Stochastik Wintersemester 2009
Leibniz Universität Hannover
Vorlesungsmitschrift

Dozent: Prof. Dr. L. Baringhaus

Mitschrift von
Ronald Becher

30. Oktober 2009

Zusammenfassung

Diese Mitschrift wird erstellt im Zuge der Vorlesung Stochastik A im Wintersemester 2009 an der Leibniz Universität Hannover. Obwohl mit großer Sorgfalt geschrieben, werden sich sicherlich Fehler einschleichen. Diese bitte ich zu melden, damit sie korrigiert werden können. Ihr könnt euch dazu an jeden der genannten Autoren (mit Ausnahme des Dozenten) wenden.

Dieses Skript wird primär über Github verteilt und „gepflegt“. Dort kann man es auch forken, verbessern und dann (idealerweise) ein „Pull Request“ losschicken. Siehe auch Github Guides. Auch wenn ihr gute Grafiken zur Verdeutlichung beitragen könnt, dürft ihr diese gerne schicken oder selbst einfügen (am besten als LaTeX-geeignete Datei (Tikz, SVG, PNG, . . .)).

Hinweis: Es wird o.B.d.A. davon abgeraten, die Vorlesung zu versäumen, nur weil eine Mitschrift angefertigt wird! Selbiges gilt für den Übungsbetrieb! Stochastik besteht ihr nur, wenn ihr auch zu der Veranstaltung geht!

Inhaltsverzeichnis

Teil I
Einführung / Organisatorisches

Die Klausur wird am **01.02.2010 von 9:15 - 10:45** stattfinden, es wird wöchentliche Haus- und Stundenübungen geben.
Die abgegebenen Hausübungen werden korrigiert und in den Übungsgruppen zurückgegeben werden, es gibt allerdings keine Bonuspunkteregelung. Übungsgruppen wird es folgende geben:

- Mo, 12:15 - 13:00 Uhr (F309)
- Mo, 14:15 - 15:00 Uhr (G123) [max. 20 Teilnehmer]
- Di, 10:15 - 11:00 Uhr (F428)
- Di, 12:15 - 13:00 (F442)
- Mi, 12:15 - 13:00 (B305)

Teil II
Grundlegende Wahrscheinlichkeitstheorie

„Was ist das richtige Modell, um ein Experiment zu beschreiben?“ ist die wichtigste Frage, die die mathematische Statistik zu beantworten hat, und diese Modelle an die Wirklichkeit anzupassen, ist die hauptsächliche Aufgabe der Stochastik. Der letzte Teil wird vor allem in der Vorlesung **Stochastik B** behandelt.

1 Wahrscheinlichkeitsräume

Zufallsexperimente werden beschrieben durch

- die Menge der möglichen Ergebnisse Ω
  Im **Würfelwurf** sei beispielsweise $\Omega := \{1, 2, 3, 4, 5, 6\}$
- die Menge $\mathfrak{A} \subset \mathfrak{P}(\Omega)$[1] $:= \{A; A \subset \Omega\}$ als Menge der interessierenden Ergebnisse.
  $\mathfrak{A}$ ist eine so genannte σ-Algebra auf Ω, d.h.

[1] Potenzmenge

1. $$\Omega \in \mathfrak{A} \tag{1}$$
   (Die Grundmenge Ω ist in $\mathfrak{A}$ enthalten)
2. $$A \in \mathfrak{A} \iff A^c \in \mathfrak{A} \tag{2}$$
   (Wenn $\mathfrak{A}$ eine Teilmenge $A \in \Omega$ als mögliches Ergebnis enthält, dann auch deren Komplement $A^c = \Omega \setminus A$)
3. $$A_1, A_2, \dots \in \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A} \tag{3}$$
   (Wenn für jede natürliche Zahl n die Menge $A_n \in \mathfrak{A}$ ist, so ist auch die abzählbare Vereinigung aller $A_n \in \mathfrak{A}$)

Im **Würfelwurf** seien beispielsweise gültige Ereignisse:

- $A := \{2, 4, 6\}$ das Ergeignis „Es fällt eine gerade Zahl“, und
- $A := \{5, 6\}$ das Ergeignis „Es fällt eine Zahl größer gleich 5“.

- durch ein Wahrscheinlichkeitsmaß P (lat: „**P**robabilitas“=Glaubwürdigkeit) mit $P : \mathfrak{A} \to [0, 1]$, also eine Mengenfunktion auf $\mathfrak{A}$ mit den Eigenschaften

1. $$P(\Omega) = 1 \tag{4}$$
   (Die so genannte “Normiertheit”).
2. Für jede Folge von **paarweise disjunkten** $A_1, A_2, \dots \in \mathfrak{A}$ gilt:
   $$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \tag{5}$$
   σ-Additivität von P

   Für $A \in \mathfrak{A}$ ist $P(A)$ die Wahrscheinlichkeitsfunktion (für das Eintreten) von A

Das Tripel $(\Omega, \mathfrak{A}, P)$ heißt **Wahrscheinlichkeitsraum**.

$\omega \in \Omega$ nennen wir das Ergebnis des Zufallsexperimentes. Das Ereignis $A \in \mathfrak{A}$ tritt genau dann ein, wenn $\omega \in A$.

$$\omega \in \bigcup_{n=1}^{\infty} A_n \iff \exists n \in \mathbb{N} \wedge \omega \in A_n \tag{6}$$

Das durch $\bigcup_{n=1}^{\infty} A_n$ beschriebene Ereignis tritt genau dann ein, wenn mindestens eines der Ereignisse $A_n,\ n \in \mathbb{N}$ eintritt.

Auf die Frage, ob auch $\emptyset$ eine gültige Menge ist, sei gesagt, dass

$$\emptyset = \Omega^c \in \mathfrak{A}$$

das „unmögliche Ereignis“ genannt wird (Die Antwort ist also “ja”).

