.

Satz 3.2 $X, Y : \Omega \to \mathbb{R}\, Zufallsvariablen\, mit\, E(|X|), E(|Y|) < \infty,\ \alpha \in \mathbb{R}$
$\Rightarrow$

i) $E(|X + \alpha Y|) < \infty$

ii) $E(X + \alpha Y) = E(X) + \alpha E(Y)$

Beweis

i)
$$\begin{array}{rcl} E(|X + \alpha Y|) & = & \sum_{\omega\in\Omega} |X(\omega) + \alpha Y(\omega)|\, p_\omega \\ & \leq & \sum_{\omega\in\Omega} (|X(\omega)| + |\alpha||Y(\omega)|)\, p_\omega \\ & = & E(|X|) + |\alpha| E(|Y|) \\ & < & \infty \end{array}$$

ii)
$$\begin{array}{rcl} E(X + \alpha Y) & = & \sum_{\omega\in\Omega} (X(\omega) + \alpha Y(\omega))\, p_\omega \\ & = & \sum_{\omega\in\Omega} X(\omega)\, p_\omega + \alpha \sum_{\omega\in\Omega} Y(\omega)\, p_\omega \\ & = & E(X) + \alpha E(Y) \end{array}$$

Beispiel 3.3 *Bernoulli Experiment*

[18] das liest sich so: wenn $E(|X|) < \infty$, dann ist $E(X)$ ...

$\Omega_0 = \{a,b\}$ (bzw. $\{0,1\}$), $p = P(\{a\}),\ q = P(\{b\}) = 1-p$
n-malige Wiederholung: $\Omega = \Omega_0^n \ni (\omega_1,\ldots,\omega_n)$
$A \subset \Omega,\ A = \{\omega | Anz(\{j|\omega_j = a\}) = k\},\ P(A) = \binom{n}{k} p^k q^{n-k}$

Frage: 1) Wie oft kommen k Erfolge?
2) Wie groß ist die Zahl der Erfolge im Mittel?

$X(\omega) = Anz(\{j|\omega_j = a\})$

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\, P(\{\omega | Anz(\{j|\omega_j = a\}) = k\}) \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k} \qquad \text{Summand für } k=0 \text{ weglassen} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!}\, p\, p^{k-1} q^{(n-1)-(k-1)} \\
&\overset{j=k-1}{=} n \cdot p \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \qquad \text{binomischer Lehrsatz} \\
&= n \cdot p\, (p+q)^{n-1} \\
&= n \cdot p
\end{aligned}$$

Das heißt: Der Erwartungswert für die Zufallsvariable, wie oft Erfolg bei n Wiederholungen Erfolg kommt, innerhalb der Binomialverteilung ist die Anzahl der Wiederholungen multipliziert mit der Erfolgswahrscheinlichkeit.

Oder stark vereinfacht ausgedrückt: Bei n Wiederholungen ist es am wahrscheinlichsten, dass $n \cdot p$-mal Erfolg eintritt.

Natürlich ist das auch die Antwort auf Frage 2): Die Zahl der Erfolge im Mittel ist $n \cdot p$.

Einschub: Erwartungswert bei LaPlace Würfel ($X \equiv$ Augenzahl): $E(X) = \frac{21}{6}$[19] $= 3,5$

Beispiel 3.4 *geometrische Verteilung*

[19] 21 ist die Summe der Augenzahlen, jedes Ereignis ist mit $\frac{1}{6}$ gleich wahrscheinlich

$\Omega_0 = \{a, b\}$
Wahrscheinlichkeit, dass bei der k-ten Wiederholung zum ersten Mal a auftritt $= p \cdot q^{k-1}$
$\Omega_0^n \ni (\omega_j)_{j=1}^{\infty} = \omega,\ \omega_j \in \Omega_0$

Frage: Wie oft muss im Mittel das Experiment wiederholt werden damit zum ersten Mal a auftritt?

$X(\omega) = inf(\{k \in \mathbb{N} | \omega_k = a\})$ [20] [21]

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k\, p\, q^{k-1} = p \sum_{k=0}^{\infty} \frac{d}{dq}(q^k) \qquad \text{siehe Fußnote 21} \\
&= p \cdot \frac{d}{dq}(\sum_{k=0}^{\infty} q^k) \\
&= p \cdot \frac{d}{dq}(\frac{1}{1-q}) = \frac{p}{(1-q)^2} \\
&= \frac{1}{p}
\end{aligned}$$

Beispiel 3.5 *Poisson-Verteilung*

$\Omega_0 = \{a, b\},\ \Omega_0^n \ni \omega$
$P(A) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}$

Frage und Definition von X wie bei Beispiel 3.3 (Bernoulli Experiment)

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} k\, e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{Summand für } k = 0 \text{ weglassen} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \qquad \text{Exponentialreihe} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

Definition 3.6 $X : \Omega \to \mathbb{R}\, oder\, \mathbb{R}^q$

$$\mathbb{R}^q \ni t : \quad P_X(\{t\}) = P(\{\omega | X(\omega) = t\})$$

ist die Verteilung der Zufallsvariablen X
Dabei gilt: $P_X(\{t\}) = 0$ *für alle* $t \in \mathbb{R}^q$ *bis auf Ausnahme der* $t \in X(\Omega) = \{X(\omega) | \omega \in \Omega\}$

[20] Infimum, da $inf(\emptyset) = \infty$; $X \equiv$ die kleinste Stelle in ω, an der a steht
[21] da Summe absolut konvergent, kann Ableitung aus Summe gezogen werden

Sei $A \subset \mathbb{R}^q$

$$\begin{aligned} P_X(A) &:= \sum_{t\in A} P_X(\{t\}) \\ &= \sum_{t\in A} P(\{\omega|X(\omega)=t\}) \qquad \text{disjunkt} \\ &= P(\bigcup_{t\in A}\{\omega|X(\omega)=t\}) \\ &= P(X^{-1}(A)) \qquad \text{Urbildmenge} \end{aligned}$$

Konsequenz: $\sum_{t\in\mathbb{R}^q} P_X(\{t\}) = 1$

Situation $\Omega_0 = \{a,b\} \quad \Omega_0^n$
$X(\omega) = Anz(\{j|\omega_j = a\})$
$P_X(\{t\}) = 0 \quad t \notin \{0,1,\dots,n\}$
$P_X(\{k\}) = b(n,k,p)$[22] $\quad k \in \{0,1,\dots,n\}$

Bemerkung 3.7 *Situation wie eben*

i) $E(|X|) = \sum_{t\in\mathbb{R}^q} |t| P_X(\{t\})$

ii) Falls $E(|X|) < \infty \Rightarrow E(X) = \sum_{t\in\mathbb{R}^q} t\, P_X(\{t\})$

Beweis

i) $E(|X|) \overset{3.1}{=} \sum_{\omega\in\Omega} |X(\omega)| \cdot p_\omega =$ [23] $\sum_{t\in\mathbb{R}^q} \sum_{\omega|X(\omega)=t} |X(\omega)| \cdot p_\omega \overset{3.6}{=} \sum_{t\in\mathbb{R}^q} |t| P_X(\{t\})$

ii) folgt aus i) und 3.1 $\quad\square$

[22]das habe ich gedichtet, p könnte 1/2 sein, ursprünglich stand da $P_X(\{k\}) = k$
[23]man beachte, dass die Summanden für $t \notin X(\Omega) = 0$

28.04.

4 Maße und Wahrscheinlichkeitsmaße

Beispiel 4.1 *Flächenbestimmung eines rechtwinkligen Dreiecks*

Das habe ich verpasst.

Frage: Kann man dieses Verfahren auf $\mathcal{P}(\mathbb{R}^2)$ fortsetzen?

Antwort: Nein! Aber man kann es auf ein großes Teilsystem in $\mathcal{P}(\mathbb{R}^2)$ anwenden.

Es sei stets $\boxed{\Omega \neq \emptyset}$

Definition 4.2 *σ-Algebra und Maß (siehe [**?**] S. 128)*

i) $\mathcal{A} \subset \mathcal{P}(\Omega)$ *ist eine σ-Algebra* $\iff$

(1) $\emptyset, \Omega \in \mathcal{A}$

(2) $A, B \in \mathcal{A} \Rightarrow A \cap B,\ A^c,\ A \cup B,\ A \setminus B \in \mathcal{A}$

(3) $A_n \in \mathcal{A} \quad n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

ii) $\mu : A \to [0, \infty]$ *heißt Maß* $\iff$

$(A_n)_{n=1}^{\infty} \in \mathcal{A}$*($\mathcal{A}$ ist σ-Algebra) paarweise disjunkt* [24]

$\Rightarrow \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$

Gilt zusätzlich $\mu(\Omega) = 1 \Rightarrow$ ist μ ein Wahrscheinlichkeitsmaß

Beispiel $\Omega \ni \omega \mapsto \delta_\omega \geq 0$ mit $\sum_{\omega \in \Omega} \delta_\omega = 1$
$A \in \mathcal{P}(\Omega$[25]$)$: $\quad P(A) = \sum_{\omega \in A} \delta_\omega$ ist ein diskretes Wahrscheinlichkeitsmaß

Beispiel 4.3

$\mathbb{N},\ \mathcal{A} = \mathcal{P}(\mathbb{N}),\ \mu(A) = Anz(A)$[26]$,\ A_n \in \mathcal{P}(\mathbb{N}),\ A_n \cap A_m = \emptyset \quad n \neq m,\ A = \bigcup_{n=1}^{\infty} A_n$

Fall 1: $Anz(A_n) < \infty \quad \forall n \in \mathbb{N}$

[24] $A_n \cup A_m = \emptyset \quad n \neq m$
[25] hier ist $\Omega = \mathcal{A}$
[26] das ist kein Wahrscheinlichkeitsmaß , da $\mu(\Omega) \neq 1$

(a) $Anz(A) < \infty \Rightarrow$ fast alle $A_n = \emptyset$

(b) $Anz(A) = \infty \Rightarrow A_n \neq \emptyset$ für unendlich viele $n \iff Anz(A_n) \geq 1$

Fall 2: $Anz(A_n) = \infty$ für (mind.) ein $n \Rightarrow Anz(A) = \infty$

Definition 4.4

i) $\mathcal{A} \subset \mathcal{P}(\Omega)$ *mit (1), (2) von 4.2 heißt Algebra*

ii) μ *ist additiv* $\iff \mu : \mathcal{A} \to [0, \infty],\ A, B \in \mathcal{A},\ A \cap B = \emptyset \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$

iii) μ *ist* σ*-additiv* $\iff A, A_n \in \mathcal{A},\ A_n \cap A_m = \emptyset \quad n \neq m,\ A = \bigcup_{n=1}^{\infty} A_n \Rightarrow \mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$

Beispiel $\mathbb{R}^q$, $\mathcal{I}^q \equiv$ Intervalle des $\mathbb{R}^q \to$ achsenparallele Rechtecke
$\mathcal{F}^q \equiv$ Vereinigung endlich vieler disjunkter Intervalle

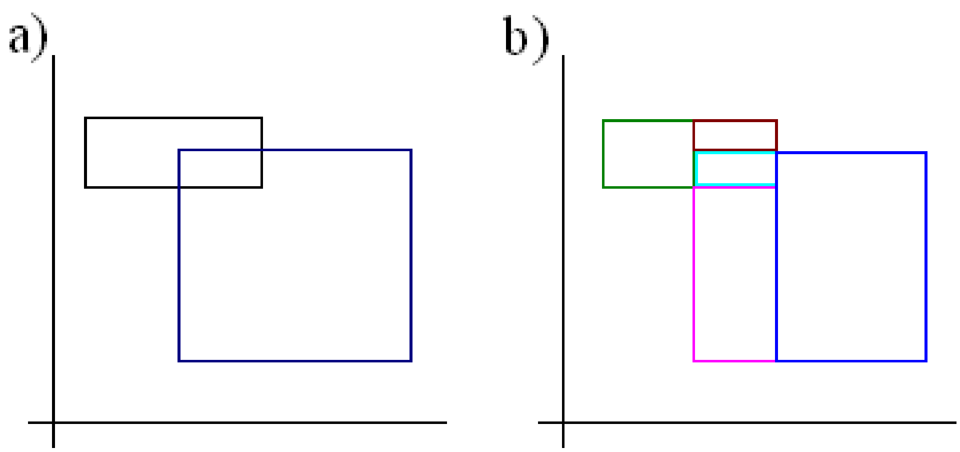

Abbildung 2: Zerlegung der Vereinigung zweier Intervalle in die Vereinigung vieler disjunkter Intervalle

