

# Stochastik für InformatikerInnen

Wintersemester 2020/21

Wolfgang Kössler

Humboldt-Universität zu Berlin, Institut für Informatik

25. Februar 2021

# Inhalt (1)

- 1 Grundbegriffe
- 2 Kombinatorik
- 3 Bedingte Wahrscheinlichkeit, Unabhängigkeit
- 4 Klassische Wahrscheinlichkeitsräume
- 5 Zufallsvariablen (allgemein)

# Inhalt (2)

- 6 Diskrete zufällige Variablen
- 7 Charakteristika von Verteilungsfunktionen
- 8 Die Exponentialverteilung
- 9 Die Normalverteilung
- 10 Transformation von Zufallsvariablen

# Inhalt (3)

11 Zufallsvektoren

12 Korrelation

13 Ungleichungen

14 Grenzwertsätze

15 Schätzmethoden

# Inhalt (4)

16 Grundlagen der Simulation

17 Markov'sche Ketten

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit

# Geschichte (1)

- antikes Griechenland

Begriff der Wahrscheinlichkeit

Naturgesetze drücken sich durch eine Vielzahl von zufälligen Erscheinungen aus.

# Geschichte (1)

- antikes Griechenland

Begriff der Wahrscheinlichkeit

Naturgesetze drücken sich durch eine Vielzahl von zufälligen Erscheinungen aus.

- 1654, Chevalier de Méré, Pascal

Würfelspiele, Würfe mit 2 Würfeln. Wenn in 25 Würfen einmal eine Doppelsechs so hat C.d.M. gewonnen, sonst sein Gegner.



# Geschichte (2)



# Geschichte (3)

Pascal, Fermat (Briefwechsel)

2 Personen-Spiele. Gespielt wird eine Serie von Partien, z.B. Schach (nur 0,1). Gewinnen soll der Spieler, der zuerst  $S$  Partien gewonnen hat, d.h. dieser Spieler erhält den vollen Einsatz.

# Geschichte (3)

Pascal, Fermat (Briefwechsel)

2 Personen-Spiele. Gespielt wird eine Serie von Partien, z.B. Schach (nur 0,1). Gewinnen soll der Spieler, der zuerst  $S$  Partien gewonnen hat, d.h. dieser Spieler erhält den vollen Einsatz. **Abbruch** des Spiels (z.B. wegen Zeitmangel)

A hat  $a$  Gewinnpartien,  $a < S$

B hat  $b$  Gewinnpartien,  $b < S$

Wie ist der Einsatz gerecht zu verteilen?

# Geschichte (3)

Pascal, Fermat (Briefwechsel)

2 Personen-Spiele. Gespielt wird eine Serie von Partien, z.B. Schach (nur 0,1). Gewinnen soll der Spieler, der zuerst  $S$  Partien gewonnen hat, d.h. dieser Spieler erhält den vollen Einsatz. **Abbruch** des Spiels (z.B. wegen Zeitmangel)

A hat  $a$  Gewinnpartien,  $a < S$

B hat  $b$  Gewinnpartien,  $b < S$

Wie ist der Einsatz gerecht zu verteilen?

Variante:  $\frac{a}{b}$ , aber  $S$  wird nicht berücksichtigt!

# Geschichte (3)

Pascal, Fermat (Briefwechsel)

2 Personen-Spiele. Gespielt wird eine Serie von Partien, z.B. Schach (nur 0,1). Gewinnen soll der Spieler, der zuerst  $S$  Partien gewonnen hat, d.h. dieser Spieler erhält den vollen Einsatz. **Abbruch** des Spiels (z.B. wegen Zeitmangel)

A hat  $a$  Gewinnpartien,  $a < S$

B hat  $b$  Gewinnpartien,  $b < S$

Wie ist der Einsatz gerecht zu verteilen?

Variante:  $\frac{a}{b}$ , aber  $S$  wird nicht berücksichtigt!

Es wäre also der weitere mögliche Verlauf nach dem Abbruch zu analysieren.

# Geschichte (4)

- 1662, Graunt; 1693 Halley

Sterlichkeitstafeln (Überlebenswkt. in Abhängigkeit vom Lebensalter) → Rentenberechnung, Schiffsversicherung

# Geschichte (4)

- 1662, Graunt; 1693 Halley

Sterlichkeitstafeln (Überlebenswkt. in Abhängigkeit vom Lebensalter)  $\rightarrow$  Rentenberechnung, Schiffsversicherung

- 1713, Jacob Bernoulli

“Ars conjectandi”: 1. Lehrbuch der Wahrscheinlichkeitsrechnung

Bernoulli-Gesetz der Großen Zahlen,  $p = P(A)$

$h_n(A) = \frac{1}{n}$  Häufigkeit des Auftreten von  $A$ ,  $h_n(A) - p \rightarrow_{n \rightarrow \infty} 0$

# Geschichte (4)

- 1662, Graunt; 1693 Halley

Sterlichkeitstafeln (Überlebenswkt. in Abhängigkeit vom Lebensalter)  $\rightarrow$  Rentenberechnung, Schiffsversicherung

- 1713, Jacob Bernoulli

“Ars conjectandi”: 1. Lehrbuch der Wahrscheinlichkeitsrechnung

Bernoulli-Gesetz der Großen Zahlen,  $p = P(A)$

$h_n(A) = \frac{1}{n}$  Häufigkeit des Auftreten von  $A$ ,  $h_n(A) - p \rightarrow_{n \rightarrow \infty} 0$

- 1733, Moivre  $\left( \bar{X} = h_n(A), \mu = p, \sigma^2 = np(1 - p) \right)$

Grenzwertsatz von Moivre-Laplace

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma} \rightarrow \mathcal{N}(0, 1)$$



JACOBI BERNOULLI,  
Profess. Basil. & utriusque Societ. Reg. Scientiar.  
Gall. & Pruss. Sodal.  
MATHEMATICI CELEBERRIMI,  
**ARS CONJECTANDI,**  
OPUS POSTHUMUM.

*Accedit*  
TRACTATUS  
DE SERIEBUS INFINITIS,

Et EPISTOLA Gallicè scripta  
DE LUDO PILÆ  
RETICULARIS.



BASILEÆ,  
Impensis THURNISIORUM, Fratrum.  
clb lccc xiii.

# Geschichte (6)

- 1812, Laplace

klassische Definition der Wahrscheinlichkeit

$$P(A) = \frac{\text{\#für } A \text{ günstigen Elementarereignisse}}{\text{\#möglichen Elementarereignisse}}$$

# Geschichte (6)

- 1812, Laplace

klassische Definition der Wahrscheinlichkeit

$$P(A) = \frac{\text{\#für } A \text{ günstigen Elementarereignisse}}{\text{\#möglichen Elementarereignisse}}$$

- 1800, Laplace, Gauss

Untersuchung von Beobachtungsfehlern

Kleinste Quadrat-Schätzung

# Geschichte (6)

- 1812, Laplace

klassische Definition der Wahrscheinlichkeit

$$P(A) = \frac{\text{\#für } A \text{ günstigen Elementarereignisse}}{\text{\#möglichen Elementarereignisse}}$$

- 1800, Laplace, Gauss

Untersuchung von Beobachtungsfehlern

Kleinste Quadrat-Schätzung

- um 1800, Bessel

Annahme Normalverteilung ( $\bar{X}$  rechtfertigen)

# Geschichte (6)

- 1812, Laplace

klassische Definition der Wahrscheinlichkeit

$$P(A) = \frac{\text{\#für } A \text{ günstigen Elementarereignisse}}{\text{\#möglichen Elementarereignisse}}$$

- 1800, Laplace, Gauss

Untersuchung von Beobachtungsfehlern

Kleinste Quadrat-Schätzung

- um 1800, Bessel

Annahme Normalverteilung ( $\bar{X}$  rechtfertigen)

- Quetelet (1796-1874):

Normalverteilung sei allgemeingültig

# Geschichte (7)

- Ende 19. Jh., Tschebyschev, Markov, Ljapunov

# Geschichte (7)

- Ende 19. Jh., Tschebyshev, Markov, Ljapunov
- Ende 19. Jh., v. Bortkiewicz

Anzahl der tödlichen Unfälle bei Pferdetritten

# Geschichte (7)

- Ende 19. Jh., Tschebyschev, Markov, Ljapunov
- Ende 19. Jh., v. Bortkiewicz  
Anzahl der tödlichen Unfälle bei Pferdetritten
- Ende 19. Jh., Galton  
Begriffe Regression, Korrelation



# Geschichte (7)

- Ende 19. Jh., Tschebyshev, Markov, Ljapunov
- Ende 19. Jh., v. Bortkiewicz  
Anzahl der tödlichen Unfälle bei Pferdetritten
- Ende 19. Jh., Galton  
Begriffe Regression, Korrelation
- 1900, David Hilbert  
(2. Internationaler Mathematikerkongress in Paris)  
23 Probleme der Mathematik,  
u.a. Axiomatik der Wahrscheinlichkeitsrechnung.

# Geschichte (8)

- 1919 R.v. Mises

statistische Definition der Wahrscheinlichkeit,

Erfahrung:  $P(A) := \lim_{n \rightarrow \infty} h_n(A)$

Existiert der Grenzwert?

# Geschichte (8)

- 1919 R.v. Mises

statistische Definition der Wahrscheinlichkeit,

Erfahrung:  $P(A) := \lim_{n \rightarrow \infty} h_n(A)$

Existiert der Grenzwert?

- 1933, A.N. Kolmogorov

Axiomensystem der Wahrscheinlichkeitsrechnung

# Stochastik

- Statistik:

Gesamtheit aller Methoden zur Analyse zufallsbehafteter Datenmengen

→ Aussagen über die zugrundeliegende Grundgesamtheit treffen.

- Wahrscheinlichkeitsrechnung:

gegebene Grundgesamtheit (Verteilung)

→ Aussagen über Realisierungen einer Zufallsvariablen treffen.

---

- Stochastik: (grch.) im Rechnen geschickt.

# Literatur

Mathar, R. und Pfeiffer, D. (1990) Stochastik für Informatiker, Stuttgart

Pflug, G. (1986). Stochastische Modelle in der Informatik, Stuttgart

Greiner, M. und Tinhofer, G. (1996) Stochastik für Studienanfänger der  
Informatik, München

Steland, A. (2013). Basiswissen Statistik, Springer

Henze, N. (2004), Stochastik für Einsteiger, Wiesbaden

Dehling, H., Haupt, B. (2003). Einführung in die  
Wahrscheinlichkeitsrechnung, Springer

Büchter, A., Henn, H.-W. (2005). Elementare Stochastik, Springer

Rosanol, J.A. (1970). Wahrscheinlichkeitstheorie, Berlin

Flachsmeyer, J. (1970). Kombinatorik, Berlin

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit

### Def. 1 Ein zufälliger Versuch (Experiment)

ist ein Versuch mit ungewissem Ausgang.

Beispiel: Glücksspiele.

Wichtig bei solchen Experimenten ist:

- die Beschreibung des Experiments (Kartenspiele, Münzwurf),
- die Erfassung der Menge aller möglichen Ausgänge des Experiments.

# Zufällige Ereignisse (2)

## Def. 2 (Grundbegriffe)

- $\Omega$ : Menge aller möglichen Versuchsausgänge, sicheres Ereignis
- Ereignis  $A$ : Teilmenge von  $\Omega$ ,  $A \subseteq \Omega$ .
- Elementarereignis: einelementige Teilmenge von  $\Omega$
- unmögliches Ereignis:  $\emptyset$ .
- Komplementärereignis:  $\bar{A} = \Omega \setminus A$

Ein Experiment kann diskret sein, d.h. endlich oder abzählbar viele Ausgänge besitzen, oder es kann überabzählbar viele Ausgänge haben.



# Zufällige Ereignisse (3)

## Experimente mit einer endlichen Anzahl von Elementarereignissen

- Münzwurf
  - zwei Elementarereignisse:  $\{\text{Zahl } (z)\}, \{\text{Wappen } (w)\}$ ;
  - das unmögliche Ereignis  $\emptyset = \{z\} \cap \{w\}$ ;
  - das sichere Ereignis  $\Omega := \{z, w\}$ .

Die Menge der auftretenden Ereignisse ist

$$\mathcal{P}(\Omega) := \{\emptyset, \{z\}, \{w\}, \Omega\},$$

die Potenzmenge von  $\Omega$ .

# Zufällige Ereignisse (4)

## Würfel (1 mal)

Elementarereignisse:  $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\},$

$\Omega = \{1, 2, 3, 4, 5, 6\}.$

Damit erhalten wir für paarweise verschiedene

$i, j, k, l, m \in \{1, 2, 3, 4, 5, 6\}$  die möglichen Ereignisse :

|             |             |         |            |               |                  |                     |          |
|-------------|-------------|---------|------------|---------------|------------------|---------------------|----------|
| Ereignistyp | $\emptyset$ | $\{i\}$ | $\{i, j\}$ | $\{i, j, k\}$ | $\{i, j, k, l\}$ | $\{i, j, k, l, m\}$ | $\Omega$ |
|-------------|-------------|---------|------------|---------------|------------------|---------------------|----------|

|        |   |   |    |    |    |   |   |
|--------|---|---|----|----|----|---|---|
| Anzahl | 1 | 6 | 15 | 20 | 15 | 6 | 1 |
|--------|---|---|----|----|----|---|---|

Also insgesamt  $2^6 = 64$  mögliche Ereignisse.

# Zufällige Ereignisse (5)

## Experimente mit abzählbar vielen Elementarereignissen

- ① Werfen einer Münze, bis zum ersten Mal die Zahl fällt

$$\Omega = \{z, wz, wwz, wwz, wwwz, \dots\}.$$

- ② Anzahl der ankommenden Fahrzeuge an einer Kreuzung in einem bestimmten Zeitbereich

$$\Omega = \{0, 1, 2, \dots\}.$$

# Zufällige Ereignisse (6)

## Experimente mit überabzählbar vielen Elementarereignissen

- Lebensdauer einer Glühbirne

$$\Omega = [0, \infty[.$$

Ereignisse sind z.B. Intervalle und Punkte.

Es gilt beispielsweise:  $\emptyset = [0, 1] \cap [3, 5]$ .

Das Ereignis  $A = [0.4, 3.1] \cup \{7\}$  bedeutet, daß die Glühbirne eine Lebensdauer von 7s oder eine Lebensdauer zwischen 0.4s und 3.1s hat.

Anmerkung: Die Aussage gilt in Unkenntnis physikalischer Gesetze über diskrete Zeittakte.

# Zufällige Ereignisse (7)

## Experiment mit überabzählbar vielen Elementarereignisse

- Messung einer physikalischen Konstante

$$\underbrace{y}_{\text{Messwert}} = \underbrace{m}_{\text{Konstante}} + \underbrace{\varepsilon}_{\text{Messfehler}} .$$

Die Messfehler sind die Elementarereignisse. Ereignisse sind beispielsweise Intervalle.

- Experimente, deren Ausgänge Funktionen der Zeit sind,  
 $\Omega = \Omega_0 \times T$ . Ereignisse im Experiment sind dann bestimmte Funktionsverläufe  $\implies$  stochastische Prozesse.

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit

Ein Ereignisfeld  $\mathcal{E}$  ist (grob) ein System von Teilmengen der Menge  $\Omega$ . Es gilt:  $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ .

### Def. 3 ( $\cup, \cap$ , Komplement)

Es seien  $A_1 \in \mathcal{E}$  und  $A_2 \in \mathcal{E}$  Ereignisse. Dann

- $A_3 := A_1 \cap A_2 = \{\omega \in \Omega: \omega \in A_1 \text{ und } \omega \in A_2\}$  das Ereignis, bei dem  $A_1$  und  $A_2$  eintreten;
- $A_3 := A_1 \cup A_2 = \{\omega \in \Omega: \omega \in A_1 \text{ oder } \omega \in A_2\}$  das Ereignis, bei dem  $A_1$  oder  $A_2$  eintreten;
- $\overline{A_1} = \Omega \setminus A_1 = \{\omega \in \Omega: \omega \notin A_1\}$  das zu  $A_1$  komplementäre Ereignis.

## Ereignisfeld (2)

Es gilt offenbar:

- $A \cup \bar{A} = \Omega$  (sicheres Ereignis),
- $A \cap \bar{A} = \emptyset$  (unmögliches Ereignis).



# Ereignisfeld (3)

## Satz (Rechenregeln für Ereignisse)

- (i)  $A \cup B = B \cup A$  (Kommutativgesetz)
- (ii)  $(A \cup B) \cup C = A \cup (B \cup C)$  (Assoziativgesetz)
- (iii)  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- (iv)  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$   
(Distributivgesetze)
- (v) (De'Morgansche Regeln)

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

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

# Ereignisfeld (4)

## Def. 4

Seien  $A_1, \dots, A_n, \dots$  Ereignisse. Die Vereinigung  $\bigcup_{i=1}^{\infty} A_i$  ist das Ereignis, das eintritt, wenn mindestens eines Ereignisse  $A_1, A_2, A_3, \dots$  eintritt.

Der Durchschnitt  $\bigcap_{i=1}^{\infty} A_i$  ist das Ereignis, das eintritt, wenn alle Ereignisse  $A_1, A_2, A_3, \dots$  eintreten.

# Ereignisfeld (5)

## Verallgemeinerungen der Rechenregeln

Seien  $A, A_1, \dots$  Ereignisse.

$$(iii) \quad A \cap (\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} (A \cap A_i)$$

$$(iv) \quad A \cup (\bigcap_{i=1}^{\infty} A_i) = \bigcap_{i=1}^{\infty} (A \cup A_i)$$

$$(v)$$

$$\overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \overline{A_i}$$

$$\overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \overline{A_i}$$

# Ereignisfeld (6)

Def. 5  $\mathcal{E} \subseteq \mathcal{P}(\Omega)$  heißt Ereignisfeld über  $\Omega$

falls folgendes gilt:

- ①  $\Omega \in \mathcal{E}$ ;
- ② Gilt  $A_i \in \mathcal{E}$  für  $i \in \mathbf{N}$ , dann folgt  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{E}$ ;
- ③  $A \in \mathcal{E} \implies \bar{A} \in \mathcal{E}$ .

$\mathcal{E}$  heißt auch  $\sigma$ -Algebra über  $\Omega$ .

# Ereignisfeld (7)

## Grundlegende Eigenschaften

- Elementarereignisse schließen sich gegenseitig aus.
- Es tritt immer nur genau ein Elementarereignis ein.
- Ein Ereignis tritt genau dann ein, wenn eines seiner Elementarereignisse eintritt.

## Folgerung

- 1 Ist  $A_i \in \mathcal{E} \quad \forall i \in \mathbf{N}$ , so folgt:  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$ .
- 2 Für das unmögliche Ereignis gilt:  $\emptyset \in \mathcal{E}$ .

# Ereignisfeld (8)

## Beweis der Folgerung

1

$$A_i \in \mathcal{E}, \quad \forall i \in \mathbf{N} \implies \overline{A_i} \in \mathcal{E}, \quad \forall i \in \mathbf{N} \quad (\text{Def. 5.3})$$

$$\implies \bigcap_{i=1}^{\infty} \overline{A_i} \in \mathcal{E} \quad (\text{Def. 5.2})$$

$$\implies \overline{\bigcup_{i=1}^{\infty} A_i} \in \mathcal{E} \quad (\text{de Morgan})$$

$$\implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{E} \quad (\text{Def. 5.3})$$

- 2 Nach Def. 5.1 gilt:  $\Omega \in \mathcal{E}$ . Wegen  $\emptyset = \overline{\Omega}$  und Def. 5.3 folgt dann:  $\emptyset \in \mathcal{E}$ .

# Ereignisfeld (9)

Def. 6 Zwei Ereignisse  $A_1, A_2 \in \mathcal{E}$  heißen

unvereinbar (disjunkt), falls  $A_1 \cap A_2 = \emptyset$  gilt. Wir sagen dann auch, diese beiden Ereignisse schließen einander aus.

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit



**Def. 7 (Wahrscheinlichkeit)** Sei  $\mathcal{E}$  ein Ereignisfeld.

Eine Abbildung  $P: \mathcal{E} \rightarrow \mathbb{R}$  heißt Wahrscheinlichkeit, falls sie die folgenden Eigenschaften hat:

- ❶ Für alle  $A \in \mathcal{E}$  gilt:  $0 \leq P(A) \leq 1$ ;
- ❷  $P(\Omega) = 1$ ;
- ❸ Sind die Ereignisse  $A_1, A_2, \dots$  paarweise unvereinbar (d.h.  $A_i \cap A_j = \emptyset$  für  $i \neq j, i, j \in \mathbb{N}$ ), so gilt die sogenannte  $\sigma$ -Additivitätseigenschaft:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

# Kolmogorov'sches Axiomensystem (2)

## Def. 8 (Wahrscheinlichkeitsraum)

Sei  $\Omega$  die Menge der Elementarereignisse,  $\mathcal{E}$  ein Ereignisfeld über  $\Omega$  ( $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ ) und  $P$  genüge den KOLMOGOROV-Axiomen, dann heißt das Tripel  $(\Omega, \mathcal{E}, P)$  Wahrscheinlichkeitsraum.

Mittels dieses Begriffes ist eine vollständige Beschreibung eines zufälligen Experimentes möglich.

## Kolmogorov'sches Axiomensystem (3)

Wir betrachten nun  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ , ein System von Teilmengen der Menge  $\Omega$ . Dann können wir die folgende Menge bilden:

$$\mathcal{E}(\mathcal{A}) = \{\mathcal{E} : \mathcal{A} \subseteq \mathcal{E}, \mathcal{E} \text{ ist Ereignisfeld}\}.$$

Dann ist die Menge

$$\mathcal{E}_{\mathcal{A}} = \bigcap_{\mathcal{E} \in \mathcal{E}(\mathcal{A})} \mathcal{E}$$

die von  $\mathcal{A}$  erzeugte  $\sigma$ -Algebra (Ereignisfeld) bzw. die kleinste  $\sigma$ -Algebra über  $\Omega$ , die  $\mathcal{A}$  enthält.

# Kolmogorov'sches Axiomensystem (4)

Beispiele für Wahrscheinlichkeitsräume  $(\Omega, \mathcal{E}, P)$

## Def. 9 (klassische Definition der Wahrscheinlichkeit)

$$\Omega = \{\omega_1, \dots, \omega_N\}, \quad \mathcal{E} = \mathcal{P}(\Omega).$$

$P(\omega) = P(\{\omega_i\}) = \frac{1}{N} \quad \forall i = 1, \dots, N$ . D.h. alle Elementarereignisse sind gleichwahrscheinlich.

Sei  $A \in \mathcal{E}$ .

$$P(A) = \frac{\#\{\omega, \omega \in A\}}{N} = \frac{\#\text{für } A \text{ günstigen Elementarereignisse}}{\#\text{möglichen Elementarereignisse}}$$

ist die klassische Definition der Wahrscheinlichkeit.

# Kolmogorov'sches Axiomensystem (5)

## Def. 10 (Borel-Mengen)

Es sei  $\Omega = \mathbb{R}$  und

$$\mathcal{A} = \{[a, b[: -\infty < a < b < \infty\} \subseteq \mathcal{P}(\Omega).$$

die Menge der halboffenen Intervalle. Dann ist  $\mathcal{B}^1 := \mathcal{E}_{\mathcal{A}}$  die  $\sigma$ -Algebra der BOREL-Mengen.  $(\mathbb{R}, \mathcal{B}^1, P)$  ist dann ein Wahrscheinlichkeitsraum mit irgendeiner Wahrscheinlichkeit  $P$ .

# Kolmogorov'sches Axiomensystem (6)

Es sei  $\Omega = [0, 1]$ . Weiterhin betrachten wir:

$$\mathcal{E} = \{A: A = B \cap [0, 1], B \in \mathcal{B}^1\}.$$

die Menge der Borelmengen auf dem Intervall  $[0, 1]$ .

$$P: A \longrightarrow \mathbb{R} \text{ mit } P(A) := \int_A dx.$$

$$P(\Omega) = \int_0^1 dx = 1; \quad P\left(\left[\frac{1}{2}, \frac{3}{4}\right]\right) = \frac{1}{4}; \quad P\left(\left\{\frac{1}{2}\right\}\right) = \int_{\frac{1}{2}}^{\frac{1}{2}} dx = 0$$

# Kolmogorov'sches Axiomensystem (7)

$$Q: A \longrightarrow \mathbb{R} \text{ mit } Q(A) := \int_A \frac{3}{2}(1 - x^2)dx$$

$$Q(\Omega) = \int_0^1 \frac{3}{2}(1 - x^2)dx = \frac{3}{2} \left( x - \frac{x^3}{3} \right) \Big|_0^1 = 1$$

$(\Omega, \mathcal{E}, P)$  und  $(\Omega, \mathcal{E}, Q)$  sind Wahrscheinlichkeitsräume.

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit



## Folgerung:

Sei  $(\Omega, \mathcal{E}, P)$  Wahrscheinlichkeitsraum und  $A, B$  Ereignisse.

①  $P(\overline{A}) = 1 - P(A).$

②  $P(\emptyset) = 0.$

③ Sei  $A \subseteq B$ . Dann gilt:

①  $B \setminus A \in \mathcal{E};$

②  $P(B \setminus A) = P(B) - P(A)$  (Subtraktivität);

③  $P(A) \leq P(B)$  (Monotonie der Wahrscheinlichkeit).

④  $P(A \cup B) = P(A) + P(B) - P(A \cap B),$

$$P(A \cup B) \leq P(A) + P(B).$$

Sind  $A$  und  $B$  unvereinbar, so gilt die Gleichheit.

## Folgerungen (2)

Es sei  $\{A_n: n \in \mathbb{N}\}$  eine Folge von Ereignissen.

- 5 Es sei  $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$ . Dann gilt:

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

„Stetigkeit (des Wahrscheinlichkeitsmaßes) von unten“

- 6 Es sei  $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$ . Dann gilt:

$$P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

„Stetigkeit (des Wahrscheinlichkeitsmaßes) von oben“

# Beweis Folgerungen 1 und 2

- ① Es gilt:  $\Omega = A \cup (\Omega \setminus A) = A \cup \bar{A}$ , für alle  $A \in \mathcal{E}$ .

Wegen  $A \cap \bar{A} = \emptyset$  folgt:

$$\begin{aligned} 1 &= P(\Omega) = P(A \cup \bar{A}) \\ &= P(A) + P(\bar{A}) \end{aligned}$$

Wir stellen um und erhalten:  $P(\bar{A}) = 1 - P(A)$ .

- ② Wegen  $\emptyset = \Omega \setminus \Omega = \bar{\Omega}$  folgt aus Aussage 1:

$$P(\emptyset) = 1 - P(\Omega) = 0.$$

# Beweis Folgerungen 3

③ Es seien  $A, B \in \mathcal{E}$  zwei Ereignisse mit  $A \subseteq B$ .

① Es gilt:

$$B \setminus A = B \cap \bar{A}.$$

Wegen  $B \in \mathcal{E}$  und  $\bar{A} \in \mathcal{E}$  folgt nach Def. 5.(2.), dass auch die Menge  $B \setminus A \in \mathcal{E}$  ist.

② Aus  $B = A \cup (B \setminus A)$  und  $A \cap (B \setminus A) = \emptyset$  folgt:

$$\begin{aligned} P(B) &= P(A \cup (B \setminus A)) \\ &= P(A) + P(B \setminus A) \end{aligned}$$

Wir stellen um und erhalten:

$$P(B) - P(A) = P(B \setminus A).$$

# Beweis Folgerungen 3.-4.

## 3 Folgerung 3.3

- 1 Wenn wir die Subtraktivitätsgleichung etwas umstellen, erhalten wir:

$$P(B) = P(A) + P(B \setminus A).$$

Wegen Definition 7.(1.) folgt daraus sofort:

$$P(A) \leq P(B).$$

- 4 Es gilt  $A \cup B = A \cup (B \setminus A)$  und da die Ereignisse  $A$  und  $B \setminus A$  unvereinbar sind folgt die erste Behauptung aus Axiom 3 und der Subtraktivität.

Die zweite Behauptung folgt aus  $P(B \setminus A) \geq 0$ .

## Beweis Folgerung 5 (1)

- 5 Es sei nun  $\{A_n : n \in \mathbb{N}\}$  eine Folge von Ereignissen mit  $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$ . Damit gilt:

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k.$$

Wir definieren die Ereignisse

$$B_1 := A_1, \quad B_2 := A_2 \setminus A_1, \quad \dots \quad B_n := A_n \setminus A_{n-1} \quad \dots$$

Offenbar gilt für alle  $i, j \in \mathbb{N}$  mit  $i \neq j$ :

$$B_i \cap B_j = \emptyset \quad \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k.$$

# Beweis Folgerung 5 (2)

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right) \\ &= \sum_{k=1}^{\infty} P(B_k) \quad (\text{Definition 7.(3.)}) \end{aligned}$$

# Beweis Folgerung 5 (2)

$$\begin{aligned}P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right) \\&= \sum_{k=1}^{\infty} P(B_k) \quad (\text{Definition 7.(3.)}) \\&= P(A_1) + \sum_{k=2}^{\infty} P(A_k \setminus A_{k-1}) \\&= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=2}^n P(A_k \setminus A_{k-1})\end{aligned}$$



## Beweis Folgerung 5 (2)

$$\begin{aligned}P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right) \\&= \sum_{k=1}^{\infty} P(B_k) \quad (\text{Definition 7.(3.)}) \\&= P(A_1) + \sum_{k=2}^{\infty} P(A_k \setminus A_{k-1}) \\&= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=2}^n P(A_k \setminus A_{k-1}) \\&= \lim_{n \rightarrow \infty} \left( P(A_1) + \sum_{k=2}^n (P(A_k) - P(A_{k-1})) \right) \\&= \lim_{n \rightarrow \infty} P(A_n).\end{aligned}$$

# Beweis Folgerung 6 (1)

- 6 Es sei nun  $\{A_n: n \in \mathbb{N}\}$  eine Folge von Ereignissen mit der Eigenschaft  $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$ .

Dann gilt:

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{k=1}^{\infty} A_k.$$

Unter Anwendung der DE MORGAN'schen Regeln erhalten wir:

$$\lim_{n \rightarrow \infty} A_n = \overline{\bigcup_{k=1}^{\infty} \overline{A_k}}.$$

Außerdem gilt:  $\overline{A_k} \subseteq \overline{A_{k+1}}$ . Dann

# Beweis Folgerung 6 (2)

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} A_n\right) &= P\left(\overline{\bigcup_{k=1}^{\infty} \overline{A_k}}\right) \\ &= 1 - P\left(\bigcup_{k=1}^{\infty} \overline{A_k}\right) \quad (\text{Aussage 1}) \\ &= 1 - P\left(\lim_{n \rightarrow \infty} \overline{A_n}\right) \\ &= 1 - \lim_{n \rightarrow \infty} P(\overline{A_n}) \quad (\text{Aussage 5}) \\ &= 1 - \lim_{n \rightarrow \infty} (1 - P(A_n)) \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

# Subadditivität von P

Satz: Seien  $A_1, A_2, \dots$  Ereignisse. Dann gilt:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

Beweis:

$$B_1 := A_1$$

$$B_2 := A_2 \setminus A_1$$

$$B_3 := A_3 \setminus (A_1 \cup A_2)$$

...

$$B_i := A_i \setminus \left(\bigcup_{j < i} A_j\right) \dots$$

$B_i$  paarw. disjunkt,  $B_i \subseteq A_i$ .

# Subadditivität von P

Satz: Seien  $A_1, A_2, \dots$  Ereignisse. Dann gilt:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

Beweis:

$$B_1 := A_1$$

$$B_2 := A_2 \setminus A_1$$

$$B_3 := A_3 \setminus (A_1 \cup A_2)$$

...

$$B_i := A_i \setminus \left(\bigcup_{j < i} A_j\right) \dots$$

$$\begin{aligned} \bigcup_{i \geq 1} B_i &= \bigcup_{i \geq 1} A_i \Rightarrow \\ P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} P(B_i) \quad (3.Ax.) \\ &\leq \sum_{i=1}^{\infty} P(A_i) \quad (\text{Mon.}) \end{aligned}$$

$B_i$  paarw. disjunkt,  $B_i \subseteq A_i$ .

# Folgerungen (8)

## Siebformel, Prinzip von Inklusion und Exklusion

Seien  $A_1, \dots, A_n$  Ereignisse. Dann gilt:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|-1} P\left(\bigcap_{i \in I} A_i\right) \\ &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \dots \\ &\quad (-1)^{n+1} \sum_{i_1 < i_2 < \dots < i_n} P\left(\bigcap_{\nu=1}^n A_{i_\nu}\right) \end{aligned}$$

auch: Formel von Poincare-Sylvester  
(Montmort: Briefwechsel mit Bernoulli)

# Siebformel

Beweis: (Induktion nach  $n$ )

- ① IA  $n = 1$  trivial, ( $n = 2$  : Subtraktivität)

$$\begin{aligned}P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \\&= \sum_{i=1}^2 P(A_i) - \sum_{I=\{1,2\}} P(A_i \cap A_j) \\&= \sum_{I \subseteq \{1, \dots, n\}, I \neq \emptyset} (-1)^{|I|-1} P\left(\bigcap_{i \in I} A_i\right)\end{aligned}$$

- ② IS: Aussage der Folgerung gelte für  $n$ . Dann

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right)$$

wegen Subtraktivität.

# Siebformel

## Beweis (2)

Auf den ersten und dritten Summanden wird jeweils die IV angewendet. Der dritte Summand ist gleich

$$\begin{aligned}
 -P\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) &= - \sum_{J \subseteq \{1, \dots, n\}, J \neq \emptyset} (-1)^{|J|-1} P\left(\bigcap_{i \in J} (A_i \cap A_{n+1})\right) \\
 &= \sum_{\{n+1\} \subseteq J \subseteq \{1, \dots, n+1\}, J \neq \{n+1\}} (-1)^{|J|-1} P\left(\bigcap_{i \in J} A_i\right).
 \end{aligned}$$



# Siebformel

## Beweis (3)

Untersuchung der Indexmengen:

1. Summe: alle nichtleeren Teilmengen von  $\{1, \dots, n\}$
3. Summe: alle nicht-1-Element. Teilmengen von  $\{1, \dots, n+1\}$ , die das Element  $n+1$  enthalten
2. Summe: das Element  $n+1$ .

Damit tauchen alle nichtleeren Teilmengen von  $\{1, \dots, n+1\}$  in einer der Summanden auf.

Alle Summanden haben die gleiche Form, wie in der Siebformel.

# Beispiele zur Siebformel (1)

## Rencontre-Problem

$n$  Studenten sollen schriftlich von einer Änderung des Vorlesungstermins benachrichtigt werden. Im irrtümlichen Glauben, daß jeder der  $n$  Briefe den gleichen Inhalt aufweist, verteilt eine Sekretärin die Briefe willkürlich in die verschiedenen Umschläge. Wie groß ist die Wahrscheinlichkeit, daß mindestens ein Brief in den richtigen Umschlag gelangt? Welchen Wert erhält man für  $n \rightarrow \infty$ ?  
Lösung: Übung.

# Beispiele zur Siebformel (2)

## Sortierprobleme

geg.: Feld der Länge  $n$

Daten zufällig angeordnet, gleichverteilt mit Wahrscheinlichkeit  $\frac{1}{n!}$ .

Wie groß ist die Wahrscheinlichkeit, daß mindestens ein Feldelement schon an der richtigen Stelle liegt.? Welchen Wert erhält man für  $n \rightarrow \infty$ ?

das ist dasselbe wie beim Rencontre-Problem.

Wie groß ist die Wahrscheinlichkeit, daß genau  $k$  Elemente bereits am richtigen Platz stehen?  $\rightarrow$  Übung

# Folgerungen aus der Siebformel

## Bonferroni-Ungleichungen (1)

Die Ungleichung

$$P(A \cup B) \leq P(A) + P(B)$$

heißt Bonferroni-Ungleichung.

Weitere (Bonferroni)- Ungleichungen erhält man durch Abbruch der Siebformel nach Gliedern mit positivem ( $\leq$ ) bzw. negativem ( $\geq$ ) Vorzeichen.

# Folgerungen aus der Siebformel

## Bonferroni-Ungleichungen (2)

$$P(A \cup B \cup C) \leq P(A) + P(B) + P(C) \quad (n = 1)$$

$$P(A \cup B \cup C) \geq P(A) + P(B) + P(C) \quad (n = 2) \\ - P(A \cap B) - P(A \cap C) - P(B \cap C)$$

$$P(A \cup B \cup C) \leq P(A) + P(B) + P(C) \\ - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ + P(A \cap B \cap C)$$

(n=3, es gilt hier sogar Gleichheit)

# Inhalt

- Einleitung, Geschichte
- Zufällige Ereignisse
- Ereignisfeld
- Kolmogorov'sches Axiomensystem
- Folgerungen aus dem Kolmogorov-Axiomensystem
- Die klassische Definition der Wahrscheinlichkeit

Wir betrachten für ein zufälliges Experiment die Menge der Elementarereignisse  $\Omega = \{\omega_1, \dots, \omega_N\}$ . Sei  $\mathcal{E} = \mathcal{P}(\Omega)$  und  $P(\{\omega_i\}) = \frac{1}{N}, \forall i = 1, \dots, N$ .

$$\begin{aligned} P(A) &= \frac{\#\{\omega: \omega \in A\}}{N} = \frac{n(A)}{N} \\ &= \frac{\# \text{ der für } A \text{ günstigen Elementarereignisse}}{\# \text{ der möglichen Elementarereignisse}} \end{aligned}$$

# DE MÉRÉ (1)

## Würfel mit 3 Würfeln

Folgende Ereignisse werden betrachtet:

$A$  = Es fallen 11 Augen.

$B$  = Es fallen 12 Augen.

Frage:  $P(A), P(B)$ ?



# DE MÉRÉ (1)

## Würfel mit 3 Würfeln

Folgende Ereignisse werden betrachtet:

$A$  = Es fallen 11 Augen.

$B$  = Es fallen 12 Augen.

Frage:  $P(A), P(B)$ ?

Die Menge der Elementarereignisse ist

$$\Omega = \{(i, j, k) : 1 \leq i, j, k \leq 6\}.$$

Anzahl der Elementarereignisse  $N := 6^3 = 216$ ,  $P((i, j, k)) = \frac{1}{216}$ .

# DE MÉRÉ (2)

Anzahl der Ereignisse

| <b>A</b> (11 Augen) |   | <b>B</b> (12 Augen) |   |
|---------------------|---|---------------------|---|
| 6-4-1               | 6 | 6-5-1               | 6 |
| 6-3-2               | 6 | 6-4-2               | 6 |
| 5-5-1               | 3 | 6-3-3               | 3 |
| 5-4-2               | 6 | 5-5-2               | 3 |
| 5-3-3               | 3 | 5-4-3               | 6 |
| 4-4-3               | 3 | 4-4-4               | 1 |
| $n(A)=27$           |   | $n(B)=25$           |   |

# DE MÉRÉ (2)

Anzahl der Ereignisse

| <b>A</b> (11 Augen) |   | <b>B</b> (12 Augen) |   |
|---------------------|---|---------------------|---|
| 6-4-1               | 6 | 6-5-1               | 6 |
| 6-3-2               | 6 | 6-4-2               | 6 |
| 5-5-1               | 3 | 6-3-3               | 3 |
| 5-4-2               | 6 | 5-5-2               | 3 |
| 5-3-3               | 3 | 5-4-3               | 6 |
| 4-4-3               | 3 | 4-4-4               | 1 |
| n(A)=27             |   | n(B)=25             |   |

$$P(A) = \frac{27}{216} > \frac{25}{216} = P(B).$$

# Inhalt

- Klassische kombinatorische Probleme
- Beispiele
- Arithmetische Beziehungen zwischen den Binomialkoeffizienten
- Die Stirling Formel

# Aufgabenstellung

Anzahl der verschiedenen Zusammenstellungen von Objekten. Je nach Art der zusätzlichen Forderungen, ist zu unterscheiden, welche Zusammenstellungen als gleich, und welche als verschieden angesehen werden.

- Permutation (ohne Wiederholung)
- Permutation mit Wiederholung
- Variation ohne Wiederholung
- Variation mit Wiederholung
- Kombination (ohne Wiederholung)
- Kombination mit Wiederholung

# Klassische kombinatorische Probleme (1)

## Permutation (ohne Wiederholung)

Jede eindeutige Abbildung  $\pi$  der geordneten Menge  $\{1, \dots, n\}$  auf eine  $n$ -elementige Menge  $M = \{s_1, \dots, s_n\}$  heißt Permutation oder Permutation ohne Wiederholung,

$$\forall i \in \{1, \dots, n\} : \pi(i) = s_i, s_i \in M, s_i \neq s_j (i \neq j)$$

Anzahl:

$$N = n!$$

Wiewiel Möglichkeiten gibt es, die Eisenbahnwagen 32,33,34,35,36,37 hintereinander zu hängen?

# Klassische kombinatorische Probleme (1)

## Permutation (ohne Wiederholung)

Jede eindeutige Abbildung  $\pi$  der geordneten Menge  $\{1, \dots, n\}$  auf eine  $n$ -elementige Menge  $M = \{s_1, \dots, s_n\}$  heißt Permutation oder Permutation ohne Wiederholung,

$$\forall i \in \{1, \dots, n\} : \pi(i) = s_i, s_i \in M, s_i \neq s_j (i \neq j)$$

Anzahl:

$$N = n!$$

Wiewiel Möglichkeiten gibt es, die Eisenbahnwagen 32,33,34,35,36,37 hintereinander zu hängen?

$$N = 6!$$

# Klassische kombinatorische Probleme (2)

## Permutation mit Wiederholung

Sei  $M = \{s_1, \dots, s_k\}$ ,  $k_i > 0 \forall i = 1, \dots, k$  mit  $\sum_{i=1}^k k_i = n$ . Jedes geordnete  $n$ -Tupel von Elementen aus  $M$ , wobei jedes Element  $s_i$  genau  $k_i$  mal vorkommt, heißt Permutation mit Wiederholung.

Anzahl:

$$N = \frac{n!}{k_1! \dots k_k!}$$

Wieviele Möglichkeiten gibt es, die Karten beim Skatspiel zu vergeben?



# Klassische kombinatorische Probleme (2)

## Permutation mit Wiederholung

Sei  $M = \{s_1, \dots, s_k\}$ ,  $k_i > 0 \forall i = 1, \dots, k$  mit  $\sum_{i=1}^k k_i = n$ . Jedes geordnete  $n$ -Tupel von Elementen aus  $M$ , wobei jedes Element  $s_i$  genau  $k_i$  mal vorkommt, heißt Permutation mit Wiederholung.

Anzahl:

$$N = \frac{n!}{k_1! \dots k_k!}$$

Wieviele Möglichkeiten gibt es, die Karten beim Skatspiel zu vergeben?

$$N = \frac{32!}{10!10!10!2!}$$

# Klassische kombinatorische Probleme (3)

## Variation ohne Wiederholung

Sei  $M = \{s_1, \dots, s_n\}$ . Jedes geordnete  $k$ -Tupel,  $k \leq n$  von verschiedenen Elementen aus  $M$  heißt Variation ohne Wiederholung.

Anzahl: 
$$N = n(n-1) \cdots (n-k+1)$$

Aufteilung von  $k$  Elementen auf  $n$  Fächer.

Wieviele Möglichkeiten für die drei Erstplatzierten im 100m Endlauf gibt es?

# Klassische kombinatorische Probleme (3)

## Variation ohne Wiederholung

Sei  $M = \{s_1, \dots, s_n\}$ . Jedes geordnete  $k$ -Tupel,  $k \leq n$  von verschiedenen Elementen aus  $M$  heißt Variation ohne Wiederholung.

Anzahl: 
$$N = n(n-1) \cdots (n-k+1)$$

Aufteilung von  $k$  Elementen auf  $n$  Fächer.

Wieviele Möglichkeiten für die drei Erstplatzierten im 100m Endlauf gibt es?

$$N = 8 \cdot 7 \cdot 6 = 336.$$

# Klassische kombinatorische Probleme (4)

## Variation mit Wiederholung

Auswahl von  $k$  Elementen aus einer Menge  $M = \{s_1, \dots, s_n\}$  mit Zurücklegen. Die Frage ist:

Wieviel verschiedene Möglichkeiten gibt es,  $k$  Elemente aus dieser Menge zu entnehmen, wobei Elemente mehrfach entnommen werden können?

$$N = n^k.$$

Anzahl der 10stelligen Dualzahlen:

# Klassische kombinatorische Probleme (4)

## Variation mit Wiederholung

Auswahl von  $k$  Elementen aus einer Menge  $M = \{s_1, \dots, s_n\}$  mit Zurücklegen. Die Frage ist:

Wieviel verschiedene Möglichkeiten gibt es,  $k$  Elemente aus dieser Menge zu entnehmen, wobei Elemente mehrfach entnommen werden können?

$$N = n^k.$$

Anzahl der 10stelligen Dualzahlen:

$$N = 2^{10}.$$

# Klassische kombinatorische Probleme (5)

## Kombinationen (ohne Wiederholung)

Jede  $k$ -elementige Teilmenge aus einer  $n$ -elementigen Menge  $M$  heißt Kombination (ohne Wiederholung) (von  $k$  aus  $n$  Elementen). Dabei sind Wiederholungen nicht erlaubt und die Reihenfolge der  $k$  Elemente wird nicht berücksichtigt.

$$N = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} = \binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Anzahl der 5er im Lotto: ÜA

# Klassische kombinatorische Probleme (6)

## Kombination (mit Wiederholung)

Fasst man alle  $n^k$  Variationen mit Wiederholung ( $n$  Elemente, Ordnung  $k$ ) zu Äquivalenzklassen zusammen, so daß sie aus den gleichen Elementen der gleichen Anzahl bestehen, so heißt jede solche Klasse Kombination mit Wiederholung. Die Anzahl solcher Äquivalenzklassen ist dann

$$N = \binom{n+k-1}{k}$$

$n = 2, k = 3$ : 4 Klassen:

$\{aaa\}$ ,  $\{aab, aba, baa\}$ ,  $\{abb, bab, bba\}$ ,  $\{bbb\}$  werden jeweils zu einer Klasse zusammengefaßt.

# Klassische kombinatorische Probleme (6a)

Zur Erläuterung betrachten wir ein Gefäß (oder ein Nagelbrett) das geordnet in  $k$  Teile unterteilt ist und werfen  $n$  Kugeln hinein und zählen jeweils die Anzahlen der Kugeln in den einzelnen Teilen. Die Anzahl der möglichen Anordnungen der  $n$  Kugeln und  $(k - 1)$  Trennwände ist  $(n + k - 1)!$ . Die Reihenfolge der  $n$  Kugeln und der  $k - 1$  Trennwände interessiert hier jedoch nicht. Deshalb haben wir eine Permutation mit Wiederholung, deren Anzahl ist

$$\frac{(n + k - 1)!}{n!(k - 1)!} = \binom{n + k - 1}{k - 1} = \binom{n + k - 1}{n}$$



Alternativ betrachten wir wieder eine Urne mit  $n$  weißen Kugeln, nummeriert von 1 bis  $n$ , und legen  $k - 1$  neue (rote) Kugeln mit den Nummern  $1-(k-1)$  hinzu. Aus der Urne mit jetzt  $n + k - 1$  verschiedenen Kugeln ziehen wir jetzt  $k$  mal ohne Zurücklegen.

Jetzt gibt es folgende Fälle:

1. Es sind nur weiße Kugeln gezogen mit den Nummern  $i_1 < \dots < i_k$   
d.h. jede der Nummern ist einfach
2. Es ist genau eine rote Kugel mit Nummer  $j$  gezogen. Dann kommt die Kugel mit der  $j$ -größten Nummer zweimal vor, alle anderen einfach. ( $j \leq k - 1$ )
3. Es sind genau zwei rote Kugeln mit Nummern  $j_1 < j_2$  gezogen. Dann kommen die Kugeln mit den  $j_1$  und  $j_2$ -größten Nummern je zweimal vor, alle anderen einfach. ( $j_1, j_2 \leq k - 2$ )

4. Ist im Fall 3  $j_2 = k - 1$  dann kommt die weiße Kugel mit Nummer  $j_1$  genau dreimal vor.
5. Analog gehen wir vor, wenn wir mehr als zwei rote Kugeln gezogen haben.

Weitere Erläuterung zur Kombination mit Wiederholung: siehe Beispiele 4, 5 und 6.

# Klassische kombinatorische Probleme (7)

## Kombination von Elementen aus mehreren Mengen

Wir betrachten beliebige Mengen  $S_1, \dots, S_k$ , wobei

$S_i = \{s_{i1}, \dots, s_{in_i}\}$  ( $i = 1, \dots, k$ ) gilt.

*Wieviele verschiedene Kombinationen von je einem Element der Mengen  $S_1, \dots, S_k$  können gebildet werden?*

Solche Kombinationen haben die Form  $(s_{1i_1}, \dots, s_{ki_k})$ , wobei  $s_{ki_k} \in S_k$  gilt für alle  $i = 1, \dots, k$ .

Anzahl:

$$N = n_1 \cdot n_2 \cdot \dots \cdot n_k.$$

# Inhalt

- Klassische kombinatorische Probleme
- Beispiele
- Arithmetische Beziehungen zwischen den Binomialkoeffizienten
- Die Stirling Formel

# Beispiele (1)

## Eine Gruppe von $r$ Studenten verweist in einem Zug

Die Studenten verteilen sich zufällig auf  $n \geq r$  Abteile. Es sei  $A$  das Ereignis, daß alle Studenten in verschiedenen Abteilen sitzen.

$$P(A) = \frac{n(A)}{N}.$$

$N = n^r = \# \text{Möglichkeiten für die Verteilung der } r \text{ Studenten auf die } n \text{ Abteile}$

$$n(A) = n \cdot (n-1) \cdot \dots \cdot (n-r+1)$$

$$P(A) = \frac{n(A)}{N} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-r+1)}{n^r}.$$

# Beispiele (2)

## Ziehen von Kugeln

In einer Urne sollen sich  $n$  Kugeln befinden. Von diesen seien  $n_1$  schwarz,  $n - n_1$  dagegen weiß. Nun werden  $k$  Kugeln (zufällig) entnommen, und zwar ohne Zurücklegen.

A: “von diesen  $k$  Kugeln genau  $k_1$  schwarz”

$$P(A) = \frac{n(A)}{N}.$$

$N = \binom{n}{k} = \#$  Möglichkeiten,  $k$  Kugeln aus  $n$  Kugeln auszuwählen.

# Beispiele (2a)

## Ziehen von Kugeln (Fortsetzung)

$n(A)$  = Anzahl der Möglichkeiten zur Entnahme von  $k$  Kugeln, bei denen genau  $k_1$  schwarze Kugeln ausgewählt werden.

In einem solchen Fall sind dann auch genau  $k - k_1$  weiße Kugeln entnommen worden. Also

- 1 Die Anzahl der Möglichkeiten, aus  $n_1$  schwarzen Kugeln  $k_1$  schwarze auszuwählen (ohne Wiederholung und ohne Berücksichtigung der Reihenfolge) ist  $\binom{n_1}{k_1}$ .

# Beispiele (2b)

## Ziehen von Kugeln (Fortsetzung)

- ① Die Anzahl der Möglichkeiten, aus  $n - n_1$  weißen Kugeln  $k - k_1$  weiße auszuwählen (ebenfalls ohne Wiederholung und ohne Berücksichtigung der Reihenfolge) ist  $\binom{n-n_1}{k-k_1}$ .

$$\# \text{günstige Ereignisse} = n(A) = \binom{n_1}{k_1} \cdot \binom{n - n_1}{k - k_1}$$

$$P(A) = \frac{n(A)}{N} = \frac{\binom{n_1}{k_1} \cdot \binom{n-n_1}{k-k_1}}{\binom{n}{k}}$$

*Hypergeometrische Wahrscheinlichkeit.*



# Beispiele (3)

## Lotto 6 aus 49

Wenn wir uns die Zahlen als Kugeln denken, die aus einer Urne entnommen werden, und außerdem gezogene Zahlen im nachhinein als schwarze Kugeln ansehen, so kann jeder Tip durch die Entnahme von 6 Kugeln verkörpert werden.  $A$ : Ereignis, dass vier Richtige getippt werden.

$$n = 49, \quad n_1 = 6, \quad k = 6, \quad k_1 = 4,$$

$$P(A) = \frac{\binom{6}{4} \cdot \binom{49-6}{6-4}}{\binom{49}{6}}$$

## Beispiele (4)

### Zwei nicht unterscheidbare Würfel

Wie groß ist die Anzahl der Würfe mit 2 nicht zu unterscheidenden Würfeln?

Seien  $i, j$  die Augenzahlen und o.B.d.A.  $i \leq j$ .

Wir vergeben die Tupel  $(i, j)$ , wenn  $i \neq j$ .

Wir vergeben die Tupel  $(i, j)$ , wenn  $i = j$ .

Die gesuchte Anzahl ist die Anzahl der möglichen Auswahlen aus der Menge  $\{1, \dots, 7\}$ , d.h.  $\binom{7}{2}$ .

# Beispiele (5)

Wie groß ist die Anzahl der Würfe mit 3

nicht zu unterscheidenden Würfeln?

Seien  $i, j, k$  die Augenzahlen und o.B.d.A.  $i \leq j \leq k$ . Wir vergeben die Tripel

$(i, j, k)$ , wenn  $i < j < k$ .

$(i, k, 7)$ , wenn  $i = j < k$ .

$(i, j, 8)$ , wenn  $i < j = k$ .

$(i, 7, 8)$ , wenn  $i = j = k$ .

Die gesuchte Anzahl ist die Anzahl der möglichen Auswahlen aus der Menge  $\{1, \dots, 8\}$ , d.h.  $\binom{8}{3}$ .

# Beispiele (6)

Verteilen von  $n$  Geldstücken an  $k$  Studenten ( $k \leq n$ )

Auf wieviele Weisen ist das möglich?

# Beispiele (6)

## Verteilen von $n$ Geldstücken an $k$ Studenten ( $k \leq n$ )

Auf wieviele Weisen ist das möglich?

- a) jeder Student bekommt mindestens ein Stück.

Geldstücke nebeneinander legen und  $k - 1$  Trennstriche verteilen  
unter  $n - 1$  möglichen

$$N = \binom{n-1}{k-1}$$

# Beispiele (6a)

## Verteilen von $n$ Geldstücken

- b) es wird zugelassen, dass Studenten nichts erhalten.

# Beispiele (6a)

## Verteilen von $n$ Geldstücken

- b) es wird zugelassen, dass Studenten nichts erhalten.

Trick: Borgen von  $k$  Stücken  $\longrightarrow n + k$  Stück

$k - 1$  Trennstriche verteilen unter den jetzt  $n + k - 1$  möglichen

$$N = \binom{n+k-1}{k-1}$$

Dann gibt jeder Student genau ein Stück zurück.

## Beispiele (6b)

ein weiterer Zugang:

### Verteilen von $n$ Geldstücken an $k$ Studenten

Wir basteln einen Würfel mit  $k$  Flächen und würfeln  $n$  mal.

Beim  $i$ -ten Wurf bekommt der Student das Geldstück, dessen Nummer gewürfelt wurde.

Die gesuchte Anzahl ist dieselbe wie bei Würfeln mit  $n$  nicht unterscheidbaren Würfeln.

$$N = \binom{n+k-1}{n} = \binom{n+k-1}{k-1}$$



# Beispiele (7)

## Hashing

Beobachtungen (oder Daten) abspeichern auf einem Feld.

$k$ : Anzahl der Beobachtungen

$n$ : Feldlänge ( $k \leq n$ )

Das Abspeichern geschieht mit Hilfe von Hashfunktionen (oder Hashtafeln).

zufällige Daten: Kollisionen können auftreten.

$A_{k,n}$ : Ereignis, daß Kollisionen auftreten. ges.:  $P(A_{k,n})$

# Beispiele (7a)

## Hashing (Fortsetzung)

$$P(\overline{A}_{k,n}) = \frac{n(n-1) \cdots (n-k+1)}{n^k}$$

# Beispiele (7a)

## Hashing (Fortsetzung)

$$P(\bar{A}_{k,n}) = \frac{n(n-1) \cdots (n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right)$$

# Beispiele (7a)

## Hashing (Fortsetzung)

$$\begin{aligned}P(\bar{A}_{k,n}) &= \frac{n(n-1)\cdots(n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\&= \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right)\end{aligned}$$

# Beispiele (7a)

## Hashing (Fortsetzung)

$$\begin{aligned}P(\bar{A}_{k,n}) &= \frac{n(n-1) \cdots (n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\&= \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right) \\&\leq \exp\left(-\sum_{i=0}^{k-1} \frac{i}{n}\right)\end{aligned}$$

# Beispiele (7a)

## Hashing (Fortsetzung)

$$\begin{aligned}P(\bar{A}_{k,n}) &= \frac{n(n-1) \cdots (n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\&= \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right) \\&\leq \exp\left(-\sum_{i=0}^{k-1} \frac{i}{n}\right) \\&= \exp\left(-\frac{(k-1)k}{2n}\right) \approx \exp\left(-\frac{k^2}{2n}\right)\end{aligned}$$

$$\ln(1-x) < -x \text{ für } x < 1$$

# Beispiele (8)

Suche von Elementen. Sei  $n = |\Omega|$

Greifen zufällig eine  $k$ -elementige Teilmenge  $A \subseteq \Omega$  heraus.

$\omega_1, \dots$ : Schlüsselemente (vorgegeben),  $\omega_1, \dots \in \Omega$

Frage: Mit welcher Wkt.  $\omega_1 \in A$ ?

# Beispiele (8)

Suche von Elementen. Sei  $n = |\Omega|$

Greifen zufällig eine  $k$ -elementige Teilmenge  $A \subseteq \Omega$  heraus.

$\omega_1, \dots$ : Schlüsselemente (vorgegeben),  $\omega_1, \dots \in \Omega$

Frage: Mit welcher Wkt.  $\omega_1 \in A$ ?

$$P(A) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$



# Beispiele (8)

Suche von Elementen. Sei  $n = |\Omega|$

Greifen zufällig eine  $k$ -elementige Teilmenge  $A \subseteq \Omega$  heraus.

$\omega_1, \dots$ : Schlüsselemente (vorgegeben),  $\omega_1, \dots \in \Omega$

Frage: Mit welcher Wkt.  $\omega_1 \in A$ ?

$$P(A) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

Frage: Mit welcher Wkt.  $\omega_1, \dots, \omega_r \in A$ ?

# Beispiele (8)

Suche von Elementen. Sei  $n = |\Omega|$

Greifen zufällig eine  $k$ -elementige Teilmenge  $A \subseteq \Omega$  heraus.

$\omega_1, \dots$ : Schlüsselemente (vorgegeben),  $\omega_1, \dots \in \Omega$

Frage: Mit welcher Wkt.  $\omega_1 \in A$ ?

$$P(A) = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

Frage: Mit welcher Wkt.  $\omega_1, \dots, \omega_r \in A$ ?

$$P(A) = \frac{\binom{n-r}{k-r}}{\binom{n}{k}} = \frac{k(k-1) \cdots (k-r+1)}{n(n-1) \cdots (n-r+1)}$$

# Beispiele (8a)

## Suche von Elementen (Fortsetzung)

Sei die Anzahl  $r$  der Schlüsselemente fest,  $\frac{k}{n} \rightarrow p$ :  $P(A) \sim p^r$

$$P(A) \gtrsim \frac{1}{2}, \quad \text{falls} \quad p^r \geq \frac{1}{2} \quad \text{falls} \quad k \geq \frac{n}{2^{1/r}}$$

Soll also die Wahrscheinlichkeit, daß alle  $r$  Schlüsselemente in der Teilmenge enthalten sind, größer als  $\frac{1}{2}$  sein, so muss

$$k \geq \frac{n}{2^{1/r}}$$

gewählt werden.

# Kombinatorik

## Zusammenfassung

$n$ : # Elemente =  $|\Omega|$

$k$ : # auszuwählende Elemente

$k_1, \dots, k_m$ : Häufigkeit der einzelnen Elemente

ohne Wiederholung

mit Wiederholung

Permutationen

$$n!$$

$$\frac{n!}{k_1! \cdots k_m!}$$

Variationen

$$n(n-1) \cdots (n-k+1)$$

$$n^k$$

Kombinationen

$$\binom{n}{k}$$

$$\binom{n+k-1}{k}$$

# Inhalt

- Klassische kombinatorische Probleme
- Beispiele
- Arithmetische Beziehungen zwischen den Binomialkoeffizienten
- Die Stirling Formel

# Arithmetische Beziehungen zwischen den Binomialkoeffizienten (1)

• 1.

$$\binom{n}{k} = \binom{n}{n-k}$$

• 2.

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

• 3.

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

• 4.

$$\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$$

# Arithmetische Beziehungen zwischen den Binomialkoeffizienten (2)

• 5.

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$$

• 6.

$$\sum_{i=0}^n \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k}$$

• 7.

$$\sum_{k=1}^n k \binom{n}{k} = n \cdot 2^{n-1}$$

# Arithmetische Beziehungen zwischen den Binomialkoeffizienten (3)

- 8. Definieren die Folge

$$S_n = \sum_{k=0}^{\lfloor \frac{n+1}{2} \rfloor} \binom{n-k}{k}$$

Zeigen Sie:  $S_{n+1} = S_n + S_{n-1}$ .

**Beweis:** 3 Methoden,

vollständige Induktion

algebraisch

kombinatorisch

teilweise Übungsaufgabe, teilweise Übung





# Inhalt

- Klassische kombinatorische Probleme
- Beispiele
- Arithmetische Beziehungen zwischen den Binomialkoeffizienten
- Die Stirling Formel

Satz: Es gilt

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

**Beweis:** Die Aussage des Satzes ist äquivalent zu

$$\ln n! \sim \ln \sqrt{2\pi} + \left(n + \frac{1}{2}\right) \ln n - n.$$

Sei

$$d_n := \ln n! - \left(n + \frac{1}{2}\right) \ln n + n.$$

Es genügt zu zeigen,

$$\lim_{n \rightarrow \infty} d_n = \ln \sqrt{2\pi}.$$

## Beweis der Stirling-Formel (2)

Wir schätzen die Differenz  $d_n - d_{n+1}$  ab, dann das Verhalten der Folge  $\{d_n\}$  und versuchen den Grenzwert zu bestimmen. Die Differenz  $d_n - d_{n+1}$  ist

$$\begin{aligned} &= \ln n! - \ln(n+1)! \\ &\quad - \left(n + \frac{1}{2}\right) \ln n + \left(n + 1 + \frac{1}{2}\right) \ln(n+1) + n - (n+1) \end{aligned}$$

## Beweis der Stirling-Formel (2)

Wir schätzen die Differenz  $d_n - d_{n+1}$  ab, dann das Verhalten der Folge  $\{d_n\}$  und versuchen den Grenzwert zu bestimmen. Die Differenz  $d_n - d_{n+1}$  ist

$$\begin{aligned} &= \ln n! - \ln(n+1)! \\ &\quad - \left(n + \frac{1}{2}\right) \ln n + \left(n + 1 + \frac{1}{2}\right) \ln(n+1) + n - (n+1) \\ &= \ln \frac{n!}{(n+1)!} + \left(n + \frac{1}{2}\right)(\ln(n+1) - \ln n) + \ln(n+1) - 1 \end{aligned}$$

## Beweis der Stirling-Formel (2)

Wir schätzen die Differenz  $d_n - d_{n+1}$  ab, dann das Verhalten der Folge  $\{d_n\}$  und versuchen den Grenzwert zu bestimmen. Die Differenz  $d_n - d_{n+1}$  ist

$$\begin{aligned} &= \ln n! - \ln(n+1)! \\ &\quad - \left(n + \frac{1}{2}\right) \ln n + \left(n + 1 + \frac{1}{2}\right) \ln(n+1) + n - (n+1) \\ &= \ln \frac{n!}{(n+1)!} + \left(n + \frac{1}{2}\right)(\ln(n+1) - \ln n) + \ln(n+1) - 1 \\ &= -\ln(n+1) + \left(n + \frac{1}{2}\right) \ln \frac{n+1}{n} + \ln(n+1) - 1 \end{aligned}$$

## Beweis der Stirling-Formel (2)

Wir schätzen die Differenz  $d_n - d_{n+1}$  ab, dann das Verhalten der Folge  $\{d_n\}$  und versuchen den Grenzwert zu bestimmen. Die Differenz  $d_n - d_{n+1}$  ist

$$\begin{aligned} &= \ln n! - \ln(n+1)! \\ &\quad - \left(n + \frac{1}{2}\right) \ln n + \left(n + 1 + \frac{1}{2}\right) \ln(n+1) + n - (n+1) \\ &= \ln \frac{n!}{(n+1)!} + \left(n + \frac{1}{2}\right)(\ln(n+1) - \ln n) + \ln(n+1) - 1 \\ &= -\ln(n+1) + \left(n + \frac{1}{2}\right) \ln \frac{n+1}{n} + \ln(n+1) - 1 \\ &= \frac{2n+1}{2} \ln \frac{n+1}{n} - 1 \\ &= (2n+1) \cdot \frac{1}{2} \ln \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}} - 1. \end{aligned}$$

# Beweis der Stirling-Formel (3)

Es gilt für  $-1 < x < 1$ :

$$\ln(1+x) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{x^i}{i}$$

$$\ln(1-x) = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{(-x)^i}{i}$$

$$\ln \frac{1+x}{1-x} = \ln(1+x) - \ln(1-x) = 2 \sum_{i=0}^{\infty} \frac{x^{2i+1}}{2i+1}$$

# Beweis der Stirling-Formel (4)

Setzen  $x := \frac{1}{2n+1}$  und erhalten ( $x \neq 0$ )

$$\begin{aligned}d_n - d_{n+1} &= \frac{1}{x} \cdot \frac{1}{2} \cdot 2 \left( x + \sum_{i=1}^{\infty} \frac{1}{2i+1} x^{2i+1} \right) - 1 \\&= \sum_{i=1}^{\infty} \frac{1}{2i+1} \frac{1}{(2n+1)^{2i}} \\&< \sum_{i=1}^{\infty} \frac{1}{3} \frac{1}{(2n+1)^{2i}} \\&= \frac{1}{3} \left( \frac{1}{1-q} - 1 \right) \quad \text{wobei} \quad q = \frac{1}{(2n+1)^2} \\&= \frac{1}{3((2n+1)^2 - 1)}\end{aligned}$$



# Beweis der Stirling-Formel (5)

Offenbar gilt auch

$$\frac{1}{3(2n+1)^2} = \sum_{i=1}^1 \frac{1}{2i+1} \cdot \frac{1}{(2n+1)^{2i}} < d_n - d_{n+1},$$

also

$$\frac{1}{3(2n+1)^2} < d_n - d_{n+1} < \frac{1}{3((2n+1)^2 - 1)}.$$

# Beweis der Stirling-Formel (6)

Abschätzung der Schranken

$$\begin{aligned}
 \frac{1}{3 \frac{((2n+1)^2-1)}{1}} &= \frac{1}{12n \frac{(n+1)}{1}} = \frac{1}{12n} - \frac{1}{12 \frac{(n+1)}{12}} \\
 \frac{1}{3(2n+1)^2} &= \frac{1}{12n(n+1) + 3} = \frac{1}{12(12n(n+1) + 3)} \\
 &= \frac{1}{12 \cdot 12n(n+1) + 36} \\
 &> \frac{1}{12 \cdot 12n^2 + 12 \cdot 12n + 24n + 13} \\
 &= \frac{1}{12 \cdot 12n^2 + 12 \cdot 14n + 13} \\
 &= \frac{1}{(12n+1)(12n+13)} \\
 &= \frac{1}{12n+1} - \frac{1}{12(n+1) + 1}
 \end{aligned}$$

# Beweis der Stirling-Formel (7)

Beide Ungleichungen zusammen

$$\frac{1}{12n+1} - \frac{1}{12(n+1)+1} < d_n - d_{n+1} < \frac{1}{12n} - \frac{1}{12(n+1)}$$

$$\begin{aligned} (d_n - \frac{1}{12n}) - (d_{n+1} - \frac{1}{12(n+1)}) &< 0 < \\ (d_n - \frac{1}{12n+1}) - (d_{n+1} - \frac{1}{12(n+1)+1}) \end{aligned}$$

Folge  $\{d_n - \frac{1}{12n+1}\}$  ist monoton fallend

Folge  $\{d_n - \frac{1}{12n}\}$  ist monoton wachsend.

# Beweis der Stirling-Formel (8)

Beide Folgen haben denselben Grenzwert  $c := \lim d_n$ ,

$$\begin{aligned} d_n - \frac{1}{12n} &< c < d_n - \frac{1}{12n+1} \\ c + \frac{1}{12n+1} &< d_n < c + \frac{1}{12n} \end{aligned}$$

Erinnerung:

$$\begin{aligned} d_n &= \ln n! - \left(n + \frac{1}{2}\right) \ln n + n \\ \Rightarrow e^{d_n} &= n! \left(\frac{n}{e}\right)^{-n} n^{-\frac{1}{2}} \end{aligned}$$

# Beweis der Stirling-Formel (9)

$$\begin{aligned}
 e^{c+\frac{1}{12n+1}} &< e^{d_n} < e^{c+\frac{1}{12n}} \\
 e^c e^{\frac{1}{12n+1}} &< n! \left(\frac{n}{e}\right)^{-n} n^{-\frac{1}{2}} < e^c e^{\frac{1}{12n}} \\
 e^c \sqrt{n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} &< n! < e^c \sqrt{n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}
 \end{aligned}$$

Bleibt zu zeigen

$$e^c = \sqrt{2\pi}.$$

# Beweis der Stirling-Formel (10)

Hilfsrechnungen

$$\begin{aligned} I_n &:= \int_0^{\pi/2} \sin^n x \, dx \\ I_n &= \int_0^{\pi/2} \sin^{n-1} x \cdot \sin x \, dx \\ &= \sin^{n-1} x \cdot (-\cos x) \Big|_0^{\pi/2} - \\ &\quad \int_0^{\pi/2} (n-1) \sin^{n-2} x \cos x \cdot (-\cos x) \, dx \\ &= (n-1) \int_0^{\pi/2} \sin^{n-2} x (1 - \sin^2 x) \, dx \\ &= (n-1)(I_{n-2} - I_n) \\ I_n &= \frac{n-1}{n} I_{n-2} \end{aligned}$$

# Beweis der Stirling-Formel (11)

Hilfsrechnungen (Fortsetzung, 1)

$$\begin{aligned}l_0 &= \frac{\pi}{2} & l_1 &= 1 \\l_2 &= \frac{1}{2}l_0 = \frac{1}{2} \cdot \frac{\pi}{2} & l_3 &= \frac{2}{3}l_1 = \frac{2}{3}\end{aligned}$$

$$\begin{aligned}l_{2n} &= \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} \frac{\pi}{2} \\l_{2n+1} &= \frac{2 \cdot 4 \cdot 6 \cdots (2n)}{3 \cdot 5 \cdot 7 \cdots (2n+1)}\end{aligned}$$

# Beweis der Stirling-Formel (12)

Hilfsrechnungen (Fortsetzung, 2)

$$0 < x < \frac{\pi}{2}$$

$$\Rightarrow 0 < \sin x < 1$$

$$\Rightarrow \sin^{2n-1} x > \sin^{2n} x > \sin^{2n+1} x$$

$$\Rightarrow I_{2n-1} > I_{2n} > I_{2n+1}$$

$$\Rightarrow \frac{I_{2n-1}}{I_{2n+1}} > \frac{I_{2n}}{I_{2n+1}} > 1$$

$$\Rightarrow \frac{2n+1}{2n} > \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n-1) \cdot (2n+1)}{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdots (2n) \cdot (2n)} \cdot \frac{\pi}{2} > 1$$

$$\Rightarrow \lim \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2n-1) \cdot (2n+1)}{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6 \cdots (2n) \cdot (2n)} \cdot \frac{\pi}{2} = 1$$

$$\begin{aligned} \Rightarrow \frac{\pi}{2} &= \lim \left( \frac{2 \cdot 4 \cdot 6 \cdots (2n)}{1 \cdot 3 \cdot 5 \cdots (2n-1)} \right)^2 \cdot \frac{1}{2n+1} \\ &= \lim \frac{2^{4n} (n!)^4}{((2n)!)^2 (2n+1)} \end{aligned}$$



# Beweis der Stirling-Formel (13)

$$n! = e^c \sqrt{nn^n} e^{-n} e^{\alpha_n}$$

$$(2n)! = e^c \sqrt{2n} 2^{2n} n^{2n} e^{-2n} e^{\beta_n}$$

wobei  $\lim_{n \rightarrow \infty} \alpha_n = \lim_{n \rightarrow \infty} \beta_n = 0$ .

Einsetzen oben liefert

$$e^c = \sqrt{2\pi}.$$

# Inhalt

- Einführung
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes
- Anwendung bedingter Wahrscheinlichkeiten

### 3-maliges Werfen einer Münze

Menge der Elementarereignisse:

$$\Omega = \{zzz, zzw, zwz, wzz, zww, wz w, wwz, www\}. |\Omega| = 2^3 = 8 = N$$

Wir definieren zwei Ereignisse:

**A:** Das Wappen fällt genau einmal, d.h.

$$A = \{zzw, zwz, wzz\}. \quad P(A) = \frac{n(A)}{N} = \frac{3}{8}.$$

**B:** # Wappenwürfe ungerade, d.h.:

$$B = \{zzw, zwz, wzz, www\}. \quad P(B) = \frac{n(B)}{N} = \frac{4}{8} = \frac{1}{2}.$$

Offenbar  $A \subset B$ .

## 3-maliges Werfen einer Münze (Fortsetzung)

Angenommen, Ereignis  $B$  sei bereits eingetreten.

Wahrscheinlichkeit, daß unter dieser Bedingung das Ereignis  $A$  eintritt?

Bei diesem Experiment ist die Menge der Elementarereignisse die Menge  $B$ . Damit gilt  $N = 4$ . Folglich erhalten wir:

$$P(A, \text{ falls } B \text{ bereits eingetreten ist}) = P(A/B) = \frac{3}{4}.$$

# Bedingte Wahrscheinlichkeit

## Einführung (2)

### Def. 11 (Bedingte Wahrscheinlichkeit)

Es seien  $A, B \in \mathcal{E}$  zwei zufällige Ereignisse und es gelte  $P(B) > 0$ .

Dann wird

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

als bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$  bezeichnet.

**Bem.:** Oft wird auch die folgende Bezeichnung verwendet:

$$P_B(A) := P(A/B).$$

# Bedingte Wahrscheinlichkeit

## Einführung (3)

**Bem.:** Wir unterscheiden folgende Fälle:

①  $A \supseteq B$ : Dann gilt:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

②  $A \subseteq B$ : Dann gilt:

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

③  $A \cap B \neq \emptyset$  (teilweise Überschneidung):

Dann gilt:

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

# Unabhängigkeit

## Definition

### Def. 12 (Unabhängigkeit)

Zwei Ereignisse  $A, B \in \mathcal{E}$  heißen  
unabhängig, wenn gilt:

$$P(A/B) = P(A).$$

**Bem.:** Für zwei unabhängige Ereignisse gilt:

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

# Unabhängigkeit

## Beispiel

### Skatspiel mit 32 Karten

Daraus wird eine Karte gezogen. ( $N = |\Omega| = 32$ ).

Wir betrachten die zufälligen Ereignisse:

**A:** Ziehen eines Königs.

$$P(A) = \frac{n(A)}{N} = \frac{4}{32} = \frac{1}{8}.$$

**B:** Ziehen einer Herzkarte.

$$P(B) = \frac{n(B)}{N} = \frac{8}{32} = \frac{1}{4}.$$

Sind diese beiden Ereignisse voneinander unabhängig?



# Unabhängigkeit

## Beispiel (Fortsetzung)

### Skatspiel mit 32 Karten, Fortsetzung

Offenbar  $P(B) > 0$ . Es sei eine Herzkarte gezogen worden (Ereignis  $B$  also eingetreten). Wahrscheinlichkeit, daß dann der Herzkönig gezogen wurde:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{8} = P(A).$$

Folglich sind nach Definition die Ereignisse  $A$  und  $B$  voneinander unabhängig.

# $P_B$ ist Wahrscheinlichkeit

## Satz:

Es seien  $A, B \in \mathcal{E}$  zwei Ereignisse, wobei  $P(B) > 0$  gelte. Dann genügt die bedingte Wahrscheinlichkeit  $P_B$  den KOLMOGOROV-Axiomen. D.h. das Tripel  $(\Omega, \mathcal{E}, P_B)$  ist ein Wahrscheinlichkeitsraum.

# $P_B$ ist Wahrscheinlichkeit

## Satz:

Es seien  $A, B \in \mathcal{E}$  zwei Ereignisse, wobei  $P(B) > 0$  gelte. Dann genügt die bedingte Wahrscheinlichkeit  $P_B$  den KOLMOGOROV-Axiomen. D.h. das Tripel  $(\Omega, \mathcal{E}, P_B)$  ist ein Wahrscheinlichkeitsraum.

**Beweis:** Wir zeigen stellvertretend Axiom 2. Es gilt:

$$\begin{aligned} P_B(\Omega) &= P(\Omega/B) \\ &= \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \end{aligned}$$

Die anderen beiden Axiome (vgl. Definition 7) sind ebenfalls erfüllt.  $\square$

# Bedingte Wahrscheinlichkeit

## Satz

Es seien  $A, B, C \in \mathcal{E}$  drei Ereignisse. Dann gilt:

$$P_B(A/C) = P(A/B \cap C).$$

# Bedingte Wahrscheinlichkeit

## Satz

Es seien  $A, B, C \in \mathcal{E}$  drei Ereignisse. Dann gilt:

$$P_B(A/C) = P(A/B \cap C).$$

**Beweis:** Es gilt:

$$\begin{aligned} P_B(A/C) &= \frac{P_B(A \cap C)}{P_B(C)} \\ &= \frac{P(A \cap C/B)}{P(C/B)} \\ &= \frac{P(A \cap B \cap C) \cdot P(B)}{P(B) \cdot P(B \cap C)} \\ &= \frac{P(A \cap B \cap C)}{P(B \cap C)} = P(A/B \cap C) \end{aligned}$$

# Unabhängigkeit

Fortsetzung (1)

## Lemma

Es seien  $A, B \in \mathcal{E}$  zwei unabhängige Ereignisse. Dann sind die Ereignisse  $A$  und  $\bar{B}$  ebenfalls unabhängig. Gleiches gilt für die Ereignisse  $\bar{A}$  und  $B$  sowie für  $\bar{A}$  und  $\bar{B}$ .

**Beweis:** Wir zeigen die Aussage am Beispiel der Ereignisse  $A$  und  $\bar{B}$ .

Es gilt:

# Unabhängigkeit

Fortsetzung (2)

## Beweis des Lemma, Fortsetzung

$$\begin{aligned}P(A/\overline{B}) &= \frac{P(A \cap \overline{B})}{P(\overline{B})} \\&= \frac{P(A \setminus (A \cap B))}{1 - P(B)} \quad (\text{Folgerung ??.1))} \\&= \frac{P(A) - P(A \cap B)}{1 - P(B)} \quad (\text{Folgerung ??.3b))} \\&= \frac{P(A) - P(A)P(B)}{1 - P(B)} \\&= \frac{P(A)(1 - P(B))}{1 - P(B)} = P(A)\end{aligned}$$

# Unabhängigkeit

Fortsetzung (3)

## Beweis des Lemma, Fortsetzung

Zusammenfassend gilt

$$P(A/B) = P(A) \iff P(A/\bar{B}) = P(A)$$

$$\iff P(\bar{A}/\bar{B}) = P(\bar{A})$$

$$\iff P(\bar{A}/B) = P(\bar{A})$$



# Inhalt

- Einführung
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes
- Anwendung bedingter Wahrscheinlichkeiten

### Def. 13 (Vollständigkeit)

Es sei  $(\Omega, \mathcal{E}, P)$  ein Wahrscheinlichkeitsraum. Eine Folge von Ereignissen

$$\{A_n\}_{n=1}^{\infty} \quad (A_n \in \mathcal{E}, \forall n \in \mathbb{N})$$

heißt vollständig (oder ausschöpfend), falls folgende Bedingungen erfüllt sind:

- ①  $\bigcup_{n=1}^{\infty} A_n = \Omega;$
- ②  $A_i \cap A_j = \emptyset$ , für alle  $i \neq j$ .

# Satz der Totalen Wahrscheinlichkeit

## Satz

Es sei  $A_1, A_2, \dots$  eine vollständige Folge von Ereignissen. Weiterhin sei  $B$  ein beliebiges Ereignis und es gelte  $P(A_i) \neq 0$  für alle  $i$ . Dann gilt:

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

Dieser Ausdruck heißt

Formel der totalen Wahrscheinlichkeit.

# Satz der Totalen Wahrscheinlichkeit

Es sei  $A_1, A_2, \dots$  eine vollständige Folge von Ereignissen. Weiterhin sei  $B$  ein beliebiges Ereignis und es gelte  $P(A_i) \neq 0$  für alle  $i$ . Dann gilt:

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

**Beweis:** Aus  $B = B \cap (\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} (B \cap A_i)$  folgt (da die  $(B \cap A_i)$  ebenfalls unvereinbar sind):

# Satz der Totalen Wahrscheinlichkeit

Es sei  $A_1, A_2, \dots$  eine vollständige Folge von Ereignissen. Weiterhin sei  $B$  ein beliebiges Ereignis und es gelte  $P(A_i) \neq 0$  für alle  $i$ . Dann gilt:

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

**Beweis:** Aus  $B = B \cap (\bigcup_{i=1}^{\infty} A_i) = \bigcup_{i=1}^{\infty} (B \cap A_i)$  folgt (da die  $(B \cap A_i)$  ebenfalls unvereinbar sind):

$$P(B) = P\left(\bigcup_{i=1}^{\infty} (B \cap A_i)\right) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$$



# Satz der Totalen Wahrscheinlichkeit

## Beispiel

### Binärkanal

Bei der Übertragung auf einem binären Kanal kommen die Zeichen '0' und '1' im Verhältnis 3:4 vor.

Ein '0' wird mit Wahrscheinlichkeit von 0.2 fehlerhaft übertragen

Ein '1' wird mit Wahrscheinlichkeit von 0.3 fehlerhaft übertragen

gesucht:

- Wahrscheinlichkeit für eine fehlerhafte Übertragung?
- Wahrscheinlichkeit, dass ein '0' empfangen wird?

Ereignisse:

# Satz der Totalen Wahrscheinlichkeit

## Beispiel

### Binärkanal

Bei der Übertragung auf einem binären Kanal kommen die Zeichen '0' und '1' im Verhältnis 3:4 vor.

Ein '0' wird mit Wahrscheinlichkeit von 0.2 fehlerhaft übertragen

Ein '1' wird mit Wahrscheinlichkeit von 0.3 fehlerhaft übertragen

gesucht:

- Wahrscheinlichkeit für eine fehlerhafte Übertragung?
- Wahrscheinlichkeit, dass ein '0' empfangen wird?

Ereignisse:

- $S_0$ : '0' wird gesendet,  $P(S_0) = \frac{3}{7}$
- $E_0$ : '0' wird empfangen,
- $S_1$ : '1' wird gesendet,  $P(S_1) = \frac{4}{7}$
- $E_1$ : '1' wird empfangen

# Satz der Totalen Wahrscheinlichkeit

## Beispiel

### Binärkanal, Fortsetzung

$$P(E_1|S_0) = 0.2, \quad P(E_0|S_1) = 0.3$$

$F$ : Ereignis, das ein Übertragungsfehler vorliegt

$$\begin{aligned} P(F) &= P(E_1, S_0) + P(E_0, S_1) \\ &= P(E_1|S_0) \cdot P(S_0) + P(E_0|S_1) \cdot P(S_1) \\ &= \frac{1}{5} \cdot \frac{3}{7} + \frac{3}{10} \cdot \frac{4}{7} = \frac{18}{70} \approx 0.2571 \\ P(E_0) &= P(E_0|S_0) \cdot P(S_0) + P(E_0|S_1) \cdot P(S_1) \\ &= \frac{8}{10} \cdot \frac{3}{7} + \frac{3}{10} \cdot \frac{4}{7} = \frac{18}{35} \approx 0.5143 \end{aligned}$$



# Inhalt

- Einführung
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes
- Anwendung bedingter Wahrscheinlichkeiten

Gegeben:  $P(A_i)$  und  $P(A/A_i)$ , ( $i \in \mathbb{N}$ ).

Gesucht:  $P(A_i/A)$ .

Unter Benutzung der Definition der bedingten Wahrscheinlichkeit und der Formel für die totale Wahrscheinlichkeit erhalten wir:

# Satz von Bayes

$$\begin{aligned}P(A_i/A) &= \frac{P(A_i \cap A)}{P(A)} \\&= \frac{P(A_i) \cdot P(A/A_i)}{P(A)}\end{aligned}$$

Wenden die Formel der totalen Wahrscheinlichkeit an,

Satz von BAYES, Formel von BAYES

$$P(A_i/A) = \frac{P(A_i) \cdot P(A/A_i)}{\sum_{j=1}^{\infty} (P(A/A_j) \cdot P(A_j))}$$

# Satz von Bayes

## Beispiel

### Binärkanal, Fortsetzung

$$\begin{aligned}P(S_0|E_0) &= \frac{P(E_0|S_0)P(S_0)}{P(E_0|S_0)P(S_0) + P(E_0|S_1)P(S_1)} \\&= \frac{\frac{8}{10} \cdot \frac{3}{7}}{\frac{8}{10} \cdot \frac{3}{7} + \frac{3}{10} \cdot \frac{4}{7}} = \frac{24}{24 + 12} = \frac{2}{3}\end{aligned}$$

$$\begin{aligned}P(S_1|E_1) &= \frac{P(E_1|S_1)P(S_1)}{P(E_1|S_0)P(S_0) + P(E_1|S_1)P(S_1)} \\&= \frac{\frac{7}{10} \cdot \frac{4}{7}}{\frac{2}{10} \cdot \frac{3}{7} + \frac{7}{10} \cdot \frac{4}{7}} = \frac{28}{28 + 6} = \frac{14}{17}\end{aligned}$$

# Inhalt

- Einführung
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes
- Anwendung bedingter Wahrscheinlichkeiten

# Expertensystem

## Aufbau der Wissensbasis:

$K_i$  – bestimmte Ereignisse (z.B. Krankheiten)

$P_0(K_i)$  – a-priori-Wahrscheinlichkeit für  $K_i$

$S_j$  – bestimmte Symptome

$P(S/K)$  – Wkt für Symptom  $S$ , falls  $K$  vorliegt

$P(S/\overline{K})$  – Wkt für Symptom  $S$ , falls  $K$  nicht vorliegt

# Expertensystem (2)

## “Inferenzmaschine”

$$P(K|S) = \frac{P(S|K) \cdot P(K)}{P(S)}$$

$$P(K|\bar{S}) = \frac{P(\bar{S}|K) \cdot P(K)}{P(\bar{S})}$$

$$P(S) = P(S|K) \cdot P(K) + P(S|\bar{K}) \cdot P(\bar{K})$$

$$P(\bar{S}) = 1 - P(S)$$

## Expertensystem (3)

### Arbeitsweise:

Krankheiten  $K_1, \dots, K_K$

Symptome  $S_1, \dots, S_S$

$I_0 = \{1, \dots, K\}$  Indexmenge der möglichen Krankheiten (wird laufend aktualisiert)

$J = \{1, \dots, S\}$  Indexmenge der Symptome

$l$ : laufender Index

$l = 0$ ; ärztliches (Basis-)Wissen

$P_0 = P$ ;  $P_0(S|K) = P(S|K), P_0(S|\overline{K}) = P(S|\overline{K})$



## Expertensystem (4)

### Berechnen der bedingten Wahrscheinlichkeiten

$\forall (i, j) \in I \times J$ :

$$P_I(S_j) := P_I(S_j|K_i) \cdot P_I(K_i) + P_I(S_j|\bar{K}_i) \cdot P_I(\bar{K}_i)$$

$$P_I(K_i|S_j) = \frac{P(S_j|K_i) \cdot P_I(K_i)}{P(S_j)}$$

$$P_I(K_i|\bar{S}_j) = \frac{P(\bar{S}_j|K_i) \cdot P_I(K_i)}{P_I(\bar{S}_j)}$$

$$P_{I+1}(S_j|K_i) = \frac{P_I(K_i|S_j) \cdot P_I(S_j)}{P_I(K_i|S_j) \cdot P_I(S_j) + P_I(K_i|\bar{S}_j)P_I(\bar{S}_j)}$$

# Expertensystem (5)

A. Bestimmen des Symptoms, das am besten die Menge der Krankheiten charakterisiert

$$r(j) := \sum_{i \in I_l} |P_l(K_i|S_j) - P(K_i|\bar{S}_j)| \quad \forall \quad j \in J;$$

$j_l := \operatorname{argmax}_{j \in J} r(j)$  das Symptom mit dem größten  $r(j)$ .

# Expertensystem (6)

## B. Frage an den Patienten nach Symptom $S_{j_i}$

$P(K_i)$  wird aktualisiert:

$$P_{I+1}(K_i) = \begin{cases} P_I(K_i|S_{j_i}) & \text{falls JA} \\ P_I(K_i|\bar{S}_{j_i}) & \text{falls NEIN} \\ P_I(K_i) & \text{falls WEIS NICHT} \end{cases}$$

# Expertensystem (7)

Aktualisieren der bedingten Wahrscheinlichkeiten.

