Nachname:	Vorname:	Matrikelnummer:

Formelsammlung zur Vorlesung

Statistik II für Studierende der Soziologie und Nebenfachstudierende

Prof. Dr. Thomas Augustin, Johanna Brandt, Julia Plaß
SoSe 2014

Handschriftliche Kommentare im Original sind nur auf den bedruckten Seiten erlaubt!

1 Wahrscheinlichkeitsrechnung

1.1 Mengen und elementare Mengenoperationen

Definition 1.1 Eine *Menge* ist eine Zusammenfassung verschiedener Objekte zu einem Ganzen. Die einzelnen Objekte einer Menge werden *Elemente* genannt.

Grundlegende Begriffe der Mengenlehre

- **Standardmengen:**

$\mathbb{N} = \{1, 2, 3, \ldots\}$	Menge der natürlichen Zahlen
$\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$	Menge der natürlichen Zahlen inklusive 0
$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$	Menge der ganzen Zahlen
$\mathbb{R} = (-\infty, \infty)$	Menge der reellen Zahlen
$\emptyset$	leere Menge

- **Elementeigenschaft:**
  x ist Element der Menge A: $x \in A$
  x ist nicht Element der Menge A: $x \notin A$

- **Teilmengen:** A ist Teilmenge von B, in Zeichen $A \subset B$, wenn jedes Element von A auch in B ist.

- **Schnittmenge:** Die Schnittmenge $A \cap B$ ist die Menge aller Elemente, die sowohl in A als auch in B enthalten sind:

$$A \cap B = \{x | x \in A \text{ und } x \in B\}$$

Eigenschaften:

  - Gilt $A \subset B$, so ist $A \cap B = A$.
  - Für jede Menge A gilt: $A \cap A = A$ und $A \cap \emptyset = \emptyset$.
  - Zwei Mengen A und B mit $A \cap B = \emptyset$, d.h. zwei Mengen, die kein gemeinsames Element haben, heißen *disjunkt*.
  - Verallgemeinerung: Die Schnittmenge aus n Mengen $A_1, \ldots, A_n$ enthält alle Elemente, die in jeder der Mengen $A_1, \ldots, A_n$ enthalten sind und wird bezeichnet mit

$$\bigcap_{i=1}^{n} A_i := A_1 \cap A_2 \cap \ldots \cap A_n.$$

- **Vereinigungsmenge:** Die Vereinigungsmenge $A \cup B$ ist die Menge aller Elemente, die in A oder B enthalten sind:

$$A \cup B = \{x | x \in A \text{ oder } x \in B\}$$

Verallgemeinerung: Die Vereinigungsmenge aus n Mengen $M_1, \ldots, M_n$ enthält alle Elemente, die in mindestens einer der Mengen $M_1, \ldots, M_n$ enthalten sind und wird bezeichnet mit

$$\bigcup_{i=1}^{n} M_i := M_1 \cup M_2 \cup \ldots \cup M_n$$

- **Differenzmenge:** Die Differenzmenge $A \setminus B$ ist die Menge aller Elemente, die in A, aber nicht in B enthalten sind:

$$A \setminus B = \{x | x \in A \text{ aber } x \notin B\}$$

- **Komplementärmenge:** Die Komplementärmenge $\overline{A} = A^C$ bezüglich einer Grundmenge Ω ist die Menge aller Elemente von Ω, die nicht in A sind:

$$\overline{A} = A^C = \{x \in \Omega | x \notin A\} = \{x : x \notin A\}$$

- **Potenzmenge:** Die Potenzmenge $\mathcal{P}(A)$ ist die Menge aller Teilmengen von A:

$$\mathcal{P}(A) = \{M | M \subset A\}.$$

- **Mächtigkeit:** Die Mächtigkeit $|A|$ einer Menge A ist die Anzahl der Elemente von A

Rechenregeln für Mengen

a) Kommutativgesetze (Vertauschung):

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

b) Assoziativgesetze (Zusammenfassen):

$$(A \cap B) \cap C = A \cap (B \cap C).$$
$$(A \cup B) \cup C = A \cup (B \cup C).$$

c) Distributivgesetze (Ausklammern/Ausmultiplizieren):

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

d) De Morgansche Regeln:

$$(\overline{A \cup B}) = \overline{A} \cap \overline{B}$$
$$(\overline{A \cap B}) = \overline{A} \cup \overline{B}$$

e) Aus $A \subset B$ folgt $\overline{B} \subset \overline{A}$.

f) Für die Differenzmenge gilt $A \setminus B = A \cap \overline{B}$.

g) Für die Potenzmenge gilt $|\mathcal{P}(A)| = 2^{|A|}$.

