Einführung in die Stochastik
für Studierende des Lehramts Mathematik
Sommersemester 2015

Norbert Henze, Institut für Stochastik, Email: Norbert.Henze@kit.edu

Karlsruher Institut für Technologie

Wichtige Hinweise:

- Hörerkreis: Lehramtsstudierende Mathematik ab dem 4. Semester
- Umfang der Lehrveranstaltung: 4 V + 2 Ü + 2 T
- Vorlesung: Di 8.00 - 9.30 Uhr, Chemie-HS III, Fr 9.45-11.15 Uhr, Hertz-HS
- Übung: Mi 15.45-17.15 Uhr, Redtenbacher HS
- Tutorien in Kleingruppen, Anmeldung über ILIAS

  Di 11.30-13.00 NH, Di 17.30-19.00 R 2.067, Mi 9.45-11.15 R -1.013
- Übungsblätter und Lösungen auf Lehrplattform ILIAS
- Hauptbegleittext: Stochastik für Einsteiger
- Sprechstd.: Mo 10-11 und n.V.
- Klausuren: 28. Juli 2015 ab 9.00 Uhr (getrennte Anmeldung!)
  - Einführung in die Stochastik für LA: 9-11 Uhr (6LP)
  - Ergänzungen zu Einführung in die Stochastik für LA: 12 -13 Uhr (3LP)

Inhaltsverzeichnis eines Schulbuchs für Gymnasien (Auszug):

Simulationen, Zufallszahlen, Laplace-Versuch, Pfadregeln,
empirisches Gesetz der großen Zahlen,
Prognoseintervalle für relative Häufigkeiten,
Ereignisse und Ereigniswahrscheinlichkeiten, Urnenmodelle,
bedingte Wahrscheinlichkeit, stochastische Unabhängigkeit,
Umgang mit Daten (u.a. Regressionsgerade und Korrelationskoeffizient),
Zufallsgrößen, Erwartungswert und Standardabweichnung,
Binomialverteilung, Bernoulli-Kette, geometrische Verteilung,
stetige Zufallsgrößen, Normalverteilung, Zentraler Grenzwertsatz,
Sigma-Regeln, Schätzen von Anteilen, Konfidenzintervalle,
Testen von Hypothesen, Signifkanztest, Operationscharakteristik,
Vier-Felder-Test, Chi-Quadrat-Anpassungstest

Stochastik: Die Lehre von den Gesetzmäßigkeiten des Zufalls

στόχοσ	das Ziel, die Mutmaßung
στοχαστικός	scharfsinnig im Vermuten
στοχάςομαι	etwas erraten, erkennen, beurteilen

- Wahrscheinlichkeitstheorie
- deskriptive (beschreibende) Statistik
- induktive (schließende) Statistik

Inhalt

Inhalt

1 Grundräume und Ereignisse

1.1 Definition (Grundraum)

Sei $\Omega := \{\omega_1, \omega_2, \ldots\}$ eine abzählbare Menge.

Ω heißt (elementarer) *Grundraum* (*Ergebnisraum*, *Merkmalraum*).

Ω steht für die Menge der Ergebnisse eines stochastischen Vorgangs.

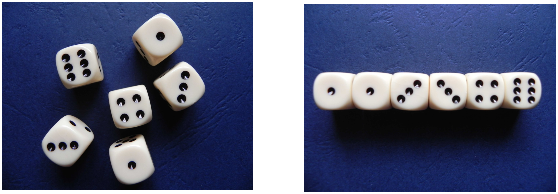
Sprechweise: $\omega \in \Omega$ heißt *Ergebnis.*

1.2 Beispiel (n-facher Würfelwurf)

$$\begin{aligned} \Omega \quad &:= \quad \{(a_1, \ldots, a_n) : a_j \in \{1, \ldots, 6\} \text{ für } j = 1, \ldots, n\} \\ &= \quad \{1, 2, 3, 4, 5, 6\}^n \end{aligned}$$

a_j beschreibt das Ergebnis des j-ten Wurfs.

1.3 Beispiel (n nicht unterscheidbare Würfel gleichzeitig werfen)



$$\Omega \ := \ \{(b_1, \ldots, b_n) : 1 \leq b_1 \leq b_2 \leq \ldots \leq b_n \leq 6\}$$

b_j steht für die j-kleinste Augenzahl.

1.4 Definition (Ereignis)

Jede Teilmenge A von Ω heißt *Ereignis*.

Ω heißt *sicheres Ereignis*.

$\emptyset$ heißt *unmögliches Ereignis*.

$\{\omega\}$ heißt *Elementarereignis*, $\omega \in \Omega$.

Übliche Buchstaben für Ereignisse: $A, B, C, D, A_1, A_2, \ldots$.

Sprechweise: „A tritt ein“ $\Longleftrightarrow \omega \in A$

1.5 Beispiel (n-facher Würfelwurf)

Ereignis *verbal*: „Es tritt mindestens eine Sechs auf“.

Ereignis als Teilmenge von $\Omega = \{1, \ldots, 6\}^n$:

$$A = \Big\{(a_1, \ldots, a_n) \in \Omega : \max_{j=1\ldots,n} a_j = 6\Big\}$$

1.6 Mengentheoretische Verknüpfungen von Ereignissen

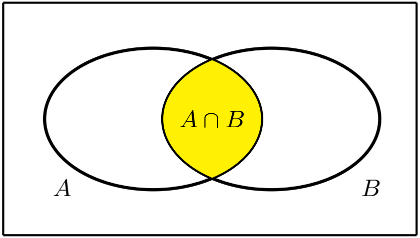
Seien $A, B \subseteq \Omega$ Ereignisse.

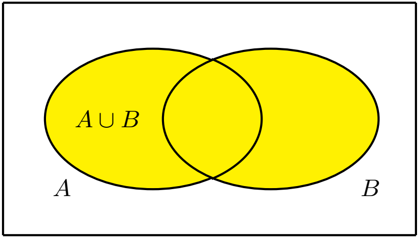
$$A \cap B \quad = \quad \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\} \quad (\textit{Durchschnitt} \text{ von } A \text{ und } B)$$

$verbal$: A *und* B treten beide ein

$$A \cup B \quad = \quad \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\} \quad (\textit{Vereinigung} \text{ von } A \text{ und } B)$$

$verbal$: A *oder* B tritt ein (evtl. beide!)





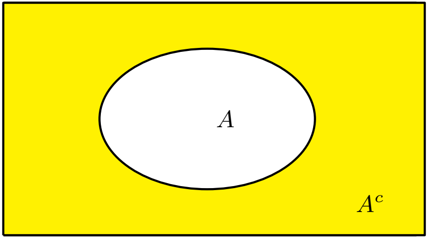
$$A^c = \{\omega \in \Omega : \omega \notin A\}$$ (*Komplement* von A)

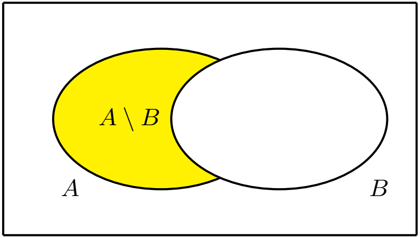
in Worten : A tritt *nicht* ein

$$A \backslash B = A \cap B^c$$ ((*Mengen*)-*Differenz* von A und B)

in Worten : A tritt ein *und* (*aber*) B *nicht*

(„A ohne B“)



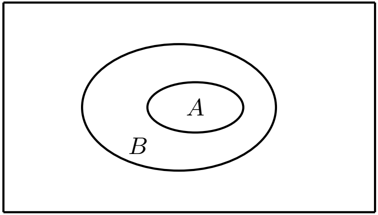


Im Fall $A \subseteq B$ Sprechweisen:

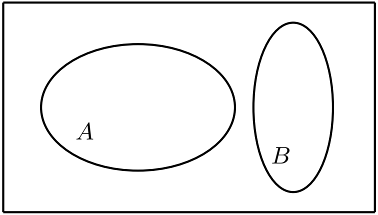
- „Aus A folgt B“.
- „Wenn A eintritt, so auch B“.
- „Das Eintreten von A zieht das Eintreten von B nach sich“.

Im Fall $A \cap B = \emptyset$ Sprechweisen:

- „A und B *disjunkt*“,
- „A und B *unvereinbar*“,
- „A und B schließen sich gegenseitig aus“.



$A \subseteq B$



$A \cap B = \emptyset$

Seien $A_1, A_2, \ldots \subseteq \Omega$ Ereignisse

$A_1 \cap \ldots \cap A_n$	*verbal*: *jedes* der Ereignisse $A_1, \ldots, A_n$ tritt ein
$\bigcap_{j=1}^{\infty} A_j$	*verbal*: *jedes* der Ereignisse $A_1, A_2, \ldots$ tritt ein
$A_1 \cup \ldots \cup A_n$	*verbal*: *mindestens eines* der Ereignisse $A_1, \ldots, A_n$ tritt ein
$\bigcup_{j=1}^{\infty} A_j$	*verbal*: *mindestens eines* der Ereignisse $A_1, A_2, \ldots$ tritt ein

$A_1, A_2, \ldots$ *(paarweise) disjunkt* $:\Longleftrightarrow$ $A_i \cap A_j = \emptyset$ für jedes Paar i, j mit $i \neq j$

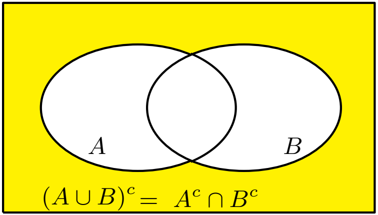
Nur für disjunkte Ereignisse A und B bzw. paarweise disjunkte Ereignisse $A_1, A_2, \ldots, A_n$ bzw. $A_1, A_2, \ldots$ setzen wir

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

Rechenregeln der Mengenlehre beachten! Z.B. *de Morgansche Regeln*

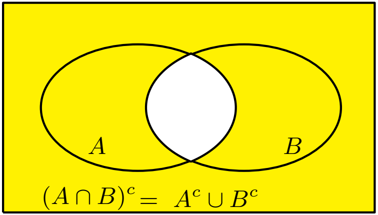
$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c$$

$$\left(\bigcup_{j=1}^{n} A_j \right)^c = \bigcap_{j=1}^{n} A_j^c, \qquad \left(\bigcap_{j=1}^{n} A_j \right)^c = \bigcup_{j=1}^{n} A_j^c$$



$$(A \cup B)^c = A^c \cap B^c$$

Verbal: Es tritt genau dann nicht mindestens eines der Ereignisse A und B ein, wenn keines dieser Ereignisse eintritt.



$$(A \cap B)^c = A^c \cup B^c$$

Verbal: Es treten genau dann nicht beide der Ereignisse A und B ein, wenn mindestens eines dieser Ereignisse nicht eintritt.

2 Zufallsvariablen

2.1 Definition (Zufallsvariable)

Jede Abbildung

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

heißt (reelle) *Zufallsvariable*.

Für $\omega \in \Omega$ heißt $X(\omega)$ *Realisierung* von X (zum Ausgang ω).

Übliche Buchstaben für Zufallsvariablen: Z, Y, X, W, V, U, T.

Zufallsvariablen beschreiben reellwertigen Aspekt eines stochastischen Vorgangs, z.B. bei $\Omega = \{1,2,3,4,5,6\}^n$:

- $X(a_1, \ldots, a_n) := a_1 + \ldots + a_n$ (Augensumme beim n-fachen Würfelwurf),
- $X(a_1, \ldots, a_n) := \max_{1 \le j \le n} a_j$ (größte Augenzahl beim n-fachen WW),
- $X(a_1, \ldots, a_n) := a_2$ (zweite Augenzahl beim n-fachen WW).

Memo: $X : \Omega \to \mathbb{R}$ Zufallsvariable

Zufallsvariablen beschreiben Ereignisse!

Sei $M \subseteq \mathbb{R}$,

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

$$=: \{X \in M\} \quad \textbf{(!!!)}$$

in Worten: „X liegt in M, X nimmt Wert in M an“

$$X^{-1} : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\Omega) \qquad \textit{Urbildabbildung} \text{ zu } X$$

X^{-1} ist verträglich mit allen mengentheoretischen Operationen, d.h. :

$$X^{-1}\left(\sum_{j=1}^{\infty} M_j\right) = \sum_{j=1}^{\infty} X^{-1}(M_j), \quad X^{-1}\left(\bigcap_{j=1}^{\infty} M_j\right) = \bigcap_{j=1}^{\infty} X^{-1}(M_j)$$

$$X^{-1}(M^c) = \left(X^{-1}(M)\right)^c$$

Memo: $X : \Omega \to \mathbb{R}$ Zufallsvariable

Schreibweisen: Sei $t \in \mathbb{R}$,

$$\begin{aligned} \{X = t\} &:= X^{-1}(\{t\}) = \{\omega \in \Omega : X(\omega) = t\} \\ \{X \leq t\} &:= X^{-1}((-\infty, t]) = \{\omega \in \Omega : X(\omega) \leq t\} \\ \{X < t\} &:= X^{-1}((-\infty, t)) = \{\omega \in \Omega : X(\omega) < t\} \end{aligned}$$

Ebenso: $\{X \geq t\}$, $\{X > t\}$, $\{X \neq t\}$, $\{a \leq X < b\}$ usw.

2.2 Beispiel (Zweifacher Würfelwurf, $X :=$ Augensumme)

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$\{X = 5\} = \{(4,1), (3,2), (2,3), (1,4)\}$

Memo: $X : \Omega \to \mathbb{R}$ Zufallsvariable

Die Menge der Zufallsvariablen auf Ω ist ein Vektorraum über $\mathbb{R}$ bezüglich Addition und skalarer Multiplikation.

Auch:

$$\begin{aligned} (X \cdot Y)(\omega) &:= X(\omega) \cdot Y(\omega), \qquad \omega \in \Omega, \\ \max(X, Y)(\omega) &:= \max(X(\omega), Y(\omega)), \qquad \omega \in \Omega, \\ \min(X, Y)(\omega) &:= \min(X(\omega), Y(\omega)), \qquad \omega \in \Omega. \end{aligned}$$

Hiermit neue Ereignisse, z.B.

$$\begin{aligned} \{X \le Y\} &:= \{\omega \in \Omega : X(\omega) \le Y(\omega)\}, \\ \{X - 2Y > 0\} &:= \{\omega \in \Omega : X(\omega) - 2Y(\omega) > 0\} \text{ usw.} \end{aligned}$$

Beispiel: Zweifacher Würfelwurf, X und Y seien die Augenzahlen des ersten bzw. zweiten Wurfs.

$$\{X - 2Y > 0\} = \{X > 2Y\} = \{(6,2), (6,1), (5,2), (5,1), (4,1), (3,1)\}$$

2.3 Definition (Indikator(funktion))

Sei $A \subseteq \Omega$ ein Ereignis. Die durch

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

definierte Zufallsvariable $\mathbf{1}_A : \Omega \to \mathbb{R}$ heißt *Indikatorfunktion* von A oder *Indikator* von A.

Die Realisierung von $\mathbf{1}_A$ zeigt an, ob A eintritt oder nicht.

(lat. *indicare* = anzeigen).

Falls Indizes auftreten, so auch $\mathbf{1}\{A\} := \mathbf{1}_A$.

2.4 Beispiel (Zweifacher Würfelwurf, $X :=$ Augensumme)

$$\begin{array}{cccccc}
(1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\
(2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\
(3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\
(4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\
(5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\
(6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6)
\end{array}$$

$$A \;:=\; \{X = 5\} \;=\; \{(4,1),(3,2),(2,3),(1,4)\}$$

Für $(i,j) \in \Omega \;=\; \{1,2,3,4,5,6\}^2$ ist

$$\mathbf{1}_A(i,j) \;:=\; \mathbf{1}_A((i,j)) \;=\; \begin{cases} 1, & \text{falls } i+j=5, \\ 0, & \text{sonst .} \end{cases}$$

Memo: $\mathbf{1}_A(\omega) = 1$, falls $\omega \in A$; $\mathbf{1}_A(\omega) = 0$, falls $\omega \notin A$

2.5 Satz (Rechenregeln für Indikatorfunktionen)

a) $\mathbf{1}_\emptyset \equiv 0, \quad \mathbf{1}_\Omega \equiv 1,$

b) $\mathbf{1}_A^2 = \mathbf{1}_A,$

c) $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A,$

d) $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B,$

e) $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B},$

f) $A \subseteq B \iff \mathbf{1}_A \leq \mathbf{1}_B.$

Memo: $\mathbf{1}_A(\omega) = 1$, falls $\omega \in A$; $\mathbf{1}_A(\omega) = 0$, falls $\omega \notin A$

2.6 Definition (Indikatorsumme, Zählvariable)

Es seien $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse.

Die Zufallsvariable $X := \sum_{j=1}^{n} \mathbf{1}\{A_j\}$ heißt *Indikatorsumme* oder *Zählvariable*.

X gibt an, wie viele der A_j eintreten.

$$\{X = 0\} = A_1^c \cap A_2^c \cap \ldots \cap A_n^c, \quad \{X = n\} = A_1 \cap A_2 \cap \ldots \cap A_n.$$

Allgemein gilt:

$$\{X = k\} = \sum_{T \subseteq \{1,\ldots,n\}:|T|=k} \left(\bigcap_{j \in T} A_j \cap \bigcap_{j \notin T} A_j^c \right), \qquad k \in \{0, 1, \ldots, n\}.$$

Beachte: $\displaystyle\bigcap_{j \in \emptyset} A_j \; := \; \Omega$

2.7 Beispiel (Trefferanzahl)

$$\Omega \;:=\; \{0,1\}^n \;=\; \{(a_1,\ldots,a_n) : a_j \in \{0,1\} \text{ für } j=1,\ldots,n\}.$$

$$a_j \;=\; \begin{cases} 1 \text{ bedeutet } \textit{Treffer} \text{ im } j\text{-ten Versuch}, \\ 0 \text{ bedeutet } \textit{Niete} \text{ im } j\text{-ten Versuch} \end{cases}$$

$$A_j \;:=\; \{(a_1,\ldots,a_n) \in \Omega : a_j = 1\} \qquad (\text{Treffer im } j\text{-ten Versuch})$$

$$X \;:=\; \mathbf{1}\{A_1\} + \ldots + \mathbf{1}\{A_n\} \quad (\text{ modelliert Trefferanzahl in } n \text{ Versuchen})$$

$$X(\omega) \;=\; a_1 + \ldots + a_n, \qquad \omega = (a_1,\ldots,a_n).$$

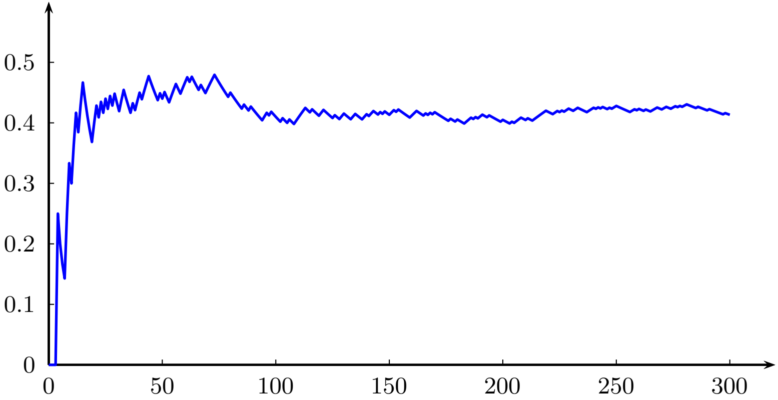
Wichtig!