$\forall (i, j) \in I \times J$ :

$$P_{l+1}(S_j) := P(S_j|K_i) \cdot P_{l+1}(K_i) + P(S_j|\bar{K}_i) \cdot P_{l+1}(\bar{K}_i)$$

$$P_{l+1}(K_i|S_j) := \frac{P(S_j|K_i) \cdot P_{l+1}(K_i)}{P(S_j)}$$

$$P_{l+1}(K_i|\bar{S}_j) := \frac{P(\bar{S}_j|K_i) \cdot P_{l+1}(K_i)}{P(\bar{S}_j)}$$

# Expertensystem (8)

C: Bestimmen des Symptoms, das am besten die Krankheit  $i$  charakterisiert

$$m_i := \max_{j \in J} |P_{I+1}(K_i|S_j) - P_{I+1}(K_i|\bar{S}_j)|, \quad \forall i \in I_I$$

## Expertensystem (9)

Krankheiten mit zu kleinen Abständen werden aus der Indexmenge entfernt.

Symptom  $j_l$  ist abgearbeitet.

$$I_{l+1} = I_l \setminus \{i \in I_l : m_i < c\}$$

$$J_{l+1} = J_l \setminus \{j_l\};$$

$$l := l + 1;$$

Abbruchbedingung nicht erfüllt: goto A.

Abbruchbedingung, z.B.

$$I_l = I_{l+1}, S_{j_l} = S_{j_{l+1}}, I_{l+1} = \{i\} \text{ oder } J_{l+1} = \emptyset$$

end.

# Ein-Prozessorsystem mit I/O-Einheit

## Langzeitverhalten eines Ein-Prozessorsystems mit einer I/O-Einheit

Wir betrachten ein Ein-Prozessorsystem, das auf folgende Weise arbeiten soll: Wenn ein Programm beendet wird, so wird mit Wahrscheinlichkeit  $p$  ( $0 < p < 1$ ) die I/O-Einheit aktiviert, und mit Wahrscheinlichkeit  $q = 1 - p$  erfolgt ein erneuter Programmstart. Nach Beendigung eines I/O-Vorgangs wird immer ein neues Programm gestartet.

# Ein-Prozessorsystem mit I/O-Einheit

(2)

Frage: Mit welcher Wahrscheinlichkeit befindet sich das System im  $n$ -ten Zyklus im Programmzustand?

Wir legen fest ( $n = 1, 2, 3, \dots$ ):

$A_n$  - Ereignis, daß im  $n$ -ten Zyklus ein Programm startet

$\overline{A_n}$  - Ereignis, daß im  $n$ -ten Zyklus die I/O-Einheit aktiviert wird

gesucht:  $P(A_n)$  . Langzeitverhalten ( $\lim_{n \rightarrow \infty} P(A_n)$ ).



# Ein-Prozessorsystem mit I/O-Einheit

(3)

$P(A_1) = 1$ , denn es wird beim Einschalten des Systems immer mit einem Programm begonnen.

Aus der angegebenen Beschreibung der Arbeitsweise des Systems folgt:

$$P(A_{n+1}/A_n) = q = 1 - p$$

$$P(\overline{A_{n+1}}/A_n) = p$$

$$P(\overline{A_{n+1}}/\overline{A_n}) = 0$$

$$P(A_{n+1}/\overline{A_n}) = 1$$

$q_n := P(A_n)$ . Die ersten drei Werte sind:

# Einprozessorsystem mit I/O-Einheit

(4)

$$q_1 = P(A_1) = 1$$

$$q_2 = P(A_2)$$

$$= P(A_2/A_1) \cdot P(A_1) + \underbrace{P(A_2/\overline{A_1}) \cdot P(\overline{A_1})}_{=0} \quad \text{totale W.}$$

$$= q = 1 - p$$

$$q_3 = P(A_3)$$

$$= P(A_3/A_2) \cdot P(A_2) + P(A_3/\overline{A_2}) \cdot P(\overline{A_2})$$

$$= q \cdot q + 1 \cdot (1 - q) = (1 - p)^2 + p = 1 - p + p^2$$

# Einprozessorsystem mit I/O–Einheit

(5)

Vermutung:

$$q_n = P(A_n) = \sum_{i=0}^{n-1} (-p)^i.$$

# Einprozessorsystem mit I/O-Einheit

(5)

Vermutung:

$$q_n = P(A_n) = \sum_{i=0}^{n-1} (-p)^i.$$

**Beweis:** (vollständige Induktion):

**IA:** Es sei  $n = 1$ :  $q_1 = 1$ .

**IS:** Wir nehmen an, daß die Formel für  $n$  gilt. Wir zeigen die Gültigkeit für  $n + 1$ :

# Einprozessorsystem mit I/O-Einheit

(6)

$$\begin{aligned}q_{n+1} &= P(A_{n+1}) \\&= P(A_{n+1}/A_n) \cdot P(A_n) + P(A_{n+1}/\overline{A_n}) \cdot P(\overline{A_n}) \\&= q \cdot q_n + 1 \cdot (1 - q_n) = 1 + (q - 1) \cdot q_n \\&= 1 - p \cdot q_n \\&= 1 - p \cdot \sum_{i=0}^{n-1} (-p)^i \quad (\text{nach IV}) \\&= 1 + \sum_{i=1}^n (-p)^i = \sum_{i=0}^n (-p)^i\end{aligned}$$

# Einprozessorsystem I/O-Einheit

(7)

Untersuchen wir noch das Langzeitverhalten:

$$\begin{aligned}\lim_{n \rightarrow \infty} P(A_n) &= \lim_{n \rightarrow \infty} q_n \\ &= \sum_{i=0}^{\infty} (-p)^i \\ &= \frac{1}{1 - (-p)} = \frac{1}{1 + p},\end{aligned}$$

geometrische Reihe mit  $|-p| < 1$ .

Frage: Sind die Ereignisse  $A_{n+1}$  und  $A_n$  unabhängig?

# Einprozessorsystem I/O-Einheit

(8)

Sind die Ereignisse  $A_{n+1}$  und  $A_n$  unabhängig?

$$\begin{aligned}P(A_{n+1} \cap A_n) &= P(A_{n+1}/A_n) \cdot P(A_n) \\ &= q \cdot q_n\end{aligned}$$

Angenommen, die beiden Ereignisse seien unabhängig,

$$\begin{aligned}P(A_{n+1}/A_n) &= P(A_{n+1}) \\ q &= q_{n+1}\end{aligned}$$

# Einprozessorsystem I/O-Einheit

(8)

Sind die Ereignisse  $A_{n+1}$  und  $A_n$  unabhängig?

$$\begin{aligned}P(A_{n+1} \cap A_n) &= P(A_{n+1}/A_n) \cdot P(A_n) \\&= q \cdot q_n\end{aligned}$$

Angenommen, die beiden Ereignisse seien unabhängig,

$$\begin{aligned}P(A_{n+1}/A_n) &= P(A_{n+1}) \\q &= q_{n+1}\end{aligned}$$

Aber, für  $n \geq 2$  gilt  $q \neq q_{n+1}$ .

Also sind die Ereignisse  $A_n$  und  $A_{n+1}$  nicht unabhängig.



# Einprozessorsystem I/O–Einheit

(9)

Der gesamte Ablauf läßt sich eindeutig in Matrixform darstellen:

|     | I/O | A       |
|-----|-----|---------|
| I/O | 0   | 1       |
| A   | $p$ | $1 - p$ |

# Weitere Anwendungen (1)

## Zuverlässigkeitstheorie

Wir betrachten ein Reihen-System mit 2 Bauteilen, die unabhängig voneinander ausfallen,

Angenommen, das System fällt (innerhalb eines bestimmten Zeitraumes) aus.

Wie groß ist Wahrscheinlichkeit, dass genau das erste Bauteil ausgefallen ist?

# Zuverlässigkeitstheorie

Beispiel, Fortsetzung

$A_i$ : Ereignis, dass Bauteil  $i$  ausfällt.

geg.:  $P(A_i) = p_i, i = 1, 2$

ges.:

# Zuverlässigkeitstheorie

Beispiel, Fortsetzung

$A_i$ : Ereignis, dass Bauteil  $i$  ausfällt.

geg.:  $P(A_i) = p_i, i = 1, 2$

ges.:  $P(A_1 \cap \bar{A}_2 | A_1 \cup A_2)$ ?

$$\begin{aligned} P(A_1 \cap \bar{A}_2 | A_1 \cup A_2) &= \frac{P((A_1 \cap \bar{A}_2) \cap (A_1 \cup A_2))}{P(A_1 \cup A_2)} \\ &= \frac{P(A_1 \cap \bar{A}_2)}{P(A_1 \cup A_2)} \\ &= \frac{P(A_1) \cdot P(\bar{A}_2)}{P(A_1) + P(A_2) - P(A_1 \cap A_2)} \quad \text{Unabh., Subtraktivität} \\ &= \frac{p_1(1 - p_2)}{p_1 + p_2 - p_1 p_2} \end{aligned}$$

# Zuverlässigkeitstheorie

## Beispiel, Fortsetzung 2

Analog

$$P(A_2 \cap \bar{A}_1 | A_1 \cup A_2) = \frac{p_2(1 - p_1)}{p_1 + p_2 - p_1 p_2}$$

Wahrscheinlichkeit für Ausfall beider Bauteile: ÜA

## Weitere Anwendungen (2)

### Münzwurf-Spiel

A und B spielen: Münze wird abwechselnd geworfen. Es gewinnt, wer zuerst Blatt hat.

$B$ : Ereignis, dass bei einem Wurf Blatt kommt

$Z$ : Ereignis, dass bei einem Wurf Zahl kommt

$E$ : Ereignis, dass A gewinnt

$F$ : Ereignis, dass B gewinnt

$G$ : Spiel endet nicht.

# Münzwurf-Spiel

(Fortsetzung)

## Münzwurf-Spiel (Fortsetzung)

$$P(E) = P(B) + P(ZZB) + P(ZZZZB) + \dots$$

# Münzwurf-Spiel

(Fortsetzung)

## Münzwurf-Spiel (Fortsetzung)

$$\begin{aligned}P(E) &= P(B) + P(ZZB) + P(ZZZB) + \dots \\&= \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \dots = \frac{1}{2} \sum_{i=0}^{\infty} \frac{1}{4^i} \\&= \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{2}{3}\end{aligned}$$

$$P(F) = P(ZB) + P(ZZZB) + P(ZZZZZB) + \dots$$



# Münzwurf-Spiel

(Fortsetzung)

## Münzwurf-Spiel (Fortsetzung)

$$\begin{aligned}P(E) &= P(B) + P(ZZB) + P(ZZZZB) + \dots \\&= \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \dots = \frac{1}{2} \sum_{i=0}^{\infty} \frac{1}{4^i} \\&= \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{2}{3}\end{aligned}$$

$$\begin{aligned}P(F) &= P(ZB) + P(ZZZB) + P(ZZZZZB) + \dots \\&= \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots \\&= \frac{1}{4} \sum_{i=0}^{\infty} \frac{1}{4^i} = \frac{1}{3}\end{aligned}$$

# Weitere Anwendungen

(Fortsetzung, 2)

## Münzwurf-Spiel (Fortsetzung)

oder (unter Anwendung der bedingten Wahrscheinlichkeiten.)

$$\begin{aligned}P(F) &= P(F|B) \cdot P(B) + P(F|Z) \cdot P(Z) \\&= 0 \cdot \frac{1}{2} + P(E) \cdot \frac{1}{2} \quad \text{2. wird 1. Spieler}\end{aligned}$$

$$\begin{aligned}P(E) &= P(E|B) \cdot P(B) + P(E|Z) \cdot P(Z) \\&= 1 \cdot \frac{1}{2} + P(F) \cdot \frac{1}{2}\end{aligned}$$

lineares Gleichungssystem lösen  $\rightarrow$  obiges Ergebnis.

## Weitere Anwendungen (3)

### Ruin des Spielers

Irrfahrt auf der Geraden mit 2 absorbierenden Zuständen, 0 und  $a + b$

$a$ : Startkapital Spieler A

$b$ : Startkapital Spieler B

Frage: Mit welcher Wahrscheinlichkeit wird Spieler A ruiniert?

## Weitere Anwendungen (3)

### Ruin des Spielers

Irrfahrt auf der Geraden mit 2 absorbierenden Zuständen, 0 und  $a + b$

$a$ : Startkapital Spieler A

$b$ : Startkapital Spieler B

Frage: Mit welcher Wahrscheinlichkeit wird Spieler A ruiniert?

$E_k$ : Ereignis, dass der Spieler, der  $k$  Euro besitzt, ruiniert wird,

$$p_k = P(E_k)$$

$A_{-1}$ : Ereignis, im nächsten Schritt einen Euro zu verlieren.

$A_{+1}$ : Ereignis, im nächsten Schritt einen Euro zu gewinnen.

# Ruin des Spielers

(Fortsetzung)

Nach dem Satz der Totalen Wahrscheinlichkeit gilt:

$$p_k = P(E_k|A_{-1}) \cdot P(A_{-1}) + P(E_k|A_{+1}) \cdot P(A_{+1})$$

# Ruin des Spielers

(Fortsetzung)

Nach dem Satz der Totalen Wahrscheinlichkeit gilt:

$$\begin{aligned} p_k &= P(E_k|A_{-1}) \cdot P(A_{-1}) + P(E_k|A_{+1}) \cdot P(A_{+1}) \\ &= \frac{1}{2}(p_{k-1} + p_{k+1}) \end{aligned}$$

# Ruin des Spielers

(Fortsetzung)

Nach dem Satz der Totalen Wahrscheinlichkeit gilt:

$$\begin{aligned} p_k &= P(E_k|A_{-1}) \cdot P(A_{-1}) + P(E_k|A_{+1}) \cdot P(A_{+1}) \\ &= \frac{1}{2}(p_{k-1} + p_{k+1}) \end{aligned}$$

Daraus folgt:

$$\begin{aligned} 2p_k &= p_{k+1} + p_{k-1} \\ p_{k+1} - p_k &= p_k - p_{k-1} =: d \end{aligned}$$

# Ruin des Spielers

(Fortsetzung, 2)

Offenbar:  $p_0 = 1$ ,  $p_{a+b} = 0$

$$p_k = \underbrace{p_k - p_{k-1}}_{=d} + p_{k-1} - + \cdots + \underbrace{p_1 - p_0}_{=d} + p_0$$



# Ruin des Spielers

(Fortsetzung, 2)

Offenbar:  $p_0 = 1$ ,  $p_{a+b} = 0$

$$\begin{aligned}
 p_k &= \underbrace{p_k - p_{k-1}}_{=d} + p_{k-1} - + \cdots + \underbrace{p_1 - p_0}_{=d} + p_0 \\
 &= kd + 1
 \end{aligned}$$

# Ruin des Spielers

(Fortsetzung, 2)

Offenbar:  $p_0 = 1$ ,  $p_{a+b} = 0$

$$\begin{aligned} p_k &= \underbrace{p_k - p_{k-1}}_{=d} + p_{k-1} - + \cdots + \underbrace{p_1 - p_0}_{=d} + p_0 \\ &= kd + 1 \end{aligned}$$

$$p_{a+b} = (a+b)d + 1 = 0 \Rightarrow d = -\frac{1}{a+b}$$

$$p_k = 1 - \frac{k}{a+b}$$

$$p_a = 1 - \frac{a}{a+b} = \frac{b}{a+b}$$

$$p_b = \frac{a}{a+b}$$

# Inhalt

- Binomiale Wahrscheinlichkeiten
- Multinomiale Wahrscheinlichkeiten
- Poisson-Wahrscheinlichkeiten

Versuche mit zwei möglichen Ausgängen:

$A$  (gut) und  $\bar{A}$  (schlecht).

$$\Omega = \{A, \bar{A}\} = \{\text{„gut“}, \text{„schlecht“}\}$$

$$\mathcal{E} = \{\emptyset, A, \bar{A}, \Omega\}$$

$$P(A) = p$$

$$P(\bar{A}) = q = 1 - p$$

### Beispiele

Münzwurf:  $p = \frac{1}{2}$

Würfeln:  $p = \frac{1}{6}$

Qualitätskontrolle:  $p \cdot 100\%$  die Ausschußquote.

# Binomiale Wahrscheinlichkeiten (2)

## 2-malige Durchführung (unabhängig voneinander)

Elementarereignisse:  $(A, A), (A, \bar{A}), (\bar{A}, A), (\bar{A}, \bar{A})$  mit den Wahrscheinlichkeiten

$$P((A, A)) = p^2$$

$$P((A, \bar{A})) = p \cdot (1 - p)$$

$$P((\bar{A}, A)) = p \cdot (1 - p)$$

$$P((\bar{A}, \bar{A})) = (1 - p)^2$$

# Binomiale Wahrscheinlichkeiten

(Zweifaches Bernoulli-Schema)

$B_k$ : Ereignis, daß  $A$   $k$ -mal auftritt, wobei  $k = 0, 1, 2$ .

$$P(B_0) = (1 - p)^2$$

$$P(B_1) = 2 \cdot (p \cdot (1 - p))$$

$$P(B_2) = p^2$$

# Binomiale Wahrscheinlichkeiten

(Zweifaches Bernoulli-Schema)

$B_k$ : Ereignis, daß  $A$   $k$ -mal auftritt, wobei  $k = 0, 1, 2$ .

$$P(B_0) = (1 - p)^2$$

$$P(B_1) = 2 \cdot (p \cdot (1 - p))$$

$$P(B_2) = p^2$$

bzw.

$$P(B_k) = \binom{2}{k} p^k (1 - p)^{2-k}.$$

# Binomiale Wahrscheinlichkeiten

( $n$ -faches Bernoulli-Schema)

## Def. 14 (Binomialwahrscheinlichkeit)

Führen das Experiment jetzt  $n$ -mal unabhängig voneinander durch. Analog zum vorigen Experiment sei jetzt  $B_k$  das Ereignis, daß  $A$  genau  $k$ -mal auftritt,  $k = 0, \dots, n$

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Formel für das  $n$ -fache BERNOULLI-Schema.

Bezeichnung:  $B(p, n)$  oder auch  $Bi(p, n)$

Die Wahrscheinlichkeiten  $P(B_k)$  bezeichnen wir auch als  
Binomialwahrscheinlichkeiten.



## $n$ -faches Bernoulli-Schema (2)

Offenbar:

$$\begin{aligned}\sum_{i=0}^n P(B_i) &= \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} \\ &= (p + 1 - p)^n = 1\end{aligned}$$

# Binomiale Wahrscheinlichkeiten

## Beispiel

### Fünfmal eine Münze werfen

$A$ : das Ereignis, daß bei einem Wurf „Zahl“ fällt,  $P(A) = p = \frac{1}{2}$

$B_3$ : Ereignis, daß  $A$  genau dreimal auftritt:

# Binomiale Wahrscheinlichkeiten

## Beispiel

### Fünfmal eine Münze werfen

$A$ : das Ereignis, daß bei einem Wurf „Zahl“ fällt,  $P(A) = p = \frac{1}{2}$

$B_3$ : Ereignis, daß  $A$  genau dreimal auftritt:

$$\begin{aligned}P(B_3) &= \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{5-3} \\&= \binom{5}{3} \left(\frac{1}{2}\right)^5 \\&= \frac{5}{16}.\end{aligned}$$

# Inhalt

- Binomiale Wahrscheinlichkeiten
- Multinomiale Wahrscheinlichkeiten
- Poisson-Wahrscheinlichkeiten

Wir betrachten ein zufälliges Experiment mit den Ausgängen  $A_1, A_2, \dots, A_l$ . Wir setzen  $p_i = P(A_i)$ ,  $\sum_{i=1}^l p_i = 1$ .

Es sei ein Behälter mit  $k$  Kugeln in  $l$  verschiedenen Farben gegeben, wobei  $k_i$  Kugeln die Farbe  $i$  ( $i = 1, \dots, l$ ) besitzen,  $\sum_{i=1}^l k_i = k$ . Wahrscheinlichkeit, mit der eine Kugel einer bestimmten Farbe aus dem Behälter entnommen wird:

Wir betrachten ein zufälliges Experiment mit den Ausgängen  $A_1, A_2, \dots, A_l$ . Wir setzen  $p_i = P(A_i)$ ,  $\sum_{i=1}^l p_i = 1$ .

Es sei ein Behälter mit  $k$  Kugeln in  $l$  verschiedenen Farben gegeben, wobei  $k_i$  Kugeln die Farbe  $i$  ( $i = 1, \dots, l$ ) besitzen,  $\sum_{i=1}^l k_i = k$ .

Wahrscheinlichkeit, mit der eine Kugel einer bestimmten Farbe aus dem Behälter entnommen wird:

$$P(\text{Kugel der Farbe } i) = p_i = \frac{k_i}{k}.$$

## Multinomiale Wahrscheinlichkeiten (2)

### Def. 15 (Multinomiale Wahrscheinlichkeit)

Das Experiment soll nun  $n$ -mal wiederholt werden.  $B_{n_1, n_2, \dots, n_I}$ : das Ereignis, daß die Ereignisse  $A_1$   $n_1$ -mal,  $A_2$   $n_2$ -mal,  $\dots$ , und  $A_I$   $n_I$ -mal eintreten.

$$P(B_{n_1, n_2, \dots, n_I}) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_I!} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdot \dots \cdot p_I^{n_I}.$$

Derartige Wahrscheinlichkeiten bezeichnen wir auch als multinomiale Wahrscheinlichkeiten (polynomiale Wktn.)

# Potenzen von Summen

Vergleichen Sie:

$$(a_1 + \dots + a_l)^n = \sum \frac{n!}{n_1! \dots n_l!} a_1^{n_1} \dots a_l^{n_l}$$

wobei die Summe über alle Tupel  $(n_1, \dots, n_l)$  gebildet wird mit  $\sum_{i=1}^l n_i = n$ .



# Multinomiale Wahrscheinlichkeiten

## Beispiel

### Fragebogen

Bei einem Fragebogen wird (u.a.) nach dem Alter der befragten Personen gefragt. Das Alter sei in Klassen eingeteilt, 10-20, 21-40, 41-60, über 60 Jahre. Der Bevölkerungsanteil beträgt jeweils  $p_i$  für die  $i$ -te Altersklasse,  $i = 1, \dots, 4$ ,  $\sum_i p_i = 1$ .

Es werden  $n=1000$  Personen befragt.

Wie groß ist die Wahrscheinlichkeit, dass höchstens 10% der befragten bis zu 20 Jahre, und außerdem bis zu 10% der Befragten älter als 60 Jahre alt waren?

# Multinomiale Wahrscheinlichkeiten

## Beispiel, Fortsetzung

Sei  $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$ , wobei

$X_{ij} = 1$  falls Person  $i$  zur  $j$ -ten Altersklasse gehört,

und  $X_{ij} = 0$  sonst. Dann ist

$$\mathbf{Y} = \sum_{i=1}^n \mathbf{X}_i =: (Y_1, \dots, Y_4) \sim \text{Mult}(n, p_1, p_2, p_3, p_4)$$

# Multinomiale Wahrscheinlichkeiten

Beispiel, Fortsetzung

Sei  $a := 100$

$$P(Y_1, Y_4 \leq a) =$$

$$= P(Y_1 \leq a, Y_2 + Y_3 = n - Y_1 - Y_4, Y_4 \leq a)$$

# Multinomiale Wahrscheinlichkeiten

Beispiel, Fortsetzung

Sei  $a := 100$

$$P(Y_1, Y_4 \leq a) =$$

$$\begin{aligned} &= P(Y_1 \leq a, Y_2 + Y_3 = n - Y_1 - Y_4, Y_4 \leq a) \\ &= \sum_{i=0}^a \sum_{j=0}^a P(Y_1 = i, Y_2 + Y_3 = n - i - j, Y_4 = j) \\ &= \sum_{i=0}^a \sum_{j=0}^a \frac{n!}{i!j!(n-i-j)!} p_1^i p_4^j (p_2 + p_3)^{n-i-j} \end{aligned}$$

# Inhalt

- Binomiale Wahrscheinlichkeiten
- Multinomiale Wahrscheinlichkeiten
- Poisson-Wahrscheinlichkeiten

Beispiele, bei denen POISSON-Wahrscheinlichkeiten auftreten, sind

- die Anzahl von Verkehrsunfällen in einem Ort in einem bestimmten Zeitintervall,
- die Ankünfte von Kunden an einem Schalter oder
- der radioaktive Zerfall von  $\alpha$ -Teilchen.
- In einer Telefonzentrale wird ermittelt, wieviel Anrufe in einer bestimmten Zeiteinheit ankommen.

# Poisson-Wahrscheinlichkeiten

Elementarereignisse sind hier Anzahlen, z.B. das Ereignis, dass in einer Zeiteinheit genau  $i$  Anrufe eintreffen.

Def. 16 (Poisson-Wahrscheinlichkeit)

$$P(\omega_i) = \frac{\lambda^i}{i!} e^{-\lambda}.$$

$\lambda$  ist dabei ein noch unbestimmter Parameter. Er kann als mittlere Rate aufgefasst werden.

$$P(\Omega) = \sum_{i=0}^{\infty} P(\omega_i) = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{=e^{\lambda}} = 1$$

Wir werden später sehen, daß diese Verteilung “natürlich” ist.

# Inhalt

- Grundbegriffe
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- Allgemeine Eigenschaften einer Verteilungsfunktion



### Def. 17 (Messbarkeit von Abbildungen)

Es seien  $(\Omega_1, \mathcal{E}_1, P_1)$  und  $(\Omega_2, \mathcal{E}_2, P_2)$  Wahrscheinlichkeitsräume. Eine Abbildung

$$X: \Omega_1 \longrightarrow \Omega_2$$

heißt  $\mathcal{E}_1$ - $\mathcal{E}_2$ -messbar, falls für alle Ereignisse  $A \in \mathcal{E}_2$  gilt:

$$X^{-1}(A) = \{\omega \in \Omega_1: X(\omega) \in A\} \in \mathcal{E}_1.$$

**Bem.:** Oftmals wird die Menge  $\mathcal{B}^1$  der BOREL-Mengen als Ereignisfeld  $\mathcal{E}_2$  betrachtet.

# Zufällige Variable

## Def. 18 (Zufällige Variable, Zufallsgröße)

Es sei  $(\Omega, \mathcal{E}, P)$  ein Wahrscheinlichkeitsraum.

Eine  $\mathcal{E}$ - $\mathcal{B}^1$ -meßbare Abbildung  $X$  von  $\Omega$  in  $\mathbb{R}$  heißt  
(reellwertige) zufällige Variable oder Zufallsgröße.

**Bem.:**  $(\mathbb{R}, \mathcal{B}^1, P')$  bildet hier einen zweiten

Wahrscheinlichkeitsraum, wobei  $P'$  eine Abbildung von  $\mathcal{B}^1$  in  $\mathbb{R}$  ist,  
die den KOLMOGOROV-Axiomen genügt.

# Zufällige Variable

## Beispiel (1)

### Augensumme beim dreimaligen Würfeln

$\Omega = \{(i, j, k), 1 \leq i, j, k \leq 6\}$ : Tripel von Augenzahlen

$\mathcal{E} = \mathcal{P}(\Omega)$ : Ereignisfeld

$P(\omega) = P(i, j, k) = \frac{1}{6^3}$ : Laplace-Wahrscheinlichkeit

$$X: \Omega \rightarrow \Omega'$$

$\Omega' = \{S : 3 \leq S \leq 18\}$  oder  $\Omega' = \mathbb{R}$ ,  $S$ : Augensumme

$\mathcal{E}' = \mathcal{P}(\Omega')$  oder  $\mathcal{E}' = \mathcal{B}$ : Ereignisfeld

$$P(\omega') = P'(S = s) = \frac{\#(i, j, k) : i + j + k = s}{6^3} = \frac{|X^{-1}(s)|}{6^3}$$

Bedingung z.B.:  $X^{-1}(s) \in \mathcal{E}$  oder  $X^{-1}(\{s_1, s_2\}) \in \mathcal{E}$

# Zufällige Variable

## Beispiel (2)

### Die Indikatorfunktion ist Zufallsvariable

Sei  $A$  ein Ereignis,  $\Omega = \{A, \bar{A}\}$  und  $\mathcal{E} = \{A, \bar{A}, \emptyset, \Omega\}$ . Die Abbildung

$$I_A(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases}$$

ist messbar, und also Zufallsvariable, denn

$$\begin{aligned} I_A^{-1}(1) &= A \in \mathcal{E}, & I_A^{-1}(0) &= \bar{A} \in \mathcal{E}, \\ I_A^{-1}(\{0, 1\}) &= \Omega \in \mathcal{E}, & I_A^{-1}(y) &= \emptyset \in \mathcal{E} (y \neq 0, 1), \end{aligned}$$

# Zufällige Variable

## Fortsetzung

$X: \Omega \longrightarrow \mathbb{R}$  sei eine zufällige Variable,  $X: (\Omega, \mathcal{E}, P) \longrightarrow (\mathbb{R}, \mathcal{B}^1, P_X)$ .

Sei  $x \in \mathbb{R}$  beliebig, aber fest. Betrachten das zufällige Ereignis

$$B = (-\infty, x) = \{X < x\} := \{\omega \in \Omega: X(\omega) < x\} \in \mathcal{B}^1.$$

Für die Wahrscheinlichkeit dieses Ereignisses gilt:

$$\begin{aligned} P(X < x) &= P(\{\omega: X(\omega) < x\}) = P(\{\omega: X(\omega) \in B\}) \\ &= P(X^{-1}(B)) =: P_X(B) \end{aligned}$$

# Verteilungsfunktion

Def. 19 (Verteilungsfunktion von  $X$ )

$$F_X(x) := P(X < x) = P_X((-\infty, x))$$

**Bem.:** Der Einfachheit halber werden wir die Funktion  $F_X$  einfach nur mit  $F$  bezeichnen.

**Bem.:** Manchmal wird die Verteilungsfunktion auch durch

$$F_X(x) = P(X \leq x)$$

definiert (bei SAS oder Mathematica z.B.)

# Inhalt

- Grundbegriffe
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- Allgemeine Eigenschaften einer Verteilungsfunktion

## Eine diskrete Zufallsgröße

$$X: \Omega \longrightarrow \{x_i: i \in \mathbb{N}\} =: W \subset \mathbb{R}.$$

nimmt höchstens abzählbar viele verschiedene Werte mit positiver Wahrscheinlichkeit an.

Notation:

$$X: \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

$x_i \in \mathbb{R}$ : Werte, die die Zufallsgröße annehmen kann  $p_i$ : die entsprechenden Wahrscheinlichkeiten.



# Diskrete Zufallsvariablen

Fortsetzung

Es gilt:

$$p_i \geq 0, \quad \sum_{i=1}^{\infty} p_i = 1, \quad p_i = P(X = x_i).$$

Wenn wir Mengen  $A_i$  definieren durch

$$A_i := \{\omega : X(\omega) = x_i\}, \quad \forall i \in \mathbb{N},$$

so gilt offenbar:  $A_i \cap A_j = \emptyset$ ,  $\forall i, j \in \mathbb{N}, i \neq j$ . Allgemein gilt dann:

$$P(X = x) = \begin{cases} p_i, & \text{falls } x = x_i \\ 0, & \text{falls } x \neq x_i \end{cases} \quad \forall x_i \in W, i \in \mathbb{N}.$$

# Diskrete Zufallsvariablen

## Verteilungsfunktion

$$\begin{aligned} F(x) &= P(X < x) = P\left(\bigcup_{i: x_i < x} A_i\right) \\ &= \sum_{i: x_i < x} P(A_i) = \sum_{i: x_i < x} p_i \end{aligned}$$

D.h.: Eine diskrete Zufallsgröße, die die Werte  $\{x_i: i \in \mathbb{N}\}$  annimmt, wobei  $x_1 < x_2 < x_3 < \dots$  gilt, hat die folgende Verteilungsfunktion:

$$F(x) = \begin{cases} 0, & \text{falls } x \leq x_1 \\ \sum_{i: x_i < x} p_i, & \text{falls } x_1 < x \end{cases}$$

# Diskrete Zufallsvariablen

## Beispiele (1)

### Diskrete Gleichverteilung

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}$$

# Diskrete Zufallsvariablen

## Beispiele (2)

Binomialverteilung,  $X \sim B(p, n)$  oder  $X \sim Bi(p, n)$ .

$$X : \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix}$$

$$P(X = i) = p_i = \binom{n}{i} p^i \cdot (1 - p)^{n-i} > 0, \quad 0 < p < 1.$$

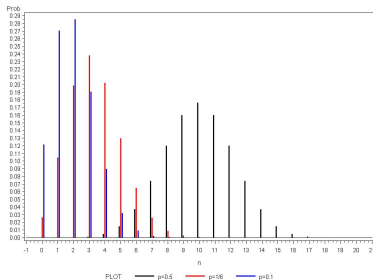
Wir haben oben gesehen, dass

$$\sum_{i=0}^n p_i = \sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = (p + 1 - p)^n = 1.$$

# Diskrete Zufallsvariablen, Wahrscheinlichkeitsfunktionen

## Binomial

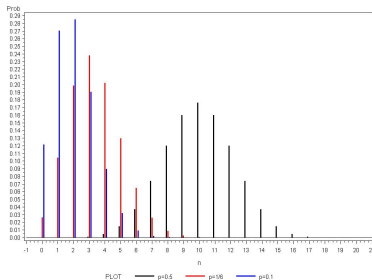
Binomial-Verteilung mit  $n=20$  und  $p=0.5, 1/6, 0.1$



# Diskrete Zufallsvariablen, Wahrscheinlichkeitsfunktionen

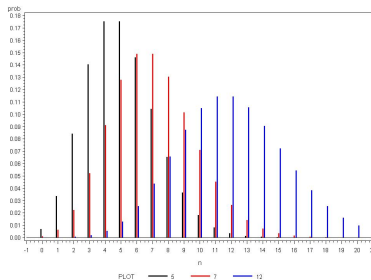
## Binomial

Binomial-Verteilung mit  $n=20$  und  $p=0.5, 1/6, 0.1$



## Poisson

Poisson-Verteilung mit  $\lambda=5, 7, 12$



# Diskrete Zufallsvariablen

## Beispiele (3)

### POISSON-Verteilung, $X \sim \text{Poi}(\lambda)$

Es sei  $X$  eine diskrete Zufallsgröße,

$$X : \begin{pmatrix} 0 & 1 & \dots & n & \dots \\ p_0 & p_1 & \dots & p_n & \dots \end{pmatrix}$$

$$P(X = n) = p_n = \frac{\lambda^n}{n!} e^{-\lambda}, \quad \lambda > 0.$$

Wir haben oben gesehen, dass

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}}_{=e^{\lambda}} = 1$$

# Inhalt

- Grundbegriffe
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- Allgemeine Eigenschaften einer Verteilungsfunktion



### Def. 20 (Dichtefunktion)

Eine Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  heißt Dichtefunktion, falls sie die folgenden Eigenschaften hat:

- 1 Für alle  $x \in \mathbb{R}$  gilt:  $f(x) \geq 0$ .
- 2 Es gilt:  $\int_{\mathbb{R}} f(x) dx = 1$ .

### Def. 21 (Stetige Zufallsvariable)

Eine zufällige Variable  $X$  heißt stetig, falls eine Dichtefunktion  $f$  existiert, so dass gilt:

$$P(X < x) = F(x) = \int_{-\infty}^x f(t) dt.$$

Falls die Funktion  $f$  stetig ist, gilt:  $F'(x) = f(x)$ .

# Stetige Zufallsvariablen

**Bem.:** Für die Wahrscheinlichkeit  $P(X = x)$  gilt

$$P(X = x) = \int_x^x f(t) dt = 0,$$

sogar wenn  $X$  den Wert  $x$  tatsächlich annehmen kann! D.h. z.B.

$$P(X \leq x) = P(X < x).$$

Außerdem gilt:

$$P(a \leq X \leq b) = \int_a^b f(t) dt.$$

# Stetige Zufallsvariablen

## Veranschaulichung der Dichtefunktion

Es sei  $X$  eine stetige Zufallsgröße. Wir teilen den Wertebereich von  $X$  in Intervalle  $I_j$  ein und beobachten für jeden der Versuche  $X_i$ , in welches der Intervalle  $I_j$  der Wert  $X_i$  ( $i = 1, \dots, n$ ) fällt. Es sei  $n_j = \#\{X_i \in I_j\}$ .  $\Delta_j$ : Länge eines Intervalls  $I_j$ . Sei  $\Delta_0 = \max_j \{\Delta_j\}$ .

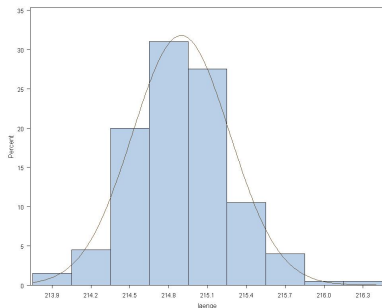
$$f_{\text{emp.}}(x) := \frac{\frac{n_j}{n}}{\Delta_j}, \quad \forall x \in I_j.$$

Dann gilt:

$$f(x) = \lim_{\substack{n \rightarrow \infty \\ \Delta_0 \rightarrow 0}} f_{\text{emp.}}(x).$$

# Stetige Zufallsvariablen

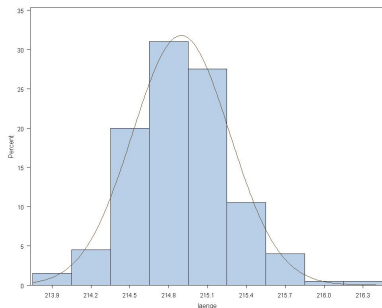
## Veranschaulichung der Dichtefunktion (2)



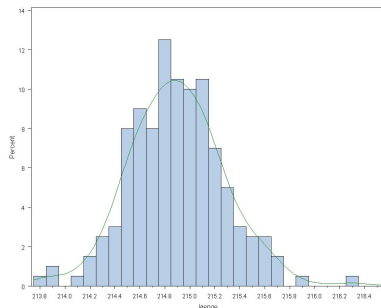
$\Delta_0$  groß

# Stetige Zufallsvariablen

## Veranschaulichung der Dichtefunktion (2)



$\Delta_0$  groß



$\Delta_0$  klein

# Stetige Zufallsvariablen, Beispiele (1)

Gleichverteilung, bez.  $X \sim R(0, 1)$  oder  $X \sim U(0, 1)$

Es sei die Zufallsvariable  $X$  auf dem Intervall  $[0, 1[$  definiert mit der Verteilungsfunktion

$$F(x) = \begin{cases} 0, & \text{falls } x < 0 \\ x, & \text{falls } 0 \leq x < 1 \\ 1, & \text{falls } x \geq 1 \end{cases} .$$

Die Dichtefunktion ist die Funktion  $f$ ;

$$f(x) = \begin{cases} 1, & \text{falls } 0 \leq x < 1 \\ 0, & \text{sonst} \end{cases} .$$

# Stetige Zufallsvariablen, Beispiele (2)

Gleichverteilung, bez.  $X \sim R(a, b)$  oder  $X \sim U(a, b)$

Sei  $X$  gleichverteilt auf dem Intervall  $[a, b]$ ,  $X \sim R(a, b)$ , dann hat  $X$  die Dichtefunktion:

$$f(x) = \begin{cases} 0, & \text{falls } x < a \\ \frac{1}{b-a}, & \text{falls } a \leq x < b \\ 0, & \text{falls } x \geq b \end{cases}$$

$$\begin{aligned} P(\{\omega: X(\omega) \in [a, b]\}) &= P(a \leq X \leq b) \\ &= \int_a^b f(x) dx = \frac{1}{b-a} \int_a^b dx = 1 \end{aligned}$$

## Stetige Zufallsvariablen, Beispiele (3)

### Exponentialverteilung, $X \sim \text{Exp}(\lambda)$

Die Zufallsvariable  $X$  habe die Verteilungsfunktion

$$F(x) = \begin{cases} 1 - e^{-\lambda \cdot x}, & \text{falls } x \geq 0 \\ 0, & \text{falls } x < 0 \end{cases}.$$

Die Dichtefunktion ist

$$f(x) = F'(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, & \text{falls } x \geq 0 \\ 0, & \text{falls } x < 0 \end{cases}.$$

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$



# Stetige Zufallsvariablen, Beispiele (4)

Normalverteilung,  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$X : (\Omega, \mathcal{E}, P) \rightarrow (\mathbb{R}^1, \mathcal{B}^1, P_X)$$

sei der Messfehler bei Messung einer physikalischen Konstanten.

Der Wahrscheinlichkeitsraum  $(\Omega, \mathcal{E}, P)$  ist ein Modell eines im Hintergrund wirkenden Zufallsmechanismus, der nicht näher beschrieben werden kann, Fehler im Meßinstrument; zufällige äußere Einflüsse.

Er enthält alle nicht näher bestimmbaren zufälligen Effekte. Zur Beschreibung dient der Bildraum  $(\mathbb{R}^1, \mathcal{B}^1, P_X)$ .

# Stetige Zufallsvariablen, Beispiele (4a)

## Normalverteilung, $X \sim \mathcal{N}(\mu, \sigma^2)$

Die Zufallsvariable  $X$  mit der Verteilungsfunktion

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

heißt normalverteilt mit den Parametern  $(\mu, \sigma^2)$ . Die zugehörige Dichtefunktion hat die Form:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \sigma > 0.$$

# Stetige Zufallsvariablen, Beispiele (4b)

Satz:  $f(x)$  ist eine Dichtefunktion

Offensichtlich ist  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$  und  $\sigma > 0$ . Es bleibt zu zeigen

$$\lim_{x \rightarrow \infty} F(x) = \int_{-\infty}^{+\infty} f(t) dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = 1.$$

Wir bezeichnen

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =: I.$$

# Stetige Zufallsvariablen, Beispiele (4c)

$$I^2 = \left( \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right)^2$$

# Stetige Zufallsvariablen, Beispiele (4c)

$$\begin{aligned} I^2 &= \left( \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right)^2 \\ &= \frac{1}{2\pi\sigma^2} \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \right) \end{aligned}$$

# Stetige Zufallsvariablen, Beispiele (4c)

$$\begin{aligned} I^2 &= \left( \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right)^2 \\ &= \frac{1}{2\pi\sigma^2} \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \right) \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \end{aligned}$$

# Stetige Zufallsvariablen, Beispiele (4c)

$$\begin{aligned} I^2 &= \left( \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right)^2 \\ &= \frac{1}{2\pi\sigma^2} \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \right) \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right) e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dx dy \end{aligned}$$

# Stetige Zufallsvariablen, Beispiele (4d)

Substitution:  $s := \frac{x-\mu}{\sigma}$        $t := \frac{y-\mu}{\sigma}$ . Dann gilt:

$$x = s\sigma + \mu \quad y = t\sigma + \mu,$$

$$dx = \sigma ds \quad dy = \sigma dt.$$

$$\begin{aligned} I^2 &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}s^2} e^{-\frac{1}{2}t^2} \sigma^2 ds dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(s^2+t^2)} ds dt \end{aligned}$$



## Stetige Zufallsvariablen, Beispiele (4e)

Wir führen eine weitere Substitution durch, Polarkoordinaten:

$$s = r \cos \varphi \quad t = r \sin \varphi.$$

Dann gilt allgemein nach der Substitutionsregel:

$$\int \int g(s, t) ds dt = \int \int g(s(r, \varphi), t(r, \varphi)) \det J dr d\varphi,$$

wobei  $J$  die Jacobi-Matrix ist.

# Stetige Zufallsvariablen, Beispiele (4f)

$$\det J = |J| = \begin{vmatrix} \frac{\partial s}{\partial r} & \frac{\partial s}{\partial \varphi} \\ \frac{\partial t}{\partial r} & \frac{\partial t}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix}$$

# Stetige Zufallsvariablen, Beispiele (4f)

$$\begin{aligned}\det J = |J| &= \begin{vmatrix} \frac{\partial s}{\partial r} & \frac{\partial s}{\partial \varphi} \\ \frac{\partial t}{\partial r} & \frac{\partial t}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} \\ &= r \cos^2 \varphi + r \sin^2 \varphi \\ &= r(\cos^2 \varphi + \sin^2 \varphi) = r\end{aligned}$$

$$\begin{aligned}I^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}(r^2 \cos^2 \varphi + r^2 \sin^2 \varphi)} r \, dr \, d\varphi \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r \, dr \, d\varphi\end{aligned}$$

# Stetige Zufallsvariablen, Beispiele (4g)

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r \, dr \, d\varphi$$

# Stetige Zufallsvariablen, Beispiele (4g)

$$\begin{aligned} I^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r \, dr \, d\varphi \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left[ -e^{-\frac{r^2}{2}} \right]_0^{\infty} d\varphi \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\varphi = \frac{1}{2\pi} 2\pi = 1 \end{aligned}$$

$\implies I = 1$ , d.h.  $f$  ist eine Dichtefunktion.

# Zufallsvariable, Grundbegriffe

## Zusammenfassung (1)

Eine Zufallsvariable ist eine (meßbare) Abbildung

$$X : \Omega \longrightarrow \mathbb{R}$$

Jedem Element  $\omega$  des Stichprobenraumes  $\Omega$  wird eine reelle Zahl zugeordnet.

Die Zufallsvariable  $X$  heißt diskret, wenn  $X$  nur endlich viele oder abzählbar unendlich viele Werte  $x_i$  annehmen kann. Jeder dieser Werte kann mit einer gewissen Wkt.  $p_i = P(X = x_i)$  auftreten.

geografische Lage (N,O,S,W); Länge einer Warteschlange; Anzahl der Punkte in der Klausur.

# Zufallsvariable, Grundbegriffe

## Zusammenfassung (2)

Die Zufallsvariable  $X$  heißt stetig, falls  $X$  beliebige Werte in einem Intervall  $(a, b)$ ,  $[a, b]$ ,  $(a, b]$ ,  $(a, b)$ ,  $(-\infty, a)$ ,  $(b, \infty)$ ,  $(-\infty, a]$ ,  $[b, \infty)$ ,  $(-\infty, \infty)$  annehmen kann.

Bem.: Jeder einzelne Wert  $x_i \in (a, b)$  (oder in einem der anderen Intervalle) hat die Wkt. Null.

Die Verteilungsfunktion  $F$  wird dann durch die sogenannte Dichtefunktion  $f$  beschrieben,

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

# Inhalt

- Grundbegriffe
- Diskrete Zufallsvariablen
- Stetige Zufallsvariablen
- Allgemeine Eigenschaften einer Verteilungsfunktion



Satz: Sei  $X$  eine Zufallsvariable mit der Verteilungsfunktion

$$F(x) = P(X < x) = P(\{\omega : X(\omega) < x\}) = P_X((-\infty, x)).$$

Dann gelten die folgenden Aussagen:

- ❶ Die Funktion  $F(x)$  ist monoton wachsend.
- ❷  $\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$
- ❸ Die Funktion  $F(x)$  ist linksseitig stetig. Es gilt also:  
 $\lim_{x \rightarrow x_0-} F(x) = F(x_0).$
- ❹  $P(a \leq X < b) = F(b) - F(a).$

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (1)

① Es sei  $x_1 < x_2 < x$ . Wir definieren zwei Mengen:

$$A := \{\omega : X(\omega) < x_1\},$$

$$B := \{\omega : X(\omega) < x_2\}.$$

Dann gilt:

$$F(x_1) = P(\{\omega : X(\omega) < x_1\}) = P(A),$$

$$F(x_2) = P(\{\omega : X(\omega) < x_2\}) = P(B).$$

Wegen  $A \subseteq B$  folgt:  $P(A) \leq P(B)$ , d.h.

$$F(x_1) \leq F(x_2),$$

d.h. die Funktion  $F(x)$  monoton wachsend.

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (2)

- 2 Sei  $(x_n)$  eine monoton fallende Folge mit  $x_n \rightarrow -\infty$  und  $(y_n)$  eine monoton wachsende Folge mit  $y_n \rightarrow \infty$ . Wir definieren:

$$A_n := \{\omega : X(\omega) < x_n\},$$

$$B_n := \{\omega : X(\omega) < y_n\}.$$

Für die Folgen  $(A_n)$  und  $(B_n)$  gilt:

$(A_n)$  ist monoton fallend  $(A_n \supseteq A_{n+1}, \forall n \in \mathbb{N})$ ,

$(B_n)$  monoton wachsend  $(B_n \subseteq B_{n+1}, \forall n \in \mathbb{N})$ . Offensichtlich gilt:

$$F(x_n) = P(A_n), \quad F(y_n) = P(B_n).$$

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (3)

Wegen der Stetigkeit der Wahrscheinlichkeit von oben und unten ist

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\lim_{n \rightarrow \infty} A_n\right) = P(X < -\infty) = 0.$$

$$\lim_{n \rightarrow \infty} P(B_n) = P\left(\lim_{n \rightarrow \infty} B_n\right) = P(X < +\infty) = 1.$$

Das ist äquivalent zu:

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow \infty} F(x_n) = 0,$$

$$\lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow \infty} F(y_n) = 1.$$

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (4)

3 Wir definieren eine Menge

$$A = \{\omega : X(\omega) < x_0\}$$

und eine Folge von Mengen

$$A_n = \{\omega : X(\omega) < x_n\},$$

wobei  $(x_n)$  eine monotone Folge ist, die von links gegen  $x_0$  konvergiert ( $x_n \rightarrow x_0 - 0$ ). Offenbar ist die Folge  $(A_n)$  monoton wachsend ( $A_n \subseteq A_{n+1}$ ). Außerdem gilt:

$$\lim_{n \rightarrow \infty} A_n = A.$$

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (5)

Damit folgt:

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(X < x_n)$$

# Eigenschaften der Verteilungsfunktion

Beweis des Satzes (5)

Damit folgt:

$$\begin{aligned}\lim_{n \rightarrow \infty} F(x_n) &= \lim_{n \rightarrow \infty} P(X < x_n) = \lim_{n \rightarrow \infty} P(A_n) \\ &= P\left(\lim_{n \rightarrow \infty} A_n\right) = P(A) = P(X < x_0) \\ &= F(x_0)\end{aligned}$$

D.h.:

$$\lim_{x \rightarrow x_0 -} F(x) = F(x_0).$$

# Eigenschaften der Verteilungsfunktion

## Beweis des Satzes (6)

④ Es gilt:

$$\begin{aligned}P(a \leq X < b) &= P(\{X < b\} \setminus \{X < a\}) \\&= P(X < b) - P(X < a) \\&\quad \text{(Subtraktivität der Wahrscheinlichkeit)} \\&= F(b) - F(a)\end{aligned}$$



# Inhalt

- Allgemeine Übersicht
- Binomialverteilung
- Geometrische Verteilung
- Poisson-Verteilung
- Negative Binomialverteilung

Erinnerung: Wir beschreiben diskrete Zufallsvariablen durch

$$X : \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n & \cdots \\ p_1 & p_2 & p_3 & \cdots & p_n & \cdots \end{pmatrix}$$

$$p_i = P(X = x_i) > 0, \quad i = 1, 2, 3, \dots \quad \sum_{i=1}^{\infty} p_i = 1$$

Def. 22 (Wahrscheinlichkeitsfunktion, Zähldichte)

Die Funktion

$$f(x_i) = p_i$$

heißt Wahrscheinlichkeitsfunktion.

# Binomialwahrscheinlichkeit

## a) Zweimaliges Werfen einer Münze

$$\Omega = \{ZZ, ZB, BZ, BB\}$$

$X :=$  Anzahl von Blatt

$$X : \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

# Binomialwahrscheinlichkeit

## a) Zweimaliges Werfen einer Münze

$$\Omega = \{ZZ, ZB, BZ, BB\}$$

$X :=$  Anzahl von Blatt

$$X : \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

## b) Erfolge bei $n$ Versuchen

$X$ : Anzahl der “Erfolge” bei  $n$  Versuchen, wobei jeder der  $n$  Versuche eine Erfolgswahrscheinlichkeit  $p$  hat.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{Binomialwkt.}$$

$$F_X(k) = P(X < k) = \sum_{i=0}^{k-1} \binom{n}{i} p^i (1 - p)^{n-i}$$

# Binomialwahrscheinlichkeit, Beispiele (1)

Es seien  $p = \frac{1}{2}$  und  $n = 5$ . Für  $x = 2.5$  gilt:

$$\begin{aligned} F(2.5) &= \sum_{i: i < 2.5} p_i \\ &= p_0 + p_1 + p_2 \end{aligned}$$

# Binomialwahrscheinlichkeit, Beispiele (1)

Es seien  $p = \frac{1}{2}$  und  $n = 5$ . Für  $x = 2.5$  gilt:

$$\begin{aligned} F(2.5) &= \sum_{i: i < 2.5} p_i \\ &= p_0 + p_1 + p_2 \\ &= \binom{5}{0} \left(\frac{1}{2}\right)^5 + \binom{5}{1} \left(\frac{1}{2}\right)^5 + \binom{5}{2} \left(\frac{1}{2}\right)^5 \\ &= \frac{1}{32} + \frac{5}{32} + \frac{10}{32} \\ &= 0.5 \end{aligned}$$

# Binomialwahrscheinlichkeit, Beispiele (2)

Würfeln 20 mal. Wie groß ist die Wahrscheinlichkeit für mindestens 4 Sechsen?

## Binomialwahrscheinlichkeit, Beispiele (2)

Würfeln 20 mal. Wie groß ist die Wahrscheinlichkeit für mindestens 4 Sechsen?

$X$ : Anzahl der Sechsen.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) = 1 - F_X(4) \\ &= 1 - \sum_{i=0}^3 P(X = i) = \end{aligned}$$



## Binomialwahrscheinlichkeit, Beispiele (2)

Würfeln 20 mal. Wie groß ist die Wahrscheinlichkeit für mindestens 4 Sechsen?

$X$ : Anzahl der Sechsen.

$$\begin{aligned}P(X \geq 4) &= 1 - P(X < 4) = 1 - F_X(4) \\&= 1 - \sum_{i=0}^3 P(X = i) =\end{aligned}$$

$$\begin{aligned}1 - \left(\frac{5}{6}\right)^{20} - 20\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{19} - \frac{20 \cdot 19}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{18} - \\- \frac{20 \cdot 19 \cdot 18}{6}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{17}\end{aligned}$$

$\approx 0.43$ .

# Poisson-Wahrscheinlichkeit

## Beispiel

Telefonzentrale,  $X \sim \text{Poi}(\lambda)$

$X$ : Anzahl der Anrufe, die pro Zeiteinheit von einer Telefonzentrale vermittelt werden.

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \end{pmatrix}$$

$$P(X = i) = p_i = \frac{\lambda^i}{i!} e^{-\lambda}, \quad \lambda > 0$$

$$\sum_{i=0}^{\infty} p_i = \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}} e^{-\lambda} = 1.$$

# Binomial und Poisson

Satz: Seien  $X_n \sim Bi(n, p)$ ,  $Y \sim Poi(\lambda)$

Für  $n \cdot p = \lambda$  gilt:  $P(X_n = k) \xrightarrow{n \rightarrow \infty} P(Y = k)$ .

# Binomial und Poisson

Satz: Seien  $X_n \sim Bi(n, p)$ ,  $Y \sim Poi(\lambda)$

Für  $n \cdot p = \lambda$  gilt:  $P(X_n = k) \xrightarrow{n \rightarrow \infty} P(Y = k)$ .

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

# Binomial und Poisson

Satz: Seien  $X_n \sim Bi(n, p)$ ,  $Y \sim Poi(\lambda)$

Für  $n \cdot p = \lambda$  gilt:  $P(X_n = k) \xrightarrow{n \rightarrow \infty} P(Y = k)$ .

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!(n-\lambda)^k} \frac{(n-\lambda)^k \lambda^k}{n^k} \frac{(n-\lambda)^{n-k}}{n^{n-k}} \end{aligned}$$

# Binomial und Poisson

Satz: Seien  $X_n \sim Bi(n, p)$ ,  $Y \sim Poi(\lambda)$

Für  $n \cdot p = \lambda$  gilt:  $P(X_n = k) \xrightarrow{n \rightarrow \infty} P(Y = k)$ .

$$\begin{aligned}
 P(X_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{n(n-1) \cdots (n-k+1)}{k!(n-\lambda)^k} \frac{(n-\lambda)^k \lambda^k}{n^k} \frac{(n-\lambda)^{n-k}}{n^{n-k}} \\
 &= \frac{1}{k!} \underbrace{\frac{n(n-1) \cdots (n-k+1)}{(n-\lambda)^k}}_{\rightarrow 1} \lambda^k \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}}
 \end{aligned}$$

# Binomial und Poisson

Satz: Seien  $X_n \sim Bi(n, p)$ ,  $Y \sim Poi(\lambda)$

Für  $n \cdot p = \lambda$  gilt:  $P(X_n = k) \xrightarrow{n \rightarrow \infty} P(Y = k)$ .

$$\begin{aligned}
 P(X_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{n(n-1) \cdots (n-k+1)}{k!(n-\lambda)^k} \frac{(n-\lambda)^k \lambda^k}{n^k} \frac{(n-\lambda)^{n-k}}{n^{n-k}} \\
 &= \frac{1}{k!} \underbrace{\frac{n(n-1) \cdots (n-k+1)}{(n-\lambda)^k}}_{\rightarrow 1} \lambda^k \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \\
 &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} = P(Y = k)
 \end{aligned}$$

# Geometrische Verteilung

d) Münzwurf solange bis B(Blatt) kommt

$$\Omega = \{B, ZB, ZZB, \dots\}$$

$X :=$  Anzahl der Würfe bis zum ersten Blatt.

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n & \dots \\ \frac{1}{2} & (\frac{1}{2})^2 & (\frac{1}{2})^3 & (\frac{1}{2})^4 & \dots & (\frac{1}{2})^n & \dots \end{pmatrix}$$



# Geometrische Verteilung

d) Münzwurf solange bis B(Blatt) kommt

$$\Omega = \{B, ZB, ZZB, \dots\}$$

$X :=$  Anzahl der Würfe bis zum ersten Blatt.

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n & \dots \\ \frac{1}{2} & (\frac{1}{2})^2 & (\frac{1}{2})^3 & (\frac{1}{2})^4 & \dots & (\frac{1}{2})^n & \dots \end{pmatrix}$$

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} (1/2)^i = \frac{1}{1 - \frac{1}{2}} - 1 = 1$$

geometrische Reihe

geometrische Verteilung mit  $p=1/2$ ,  $p_i = (1/2)^i$ .

# Geometrische Verteilung

## Def. 23 (Geometrische Verteilung)

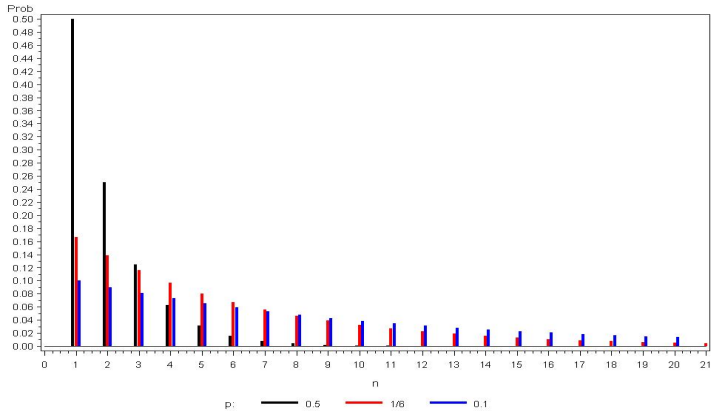
Eine Zufallsvariable  $X$  mit

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

heißt geometrisch verteilt, bez.  $X \sim \text{Geo}(p)$

Anzahl der Schritte bis zum ersten “Erfolg”.

## Geometrische Verteilung mit $p=0.5, 1/6, 0.1$



# Hypergeometrische Verteilung

## e) Qualitätskontrolle

Gegeben sei eine Grundgesamtheit (z.B. eine Warenlieferung) mit  $N$  Stücken, von denen genau  $n$  schlecht seien. Wie groß ist die Wahrscheinlichkeit, dass in einer Stichprobe vom Umfang  $m$  höchstens  $k$  Stück schlecht sind?

$X$ : zufällige Anzahl der schlechten Stücke in der Stichprobe.

$$P(X = x) = \frac{\binom{n}{x} \cdot \binom{N-n}{m-x}}{\binom{N}{m}}$$

# Hypergeometrische Verteilung

Fortsetzung

$\binom{N}{m}$ : # möglichen Stichproben.

$\binom{n}{x}$ : # Möglichkeiten, aus  $n$  schlechten Stücken in der Grundgesamtheit  $x$  schlechte Stücke zu ziehen.

$\binom{N-n}{m-x}$ : # Möglichkeiten, aus  $N - n$  guten Stücken in der Grundgesamtheit  $m - x$  gute Stücke zu ziehen.

Offenbar:

$$0 \leq x \leq \min(n, m)$$

$$m - x \leq N - n.$$

# Hypergeometrische Verteilung

## Def. 24 (Hypergeometrische Verteilung)

Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$f(x|H_{N,n,m}) = \frac{\binom{n}{x} \cdot \binom{N-n}{m-x}}{\binom{N}{m}}$$

heißt hypergeometrisch verteilt.

**Bez.:**  $X \sim H_{N,n,m}$ . Verteilungsfunktion:

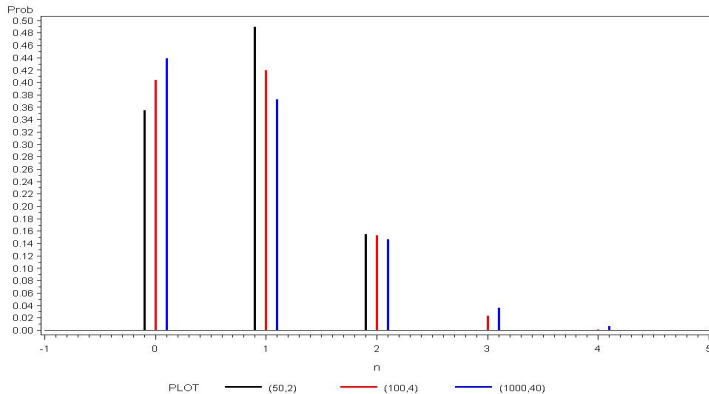
$$F(k|H_{N,n,m}) = \sum_{x=0}^{k-1} \frac{\binom{n}{x} \cdot \binom{N-n}{m-x}}{\binom{N}{m}}$$

**Satz:** Für  $N \rightarrow \infty$ ,  $n \rightarrow \infty$ ,  $\frac{n}{N} \rightarrow p$  gilt:

$$f(x|H_{N,n,m}) \rightarrow \binom{m}{x} p^x (1-p)^{m-x} = f(x|Bi(m, p))$$

## Hypergeometrische Verteilung mit $m=20$

und  $(N,n)=(1000,40), (100,4), (50,2)$



# Inhalt

- Allgemeine Übersicht
- Binomialverteilung
- Geometrische Verteilung
- Poisson-Verteilung
- Negative Binomialverteilung



## 6.2 Binomialverteilung

### Weitere Beispiele (1)

#### Kommunikationskanal

Schicken Binärzahlen durch einen Kommunikationskanal.

$p$ : Wahrscheinlichkeit einer fehlerhaften Übertragung

$n$ : Anzahl der übertragenen Zeichen

$X$ : Anzahl der Fehler:

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$$

# Binomialverteilung

## Weitere Beispiele (2)

### Qualitätskontrolle

Stichprobe (hier: mit Zurücklegen) von 10 Computerchips aus einer sehr großen Lieferung (Los). Wenn keine defekt, so wird die Lieferung angenommen, sonst nicht.

$p$ : Wahrscheinlichkeit, ein zufällig ausgewählter Chip ist defekt.

$X$ : Anzahl der intakten Stücke,  $X \sim Bi(10, 1 - p)$

$$P(\text{Los angenommen}) = P(X = 10) = (1 - p)^{10}$$

# Binomialverteilung

## Weitere Beispiele (3)

### $k$ aus $n$ Systeme

Jede Komponente habe die Intaktwahrscheinlichkeit  $p$ .

$X$ : Anzahl der ausfallenden  $i$  Komponenten.

$$P(X = i) = \binom{n}{i} p^{n-i} (1-p)^i$$

Wahrscheinlichkeit, daß höchstens  $k$  Komponenten ausfallen:

$$\begin{aligned} P(X \leq k) &= \sum_{i=0}^k \binom{n}{i} p^{n-i} (1-p)^i \\ &= \sum_{i=n-k}^n \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

# Inhalt

- Allgemeine Übersicht
- Binomialverteilung
- Geometrische Verteilung
- Poisson-Verteilung
- Negative Binomialverteilung

# Geometrische Verteilung (1)

Sei  $Y \sim \text{Geo}(p)$ , d.h.

$$P(Y > s) = 1 - \sum_{i=1}^s (1-p)^{i-1} \cdot p = (1-p)^s$$

$$P(Y > t) = 1 - \sum_{i=1}^t (1-p)^{i-1} \cdot p = (1-p)^t$$

# Geometrische Verteilung (1)

Sei  $Y \sim \text{Geo}(p)$ , d.h.

$$P(Y > s) = 1 - \sum_{i=1}^s (1-p)^{i-1} \cdot p = (1-p)^s$$

$$P(Y > t) = 1 - \sum_{i=1}^t (1-p)^{i-1} \cdot p = (1-p)^t$$

$$\begin{aligned} P(Y > s) \cdot P(Y > t) &= (1-p)^{s+t} \\ &= P(Y > s+t). \end{aligned}$$

## Geometrische Verteilung (2)

also:

$$\begin{aligned}P(Y > s + t | Y > t) &= \frac{P(Y > s + t, Y > t)}{P(Y > t)} \\&= \frac{P(Y > s + t)}{P(Y > t)} \\&= P(Y > s)\end{aligned}$$

Def. 25 (Markov-Eigenschaft, Gedächtnislosigkeit)

Verteilungen mit der Markov-Eigenschaft

$$P(Y > s + t | Y > t) = P(Y > s)$$

heißen *gedächtnislos*.

## Geometrische Verteilung (3)

Satz: Sei  $X$  diskrete Zufallsvariable mit Werten in  $\mathbb{N}^+$  und  $X$  habe die Markov-Eigenschaft. Dann ist



## Geometrische Verteilung (3)

**Satz:** Sei  $X$  diskrete Zufallsvariable mit Werten in  $\mathbb{N}^+$

und  $X$  habe die Markov-Eigenschaft. Dann ist  $X \sim \text{Geo}(p)$  für ein  $p, p \in (0, 1)$

**Beweis:** Sei

$$X : \begin{pmatrix} 1 & 2 & 3 & \dots \\ p_1 & p_2 & p_3 & \dots \end{pmatrix}$$

## Geometrische Verteilung (3)

**Satz:** Sei  $X$  diskrete Zufallsvariable mit Werten in  $\mathbb{N}^+$

und  $X$  habe die Markov-Eigenschaft. Dann ist  $X \sim \text{Geo}(p)$  für ein  $p, p \in (0, 1)$

**Beweis:** Sei

$$X : \begin{pmatrix} 1 & 2 & 3 & \dots \\ p_1 & p_2 & p_3 & \dots \end{pmatrix}$$

Aus der Markov-Eigenschaft folgt:

$$\begin{aligned} P(X > s) \cdot P(X > t) &= P(X > s + t) \quad \forall s, t \\ \left(1 - \sum_{i=1}^s p_i\right) \left(1 - \sum_{i=1}^t p_i\right) &= 1 - \sum_{i=1}^{s+t} p_i \end{aligned}$$

# Geometrische Verteilung (4)

$$(1 - \sum_{i=1}^s p_i)(1 - \sum_{i=1}^t p_i) = 1 - \sum_{i=1}^{s+t} p_i$$

Setzen  $p := p_1$ . Einsetzen von

$s = 1, t = 1$  liefert  $(1 - p)^2 = (1 - p - p_2)$ ;  $p_2 = p(1 - p)$ .

# Geometrische Verteilung (4)

$$(1 - \sum_{i=1}^s p_i)(1 - \sum_{i=1}^t p_i) = 1 - \sum_{i=1}^{s+t} p_i$$

Setzen  $p := p_1$ . Einsetzen von

$s = 1, t = 1$  liefert  $(1 - p)^2 = (1 - p - p_2)$ ;  $p_2 = p(1 - p)$ .

$s = 1, t = 2$  liefert  $(1 - p)(1 - p - p_2) = (1 - p - p_2 - p_3)$ ;

$$(1 - p - p_2)(1 - p - 1) = -p_3; \quad \text{also } p_3 = p(1 - p)^2$$

usw.

# Geometrische Verteilung (5)

## Qualitätskontrolle

Wahrscheinlichkeit, daß das  $i$ -te Item das erste defekte ist.

# Geometrische Verteilung (5)

## Qualitätskontrolle

Wahrscheinlichkeit, daß das  $i$ -te Item das erste defekte ist.

## Time-sharing computer system

mit festen Zeitscheiben.

Programm wird in der Zeitscheibe vollständig abgearbeitet mit

Wahrscheinlichkeit  $p$

Wenn nicht, neuer Versuch in der neuen Zeitscheibe

# Geometrische Verteilung (5)

## Qualitätskontrolle

Wahrscheinlichkeit, daß das  $i$ -te Item das erste defekte ist.

## Time-sharing computer system

mit festen Zeitscheiben.

Programm wird in der Zeitscheibe vollständig abgearbeitet mit

Wahrscheinlichkeit  $p$

Wenn nicht, neuer Versuch in der neuen Zeitscheibe

$X$ : # benötigten Zeitscheiben

$$X \sim \text{Geo}(p).$$

# Geometrische Verteilung (6)

## Repeat-Schleife

$A$ : aussagenlogischer Ausdruck,  $A = \text{true}$  mit Wahrscheinlichkeit  $p$ .  
repeat  $S$  until  $A$ .



# Geometrische Verteilung (6)

## Repeat-Schleife

$A$ : aussagenlogischer Ausdruck,  $A = \text{true}$  mit Wahrscheinlichkeit  $p$ .

repeat  $S$  until  $A$ .

$X = \#$  der Durchläufe von  $S$ :  $\sim \text{Geo}(p)$ .

# Inhalt

- Allgemeine Übersicht
- Binomialverteilung
- Geometrische Verteilung
- Poisson-Verteilung
- Negative Binomialverteilung

## 6.4 Poisson-Verteilung

Vorbemerkung, Definition Unabhängigkeit von Zufallsvariablen

Erinnerung: Unabhängigkeit von Ereignissen

Die Ereignisse  $A$  und  $B$  heißen unabhängig, falls

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

Def. 26 (Unabhängigkeit von Zufallsvariablen)

Zwei Zufallsvariablen  $X$  und  $Y$  heißen unabhängig, falls

$$\forall A, B \in \mathcal{B}; \quad P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

## Poisson-Verteilung (2)

Sei  $\{N_t\}_{t \in T}$  eine Menge von Zufallsvariablen (ein stochastischer Prozess ) mit Werten in  $\mathbb{N}$  und mit folgenden Eigenschaften:

V1: Zuwächse sind unabhängig, dh. die Zufallsvariablen

$N_{t+h} - N_t$  und  $N_t - N_{t-h}$  sind unabhängig.

V2: es ist egal wo wir Zeitintervall betrachten, dh.

$N_{t+h} - N_t$  und  $N_t - N_{t-h}$  haben dieselbe Verteilung

V3: Wahrscheinlichkeit, daß mindestens ein Ereignis in der Zeit  $h$  eintritt, z.B. ein Kunde ankommt.

$$p(h) = a \cdot h + o(h), \quad a > 0, h \rightarrow 0$$

V4: Wahrscheinlichkeit für  $\geq 2$  Ereignisse in der Zeit  $h$ :  $o(h)$

# Poisson-Verteilung (3)

$N_t$ : Anzahl der bis zum Zeitpunkt  $t$  eintretenden Ereignisse (z.B. eingetroffene Kunden, zerfallene Teilchen)

$$P_k(t) := P(N_t = k), \quad P_k(t) := 0 \quad \text{für} \quad k < 0$$

$$p(h) := \sum_{k=1}^{\infty} P_k(h) \geq 1 \text{ Ereignis tritt ein}$$

$$1 = \sum_{k=0}^{\infty} P_k(t)$$

$$V3 \Rightarrow P_0(h) = 1 - p(h) = 1 - ah + o(h)$$

$$V4 \Rightarrow \sum_{k=2}^{\infty} P_k(h) = o(h), \quad (h \rightarrow 0)$$

# Poisson-Verteilung (4)

1. Schritt: Bestimmen  $P_0(t)$ .

$$\begin{aligned}P_0(t+h) &= P(N_t = 0, N_{t+h} - N_t = 0) \\&= P_0(t)P(N_{t+h} - N_t = 0) \quad \text{wegen V1} \\&= P_0(t)P(N_h - N_0 = 0) \quad \text{wegen V2} \\&= P_0(t)P_0(h) \quad \text{wegen } N_0 = 0 \\&= P_0(t)(1 - p(h)) \\&= P_0(t)(1 - ah + o(h)) \quad \text{wegen V4}\end{aligned}$$

# Poisson-Verteilung (5)

$$P_0(t + h) = P_0(t)(1 - ah + o(h))$$

Nacheinander folgt:

$$\frac{P_0(t + h) - P_0(t)}{h} = P_0(t)\left(-a + \frac{o(h)}{h}\right)$$

# Poisson-Verteilung (5)

$$P_0(t+h) = P_0(t)(1 - ah + o(h))$$

Nacheinander folgt:

$$\begin{aligned}\frac{P_0(t+h) - P_0(t)}{h} &= P_0(t)\left(-a + \frac{o(h)}{h}\right) \\ \lim_{h \rightarrow 0} \frac{P_0(t+h) - P_0(t)}{h} &= \lim_{h \rightarrow 0} P_0(t)\left(-a + \frac{o(h)}{h}\right) \\ P'_0(t) &= -aP_0(t) \\ P_0(t) &= ce^{-at}\end{aligned}$$

Wegen  $P_0(0) = 1$  folgt:  $c = 1$  und

$$P_0(t) = e^{-at}$$



# Poisson-Verteilung (6)

2. Schritt: Bestimmen  $P_k(t)$ .

Zerlegen das Ereignis  $\{N_{t+h} = k\}$  in disjunkte Teilereignisse.

$$\begin{aligned}\{N_{t+h} = k\} &= \{N_t = 0, N_{t+h} - N_t = k\} \cup \\ &\quad \{N_t = 1, N_{t+h} - N_t = k - 1\} \cup \\ &\quad \{N_t = 2, N_{t+h} - N_t = k - 2\} \cup \dots \cup \\ &\quad \{N_t = k, N_{t+h} - N_t = 0\}\end{aligned}$$

# Poisson-Verteilung (7)

$$P_k(t+h) = \sum_{j=0}^k P(N_t = k-j, N_{t+h} - N_t = j)$$

# Poisson-Verteilung (7)

$$\begin{aligned}P_k(t+h) &= \sum_{j=0}^k P(N_t = k-j, N_{t+h} - N_t = j) \\&= \sum_{j=0}^k P_{k-j}(t) \underbrace{P(N_{t+h} - N_t = j)}_{=P(N_h - N_0=j)} \quad \text{wegen V1}\end{aligned}$$

# Poisson-Verteilung (7)

$$\begin{aligned}
 P_k(t+h) &= \sum_{j=0}^k P(N_t = k-j, N_{t+h} - N_t = j) \\
 &= \sum_{j=0}^k P_{k-j}(t) \underbrace{P(N_{t+h} - N_t = j)}_{=P(N_h - N_0=j)} \quad \text{wegen V1} \\
 &= \sum_{j=0}^k P_{k-j}(t) P_j(h) \quad \text{wegen V2} \\
 &= P_k(t) P_0(h) + P_{k-1}(t) \underbrace{P_1(h)}_{ah+o(h)} + \underbrace{\sum_{j=2}^k P_{k-j}(t) P_j(h)}_{o(h)} \\
 &= P_k(t) P_0(h) + P_{k-1}(t) \cdot (ah + o(h)) + o(h)
 \end{aligned}$$

# Poisson-Verteilung (8)

$$\begin{aligned}P_1(h) &= \sum_{j=1}^{\infty} P_j(h) - \sum_{j=2}^{\infty} P_j(h) \\&= p(h) + o(h) \\&= ah + o(h)\end{aligned}$$

$$\sum_{j=2}^{\infty} P_{k-j}(t) P_j(h) \leq \sum_{j=2}^{\infty} P_j(h) = o(h) \quad \text{wegen V4}$$

# Poisson-Verteilung (9)

Nacheinander folgt:

$$P_k(t+h) = P_k(t)P_0(h) + P_{k-1}(t) \cdot (ah + o(h)) + o(h)$$

$$\begin{aligned} P_k(t+h) - P_k(t) &= (P_0(h) - 1)P_k(t) + P_{k-1}(t) \cdot ah + o(h) \\ &= -ahP_k(t) + ahP_{k-1}(t) + o(h) \end{aligned}$$

# Poisson-Verteilung (9)

Nacheinander folgt:

$$P_k(t+h) = P_k(t)P_0(h) + P_{k-1}(t) \cdot (ah + o(h)) + o(h)$$

$$\begin{aligned} P_k(t+h) - P_k(t) &= (P_0(h) - 1)P_k(t) + P_{k-1}(t) \cdot ah + o(h) \\ &= -ahP_k(t) + ahP_{k-1}(t) + o(h) \end{aligned}$$

$$\frac{P_k(t+h) - P_k(t)}{h} = -aP_k(t) + aP_{k-1}(t) + \frac{o(h)}{h}$$

$$P'_k(t) = -aP_k(t) + aP_{k-1}(t), \quad P_k(0) = 0$$

Die letzte Gleichung folgt durch Grenzwertbildung ( $h \rightarrow 0$ ).

# Poisson-Verteilung (10)

$$Q_k(t) := P_k(t)e^{at}$$

$$Q'_k(t) = P'_k(t)e^{at} + P_k(t)ae^{at}$$

$$\begin{aligned} Q'_k(t) &= e^{at} \underbrace{(-aP_k(t) + aP_{k-1}(t) + aP_k(t))}_{P'_k(t)} \\ &= aQ_{k-1}(t) \end{aligned}$$



# Poisson-Verteilung (10)

$$Q_k(t) := P_k(t)e^{at}$$

$$Q'_k(t) = P'_k(t)e^{at} + P_k(t)ae^{at}$$

$$Q'_k(t) = e^{at} \underbrace{(-aP_k(t) + aP_{k-1}(t) + aP_k(t))}_{P'_k(t)}$$

$$= aQ_{k-1}(t)$$

$$Q'_1(t) = aQ_0(t) = ae^{-at}e^{at} = a \Rightarrow Q_1(t) = at$$

$$Q'_2(t) = aQ_1(t) = a^2t \Rightarrow Q_2(t) = \frac{a^2t^2}{2}$$

Durch vollständige Induktion:

$$Q_k(t) = \frac{a^k t^k}{k!} \quad P_k(t) = \frac{a^k t^k}{k!} e^{-at}$$

Poisson-Verteilung mit Parameter  $\lambda = at$ .

# Inhalt

- Allgemeine Übersicht
- Binomialverteilung
- Geometrische Verteilung
- Poisson-Verteilung
- Negative Binomialverteilung

### Def. 27 (Negative Binomialverteilung)

Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P(X = m + k) = \binom{m + k - 1}{m - 1} p^m (1 - p)^k$$

heißt negativ Binomialverteilt mit Parametern  $(m, p)$

### Qualitätskontrolle

Prüfen solange bis wir  $m$  defekte Stücke entdecken. Wenn  $m + k$  “klein”  $\rightarrow$  Los ablehnen

Wenn  $m + k$  “groß”  $\rightarrow$  Los annehmen

(hier kann die Prüfung evtl. vorzeitig abgebrochen werden.)

## Negative Binomialverteilung (2)

Diese Verteilung entsteht auch, wenn man Poisson-Verteilung mit einer Gamma-Verteilung mischt.

Deshalb wird sie verwendet, wenn sich Zählraten aus verschiedenen Quellen zusammensetzen (und Poisson nicht geeignet scheint).

## Negative Binomialverteilung (2)

Diese Verteilung entsteht auch, wenn man Poisson-Verteilung mit einer Gamma-Verteilung mischt.

Deshalb wird sie verwendet, wenn sich Zählraten aus verschiedenen Quellen zusammensetzen (und Poisson nicht geeignet scheint).

### File-Dokumentenserver

Die Gesamt-Anzahl der Zugriffe auf ein bestimmtes Dokument setzt sich aus Teil-Anzahlen von vielfältigen Zugriffen aus verschiedenartigen Quellen zusammen.

Bem: In den Wahrscheinlichkeiten können Parameter auftreten, die in der Regel unbekannt sind.

Die Parameter sind anhand der Beobachtungen (der Daten) zu bestimmen/zu schätzen!

→ Aufgabe der Statistik

# Inhalt

- Der Erwartungswert
- Moment und Varianz
- Schiefe und Exzess
- Charakteristische Funktionen

## 7. Charakteristika von Verteilungsfunktionen

Eine Münze wird 3 mal geworfen.

Wie oft können wir erwarten, daß Blatt oben liegt?

Wie oft wird im Mittel Blatt oben liegen?

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{pmatrix}$$

Erwartungswert:  $0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$

D.h. bei 10maliger Durchführung des Experiments können wir im Mittel mit 15mal Blatt rechnen.



## 7.1 Der Erwartungswert

Sei  $X$  diskrete Zufallsvariable,

$$X : \begin{pmatrix} x_1 & \dots & x_n & \dots \\ p_1 & \dots & p_n & \dots \end{pmatrix}$$

**Def. 28** (Erwartungswert,  $X$  diskret)

Die reelle Zahl

$$\mathbf{E}X = \sum_{i=1}^{\infty} p_i x_i$$

heißt Erwartungswert von  $X$

# Der Erwartungswert, Beispiele (1)

a)  $X \sim \text{Poisson}(\lambda)$

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & \dots \\ p_0 & p_1 & p_2 & p_3 & \dots \end{pmatrix}$$

$$p_i = \frac{\lambda^i}{i!} e^{-\lambda}$$

$$\mathbf{E}X =$$

# Der Erwartungswert, Beispiele (1)

a)  $X \sim \text{Poisson}(\lambda)$

$$X : \begin{pmatrix} 0 & 1 & 2 & 3 & \dots \\ p_0 & p_1 & p_2 & p_3 & \dots \end{pmatrix}$$

$$p_i = \frac{\lambda^i}{i!} e^{-\lambda}$$

$$\mathbf{E}X = \sum_{i=0}^{\infty} p_i i = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \cdot i = \lambda \underbrace{\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}}_{e^{\lambda}} e^{-\lambda} = \lambda.$$

z.B. mittlere Ankunftsrate.

# Der Erwartungswert, Beispiele (2)

b)  $X \sim B(n, p)$

$$\mathbf{EX} =$$

## Der Erwartungswert, Beispiele (2)

b)  $X \sim B(n, p)$

$$\begin{aligned}\mathbf{E}X &= \sum_{k=0}^n k \binom{n}{k} p^k \cdot (1-p)^{n-k} \\&= p \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\&= p \cdot n \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\&= p \cdot n \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i}, \quad k = i + 1 \\&= n \cdot p.\end{aligned}$$

## Der Erwartungswert, Beispiele (3)

c)  $X \sim \text{Geo}(p)$

$$X : \begin{pmatrix} 1 & 2 & 3 & \dots & k & \dots \\ p & pq & pq^2 & \dots & pq^{k-1} & \dots \end{pmatrix} \quad q = 1 - p$$

$$\mathbf{E}X = \sum_{k=0}^{\infty} x_k p_k = \sum_{k=1}^{\infty} k p q^{k-1} = p \cdot \sum_{k=1}^{\infty} k q^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Beweis des vorletzten Gleichheitszeichens:

## Der Erwartungswert, Beispiele (3)

c)  $X \sim \text{Geo}(p)$

$$X : \begin{pmatrix} 1 & 2 & 3 & \dots & k & \dots \\ p & pq & pq^2 & \dots & pq^{k-1} & \dots \end{pmatrix} \quad q = 1 - p$$

$$\mathbf{E}X = \sum_{k=0}^{\infty} x_k p_k = \sum_{k=1}^{\infty} k p q^{k-1} = p \cdot \sum_{k=1}^{\infty} k q^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Beweis des vorletzten Gleichheitszeichens:

a) durch vollständige Induktion

b) Differenzieren der geometrischen Reihe

# Erwartungswert

## Def. 29 (Erwartungswert, $X$ stetig)

Sei  $X$  stetig mit Dichtefunktion  $f(x)$ . Die reelle Zahl

$$\mathbf{E}X = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

heißt Erwartungswert von  $X$ .



# Der Erwartungswert, Beispiele (4)

a)  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbf{E}X = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-(\frac{x-\mu}{\sigma})^2/2} dx$$

# Der Erwartungswert, Beispiele (4)

a)  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned}\mathbf{E}X &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-(\frac{x-\mu}{\sigma})^2/2} dx \\ &= \int_{-\infty}^{\infty} (\sigma t + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\end{aligned}$$

# Der Erwartungswert, Beispiele (4)

a)  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned}\mathbf{E}X &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-(\frac{x-\mu}{\sigma})^2/2} dx \\ &= \int_{-\infty}^{\infty} (\sigma t + \mu) \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\ &= \mu + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma \cdot t \cdot e^{\frac{-t^2}{2}} dt =\end{aligned}$$

# Der Erwartungswert, Beispiele (4)

a)  $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned}\mathbf{E}X &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-(\frac{x-\mu}{\sigma})^2/2} dx \\&= \int_{-\infty}^{\infty} (\sigma t + \mu) \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt \\&= \mu + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma \cdot t \cdot e^{\frac{-t^2}{2}} dt = \mu.\end{aligned}$$

$$\frac{x-\mu}{\sigma} = t, \quad dt = \frac{1}{\sigma} dx$$

# Der Erwartungswert, Beispiele (5)

b)  $X \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$

$$\mathbf{E}X = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda \cdot x} dx = \frac{1}{\lambda}$$

## Der Erwartungswert, Beispiele (6)

c)  $X \sim R(a, b)$ , gleichverteilt auf dem Intervall  $(a, b)$

$$\begin{aligned} \mathbf{E}X &= \frac{1}{b-a} \int_a^b x \, dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}. \end{aligned}$$

Bemerkung: Die Erwartungswerte sind für stetige und diskrete Zufallsgrößen zweckmäßigerweise unterschiedlich definiert. Sie lässt sich jedoch (maßtheoretisch) vereinheitlichen.

# Eigenschaften des Erwartungswertes

## Satz

Seien  $X, X_1$  und  $X_2$  zufällige Variablen und  $a, b, c \in \mathbb{R}$  beliebig. Dann gelten folgende Aussagen:

- ① Wenn  $P(X = c) = 1$ , d.h. nimmt die zufällige Variable  $X$  genau einen festen Wert an, so folgt  $\mathbf{E}X = \mathbf{E}c = c$ .
- ② Wenn  $P(X \geq c) = 1$ , so  $\mathbf{E}X \geq c$ .
- ③  $\mathbf{E}(c \cdot X) = c \cdot \mathbf{E}X$ .
- ④  $\mathbf{E}(X + c) = \mathbf{E}X + \mathbf{E}c = \mathbf{E}X + c$ .
- ⑤  $\mathbf{E}(a \cdot X_1 + b \cdot X_2) = a \cdot \mathbf{E}X_1 + b \cdot \mathbf{E}X_2$ .

# Eigenschaften des Erwartungswertes

Beweis des Satzes

**Beweis:** Wir beweisen stellvertretend Aussage 2.

- Es sei  $X$  eine diskrete Zufallsgröße,

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

Nach Voraussetzung:  $c = x_1 < x_2 < \dots < x_n < \dots$ . Daraus folgt:

$$\mathbf{E}X = \sum_{i \in \mathbb{N}} x_i \cdot p_i \geq \sum_{i \in \mathbb{N}} c \cdot p_i = c \cdot \sum_{i \in \mathbb{N}} p_i = c.$$





# Eigenschaften des Erwartungswertes

Beweis des Satzes (Fortsetzung)

- Es sei  $X$  eine stetige zufällige Variable mit der Dichtefunktion  $f$ .

Dann gilt:

$$P(X \geq c) = \int_c^{+\infty} f(x) dx = 1. \quad \Rightarrow$$

$$P(X < c) = \int_{-\infty}^c f(x) dx = 0. \quad \Rightarrow$$

$$\mathbf{E}X = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_c^{+\infty} x \cdot f(x) dx \geq c \cdot \underbrace{\int_c^{+\infty} f(x) dx}_{=1} = c$$

# Eigenschaften des Erwartungswertes

## Ergänzungen

- Aus Aussage 4 folgt:

$$\mathbf{E}(X - \mathbf{E}X) = \mathbf{E}X - \mathbf{E}(\mathbf{E}X) = 0.$$

- Aussage 5 besagt, daß der Erwartungswert eine linearer Operator ist.

# Erwartungswert von Funktionen von Zufallsvariablen

Frage: Wie berechnen wir  $\mathbf{E}(g(X))$ ?

$X$  diskret Dann ist  $Y = g(X)$  gegeben durch

$$Y : \begin{pmatrix} g(x_1) & g(x_2) & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$$

$$\mathbf{E}(g(X)) = \sum_{i=0}^{\infty} g(x_i) p_i$$

# Erwartungswert von Funktionen von Zufallsvariablen

Frage: Wie berechnen wir  $\mathbf{E}(g(X))$ ?

$X$  diskret Dann ist  $Y = g(X)$  gegeben durch

$$Y : \begin{pmatrix} g(x_1) & g(x_2) & \dots \\ p_1 & p_2 & \dots \end{pmatrix}$$
$$\mathbf{E}(g(X)) = \sum_{i=0}^{\infty} g(x_i) p_i$$

$X$  stetig 1. Variante: Dichte  $f_Y$  von  $Y = g(X)$  ausrechnen. Wie man das macht, sehen wir später. Dann

$$\mathbf{E}(Y) = \int y f_Y(y) dy.$$

# Erwartungswert von Funktionen von Zufallsvariablen (2)

## 2. Variante: Satz (Regel des Faulen Statistikers)

Seien  $X$  und  $Y = g(X)$  Zufallsgrößen. Dann gilt:

$$\mathbf{E}(g(X)) = \begin{cases} \sum_{i=0}^{\infty} g(x_i) p_i, & \text{falls } X \text{ diskret} \\ \int_{-\infty}^{\infty} g(x) f(x) dx, & \text{falls } X \text{ stetig} \end{cases}$$

vorausgesetzt die Erwartungswerte existieren.