Das kartesische Produkt

Das *kartesische Produkt* zweier Mengen

$$A = \{a_1, a_2, a_3, \ldots, a_k\}$$
$$B = \{b_1, b_2, b_3, \ldots, b_m\}$$

ist die Menge

$$A \times B := \Big\{(a_i, b_j) \;\Big|\; i = 1, \ldots, k, j = 1, \ldots, m\Big\}$$

Sie besteht also aus allen möglichen Kombinationen der Elemente von A und B:

$$\begin{aligned} A \times B = \big\{&(a_1, b_1), (a_1, b_2), (a_1, b_3), \ldots, (a_1, b_m), \\ &(a_2, b_1), (a_2, b_2), (a_2, b_3), \ldots, (a_2, b_m), \\ &\vdots \\ &(a_k, b_1), (a_k, b_2), (a_k, b_3), \ldots, (a_k, b_m)\big\} \end{aligned}$$

Verallgemeinerungen:

- Das kartesische Produkt der Mengen $\Omega_1, \Omega_2, \ldots, \Omega_n$ wird mit

$$\bigtimes_{i=1}^{n} \Omega_i = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$$

  bezeichnet und besteht aus allen möglichen n-Tupeln, die sich (unter Beachtung der Reihenfolge) aus Elementen aus $\Omega_1, \Omega_2, \ldots, \Omega_n$ bilden lassen.

- Die Mengen $\Omega_1, \Omega_2, \ldots, \Omega_n$ müssen nicht endlich sein; für endliche Mengen gilt

$$\left| \bigtimes_{i=1}^{n} \Omega \right| = |\Omega_1| \cdot |\Omega_2| \cdot \ldots \cdot |\Omega_n|$$

- Kartesische Produkte werden verwendet, um Ergebnisse komplexer Experimente aus Einzelexperimenten zusammenzusetzen.

1.2 Wahrscheinlichkeit – Ein komplexer Begriff und seine Formalisierung

1.2.1 Zufallsvorgänge

1.2.2 Laplace-Wahrscheinlichkeiten und Urnenmodelle

Abzählregel

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}}$$

Laplace-Wahrscheinlichkeit

In einem Laplace-Experiment gilt für $P(A)$ mit $|A| = M$ und $|\Omega| = N$:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{M}{N}.$$

Urnenmodell

- Grundgesamtheit: Urne mit N nummerierten Kugeln
- Stichprobe: Zufälliges Ziehen von n Kugeln aus der Urne

Ziehen mit Zurücklegen mit Berücksichtigung der Reihenfolge

- Ziehe Stichprobe vom Umfang n **mit** Zurücklegen.
- $\Omega = \{(\omega_1, \ldots, \omega_n) | \omega_j \in \{1, \ldots, N\}\}$
  (Das selbe Element kann mehrfach vorkommen.)
- Anzahl möglicher Stichproben:

$$|\Omega| = \underbrace{N \cdot N \cdot \ldots \cdot N}_{n\text{-Faktoren}} = N^n$$

Ziehen ohne Zurücklegen mit Berücksichtigung der Reihenfolge

- Ziehe Stichprobe vom Umfang n **ohne** Zurücklegen.
- $\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_j \in \{1, \ldots, N\},\, \omega_j \neq \omega_i \text{ für } i \neq j\}$
  (Jedes Element kann nur einmal vorkommen.)
- Anzahl möglicher Stichproben:

$$|\Omega| = N \cdot (N-1) \cdot \ldots \cdot N - n + 1 = \frac{N!}{(N-n)!}$$

Wiederholung Fakultät: Die *Fakultät* einer natürlichen Zahl k ist definiert als

$$k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 2 \cdot 1.$$

Es gilt:

$$1! = 1, \quad 0! = 1.$$

Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge

- Ziehe Stichprobe vom Umfang n **ohne** Zurücklegen.
- $\Omega = \{\{\omega_1, \ldots, \omega_n\} \,:\, \omega_j \in \{1, \ldots, N\},\, \omega_j \neq \omega_i \text{ für } j \neq i\}$
- Anzahl der möglichen Stichproben:

$$|\Omega| = \frac{N!}{(N-n)!n!} = \binom{N}{n}$$

Wiederholung Binomialkoeffizient: Der *Binomialkoeffizient* $\binom{N}{n}$ ist definiert als

$$\binom{N}{n} = \frac{N!}{(N-n)! \cdot n!}\,.$$

Es gilt:

$$\binom{N}{0} = 1, \binom{N}{1} = N, \binom{N}{N} = 1\,, \quad \binom{N}{n} = 0, \text{ falls } N < n\,.$$

1.2.3 Die „induktive Brücke“ I

1.2.4 Das Axiomensystem von Kolmogoroff (1933) und wichtige Rechenregeln

Definition 1.2 Eine Funktion P (P steht für Probability), die Ereignissen aus Ω reelle Zahlen zuordnet, heißt *Wahrscheinlichkeit*, wenn gilt

(K1) $P(A) \geq 0$ für alle Ereignisse $A \subset \Omega$.

(K2) $P(\Omega) = 1$.

(K3) Falls $A \cap B = \emptyset$, dann gilt $P(A \cup B) = P(A) + P(B)$

Rechenregeln für Wahrscheinlichkeiten

- $P(\overline{A}) = 1 - P(A)$
- Für beliebige Mengen A, B gilt: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Falls $A_1, A_2, \ldots, A_n$ paarweise disjunkt sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

Vollständige Zerlegung: Ist $A_1, A_2, \ldots, A_k$ eine *vollständige Zerlegung* von Ω, d.h. gilt

$$\bigcup_{i=1}^{k} A_i = \Omega \text{ und } A_i \cap A_j = \emptyset \text{ für } i \neq j,$$

so gilt für jedes Ereignis B:

$$P(B) = \sum_{i=1}^{k} P(B \cap A_i).$$

1.2.5 Grundlegendes zum Begriff „Wahrscheinlichkeit“

1.3 Stochastische Unabhängigkeit und bedingte Wahrscheinlichkeiten

1.3.1 Stochastische Unabhängigkeit

Definition 1.3 Zwei Ereignisse A und B heißen *(stochastisch) unabhängig*, wenn gilt

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

andernfalls heißen sie *stochastisch abhängig*.

Definition 1.4 Ereignisse $A_1, A_2, \ldots, A_n$ heißen (vollständig) stochastisch unabhängig, wenn für alle $I \subseteq \{1, \ldots, n\}$ gilt

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

Achtung: Aus der paarweisen Unabhängigkeit

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{für alle } i, j$$

folgt nicht die vollständige Unabhängigkeit.

1.3.2 Bedingte Wahrscheinlichkeiten

Definition 1.5 Gegeben seien zwei Ereignisse A und B mit $P(B) > 0$. Dann heißt:

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

bedingte Wahrscheinlichkeit von A gegeben B oder *bedingte Wahrscheinlichkeit von A unter der Bedingung B.*

1.3.3 Koppelung unabhängiger Experimente (unabhängige Wiederholungen)

Unabhängige Koppelung mehrerer Experimente: Gegeben sei eine Menge von Zufallsexperimenten, beschrieben durch die Ergebnisräume Ω_i, $i = 1, \ldots, n$ und die Wahrscheinlichkeitsbewertungen P_i, $i = 1, \ldots, n$. Fasst man die Experimente zusammen, so ergibt sich der Ergebnisraum

$$\Omega := \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$$

mit den Elementen

$$\omega := (\omega_1, \omega_2, \ldots, \omega_n).$$

Sind die Experimente unabhängig (Dies ist inhaltlich zu entscheiden!), so setzt man für beliebige $A_i \subset \Omega_i, i = 1, \ldots, n$,

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P_1(A_1) \cdot P_2(A_2) \cdot \ldots \cdot P_n(A_n).$$

Dies beschreibt ein Wahrscheinlichkeitsmaß auf Ω, bei dem – per Konstruktion – beliebige Ereignisse aus den einzelnen Ω_i voneinander unabhängig sind.

