Einführung in die Stochastik

Kapitel 1-21

Steffen Winter | Wintersemester 2024/2025

Kapitel 1: Klassische Beispiele und Paradoxa

- Würfelproblem
- Teilungsproblem
- Bertrandsches Paradoxon
- Efrons Würfel
- Heikle Fragen

Ein Würfelproblem

Das Problem wird dem Schriftsteller, Edelmann und Spieler Antoine Gombaud, genannt Chevalier de Méré (17. Jh.) zugeschrieben, der mit dem Mathematiker Pierre de Fermat befreundet war.

- Werfe einen fairen Würfel viermal, gewünscht ist eine “6”. 
- Werfe ein Würfelpaar 24 mal, gewünscht ist eine “66”. 

Welches Ziel wird eher erreicht? (Was genau ist hier die Frage?)

(1) Wahrscheinlichkeit für mindestens eine “6" bei vier Würfen bzw. für mindestens eine “66" bei 24 Würfen mit Würfelpaar?

(2) Mittlere Anzahl von “6" bzw. “66" beim wiederholten Viererwurf bzw. 24-er Wurf?

Die Antwort ist in der Tat abhängig von der Interpretation der Frage.

Informelle Analyse

Unter den “üblichen Annahmen” gilt:

$$\text{Wkeit (mind. eine “6”)} \;=\underbrace{1-\left(\frac{5}{6}\right)^{4}}_{\approx 0.52} > \text{Wkeit (mind. eine “66”)} \;=\underbrace{1-\left(\frac{35}{36}\right)^{24}}_{\approx 0.49}$$

Dagegen erhält man für die mittleren Anzahlen:

$$\text{Mittlere Anzahl “6”} \;= 4\cdot\frac{1}{6}=\frac{2}{3} = \text{Mittlere Anzahl “66”} \;= 24\cdot\frac{1}{36}=\frac{2}{3}.$$

Teilungsproblem

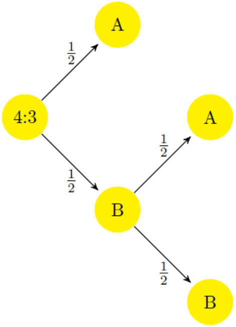
Das Problem wird dem Mathematiker und Franziskaner Luca Pacioli (1494) zugeschrieben, der als erster die doppelte Buchführung vollständig beschrieben hat.

- Zwei Spieler A, B mit z.B. gleichen Gewinnchancen treten gegeneinander in wiederholten Partien an.
- Gewinner ist, wer zuerst fünf Partien gewonnen hat.
- Einsatz wird zu Beginn des Wettstreits eingezahlt.
- Der Wettkampf muss beim Stand 4 : 3 abgebrochen werden.

Frage: Wie soll der Einsatz aufgeteilt werden?

Pierre de Fermat und Blaise Pascal (17. Jh.): Teile im Verhältnis der Gewinnwahrscheinlichkeiten, wenn beim Stand 4 : 3 weitergespielt würde.

Aufgabe: Bestimme die Wahrscheinlichkeit, dass A beim Stand von 4 : 3 gewinnt.

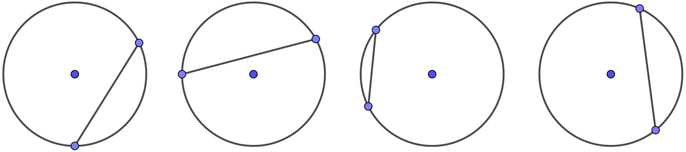


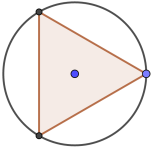
$$\text{Wkeit(A gewinnt)} = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$$

Bertrandsches Paradoxon

(nach Joseph Bertrand (1888), frz. Pädagoge und Mathematiker)

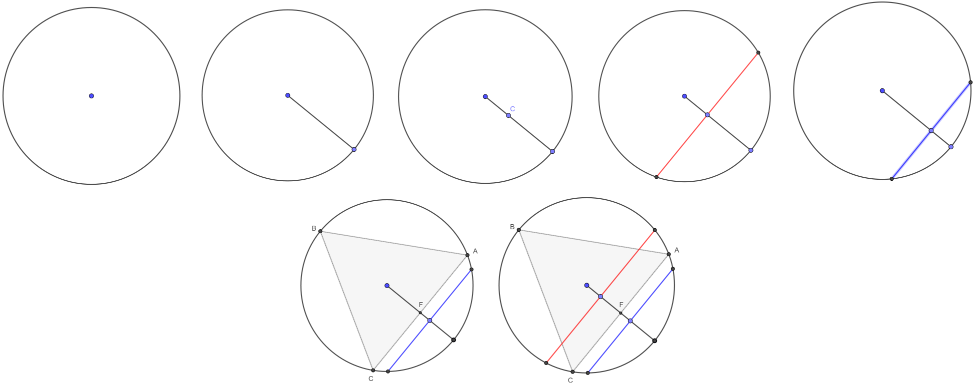
In einem Einheitskreis wird “zufällig” eine Sehne gezogen. Wie groß ist die Wahrscheinlichkeit, dass die Sehne länger ist als die Seite eines einbeschriebenen regulären Dreiecks?





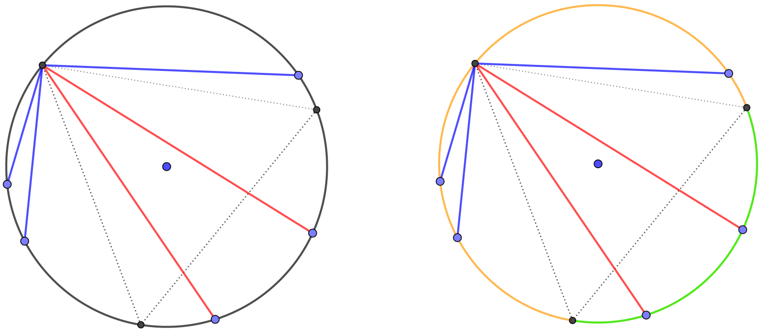
Lösung 1: Wahrscheinlichkeit $\frac{1}{2}$

Eine Sehne ist bestimmt durch ihren Mittelpunkt bzw. durch Abstand und Richtung desselben vom Ursprung. Wähle Richtung und Abstand zufällig.



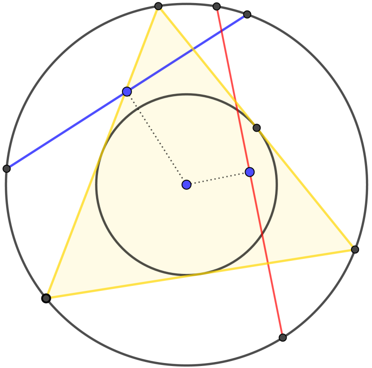
Lösung 2: Wahrscheinlichkeit $\frac{1}{3}$

Eine Sehne ist bestimmt durch ihre Endpunkte. Wähle diese zufällig auf dem Kreisrand.



Lösung 3: Wahrscheinlichkeit $\frac{1}{4}$

Eine Sehne ist bestimmt durch ihren Mittelpunkt. Wähle diesen zufällig in der Kreisscheibe.



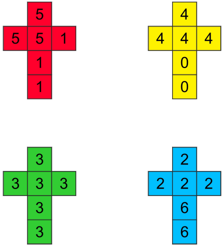
Ein Problem – drei Antworten?

Nein! Drei Modellierungen – drei Antworten!

Efrons Würfel – Wer höflich ist, gewinnt!

Bradley Efron (1938 –): US-amerikanischer Statistiker, bekannt für das Bootstrap-Verfahren in der nichtparametrischen Statistik.

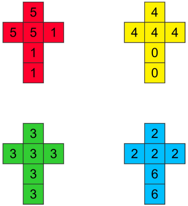
Vier Nicht-Standard-Würfel mit Farben rot, gelb, grün, blau sind durch ihre Netze gegeben.



Zwei Spieler A, B treten wiederholt gegeneinander an. Zunächst wählt A einen beliebigen Würfel, dann B einen der verbleibenden drei Würfel. Die größere der jeweils gewürfelten Augenzahlen gewinnt.

- Ist das Spiel fair?
- Gibt es einen besten Würfel?

Ein Beispiel



Beispiel: Spieler 1 wählt rot. Dann wählt Spieler 2 blau. Spieler 1 gewinnt nur dann, wenn er zunächst eine 5 würfelt, das geschieht mit Wkeit $\frac{1}{2}$, und Spieler 2 danach eine 2 würfelt, das geschieht mit Wkeit $\frac{2}{3}$. Spieler 1 gewinnt also (lediglich) mit Wkeit $\frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$.

Man kann zeigen:

$$\text{rot} \succ \text{gelb} \succ \text{grün} \succ \text{blau} \succ \text{rot}.$$

Ehrliche Antworten auf heikle Fragen?

- Haben Sie eine Corona-Schutzimpfung erhalten?
- **Randomisierte Antwort:** Anonyme Befragung, n Personen
  Jede Person wählt aus den drei Karten

Sind Sie gegen Corona geimpft?	Steht auf dieser Karte die Ziffer Eins? **1**	Steht auf dieser Karte die Ziffer Eins?

zufällig eine aus (Interviewer sieht diese Karte nicht).
Die Frage wird wahrheitsgemäß beantwortet, die Karte zurückgelegt. Anonymität ist gewahrt. Durch

$$\frac{3}{n} \cdot \text{“Anzahl Ja”} - 1$$

wird der Anteil der Befragten geschätzt, die eine Corona-Schutzimpfung erhalten haben.

Fazit

- Die Vorlesung Einführung in die Stochastik wird ein solides Fundament legen, auf dessen Basis die vorstehend angesprochenen Beispiele und Probleme rigoros behandelt werden können.
- Entscheidend für die Entwicklung der Wahrscheinlichkeitstheorie waren
  - einerseits viele einzelne Probleme und Anwendungen,
  - ganz wesentlich jedoch auch die axiomatische Begründung der Wahrscheinlichkeitstheorie – auf Grundlage der allgemeinen Maß- und Integrationstheorie ($\rightarrow$ Analysis 3) – durch *Andrei Nikolajewitsch Kolmogorow* (1903 – 1987).

Kapitel 2: Merkmalräume und Ereignisse

In der Stochastik werden **Zufallsexperimente** modelliert und untersucht.

Was sind Zufallsexperimente?

Einen stochastischen Vorgang nennt man ideales Zufallsexperiment, wenn folgende Gegebenheiten vorliegen:

- Das Experiment wird unter genau festgelegten Bedingungen (Versuchsbedingungen) durchgeführt.
- Die Menge der möglichen Ergebnisse (Grundgesamtheit) ist vor Durchführung des Experimentes bekannt.
- Das Experiment kann (zumindest prinzipiell) beliebig oft unter gleichen Bedingungen wiederholt werden.
- Das Ergebnis des Experiments ist (in der Regel) nicht eindeutig festgelegt.

Beispiele für Zufallsexperimente

Beispiele (für Zufallsexperimente)

- das Werfen eines oder mehrerer Würfel oder Münzen,
- die Beobachtung der Anzahl von „Jobs“, die an einem Prozessor innerhalb einer Minute ankommen,
- die Beobachtung von Lebensdauern von Geräten,
- die Beobachtung der Anzahl der Einsen in einer Folge von 32 zufälligen Binärzahlen,
- Zeitpunkte radioaktiven Zerfalls,
- Zeitpunkte des Eintretens von Versicherungsfällen (oder deren Anzahl in einem vorgegebenen Zeitintervall),
- Status eines großen physikalischen Systems (interagierender Partikel),
- Brownsche Bewegung,
- Zustandsbestimmung eines quantenphysikalischen Systems . . .

Merkmalräume und Ereignisse

Definition 2.1 (Notation)

- Die Menge Ω aller möglichen Ergebnisse eines Zufallsexperiments heißt Merkmalraum, Grundraum oder Ergebnisraum. Die Elemente von Ω werden auch Ergebnisse genannt und z.B. mit $\omega, \omega', \omega_1, \omega_2, \dots$ bezeichnet.
- Teilmengen $A \subset \Omega$ heißen Ereignisse. Ereignisse werden z.B. mit $A, B, A_1, A_2, \dots$ bezeichnet.
- Spezielle Ereignisse sind
    - die Elementarereignisse $\{\omega\}$ für $\omega \in \Omega$,
    - das sichere Ereignis Ω,
    - die leere Menge $\emptyset$ als unmögliches Ereignis.

Einfacher Würfelwurf

Beispiel

Um den Wurf eines Würfels als Zufallsexperiment zu beschreiben, verwenden wir als Grundmenge

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

- Ereignisse sind z.B. $A = \{1, 2, 5\}$, $\{2, 1, 3, 4\}$, nicht dagegen $\{1, 7\}$.
- Elementarereignisse sind z.B. $\{1\}$ oder $\{5\}$.
- Wie viele verschiedene Ereignisse gibt es zu dieser Grundmenge Ω?

  Insgesamt 2^6, einschließlich $\emptyset$ und Ω.

Eintreten von Ereignissen

Definition 2.2 (Sprechweise)

Ist ω Ergebnis eines Zufallsexperiments und A ein Ereignis, so sagt man

- A „tritt ein“, falls $\omega \in A$ gilt.
- A „tritt nicht ein“, falls $\omega \notin A$ gilt.

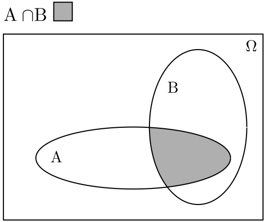
Durchschnitt von Ereignissen

Definition 2.3

Es seien $A, B \subset \Omega$ Ereignisse. Dann ist der Durchschnitt von A und B

$$A \cap B := \{\omega \in \Omega : \omega \in A \textbf{ und } \omega \in B\}$$

wieder ein Ereignis. $A \cap B$ tritt ein genau dann, wenn A **und** B eintreten.



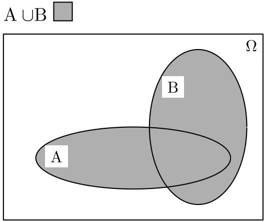
Vereinigung von Ereignissen

Definition 2.4

Seien $A, B \subset \Omega$ Ereignisse. Dann ist die Vereinigung von A und B

$$A \cup B := \{\omega \in \Omega : \omega \in A \textbf{ oder } \omega \in B\}$$

wieder ein Ereignis. $A \cup B$ tritt ein genau dann, wenn A **oder** B eintritt (eventuell beide).



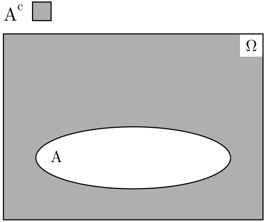
Komplement von Ereignissen

Definition 2.5

Für ein Ereignis $A \subset \Omega$ nennt man das Ereignis

$$A^{\mathsf{c}} := \{\omega \in \Omega : \omega \notin A\}$$

das Komplement von A in Ω oder auch das Gegenereignis von A. Es tritt ein genau dann, wenn A **nicht** eintritt.



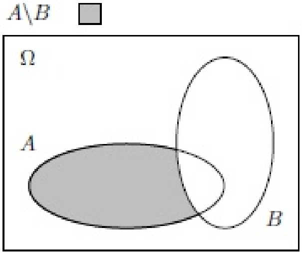
Mengendifferenz von Ereignissen

Definition 2.6

Seien $A, B \subset \Omega$ Ereignisse. Dann heißt das Ereignis

$$A \setminus B := \{\omega \in \Omega : \omega \in A \textbf{ und } \omega \notin B\} = A \cap B^c$$

Mengen-Differenz von A und B. $A \setminus B$ tritt ein, falls A eintritt aber B **nicht**.



Interpretation

Für Ereignisse $A_1, A_2, \ldots$ bedeutet

- $A_1 \cap \ldots \cap A_n$, dass jedes der Ereignisse $A_1, \ldots, A_n$ eintritt.
- $A_1 \cap A_2 \cap \ldots$, dass jedes der Ereignisse $A_1, A_2, \ldots$ eintritt.
- $A_1 \cup \ldots \cup A_n$, dass mindestens eines der Ereignisse $A_1, \ldots, A_n$ eintritt.
- $A_1 \cup A_2 \cup \ldots$, dass mindestens eines der Ereignisse $A_1, A_2, \ldots$ eintritt.

Disjunkte Ereignisse

Definition 2.7

- Zwei Ereignisse A und B heißen disjunkt, wenn A und B kein Element gemeinsam haben, d.h. wenn $A \cap B = \emptyset$ gilt.
- Ereignisse $A_1, \ldots, A_n$ bzw. $A_1, A_2, \ldots$ heißen paarweise disjunkt, wenn $A_i \cap A_j = \emptyset$ für jedes Paar i, j mit $i \neq j$ gilt.
- **Nur für disjunkte** Ereignisse A und B oder **paarweise disjunkte** Ereignisse $A_1, A_2, \ldots, A_n$ bzw. $A_1, A_2, \ldots$ verwenden wir die Bezeichnungen:

$$\begin{aligned} A + B &:= A \cup B \\ \sum_{j=1}^{n} A_j &:= A_1 \cup A_2 \cup \ldots \cup A_n \\ \sum_{j=1}^{\infty} A_j &:= A_1 \cup A_2 \cup \ldots \end{aligned}$$

Einfacher Würfelwurf

Beispiel

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

Beispiele für Ereignisse:

- $A =$ „Augenzahl gerade“ $= \{2, 4, 6\}$,
- $B =$ „Augenzahl mindestens 4“ $= \{4, 5, 6\}$,
- $C =$ „Augenzahl gleich 1“ $= \{1\}$

- $A \cup B =$ „Augenzahl gerade oder ≥ 4“ $= \{2, 4, 5, 6\}$
- $A \cap B =$ „Augenzahl gerade und ≥ 4“ $= \{4, 6\}$
- $B^c =$ „Augenzahl nicht ≥ 4“ $= \{1, 2, 3\}$
- $A \setminus B =$ „Augenzahl gerade, aber nicht ≥ 4“ $= \{2\}$
- $A + C =$ „Augenzahl gerade oder gleich 1“ $= \{1, 2, 4, 6\}$
- $(A \cap B) \cup C = \{4, 6\} \cup \{1\} = \{1, 4, 6\}$

Zweifacher Würfelwurf

Beispiel

Ein roter und ein grüner Würfel werden gleichzeitig geworfen. Ein geeigneter Grundraum ist dann

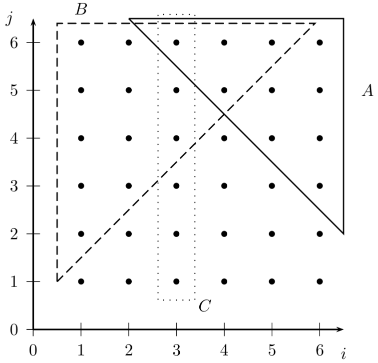
$$\Omega = \{(i,j) : i,j \in \{1,2,\ldots,6\}\},$$

wobei i das Ergebnis des roten und j das Ergebnis des grünen Würfels bezeichnet. Die Menge

$$\begin{aligned} A &= \{(i,j) \in \Omega : i+j \geq 9\} \\ &= \{(3,6),(4,5),(4,6),(5,4),(5,5),(5,6),(6,3),(6,4),(6,5),(6,6)\} \end{aligned}$$

beschreibt das Ereignis „Die Augensumme beträgt mindestens 9“. Sei

$$\begin{aligned} B &= \text{„Der grüne Würfel zeigt eine höhere Zahl als der rote“} \\ &= \{(i,j) \in \Omega : i < j\} \\ C &= \text{„Der rote Würfel hat die Augenzahl 3“} = \{(i,j) \in \Omega : i = 3\}. \end{aligned}$$



Zweifacher Würfelwurf (Forts.)

Beispiel

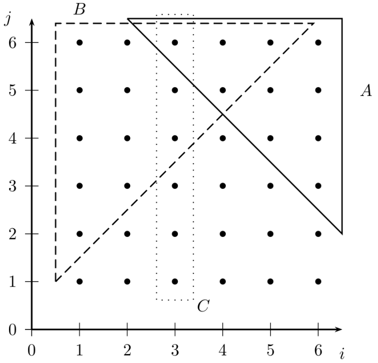
Dann gilt:

$$A \cap B = \text{„Die Augensumme ist} \geq 9 \text{ und der grüne Würfel zeigt mehr als der rote“} = \{(3,6),(4,5),(4,6),(5,6)\}$$

$$B \cup C = \text{„Der rote Würfel zeigt eine 3 oder der grüne Würfel zeigt mehr als der rote“}$$

$$C^c = \text{„Der rote Würfel zeigt keine 3“}$$

$A \cap C$ ist das Elementarereignis $\{(3,6)\}$.

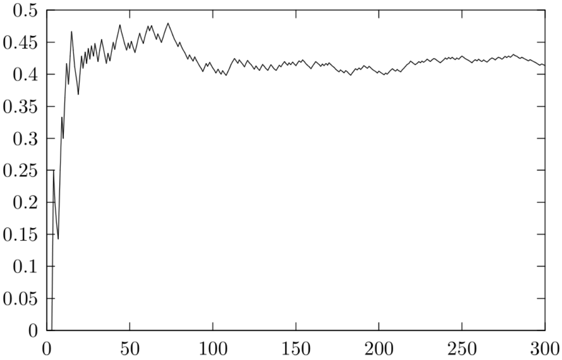


Kapitel 3: Wahrscheinlichkeitsräume

Motivation: Wahrscheinlichkeiten modellieren relative Häufigkeiten!

Empirisches Gesetz über die Stabilisierung relativer Häufigkeiten:

Wächst bei einem idealen Zufallsexperiment die Anzahl *n* der Wiederholungen, so stabilisieren sich die relativen Häufigkeiten des Eintretens eines festen Ereignisses erfahrungsgemäß um einen gewissen (unbekannten) Wert.



Stabilisierung relativer Häufigkeiten

Sei $A \subset \Omega$. Für eine feste *Stichprobe* $x = (x_1, \dots, x_n)$ mit $x_i \in \Omega$ besitzen die relativen Häufigkeiten

$$h_x(A) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}\{x_j \in A\}$$

die folgenden Eigenschaften:

- $0 \leq \; h_x(A) \leq 1, \qquad A \subset \Omega,$
- $h_x(\Omega) = 1,$
- $A \cap B = \emptyset \quad \Longrightarrow \quad h_x(A + B) = h_x(A) + h_x(B)$

Bezeichnen wir den (von A abhängenden) „Grenzwert“ im empirischen Gesetz über die Stabilisierung relativer Häufigkeiten mit $\mathbb{P}(A)$, so sollten für die Zuordnung $A \mapsto \mathbb{P}(A)$ die folgenden Eigenschaften gelten:

- $0 \leq \; \mathbb{P}(A) \leq 1, \qquad A \subset \Omega,$
- $\mathbb{P}(\Omega) = 1,$
- $A \cap B = \emptyset \quad \Longrightarrow \quad \mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B).$

Diskrete Wahrscheinlichkeitsräume

Für eine endliche oder abzählbare Menge Ω bezeichnen wir mit $\mathcal{P}(\Omega)$ die Potenzmenge von Ω.

Definition 3.1

Sei $\Omega = \{\omega_1, \omega_2, \dots\}$ ein höchstens abzählbarer, nichtleerer Grundraum und $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ eine Abbildung mit den Eigenschaften

(P1) $\mathbb{P}(\Omega) = 1$,

(P2) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$, falls $A_1, A_2, \dots \subset \Omega$ paarweise disjunkt sind.

Dann heißt $\mathbb{P}$ eine Wahrscheinlichkeitsverteilung oder ein Wahrscheinlichkeitsmaß auf Ω (bzw. auf $\mathcal{P}(\Omega)$), $\mathbb{P}(A)$ heißt Wahrscheinlichkeit von $A \subset \Omega$.

Die Eigenschaft (P2) wird als σ-Additivität von $\mathbb{P}$ bezeichnet. Das Paar $(\Omega, \mathbb{P})$ (bzw. das Tripel $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$) heißt diskreter Wahrscheinlichkeitsraum.

Bemerkung: Ist die Grundmenge Ω endlich, so spricht man auch von einem endlichen W.Raum. Axiom (P2) kann dann ersetzt werden durch das folgende schwächere Axiom:

(P2*) $A, B \subset \Omega, A \cap B = \emptyset \Longrightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Eigenschaften diskreter Wahrscheinlichkeitsräume

Satz 3.2

Sei $(\Omega, \mathbb{P})$ ein diskreter W.Raum. Seien $A, B, A_1, A_2, \dots, A_n \subset \Omega$. Dann gelten

(a) $\mathbb{P}(\emptyset) = 0$,

(b) $\mathbb{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ (falls $A_1, \dots, A_n$ paarweise disjunkt), (Additivität)

(c) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

(d) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, (Monotonie)

(e) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, (Additionsgesetz)

(f) $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$. (Subadditivität)

(g) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$. ($\sigma$-Subadditivität)

Beweis von Satz 3.2

Beweis

(a) Es gilt $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset \cup \ldots) \overset{\text{(P2)}}{=} \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \ldots$ und damit $\mathbb{P}(\emptyset) = 0,$ da $\mathbb{P}(\emptyset) \in \mathbb{R}$.

(b) Wähle zu $A_1, \ldots, A_n$ die Mengen $A_k := \emptyset$ für jedes $k > n$. Nutze (a) und (P2).

(c) Aus $\Omega = A + A^c$ folgt $1 \overset{\text{(P1)}}{=} \mathbb{P}(\Omega) \overset{\text{(b)}}{=} \mathbb{P}(A) + \mathbb{P}(A^c)$.

(d) Wegen $B = A + (B \setminus A)$ folgt $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.

(e) Es gilt (wegen (b)) $\mathbb{P}(A \cup B) = \mathbb{P}(A + (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ sowie $\mathbb{P}(B) = \mathbb{P}((B \cap A) + (B \setminus A)) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$. Umstellen der zweiten Gleichung und einsetzen in die erste liefert die Behauptung.

(f) Für $n = 2$ folgt dies aus (e), sodann für beliebiges n durch vollständige Induktion. (Alternativ kann (f) als Spezialfall von (g) angesehen werden.)

Beweis von Satz 3.2(g)

Beweis

Es bleibt (g) zu zeigen: Seien $A_1, A_2, \ldots \subset \Omega$. Definiere Mengen

$$B_1 := A_1,\ B_2 := A_2 \setminus A_1, \ldots, B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) \text{ für } n \geq 2.$$

Dann gilt

$$B_n \subset A_n,\ A_1 \cup \ldots \cup A_n = B_1 \cup \ldots \cup B_n, \quad n \in \mathbb{N},$$

$$B_k \cap B_l = \emptyset \quad \text{für } k \neq l,\ \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$

Es folgt

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

Warum nur diskrete W.Räume?

→ Problem der Existenz von $\mathbb{P}$ auf $\mathcal{P}(\Omega)$ mit (P1) und (P2).

→ **Ausweg:** Man erklärt $\mathbb{P}$ nur auf Teilmengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ mit

  - $\emptyset \in \mathcal{A}$,
  - $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
  - $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Ein Mengensystem $\mathcal{A}$ über $\Omega \neq \emptyset$ mit diesen Eigenschaften nennt man eine σ-Algebra.

- $\mathcal{P}(\Omega)$ ist stets eine σ-Algebra (aber im Allgemeinen zu groß, wenn Ω nicht abzählbar ist)
- Die systematische Konstruktion von σ-Algebren und Maßen (Wahrscheinlichkeitsmaßen) auf diesen erfolgt in der Maßtheorie (→ Analysis 3, später in dieser VL und in der Wahrscheinlichkeitstheorie, . . .)

Weitere Eigenschaften von W.Maßen

Satz 3.3 (Stetigkeit von W.Maßen)

Seien $(\Omega, \mathbb{P})$ ein diskreter W.Raum und $A_n \subset \Omega$ für $n \in \mathbb{N}$.

(a) Gilt $A_n \subset A_{n+1}$ für alle $n \in \mathbb{N}$, so folgt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mathbb{P}(A_i). \qquad \text{(Stetigkeit von unten)}$$

(b) Gilt $A_n \supset A_{n+1}$ für alle $n \in \mathbb{N}$, so folgt

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mathbb{P}(A_i). \qquad \text{(Stetigkeit von oben)}$$

Beweis von Satz 3.3

Beweis

(a) Setze $B_1 := A_1,\ B_2 := A_2 \setminus A_1, \ldots, B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1})$ für $n \geq 2$. Dann sind $B_1, B_2, \ldots$ paarweise disjunkt und es gilt für jedes $n \in \mathbb{N}$

$$A_n = \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i.$$

Deshalb folgt mit Hilfe der σ-Additivität und von Satz 3.2(b)

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(b) Folgt aus (a) mit Hilfe von Komplementbildung.

Einschluss-Ausschluss-Formel

Gegeben sei ein diskreter W.Raum Ω und Ereignisse $A_1, A_2, A_3 \subset \Omega$. Dann gilt

$$\begin{aligned}
\mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \\
\mathbb{P}((A_1 \cup A_2) \cup A_3) &= \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}((A_1 \cup A_2) \cap A_3) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_3) - \mathbb{P}((A_1 \cap A_3) \cup (A_2 \cap A_3)) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\
&\qquad - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3).
\end{aligned}$$

Satz 3.4 (Einschluss-Ausschluss-Formel)

Sei $(\Omega, \mathbb{P})$ ein (diskreter) W.Raum, $A_1, \ldots, A_n \subset \Omega$. Dann gilt

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Beweis: ... (Übung)

Bonferroni-Ungleichungen

Satz 3.5 (Einschluss-Ausschluss-Formel, Bonferroni-Ungleichungen)

Seien $(\Omega, \mathbb{P})$ ein (diskreter) W.Raum, $A_1, \ldots, A_n \subset \Omega$ und für $k = 1, \ldots, n$ sei

$$S_k := \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Dann gelten:

$$\text{(a)} \quad \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{k=1}^{n} (-1)^{k-1} S_k,$$

$$\text{(b)} \quad \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq \sum_{k=1}^{2s+1} (-1)^{k-1} S_k, \qquad s = 0, 1, \ldots, \left\lfloor \frac{n-1}{2} \right\rfloor,$$

$$\text{(c)} \quad \mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \geq \sum_{k=1}^{2s} (-1)^{k-1} S_k, \qquad s = 0, 1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor.$$

(ohne Beweis)

Festlegung von diskreten W.Maßen

Definition 3.6

Sei $(\Omega, \mathbb{P})$ ein diskreter W.Raum. Dann nennt man die durch

$$\mathrm{p}(\omega) := \mathbb{P}(\{\omega\}), \quad \omega \in \Omega,$$

erklärte Funktion $\mathrm{p} : \Omega \to [0,1]$ Wahrscheinlichkeitsfunktion oder Zähldichte von $\mathbb{P}$.

Ist $A \subset \Omega$ (und damit abzählbar), so gilt

$$\mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega \ : \ \omega \in A\}) = \mathbb{P}\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} \mathrm{p}(\omega).$$

Also ist $\mathbb{P}$ durch seine Zähldichte p schon vollständig festgelegt. Wegen $\mathbb{P}(\Omega) = 1$ folgt außerdem

$$\sum_{\omega \in \Omega} \mathrm{p}(\omega) = 1.$$

Zähldichten und W.Maße

Satz 3.7

Es seien Ω ein diskreter Grundraum und $\mathrm{p} : \Omega \to [0, \infty)$ eine Abbildung mit

$$\sum_{\omega \in \Omega} \mathrm{p}(\omega) = 1.$$

Dann ist durch

$$\mathbb{P}(A) := \sum_{\omega \in A} \mathrm{p}(\omega), \quad A \subset \Omega$$

eine Wahrscheinlichkeitsmaß $\mathbb{P}$ auf Ω definiert. $\mathbb{P}$ hat die Zähldichte p.

Beweis von Satz 3.7

Beweis

$\mathbb{P}(A) \geq 0$ für $A \subset \Omega$ ist klar, ebenso $\mathbb{P}(\Omega) = 1$.
Seien $A_1, A_2, \ldots \subset \Omega$ paarweise disjunkt. Dann gilt:

$$\mathbb{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{\omega \in \sum_{i=1}^{\infty} A_i} \mathrm{p}(\omega) \overset{(\star)}{=} \sum_{i=1}^{\infty} \sum_{\omega \in A_i} \mathrm{p}(\omega) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Bemerkung zu $(\star)$: Hier wurde der Umordnungssatz für Reihen mit nichtnegativen Summanden (Analysis 1) verwendet: Konvergenz und Wert sind unabhängig von der Summationsreihenfolge.

Beispiele für Zähldichten

Beispiel

Sei $\Omega = \mathbb{N}_0$ und seien $(p_k)_{k\in\mathbb{N}_0}$ wie folgt festgelegt:

(a) $p_k := \frac{1}{(k+1)(k+2)}$.

(b) $p_k := (1-a)a^k$, $a \in (0,1)$.

(c) $p_k := \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$, $\lambda > 0$.

(d) $p_k := \binom{n}{k}p^k(1-p)^{n-k}$ für $k \in \{0,\ldots,n\}$ und $p_k := 0$ sonst ($p \in [0,1]$).

Dann ist jeweils durch $\mathrm{p}(k) := p_k$, $k \in \mathbb{N}_0$ eine Zähldichte auf $\mathbb{N}_0$ festgelegt (und damit ein W.maß).

Nachweise ... an der Tafel bzw. Übung

Laplacescher Wahrscheinlichkeitsraum

Notation: $|A|$ bezeichnet die Anzahl der Elemente einer endlichen Menge A. ($\rightarrow$ Definition 4.1 allgemein!)

Beispiel (und Definition)

Sei $\Omega \neq \emptyset$ endlich. Für $A \subset \Omega$ setze

$$\mathbb{P}(A) := \frac{|A|}{|\Omega|}.$$

Man nennt $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ Laplaceschen Wahrscheinlichkeitsraum und $\mathbb{P}$ diskrete Gleichverteilung auf Ω.
Für die Zähldichte p von $\mathbb{P}$ gilt

$$\mathrm{p}(\omega) := \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

Zweifacher Würfelwurf

Beispiel

Modellierung mittels $\Omega := \{1, \dots, 6\}^2$, $\mathbb{P}(\{(i,j)\}) = \frac{1}{36} = \mathrm{p}(i,j), \quad (i,j) \in \Omega$.

Sei $A_5 :=$ "Augensumme ist 5" $= \{(i,j) \in \Omega : i + j = 5\}$.

Dann gilt

$$\mathbb{P}(A_5) = \sum_{\substack{(i,j)\in\Omega \\ i+j=5}} \mathrm{p}(i,j) = \mathrm{p}(1,4) + \mathrm{p}(3,2) + \mathrm{p}(2,3) + \mathrm{p}(4,1) = \frac{4}{36} = \frac{1}{9}$$

sowie

$$\begin{aligned} &\mathbb{P}(A_2) = \frac{1}{36}, \qquad \mathbb{P}(A_3) = \frac{2}{36}, \qquad \mathbb{P}(A_4) = \frac{3}{36}, \\ &\mathbb{P}(A_6) = \frac{5}{36}, \qquad \mathbb{P}(A_7) = \frac{6}{36} = \frac{1}{6}, \quad \dots \quad , \mathbb{P}(A_{12}) = \frac{1}{36}. \end{aligned}$$

Dreifacher Würfelwurf

Beispiel

Modellierung mittels $\Omega := \{1, \ldots, 6\}^3$, $\mathbb{P}(\{\omega\}) = \mathrm{p}(\omega) = \frac{1}{6^3} = \frac{1}{216}, \quad \omega \in \Omega$.

Betrachte die Ereignisse $A :=$ “Augensumme ist 11” und $B :=$ “Augensumme ist 12”.

A	641	632	551	542	533	443
#	6	6	3	6	3	3

$\rightarrow \sum = 27 = |A|$

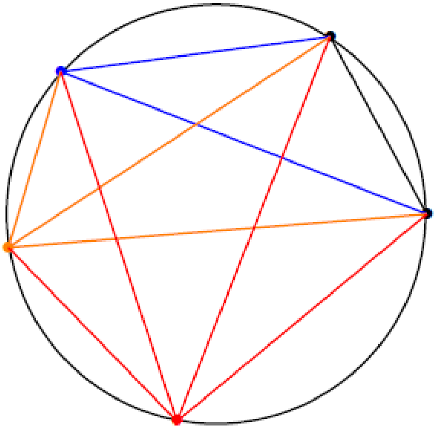
B	651	642	633	552	543	444
#	6	6	3	3	6	1

$\rightarrow \sum = 25 = |B|$

Kapitel 4: Kombinatorik

Kombinatorik ist die Kunst des (systematischen) Abzählens.

n Punkte auf Kreisrand
jeden mit jedem verbinden
Wie viele ($=: a_n$) Teile entstehen?
$a_1 = 1$
$a_2 = 2$
$a_3 = 4$
$a_4 = 8$
$a_5 = 16$
$a_6 \neq 32$



Geschlossener Ausdruck für a_n ?

Verbindung zur Stochastik

Erscheint ein Laplace–Modell in einer Situation angemessen, so ist die Wahrscheinlichkeit eines Ereignisses A gegeben durch

$$\mathbb{P}(A) \;=\; \frac{\text{Anzahl der für } A \text{ „günstigen“ Fälle}}{\text{Anzahl aller möglichen Fälle}}$$

Die Lehre vom systematischen Abzählen endlicher „strukturierter“ Mengen heißt Kombinatorik.

Kardinalität

Abzählen bedeutet, eine Bijektion zu einem Anfangsstück der natürlichen Zahlen anzugeben.

Definition 4.1 (Kardinalität)

Zwei Mengen A, B haben die gleiche Kardinalität, wenn es eine Bijektion zwischen A und B gibt, d.h.,

$$|A| = |B| :\Longleftrightarrow \exists f : A \to B,\ f \text{ bijektiv.}$$

Es ist

$$|A| = n \in \mathbb{N} :\Leftrightarrow |A| = |\{1, \dots, n\}|.$$

Setze $[n] := \{1, \dots, n\}$ für $n \in \mathbb{N}$.

Additionsregel oder Prinzip 1

Sind $M_1, \dots, M_k$ paarweise disjunkte Mengen, dann gilt

$$|M_1 \cup \ldots \cup M_k| = \sum_{i=1}^{k} |M_i|.$$

Für einen formalen Beweis würde man aus den Bijektionen zwischen M_i und $[|M_i|]$ eine Bijektion zwischen $M_1 \cup \ldots \cup M_k$ und $[|M_1| + \cdots + |M_k|]$ konstruieren. Man kann induktiv argumentieren, der wesentliche Punkt betrifft den Fall $k = 2$.

Für das auf der nächsten Folie formulierte Abzählprinzip geben wir eine anschauliche Beschreibung. Die Aussage kann wiederum mit vollständiger Induktion formal bewiesen werden.

Multiplikationsregel oder Prinzip 2

Gegeben seien endliche Mengen M_i und natürliche Zahlen $1 \leq m_i \leq |M_i|$ für $i = 1, \ldots, k$. Der Reihe nach werden Elemente $a_i \in M_i$ so ausgewählt, dass es für das

1. Element	$a_1 \in M_1$ stets m_1 Möglichkeiten,
2. Element	$a_2 \in M_2$ bei festgelegtem a_1 stets m_2 Möglichkeiten,
3. Element	$a_3 \in M_3$ bei festgelegten a_1, a_2 stets m_3 Möglichkeiten,
	$\vdots$
k. Element	$a_k \in M_k$ bei festgelegten $a_1, \ldots, a_{k-1}$ stets m_k Möglichkeiten

gibt. Dann gilt (mit vollst. Ind.):

Die Anzahl verschiedener solcher k–Tupel $(a_1, \ldots, a_k) \in M_1 \times \cdots \times M_k$ ist

$$\boxed{m_1 \cdot m_2 \ \cdots \ m_k.}$$

Typische Abzählprobleme

- Abbildungen $[k] \to [n]$, die injektiv / surjektiv / monoton / streng monoton sind
- Lotto mit / ohne Reihenfolge
- $\to$ 4 Grundtypen von “kombinatorischen Problemen”:
  mit/ohne Reihenfolge, mit/ohne Wiederholungen.

Mit Berücksichtigung der Reihenfolge: Permutationen

Definition 4.2 (k-Permutationen)

Sei M eine endliche Menge, $|M| = n$.

(a) Ein k-Tupel $(a_1, \dots, a_k) \in M^k$ heißt k-Permutation aus M mit Wiederholungen (mW), die Menge aller solcher k-Permutationen sei

$$\text{Per}_k^M(mW) := M^k.$$

(b) Ein k-Tupel $(a_1, \dots, a_k) \in M^k$ mit $a_i \neq a_j$ für alle $i \neq j$ heißt k-Permutation aus M ohne Wiederholungen (oW), die Menge aller solcher k-Permutationen sei

$$\text{Per}_k^M(oW).$$

Man schreibt auch $\text{Per}_k^n(mW)$ bzw. $\text{Per}_k^n(oW)$, falls $M = \{1, \dots, n\}$ gilt.

Ohne Berücksichtigung der Reihenfolge: Kombinationen

Definition 4.3 (k-Kombinationen)

Sei $M = \{a_1, \ldots, a_n\}$, $|M| = n$.

(a) Ein k-Tupel $(a_{i_1}, \ldots, a_{i_k}) \in M^k$ mit $1 \leq i_1 \leq \ldots \leq i_k \leq n$ heißt k-Kombination aus M mit Wiederholungen, die Menge aller solcher k-Kombinationen sei

$$\mathrm{Kom}_k^M(mW).$$

(b) Ein k-Tupel $(a_{i_1}, \ldots, a_{i_k}) \in M^k$ mit $1 \leq i_1 < \ldots < i_k \leq n$ heißt k-Kombination aus M ohne Wiederholungen, die Menge aller solcher k-Kombinationen sei

$$\mathrm{Kom}_k^M(oW).$$

Man schreibt auch $\mathrm{Kom}_k^n(mW)$ bzw. $\mathrm{Kom}_k^n(oW)$, falls $M = \{1, \ldots, n\}$ gilt.

Bezeichnungen

Für $n \in \mathbb{N}$ und $k \in \{0, \dots, n\}$ sei

$$0! := 1, \quad n! := 1 \cdots n \quad \text{sowie} \quad \binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

Dann gilt

$$\binom{n}{k} = \binom{n}{n-k}, \quad \binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n$$

sowie

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

Im Falle $k = 0$ setzt man dafür $n \cdots (n-k+1) := 1$ ("leeres Produkt").
Für $n \in \mathbb{N}$ und $k \in \{1, \dots, n-1\}$ gilt

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \to \text{ Pascalsches Dreieck}$$

→ kombinatorische Begründung folgt!

Basiswissen Anzahlbestimmung

Satz 4.4

Sei M eine Menge, $|M| = n$. Dann gilt

(a) $|\,\mathrm{Per}_k^M(mW)| = n^k, \;\; k \in \mathbb{N},$

(b) $|\,\mathrm{Per}_k^M(oW)| = n^{\underline{k}} := \frac{n!}{(n-k)!}, \;\; k \in \mathbb{N}, k \leq n,$

(c) $|\,\mathrm{Kom}_k^M(mW)| = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}, \;\; k \in \mathbb{N},$

(d) $|\,\mathrm{Kom}_k^M(oW)| = \binom{n}{k}, \;\; 1 \leq k \leq n.$

(a), (b) folgen sofort aus Prinzip 2.

Beweis von Satz 4.4 (d)

Beweis

Sei $M = \{1, \dots, n\}$. Für $(i_1, \dots, i_k) \in \mathrm{Kom}_k^n(oW)$ setze

$$M(i_1, \dots, i_k) := \{(i_{\pi(1)}, \dots, i_{\pi(k)}) \ : \ \pi \in \mathrm{Per}_k^k(oW)\} \qquad \text{(Bijektion zu } [k])$$

$\Rightarrow |M(i_1, \dots, i_k)| = k!$

Ferner ist $\mathrm{Per}_k^n(oW) = \sum_{(i_1,\dots,i_k) \in \mathrm{Kom}_k^n(oW)} M(i_1, \dots, i_k)$, also

$$\frac{n!}{(n-k)!} = |\,\mathrm{Per}_k^n(oW)| = \sum_{(i_1,\dots,i_k) \in \mathrm{Kom}_k^n(oW)} \underbrace{|M(i_1, \dots, i_k)|}_{=k!}$$

$$= |\,\mathrm{Kom}_k^n(oW)| \cdot k!$$

$$\Rightarrow \mathrm{Kom}_k^n(oW) = \frac{n!}{k!(n-k)!}.$$

Beweis von Satz 4.4 (c)

Beweis

Betrachte

$$f : \mathrm{Kom}_k^n(mW) \to \mathrm{Kom}_k^{n+k-1}(oW),$$
$$(a_1, \ldots, a_k) \mapsto (b_1, \ldots, b_k),$$

$$\text{wobei } 1 \leq a_1 \leq \ldots \leq a_k \leq n, \quad 1 \leq b_1 < \ldots < b_k \leq n+k-1$$

mit $b_1 := a_1, b_2 := a_2 + 1, \ldots, b_i := a_i + i - 1, \; i = 2, \ldots, k$.

Die Abbildung ist wohldefiniert und bijektiv. $\rightarrow$ Behauptung.

Alternative Sichtweisen

Zu (d): Es gilt

$$\mathrm{Kom}_k^{[n]}(oW) = \mathrm{Kom}_k^{[n-1]}(oW) + \left(\mathrm{Kom}_{k-1}^{[n-1]}(oW) \times \{n\}\right).$$

Induktiv folgt damit

$$\left|\mathrm{Kom}_k^{[n]}(oW)\right| = \binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}.$$

Zu (c): Jedes k-Tupel aus $\mathrm{Kom}_k^M(mW)$ lässt sich (eindeutig) als Punkt-Strich-Code der Länge $k+n-1$ schreiben (Bijektion!). Beispiel:

$$n = 8, k = 7 : (1,1,1,3,4,6,6) \quad \rightarrow \quad \cdot\cdot\cdot||\cdot|\cdot||\cdot\cdot||$$

Im Code kommt genau k mal “$\cdot$” und $(n-1)$ mal “$|$” vor. Also ist $|\,\mathrm{Kom}_k^M(mW)|$ die Anzahl aller Möglichkeiten, die k Punkte auf $k+n-1$ Plätze zu verteilen, d.h., k aus $k+n-1$ auszuwählen.

Binomischer Lehrsatz

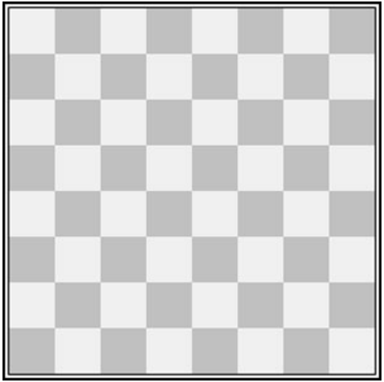
Beispiel

$$
\begin{aligned}
(x+y)^n &= (x+y)\cdots(x+y) \\
&= \sum_{I\subset[n]} x^{|I|}\cdot y^{n-|I|} = \sum_{k=0}^{n}\sum_{\substack{|I|=k\\ I\subset[n]}} x^k y^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.
\end{aligned}
$$

Türme auf dem Schachbrett

Beispiel

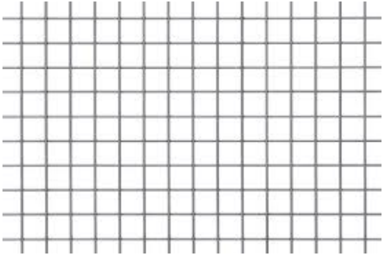
Wieviele Möglichkeiten gibt es, 8 Türme (unterscheidbare / nicht unterscheidbare) auf einem Schachbrett so aufzustellen, dass diese sich gegenseitig nicht schlagen können?



Anzahlen von Wegen

Beispiel

Wieviele Wege minimaler Länge gibt es zwischen zwei Punkten auf einem ganzzahligen Gitter?



Ganzzahlige Lösungen von Gleichungen

Beispiel

Es seien $n, k \in \mathbb{N}$ mit $k \leq n$. Wieviele verschiedene Lösungen hat die folgende Gleichung?

$$x_1 + x_2 + \cdots + x_k = n, \quad x_i \in \mathbb{N}_0$$

Codiere wieder mit Strichen und Punkten. Beispiel: $x_1 = 3, x_2 = 0, \ldots, x_k = 1$

$$\left.\begin{array}{r}\rightarrow 1+1+1|0| \quad \ldots \quad |1 \\ \rightarrow \cdot\cdot\cdot|| \quad \ldots \quad |\cdot\end{array}\right\} n\text{-mal } “\cdot”, k-1\text{-mal } “|”$$

$$\Rightarrow \binom{n+k-1}{k-1} \text{ verschiedene Lösungen.}$$

Klassenzerlegungen

Beispiel (Klassenzerlegungen)

Sei $M \neq \emptyset$ eine n-elementige Menge. Eine Familie $\{K_1, \ldots, K_r\}$ von Teilmengen $K_1, \ldots, K_r$ von M mit

$$K_i \neq \emptyset$$

$$K_i \cap K_j = \emptyset \text{ für } i \neq j \text{ und}$$

$$M = K_1 \cup \ldots \cup K_r,\ r \in \{1, \ldots, n\}$$

heißt Klassenzerlegung von M.
Sei k_n die Anzahl der Klassenzerlegungen von M. (Es sei o.B.d.A. $M = [n]$.)

Es gilt: $k_1 = 1$, $k_2 = 2$, $k_3 = 5$:
$\{\{1, 2, 3\}\}$, $\{\{1, 2\}, \{3\}\}$, $\{\{1, 3\}, \{2\}\}$, $\{\{2, 3\}, \{1\}\}$, $\{\{1\}, \{2\}, \{3\}\}$.

Rekursionsformel: Unterscheide danach, welcher r-elementigen Menge "$n+1$" zugeordnet wird. Es gibt $\binom{n}{r}$ Möglichkeiten, eine r-elementige Menge aus $[n]$ auszuwählen, der "$n+1$" zugeordnet wird. Dann gibt es noch k_{n-r} verschiedene Klassenzerlegungen der restlichen $n-r$ Elemente. Mit $k_0 := 1$ gilt dann :
$k_{n+1} = \sum\limits_{r=0}^{n} \binom{n}{r} k_{n-r}$. → Übung !

Geordnete Zerlegungen

Beispiel (Geordnete Zerlegungen einer Menge M)

Sei $n \in \mathbb{N}$, $M = [n]$, $r \in \mathbb{N}$, $n_1, \ldots, n_r \in \mathbb{N}_0$, $n_1 + \cdots + n_r = n$. Setze

$$\mathcal{M}^n_{n_1,\ldots,n_r} := \left\{ (A_1, \ldots, A_r) \ : \ M = \sum_{i=1}^{r} A_i, |A_i| = n_i, i = 1, \ldots, r \right\}.$$

$$\begin{aligned} \Rightarrow \left| \mathcal{M}^n_{n_1,\ldots,n_r} \right| &= \binom{n}{n_1} \binom{n - n_1}{n_2} \cdots \binom{n - n_1 - \cdots - n_{r-1}}{n_r} = \frac{n!}{n_1! \cdots n_r!} \\ &=: \binom{n}{n_1, \ldots, n_r} \qquad \text{(Multinomialkoeffizient)} \end{aligned}$$

Dieser Multinomialkoeffizient ist als 0 definiert, falls $n_i \notin \mathbb{N}_0$ oder $n_1 + \cdots + n_r \neq n$. Setze

$$\mathcal{M}^n_r := \sum_{\substack{n_1,\ldots,n_r=0 \\ n_1+\cdots+n_r=n}}^{n} \mathcal{M}^n_{n_1,\ldots,n_r}.$$

Multinomialformel

Beispiel (Multinomialformel)

$$\begin{aligned}
(x_1 + \cdots + x_r)^n &= \sum_{(A_1,\ldots,A_r)\in\mathcal{M}_r^n} x_1^{|A_1|}\cdots x_r^{|A_r|} \\
&= \sum_{\substack{n_1,\ldots,n_r\in\mathbb{N}_0 \\ n_1+\cdots+n_r=n}} \; \sum_{(A_1,\ldots,A_r)\in\mathcal{M}^n_{n_1,\ldots,n_r}} x_1^{n_1}\cdots x_r^{n_r} \\
&= \sum_{n_1,\ldots,n_r} \binom{n}{n_1,\ldots,n_r} x_1^{n_1}\cdots x_r^{n_r}.
\end{aligned}$$

Laplace-Würfel

Beispiel (Laplace-Würfel)

Ein fairer Würfel wird n-mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass n_1-mal $1, \dots, n_6$-mal 6 fällt?

$$\to \Omega = [6]^n, A = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega \ : \ |\{i \in [n] \ : \omega_i = j\}| = n_j, j = 1, \dots, 6\}$$

Dann ist

$$A \overset{\text{bijektiv}}{\leftrightarrow} \mathcal{M}^n_{n_1,\dots,n_6} \Rightarrow |A| = \frac{n!}{n_1! \cdots n_6!}.$$

Laplace $\Rightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{n!}{n_1! \ \cdots \ n_6! \cdot 6^n}$.

Geburtstags-Problem

Beispiel ("Geburtstags-Problem")

Es seien k Personen im Raum.

$A =$ "Mindestens zwei haben am gleichen Tag Geburtstag"

Modell: Laplace-Modell. Realistisch?

$$\Omega = \{1, \ldots, 365\}^k, \; |\Omega| = 365^k$$
$$A = \{\omega \in \Omega \; : \; \omega_i = \omega_j \text{ für ein Paar } (i, j) \text{ mit } i \neq j\}.$$

$$A^c = \{\omega \in \Omega \; : \; \omega_i \neq \omega_j \text{ für } i \neq j\}$$
$$|A^c| = 365 \cdot (365 - 1) \cdots (365 - k + 1), \text{ falls } k \leq 365$$
$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{|A^c|}{|\Omega|} = 1 - (1 - \tfrac{1}{365}) \cdots (1 - \tfrac{k-1}{365})$$

k	20	23	40	150
$\mathbb{P}(A)$	0,411	0,507	0,891	$1 - 10^{-15}$

Vergleiche: k-maliges Ziehen aus einer Urne mit 365 Kugeln und Zurücklegen.

Weitere Beispiele: Übung/Tutorium

- Geburtstagsproblem
- Stimmzettelprobleme / Irrfahrten auf Gitter
- Qualitätskontrolle
- Vertauschte Briefe / Fixpunktfreie Permutation
- Erste Kollision, Sammelbilderprobleme . . .