# Erwartungswert von Funktionen von Zufallsvariablen (3)

## Intuitive Erläuterung: Spiel

wobei wir  $X$  zufällig ziehen. Dann zahle ich den 'Gewinn'  $Y = g(X)$ .  
Ihr erwartetes Einkommen ist

$$\sum_x g(x)P(X = x) \quad \text{bzw.} \quad \int g(x)f(x) dx.$$

Spezialfall:  $g(x) = I_A(x)$  Indikatorfunktion eines Ereignisses  $A$

$$\begin{aligned} \mathbf{E}(I_A(X)) &= \int I_A(x)f_X(x) dx = \int_A f_X(x) dx \\ &= P(X \in A) = P(A). \end{aligned}$$

Die Wahrscheinlichkeit ist ein Spezialfall eines Erwartungswertes!

# Regel des Faulen Statistikers

## Beispiele (1)

Sei  $X \sim R(0, 1)$  und  $Y = g(X) = e^X$ . Dann

$$\mathbf{E}(Y) = \int_0^1 e^x f(x) dx = \int_0^1 e^x dx = e - 1.$$

# Regel des Faulen Statistikers

Beispiele (2)

Stab der Länge 1 zufällig brechen

Sei  $Y$  die Länge des längeren Stücks. Gesucht ist die erwartete Länge  $E(Y)$ .



# Regel des Faulen Statistikers

## Beispiele (2)

### Stab der Länge 1 zufällig brechen

Sei  $Y$  die Länge des längeren Stücks. Gesucht ist die erwartete Länge  $\mathbf{E}(Y)$ .

Wenn  $X$  der zufällige Bruchpunkt ist, dann  $X \sim R(0, 1)$  und  $Y = g(X) = \max(X, 1 - X)$ . D.h.

$$g(x) = \begin{cases} 1 - x & \text{falls } 0 < x < 0.5 \\ x & \text{falls } 0.5 < x < 1 \end{cases}$$

# Regel des Faulen Statistikers

## Beispiele (2)

### Stab der Länge 1 zufällig brechen

Sei  $Y$  die Länge des längeren Stücks. Gesucht ist die erwartete Länge  $\mathbf{E}(Y)$ .

Wenn  $X$  der zufällige Bruchpunkt ist, dann  $X \sim R(0, 1)$  und  $Y = g(X) = \max(X, 1 - X)$ . D.h.

$$g(x) = \begin{cases} 1 - x & \text{falls } 0 < x < 0.5 \\ x & \text{falls } 0.5 < x < 1 \end{cases}$$

$$\mathbf{E}(Y) = \int_0^1 g(x)f(x) dx =$$

# Regel des Faulen Statistikers

## Beispiele (2)

### Stab der Länge 1 zufällig brechen

Sei  $Y$  die Länge des längeren Stücks. Gesucht ist die erwartete Länge  $\mathbf{E}(Y)$ .

Wenn  $X$  der zufällige Bruchpunkt ist, dann  $X \sim R(0, 1)$  und  $Y = g(X) = \max(X, 1 - X)$ . D.h.

$$g(x) = \begin{cases} 1 - x & \text{falls } 0 < x < 0.5 \\ x & \text{falls } 0.5 < x < 1 \end{cases}$$

$$\mathbf{E}(Y) = \int_0^1 g(x)f(x) dx = \int_0^{0.5} (1 - x) dx + \int_{0.5}^1 x dx = \frac{3}{4}.$$

# Regel des Faulen Statistikers

Beweis (1)

Wir zeigen die letzte Behauptung unter der Annahme  $g: R \rightarrow R$  differenzierbar,  $g'(x) \neq 0 \quad \forall x$ .

Wir wollen o.B.d.A. annehmen, dass die Zufallsvariablen  $X$  und  $g(X)$  auf  $(-\infty, \infty)$  definiert sind. Nach der Definition des Erwartungswertes gilt:

$$\mathbf{E}(g(X)) = \int_{-\infty}^{\infty} y \cdot h(y) dy,$$

wobei  $h(y)$  die Dichte von  $Y = g(X)$  ist.

# Regel des Faulen Statistikers

Beweis (2)

Wir bestimmen jetzt  $h(y)$ :

1. Fall: Sei  $g$  monoton wachsend.

$$\begin{aligned} F_Y(t) &= F_{g(X)}(t) = \\ P(g(X) < t) &= P(X < g^{-1}(t)) = \int_{-\infty}^{g^{-1}(t)} f(x) dx \end{aligned}$$

# Regel des Faulen Statistikers

Beweis (2)

Wir bestimmen jetzt  $h(y)$ :

1. Fall: Sei  $g$  monoton wachsend.

$$F_Y(t) = F_{g(X)}(t) =$$

$$P(g(X) < t) = P(X < g^{-1}(t)) = \int_{-\infty}^{g^{-1}(t)} f(x) dx$$

Substitution:  $g(x) = y$ ,  $g'(x)dx = dy$ .

$$F_{g(X)}(t) = \int_{-\infty}^t \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} dy$$

$$\Rightarrow \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = h(y) \quad \text{ist Dichte von } g(X)$$

# Regel des Faulen Statistikers

Beweis (3)

2. Fall: Sei  $g$  monoton fallend.

$$\begin{aligned} F_Y(t) &= F_{g(X)}(t) = \\ P(g(X) < t) &= P(X > g^{-1}(t)) = \int_{g^{-1}(t)}^{\infty} f(x) dx \end{aligned}$$

Substitution:  $g(x) = y$ ,  $g'(x)dx = dy$ ,  $g(\infty) = -\infty$

# Regel des Faulen Statistikers

Beweis (4)

$$\begin{aligned} F_{g(X)}(t) &= \int_t^{-\infty} \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} dy = - \int_t^{-\infty} \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} dy \\ &= \int_{-\infty}^t \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} dy \end{aligned}$$

$$\Rightarrow \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} = h(y) \quad \text{ist Dichte von } g(X)$$



# Regel des Faulen Statistikers

Beweis (5)

$$\Rightarrow \mathbf{E}(g(X)) = \int_{-\infty}^{\infty} y \cdot h(y) dy = \int_{-\infty}^{\infty} y \cdot \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|} dy$$

Substitution:  $y = g(x)$ ,  $dy = g'(x)dx$

$$\mathbf{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

# Regel des Faulen Statistikers

Beispiele (Fortsetzung). Verwenden die Dichte von  $g(X)$ .

Fortsetzung von Bsp.  $X \sim \mathbf{R}(0, 1)$ ,  $Y = g(X) = e^X$

$$g(x) = e^x, \quad g'(x) = e^x, \quad g^{-1}(y) = \ln y$$

Also

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{1}{e^{\ln y}} = \frac{1}{y}, \quad 1 \leq y \leq e.$$

# Regel des Faulen Statistikers

Beispiele (Fortsetzung). Verwenden die Dichte von  $g(X)$ .

Fortsetzung von Bsp.  $X \sim \mathbf{R}(0, 1)$ ,  $Y = g(X) = e^X$

$$g(x) = e^x, \quad g'(x) = e^x, \quad g^{-1}(y) = \ln y$$

Also

$$h(y) = \frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{1}{e^{\ln y}} = \frac{1}{y}, \quad 1 \leq y \leq e.$$

$$\mathbf{E}(Y) = \int_1^e y \cdot h(y) dy = \int_1^e y \cdot \frac{1}{y} dy = \int_1^e 1 dy = e - 1$$

dasselbe Resultat wie mit der Regel des Faulen Statistikers.

# Regel des Faulen Statistikers

Beispiele (Fortsetzung von Bsp. Gebrochener Stab)

Es war  $X \sim \mathbf{R}(0, 1)$ ,  $Y = g(X) = \max(X, 1 - X)$ .

$g(x) = \max(x, 1 - x)$  ist stückweise differenzierbar.

$$g'(x) = \begin{cases} 1, & x > 0.5 \\ -1, & x < 0.5. \end{cases} \quad \begin{aligned} g^{-1}(y) &= \{y, 1 - y\} \\ g'(g^{-1}(y)) &= \{1, -1\} \end{aligned}$$

$$h(y) = \frac{f(y) + f(1 - y))}{|g'(g^{-1}(y))|} = \frac{1 + 1}{1} = 2, \quad y \in (0.5, 1)$$

$$\mathbf{E}(Y) = \int_{0.5}^1 y \cdot h(y) dy = \int_{0.5}^1 y \cdot 2 dy = 2 \cdot \frac{1}{2} y^2 \Big|_{0.5}^1 = \frac{3}{4}$$

Also wieder dasselbe Resultat wie mit der Regel des Faulen Statistikers.

# Inhalt

- Der Erwartungswert
- Moment und Varianz
- Schiefe und Exzess
- Charakteristische Funktionen

## 7.2 Moment und Varianz

Es sei  $X$  eine zufällige Variable.

### Def. 30 (Moment und Zentrales Moment)

Falls  $\mathbf{E}(|X|^p) < \infty$ , heißt der Erwartungswert  $\mathbf{E}X^p$   $p$ -tes Moment

$$\mathbf{E}X^p = \begin{cases} \int_{-\infty}^{+\infty} x^p \cdot f(x) dx, & \text{falls } X \text{ stetig ist} \\ \sum_{i \in \mathbb{N}} x_i^p \cdot p_i, & \text{falls } X \text{ diskret ist} \end{cases}$$

$\mathbf{E}(X - \mathbf{E}X)^p$  heißt  $p$ -tes zentrales Moment.

# Varianz und Standardabweichung

Def. 31 (Varianz), bez.  $\text{Var } X$  oder  $\sigma_X^2$

Das zweite zentrale Moment  $\mathbf{E}(X - \mathbf{E}X)^2$  nennen wir auch Streuung oder Varianz der Zufallsgröße  $X$ .

Def. 32 (Standardabweichung),  $\sigma, \sigma_X$

$$\sigma = \sqrt{\text{Var}(X)}$$

**Bem.:**  $\text{Var}(X)$ : mittlere quadratische Abweichung zwischen  $X$  und  $\mathbf{E}X$ .

# Varianz

## Satz (Eigenschaften der Varianz):

- ❶ Sei  $c \in \mathbb{R}$ . Wenn  $P(X = c) = 1$ , so  $\text{Var } X = 0$ . Ist umgekehrt  $\text{Var } X = 0$ , so existiert ein  $c \in \mathbb{R}$ , so daß gilt:  $P(X = c) = 1$ .
- ❷ Für beliebige  $c \in \mathbb{R}$  gilt:  $\text{Var } (X + c) = \text{Var } X$ .
- ❸ Für beliebige  $a \in \mathbb{R}$  gilt:  $\text{Var } (a \cdot X) = a^2 \cdot \text{Var } X$ .
- ❹ Für zwei zufällige Variablen  $X_1$  und  $X_2$  gilt:  
$$\text{Var } (X_1 + X_2) = \text{Var } X_1 + \text{Var } X_2 + 2 \cdot \text{cov } (X_1, X_2).$$



# Eigenschaften der Varianz

## Beweis (1)

Es seien  $X$ ,  $X_1$  und  $X_2$  beliebige zufällige Variablen.  $a, c \in \mathbb{R}$  seien ebenfalls beliebig gewählt. Die folgenden Aussagen folgen aus dem Satz über die Eigenschaften des Erwartungswertes.

① Es gelte:  $P(X = c) = 1$ . Daraus folgt  $\mathbf{E}X = c$ .

$$\text{Var } X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X - c)^2 = \mathbf{E}(c - c)^2 = 0$$

Es sei nun  $\text{Var } X = 0 = \mathbf{E}(X - \mathbf{E}X)^2 = 0$ . Allgemein gilt für  $c \in \mathbb{R}$ :  $\mathbf{E}(X - c)^2 \geq 0$ . Also,  $P(X - EX = 0) = 1$ . und  $c := \mathbf{E}X$  leistet das Verlangte.

# Eigenschaften der Varianz

Beweis (2)

2

$$\text{Var}(X + c) = \mathbf{E}(X + c - \mathbf{E}(X + c))^2$$

# Eigenschaften der Varianz

Beweis (2)

2

$$\begin{aligned}\text{Var}(X + c) &= \mathbf{E}(X + c - \mathbf{E}(X + c))^2 \\ &= \mathbf{E}(X + c - \mathbf{E}X - \mathbf{E}c)^2 \\ &= \mathbf{E}(X + c - \mathbf{E}X - c)^2 \\ &= \mathbf{E}(X - \mathbf{E}X)^2 = \text{Var } X\end{aligned}$$

# Eigenschaften der Varianz

Beweis (3)

3

$$\text{Var}(a \cdot X) = \mathbf{E}(a \cdot X - \mathbf{E}(a \cdot X))^2$$

# Eigenschaften der Varianz

Beweis (3)

3

$$\begin{aligned}\text{Var}(a \cdot X) &= \mathbf{E}(a \cdot X - \mathbf{E}(a \cdot X))^2 \\ &= \mathbf{E}(a \cdot X - a \cdot \mathbf{E}X)^2 \\ &= \mathbf{E}(a \cdot (X - \mathbf{E}X))^2 \\ &= \mathbf{E}(a^2 \cdot (X - \mathbf{E}X)^2) \\ &= a^2 \cdot \mathbf{E}(X - \mathbf{E}X)^2 \\ &= a^2 \cdot \text{Var } X\end{aligned}$$

# Eigenschaften der Varianz

Beweis (4)

4

$$\text{Var}(X_1 + X_2) = \mathbf{E}(X_1 + X_2 - \mathbf{E}(X_1 + X_2))^2$$

# Eigenschaften der Varianz

Beweis (4)

4

$$\begin{aligned}\text{Var}(X_1 + X_2) &= \mathbf{E}(X_1 + X_2 - \mathbf{E}(X_1 + X_2))^2 \\ &= \mathbf{E}(X_1 + X_2 - \mathbf{E}X_1 - \mathbf{E}X_2)^2\end{aligned}$$

# Eigenschaften der Varianz

Beweis (4)

4

$$\begin{aligned}\text{Var}(X_1 + X_2) &= \mathbf{E}(X_1 + X_2 - \mathbf{E}(X_1 + X_2))^2 \\ &= \mathbf{E}(X_1 + X_2 - \mathbf{E}X_1 - \mathbf{E}X_2)^2 \\ &= \mathbf{E}((X_1 - \mathbf{E}X_1) + (X_2 - \mathbf{E}X_2))^2\end{aligned}$$



# Eigenschaften der Varianz

Beweis (4)

4

$$\begin{aligned}\text{Var}(X_1 + X_2) &= \mathbf{E}(X_1 + X_2 - \mathbf{E}(X_1 + X_2))^2 \\&= \mathbf{E}(X_1 + X_2 - \mathbf{E}X_1 - \mathbf{E}X_2)^2 \\&= \mathbf{E}((X_1 - \mathbf{E}X_1) + (X_2 - \mathbf{E}X_2))^2 \\&= \mathbf{E}((X_1 - \mathbf{E}X_1)^2 + (X_2 - \mathbf{E}X_2)^2 \\&\quad + 2 \cdot (X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2)) \\&= \mathbf{E}(X_1 - \mathbf{E}X_1)^2 + \mathbf{E}(X_2 - \mathbf{E}X_2)^2 \\&\quad + 2 \cdot \underbrace{\mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))}\end{aligned}$$

# Eigenschaften der Varianz

Beweis (4)

4

$$\begin{aligned}\text{Var}(X_1 + X_2) &= \mathbf{E}(X_1 + X_2 - \mathbf{E}(X_1 + X_2))^2 \\&= \mathbf{E}(X_1 + X_2 - \mathbf{E}X_1 - \mathbf{E}X_2)^2 \\&= \mathbf{E}((X_1 - \mathbf{E}X_1) + (X_2 - \mathbf{E}X_2))^2 \\&= \mathbf{E}((X_1 - \mathbf{E}X_1)^2 + (X_2 - \mathbf{E}X_2)^2 \\&\quad + 2 \cdot (X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2)) \\&= \mathbf{E}(X_1 - \mathbf{E}X_1)^2 + \mathbf{E}(X_2 - \mathbf{E}X_2)^2 \\&\quad + 2 \cdot \underbrace{\mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))}\end{aligned}$$

# Kovarianz und Unabhängigkeit

Def. 33 Kovarianz der zufälligen Variablen  $X_1$  und  $X_2$

$$\text{cov}(X_1, X_2) := \mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))$$

$$\text{cov}(X_1, X_2) =$$

$$= \mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))$$

$$= \mathbf{E}(X_1 \cdot X_2 - X_1 \cdot \mathbf{E}X_2 - X_2 \cdot \mathbf{E}X_1 + \mathbf{E}X_1 \cdot \mathbf{E}X_2)$$

$$= \mathbf{E}(X_1 \cdot X_2) - \mathbf{E}(X_1 \cdot \mathbf{E}X_2) - \mathbf{E}(X_2 \cdot \mathbf{E}X_1) + \mathbf{E}X_1 \cdot \mathbf{E}X_2$$

$$= \mathbf{E}(X_1 \cdot X_2) - \mathbf{E}X_1 \cdot \mathbf{E}X_2$$

# Kovarianz und Unabhängigkeit

Erinnerung:

## Def. 34 Unabhängigkeit

Zwei Zufallsvariablen  $X_1$  und  $X_2$  heißen unabhängig, falls für alle  $x_1, x_2 \in \mathbf{R}$  gilt:

$$P(X_1 < x_1, X_2 < x_2) = P(X_1 < x_1) \cdot P(X_2 < x_2)$$

## Lemma

Es seien  $X_1$  und  $X_2$  zwei unabhängige Zufallsgrößen. Dann gilt:

$$\text{cov}(X_1, X_2) = 0.$$

Def. 35 Zwei Zufallsvariablen  $X_1$  und  $X_2$  heißen unkorreliert falls  $\text{cov}(X_1, X_2) = 0$ .

# Kovarianz und Unabhängigkeit

Beweis des Lemmas (1)

**Beweis:** Wir betrachten den zufälligen Vektor  $\mathbf{X} = (X_1, X_2)^T$  und betrachten nur den Fall, dass die beiden Zufallsgrößen  $X_1$  und  $X_2$  stetig sind. Für den diskreten Fall verfährt man analog.

Es sei  $f(x_1, x_2)$  die Dichtefunktion des zufälligen Vektors  $\mathbf{X}$ .

Wir definieren eine Funktion  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  durch:

$$g(X_1, X_2) := (X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2).$$

Offenbar,

$$\text{cov}(X_1, X_2) = \mathbf{E}g(X_1, X_2).$$

# Kovarianz und Unabhängigkeit

Beweis des Lemmas (2)

Außerdem ist:

$$\mathbf{E}g(X_1, X_2) = \int_{\mathbb{R}^2} (x_1 - \mathbf{E}X_1) \cdot (x_2 - \mathbf{E}X_2) \cdot f(x_1, x_2) dx_1 dx_2.$$

Nach Voraussetzung sind die zufälligen Variablen  $X_1$  und  $X_2$  unabhängig, also

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

(das folgt unmittelbar durch zweimaliges Differenzieren, nach  $x_1$  und nach  $x_2$ , der Gleichung  $F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1)F_{X_2}(x_2)$ )

Somit gilt dann:

# Kovarianz und Unabhängigkeit

Beweis des Lemmas (3)

$$\text{cov}(X_1, X_2) =$$

$$= \int_{\mathbb{R}^2} (x_1 - \mathbf{E}X_1) \cdot (x_2 - \mathbf{E}X_2) \cdot f_{X_1}(x_1) \cdot f_{X_2}(x_2) dx_1 dx_2$$

$$= \int_{\mathbb{R}} (x_1 - \mathbf{E}X_1) \cdot f_{X_1}(x_1) dx_1 \cdot \int_{\mathbb{R}} (x_2 - \mathbf{E}X_2) \cdot f_{X_2}(x_2) dx_2$$

$$= \mathbf{E}(X_1 - \mathbf{E}X_1) \cdot \mathbf{E}(X_2 - \mathbf{E}X_2)$$

$$= 0$$

# Korrelation und Unabhängigkeit

Die Umkehrung der Aussage des Lemmas gilt im allgemeinen nicht, wie das folgende Beispiel zeigt:

Es sei  $X_1 \sim R(0, \pi)$

$$f_{X_1}(x) = \begin{cases} \frac{1}{\pi} & , \text{ falls } 0 \leq x < \pi \\ 0 & , \text{ sonst} \end{cases} .$$

Die Zufallsgröße  $X_2$  definieren wir durch  $X_2 = \sin X_1$ . Offenbar,  $X_1$  und  $X_2$  sind streng abhängig.

Wir berechnen die Kovarianz.



# Korrelation und Unabhängigkeit

Beispiel (Fortsetzung, 1)

Nun gilt für die Erwartungswerte  $\mathbf{E}X_1$  und  $\mathbf{E}X_2$ :

$$\begin{aligned}\mathbf{E}X_1 &= \int_{-\infty}^{+\infty} x \cdot f_{X_1}(x) dx = \int_0^{\pi} x \cdot \frac{1}{\pi} dx \\ &= \frac{1}{\pi} \cdot \left[ \frac{x^2}{2} \right]_0^{\pi} = \frac{1}{\pi} \cdot \frac{\pi^2}{2} = \frac{\pi}{2} \\ \mathbf{E}X_2 &= \mathbf{E}(\sin X_1) = \int_{-\infty}^{+\infty} \sin x \cdot f_{X_1}(x) dx \\ &= \int_0^{\pi} \sin x \cdot \frac{1}{\pi} dx = \frac{1}{\pi} \cdot [-\cos x]_0^{\pi} = \frac{2}{\pi}\end{aligned}$$

# Korrelation und Unabhängigkeit

Beispiel (Fortsetzung, 2)

Für den Erwartungswert  $\mathbf{E}(X_1 \cdot X_2)$  gilt nach der Regel des Faulen Statistikers

$$\begin{aligned}\mathbf{E}(X_1 \cdot X_2) &= \mathbf{E}(X_1 \cdot \sin X_1) = \int_0^{\pi} x \cdot \sin x \cdot \frac{1}{\pi} dx \\&= \left[ -\frac{1}{\pi} \cdot x \cdot \cos x \right]_0^{\pi} + \frac{1}{\pi} \cdot \underbrace{\int_0^{\pi} \cos x \, dx}_{=0} \\&= -\frac{1}{\pi} \cdot (-1)\pi - 0 = 1\end{aligned}$$

Wir setzen alle diese Werte in die Ausgangsgleichung ein und erhalten:

# Korrelation und Unabhängigkeit

Beispiel (Fortsetzung, 3)

$$\begin{aligned}\text{COV}(X_1, X_2) &= \mathbf{E}(X_1 \cdot X_2) - \mathbf{E}X_1 \cdot \mathbf{E}X_2 \\ &= 1 - \frac{\pi}{2} \cdot \frac{2}{\pi} = 0\end{aligned}$$

Trotz der Abhängigkeit der beiden Zufallsgrößen  $X_1$  und  $X_2$  ist ihre Kovarianz gleich Null.

# Korrelation und Unabhängigkeit

Beispiel (Fortsetzung, 3)

$$\begin{aligned}\text{cov}(X_1, X_2) &= \mathbf{E}(X_1 \cdot X_2) - \mathbf{E}X_1 \cdot \mathbf{E}X_2 \\ &= 1 - \frac{\pi}{2} \cdot \frac{2}{\pi} = 0\end{aligned}$$

Trotz der Abhängigkeit der beiden Zufallsgrößen  $X_1$  und  $X_2$  ist ihre Kovarianz gleich Null.

## Folgerung

Falls zwei zufällige Variablen  $X_1$  und  $X_2$  unabhängig sind, gilt für die Varianz ihrer Summe:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

# Varianz, Beispiele (1)

a) Poisson-Verteilung,  $X \sim Poi(\lambda)$

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

$$\text{Var}(X) = \mathbf{E}(X - \mathbf{E}X)^2 = \sum_{i=0}^{\infty} (i - \lambda)^2 p_i$$

# Varianz, Beispiele (1)

a) Poisson-Verteilung,  $X \sim Poi(\lambda)$

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

$$\begin{aligned} \text{Var}(X) &= \mathbf{E}(X - \mathbf{E}X)^2 = \sum_{i=0}^{\infty} (i - \lambda)^2 p_i \\ &= \sum_{i=2}^{\infty} i \cdot (i - 1) p_i + \sum_{i=0}^{\infty} i p_i - 2\lambda \sum_{i=0}^{\infty} i p_i + \lambda^2 \sum_{i=0}^{\infty} p_i \end{aligned}$$

# Varianz, Beispiele (1)

## a) Poisson-Verteilung, $X \sim \text{Poi}(\lambda)$

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

$$\begin{aligned} \text{Var}(X) &= \mathbf{E}(X - \mathbf{E}X)^2 = \sum_{i=0}^{\infty} (i - \lambda)^2 p_i \\ &= \sum_{i=2}^{\infty} i \cdot (i - 1) p_i + \sum_{i=0}^{\infty} i p_i - 2\lambda \sum_{i=0}^{\infty} i p_i + \lambda^2 \sum_{i=0}^{\infty} p_i \\ &= \lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} e^{-\lambda} + \lambda - 2\lambda^2 + \lambda^2 = \lambda. \end{aligned}$$

## Varianz, Beispiele (2)

b) Binomialverteilung,  $X \sim B(n, p)$ .

$$\text{Var}(X) = np(1 - p).$$

(ohne Beweis, ÜA)



## Varianz, Beispiele (3)

c) Gleichverteilung auf  $(a, b)$ ,  $X \sim R(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{sonst.} \end{cases} \quad \mathbf{E}X = \frac{a+b}{2}.$$

## Varianz, Beispiele (3)

### c) Gleichverteilung auf $(a, b)$ , $X \sim R(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{sonst.} \end{cases} \quad \mathbf{E}X = \frac{a+b}{2}.$$

$$\begin{aligned} \mathbf{E}X^2 &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3} x^3 \Big|_a^b \cdot \frac{1}{b-a} \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

## Varianz, Beispiele (3)

### c) Gleichverteilung auf $(a, b)$ , $X \sim R(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{sonst.} \end{cases} \quad \mathbf{E}X = \frac{a+b}{2}.$$

$$\begin{aligned} \mathbf{E}X^2 &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3} x^3 \Big|_a^b \cdot \frac{1}{b-a} \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \frac{1}{12} (4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2) \\ &= \frac{1}{12} (a^2 - 2ab + b^2) = \frac{(b-a)^2}{12}. \end{aligned}$$

# Varianz, Beispiele (4)

## d) Exponentialverteilung

$$f(x) = \begin{cases} \lambda e^{-\lambda \cdot x} & \text{falls } x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

$$\mathbf{E}X = \frac{1}{\lambda}.$$

$$\mathbf{E}X^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda \cdot x} dx = \frac{2}{\lambda^2} \quad (\text{ÜA}).$$

$$\text{Var}(X) = \frac{1}{\lambda^2}.$$

# Varianz, Beispiele (5a)

## e) Normalverteilung

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

$$\mathbf{E}(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

# Varianz, Beispiele (5a)

## e) Normalverteilung

$$\begin{aligned}f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\ \mathbf{E}(X - \mu)^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \sigma^2 \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\end{aligned}$$

# Varianz, Beispiele (5a)

## e) Normalverteilung

$$\begin{aligned}f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\ \mathbf{E}(X - \mu)^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \sigma^2 \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \sigma^2 \int_{-\infty}^{\infty} (-t) \left(-t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\right) dt\end{aligned}$$

# Varianz, Beispiele (5b)

## e) Normalverteilung

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left( -te^{-t^2/2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1)e^{-\frac{t^2}{2}} dt \right)$$



# Varianz, Beispiele (5b)

## e) Normalverteilung

$$\begin{aligned} &= \frac{\sigma^2}{\sqrt{2\pi}} \left( -te^{-t^2/2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1)e^{-\frac{t^2}{2}} dt \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \\ &= \sigma^2. \end{aligned}$$

$$t = \frac{x-\mu}{\sigma}, \quad \sigma dt = dx$$

# Varianz, Beispiele (5b)

## e) Normalverteilung

$$\begin{aligned} &= \frac{\sigma^2}{\sqrt{2\pi}} \left( -te^{-t^2/2} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1)e^{-\frac{t^2}{2}} dt \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \\ &= \sigma^2. \end{aligned}$$

$$t = \frac{x-\mu}{\sigma}, \quad \sigma dt = dx$$

Bei Normalverteilung sind also die Parameter  $\mu$  und  $\sigma^2$  Erwartungswert und Varianz.

# Inhalt

- Der Erwartungswert
- Moment und Varianz
- Schiefe und Exzess
- Charakteristische Funktionen

## 7.3 Schiefe und Exzess

Angenommen, das 4. Moment existiert.

Def. 36 (Schiefe und Kurtosis)

$$\sigma_X = \sqrt{\text{Var}(X)} \quad (\text{Standardabweichung})$$

$$\text{Schiefe} \quad \gamma_1 = \frac{\mathbf{E}(X - \mathbf{E}X)^3}{(\text{Var}X)^{3/2}}$$

$$\text{Kurtosis} \quad \gamma_2 = \frac{\mathbf{E}(X - \mathbf{E}X)^4}{(\text{Var}X)^2}$$

Exzess:  $\gamma_2 - 3$ .

# Schiefe und Exzess, Versuch einer Klassifikation

$\gamma_1 > 0$ : rechtsschiefe Verteilung

$\gamma_1 = 0$ : symmetrische Verteilung

$\gamma_1 < 0$ : linksschiefe Verteilung

$\gamma_2 > 3$ : starke Tails

$\gamma_2 = 3$ : Wölbung wie bei Normalverteilung

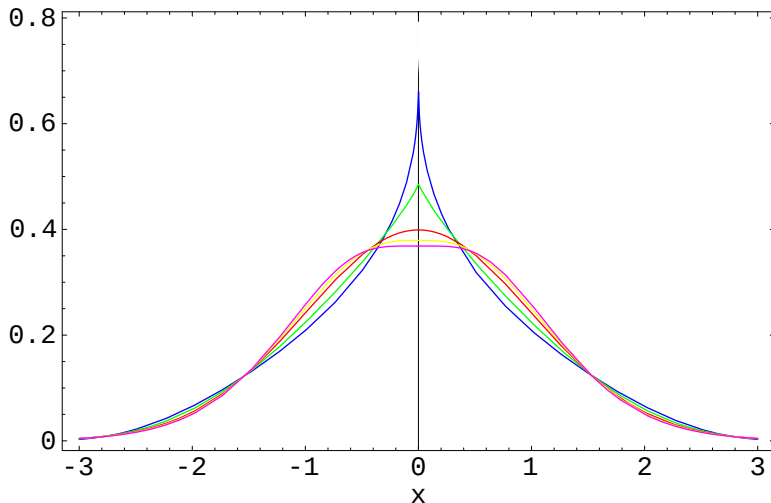
$\gamma_2 < 3$ : schwache Tails

**Bem.:** Diese Klassifikation ist recht vage. Es gibt mehrere Verteilungen mit gleichem Erwartungswert, gleicher Varianz, gleicher Schiefe und gleicher Kurtosis, die aber recht unterschiedlich aussehen.

# Schiefe und Exzess

$$\mathbb{E}(X) = 0, \text{var}(X) = 1, \gamma_1 = 0, \gamma_2 = 3$$

Dichte



# Inhalt

- Der Erwartungswert
- Moment und Varianz
- Schiefe und Exzess
- Charakteristische Funktionen

Charakteristische Funktionen werden wir brauchen um einen wichtigen Satz, den Zentralen Grenzwertsatz (siehe unten), zu beweisen.

Sei  $X$  Zufallsvariable mit Dichtefunktion  $f_X$  (falls  $X$  stetig) oder Wkt.funktion  $p_j$  (falls  $X$  diskret).

Def. 37 (charakteristische Funktion von  $X$ )

$$\phi_X(t) := \mathbf{E}e^{itX} = \begin{cases} \int_{-\infty}^{\infty} e^{itx} f_X(x) dx & \text{falls } X \text{ stetig} \\ \sum_{j=1}^{\infty} e^{itx_j} p_j & \text{falls } X \text{ diskret} \end{cases}$$

**Bem.:** Die Funktion  $\phi_X$  ist (bis auf den Faktor  $\sqrt{2\pi}$ ) die Fourier-Transformierte von  $f_X$ .

**Bem.:** Die charakterische Funktion existiert.



# Charakteristische Funktionen

## Satz (Eigenschaften)

(i)  $\phi_X(t)$  ist in  $-\infty < t < \infty$  gleichmäßig stetig.

$$\begin{aligned} |\phi_X(t)| &\leq 1 & \phi_X(0) &= 1 \\ \phi_X(-t) &= \overline{\phi_X(t)} \end{aligned}$$

(ii) Die Zufallsvariable  $Y = aX + b$  hat die charakteristische Funktion

$$\phi_Y(t) = \phi_X(at)e^{ibt}$$

(iii)  $\phi_X(t)$  ist reellwertig  $\Leftrightarrow X$  bzgl.  $x = 0$  symmetrisch ist.

**Beweis:** ÜA, Eigenschaften der Fkt.  $e^{it}$ .

# Charakteristische Funktionen

## Satz (Multiplikationssatz)

Seien die Zufallsvariablen  $X_1$  und  $X_2$  unabhängig mit den charakteristischen Funktionen  $\phi_1$  und  $\phi_2$ . Dann hat die Zufallsvariable  $X_1 + X_2$  die charakteristische Funktion  $\phi_1 \cdot \phi_2$ .

**Beweis:** Es gilt:

$$\phi_{X_1+X_2}(t) = \mathbf{E}e^{it(X_1+X_2)} = \mathbf{E}e^{itX_1} \cdot \mathbf{E}e^{itX_2} = \phi_1(t) \cdot \phi_2(t)$$



# Charakteristische Funktionen

## Satz (Eindeutigkeitssatz)

Die Beziehung  $F_X \Leftrightarrow \phi_X$  ist eineindeutig.

Für  $X$  stetig gilt:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt$$

Für  $X$  diskret gilt:

$$p_j = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T e^{-itx_j} \phi_X(t) dt$$

**Beweis:** siehe z.B. Günther, Grundkurs Analysis, Teil 3.

□

# Charakteristische Funktionen

## Satz (Konvergenzsatz)

Seien  $X_n$  Zufallsvariablen mit  $X_n \sim F_n$ . Dann gilt

$$F_n \rightarrow F \quad \Leftrightarrow \quad \phi_n \rightarrow \phi, \quad \phi \text{ stetig in } t = 0.$$

# Charakteristische Funktionen

Wozu brauchen wir sie?

## Zum Beweis des Zentralen Grenzwertsatzes:

Die Summe von unabhängigen, identisch verteilten Zufallsgrößen ist (oft) asymptotisch normalverteilt (siehe Abschnitt Grenzwertsätze).

- 1. charakteristische Funktion der Summe (Multiplikationssatz)
- 2. diese konvergiert gegen charakteristische Funktion der Normalverteilung (s. unten)
- 3. Konvergenz der Summe folgt aus dem Konvergenzsatz

# Charakteristische Funktionen

## Satz (Erzeugung der Momente)

Sei  $\mathbf{E}X^k < \infty$ . Dann gilt:

$$\alpha_k := \mathbf{E}X^k = \frac{1}{i^k} \phi_X^{(k)}(0)$$

**Beweis:** Vertauschen von Integration und Differentiation. □

Die charakteristische Funktion hat also die Taylor-Entwicklung

$$\phi_X(t) = \mathbf{E}e^{itX} = \mathbf{E}\left(\sum_{j=0}^{\infty} \frac{(it)^j}{j!} X^j\right) = \sum_{j=0}^k \alpha_j \frac{(it)^j}{j!} + o(t^k), \quad t \rightarrow 0.$$

# Charakteristische Funktionen

$$X \sim \mathcal{N}(0, 1)$$

$$\begin{aligned} \mathbf{E}e^{itX} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2 - 2itx + (it)^2 - (it)^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-it)^2}{2}} dx \quad z = x - it \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{-\infty+it}^{\infty+it} e^{-\frac{z^2}{2}} dz = e^{-\frac{t^2}{2}}. \end{aligned}$$

$$Y \sim \mathcal{N}(\mu, \sigma^2):$$

$$\mathbf{E}e^{itY} = \mathbf{E}e^{it(\sigma X + \mu)} = e^{it\mu} \phi_X(\sigma t)$$

# Inhalt

- Einführung
- Gedächtnislosigkeit
- Zuverlässigkeitsmodelle
- Bedienungstheorie
- Zusammenfassung (Exponentialverteilung)



**Def. 38 (Exponentialverteilung),  $X \sim EX(\lambda)$** 

Sei  $X$  eine Zufallsvariable mit Werten in  $[0, \infty)$ . Sie heißt exponentialverteilt mit dem Parameter  $\lambda$ ,  $\lambda > 0$ , falls die Verteilungsfunktion beschrieben wird durch

$$F(t) = P(X < t) = \begin{cases} 1 - e^{-\lambda t} & \text{falls } t \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Die Dichte der Exponentialverteilung ist

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{falls } t \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

# Die Exponentialverteilung

Erwartungswert

$$\begin{aligned}
 \mathbf{E}X &= \int_0^{\infty} x \cdot f(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx \\
 &\quad \quad \quad u \quad v' \\
 &= x \cdot (-e^{-\lambda x}) \Big|_0^{\infty} - \int_0^{\infty} 1 \cdot (-e^{-\lambda x}) dx \\
 &\quad \quad \quad u \quad v \quad \quad \quad u' \quad v \\
 &= 0 + \int_0^{\infty} e^{-\lambda x} dx = \frac{-1}{\lambda} \cdot e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}.
 \end{aligned}$$

# Die Exponentialverteilung

Varianz, Schiefe, Exzess

$$\text{Var}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\sigma_X = \frac{1}{\lambda} \quad (\text{Standardabweichung})$$

$$\text{Schiefe} = \frac{\mathbf{E}(X - \mathbf{E}X)^3}{(\text{Var}X)^{3/2}} = 2$$

$$\text{Kurtosis} = \frac{\mathbf{E}(X - \mathbf{E}X)^4}{(\text{Var}X)^2} = 9$$

# Die Exponentialverteilung

## Beispiel

### Die zufällige Wartezeit eines Kunden

am Schalter sei exponentialverteilt mit einem Erwartungswert von 10 min.

Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens 15 min. warten müssen?

# Die Exponentialverteilung

## Beispiel

### Die zufällige Wartezeit eines Kunden

am Schalter sei exponentialverteilt mit einem Erwartungswert von 10 min.

Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens 15 min. warten müssen?

$X$ : zufällige Wartezeit eines Kunden am Schalter,

$X \sim \text{Exp}(\lambda)$ ,  $\lambda = \frac{1}{10}$ . Frage:  $P(X > 15)$  ?

$$\begin{aligned} P(X > 15) &= e^{-15\lambda} \\ &= e^{-1.5} \approx 0.220. \end{aligned}$$

# Inhalt

- Einführung
- Gedächtnislosigkeit
- Zuverlässigkeitsmodelle
- Bedienungstheorie
- Zusammenfassung (Exponentialverteilung)

### Def. 39 (Gedächtnislosigkeit)

Eine Verteilung  $P$  (mit Verteilungsfunktion  $F$ ) heißt gedächtnislos, wenn für alle  $s, t \geq 0$ , gilt:

$$P(X > s + t | X > t) = P(X > s).$$

**Bem.:** Bei stetigen Verteilungen ist das äquivalent zu

$$P(X \geq s + t | X \geq t) = P(X \geq s).$$

Es gilt (Definition der bedingten Wahrscheinlichkeit)

$$\begin{aligned} P(X \geq s + t | X \geq t) &= \frac{P(\{X \geq s + t\} \cap \{X \geq t\})}{P(X \geq t)} \\ &= \frac{P(X \geq s + t)}{P(X \geq t)}. \end{aligned}$$

## Gedächtnislosigkeit (2)

Eine Verteilung(sfunktion) ist also gedächtnislos, genau dann wenn

$$\frac{P(X \geq s + t)}{P(X \geq t)} = P(X \geq s)$$

bzw.

$$\frac{1 - F(s + t)}{1 - F(t)} = 1 - F(s).$$

Def. 40 Überlebensfunktion (oder Zuverlässigkeitsfunktion)

$$G(t) = 1 - F(t)$$



## Gedächtnislosigkeit (3)

Die Verteilungsfunktion  $F$  (mit der Überlebensfunktion  $G$ ) ist also gedächtnislos genau dann wenn

$$G(s + t) = G(s) \cdot G(t) \quad \text{für alle } s, t \geq 0$$

Cauchy- Funktionalgleichung

# Gedächtnislosigkeit (4)

**Satz:** Die Exponentialverteilung ist gedächtnislos.

**Beweis:** Die Verteilungsfunktion ist

$$F(t) = P(X < t) = \begin{cases} 1 - e^{-\lambda t} & \text{falls } t \geq 0 \\ 0 & \text{sonst,} \end{cases}$$

und die Überlebensfunktion

$$G(t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

Folglich erhalten wir

$$G(s + t) = e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t} = G(s) \cdot G(t).$$

# Gedächtnislosigkeit (5)

Satz: Sei  $F$  eine stetige Verteilungsfunktion mit  $F(0) = 0$  und  $G(t) = 1 - F(t)$ .

Es gelte die Cauchy-Funktionalgleichung

$$G(s + t) = G(s) \cdot G(t) \quad \text{für alle } s, t \geq 0. \quad (1)$$

Dann gilt für alle  $t, t > 0$ ,

$$F(t) = 1 - e^{-\lambda t},$$

wobei  $\lambda > 0$ . D.h.  $F$  ist Exponential-Verteilungsfunktion.

# Gedächtnislosigkeit (6)

Beweis des Satzes

① Es gilt:

$$G(t) = G\left(\frac{t}{2} + \frac{t}{2}\right) = \left(G\left(\frac{t}{2}\right)\right)^2 \geq 0,$$

d.h.  $G(t) \geq 0$  für alle  $t$ .

Angenommen, es existiert ein  $t_0$  mit  $G(t_0) = 0$ , dann folgt:

$$G(t) = G(t - t_0 + t_0) = G(t - t_0) \cdot G(t_0) = 0$$

für alle  $t$ , d.h. wir erhalten die triviale Lösung für die obige Cauchy-Funktionalgleichung, die jedoch wegen  $G(0) = 1 - F(0) = 1$  nicht zugelassen ist.

## Gedächtnislosigkeit (7)

2 Es gilt also  $G(t) > 0$  für alle  $t$ .

Sei  $m, m > 0$ , eine natürliche Zahl. Dann folgt aus (1) für alle  $t > 0$ :

$$G(t) = G(\underbrace{\frac{t}{m} + \dots + \frac{t}{m}}_{m \text{ mal}}) = \left(G\left(\frac{t}{m}\right)\right)^m,$$

insbesondere

$$G(1) = \left(G\left(\frac{1}{m}\right)\right)^m \quad \text{oder} \quad G\left(\frac{1}{m}\right) = \left(G(1)\right)^{\frac{1}{m}}$$

# Gedächtnislosigkeit (8)

3 Für rationale Zahlen  $r = \frac{n}{m}$  erhalten wir

$$\begin{aligned} G(r) &= G\left(\frac{n}{m}\right) = G\left(\underbrace{\frac{1}{m} + \dots + \frac{1}{m}}_{n \text{ mal}}\right) \\ &= \left(G\left(\frac{1}{m}\right)\right)^n \\ &= \left(G(1)\right)^{\frac{n}{m}} \\ &= \left(G(1)\right)^r. \end{aligned}$$

## Gedächtnislosigkeit (9)

- ④ Da die Funktion  $(G(1))^t$  stetig ist auf  $\mathbb{R}^+$  folgt für alle  $t > 0$ :

$$G(t) = G(1)^t = e^{t \cdot \ln(G(1))}$$

- ⑤ Wir setzen  $\lambda := -\ln G(1)$ .

Da  $F$  als Verteilungsfunktion monoton wachsend ist, ist  $G$  monoton fallend, d.h.  $\ln G(1) < 0$  und  $\lambda > 0$ . Wir erhalten demnach

$$G(t) = e^{-\lambda \cdot t},$$

also

$$F(t) = 1 - e^{-\lambda \cdot t}.$$

# Gedächtnislosigkeit (10)

**Bem.:** Unter den diskreten Verteilungen hat nur die geometrische Verteilung diese Eigenschaft (siehe dort)

## Fortsetzung von Beispiel 1

Der Kunde hat schon 10 min. gewartet.

Wie groß ist die Wahrscheinlichkeit, dass er insgesamt länger als 15 min. warten muss?



# Gedächtnislosigkeit (10)

**Bem.:** Unter den diskreten Verteilungen hat nur die geometrische Verteilung diese Eigenschaft (siehe dort)

## Fortsetzung von Beispiel 1

Der Kunde hat schon 10 min. gewartet.

Wie groß ist die Wahrscheinlichkeit, dass er insgesamt länger als 15 min. warten muss?

$$\begin{aligned} P(X > 15 | X > 10) &= P(X > 5) = e^{-5\lambda} = e^{-0.5} \\ &\approx 0.604. \end{aligned}$$

## Gedächtnislosigkeit (12)

Postschalter mit 2 Personen besetzt. Die Bedienungszeit sei zufällig, exponential verteilt, mit Erwartungswert  $\frac{1}{\lambda}$ . Es werden gerade zwei Kunden bedient, Sie sind der nächste.

Wahrscheinlichkeit dafür, dass Sie nicht der letzte der 3 Kunden sind?

## Gedächtnislosigkeit (12)

Postschalter mit 2 Personen besetzt. Die Bedienungszeit sei zufällig, exponential verteilt, mit Erwartungswert  $\frac{1}{\lambda}$ . Es werden gerade zwei Kunden bedient, Sie sind der nächste.

Wahrscheinlichkeit dafür, dass Sie nicht der letzte der 3 Kunden sind? Antwort: Sie werden bedient, sobald der erste Platz frei wird. Wegen der Gedächtnislosigkeit der Exponentialverteilung hat die Bedienungszeit des anderen Kunden dieselbe Verteilung wie Ihre.

$$P = 0.5.$$

# Inhalt

- Einführung
- Gedächtnislosigkeit
- Zuverlässigkeitsmodelle
- Bedienungstheorie
- Zusammenfassung (Exponentialverteilung)

### Def. 41 Die Zuverlässigkeit eines Systems $\zeta$

ist die Wahrscheinlichkeit, dass das System zum Zeitpunkt  $t$  intakt ist:

$$\text{Rel}(\zeta) = P(X \geq t).$$

Annahmen:

Das System besteht aus mehreren Komponenten

Die Komponenten sind unabhängig

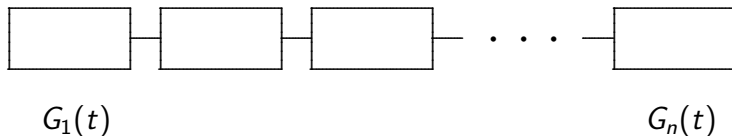
$$X_i \sim \text{Exp}(\lambda_i).$$

# Zuverlässigkeitsmodelle

- Reihensystem
- Parallelsystem
- $k$  aus  $n$  System
- Provisionswahrscheinlichkeit
- Faltung

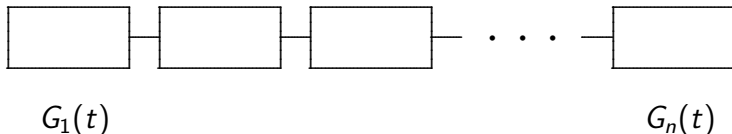
# Zuverlässigkeitsmodelle

Reihensystem  $\zeta_R$



# Zuverlässigkeitsmodelle

## Reihensystem $\zeta_R$



$$\begin{aligned}\text{Rel}(\zeta_R) &= P(X_R \geq t) = P(X_1 \geq t, \dots, X_n \geq t) = \\ &= \prod_{i=1}^n P(X_i \geq t) = \prod_{i=1}^n G_i(t) = \\ &= \prod_{i=1}^n e^{-\lambda_i t} = \exp\left(-\sum_{i=1}^n \lambda_i t\right).\end{aligned}$$



# Reihensystem

Die zufällige Lebensdauer  $X_R$  des Reihensystems ist

$$X_R \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right).$$

# Reihensystem

Die zufällige Lebensdauer  $X_R$  des Reihensystems ist

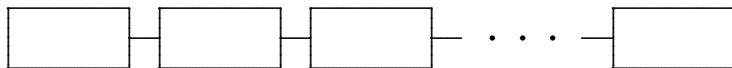
$$X_R \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right).$$

Die mittlere Lebensdauer des Reihensystems ist

$$\mathbf{E}X_R = \frac{1}{\sum_{i=1}^n \lambda_i}.$$

# Zuverlässigkeitsmodelle

## Reihensystem:



$$\text{Rel}(\zeta_R) = e^{\left(-\sum_{i=1}^n \lambda_i t\right)},$$

$$\begin{aligned} n \rightarrow \infty, \lambda_i = \lambda : \quad \text{Rel}(\zeta_R) &\rightarrow 0. \\ n \rightarrow \infty, \sum_{i=1}^n \lambda_{i,n} \rightarrow \lambda < \infty : \quad \text{Rel}(\zeta_R) &\rightarrow e^{-\lambda t} \end{aligned}$$

Die Lebensdauer  $X_R$  des Reihensystems ist asymptotisch wieder exponentialverteilt.

Die Exponentialverteilung ist eine sogenannte Extremwertverteilung.

# Zuverlässigkeitsmodelle

## Reihensystem

**Bem.:** Die Lebensdauer  $X_R$  des Reihensystems kann beschrieben werden durch

$$X_R = \min_i X_i.$$

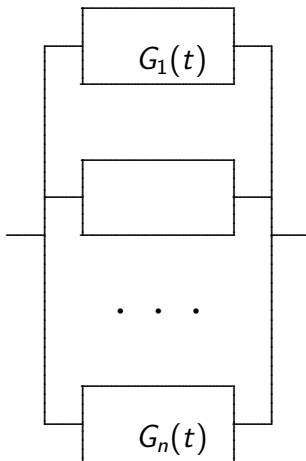
Die Zufallsvariable  $X_R$  hat oft (auch dann wenn nicht  $X_i \sim \text{Exp}(\lambda)$ ) asymptotisch eine Weibull-Verteilung mit der Dichte

$$f(t) = b(\lambda t)^{b-1} e^{-(\lambda t)^b}, \quad t > 0, b > 0, \lambda > 0.$$

Das ist dann der Fall, wenn die Dichte der unabhängig und identisch verteilten Zufallsvariablen  $X_i$  'kurze' Tails hat.

# Zuverlässigkeitsmodelle

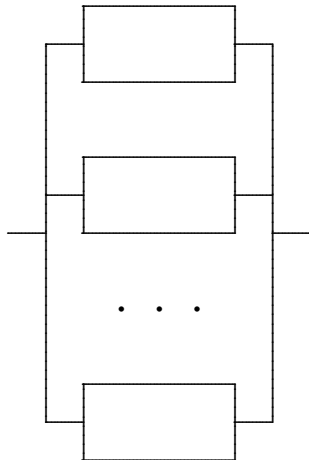
## Parallelsystem $\zeta_P$



# Parallelsystem

$$\begin{aligned}\text{Rel}(\zeta_P) &= P(X_P \geq t) = 1 - P(X_P < t) \\ &= 1 - \underbrace{P(X_1 < t, \dots, X_n < t)}_{\substack{\text{alle Komponenten sind} \\ \text{vor dem Zeitpunkt } t \\ \text{ausgefallen}}} = \\ &= 1 - \prod_{i=1}^n P(X_i < t) = 1 - \prod_{i=1}^n F_i(t) \\ &= 1 - (1 - e^{-\lambda t})^n \quad \text{wenn} \quad \lambda_i = \lambda \quad \forall i\end{aligned}$$

# Parallelsystem



Parallelsystem

$$\text{Rel}(\zeta_P) = 1 - (1 - e^{-\lambda t})^n$$

# Parallelsystem

$$n \rightarrow \infty, \lambda_i = \lambda : \text{Rel}(\zeta_P) \rightarrow 1$$

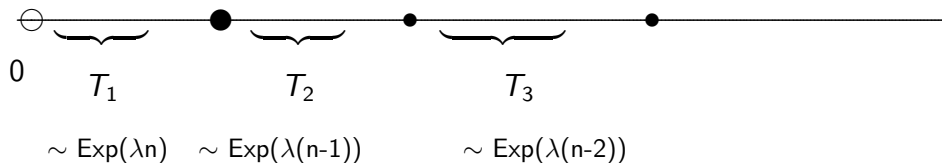
**Bem.:** Die Lebensdauer  $X_P$  des Parallelsystems kann beschrieben werden durch

$$X_P = \max_i X_i.$$



# Zuverlässigkeitsmodelle

Mittlere Lebensdauer des Parallelsystems ( $\lambda_i = \lambda$ )



$T_1$ : Wartezeit bis zum 1. Ausfall einer Komponente

$T_i$ : Wartezeit zwischen  $(i-1)$ -tem und  $i$ -tem Ausfall einer Komponente

$$X_P = \sum_{i=1}^n T_i.$$

# Parallelsystem

mittlere Lebensdauer (2)

Zwischen  $(i - 1)$ -tem und  $i$ -tem Ausfall einer Komponente arbeiten genau  $n - i + 1$  Komponenten gleichzeitig. Die Lebensdauer dieses Teilsystems aus  $n - i + 1$  Komponenten (Reihensystem) hat eine Exponentialverteilung mit Parameter  $\mu_i = (n - i + 1) \cdot \lambda$ ,

$$\mathbf{E}T_i = \frac{1}{\mu_i} = \frac{1}{n - i + 1} \cdot \frac{1}{\lambda}$$

$$\mathbf{E}X_P = \sum_{i=1}^n \frac{1}{\mu_i} = \frac{1}{\lambda} \sum_{i=1}^n \frac{1}{n - i + 1} = \frac{1}{\lambda} \sum_{i=1}^n \frac{1}{i}.$$

# Zuverlässigkeitsmodelle

## $k$ aus $n$ Systeme

Das System fällt aus, wenn  $k$  Komponenten ausgefallen sind.

Lebensdauer:  $T = \sum_{i=1}^k T_i$ .

Mittlere Lebensdauer:

$$\begin{aligned}\mathbf{E} T &= \sum_{i=1}^k \frac{1}{\mu_i} = \frac{1}{\lambda} \sum_{i=1}^k \frac{1}{n-i+1} \\ &= \frac{1}{\lambda} \left( \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-k+1} \right).\end{aligned}$$

$n$  aus  $n$ -System: Parallelsystem

1 aus  $n$ -System: Reihensystem

# Zuverlässigkeitsmodelle

## Proversionswahrscheinlichkeiten

Problem: Reihensystem mit 2 Komponenten und der zufälligen Lebensdauer  $X_1, X_2$ :

$$X_1 \sim \text{Exp}(\lambda_1), \quad X_2 \sim \text{Exp}(\lambda_2).$$

System fällt aus.

Mit welcher Wahrscheinlichkeit liegt das an der ersten Komponente?

# Proversionswahrscheinlichkeiten

$$P(X_1 < X_2) = \int_0^{\infty} P(X_1 < X_2 | X_2 = t) f_2(t) dt$$

# Proversionswahrscheinlichkeiten

$$\begin{aligned}P(X_1 < X_2) &= \int_0^{\infty} P(X_1 < X_2 | X_2 = t) f_2(t) dt \\&= \int_0^{\infty} P(X_1 < t) \cdot \lambda_2 e^{-\lambda_2 t} dt \\&= \int_0^{\infty} (1 - e^{-\lambda_1 t}) \cdot \lambda_2 e^{-\lambda_2 t} dt \\&= 1 - \int_0^{\infty} \lambda_2 e^{-(\lambda_1 + \lambda_2)t} dt\end{aligned}$$

# Proversionswahrscheinlichkeiten

$$\begin{aligned}P(X_1 < X_2) &= \int_0^{\infty} P(X_1 < X_2 | X_2 = t) f_2(t) dt \\&= \int_0^{\infty} P(X_1 < t) \cdot \lambda_2 e^{-\lambda_2 t} dt \\&= \int_0^{\infty} (1 - e^{-\lambda_1 t}) \cdot \lambda_2 e^{-\lambda_2 t} dt \\&= 1 - \int_0^{\infty} \lambda_2 e^{-(\lambda_1 + \lambda_2)t} dt \\&= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}.\end{aligned}$$

# Proversionswahrscheinlichkeiten

bei Exponentialverteilung

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

$$\frac{1}{\lambda_1} = 1000h, \frac{1}{\lambda_2} = 500h :$$

$$P(X_1 < X_2) = \frac{1}{3}.$$



# Faltung der Exponentialverteilung

System mit 2 Komponenten: Zunächst ist nur die erste Komponente eingeschaltet. Wenn diese ausfällt, wird automatisch die 2. Komponente zugeschaltet. Das System fällt aus, wenn beide Komponenten defekt sind.

Die Lebensdauern  $X_1, X_2$  seien unabhängig und exponential,  $X_1, X_2 \sim \text{Exp}(\lambda)$  verteilt.

Frage: Wahrscheinlichkeit für Systemausfall?

# Faltung der Exponentialverteilung

$$F_{X_1+X_2}(t) = P(X_1 + X_2 < t)$$

# Faltung der Exponentialverteilung

$$\begin{aligned} F_{X_1+X_2}(t) &= P(X_1 + X_2 < t) \\ &= \int_0^\infty P(X_1 + X_2 < t | X_2 = s) f(s) ds \end{aligned}$$

# Faltung der Exponentialverteilung

$$\begin{aligned}F_{X_1+X_2}(t) &= P(X_1 + X_2 < t) \\&= \int_0^\infty P(X_1 + X_2 < t | X_2 = s) f(s) ds \\&= \int_0^\infty P(X_1 < t - s) f(s) ds\end{aligned}$$

# Faltung der Exponentialverteilung

$$\begin{aligned}F_{X_1+X_2}(t) &= P(X_1 + X_2 < t) \\&= \int_0^\infty P(X_1 + X_2 < t | X_2 = s) f(s) ds \\&= \int_0^\infty P(X_1 < t - s) f(s) ds \\&= \int_0^\infty F(t - s) f(s) ds\end{aligned}$$

# Faltung der Exponentialverteilung

$$\begin{aligned}F_{X_1+X_2}(t) &= P(X_1 + X_2 < t) \\&= \int_0^\infty P(X_1 + X_2 < t | X_2 = s) f(s) ds \\&= \int_0^\infty P(X_1 < t - s) f(s) ds \\&= \int_0^\infty F(t - s) f(s) ds \\&= \int_0^t (1 - e^{-\lambda(t-s)}) \lambda e^{-\lambda s} ds \\&= \int_0^t \lambda e^{-\lambda s} ds - \int_0^t \lambda e^{-\lambda t} ds \\&= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}.\end{aligned}$$

# Faltung der Exponentialverteilung

Erlang-Verteilung

Dichte ( $t > 0$ ):

$$\begin{aligned} f(t) = F'(t) &= \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t} - \lambda e^{-\lambda t} \\ &= \lambda^2 \cdot t \cdot e^{-\lambda t} \end{aligned}$$

# Faltung der Exponentialverteilung

Erlang-Verteilung

Dichte ( $t > 0$ ):

$$\begin{aligned} f(t) = F'(t) &= \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t} - \lambda e^{-\lambda t} \\ &= \lambda^2 \cdot t \cdot e^{-\lambda t} \end{aligned}$$

Erlang-Verteilung mit Parameter  $(2, \lambda)$ .

**Satz:** Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim \text{Exp}(\lambda)$

Dann ist

$$X_1 + X_2 + \dots + X_n \sim \text{Erlang}(n, \lambda),$$



# Faltung der Exponentialverteilung

Erlang-Verteilung

Dichte ( $t > 0$ ):

$$\begin{aligned} f(t) = F'(t) &= \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t} - \lambda e^{-\lambda t} \\ &= \lambda^2 \cdot t \cdot e^{-\lambda t} \end{aligned}$$

Erlang-Verteilung mit Parameter  $(2, \lambda)$ .

**Satz:** Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim \text{Exp}(\lambda)$

Dann ist

$$X_1 + X_2 + \dots + X_n \sim \text{Erlang}(n, \lambda),$$

Erlang verteilt mit Parametern  $(n, \lambda)$  und Dichte:

$$f_{\text{Erl}}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

# Zuverlässigkeitsmodelle

## Ausfallrate

### Def. 42 Ausfallrate-Funktion (oder Hazardrate-Funktion)

$$\mu(t) = \frac{f(t)}{1 - F(t)}$$

( $F$  eine Verteilungsfunktion mit Dichte  $f$ )

Interpretation: Die Zufallsvariable  $X$  habe bereits die Zeit  $t$  überlebt.

Frage: Wie groß ist die Wahrscheinlichkeit, dass  $X$  den Zeitraum  $[t, t + dt]$  nicht überlebt

## Ausfallrate-Funktion (2)

Frage: Wie groß ist die Wahrscheinlichkeit, dass  $X$  den Zeitraum  $[t, t + dt]$  nicht überlebt, also

$$\begin{aligned} P(X \leq t + dt | X > t) &= \frac{P(X \in [t, t + dt])}{P(X > t)} \\ &= \frac{\int_t^{t+dt} f(x) dx}{1 - F(t)} \\ &= \frac{F(t + dt) - F(t)}{1 - F(t)} \\ &\approx \frac{f(t)dt}{1 - F(t)} = \mu(t)dt. \end{aligned}$$

$\mu(t)$ : Rate mit der ein Bauteil, das  $t$  alt ist, ausfällt.

## Ausfallrate-Funktion (3)

$$F(t) = 1 - e^{-\lambda t}$$

$$\mu(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Bei Exponentialverteilung ist die Ausfallrate konstant, sie hängt nicht vom Zeitpunkt ab!

## Ausfallrate-Funktion (3)

$$F(t) = 1 - e^{-\lambda t}$$
$$\mu(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Bei Exponentialverteilung ist die Ausfallrate konstant, sie hängt nicht vom Zeitpunkt ab!

ÜA: Sei  $F$  eine stetige Verteilungsfunktion mit Dichte  $f$  und konstanter Ausfallrate. Zeigen Sie, dass  $f$  Exponential-Dichte ist.

Hinweis: Setzen Sie  $u(t) := 1 - F(t)$  und lösen Sie die Differentialgleichung  $u' + \lambda u = 0$ .

# Ausfallrate-Funktion (4)

## Def. 43 (IFR, DFR)

- Eine Verteilungsfunktion  $F$  hat Increasing Failure Rate (IFR), falls  $\mu(t)$  monoton wachsend ist.
- $F$  hat Decreasing Failure Rate (DFR), falls  $\mu(t)$  monoton fallend ist.

## Weibull-Verteilung

Verteilungsfunktion:  $F(t) = 1 - e^{-(\lambda t)^b}, \quad t, \lambda, b > 0,$

Dichtefunktion:  $f(t) = b\lambda^b t^{b-1} e^{-(\lambda t)^b}$

# Ausfallrate-Funktion (5)

Weibull-Verteilung

$$\mu(t) = \frac{f(t)}{1 - F(t)} = \frac{b\lambda^b t^{b-1} e^{-(\lambda t)^b}}{e^{-(\lambda t)^b}} = b\lambda^b t^{b-1}$$

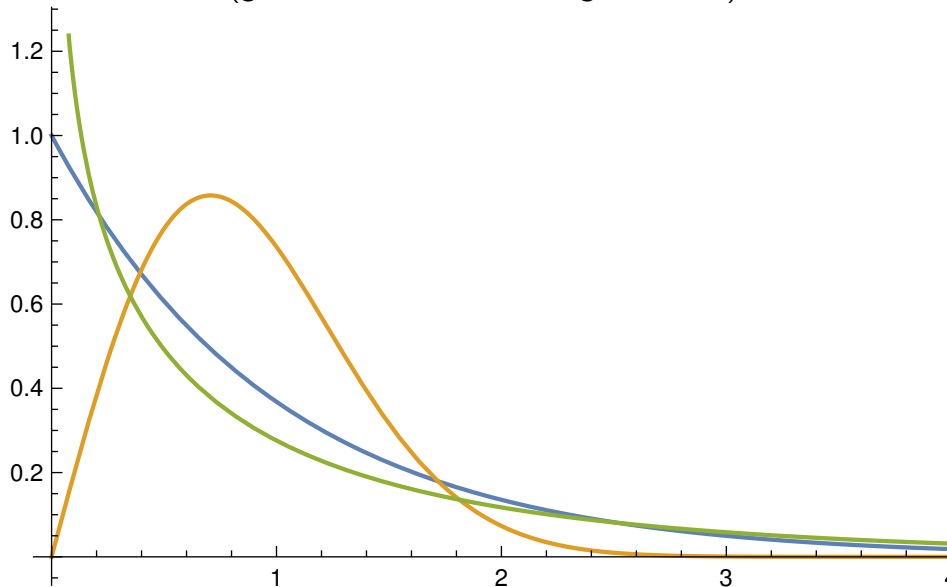
IFR            falls    $b > 1$

IFR, DFR   falls    $b = 1$    (exp)

DFR           falls    $b < 1$

- System mit verdeckten Mängeln, aber langsamen “Altern” → Ausfallrate sinkt → Weibull,  $b < 1$
- System mit wenig verdeckten Mängeln, aber schnellem “Altern” → Ausfallrate steigt → Weibull,  $b > 1$

Weibull Dichten (grün:  $b < 1$ , blau:  $b = 1$ , gelb:  $b > 1$ )





# Ausfallrate-Funktion

## Hjorth-Verteilung

$$\text{Verteilungsfkt.: } F(t) = 1 - \frac{e^{-\lambda t^2/2}}{(1 + bt)^{\gamma/b}}, \quad t, \lambda, \gamma, b > 0,$$

$$\text{Dichtefkt.: } f(t) = \frac{\lambda t(1 + bt) + \gamma}{(1 + bt)^{\gamma/b+1}} e^{-\lambda t^2/2}$$

$$\mu(t) = \frac{f(t)}{1 - F(t)} = \lambda t + \frac{\gamma}{1 + bt}$$

fallend für  $\lambda = 0$

badewannenförmig für  $0 < \lambda < b\gamma$ .

# Ausfallrate-Funktion

Die Hjorth-Verteilung modelliert also badewannenförmige Ausfallraten.

- zunächst fallen viele Objekte aus (Kinderkrankheiten)
- dann Ausfallrate zeitweilig konstant
- schließlich mehren sich die Ausfälle aufgrund von Alterungserscheinungen.

# Kumulierte Hazardfunktion

$$H(t) = \int_0^t \mu(s) ds = -\log G(t)$$

“Ansammlung” von Risiko (Hazard).

# Inhalt

- Einführung
- Gedächtnislosigkeit
- Zuverlässigkeitsmodelle
- Bedienungstheorie
- Zusammenfassung (Exponentialverteilung)

Es werden ganz kurz einige Fragestellungen skizziert.

### M/M/s - Wartesystem

- $X \sim \text{Exp}(\lambda)$  Zeit zwischen Ankünften/Anforderungen
- Forderungen reihen sich in eine Warteschlange ein.
- $B \sim \text{Exp}(\mu)$  Bedienungszeiten, unabhängig
- s parallele Bedienungsplätze
- Bei frei werdendem Bedienungsplatz wird die nächste Forderung sofort bedient.

# M/M/s - Wartesystem

Fragestellungen:

- Mittlere Anzahl der Forderungen im System
- Mittlere Warteschlangenlänge
- Mittlere Wartezeit  $\mathbf{EW}$
- Besetzungswahrscheinlichkeit  $P_B$
- Wartezeitverteilung (ohne Beweis)

$$P(W \leq u) = 1 - P_B e^{-(s\mu - \lambda)u}$$

$$\mathbf{EW} = \frac{P_B}{s\mu - \lambda}.$$

Stationärer Fall, wenn  $\frac{1}{s\mu} < \frac{1}{\lambda}$ .

# Bedienungstheorie

## M/M/s - Verlustsystem

- $X \sim \text{Exp}(\lambda)$  Zeit zwischen Ankünften/Anforderungen
- Eine ankommende Forderung wird sofort bedient, wenn ein Bedienungsplatz frei ist, ansonsten geht sie verloren.
- $B \sim \text{Exp}(\mu)$  Bedienungszeiten, unabhängig
- $s$  parallele Bedienungsplätze

### Fragestellungen:

- Verlustwahrscheinlichkeit
- Mittlere Anzahl der besetzten Bedienungsplätze

# Inhalt

- Einführung
- Gedächtnislosigkeit
- Zuverlässigkeitsmodelle
- Bedienungstheorie
- Zusammenfassung (Exponentialverteilung)



- Exponentialdichte

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{else.} \end{cases}$$

- Erwartungswert

$$\mathbf{E}X = \frac{1}{\lambda}.$$

- Überlebensfunktion

$$G(t) = 1 - F(t) = e^{-\lambda t}.$$

- Cauchy-Funktionalgleichung

$$G(s + t) = G(s) \cdot G(t).$$

## Zusammenfassung (Exponentialverteilung, 2)

- Die Exponential-Verteilung ist gedächtnislos.
- Die einzige gedächtnislose stetige Verteilung ist die Exponential-Verteilung
- Exponential-Verteilung ist eine Extremwertverteilung.
- Anwendungen in der Zuverlässigkeitstheorie  
Reihensystem, Parallelsystem
- Ausfallrate-Funktion

$$\mu(t) = \frac{f(t)}{1 - F(t)}.$$

- Die Ausfallratefunktion der Exponentialverteilung ist konstant.

## Ergänzung (Exponentialverteilung)

**Satz:** Sei  $\mathbf{X}$  Zufallsvariable auf  $[0, \infty)$  mit  $\mathbf{E}(X) = \lambda$ .

Dann ist die Entropie

$$H(p) = - \int_0^{\infty} p(x) \log(p(x)) dx$$

maximal, falls

$$p(x) = \lambda e^{-\lambda x}$$

d.h.  $p(x)$  ist Exponentialdichte.

**Beweis:** der Beweis ist analog zu dem entsprechenden Satz bei Normalverteilung, siehe Abschnitt Normalverteilung.

(Man setzt dort  $\log q(x) = \alpha + \beta x$ .)

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

## Def. 44 (Dichte)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-(x-\mu)^2/2\sigma^2}, \quad \mu \in \mathbb{R}, \sigma > 0$$

Def. 45 (Standard-Normaldichte,  $\mu = 0$ ,  $\sigma^2 = 1$ )

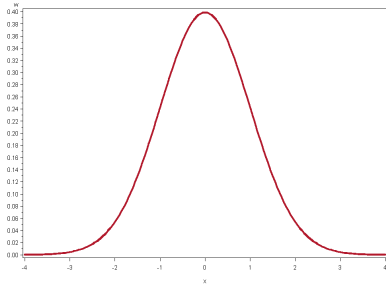
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \quad \text{Dichte}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad \text{Verteilungsfunktion}$$

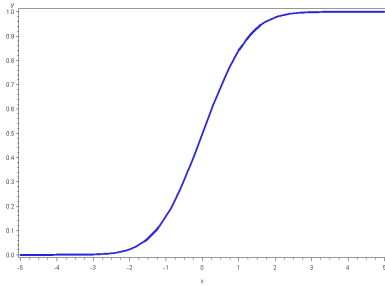
$\varphi(x)$ ,  $\Phi(x)$  sind tabelliert!

$$\varphi(x) = \varphi(-x) \quad \Phi(x) = 1 - \Phi(-x)$$

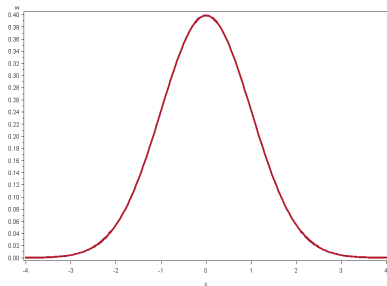
Dichtefunktion der Standard-Normalverteilung



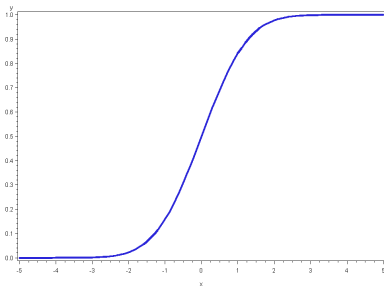
Verteilungsfunktion der Standard-Normalverteilung



Dichtefunktion der Standard-Normalverteilung

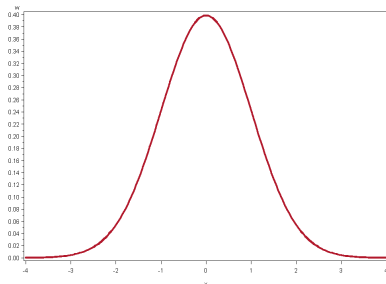


Verteilungsfunktion der Standard-Normalverteilung

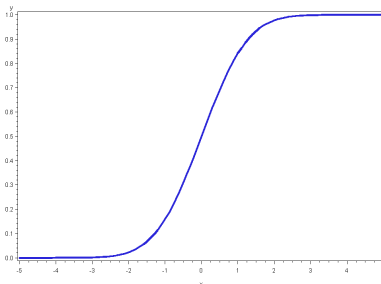


$$X \sim \mathcal{N}(0, 1) : \quad P(a < X < b) = \Phi(b) - \Phi(a).$$

Dichtefunktion der Standard-Normalverteilung



Verteilungsfunktion der Standard-Normalverteilung



$$X \sim \mathcal{N}(0, 1) : \quad P(a < X < b) = \Phi(b) - \Phi(a).$$

Frage: Für welches  $x$  gilt:  $\Phi(x) = \alpha$ ?

$x = \Phi^{-1}(\alpha)$   $\alpha$ -Quantil.

$\Phi^{-1}(\alpha)$  als Funktion: Quantilfunktion



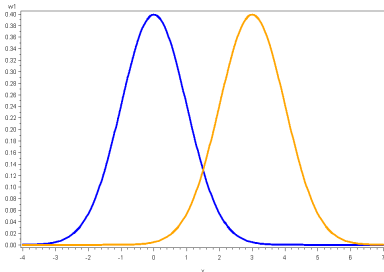
# Die Normalverteilung

## Vergleichen Sie

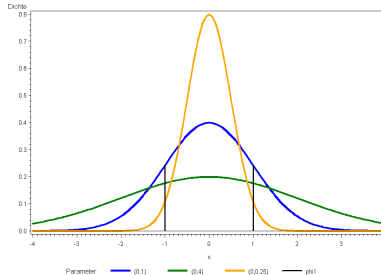
a)  $\sigma^2$  fest,  $\mu$  verschieden

b)  $\mu$  fest,  $\sigma^2$  verschieden

Dichtefunktion verschiedener Normalverteilungen  
Lageunterschied



Dichtefunktion verschiedener Normalverteilungen  
Skalenunterschied



# Die Normalverteilung

Satz: Es gilt:

$$X \sim \mathcal{N}(0, 1) \iff \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \iff \alpha X + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$$

$$X \sim \mathcal{N}(\mu, \sigma^2) \iff \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

**Beweis:** : Wir zeigen nur 1. ( $\rightarrow$ ). Sei  $X \sim \mathcal{N}(0, 1)$ .

$$\begin{aligned} P(\sigma X + \mu \leq x) &= P(X \leq \frac{x - \mu}{\sigma}) = \Phi\left(\frac{x - \mu}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{x - \mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u - \mu)^2/(2\sigma^2)} du \end{aligned}$$

$$\frac{u - \mu}{\sigma} = t, \quad \frac{1}{\sigma} du = dt.$$



# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

Satz: Sei  $X_1 \sim \mathcal{N}(\mu, \sigma_1^2)$ ,  $X_2 \sim \mathcal{N}(\mu, \sigma_2^2)$ ,

$\sigma_1^2 < \sigma_2^2$  und  $a > 0$ . Dann gilt:

$$P(\mu - a < X_1 < \mu + a) > P(\mu - a < X_2 < \mu + a).$$

**Beweis:**

$$\begin{aligned} P(\mu - a < X_1 < \mu + a) &= P\left(\frac{-a}{\sigma_1} < \frac{X_1 - \mu}{\sigma_1} < \frac{a}{\sigma_1}\right) \\ &= \Phi\left(\frac{a}{\sigma_1}\right) - \Phi\left(-\frac{a}{\sigma_1}\right) \\ &> \Phi\left(\frac{a}{\sigma_2}\right) - \Phi\left(-\frac{a}{\sigma_2}\right) \\ &= P(\mu - a < X_2 < \mu + a). \end{aligned}$$

□

# Beispiel

$$X_1 \sim \mathcal{N}(10, 4), X_2 \sim \mathcal{N}(10, 9), a = 1.$$

$$P(9 < X_1 < 11) =$$

$$= \Phi\left(\frac{11-10}{2}\right) - \Phi\left(\frac{9-10}{2}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{2}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - (1 - \Phi\left(\frac{1}{2}\right))$$

$$= 2 \cdot \Phi\left(\frac{1}{2}\right) - 1$$

$$= 2 \cdot 0.6915 - 1 = 0.383.$$

# Beispiel

$$X_1 \sim \mathcal{N}(10, 4), X_2 \sim \mathcal{N}(10, 9), a = 1.$$

$$P(9 < X_1 < 11) =$$

$$= \Phi\left(\frac{11-10}{2}\right) - \Phi\left(\frac{9-10}{2}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{2}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - (1 - \Phi\left(\frac{1}{2}\right))$$

$$= 2 \cdot \Phi\left(\frac{1}{2}\right) - 1$$

$$= 2 \cdot 0.6915 - 1 = 0.383.$$

$$P(9 < X_2 < 11) =$$

$$= \Phi\left(\frac{11-10}{3}\right) - \Phi\left(\frac{9-10}{3}\right)$$

$$= \Phi\left(\frac{1}{3}\right) - \Phi\left(-\frac{1}{3}\right)$$

$$= \Phi\left(\frac{1}{3}\right) - (1 - \Phi\left(\frac{1}{3}\right))$$

$$= 2 \cdot \Phi\left(\frac{1}{3}\right) - 1$$

$$= 2 \cdot 0.63056 - 1 = 0.26112.$$

Für die Berechnung der Wahrscheinlichkeiten  $\Phi(x)$  existieren Programme und Tabellen.

- $x \geq 0$ . In diesem Fall kann der Wert für  $P(X < x)$  direkt aus der Tabelle abgelesen werden.

Für die Berechnung der Wahrscheinlichkeiten  $\Phi(x)$  existieren Programme und Tabellen.

- $x \geq 0$ . In diesem Fall kann der Wert für  $P(X < x)$  direkt aus der Tabelle abgelesen werden.
- $x < 0$ .  $P(X < x) = \Phi(x) = 1 - \Phi(-x)$ , z.B.

$$P(X < -1) = \Phi(-1) = 1 - \Phi(1) \approx 0.15.$$



Für die Berechnung der Wahrscheinlichkeiten  $\Phi(x)$  existieren Programme und Tabellen.

- $x \geq 0$ . In diesem Fall kann der Wert für  $P(X < x)$  direkt aus der Tabelle abgelesen werden.
- $x < 0$ .  $P(X < x) = \Phi(x) = 1 - \Phi(-x)$ , z.B.

$$P(X < -1) = \Phi(-1) = 1 - \Phi(1) \approx 0.15.$$

- $P(a < X < b) = \Phi(b) - \Phi(a)$ , z.B.

$$\begin{aligned} P(-1 \leq x \leq 1) &= \Phi(1) - \Phi(-1) = \\ &= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 \approx 0.68. \end{aligned}$$

# Berechnen von Wahrscheinlichkeiten

## Beispiele

- $Y \sim \mathcal{N}(0, 1)$ :  $P(Y < 0) = \frac{1}{2}$  (laut Tabelle oder wegen Symmetrie);
- $X \sim \mathcal{N}(1, 2^2)$ :  $P(X < 0) = \Phi\left(\frac{0-1}{2}\right) = \Phi\left(-\frac{1}{2}\right) = 1 - \Phi\left(\frac{1}{2}\right) \approx 1 - 0.691 = 0.309$ .

# Berechnen von Wahrscheinlichkeiten

## Def. 46 (p-Quantil)

Sei die Verteilungsfunktion  $F$  und die Wahrscheinlichkeiten  $p$  gegeben. Ein Wert  $x_p$  mit

$$p = P(X < x_p) = F(x_p)$$

heißt  $p$ -Quantil der Zufallsvariablen  $X$ , der Verteilungsfunktion (oder nur der Verteilung)  $F$ .

Sei  $Y \sim \mathcal{N}(0, 1)$ .

Gesucht ist das  $p = 0.95$ -Quantil von  $Y$ .

Tabelle für  $p = 0.95$ :  $x_p(0, 1) \approx 1.645$

# Berechnen von Wahrscheinlichkeiten

Sei  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Bestimmen das  $p$ -Quantil  $x_p(\mu, \sigma^2)$ :

$$\begin{aligned} p &= P(X < x_p(\mu, \sigma)) = P\left(\frac{X - \mu}{\sigma} < \frac{x_p(\mu, \sigma) - \mu}{\sigma}\right) \\ &= P(Y < x_p(0, 1)), \quad Y \sim \mathcal{N}(0, 1). \end{aligned}$$

D.h.

$$x_p(0, 1) = \frac{x_p(\mu, \sigma) - \mu}{\sigma},$$

woraus durch Umstellen folgt:

$$x_p(\mu, \sigma) = \sigma \cdot x_p(0, 1) + \mu.$$

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

**Def. 47** ( $k \cdot \sigma$ -Intervalle)

Für eine normalverteilte Zufallsvariable  $X \sim \mathcal{N}(\mu, \sigma^2)$  ist  $[\mu - k\sigma, \mu + k\sigma]$  ein  $k \cdot \sigma$ -Intervall,  $k \in \mathbb{Z}^+$ . Interessant sind dabei die Wahrscheinlichkeiten:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma).$$

$$\begin{aligned} P(X \in [\mu - k\sigma, \mu + k\sigma]) &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi(k) - \Phi(-k) \\ &= \Phi(k) - (1 - \Phi(k)) \\ &= 2 \cdot \Phi(k) - 1 \end{aligned}$$

# $k \cdot \sigma$ -Intervalle

$k \cdot \sigma$ -Intervalle für  $k = 1, \dots, 5$

| $k$ | $2 \cdot \Phi(k) - 1$ |
|-----|-----------------------|
| 1   | 0.6827                |
| 2   | 0.9545                |
| 3   | 0.9973                |
| 4   | 0.99997               |
| 5   | 0.9999994             |
| 6   | 0.999999998           |

## $k \cdot \sigma$ -Intervalle

Ein Zeitungsverkäufer sieht die Nachfrage  $X$  nach einer Tageszeitung als angenähert normalverteilt an. Das  $2 \cdot \sigma$ -Intervall sei  $[322, 408]$ . Wie groß ist die Wahrscheinlichkeit, dass mindestens 400 Exemplare der Zeitung verkauft werden?



## $k \cdot \sigma$ -Intervalle

Ein Zeitungsverkäufer sieht die Nachfrage  $X$  nach einer Tageszeitung als angenähert normalverteilt an. Das  $2 \cdot \sigma$ -Intervall sei  $[322, 408]$ . Wie groß ist die Wahrscheinlichkeit, dass mindestens 400 Exemplare der Zeitung verkauft werden?

Die Frage ist also:  $P(X \geq 400) = ?$

## $k \cdot \sigma$ -Intervalle

Ein Zeitungsverkäufer sieht die Nachfrage  $X$  nach einer Tageszeitung als angenähert normalverteilt an. Das  $2 \cdot \sigma$ -Intervall sei  $[322, 408]$ . Wie groß ist die Wahrscheinlichkeit, dass mindestens 400 Exemplare der Zeitung verkauft werden?

Die Frage ist also:  $P(X \geq 400) = ?$

Nach Voraussetzung gilt:

$$322 = \mu - 2\sigma, \quad 408 = \mu + 2\sigma.$$

Lösung des linearen Gleichungssystems liefert

$$730 = 2\mu \quad \Rightarrow \quad \mu = 365, \quad 86 = 4\sigma \quad \Rightarrow \quad \sigma = 21,5.$$

## $k \cdot \sigma$ -Intervalle

$$P(X \geq 400) = 1 - P(X < 400) =$$

## $k \cdot \sigma$ -Intervalle

$$\begin{aligned}P(X \geq 400) &= 1 - P(X < 400) = 1 - \Phi\left(\frac{400 - \mu}{\sigma}\right) \\&= 1 - \Phi\left(\frac{400 - 365}{21.5}\right) \approx 1 - \Phi(1.63) \\&\approx 1 - 0.95 = 0.05\end{aligned}$$

Hat man ein  $k \cdot \sigma$ -Intervall gegeben (und es wird Normalverteilung angenommen), so ist es möglich, jede andere Wahrscheinlichkeit auszurechnen.

Anwendung z.B. bei der Untersuchung von Toleranzen bei Werkstückmaßen oder bei Gewichtseinlagen von Gerichten.

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

## Zentraler Grenzwertsatz

Seien  $X_i$  unabhängig, identisch verteilt,

$$\mathbf{E}X_i = \mu, \text{Var } X_i = \sigma^2.$$

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

$$Z_n := \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{n \rightarrow \infty} Z, \quad Z \sim \mathcal{N}(0, 1).$$

Beweis: siehe Grenzwertsätze.

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

## Satz

Fehler sind unter folgenden Annahmen (asymptotisch) normalverteilt:

V1: Jeder Fehler ist Summe einer sehr großen Anzahl sehr kleiner, gleich großer Fehler, die verschiedene Ursachen haben.

V2: Die verschiedenen Fehlerkomponenten sind unabhängig.

V3: Jede Fehlerkomponente ist mit Wkt. 0.5 positiv und mit Wkt. 0.5 negativ.



# Fehlertheorie

## Beweis des Satzes

Seien  $\epsilon_j, j = 1, \dots, n$  die Fehlerkomponenten.

V3  $\Rightarrow P(\epsilon_j = \pm\epsilon) = \frac{1}{2}$ , d.h.  $\mathbf{E}\epsilon_j = 0$ ,  $\text{var}\epsilon_j = \epsilon^2$

V1  $\Rightarrow$  Gesamtfehler  $X = \sum_j \epsilon_j$ , also

$$\mathbf{E}(X) = \sum_{j=1}^n \mathbf{E}(\epsilon_j) = 0$$

$$\text{var}(X) = \sum_{j=1}^n \text{var}(\epsilon_j) = n\epsilon^2 =: \sigma^2$$

# Fehlertheorie

Beweis des Satzes (2)

Charakteristische Funktion von  $\epsilon_j$ :

$$\phi_{\epsilon_j}(t) = \mathbf{E}(e^{it\epsilon_j}) = \frac{1}{2}(e^{it\epsilon} + e^{-it\epsilon}) = \sum_{k=0}^{\infty} \frac{(it\epsilon)^{2k}}{(2k)!}$$

Charakteristische Funktion von  $X$ :

$$\begin{aligned}\phi_X(t) &= \prod_{j=1}^n \phi_{\epsilon_j}(t) = \left(1 - \frac{t^2}{2!}\epsilon^2 + \frac{t^4}{4!}\epsilon^4 - + \dots\right)^n \\ &= \left(1 - \frac{t^2}{2!} \frac{\sigma^2}{n} + o\left(\frac{1}{n}\right)\right)^n \\ &\xrightarrow{n \rightarrow \infty} e^{-t^2 \sigma^2 / 2}\end{aligned}$$

Das ist die charakteristische Funktion von  $\mathcal{N}(0, \sigma^2)$ .

# Fehlertheorie

Beweis des Satzes (3)

Den Grenzwert kann man sich wie folgt überlegen:

Sei  $a_n = 1 - \frac{t^2 \sigma^2 / 2!}{n}$  und  $b_n = o(\frac{1}{n})$ . Dann ist

$$\begin{aligned}
 \log \phi_X(t) &= \log(a_n + b_n)^n = n \log(a_n + b_n) \\
 &= n(\log(a_n + b_n) - \log a_n + \log a_n) \\
 &= \underbrace{nb_n}_{o(1)} \underbrace{\frac{\log(a_n + b_n) - \log a_n}{b_n}}_{\approx 1/a_n} + n \log a_n \\
 &= o(1) + n \log a_n \\
 \phi_X(t) &= \underbrace{e^{o(1)}}_{\rightarrow 1} \cdot a_n^n \rightarrow e^{-t^2 \sigma^2 / 2}
 \end{aligned}$$

Die Behauptung folgt dann aus dem Konvergenzsatz.

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

## Def. 48 (Entropie)

$$H(f) := - \int f(x) \log f(x) dx$$

Maximale Entropie bei gegebenen Erwartungswert  $\mu$  und Varianz  $\sigma^2$ .

$f$ : Wahrscheinlichkeitsdichte auf  $(-\infty, \infty)$ .

$$(*) \quad \int x f(x) dx = \mu, \quad \int (x - \mu)^2 f(x) dx = \sigma^2$$

Die Entropie ist zu maximieren unter den obigen Bedingungen (\*).

# Maximale Entropie (2)

## Satz:

Eine Dichtefunktion, die die Entropie unter den obigen Bedingungen maximiert ist normal.