$\Rightarrow$ ist das Flächenmaß λ^q additiv

Bemerkung 4.5
$\mathcal{A} \subset \mathcal{P}(\Omega)$ *sei eine Algebra,* $\mu : \mathcal{A} \to [0, \infty]$ *sei additiv*

i) $A, B \in \mathcal{A},\ A \subset B \Rightarrow \mu(A) \leq \mu(B)$

ii) $A, B \in \mathcal{A} \Rightarrow \mu(A \cup B) \leq \mu(A) + \mu(B)$

Beweis

i) $B = A \cup (B \setminus A) \Rightarrow A \cap (B \setminus A) = \emptyset$
wegen additiv: $\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A) + \underbrace{\mu(B \setminus A)}_{\geq 0} \geq \mu(A)$

$\Rightarrow$ größere Menge $\rightarrow$ größere Fläche

ii) $A \cup B = A \cup (B \setminus A)$
$\Rightarrow \mu(A \cup B) = \mu(A \cup (B \setminus A)) \overset{additiv}{=} \mu(A) + \mu(B \setminus A) \leq \mu(A) + \mu(B)$ $\square$

Satz 4.6
$\mathcal{A} \subset \mathcal{P}(\Omega)$ sei eine Algebra, $\mu : \mathcal{A} \rightarrow [0, \infty]$ sei additiv
$\Rightarrow$ sind äquivalent:

i) μ *ist* σ*-additiv*

ii) $(A_n)_{n=1}^{\infty} \subset \mathcal{A},\ A \in \mathcal{A}$ *mit* $A \subset \bigcup_{n=1}^{\infty} A_n$
$\Rightarrow \mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$

iii) $(A_n)_{n=1}^{\infty} \subset \mathcal{A}$ *mit* $A_n \subset A_{n+1}\ \forall(n \in \mathbb{N}),\ A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$
$\Rightarrow \mu(A_n) \rightarrow \mu(A)$ *für* $(n \rightarrow \infty)$

Zusatz: falls $\mu(\Omega) < \infty \iff$

iv) $(B_n)_{n=1}^{\infty} \subset \mathcal{A},\ B_n \supset B_{n+1}\ \forall(n \in \mathbb{N}),\ \bigcap_{n=1}^{\infty} B_n = \emptyset$
$\Rightarrow \mu(B_n) \rightarrow 0$ *für* $(n \rightarrow \infty)$

Beispiel zu iv) (als Beispiel dafür, dass man die zusätzliche Bedingung benötigt, da in diesem Beispiel zwar $\bigcap_{n=1}^{\infty} B_n = \emptyset$, aber nicht $\mu(B_n) \rightarrow 0$):

$$B_n = [n, \infty[$$

Beweis

i) $\Rightarrow$ ii) $B_1 = A_1 \cap A,\ B_{n+1} = (A \cap A_n) \setminus (B_1 \cup \cdots \cup B_n)$
$\Rightarrow \bigcup_{n=1}^{\infty} B_n = A$ (weil $A \subset \bigcup_{n=1}^{\infty} A_n$)
$\overset{4.4}{\Rightarrow} \mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$

ii) $\Rightarrow$ i) $A = \bigcup_{k=1}^{\infty} A_k \in \mathcal{A},\ A_k \cap A_j = \emptyset\ k \neq j$

$\Rightarrow \sum_{k=1}^{n} \mu(A_k) \overset{4.2}{=} \mu(\bigcup_{k=1}^{n} A_k) \leq \mu(A) \overset{ii)}{\leq} \sum_{k=1}^{\infty} \mu(A_k)$
für $(n \to \infty)$ werden dann die "$\leq$" zu "$=$"

i) $\Rightarrow$ iii) $A_n \subset A_{n+1}$
$B_1 = A_1,\ B_2 = A_2 \setminus B_1,\ B_n = A_n \setminus A_{n-1},\ A_n = B_1 \cup \cdots \cup B_n$
$A = \bigcup_{n=1}^{\infty} B_n \Rightarrow \mu(A) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{k\to\infty} \sum_{n=1}^{\infty} \mu(B_n) = \lim_{k\to\infty} \mu(A_k)$

iii) $\Rightarrow$ i) $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A},\ A_n \cap A_m = \emptyset\ n \neq m,\ B_n = A_1 \cup \cdots \cup A_n$
$\mu(B_n) \overset{4.2}{=} \mu(A_1) + \cdots + \mu(A_n) \leq \mu(A)$
da nach iii) $\mu(B_n) \to \mu(A)$ für $(n \to \infty)$, folgt: $\mu(A) = \sum_{k=1}^{\infty} A_k$

Satz 4.7
$\lambda([a,b]) = b - a \Rightarrow \lambda$ ist auf $\mathcal{F}^1 \sigma$-additiv (ohne Beweis)

Bemerkung 4.8
$\mathcal{E} \subset \mathcal{P}(\Omega),\ \mathcal{A}(\mathcal{E}) = \bigcap\{\mathcal{A} | \mathcal{A} \subset \mathcal{P}(\Omega)$ ist σ-Algebra, $\mathcal{E} \subset \mathcal{A}\}$
$\equiv$ die von $\mathcal{E}$ erzeugte σ-Algebra (die kleinste σ-Algebra, die $\mathcal{E}$ enthält)
Dann gilt: $\mathcal{B}^q := \mathcal{A}(\mathcal{F}^q) = \mathcal{A}(\mathcal{I}^q) = \mathcal{A}(\mathcal{O}^q)$ [27]

Satz 4.9 *Eindeutigkeitssatz*
$\mathcal{A}$ ist σ-Algebra, $\mathcal{F} \subset \mathcal{P}(\Omega)$ mit $\mathcal{A} = \mathcal{A}(\mathcal{F}),\ A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$
Seien μ, ν Maße auf $\mathcal{A}$ mit $\mu(A) = \nu(A) < \infty\ \forall (A \in \mathcal{F}),\ A_n \in \mathcal{F}$ seien geeignet gewählt, sodass $\Omega = \bigcup_{n=1}^{\infty} A_n$

$$\Rightarrow \mu = \nu \qquad \textit{auf ganz } \mathcal{A}$$

Satz 4.10
$\mathcal{A}_0$ ist eine Algebra, $\mu : A_0 \to [0, \infty]$ sei σ-additiv
$A_n \in \mathcal{A}_0,\ \mu(A_n) < \infty,\ \Omega = \bigcup_{n=1}^{\infty} A_n$

$\Rightarrow \mu$ lässt sich eindeutig zu einem Maß auf $\mathcal{A}(\mathcal{A}_0)$ fortsetzen

Auf $\mathcal{B}^q \mapsto \lambda^q$ (auf $\mathcal{B}^q$ führt das zum Flächenmaß λ^q)

05.05.

4.1 Kontinuierliche Theorie

Erwartungswerte = Integrale = Summen (bzw. Reihen) [28]

[27] $\mathcal{O}^q$ ist die offene Menge des $\mathbb{R}^q$
[28] im diskreten Fall

Diskrete Theorie:
$\Omega \ni \omega \mapsto p_\omega \geq 0,\ P(A) = \sum_{\omega \in A} p_\omega \quad A \in \mathcal{P}(\Omega)$

Kontinuierliche Theorie:
$\Omega \neq \emptyset,\ \mathcal{A} \subset \mathcal{P}(\Omega)$ ist eine σ-Algebra (siehe 4.2)
$P : \mathcal{A} \to [0, \infty],\ A_n \cap A_m = \emptyset\ n \neq m \Rightarrow P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (siehe 4.4)

Name: eine *Borel Algebra* $\mathcal{B}^1$ *(*$\mathcal{B}^q$*)* ist die kleinste σ-Algebra, die alle Intervalle (offene Mengen) $\subset \mathbb{R}$ ($\mathbb{R}^q$) enthält
$\lambda^1(I) = sup(I) - inf(I)$ [29]

An dieser Stelle wurde der Eindeutigkeitssatz (4.9) noch einmal ins Gedächtnis zurückgerufen.

Definition 4.11 *Messbarkeit (siehe [?] S. 140)*
Eine Zufallsvariable $f : (\Omega, \mathcal{A}) \to (\Omega_1, \mathcal{A}_1) \quad (\Omega, \Omega_1 \neq \emptyset,\ \mathcal{A}, \mathcal{A}_1 \sigma$*-Algebren) heißt messbar* $\iff$

$$A_1 \in \mathcal{A}_1 \Rightarrow f^{-1}(A_1) \in \mathcal{A}$$

Analoge Eigenschaft: X, Y seien metrische Räume (etwa $\mathbb{R}^q$), dann ist $f : X \to Y$ stetig $\iff$ $f^{-1}(U)$ offen ist für alle offenen $U \subset Y$

$\Rightarrow$ Messbarkeit gilt für fast alle Funktionen (ist in Praxis kein Problem)

Bemerkung 4.12
Wenn $\mathcal{F}_1 \subset \mathcal{A}_1,\ \mathcal{A}_1 = \mathcal{A}(\mathcal{F}_1)$ *(die kleinste* σ*-Algebra, die* $\mathcal{F}_1$ *enthält) und* $f : (\Omega, \mathcal{A}) \to (\Omega_1, \mathcal{A}_1)$*, dann sind äquivalent:*

i) f *ist messbar*

ii) $f^{-1}(A_1) \in \mathcal{A}\ \forall (A_1 \in \mathcal{F}_1)$

Beweis

"$\Rightarrow$" $\checkmark$ [30]

"$\Leftarrow$" Sei $\widetilde{\mathcal{A}} = \{A_1 | f^{-1}(A_1) \in \mathcal{A}\} (\ni \Omega_1, \emptyset$[31]$)$
da f^{-1} verträglich mit $\cap, \cup, ()^c \Rightarrow f^{-1}(A_1)$ ist σ-Algebra $\quad \widetilde{\mathcal{A}} \supset \mathcal{F}_1 \quad \square$

[29] λ^1 ist ein Flächenmaß
[30] da laut Messbarkeit ii) auf ganz $\mathcal{A}_1$ gilt muss es auch auf $\mathcal{F}_1 \subset \mathcal{A}_1$ gelten
[31] das Urbild der leeren Menge ist die leere Menge

Definition 4.13 *Indikatorfunktion (charakteristische Funktion)*

$$A \subset \Omega \qquad 1_A(\omega) = \chi_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & sonst \end{cases}$$

In der Wahrscheinlichkeitstheorie schreibt man eher 1_A (Indikatorfunktion) in der Analysis eher χ_A (charakteristische Funktion)

Im diskreten Fall ergibt sich:
$f(\Omega) = \{\alpha_1, \dots, \alpha_r\}$
$A_j = [f = \alpha_j]$[32]
$f = \alpha_1 1_{A_1} + \cdots + \alpha_r 1_{A_r}$
$E(f) = \alpha_1 P(A_1) + \cdots + \alpha_r P(A_r)$

$A_1, \dots, A_r \in \mathcal{A}$ paarweise disjunkt, $\mu(A_j) < \infty$
$f = \sum_{j=1}^r \alpha_j \chi_{A_j} = \sum_{j=1}^r \alpha_j 1_{A_j}$ ist eine einfache Treppenfunktion $\mathcal{T}_0(\mathcal{A}, \mathbb{R})$ (oder alternativ z.B.: $\mathcal{T}_0(\mathcal{A}, \mathbb{C})_+$)

damit ist der Erwartungswert:
$E(f) = \sum_{j=1}^r \alpha_j \mu(A_j) = \int f \, d\mu$

Satz 4.14
$f, g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})$ $(\in \mathcal{T}_0(\mathcal{A}, \mathbb{C}))$
$\Rightarrow$

$$E(\beta f + g) = \beta E(f) + E(g)$$

$$f \le g \Rightarrow E(f) \le E(g)$$

$$|E(f)| \le E(|f|)$$

Dazu Zerlegungsprinzip (Abb. 3)

Zusammenhang mit messbaren Funktionen:
$f : (\Omega, \mathcal{A}) \to \mathbb{R} = (\mathbb{R}, \mathcal{B})$ ist messbar $\iff$ $f^{-1}(]-\infty, \alpha]) = [f \le \alpha] \in \mathcal{A} \; \forall \alpha \in \mathbb{R}$

$$\begin{aligned} [f < \alpha] &= [f \ge \alpha]^c \\ [f < \infty] &= \bigcup_{n=1}^{\infty} [f \le \alpha - \tfrac{1}{n}] \end{aligned}$$

$f_n : \Omega \to \mathbb{R}$ messbar $\Rightarrow \liminf_{n\to\infty} f_n, \; \limsup_{n\to\infty} f_n, \; \sup_n f_n, \; \inf_n f_n$ ist messbar

[32] $\equiv A_j$ ist die Menge, die alle ω enthält, für die gilt $f(\omega) = \alpha_j$