Es können „Abhängigkeiten“ zwischen den „Versuchen“ vorhanden sein, z.B.:

Ich erhalte der Reihe nach 10 Spielkarten;

$a_j := 1$ bzw. $a_j := 0$, falls j-te Karte ein As bzw. kein As ist.

3 Diskrete Wahrscheinlichkeitsräume



Fortlaufend notierte relative Häufigkeiten für „Spitze nach oben“ beim Werfen einer Reißzwecke

Ideales Zufallsexperiment:

Experiment mit *zufälligem Ausgang*, das *beliebig oft* unter *gleichen, sich gegenseitig nicht beeinflussenden Bedingungen* wiederholt werden kann, z.B.

- Münz- oder Würfelwurf,
- Ziehen aus einer Urne mit Zurücklegen,
- Drehen eines Glücksrades,
- Roulette,
- Kartenverteilungen,
- Ausspielungen beim Lotto.

Empirisches Gesetz über die Stabilisierung relativer Häufigkeiten:

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

Ideales Zufallsexperiment, Ergebnise modelliert durch Grundraum Ω.

n mal „in unabhängiger Folge“ wiederholen

Ergebnisse $\in \Omega^n \; = \; \{(a_1, \ldots, a_n) : a_j \in \Omega$ für $j = 1, \ldots, n\}$

Sei $A \subseteq \Omega$, $(a_1, \ldots, a_n) \in \Omega^n$ fest

$$r_n(A) \; := \; \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_A(a_j)$$ (relative Häufigkeit von A zu $(a_1, \ldots, a_n)$)

Für die relative Häufigkeitsfunktion $r_n : \mathcal{P}(\Omega) \to \mathbb{R}$ gelten:

- $0 \le r_n(A) \; \le \; 1, \qquad A \subseteq \Omega$,
- $r_n(\Omega) \; = \; 1$,
- $r_n(A + B) \; = \; r_n(A) + r_n(B) \qquad (\text{falls } A \cap B \; = \; \emptyset)$,
- $r_n(A) \; \rightsquigarrow \; ?$ bei $n \to \infty$.

3.1 Definition (diskreter Wahrscheinlichkeitsraum, vorläufig)

$(\Omega, \mathbb{P})$ heißt *diskreter Wahrscheinlichkeitsraum* $:\Longleftrightarrow$

- $\Omega \neq \emptyset$ elementarer Grundraum
- $\mathbb{P} : \mathcal{P}(\Omega) \to \mathbb{R}$ mit folgenden Eigenschaften:

(P1) $\mathbb{P}(A) \geq 0, \qquad A \subseteq \Omega,$

(P2) $\mathbb{P}(\Omega) = 1,$

(P3) $A_1, A_2, \ldots \subseteq \Omega$ *paarweise disjunkt* $\Longrightarrow \mathbb{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$

(sog. *σ-Additivität* von $\mathbb{P}$)

$\mathbb{P}$ heißt *Wahrscheinlichkeitsverteilung* oder *Wahrscheinlichkeitsmaß* auf $\mathcal{P}(\Omega)$.

$\mathbb{P}(A)$ heißt *Wahrscheinlichkeit* von A.

W kurz für *Wahrscheinlichkeit(s)* und W-Verteilung *auf* Ω (anstelle $\mathcal{P}(\Omega)$).

Memo: (P1) $\mathbb{P}(A) \geq 0$, (P2) $\mathbb{P}(\Omega) = 1$, (P3) $\mathbb{P}(\sum_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$

3.2 Folgerungen Für $A, B, A_1, A_2, \ldots \subseteq \Omega$ gelten:

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

b) $\mathbb{P}\left(\sum_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbb{P}(A_j)$ (*endliche Additivität*),

c) $0 \leq \mathbb{P}(A) \leq 1$,

d) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ (*Regel von der komplementären W'*),

e) $A \subseteq B \Longrightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (*Monotonie*),

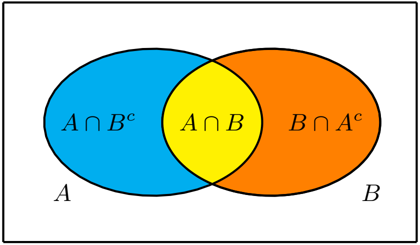
f) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (*Additionssatz*),

g) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ (*σ-Subadditivität*)

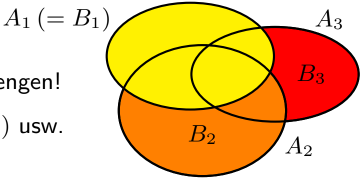
Memo: 3.2.b): $\mathbb{P}\left(\sum_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbb{P}(A_j)$

f) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

$$\begin{aligned} A \cup B &= A + B \cap A^c, \\ \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \\ &= \mathbb{P}(A) + \mathbb{P}(B \cap A^c) + \mathbb{P}(B \cap A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{aligned}$$



g) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n)$



Schreibe $\bigcup_{n=1}^{\infty} A_n$ als Vereinigung disjunkter Mengen!

$B_1 := A_1$, $B_2 := A_2 \setminus A_1$, $B_3 := A_3 \setminus (A_1 \cup A_2)$ usw.

Allgemein:

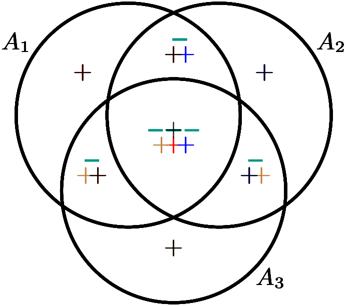
$$B_n := A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) = A_n \cap A_{n-1}^c \cap \ldots \cap A_2^c \cap A_1^c, \qquad n \ge 2.$$

$$\begin{aligned} B_n &\subseteq A_n, \\ B_n \cap B_m &= \emptyset \ \forall m \neq n, \\ \bigcup_{n=1}^{\infty} A_n &= \sum_{n=1}^{\infty} B_n \quad \text{(„}\subseteq\text{“ mit Prinzip des kleinsten Täters!)} \end{aligned}$$

$$\Longrightarrow \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \le \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Memo: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

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



Memo: $\mathbb{P}(\cup_{j=1}^3 A_j) = \sum_{j=1}^3 \mathbb{P}(A_j) - \sum_{1\le i<j\le 3} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$

3.3 Satz (Formel des Ein- und Ausschließens, Bonferroni-Ungleichungen)

Es seien $A_1, \ldots, A_n$ Ereignisse und für $k = 1, \ldots, n$

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

Dann gelten:

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

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

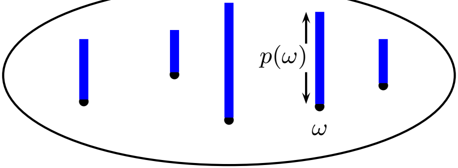
$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) \ge \sum_{k=1}^{2s} (-1)^{k-1} S_k, \qquad s = 0, 1, \ldots, \left\lfloor \frac{n}{2} \right\rfloor.$$

3.4 Definition (Wahrscheinlichkeitsfunktion, Zähldichte)

Sei $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und

$$p(\omega) \;:=\; \mathbb{P}(\{\omega\}), \qquad \omega \in \Omega.$$

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



Beachte: $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \qquad A \subseteq \Omega.$

Memo: (P1) $\mathbb{P}(A) \geq 0$, (P2) $\mathbb{P}(\Omega) = 1$, (P3) $\mathbb{P}(\sum_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$

3.5 Satz ($\mathbb{P}$ ist durch p festgelegt)

Seien Ω ein Grundraum und $p : \Omega \rightarrow [0, \infty)$ mit $\sum_{\omega \in \Omega} p(\omega) = 1$. Sei

$$\mathbb{P}(A) \; := \; \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega.$$

Dann ist $\mathbb{P}$ ein W-Maß auf $\mathcal{P}(\Omega)$ mit Wahrscheinlichkeitsfunktion p.

BEWEIS: $\mathbb{P}(A) \geq 0$ klar, $\mathbb{P}(\Omega) = 1$ klar,

$$\mathbb{P}\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{\omega \in \sum_{j=1}^{\infty} A_j} p(\omega) = \sum_{j=1}^{\infty} \left(\sum_{\omega \in A_j} p(\omega)\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

3.6 Beispiel Seien

- $\Omega := \mathbb{N}_0$,
- $(p_k)_{k \geq 0}$ eine reelle Folge mit $p_k \geq 0$ für $k \geq 0$,
- $\sum\limits_{k=0}^{\infty} p_k = 1$,

z.B.

$$\begin{aligned} p_k &= \frac{1}{(k+1)(k+2)}, \\ p_k &= (1-a)a^k, \quad a \in (0,1), \\ p_k &= e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \qquad \lambda > 0. \end{aligned}$$

3.7 Definition (Endlicher W-Raum, Laplacescher W-Raum)

Es sei $(\Omega, \mathbb{P})$ ein diskreter W-Raum. Ist Ω endlich, so heißt $(\Omega, \mathbb{P})$ *endlicher W-Raum*.

Gilt für jedes $A \subseteq \Omega$

$$\mathbb{P}(A) \ := \ \frac{|A|}{|\Omega|} \ = \ \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}},$$

so heißt $(\Omega, \mathbb{P})$ *Laplacescher W-Raum (der Ordnung* $|\Omega|$*)*.

$\mathbb{P}$ heißt (diskrete) *Gleichverteilung* auf Ω.

$(\Omega, \mathbb{P})$ heißt auch *Laplace-Modell*.

Es gilt dann insbesondere

$$\mathbb{P}(\{\omega\}) \ = \ \frac{1}{|\Omega|}, \quad \omega \in \Omega.$$

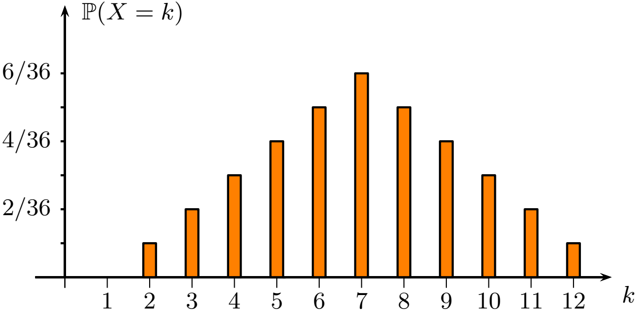
Sprechweisen: Echte Münze, fairer (homogener) Würfel, rein zufälliges Ziehen

3.8 Beispiel (Zweifacher Würfelwurf: Augensumme)

$$\Omega \ := \ \{(i,j) : 1 \le i,j \le 6\}, \quad \mathbb{P}(\{(i,j)\}) := \frac{1}{36} =: p(i,j), \qquad 1 \le i,j \le 6.$$

X = Augensumme, d.h. $X(i,j) := i + j$, $\omega \ = \ (i,j) \in \Omega$.

$$\mathbb{P}(X = 5) = p(1,4) + p(2,3) + p(3,2) + p(4,1) = 4 \cdot \frac{1}{36} = \ \frac{1}{9}.$$



Stabdiagramm der *Verteilung (?) von* X

Seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\}$$

ist abzählbare Teilmenge der überabzählbaren Menge $\mathbb{R}$.

Für $M \subseteq \mathbb{R}$ sei

$$\mathbb{P}^X(M) \ := \ \mathbb{P}\left(X^{-1}(M)\right) \ (= \mathbb{P}(X \in M) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in M\})).$$

Für $M, M_1, M_2, \ldots \subseteq \mathbb{R}$ gelten

$$\begin{aligned} \mathbb{P}^X(M) &\geq 0, \qquad M \subseteq \mathbb{R}, \\ \mathbb{P}^X(\mathbb{R}) &= 1, \\ \mathbb{P}^X\left(\sum_{j=1}^{\infty} M_j\right) &= \sum_{j=1}^{\infty} \mathbb{P}^X(M_j), \\ \mathbb{P}^X(X(\Omega)) &= 1 \end{aligned}$$

Folgende Verallgemeinerung eines diskreten W-Raums sinnvoll:

3.9 Definition (Diskreter W-Raum, endgültig)

Seien $\Omega \neq \emptyset$ eine beliebige Menge und $\mathbb{P} : \mathcal{P}(\Omega) \to \mathbb{R}$.

$(\Omega, \mathbb{P})$ heißt *diskreter Wahrscheinlichkeitsraum*, falls gilt:

(P1) $\mathbb{P}(A) \geq 0, \qquad A \subseteq \Omega,$

(P2) $\mathbb{P}(\Omega) = 1,$

(P3) $\mathbb{P}\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n),$

(P4) $\mathbb{P}(\Omega_0) = 1$ für eine abzählbare Menge $\Omega_0 \subseteq \Omega$ (sog. Träger von $\mathbb{P}$) .

Mit $p(\omega) := \mathbb{P}(\{\omega\}),\ \omega \in \Omega$, gilt $\sum_{\omega \in \Omega_0} p(\omega) = 1$ und

$$\mathbb{P}(A) = \sum_{\omega \in A \cap \Omega_0} p(\omega), \qquad A \subseteq \Omega.$$

Weiterhin heißt jede Abbildung $X : \Omega \to \mathbb{R}$ *Zufallsvariable*.

Memo: (P1) $\mathbb{P}^X(M) \geq 0$, (P2) $\mathbb{P}^X(\mathbb{R}) = 1$, (P3) $\mathbb{P}^X\Big(\sum_{j=1}^{\infty} M_j\Big) = \sum_{j=1}^{\infty} \mathbb{P}(M_j)$

3.10 Satz und Definition (Verteilung einer Zufallsvariablen)

Seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.

Dann ist $(\mathbb{R}, \mathbb{P}^X)$ ein diskreter W-Raum.

Das Wahrscheinlichkeitsmaß $\mathbb{P}^X$ auf $\mathcal{P}(\mathbb{R})$ heißt *Verteilung* von X.

Beachte: $\exists\, \Omega_0 \subseteq \Omega$, Ω_0 abzählbar, $\mathbb{P}(\Omega_0) = 1$

$\Longrightarrow \mathbb{P}^X(X(\Omega_0)) = 1$, $X(\Omega_0) \subseteq \mathbb{R}$ abzählbar

$$\Longrightarrow \mathbb{P}^X(M) = \sum_{t \in M \cap M_0} \mathbb{P}(X = t), \qquad M \subseteq \mathbb{R},$$

d.h. $\mathbb{P}^X$ ist durch die sog. *Wahrscheinlichkeitsfunktion*

$$t \mapsto \mathbb{P}(X = t), \qquad t \in \mathbb{R},$$

von $\mathbb{P}^X$ (von X) eindeutig festgelegt.

Deshalb oft synonym:

$$\{(t, \mathbb{P}(X = t)) : t \in \mathbb{R}, \mathbb{P}(X = t) > 0\} = \textit{ Verteilung } \text{ von } X.$$

Manchmal einfache geschlossene Form hierfür möglich,

z.B. für Verteilung der Augensumme X beim zweifachen Würfelwurf:

$$\mathbb{P}(X = k) \; = \; \frac{6 - |7 - k|}{36}, \qquad k = 2, 3, \ldots, 12$$

Plot von $(t, \mathbb{P}(X = t))$ in Form von *Stab-* oder *Balkendiagrammen*.

Memo: z.z: (P1) $Q(M) \geq 0$, (P2) $Q(\mathbb{R}) = 1$, (P3) $Q(\sum_{j=1}^{\infty} M_j) = \sum_{j=1}^{\infty} Q(M_j)$

Sei

- $T_0 \subseteq \mathbb{R}$ abzählbar,
- $f : \mathbb{R} \to [0, \infty)$ eine Funktion mit
- $\displaystyle\sum_{t \in T_0} f(t) = 1, \qquad f(t) = 0$ für $t \in \mathbb{R} \setminus T_0$.

Sei $\displaystyle Q(M) := \sum_{t \in M \cap T_0} f(t), \qquad M \subseteq \mathbb{R}$.

Dann ist Q eine Wahrscheinlichkeitsverteilung auf $\mathcal{P}(\mathbb{R})$.

Beweis: (P1) klar, (P2) klar,

$$Q\left(\sum_{j=1}^{\infty} M_j\right) = \sum_{t \in \left(\sum_{j=1}^{\infty} M_j\right) \cap T_0} f(t) = \sum_{j=1}^{\infty} \left(\sum_{t \in M_j \cap T_0} f(t) \right) = \sum_{j=1}^{\infty} Q(M_j)$$

Sei Q ein W-Maß auf $\mathcal{P}(\mathbb{R})$.

Gibt es einen diskreten W-Raum $(\Omega, \mathbb{P})$ und eine Zufallsvariable $X : \Omega \to \mathbb{R}$ mit

$$\mathbb{P}^X \; = \; Q? \qquad \text{(d.h.: } X \text{ hat die Verteilung } Q\text{)}$$

Antwort: Ja! Setze

$$\begin{aligned} \Omega &:= \mathbb{R}, \\ \mathbb{P} &:= Q, \\ X &:= \mathrm{id}_\Omega. \end{aligned}$$

$$\mathbb{P}^X(M) \; = \; \mathbb{P}(X^{-1}(M)) \; = \; \mathbb{P}(M) \; = \; Q(M), \qquad M \subseteq \mathbb{R}.$$

Konsequenz: $(\Omega, \mathbb{P})$ und X kanonisch konstruierbar, wenn es um Verteilungen geht.

Sei X eine (auf irgendeinem W-Raum $(\Omega, \mathbb{P})$ definierte) Zufallsvariable.

Besitzt X die Verteilung Q, so schreiben wir hierfür

$$X \ \sim \ Q \quad (:\Longleftrightarrow \mathbb{P}^X = Q).$$

Manche Verteilungen Q (als W-Maße auf $\mathcal{P}(\mathbb{R})$) sind so wichtig, dass sie eigene Namen und Bezeichungen erhalten, z.B.