Zum Beweis verwenden wir die Jensensche Ungleichung:

## Jensensche Ungleichung für konkave Funktionen

Es sei  $g$  eine differenzierbare und konkave Funktion, und sei  $X$  eine zufällige Variable. Dann gilt:

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

# Maximale Entropie

Beweis der Jensenschen Ungleichung

**Beweis:** Sei  $T(x)$  die Tangente an die Kurve der Funktion  $g$  im Punkt  $x_0$ ,

$$g(x) \leq T(x) = g(x_0) + \underbrace{g'(x_0)}_{\text{Anstieg der Kurve in } x_0} \cdot (x - x_0).$$

Wir setzen nun  $x := X$  und  $x_0 := \mathbf{E}X$  und erhalten:

$$g(X) \leq g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X).$$

Daraus folgt:

$$\begin{aligned} \mathbf{E}g(X) &\leq \mathbf{E}(g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X)) \\ &= g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot \underbrace{\mathbf{E}(X - \mathbf{E}X)}_{=0} = g(\mathbf{E}X) \end{aligned}$$

# Maximale Entropie

## Beweis des Satzes

Seien  $p$  und  $q$  beliebige Dichten. Da die Logarithmus-Funktion konkav ist folgt aus der Jensenschen Ungleichung:

$$\begin{aligned}\int \ln\left(\frac{q}{p}(x)\right)p(x) dx &= \mathbf{E}_p \ln\left(\frac{q}{p}(X)\right) \\ &\leq \ln \mathbf{E}_p\left(\frac{q}{p}(X)\right) \\ &= \ln \int \left(\frac{q}{p}(x)\right)p(x) dx \\ &= \ln\left(\int q(x) dx\right) = \ln 1 = 0.\end{aligned}$$

Daraus folgt:



# Maximale Entropie

Beweis des Satzes (2)

$$H(p) = - \int p \ln p \, dx \leq - \int p \ln q \, dx$$

Sei  $q$  wie folgt definiert:

$$\ln q = \alpha + \beta(x - \mu) + \gamma(x - \mu)^2,$$

wobei  $\alpha, \beta, \gamma$  so dass  $q$  Dichte,  $q \sim (\mu, \sigma^2)$ .

$$\begin{aligned} H(p) &\leq - \int p \ln q \, dx \\ &= - \int p(x) (\alpha + \beta(x - \mu) + \gamma(x - \mu)^2) \, dx \\ &= -(\alpha + \gamma\sigma^2) \end{aligned}$$

feste obere Schranke für die Entropie.

# Maximale Entropie

Beweis des Satzes (3)

Diese Schranke wird angenommen für  $p = q$ , also

$$\ln p = \alpha + \beta(x - \mu) + \gamma(x - \mu)^2$$

$$p = e^{\alpha + \beta(x - \mu) + \gamma(x - \mu)^2}$$

# Maximale Entropie

Beweis des Satzes (3)

Diese Schranke wird angenommen für  $p = q$ , also

$$\ln p = \alpha + \beta(x - \mu) + \gamma(x - \mu)^2$$

$$p = e^{\alpha + \beta(x - \mu) + \gamma(x - \mu)^2}$$

Offen: Gibt es  $\alpha, \beta, \gamma$  mit  $p$  Dichte und  $p \sim (\mu, \sigma^2)$ ?

Antwort: ja,  $\alpha = -\ln(\sqrt{2\pi}\sigma)$ ,  $\beta = 0$ ,  $\gamma = -\frac{1}{2\sigma^2}$ .

Die Lösung ist auch (i.W.) eindeutig, da in der Jensenschen

Ungleichung das Gleichheitszeichen nur gilt, wenn fast überall  $\frac{p}{q} = 1$  gilt.

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

Satz: Seien  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$       $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

unabhängig. Dann:

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

**Beweis:** : (allgemeiner für  $n$  Zufallsvariablen)

Seien  $X_j$  u.a. Zufallsvariablen mit  $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ .

Charakteristische Funktion von  $X = \sum_{j=1}^n X_j$ :

$$\phi_X(t) = \prod_{j=1}^n e^{it\mu_j - \sigma_j^2 t^2/2} = e^{it\mu - \sigma^2 t^2/2}$$

wobei  $\mu = \sum \mu_j, \sigma^2 = \sum \sigma_j^2 \Rightarrow X \sim \mathcal{N}(\mu, \sigma^2)$

□

# Inhalt

- Standard-Normalverteilung
- Berechnen von Wahrscheinlichkeiten
- $k \cdot \sigma$ -Intervalle
- Zentraler Grenzwertsatz
- Fehlertheorie
- Maximale Entropie
- Summe normalverteilter Zufallsvariablen
- Treffen einer Zielscheibe

Satz: Sei  $(X, Y)$  zweidimensionale Zufallsvariable.

Folgende Annahmen seien erfüllt:

- V1: Die Randverteilungen von  $X$  und  $Y$  seien stetig.
- V2: Die Dichte  $h(x, y)$  von  $(X, Y)$  hängt nur vom Abstand  $\sqrt{x^2 + y^2}$  vom Nullpunkt ab (Radialsymmetrie).
- V3: Die Fehler in  $x$ - und  $y$ -Richtung sind unabhängig.

# Treffen einer Zielscheibe

Sei  $Z$  die zufällige Abweichung in beliebiger Richtung. Dann gilt

$$Z \sim \mathcal{N}(0, \sigma^2).$$

**Beweis:** siehe Abschnitt Transformationsformel



.



Sei  $X: \Omega \longrightarrow \mathbb{R}$  eine Zufallsvariable mit Verteilungsfunktion  $F_X(x) = P(X < x)$ .

Wir betrachten eine Funktion  $g: \mathbb{R} \longrightarrow \mathbb{R}$  und eine Zufallsvariable  $Y: \Omega \longrightarrow \mathbb{R}$  mit  $Y = g(X)$ .

$$Y: \quad \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}.$$

$$Y(\omega) = g(X(\omega)), \forall \omega \in \Omega.$$

# Transformation von Zufallsvariablen

Die zufällige Variable  $Y = g(X)$  besitzt die Verteilungsfunktion

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(\{\omega: Y(\omega) < y\}) \\ &= P(\{\omega: g(X(\omega)) < y\}) \\ &= P(X \in \underbrace{\{x: g(x) < y\}}_{\in \mathcal{B}^1}) = P(g(X) < y) \end{aligned}$$

**Bem.:**  $\{x: g(x) < y\} \in \mathcal{B}^1$  gilt, wenn die Funktion  $g$  messbar ist.

# Transformation von Zufallsvariablen

Frage: Wie berechnen wir  $F_Y(y)$ ?

Fall 1:  $F$  diskret.

$$\begin{aligned}P(Y = y) &= P(g(X) = y) \\&= P(x : g(x) = y) \\&= P(x : x = g^{-1}(y)) \\&= P(X \in g^{-1}(y)) \\&= P(g^{-1}(y))\end{aligned}$$

# Transformation von Zufallsvariablen

$F$  diskret, Beispiel

Sei  $Y = X^2$ , wobei

$$X = \begin{cases} 1 & \text{mit Wkt. } \frac{1}{4} \\ 0 & \text{mit Wkt. } \frac{1}{2} \\ -1 & \text{mit Wkt. } \frac{1}{4} \end{cases}$$

also  $g(x) = x^2$ ,  $g^{-1}(y) = \pm\sqrt{y} = \{-\sqrt{y}, \sqrt{y}\}$ .

$$P(Y = 0) = P(X = 0) = \frac{1}{2}$$

# Transformation von Zufallsvariablen

$F$  diskret, Beispiel

Sei  $Y = X^2$ , wobei

$$X = \begin{cases} 1 & \text{mit Wkt. } \frac{1}{4} \\ 0 & \text{mit Wkt. } \frac{1}{2} \\ -1 & \text{mit Wkt. } \frac{1}{4} \end{cases}$$

also  $g(x) = x^2$ ,  $g^{-1}(y) = \pm\sqrt{y} = \{-\sqrt{y}, \sqrt{y}\}$ .

$$P(Y = 0) = P(X = 0) = \frac{1}{2}$$

$$\begin{aligned} P(Y = 1) &= P(X \in \sqrt{1}) = P(X = 1 \vee X = -1) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \end{aligned}$$

# Transformation von Zufallsvariablen

Fall 2:  $F$  stetig.

1. Finde für jedes  $y$ :

$$A_y = \{x : g(x) < y\}.$$

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(g(X) < y) \\ &= P(x : g(x) < y) = P(A_y) = \int_{A_y} f_X(x) dx \end{aligned}$$

# Transformation von Zufallsvariablen

Fall 2:  $F$  stetig.

1. Finde für jedes  $y$ :

$$A_y = \{x : g(x) < y\}.$$

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(g(X) < y) \\ &= P(x : g(x) < y) = P(A_y) = \int_{A_y} f_X(x) dx \end{aligned}$$

3. Dichte von  $Y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

# Transformation von Zufallsvariablen

## Beispiel 1

$X \sim R(0, \frac{\pi}{2})$ ., d.h.  $X$  hat die Dichte

$$f(x) = \begin{cases} \frac{2}{\pi} & , \text{ falls } 0 \leq x < \frac{\pi}{2} \\ 0 & , \text{ sonst} \end{cases}.$$

Welche Verteilung hat die Zufallsvariable  $Y = \sin(X)$  ?

1. Finde für jedes  $y$ ,  $y \in (0, 1)$

$$\begin{aligned} A_y &= \{x : g(x) < y\} = \{x : \sin(x) < y\} \\ &= \{x : x < \arcsin(y)\} \end{aligned}$$

Offenbar  $A_y = \emptyset$  für  $y \leq 0$  und  $A_y = \mathbb{R}$  für  $y \geq 1$



# Transformation von Zufallsvariablen

## Beispiel 1 (Fortsetzung)

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = \int_{A_Y} f_X(x) dx \\ &= \frac{2}{\pi} \int_0^{\arcsin(y)} dx = \frac{2}{\pi} \arcsin(y) \end{aligned}$$

3. Dichte von  $Y$ :

# Transformation von Zufallsvariablen

## Beispiel 1 (Fortsetzung)

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = \int_{A_y} f_X(x) dx \\ &= \frac{2}{\pi} \int_0^{\arcsin(y)} dx = \frac{2}{\pi} \arcsin(y) \end{aligned}$$

3. Dichte von  $Y$ :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{2}{\pi} \frac{1}{\sqrt{1-y^2}} & y \in (0, 1) \\ 0 & \text{sonst.} \end{cases}$$

# Transformation von Zufallsvariablen

## Beispiel 2

Sei  $X$  stetig und  $X \sim F_X$  mit Dichte  $f_X$ .

Welche Verteilung hat die Zufallsvariable  $Y = F_X(X)$  ?

$$1. \quad A_y = \{x : F_X(x) < y\} = \{x : x < F_X^{-1}(y)\}$$

Offenbar  $A_y = \emptyset$  für  $y \leq 0$  und  $A_y = \mathbb{R}$  für  $y \geq 1$

2.

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(F_X(X) < y) = P(X < F_X^{-1}(y)) \\ &= \int_{A_y} f_X(x) dx = \int_{-\infty}^{F_X^{-1}(y)} f_X(x) dx \end{aligned}$$

# Transformation von Zufallsvariablen

## Beispiel 2

Sei  $X$  stetig und  $X \sim F_X$  mit Dichte  $f_X$ .

Welche Verteilung hat die Zufallsvariable  $Y = F_X(X)$  ?

$$1. \quad A_y = \{x : F_X(x) < y\} = \{x : x < F_X^{-1}(y)\}$$

Offenbar  $A_y = \emptyset$  für  $y \leq 0$  und  $A_y = \mathbb{R}$  für  $y \geq 1$

2.

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(F_X(X) < y) = P(X < F_X^{-1}(y)) \\ &= \int_{A_y} f_X(x) dx = \int_{-\infty}^{F_X^{-1}(y)} f_X(x) dx \\ &= F_X(F_X^{-1}(y)) - F_X(-\infty) = y \end{aligned}$$

# Transformation von Zufallsvariablen

## Beispiel 2 (Fortsetzung)

### 3. Dichte von $Y$ :

$$f_Y(y) = \begin{cases} 1 & y \in (0, 1) \\ 0 & \text{sonst.} \end{cases}$$

D.h.  $Y \sim R(0, 1)$

# Transformation von Zufallsvariablen

## Beispiel 3

Sei umgekehrt  $U \sim R(0, 1)$  und  $F$  eine Verteilungsfunktion mit Dichte  $f$ .

Welche Verteilung hat die Zufallsvariable  $Y = F^{-1}(U)$  ?

1. Finde für jedes  $y$

$$\begin{aligned} A_y &= \{u : F^{-1}(u) < y\} \\ &= \{u : u < F(y)\} = (0, F(y)). \end{aligned}$$

# Transformation von Zufallsvariablen

## Beispiel 3 (Fortsetzung)

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(F^{-1}(U) < y) \\ &= P(U < F(y)) \\ &= \int_{A_y} f_U(u) \, du = \int_0^{F(y)} f_U(u) \, du \end{aligned}$$

# Transformation von Zufallsvariablen

## Beispiel 3 (Fortsetzung)

2. Die Verteilungsfunktion von  $Y$  ist

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(F^{-1}(U) < y) \\ &= P(U < F(y)) \\ &= \int_{A_y} f_U(u) \, du = \int_0^{F(y)} f_U(u) \, du \\ &= \int_0^{F(y)} du = F(y). \end{aligned}$$

Also  $Y \sim F$ .



# Transformation von Zufallsvariablen

Unter gewissen Zusatzannahmen gilt

## Transformationssatz:

Sei  $X$  eine, auf  $(a, b)$  definierte ( $a = -\infty, b = +\infty$  ist erlaubt) Zufallsgröße mit Dichtefunktion  $f$ . Die Funktion  $g : (a, b) \rightarrow \mathbb{R}$  sei differenzierbar mit  $g'(x) \neq 0$  für alle  $x \in (a, b)$ . Dann hat die zufällige Variable  $Y = g(X)$  auf dem Definitionsbereich von  $g^{-1}$  die Dichtefunktion

$$h(y) = f(g^{-1}(y)) \cdot \left| \frac{dg^{-1}}{dy}(y) \right| = \frac{f(g^{-1}(y))}{|g'(g^{-1}(y))|}.$$

# Transformationssatz

Beweis, Fall 1:  $g'(x) > 0$

**Bem.:** Die Voraussetzung  $g'(x) \neq 0$  für alle  $x \in (a, b)$  bewirkt, dass die Funktion  $g$  auf dem Intervall  $(a, b)$  streng monoton ist.

**Fall 1:** Es sei  $g'(x) > 0 \forall x \in (a, b)$  und  $y \in Db(g^{-1})$ . Da  $g$  streng monoton wachsend ist, ist die Menge  $A_y = (a, g^{-1}(y))$  ein Intervall und die Dichte von  $Y$  ist gegeben durch

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}(F_X(g^{-1}(y)) - F_X(a)).$$

Anwendung der Kettenregel liefert die Behauptung.

# Transformationssatz

Beweis, Fall 2:  $g'(x) < 0$

**Fall 2:** Es gilt  $g'(x) < 0$ , für alle  $x \in (a, b)$ . Da also die Funktion  $g$  streng monoton fallend ist, ist die Menge  $A_y = (g^{-1}(y), b)$  ein Intervall und die Dichte von  $Y$  ist gegeben durch

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} (F_X(b) - F_X(g^{-1}(y))).$$

Anwendung der Kettenregel liefert die Behauptung.

**Bem.:** Beachten Sie, dass in der Formel des Satzes Betragsstriche stehen.

# Transformationsformel

## Beispiel 1

Die folgenden drei Beispiele wurden bereits oben behandelt. Sie folgen jetzt nochmal, diesmal direkte Anwendung des Satzes.

Es sei  $X \sim R(0, \frac{\pi}{2})$ ., d.h.  $X$  hat die Dichte

$$f(x) = \begin{cases} \frac{2}{\pi} & , \text{ falls } 0 \leq x < \frac{\pi}{2} \\ 0 & , \text{ sonst} \end{cases} .$$

$$y = g(x) = \sin x.$$

Für alle  $x \in [0, \frac{\pi}{2}[$  gilt:  $0 \leq g(x) < 1$ ,  $g^{-1}(y) = \arcsin y$ .

# Transformationsformel

## Beispiel 1 (Fortsetzung)

Die Dichte von  $Y = \sin X$  ist nach Transformationsformel

$$\begin{aligned} h(y) &= f(\arcsin y) \cdot \left| \frac{d \arcsin}{dy}(y) \right| \\ &= f(\arcsin y) \cdot \frac{1}{\sqrt{1-y^2}} \\ &= \begin{cases} \frac{2}{\pi} \frac{1}{\sqrt{1-y^2}} & , \text{ falls } 0 \leq y < 1 \\ 0 & , \text{ sonst.} \end{cases} \end{aligned}$$

# Transformationsformel

## Beispiel 2

Es sei  $X$  Zufallsvariable mit Verteilungsfunktion  $F(x) = P(X < x) \in [0, 1[$  und Dichte  $f$ .

Die Dichte der Zufallsvariablen  $Y = F(X)$  ist mittels Transformationsformel ( $y \in (0, 1)$ )

$$\begin{aligned} h(y) &= f(F^{-1}(y)) \cdot \frac{dF^{-1}}{dy}(y) \\ &= f(F^{-1}(y)) \cdot \frac{1}{F'(F^{-1}(y))} \\ &= \frac{f(F^{-1}(y))}{f(F^{-1}(y))} = 1 \end{aligned}$$

Folglich gilt:  $Y \sim R(0, 1)$

# Transformationsformel

## Beispiel 2 (Fortsetzung)

**Bem.:** Wir haben also gezeigt: Wenn  $X \sim F$  so ist die transformierte Zufallsvariable

$$Y = F(X) \sim R(0, 1)$$

Umgekehrt gilt: Ist  $U \sim R(0, 1)$  und ist  $F$  eine beliebige Verteilungsfunktion, so ist  $Y = F^{-1}(U) \sim F$ .

Anwendung: Zufallszahlen (siehe später).

# Transformationsformel

## Beispiel 4

Es sei  $X : \Omega \rightarrow \mathbb{R}$  mit  $X \sim \text{Exp}(\lambda)$ , d.h.

$$F(x) = 1 - e^{-\lambda \cdot x}, \quad x \geq 0.$$

Wegen  $U := F(X) \sim R(0, 1)$  erhalten wir eine exponentialverteilte Zufallsvariable wie folgt:



# Transformationsformel

## Beispiel 4

Es sei  $X : \Omega \rightarrow \mathbb{R}$  mit  $X \sim \text{Exp}(\lambda)$ , d.h.

$$F(x) = 1 - e^{-\lambda \cdot x}, \quad x \geq 0.$$

Wegen  $U := F(X) \sim R(0, 1)$  erhalten wir eine exponentialverteilte Zufallsvariable wie folgt:

$$\begin{aligned} u &= F(x) = 1 - e^{-\lambda \cdot x} \\ e^{-\lambda \cdot x} &= 1 - u \\ x &= -\frac{1}{\lambda} \ln(1 - u) \end{aligned}$$

Die Zufallsgröße  $X = -\frac{1}{\lambda} \ln(1 - U) \sim \text{Exp}(\lambda)$ , d.h.  $X$  ist exponentialverteilt mit dem Parameter  $\lambda$ .

# Transformationsformel

## Beispiel 5

Es sei  $X$  eine Zufallsgröße mit der Dichtefunktion  $f$ .

Weiter sei  $g$  die wie folgt definierte Funktion:

$$g(x) = ax + b.$$

Wir betrachten die Zufallsgröße  $Y$ ,

$$Y = g(X) = aX + b, \quad a \neq 0$$

und bezeichnen  $y := g(x)$ . Dann gilt:

$$g^{-1}(y) = x = \frac{y - b}{a}.$$

# Transformationsformel

## Beispiel 5 (Fortsetzung)

Für die Dichte der Zufallsvariable  $Y$  gilt nach dem Transformationssatz

$$h(y) = f(g^{-1}(y)) \cdot \left| \frac{dg^{-1}}{dy}(y) \right| = f\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|}$$

# Transformationsformel

## Beispiel 5 (Fortsetzung)

Für die Dichte der Zufallsvariable  $Y$  gilt nach dem Transformationssatz

$$h(y) = f(g^{-1}(y)) \cdot \left| \frac{dg^{-1}}{dy}(y) \right| = f\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|}$$

**Bem.:** Im Fall der Normalverteilung,  $X \sim \mathcal{N}(\mu, \sigma^2)$ ,  $\sigma > 0$ , haben wir dieses Ergebnis bereits früher erhalten.

# Transformationsformel

Lineare Transformation, Normal

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Es sei  $(a = \frac{1}{\sigma}, b = \frac{\mu}{\sigma})$

$$Y = \frac{X - \mu}{\sigma} \quad \text{bzw.} \quad X = \sigma Y + \mu.$$

Nach der in diesem Abschnitt hergeleiteten Formel ergibt sich die Dichtefunktion  $h$  der Zufallsgröße  $Y$ :

# Transformationsformel

Lineare Transformation, Normal

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Es sei  $(a = \frac{1}{\sigma}, b = \frac{\mu}{\sigma})$

$$Y = \frac{X - \mu}{\sigma} \quad \text{bzw.} \quad X = \sigma Y + \mu.$$

Nach der in diesem Abschnitt hergeleiteten Formel ergibt sich die Dichtefunktion  $h$  der Zufallsgröße  $Y$ :

$$\begin{aligned} h(y) &= \frac{1}{|a|} f\left(\frac{y-b}{a}\right) = \frac{1}{\left|\frac{1}{\sigma}\right|} f\left(\frac{y + \frac{\mu}{\sigma}}{\frac{1}{\sigma}}\right) = \sigma f(\sigma y + \mu) \\ &= \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\sigma y + \mu - \mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \end{aligned}$$

Dichtefunktion einer Normal mit  $\mu = 0$  und  $\sigma^2 = 1$ .

# Transformationsformel

Lineare Transformation, Normal (Fortsetzung)

D.h. Eine normalverteilte Zufallsgröße wird in eine standard-normalverteilte Zufallsgröße transformiert, indem der Parameter  $\mu$  subtrahiert und anschließend durch den Parameter  $\sigma$  dividiert wird. Sei also  $X \sim \mathcal{N}(\mu, \sigma^2)$ ,

$$\begin{aligned} F(x) &= P(X < x) = P\left(\underbrace{\frac{X - \mu}{\sigma}}_{=Y} < \frac{x - \mu}{\sigma}\right) \\ &= P\left(Y < \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

Es gilt:  $Y \sim \mathcal{N}(0, 1)$ . (vgl. auch Abschnitt Normalverteilung)

# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte



### Def. 49 (zufälliger Vektor)

Es seien  $X_i$ ,  $i = 1, \dots, p$ , reellwertige, zufällige Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{E}, P)$ . Dann heißt

$$\mathbf{X} = (X_1, \dots, X_p)^T: \quad \Omega \longrightarrow \mathbb{R}^p$$

zufälliger Vektor.

Er transformiert den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{E}, P)$  in den Wahrscheinlichkeitsraum  $(\mathbb{R}^p, \mathcal{B}^p, P_X)$ , wobei  $\mathcal{B}^p$  die  $\sigma$ -Algebra der  $p$ -dimensionalen Borelmengen ist.

# Mehrdimensionale Zufallsvariablen

## Def. 50 (Mehrdimensionale Verteilungsfunktion)

Die Funktion

$$F_{\mathbf{X}}(x_1, \dots, x_p) := P(\{\omega : X_1(\omega) < x_1, \dots, X_p(\omega) < x_p\})$$

heißt Verteilungsfunktion des zufälligen Vektors  $\mathbf{X}$ . Sie wird auch mit  $F_{X_1, \dots, X_p}(x_1, \dots, x_p)$  bezeichnet.

Es gilt:

$$F_{X_1, \dots, X_p}(x_1, \dots, x_p) = P\left(\bigcap_{i=1}^p \{\omega \in \Omega : X_i(\omega) < x_i\}\right).$$

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion

- 1 Invarianz gegenüber Permutationen, d.h.

$$F_{X_1, \dots, X_p}(x_1, \dots, x_p) = F_{X_{i_1}, \dots, X_{i_p}}(x_{i_1}, \dots, x_{i_p})$$

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion

- ① Invarianz gegenüber Permutationen, d.h.

$$F_{X_1, \dots, X_p}(x_1, \dots, x_p) = F_{X_{i_1}, \dots, X_{i_p}}(x_{i_1}, \dots, x_{i_p})$$

- ②  $\lim_{x_p \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_p) = F_{X_1, \dots, X_{p-1}}(x_1, \dots, x_{p-1});$   

$$F_X(x_1, \dots, x_p) = P(\underbrace{\{X_1 < x_1, \dots, X_{p-1} < x_{p-1}\}}_{=: A} \cap \underbrace{X_p < x_p}_{\rightarrow_{x_p \rightarrow \infty} \Omega}).$$

$$\begin{aligned} \lim_{x_p \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_p) &= P(A \cap \Omega) = P(A) \\ &= F_{X_1, \dots, X_{p-1}}(x_1, \dots, x_{p-1}). \end{aligned}$$

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion (2)

③  $\lim_{x_p \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_p) = 0;$

**Bem.:** Man kann wegen 1. auch jede beliebige Komponente wählen!

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion (2)

$$\textcircled{3} \quad \lim_{x_p \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_p) = 0;$$

**Bem.:** Man kann wegen 1. auch jede beliebige Komponente wählen!

$$\textcircled{4} \quad \lim_{(x_1, \dots, x_p) \rightarrow (\infty, \dots, \infty)} F_{\mathbf{X}}(x_1, \dots, x_p) = 1;$$

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion (2)

$$\textcircled{3} \quad \lim_{x_p \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_p) = 0;$$

**Bem.:** Man kann wegen 1. auch jede beliebige Komponente wählen!

$$\textcircled{4} \quad \lim_{(x_1, \dots, x_p) \rightarrow (\infty, \dots, \infty)} F_{\mathbf{X}}(x_1, \dots, x_p) = 1;$$

$$\textcircled{5} \quad F_{\mathbf{X}}(x_1, \dots, x_p) \text{ ist in jedem Argument monoton wachsend;}$$

# Mehrdimensionale Zufallsvariablen

## Eigenschaften der Verteilungsfunktion (2)

$$\textcircled{3} \quad \lim_{x_p \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_p) = 0;$$

**Bem.:** Man kann wegen 1. auch jede beliebige Komponente wählen!

$$\textcircled{4} \quad \lim_{(x_1, \dots, x_p) \rightarrow (\infty, \dots, \infty)} F_{\mathbf{X}}(x_1, \dots, x_p) = 1;$$

$$\textcircled{5} \quad F_{\mathbf{X}}(x_1, \dots, x_p) \text{ ist in jedem Argument monoton wachsend;}$$

$$\textcircled{6} \quad F_{\mathbf{X}}(x_1, \dots, x_p) \text{ ist in jedem Argument linksseitig stetig.}$$



# Mehrdimensionale Zufallsvariablen

## Stetige Verteilung

Def. 51 Ein zufälliger Vektor  $\mathbf{X} = (X_1, \dots, X_p)^T$  heißt stetig,

wenn seine Verteilungsfunktion charakterisiert ist durch:

$$F(x_1, \dots, x_p) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} \underbrace{f(t_1, \dots, t_p)}_{\text{Dichtefunktion}} dt_p \dots dt_1,$$

wobei für die Funktion  $f$  gilt:

- ①  $f(x_1, \dots, x_p) \geq 0, \forall x_1, \dots, x_p;$
- ②  $\int_{\mathbb{R}^p} f(x_1, \dots, x_p) dx_1 \dots dx_p = 1.$

# Mehrdimensionale Zufallsvariablen

## Stetige Verteilung (2)

Die Funktion  $f(x_1, \dots, x_p)$  heißt dann Dichtefunktion des zufälligen Vektors  $\mathbf{X}$ .

Falls die Dichtefunktion  $f(x_1, \dots, x_p)$  stetig ist, so gilt:

$$f(x_1, \dots, x_p) = \frac{\partial^p F_X(x_1, \dots, x_p)}{\partial x_1 \dots \partial x_p}.$$

# Mehrdimensionale Zufallsvariablen

Def. 52 Ein zufälliger Vektor  $\mathbf{X} = (X_1, \dots, X_p)^T$

heißt

- diskret, falls jede Komponente von  $\mathbf{X}$  diskret ist, d.h. jedes  $X_i$  besitzt höchstens abzählbar viele Argumente.
- gemischt, falls einige seiner Komponenten diskret, die restlichen dagegen stetig sind.
- stetig, falls alle Komponenten von  $\mathbf{X}$  stetige Zufallsgrößen sind.

# Mehrdimensionale Zufallsvariablen

$X$  diskret

Es sei  $\mathbf{X} = (X_1, \dots, X_p)^T$  ein diskreter zufälliger Vektor. Für  $i = 1, \dots, p$  habe  $X_i$  den Wertevorrat  $\{x_{i1}, \dots, x_{ik}, \dots\}$ . Dann definieren wir:

$$p_{j\dots k} = P(X_1 = x_{1j}, \dots, X_p = x_{pk}).$$

Verteilungsfunktion des zufälligen Vektors  $\mathbf{X}$ :

$$\begin{aligned} F(x_1, \dots, x_p) &= P\left(\bigcap_{i=1}^p \{\omega \in \Omega : X_i(\omega) < x_i\}\right) \\ &= \sum_{\substack{j: x_{1j} < x_1 \\ \dots \\ k: x_{pk} < x_p}} p_{j\dots k} \end{aligned}$$

# Mehrdimensionale Zufallsvariablen

$X$  diskret,  $p = 2$

Es sei  $p = 2$  und  $\mathbf{X} = (X_1, X_2)^T$ .

$$X_1 : \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}$$

$$X_2 : \begin{pmatrix} y_1 & y_2 & \dots & y_n & \dots \\ q_1 & q_2 & \dots & q_n & \dots \end{pmatrix}$$

$$p_{ij} = P(X_1 = x_i, X_2 = y_j) = P(\mathbf{X} = (x_i, y_j)).$$

# Mehrdimensionale Zufallsvariablen

$X$  diskret,  $p = 2$  (2)

Weiterhin gilt:

$$P(X_1 \in \{x_i : i \in \mathbb{N}\}) = 1$$

$$P(X_2 \in \{y_j : j \in \mathbb{N}\}) = 1$$

Wir bezeichnen:

$$\mathcal{X} := \{x_i : i \in \mathbb{N}\}, \quad \mathcal{Y} := \{y_j : j \in \mathbb{N}\}.$$

Der zufällige Vektor  $\mathbf{X}$  kann Werte der Form  $(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}$  annehmen,

$$P(\mathbf{X} \in \mathcal{X} \times \mathcal{Y}) = P(X_1 \in \mathcal{X}, X_2 \in \mathcal{Y}) = \sum_{i,j \in \mathbb{N}} p_{ij} = 1.$$

# Mehrdimensionale Zufallsvariablen

$X$  diskret,  $p = 2$  (3)

$$\begin{aligned} P(X_1 = x_i) &= \\ &= P(\{X_1 = x_i\} \cap \Omega) \\ &= P(\{X_1 = x_i\} \cap \underbrace{\{(X_2 = y_1) \vee (X_2 = y_2) \vee \dots \vee (X_2 = y_n) \vee \dots\}}_{= \bigcup_{j \in \mathbb{N}} \{X_2 = y_j\} = \Omega}) \\ &= P(\{X_1 = x_i\} \cap \left( \bigcup_{j \in \mathbb{N}} \{X_2 = y_j\} \right)) \\ &= P\left( \bigcup_{j \in \mathbb{N}} \{(X_1 = x_i) \wedge (X_2 = y_j)\} \right) = \sum_{j \in \mathbb{N}} p_{ij} =: p_{i.} \end{aligned}$$

# Mehrdimensionale Zufallsvariablen

$X$  diskret,  $p = 2$  (4)

Wir erhalten also:

$$p_{i\cdot} = P(X_1 = x_i).$$

Analog:

$$p_{\cdot j} = P(X_2 = y_j).$$

## Def. 53 (Randwahrscheinlichkeiten)

Die Wahrscheinlichkeiten  $p_{i\cdot}$  bzw.  $p_{\cdot j}$  ( $i, j \in \mathbb{N}$ ) nennen wir die Randwahrscheinlichkeiten des zufälligen Vektors  $X = (X_1, X_2)^T$ .

Die Zusammenhänge zwischen den einzelnen Wahrscheinlichkeiten werden in einer Kontingenztafel schematisiert.



# Mehrdimensionale Zufallsvariablen

Kontingenztafel

| $x_1 \backslash x_2$ | $y_1$         | $y_2$         | $y_3$         | $\dots$ | $y_j$         | $\dots$ | $y_n$         | $\dots$ | $\Sigma$     |
|----------------------|---------------|---------------|---------------|---------|---------------|---------|---------------|---------|--------------|
| $x_1$                | $p_{11}$      | $p_{12}$      | $p_{13}$      | $\dots$ | $p_{1j}$      | $\dots$ | $p_{1n}$      | $\dots$ | $p_{1\cdot}$ |
| $x_2$                | $p_{21}$      | $p_{22}$      | $p_{23}$      | $\dots$ | $p_{2j}$      | $\dots$ | $p_{2n}$      | $\dots$ | $p_{2\cdot}$ |
| $x_3$                | $p_{31}$      | $p_{32}$      | $p_{33}$      | $\dots$ | $p_{3j}$      | $\dots$ | $p_{3n}$      | $\dots$ | $p_{3\cdot}$ |
| $\vdots$             | $\vdots$      | $\vdots$      | $\vdots$      |         | $\vdots$      |         | $\vdots$      |         | $\vdots$     |
| $x_i$                | $p_{i1}$      | $p_{i2}$      | $p_{i3}$      | $\dots$ | $p_{ij}$      | $\dots$ | $p_{in}$      | $\dots$ | $p_{i\cdot}$ |
| $\vdots$             | $\vdots$      | $\vdots$      | $\vdots$      |         | $\vdots$      |         | $\vdots$      |         | $\vdots$     |
| $\Sigma$             | $p_{\cdot 1}$ | $p_{\cdot 2}$ | $p_{\cdot 3}$ | $\dots$ | $p_{\cdot j}$ | $\dots$ | $p_{\cdot n}$ | $\dots$ | 1            |

# Mehrdimensionale Zufallsvariablen

## Beispiel 1

### Umfrage zum Thema "Sport"

Dabei werden Männer und Frauen darüber befragt, ob sie Sportler oder Nichtsportler sind. Das ergibt die beiden folgenden Zufallsvariablen:

$$X_1 = \begin{cases} 1 & , \text{ falls weiblich} \\ 2 & , \text{ falls männlich} \end{cases}$$
$$X_2 = \begin{cases} 1 & , \text{ falls Sportler} \\ 2 & , \text{ falls Nichtsportler} \end{cases}$$

# Mehrdimensionale Zufallsvariablen

## Beispiel 1 (Fortsetzung)

Schema für den zufälligen Vektor

$$X = (X_1, X_2)^T:$$

| $x_1 \backslash x_2$ | 1        | 2        |          |
|----------------------|----------|----------|----------|
| 1                    | $p_{11}$ | $p_{12}$ | $p_{1.}$ |
| 2                    | $p_{21}$ | $p_{22}$ | $p_{2.}$ |
|                      | $p_{.1}$ | $p_{.2}$ | 1        |

# Mehrdimensionale Zufallsvariablen

## Beispiel 1 (Fortsetzung)

Schema für den zufälligen Vektor

$$X = (X_1, X_2)^T:$$

| $x_1 \backslash x_2$ | 1        | 2        |          |
|----------------------|----------|----------|----------|
| 1                    | $p_{11}$ | $p_{12}$ | $p_{1.}$ |
| 2                    | $p_{21}$ | $p_{22}$ | $p_{2.}$ |
|                      | $p_{.1}$ | $p_{.2}$ | 1        |

$2 \times 2$ -Kontingenztafel:

| $x_1 \backslash x_2$ | 1        | 2        |          |
|----------------------|----------|----------|----------|
| 1                    | $n_{11}$ | $n_{12}$ | $n_{1.}$ |
| 2                    | $n_{21}$ | $n_{22}$ | $n_{2.}$ |
|                      | $n_{.1}$ | $n_{.2}$ | $n_{..}$ |

Dabei bedeuten:

# Mehrdimensionale Zufallsvariablen

## Beispiel 1 (Fortsetzung, 2)

$n_{ij}$  – die Anzahl der Personen mit dem Geschlecht  $i$  und dem Sportverhalten  $j$ ;

$n_{.1}$  – die Anzahl der Sportler;

$n_{.2}$  – die Anzahl der Nichtsportler;

$n_{1.}$  – die Anzahl der Frauen;

$n_{2.}$  – die Anzahl der Männer;

$n_{..}$  – die Gesamtzahl der Befragten.

Mit  $\hat{p}_{ij} = \frac{n_{ij}}{n_{..}}$  ergibt sich nun eine Schätzung für die Wahrscheinlichkeit  $p_{ij}$ .

# Mehrdimensionale Zufallsvariablen

## Beispiel 2

### Werfen zweier Würfel

Wir betrachten den zufälligen Vektor  $X = (X_1, X_2)^T$ , wobei  $X_1$  die Augenzahl des ersten Würfels ist und  $X_2$  die des zweiten. Für die zufälligen Variablen  $X_1$  und  $X_2$  gilt:

$$X_1, X_2 : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Da die Würfel voneinander unabhängig sind, gilt

$$p_{ij} = P(X_1 = i, X_2 = j) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

# Mehrdimensionale Zufallsvariablen

## Beispiel 2 (Fortsetzung)

Damit erhalten wir das folgende Schema:

| $x_1 \backslash x_2$ | 1              | 2              | 3              | 4              | 5              | 6              |
|----------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 1                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 2                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 3                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 4                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 5                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 6                    | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |

# Mehrdimensionale Zufallsvariablen

## Beispiel 2 (Fortsetzung)

$$P(X_1 < 4, X_2 < 3) = \sum_{i < 4; j < 3} p_{ij} = \frac{6}{36} = \frac{1}{6}$$

Die hier addierten Wahrscheinlichkeiten sind in dem oben angegebenen Schema eingerahmt.

Die Aussagen zu zweidimensionalen zufälligen Vektoren, die wir bis hierher gemacht haben, gelten analog erweitert auch für höherdimensionale zufällige Vektoren.



# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$

## Zweidimensionale Dichtefunktion $f(x, y)$

- 1  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx \, dy = 1;$
- 2  $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2.$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$

## Zweidimensionale Dichtefunktion $f(x, y)$

- ❶  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1;$
- ❷  $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2.$

## Zweidimensionale Verteilungsfunktion

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = P(X_1 < x, X_2 < y).$$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$

## Zweidimensionale Dichtefunktion $f(x, y)$

- 1  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1;$
- 2  $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2.$

## Zweidimensionale Verteilungsfunktion

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = P(X_1 < x, X_2 < y).$$

Da  $f(x, y)$  stetig ist, gilt weiterhin:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$  (Fortsetzung)

$$\lim_{y \rightarrow \infty} F(x, y) = F_{X_1}(x) = P(X_1 < x).$$

$$\lim_{x \rightarrow \infty} F(x, y) = F_{X_2}(y) = P(X_2 < y).$$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$  (Fortsetzung)

$$\lim_{y \rightarrow \infty} F(x, y) = F_{X_1}(x) = P(X_1 < x).$$

$$\lim_{x \rightarrow \infty} F(x, y) = F_{X_2}(y) = P(X_2 < y).$$

## Randverteilungen, Randverteilungsfunktionen

Die Verteilungsfunktionen  $F_{X_1}$  und  $F_{X_2}$  bezeichnen wir als Randverteilungen von  $X_1$  bzw.  $X_2$ .

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$  (Fortsetzung)

Integrieren wir die Dichtefunktion nur nach einer der beiden Variablen, so erhalten wir:

$$\int_{-\infty}^{+\infty} f(x, y) dy = \frac{dF_{X_1}(x)}{dx} =: f_{X_1}(x)$$
$$\int_{-\infty}^{+\infty} f(x, y) dx = \frac{dF_{X_2}(y)}{dy} =: f_{X_2}(y)$$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$  (Fortsetzung, 2)

## Def. 54 (Randdichten)

Die Funktionen  $f_{X_1}$  und  $f_{X_2}$  bezeichnen wir als Randdichten von  $X_1$  bzw.  $X_2$ .

Offenbar,

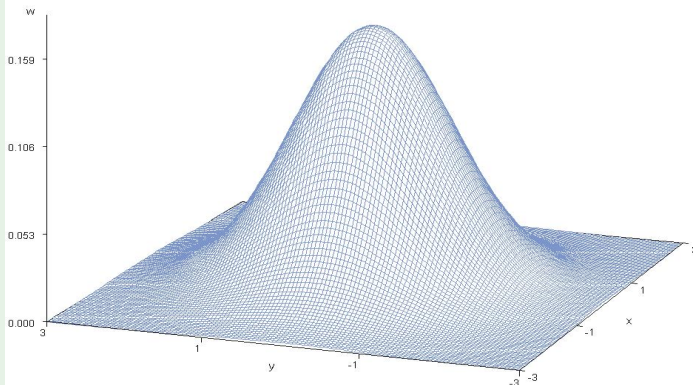
$$F_{X_1}(x) = \int_{-\infty}^x f_{X_1}(t) dt$$
$$F_{X_2}(y) = \int_{-\infty}^y f_{X_2}(t) dt$$

# Mehrdimensionale Zufallsvariablen

$X$  stetig,  $p = 2$  (Fortsetzung)

## Zweidimensionale Normalverteilung

**Dichtefunktion der 2-dimensionalen Standard-Normalverteilung**





# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

### Def. 55 (Unabhängigkeit)

Es seien  $X_1$  und  $X_2$  zwei zufällige Variablen auf dem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{E}, P)$ . Diese beiden zufälligen Variablen  $X_1$  und  $X_2$  heißen stochastisch unabhängig, wenn für alle  $A, B \in \mathcal{B}^1$  gilt:

- $P(X_1 \in A, X_2 \in B) = P(X_1 \in A) \cdot P(X_2 \in B)$ ;  
oder kürzer:
- $P(A \cap B) = P(A) \cdot P(B)$ .

# Unabhängigkeit von Zufallsgrößen

## Verteilungsfunktion

Es sei  $X = (X_1, X_2)^T$  ein zufälliger Vektor, deren Komponenten  $X_1$  und  $X_2$  stochastisch unabhängig sind. Dann gilt:

$$\begin{aligned} F_{X_1, X_2}(x_1, x_2) &= P(X_1 < x_1, X_2 < x_2) \\ &= P(X_1 \in \underbrace{(-\infty, x_1)}_{A \in \mathcal{B}^1}, X_2 \in \underbrace{(-\infty, x_2)}_{B \in \mathcal{B}^1}) \\ &= P(X_1 \in (-\infty, x_1)) \cdot P(X_2 \in (-\infty, x_2)) \\ &= F_{X_1}(x_1) \cdot F_{X_2}(x_2) \end{aligned}$$

# Unabhängigkeit von Zufallsgrößen

$F$  stetig

Sei  $F$  stetig mit Dichte  $f_{\mathbf{X}}(t_1, t_2)$ .

Aus der letzten Aussage folgt:

$$\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{\mathbf{X}}(t_1, t_2) dt_1 dt_2 = \int_{-\infty}^{x_1} f_{X_1}(t_1) dt_1 \int_{-\infty}^{x_2} f_{X_2}(t_2) dt_2$$

$$\int_{-\infty}^{x_1} \int_{-\infty}^{x_2} (f_{\mathbf{X}}(t_1, t_2) - f_{X_1}(t_1)f_{X_2}(t_2)) dt_1 dt_2 = 0 \quad \forall x_1, x_2 \in \mathbb{R}$$

$$\text{D.h.} \quad f_{\mathbf{X}}(t_1, t_2) = f_{X_1}(t_1)f_{X_2}(t_2) \quad \forall t_1, t_2$$

# Unabhängigkeit von Zufallsgrößen

- Ist der zufällige Vektor  $X = (X_1, X_2)^T$  stetig, so

$$\underbrace{f_{X_1, X_2}(x_1, x_2)}_{\text{zweidimensionale Dichte}} = \underbrace{f_{X_1}(x_1) \cdot f_{X_2}(x_2)}_{\text{Randdichten}}.$$

- Ist der zufällige Vektor  $X = (X_1, X_2)^T$  diskret, so folgt für alle  $i, j = 1, \dots$ :

$$p_{ij} = p_{i.} \cdot p_{.j}.$$

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3 | $p_{i.}$    |
|-----------------|-------------|-------------|---|-------------|
| -1              |             | <b>0.06</b> |   |             |
| 0               | <b>0.03</b> |             |   |             |
| 1               |             |             |   | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> |   | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3 | $p_{i.}$    |
|-----------------|-------------|-------------|---|-------------|
| -1              |             | <b>0.06</b> |   | <b>0.20</b> |
| 0               | <b>0.03</b> |             |   |             |
| 1               |             |             |   | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> |   | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3 | $p_{i.}$    |
|-----------------|-------------|-------------|---|-------------|
| -1              |             | <b>0.06</b> |   | 0.20        |
| 0               | <b>0.03</b> |             |   | 0.30        |
| 1               |             |             |   | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> |   | 1           |



# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3 | $p_{i.}$    |
|-----------------|-------------|-------------|---|-------------|
| -1              |             | <b>0.06</b> |   | 0.20        |
| 0               | <b>0.03</b> | 0.09        |   | 0.30        |
| 1               |             |             |   | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> |   | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              |             | <b>0.06</b> |      | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             |      | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> |      | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              |             | <b>0.06</b> |      | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             |      | <b>0.50</b> |
| $p_{.j}$        |             | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              |             | <b>0.06</b> |      | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             |      | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              | 0.02        | <b>0.06</b> |      | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             |      | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              | 0.02        | <b>0.06</b> | 0.12 | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             |      | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              | 0.02        | <b>0.06</b> | 0.12 | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             |             | 0.30 | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              | 0.02        | <b>0.06</b> | 0.12 | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               |             | 0.15        | 0.30 | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |



# Unabhängigkeit von Zufallsgrößen

## Beispiel

Es seien einige Einzelwahrscheinlichkeiten  $p_{ij}$  einer diskreten zweidimensionalen Zufallsvariablen  $(X, Y)$  bekannt (fett eingetragen).

Die Komponenten  $X$  und  $Y$  seien unabhängig. Bestimmen Sie die restlichen Einträge!

| $X \setminus Y$ | 1           | 2           | 3    | $p_{i.}$    |
|-----------------|-------------|-------------|------|-------------|
| -1              | 0.02        | <b>0.06</b> | 0.12 | 0.20        |
| 0               | <b>0.03</b> | 0.09        | 0.18 | 0.30        |
| 1               | 0.05        | 0.15        | 0.30 | <b>0.50</b> |
| $p_{.j}$        | 0.10        | <b>0.30</b> | 0.60 | 1           |

# Unabhängigkeit von Zufallsgrößen

Beispiel (Fortsetzung)

$$\mathbf{E}X = -1 \cdot 0.2 + 0 \cdot 0.3 + 1 \cdot 0.5 = 0.3$$

$$\mathbf{E}Y = 1 \cdot 0.1 + 2 \cdot 0.3 + 3 \cdot 0.6 = 2.5$$

# Unabhängigkeit von Zufallsgrößen

Beispiel (Fortsetzung)

$$\mathbf{E}X = -1 \cdot 0.2 + 0 \cdot 0.3 + 1 \cdot 0.5 = 0.3$$

$$\mathbf{E}Y = 1 \cdot 0.1 + 2 \cdot 0.3 + 3 \cdot 0.6 = 2.5$$

$$\begin{aligned}\mathbf{E}(X \cdot Y) &= -0.02 - 2 \cdot 0.06 - 3 \cdot 0.12 + 0 \cdot (\dots) \\ &\quad + 1 \cdot 0.05 + 2 \cdot 0.15 + 3 \cdot 0.3 = 0.75\end{aligned}$$

$$\text{cov}(X, Y) = \mathbf{E}(X \cdot Y) - (\mathbf{E}X)(\mathbf{E}Y) = 0.75 - 0.75 = 0.$$

Merkwürdig?

# Unabhängigkeit von Zufallsgrößen

## Satz

Es seien  $X_1$  und  $X_2$  zwei zufällige Variablen.

$\varphi$  und  $\psi$  seien zwei beliebige ( $\mathcal{B}^1$ -messbare) Transformationen dieser beiden Variablen,

$$X'_1 = \varphi(X_1), \quad X'_2 = \psi(X_2).$$

Die zufälligen Variablen  $X_1$  und  $X_2$  sind genau dann stochastisch unabhängig, wenn die Zufallsgrößen  $X'_1$  und  $X'_2$ , für alle Transformationen  $\varphi$  und  $\psi$ , unabhängig sind.

# Unabhängigkeit von Zufallsgrößen

Beweis des Satzes, Anmerkungen

Die Funktionen  $\varphi$  und  $\psi$  seien auf der Menge  $\mathbb{R}$  definiert und reellwertig. Dann gilt für die jeweilige Umkehrfunktion genau dann

$$\begin{aligned}\varphi^{-1}(A) &= \{x: \varphi(x) \in A\} \in \mathcal{B}^1, \quad \forall A \in \mathcal{B}^1 \\ \psi^{-1}(B) &= \{y: \psi(y) \in B\} \in \mathcal{B}^1, \quad \forall B \in \mathcal{B}^1,\end{aligned}$$

wenn  $\varphi$  und  $\psi$   $\mathcal{B}^1$ -messbar sind.

# Unabhängigkeit von Zufallsgrößen

Beweis des Satzes ( $\implies$ )

Es seien die zufälligen Variablen  $X_1$  und  $X_2$  stochastisch unabhängig.

Wir zeigen, dass  $\varphi(X_1)$  und  $\psi(X_2)$  unabhängig sind. Da die

Funktionen  $\varphi$  und  $\psi$   $\mathcal{B}^1$ -messbar sind, gilt

$$P(\varphi(X_1) \in A, \psi(X_2) \in B)$$

$$= P(X_1 \in \varphi^{-1}(A), X_2 \in \psi^{-1}(B))$$

$$= P(X_1 \in \varphi^{-1}(A)) \cdot P(X_2 \in \psi^{-1}(B))$$

$$= P(\varphi(X_1) \in A) \cdot P(\psi(X_2) \in B)$$

D.h. die zufälligen Variablen  $\varphi(X_1)$  und  $\psi(X_2)$  sind unabhängig.

# Unabhängigkeit von Zufallsgrößen

Beweis des Satzes ( $\Leftarrow$ )

Es gelte also, daß für alle  $\mathcal{B}^1$ -messbaren Funktionen  $\varphi$  und  $\psi$  die zufälligen Variablen  $\varphi(X_1)$  und  $\psi(X_2)$  unabhängig sind. Insbesondere ist das dann auch der Fall für die Funktionen  $\varphi(x) \equiv \psi(x) \equiv x$ . D.h.

$$X_1 = \varphi(X_1), \quad X_2 = \psi(X_2).$$

Folglich sind auch die zufälligen Variablen  $X_1$  und  $X_2$  unabhängig.

# Unabhängigkeit von Zufallsgrößen

## Beispiel 2

Sei  $X \sim \mathcal{N}(0, 1)$ .

- $X$  und  $Y = X^2$  sind nicht unabhängig, sogar funktional abhängig



# Unabhängigkeit von Zufallsgrößen

## Beispiel 2

Sei  $X \sim \mathcal{N}(0, 1)$ .

- $X$  und  $Y = X^2$  sind nicht unabhängig, sogar funktional abhängig
- $X$  und  $Y$  sind unkorreliert, wegen  $\mathbf{E}X = 0$  und

$$\text{cov}(X, Y) = \mathbf{E}(X \cdot X^2) - \mathbf{E}X \cdot \mathbf{E}Y = \mathbf{E}X^3 = 0,$$

da  $X$  symmetrisch ist.

Die Aussage gilt also für beliebige symmetrische Zufallsvariablen  $X$  mit endlicher Varianz.

# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

Es sei  $\mathbf{X} = (X_1, \dots, X_p)^T$  ein zufälliger Vektor mit der Dichtefunktion  $f(x_1, \dots, x_p)$ . Es sei  $g: \mathbb{R}^p \longrightarrow \mathbb{R}^p$  eine umkehrbar eindeutige Abbildung. Sie ordnet einem Vektor  $\mathbf{x} = (x_1, \dots, x_p)^T$  einen Vektor  $\mathbf{y} = (y_1, \dots, y_p)^T$  zu und besteht aus Teilabbildungen  $g_1, \dots, g_p$  mit  $g_i: \mathbb{R}^p \longrightarrow \mathbb{R}$  (für alle  $i = 1, \dots, p$ ).

### Beispiel

$\mathbf{y} = g(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x}$ , wobei  $\mathbf{A}$  reguläre  $(p, p)$ -Matrix.

## Transformationssatz für Zufallsvektoren (2)

Die Umkehrabbildung  $g^{-1}: \mathbb{R}^p \longrightarrow \mathbb{R}^p$  ist durch Funktionen  $x_i = \psi_i(y_1, \dots, y_p)$  definiert ( $i = 1, \dots, p$ ). Die Funktionen  $\psi_i$  ( $i = 1, \dots, p$ ) existieren wegen der umkehrbaren Eindeutigkeit der Funktion  $g$ .

$$g^{-1}(\mathbf{y}) = g^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \psi_1(y_1, \dots, y_p) \\ \vdots \\ \psi_p(y_1, \dots, y_p) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

$$g^{-1}(\mathbf{y}) = (\psi_1(\mathbf{y}), \dots, \psi_p(\mathbf{y}))^T = \mathbf{x} \quad \text{Kurzform}$$

## Transformationssatz für Zufallsvektoren (3)

Wir definieren einen weiteren zufälligen Vektor

$\mathbf{Y} = (Y_1, \dots, Y_p)^T$  wie folgt:

$$\mathbf{Y} = g(\mathbf{X}) := (g_1(X_1, \dots, X_p), \dots, g_p(X_1, \dots, X_p))^T$$

und nehmen an, die  $g_i$  ( $i = 1, \dots, p$ ) besitzen stetige partielle Ableitungen nach allen Argumenten.

Für den zufälligen Vektor  $\mathbf{X}$  gilt umgekehrt:

$$\begin{aligned}\mathbf{X} &= (X_1, \dots, X_p)^T \\ &= (\psi_1(Y_1, \dots, Y_p), \dots, \psi_p(Y_1, \dots, Y_p))^T \\ &= g^{-1}(Y_1, \dots, Y_p) = g^{-1}(\mathbf{Y}).\end{aligned}$$

# Transformationssatz für Zufallsvektoren (4)

Satz (Dichte von  $\mathbf{Y} = g(\mathbf{X})$ ), ohne Beweis

Die Zufallsvariable  $\mathbf{X}$  habe die Dichte  $f$ . Die Dichte der Zufallsvariablen  $\mathbf{Y} = g(\mathbf{X})$  ist

$$h_Y(y_1, \dots, y_p) = f(\psi_1(y_1, \dots, y_p), \dots, \psi_p(y_1, \dots, y_p)) \cdot |J|,$$

wobei

$$J = \det \left( \frac{\partial \psi_i(y_1, \dots, y_p)}{\partial y_j} \right)_{i,j=1,\dots,p}$$

die sogenannte Jacobi-Determinante ist.

# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

## BOX-MÜLLER-Transformation

Es seien  $U_1$  und  $U_2$  zwei unabhängige, über dem Intervall  $[0, 1[$  gleichverteilte Zufallsgrößen ( $U_i \sim R(0, 1)$ ,  $i = 1, 2$ ),  $\mathbf{U} = (U_1, U_2)^T$  ein zufälliger Vektor. Wir betrachten den zufälligen Vektor  $\mathbf{V} = g(\mathbf{U}) = (X, Y)^T$ , wobei:

$$X = g_1(U_1, U_2) = \sqrt{-2 \ln U_1} \cdot \cos 2\pi U_2$$

$$Y = g_2(U_1, U_2) = \sqrt{-2 \ln U_1} \cdot \sin 2\pi U_2$$

Wir suchen die Dichtefunktionen für die zufälligen Variablen  $X$  und  $Y$ .



## Box-Müller Transformation (2)

Wir bestimmen zunächst die Umkehrfunktion zur Abbildung  $g$ . Es gilt:

$$\mathbf{U} = g^{-1}(\mathbf{V}) = (\psi_1(X, Y), \psi_2(X, Y)).$$

Zur Bestimmung der  $\psi_1$  und  $\psi_2$  berechnen wir

## Box-Müller Transformation (2)

Wir bestimmen zunächst die Umkehrfunktion zur Abbildung  $g$ . Es gilt:

$$\mathbf{U} = g^{-1}(\mathbf{V}) = (\psi_1(X, Y), \psi_2(X, Y)).$$

Zur Bestimmung der  $\psi_1$  und  $\psi_2$  berechnen wir

$$\begin{aligned} X^2 + Y^2 &= (-2 \ln U_1 \cdot \cos^2(2\pi U_2)) + \\ &\quad (-2 \ln U_1 \cdot \sin^2(2\pi U_2)) \\ &= (-2 \ln U_1) \cdot (\cos^2(2\pi U_2) + \sin^2(2\pi U_2)) \\ &= -2 \ln U_1 \end{aligned}$$

## Box-Müller Transformation (3)

Durch Umstellen erhalten wir:

$$U_1 = \psi_1(X, Y) = e^{-\frac{1}{2}(X^2+Y^2)}.$$

Die zweite Komponente erhalten wir durch

## Box-Müller Transformation (3)

Durch Umstellen erhalten wir:

$$U_1 = \psi_1(X, Y) = e^{-\frac{1}{2}(X^2 + Y^2)}.$$

Die zweite Komponente erhalten wir durch

$$\frac{Y}{X} = \tan 2\pi U_2.$$

Daraus folgt:

$$U_2 = \psi_2(X, Y) = \frac{1}{2\pi} \arctan \left( \frac{Y}{X} \right).$$

# Box-Müller Transformation (4)

Bestimmung von  $|J|$ .

$$|J| = \begin{vmatrix} \frac{\partial \psi_1}{\partial x} & \frac{\partial \psi_1}{\partial y} \\ \frac{\partial \psi_2}{\partial x} & \frac{\partial \psi_2}{\partial y} \end{vmatrix}$$

# Box-Müller Transformation (4)

Bestimmung von  $|J|$ .

$$\begin{aligned}
 |J| &= \left\| \begin{array}{cc} \frac{\partial \psi_1}{\partial x} & \frac{\partial \psi_1}{\partial y} \\ \frac{\partial \psi_2}{\partial x} & \frac{\partial \psi_2}{\partial y} \end{array} \right\| \\
 &= \left\| \begin{array}{cc} -x \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) & -y \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \\ \frac{1}{2\pi} \cdot \frac{-y}{\left(1 + \frac{y^2}{x^2}\right) \cdot x^2} & \frac{1}{2\pi} \cdot \frac{1}{\left(1 + \frac{y^2}{x^2}\right) \cdot x} \end{array} \right\|
 \end{aligned}$$

# Box-Müller Transformation (4)

Bestimmung von  $|J|$ .

$$\begin{aligned}
 |J| &= \left\| \begin{array}{cc} \frac{\partial \psi_1}{\partial x} & \frac{\partial \psi_1}{\partial y} \\ \frac{\partial \psi_2}{\partial x} & \frac{\partial \psi_2}{\partial y} \end{array} \right\| \\
 &= \left\| \begin{array}{cc} -x \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) & -y \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \\ \frac{1}{2\pi} \cdot \frac{-y}{\left(1 + \frac{y^2}{x^2}\right) \cdot x^2} & \frac{1}{2\pi} \cdot \frac{1}{\left(1 + \frac{y^2}{x^2}\right) \cdot x} \end{array} \right\| \\
 &= \left| -\frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \cdot \left(\frac{x^2}{x^2 + y^2} + \frac{y^2}{x^2 + y^2}\right) \right| \\
 &= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)}
 \end{aligned}$$

## Box-Müller Transformation (5)

Für die Dichtefunktion des zufälligen Vektors  $\mathbf{V}$  gilt nach der Transformationsformel:

$$f_{\mathbf{V}}(x, y) = f_{\mathbf{U}}(\psi_1(x, y), \psi_2(x, y)) \cdot |J|.$$

Da die Zufallsgrößen  $U_1$  und  $U_2$  unabhängig sind, gilt:

$$f_{\mathbf{V}}(x, y) = f_{U_1}(\psi_1(x, y)) \cdot f_{U_2}(\psi_2(x, y)) \cdot |J|.$$

Nun sind  $U_1, U_2 \sim R(0, 1)$ . Daraus folgt:

$$\begin{aligned} f_{\mathbf{V}}(x, y) &= |J| = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \\ &= f_X(x) \cdot f_Y(y). \end{aligned}$$



# Box-Müller Transformation (6)

mit

$$\begin{aligned}f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \\f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\end{aligned}$$

d.h. die Zufallsgrößen  $X$  und  $Y$  sind unabhängig und standardnormalverteilt,

$$X \sim \mathcal{N}(0, 1), \quad Y \sim \mathcal{N}(0, 1).$$

# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

# Transformationssatz

Treffen einer Zielscheibe\*

Es seien folgende Bedingungen erfüllt

- V1: Die Randverteilungen von  $X$  und  $Y$  seien stetig
- V2: Die Dichte  $h(x, y)$  von  $(X, Y)$  hängt nur vom Abstand  $\sqrt{x^2 + y^2}$  vom Nullpunkt ab (Radialsymmetrie)
- V3: Die Fehler in  $x$ - und  $y$ -Richtung sind unabhängig.

Sei  $Z$  die zufällige Abweichung in beliebiger Richtung. Dann ist  $Z \sim \mathcal{N}(0, \sigma^2)$ .

# Treffen einer Zielscheibe, Beweis des Satzes (1)

Seien  $p(x)$  und  $q(y)$  Randdichten von  $(X, Y)$ . Aus V2 und V3 folgt

$$p(x)q(y) = s(r), \quad r^2 = x^2 + y^2 \quad (2)$$

Substitutionsmethode:

$$x = 0: p(0)q(y) = s(y), \quad p(0) \neq 0$$

$$y = 0: q(0)p(x) = s(x), \quad q(0) \neq 0$$

$$x \neq y: p(x)q(y) = p(y)q(x) \quad \forall x, y,$$

und damit  $p(x) = q(x)$  und  $p(0)p(y) = s(y)$ .

Teilen obige Funktionalgleichung durch  $p(0)^2$ ,

$$\frac{p(x)}{p(0)} \frac{p(y)}{p(0)} = \frac{s(r)}{p(0)^2} = \frac{p(r)}{p(0)}$$

# Treffen einer Zielscheibe

Beweis des Satzes (2)

Logarithmieren

$$\ln\left(\frac{p(x)}{p(0)}\right) + \ln\left(\frac{p(y)}{p(0)}\right) = \ln\left(\frac{p(r)}{p(0)}\right)$$

Mit  $f(x) := \ln\left(\frac{p(x)}{p(0)}\right)$ :

$$f(x) + f(y) = f(r), \quad r^2 = x^2 + y^2$$

weiter Substitutionsmethode.

$$y = 0: f(-x) = f(|x|) \quad \text{wegen } f(0) = 0.$$

# Treffen einer Zielscheibe

Beweis des Satzes (3)

$$x^2 = x_1^2 + x_2^2:$$

$$f(r) = f(y) + f(x_1) + f(x_2), \quad r^2 = y^2 + x_1^2 + x_2^2$$

Wiederholtes Einsetzen:

$$f(r) = f(x_1) + f(x_2) + \dots + f(x_k), \quad r^2 = \sum_{i=1}^k x_i^2$$

$$k = n^2, x = x_1 = \dots = x_k:$$

$$f(nx) = n^2 f(x) \Rightarrow_{x=1} f(n) = n^2 f(1)$$

$$x = \frac{m}{n}, m \in \mathbb{Z}:$$

$$n^2 f\left(\frac{m}{n}\right) = f\left(n \frac{m}{n}\right) = f(m) = m^2 f(1)$$

# Treffen einer Zielscheibe

Beweis des Satzes (4)

$$\Rightarrow f\left(\frac{m}{n}\right) = f(1)\left(\frac{m}{n}\right)^2 \Rightarrow$$

$$f(x) = cx^2, \quad c = f(1)$$

für alle rationalen  $x$ . Wegen der Stetigkeit (V1) folgt diese Relation für alle  $x \in \mathbb{R}$ .

$$p(x) = p(0)e^{cx^2}$$

$p(x) > 0$  da Wkt.dichte,  $c < 0$ ,  $c := -\frac{1}{2\sigma^2}$ .

$$1 = \int_{-\infty}^{\infty} p(x) dx = p(0) \int_{-\infty}^{\infty} e^{cx^2} dx = p(0)\sigma\sqrt{2\pi}$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

# Treffen einer Zielscheibe

Beweis des Satzes (5)

Gemeinsame Dichte von  $(X, Y)$ :

$$p(x)p(y) = \frac{1}{\sigma^2 2\pi} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Fehler in einer beliebigen Richtung  $\theta$ ,  $0 \leq \theta \leq 2\pi$ :

$$Z = X \cos(\theta) + Y \sin(\theta)$$

Variablentransformation

$$z = x \cos(\theta) + y \sin(\theta)$$

$$u = x \sin(\theta) - y \cos(\theta)$$

$$\text{Jacobi-Determinante } J = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{vmatrix} = |-1| = 1$$



# Treffen einer Zielscheibe

Beweis des Satzes (6)

Quadrieren von  $z$  und  $u$  liefert

$$z^2 = x^2 \cos^2(\theta) + y^2 \sin^2(\theta) + 2xy \cos(\theta) \sin(\theta)$$

$$u^2 = x^2 \sin^2(\theta) + y^2 \cos^2(\theta) - 2xy \cos(\theta) \sin(\theta)$$

Addition:  $x^2 + y^2 = z^2 + u^2$  also gemeinsame Dichte von  $(Z, U)$ :

$$h_1(z, u) = \frac{1}{\sigma^2 2\pi} e^{-\frac{z^2+u^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}$$

d.h.  $Z$  und  $U$  sind unabhängig,  $h_1(z, u) = h_Z(z)h_U(u)$  und

$$h_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$$

# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

Seien zwei unabhängige stetige Zufallsvariablen  $X_1, X_2$  mit Dichte  $f_{X_1}$  bzw.  $f_{X_2}$  gegeben. Gesucht ist die Dichte der Zufallsvariablen  $X_1 + X_2$ . Diese liefert die Faltungsformel. Wir leiten die Faltungsformel zunächst mit Hilfe des Transformationssatzes her. Später werden wir noch einen anderen Beweis kennen lernen, der den Satz der Totalen Wahrscheinlichkeit für stetige Zufallsvariablen verwendet.

## Faltung

Es sei  $\mathbf{X} = (X_1, X_2)^T$  ein zufälliger Vektor ( $p = 2$ ), mit unabhängigen Komponenten  $X_1$  und  $X_2$ . Die Dichte  $f_{X_1, X_2}$  von  $\mathbf{X}$  ist  $f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$ . Es sei  $\mathbf{Y} = g(\mathbf{X})$ ,

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} g_1(X_1, X_2) \\ g_2(X_1, X_2) \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_2 \end{pmatrix}.$$

## Faltung (2)

Wir suchen die Dichte des zufälligen Vektors  $\mathbf{Y} = (Y_1, Y_2)$ . Die beiden Teilkomponenten von  $g$  sind

$$g_1(x_1, x_2) = y_1 = x_1 + x_2$$

$$g_2(x_1, x_2) = y_2 = x_2$$

Die Umkehrfunktion  $g^{-1}$  besteht aus den beiden Teilfunktionen:

$$\psi_1(y_1, y_2) = x_1 = y_1 - y_2$$

$$\psi_2(y_1, y_2) = x_2 = y_2$$

$$|J| = \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} \right| = \left| \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right| = 1$$

## Faltung (3)

Dichte des zufälligen Vektors  $\mathbf{Y} = (X_1 + X_2, X_2)$ :

$$\begin{aligned}h_{\mathbf{Y}}(y_1, y_2) &= f_{X_1, X_2}(\psi_1(y_1, y_2), \psi_2(y_1, y_2)) \cdot |J| \\&= f_{X_1, X_2}(y_1 - y_2, y_2) \\&= f_{X_1}(y_1 - y_2) \cdot f_{X_2}(y_2)\end{aligned}$$

## Faltung (3)

Dichte des zufälligen Vektors  $\mathbf{Y} = (X_1 + X_2, X_2)$ :

$$\begin{aligned}h_{\mathbf{Y}}(y_1, y_2) &= f_{X_1, X_2}(\psi_1(y_1, y_2), \psi_2(y_1, y_2)) \cdot |J| \\&= f_{X_1, X_2}(y_1 - y_2, y_2) \\&= f_{X_1}(y_1 - y_2) \cdot f_{X_2}(y_2)\end{aligned}$$

Randdichte für  $Y_1 = X_1 + X_2$ :

$$\begin{aligned}h_{Y_1}(y_1) &= \int_{-\infty}^{+\infty} h_{\mathbf{Y}}(y_1, y_2) dy_2 \\&= \int_{-\infty}^{+\infty} f_{X_1}(y_1 - y_2) \cdot f_{X_2}(y_2) dy_2 =: f_{X_1} * f_{X_2}(y)\end{aligned}$$

# Faltung (4)

## Def. 56 (Faltung)

Die Verknüpfung  $f_{X_1} * f_{X_2}$  zweier Funktionen  $f_1$  und  $f_2$  heißt Faltung aus  $f_1$  und  $f_2$ .

**Bem.:** Die Dichte der Summe zweier unabhängiger Zufallsvariablen ist Faltung der beiden Einzeldichten.

$X_1, X_2 \sim R(0, 1)$ ,  $\mathbf{Y}$  wie im letzten Beispiel

Dichtefunktion von  $Y_1 = X_1 + X_2$ :

$$h_{Y_1}(y) = \int_0^1 f_{X_1}(y-x) \cdot \underbrace{f_{X_2}(x)}_{\equiv 1} dx = \int_0^1 f_{X_1}(y-x) dx$$

## Faltung (5)

Es gilt:  $0 \leq X_i < 1$ ,  $i = 1, 2$ , d.h.

$$0 \leq X_1 + X_2 = Y_1 < 2.$$

und für die Funktion  $f_{X_1}$ :

$$\begin{aligned} f_{X_1}(y - x) &= \begin{cases} 1 & , \text{ falls } 0 \leq y - x \leq 1 \\ 0 & , \text{ sonst} \end{cases} \\ &= \begin{cases} 1 & , \text{ falls } 0 \leq x < y \leq 1 \\ 1 & , \text{ falls } 0 \leq y - 1 \leq x \leq 1 < y \\ 0 & , \text{ falls } y - x \notin [0, 1[ \end{cases} \end{aligned}$$



# Faltung (6)

Randdichte  $Y_1$  von  $\mathbf{Y}$

$$\begin{aligned} h_{Y_1}(y) &= \int_0^1 f_{X_1}(y-x) dx \\ &= \begin{cases} \int_0^y dx & , \text{ falls } 0 \leq y \leq 1 \\ \int_{y-1}^1 dx & , \text{ falls } 1 < y < 2 \\ 0 & , \text{ falls } y \notin [0, 2[ \end{cases} \\ &= \begin{cases} y & , \text{ falls } 0 \leq y \leq 1 \\ 2-y & , \text{ falls } 1 < y < 2 \\ 0 & , \text{ falls } y \notin [0, 2[ \end{cases} \end{aligned}$$

## Faltung (7)

Wir addieren drei zufällige Variablen  $X_1, X_2, X_3$ ,  $X_i \sim R(0, 1)$ ,

$$Y_3 = (X_1 + X_2) + X_3.$$

## Faltung (7)

Wir addieren drei zufällige Variablen  $X_1, X_2, X_3$ ,  $X_i \sim R(0, 1)$ ,

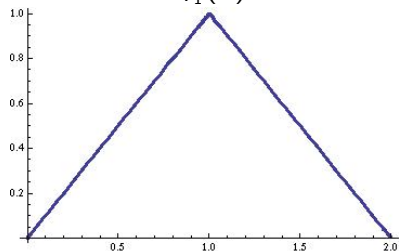
$$Y_3 = (X_1 + X_2) + X_3.$$

Für die Dichtefunktion der Zufallsgröße  $Y_3$  gilt dann nach der Faltungsformel:

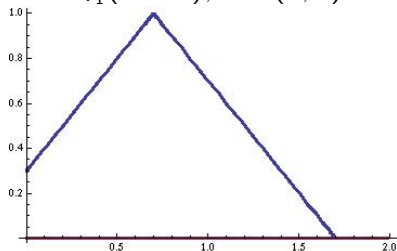
$$\begin{aligned} h_{Y_3}(z) &= h_{Y_1} * f_{X_3}(z) = \int_{-\infty}^{+\infty} h_{Y_1}(z-x) \cdot f_{X_3}(x) dx \\ &= \int_0^1 h_{Y_1}(z-x) \cdot f_{X_3}(x) dx = \int_0^1 h_{Y_1}(z-x) dx \end{aligned}$$

## Faltung (8)

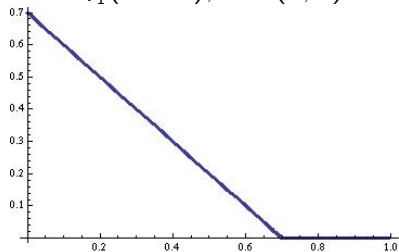
$$h_{Y_1}(x)$$



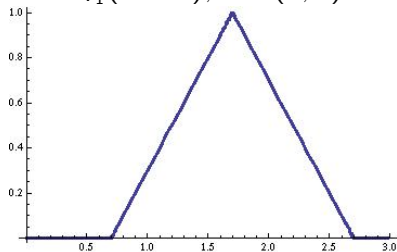
$$h_{Y_1}(z-x), z \in (1, 2)$$



$$h_{Y_1}(z-x), z \in (0, 1)$$



$$h_{Y_1}(z-x), z \in (2, 3)$$



# Faltung (9)

Das letzte Integral ist gleich

$$\textcircled{1} \text{ Fall 1: } 0 < z < 1 \quad = \int_0^z (z - x) dx = \frac{z^2}{2}$$

# Faltung (9)

Das letzte Integral ist gleich

$$\textcircled{1} \text{ Fall 1: } 0 < z < 1 \quad = \int_0^z (z - x) dx = \frac{z^2}{2}$$

$$\textcircled{2} \text{ Fall 2: } 1 < z < 2$$

# Faltung (9)

Das letzte Integral ist gleich

❶ Fall 1:  $0 < z < 1$        $= \int_0^z (z - x) dx = \frac{z^2}{2}$

❷ Fall 2:  $1 < z < 2$

$$\begin{aligned} & \int_0^{z-1} (2 - z + x) dx + \int_{z-1}^1 (z - x) dx \\ &= \int_{2-z}^1 t dt - \int_1^{z-1} t dt \\ &= \frac{1}{2}(1 - (2 - z)^2 - (z - 1)^2 + 1) \end{aligned}$$

# Faltung (9)

Das letzte Integral ist gleich

❶ Fall 1:  $0 < z < 1$        $= \int_0^z (z - x) dx = \frac{z^2}{2}$

❷ Fall 2:  $1 < z < 2$

$$\begin{aligned} & \int_0^{z-1} (2 - z + x) dx + \int_{z-1}^1 (z - x) dx \\ &= \int_{2-z}^1 t dt - \int_1^{z-1} t dt \\ &= \frac{1}{2}(1 - (2 - z)^2 - (z - 1)^2 + 1) \end{aligned}$$

❸ Fall 3:  $2 < z < 3$

$$\int_{z-2}^1 (x - (z - 2)) dx = \int_0^{3-z} t dt = \frac{(3 - z)^2}{2}$$



## Faltung (10)

Wegen  $0 \leq X_i < 1$  folgt dann:

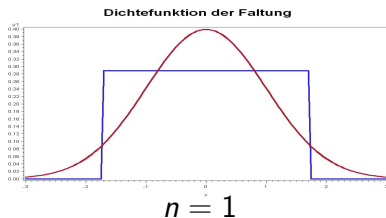
$$0 \leq (X_1 + X_2) + X_3 = Y_1 + X_3 = Y_3 < 3.$$

Für die Dichte der Summe der drei Zufallsgrößen  $X_1$ ,  $X_2$  und  $X_3$  gilt also:

$$h_{Y_3}(z) = \begin{cases} 0 & , \text{ falls } z \notin [0, 3[ \\ \frac{z^2}{2} & , \text{ falls } 0 \leq z \leq 1 \\ 1 - \frac{(z-1)^2}{2} - \frac{(2-z)^2}{2} & , \text{ falls } 1 < z \leq 2 \\ \frac{(3-z)^2}{2} & , \text{ falls } 2 < z < 3 \end{cases}$$

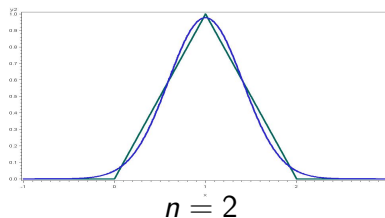
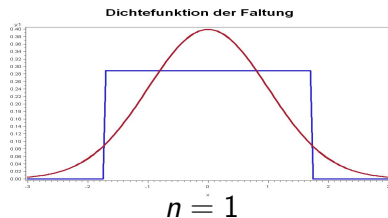
# Faltung (Veranschaulichung)

Seien  $X_i \sim R(0, 1)$ ,  $i = 1, 2, 3$ . Verteilungsfunktion von  $\sum_{i=1}^n X_i$ :



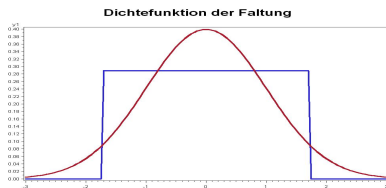
# Faltung (Veranschaulichung)

Seien  $X_i \sim R(0, 1)$ ,  $i = 1, 2, 3$ . Verteilungsfunktion von  $\sum_{i=1}^n X_i$ :

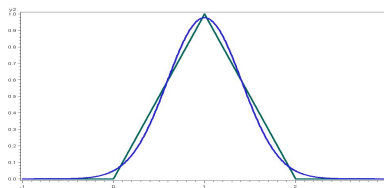


# Faltung (Veranschaulichung)

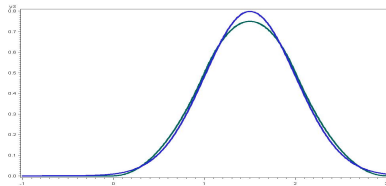
Seien  $X_i \sim R(0, 1)$ ,  $i = 1, 2, 3$ . Verteilungsfunktion von  $\sum_{i=1}^n X_i$ :



$n = 1$



$n = 2$



$n = 3$

# Faltung (10)

## Vermutung:

Die Summe unabhängiger Zufallsgrößen nähert sich bei wachsender Zahl der Zufallsgrößen einer Normalverteilung.

## Diese Vermutung ist richtig.

Sie gilt sogar (unter sehr allgemeinen Voraussetzungen, wie  $\text{var}(X_i) < \infty$ ) unabhängig davon, welche Verteilung diese Zufallsgrößen vorher hatten (Normal-, Gleich-, Exponentialverteilung oder diskret). Wir kommen später beim Zentralen Grenzwertsatz noch einmal darauf zurück.

# Satz der Totalen Wahrscheinlichkeit

für stetige Zufallsvariablen

Sei  $A$  ein Ereignis, das (unter Umständen) von den Werten der stetigen Zufallsvariable  $X$  mit Dichte  $f_X$  abhängt. Dann gilt

Satz der Totalen Wahrscheinlichkeit für stetige Zufallsvariablen

$$P(A) = \int P(A|X = t) f_X(t) dt,$$

wobei über den Definitionsbereich von  $f_X$  integriert wird.

**Beweis:** Sei  $F_X$  die Verteilungsfunktion von  $X$  und

$a_0 < a_{1,n} < \dots < a_{n,n}$  eine Zerlegung des Definitionsbereiches

( $a_0 = -\infty$ ,  $a_n = \infty$  ist erlaubt). Sei  $\forall i = 1, \dots, n-2$ :

$a_{i+1} - a_i = \Delta t$ . Dann gilt nach dem Satz der Totalen

Wahrscheinlichkeit

# Satz der Totalen Wahrscheinlichkeit

für stetige Zufallsvariablen, Beweis, 2

$$\begin{aligned}
 P(A) &= \sum_{i=0}^{n-1} P(A|X \in [a_{i,n}, a_{i+1,n}])P(X \in [a_{i,n}, a_{i+1,n}]) \\
 &= \sum_{i=0}^{n-1} P(A|X \in [a_{i,n}, a_{i+1,n}]) (F_X(a_{i+1,n}) - F_X(a_{i,n})) \\
 &= \sum_{i=0}^{n-1} P(A|X \in [a_{i,n}, a_{i+1,n}]) f_X(t_{i,n}^*) (a_{i+1,n} - a_{i,n}) \quad t_{i,n}^* \in [a_{i,n}, a_{i+1,n}] \\
 &= \lim_{n \rightarrow \infty, \Delta t \rightarrow 0} \sum_{i=0}^{n-1} P(A|X \in [a_{i,n}, a_{i+1,n}]) f_X(t_{i,n}^*) \Delta t \\
 &= \int P(A|X = t) f_X(t) dt.
 \end{aligned}$$

# Anwendung auf Faltung

Seien  $f_{X_1}$  und  $f_{X_2}$  Dichten von  $X_1$  bzw.  $X_2$  und sei das Ereignis  $A := \{X_1 + X_2 < t\}$ ,  $X_1$  und  $X_2$  unabhängig.

$$\begin{aligned} F_{X_1+X_2}(t) &= P(A) = \int P(A|X_2 = s) f_{X_2}(s) ds \\ &= \int P(X_1 + X_2 < t | X_2 = s) f_{X_2}(s) ds \\ &= \int P(X_1 < t - s) f_{X_2}(s) ds \quad X_1, X_2 \text{ unabhängig} \\ f_{X_1+X_2}(t) &= \frac{d}{dt} F_{X_1+X_2}(t) = \int \frac{d}{dt} F_{X_1}(t - s) f_{X_2}(s) ds \\ &= \int f_{X_1}(t - s) f_{X_2}(s) ds \end{aligned}$$



# Inhalt

- Begriffe
- Unabhängigkeit von Zufallsgrößen
- Transformationssatz für Zufallsvektoren
- Box-Müller Transformation
- Treffen einer Zielscheibe
- Faltung
- Transformationssatz für Erwartungswerte

Satz. Es sei  $\mathbf{X} = (X_1, \dots, X_p)^T$  ein zufälliger Vektor und  $g: \mathbb{R}^p \rightarrow \mathbb{R}$  eine Abbildung.

a)  $\mathbf{X}$  diskret mit Wkt.funktion (Zähldichte)  $f$ . Falls

$$\sum_{\mathbf{x}} |g(\mathbf{x})| f(\mathbf{x}) < \infty \quad \text{so:} \quad \mathbf{E}(g(\mathbf{X})) = \sum_{\mathbf{x}} g(\mathbf{x}) f(\mathbf{x}).$$

b)  $\mathbf{X}$  stetig mit Dichtefunktion  $f$ .

$$\mathbf{E}g(\mathbf{X}) = \int_{\mathbb{R}^p} g(x_1, \dots, x_p) \cdot f(x_1, \dots, x_p) dx_1 \dots dx_p,$$

falls das Integral  $\int |g(\mathbf{x})| f(\mathbf{x}) d\mathbf{x}$  existiert.

# Transformationssatz für Erwartungswerte

## Beispiel

Es sei  $\mathbf{X} = (X_1, X_2)^T$  ein stetiger zufälliger Vektor mit Dichtefunktion  $f(x_1, x_2)$ . Wir definieren die Funktion  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  durch  $g(\mathbf{X}) := X_1 + X_2$ . Dann gilt:

$$\mathbf{E}(X_1 + X_2) = \mathbf{E}X_1 + \mathbf{E}X_2$$

Allgemeiner,

$$\mathbf{E}(c \cdot X_1 + d \cdot X_2) = c \cdot \mathbf{E}X_1 + d \cdot \mathbf{E}X_2.$$

# Transformationssatz für Erwartungswerte

Beispiel (Fortsetzung)

$$\begin{aligned}\mathbf{E}g(\mathbf{X}) &= \mathbf{E}(X_1 + X_2) = \int_{\mathbb{R}^2} g(x_1, x_2) \cdot f(x_1, x_2) dx_1 dx_2 \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) \cdot f(x_1, x_2) dx_1 dx_2 \\&= \int_{\mathbb{R}^2} x_1 f(x_1, x_2) dx_1 dx_2 + \int_{\mathbb{R}^2} x_2 f(x_1, x_2) dx_1 dx_2 \\&= \int_{-\infty}^{\infty} x_1 \left( \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \right) dx_1 + \int_{-\infty}^{\infty} x_2 \left( \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 \right) dx_2 \\&= \int_{-\infty}^{+\infty} x_1 \cdot f_{X_1}(x_1) dx_1 + \int_{-\infty}^{+\infty} x_2 \cdot f_{X_2}(x_2) dx_2 = \mathbf{E}X_1 + \mathbf{E}X_2\end{aligned}$$

### Def. 57 (Korrelationskoeffizient)

Es seien  $X_1$  und  $X_2$  zwei zufällige Variablen, für die gilt:

$0 < \sigma_{X_1}, \sigma_{X_2} < \infty$ . Dann heißt der Quotient

$$\varrho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}}$$

Korrelationskoeffizient der Zufallsgrößen  $X_1$  und  $X_2$ .

Ist  $\text{cov}(X_1, X_2) = 0$  dann heißen die beiden Zufallsgrößen unkorreliert.

**Bem.:**  $X_1$  und  $X_2$  unabhängig  $\Rightarrow \text{cov}(X_1, X_2) = 0$ . Die Umkehrung der Aussage gilt i.a. nicht.

# Korrelation

## 2x2 Tafel

| Y \ X | 0(Sportler) | 1(Nichtsportler) | Summe    |
|-------|-------------|------------------|----------|
| 0(w)  | $p_{11}$    | $p_{12}$         | $p_{1.}$ |
| 1(m)  | $p_{21}$    | $p_{22}$         | $p_{2.}$ |
| Summe | $p_{.1}$    | $p_{.2}$         | 1        |

$$X \sim Bi(1, p_{.2}) \quad Y \sim Bi(1, p_{2.})$$

# Korrelation

## 2x2 Tafel

| Y \ X | 0(Sportler) | 1(Nichtsportler) | Summe    |
|-------|-------------|------------------|----------|
| 0(w)  | $p_{11}$    | $p_{12}$         | $p_{1.}$ |
| 1(m)  | $p_{21}$    | $p_{22}$         | $p_{2.}$ |
| Summe | $p_{.1}$    | $p_{.2}$         | 1        |

$$X \sim Bi(1, p_{.2}) \quad Y \sim Bi(1, p_{2.})$$

$$\mathbf{E}(X) = p_{.2} \quad \text{var}(X) = p_{.2}(1 - p_{.2}) = p_{.2}p_{.1}$$

$$\mathbf{E}(Y) = p_{2.} \quad \text{var}(Y) = p_{2.}(1 - p_{2.}) = p_{2.}p_{1.}$$

# Korrelation

2x2 Tafel

$$\text{cov}(X, Y) = \mathbf{E}(X \cdot Y) - \mathbf{E}(X)\mathbf{E}(Y) = p_{22} - p_{.2}p_{2.}$$

Korrelationskoeffizient:

$$\rho = \frac{p_{22} - p_{.2}p_{2.}}{\sqrt{p_{.2}p_{1.}p_{2.}p_{1.}}} = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{.2}p_{2.}p_{1.}p_{1.}}}$$



# Korrelation

2x2 Tafel

$$\text{cov}(X, Y) = \mathbf{E}(X \cdot Y) - \mathbf{E}(X)\mathbf{E}(Y) = p_{22} - p_{.2}p_{2.}$$

Korrelationskoeffizient:

$$\rho = \frac{p_{22} - p_{.2}p_{2.}}{\sqrt{p_{.2}p_{1.}p_{2.}p_{1.}}} = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{.2}p_{2.}p_{1.}p_{1.}}}$$

$$\begin{aligned} p_{22} - p_{.2}p_{2.} &= p_{22} - (p_{21} + p_{22})(p_{12} + p_{22}) \\ &= p_{22} - (p_{21}p_{12} + p_{22}p_{12} + p_{21}p_{22} + p_{22}^2) \\ &= p_{22}(1 - p_{12} - p_{21} - p_{22}) - p_{21}p_{12} \\ &= p_{22}p_{11} - p_{21}p_{12} \end{aligned}$$

# Korrelationskoeffizient

## Satz

Es seien  $X_1$  und  $X_2$  zwei Zufallsgrößen mit  $\sigma_{X_1}, \sigma_{X_2} > 0$ . Dann gilt für den Korrelationskoeffizienten:

$$-1 \leq \varrho(X_1, X_2) \leq 1.$$

# Korrelationskoeffizient

## Satz

Es seien  $X_1$  und  $X_2$  zwei Zufallsgrößen mit  $\sigma_{X_1}, \sigma_{X_2} > 0$ . Dann gilt für den Korrelationskoeffizienten:

$$-1 \leq \varrho(X_1, X_2) \leq 1.$$

**Beweis:** Wir definieren eine Funktion  $A$  wie folgt:

$$A(t, u) := \mathbf{E}[t \cdot (X_1 - \mathbf{E}X_1) + u \cdot (X_2 - \mathbf{E}X_2)]^2 \geq 0.$$

Es gilt für alle  $t, u \in \mathbb{R}$ :

# Korrelationskoeffizient

Beweis des Satzes (Fortsetzung,1)

$$\begin{aligned} 0 &\leq A(t, u) = \mathbf{E}[t \cdot (X_1 - \mathbf{E}X_1) + u \cdot (X_2 - \mathbf{E}X_2)]^2 \\ &= \mathbf{E}(t^2(X_1 - \mathbf{E}X_1)^2 + u^2(X_2 - \mathbf{E}X_2)^2 \\ &\quad + 2tu\mathbf{E}(X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2)) \\ &= t^2\mathbf{E}(X_1 - \mathbf{E}X_1)^2 + u^2\mathbf{E}(X_2 - \mathbf{E}X_2)^2 \\ &\quad + 2tu\mathbf{E}((X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2)) \\ &= t^2\text{Var } X_1 + 2 \cdot t \cdot u \cdot \text{cov}(X_1, X_2) + u^2\text{Var } X_2 \end{aligned}$$

# Korrelationskoeffizient

Beweis des Satzes (Fortsetzung, 2)

Wir setzen  $t := \sigma_{X_2}$ ,  $u := \sigma_{X_1}$  und dividieren durch  $\sigma_{X_1} \cdot \sigma_{X_2}$ :

$$\begin{aligned} \frac{A(\sigma_{X_2}, \sigma_{X_1})}{\sigma_{X_1} \cdot \sigma_{X_2}} &= \frac{\sigma_{X_2}^2 \sigma_{X_1}^2 + 2\sigma_{X_1} \sigma_{X_2} \text{cov}(X_1, X_2) + \sigma_{X_1}^2 \sigma_{X_2}^2}{\sigma_{X_1} \cdot \sigma_{X_2}} \\ &= \sigma_{X_1} \cdot \sigma_{X_2} + 2 \cdot \text{cov}(X_1, X_2) + \sigma_{X_1} \cdot \sigma_{X_2} \\ &= 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} + 2 \cdot \text{cov}(X_1, X_2) \geq 0 \end{aligned}$$

Also:

$$\underline{\sigma_{X_1} \cdot \sigma_{X_2} + \text{cov}(X_1, X_2) \geq 0.}$$

Andererseits gilt aber auch mit  $t := -\sigma_{X_2}$  und  $u := \sigma_{X_1}$  sowie derselben Herleitung wie oben:

# Korrelationskoeffizient

Beweis des Satzes (Fortsetzung,3)

$$\begin{aligned} \frac{A(-\sigma_{X_2}, \sigma_{X_1})}{\sigma_{X_1} \sigma_{X_2}} &= \frac{\sigma_{X_2}^2 \sigma_{X_1}^2 - 2\sigma_{X_1} \sigma_{X_2} \cdot \text{COV}(X_1, X_2) + \sigma_{X_1}^2 \sigma_{X_2}^2}{\sigma_{X_1} \cdot \sigma_{X_2}} \\ &= 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} - 2 \cdot \text{COV}(X_1, X_2) \geq 0 \end{aligned}$$

Also:

$$\underline{\sigma_{X_1} \cdot \sigma_{X_2} - \text{COV}(X_1, X_2) \geq 0.}$$

Beides zusammen ergibt

$$-\sigma_{X_1} \cdot \sigma_{X_2} \leq \text{COV}(X_1, X_2) \leq \sigma_{X_1} \cdot \sigma_{X_2}.$$

Wir stellen etwas um und erhalten:

$$-1 \leq \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \varrho(X_1, X_2) \leq 1.$$

# Korrelationskoeffizient

**Bem.:** Die Ungleichung kann auch direkt aus der Cauchy-Schwarz'schen Ungleichung hergeleitet werden.

## Satz

Es seien  $X_1$  und  $X_2$  zwei Zufallsgrößen, für die  $\sigma_{X_1}, \sigma_{X_2} > 0$  ist. Dann gilt  $|\varrho(X_1, X_2)| = 1$  genau dann, wenn es Zahlen  $a, b \in \mathbb{R}$  ( $a \neq 0$ ) gibt, so daß gilt:  $P(X_1 = a \cdot X_2 + b) = 1$ .