Modellwechsel

$\mathrm{Per}_k^n, \mathrm{Kom}_k^n$ lassen sich beschreiben im

Urnenmodell	k-maliges Ziehen aus Urne mit n Kugeln $K_1, \ldots, K_n$ mit/ohne Beachten der Reihenfolge mit/ohne Zurücklegen ($\widehat{=}$ Wiederholungen)
Teilchen-Fächer -Modell	Verteilen von k Teilchen auf n Fächer $F_1, \ldots, F_n$ unterscheidbare/nicht unterscheidbare Teilchen Mehrbelegung von Fächern erlaubt/nicht erlaubt

Beispiel. Rückblick auf Geburtstagsproblem mit $k \leq 365$ Personen:

- Personen $P_1, \ldots, P_k$ (Teilchen, unterscheidbar) werden auf Tage (Fächer) $T_1, \ldots, T_{365}$ verteilt (Mehrfachbelegung erlaubt).
- Aus Urne mit Kugeln $T_1, \ldots, T_{365}$ werden k Kugeln (Geburtstage der Personen $P_1, \ldots, P_k$) gezogen (mit Berücksichtigung der Reihenfolge und mit Zurücklegen).

Interpretation (Physik)

Statistische Physik	k unterscheidbare Teilchen	k nicht unterscheidbare Teilchen	
Mehrfachbelegung ok (kein Pauliprinzip) Fächer $1, \dots, n$	n^k Maxwell - Boltzmann	$\binom{n+k-1}{k}$ Bose - Einstein	Ziehen mit Zurücklegen aus Kugeln $1, \dots, n$
keine Mehrfachbelegung (Pauliprinzip) Fächer $1, \dots, n$	$n^{\underline{k}}$	$\binom{n}{k}$ Fermi - Dirac	Ziehen ohne Zurücklegen aus Kugeln $1, \dots, n$
	geordnete Stichprobe (Beachten der Reihenfolge) der Länge k	ungeordnete Stichprobe der Länge k	Urnenmodell "Ziehen aus Urne"

Bosonen: Photon, Graviton (Elementarteilchen); Cooper-Paare (Elektron-Phonon), Atomkerne mit gerader Nukleonenzahl, Mesonen (Quark-Antiquark-Paare), flüssiges Helium

Fermionen: Leptonen (Elektronen, Neutrinos), Quarks; Baryonen (Proton, Neutron), Pentaquarks

Kapitel 5: Zufallsvariablen und Indikatorfunktionen

Motivation für die Einführung von Zufallsvariablen (speziell von Indikatorfunktionen):

theoretisch

- Funktionen als analytisches Hilfsmittel nutzen,
- eine stukturierte Beschreibung von Ereignissen/Mengen ermöglichen.

praktisch

Oft ist bei einem Zufallsexperiment nicht das Ergebnis ω selbst interessant, sondern eine Funktion $X(\omega)$ davon, z.B. die Augensumme beim zweifachen Würfeln.

Zufallsvariablen

Definition 5.1

Seien $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $\Omega' \neq \emptyset$ eine beliebige Menge. Eine Abbildung

$$X : \Omega \to \Omega'$$

heißt Ω'-wertige Zufallsvariable. Für $\Omega' = \mathbb{R}^d$ heißt X d-dimensionaler Zufallsvektor, für $\Omega' = \mathbb{R}$ reelle Zufallsvariable oder kurz Zufallsvariable.

Bemerkungen.

(1) Das W.Maß $\mathbb{P}$ ist für die hier gegebene Definition nicht erforderlich.

(2) In einem diskreten W.Raum ist das Bild $X(\Omega)$ der ZV X stets abzählbar.

(3) Später: Zufallsvariablen auf allgemeinen Wahrscheinlichkeitsräumen $(\Omega, \mathcal{A}, \mathbb{P})$ beziehungsweise auf Messräumen $(\Omega, \mathcal{A})$ müssen eine Kompatibilitätsbedingung erfüllen: Messbarkeit! (vgl. Analysis 3)

(4) Notation: vorzugsweise große lateinische Buchstaben vom Ende des Alphabets

Beispiel n-facher Würfelwurf

Beispiel

Sei $\Omega = \{1, \dots, 6\}^n$. Beispiele für (reelle) Zufallsvariablen sind

- $X(\omega) := \omega_1 + \cdots + \omega_n$, $\omega = (\omega_1, \dots, \omega_n) \in \Omega$,
- $Y(\omega) := \max\{\omega_i \; : \; 1 \leq i \leq n\}$,
- $Z(\omega) := \omega_i$ für ein festes $i \in \{1, \dots, n\}$.

Urbilder

Es sei $X : \Omega \to \Omega'$ eine Ω'-wertige Zufallsvariable. Für $M \subset \Omega'$ sei

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

das Urbild von M unter X.

Die Urbildabbildung $X^{-1} : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega)$ hat die folgenden Eigenschaften: (Hierbei sei I eine abzählbare Indexmenge und $M_i \subset \Omega', i \in I$.)

- $X^{-1}(X(\Omega)) = \Omega$ und $X^{-1}(X(\emptyset)) = \emptyset$,
- $X^{-1}\left(\bigcup_{i\in I} M_i\right) = \bigcup_{i\in I} X^{-1}(M_i)$,
- $X^{-1}\left(\sum_{i\in I} M_i\right) = \sum_{i\in I} X^{-1}(M_i)$,
- $X^{-1}\left(\bigcap_{i\in I} M_i\right) = \bigcap_{i\in I} X^{-1}(M_i)$,
- $X^{-1}(M^{\mathrm{c}}) = (X^{-1}(M))^{\mathrm{c}}$.

Rechnen mit reellen Zufallsvariablen

Notation: Für eine ZV $X : \Omega \to \mathbb{R}$ und $t \in \mathbb{R}$ sei

$$\{X = t\} := X^{-1}(\{t\}) = \{\omega \in \Omega \; : \; X(\omega) = t\}.$$

Analog erklärt man die Ereignisse $\{X \leq t\}, \{X < t\}$ etc.

Operationen: Für reelle Zufallsvariablen X, Y auf Ω erklärt man

- die ZV $X \cdot Y$ durch $(X \cdot Y)(\omega) := X(\omega)Y(\omega)$, $\omega \in \Omega$,
- die ZV $\max\{X, Y\}$ durch $\max\{X, Y\}(\omega) := \max\{X(\omega), Y(\omega)\}$, $\omega \in \Omega$,
- das Ereignis $\{X \leq Y\} := \{\omega \in \Omega \; : \; X(\omega) \leq Y(\omega)\}$.

Beispiel

Zweifacher Würfelwurf Sei $\Omega = \{1, \dots, 6\}^2$ und $X(i,j) := i$, $Y(i,j) := j$. Dann ist

$$\{X \leq Y\} = \{(i,j) \in \Omega \; : \; i \leq j\}, \qquad \{X - 2Y > 0\} = \{(i,j) \in \Omega \; : \; 2j < i\}.$$

Indikatorfunktionen

Definition 5.2 (Indikatorfunktion)

Sei $A \subset \Omega$. Durch

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A, \end{cases}$$

wird eine Zufallsvariable $\mathbb{1}_A : \Omega \to \mathbb{R}$ erklärt, die Indikatorfunktion von A.
Alternative Notation : $\mathbb{1}\{\omega \in A\} := \mathbb{1}_A(\omega)$

Was ist der Zusammenhang zwischen Zufallsvariablen und Indikatorfunktionen?

Rechenregeln für Indikatorfunktionen

Satz 5.3 (Rechenregeln)

Seien $A, B \subset \Omega$. Dann gilt

(1) $\mathbb{1}_\emptyset \equiv 0, \mathbb{1}_\Omega \equiv 1$,

(2) $\mathbb{1}_A^2 = \mathbb{1}_A$,

(3) $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$,

(4) $\mathbb{1}_{A\cap B} = \mathbb{1}_A \cdot \mathbb{1}_B, \mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}$,

(5) $A \subset B \Rightarrow \mathbb{1}_A \leq \mathbb{1}_B$.

Beweis: siehe Tutorium

Indikatorsummen

Definition

Seien $A_1, \dots, A_n \subset \Omega$ Ereignisse. Man nennt die ZV

$$X := \sum_{j=1}^{n} \mathbb{1}_{A_j} = \sum_{j=1}^{n} \mathbb{1}\{A_j\}$$

Indikatorsumme oder Zählvariable.

Es gilt

- $\{X = 0\} = A_1^c \cap \ldots \cap A_n^c$,
- $\{X = n\} = A_1 \cap \ldots \cap A_n$,
- $\{X \neq 0\} = A_1 \cup \ldots \cup A_n$,
- $\{X = k\} = \sum_{T \subset [n], |T| = k} \left(\bigcap_{j \in T} A_j \cap \bigcap_{j \notin T} A_j^c \right)$ mit $\bigcap_{j \in \emptyset} A_j := \Omega, k \in \mathbb{N}$.

Einschluss-Ausschluss-Formel für Indikatoren

Satz 5.4 (Einschluss-Ausschluss-Formel)

Seien $A_1, \dots, A_n \subset \Omega$. Dann gilt

$$\mathbb{1}_{A_1 \cup \ldots \cup A_n} = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbb{1}_{A_{i_1} \cap \ldots \cap A_{i_k}} \, .$$

Beweis: siehe Tutorium

Verteilung einer ZV

Definition und Satz 5.5

Es sei $(\Omega, \mathbb{P})$ ein diskreter W.Raum und $X : \Omega \to \Omega'$ eine Ω'-wertige ZV. Sei $\widetilde{\Omega} := X(\Omega)$. Durch $\mathbb{P}^X : \mathcal{P}(\widetilde{\Omega}) \to [0, 1]$,

$$A' \mapsto \mathbb{P}^X(A') := \mathbb{P}\Big(X^{-1}(A')\Big)$$

wird ein Wahrscheinlichkeitsmaß auf $\widetilde{\Omega}$ definiert. $\mathbb{P}^X$ heißt Verteilung von X (unter $\mathbb{P}$).

Bemerkungen:

- Die Bildmenge $\widetilde{\Omega}$ ist stets abzählbar.
- Das Paar $(\widetilde{\Omega}, \mathbb{P}^X)$ ist wieder ein diskreter W.raum.

Beweis von Satz 5.5

Beweis

Wir rechnen die Eigenschaften eines Wahrscheinlichkeitsmaßes nach.
Es ist $\mathbb{P}^X(\cdot) \in [0,1]$ und $\mathbb{P}^X(\widetilde{\Omega}) = \mathbb{P}\Big(X^{-1}(\widetilde{\Omega})\Big) = \mathbb{P}(\Omega) = 1$.
Sind $\widetilde{A}_1, \widetilde{A}_2, \ldots \subset \widetilde{\Omega}$ paarweise disjunkt, so folgt:

$$\begin{aligned}\mathbb{P}^X\Big(\sum_{j=1}^{\infty} \widetilde{A}_i\Big) &= \mathbb{P}\Big(X^{-1}(\sum_{j=1}^{\infty} \widetilde{A}_j)\Big) = \mathbb{P}\Big(\sum_{j=1}^{\infty} X^{-1}(\widetilde{A}_j)\Big) \\ &= \sum_{j=1}^{\infty} \mathbb{P}\Big(X^{-1}(\widetilde{A}_j)\Big) = \sum_{j=1}^{\infty} \mathbb{P}^X(\widetilde{A}_j),\end{aligned}$$

also ist $\mathbb{P}^X$ auch σ-additiv.

Bemerkung

Die Abbildung X bewirkt einen **Maß-Transport**:

$$(\Omega, \mathbb{P}) \quad \xrightarrow{X} \quad (\widetilde{\Omega}, \mathbb{P}^X).$$

Aufgrund der Eigenschaften von Abbildungen kann $(\widetilde{\Omega}, \mathbb{P}^X)$ als Vergröberung von $(\Omega, \mathbb{P})$ angesehen werden. Die disjunkten Mengen

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

bilden wegen

$$\Omega = \sum_{\omega' \in \widetilde{\Omega}} X^{-1}(\{\omega'\})$$

eine Zerlegung von Ω. Ist man nur an der ZV X interessiert, so kann man auch mit $(\widetilde{\Omega}, \mathbb{P}^X)$ arbeiten.

Zweifacher Würfelwurf

Beispiel

Kanonischer Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ für den zweifachen Würfelwurf:

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

und $\mathbb{P}$ ist die Gleichverteilung auf Ω.

Augensumme als Zufallsvariable: $X : \Omega \to \mathbb{R}, X\big((i,j)\big) = i + j$.

Dann ist $\widetilde{\Omega}(= X(\Omega)) = \{2, 3, \ldots, 12\}$ und für $k \in \tilde{\Omega}$ gilt

$$\mathbb{P}^X(\{k\}) = \mathbb{P}\Big(\{(i,j) : i + j = k\}\Big) = \frac{6 - |k - 7|}{36}.$$

Nochmal Diskrete W.Räume

Manchmal ist es nützlich, für Ω auch überabzählbare Mengen zuzulassen (etwa $\Omega = \mathbb{R}$).

Definition 5.6 (Diskreter Wahrscheinlichkeitsraum erweitert)

Sei $\Omega \neq \emptyset$ und $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ eine Abbildung mit

(P1) Es gibt $\Omega_0 \subset \Omega$ abzählbar mit $\mathbb{P}(\Omega_0) = 1$.

(P2) $\mathbb{P}\left(\sum_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mathbb{P}(A_i)$ für (paarweise disjunkte) $A_1, A_2, \ldots \subset \Omega$.

Dann nennt man $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ einen diskreten Wahrscheinlichkeitsraum und $\mathbb{P}$ ein diskretes Wahrscheinlichkeitsmaß.

Beobachtungen:

- Vergleich mit Def 3.1: Ist Ω abzählbar, so sind beide Definitionen äquivalent.
- $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, endliche Additivität, Monotonie, ... gelten wie vorher.
- $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega_0)$ für $A \subset \Omega$.
- Ω_0 ist nicht eindeutig. Aber es ist eine (eindeutige) minimale Wahl möglich!
- $(\Omega_0, \mathbb{P}_0)$ mit $\mathbb{P}_0(A) := \mathbb{P}(A)$, $A \subset \Omega_0$ ist diskreter W.Raum im Sinne der früheren Definition.

Verteilung reeller Zufallsvariablen

Im Rahmen dieser erweiteren Definition diskreter W.Räume lassen sich etwa die Verteilungen reeller ZV als Wahrscheinlichkeitsmaße auf $\mathbb{R}$ interpretieren.

Satz und Definition 5.7

Für einen diskreten W.Raum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ und eine Zufallsvariable $X : \Omega \to \mathbb{R}$ sei

$$\mathbb{P}^X(M) := \mathbb{P}(X^{-1}(M)) = \mathbb{P}(\{X \in M\}) \quad \text{für } M \subset \mathbb{R}.$$

Dann ist $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathbb{P}^X)$ ein diskreter Wahrscheinlichkeitsraum. Man nennt auch dieses diskrete Wahrscheinlichkeitsmaß $\mathbb{P}^X$ die Verteilung von X.

Verteilung reeller Zufallsvariablen

Wir lassen den Zusatz “diskret” häufig weg, wenn dies aus dem Kontext klar ist.

Ist Q ein beliebiges diskretes W.maß auf $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$, so kann man Q stets als Verteilung einer ZV X auf einem diskreten W.raum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ auffassen: Wähle $\Omega := \mathbb{R}$, $\mathbb{P} := Q$, $X := \mathrm{id}_\Omega$. Dann gilt $\mathbb{P}^X = Q$.

Definition 5.8

(a) Sei X eine reelle ZV auf einem diskreten W.raum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ und Q ein diskretes W.maß auf $\mathbb{R}$. Falls $Q = \mathbb{P}^X$ gilt, sagen wir *X hat die Verteilung Q* und schreiben hierfür auch $X \sim Q$.

(b) Seien X, Y reelle ZV auf diskreten W.räumen $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ bzw. $(\Omega', \mathcal{P}(\Omega'), \mathbb{P}')$. Falls $\mathbb{P}^X = \mathbb{P}^Y$ gilt, sagen wir *X und Y haben dieselbe Verteilung* und schreiben $X \sim Y$ (oder $X \overset{\mathcal{D}}{=} Y$).

Binomialverteilung

Beispiel

Für $n \in \mathbb{N}_0$, $p \in [0,1]$ betrachten wir den diskreten W.raum $(\Omega, \mathbb{P})$ mit $\Omega = \mathbb{R}$ und

$$\mathbb{P}(\{k\}) := \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in \{0, \dots, n\}, \\ 0, & \text{sonst.} \end{cases}$$

(Wie üblich vereinbaren wir $0^0 := 1$.)
Hier kann man $\Omega_0 = \{0, \dots, n\}$ oder $\Omega_0 = \mathbb{N}_0$ wählen. Dann gilt $\mathbb{P}(\Omega_0) = 1$.
Sei X Zufallsvariable auf $(\Omega, \mathbb{P})$ mit $X = \mathrm{id}_{\mathbb{R}}$. Dann gilt $X \sim \mathbb{P}$ bzw. $X \sim \mathrm{Bin}(n, p)$.

Poissonverteilung

Beispiel

Sei $\lambda > 0$ und $(\Omega, \mathbb{P})$ mit $\Omega = \mathbb{R}$ und

$$\mathbb{P}(\{k\}) := \begin{cases} \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, & k \in \mathbb{N}_0, \\ 0, & \text{sonst.} \end{cases}$$

Hier kann zum Beispiel $\Omega_0 = \mathbb{N}_0$ gewählt werden. $\mathbb{P}$ ist selbst ein diskretes W.maß auf $\mathbb{R}$.

Sei Y eine Zufallsvariable auf Ω mit $Y = \mathrm{id}_{\mathbb{R}}$. Dann gilt $\mathbb{P}^Y = \mathbb{P}$ d.h. $Y \sim \mathbb{P}$ bzw. $Y \sim \mathrm{Po}(\lambda)$.

Sei weiterhin $(\Omega', \mathbb{P}')$ gegeben mit $\Omega' := \mathbb{N}_0$ und

$$\mathbb{P}'(\{k\}) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad k \in \mathbb{N}_0.$$

Betrachte die ZV $Z : \Omega' \to \mathbb{R}$, $\omega \mapsto \omega$. Dann gilt $\mathbb{P}^Z(\{k\}) = \mathbb{P}(\{k\})$, $k \in \mathbb{N}_0$ sowie $\mathbb{P}^Z(\{k\}) = 0$ sonst (d.h. für $k \in \mathbb{R} \setminus \mathbb{N}_0$). Also gilt $\mathbb{P}^Z = \mathbb{P} = \mathbb{P}^Y$ und somit

$$Z \sim Y.$$

Kapitel 6: Der Erwartungswert

Der Erwartungswert ist eine wichtige (einfache) Kenngröße für ZVen auf einem W.raum $(\Omega, \mathbb{P})$.

Motivation: Ein Zufallsexperiment (etwa ein Glücksspiel) mit der Ergebnismenge $\Omega = \{\omega_1, \ldots, \omega_s\}$ wird wiederholt durchgeführt. Sei $X(\omega_i)$ die Auszahlung bei Ergebnis ω_i.

Frage: Wie groß ist die durchschnittliche Auszahlung bei wiederholter Durchführung des Experiments?

Analyse: n Durchführungen, bei H_j davon trete ω_j ein.

Gesamtgewinn: $X(\omega_1) \cdot H_1 + X(\omega_2) \cdot H_2 + \cdots + X(\omega_s) \cdot H_s$

Durchschnittlicher Gewinn:

$$X(\omega_1) \cdot \underbrace{\frac{H_1}{n}}_{\approx \mathbb{P}(\{\omega_1\})} + \cdots + X(\omega_s) \cdot \underbrace{\frac{H_s}{n}}_{\approx \mathbb{P}(\{\omega_s\})} \quad \leftarrow \text{ empirisch}$$

Definition des Erwartungswerts

Definition 6.1

Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum, $\Omega_0 \subset \Omega$ abzählbar, $\mathbb{P}(\Omega_0) = 1$. Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Man sagt, der Erwartungswert von X existiert, falls

$$\sum_{\omega \in \Omega_0} |X(\omega)| \, \mathbb{P}(\{\omega\}) < \infty \tag{$\star$}$$

gilt. In diesem Fall heißt

$$\mathbb{E}[X] := \sum_{\omega \in \Omega_0} X(\omega) \, \mathbb{P}(\{\omega\})$$

der Erwartungswert von X bezüglich $\mathbb{P}$.

Notation: $\mathbb{E}X, \mathbb{E}(X), \mathbb{E}_{\mathbb{P}}(X), \ldots$

- Ist Ω endlich, so ist $(\star)$ stets erfüllt.
- Ist $X \geq 0$, so erklärt man $\mathbb{E}[X]$ wie oben und lässt $\mathbb{E}[X] = \infty$ zu. Dann gilt: der Erwartungswert von X existiert genau dann, wenn $\mathbb{E}[|X|] < \infty$.
- Die Definition ist unabhängig von der Wahl von Ω_0.

Eigenschaften von Erwartungswerten

Satz 6.2 (Eigenschaften von Erwartungswerten)

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$. Sei $L^1 := \{X : \Omega \to \mathbb{R} \ : \ \mathbb{E}|X| < \infty\}$. Dann ist L^1 ein reeller Vektorraum. Für $X, Y \in L^1$, $a \in \mathbb{R}$ gilt:

$$\left.\begin{array}{ll}\text{(a)} & \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y], \\ \text{(b)} & \mathbb{E}[a\,X] = a\,\mathbb{E}[X],\end{array}\right\} \text{(Linearität)}$$

(c) $X \leq Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y]$, (Monotonie)

(d) $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ für $A \subset \Omega$,

(e) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Beweis von Satz 6.2

Beweis

Die Menge $\mathbb{R}^\Omega := \{X : \Omega \to \mathbb{R}\}$ ist bekanntlich ein $\mathbb{R}$-Vektorraum, und L^1 ist ein Unterraum von $\mathbb{R}^\Omega$ (insbesondere gilt $\mathbb{E}[|X+Y|] < \infty$, falls $X, Y \in L^1$). (Warum?)
Sei $\Omega_0 \subset \Omega$ abzählbar mit $\mathbb{P}(\Omega_0) = 1$. Nach den Rechenregeln für absolut konvergente Reihen gilt

$$(a) \qquad \begin{aligned}\mathbb{E}[X+Y] &= \sum_{\omega\in\Omega_0} (X+Y)(\omega)\,\mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega_0} (X(\omega)+Y(\omega))\,\mathbb{P}(\{\omega\}) \\ &= \sum_{\omega\in\Omega_0} X(\omega)\,\mathbb{P}(\{\omega\}) + \sum_{\omega\in\Omega_0} Y(\omega)\,\mathbb{P}(\{\omega\}) = \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

$$(b) \qquad \mathbb{E}[\alpha X] = \sum_{\omega\in\Omega_0} \alpha X(\omega)\,\mathbb{P}(\{\omega\}) = \alpha \sum_{\omega\in\Omega_0} X(\omega)\,\mathbb{P}(\{\omega\}) = \alpha\mathbb{E}[X].$$

(c) Aus $X \le Y$ (d.h., $X(\omega) \le Y(\omega)$ für $\omega \in \Omega$) folgt

$$\mathbb{E}[X] = \sum_{\omega\in\Omega_0} X(\omega)\,\mathbb{P}(\{\omega\}) \le \sum_{\omega\in\Omega_0} Y(\omega)\,\mathbb{P}(\{\omega\}) = \mathbb{E}[Y].$$

(d) Es gilt $\mathbb{E}[\mathbb{1}_A] = \sum_{\omega\in\Omega_0} \mathbb{1}_A(\omega)\,\mathbb{P}(\{\omega\}) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}) = \mathbb{P}(A)$. (e) Δ-Ungl. für absolut konv. Reihen.

Folgerung (Zählvariablen)

Seien $A_1, \dots, A_n \subset \Omega$ Ereignisse und $X := \sum_{j=1}^{n} \mathbb{1}_{A_j}$. Dann gilt

$$\mathbb{E}X = \sum_{j=1}^{n} \mathbb{P}(A_j)$$

und insbesondere wieder

$$\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A), \qquad A \subset \Omega.$$

Anzahl von Rekorden

Beispiel (Rekorde)

$$\begin{array}{cccccccccccc} 5 & 7 & 3 & 2 & 6 & 8 & 13 & 12 & 17 & 14 & 11 & 4\ldots \\ \uparrow & \uparrow & & & & \uparrow & \uparrow & & \uparrow & & & \\ \text{Rekorde} & & & & & & & & & & & \end{array}$$

Modell: Sei $\Omega = \mathrm{Per}_n^n(oW)$ und $\mathbb{P}$ die Gleichverteilung auf Ω. Für $j = 1, \ldots, n$ betrachten wir das Ereignis

$$A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j = \max\{a_1, \ldots, a_j\}\} \quad \text{(Rekord an Stelle } j\text{)}.$$

Dann beschreibt $X_n := \sum_{j=1}^{n} \mathbb{1}_{A_j}$ die Anzahl der Rekorde.

Es gilt $\mathbb{P}(A_j) = \frac{|A_j|}{|\Omega|} = \frac{1}{n!} \cdot \binom{n}{j} \cdot (j-1)!(n-j)! = \frac{1}{j}$. (intuitiv?)

$\Rightarrow \mathbb{E}[X_n] = \sum_{j=1}^{n} \frac{1}{j}, \quad \mathbb{E}(X_{32}) \approx 4,06, \quad \mathbb{E}(X_{10^9}) \approx 20,3$

Transformationssatz

Satz 6.3

Sei $(\Omega, \mathbb{P})$ ein diskreter W.raum und X eine reelle Zufallsvariable auf Ω. Sei $g : \mathbb{R} \to \mathbb{R}$ eine Abbildung. Dann ist $g(X) = g \circ X$ eine reelle ZV. Der Erwartungswert $\mathbb{E}[g(X)]$ existiert genau dann, wenn

$$\sum_{t \in \mathbb{R} \,:\, \mathbb{P}(X=t)>0} |g(t)| \, \mathbb{P}(X = t) < \infty.$$

In diesem Fall gilt

$$\mathbb{E}[g(X)] = \sum_{t \in \mathbb{R} \,:\, \mathbb{P}(X=t)>0} g(t) \, \mathbb{P}(X = t).$$

Speziell für $g(t) := t$, $t \in \mathbb{R}$ erhält man

$$\mathbb{E}[X] = \sum_{t \in \mathbb{R} \,:\, \mathbb{P}(X=t)>0} t \cdot \mathbb{P}(X = t).$$

Beachte:

- Der Grundraum Ω tritt in den Hintergrund.
- Additivität des Erwartungswertes so nicht offensichtlich.

Beweis von Satz 6.3

Beweis

Die Bedingung $\mathbb{P}(X = t) > 0$ sichert, dass es nur abzählbare viele Summanden gibt: Die Mengen $\{X = t\}$ sind paarweise disjunkt für verschiedene $t \in \mathbb{R}$, also gibt es nur endlich viele $t \in \mathbb{R}$ mit $\mathbb{P}(X = t) \geq \frac{1}{n}$, $n \in \mathbb{N}$.

Existenz des Erwartungswertes:

$$\begin{aligned}
\mathbb{E}\,|g(X)| = \sum_{\omega \in \Omega_0} |g(X(\omega))|\,\mathbb{P}(\{\omega\}) &= \sum_{\substack{\omega \in \Omega_0, X(\omega)=t, \\ t \in \mathbb{R}\,:\,\mathbb{P}(X=t)>0}} |g(X(\omega))|\,\mathbb{P}(\{\omega\}) \\
&= \sum_{\substack{t \in \mathbb{R} \\ \mathbb{P}(X=t)>0}} \sum_{\substack{\omega \in \Omega_0 \\ X(\omega)=t}} |g(t)|\,\mathbb{P}(\{\omega\}) \\
&= \sum_{\substack{t \in \mathbb{R} \\ \mathbb{P}(X=t)>0}} |g(t)| \sum_{\substack{\omega \in \Omega_0 \\ X(\omega)=t}} \mathbb{P}(\{\omega\}) = \sum_{t \in \mathbb{R}} |g(t)|\,\mathbb{P}(X = t).
\end{aligned}$$

Wiederhole das Argument ohne Betragsstriche.

Beispiel zur Transformationsformel

Beispiel

Sei $(\Omega, \mathbb{P})$ ein beliebiger diskreter W.raum und X eine reelle Zufallsvariable mit

$$\mathbb{P}^X(\{j\}) = \mathbb{P}(X = j) = \frac{1}{k}, \ j = 1, \ldots, k.$$

Dann gilt

$$\mathbb{E}X = \sum_{j=1}^{k} j \cdot \mathbb{P}(X = j) = \frac{1}{k} \cdot \sum_{j=1}^{k} j = \frac{1}{k} \cdot \frac{1}{2} k(k+1) = \frac{k+1}{2}.$$

Bemerkungen:

- Ω und $\mathbb{P}$ sind oft nicht relevant, $\mathbb{P}^X$ genügt.
- $\mathbb{E}X$ muss nicht Element der Menge $X(\Omega)$ sein.
- $\mathbb{E}X$ ist als Schwerpunkt interpretierbar:

$$\sum_j (j - \mathbb{E}X) \cdot \underbrace{\mathbb{P}(X = j)}_{\substack{\text{Masse an} \\ \text{Position } j}} = 0.$$

Jordan-Formel

Notation: $[n] = \{1, 2, \ldots, n\}$

Satz 6.4 (Jordan-Formel)

Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum, seien $A_1, \ldots, A_n \subset \Omega$ und sei $X := \sum_{j=1}^{n} \mathbb{1}_{A_j}$. Setze

$$S_j := \sum_{\substack{I \subset [n] \\ |I| = j}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right), \quad j = 0, \ldots, n$$

(also insbesondere $S_0 = 1$). Dann gilt:

$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j, \quad k = 0, \ldots, n. \tag{$\star$}$$

Spezialfälle

$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j, \quad k = 0, \dots, n.$$

- Hängt $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_j})$ nicht von der Indexmenge $I = \{i_1, \dots, i_j\}$ mit $|I| = j$, sondern nur von j ab, so gilt

$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \mathbb{P}(A_1 \cap \ldots \cap A_j).$$

- Einschluss-Ausschluss Formel:

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = \sum_{j=1}^{n} (-1)^{j-1} S_j.$$

- Die alternierenden Summen in $(\star)$ sind ≥ 0 und addieren sich zu 1 (wirklich?!).

Beweis von Satz 6.4

Beweis

Es sei $N := [n]$ und für $r \in \mathbb{N}_0$ sei $\{N\}_r := \{A \subset N : |A| = r\}$. Zunächst gilt

$$\mathbb{1}\{X = k\} = \sum_{T \in \{N\}_k} \left(\prod_{i \in T} \mathbb{1}_{A_i} \right) \cdot \prod_{\ell \in N \setminus T} (1 - \mathbb{1}_{A_\ell}) \quad \text{und} \quad \prod_{\ell \in N \setminus T} (1 - \mathbb{1}_{A_\ell}) = \sum_{r=0}^{n-k} (-1)^r \cdot \sum_{U \in \{N \setminus T\}_r} \prod_{i \in U} \mathbb{1}_{A_i} .$$

Einsetzen ergibt

$$\begin{aligned}
\mathbb{1}\{X = k\} &= \sum_{r=0}^{n-k} (-1)^r \sum_{T \in \{N\}_k} \sum_{U \in \{N \setminus T\}_r} \left(\prod_{i \in T} \mathbb{1}_{A_i} \right) \left(\prod_{i \in U} \mathbb{1}_{A_i} \right) \\
&= \sum_{r=0}^{n-k} (-1)^r \sum_{T \in \{N\}_k} \sum_{U \in \{N \setminus T\}_r} \prod_{i \in T+U} \mathbb{1}_{A_i} \\
&= \sum_{r=0}^{n-k} (-1)^r \sum_{V \in \{N\}_{k+r}} \binom{k+r}{k} \prod_{i \in V} \mathbb{1}_{A_i} \\
&\overset{j=k+r}{=} \sum_{j=k}^{n} (-1)^{j-k} \sum_{V \in \{N\}_j} \binom{j}{k} \mathbb{1}_{\bigcap_{i \in V} A_i} .
\end{aligned}$$

Anwenden von $\mathbb{E}$ (und Ausnutzung von $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$ und der Linearität des EW) ergibt die Behauptung.

Freie Fächer

Beispiel

Es werden $s \geq 1$ Teilchen (unterscheidbar) rein zufällig auf $n \geq 1$ Fächer verteilt, hierbei sind Mehrfachbelegungen zugelassen.

Gesucht: Verteilung der Anzahl freier Fächer

Modell:

$$\Omega = \text{Per}_s^n(mW) = [n]^s, \mathbb{P} = \text{Gleichverteilung auf } \Omega.$$

$$A_i := \{(a_1, \ldots, a_s) \in \Omega : a_l \neq i \text{ für } l = 1, \ldots, s\} \quad \text{“Fach } i \text{ ist leer”.}$$

$$X = \sum_{i=1}^{n} \mathbb{1}_{A_i} \to \text{Anzahl der freien Fächer.}$$

Für $j = 0, \ldots, n$ und $I \subset \{1, \ldots, n\}$ mit $|I| = j$ gilt

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \left(\frac{n-j}{n}\right)^s \quad \text{(“}j\text{ Fächer sind nicht belegt”).}$$

Freie Fächer

Beispiel (fortgesetzt)

Mit der Formel von Jordan folgt

$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \left(\frac{n-j}{n} \right)^{s}, \quad k = 0, \ldots, n.$$

Die Zufallsvariable $Y := n - X$ bezeichnet die Anzahl der besetzten Fächer. Es gilt

$$\mathbb{P}(Y = m) = \mathbb{P}(X = n - m)$$

und

$$\begin{aligned} \mathbb{E}[Y] = n - \mathbb{E}[X] &= n - \sum_{j=1}^{n} \left(\frac{n-1}{n} \right)^{s} \\ &= n - n \cdot \left(\frac{n-1}{n} \right)^{s} = n \cdot \left[1 - \left(1 - \frac{1}{n} \right)^{s} \right]. \end{aligned}$$

Zweidrittel-Gesetz beim Roulette

Beispiel

Bestimmung der mittleren Anzahl der Zahlen, die beim Roulette nach s Runden aufgetreten sind. Hier ist $n = 37$ wegen $0, 1, \ldots, 36$ als Nummern der verfügbaren Fächer. Also folgt

$$\mathbb{E}[Y] = 37\left(1 - \left(\frac{36}{37}\right)^s\right).$$

Für $s = 37$ etwa erhält man $\mathbb{E}Y \approx 23,58$. Wegen $\frac{2}{3} \cdot 37 \approx 24,6$ besagt dies, dass nach 37 Spielen / Würfen "im Mittel" knapp $\frac{2}{3}$ der Zahlen $0, 1, \ldots, 36$ aufgetreten sind. (Vorsicht bei der Interpretation!)

Das Phänomen ist als $\frac{2}{3}$-Gesetz bei Roulette-Spielern bekannt.

Kapitel 7: Binomial- und hypergeometrische Verteilung

Motivation der Binomialverteilung

Problem: In einer Urne seien r rote und s schwarze Kugeln. Die Kugeln seien unterscheidbar. Wir ziehen n Kugeln nacheinander mit Zurücklegen und beachten die Reihenfolge. Wieviele der gezogenen Kugeln sind rot?

Modellraum: $\Omega = \mathrm{Per}_n^{r+s}(mW)$, $\mathbb{P}$: Gleichverteilung auf Ω.

O.B.d.A. seien die Kugeln so durchnummeriert, dass $1, \dots, r$ die roten und $r+1, \dots, r+s$ die schwarzen Kugeln sind.

$$A_j := \{(a_1, \dots, a_n) \in \Omega : a_j \leq r\} \qquad \text{“}j\text{-te gezogene Kugel ist rot”}$$

$$X := \sum_{j=1}^{n} \mathbb{1}_{A_j} \qquad \text{“Anzahl gezogener roter Kugeln”}$$

Es gilt $|\Omega| = (r+s)^n, |A_j| = r \cdot (r+s)^{n-1}$ und somit $\mathbb{P}(A_j) = \frac{r(r+s)^{n-1}}{(r+s)^n} = \frac{r}{r+s}$. Es folgt

$$\mathbb{P}(X = k) = \frac{\binom{n}{k} \cdot r^k \cdot s^{n-k}}{(r+s)^n} = \binom{n}{k} \left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}, \qquad k \in \{0, \dots, n\}.$$

Binomialverteilung

Definition 7.1 (Binomialverteilung)

Sei $n \in \mathbb{N}_0$, $p \in [0,1]$. Durch

$$\mathrm{p}(k) := \binom{n}{k} p^k \cdot (1-p)^{n-k}, \quad k \in \{0, \dots, n\}$$

wird eine Zähldichte auf $\{0, \dots, n\}$ erklärt. Das zugehörige Wahrscheinlichkeitsmaß auf dem Grundraum $\{0, \dots, n\}$ heißt Binomialverteilung mit den Parametern n und p und wird mit $\mathrm{Bin}(n,p)$ bezeichnet.

Ist $X : \Omega \to \mathbb{R}$ eine ZV auf einem (diskreten) Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ mit

$$\mathbb{P}^X(\{k\}) = \mathrm{p}(k) \quad \text{für } k \in \{0, \dots, n\},$$

so sagt man, dass X eine Binomialverteilung mit den Parametern n, p hat und schreibt hierfür $X \sim \mathrm{Bin}(n,p)$. (Man sagt auch, X ist binomialverteilt.)

Bemerkungen

Man kann bei einer binomialverteilten ZV $X \sim \mathrm{Bin}(n, p)$ den Wertebereich auch als Teilmenge von $\mathbb{R}$ wählen, solange zumindest $\{0, \dots, n\} \subset X(\Omega)$ erfüllt ist, d.h. $X : \Omega \to W \subset \mathbb{R}$ mit $\{0, \dots, n\} \subset W$.

Beachte:

- $\sum\limits_{k=0}^{n} \mathrm{p}(k) = \sum\limits_{k=0}^{n} \binom{n}{k} p^k \cdot (1-p)^{n-k} = (p + (1-p))^n = 1$

  Es handelt sich also bei p in der Tat um eine Zähldichte.
- Für $p = 0$ ist $\mathbb{P}(\{0\}) = 1$ und $\mathbb{P}(\{k\}) = 0$ für $k \in \{1, \dots, n\}$.

  Für $p = 1$ ist $\mathbb{P}(\{n\}) = 1$ und $\mathbb{P}(\{k\}) = 0$ für $k \in \{0, \dots, n-1\}$.

Erwartungswert binomialverteilter ZV

Satz 7.2

Sei $X \sim \mathrm{Bin}(n, p)$. Dann gilt $\mathbb{E}[X] = np$.

Beweis

Man kann dies zum Beispiel einfach nachrechnen. Für $p \in (0, 1]$ gilt

$$\begin{aligned}\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} &= p \cdot \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \cdot (p + (1-p))^{n-1} = np.\end{aligned}$$

Der Fall $p = 0$ ist klar.

Zurück zum Eingangsproblem

Wir hatten

$$X := \sum_{j=1}^{n} \mathbb{1}_{A_j} \qquad \text{“Anzahl gezogener roter Kugeln”}$$

und

$$\mathbb{P}(A_j) = \frac{r}{r+s}.$$

Ferner galt $X \sim \mathrm{Bin}(n, p)$ mit $p = \frac{r}{r+s}$.

In dieser Situation ist also (fast ohne Rechnen!)

$$\mathbb{E}X = \sum_{j=1}^{n} \mathbb{E}\,\mathbb{1}_{A_j} = \sum_{j=1}^{n} \mathbb{P}(A_j) = np.$$

Folgt der allgemeine Fall mittels “Approximation”?

Bemerkung 1

Mit der Notation von Folie 7.1 gilt

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_j}) = \frac{r^j (r+s)^{n-j}}{(r+s)^n} = p^j, \quad 1 \le i_1 < \ldots < i_j \le n.$$

Also erhält man mit der Jordanschen Formel

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} p^j = \sum_{j=k}^{n} (-1)^{j-k} \frac{n!}{k!(j-k)!(n-j)!} p^j \\
&= \sum_{j=k}^{n} (-1)^{j-k} \binom{n}{k} \binom{n-k}{j-k} p^j = \binom{n}{k} \sum_{j=k}^{n} (-1)^{j-k} \binom{n-k}{j-k} p^{j-k+k} \\
&= \binom{n}{k} p^k \sum_{\ell=0}^{n-k} \binom{n-k}{\ell} (-p)^\ell = \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

Bemerkung 2

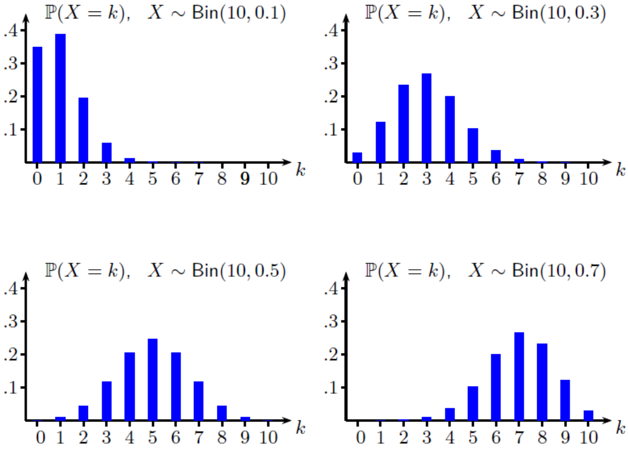
Für $I \subset [n], |I| = k$ gilt

$$\mathbb{P}\left(\bigcap_{i\in I} A_i \cap \bigcap_{j\in I^{\mathrm{c}}} A_j^c\right) = \frac{r^k \cdot s^{n-k}}{(r+s)^n}.$$

Also erhält man durch direkte Rechnung (viel schneller und einfacher als mit der Jordanschen Formel)

$$\begin{aligned}\mathbb{P}(X=k) &= \mathbb{P}\left(\sum_{I\subset[n],|I|=k} \bigcap_{i\in I} A_i \cap \bigcap_{j\in I^{\mathrm{c}}} A_j^c\right) = \sum_{I\subset[n],|I|=k} \mathbb{P}\left(\bigcap_{i\in I} A_i \cap \bigcap_{j\in I^{\mathrm{c}}} A_j^c\right)\\ &= \sum_{|I|=k} p^k(1-p)^{n-k}\\ &= \binom{n}{k} p^k(1-p)^{n-k}.\end{aligned}$$

Plots von $\text{Bin}(n, p)$ **für** $n = 10$ **und** $p = 0.1; 0.3; 0.5; 0.7$



Hypergeometrische Verteilung

Motivation: Aus einer Urne mit r roten und s schwarzen unterscheidbaren Kugel werden erneut rein zufällig n Kugeln gezogen mit Berücksichtigung der Reihenfolge, jetzt aber ohne Zurücklegen. (Dann muss notwendig $n \leq r + s$ gelten!) Wieviele der gezogenen Kugeln sind jetzt rot?

Modellraum: Sei $\Omega = \mathrm{Per}_n^{r+s}(oW)$ und $\mathbb{P}$ die Gleichverteilung auf Ω.

Sei A_j wie oben und $X = \sum\limits_{j=1}^{n} \mathbb{1}_{A_j}$. Wir betrachten das Ereignis $B_k := \{X = k\}$ für $k \in \{0, \dots, n\}$.

Dann gilt

$$\begin{aligned} |B_k| = |\{X = k\}| &= \binom{n}{k} \cdot r(r-1)\cdots(r-k+1) \cdot s(s-1)\cdots(s-(n-k)+1) \\ &= \binom{n}{k} \frac{r!}{(r-k)!} \cdot \frac{s!}{(s-(n-k))!} \end{aligned}$$

und daher

$$\mathbb{P}(X = k) = \binom{n}{k} \frac{r!}{(r-k)!} \cdot \frac{s!}{(s-(n-k))!} \cdot \frac{(r+s-n)!}{(r+s)!} = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}},$$

falls $0 \leq k \leq r$, $0 \leq n-k \leq s$ (sonst sind beide Seiten der Gleichung ohnehin 0).

Hypergeometrische Verteilung

Definition und Behauptung 7.3 (Hypergeometrische Verteilung)

Seien $n \in \mathbb{N}_0$, $r, s \in \mathbb{N}_0$ mit $n \leq r + s$. Dann wird durch

$$\mathrm{p}(k) := \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, \quad k \in \{0, \ldots, n\}$$

eine Zähldichte auf $\{0, \ldots, n\}$ erklärt. Das zugehörige Maß auf dem Grundraum $\{0, \ldots, n\}$ heißt hypergeometrische Verteilung mit den Parametern n, r, s und wird mit $\mathrm{Hyp}(n, r, s)$ bezeichnet.

Ist $X : \Omega \to \mathbb{R}$ eine ZV auf einem (diskreten) Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ mit

$$\mathbb{P}^X(\{k\}) = \mathrm{p}(k) \quad \text{für } k \in \{0, \ldots, n\},$$

so sagt man, dass X eine hypergeometrische Verteilung (mit den Parametern n, r, s) hat und schreibt hierfür $X \sim \mathrm{Hyp}(n, r, s)$. (Man sagt auch, X ist hypergeometrisch verteilt.)

Konvention (Erinnerung): $\binom{a}{b} = 0$ für $b > a$ oder $b < 0$, $a, b \in \mathbb{Z}$.
In der Tat ist p eine Zähldichte, was durch die Vorüberlegungen bewiesen ist. (Es wird im Tutorium nochmals auf unabhängige Weise nachgewiesen.)

Ziehen mit einem Griff

Alternative Modellierung: Das Ereignis $\{X = k\}$ kann auch als Teilmenge der Menge der n-elementigen Teilmengen der $(r+s)$-elementigen Menge aller (unterscheidbaren) Kugeln (also ohne Berücksichtigung der Reihenfolge und ohne Wiederholungen) modelliert werden.

“Ziehen mit einem Griff”, d.h. Modellierung in $\mathrm{Kom}_n^{r+s}(oW)$.

Mit der Gleichverteilung auf diesem Grundraum wird man auch in diesem Fall auf die hypergeometrische Verteilung geführt, und zwar mit denselben Parametern, d.h. auch in diesem Modell gilt

$$\mathbb{P}(X = k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, \quad k \in \{0, \dots, n\}.$$

Erwartungswert hypergeometrisch verteilter ZV

Satz 7.4

Ist $X \sim \mathrm{Hyp}(n, r, s)$, so gilt $\mathbb{E}(X) = n \cdot \frac{r}{r+s}$.

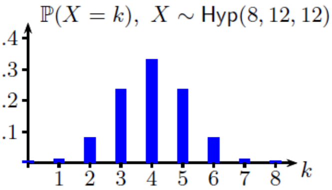
Beweis

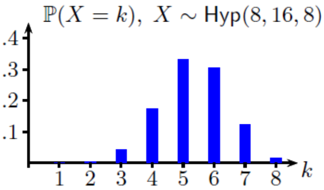
Dies folgt direkt aus $X = \sum_{j=1}^{n} \mathbb{1}_{A_j}$ und

$$\mathbb{P}(A_j) = \frac{|A_j|}{|\Omega|} = \frac{r}{r+s} \quad \text{für } j = 1, \ldots, n. \quad \text{(Warum gilt das?)}$$

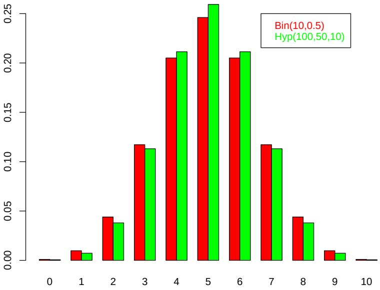
(Man kann den Erwartungswert auch direkt ausrechnen. → Übung, Tutorium!)

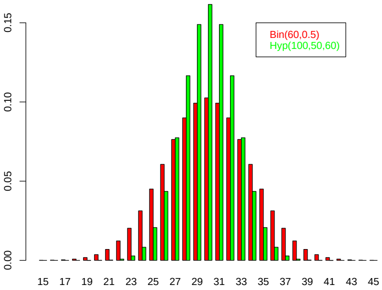
Plots von Hyp$(8, 12, 12)$, Hyp$(8, 16, 8)$





Für $r, s >> n$ ist Hyp$(n, r, s) \approx$ Bin$(n, \frac{r}{r+s})$.





Kapitel 8: Modellierung mehrstufiger Zufallsexperimente

Häufig beruhen Zufallsexperimente auf Teilexperimenten in mehreren Stufen. Ereignisse eines Teilexperiments können von Ausgängen der vorangehenden Experimente abhängen.

Modellierung: $\Omega_1, \ldots, \Omega_n$ seien die Mengen der möglichen Ausgänge des 1. bis n. Teilexperiments. Der Grundraum für das Gesamtexperiment sei

$$\Omega := \Omega_1 \times \cdots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \Omega_i \text{ für } i = 1, \ldots, n\}.$$

Beispiel (Polya-Urnenschema)

In einer Urne seien 1 rote und 3 **schwarze** Kugeln. Es wird rein zufällig eine Kugel gezogen, deren Farbe notiert, und diese wird zusammen mit einer weiteren Kugel gleicher (!) Farbe zurückgelegt. Dann wird erneut rein zufällig eine Kugel gezogen. Wie hoch ist die Wahrscheinlichkeit, dass diese rot ist?

Polya's Urnenschema

Beispiel (Polya-Urnenschema)

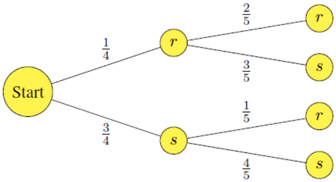
In einer Urne seien 1 rote und 3 **schwarze** Kugeln. Es wird rein zufällig eine Kugel gezogen, deren Farbe notiert, und diese wird zusammen mit einer weiteren Kugel gleicher (!) Farbe zurückgelegt. Dann wird erneut rein zufällig eine Kugel gezogen. Wie hoch ist die Wahrscheinlichkeit, dass diese rot ist?

Modellraum zu Polyas Urnenschema:

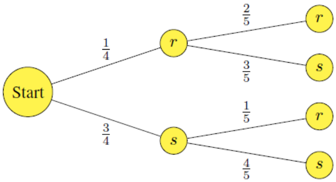
$$\Omega_1 = \Omega_2 = \{r, s\},\ \Omega = \Omega_1 \times \Omega_2,\ B := \{(r,r),(s,r)\}.$$

Festlegung von W.maß / Zähldichte auf Ω mittels "Baumdiagramm":

$$\begin{aligned} \mathrm{p}(r,r) &= \tfrac{1}{4} \cdot \tfrac{2}{5}, \\ \mathrm{p}(r,s) &= \tfrac{1}{4} \cdot \tfrac{3}{5}, \\ \mathrm{p}(s,r) &= \tfrac{3}{4} \cdot \tfrac{1}{5}, \\ \mathrm{p}(s,s) &= \tfrac{3}{4} \cdot \tfrac{4}{5}. \end{aligned}$$



Pfadregeln im Beispiel



1. Pfadregel: Die Übergangswahrscheinlichkeiten entlang eines Pfades werden multipliziert.

$$\mathrm{p}(r,r) = \frac{1}{4} \cdot \frac{2}{5}, \qquad \mathrm{p}(s,r) = \frac{3}{4} \cdot \frac{1}{5}$$

Im Beispiel ist die “erste Pfadregel” konsistent mit relativen Häufigkeiten!

2. Pfadregel: Wahrscheinlichkeiten der Pfade, die zum Ereignis gehören, werden aufsummiert.

$$\mathbb{P}(B) = \mathrm{p}(r,r) + \mathrm{p}(s,r) = \frac{2}{20} + \frac{3}{20} = \frac{1}{4}.$$

Das Beispiel legt nahe, wie aus **Startverteilung** und **Übergangswahrscheinlichkeiten** schließlich W.maße erhalten werden.

Zweistufige Experimente

Definition und Satz 8.1

Seien Ω_1, Ω_2 abzählbare Grundräume, $\mathbb{P}_1$ ein W.maß auf Ω_1 (die Startverteilung) mit Zähldichte $\mathrm{p}_1(\omega_1) := \mathbb{P}_1(\{\omega_1\})$ für $\omega_1 \in \Omega_1$. Sei ferner $\mathrm{p}_{12} : \Omega_1 \times \Omega_2 \to \mathbb{R}$ eine Funktion mit $\mathrm{p}_{12}(\omega_1, \omega_2) \geq 0$ für $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ und

$$\sum_{\omega_2 \in \Omega_2} \mathrm{p}_{12}(\omega_1, \omega_2) = 1, \quad \omega_1 \in \Omega_1.$$

Man nennt p_{12} eine Übergangszähldichte von Ω_1 nach Ω_2. Dann definiert

$$\mathrm{p}(\omega_1, \omega_2) := \mathrm{p}_1(\omega_1) \cdot \mathrm{p}_{12}(\omega_1, \omega_2), \quad (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$$

eine Zähldichte bzw. Wahrscheinlichkeitsfunktion auf $\Omega := \Omega_1 \times \Omega_2$. Das durch p definierte W.maß

$$\mathbb{P}(A) := \sum_{\omega \in A} \mathrm{p}(\omega), \quad A \subset \Omega,$$

heißt Kopplung von p_1 und p_{12}.

Beweis von Satz 8.1

Beweis

Es gilt

$$\begin{aligned}\sum_{\omega\in\Omega} p(\omega) &= \sum_{(\omega_1,\omega_2)\in\Omega_1\times\Omega_2} p_1(\omega_1)\cdot p_{12}(\omega_1,\omega_2)\\ &= \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2} p_1(\omega_1)\cdot p_{12}(\omega_1,\omega_2)\\ &= \sum_{\omega_1\in\Omega_1} p_1(\omega_1)\sum_{\omega_2\in\Omega_2} p_{12}(\omega_1,\omega_2)\\ &= \sum_{\omega_1\in\Omega_1} p_1(\omega_1)\cdot 1 = 1.\end{aligned}$$

Bemerkungen

Sei im obigen Modell etwa $A_2 \subset \Omega_2, A := \Omega_1 \times A_2$. Dann ist

$$\mathbb{P}(A) = \sum_{\omega_2 \in A_2} \left(\sum_{\omega_1 \in \Omega_1} \mathrm{p}_1(\omega_1) \cdot \mathrm{p}_{12}(\omega_1, \omega_2) \right) .$$

Man erhält also $\mathbb{P}(A)$ durch “Summation der W. aller Pfade mit Endknoten in A_2”.
Dies ist eine spezielle Form der zweiten Pfadregel.

Generell: Die Pfadregeln sind eine Folge unserer Definitionen der W. von Elementarereignissen bei der Modellierung als zweistufiges Experiment.