Abbildung 3: Zerlegungsprinzip

Beweis:

$$[\sup f_n > \alpha] = \underbrace{\bigcup_{n=1}^{\infty} \underbrace{[f > \alpha]}_{\in\mathcal{A}}}_{\in\mathcal{A}} \Rightarrow [\sup f_n > \alpha] \text{ messbar}$$

$$\limsup_{n\to\infty} f_n = \underbrace{\inf_{n\in\mathbb{N}} \underbrace{(\sup_{m\leq n} f_m)}_{messbar}}_{messbar} \Rightarrow \limsup_{n\to\infty} f_n \text{ messbar}$$

$f : \Omega \to \mathbb{R}$ messbar $\iff$ $\exists$ Treppenfunktion f_n mit $f_n \to f\ (n \to \infty)$
Beweis (Abb. 4):

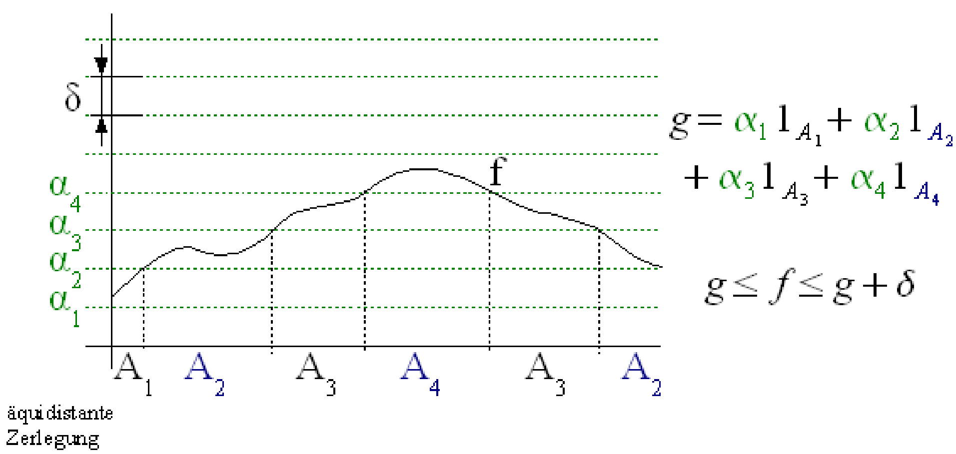

Abbildung 4: Beweis

Tja, was hat das nun damit zu tun, dass f genau dann messbar ist, wenn $f_n \to f\ (n \to \infty)$?

Sei $f(\Omega, \mathcal{A}) \to [0, \infty]$ messbar und f_n Treppenfunktion mit $f_n < f_{n+1}$ und $f_n \uparrow f$, dann gilt

$$\boxed{E(f) = \lim E(f_n) = \int f\, d\mu}$$

Situation:
$f : [a, b] \to \mathbb{R}$ stetig, $t \in [a, b] : f(t) = 0$
$\int_a^b f(t)dt = \int f\, d\lambda = E(f)\checkmark_{(}ganz, ganzbillig)$

Ω diskret: $P(A) = \sum_{\omega \in A} p_\omega \Rightarrow E(f) = \sum_{\omega \in \Omega} f(\omega) p_\omega$

Auf $\mathbb{R}$ sind Maße oft von der Art: $A \mapsto \int_A h(t)dt$
Beispiel: $A = [\alpha, \beta]$
$P(A) = \int_\alpha^\beta h(t)dt$
$h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ (Standardnormalverteilung [33])

08.05.

Beispiel: $\mathbb{R}^q \qquad \mathcal{A}(\mathcal{I}^q) = \mathcal{B}^q$ [34]
dann ist: $\lambda^q(I) = (b_1 - a_1) \dots (b_q - a_q) \quad I = [a_1, b_1] \times \dots \times [a_q, b_q]$

ein einfaches Maß: $\mathbb{N},\ \mathcal{P}(\mathbb{N}),\ \mu$ Zahlenmaß: $\mu(A) = Anz(A)$

Beispiel 4.15

$f : [a, b] \to \mathbb{R}$ stetig
siehe Abb. 5
$E(g) = \int g\, d\lambda = \alpha_1 l(A_1) + \dots + \alpha_r l(A_r) \leq \int f(t)dt$

Kopie dieser konkreten Situation auf abstrakte Situation:
$f : \Omega \to \mathbb{R}_+$ messbar (d.h. $[f \leq \alpha] \in \mathcal{A}\ \forall \alpha \in \mathbb{R}$) $f_n = \sum_{i=1}^r \alpha_i 1_{A_i} \quad A_1, \dots, A_r$ disjunkt $\in \mathcal{A},\ \alpha_i \leq 0$
$f_n \uparrow f$ [35]

$$\boxed{\begin{array}{c} E(f_n) = \sum_{i=1}^r \alpha_i \mu(A_i) = \int f_n d\mu \\ E(f) = \lim_{n\to\infty} E(f_n) \end{array}}$$

[33] für die Standardnormalverteilung kann man keine Stammfunktion finden
[34] $\mathcal{I}^q$ ist eine Familie von Intervallen, $\mathcal{B}^q$ ist die borelsche σ-Algebra
[35] f_n geht monoton von unten gegen f

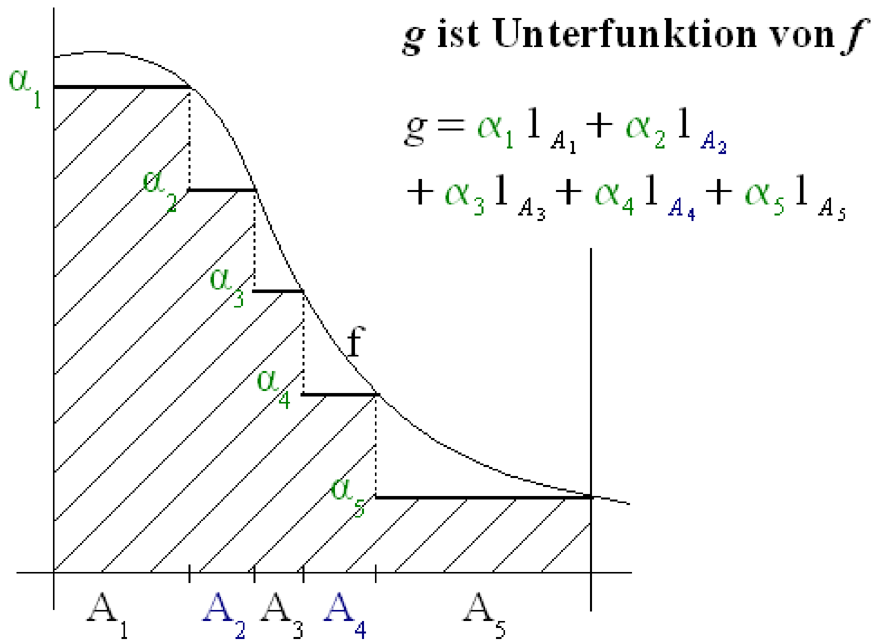

Abbildung 5: Beispiel einer Treppenfunktion

Satz 4.16
$E(f)$ ist unabhängig von der approximierenden Folge (f_n) von Treppenfunktionen

Beweisidee $0 \leq f_n \uparrow f,\ f_n = \sum_{j=1}^r \alpha_j 1_{A_j}$ $\quad A_1, \ldots, A_r$ disjunkt $\in \mathcal{A},\ 0 \leq \beta,\ 1_B \leq f,\ \varepsilon > 0$
zu zeigen: $1_B \cdot f_n = \sum_{j=1}^r \alpha_j \underbrace{1_{A_j} \cdot 1_B}_{1_{A_j \cap B}} \hookrightarrow (\beta - \varepsilon)\mu(B) \leq \lim E(f_n \cdot 1_B)$
$C_n = [f_n \geq \beta - \varepsilon] \cap B \Rightarrow C_n \subset C_{n+1},\ \bigcup_{n=1}^\infty C_n = B$
$\Rightarrow \mu(C_n) \to \mu(B)$

Satz 4.17 [36]
$g, f : \Omega \to [0, \infty[$ messbar, $\alpha \geq 0$

i) $E(f + \alpha g) = E(f) + \alpha E(g)$

ii) $E(f) \leq E(g)$ *falls* $f \leq g$

[36] wer den Unterschied zu Satz 4.14 erklärt, bekommt ein Bonbon

Erwartungswert von $f : \Omega \to \mathbb{C}\,(\mathbb{R})$ existiert, wenn $E(|f|) < \infty$; dann ($E(f)$ existiert):

$$E(f) = E((Re\,f)^+) - E((Re\,f)^-) + iE((Im\,f)^+) - iE((Im\,f)^-)$$

$$\left.\begin{array}{rcll} f & = & u - v & u, v \geq 0 \\ & = & g - h & g, h \geq 0 \end{array}\right\} u + h = g + v$$
$E(u) + E(h) = E(g) + E(v) \Rightarrow E(u) - E(v) = E(g) - E(h)$

$f : [a, b] \to \mathbb{R},\ \int_a^b f(t)dt = \int f\,d\lambda \Rightarrow t \notin [a, b] : f(t) = 0$

Satz 4.18
$\mathbb{N},\ \mathcal{P}(\mathbb{N}),\ \mu$ *(Zählmaß),*
$X : \mathbb{N} \to \mathbb{C}(\mathbb{R})$ *Zufallsvariable,* $X(n) = a_n$

$$(E(|X|) < \infty \iff \sum_{n=1}^{\infty} |a_n| < \infty) \Rightarrow E(X) = \sum_{n=1}^{\infty} a_n$$

Beweis $X_m = \sum_{n=1}^{m} a_n 1_{\{n\}} \quad \uparrow X$
$a_n \geq 0$ [37]
$\Rightarrow E(X_m) = \sum_{n=1}^{m} a_n \cdot \mu(\{n\}) =$ [38] $\sum_{n=1}^{m} a_n \cdot 1$
$(m \to \infty) \Rightarrow E(X) = \sum_{n=1}^{\infty} a_n$

4.2 Unabhängigkeit von Ereignissen und Zufallsvariablen

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum [39]
$J \neq \emptyset$ *endliche* Indexmenge
$j \in J \mapsto G_j \subset \mathcal{A},\ A_{i_1} \in G_{i_1}, \ldots, A_{i_r} \in G_{i_r}$

$$P(A_{i_1} \cap \cdots \cap A_{i_r}) = P(A_{i_1}) \ldots P(A_{i_r}) \Rightarrow G_j | j \in J \text{ unabhängig}$$

für endliches J folgende Situation: $G_j \ni A_j = \Omega \times \cdots \times \Omega \times \overset{j}{A} \times \Omega \ldots$

[37] an dieser Stelle sei darauf hingewiesen, dass der Bequemlichkeit halber der Beweis innerhalb der reellen Zahlen geführt wird anstatt innerhalb der komplexen, dies ist leicht an der Forderung $a_n \geq 0$ zu erkennen, da ja bekanntlich auf $\mathbb{C}$ keine vernünftige Ordnung definiert ist; der Beweis funktioniert auch in $\mathbb{C}$, ist dann aber komplizierter

[38] $\mu(\{n\}) = 1$, da einelementige Menge

[39] d.h. $\Omega \neq \emptyset,\ \mathcal{A} \subset \mathcal{P}(\Omega) \sigma$-Algebra, $P : \mathcal{A} \to [0, 1]$ Maß, $P(\Omega) = 1$

Satz 4.19
$j \in J,\ \mathcal{D}_j \subset \mathcal{A},\ A_1, A_2 \in \mathcal{D}_j \Rightarrow A_1 \cap A_2 \in \mathcal{D}_j$
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$\{\mathcal{D}_j | j \in J\}$ *unabhängig,* $\mathcal{A}_j = \mathcal{A}(\mathcal{D}_j)$

$$\Rightarrow \{\mathcal{A}_j | j \in J\} \text{ unabhängig}$$

Wie erhält man so eine σ-Algebra?:

(1) $\mathcal{D}_j$ wie oben, $A_1 \in \mathcal{D}_1, \dots, A_n \in \mathcal{D}_n$ seien fixiert
$\mu(A) = P(A \cap A_1 \cap \dots \cap A_n)$
$\nu(A) = P(A) \cdot P(A_1) \dots P(A_n)$
$A \mapsto \mu(A)$, $\nu(A)$ Maße auf $\mathcal{A}$
$\nu(A) = \mu(A)\ \forall A \in \mathcal{D}_0$
also folgt aus dem Eindeutigkeitssatz (4.9):

$$\nu(A) = \mu(A)\ \forall A \in \mathcal{A}(\mathcal{D}_0) = \mathcal{A}_0$$

(2) $\mathcal{A}_0,\ \mathcal{D}_1,\ \dots,\ \mathcal{D}_n$
Schritt (1) wird mit dem nächsten Element wiederholt bis alle $\mathcal{D}_j$ mit der entsprechenden $\mathcal{A}_j$ ersetzt wurden

Man kann also bei Unabhängigkeit (und vollständiger Durchschnittsstabilität) davon ausgehen, dass σ-Algebren vorliegen.