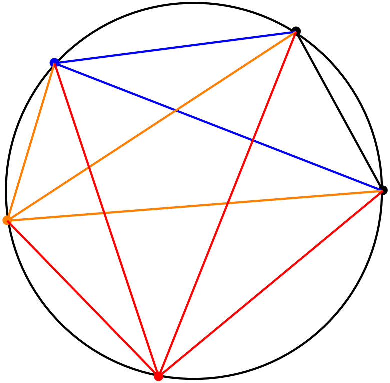
$$X \ \sim \ \mathsf{Po}(\lambda) \ :\Longleftrightarrow \ \mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

(sogenannte *Poisson-Verteilung* mit Parameter λ, $0 < \lambda < \infty$).

Besitzen zwei Zufallsvariablen X und Y (die auf unterschiedlichen W-Räumen $(\Omega, \mathbb{P})$ und $(\widetilde{\Omega}, \widetilde{\mathbb{P}})$ definiert sein können) *dieselbe Verteilung*, so schreiben wir hierfür

$$X \ \sim \ Y \quad (:\Longleftrightarrow \mathbb{P}^X = \widetilde{\mathbb{P}}^Y).$$

4 Kombinatorik



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 ?

4.1 Satz (Erstes Fundamentalprinzip des Zählens)

Es seien M und N endliche Mengen. Dann gilt:

$$|M| = |N| \iff \exists\, f : M \to N,\ f \text{ bijektiv.}$$

4.2 Satz (Zweites Fundamentalprinzip des Zählens, Multiplikationsregel)

Seien M eine endliche Menge und $m_1, \ldots, m_k \in \{1, \ldots, |M|\}$.

Sukzessive werden Elemente $a_1, a_2, \ldots, a_k \in M$ so ausgewählt, dass es für

das 1. Element a_1 stets m_1 Möglichkeiten,

das 2. Element a_2 bei festgelegtem a_1 stets m_2 Möglichkeiten,

das 3. Element a_3 bei festgelegten a_1, a_2 stets m_3 Möglichkeiten,

$\vdots$

das k. Element a_k bei festgelegten $a_1, \ldots, a_{k-1}$ stets m_k Möglichkeiten gibt.

Die Anzahl verschiedener k-Tupel $(a_1, \ldots, a_k)$ ist dann das Produkt

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

Wie viele Abbildungen gibt es von einer 4-elementigen in eine 6-elementige Menge?

Wie viele Abbildungen gibt es von einer k-elementigen in eine n-elementige Menge?

Wie viele injektive Abbildungen gibt es von einer 4-elementigen in eine 6-elementige Menge?

Wie viele injektive Abbildungen gibt es von einer k-elementigen in eine n-elementige Menge?

4.3 Definition (k-Permutationen)

Sei M eine n-elementige Menge.

a) Ein k-Tupel $(a_1, \ldots, a_k)$ mit $a_j \in M \quad \forall j = 1, \ldots, k$ heißt *k-Permutation aus M mit Wiederholung (m.W.)*.

$$Per_k^M(mW) := M^k$$

sei die Menge aller k-Permutationen m.W. aus M.

b) Gilt zusätzlich $a_i \neq a_j$ für alle $i \neq j$, so heißt $(a_1, \ldots, a_k)$ eine *k-Permutation aus M ohne Wiederholung (o.W.)*.
Hierfür muss $k \leq n$ gelten.

$$Per_k^M(oW) := \{(a_1, \ldots, a_k) \in M^k : a_i \neq a_j \ \forall \ i \neq j\}$$

sei die Menge aller k-Permutationen o.W. aus M.

Die n-Permutationen aus M heißen kurz *Permutationen von M*.

4.4 Definition (k-Kombinationen)

Sei M durch eine Relation $\leq$ geordnet.

a) Ein k-Tupel $(a_1, \ldots, a_k)$ aus M^k mit $a_1 \leq a_2 \leq \ldots \leq a_k$ heißt *k-Kombination aus M mit Wiederholung*.

$$Kom_k^M(mW) := \{(a_1, \ldots, a_k) \in M^k : a_1 \leq a_2 \leq \ldots \leq a_k\}$$

sei die Menge aller k-Kombinationen m.W. aus M.

b) Ein k-Tupel $(a_1, \ldots, a_k)$ aus M^k mit $a_1 < a_2 < \ldots < a_k$ heißt *k-Kombination aus M ohne Wiederholung*.

$$Kom_k^M(oW) := \{(a_1, \ldots, a_k) \in M^k : a_1 < a_2 < \ldots < a_k\}$$

bezeichne die Menge aller k-Kombinationen aus M ohne Wiederholung.

Ist speziell $M = \{1, \ldots, n\}$, so

$$Per_k^n := Per_k^M, \qquad Kom_k^n := Kom_k^M.$$

Memo: $Per_k^n(oW) = \{(a_1, \ldots, a_k) \in \{1, \ldots, n\}^k : a_i \neq a_j \ \forall i \neq j\}$

Memo: $Per_k^n(mW) = \{1, \ldots, n\}^k$

Memo: $Kom_k^n(oW) = \{(a_1, \ldots, a_k) \in \{1, \ldots, n\}^k : 1 \leq a_1 < a_2 < \ldots < a_k\}$

Memo: $Kom_k^n(mW) = \{(a_1, \ldots, a_k) \in \{1, \ldots, n\}^k : a_1 \leq a_2 \leq \ldots \leq a_k\}$

4.5 Satz (Grundformeln der Kombinatorik)

Die Anzahl der k-Permutationen mit/ohne Wiederholung und der k-Kombinationen mit/ohne Wiederholung aus M ist

	m.W.	o.W. ($k \leq n$)
k-Permutationen	n^k	$n^{\underline{k}}$
k-Kombinationen	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Hierbei ist $\quad n^{\underline{k}} = n\,(n-1)\,\ldots\,(n-(k-1)) = \frac{n!}{(n-k)!}$

$$\binom{m}{l} := \frac{m!}{l!\,(m-l)!} \qquad m! = 1 \cdot 2 \cdot \ldots \cdot m, \quad 0! = 1$$

Memo: $|Kom_s^r(oW)| = \binom{r}{s}$ Behauptung: $|Kom_k^n(mW)| = \binom{n+k-1}{k}$

BEWEIS: Sei $a = (a_1, \ldots, a_k) \in Kom_k^n(mW)$, also $1 \leq a_1 \leq a_2 \leq \ldots \leq a_k \leq n$.

Idee: Ziehe Komponenten $a_1, \ldots, a_k$ auseinander

$b_1 := a_1$, $b_2 := a_2 + 1$, $b_3 := a_3 + 2$, $\ldots$, $b_k := a_k + k - 1$

$$\Longrightarrow b := (b_1, \ldots, b_k) =: f(a) \in Kom_k^{n+k-1}(oW)$$

$f : Kom_k^n(mW) \to Kom_k^{n+k-1}(oW)$ ist injektiv!

f ist surjektiv! (da $b_j = a_j + j - 1 \Longleftrightarrow a_j = b_j - j + 1$)

$$\Longrightarrow |Kom_k^n(mW)| = |Kom_k^{n+k-1}(oW)| = \binom{n+k-1}{k}$$

Alternative Herleitung der Formel

$$|Kom_k^n(mW)| \; = \; \binom{n+k-1}{k} \; \left(= \binom{n+k-1}{n-1} \right)$$

$n+k-1$ Kreise , davon $n-1$ mit Strich markieren: im Bild: $n=5$, $k=7$

Striche als Trennstriche interpretieren!

Konfiguration entspricht $(1, 1, 1, 5, 5, 5, 5) \in Kom_7^5(mW)$

Konfiguration entspricht $(1, 2, 2, 4, 5, 5, 5) \in Kom_7^5(mW)$

4.6 Bemerkungen

a) $\binom{n}{k} = |\{M : M \subseteq \{1,\ldots,n\},\ |M| = k\}|$

b) $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$ (begrifflicher Beweis?)

$$\begin{array}{ccccccccccccccc}
 & & & & & & & 1 & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}$$

c) Es gilt $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k\, y^{n-k}$ (begrifflicher Beweis?)

4.7 Beispiel (Zwillinge beim Lotto)

Sei $\Omega := Kom_6^{49}(oW) = \{a = (a_1, \ldots, a_6) : 1 \leq a_1 < \ldots < a_6 \leq 49\}$

$$A := \{a \in \Omega : \exists i \in \{2, \ldots, 6\} \text{ mit } a_i - a_{i-1} = 1\}$$

(mindestens zwei direkt benachbarte Zahlen, also ein „Zwilling“)

$\mathbb{P}$ sei die Gleichverteilung auf Ω. $\mathbb{P}(A) = ?$

Sei $a \in A^c$, d.h. $a_i - a_{i-1} \geq 2$, $i = 2, \ldots, 6$. Reihe komprimieren!

$$b_1 := a_1,\ b_2 := a_2 - 1,\ b_3 := a_3 - 2,\ b_4 := a_4 - 3,\ b_5 := a_5 - 4,\ b_6 := a_6 - 5$$

$$\Longrightarrow 1 \leq b_1 < b_2 < \ldots < b_6 \leq 44 \text{ , d.h. } (b_1, \ldots, b_6) \in Kom_6^{44}(oW)$$

Zuordnung $A^c \ni (a_1, \ldots, a_6) \mapsto (b_1, \ldots, b_6) \in Kom_6^{44}(oW)$ ist bijektiv!

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{\binom{44}{6}}{\binom{49}{6}} = 0.495\ldots$$

4.8 Beispiel (Kartenverteilungen beim Skatspiel)

(⟶ Youtube Stochastikclips Skatspiel)

3 Spieler erhalten je 10 Karten; 2 Karten kommen in den *Skat*.

Modellierung der verschiedenen möglichen Kartenverteilungen:

Möglicher Grundraum: Sei

$$K := \{\diamondsuit 7, \diamondsuit 8, \diamondsuit 9, \diamondsuit 10, \diamondsuit B, \diamondsuit D, \diamondsuit K, \diamondsuit A, \heartsuit 7, \heartsuit 8, \heartsuit 9, \heartsuit 10, \heartsuit B, \heartsuit D, \heartsuit K, \heartsuit A,$$
$$\spadesuit 7, \spadesuit 8, \spadesuit 9, \spadesuit 10, \spadesuit B, \spadesuit D, \spadesuit K, \spadesuit A, \clubsuit 7, \clubsuit 8, \clubsuit 9, \clubsuit 10, \clubsuit B, \clubsuit D, \clubsuit K, \clubsuit A\}$$

$$\Omega := \{(K_1, K_2, K_3) : |K_1| = |K_2| = |K_3| = 10, K_1 + K_2 + K_3 \subseteq K\}.$$

K_j steht für die Menge der Karten von Spieler j; der *Skat* ist redundant!

$$|\Omega| = \binom{32}{10} \cdot \binom{22}{10} \cdot \binom{12}{10}$$

Im Folgenden Laplace-Annahme, d.h. Gleichverteilung auf Ω.

Sei A das Ereignis „Spieler 1 erhält genau zwei Buben“.

Sei $\mathbb{B} := \{\diamondsuit B, \heartsuit B, \spadesuit B, \clubsuit B\}$.

$$A = \{(K_1, K_2, K_3) \in \Omega : |K_1 \cap \mathbb{B}| = 2\}.$$

$$|A| = \binom{4}{2} \cdot \binom{28}{8} \cdot \binom{22}{10} \cdot \binom{12}{10}$$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\binom{4}{2} \cdot \binom{28}{8} \cdot \binom{22}{10} \cdot \binom{12}{10}}{\binom{32}{10} \cdot \binom{22}{10} \cdot \binom{12}{10}}$$

$$= \frac{6 \cdot 28! \cdot 22! \cdot 10!}{8! \cdot 20! \cdot 32!} = \frac{6 \cdot 22 \cdot 21 \cdot 10 \cdot 9}{29 \cdot 30 \cdot 31 \cdot 32}$$

$$= \frac{2079}{7192} = 0.289...$$

$\mathbb{P}$(„mindestens ein Spieler hat genau zwei Buben“) $=$?

$$\begin{aligned}\text{Sei } A_j \quad &:= \quad \{\text{„Spieler } j \text{ hat genau zwei Buben“}\} \\ &= \quad \{(K_1, K_2, K_3) \in \Omega : |K_j \cap \mathbb{B}| = 2\}, \qquad j = 1, 2, 3.\end{aligned}$$

Gesucht ist $\mathbb{P}(A_1 \cup A_2 \cup A_3)$. Es gilt

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) \quad &= \quad \sum_{j=1}^{3} \mathbb{P}(A_j) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3) \\ &= \quad 3 \cdot \mathbb{P}(A_1) - 3 \cdot \mathbb{P}(A_1 \cap A_2)\end{aligned}$$

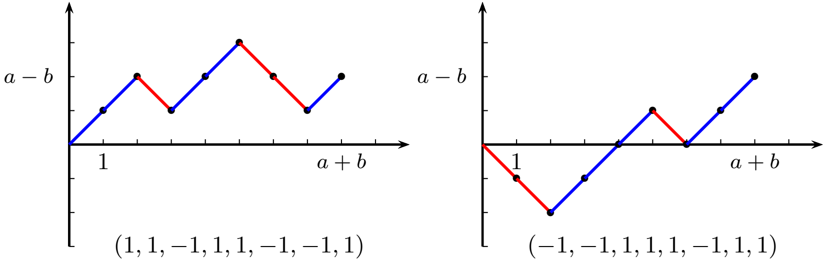
$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) \quad &= \quad \frac{|A_1 \cap A_2|}{|\Omega|} = \frac{\binom{4}{2} \cdot \binom{28}{8} \cdot \binom{20}{8} \cdot \binom{12}{10}}{\binom{32}{10} \cdot \binom{22}{10} \cdot \binom{12}{10}} \\ &= \quad \frac{405}{7192} = 0.0563\ldots\end{aligned}$$

$$\mathbb{P}\left(\cup_{j=1}^{3} A_j\right) = 3 \cdot \frac{2079}{7192} - 3 \cdot \frac{405}{7192} = \frac{5022}{7192} = 0.698\ldots$$

4.9 Beispiel (Das Stimmzettelproblem (ballot problem))

Bei einer Wahl erhalten A und B am Ende a bzw. b Stimmen, wobei $a > b$.

$\mathbb{P}$(A führt während der gesamten Stimmauszählung) = ?



Sei $n := a + b$. Stimmauszählung ist n-tupel $(c_1, \ldots, c_n)$, in dem a Komponenten gleich 1 und b Komponenten gleich -1 sind.

$$\Omega \ := \ \left\{(c_1, \ldots, c_n) \in \{-1, 1\}^n : \sum_{j=1}^{n} \mathbf{1}\{c_j = 1\} = a, \ \sum_{j=1}^{n} \mathbf{1}\{c_j = -1\} = b\right\}$$

$$|\Omega| = \binom{a+b}{a} \qquad \mathbb{P} := \text{Gleichverteilung auf } \Omega$$

Günstige Fälle?

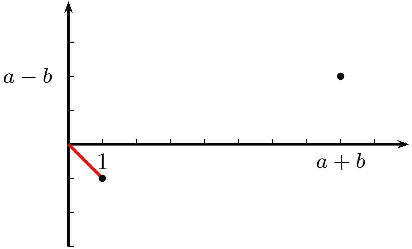
$$\begin{aligned} D \quad &:= \quad \{(c_1,\ldots,c_n) \in \Omega : c_1 + \ldots + c_k \geq 1 \ \text{ für jedes } k = 1,\ldots,n-1\} \\ &= \quad \{\text{„A liegt immer in Führung“}\} \end{aligned}$$

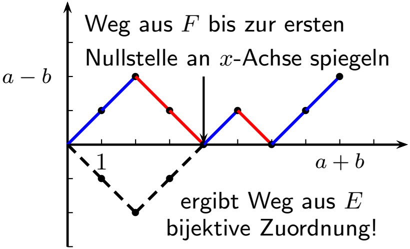
Es gilt $\Omega = D + E + F$, wobei

$E := \{(c_1,\ldots,c_n) \in \Omega : c_1 = -1\}$ (1. Zettel für B)

$F := \{(c_1,\ldots,c_n) \in \Omega : c_1 = 1$ und $c_1 + \ldots + c_k \leq 0$ für ein $k \geq 2\}$

(1. Zettel für A und A führt nicht immer)





Weg aus F bis zur ersten Nullstelle an x-Achse spiegeln

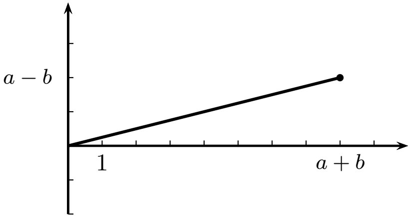
ergibt Weg aus E bijektive Zuordnung!

$$|E| \;=\; \binom{a+b-1}{a} \;=\; |F|$$

Memo: $\Omega = D + E + F, \ |E| = |F|, \ |\Omega| = \binom{a+b}{a}, \ |E| = \binom{a+b-1}{a}$

Folglich gilt

$$\begin{aligned}\mathbb{P}(D) &= 1 - \mathbb{P}(D^c) = 1 - \frac{|E| + |F|}{|\Omega|} = 1 - 2 \cdot \frac{|E|}{|\Omega|} \\ &= 1 - 2 \cdot \frac{b}{a+b} = \frac{a-b}{a+b}.\end{aligned}$$



$\mathbb{P}(D)$ ist die Steigung der Geraden von $(0,0)$ nach $(a+b, a-b)$.

5 Urnen- und Fächer-Modelle

5.1 Beispiel (Qualitätskontrolle)

Werkstatt kauft $n = 10000$ Schrauben. Lieferfirma behauptet, höchstens 5% der Schrauben hielten die vorgeschriebenen Maßtoleranzen nicht ein.

Unter $k = 30$ rein zufällig entnommenen Schrauben sind $m = 6$ unbrauchbare. Soll die Sendung reklamiert werden?

Annahme: 5% der gelieferten Schrauben, d.h $r := 500$ Stück, halten die vorgeschriebenen Maßtoleranzen nicht ein. Berechne $\mathbb{P}(A)$ für das Ereignis

$$A := \quad \text{„Unter } k = 30 \text{ rein zufällig ausgewählten Schrauben sind mindestens 6 unbrauchbare“.}$$

Möglicher Grundraum: $\Omega = Per_k^n(oW)$. Dabei: $n = 10000$, $k = 30$.

Schrauben gedanklich von 1 bis n durchnummerieren.

Für $(a_1, \ldots, a_k) \in \Omega$ ist a_j die Nummer der j-ten entnommenen Schraube.

5.2 Urnenmodelle

In einer *Urne* liegen gleichartige, von 1 bis n nummerierte Kugeln. Wir betrachten vier verschiedene Arten, k Kugeln aus dieser Urne zu ziehen.