Seien $A_1, A_2, \ldots \in \mathfrak{A}$. Dann

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathfrak{A} \tag{7}$$

Anmerkung: $\bigcap_{n=1}^{\infty} A_n$ tritt genau dann ein, wenn alle $A_n, n \in \mathbb{N}$ eintreten.

Wenn nun $A, B \in \mathfrak{A}$, dann gilt

- $$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \cdots \in \mathfrak{A}$$
  (Die Vereinigung von A, B ist auch ein interessierendes Ereignis und steht für “das Ereignis A oder B treten ein”)

- $$A \cap B = A \cap B \cap \Omega \cap \Omega \cap \cdots \in \mathfrak{A}$$
  (Die Schnittmenge von A, B ist auch ein interessierendes Ereignis und steht für “das Ereignis A **und** B treten ein”)

- $$A \Delta B = (A \cap B^c) \cup (B \cap A^c)$$
  (Die “symmetrische Differenz” beschreibt das Ereignis “Entweder A oder B”)

- $A \subseteq B$ heißt: “Das Ereignis A hat das Ereignis B zur Folge”.
  (Kurz erklärt: Wenn A eintritt, folgt $\omega \in A \;\wedge\; A \subseteq B \Rightarrow \omega \in B$)

1.1 Eigenschaften von Wahrscheinlichkeits-Maßen

- Aus $P(\emptyset) = P(\emptyset \cup \emptyset \cup \ldots) = \sum_{n=1}^{\infty} P(\emptyset)$ folgt $P(\emptyset) = 0$.

- Für $A, B \in \mathfrak{A}$ mit $A \subseteq B$ ist

  $$P(B) = P(A \cup B \cap A^c) = P(A \cup B \cap A^c \cup \emptyset \cup \ldots) = P(A) + P(B \cap A^c) \geq P(A) \tag{8}$$

  Also gelten folgende Eigenschaften / Gesetzmäßigkeiten:

  – $$P(A) \leq P(B) \tag{9}$$

(Isotonie)

–

$$P(B \cap A^c) = P(B) - P(A) \tag{10}$$

(Subtraktivität)

1.2 Beispiel

$\emptyset \neq \Omega$ sei endlich, $|\Omega|$ Anzahl der Elemente von Ω und $\mathfrak{A} = \mathfrak{P}(\Omega)$.
Sei P definiert durch

$$P(A) = \frac{|A|}{|\Omega|},\ A \subset \Omega \tag{11}$$

P heißt „diskretes Laplace'sches W-Maß„ auf Ω.
Es ist

$$P(\{\omega\}) = \frac{1}{|\Omega|},\ \omega \in \Omega \tag{12}$$

also folgert man: „Alle möglichen Ergebnisse sind **gleich wahrscheinlich**.“

2 Grundformeln der Kombinatorik

$n, r \in \mathbb{N}$, M_n ist eine n-elementige Menge, o.B.d.A. $M_n = \{1, \dots, n\}$.

- $$P_n^r = \{(x_1, \dots, x_r);\ x_i \in M_n, 1 \leq i \leq r\} \tag{13}$$
  das sind die r-Permutationen aus M_n mit Wiederholung.

- $$P_n^{(r)} = \{(x_1, \dots, x_r);\ x_i \in M_n, 1 \leq i \leq r, \text{paarweise verschieden}\}\ (r \leq n) \tag{14}$$
  das sind die r-Permutationen aus M_n ohne Wiederholung.

- $$K_n^{(r)} = \{(x_1, \dots, x_r);\ x_i \in M_n, 1 \leq i \leq r,\ x_1 < x_2 < \dots < x_n\}\ (r \leq n) \tag{15}$$
  das sind die r-Kombinationen aus M_n ohne Wiederholung.

- $$K_n^r = \{(x_1, \dots, x_r);\ x_i \in M_n, 1 \leq i \leq r,\ x_1 \leq x_2 \leq \dots \leq x_n\}\ (r \leq n) \tag{16}$$
  das sind die r-Kombinationen aus M_n mit Wiederholung.

2.1 Zur Erklärung

1. Tupel in Permutationen sind nicht angeordnet, hier ist die Reihenfolge wichtig!
2. Tupel in Kombinationen sind o.B.d.A. angeordnet, wir können sie also "vergleichen".
3. Bei Varianten **mit Wiederholung**, dürfen also zwei Elemente "nebeneinander" auch gleich sein, bei Varianten **ohne Wiederholung** ist das nicht der Fall, hier müssen die Elemente also echt unterschiedlich sein.
4. Bei einer r-Permutation/Kombination haben wir eine r-elementige Teilmenge (die also kleiner oder gleich groß der "großen" Menge M_n ist).
5. "mit Wiederholung" nennt man auch "mit Zurücklegen"

2.2 Anschauliche Beispiele

- Bei Kombinationen interessiert uns ausschließlich, welche Elemente wir aus M_n bekommen, ein gutes Beispiel für eine Kombination

  ohne Wiederholung ist also das übliche Lotto-Spiel "6 aus 49". Hier ist es uns egal, ob die Reihenfolge $(1, 2, 3, 4, 5, 6)$ ist oder $(6, 5, 4, 3, 2, 1)$.

mit Wiederholung ist die Frage, ob wir mindestens zwei Mal in fünf Versuchen eine Zahl ≥ 5 würfeln

- Bei Permutationen interessiert uns neben der Auswahl auch die Anordnung (man sagt: die Elemente sind "unterscheidbar"). Eine Permutation

ohne Wiederholung ist z.B. ein Kinobesuch. Ob Tim zwischen Tina und Claudia sitzt, ist für ihn nicht gleichwertig mit der Variante, zwischen Klaus und Jochen zu sitzen.
Anders ausgedrückt gilt in Permutationen

$$(Claudia, Tim, Tina, Klaus, Jochen) \neq (Claudia, Tina, Klaus, Tim, Jochen)$$

mit Wiederholung ist die Frage, ob ich beim Roulette einen aufsteigenden Run habe, also erst eine 0 fällt, dann eine 1, dann eine 2, usw. Das Standardbeispiel hierbei ist eine Urne, wo ich Kugeln oder Zettel heraushole und wieder hereinlege.
Eine Permutation mit Wiederholung ist also eine "Auswahl mit Wiederholungen unter Beachtung der Reihenfolge".