Definition 4.20
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$X_j : \Omega \to \mathbb{R}$ *Zufallsvariable* $(j \in J)$ *heißen unabhängig* $\iff$

$$G_j = \{[X_j \leq \alpha_j] | \alpha_j \in \mathbb{R}\} \text{ unabhängig}$$

$\alpha_j \leq \beta_j \Rightarrow [X_j \leq \alpha_j] \cap [X_j \leq \beta_j] = [X_j \leq \alpha_j]$

Bemerkung 4.21
$\mathcal{A}_j = \mathcal{A}(G_j)$ *unabhängig* $\Rightarrow X_j(\Omega, \mathcal{A}_j) \to \mathbb{R}, \mathcal{B}^1$ *messbar*

$X_n^{(j)} \overset{gegen}{\to} X_j \qquad X_n^{(j)} = \sum_{k=1}^{r_n} \alpha_k 1_{A_{k,j}} \quad A_{k,j} \in \mathcal{A}_j$
$Anz(J) = 2 : X_0, X_1 \geq 0$ messbar, unabhängig

$X_t^{(0)} = \sum_{k=1}^r \alpha_k 1_{A_{k,0}},\ X_l^{(1)} = \sum_{m=1}^n \beta_m 1_{A_{m,1}}$ [40]

$$\begin{aligned} E(X_t^{(0)} X_l^{(1)}) &= \sum_{k=1}^r \alpha_k \sum_{m=1}^n \beta_m E(1_{A_{k,0}} \cdot 1_{A_{m,1}}) \\ &= \sum_{k=1}^r \alpha_k \sum_{m=1}^n \beta_m P(A_{k,0}) \cdot P(A_{m,1}) \\ &= E(X_t^{(0)}) \cdot E(X_l^{(1)}) \end{aligned}$$

Daraus ergibt sich folgender Satz:

Satz 4.22
X, Y integrierbare Zufallsvariable, unabhängig

$$\Rightarrow \underbrace{E(X \cdot Y) = E(X) \cdot E(Y)}_{unkorreliert}$$

Unabhängige Zufallsvariable sind unkorreliert.

[40] die beiden Reihen approximieren X_0 und X_1 und konvergieren gegen sie

5 Grenzwertsätze und Produkträume

in der Analysis: Integrale durch gleichmäßige Konvergenz über $[a, b]$
Nachteile:

- beschränkt auf (kompakte) Intervalle
- Gleichmäßigkeit oft verletzt oder nicht nachprüfbar (siehe Abb. 6)

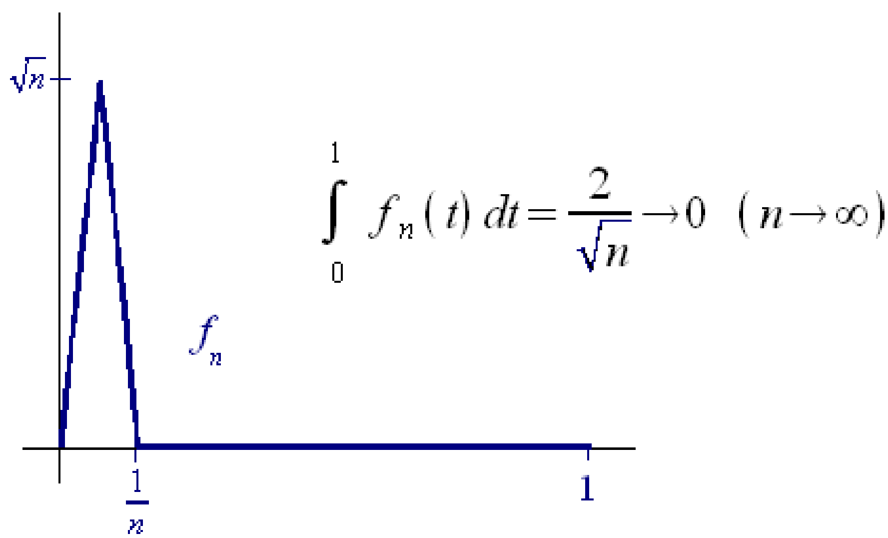

Abbildung 6: Beispiel für Probleme mit Gleichmäßigkeit?

12.05.

$(\Omega, \mathcal{A}, \mu)$ sei irgendein Maßraum (μ ist ein Maß)
Fall: $\mu(\Omega) = 1$, dann handelt es sich um einen Wahrscheinlichkeitsraum

Bezeichnung
$A \in \Omega$ heißt μ-Nullmenge $\iff \exists B \in \mathcal{A}$ mit $A \subset B,\ \mu(B) = 0$

Beispiel: $Q \subset (\mathbb{R}, \mathcal{B}, \lambda)$ ist eine Nullmenge
Warum?: $Q = \{a_n | n \in \mathbb{N}\}$[41], $a \in \mathbb{R} \Rightarrow \lambda(\{a\}) = 0$
$\Rightarrow \lambda(Q) = \sum_{n=1}^{\infty} \lambda(\{a_n\}) = 0$

[41] weil Q abzählbar, kann man es hier mit Hilfe der natürlichen Zahlen schreiben

Bezeichnung
Eigenschaft (*) gilt *μ-fast überall* auf $\Omega \iff \exists A \in \mathcal{A}, \mu(A) = 0$, sodass (*) auf A^c gilt

Beispiel: $f(t) = t^2(t-1)^2 > 0$ λ-fast überall [42]

Bemerkung 5.1
$f : \Omega \to \mathbb{R}\ (\mathbb{C}),\ f \geq 0$ messbar

i) f integrierbar $\Rightarrow f(t) < \infty$ für fast alle $t \in \Omega$

ii) $\int f\, d\mu = 0 \iff f = 0$ ist μ-fast überall

iii) $f = 0$ μ-fast überall $\iff \int_A f\, d\mu = \int 1_A f\, d\mu = 0\ \forall A \in \mathcal{A}$

Beweis

i) $A = [f = \infty]$
$n \cdot 1_A(t) \leq f(t)\ \forall t \in \Omega$
$\boxed{n \cdot \mu(A) = \int n 1_A d\mu} \leq \int f\, d\mu < \infty\ \forall n \in \mathbb{N}$ [43]
$n \cdot \mu(A) < \infty\ \forall n \in \mathbb{N} \Rightarrow \mu(A) = 0$

ii) "$\Leftarrow$" $f = 0$ fast überall $\Rightarrow \mu(\underbrace{[f \neq 0]}_{A}) = 0 \Rightarrow 0 \leq f \leq \infty \cdot 1_A$
$0 \leq \int f\, d\mu \leq \infty \cdot \mu(A) = 0$
"$\Rightarrow$" sei $A_n = [f \geq \frac{1}{n}]$ dann ist $\frac{1}{n} 1_{A_n} \leq f$
$\Rightarrow \frac{1}{n}\mu(A_n) \leq \int f\, d\mu = 0$
$[f \neq 0] = \bigcup_{n=1}^{\infty} A_n \overset{A_n \subset A_{n+1}}{\Rightarrow} \mu([f \neq 0]) = 0$ $\square$

iii) ähnlich ;)

Satz 5.2 *monotone Konvergenz - Beppo Levi*
$f_n \geq 0$ messbar ($[f \leq \alpha] \in \mathcal{A}\ \forall \alpha \in \mathbb{R}$)
$f_n \leq f_{n+1}$ gelte fast überall auf Ω für alle $n \in \mathbb{N}$

$$f = \lim_{n\to\infty} f_n \text{[44]} \Rightarrow \int f_n d\mu \to \int f\, d\mu \in [0, \infty]\ (n \to \infty)$$

$\Rightarrow$ man kann Grenzvorgang gegen Integral austauschen

[42] es gibt zwei Punkte {0,1}, für die $f(t) > 0$ nicht gilt, $\lambda(\{0,1\}) = 0$, für $\{0,1\}^c$ gilt die Eigenschaft
[43] weil das Integral mononton ist
[44] eine monotone Folge konvergiert eigentlich oder uneigentlich

Beweis Wird vielleicht, wohl aber eher unwahrscheinlich, nachgereicht

Satz 5.3 *(majorisierte Konvergenz - Lebesgue)*
$f_n, f : \Omega \to \mathbb{C}$ *messbar*
$f_n(t) \to f(t)$ $(n \to \infty)$ μ*-fast alle* $t \in \Omega$
$h : \Omega \to \mathbb{R}_+$ *integrierbar*
$|f_n(t)| \leq h(t)$ *fast alle* $t \in \Omega$ $n \in \mathbb{N}$

$$\Rightarrow \quad f_n, f \text{ integrierbar und}$$
$$|\int f_n d\mu - \int f\, d\mu| \leq \int |f - f_n| d\mu \ \to 0 \ (n \to \infty)$$

Beweis
oBdA sei $f, f_n : \Omega \to \mathbb{R}$

i) $g_n : \Omega \to [0, \infty]$ messbar $\Rightarrow \int (\underbrace{\liminf_{n\to\infty} g_n}_{=g}) d\mu \leq \liminf_{n\to\infty} \int g_n d\mu$

   Beweis:
   $h_m = \inf\{g_n | n \geq m\} \leq$ [45] $h_m \uparrow g$
   nach Beppo Levi:
   $\Rightarrow \int h_m d\mu \to \int g\, d\mu$
   und aus $\int g_n d\mu \geq \int h_m d\mu$ folgt die Aussage

ii) f_n ist integrierbar, da $\int |f_n| d\mu \leq \int h\, d\mu < \infty$, für f gilt das gleiche
   $g_n = 2h - |f - f_n| \geq 0$ und messbar
   $g_n(t) \to 2h(t)$ $(n \to \infty)$ $\forall t \in \Omega$
   aus i):
   $\int 2h\, d\mu = \int (\liminf_{n\to\infty} g_n) d\mu \leq \liminf_{n\to\infty} (\int g_n d\mu) = \int 2h\, d\mu - \limsup_{n\to\infty} (\int |f - f_n| d\mu)$

Anwendung
$g(t) = \sum_{n=0}^{\infty} a_n t^n$ mit Konvergenzradius $R > 0$
$\Rightarrow$
$g'(t) = \sum_{n=1}^{\infty} n a_n t^{n-1}$ bei gleichem Konvergenzradius R, da $n^{\frac{1}{n}} \to 1$ $(n \to \infty)$
Beweis: $\mathbb{N}_0$, μ Zählmaß, t_0, $|t_0 \pm \delta| \leq r < R$
$\frac{g(t_n) g(t_0)}{t_n - t_0}$ [46] $= \sum_{k=0}^{\infty} \frac{a_k (t_n^k - t_0^k)}{t_n - t_0}$ [47] $\to \sum_{k=1}^{\infty} k a_k t_0^{k-1} = g'(t_0)$

[45] da g der $\liminf_{n\to\infty} g_n$ ist und für eine nach unten beschränkte Folge a_n gilt: $\liminf_{n\to\infty} a_n = \sup_{m\in\mathbb{N}}\{\inf_{n\geq m} a_n\}$, folgt die "$\leq$"-Beziehung
[46] $g(t_n) \to g(t_0)$ $(n \to \infty)$
[47] $\sum < \infty$, da $|| \leq |a_k| n r^{n-1}$

Satz 5.4
$f : I \times \Omega \to \mathbb{R} \quad I \subset \mathbb{R}$ Intervall
D_1[48] $f : I \times \Omega \to \mathbb{R}$ existiere und
$K \subset I$ kompakt
$\Rightarrow \exists h : \Omega \to [0, \infty]$ integrierbar mit $|D_1 f(t,\omega)| \leq h(\omega)\ \forall t \in K,\ \omega \in \Omega$

$$g(t) = \int f(t,\omega)d\mu(\omega) \Rightarrow g'(t) = \int D_1 f(t,\omega)d\mu(\omega)$$

Beispiel 5.5

$f(t) = \int_0^\infty \frac{\sin(s^2 t)}{s} e^{-s^2} ds = Im \int_0^\infty \frac{e^{-s^2(1-it)}}{s} ds$
$\frac{e^{-s^2(1-it)}}{s}(+s^2 i) = ise^{-s^2(1-it)}$

$$\begin{aligned} f'(t) &= Im \frac{-i}{2-2it} \int_0^\infty -2s(1-it)e^{-s^2(1-it)} ds \\ &= Im \frac{-i}{2-2it} \int_0^\infty \left[\frac{d}{ds}(e^{-s^2(1-it)}) \right] ds \\ &= Im \frac{-i}{2-2it} \\ &= \frac{1}{2+2t^2} \end{aligned}$$