(1) *Ziehen unter Beachtung der Reihenfolge mit Zurücklegen*

Nach jedem Zug Kugel-Nummer notieren und Kugel zurücklegen.
a_j sei die Nummer der beim j-ten Zug erhaltenen Kugel.
Geeigneter Ergebnisraum:

$$Per_k^n(mW) = \{(a_1, \ldots, a_k) : 1 \leq a_j \leq n \text{ für } j = 1, \ldots, k\}$$

(k–*Permutationen* aus $1, 2, \ldots, n$ *mit Wiederholung*)

(2) *Ziehen unter Beachtung der Reihenfolge ohne Zurücklegen*

Sei $k \leq n$. Ziehen wie oben, aber ohne Zurücklegen.
Geeigneter Ergebnisraum:

$$Per_k^n(oW) = \{(a_1, \ldots, a_k) \in \{1, 2, \ldots, n\}^k : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq k\}$$

(k-*Permutationen* aus $1, \ldots, n$ *ohne Wiederholung*)

(3) *Ziehen ohne Beachtung der Reihenfolge mit Zurücklegen*

Ziehen mit Zurücklegen; am Ende nur Info, *wie oft* jede der n Kugeln gezogen wurde. Geeigneter Ergebnisraum:

$$Kom_k^n(mW) \quad = \quad \{(a_1, \ldots, a_k) \in \{1, 2, \ldots, n\}^k : a_1 \leq \ldots \leq a_k\}$$

(k-*Kombinationen* aus $1, \ldots, n$ *mit Wiederholung*).

a_j ist j-kleinste Nummer der gezogenen Kugeln.

(4) *Ziehen ohne Beachtung der Reihenfolge ohne Zurücklegen*

Ziehen wie in (3), aber ohne Zurücklegen (vgl. Lotto), $k \leq n$
Geeigneter Ergebnisraum:

$$Kom_k^n(oW) \quad = \quad \{(a_1, \ldots, a_k) \in \{1, 2, \ldots, n\}^k : a_1 < \ldots < a_k\}$$

(k–*Kombinationen* aus $1, \ldots, n$ *ohne Wiederholung*)

5.3 Fächer–Modelle

Es werden k Teilchen auf n von 1 bis n nummerierte Fächer verteilt.

(1) *Unterscheidbare Teilchen*, *Mehrfachbesetzungen zugelassen*

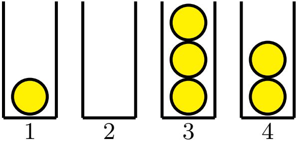
Geeigneter Grundraum $= Per_k^n(mW)$.
$a_j =$ Nummer des Fachs, in dem das j–te Teilchen liegt.

(2) *Unterscheidbare Teilchen*, *keine Mehrfachbesetzungen*

Geeigneter Merkmalraum $= Per_k^n(oW)$.

(3) *Nichtunterscheidbare Teilchen*, *Mehrfachbesetzungen zugelassen*

Geeigneter Merkmalraum $= Kom_k^n(mW)$.



$(1,3,3,3,4,4) \in Kom_6^4(mW)$

(4) *Nichtunterscheidbare Teilchen*, *keine Mehrfachbesetzungen*

Geeigneter Merkmalraum $= Kom_k^n(oW)$

5.4 Beispiel (Fächer–Modelle in der Physik)

Die Modelle (1), (3) und (4) finden in der statistischen Physik Anwendung.

Teilchen sind Gasmoleküle, Photonen, Elektronen, Protonen o.ä.

Phasenraum wird in Zellen (Fächer) unterteilt.

Je nach Gleichverteilungsannahme ergeben sich verschiedene Verteilungen (sog. „Statistiken“).

- *Maxwell–Boltzmann-Statistik* (Modell (1), unter anderem bei Gasen unter mittleren und hohen Temperaturen),
- *Bose–Einstein-Statistik* (Modell (3), für Photonen und He-4-Kerne)
- *Fermi–Dirac-Statistik* (Modell (4), für Elektronen, Neutronen und Protonen (höchstens ein Teilchen in einer Zelle, sog. *Pauli-Verbot*)).

Beachte: Urnen- und Fächer-Modelle sind begrifflich gleichwertig!

Teilchen in Fach Nr. j legen $\Longleftrightarrow$ Kugel Nr. j ziehen.

Teilchen unterscheidbar $\Longleftrightarrow$ Reihenfolge beachten.

Mit Zurücklegen $\Longleftrightarrow$ Mehrfachbesetzungen zugelassen.

- Kollisionsprobleme (Geburtstagsproblem, Lotto-Gewinnreihenwiederholung)
- Zwei-Drittel-Gesetz beim Roulette
- Vollständige Serien, Sammelbilder-Probleme

5.5 Beispiel (Das Paradoxon der ersten Kollision)

Erstmals im Lotto dieselbe Zahlenreihe

> **Stuttgart** *(dpa/lsw). Die Staatliche Toto-Lotto GmbH in Stuttgart hat eine Lottosensation gemeldet: Zum ersten Mal in der 40jährigen Geschichte des deutschen Zahlenlottos wurden zwei identische Gewinnreihen festgestellt. Am 21. Juni dieses Jahres kam im Lotto am Mittwoch in der Ziehung A die Gewinnreihe 15–25–27–30–42–48 heraus. Genau dieselben Zahlen wurden bei der 1628. Ausspielung im Samstaglotto schon einmal gezogen, nämlich am 20. Dezember 1986. Welch ein Lottozufall: Unter den 49 Zahlen sind fast 14 Millionen verschiedene Sechserreihen möglich.*

In der 3016. Ausspielung war zum ersten Mal eine Gewinnreihenwiederholung aufgetreten!

Es gibt

$$n := \binom{49}{6} = 13\ 983\ 816$$

mögliche Gewinnreihen. Gedankliche Durchnummerierung:

Nr. 1: 1 - 2 - 3 - 4 - 5 - **6**
Nr. 2: 1 - 2 - 3 - 4 - 5 - **7**
Nr. 3: 1 - 2 - 3 - 4 - 5 - **8**
⋮ ⋮ ⋮
Nr. 44: 1 - 2 - 3 - 4 - 5 - **49**
Nr. 45: 1 - 2 - 3 - 4 - **6** - **7**
Nr. 46: 1 - 2 - 3 - 4 - 6 - **8**
⋮ ⋮ ⋮
Nr. n: 44 - 45 - 46 - 47 - 48 - **49**

Gewinnreihenermittlung ist rein zufälliges Besetzen eines von insgesamt n verschiedenen Fächern.

Modellierung: Sei

$$X_n \quad := \quad \text{Zeitpunkt der ersten Kollision beim sukzessiven} \\ \text{rein zufälligen Besetzen von } n \text{ Fächern.}$$

Welche Werte nimmt X_n an? Antwort: $2, 3, \ldots, n+1$.

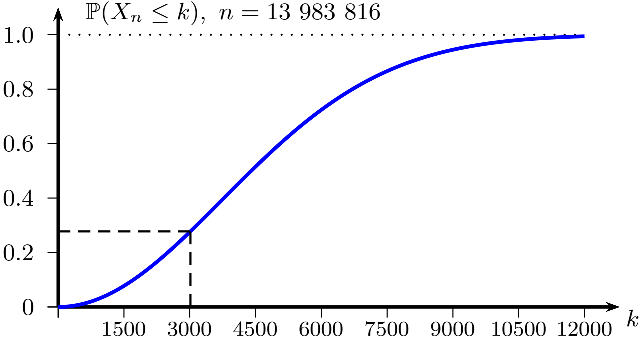
$$X_n \geq k+1 \iff \text{die ersten } k \text{ Teilchen fallen in verschiedene Fächer}$$

$$\begin{aligned} \mathbb{P}(X_n \geq k+1) &= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ \ldots \ \cdot (n-k+1)}{n^k} \\ &= \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) \end{aligned}$$

$k = 1, 2, \ldots, n+1$. Komplementbildung liefert

$$\mathbb{P}\,(X_n \le k) \;=\; 1 - \prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right), \qquad k = 1, 2, 3, \ldots, n+1.$$

Für $n = 13\ 983\ 816$ gilt $\mathbb{P}(X_n \le 3016) \;=\; 0.2775\ldots \approx 10/36$.



Wahrscheinlichkeit für die erste Gewinnreihenwiederholung im Lotto nach höchstens k Ziehungen

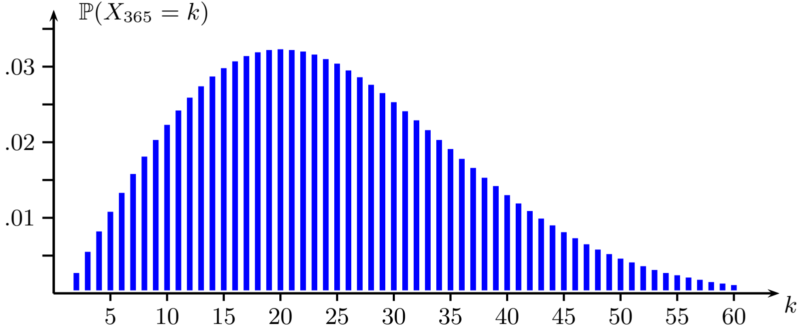
Memo: $\mathbb{P}\left(X_n \le k\right) \;=\; 1 - \prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right), \qquad k = 1, 2, 3, \ldots, n+1.$

Beachte: $\{X_n \le k-1\} + \{X_n = k\} = \{X_n \le k\}$. Also:

$$\begin{aligned}
\mathbb{P}(X_n = k) \quad &= \quad \mathbb{P}(X_n \le k) - \mathbb{P}(X_n \le k-1) \\
&= \quad 1 - \prod_{j=1}^{k-1}\left(1 - \frac{j}{n}\right) - \left(1 - \prod_{j=1}^{k-2}\left(1 - \frac{j}{n}\right)\right) \\
&= \quad \prod_{j=1}^{k-2}\left(1 - \frac{j}{n}\right)\left\{1 - \left(1 - \frac{k-1}{n}\right)\right\} \\
&= \quad \frac{k-1}{n} \cdot \prod_{j=1}^{k-2}\left(1 - \frac{j}{n}\right),
\end{aligned}$$

$k = 2, 3, \ldots, n+1.$

$n = 365$ (Tage des Jahres, ohne 29. Februar)



$\mathbb{P}(X_{365} \leq 23) \approx 0.503$ (Geburtstags-Paradoxon)

Verhalten von X_n für $n \to \infty$?

$\sqrt{365} \approx 19.1$, $\sqrt{13\,983\,816} \approx 3739.5$. Wächst X_n wie $\sqrt{n}$?

Sei $t > 0$ beliebig. Für jede genügend große Zahl n existiert ein $k_n \in \mathbb{N}$ mit

$$2 \;\le\; k_n \;\le\; \sqrt{n} \cdot t \;\le\; k_n + 1 \;\le\; n+1 \qquad \left(\Longrightarrow \frac{k_n}{\sqrt{n}} \to t \right)$$

$$\Longrightarrow\; \mathbb{P}(X_n \le k_n) \;\le\; \mathbb{P}(X_n \le \sqrt{n} \cdot t) \;\le\; \mathbb{P}(X_n \le k_n + 1).$$

Es gilt (mit $1 - x \le \mathrm{e}^{-x}$)

$$\begin{aligned} \mathbb{P}(X_n \le k_n) \quad &= \quad 1 - \prod_{j=1}^{k_n - 1} \left(1 - \frac{j}{n}\right) \;\ge\; 1 - \exp\left(-\sum_{j=1}^{k_n-1} \frac{j}{n}\right) \\ &= \quad 1 - \exp\left(-\frac{1}{2}\, \frac{k_n(k_n-1)}{n}\right) \\ &\to \quad 1 - \exp\left(-\frac{t^2}{2}\right) \end{aligned}$$

$$\Longrightarrow \liminf_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \le t\right) \;\ge\; 1 - \exp\left(-\frac{t^2}{2}\right).$$

Abschätzung von $\mathbb{P}(X_n \le \sqrt{n} \cdot t)$ nach oben? Es gilt $\log y \le y - 1 \Longrightarrow$ $\log y \ge 1 - 1/y \Longrightarrow y \ge \exp((y-1)/y) \Longrightarrow 1 - x \ge \exp(-\frac{x}{1-x})$ für $x < 1 \Longrightarrow$

$$\begin{aligned}
\mathbb{P}(X_n \le k_n + 1) \quad &= \quad 1 - \prod_{j=1}^{k_n} \left(1 - \frac{j}{n}\right) \;\le\; 1 - \prod_{j=1}^{k_n} \exp\left(-\frac{j/n}{1 - j/n}\right) \\
&= \quad 1 - \prod_{j=1}^{k_n} \exp\left(-\frac{j}{n-j}\right) \;\le\; 1 - \prod_{j=1}^{k_n} \exp\left(-\frac{j}{n-k_n}\right) \\
&= \quad 1 - \exp\left(-\sum_{j=1}^{k_n} \frac{j}{n-k_n}\right) \\
&= \quad 1 - \exp\left(-\frac{1}{2}\,\frac{k_n\,(k_n+1)}{n-k_n}\right) \to 1 - \exp\left(-\frac{t^2}{2}\right)
\end{aligned}$$

$$\Longrightarrow \limsup_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \le t\right) \;\le\; 1 - \exp\left(-\frac{t^2}{2}\right)$$

5.6 Satz (Grenzverteilung für die Kollisionszeit)
Für jedes $t > 0$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_n}{\sqrt{n}} \le t\right) \;=\; 1 - \exp\left(-\frac{t^2}{2}\right).$$

Setze

$$\frac{1}{2} \;=\; \exp\left(-\frac{t^2}{2}\right) \;\Longleftrightarrow\; t = \sqrt{2\,\log 2}$$

Es folgt

$$\lim_{n\to\infty} \mathbb{P}(X_n \le \sqrt{2n\,\log 2}) \;=\; \frac{1}{2}$$

$\Longrightarrow\; \mathbb{P}\left(X_{13983816} \le 4403\right) \approx 0.5.$

6 Der Erwartungswert

Motivation: Stochastischer Vorgang (z.B. Glücksspiel) mit Ergebnismenge $\Omega = \{\omega_1, \ldots, \omega_s\}$.

Sei $X(\omega_j)$ der Gewinn bei Ausgang ω_j.

n Spiele unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen durchführen.

h_j mal trete der Ausgang ω_j auf, $j = 1, \ldots, s$.

Gesamtgewinn: $X(\omega_1) \cdot h_1 + \ldots + X(\omega_s) \cdot h_s$.

Durchschnittlicher Gewinn pro Spiel:

$$X(\omega_1) \cdot \underbrace{\frac{h_1}{n}}_{\rightsquigarrow \mathbb{P}(\{\omega_1\})} + \ \ldots + \ X(\omega_s) \cdot \underbrace{\frac{h_s}{n}}_{\rightsquigarrow \mathbb{P}(\{\omega_s\})}.$$

(empirisches Gesetz über die Stabilisierung relativer Häufigkeiten)

6.1 Definition (Erwartungswert)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum mit $\mathbb{P}(\Omega_0) = 1$, wobei $\Omega_0 \subseteq \Omega$ abzählbar, sowie $X : \Omega \to \mathbb{R}$ eine Zufallsvariable.

Der Erwartungswert von X existiert, falls gilt:

$$\sum_{\omega \in \Omega_0} |X(\omega)| \, \mathbb{P}(\{\omega\}) \; < \; \infty. \tag{6.1}$$

In diesem Fall heißt

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

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

Beachte: Bedingung (6.1) bei endlichem Ω_0 (trivialerweise) erfüllt.

Meist: $\mathbb{E}\, X \; = \; \mathbb{E}(X)$.

Für eine nichtnegative Zufallsvariable Y definiert man

$$\mathbb{E}\,Y := \sum_{\omega\in\Omega_0} Y(\omega)\,\mathbb{P}(\{\omega\}) \quad (\leq \infty).$$

Dann gilt: Der Erwartungswert von X existiert $\iff \mathbb{E}|X| < \infty$, denn:

$$\begin{aligned} \mathbb{E}|X| &= \sum_{\omega\in\Omega_0} |X|(\omega)\,\mathbb{P}(\{\omega\}) \\ &= \sum_{\omega\in\Omega_0} |X(\omega)|\,\mathbb{P}(\{\omega\}). \end{aligned}$$

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

6.2 Satz (Strukturelle Eigenschaften der Erwartungswertbildung)

Sei $L^1 := \{X : \Omega \to \mathbb{R} | \sum_{\omega \in \Omega_0} |X(\omega)|\,\mathbb{P}(\{\omega\}) < \infty\}$.

Dann ist L^1 ein Vektorraum über $\mathbb{R}$, und die Zuordnung $X \mapsto \mathbb{E}(X)$ ist additiv, homogen (und damit linear) sowie monoton auf L^1, d.h. es gelten für $X, Y \in L^1$ und $a \in \mathbb{R}$:

$$\left.\begin{aligned} &\textbf{a)}\ \mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y, \\ &\textbf{b)}\ \mathbb{E}(aX) = a\mathbb{E}\,X, \end{aligned}\right\} \text{(\textit{Linearität von} } \mathbb{E}(\cdot)\text{)}$$

c) $X \le Y \implies \mathbb{E}\,X \le \mathbb{E}\,Y$. (*Monotonie von* $\mathbb{E}(\cdot)$)

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

6.3 Satz (Weitere Eigenschaften der Erwartungswertbildung)

Es gelten:

a) $\mathbb{E}(\mathbf{1}_A) \; = \; \mathbb{P}(A), \qquad A \subseteq \Omega,$

b) $|\mathbb{E}\, X| \; \leq \; \mathbb{E}|X|, \qquad X \in L^1.$ (Dreiecksungleichung)

Beweis: a)

$$\mathbb{E}(\mathbf{1}_A) \; = \; \sum_{\Omega_0 \cap A} 1 \cdot \mathbb{P}(\{\omega\}) \; = \; \mathbb{P}(\Omega_0 \cap A) \; = \; \mathbb{P}(A).$$

b) Sei $\Omega_1 \subseteq \Omega_0$, Ω_1 endlich; Dreiecksungleichung $\Longrightarrow$

$$\left| \sum_{\omega \in \Omega_1} X(\omega)\, \mathbb{P}(\{\omega\}) \right| \; \leq \; \sum_{\omega \in \Omega_1} |X(\omega)|\, \mathbb{P}(\{\omega\}) \; \leq \; \mathbb{E}|X| \; \surd$$

Memo: $\mathbb{E}(X+Y) = \mathbb{E}\,X + \mathbb{E}\,Y, \quad \mathbb{E}\,\mathbf{1}_A = \mathbb{P}(A)$

6.4 Folgerung (Erwartungswert einer Zählvariablen)

Seien $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse und

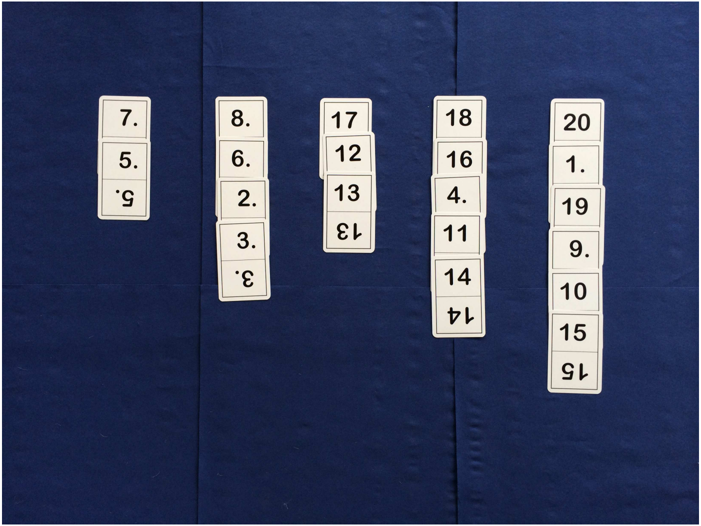
$$X \;=\; \sum_{j=1}^{n} \mathbf{1}\{A_j\}$$

die Anzahl der eintretenden A_j. Dann gilt

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

Gilt speziell $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \ldots = \mathbb{P}(A_n) =: p$, so folgt $\mathbb{E}\,X = n\,p$.