2.3 Rechenbeispiele

Es gilt

- $$|P_n^r| = n^r \tag{17}$$

  (Erklärung: Man greift r Mal in die Urne, schaut sich einen der n Zettel an und legt ihn wieder herein, es gibt also danach genauso viele Möglichkeiten zu ziehen (also n))

- $$|P_n^{(r)}| = \begin{cases} r \leq n & n \cdot (n-1) \cdot \ldots \cdot (n-r+1) \\ r = n & r! \end{cases} \tag{18}$$

  (Erklärung: Lege ich meine Zettel nicht wieder in die Urne zurück, habe ich beim nächsten Versuch natürlich eine Möglichkeit weniger, welche Zettel ich ziehen könnte)

- $$r! \cdot |K_n^{(r)}| = |P_n^{(r)}| \tag{19}$$
$$\Updownarrow$$
$$|K_n^{(r)}| = \frac{n(n-1)\dots(n-r+1)}{r!} = \binom{n}{r} \tag{20}$$
$$\Downarrow$$
$$|\{A; A \subset M_n, |A| = r\}| = \binom{n}{r} \tag{21}$$
$$\Downarrow$$
$$|\mathfrak{P}(M_n)| = \sum_{r=0}^{r} \binom{n}{r} 1^r 1^{n-r} = (1+1)^n = 2^n \tag{22}$$

  (21) Gilt auch noch für $r = 0$.
  Erklärungen:
  (19) *und äquivalent* (20) *kann man sich so vorstellen, dass man sich seine r-elementige Teilmenge (ohne Zurücklegen) zusammensuchen kann, aber die verschiedenen Möglichkeiten, sie zu permutieren ("die Kinobesucher umzusetzen"), interessieren den Betrachter hier nicht, also müssen wir die aus der Rechnung rausnehmen.*
  (21) *ist die Anzahl aller r-elementigen Teilmengen von M_n hoch r, und wegen* (20) *ebenfalls $\binom{n}{r}$.*
  (22) *ist die Länge der Potenzmenge, also aller möglichen Kombinationen von Elementen aus M_n*

- $K_n^r \to (x_1, \dots, x_r) \underbrace{\longrightarrow}_{\textit{ist bijektiv}} (x_1, x_1 + 1, \dots, x_r + r - 1) \in K_{n+r-1}^{(r)}$
  $\Rightarrow |K_n^r| = |K_{n+r-1}^{(r)}| = \binom{n+r-1}{r}$
  Identifiziere $(x_1, \dots, x_r) \in K_n^r$ mit dem Besetzungszahlvektor $(k_1, \dots, k_n)$, wobei $k_i = |\{j \in \{1, \dots, r\}; x_j = i\}|,\ 1 \le i \le n$. Dann ist $K - 1 + \dots + k_n = r$.
  Dann ist $|\{(K_1, \dots, k_n) \in \mathbb{N}_0^n; k_1 + \dots + k_n = r\}| = \binom{n+r-1}{r}$
  Folgerung: $|\{(K_1, \dots, k_n) \in \mathbb{N}^n; k_1 + \dots + k_n = r\}| = \binom{n+(r-n)-1}{r-n} = \binom{r-1}{r-n} = \binom{r-1}{n-1}, r \ge n$

Aus der Menge der Permutationen betrachten wir eine spezifische Teilmenge.
Sei $k_1, \ldots, k_n \in \mathbb{N}_0$ mit $k_1 + \ldots + k_n = r$. Einen so genannter "Besetzungszahlvektor" drückt man beispielsweise so aus:

$$\begin{aligned} &\{(x_1, \ldots, x_n) \in \mathbb{P}_n^r;\ |\{j; j \in \{1, \ldots, r\}, x_j = i\}| = k_i, 1 \leq i \leq n\}| \\ =&\{(x_1, \ldots, x_n) \in \mathbb{P}_n^r;\ \text{Genau } k_i \text{ der } r \text{ Komponenten } (x_1, \ldots, x_r) \text{ sind gleich } i, 1 \leq i \leq n\} \end{aligned} \tag{23}$$

Die Länge oder Mächtigkeit dieser Menge ist wie folgt:

$$\begin{aligned} |23| &= \binom{r}{k_1}\binom{r-k_1}{k_2}\binom{r-k_1-k_2}{k_3}\cdots\binom{r-k1-\ldots-k_{n-1}}{k_n} \\ &= \frac{r!}{k_1!\cancel{(r-k_1)!}} \cdot \frac{\cancel{(r-k_1)!}}{k_2!\cancel{(r-k_1-k_2)!}} \cdots \frac{\cancel{(r-k_1-\ldots-k_n)!}}{k_n!} \\ &= \frac{r!}{k_1!\cdots k_n!} \end{aligned}$$

2.4 Binomischer Lehrsatz

Es folgt aus dem (bekannten) binomischen Lersatz durch Verallgemeinerung der multinomiale Lehrsatz.

$$n^r = \sum_{\substack{(k_1,\ldots,k_n)\in\mathbb{N}_0^n \\ k_1+\ldots+k_n=r}} \frac{r!}{k_1!\cdots k_n!} \tag{24}$$

$$|24| = \binom{n+r-1}{r} \tag{25}$$

2.5 Beispiel

O.B.d.A. sind r Personen im Raum, $n = 365$ seien die Anzahle der Tage im Jahr und $r \leq n$.
Gesucht: Wahrscheinlickeit, dass mindestens 2 der r Personen am gleichen Tag Geburtstag haben.