# Korrelationskoeffizient

Beweis des Satzes ( $\Leftarrow$ )

Seien  $a, b \in \mathbb{R}$  so, daß  $P(X_1 = a \cdot X_2 + b) = 1$ . Für Erwartungswert und Varianz von  $X_1$  gilt dann:

$$\mathbf{E}X_1 = \mathbf{E}(a \cdot X_2 + b) = a \cdot \mathbf{E}X_2 + b, \quad \sigma_{X_1}^2 = a^2 \cdot \sigma_{X_2}^2$$



# Korrelationskoeffizient

Beweis des Satzes ( $\Leftarrow$ )

Seien  $a, b \in \mathbb{R}$  so, daß  $P(X_1 = a \cdot X_2 + b) = 1$ . Für Erwartungswert und Varianz von  $X_1$  gilt dann:

$$\mathbf{E}X_1 = \mathbf{E}(a \cdot X_2 + b) = a \cdot \mathbf{E}X_2 + b, \quad \sigma_{X_1}^2 = a^2 \cdot \sigma_{X_2}^2$$

$$\varrho(X_1, X_2) = \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\mathbf{E}((X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2} \cdot \sigma_{X_2}}$$

# Korrelationskoeffizient

Beweis des Satzes ( $\Leftarrow$ )

Seien  $a, b \in \mathbb{R}$  so, daß  $P(X_1 = a \cdot X_2 + b) = 1$ . Für Erwartungswert und Varianz von  $X_1$  gilt dann:

$$\mathbf{E}X_1 = \mathbf{E}(a \cdot X_2 + b) = a \cdot \mathbf{E}X_2 + b, \quad \sigma_{X_1}^2 = a^2 \cdot \sigma_{X_2}^2$$

$$\begin{aligned} \varrho(X_1, X_2) &= \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\mathbf{E}((X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2} \cdot \sigma_{X_2}} \\ &= \frac{\mathbf{E}([(aX_2 + b) - (a\mathbf{E}X_2 + b)](X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2}^2} \\ &= \frac{a \cdot \mathbf{E}(X_2 - \mathbf{E}X_2)^2}{|a| \cdot \sigma_{X_2}^2} = \begin{cases} 1 & , \text{ falls } a > 0 \end{cases} \end{aligned}$$

# Korrelationskoeffizient

Beweis des Satzes ( $\Leftarrow$ )

Seien  $a, b \in \mathbb{R}$  so, daß  $P(X_1 = a \cdot X_2 + b) = 1$ . Für Erwartungswert und Varianz von  $X_1$  gilt dann:

$$\mathbf{E}X_1 = \mathbf{E}(a \cdot X_2 + b) = a \cdot \mathbf{E}X_2 + b, \quad \sigma_{X_1}^2 = a^2 \cdot \sigma_{X_2}^2$$

$$\begin{aligned} \varrho(X_1, X_2) &= \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\mathbf{E}((X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2} \cdot \sigma_{X_2}} \\ &= \frac{\mathbf{E}([(aX_2 + b) - (a\mathbf{E}X_2 + b)](X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2}^2} \\ &= \frac{a \cdot \mathbf{E}(X_2 - \mathbf{E}X_2)^2}{|a| \cdot \sigma_{X_2}^2} = \begin{cases} 1 & , \text{ falls } a > 0 \\ -1 & , \text{ falls } a < 0 \end{cases} \end{aligned}$$

# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ )

Es gelte  $|\varrho(X_1, X_2)| = 1$ . Dann gilt:

$$\begin{aligned}\varrho(X_1, X_2) &= \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))}{\sigma_{X_1} \cdot \sigma_{X_2}} \\ &= \mathbf{E}\left(\frac{X_1 - \mathbf{E}X_1}{\sigma_{X_1}} \cdot \frac{X_2 - \mathbf{E}X_2}{\sigma_{X_2}}\right) = \mathbf{E}(X_1^* \cdot X_2^*),\end{aligned}$$

wobei

$$X_1^* := \frac{X_1 - \mathbf{E}X_1}{\sigma_{X_1}}, \quad X_2^* := \frac{X_2 - \mathbf{E}X_2}{\sigma_{X_2}}.$$

Für die Varianz der Zufallsgrößen  $X_i^*$  ( $i = 1, 2$ ) gilt:

# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ ) (2)

$$\begin{aligned}
 \sigma_{X_i^*}^2 &= \mathbf{E} (X_i^* - \mathbf{E}X_i^*)^2 = \mathbf{E} (X_i^*)^2 - (\mathbf{E}X_i^*)^2 \\
 &= \mathbf{E} \left( \frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}} \right)^2 - \left( \mathbf{E} \left( \frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}} \right) \right)^2 \\
 &= \frac{1}{\sigma_{X_i}^2} \cdot \left( \mathbf{E}(X_i - \mathbf{E}X_i)^2 - (\mathbf{E}(X_i - \mathbf{E}X_i))^2 \right) \\
 &= \frac{1}{\sigma_{X_i}^2} \cdot \sigma_{X_i - \mathbf{E}X_i}^2 = \frac{1}{\sigma_{X_i}^2} \cdot \sigma_{X_i}^2 = 1
 \end{aligned}$$

# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ ), (3)

Offenbar gilt für die Erwartungswerte ( $i = 1, 2$ ):

$$\begin{aligned}\mathbf{E}X_i^* &= \mathbf{E}\left(\frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}}\right) = \frac{1}{\sigma_{X_i}} \cdot (\mathbf{E}X_i - \mathbf{E}(\mathbf{E}X_i)) \\ &= \frac{1}{\sigma_{X_i}} \cdot (\mathbf{E}X_i - \mathbf{E}X_i) = 0\end{aligned}$$

Daraus folgt:  $\varrho(X_1, X_2) = \mathbf{E}(X_1^* \cdot X_2^*)$ .

Wir unterscheiden zwei Fälle:

$$\underline{\varrho(X_1, X_2) = 1} \text{ und } \underline{\varrho(X_1, X_2) = -1}$$

# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ ), (4),  $\underline{\varrho(X_1, X_2) = 1}$

Wir untersuchen die Varianz der Zufallsgröße  $X_1^* - X_2^*$ :

$$\begin{aligned}\sigma_{X_1^* - X_2^*}^2 &= \mathbf{E}((X_1^* - X_2^*) - \mathbf{E}(X_1^* - X_2^*))^2 = \mathbf{E}(X_1^* - X_2^*)^2 \\ &= \mathbf{E}(X_1^*)^2 - 2 \cdot \mathbf{E}(X_1^* \cdot X_2^*) + \mathbf{E}(X_2^*)^2 \\ &= 1 - 2 \cdot \varrho(X_1, X_2) + 1 = 0\end{aligned}$$

Nun gilt aber  $\sigma_{X_1^* - X_2^*}^2 = 0$  genau dann, wenn es ein  $c \in \mathbb{R}$  gibt, so daß  $P(X_1^* - X_2^* = c) = 1$  ist. D.h.  $\mathbf{E}(X_1^* - X_2^*) = c$ .

Wegen  $\mathbf{E}X_1^* = \mathbf{E}X_2^* = 0$  ist  $c = 0$ , woraus folgt

$$P(X_1^* = X_2^*) = 1.$$

# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ ), (5),  $\varrho(X_1, X_2) = 1$

Dann gilt:

$$\begin{aligned}
 1 &= P(X_1^* = X_2^*) \\
 &= P\left(\frac{X_1 - \mathbf{E}X_1}{\sigma_{X_1}} = \frac{X_2 - \mathbf{E}X_2}{\sigma_{X_2}}\right) \\
 &= P\left(X_1 = \frac{\sigma_{X_1} \cdot X_2 - \sigma_{X_1} \cdot \mathbf{E}X_2}{\sigma_{X_2}} + \mathbf{E}X_1\right) \\
 &= P\left(X_1 = \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot X_2 - \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot \mathbf{E}X_2 + \mathbf{E}X_1\right)
 \end{aligned}$$

Wir definieren  $a := \frac{\sigma_{X_1}}{\sigma_{X_2}} > 0$  und  $b := \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot \mathbf{E}X_2 + \mathbf{E}X_1$ , und die Aussage ist für diesen Fall gezeigt.



# Korrelationskoeffizient

Beweis des Satzes ( $\implies$ ), (6),  $\varrho(X_1, X_2) = -1$

Sei  $\varrho(X_1, X_2) = -1$ : Hier untersucht man die Varianz der Zufallsgröße  $X_1^* + X_2^*$  und zeigt, dass sie ebenfalls gleich Null ist. Danach verläuft der Beweis völlig analog zum Fall  $\varrho(X_1, X_2) = 1$ .

## Def. 58 (standardisierte Zufallsgröße)

Eine Zufallsgröße, deren Erwartungswert gleich Null und deren Varianz gleich Eins sind, heißt standardisierte Zufallsgröße.

# Korrelationskoeffizient

Seien  $X, Y \sim (0, 1)$ ,  $X$  und  $Y$  unabhängig.

$$X^* = X$$

$$Y^* = \rho X + \sqrt{1 - \rho^2} Y$$

Offenbar

$$\text{var}X^* = \text{var}Y^* = 1$$

$$\text{cov}(X^*, Y^*) = \rho.$$

# Korrelationskoeffizient

## Zweidimensionale Normalverteilung

Seien  $X, Y \sim \mathcal{N}(0, 1)$ , unabhängig, d.h. die gemeinsame Dichte ist

$$f(x, y) = \phi(x) \cdot \phi(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

$$X^* = X$$

$$Y^* = \rho X + \sqrt{1 - \rho^2} Y$$

Wir suchen die gemeinsame Verteilung von  $(X^*, Y^*)$ .

Transformation:

$$g_1(x, y) = x$$

$$g_2(x, y) = \rho x + \sqrt{1 - \rho^2} y$$

# Korrelationskoeffizient

Zweidimensionale Normalverteilung

Inverse Transformation:

$$\begin{aligned}\psi_1(x^*, y^*) &= x^* \\ \psi_2(x^*, y^*) &= \frac{y^* - \rho x^*}{\sqrt{1 - \rho^2}}\end{aligned}$$

Jacobi-Determinante

$$\det(J) = \det \begin{pmatrix} 1 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} = \frac{1}{\sqrt{1-\rho^2}}$$

# Korrelationskoeffizient

Zweidimensionale Normalverteilung, Dichte

$$\begin{aligned}
 h(x^*, y^*) &= f(\psi_1(x^*, y^*), \psi_2(x^*, y^*)) \cdot |\det(J)| \\
 &= f\left(x^*, \frac{y^* - \rho x^*}{\sqrt{1 - \rho^2}}\right) \cdot \frac{1}{\sqrt{1 - \rho^2}} \\
 &= \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{1}{2}\left(x^{*2} + \left(\frac{y^* - \rho x^*}{\sqrt{1 - \rho^2}}\right)^2\right)} \\
 &= \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1 - \rho^2)}(x^{*2} - 2\rho x^* y^* + y^{*2})}
 \end{aligned}$$

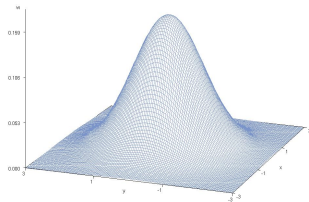
da der Exponent

$$x^{*2} + \left(\frac{y^* - \rho x^*}{\sqrt{1 - \rho^2}}\right)^2 = \frac{1}{\sqrt{1 - \rho^2}^2} \left( (1 - \rho^2)x^{*2} + (y^* - \rho x^*)^2 \right)$$

# Zweidimensionale Normalverteilung, Dichte

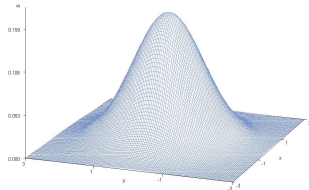
$$\rho = 0$$

Dichtefunktion der 2-dimensionalen Standard-Normalverteilung



$$\rho = 0.5$$

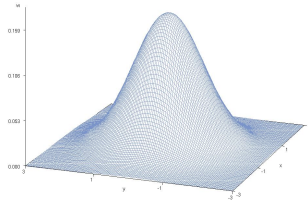
Dichtefunktion einer 2-dimensionalen Normalverteilung  
( $\rho=0.5$ )



# Zweidimensionale Normalverteilung, Dichte

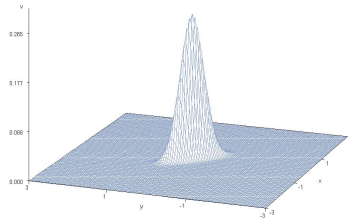
$$\rho = 0$$

Dichtefunktion der 2-dimensionalen Standard-Normalverteilung



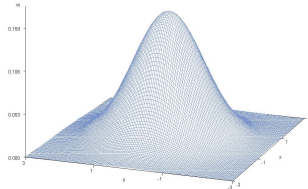
$$\rho = 0.8$$

Dichtefunktion einer 2-dimensionalen Normalverteilung  
(rho=0.8)



$$\rho = 0.5$$

Dichtefunktion einer 2-dimensionalen Normalverteilung  
(rho=0.5)



# Normalverteilung

$X$  und  $Y$  sind unabhängig gdw.  $X$  und  $Y$  sind unkorreliert

**Satz:** Es seien  $X, Y$  normalverteilt. Dann sind  $X$  und  $Y$  unabhängig gdw. sie unkorreliert sind.

**Beweis:** 1. Seien  $X$  und  $Y$  unabhängig. Dann sind sie unkorreliert (das gilt immer).

2. Seien  $X$  und  $Y$  unkorreliert,  $\rho = 0$ , und normalverteilt. Setzen wir in der letzten Formel für  $h(x^*, y^*)$ :  $\rho = 0$  ein, erhalten wir

$$h(x^*, y^*) = f_{X^*}(x^*)f_{Y^*}(y^*).$$



# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

Satz: Es sei  $X$  eine zufällige Variable. Dann gilt:

$$\min_{c \in \mathbb{R}} \mathbf{E}(X - c)^2 = \text{Var } X.$$

**Beweis:** Für alle reellen Zahlen  $c \in \mathbb{R}$  gilt:

$$\begin{aligned} \mathbf{E}(X - c)^2 &= \mathbf{E}(X - \mathbf{E}X + \mathbf{E}X - c)^2 \\ &= \mathbf{E}(X - \mathbf{E}X)^2 + 2\mathbf{E}((X - \mathbf{E}X)(\mathbf{E}X - c)) + (\mathbf{E}X - c)^2 \end{aligned}$$

Satz: Es sei  $X$  eine zufällige Variable. Dann gilt:

$$\min_{c \in \mathbb{R}} \mathbf{E}(X - c)^2 = \text{Var } X.$$

**Beweis:** Für alle reellen Zahlen  $c \in \mathbb{R}$  gilt:

$$\begin{aligned} \mathbf{E}(X - c)^2 &= \mathbf{E}(X - \mathbf{E}X + \mathbf{E}X - c)^2 \\ &= \mathbf{E}(X - \mathbf{E}X)^2 + 2\mathbf{E}((X - \mathbf{E}X)(\mathbf{E}X - c)) + (\mathbf{E}X - c)^2 \\ &= \mathbf{E}(X - \mathbf{E}X)^2 + 2(\mathbf{E}X - c) \underbrace{\mathbf{E}(X - \mathbf{E}X)}_{=0} + (\mathbf{E}X - c)^2 \\ &= \text{Var } X + (\mathbf{E}X - c)^2 \geq \text{Var } X \end{aligned}$$

Setzen wir  $c := \mathbf{E}X$  erhalten wir Gleichheit.



# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

### Satz (Ungleichung von JENSEN)

Sei  $X$  eine zufällige Variable mit  $\mathbf{E}X < \infty$  und  $g$  eine differenzierbare und konvexe Funktion. Dann gilt:

$$\mathbf{E}g(X) \geq g(\mathbf{E}X).$$

**Beweis:** Sei  $T(x)$  die Tangente an die Kurve der Funktion  $g$  im Punkt  $x_0$ ,

$$g(x) \geq T(x) = g(x_0) + \underbrace{g'(x_0)}_{\text{Anstieg der Kurve in } x_0} \cdot (x - x_0).$$

# Jensen-Ungleichung

Wir setzen  $x := X$  und  $x_0 := \mathbf{E}X$  und erhalten:

$$g(X) \geq g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X).$$

# Jensen-Ungleichung

Wir setzen  $x := X$  und  $x_0 := \mathbf{E}X$  und erhalten:

$$g(X) \geq g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X).$$

Daraus folgt:

$$\begin{aligned} \mathbf{E}g(X) &\geq \mathbf{E}(g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X)) \\ &= g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot \underbrace{\mathbf{E}(X - \mathbf{E}X)}_{=0} \\ &= g(\mathbf{E}X) \end{aligned}$$

# Jensen-Ungleichung

## Folgerung

Es sei  $g$  differenzierbar und konkav. Weiterhin sei  $X$  eine zufällige Variable. Dann gilt:

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$



# Jensen-Ungleichung

## Folgerung

Es sei  $g$  differenzierbar und konkav. Weiterhin sei  $X$  eine zufällige Variable. Dann gilt:

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

**Beweis:** Da die Funktion  $g$  nach Voraussetzung konkav ist, ist die Funktion  $(-g)$  konvex. Dann gilt nach der Jensen-Ungleichung:

$$\mathbf{E}((-g)(X)) \geq (-g)(\mathbf{E}X).$$

Daraus folgt die Behauptung. □

# Jensen-Ungleichung

## Beispiele

- ❶ Es sei  $g(x) = x^2$ . Dann gilt  $\mathbf{E}X^2 \geq (\mathbf{E}X)^2$ . Daraus folgt (die schon bekannte Aussage):

$$\text{Var } X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 \geq 0.$$

- ❷ Es sei  $g(x) = |x|$ . Dann gilt

$$\mathbf{E}|X| \geq |\mathbf{E}X|.$$

- ❸ Es sei  $g(x) = \ln x$ . Diese Funktion ist konkav. Also gilt

$$\mathbf{E}(\ln X) \leq \ln(\mathbf{E}X).$$

# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

## Satz (Ungleichung von MARKOV)

Sei  $X$  eine Zufallsgröße und  $c > 0$ . Dann gilt:

$$P(|X| \geq c) \leq \frac{\mathbf{E}|X|}{c}.$$

## Satz (Ungleichung von MARKOV)

Sei  $X$  eine Zufallsgröße und  $c > 0$ . Dann gilt:

$$P(|X| \geq c) \leq \frac{\mathbf{E}|X|}{c}.$$

**Beweis:** Wir definieren eine Zufallsgröße  $Y$  wie folgt

$$Y = \begin{cases} 0 & \text{falls } |X| < c \\ c & \text{falls } |X| \geq c \end{cases} \quad \text{bzw.} \quad Y : \begin{pmatrix} 0 & c \\ P(|X| < c) & P(|X| \geq c) \end{pmatrix}$$

Offenbar gilt  $0 \leq Y \leq |X|$  bzw.  $P(|X| - Y \geq 0) = 1$  und

$$\mathbf{E}(|X| - Y) \geq 0 \quad \text{bzw.} \quad \mathbf{E}|X| \geq \mathbf{E}Y$$

$$\mathbf{E}Y = 0 \cdot P(|X| < c) + c \cdot P(|X| \geq c)$$

$$= c \cdot P(|X| \geq c) \leq \mathbf{E}|X|$$

Division durch  $c$  liefert die Behauptung.

# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

## Satz (Ungleichung von TSCHEBYCHEV)

Es sei  $\varepsilon > 0$  und sei  $Y$  eine Zufallsgröße. Dann gilt:

$$P(|Y - \mathbf{E}Y| \geq \varepsilon) \leq \frac{\text{Var } Y}{\varepsilon^2}.$$

### Satz (Ungleichung von TSCHEBYCHEV)

Es sei  $\varepsilon > 0$  und sei  $Y$  eine Zufallsgröße. Dann gilt:

$$P(|Y - \mathbf{E}Y| \geq \varepsilon) \leq \frac{\text{Var } Y}{\varepsilon^2}.$$

**Beweis:** Wir verwenden die Markov-Ungleichung:

$$P(|X| \geq c) \leq \frac{\mathbf{E}|X|}{c}.$$

und setzen

$$X := (Y - \mathbf{E}Y)^{2i} \geq 0, \quad c := \varepsilon^{2i} \quad (i \in \mathbb{N}).$$



# Tschebychev-Ungleichung

Beweis, Fortsetzung

Da  $\varepsilon > 0$  gilt, ist die Voraussetzung der MARKOV- Ungleichung erfüllt. Wir erhalten:

$$P(|Y - \mathbf{E}Y| \geq \varepsilon) = P\left((Y - \mathbf{E}Y)^{2i} \geq \varepsilon^{2i}\right) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^{2i}}{\varepsilon^{2i}}.$$

Für  $i := 1$  ergibt sich:

$$P(|Y - \mathbf{E}Y| \geq \varepsilon) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^2}{\varepsilon^2} = \frac{\text{Var } Y}{\varepsilon^2}.$$

# Tschebyschev-Ungleichung

## 2. Formulierung

**Bem.:** Aus der TSCHEBYSCHEV-Ungleichung folgt:

$$P(|Y - EY| < \varepsilon) \geq 1 - \frac{\text{Var } Y}{\varepsilon^2}.$$

# Tschebyschev-Ungleichung

## 2. Formulierung

**Bem.:** Aus der TSCHEBYCHEV-Ungleichung folgt:

$$P(|Y - EY| < \varepsilon) \geq 1 - \frac{\text{Var } Y}{\varepsilon^2}.$$

Es sei  $X \sim (\mu, \sigma^2)$ , also  $EX = \mu$ ,  $\text{Var } X = \sigma^2$ .

Wir setzen  $\varepsilon := k \cdot \sigma$  ( $k \in \mathbb{N}$ ) und erhalten dann mit der Ungleichung von TSCHEBYCHEV:

$$P(|X - \mu| < k \cdot \sigma) \geq 1 - \frac{\sigma^2}{k^2 \cdot \sigma^2} = 1 - \frac{1}{k^2}.$$

# Tschebyschev-Ungleichung

Normalverteilung,  $k\sigma$ -Intervalle, Vergleich mit exakten Wahrscheinlichkeiten

| $k$ | exakt<br>$\Phi(k) - \Phi(-k)$ | Tschebychev-Ungleichung<br>$1 - \frac{1}{k^2}$ |
|-----|-------------------------------|------------------------------------------------|
| 1   | 0.68629                       | 0                                              |
| 2   | 0.9545                        | 0.75                                           |
| 3   | 0.9973                        | 0.89                                           |
| 4   | 0.99997                       | 0.93                                           |
| 5   | $\approx 1$                   | 0.96                                           |

# Tschebyshev-Ungleichung

**Bem.:** Die Tschebyshev-Ungleichung gilt für beliebig verteilte Zufallsvariablen, die Erwartungswert und Varianz besitzen, insbesondere liegt die Zufallsvariable  $X$  mit Wahrscheinlichkeit  $\geq 0.89$  im  $3\sigma$ -Intervall.

# Tschebyschev-Ungleichung, Beispiel

## Median der Zufallsvariablen $X$

Die Zahl  $med = med(X)$  heißt Median, falls

$$P(X \leq med) \geq \frac{1}{2} \quad \text{und} \quad P(X \geq med) \geq \frac{1}{2}$$

Sei  $P(X > 0) = 1$ . Aus der Markov-Ungleichung folgt:

$$\frac{1}{2} \leq P(X \geq med) \leq \frac{\mathbf{E}|X|}{med}, \quad \text{d.h. } med \leq 2 \cdot \mathbf{E}|X|$$

ÜA: Berechnen Sie Median und Erwartungswert von

$$X : \begin{pmatrix} 1 & 100 \\ 0.49 & 0.51 \end{pmatrix}$$

# Tschebyschev-Ungleichung

Die Tschebyschev-Ungleichung kann nicht verschärft werden

$$X : \begin{pmatrix} -\varepsilon & 0 & \varepsilon \\ \frac{1}{2\varepsilon^2} & 1 - \frac{1}{\varepsilon^2} & \frac{1}{2\varepsilon^2} \end{pmatrix}$$

$$\mathbf{E}X = 0, \quad \text{var}(X) = 1 \quad (\ddot{U}A)$$

Offenbar:

$$P(|X - \mathbf{E}X| \geq \varepsilon) = P(|X| \geq \varepsilon) = \frac{1}{\varepsilon^2}$$

rechte Seite bei der Tschebyschev-Ungleichung.

# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung



### Satz (Hoeffding-Ungleichung)

Seien  $Y_1, \dots, Y_n$  unabhängig und so dass  $\mathbf{E}Y_i = 0$  und  $a_i \leq Y_i \leq b_i$ .

Sei  $\epsilon > 0$ . Dann gilt  $\forall t > 0$ :

$$P\left(\sum_{i=1}^n Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(b_i - a_i)^2/8},$$

### Satz (Hoeffding-Ungleichung)

Seien  $Y_1, \dots, Y_n$  unabhängig und so dass  $\mathbf{E}Y_i = 0$  und  $a_i \leq Y_i \leq b_i$ .

Sei  $\epsilon > 0$ . Dann gilt  $\forall t > 0$ :

$$P\left(\sum_{i=1}^n Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(b_i - a_i)^2/8},$$

### Satz (Hoeffding-Ungleichung für Bernoulli Zufallsvariablen)

Seien  $X_1, \dots, X_n \sim \text{Bi}(1, p)$ . Dann gilt  $\forall \epsilon > 0$ :

$$P(|\bar{X}_n - p| \geq \epsilon) \leq 2e^{-2n\epsilon^2},$$

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

# Hoeffding-Ungleichung

## Beispiel

Seien  $X_1, \dots, X_n \sim \text{Bi}(1, p)$ , unabhängig

d.h. die Zufallsvariablen  $X_i$  sind Bernoulli verteilt:  $X_i = 1$  mit Wkt.  $p$ ,  
 $X_i = 0$  sonst.

$n = 100, \epsilon = 0.2$ .

Tschebyschev:

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{\text{var} \bar{X}_n}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} = 0.0625.$$

Hoeffding:

# Hoeffding-Ungleichung

## Beispiel

Seien  $X_1, \dots, X_n \sim \text{Bi}(1, p)$ , unabhängig

d.h. die Zufallsvariablen  $X_i$  sind Bernoulli verteilt:  $X_i = 1$  mit Wkt.  $p$ ,  
 $X_i = 0$  sonst.

$n = 100, \epsilon = 0.2$ .

Tschebyschev:

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{\text{var} \bar{X}_n}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} = 0.0625.$$

Hoeffding:

$$P(|\bar{X}_n - p| \geq \epsilon) \leq 2e^{-2 \cdot 100 \cdot 0.2^2} = 0.00067.$$

# Hoeffding-Ungleichung

Es geht sogar noch besser:

Zentraler Grenzwertsatz (siehe Kapitel Grenzwertsätze )

$$P(|\bar{X}_n - p| \geq \epsilon) = P\left(\underbrace{\left| \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \right|}_{\sim \mathcal{N}(0,1)} \geq \frac{n\epsilon}{\sqrt{np(1-p)}}\right)$$

approx.

$$\begin{aligned} &\approx \left(1 - \Phi\left(\frac{n\epsilon}{\sqrt{np(1-p)}}\right)\right) + \Phi\left(-\frac{n\epsilon}{\sqrt{np(1-p)}}\right) \\ &= 2\Phi\left(-\frac{n\epsilon}{\sqrt{np(1-p)}}\right) \\ &\leq 2\Phi\left(\frac{-n\epsilon}{\sqrt{n\frac{1}{4}}}\right) = 2\Phi(-2\epsilon\sqrt{n}) = 2\Phi(-4) \approx 10^{-4}. \end{aligned}$$

# Hoeffding-Ungleichung: $(1 - \alpha)$ Konfidenzintervall

Sei  $\alpha > 0$  und  $\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}$ .

Hoeffding:

$$P(|\bar{X}_n - p| \geq \epsilon_n) \leq$$

# Hoeffding-Ungleichung: $(1 - \alpha)$ Konfidenzintervall

Sei  $\alpha > 0$  und  $\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}$ .

Hoeffding:

$$P(|\bar{X}_n - p| \geq \epsilon_n) \leq 2e^{-2n\epsilon_n^2} = \alpha.$$

Sei  $C = (\bar{X}_n - \epsilon_n, \bar{X}_n + \epsilon_n)$ .

$$P(p \notin C) = P(|\bar{X}_n - p| \geq \epsilon_n) \leq \alpha$$

$$P(p \in C) \geq 1 - \alpha$$

D.h. das zufällige Intervall  $C$  überdeckt den wahren Parameter  $p$  mit Wahrscheinlichkeit  $\geq 1 - \alpha$ .

# Schätzung von Binomialwahrscheinlichkeiten (1)

Vorgabe:  $\epsilon, \alpha$ .

Gesucht: notwendiger Stichprobenumfang um für die Schätzung

$$\hat{p} = \bar{X}$$

$$P(|\hat{p} - p| \geq \epsilon) < \alpha$$

zu sichern.



# Schätzung von Binomialwahrscheinlichkeiten (1)

Vorgabe:  $\epsilon, \alpha$ .

Gesucht: notwendiger Stichprobenumfang um für die Schätzung

$$\hat{p} = \bar{X}$$

$$P(|\hat{p} - p| \geq \epsilon) < \alpha$$

zu sichern.

Hoeffding: Es genügt:

$$2 \cdot e^{-2n\epsilon^2} < \alpha$$

also

$$n > \frac{-\ln \alpha / 2}{2\epsilon^2} = \frac{\ln(2/\alpha)}{2\epsilon^2}.$$

# Schätzung von Binomialwahrscheinlichkeiten (2)

## Zentraler Grenzwertsatz

$$\begin{aligned}P(|\hat{p} - p| \geq \epsilon) &= P\left(\frac{n|\hat{p} - p|}{\sqrt{np(1-p)}} \geq \frac{n\epsilon}{\sqrt{np(1-p)}}\right) \\&\approx 2\Phi\left(-\frac{n\epsilon}{\sqrt{np(1-p)}}\right) < \alpha \\-\frac{n\epsilon}{\sqrt{np(1-p)}} &< \Phi^{-1}\left(\frac{\alpha}{2}\right) \\\sqrt{n} &> \frac{-\Phi^{-1}\left(\frac{\alpha}{2}\right)}{\epsilon} \sqrt{p(1-p)} \\n &> \frac{\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)^2}{4\epsilon^2}\end{aligned}$$

# Schätzung von Binomialwahrscheinlichkeiten (3)

Vergleich Hoeffding - ZGWS. Vorgabe:  $P(|\hat{p} - p| > 0.01) < \alpha$

|          | Notwendige Stichprobenumfänge                           |                                              |
|----------|---------------------------------------------------------|----------------------------------------------|
|          | ZGWS                                                    | Hoeffding                                    |
| $\alpha$ | $\frac{1}{4\epsilon^2} \Phi^{-1}(1 - \frac{\alpha}{2})$ | $\frac{1}{2\epsilon^2} \ln \frac{2}{\alpha}$ |
| 0.1      | 6765                                                    | 15000                                        |
| 0.05     | 9640                                                    | 18450                                        |
| 0.01     | 16641                                                   | 26490                                        |
| 0.001    | 27225                                                   | 38000                                        |

# Hoeffding-Ungleichung

Beweis (1)

Sei  $t > 0$ . Aus der Markov-Ungleichung folgt:

$$\begin{aligned} P\left(\sum_{i=1}^n Y_i \geq \epsilon\right) &= P\left(t \sum_{i=1}^n Y_i \geq t\epsilon\right) = P\left(e^{t \sum_{i=1}^n Y_i} \geq e^{t\epsilon}\right) \\ &\leq e^{-t\epsilon} \mathbf{E}\left(e^{t \sum_{i=1}^n Y_i}\right) = e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}\left(e^{t Y_i}\right). \end{aligned}$$

Da  $a_i \leq Y_i \leq b_i$  lässt sich  $Y_i$  als konvexe Kombination von  $a_i$  und  $b_i$  schreiben,

$$Y_i = \alpha b_i + (1 - \alpha) a_i,$$

wobei  $\alpha = \frac{Y_i - a_i}{b_i - a_i}$ .

# Hoeffding-Ungleichung

Beweis (2)

NR.: Für konvexe Funktionen  $f(x)$ ,  $x \in (a, b)$  gilt:

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a) = \alpha f(b) + (1 - \alpha)f(a)$$

(Die Kurve  $f$  liegt unterhalb der Sekante,  $\alpha = \frac{x-a}{b-a}$ .).

# Hoeffding-Ungleichung

Beweis (2)

NR.: Für konvexe Funktionen  $f(x)$ ,  $x \in (a, b)$  gilt:

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a) = \alpha f(b) + (1 - \alpha)f(a)$$

(Die Kurve  $f$  liegt unterhalb der Sekante,  $\alpha = \frac{x-a}{b-a}$ ). Da die Exponentialfunktion konvex ist:

$$\begin{aligned} e^{tY_i} &\leq \alpha e^{tb_i} + (1 - \alpha)e^{ta_i} \\ &= \frac{Y_i - a_i}{b_i - a_i} e^{tb_i} + \frac{b_i - Y_i}{b_i - a_i} e^{ta_i} \\ \mathbf{E}(e^{tY_i}) &\leq \frac{-a_i}{b_i - a_i} e^{tb_i} + \frac{b_i}{b_i - a_i} e^{ta_i} = e^{g(u)} \end{aligned}$$

# Hoeffding-Ungleichung

Beweis (3)

wegen  $\mathbf{E} Y_i = 0$ . Dabei ist

$$u = t(b_i - a_i)$$

$$g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$$

$$\gamma = \frac{-a_i}{b_i - a_i}, \quad \gamma \in (0, 1) \text{ da } a_i < 0 < b_i$$

$$g'(u) = -\gamma + \frac{\gamma e^u}{1 - \gamma + \gamma e^u}$$

$$g''(u) = \frac{\gamma e^u (1 - \gamma)}{(1 - \gamma + \gamma e^u)^2} =: \frac{xy}{(x + y)^2}$$

$$g(0) = g'(0) = 0, \quad g''(u) \leq \frac{1}{4} \quad \forall u > 0.$$

wobei  $x = \gamma e^u$ ,  $y = 1 - \gamma$ .

# Hoeffding-Ungleichung

Beweis (4)

Die Aussage  $0 \leq g''(u) = \frac{xy}{(x+y)^2} \leq \frac{1}{4}$  folgt aus

$$0 \leq (x - y)^2 \quad \text{gdw.}$$

$$4xy \leq x^2 + 2xy + y^2 = (x + y)^2$$



# Hoeffding-Ungleichung

Beweis (4)

Die Aussage  $0 \leq g''(u) = \frac{xy}{(x+y)^2} \leq \frac{1}{4}$  folgt aus

$$0 \leq (x - y)^2 \quad \text{gdw.}$$

$$4xy \leq x^2 + 2xy + y^2 = (x + y)^2$$

Satz von Taylor: es ex. ein  $\xi \in (0, u)$ :

$$\begin{aligned} g(u) &= g(0) + ug'(0) + \frac{u^2}{2}g''(\xi) \\ &= \frac{u^2}{2}g''(\xi) \leq \frac{u^2}{8} = \frac{t^2(b_i - a_i)^2}{8} \end{aligned}$$

# Hoeffding-Ungleichung

Beweis (5)

Daraus folgt:

$$\mathbf{E}(e^{tY_i}) \leq e^{g(u)} \leq e^{t^2(b_i - a_i)^2/8}.$$

Damit:

$$P\left(\sum_{i=1}^n Y_i \geq \epsilon\right) = e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}(e^{tY_i}) \leq e^{-t\epsilon} \prod_{i=1}^n e^{t^2(b_i - a_i)^2/8}.$$

# Hoeffding-Ungleichung für Bernoulli

Beweis (1):

Sei  $Y_i = \frac{1}{n}(X_i - p)$ . Dann gilt  $\mathbf{E} Y_i = 0$  und  $a \leq Y_i \leq b$ , wobei  $a = -p/n$  und  $b = (1 - p)/n$ .

Also  $(b - a)^2 = 1/n^2$ . Aus der Hoeffding-Ungleichung folgt:

$$P(\bar{X}_n - p \geq \epsilon) = P\left(\sum_{i=1}^n Y_i \geq \epsilon\right) \leq e^{-t\epsilon} e^{t^2/(8n)},$$

für jedes  $t > 0$ . Setze  $t = 4n\epsilon$ :

$$P(\bar{X}_n - p > \epsilon) \leq e^{-2n\epsilon^2}.$$

# Hoeffding-Ungleichung für Bernoulli

Beweis (2)

Analog:

$$P(\overline{X}_n - p \leq -\epsilon) \leq e^{-2n\epsilon^2}.$$

Beides zusammen:

$$P(|\overline{X}_n - p| \geq \epsilon) \leq 2e^{-2n\epsilon^2}.$$

# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

Während die Hoeffding-Ungleichung Abschätzungen für Wahrscheinlichkeiten von absoluten Fehlern angibt, gibt die Chernov-Ungleichung solche für relative Fehler an.

### Satz (Chernov-Ungleichung)

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $X_i \sim Bi(1, p_i)$ , d.h.

$P(X_i = 1) = \mathbf{E}(X_i) = p_i$ . Dann gilt  $\forall \delta \in (0, 1)$ :

$$P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) < \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^\mu$$
$$P\left(-\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) < \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^\mu$$

wobei  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\bar{p}_n = \frac{1}{n} \sum_{i=1}^n p_i$  und  $\mu = n\bar{p} = \sum_{i=1}^n p_i$ .

# Chernov Ungleichung, Beweis

**Beweis:** Mit Hilfe der Markov-Ungleichung erhalten wir die folgende Ungleichung  $\forall t > 0$

$$P\left(\sum_{i=1}^n X_i \geq \epsilon\right) = P\left(e^t \sum_{i=1}^n X_i \geq e^{t\epsilon}\right) \leq e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}(e^{tX_i}).$$

Daraus folgt  $\forall t > 0$

$$P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) = P(\bar{X}_n > (1 + \delta)\bar{p}) = P\left(\sum_{i=1}^n X_i > \underbrace{(1 + \delta)\mu}_{=\epsilon}\right)$$

# Chernov Ungleichung, Beweis

**Beweis:** Mit Hilfe der Markov-Ungleichung erhalten wir die folgende Ungleichung  $\forall t > 0$

$$P\left(\sum_{i=1}^n X_i \geq \epsilon\right) = P\left(e^{t \sum_{i=1}^n X_i} \geq e^{t\epsilon}\right) \leq e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}(e^{tX_i}).$$

Daraus folgt  $\forall t > 0$

$$\begin{aligned} P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) &= P(\bar{X}_n > (1 + \delta)\bar{p}) = P\left(\sum_{i=1}^n X_i > \underbrace{(1 + \delta)\mu}_{=\epsilon}\right) \\ &\leq \frac{\prod_{i=1}^n \mathbf{E}(e^{tX_i})}{e^{t(1+\delta)\mu}} \end{aligned}$$



# Chernov Ungleichung, Beweis

**Beweis:** Mit Hilfe der Markov-Ungleichung erhalten wir die folgende Ungleichung  $\forall t > 0$

$$P\left(\sum_{i=1}^n X_i \geq \epsilon\right) = P\left(e^{t \sum_{i=1}^n X_i} \geq e^{t\epsilon}\right) \leq e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}(e^{tX_i}).$$

Daraus folgt  $\forall t > 0$

$$\begin{aligned} P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) &= P(\bar{X}_n > (1 + \delta)\bar{p}) = P\left(\sum_{i=1}^n X_i > \underbrace{(1 + \delta)\mu}_{=\epsilon}\right) \\ &\leq \frac{\prod_{i=1}^n \mathbf{E}(e^{tX_i})}{e^{t(1+\delta)\mu}} \leq \frac{\prod_{i=1}^n (p_i e^t + (1 - p_i))}{e^{t(1+\delta)\mu}} \end{aligned}$$

# Chernov Ungleichung, Beweis

**Beweis:** Mit Hilfe der Markov-Ungleichung erhalten wir die folgende Ungleichung  $\forall t > 0$

$$P\left(\sum_{i=1}^n X_i \geq \epsilon\right) = P\left(e^{t \sum_{i=1}^n X_i} \geq e^{t\epsilon}\right) \leq e^{-t\epsilon} \prod_{i=1}^n \mathbf{E}(e^{tX_i}).$$

Daraus folgt  $\forall t > 0$

$$\begin{aligned} P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) &= P(\bar{X}_n > (1 + \delta)\bar{p}) = P\left(\sum_{i=1}^n X_i > \underbrace{(1 + \delta)\mu}_{=\epsilon}\right) \\ &\leq \frac{\prod_{i=1}^n \mathbf{E}(e^{tX_i})}{e^{t(1+\delta)\mu}} \leq \frac{\prod_{i=1}^n (p_i e^t + (1 - p_i))}{e^{t(1+\delta)\mu}} \end{aligned}$$

Schreiben wir  $p_i e^t + (1 - p_i) = \underbrace{p_i(e^t - 1)}_{=:x} + 1$  und bemerken, dass

$1 + x < e^x \quad \forall x > 0$  so erhalten wir

$$\begin{aligned}
 P\left(\frac{\overline{X}_n - \bar{p}}{\bar{p}} > \delta\right) &< \frac{\prod_{i=1}^n \exp(p_i(e^t - 1))}{e^{t(1+\delta)\mu}} \\
 &= \frac{\exp((e^t - 1) \sum_{i=1}^n p_i)}{e^{t(1+\delta)\mu}} = \frac{\exp((e^t - 1)\mu)}{e^{t(1+\delta)\mu}}.
 \end{aligned}$$

Setzen wir nun  $t = \log(1 + \delta)$  ( $t > 0$  für  $\delta > 0$ ) erhalten wir für den letzten Bruch

$$\frac{\exp((e^t - 1)\mu)}{e^{t(1+\delta)\mu}} = \left( \frac{e^{1+\delta} - 1}{(1 + \delta)^{1+\delta}} \right)^\mu = \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu$$

Damit ist die erste Ungleichung bewiesen.

Analog (beachte  $e^{-x} > 1 - x \quad \forall x \neq 0$ ) folgt  $\forall t > 0$

$$\begin{aligned}
 P\left(-\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) &= P\left(\sum_{i=1}^n X_i < (1 - \delta)\mu\right) \\
 &= P(e^{-t \sum_{i=1}^n X_i} > e^{-t(1-\delta)\mu}) \leq \frac{\prod_{i=1}^n \mathbf{E}(e^{-tX_i})}{e^{-t(1-\delta)\mu}} \\
 &\leq \frac{\prod_{i=1}^n (p_i e^{-t} + (1 - p_i))}{e^{t(1-\delta)\mu}} \\
 &= \frac{\prod_{i=1}^n (1 - p_i(1 - e^{-t}))}{e^{-t(1-\delta)\mu}} < \frac{\exp((1 - e^{-t})\mu)}{e^{-t(1-\delta)\mu}}
 \end{aligned}$$

Setzen wir diesmal  $t = -\log(1 - \delta)$  ( $t > 0$  für  $\delta > 0$ ) erhalten wir für den letzten Bruch

$$\frac{\exp((1 - e^{-t})\mu)}{e^{-t(1-\delta)\mu}} = \frac{e^{-\delta\mu}}{(1 - \delta)^{(1-\delta)\mu}}$$

Damit ist auch die zweite Ungleichung bewiesen.

### Satz (Chernov-Ungleichung, vereinfacht)

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $X_i \sim Bi(p_i)$ , d.h.

$P(X_i = 1) = \mathbf{E}(X_i) = p_i$ . Dann gilt  $\forall \delta \in (0, 1)$ :

$$P\left(\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) \leq e^{-\mu \frac{\delta^2}{2+\delta}} \leq e^{-\mu \frac{\delta^2}{3}}$$

$$P\left(-\frac{\bar{X}_n - \bar{p}}{\bar{p}} > \delta\right) \leq e^{-\mu \frac{\delta^2}{2}}$$

wobei  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\bar{p}_n = \frac{1}{n} \sum_{i=1}^n p_i$  und  $\mu = \sum_{i=1}^n p_i$ .

# Beweis:

Logarithmieren die rechte Seite der Chernov-Ungleichung und wenden die Ungleichungen  $\log(1+x) \geq \frac{x}{1+x/2}$  (auf die erste Chernov-Ungleichung) und  $\log(1-x) \geq -x - x^2/2$  (auf die zweite Chernov-Ungleichung) (Beweis unter Verwendung der Reihenentwicklung) an und erhalten ( $0 < \delta < 1$ )

$$\begin{aligned}\log\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu &= \mu(\delta - (1+\delta)\log(1+\delta)) \leq -\frac{\delta^2}{2+\delta}\mu \\ \log\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu &= \mu(-\delta - (1-\delta)\log(1-\delta)) \leq -\frac{\delta^2}{2}\mu \leq e^{-\mu\frac{\delta^2}{3}}\end{aligned}$$

# Inhalt

- Varianz-Ungleichung
- Jensen-Ungleichung
- Markov-Ungleichung
- Tschebychev-Ungleichung
- Hoeffding-Ungleichung
- Chernov-Ungleichung
- Mill Ungleichung

Satz (Mill-Ungleichung). Sei  $Z \sim \mathcal{N}(0, 1)$ . Dann

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} = \frac{2\phi(t)}{t}.$$



Satz (Mill-Ungleichung). Sei  $Z \sim \mathcal{N}(0, 1)$ . Dann

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} = \frac{2\phi(t)}{t}.$$

**Beweis:** Es gilt

$$P(|Z| > t) = 2P(Z > t) = 2 \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Satz (Mill-Ungleichung). Sei  $Z \sim \mathcal{N}(0, 1)$ . Dann

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} = \frac{2\phi(t)}{t}.$$

**Beweis:** Es gilt

$$\begin{aligned} P(|Z| > t) &= 2P(Z > t) = 2 \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \sqrt{\frac{2}{\pi}} \int_t^\infty \left(-\frac{1}{x}\right) (-xe^{-\frac{x^2}{2}}) dx \end{aligned}$$

Satz (Mill-Ungleichung). Sei  $Z \sim \mathcal{N}(0, 1)$ . Dann

$$P(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} = \frac{2\phi(t)}{t}.$$

**Beweis:** Es gilt

$$\begin{aligned} P(|Z| > t) &= 2P(Z > t) = 2 \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \sqrt{\frac{2}{\pi}} \int_t^\infty \left(-\frac{1}{x}\right) (-xe^{-\frac{x^2}{2}}) dx \\ &= \sqrt{\frac{2}{\pi}} \left(-\frac{1}{x}\right) e^{-\frac{x^2}{2}} \Big|_t^\infty - \underbrace{\int_t^\infty \frac{1}{x^2} e^{-\frac{x^2}{2}} dx}_{\geq 0} \\ &\leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} \end{aligned}$$

# Inhalt

- Das Gesetz der Großen Zahlen
- Der Satz von GLIVENKO–CANTELLI
- Konvergenz von Folgen zufälliger Variablen
- Der zentrale Grenzwertsatz

# Das Gesetz der Großen Zahlen: Motivation

Der Erwartungswert einer zufälligen Variablen  $X$  ist in der Praxis meist nicht bekannt. Um ihn zu schätzen, sammelt man Beobachtungen  $X_1, X_2, \dots, X_n$ , und bildet dann das arithmetische Mittel:

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

Beachten: die Beobachtungen  $X_1, \dots, X_n$  müssen unabhängig oder wenigstens unkorreliert sein.

# Das Gesetz der Großen Zahlen

## Satz (Schwaches Gesetz der Großen Zahlen)

Es seien  $X_1, \dots, X_n$  unkorrelierte zufällige Variablen mit  $\mu := \mathbf{E}X_i$  und  $\sigma^2 := \text{Var } X_i < \infty$  (für alle  $i = 1, \dots, n$ ). Dann gilt für alle  $\varepsilon > 0$ :

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

# Das Gesetz der Großen Zahlen

Beweis

**Beweis:** Da die Zufallsgrößen  $X_1, \dots, X_n$  unkorreliert sind, gilt

$$\mathbf{E}\bar{X} = \mu, \quad \text{Var } \bar{X} = \frac{\sigma^2}{n}$$

Mittels der TSCHEBYCHEV-Ungleichung erhalten wir:

$$\begin{aligned} P(|\bar{X}_n - \mu| > \varepsilon) &= P(|\bar{X} - \mathbf{E}\bar{X}| > \varepsilon) \\ &\leq \frac{\text{Var } \bar{X}}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$



# Das Gesetz der Großen Zahlen

**Bem.:** Aus dem Beweis erkennen wir, daß die Voraussetzungen etwas abgeschwächt werden können, anstelle  $\text{Var } X_i = \sigma^2$  genügt die Forderung

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i = 0.$$



# Das Gesetz der Großen Zahlen

**Bem.:** Aus dem Beweis erkennen wir, daß die Voraussetzungen etwas abgeschwächt werden können, anstelle  $\text{Var } X_i = \sigma^2$  genügt die Forderung

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i = 0.$$

**Bem.:** Die Voraussetzung der endlichen Varianz kann auch fallen gelassen werden. Dann können wir aber zum Beweis nicht mehr die Tschebyschev-Ungleichung anwenden. Der Beweis geht dann über charakteristische Funktionen. Und wir brauchen die Unabhängigkeit.

**Bem.:** Auf die Unkorreliertheit kann nicht verzichtet werden. Sei etwa  $EX_i = 0$ ,  $\text{var}X_i = 1$ ,  $\text{cov}(X_i, X_j) = \rho$  ( $i \neq j$ ). Dann

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \left( \sum_{i=1}^n \text{Var } X_i + \sum_{i \neq j} \text{cov}(X_i, X_j) \right) = \lim_{n \rightarrow \infty} \frac{n + n(n-1)\rho}{n^2} = \rho.$$

# Das Gesetz der Großen Zahlen

## Stochastischer Grenzwert

Wenn  $\lim_{n \rightarrow \infty} P(|Y_n - Y_0| > \varepsilon) = 0 \quad \forall \varepsilon > 0$

dann heißt  $Y_0$  stochastischer Grenzwert der Folge  $\{Y_n\}$  und man schreibt  $p - \lim Y_n = Y_0$  oder  $Y_n \rightarrow_p Y_0$ .

# Das Gesetz der Großen Zahlen

## Beispiel 1

Es seien  $X_i \sim Bi(1, p)$

$$X_i : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

$$\mu := \mathbf{E}X = \mathbf{E}X_i = p \quad \sigma^2 = p \cdot (1-p) < \infty$$

Nach dem Schwachen Gesetz der Großen Zahlen folgt:

$$P \left( \left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0.$$

# Das Gesetz der Großen Zahlen

## Beispiel 2

Es sei  $A$  ein Ereignis,  $p = P(A)$  sei unbekannt.

Zur Schätzung von  $p$  führen wir eine Reihe von unabhängigen Experimenten durch, bei denen  $A$  und  $\bar{A}$  die einzig möglichen Ausgänge seien.

$n$ : # der Experimente, die durchgeführt werden.

$n(A)$ : # Auftretens des Ereignisses  $A$ .

$$\hat{p}_n = \frac{n(A)}{n}$$

die relative Häufigkeit des Ereignisses  $A$ .

Frage:  $\hat{p}_n \xrightarrow{n \rightarrow \infty} p?$

# Das Gesetz der Großen Zahlen

Beispiel 2, Fortsetzung

Dazu definieren wir Zufallsgrößen  $X_i$  ( $i = 1, \dots, n$ ),

$$X_i := \begin{cases} 1 & , A \text{ im } i\text{-ten Experiment eintritt} \\ 0 & , A \text{ im } i\text{-ten Experiment nicht eintritt} \end{cases}$$

Dann gilt für alle  $i = 1, \dots, n$ :

$$X_i \sim Bi(1, p)$$

und  $P(X_i = 1) = p$  sowie  $P(X_i = 0) = 1 - p$ .

$$\mu = \mathbf{E}X_i = p \quad \sigma^2 = \text{Var } X_i = p \cdot (1 - p)$$

$$\bar{X} := \frac{1}{n} \cdot \sum_{i=1}^n X_i = \frac{1}{n} \cdot n(A) = \hat{p}_n$$

# Das Gesetz der Großen Zahlen

## Beispiel 2, Fortsetzung

Wenden das Schwache Gesetz der Großen Zahlen an und erhalten:

$$\begin{aligned}\lim_{n \rightarrow \infty} P(|\hat{p}_n - p| > \varepsilon) &= \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) \\ &= 0, \quad \forall \varepsilon > 0\end{aligned}$$

Folglich gilt:  $\hat{p}_n \xrightarrow{n \rightarrow \infty} p$  oder, genauer,  $\hat{p}_n \rightarrow_p p$

**Bem.:** Schätzungen  $\hat{p}_n$ , die gegen den zu schätzenden Parameter konvergieren heißen (schwach) konsistent.

# Starkes Gesetz der Großen Zahlen

## Satz (Gesetz der Großen Zahlen), ohne Beweis

Seien die Zufallsvariablen  $X_1, \dots, X_n$  identisch verteilt und unabhängig,  $\mathbf{E}|X_i| < \infty$ ,  $\mathbf{E}X_i = \mu$ . Dann gilt

$$P(\omega : \lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1.$$

**Bem.:** Schwaches Gesetz der Großen Zahlen: Seien die  $X_1, \dots, X_n$  identisch verteilt,  $\mathbf{E}X_i = \mu$  und unkorreliert ( $\text{cov}(X_i, X_j) = \sigma^2 \delta_{ij}$ ).

Dann gilt

$$\Rightarrow p - \lim \bar{X}_n = \mu.$$



# Gesetz der Großen Zahlen

## Anwendung 1

Das Gesetz der großen Zahlen eignet sich also z.B. zum Schätzen von Erwartungswerten.

Sei  $X \sim F$  mit Dichte  $f(x)$ , den Beobachtungen  $x_1, \dots, x_n$  und  $g(\cdot)$  eine beliebige Funktion.

Der Erwartungswert

$$\mathbf{E}(g(X)) = \int g(x)f(x) dx$$

wird (falls er existiert) geschätzt durch

$$\hat{l} = \frac{1}{n} \sum_{i=1}^n g(x_i)$$

# Gesetz der Großen Zahlen

## Anwendung 2

Das Gesetz der großen Zahlen eignet sich auch zur Approximation von Integralen.

Ist  $f > 0$  kann das Integral

$$I = \int g(x) dx$$

(falls es existiert) geschätzt werden durch

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \frac{g(x_i)}{f(x_i)},$$

wobei die Beobachtungen  $x_i$  aus einer Population mit Dichte  $f$  stammen.

# Inhalt

- Das Gesetz der Großen Zahlen
- Der Satz von GLIVENKO–CANTELLI
- Konvergenz von Folgen zufälliger Variablen
- Der zentrale Grenzwertsatz

### Def. 59 (Empirische Verteilungsfunktion)

Seien  $X_1, \dots, X_n$  unkorreliert,  $X_i \sim F$ , und  $X_{(1)}, \dots, X_{(n)}$ ,  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  die geordnete Stichprobe. Die Funktion

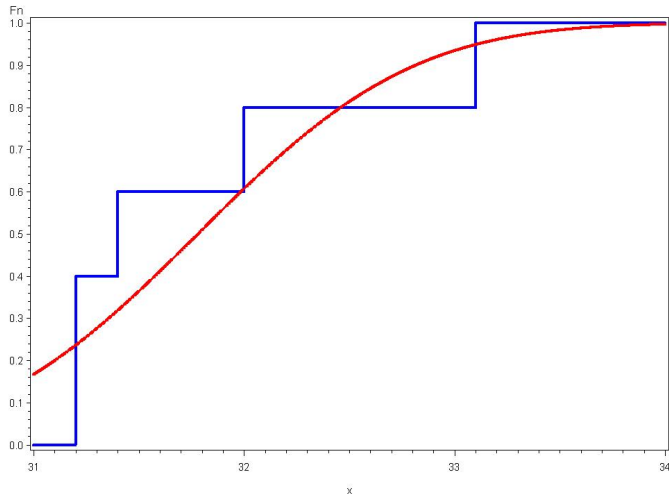
$$\begin{aligned} F_n(x) &= \frac{\#\{X_i: X_i < x, i = 1, \dots, n\}}{n} \\ &= \begin{cases} 0 & \text{falls } x \leq X_{(1)} \\ \frac{i}{n} & \text{falls } X_{(i)} < x \leq X_{(i+1)} \\ 1 & \text{falls } X_{(n)} < x \end{cases} \end{aligned}$$

heißt empirische Verteilungsfunktion.

# Der Satz von GLIVENKO–CANTELLI

Veranschaulichung der empirischen Verteilungsfunktion

## Empirische Verteilungsfunktion



# Der Satz von GLIVENKO–CANTELLI

## Satz von GLIVENKO–CANTELLI (1)

Seien  $X_1, \dots, X_n$  unkorreliert. Es gilt:

$$\lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| > \varepsilon) = 0 \quad \forall x \in \mathbb{R}.$$

**Beweis:** Wir definieren Zufallsgrößen  $Y_{ix}$  ( $i = 1, \dots, n, x \in \mathbb{R}$ ) durch:

$$Y_{ix} = \begin{cases} 1 & , \text{ falls } X_i < x \\ 0 & , \text{ sonst} \end{cases}$$

# Der Satz von GLIVENKO–CANTELLI

Beweis (Fortsetzung)

Dann gilt offensichtlich für alle  $i = 1, \dots, n$  und  $x \in \mathbb{R}$ :

$$Y_{ix} : \begin{pmatrix} 0 & 1 \\ 1 - F(x) & F(x) \end{pmatrix}$$

D.h.  $Y_{ix} \sim \text{Bi}(1, F(x))$ . Sei, für alle  $x \in \mathbb{R}$ ,

$$\overline{Y}_x := \frac{1}{n} \sum_{i=1}^n Y_{ix}.$$

Vergleichen wir die Zufallsgrößen  $F_n(x)$  und  $\overline{Y}_x$ :

$$\overline{Y}_x = F_n(x).$$

# Der Satz von GLIVENKO–CANTELLI

Beweis (Fortsetzung)

Aus dem letzten Beispiel folgt,  $\mu := \mathbf{E}Y_{ix} = F(x)$ . Deshalb folgt aus dem schwachen Gesetz der großen Zahlen:

$$\lim_{n \rightarrow \infty} P(|\overline{Y}_x - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

D.h. für alle  $\varepsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| > \varepsilon) = 0$$



# Der Satz von GLIVENKO–CANTELLI

Verschärfung:

## Satz von GLIVENKO–CANTELLI (2)

Es seien  $X_1, \dots, X_n$  unabhängige zufällige Variablen. Dann gilt:

$$P\left(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Dieser Satz wird auch oft als der Hauptsatz der Statistik bezeichnet.

# Inhalt

- Das Gesetz der Großen Zahlen
- Der Satz von GLIVENKO–CANTELLI
- Konvergenz von Folgen zufälliger Variablen
- Der zentrale Grenzwertsatz

### Def. 60 (Stochastische Konvergenz)

Eine Folge  $\{X_n\}_{n \in \mathbb{N}}$  zufälliger Variablen  
konvergiert stochastisch (in Wkt.) gegen eine zufällige Variable  $X$ ,  
falls für alle  $\varepsilon > 0$  gilt:

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Wir bezeichnen dann:  $p\text{-}\lim X_n = X$ .

$X$  heißt stochastischer Grenzwert der Folge  $\{X_n\}$ .

# Konvergenz (2)

## Def. 61 (fast sichere Konvergenz)

Eine Folge  $\{X_n\}_{n \in \mathbb{N}}$  zufälliger Variablen heißt fast sicher konvergent gegen eine zufällige Variable  $X$ , falls gilt:

$$P\left(\left\{\omega: \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

Wir bezeichnen dann:  $\lim X_n = X$  f.s.

$X$  heißt f.s. Grenzwert der Folge  $\{X_n\}$ .

# Konvergenz (3)

## Def. 62 (Konvergenz im $p$ -ten Mittel)

Es seien  $X_1, \dots, X_n, X$  zufällige Variablen mit

$$\mathbf{E}|X_i|^p < \infty, \mathbf{E}|X|^p < \infty.$$

$\{X_n\}$  konvergiert im  $p$ -ten Mittel gegen  $X$ , falls

$$\lim_{n \rightarrow \infty} \mathbf{E}|X_n - X|^p = 0.$$

Wir bezeichnen dann:  $\lim_{n \rightarrow \infty} X_n = X$  p.m.

(q.m. wenn  $p = 2$ ).

# Konvergenz (4)

## Def. 63 (Konvergenz in Verteilung)

Es sei  $\{X_n\}_{n \in \mathbb{N}}$  eine Folge von zufälligen Variablen.  $X$  sei eine Zufallsgröße mit der Verteilungsfunktion  $F(x) = P(X < x)$ .

Die Folge  $\{X_n\}_{n \in \mathbb{N}}$  konvergiert in Verteilung gegen die Zufallsgröße  $X$ , wenn für alle  $x \in \mathbb{R}$ , in denen die Funktion  $F$  stetig ist, gilt:

$$\lim_{n \rightarrow \infty} P(X_n < x) = F(x).$$

Wir bezeichnen dann:  $X_n \xrightarrow{D} X$ .

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (1)

**Lemma:** Sei  $X$  eine Zufallsvariable mit

$\mathbf{E}|X|^p < \infty$ ,  $p' < p$ . Dann gilt

$$\left(\mathbf{E}|X|^{p'}\right)^{\frac{1}{p'}} \leq \left(\mathbf{E}|X|^p\right)^{\frac{1}{p}}.$$

**Beweis:** Die Funktion  $g(x) = |x|^t$  ist konvex für  $t \geq 1$ . Für eine beliebige Zufallsvariable  $Y$  gilt (Jensens Ungleichung)

$$|\mathbf{E}Y|^t \leq \mathbf{E}|Y|^t.$$

Sei  $Y = |X|^{p'}$ ,  $t = \frac{p}{p'} \geq 1$ . Daraus folgt

$$\left(\mathbf{E}|X|^{p'}\right)^{\frac{p}{p'}} \leq \mathbf{E}\left(\left(|X|^{p'}\right)^{\frac{p}{p'}}\right) = \mathbf{E}|X|^p.$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (2)

## Folgerung

Sei  $p' < p$ .

$$\lim_{n \rightarrow \infty} X_n = X \quad p.m. \Rightarrow \lim_{n \rightarrow \infty} X_n = X \quad p'.m.$$

**Beweis:** Wegen dem letzten Lemma gilt:

$$\left( \mathbf{E} |X_n - X|^{p'} \right)^{\frac{1}{p'}} \leq \left( \mathbf{E} |X_n - X|^p \right)^{\frac{1}{p}}.$$





# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (3)

## Lemma

Sei  $p \geq 1$ . Dann gilt

$$\lim_{n \rightarrow \infty} X_n = X \quad p.m. \Rightarrow p\text{-}\lim_{n \rightarrow \infty} X_n = X.$$

**Beweis:** Sei  $\varepsilon > 0$ . Es gilt für alle  $n$ :

$$\begin{aligned} P(|X_n - X| > \varepsilon) &= P(|X_n - X|^p > \varepsilon^p) \\ &\leq \frac{\mathbf{E}|X_n - X|^p}{\varepsilon^p} \end{aligned}$$

Markov-Ungleichung

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\mathbf{E}|X_n - X|^p}{\varepsilon^p} = 0.$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (4)

Die Umkehrung dieses Satzes gilt nicht:

Seien  $X, \{X_n\}_{n \in \mathbb{N}}$  Zufallsgrößen mit

$$P(X_n = n^\alpha) = \frac{1}{n}, \quad P(X_n = 0) = 1 - \frac{1}{n}.$$

$\forall \epsilon \in (0, 1) : \quad P(|X_n| > \epsilon) = P(X_n = n^\alpha) = \frac{1}{n} \rightarrow 0$ , also

$p - \lim X_n = 0$ .

Andererseits:  $\mathbf{E}|X_n|^p = n^{\alpha p - 1}$  konvergiert nicht für  $\alpha p \geq 1$ .

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (5)

**Satz:** Seien  $X, X_1, X_2, \dots$  Zufallsgrößen

$$\lim X_n = X \quad \text{f.s.} \Rightarrow \text{p-lim } X_n = X.$$

**Beweis:** Es sei  $\varepsilon > 0$  beliebig und sei die Ereignisse

$A_n = \{\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\}$ . Dann gilt:

$$0 \leq \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) \leq \lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n)$$

(Stetigkeit des Wahrscheinlichkeitsmaßes von unten)

$$= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\}\right) = P(|X_n - X| > \varepsilon \text{ für } \infty \text{ viele } n)$$

$$\leq P(\overline{\{\lim_{n \rightarrow \infty} X_n = X\}}) = 0$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (6)

Das folgende Beispiel zeigt, daß stochastische und fast sichere Konvergenz nicht identisch sind.

Konstruktion einer Folge  $\{X_n\}_{n \in \mathbb{N}}$  zufälliger Variablen mit  $p\text{-}\lim X_n = 0$ , nicht aber  $\lim X_n = 0$  f.s.

Es seien  $\Omega = [0, 1]$  und  $\mathcal{E} = [0, 1] \cap \mathcal{B}^1$  gegeben. Für alle Ereignisse  $A \subseteq [0, 1]$  gelte:

$$0 \leq P(A) = \int_A 1 \, dx \leq 1.$$

Sei  $\{A_n\}_{n \in \mathbb{N}}$  eine Folge von Ereignissen im Ereignisfeld  $\mathcal{E}$ ,

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (7)

$$A_n := [k \cdot 2^{-h}, (k+1) \cdot 2^{-h}], \quad \forall n \in \mathbb{N}$$

wobei für die Zahlen  $h$  und  $k$  folgendes gelte:

- $h, k \in \mathbb{Z}^+ \cup \{0\}$ ;
- $n = 2^h + k; \quad (n \leq 2 \cdot 2^h)$
- $0 \leq k < 2^h$ .

Die Folge  $\{X_n\}_{n \in \mathbb{N}}$  definieren wir wie folgt:

$$X_n(\omega) = \begin{cases} 1 & , \text{ falls } \omega \in A_n \\ 0 & , \text{ sonst} \end{cases}$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (8)

Untersuchen wir die stochastische Konvergenz von  $\{X_n\}$ :

Nach Definition der Folge  $\{X_n\}_{n \in \mathbb{N}}$  gilt:

$$\begin{aligned} P(|X_n| > \varepsilon) &= P(|X_n| = 1) = P(A_n) \\ &= (k+1) \cdot 2^{-h} - k \cdot 2^{-h} \\ &= 2^{-h} \leq \frac{2}{n} \rightarrow 0, \end{aligned}$$

d.h.  $\text{p-lim } X_n = 0$ .

# Konvergenz

Zusammenhänge (9), Die Intervalle  $A_n = [k \cdot 2^{-h}, (k+1) \cdot 2^{-h}]$

| $n = 2^h + k$ | $h$ | $k$ | $A_n$              |
|---------------|-----|-----|--------------------|
| $1 = 2^0 + 0$ | 0   | 0   | $[0, 1]$           |
| $2 = 2^1 + 0$ | 1   | 0   | $[0, \frac{1}{2}]$ |
| $3 = 2^1 + 1$ | 1   | 1   | $[\frac{1}{2}, 1]$ |
| $4 = 2^2 + 0$ | 2   | 0   | $[0, \frac{1}{4}]$ |

# Konvergenz

Zusammenhänge (9), Die Intervalle  $A_n = [k \cdot 2^{-h}, (k+1) \cdot 2^{-h}]$

| $n = 2^h + k$ | $h$ | $k$ | $A_n$              |
|---------------|-----|-----|--------------------|
| $1 = 2^0 + 0$ | 0   | 0   | $[0, 1]$           |
| $2 = 2^1 + 0$ | 1   | 0   | $[0, \frac{1}{2}]$ |
| $3 = 2^1 + 1$ | 1   | 1   | $[\frac{1}{2}, 1]$ |
| $4 = 2^2 + 0$ | 2   | 0   | $[0, \frac{1}{4}]$ |

| $n = 2^h + k$ | $h$ | $k$ | $A_n$                        |
|---------------|-----|-----|------------------------------|
| $5 = 2^2 + 1$ | 2   | 1   | $[\frac{1}{4}, \frac{1}{2}]$ |
| $6 = 2^2 + 2$ | 2   | 2   | $[\frac{1}{2}, \frac{3}{4}]$ |
| $7 = 2^2 + 3$ | 2   | 3   | $[\frac{3}{4}, 1]$           |
| $8 = 2^3 + 0$ | 3   | 0   | $[0, \frac{1}{8}]$           |



# Konvergenz

Zusammenhänge (9), Die Intervalle  $A_n = [k \cdot 2^{-h}, (k+1) \cdot 2^{-h}]$

| $n = 2^h + k$ | $h$ | $k$ | $A_n$              |
|---------------|-----|-----|--------------------|
| $1 = 2^0 + 0$ | 0   | 0   | $[0, 1]$           |
| $2 = 2^1 + 0$ | 1   | 0   | $[0, \frac{1}{2}]$ |
| $3 = 2^1 + 1$ | 1   | 1   | $[\frac{1}{2}, 1]$ |
| $4 = 2^2 + 0$ | 2   | 0   | $[0, \frac{1}{4}]$ |

| $n = 2^h + k$ | $h$ | $k$ | $A_n$                        |
|---------------|-----|-----|------------------------------|
| $5 = 2^2 + 1$ | 2   | 1   | $[\frac{1}{4}, \frac{1}{2}]$ |
| $6 = 2^2 + 2$ | 2   | 2   | $[\frac{1}{2}, \frac{3}{4}]$ |
| $7 = 2^2 + 3$ | 2   | 3   | $[\frac{3}{4}, 1]$           |
| $8 = 2^3 + 0$ | 3   | 0   | $[0, \frac{1}{8}]$           |

Die Folge  $\{A_n\}_{n \in \mathbb{N}}$  ist nirgends konvergent. Also

$$P\left(\left\{\omega: \lim_{n \rightarrow \infty} X_n(\omega) = 0\right\}\right) = 0 \neq 1.$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (10)

## Satz

Es sei  $\{X_n\}_{n \in \mathbb{N}}$  eine Folge von zufälligen Variablen, für die es zwei Zufallsgrößen  $X$  und  $Y$  gibt, so dass gilt:

$$X = \text{p-lim } X_n \quad \text{und} \quad Y = \text{p-lim } X_n.$$

Dann folgt daraus:

$$P(X = Y) = 1.$$

**Beweis:** Es sei  $\varepsilon > 0$  beliebig. Dann berechnen wir

$$P(\{\omega: |X(\omega) - Y(\omega)| > \varepsilon\}) = (*)$$

# Konvergenz

Beweis des Satzes, (\*)  $= P(|X - Y| > \varepsilon)$

$$\begin{aligned} &= P(|X - X_n + X_n - Y| > \varepsilon) \\ &\leq P(|X - X_n| + |X_n - Y| > \varepsilon) \\ &\leq P\left(\left\{|X_n - X| > \frac{\varepsilon}{2}\right\} \cup \left\{|X_n - Y| > \frac{\varepsilon}{2}\right\}\right) \\ &\leq P\left(\left\{|X_n - X| > \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{|X_n - Y| > \frac{\varepsilon}{2}\right\}\right) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

D.h.

$$P(|X - Y| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

$$P(\{\omega: X(\omega) = Y(\omega)\}) = 1.$$

# Konvergenz

Zusammenhänge zwischen den Konvergenzbegriffen (11)

## Lemma

$$p\text{-}\lim_{n \rightarrow \infty} X_n = X \Rightarrow X_n \rightarrow^D X$$

**Beweis:** Seien  $x' < x < x'' \in \mathbb{R}$  sowie  $F$  und  $F_n$  die Verteilungsfunktionen von  $X$  bzw.  $X_n$ . Es gilt:

$$\begin{aligned} \{X < x'\} &= \{X < x', X_n < x\} \cup \{X < x', X_n \geq x\} \\ &\subseteq \{X_n < x\} \cup \{X < x', X_n \geq x\} \Rightarrow \\ F(x') &\leq F_n(x) + \underbrace{P(|X_n - X| \geq x - x')}_{\rightarrow 0 \text{ wegen } X_n \rightarrow_p X} \\ F(x') &\leq \lim_{n \rightarrow \infty} F_n(x) \end{aligned}$$

Beweis von  $p\text{-}\lim_{n \rightarrow \infty} X_n = X \Rightarrow X_n \xrightarrow{D} X$  (2)

Weiterhin

$$\{X_n < x\} = \{X < x'', X_n < x\} \cup \{X \geq x'', X_n < x\}$$

$$\subseteq \{X < x''\} \cup \{X \geq x'', X_n < x\} \Rightarrow$$

$$F_n(x) \leq F(x'') + \underbrace{P(|X_n - X| \geq x'' - x)}_{\rightarrow 0 \text{ wegen } X_n \rightarrow_p X}$$

$$\overline{\lim}_{n \rightarrow \infty} F_n(x) \leq F(x'')$$

Beides zusammen:

$$F(x') \leq \underline{\lim}_{n \rightarrow \infty} F_n(x) \leq \overline{\lim}_{n \rightarrow \infty} F_n(x) \leq F(x'')$$

Beweis von  $\text{p-lim}_{n \rightarrow \infty} X_n = X \Rightarrow X_n \rightarrow^D X$  (3)

Wenn jetzt  $x$  Stetigkeitsstelle und  $x' \rightarrow x - 0$  und  $x'' \rightarrow x + 0$  so  
 $F(x') \rightarrow F(x)$  und  $F(x'') \rightarrow F(x)$  und

$$\lim F_n(x) = F(x).$$

Die Rückrichtung gilt i.A. nicht:

$$X \sim \text{Bi}(1, \frac{1}{2}), X_n = 1 - X \quad \forall n \in \mathbb{N}$$

$X$  und  $X_n$  besitzen dieselbe Verteilung  $\text{Bi}(1, \frac{1}{2})$ ,  $X_n \rightarrow^D X$ .

Es gilt aber nicht:  $X_n \rightarrow_p X$ , da

$$|X_n - X| = |1 - 2X| = 1 \quad \forall n \in \mathbb{N}$$

# Konvergenzarten

Die verschiedenen Arten der Konvergenz einer Folge von Zufallsgrößen gegen eine zufällige Variable bilden z.T. eine gewisse Hierarchie:

$$\lim X_n = X \text{ f.s.} \implies \text{p-lim } X_n = X$$

$$\implies X_n \xrightarrow{D} X$$

$$\lim X_n = X \text{ q.m.} \implies \text{p-lim } X_n = X$$

$$\lim X_n = X \text{ p.m.} \implies \text{p-lim } X_n = X \quad (p \geq 1)$$

Die Umkehrungen gelten im allgemeinen nicht.

# Konvergenz in Verteilung

## Beispiel

$$X_n \sim Bi(n, p_n), \quad \lim np_n = \lambda, \quad Y \sim Poi(\lambda) \quad \Rightarrow X_n \rightarrow^D Y.$$

Diese Aussage kennen wir schon von früher.

Weitere werden wir im nächsten Abschnitt kennenlernen.



# Inhalt

- Das Gesetz der Großen Zahlen
- Der Satz von GLIVENKO–CANTELLI
- Konvergenz von Folgen zufälliger Variablen
- Der zentrale Grenzwertsatz

## Der Zentrale Grenzwertsatz

Es seien  $X_1, \dots, X_n$  unabhängige, identisch verteilte Zufallsvariablen mit  $\mu := \mathbf{E}X_i$ ;  $\sigma^2 := \text{Var } X_i < \infty$ . Seien Zufallsgrößen  $Z_n$ ,  $\bar{Z}_n$  und  $Y_n$  definiert durch:  $Z_n := \sum_{i=1}^n X_i$  bzw.  $\bar{Z}_n := \frac{Z_n}{n}$  und

$$Y_n = \sqrt{n} \cdot \frac{\bar{Z}_n - \mu}{\sigma} = \frac{Z_n - n\mu}{\sqrt{n}\sigma}$$

Dann gilt für alle reellen  $x$ :

$$\lim_{n \rightarrow \infty} P\left(\frac{Z_n - n\mu}{\sqrt{n}\sigma} < x\right) = \lim_{n \rightarrow \infty} P(Y_n < x) = \Phi(x)$$

Erinnerung:  $\Phi(x)$  ist die Verteilungsfunktion der Standardnormalverteilung.

# Der zentrale Grenzwertsatz

**Beweis:** Ich biete Ihnen unten 2 völlig verschiedene Beweise an. Der erste verwendet als Hilfsmittel charakteristische Funktionen (Fourier-Transformierte), den zweiten lege ich Ihnen nahe, da er interessante Ideen beinhaltet und gleichzeitig einigen Stoff wiederholt.

**Bem.:** Die Folge  $\{Y_n\}_{n \in \mathbb{N}}$  konvergiert in Verteilung gegen eine Zufallsgröße  $Z$ ,  $Y_n \xrightarrow{D} Z$ ,  $Z \sim \mathcal{N}(0, 1)$ .

Anwendungen:

- Simulation bei der Erzeugung einer normalverteilten Zufallsgröße aus Pseudozufallszahlen
- Approximation von Wahrscheinlichkeitsverteilungen (insbesondere von Schätz- und Teststatistiken)

# Der zentrale Grenzwertsatz

## Genauigkeitsabschätzung (1)

### Satz (BERRY-ESSÉEN SCHRANKE)

Es seien die Voraussetzungen des zentralen Grenzwertsatzes erfüllt und  $M := \mathbf{E}|X_i - \mu|^3 < \infty$ . Dann gilt:

$$\left| P\left(\frac{Z_n - n\mu}{\sqrt{n}\sigma} < x\right) - \Phi(x) \right| < \frac{0.4748 \cdot M}{\sigma^3 \cdot \sqrt{n}} =: K,$$

**Bem.:** Die Schranke wird laufend verbessert (aktueller Stand: 2019).

# Der zentrale Grenzwertsatz

Genauigkeitsabschätzung nach Berry-Esséen (2)

Es seien  $X_i \sim R(0, 1)$ ,  $\mu = \frac{1}{2}$ ,  $\sigma^2 = \frac{1}{12}$

$$\begin{aligned} M &= \mathbf{E}|X_i - \mu|^3 = \int_{-\infty}^{+\infty} |x - \mu|^3 \cdot f(x) dx \\ &= \int_0^1 |x - \frac{1}{2}|^3 dx = 2 \cdot \int_{\frac{1}{2}}^1 (x - \frac{1}{2})^3 dx = \frac{1}{32} \end{aligned}$$

| $n$ | 12    | 100   | 1000   |
|-----|-------|-------|--------|
| $K$ | 0.178 | 0.062 | 0.0195 |

# Der zentrale Grenzwertsatz

## Genauigkeitsabschätzung (3)

Seien  $X_i \sim \text{Poi}(\lambda)$ ,  $\mathbf{E}X_i = \text{Var } X_i = \lambda$

$$\begin{aligned} M^{\frac{1}{3}} &= \left( \mathbf{E}|X_i - \lambda|^3 \right)^{\frac{1}{3}} \leq \left( \mathbf{E}|X_i - \lambda|^4 \right)^{\frac{1}{4}} \\ &= \left( \mathbf{E}(X_i - \lambda)^4 \right)^{\frac{1}{4}} = (\lambda + 3\lambda^2)^{\frac{1}{4}} \end{aligned}$$

Berry-Esseen Schranke:

$$K \leq \frac{0.4748(\lambda + 3\lambda^2)^{\frac{3}{4}}}{\lambda^{\frac{3}{2}}\sqrt{n}} \rightarrow_{\lambda \rightarrow \infty} \frac{0.4748 \cdot 3^{\frac{3}{4}}}{\sqrt{n}} =: K'$$

|      |       |       |        |
|------|-------|-------|--------|
| $n$  | 12    | 100   | 1000   |
| $K'$ | 0.312 | 0.108 | 0.0342 |

# Der zentrale Grenzwertsatz

$X_i$  Bernoulli

## Satz (MOIVRE–LAPLACE)

Es seien  $X_i \sim Bi(1, p)$ , unabhängig. Dann gilt für  $Z_n = \sum_{i=1}^n X_i$  ( $\sim Bi(n, p)$ ):

$$Z_n \xrightarrow{D} Z \sim \mathcal{N}(np, np(1-p))$$

**Bem.:** Für ausreichend großes  $n \in \mathbb{N}$  kann also die Binomialverteilung durch eine Normalverteilung ersetzt werden,

$$P(Z_n < y) \approx \Phi \left( \frac{y - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \right).$$

# Satz von MOIVRE–LAPLACE

## Beweis

**Beweis:** Mit  $\mathbf{E}Z_n = np$  und  $\text{Var } Z_n = np(1 - p)$  folgt unter Anwendung des Zentralen Grenzwertsatzes:

$$\begin{aligned} F_{Z_n}(y) &= P(Z_n < y) \\ &= P\left(\frac{Z_n - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}} < \frac{y - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}\right) \\ &\rightarrow \Phi\left(\frac{y - n \cdot p}{\sqrt{n \cdot p \cdot (1 - p)}}\right) \end{aligned}$$





# Der zentrale Grenzwertsatz

## Satz von MOIVRE–LAPLACE

Es seien  $X_i \sim Bi(1, p)$ ,  $n = 1000$  und  $p = 0.4$ . Gesucht werde die Wahrscheinlichkeit  $P(Z_n < 300)$ . Es gilt:

$$\begin{aligned} P(Z_n < 300) &= \sum_{x < 300} P(Z_n = x) \\ &= \sum_{i=0}^{299} \binom{1000}{i} 0.4^i (1 - 0.4)^{1000-i} \end{aligned}$$

großer Rechenaufwand.

besser: Anwendung des Satzes von MOIVRE–LAPLACE.

# Satz von MOIVRE–LAPLACE

Beispiel, Fortsetzung

Es gilt:

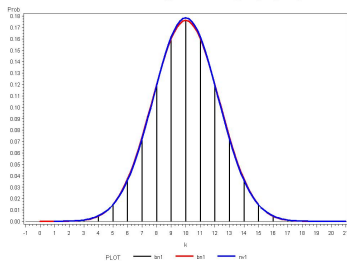
$$\begin{aligned}P(Z_n < 300) &\approx \Phi\left(\frac{300 - 1000 \cdot 0.4}{\sqrt{1000 \cdot 0.4 \cdot (1 - 0.4)}}\right) \\&= \Phi\left(\frac{-100}{\sqrt{240}}\right) \approx \Phi\left(\frac{-100}{15.49}\right) \\&= \Phi(-6.45) = 1 - \underbrace{\Phi(6.45)}_{\approx 1} \approx 0\end{aligned}$$

**Bem.:** Die Anwendung des Satzes von MOIVRE–LAPLACE setzt voraus, dass  $n \in \mathbb{N}$  hinreichend groß ist.

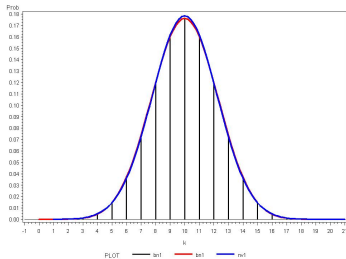
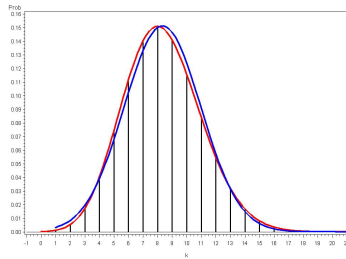
Faustregel:  $n \cdot p \geq 10$  und  $n \cdot (1 - p) \geq 10$ .

# Satz von MOIVRE–LAPLACE

Binomialverteilung-Verteilung  $B(20,0.5)$

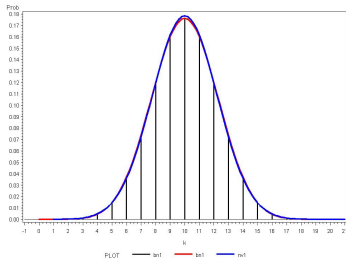


# Satz von MOIVRE–LAPLACE

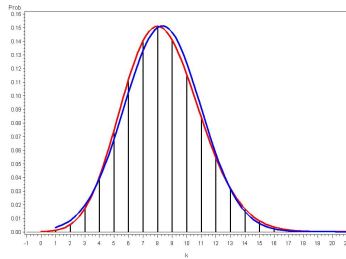
Binomialverteilung-Verteilung  $B(20,0.5)$ Binomialverteilung-Verteilung  $B(50,1/6)$ 

# Satz von MOIVRE–LAPLACE

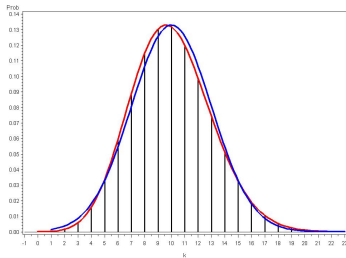
Binomialverteilung-Verteilung B(20,0.5)



Binomialverteilung-Verteilung B(50,1/6)



Binomialverteilung-Verteilung B(100,0.1)



# Der zentrale Grenzwertsatz

$X_i$  Poisson

Seien  $X_i$  unabhängig,  $X_i \sim \text{Poi}(\lambda_i)$ ,  $i = 1, \dots, n$

$$X_i : \begin{pmatrix} 0 & 1 & 2 & \dots & k & \dots \\ p_{0i} & p_{1i} & p_{2i} & \dots & p_{ki} & \dots \end{pmatrix} \quad Z_n := \sum_{i=1}^n X_i$$

mit  $p_{ji} = \frac{\lambda_i^j}{j!} \cdot e^{-\lambda_i}$ ,  $\mathbf{E}X_i = \text{Var } X_i = \lambda_i$ .

Für den Erwartungswert von  $Z_n$  gilt:

$$\mathbf{E}Z_n = \mathbf{E} \left( \sum_{i=1}^n X_i \right) = \sum_{i=1}^n \mathbf{E}X_i = \sum_{i=1}^n \lambda_i$$

# Der zentrale Grenzwertsatz

Poisson

## Lemma

Es seien  $X_1$  und  $X_2$  unabhängig,

$X_1, X_2 \sim \text{Poi}(\lambda_i)$ ,  $i = 1, 2$ . Dann ist die Zufallsgröße  $Z_2 := X_1 + X_2$  ebenfalls POISSON-verteilt und es gilt:  $Z_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$ .

**Bem:** Vergleichen Sie die folgende Formel mit der Faltungsformel für stetige Zufallsvariablen. Erinnerung:  $\mathbf{E}X_i = \lambda_i$ ,  $\text{Var } X_i = \lambda_i$ .