Man interpretiert $\mathrm{p}_{12}(\omega_1, \omega_2)$ als die Wahrscheinlichkeit, dass das 2. Teilexperiment den Ausgang ω_2 hat, wenn das erste Teilexperiment den Ausgang ω_1 hatte.

n-stufige Experimente

Definition und Satz 8.2 (3-stufige Experimente)

Gegeben seien abzählbare Grundräume $\Omega_1, \ldots, \Omega_3$. Ferner seien gegeben eine Startverteilung mit Zähldichte p_1, eine Übergangszähldichte p_{12} sowie eine Funktion $\mathrm{p}_{123} : \Omega_1 \times \Omega_2 \times \Omega_3 \to \mathbb{R}$ mit

$$\mathrm{p}_{123}(\omega_1, \omega_2, \omega_3) \geq 0, \quad (\omega_1, \omega_2, \omega_3) \in \Omega_1 \times \Omega_2 \times \Omega_3,$$

$$\sum_{\omega_3 \in \Omega_3} \mathrm{p}_{123}(\omega_1, \omega_2, \omega_3) = 1, \quad (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2.$$

Dann definiert

$$\mathrm{p}(\omega_1, \omega_2, \omega_3) := \mathrm{p}_1(\omega_1)\mathrm{p}_{12}(\omega_1, \omega_2)\mathrm{p}_{123}(\omega_1, \omega_2, \omega_3), \quad (\omega_1, \omega_2, \omega_3) \in \Omega_1 \times \Omega_2 \times \Omega_3$$

eine Zähldichte auf $\Omega_1 \times \Omega_2 \times \Omega_3$. Das zugehörige W.maß heißt Kopplung von $\mathrm{p}_1, \mathrm{p}_{12}, \mathrm{p}_{123}$.

n-stufige Experimente

Definition und Satz 8.3 (n-stufige Experimente)

Gegeben seien abzählbare Grundräume $\Omega_1, \dots, \Omega_n$.Allgemein gelte für $\mathrm{p}_j : \Omega_1 \times \cdots \times \Omega_j \to \mathbb{R}$ mit $j \in \{2, \dots, n\}$,

$$\mathrm{p}_j(\omega_1, \dots, \omega_j) \geq 0, \quad (\omega_1, \dots, \omega_j) \in \Omega_1 \times \cdots \times \Omega_j,$$

$$\sum_{\omega_j \in \Omega_j} \mathrm{p}_j(\omega_1, \dots, \omega_j) = 1, \quad (\omega_1, \dots, \omega_{j-1}) \in \Omega_1 \times \cdots \times \Omega_{j-1}.$$

Dann definiert

$$\mathrm{p}(\omega_1, \dots, \omega_n) := \mathrm{p}_1(\omega_1) \cdot \mathrm{p}_2(\omega_1, \omega_2) \cdots \mathrm{p}_n(\omega_1, \dots, \omega_n), \quad (\omega_1, \dots, \omega_n) \in \Omega_1 \times \cdots \times \Omega_n$$

eine Zähldichte auf $\Omega_1 \times \cdots \times \Omega_n$. Das zugehörige W.maß heißt Kopplung von $\mathrm{p}_1, \dots, \mathrm{p}_n$. Die Funktionen p_j, $j = 2, \dots, n$ werden als Übergangszähldichten von $\Omega_1 \times \cdots \times \Omega_{j-1}$ nach Ω_j bezeichnet.

Bemerkungen

(1) Die Behauptungen in Definition und Satz 8.2/8.3 können ebenso wie in Definition und Satz 8.1 nachgewiesen werden.

(2) Anstelle von $\mathrm{p}_j(\omega_1, \dots, \omega_j)$ schreibt man manchmal auch suggestiv

$$\mathrm{p}_j(\omega_j \mid \omega_1, \dots, \omega_{j-1})$$

für $j \geq 2$. Diese Notation passt gut zu den bedingten Wahrscheinlichkeiten, die im nächsten Kapitel eingeführt werden. Man interpretiert dabei $\mathrm{p}_j(\omega_j \mid \omega_1, \dots, \omega_{j-1})$ (d.h., die Übergangswahrscheinlichkeit von $(\omega_1, \dots, \omega_{j-1})$ nach ω_j) als die Wahrscheinlichkeit, dass ω_j im j. Teilexperiment eintritt, wenn die vorigen Teilexperimente die Ergebnisse $\omega_1, \dots, \omega_{j-1}$ haben.

Zwei Spezialfälle

(1) **Produktexperimente**: Hängen die Übergangswahrscheinlichkeiten nicht von den Ergebnissen der vorigen Schritte ab, d.h., gilt $\mathrm{p}_j(\omega_1, \dots, \omega_{j-1}, \omega_j) = \bar{\mathrm{p}}_j(\omega_j)$ für $j = 1, \dots, n$, so folgt

$$\mathrm{p}(\omega) = \bar{\mathrm{p}}_1(\omega_1) \cdots \bar{\mathrm{p}}_n(\omega_n), \quad \omega = (\omega_1, \dots, \omega_n) \in \Omega.$$

Gilt speziell $\Omega_i = \Omega_1$, $i = 1, \dots, n$ und $\bar{\mathrm{p}}_j = \bar{\mathrm{p}}$, so wird hiermit die n-fache “unabhängige” Wiederholung eines Experiments modelliert.

(2) **Markov-Kette:** Es sei $S := \Omega_1 = \cdots = \Omega_n$ und $\mathrm{p}_j(\omega_1, \dots, \omega_{j-1}, \omega_j)$ hänge nur von ω_{j-1} ab, nicht von $\omega_1, \dots, \omega_{j-2}$ für alle $j = 2, 3, \dots$.

Sei $X_j(\omega) := \omega_j$ für $\omega = (\omega_1, \dots, \omega_n) \in S^n$. Dann ist $X_1, \dots, X_n$ eine (sogenannte) Markov-Kette mit Zustandsraum S.

Kapitel 9: Bedingte Wahrscheinlichkeiten

Leitfrage: Wie lässt sich Vorinformation bei der Modellierung / Festlegung eines W.raumes angemessen verwenden? Ein typisches Beispiel hierzu ist das (sogenannte) Ziegen-Problem.

Konkret: Wie beeinflusst die Kenntnis, dass ein Ereignis B schon eingetreten ist, die Aussichten auf das Eintreten eines Ereignisses A?

$$\frac{H_n(A\cap B)}{H_n(B)} = \frac{\frac{1}{n}H_n(A\cap B)}{\frac{1}{n}H_n(B)} \begin{matrix}\rightsquigarrow\\ \rightsquigarrow\end{matrix} \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

Anteil derjenigen Fälle unter allen Fällen, in denen B eintritt, in denen auch noch A eintritt.

Bedingte Wahrscheinlichkeit

Definition 9.1 (Bedingte Wahrscheinlichkeit, bedingte Verteilung)

Seien $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum, $A, B \subset \Omega$ Ereignisse mit $\mathbb{P}(B) > 0$. Dann heißt

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

die bedingte Wahrscheinlichkeit von A unter der Bedingung B (bzw. Wahrscheinlichkeit von A gegeben B). Die Funktion

$$\mathbb{P}_B : \mathcal{P}(\Omega) \to [0,1],\ A \mapsto \mathbb{P}_B(A) := \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

heißt bedingte Verteilung von $\mathbb{P}$ unter der Bedingung B (bzw. gegeben B).

Bemerkung. Anschaulich wird die auf B vorhandene “Wahrscheinlichkeitsmasse” auf 1 normiert und die Masse außerhalb von B wird “vergessen”. Es gilt

$$\mathbb{P}(B \mid B) = 1,\ \text{falls } \mathbb{P}(B) > 0,\ \text{sowie } \mathbb{P}(\mathrm{A} \mid \mathrm{B}) = 0 \text{ für } \mathrm{A} \subset \mathrm{B}^{\mathrm{c}}.$$

Bedingte Verteilung ist Wahrscheinlichkeitsmaß

Satz 9.2

Sei die Situation wie in Definition 9.1 gegeben. Dann ist $\mathbb{P}_B$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega))$.

Beweis

Für beliebige $A \subset \Omega$ gilt offenbar $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1$ (sowie ≥ 0).

Außerdem gilt $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = 1$.

Seien $A_1, A_2, \ldots$ paarweise disjunkt. Dann sind auch die Mengen $A_1 \cap B, A_2 \cap B, \ldots$ paarweise disjunkt und aus der σ-Additivität von $\mathbb{P}$ folgt

$$\mathbb{P}_B\left(\sum_{i=1}^{\infty} A_i\right) = \frac{\mathbb{P}\left(\left(\sum_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\sum_{i=1}^{\infty}(A_i \cap B)\right)}{\mathbb{P}(B)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}\left(A_i \cap B\right)}{\mathbb{P}(B)} = \sum_{i=1}^{\infty} \mathbb{P}_B(A_i).$$

Dies zeigt die σ-Additivität von $\mathbb{P}_B$.

Übergangswahrscheinlichkeiten sind bedingte W.

Es liege die Situation von Definition 8.1 vor, also

$$\Omega := \Omega_1 \times \Omega_2,\ \omega = (\omega_1, \omega_2),\ \mathrm{p}(\omega) = \mathrm{p}_1(\omega_1)\mathrm{p}_{12}(\omega_1, \omega_2), \mathbb{P}(C) = \sum_{\omega \in C} \mathrm{p}(\omega), \quad C \subset \Omega.$$

Ziel: Interpretation der Übergangszähldichte p_{12} als bedingte W.

Seien $a_1 \in \Omega_1$, $a_2 \in \Omega_2$, $A_1 := \{a_1\} \times \Omega_2$, $A_2 = \Omega_1 \times \{a_2\}$. Also: $A_1 \cap A_2 = \{(a_1, a_2)\}$ und $\mathbb{P}(A_1 \cap A_2) = \mathrm{p}_1(a_1)\mathrm{p}_{12}(a_1, a_2)$. Ist $\mathrm{p}_1(a_1) > 0$, so folgt

$$\mathbb{P}(A_1) = \sum_{\omega_2 \in \Omega_2} \mathrm{p}_1(a_1)\mathrm{p}_{12}(a_1, \omega_2) = \mathrm{p}_1(a_1)$$

und damit

$$\mathbb{P}(A_2 \mid A_1) = \frac{\mathrm{p}_1(a_1)\mathrm{p}_{12}(a_1, a_2)}{\mathrm{p}_1(a_1)} = \mathrm{p}_{12}(a_1, a_2).$$

Allgemeiner gelten entsprechende Aussagen bei n-stufigen Experimenten.

Multiplikationsformel

Sei $\mathbb{P}(A_1) > 0$. Dann gilt

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1)$$

Allgemeinere Aussage:

Satz 9.3 (Multiplikationsformel)

Sei $(\Omega, \mathbb{P})$ ein diskreter W.raum. Seien $A_1, \ldots, A_n \subset \Omega$ Ereignisse mit $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt

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

Beweis

Für $n = 2$: klar. Nach Definition gilt für $n \geq 3$:

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) \cdot \mathbb{P}(A_n \mid A_1 \cap \ldots \cap A_{n-1}).$$

Die Aussage ergibt sich nun durch vollständige Induktion über n.

Totale Wahrscheinlichkeit und Bayes-Formel

Satz 9.4 (Totale Wahrscheinlichkeit und Bayes-Formel)

Sei $(\Omega, \mathbb{P})$ ein diskreter WRaum. Seien $A_1, A_2, \ldots \subset \Omega$ paarweise disjunkte Ereignisse mit $\sum_{j=1}^{\infty} A_j = \Omega$ sowie B ein Ereignis. Nachfolgend sei $\mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j) := 0$, falls $\mathbb{P}(A_j) = 0$. Dann gilt:

(a)

$$\mathbb{P}(B) \; = \; \sum_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j) \qquad \text{(Formel von der totalen Wahrscheinlichkeit)}$$

(b) Falls $\mathbb{P}(B) > 0$, so gilt für jedes $k \in \mathbb{N}$

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k) \cdot \mathbb{P}(B|A_k)}{\sum\limits_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j)}. \qquad \text{(Formel von Bayes)}$$

Beweis von Satz 9.4

Beweis

(a) folgt aus

$$\mathbb{P}(B) = \mathbb{P}(\Omega \cap B) = \mathbb{P}\left(\left(\sum_{j=1}^{\infty} A_j\right) \cap B\right) = \mathbb{P}\left(\sum_{j=1}^{\infty}(A_j \cap B)\right)$$

$$= \sum_{j=1}^{\infty} \mathbb{P}(A_j \cap B) = \sum_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j),$$

wobei die getroffene Konvention zu beachten ist.

(b) folgt aus

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_k) \cdot \mathbb{P}(B \mid A_k)}{\mathbb{P}(B)}$$

und Anwendung von (a) im Nenner unter Beachtung der Konvention.

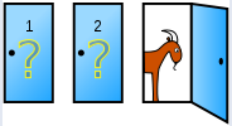
Beispiel: Ziegenproblem

Beispiel

Hinter einer von drei Türen befindet sich der Hauptgewinn, hinter den beiden anderen jeweils eine Ziege. Der Kandidat zeigt auf Tür 1; diese bleibt zunächst verschlossen.

Der Moderator weiß, hinter welcher Tür sich der Hauptgewinn befindet. Er darf die Gewinntür nicht öffnen, muss aber eine Ziege zu erkennen geben.

Der Moderator öffnet Tür 3 und bietet an, von Tür 1 zu Tür 2 zu wechseln. ▶ Soll man das tun?



Beispiel: Ziegenproblem

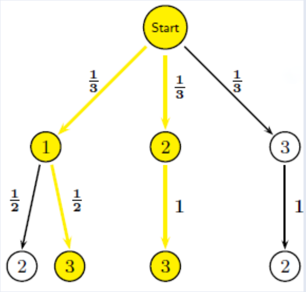
Beispiel

Wir modellieren:

$$A_j = \text{„Gewinn hinter Tür } j\text{.“} \quad \rightsquigarrow \quad \mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3}$$

$$B_j = \text{„Moderator öffnet Tür } j\text{.“} \quad \rightsquigarrow \quad \mathbb{P}(B_2|A_1) = \mathbb{P}(B_3|A_1) = \frac{1}{2}$$

$$\mathbb{P}(B_3|A_2) = \mathbb{P}(B_2|A_3) = 1.$$



- $$\mathbb{P}(A_1 \cap B_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(B_3|A_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

$$\mathbb{P}(A_2|B_3) = \frac{\mathbb{P}(A_2)\mathbb{P}(B_3|A_2)}{\mathbb{P}(A_1)\mathbb{P}(B_3|A_1) + \mathbb{P}(A_2)\mathbb{P}(B_3|A_2)} = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

Beispiel: Test auf eine seltene Krankheit

Beispiel (Test auf eine seltene Krankheit)

Für eine Krankheit, die mit (evtl. unbekannter) Häufigkeit $q \in [0, 1]$ in einer Population auftritt, steht ein Test zur Verfügung.

Modellierung: Setze $\Omega := \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Dabei bedeutet eine
- 1 (bzw. 0) in der ersten Komponente: krank (bzw. gesund),
- 1 (bzw. 0) in der zweiten Komponente: positives (bzw. negatives) Testergebnis.

Also ist

$$\begin{aligned} K &:= \{(1, 0), (1, 1)\} && (\text{krank}) \\ K^c &:= \{(0, 1), (0, 0)\} && (\text{gesund}) \\ N &:= \{(1, 0), (0, 0)\} && (\text{negatives Testergebnis}) \\ N^c &:= \{(0, 1), (1, 1)\} && (\text{positives Testergebnis}) \end{aligned}$$

Beispiel: Test auf eine seltene Krankheit

Es gilt dann:

$$
\begin{aligned}
\mathbb{P}(K) &= q, && \text{(Prävalenz)}\\
\mathbb{P}(N^c|K) &= p_{se}, && \text{(Sensitivität)}\\
\mathbb{P}(N|K^c) &= p_{sp}, && \text{(Spezifität)}\\
\mathbb{P}(K|N^c) & && \text{(positiver Vorhersagewert)}\\
\mathbb{P}(K^c|N) & && \text{(negativer Vorhersagewert)}
\end{aligned}
$$

- Die **Sensitivität** und die **Spezifität** sind Gütekriterien für das medizinische Testverfahren und würden idealerweise 100 % betragen, was praktisch aber nicht erreicht werden kann.
- Die **Prävalenz** ist die Häufigkeit mit der die Krankheit innerhalb der betrachteten Population auftritt (oft unbekannt). Sie beeinflusst stark die Vorhersagewerte des Tests. Sie kann ggf. sehr klein sein.

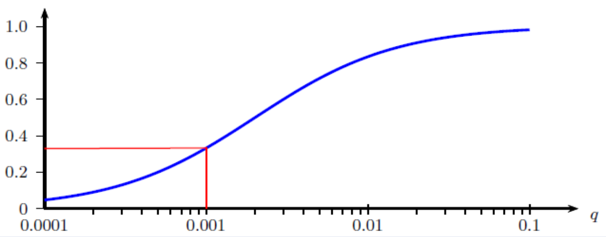
Beispiel: Test auf eine seltene Krankheit

Die Bayes-Formel ergibt für den positiven Vorhersagewert

$$\mathbb{P}(K|N^c) \;=\; \frac{\mathbb{P}(K)\cdot\mathbb{P}(N^c|K)}{\mathbb{P}(K)\cdot\mathbb{P}(N^c|K)\;+\;\mathbb{P}(K^c)\cdot\mathbb{P}(N^c|K^c)} = \frac{q\cdot p_{se}}{q\cdot p_{se}\;+\;(1-q)\cdot(1-p_{sp})}\;.$$

wegen $\mathbb{P}(K^c) = 1-q$ und $\mathbb{P}(N^c|K^c) = 1-p_{sp}$.

Plot von $q \mapsto \mathbb{P}(K|N^c)$ für $p_{se} = p_{sp} = 0.998$ (beachte die logarithmische Skala auf der q-Achse):



Beispiel: Test auf eine seltene Krankheit

konkrete Zahlen:

- 1 000 000 Menschen werden getestet, davon sind 1000 krank ($q = 0.001$).
- Von den 1000 Kranken erhalten ca. 998 ein positives Ergebnis (wegen hoher Sensitivität, $p_{se} = 0.998$).
- Von den 999 000 Gesunden erhalten ca. 1998 ein positives Ergebnis (trotz hoher Spezifität, $p_{sp} = 0.998$).
- Insgesamt gibt es also ca. 2996 Personen mit positivem Testergebnis.
- Davon ist nur ca. ein Drittel krank ($\mathbb{P}(K|N^c)$).
- generelles Problem von Reihenuntersuchungen!

Beispiel: Simpson-Paradoxon

Beispiel

Bei den 6 Fächern mit den höchsten Bewerberzahlen wurden 1973 an der Universität Berkeley ca. $44,5\%$ der männlichen, aber nur etwa $30,3\%$ der weiblichen Bewerber zugelassen.

	Männer		Frauen	
Fach	Anzahl der Bewerber	Zulassungsquote (in %)	Anzahl der Bewerberinnen	Zulassungsquote (in %)
1	825	62	108	82
2	560	63	25	68
3	325	37	593	34
4	417	33	375	35
5	191	28	393	24
6	373	6	341	7
insg.	2691	**44,5**	1835	**30,3**

- Ist die geringere Annahmequote bei Frauen ein Zeichen für Geschlechterdiskriminierung? Wir setzen natürlich voraus, dass bei beiden Geschlechtern der Anteil Qualifizierter gleich hoch ist.
- Nein, die Frauen haben sich bevorzugt in Fächern mit hoher Ablehnungsquote beworben.

Beispiel: Simpson-Paradoxon

Was ist der Zusammenhang zur bedingten Wahrscheinlichkeit?

Wir modellieren der Einfachheit halber die Situation für Frauen und Männer getrennt (aber analog). Hier sei das Modell für eine zufällig unter allen Bewerberinnen ausgewählte Frau vorgestellt.

Setze $\Omega = \{1, \dots, 6\} \times \{1, 0\}$. Für $(f, z) \in \Omega$ gibt f das gewählte Fach an, $z = 1$ bedeutet, dass die Bewerberin zugelassen, und $z = 0$, dass sie abgelehnt wurde.
Dann ist

$$\begin{aligned} F_i &:= \{(i,1),(i,0)\} \hat{=} \text{„sie hat sich auf Fach } i \text{ beworben“}, 1 \leq i \leq 6, \\ Z &:= \{(i,1) \mid 1 \leq i \leq 6\} \hat{=} \text{„sie wurde zugelassen“}. \end{aligned}$$

Das W.maß $\mathbb{P}_w$, das die Situation der Frauen beschreibt, ist vollständig bestimmt durch die Angaben

$$\begin{aligned} \mathbb{P}_w(F_i) &= \text{Anteil der Frauen, die sich auf Fach } i \text{ beworben haben}, \\ \mathbb{P}_w(Z|F_i) &= \text{Annahmequote für Frauen im Fach } i. \end{aligned}$$

(Warum?)

Beispiel: Simpson-Paradoxon

$$\begin{aligned} F_i &:= \{(i,1),(i,0)\} \hat{=} \text{„hat sich auf Fach } i \text{ beworben“}, 1 \le i \le 6,\\ Z &:= \{(i,1) \mid 1 \le i \le 6\} \hat{=} \text{„wurde zugelassen“},\\ \mathbb{P}_w(F_i) &= \text{Anteil der Frauen, die sich auf Fach } i \text{ beworben haben},\\ \mathbb{P}_w(Z|F_i) &= \text{Annahmequote für Frauen im Fach } i. \end{aligned}$$

Sei $(\Omega, \mathbb{P}_m)$ das analoge Modell für die Männer.
▶ Die Zulassungswahrscheinlichkeit beträgt (Satz von der totalen W.) für Frauen/Männer:

$$\mathbb{P}_w(Z) = \sum_{i=1}^{6} \mathbb{P}_w(Z|F_i) \cdot \mathbb{P}_w(F_i) \quad \text{bzw.} \quad \mathbb{P}_m(Z) = \sum_{i=1}^{6} \mathbb{P}_m(Z|F_i) \cdot \mathbb{P}_m(F_i).$$

▶ $\mathbb{P}_w(Z) < \mathbb{P}_m(Z)$ kann auftreten, obwohl $\mathbb{P}_w(Z|F_i) \ge \mathbb{P}_m(Z|F_i)$ für die meisten (oder sogar alle!) i gilt, wenn nur die kleinen bedingten W. $\mathbb{P}_w(Z|F_i)$ mit großen Gewichten $\mathbb{P}_w(F_i)$ in die Summe eingehen.

Beispiel: Simpson-Paradoxon

▶ Das „Paradoxon“ besteht also nur in einer unzulässigen Gleichsetzung von bedingten und „normalen/unbedingten“ Wahrscheinlichkeiten!

▶ Alle relevanten Einflussfaktoren (hier: die Fächerwahl) müssen berücksichtigt werden, wenn man den Einfluss eines Merkmals (hier: das Geschlecht) auf eine Zielgröße (hier: die Zulassungsquote) untersuchen möchte, es sei denn, man kann sicherstellen, dass diese weiteren Einflussfaktoren für alle Ausprägungen des eigentlich interessierenden Merkmals gleich sind (im vorliegenden Fall also $\mathbb{P}_w(F_i) = \mathbb{P}_m(F_i)$ gelten würde).

Anwendung in medizinischen Studien: Beim Vergleich der Wirksamkeit eines neuen Medikaments mit der eines anderen Medikaments oder eines Placebos werden die teilnehmenden Patienten zufällig auf die beiden Gruppen (also diejenigen, die das neue Medikament erhalten, und die Kontrollgruppe derjenigen, die das andere Medikament oder das Placebo erhalten) aufgeteilt („randomized trial“).

Kapitel 10: Stochastische Unabhängigkeit

Ergebnisse von je 25 unabhängigen Würfelwürfen?

2 5 3 5 4 1 2 6 3 6 5 3 1 4 2 3 5 4 1 4 2 6 4 1 3

4 3 3 4 4 6 1 2 3 4 5 4 5 6 3 3 4 1 3 6 2 6 3 6 5

3 6 4 5 1 2 3 6 4 5 3 2 3 4 6 4 2 3 5 6 2 1 4 6 5

2 2 6 2 3 3 6 3 6 2 6 4 4 1 4 4 5 5 3 3 3 5 1 5 3

Weitere Beispielsituationen

- Stein-Schere-Papier (Kann ich unabhängig wechseln?)
- Roulette (Werden Zahlen bei längerem Ausbleiben wahrscheinlicher?)
- Lotto (Sind manche Kombinationen wahrscheinlicher als andere?)

Unabhängige Ereignisse

Die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$ von A unter der Bedingung B ist

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

Die Kenntnis des Eintretens des Ereignisses B führt oft zu Wahrscheinlichkeiten $\mathbb{P}(A|B)$, die verschieden von der „unbedingten" Wahrscheinlichkeit $\mathbb{P}(A)$ sind.

Falls jedoch gilt

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

so hat das Eintreten von B **wahrscheinlichkeitstheoretisch** keinen Einfluss auf das Eintreten von A. Hierdurch wird keine Aussage über kausale Abhängigkeiten getroffen.

Abhängig oder unabhängig?

Beispiel: Zweimaliges Ziehen aus Urne ohne Zurücklegen

Wir ziehen zweimal aus einer Urne mit r roten und s schwarzen Kugeln ohne Zurücklegen und betrachten die Ereignisse

$$A := \text{„zweite Kugel rot“}, \qquad B := \text{„erste Kugel rot“}.$$

Dann gilt

$$\mathbb{P}(A) = \frac{r(r+s-1)}{(r+s)(r+s-1)} = \frac{r}{r+s} = \mathbb{P}(B),$$

$$\mathbb{P}(A \cap B) = \frac{r(r-1)}{(r+s)(r+s-1)} = \mathbb{P}(B)\frac{r-1}{r+s-1}$$

und daher

$$\mathbb{P}(A|B) = \frac{r-1}{r+s-1} \neq \mathbb{P}(A), \qquad \mathbb{P}(B|A) \neq \mathbb{P}(B).$$

Unabhängige Ereignisse

In gleicher Weise bedeutet

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

dass die Wahrscheinlichkeit des Eintretens von B „unabhängig“ ist von der Information „A tritt ein“.

Jede der Gleichungen $\mathbb{P}(A|B) = \mathbb{P}(A)$ und $\mathbb{P}(B|A) = \mathbb{P}(B)$ ist äquivalent zu

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

In diesem Fall heißen die Ereignisse A und B (stochastisch) unabhängig.

Dabei sind auch die Fälle $\mathbb{P}(A) = 0$ oder $\mathbb{P}(B) = 0$ zugelassen.

Diskussion: Die Unabhängigkeit von A und B im Fall $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$ bedeutet anschaulich, dass A und B wahrscheinlichkeitstheoretisch in dem Sinne keinerlei Einfluss aufeinander ausüben, dass jede der beiden Informationen „A tritt ein“ oder „B tritt ein“ die Aussicht auf das Eintreten des jeweils anderen Ereignisses unverändert lässt.

Stochastische Unabhängigkeit

Definition 10.1 (Stoch. Unabhängigkeit von endlich vielen Ereignissen)

Sei $(\Omega, \mathbb{P})$ ein diskreter W.raum. Ereignisse $A_1, \ldots, A_n \subset \Omega$ heißen (stochastisch) unabhängig, falls

$$\mathbb{P}\left(\bigcap_{j \in T} A_j\right) = \prod_{j \in T} \mathbb{P}(A_j)$$

für jede Menge $T \subset \{1, 2, \ldots, n\}$ (mit $|T| \geq 2$) gilt.

Für drei Ereignisse A, B und C ist die Unabhängigkeit gleichbedeutend mit der Gültigkeit der **vier** Gleichungen

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B) \cdot \mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C). \end{aligned}$$

Es sind dann auch A und B stochastisch unabhängig, ebenso A und C usw.

Bemerkungen zur stoch. Unabhängigkeit

- A, B, C paarweise unabhängig $\not\Longrightarrow$ A, B, C unabhängig (Übungsaufgabe).
- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ $\not\Longrightarrow$ A, B, C unabhängig (Übungsaufgabe).
- Unabhängigkeit hat nichts mit Disjunktheit zu tun!
- Unabhängigkeit ist von realer Beeinflussung zu unterscheiden!

Beispiel:

In einer Urne seien 1 rote und 2 schwarze Kugeln. Wir ziehen zwei Kugeln ohne Zurücklegen. Seien

$$B := \text{„erste Kugel rot“}, \text{ und } A := \text{„zweite Kugel rot“}$$

Dann gilt $\mathbb{P}(A) = 1/3$ und $\mathbb{P}(A|B) = 0$. A und B sind also nicht unabhängig. A ist real beeinflusst von B, aber nicht B von A !

Komplemente

Mit A und B sind auch A und B^c unabhängig, da

$$\begin{aligned}
\mathbb{P}(A \cap B^c) &= \mathbb{P}(A \setminus (A \cap B)) \\
&= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) \\
&= \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A) \cdot \mathbb{P}(B^c).
\end{aligned}$$

Damit sind auch A^c und B^c unabhängig!

Mit vollständiger Induktion erhält man den folgenden Satz.

Komplemente

Satz 10.2 (Unabhängigkeit und Komplementbildung)

Es seien $A_1, \ldots, A_n$, $n \geq 2$, Ereignisse in einem diskreten W.raum $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Dann sind folgende Aussagen äquivalent:

(a) $A_1, \ldots, A_n$ sind stochastisch unabhängig.

(b)

$$\mathbb{P}\left(\bigcap_{i\in I} A_i \cap \bigcap_{j\in J} A_j^c\right) = \prod_{i\in I} \mathbb{P}(A_i) \cdot \prod_{j\in J} \mathbb{P}\left(A_j^c\right)$$

für jede Wahl disjunkter Teilmengen I und J aus $\{1, 2, \ldots, n\}$.

Dabei gilt

$$\bigcap_{i\in\emptyset} A_i := \bigcap_{j\in\emptyset} A_j^c := \Omega, \qquad \prod_{i\in\emptyset} \mathbb{P}(A_i) := \prod_{j\in\emptyset} \mathbb{P}\left(A_j^c\right) := 1.$$

Beachte: (b) $\Longrightarrow$ (a) folgt mit $J = \emptyset$.

(a) $\Longrightarrow$ (b) folgt durch Induktion über $|J|$, analog zum Beispiel.

Produktexperimente

Beispiel: Unabhängigkeit und Produktexperimente

Seien $(\Omega_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathbb{P}_n)$ diskrete W.räume, $\Omega := \Omega_1 \times \cdots \times \Omega_n$,

$$\mathbb{P}(\{\omega\}) \;=\; \mathrm{p}(\omega) \;:=\; \prod_{j=1}^{n} \mathrm{p}_j(a_j) \;:=\; \prod_{j=1}^{n} \mathbb{P}_j(\{a_j\}), \qquad \omega = (a_1, \ldots, a_n) \in \Omega.$$

Dann ist $(\Omega, \mathbb{P})$ ein diskreter W.raum und p die Zähldichte von $\mathbb{P}$.

Für $j \in \{1, \ldots, n\}$ sei $A_j^* \subset \Omega_j$ beliebig und

$$A_j \;:=\; \bigtimes_{m=1}^{j-1} \Omega_m \times A_j^* \times \bigtimes_{m=j+1}^{n} \Omega_m \;=\; \{\omega = (a_1, \ldots, a_n) \in \Omega : a_j \in A_j^*\}.$$

Dann sind $A_1, \ldots, A_n$ stochastisch unabhängig.

Produktexperimente

Beispiel: Unabhängigkeit und Produktexperimente

Nachweis: Sind $C_1 \subset \Omega_1,\ C_2 \subset \Omega_2,\ \dots,\ C_n \subset \Omega_n$ und ist $\omega = (a_1, \dots, a_n)$, so gilt

$$\mathbb{P}(C_1 \times \cdots \times C_n) = \sum_{\omega \in C_1 \times \cdots \times C_n} \mathrm{p}(\omega) \ = \sum_{a_1 \in C_1} \cdots \sum_{a_n \in C_n} \prod_{j=1}^{n} \mathrm{p}_j(a_j) = \Big(\sum_{a_1 \in C_1} \mathrm{p}_1(a_1) \Big) \cdots \Big(\sum_{a_n \in C_n} \mathrm{p}_n(a_n) \Big)$$
$$= \mathbb{P}_1(C_1) \cdots \mathbb{P}_n(C_n). \qquad (*)$$

Ferner: Für $\emptyset \neq T \subset \{1, \dots, n\}$ ist

$$\bigcap_{j \in T} A_j \ = \ \bigtimes_{j=1}^{n} B_j^*,$$

wobei $B_j^* := A_j^*$ für $j \in T$ und $B_j^* := \Omega_j$ für $j \notin T$. Es folgt

$$\mathbb{P}\Big(\bigcap_{j \in T} A_j \Big) \ = \mathbb{P}\Big(\bigtimes_{j=1}^{n} B_j^* \Big) \overset{(*)}{=} \prod_{j=1}^{n} \mathbb{P}_j(B_j^*) \ = \prod_{j \in T} \mathbb{P}_j(A_j^*) \overset{(*)}{=} \prod_{j \in T} \mathbb{P}(A_j).$$

Erzeugungsweise der Binomialverteilung

Satz 10.3

Es seien $A_1, \dots, A_n$ unabhängige Ereignisse in einem W.raum $(\Omega, \mathbb{P})$ mit $\mathbb{P}(A_j) =: p$ für $j = 1, \dots, n$. Dann gilt

$$X := \sum_{j=1}^{n} \mathbf{1}\{A_j\} \sim \mathrm{Bin}(n, p).$$

Beweis von Satz 10.3

Beweis

Sei $N := [n]$, $\{N\}_k := \{T \subset N : |T| = k\}$. Für $k \in [n]$ gilt $|\{N\}_k| = \binom{n}{k}$ und

$$\begin{aligned}\mathbb{P}(\{X = k\}) &= \mathbb{P}\left(\sum_{T \in \{N\}_k} \left(\bigcap_{i \in T} A_i \cap \bigcap_{j \in N \setminus T} A_j^c\right)\right) = \sum_{T \in \{N\}_k} \mathbb{P}\left(\bigcap_{i \in T} A_i \cap \bigcap_{j \in N \setminus T} A_j^c\right) \\ &= \sum_{T \in \{N\}_k} \prod_{i \in T} \mathbb{P}(A_i) \cdot \prod_{j \in N \setminus T} \mathbb{P}(A_j^c) = \sum_{T \in \{N\}_k} p^k \cdot (1-p)^{n-k} \\ &= \binom{n}{k} p^k \cdot (1-p)^{n-k}.\end{aligned}$$

Beispiel: Bernoulli-Ketten

Beispiel

Die Situation von Satz 10.3 wird hergestellt durch die Wahl (vgl. Folien 10.9/10.10):

- $\Omega \ := \ \Omega_1 \times \cdots \times \Omega_n$,
- $\Omega_j \ := \ \{0,1\}$ für $j = 1, \ldots, n$,
- $\mathbb{P}_j(\{1\}) := p = 1 - \mathbb{P}_j(\{0\})$ für $j = 1, \ldots, n$,
- für $\omega = (a_1, \ldots, a_n) \in \Omega$ sei

$$\mathbb{P}(\{(a_1, \ldots, a_n)\}) := \prod_{i=1}^{n} \mathbb{P}_i(\{a_i\}) = \prod_{i=1}^{n} \left[p^{a_i}(1-p)^{1-a_i}\right] = p^{a_1+\cdots+a_n}(1-p)^{n-a_1-\cdots-a_n},$$

- $A_j \ := \ \{(a_1, \ldots, a_n) \in \Omega : a_j = 1\}$.

Der W.raum $(\Omega, \mathbb{P})$ (und das damit einhergehende Experiment) heißt Bernoulli-Kette der Länge n mit Trefferwahrscheinlichkeit p.

Dabei steht 1 für Treffer und 0 wird als Niete interpretiert.

Beispiel: Das Zwei-Finger-Morra

Beispiel: Zwei-Finger-Morra

Zwei Spieler A und B heben gleichzeitig jeweils einen oder zwei Finger hoch.

Stimmen die Anzahlen der gezeigten Finger überein, so erhält A von B so viele Euro, wie insgesamt Finger gezeigt wurden (also 2 oder 4).

Stimmen sie nicht überein, so zahlt A 3 Euro an B.

Annahmen: A hebt mit W. a einen Finger und mit W. $1 - a$ zwei Finger.
B hebt mit W. b einen Finger und mit W. $1 - b$ zwei Finger.
A und B treffen ihre Wahl unabhängig voneinander.

Modell: $\Omega := \{(1,1),(1,2),(2,1),(2,2)\}$ und

$$p(1,1) = ab, \quad p(1,2) = a(1-b), \quad p(2,1) = (1-a)b, \quad p(2,2) = (1-a)(1-b).$$

Sei X der Spielgewinn von Spieler A. Es gilt:

$$\begin{aligned} \mathbb{P}(X=2) &= ab, \\ \mathbb{P}(X=-3) &= a(1-b) + (1-a)b, \\ \mathbb{P}(X=4) &= (1-a)(1-b). \end{aligned}$$

Beispiel: Das Zwei-Finger-Morra

Erwartungswert von X ist von Spielstrategien a und b abhängig!

$$\mathbb{E}_{a,b}(X) = 2ab - 3[a(1-b) + (1-a)b] + 4(1-a)(1-b) = 4 - 7b + (12b-7)a.$$

Ist das Spiel fair? Falls $a = b = 1/2$ gilt, so folgt $\mathbb{E}_{a,b}(X) = 0$.

Wählt Spieler B die Strategie $b_0 := 7/12$, so folgt

$$\mathbb{E}_{a,b_0}(X) = 4 - \frac{49}{12} + 0 = -\frac{1}{12} \qquad \text{unabhängig von } a!$$

Kann B vielleicht noch etwas besser agieren? Beachte:

$$\max_{0 \le a \le 1} \mathbb{E}_{a,b}(X) = \begin{cases} 5b - 3, & \text{falls} \quad b > 7/12, \\ 4 - 7b, & \text{falls} \quad b < 7/12, \\ -\frac{1}{12}, & \text{falls} \quad b = 7/12, \end{cases}$$

$$\Longrightarrow \min_{0 \le b \le 1} \max_{0 \le a \le 1} \mathbb{E}_{a,b}(X) = \max_{0 \le a \le 1} \mathbb{E}_{a,b_0}(X) = -\frac{1}{12}.$$

Spieler B ist im Vorteil!

Unabhängig oder nicht?

Beispiel: Unabhängigkeit und Gerichts-(Fehl)-Urteile

Sally Clark verliert zwei Kinder durch plötzlichen Kindstod. Nach dem Tod des zweiten Kindes wird sie wegen zweifachen Mordes verurteilt.

Sei A_j das Ereignis, dass in einer wohlhabenden Nichtraucherfamilie das j-te Kind durch plötzlichen Kindstod stirbt.

$$\mathbb{P}(A_j) \approx \frac{1}{8500}. \qquad \text{(aufgrund empirischer Daten)}$$

Urteil stützte sich maßgeblich auf die Annahme, A_1 und A_2 seien unabhängig.

$$\Longrightarrow \quad \mathbb{P}(A_1 \cap A_2) \approx \frac{1}{8500} \cdot \frac{1}{8500} \approx \frac{1}{72000000}.$$

Jury interpretierte diese W. zudem fälschlicherweise als W. für die Unschuld der Mutter.
Die Royal Statistical Society schaltete sich ein. Keine Unabhängigkeit!

$$\mathbb{P}(A_2 | A_1) \gg \mathbb{P}(A_1)!$$

Sally Clark wurde im zweiten Berufungsverfahren freigesprochen. (https://en.wikipedia.org/wiki/Sally_Clark)

Kapitel 11: Zufallsvektoren, gemeinsame Verteilung

Beispiel (Zweifacher Würfelwurf, erste und größte Augenzahl)

Es seien $\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\}$, $\mathbb{P}$ die Gleichverteilung auf Ω, $X(i,j) := i$ sowie $Y(i,j) := \max(i,j)$. Bestimme $\mathbb{P}(\{X = k\} \cap \{Y = \ell\})$.

ℓ

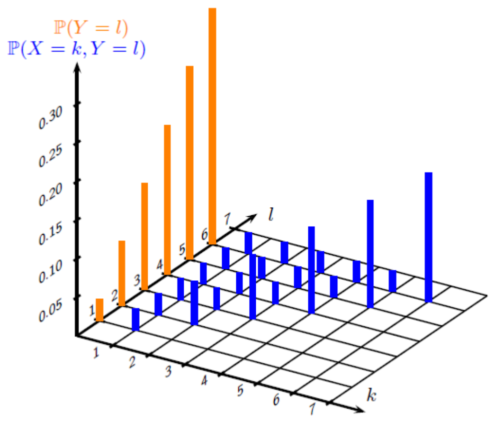
k

	1	2	3	4	5	6	Σ
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	0	2/36	1/36	1/36	1/36	1/36	1/6
3	0	0	3/36	1/36	1/36	1/36	1/6
4	0	0	0	4/36	1/36	1/36	1/6
5	0	0	0	0	5/36	1/36	1/6
6	0	0	0	0	0	6/36	1/6
Σ	1/36	3/36	5/36	7/36	9/36	11/36	1

$\mathbb{P}(X = k)$

$\mathbb{P}(Y = \ell)$

Erste und größte Augenzahl beim zweifachen Würfelwurf



Stabdiagramm der gemeinsamen Verteilung von erster und größter Augenzahl beim zweifachen Würfelwurf (blau) sowie der Verteilung des Maximums (orange)

Gemeinsame Verteilung

Definition und Behauptung 11.1 (Zufallsvektor, gemeinsame Verteilung)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W.raum und $X_j : \Omega \to \mathbb{R}$, $j = 1, \ldots, n$, Zufallsvariablen. Dann heißt die durch

$$X(\omega) \ := \ (X_1(\omega), \ldots, X_n(\omega)), \quad \omega \in \Omega,$$

definierte Abbildung $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ (n-dimensionaler) Zufallsvektor mit Komponenten $X_1, \ldots, X_n$.

Das durch

$$\mathbb{P}^X(M) \ := \ \mathbb{P}\left(X^{-1}(M)\right), \qquad M \subset \mathbb{R}^n,$$

definierte Wahrscheinlichkeitsmaß $\mathbb{P}^X : \mathcal{P}(\mathbb{R}^n) \to [0, 1]$ heißt Verteilung von X oder gemeinsame Verteilung von $X_1, \ldots, X_n$. Desweiteren heißt die durch

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

definierte Funktion $f_{X_1,\ldots,X_n}$ gemeinsame Zähldichte von $X_1, \ldots, X_n$.
Die Verteilung von X_j heißt j-te Marginalverteilung (oder Randverteilung) von X.

Bem.: Zu zeigen ist nur, dass $\mathbb{P}^X$ ein W.maß ist. Das Argument ist dasselbe wie im reellwertigen Fall.

Erklärungen

- Kurze Notation: $\mathbb{P}(X \in M)$ für $\mathbb{P}(\{X \in M\})$, $\mathbb{P}(X_1 \in M_1, X_2 \in M_2)$ für $\mathbb{P}(\{X_1 \in M_1\} \cap \{X_2 \in M_2\})$ etc.
- Da der W.raum $(\Omega, \mathbb{P})$ diskret ist, gibt es ein $\Omega_0 \subset \Omega$ mit $\mathbb{P}(\Omega_0) = 1$. Deshalb ist für $X : \Omega \to \mathbb{R}^n$ die Menge $M_0 := X(\Omega_0) \subset \mathbb{R}^n$ abzählbar.
- Ferner gilt $\mathbb{P}(X \in M_0) = 1$ und

$$0 \leq \mathbb{P}(X^{-1}(M_0^c)) \leq \mathbb{P}(\Omega \setminus \Omega_0) = 1 - 1 = 0.$$

- Man kann auch $M_0^* := \{t \in M_0 : \mathbb{P}(X = t) > 0\}$ betrachten. Dann ist M_0^* abzählbar,

$$\mathbb{P}(X \in M_0 \setminus M_0^*) = \sum_{t \in M_0 \setminus M_0^*} \mathbb{P}(X = t) = 0,$$

und daher $\mathbb{P}(X \in M_0^*) = 1$ und $\mathbb{P}(X \in (M_0^*)^c) = 0$.

Gemeinsame Verteilung vs. Randverteilungen

Beobachtung

In der Situation von Definition 11.1 lassen sich aus $\mathbb{P}^X$ die Marginalverteilungen von $X_1, \ldots, X_n$ bestimmen.

Zur Erinnerung: Sind $B_1, B_2, \ldots \subset \Omega$ paarweise disjunkt mit $\mathbb{P}\left(\sum_{i\geq 1} B_i\right) = 1$, so gilt

$$\mathbb{P}(A) \;=\; \sum_{i\geq 1} \mathbb{P}(A \cap B_i), \quad A \subset \Omega.$$

Hier: Zu jedem $j \in \{1, \ldots, n\}$ existiert eine abzählbare Menge $M_j \subset \mathbb{R}$ mit $\mathbb{P}(X_j \in M_j) = 1$. Dann gilt für $n = 2$

$$\mathbb{P}(X_1 = t_1) \;=\; \sum_{t_2 \in M_2} \mathbb{P}(X_1 = t_1, X_2 = t_2), \qquad t_1 \in M_1$$

und allgemein:

$$\mathbb{P}(X_1 = t_1) \;=\; \sum_{t_2 \in M_2} \cdots \sum_{t_n \in M_n} \mathbb{P}(X_1 = t_1, X_2 = t_2, \ldots, X_n = t_n).$$

(Summe kann weiter eingeschränkt werden auf $\{(t_2, \ldots, t_n) \in M_2 \times \ldots \times M_n : \mathbb{P}(X_2 = t_2, \ldots, X_n = t_n) > 0\}$)

Gemeinsame Verteilung vs. Randverteilungen

Die gemeinsame Verteilung legt die Randverteilungen fest.

Aber: Die gemeinsame Verteilung ist i.A. **nicht** durch die Marginalverteilungen bestimmt!

Beispiel

Seien $\Omega := \{(1,1),(1,2),(2,1),(2,2)\}$, $X(i,j) := i$, $Y(i,j) := j$, $c \in [0,1/2]$ ein freier Parameter und die gemeinsame Verteilung von X und Y sei gegeben durch die folgende Tabelle:

j

i

	1	2	Σ
1	$\mathbf{c}$	$\frac{1}{2} - \mathbf{c}$	$\frac{1}{2}$
2	$\frac{1}{2} - \mathbf{c}$	$\mathbf{c}$	$\frac{1}{2}$
Σ	$\frac{1}{2}$	$\frac{1}{2}$	1

$\mathbb{P}(X = i)$

$\mathbb{P}(Y = j)$

Verschiedene gemeinsame Verteilungen besitzen die gleichen Marginalverteilungen!

Unabhängigkeit von Zufallsvariablen

Definition 11.2 (Unabhängigkeit von Zufallsvariablen)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W.raum und $X_1, \ldots, X_n$ reelle Zufallsvariablen auf Ω. Die ZVen $X_1, \ldots, X_n$ heißen stochastisch unabhängig, falls gilt:

$\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ sind unabhängig für alle $B_1, \ldots, B_n \subset \mathbb{R}$.

Bemerkung: Die ZVen $X_1, \ldots, X_n$ können auch allgemeiner Zufallsvektoren mit Werten in Räumen unterschiedlicher Dimensionen sein, X_j kann etwa $\mathbb{R}^{k_j}$-wertig sein für $j = 1, \ldots, n$. In diesem Fall ist $B_j \subset \mathbb{R}^{k_j}$, $j = 1, \ldots, n$ zu wählen.

Unabhängigkeit von Zufallsvariablen

Satz 11.3 (Kriterien für stoch. Unabhängigkeit von ZVen)

Folgende Aussagen sind äquivalent für Zufallsvariablen $X_1, \dots, X_n$ auf einem W.raum $(\Omega, \mathbb{P})$:

(a) $X_1, \dots, X_n$ sind stochastisch unabhängig,

(b) $\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) \;=\; \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j) \quad$ für $B_1, \dots, B_n \subset \mathbb{R}$,

(c) $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \;=\; \prod_{j=1}^{n} \mathbb{P}(X_j = x_j) \quad$ für $x_1, \dots, x_n \in \mathbb{R}$.

Beweis

„(a) $\Longrightarrow$ (b)“ ist direkte Folge der Definition der Unabhängigkeit von Ereignissen.

„(b) $\Longrightarrow$ (a)“: Setzt man $B_j = \mathbb{R}$, so ist $\{X_j \in B_j\} = \Omega$ und $\mathbb{P}(X_j \in B_j) = 1$.

„(b) $\Longrightarrow$ (c)“: Setze in (b) $B_j = \{x_j\}$ für $j = 1, \dots, n$.

„(c) $\Longrightarrow$ (b)“: Summation! (nächste Folie)

Beweis von Satz 11.3 ((c) $\Longrightarrow$ (b))

Beweis

Sei $M_j \subset \mathbb{R}$ abzählbar mit $\mathbb{P}(X_j \in M_j) = 1$ für $j = 1, \dots, n$.

Seien $B_1, \dots, B_n \subset \mathbb{R}$ und $B_j^* := B_j \cap M_j \subset B_j$, also $\mathbb{P}(X_j \in B_j \setminus B_j^*) = 0$. Dann folgt

$$\begin{aligned}
\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) &= \mathbb{P}(X_1 \in B_1^*, \dots, X_n \in B_n^*) \\
&= \sum_{(x_1,\dots,x_n) \in B_1^* \times \cdots \times B_n^*} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_1 \in B_1^*} \cdots \sum_{x_n \in B_n^*} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) \\
&= \left(\sum_{x_1 \in B_1^*} \mathbb{P}(X_1 = x_1) \right) \cdots \left(\sum_{x_n \in B_n^*} \mathbb{P}(X_n = x_n) \right) \\
&= \mathbb{P}(X_1 \in B_1^*) \cdots \mathbb{P}(X_n \in B_n^*) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).
\end{aligned}$$

Gemeinsame Verteilung unabhängiger Zufallsvariablen

Folgerung 11.4

Die gemeinsame Verteilung von stochastisch unabhängigen ZVen $X_1, \dots, X_n$ auf einem diskreten W.raum $(\Omega, \mathbb{P})$ ist durch ihre Randverteilungen eindeutig bestimmt.

Blockungslemma

Satz 11.5 (Blockungslemma für Zufallsvariablen)

Es seien $X_1, \ldots, X_n$ unabhängige reellwertige Zufallsvariablen auf einem gemeinsamen diskreten W.raum $(\Omega, \mathbb{P})$. Sei $k \in \{1, \ldots, n-1\}$. Weiter seien $g : \mathbb{R}^k \to \mathbb{R}$ und $h : \mathbb{R}^{n-k} \to \mathbb{R}$ Funktionen. Dann sind die Zufallsvariablen $g(X_1, \ldots, X_k)$ und $h(X_{k+1}, \ldots, X_n)$ stochastisch unabhängig.

Die Aussage des Blockungslemmas bleibt für Funktionen von mehr als zwei disjunkten Blöcken unabhängiger Zufallsvariablen gültig.

Beispiele:

- X_1, X_2, X_3, X_4 unabhängig $\implies$ $\sin(X_1 + X_2)$, $X_3 - 2X_4$ unabhängig
- X_1, X_2, X_3, X_4, X_5 unabhängig $\implies$ $X_1^2 + \sqrt{X_3 + X_5}$, X_2/X_4 unabhängig

Frage: Sind $X_1 + X_2$ und $X_1 - X_2$ stochastisch unabhängig, wenn X_1, X_2 stochastisch unabhängig sind?

Beweis von Satz 11.5

Beweis

Seien $Y_1 := g(X_1, \ldots, X_k)$, $Y_2 := h(X_{k+1}, \ldots, X_n)$. Seien $y_1, y_2 \in \mathbb{R}$. Wir wählen abzählbare Mengen M_i wie im Beweis von Satz 11.3. Nachfolgend ist stets $x_i \in M_i$ für $i = 1, \ldots, n$. Dann gilt

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2) = \sum_{\substack{(x_1,\ldots,x_n): \\ g(x_1,\ldots,x_k)=y_1, h(x_{k+1},\ldots,x_n)=y_2}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \sum_{\substack{(x_1,\ldots,x_n): \\ g(x_1,\ldots,x_k)=y_1, h(x_{k+1},\ldots,x_n)=y_2}} \prod_{j=1}^{n} \mathbb{P}(X_j = x_j)$$

$$= \sum_{\substack{(x_1,\ldots,x_k): \\ g(x_1,\ldots,x_k)=y_1}} \sum_{\substack{(x_{k+1},\ldots,x_n): \\ h(x_{k+1},\ldots,x_n)=y_2}} \prod_{j=1}^{k} \mathbb{P}(X_j = x_j) \prod_{j=k+1}^{n} \mathbb{P}(X_j = x_j)$$

$$= \left[\sum_{\substack{(x_1,\ldots,x_k): \\ g(x_1,\ldots,x_k)=y_1}} \prod_{j=1}^{k} \mathbb{P}(X_j = x_j) \right] \left[\sum_{\substack{(x_{k+1},\ldots,x_n): \\ h(x_{k+1},\ldots,x_n)=y_2}} \prod_{j=k+1}^{n} \mathbb{P}(X_j = x_j) \right] = \mathbb{P}(Y_1 = y_1)\mathbb{P}(Y_2 = y_2).$$

Die Behauptung folgt nun mit Hilfe von Satz 11.3.

Erwartungswerte, allg. Transformationsformel

Wir verallgemeinern nun Satz 6.3 auf Zufallsvektoren.

Satz 11.6 (Die allgemeine Transformationsformel)

Sei $Z : \Omega \to \mathbb{R}^n$ ein n-dimensionaler Zufallsvektor auf dem diskreten W.raum $(\Omega, \mathbb{P})$. Sei $g : \mathbb{R}^n \to \mathbb{R}$ eine reelle Funktion. Setze $M_0 := \{z \in \mathbb{R}^n : \mathbb{P}(Z = z) > 0\}$. Genau dann existiert der Erwartungswert von $g(Z)$, wenn $\sum_{z \in M_0} |g(z)| \, \mathbb{P}(Z = z) < \infty$. In diesem Fall gilt

$$\mathbb{E}\, g(Z) \; = \; \sum_{z \in M_0} g(z) \, \mathbb{P}(Z = z).$$

Beweis von Satz 11.6

Zur Erinnerung: $\mathbb{E}X$ existiert, falls $\mathbb{E}|X| < \infty$, d.h., $\sum_{\omega\in\Omega_0} |X(\omega)|\mathbb{P}(\{\omega\}) < \infty$.

Beweis

Sei $\Omega_0 \subset \Omega$ abzählbar mit $\mathbb{P}(\Omega_0) = 1$. Zunächst ist

$$\begin{aligned}
\mathbb{E}|g(Z)| &= \sum_{\omega\in\Omega_0} |g(Z(\omega))|\,\mathbb{P}(\{\omega\}) = \sum_{\omega\in\Omega_0}\sum_{z\in M_0} \mathbf{1}\{Z(\omega) = z\}|g(Z(\omega))|\,\mathbb{P}(\{\omega\}) \\
&= \sum_{z\in M_0}\sum_{\omega\in\Omega_0} |g(z)|\,\mathbf{1}\{Z(\omega) = z\}\,\mathbb{P}(\{\omega\}) = \sum_{z\in M_0} |g(z)| \sum_{\omega\in\Omega_0 : Z(\omega)=z} \mathbb{P}(\{\omega\}) \\
&= \sum_{z\in M_0} |g(z)|\,\mathbb{P}(Z = z).
\end{aligned}$$

Der Erwartungswert $\mathbb{E}(g(Z))$ existiert also (d.h., $\mathbb{E}|g(Z)| < \infty$) genau dann, wenn der letzte Ausdruck endlich ist. Dies zeigt die erste Behauptung.
Wiederholung des Arguments ohne $|\cdot|$ ergibt die zweite Behauptung.