1.3.4 Koppelung abhängiger Experimente

Satz 1.6 (Satz von der totalen Wahrscheinlichkeit) Gegeben sei eine vollständige Zerlegung $A_1, A_2, \ldots, A_k$ von Ω. Dann gilt für jedes Ereignis B

$$P(B) = \sum_{j=1}^{k} P(B|A_j) \cdot P(A_j) = \sum_{j=1}^{k} P(B \cap A_j).$$

Koppelung abhängiger Experimente: Gegeben seien n Experimente, beschrieben durch die Grundräume $\Omega_i = \{a_{i1}, \ldots, a_{ik_i}\}$ und die Wahrscheinlichkeitsbewertungen $P_i, i = 1, \ldots, n$. Bezeichnet man für beliebiges $i = 1, \ldots, n$ und $j = 1, \ldots, k_i$, mit A_{ij} jeweils das zu a_{ij} gehörige Elementarereignis (also das Ereignis „a_{ij} tritt ein“), so gilt:

$$\begin{aligned} P\left(A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{nj_n}\right) =& P_1(A_{1j_1}) \cdot P_2\left(A_{2j_2}|A_{1j_1}\right) \cdot P_3\left(A_{3j_3}|A_{1j_1} \cap A_{2j_2}\right) \\ & \cdot \ldots \cdot P_n\left(A_{nj_n}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{n-1j_{n-1}}\right) \end{aligned}$$

Häufig werden die Indizes bei P weggelassen.

Korollar 1.7 Sei $A_1, A_2, \ldots, A_k$ eine vollständige Zerlegung. Dann gilt für beliebige Ereignisse B und C mit $P(C) > 0$

$$P(B|C) = \sum_{j=1}^{k} P(B|(A_j \cap C)) \cdot P(A_j|C)$$

Markovmodelle

Hier interpretiert man den Laufindex als Zeit. Gilt in der Koppelung abhängiger Experiment $\Omega_1 = \Omega_2 = \ldots = \Omega_n = \{a_1, \ldots, a_k\}$ und sind alle bedingten Wahrscheinlichkeiten nur vom jeweils unmittelbar vorhergehenden Zeitpunkt abhängig, d.h. gilt

$$P(A_{i+1,j_{i+1}}|A_{1j_1} \cap A_{2j_2} \cap \ldots \cap A_{ij_i}) = P(A_{i+1,j_{i+1}}|A_{ij_i}), \qquad (*)$$

so spricht man von einem *Markovmodell mit den Zuständen* $a_1, \ldots, a_k$*1*.
Sind die sogenannten *Übergangswahrscheinlichkeiten* in (*) unabhängig von der Zeit, gilt also $P(A_{i+1,j}|A_{il}) \equiv p_{jl}$ für alle i, j, l, so heißt das Markovmodell *homogen*.

1.3.5 Das Theorem von Bayes

Satz 1.8 Sei $A_1, \ldots A_k$ eine vollständige Zerlegung von Ω, wobei $P(A_i) > 0$, $P(B|A_i) > 0$, $i = 1, \ldots k$ und $P(B) > 0$ erfüllt seien. Dann gilt

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)}.$$

1.4 Zufallsvariablen und ihre Verteilung

1.4.1 Diskrete Zufallsvariablen

Definition 1.9 Gegeben seien ein diskreter, d.h. höchstens abzählbarer, Ergebnisraum Ω und die Wahrscheinlichkeit P auf Ω. Jede Abbildung

$$\begin{aligned} X : \Omega &\mapsto \Omega_X \\ \omega &\mapsto X(\omega) \end{aligned}$$

heißt Zufalls*element*. Setzt man für jede *Realisation* $x \in \Omega_X$

$$P_X(\{x\}) := P(\{X = x\}) := P(\{\omega | X(\omega) = x\})$$

so erhält man eine Wahrscheinlichkeit auf Ω_X. (Oft wird auch $P(X = x)$ statt $P(\{X = x\})$ geschrieben.)