# Der zentrale Grenzwertsatz

Poisson, Beweis des Lemma

**Beweis:** Es gilt für alle  $k \in \mathbb{N}$ :

$$P(Z_2 = k) = \sum_{t=0}^k p_1(t) \cdot p_2(k - t)$$



# Der zentrale Grenzwertsatz

Poisson, Beweis des Lemma

**Beweis:** Es gilt für alle  $k \in \mathbb{N}$ :

$$\begin{aligned} P(Z_2 = k) &= \sum_{t=0}^k p_1(t) \cdot p_2(k-t) \\ &= \sum_{t=0}^k \left( \frac{\lambda_1^t}{t!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{k-t}}{(k-t)!} \cdot e^{-\lambda_2} \right) \end{aligned}$$

# Der zentrale Grenzwertsatz

Poisson, Beweis des Lemma

**Beweis:** Es gilt für alle  $k \in \mathbb{N}$ :

$$\begin{aligned}P(Z_2 = k) &= \sum_{t=0}^k p_1(t) \cdot p_2(k-t) \\&= \sum_{t=0}^k \left( \frac{\lambda_1^t}{t!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{k-t}}{(k-t)!} \cdot e^{-\lambda_2} \right) \\&= \sum_{t=0}^k \left( \frac{\lambda_1^t \cdot \lambda_2^{k-t}}{t! \cdot (k-t)!} \cdot e^{-(\lambda_1 + \lambda_2)} \right) \\&= e^{-(\lambda_1 + \lambda_2)} \cdot \frac{1}{k!} \cdot \sum_{t=0}^k \frac{\lambda_1^t \cdot \lambda_2^{k-t} \cdot k!}{t! \cdot (k-t)!}\end{aligned}$$

# Der zentrale Grenzwertsatz

Poisson, Beweis des Lemma

**Beweis:** Es gilt für alle  $k \in \mathbb{N}$ :

$$\begin{aligned}P(Z_2 = k) &= \sum_{t=0}^k p_1(t) \cdot p_2(k-t) \\&= \sum_{t=0}^k \left( \frac{\lambda_1^t}{t!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{k-t}}{(k-t)!} \cdot e^{-\lambda_2} \right) \\&= \sum_{t=0}^k \left( \frac{\lambda_1^t \cdot \lambda_2^{k-t}}{t! \cdot (k-t)!} \cdot e^{-(\lambda_1 + \lambda_2)} \right) \\&= e^{-(\lambda_1 + \lambda_2)} \cdot \frac{1}{k!} \cdot \sum_{t=0}^k \frac{\lambda_1^t \cdot \lambda_2^{k-t} \cdot k!}{t! \cdot (k-t)!} \\&= \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \cdot (\lambda_1 + \lambda_2)^k \quad (\text{Binomische Formel})\end{aligned}$$

# Der zentrale Grenzwertsatz

Poisson

Sei  $\lambda_i = \lambda$  ( $i = 1, \dots, n$ ). Dann

$$Z_n = \sum_{i=1}^n X_i \sim \text{Poi}(n \cdot \lambda).$$

Anwendung des Zentralen Grenzwertsatz liefert ( $\lambda' := n \cdot \lambda$ ):

$$P\left(\frac{Z_n - n \cdot \mu}{\sqrt{n \cdot \sigma}} < x\right) = P\left(\frac{Z_n - \lambda'}{\sqrt{\lambda'}} < x\right) \rightarrow \Phi(x).$$

Also kann auch eine POISSON-Verteilung durch eine Normalverteilung approximiert werden, falls der Parameter  $\lambda' (= n\lambda)$  hinreichend groß ist (etwa  $\lambda' \geq 10$ ).

# $\chi^2$ -Verteilung

## Def. 64 ( $\chi^2$ -Verteilung)

Seien  $X_i$  unabhängig,  $X_i \sim \mathcal{N}(0, 1)$ ,  $i = 1, \dots, n$ .

$$Y = \sum_{i=1}^n X_i^2 \sim \chi_n^2,$$

heißt  $\chi^2$  verteilt mit  $n$  Freiheitsgraden.

Dichte:

$$f_Y(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n-2}{2}} e^{-\frac{x}{2}}, & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

# $\chi^2$ -Verteilung

$$\mathbf{E}Y = n\mathbf{E}X_i^2 = n$$

# $\chi^2$ -Verteilung

$$\mathbf{E}Y = n\mathbf{E}X_i^2 = n$$

$$\text{Var } Y = \mathbf{E}(Y - n)^2 = \mathbf{E}\left(\sum_{i=1}^n (X_i^2 - 1)\right)^2 = n\mathbf{E}(X_1^2 - 1)^2$$

# $\chi^2$ -Verteilung

$$\mathbf{E}Y = n\mathbf{E}X_i^2 = n$$

$$\begin{aligned}\text{Var } Y &= \mathbf{E}(Y - n)^2 = \mathbf{E}\left(\sum_{i=1}^n (X_i^2 - 1)\right)^2 = n\mathbf{E}(X_1^2 - 1)^2 \\ &= n\mathbf{E}(X_1^4 - 2\mathbf{E}X_1^2 + 1) = n(\underbrace{3}_{s.f.S.} - 2 + 1) = 2n.\end{aligned}$$

$$\Rightarrow \lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}} < y\right) = \Phi(y)$$

$$P\left(\sum_{i=1}^n X_i^2 < x\right) \approx \Phi\left(\frac{x - n}{\sqrt{2n}}\right)$$



# $\chi^2$ -Verteilung

$$n = 30, x = 23.364: P(\sum_{i=1}^n X_i^2 < x) = 0.2$$

Approximation durch eine Normalverteilung:

$$\Phi\left(\frac{x - n}{\sqrt{2n}}\right) = \Phi(-0.8567) = 1 - 0.8042 = 0.1958.$$

# $\chi^2$ -Verteilung, Fortsetzung

bleibt z.z.:  $\mathbf{E}X_i^4 = 3$ .

$$\begin{aligned}\sqrt{2\pi}\mathbf{E}X_i^4 &= \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} dx \\&= 2 \int_0^{\infty} x^4 e^{-\frac{x^2}{2}} dx, \quad t = x^2, \quad dx = \frac{1}{2} t^{-\frac{1}{2}} dt \\&= \int_0^{\infty} t^{\frac{3}{2}} e^{-\frac{t}{2}} dt = \int_0^{\infty} t^{\frac{5}{2}-1} e^{-\frac{t}{2}} dt, \quad s = \frac{t}{2} \\&= \int_0^{\infty} (2s)^{\frac{5}{2}-1} e^{-s} \cdot 2 ds = \Gamma\left(\frac{5}{2}\right) 2^{\frac{5}{2}} = \Gamma\left(2 + \frac{1}{2}\right) 2^{\frac{5}{2}} \\&= 1 \cdot 3 \cdot \frac{\sqrt{\pi}}{4} \cdot 2^{\frac{5}{2}} = 3 \cdot \sqrt{2\pi} \\ \mathbf{E}X_i^4 &= 3.\end{aligned}$$

# $\chi^2$ -Verteilung, Fortsetzung

Dabei haben wir verwendet:

$$\int_0^{\infty} t^{\lambda-1} e^{-\alpha t} dt = \frac{\Gamma(\lambda)}{\alpha^{\lambda}}$$

$$\Gamma(n+1) = n\Gamma(n) = n!$$

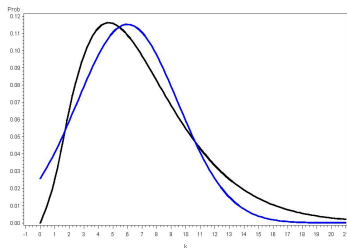
$$\Gamma\left(n + \frac{1}{2}\right) = 1 \cdot 3 \cdot 5 \cdots (2n-1) \frac{\sqrt{\pi}}{2^n}$$

# $\chi^2$ -Verteilung

Veranschaulichung für verschiedene  $n$ ,  $n = 6, 10, 20, 40$

Verteilung der Summe von Quadraten von Zufallsgrößen

$X \sim \text{standardnormal}, n=6$

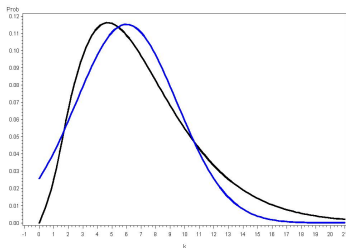


# $\chi^2$ -Verteilung

Veranschaulichung für verschiedene  $n$ ,  $n = 6, 10, 20, 40$

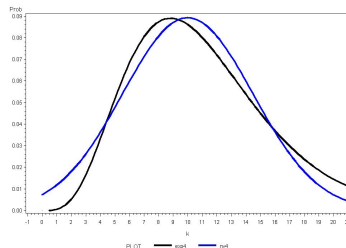
Verteilung der Summe von Quadraten von Zufallsgrößen

$\chi^2$  standardnormal,  $n=6$



Verteilung der Summe von Quadraten von Zufallsgrößen

$\chi^2$  standardnormal,  $n=10$

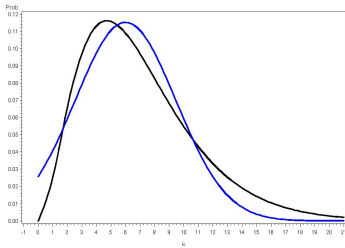


# $\chi^2$ -Verteilung

Veranschaulichung für verschiedene  $n$ ,  $n = 6, 10, 20, 40$

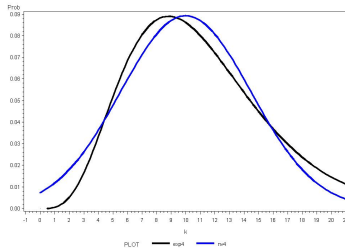
Verteilung der Summe von Quadraten von Zufallsgrößen

$\chi^2$  standardnormal,  $n=6$



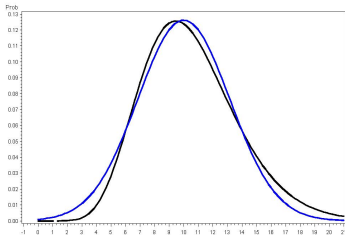
Verteilung der Summe von Quadraten von Zufallsgrößen

$\chi^2$  standardnormal,  $n=10$



Verteilung der Summe von Quadraten von Zufallsgrößen

$\chi^2$  standardnormal,  $n=20$

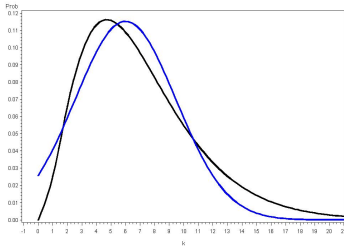


# $\chi^2$ -Verteilung

Veranschaulichung für verschiedene  $n$ ,  $n = 6, 10, 20, 40$

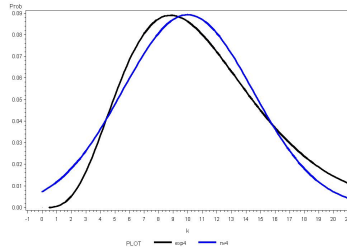
Verteilung der Summe von Quadraten von Zufallsgrößen

$X_i$  standardnormal,  $n=6$



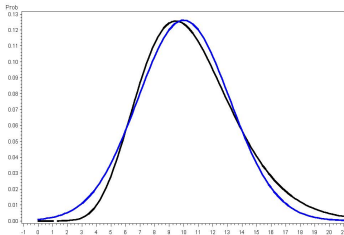
Verteilung der Summe von Quadraten von Zufallsgrößen

$X_i$  standardnormal,  $n=10$



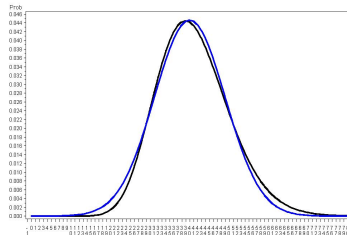
Verteilung der Summe von Quadraten von Zufallsgrößen

$X_i$  standardnormal,  $n=20$



Verteilung der Summe von Quadraten von Zufallsgrößen

$X_i$  standardnormal,  $n=40$



## \*Beweis des Zentralen Grenzwertsatzes

Sei  $\phi_{X-\mu}$  die charakteristische Funktion von  $X_i - \mu$ . Da die ersten beiden Momente  $(\mu, \sigma^2)$  existieren,  $\mathbf{E}(X_i - \mu) = 0$ ,  $\mathbf{E}(X_i - \mu)^2 = \sigma^2$ , folgt aus der Taylorreihendarstellung

$$\phi_{X-\mu}(t) = \sum_{j=0}^k \mathbf{E}(X_i - \mu)^j \frac{(it)^j}{j!} + o(t^k) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2)$$

Die Zufallsvariablen

$$\frac{X_i - \mu}{\sqrt{n}\sigma}$$

haben die charakteristische Funktion

$$\phi_{X-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right) = 1 - \frac{1}{2n}\sigma^2 t^2 + o(t^2)$$



## \*Beweis des Zentralen Grenzwertsatzes (2)

Die Zufallsvariable  $Y_n = \sum_{i=1}^n \frac{X_i - \mu}{\sqrt{n}\sigma}$  hat also die charakteristische Funktion

$$\phi_{Y_n}(t) = \left( \phi_{X-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right) \right)^n = \left( 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)^n.$$

Es gilt:

$$\ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = n \ln\left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right) \rightarrow -\frac{t^2}{2}.$$

(vgl. Taylorreihenentwicklung des Logarithmus)

## \*Beweis des Zentralen Grenzwertsatzes (3)

$$\ln \phi_{Y_n}(t) \rightarrow -\frac{t^2}{2}$$

$$\phi_{Y_n}(t) \rightarrow e^{-\frac{t^2}{2}}.$$

D.h. die charakteristische Funktion von  $Y_n$  konvergiert gegen die charakteristische Funktion der Standard-Normalverteilung (sogar gleichmäßig).

Aus dem Konvergenzsatz folgt:  $Y_n \rightarrow Z \sim \mathcal{N}(0, 1)$ .

# Weiterer Beweis des Zentralen Grenzwertsatzes

ohne Verwendung komplexer Funktionen

Ich mache den Beweis mit Ihnen, weil er einige wichtige Kenntnisse wiederholt (insbesondere Rechnen mit Erwartungswerten).

Wir verschärfen die Voraussetzungen etwas, und fordern zusätzlich

$$\mathbf{E}|X_i|^3 < \infty.$$

Erinnerung: Voraussetzungen:  $X_1, \dots, X_n$  unabhängig, identisch verteilt und  $\mathbf{E}|X_i|^2 < \infty$ .

$$P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} < x\right) \rightarrow_{n \rightarrow \infty} P(Z < x)$$

wobei  $Z \sim \mathcal{N}(0, 1)$ . Wir setzen  $Y_i = \frac{X_i - \mu}{\sigma}$ . Es genügt zu zeigen

$$P\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} < x\right) \rightarrow_{n \rightarrow \infty} P(Z < x).$$

1. Idee. Die Wahrscheinlichkeit ist ein Spezialfall des Erwartungswertes Sei  $I_A$  Indkatorfunktion der Menge  $A$ , insbesondere gilt für die Zufallsvariable

$$I_{(-\infty, x)}(X) = \begin{cases} 1 & \text{falls } X < x \\ 0 & \text{sonst} \end{cases} \sim Bi(1, p) \quad \text{mit } p = P(X < x)$$

und deshalb  $P(X < x) = \mathbf{E}(I_{(-\infty, x)}(X))$ , insbesondere

$$P\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}} < x\right) = \mathbf{E}\left(I_{(-\infty, x)}\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\right)\right)$$

2. Idee. Die Indikatorfunktion kann durch eine glatte Funktion approximiert werden (aus der Analysis bzw. Angewandten Mathematik bekannt)

Sei die Funktion  $h(x)$  eine glatte Approximation der Indikatorfunktion, “glatt” heißt hier: 3 mal stetig differenzierbar und  $|h'(x)|, |h''(x)|, |h'''(x)|$  beschränkt.

Es genügt also zu zeigen, dass für glatte Funktionen  $h$  gilt

$$\mathbf{E}h\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\right) \rightarrow \mathbf{E}h(Z)$$

( $h(x) = I_{(-\infty, x)}(x)$  ist ein Spezialfall,  $Z \sim \mathcal{N}(0, 1)$ ).

3. Idee. Die Beobachtungen  $Y_i$  werden gekoppelt mit Beobachtungen  $Z_i \sim \mathcal{N}(0, 1)$ , wobei alle Beobachtungen  $Y_1, \dots, Y_n, Z_1, \dots, Z_n$  unabhängig sind.

4. Erinnerung: Die Summe normalverteilter Zufallsvariablen ist normal.

$$\frac{\sum_{i=1}^n Z_i}{\sqrt{n}} \sim \mathcal{N}(0, 1), \text{ also } \mathbf{E}h\left(\frac{\sum_{i=1}^n Z_i}{\sqrt{n}}\right) = \mathbf{E}h(Z)$$

(hier gilt das Gleichheitszeichen)

5. Linearität des Erwartungswertes

$$\mathbf{E}h\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\right) - \mathbf{E}h\left(\frac{\sum_{i=1}^n Z_i}{\sqrt{n}}\right) = \mathbf{E}\left(h\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\right) - h\left(\frac{\sum_{i=1}^n Z_i}{\sqrt{n}}\right)\right)$$

und es genügt zu zeigen, dass der Erwartungswert auf der rechten Seite gegen Null geht.

6. Idee. Wir schreiben die Differenz als sogenannte Teleskopsumme

$$\begin{aligned}
 & h\left(\frac{\sum_{i=1}^n Y_i}{\sqrt{n}}\right) - h\left(\frac{\sum_{i=1}^n Z_i}{\sqrt{n}}\right) \\
 = & h\left(\frac{Y_1 + \dots + Y_{n-1} + \textcolor{blue}{Y}_n}{\sqrt{n}}\right) - h\left(\frac{Y_1 + \dots + Y_{n-1} + \textcolor{blue}{Z}_n}{\sqrt{n}}\right) \\
 + & h\left(\frac{Y_1 + \dots + \textcolor{blue}{Y}_{n-1} + Z_n}{\sqrt{n}}\right) - h\left(\frac{Y_1 + \dots + Y_{n-2} + \textcolor{blue}{Z}_{n-1} + Z_n}{\sqrt{n}}\right) \\
 + & \dots \\
 + & h\left(\frac{Y_1 + \textcolor{blue}{Y}_2 + Z_3 + \dots + Z_n}{\sqrt{n}}\right) - h\left(\frac{Y_1 + \textcolor{blue}{Z}_2 + Z_3 + \dots + Z_n}{\sqrt{n}}\right) \\
 + & h\left(\frac{\textcolor{blue}{Y}_1 + Z_2 + \dots + Z_n}{\sqrt{n}}\right) - h\left(\frac{\textcolor{blue}{Z}_1 + Z_2 + \dots + Z_n}{\sqrt{n}}\right)
 \end{aligned}$$

7. Idee. Jede dieser Zeilen entwickeln wir in eine Taylorreihe, an den Stellen  $Z_i$  ( $i = n, \dots, 1$  (wenn der Index kleiner Null oder größer  $n$  dann existiert die entsprechende Summe nicht)

$$\begin{aligned}
 & h\left(\frac{Y_1 + \dots + Y_i + Z_{i+1} + \dots + Z_n}{\sqrt{n}}\right) - h\left(\frac{Y_1 + \dots + Y_{i-1} + Z_i + \dots + Z_n}{\sqrt{n}}\right) \\
 &= \frac{Y_i - Z_i}{\sqrt{n}} h'\left(\frac{Y_1 + \dots + Y_{i-1} + Z_i + \dots + Z_n}{\sqrt{n}}\right) \\
 &\quad + \frac{(Y_i - Z_i)^2}{2n} h''\left(\frac{Y_1 + \dots + Y_{i-1} + Z_i + \dots + Z_n}{\sqrt{n}}\right) \\
 &\quad + \frac{(Y_i - Z_i)^3}{6n^{3/2}} h'''(Y_i^*) \quad \text{wobei } Y_i^* \text{ zwischen } Y_i \text{ und } Z_i \text{ liegt.}
 \end{aligned}$$



8. Wegen  $\mathbf{E}Y_i = \mathbf{E}Z_i = 0$  und  $\mathbf{E}Y_i^2 = \mathbf{E}Z_i^2 = 1$  sind die Erwartungswerte der ersten beiden Summanden Null. Es bleibt

$$\begin{aligned} \left| \mathbf{E} \left( h \left( \frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \right) \right) - \mathbf{E} \left( h \left( \frac{\sum_{i=1}^n Z_i}{\sqrt{n}} \right) \right) \right| &\leq \sum_{i=1}^n \left| \frac{\mathbf{E}(Y_i - Z_i)^3}{6n^{3/2}} h'''(Y'_i) \right| \\ &\leq \frac{C}{6\sqrt{n}} \rightarrow_{n \rightarrow \infty} 0 \end{aligned}$$

mit Konstante  $C$ , da wir die beiden Faktoren im Zähler als beschränkt vorausgesetzt haben.

# Zentraler Grenzwertsatz

## Beispiele

Münzwurf: 1000 mal. Wie groß ist die Wahrscheinlichkeit, dass weniger als 475 mal Zahl fällt?

# Zentraler Grenzwertsatz

## Beispiele

Münzwurf: 1000 mal. Wie groß ist die Wahrscheinlichkeit, dass weniger als 475 mal Zahl fällt?

$X_i = 1$  falls Zahl,  $X_i = 0$  sonst.

$$P(\sum_{i=1}^{1000} X_i < 475) =$$

$$\begin{aligned} & P\left(\underbrace{\frac{\sum X_i - \frac{1}{2}}{\sqrt{\frac{1}{4}}}}_{\sim \mathcal{N}(0,1)} \leq \frac{475 - \frac{1}{2}}{\sqrt{\frac{1}{4}}}\right) \\ & \approx \Phi\left(\sqrt{1000} \frac{0.475 - 0.5}{\frac{1}{2}}\right) \\ & = \Phi(-1.58) \approx 0.057. \end{aligned}$$

# Bedeutung des ZGWS in der Statistik

beim Schätzen

Gesetz der Großen Zahlen:  $\bar{X} \rightarrow \mu = \mathbf{E}(X)$ .

Frage: Wie groß ist der Stichprobenumfang zu wählen, um eine bestimmte Genauigkeit zu erreichen?

$\varepsilon, \delta$  vorgegeben, klein ( $\varepsilon, \delta < 0.5$ ).

$n$  ist so zu wählen, dass

$$P(|\bar{X} - \mu| \leq \varepsilon) \geq 1 - \delta$$

# Bedeutung des ZGWS beim Schätzen

Fortsetzung

$$\begin{aligned} 1 - \delta &\leq P(|\bar{X} - \mu| \leq \varepsilon) \\ &= P\left(\sqrt{n} \frac{|\bar{X} - \mu|}{\sqrt{\text{Var}X}} \leq \sqrt{n} \frac{\varepsilon}{\sqrt{\text{Var}X}}\right) \\ &= P\left(\sqrt{n} \frac{|\bar{X} - \mu|}{\sigma} \leq \sqrt{n} \frac{\varepsilon}{\sigma}\right) \\ &\approx \Phi\left(\sqrt{n} \frac{\varepsilon}{\sigma}\right) - \Phi\left(-\sqrt{n} \frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\sqrt{n} \frac{\varepsilon}{\sigma}\right) - 1 \end{aligned}$$

Es genügt etwa

$$\begin{aligned} \Phi^{-1}\left(1 - \frac{\delta}{2}\right) &\leq \sqrt{n} \frac{\varepsilon}{\sigma} \\ n &\geq \left(\frac{\sigma \Phi^{-1}\left(1 - \frac{\delta}{2}\right)}{\varepsilon}\right)^2 \end{aligned}$$

# Bedeutung des ZGWS in der Statistik

beim Testen

$\mu := \mathbf{E}X$ , und nehmen hier an,  $\sigma^2 = \text{Var } X$  ist bekannt. Wir testen z.B.

$$H_0 : \mu \leq \mu_0 \quad \text{gegen} \quad H_1 : \mu > \mu_0$$

Teststatistik:

$$T_n = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$$

$T_n$  klein spricht für  $H_0$ ,  $T_n$  groß gegen  $H_0$ .

Fehler 1. Art:  $H_0$  ablehnen, obwohl richtig möchte man begrenzen ( $\leq \alpha$ )

Fehler 2. Art:  $H_0$  annehmen, obwohl falsch sollte auch klein sein ( $\leq \beta$ )

# Bedeutung des ZGWS beim Testen

Fortsetzung

$$P_{\mu_0}(T_n \geq u_{1-\alpha}) \rightarrow \alpha \quad \text{nach ZGWS}$$

denn

$$P_{\mu_0}(T_n < u_{1-\alpha}) \rightarrow \Phi(u_{1-\alpha}) = 1 - \alpha$$

(wenn  $\mu < \mu_0$  so  $P_{\mu}(T_n < u_{1-\alpha}) > P_{\mu_0}(T_n < u_{1-\alpha})$ )

Wenn also  $T_n > u_{1-\alpha}$  so lehnen wir die Nullhypothese ab!

# Bedeutung des ZGWS beim Testen

## Beispiel

In der BRD gab es im Zeitraum 1970-1990 insgesamt 25 171 123 registrierte Lebendgeburten, davon waren 12 241 392 Mädchen. Berechnen Sie die ein 95% Vertrauensintervall für die Wahrscheinlichkeit einer Mädchengeburt!



# Bedeutung des ZGWS beim Testen

## Beispiel

In der BRD gab es im Zeitraum 1970-1990 insgesamt 25 171 123 registrierte Lebendgeburten, davon waren 12 241 392 Mädchen. Berechnen Sie die ein 95% Vertrauensintervall für die Wahrscheinlichkeit einer Mädchengeburt!

Das zufällige Ereignis einer Mädchengeburt wird dargestellt durch eine Bernoulli-verteilte Zufallsvariable,  $X_i \sim Bi(1, p)$ . Sei  $n = 25171123$  und

$$S_n = \sum_{i=1}^n X_i \quad \text{die zufällige Anzahl der Mädchengeburten.}$$

# Bedeutung des ZGWS beim Testen

Beispiel, Fortsetzung

Wir wissen,  $\mathbf{E}S_n = n \cdot p$  und  $\text{Var } S_n = n \cdot p \cdot (1 - p)$ .

Weiter sei  $u_{0.975}$  das 0.975-Quantil von  $\mathcal{N}(0, 1)$ ,

$$\Phi(u_{0.975}) = 0.975.$$

Nachsehen in der Tabelle liefert  $u_{0.975} \approx 1.96$ .

Aus dem Zentralen Grenzwertsatz folgt

$$P\left(\frac{|S_n - np|}{\sqrt{\text{Var } S_n}} \leq u_{0.975}\right) \approx 0.95.$$

# Bedeutung des ZGWS beim Testen

Beispiel, Fortsetzung, 2

Die folgenden Ungleichungen gelten jeweils mit Wkt. etwa 0.95:

$$|S_n - np| \leq 1.96 \cdot \sqrt{np(1-p)}$$

$$(S_n - np)^2 \leq 1.96^2 np(1-p)$$

$$n^2 p^2 - 2S_n np + S_n^2 \leq 1.96^2 np - 1.96^2 np^2$$

$$(n^2 + 1.96^2 n)p^2 - (1.96^2 n + 2nS_n)p + S_n^2 \leq 0$$

# Bedeutung des ZGWS beim Testen

Beispiel, Fortsetzung, 3

bzw. wenn wir die Schätzung

$$\hat{p} = \frac{S_n}{n}$$

für die relative Anzahl der Mädchengeburten einsetzen,

für die Randpunkte des Vertrauensintervalls

$$p_{1,2} = \frac{1}{n + 1.96^2} \left( n\hat{p} + \frac{1.96^2}{2} \pm 1.96 \sqrt{n\hat{p}(1 - \hat{p}) + \frac{1.96^2}{4}} \right).$$

Hier haben wir

$$\hat{p} = \frac{S_n}{n} = \frac{12241392}{25171123} = 0.48633$$

95%-Vertrauensintervall:  $[0.48613, 0.48652]$ .

# Bedeutung des ZGWS beim Testen

Beispiel, Fortsetzung, 4

## Fortsetzung des vorigen Beispiels

Angenommen, es würde gelten  $p = \frac{1}{2}$ . Mit welcher Wkt. würden dann höchstens 12 241 392 Mädchengeburten auftreten?

$$\begin{aligned}P(S_n \leq 12241392) &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{12241392 - np}{\sqrt{np(1-p)}}\right) \\&\approx \Phi\left(\frac{12241392 - np}{\sqrt{np(1-p)}}\right) \\&= \Phi(-137.2) \leq 3 \cdot 10^{-4091}.\end{aligned}$$

D.h. wir lehnen die Nullhypothese

$H_0 : p = \frac{1}{2}$  gegen  $H_1 : p \neq \frac{1}{2}$  ab.

# Bedeutung des ZGWS

## Beispiel

### Roulette

Beim Roulette gibt es 37 Zahlen, 18 davon sind schwarz, 18 sind rot, dazu die 0, die ist grün. Bei Setzen der richtigen Farbe gibt es den doppelten Einsatz, bei Setzen der richtigen Zahl den 36 fachen Einsatz. Zwei Spieler A und B spielen folgende Strategie: A setzt auf Farbe, B auf Zahl. Beide spielen 100 mal, und jetzen jeweils 10 Euro. Wie groß ist die Wkt., dass sie nach  $n = 100$  Spielen mindestens 40 Euro gewonnen haben?

# Roulette, Fortsetzung

Wir beschreiben die Gewinne/Verluste im  $i$ -ten Spiel durch Bernoulli-Zufallsvariablen,

$$X_i : \begin{pmatrix} 10 & -10 \\ \frac{18}{37} & \frac{19}{37} \end{pmatrix}, \quad Y_i : \begin{pmatrix} 350 & -10 \\ \frac{1}{37} & \frac{36}{37} \end{pmatrix}$$

# Roulette, Fortsetzung, 2

$$\mathbf{E}X_i = 10 \cdot \frac{18}{37} - 10 \cdot \frac{19}{37} = -\frac{10}{37} =: \mu_A$$



# Roulette, Fortsetzung, 2

$$\mathbf{E}X_i = 10 \cdot \frac{18}{37} - 10 \cdot \frac{19}{37} = -\frac{10}{37} =: \mu_A$$

$$\text{Var}X_i = \mathbf{E}X_i^2 - (\mathbf{E}X_i)^2 = 100 - \left(\frac{10}{37}\right)^2 =: \sigma_A^2 \approx 100$$

# Roulette, Fortsetzung, 2

$$\mathbf{E}X_i = 10 \cdot \frac{18}{37} - 10 \cdot \frac{19}{37} = -\frac{10}{37} =: \mu_A$$

$$\text{Var}X_i = \mathbf{E}X_i^2 - (\mathbf{E}X_i)^2 = 100 - \left(\frac{10}{37}\right)^2 =: \sigma_A^2 \approx 100$$

$$\mathbf{E}Y_i = 350 \cdot \frac{1}{37} - 10 \cdot \frac{36}{37} = -\frac{10}{37} =: \mu_B$$

$$\begin{aligned}\text{Var}Y_i &= \mathbf{E}Y_i^2 - (\mathbf{E}Y_i)^2 = 350^2 \frac{1}{37} + (-10)^2 \frac{36}{37} - \left(\frac{10}{37}\right)^2 =: \sigma_B^2 \\ &\approx 3200\end{aligned}$$

# Roulette, Fortsetzung, 3

$$P\left(\sum_{i=1}^{100} X_i \geq 40\right) = P\left(\frac{\sum_{i=1}^{100} X_i - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}} \geq \frac{40 - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}}\right)$$

# Roulette, Fortsetzung, 3

$$\begin{aligned} P\left(\sum_{i=1}^{100} X_i \geq 40\right) &= P\left(\frac{\sum_{i=1}^{100} X_i - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}} \geq \frac{40 - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}}\right) \\ &\approx 1 - \Phi\left(\frac{40 - n\mu_A}{\sqrt{n}\sigma_A}\right) \\ &= 1 - \Phi(0.67) = 0.25 \end{aligned}$$

# Roulette, Fortsetzung, 3

$$\begin{aligned}P\left(\sum_{i=1}^{100} X_i \geq 40\right) &= P\left(\frac{\sum_{i=1}^{100} X_i - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}} \geq \frac{40 - n\mu_A}{\sqrt{n}\sqrt{\text{Var}X_i}}\right) \\&\approx 1 - \Phi\left(\frac{40 - n\mu_A}{\sqrt{n}\sigma_A}\right) \\&= 1 - \Phi(0.67) = 0.25\end{aligned}$$

$$\begin{aligned}P\left(\sum_{i=1}^{100} Y_i \geq 40\right) &= P\left(\frac{\sum_{i=1}^{100} Y_i - n\mu_B}{\sqrt{n}\sqrt{\text{Var}Y_i}} \geq \frac{40 - n\mu_B}{\sqrt{n}\sqrt{\text{Var}Y_i}}\right) \\&\approx 1 - \Phi\left(\frac{40 - n\mu_B}{\sqrt{n}\sigma_B}\right) \\&= 1 - \Phi(0.12) = 0.45\end{aligned}$$

# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

# Eigenschaften von Schätzungen $\hat{\theta}$

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $X_i \sim F_\theta$  mit bekannter oder unbekannter Verteilungsfunktion  $F_\theta$  und unbekanntem Parameter  $\theta$ .

Ziel: “Näherung” für  $\theta$ .

Sei  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  eine Schätzung eines Parameters  $\theta$ , die auf  $n$  Beobachtungen beruht.

# Eigenschaften von Schätzungen $\hat{\theta}$

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $X_i \sim F_\theta$  mit bekannter oder unbekannter Verteilungsfunktion  $F_\theta$  und unbekanntem Parameter  $\theta$ .

Ziel: “Näherung” für  $\theta$ .

Sei  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  eine Schätzung eines Parameters  $\theta$ , die auf  $n$  Beobachtungen beruht.

- $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  “Konsistenz” (Minimalforderung)



# Eigenschaften von Schätzungen $\hat{\theta}$

Seien  $X_1, \dots, X_n$  Zufallsvariablen mit  $X_i \sim F_\theta$  mit bekannter oder unbekannter Verteilungsfunktion  $F_\theta$  und unbekanntem Parameter  $\theta$ .

Ziel: “Näherung” für  $\theta$ .

Sei  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  eine Schätzung eines Parameters  $\theta$ , die auf  $n$  Beobachtungen beruht.

- $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  “Konsistenz” (Minimalforderung)
- $E\hat{\theta}_n = \theta$  “Erwartungstreue”  
 $E\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$  “Asymptotische Erwartungstreue”

# Eigenschaften von Schätzungen (2)

- $\text{var } \hat{\theta}_n$  möglichst klein: “gute”, “effiziente” Schätzung

# Eigenschaften von Schätzungen (2)

- $\text{var } \hat{\theta}_n$  möglichst klein: “gute”, “effiziente” Schätzung
- wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für alle erwartungstreuen Schätzungen,  $\hat{\theta}_n$ : “optimale Schätzung”
- $\text{MSE} = \text{var } \hat{\theta}_n + \text{bias}^2 \hat{\theta}_n = \text{var } \hat{\theta}_n + (E\hat{\theta}_n - \theta)^2$   
→ minimal oder möglichst klein.

# Eigenschaften von Schätzungen (2)

- $\text{var } \hat{\theta}_n$  möglichst klein: “gute”, “effiziente” Schätzung
- wenn  $\text{var } \hat{\theta}_n$  den kleinstmöglichen Wert annimmt für alle erwartungstreuen Schätzungen,  $\hat{\theta}_n$ : “optimale Schätzung”
- $\text{MSE} = \text{var } \hat{\theta}_n + \text{bias}^2 \hat{\theta}_n = \text{var } \hat{\theta}_n + (E\hat{\theta}_n - \theta)^2$   
→ minimal oder möglichst klein.
- Eigenschaften sollten “möglichst” auch bei (kleinen) Abweichungen von der (Normal-)Verteilungsannahme gelten  
→ robuste Schätzung.

## Eigenschaften von Schätzungen (3), Anmerkungen

- $\hat{\theta} = \hat{\theta}_n$  ist hier eine Zufallsvariable die auf  $n$  zufälligen Beobachtungen beruht. Die Bezeichnung  $\hat{\theta}$  wird auch für die Realisierung der Zufallsvariablen verwendet. Was gemeint ist wird i.A. aus dem Kontext klar.
- Mit “Konsistenz” ist hier schwache Konsistenz gemeint, d.h. die Zufallsvariable  $\hat{\theta}_n$  konvergiert schwach (stochastisch) gegen  $\theta$ .
- Im Gegensatz dazu haben wir bei der “Asymptotischen Erwartungstreue” eine Konvergenz reeller Zahlen.
- Für erwartungstreue (engl. unbiased) Schätzungen gilt offenbar  $\text{bias}(\hat{\theta}) = \mathbf{E}(\hat{\theta}_n - \theta) = 0$
- Die Eigenschaften der Schätzung hängen davon ab, was man über die zugrundeliegende Verteilung  $F_\theta$  annimmt.

# Konsistenz, asymptotische Erwartungstreue

## Konsistente, nicht asymptotisch erwartungstreue Schätzung

Seien  $X_1, \dots, X_n \sim F_\theta$  mit  $P(X_i = \theta) = P(X_i = 1 + \theta) = \frac{1}{2}$ .

Sei  $Y_n = 1 - (\max(X_1, \dots, X_n) - \min(X_1, \dots, X_n))$ . Offenbar

$$P(Y_n = 1) = \frac{2}{2^n} = \frac{1}{2^{n-1}}, \quad P(Y_n = 0) = 1 - \frac{1}{2^{n-1}}.$$

Setzen  $\hat{\theta}_n = Y_n \cdot 2^{n-1} \cdot n + \min(X_1, \dots, X_n)$ .

$$\hat{\theta}_n : \begin{pmatrix} \theta & \theta + 2^{n-1} \cdot n & \theta + 1 + 2^{n-1} \cdot n \\ 1 - \frac{1}{2^{n-1}} & \frac{1}{2^n} & \frac{1}{2^n} \end{pmatrix}$$

$P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{2^{n-1}} \rightarrow 0$       Konsistenz, aber keine asy. e-treue:

$$\mathbf{E}(\hat{\theta}_n) = \theta + \frac{1}{2^n}(1 + 2 \cdot 2^{n-1} \cdot n) = \theta + n + \frac{1}{2^n} \rightarrow \infty.$$

# Konsistenz, Erwartungstreue

## nicht Konsistente, aber erwartungstreue Schätzung

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$  und  $\hat{\mu}_n := X_1$ . Es gilt

$\mathbf{E}(\hat{\mu}_n) = \mathbf{E}(X_1) = \mu$  aber  $\hat{\mu}_n$  konvergiert i.A. nicht gegen  $\mu$ .

## Momentenmethode

Man drückt den zu schätzenden Parameter durch die Momente, z.B.  $E(X)$ , aus.

Dann werden die Momente durch die entsprechenden *empirischen* Momente,  
z.B. der Erwartungswert durch  $\bar{X}$ , ersetzt.



## Momentenmethode

Man drückt den zu schätzenden Parameter durch die Momente, z.B.  $E(X)$ , aus.

Dann werden die Momente durch die entsprechenden *empirischen* Momente,  
z.B. der Erwartungswert durch  $\bar{X}$ , ersetzt.

## Maximum-Likelihood-Schätzung (ML-Schätzung)

Es wird der Schätzwert für den unbekannten Parameter ermittelt, der anhand der vorliegenden Daten, am meisten für diesen Parameter spricht (most likely).

## Kleinste-Quadrat-Schätzung (KQS)

Sei  $\theta$  der zu schätzende Parameter. Man geht aus von einem Modell, z.B.

$$Y_i = g(\theta, X_i) + \epsilon_i$$

Dann versucht man die Summe der Fehlerquadrate

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - g(\theta, X_i))^2.$$

zu minimieren (Kleinste Quadrate).

# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

## Momentenschätzung bei Normalverteilung

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ .

$$\mu = \mathbf{E}X_i \implies \hat{\mu} = \bar{X}$$

## Momentenschätzung bei Normalverteilung

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ .

$$\mu = \mathbf{E}X_i \implies \hat{\mu} = \bar{X}$$

$$\sigma^2 = \mathbf{E}(X - \mathbf{E}X)^2 \implies \hat{\sigma}^2 = \overline{(X_i - \bar{X})^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

## Momentenschätzung bei Normalverteilung

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ .

$$\mu = \mathbf{E}X_i \implies \hat{\mu} = \bar{X}$$

$$\sigma^2 = \mathbf{E}(X - \mathbf{E}X)^2 \implies \hat{\sigma}^2 = \overline{(X_i - \bar{X})^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

## Momentenschätzung bei Exponentialverteilung

Seien  $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ .

$$\lambda = \frac{1}{\mathbf{E}X_i} \implies$$

## Momentenschätzung bei Normalverteilung

Seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ .

$$\mu = \mathbf{E}X_i \implies \hat{\mu} = \bar{X}$$

$$\sigma^2 = \mathbf{E}(X - \mathbf{E}X)^2 \implies \hat{\sigma}^2 = \overline{(X_i - \bar{X})^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

## Momentenschätzung bei Exponentialverteilung

Seien  $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ .

$$\lambda = \frac{1}{\mathbf{E}X_i} \implies \hat{\lambda} = \frac{1}{\bar{X}}$$

## Momentenschätzung bei Binomialverteilung

Seien  $X_1, \dots, X_n \sim \text{Bi}(1, p)$ .

$$p = \mathbf{E}X_i \implies \hat{p} = \overline{X}$$

der relative Anteil der Realisierungen  $x_i = 1$ .



# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

Seien  $x_1, \dots, x_n$  i.i.d. Beobachtungen mit der (Zähl-)Dichte  $f(x; \theta)$

### Def. 65 (Likelihood-Funktion, Log-Likelihood Funktion)

$$L_n(\theta) = \prod_{i=1}^n f(x_i, \theta), \quad l_n(\theta) = \log(L(\theta))$$

Die Likelihood-Funktion ist die Dichte der Daten, sie wird aber als Funktion des Parameters  $\theta$  aufgefasst.

### Def. 66 (Maximum-Likelihood Schätzung)

Die Maximum-Likelihood-Schätzung ist der Wert  $\hat{\theta}$ , der  $L_n(\theta)$  maximiert.

Es ist also die Likelihood-Funktion (oder deren Logarithmus) zu maximieren.

## ML-Schätzung bei Binomialverteilung

Beobachten  $n=1000$  Jugendliche. Stichprobe  $(X_1, \dots, X_n)$

$X_i = 1$  falls Übergewicht festgestellt

$X_i = 0$  sonst.

Die Wahrscheinlichkeit, dass die beobachtete Stichprobe auftritt, wenn der Parameter  $p$  vorliegt ist

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= p^k (1-p)^{n-k}, \quad \text{wobei} \quad k = \sum_{i=1}^n x_i. \end{aligned}$$

# ML Schätzung bei Binomialverteilung

Der ML-Schätzer ist der Wert, der diese Funktion,  $L_n(p)$ ,  
Likelihood-Funktion genannt, bzgl.  $p$  maximiert. Maximieren statt  
 $L_n(p)$ :  $\log L_n(p)$  (Arg.Max. ist dasselbe).

$$\begin{aligned}\ln L_n(p) &= \ln(p^k(1-p)^{n-k}) \\ &= k \ln p + (n-k) \ln(1-p).\end{aligned}$$

# ML Schätzung bei Binomialverteilung

Der ML-Schätzer ist der Wert, der diese Funktion,  $L_n(p)$ , Likelihood-Funktion genannt, bzgl.  $p$  maximiert. Maximieren statt  $L_n(p)$ :  $\log L_n(p)$  (Arg.Max. ist dasselbe).

$$\begin{aligned}\ln L_n(p) &= \ln(p^k(1-p)^{n-k}) \\ &= k \ln p + (n-k) \ln(1-p).\end{aligned}$$

Ableiten nach  $p$  und Nullsetzen liefert:

$$\frac{k}{p} - \frac{n-k}{1-p} = 0$$

## ML Schätzung bei Binomialverteilung (2)

Die einzige Lösung ist:

$$\hat{p} = \frac{k}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Für ein relatives Extremum in  $(0,1)$  kommt nur dieser Wert in Betracht. Müssen aber noch die Likelihood-Funktion an den Rändern betrachten:

Für  $p = 0$  und  $p = 1$  wird  $\ln L(p) = -\infty$ . Also:

$$\hat{p}_{ML} = \frac{k}{n}.$$

# ML-Schätzung bei Normalverteilung

$\mu$  unbekannt,  $\sigma^2$  bekannt

Likelihood:  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ , die gemeinsame Dichtefunktion der  $X_i$ .

Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim \mathcal{N}(\mu, 1)$ .

Likelihood:

$$\begin{aligned} L_n(\mu) &= \prod_{i=1}^n f_{X_i}(x_i) \quad (\text{Unabhängigkeit}) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2/2} \end{aligned}$$

# ML-Schätzung bei Normalverteilung, 2

$$L_n(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2/2}$$

$$\ln L_n(\mu) = -n \ln(\sqrt{2\pi}) + \sum_{i=1}^n \left( -\frac{(x_i - \mu)^2}{2} \right)$$

$$\frac{\partial L_n(\mu)}{\partial \mu} = \sum_{i=1}^n (x_i - \mu)$$



## ML-Schätzung bei Normalverteilung, 2

$$L_n(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu)^2/2}$$

$$\ln L_n(\mu) = -n \ln(\sqrt{2\pi}) + \sum_{i=1}^n \left( -\frac{(x_i - \mu)^2}{2} \right)$$

$$\frac{\partial L_n(\mu)}{\partial \mu} = \sum_{i=1}^n (x_i - \mu)$$

Nullsetzen liefert die Maximum-Likelihood-Schätzung

$$\hat{\mu} = \bar{X}.$$

# Normalverteilung, $\mu$ und $\sigma^2$ unbekannt

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\begin{aligned} L_n(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(X_i - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi}^n \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right) \end{aligned}$$

# Normalverteilung, $\mu$ und $\sigma^2$ unbekannt

$$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\begin{aligned} L_n(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(X_i - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi}^n \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right) \\ &= \frac{1}{\sqrt{2\pi}^n \sigma^n} \exp\left(-\frac{nS^2}{2\sigma^2}\right) \exp\left(-\frac{n(\bar{X} - \mu)^2}{2\sigma^2}\right) \end{aligned}$$

wobei  $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

Die letzte Gleichung folgt aus:

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 = nS^2 + n(\bar{X} - \mu)^2$$

# ML-Schätzung bei Normalverteilung, Fortsetzung

Log-Likelihood:

$$\ln L(\mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{nS^2}{2\sigma^2} - \frac{n(\bar{X} - \mu)^2}{2\sigma^2}$$

# ML-Schätzung bei Normalverteilung, Fortsetzung

Log-Likelihood:

$$\ln L(\mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{nS^2}{2\sigma^2} - \frac{n(\bar{X} - \mu)^2}{2\sigma^2}$$

Lösen des Gleichungssystems

$$0 = \frac{\partial \ln L(\mu, \sigma)}{\partial \mu} = \frac{\bar{X} - \mu}{\sigma^2}$$

$$0 = \frac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{nS^2}{\sigma^3} + \frac{n(\bar{X} - \mu)^2}{\sigma^3}$$

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = S^2$$

# ML-Schätzung bei Gleichverteilung

## ML-Schätzung bei Gleichverteilung auf $(0, \theta)$

Likelihood:  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ ,

die gemeinsame Dichtefunktion der  $X_i$ .

Seien  $X_1, \dots, X_n$  unabhängig,  $X_i \sim R(0, \theta)$ , d.h.

$$f_{X_i}(x_i) = \begin{cases} \frac{1}{\theta} & \text{falls } 0 \leq x_i \leq \theta \\ 0 & \text{sonst} \end{cases}$$

# ML-Schätzung bei Gleichverteilung, 2

Likelihood:

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n f_{X_i}(x_i) && \text{(Unabhängigkeit)} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{falls } 0 \leq x_i \leq \theta \quad \forall x_i \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

# ML-Schätzung bei Gleichverteilung, 2

Likelihood:

$$\begin{aligned} L_n(\theta) &= \prod_{i=1}^n f_{X_i}(x_i) \quad (\text{Unabhängigkeit}) \\ &= \begin{cases} \frac{1}{\theta^n} & \text{falls } 0 \leq x_i \leq \theta \quad \forall x_i \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Maximal, wenn  $\theta \geq x_1, \dots, x_n$ , und wenn  $\theta$  möglichst klein, also

$$\hat{\theta} = \max(x_1, \dots, x_n).$$



# Maximum-Likelihood-Schätzung

## Gemischte Normalverteilung

Dichte ( $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p)$ ):

$$f(x; \theta) = (1 - p)\phi\left(\frac{x - \mu_1}{\sigma_1}\right) + p\phi\left(\frac{x - \mu_2}{\sigma_2}\right)$$

$X_i \sim \mathcal{N}(\mu_1, \sigma_1^2)$  mit Wahrscheinlichkeit  $(1 - p)$  und  $X_i \sim \mathcal{N}(\mu_2, \sigma_2^2)$  mit Wahrscheinlichkeit  $p$ , aber welche ist nicht bekannt.

Likelihood:

$$L(\theta) = \prod_{i=1}^n \left( (1 - p)\phi\left(\frac{x_i - \mu_1}{\sigma_1}\right) + p\phi\left(\frac{x_i - \mu_2}{\sigma_2}\right) \right)$$

Maximieren des (log-)Likelihood ist schwer.

# Lösungsverfahren

Newton-Raphson, allgemein (aber eindimensional)

Taylor-Entwicklung von  $l'(\theta) = \frac{\partial l(\theta)}{\partial \theta}$  an der Stelle  $\theta^j$

# Lösungsverfahren

Newton-Raphson, allgemein (aber eindimensional)

Taylor-Entwicklung von  $l'(\theta) = \frac{\partial l(\theta)}{\partial \theta}$  an der Stelle  $\theta^j$  und Nullsetzen

$$0 = l'(\hat{\theta}) \approx l'(\theta^j) + (\hat{\theta} - \theta^j)l''(\theta^j)$$

Lösung:

$$\hat{\theta} \approx \theta^j - \frac{l'(\theta^j)}{l''(\theta^j)}$$

Iterationsverfahren

$$\theta^{j+1} = \theta^j - \frac{l'(\theta^j)}{l''(\theta^j)}$$

# Lösungsverfahren

Newton-Raphson, allgemein (aber eindimensional)

Taylor-Entwicklung von  $l'(\theta) = \frac{\partial l(\theta)}{\partial \theta}$  an der Stelle  $\theta^j$  und Nullsetzen

$$0 = l'(\hat{\theta}) \approx l'(\theta^j) + (\hat{\theta} - \theta^j)l''(\theta^j)$$

Lösung:

$$\hat{\theta} \approx \theta^j - \frac{l'(\theta^j)}{l''(\theta^j)}$$

Iterationsverfahren

$$\theta^{j+1} = \theta^j - \frac{l'(\theta^j)}{l''(\theta^j)}$$

Verallgemeinerung auf  $k$ -Vektor

$$\theta^{j+1} = \theta^j - \mathbf{H}^{-1}l'(\theta^j) \quad \mathbf{H} : \text{Matrix der 2. Ableitungen}$$

# Lösungsverfahren, Anmerkungen

Erinnerung: Konvergenz nur wenn Startwert  $\theta^0$  in der “Nähe” der Nullstelle.

$$\theta^j \rightarrow_{j \rightarrow \infty} \hat{\theta}_n \rightarrow_{n \rightarrow \infty} \theta$$

# Eigenschaften von ML-Schätzungen

Seien Regularitätsvoraussetzungen erfüllt

- Sie sind konsistent,  $\hat{\theta}_n \rightarrow \theta$ .
- Wenn  $\hat{\theta}_n$  ML-Schätzung für  $\theta$  dann ist  $g(\hat{\theta}_n)$  ML-Schätzung für  $g(\theta)$ .
- Die ML-Schätzung ist asymptotisch normal verteilt, d.h.  
 $\hat{\theta}_n \rightarrow_D Z \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\sigma^2 = \lim_{n \rightarrow \infty} \text{var} \hat{\theta}_n$ .
- Die ML-Schätzung ist asymptotisch optimal, d.h.  
 $\text{MSE}(\hat{\theta}_n) \rightarrow_{n \rightarrow \infty} \min \text{MSE}(\hat{\theta}'_n)$ , wobei das Minimum unter allen konsistenten Schätzungen  $\hat{\theta}'_n$  genommen wird.
- Wenn für die MLS  $\hat{\theta}_n$  gilt,  $\mathbf{E}(\hat{\theta}_n) = \theta$  dann ist sie optimal, d.h. sie hat minimale Varianz unter allen Schätzungen. Diese Varianz ist aus der Cramér-Rao-Ungleichung abzulesen (s.u.)

# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

# Allgemeine Idee des EM Algorithmus

E: Expectation      M: Maximization

Iterieren fortlaufend, und berechnen abwechselnd **E** und Max.

Angenommen, die Daten  $Y$  kommen aus einer Population, für die direkte Maximierung des Log-Likelihood schwer ist.

**Idee:** Ergänzen diese Daten um zusätzliche (natürlich unbekannte) Daten  $Z$ , so dass  $f(y; \theta) = \int f(y, z; \theta) dz$  und das auf  $f(y, z; \theta)$  basierende Likelihood leicht zu maximieren ist.



# Allgemeine Idee des EM-Algorithmus (Fortsetzung)

Das interessierende komplizierte  $f(y; \theta)$  ist also **Rand**dichte des Modells mit einfacherem Likelihood.

$Y$ : beobachtete Daten,

$Z$ : versteckte (latente, fehlende) Daten.

Wenn wir die fehlenden Daten irgendwie “auffüllen” können, haben wir ein leichtes Problem.

Der EM-Algorithmus versucht, iterativ, die fehlenden Daten aufzufüllen.

## zur Illustration des EM-Algorithmus: Vereinfachung

Wir betrachten gemischte Normalverteilung und nehmen an,  $p = \frac{1}{2}$  und  $\sigma_1^2 = \sigma_2^2 = 1$ .

Direkte Maximierung der Likelihood ist schwer.

Führen latente Variablen ein,

$$Z_i = \begin{cases} 0 & \text{falls } X_i \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ 1 & \text{falls } X_i \sim \mathcal{N}(\mu_2, \sigma_2^2) \end{cases}$$

$$P(Z_i = 0) = P(Z_i = 1) = p = \frac{1}{2}$$

$$f(x_i|Z_i = 0) = \phi\left(\frac{x_i - \mu_1}{\sigma_1}\right), \quad f(x_i|Z_i = 1) = \phi\left(\frac{x_i - \mu_2}{\sigma_2}\right)$$

Damit gemischte Normal:  $f(x) = \sum_{z=0}^1 f(x, z)$

$$f(x, z) = f(z)f(x|z) = \frac{1}{2}\phi(x - \mu_1)^{1-z}\phi(x - \mu_2)^z$$

# Maximum-Likelihood-Schätzung

Gemischte Normalverteilung

vollständige Likelihood  $(x_i, z_i)$

$$L = \frac{1}{2^n} \prod_{i=1}^n \phi(x_i - \mu_1)^{1-z_i} \phi(x_i - \mu_2)^{z_i}$$

vollständige Log-Likelihood (ohne Konstante)

$$\ln L = \tilde{l} = -\frac{1}{2} \sum_{i=1}^n (1 - z_i)(x_i - \mu_1)^2 - \frac{1}{2} \sum_{i=1}^n z_i(x_i - \mu_2)^2$$

# Maximum-Likelihood-Schätzung

Gemischte Normalverteilung

Sei  $\theta^j$  wieder die  $j$ te Iterierte,  $\theta^0$  Startwert.

Bedingtes erwartetes Log-Likelihood, unter der Bedingung Daten

$\mathbf{x} = (x_1, \dots, x_n)$  und iteriertem Parametervektor  $\theta^j$

$$\mathbf{E}(\tilde{l} | \mathbf{x}, \theta^j) =$$

$$-\frac{1}{2} \sum_{i=1}^n (1 - \mathbf{E}(Z_i | \mathbf{x}, \theta^j)) (x_i - \mu_1)^2 - \frac{1}{2} \sum_{i=1}^n \mathbf{E}(Z_i | \mathbf{x}, \theta^j) (x_i - \mu_2)^2$$

ist eine Funktion von  $\theta^j$  und  $\theta$ , hier  $\theta^j = (\mu_1^j, \mu_2^j)$  und  $\theta = (\mu_1, \mu_2)$ .

Bezeichnen diese mit  $J(\theta | \theta^j)$ .

Diese Funktion hängt also vom wahren Parametervektor  $\theta$  und von der  $j$ -ten Iterierten  $\theta^j$  ab,

# Maximum-Likelihood-Schätzung

## Gemischte Normalverteilung

$Z_i$  ist binär, deshalb  $\mathbf{E}(Z_i|\mathbf{x}, \boldsymbol{\theta}^j) = P(Z_i = 1|\mathbf{x}, \boldsymbol{\theta}^j)$

Satz von Bayes:  $P(Z_i = 1|\mathbf{x}, \boldsymbol{\theta}^j) =$

$$\begin{aligned} & \frac{f(\mathbf{x}|Z_i = 1; \boldsymbol{\theta}^j)P(Z_i = 1)}{f(\mathbf{x}|Z_i = 1; \boldsymbol{\theta}^j)P(Z_i = 1) + f(\mathbf{x}|Z_i = 0; \boldsymbol{\theta}^j)P(Z_i = 0)} \\ = & \frac{\phi(x_i - \mu_2^j)\frac{1}{2}}{\phi(x_i - \mu_2^j)\frac{1}{2} + \phi(x_i - \mu_1^j)\frac{1}{2}} \\ = & \frac{\phi(x_i - \mu_2^j)}{\phi(x_i - \mu_2^j) + \phi(x_i - \mu_1^j)} =: \tau_{ij} \end{aligned}$$

# Maximum-Likelihood-Schätzung

Gemischte Normalverteilung

Damit (E-Schritt)

$$J(\boldsymbol{\theta}|\boldsymbol{\theta}^j) = -\frac{1}{2} \sum_{i=1}^n (1 - \tau_{ij})(x_i - \mu_1)^2 - \frac{1}{2} \sum_{i=1}^n \tau_{ij}(x_i - \mu_2)^2$$

Zur Maximierung von  $J$  (M-Schritt) leiten wir ab nach  $\mu_1$  und  $\mu_2$  und setzen Null. Dann

$$\begin{aligned}\hat{\mu}_2^{j+1} &= \frac{\sum_{i=1}^n \tau_{ij} x_i}{\sum_{i=1}^n \tau_{ij}} \\ \hat{\mu}_1^{j+1} &= \frac{\sum_{i=1}^n (1 - \tau_{ij}) x_i}{\sum_{i=1}^n (1 - \tau_{ij})}\end{aligned}$$

Startschätzung  $\boldsymbol{\theta}_0$ : z.B. nach Momentenmethode.

Iteration bis das Verfahren "steht".

# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

## Kleinste Quadrat Schätzung des Lageparameters

Modell:

$$Y_i = \mu + \epsilon_i$$

Die Summe der Fehlerquadrate

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (Y_i - \mu)^2.$$

minimieren: Differenzieren und Nullsetzen liefert:

$$\hat{\mu}_{KQS} = \bar{Y}.$$



# Kleinste Quadrat-Schätzung

## KQS im einfachen linearen Regressionsmodell

$$Y_i = \theta_2 + \theta_1 X_i + \epsilon_i \quad f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$$

$$\frac{\partial f}{\partial \theta_1} = X \quad \frac{\partial f}{\partial \theta_2} = 1$$

Minimiere Summe der Fehlerquadrate  $\sum_{i=1}^n (Y_i - f(X_i, \theta_1, \theta_2))^2$ :

# Kleinste Quadrat-Schätzung

## KQS im einfachen linearen Regressionsmodell

$$Y_i = \theta_2 + \theta_1 X_i + \epsilon_i \quad f(X, \theta_1, \theta_2) = \theta_1 X + \theta_2$$

$$\frac{\partial f}{\partial \theta_1} = X \quad \frac{\partial f}{\partial \theta_2} = 1$$

Minimiere Summe der Fehlerquadrate  $\sum_{i=1}^n (Y_i - f(X_i, \theta_1, \theta_2))^2$ :

$$\sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot X_i = 0$$

$$\sum_{i=1}^n (Y_i - (\theta_1 X_i + \theta_2)) \cdot 1 = 0$$

# Kleinste Quadrat-Schätzung

⇒

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \theta_2 \sum_i X_i = 0$$

$$\sum_i Y_i - \theta_1 \sum_i X_i - \theta_2 \cdot n = 0$$

Die zweite Gleichung nach  $\theta_2$  auflösen:

$$\theta_2 = \frac{1}{n} \sum_i Y_i - \theta_1 \frac{1}{n} \sum_i X_i$$

und in die erste einsetzen:

# Kleinste Quadrat-Schätzung

$$\sum_i X_i Y_i - \theta_1 \sum_i X_i^2 - \frac{1}{n} \sum_i Y_i \sum_i X_i + \theta_1 \frac{1}{n} \sum_i X_i \sum_i X_i = 0$$

$$\sum_i X_i Y_i - \frac{1}{n} \sum_i Y_i \sum_i X_i - \theta_1 \left( \sum_i X_i^2 - \frac{1}{n} \sum_i X_i \sum_i X_i \right) = 0$$

$\Rightarrow$

$$\hat{\theta}_1 = \frac{\sum_i X_i Y_i - \frac{1}{n} \sum_i X_i \sum_i Y_i}{\sum_i X_i^2 - \frac{1}{n} (\sum_i X_i)^2} = \frac{S_{XY}}{S_X^2} = \frac{S_{XY}}{S_X S_Y} \cdot \frac{S_Y}{S_X} = r_{XY} \cdot \frac{S_Y}{S_X}$$

$$\hat{\theta}_2 = \frac{1}{n} \left( \sum_i Y_i - \hat{\theta}_1 \sum_i X_i \right)$$

# Inhalt

- Einführung
- Momentenschätzung
- Maximum-Likelihood-Schätzung
- EM-Algorithmus
- Kleinste Quadrat Schätzung
- Die Cramer-Rao Ungleichung

## \*Die Cramer-Rao Ungleichung

Sei  $\theta$  ein zu schätzender Parameter einer Population mit Dichte  $f$ .

Sei  $\hat{\theta} = \theta_n$  eine erwartungstreue Schätzung von  $\theta$ .

## \*Die Cramer-Rao Ungleichung

Sei  $\theta$  ein zu schätzender Parameter einer Population mit Dichte  $f$ .

Sei  $\hat{\theta} = \theta_n$  eine erwartungstreue Schätzung von  $\theta$ .

### Cramer-Rao-Ungleichung

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)}, \quad \text{wobei}$$

$$\begin{aligned} I(f, \theta) &= \mathbf{E} \left( \frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2 \\ &= \int \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x, \theta) dx \end{aligned}$$

die sogenannte Fisher-Information ist.

# Maximum-Likelihood-Schätzung ist optimal

Seien die Regularitätsbedingungen erfüllt.

**Satz:** Existiert eine erwartungstreue Schätzung,

die die Cramér-Rao-Ungleichung annimmt, d.h.  $\text{var}(\hat{\theta}) = \frac{1}{nI(f, \theta)}$  dann ist  $\hat{\theta}$  auch ML-Schätzung.

**Bedeutung des Satzes:** Praktische Berechnung einer Schätzung mit minimaler Varianz:

- Berechne ML-Schätzung  $\hat{\theta}_{ML}$ .
- Prüfe erwartungstreue, wenn ja: Berechne  $\text{var}\hat{\theta}_{ML}$ .
- Vergleiche mit der Cramér-Rao-Schranke,  $\frac{1}{nI(f, \theta)}$ .
- wenn  $=$  so beste Schätzung gefunden
- wenn  $\neq$  dann gibt es keine bessere erwartungstreue Schätzung.



# Cramer-Rao-Ungleichung

## Beispiele

*f* normal

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ln f(x, \mu) = -\ln(\sqrt{2\pi}\sigma) - \frac{(x - \mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x - \mu}{\sigma} \cdot \frac{1}{\sigma}$$

# Cramer-Rao-Ungleichung

## Beispiele

*f* normal

$$f(x, \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\ln f(x, \mu) = -\ln(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x-\mu}{\sigma} \cdot \frac{1}{\sigma}$$

$$I(f, \mu) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 \cdot f(x, \mu) dx = \frac{1}{\sigma^2}.$$

$$\text{Also : } \operatorname{var} \hat{\mu} \geq \frac{\sigma^2}{n}, \quad \text{vgl. mit } \operatorname{var} \bar{X} = \frac{\sigma^2}{n}.$$

# Cramer-Rao-Ungleichung

## Beispiele

*f* exponential

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

# Cramer-Rao-Ungleichung

## Beispiele

### *f* exponential

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt:

$$I(f, \lambda) = \frac{1}{\lambda^2} \quad (\ddot{U}A)$$

# Cramer-Rao-Ungleichung

## Beispiele

### $f$ exponential

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt:

$$I(f, \lambda) = \frac{1}{\lambda^2} \quad (\ddot{U}A)$$

Die Cramer-Rao-Schranke ist also:

$$\frac{1}{nI(f, \lambda)} = \frac{\lambda^2}{n}.$$

Andererseits:

$$\text{var} \bar{X} = \frac{1}{n\lambda^2} = \frac{1}{nI(f, \lambda^{-1})} = \frac{1}{nI(f, \mathbf{E}X)}.$$

# Cramer-Rao-Ungleichung

## Beispiele (3)

### F Doppelsexponential (=Laplace), Skalenparameter

$$f(x, \lambda) = \frac{1}{2} \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ \lambda e^{\lambda x} & \text{falls } x < 0 \end{cases}$$

$$\ln f(x, \lambda) = -\ln 2 + \ln \lambda + \lambda x \begin{cases} -1 & \text{falls } x \geq 0 \\ 1 & \text{falls } x < 0 \end{cases}$$

$$\frac{\partial \ln f(x, \lambda)}{\partial \lambda} = \frac{1}{\lambda} - x \begin{cases} 1 & \text{falls } x \geq 0 \\ -1 & \text{falls } x < 0 \end{cases}$$

# Cramer-Rao-Ungleichung

Beispiele (3), Fortsetzung

$$\begin{aligned} I(f, \lambda) &= \frac{1}{2} \left( \int_0^{\infty} \left( \frac{1}{\lambda} - x \right)^2 \cdot \lambda e^{-\lambda x} dx + \right. \\ &\quad \left. \int_{-\infty}^0 \left( \frac{1}{\lambda} + x \right)^2 \cdot \lambda e^{\lambda x} dx \right) \\ &= \int_0^{\infty} \left( \frac{1}{\lambda^2} - \frac{2x}{\lambda} + x^2 \right) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

# Cramer-Rao-Ungleichung

Beispiele (3), Fortsetzung, 2

Cramer-Rao-Schranke

$$\frac{\lambda^2}{n} = \frac{1}{nl(f, \lambda^{-1})}.$$

Vergleichen Sie mit (ÜA)  $\mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n |X_i|\right) = \frac{1}{\lambda}$  und

$$\text{var}|\overline{X}| = \frac{1}{n^2} \sum_{i=1}^n \text{var}|X_i| = \frac{1}{\lambda^2 n} = \frac{1}{nl(f, \lambda^{-1})}.$$



# Cramer-Rao-Ungleichung

Beispiele (3a)

**F** Doppelsexponential (=Laplace), Lageparameter

$$f(x, \lambda, \mu) = \frac{1}{2} \begin{cases} \lambda e^{-\lambda(x-\mu)} & \text{falls } x \geq \mu \\ \lambda e^{\lambda(x-\mu)} & \text{falls } x < \mu \end{cases}$$

$$\ln f(x, \lambda, \mu) = -\ln 2 + \ln \lambda + \lambda(x - \mu) \begin{cases} -1 & \text{falls } x \geq 0 \\ 1 & \text{falls } x < 0 \end{cases}$$

$$\frac{\partial \ln f(x, \lambda, \mu)}{\partial \mu} = \lambda \begin{cases} 1 & \text{falls } x \geq 0 \\ -1 & \text{falls } x < 0 \end{cases}$$

$$I(f, \mu) = \lambda^2$$

## Cramer-Rao-Schranke

$$\text{var} \hat{\mu} \geq \frac{1}{n\lambda^2}$$

Die Varianz von  $\bar{X}$  ist:  $\text{var}(\bar{X}) = \frac{1}{n} 2 \frac{1}{2} \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{1}{n} \frac{2}{\lambda^2}$ .

Die *asymptotische* Varianz des Medians  $x_{(n/2)}$  ist:

$$\text{var}(X_{(n/2)}) = \frac{1}{n\lambda^2}$$

(siehe z.B. Serfling, 1980, S.79.), also halb soviel wie bei  $\bar{X}$ .

Für die *exakte* Varianz siehe Johnson, Kotz: Continuous univariate distributions 2, S.25.

# Cramer-Rao-Ungleichung

## Satz: (Cramer-Rao-Ungleichung)

Sei  $f$  Dichte der Population, und  $\hat{\theta}$  eine erwartungstreue Schätzung des Parameters  $\theta$ . Dann gilt:

$$\text{var}(\hat{\theta}) \geq \frac{1}{nI(f, \theta)},$$

wobei

$$I(f, \theta) = \mathbf{E} \left( \frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2$$

falls der Erwartungswert existiert.

# Cramer-Rao-Ungleichung

Beweis

Sei  $\mathbf{x} = (x_1, \dots, x_n)$  eine unabhängige Stichprobe und

$$L(\mathbf{x}, \theta) := \prod_{i=1}^n f(x_i, \theta)$$

die Likelihood der Stichprobe.

Offenbar gilt

$$\int_{\mathbb{R}^n} L(\mathbf{x}, \theta) d\mathbf{x} = 1.$$

und (wir setzen voraus, Differentiation und Integration dürfen vertauscht werden.)

$$\frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\mathbf{x}, \theta) d\mathbf{x} = \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} = 0$$

# Cramer-Rao-Ungleichung

Beweis, Fortsetzung (1)

Weiter gilt, da  $\hat{\theta}$  erwartungstreu,

$$\begin{aligned}\mathbf{E}\hat{\theta} &= \int_{\mathbb{R}^n} \hat{\theta} L(\mathbf{x}, \theta) d\mathbf{x} = \theta \\ \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\theta} L(\mathbf{x}, \theta) d\mathbf{x} &= \int_{\mathbb{R}^n} \hat{\theta} \underbrace{\frac{\partial L(\mathbf{x}, \theta)}{\partial \theta}} d\mathbf{x} = 1 \\ \int_{\mathbb{R}^n} \hat{\theta} \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} &= 1 \\ \mathbf{E}\left(\hat{\theta} \frac{\partial \ln L(\mathbf{X}, \theta)}{\partial \theta}\right) &= 1\end{aligned}$$

Auf den linken Term in der vorletzten Gleichung wenden wir die Cauchy-Schwarzsche Ungleichung an,

# Cramer-Rao-Ungleichung

Beweis, Fortsetzung (2)

$$\begin{aligned}
 1 &= \int_{\mathbb{R}^n} \hat{\theta} \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} - \underbrace{\theta \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x}}_{=0} \\
 &= \int_{\mathbb{R}^n} (\hat{\theta} - \theta) \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} L(\mathbf{x}, \theta) d\mathbf{x} \\
 &\leq \int_{\mathbb{R}^n} (\hat{\theta} - \theta)^2 L(\mathbf{x}, \theta) d\mathbf{x} \cdot \int_{\mathbb{R}^n} \left( \frac{\partial \ln L(\mathbf{x}, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x} \\
 &= \text{var}(\hat{\theta}) \cdot \int_{\mathbb{R}^n} \left( \frac{\partial \sum_{i=1}^n \ln f(x_i, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x} \\
 &= \text{var}(\hat{\theta}) \cdot \sum_{i=1}^n \int_{\mathbb{R}^n} \left( \frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 L(\mathbf{x}, \theta) d\mathbf{x}
 \end{aligned}$$

# Cramer-Rao-Ungleichung

Beweis, Fortsetzung (3)

Der Term auf der rechten Seite ist  $\text{var} \hat{\theta} \cdot n \cdot I(f)$ . Die zu den gemischten Summanden gehörenden Integrale sind alle Null, ( $i \neq j$ ):

$$\begin{aligned} & \int_{\mathbb{R}^2} \left( \frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right) \left( \frac{\partial \ln f(x_j, \theta)}{\partial \theta} \right) f(x_i, \theta) f(x_j, \theta) dx_i dx_j \\ &= \int_{\mathbb{R}^2} \frac{\partial f(x_i, \theta)}{\partial \theta} \frac{\partial f(x_j, \theta)}{\partial \theta} dx_i dx_j = 0. \end{aligned}$$

da alle Beobachtungen unabhängig, Differentiation und Integration vertauschbar und  $\int f(x_i, \theta) dx_i = 1$ .

# Cramer-Rao Ungleichung

**Anmerkung:** Wenn die Regularitätsvoraussetzungen nicht erfüllt sind, muss die Cramer-Rao Ungleichung nicht unbedingt gelten.

Die Dichte

$$f(x, \theta) = \begin{cases} c_{\theta} x^{-\theta} \log^{-3} x & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}$$

wobei  $c_{\theta}$  so dass  $f(x, \theta)$  Dichte ist, hat Fisher-Information  $I(f, \theta) = \infty$  für  $\theta = 1$ . (ÜA)



# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

**Komplexe Problemstellungen**, die einer analytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B.  
Travelling Salesman Problem

**Komplexe Problemstellungen**, die einer analytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B.  
Travelling Salesman Problem
- Berechnung von Integralen

**Komplexe Problemstellungen**, die einer analytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B.  
Travelling Salesman Problem
- Berechnung von Integralen
- Untersuchung des Verhaltens von Algorithmen, z.B. Sortier- und Suchverfahren

**Komplexe Problemstellungen**, die einer analytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B.  
Travelling Salesman Problem
- Berechnung von Integralen
- Untersuchung des Verhaltens von Algorithmen, z.B. Sortier- und Suchverfahren
- Theorie oft nur asymptotisch. Verhalten im Endlichen?

**Komplexe Problemstellungen**, die einer analytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B.  
Travelling Salesman Problem
- Berechnung von Integralen
- Untersuchung des Verhaltens von Algorithmen, z.B. Sortier- und Suchverfahren
- Theorie oft nur asymptotisch. Verhalten im Endlichen?
- “Wer nix kapiert, der simuliert”.

# Grundlagen der Simulation

## Einführung (2)

### **Stochastische Optimierungsverfahren**

- Mutation und Selektion
- Simulated Annealing
- Genetische Algorithmen

Allen diesen Verfahren ist gemeinsam, dass Zustandsübergänge zufällig geschehen und zwischenzeitlich auch mit gewissen (kleinen) Wahrscheinlichkeiten auch schlechtere Lösungen akzeptiert werden.

Vorteil: "Optimum" wird in Polynomialzeit gefunden.

Nachteil: "Optimum" nur mit hoher Wkt. gefunden.

# Grundlagen der Simulation

## Einführung (3)

Grundlage aller Simulationverfahren sind gleichverteilte Zufallsgrößen

$X \sim R(0, 1)$ ,

$$P(X < x) = \int_0^x dt = x,$$

d.h.  $X$  hat die Dichtefunktion:

$$f(x) = \begin{cases} 1 & \text{falls } 0 \leq x < 1 \\ 0 & \text{sonst.} \end{cases}$$

Das Kernproblem der Simulation ist deshalb die Erzeugung von Folgen unabhängiger gleichverteilter Zufallsgrößen  $X_i$ .

**Bez.:** Zufallszahlen.



# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Exakte Methoden von Hand

**Methode 1:** Es werden zufällig, gleichverteilt, die Zahlen  $0, 1, \dots, 9$  erzeugt.

$$X : \begin{pmatrix} 0 & 1 & \dots & 8 & 9 \\ \frac{1}{10} & \frac{1}{10} & \dots & \frac{1}{10} & \frac{1}{10} \end{pmatrix}.$$

Realisierung:

Es werden Karten mit den Zahlen 0 bis 9 beschriftet. Für jede Zahl ist dabei die Anzahl der Karten gleich. Nun zieht man zufällig Karten und legt sie wieder zurück. Die sich ergebende Folge von Ziffern kann man in einer Tabelle aufschreiben:

# Erzeugung von Zufallszahlen

## Exakte Methoden von Hand (2)

$$\left| \begin{array}{cccccc} 3 & 8 & 7 & 0 & 9 & 1 & \dots \\ \hline 2 & 4 & 9 & 1 & 3 & 2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array} \right|$$

Nun wählen wir zufällig Fünferblocks (es können auch Blocks von mehr Zahlen sein) aus und betrachten diese als Dezimalstellen, d.h. wir erhalten beispielsweise die Zahl 0,87091. Auf diese Weise erhalten wir Zufallszahlen auf dem Intervall  $[0, 1[$ .

# Erzeugung von Zufallszahlen

Exakte Methoden von Hand (3)

**Methode 2:** Wir erzeugen zufällig die Ziffern 0 und 1, beispielsweise mittels Münzwurf, d.h. Realisierungen der Zufallsgröße

$$X : \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Wir erhalten eine Folge  $d_1 d_2 d_3 \dots d_n \dots$  von Nullen und Einsen. Dann ermitteln wir:

$$z := \sum_{i=1}^n d_i \cdot 2^{-i} \leq 1 - \left(\frac{1}{2}\right)^n$$

Für die so erhaltene Zahl  $z$  gilt:  $0 \leq z < 1$ .

# Erzeugung von Zufallszahlen

Exakte Methoden von Hand (4)

## **Methode 3:** (4–Würfel–Spezialwürfeln)

Wir beschriften vier Würfel nach folgender Vorschrift:

1. Würfel: 0, 1, 2, 3, 4, 5

2. Würfel: 0, 6, 12, 18, 24, 30

3. Würfel: 0, 36, 72, 108, 144, 180

4. Würfel: 0, 216, 432, 648, 864, 1080

Wir werfen diese Würfel gleichzeitig und bilden die Summe der Augen. Das ergibt eine Zahl  $k$ , für die gilt:  $0 \leq k \leq 1295$ . Die Zufallsgröße  $X := \frac{k}{1295} \sim R(0, 1)$  annähernd.

# Erzeugung von Zufallszahlen

## Elektronische Erzeugung

In elektronischen Geräten fließen auch im Ruhezustand Ströme deren Spannungen zeitlich zufällig schwanken (weißes Rauschen). Nun kann man innerhalb von Zeitintervallen gleicher Länge zählen, wie oft ein kritischer Spannungswert (Schwellenwert) überschritten wird. Z.B. läßt sich bei jedem Überschreiten des Wertes ein Impuls auslösen. Diese Impulse können dann gezählt werden. Im Falle einer geraden Anzahl von Impulsen wird als Zufallsziffer eine 1 realisiert, andernfalls eine 0. Aus der resultierenden 0–1–Folge erhält man nach obigem Muster eine Zufallszahl.

# Erzeugung von Zufallszahlen

## Kongruenzmethoden

Die bisher betrachteten Verfahren sind alle sehr aufwendig (?) und deshalb praktisch schwer anwendbar. Aus diesem Grunde spielen in der Simulation nur die mathematischen Methoden (Algorithmen) zur Erzeugung von Zufallszahlen eine Rolle. Die mit diesen Methoden generierten Zufallszahlen (gewissermaßen ein Ersatz für Zufallszahlen) werden auch als Pseudozufallszahlen bezeichnet. Algorithmen, die Pseudozufallszahlen erzeugen, werden auch Zufallszahlengeneratoren genannt.

# Die multiplikative Kongruenzmethode

Wir geben die Parameter  $m, a \in \mathbb{Z}^+$  und den Startwert  $z_0 \in \mathbb{Z}^+$  vor, und definieren die Folge

$$z_{i+1} := a \cdot z_i \pmod{m}.$$

Offenbar:

$$a \cdot z_i = k \cdot m + z_{i+1}; \quad 0 \leq z_{i+1} < m \quad (k \in \mathbb{N}, i = 1, 2, \dots).$$

$$u_i = \frac{z_i}{m}, \quad (i = 1, 2, \dots)$$

ist eine Folge von Pseudozufallszahlen zwischen 0 und 1.



## Die multiplikative Kongruenzmethode (2)

Frage: Sind diese  $u_i$  annähernd eine Folge unabhängiger, auf dem Intervall  $[0, 1[$  gleichverteilter Zufallszahlen?

Frage: Geeignete Wahl der Zahlen  $a$ ,  $m$  und  $z_0$ .

### Zufallszahlengeneratoren

- RANDU (IBM):  $m = 2^{31}$ ,  $a = 2^{16} + 3$ ;
- RANDA (PRIME):  $m = 2^{31} - 1$ ,  $a = 16807$ ;
- SIMULA (CDC):  $m = 2^{59}$ ,  $a = 5^{11}$ .
- SAS 8, SAS 9.4 (ranuni-Funktion):  $m = 2^{31} - 1$ ,  $a = 397204094$ .

# Verallgemeinerung: Die lineare Kongruenzmethode

Wir geben wieder Werte vor:  $m, a, r, z_0 \in \mathbb{Z}^+$  und definieren die Folge

$$z_{i+1} = (a \cdot z_i + r) \pmod{m}$$

und die Folge von Zufallszahlen ist

$$u_i := \frac{z_i}{m} \quad (i \in \mathbb{N}).$$

Turbo-Pascal 5.0:

$$z_{n+1} = 134775813z_n + 1 \pmod{2^{32}}$$

Java, C++ (Jonischkat, Klostermann, Algorithmus der Woche, 2006)

$$z_{n+1} = 25149003917z_n + 11 \pmod{2^{48}} \quad \text{hat volle Periode}$$

# Die mehrfache lineare Kongruenzmethode

Parameter:  $m, a_1, \dots, a_k, r \in \mathbb{Z}^+$  Startwerte :  $z_0, \dots, z_{(k-1)} \in \mathbb{Z}^+$ .

Wir definieren die Folge für  $n > (k - 1)$

$$z_n = \left( \sum_{l=1}^k a_l \cdot z_{n-l} + r \right) \pmod{m}.$$

Die Zufallszahlenfolge ist dann wieder

$$u_n := \frac{z_n}{m}.$$

# Wünschenswerte Eigenschaften von Pseudozufallszahlen

- Einfacher Algorithmus, wenig Rechenzeit.

# Wünschenswerte Eigenschaften von Pseudozufallszahlen

- Einfacher Algorithmus, wenig Rechenzeit.
  - möglichst viele verschiedene Zufallszahlen
- ⇒ lange Periode.
- ⇒  $m$  möglichst groß (etwa in der Nähe der oberen Grenze des INTEGER-Bereichs)

# Wünschenswerte Eigenschaften von Pseudozufallszahlen

- Einfacher Algorithmus, wenig Rechenzeit.
- möglichst viele verschiedene Zufallszahlen  
⇒ lange Periode.  
⇒  $m$  möglichst groß (etwa in der Nähe der oberen Grenze des INTEGER-Bereichs)
- $k$ -Tupel  $(U_1, \dots, U_k) \sim R(0, 1)^k$ ,  $k \leq 10$   
⇒ Test auf Gleichverteilung.

# Wünschenswerte Eigenschaften von Pseudozufallszahlen

- Einfacher Algorithmus, wenig Rechenzeit.
- möglichst viele verschiedene Zufallszahlen  
⇒ lange Periode.  
⇒  $m$  möglichst groß (etwa in der Nähe der oberen Grenze des INTEGER-Bereichs)
- $k$ -Tupel  $(U_1, \dots, U_k) \sim R(0, 1)^k$ ,  $k \leq 10$   
⇒ Test auf Gleichverteilung.
- “Unabhängigkeit” ⇒ Test auf Autokorrelation  
Plot der Punkte  $(U_i, U_{i+k})$ ,  $k = 1, 2, \dots$   
es sollten keine Muster zu erkennen sein.

# Multiplikative Generatoren (1)

## Ein schlechter Generator

Wir wählen  $m = 2^4$ ,  $a = 11$ ,  $z_0 = 3$ .

$$z_1 = 11 \cdot 3 \pmod{16} = 1$$

$$z_2 = 11 \cdot 1 \pmod{16} = 11$$

$$z_3 = 11 \cdot 11 \pmod{16} = 9$$

$$z_4 = 11 \cdot 9 \pmod{16} = 3$$

Dann gilt:  $z_5 = z_1$  und die Folge wiederholt sich.

Periodenlänge = 4 statt gleich 16 (wie theoretisch möglich)



# Multiplikative Generatoren (2)

$$z_{i+1} = a \cdot z_i \pmod{m}$$

## Satz

Wenn  $m = 2^k$ ,  $a \bmod 8 \in \{3, 5\}$ ,  $z_0$  ungerade und  $r = 0$  sind, so hat die multiplikative Kongruenzmethode die maximal mögliche Periodenlänge  $2^{k-2}$ .

In allen anderen Fällen gilt, daß die Periodenlänge kleiner als  $2^{k-2}$  ist.

# Lineare Generatoren

$$z_{i+1} = a \cdot z_i + r \pmod{m}$$

## Satz

Die lineare Kongruenzmethode besitzt genau dann die volle Periodenlänge  $m$ , falls die folgenden Bedingungen erfüllt sind:

- 1  $\text{ggT}(r, m) = 1$  ( $\text{ggT}(0, m) := m$ ),  
d.h. für  $r = 0$  ist diese Bedingung nicht erfüllt;
- 2  $a \bmod p = 1$ , für alle Primfaktoren  $p$  von  $m$ ;
- 3  $a \bmod 4 = 1$ , falls  $m$  ein Vielfaches von 4 ist.

# Beurteilung der Generatoren

## Punkteplots in $\mathbb{R}^2$

Bilden wir Paare  $(u_1, u_2)$ ,  $(u_3, u_4)$ ,  $(u_5, u_6)$ , usw. aufeinanderfolgender Zufallszahlen und tragen sie in das Einheitsquadrat ein. Es entsteht ein (zweidimensionales) Scatterplot von Punkten. Die Pseudozufallszahlen sind evtl. dann akzeptabel, wenn sich hier eine gleichmäßige Verteilung ergibt und keine Struktur erkennbar ist. Entstehen dagegen (Linien)muster, so ist der Zufallszahlengenerator schlecht.

Verallgemeinerung auf  $k$ -Tupel möglich.

# Punkteplots in $\mathbb{R}^k$

Es sei  $\{z_i\}_{i \in \mathbb{N}}$  eine Folge von Werten, die mit der multiplikativen Kongruenzmethode mit

$m = 2^t$ ,  $a = 5 \pmod{8}$  und  $z_0 = 1 \pmod{4}$

ermittelt wurden, d.h.:

$$z_{i+1} = a \cdot z_i \pmod{2^t}.$$

$$u_i = \frac{z_i}{2^t}.$$

Wir bilden nun  $k$ -Tupel von aufeinanderfolgenden

Pseudozufallszahlen:

$$\mathbf{u}_{(k)} = (u_l, \dots, u_{l+k-1}) = \left( \frac{z_l}{2^t}, \dots, \frac{z_{l+k-1}}{2^t} \right).$$

# Gitter von Zufallszahlen (1)

Sei  $u_0$  die erste Zufallszahl. Die ersten  $k$  Zufallszahlen haben die Form

$$u_0 \cdot ((1, a, \dots, a^{k-1})(\bmod m))/m = u_0 \cdot \frac{\mathbf{b}_1}{4} + g,$$

wobei

$$\mathbf{b}_1 = \frac{1}{2^{t-2}} \cdot (1, a, \dots, a^{k-1})$$

und  $g \in G$  ein geeigneter Vektor ist, so dass die  $u_l, l = 1, \dots, k$ , auch im Intervall  $(0, 1)$  liegen.

Anstelle der ersten kann mit einer beliebigen Zufallszahl begonnen werden.

## Gitter von Zufallszahlen (2)

Für diese  $k$ -Tupel von Pseudozufallszahlen gilt:

$$\mathbf{u}_{(k)} \in \left( \frac{1}{4} \cdot b_1 + G \right) \cap [0, 1]^k.$$

Dabei ist:

$$G = \left\{ \sum_{i=1}^k q_i \cdot \mathbf{b}_i : q_1, \dots, q_k \in \mathbb{Z} \right\}$$
$$\mathbf{b}_1^T = \frac{1}{2^{t-2}} \cdot \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{k-1} \end{pmatrix}, \mathbf{b}_2 = \mathbf{e}_2, \dots, \mathbf{b}_k = \mathbf{e}_k.$$

# Ein alter Zufallszahlengenerator

RANDU  $m = 2^{31}, \quad a = 2^{16} + 3, \quad r = 0$

$$X_{i+2} = (2^{16} + 3)X_{i+1} + c_1 2^{31}$$

# Ein alter Zufallszahlengenerator

RANDU  $m = 2^{31}, \quad a = 2^{16} + 3, \quad r = 0$

$$\begin{aligned} X_{i+2} &= (2^{16} + 3)X_{i+1} + c_1 2^{31} \\ &= (2^{16} + 3)^2 X_i + c_2 2^{31} (2^{16} + 3) + c_1 2^{31} \end{aligned}$$



# Ein alter Zufallszahlengenerator

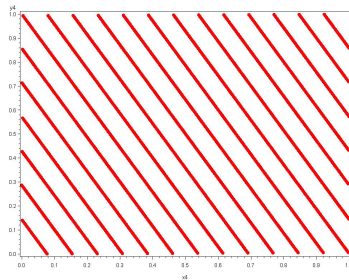
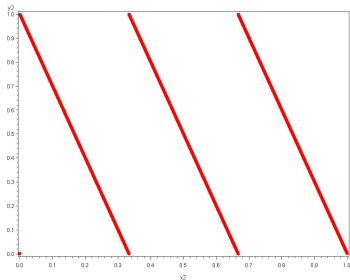
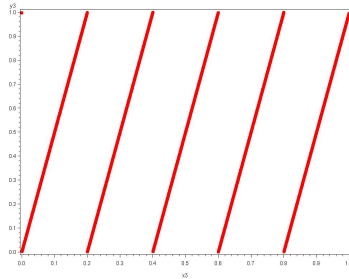
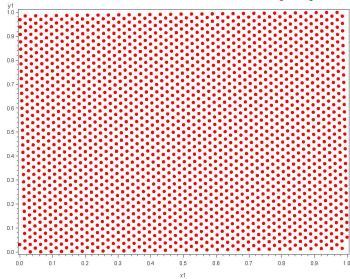
RANDU  $m = 2^{31}$ ,  $a = 2^{16} + 3$ ,  $r = 0$

$$\begin{aligned}X_{i+2} &= (2^{16} + 3)X_{i+1} + c_1 2^{31} \\&= (2^{16} + 3)^2 X_i + c_2 2^{31} (2^{16} + 3) + c_1 2^{31} \\&= (6 \cdot 2^{16} + 9)X_i + 2^{31} (2X_i + (2^{16} + 3)c_2 + c_1) \\&= 6(2^{16} + 3)X_i - 9X_i + c_3 2^{31} \\&= 6X_{i+1} - 9X_i + c_4 2^{31}\end{aligned}$$

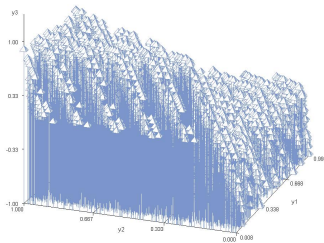
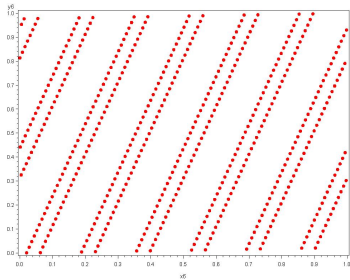
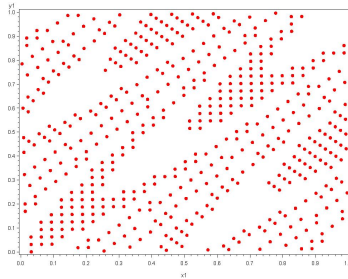
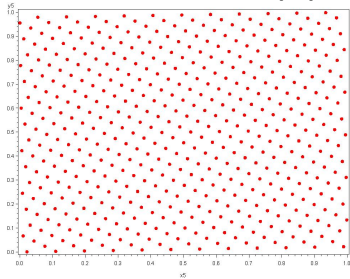
$c_i \in \mathbb{Z}, i = 1, \dots, 4.$  Daraus folgt:

$$X_{i+2} - 6X_{i+1} + 9X_i \in \mathbb{Z}.$$

# Beispielmuster (1)



# Beispielmuster (2)



# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

Def. 67 Ein Test ist eine Entscheidungsvorschrift, die über die Akzeptanz genau einer von zwei alternativen Hypothesen entscheidet.