Multiplikationsformel für Erwartungswerte

Satz 11.7

Seien X und Y unabhängige Zufallsvariablen auf einem gemeinsamen diskreten W.raum $(\Omega, \mathbb{P})$ mit $\mathbb{E}|X| < \infty$ und $\mathbb{E}|Y| < \infty$. Dann existiert auch der Erwartungswert des Produktes XY, und es gilt

$$\mathbb{E}\,(XY) \;=\; \mathbb{E}\,X \cdot \mathbb{E}\,Y.$$

Beweis von Satz 11.7

Beweis

Setze $Z := (X, Y)$, $z := (x, y)$ und $g(x, y) := x\,y$. Dann gilt $\mathbb{P}(Z = z) = \mathbb{P}(X = x, Y = y)$. Mit den Sätzen 11.6 und 11.3 folgt

$$\begin{aligned}
\mathbb{E}[|XY|] = \mathbb{E}|g(Z)| &= \sum_{(x,y)\in\mathbb{R}^2 :\, \mathbb{P}(X=x,Y=y)>0} |x|\,|y|\,\mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y)\in\mathbb{R}^2 :\, \mathbb{P}(X=x)\mathbb{P}(Y=y)>0} |x|\,|y|\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y) \\
&= \left(\sum_{x :\, \mathbb{P}(X=x)>0} |x|\,\mathbb{P}(X=x) \right) \left(\sum_{y :\, \mathbb{P}(Y=y)>0} |y|\,\mathbb{P}(Y=y) \right) = \mathbb{E}[|X|] \cdot \mathbb{E}[|Y|] < \infty.
\end{aligned}$$

Wiederholung des Arguments ohne $|\cdot|$ ergibt die zweite Behauptung.

Produkte ohne Unabhängigkeit

Nach Satz 11.7 gilt: Sind X, Y unabhängig mit $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$, so gilt $\mathbb{E}\,|XY| < \infty$.

Frage: Gilt diese Implikation auch, falls X und Y nicht stochastisch unabhängig sind? Antwort: Nein!

Gegenbeispiel

Sei $c := \sum_{k=1}^{\infty} \frac{1}{k^3} < \infty$. Dann definiert

$$p_k \; := \; \frac{1}{c\,k^3}, \qquad k \in \mathbb{N},$$

eine Verteilung auf $\mathbb{N}$.

Sei X eine Zufallsvariable mit $\mathbb{P}(X = k) = p_k$, $k \in \mathbb{N}$. Sei $Y := X$. Dann gilt

$$\mathbb{E}|X| = \mathbb{E}|Y| = \sum_{k=1}^{\infty} k\,p_k \; = \; \frac{1}{c} \sum_{k=1}^{\infty} \frac{1}{k^2} \; < \; \infty,$$

$$\mathbb{E}[|XY|] = \mathbb{E}X^2 = \; = \sum_{k=1}^{\infty} k^2\,p_k \; = \; \frac{1}{c} \sum_{k=1}^{\infty} \frac{1}{k} \; = \; \infty.$$

Diskrete Faltungsformel

Satz 11.8

Es seien X und Y unabhängige reelle Zufallsvariablen auf einem gemeinsamen diskreten W.raum $(\Omega, \mathbb{P})$. Sei $M_0 \subset \mathbb{R}$ eine abzählbare Menge mit $\mathbb{P}(X \in M_0) = 1$. Dann gilt

$$\mathbb{P}(X+Y=t) \;=\; \sum_{x \in M_0} \mathbb{P}(X=x)\,\mathbb{P}(Y=t-x), \qquad t \in \mathbb{R}.$$

Beweis

Sei $M_0 \subset \mathbb{R}$ abzählbar mit $\mathbb{P}(X \in M_0) = 1$. Dann ist $\mathbb{P}(X \in M_0^c) = 0$ und

$$\{X+Y=t\} = \{X \in M_0, Y = t - X\} \;+ \{X \in M_0^c, X+Y=t\}.$$

Es folgt $\mathbb{P}(X+Y=t) = \mathbb{P}(X \in M_0, Y = t-X)$ und daher

$$\begin{aligned}\mathbb{P}(X+Y=t) &= \mathbb{P}\Big(\sum_{x \in M_0} \{X=x, Y=t-x\}\Big) = \sum_{x \in M_0} \mathbb{P}(X=x, Y=t-x) \\ &= \sum_{x \in M_0} \mathbb{P}(X=x)\,\mathbb{P}(Y=t-x).\end{aligned}$$

Faltung von Gleichverteilungen

Beispiel

Seien X, Y stochastisch unabhängig und je gleichverteilt auf $1, 2, \ldots, n$, also

$$\mathbb{P}(X = j) \;=\; \mathbb{P}(Y = j) \;=\; \frac{1}{n}, \qquad j = 1, \ldots, n.$$

Für $k \in \{2, 3, \ldots, 2n\}$ gilt direkt (erste Gleichung) oder mittels Faltungsformel (zweite Gleichung)

$$\begin{aligned}
\mathbb{P}(X + Y = k) \quad &= \quad \frac{1}{n^2} \left|\{(i,j) \in \{1, \ldots, n\}^2 : i + j = k\}\right| \\
&= \quad \sum_{i=1}^{k} \mathbb{P}(X = i)\, \mathbb{P}(Y = k - i) \\
&= \quad \frac{1}{n^2} \cdot \begin{cases} k - 1, & \text{falls } 2 \le k \le n, \\ n - (k - n) + 1, & \text{falls } n + 1 \le k \le 2n, \end{cases} \\
&= \quad \frac{n - |k - n - 1|}{n^2}.
\end{aligned}$$

Für $n = 6$ ergibt sich die Verteilung der Augensumme beim zweifachen Würfelwurf.

Additionsgesetz für die Binomialverteilung

Satz 11.9

Seien X und Y unabhängige Zufallsvariablen mit $X \sim \mathrm{Bin}(m, p)$ und $Y \sim \mathrm{Bin}(n, p)$. Dann gilt

$$X + Y \sim \mathrm{Bin}(m + n, p).$$

Beweis: siehe Tutorium

Kapitel 12: Varianz, Kovarianz und Korrelation

Der Erwartungswert ist ein *Lagemaß* für eine Verteilung bzw. eine ZV. Weitere Information liefern *Streuungsmaße*.

Definition 12.1 (Varianz, Standardabweichung)

Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann heißt die Zahl

$$\mathbb{V}(X) := \mathbb{E}(X - \mathbb{E}X)^2 \qquad\qquad (= \mathbb{E}[(X - \mathbb{E}(X))^2])$$

Varianz (der Verteilung) von X. Die Zahl $\sqrt{\mathbb{V}(X)}$ heißt Standardabweichung (oder auch Streuung) (der Verteilung) von X.

Beachte:

- Es gilt $|x| \leq 1 + x^2$ für alle $x \in \mathbb{R}$, also $|X| \leq 1 + X^2$. Die Voraussetzung impliziert also insbes. $\mathbb{E}|X| < \infty$, d.h. $\mathbb{E}X$ existiert.
- Analog gilt $(X - a)^2 \leq X^2 + 2\,|a|\,|X| + a^2, \quad a \in \mathbb{R}$. Also existiert $\mathbb{V}(X)$.
- Sei $g(t) := (t - \mathbb{E}X)^2,\ t \in \mathbb{R}$. Dann ist $\mathbb{V}(X) = \mathbb{E}(g(X))$.
- Wir setzen nachfolgend (mitunter stillschweigend) $\mathbb{E}X^2 < \infty$ voraus.

Darstellungsformel für die Varianz

Satz 12.2

Sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann gilt:

(a) $\mathbb{V}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

(b) Ist $M_0 \subset \mathbb{R}$ abzählbar mit $\mathbb{P}(X \in M_0) = 1$, so folgt

$$\mathbb{V}(X) = \sum_{x \in M_0} (x - \mathbb{E}X)^2 \cdot \mathbb{P}(X = x) = \sum_{x \in M_0} x^2 \cdot \mathbb{P}(X = x) - (\mathbb{E}X)^2.$$

Beweis

(a) Wegen $(X - \mathbb{E}(X))^2 = X \cdot X - 2\mathbb{E}(X)X + (\mathbb{E}(X))^2$ folgt mit der Additivität des Erwartungswerts

$$\mathbb{V}(X) = \mathbb{E}(X \cdot X) - 2\,\underbrace{\mathbb{E}(\mathbb{E}(X) \cdot X)}_{=\mathbb{E}(X) \cdot \mathbb{E}(X)} + (\mathbb{E}(X))^2.$$

(b) Transformationsformel für den Erwartungswert.

Beispiel: Varianz der Gleichverteilung

Beispiel

Sei $\Omega = \{1, \ldots, k\}$, $k \in \mathbb{N}$, $\mathbb{P}(\{\omega\}) = \frac{1}{k}$, $\omega \in \Omega$, $X = \mathrm{id}_\Omega$. Dann ist

$$\mathbb{E}(X) = \frac{1}{k} \cdot (1 + \cdots + k) = \frac{k+1}{2}$$

und

$$\mathbb{V}(X) = \sum_{j=1}^{k} j^2 \cdot \frac{1}{k} - \left(\frac{k+1}{2}\right)^2 = \frac{1}{6} k(k+1)(2k+1)\frac{1}{k} - \frac{(k+1)^2}{4} = \frac{k^2-1}{12}.$$

Allgemeiner gilt: Ist X gleichverteilt auf $\{a, a+1, \ldots, b\}$ mit $a, b \in \mathbb{N}$ und $a \leq b$, dann gilt

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

Konvention: Wir sagen, ein Ereignis $A \subset \Omega$ gilt $\mathbb{P}$-fast sicher, falls $\mathbb{P}(A) = 1$ oder (in der Stochastik gleichwertig) $\mathbb{P}(A^c) = 0$. **Etwa:** $X = a$ gilt $\mathbb{P}$-fast sicher, falls $\mathbb{P}(X \neq a) = 0$. Letzteres ist konsistent mit Sprechweisen in der allgemeinen Maßtheorie, wo man auch unendliche Maße betrachtet.

Eigenschaften der Varianz

Satz 12.3

Es seien X eine Zufallsvariable auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}X^2 < \infty$ und $A \subset \Omega$ ein Ereignis. Dann gilt:

(a) $\mathbb{V}(X) \geq 0$ und $\mathbb{V}(X) = 0 \iff \mathbb{P}(X = \mathbb{E}X) = 1$, d.h. $X = \mathbb{E}X$ gilt $\mathbb{P}-$f.s.,

(b) $\mathbb{V}(aX + b) = a^2\,\mathbb{V}(X),\ a, b \in \mathbb{R}$,

(c) $\mathbb{V}(X) = \mathbb{E}(X - t)^2 - (\mathbb{E}X - t)^2,\ t \in \mathbb{R}$,

(d) $\mathbb{V}(\mathbb{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A)) = \mathbb{P}(A)\mathbb{P}(A^c)$.

Es seien $X_1, \ldots, X_n$ ZVen auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}X_i^2 < \infty$. Dann gilt:

(e) $\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \left[\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right].$

(f) Sind speziell $X_1, \ldots, X_n$ paarweise stochastisch unabhängig, so gilt

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Beweis von Satz 12.3

Beweis

(c) $\mathbb{E}\left[(X-t)^2\right] - (\mathbb{E}X - t)^2 = \mathbb{V}(X-t) = \mathbb{V}(X)$.

(e) Es gilt

$$\begin{aligned}\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right)^2\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mathbb{E}X_i)^2\right] + \sum_{i \neq j} \underbrace{\mathbb{E}\left[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right]}_{=:E}.\end{aligned}$$

Wegen

$$\begin{aligned}E &= \mathbb{E}[X_i X_j] - \mathbb{E}(X_j)\mathbb{E}(X_i) - \mathbb{E}(X_i)\mathbb{E}(X_j) + \mathbb{E}(X_i)\mathbb{E}(X_j) \\ &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]\end{aligned}$$

folgt die Behauptung aus der Symmetrie.

Minimalitätseigenschaft des Erwartungswerts

Folgerung

Nach Satz 12.3(c) gilt

$$\mathbb{E}(X-t)^2 = \mathbb{V}(X) + (\mathbb{E}X - t)^2, \quad t \in \mathbb{R},$$

und somit

$$\mathbb{V}(X) = \min\left\{\mathbb{E}(X-t)^2 : t \in \mathbb{R}\right\},$$

d.h., $t = \mathbb{E}X$ ist Minimalstelle des Funktionals $t \mapsto \mathbb{E}(X-t)^2$ und $\mathbb{V}(X)$ das zugehörige Minimum.

Varianz von Zählvariablen

Satz 12.4

Es seien $(\Omega, \mathbb{P})$ ein W.raum, $A_1, \ldots, A_n \subset \Omega$ Ereignisse und $X := \sum_{i=1}^{n} \mathbb{1}_{A_i}$. Dann gilt

$$\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i)(1 - \mathbb{P}(A_i)) + 2 \sum_{1 \leq i < j \leq n} (\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)).$$

Gilt speziell $\mathbb{P}(A_j) = \mathbb{P}(A_1)$ und $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_1 \cap A_2)$ für alle $i, j = 1, \ldots, n$ mit $i \neq j$, so folgt

$$\mathbb{V}(X) = n \left[\mathbb{P}(A_1)(1 - \mathbb{P}(A_1)) + (n-1) \left(\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)^2\right)\right].$$

Gilt noch spezieller paarweise stochastische Unabhängigkeit der Ereignisse und ferner $\mathbb{P}(A_i) = p$, dann folgt

$$\mathbb{V}(X) = np(1-p).$$

Beweis: Setze $X_i := \mathbb{1}_{A_i}$, $i = 1, \ldots, n$ und wende Satz 12.3 (e) an.

Varianz von Binomial- und hypergeometrischer Verteilung

Beispiel (Binomialverteilung)

Für $X \sim \mathrm{Bin}(n,p)$, gilt $\mathbb{E}X = np$ und

$$\mathbb{V}(X) = np(1-p).$$

Dies ergibt sich aus Satz 12.4. Setze dafür A_j an als Treffer im j. Versuch, $j = 1, \ldots, n$. Dann gilt $\mathbb{P}(A_i) = p$ und A_i, A_j sind stochastisch unabhängig für $i \neq j$.

Beispiel (Hypergeometrische Verteilung)

Für $X = \sum\limits_{i=1}^{n} \mathbb{1}_{A_i} \sim \mathrm{Hyp}(n,r,s)$ mit $\mathbb{P}(A_i) = \frac{r}{r+s} =: p$ ist $\mathbb{E}X = np$. Aus $\mathbb{P}(A_i \cap A_j) = \frac{r(r-1)}{(r+s)(r+s-1)} = p\frac{r-1}{r+s-1}$ für $i \neq j$ folgt

$$\begin{aligned}\mathbb{V}(X) &= np(1-p) + n(n-1)\left[p\frac{r-1}{r+s-1} - p^2\right] = np(1-p) + np(n-1)\Big[\ \underbrace{\frac{r-1}{r+s-1} - \frac{r}{r+s}}_{-\frac{s}{(r+s)(r+s-1)} = -(1-p)\frac{1}{r+s-1}}\ \Big] \\ &= np(1-p)\left[1 - \frac{(n-1)}{r+s-1}\right].\end{aligned}$$

Anzahl der Rekorde

Beispiel

Sei $\Omega = \mathrm{Per}_n^n(oW)$ und $\mathbb{P}$ die Gleichverteilung auf Ω. Zur Erinnerung:

$$A_j = \{a \in \Omega : a_j = \max\{a_1, \dots, a_j\}\},\ j = 1, \dots, n$$

$$X_n = \sum_{j=1}^{n} \mathbb{1}_{A_j} \qquad \text{Anzahl Rekorde.}$$

Wir hatten schon gezeigt, dass $\mathbb{P}(A_j) = \frac{1}{j}$, $j = 1, \dots, n$ gilt. Für $1 \leq j < k \leq n$ gilt

$$\mathbb{P}(A_j \cap A_k) \overset{(!)}{=} \frac{1}{n!}\binom{n}{k}\binom{k-1}{j}(j-1)!(k-1-j)!(n-k)! = \frac{1}{j \cdot k} = \mathbb{P}(A_j)\mathbb{P}(A_k),$$

das heißt, $A_1, \dots, A_n$ sind paarweise stochastisch unabhängig. Es folgt

$$\mathbb{V}(X_n) = \sum_{j=1}^{n} \mathbb{P}(A_j)(1 - \mathbb{P}(A_j)) = \sum_{j=1}^{n} \frac{1}{j}\left(1 - \frac{1}{j}\right) = \sum_{j=1}^{n} \frac{1}{j} - \sum_{j=1}^{n} \frac{1}{j^2} \ \sim \ \log(n).$$

Übung: $A_1, \dots, A_n$ sind stochastisch unabhängig.

Tschebyscheff-Ungleichung

Satz 12.6 (Tschebyscheff-Ungleichung)

Es sei X eine Zufallsvariable auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}X^2 < \infty$. Für jedes $\varepsilon > 0$ gilt

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

Beweis

Definiere Funktionen $g, h : \mathbb{R} \to [0, \infty)$ durch

$$g(t) := \mathbb{1}\{|t - \mathbb{E}X| \geq \varepsilon\} \leq h(t) := \frac{1}{\varepsilon^2}(t - \mathbb{E}X)^2, \qquad t \in \mathbb{R}.$$

Dann gilt

$$g(X) \leq h(X), \quad \text{also} \quad \mathbb{E}[g(X)] \leq \mathbb{E}[h(X)].$$

Standardisierung von Zufallsvariablen

Definition und Behauptung 12.5

Es sei X eine Zufallsvariable mit $\mathbb{V}(X) > 0$. Dann heißt die Zufallsvariable

$$\widetilde{X} := \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}}$$

Standardisierung von X.

Es gilt $\mathbb{E}\widetilde{X} = 0$ und $\mathbb{V}(\widetilde{X}) = 1$.

Bemerkungen und Beispiele:

- zur Erinnerung: $\sqrt{\mathbb{V}(X)}$ ist die Standardabweichung von X.
- Es gelte $X \sim \mathrm{Bin}(n, p)$. Dann ist $\widetilde{X} = \dfrac{X - np}{\sqrt{np(1-p)}}$ die Standardisierung von X.
- Ist $\widetilde{X}$ die Standardisierung von X, so gilt $\mathbb{P}(|\widetilde{X}| \geq \varepsilon) \leq \varepsilon^{-2}$ für jedes $\varepsilon > 0$.

Kovarianz und (Un-)Korreliertheit

Für beliebige Zufallsvariablen X, Y gilt nach Satz 12.3 (e)

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\underbrace{(\mathbb{E}[X \cdot Y] - \mathbb{E}X \cdot \mathbb{E}Y)}_{=:\mathbb{C}(X,Y)}.$$

Sind X, Y stochastisch unabhängig, so ist nach Satz 11.7 (Multiplikationsformel) $\mathbb{C}(X, Y) = 0$.

Frage: Ist $\mathbb{C}(X, Y)$ ein Maß für stochastische Abhängigkeit?

Definition und Behauptung 12.7

Seien X, Y Zufallsvariablen auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$. Dann heißt

$$\mathbb{C}(X, Y) := \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right] \overset{!}{=} \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

die Kovarianz von X und Y. Man nennt X, Y unkorreliert, wenn $\mathbb{C}(X, Y) = 0$, andernfalls korreliert.

Bem.: Wegen $|u \cdot v| \leq \frac{1}{2}(u^2 + v^2)$ für beliebige $u, v \in \mathbb{R}$ gilt $|X \cdot Y| \leq \frac{1}{2}(X^2 + Y^2)$. Also existiert $\mathbb{E}[XY]$!

Eigenschaften der Kovarianz

Satz 2.8

Sei $(\Omega, \mathbb{P})$ ein diskreter W.raum. Von allen im Folgenden betrachteten Zufallsvariablen sei angenommen, dass die zweiten Momente existieren. Dann gilt:

(a) $\mathbb{C}(X,Y) = \mathbb{C}(Y,X)$ und $\mathbb{C}(X,X) = \mathbb{V}(X)$,

(b) $\mathbb{C}(\alpha X + a, \beta Y + b) = \alpha\beta\,\mathbb{C}(X,Y), \qquad a,b \in \mathbb{R},\ \alpha,\beta \in \mathbb{R}$,

(c) $\mathbb{C}(X + \tilde{X}, Y + \tilde{Y}) = \mathbb{C}(X,Y) + \mathbb{C}(X,\tilde{Y}) + \mathbb{C}(\tilde{X},Y) + \mathbb{C}(\tilde{X},\tilde{Y})$,

(d) $\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{1 \le i < j \le n} \mathbb{C}(X_i, X_j)$,

(e) $\mathbb{C}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j\, \mathbb{C}(X_i, Y_j)$.

Fazit: Die Kovarianz $\mathbb{C}(\cdot,\cdot)$ ist ein bilineares Funktional auf dem Vektorraum

$$\mathrm{L}^2(\Omega,\mathbb{P}) := \{X : \Omega \to \mathbb{R} \quad | \ \mathbb{E}X^2 < \infty\}.$$

Beweis: Nachrechnen!

Unkorreliertheit vs. Unabhängigkeit

Sind X, Y unabhängige ZV, so gilt nach Satz 11.7 (Multiplikationsformel) $\mathbb{C}(X, Y) = 0$.

Aus stochastischer Unabhängigkeit folgt also Unkorreliertheit. **Die Umkehrung gilt im Allgemeinen nicht!**

Beispiel

Seien X, Y (stochastisch unabhängige) Zufallsvariablen auf $(\Omega, \mathbb{P})$ mit $X \sim Y$. Dann folgt

$$\mathbb{C}(X - Y, X + Y) = \mathbb{C}(X, X) + \mathbb{C}(X, Y) - \mathbb{C}(Y, X) - \mathbb{C}(Y, Y) = 0,$$

das heißt, $X - Y, X + Y$ sind unkorreliert.

Speziell: Seien X, Y stochastisch unabhängig und $\mathbb{P}(X = j) = \mathbb{P}(Y = j) = \frac{1}{6}$ für $j = 1, \dots, 6$. Dann gilt

$$\frac{1}{36} = \mathbb{P}(X + Y = 12, X - Y = 0) \neq \underbrace{\mathbb{P}(X + Y = 12)}_{=\frac{1}{36}} \cdot \underbrace{\mathbb{P}(X - Y = 0)}_{=\frac{1}{6}}.$$

Also sind $X - Y, X + Y$ in diesem Beispiel **nicht** stochastisch unabhängig, aber dennoch unkorreliert!

Weitere Beispiele und Darstellungsformel

Beispiel

(a) $\mathbb{C}(2X - 3Y, 3X - 4Y) = \ldots = 6\,\mathbb{V}(X) - 17\,\mathbb{C}(X, Y) + 12\,\mathbb{V}(Y)$.

(b) $\mathbb{C}(aX, -aY + c) = -a^2\,\mathbb{C}(X, Y)$, a, c konstant.

(c) $X_1, \ldots, X_n$ unabhängig $\Longrightarrow \mathbb{C}\left(X_1, \frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\,\mathbb{V}(X_1)$.

Satz 12.9 (Darstellungsformel für die Kovarianz)

Sei $M_0 \subset \mathbb{R}$ abzählbar mit $\sum_{x \in M_0} \mathbb{P}(X = x) = 1 = \sum_{y \in M_0} \mathbb{P}(Y = y)$. Dann gilt

$$\mathbb{C}(X, Y) = \sum_{x,y \in M_0} xy\mathbb{P}(X = x, Y = y) - (\mathbb{E}X)(\mathbb{E}Y).$$

Beweis: Nach Satz 11.6 (Transformationsformel) gilt $\mathbb{E}(XY) = \sum_{x,y \in M_0} xy\mathbb{P}(X = x, Y = y)$.

Beispiel zur Korreliertheit

Beispiel (vgl. Folie 11.6)

Die gemeinsame Verteilung von Zufallsvariablen X, Y sei gegeben durch

j

i

	1	2	Σ
1	c	$\frac{1}{2}-c$	$\frac{1}{2}$
2	$\frac{1}{2}-c$	c	$\frac{1}{2}$
Σ	$\frac{1}{2}$	$\frac{1}{2}$	1

$\mathbb{P}(X=i)$

$\mathbb{P}(Y=j)$

wobei $c \in [0,1/2]$ ein freier Parameter sei. Dann gilt

$$\mathbb{E}X = \mathbb{E}Y = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{2} = \tfrac{3}{2},$$

$$\mathbb{C}(X,Y) = 1 \cdot 1 \cdot c + 1 \cdot 2 \cdot \left(\tfrac{1}{2} - c\right) + 2 \cdot 1 \cdot \left(\tfrac{1}{2} - c\right) + 2 \cdot 2 \cdot c - \left(\tfrac{3}{2}\right)^2 = c - \tfrac{1}{4}.$$

Methode der kleinsten Quadrate, lineare Regression

Seien X, Y Zufallsvariablen auf $(\Omega, \mathbb{P})$.

Ziel: Approximation von Y (auf Basis von X) mittels einer optimalen, affinen Vorhersagefunktion

$$g_{a,b}(t) := a + bt,\ t \in \mathbb{R},\ a, b \in \mathbb{R},$$

wobei a, b fest zu wählen sind (d.h., Y wird durch $\hat{Y} := a + bX$ approximiert).

Die Güte der Approximation wird durch die mittlere quadratische Abweichung $\mathbb{E}[(Y - a - bX)^2]$ gemessen.

Lineare Regression

Satz 12.10

(a) Sei $\mathbb{V}(X) > 0$. Dann hat das Optimierungsproblem

$$f(a,b) := \mathbb{E}[(Y - a - bX)^2] = \min_{a,b\in\mathbb{R}} !$$

die eindeutige Lösung

$$b^* = \frac{\mathbb{C}(X,Y)}{\mathbb{V}(X)}, \;\; a^* = \mathbb{E}(Y) - b^*\mathbb{E}(X)$$

mit Minimalwert

$$M^* := f(a^*, b^*) = \mathbb{V}(Y) - \frac{\mathbb{C}(X,Y)^2}{\mathbb{V}(X)} \geq 0.$$

Ist $M^* = 0$, so ist $Y = a^* + b^* X$ $\mathbb{P}$-fast sicher, d.h., $\mathbb{P}$-fast sicher gilt

$$Y - \mathbb{E}(Y) = \frac{\mathbb{C}(X,Y)}{\mathbb{V}(X)}(X - \mathbb{E}(X)) \quad \text{(und umgekehrt).}$$

Lineare Regression

Satz 12.10

(b) Ist $\mathbb{V}(X)\,\mathbb{V}(Y) > 0$, so ist

$$M^* = \mathbb{V}(Y)\left(1 - \varrho^2(X,Y)\right) \geq 0 \qquad \text{mit } \varrho(X,Y) := \frac{\mathbb{C}(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}.$$

Es gilt $M^* = 0$ genau dann, wenn $\mathbb{P}$-fast sicher $\widetilde{X} = \widetilde{Y}$ gilt (nämlich im Fall $\varrho(X,Y) = 1$) oder wenn $\mathbb{P}$-fast sicher gilt $\widetilde{X} = -\widetilde{Y}$ (nämlich im Fall $\varrho(X,Y) = -1$).

(c) Ist $\mathbb{V}(X) = 0$, so ist $X = \mathbb{E}(X)$ $\mathbb{P}$-fast sicher und $b^* = 0$, $a^* = \mathbb{E}(Y)$ löst das Optimierungsproblem.

Ist $\mathbb{V}(Y) = 0$, so ist $Y = \mathbb{E}(Y)$ $\mathbb{P}$-fast sicher und $a^* = \mathbb{E}(Y)$, $b^* = 0$ löst das Optimierungsproblem.

Beweis von Satz 12.10

Beweis

Man erhält für f mit $a, b \in \mathbb{R}$:

$$f(a,b) = \mathbb{E}(Y^2) + a^2 + b^2\mathbb{E}(X^2) - 2a\mathbb{E}(Y) - 2b\mathbb{E}(XY) + 2ab\mathbb{E}(X).$$

Für die Ableitungen von f gilt

$$f_a(a,b) = 0 \Leftrightarrow \mathbb{E}(Y) = a + b\mathbb{E}(X), \tag{1}$$

$$f_b(a,b) = 0 \Leftrightarrow b\mathbb{E}(X^2) - \mathbb{E}(XY) + a\mathbb{E}(X) = 0. \tag{2}$$

Einsetzen von a aus (1) in (2) liefert b^*, daraus folgt a^*.

Für die Existenz einer (eindeutig bestimmten) Minimalstelle zeigen wir, dass

$$f(a,b) \to \infty \quad \text{für} \quad a^2 + b^2 \to \infty.$$

Beweis von Satz 12.10

Beweis

Hierzu schreiben wir $f(a,b)$ in der Form

$$\begin{aligned} f(a,b) = \mathbb{E}[(Y-a-bX)^2] &= \mathbb{E}\left[((Y-a-b\mathbb{E}(X)) - b(X-\mathbb{E}(X)))^2\right] \\ &= \mathbb{E}\left[(Y-a-b\mathbb{E}(X))^2\right] + b^2\,\mathbb{V}(X) - 2b\mathbb{E}\left[(Y-a-b\mathbb{E}(X))(X-\mathbb{E}(X))\right] \\ &= \mathbb{E}\left[(Y-a-b\mathbb{E}(X))^2\right] + b^2\,\mathbb{V}(X) - 2b\mathbb{E}[Y(X-\mathbb{E}(X))]. \end{aligned}$$

Wegen $\mathbb{V}(X) > 0$ strebt $f(a,b)$ gegen ∞, falls $b \to \infty$, da der quadratische Term $b^2\,\mathbb{V}(X)$ dominiert. Bleibt dagegen b beschränkt, so sieht man an

$$f(a,b) = \mathbb{E}(Y^2) + a^2 + b^2\mathbb{E}(X^2) - 2a\mathbb{E}(Y) - 2b\mathbb{E}(XY) + 2ab\mathbb{E}(X),$$

dass $f(a,b) \to \infty$, falls $a \to \infty$, da dann a^2 dominiert.

Beweis von Satz 12.10

Beweis

Ferner ist (Memo: $b^* = \frac{\mathbb{C}(X,Y)}{\mathbb{V}(X)}$, $a^* = \mathbb{E}(Y) - b^*\mathbb{E}(X)$)

$$\begin{aligned} M^* &= \mathbb{E}[(Y - a^* - b^*X)^2] = \mathbb{E}\left[(Y - \mathbb{E}(Y) + b^*\mathbb{E}(X) - b^*X)^2\right] \\ &= \mathbb{V}(Y) + b^{*2}\,\mathbb{V}(X) - 2b^*\,\mathbb{C}(X,Y) \\ &= \mathbb{V}(Y) - \frac{\mathbb{C}(X,Y)^2}{\mathbb{V}(X)}. \end{aligned}$$

Ist $M^* = 0$ (und $\mathbb{V}(X) > 0$), so ist $\mathbb{E}[(Y - a^* - b^*X)^2] = 0$, das heißt $Y = a^* + b^*X$ $\mathbb{P}$-fast sicher. Dies zeigt $Y - \mathbb{E}(Y) = b^*(X - \mathbb{E}(X))$ $\mathbb{P}$-fast sicher. Dann ist aber $\mathbb{C}(X,Y) = b^*\,\mathbb{V}(X)$. Die Umkehrung sieht man leicht.

Dies zeigt (a) und (b). Aussage (c) ist klar.

Korrelationskoeffizient

Definition 12.11

Für Zufallsvariablen X, Y gelte $\mathbb{V}(X)\,\mathbb{V}(Y) > 0$. Man nennt

$$\varrho(X,Y) := \frac{\mathbb{C}(X,Y)}{\sigma_X \sigma_Y} \qquad \text{mit } \sigma_X = \sqrt{\mathbb{V}(X)}, \sigma_Y = \sqrt{\mathbb{V}(Y)}$$

den (Pearson-)Korrelationskoeffizienten von X und Y. Ferner heißen X und Y

$$\left\{\begin{matrix} \text{positiv} \\ \text{un}- \\ \text{negativ} \end{matrix}\right\} \text{korreliert, falls } \varrho(X,Y) \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 0.$$

Interpretation des Korrelationskoeffizienten

Folgerung 12.12

Für Zufallsvariablen X, Y auf $(\Omega, \mathbb{P})$ mit $\mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$ gelten:

(a)

$$\mathbb{C}(X, Y)^2 \leq \mathbb{V}(X)\,\mathbb{V}(Y).$$

Falls $\mathbb{V}(X) > 0$, so gilt "$=$" genau dann, wenn es $a, b \in \mathbb{R}$ gibt mit $Y = a + bX$ $\mathbb{P}$-fast sicher. Ist $\mathbb{V}(X) = 0$, so ist auch $\mathbb{C}(X, Y) = 0$ und "$=$" gilt für beliebige Y.

(b) Ist $\mathbb{V}(X)\,\mathbb{V}(Y) > 0$, so gilt

$$-1 \leq \varrho(X, Y) \leq 1.$$

Weiterhin gilt

$$\begin{aligned} \varrho(X, Y) = 1 \quad &\Longleftrightarrow \quad \mathbb{C}(X, Y) = \sigma_X \sigma_Y \quad \Longleftrightarrow \quad \widetilde{X} = \widetilde{Y}\ \mathbb{P}\text{-fast sicher} \quad \text{und} \\ \varrho(X, Y) = -1 \quad &\Longleftrightarrow \quad -\mathbb{C}(X, Y) = \sigma_X \sigma_Y \quad \Longleftrightarrow \quad \widetilde{X} = -\widetilde{Y}\ \mathbb{P}\text{-fast sicher.} \end{aligned}$$

Kapitel 13: Deskriptive Statistik

- Ergebnisse von Zufallsexperimenten ergeben Daten.
- Daten werden auf vielfältige Weise erzeugt und festgehalten.
- Wie können diese Daten sinnvoll und strukturiert dargestellt, analysiert und verstanden werden?

Was sind Zufallsexperimente? Wie kommen hierbei Daten zustande? Zufall und Komplexität? Keine strikten Definitionen, aber einige Hinweise:

Wdh.: Ideale Zufallsexperimente

- Experiment wird unter genau festgelegten Bedingungen (*Versuchsbedingungen*) durchgeführt,
- Menge der möglichen Ergebnisse (Grundgesamtheit) ist vor Durchführung des Experimentes bekannt,
- Experiment kann (zumindest prinzipiell) beliebig oft unter gleichen Bedingungen wiederholt werden,
- das Ergebnis eines Experiments ist (in der Regel) nicht (stets) eindeutig festgelegt.

Beispiele (Wdh.)

Beispiele (für Zufallsexperimente)

- das Werfen eines oder mehrerer Würfel oder Münzen,
- die Beobachtung der Anzahl von „Jobs“, die an einem Prozessor innerhalb einer Minute ankommen,
- die Beobachtung von Lebensdauern von Geräten,
- die Beobachtung der Anzahl der Einsen in einer Folge von 32 zufälligen Binärzahlen,
- Zeitpunkte radioaktiven Zerfalls,
- Zeitpunkte des Eintretens von Versicherungsfällen (oder deren Anzahl in einem vorgegebenen Zeitintervall),
- Status eines großen physikalischen Systems (interagierender Partikel),
- Brownsche Bewegung,
- Zustandsbestimmung eines quantenphysikalischen Systems . . .

Merkmale

Die bei einem Zufallsexperiment oder bei vorliegenden Untersuchungseinheiten interessierenden (beobachtbaren) Größen einer Grundgesamtheit heißen Merkmale. Man unterscheidet (nicht trennscharf)

- quantitative Merkmale (in natürlicher Weise zahlenmäßig erfassbar)
  - stetige Merkmale (prinzipiell jeder Wert in einem Intervall möglich)
    Größe, Gewicht und Länge
  - diskrete Merkmale (Ausprägungen sind isolierte Zahlenwerte)
    Anzahlen, Alter in Jahren
- qualitative Merkmale (artmäßig erfassbar)
  - ordinale Merkmale (Ausprägungen weisen Rangfolge auf)
    Zeugnisnoten, Priorität eines Prozesses
  - nominale Merkmale (Klassifizierung nach rein qualitativen Gesichtspunkten)
    Prozessortyp, Geschlecht einer Person

Von Merkmalen angenommene Werte heißen Merkmalsausprägungen oder Merkmalswerte.

Beispiele für Merkmale

Untersuchungseinheit	**Merkmal**	**Ausprägungen**
Baum	Baumart	Eiche, Buche, ...
Baum	Schadstufe	0, 1, 2, 3, 4
Neugeborenes	Größe (in cm)	..., 49.5, 50, 50.5, ...
arbeitslose Person	Schulabschluss	keiner, Sonderschule, Hauptschule, Realschule, Gymnasium, ...
vollzeiterwerbstätige Person	Bruttoeinkommen im Jahr 2014 (in Euro)	..., 39999, 40000, 40001, ...
Betonwürfel	Druckfestigkeit (in 0.1 $\mathrm{N/mm^2}$)	..., 399, 400, 401, ...

Grundgesamtheit und Stichprobe

Grundgesamtheit (Population):

Als Grundgesamtheit (GG) zur Beschreibung einer Datenmenge oder für ein Zufallsexperiment wählt man die Menge aller denkbaren oder relevanten Untersuchungseinheiten oder Ergebnisse (endlich oder unendlich groß, eventuell fiktiv; vgl. Ergebnismenge Ω im W.raum).

Geordnete Stichprobe vom Umfang n:

Zufällig oder deterministisch gewonnene, endliche Folge der Länge $n \in \mathbb{N}$ aus der Grundgesamtheit,

etwa die durch n-fache Wiederholung eines Zufallsexperiments gewonnenen Ergebnisse.

Empirische Häufigkeiten

Sei $x := (x_1, \dots, x_n)$ eine Stichprobe vom Umfang n. Wir betrachten dabei ein festes Merkmal. Jedes x_i gibt genau eine Merkmalsausprägung in Bezug auf dieses Merkmal an (d.h., einen möglichen Wert, den dieses Merkmal annehmen kann).

Definition 13.1

Ist a ein möglicher Merkmalswert in der Stichprobe x, so heißt die **Anzahl** der in der Stichprobe x vorkommenden Stichprobenelemente x_i mit Merkmalsausprägung a, d.h., die Größe

$$H_x(a) \; := \; \sum_{i=1}^{n} \mathbf{1}\{x_i = a\}$$

die absolute Häufigkeit von a in x, und der **Anteil** der in x vorkommenden Stichprobenelemente x_i mit Merkmalsausprägung a, d.h., die Größe

$$h_x(a) \; := \; \frac{H_x(a)}{n} \; = \; \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{x_i = a\}$$

die relative Häufigkeit von a in x.

Diagramme

Empirischen Häufigkeitsverteilungen können als Stabdiagramme oder Kreisdiagramme dargestellt werden.

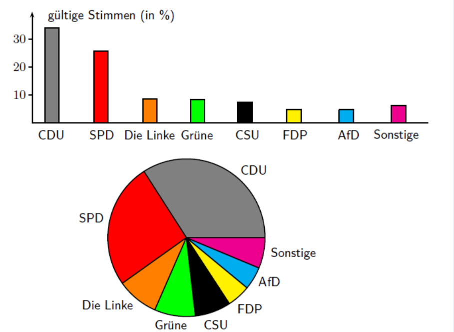
Stabdiagramm: Absolute bzw. relative Häufigkeiten werden als Funktion der Merkmalsausprägungen angezeigt.

Kreisdiagramm: Aufteilung der Kreisfläche in Sektoren, deren Flächen proportional zu den relativen Häufigkeiten der Ausprägungen sind.

Beispiel: Bundestagswahl 2013, Zweitstimmen

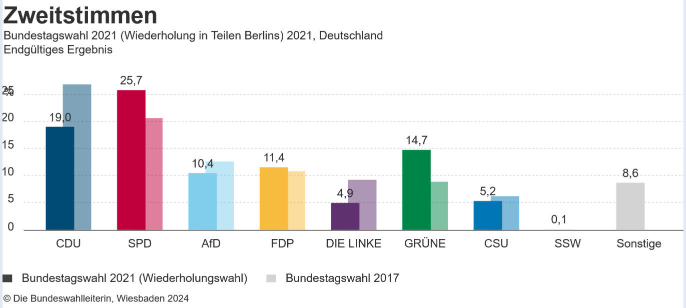
Partei	Zweitstimmen	in Prozent
CDU	14 921 877	34.1
SPD	11 252 215	25.7
Die Linke	3 755 699	8.6
Grüne	3 694 057	8.4
CSU	3 243 569	7.4
FDP	2 083 533	4.8
AfD	2 056 985	4.7
Sonstige	2 718 951	6.2

Untersuchungseinheit Stimmzettel, Merkmal Partei



Vergleich der Bundestagswahlen 2021 und 2017

Beispiel



Zweitstimmen

Bundestagswahl 2021 (Wiederholung in Teilen Berlins) 2021, Deutschland
Endgültiges Ergebnis

© Die Bundeswahlleiterin, Wiesbaden 2024

Empirische Verteilungsfunktion

Definition 13.2

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe. Die zugehörigen **Merkmalswerte** seien **reelle Zahlen**. Die Funktion $F_x \colon \mathbb{R} \to [0,1]$,

$$t \mapsto F_x(t) := \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{1}\{x_i \leq t\}, \; t \in \mathbb{R},$$

heißt empirische Verteilungsfunktion von x.

$F_x(t)$ ist der Anteil der Stichprobenelemente, die kleiner oder gleich t sind.

Alternativ: $$F_x(t) = \sum_{a \leq t} h_x(a), \qquad t \in \mathbb{R}.$$

Beispiel (diskretes Merkmal)

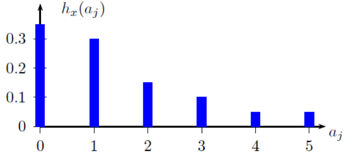
Beispiel

Bei einer Produktion von Werkstücken werden $\mathbf{n} = \mathbf{20}$ Proben zu je 15 Teilen entnommen und jeweils die Anzahl defekter Teile festgestellt. Es sei $x_i \in \{0, \ldots, 15\}$ die Anzahl der defekten Teile in der i-ten Probe, $i = 1, \ldots, 20$.

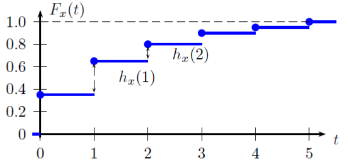
$$x = (\,0 \quad 4 \quad 2 \quad 1 \quad 1 \quad 0 \quad 0 \quad 2 \quad 3 \quad 1 \quad 0 \quad 5 \quad 3 \quad 1 \quad 1 \quad 2 \quad 0 \quad 0 \quad 1 \quad 0\,).$$

Häufigkeitstabelle:

a_j	$H_x(a_j)$	$h_x(a_j)$	$F_x(a_j)$
0	7	0.35	0.35
1	6	0.30	0.65
2	3	0.15	0.80
3	2	0.10	0.90
4	1	0.05	0.95
5	1	0.05	1.00



Stabdiagramm



Empirische Verteilungsfunktion $F_x(t)$

Beispiel (stetiges Merkmal)

Beispiel

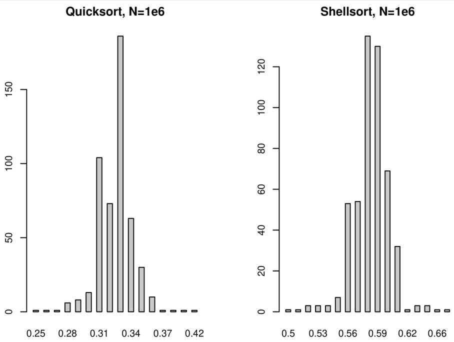
Bei zwei Sortieralgorithmen wurde jeweils $\mathbf{n = 500}$ mal die Laufzeit (in *sec*) zum Sortieren eines Vektors der Länge 10^6 bestimmt, wobei sich folgende Urlisten ergaben:

Quicksort: $x = (\ 0.34 \quad 0.33 \quad 0.38 \quad 0.35 \quad 0.35 \quad 0.32 \quad 0.31 \quad \ldots)$

Kleinster Wert: 0.25 *sec*, größter Wert: 0.42 *sec*, Abstufung: 0.01 *sec*.

Shellsort: $y = (\ 0.58 \quad 0.60 \quad 0.58 \quad 0.58 \quad 0.57 \quad 0.59 \quad 0.58 \quad \ldots)$

Kleinster Wert: 0.50 *sec*, größter Wert: 0.70 *sec*, Abstufung: 0.01 *sec*.



Empirische Häufigkeitsverteilung (**Quicksort**):

$\mathbf{a_j}$	.25	.26	.27	.28	.29	.30	.31	.32	.33	.34
$\mathbf{H_x(a_j)}$	1	1	1	6	8	13	104	73	186	63
$\mathbf{h_x(a_j)}$	.002	.002	.002	.012	.016	.026	.208	.146	.372	.126

$\mathbf{a_j}$	.35	.36	.37	.38	.39	.40	.41	.42
$\mathbf{H_x(a_j)}$	30	10	1	1	1	0	0	1
$\mathbf{h_x(a_j)}$	.060	.020	.002	.002	.002	.000	.000	0.002

Klassenbildung und Histogramme

Klassen sind Intervalle der Form $(a, b] := \{x \in \mathbb{R} : a < x \leq b\}$.

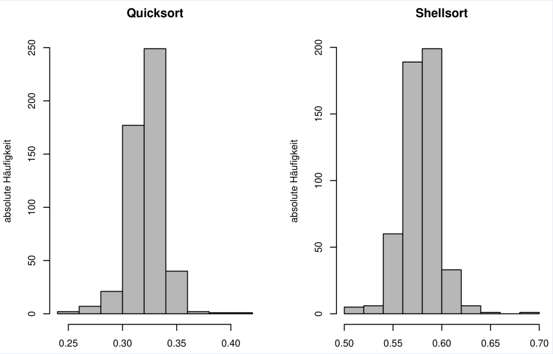
Beispiel (Sortieralgorithmen)

Einteilung der Ergebnismenge in 9 Klassen:

Klasse 1:	$(0.24, 0.26]$
Klasse 2:	$(0.26, 0.28]$
Klasse 3:	$(0.28, 0.30]$
Klasse 4:	$(0.30, 0.32]$
…	

Klasse	1	2	3	4	5
H_j	2	7	21	177	249
h_j	.004	.014	.042	.354	.498

Klasse	6	7	8	9
H_j	40	2	1	1
h_j	.080	.004	.002	.002



Histogramm mit absoluten Häufigkeiten, nur bei äquidistanten Klassen verwenden!

Ansonsten: Wähle Höhe d_j des Balkens über dem Intervall $(t_j, t_{j+1}]$ so, dass $d_j \cdot (t_{j+1} - t_j) = h_j$ gilt. (Die Gesamtfläche ist dann 1!)

Klasseneinteilung allgemein

- Zweck: übersichtliche Darstellung großer Datenmengen bei stetigen und diskreten Merkmalen mit vielen Ausprägungen
- Klassen sind halboffene Intervalle der Form $(a, b]$

Vorgehen bei s Klassen $(a_1,\ a_2],\ (a_2,\ a_3], \ldots,\ (a_s,\ a_{s+1}]$ mit

$$a_1 < a_2 < \ldots < a_s < a_{s+1}, \quad a_1 < \min_{1 \le j \le n} x_j, \quad \max_{1 \le j \le n} x_j \le a_{s+1} :$$

Bilde über $(a_j, a_{j+1}]$ ein Rechteck, dessen Fläche gleich der zugehörigen relativen Klassenhäufigkeit

$$k_j \ := \ \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{a_j < x_i \le a_{j+1}\}$$

ist. Die Rechteckhöhe d_j ist also durch die Gleichung $d_j\,(a_{j+1} - a_j) = k_j$ bestimmt.

Beispiel zur Klasseneinteilung

Beispiel

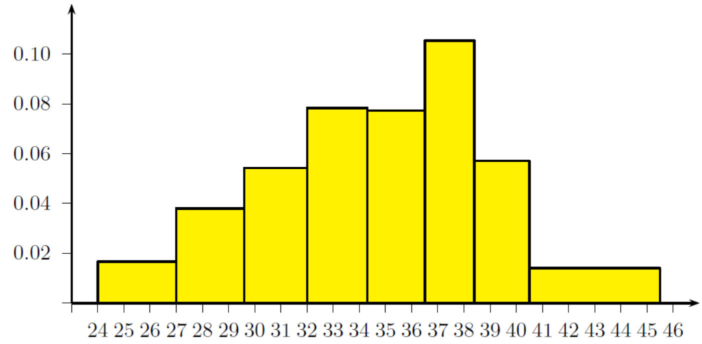
Gegeben sei eine Stichprobe vom Umfang $n = 100$:

37.4	37.8	29.0	35.1	30.9	28.5	38.4	34.7	36.3	30.4
39.1	37.3	45.3	32.2	27.4	37.0	25.1	30.7	37.1	37.7
26.4	39.7	33.0	32.5	24.7	35.1	33.2	42.4	37.4	37.2
37.5	44.2	39.2	39.4	43.6	28.0	30.6	38.5	31.4	29.9
34.5	34.3	35.0	35.5	32.6	33.7	37.7	35.3	37.0	37.8
32.5	32.9	38.0	36.0	35.3	31.3	39.3	34.4	37.2	39.0
41.8	32.7	33.6	43.4	30.4	25.8	28.7	31.1	33.0	39.0
37.1	36.2	28.4	37.1	37.4	30.8	41.6	33.8	35.0	37.4
33.7	33.8	30.4	37.4	39.3	30.7	30.6	35.1	33.7	32.9
35.7	32.9	39.2	37.5	26.1	29.2	34.8	33.3	28.8	38.9

Beispiel zur Klasseneinteilung

Beispiel

Mit $s = 8$ Klassen und $a_1 := 24$, $a_2 := 27$, $a_3 := 29.6$, $a_4 := 32$, $a_5 := 34.3$, $a_6 := 36.5$, $a_7 := 38.4$, $a_8 := 40.5$, $a_9 := 45.5$ ergibt sich folgendes Histogramm:



Kenngrößen: Lage- und Streumaße

Die Beschreibung von Daten/Stichproben mit Hilfe von Histogrammen ist häufig zu aufwändig. Oft genügen einfachere Kenngrößen zur Beschreibung der Daten.

Hierzu gehören

- verschiedene Mittelwerte,
- Stichprobenvarianzen.

Definition 13.3

Es sei $x = (x_1, \dots, x_n)$ eine Stichprobe vom Umfang $n \geq 2$ mit Werten in $\mathbb{R}$. Dann nennt man

$$\bar{x} := \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{x_1 + \cdots + x_n}{n}, \qquad \text{Stichproben-Mittel}$$

$$s_x^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1}, \qquad \text{Stichproben-Varianz}$$

$$s_x := \sqrt{s_x^2} \qquad \text{Stichproben-Standardabweichung}$$

von x und bezeichnet für $x_1, \dots, x_n > 0$ die Größe $v_x := \frac{s_x}{\bar{x}}$ als Stichproben-Variationskoeffizient.

Beispiele

Beispiel

- Quicksort: $\bar{x} = 0.325, \quad s_x = 0.017, \quad v_x = \frac{s_x}{\bar{x}} = 0.051$
- Shellsort: $\bar{x} = 0.584, \quad s_x = 0.019, \quad v_x = \frac{s_x}{\bar{x}} = 0.032$

Beispiel

Vergleiche die Varianzen und die Variationskoeffizienten für folgende Datensätze:

- $x = (1 \quad 2 \quad 3)$
- $x = (99 \quad 100 \quad 101)$
- $x = (50 \quad 100 \quad 150)$
- $x = (1 \quad 100 \quad 199)$

Mittelwert und Varianz m.H. relativer Häufigkeiten

Satz 13.4

Sei $x = (x_1, \ldots, x_n)$ eine Stichprobe mit den möglichen **reellen** Merkmalswerten $a_1, \ldots, a_k$. Dann gilt:

(1) $\bar{x} = \sum_{j=1}^{k} a_j \cdot h_x(a_j)$

(2)
$$\begin{aligned} s_x^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n}{n-1} \cdot \sum_{j=1}^{k} (a_j - \bar{x})^2 \cdot h_x(a_j) \\ &= \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2 \right] = \frac{n}{n-1} \left[\sum_{j=1}^{k} a_j^2 \cdot h_x(a_j) - \left(\sum_{j=1}^{k} a_j \cdot h_x(a_j) \right)^2 \right]. \end{aligned}$$

(3) Wird $y = (y_1, \ldots, y_n)$ aus x durch die Transformation $y_i = a \cdot x_i + b$, $i = 1, \ldots, n$ gewonnen, so ist

$$\bar{y} = a \cdot \bar{x} + b, \qquad s_y^2 = a^2 \cdot s_x^2, \qquad s_y = |a| \cdot s_x.$$

Geordnete Stichproben und Empirischer Median

Definition 13.5

Gegeben eine Stichprobe $x = (x_1, \dots, x_n)$ mit reellen Merkmalswerten x_i.

- Die aus den aufsteigend sortierten Elementen $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ von $x_1, \dots, x_n$ bestehende Stichprobe

  $$x_{()} := (x_{(1)}, x_{(2)}, \dots, x_{(n)})$$

  heißt die geordnete Stichprobe (zu x).

- Die reelle Zahl

  $$\tilde{x} := \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{falls } n \text{ ungerade,} \\ \frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right), & \text{falls } n \text{ gerade} \end{cases}$$

  heißt der Stichproben-Median (Zentralwert, empirischer Median) von x.