- $\Omega = \{(x_1, \ldots, x_r); x_i\ \}$
- $\mathfrak{A} = \mathfrak{P}(\Omega)$
- $P(A) = \frac{|A|}{|\Omega|}, A \subseteq \Omega$
- Alle möglichen r-Tupel von Geburtstagen sind gleich wahrscheinlich, also
- $A = \{(x_1, \ldots, x_r) \in \Omega; i, j \in \{1, \ldots, r\}, i \neq j, x_i = x_j\}$

- $P(A) = 1 - P(A^c)$
  In vielen Fällen ist es ein valider "Trick", statt der Ergebnismenge das Gegenergebnis zu berechnen
- $A^c = \{(x_1, \dots, x_r) \in \Omega; x_1, \dots, x_r$ paarweise verschieden$\}$
- $|A^c| = n(n-1)\cdots(n-r+1)$
- $P(A) = 1 - \frac{n(n-1)\cdots(n-r+1)}{n^r}$

$$P(A) = \begin{cases} > 1/2 & \text{wenn } r > 23, \\ 0,706 & \text{wenn } r = 30, \\ 0,994 & \text{wenn } r = 60. \end{cases}$$

2.6 Die Siebformel

Seien A_1, A_n Ereignisse in W-Raum $(\Omega, \mathfrak{A}, P)$

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1 \cup (A_2 \cap (A_1 \cap A_2)^c)) \\ &= P(A_1) + P(A_2 \cap (A_1 \cap A_2)^c) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \end{aligned} \tag{26}$$

Dies führt uns zu folgendem: Seien $A_1, \dots, A_n$ Ereignisse. Dann

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k-1}^{n} (-1)^{k-1} \sum_{1 \leq j_1 < \ldots < j_k \leq n} P(A_{j_1} \cap \ldots \cap A_{j_k}) \tag{27}$$

Wie leicht überprüfbar ist, ist (26) eine Anwendung von (27).

2.7 Beispiel

Sei

- $\Omega = \{\pi; \pi : \{1, \dots, n\} \to \{1, \dots, n\}$ bijektiv
- $|\Omega| = n!$
- $\mathfrak{A} = \mathfrak{P}(\Omega)$
- P das diskrete Laplace'sche W-Maßauf Ω
- $A_i = \{\pi \in \Omega; \pi(i) = i\},\ 1 \leq i \leq n\}$
- $\bigcup_{i=1}^{n} A_i = \{\pi \in \Omega;$ Es gibt ein $i \in \{1, \dots, n\}$ mit $\pi(i) = i\}$

Dann

$$\begin{aligned}P(\bigcup_{i=1}^{n} A_i) &= \sum_{k-1}^{n} (-1)^{k-1} \sum_{1\leq j_1<\ldots<j_k\leq n} \frac{|A_{j_1} \cap \ldots \cap A_{j_k}|}{|\Omega|} \\ &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{\ldots} \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \cdot \frac{(n-k)!}{n!} \\ &= \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!}\end{aligned} \tag{28}$$

und wegen

$$\begin{aligned}\bigcap_{i=1}^{n} A_i^c &= \left(\bigcup_{i=1}^{n} A_i\right)^c \\ &= \{\pi \in \Omega; \forall i \in \{1,\ldots,n\} \text{ ist } \pi(i) \neq i\}\end{aligned} \tag{29}$$

ist

$$\begin{aligned}P\left(\bigcap_{i=1}^{n} A_i^c\right) &= 1 - \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!} \\ &= \sum_{k=0}^{n} (-1)^{k} \frac{1}{k!} \\ &\xrightarrow[n\to\infty]{} \sum_{k=0}^{\infty} (-1)^{k} \frac{1}{k!} \\ &= e^{-1} \\ &\approx 0.37\end{aligned} \tag{30}$$

2.8 Weitere Beispiele für W-Maße

Sei $\emptyset \neq \Omega$, $\mathfrak{A}$ σ-Algebra mit $\{\omega\} \in \mathfrak{A}$ für jedes $\omega \in \Omega$.
Sei P W-Maß auf Ω mit der Eigenschaft, dass eine abzählbare Menge $\Omega_o \subseteq \Omega$ existiert mit $P(\Omega_0) = 1$
Dann

$$P(\Omega_0^c) = 1 - P(\Omega_0^c) = 0$$

Für $A \in \mathfrak{A}$ gilt

$$\begin{aligned} P(A) &= P\left((A \cap \Omega_0) + (A \cap \Omega_0^c)\right) \\ &= P(A \cap \Omega_0) + P(A \cap \Omega_0^c) \\ &= P\left(\sum_{\omega \in A \cap \Omega_0} \{\omega\}\right) \\ &= \sum_{\omega \in A \cap \Omega_o} P\left(\{\omega\}\right) \\ &= \sum_{\omega \in \Omega} P\left(\{\omega\}\right) \delta_\omega(A) \end{aligned} \tag{31}$$

Für $\omega \in \Omega$ heißt $\delta_\omega : \mathfrak{A} \to [0,1]$, definiert durch

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

das Einpunktmaß oder **Dirac-Maß in** ω.

Es gilt also

$$P = \sum_{\omega \in \Omega_0} P(\{\omega\}) \delta_\omega.$$

Solche W-Maße heißen **diskrete W-Maße**.