6.5 Beispiel (Anzahl der Rekorde in rein zufälliger Permutation)



Sei $\Omega := Per_n^n(oW)$ die Menge der Permutationen von $1, 2, \ldots, n$.

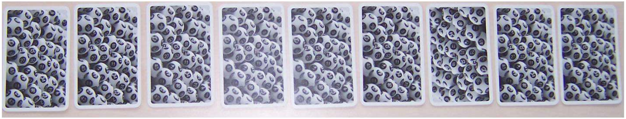
$\mathbb{P}$ sei die Gleichverteilung auf Ω.

Für $j = 1, \ldots, n$ sei

$$A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j = \max(a_1, a_2, \ldots, a_j)\}.$$

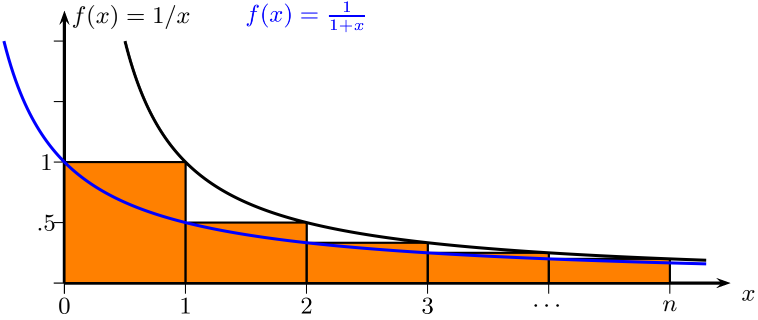
Sei $X_n := \mathbf{1}\{A_1\} + \ldots + \mathbf{1}\{A_n\}$ die *Anzahl der Rekorde*.

$$\mathbb{P}(A_j) \ = \ ?\frac{1}{j} \quad (!)$$



$$\Longrightarrow \ \mathbb{E}(X_n) \ = \ 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$$

$\mathbb{E}(X_{32}) \approx 4.06, \qquad \mathbb{E}(X_{1000000000}) \approx 20.3.$



$$\mathbb{E}(X_n) = \sum_{j=1}^{n} \frac{1}{j} \le 1 + \int_1^n \frac{1}{x}\,\mathsf{d}x = 1 + \log n$$

$$\mathbb{E}(X_n) = \sum_{j=1}^{n} \frac{1}{j} \ge \int_0^n \frac{1}{1+x}\,\mathsf{d}x = \log(1+x)\Big|_0^n = \log(n+1)$$

$$\mathbb{E}(X_n) = \log n + \gamma + o(1), \quad \gamma = 0.57721... \quad \text{(Euler-Mascheronische Konstante)}$$

6.6 Satz (Transformationsformel für $\mathbb{E}g(X)$)

Es sei $g : \mathbb{R} \to \mathbb{R}$ und $g(X) := g \circ X$, also

$$g(X)(\omega) \ := \ g(X(\omega)), \quad \omega \in \Omega.$$

Dann existiert der Erwartungswert von $g(X)$ genau dann, wenn gilt:

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

In diesem Fall gilt $\mathbb{E}\, g(X) \ = \displaystyle\sum_{t \in \mathbb{R}:\mathbb{P}(X=t)>0} g(t)\, \mathbb{P}(X = t).$

Speziell gilt also $\mathbb{E}\, X \ = \displaystyle\sum_{t \in \mathbb{R}:\mathbb{P}(X=t)>0} t\, \mathbb{P}(X = t).$

Folg.: $\mathbb{E}\, X$ hängt nur von Verteilung $\mathbb{P}^X$, aber nicht vom konkreten Ω ab!

Beweis:

$$\mathbb{E}|g(X)| = \sum_{\omega \in \Omega_0} |g(X(\omega))| \, \mathbb{P}(\{\omega\})$$

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

Beachte: Das erste Gleichheitszeichen gilt wegen des Großen Umordnungssatzes!

Weglassen der Betragsstriche liefert dann die Behauptung.

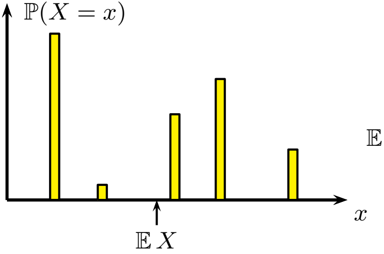
6.7 Beispiel

Die Zufallsvariable X besitze eine Gleichverteilung auf $\{1, 2, \ldots, k\}$, d.h. es gilt

$$\mathbb{P}(X = j) \; = \; \frac{1}{k}, \quad j = 1, \ldots, k. \qquad (\text{Beachte: } \Omega \text{ wird nicht spezifiziert!})$$

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

Der Erwartungswert von X muss also keine Realisierung von X sein!



$\mathbb{E}\,X$ ist physikalischer Schwerpunkt!

6.8 Satz (Die Jordan-Formel)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse.

$X := \sum_{j=1}^{n} \mathbf{1}\{A_j\}$ sei die Anzahl der eintretenden A_j.

Sei $S_0 := 1$,

$$S_j \;:=\; \sum_{1 \le i_1 < \ldots < i_j \le n} \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_j}\right), \qquad 1 \le j \le n.$$

Dann gilt:

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

6.9 Folgerung Für jedes $j \in \{1, \ldots, n\}$ hänge $\mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_j}\right)$ nur von j, aber nicht von der speziellen Wahl der Indizes $i_1, \ldots, i_j$ ab. Dann gilt:

$$\mathbb{P}(X = k) \;=\; \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \mathbb{P}(A_1 \cap \ldots \cap A_j), \qquad k = 0, 1, \ldots, n.$$

Beweis der Jordan-Formel:

Sei $N := \{1, \ldots, n\}$ und allgemein $\{M\}_s := \{T \subseteq M : |T| = s\}$.

Memo: $\{X = k\} = \sum_{T \in \{N\}_k} \left(\bigcap_{j \in T} A_j \cap \bigcap_{l \in N \setminus T} A_l^c \right), \qquad k \in \{0, 1, \ldots, n\}.$

Memo: $\mathbf{1}_{A+B} = \mathbf{1}_A + \mathbf{1}_B, \quad \mathbf{1}_{A \cap B} = \mathbf{1}_A \, \mathbf{1}_B, \quad \mathbf{1}_{A^c} = 1 - \mathbf{1}_A$

$$\Longrightarrow \mathbf{1}\{X = k\} = \sum_{T \in \{N\}_k} \prod_{j \in T} \mathbf{1}_{A_j} \prod_{l \in N \setminus T} (1 - \mathbf{1}_{A_l})$$

$$\prod_{l \in N \setminus T} (1 - \mathbf{1}_{A_l}) = \sum_{r=0}^{n-k} (-1)^r \sum_{U \in \{N \setminus T\}_r} \prod_{j \in U} \mathbf{1}_{A_j}$$

$$\Longrightarrow \mathbf{1}\{X = k\} = \sum_{r=0}^{n-k} (-1)^r \sum_{T \in \{N\}_k} \sum_{U \in \{N \setminus T\}_r} \prod_{j \in T \cup U} \mathbf{1}_{A_j}$$

$$= \sum_{r=0}^{n-k} (-1)^r \sum_{V \in \{N\}_{k+r}} \prod_{j \in V} \mathbf{1}_{A_j} \binom{k+r}{k}.$$

Memo: $\mathbf{1}\{X=k\} \;=\; \sum_{r=0}^{n-k} (-1)^r \binom{k+r}{k} \sum_{V\in\{N\}_{k+r}} \prod_{j\in V} \mathbf{1}_{A_j}$

Memo: $\mathbb{E}\mathbf{1}_B = \mathbb{P}(B),\;\; \mathbf{1}_A\,\mathbf{1}_B = \mathbf{1}_{A\cap B},\;\; \mathbb{E}(aU+bV) = a\mathbb{E}U + b\mathbb{E}V$

$$\begin{aligned}
\Longrightarrow\;\; \mathbb{P}(X=k) &= \mathbb{E}\mathbf{1}\{X=k\} \\
&= \sum_{r=0}^{n-k} (-1)^r \binom{k+r}{k} \sum_{V\in\{N\}_{k+r}} \mathbb{P}\left(\bigcap_{j\in V} A_j\right) \\
&= \sum_{r=0}^{n-k} (-1)^r \binom{k+r}{k} S_{k+r} \\
&= \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j. \quad \surd
\end{aligned}$$

6.10 Beispiel (Anzahl der freien bzw. besetzten Fächer)

s Teilchen rein zufällig auf n Fächer verteilen (Fächermodell 5.3 (1) mit $k = s$). $\Omega := Per_s^n(mW)$, $\mathbb{P} :=$ Gleichverteilung auf Ω.

$A_i := \{(a_1, \ldots, a_s) \in \Omega : a_m \neq i$ für $m = 1, \ldots, s\}$ („Fach Nr. i frei")

Sei $X := \sum_{i=1}^n \mathbf{1}\{A_i\}$ (Anzahl der freien Fächer).

Es gilt für jedes $j = 1, \ldots, n$ und $i_1, \ldots, i_j$ mit $1 \leq i_1 < \ldots < i_j \leq n$

$$\mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_j}\right) = \frac{(n-j)^s}{n^s}. \quad \text{Folgerung 6.9} \implies$$

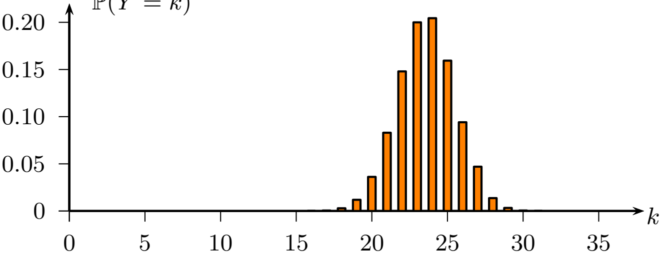
$$\mathbb{P}(X = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \left(\frac{n-j}{n}\right)^s, \qquad k = 0, 1, \ldots, n.$$

Beachte: $Y := n - X$ ist die Anzahl der besetzten Fächer.

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

Spezialfall: $n = s = 37$ (Zwei-Drittel-Gesetz beim Roulette)

$Y =$ Anzahl der verschiedenen Zahlen, die nach 37 Ausspielungen im Roulette aufgetreten sind.



Verteilung der Anzahl Y der besetzten Fächer im Fall $n = s = 37$

Wegen $Y = 37 - \sum_{j=1}^{37} \mathbf{1}\{A_j\}$ gilt

$$\mathbb{E}\,Y \;=\; 37 - 37 \cdot \mathbb{P}(A_1) \;=\; 37\left(1 - \left(\frac{36}{37}\right)^{37}\right) \;\approx\; 23.58.$$

Memo: $X = \sum_{j=1}^{n} \mathbf{1}\{A_j\} \implies \mathbb{P}(X = k) \; = \; \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} S_j$

$$S_0 \; := \; 1, \; S_j = \sum_{1 \le i_1 < \ldots < i_j \le n} \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_j}\right)$$

Setze speziell $k = 0$.

$$\mathbb{P}(X = 0) \; = \; \sum_{j=0}^{n} (-1)^j S_j \quad \implies$$

$$\begin{aligned} \mathbb{P}(X \ge 1) \quad &= \quad 1 - \mathbb{P}(X = 0) \; = \; 1 - \sum_{j=0}^{n} (-1)^j S_j \; = \; \sum_{j=1}^{n} (-1)^{j-1} S_j \\ &= \quad \mathbb{P}(A_1 \cup \ldots \cup A_n) \end{aligned}$$

Aus der Jordan-Formel folgt die Formel des Ein- und Ausschließens!

Das Schnur-Orakel

n Schnüre (im Bild $n = 4$) werden in der Mitte festgehalten, so dass $2n$ Enden frei sind. Diese Enden werden rein zufällig verknotet.

Welchen Erwartungswert (welche Verteilung) besitzt die Anzahl R_n der dabei entstehenden (geschlossenen) Ringe?

$$\begin{aligned} \text{Sei } A_j \quad &:= \quad \{j\text{-te Verknotung führt zu einem Ring}\}, \quad j = 1, \ldots, n \\ \Longrightarrow \; R_n \quad &= \quad \sum_{j=1}^{n} \mathbf{1}\{A_j\} \end{aligned}$$

Bei $n = 4$ Schnüren: $\mathbb{P}(A_1) \; = \; \frac{1}{7}, \;\; \mathbb{P}(A_2) \; = \; \frac{1}{5}, \;\; \mathbb{P}(A_3) \; = \; \frac{1}{3}, \;\; \mathbb{P}(A_4) \; = \; 1$

$$\text{Allg.: } \mathbb{P}(A_1) = \frac{1}{2n-1}, \; \mathbb{P}(A_2) = \frac{1}{2n-3}, \ldots, \mathbb{P}(A_{n-1}) = \frac{1}{3}, \; \mathbb{P}(A_n) = 1.$$

$$\mathbb{E}\, R_n \;=\; \sum_{j=1}^{n} \mathbb{P}(A_j) = 1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2n-1} = \sum_{k=1}^{2n} \frac{1}{k} - \sum_{j=1}^{n} \frac{1}{2j}$$

Sei $H_m := \sum_{k=1}^{m} \frac{1}{k} \approx \log m + \gamma, \;\; \gamma = 0.57721\ldots.$

$$\Longrightarrow \mathbb{E} R_n = H_{2n} - \frac{1}{2} \cdot H_n \approx \log(2n) + \gamma - \frac{1}{2}\,(\log n + \gamma) \;\approx\; \frac{\log n}{2} + 0.98175.$$

n	3	10	100	1000	10^6	10^9
$\mathbb{E}\, R_n$	1.53	2.18	3.28	4.44	7.89	11.34

7 Binomialverteilung und hypergeometrische Verteilung

Situation: Urne mit r roten und s schwarzen Kugeln.

Die Kugeln seien gedanklich nummeriert: $1, 2, \ldots, r, r+1, \ldots, r+s$.

Es wird n mal rein zufällig mit Zurücklegen gezogen.

$$\Omega := Per_n^{r+s}(mW), \quad \mathbb{P} := \text{ Gleichverteilung auf } \Omega$$

$$A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j \leq r\} \quad (j\text{-te Kugel rot})$$

$$X := \sum_{j=1}^{n} \mathbf{1}\{A_j\} \qquad (= \text{Anzahl der gezogenen roten Kugeln})$$

$$|\Omega| = (r+s)^n$$

$$|A_j| = r \cdot (r+s)^{n-1}$$

$$\mathbb{P}(A_j) = \frac{|A_j|}{|\Omega|} = \frac{r}{r+s}, \qquad j = 1, 2, \ldots, n$$

Memo: $X = \sum_{j=1}^{n} \mathbf{1}\{A_j\}$, $\mathbb{P}(A_j) = p$

7.1 Definition (Binomialverteilung)

Es sei $p := \dfrac{r}{r+s}$.

Die Verteilung von X heißt *Binomialverteilung* mit Parametern n und p, kurz

$$X \sim \mathrm{Bin}(n,p) \quad \text{oder} \quad \mathbb{P}^X = \mathrm{Bin}(n,p).$$

7.2 Satz Im Fall $X \sim \mathrm{Bin}(n,p)$ gelten:

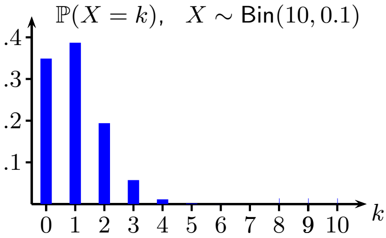
a) $\mathbb{E}\, X \; = \; n\, p$,

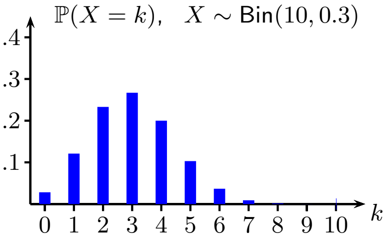
b) $\mathbb{P}(X = k) \; = \; \binom{n}{k} p^k \, (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$

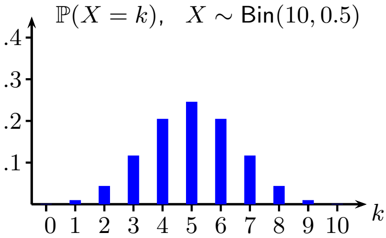
BEWEIS von b)

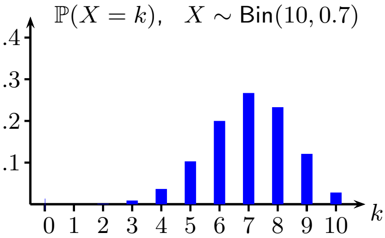
Memo: $X = \sum_{j=1}^{n} \mathbf{1}\{A_j\}, \quad A_j = \{(a_1, \ldots, a_n) \in Per_n^{r+s}(mW) : a_j \le r\}$

$$
\begin{aligned}
|\{X = k\}| \quad &= \quad |\{(a_1, \ldots, a_n) : \text{ genau } k \text{ der } a_j \text{ sind } \; \le r\}| \\
&= \quad \binom{n}{k} \cdot |\{(a_1, \ldots, a_n) : a_1 \le r, \ldots, a_k \le r, a_{k+1} > r, \ldots, a_n > r\}| \\
&= \quad \binom{n}{k} \cdot r^k \cdot s^{n-k} \\
\mathbb{P}(X = k) \quad &= \quad \frac{|\{X = k\}|}{(r+s)^n} \; = \; \binom{n}{k} \cdot \left(\frac{r}{r+s}\right)^k \cdot \left(\frac{s}{r+s}\right)^{n-k} \\
&= \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}
\end{aligned}
$$