Beispiele und Erläuterungen zum Median

Beispiel

- $x = (3, 2, 2, 1)$, $x_{()} = (1, 2, 2, 3)$, $\tilde{x} = \frac{1}{2}(2 + 2) = 2$
- $x = (3, 3, 2, 1)$, $x_{()} = (1, 2, 3, 3)$, $\tilde{x} = \frac{1}{2}(2 + 3) = 2.5$
- $x = (5, 3, 3, 1, 2)$, $x_{()} = (1, 2, 3, 3, 5)$, $\tilde{x} = x_{(3)} = 3$

- Median ist „*Hälftigkeitswert*“ (mindestens 50% aller x_j sind kleiner oder gleich $\tilde{x}$ und mindestens 50% aller x_j sind größer oder gleich $\tilde{x}$).
- Der Median ist robust gegenüber Ausreißern!
- Der Median $\tilde{x}$ minimiert die *Abstands-Summe* als Funktion von $t \in \mathbb{R}$:

$$t \mapsto s(t) \ := \sum_{j=1}^{n} |x_j - t| \, .$$

α-Quantile

Für $y \in \mathbb{R}$ sei $\lfloor y \rfloor$ die größte ganze Zahl, welche kleiner oder gleich y ist:

$$\lfloor y \rfloor := \max\{k \in \mathbb{Z} : k \leq y\}.$$

Beispiele: $\lfloor 1.2 \rfloor = 1$, $\lfloor -0.3 \rfloor = -1$, $\lfloor 5 \rfloor = 5$

Definition 13.6 (α-Quantil)

Sei $\alpha \in (0,1)$, $x = (x_1, \ldots, x_n)$ und $k := \lfloor n \cdot \alpha \rfloor$. Dann heißt

$$\tilde{x}_\alpha := \begin{cases} x_{(k+1)}, & \text{falls } n \cdot \alpha \notin \mathbb{N} \\ \frac{1}{2} \cdot \left(x_{(k)} + x_{(k+1)}\right), & \text{sonst} \end{cases}$$

das Stichproben-α-Quantil. Es gilt $\tilde{x} = \tilde{x}_{\frac{1}{2}}$. Speziell heißen

$\tilde{x}_{0.25}$ das untere Quartil und

$\tilde{x}_{0.75}$ das obere Quartil.

Eigenschaften des α-Quantils

- Mindestens $\alpha \cdot 100\%$ aller x_j sind kleiner oder gleich $\tilde{x}_\alpha$.
- Mindestens $(1-\alpha) \cdot 100\%$ aller x_j sind größer oder gleich $\tilde{x}_\alpha$.
- Merke (fast korrekt): $\tilde{x}_\alpha$ teilt die geordnete Stichprobe im Verhältnis α zu $1-\alpha$.

Beispiel

j	1	2	3	4	5	6	7	8	9	10
x_j	8.5	1.5	75	4.5	6.0	3.0	3.0	2.5	6.0	9.0
$x_{(j)}$	1.5	2.5	3.0	3.0	4.5	6.0	6.0	8.5	9.0	75

Es gilt $\tilde{x}_{0.25} = x_{(\lfloor 2.5 \rfloor + 1)} = x_{(3)} = 3.0$, da $10 \cdot 0.25 = 2.5 \notin \mathbb{N}$.

Weitere Lagemaße: Getrimmte Mittel

Definition 13.7 (α-getrimmtes Stichproben-Mittel)

Sei $\alpha \in [0, 0.5)$ und $k := \lfloor n \cdot \alpha \rfloor$. Dann heißt

$$\bar{x}_\alpha := \frac{1}{n - 2 \cdot k} \cdot \left(x_{(k+1)} + \cdots + x_{(n-k)}\right)$$

das α-getrimmte (gestutzte) Stichproben-Mittel.

Insbesondere ist $\bar{x} = \bar{x}_0$.

Beispiel

$\alpha = 0.1, n = 500$. Dann $k = \lfloor 50.0 \rfloor = 50$.

$$\text{Quicksort:} \quad \bar{x}_{0.1} = \frac{x_{(51)} + \cdots + x_{(450)}}{400} = 0.325,$$

$$\text{Shellsort:} \quad \bar{x}_{0.1} = \frac{x_{(51)} + \cdots + x_{(450)}}{400} = 0.584.$$

Weitere Lagemaße: Getrimmte Mittel

Für $0 < \alpha < 1/2$ und $k := \lfloor n \cdot \alpha \rfloor$ ist also

$$\begin{aligned} \bar{x}_\alpha \quad &:= \quad \frac{1}{n - 2\,k} \left(x_{(k+1)} + x_{(k+2)} + \cdots + x_{(n-k-1)} + x_{(n-k)} \right) \\ &= \quad \frac{1}{n - 2\,k} \sum_{j=k+1}^{n-k} x_{(j)} \end{aligned}$$

das α-getrimmte Mittel oder $\alpha \cdot 100\%$-getrimmte Mittel von $x_1, \ldots, x_n$.

- $\bar{x}_\alpha$ ist „Kompromiss“ zwischen arithmetischem Mittel ($\alpha \downarrow 0$) und Median ($\alpha \uparrow 1/2$).
- $\bar{x}_\alpha$ ignoriert je $\alpha \cdot 100\%$ der Daten in den Enden der geordneten Stichprobe.
- $\bar{x}_\alpha$ ist flexibles Instrument gegenüber Ausreißern.

Weitere Streuungsmaße

Definition 13.8

Die folgenden Größen werden ebenfalls gelegentlich als Streuungsmaße einer Stichprobe x vom Umfang n verwendet:

- Quartilsabstand: $\tilde{x}_{0.75} - \tilde{x}_{0.25}$,

  Spannweite: $x_{(n)} - x_{(1)}$,

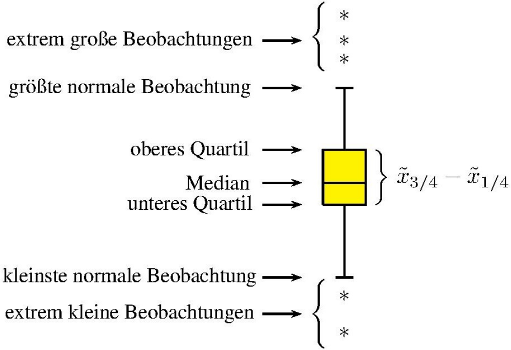
  mittlere absolute Abweichung: $\frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \tilde{x}|$.

Einige dieser Kenngrößen werden bei der graphischen Veranschaulichung von Daten mit Hilfe von **Box-Plots** verwendet.

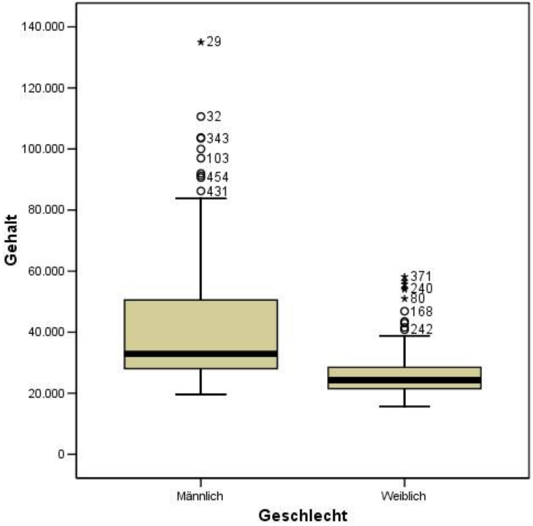
Box-Plots

Zweck: Schneller visueller Vergleich verschiedener Stichproben.

- Box vom unteren zum oberen Quartil, beim Median unterteilt.
- Stab bis zum größten x_j mit $x_j \leq \tilde{x}_{3/4} + 1.5 \cdot (\tilde{x}_{3/4} - \tilde{x}_{1/4})$ (größte normale Beob.)
- Stab bis zum kleinsten x_j mit $x_j \geq \tilde{x}_{1/4} - 1.5 \cdot (\tilde{x}_{3/4} - \tilde{x}_{1/4})$ (kleinste normale Beob.)



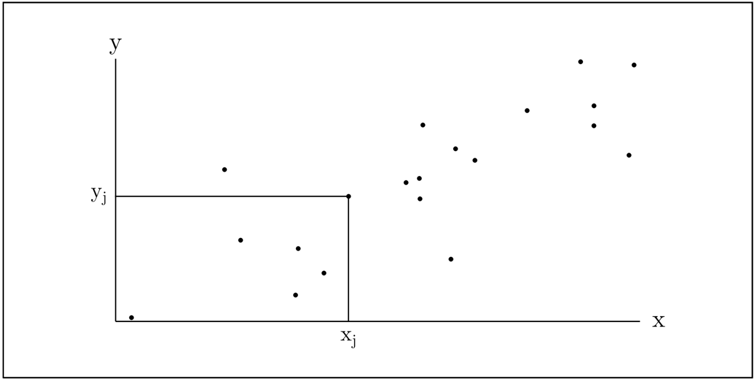
Beispiel Box-Plot



Parametrische Regression

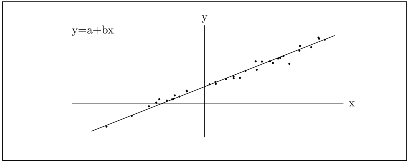
Gegeben: Stichprobe zweidimensionaler Daten $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathbb{R}^2$ vom Umfang $n \geq 2$, $x_1, \ldots, x_n$ nicht alle gleich, $y_1, \ldots, y_n$ nicht alle gleich.

Darstellung: Punktwolke oder Streuungsdiagramm:

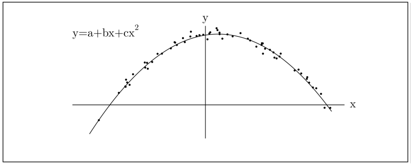


Parametrische Regression

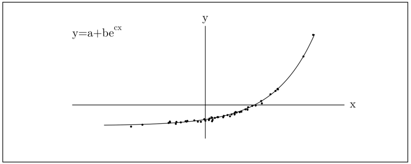
Ein (parametrisches) Regressionsmodell legt den Typ einer Regressionsfunktion $y = f_\theta(x)$ bis auf gewisse Parameter $\theta = (a, b, \ldots) \in \Theta$ (ein geeigneter Parameterraum) fest.



Regressionsgerade
$y = a + bx$



quadratische Regessionsfunktion
$y = a + bx - cx^2$



exponentielle Regressionsfunktion
$y = a + be^{cx}$

Allgemeines Regressionsproblem: Finde (zu den gegebenen Daten) innerhalb der vorgegebenen (parametrischen) Klasse von Funktionen $\{f_\theta : \theta \in \Theta\}$ diejenige Funktion f_{θ^*}, für welche die Summe der quadratischen Abstände

$$\sum_{j=1}^{n} (y_j - f_\theta(x_j))^2$$

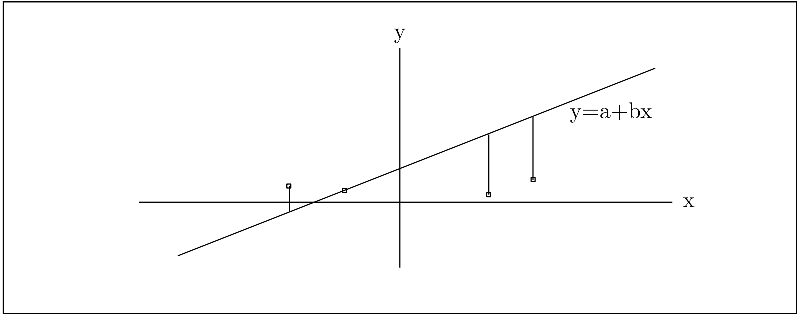
minimal wird.

Lineare Regression, Ausgleichsgerade

Spezialfall: Die Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x ist diejenige Gerade, für welche die Summe

$$\sum_{j=1}^{n} (y_j - a - b \cdot x_j)^2$$

minimal ist.



Zusammenhang mit Satz 12.10

Seien $(x_j, y_j) \in \mathbb{R}^2$ für $j = 1, \dots, n$ paarweise verschieden. Wir definieren

$$\mathbb{P}(X = x_j, Y = y_j) := \frac{1}{n}, \quad j = 1, \dots, n.$$

Hierdurch ist die gemeinsame Verteilung von X und Y festgelegt.

(Formal: $\Omega := \{(x_i, y_i) : i = 1, \dots, n\}$, $X := \pi_1$, $Y := \pi_2$ mit $\pi_1(x, y) := x$ und $\pi_2(x, y) := y$.)

Man erhält

$$\mathbb{E}\left[(Y - a - bX)^2\right] = \frac{1}{n} \sum_{i=1}^{n} (y_i - a - bx_i)^2 .$$

Damit ist das Problem der Ausgleichsgeraden der beschreibenden Statistik in den Rahmen der linearen Regression in Kapitel 12 (vgl. Folien 12.17ff) eingeordnet. Die optimalen Parameter a^* und b^* ergeben sich direkt aus Satz 12.10.

Ausgleichsgerade

Es ergeben sich die folgenden optimalen Parameter a^* und b^*:

$$b^* := \frac{\sum_{j=1}^n (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}, \qquad a^* := \bar{y} - b^* \cdot \bar{x}.$$

Mit

$$r_{xy} := \frac{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x}) \cdot (y_j - \bar{y})}{s_x \cdot s_y} =: \frac{\widetilde{s}_{xy}}{s_x s_y} \qquad (\text{ da } s_x, s_y > 0)$$

gilt

$$b^* = r_{xy} \cdot \frac{s_y}{s_x} = \frac{\widetilde{s}_{xy}}{s_x^2}.$$

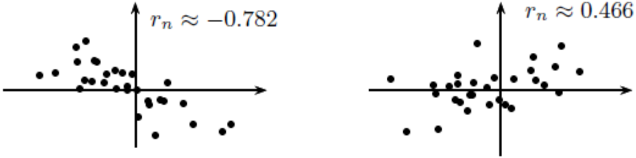
- r_{xy} heißt (empirischer) (Pearson-) Korrelationskoeffizient der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$.
- r_{xy} misst die Stärke des affinen Zusammenhangs von Merkmal 1 und Merkmal 2.
- $\widetilde{s}_{xy}$ heißt auch empirische Kovarianz oder Stichprobenkovarianz

Eigenschaften von $r_{x,y}$

(a) Es gilt $-1 \leq r_{xy} \leq +1$.

(b) Falls $r_{xy} \approx +1$, so liegt ein deutlich ansteigender linearer Trend vor.
Falls $r_{xy} \approx -1$, so liegt ein deutlich fallender linearer Trend vor.

(c) Ist $r_{xy} \approx 0$, so gibt es keinen statistischen Zusammenhang in Form einer linearen/affinen Beziehung zwischen den betrachteten Merkmalen.



(d) Ist $r_{xy} > 0$, so entsprechen wachsenden x_i-Werten „im Mittel“ wachsende y_i-Werte. Die Merkmale heißen positiv korreliert. Ist $r_{xy} < 0$, so entsprechen wachsenden x_i-Werten „im Mittel“ fallende y_i-Werte. Die Merkmale sind negativ korreliert.

(e) Bei affinen Daten-Transformationen der Form

$$\tilde{x}_j \;=\; a \cdot x_j + b, \qquad \tilde{y}_j \;=\; c \cdot y_j + d$$

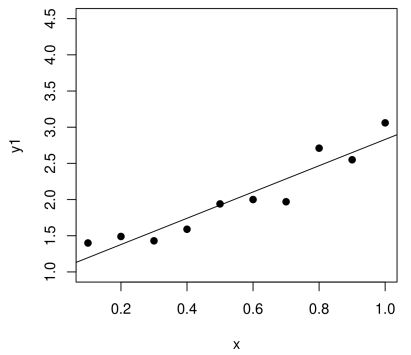
(mit $a > 0, c > 0$) ändert sich r_{xy} nicht, d.h. es gilt $r_{\tilde{x}\tilde{y}} = r_{xy}$.

Beispiele zur linearen Regression

Beispiel 1

$$x = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$$
$$y_1 = (1.40, 1.49, 1.43, 1.59, 1.94, 2.00, 1.97, 2.71, 2.55, 3.06)$$

Regressionsgerade: $y_1 \; = \; 1.02 \; + \; 1.82 \cdot x, \quad r_{xy_1} \; = \; 0.94$

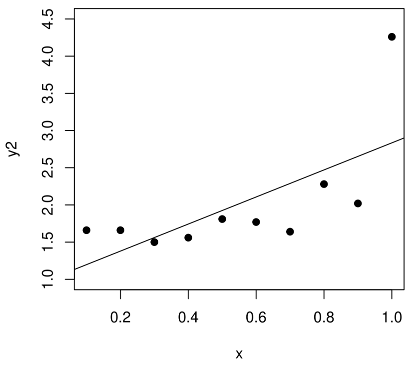


Beispiele zur linearen Regression

Beispiel 2

$$x = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$$
$$y_2 = (1.66, 1.66, 1.50, 1.56, 1.81, 1.77, 1.64, 2.28, 2.02, 4.26)$$

Regressionsgerade: $y_2 \ = \ 1.02 \ + \ 1.82 \cdot x, \quad r_{xy_2} \ = \ 0.67$



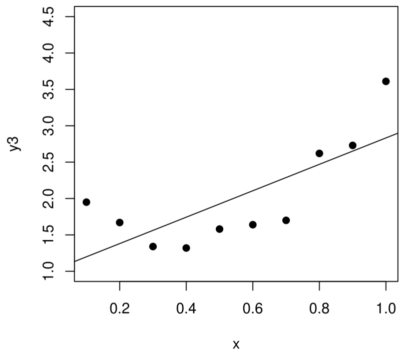
Beispiele zur linearen Regression

Beispiel 3

$$x = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$$

$$y_3 = (1.95, 1.67, 1.34, 1.32, 1.58, 1.64, 1.70, 2.62, 2.73, 3.61)$$

Regressionsgerade: $y_3 \; = \; 1.02 \; + \; 1.82 \cdot x, \quad r_{xy_3} \; = \; 0.74$



Kapitel 14: Die Multinomialverteilung

- Ein Zufallsexperiment habe s mögliche Ausgänge $1, \ldots, s$.
- Ausgang j bezeichnet man als Treffer j-ter Art (Ausgang vom Typ j).
- Das Zufallsexperiment wird n-mal unabhängig wiederholt.
- Die einzelnen Ausgänge müssen nicht gleich wahrscheinlich sein, die Wahrscheinlichkeiten ändern sich jedoch nicht im Verlauf der Wiederholungen.

Beispiele

- Würfelwurf ($s = 6$),
- Augensumme beim zweifachen Würfelwurf ($s = 11$),
- Verteilen einer Kugel auf eines von s Fächern,
- Ziehen (mit Zurücklegen) einer Kugel aus einer Urne, die verschiedenfarbige Kugeln (mit s verschiedenen Farben) enthält.

Es wird jeweils der Ausgang notiert, und der Versuch wird ingesamt n-mal unabhängig durchgeführt.

Modellierung

Sei $\Omega := \{1, \ldots, s\}^n$, d.h., $\omega = (a_1, \ldots, a_n) \in \Omega$ mit $a_i \in \{1, \ldots, s\}$.

$a_i = j$ bedeutet *Treffer j-ter Art* (oder *Treffer vom Typ j*) *im i-ten Versuch.*

Sei p_j die Wahrscheinlichkeit für einen Treffer j-ter Art. Dann gilt $\sum_{j=1}^{s} p_j = 1$.

Aus der geforderten Unabhängigkeit folgt der Ansatz

$$p(\omega) = \prod_{j=1}^{s} p_j^{k_j}, \quad k_j = \sum_{i=1}^{n} \mathbb{1}\{a_i = j\}, \quad j = 1, \ldots, s.$$

Sei

$$A_{ij} := \{(a_1, \ldots, a_n) \in \Omega : a_i = j\} \quad \text{“Treffer } j\text{-ter Art im } i\text{-ten Versuch”.}$$

Bsp. nach Satz 10.2 zeigt: $\mathbb{P}(A_{ij}) = p_j$ und für $j_1, \ldots, j_n \in [s]$ sind $A_{1j_1}, \ldots, A_{nj_n}$ stochastisch unabhängig.

Ferner gilt für die Anzahl der Treffer j-ter Art

$$X_j := \sum_{i=1}^{n} \mathbb{1}_{A_{ij}} \sim \mathrm{Bin}(n, p_j).$$

Modellierung

Gesucht: Verteilung des Zufallsvektors $(X_1, \ldots, X_s)$.

Beachte: $X_1 + \cdots + X_s = n$, das heißt $X_1, \ldots, X_s$ sind "abhängig".

Seien $k_1, \ldots, k_s \in \mathbb{N}_0$ mit $k_1 + \cdots + k_s = n$. Dann ist die W. des folgenden Ereignisses gesucht:

$$\{X_1 = k_1, \ldots, X_s = k_s\} = \{\omega \in \Omega : \text{Anzahl Treffer } j\text{-ter Art ist } k_j, 1 \le j \le s\}.$$

Jedes Tupel ω in dieser Menge erfüllt

$$\mathbb{P}(\{\omega\}) = p_1^{k_1} \cdots p_s^{k_s}.$$

Die Anzahl dieser Tupel ist

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{n-k_1-\ldots-k_{s-1}}{k_s} = \frac{n!}{k_1! \cdots k_s!} = \binom{n}{k_1, \ldots, k_s}.$$

Der Zufallsvektor $(X_1, \ldots, X_s)$ hat also die in der folgenden Definition angegebene Verteilung.

Multinomialverteilung

Definition und Behauptung 14.1

Ein auf einem diskreten W.raum erklärter Zufallsvektor $(X_1, \ldots, X_s)$ mit Werten in $\mathbb{N}_0^s$ hat eine Multinomialverteilung mit den Parametern $s, n \in \mathbb{N}$ und $p_1, \ldots, p_s \in [0, 1]$, wobei $s \geq 2$, $n \geq 1$ und $\sum_{j=1}^{s} p_j = 1$ gelten, wenn

$$\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) = \binom{n}{k_1, \ldots, k_s} p_1^{k_1} \cdots p_s^{k_s} \qquad (\star)$$

gilt mit der Konvention $\binom{n}{k_1, \ldots, k_s} = 0$, falls $k_1, \ldots, k_s \notin \mathbb{N}_0$ oder $\sum_{j=1}^{s} k_j \neq n$.

Notation:

$$(X_1, \ldots, X_s) \sim \mathrm{Mult}(n; p_1, \ldots, p_s).$$

Bemerkung: Die Multinomialformel zeigt, dass durch $(\star)$ in der Tat eine Wahrscheinlichkeitsverteilung erklärt wird. Es gilt $\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) > 0$ nur für $k_1, \ldots, k_s \in \mathbb{N}_0$ mit $k_1 + \cdots + k_s = n$.

6-facher Würfelwurf

Beispiel

Ein fairer Würfel wird sechsmal geworfen, X_j sei die Anzahl der Ausgänge mit Augenzahl j. Dann gilt $(X_1, \dots, X_6) \sim \mathrm{Mult}(6; \frac{1}{6}, \dots, \frac{1}{6})$.

(a) Was ist die Wahrscheinlichkeit, dass jede Augenzahl genau einmal auftritt?

$$\mathbb{P}(X_1 = 1, \dots, X_6 = 1) = \binom{6}{1, \dots, 1} \left(\frac{1}{6}\right)^6 = 6!\ 6^{-6} \approx 0.0154.$$

(b) Was ist die Wahrscheinlichkeit, dass irgendwelche zwei Augenzahlen je dreimal auftreten?

$$\mathbb{P}(X_1 = 3, X_2 = 3, X_4 = \cdots = X_6 = 0) + \cdots + \mathbb{P}(X_1 = \cdots = X_4 = 0, X_5 = X_6 = 3) = \binom{6}{2} \frac{6!}{3!\ 3!\ 0!\ 0!\ 0!\ 0!} \left(\frac{1}{6}\right)^6 \approx 0.00643.$$

(c) Was ist die Wahrscheinlichkeit, dass eine Augenzahl zweimal und vier Augenzahlen je einmal auftreten?

$$6 \cdot 5 \cdot \mathbb{P}(X_1 = 2, X_2 = 0, X_3 = \cdots = X_6 = 1) = 30 \cdot \frac{6!}{2!0!1!^4} \cdot \left(\frac{1}{6}\right)^6 \approx 0.231.$$

Randverteilungen der Multinomialverteilung

Aufgabe: Bestimme alle (partiellen) Randverteilungen der Multinomialverteilung $(X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s)$.

Es gilt $X_j \sim \text{Bin}(n, p_j)$ (klar; siehe Tutorium).

Allgemeiner: Trefferarten zu paarweise disjunkten Gruppen zusammenfassen! Die Gruppentrefferzahlen folgen dann wiederum einer Multinomialverteilung.

Satz 14.2

Sei $(X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s)$. Weiter seien $T_1 + \cdots + T_\ell = \{1, \ldots, s\}$ mit $\ell \geq 2$, $T_r \neq \emptyset$ für $r = 1, \ldots, \ell$ und

$$Y_r := \sum_{k \in T_r} X_k, \; q_r := \sum_{k \in T_r} p_k, \quad r = 1, \ldots, \ell.$$

Dann gilt

$$(Y_1, \ldots, Y_\ell) \sim \text{Mult}(n; q_1, \ldots, q_\ell).$$

Insbesondere folgt $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$ für $i \neq j$.

Beweis von Satz 14.2

Beweis

Seien $k_1, \dots, k_\ell \in \mathbb{N}_0$, $k_1 + \cdots + k_\ell = n$. Sei

$$R := \left\{ r = (r_1, \dots, r_s) \in \mathbb{N}_0^s : \sum_{i \in T_j} r_i = k_j \text{ für } j = 1, \dots, \ell \right\}.$$

Dann erhält man (da $(X_1, \dots, X_s) \sim \mathrm{Mult}(n; p_1, \dots, p_s)$)

$$\begin{aligned} \mathbb{P}(Y_1 = k_1, \dots, Y_\ell = k_\ell) &= \mathbb{P}\left(\sum_{r \in R} \{X_1 = r_1, \dots, X_s = r_s\} \right) = \sum_{r \in R} \mathbb{P}(X_1 = r_1, \dots, X_s = r_s) \\ &= \sum_{r \in R} \binom{n}{r_1, \dots, r_s} p_1^{r_1} \cdots p_s^{r_s} \\ &= n! \sum_{r \in R} \prod_{j=1}^{\ell} \left(\frac{1}{\prod_{i \in T_j} r_i!} \cdot \prod_{i \in T_j} p_i^{r_i} \right) \end{aligned}$$

Beweis von Satz 14.2

Beweis

und mit Hilfe von Distributivgesetz und Multinomialformel folgt nun

$$\begin{aligned}\mathbb{P}(Y_1 = k_1, \ldots, Y_\ell = k_\ell) &= n! \sum_{r \in R} \prod_{j=1}^{\ell} \left(\frac{1}{\prod\limits_{i \in T_j} r_i!} \cdot \prod_{i \in T_j} p_i^{r_i} \right) = \frac{n!}{\prod\limits_{j=1}^{\ell} k_j!} \sum_{r \in R} \left[\prod_{j=1}^{\ell} \frac{k_j!}{\prod\limits_{i \in T_j} r_i!} \prod_{i \in T_j} p_i^{r_i} \right] \\ &= \binom{n}{k_1, \ldots, k_\ell} \prod_{j=1}^{\ell} \left[\sum_{\sum\limits_{i \in T_j} r_i = k_j} \frac{k_j!}{\prod\limits_{i \in T_j} r_i!} \prod_{i \in T_j} p_i^{r_i} \right] \\ &= \binom{n}{k_1, \ldots, k_\ell} \prod_{j=1}^{\ell} \left(\sum_{i \in T_j} p_i \right)^{k_j} = \binom{n}{k_1, \ldots, k_\ell} q_1^{k_1} \cdots q_\ell^{k_\ell}.\end{aligned}$$

Kovarianz und Korrelation der Multinomialverteilung

Satz 14.3 (Kovarianz, Korrelation von Mult)

Sei $(X_1, \dots, X_s) \sim \text{Mult}(n; p_1, \dots, p_s)$. Dann gilt

(a) $\mathbb{C}(X_i, X_j) = -np_ip_j$ für $i \neq j$;

(b) $\varrho(X_i, X_j) = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}, \quad i \neq j,\ p_ip_j > 0.$

Beweis

Aufgrund des vorherigen Satzes gilt $X_i \sim \text{Bin}(n, p_i)$ und $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$ für $i \neq j$. Damit erhält man einerseits

$$\mathbb{V}(X_i + X_j) = n(p_i + p_j)(1 - (p_i + p_j)),$$

und andererseits

$$\begin{aligned}\mathbb{V}(X_i + X_j) &= \mathbb{V}(X_i) + \mathbb{V}(X_j) + 2\mathbb{C}(X_i, X_j)\\ &= np_i(1 - p_i) + np_j(1 - p_j) + 2\mathbb{C}(X_i, X_j).\end{aligned}$$

Gleichsetzen ergibt Behauptung (a), und (b) folgt aus (a).

Mehrdimensionale hypergeometrische Verteilung

Die Multinomialverteilung verallgemeinert die Binomialverteilung auf den Fall von mehr als zwei möglichen Ausgängen in jedem Teilexperiment. Hierbei werden Unabhängigkeit und Erhalt der Wahrscheinlichkeiten $p_1, \ldots, p_s$ der einzelnen Ausgänge $1, \ldots, s$ in den n Teilexperimenten vorausgesetzt.

Wir betrachten nun das folgende Modell: Eine Urne enthalte r_j Kugeln der Farbe j mit $j = 1, \ldots, s$ und $s \geq 2$. Das Zufallsexperiment besteht im n-maligen Ziehen jeweils einer Kugel **ohne** Zurücklegen. (Hierbei wird $n \leq \sum_{j=1}^{s} r_j$ angenommen.)

Sei $X_j =$ Anzahl der insgesamt gezogenen Kugeln der Farbe j.

Man nennt die Verteilung von $(X_1, \ldots, X_s)$ eine **mehrdimensionale hypergeometrische Verteilung**. Es gilt

$$\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) = \frac{\binom{r_1}{k_1} \cdots \binom{r_s}{k_s}}{\binom{r_1 + \cdots + r_s}{n}}, \quad (k_1, \ldots, k_s) \in \mathbb{N}_0^s, \ \sum_{i=1}^{s} k_i = n,$$

mit der üblichen Konvention $\binom{r_i}{k_i} = 0$ falls $k_i > r_i$ etc.

Erfolgt das Ziehen **mit** Zurücklegen, so ist $(X_1, \ldots, X_s) \sim \mathrm{Mult}(n; p_1, \ldots, p_s)$ mit $p_i = \frac{r_i}{r_1 + \cdots + r_s}$, $i = 1, \ldots, s$. Man erhält also eine Multinomialverteilung.

Kapitel 15: Wartezeitverteilungen

Betrachte eine Bernoulli-Kette mit $p \in (0,1)$, d.h., eine unabhängige Folge von Zufallsexperimenten mit Werten in $\{0,1\}$ und $\mathbb{P}(\{1\}) = p \in (0,1)$. Es bezeichne X die Anzahl Nieten (0) vor dem ersten Treffer (1). Was ist die Verteilung der Zufallsvariable X?

$$\{X = k\} \text{ bedeutet:} \qquad \underbrace{0\ldots0}_{k-mal}1 \qquad \to \qquad \mathbb{P}(X = k) = (1-p)^k p$$

Grundraum: $\Omega = \{1, 01, 001, 0001, \ldots\}$.

Definition und Behauptung 15.1

Sei $p \in (0,1)$ fest. Sei $(\Omega, \mathbb{P})$ ein diskreter W.raum und X eine reelle Zufallsvariable auf Ω mit Werten in $\mathbb{N}_0$ und

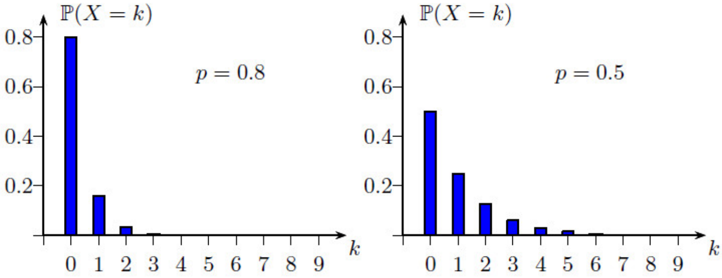
$$\mathbb{P}(X = k) = (1-p)^k p, \quad k \in \mathbb{N}_0.$$

Dann sagt man, dass X eine geometrische Verteilung hat.

Notation: $X \sim \mathrm{G}(p)$.

Bemerkungen zur geometrischen Verteilung

- Die Wahrscheinlichkeiten summieren sich in der Tat zu 1 (siehe den Beweis des folgenden Satzes).
- $X + 1$ beschreibt dann die Anzahl der **Versuche** bis (einschließlich) zum ersten Treffer.
- $k \mapsto (1-p)^k p$ ist (streng) monoton fallend
- Stabdiagramme:



Stabdiagramme geometrischer Verteilungen

Erwartungswert und Varianz der geometrischen Verteilung

Satz 15.2

Sei $X \sim \mathrm{G}(p)$. Dann gilt

(a) $\mathbb{E}(X) = \frac{1-p}{p} = \frac{1}{p} - 1$,

(b) $\mathbb{V}(X) = \frac{1-p}{p^2}$.

Beweis von Satz 15.2

Beweis

Es liegt in der Tat eine W.verteilung vor, denn $\sum_{k=0}^{\infty}(1-p)^k p = p\frac{1}{1-(1-p)} = 1$.

(a)
$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k(1-p)^k p \;=\; p(1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1} = p(1-p)\left(\frac{d}{dt}\sum_{k=0}^{\infty} t^k\right)\Bigg|_{t=1-p}$$
$$= p(1-p)\left(\frac{d}{dt}\frac{1}{1-t}\right)\Bigg|_{t=1-p} = p(1-p)\left(\frac{1}{1-t}\right)^2\Bigg|_{t=1-p} = \frac{1-p}{p}.$$

(b) Beachte zunächst, dass $\sum\limits_{k=2}^{\infty} k(k-1)(1-p)^{k-2} = \frac{d^2}{dt^2}\sum\limits_{k=0}^{\infty} t^k\Big|_{t=1-p} = \frac{2}{p^3}$. Damit folgt

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)(1-p)^k p = p(1-p)^2\frac{2}{p^3} = \frac{2(1-p)^2}{p^2}.$$

Wegen

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}X)^2$$

erhält man die Behauptung durch Nachrechnen.

Sechser im Lotto

Beispiel

Ein Lottospieler gibt regelmäßig 10 verschiedene Tippreihen ab. Pro Woche gibt es zwei Ausspielungen. Wie groß ist der Erwartungswert in Jahren bis zum ersten Sechser?

Modell: Bernoulli-Kette mit $p = \frac{10}{\binom{49}{6}}$.

Sei X die Anzahl der Ausspielungen bis (einschließlich) zum ersten Erfolg. Dann ist $X - 1 \sim \mathrm{G}(p)$ und somit

$$\mathbb{E}(X) = (\frac{1}{p} - 1) + 1 = \frac{1}{p} \approx 1.398.381,6.$$

Bei circa 104 Ausspielungen pro Jahr ergibt sich eine mittlere Wartezeit von circa 13.446 Jahren!

Gedächtnislosigkeit der geometrischen Verteilung

Satz 15.3

(a) Für $X \sim \mathrm{G}(p)$ gilt

$$\mathbb{P}(X = m + k \mid X \geq k) = \mathbb{P}(X = m), \quad m, k \in \mathbb{N}_0. \tag{3}$$

(b) Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable mit der Eigenschaft (3) und $0 < \mathbb{P}(X = 0) < 1$. Dann gilt $X \sim \mathrm{G}(p)$ für ein $p \in (0, 1)$.

Beweis von Satz 15.3

Beweis

(a) Sei $X \sim \mathrm{G}(p)$. Dann gilt für beliebige $k, m \in \mathbb{N}_0$

$$\begin{aligned}\mathbb{P}(X = m + k \mid X \geq k) &= \frac{\mathbb{P}(X = m + k, X \geq k)}{\mathbb{P}(X \geq k)} = \frac{\mathbb{P}(X = m + k)}{\mathbb{P}(X \geq k)} = \frac{(1-p)^{m+k} p}{\sum_{\ell=k}^{\infty} (1-p)^{\ell} p} \\ &= \frac{(1-p)^{m+k}}{(1-p)^k \sum_{i=0}^{\infty} (1-p)^i} = \frac{(1-p)^m}{\frac{1}{1-(1-p)}} = p(1-p)^m = \mathbb{P}(X = m).\end{aligned}$$

(b) Die Voraussetzung (3) wird nur für $k = 1$ und beliebiges $m \in \mathbb{N}_0$ benötigt. Dann ist

$$\mathbb{P}(X = m) = \mathbb{P}(X = m + 1 \mid X \geq 1) = \frac{\mathbb{P}(X = m + 1)}{\mathbb{P}(X \geq 1)}, \qquad m \in \mathbb{N}_0.$$

Setze $p_m := \mathbb{P}(X = m)$, $m \in \mathbb{N}_0$ mit $p_0 \in (0, 1)$ nach Voraussetzung. Es folgt

$$p_m = \frac{p_{m+1}}{1 - p_0}, \quad m \in \mathbb{N}_0, \text{ das heißt } p_{m+1} = (1 - p_0) p_m = \ldots = (1 - p_0)^{m+1} p_0.$$

Wartezeit bis zum r-ten Treffer

Verallgemeinerung: Wir betrachten eine Bernoulli-Kette. Für ein festes $r \in \mathbb{N}$ sei X die Anzahl der Nieten vor dem r-ten Treffer.

$$\{X = k\}: \qquad \underbrace{***\ldots **}_{k\text{-mal } 0 \text{ und } (r-1)-\text{mal } 1}1 \;\rightarrow\; (1-p)^k p^r.$$

Anzahl solcher Folgen: $\binom{r+k-1}{k}$.

Dies motiviert die folgende Definition.

Definition und Behauptung 15.4 (Negative Binomialverteilung)

Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und X eine reelle Zufallsvariable mit

$$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \qquad k \in \mathbb{N}_0,$$

wobei $r \in \mathbb{N}$ und $p \in (0,1)$ fest seien. Dann sagt man, dass X eine negative Binomialverteilung (mit den Parametern r und p) hat.

Notation: $X \sim \mathrm{Nb}(r,p)$.

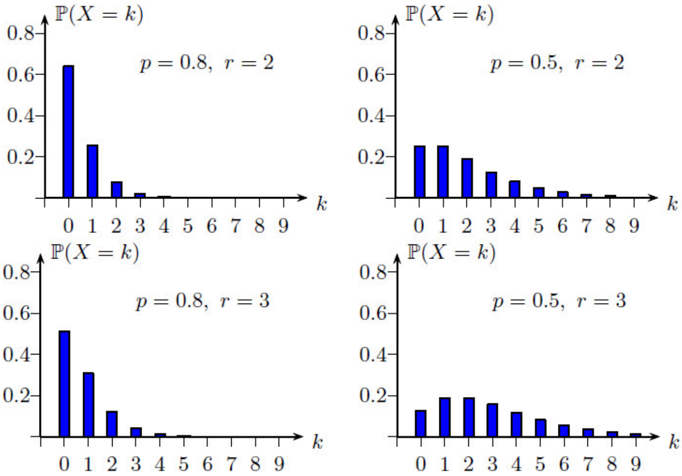
Bemerkungen zur negativen Binomialverteilung

- Es gilt $\binom{k+r-1}{k} = \frac{(k+r-1)\cdots(r+1)r}{k!} = (-1)^k\binom{-r}{k}$, wobei $\binom{a}{k} := \frac{a(a-1)\cdots(a-k+1)}{k!}$, $a \in \mathbb{C}$. Damit folgt (Binomialreihe!)

$$\sum_{k=0}^{\infty} \mathbb{P}(X=k) = \sum_{k=0}^{\infty} \binom{k+r-1}{k} p^r (1-p)^k = p^r \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} (1-p)^k$$
$$= p^r (1-(1-p))^{-r} = p^r (p^{-r}) = 1.$$

  Es handelt sich also tatsächlich um eine Zähldichte.
- Für $r = 1$ ergibt sich die geometrische Verteilung, d.h. $\mathrm{Nb}(1,p) = \mathrm{G}(p)$.
- $\mathbb{P}(X=0) < \mathbb{P}(X=1) \Leftrightarrow\ r(1-p) > 1$.
- Monotonieverhalten von $k \mapsto \mathbb{P}(X=k)$: $\searrow$ oder $\nearrow\searrow$

Stabdiagramme negativer Binomialverteilungen



Stabdiagramme von negativen Binomialverteilungen

Erwartungswert und Varianz der neg. Binomialverteilung

Satz 15.5

Sei $r \in \mathbb{N}$ und $p \in (0,1)$. Für $X \sim \mathrm{Nb}(r,p)$ gilt

(a) $\mathbb{E}(X) = r\frac{1-p}{p}$,

(b) $\mathbb{V}(X) = r\frac{1-p}{p^2}$.

Beweis: Man kann Eerwartungswert und Varianz direkt nachrechnen (siehe Übung/Tutorium). Alternativ ergeben sie sich auch aus dem nachfolgenden Satz.

Additionsgesetze für negative Binomialverteilungen

Intuitiv: $\mathrm{G}(p) = \mathrm{Nb}(1, p)$ und für $r \geq 2$:

$$\underbrace{00}_{X_1}\,1\,\underbrace{0000}_{X_2}\,1\,\underbrace{0}_{X_3}\,1\,\underbrace{000}_{X_4}\,1\,\underbrace{}_{X_5}\,1\,\underbrace{0000000}_{X_6}\,1\ldots$$

Ist $X \sim \mathrm{Nb}(r, p)$, so sollte $X \sim X_1 + \cdots + X_r$ mit $X_i \sim \mathrm{G}(p)$ gelten mit stochastisch unabhängigen Zufallsvariablen $X_1, \ldots, X_r$.

Satz 15.6 (Struktur- und Additionsgesetze für Nb)

Seien $r, s \in \mathbb{N}$ und $p \in (0, 1)$.

(a) Sind $X_1, \ldots, X_r$ unabhängig, $X_j \sim \mathrm{G}(p)$, $j = 1, \ldots, r$, so ist $\sum_{j=1}^{r} X_j \sim \mathrm{Nb}(r, p)$.

(b) Sind X, Y unabhängig, $X \sim \mathrm{Nb}(r, p)$, $Y \sim \mathrm{Nb}(s, p)$, so ist $X + Y \sim \mathrm{Nb}(r + s, p)$.

Beweis von Satz 15.5 mit Hilfe von Satz 15.6

Beweis

Wir betrachten unabhängige Zufallsvariablen $X_1, \dots, X_r$ mit $X_i \sim \mathrm{G}(p),\ i \in \{1, \dots, r\}$ auf einem geeigneten Wahrscheinlichkeitsraum. Dann ist (nach Satz 15.6) $X' := X_1 + \cdots + X_r \sim \mathrm{Nb}(r, p)$, und deshalb folgt (nach Satz 15.2 (a), (b)):

$$\mathbb{E}[X'] = \sum_{i=1}^{r} \mathbb{E}X_i = r\frac{1-p}{p} \quad \text{sowie} \quad \mathbb{V}(X') = \sum_{i=1}^{r} \mathbb{V}(X_i) = r\frac{1-p}{p^2}.$$

(Nur bei der zweiten Gleichung geht die Unabhängigkeit ein.)

Beachte: $\mathbb{E}, \mathbb{V}$ hängen nur von der Verteilung ab!

Hilfsaussage zum Beweis von Satz 15.6

Lemma

Sind $a, b \in \mathbb{R}$, $k \in \mathbb{N}_0$, so gilt

$$\binom{a+b}{k} = \sum_{j=0}^{k} \binom{a}{j}\binom{b}{k-j}.$$

Beweisidee

Vergleiche die Binomialreihen

$$(1+x)^a, \quad (1+x)^b, \quad (1+x)^{a+b}$$

für $|x| < 1$. Bilde hierzu das Cauchy-Produkt der Reihen $(1+x)^a$, $(1+x)^b$ und nimm einen Koeffizientenvergleich vor (Eindeutigkeitssatz für Potenzreihen).

Beweis von Satz 15.6

Beweis

Teil (a) folgt durch wiederholte Anwendung von (b) (vollständige Induktion).

(b) Wir verwenden die Faltungsformel, das Lemma und $\binom{r+k-1}{k} = (-1)^k\binom{-r}{k}$. Damit gilt

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{j=0}^{k} \mathbb{P}(X=j)\mathbb{P}(Y=k-j) \\
&= \sum_{j=0}^{k} \binom{j+r-1}{j}(1-p)^j p^r \binom{k-j+s-1}{k-j}(1-p)^{k-j}p^s \\
&= (1-p)^k p^{r+s} \sum_{j=0}^{k} (-1)^j \binom{-r}{j}(-1)^{k-j}\binom{-s}{k-j} \\
&= (1-p)^k p^{r+s}(-1)^k \binom{-(r+s)}{k} \\
&= (1-p)^k p^{r+s} \binom{r+s+k-1}{k}.
\end{aligned}$$

Wartezeiten bei Würfelwürfen

Beispiel

Ein echter Würfel wird in unabhängiger Folge geworfen. Wie hoch ist die Wahrscheinlichkeit, dass die dritte Sechs im 16. Wurf auftritt ?

$$\text{Antwort:} \quad \binom{3+13-1}{13}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{13} = \binom{15}{2}\frac{5^{13}}{6^{16}} \approx 0.054.$$

Beispiel

Ein Würfel wird solange geworfen, bis die erste Sechs auftritt. Wie hoch ist die Wahrscheinlichkeit, dass davor genau $\ell \in \mathbb{N}_0$-mal die Drei auftritt?

$$\text{Antwort (intuitiv):} \quad \left(\frac{1}{2}\right)^{\ell+1} \quad \text{Formal? (Tutorium)}$$

Kapitel 16: Die Poissonverteilung

- Wichtige diskrete Verteilung, insbesondere für Zählvorgänge seltener Ereignisse,
- Träger ist $\mathbb{N}_0$,
- Ausgangspunkt: Ideales Zufallsexperiment mit Erfolgswahrscheinlichkeit p_n (klein) wird n mal (oft) durchgeführt.

Poisson-Verteilung als Grenzfall von $\mathrm{Bin}(n, p_n)$-Verteilungen:

Wir betrachten eine Folge von Verteilungen $\mathrm{Bin}(n, p_n)$, $n \in \mathbb{N}$, mit $p_n \searrow 0$ und $n \cdot p_n \to \lambda > 0$ für $n \to \infty$.
Für $n \to \infty$ verschwindet die Wahrscheinlichkeit für Treffer, $p_n \to 0$, aber für die mittlere Trefferzahl gilt $n \cdot p_n \to \lambda$.
Wahrscheinlichkeit für genau $k \in \mathbb{N}_0$ Treffer bei n Versuchen, wenn $n \to \infty$:

$$\binom{n}{k} p_n^k \, (1 - p_n)^{n-k} = \frac{1}{k!} \underbrace{(np_n)^k}_{\longrightarrow \lambda^k} \underbrace{\frac{n^{\underline{k}}}{n^k}}_{\longrightarrow 1} \underbrace{\left(1 - \frac{np_n}{n}\right)^n}_{\longrightarrow e^{-\lambda}} \underbrace{(1 - p_n)^{-k}}_{\longrightarrow 1} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Gesetz seltener Ereignisse und Poissonverteilung

Satz 16.1 (Gesetz seltener Ereignisse)

Sei $(p_n)_{n\in\mathbb{N}}$ eine Folge in $[0,1]$ mit $\lim_{n\to\infty} np_n = \lambda \in (0,\infty)$. Dann gilt

$$\binom{n}{k} p_n^k \left(1-p_n\right)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k \in \mathbb{N}_0.$$

Definition und Behauptung 16.2

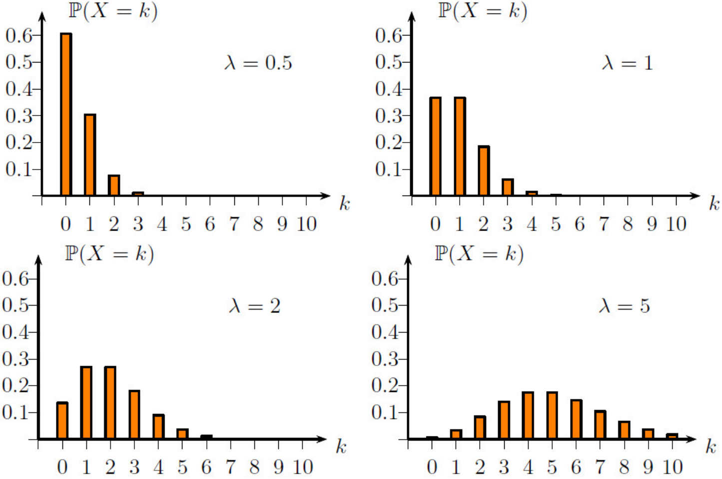
Sei $\lambda > 0$ und X eine reelle Zufallsvariable auf einem W.raum $(\Omega, \mathbb{P})$ mit

$$\mathbb{P}^X(k) = \mathbb{P}(X = k) := f_\lambda(k) := e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0,$$

so sagt man, X hat eine Poissonverteilung mit Parameter λ.

Notation: $X \sim \mathrm{Po}(\lambda)$.

Zähldichten von Poissonverteilungen



Stabdiagramme von Poisson-Verteilungen

Vergleich von Binomial- und Poissonverteilungen

(mit gleichem Erwartungswert)

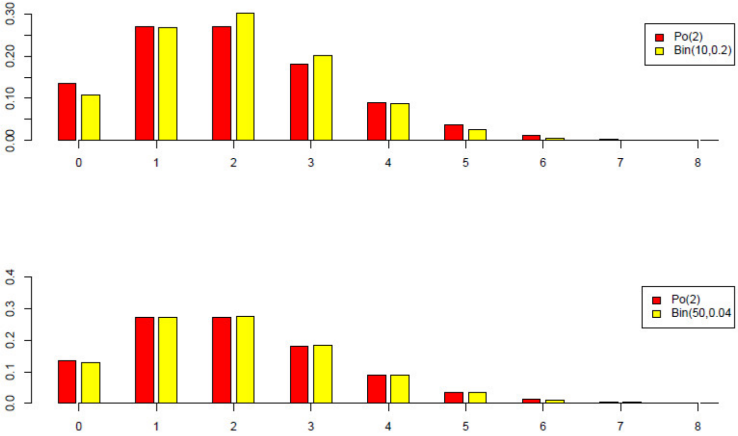


Abbildung: Vergleich Bin(n, p) und Po(np).

Eigenschaften der Zähldichten

Sei $k \mapsto f_\lambda(k)$, $k \in \mathbb{N}_0$ die Zähldichte von $\mathrm{Po}(\lambda)$.

- Ist $0 < \lambda < 1$, so ist $k \mapsto f_\lambda(k)$ streng monoton fallend.
- Ist $\lambda \geq 1$ und $\lambda \notin \mathbb{N}$, so ist $k \mapsto f_\lambda(k)$ streng monoton steigend bis $k = \lceil \lambda - 1 \rceil$, dann streng monoton fallend.
- Ist $\lambda \geq 1$ und $\lambda \in \mathbb{N}$, so ist $k \mapsto f_\lambda(k)$ streng monoton steigend bis $k = \lambda - 1$, $f_\lambda(\lambda - 1) = f_\lambda(\lambda)$, danach streng monoton fallend, d.h.

$$\ldots f_\lambda(\lambda - 2) < f_\lambda(\lambda - 1) = e^{-\lambda} \frac{\lambda^{\lambda - 1}}{(\lambda - 1)!} = f_\lambda(\lambda) > f_\lambda(\lambda + 1) \ldots$$

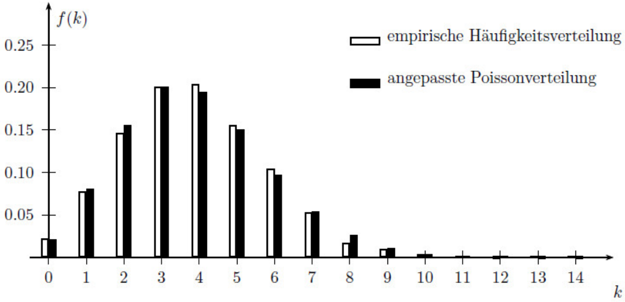
- $f_\lambda(k) \to 0$ für $k \to \infty$.

Das Rutherford-Geiger-Experiment

Im Jahre 1910 untersuchten Rutherford und Geiger ein radioaktives Präparat über 2608 Zeitintervalle von je 7.5 Sekunden Länge. Beobachtet wurden insgesamt 10097 Zerfälle, also im Durchschnitt 3.87 Zerfälle innerhalb von 7.5 Sekunden. Es sei n_k die Anzahl der Zeitintervalle, in denen k Zerfälle beobachtet wurden.

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_k	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1

Vergleich mit „angepasster“ Poisson–Verteilung (mit dem Parameter $\lambda = 3.87$):



Erwartungswert und Varianz der Poissonverteilung

Satz 16.3

Seien $X \sim \text{Po}(\lambda)$, $Y \sim \text{Po}(\mu)$ und X, Y stochastisch unabhängig. Dann gilt

(a) $\mathbb{E}(X) = \lambda = \mathbb{V}(X)$;

(b) $X + Y \sim \text{Po}(\lambda + \mu)$.

Beweis: Übung / Tutorium.
(Hinweise: (a) direkte Rechnung (Exponentialreihe, $\mathbb{V}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$); (b) Faltungsformel.)

Beispiel

Fazit: Zählt eine Zufallsvariable X, wieviele Ereignisse aus einer großen Zahl von seltenen (d.h. mit kleiner W.) Ereignissen eintreten, die kaum voneinander abhängen, so liegt es nahe, die Verteilung von X mit einer Poissonverteilung Po(λ) mit $\lambda = \mathbb{E}(X)$ zu beschreiben.

Beispiel

Der Lektor eines Verlags stellt fest, dass eine zufällig ausgewählte Seite eines von ihm begutachteten Buches mit Wahrscheinlichkeit 0.9 keine Tippfehler enthält. Schätze die Wahrscheinlichkeit, dass eine zufällig ausgewählte Seite genau zwei Tippfehler enthält.

Sei X die Anzahl der Tippfehlern auf einer zufällig ausgewählten Seite. Jedes Zeichen auf einer Seite ist mit sehr kleiner Wahrscheinlichkeit ein Tippfehler. Die Anzahl der potentiell betroffenen Zeichen ist groß. Wir unterstellen, dass Fehler unabhängig voneinander auftreten. Daher nehmen wir $X \sim \text{Po}(\lambda)$ an. λ ist nicht direkt gegeben.

Aber:

$$\mathbb{P}(X = 0) = e^{-\lambda}\frac{\lambda^0}{0!} = 0.9, \qquad \text{also } \lambda \approx 0.105.$$

Daher

$$\mathbb{P}(X = 2) = e^{-\lambda}\frac{\lambda^2}{2!} \approx 0.005.$$

Kapitel 17: Erzeugende Funktionen

Erzeugende Funktionen sind ein analytisches Hilfsmittel zur Untersuchung von Verteilungen in der elementaren Stochastik.

Erzeugende Funktion einer Folge

Definition 17.1

Für eine reelle Zahlenfolge $a := (a_k)_{k \geq 0}$ heißt die (Summenfunktion der) Potenzreihe

$$g_a(t) := \sum_{k=0}^{\infty} a_k t^k, \qquad t \in (-\rho, \rho),$$

auch die erzeugende Funktion von a. Hierbei bezeichnet ρ den Konvergenzradius von g_a.