$\Rightarrow f(t) = \frac{1}{2} \arctan(t)$

wird noch vervollständigt

15.05.

wird nachgereicht

19.05.

ein paar Zeilen fehlen noch

[48] Ableitung nach der ersten Variablen

6 Erwartungswert, Varianz und Verteilungen

(siehe [?] S. 140)

$(\Omega, \mathcal{A}, P)$ W-Raum
$(\Omega_1, \mathcal{A}_1)$ oft $\mathbb{R}, \mathcal{B}^1$
$X : \Omega \to \mathbb{R}$ Zufallsvariable (also verträglich mit oberen Strukturen $\to$ messbar)

Bezeichnung

$$\begin{array}{lll} E(|X|) = \int |X| dP < \infty & \equiv X \in \mathcal{L}^1(P) & \text{(Schreibweise)} \\ E(|X|^2) < \infty & \equiv X \in \mathcal{L}^2(P) & \\ \text{usw.} & & \end{array}$$

Bemerkung 6.1
$X : \Omega \to \mathbb{R},\ X \in \mathcal{L}^2(P)$

$$\Rightarrow \mathcal{L}^1(P)$$

Beweis
$|X| \leq 1 + |X|^2$
$E(|X|) \leq E(1 + |X|^2) = 1 + E(|X|^2) < \infty \quad \square$

Bemerkung 6.2 *(Verteilung)*
$\phi : (\Omega, \mathcal{A}) \to (\Omega_1, \mathcal{A}_1)$ *messbar (d.h. Zufallsvariable) (* $\iff \phi^{-1}(A_1) \in \mathcal{A}\ \forall A_1 \in \mathcal{A}_1$*)*
$\mathcal{A}_1 \ni A_1 \mapsto \phi^{-1}(A_1) \in \mathcal{A}$

$$\Rightarrow P(\phi^{-1}(A_1)) = P_\phi(A_1) \textit{ ist ein Maß auf } \mathcal{A}_1 \text{[49]}$$
$$\equiv \textit{Verteilung (oder Bildmaß)}$$

Satz 6.3 *(Integration bezüglich Verteilung)*
$\phi : (\Omega, \mathcal{A}, P) \to (\Omega_1, \mathcal{A}_1)$ *messbar*
$\nu(A_1) = P_\phi(A_1) = P(\phi^{-1}(A_1))$
$f : \Omega_1 \to \mathbb{R}\ (\mathbb{C})$ *messbar* $f \in \mathcal{L}^1(\nu) \iff f \circ \phi \in \mathcal{L}^1(P)$

$$\Rightarrow \int f\, d\nu = \int f \circ \phi\, dP$$

[49] ist leicht nachzuprüfen, aber viel Schreibarbeit, weswegen der Beweis weggelassen wurde

Beweis Maßtheoretische Induktion
...

Bemerkung 6.4 *(Verteilungsfunktion)*
$X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ Zufallsvariable (also messbar)
$F_X(t) := P(X \le t) \quad (= P(\{\omega | X(\omega) \le t\}))$

$$F_X(t) \to \begin{cases} 0 & bei\, t \to -\infty \\ 1 & bei\, t \to \infty \end{cases} \qquad \textit{monoton wachsend und rechtsseitig stetig}$$

Beweis rechtsseitige Stetigkeit:
$t \downarrow \tau \Rightarrow A_t = [X \le t] \downarrow A_\tau \Rightarrow P(A_t) \to P(A_\tau)$

Diskrete Situation
$A_m(\Omega) = m < \infty$
$X : \Omega \to \mathbb{R}$ Zufallsvariable mit Werten $x_1 < \cdots < x_r \in \mathbb{R}$
$v_i = P(X = x_i), \quad w_i = v_1 + \cdots + v_i = P(X \le x_i)$
BILD

F_X ist Verteilungsfunktion und sei differenzierbar, dann

$$0 \le (F_X)' = g \qquad \equiv \textit{Dichte} \text{ (siehe [?] S. 129)}$$

$\alpha < \beta$
$F_X(\beta) - F_X(\alpha) = \int_\alpha^\beta g(t)dt$

$$\Rightarrow F_X(\beta) = \int_{-\infty}^{\beta} g(t)dt$$

also gilt für die Verteilung ($A_1 \in \mathcal{B}_1,\ A_1 =]\alpha, \beta]$):

$$\boxed{P_X(A_1)} = F_X(\beta) - F_X(\alpha) = \int_{]\alpha,\beta]} g(t)dt = \boxed{\int_{A_1} g(t)dt}$$

$P_X : \mathcal{B}_1 \in A_1 \to \int_{A_1} g(t)dt$ ist Maß auf $\mathcal{B}_1$[50] $(= \nu?)$
f irgendeine (messbare) Funktion (?)

$$\int f\, d\nu = \int_{-\infty}^{\infty} f(t)g(t)dt = \int f(X)dP$$

[50]die σ-Additivität kann man mit Hilfe des Satzes von Beppo Levi beweisen

Folgerung 6.5
$X : (\Omega, \mathcal{A}, P) \to \mathbb{R}$ besitze eine differenzierbare Verteilungsfunktion F_X mit Ableitung $= g$
$f : \mathbb{R} \to \mathbb{C}$ messbar
$f \cdot g \in \mathcal{L}^{1\,(2)}(\lambda) \iff f \circ X \in \mathcal{L}^{1\,(2)}(P)$

$$\boxed{\Rightarrow \int_{-\infty}^{\infty} f(t)g(t)dt = E(f \circ X)\ (= \int f(X)dP)}$$

Speziell:

$$\int_{-\infty}^{\infty} t\, g(t)dt = E(X) \qquad \int_{-\infty}^{\infty} t^2 g(t)dt = E(X^2)$$

Beispiel 6.6 *Normalverteilung (siehe [**?**] S. 133)*

$\mu,\ \sigma > 0$

$$g(t) = \frac{1}{\sqrt{2\pi} \cdot \sigma}\, e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$ [51]

Es folgt:

i) $\displaystyle\int_{-\infty}^{\infty} g(t)dt = 1$

ii) $\displaystyle\int_{-\infty}^{\infty} t\, g(t)dt = \mu, \quad \int_{-\infty}^{\infty} (t-\mu)^2 g(t)dt = \sigma^2$

wörtlich: μ ist der Erwartungswert und σ^2 die Varianz der Normalverteilung

Beweis zu i)
...

22.05.

Wiederholung / Zusammenfassung
$(\Omega, \mathcal{A}, P)$ ist Wahrscheinlichkeitsraum
$X : (\Omega, \mathcal{A}) \to (\Omega_0, \mathcal{A}_0)$ ist Zufallsvariable (wenn messbar)
$P(X^{-1}(A_0))$ ist ein Maß auf $\mathcal{A}_0 \quad = P_X(A_0)$ und heißt Verteilung von X
($\mathcal{B}$ erzeugt $]-\infty, t]$)(?)
für $X : \Omega \to \mathbb{R}$ gilt des weiteren:

[51] ist die Dichte der Normalverteilung

$F_X(t) = P(X \leq t)$ ist die Verteilungsfunktion von X und ist monoton wachsend, rechtsseitig stetig und $F_X(t) \Rightarrow 1$ $(t \to \infty)$, $F_X(t) \Rightarrow 0$ $(t \to \infty)$
$F'_X = g \Rightarrow g \geq 0 \Rightarrow \int_{-\infty}^{t} g(s)ds = F_X(t)$
$E(f(X)) = \int_{-\infty}^{\infty} f(t)g(t)dt$, wenn eine der Seiten existiert
$\Rightarrow$ unabhängig von der speziellen Struktur des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{A}, P)$ können Erwartungswert und höhere Momente ausgerechnet werden

Aber was ist, wenn F_X nicht differenzierbar ist? Dann kann man nur noch auf den *diskreten Fall* ausweichen:
$Anz(\Omega) < \infty$
$X \rightsquigarrow F_X$[52] ist stückweise konstant
$\alpha_1 < \cdots < \alpha_n$ Sprungstellen
$\delta_1, \ldots, \delta_n > 0$ Sprunghöhen $\sum \delta_i = 1$
BILD
$X = \sum_{i=1}^{n} \alpha_i 1_{A_i},\ A_i = [X = \alpha_i]$
$E(X) = \sum_{i=1}^{n} \alpha_i P(A_i) = \sum_{i=1}^{n} \alpha_i \delta_i$
$P(X \leq \alpha_i) = P(A_1 \cup \cdots \cup A_i)$
$E(f(X)) = \sum_{i=1}^{n} f(\alpha_i)\delta_i$

Beispiel 6.7 *Binomialverteilung*

$b(n,k,p) = \binom{n}{k} p^k q^{n-k} \quad q = 1 - p$
$\alpha_0 = 0 < \alpha_1 = 1 < \cdots < \alpha_n = n$
$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$

$$E(f(X)) = \sum_{k=0}^{n} f(k) \binom{n}{k} p^k q^{n-k}$$

Situation
$X_1, \ldots, X_q : (\Omega, \mathcal{A}, P) \to \mathbb{R} \qquad X = (X_1, \ldots, X_q) : \Omega \to \mathbb{R}^q$
$t = (t_1, \ldots, t_q) \in \mathbb{R}^q$
$F_X(t) = (F_{X_1}(t_1), \ldots, F_{X_q}(t_q))$ [53]
falls differenzierbar:
$F'_X(t) = g(t) = (g_1(t_1), \ldots, g_q(t_q))$

Definition 6.8 *gemeinsame Verteilung*
wenn $X_1, \ldots, X_q$ *unabhängig*

$$P([X_1 \leq t_1] \cap \cdots \cap [X_q \leq t_q]) := F_{X_1}(t_1) \cdot \ldots \cdot F_{X_q}(t_q)$$

[52] $X \rightsquigarrow Y$ soll so etwas heißen wie: Y wird von X gebildet
[53] jedes Element ist nur von einer Variable (t_i) abhängig

Zufallsvariable $X \rightsquigarrow E(X) \sim$ Mittelwert

Beispiel 6.9 *verschiedene Abweichungen von $E(X)$*

...

Definition 6.10 ***(Varianz)***
zur Messung der Güte des Erwartungswertes $E(X)$:
$X \in \mathcal{L}^2(P)$

$$V(X) := E((X - E(X))^2) \qquad \equiv Varianz$$

Folgerung 6.11 $V(X) = 0 \iff X = E(X)$ ✓

Bemerkung 6.12 *(Rechenregeln für Varianz)*
$X_1, \dots, X_n \in \mathcal{L}^2(P)$

i) $a \in \mathbb{R} \Rightarrow V(aX_1) = a^2 V(X_1)$

ii) $X_1, \dots, X_n$ *unkorreliert* [54]

$$\Rightarrow V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n)$$

iii) $V(X) = E(X^2) - E(X)^2$

Beweis
...

Beispiel 6.13 *Binomialverteilung*

$$V(X) = V(X_1 + \cdots + X_n) = n \cdot p \cdot q$$

$\Rightarrow$

$$E(X^2) = V(X) - E(X)^2 = npq - (np)^2$$

Beispiel 6.14 *Normalverteilung*

...

Definition 6.15 ***(Median)***
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$X : \Omega \to \mathbb{R}$ *Zufallsvariable*
F_X *Verteilungsfunktion*

$m \in \mathbb{R}$ *heißt Median (Zentralwert) von X* $\iff P(X \leq m) \wedge P(X \geq m)$

[54] d.h. $E(X_i X_j) = E(X_i) \cdot E(X_j)$, z.B. wenn unabhängig

Der Median ist nicht eindeutig
siehe BILD

Beispiel 6.16 *Exponentialverteilung*

Dichte:

$$g(t) = \begin{cases} 0 & t \leq 0 \\ \lambda e^{-\lambda t} & t > 0 \end{cases}$$

$$\int_{-\infty}^{\infty} g(t)\,dt = \int_0^{\infty} g(t)\,dt = \int_0^{\infty} \lambda e^{-\lambda t}\,dt = 1$$

$\Rightarrow$

$$\boxed{F_X(x)} = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda x} + e^0 = \boxed{1 - e^{-\lambda x}}$$

Erwartungswert:

$$\begin{aligned} E(X) &= \int_0^{\infty} t \cdot g(t)\,dt \\ &= t \cdot (-e^{-\lambda t}) \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = 0 + \left(-\frac{1}{\lambda} e^{-\lambda t} \Big|_0^{\infty} \right) \\ &= \frac{1}{\lambda} \end{aligned}$$

Varianz:

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\begin{aligned} E(X^2) &= \int_0^{\infty} t^2 \cdot \lambda e^{-\lambda t} dt = t^2(-e^{-\lambda t}) \Big|_0^{\infty} + \int_0^{\infty} 2t \cdot e^{-\lambda t} dt \\ &= 0 + 2t \cdot \left(-\frac{1}{\lambda} \right) e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} 2 \cdot \frac{1}{\lambda} e^{-\lambda t} dt = 0 + 0 + \left(-\frac{2}{\lambda^2} e^{-\lambda t} \Big|_0^{\infty} \right) \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Median:

$$F_X(m) = \frac{1}{2} = \int_0^m g(t) dt = 1 - e^{-\lambda m} = e^{-\lambda m}$$

$\Rightarrow$

$$m = \frac{\ln(2)}{\lambda}$$

Definition 6.17 *(Kovarianz)*
$X, Y \in \mathcal{L}^2(P)$

$$Cov(X,Y) := E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

$\Rightarrow$ ist 0, wenn X, Y unkorreliert

26.05.