Definition 1.10 Gegeben sei eine diskrete Zufallsvariable X mit der Wahrscheinlichkeitsverteilung P. Die Menge

$$\mathcal{X} := \{x \in \mathbb{R} | P(\{x\}) > 0\}$$

heißt *Träger von X*.

Definition 1.11 Die *Wahrscheinlichkeitsfunktion* $f(x)$ einer diskreten Zufallsvariable X ist für $x \in \mathbb{R}$ definiert durch

$$f(x) = \begin{cases} P(X = x_i) = p_i, & x = x_i \in \mathcal{X} = \{x_1, x_2, \ldots, x_k, \ldots\} \\ 0, & \text{sonst.} \end{cases}$$

1.4.2 Verteilungsfunktion

Definition 1.12 Sei X eine Zufallsvariable. Die Funktion

$$\begin{aligned} F : \mathbb{R} &\mapsto [0; 1] \\ x &\mapsto F(x) \end{aligned}$$

$$F(x) := P(X \leq x)$$

heißt *Verteilungsfunktion*.

Satz 1.13 Für beliebige reelle Zahlen a und b mit $a < b$ gilt: $P(a < X \leq b) = F(b) - F(a)$.

Satz 1.14 Unter Regularitätsbedingungen gilt: Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen X kann man durch die Verteilungsfunktion eineindeutig erklären.

1.4.3 Stetige Zufallsvariablen

Definition 1.15 Gegeben sei eine stetige Zufallsvariable X mit differenzierbarer Verteilungsfunktion $F(x)$. Dann heißt die Ableitung von $F(x)$ nach x, also

$$f(x) = \frac{dF(x)}{dx},$$

Dichte der Zufallsvariablen X.

Satz 1.16 In der Situation der obigen Definition gilt

$$F(x) = \int_{-\infty}^{x} f(u)\ du$$

und damit für beliebige reelle Zahlen a und b mit $a < b$

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) = P(a < X < b) \\ &= P(a \leq X < b) \\ &= \int_a^b f(x)\, dx. \end{aligned}$$

Jede Funktion f auf $\mathbb{R}$ mit $f(x) \geq 0$ für alle x und

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

kann als Dichte verwendet werden.

Alternative Definition: Eine Zufallsvariable X heißt stetig, wenn es eine Funktion $f(x) \geq 0$ gibt, so dass für jedes Intervall $[a, b]$

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

gilt.

1.4.4 Lebensdauern; Hazardrate und Survivorfunktion

Satz 1.17 Die Verteilung einer nicht negativen, stetigen Zufallsvariable X wird eineindeutig durch sowohl die *Überlebensfunktion* (Survivorfunktion)

$$S(x) := P(X \geq x),$$

als auch durch die *Hazardrate*

$$\lambda(x) := \lim_{h \to 0} \frac{P(x \leq X \leq x + h | X \geq x)}{h}$$

beschrieben.

Es gelten folgende Zusammenhänge:

$$\begin{aligned} S(x) &= 1 - F(x) \\ S(x) &= \exp\left(-\int_0^x \lambda(u)du\right) \\ F(x) &= 1 - \exp\left(-\int_0^x \lambda(u)du\right) \\ f(x) &= \lambda(x) \cdot S(x) \end{aligned}$$

1.4.5 Unabhängigkeit von Zufallsvariablen

Definition 1.18 Zwei Zufallsvariablen X und Y mit den Verteilungsfunktionen F_X und F_Y heißen *stochastisch unabhängig*, falls für alle x und y gilt

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\}) \cdot P(\{Y \leq y\}) = F_X(x) \cdot F_Y(y),$$

andernfalls heißen sie stochastisch abhängig.

1.5 Erwartungswert und Varianz

1.5.1 Diskrete Zufallsvariablen

Definition 1.19 Gegeben sei eine diskrete Zufallsvariable X mit Träger $\mathcal{X}$. Dann heißt

$$\mathrm{E}\,X := \mathrm{E}(X) := \mathbb{E}(X) := \sum_{x \in \mathcal{X}} x \cdot P(X = x)$$

Erwartungswert von X,

$$\begin{aligned} \mathrm{Var}\,X := \mathrm{Var}(X) := \mathbb{V}(X) &:= \mathrm{E}((X - \mathrm{E}(X))^2) \\ &= \sum_{x \in \mathcal{X}} (x - \mathrm{E}(X))^2 \cdot P(X = x) \end{aligned}$$

Varianz von X und

$$\sigma_X := \sqrt{\mathrm{Var}(X)}$$

Standardabweichung von X.