Bemerkungen:

(1) $\rho := \dfrac{1}{\limsup_{k\to\infty} |a_k|^{\frac{1}{k}}} \in [0, \infty]$. Nachfolgend sei $\rho > 0$.

(2) g_a ist in $(-\rho, \rho)$ beliebig oft gliedweise differenzierbar.

(3) g_a legt die Folge a fest, da $a_n = \frac{g_a^{(n)}(0)}{n!}$.

(4) Konvergiert g_a an der Stelle ρ, so ist g_a in ρ linksseitig stetig (Abelscher Grenzwertsatz).

Erzeugende Funktion einer $\mathbb{N}_0$-wertigen Zufallsvariable

Definition 17.2

Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable, das heißt $\mathbb{P}(X \in \mathbb{N}_0) = 1$. Dann nennt man

$$g_X(t) := \sum_{k=0}^{\infty} \mathbb{P}(X = k) t^k, \quad |t| \leq 1,$$

die erzeugende Funktion (der Verteilung) von X.

Bemerkungen:

- $g_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$, das heißt $\rho \geq 1$, und $|g_X(-1)| \leq 1$.
- g_X legt die Verteilung von X fest.
- $g_X(t) = \mathbb{E}(t^X) \in [-1, 1]$ für $|t| \leq 1$.

Erzeugende Funktionen bekannter Verteilungen

Beispiele

(a) Ist $X \sim \operatorname{Bin}(n, p)$ für $p \in [0,1]$ und $n \in \mathbb{N}$, so gilt

$$g_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} t^k = (1 - p + tp)^n, \quad t \in \mathbb{R}.$$

(b) Ist $X \sim \operatorname{Po}(\lambda)$ mit $\lambda \geq 0$, so gilt

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

(c) Ist $X \sim \operatorname{Nb}(r, p)$ mit $p \in (0,1)$, $r \in \mathbb{N}$, so gilt

$$\begin{aligned} g_X(t) &= \sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r (1-p)^k t^k = p^r \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} (t(1-p))^k \\ &= p^r (1 - t(1-p))^{-r} = \left(\frac{p}{1-(1-p)t} \right)^r, \quad |t| < \frac{1}{1-p}. \end{aligned}$$

Eindeutigkeitssatz

Satz 17.3

Es seien X, Y je $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugenden Funktionen g_X, g_Y. Dann gilt

$$g_X = g_Y \quad \Longleftrightarrow \quad X \sim Y \quad (\text{d.h. } \mathbb{P}(X = k) = \mathbb{P}(Y = k) \quad \forall k \in \mathbb{N}_0).$$

Multiplikationsformel

Satz 17.4

Seien X, Y stochastisch unabhängige, $\mathbb{N}_0$-wertige Zufallsvariablen. Dann gilt

$$g_{X+Y}(t) = g_X(t) \cdot g_Y(t), \quad |t| \leq 1.$$

Beweis

Für $|t| \leq 1$ gilt

$$g_{X+Y}(t) = \mathbb{E}[t^{X+Y}] = \mathbb{E}[t^X \cdot t^Y] = \mathbb{E}[t^X] \cdot \mathbb{E}[t^Y] = g_X(t) \cdot g_Y(t),$$

da mit X, Y auch t^X, t^Y stochastisch unabhängig sind für $t \in \mathbb{R}$ (Blockungslemma).

Verallgemeinerung: Sind $X_1, \ldots, X_n$ $\mathbb{N}_0$-wertige, stochastisch unabhängige Zufallsvariablen, dann gilt

$$g_{X_1+\cdots+X_n} = g_{X_1} \cdots g_{X_n}.$$

Man zeigt dies mit vollständiger Induktion.

Anwendung: Additionsgesetze

Folgerung 17.5

Aus den Beispielen und den Sätzen 17.4 und 17.3 erhält man direkt die Additionsgesetze für Binomial-, Poisson-, und negative Binomial-Verteilungen.

Seien etwa $X \sim \mathrm{Bin}(n, p)$, $Y \sim \mathrm{Bin}(m, p)$ stochastisch unabhängig. Dann gilt nach Satz 17.4

$$g_{X+Y}(t) = g_X(t) g_Y(t) = (1-p+pt)^n (1-p+pt)^m = (1-p+pt)^{n+m}$$

und nach Satz 17.3 folgt damit $X + Y \sim \mathrm{Bin}(n+m, p)$.

Erzeugende Funktionen und Momente

Satz 17.6

Seien X eine $\mathbb{N}_0$-wertige Zufallsvariable mit erzeugender Funktion g und $r \in \mathbb{N}$. Dann sind äquivalent:

(a) $\mathbb{E}[X(X-1)\cdots(X-r+1)]$ existiert in $\mathbb{R}$;

(b) $g^{(r)}(1-) := \lim_{t \nearrow 1} g^{(r)}(t)$ existiert in $\mathbb{R}$.

Falls (a) oder (b) gilt, so ist $\mathbb{E}[X(X-1)\cdots(X-r+1)] = g^{(r)}(1-)$.

Bemerkung: Die Größe $\mathbb{E}[X(X-1)\cdots(X-r+1)]$ heißt auch *r-tes faktorielles Moment* von X.

Beweis für (a) ⇔ (b)

$$(a) \iff \sum_{k=r}^{\infty} k(k-1)\cdots(k-r+1)\mathbb{P}(X=k) < \infty$$

$$\iff \underbrace{\sum_{k=r}^{\infty} k(k-1)\cdots(k-r+1)\mathbb{P}(X=k)t^{k-r}}_{=g^{(r)}(t) \text{ für } 0<t<1} \quad \text{konvergiert für } t=1 \iff \text{(b)}.$$

Hilfsaussage zum Beweis von Satz 17.6

Lemma 17.7

Sei $a_n \geq 0$ für $n \in \mathbb{N}_0$ und

$$f(t) := \sum_{n=0}^{\infty} a_n t^n, \qquad t \in [0,1].$$

Dann sind äquivalent:

(a) $f(1) < \infty$;

(b) $\lim_{t \nearrow 1} f(t) < \infty$.

Ist eine der Bedingungen (und damit beide) erfüllt, dann gilt $\lim_{t \nearrow 1} f(t) = f(1)$.

Beweis von Satz 17.6

Sei $r \in \mathbb{N}$ und X eine ZV mit $g_X = g$. Setze

$$a_{k+r} := k(k-1)\cdots(k-r+1)\mathbb{P}(X=k), k \in \mathbb{N}_0 \text{ sowie } a_k := 0, k < r.$$

Dann gilt formal $f(t) = g^{(r)}(t)$, $t \in [0,1]$ und insbes. $f(1) = g^{(r)}(1) = \mathbb{E}[X(X-1)\cdots(X-r+1)]$.

Beweis von Lemma 17.7

Beweis

(a) $\Rightarrow$ (b). Es gelte (a). Da $a_n \geq 0$ gilt, ist f monoton wachsend in $[0,1]$ und $f(t) \leq f(1) < \infty$ für $t \in [0,1]$. Dies zeigt (b).

(b) $\Rightarrow$ (a). Es gelte (b). Sei $\infty > M := \lim_{t \nearrow 1} f(t)$. Dann folgt $M \geq f(s)$ für alle $s \in [0,1)$. Für beliebige $N \in \mathbb{N}$ erhält man

$$\sum_{k=0}^{N} a_k s^k \leq f(s) \leq M, \qquad s \in [0,1),$$

und daher auch $\sum_{k=0}^{N} a_k \leq M$. Da N beliebig war, folgt

$$f(1) = \sum_{k=0}^{\infty} a_k \leq M.$$

Also gilt (a).

Beweis von Lemma 17.7

Beweis (Fortsetzung)

Jetzt mögen (a) und (b) gelten. Für $t \in [0,1)$ folgt

$$0 \le f(1) - f(t) = \sum_{k=0}^{\infty}(1-t^k)a_k = \sum_{k=0}^{N}(1-t^k)a_k + \sum_{k=N+1}^{\infty}(1-t^k)a_k.$$

Wegen $f(1) = \sum_{k=0}^{\infty} a_k < \infty$ gibt es zu $\varepsilon > 0$ ein $N \in \mathbb{N}$, so dass $\sum_{k=N+1}^{\infty} a_k < \varepsilon/2$, also

$$0 \le f(1) - f(t) \le \sum_{k=0}^{N}(1-t^k)a_k + \frac{\varepsilon}{2}.$$

Für $t \in [0,1)$ mit $t \ge t_0(\varepsilon)$ gilt

$$\sum_{k=0}^{N}(1-t^k)a_k \le \frac{\varepsilon}{2},$$

also $0 \le f(1) - f(t) \le \varepsilon$. Dies zeigt die restliche Behauptung.

Erwartungswert und Varianz m.H. erzeugender Funktionen

Folgerung

Sei $g = g_X, X$ eine $\mathbb{N}_0$-wertige Zufallsvariable, für die das erste beziehungsweise das zweite Moment existieren. Dann gelten:

- $\mathbb{E}(X) = g'(1-)$,
- $\mathbb{E}(X(X-1)) = g''(1-)$,
- $\mathbb{V}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = g''(1-) + g'(1-) - (g'(1-))^2$.

Beispiel

Sei $X \sim \mathrm{Po}(\lambda)$, also $g(t) := g_X(t) = \mathrm{e}^{\lambda(t-1)}$ für $t \in \mathbb{R}$. Für $r \in \mathbb{N}_0$ erhält man $g^{(r)}(t) = \lambda^r g(t)$, $t \in \mathbb{R}$, also

$$\mathbb{E}(X \cdots (X - r + 1)) = \lambda^r.$$

Es folgt

$$\mathbb{E}(X) = \lambda, \quad \mathbb{V}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda, \quad \mathbb{E}(X(X-1)(X-2)) = \lambda^3.$$

Kapitel 18: Allgemeine Modelle

18.1 Maßtheoretische Grundlagen

Bislang: diskreter Wahrscheinlichkeitsraum Ω und Ereignisraum $\mathcal{P}(\Omega)$.

Es ist nicht möglich, Wahrscheinlichkeiten für beliebige Teilmengen auf beliebigen Räumen Ω in konsistenter Weise zu erklären.

Es folgt daher ein maßtheoretisch begründeter Aufbau der Wtheorie.

Vorgehen ist analog zu dem in der Analysis III.

Henze, N.: Stochastik: Eine Einführung mit Grundzügen der Maßtheorie, 2024, Springer eBooks.
https://link.springer.com/book/10.1007/978-3-662-68649-2

Elstrodt, J.: Maß- und Integrationstheorie, Springer, 2018, 8. Auflage.
https://link.springer.com/book/10.1007/978-3-662-57939-8

σ-Algebren

Als σ-Algebren bezeichnete Mengensysteme sollen

- hinreichend reichhaltig sein, um die Beschreibung aller relevanten Ereignisse zu ermöglichen,
- es erlauben, erforderliche Beschränkungen vorzunehmen, um Inkonsistenzen zu vermeiden,
- es ermöglichen, den Informationsstand in einer konkreten Situation bzw. zu einer bestimmten Zeit zu modellieren.

Definition 18.1

Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt σ-Algebra über Ω, falls

(a) $\emptyset \in \mathcal{A}$,

(b) $A \in \mathcal{A} \Rightarrow A^{\mathsf{c}} \in \mathcal{A}$,

(c) $A_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Beispiel

Sei $\Omega \neq \emptyset$. Dann gilt:

(a) $\mathcal{P}(\Omega)$ und $\{\emptyset, \Omega\}$ sind σ-Algebren über Ω (die größtmögliche bzw. die kleinstmögliche).

(b) Ist $A \subset \Omega$, so ist $\{\emptyset, A, A^{\mathsf{c}}, \Omega\}$ eine σ-Algebra über Ω.

Eigenschaften von σ-Algebren

Eigenschaften

Sei $\Omega \neq \emptyset$ und $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine σ-Algebra.

- $\Omega = \emptyset^{\mathsf{c}} \in \mathcal{A}$.
- Ist $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, so auch
$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^{\mathsf{c}} \right)^{\mathsf{c}} \in \mathcal{A}.$$
- Jede σ-Algebra ist abgeschlossen bezüglich endlicher Vereinigungen und Schnitte.
- Wenn $A, B \in \mathcal{A}$, dann ist $A \setminus B = (A^{\mathsf{c}} \cup B)^{\mathsf{c}} \in \mathcal{A}$.

Ferner gilt: Sind $\mathcal{A}_i$, $i \in I$, σ-Algebren über einem gemeinsamen Grundraum Ω, so ist auch $\bigcap_{i \in I} \mathcal{A}_i$ eine σ-Algebra über dem Grundraum Ω.

Erzeugendensysteme

σ-Algebren können durch Erzeugendensysteme festgelegt werden.

Definition und Behauptung 18.2

Sei $\emptyset \neq \mathcal{M} \subset \mathcal{P}(\Omega)$. Dann existiert genau eine σ-Algebra $\sigma(\mathcal{M})$ über Ω mit

(a) $\mathcal{M} \subset \sigma(\mathcal{M})$,

(b) ist $\mathcal{A}$ eine beliebige σ-Algebra über Ω mit $\mathcal{M} \subset \mathcal{A}$, so gilt $\sigma(\mathcal{M}) \subset \mathcal{A}$.

Man nennt $\sigma(\mathcal{M})$ die von $\mathcal{M}$ erzeugte σ-Algebra, $\mathcal{M}$ wird als *ein* Erzeuger (ein Erzeugendensystem) von $\sigma(\mathcal{M})$ bezeichnet.

Beweis

Existenz:

$$\sigma(\mathcal{M}) := \bigcap\{\mathcal{A} : \mathcal{A} \subset \mathcal{P}(\Omega),\ \mathcal{A} \text{ ist } \sigma\text{-Algebra},\ \mathcal{M} \subset \mathcal{A}\}.$$

Dann ist $\sigma(\mathcal{M})$ eine σ-Algebra, die $\mathcal{M}$ enthält und nach Definition die geforderte Minimalitätseigenschaft hat. Beachte, dass $\mathcal{M} \subset \mathcal{P}(\Omega)$ und $\mathcal{P}(\Omega)$ eine σ-Algebra ist.

Eindeutigkeit: klar.

Eigenschaften erzeugter σ-Algebren

Folgerung 18.3

Für $\emptyset \neq \mathcal{M}, \mathcal{M}_1, \mathcal{M}_2 \subset \mathcal{P}(\Omega)$ gelten folgende Aussagen.

(a) $\mathcal{M}_1 \subset \mathcal{M}_2 \Rightarrow \sigma(\mathcal{M}_1) \subset \sigma(\mathcal{M}_2)$.

(b) Ist $\mathcal{A}$ eine σ-Algebra, so ist $\sigma(\mathcal{A}) = \mathcal{A}$, insbesondere gilt $\sigma(\sigma(\mathcal{M})) = \sigma(\mathcal{M})$.

(c) $\mathcal{M}_1 \subset \sigma(\mathcal{M}_2)$ und $\mathcal{M}_2 \subset \sigma(\mathcal{M}_1) \;\Leftrightarrow\; \sigma(\mathcal{M}_1) = \sigma(\mathcal{M}_2)$.

Borelsche σ-Algebra

Definition 18.4

Sei $\mathcal{O}^k$ das System der offenen Mengen des $\mathbb{R}^k$. Dann nennt man

$$\mathcal{B}^k := \mathcal{B}(\mathbb{R}^k) := \sigma(\mathcal{O}^k)$$

die Borelsche σ-Algebra (σ-Algebra der Borelmengen) des $\mathbb{R}^k$. Statt $\mathcal{B}^1$ schreiben wir nur $\mathcal{B}$.

Bemerkungen:

(1) Analog definiert man für einen topologischen Raum $(T, \mathcal{O})$ mit dem System $\mathcal{O}$ der offenen Mengen die Borelsche σ-Algebra $\mathcal{B}(T) := \sigma(\mathcal{O})$.

(2) Es gilt $\mathcal{B}^k \neq \mathcal{P}(\mathbb{R}^k)$.

Borelsche σ-Algebra

Bemerkungen:

(3) Weitere Mengensysteme: $\mathcal{A}^k$ - abgeschlossene Mengen des $\mathbb{R}^k$, $\mathcal{C}^k$ - kompakte Mengen des $\mathbb{R}^k$, $\mathcal{I}^k$ - halboffene verallgemeinerte Intervalle (Quader). Hierzu sei $x < y$ für $x := (x_1, \dots, x_k)^\top \in \mathbb{R}^k$, $y := (y_1, \dots, y_k)^\top \in \mathbb{R}^k$ erklärt durch $x_i < y_i$ für $i = 1, \dots, k$ und

$$\begin{aligned}
\mathcal{I}^k &:= \{(x, y] : x, y \in \mathbb{R}^k, x < y\}, \\
\mathcal{J}^k &:= \{(-\infty, y] : y \in \mathbb{R}^k\}, \\
(x, y] &:= \{z \in \mathbb{R}^k : x_i < z_i \leq y_i \text{ für } i = 1, \dots, k\}, \\
(-\infty, y] &:= \{z \in \mathbb{R}^k : -\infty < z_i \leq y_i \text{ für } i = 1, \dots, k\}.
\end{aligned}$$

Satz 18.5

Es gilt $\mathcal{B}^k = \sigma(\mathcal{A}^k) = \sigma(\mathcal{C}^k) = \sigma(\mathcal{I}^k) = \sigma(\mathcal{J}^k)$.

Beweis

Analysis III/Elstrodt, Satz 4.3/Henze, S.313

Messräume und Maßräume

Definition 18.6 (Messraum, Maßraum)

Ein Tripel $(\Omega, \mathcal{A}, \mu)$ heißt Maßraum, falls gilt:

(a) $\Omega \neq \emptyset, \mathcal{A} \subset \mathcal{P}(\Omega)$ ist eine σ-Algebra; das Paar $(\Omega, \mathcal{A})$ nennt man einen Messraum (messbaren Raum);

(b) $\mu : \mathcal{A} \to [0, \infty]$ ist ein Maß auf $\mathcal{A}$, das heißt

  (i) $\mu(\emptyset) = 0$,

  (ii) $\mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$ für jede Folge paarweise disjunkter Mengen $A_n \in \mathcal{A}, n \in \mathbb{N}$.

- Die Eigenschaft in (ii) wird als σ-Additivität von μ bezeichnet.
- Man nennt μ ein endliches Maß (endlich), falls $\mu(\Omega) < \infty$.
- Man nennt μ ein σ-endliches Maß (σ-endlich), falls eine aufsteigende Folge von Mengen $A_n \in \mathcal{A}, n \in \mathbb{N}$, existiert mit $\Omega = \bigcup_{n \geq 1} A_n$ und $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$.

Wahrscheinlichkeitsraum

Definition 18.7 (Wahrscheinlichkeitsraum, Axiomensystem von Kolmogorov (1933))

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$, wobei $\Omega \neq \emptyset$ eine Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine σ-Algebra über Ω ist. Ferner sei $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:

(a) $\mathbb{P}(A) \geq 0$ für alle $A \in \mathcal{A}$,

(b) $\mathbb{P}(\Omega) = 1$,

(c) $A_j \in \mathcal{A}, j \in \mathbb{N}$, paarweise disjunkt $\Rightarrow \mathbb{P}\left(\sum_{j=1}^{\infty} A_j\right) = \sum\limits_{j=1}^{\infty} \mathbb{P}(A_j)$.

Man nennt die Elemente von $\mathcal{A}$ Ereignisse und $\mathbb{P}$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{A}$.

Beachte:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset \cup \emptyset \cup \ldots) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \ldots \geq 1 + \mathbb{P}(\emptyset),$$

also $0 \leq \mathbb{P}(\emptyset) \leq 0$, d.h. $\mathbb{P}(\emptyset) = 0$.

Wahrscheinlichkeitsraum – diskret vs. allgemein

Bemerkungen:

(1) Die bisher abgeleiteten, allgemeinen Aussagen (etwa die Sätze 3.2, 3.3, 3.4, 9.2, 9.3, 9.4, 10.2, 10.3) in diskreten W.räumen bleiben sinngemäß erhalten (alle Mengen müssen in der σ-Algebra liegen), ebenso frühere Definitionen (bedingte Wahrscheinlichkeiten, Unabhängigkeit).

(2) Resultate der allgemeinen Maß- und Integrationstheorie (Analysis III) werden häufig nur im Fall von W.maßen benötigt.

(3) Wir erhalten jetzt Zugang zu nicht-diskreten Wahrscheinlichkeitsmaßen wie etwa Wahrscheinlichkeitsmaßen mit Dichten bezüglich des Lebesguemaßes (genaueres später).

Beispiele für Maße und Wahrscheinlichkeitsmaße

Beispiele

Sei $(\Omega, \mathcal{A})$ ein Messraum.

(1) Für ein beliebiges, festes $\omega \in \Omega$ sei $\delta_\omega(A) := \mathbb{1}\{\omega \in A\}$ für $A \in \mathcal{A}$. Man nennt das W.maß δ_ω Dirac-Maß oder Einpunktverteilung im Punkt ω.

(2) Sei $\Omega_0 \subset \Omega$ abzählbar und $p(\omega) \in [0, \infty)$ für $\omega \in \Omega_0$. Es gelte $\sum_{\omega \in \Omega_0} p(\omega) = 1$. Dann ist ein W.maß auf $\mathcal{A}$ gegeben durch

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

(3) Alle bisherigen Beispiele diskreter Verteilungen sind Spezialfälle von (2). So etwa $\Omega = \mathbb{R}, \mathcal{A} = \mathcal{B}^1, \Omega_0 = \{0, 1, \ldots, n\}$:

$$\mathrm{Bin}(n, p) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k, \quad p \in [0,1], n \in \mathbb{N}.$$

(4) Sei $\emptyset \neq T \subset \Omega$. Setze

$$\mu_T(A) := \begin{cases} |A \cap T|, & \text{falls } A \cap T \text{endlich}, \\ \infty, & \text{sonst}, \end{cases}$$

für $A \in \mathcal{A}$. Man nennt μ_T das T-Zählmaß auf $\mathcal{A}$.

Lebesgue-Maß

Beispiel

Sei $\Omega = \mathbb{R}^k$, $\mathcal{A} = \mathcal{B}^k$, λ^k das Lebesgue-Maß auf $\mathcal{B}^k$. Beachte, dass das Lebesgue-Maß durch

$$\lambda^k((x,y]) = \prod_{j=1}^{k} (y_j - x_j), \quad x < y,\ x, y \in \mathbb{R}^k,$$

eindeutig festgelegt wird. (→ Existenz- und Eindeutigkeitssätze für die Konstruktion von Maßen)

Dynkin-Systeme

Ein sehr nützliches Konzept für viele Argumentationen im Zusammenhang mit σ-Algebren wird in der folgenden Definition eingeführt. Typischerweise wird es verwendet, um Aussagen, die zunächst für speziellere Mengensysteme bekannt sind oder bewiesen werden können, schließlich auf den Fall allgemeinerer Mengensysteme auszudehnen.

Definition 18.8 (Dynkin-System)

Ein Mengensystem $\mathcal{D} \subset \mathcal{P}(\Omega)$ heißt Dynkin-System über Ω, falls

(a) $\Omega \in \mathcal{D}$,

(b) $D, E \in \mathcal{D},\ D \subset E \Rightarrow E \setminus D \in \mathcal{D}$,

(c) $D_1, D_2, \ldots \in \mathcal{D}$ paarweise disjunkt $\Rightarrow \bigcup_{n \geq 1} D_n \in \mathcal{D}$.

Bemerkung: (b') $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D}$. Man kann (b) durch (b') ersetzen wegen $E \setminus D = (D \cup E^c)^c$.

Beispiel

Für $\Omega = \{0, 1, 2, 3\}$, ist das Mengensystem $\{\varnothing, \{0, 1\}, \{1, 2\}, \{0, 3\}, \{2, 3\}, \{0, 1, 2, 3\}\}$ ist ein Dynkin-System, aber keine σ-Algebra.

Eigenschaften von Dynkin-Systemen

Bemerkungen: (Übung/Tutorium)

(a) Ist $\mathcal{D}$ eine σ-Algebra, so ist $\mathcal{D}$ ein Dynkin-System.

(b) Jedes $\cap$-stabile Dynkin-System ist eine σ-Algebra.

(c) Beliebige Durchschnitte von Dynkin-Systemen über Ω sind wieder ein Dynkin-System.

Definition 18.9

Sei $\emptyset \neq \mathcal{M} \subset \mathcal{P}(\Omega)$. Dann heißt

$$\delta(\mathcal{M}) := \bigcap \{\mathcal{D} : \mathcal{M} \subset \mathcal{D} \subset \mathcal{P}(\Omega),\ \mathcal{D} \text{ ist Dynkin-System}\}$$

das von $\mathcal{M}$ erzeugte Dynkin-System, $\mathcal{M}$ wird als ein Erzeuger von $\delta(\mathcal{M})$ bezeichnet.

Bemerkung: Offenbar gilt stets $\delta(\mathcal{M}) \subset \sigma(\mathcal{M})$.

Beispiel

Sei $\Omega = \{0, 1, 2, 3\}$ sowie $\mathcal{M} = \{\varnothing, \{0, 1\}, \{1, 2\}\}$. Dann gilt

$$\delta(\mathcal{M}) = \{\varnothing, \{0, 1\}, \{1, 2\}, \{0, 3\}, \{2, 3\}, \{0, 1, 2, 3\}\} \neq \sigma(\mathcal{M}).$$

Schnittstabile Dynkin-Systeme

Lemma 18.10

Ist $\emptyset \neq \mathcal{M}$ $\cap$-stabil, so gilt $\delta(\mathcal{M}) = \sigma(\mathcal{M})$.

Beweis

Es genügt zu zeigen, dass $\delta(\mathcal{M})$ $\cap$-stabil ist, denn dann ist $\delta(\mathcal{M})$ eine σ-Algebra, also $\sigma(\mathcal{M}) \subset \delta(\mathcal{M})$.

Sei hierzu $A \in \delta(\mathcal{M})$ beliebig. Setze

$$\mathcal{D}_A := \{B \subset \Omega : A \cap B \in \delta(\mathcal{M})\}.$$

Wir zeigen $\delta(\mathcal{M}) \subset \mathcal{D}_A$. Daraus folgt die Behauptung (denn für $A, B \in \delta(\mathcal{M})$ gilt dann $B \in \mathcal{D}_A$, also $A \cap B \in \delta(\mathcal{M})$, was die $\cap$-Stabilitität von $\delta(\mathcal{M})$ zeigt).

Zunächst sieht man leicht, dass $\mathcal{D}_A$ ein Dynkin-System ist. (In der Tat ist $\Omega \in \mathcal{D}_A$ und mit $D, E \in \mathcal{D}_A$, $D \subset E$ folgt $E \setminus D \in \mathcal{D}_A$. Sind $D_1, D_2, \ldots \in \mathcal{D}_A$ paarweise disjunkt, so ist $\bigcup_n D_n \in \mathcal{D}_A$.)

Ist speziell $A \in \mathcal{M}$, so gilt $\mathcal{M} \subset \mathcal{D}_A$, da $\mathcal{M}$ $\cap$-stabil ist. Da $\mathcal{D}_A$ ein Dynkin-System ist, folgt $\delta(\mathcal{M}) \subset \mathcal{D}_A$.

Ist nun $A \in \delta(\mathcal{M})$ und $B \in \mathcal{M}$, so folgt $A \cap B \in \delta(\mathcal{M})$ (wegen $\delta(\mathcal{M}) \subset \mathcal{D}_B$), das heißt $\mathcal{M} \subset \mathcal{D}_A$. Also gilt auch $\delta(\mathcal{M}) \subset \mathcal{D}_A$ für eine beliebiges $A \in \delta(\mathcal{M})$.

Der Eindeutigkeitssatz für (W.-)Maße

Eine typische Anwendung des Lemmas ist der Eindeutigkeitssatz für Maße. Wir verwenden im Beweis, dass beliebige Maße von unten stetig sind. Wahrscheinlichkeitsmaße sind auch von oben stetig (dies gilt jedoch nicht für beliebige Maße für alle Mengen), vgl. Kap. 3, Satz 3.3 und Ana III.

Satz 18.11 (Eindeutigkeitssatz für Maße)

Sei $(\Omega, \mathcal{A})$ ein Messraum. Sei $\mathcal{M} \subset \mathcal{P}(\Omega)$ ein $\cap$-stabiles Erzeugendensystem von $\mathcal{A}$ (das heißt $\sigma(\mathcal{M}) = \mathcal{A}$). Seien μ_1, μ_2 Maße auf $\mathcal{A}$ mit $\mu_1(M) = \mu_2(M)$ für alle $M \in \mathcal{M}$. Gibt es eine Folge M_n in $\mathcal{M}$ mit $M_n \nearrow \Omega$ und $\mu_1(M_n) < \infty$ für alle $n \in \mathbb{N}$, so gilt $\mu_1 = \mu_2$.

Folgerung 18.12 (Eindeutigkeitssatz für Wahrscheinlichkeitsmaße)

Stimmen zwei Wahrscheinlichkeitsmaße auf einem $\cap$-stabilen Erzeugendensystem $\mathcal{M}$ überein und gibt es eine Folge $M_n \in \mathcal{M}$, $n \in \mathbb{N}$, mit $M_n \nearrow \Omega$, so stimmen die Maße auf der von $\mathcal{M}$ erzeugten σ-Algebra überein.

Beweis von Satz 18.11

Beweis

Sei $B \in \mathcal{M}$ mit $\mu_1(B) = \mu_2(B) < \infty$. Setze

$$\mathcal{D}_B := \{A \in \mathcal{A} : \mu_1(B \cap A) = \mu_2(B \cap A)\}.$$

Dann ist $\mathcal{D}_B$ ein Dynkin-System (Tutorium!).

Nach Voraussetzung gilt $\mathcal{M} \subset \mathcal{D}_B$, also $\delta(\mathcal{M}) \subset \mathcal{D}_B$.

Da $\mathcal{M}$ $\cap$-stabil ist, gilt $\delta(\mathcal{M}) = \sigma(\mathcal{M}) = \mathcal{A} \subset \mathcal{D}_B$. Speziell erhält man $\mathcal{A} \subset \mathcal{D}_{M_n},\ n \in \mathbb{N}$.

Wegen $A \cap M_n \nearrow A,\ A \in \mathcal{A}$, liefert die Stetigkeit von unten

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A \cap M_n) = \lim_{n\to\infty} \mu_2(A \cap M_n) = \mu_2(A),$$

also die Behauptung.

Maßdefinierende Funktionen, Verteilungsfunktionen

Definition 18.13 (Maßdefinierende Funktion)

Eine Funktion $G : \mathbb{R} \to \mathbb{R}$ heißt maßdefinierend (oder maßerzeugend), falls

(a) G (schwach) monoton wächst, das heißt $x \leq y \Rightarrow G(x) \leq G(y)$,

(b) G rechtsseitig stetig ist.

Gilt zudem

(c) $G(\infty) := \lim_{x \to \infty} G(x) = 1,\ G(-\infty) := \lim_{x \to -\infty} G(x) = 0,$

so heißt G Verteilungsfunktion.

Maßdefinierende Funktionen definieren Maße

- Ist G maßerzeugend, so existiert genau ein Maß μ_G auf $\mathcal{B}$ mit

  $$\mu_G((a,b]) = G(b) - G(a) \text{ für } a, b \in \mathbb{R} \text{ mit } a \leq b.$$

  Das Maß μ_G ist σ-endlich. (Die Eindeutigkeit folgt aus Satz 18.11, die Existenz aus dem (hier nicht diskutierten) Fortsetzungssatz von Carathéodory (siehe z.B. Elstrodt, Satz 4.5, S.57 oder Henze, S.322; vgl. auch die Konstruktion des Lebesgue-Maßes in Analysis III).)
- Ist G eine Verteilungsfunktion, so ist μ_G ein Wahrscheinlichkeitsmaß. Man nennt μ_G das Lebesgue–Stieltjes-Maß zu G.
- **Umgekehrt:** Ist $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, d.h. $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{B}$, dann nennt man die durch

  $$F(x) := \mathbb{P}((-\infty, x]), \quad x \in \mathbb{R},$$

  erklärte Funktion $F : \mathbb{R} \to \mathbb{R}$ die Verteilungsfunktion zu $\mathbb{P}$. (Man weise nach, dass F (a), (b), (c) erfüllt.) Das Lebesgue–Stieltjes-Maß zu F ist gerade $\mathbb{P}$.

18.2 Vom Maß zum Integral

Ziel: Messbarkeit von Abbildungen, maßtheoretischer Zugang zur Integration
Motivation: Zufallsvariablen und u.a. deren Erwartungswerte im allg. Modell

Sprechweisen

Sei $(\Omega, \mathcal{A})$ ein Messraum. Elemente von $\mathcal{A}$ heißen $\mathcal{A}$-messbare Mengen (kurz: messbare Mengen). Wenn $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum ist, so nennen wir Elemente von $\mathcal{A}$ weiterhin Ereignisse.

Definition 18.14 (Messbare Abbildung)

Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Messräume. Eine Abbildung $f : \Omega \to \Omega'$ heißt $(\mathcal{A}, \mathcal{A}')$-messbar, falls $f^{-1}(A') \in \mathcal{A}$ für alle $A' \in \mathcal{A}'$ gilt.

Bemerkungen:

(1) Jeder Maßraum $(\Omega, \mathcal{A}, \mu)$ ist nach Definition ein Messraum.

(2) Sind $\mathcal{A}, \mathcal{A}'$ aus dem Kontext klar, spricht man lediglich von der Messbarkeit von f.

(3) In Analysis III wurde ausgiebig der Fall $(\Omega', \mathcal{A}') = (\overline{\mathbb{R}}, \mathcal{B})$ betrachtet.

Bildmaß

Definition und Behauptung 18.15

Sind $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\Omega', \mathcal{A}')$ ein Messraum und ist die Abbildung $f : (\Omega, \mathcal{A}) \to (\Omega, \mathcal{A}')$ messbar, dann wird durch

$$\mu^f(A') := \mu(f^{-1}(A')), \quad A' \in \mathcal{A}'$$

ein Maß $\mu^f : \mathcal{A}' \to [0, \infty]$ auf $\mathcal{A}'$ erklärt. Es heißt das Bildmaß von μ unter der Abbbildung f.

Gebräuchliche Notationen: $\mu^f = \mu \circ f^{-1} = f(\mu) = f_\sharp \mu$.

Man vergewissere sich, dass μ^f tatsächlich ein Maß ist.

Beispiel

Seien $(\Omega, \mathcal{A}) = (\Omega', \mathcal{A}') = (\mathbb{R}^k, \mathcal{B}^k)$, $\mu = \lambda^k$ und $f : \mathbb{R}^k \to \mathbb{R}^k$ eine Isometrie, das heißt $f(x) := Ux + b$ für $x \in \mathbb{R}^k$, wobei $U \in O(k)$, $b \in \mathbb{R}^k$. Dann gilt $f(\lambda^k) = \lambda^k$. (Vgl. Analysis III)

Induzierte σ-Algebren

Für eine Abbildung $f : \Omega \to \Omega'$ und $\mathcal{M}' \subset \mathcal{P}(\Omega')$ erklären wir

$$f^{-1}(\mathcal{M}') := \{f^{-1}(A') : A' \in \mathcal{M}'\}.$$

Satz 18.16

Sei $f : \Omega \to \Omega'$ eine Abbildung. Dann gelten:

(a) Ist $\mathcal{A}'$ eine σ-Algebra über Ω', so ist $f^{-1}(\mathcal{A}')$ eine σ-Algebra über Ω.

(b) Ist $\mathcal{A}$ eine σ-Algebra über Ω, so ist $\mathcal{A}_f := \{A' \subset \Omega' : f^{-1}(A') \in \mathcal{A}\}$ eine σ-Algebra über Ω'.

(c) Wird $\mathcal{A}'$ von $\mathcal{M}'$ erzeugt, so wird $f^{-1}(\mathcal{A}')$ von $f^{-1}(\mathcal{M}')$ erzeugt:

$$\mathcal{A}' = \sigma(\mathcal{M}') \Rightarrow f^{-1}(\mathcal{A}') = \sigma(f^{-1}(\mathcal{M}')).$$

Bemerkungen:

- Aussage (c) ist häufig nützlich: Um die Messbarkeit einer Abbildung $f : \Omega \to \Omega'$ bezüglich der σ-Algebren $\mathcal{A}$ auf Ω und $\mathcal{A}' = \sigma(\mathcal{M}')$ auf Ω' zu zeigen, g.z.z.: $f^{-1}(\mathcal{M}') \subset \mathcal{A}$. Dann folgt
$$f^{-1}(\mathcal{A}') = f^{-1}(\sigma(\mathcal{M}')) = \sigma(f^{-1}(\mathcal{M}')) \subset \sigma(\mathcal{A}) = \mathcal{A}.$$
- Aussage (a) kann genutzt werden, um geeignete σ-Algebren zu konstruieren (kleinste σ-Algebren, so dass vorgeg. Abb. messbar werden).

Beweis von Satz 18.16

Beweis

(a) und (b) sind leicht zu sehen.
(c) Klar ist $\sigma(f^{-1}(\mathcal{M}')) \subset f^{-1}(\mathcal{A}')$ wegen (a). Setze nun

$$\mathcal{F} := \{F \subset \Omega' : f^{-1}(F) \in \sigma(f^{-1}(\mathcal{M}'))\}.$$

Dann ist $\mathcal{F}$ eine σ-Algebra (nach (b)) und $\mathcal{M}' \subset \mathcal{F}$, also $\sigma(\mathcal{M}') \subset \mathcal{F}$. Folglich ist

$$f^{-1}(\sigma(\mathcal{M}')) \subset f^{-1}(\mathcal{F}) \subset \sigma(f^{-1}(\mathcal{M}')).$$

Reellwertige und numerische Funktionen

Häufige Situation: $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$ oder $\Omega' = \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ mit

$$\mathcal{A}' = \overline{\mathcal{B}} := \{B \cup E : B \in \mathcal{B}, E \subset \{-\infty, \infty\}\}.$$

Definition 18.17

Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine $(\mathcal{A}, \mathcal{B})$-messbare Funktion $f : \Omega \to \mathbb{R}$ heißt (Borel-)messbar. Eine Funktion $f : \Omega \to \overline{\mathbb{R}}$ heißt numerische Funktion. Messbarkeit für numerische Funktionen meint $(\mathcal{A}, \overline{\mathcal{B}})$-Messbarkeit.

Das Konzept der Messbarkeit für reellwertige und numerische Funktionen und deren Verknüpfungen wurde in der Analysis III ausgiebig behandelt.

Wir verwenden im Folgenden, dass alle betrachteten Verknüpfungen messbarer Funktionen oder Ereignisse, die auf Relationen zwischen messbaren Funktionen beruhen, messbar sind.

Zufallsvariablen und deren Verteilung

Definition 18.18

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ ein Messraum.

- Eine (Ω'-wertige) Zufallsvariable (oder ein Ω'-wertiges Zufallselement) ist eine $(\mathcal{A}, \mathcal{A}')$-messbare Abbildung $X : \Omega \to \Omega'$.
- Das Bildmaß $\mathbb{P}^X = \mathbb{P}_X = X(\mathbb{P}) = \mathbb{P} \circ X^{-1}$ von $\mathbb{P}$ unter X heißt Verteilung von X unter $\mathbb{P}$, also

$$\mathbb{P}^X(A') = \mathbb{P}(X^{-1}(A')) = \mathbb{P}(X \in A'), \quad A' \in \mathcal{A}'.$$

- Im Fall $(\Omega', \mathcal{A}') = (\mathbb{R}^k, \mathcal{B}^k)$ heißt $X = (X_1, \ldots, X_k)$ k-dimensionaler Zufallsvektor mit den Komponenten $X_1, \ldots, X_k$, man bezeichnet $\mathbb{P}^X$ als die gemeinsame Verteilung von $X_1, \ldots, X_k$.
- Für $k = 1$ heißt X reellwertige Zufallsvariable. Für $(\Omega', \mathcal{A}') = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ spricht man manchmal von numerischen Zufallsvariablen.

Integrale in Maßräumen

Zu einem gegebenen Maß μ eines Maßraumes $(\Omega, \mathcal{A}, \mu)$ kann man das Integral von messbaren, numerischen Funktionen $f : \Omega \to \overline{\mathbb{R}}$ erklären, das heißt eine Integration bezüglich μ.

Ziel ist es,

$$\int_\Omega f \, d\mu = \int_\Omega f(\omega)\, \mu(d\omega) \qquad (\text{ das } \mu\text{-Integral von } f \text{ über } \Omega \,)$$

in geeigneter Weise zu konstruieren.

Man geht dabei aus von $f = \mathbb{1}_A$ mit $A \in \mathcal{A}$ und erklärt zunächst

$$\int_\Omega \mathbb{1}_A \, d\mu := \mu(A) \in [0, \infty].$$

Das weitere Vorgehen besteht darin, den Rahmen schrittweise zu erweitern!

Definition des Maß-Integrals

Schritt 1. $f \in \mathcal{E} := \{f : \Omega \to \mathbb{R} : f \geq 0, f \text{ messbar}, \ |f(\Omega)| < \infty\}$.

Man nennt Funktionen in $\mathcal{E}$ Elementarfunktionen.

Dann gibt es zu f ein $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}_{\geq 0}$, paarweise disjunkte Mengen $A_1, \ldots, A_n \in \mathcal{A}$ mit $A_1 + \cdots + A_n = \Omega$, so dass

$$f = \sum_{i=1}^{n} \alpha_i \, \mathbb{1}_{A_i} \, .$$

Jede solche Darstellung einer Elementarfunktion wird als eine Normaldarstellung von f bezeichnet. Diese ist durch f jedoch nicht eindeutig festgelegt. Wir erklären nun

$$\int_\Omega f \, d\mu := \sum_{j=1}^{n} \alpha_j \mu(A_j).$$

<u>Nachweis</u>: Wohldefiniertheit (siehe Analysis III)

Definition des Maß-Integrals

Schritt 2. Sei $f \in \mathcal{E}^* := \{f : \Omega \to \overline{\mathbb{R}} : f \geq 0, f \text{ messbar}\}$.

Zu $f \in \mathcal{E}^*$ existiert (siehe Analysis III) eine Folge $(u_n)_{n\in\mathbb{N}}$ mit $u_n \in \mathcal{E}$ und $u_n \nearrow f$.

Definiere

$$\int_\Omega f \, d\mu := \lim_{n\to\infty} \int_\Omega u_n \, d\mu \in [0, \infty].$$

Nachweis: Wohldefiniertheit (siehe Analysis III)

Definition des Maß-Integrals

Schritt 3. Sei $f : \Omega \to \overline{\mathbb{R}}$ messbar. Dann ist $f = f^+ - f^-$ mit $f^+ := \max\{f, 0\}$ und $f^- := -\min\{f, 0\}$. Falls

$$\int_\Omega f^+ \, d\mu < \infty \quad \text{und} \quad \int_\Omega f^- \, d\mu < \infty,$$

so sagt man, f sei μ-integrierbar und erklärt

$$\int_\Omega f \, d\mu := \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu \in \mathbb{R}.$$

Ist lediglich

$$\int_\Omega f^+ \, d\mu < \infty \quad \text{oder} \quad \int_\Omega f^- \, d\mu < \infty,$$

so erklärt man

$$\int_\Omega f \, d\mu := \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu \in \overline{\mathbb{R}}$$

und nennt f quasi-integrierbar.

Definition des Maß-Integrals

Bemerkungen:

(a) Im Folgenden geben wir den Integrationsbereich unter dem Integralzeichen oft nicht an, wenn dieser im Kontext klar ist. In einem Maßraum $(\Omega, \mathcal{A}, \mu)$ und für eine messbare Funktion f schreiben wir also kurz

$$\int f \; d\mu \qquad \text{statt} \qquad \int_\Omega f \; d\mu.$$

(b) Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $X(\Omega) = \{x_1, \ldots, x_n\}$, wobei $x_1, \ldots, x_n$ paarweise verschieden seien. Setze $A_j := \{X = x_j\}$. Dann gilt

$$\int_\Omega X \; d\mathbb{P} = \sum_{j=1}^{n} x_j \mathbb{P}(A_j) = \sum_{j=1}^{n} x_j \mathbb{P}(X = x_j) = \mathbb{E}(X).$$

Eigenschaften des Maß-Integrals

Satz 18.19

Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f, g integrierbar. Dann gelten:

(a) $$\int_\Omega (\alpha f + \beta g)\,\mathrm{d}\mu \;=\; \alpha \int_\Omega f\,\mathrm{d}\mu + \beta \int_\Omega g\,\mathrm{d}\mu, \quad \alpha, \beta \in \mathbb{R}, \qquad \text{(Linearität)}$$

(b) $$f \leq g \implies \int_\Omega f\,\mathrm{d}\mu \leq \int_\Omega g\,\mathrm{d}\mu, \qquad \text{(Monotonie)}$$

(c) $$\left| \int_\Omega f\,\mathrm{d}\mu \right| \leq \int_\Omega |f|\,\mathrm{d}\mu, \qquad \text{(Dreiecksungleichung)}$$

(d) f μ-integrierbar $\iff \displaystyle\int_\Omega |f|\,\mathrm{d}\mu < \infty$.

Die Strategie zum Beweis von Satz 18.19 besteht darin, die Aussagen zunächst für Elementarfunktionen zu zeigen. Dazu betrachtet man Normaldarstellungen der involvierten Funktionen sowie gemeinsame Verfeinerungen dieser Normaldarstellungen. Die schrittweise Erweiterung ist dann (eher) eine Routineaufgabe.

Algebraische Induktion

Soll eine Aussage E für messbare Funktionen bewiesen werden, so geht man häufig wie folgt vor und zeigt der Reihe nach:

(a) E gilt für Elementarfunktionen, insbesondere für $f = \mathbf{1}_A$.

(b) E gilt für nichtnegative messbare Funktionen f mittels Approximation $u_n \uparrow f$.

(c) E gilt allgemein mittels der Zerlegung $f = f^+ - f^-$.

Man nennt diese Vorgehensweise auch Beweisprinzip der **algebraischen Induktion** (oder manchmal auch **Lebesguesche Leiter**).

Diracmaß

Beispiel

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sei δ_x das Diracmaß in $x \in \Omega$. Sei $f : \Omega \to \mathbb{R}$ eine messbare Funktion. Dann gilt

$$\int_\Omega f \, d\delta_x = \int_\Omega f(\omega) \, \delta_x(d\omega) = f(x).$$

Erwartungswert einer Zufallsvariablen

Wir sind nun in der Lage, den Erwartungswert einer Zufallsvariable in einem allgemeinen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ zu erklären.

Definition 18.20

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \overline{\mathbb{R}}$ eine Zufallsvariable.
Man sagt, der Erwartungswert von X existiert, genau dann, wenn gilt

$$\int_\Omega |X| \,\mathrm{d}\mathbb{P} < \infty.$$

In diesem Fall nennt man

$$\mathbb{E}(X) := \int_\Omega X \,\mathrm{d}\mathbb{P} = \int_\Omega X(\omega)\,\mathbb{P}(\mathrm{d}\omega)$$

den Erwartungswert von X.

Eigenschaften des Erwartungswerts

Die Eigenschaften des Integrals übertragen sich nun sofort in die Sprache der Erwartungswerte von Zufallsvariablen.

Satz 18.21

Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und X, Y Zufallsvariablen mit $\mathbb{E}|X| < \infty$ und $\mathbb{E}|Y| < \infty$. Dann gelten:

(a) $\mathbb{E}\left(\alpha X + \beta Y\right) \;=\; \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y), \quad \alpha, \beta \in \mathbb{R},$ (Linearität),

(b) $X \leq Y \implies \mathbb{E}(X) \leq \mathbb{E}(Y),$ (Monotonie)

(c) $\left|\mathbb{E}(X)\right| \leq \mathbb{E}|X|.$ (Dreiecksungleichung)

Satz von der monotonen Konvergenz

Neben dem Satz von der dominierten Konvergenz (Analysis III, W.theorie) ist insbesondere der Satz von der monotonen Konvergenz häufig hilfreich.

Er ergibt sich in Verallgemeinerung der Konstruktion des Integrals einer nicht-negativen Funktion.

Satz 18.22 (monotone Konvergenz, Beppo Levi, MON)

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und seien $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$ messbare Funktionen. Falls $f_n \geq 0$ μ-fast überall und $f_n \leq f_{n+1}$ μ-fast überall für $n \geq 1$, so gilt

$$\int_\Omega \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu.$$

Beweis: Analysis III

Summen von Maßen

Satz 18.23

Seien $\mu_1, \mu_2, \dots$ Maße auf $(\Omega, \mathcal{A})$. Dann ist $\mu := \sum_{k \geq 1} \mu_k$ ein Maß auf $(\Omega, \mathcal{A})$. Sei $f : \Omega \to \overline{\mathbb{R}}$ eine messbare Funktion. Dann gilt:

$$f \text{ ist } \mu\text{-integrierbar} \iff \sum_{k \geq 1} \int |f| \, \mathrm{d}\mu_k < \infty.$$

In diesem Fall gilt

$$\int f \, \mathrm{d}\mu = \sum_{k \geq 1} \int f \, \mathrm{d}\mu_k.$$

Beweis

Algebraische Induktion. (siehe z.B. Henze, Beispiel auf S.342) Schritt 1: Beh. gilt für Elementarfunktionen (großer Umordnungssatz). Schritt 2: Für eine nichtneg. numerische Funktionen f approximiere f durch eine isotone Folge von Elementarfunktionen. ... u.s.w.

Transformationsformel

Satz 18.24 (messbare Transformation)

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\Omega', \mathcal{A}')$ ein Messraum und $f : \Omega \to \Omega'$ sei $(\mathcal{A}, \mathcal{A}')$-messbar. Es bezeichne $f(\mu)$ das Bildmaß von μ unter f. Sei $h : \Omega' \to \overline{\mathbb{R}}$ eine $(\mathcal{A}', \mathcal{B})$-messbare numerische Funktion.

(a) Ist h nichtnegativ, so gilt

$$\int_{\Omega'} h \, df(\mu) = \int_{\Omega} h \circ f \, d\mu. \tag{$\star$}$$

(b) Genau dann ist h eine $f(\mu)$-integrierbare numerische Funktion, wenn $h \circ f$ eine μ-integrierbare numerische Funktion ist. In diesem Fall gilt ebenfalls $(\star)$.

Beweis

Für $h = \mathbb{1}_{A'}$ mit $A' \in \mathcal{A}'$ gilt

$$\int_{\Omega'} \mathbb{1}_{A'} \, df(\mu) = f(\mu)(A') = \mu(f^{-1}(A')) = \int_{\Omega} \mathbb{1}_{f^{-1}(A')} \, d\mu = \int_{\Omega} \mathbb{1}_{A'} \circ f \, d\mu.$$

Jetzt: algebraische Induktion (Linearität des Integrals und monotone Konvergenz).

Erwartungswert und Verteilung

Folgerung

Für eine Zufallsvariable X auf einem beliebigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ hängt der Erwartungswert $\mathbb{E}(X)$ nur von der Verteilung $\mathbb{P}^X$ von X ab.

Beweis

Wende Satz 18.24 an auf $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$ und $h : \mathbb{R} \to \mathbb{R}, h(t) := t$, d.h. $h = id_\mathbb{R}$. Zunächst gilt:

$$\begin{aligned} \mathbb{E}(X) \text{ existiert} &\Leftrightarrow \int_\Omega |X| d\mathbb{P} < \infty && (X \text{ ist } \mathbb{P}\text{-integrierbar}) \\ &\Leftrightarrow \int_\Omega |id_\mathbb{R} \circ X| d\mathbb{P} < \infty && (id_\mathbb{R} \circ X \text{ ist } \mathbb{P}\text{-integrierbar}) \\ &\Leftrightarrow \int_\mathbb{R} |id_\mathbb{R}| d \underbrace{\mathbb{P}^X}_{X(\mathbb{P})} < \infty && (id_\mathbb{R} \text{ ist } \mathbb{P}^X\text{-integrierbar}) \\ &\Leftrightarrow \int_\mathbb{R} |x| \mathbb{P}^X(dx) < \infty \end{aligned}$$

In diesem Fall gilt $\mathbb{E}(X) = \int_\Omega X \; d\mathbb{P} = \int_\mathbb{R} x \; \mathbb{P}^X(dx)$, d.h., $\mathbb{E}(X)$ hängt nur von $\mathbb{P}^X$ ab.

Fast sichere Aussagen

Definition 18.25

Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und für jedes $\omega \in \Omega$ sei $E(\omega)$ eine Aussage, die gilt (wahr ist) oder nicht.

- Wir sagen, E gilt μ-fast überall, wenn es eine Menge $N \in \mathcal{A}$ gibt mit $\mu(N) = 0$ und so, dass $E(\omega)$ für alle $\omega \in \Omega \setminus N$ wahr ist.
- Falls $\mu = \mathbb{P}$ ein Wahrscheinlichkeitsmaß ist, sagen wir, E gilt $\mathbb{P}$-fast sicher, wenn E im obigen Sinn $\mathbb{P}$-fast überall gilt (das heißt, es gibt $A \in \mathcal{A}$ mit $E(\omega)$ ist wahr für alle $\omega \in A$ und $\mathbb{P}(A) = 1$).

Beispiel

Seien $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}), f(x) = x^2, x \in \mathbb{R}$. Dann gilt

- $f \neq 0$ λ^1-fast überall,
- aber $f = 0$ gilt δ_0-fast überall.

Das zugrunde liegende Maß ist also von entscheidender Bedeutung.

Markov-Ungleichung

Satz 18.26

Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : \Omega \to \overline{\mathbb{R}}_{\geq 0} = [0, \infty]$ eine $\mathcal{A}$-messbare Abbildung. Dann gilt

$$\mu(f \geq t) \;\leq\; \frac{1}{t} \int_\Omega f \, d\mu, \qquad t > 0,$$

wobei $\mu(f \geq t) = \mu(\{\omega \in \Omega : f(\omega) \geq t\})$.

Beweis

Sei $t > 0$. Für jedes Argument gilt $\mathbb{1}\{f \geq t\} \;\leq\; \frac{f}{t}$. Also

$$\mu(f \geq t) = \int_\Omega \mathbb{1}\{f \geq t\} \, d\mu \;\leq\; \int_\Omega \frac{1}{t} f \, d\mu = \frac{1}{t} \int_\Omega f \, d\mu,$$

was die Behauptung beweist.

Markov-Ungleichung

Bemerkungen:

- In der üblichen Notation der Stochastik ($\mu = \mathbb{P}$, $f = X$) lautet die Markov-Ungleichung folgendermaßen: Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und X eine nichtnegative, numerische Zufallsvariable. Dann gilt für $t > 0$

$$\mathbb{P}(X \geq t) \; \leq \; \frac{1}{t} \int_\Omega X \, d\mathbb{P} = \frac{1}{t} \mathbb{E}(X).$$

- Mit der speziellen Wahl $\mu = \mathbb{P}$, $f(\omega) = |X(\omega) - \mathbb{E}(X)|^2$, $t = \varepsilon^2$, $\varepsilon > 0$ ergibt sich

$$\mathbb{P}(|X(\omega) - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \int_\Omega |X - \mathbb{E}(X)|^2 \, d\mathbb{P} = \frac{1}{\varepsilon^2} \mathbb{V}(X),$$

  also die Tschebyscheff-Ungleichung (vgl. Kap. 12) als Folgerung der Markov-Ungleichung.