### Analogie zur Qualitätskontrolle

Ein Käufer soll anhand einer Stichprobe entscheiden, ob er einen Warenbestand kauft oder nicht. Wir haben zwei Hypothesen, die Null- und die Alternativhypothese:

$H_0$ : Die Ware ist in Ordnung, z.B. der Ausschußanteil  $p$  ist kleiner oder gleich 2%.

$H_A$ : Die Ware ist schlecht, d.h.  $p > 2\%$ .

# Analogie zur Qualitätskontrolle

Der Kunde führt nun bei  $n$  Produkten eine Kontrolle durch,

$$x_i = \begin{cases} 0 & , \text{ falls das Produkt } i \text{ gut ist,} \\ 1 & , \text{ falls das Produkt } i \text{ schlecht ist.} \end{cases}$$

Dann ist  $z = \sum_{i=1}^n x_i$  die Anzahl der fehlerhaften Produkte, die der Kunde gefunden hat. Nun wird vor dem Test ein kritischer Wert  $z_\alpha$  festgelegt

- Ist  $z > z_\alpha$ , so wird die Hypothese  $H_0$  abgelehnt;
- Ist  $z \leq z_\alpha$ , so wird die Hypothese  $H_0$  für richtig befunden.

# Statistische Tests von Pseudozufallszahlen

## Fehlerwahrscheinlichkeiten

- 1  $P(Z > z_\alpha | H_0 \text{ ist wahr})$  – die Wahrscheinlichkeit also, dass der Käufer die Ware für schlecht befindet und ablehnt, obwohl sie doch in Ordnung ist. Diese Wahrscheinlichkeit spiegelt das „Risiko des Produzenten“ wider.
- 2  $P(Z \leq z_\alpha | H_0 \text{ ist falsch})$  – die Wahrscheinlichkeit also, daß der Käufer die Ware nimmt, obwohl ihre Qualität stark zu wünschen übrig lässt. Diese Wahrscheinlichkeit spiegelt das „Risiko des Käufers“ wider.

# Statistische Tests von Pseudozufallszahlen

Die Entscheidung für  $H_A$  oder für  $H_0$  wird anhand einer Teststatistik

$$Z = Z(x_1, \dots, x_n)$$

gefällt. Falls  $Z \in K$  (kritischen Bereich, Ablehnungsbereich), dann wird  $H_0$  abgelehnt, sonst nicht.

Bei jeder dieser Entscheidungen kann man Fehlentscheidungen treffen:

- Entscheidung für  $H_A$  obwohl  $H_0$  richtig ist: Fehler 1.Art
- Entscheidung für  $H_0$  obwohl  $H_A$  richtig ist: Fehler 2.Art



# (Fehl-)Entscheidungstabelle

|               | Entscheidung<br>für $H_0$                  | Entscheidung<br>für $H_A$              |
|---------------|--------------------------------------------|----------------------------------------|
| $H_0$ richtig | richtig, Sicher-<br>heitswkt. $1 - \alpha$ | Fehler 1. Art<br>Fehlerwkt. $\alpha$ . |
| $H_A$ richtig | Fehler 2.Art<br>Fehlerwkt. $1-\beta$       | richtig,<br>Güte $\beta$               |

**Bem.:** Entscheidung für  $H_0$  heißt nicht notwendig, dass  $H_0$  richtig ist.

# Statistische Tests von Pseudozufallszahlen

Der Parameter  $\alpha := P(Z > Z_\alpha | H_0 \text{ ist wahr})$  ist meist vorgegeben. Übliche Werte für  $\alpha$  sind 0.05 oder 0.01. Gesucht ist eine Testvorschrift, die zur Minimierung des „Risikos des Käufers“ führt.

## Anwendung auf Pseudozufallszahlen

zu testen:

- Gleichverteilung der Pseudozufallszahlen über dem Intervall  $[0, 1[$ ;
- Unabhängigkeit der Pseudozufallszahlen.

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Der $\chi^2$ -Anpassungs-Test

$\chi^2$ -Verteilung, Erinnerung,  $Y \sim \chi_k^2$

$Y_1, \dots, Y_k$  seien unabhängig, identisch verteilte Zufallszahlen mit  $Y_i \sim \mathcal{N}(0, 1)$ .

Dann heißt die Zufallsvariable  $Y$  mit

$$Y = \sum_{i=1}^k Y_i^2$$

$\chi^2$ -verteilt mit  $k$  Freiheitsgraden.

## Der $\chi^2$ -Anpassungs-Test (2)

Es seien jetzt  $X_i$  ( $i = 1, \dots, n$ ) beliebige unabhängig und identisch verteilte Zufallsgrößen

$$B = [0, 1)$$

$$A_j = \left[ \frac{j-1}{k}, \frac{j}{k} \right) \quad n \geq 5k$$

$$p_j = P(X \in A_j) = \frac{1}{k}$$

Wir testen

$$H_0 : p_j = \frac{1}{k} \quad j = 1, \dots, k$$

$$H_A : p_j \neq \frac{1}{k} \quad \text{für ein } j$$

## Der $\chi^2$ -Anpassungs-Test (3)

Dazu bilden wir

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j} \quad n_j = \#\{X_i : X_i \in A_j\}$$

Wenn  $H_0$  zutrifft, gilt für große  $n$  dann approximativ,

$$\chi^2 \sim \chi_{k-1}^2.$$

Wenn  $H_0$  richtig ist, gilt wegen dem schwachen Gesetz großer Zahlen

$$n_j \approx n \cdot p_j$$

Offenbar,  $0 \leq \chi^2$ .

Wenn  $\chi^2 \leq c_\alpha$  wollen wir Hypothese  $H_0$  annehmen, wenn  $\chi^2 > c_\alpha$  lehnen wir diese ab.

## Der $\chi^2$ -Anpassungs-Test (4)

$c_\alpha$  wird wie folgt festgelegt:

$$P(\chi^2 > c_\alpha | H_0 \text{ richtig}) = \alpha$$

ist die Wahrscheinlichkeit (bzw. das Risiko) dafür, das trotz “guter” Verteilung (Gleichverteilung) der Zufallszahlen wir die Hypothese  $H_0$  ablehnen, d.h. die Nicht-Gleichverteilung annehmen.

# Auf der empirischen Verteilungsfunktion beruhende Tests (EDF-Tests, allgemein)

Erinnerung (empirische Verteilungsfunktion):

Seien  $X_1, \dots, X_n$  unabh. Beobachtungen,

$X_{(1)} \leq \dots \leq X_{(n)}$  die geordneten Beob. Die Funktion

$$F_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{i}{n} & X_{(i)} \leq x < X_{(i+1)} \\ 1 & X_{(n)} \leq x \end{cases} \quad i = 1 \dots n$$

heißt empirische Verteilungsfunktion.

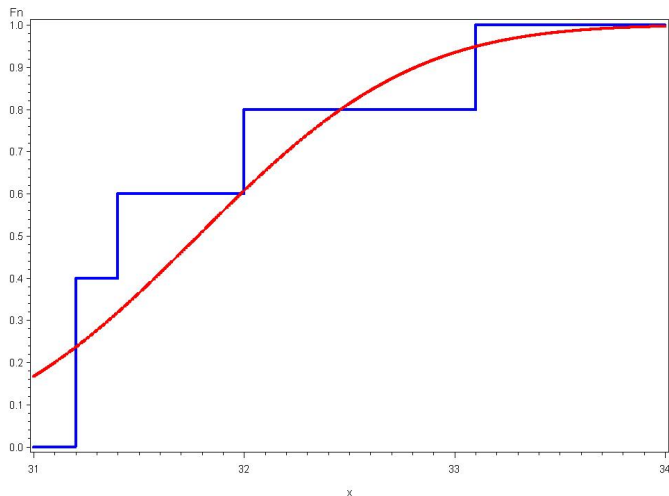
Satz v. Glivenko-Cantelli:  $F_n(x) \rightarrow F(x)$ .



# Der Satz von GLIVENKO–CANTELLI

Wiederholung

## Empirische Verteilungsfunktion



# Drei EDF-Tests

## Kolmogorov-Smirnov-Test

$$D = \sup_x |F_n(x) - F_0(x)|$$

## Cramer-von Mises-Test\*

$$W^2 = n \int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 dF_0(x)$$

## Anderson-Darling-Test\*

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x)$$

hier:  $F_0(x) = x$ .

# EDF-Tests, nur zur Info.

## Modifikationen für endliche Stichproben

$$D: D \cdot (\sqrt{n} - 0.01 + 0.85/\sqrt{n})$$

$$A^2: AD^2 \cdot (1.0 + 0.75/n + 2.25/n^2)$$

$$W^2: CM^2 \cdot (1.0 + 0.5/n)$$

Kritische Werte

$W^2$ : D'Agostino, Stephens (1986), S. 123.

$A^2$ : Crawford Moss u.a. (1990)

# Der Kolmogorov–Smirnov–Test

Erinnerung:

$$\lim_{n \rightarrow \infty} D_n = \lim_{n \rightarrow \infty} \sup_x |F_n(x) - x| = 0$$

## Satz (KOLMOGOROV–SMIRNOV)

Es gilt für  $x > 0$ :

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\sqrt{n} \cdot D_n < x) &= 1 + 2 \sum_{i=1}^{\infty} (-1)^i \cdot e^{-2 \cdot i^2 \cdot x^2} \\ &=: Q(x) \end{aligned}$$

$Q(x)$  ist die Verteilungsfunktion der Kolmogorov-Verteilung.

# Der Kolmogorov–Smirnov–Test

## Praktische Durchführung

- 1 Die Pseudozufallszahlen werden der Größe nach geordnet,

$$u_{(1)} < u_{(2)} < \dots < u_{(n)}.$$

# Der Kolmogorov–Smirnov–Test

## Praktische Durchführung

- 1 Die Pseudozufallszahlen werden der Größe nach geordnet,

$$u_{(1)} < u_{(2)} < \dots < u_{(n)}.$$

- 2 EDF:  $F_n(x) = \frac{\#\{u_i: u_i \leq x, 0 \leq x < 1\}}{n}.$

# Der Kolmogorov–Smirnov–Test

## Praktische Durchführung

- 1 Die Pseudozufallszahlen werden der Größe nach geordnet,

$$u_{(1)} < u_{(2)} < \dots < u_{(n)}.$$

- 2 EDF:  $F_n(x) = \frac{\#\{u_i: u_i \leq x, 0 \leq x < 1\}}{n}.$

- 3 Wir ermitteln die Zahl

$$D_n := \sup_x |F_n(x) - x| = \max \left\{ \max_{1 \leq i \leq n} a_i, \max_{1 \leq i \leq n} b_i \right\},$$

$$a_i := \left| u_{(i)} - \frac{i}{n} \right|, \quad b_i := \left| u_{(i)} - \frac{i-1}{n} \right|.$$

# Der Kolmogorov–Smirnov–Test

## Praktische Durchführung

- 1 Die Pseudozufallszahlen werden der Größe nach geordnet,

$$u_{(1)} < u_{(2)} < \dots < u_{(n)}.$$

- 2 EDF:  $F_n(x) = \frac{\#\{u_i: u_i \leq x, 0 \leq x < 1\}}{n}.$

- 3 Wir ermitteln die Zahl

$$D_n := \sup_x |F_n(x) - x| = \max \left\{ \max_{1 \leq i \leq n} a_i, \max_{1 \leq i \leq n} b_i \right\},$$

$$a_i := \left| u_{(i)} - \frac{i}{n} \right|, \quad b_i := \left| u_{(i)} - \frac{i-1}{n} \right|.$$

- 4  $c_\alpha$ :  $1 - \alpha$ -Quantil der Kolmogorov-Verteilung.

$$\sqrt{n} \cdot D_n > c_\alpha \implies \text{Ablehnung der Hypothese } H_0$$

$$\sqrt{n} \cdot D_n \leq c_\alpha \implies \text{Annahme der Hypothese } H_0$$



# Der Kolmogorov–Smirnov–Test (2)

Dabei ist

$$\alpha = P(H \text{ abgelehnt} | H_0) = P(\sqrt{n} \cdot D_n > c_\alpha | H_0).$$

$$\text{D.h.} \quad Q(c_\alpha) = \lim_{n \rightarrow \infty} P(\sqrt{n} \cdot D_n < c_\alpha) = 1 - \alpha.$$

| $\alpha$ | $c_\alpha$ (gerundet) |
|----------|-----------------------|
| 0.01     | 1.63                  |
| 0.05     | 1.36                  |
| 0.1      | 1.22                  |

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Der Run-Test

## Def. 68 (Run)

Jeder Teilabschnitt einer Folge unabhängiger, identisch verteilter Zufallszahlen, in dem die Zufallszahlen aufsteigend geordnet sind.

Wir teilen eine Folge in Runs ein:

|                |                       |     |  |  |      |  |     |  |  |  |    |  |
|----------------|-----------------------|-----|--|--|------|--|-----|--|--|--|----|--|
| Folge          | 2 1 2 3 2 4 1 7 8 9 0 |     |  |  |      |  |     |  |  |  |    |  |
| Run            | I.                    | II. |  |  | III. |  | IV. |  |  |  | V. |  |
| Länge des Runs | 1                     | 3   |  |  | 2    |  | 4   |  |  |  | 1  |  |

# Run-Test (2)

## Satz

Es sei  $u_1, \dots, u_n$  eine Folge unabhängiger Zufallsgrößen mit  $u_i \sim U(0, 1)$  ( $i = 1, \dots, n$ ). Dann gilt für die zufällige Länge  $R$  eines Runs:

$$P(R = r) = \frac{r}{(r+1)!}.$$

Wir beschreiben  $R$  also durch:

$$R : \begin{pmatrix} 1 & 2 & \dots & r & \dots \\ \frac{1}{2} & \frac{1}{3} & \dots & \frac{r}{(r+1)!} & \dots \end{pmatrix}.$$

# Run-Test (3)

Beweis des Satzes

Wegen der Unabhängigkeit und der identischen Verteilung genügt es, die ersten  $r + 1$  Zufallsvariablen zu betrachten. Es gilt:

$$P(R = r) = P(U_1 \leq \dots \leq U_r > U_{r+1})$$

# Run-Test (3)

Beweis des Satzes

Wegen der Unabhängigkeit und der identischen Verteilung genügt es, die ersten  $r + 1$  Zufallsvariablen zu betrachten. Es gilt:

$$\begin{aligned}P(R = r) &= P(U_1 \leq \dots \leq U_r > U_{r+1}) \\&= P(U_1 \leq \dots \leq U_r) - P(U_1 \leq \dots \leq U_r \leq U_{r+1}) \\&= \frac{1}{r!} - \frac{1}{(r+1)!} = \frac{r}{(r+1)!}\end{aligned}$$

# Run-Test (3)

Beweis des Satzes

Wegen der Unabhängigkeit und der identischen Verteilung genügt es, die ersten  $r + 1$  Zufallsvariablen zu betrachten. Es gilt:

$$\begin{aligned}P(R = r) &= P(U_1 \leq \dots \leq U_r > U_{r+1}) \\&= P(U_1 \leq \dots \leq U_r) - P(U_1 \leq \dots \leq U_r \leq U_{r+1}) \\&= \frac{1}{r!} - \frac{1}{(r+1)!} = \frac{r}{(r+1)!}\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{\infty} P(R = i) &= \sum_{i=1}^{\infty} \left( \frac{1}{i!} - \frac{1}{(i+1)!} \right) = \sum_{i=1}^{\infty} \frac{1}{i!} - \sum_{i=1}^{\infty} \frac{1}{(i+1)!} \\&= \left( \sum_{i=0}^{\infty} \frac{1}{i!} - 1 \right) - \left( \sum_{i=0}^{\infty} \frac{1}{(i+1)!} - 1 \right) = 1.\end{aligned}$$

## Run-Test (4)

Seien  $u_1, \dots, u_n$  Pseudozufallszahlen. Wir testen

$H_0 :$          $u_1, \dots, u_n$    sind unabhängig        gegen

$H_1 :$          $u_1, \dots, u_n$    sind abhängig.

$R_1, \dots, R_m$  sei die Folge der Längen der auftretenden Runs.

Diese Folgen sind jedoch nicht unabhängig (auch nicht, wenn  $X_i$  stochastisch unabhängig sind) Deshalb streichen wir nach jedem Run die nächste Zufallszahl, und berechnen die nachfolgenden Runlängen von neuem.



## Run-Test (5)

Es entstehen die Größen  $R_1^*, \dots, R_m^*$ , die unabhängig sind  
(Mathar/Pfeiffer, Lemma 6.2.2)

Formal sieht das folgendermaßen aus:

Seien die  $S_i$  die Stellen an denen ein Run zuende ist,

$$S_1 = \inf\{n \in \mathbb{N}: u_{n+1} < u_n\}$$

$$S_2 = \inf\{n \in \mathbb{N}: n > S_1 + 1, u_{n+1} < u_n\}$$

$$\vdots$$

$$S_{k+1} = \inf\{n \in \mathbb{N}: n > S_k + 1, u_{n+1} < u_n\}$$

## Run-Test (6)

Dann definieren wir:

$$\begin{aligned}R_1^* &:= S_1 \\R_2^* &:= S_2 - S_1 - 1 \\&\vdots \\R_{k+1}^* &:= S_{k+1} - S_k - 1\end{aligned}$$

Wenn nun die Hypothese  $H_0$  gilt, dann ist:

$$P(R^* = r) = \frac{r}{(r+1)!},$$

und die  $R_i^*$  ( $i = 1, \dots, m$ ) sind unabhängig.

**Run-Test:** Anpassungstest auf diese Verteilung

# Run-Test (7)

Teilen  $\mathbb{Z}^+$  in  $k$  disjunkte Teilintervalle auf:

$$[i_1 + 1, i_2], [i_2 + 1, i_3], \dots, [i_k + 1, \infty)$$

$$p_j^* = \sum_{m=i_j+1}^{i_{j+1}} P(R^* = m) = P(i_j + 1 \leq R^* \leq i_{j+1})$$

$$n_j = \#_{i=1, \dots, m} \{R_i^* : i_j + 1 \leq R_i^* \leq i_{j+1}\}$$

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - mp_j^*)^2}{mp_j^*} \sim \chi_{k-1}^2$$

Falls  $\chi^2 >$  kritischer Wert, lehnen wir die Unabhängigkeitshypothese ab.

# Run-Test (8)

Gesamtumfang der zu erzeugenden Zufallszahlen sollte  $\geq 4000$  sein.  
Wir haben hier einen Anpassungstest auf eine gegebene diskrete Verteilung gemacht.

$\chi^2$ -Anpassungstests (auf eine stetige Verteilung, hier Gleichverteilung) sollten, u.a. wegen der Willkür der Klasseneinteilung mit Vorsicht betrachtet werden.

# Autokorrelationstest

Sei  $U_1, \dots, U_n$  eine Folge von zufälligen Variablen. Für alle  $m$  können wir nun bilden:

$$\rho_m(k) = \frac{\text{COV}(U_m, U_{m+k})}{\sigma_{U_m} \sigma_{U_{m+k}}}$$

wobei  $1 \leq k \leq \frac{n}{2}$  Wenn  $U_1, \dots, U_n$  identisch verteilt so  $\sigma_{U_j} = \sigma \quad \forall j$   
und

$$\text{COV}(U_m, U_{m+k}) = \text{COV}(U_1, U_{k+1})$$

Autokorrelation  $k$ -ter Ordnung

$$\sigma_m(k) = \rho(k) = \frac{E(U_m \cdot U_{m+k}) - (EU_m)^2}{\sigma^2}$$

$$\forall m, \quad k = 1, \dots, \left\lfloor \frac{n}{2} \right\rfloor.$$

# Autokorrelationstest (2)

Sei  $u_1, \dots, u_n$  eine Folge von Realisierungen.

$$\hat{\rho}(k) = \frac{\frac{1}{n-k} \sum_{i=1}^{n-k} u_i \cdot u_{i+k} - \left( \frac{1}{n-k} \sum_{i=1}^{n-k} u_i \right)^2}{\frac{1}{n-k} \sum_{i=1}^{n-k} u_i^2 - \left( \frac{1}{n-k} \sum_{i=1}^{n-k} u_i \right)^2}$$

ist die *empirische Autokorrelation*  $k$ -ter Ordnung.

## Autokorrelationstest (3)

$\rho(k)$  ist die Pearson-Korrelation zwischen  $U_i$  und  $U_{i+k}$ .

Offenbar,  $\rho(k) = 0 \quad \forall k \geq 1$ , wenn die Zufallszahlen keine Autokorrelation besitzen. Für die  $u_1, \dots, u_n$  sollte dann gelten:  $\hat{\rho}(k) \approx 0$ .

Ersetzen wir die

$U_i$  durch ihre Ränge  $R_1, \dots, R_n$  und die

$U_{i+k}$  durch ihre Ränge  $S_1, \dots, S_n$

dann erhalten wir den Spearman-Rang-Korrelationskoeffizient  $r_S$ .

## Autokorrelationstest (4)

Es gilt asymptotisch (wenn  $H_0$  richtig ist)

$$r_S \sim \mathcal{N}\left(0, \frac{1}{n-1}\right).$$

Die Nullhypothese

$H_0$ : keine Autokorrelation

wird also abgelehnt, wenn

$$\sqrt{n-1}|r_S| \geq z_{1-\alpha/2}$$

$z_{1-\alpha/2}$ :  $1 - \alpha/2$ -Quantil der Standard-Normalverteilung,

$$z_{0.975} = 1.96.$$



# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# diskrete Zufallsvariablen, Intervallmethode

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}.$$

Zerlegen das Intervall  $[0, 1]$  in Teilintervalle  $I_j$ ,

$$I_j = \left[ \sum_{k=0}^{j-1} p_k, \sum_{k=0}^j p_k \right], \quad (p_0 = 0)$$

Sei  $u$  eine Pseudozufallszahl. Wir setzen

$$X = x_j \quad \text{falls} \quad u \in I_j$$

# Erzeugung stetiger Zufallsvariablen

## Quantilmethode

Es sei  $U \sim R(0, 1)$ . Wir betrachten die Transformation

$$X := \varphi(U),$$

wobei  $\varphi$  monoton wachsend sei. Die Zufallsgröße  $X$  ist ebenfalls stetig, und für ihre Dichte gilt (nach der Transformationsformel für Dichten)

$$f_X(x) = h\left(\varphi^{-1}(x)\right) \cdot \left| \frac{d\varphi^{-1}(x)}{dx} \right|.$$

Wir wählen nun  $\varphi := F^{-1}$ . Dann erhalten wir:

$$f_X(x) = h(F(x)) \cdot \frac{dF(x)}{dx} = f(x).$$

$$X = F^{-1}(U) \sim F.$$

# Erzeugung einer normalverteilten Zufallsvariablen

Ziel:  $X \sim \mathcal{N}(0, 1)$  erzeugen,

$$F(x) := \Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Erzeugung einer solchen Zufallsgröße:

- Quantilmethode (siehe oben)
- Zentraler Grenzwertsatz
- Box-Müller Transformation
- Akzeptanzmethode (siehe unten)
- Ziggurat-Algorithmus (siehe unten)

# Erzeugung einer normalverteilten Zufallsvariablen

## (2)

### Quantilmethode

$U \sim R(0, 1)$ .  $X := \Phi^{-1}(u) \sim \mathcal{N}(0, 1)$ , denn

$$f_X(x) = h(\Phi(x)) \cdot \frac{d\Phi(x)}{dx} = \frac{d\Phi(x)}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Problem: Berechnung von  $\Phi^{-1}(u)$  ist aufwendig.

Ziel:  $X \sim \mathcal{N}(\mu, \sigma^2)$  erzeugen,

$$Y := \mu + \sigma \cdot \Phi^{-1}(U) \sim \mathcal{N}(\mu, \sigma^2).$$

# Erzeugung einer normalverteilten Zufallsvariablen

(3)

## Zentraler Grenzwertsatz (1)

$U_1, \dots, U_n \sim R(0, 1)$  unabhängig. Erwartungswert und Varianz sind

$$\mu := EU_i = \int_0^1 x \, dx = \frac{1}{2} \quad \sigma^2 := E \left( U_i - \frac{1}{2} \right)^2 = \frac{1}{12}$$

$$\lim_{n \rightarrow \infty} P \left( \frac{\sum_{i=1}^n U_i - n \cdot \mu}{\sqrt{n} \cdot \sigma} < x \right) = \Phi(x).$$

Einsetzen:

$$\lim_{n \rightarrow \infty} P \left( \frac{\sum_{i=1}^n U_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} < x \right) = \Phi(x).$$

# Erzeugung einer normalverteilten Zufallsvariablen (4)

## Zentraler Grenzwertsatz (2)

Es sei  $n = 12$ .

Wir erhalten dann folgende Zufallsgröße  $X$ :

$$X = \sum_{i=1}^{12} U_i - 6.$$

Diese Approximation ist in der Regel ausreichend. Man braucht jedoch 12 Pseudozufallszahlen, um eine standardnormalverteilte Zufallsgröße zu erhalten.

# Erzeugung einer normalverteilten Zufallsvariablen (5)

## Satz BOX-MÜLLER-Transformation

Seien  $U, V \sim R(0, 1)$  unabhängig. Dann sind die Zufallsgrößen

$$X = \sqrt{-2 \cdot \ln U} \cdot \cos(2\pi V)$$

$$Y = \sqrt{-2 \cdot \ln U} \cdot \sin(2\pi V)$$

unabhängig und standardnormalverteilt,  $X, Y \sim \mathcal{N}(0, 1)$ .

**Beweis:** vgl. Abschnitt Transformationsformel





# Erzeugung exponentialverteilter Zufallsvariablen

Es sei  $U \sim R(0, 1)$  eine Pseudozufallszahl. Erzeugt werden soll eine Zufallsgröße  $X \sim \text{Exp}(\lambda)$  mit der Verteilungsfunktion:

$$F(x) = \begin{cases} 1 - e^{-\lambda \cdot x} & , \text{ falls } x \geq 0; \\ 0 & , \text{ sonst.} \end{cases}$$

Dazu wird folgende Transformation verwendet

$$X := F^{-1}(U) = -\frac{1}{\lambda} \cdot \ln(1 - u) \geq 0.$$

# Erzeugung binomialverteilter Zufallsvariablen

**Variante 1:** Seien  $X_i \sim Bi(1, p)$ . Dann ist  $X = \sum_{i=1}^n X_i$  binomialverteilt mit Parametern  $(n, p)$ .

# Erzeugung binomialverteilter Zufallsvariablen

**Variante 1:** Seien  $X_i \sim Bi(1, p)$ . Dann ist  $X = \sum_{i=1}^n X_i$  binomialverteilt mit Parametern  $(n, p)$ .

**Variante 2:** (Intervallmethode)

Zerlegen das Intervall  $(0, 1)$  in disjunkte Teilintervalle der Länge der Einzelwahrscheinlichkeiten,

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

$$(0, 1) = \bigcup_{i=0}^n I_i = (0, p_0] \cup (p_0, p_0 + p_1] \cup (p_0 + p_1, p_0 + p_1 + p_2] \cup \dots \cup (1 - \sum_{i=0}^{n-1} p_i, 1)$$

Sei  $U \sim R(0, 1)$ .  $X = i$  falls  $U \in I_i$ .

# Erzeugung POISSON-verteilter Zufallsvariablen

## (1)

Es ist eine POISSON-verteilte Zufallsgröße  $X$  zu erzeugen, d.h.

$$P(X = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda} \quad (i = 0, 1, 2, \dots).$$

**Variante 1:** Intervallmethode

**Variante 2:** (Über die Exponentialverteilung)

# Erzeugung POISSON-verteilter Zufallsvariablen (2)

## Satz

Es seien  $Y_1, \dots, Y_k$  unabhängige exponentialverteilte Zufallsgrößen und  $Y^{(k)} := \sum_{i=1}^k Y_i$ . Dann gilt für die Dichte der Zufallsvariable  $Y^{(k)}$ :

$$f_{Y^{(k)}}(y) = \begin{cases} \frac{\lambda^k}{(k-1)!} \cdot y^{k-1} \cdot e^{-\lambda \cdot y} & , \text{ falls } y \geq 0; \\ 0 & , \text{ sonst.} \end{cases}$$

Diese Funktion ist die Dichte der ERLANG-Verteilung mit Parametern  $(k, \lambda)$ .

# Erzeugung POISSON-verteilter Zufallsvariablen (3)

**Beweis.** Wir beweisen die Aussage mittels vollständiger Induktion. Es sei  $y \geq 0$ .

**IA:**  $Y^{(1)} = Y_1 \sim \text{Exp}(\lambda) = \text{Erl}(1, \lambda)$

**IV:** Es sei die Aussage für  $k$  gültig.

**IS:** Wir zeigen sie für  $k + 1$ . Es gilt:

$$Y^{(k+1)} = Y^{(k)} + Y_{k+1}.$$

Bestimmen die Dichtefunktion  $f_{Y^{(k+1)}}$  mittels Faltung der Dichtefunktionen  $f_{Y^{(k)}}$  und  $f_{Y^{(1)}}$ .

# Erzeugung POISSON-verteilten Zufallsvariablen

## (4)

Zum Beweis des Satzes:

$$\begin{aligned}f_{Y^{(k+1)}}(y) &= \int_0^{\infty} f_{Y^{(k)}}(x) \cdot f_{Y^{(1)}}(y-x) dx \\&= \int_0^y \frac{\lambda^k}{(k-1)!} \cdot x^{k-1} \cdot e^{-\lambda \cdot x} \cdot \lambda \cdot e^{-\lambda \cdot (y-x)} dx \\&= \int_0^y \frac{\lambda^{k+1}}{(k-1)!} \cdot x^{k-1} \cdot e^{-\lambda \cdot y} dx \\&= \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda y} \int_0^y x^{k-1} dx = \frac{\lambda^{k+1}}{k!} y^k e^{-\lambda y}\end{aligned}$$

# Erzeugung einer POISSON-Verteilten Zufallsvariable (5)

## Satz

Sind  $Y_i$  ( $i \in \mathbb{N}$ ) unabhängige, exponentialverteilte Zufallsgrößen ( $Y_i \sim \text{EX}(\lambda)$ ,  $i \in \mathbb{N}$ ), so ist die wie folgt definierte Zufallsvariable  $Y$  POISSON-verteilt mit Parameter  $\lambda$ :

$$Y := \inf \left\{ k : \sum_{i=1}^{k+1} Y_i > 1 \right\} \sim \text{Poi}(\lambda).$$

Es gilt also:

$$P(Y = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda} \quad (i = 1, 2, \dots).$$



# Erzeugung einer POISSON-Verteilten Zufallsvariable (6)

**Beweis:** Es gilt:

$$P(Y = k)$$

# Erzeugung einer POISSON-Verteilten Zufallsvariable (6)

**Beweis:** Es gilt:

$$P(Y = k) = P\left(\sum_{i=1}^k Y_i \leq 1, \sum_{i=1}^{k+1} Y_i > 1\right)$$

# Erzeugung einer POISSON-Verteilten Zufallsvariable (6)

**Beweis:** Es gilt:

$$\begin{aligned}P(Y = k) &= P\left(\sum_{i=1}^k Y_i \leq 1, \sum_{i=1}^{k+1} Y_i > 1\right) \\&= P\left(\underbrace{\sum_{i=1}^k Y_i}_{=: T} \leq 1, Y_{k+1} > 1 - \sum_{i=1}^k Y_i\right) \\&= \int_0^1 P(Y_{k+1} > 1 - T | T = t) f_T(t) dt \\&= \int_0^1 P(Y_{k+1} > 1 - t) f_T(t) dt\end{aligned}$$

# Erzeugung einer POISSON-Verteilten Zufallsvariable (7)

Da  $T = Y^{(k)} = \sum_{i=1}^k Y_i$  Erlang-verteilt ist, folgt

$$\begin{aligned} P(Y = k) &= \int_0^1 e^{-\lambda(1-t)} \cdot \frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t} dt \\ &= e^{-\lambda} \lambda^k \int_0^1 \frac{t^{k-1}}{(k-1)!} dt \\ &= e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

# Erzeugung einer geometrisch verteilten Zufallsvariable

**Variante 1:** Intervallmethode

# Erzeugung einer geometrisch verteilten Zufallsvariable

**Variante 1:** Intervallmethode

**Variante 2:** Zur Erzeugung einer geometrisch verteilten Zufallsvariablen  $X \sim \text{Geo}(p)$  seien  $Y_i \sim \text{Bi}(1, p)$  Bernoulli verteilte Zufallsvariablen und

$$X = \min\{n : Y_n = 1\}$$

# Erzeugung einer geometrisch verteilten Zufallsvariable

**Variante 1:** Intervallmethode

**Variante 2:** Zur Erzeugung einer geometrisch verteilten Zufallsvariablen  $X \sim \text{Geo}(p)$  seien  $Y_i \sim \text{Bi}(1, p)$  Bernoulli verteilte Zufallsvariablen und

$$X = \min\{n : Y_n = 1\}$$

**Variante 3:** Sei  $Y \sim \text{Exp}(\lambda)$ , d.h.  $F(y) = 1 - e^{-\lambda y}$ . Die Zufallsvariable  $\lfloor Y \rfloor + 1$  ist geometrisch verteilt mit  $p = 1 - e^{-\lambda}$ .

# Erzeugung einer geometrisch verteilten Zufallsvariable (2)

**Beweis:** Es gilt:

$$\begin{aligned}P(\lfloor Y \rfloor = k) &= P(k \leq Y < k + 1) \\&= F(k + 1) - F(k) \\&= (1 - e^{-\lambda(k+1)}) - (1 - e^{-\lambda k}) \\&= e^{-\lambda k}(1 - e^{-\lambda}) = (1 - p)^k p\end{aligned}$$





# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

Sei  $F$  eine Linearkombination von mehreren Verteilungsfunktionen  $F_i$ ,

$$F = \sum_{i=1}^k \epsilon_i F_i, \quad \sum_{i=1}^k \epsilon_i = 1, \quad \epsilon_i \geq 0.$$

Algorithmus:

Erzeuge gleichverteilte Zufallszahl  $U$ ,

falls  $U \in [\sum_{j=1}^{i-1} \epsilon_j, \sum_{j=1}^i \epsilon_j)$  simuliere aus  $F_i$ .

Es folgen zwei Beispiele.

# Kompositionsmethode (2)

## Kontaminierte Normalverteilung

$$F(x) = (1 - \epsilon)\Phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \epsilon\Phi\left(\frac{x - \mu_2}{\sigma_2}\right)$$

# Kompositionsmethode (2)

## Kontaminierte Normalverteilung

$$F(x) = (1 - \epsilon)\Phi\left(\frac{x - \mu_1}{\sigma_1}\right) + \epsilon\Phi\left(\frac{x - \mu_2}{\sigma_2}\right)$$

## Doppelexponential (Laplace)

$$X_1 \sim \exp(\lambda)$$

$$X = \begin{cases} X_1 & \text{falls } U \leq \frac{1}{2} \\ -X_1 & \text{falls } U > \frac{1}{2} \end{cases}$$

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Verwerfungsmethode

oder Akzeptanzmethode oder Accept-Reject Sampling

$F$  habe Dichte  $f$ , aber die Zufallszahlen seien schwierig direkt zu erzeugen.

Erzeugung von Zufallszahlen mit der Dichte  $g$  sei “leicht”.

$$M := \sup_x \frac{f(x)}{g(x)} < \infty$$

Algorithmus:

1. Simuliere  $U \sim R(0, 1)$
2. Simuliere  $Y \sim g$
3. Akzeptiere  $X = Y$ , falls  $U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}$  sonst gehe nach 1. (neuer Versuch)

## Verwerfungsmethode (2)

Berechnen die Wahrscheinlichkeit, dass eine Zufallszahl akzeptiert wird,  $U \sim R(0, 1)$ ,  $Y \sim g$ :

## Verwerfungsmethode (2)

Berechnen die Wahrscheinlichkeit, dass eine Zufallszahl akzeptiert wird,  $U \sim R(0, 1)$ ,  $Y \sim g$ :

$$P(Y_{\text{akzeptiert}}) = P\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right)$$



## Verwerfungsmethode (2)

Berechnen die Wahrscheinlichkeit, dass eine Zufallszahl akzeptiert wird,  $U \sim R(0, 1)$ ,  $Y \sim g$ :

$$\begin{aligned} P(Y_{\text{akzeptiert}}) &= P\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) \\ &= \int P\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)} \mid Y = y\right) g(y) dy \end{aligned}$$

## Verwerfungsmethode (2)

Berechnen die Wahrscheinlichkeit, dass eine Zufallszahl akzeptiert wird,  $U \sim R(0, 1)$ ,  $Y \sim g$ :

$$\begin{aligned}P(Y_{\text{akzeptiert}}) &= P\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)}\right) \\&= \int P\left(U \leq \frac{1}{M} \frac{f(Y)}{g(Y)} \mid Y = y\right) g(y) dy \\&= \int \frac{1}{M} \frac{f(y)}{g(y)} \cdot g(y) dy = \frac{1}{M}\end{aligned}$$

wenn die Definitionsbereiche von  $X$  und  $Y$  übereinstimmen.

(Integration über den Definitionsbereich von  $Y$ )

Im Mittel müssen also  $M$  Zufallszahlen  $Y$  erzeugt werden.

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$P(X \leq x | Y_{\text{akzept.}})$$

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$P(X \leq x | Y_{\text{akzept.}}) = \int_{-\infty}^x P(X = Y = y | Y_{\text{akzept.}}) g(y) dy$$

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$\begin{aligned}P(X \leq x | Y_{\text{akzept.}}) &= \int_{-\infty}^x P(X = Y = y | Y_{\text{akzept.}}) g(y) dy \\&= \int_{-\infty}^x \frac{P(Y_{\text{akzept.}}, Y=y)}{P(Y_{\text{akzept.}})} g(y) dy\end{aligned}$$

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$\begin{aligned}P(X \leq x | Y_{\text{akzept.}}) &= \int_{-\infty}^x P(X = Y = y | Y_{\text{akzept.}}) g(y) dy \\&= \int_{-\infty}^x \frac{P(Y_{\text{akzept.}}, Y=y)}{P(Y_{\text{akzept.}})} g(y) dy \\&= \int \frac{P\left(U \leq \frac{1}{M} \frac{f(y)}{g(y)}\right)}{P(Y_{\text{akzept.}})} g(y) dy\end{aligned}$$

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$\begin{aligned}P(X \leq x | Y_{\text{akzept.}}) &= \int_{-\infty}^x P(X = Y = y | Y_{\text{akzept.}}) g(y) dy \\&= \int_{-\infty}^x \frac{P(Y_{\text{akzept.}}, Y=y)}{P(Y_{\text{akzept.}})} g(y) dy \\&= \int \frac{P\left(U \leq \frac{1}{M} \frac{f(y)}{g(y)}\right)}{P(Y_{\text{akzept.}})} g(y) dy \\&= M \int_{-\infty}^x \frac{1}{M} \frac{f(y)}{g(y)} g(y) dy\end{aligned}$$

## Verwerfungsmethode (3)

Die Methode ist korrekt, denn:

$$\begin{aligned}P(X \leq x | Y_{\text{akzept.}}) &= \int_{-\infty}^x P(X = Y = y | Y_{\text{akzept.}}) g(y) dy \\&= \int_{-\infty}^x \frac{P(Y_{\text{akzept.}}, Y=y)}{P(Y_{\text{akzept.}})} g(y) dy \\&= \int \frac{P\left(U \leq \frac{1}{M} \frac{f(y)}{g(y)}\right)}{P(Y_{\text{akzept.}})} g(y) dy \\&= M \int_{-\infty}^x \frac{1}{M} \frac{f(y)}{g(y)} g(y) dy \\&= F(x).\end{aligned}$$



# Verwerfungsmethode (4)

## Simulation einer Standardnormal

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (\text{Normal})$$

$$g(x) = \frac{1}{2} e^{-|x|} \quad (\text{Doppelexp})$$

# Verwerfungsmethode (4)

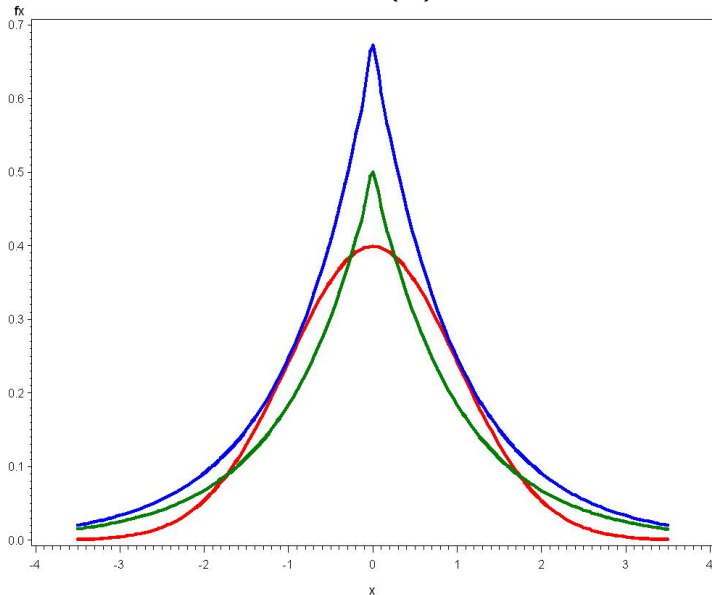
## Simulation einer Standardnormal

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (\text{Normal})$$

$$g(x) = \frac{1}{2} e^{-|x|} \quad (\text{Doppelexp})$$

$$\begin{aligned} \sup_x \frac{f(x)}{g(x)} &= \sup_x \sqrt{\frac{2}{\pi}} e^{-x^2/2+|x|} \\ &= \sqrt{\frac{2}{\pi}} \sup_x e^{(-x^2+2|x|-1+1)/2} \\ &= \sqrt{\frac{2}{\pi}} e^{1/2} \sup_{x, x \geq 0} e^{-\frac{1}{2}(x-1)^2} = \sqrt{\frac{2}{\pi}} e^{1/2} \approx 1.315. \end{aligned}$$

# Verwerfungsmethode (5)



# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Anwendung der Verwerfungsmethode

## Der Ziggurat-Algorithmus

Sei  $f$  die zu simulierende Verteilung, z.B.  $f$ =Normaldichte. Wir simulieren der Einfachheit halber nur den positiven Teil.

Idee ist, die Fläche unter der Dichte (möglichst knapp) zu überdecken durch Rechtecke. Dann wird ein zufälliger Punkt in dieser Überdeckung generiert, und wenn er in der Fläche liegt akzeptiert, sonst nicht.

### Matlab 7.4

wird z.B. in Matlab 7.4 zur Erzeugung normalverteilter Zufallszahlen als Standard verwendet.

# Der Ziggurat-Algorithmus

## Vorbereitungen für den Ziggurat-Algorithmus

Sei  $n$  die Anzahl der Streifen. Dann werden die zur x-Achse parallelen Streifen sukzessive so definiert, dass die Flächeninhalte der Rechtecke und der Flächeninhalt des Basistreifens unter der Kurve jeweils gleich  $v$  sind. ( $v$  ist noch zu bestimmen!)

Der Flächeninhalt des Basistreifens (der durch die x-Achse, die Parallele  $y = y_n = f(x_n)$  und den Tail von  $f$  begrenzt ist.) bzw. der der Rechtecke ist ( $r = x_{n-1}$ )

$$v := rf(r) + \int_r^\infty f(t) dt \quad \text{bzw.} \quad v = x_{i+1}(f(x_i) - f(x_{i+1})).$$

# Der Ziggurat-Algorithmus

## Vorbereitungen für den Ziggurat-Algorithmus, 2

Zu bestimmen sind  $v$  sowie die Punkte  $x_i, i = 255, \dots, 0$ . Dies geschieht rekursiv, indem man  $r = x_{255}$  geschickt rät, den Algorithmus laufen lässt, und (hoffentlich) mit  $x_0 \approx 0$  endet.

### Algorithmus zur Bestimmung der $x_i$ :

- 1. Rate  $r$ . Bei  $n = 256$  etwa  $r = 3.5$ . Das bekommt man etwa, wenn man die Flächeninhalte  $v = v_{start}$  etwa auf  $\frac{1}{256}$  setzt.
- 2. **for**  $i$  **from**  $n - 1$  **to**  $0$  **do**  $x_i = f^{-1}(\frac{v}{x_{i+1}} + f(x_{i+1}))$
- 3. Wenn  $x_0 \approx 0$  nehme die berechneten  $x_i$  **return**
- 4. Aktualisiere  $r$  und gehe zu 1. Wenn  $x_0 > 0$  so  $r$  verkleinern ( $v$  war zu klein und  $r$  zu groß), sonst vergrößern.

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$



## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**  
Wenn  $i = n + 1$  gehe nach 6. (der unterste Streifen)

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**  
Wenn  $i = n + 1$  gehe nach 6. (der unterste Streifen)
- 4. Sei  $V \sim R(0, 1)$  und  $y = y_i + V(y_{i-1} - y_i)$ ,  $y_i = f(x_i)$ .

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**  
Wenn  $i = n + 1$  gehe nach 6. (der unterste Streifen)
- 4. Sei  $V \sim R(0, 1)$  und  $y = y_i + V(y_{i-1} - y_i)$ ,  $y_i = f(x_i)$ .
- 5. Berechne  $f(x)$ , Wenn  $f(x) > y$  **akzeptiere**  $x$ , **return**

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**  
Wenn  $i = n + 1$  gehe nach 6. (der unterste Streifen)
- 4. Sei  $V \sim R(0, 1)$  und  $y = y_i + V(y_{i-1} - y_i)$ ,  $y_i = f(x_i)$ .
- 5. Berechne  $f(x)$ , Wenn  $f(x) > y$  **akzeptiere**  $x$ , **return**  
sonst gehe nach 1. zurück.

## Ziggurat-Algorithmus

- 1. Wähle eine zufällige ganze Zahl  $i$ ,  $1 \leq i \leq n + 1$
- 2. Sei  $U \sim R(0, 1)$  und  $x = Ux_i$
- 3. Wenn  $x < x_{i-1}$  so **akzeptiere**  $x$ , **return**  
Wenn  $i = n + 1$  gehe nach 6. (der unterste Streifen)
- 4. Sei  $V \sim R(0, 1)$  und  $y = y_i + V(y_{i-1} - y_i)$ ,  $y_i = f(x_i)$ .
- 5. Berechne  $f(x)$ , Wenn  $f(x) > y$  **akzeptiere**  $x$ , **return**  
sonst gehe nach 1. zurück.
- 6. Jetzt ist nur noch eine Beobachtung aus dem Tail der Verteilung,  $x > x_n =: r$  zu generieren. Dazu wird die Akzeptanzmethode verwendet.

# Ziggurat-Algorithmus

## Anmerkungen

- zu 1. Hier wird der Streifen ausgewählt. Je größer  $n$ , desto schmaler die Streifen, und desto knapper die Überdeckung von  $f$ , und desto mehr Zufallszahlen werden akzeptiert. (Marsaglia:  $n=255$ )
- Die Werte von  $(x_i, y_i)$  sind in einer Tabelle abzuspeichern.
- zu 3. Wenn  $x < x_{i-1}$  so liegt der Punkt  $(x, y)$  sicher in der Fläche unterhalb  $f$ .
- zu 5. Wenn  $f(x) > y$  dann liegt der Punkt  $(x, y)$  in der Fläche unterhalb  $f$ .



# Ziggurat-Algorithmus

Der Fall des Basisstreifens ( $i = n + 1$ ),  $f$  Standardnormal

Zu generieren ist eine Beobachtung aus dem Tail der Verteilung  $f$  (normal). Die bedingte Dichte, unter der Bedingung  $X > r$ , ist dann  $f/(1 - \Phi(r))$ .

Proposal Verteilung sei

# Ziggurat-Algorithmus

Der Fall des Basisstreifens ( $i = n + 1$ ),  $f$  Standardnormal

Zu generieren ist eine Beobachtung aus dem Tail der Verteilung  $f$  (normal). Die bedingte Dichte, unter der Bedingung  $X > r$ , ist dann  $f/(1 - \Phi(r))$ .

Proposal Verteilung sei z.B. verschobene Exponential,  
 $g(x) = re^{-r(x-r)}, x \geq r$ . (bei  $n = 256$  ist  $r \approx 3.65$ .)

# Ziggurat-Algorithmus

Der Fall des Basisstreifens ( $i = n + 1$ ),  $f$  Standardnormal

Zu generieren ist eine Beobachtung aus dem Tail der Verteilung  $f$  (normal). Die bedingte Dichte, unter der Bedingung  $X > r$ , ist dann  $f/(1 - \Phi(r))$ .

Proposal Verteilung sei z.B. verschobene Exponential,

$g(x) = re^{-r(x-r)}, x \geq r$ . (bei  $n = 256$  ist  $r \approx 3.65$ .)

$$\begin{aligned}
 M &:= \sup_{x \geq r} \frac{f(x)/(1 - \Phi(r))}{g(x)} = \sup_{x \geq r} \frac{1}{(1 - \Phi(r))\sqrt{2\pi}r} e^{-\frac{x^2}{2} + rx - r^2} \\
 &= \frac{e^{-\frac{r^2}{2}}}{(1 - \Phi(r))\sqrt{2\pi}r} \sup_{x \geq r} e^{-\frac{1}{2}(x^2 - 2rx + r^2)} = \frac{e^{-\frac{r^2}{2}}}{(1 - \Phi(r))\sqrt{2\pi}r} \\
 &\approx 1.06
 \end{aligned}$$

# Ziggurat-Algorithmus

Der Fall des Basisstreifens ( $i = n + 1$ ),  $f$  Standardnormal

Zu generieren ist eine Beobachtung aus dem Tail der Verteilung  $f$  (normal). Die bedingte Dichte, unter der Bedingung  $X > r$ , ist dann  $f/(1 - \Phi(r))$ .

Proposal Verteilung sei z.B. verschobene Exponential,

$g(x) = re^{-r(x-r)}, x \geq r$ . (bei  $n = 256$  ist  $r \approx 3.65$ .)

$$\begin{aligned} M &:= \sup_{x \geq r} \frac{f(x)/(1 - \Phi(r))}{g(x)} = \sup_{x \geq r} \frac{1}{(1 - \Phi(r))\sqrt{2\pi}r} e^{-\frac{x^2}{2} + rx - r^2} \\ &= \frac{e^{-\frac{r^2}{2}}}{(1 - \Phi(r))\sqrt{2\pi}r} \sup_{x \geq r} e^{-\frac{1}{2}(x^2 - 2rx + r^2)} = \frac{e^{-\frac{r^2}{2}}}{(1 - \Phi(r))\sqrt{2\pi}r} \\ &\approx 1.06 \end{aligned}$$

(Das ist die Version von Marsaglia)

## Akzeptanzmethode für den Basisstreifen

- Erzeuge  $U \sim R(0, 1)$
- Erzeuge  $V \sim g$ , d.h.  $V = r - \frac{\ln V_1}{r}$  und  $V_1 \sim R(0, 1)$ .
- Akzeptiere falls

$$U \leq \frac{1}{M} \frac{f(V)}{g(V)} = \frac{1}{M} M e^{-\frac{1}{2}(V-r)^2} \quad \text{gdw.}$$

$$-\ln U \geq \frac{1}{2}(V-r)^2 \quad \text{gdw.}$$

$$2(-\ln U) \geq \left(-\frac{\ln V_1}{r}\right)^2 \quad \text{gdw.}$$

$$2Y \geq X^2,$$

wobei  $Y \sim \text{Exp}(1)$ ,  $X \sim \frac{1}{r} \text{Exp}(1)$ .

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Erzeugung von korrelierten Zufallsgrößen

Es seien  $X$  und  $Y$  zwei unabhängige, standardisierte Zufallsgrößen ( $X, Y \sim (0, 1)$ ). Seien

$$X^* := X$$

$$Y^* := \varrho \cdot X + \sqrt{1 - \varrho^2} \cdot Y \quad (\varrho \in [-1, 1])$$

**Beh.:**  $\varrho$  ist der gewünschte Korrelationskoeffizient zwischen  $X^*$  und  $Y^*$  (s. Abschnitt Korrelation).

Ist  $\varrho = 1$ , dann gilt  $Y^* = X^* = X$ , d.h. die beiden Zufallsgrößen sind identisch. Wird  $\varrho = 0$  gewählt, so sind beide Zufallsvariablen unabhängig.

# Erzeugung von zweidimensionalen Zufallsgrößen

Ziel: Erzeugung von Zufallsvariablen mit Dichte  $h(x, y)$

Seien  $F_Y(y)$  die Verteilungsfunktion von  $Y$ , und  $F_{X|Y}(x|y)$  die bedingte Verteilungsfunktion von  $X$  unter der Bedingung  $Y$ .

- Erzeuge  $U, V \sim R(0, 1)$
- $X = F_{X|Y}^{-1}(U|V) =: g_1(U, V)$
- $Y = F_Y^{-1}(V) =: g_2(U, V)$

Offenbar

$$g^{-1}(x, y) = (F_{X|Y}(x|y), F_Y(y))$$



Die Jacobi-Matrix von  $g^{-1}$  ist

$$J = \begin{pmatrix} \frac{\partial F_{X|Y}(x|y)}{\partial x} & \frac{\partial F_{X|Y}(x|y)}{\partial y} \\ \frac{\partial F_Y(y)}{\partial x} & \frac{\partial F_Y(y)}{\partial y} \end{pmatrix} = \begin{pmatrix} f_{X|Y}(x|y) & \frac{\partial F_{X|Y}(x|y)}{\partial y} \\ 0 & f_Y(y) \end{pmatrix}$$

Die gemeinsame Dichte  $h(x, y)$  von  $(X, Y)$  ist nach dem Transformationssatz

$$h(x, y) = \underbrace{f_{U,V}(F_{X|Y}(y|x), F_Y(y))}_{=1} |det J| = f_{X|Y}(x|y) f_Y(y)$$

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen

# Importance Sampling

Ziel: Berechnung (Schätzung) von Integralen

$$I = \int h(x)f(x) dx,$$

wobei  $f$  eine Dichte ist.

## 1. Methode: Monte Carlo

Simulieren Realisierungen  $x_1, \dots, x_n$  aus einer Population mit Dichte  $f$ .

Schätzen den Erwartungswert  $I$  durch das arithmetische Mittel

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n h(x_i)$$

# Berechnung (Schätzung) von Integralen (2)

Aber, was wenn Simulation von  $f$  schwer ist?

## Berechnung (Schätzung) von Integralen (2)

Aber, was wenn Simulation von  $f$  schwer ist?

### 2. Methode: Importance Sampling

Suchen uns eine Dichte  $g$ , die “leicht” zu simulieren ist. Dann wird

$$I = \int h(x) \frac{f(x)}{g(x)} g(x) dx$$

geschätzt durch

$$\hat{I}_1 = \frac{1}{n} \sum_{i=1}^n h(x_i) \frac{f(x_i)}{g(x_i)},$$

wobei  $x_1, \dots, x_n$  jetzt Realisierungen aus einer Population mit Dichte  $g$  sind.

# Berechnung (Schätzung) von Integralen (3)

Problem:  $\hat{I}_1$  kann unendliche Varianz haben

Lösung:  $g$  "nahe"  $f$  (dann der Quotient nahe 1) und

## Berechnung (Schätzung) von Integralen (3)

Problem:  $\hat{I}_1$  kann unendliche Varianz haben

Lösung:  $g$  "nahe"  $f$  (dann der Quotient nahe 1) und  $g$  "dickere" Tails als  $f$ , der Quotient ist dann in den Tails  $< 1$ .

# Inhalt

- Einführung
- Erzeugung von Zufallszahlen
- Statistische Tests
- Test auf Gleichverteilung
- Test auf Unabhängigkeit
- Erzeugung diskreter und stetiger Zufallsvariablen
- Kompositionsmethode
- Verwerfungsmethode
- Ziggurat-Algorithmus
- Korrelierte Zufallsgrößen
- Importance Sampling
- Ergänzungen



# Das Buffonsche Nadelproblem (1777)

In der Ebene seien zwei parallele Geraden im Abstand  $a$  gezogen. Auf die Ebene wird zufällig eine Nadel der Länge  $l$ , ( $l \leq a$ ) geworfen.

**Frage:** Wie groß ist die Wahrscheinlichkeit, daß die Nadel eine der Geraden schneidet?

# Das Buffonsche Nadelproblem (1777)

In der Ebene seien zwei parallele Geraden im Abstand  $a$  gezogen. Auf die Ebene wird zufällig eine Nadel der Länge  $l$ , ( $l \leq a$ ) geworfen.

**Frage:** Wie groß ist die Wahrscheinlichkeit, daß die Nadel eine der Geraden schneidet?

Was heißt Nadel zufällig werfen?

# Das Buffonsche Nadelproblem (1777)

In der Ebene seien zwei parallele Geraden im Abstand  $a$  gezogen. Auf die Ebene wird zufällig eine Nadel der Länge  $l$ , ( $l \leq a$ ) geworfen.

**Frage:** Wie groß ist die Wahrscheinlichkeit, daß die Nadel eine der Geraden schneidet?

Was heißt Nadel zufällig werfen?

$X$ : Abstand des Nadelmittelpunkts von der nächstgelegenen Geraden,  
$$0 \leq X \leq \frac{a}{2}.$$

$\phi$ : Winkel zwischen Nadel und Geraden,  $0 < \phi \leq \pi$ .

# Das Buffonsche Nadelproblem (2)

Nadel zufällig werfen:

$$X \sim R(0, \frac{a}{2}), \quad \phi \sim R(0, \pi).$$

Wann schneidet die Nadel eine Parallele? gdw.

# Das Buffonsche Nadelproblem (2)

Nadel zufällig werfen:

$$X \sim R(0, \frac{a}{2}), \quad \phi \sim R(0, \pi).$$

Wann schneidet die Nadel eine Parallele? gdw.

$$X \leq \frac{l}{2} \sin \phi \quad \text{gdw.}$$

der Punkt  $(\phi, X)$  unterhalb des Sinusbogens liegt.

# Das Buffonsche Nadelproblem (2)

Nadel zufällig werfen:

$$X \sim R(0, \frac{a}{2}), \quad \phi \sim R(0, \pi).$$

Wann schneidet die Nadel eine Parallele? gdw.

$$X \leq \frac{l}{2} \sin \phi \quad \text{gdw.}$$

der Punkt  $(\phi, X)$  unterhalb des Sinusbogens liegt.

$$\begin{aligned} P &= \frac{\text{Fläche unterhalb des Sinusbogens}}{\text{Fläche des Rechtecks}[0, \pi] \times [0, \frac{a}{2}]} \\ &= \frac{\int_0^\pi \frac{l}{2} \sin \phi \, d\phi}{\pi \cdot \frac{a}{2}} = \frac{2l}{\pi a} \end{aligned}$$

# Das Buffonsche Nadelproblem (3)

Insbesondere:  $a = 2l$ :

$$P = \frac{1}{\pi}.$$

Schätzung für  $\pi$ :

$$\hat{\pi} = \frac{\# \text{Würfe}}{\# \text{Treffer}}$$

# Simulation einer Markov'schen Kette

gegeben: Zustandsraum:  $S = \{1, 2, \dots\}$

Anfangsverteilung:  $\{p_j^0\}_{j=1,2,\dots}$ ,  $(p_0^0 = 0)$

Übergangsmatrix:

$$\left( p_{ij} \right)_{\substack{i=1,2,\dots \\ j=1,2,\dots}}$$

1. Schritt: Erzeuge eine Pseudozufallszahl  $U_0$ . Falls

$$\sum_{k=0}^{i-1} p_k^0 \leq U_0 < \sum_{k=0}^i p_k^0$$

so starte im Zustand "i".



## Simulation einer Markov'schen Kette (2)

$n$ -ter Schritt: Im  $n - 1$ ten Schritt sei der Zustand "i" erreicht worden.  
Erzeuge eine Pseudozufallszahl  $U_n$ . Falls

$$\sum_{k=0}^{j-1} p_{ik} \leq U_n < \sum_{k=0}^j p_{ik}$$

so gehe in den Zustand "j".

# \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Satz: Seien  $X_i \sim \mathcal{N}(0, 1)$ , i.i.d.  $i = 1, \dots, n$ , und

$$Y_i = \frac{X_i}{R}, \quad i = 1, \dots, n,$$

wobei

$$R^2 = \sum_{i=1}^n X_i^2.$$

Dann gilt

$$Y_i \sim R(K_n^O(0, 1)),$$

wobei  $K_n^O(0, 1)$  die Oberfläche der  $n$ -dimensionalen Einheitskugel ist.

## \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Sei  $K_{n-1}(0, 1)$  die  $n - 1$  dim. Einheitsvollkugel. Wir betrachten die Transformation

$$G : \mathbb{R}^{n-1} \times \mathbb{R}^+ \rightarrow K_{n-1}(0, 1) \times \mathbb{R}^+$$

$$y_2 = \frac{x_2}{r}$$

...

$$y_n = \frac{x_n}{r}$$

$$r = r$$

# \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Diese Abbildung ist injektiv und es gilt für  $G^{-1}$ :

$$x_2 = r \cdot y_2$$

...

$$x_n = r \cdot y_n$$

$$r = r$$

# \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Die Jacobi-Matrix ist

$$J := \frac{\partial G^{-1}(y_2, \dots, y_n, r)}{\partial (y_2, \dots, y_n, r)} = \begin{pmatrix} r & 0 & \dots & 0 & y_2 \\ 0 & r & \dots & 0 & y_3 \\ & & \dots & & \\ 0 & 0 & \dots & r & y_n \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Also:  $\det J = r^{n-1}$ .

# \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Die gemeinsame Dichte von  $(\mathbf{Y}, R) = (Y_1, Y_2, \dots, Y_n, R)$  ist

$$f_{\mathbf{Y}, R}(y_1, \dots, y_n, r) =$$

$$= \begin{cases} f_{\mathbf{X}, R}(ry_1, G^{-1}(y_2, \dots, y_n, r)) \det J, & y_1^2 = 1 - \sum_{j=2}^n y_j^2 \\ 0 & \text{sonst} \end{cases}$$

$$= \begin{cases} \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{j=1}^n e^{-\frac{r^2 y_j^2}{2}} \cdot r^{n-1}, & y_n^2 = 1 - \sum_{j=1}^{n-1} y_j^2 \\ 0, & \text{sonst} \end{cases}$$

$$= \begin{cases} \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{r^2}{2}} \cdot r^{n-1} & \text{falls } y_n^2 = 1 - \sum_{j=1}^{n-1} y_j^2 \\ 0 & \text{sonst} \end{cases}$$

## \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

Die Zufallsvektoren  $(Y_1, \dots, Y_n)$  und  $R$  sind also unabhängig und wegen

$$\frac{e^{-\frac{r^2}{2}} \cdot r^{n-1}}{(2\pi)^{n/2}} = \frac{r^{n-1} e^{-\frac{r^2}{2}}}{2^{\frac{n}{2}-1} \Gamma(\frac{n}{2})} \cdot \frac{\Gamma(\frac{n}{2})}{2\pi^{\frac{n}{2}}} = f_{\chi_n}(r) \cdot \frac{1}{A_{K_n^O(0,1)}}$$

ist

$$R \sim \chi_n \quad \text{und} \quad \mathbf{Y} \sim R(K_n^O(0,1))$$

mit der Dichte  $\frac{1}{A_{K_n^O(0,1)}}$ , wobei

$$A_{K_n^O(0,1)} = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$$

die Fläche der  $n$ -dimensionalen Einheitskugel ist.

## \*Simulation von auf der Kugeloberfläche

gleichverteilten Zufallsvariablen

**Bem.:** Die Fläche der  $n$ -dimensionalen Kugeloberfläche ist, vgl. Fichtenholz 3, S.389,

$$A_{K_n^O(0,r)} = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} r^{n-1}$$

$$n = 2: \quad 2\pi r$$

$$n = 3: \quad 4\pi r^2 \quad \left( \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \right)$$

$$n = 4: \quad 4\pi^2 r^3$$



# Markov'sche Ketten

## Beispiele

Irrfahrten (auf der Geraden, der Ebene, im Raum)

Ruin des Spielers

Markov Chain Monte Carlo (z.B. Simulated Annealing)

## Fragestellungen

Rückkehr-, Absorptionswahrscheinlichkeiten

Erste Rückkehr

Stationäre Verteilungen

# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo

$\{X_t\}_{t \in T}$ : Familie von Zufallsgrößen.

$T$ : total geordnete Menge (mit kleinstem Element  $t_0$ ).

$T$  endlich, o.B.d.A.  $T = \{0, 1, 2, \dots, k\}$  oder

$T$  abzählbar, o.B.d.A.  $T \in \{0, 1, 2, \dots\} = \mathbb{N}$

Wir betrachten ein System, das aus einem Anfangszustand für  $t = t_0$   
schrittweise übergeht in Zustände für  $t = t_1, t = t_2, \dots$

Menge der Zustände: Zustandsraum  $S$ ,

$S = \{1, 2, \dots, m\}$  oder  $S = \mathbb{N}$  oder  $S = \mathbb{Z}$ .

## Definitionen (2)

Für jedes  $t$  wird der (aktuelle) Zustand durch eine Zufallsvariable  $X_t$  beschrieben,

$$P(X_t \in S) = 1, \quad F_t(x) := P(X_t < x)$$

### Def. 69 (Markov Kette)

Eine Familie  $\{X_t\}_{t \in T}$  Zufallsgrößen heißt Markov Kette, falls gilt:

$$\begin{aligned} P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) &= P(X_{t+1} = j | X_t = i) \\ &= p_{ij}^{(t)}. \end{aligned}$$

Die Anfangsverteilung der MARKOV-Kette bezeichnen wir mit  $p_i^{(0)} = P(X_0 = i)$ .

## Definitionen (3)

**Bem.:** Wir stellen uns also vor, dass wir, beginnend im Zustand  $i_0$ , über die Zustände  $i_1, \dots, i_{t-1}$  in den Zustand  $i$  gelangt sind und nun in einen weiteren Zustand übergehen wollen. Eine Familie von Zufallsgrößen ist eine MARKOV'sche Kette, wenn für den Übergang in diesen Zustand nur der unmittelbar vorangegangene Zustand, also der Zustand  $i$ , relevant ist. (Markov-Eigenschaft)

## Definitionen (4)

### Def. 70 (homogene Markov-Kette)

Eine MARKOV-Kette heißt homogen, wenn für alle  $i, j \in S$  und für alle  $t \in T$  gilt, daß  $p_{ij}^{(t)} = p_{ij}$ , d.h. wenn die Übergangswahrscheinlichkeiten unabhängig vom jeweiligen Schritt  $t$  sind.

$p_{ij}$  heißt Übergangswahrscheinlichkeit vom Zustand  $i$  in den Zustand  $j$ .

## Definitionen (4)

### Def. 71 (Übergangsmatrix)

Die Matrix  $\mathbf{M} = (p_{ij})_{i,j \in S}$ ,

$$\mathbf{M} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \dots \\ p_{21} & p_{22} & p_{23} & \dots \\ p_{31} & p_{32} & p_{33} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

heißt Übergangsmatrix, falls

$$p_{ij} \geq 0, \quad \forall i, j \in S \text{ und } \sum_{j \in S} p_{ij} = 1 \quad \forall i \in S,$$

## Definitionen (5)

Wir werden ausschließlich homogene MARKOV-Ketten betrachten.

### Def. 72 (n-Schritt Übergangswahrscheinlichkeit)

Es sei  $\{X_t\}_{t \in T}$  eine solche homogene MARKOV-Kette. Wir definieren:

$$p_{ij}(n) := P(X_{m+n} = j \mid X_m = i).$$

Das ist die Wahrscheinlichkeit, daß man nach  $n$  Schritten aus dem Zustand  $i$  in den Zustand  $j$  gelangt.

Da die Kette homogen ist, gilt:

$$p_{ij}(n) = P(X_n = j \mid X_0 = i).$$



# Einfache Zusammenhänge (1)

Wie kann man die Matrix für die Wahrscheinlichkeiten  $p_{ij}(n)$  aus der (Ein-Schritt-)Übergangsmatrix berechnen?

# Einfache Zusammenhänge (1)

Wie kann man die Matrix für die Wahrscheinlichkeiten  $p_{ij}(n)$  aus der (Ein-Schritt-)Übergangsmatrix berechnen?

$$p_{ij}(0) = \begin{cases} 1 & \text{falls } i = j; \\ 0 & \text{sonst.} \end{cases}$$

# Einfache Zusammenhänge (1)

Wie kann man die Matrix für die Wahrscheinlichkeiten  $p_{ij}(n)$  aus der (Ein-Schritt-)Übergangsmatrix berechnen?

$$p_{ij}(0) = \begin{cases} 1 & \text{falls } i = j; \\ 0 & \text{sonst.} \end{cases}$$

$$p_{ij}(1) = p_{ij}$$

$$\begin{aligned} p_{ij}(2) &= P(X_2 = j \mid X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j, X_1 = k \mid X_0 = i) \end{aligned}$$

# Einfache Zusammenhänge (2)

Wenden die Formel der Totalen Wahrscheinlichkeit an,

- $A_i := \{X_1 = i\}$ , für alle  $i \in S$ , denn:  $\bigcup_{i \in S} A_i = \Omega$  und  $A_i \cap A_j = \emptyset$ , für alle  $i, j \in S$  mit  $i \neq j$ ;

$$\begin{aligned} p_{ij}(2) &= \sum_{k \in S} P(X_2 = j | X_1 = k, X_0 = i) \cdot P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) \cdot P(X_1 = k | X_0 = i) \end{aligned}$$

# Einfache Zusammenhänge (2)

Wenden die Formel der Totalen Wahrscheinlichkeit an,

- $A_i := \{X_1 = i\}$ , für alle  $i \in S$ , denn:  $\bigcup_{i \in S} A_i = \Omega$  und  $A_i \cap A_j = \emptyset$ , für alle  $i, j \in S$  mit  $i \neq j$ ;

$$\begin{aligned} p_{ij}(2) &= \sum_{k \in S} P(X_2 = j | X_1 = k, X_0 = i) \cdot P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) \cdot P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} p_{kj} \cdot p_{ik} = \end{aligned}$$

# Einfache Zusammenhänge (2)

Wenden die Formel der Totalen Wahrscheinlichkeit an,

- $A_i := \{X_1 = i\}$ , für alle  $i \in S$ , denn:  $\bigcup_{i \in S} A_i = \Omega$  und  $A_i \cap A_j = \emptyset$ , für alle  $i, j \in S$  mit  $i \neq j$ ;

$$\begin{aligned} p_{ij}(2) &= \sum_{k \in S} P(X_2 = j | X_1 = k, X_0 = i) \cdot P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) \cdot P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} p_{kj} \cdot p_{ik} = (\mathbf{M}^2)_{ij} \end{aligned}$$

# Einfache Zusammenhänge (3)

## Rekursion von Chapman–Kolmogorov

$$\begin{aligned}\mathbf{M}_n &= \mathbf{M}^n \\ p_{ij}(n) &= \sum_{k \in S} p_{ik}(n-m) \cdot p_{kj}(m) \\ &= \sum_{k \in S} p_{ik}(n-1) \cdot p_{kj}, \quad (m=1).\end{aligned}$$

# Einfache Zusammenhänge (4)

## Folgerung

$$P(X_n = j) = \sum_k p_{kj}(n) \cdot p_k^0.$$

**Beweis:** Seien  $\pi_j = P(X_n = j)$ ,  $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots)$  und  $\mathbf{p}^{0^T} = (p_1^0, p_2^0, \dots)$ . Es gilt

$$P(X_n = j) = \sum_k P(X_n = j, X_0 = k)$$



# Einfache Zusammenhänge (4)

## Folgerung

$$P(X_n = j) = \sum_k p_{kj}(n) \cdot p_k^0.$$

**Beweis:** Seien  $\pi_j = P(X_n = j)$ ,  $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \dots)$  und  $\mathbf{p}^{0^T} = (p_1^0, p_2^0, \dots)$ . Es gilt

$$\begin{aligned} P(X_n = j) &= \sum_k P(X_n = j, X_0 = k) \\ &= \sum_k P(X_n = j | X_0 = k) \cdot P(X_0 = k) \\ &= \sum_k p_{kj}(n) \cdot p_k^0 \\ \boldsymbol{\pi} &= \mathbf{M}^{n^T} \cdot \mathbf{p}^0, \quad \boldsymbol{\pi}^T = \mathbf{p}^{0^T} \cdot \mathbf{M}^n \end{aligned}$$

# Beispiele

## Ein-Prozessorsystem mit einer I/O–Einheit

$$S = \{1, 2\}$$

- 1: Programmstatus, in dem sich das System befindet, wenn es ein Programm abarbeitet (Prozessor aktiv)
- 2: I/O–Status, der dann angenommen wird, wenn die I/O–Einheit aktiviert wird.

Für jeden Schritt  $n$ , den das System macht, definieren wir eine Zufallsgröße  $X_n$ ,  $X_n = i$ ,  $i \in S$ .

## Ein-Prozessorsystem (2)

$X_n = 1 \implies X_{n+1} = 1$ , mit Wahrscheinlichkeit  $1 - p$

$X_n = 1 \implies X_{n+1} = 2$ , mit Wahrscheinlichkeit  $p$

$X_n = 2 \implies X_{n+1} = 1$ , mit Wahrscheinlichkeit 1

$X_n = 2 \implies X_{n+1} = 2$ , mit Wahrscheinlichkeit 0

Übergangsmatrix:

$$\mathbf{M} = \begin{pmatrix} 1-p & p \\ 1 & 0 \end{pmatrix}.$$

# Ein-Prozessorsystem (3)

Anfangsverteilung  $p_i^{(0)} = P(X_0 = i)$ :

- $p_1^{(0)} = 1$ , d.h. die erste Aktion ist mit Wahrscheinlichkeit Eins die Ausführung eines Programms;
- $p_2^{(0)} = 0$ , d.h. die erste Aktion ist mit Wahrscheinlichkeit Null die Aktivierung der I/O-Einheit.

$$\mathbf{M}_2 =$$

# Ein-Prozessorsystem (3)

Anfangsverteilung  $p_i^{(0)} = P(X_0 = i)$ :

- $p_1^{(0)} = 1$ , d.h. die erste Aktion ist mit Wahrscheinlichkeit Eins die Ausführung eines Programms;
- $p_2^{(0)} = 0$ , d.h. die erste Aktion ist mit Wahrscheinlichkeit Null die Aktivierung der I/O-Einheit.

$$\mathbf{M}_2 = \begin{pmatrix} (1-p)^2 + p & p(1-p) \\ 1-p & p \end{pmatrix}$$

# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo

### Def. 73 (Erreichbarkeit)

Ein Zustand  $j$  heißt vom Zustand  $i$  aus erreichbar, wenn es eine Zahl  $n$  gibt, so daß gilt:  $p_{ij}(n) > 0$ .

Bez.:  $i \longrightarrow j$ .

### Def. 74 (Kommunikation)

Zwei Zustände  $i$  und  $j$  kommunizieren, wenn gilt:  $i \longrightarrow j$  und  $j \longrightarrow i$ .

Wir schreiben dann:  $i \longleftrightarrow j$ .

# Klassifikation der Zustände

Die Relation „ $\longleftrightarrow$ “ ist eine Äquivalenzrelation:

- 1 Sie ist **reflexiv**. Es gilt:  $i \longleftrightarrow i$  wegen  $p_{ii}(0) = 1$ .



# Klassifikation der Zustände

Die Relation „ $\longleftrightarrow$ “ ist eine Äquivalenzrelation:

- 1 Sie ist **reflexiv**. Es gilt:  $i \longleftrightarrow i$  wegen  $p_{ii}(0) = 1$ .
- 2 Sie ist **symmetrisch**.  $i \longleftrightarrow j$  gdw.  $j \longleftrightarrow i$ .

# Klassifikation der Zustände

Die Relation „ $\longleftrightarrow$ “ ist eine Äquivalenzrelation:

- ① Sie ist **reflexiv**. Es gilt:  $i \longleftrightarrow i$  wegen  $p_{ii}(0) = 1$ .
- ② Sie ist **symmetrisch**.  $i \longleftrightarrow j$  gdw.  $j \longleftrightarrow i$ .
- ③ Sie ist **transitiv**. Es gelte  $i \longleftrightarrow j$  und  $j \longleftrightarrow k$ .

D.h. es existieren Zahlen  $m, n \geq 0$ , so dass gilt:

$$p_{ij}(m) > 0, \quad p_{jk}(n) > 0.$$

# Klassifikation der Zustände

Die Relation „ $\longleftrightarrow$ “ ist eine Äquivalenzrelation:

- ① Sie ist **reflexiv**. Es gilt:  $i \longleftrightarrow i$  wegen  $p_{ii}(0) = 1$ .
- ② Sie ist **symmetrisch**.  $i \longleftrightarrow j$  gdw.  $j \longleftrightarrow i$ .
- ③ Sie ist **transitiv**. Es gelte  $i \longleftrightarrow j$  und  $j \longleftrightarrow k$ .

D.h. es existieren Zahlen  $m, n \geq 0$ , so dass gilt:

$$p_{ij}(m) > 0, \quad p_{jk}(n) > 0.$$

Dann folgt aus Chapman–Kolmogorov

$$\begin{aligned} p_{ik}(m+n) &= \sum_{l \in S} p_{il}(m) \cdot p_{lk}(n) \\ &\geq p_{ij}(m) \cdot p_{jk}(n) > 0. \end{aligned}$$

# Klassifikation der Zustände

Nach  $m + n$  Schritten erreicht man folglich vom Zustand  $i$  aus den Zustand  $k$ . Es gilt also:  $i \longrightarrow k$ . Mit Hilfe der Symmetrieeigenschaft der Relation „ $\longleftrightarrow$ “, angewendet auf die Voraussetzung, folgt  $k \longrightarrow i$ .

## Folgerung

Es sei  $S$  der Zustandsraum einer MARKOV'schen Kette. Es gibt eine Zerlegung von  $S$  in Äquivalenzklassen bzgl. der Relation „ $\longleftrightarrow$ “.

# Klassifikation der Zustände

Die Menge der Zustände lässt sich weiter unterteilen.

## Def. 75 (wesentliche und unwesentliche Zustände)

Gibt es für einen Zustand  $i$  einen Zustand  $j$  und eine Zahl  $n \geq 0$ , so dass

$$p_{ij}(n) > 0, \text{ aber } p_{ji}(m) = 0, \quad \forall m \in \mathbb{N}$$

gilt, so heißt  $i$  unwesentlicher oder auch vorübergehender Zustand.

Andernfalls heißt  $i$  wesentlicher Zustand.

# Klassifikation der Zustände

## Beispiel

Wir betrachten den Zustandsraum  $S = \{1, 2, 3, 4\}$

und eine MARKOV- Kette mit der Übergangsmatrix

$$\mathbf{M} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Zustände 1 und 2:

# Klassifikation der Zustände

## Beispiel

Wir betrachten den Zustandsraum  $S = \{1, 2, 3, 4\}$

und eine MARKOV- Kette mit der Übergangsmatrix

$$\mathbf{M} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Zustände 1 und 2: unwesentlich. Für den Zustand 1 existiert der Zustand 3, für den gilt, daß  $p_{13}(1) = \frac{1}{2} > 0$  ist. Eine Zahl  $m$ , für die  $p_{31}(m) > 0$  ex. nicht.

Zustände 2 mit 4: analog.

# Klassifikation der Zustände

## Fortsetzung des Beispiels

Die Zustände 3 und 4 sind dagegen wesentlich.

An der Matrix **M** (vgl. folgende Folie) kann man die Klassen ablesen.

Die Elemente des Zustandsraumes sind in hier bereits so sortiert, daß die unwesentlichen Zustände vorn stehen. In der Matrix stehen in den ersten beiden Spalten im unteren Bereich nur noch Nullen. Sie zeigen an, daß man aus den durch die Zeilennummern bezeichneten Zuständen nicht mehr in die Zustände, die durch die betreffenden Spaltennummern gekennzeichnet werden, zurückkehren kann.



# Übergangsmatrix, geordnet

| Zustände     | unwesentliche | wesentliche |      |       |
|--------------|---------------|-------------|------|-------|
|              | $S_0$         | $S_1$       | ...  | $S_k$ |
| unwesentlich |               |             |      |       |
| wesentlich   | 0..0          |             | 0..0 | 0..0  |
|              | 0..0          | 0..0        |      | 0..0  |
|              | 0..0          | 0..0        | 0..0 |       |

$S_i$ : die Zustandsklassen, in die der Zustandsraum  $S$  bzgl. der Äquivalenzrelation „ $\longleftrightarrow$ “ zerlegt werden kann.

# Klassifikation der Zustände

$S_0$  ist die Klasse der unwesentlichen Zustände,

$S_i$  ( $i \geq 1$ ) sind die Klassen der wesentlichen Zustände.

Die leeren Felder sind Teilmatrizen der Matrix  $M$ , die mit Übergangswahrscheinlichkeiten bestzt sind.

Man sieht, dass in  $S_i$ ,  $i \geq 1$ , Übergänge nur innerhalb einer Zustandsklasse möglich sind.

## Def. 76 (absorbierender Zustand)

Besteht eine Äquivalenzklasse  $s_i$  bzgl. „ $\longleftrightarrow$ “ nur aus einem einzigen Zustand ( $s_i = \{j_i\}$ ), so heißt dieser Zustand absorbierender Zustand.

# Klassifikation der Markov-Kette

## Def. 77 (Irreduzibilität)

Eine MARKOV'sche Kette heißt irreduzibel oder unzerlegbar, wenn der Zustandsraum  $S$  aus genau einer Klasse besteht, und alle Zustände wesentlich sind.

$S = \{1, 2\}$ , Übergangsmatrix:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = M_n \quad \forall n \geq 1.$$

$\{X_t\}$  ist reduzibel! Zustand 1 ist absorbierend!

Zustand 2 ist unwesentlich.

# Beispiel einer irreduziblen MK

Sei  $S = \{1, 2, 3\}$ , Übergangsmatrix:

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

# Beispiel einer irreduziblen MK

Sei  $S = \{1, 2, 3\}$ , Übergangsmatrix:

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

$$\mathbf{M}_2 = \mathbf{M}^2 = \begin{pmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{3}{8} & \frac{19}{48} & \frac{11}{48} \\ \frac{1}{6} & \frac{11}{36} & \frac{19}{36} \end{pmatrix}$$

$p_{ij}(2) > 0 \quad \forall i, j \in S.$   $\{X_t\}$  ist irreduzibel!

Alle Zustände kommunizieren miteinander.

# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo

Sei  $i$  fest und

$$f_i(n) = P(X_n = i, X_{n-1} \neq i, \dots, X_1 \neq i, X_0 = i)$$

die Wahrscheinlichkeit, dass nach  $n$  Schritten erstmalig wieder der Zustand  $i$  erreicht wird. Es gilt:

$$f_i(0) := 0 \text{ und } f_i(1) = p_{ii}.$$

$B_k$ : Ereignis, erstmals nach  $k$  Schritten wieder in  $i$  ( $k = 1, \dots, n$ ).

$$B_k = \{X_k = i, X_\nu \neq i \quad \forall \nu = 1, \dots, k-1 | X_0 = i\}$$

$B_{n+1} = \{\text{System befand sich während der ersten } n \text{ Schritte nie im Zustand } i\}.$

# Rekurrente und transiente Zustände

Offenbar

$$\bigcup_{l=1}^{n+1} B_l = \Omega, \quad B_l \cap B_{l'} = \emptyset \quad (l \neq l').$$

$\Omega = \{\omega_i, i \in \mathbb{N} \cup \{\infty\}\}$  und  $\omega_i$  ist das Elementarereignis, dass genau  $i$  Schritte bis zur ersten Rückkehr erfolgen.

Dann gilt

$$p_{ii}(n) = P(X_n = i | X_0 = i)$$



# Rekurrente und transiente Zustände

Offenbar

$$\bigcup_{l=1}^{n+1} B_l = \Omega, \quad B_l \cap B_{l'} = \emptyset \quad (l \neq l').$$

$\Omega = \{\omega_i, i \in \mathbb{N} \cup \{\infty\}\}$  und  $\omega_i$  ist das Elementarereignis, dass genau  $i$  Schritte bis zur ersten Rückkehr erfolgen.

Dann gilt

$$\begin{aligned} p_{ii}(n) &= P(X_n = i | X_0 = i) \\ &= \sum_{k=1}^{n+1} P(X_n = i | B_k) \cdot P(B_k) \\ &= \sum_{k=1}^n p_{ii}(n-k) f_i(k) + \underbrace{P(X_n = i | B_{n+1})}_{=0} \cdot P(B_{n+1}) \end{aligned}$$

# Rekurrente und transiente Zustände

Wegen  $P(X_n = i | B_{n+1}) = 0$  folgt

$$p_{ii}(n) = \sum_{k=1}^n f_i(k) \cdot p_{ii}(n-k) \quad (n \geq 1).$$

Damit lässt sich  $f_i(k)$  rekursiv berechnen:

$$f_i(0) = 0, \quad f_i(1) = p_{ii}$$

$$\begin{aligned} p_{ii}(2) &= f_i(1) \cdot p_{ii}(1) + f_i(2) \cdot p_{ii}(0) \\ &= p_{ii}^2 + f_i(2) \end{aligned}$$

$$f_i(2) = p_{ii}(2) - p_{ii}^2 \quad \text{usw.}$$

$$(p_{ii}(2) = \sum_k p_{ik} p_{ki} \geq p_{ii}^2).$$

# Rekurrente und transiente Zustände

Wir bezeichnen mit

$$F_i := \sum_{j=1}^{\infty} f_i(j)$$

die Wahrscheinlichkeit, dass man irgendwann in den Zustand  $i$  zurückkehrt.

## Def. 78 (rekurrente und transiente Zustände)

Ein Zustand  $i \in S$  heißt rekurrent, wenn  $F_i = 1$  gilt. Ist dagegen  $F_i < 1$ , so heißt er transient.

# Rekurrente und transiente Zustände

## Satz

Zustand  $i$  rekurrent  $\Rightarrow$  er wird unendlich oft erreicht mit Wahrscheinlichkeit 1.

Zustand  $i$  transient  $\Rightarrow$  er kann höchstens endlich oft erreicht werden.

# Beweis des Satzes (1)

Sei  $r_i(k)$  die Wahrscheinlichkeit, dass die Markov-Kette mindestens  $k$  mal nach  $i$  zurückkehrt.

$$\begin{aligned} r_i(k) &= \sum_{n=1}^{\infty} P(\geq k \text{ mal zurück} | \text{erstmal nach } n \text{ Schritten zurück}) \cdot \\ &\quad P(\text{erstmal nach } n \text{ Schritten zurück}) \\ &= \sum_{n=1}^{\infty} r_i(k-1) f_i(n) \\ &= r_i(k-1) \sum_{n=1}^{\infty} f_i(n) = r_i(k-1) F_i \end{aligned}$$

$$\Rightarrow r_i(k) = F_i^k$$

## Beweis des Satzes (2)

Ist  $i$  rekurrent, also  $F_i = 1$ , dann  $r_i(k) = 1 \quad \forall k \in \mathbb{N}$ .

Sei  $i$  transient, d.h.  $F_i < 1$ .

Sei  $Z_i$  die Anzahl der Besuche in  $i$ .

$$P(Z_i = k) =$$

## Beweis des Satzes (2)

Ist  $i$  rekurrent, also  $F_i = 1$ , dann  $r_i(k) = 1 \quad \forall k \in \mathbb{N}$ .

Sei  $i$  transient, d.h.  $F_i < 1$ .