Zur Berechnung der Varianz ist der sogenannte *Verschiebungssatz* sehr praktisch:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}\,X)^2$$

1.5.2 Stetige Zufallsvariablen

Definition 1.20 Sei X eine stetige Zufallsvariable mit Dichte $f(x)$. Dann heißt

$$\mathrm{E}\,X := \mathrm{E}(X) := \mathbb{E}(X) := \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

Erwartungswert von X,

$$\begin{aligned}\operatorname{Var} X := \operatorname{Var}(X) := \mathbb{V}(X) &:= \mathrm{E}((X - \mathrm{E}(X))^2 \\ &= \int_{-\infty}^{\infty} (x - \mathrm{E}(X))^2 \cdot f(x)\,dx\end{aligned}$$

Varianz von X und

$$\sigma_X := \sqrt{\operatorname{Var}(X)}$$

Standardabweichung von X.

1.5.3 Allgemeine Rechenregeln für Erwartungswert und Varianz

Satz 1.21 Seien X und Y diskrete oder stetige Zufallsvariablen (mit existierenden Erwartungswerten und Varianzen). Dann gilt:

a) $\mathrm{E}(aX + bY) = a \cdot \mathrm{E}(X) + b \cdot \mathrm{E}(Y)$ und insbesondere auch

$$\begin{aligned}\mathrm{E}(a) &= a, \\ \mathrm{E}(aX) &= a \cdot \mathrm{E}(X) \\ \mathrm{E}(X + Y) &= \mathrm{E}(X) + \mathrm{E}(Y)\end{aligned}$$

b) $\operatorname{Var}(aX + b) = a^2 \cdot \operatorname{Var}$.

Sind X und Y zusätzlich unabhängig, so gilt

$$\begin{aligned}\mathrm{E}(X \cdot Y) &= \mathrm{E}(X) \cdot \mathrm{E}(Y) \\ \operatorname{Var}(X + Y) &= \operatorname{Var}(X) + \operatorname{Var}(Y)\end{aligned}$$

Definition 1.22 Die Zufallsvariable

$$Z := \frac{X - \mathrm{E}(X)}{\sqrt{\operatorname{Var}(X)}}$$

heißt *standardisierte Zufallsvariable*. Es gilt

$$\mathrm{E}(Z) = 0 \;\text{ und }\; \operatorname{Var}(Z) = 1.$$

1.6 Wichtige Verteilungsmodelle

1.6.1 Binomialverteilung

Definition 1.23 Eine Zufallsvariable X heißt *binomialverteilt* mit dem Parametern π bei n Versuchen, kurz $X \sim B(n, \pi)$, wenn sie die Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \binom{n}{x} \pi^x (1-\pi)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Alternative Darstellung: Die Zufallsvariable X kann als

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

mit den binären Variablen

$$X_i = \begin{cases} 1 & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0 & \text{sonst.} \end{cases}$$

dargestellt werden. Die X_i seien stochastisch unabhängig mit $X_i \sim B(1, \pi)$. $B(1, \pi)$ heißt *Bernoulliverteilung*.

Erwartungswert und Varianz

- Erwartungswert der Binomialverteilung:

$$\mathrm{E}(X) = \mathrm{E}(X_1 + \ldots + X_n) = \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_n) = n\pi$$

- Varianz der Binomialverteilung:

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = n\pi(1-\pi)$$

Eigenschaften:

- Symmetrieeigenschaft:
  Sei $X \sim B(n, \pi)$ und $Y = n - X$. Dann gilt $Y \sim B(n, 1 - \pi)$.

- Summeneigenschaft:
  Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$. Sind X und Y unabhängig, so gilt

$$X + Y \sim B(n + m, \pi).$$

1.6.2 Poisson-Verteilung

Definition 1.24 Eine Zufallsvariable X mit der Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!}\,\mathrm{e}^{-\lambda}, & x = 0, 1, \ldots, n \\ 0, & \text{sonst} \end{cases}$$

heißt *Poisson-verteilt* mit Parameter (oder Rate) $\lambda > 0$, kurz $X \sim Po(\lambda)$.