Integrale über Nullmengen

Lemma A

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Ist f $\mathcal{A}$-messbar und $N \in \mathcal{A}$ mit $\mu(N) = 0$, so ist $|f|\,\mathbb{1}_N$ $\mathcal{A}$-messbar und es gilt

$$\int |f|\,\mathbb{1}_N \, d\mu = 0.$$

Beweis

Messbarkeit ist klar. Ist f reellwertig, so gilt

$$\begin{aligned} 0 &\leq \int f^{\pm}\,\mathbb{1}_N \, d\mu \leq \int f^{\pm}\,\mathbb{1}_N \sum_{n\geq 1} \mathbb{1}\{f^{\pm} \leq n\} \, d\mu = \sum_{n\geq 1} \int f^{\pm}\,\mathbb{1}\{N \cap \{f^{\pm} \leq n\}\} \, d\mu \\ &\leq \sum_{n\geq 1} \int n\,\mathbb{1}\{N \cap \{f^{\pm} \leq n\}\} \, d\mu = \sum_{n\geq 1} n\mu(N \cap \{f^{\pm} \leq n\}) \leq \sum_{n\geq 1} n\mu(N) = 0. \end{aligned}$$

Ist f eine numerische Funktion, so beachte, dass $f = f^+ - f^-$ und $\int \infty\,\mathbb{1}_N \, d\mu = 0$ nach Definition des Integrals (Schritt 2) gilt.

Nichtnegativität

Lemma B

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sei f $\mathcal{A}$-messbar. Ist $f \geq 0$ μ-fast überall, so ist $\int f \, d\mu \geq 0$.

Beweis

Mit Lemma A gilt

$$\begin{aligned} \int f \, d\mu &= \int f \cdot \mathbb{1}\{f \geq 0\} + f \cdot \mathbb{1}\{f < 0\} \, d\mu \\ &= \int f \cdot \mathbb{1}\{f \geq 0\} \, d\mu + \int f \cdot \mathbb{1}\{f < 0\} \, d\mu = \int f \cdot \mathbb{1}\{f \geq 0\} \, d\mu \geq 0, \end{aligned}$$

da $\mu(f < 0) = 0$ nach Voraussetzung, $f \cdot \mathbb{1}\{f \geq 0\} \geq 0$ und wegen der Additivität und Monotonie des Integrals.

Stochastische Interpretation: Gilt $X \geq 0$ $\mathbb{P}$-f.s., so folgt $\mathbb{E}X \geq 0$.

Nullfunktionen

Lemma C

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sei f $\mathcal{A}$-messbar. Ist $f \geq 0$ μ-fast überall, so gilt

$$\int_\Omega f\, d\mu = 0 \quad \Longleftrightarrow \quad f = 0 \ \mu\text{-fast überall.}$$

Beweis

“$\Rightarrow$”: Mit der Markov-Ungleichung (Satz 18.26) folgt

$$0 \leq \mu(f > 0) \leq \sum_{n \geq 1} \mu(f \geq 1/n) \leq \sum_{n \geq 1} n \int f\, d\mu = 0.$$

“$\Leftarrow$”: Nach Vor. gilt $f \geq 0$ μ-f.ü., also ist wegen Lemma B $\int f\, d\mu \geq 0$.
Nun gilt aber auch $-f \geq 0$ μ-f.ü., also $-\int f\, d\mu = \int (-f)\, d\mu \geq 0$ wegen Lemma B. Zusammen folgt die Behauptung.

Stochastische Interpretation: Für eine nichtneg. ZV X gilt: $\mathbb{E}X = 0$ gdw. $X = 0$ $\mathbb{P}$-f.s..

Endlichkeit fast überall

Lemma D

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Sei f $\mathcal{A}$-messbar. Ist $\int_\Omega |f| \, d\mu < \infty$, so gilt $|f| < \infty$ μ-fast überall.

Beweis

Mit der Markov-Ungleichung (Satz 18.26) folgt

$$\mu(|f| = \infty) = \mu\left(\bigcap_{n \geq 1} \{|f| \geq n\}\right) \leq \mu(|f| \geq n) \leq \frac{1}{n} \int |f| \, d\mu.$$

Da das Integral endlich ist und $n \geq 1$ beliebig, folgt die Behauptung.

Stochastische Interpretation: Existiert der Erwartungswert von X, so gilt $|X| < \infty$ $\mathbb{P}$-fast sicher.

Nullmengen-Unempfindlichkeit des Integrals

Lemma E

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Seien $f, g : \Omega \to \overline{\mathbb{R}}$ messbare Funktionen mit $f = g$ μ-fast überall. Dann gilt

$$f \text{ ist } \mu\text{-integrierbar} \iff g \text{ ist } \mu\text{-integrierbar.}$$

In diesem Fall ist ferner

$$\int_\Omega f \, d\mu = \int_\Omega g \, d\mu.$$

Beweis

Die Vor. liefert, dass auch $f^\pm = g^\pm$ μ-fast überall gilt. Daher können wir o.B.d.A. $f, g \geq 0$ annehmen. Setze $G := \{f = g\}$, $N := G^c$, also $\mu(N) = 0$. Nun folgt mit Lemma C und der Additivität des Integrals:

$$\begin{aligned} \int f \, d\mu &= \int f \mathbb{1}_G + f \mathbb{1}_N \, d\mu = \int f \mathbb{1}_G \, d\mu + \int f \mathbb{1}_N \, d\mu \\ &= \int f \mathbb{1}_G \, d\mu = \int g \mathbb{1}_G \, d\mu = \cdots = \int g \, d\mu, \end{aligned}$$

woraus alle Beh. folgen.

Integral über Teilmengen

Definition 18.27

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Maßraum. Sei $f : \Omega \to \overline{\mathbb{R}}$ messbar und $f \geq 0$ oder aber f eine μ-integrierbare numerische Funktion. Für $A \in \mathcal{A}$ heißt

$$\int_A f \, d\mu := \int_\Omega f \cdot \mathbb{1}_A \, d\mu$$

das μ-Integral von f über A.

Maße mit Dichten

Definition und Satz 18.28

Es seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : \Omega \to [0, \infty]$ messbar.

(1) Auf $\mathcal{A}$ ist ein Maß erklärt durch

$$\nu(A) := \int_A f \, d\mu, \qquad A \in \mathcal{A}.$$

Es heißt Maß mit der Dichte f bezüglich μ und wird mit $\nu =: f\mu$ bezeichnet. Im Fall $\int_\Omega f \, d\mu = 1$ ist ν ein Wahrscheinlichkeitsmaß und man nennt f eine Wahrscheinlichkeitsdichte.

(2) Falls $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^k, \mathcal{B}^k, \lambda^k)$, so heißt f (Lebesgue-)Dichte von ν.
Falls $\nu = \mathbb{P}^X$ für einen k-dimensionalen Zufallsvektor $X = (X_1, \ldots, X_k)$ gilt, so heißt f μ-Dichte von X oder gemeinsame μ-Dichte von $X_1, \ldots, X_k$.

Beweis: Seien $A_1, A_2, \ldots \in \mathcal{A}$ paarweise disjunkt. Dann gilt nach Satz 18.22 (MON)

$$\nu\left(\sum_j A_j\right) = \int f \cdot \mathbb{1}_{\sum_j A_j} \, d\mu = \int \sum_j \left(f \cdot \mathbb{1}_{A_j}\right) d\mu = \sum_j \nu(A_j). \quad \square$$

Bemerkung: Sind f, g zwei μ-Dichten von ν (d.h. Dichten von ν bezüglich μ), so gilt $f = g$ μ-fast überall. Dies kann man mit Hilfe von Lemma C einsehen.

Stetige Zufallsvariablen

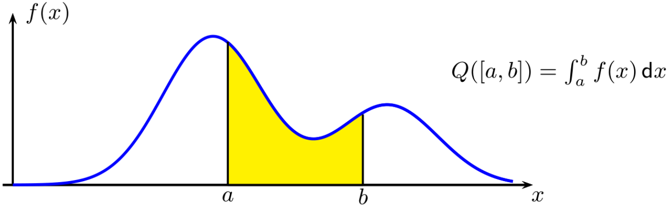
Definition 18.29 (Spezialfall 1: absolut stetige Verteilung)

Ein k-dimensionaler Zufallsvektor $X = (X_1, \ldots, X_k)$ (auf einem W.raum $(\Omega, \mathcal{A}, \mathbb{P})$) heißt (absolut) stetig (verteilt), falls X eine Lebesgue-Dichte besitzt, falls also eine messbare Funktion $f : \mathbb{R}^k \to [0, \infty)$ existiert mit

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \int_B f(x)\ \lambda^k(dx), \qquad B \in \mathcal{B}^k.$$

Insbesondere spricht man im Fall $k = 1$ von einer stetigen Zufallsvariablen.

Bemerkung: Im Fall $k = 1$ kann die Wahrscheinlichkeit $\mathbb{P}(X \in B)$ als Fläche zwischen dem Graphen von f und der x-Achse interpretiert werden, für $k > 1$ entsprechend.



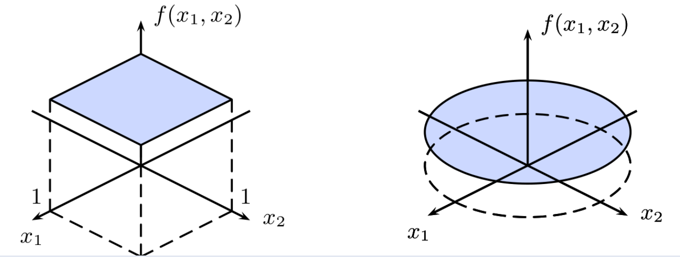
Stetige Gleichverteilung

Beispiel

Sei $B \subset \mathbb{R}^k$ eine beschränkte Borelmenge mit $\lambda^k(B) > 0$. Setze

$$f(x) := \frac{1}{\lambda^k(B)} \cdot \mathbb{1}_B(x), \qquad x \in \mathbb{R}^k.$$

Dann ist f eine Wahrscheinlichkeitsdichte bezüglich λ^k, da $\int_{\mathbb{R}^k} f(x)\, \lambda^k(dx) = 1$. Ein Zufallsvektor X mit dieser Dichte f bezüglich λ^k heißt gleichverteilt auf B, kurz $X \sim \mathrm{U}_B$.



Gleichverteilung auf einen Quadrat bzw. einer Kreisscheibe

Spezialfall: $k = 1$, $B = [0, 1]$: $\mathrm{U}_{[0,1]}$ Gleichverteilung auf dem Einheitsintervall

Diskrete Zufallsvariablen

Beispiel (Spezialfall 2: diskrete Verteilung als Maß mit Dichte)

Ein k-dimensionaler Zufallsvektor $X = (X_1, \dots, X_k)$ (auf einem W.raum $(\Omega, \mathcal{A}, \mathbb{P})$) heißt (bekanntlich) diskret (verteilt), falls es eine abzählbare Menge $D \subset \mathbb{R}^k$ mit $\mathbb{P}(X \in D) = 1$ gibt.

Ist μ_D das Zählmaß auf D , d.h., $\mu_D := \sum\limits_{y \in D} \delta_y$, und setzt man

$$f(x) := \begin{cases} \mathbb{P}(X = x), & x \in D, \\ 0, & x \in \mathbb{R}^k \setminus D, \end{cases}$$

so gilt

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \int_B f(x)\ \mu_D(dx) = \int_B f\ d\mu_D, \qquad B \in \mathcal{B}^k.$$

In der Tat folgt nach Satz 18.23

$$\int_B f\ d\mu_D = \sum_{y \in D} f(y)\, \mathbb{1}_B(y) = \sum_{y \in B \cap D} f(y) = \sum_{y \in B \cap D} \mathbb{P}(X = y) = \mathbb{P}(X \in B \cap D) = \mathbb{P}(X \in B),$$

Die Zähldichte f von X ist also eine Dichte von X bzgl. des Maßes μ_D.

Diskrete Gleichverteilung

Beispiel

In einem W.raum $(\mathbb{R}^k, \mathcal{B}^k, \mathbb{P})$ sei $\emptyset \neq D \subset \mathbb{R}^k$ eine endliche Menge und μ_D das Zählmaß auf D, d.h.

$$\mu_D = \sum_{x \in D} \delta_x .$$

Dann ist die Gleichverteilung U_D auf D erklärt als das Maß

$$\mathrm{U}_D(A) := \frac{\mu_D(A)}{\mu_D(\Omega)} = \frac{1}{|D|}\,\mu_D(A) = \int_A \frac{1}{|D|}\,d\mu_D = \frac{|A \cap D|}{|D|}, \quad A \in \mathcal{B}^k .$$

In diesem Fall hat U_D die konstante Dichte $f(\omega) = 1/|D|$ bezüglich μ_D.

Integration bezüglich $\nu = f\mu$

Satz 18.30

Seien $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\nu = f\mu$ das Maß mit Dichte $f \geq 0$ bezüglich μ. Dann gelten:

(a) Ist $\varphi : \Omega \to \overline{\mathbb{R}}_{\geq 0}$ messbar, so gilt

$$\int_\Omega \varphi \, d\nu = \int_\Omega \varphi \cdot f \, d\mu. \tag{$\star$}$$

(b) Ist $\varphi : \Omega \to \overline{\mathbb{R}}$ messbar, so gilt:

$$\varphi \text{ ist } \nu\text{-integrierbar} \Leftrightarrow \varphi \cdot f \text{ ist } \mu\text{-integrierbar}.$$

In diesem Fall gilt ebenfalls $(\star)$.

Beweis

(a) Ist $\varphi = \mathbb{1}_A, A \in \mathcal{A}$, so gilt $(\star)$ nach Definition, denn $\int \mathbb{1}_A \, d\nu = \nu(A) = \int \mathbb{1}_A \cdot f \, d\mu$. Für Elementarfunktionen φ folgt $(\star)$ dann wegen der Linearität des Integrals. Der allgemeine Fall folgt nach dem Satz von der monotonen Konvergenz.

(b) Anwendung von (a) auf φ^+ und φ^-.

Berechnung des Erwartungswerts

Folgerung 18.31

Die reelle Zufallsvariable X besitze die μ-Dichte f. Dann gilt:

Der Erwartungswert von X existiert genau dann, wenn $\displaystyle\int_{\mathbb{R}} |x| f(x)\mu(dx) < \infty$.

In diesem Fall gilt $\mathbb{E}(X) = \displaystyle\int_{\mathbb{R}} x \cdot f(x)\mu(dx)$, also speziell

$$\mathbb{E}(X) = \begin{cases} \displaystyle\int_{\mathbb{R}} x \cdot f(x)\, dx, & \text{falls } \mu = \lambda^1, \\ \displaystyle\sum_{x \in D} x \cdot \mathbb{P}(X = x), & \text{falls } \mu = \mu_D = \sum_{x \in D} \delta_x. \end{cases}$$

Beweis: $\mathbb{E}(X)$ existiert, falls $\int_{\mathbb{R}} |id_{\mathbb{R}}| d\mathbb{P}^X < \infty$ (vgl. Folie 18.39). Setze nun $\Omega = \mathbb{R}$, $\nu = \mathbb{P}^X (= f\mu)$ und $\varphi = id_X$. Dann gilt nach Satz 18.30:

$$\mathbb{E}(X) = \int_{\mathbb{R}} id_{\mathbb{R}}\, d\mathbb{P}^X = \int_{\mathbb{R}} id_{\mathbb{R}} \cdot f\, d\mu = \int_{\mathbb{R}} id_{\mathbb{R}}(x) \cdot f(x)\mu(dx) = \int_{\mathbb{R}} x \cdot f(x)\mu(dx) \qquad \square$$

Produktmaße

Satz und Definition 18.32

Es seien $(\Omega_j, \mathcal{A}_j, \mu_j), 1 \leq j \leq n, n \geq 2, \sigma$-endliche Maßräume. Sei $\Omega = \Omega_1 \times \cdots \times \Omega_n$ und

$$\mathcal{A} := \bigotimes_{j=1}^{n} \mathcal{A}_j = \sigma(\{A_1 \times \cdots \times A_n : A_i \in \mathcal{A}_i, 1 \leq i \leq n\})$$

die Produkt-σ-Algebra von $\mathcal{A}_1, \ldots, \mathcal{A}_n$. Dann existiert genau ein Maß μ auf $\mathcal{A}$ mit

$$\mu(A_1 \times \cdots \times A_n) = \prod_{j=1}^{n} \mu_j(A_j), \qquad A_i \in \mathcal{A}_i, 1 \leq i \leq n.$$

Man nennt μ das Produkt von $\mu_1, \ldots, \mu_n$ und schreibt $\mu = \mu_1 \otimes \cdots \otimes \mu_n = \bigotimes_{j=1}^{n} \mu_j$. Der Maßraum

$$\bigotimes_{j=1}^{n} (\Omega_j, \mathcal{A}_j, \mu_j) := \left(\times_{j=1}^{n} \Omega_j, \bigotimes_{j=1}^{n} \mathcal{A}_j, \bigotimes_{j=1}^{n} \mu_j \right)$$

heißt das Produkt von $(\Omega_j, \mathcal{A}_j, \mu_j), 1 \leq j \leq n$.

Beweis: Existenz: Fortsetzungssatz von Carathéodory, Eindeutigkeit: Satz 18.11

Der Satz von Fubini

Satz 18.33 (Fubini)

Sei $n = 2$ in der Situation von Satz 18.32. Sei $f : \Omega \to \overline{\mathbb{R}}$ messbar. Dann sind die Schnittfunktionen

$$\begin{aligned} f^2_{\omega_1} &: \Omega_2 \to \mathbb{R}, \quad f^2_{\omega_1}(\omega_2) := f(\omega_1, \omega_2), \qquad (\omega_1 \in \Omega_1) \\ f^1_{\omega_2} &: \Omega_1 \to \mathbb{R}, \quad f^1_{\omega_2}(\omega_1) := f(\omega_1, \omega_2), \qquad (\omega_2 \in \Omega_2) \end{aligned}$$

jeweils $(\mathcal{A}_2, \overline{\mathcal{B}})$- beziehungsweise $(\mathcal{A}_1, \overline{\mathcal{B}})$-messbar.

(a) Ist $f \geq 0$, so sind

$$\Omega_2 \ni \omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2)\ \mu_1(d\omega_1), \qquad \Omega_1 \ni \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\ \mu_2(d\omega_2)$$

jeweils $(\mathcal{A}_2, \overline{\mathcal{B}})$- beziehungsweise $(\mathcal{A}_1, \overline{\mathcal{B}})$-messbar. Ferner gilt

$$\begin{aligned} \int_\Omega f\ d(\mu_1 \otimes \mu_2) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\ \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \qquad (\star) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\ \mu_2(d\omega_2) \right) \mu_1(d\omega_1). \qquad (\star\star) \end{aligned}$$

Der Satz von Fubini

Satz 18.33 (Fubini, Fortsetzung)

(b) Ist f eine $(\mu_1 \otimes \mu_2)$-integrierbare, numerische Funktion, dann gilt:

- $f(\omega_1, \cdot)$ ist μ_2-integrierbar für μ_1-fast alle $\omega_1 \in \Omega_1$,
- $f(\cdot, \omega_2)$ ist μ_1-integrierbar für μ_2-fast alle $\omega_2 \in \Omega_2$,
- die μ_1-fast überall definierte Funktion

$$\omega_1 \mapsto \int |f(\omega_1, \omega_2)| \; \mu_2(d\omega_2) \;\; \text{ist } \mu_1\text{-integrierbar},$$

- die μ_2-fast überall definierte Funktion

$$\omega_2 \mapsto \int |f(\omega_1, \omega_2)| \; \mu_1(d\omega_1) \;\; \text{ist } \mu_2\text{-intergrierbar}.$$

(und ebenso ohne Betragsstriche). Ferner gelten $(\star)$ und $(\star\star)$.

Entsprechende Aussagen gelten allgemeiner für $n \geq 2$.

Stochastische Unabhängigkeit und Produktmaße

Definition und Satz 18.34

Es seien $X_1, \dots, X_n$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Die Zufallsvariablen $X_1, \dots, X_n$ heißen (stochastisch) unabhängig (bezüglich $\mathbb{P}$), falls gilt

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j), \qquad B_1, \dots, B_n \in \mathcal{B}.$$

Mit $X := (X_1, \dots, X_n)$ sind $X_1, \dots, X_n$ genau dann unabhängig, wenn

$$\mathbb{P}^X = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}.$$

Die Behauptung folgt mit Hilfe des Eindeutigkeitssatzes für Maße (Folgerung 18.12).

Bemerkung. Die Definition der Unabhängigkeit gilt analog, wenn $X_1, \dots, X_n$ Werte in verschiedenen und allgemeinen messbaren Räumen annehmen, zum Beispiel X_j in $(\mathbb{R}^{l_j}, \mathcal{B}^{l_j})$.

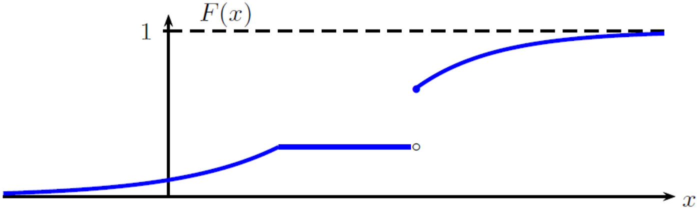
19. Verteilungsfunktionen und Lebesgue-Dichten

Definition 19.1

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable auf Ω. Die durch

$$F(x) := \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

erklärte Funktion $F : \mathbb{R} \to [0, 1]$ heißt Verteilungsfunktion von X.



Eigenschaften von Verteilungsfunktionen

Satz 19.2

Für jede Verteilungsfunktion F gilt:

(a) F ist monoton wachsend;

(b) F ist rechtsseitig stetig;

(c) $\lim\limits_{x \to -\infty} F(x) = 0$ und $\lim\limits_{x \to +\infty} F(x) = 1$;

(d) $\mathbb{P}(a < X \leq b) = F(b) - F(a)$ für $a, b \in \mathbb{R}$ mit $a \leq b$;

(e) $\mathbb{P}(X = x) = F(x) - F(x-)$, $x \in \mathbb{R}$, wobei $F(x-) := \lim\limits_{y \nearrow x} F(y)$ der linksseitige Grenzwert ist;

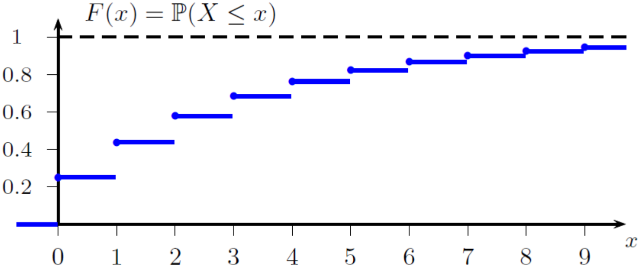
(f) F hat höchstens abzählbar viele Unstetigkeitsstellen, welche allesamt Sprungstellen sind.

Bemerkungen: (1) Wegen der Eigenschaften (a)-(c) ist die Definition der VF einer ZV konsistent mit der früheren Definition 18.13 einer (abstrakten) VF.
(2) Jede Verteilung auf $\mathbb{R}$ besitzt eine Verteilungsfunktion. Nach Folie 18.19 gilt umgekehrt: Jede Funktion $F : \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften (a)-(c) definiert genau ein Wahrscheinlichkeitsmaß μ_F auf $\mathbb{R}$, d.h. eine Verteilung.
(3) **Folgerung**: $X \sim Y \iff F_X = F_Y$

Verteilungsfunktion einer diskreten ZV

Gegeben seien $(\Omega, \mathcal{A}, \mathbb{P})$ und eine ZV X mit $\mathbb{P}(X \in D) = 1$ für eine abzählbare Menge $D \subset \mathbb{R}$. Dann ist die Verteilungsfunktion von X diskret (hat abzählbar viele Werte) und von der Form

$$F(x) = \mathbb{P}(X \leq x) = \sum_{y \in D \,:\, y \leq x} \mathbb{P}(X = y), \qquad x \in \mathbb{R}.$$



Verteilungsfunktion einer diskreten Verteilung (G(1/4))

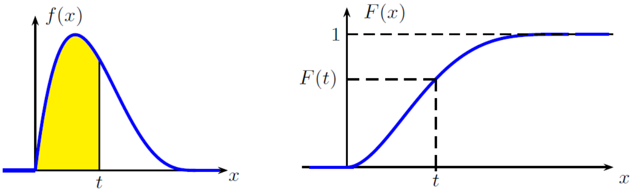
Verteilungsfunktion einer stetigen ZV

Hat X die Lebesgue-Dichte f (vgl. Def. 18.28/18.29), so gilt

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \qquad x \in \mathbb{R}. \tag{$\star$}$$

Die **Umkehrung** gilt ebenfalls: Gilt $(\star)$ für die Verteilungsfunktion F von X mit einer messbaren Funktion $f \geq 0$, so hat X die Lebesgue-Dichte f, das heißt

$$\mathbb{P}(X \in B) = \int_{B} f(t)\, dt, \qquad B \in \mathcal{B}. \tag{$\star\star$}$$



Dichte und Verteilungsfunktion einer stetigen Zufallsvariablen

Verteilungsfunktion einer stetigen ZV

Hintergrund: Sei $f \geq 0$ messbar. Genau dann gilt

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \quad \text{für alle } \; x \in \mathbb{R}, \tag{$\star$}$$

wenn

$$\mathbb{P}(X \in B) = \int_B f(t)\, dt \quad \text{für alle } \; B \in \mathcal{B} \text{ gilt.} \tag{$\star\star$}$$

Beweis: "$\Leftarrow$" klar. "$\Rightarrow$": Folgt aus dem Eindeutigkeitssatz für W.maße (Folgerung 18.12) mit $\mathcal{J} := \{(-\infty, x] : x \in \mathbb{R}\}$ als $\cap$-stabilem Erzeuger. ☐

Bemerkungen:

(1) Die Dichte f in $(\star\star)$ ist nur bis auf eine λ^1-Nullmenge bestimmt.

(2) Im Folgenden werden nur f betrachtet, die bis auf endlich viele Stellen stetig gewählt werden können.

(3) Ist x eine Stetigkeitsstelle von f, so gilt $F'(x) = f(x)$. (Hauptsatz der DIR)

(4) Ist F (stetig und) außerhalb einer endlichen Menge M stetig differenzierbar, dann wird durch $f(x) := F'(x),\ x \in \mathbb{R} \setminus M$, und $f(x) := 0,\ x \in M$, eine Lebesgue-Dichte zu F definiert, das heißt

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad x \in \mathbb{R}.$$

(5) **Achtung!** Nicht jede VF besitzt eine Dichte. Es gibt F mit $F'(t) = 0$ für λ^1-f.a. $t \in \mathbb{R}$ (etwa diskrete VF, "devil's staircase").

19.1 Beispiele stetiger Verteilungen

Definition 19.3 (Gleichverteilung auf einem Intervall)

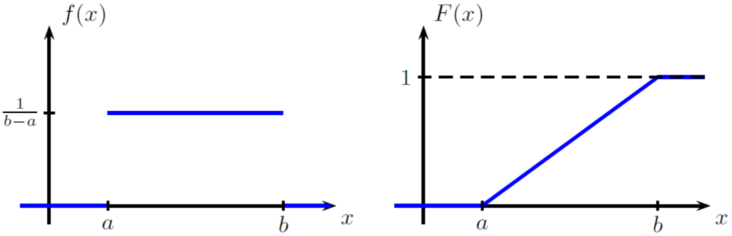
Seien $a, b \in \mathbb{R}$, $a < b$ und

$$f(x) := \begin{cases} \frac{1}{b-a}, & x \in (a,b), \\ 0, & \text{sonst.} \end{cases}$$

Man nennt eine Zufallsvariable X gleichverteilt auf (a, b), kurz $X \sim \mathrm{U}_{(a,b)}$, falls X die Lebesgue-Dichte f hat.

Es gilt dann:

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



Dichte und Verteilungsfunktion der Gleichverteilung $\mathrm{U}(a, b)$

Eigenschaften der Gleichverteilung

Hinweis: Pseudo-Zufallszahlengeneratoren simulieren die Verteilung $\mathrm{U}_{(0,1)}$ zum Beispiel mit Hilfe von Kongruenzgeneratoren. Mit Hilfe von $\mathrm{U}_{(0,1)}$-verteilten Zufallsvariablen lassen sich dann weitere Verteilungen erzeugen.

Es gilt

$$\mathbb{E}(X) = \frac{1}{b-a}\int_a^b x\ dx = \frac{a+b}{2},$$

$$\mathbb{E}(X^2) = \frac{1}{b-a}\int_a^b x^2\ dx = \frac{a^2+ab+b^2}{3},$$

$$\mathbb{V}(X) = \frac{(a-b)^2}{12}.$$

Satz 19.4

$X \sim \mathrm{U}_{(0,1)} \Longleftrightarrow a + (b-a)X \sim \mathrm{U}_{(a,b)}$

Beweis: Vergleich der Verteilungsfunktion von $a + (b-a)X$ mit der von $Y \sim \mathrm{U}_{(a,b)}$. □

Exponentialverteilung

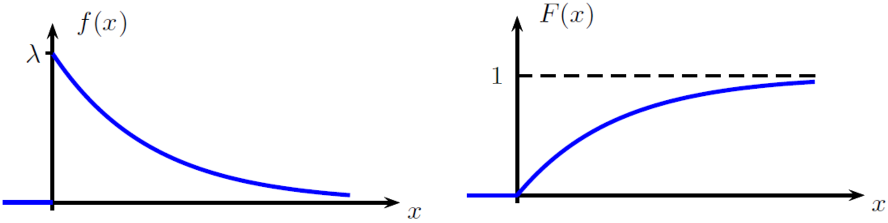
Definition 19.5

Sei $\lambda > 0$. Eine Zufallsvariable X hat eine Exponentialverteilung mit Parameter λ, kurz $X \sim \mathrm{Exp}(\lambda)$, wenn X die folgende Lebesgue-Dichte hat:

$$f(x) := \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Es gilt dann

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$



Dichte und Verteilungsfunktion der Exponentialverteilung

Eigenschaften der Exponentialverteilung

- Für $t, h > 0$ und $X \sim \text{Exp}(\lambda)$ gilt (die Gedächtnislosigkeit)

$$\mathbb{P}(X \geq t + h \mid X \geq t) = \frac{\mathbb{P}(X \geq t + h)}{\mathbb{P}(X \geq t)} = \frac{1 - F(t + h)}{1 - F(t)} = \frac{\mathrm{e}^{-\lambda(t+h)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda h} = \mathbb{P}(X \geq h).$$

- Für $X \sim \text{Exp}(\lambda)$ gilt (siehe Tutorium):

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathbb{E}(X^2) = \frac{2}{\lambda^2}, \qquad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

- Skalierungsverhalten: Ist $X \sim \text{Exp}(1)$, so ist $\lambda^{-1} \cdot X \sim \text{Exp}(\lambda)$.
- Zusammenhang mit der Gleichverteilung (Übung):

  Ist $U \sim \mathrm{U}_{(0,1)}$, so gilt $X := -\frac{1}{\lambda} \ln(U) \sim \text{Exp}(\lambda)$.

Die Gammafunktion

Zur Erinnerung (Analysis): Die Gammafunktion $\Gamma : (0, \infty) \to \mathbb{R}$ ist erklärt durch

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \mathrm{e}^{-t} \, dt, \quad \alpha > 0.$$

Nützlich ist oft die Gleichung

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha > 0,$$

speziell ist

$$\Gamma(2) = \Gamma(1) = 1, \quad \Gamma(n+1) = n! \quad \text{sowie} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Normalverteilung

Definition 19.6

Seien $\mu \in \mathbb{R},\ \sigma > 0$ und

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Eine Zufallsvariable X hat eine Normalverteilung mit den Parametern μ und σ^2, kurz: $X \sim \mathrm{N}(\mu, \sigma^2)$, wenn X die Lebesgue-Dichte f besitzt. Man nennt $\mathrm{N}(0,1)$ die Standard-Normalverteilung oder standardisierte Normalverteilung. Deren Dichte und Verteilungsfunktion werden mit φ bzw. Φ bezeichnet.

Mit der "Gaußschen Glockenkurve" $\varphi(x) = \sqrt{2\pi}^{-1} \mathrm{e}^{-x^2/2}$, $x \in \mathbb{R}$ gilt also für die Dichte von $X \sim \mathrm{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \qquad x \in \mathbb{R}$$

und daher

$$\int_{\mathbb{R}} f(x)\ dx = \frac{1}{\sigma} \int_{-\infty}^{\infty} \varphi\left(\frac{x-\mu}{\sigma}\right)\ dx = \int_{\mathbb{R}} \varphi(t)\ dt = 1.$$

Bezeichnet F die Verteilungsfunktion von X, so gilt

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right)\ dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(z)\ dz = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Eigenschaften der Normalverteilung

Satz 19.7

Für $\mu \in \mathbb{R}, \sigma > 0$ gilt

(a) $X \sim \mathrm{N}(0,1) \Longrightarrow \mu + \sigma X \sim \mathrm{N}(\mu, \sigma^2)$;

(b) $X \sim \mathrm{N}(\mu, \sigma^2) \Longrightarrow \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

Beweis

(a) Für die VF F_Y von $Y := \mu + \sigma X$ gilt für jedes $x \in \mathbb{R}$

$$F_Y(x) = \mathbb{P}(\mu + \sigma X \leq x) = \mathbb{P}\left(X \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\tfrac{x-\mu}{\sigma}\right) = F_{\mathrm{N}(\mu,\sigma^2)}(x).$$

(b) Analog gilt für die VF F_Z von $Z := \frac{X-\mu}{\sigma}$ für jedes $x \in \mathbb{R}$

$$F_Z(x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq x\right) = \mathbb{P}(X \leq \mu + \sigma x) = F_{\mathrm{N}(\mu,\sigma^2)}(\mu + \sigma x) = \Phi\left(\frac{\mu + \sigma x - \mu}{\sigma}\right) = \Phi(x).$$

Erwartungswert und Varianz der Normalverteilung

Satz 19.8

Sei $\mu \in \mathbb{R}, \sigma > 0$ und $X \sim \mathrm{N}(\mu, \sigma^2)$. Dann gilt $\mathbb{E}(X) = \mu$ und $\mathbb{V}(X) = \sigma^2$.

Beweis

Für $X \sim \mathrm{N}(0,1)$ erhält man

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\varphi(x)\ dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\mathrm{e}^{-\frac{x^2}{2}}\ dx = 0$$

und

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2\varphi(x)\ dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty x^2\mathrm{e}^{-\frac{x^2}{2}}\ dx = \sqrt{\frac{2}{\pi}} \int_0^\infty \sqrt{2}t^{\frac{1}{2}}\mathrm{e}^{-t}\ dt = \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 1$$

und daher $\mathbb{V}(X) = 1$. Für $Y \sim \mathrm{N}(\mu, \sigma^2)$ folgt nun wegen $Y \sim \mu + \sigma X,\ X \sim \mathrm{N}(0,1)$:

$$\mathbb{E}(Y) = \mu + \sigma\mathbb{E}(X) = \mu \quad \text{und} \quad \mathbb{E}(Y^2) = \mathbb{E}((\mu + \sigma X)^2) = \mu^2 + \sigma^2 + 2\mu\sigma\mathbb{E}(X) = \mu^2 + \sigma^2,$$

also $\mathbb{V}(Y) = \sigma^2$.

Transformation von Zufallsvariablen

- $X \mapsto \mu + \sigma X$ ist eine (lineare) Transformation der Zufallsvariablen X.

Allgemeiner: Sei X eine ZV mit Verteilungsfunktion F und Dichte f, und sei $T : \mathbb{R} \to \mathbb{R}$ eine messbare Funktion.

Frage: Welche Verteilungsfunktion besitzt $Y := T(X)$? Besitzt Y eine Dichte?

Grundlegende Vorgehensweise: Sei $G(y) := \mathbb{P}(Y \leq y)$, $y \in \mathbb{R}$, die Verteilungsfunktion von Y. Schreibe das Ereignis $\{Y \leq y\} = \{T(X) \leq y\}$ in ein Ereignis für X um, dessen W. mit Hilfe der Verteilungsfunktion F von X ausgerechnet werden kann.

Beispiele

- $T(X) := X^4$: $\{X^4 \leq y\} = \{-y^{1/4} \leq X \leq y^{1/4}\}$ für $y \geq 0$.
  $G(y) := \mathbb{P}(X^4 \leq y) = \mathbb{P}(-y^{1/4} \leq X \leq y^{1/4}) = F(y^{1/4}) - F(-y^{1/4})$,
  da $\mathbb{P}(X = -y^{1/4}) = 0$.
- $T(X) = \mathrm{e}^X$: $\{\mathrm{e}^X \leq y\} = \{X \leq \log y\}$ für $y > 0$,
  also $G(y) \;=\; \mathbb{P}(Y \leq y) \;=\; F(\log y)$, $y > 0$; $G(y) = 0$, sonst.

Streng monotone Transformationen

Satz 19.9

Es sei X eine ZV mit Verteilungsfunktion F und einer bis auf endlich viele Stellen stetigen Dichte f. Es gelte $\mathbb{P}(X \in O) = 1$ für ein offenes Intervall O.
Die Einschränkung der Abbildung $T : \mathbb{R} \to \mathbb{R}$ auf O sei stetig differenzierbar und streng monoton mit $T'(x) \neq 0, x \in O$. Seien $T^{-1} : T(O) \to O$ die Inverse von T auf $T(O)$ und G die Verteilungsfunktion von $Y = T(X)$. Dann gelten:

(a) Ist T streng monoton wachsend, so gilt $G(y) \; = \; F(T^{-1}(y)), \quad y \in T(O)$.

(b) Ist T streng monoton fallend, so gilt $G(y) \; = \; 1 - F(T^{-1}(y)), \quad y \in T(O)$.

(c) In beiden Fällen besitzt Y die Dichte

$$g(y) \; := \; \frac{f(T^{-1}(y))}{|T'(T^{-1}(y))|}, \quad y \in T(O), \text{ und } g(y) := 0, \text{ sonst.}$$

Beweis von Satz 19.9

Beweis

Sei T streng monoton wachsend. Für $y \in T(O)$ gilt

$$G(y) \;=\; \mathbb{P}(Y \leq y) \;=\; \mathbb{P}(T(X) \leq y) \;=\; \mathbb{P}(X \leq T^{-1}(y)) = F(T^{-1}(y)).$$

Differentiation (in jedem Stetigkeitspunkt der Ableitung!) ergibt

$$g(y) \;=\; G'(y) \;=\; \frac{F'(T^{-1}(y))}{T'(T^{-1}(y))} \;=\; \frac{f(T^{-1}(y))}{T'(T^{-1}(y))}.$$

Der zweite Fall folgt analog.

Lognormalverteilung

Beispiel

Seien $X \sim \mathrm{N}(\mu, \sigma^2)$ und $T(x) := \mathrm{e}^x$, $x \in \mathbb{R}$. Was ist die Verteilung der ZV $Y := T(X) = \mathrm{e}^X$?

Sei $G(y) := \mathbb{P}(Y \leq y)$, $y \in \mathbb{R}$ die Verteilungsfunktion von Y.

- Für $y \leq 0$ gilt $G(y) = 0$
- Für $y > 0$ ist

$$G(y) \;=\; \mathbb{P}(Y \leq y) \;=\; \mathbb{P}(\mathrm{e}^X \leq y) \;=\; \mathbb{P}(X \leq \log y) \;=\; \Phi\left(\frac{\log y - \mu}{\sigma}\right)$$

- Differentiation (Kettenregel) liefert wegen $\Phi' = \varphi$

$$g(y) \;=\; \varphi\left(\frac{\log y - \mu}{\sigma}\right)\frac{1}{\sigma y} \;=\; \frac{1}{\sigma y\sqrt{2\pi}} \cdot \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)$$

für $y > 0$ ($g(y) := 0$, sonst).

Lognormalverteilung

Definition 19.10

Eine positive Zufallsvariable X besitzt eine Lognormalverteilung mit den Parametern μ und σ^2 ($\mu \in \mathbb{R}, \sigma^2 > 0$), falls X die Dichte

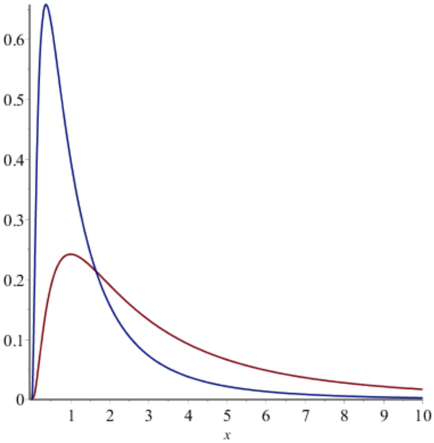
$$f(x) \;=\; \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2} \right)$$

für $x > 0$ und $f(x) = 0$, sonst, besitzt.

Kurz $X \sim \mathrm{LN}(\mu, \sigma^2)$.

Beachte: $X \sim \mathrm{LN}(\mu, \sigma^2) \iff \ln(X) \sim \mathrm{N}(\mu, \sigma^2)$.

Dichten von $\mathrm{LN}(0,1)$ und $\mathrm{LN}(1,1)$



19.2 Gemeinsame Verteilungen mit Lebesgue-Dichten

Wenn nicht etwas anderes vereinbart wird, sind in diesem Abschnitt mit Dichten stets Dichten bezüglich des Lebesgue-Maßes gemeint.

Zur Erinnerung: ein k-dimensionaler Zufallsvektor X auf einem W.raum $(\Omega, \mathcal{A}, \mathbb{P})$ heißt (absolut) stetig (verteilt), wenn $\mathbb{P}^X$ eine Lebesgue-Dichte (kurz L-Dichte) hat, d.h.

$$\mathbb{P}^X(B) = \int \mathbf{1}\{x \in B\}\, f(x)\, \lambda^k(dx), \quad B \in \mathcal{B}^k,$$

mit einer messbaren, nichtnegativen Funktion f (die Lebesgue-Dichte von $\mathbb{P}^X$).

Je zwei L-Dichten von $\mathbb{P}^X$ stimmen λ^k-f.ü. überein.

Marginalverteilungen bei Dichten

Satz 19.11

Sei $X = (X_1, \ldots, X_k)$ ein Zufallsvektor mit L-Dichte f. Dann haben auch $X_1, \ldots, X_k$ L-Dichten. Eine L-Dichte f_j von X_j ist gegeben durch

$$f_j(t) = \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_k)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_k, \quad t \in \mathbb{R}.$$

Beweis

Mit $B_j \in \mathcal{B}$ und $\tilde{B}_j := \mathbb{R}^{j-1} \times B_j \times \mathbb{R}^{k-j}$ gilt

$$\mathbb{P}(X_j \in B_j) = \mathbb{P}(X \in \mathbb{R}^{j-1} \times B_j \times \mathbb{R}^{k-j}) = \int_{\tilde{B}_j} f(x)\, \lambda^k(dx).$$

Die Aussage folgt nun mit dem Satz von Fubini, das heißt

$$\mathbb{P}(X_j \in B_j) = \int_{B_j} \int_{\mathbb{R}^{k-1}} f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_k)\ \lambda^{k-1}(d(x_1, \ldots, x_{j-1}, x_{j+1} \ldots, x_k))\, dt. \quad \square$$

Unabhängigkeit und Dichten

Satz 19.12

Sei $X = (X_1, \dots, X_k)$ ein k-dimensionaler Zufallsvektor mit L-Dichte f. Seien $f_1, \dots, f_k$ marginale L-Dichten von $X_1, \dots, X_k$. Genau dann sind $X_1, \dots, X_k$ stochastisch unabhängig, wenn gilt

$$f(x) = \prod_{j=1}^{k} f_j(x_j) \qquad \text{für } \lambda^k\text{-fast alle } \ x = (x_1, \dots, x_k) \in \mathbb{R}^k.$$

Beweis

Für $B_1, \dots, B_k \in \mathcal{B}$ gilt einerseits

$$\left(\bigotimes_{j=1}^{k} \mathbb{P}^{X_j}\right)(B_1 \times \cdots \times B_k) = \prod_{j=1}^{k} \mathbb{P}^{X_j}(B_j) = \prod_{j=1}^{k} \int_{B_j} f_j(x_j)\, dx_j = \int_{B_1 \times \cdots \times B_k} \prod_{j=1}^{k} f_j(x_j)\, d(x_1, \dots, x_k)$$

und andererseits

$$\mathbb{P}^X(B_1 \times \cdots \times B_k) = \int_{B_1 \times \cdots \times B_k} f(x)\, dx.$$

Die Aussage folgt nun mit Satz 18.34 ($\mathbb{P}^X = \mathbb{P}^{X_1} \otimes \cdots \otimes \mathbb{P}^{X_n}$), vgl. auch die Bem. auf Folie 18.49. ☐

Unabhängigkeit und Dichten

Aus dem vorangehenden Argument erhält man auch folgende partielle Umkehrung von Satz 19.11.

Satz 19.13

Sei $X = (X_1, \dots, X_k)$ ein k-dimensionaler Zufallsvektor mit stochastisch unabhängigen Komponenten $X_1, \dots, X_k$, die L-Dichten $f_1, \dots, f_k$ besitzen. Dann besitzt auch X eine L-Dichte f und für diese gilt

$$f(x) = \prod_{j=1}^{k} f_j(x_j) \quad \text{für } \lambda^k\text{-fast alle } x = (x_1, \dots, x_k) \in \mathbb{R}^k.$$

Anwendung von Satz 19.12

Beispiel

Sei $X = (X_1, X_2) \sim \mathrm{U}_B$ mit $B = \left[0, \frac{1}{2}\right]^2 \cup \left[\frac{1}{2}, 1\right]^2$.

Also ist $f(x_1, x_2) = 2 \cdot \mathbb{1}_B(x_1, x_2) = 2 \cdot \mathbb{1}\{x_1, x_2 \in [0, \frac{1}{2}]$ oder $x_1, x_2 \in [\frac{1}{2}, 1]\}$ die L-Dichte von X.

Es gilt

$$f_1(x_1) = \int_0^1 f(x_1, x_2)\, dx_2 = 1 \quad \text{für } x_1 \in [0, 1],$$

und ebenso $f_2(x_2) = \mathbb{1}_{[0,1]}(x_2)$, d.h. $X_1 \sim \mathrm{U}_{(0,1)}$, $X_2 \sim \mathrm{U}_{(0,1)}$. Wegen

$$\mathbb{P}\left(X_1 \in \left[0, \frac{1}{2}\right], X_2 \in \left[0, \frac{1}{2}\right]\right) = 2 \cdot \frac{1}{4} = \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}\left(X_1 \in \left[0, \frac{1}{2}\right]\right) \cdot \mathbb{P}\left(X_2 \in \left[0, \frac{1}{2}\right]\right)$$

bzw.

$$\{0, 2\} \ni f(x_1, x_2) \neq f_1(x_1) \cdot f_2(x_2) \equiv 1 \quad \text{für } x_1, x_2 \in (0, 1),$$

sind X_1, X_2 **nicht** stochastisch unabhängig.

Transformation von Zufallsvektoren

Voraussetzungen:

- Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.Raum und $X : \Omega \to \mathbb{R}^k$ ein Zufallsvektor mit Verteilung $\mathbb{P}^X = f\,\lambda^k$, d.h. L-Dichte f.
- Seien $T : \mathbb{R}^k \to \mathbb{R}^k$ messbar, $U, V \subset \mathbb{R}^k$ offen und $T|_U : U \to V$ ein Diffeomorphismus.
- $\mathbb{P}$–fast sicher gelte $X \in U$, d.h. (o.B.d.A.) $f = 0$ auf $\mathbb{R}^k \setminus U$.

Aufgabe: Bestimme die Verteilung von $Y := T(X)$.

Satz 19.14

Mit den obigen Vorausetzungen besitzt $Y := T(X)$ die L-Dichte

$$g(y) = \frac{f \circ T^{-1}(y)}{|\det DT(T^{-1}(y))|}, \quad y \in V,$$

und $g(y) = 0$ für $y \notin V$.

Beweis von Satz 19.14

Beweisidee

Sei $A \subset V$ Borelmenge. Dann ist

$$\begin{aligned}\mathbb{P}(Y \in A) &= \mathbb{P}(X \in T^{-1}(A)) = \mathbb{P}(X \in U \cap T^{-1}(A)) \\ &= \int_U \mathbf{1}_{T^{-1}(A)}(x) f(x)\, \lambda^k(dx) \\ &= \int_V \mathbf{1}_{T^{-1}(A)}(T^{-1}(y)) f(T^{-1}(y)) |\det DT^{-1}(y)|\, \lambda^k(dy) \\ &= \int_V \mathbf{1}_A(y) \frac{f \circ T^{-1}(y)}{|\det DT(T^{-1}(y))|}\, \lambda^k(dy),\end{aligned}$$

wobei für die dritte Gleichung der Transformationssatz für Gebietsintegrale (Analysis 3) verwendet wurde.

Anwendung des Transformationssatzes

Beispiel

Seien $X_1, X_2 \sim \mathrm{U}_{(0,1)}$ und stochastisch unabhängig. Dann sind

$$Y_1 := \sqrt{-2 \ln X_1} \cdot \cos(2\pi X_2), \quad Y_2 := \sqrt{-2 \ln X_1} \cdot \sin(2\pi X_2)$$

stochastisch unabhängig und $\mathrm{N}(0,1)$-verteilt.

Nachweis: Es ist $X = (X_1, X_2) \sim \mathrm{U}_{(0,1)^2}$, d.h. X hat die L-Dichte $f(x_1, x_2) = \mathbf{1}\{(x_1, x_2) \in (0,1)^2\}$ auf $\mathbb{R}^2$. Mit

$$(y_1, y_2) = T(x_1, x_2) := \left(\sqrt{-2 \ln x_1} \cdot \cos(2\pi x_2), \sqrt{-2 \ln x_1} \cdot \sin(2\pi x_2)\right)$$

für $(x_1, x_2) \in (0,1)^2$ und $T(x_1, x_2) := 0$ sonst bildet T die Menge $(0,1)^2$ diffeomorph ab auf $\mathbb{R}^2 \setminus ([0, \infty) \times \{0\})$, d.h. bis auf eine λ^2-Nullmenge ist das $\mathbb{R}^2$. Es gilt

$$|\det(DT(x_1, x_2))| = \frac{2\pi}{x_1}, \quad x_1 = \exp\left(-\frac{1}{2}(y_1^2 + y_2^2)\right)$$

für $(x_1, x_2) \in (0,1)^2$ und $(y_1, y_2) = T(x_1, x_2)$. Für die L-Dichte g von $(Y_1, Y_2) = T(X)$ folgt

$$g(y_1, y_2) = \frac{x_1}{2\pi} = \varphi(y_1) \cdot \varphi(y_2), \qquad \text{woraus sich alle Beh. ergeben, da } f(T^{-1}(y)) = 1.$$

19.3 Summen von Zufallsvariablen

Satz 19.15 (Faltungsformel für stetige Zufallsvariablen)

Seien X, Y unabhängig mit Dichten f_X bzw. f_Y. Dann hat $X+Y$ die Dichte

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\,\mathrm{d}s, \qquad t \in \mathbb{R}.$$

- Die durch

$$(f_X * f_Y)(t) := \int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\,\mathrm{d}s = \int_{-\infty}^{\infty} f_X(t-s)\, f_Y(s)\,\mathrm{d}s, \quad t \in \mathbb{R},$$

  definierte Funktion $f_X * f_Y : \mathbb{R} \to [0,\infty]$ nennt man die Faltung von f_X und f_Y. Obige Formel besagt also

$$f_{X+Y} = f_X * f_Y.$$

- Vgl. mit diskreter Faltungsformel

$$\mathbb{P}(X+Y=t) = \sum_{s:\mathbb{P}(X=s)>0} \mathbb{P}(X=s)\mathbb{P}(Y=t-s).$$

- Die Dichte der Summe von mehr als zwei unabhängigen Zufallsvariablen ergibt sich induktiv.

Beweis von Satz 19.15

Beweis

Für $z \in \mathbb{R}$ sei $B_z := \{(x, y) : x + y \leq z\}$. Mit $t := y + s$ folgt

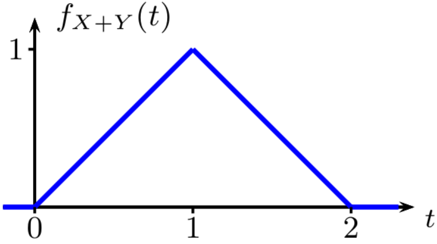
$$\begin{aligned}
\mathbb{P}(X + Y \leq z) &= \mathbb{P}((X, Y) \in B_z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-s} f_Y(y) \, \mathrm{d}y \right) f_X(s) \, \mathrm{d}s \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(t - s) \, \mathrm{d}t \right) f_X(s) \, \mathrm{d}s \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(s) \cdot f_Y(t - s) \, \mathrm{d}s \right) \mathrm{d}t.
\end{aligned}$$

Summe gleichverteilter ZV

Beispiel

X, Y unabhängig und je $\mathrm{U}_{(0,1)}$-verteilt. Hier ist $f_X(t) = f_Y(t) = \mathbb{1}_{(0,1)}(t)$. Man erhält

$$f_{X+Y}(t) = \int_{\mathbb{R}} \mathbb{1}_{(0,1)}(s)\,\mathbb{1}_{(0,1)}(t-s)\,ds = \begin{cases} \int_0^t 1\,ds = t, & \text{falls } 0 < t \leq 1, \\ \int_{t-1}^1 1\,ds = 2-t, & \text{falls } 1 \leq t < 2, \\ 0, & \text{sonst.} \end{cases}$$



Additionsgesetz für Normalverteilungen

Satz 19.16

Es seien X, Y stochastisch unabhängig mit $X \sim \mathrm{N}(\mu, \sigma^2)$, $Y \sim \mathrm{N}(\nu, \tau^2)$, $\mu, \nu \in \mathbb{R}$, $\sigma^2, \tau^2 > 0$. Dann gilt

$$X + Y \sim \mathrm{N}(\mu + \nu, \sigma^2 + \tau^2).$$

Beweis: siehe Tutorium

19.4 Momente und Quantile

Ist $X : \Omega \to \mathbb{R}^n$ auf $(\Omega, \mathcal{A}, \mathbb{P})$ ein Zufallsvektor und $h : \mathbb{R}^n \to \mathbb{R}$ eine messbare Funktion, so ist $\mathbb{E}[h \circ X]$ erklärt, falls

$$\mathbb{E}[|h \circ X|] = \int_{\mathbb{R}^n} |h| \, d\mathbb{P}^X < \infty.$$

In diesem Fall gilt

$$\mathbb{E}[h \circ X] = \int_{\mathbb{R}^n} h \, d\mathbb{P}^X = \int_{\mathbb{R}^n} h(x_1, \dots, x_n) \, \mathbb{P}^X(d(x_1, \dots, x_n)).$$

Falls $\mathbb{P}^X$ die Lebesgue-Dichte f hat, so folgt

$$\mathbb{E}[h \circ X] = \int_{\mathbb{R}^n} h(x) f(x) \, \lambda^n(dx).$$

Momente

Für reelle Zufallsvariablen X, Y können wir nun wie im diskreten Fall auch allgemein die folgenden Kenngrößen erklären (falls diese existieren):

- $\mathbb{E}[X^k]$, $\mathbb{E}[|X|^k]$, $\mathbb{E}[(X - \mathbb{E}X)^k]$ (k-tes Moment, absolutes k-tes Moment, zentriertes k-tes Moment)
- $\mathbb{V}(X)$, $\sigma_X = \sqrt{\mathbb{V}(X)}$, $\mathbb{C}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$ (Varianz, Standardabw., Kovarianz)
- $\varrho(X, Y) = \frac{\mathbb{C}(X,Y)}{\sigma_X \sigma_Y}$ (Korrelationskoeffizient).