7 Grenzwertaussagen

(siehe [?] S. 153)

$\boxed{\text{Folge von Zufallsvariablen} \overset{unendliches Produkt}{\rightsquigarrow} \text{Wahrscheinlichkeitsraum } (\Omega, \mathcal{A}, P)}$ [55]

Satz 7.1 *(Lemma von Borel-Cantelli)*
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$(A_n)_{n=1}^{\infty} \subset \mathcal{A}$
$A = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}(\bigcup_{k=n}^{\infty} A_k) = \{\omega \in \Omega | \omega\, ist\, in\, \infty\, vielen\, A_n\, enthalten\}$

i) $\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A) = 0$

ii) $\{A_k | k \in \mathbb{N}\}$ *unabhängig*, $\sum_{k=0}^{\infty} P(A_k) = \infty \Rightarrow P(A) = 1$

Beweis
...

Satz 7.2 *(Tschebyschew[56] Ungleichung)*
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$X \in \mathcal{L}^2(P) (\iff E(X^2) < \infty)$
$\varepsilon > 0$

$$\Rightarrow P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} V(X)$$

Beweis
$A = [|X - E(X)| \geq \varepsilon] \in \mathcal{A} \quad \checkmark$
$0 \leq 1_A \cdot \varepsilon \leq |X - E(X)|$, also auch: $0 \leq 1_A \cdot \varepsilon^2 \leq |X - E(X)|^2$
$\varepsilon^2 P(A) = E(1_A \cdot \varepsilon^2) \leq E(|X - E(X)|^2) = V(X) \qquad \square$

Lemma 7.3 *(Schwaches Gesetz der großen Zahlen)* [57]
$(X_n)_1^{\infty} \subset \mathcal{L}^2(P)$ *unkorreliert (z.B. unabhängig)*

---

[55] oft wird der resultierende Wahrscheinlichketisraum durch eine Folge von unabhängigen zufälligen Experimenten gebildet

[56] weitere mögliche Schreibweisen: Tschebeyscheff, Tchebychev, Tchebykhev, Tchebykhov

[57] die hier vorliegende ist nicht die Originalformulierung

$M := \sup_{n\in\mathbb{N}} V(X_n) < \infty$ *(Varianzen sollen beschränkt sein)*
$\varepsilon > 0, n \in \mathbb{N}$

$$\Rightarrow P(|\frac{1}{n}\sum_{k=1}^{n}(X_k - E(X_k))| \geq \varepsilon) \leq \frac{M}{n\varepsilon^2}$$

Beweis
...

Satz 7.4 *(Starkes Gesetz der großen Zahlen)*
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$(X_n)_1^\infty \subset \mathcal{L}^2(P)$ *unkorreliert*
$M := \sup_{n\in\mathbb{N}} V(X_n) < \infty$

$$\Rightarrow \frac{1}{n}\sum_{k=1}^{n}(X_k - E(X_k)) \to 0 \quad (n \to \infty) \quad \textit{fast sicher (P-fast-überall)}$$

Beweis
...

Unterschied starkes und schwaches Gesetz der großen Zahlen:

- im schwachen haben wir stochastische Konvergenz:

$$Z_n \to 0 \text{ stochastisch} \iff \forall \varepsilon > 0 : P(|Z_n| \geq \varepsilon) \to 0 \quad (n \to \infty)$$

- im starken fast-sichere Konvergenz:

$$Z_n \to 0 \text{ fast sicher} \iff \exists B \in \mathcal{A} \text{ mit } P(B) = 1 \text{ sodass}$$
$$\frac{1}{n}\sum_{k=1}^{n}(X_k(\omega) - E(X_k)) \to 0 \quad (n \to \infty)\, \forall \omega \in B$$

Bemerkung 7.5
Folge Z_n von Zufallsvariablen
Z_n konvergiert fast sicher gegen 0

$$\Rightarrow Z_n \textit{ konvergiert stochastisch gegen } 0$$

Beispiel 7.6 *fairer Würfel*

Man will herausbekommen, ob ein Würfel fair ist.
Dazu würfelt man n-mal (n groß) und zählt die gewurfenen Sechsen:
$Z_n = \frac{X_1+\cdots+X_n}{n} \quad (X_k = 1$ falls 6 gewurfen, 0 sonst)
Aus dem Gesetz der großen Zahlen folgt dann:
$Z_n \to \frac{1}{6} \quad (n \to \infty)$ *fast sicher*
D.h. sollte der tatsächliche Wert von $\frac{1}{6}$ abweichen, dann ist der Würfel nicht fair.
Das ist allerdings nur eine qualitative Aussage (quantitativ wäre zu wissen, wie oft man würfeln muss um entscheiden zu können, ob der Würfel fair ist)

02.06.

7.1 Zentraler Grenzwertsatz und Normalverteilung

Normalverteilung
Standardnormalverteilung (siehe Abb. 7):

$$N_{0,1}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \varphi_{0,1}(s)ds = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{s^2}{2}} ds = \Phi(t)$$

allgemein:

$$N_{\mu,\sigma^2}(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{t} \varphi_{\mu,\sigma^2}(s)ds, \quad \varphi_{\mu,\sigma^2}(s) = e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

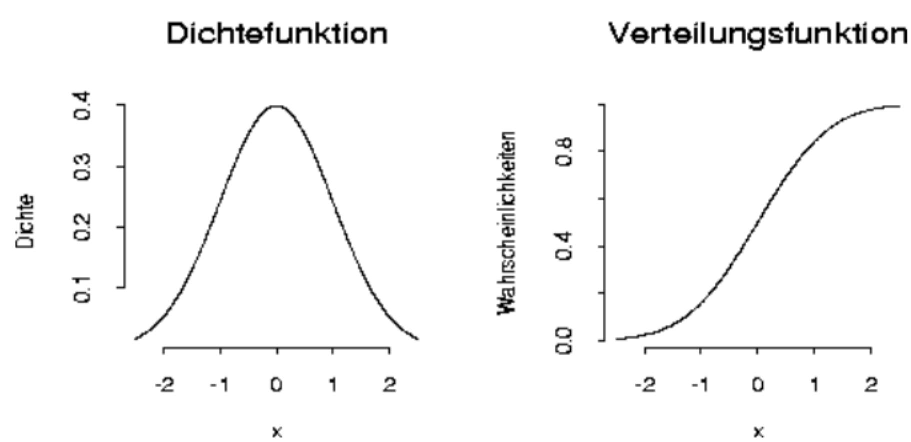

Abbildung 7: Standardnormalverteilung

Bemerkung 7.7
$(\Omega, \mathcal{A}, P)$ Wahrscheinlichkeitsraum
$X_1, \ldots, X_n : \Omega \to \mathbb{R}$ unabhängige Zufallsvariable
$F_i = F_{X_i}$ Verteilungsfunktion mit Dichte f_i: $F_i(t) = \int_{-\infty}^t f_i(s)ds$
$X := (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ mit Dichte $f(X_1, \ldots, X_n) = f_1(X_1) \cdot \ldots \cdot f_n(X_n)$

$$\Rightarrow \quad \begin{array}{l} P_X = P(X^{-1}(A)),\ A \subset \mathbb{R}^n,\ A \in \mathcal{B}^n \\ d.h.\ P_X(A) = \int \ldots \int_A f_1(X_1) \ldots f_n(X_n) dx_1 \ldots dx_n \end{array}$$

Beweis
...

Beispiel 7.8

$X_1, X_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariable
Verteilungen mit Dichten f_1, f_2 (d.h. $P(X_i \leq t) = \int_{-\infty}^t f_i(s)ds$)
$Y = X_1 + X_2$ setze $B = \{(X_1, X_2) \in \mathbb{R}^2 | X_1 + X_2 \leq t\}$

$$\begin{aligned} \Rightarrow \quad P(Y \leq t) &= P_{(X_1,X_2)}(B) \\ &= \iint_B f_1(x_1) f_2(x_2) dx_1 dx_2 \qquad | x_1 + x_2 = s,\ x_2 = u \\ &= \int_{-\infty}^t (\underbrace{\int_{-\infty}^{\infty} f_1(s-u) f_2(u) du}_{f_1 * f_2(s) = f_2 * f_1(s) \equiv \textbf{\textit{Faltung}}}) ds \\ &= \int_{-\infty}^t f_1 * f_2(s) ds \end{aligned}$$

Satz 7.9
X_1, X_2 unabhängige Zufallsvariable mit Dichten der Verteilungsfunktionen f_1, f_2
$\Rightarrow$

i) $X_1 + X_2$ besitzt Dichte $f_1 * f_2$

ii) $f_i = \varphi_{\mu_i, \sigma_i^2} \Rightarrow f = \varphi_{\mu, \sigma^2}$ Dichte von $X_1 + X_2$ mit $\mu = \mu_1 + \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$

Beweis

i) siehe 7.8

ii) ...

Satz 7.10 ***(Zentraler Grenzwertsatz)***
$(\Omega, \mathcal{A}, P)$ *Wahrscheinlichkeitsraum*
$X_n : \Omega \to \mathbb{R}$ *Zufallsvariable, identisch verteilt* [58], *unabhängig,* $X_n \in \mathcal{L}^2(P)$
$\mu = E(X_n),\ \sigma^2 = V(X_n)$
$S_n = (X_1 + \cdots + X_n)$

$$S_n^* = \frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu)$$
$$\Rightarrow d(S_n^*, \Phi) \to 0 \quad (n \to \infty)$$

$d(S_n^*, \Phi) = ||F_n^* - \Phi||_\infty = \sup\{|F_n^*(t) - \Phi(t)| \; |t \in \mathbb{R}\}$
wobei F_n^* *Verteilungsfunktion von* S_n^* *und* Φ *Verteilungsfunktion der Standardnormalverteilung*

Beweis
Es wird der elementare Beweis vorgeführt, der auf zwei zusätzlichen Lemmata basiert.

Lemma 7.11
$X, Y : \Omega \to \mathbb{R}$ *Zufallsvariablen*
$E(Y^2) \leq v$
$$\Rightarrow d(X + Y, \Phi) \leq d(X, \Phi) + 2v^{\frac{1}{3}}$$

Beweis
...

Lemma 7.12
$X_n : \Omega \to \mathbb{R}$ *Zufallsvariable mit endlichem Wertebereich*
$E(X_n) = 0,\ V(X_n) = 1,\ |X_n| \leq M$ *für ein* $M > 0$
$S_n^* = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$

$$\Rightarrow d(S_n^*, \Phi) \to 0 (n \to \infty)$$

[58] d.h. alle X_n haben die gleiche Verteilungsfunktion

Beweis
...

Somit kann man den allgemeinen Fall (für den zentralen Grenzwertsatz) durch Abschneiden und Stückeln auf den Fall in Lemma 7.12 zurückführen.