2.9 Die Borelsche σ-Algebra

Betrachte den Fall $\Omega = \mathbb{R}$

$$\mathfrak{B} = \mathfrak{A} = \bigcap_{\substack{\mathfrak{A}^* \ \sigma\text{-Algebra auf } \mathbb{R} \\ \mathfrak{A}^* \supseteq \{(-\infty, x]; x \in \mathbb{R}\}}} \mathfrak{A}^*$$

ist die Borelsche σ-Algebra auf $\mathbb{R}$.
Beachte

$$\mathfrak{B} \neq \mathfrak{P}(\mathbb{R})$$

$\mathfrak{B}$ enthält alle offenen, abgeschlossenen und ohne Intervalle Teilmengen von R.

2.10 Verteilungsfunktionen

Sei $F : \mathbb{R} \to \mathbb{R}$ Funktion mit folgenden Eigenschaften:

1. F ist monoton wachsend

2. F ist rechtsseitig stetig

3. $$\lim_{x \to -\infty} F(x) = 0$$

4. $$\lim_{x \to +\infty} F(x) = 1$$

Es existiert genau ein W-Maß P auf $\mathfrak{B}$ (auf $\mathbb{R}$) mit der Eigenschaft, dass

$$P\left((-\infty, x]\right)) = F(x), \ \forall x \in \mathbb{R} \tag{32}$$

F heißt die zu P gehörige **Verteilungsfunktion**.
Für $(a, b]$ mit $-\infty < a < b < +\infty$ ist

$$P((a, b]) = P((-\infty, b]) - P(-\infty, a]) = F(b) - F(a) \tag{33}$$

2.11 Verteilungsdichten

Sei $f : \mathbb{R} \to \mathbb{R}_+$ (uneigentlich) Riemann-integrierbar mit

$$\int_{-\infty}^{+\infty} f(t)dt = 1 \tag{34}$$

Setze

$$F(x) = \int_{-\infty}^{+\infty} f(t)dt, \ -\infty < x < +\infty s \tag{35}$$

F ist Verteilungsfunktion eines W-Maßes P auf $\mathbb{R}$.
f heißt **Wahrscheinlichkeitsdichte** (oder einfach „Dichte“) von P.

2.11.1 Spezialfälle

- $$f(t) = \begin{cases} 0 & , \ x < a, \\ \frac{1}{b-a} & , \ a \leq x \leq b, \\ 0 & , \ x > b. \end{cases}$$

  $$F(x) = \begin{cases} 0 & , \ x < a, \\ \frac{x-a}{b-a} & , \ a \leq x \leq b, \\ 1 & , \ x > b. \end{cases}$$

- $$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{(t-a)^2}{\sigma^2}), -\infty < t < +\infty, (a \in R, \sigma^2 > 0)$$

(Anmerkung: Diese Verteilungsdichte wird auch Standardnormalverteilung genannt. Die Verteilungsfunktion dafür produziert die so genannte Gauß'sche Glockenkurve)
Dann ist

$$\int_{-\infty}^{+\infty} f(t)dt = 1$$

und somit

$$\begin{aligned} F(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{(t-a)^2}{\sigma^2})dt \\ &= \phi(\frac{x-a}{\sigma}) \\ \phi(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{t^2}{2})dt, x \in \mathbb{R} \end{aligned}$$

Sei $(R, \mathfrak{B}, P)$ das W-Raum unserer Wahl. P habe Dichte f, also eine Teilfunktion $f : R \rightarrow, \int_{-\infty}^{+\infty} f(t)dt) = 1$. $F(x) = P((-\infty, x]) = \int_{-\infty}^{+\infty} f(t)dt), \ x \in R$.

Allgemein gilt $P(E) = \int_E f(t)dt), \ E \in \mathfrak{B}(geeignet)$.

Beispiel:

$$f(t) = \begin{cases} 0 & t < 0 \\ \lambda \cdot exp(-\lambda t) & t \geq 0 \end{cases}$$

Aus

$$\int_0^x \lambda exp(-\lambda t)dt = [-exp(-\lambda t)]_o^x = 1 - exp(-\lambda x)$$

folgt

$$F(x) = \int_{-\infty}^{+\infty} f(t)dt) = \begin{cases} 0 & x < 0 \\ 1 - exp(-\lambda x) & x \geq 0 \end{cases}$$

3 Weitere Beispele von Wahrscheinlichkeitsmaßen

Sei $(R, \mathfrak{A}, P)$ W-Raum. Seien $A_1, A_2, \ldots \in \mathfrak{A}$. Dann gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \tag{36}$$

Diese Eigenschaft wird als **Sub-σ-Additivität** bezeichnet.

$$\bigcup_{n=1}^{\infty} A_n = A_1 + \sum_{n=2}^{\infty} (A_n \cap A_1^c \cap \ldots \cap A_{n-1}^c) \tag{37}$$

Wir können die Vereinigung auch als Summe von disjunkten Ereignissen auffassen und gelangen zu dieser Schreibweise in BEFORE

Deshalb gilt auch BEFORE+1

$$\begin{aligned} &P\left(\bigcup_{n=1}^{\infty} A_n\right) \\ &= \sum P(A_1) + \sum_{n=2}^{\infty} P(A_n \cap A_1^c \cap \ldots \cap A_{n-1}^c) \\ &\leq \sum_{n=1}^{\infty} P(A_n) \end{aligned} \tag{38}$$

Speziell folgt hieraus die so genannten “Sub-Additivität von P”:

$$P\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{j=1}^{n} P(A_j) \tag{39}$$

Seien $A_n \in \mathfrak{A}, n = 1, 2, \ldots, A \in \mathfrak{A}$. Wir schreiben

- $A_n \uparrow A$ für $A_1 \subset A_2 \subset \ldots, A = P\left(\bigcup_{j=1}^{n} A_j\right)$
  Man sagt “A_n konvergiert **isoton** gegen A”)

- $A_n \downarrow A$ für $A_1 \subset A_2 \subset \ldots, A = P\left(\bigcap_{j=1}^{n} A_j\right)$
  Man sagt “A_n konvergiert **antiton** gegen A”)