Alle im diskreten Fall zuvor in Kapitel 12 bewiesenen Eigenschaften und Aussagen bleiben gültig, insbesondere bei Spezialisierung auf den Fall stetig verteilter Zufallsvariablen.

Beispiel: Multiplikationsformel

Satz 19.17

Sind X, Y unabhängige, reelle Zufallsvariablen mit existierenden Erwartungswerten, so existiert auch der Erwartungswert von XY, und es gilt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, das heißt $\mathbb{C}(X, Y) = 0$.

Beweis

Zunächst gilt

$$\begin{aligned}
\mathbb{E}(|XY|) &= \int_{\mathbb{R}^2} |xy| \, \mathbb{P}^{(X,Y)}(d(x,y)) \\
&= \int_{\mathbb{R}^2} |x| \cdot |y| \, (\mathbb{P}^X \otimes \mathbb{P}^Y)(d(x,y)) && \text{(Unabhängigkeit)} \\
&= \left(\int_{\mathbb{R}} |x| \, \mathbb{P}^X(dx) \right) \left(\int_{\mathbb{R}} |y| \, \mathbb{P}^Y(dy) \right) && \text{(Fubini)} \\
&= \mathbb{E}(|X|)\mathbb{E}(|Y|) < \infty. && \text{(EW existieren)}
\end{aligned}$$

Wiederholung der Rechung ohne Betragsstriche ergibt die Behauptung.

Momente der Standard-Normalverteilung

Beispiel

Sei $X \sim \mathrm{N}(0,1)$ und $p > 0$. Dann gilt

$$\begin{aligned}\mathbb{E}\left(|X|^p\right) &= \int_{\mathbb{R}} |x|^p \varphi(x)\, dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty x^p \mathrm{e}^{-\frac{x^2}{2}}\, dx \\ &= \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \int_0^\infty y^{\frac{p+1}{2}-1} \mathrm{e}^{-y}\, dy \\ &= \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right).\end{aligned}$$

Falls $p = 2k,\ k \in \mathbb{N}$, so folgt mit der Rekursionsformel zur Berechnung der Gammafunktion

$$\mathbb{E}\left(X^{2k}\right) = \prod_{j=1}^{k} (2j-1).$$

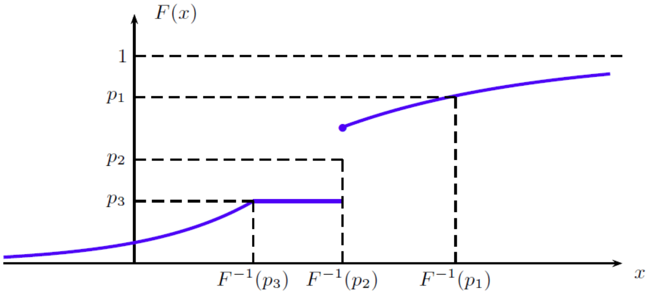
Quantile von reellen Zufallsvariablen

Definition 19.18

Seien X eine Zufallsvariable mit Verteilungsfunktion F und $p \in (0,1)$. Die Zahl

$$F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}$$

heißt p-Quantil der Verteilung von F (beziehungsweise von X).
Die Funktion $F^{-1} : (0,1) \to \mathbb{R}$ heißt Quantilfunktion zu F.



Bemerkungen zur Quantilfunktion

(1) F^{-1} ist wohldefiniert, da $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to \infty} F(x) = 1$.

(2) F^{-1} is monoton wachsend und linksseitig (!) stetig.

(3) Die Infimum-Bildung ist erforderlich! Insbesondere ist F^{-1} i.A. nicht die Umkehrfunktion zu F. Falls F streng monoton und stetig ist, so ist F^{-1} in der Tat die Umkehrfunktion zu F.

(4) Analog zur Situation bei empirischen Quantilen werden die folgenden Begriffe verwendet:

$F^{-1}(\frac{1}{2})$: Median,

$F^{-1}(\frac{1}{4})$: unteres Quartil,

$F^{-1}(\frac{3}{4})$: oberes Quartil,

$F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})$: Quartilsabstand.

Beispiel zur Quantilfunktion

Beispiel (Exponentialverteilung)

Sei $X \sim \mathrm{Exp}(\lambda)$, also

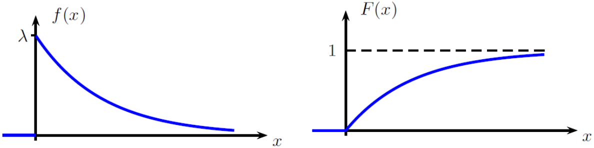
$$F(x) = \begin{cases} 1 - \exp(-\lambda x), & \text{falls } x \geq 0, \\ 0, & \text{sonst.} \end{cases}$$

F ist auf $\{x : 0 < F(x) < 1\}$ stetig und streng monoton wachsend. Also gilt

$$p = F(F^{-1}(p)) = 1 - \exp(-\lambda F^{-1}(p))$$

und somit

$$F^{-1}(p) = -\frac{1}{\lambda} \ln(1 - p).$$



Dichte und Verteilungsfunktion der Exponentialverteilung

Quantil-Transformation

Satz 19.19

Sei F eine Verteilungsfunktion und sei $U \sim \mathrm{U}_{(0,1)}$. Dann hat

$$X := F^{-1}(U)$$

die Verteilungsfunktion F.

Bemerkung: Die Quantiltransformation dient zur Erzeugung einer Pseudozufallszahl x mit Verteilungsfunktion F aus einer Pseudozufallszahl u mit Gleichverteilung $\mathrm{U}_{(0,1)}$.

Beispiel: $x = -\lambda^{-1} \ln(1-u)$ erzeugt eine $\mathrm{Exp}(\lambda)$-Zufallszahl.

Beweis

Für $x \in \mathbb{R}$ und $p \in (0,1)$ gilt

$$F(x) \geq p \Leftrightarrow x \geq F^{-1}(p).$$

Daraus folgt

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Kapitel 20: Grenzwertsätze

Satz 20.1 (Schwaches Gesetz der großen Zahlen)

Seien $X_1, X_2, \ldots$ paarweise unkorrelierte Zufallsvariablen mit gleichem Erwartungswert $\mu = \mathbb{E}(X_1)$ und gleicher endlicher Varianz $\sigma^2 := \mathbb{V}(X_1) < \infty$. Mit $\overline{X}_n := \frac{1}{n} S_n := \frac{1}{n}(X_1 + \cdots + X_n)$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{X}_n - \mu| \geq \varepsilon\right) = 0 \quad \text{für jedes } \varepsilon > 0.$$

Bemerkung: Unabhängigkeit impliziert Unkorreliertheit. Also gilt der Satz inbesondere für Folgen unabhängiger Zufallsvariablen.

Beweis

Nach Vor. gilt $\mathbb{E}(\overline{X}_n) = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) = \mu$ und $\mathbb{V}(\overline{X}_n) = \frac{1}{n^2}\mathbb{V}(X_1 + \ldots + X_n) = \frac{1}{n}\mathbb{V}(X_1) = \frac{\sigma^2}{n}$. Somit folgt mit der Tschebyscheff-Ungleichung

$$\mathbb{P}\left(|\overline{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Konvergenzbegriff im Hintergrund: Seien $Y, Y_1, Y_2, \ldots$ ZV auf einem gemeinsamen W.raum. Dann sagt man, $(Y_n)_{n\geq 1}$ konvergiert stochastisch gegen Y, falls gilt: $\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) = 0 \quad \text{für jedes } \varepsilon > 0.$

Schwaches Gesetz großer Zahlen

Folgerung

Sind $A_1, A_2, \ldots$ paarweise unabhängige Ereignisse mit $\mathbb{P}(A_i) = p$ für $i \in \mathbb{N}$, so gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{A_i} - p \right| \geq \varepsilon \right) = 0 \qquad \text{für} \quad \varepsilon > 0.$$

Hier ist $\frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}_{A_i}$ die (zufällige) relative Erfolgshäufigkeit.

Äquivalent: Zu $\varepsilon > 0$ und $\eta > 0$ existiert ein $n_0 = n_0(\varepsilon, \eta)$, so dass für $n \geq n_0$ gilt

$$1 \geq \mathbb{P}(|\overline{X}_n - \mu| < \varepsilon) \geq 1 - \eta.$$

Die Ereignisse $\{\omega \in \Omega : |\overline{X}_n - \mu| < \varepsilon\}$ haben also Wahrscheinlichkeiten zunehmend nahe 1 und die Ereignisse $\{\omega \in \Omega : |\overline{X}_n - \mu| \geq \varepsilon\}$ haben Wahrscheinlichkeiten zunehmend nahe 0 (jeweils für $n \to \infty$).

Analog lässt sich das für die Folgerung formulieren.

Ausblick: Starkes Gesetz großer Zahlen

Satz 20.2 (Starkes Gesetz der großen Zahlen)

Seien $X_1, X_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}|X_1| < \infty$ (und $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$). Dann gilt

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{X}_n = \mathbb{E}X_1\right) = 1,$$

das heißt, für $\mathbb{P}$-fast alle $\omega \in \Omega$ gilt $\lim\limits_{n\to\infty} \overline{X}_n(\omega) = \mathbb{E}X_1$.

Diese Aussage ist stärker als Satz 20.1 und verknüpft mit dem Konvergenzbegriff der **fast sicheren Konvergenz**:

Seien Y_n, $n \in \mathbb{N}$ und Y Zufallsvariablen auf einem gemeinsamen W.raum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann sagen wir: $Y_n \to Y$ $\mathbb{P}$-fast sicher, falls

$$\mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1.$$

Eine detailierte Betrachtung erfolgt in der *Wahrscheinlichkeitstheorie*.

Zentrale Grenzwertsätze

Fragestellung: Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt (u.i.v.) mit $\mathbb{E}(X_1^2) < \infty$. Was lässt sich über das Verhalten der Zufallsvariablen $S_n = X_1 + \cdots + X_n$ für $n \to \infty$ aussagen?

Setze $\mu := \mathbb{E}(X_1)$ und $\sigma := \sqrt{\mathbb{V}(X_1)}$. Wegen $\mathbb{E}(S_n) = n\mu$ und $\mathbb{V}(S_n) = n\,\mathbb{V}(X_1)$ ist eine Standardisierung erforderlich.

Aufgrund der Tschebyscheff-Ungleichung gilt

$$\mathbb{P}\left(\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| \leq b\right) \geq 1 - \frac{1}{b^2}.$$

Im Folgenden soll nicht nur eine Abschätzung betrachtet werden, sondern das präzise asymptotische Verhalten von Wahrscheinlichkeiten

$$\mathbb{P}\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \qquad \text{für } n \to \infty$$

für beliebige $a \leq b$ bestimmt werden.

Zentrale Grenzwertsätze

Man kann die zu untersuchenden Wahrscheinlichkeiten auch schreiben als

$$\mathbb{P}(n\mu + a\sigma\sqrt{n} \leq S_n \leq n\mu + b\sigma\sqrt{n}).$$

Das sind die W. der Fluktuationen der Zufallsvariable S_n um ihren Erwartungswert $n\mu$ von der Größenordnung maximal $\sqrt{n}$.

Alternativ:

$$\mathbb{P}\left(\mu + \frac{a\sigma}{\sqrt{n}} \leq \frac{1}{n}S_n \leq \mu + \frac{b\sigma}{\sqrt{n}}\right)$$

ist die W. der Fluktuationen der Zufallsvariable $\frac{1}{n}S_n$ um ihren Erwartungswert μ von der Größenordnung maximal $\frac{1}{\sqrt{n}}$.

Die vorgelegte Fragestellung ist extrem allgemein: keine konkreten Verteilungsannahmen, lediglich strukturelle Eigenschaften (Unabhängigkeit, gleiche Verteilungen, Existenz des zweiten Moments).

ZGWS von de Moivre-Laplace

Wir untersuchen zunächst den Spezialfall $X_1 \sim \mathrm{Bin}(1, p)$ mit $p \in (0, 1)$ und $q := 1 - p$. Dann gilt $S_n \sim \mathrm{Bin}(n, p)$. Wegen $S_n \in \{0, 1, \ldots, n\}$ gilt für die Standardisierung $\widetilde{S}_n$ von S_n dann

$$\widetilde{S}_n = \frac{S_n - np}{\sqrt{npq}} \in \left\{ -\sqrt{n}\sqrt{\frac{p}{q}}, \ldots, \frac{k - np}{\sqrt{npq}}, \ldots \sqrt{n}\sqrt{\frac{q}{p}} \right\}$$

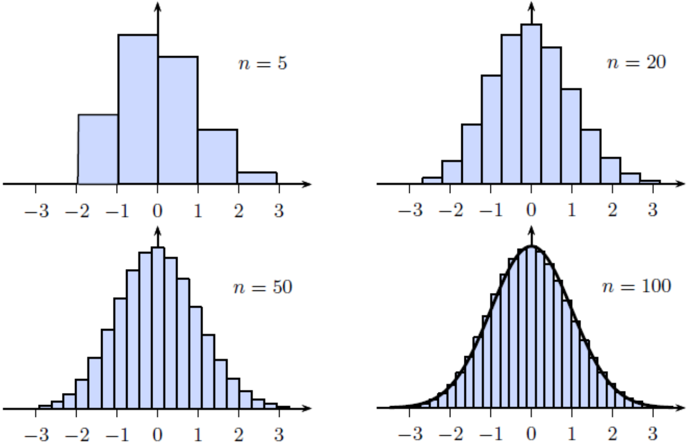
sowie $\mathbb{E}(\widetilde{S}_n) = 0$ und $\mathbb{V}(\widetilde{S}_n) = 1$.

Wir setzen

$$\begin{aligned} t_{k,n} &:= \frac{k - np}{\sqrt{npq}}, && k = 0, \ldots, n, \\ \Delta t_{k,n} &:= t_{k+1,n} - t_{k,n} = \frac{1}{\sqrt{npq}} && k = 0, \ldots, n - 1. \end{aligned}$$

Nachfolgende Folie zeigt Histogramme der Verteilungen der standardisierten ZV $\widetilde{S}_n$ für verschiedene n. Dabei sind die $t_{n,k}$ als Klassenmittelpunkte gewählt.

ZGWS von de Moivre-Laplace



Histogramme standardisierter Binomialverteilungen für $p = 0.3$

ZGWS von de Moivre-Laplace

Wegen

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}, \qquad k \in \{0, \dots, n\},$$

gilt für $-\infty < a \leq b < \infty$ dann

$$\begin{aligned} \mathbb{P}(a < \widetilde{S}_n \leq b) &= \sum_{(*)} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{(*)} \sqrt{npq} \binom{n}{k} p^k q^{n-k} \frac{1}{\sqrt{npq}} \\ &\overset{(1)}{\sim} \sum_{(*)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t_{k,n}^2}{2}} \Delta t_{k,n} \\ &\overset{(2)}{\sim} \int_a^b \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} dt = \int_a^b \varphi(t) dt = \Phi(b) - \Phi(a) \end{aligned}$$

wobei $(*)$ die Summation über alle $k \in \{0, \dots, n\}$ mit $a < \frac{k-np}{\sqrt{npq}} \leq b$ beschreibt, d.h. mit

$$np + a\sqrt{npq} < k \leq np + b\sqrt{npq}.$$

ZGWS von de Moivre-Laplace

Für (2) verwendet man eine Riemann-Summen-Näherung des Integrals. Für (1) zeigt man genauer folgende lokale Aussage:

Satz 20.3 (Lokaler zentraler Grenzwertsatz)

Sei $p \in (0,1)$ und $\nu(n) = o(n^{\frac{2}{3}})$. Dann gilt

$$\sup_{k \in \mathbb{N}_0:\ |k-np| \leq \nu(n)} \left| \frac{\sqrt{npq}\,\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(k-np)^2}{2npq}}} - 1 \right| \to 0 \quad \text{für } n \to \infty.$$

Bemerkung zum Beweis: Wichtiges Hilfsmittel ist die Stirlingsche Formel

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \mathrm{e}^{\frac{\theta_n}{12n}} \text{ mit } \theta_n \in (0,1).$$

(siehe z.B.: Fichtenholz, Differentialrechnung und Integralrechnung, Band 2, VEB Deutscher Verlag der Wissenschaften, Berlin, 1979, S. 384)

Die Schwierigkeit ist, dass man $k!, (n-k)!$ und $n!$ zugleich nähern muss und die Aussagen im betrachteten Bereich für k gleichmäßig gelten müssen.

Zentraler Grenzwertsatz von de Moivre-Laplace

Satz (Zentraler Grenzwertsatz von de Moivre-Laplace)

Sei $S_n \sim \mathrm{Bin}(n,p)$, $0 < p < 1$. Dann gelten für die standardisierten Zufallsvariablen $\widetilde{S}_n := \frac{S_n - np}{\sqrt{np(1-p)}}$:

a) $$\lim_{n\to\infty} \mathbb{P}\left(a \le \widetilde{S}_n \le b\right) = \int_a^b \varphi(t)\,\mathrm{d}t, \quad -\infty < a < b < \infty,$$

b) $$\lim_{n\to\infty} \mathbb{P}\left(\widetilde{S}_n \le x\right) = \int_{-\infty}^x \varphi(t)\,\mathrm{d}t, \quad x \in \mathbb{R}.$$

Bemerkungen:

- in a) und b) kann jedes der $\le$-Zeichen durch das $<$-Zeichen ersetzt werden, ohne den jeweiligen Grenzwert zu ändern.
- a) folgt aus b): $\mathbb{P}(a < \widetilde{S}_n \le b) = \mathbb{P}(\widetilde{S}_n \le b) - \mathbb{P}(\widetilde{S}_n \le a)$, $\mathbb{P}(\widetilde{S}_n = a) \to 0$
- Konvergenzbegriff im Hintergrund: Konvergenz in Verteilung $\widetilde{S}_n \xrightarrow{d} S \sim \mathrm{N}(0,1)$

Praktische Anwendung des ZGWS von de Moivre-Laplace

Sei $S_n \sim \mathrm{Bin}(n,p)$, n groß. Seien $k, \ell \in \mathbb{N}_0$, $0 \leq k < \ell \leq n$.

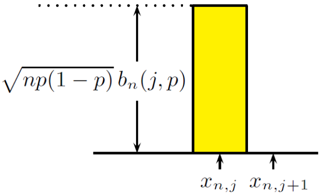
$$\begin{aligned}
\mathbb{P}(k \leq S_n \leq \ell) &= \sum_{j=k}^{\ell} \binom{n}{j} p^j (1-p)^{n-j} \\
&= \mathbb{P}\Bigg(\underbrace{\frac{k-np}{\sqrt{np(1-p)}}}_{=:a} \leq \underbrace{\frac{S_n-np}{\sqrt{np(1-p)}}}_{\widetilde{S}_n} \leq \underbrace{\frac{\ell-np}{\sqrt{np(1-p)}}}_{=:b} \Bigg) \\
&\approx \Phi(b) - \Phi(a)
\end{aligned}$$

Vielfach bessere Näherung (sog. Stetigkeitskorrektur):

$$\begin{aligned}
\mathbb{P}(k \leq S_n \leq \ell) &\approx \Phi\left(\frac{\ell - np + \frac{1}{2}}{\sqrt{np(1-p)}} \right) - \Phi\left(\frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}} \right) \\
&= \Phi\left(b + \frac{1}{2\sqrt{np(1-p)}} \right) - \Phi\left(a - \frac{1}{2\sqrt{np(1-p)}} \right).
\end{aligned}$$

Zur Stetigkeitsapproximation

$$x_{n,j} := \frac{j - np}{\sqrt{np(1-p)}}, \quad b_n(j,p) \ := \ \binom{n}{j} p^j \, (1-p)^{n-j}$$

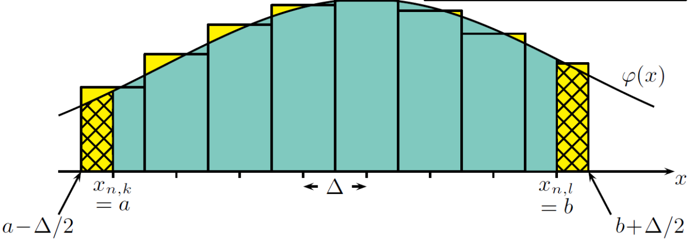


$$x_{n,j+1} - x_{n,j} = \frac{1}{\sqrt{np(1-p)}} =: \Delta$$

Rechteckfläche $= b_n(j,p)$

$\mathbb{P}(k \le S_n \le l) = \sum_{j=k}^{l} b_n(j,p)$

$$\boxed{\int_a^b \varphi(x)\,\mathrm{d}x = \Phi(b) - \Phi(a)}$$



Beispiel

Beispiel (Würfelwurf)

Ein echter Würfel wird $n = 600$ mal geworfen. Es sei S_n die Anzahl der geworfenen Sechsen.

Gesucht: $\mathbb{P}(90 \leq S_n \leq 110)$?

Es ist $p = 1/6$, $S_n \sim \mathrm{Bin}(n, p)$ und damit $np = 100$ und $\sigma_n := \sqrt{np(1-p)} \approx 9.13$.

$$\begin{aligned}
\mathbb{P}(90 \leq S_n \leq 110) &= \mathbb{P}\left(\frac{90-100}{\sigma_n} \leq \frac{S_n - 100}{\sigma_n} \leq \frac{110-100}{\sigma_n}\right) \\
&\approx \Phi\left(\frac{10}{9.13}\right) - \Phi\left(\frac{-10}{9.13}\right) \\
&\approx \Phi(1.1) - (1 - \Phi(1.1)) = 2 \cdot \Phi(1.1) - 1 \\
&\approx 2 \cdot 0.864 - 1 = 0.728.
\end{aligned}$$

Die verbesserte Approximation (Stetigkeitskorrektur) ergibt

$$\mathbb{P}(90 \leq S_n \leq 110) \approx \Phi\left(\frac{10+0.5}{9.13}\right) - \Phi\left(\frac{-10-0.5}{9.13}\right) \approx 2 \cdot \Phi(1.15) - 1 \approx 0.75.$$

Der exakte Wert (Maple) ist 0.753.

Ein allgemeinerer ZGWS

Der bisher behandelte Fall $X_1 \sim \mathrm{Bin}(1, p)$ ist ein Spezialfall des folgenden Satzes.

Satz 20.5 (Lindeberg–Lévy (1922))

Seien $X_1, X_2, \ldots$ u.i.v. Zufallsvariablen mit endlicher Varianz $\sigma^2 = \mathbb{V}(X_1) \in (0, \infty)$ und $\mu = \mathbb{E}(X_1)$. Dann gilt

$$\lim_{n \to \infty} \mathbb{P}(\widetilde{S}_n \leq x) = \Phi(x), \qquad x \in \mathbb{R},$$

wobei $\widetilde{S}_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ mit $S_n = X_1 + \cdots + X_n$ ist.

Ein Beweis und Verallgemeinerungen dieses Satzes werden in der Vorlesung *Wahrscheinlichkeitstheorie* diskutiert. Interessant ist auch die Frage nach der Konvergenzgeschwindigkeit, d.h. nach der Güte der Approximation.

Anwendung

Beispiel

Es seien $X_1, X_2, \dots, X_n$ u.i.v. mit $X_1 \sim \mathrm{G}(p)$, also insbes. $\mathbb{E}X_1 = \frac{1-p}{p}$ und $\mathbb{V}(X_1) = \frac{1-p}{p^2}$.

Dann gilt offenbar $S_n = X_1 + \cdots + X_n \sim \mathrm{Nb}(n, p)$, sowie $\mathbb{E}S_n = n\frac{1-p}{p}$ und $\mathbb{V}(S_n) = n\frac{1-p}{p^2}$.

Es folgt

$$\lim_{n\to\infty} \mathbb{P}\left(S_n \le \frac{1-p}{p}\, n + \frac{\sqrt{1-p}}{p}\sqrt{n} \cdot x\right) = \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\frac{1-p}{p}}{\frac{\sqrt{1-p}}{p}\sqrt{n}} \le x\right) \;=\; \Phi(x)$$

für $x \in \mathbb{R}$.

Kapitel 21: Induktive Statistik - Punktschätzung

Bislang:

- $(\Omega, \mathcal{A}, \mathbb{P})$ ist W.Raum, $\mathbb{P}$ bekannt/gegeben,
- $\omega \in \Omega$ ist Ergebnis eines stochastischen Vorgangs, $\mathbb{P}$ „bewertet/steuert das Auftreten von Daten“,
- ein n-dimensionaler Zufallsvektor $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ ergibt Daten in Form von Realisierungen $x = (x_1, \ldots, x_n) = X(\omega)$ von X.
- Wir bestimmen W. von Ereignissen, etwa $\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$ in Abhängigkeit von x.

Im Folgenden:

- $\mathbb{P}$ ist nicht vollständig bekannt.
- Aufgrund von beobachteten Daten $x = (x_1, \ldots, x_n)$ ist eine begründete Aussage über ein zugrunde liegendes $\mathbb{P}$ zu treffen.

21.1 Allgemeiner Modellrahmen

- Sei $\Theta \neq \emptyset$ (sogenannter Parameterraum),
- sei $\mathcal{X} \neq \emptyset$ (sogenannter Stichprobenraum),
- für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta$ ein W.maß auf $\mathcal{P}(\mathcal{X})$ bzw. einer geeigneten σ-Algebra $\mathcal{A}$.
- Die Zuordnung $\Theta \ni \vartheta \mapsto \mathbb{P}_\vartheta$ sei injektiv.
- Das Tripel $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ heißt statistisches Modell.

Im Folgenden:

- $\Theta \subseteq \mathbb{R}^s$ für $s \geq 1$ (vorläufig: $s = 1$), $\mathcal{X} \subseteq \mathbb{R}^n$,
- $x = (x_1, \ldots, x_n) \in \mathcal{X}$ wird aufgefasst als Realisierung eines Zufallsvektors $X = (X_1, \ldots, X_n)$ (zumeist mit unabhängigen Komponenten). Kanonische Konstruktion: $\Omega := \mathcal{X}$, $X = \mathrm{id}_\Omega$
- Diskrete Verteilung: $x \mapsto \mathbb{P}_\vartheta(\{x\}) = \mathbb{P}_\vartheta(X = x) = f_\vartheta(x)$ ist Zähldichte.
- Stetige Verteilung: $x \mapsto f_\vartheta(x)$ ist L-Dichte zu $\mathbb{P}_\vartheta$.

Statistisches Modell für Bernoulli-Kette

Beispiel (Bernoulli-Schema)

- Parameterraum: $\Theta \; := \; (0,1)$,
- Stichprobenraum: $\mathcal{X} \; := \; \{0,1\}^n$,
- $X = (X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ unabhängig und je $\sim \text{Bin}(1, \vartheta)$
- Für $x = (x_1 \ldots, x_n) \in \mathcal{X}$ gilt dann

$$\mathbb{P}_\vartheta(X = x) \; = \; \vartheta^k (1-\vartheta)^{n-k}, \quad k := x_1 + \cdots + x_n.$$

- Unter $\mathbb{P}_\vartheta$ gilt: $S_n := X_1 + \cdots + X_n \sim \text{Bin}(n, \vartheta)$, d.h.

$$\mathbb{P}_\vartheta(S_n = k) \; = \; \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Punktschätzer

Definition 21.1

Es sei $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell. Ein (Punkt-)Schätzer für ϑ ist eine (messbare) Abbildung $T : \mathcal{X} \to \widetilde{\Theta}$, wobei $\widetilde{\Theta} \supseteq \Theta$. Der Wert $T(x)$ heißt konkreter Schätzwert für ϑ zu $x \in \mathcal{X}$.

Beachte:

- Häufig ist $\Theta = \widetilde{\Theta}$. Aus mathematischen Gründen ist manchmal eine echte Obermenge nötig.
- Allgemeiner: Manchmal sind Schätzer für $\gamma(\vartheta)$ gesucht, wobei $\gamma : \Theta \to \mathbb{R}^k$ eine (messbare) Funktion der Parameter ist.

Beispiel (Bernoulli-Schema, Forts.)

In der Situation des obigen Beispiels ist $T : \mathcal{X} \to \widetilde{\Theta} := [0, 1]$ mit

$$T(x) := \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad x = (x_1, \ldots, x_n) \in \{0, 1\}^n = \mathcal{X}$$

ein Schätzer für ϑ. Hier wird meist $\Theta := (0, 1)$ angenommen.

Bemerkungen zu Punktschätzern

- Allgemein heißt eine auf $\mathcal{X}$ definierte Abbildung Stichprobenfunktion oder Statistik.
- Das W.maß $\mathbb{P}_\vartheta$ „steuert“ das Auftreten von x und damit von $T(x)$.
- T ist (im Prinzip) eine (beliebige) auf $\mathcal{X}$ definierte Zufallsvariable mit Werten in $\widetilde{\Theta}$.
- Die Verteilung von T unter $\mathbb{P}_\vartheta$ ist im diskreten Fall bestimmt durch

$$t \mapsto \mathbb{P}_\vartheta(T = t) \ = \ \mathbb{P}_\vartheta(\{x \in \mathcal{X} : T(x) = t\}).$$

- Ideal wäre $\mathbb{P}_\vartheta(T = \vartheta) = 1$ für jedes $\vartheta \in \Theta$, wenn T Schätzer für ϑ.
- Realistischer und wünschenswert ist $\mathbb{P}_\vartheta(|T - \vartheta)| \leq \varepsilon) \approx 1$ für kleines ε, für jedes $\vartheta \in \Theta$.
  Die Verteilung von T sollte also stark um den unbekannten Wert ϑ konzentriert sein, falls tatsächlich $\mathbb{P}_\vartheta$ zugrunde liegt.

Mittlere quadratische Abweichung und Verzerrung

Zur Beurteilung der Güte von Schätzern sind folgende Größen nützlich.

Definition 21.2 (Mittlere quadratische Abweichung, Verzerrung)

Sei $T : \mathcal{X} \to \widetilde{\Theta}$ mit $\widetilde{\Theta} \subseteq \mathbb{R}$. Es gelte

$$\mathbb{E}_\vartheta(T^2) \ < \ \infty, \qquad \vartheta \in \Theta.$$

a) $\mathrm{MQA}_T(\vartheta) \ := \ \mathbb{E}_\vartheta \left[(T - \vartheta)^2 \right]$ heißt mittlere quadratische Abweichung von T an der Stelle ϑ.

b) $\mathrm{b}_T(\vartheta) \ := \ \mathbb{E}_\vartheta(T) - \vartheta$ heißt Verzerrung (engl.: bias) von T an der Stelle ϑ.

c) T heißt erwartungstreu, falls $\mathbb{E}_\vartheta(T) = \vartheta$ (d.h., $\mathrm{b}_T(\vartheta) = 0$) für $\vartheta \in \Theta$ gilt.

Zur Güte von Schätzern

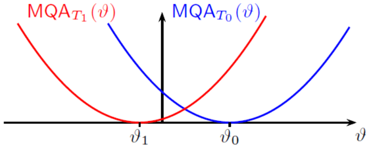
Wünschenswert ist eine möglichst kleine mittlere quadratische Abweichung $\mathsf{MQA}_T(\vartheta) := \mathbb{E}_\vartheta\left[(T-\vartheta)^2\right]$

- Nach dem Verschiebungssatz (d.h. Satz 12.3 (c) mit $X = T$, $t = \vartheta$) gilt

$$\mathsf{MQA}_T(\vartheta) \;=\; \mathbb{V}_\vartheta(T) + (\mathsf{b}_T(\vartheta))^2 \,.$$

- Sei $\vartheta_0 \in \Theta$ fest. Für den unsinnigen Schätzer $T_0(x) := \vartheta_0$ für $x \in \mathcal{X}$ (ignoriert die Daten!) gilt $\mathbb{V}_\vartheta(T_0) = 0$, $\vartheta \in \Theta$, sowie $\mathbb{E}_\vartheta(T_0) = \vartheta_0$, $\vartheta \in \Theta$,

$$\Longrightarrow \quad \mathsf{MQA}_{T_0}(\vartheta) \;= (\mathsf{b}_T(\vartheta))^2 = (\vartheta_0 - \vartheta)^2$$



Man schließt solche extremen Schätzer aus, z.B. durch Einschränkung auf erwartungstreue Schätzer.

Ein Schätzer für den Binomialfall

$$\mathrm{MQA}_T(\vartheta) := \mathbb{E}_\vartheta\left[(T-\vartheta)^2\right], \quad \mathrm{b}_T(\vartheta) = \mathbb{E}_\vartheta(T) - \vartheta$$

Beispiel (Binomial-Fall)

Seien $X_1, \ldots, X_n$ unabhängig und je $\mathrm{Bin}(1, \vartheta)$-verteilt, wobei $\vartheta \in \Theta := [0,1]$, $\mathcal{X} = \{0,1\}^n$. Wir betrachten den Schätzer

$$T_n(x_1, \ldots, x_n) \ := \ \frac{1}{n}\sum_{j=1}^{n} x_j, \quad x = (x_1, \ldots, x_n) \in \mathcal{X}.$$

Sei $T_n^* := T_n(X_1, \ldots, X_n)$. Dann gilt

$$\mathbb{E}_\vartheta(T_n^*) \ = \ \vartheta \ \text{ für } \ \vartheta \in \Theta \ \text{ und somit } \ \mathrm{b}_{T_n^*}(\vartheta) \ = \ 0 \ \text{ für } \ \vartheta \in \Theta$$

Also ist T_n erwartungstreu. Weiter gilt

$$\mathbb{V}_\vartheta(T_n^*) \ = \ \frac{\vartheta(1-\vartheta)}{n} \ \text{ für } \ \vartheta \in \Theta, \ \text{ also } \ \mathrm{MQA}_{T_n^*}(\vartheta) \ = \ \frac{\vartheta(1-\vartheta)}{n}.$$

21.2 Maximum-Likelihood-Methode

Definition 21.3 (Maximum-Likelihood-Schätzung, diskreter Fall)

Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\mathbb{P}_\vartheta$ diskret. Sei $x \in \mathcal{X}$.

- Die Funktion
$$L_x : \Theta \to [0,1], \quad \vartheta \mapsto L_x(\vartheta) := \mathbb{P}_\vartheta(X = x) = \mathbb{P}_\vartheta(\{x\})$$
heißt Likelihood-Funktion (Plausibilitätsfunktion) für ϑ zur Beobachtung $X = x$.
- Existiert ein $\widehat{\vartheta}(x) \in \Theta$ (evtl. $\widehat{\vartheta}(x) \in \widetilde{\Theta} \supseteq \Theta$) mit
$$L_x\left(\widehat{\vartheta}(x)\right) = \sup_{\vartheta\in\Theta} L_x(\vartheta), \tag{4}$$
so heißt $\widehat{\vartheta}(x)$ Maximum-Likelihood-Schätzwert (kurz: ML-Schätzwert) für ϑ zu x.
- Ein Schätzer $\widehat{\vartheta} : \mathcal{X} \to \widetilde{\Theta}$ mit (4) für jedes $x \in \mathcal{X}$ heißt Maximum-Likelihood-Schätzer (kurz: ML-Schätzer) für ϑ.

Log-Likelihood-Funktion

Für $x \in \mathcal{X}$ heißt die durch

$$\Theta \ni \vartheta \mapsto \ln L_x(\vartheta) \;=\; \ln \mathbb{P}_\vartheta(X = x)$$

definierte Funktion $\ln L_x$ die Log-Likelihood-Funktion zu x.

Beachte: $\ln L_x(\cdot)$ und $L_x(\cdot)$ nehmen Maxima an der gleichen Stelle an!

Vorteilhaft, wenn Θ Intervall und L_x differenzierbar!

Falls $X = (X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ stochastisch unabhängig, so

$$\begin{aligned} L_x(\vartheta) \;&=\; \mathbb{P}_\vartheta(X = x) \;=\; \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^{n} \mathbb{P}_\vartheta(X_j = x_j) \\ \Rightarrow\; \ln L_x(\vartheta) \;&=\; \sum_{j=1}^{n} \ln \mathbb{P}_\vartheta(X_j = x_j). \end{aligned}$$

Summen sind oft leichter zu differenzieren als ein Produkt!

ML-Schätzung bei geometrischer Verteilung

Beispiel

Seien $X_1, \ldots, X_n$ unabhängig und je $\mathrm{G}(\vartheta)$-verteilt, wobei $\vartheta \in \Theta = (0,1)$.

Aufgabe: ML-Schätzung von ϑ aufgrund von $X_1, \ldots, X_n$.

Seien $x_1, \ldots, x_n \in \mathbb{N}_0$, $x := (x_1, \ldots, x_n)$, $X := (X_1, \ldots, X_n)$. Dann ist

$$L_x(\vartheta) = \mathbb{P}_\vartheta(X = x) = \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n),$$

also

$$L_x(\vartheta) = \prod_{j=1}^n \mathbb{P}_\vartheta(X_j = x_j) \;=\; \prod_{j=1}^n \left\{ (1-\vartheta)^{x_j} \vartheta \right\} \;=\; \vartheta^n (1-\vartheta)^{x_1 + \cdots + x_n} \;\Rightarrow \ln L_x(\vartheta) = n \ln \vartheta + \left(\sum_{j=1}^n x_j \right) \ln(1-\vartheta)$$

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta} \ln L_x(\vartheta) \;=\; \frac{n}{\vartheta} - \frac{1}{1-\vartheta} \sum_{j=1}^n x_j \;\overset{!}{=}\; 0 \iff \widehat{\vartheta}(x) = \frac{1}{1 + \frac{1}{n} \sum_{j=1}^n x_j}.$$

Der ML-Schätzer für ϑ ist also

$$\widehat{\vartheta}_n \;:=\; \frac{1}{1 + \overline{X}_n}, \text{ wobei } \quad \overline{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Schätzfolgen

Definition 21.4 (Schätzfolge, asymptotische E-Treue, Konsistenz)

Seien $X_1, X_2, \dots$ unabhängig und identisch verteilt mit Verteilung $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$, auf $(\Omega, \mathcal{A})$.

Sei $\mathcal{X} \subseteq \mathbb{R}$ der Stichprobenraum für jedes einzelne X_j und $\mathcal{X}^n \subseteq \mathbb{R}^n$ der Stichprobenraum für $(X_1, \dots, X_n)$.

Für $n \in \mathbb{N}$ sei $T_n : \mathcal{X}^n \to \widetilde{\Theta} \supseteq \Theta$ ein Schätzer für ϑ. Dann heißt $(T_n)_{n\geq 1}$ eine Schätzfolge.

Beachte: $T_n^* := T_n(X_1, \dots, X_n)$ ist eine Zufallsvariable.

Die Schätzfolge (T_n) heißt

- asymptotisch erwartungstreu für ϑ, falls

$$\lim_{n\to\infty} \mathbb{E}_\vartheta\left(T_n^*\right) = \vartheta \quad \forall \vartheta \in \Theta,$$

- konsistent für ϑ, falls

$$\lim_{n\to\infty} \mathbb{P}_\vartheta\left(|T_n^* - \vartheta| \geq \varepsilon\right) = 0 \quad \forall \varepsilon > 0,\ \forall \vartheta \in \Theta.$$

Fortsetzung des Beispiels zur geometrischen Verteilung

Beispiel

Seien $X_1, X_2, \dots$ unabhängig und je G(ϑ)-verteilt. Der zuvor ermittelte auf $X_1, \dots, X_n$ basierende ML-Schätzer war

$$\widehat{\vartheta}_n \;:=\; \frac{1}{1+\overline{X}_n}, \qquad \text{mit } \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Nach dem Gesetz großer Zahlen gilt

$$\lim_{n\to\infty} \mathbb{P}_\vartheta(|\overline{X}_n - \mathbb{E}_\vartheta(X_1)| \geq \varepsilon) = 0 \text{ mit } \; \mathbb{E}_\vartheta(X_1) = \frac{1-\vartheta}{\vartheta} \;=:\; a, \qquad \vartheta = \frac{1}{1+a}.$$

Da die Funktion $g(t) := \frac{1}{1+t}$, $t \geq 0$, stetig ist, folgt

$$\lim_{n\to\infty} \mathbb{P}_\vartheta(|\widehat{\vartheta}_n - \vartheta| \geq \varepsilon) = \lim_{n\to\infty} \mathbb{P}_\vartheta(|g(\overline{X}_n) - g(\mathbb{E}_\vartheta(X_1))| \geq \varepsilon) = 0$$

Somit ist die Schätzfolge $(\widehat{\vartheta}_n)$ konsistent für ϑ.

Man kann zeigen: Der Schätzer $\widehat{\vartheta}_n$ ist nicht erwartungstreu für ϑ (genauer: ϑ wird systematisch überschätzt), die Schätzfolge $(\widehat{\vartheta}_n)$ ist aber asymptotisch erwartungstreu mit $\mathbb{E}(\widehat{\vartheta}_n) \in [\vartheta, \frac{n}{n-1}\vartheta]$ für $n \geq 2$.

Maximum-Likelihood-Schätzung, stetiger Fall

Definition 21.5

Sei $(\mathcal{X}, \mathcal{A}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathbb{P}_\vartheta = f_\vartheta \lambda^n$. Sei $x \in \mathcal{X}$.

- Die Funktion
$$L_x : \Theta \to [0, \infty), \quad \vartheta \mapsto L_x(\vartheta) := f_\vartheta(x)$$
heißt Likelihood-Funktion für ϑ zur Beobachtung $X = x$.

- Existiert ein $\widehat{\vartheta}(x) \in \Theta$ (evtl. $\widehat{\vartheta}(x) \in \widetilde{\Theta} \supseteq \Theta$) mit
$$L_x\left(\widehat{\vartheta}(x)\right) = \sup_{\vartheta\in\Theta} L_x(\vartheta), \tag{5}$$
so heißt $\widehat{\vartheta}(x)$ Maximum-Likelihood-Schätzwert (kurz: ML-Schätzwert) für ϑ zu x.

- Ein Schätzer $\widehat{\vartheta} : \mathcal{X} \to \widetilde{\Theta}$ mit (5) für $x \in \mathcal{X}$ heißt Maximum-Likelihood-Schätzer (kurz: ML-Schätzer) für ϑ.

Maximum-Likelihood-Schätzung, stetiger Fall

- Beachte, dass f_ϑ keine Wahrscheinlichkeit ist.
- Im Fall, dass die Beobachtungen als Realisierungen von n unabhängigen, gleichverteilten Zufallsvariablen $X_1, \dots, X_n$ zustande kommen, gilt

$$f_\vartheta(x) = \prod_{i=1}^{n} f_\vartheta(x_i), \qquad x = (x_1, \dots, x_n).$$

- Man kann den diskreten und stetigen Fall einheitlich behandeln und als Sonderfälle von Verteilungsfamilien ansehen, die von einem gemeinsamen Maß dominiert werden. Im diskreten Fall ist dieses dominierende Maß das Zählmaß, im stetigen Fall das Lebesguemaß.
- Man geht dann ganz analog wie im diskreten Fall vor und betrachtet häufig vorteilhaft die Log-Likelihood-Funktion zu f_ϑ.

Beispiel: Exponential-Verteilung

Beispiel

Für stochastisch unabhängige, exponentialverteilte Stichprobenvariablen $X_1, \ldots, X_n \sim Exp(\vartheta)$ mit $\vartheta > 0$ ist $\hat{\vartheta}_n(x) \ := \ \bar{x}^{-1}$ der Maximum-Likelihood-Schätzer für ϑ (mit $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$). Wir setzen hier $\bar{x} > 0$ voraus.

Nachweis: (zur Erinnerung: $f_\vartheta(t) = \vartheta e^{-\vartheta t}$)

$$\begin{aligned} L_x(\vartheta) &= \prod_{i=1}^{n} f_\vartheta(x_i) = \vartheta^n e^{-\vartheta \sum_{i=1}^{n} x_i}, \\ M_x(\vartheta) &= \ln L_x(\vartheta) = n \ln \vartheta - \vartheta \sum_{i=1}^{n} x_i, \\ M_x'(\vartheta) &= \frac{n}{\vartheta} - \sum_{i=1}^{n} x_i, \qquad M_x''(\vartheta) = -\frac{n}{\vartheta^2} < 0. \end{aligned}$$

Hieraus leitet man die Behauptung leicht ab.

Man kann zeigen: $\hat{\vartheta}_n$ ist zwar nicht erwartungstreu, aber asymptotisch erwartungstreu.

Beispiel: Normalverteilte Stichprobenvariablen

Beispiel

Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ stochastisch unabhängig mit $\vartheta = (\mu, \sigma^2) \in \Theta := \mathbb{R} \times (0, \infty)$. Dann ist

$$
\begin{aligned}
L_x(\vartheta) = \prod_{j=1}^{n} f_\vartheta(x_j) \quad &= \quad \prod_{j=1}^{n} \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_j - \mu)^2}{2\sigma^2} \right) \right\} = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot \sum_{j=1}^{n} (x_j - \mu)^2 \right), \\
M_x(\vartheta) \quad &= \quad C - n \ln \sigma - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^{n} (x_j - \mu)^2 .
\end{aligned}
$$

Definiere für festes σ^2 die Funktion $h(\mu) := C - n \ln \sigma - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^{n} (x_j - \mu)^2$. Dann gilt

$$
\begin{aligned}
h'(\mu) \quad &= \quad \frac{1}{\sigma^2} \cdot \sum_{j=1}^{n} (x_j - \mu) = \frac{n}{\sigma^2} \frac{1}{n} \sum_{j=1}^{n} x_j - \frac{1}{\sigma^2} \sum_{j=1}^{n} \mu \\
&= \quad \frac{n}{\sigma^2} \overline{x} - \frac{n}{\sigma^2} \mu \overset{!}{=} 0 \quad \Longleftrightarrow \quad \mu = \overline{x}.
\end{aligned}
$$

Bei festem σ^2 ist M_x maximal für $\mu = \bar{x}$ (denn h' wechselt dort passend das Vorzeichen!).

Beispiel: Normalverteilte Stichprobenvariablen

Beispiel (Fortsetzung)

Bleibt die Aufgabe, mit $s := \sigma^2$ die Funktion

$$g(s) := C - \frac{n}{2} \ln s - \frac{1}{2s} \cdot \sum_{j=1}^{n} (x_j - \overline{x})^2$$

zu maximieren.

$$g'(s) \quad = \quad -\frac{n}{2s} + \frac{1}{2s^2} \sum_{j=1}^{n} (x_j - \overline{x})^2 \overset{!}{=} 0 \quad \Longleftrightarrow \quad s = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2$$

g ist maximal für $s = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2$ (denn g' wechselt dort passend das Vorzeichen!). Also ist

$$\widehat{\vartheta}(x) \quad = \quad \left(\bar{x}, \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 \right) = \left(\bar{x}, \frac{n-1}{n} \cdot s_x^2 \right)$$

der Maximum-Likelihood-Schätzwert von ϑ.

Beachte: Bei der ML-Schätzung der Varianz unter Normalverteilung wird durch n und nicht durch $n-1$ dividiert!

Beispiel: Gleichverteilung

Die „Ableitungsmethode“ versagt, wenn die Funktion M_x nicht stetig ist.

Beispiel (Gleichverteilung)

Unabhängige Stichproben: $X_1, \ldots, X_n \sim \mathrm{U}_{(0,\vartheta)}$ mit $\vartheta \in \Theta := (0, \infty)$.

Erinnerung: $f_\vartheta(t) = \frac{1}{\vartheta} 1_{(0,\vartheta)}(t)$ ist die Dichte der $\mathrm{U}_{(0,\vartheta)}$-Verteilung.

Bei Vorliegen der Stichprobe $x = (x_1, \ldots, x_n)$ mit $x_j > 0$ ist

$$L_x(\vartheta) = \prod_{i=1}^{n} f_\vartheta(x_i) = \left\{ \begin{array}{cl} 0 & , \ \text{falls } \vartheta < \max\{x_1, \ldots, x_n\} \\ \frac{1}{\vartheta^n} & , \ \text{falls } \vartheta \geq \max\{x_1, \ldots, x_n\} \end{array} \right\} = \frac{1}{\vartheta^n} \mathbb{1}_{[\max(x_1,\ldots,x_n),\infty)}(\vartheta)$$

an der Stelle $\widehat{\vartheta}(x) \ = \ \max\{x_1, \ldots, x_n\}$ unstetig und hat dort das absolute Maximum.

21.3 Die Momentenmethode

Sei $\vartheta = (\vartheta_1, \ldots, \vartheta_s)$ der unbekannte, s-dimensionale Parameter. Schätze das k-te Moment

$$m_k(\vartheta) := \mathbb{E}_\vartheta(X_1^k) \qquad \text{für } k = 1, \ldots, s$$

durch das k-te Stichprobenmoment

$$\widehat{m}_k(x) = \frac{1}{n} \sum_{j=1}^{n} x_j^k .$$

Definition 21.6 (Momentenschätzer)

Der Momentenschätzer

$$\widehat{\vartheta}(x) = (\widehat{\vartheta}_1(x), \ldots, \widehat{\vartheta}_s(x))$$

ergibt sich durch Auflösen der s Gleichungen

$$m_k(\vartheta) = \widehat{m}_k(x), \quad k = 1, \ldots, s,$$

nach ϑ.

Beispiel: Gammaverteilung

Beispiel

Sei $\mathbb{P}_\vartheta = \Gamma(\alpha, \beta)$ für $\vartheta = (\alpha, \beta) \in \Theta := (0, \infty)^2$. Das 1. und 2. Moment sind

$$m_1(\alpha, \beta) = \frac{\alpha}{\beta}, \qquad m_2(\alpha, \beta) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

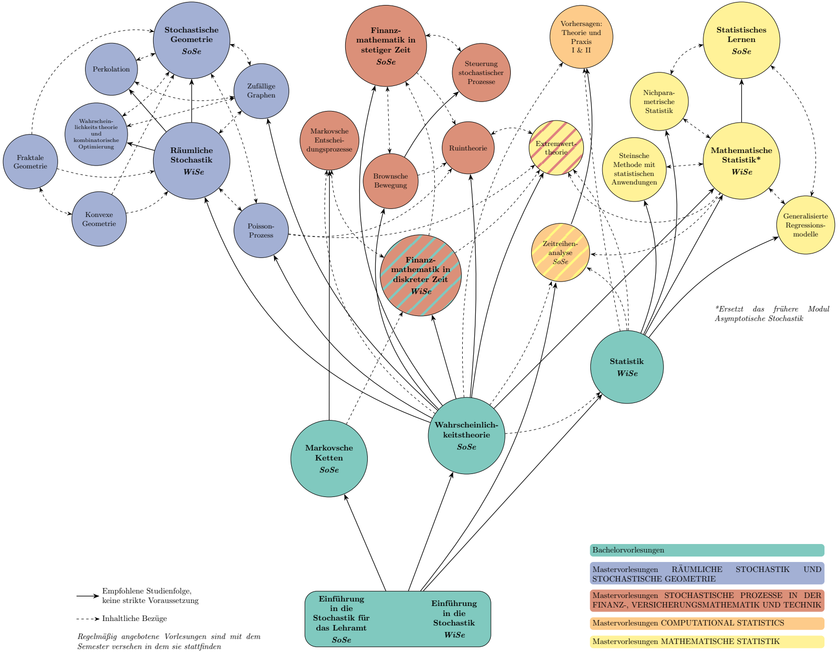
Resultierende („Schätz“-)Gleichungen:

$$\frac{\alpha}{\beta} \overset{!}{=} \overline{x}, \qquad \frac{\alpha(\alpha+1)}{\beta^2} \overset{!}{=} \frac{1}{n}\sum_{j=1}^{n} x_j^2.$$

Nach einigen Umformungen folgt

$$\widehat{\beta}(x) = \frac{\overline{x}}{\tilde{s}_x^2}, \qquad \widehat{\alpha}(x) = \frac{(\overline{x})^2}{\tilde{s}_x^2}$$

mit $\tilde{s}_x^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2 = \frac{n-1}{n}\, s_x^2.$



Stochastische Geometrie ***SoSe***

Perkolation

Zufällige Graphen

Wahrscheinlichkeitstheorie und kombinatorische Optimierung

Fraktale Geometrie

Räumliche Stochastik ***WiSe***

Konvexe Geometrie

Poisson-Prozess

Finanzmathematik in stetiger Zeit ***SoSe***

Steuerung stochastischer Prozesse

Markovsche Entscheidungsprozesse

Brownsche Bewegung

Ruintheorie

Finanzmathematik in diskreter Zeit ***WiSe***

Vorhersagen: Theorie und Praxis I & II

Extremwerttheorie

Zeitreihenanalyse *SoSe*

Statistisches Lernen ***SoSe***

Nichtparametrische Statistik

Steinsche Methode mit statistischen Anwendungen

Mathematische Statistik* ***WiSe***

Generalisierte Regressionsmodelle

**Ersetzt das frühere Modul Asymptotische Stochastik*

Statistik ***WiSe***

Markovsche Ketten ***SoSe***

Wahrscheinlichkeitstheorie ***SoSe***

Einführung in die Stochastik für das Lehramt ***SoSe***

Einführung in die Stochastik ***WiSe***

⟶ Empfohlene Studienfolge, keine strikte Voraussetzung

⇢ Inhaltliche Bezüge

Regelmäßig angebotene Vorlesungen sind mit dem Semester versehen in dem sie stattfinden

Bachelorvorlesungen

Mastervorlesungen RÄUMLICHE STOCHASTIK UND STOCHASTISCHE GEOMETRIE

Mastervorlesungen STOCHASTISCHE PROZESSE IN DER FINANZ-, VERSICHERUNGSMATHEMATIK UND TECHNIK

Mastervorlesungen COMPUTATIONAL STATISTICS

Mastervorlesungen MATHEMATISCHE STATISTIK