Erwartungswert und Varianz

- Erwartungswert der Poissonverteilung:
$$\mathrm{E}(X) = \lambda$$
- Varianz der Poissonverteilung:
$$\mathrm{Var}(X) = \lambda$$

Poisson-Verteilung als Näherungsmodell für die Binomialverteilung:

$$X \sim B(n, \pi) \underset{\substack{n \text{ groß} \\ \pi \text{ klein}}}{\Longrightarrow} X \approxeq Po(n \cdot \pi)$$

Satz 1.25 (Addition von Poisson-verteilten Zufallsvariablen) Sind $X \sim Po(\lambda_X), Y \sim Po(\lambda_Y)$ voneinander unabhängig, so gilt

$$X + Y \sim Po(\lambda_X + \lambda_Y).$$

1.6.3 Normalverteilung

Definition 1.26 Eine stetige Zufallsvariable X heißt *normalverteilt* mit den Parametern μ und σ^2, kurz $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn für ihre Dichte gilt :

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \; x \in \mathbb{R}$$

und *standardnormalverteilt*, in Zeichen $X \sim \mathcal{N}(0; 1)$, falls $\mu = 0$ und $\sigma^2 = 1$ gilt (π ist hier die Kreiszahl $\pi = 3.14\ldots$).

Wichtige Formeln

- Die *Dichte der Standardnormalverteilung* wird oft mit $\varphi(x)$ bezeichnet, also
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$
und die zugehörige Verteilungsfunktion mit
$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)du$$

- $\Phi(x)$ lässt sich nicht in geschlossener Form durch bekannte Funktionen beschreiben $\Longrightarrow$ numerische Berechnung, Tabellierung.
- μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann
$$\mathrm{E}(X) = \mu \quad \text{und} \quad \mathrm{Var}(X) = \sigma^2.$$
- Die Dichte ist symmetrisch um μ, d.h.
$$f(\mu - x) = f(\mu + x)\,.$$
- Funktionswerte für negative Argumente::
$$\Phi(-x) = 1 - \Phi(x)$$
- Abgeschlossenheit gegenüber Linearkombinationen:

  Seien X_1 und X_2 unabhängig und $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2),\ i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt
$$\begin{aligned} Y_1 &:= a_1X_1 + b \sim \mathcal{N}(a_1\mu_1 + b; a_1^2\sigma_1^2) \\ Y_2 &:= a_1X_1 + a_2X_2 \sim \mathcal{N}(a_1\mu_1 + a_2\mu_2; a_1^2\sigma_1^2 + a_2^2\sigma_2^2). \end{aligned}$$

Standardisierung: Gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die zugehörige standardisierte Zufallsvariable
$$Z = \frac{X - \mu}{\sigma}$$
standardnormalverteilt.

1.7 Grenzwertsätze und Approximationen

1.7.1 Das i.i.d.-Modell

1.7.2 Das schwache Gesetz der großen Zahlen

Satz 1.27 (Theorem von Bernoulli) Seien $X_1, \ldots, X_n$, i.i.d. mit $X_i \in \{0, 1\}$ und $P(X_i = 1) = \pi$. Dann gilt für
$$H_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$
und jedes $\epsilon > 0$
$$\lim_{n\to\infty} P(|H_n - \pi| \le \epsilon) = 1$$

Schwaches Gesetz der großen Zahl: Gegeben seien $X_1, \ldots, X_n$ i.i.d. Zufallsvariablen mit (existierendem) Erwartungswert μ und (existierender) Varianz σ^2. Dann gilt für

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

und beliebiges $\epsilon > 0$:

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1$$

Schreibweise:

$$\bar{X}_n \xrightarrow{P} \mu$$

(„Stochastische Konvergenz", „X_n konvergiert in Wahrscheinlichkeit gegen μ".)