Als direkte Konsequenz aus dem zentralen Grenzwertsatz folgt:

Satz 7.13 *(Satz von de Moivre-Laplace)*

Approximation der Binomialverteilung durch (Standard-)Normalverteilung

Bernoulli-Experiment:
$\Omega_0 = \{a, b\},\ X_n(a) = 1,\ X_n(b) = 0,\ P(\{a\}) = p \in]0,1[$
$S_n = X_1 + \cdots + X_n$
$P(S_n = k) = \binom{n}{k} p^k q^{n-k} \qquad (0 \leq k \leq n)$
$F_n(t) = \sum_{k \leq t} \binom{n}{k} p^k q^{n-k}$
$S_n^* = \frac{1}{\sqrt{n \cdot p \cdot q}}(X_1 + \cdots + X_n - n \cdot p) = \frac{1}{\sqrt{V(S_n)}}(S_n - E(S_n))$
$d(S_n^*, \Phi) \to 0 \quad (n \to \infty)$

$$\Rightarrow P(k_a \leq X_1 + \cdots + X_n \leq k_b) = \sum_{k=k_a}^{k_b} \binom{n}{k} p^k q^{n-k} \approx \Phi(b) - \Phi(a)$$

$$a = \frac{1}{\sqrt{npq}}(k_a - np) \quad b = \frac{1}{\sqrt{npq}}(k_b - np)$$

Konvergenzverbesserung
Problem: $P(X_1 + \cdots + X_n \geq k)$ für ein $k \in \{1 \ldots n-1\}$
Komplementärereignis: $P(X_1 + \cdots + X_n \leq k-1)$
siehe Abb. 8
Lösung: berechne $P(X_1 + \cdots + X_n \geq k - \frac{1}{2})$

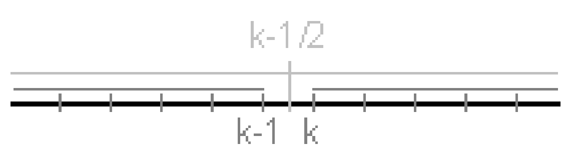

Abbildung 8: Veranschaulichung zu Konvergenzverbesserung

Beispiel

$p = q = \frac{1}{2},\ n = 20,\ k_a = 7,\ k_b = 11$
$P(k_a \leq S_n \leq k_b) = 0,69062$ (exakt)

ohne Korrektur:	$a = -1,3416,\ b = 0,4472$
	$\Rightarrow \Phi(b) - \Phi(a) = 0,5828$
mit Korrektur:	$P(k_a - \frac{1}{2} \leq S_n \leq k_b + \frac{1}{2})$
	$\tilde{a} = -1,5652,\ \tilde{b} = 0,6708$
	$\Rightarrow \Phi(b) - \Phi(a) = 0,6901$

Andere Approximation der Binomialverteilung: Poisson-Verteilung

$$e^{-\alpha}\frac{\alpha^k}{k!} \quad \alpha = n \cdot p \qquad \text{(siehe 1.18)}$$

Erwartungswert und Varianz der Poisson-Verteilung:

$$E(X) = \alpha \qquad V(X) = \alpha$$

16.06.

8 Parameterschätzungen und grundlegende Fragestellungen der Statistik

Beispiel 8.1 *Kontrolle der Produktqualität*

$N = 100000$ Schrauben
Stichprobe: $n = 100$ Stück werden gezogen, davon sind x (zufällige Zahl) defekt
Frage: Welche Rückschlüsse auf die wahre Zahl defekter Schrauben (W) sind möglich?

1. Ansatz: $W(x) = x \cdot \frac{N}{n}$ als *Schätzung* $\to$ zufällige Größe

2. Ansatz: Angabe eines *Konfidenzintervalls*
   d.h. Angabe $C(x) \subset \mathbb{R}$, in dem der Wert mit großer Sicherheit liegt
   x zufällig $\Rightarrow C(x)$ zufällig
   $P_w(\{x | W \in C(x)\}) \geq 1 - \alpha \qquad \alpha$: Irrtumsniveau
   d.h. im Beispiel:

$$\sum_{x | W \in C(x)} \frac{\binom{W}{x}\binom{N-W}{n-x}}{\binom{N}{n}} \geq 1 - \alpha \quad \forall W \in \{0, \ldots, N\}$$

3. Ansatz: Zusatzvoraussetzung
   der vereinbarte Preis wird nur bezahlt im Fall $W \leq 5000$ (nicht mehr als 5% defekt)
   Hypothese H_0: $W \in \{0, \ldots, 5000\}$
   Alternative H_1: $W \notin \{0, \ldots, 5000\}$
   Entscheidungsverfahren: $c \in \mathbb{R}$ so zu bestimmen, dass im Fall

$$x \leq c \Rightarrow \text{Entscheidung für } H_0$$
$$x > c \Rightarrow \text{Entscheidung für } H_1$$

   $\equiv$ *Test*

Formalisierung der Situation
$\mathcal{X}$ Stichprobenraum, d.h. die Gesamtheit aller möglichen Beobachtungen

Anmerkung:

Ausgangspunkt ist ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ mit einer Zufallsvariable $X : \Omega \to \mathcal{X}$, wobei $\mathcal{X}$ die real beobachtbaren Ereignisse enthält. Ω und X spielen dann keine weitere Rolle mehr, nur noch die Verteilung F_X.

Gesucht ist: Das zugeordnete Wahrscheinlichkeitsmaß auf $\mathcal{X}$ (bzw. die zugehörige Verteilung)

Definition 8.2 *Statistisches Modell*

$$(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$$

$\mathcal{X}$ *Stichprobenraum*
$\mathcal{F}$ *σ-Algebra* $\subset \mathcal{P}(\mathcal{X})$
P_ϑ *Wahrscheinlichkeitsmaß auf* $\mathcal{F}$ $\quad \forall \vartheta \in \Theta$[59], $Anz(\Theta) \geq 2$
E_ϑ *Erwartungsert bezüglich* P_ϑ
V_ϑ *Varianz bezüglich* P_ϑ

ein statistisches Modell heißt parametrisch $\iff$ $\Theta \subset \mathbb{R}^q$ *(ist $q = 1$, dann nennt man es einparametrisch)*

Statistische Standardmodelle:

i) $\mathcal{X}$ *diskret,* $\mathcal{F} = \mathcal{P}(\mathcal{X})$, P_ϑ *Zähldichte* $(\varrho_\vartheta) \Rightarrow$ *diskretes Modell*

ii) $\mathcal{X} \in \mathcal{B}^q$, $\mathcal{F} = \mathcal{B}^q_{\mathcal{X}}$, *jedes* P_ϑ *hat Dichtefunktion* $\varrho_\vartheta : P_\vartheta(A) = \int_A \varrho_\vartheta(\omega) d\omega \Rightarrow$ *stetiges Modell*

Definition 8.3 *Produktmodell*
$(E, G, Q_\vartheta | \vartheta \in \Theta)$ *statistisches Modell,* $n \in \mathbb{N}$, $n \geq 2$
Setze

$$(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta) = (E^n, G^{\otimes n}, Q_\vartheta^{\otimes n} | \vartheta \in \Theta)$$

ist das n-fache Produktmodell

die oben für stastische Modelle angegebenen Eigenschaften übertragen sich sinngemäß (parametrisch, diskret, stetig)

[59] Θ kann eine beliebige Menge sein, mit deren Elementen die für möglich gehaltenen Verteilungen parametrisiert sind; meist Intervall in $\mathbb{R}, \mathbb{R}^d, \mathbb{N}$

8.1 Parameterschätzungen

(siehe [?] S. 60 und S. 164)

Definition 8.4 *Statistik, Schätzer*
$(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$ *statistisches Modell,* $\quad (\mathcal{E}, \mathcal{S})$ [60]

i) $S : (\mathcal{X}, \mathcal{F}) \to (\mathcal{E}, \mathcal{S})$ *messbar* $\Rightarrow$ S *heißt Statistik*

ii) $\tau : \Theta \to \mathcal{E}$ *und* $T : (\mathcal{X}, \mathcal{F}) \to (\mathcal{E}, \mathcal{S})$ *ist Statistik* $\Rightarrow$ T *heißt Schätzer für* τ

Einschub: Gleichverteilung in einem Intervall $[a, b] \subset \mathbb{R}$
(siehe [?] S. 132)
Die Gleichverteilung in $I = [a, b]$ hat die Dichte, die innerhalb von I den Wert $\frac{1}{b-a}$ annimmt und außerhalb 0:

$$g(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & sonst \end{cases}$$

$\Rightarrow$ Die Wahrscheinlichkeit eines Teilintervalls $J \subset I$ steigt also proportional mit der Länge von J.
siehe Abb. 9
Erwartungswert und Varianz der Gleichverteilung:

$$E(X) = \frac{1}{2}(a + b) \qquad V(X) = \frac{1}{12}(b - a)^2$$

Wenn man sagt, dass zufällig ein Punkt in I gewählt wird, dann meint man als Wahrscheinlichkeitsmaß dafür die Gleichverteilung.

Beispiel 8.5

$\vartheta \in]0, \infty[$ unbekannt
$n \in \mathbb{N},\ X_1, \ldots, X_n \in [0, \vartheta]$ unabhängig ausgewählt
Frage: Wie groß ist ϑ?

Statistisches Modell: $([0, \infty[^n, \mathcal{B}^n_{[0,\infty[^n}, Q_\vartheta^{\otimes n} | \vartheta > 0)$
Q_ϑ ist Gleichverteilung auf $[0, \vartheta]$
$E(Q_\vartheta) = \int_0^\vartheta t \cdot \frac{1}{\vartheta} dt = \frac{\vartheta}{2}$

[60] ist oft $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}))$ oder $\mathbb{R}, \mathcal{B}$

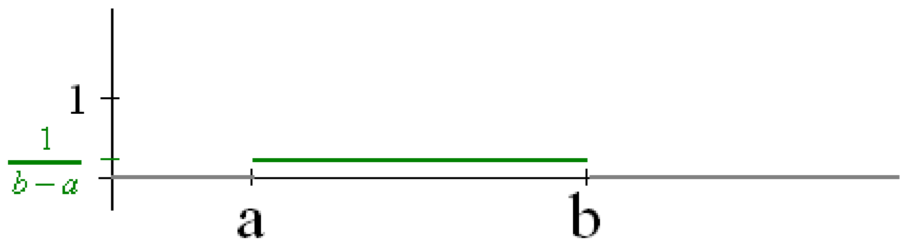

Abbildung 9: Dichte der Gleichverteilung

1. Fall: $T_n = 2M = \frac{2}{n}\sum_{k=1}^{n} X_k$ (doppelter Mittelwert)
laut dem (schwachen) Gesetz der großen Zahlen gilt:

$$P_\vartheta(|T_n - \vartheta| > \epsilon) \to 0 \quad (n \to \infty) \quad \forall \varepsilon > 0$$

$\Rightarrow T_n$ ist Schätzer

2. Fall: $\widetilde{T}_n = max\{X_1, \dots, X_n\}$

$$\begin{aligned} P_\vartheta(\widetilde{T}_n \le \vartheta - \varepsilon) &= P_\vartheta([X_1 \le \vartheta - \varepsilon] \cap \dots \cap [X_n \le \vartheta - \varepsilon]) \\ &= P_\vartheta(X_1 \le \vartheta - \varepsilon) \cdot \ldots \cdot P(X_n \le \vartheta - \varepsilon) \\ &= (\tfrac{\vartheta - \varepsilon}{\vartheta})^n \to 0 \quad (n \to \infty) \end{aligned}$$

[61] $\Rightarrow \widetilde{T}_n$ ist ebenfalls Schätzer

Welches ist der bessere Schätzer?

Vergleich der Schätzer:

I beide sind *konsistent*, d.h. $T_n \to \vartheta$, $\widetilde{T}_n \to \vartheta$ bei $(n \to \infty)$ stochastisch

II Der Schätzer T_n ist *erwartungstreu*, d.h. $E_\vartheta(T_n) = \vartheta$
Beweis:
$E_\vartheta(T_n) = \frac{2}{n}\sum_{k=1}^{n} E_\vartheta(X_k) = \frac{2}{n} \cdot n \cdot \frac{\vartheta}{2} = \vartheta$

III Der Schätzer $\widetilde{T}_n$ ist nicht erwartungstreu, aber asymptotisch erwartungstreu, d.h. $E_\vartheta(\widetilde{T}_n) \to \vartheta \quad (n \to \infty)$
Beweis:

[61] soll das Maximum aller Durchgänge $\le \vartheta - \varepsilon$ sein, dann muss $\le \vartheta - \varepsilon$ für alle Durchgänge gelten; $P_\vartheta(X_1 \le \vartheta - \varepsilon) = \frac{\vartheta - \varepsilon}{\vartheta}$ folgt aus der Gleichverteilung, denn es gilt hier $P_\vartheta(X_1 \le \vartheta - \varepsilon) = \frac{1}{\vartheta}(\vartheta - \varepsilon)$

$x \in [0, \vartheta],\ P_\vartheta(\widetilde{T}_n \leq x) = (\frac{x}{\vartheta})^n \rightsquigarrow$ Dichte $(\frac{nx^{n-1}}{\vartheta^n})$
$\Rightarrow E_\vartheta(\widetilde{T}_n) = \int_0^\vartheta t \frac{nx^{n-1}}{\vartheta^n} dt = \frac{n}{n+1}\vartheta$

neuer Schätzer $T_n^* = \frac{n+1}{n}\widetilde{T}_n$ ist erwartungstreu

IV Betrachtung der Varianz $\equiv$ Maß für die Güte des Schätzers [62]

$$\begin{aligned}
V_\vartheta(T_n) &= (\tfrac{2}{n})^2 V_\vartheta(X_1 + \cdots + X_n) \overset{(unabh.)}{=} \tfrac{4}{n^2} n V_\vartheta(X_1) \\
&= \tfrac{4}{n\vartheta} \int_0^\vartheta (t - \tfrac{\vartheta}{2})^2 dt \\
&= \tfrac{\vartheta^3}{3n} \quad \Rightarrow \text{starke Streuung} \\
V_\vartheta(\widetilde{T}_n) &= E_\vartheta(\widetilde{T}_n^2) - E_\vartheta(\widetilde{T}_n)^2 = \int_0^\vartheta t^2 \tfrac{nt^{n-1}}{\vartheta^n} dt - (\tfrac{n\vartheta}{n+1})^2 \\
&= \tfrac{n}{(n+1)^2(n+2)} \vartheta^2 \\
V_\vartheta(T_n^*) &= \tfrac{\vartheta^2}{n(n+2)}
\end{aligned}$$

$\Rightarrow T_n^*$ ist bester Schätzer für dieses Problem

19.06.