4 Stetigkeitseigenschaften von W-Maßen

1. $A_n \uparrow A \Longrightarrow \lim_{n\to\infty} P(A_n) = P(A)$

2. $A_n \downarrow A \Longrightarrow \lim_{n\to\infty} P(A_n) = P(A)$

4.1 Beweis zu (1)

$$A = \bigcup_{n=1}^{\infty} A_n = A_1 + \sum_{n=2}^{\infty} (A_n \cap A_{n-1}^c) \tag{40}$$

Also

$$\begin{aligned}
P(A) &= P(A) + \sum_{n=2}^{\infty} P(A_n \cap A_{n-1}^c) \\
&= P(A) + \sum_{n=2}^{\infty} P(A_n - A_{n-1}^c) \\
&= P(A) + \lim_{m\to\infty} \sum_{n=2}^{m} P(A_n - A_{n-1}) \\
&= P(A_1) + \lim_{m\to\infty} (P(A_m - A_1)) \\
&= \lim_{m\to\infty} P(A_m)
\end{aligned} \tag{41}$$

4.2 Beweis zu (2)

Mit Gegenmenge ...

5 Beispiel

$\Omega = R, x \in R, (x_n) \subset R$ Folge, die monoton fallend gegen x konvergiert.

$$(-\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, x_n]$$

P ist W-Maß auf R

$$F(x) = P((-\infty, x]) = \lim_{n\to\infty} P((-\infty, x_n]) = \lim_{n\to} F(x_n)$$

Hieraus folgt allgemein:
Ist $(x_n) \subset R$ eine Folge, die von rechts gegen x konvergiert (d.h. $x \leq x_n$ für jedes $n \leftarrow N$ und $\lim_{n\to} x_n = x$), so gilt

$$\lim_{n\to\infty} F(x_n) = F(x)$$

("F ist rechtsseitig stetig")

Teil III
Zufallsvariablen und ihre Verteilung

Wir gehen von einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ aus, R sei unsere gesuchte Bild-Menge, $\mathfrak{S}$ σ-Algebra auf R.

$$X : \Omega \to R$$

sei Wahrscheinlichkeitsfunktion mit der Eigenschaft, dass für jedes $B \in \mathfrak{S}$ gilt

$$X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\} \in \mathfrak{A} \tag{42}$$

heißt Zufallsvariable.

- $$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\} = X^{-1}(B) \tag{43}$$
- $$P(X \in B) \text{ für } P(\{X \in B\}) = P(X^{-1}(B)) \tag{44}$$
- Durch
  $$P^X : \mathfrak{S} \to [0,1]$$
  definiert durch
  $$P(B) = P(X^{-1}(B)) = P(X \in B),\ B \in \mathfrak{S}$$
  wird ein W-Maß P^X auf $\mathfrak{S}$ definiert.
- $$X^{-1}\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} X^{-1}(B_n)$$

P^X heißt die **Verteilung von** X.

6 Spezialfälle

1. $R = \mathbb{R}, \mathfrak{S} = \mathfrak{B}$. Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt **reelle Zufallsvariable**.
   $F(x) = P^X((-\infty, x]))P(X \in ((-\infty, x])) = P(X \leq x), x \in \mathbb{R}$ ist die Verteilungsfunktion von P^X, kurz: Die Verteilungsfunktion von X.
   Hat P^X die Dichte f, so ist
   $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, x \in \mathbb{R}$$
   Sprechweise: X hat die Dichte f.

2. $R \neq \emptyset$, $\mathfrak{S}$ enthalte die Mengen $\{x\}$ für $x \in R$.
   Es gibt eine abzählbare Menge $R_0 \subset R$ mit $P(X \in R_0) = 1$.
   Dann ist $P(X \notin R_0) = 1 - P(X \in R_0) = 0$
   Es gilt daher
   $$P(X \in R_0^c \cap B) = 0, \ \forall B \in \mathfrak{S}$$
   und
   $$\begin{aligned} P(X \in B) &= P(X \in R_0 \cap B) + \cancel{P(X \in R_0^c \cap B)} \\ &= \sum_{x \in R_0 \cap B} P(X = x) \\ &= \sum_{x \in R_0} P(X = x)\delta_x(B) \end{aligned}$$
   Also
   $$P^X = \sum_{x \in R_0} P(X = x)\delta_x$$
   X heißt dann diskret verteilt.
   P^X ist festgelegt durch die Werte $P(X = x), x \in R_0$

7 Beispiel: Würfelwurf von 2 echten Würfeln

- $\Omega = \{(\omega_1, \omega_2); \omega_1, \omega_2 \in \{1, \dots, 6\}\}$
- $|\Omega| = 6^2 = 36$
- $\mathfrak{A} = \mathfrak{P}(\Omega)$
- $P(A) = \frac{|A|}{|\Omega|}, A \subset \Omega$
- $X : \Omega \to \mathbb{R}$, $X(\omega_1, \omega_2) = \omega_1 + \omega_2, (\omega_1, \omega_2) \in \Omega$
- $P(X = k)$

Also

$$\begin{aligned} P(X = k) &= P(\{(\omega_1, \omega_2) \in \Omega; \omega_1 + \omega_2 = k\}) \\ &= \frac{|\{(\omega_1, \omega_2) \in \Omega; \omega_1 + \omega_2 = k\}|}{36} \\ &= \begin{cases} \frac{1}{36} & , k = 2 \\ \frac{2}{36} & , k = 3 \\ \frac{3}{36} & , k = 4 \\ \vdots & \\ \frac{1}{36} & , k = 12 \end{cases} \end{aligned} \tag{45}$$

Natürlich kann man das auch allgemeiner schreiben:
Sei $R = \mathbb{R}^d = \mathfrak{B}^d$. X heißt dann d-dimensionaler Zufallsvektor $X : \Omega \to \mathbb{R}^d$.
$X = (X_1, \dots, X_d)$ mit reellen Zufallsvariablen $X_1, \dots, X_n$ die Komponenten von X. X heißt nun **Zufallsvektor**.