1.7.3 Der Hauptsatz der Statistik

Satz 1.28 Seien $X_1, \ldots, X_n$ i.i.d. mit Verteilungsfunktion F und sei $F_n(x)$ die empirische Verteilungsfunktion der ersten n Beobachtungen. Mit

$$D_n := \sup_x |F_n(x) - F(x)|,$$

gilt für jedes $c > 0$

$$\lim_{n \to \infty} P(D_n > c) = 0.$$

1.7.4 Der zentrale Grenzwertsatz

Satz 1.29 Seien $X_1, \ldots, X_n$ i.i.d. mit $\mathrm{E}(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2 > 0$ sowie

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right).$$

Dann gilt: Z_n ist *asymptotisch standardnormalverteilt*, in Zeichen: $Z_n \overset{a}{\sim} \mathcal{N}(0;1)$, d.h. es gilt für jedes z

$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z).$$

Wichtige Anwendung: Approximation der Binomialverteilung

Sei $X \sim B(n, \pi)$. Falls n groß genug ist, gilt:

$$P(X \leq x) \;\approx\; \Phi\left(\frac{x - n \cdot \pi}{\sqrt{n \cdot \pi(1 - \pi)}} \right).$$

Stetigkeitskorrektur:

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right)$$

$$P(X = x) \approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right) - \Phi\left(\frac{x - 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right)$$

Es gibt verschiedene Faustregeln, ab wann diese Approximation gut ist, z.B.

$$n \cdot \pi \geq 5 \text{ und } n \cdot (1-\pi) \geq 5$$
$$n \cdot \pi(1-\pi) \geq 9$$

1.8 Mehrdimensionale Zufallsvariablen

Definition 1.30 Betrachtet werden zwei eindimensionale diskrete Zufallselemente X und Y (zu demselben Zufallsexperiment). Die Wahrscheinlichkeit

$$P(X = x_i, Y = y_j) := P(\{X = x_i\} \cap \{Y = y_j\})$$

in Abhängigkeit von x_i und y_j ($i = 1, \ldots, k$, $j = 1, \ldots, m$) heißt *gemeinsame Verteilung* der mehrdimensionalen Zufallsvariable $\binom{X}{Y}$ bzw. der Variablen X und Y ($i = 1, ..., k$, $j = 1, ..., m$).

Randwahrscheinlichkeiten:

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j)$$
$$p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^{k} P(X = x_i, Y = y_j)$$

Bedingte Verteilungen:

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$
$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

Definition 1.31 Seien X und Y zwei Zufallsvariablen. Dann heißt

$$\sigma_{X,Y} := \operatorname{Cov}(X, Y) = \operatorname{E}\big((X - \operatorname{E}(X))(Y - \operatorname{E}(Y))\big)$$

Kovarianz von X und Y.

Rechenregeln:

a) $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$

b) $\operatorname{Cov}(X, Y) = \operatorname{E}(XY) - \operatorname{E}(X) \cdot \operatorname{E}(Y)$ (Verschiebungssatz)

c) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$

d) Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist

$$\operatorname{Cov}(\tilde{X}, \tilde{Y}) = a_X \cdot a_Y \cdot \operatorname{Cov}(X, Y)$$

e) $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \cdot \operatorname{Cov}(X, Y)$

Definition 1.32 Zwei Zufallsvariablen X und Y mit $\mathrm{Cov}(X, Y) = 0$ heißen *unkorreliert*.

Satz 1.33 Stochastisch unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung gilt jedoch im Allgemeinen nicht.

Definition 1.34 Gegeben seien zwei Zufallsvariablen X und Y. Dann heißt

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

Korrelationskoeffizient von X und Y.

Eigenschaften des Korrelationskoeffizienten

- Mit $\tilde{X} = a_X X + b_X$ und $\tilde{Y} = a_Y Y + b_Y$ ist
$$|\rho(\tilde{X}, \tilde{Y})| = |\rho(X, Y)|.$$
- $-1 \leq \rho(X, Y) \leq 1$.
- $|\rho(X, Y)| = 1 \Longleftrightarrow Y = aX + b$
- Sind $\mathrm{Var}(X) > 0$ und $\mathrm{Var}(Y) > 0$, so gilt $\rho(X, Y) = 0$ genau dann, wenn $\mathrm{Cov}(X, Y) = 0$.