$\mathbb{P}(X = k), \quad X \sim \mathsf{Bin}(10, 0.1)$


$\mathbb{P}(X = k), \quad X \sim \mathsf{Bin}(10, 0.3)$


$\mathbb{P}(X = k), \quad X \sim \mathsf{Bin}(10, 0.5)$


$\mathbb{P}(X = k), \quad X \sim \mathsf{Bin}(10, 0.7)$

Jetzt: Ziehen ohne Zurücklegen, wobei $n \le r + s$.

$$\Omega \quad := \quad Per_n^{r+s}(oW), \qquad \mathbb{P} := \text{ Gleichverteilung auf } \Omega,$$

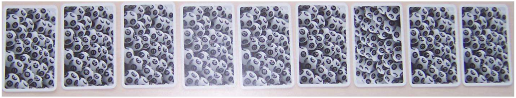
$$A_j \quad := \quad \{(a_1, \ldots, a_n) \in \Omega : a_j \le r\} \quad (j\text{-te Kugel rot})$$

$$X \quad := \quad \sum_{j=1}^{n} \mathbf{1}\{A_j\} \qquad (= \text{Anzahl der gezogenen roten Kugeln})$$

$$|\Omega| \quad = \quad (r+s) \cdot (r+s-1) \cdot \ldots \cdot (r+s-(n-1)) \; = \; (r+s)^{\underline{n}}$$

$$|A_j| \quad = \quad r \cdot (r+s-1) \cdot \ldots \cdot (r+s-(n-1)) \; = \; r \cdot (r+s-1)^{\underline{n-1}}$$

$$\mathbb{P}(A_j) \quad = \quad \frac{|A_j|}{|\Omega|} \; = \; \frac{r}{r+s}, \qquad j = 1, \ldots, n. \quad (!)$$



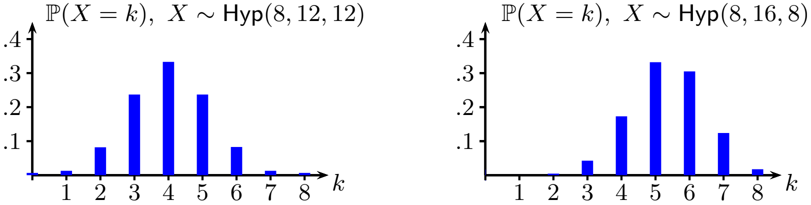
r Karten sind rot,
die übrigen s schwarz

Memo: $X = \sum_{j=1}^{n} \mathbf{1}\{A_j\}$, $A_j = \{(a_1, \ldots, a_n) \in Per_n^{r+s}(oW) : a_j \leq r\}$

7.3 Definition (Hypergeometrische Verteilung)

Die Verteilung von X heißt *hypergeometrische Verteilung* mit Parametern n, r und s, kurz:

$$X \sim \mathrm{Hyp}(n, r, s) \quad \text{bzw.} \quad \mathbb{P}^X = \mathrm{Hyp}(n, r, s).$$



7.4 Satz Falls $X \sim \mathrm{Hyp}(n, r, s)$, so gelten:

a) $\mathbb{E}\,X \;=\; n \cdot \dfrac{r}{r+s}$,

b) $\mathbb{P}(X=k) \;=\; \dfrac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{r+s}{n}}$, $\qquad k = 0, 1, \ldots, n,\ \dbinom{m}{l} := 0$, falls $m < l$.

$$
\begin{aligned}
|\{X=k\}| &= |\{(a_1,\ldots,a_n) :\ \text{genau } k \text{ der } a_j \text{ sind } \le r\}| \\
&= \binom{n}{k} \cdot |\{(a_1,\ldots,a_n) : a_1 \le r, \ldots, a_k \le r, a_{k+1} > r, \ldots, a_n > r\}| \\
&= \binom{n}{k} \cdot r(r-1)\ldots(r-(k-1)) \cdot s(s-1)\ldots(s-(n-k-1)) \\
\mathbb{P}(X=k) &= \frac{|\{X=k\}|}{(r+s)_{\underline{n}}} = \binom{n}{k} \cdot \frac{r!}{(r-k)!} \cdot \frac{s!}{(s-(n-k))!} \cdot \frac{(r+s-n)!}{(r+s)!} \\
&= \frac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{r+s}{n}}
\end{aligned}
$$

8 Modellierung mehrstufiger Experimente

Viele stochastische Vorgänge bestehen aus *Teilexperimenten* (*Stufen*).

Ergebnisse eines n-stufigen Experiments sind n-Tupel $\omega = (a_1, a_2, \ldots, a_n)$.

Dabei sei a_j der Ausgang des j-ten Teilexperiments.

Sei Ω_j die Ergebnismenge des j-ten Teilexperiments.

$$\Omega \quad := \quad \Omega_1 \times \cdots \times \Omega_n \;=\; \{\omega = (a_1, \ldots, a_n) : a_j \in \Omega_j \;\text{ für }\; j = 1, \ldots, n\}$$

ist natürlicher Grundraum für das Gesamt–Experiment.

8.1 Beispiel (Pólya-Urnenschema)

Eine Urne enthalte eine rote und drei schwarze Kugeln.

Rein zufällig Kugel ziehen und Farbe notieren.

Diese sowie eine weitere Kugel derselben Farbe in die Urne zurücklegen.

Nach Mischen wieder Kugel ziehen. Mit welcher W' ist diese rot?

Ziehen einer roten (bzw. schwarzen) Kugel: „r“ (bzw. „s“)

$\Omega := \Omega_1 \times \Omega_2$ mit $\Omega_1 = \Omega_2 = \{r, s\}$

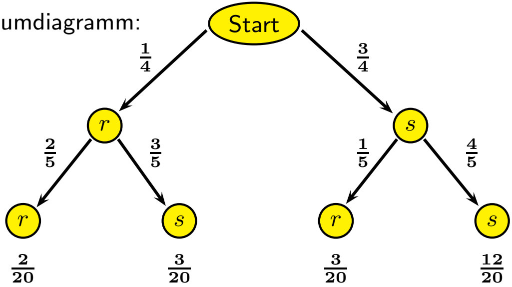
$B \; := \; \{(r,r),(s,r)\}$ (die beim zweiten Mal gezogene Kugel ist rot)

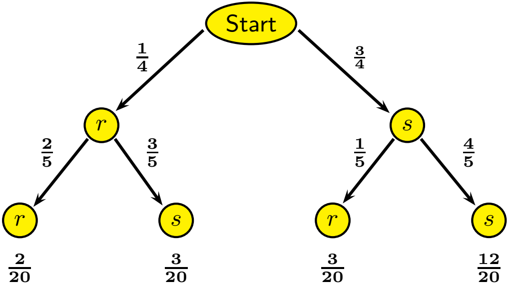
Festlegung der Wahrscheinlichkeiten $p(\omega) := \mathbb{P}(\{\omega\})$:

$$p(r,r) \; := \; \frac{1}{4} \cdot \frac{2}{5}\,, \qquad p(r,s) \; := \; \frac{1}{4} \cdot \frac{3}{5}\,,$$ („Erste Pfadregel“)

$$p(s,r) \; := \; \frac{3}{4} \cdot \frac{1}{5}\,, \qquad p(s,s) \; := \; \frac{3}{4} \cdot \frac{4}{5}\,.$$ (motiviert durch relat. Häufigkeiten!)

Baumdiagramm:





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

(„Zweite Pfadregel“)

Allgemein erhält man ein begründetes W-Maß auf (den Teilmengen von) Ω mithilfe einer *Start-Verteilung* und *Übergangswahrscheinlichkeiten*.

8.2 Satz und Definition

Ω_1 und Ω_2 seien *elementare* Grundräume, $\mathbb{P}_1$ ein W-Maß auf Ω_1,
(sog. *Startverteilung*), mit W-Funktion $p_1(\omega_1) := \mathbb{P}_1(\{\omega_1\})$, $\omega_1 \in \Omega_1$.

$p_2 : \Omega_1 \times \Omega_2 \to \mathbb{R}$ sei eine *Übergangs-W-Funktion* von Ω_1 nach Ω_2,

d.h. eine Funktion mit den Eigenschaften

$$\begin{aligned} p_2(\omega_1, \omega_2) &\geq 0, \quad (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2, \\ \sum_{\omega_2 \in \Omega_2} p_2(\omega_1, \omega_2) &= 1, \quad \omega_1 \in \Omega_1. \end{aligned}$$

Dann definiert

$$p(\omega_1, \omega_2) := p_1(\omega_1) \cdot p_2(\omega_1, \omega_2) \qquad \text{(sog. \textit{Erste Pfadregel})}$$

eine W-Funktion auf $\Omega := \Omega_1 \times \Omega_2$. Das durch

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

definierte W-Maß $\mathbb{P}$ zu p heißt *Kopplung von p_1 und p_2*.

8.3 Satz (Zweite Pfadregel)

$\mathbb{P}$ sei die Kopplung von p_1 und p_2. Sei $A_2 \subseteq \Omega_2$ und $A := \Omega_1 \times A_2$.

(A bezieht sich nur auf das zweite Teilexperiment!)

Dann gilt:

$$\mathbb{P}(A) \ = \sum_{\omega_2 \in A_2} \left(\sum_{\omega_1 \in \Omega_1} p_1(\omega_1) \cdot p_2(\omega_1, \omega_2) \right)$$

(Summation der W'en aller Pfade, die zu einem Endknoten in A_2 führen).

Interpretation der Übergangswahrscheinlichkeit $p_2(\omega_1, \omega_2)$:

W', dass das zweite Teilexperiment den Ausgang ω_2 hat, wenn das erste Teilexperiment den Ausgang ω_1 hat.

8.4 Bemerkung (n-stufige Experimente)

Die Modellierung n-stufiger Experimente geschieht induktiv mit einer *Startverteilung* p_1 und *Übergangs-W-Funktionen* p_2, $p_3, \ldots, p_n$.

Seien etwa p_1 und p_2 wie oben und p_3 eine Übergangs-W-Funktion von $\Omega_1 \times \Omega_2$ nach Ω_3, d.h. eine Abbildung $p_3 : \Omega_1 \times \Omega_2 \times \Omega_3 \to \mathbb{R}$ mit:

$$\begin{aligned} p_3(\omega_1, \omega_2, \omega_3) &\geq 0, \quad (\omega_1, \omega_2, \omega_3) \in \Omega_1 \times \Omega_2 \times \Omega_3, \\ \sum_{\omega_3 \in \Omega_3} p_3(\omega_1, \omega_2, \omega_3) &= 1, \quad (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2. \end{aligned}$$

Dann definiert die erste Pfadregel

$$p(\omega_1, \omega_2, \omega_3) := p_1(\omega_1) \cdot p_2(\omega_1, \omega_2) \cdot p_3(\omega_1, \omega_2, \omega_3)$$

eine W-Funktion auf $\Omega_1 \times \Omega_2 \times \Omega_3$.

Das zu p gehörende W-Maß heißt *Kopplung* von p_1, p_2 und p_3.

Allg.: Kopplung von $p_1,\ p_2, \ldots, p_n$.

8.5 Beispiel (Pólyasches Urnenschema, allgemein)

Urne mit r roten und s schwarzen Kugeln. Rein zufällig Kugel entnehmen, Farbe notieren. Diese sowie c weitere Kugeln derselben Farbe zurücklegen.

Dabei $c < 0$ möglich (dann $|c|$ Kugeln entnehmen).

$c = 0$ und $c = -1$ bedeuten Ziehen mit bzw. ohne Zurücklegen.

Vorgang nach gutem Mischen $n - 1$ mal wiederholen.

$\mathbb{P}$(„genau k mal eine rote Kugel ziehen“) $=$?

Modell: $\Omega \ := \ \Omega_1 \times \cdots \times \Omega_n, \quad \Omega_j := \{0, 1\}$, $1 =$ rot, $0 =$ schwarz.

Startverteilung:

$$p_1(1) \ := \ \frac{r}{r+s} \ , \qquad p_1(0) \ := \ \frac{s}{r+s}$$

Übergangswahrscheinlichkeiten?

Beachte: $\omega \ = \ (a_1, \ldots, a_n)$, $a_j = 1 \ (0)$, falls j-te Kugel rot (schwarz).

Für $j = 2, \ldots, n$ gilt:

$$\sum_{\nu=1}^{j-1} a_\nu = l \iff \text{genau } l \text{ mal "rot" in den ersten } j-1 \text{ Ziehungen.}$$

Inhalt vor j-ter Ziehung: $r + l \cdot c$ rote und $s + (j-1-l) \cdot c$ schwarze Kugeln.
Also: Falls $\sum_{\nu=1}^{j-1} a_\nu = l$, so

$$\begin{aligned} p_j(a_1, \ldots, a_{j-1}, 1) &:= \frac{r + l \cdot c}{r + s + (j-1) \cdot c}, \\ p_j(a_1, \ldots, a_{j-1}, 0) &:= \frac{s + (j-1-l) \cdot c}{r + s + (j-1) \cdot c}. \end{aligned}$$

Nach erster Pfadregel ist für $\omega = (a_1, \ldots, a_n) \in \Omega$ und $k = 0, \ldots, n$

$$p(\omega) := \frac{\prod_{j=0}^{k-1}(r + j \cdot c) \cdot \prod_{j=0}^{n-k-1}(s + j \cdot c)}{\prod_{j=0}^{n-1}(r + s + j \cdot c)}, \quad \text{falls} \quad \sum_{j=1}^{n} a_j = k.$$

Dabei: Produkt über die leere Menge als 1 definiert.

Beachte: $p(\omega)$ hängt nur von Anzahl der Einsen, also $\sum_{j=1}^{k} a_j$, ab!

Sei $A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j = 1\}$ (j-te Kugel rot),

$$X := \sum_{j=1}^{n} \mathbf{1}\{A_j\} \qquad \text{(Anzahl der gezogenen roten Kugeln)}$$

Es gilt $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r}) = \mathbb{P}(A_1 \cap \ldots \cap A_r)\ \forall r\, \forall 1 \le i_1 < \ldots < i_r \le n$ (!) sowie

$$\mathbb{P}(X = k) = \binom{n}{k} \cdot \frac{\prod_{j=0}^{k-1}(r + j \cdot c) \prod_{j=0}^{n-k-1}(s + j \cdot c)}{\prod_{j=0}^{n-1}(r + s + j \cdot c)}, \quad k = 0, \ldots, n.$$

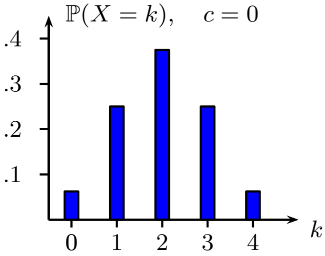
8.6 Definition und Satz

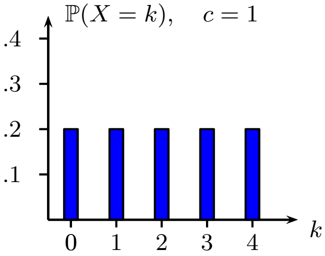
Die Verteilung von X heißt *Pólya-Verteilung* mit Parametern n, r, s und c, kurz:

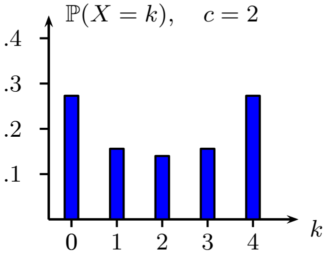
$$X \sim \mathrm{Pol}(n, r, s, c) \quad \text{oder} \quad \mathbb{P}^X = \mathrm{Pol}(n, r, s, c).$$

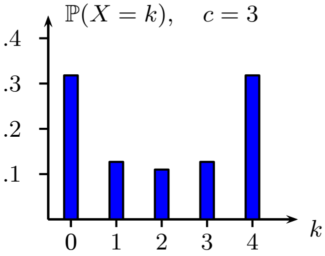
Im Fall $\sim \mathrm{Pol}(n, r, s, c)$ gilt $\mathbb{E}\, X = n \cdot \dfrac{r}{r+s}$.

Beachte: $\mathrm{Pol}(n, r, s, 0) = \mathrm{Bin}(n, r/(r+s))$, $\mathrm{Pol}(n, r, s, -1) = \mathrm{Hyp}(n, r, s)$.









Stabdiagramme der Pólya-Verteilungen Pol$(4, 1, 1, c)$ mit $c = 0, 1, 2, 3$

8.7 Produktexperimente, Markov-Ketten

Laufen die n Teilexperimente unbeeinflusst (getrennt) voneinander ab, so ist

$$p_j(a_1, \ldots, a_{j-1}, a_j) \ =: p_j(a_j) \quad \text{(unabhängig von } a_1, \ldots, a_{j-1}\text{)}$$

mit einer W-Funktion p_j auf Ω_j.

Es gilt dann

$$p(\omega) \ = \ p_1(\omega_1) \cdot \ldots \cdot p_n(\omega_n), \qquad \omega = (\omega_1, \ldots, \omega_n) \in \Omega.$$

Man spricht in diesem Fall von einem *Produktexperiment*.

Spezialfall: $\Omega_1 = \ldots = \Omega_n$, $p_1 = \ldots = p_n$

(n-fache unabhängige Wiederholung eines Experiments)

Sei $S := \Omega_1 = \ldots = \Omega_n$.

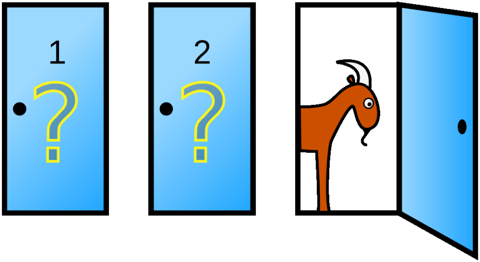
Sei $p_j(a_1, \ldots, a_{j-1}, a_j) \ \forall \, j \geq 2 \ \forall \, a_1, \ldots, a_j$ nur von a_{j-1} abhängig.

Sei $X_j(\omega) := a_j$, $\omega = (a_1, \ldots, a_n) \in S^n$.

Dann heißt $X_1, \ldots, X_n$ *Markov-Kette mit Zustandsraum S.*

9 Bedingte Wahrscheinlichkeiten

Das Ziegen-Problem (Drei-Türen-Problem, Monty-Hall-Problem)



Hinter einer von drei Türen befindet sich ein Auto, hinter den beiden anderen jeweils eine Ziege. Der Kandidat zeigt auf Tür 1; diese bleibt zunächst verschlossen. Der Moderator weiß, hinter welcher Tür sich das Auto befindet. Er darf die Auto-Tür 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?

Über einen stochastischen Vorgang sei bekannt, dass ein Ereignis B eingetreten ist. Wie beeinflusst diese Bedingung (Information) die Aussicht auf das Eintreten eines Ereignisses A?