Sei $f : \mathbb{R}^d \to \mathbb{R}_+$ mit

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_d) dt_1 \cdots dt_d = 1$$

f heißt Dichte des Zufallsvektors X, wenn gilt

$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_d) dt_1 \cdots dt_d, (x_1, \dots, x_d) \in \mathbb{R}^d$$

F heißt Verteilungsfunktion des Zufallsvektors.

8 Beispiele für Zufallsvariablen und ihre Verteilungen

Sei $(\Omega, \mathfrak{A}, P)$ W-Raum und $X : \Omega \to \mathbb{R}$ reelle Zufallsvariable. X hat dann eine Dichtefunktion (oder kurz „Dichte“) $f(\geq 0)$, wenn gilt

$$F(x) = P(X \leq x) = \int_{-\infty}^{+\infty} f(t)dt, \ -\infty < x < +\infty \tag{46}$$

....

8.1 Die Rechteck- oder Gleichverteilung

X heißt gleichverteilt oder Rechteck-verteilt auf $[a, b]$. Die Schreibweise hierfür sieht aus wie folgt

$$X \ Suett - R(a, b)$$

Soviel zu Rechteck- und Gleich-verteilung.

8.2 Die Normalverteilung

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp(-\frac{1}{2} \cdot \frac{(t-\mu)^2}{\sigma^2}), \ -\infty < t < +\infty (\mu \in \mathbb{R}, \sigma^2 > 0) \tag{47}$$

Die Dichte sieht aus wie folgt

$$F(x) = P(X \leq x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp(-\frac{1}{2} \cdot \frac{(t-\mu)^2}{\sigma^2})dt \tag{48}$$

Für

- $s := \frac{t-\mu}{\sigma}$
- $t = \mu + \sigma s$
- $dt = \sigma ds$

ist (oben)

$$\begin{aligned} () &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot exp(-\frac{1}{2} \cdot s^2)dt \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \\ \Phi(x) &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} exp(-\frac{1}{2} \cdot s^2)dt, \ -\infty < x < +\infty \end{aligned} \tag{49}$$

X heißt Normal-verteilt mit dem Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Schreibweise:

$$X \ N(\mu, \sigma^2) \tag{50}$$

8.3 Die Exponential-Verteilung

$$f(t) = \begin{cases} 0 & , \ t < 0 \\ \lambda \cdot exp(-\lambda t) & , \ t \geq 0 \end{cases} \tag{51}$$

Hierbei sei $\lambda > 0$.

Dann

$$F(x) = P(X \leq x) = \begin{cases} 0 & , \ x < 0 \\ 1 - exp(-\lambda t) & , \ x \geq 0 \end{cases} \tag{52}$$

X heißt Exponential-verteilt mit dem Parameter $\lambda > 0$. Schreibweise:

$$X \ Exp(\lambda) \tag{53}$$

$X : \Omega \rightarrow \mathbb{R}^2$ sei ein 2-dimensoinaler Zufalsvektor, $X = (X_1, X_2)$. X habe die Dichte $f(\geq 0)$, d.h.

$$\begin{aligned} P(X_1 \leq x_1, X_2 \leq x_2) &= \int_{-\infty}^{x_2} \left(\int_{-\infty}^{x_1} f(t_1, t_2) dt_2 \right) dt_1 \\ &= \int_{-\infty}^{x_1} \left(\int_{-\infty}^{x_2} f(t_1, t_2) dt_1 \right) dt_2 \end{aligned} \tag{54}$$

Dann ist

$$\begin{aligned} P(X_1 \leq x_1) &= \int_{-\infty}^{x_1} \left(\int_{-\infty}^{+\infty} f(t_1, t_2) dt_2 \right) dt_1 \\ &= \int_{-\infty}^{x_1} f_{X_1}(t_1) dt_1 \end{aligned} \tag{55}$$

Also ist

$$f_{X_1}(t_1) = \int_{-\infty}^{+\infty} f(t_1, t_2) dt_2, \ -\infty < t_1 < +\infty \tag{56}$$

Dichte von X_1

Analog errechnet sich die Dichte von X_2

Teil IV
Bedingte Wahrscheinlichkeiten

9 Bedingte Wahrscheinlichkeiten

$(\Omega, \mathfrak{A}, P)$ ist W-Raum. Zunächst $|\Omega| < \infty$, $\mathfrak{A} = \mathfrak{P}(\Omega)$, $P(A) = \frac{|A|}{|\Omega|}$, $A \subset \Omega$, $A \subset \Omega, B \subset \Omega, |B| \geq 1$. Dann

$$\frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{P(A \cap B)}{P(B)} \tag{57}$$

„Wahrscheinlichkeit für das Eintreten von A unter Bedingung, dass B eintritt", lese als „Wahrscheinlichket von A unter B"

9.1 Definition

$(\Omega, \mathfrak{A}, P)$ W-Raum (bel), $A, b \in \mathfrak{A}, P(B) > 0$. Dann heißt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{58}$$

die bedingte Wahrscheinlichkeit von A unter B („A gegeben B").