8.2 Maximum-Likelihood-Schätzer

Grundlage ist ein statistisches (Standard[63]-)Modell $(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$
Nun werde ein $x \in \mathcal{X}$ beobachtet.
Die Aufgabe ist dann mit Hilfe von x einen Parameter $\vartheta \in \Theta$ zu bestimmen (zu schätzen).
Dabei ist

$$P_\vartheta(\{x\}) = \varrho_\vartheta(x)$$

die Wahrscheinlichkeit mit der x eintritt, wenn ϑ der richtige Parameter ist.

Beobachtet man ein zufälliges Ereignis, dann ist das bestimmt keines, dass nur mit einer geringen Wahrscheinlichkeit eintritt, also eine Ausnahme ist, vielmehr ist es sicherlich eines, das sehr oft (wenn nicht am meisten) eintritt.[64]

Zumindest ist das die Annahme, die man beim Maximum-Likelihood-Schätzverfahren macht, denn hier sucht man den Parameter ϑ, der die Wahr-

[62] es versteht sich von selbst, dass eine kleine Varianz einen guten Schätzer ausmacht
[63] d.h. entweder ein diskretes oder stetiges Modell
[64] Wenn es 2 Millionen weiße Schwäne gibt und 200 schwarze und mir zufällig ein Schwan begegnet, dann ist das bestimmt kein schwarzer, oder?

scheinlichkeit, dass ein beobachtetes Ereignis x eintritt, maximiert. Symbolisch heißt das:

$$\text{wähle } \vartheta_0 \text{ so, dass } \vartheta_0 = \max_{\vartheta}\{\varrho_\vartheta(x)|\vartheta \in \Theta\}$$

ϑ_0 bildet somit einen Schätzer $T(x)$, den wir Maximum-Likelihood-Schätzer (MLE) nennen.

Definition 8.6 *Maximum-Likelihood-Schätzer*
$(\mathcal{X}, \mathcal{F}, P_\vartheta|\vartheta \in \Theta)$ *statistisches Standardmodell*

$$\varrho : (\mathcal{X} \times \Theta) \to [0, \infty[\qquad (x, \vartheta) \mapsto \varrho(x, \vartheta) = \varrho_\vartheta(x)$$

heißt Likelihood-Funktion und x heißt Beobachtungswert.

$$T : \mathcal{X} \to \Theta \text{ mit } \varrho(x, T(x)) = max\{\varrho(x, \vartheta)|\vartheta \in \Theta\} \quad \forall x \in \mathcal{X}$$

heißt Maximum-Likelihood-Schätzer (MLE). Er ist nicht (notwendigerweise) eindeutig bestimmt.

Beispiel 8.7 *Überprüfung von Produkten*

geg.: Lieferung von N gleichen Produkten
Überprüfung durch Stichprobe vom Umfang n,
Stichprobe enthalte x defekte Produkte
ges.: Gesamtzahl aller defekten Produkte $= \vartheta$

Modell: $\mathcal{X} = \{0, \ldots, n\}$ (Beobachtungsraum)
$\mathcal{F} = \mathcal{P}(\mathcal{X})$ (wird im folgenden in den diskreten Fällen nicht mehr mit angegeben)
$\Theta = \{0, \ldots, N\}$ (die Zahl der defekten Elemente muss ≥ 0 und $\leq N$ sein)
P_ϑ :

$$\varrho_\vartheta(x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} \quad \text{(hypergeometrische Verteilung)}$$

Problem: Maximum-Bestimmung von ϱ

Dazu Monnotonie-Untersuchung:
$\vartheta \in \mathbb{N},\ \vartheta \leq N \quad (\equiv \vartheta \in \Theta)$

$$\frac{\varrho_\vartheta(x)}{\varrho_{\vartheta-1}(x)} = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{\vartheta-1}{x}\binom{N-\vartheta+1}{n-x}} = \frac{\vartheta(N-\vartheta-(n-x)+1)}{(\vartheta-x)(N-\vartheta+1)} \geq 1^{65}$$

[65] ϱ_ϑ soll also monoton wachsen

$$\begin{aligned} &\iff \quad \vartheta N - \vartheta^2 - \vartheta n + \vartheta x + \vartheta \geq \vartheta N - \vartheta^2 + \vartheta - xN + \vartheta x - x \\ &\iff \quad \vartheta n \leq xN + x \end{aligned}$$

Mit

$$T(x) = \left\lfloor x\frac{N+1}{n} \right\rfloor$$ [66]

haben wir somit ein Maximum von ϱ gefunden, denn die Likelihood-Funktion $\vartheta \mapsto \varrho_\vartheta(x)$ ist monoton wachsend für alle $\vartheta \leq T(x)$ und monoton fallend für alle $\vartheta > T(x)$

Beispiel 8.8 *Schließen auf Gesamtzahl von Kugeln in Urne (vergl. [**?**] S. 60 - Schätzung eines Fischbestandes)*

geg.: es werden w Kugeln aus einer Urne mit nur schwarzen Kugeln entnommen
diese werden durch w weiße Kugeln ersetzt, dann wird durchgemischt
danch werden n Kugeln entnommen
davon seien x weiß

ges.: Gesamtzahl aller Kugeln in der Urne $= \vartheta$

Modell: $\mathcal{X} = \{0, \dots, n\}$
$\Theta = \{w, w+1, \dots\}$
P_ϑ :

$$\varrho(x, \vartheta) = \frac{\binom{w}{x}\binom{\vartheta - w}{n-x}}{\binom{\vartheta}{n}}$$

$$\frac{\varrho(x,\vartheta)}{\varrho(x,\vartheta-1)} \geq 1 \iff \vartheta \leq n\frac{w}{x}$$

$\Rightarrow$

$$T(x) = \left\lfloor n\frac{w}{x} \right\rfloor \qquad \text{(Begründung wie oben bei 8.7)}$$

Beispiel 8.9 *Schätzung der Erfolgswahrscheinlichkeit*

geg.: Bernoulli-Experimente
n-mal durchgeführt, dabei x Erfolge

ges.: Erfolgswahrscheinlichkeit $p = \vartheta$

Modell: $\mathcal{X} = \{0, \dots, n\}$
$\Theta = [0, 1]$
P_ϑ :

$$\varrho(x, \vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$$

[66] $\lfloor x \rfloor$ ist die "floor"-Funktion und gibt den ganzzahligen Anteil von x

Um das Maximum leichter berechnen zu können wird die Likelihood-Funktion logarithmiert. [67] Die resultierende Funktion nennt man dann *log-Likelihood-Funktion*:

$$\mathcal{L}_x(\vartheta) = \ln(\varrho(x,\vartheta)) = \ln\binom{n}{x} + x\ln\vartheta + (n-x)\ln(1-\vartheta)$$

$$\mathcal{L}'_x(\vartheta) = \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} \quad (=0)$$

$\Rightarrow$

$$T(x) = \frac{x}{n}$$

Definition 8.10 *Produktmodell / Gaußsches Modell*
Sei $(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$ ein statistisches Standardmodell für eine Beobachtung. Dann nennt man ein Modell über n solche Beobachtungen (das entspricht n unabhängigen Zufallsvariablen) n-faches Produktmodell:

$$(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta) \rightsquigarrow (\mathcal{X}^n, \mathcal{F}^{\otimes n}, P_\vartheta^{\otimes n} | \vartheta \in \Theta)$$

Ist die Verteilung des statistischen Modells eine Normalverteilung ($P_\vartheta = N_{\mu,v}$), so nennt man das resultierende n-fach-normalverteilte Produktmodell auch Gaußsches Modell

Beispiel 8.11 *Größe des Stichprobenraums schätzen*

geg.:	$x_1, \ldots, x_n \in [0, \vartheta]$ unabhängig
ges.:	$\vartheta > 0$

Modell:	$\mathcal{X} = [0, \infty[^n$	sollte das nicht $[0,\vartheta]^n$ sein?
	$\Theta = [0, \infty[$	
	$\mathcal{F} = \mathcal{B}^n$	
	$P_\vartheta = Q_\vartheta^n$ (Gleichverteilung auf $[0,\vartheta]^n$)	

$\int \ldots dQ_\vartheta = \frac{1}{\vartheta}\int_0^\vartheta \ldots dt$

fasst man jeden beobachteten Wert als eine Zufallsvariable auf ($x_i \rightsquigarrow X_i$), dann ist jedes X_i nach Q_ϑ verteilt:

$$P(X_i \leq t) = \begin{array}{ll} t & 0 \leq t \leq \vartheta \\ 1 & t \geq \vartheta \\ 0 & t < 0 \end{array}$$

[67] Aufgrund der Monotonie der Funktion $x \to \log t$ haben beide Funktionen das Maximum an der gleichen Stelle

$\Rightarrow$

$$E(X_i) = \frac{1}{\vartheta}\int_0^{\vartheta} t\,dt$$

$\Rightarrow$

$$\varrho(\underbrace{X_1,\ldots,X_n}_{X},\vartheta) = \begin{cases} \vartheta^{-n} & \text{falls } X_1,\ldots,X_n \le \vartheta \\ 0 & \text{sonst} \end{cases}$$

$\Rightarrow$

$$\widetilde{T}_n(X) = \max\{X_1,\ldots,X_n\} \text{ ist MLE}$$

Beispiel 8.12 *Gaußsches Modell (n-fach normalverteiltes Produktmodell)*

Ausgang eines Experiments sei $N_{\mu,v}$-verteilt (μ Mittelwert, $v = \sigma^2$ Varianz)

geg.: Ausgänge $x_1,\ldots,x_n$ unabhängig

ges.: $\mu, v \in \mathbb{R},\ v \ge 0$

Modell: $\mathcal{X} = \mathbb{R}^n$

$\Theta = \mathbb{R}\times]0,\infty[$

$\mathcal{F} = \mathcal{B}^n$

$P_\vartheta = N_{\mu,v}^{\otimes n} \mid (\mu,v) \in \Theta$

$$\varrho(x,\mu,v) = \prod_{j=1}^{n} \Phi_{\mu,v}(x_j) = \frac{1}{(2\pi v)^{\frac{n}{2}}} \exp(-\frac{\sum_{j=1}^{n}(x_j-\mu)^2}{2v})$$

log-Likelihood-Funktion:

$$\mathcal{L}_x(\vartheta) = -\frac{n}{2} ln(2\pi v) - \frac{1}{2v}\sum_{j=1}^{n}(x_j-\mu)^2$$

Schätzer für μ:

$$\frac{D\mathcal{L}_x(\vartheta)}{D\mu} = -\frac{1}{2v}\sum_{j=1}^{n} -2(x_j-\mu) = \frac{1}{v}\sum_{j=1}^{n}(x_j-\mu) = 0$$

$\Rightarrow$

$$\mu = \frac{1}{n}\sum_{j=1}^{n} x_j$$

Schätzer für v:

$$\frac{D\mathcal{L}_x(\vartheta)}{Dv} = -\frac{n}{2}\cdot\frac{1}{v} + \frac{1}{2}\cdot\frac{1}{v^2}\cdot\sum_{j=1}^{n}(x_j-\mu)^2 = 0$$

$\Rightarrow$

$$v = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)^2$$

Beim Überprüfen der hinreichenden Bedingungen bestätigt sich, dass die angegebenen Werte für μ und v maximal sind. Es folgt:

Satz 8.13
gegeben sei ein Gaußsches Modell:
$X_1, \ldots, X_n \quad N_{\mu,v}$*-verteilt*
$x = (x_1, \ldots, x_n)$ *Stichprobe*
$\Rightarrow$

$$T(x) = (\mu(x), v(x)) = \left(\frac{1}{n}\sum_{j=1}^{n} x_j,\ \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu(x))^2\right)$$

ist MLE im n-fachen Produktmodell der $N_{\mu,v}$*-Verteilung*