Motivierung bedingter Wahrscheinlichkeiten mithilfe relativer Häufigkeiten

Experiment mit Grundraum Ω n mal „in unabhängiger Folge“ wiederholen.

Seien $A, B \subseteq \Omega$.

Sei $h_n(B)$ die Anzahl der Male, bei denen das Ereignis B eintritt,

$h_n(A \cap B)$ die Anzahl der Male, bei denen sowohl A als auch B eintreten.

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

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

Memo: $$\frac{h_n(A\cap B)}{h_n(B)} = \frac{\frac{1}{n}h_n(A\cap B)}{\frac{1}{n}h_n(B)} \quad \begin{matrix}\rightsquigarrow \mathbb{P}(A\cap B)\\ \rightsquigarrow \mathbb{P}(B)\end{matrix}$$

9.1 Definition (bedingte Wahrscheinlichkeit, bedingte Verteilung)

Seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum, $A, B \subseteq \Omega$ mit $\mathbb{P}(B) > 0$. Dann heißt

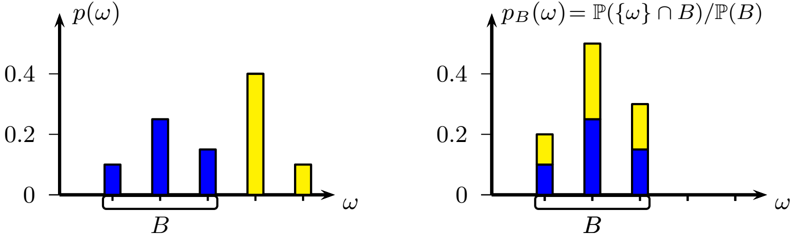
$$\mathbb{P}(A|B) \ := \ \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

die *bedingte Wahrscheinlichkeit von A unter der Bedingung B*.

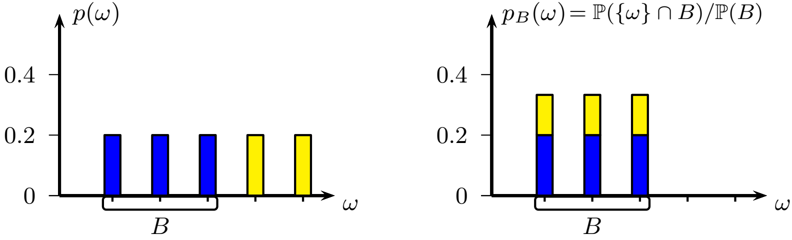
Die durch

$$\mathbb{P}_B(A) \ := \ \mathbb{P}(A|B), \qquad A \subseteq \Omega,$$

definierte Funktion $\mathbb{P}_B : \mathcal{P}(\Omega) \to \mathbb{R}$ heißt *bedingte Verteilung von $\mathbb{P}$ unter der Bedingung B*.



Übergang zur bedingten Verteilung



Bedingte Verteilung $\mathbb{P}_B$ der Gleichverteilung $\mathbb{P}$ auf Ω ist Gleichverteilung auf B.

Memo: $\mathbb{P}_B(A) \;=\; \mathbb{P}(A|B) \;=\dfrac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}, \qquad A \subseteq \Omega,$

Memo: $\mathbb{P}_B(A) \ge 0 \;\; \surd \quad \mathbb{P}_B(\Omega) = 1 \;\; \surd \quad \mathbb{P}_B\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}_B(A_j) \;\; \surd$

9.2 Satz

Die bedingte Verteilung $\mathbb{P}_B$ ist ein W-Maß auf $\mathcal{P}(\Omega)$.

BEWEIS:

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

9.3 Zusammenhang mit Übergangswahrscheinlichkeiten

Betrachte zweistufiges Experiment:

$$\Omega = \Omega_1 \times \Omega_2, \quad \omega = (a_1, a_2), \quad p(\omega) := p_1(a_1) p_2(a_1, a_2)$$

$$\mathbb{P}(A) \ = \ \sum_{\omega \in A} p(\omega), \qquad A \subseteq \Omega.$$

Seien $a_1 \in \Omega_1$, $a_2 \in \Omega_2$.

$$B \ := \ \{a_1\} \times \Omega_2 \qquad \text{(beim 1. Teilexperiment tritt Ergebnis } a_1 \text{ auf)},$$
$$A \ := \ \Omega_1 \times \{a_2\} \qquad \text{(beim 2. Teilexperiment tritt Ergebnis } a_2 \text{ auf)}.$$

Es gilt $A \cap B \ = \ \{(a_1, a_2)\} \qquad \mathbb{P}(A \cap B) = p_1(a_1) \cdot p_2(a_1, a_2)$,

$$\mathbb{P}(B) \ = \ \sum_{\widetilde{a}_2 \in \Omega_2} p_1(a_1) \cdot p_2(a_1, \widetilde{a}_2) \ = \ p_1(a_1) \cdot \underbrace{\sum_{\widetilde{a}_2 \in \Omega_2} p_2(a_1, \widetilde{a}_2)}_{=1} \ = \ p_1(a_1)$$

$$\Longrightarrow \ \mathbb{P}(A|B) \ = \ p_2(a_1, a_2).$$

Übergangswahrscheinlichkeiten sind bedingte Wahrscheinlichkeiten!

$$\frac{h_n(A\cap B)}{h_n(B)} = \frac{\frac{1}{n}h_n(A\cap B)}{\frac{1}{n}h_n(B)} \quad \begin{array}{l} \rightsquigarrow \mathbb{P}(A\cap B) \\ \rightsquigarrow \mathbb{P}(B) \end{array}$$

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$ (bedingte W' von A unter der Bedingung B)

$$\mathbb{P}_B(A) := \mathbb{P}(A|B), \qquad A\subseteq \Omega.$$

$\mathbb{P}_B$ ist ein W-Maß (bedingte Verteilung von $\mathbb{P}$ unter der Bedingung B).

Zusammenhang mit Übergangswahrscheinlichkeiten:

$$\Omega := \Omega_1\times\Omega_2, \quad \omega=(a_1,a_2), \quad p(\omega):=p_1(a_1)p_2(a_1,a_2),$$

$$B := \{a_1\}\times\Omega_2, \quad A := \Omega_1\times\{a_2\}, \quad A\cap B = \{(a_1,a_2)\}.$$

$$\mathbb{P}(A\cap B) = p_1(a_1)\,p_2(a_1,a_2), \qquad \mathbb{P}(B) = p_1(a_1)$$

$$\Longrightarrow \mathbb{P}(A|B) = p_2(a_1,a_2).$$

Beachte: Es gilt

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

Hierbei meist $\mathbb{P}(A)$ und $\mathbb{P}(B|A)$ als Modellbausteine gegeben!

9.4 Satz (Multiplikationsformel)

Es seien $A_1, \ldots, A_n$ Ereignisse mit $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt

$$\begin{aligned} & \mathbb{P}(A_1 \cap \ldots \cap A_n) \\ = \;& \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot \ldots \cdot \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}). \end{aligned}$$

Standardbeispiel ist gekoppeltes n-stufiges Experiment mit

$$A_j \;=\; \Omega_1 \times \ldots \times \Omega_{j-1} \times \{a_j\} \times \Omega_{j+1} \times \ldots \times \Omega_n$$

Dann: $\mathbb{P}(A_j|A_1 \cap \ldots \cap A_{j-1}) \;=\; p_j(a_1, \ldots, a_{j-1}, a_j)$, und die Multiplikationsregel ist die erste Pfadregel.

9.5 Satz (Formel von der totalen Wahrscheinlichkeit, Bayes-Formel)

Es seien $A_1, A_2, \ldots$ paarweise disjunkte Ereignisse mit $\sum_{j=1}^{\infty} A_j = \Omega$ (sogenannte *Zerlegung* von Ω). Weiter sei B ein Ereignis. Dann gelten:

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

Hierbei: $\mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j) := 0$, falls $\mathbb{P}(A_j) = 0$.

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

$$\mathbb{P}(A_k|B) \;=\; \frac{\mathbb{P}(A_k) \cdot \mathbb{P}(B|A_k)}{\sum_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j)}$$ (*Bayes-Formel*)

Bemerkung: Im Zusammenhang mit der Bayes-Formel nennt man

$\mathbb{P}(A_1), \ldots, \mathbb{P}(A_n) \ldots$ *a-priori-Wahrscheinlichkeiten*
$\mathbb{P}(A_1|B), \ldots, \mathbb{P}(A_n|B) \ldots$ *a-posteriori-Wahrscheinlichkeiten*.

Memo: a) z.z: $\mathbb{P}(B) = \sum_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j)$

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

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

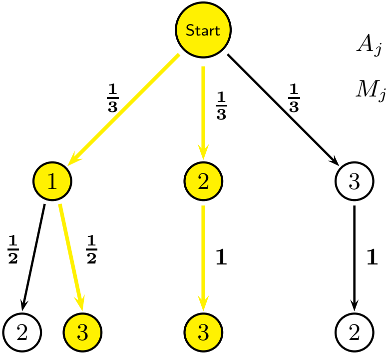
Memo: b) z.z: $\mathbb{P}(A_k|B) = \dfrac{\mathbb{P}(A_k) \cdot \mathbb{P}(B|A_k)}{\sum_{j=1}^{\infty} \mathbb{P}(A_j) \cdot \mathbb{P}(B|A_j)}$

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

9.6 Beispiel (Das Drei-Türen-Problem, „Ziegen-Problem“)

Hauptgewinn rein zufällig hinter einer von 3 Türen platziert.

Kandidat wählt Tür 1. Moderator muss Niete zeigen, darf Tür 1 nicht öffnen.
Moderator öffnet Tür 3. Soll man wechseln?



$A_j := \{$„Auto hinter Tür j“$\}$

$M_j := \{$„Moderator öffnet Tür j“$\}$

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3}$$

$$\mathbb{P}(M_2|A_1) = \frac{1}{2} = \mathbb{P}(M_3|A_1)$$

$$\mathbb{P}(M_3|A_2) = 1 = \mathbb{P}(M_2|A_3)$$

$$\mathbb{P}(A_2|M_3) = \frac{\mathbb{P}(A_2)\mathbb{P}(M_3|A_2)}{\mathbb{P}(A_1)\mathbb{P}(M_3|A_1) + \mathbb{P}(A_2)\mathbb{P}(M_3|A_2)} = \frac{2}{3}$$

→ wechseln!

9.7 Beispiel (Test auf eine seltene Krankheit)

$\Omega \ := \ \{(0,\ominus),(0,\oplus),(1,\ominus),(1,\oplus)\}$,

1 (0) in erster Komponente: krank bzw. gesund

$\oplus$ $(\ominus)$ in zweiter Komponente: positives (negatives) Testergebnis

$K := \{(1,\ominus),(1,\oplus)\}$ „krank“, $G := \{(0,\ominus),(0,\oplus)\}$ „gesund“ ,

$\ominus = \{(1,\ominus),(0,\ominus)\}$ „Test negativ“ , $\oplus = \{(1,\oplus),(0,\oplus)\}$ „Test positiv“

Modellannahmen: $\mathbb{P}(K) \ = \ q$,

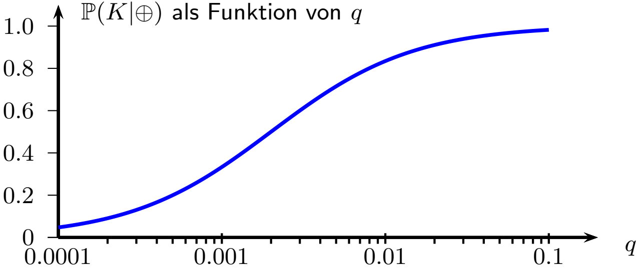
$$\mathbb{P}(\oplus|K) \ = \ p_{se}, \quad \text{„Sensitivität“}, \quad \mathbb{P}(\ominus|G) \ = \ p_{sp} \quad \text{„Spezifität“}.$$

$$\text{Bayes-Formel} \implies \mathbb{P}(K|\oplus) \ = \ \frac{\mathbb{P}(K)\cdot\mathbb{P}(\oplus|K)}{\mathbb{P}(K)\cdot\mathbb{P}(\oplus|K) \ + \ \mathbb{P}(G)\cdot\mathbb{P}(\oplus|G)}.$$

$\mathbb{P}(G) = 1-q, \quad \mathbb{P}(\oplus|G) = 1-p_{sp} \implies$

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

$p_{se} = p_{sp} = 0.998$



1 000 000 Menschen Bluttest unterziehen, davon 1000 krank.

Von den 1000 Kranken erhalten ca. 998 ein positives Ergebnis.

Von den 999 000 Gesunden erhalten ca. 1 998 ein positives Ergebnis.

Insgesamt gibt es ca. 2 996 Personen mit positivem Testergebnis.
Davon ist nur ca. ein Drittel krank.

Sinn und Unsinn von Reihenuntersuchungen (!)

9.8 Beispiel (Eine männerfeindliche Universität?)

	Frauen		Männer	
	Bewerberinnen	zugelassen	Bewerber	zugelassen
Fach 1	900	720	200	**180**
Fach 2	100	20	800	**240**
Summe	**1000**	**740**	**1000**	**420**

$$\mathbf{0.74 \;=\; 0.9 \cdot 0.8 \;+\; 0.1 \cdot 0.2, \qquad 0.42 \;=\; 0.2 \cdot 0.9 \;+\; 0.8 \cdot 0.3}$$

9.9 Beispiel (Das Simpson-Paradoxon)

Für Ereignisse $A, B \subseteq \Omega$ sowie $\Omega = K_1 + \ldots + K_n$ kann Folgendes gelten:

$$\mathbb{P}(B|A \cap K_j) > \mathbb{P}(B|A^c \cap K_j) \text{ für jedes } j = 1, \ldots, n$$

und (!) $\mathbb{P}(B|A) < \mathbb{P}(B|A^c)$. (*Simpson-Paradoxon*)

Im obigen Beispiel: $n = 2$, $K_j = \{\text{Bewerbung in Fach } j\}$,

$$\begin{aligned} B &= \{\text{aus allen 2000 Bewerbern zufällig gewählte Person wird zugelassen}\}, \\ A &= \{\text{aus allen 2000 Bewerbern zufällig gewählte Person ist männlich}\}. \end{aligned}$$

Beachte: Nach der Formel von der totalen W' für $\mathbb{P}_A$ und $\mathbb{P}_{A^c}$ gilt

$$\begin{aligned} \mathbb{P}_A(B) &= \sum_{j=1}^{n} \mathbb{P}_A(K_j) \cdot \mathbb{P}_A(B|K_j) \qquad (0.42 = 0.2 \cdot 0.9 + 0.8 \cdot 0.3) \\ \mathbb{P}_{A^c}(B) &= \sum_{j=1}^{n} \mathbb{P}_{A^c}(K_j) \cdot \mathbb{P}_{A^c}(B|K_j) \quad (0.74 = 0.9 \cdot 0.8 + 0.1 \cdot 0.2) \end{aligned}$$

Jahreseinkommen (pro Person in $)	Einkommen (in 1000 $)	gezahlte Steuer (in 1000 $)	durchschnittlicher Steueranteil
1974			
< 5000	41 651 643	2 244 467	5,4%
5000 bis 9999	146 400 740	13 646 348	9,3%
10000 bis 14999	192 688 922	21 449 597	11,1%
15000 bis 99999	470 010 790	75 038 230	16,0%
≥ 100000	29 427 152	11 311 672	38,4%
Insgesamt	880 179 247	123 690 314	14,1%
1978			
< 5000	19 879 622	689 318	3,5%
5000 bis 9999	122 853 315	8 819 461	7,2%
10000 bis 14999	171 858 024	17 155 758	10,0%
15000 bis 99999	865 037 814	137 860 951	15,9%
≥ 100000	62 806 159	24 051 698	38,3%
Insgesamt	1 242 434 934	188 577 186	15,2%

A bzw. A^c = Menge der 1974 bzw. 1978 als Einkommen erzielten Dollar
B = Menge der 1974 oder 1978 gezahlten „Steuer-Dollar“
$K_1, \ldots, K_5$ sind die Einkommenskategorien

$\mathbb{P}(B|A) < \mathbb{P}(B|A^c) \qquad \mathbb{P}(B|A \cap K_j) > \mathbb{P}(B|A^c \cap K_j)$ für $j = 1, \ldots, 5$

9.10 Beispiel (Sterbetafeln)

	männlich		weiblich	
Vollendetes Alter	Sterbew' in $[x, x+1)$	Überlebensw' in $[x, x+1)$	Sterbew' in $[x, x+1)$	Überlebensw' in $[x, x+1)$
x	q_x	p_x	q_x	p_x
53	0.0062421	0.9937579	0.0032661	0.9967339
54	0.0069945	0.9930055	0.0036235	0.9963765
55	0.0075380	0.9924620	0.0037740	0.9962260
56	0.0080322	0.9919678	0.0040909	0.9959091
57	0.0089120	0.9910880	0.0044749	0.9955251
58	0.0098247	0.9901754	0.0048468	0.9951532
59	0.0107287	0.9892713	0.0050292	0.9949708
60	0.0113516	0.9886484	0.0053715	0.9946285
61	0.0124905	0.9875095	0.0058549	0.9941451
62	0.0136614	0.9863386	0.0063575	0.9936425
63	0.0149324	0.9850676	0.0070056	0.9929944
64	0.0162704	0.9837296	0.0075398	0.9924602
65	0.0179300	0.9820700	0.0084661	0.9915339

Sterbetafel 2001/2003 für Deutschland (Quelle: Statistisches Bundesamt 2004)

10 Stochastische Unabhängigkeit

Ergebnisse von 25 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 (ausgedacht)

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 (Pseudo-Zufallszahlen)

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 (ausgedacht)

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 (gewürfelt)

Begriffliche Schwierigkeiten!

- Stein-Schere-Papier (kann ich unabhängig wechseln?)
- Roulette (werden Zahlen bei längerem Ausbleiben wahrscheinlicher?)
- Lotto (s.o.)

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

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

Meist gilt

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \;\neq\; \mathbb{P}(A).$$

Beispiel: Ziehen ohne Zurücklegen aus „r/s-Urne“.

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

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

Falls $\mathbb{P}(A|B) = \mathbb{P}(A)$, so hat Eintreten von B *wahrscheinlichkeitstheoretisch* keinen Einfluss auf das Eintreten von A.

In gleicher Weise bedeutet

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

Die Wahrscheinlichkeit des Eintretens von B ist „unabhängig“ von der Information „A geschieht“.