9.2 Satz

$(\Omega, \mathfrak{A}, P)$ W-Raum, $\emptyset \neq I$ abzählbare Indexmenge. Wir unterstellen $A_i \in \mathfrak{A}, i \in I$ seien paarweise disjunkt, $P(A_i) > 0$ für jede $i \in I$. Dann halten wir fest

1. Für $A \in \mathfrak{A}$ mit der Eigenschaft $A \in \sum_{i \in I} A_1$ ist $P(A) = \sum_{i \in I} P(A|A_i)P(A_i)$ (bekannt als „Satz von der totalen Wahrscheinlichkeit")
2. Sei $A \in \mathfrak{A}$ mit $P(A) > 0, i_0 \in I$. Dann

$$P(A_{i_0}|A) = \frac{P(A|A_{i_0})P(A_{i_0})}{\sum_{i \in I} P(A|A_i)P(A_i)} \tag{59}$$

   (bekannt als „Formel von Bayes")
3. Sei $I = \{0, 1, \dots, n\}$, $(n \in \mathbb{N}\text{fest})$. Dann gilt

$$P(A_0 \cap A_1 \cap \ldots \cap A_n) = P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_n|A_0 \cap \ldots \cap A_{n-1}) \tag{60}$$

   (bekannt als „Multiplikationsformel")

9.3 Beweis

1. Aus
$$A = \sum_{i \in I} (A \cap A_i)$$
folge sofort
$$P(A) \sum_{i \in I} P(A \cap A_i) = \sum_{i \in I} P(A|A_i) P(A_i)$$

2. folgt sofort aus 1.

3.
$$\begin{aligned} &P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_n|A_0 \cap \ldots \cap A_{n-1}) \\ =&\cancel{P(A_0)} \cdot \frac{\cancel{P(A_1 \cap A_0)}}{\cancel{P(A_0)}} \cdot \frac{P(A_0 \cap A_1 \cap A_2)}{\cancel{P(A_0 \cap A_1)}} \cdots \\ =&PP(A_0 \cap A_1 \cap \ldots \cap A_n) \end{aligned} \tag{61}$$

9.4 Beispiel: Lostrommel

Eine Lostrommel enthält a Gewinnlose und b Nieten, $A \geq 1, b \geq 2$. Es gibt 2 Spielmöglichkeiten:

1. Ein Los wird gezogen. Das gezogene Los ist ein Gewinnlos oder eine Niete. Das Spiel ist beendet.

2. Ein Los wird gezogen und unbesehen weggeworfen. Daraufhin entfernt der Losverkäufer eine Niete aus der Trommel. Ein weiteres Los wird gezogen. Dieses Los ist ein Gewinnlos oder eine Niete.

Sei A das Ereignis, dass bei der ersten Ziehung ein Gewinnlos gezogen wird.
Sei B das Ereignis, dass bei der zweiten Ziehung ein Gewinnlos gezogen wird.
Dann ergibt sich bei diesen Fällen

1.
$$P(A) = \frac{a}{a+b} \tag{62}$$

2.
$$\begin{aligned} P(B) &= P(B|A) \cdot P(A) + P(B|A^c)P(A^c) \\ &= \frac{a-1}{a+b-2} \cdot \frac{a}{a+b} + \frac{a}{a+b-2} \cdot \frac{b}{a+b} \\ &= \frac{a}{a+b} \cdot \left(\frac{a+b-1}{a+b-2}\right) > \frac{a}{a+b} > P(A) \end{aligned} \tag{63}$$

Sei nun $a = 1,\ b = 2$. Dann folgt sofort

$$\begin{aligned} P(A) &= \frac{1}{3} \\ P(B) &= \frac{1}{3} \cdot 2 = \frac{2}{3} \end{aligned}$$

Sei $A, B \in \mathfrak{A}, P(B) > 0$. Dann gilt offenbar

$$P(A|B) = P(A) \iff P(A \cap B) = P(A)P(B) \tag{64}$$

10 Stochastische Unabhängigkeit von Ereignissen und Zufallsvariablen

10.1 Definition

$(\Omega, \mathfrak{A}, P)$ W-Raum, $A, B \in \mathfrak{A}$. A, B heißen (stochastisch) unabhängig, wenn

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

gilt.

10.2 Einsicht / Beweis

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B), \text{ falls } A, B \text{ unabhängig} \\ &(1 - P(B))P(A) \\ &= P(A)P(B^c) \end{aligned} \tag{65}$$

10.3 Definition

- $\emptyset \neq I$ endliche Indexmenge, $A_i \in \mathfrak{A}, i \in I$ Ereignisse.
  A_i heißen (stochastisch) unabhängig, wenn für jede $\emptyset \neq J \subset I$ gilt

$$P\left(\bigcap_{j \in J} A_j\right) = \Pi_{j \in J} P(A_j) \tag{66}$$

- $\emptyset \neq I$ beliebige Indexmenge, $A_i \in \mathfrak{A}, i \in I$ Ereignisse.
  $\mathfrak{A}_i, i \in I$ heißen unabhängig, wenn für jede Teilmenge $\emptyset \neq J \subset I,\ J$endlich die Ereignisse $A_j, j \in J$ unabhängig sind.

10.3.1 Bemerkung

$A_i \in \mathfrak{A}, i \in I$ seien Ereignisse. Dann gilt

$$A_i, i \in I \text{ unabhängig } \iff B_i, i \in I \text{ unabhängig }, \; B_i \in \{A_i, A_i^c\} \text{ beliebig} \tag{67}$$

10.4 Definition

$(\Omega, \mathfrak{A}, P)$ W-Raum, $\emptyset \neq I$ Indexmenge, $(R_i, \mathfrak{S})$ Messräume, $i \in I$, $X_i : \Omega \to R_i$ Zufallsvariable, $i \in I$.
$X_i, i \in I$ heißen (stochastisch) unabhängig, wenn für jede Wahl von $B_i \in \mathfrak{S}_i, i \in I$ die Ereignisse $A_i = X_i^{-1}(B_i), i \in I$ unabhängig sind.

10.5 Spezialfall

$I = \{1, \ldots, n\}$.
$X_i, i \in I$ sind unabhängig $\iff P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n), \forall B_i \in \mathfrak{S}_i, i \in I$

1. $X_1, \ldots, X_n$ haben eine **diskrete Verteilung**. Dann

$$X_1, \ldots, X_n \iff P(X_1 = x, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

für jede $x_i \in R_i, i \in I$