Sei  $Z_i$  die Anzahl der Besuche in  $i$ .

$$P(Z_i = k) = F_i^k (1 - F_i)$$

geometrische Verteilung mit Parameter  $(1 - F_i)$ .

$$\mathbf{E}Z_i = \frac{1}{1 - F_i} < \infty$$

# Rekurrente und transiente Zustände

## Satz

Ein Zustand  $i$  ist genau dann rekurrent, wenn gilt:  $\sum_{n=0}^{\infty} p_{ii}(n) = \infty$ .

Er ist genau dann transient, wenn  $\sum_{n=0}^{\infty} p_{ii}(n) < \infty$  ist.

**Beweis:** (für einen anderen Beweis siehe z.B. Mathar/Pfeifer, Satz 3.2.1) □

Erinnerung:

$$p_{ii}(n) = \sum_{k=1}^n f_i(k) \cdot p_{ii}(n-k) \quad (n \geq 1)$$

Multiplizieren diese Gleichung mit  $z^n$  und summieren über  $n$ :



# Beweis des Satzes (1)

Es gilt  $P_i(z) :=$

$$\begin{aligned} &= \sum_{n=1}^{\infty} p_{ii}(n) z^n \\ &= \sum_{n=1}^{\infty} z^n \left( \sum_{k=1}^n f_i(k) \cdot p_{ii}(n-k) \right) \\ &= z f_i(1) \cdot p_{ii}(1-1) \\ &\quad + z^2 (f_i(1) \cdot p_{ii}(2-1) + f_i(2) \cdot p_{ii}(2-2)) \\ &\quad + z^3 (f_i(1) \cdot p_{ii}(3-1) + f_i(2) \cdot p_{ii}(3-2) + f_i(3) \cdot p_{ii}(3-3)) \\ &\quad + \dots \\ &\quad + z^n (f_i(1) \cdot p_{ii}(n-1) + \dots + f_i(n) \cdot p_{ii}(0)) + \dots \end{aligned}$$

# Beweis des Satzes (2)

Es gilt

$$\begin{aligned} P_i(z) &= z f_i(1) \left( 1 + \sum_{\nu=1}^{\infty} z^{\nu} p_{ii}(\nu) \right) \\ &\quad + z^2 f_i(2) \left( 1 + \sum_{\nu=1}^{\infty} z^{\nu} p_{ii}(\nu) \right) + \dots \\ &\quad + z^n f_i(n) \left( 1 + \sum_{\nu=1}^{\infty} z^{\nu} p_{ii}(\nu) \right) + \dots \\ &= \sum_{\nu=1}^{\infty} z^{\nu} f_i(\nu) \cdot (1 + P_i(z)) \\ &= F_i(z) \cdot (1 + P_i(z)) \end{aligned}$$

## Beweis des Satzes (3)

wobei

$$F_i(z) := \sum_{\nu=1}^{\infty} z^{\nu} f_i(\nu).$$

Die Funktionen  $F_i(z)$  und  $P_i(z)$  sind analytisch für  $|z| < 1$ .

$$F_i(z) = \frac{P_i(z)}{1 + P_i(z)}, \quad P_i(z) = \frac{F_i(z)}{1 - F_i(z)}$$

$$\lim_{z \rightarrow 1} F_i(z) = F_i(1) = F_i = \sum_{\nu=1}^{\infty} f_i(\nu)$$

ist die Wahrscheinlichkeit für eine Rückkehr nach  $i$ . Sei

$$\lim_{z \rightarrow 1} P_i(z) = P_i = \sum_{n=1}^{\infty} p_{ii}(n) = \infty$$

## Beweis des Satzes (4)

Daraus folgt

$$F_i = \lim_{z \rightarrow 1} \frac{P_i(z)}{1 + P_i(z)} = 1,$$

d.h.  $i$  ist rekurrent.

Sei umgekehrt  $F_i = 1$ . Dann folgt

$$P_i = \lim_{z \rightarrow 1} P_i(z) = \frac{1}{1 - \lim_{z \rightarrow 1} F_i(z)} = \infty.$$

Der zweite Teil des Satzes ist die Kontraposition des ersten Teils.

# Transiente und rekurrente Zustände

## Folgerung

Sei  $i$  transient. dann

$$F_i = \frac{P_i}{1 + P_i},$$

d.h.  $F_i$  kann mit Hilfe von  $P_i$  ausgerechnet werden.

Diese beiden Aussagen können zum Beweis des folgenden Lemmas verwendet werden.

## Lemma

Ist ein Zustand  $i$  rekurrent (transient) und kommuniziert er mit einem Zustand  $j$  ( $i \longleftrightarrow j$ ), so ist auch der Zustand  $j$  rekurrent (transient).

# Beweis des Lemmas, Rekurrente Zustände

1. Sei  $i$  rekurrent und  $i \longleftrightarrow j$ . Dann existieren  $m, k > 0$ :  $p_{ij}(k) > 0$  und  $p_{ji}(m) > 0$ . Für alle  $n \in \mathbf{N}$  gilt:

$$\begin{aligned} p_{jj}(m+n+k) &= \sum_l \left( \sum_{k'} p_{jk'}(m) p_{k'l}(n) \right) p_{lj}(k) \\ &= \sum_l p_{jl}(m+n) p_{lj}(k) \\ &\geq p_{ji}(m) p_{ii}(n) p_{ij}(k) \quad (l = i). \end{aligned}$$

Daraus folgt (da  $i$  rekurrent)

$$\sum_{n=1}^{\infty} p_{jj}(m+n+k) \geq p_{ji}(m) p_{ij}(k) \sum_{n=1}^{\infty} p_{ii}(n) = \infty.$$

## Beweis des Lemmas (2)

2. Sei  $i \longleftrightarrow j$ . und  $i$  transient. Angenommen,  $j$  wäre rekurrent, dann wäre nach 1. auch  $i$  rekurrent. Wid.

### Folgerung

Eine irreduzible MARKOV'sche Kette mit endlich vielen Zuständen hat nur rekurrente Zustände.

## Beweis des Lemmas (2)

2. Sei  $i \longleftrightarrow j$ . und  $i$  transient. Angenommen,  $j$  wäre rekurrent, dann wäre nach 1. auch  $i$  rekurrent. Wid.

### Folgerung

Eine irreduzible MARKOV'sche Kette mit endlich vielen Zuständen hat nur rekurrente Zustände.

**Beweis:** Mindestens ein Zustand muß rekurrent sein. Da alle Zustände miteinander kommunizieren, sind alle Zustände rekurrent.

□



# Beispiel

## Random Walk, eindimensionaler Fall

Der Zustandsraum ist  $S = \mathbb{Z}$ . Die Übergangswahrscheinlichkeiten sind

$$p_{i,i+1} := p$$

$$p_{i,i-1} := 1 - p$$

$$p_{ij} := 0, \quad \text{falls } |i - j| \neq 1.$$

D.h. Übergänge zwischen Zuständen, die einen Abstand ungleich Eins zueinander haben, sind nicht möglich. Die Übergangsmatrix **M** hat folgende Gestalt:

# Random Walk, Fortsetzung

$$\mathbf{M} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \dots & 0 & p & 0 & 0 & \dots \\ \dots & 1-p & 0 & p & 0 & \dots \\ \dots & 0 & 1-p & 0 & p & \dots \\ \dots & 0 & 0 & 1-p & 0 & \dots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Offenbar kommunizieren alle Zustände miteinander. Ist somit ein Zustand rekurrent, so sind es alle. Und umgekehrt.

# Random Walk, Fortsetzung, 2

Es genügt also zu untersuchen:

$$\sum_{n=1}^{\infty} p_{00}(n).$$

Dazu siehe den Abschnitt Irrfahrten!

$$\sum_{n=1}^{\infty} p_{00}(n) = \infty, \text{ wenn } p = \frac{1}{2}.$$

# Random Walk, Fortsetzung, 3

## Random Walk, zwei- und dreidimensionaler Fall

Im zweidimensionalen Fall haben wir in jedem Zustand vier mögliche Übergänge, denen die Wahrscheinlichkeiten  $p_1, p_2, p_3$  und  $p_4$  zugeordnet werden. Die Zustände sind rekurrent, wenn  $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$  gilt.

Im dreidimensionalen Fall sind in jedem Punkt im dreidimensionalen ganzzahligen Gitter sechs Übergänge möglich. Auch wenn  $p_1 = \dots = p_6 = \frac{1}{6}$ , so sind alle Zustände transient.

Dazu siehe den Abschnitt Irrfahrten!

# Transiente Zustände

Sei der Zustand  $i$  Startzustand (fest) und

$Y_1$  = zufällige Anzahl der Schritte bis zur ersten Rückkehr nach  $i$

$Y_2$  = zufällige Anzahl der Schritte bis zur zweiten Rückkehr

$Y_k$  = zufällige Anzahl der Schritte bis zur  $k$ -ten Rückkehr

## Satz (Transiente Zustände)

$i$  transient  $\implies$  nach unendlich vielen Schritten tritt  $i$  höchstens endlich oft mit Wahrscheinlichkeit 1 ein.

**Beweis:**

$$P(Y_2 < \infty) = P(Y_2 < \infty | Y_1 < \infty) \cdot P(Y_1 < \infty) = F_i^2$$

$$P(Y_k < \infty) = F_i^k$$

**Beweis:** der Folgerung, Fortsetzung. Da  $i$  transient

$$\implies F_i < 1 \implies \sum_{k=1}^{\infty} P(Y_k < \infty) = \sum_{k=1}^{\infty} F_i^k < \infty.$$

Bezeichnen wir  $A_k = \{Y_k < \infty\}$  und  $B_n = \bigcup_{k \geq n} A_k \downarrow$ .

$B_n$ : Ereignis, dass die Anzahl der Schritte bis zur  $k$ -ten Rückkehr  $\geq n$ . Dann offenbar  $A_{k+1} \subseteq A_k$  und wegen  $\sum_{k=1}^{\infty} P(A_k) < \infty$  folgt

$$0 = \lim P(A_k) = P\left(\bigcup_{k \geq n} A_k\right) = P(\lim B_n) = \lim P(B_n) = P(B)$$

$B = \{\text{unendlich viele der } A_k, k = 1, 2, \dots, \text{treten ein}\}$

$\bar{B} = \{\text{nur endlich viele der } A_k, k = 1, 2, \dots, \text{treten ein}\}$

$$P(\bar{B}) = 1$$

# Rekurrente Zustände

## Folgerung

Sei jetzt  $i$  rekurrent, d.h.  $F_i = 1$ .  $\implies i$  wird unendlich oft erreicht.

**Beweis:** Für beliebiges  $k$  gilt:  $P(Y_k < \infty) = 1$ .

$Y = \#$  der Rückkehren nach  $i$  bei unendlich vielen Schritten.

$$\{Y_k < \infty\} \Leftrightarrow \{Y \geq k\}$$

$$P(Y = \infty) = \lim_{k \rightarrow \infty} P(Y \geq k) = \lim_{k \rightarrow \infty} P(Y_k < \infty) = 1.$$



# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo



### Def. 79 (Periode)

Ein Zustand  $i$  heißt periodisch mit der Periode  $d$ , falls  $d$  größter gemeinsamer Teiler aller der Zahlen  $n \in \mathbb{Z}^+$  ist, für die  $p_{ii}(n) > 0$  gilt. Ist  $d = 1$ , so heißt der Zustand  $i$  aperiodisch. Falls für alle Zahlen  $n \in \mathbb{Z}^+$   $p_{ii}(n) = 0$  gilt, so setzen wir  $d := \infty$ .

### Satz

Es sei  $i \in S$  ein periodischer Zustand mit Periode  $d$ . Desweiteren kommuniziere er mit einem weiteren Zustand  $j$  ( $i \longleftrightarrow j$ ). Dann hat auch der Zustand  $j$  die Periode  $d$ .

# Beweis des Satzes (1)

Sei  $i$  periodischer Zustand mit Periode  $d$ . Dann lassen sich alle Zahlen  $k$  mit  $p_{ii}(k) > 0$  durch  $k = k_0 \cdot d$ , für eine Zahl  $k_0$ , darstellen. Da die Zustände  $i$  und  $j$  miteinander kommunizieren, existieren weitere Zahlen  $n$  und  $m$ , so daß gilt:

$$p_{ij}(n) > 0 \text{ und } p_{ji}(m) > 0.$$

Nach CHAPMAN–KOLMOGOROV:

$$\begin{aligned} p_{ii}(n+m) &= \sum_{l \in S} p_{il}(n) \cdot p_{li}(m) \\ &\geq p_{ij}(n) \cdot p_{ji}(m) > 0 \end{aligned}$$

## Beweis des Satzes (2)

Folglich ist  $d$  Teiler der Summe  $n + m$ .

Es gelte nun  $p_{jj}(r) > 0$  für ein gewisses  $r$ . Dann gilt:

$$\begin{aligned} p_{ii}(n + m + r) &= \sum_{l,s \in S} p_{il}(n) \cdot p_{ls}(r) \cdot p_{si}(m) \\ &\geq p_{ij}(n) \cdot p_{jj}(r) \cdot p_{ji}(m) \\ &> 0 \end{aligned}$$

Wir stellen also fest:

$$\left. \begin{array}{l} d \text{ teilt } m + n + r \\ d \text{ teilt } m + n \end{array} \right\} \Rightarrow d \text{ teilt } r.$$

## Beweis des Satzes (3)

Folglich ist der Zustand  $j$  periodisch mit Periode  $d'$ , wobei gilt:  
 $d \leq d'$ .

Da die Relation „ $\longleftrightarrow$ “ symmetrisch ist, gilt auch:  $j \longleftrightarrow i$ . Mit der gleichen Beweisführung wie oben können wir dann zeigen, daß gilt:  
 $d' \leq d$ . Daraus folgt: Die Zustände  $i$  und  $j$  haben die gleiche Periodenlänge.

# Mittlere Rückkehrzeit (1)

Es sei nun  $i \in S$  ein rekurrenter Zustand. Wir betrachten die folgende Zufallsgröße:

$$Y : \begin{pmatrix} 1 & 2 & \dots & n & \dots \\ f_i(1) & f_i(2) & \dots & f_i(n) & \dots \end{pmatrix}.$$

Def. 80 (mittlere Rückkehrzeit in den Zustand  $i$ )

$$\mu_i := \sum_{n=1}^{\infty} n \cdot f_i(n) = \mathbf{E}Y.$$

Def. 81 (Nullrekurrenz, positive Rekurrenz)

Der Zustand  $i$  heißt positiv rekurrent, falls  $\mu_i < \infty$ . Ist  $\mu_i = \infty$ , so nennen wir den Zustand  $i$  Null-rekurrent.

# Nullrekurrenz

## Satz:

Eine endliche Markov-Kette hat keine nullrekurrenten Zustände

**Beweis:** Ang.  $C$  ist eine nullrekurrente Äquivalenzklasse. Dann gilt:

$$\sum_{j \in C} p_{ij}(n) = 1 \quad \forall i \in S, \forall n \in \mathbb{N}$$

$$\Rightarrow \sum_{j \in C} \frac{1}{n} \sum_{k=1}^n p_{ij}(k) = 1 \quad \forall i \in S$$

$$\text{aber} \quad \lim_{n \rightarrow \infty} \sum_{j \in C} \frac{1}{n} \sum_{k=1}^n p_{ij}(k) = \sum_{j \in C} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \underbrace{p_{ij}(k)}_{\rightarrow 0} = 0$$

da die letzte Summe wegen der Nullrekurrenz ( $p_{ij}(k) \rightarrow 0$ , siehe folgende Seite) endlich ist (die Summe geht schneller als linear gegen Null). Widerspruch.  $\square$

## Mittlere Rückkehrzeit (2)

Es gilt für einen beliebigen Zustand  $i$  (ohne Beweis):

- $\mu_i < \infty$  genau dann, wenn  $\lim_{n \rightarrow \infty} p_{ii}(n) > 0$ ;
- $\mu_i = \infty$  genau dann, wenn  $\lim_{n \rightarrow \infty} p_{ii}(n) = 0$ .
- Ist der Zustand  $i$  positiv rekurrent und aperiodisch, so gilt:

$$\mu_i = \frac{1}{\lim_{n \rightarrow \infty} p_{ii}(n)}.$$

### Def. 82 (Ergodische Markov-Kette)

Eine MARKOV-Kette  $\{X_t\}_{t \in T}$  heißt ergodisch, falls der Zustandsraum  $S$  nur aus positiv-rekurrenten und aperiodischen Zuständen besteht.

# Stationäre Verteilung und Ergodensatz

Erinnerung:  $\pi_j = P(X_n = j)$ ,  $\pi^T = (\pi_1, \pi_2, \dots)$

## Def. 83 (Stationäre Verteilung)

$\pi$  heißt stationär, falls  $\pi = M^T \pi$

## Ergodensatz

Sei  $\{X_t\}_{t \in T}$  eine homogene, irreduzible MARKOV-Kette. Seien außerdem alle Zustände ergodisch (positiv rekurrent und aperiodisch). Dann gilt für alle Zustände  $i, j \in S$ :

$$\pi_j := \lim_{n \rightarrow \infty} p_{ij}(n) > 0.$$

Außerdem gilt  $\mu_j = \frac{1}{\pi_j}$  und  $\pi_j$  ist eindeutig bestimmt durch:

$$\pi_j = \sum_{i=1}^{\infty} \pi_i \cdot p_{ij}. \quad \text{d.h. } \pi \text{ ist stationär.}$$



## Stationäre Verteilung

Die Grenzverteilung  $\mathbf{p} = (p_1, \dots)$  ist also stationäre oder Finalverteilung. Die stationäre Verteilung kann nach obiger Gleichung ermittelt werden.

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_j \\ \vdots \end{pmatrix} = \mathbf{M}^T \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_j \\ \vdots \end{pmatrix}.$$

## Stationäre Verteilung (2)

Also gilt:  $\mathbf{M}^T \cdot \boldsymbol{\pi} = \boldsymbol{\pi} = \lambda \cdot \boldsymbol{\pi}$  mit  $\lambda = 1$ . Eigenwertgleichung für den Eigenwert 1. Der Vektor  $\boldsymbol{\pi}$  ist Eigenvektor von  $\mathbf{M}^T$  zum Eigenwert 1.

**Bem.:**  $\mathbf{M}$  und  $\mathbf{M}^T$  haben dieselben Eigenwerte.

## Stationäre Verteilung (2)

Also gilt:  $\mathbf{M}^T \cdot \boldsymbol{\pi} = \boldsymbol{\pi} = \lambda \cdot \boldsymbol{\pi}$  mit  $\lambda = 1$ . Eigenwertgleichung für den Eigenwert 1. Der Vektor  $\boldsymbol{\pi}$  ist Eigenvektor von  $\mathbf{M}^T$  zum Eigenwert 1.

**Bem.:**  $\mathbf{M}$  und  $\mathbf{M}^T$  haben dieselben Eigenwerte.

### Folgerung

Sei  $\mathbf{M}$  die Übergangsmatrix einer MARKOV'schen Kette mit endlich vielen Zuständen (in der Form, in der die Äquivalenzklassen ablesbar sind) Dann gilt: Die Vielfachheit des Eigenwertes 1 ist gleich der Anzahl der rekurrenten Äquivalenzklassen.

# Stationäre Verteilung, Beispiel

**Beweis:** Jede Teilübergangsmatrix von Äquivalenzklassen hat den einfachen Eigenwert 1 (Finalverteilung eindeutig!) □

Wir betrachten eine MARKOV'sche Kette über  $S = \{1, 2, 3\}$  mit Übergangsmatrix

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Äquivalenzklassen:  $\{1, 2\}, \{3\}$ .

# Stationäre Verteilung, Beispiel (Fortsetzung)

Wir ermitteln die Eigenwerte:

$$\begin{aligned} 0 &= \det(M - \lambda \cdot I) \\ &= \begin{vmatrix} \frac{1}{2} - \lambda & \frac{1}{2} & 0 \\ \frac{3}{4} & \frac{1}{4} - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{vmatrix} \\ &= (1 - \lambda) \cdot \left[ \left( \frac{1}{2} - \lambda \right) \cdot \left( \frac{1}{4} - \lambda \right) - \frac{3}{8} \right] \end{aligned}$$

# Stationäre Verteilung, Beispiel (Fortsetz., 2)

Der erste Eigenwert:  $\lambda_1 = 1$ . Weiter:

$$\begin{aligned} 0 &= \left(\frac{1}{2} - \lambda\right) \cdot \left(\frac{1}{4} - \lambda\right) - \frac{3}{8} \\ &= \frac{1}{8} - \frac{3}{4}\lambda + \lambda^2 - \frac{3}{8} = \lambda^2 - \frac{3}{4}\lambda - \frac{1}{4} \\ \lambda_{2,3} &= \frac{3}{8} \pm \sqrt{\frac{9}{64} + \frac{16}{64}} = \frac{3}{8} \pm \sqrt{\frac{25}{64}} \\ \lambda_2 &= \frac{3}{8} + \frac{5}{8} = 1 \quad \lambda_3 = -\frac{1}{4} \end{aligned}$$

Also: Eigenwerte:  $\lambda_1 = \lambda_2 = 1$  und  $\lambda_3 = -\frac{1}{4}$ . Der Eigenwert 1 hat folglich die Häufigkeit 2, und somit gibt es zwei rekurrente Äquivalenzklassen.

# Stationäre Verteilung uniform?

Folgerung: Sei die Markov-Kette endlich und irreduzibel. Falls

$$\sum_i p_{ij} = \sum_j p_{ij} = 1$$

so sind die stationären Verteilungen Gleichverteilungen.

# Stationäre Verteilung uniform?

Folgerung: Sei die Markov-Kette endlich und irreduzibel. Falls

$$\sum_i p_{ij} = \sum_j p_{ij} = 1$$

so sind die stationären Verteilungen Gleichverteilungen.

**Beweis:** Es gilt für die stationäre Verteilung  $(\pi_1, \dots, \pi_n)$ :

$$\sum_i \pi_i p_{ij} = \pi_j = \pi_j \sum_i p_{ij} \quad \forall j$$

$$\sum_i (\pi_i - \pi_j) p_{ij} = 0, \quad \text{insbesondere}$$

$$\sum_i (\pi_i - \pi_{j_0}) p_{ij_0} = 0, \quad j_0 = \min_j \pi_j$$

Wegen  $(\pi_i - \pi_{j_0}) \geq 0$  folgt  $\pi_{j_0} = \pi_i \quad \forall i$ , d.h.  $\pi_i = \frac{1}{n}$ .





# Ergodensatz

Veranschaulichung von  $\lim p_{jj}(n) = \pi_j = \frac{1}{\mu_j}$

$\mu_j$ : Erwartete Anzahl der Schritte bis zur Rückkehr nach  $j$

$Y$ : zufällige Anzahl der Schritte bis zur Rückkehr nach  $j$ ,

# Ergodensatz

Veranschaulichung von  $\lim p_{jj}(n) = \pi_j = \frac{1}{\mu_j}$

$\mu_j$ : Erwartete Anzahl der Schritte bis zur Rückkehr nach  $j$

$Y$ : zufällige Anzahl der Schritte bis zur Rückkehr nach  $j$ ,

$$Y \sim \text{Geo}(\pi_j) \quad (\text{etwa})$$

$$\mu_j = \mathbf{E}Y = \frac{1}{\pi_j}$$

# Ergodensatz

Veranschaulichung von  $\lim p_{jj}(n) = \frac{1}{\mu_j}$

$\{X_t\}$ : homogene Markov'sche Kette

$j$ : rekurrenter Zustand,  $X_0 = j$  ( $j$  fest).

$$Y_k = \begin{cases} 1, & \text{falls } X_k = j \\ 0, & \text{sonst.} \end{cases}$$

$$P(Y_k = 1) = p_{jj}(k), \quad \mathbf{E} Y_k = p_{jj}(k)$$

Anzahl der Wiederkehrzeitpunkte im Zeitraum  $1, \dots, N$

$$\sum_{k=1}^N Y_k = k_N.$$

# Ergodensatz

Beobachtete mittlere Anzahl der Wiederkehrpunkte pro Schritt (im Zeitraum  $1, \dots, N$ )

$$\begin{aligned}\frac{k_N}{N} &\approx \mathbf{E} \frac{k_N}{N} = \frac{1}{N} \mathbf{E} \left( \sum_{k=1}^N Y_k \right) = \frac{1}{N} \sum_{k=1}^N \mathbf{E} Y_k \\ &= \frac{1}{N} \sum_{k=1}^N p_{jj}(k)\end{aligned}$$

Mittlere beobachtete Wiederkehrzeit im Zeitraum  $1, \dots, N$

$$\frac{N}{k_N} \rightarrow \mu_j$$

# Ergodensatz

 $\implies$ 

$$\frac{1}{N} \sum_{k=1}^N p_{jj}(k) \xrightarrow{N \rightarrow \infty} \frac{1}{\mu_j}$$

Andererseits:

$$\lim_{n \rightarrow \infty} p_{jj}(k) = \pi_j \implies \frac{1}{N} \sum_{k=1}^N p_{jj}(k) \xrightarrow{N \rightarrow \infty} \pi_j = \frac{1}{\mu_j}.$$

# Ergodensatz, Beispiel

## Ein-Prozessorsystem mit mehreren E/A-Einheiten.

Ein Programm, das sich in der CPU befindet, geht mit Wahrscheinlichkeit  $q_i$  in die I/O-Einheit  $i$  über, oder endet (mit Wahrscheinlichkeit  $q_0$ ) und macht Platz für ein neues Programm in der CPU.

$$\mathbf{M} = \begin{pmatrix} q_0 & q_1 & \dots & q_m \\ 1 & 0 & \dots & 0 \\ \dots & & & \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

# Ergodensatz, Beispiel

## Ein-Prozessorsystem mit mehreren E/A-Einheiten.

Ein Programm, das sich in der CPU befindet, geht mit Wahrscheinlichkeit  $q_i$  in die I/O-Einheit  $i$  über, oder endet (mit Wahrscheinlichkeit  $q_0$ ) und macht Platz für ein neues Programm in der CPU.

$$\mathbf{M} = \begin{pmatrix} q_0 & q_1 & \dots & q_m \\ 1 & 0 & \dots & 0 \\ \dots & & & \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

Frage: Ist die zugehörige Markov-Kette irreduzibel?

# Ergodensatz, Beispiel (2)

## Ein-Prozessorsystem (Fortsetzung)

$$\mathbf{M}_2 = \mathbf{M}^2 = \begin{pmatrix} q_0^2 + \sum_{i=1}^m q_i & q_0 q_1 & \dots & q_0 q_m \\ q_0 & q_1 & \dots & q_m \\ .. & & & \\ q_0 & q_1 & \dots & q_m \end{pmatrix}$$



# Ergodensatz, Beispiel (2)

## Ein-Prozessorsystem (Fortsetzung)

$$\mathbf{M}_2 = \mathbf{M}^2 = \begin{pmatrix} q_0^2 + \sum_{i=1}^m q_i & q_0 q_1 & \dots & q_0 q_m \\ q_0 & q_1 & \dots & q_m \\ .. & & & \\ q_0 & q_1 & \dots & q_m \end{pmatrix}$$

also  $p_{ij}(2) > 0 \quad \forall i, j \implies X_t$  irreduzibel.

# Ein-Prozessorsystem

## Stationäre Verteilung

### Ein-Prozessorsystem (Fortsetzung, 2)

$$\mathbf{M}^T \begin{pmatrix} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_m \end{pmatrix} = \begin{pmatrix} \pi_0 q_0 + \sum_{i=1}^m \pi_i \\ \pi_0 q_1 \\ \vdots \\ \pi_0 q_m \end{pmatrix} = \begin{pmatrix} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_m \end{pmatrix}$$

# Ein-Prozessorsystem

## Stationäre Verteilung

$$q_0\pi_0 + 1 - \pi_0 = \pi_0$$

$$2\pi_0 - q_0\pi_0 = 1$$

$$\pi_0(2 - q_0) = 1$$

$$\pi_0 = \frac{1}{2 - q_0}$$

$$\pi_i = \pi_0 q_i = \frac{q_i}{2 - q_0}, \quad i = 1, \dots, m$$

# Ein-Prozessorsystem

## Stationäre Verteilung

$$q_0\pi_0 + 1 - \pi_0 = \pi_0$$

$$2\pi_0 - q_0\pi_0 = 1$$

$$\pi_0(2 - q_0) = 1$$

$$\pi_0 = \frac{1}{2 - q_0}$$

$$\pi_i = \pi_0 q_i = \frac{q_i}{2 - q_0}, \quad i = 1, \dots, m$$

$$\sum_{i=0}^m \pi_i = \frac{1}{2 - q_0} + \sum_{i=1}^m \frac{q_i}{2 - q_0} = \frac{1}{2 - q_0} + \frac{1 - q_0}{2 - q_0} = 1.$$

# Multiprozessorsystem

## Multiprozessorsystem

Ein "Job" (oder ein Prozessor) greift zufällig auf bestimmte Speichermodule zu.

Er wird bedient, wenn der angeforderte Speichermodul frei ist, sonst muß er warten.

Die Zeit für einen Speicherzugriff sei konstant und für alle Speichermodule gleich.

Neue Anforderungen beginnen sofort nach Abarbeitung der alten.  
 $m$  "Jobs",  $n$  Speichermodule.

# Multiprozessorsystem

## Multiprozessorsystem (2)

$N_i$ : Anzahl der "Jobs" (Wartenden) am Speichermodul  $M_i$   
(Bedienplätze) (wartend oder in Arbeit),  $i = 1, \dots, n$

Zustandsraum:

$$S = \{(N_1, N_2, \dots, N_n) \in \mathbb{Z}^+ : \sum_i N_i = m\}$$

Bsp.:  $m = n = 2$ :  $S = \{(1, 1), (0, 2), (2, 0)\}$

$q_1$ : Wahrscheinlichkeit, 1. Speichermodul wird angefordert

$q_2$ : Wahrscheinlichkeit, 2. Speichermodul wird angefordert

# Multiprozessorsystem (3)

## Übergangsmatrix

$$\mathbf{M} = \begin{pmatrix} 2q_1q_2 & q_2^2 & q_1^2 \\ q_1 & q_2 & 0 \\ q_2 & 0 & q_1 \end{pmatrix}$$

# Multiprozessorsystem (3)

## Übergangsmatrix

$$\mathbf{M} = \begin{pmatrix} 2q_1q_2 & q_2^2 & q_1^2 \\ q_1 & q_2 & 0 \\ q_2 & 0 & q_1 \end{pmatrix}$$

## Stationäre Verteilung

$$\begin{aligned} \pi^T \mathbf{M} &= \pi^T \\ (\pi_1, \pi_2, \pi_3) \begin{pmatrix} 2q_1q_2 & q_2^2 & q_1^2 \\ q_1 & q_2 & 0 \\ q_2 & 0 & q_1 \end{pmatrix} &= (\pi_1, \pi_2, \pi_3) \end{aligned}$$



# Multiprozessorsystem (4)

Stationäre Verteilung (Fortsetz.)

$$\pi_1 \cdot 2q_1q_2 + \pi_2q_1 + \pi_3q_2 = \pi_1$$

$$\pi_1 \cdot q_2^2 + \pi_2 \cdot q_2 + \pi_3 \cdot 0 = \pi_2$$

$$\pi_1 \cdot q_1^2 + \pi_2 \cdot 0 + \pi_3 \cdot q_1 = \pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

# Multiprozessorsystem (4)

Stationäre Verteilung (Fortsetz.)

$$\pi_1 \cdot 2q_1q_2 + \pi_2q_1 + \pi_3q_2 = \pi_1$$

$$\pi_1 \cdot q_2^2 + \pi_2 \cdot q_2 + \pi_3 \cdot 0 = \pi_2$$

$$\pi_1 \cdot q_1^2 + \pi_2 \cdot 0 + \pi_3 \cdot q_1 = \pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_1 \cdot q_2^2 = \pi_2(1 - q_2)$$

$$\pi_1 \cdot q_1^2 = \pi_3(1 - q_1)$$

# Multiprozessorsystem (4)

Stationäre Verteilung (Fortsetz.)

$$\pi_1 \cdot 2q_1q_2 + \pi_2q_1 + \pi_3q_2 = \pi_1$$

$$\pi_1 \cdot q_2^2 + \pi_2 \cdot q_2 + \pi_3 \cdot 0 = \pi_2$$

$$\pi_1 \cdot q_1^2 + \pi_2 \cdot 0 + \pi_3 \cdot q_1 = \pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_1 \cdot q_2^2 = \pi_2(1 - q_2)$$

$$\pi_2 = \frac{q_2^2}{1 - q_2} \cdot \pi_1$$

$$\pi_1 \cdot q_1^2 = \pi_3(1 - q_1)$$

$$\pi_3 = \frac{q_1^2}{1 - q_1} \cdot \pi_1$$

# Multiprozessorsystem (4)

Stationäre Verteilung (Fortsetz.)

$$\pi_1 \cdot 2q_1q_2 + \pi_2q_1 + \pi_3q_2 = \pi_1$$

$$\pi_1 \cdot q_2^2 + \pi_2 \cdot q_2 + \pi_3 \cdot 0 = \pi_2$$

$$\pi_1 \cdot q_1^2 + \pi_2 \cdot 0 + \pi_3 \cdot q_1 = \pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_1 \cdot q_2^2 = \pi_2(1 - q_2) \qquad \pi_2 = \frac{q_2^2}{1 - q_2} \cdot \pi_1$$

$$\pi_1 \cdot q_1^2 = \pi_3(1 - q_1) \qquad \pi_3 = \frac{q_1^2}{1 - q_1} \cdot \pi_1$$

$$\pi_1 = \frac{1}{1 + \frac{q_1^2}{1 - q_1} + \frac{q_2^2}{1 - q_2}} = \frac{q_1q_2}{1 - 2q_1q_2}$$

# Multiprozessorsystem (5)

$X$  : # erledigten Speicherplatz-Anforderungen pro Zyklus im stationären Zustand:

$$(X|(1, 1)) = 2$$

$$(X|(2, 0)) = 1$$

$$(X|(0, 2)) = 1$$

$$\begin{aligned} \mathbf{E}X &= 2 \cdot \pi_1 + 1 \cdot \pi_2 + 1 \cdot \pi_3 \\ &= \left(2 + \frac{q_1^2}{1 - q_1} + \frac{q_2^2}{1 - q_2}\right) \pi_1 = \frac{1 - q_1 q_2}{1 - 2q_1 q_2} \end{aligned}$$

# Multiprozessorsystem (5)

$X$  : # erledigten Speicherplatz-Anforderungen pro Zyklus im stationären Zustand:

$$(X|(1, 1)) = 2$$

$$(X|(2, 0)) = 1$$

$$(X|(0, 2)) = 1$$

$$\begin{aligned}\mathbf{E}X &= 2 \cdot \pi_1 + 1 \cdot \pi_2 + 1 \cdot \pi_3 \\ &= \left(2 + \frac{q_1^2}{1 - q_1} + \frac{q_2^2}{1 - q_2}\right)\pi_1 = \frac{1 - q_1 q_2}{1 - 2q_1 q_2}\end{aligned}$$

$$q_1 = q_2 = \frac{1}{2}: \quad \mathbf{E}X = \frac{3}{2}. \quad \text{maximal möglicher Wert.}$$

# Betriebssystem

Das Betriebssystem schalte zwischen den Zuständen:

- 1: Benutzerprogramm aktiv
- 2: Scheduler aktiv
- 3: Operatorkommunikation aktiv
- 4: Nullprozess

$$M = \begin{pmatrix} 0.90 & 0.04 & 0.05 & 0.01 \\ 0.94 & 0.00 & 0.05 & 0.01 \\ 0.85 & 0.10 & 0.04 & 0.01 \\ 0.75 & 0.00 & 0.05 & 0.20 \end{pmatrix}$$

$$\pi = \begin{pmatrix} 0.897 \\ 0.041 \\ 0.05 \\ 0.012 \end{pmatrix}$$

$\pi$  ist stationäre Verteilung. (ÜA)

# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo



## Ruin des Spielers

Zwei Spieler werfen abwechselnd eine (nicht manipulierte) Münze. Fällt Kopf, so erhält Spieler A den vereinbarten Einsatz (1 Euro) von Spieler B, anderenfalls erhält Spieler B denselben Einsatz von Spieler A. Zu Beginn des Spieles besitzt A  $a$  Euro und B  $b$  Euro. Das Spiel wird solange fortgesetzt, bis einer der beiden Spieler kein Geld mehr besitzt.

# Ruin des Spielers (Fortsetzung)

Zustände:  $S = \{0, 1, \dots, N\}$ ,  $N = a + b$ .

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \dots & 0 & \dots & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \dots & 0 & \dots & 0 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \dots & 0 & \dots & 1 \end{pmatrix}$$

Frage: Wie groß ist die Wahrscheinlichkeit des Ruins von Spieler A  
bzw. B?

## Ruin des Spielers (Fortsetzung, 2)

Sei  $E_i$  das Ereignis, daß ein Spieler, der genau  $i$  Euro besitzt, ruiniert wird und sei  $p_i = P(E_i)$ .

1. Die Übergangswktn. sind

$$p_{i,i+1} = p_{i,i-1} = \frac{1}{2}$$

und offenbar ist  $p_0 = 1$  und  $p_N = 0$ .

2. Satz der totalen Wkt.: Es gilt für alle  $i, i = 0, \dots, N$ :

$$\begin{aligned} p_i = P(E_i) &= P(E_i | \text{Übergang nach } i-1) \cdot p_{i,i-1} + \\ &\quad P(E_i | \text{Übergang nach } i+1) \cdot p_{i,i+1} \end{aligned}$$

# Ruin des Spielers (Fortsetzung, 3)

$$p_i = \frac{1}{2}p_{i-1} + \frac{1}{2}p_{i+1} \quad 2p_i = p_{i-1} + p_{i+1}$$

$$p_i - p_{i-1} = p_{i+1} - p_i =: d$$

$$\begin{aligned} p_i - p_0 &= \underbrace{p_i - p_{i-1}}_{=d} + \underbrace{p_{i-1} - p_{i-2}}_{=d} + p_{i-2} - + \cdots - p_1 \\ &\quad + \underbrace{p_1 - p_0}_{=d} \end{aligned}$$

$$p_i - 1 = i \cdot d$$

$$p_i = 1 + i \cdot d, \quad \text{insbesondere}$$

$$p_N = 1 + N \cdot d$$

$$d = -\frac{1}{N}, \quad N = a + b$$

# Ruin des Spielers (Fortsetzung, 4)

3.

$$p_i = 1 - i \cdot \frac{1}{a+b} = \frac{a+b-i}{a+b}$$

$$p_a = \frac{b}{a+b}, \quad p_b = \frac{a}{a+b}$$

4.  $a = b : p_a = p_b = \frac{1}{2}$   
 $a \gg b : p_a \approx 0, p_b \approx 1.$

3 Klassen von Zuständen:

# Ruin des Spielers (Fortsetzung, 4)

3.

$$p_i = 1 - i \cdot \frac{1}{a+b} = \frac{a+b-i}{a+b}$$

$$p_a = \frac{b}{a+b}, \quad p_b = \frac{a}{a+b}$$

4.  $a = b : p_a = p_b = \frac{1}{2}$   
 $a \gg b : p_a \approx 0, p_b \approx 1.$

3 Klassen von Zuständen:

$T = \{1, \dots, N-1\}$ : unwesentliche Zustände

$S_1 = \{0\}, S_2 = \{N\}$ : absorbierende Zustände

$$T^c := S_1 \cup S_2$$

# Ruin des Spielers (Fortsetzung, 5)

Umordnung von **M**:

$$\mathbf{M}^* = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{P} \end{pmatrix}$$

$$\mathbf{Q} = (p_{ij}; i, j \in T) \quad \mathbf{P} = (p_{ij}; i, j \in T^c)$$

$$\mathbf{R} = (p_{ik}; i \in T, k \in T^c)$$

Übergang von  $i \in T$  nach  $k \in T^c$  einschrittig oder nach Übergängen innerhalb von  $T$  und anschließendem Übergang von  $T$  nach  $k$ .

## Ruin des Spielers (Fortsetzung, 6)

$u_{ik}$ : Wkt. von  $i \in T$  (irgendwann) nach  $k \in T^c$  zu kommen

$$u_{ik} = \sum_{j \in T} Q_{ij} u_{jk} + p_{ik}, \quad Q_{ij} = p_{ij}$$



## Ruin des Spielers (Fortsetzung, 6)

$u_{ik}$ : Wkt. von  $i \in T$  (irgendwann) nach  $k \in T^c$  zu kommen

$$u_{ik} = \sum_{j \in T} Q_{ij} u_{jk} + p_{ik}, \quad Q_{ij} = p_{ij}$$

$\mathbf{U} = (U_{ik})_{i \in T, k \in T^c}$ : Absorptionswahrscheinlichkeit von Zustand  $k$  wenn wir in  $i$  starten.

$$\mathbf{U} = \mathbf{Q}\mathbf{U} + \mathbf{R}, \quad \text{Rekursionsformel}$$

$$\mathbf{U} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R}$$

Die Matrix  $(\mathbf{I} - \mathbf{Q})^{-1}$  existiert, falls  $T$  endlich!

Lit.: Resnick, S.I. Adventures in Stochastic Processes, 1992

$(\mathbf{I} - \mathbf{Q})^{-1}$  kann interpretiert werden als: erwartete Anzahl der Besuche in Zustand  $j \in T$  wenn in Zustand  $i$  gestartet wird.

# Ruin des Spielers (Fortsetzung, 7)

hier:

$$(\mathbf{I} - \mathbf{Q})\mathbf{U} = \mathbf{R}$$

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 & \cdots & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ & & & -\frac{1}{2} & 1 & -\frac{1}{2} \\ & & & 0 & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} u_{10} & u_{1N} \\ u_{20} & u_{2N} \\ u_{30} & u_{3N} \\ \vdots & \vdots \\ u_{N-2,0} & u_{N-2,N} \\ u_{N-1,0} & u_{N-1,N} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

# Ruin des Spielers (Fortsetzung, 8)

$$\begin{array}{rcll}
 u_{1,0} & -\frac{1}{2}u_{2,0} & & = \frac{1}{2} \\
 -\frac{1}{2}u_{1,0} & +u_{2,0} & -\frac{1}{2}u_{3,0} & = 0 \\
 & -\frac{1}{2}u_{2,0} & +u_{3,0} & -\frac{1}{2}u_{4,0} = 0 \\
 & & \dots & \\
 & & -\frac{1}{2}u_{N-3,0} & +u_{N-2,0} & -\frac{1}{2}u_{N-1,0} = 0 \\
 & & & -\frac{1}{2}u_{N-2,0} & +u_{N-1,0} = 0
 \end{array}$$

# Ruin des Spielers (Fortsetzung, 9)

$N - 1$ . Gleichung (1. **U**-Spalte)

$$u_{N-1,0} = \frac{1}{2} u_{N-2,0}$$

$N - 2$ . Gleichung (1. **U**-Spalte)

$$\begin{aligned} -\frac{1}{2} u_{N-3,0} + u_{N-2,0} - \frac{1}{2} u_{N-1,0} &= 0 \\ u_{N-2,0} - \frac{1}{4} u_{N-2,0} &= \frac{1}{2} u_{N-3,0} \\ \frac{3}{4} u_{N-2,0} &= \frac{1}{2} u_{N-3,0} \\ u_{N-2,0} &= \frac{2}{3} u_{N-3,0} \end{aligned}$$

# Ruin des Spielers (Fortsetzung, 10)

$N - 3$ . Gleichung (1. **U**-Spalte)

$$\begin{aligned}
 -\frac{1}{2}u_{N-4,0} + u_{N-3,0} - \frac{1}{2}u_{N-2,0} &= 0 \\
 u_{N-3,0} - \frac{1}{3}u_{N-3,0} &= \frac{1}{2}u_{N-4,0} \\
 \frac{2}{3}u_{N-3,0} &= \frac{1}{2}u_{N-4,0} \\
 u_{N-3,0} &= \frac{3}{4}u_{N-4,0}
 \end{aligned}$$

$N - i$ . Gleichung (1. **U**-Spalte)

$$u_{N-i,0} = \frac{i}{i+1}u_{N-(i+1),0}, \quad i = 1, \dots, N-2$$

# Ruin des Spielers (Fortsetzung, 11)

1. Gleichung:

$$u_{1,0} - \frac{1}{2}u_{2,0} = \frac{1}{2}$$

Da

$$u_{2,0} = u_{N-(N-2),0} = \frac{N-2}{N-1}u_{N-(N-1),0} = \frac{N-2}{N-1}u_{1,0}$$

folgt

# Ruin des Spielers (Fortsetzung, 12)

$$u_{1,0} - \frac{1}{2} \frac{N-2}{N-1} u_{1,0} = \frac{1}{2}$$

$$u_{1,0} \left( 1 - \frac{N-2}{2(N-1)} \right) = \frac{1}{2}$$

$$u_{1,0} \frac{N}{2(N-1)} = \frac{1}{2}$$

$$u_{1,0} = \frac{N-1}{N} = 1 - \frac{1}{N}$$

$$u_{2,0} = \frac{N-2}{N-1} u_{1,0} = \frac{N-2}{N-1} \cdot \frac{N-1}{N} = \frac{N-2}{N} = 1 - \frac{2}{N}$$

$$u_{N-i,0} = \frac{N-i}{N} = 1 - \frac{i}{N}, \quad i = 1, 2, \dots, N-1.$$

# Münzwurfspiel (1), vgl. ÜA 14

Seien die Zustände 000, 001, 010, 011, 100, 101, 110 und 111 nacheinander mit 1-8 bezeichnet. Dann hat die Übergangsmatrix die Gestalt (wir tragen nur Einträge ein, die nicht Null sind)

$$M := \frac{1}{2} \begin{pmatrix} 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & & & 1 & 1 & & \\ & & & & & & 1 & 1 \\ 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & & & 1 & 1 & & \\ & & & & & & 1 & 1 \end{pmatrix}$$



# Münzwurfspiel (1), vgl. ÜA 14

Seien die Zustände 000, 001, 010, 011, 100, 101, 110 und 111 nacheinander mit 1-8 bezeichnet. Dann hat die Übergangsmatrix die Gestalt (wir tragen nur Einträge ein, die nicht Null sind)

$$M := \frac{1}{2} \begin{pmatrix} 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & & & 1 & 1 & & \\ & & & & & & 1 & 1 \\ 1 & 1 & & & & & & \\ & & 1 & 1 & & & & \\ & & & & 1 & 1 & & \\ & & & & & & 1 & 1 \end{pmatrix}$$

Die Markov-Kette besteht aus einer Äquivalenzklasse, ist irreduzibel und aperiodisch (Diagonaleintrag  $\neq 0$ ). Alle Zustände sind positiv rekurrent.

# Münzwurfspiel (2)

vgl. ÜA 14

Seien die Zustände 000, 001, 010, 011, 100, 101, 110 und 111 nacheinander mit 1-8 bezeichnet.

Es existiert eine stationäre Verteilung

Die Markov-Kette besteht aus einer Äquivalenzklasse, ist irreduzibel und aperiodisch (Diagonaleintrag  $\neq 0$ ). Alle Zustände sind positiv rekurrent.

Berechnung der stationären Verteilung

Lösung des linearen Gleichungssystems

$$\mathbf{M}^T \mathbf{p} = \mathbf{p} \quad \longrightarrow \quad (\mathbf{M}^T - \mathbf{I}) \mathbf{p} = \mathbf{0}$$

# Münzwurfspiel (3)

$$\frac{1}{2} \begin{pmatrix} -1 & & & & & & & 1 \\ 1 & -2 & & & & & & 1 \\ & 1 & -2 & & & & & 1 \\ & 1 & & -2 & & & & 1 \\ & & 1 & & -2 & & & 1 \\ & & 1 & & & -2 & 1 \\ & & & 1 & & & -2 & 1 \\ & & & 1 & & & & -1 \end{pmatrix} \mathbf{p} = \mathbf{0}$$

Stationäre Verteilung

# Münzwurfspiel (3)

$$\frac{1}{2} \begin{pmatrix} -1 & & & & & & & 1 \\ 1 & -2 & & & & & & 1 \\ & 1 & -2 & & & & & 1 \\ & 1 & & -2 & & & & 1 \\ & & 1 & & -2 & & & 1 \\ & & 1 & & & -2 & 1 \\ & & & 1 & & & -2 & 1 \\ & & & 1 & & & & -1 \end{pmatrix} \mathbf{p} = \mathbf{0}$$

Stationäre Verteilung ist die Gleichverteilung.

# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

Zustände 7 und 4:

# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

Zustände 7 und 4: absorbierend,

# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

Zustände 7 und 4: absorbierend,  
andere Zustände:



# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

Zustände 7 und 4: absorbierend,  
andere Zustände: unwesentlich

# Münzwurfspiel (4)

vgl. ÜA 14

Spiel: 7 (110) gegen 4 (011)

Zustände 7 und 4: absorbierend,  
andere Zustände: unwesentlich

$$M_{4,7} := \frac{1}{2} \begin{pmatrix} 1 & 1 & & & & & \\ & & 1 & 1 & & & \\ & & & & 1 & 1 & \\ & & & & & 2 & \\ 1 & 1 & & & & & \\ & & 1 & 1 & & & \\ & & & & & 2 & \\ & & & & & 1 & 1 \end{pmatrix}$$

$$M'_{4,7} := \frac{1}{2} \begin{pmatrix} 1 & 1 & & & & & \\ & & 1 & & & & 1 \\ & & & 1 & 1 & & \\ 1 & 1 & & & & & \\ & & 1 & & & & 1 \\ & & & & & 1 & 1 \\ \hline & & & & & 2 & \\ & & & & & & 2 \end{pmatrix}$$

wobei  $M'_{4,7}$  die umgeordnete Matrix ist.

# Münzwurfspiel (5)

$$\mathbf{M}'_{4,7} = \frac{1}{2} \left( \begin{array}{cccccc|cc} 1 & 1 & & & & & & \\ & & 1 & & & & 1 & \\ & & & 1 & 1 & & & \\ 1 & 1 & & & & & & \\ & & 1 & & & & 1 & \\ & & & & & 1 & & 1 \\ \hline & & & & & & 2 & \\ & & & & & & & 2 \end{array} \right) = \begin{pmatrix} \mathbf{Q}_{4,7} & \mathbf{R} \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix}$$

$$\mathbf{Q}_{4,7} = \frac{1}{2} \begin{pmatrix} 1 & 1 & & & \\ & & 1 & & \\ & & & 1 & 1 \\ 1 & 1 & & & \\ & & 1 & & \\ & & & & 1 \end{pmatrix} \quad \mathbf{R} = \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Das Gleichungssystem

$$(\mathbf{I} - \mathbf{Q})\mathbf{U} = \mathbf{R}$$

ist zu lösen.

In unserem Beispiel ist die Gewinnwahrscheinlichkeit von 110 gegen 011: 0.25.

# Irrfahrten

## Irrfahrt auf der Geraden

Zustände:  $k \in \mathbb{Z}$ , Anfangszustand: 0

Bewegung: ein Schritt nach rechts mit Wkt.  $p$  oder nach links mit

Wkt.  $q = 1 - p$

$$p_{k,k+1} = p = 1 - p_{k,k-1}; \quad p_{ij} = 0, \text{ falls } |i - j| \neq 1$$

$$M = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & & \\ & 0 & q & 0 & p & 0 \\ & & 0 & q & 0 & p & 0 \\ & & & & \cdot & \cdot & \cdot \end{pmatrix}$$

# Irrfahrten, Fortsetzung, 1

$A_{n,k}$ : Ereignis, nach  $n$  Schritten im Zustand  $k$  zu sein

$$D_{n,k} := P(A_{n,k}), \quad \Omega_{n,k} = A_{n-1,k-1} \cup A_{n-1,k+1}$$

Satz der totalen Wkt. ( $k = -n, \dots, n$ ):

$$\begin{aligned} D_{n,k} &= P(A_{n,k}) \\ &= P(A_{n,k} | A_{n-1,k-1}) \cdot P(A_{n-1,k-1}) + \\ &\quad P(A_{n,k} | A_{n-1,k+1}) \cdot P(A_{n-1,k+1}) \\ &= pD_{n-1,k-1} + qD_{n-1,k+1} \end{aligned}$$

# Irrfahrten, Fortsetzung, 1

$A_{n,k}$ : Ereignis, nach  $n$  Schritten im Zustand  $k$  zu sein

$$D_{n,k} := P(A_{n,k}), \quad \Omega_{n,k} = A_{n-1,k-1} \cup A_{n-1,k+1}$$

Satz der totalen Wkt. ( $k = -n, \dots, n$ ):

$$\begin{aligned} D_{n,k} &= P(A_{n,k}) \\ &= P(A_{n,k} | A_{n-1,k-1}) \cdot P(A_{n-1,k-1}) + \\ &\quad P(A_{n,k} | A_{n-1,k+1}) \cdot P(A_{n-1,k+1}) \\ &= pD_{n-1,k-1} + qD_{n-1,k+1} \\ &= \begin{cases} \left(\frac{n+k}{2}\right) p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} & \text{falls } k = -n, -n+2, \dots, n \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

# Irrfahrten, Fortsetzung, 2

Explizite Formel:

$$D_{n,k} = \begin{cases} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} & \text{falls } k = -n, -n+2, \dots, n \\ 0 & \text{sonst} \end{cases}$$



# Irrfahrten, Fortsetzung, 3

In den Zustand  $k$  gelangt man in genau  $n$  Schritten, indem man  $\frac{n+k}{2}$  mal nach rechts und  $\frac{n-k}{2}$  mal nach links geht.

Es gibt genau  $\binom{n}{\frac{n+k}{2}}$  Möglichkeiten die Zeitpunkte für einen Schritt nach rechts auszuwählen.

Insbesondere

$$D_{2n,0} = \binom{2n}{n} p^n q^n.$$

Abschätzung: Stirling'sche Formel

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

# Irrfahrten, Fortsetzung, 4

Damit

$$\binom{2n}{n} = \frac{(2n)!}{n!n!}$$

# Irrfahrten, Fortsetzung, 4

Damit

$$\begin{aligned}\binom{2n}{n} &= \frac{(2n)!}{n!n!} \\ &\sim \frac{\sqrt{2\pi 2n} \left(\frac{2n}{e}\right)^{2n} e^{\frac{1}{12 \cdot 2n}}}{\left(\sqrt{2\pi n} \left(\frac{n}{e}\right)^n\right)^2 \left(e^{\frac{1}{12n}}\right)^2} \\ &= \frac{1}{\sqrt{\pi n}} 2^{2n} e^{-\frac{3}{4n}}\end{aligned}$$

# Irrfahrten, Fortsetzung, 4

Damit

$$\begin{aligned}
 \binom{2n}{n} &= \frac{(2n)!}{n!n!} \\
 &\sim \frac{\sqrt{2\pi 2n} \left(\frac{2n}{e}\right)^{2n} e^{\frac{1}{12 \cdot 2n}}}{\left(\sqrt{2\pi n} \left(\frac{n}{e}\right)^n\right)^2 \left(e^{\frac{1}{12n}}\right)^2} \\
 &= \frac{1}{\sqrt{\pi n}} 2^{2n} e^{-\frac{3}{4n}}
 \end{aligned}$$

$$p = q = \frac{1}{2} : \quad D_{2n,0} \sim \frac{1}{\sqrt{\pi n}} e^{-\frac{3}{4n}}$$

# Irrfahrten, Fortsetzung, 4

Damit

$$\begin{aligned}
 \binom{2n}{n} &= \frac{(2n)!}{n!n!} \\
 &\sim \frac{\sqrt{2\pi 2n} \left(\frac{2n}{e}\right)^{2n}}{\left(\sqrt{2\pi n} \left(\frac{n}{e}\right)^n\right)^2} \frac{e^{\frac{1}{12 \cdot 2n}}}{\left(e^{\frac{1}{12n}}\right)^2} \\
 &= \frac{1}{\sqrt{\pi n}} 2^{2n} e^{-\frac{3}{4n}}
 \end{aligned}$$

$$p = q = \frac{1}{2} : \quad D_{2n,0} \sim \frac{1}{\sqrt{\pi n}} e^{-\frac{3}{4n}}$$

$$p \neq q : \quad D_{2n,0} \sim \frac{1}{\sqrt{\pi n}} 4^n p^n (1-p)^n e^{-\frac{3}{4n}}.$$

# Irrfahrten, Fortsetzung, 5

Mittlere Rückkehrhäufigkeit:

$$\sum_{n=1}^{\infty} D_{2n,0} \sim \begin{cases} \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty & (p = \frac{1}{2}) \\ \sum_{n=1}^{\infty} \frac{(4p(1-p))^n}{\sqrt{\pi n}} < \infty & (p \neq \frac{1}{2}) \end{cases}$$

# Irrfahrten, Fortsetzung, 6

Der Zustand “0” (und die anderen Zustände auch) ist also falls  $p = q = \frac{1}{2}$ :

# Irrfahrten, Fortsetzung, 6

Der Zustand “0” (und die anderen Zustände auch) ist also

falls  $p = q = \frac{1}{2}$ : *rekurrent*

falls  $p \neq q$ :



# Irrfahrten, Fortsetzung, 6

Der Zustand “0” (und die anderen Zustände auch) ist also

falls  $p = q = \frac{1}{2}$ : *rekurrent*

falls  $p \neq q$ : *transient*

falls  $p = q = \frac{1}{2}$ :

# Irrfahrten, Fortsetzung, 6

Der Zustand "0" (und die anderen Zustände auch) ist also

falls  $p = q = \frac{1}{2}$ : *rekurrent*

falls  $p \neq q$ : *transient*

falls  $p = q = \frac{1}{2}$ : *nullrekurrent* da  $D_{2n,0} \rightarrow_{n \rightarrow \infty} 0$ .

$$D_{2n,0} = p_{00}(n) \rightarrow 0 \Rightarrow \mu_i = \infty$$

# Irrfahrten

## symmetrische Irrfahrt in der Ebene

Zustände:  $(k, l) \in \mathbb{Z}^2$ , Anfangszustand:  $(0, 0)$

Bewegung: Punkt  $(X, Y)$

$X$ : ein Schritt nach rechts mit Wkt.  $p = \frac{1}{2}$  oder nach links mit Wkt.  $q = \frac{1}{2}$

$Y$ : ein Schritt nach oben mit Wkt.  $p$  oder nach unten mit Wkt.  $q = \frac{1}{2}$

Die Zufallsvariablen  $X$  und  $Y$  sind unabhängig.

$B_{n,k}$ : Ereignis, nach  $n$  Schritten im Zustand  $k$  zu sein

$$E_{n,k} := P(B_{n,k})$$

# symmetrische Irrfahrt in der Ebene

$$E_{2n,0} = P(X_{2n,0} = 0 \wedge Y_{2n,0} = 0) = D_{2n,0}^2 \sim \left(\frac{1}{\sqrt{\pi n}}\right)^2$$

# symmetrische Irrfahrt in der Ebene

$$E_{2n,0} = P(X_{2n,0} = 0 \wedge Y_{2n,0} = 0) = D_{2n,0}^2 \sim \left(\frac{1}{\sqrt{\pi n}}\right)^2$$

$$\begin{aligned}\sum_{n=1}^{\infty} E_{2n,0} &\sim \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} = \infty \\ \frac{1}{\pi} \sum_{n=1}^N \frac{1}{n} &\sim \frac{\ln N}{\pi} \xrightarrow{N \rightarrow \infty} \infty.\end{aligned}$$

Der Zustand “0” (und die anderen Zustände auch) ist also *rekurrent*, falls  $p = q = \frac{1}{2}$

# Irrfahrten

## symmetrische Irrfahrt im Raum

Zustände:  $(j, k, l) \in \mathbb{Z}^3$ , Anfangszustand:  $(0, 0, 0)$

Bewegung: Punkt  $(X, Y, Z)$

$X$ : ein Schritt nach rechts mit Wkt.  $p = \frac{1}{2}$  oder nach links mit Wkt.

$$q = 1 - p$$

$Y$ : ein Schritt nach oben mit Wkt.  $p$  oder nach unten mit Wkt.

$$q = 1 - p$$

$Z$ : ein Schritt nach hinten mit Wkt.  $p$  oder nach vorn mit Wkt.

$$q = 1 - p$$

Die Zufallsvariablen  $X$ ,  $Y$  und  $Z$  sind unabhängig.

# Irrfahrten im Raum

$C_{n,k}$ : Ereignis, nach  $n$  Schritten im Zustand  $k$ .

$$F_{n,k} := P(C_{n,k})$$

$$\begin{aligned} F_{2n,0} &= P(X_{2n,0} = 0, Y_{2n,0} = 0, Z_{2n,0} = 0) = D_{2n,0}^3 \\ &\sim \left(\frac{1}{\sqrt{\pi n}}\right)^3 \end{aligned}$$

# Irrfahrten im Raum

$C_{n,k}$ : Ereignis, nach  $n$  Schritten im Zustand  $k$ .

$$F_{n,k} := P(C_{n,k})$$

$$F_{2n,0} = P(X_{2n,0} = 0, Y_{2n,0} = 0, Z_{2n,0} = 0) = D_{2n,0}^3$$

$$\sim \left(\frac{1}{\sqrt{\pi n}}\right)^3$$

$$\sum_{n=0}^{\infty} F_{2n,0} \sim \frac{1}{(\pi)^{3/2}} \sum_{n=0}^{\infty} \frac{1}{n^{3/2}} < \infty$$

Der Zustand “0” (und die anderen Zustände auch) ist also *transient*.



# Irrfahrten mit Barriere

## Irrfahrt auf der Geraden mit Barriere

Zustände:  $k \in \mathbb{N}$ , Anfangszustand: 0

Bewegung: ein Schritt nach rechts mit Wkt.  $p$  oder  
nach links mit Wkt.  $q = 1 - p$

von  $k = 0$  aus geht es nur nach rechts

$0 < p, q < 1$ .

Übergangswktn.:

$$p_{k,k+1} = p = 1 - p_{k,k-1}$$

$$p_{ij} = 0, \quad \text{falls } |i - j| \neq 1 \quad \text{und} \quad i \neq 0$$

$$p_{01} = 1$$

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{pmatrix}$$

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \end{pmatrix}$$

wenn  $p = q = \frac{1}{2}$  so

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

wenn  $p = q = \frac{1}{2}$  so alle Zustände nullrekurrent.

wenn  $p > q$  so

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

wenn  $p = q = \frac{1}{2}$  so alle Zustände nullrekurrent.

wenn  $p > q$  so alle Zustände transient.

falls  $q > p$  so

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

wenn  $p = q = \frac{1}{2}$  so alle Zustände nullrekurrent.

wenn  $p > q$  so alle Zustände transient.

falls  $q > p$  so alle Zustände positiv rekurrent.

# Irrfahrt mit Barriere

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \\ q & 0 & p & 0 & \\ 0 & q & 0 & p & 0 \\ & & & \ddots & \ddots & \ddots \\ & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

wenn  $p = q = \frac{1}{2}$  so alle Zustände nullrekurrent.

wenn  $p > q$  so alle Zustände transient.

falls  $q > p$  so alle Zustände positiv rekurrent.

Alle Zustände haben die Periode 2.

# Irrfahrt mit Barriere

Die ersten beiden Fälle sind analog zur Irrfahrt ohne Barriere (wir spiegeln ggf. in Zustand 0)

Der dritte Fall erfordert etwas Rechenaufwand.

Stationäre Verteilung  $\pi$  im Fall  $p < q$ :

Sie ist (falls sie existiert) Lösung von

$$\mathbf{M}^T \cdot \pi = \pi$$

$$\pi_0 = q\pi_1$$

$$\pi_1 = \pi_0 + q\pi_2$$

$$\pi_i = p\pi_{i-1} + q\pi_{i+1}, \quad i \geq 2$$



# Irrfahrt mit Barriere

$$1 = \sum_{j=1}^{\infty} \pi_j$$

**Behauptung:**

$$\pi_i = \frac{p^{i-1}}{q^i} \pi_0, \quad i \geq 1$$

**Beweis:** vollständige Induktion.

# Irrfahrt mit Barriere

Stationäre Verteilung

$$1 = \sum_{i=0}^{\infty} \pi_i = \pi_0 + \sum_{i=1}^{\infty} \frac{p^{i-1}}{q^i} \pi_0$$

# Irrfahrt mit Barriere

## Stationäre Verteilung

$$\begin{aligned} 1 &= \sum_{i=0}^{\infty} \pi_i = \pi_0 + \sum_{i=1}^{\infty} \frac{p^{i-1}}{q^i} \pi_0 \\ &= \pi_0 + \frac{1}{q} \sum_{i=0}^{\infty} \frac{p^i}{q^i} \pi_0 = \pi_0 + \frac{1}{q} \frac{1}{1 - \frac{p}{q}} \pi_0 \\ &= \pi_0 + \frac{1}{q - p} \pi_0 \end{aligned}$$

# Irrfahrt mit Barriere

## Stationäre Verteilung

$$\begin{aligned}1 &= \sum_{i=0}^{\infty} \pi_i = \pi_0 + \sum_{i=1}^{\infty} \frac{p^{i-1}}{q^i} \pi_0 \\&= \pi_0 + \frac{1}{q} \sum_{i=0}^{\infty} \frac{p^i}{q^i} \pi_0 = \pi_0 + \frac{1}{q} \frac{1}{1 - \frac{p}{q}} \pi_0 \\&= \pi_0 + \frac{1}{q - p} \pi_0 \\ \pi_0 &= \frac{1}{1 + \frac{1}{q-p}} = \frac{q - p}{q - p + 1} \\ \pi_i &= \frac{p^{i-1}}{q^i} \cdot \frac{q - p}{q - p + 1}\end{aligned}$$

# Das Hardy-Weinberg Gesetz

Gegeben sei eine Population mit den Genotypen  $AA$ ,  $Aa$ ,  $aa$  und den Wahrscheinlichkeiten  $p, 2q, r$ , d.h. die Zufallsvariable  $X$ , die den zufälligen Genotyp einer Person beschreibt ist

$$X : \begin{pmatrix} AA & Aa & aa \\ p & 2q & r \end{pmatrix}$$

Dann ist  $P := p + q$  der Anteil der A-Gene, und  $Q := q + r$  der Anteil der a-Gene in der Population.

$X_n$ : zufällige Genotyp in der  $n$ -ten Generation

ist eine homogene Markov-Kette mit den Zuständen  $AA$ ,  $Aa$  und  $aa$ .

# Das Hardy-Weinberg Gesetz (2)

## Übergangswahrscheinlichkeiten

$$\begin{aligned} P(AA|AA) &= P(AA|AA, AA)P(AA) + P(AA|AA, Aa)P(Aa) + P(AA|AA, aa)P(aa) \\ &= p + \frac{1}{2}2q + 0 = p + q = P \end{aligned}$$

$$\begin{aligned} P(AA|Aa) &= P(AA|Aa, AA)P(AA) + P(AA|Aa, Aa)P(Aa) + P(AA|Aa, aa)P(aa) \\ &= \frac{1}{2}p + \frac{1}{4}2q + 0 = \frac{p+q}{2} = \frac{P}{2} \end{aligned}$$

$$P(AA|aa) = 0$$

$$\begin{aligned} P(Aa|AA) &= P(Aa|AA, AA)P(AA) + P(Aa|AA, Aa)P(Aa) + P(Aa|AA, aa)P(aa) \\ &= 0 + \frac{1}{2}2q + 1 \cdot r = q + r = Q \end{aligned}$$

$$\begin{aligned} P(Aa|Aa) &= P(Aa|Aa, AA)P(AA) + P(Aa|Aa, Aa)P(Aa) + P(Aa|Aa, aa)P(aa) \\ &= \frac{1}{2}p + \frac{1}{4}2q + \frac{1}{2}r = \frac{p+2q+r}{2} = \frac{P+Q}{2} \end{aligned}$$

$$\begin{aligned} P(Aa|aa) &= P(Aa|aa, AA)P(AA) + P(Aa|aa, Aa)P(Aa) + P(Aa|aa, aa)P(aa) \\ &= 1 \cdot p + \frac{1}{2}2q + 0 = p + q = P \end{aligned}$$

$$P(aa|AA) = 0$$

# Das Hardy-Weinberg Gesetz (3)

$$\begin{aligned}
 P(aa|Aa) &= P(aa|Aa, AA)P(AA) + P(aa|Aa, Aa)P(Aa) + P(aa|Aa, aa)P(aa) \\
 &= 0 + \frac{1}{4}2q + \frac{1}{2}r = \frac{q+r}{2} = \frac{Q}{2} \\
 P(aa|aa) &= P(aa|aa, AA)P(AA) + P(aa|aa, Aa)P(Aa) + P(aa|aa, aa)P(aa) \\
 &= 0 + \frac{1}{2}2q + 1 \cdot r = q + r = Q
 \end{aligned}$$

Wir haben also die transponierte Übergangsmatrix

$$M^T = \begin{pmatrix} P & \frac{P}{2} & 0 \\ Q & \frac{P+Q}{2} & P \\ 0 & \frac{Q}{2} & Q \end{pmatrix}$$

# Das Hardy-Weinberg Gesetz (4)

Eine stationäre Verteilung erhalten wir durch Lösen des linearen Gleichungssystems  $\pi = M^T \pi$ , also:

$$\pi_1 = P\pi_1 + \frac{P}{2}\pi_2$$

$$\pi_2 = Q\pi_1 + \frac{P+Q}{2}\pi_2 + P\pi_3$$

$$\pi_3 = \frac{Q}{2}\pi_2 + Q\pi_3$$

$$1 = \pi_1 + \pi_2 + \pi_3$$

$$0 = (\mathbf{M}^T - \mathbf{I})\pi$$

$$\begin{pmatrix} -Q & \frac{P}{2} & 0 \\ Q & -\frac{1}{2} & P \\ 0 & \frac{Q}{2} & -P \end{pmatrix} \rightarrow \begin{pmatrix} -Q & \frac{P}{2} & 0 \\ 0 & -\frac{Q}{2} & P \\ 0 & \frac{Q}{2} & -P \end{pmatrix} \rightarrow \begin{pmatrix} -Q & \frac{P}{2} & 0 \\ 0 & -\frac{Q}{2} & P \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} -Q^2 & 0 & P^2 \\ 0 & -\frac{1}{2} & P \\ 0 & 0 & 0 \end{pmatrix}$$



# Das Hardy-Weinberg Gesetz (5)

$$-Q^2\pi_1 + P^2\pi_3 = 0 \implies \pi_3 = \frac{Q^2}{P^2}\pi_1$$

$$-\frac{Q}{2}\pi_2 + P\pi_3 = 0 \implies \pi_2 = \frac{2P}{Q}\pi_3 = \frac{2Q}{P}\pi_1$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_1 \underbrace{\left(1 + \frac{2Q}{P} + \frac{Q^2}{P^2}\right)}_{(1+Q/P)^2=1/P^2} = 1$$

$$\pi_1 = P^2, \quad \pi_2 = 2PQ, \quad \pi_3 = Q^2$$

dasselbe, das wir schon früher erhalten haben.

# Regen und Regenschirme (1)

Herr K besitzt  $R$  Regenschirme, deren Verteilung sich zwischen Büro und Wohnung sich wie folgt ändert:

Regnet es beim Verlassen der Wohnung nimmt er einen Schirm aus der Wohnung, soweit vorhanden.

Umgekehrt, regnet es beim Verlassen des Büros, nimmt er einen Schirm soweit vorhanden.

Regnet es nicht, nimmt er keinen Schirm.

Regenwahrscheinlichkeit (Klima) sei  $p$ .

$X_n$ : Anzahl der Schirme an dem Ort, wo der Weg beginnt.

Zustandsmenge:  $S = \{0, 1, \dots, R\}$

$X_n$  ist Markov-Kette, da die nächsten Zustände durch den aktuellen Zustand bestimmt sind.

## Regen und Regenschirme (2)

Sei  $X_n = r$ . Dann

$$X_{n+1} = \begin{cases} R - r + 1 & \text{falls es regnet, } r > 0 \\ R - r & \text{falls es nicht regnet, } r > 0 \\ R & \text{falls } r = 0 \end{cases}$$

Alle Zustände positiv rekurrent, aperiodisch

Stationäre Verteilung  $\pi$ :  $\pi^T \mathbf{M} = \pi^T$ .

# Regen und Regenschirme (3)

Übergangsmatrix:

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1-p & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 1-p & p & 0 \\ 0 & 0 & 0 & 0 & \dots & 1-p & p & 0 & 0 \\ & & & & \dots & & & & \\ 0 & 0 & 1-p & p & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-p & p & 0 & \dots & 0 & 0 & 0 & 0 \\ 1-p & p & 0 & 0 & \dots & 0 & 0 & 0 & 0 \end{pmatrix}$$

# Regen und Regenschirme (4), $R = 1, 2$

$$\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Eigenvektor zu Eigenwert  $\lambda = 1$ :  $(\frac{1-p}{2-p}, \frac{1}{2-p})^T$ .

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{pmatrix}$$

Eigenvektor zu Eigenwert  $\lambda = 1$ :  $(\frac{1-p}{3-p}, \frac{1}{3-p}, \frac{1}{3-p})^T$ .

# Regen und Regenschirme (4), Lineares Gleichungssystem

$$(1 - p)\pi_R = \pi_0$$

$$(1 - p)\pi_{R-1} + p\pi_R = \pi_1$$

$$(1 - p)\pi_{R-2} + p\pi_{R-1} = \pi_2$$

$$\dots =$$

$$(1 - p)\pi_1 + p\pi_2 = \pi_{R-1}$$

$$1\pi_0 + p\pi_1 = \pi_R$$

R+1. Gleichung:  $\pi_0 = \pi_R - p\pi_1$

0. Gleichung:  $\pi_0 = (1-p)\pi_R$     woraus folgt:  $\pi_1 = \pi_R$

R+1. Gleichung:  $\pi_0 = (1-p)\pi_1$

1. Gleichung:  $(1-p)\pi_{R-1} + p\pi_1 = \pi_1 \Rightarrow \pi_1 = \pi_{R-1} = \pi_R$

R. Gleichung:  $(1-p)\pi_1 + p\pi_2 = \pi_{R-1} \Rightarrow$

$\pi_1 = \pi_2$  und weiter  $\pi_1 = \pi_2 = \dots = \pi_R$ .

Mit  $1 = R \cdot \pi_1 + \pi_0 = R \cdot \pi_1 + (1-p)\pi_1$  folgt die stationäre

Verteilung  $\pi_1 = \frac{1}{R+1-p} = \dots = \pi_R$  und  $\pi_0 = \frac{1-p}{R+1-p}$ .

# Regen und Regenschirme (5)

Nass wird K wenn es in Zustand 0 regnet, also

$$P(\text{nass}) = \frac{p(1-p)}{R+1-p} \leq \alpha \quad \text{gdw.}$$

$$\underbrace{p + \frac{p(1-p)}{\alpha}}_{\phi(p)} \leq R+1$$

Zur Berechnung des Maximums der linken Seite bestimmen wir

$$\phi'(p) = 1 + \frac{1-2p}{\alpha} = 0 \Rightarrow p = \frac{1+\alpha}{2}$$

$$\phi(p) \leq \frac{1+\alpha}{2} + \frac{\frac{1+\alpha}{2}(1-\frac{1+\alpha}{2})}{\alpha} \leq R+1$$

Bei  $\alpha = 0.05$ :  $\frac{1+\alpha}{2} = 0.525$  und es wird die linke Seite zu  $4.5105 \leq R$ , d.h. 5 Regenschirme reichen bei jedem Klima aus.



# Inhalt

- Begriffe und einfache Zusammenhänge
- Klassifikation der Zustände
- Rekurrente und transiente Zustände
- Grenzverteilungen
- Klassische Beispiele
- Markov Chain Monte Carlo

# Markov Chain Monte Carlo, Idee

- Erinnerung: Eine irreduzible ergodische Markov-Kette  $X_n$  hat eine stationäre Verteilung,  $X_n \rightarrow X$ ,  $X \sim \pi$
- Wenn  $g$  beschränkt so (Gesetz der großen Zahlen für MK):

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \rightarrow_p \mathbf{E}_{\pi} g(X) = \sum_j g(j) \pi_j$$

wobei über alle Zustände von  $X$  summiert wird.

- Betrachten jetzt einen stetigen Zustandsraum und konstruieren eine Markov-Kette mit stationärer Verteilung  $f$ . Dann können wir z.B. das Integral  $\int h(x) f(x) dx$  approximieren:

$$\frac{1}{n} \sum_{i=1}^n h(X_i) \rightarrow_p \mathbf{E}_f h(X) = \int h(x) f(x) dx$$

# Markov Chain Monte Carlo

## Metropolis-Hastings Algorithmus

Sei  $q(y|x)$  eine beliebige leicht zu simulierende Dichte.

0. Wähle  $X_0$  beliebig. Seien  $X_0, X_1, \dots, X_i$  gegeben.  $X_{i+1}$  wird wie folgt generiert:
1. Generiere  $Y \sim q(y|X_i)$
2. Berechne  $r(X_i, Y)$ , wobei

$$r(x, y) = \min \left( \frac{f(y) q(x|y)}{f(x) q(y|x)}, 1 \right)$$

3. Setze

$$X_{i+1} = \begin{cases} Y & \text{mit Wahrscheinlichkeit } r \\ X_i & \text{mit Wahrscheinlichkeit } 1 - r \end{cases}$$

# Markov Chain Monte Carlo

Metropolis-Hastings Algorithmus, Anmerkung

Eine übliche Wahl der “freundlichen” Dichte ist  $q(y|x) : \mathcal{N}(x, b^2)$   
(Normalverteilung, zentriert auf den aktuellen Wert  $x$ )

$$q(y|x) = \frac{1}{\sqrt{2\pi}b} e^{-\frac{1}{2b^2}(y-x)^2} = q(x|y).$$

Damit vereinfacht sich  $r$  zu

$$r(x, y) = \min \left( \frac{f(y)}{f(x)}, 1 \right)$$

Wahl von  $b$ : noch offen.

# Markov Chain Monte Carlo

Metropolis-Hastings Algorithmus, Beispiel

Angenommen, wir wollen eine Cauchy-Verteilung simulieren,

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

# Markov Chain Monte Carlo

## Metropolis-Hastings Algorithmus, Beispiel

Angenommen, wir wollen eine Cauchy-Verteilung simulieren,

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

Wenn wir, wie im Vorschlag oben  $q(y|x) \sim \mathcal{N}(x, b^2)$  setzen,

$$r(x, y) = \min \left( \frac{f(y)}{f(x)}, 1 \right) = \min \left( \frac{1+x^2}{1+y^2}, 1 \right)$$

Algorithmus:

- 1. Ziehe  $Y \sim \mathcal{N}(x, b^2)$ .
- 2.

$$X_{i+1} = \begin{cases} Y & \text{mit Wahrscheinlichkeit } r(X_i, Y) \\ X_i & \text{mit Wahrscheinlichkeit } 1 - r(X_i, Y) \end{cases}$$

# Markov Chain Monte Carlo

Metropolis-Hastings Algorithmus, Wahl des Tuning-Parameters  $b$

- $b$  zu klein: nur kleine Schritte, es wird nicht der ganze Stichprobenraum untersucht
- $b$  zu groß: viele Vorschläge  $Y$ , die weit in den Tails sind, d.h.  $r$  wird klein, die Markov-Kette bleibt lange in derselben Position.
- $b$  mittel: gut.

# Markov Chain Monte Carlo

Korrektheit des Metropolis-Hastings Algorithmus (1)

Stationäre Verteilung, bei diskreten Zufallsvariablen

$$\pi = \pi \mathbf{M}$$

wobei  $\mathbf{M}$  Übergangsmatrix der Markov-Kette ist.

Def. 84 (Detailed balance)

Wir sagen, eine Markov-Kette hat Detailed balance, wenn

$$p_{ij}\pi_i = p_{ji}\pi_j \quad \forall i, j$$

Erinnerung: Wenn eine Markov-Kette  $X_n$  detailed balance mit  $\pi$  hat so ist  $\pi$  stationäre Verteilung von  $X_n$ :

$$(\pi \mathbf{M})_j = \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j.$$



# Markov Chain Monte Carlo

Korrektheit des Metropolis-Hastings Algorithmus (2)

Stationäre Verteilung, bei stetigen Zufallsvariablen

$$f(x) = \int f(y)p(y, x) dy$$

wobei  $p(x, y)$  Übergangsdichte von Zustand  $x$  in Zustand  $y$  ist.

Detailed balance, falls

$$f(x)p(x, y) = f(y)p(y, x) \quad \forall x, y$$

Satz: falls  $f$  detailed balance besitzt, so ist  $f$  stationär.

**Beweis:** Aus detailed balance folgt:

$$\int f(y)p(y, x) dy = \int f(x)p(x, y) dy = f(x) \int p(x, y) dy = f(x).$$

# Markov Chain Monte Carlo

## Korrektheit des Metropolis-Hastings Algorithmus (3)

Bleibt zu zeigen,  $f$  erfüllt detailed balance.

Seien  $x, y$  beliebige Punkte. Es gilt

$$f(x)q(y|x) > f(y)q(x|y) \text{ oder}$$

$$f(x)q(y|x) < f(y)q(x|y)$$

(oder  $f(x)q(y|x) = f(y)q(x|y)$ , aber letzteres nur mit

Wahrscheinlichkeit Null). Sei o.B.d.A.  $f(x)q(y|x) > f(y)q(x|y)$ .

Dann

$$r(x, y) = \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)} \quad \text{und} \quad r(y, x) = 1.$$

# Markov Chain Monte Carlo

Korrektheit des Metropolis-Hastings Algorithmus (4)

$p(x, y)$  ist Übergangsdichte von  $x$  nach  $y$

Forderung:

1. Vorschlagsdichte  $q(y|x)$  muss  $y$  generieren und
2.  $y$  muss akzeptiert werden.

$$p(x, y) = \underbrace{q(y|x)}_{1.} \underbrace{r(x, y)}_{2.} = q(y|x) \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)} = \frac{f(y)}{f(x)} q(x|y)$$

Daraus folgt:

$$f(x)p(x, y) = f(y)q(x|y)$$

# Markov Chain Monte Carlo

Korrektheit des Metropolis-Hastings Algorithmus (5)

$p(y, x)$  ist Übergangsdichte von  $y$  nach  $x$

Forderung:

1. Vorschlagsdichte  $q(x|y)$  muss  $x$  generieren und
2.  $x$  muss akzeptiert werden.

$$p(y, x) = \underbrace{q(x|y)}_{1.} \underbrace{r(y, x)}_{2.} = q(x|y) \Rightarrow \\ f(y)p(y, x) = f(y)q(x|y)$$

Zusammen mit der letzten Gleichung auf der vorigen Folie folgt:

$$f(x)p(x, y) = f(y)p(y, x).$$

# Gibbs Sampling

## Problemstellung

Simulation aus einer “schwierigen” zweidimensionalen Dichte

aber Simulation aus bedingten Dichten  $f_{X|Y}(x|y)$  und  $f_{Y|X}(y|x)$  sei einfach

## Gibbs Sampling Algorithmus

Sei  $(X_0, Y_0)$  beliebiger Startwert und  $(X_0, Y_0), \dots, (X_n, Y_n)$  bereits simuliert.

- $X_{n+1} \sim f_{X|Y}(x|Y_n)$
- $Y_{n+1} \sim f_{Y|X}(y|X_{n+1})$

Simulation aus den bedingten Verteilungen nicht so einfach  $\implies$   
Metropolis-Hastings Algorithmus.

- Wahrscheinlichkeitsbegriff
- Rechnen mit Wahrscheinlichkeiten

- Wahrscheinlichkeitsbegriff
- Rechnen mit Wahrscheinlichkeiten
- Einfache kombinatorische Formeln
- Stirling-Formel

- Wahrscheinlichkeitsbegriff
- Rechnen mit Wahrscheinlichkeiten
- Einfache kombinatorische Formeln
- Stirling-Formel
- Bedingte Wahrscheinlichkeiten, Unabhängigkeit
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes



- Wahrscheinlichkeitsbegriff
- Rechnen mit Wahrscheinlichkeiten
- Einfache kombinatorische Formeln
- Stirling-Formel
- Bedingte Wahrscheinlichkeiten, Unabhängigkeit
- Satz der Totalen Wahrscheinlichkeit
- Satz von Bayes
- Zufallsvariable, Verteilungsfunktion, Eigenschaften
- Erwartungswert, Varianz, Rechnen mit Erwartungswert, Varianz

# Zusammenfassung (2)

## Wahrscheinlichkeitsmodelle und Transformationen

- Diskrete Gleichverteilung
- Binomialverteilung
- Poisson-Verteilung
- Geometrische Verteilung

# Zusammenfassung (2)

## Wahrscheinlichkeitsmodelle und Transformationen

- Diskrete Gleichverteilung
- Binomialverteilung
- Poisson-Verteilung
- Geometrische Verteilung
- Gleichverteilung
- Exponentialverteilung, Anwendungen
- Normalverteilung, Eigenschaften

# Zusammenfassung (2)

## Wahrscheinlichkeitsmodelle und Transformationen

- Diskrete Gleichverteilung
- Binomialverteilung
- Poisson-Verteilung
- Geometrische Verteilung
- Gleichverteilung
- Exponentialverteilung, Anwendungen
- Normalverteilung, Eigenschaften
- Transformationssatz für eindimensionale Zufallsvariablen
- Faltungsformel

# Zusammenfassung (3)

Mehrdimensionale Verteilungen, Ungleichungen, Konvergenzarten

- Zweidimensionale Zufallsvariablen
- Unabhängigkeit und Korrelation, Berechnung von Korrelationskoeffizienten für diskrete und für stetige Zufallsvariablen

# Zusammenfassung (3)

Mehrdimensionale Verteilungen, Ungleichungen, Konvergenzarten

- Zweidimensionale Zufallsvariablen
- Unabhängigkeit und Korrelation, Berechnung von Korrelationskoeffizienten für diskrete und für stetige Zufallsvariablen
- Markov-Ungleichung, Tschebyschev-Ungleichung, Jensen-Ungleichung

# Zusammenfassung (3)

Mehrdimensionale Verteilungen, Ungleichungen, Konvergenzarten

- Zweidimensionale Zufallsvariablen
- Unabhängigkeit und Korrelation, Berechnung von Korrelationskoeffizienten für diskrete und für stetige Zufallsvariablen
- Markov-Ungleichung, Tschebyschev-Ungleichung, Jensen-Ungleichung
- Konvergenzarten: in Wahrscheinlichkeit, in Verteilung, im quadratischen Mittel, fast sicher

# Zusammenfassung (4)

Grenzwertsätze, Schätzmethoden und Zufallszahlen

- Gesetz der Großen Zahlen
- Empirische Verteilungsfunktion
- Satz von Glivenko-Cantelli
- Zentraler Grenzwertsatz
- Schätzmethoden (Momentenschätzung, Maximum-Likelihood-Methode)



# Zusammenfassung (4)

Grenzwertsätze, Schätzmethoden und Zufallszahlen

- Gesetz der Großen Zahlen
- Empirische Verteilungsfunktion
- Satz von Glivenko-Cantelli
- Zentraler Grenzwertsatz
- Schätzmethoden (Momentenschätzung, Maximum-Likelihood-Methode)
- Erzeugung und Eigenschaften von Zufallszahlen
- Statistische Tests von Zufallszahlen
- Methoden zur Erzeugung spezieller Verteilungen, Berechnung der inversen Verteilungsfunktion

# Zusammenfassung (5)

## Markov-Ketten

- Begriff der Markov'schen Kette, Eigenschaften

# Zusammenfassung (5)

## Markov-Ketten

- Begriff der Markov'schen Kette, Eigenschaften
- Klassifikation der Zustände (Kommunikation, wesentliche, unwesentliche Zustände, Periodizität)
- Positiv rekurrente, nullrekurrente und transiente Zustände, mittlere Rückkehrzeit

# Zusammenfassung (5)

## Markov-Ketten

- Begriff der Markov'schen Kette, Eigenschaften
- Klassifikation der Zustände (Kommunikation, wesentliche, unwesentliche Zustände, Periodizität)
- Positiv rekurrente, nullrekurrente und transiente Zustände, mittlere Rückkehrzeit
- Ergodensatz, stationäre Verteilung, Berechnung stationärer Verteilungen

# Zusammenfassung (5)

## Markov-Ketten

- Begriff der Markov'schen Kette, Eigenschaften
- Klassifikation der Zustände (Kommunikation, wesentliche, unwesentliche Zustände, Periodizität)
- Positiv rekurrente, nullrekurrente und transiente Zustände, mittlere Rückkehrzeit
- Ergodensatz, stationäre Verteilung, Berechnung stationärer Verteilungen
- Irrfahrten

# Zusammenfassung (6)

## Übungsaufgaben

- 8, 9 (Binomialverteilung)
- 10, 11 (Satz der Totalen Wahrscheinlichkeit, Satz von Bayes)
- 12 (Poisson-, Binomialverteilung, Satz der Totalen Wahrscheinlichkeit)
- 15 (Berechnen der Dichtefunktion, Berechnen von Wahrscheinlichkeiten)
- 16 (Geometrische Verteilung)
- 17, 18 (Rechnen mit Erwartungswert und Varianz)
- 20 (Normalverteilung)
- 21 (Rechnen mit Wahrscheinlichkeiten, Exponentialverteilung)

# Zusammenfassung (7)

## Übungsaufgaben (2)

- 22, 24a,b,c, 25 (Transformationsformel)
- 23 (Geometrische Verteilung, Rechnen mit Wahrscheinlichkeiten)
- 26 (Faltung)
- 28 (Berechnen von Erwartungswerten)
- 28, 31, 32 (Zweidimensionale Zufallsvariablen, Berechnung von Kovarianzen und Korrelationskoeffizienten)
- 30a (Randverteilungen)

# Zusammenfassung (8)

## Übungsaufgaben (3)

- 33,34 (Zentraler Grenzwertsatz, Tschebyschev-Ungleichung)
- 35,36 (Momentenschätzung, ML-Schätzung)
- 38,39 (Dichte, Zufallszahlen, Akzeptanzmethode)
- 40,41,42,43,44 (Markov-Ketten)