Beachte:

$$\begin{array}{rcl} \mathbb{P}(A|B) & = & \mathbb{P}(A) \iff \mathbb{P}(A \cap B) \;=\; \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(B|A) & = & \mathbb{P}(B) \iff \mathbb{P}(A \cap B) \;=\; \mathbb{P}(A)\mathbb{P}(B). \end{array}$$

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

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

10.1 Definition (stochastische Unabhängigkeit)

Ereignisse $A_1, \ldots, A_n$ heißen *(stochastisch) unabhängig*, falls gilt:

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

für *jede* Menge $T \subseteq \{1, 2, \ldots, n\}$ mit $|T| \geq 2$. ($2^n - n - 1$ Gleichungen)

Folgerung: Drei Ereignisse A, B und C sind unabhängig $\Longleftrightarrow$

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

Die ersten 3 Gleichungen bedeuten die *paarweise Unabhängigkeit* von A, B, C.

10.2 Wichtige Punkte im Zusammenhang mit 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: Urne mit 2 roten und einer schwarzen Kugel; Ziehen ohne Z.

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

  $\mathbb{P}(B) = 2/3, \quad \mathbb{P}(B|A) = 1/2 \Longrightarrow$ A und B *nicht* unabhängig

  B ist real beeinflusst von A, aber nicht A von B !

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

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

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

Induktiv gilt:

10.3 Satz (Unabhängigkeit und Komplementbildung)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $A_1, \ldots, A_n \subseteq \Omega$ Ereignisse, $n \geq 2$.

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:

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

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

a) $\Longrightarrow$ b) folgt durch Induktion über $|J|$.

10.4 Beispiel (Unabhängigkeit und Produktexperimente)

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

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

Sei für $j \in \{1, \ldots, n\}$ $A_j^* \subseteq \Omega_j$,

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

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

Beachte: Sind $C_1 \subseteq \Omega_1$, $C_2 \subseteq \Omega_2$, $\ldots, C_n \subseteq \Omega_n$, so gilt mit $\omega = (a_1, \ldots, a_n)$

$$\begin{aligned}
\mathbb{P}(C_1 \times \ldots \times C_n) \;&=\; \sum_{\omega \in C_1 \times \ldots \times C_n} p(\omega) \;=\; \sum_{a_1 \in C_1} \cdots \sum_{a_n \in C_n} \prod_{j=1}^{n} p_j(a_j) \\
&=\; \left(\sum_{a_1 \in C_1} p_1(a_1) \right) \cdot \ldots \cdot \left(\sum_{a_n \in C_n} p_n(a_n) \right) \\
&=\; \mathbb{P}_1(C_1) \cdot \ldots \cdot \mathbb{P}_n(C_n).
\end{aligned}$$

Memo: $\mathbb{P}(C_1 \times \ldots \times C_n) = \mathbb{P}_1(C_1) \cdot \ldots \cdot \mathbb{P}_n(C_n)$

Memo: $(C_1 \times \ldots \times C_n) \cap (D_1 \times \ldots \times D_n) = (C_1 \cap D_1) \times \ldots \times (C_n \cap D_n)$

Memo: $A_j = \bigtimes_{m=1}^{j-1} \Omega_m \times A_j^* \times \bigtimes_{m=j+1}^{n} \Omega_m$

Beachte: Für $\emptyset \neq T \subseteq \{1, \ldots, n\}$ 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

$$\mathbb{P}\left(\bigcap_{j \in T} A_j\right) \;=\; \mathbb{P}\left(\bigtimes_{j=1}^{n} B_j^*\right) \;=\; \prod_{j=1}^{n} \mathbb{P}_j(B_j^*) \;=\; \prod_{j \in T} \mathbb{P}_j(A_j^*) \;=\; \prod_{j \in T} \mathbb{P}(A_j)$$

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

10.5 Satz (Blockungslemma)

Seien $A_1, \ldots, A_n \subseteq \Omega$, $n \geq 2$, unabhängige Ereignisse sowie $k \in \{1, \ldots, n-1\}$.

Sei B eine mengentheoretische Funktion von $A_1, \ldots, A_k$,

C eine mengentheoretische Funktion von $A_{k+1}, \ldots, A_n$.

Dann sind B und C stochastisch unabhängig.

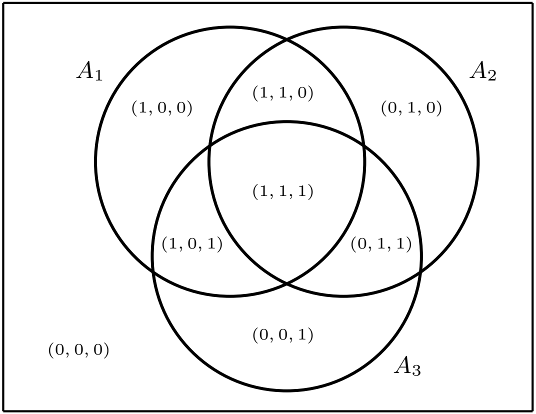
Beachte: Mengentheoretische Funktion von A_1, A_2 ist Vereinigung der paarweise disjunkten Mengen

$$A_1 \cap A_2, \qquad A_1 \cap A_2^c, \qquad A_1^c \cap A_2, \qquad A_1^c \cap A_2^c.$$

Sei allgemein $A^1 := A$ und $A^0 := A^c$.

$$A_1 \cap A_2 \leftrightarrow (1,1), \quad A_1 \cap A_2^c \leftrightarrow (1,0), \quad A_1^c \cap A_2 \leftrightarrow (0,1), \quad A_1^c \cap A_2^c \leftrightarrow (0,0).$$

$$A_1 \cup A_2 \leftrightarrow \{(1,0),(1,1),(0,1)\}$$



Mengentheoretische Funktion von A_1, A_2, A_3 $\quad\leftrightarrow\quad$ Teilmenge $R \subseteq \{0,1\}^3$

z.B.: $A_1 \cap (A_2 \cup A_3^c) \quad\leftrightarrow\quad \{(1,1,0),(1,1,1),(1,0,0)\}$

Allgemein: B mengenth. Funktion von $A_1, \ldots, A_k \Longrightarrow \exists\ R \subseteq \{0,1\}^k$ mit

$$B \;=\; \sum_{r \in R} A_1^{r_1} \cap \ldots \cap A_k^{r_k}, \qquad r := (r_1, \ldots, r_k).$$

Ebenso: C mengenth. Funktion von $A_{k+1}, \ldots, A_n \Longrightarrow \exists\ S \subseteq \{0,1\}^{n-k}$ mit

$$C \;=\; \sum_{s \in S} A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}, \qquad s := (s_1, \ldots, s_{n-k}).$$

Zu zeigen: $A_1, \ldots, A_n$ unabhängig $\Longrightarrow\ \mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C)$.

$$
\begin{aligned}
\mathbb{P}(B \cap C) &= \mathbb{P}\left(\left(\sum_{r\in R} A_1^{r_1} \cap \ldots \cap A_k^{r_k}\right) \cap \left(\sum_{s\in S} A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right)\right) \\
&= \mathbb{P}\left(\sum_{r\in R}\sum_{s\in S} A_1^{r_1} \cap \ldots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right) \text{ (Distrib.gesetz)} \\
&= \sum_{r\in R}\sum_{s\in S} \mathbb{P}\left(A_1^{r_1} \cap \ldots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right) \text{ (σ-Additivität)} \\
&= \sum_{r\in R}\sum_{s\in S} \prod_{i=1}^{k} \mathbb{P}(A_i^{r_i}) \cdot \prod_{j=1}^{n-k} \mathbb{P}(A_{k+j}^{s_j}) \qquad \text{($A_1, \ldots, A_n$ unabhängig)} \\
&= \left(\sum_{r\in R} \prod_{i=1}^{k} \mathbb{P}(A_i^{r_i})\right) \cdot \left(\sum_{s\in S} \prod_{j=1}^{n-k} \mathbb{P}(A_{k+j}^{s_j})\right) \text{ (Distributivgesetz auf $\mathbb{R}$)} \\
&= \left(\sum_{r\in R} \mathbb{P}(A_1^{r_1} \cap \ldots \cap A_k^{r_k})\right) \cdot \left(\sum_{s\in S} \mathbb{P}(A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}})\right) \text{ (A_j unabh.)} \\
&= \mathbb{P}(B) \cdot \mathbb{P}(C) \qquad \text{(σ-Additivität)} \quad \surd
\end{aligned}
$$

10.6 Satz (Erzeugungsweise der Binomialverteilung)

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

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

Beweis: Sei $N := \{1, \ldots, n\}$, $\{N\}_k := \{T \subseteq N : |T| = k\} \ \Longrightarrow$

$$\{X = k\} \ = \ \sum_{T \in \{N\}_k} \left(\bigcap_{i \in T} A_i \cap \bigcap_{j \in N \setminus T} A_j^c \right)$$

$$\mathbb{P}\left(\bigcap_{i \in T} A_i \cap \bigcap_{j \in N \setminus T} A_j^c \right) \ = \ \prod_{i \in T} \mathbb{P}(A_i) \cdot \prod_{j \in N \setminus T} \mathbb{P}(A_j^c) \ = \ p^k \cdot (1-p)^{n-k}$$

$$|\{N\}_k| \ = \ \binom{n}{k} \qquad \surd$$

10.7 Definition (Bernoulli-Kette)

Die Situation von Satz 10.6 ist gegeben mit

- $\Omega \; := \; \Omega_1 \times \ldots \times \Omega_n$,
- $\Omega_j \; := \; \{0,1\}$ für $j = 1,\ldots,n$,
- $\mathbb{P}_j(\{1\}) \; := \; p \; = \; 1 - \mathbb{P}_j(\{0\})$ für $j = 1,\ldots,n$,
- $\mathbb{P}(\{(a_1,\ldots,a_n)\}) = p^{a_1+\ldots+a_n}(1-p)^{n-a_1-\ldots-a_n}, \quad \omega = (a_1,\ldots,a_n) \in \Omega$
- $A_j \; := \; \{(a_1,\ldots,a_n) \in \Omega : a_j = 1\}$

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

Dabei steht 1 bzw. 0 für Treffer bzw. Niete.

10.8 Beispiel (Gruppenscreening)

Viele Personen mögen unabhängig voneinander mit gleicher W' p eine Krankheit besitzen, die durch Blutuntersuchung entdeckt werden kann.

Ziel: Von den Blutproben dieser Personen die Proben mit positivem Befund möglichst kostengünstig herausfinden.

Alternativen:

- alle Blutproben einzeln untersuchen
- Gruppenscreening:
  jeweils Blut von k Personen vermischen und untersuchen.
  - Falls positiver Befund, so jede Person der Gruppe einzeln untersuchen, d.h. $k+1$ Tests für Gruppe
  - Falls negativer Befund, so nur ein Test für k Personen.

Sei Y_k die Anzahl der nötigen Untersuchungen bei Gruppe von k Personen.

$$\mathbb{P}(Y_k = 1) \ = \ (1-p)^k, \qquad \mathbb{P}(Y_k = k+1) \ = \ 1-(1-p)^k$$

$$\begin{aligned} \mathbb{E}(Y_k) &= 1 \cdot (1-p)^k + (k+1) \cdot (1-(1-p)^k) \\ &= k + 1 - k\,(1-p)^k \end{aligned}$$

Ergibt sich im Mittel überhaupt eine Ersparnis durch Gruppenscreening?

$$\mathbb{E}(Y_k) < k \iff 1 < k \cdot (1-p)^k \iff 1 - p > \frac{1}{k^{1/k}} \iff p < 1 - \frac{1}{k^{1/k}}$$

Die Funktion $\mathbb{N} \ni k \mapsto 1/k^{1/k}$ nimmt ihr Minimum $(= 0.6933\ldots)$ für $k = 3$ an.

Folglich lohnt sich Gruppenscreening, falls $p < 1 - 1/3^{1/3} = 0.3066\ldots$.

Aufgabe: Funktion $\mathbb{N} \ni k \mapsto \dfrac{\mathbb{E}(Y_k)}{k}$ bzgl. k minimieren!

Optimale Gruppengröße $k_0 = k_0(p)$ hängt von p ab → Computer!

Erwartete prozentuale Ersparnis pro Person: $\left(1 - \dfrac{\mathbb{E}Y_{k_0}}{k_0}\right) \times$ 100%

p	0.2	0.1	0.05	0.01	0.005	0.001	0.0001
k_0	3	4	5	11	15	32	101
Ersparnis in %	18	41	57	80	86	94	98

Approximation für kleines p:

$$\begin{aligned}
\frac{\mathbb{E}(Y_k)}{k} &= \frac{1}{k}\cdot\Big(1\cdot(1-p)^k + (k+1)\cdot(1-(1-p)^k)\Big)\\
&= \frac{1}{k}\cdot\Big(k+1-k\,(1-p)^k\Big)\\
&= 1+\frac{1}{k}-(1-p)^k\\
&\approx 1+\frac{1}{k}-(1-k\,p)\\
&= \frac{1}{k}+k\,p \;=\; \min_k!
\end{aligned}$$

Betrachte

$$f(x) := \frac{1}{x} + x\,p, \qquad x\in(0,\infty).$$

$$f'(x) = -\frac{1}{x^2}+p$$

$f'(x)=0 \iff x=\frac{1}{\sqrt{p}}$

10.9 Beispiel (Das Zwei-Finger-Morra)

2 Spieler A und B 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.

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

$$\begin{array}{rclrcl} p(1,1) & = & ab, & p(1,2) & = & a(1-b), \\ p(2,1) & = & (1-a)b, & p(2,2) & = & (1-a)(1-b). \end{array}$$

Sei X der Spielgewinn von Spieler A. Es gilt:

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

Erwartungswert von X ist von Spielstrategien a und b abhängig!

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

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

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!

10.10 Beispiel (Unabhängigkeit und Gerichts-(Fehl)-Urteile)

Sally Clark verliert zwei Kinder durch plötzlichen Kindstod.

Nach 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 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 in einem zweiten Berufungsverfahren freigesprochen.

11 Zufallsvektoren, gemeinsame Verteilung

11.1 Beispiel (Zweifacher Würfelwurf, erste und größte Augenzahl)

Sei $\Omega := \{(i,j) : i,j \in \{1,\ldots,6\}\}$, $\quad \mathbb{P} :=$ Gleichverteilung auf Ω

$X(i,j) := i$, $Y(i,j) := \max(i,j)$, $\quad \mathbb{P}(\{X = k\} \cap \{Y = l\}) = ?$

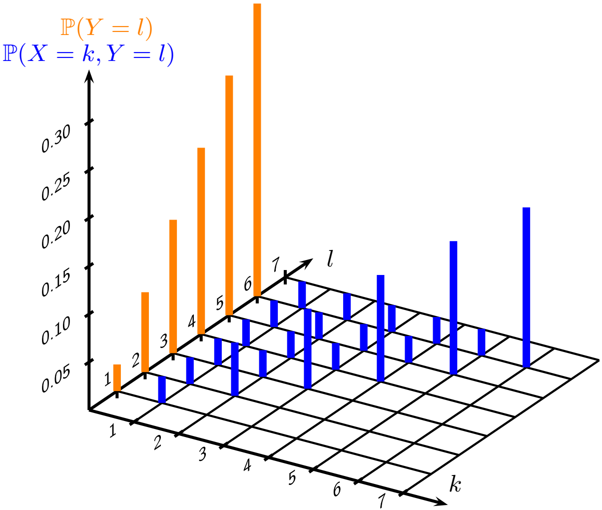
l

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

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

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



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

11.2 Definition (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

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

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

Die Verteilung von X_j heißt j-te Marginalverteilung von X.

Beachte: $\mathbb{P}^X(M_0) = 1$ für abzählbare Menge $M_0 \subseteq \mathbb{R}^n \implies$

$$\{(t, \mathbb{P}(X = t)) : t \in M_0\} \text{ legt } \mathbb{P}^X \text{ fest.}$$

Schreibweisen:

$$\{X \in M\} \quad := \quad \{\omega \in \Omega : X(\omega) \in M\},$$

$$\mathbb{P}(\{X \in M\}) \quad =: \quad \mathbb{P}(X \in M) \text{ wie früher.}$$

Speziell: Falls $M = B_1 \times \ldots \times B_n$ mit $B_j \subseteq \mathbb{R}$, so

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) \quad := \quad \mathbb{P}\left(\bigcap_{j=1}^{n} \{X_j \in B_j\}\right) \;=\; \mathbb{P}(X \in M).$$

$$\mathbb{P}(X_1 = t_1, \ldots, X_n = t_n) \quad := \quad \mathbb{P}\left(\bigcap_{j=1}^{n} \{X_j = t_j\}\right) \;=\; \mathbb{P}(X = t),$$

$t = (t_1, \ldots, t_n) \in \mathbb{R}^n$.

Aus $\mathbb{P}^X$ lassen sich die Marginalverteilungen von $X_1, \ldots, X_n$ bestimmen:

$$\text{Beachte:} \quad \mathbb{P}\left(\sum_{i\geq 1} B_i\right) = 1 \implies \mathbb{P}(A) = \sum_{i\geq 1} \mathbb{P}(A \cap B_i).$$

Hier: Zu jedem $j \in \{1, \ldots, n\}$ existiert abzählbare Menge $M_j =: \{t_{j,i} : i \geq 1\} \subset \mathbb{R}$ mit

$$\mathbb{P}(X_j \in M_j) = \sum_{i\geq 1} \mathbb{P}(X_j = t_{j,i}) = 1.$$

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

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 über $\{(t_2, \ldots, t_n) \in M_2 \times \ldots \times M_n : \mathbb{P}(X_2 = t_2, \ldots, X_n = t_n) > 0\}$)

Die gemeinsame Verteilung ist i. Allg. nicht durch die Marginalverteilungen bestimmt! $c \in [0, 1/2]$ ist ein freier Parameter!

j

		1	2	Σ	
i	1	$\mathbf{c}$	$\frac{1}{2} - \mathbf{c}$	$\frac{1}{2}$	$\mathbb{P}(X = i)$
	2	$\frac{1}{2} - \mathbf{c}$	$\mathbf{c}$	$\frac{1}{2}$	
	Σ	$\frac{1}{2}$	$\frac{1}{2}$	1	

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

Verschiedene gemeinsame Verteilungen mit gleichen Marginalverteilungen

11.3 Definition (Unabhängigkeit von Zufallsvariablen)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $X_1, \ldots, X_n$ Zufallsvariablen auf Ω.

$X_1, \ldots, X_n$ heißen *stochastisch unabhängig*, falls gilt:

$\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ sind unabhängig $\forall\ B_1, \ldots, B_n \subseteq \mathbb{R}$.

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

11.4 Bemerkung

$X_1, \ldots, X_n$ können auch allgemeiner Zufallsvektoren mit unterschiedlichen Dimensionen sein, also X_j $\mathbb{R}^{k_j}$-wertig für $j = 1, \ldots, n$.

Dann ist $B_j \subseteq \mathbb{R}^{k_j}$ für $j = 1, \ldots, n$.

Memo: $X_1, \ldots, X_n$ unabh. $\Longleftrightarrow \{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ unabh. $\forall B_1, \ldots, B_n$

11.5 Satz (Kriterien für Unabhängigkeit)

Folgende Aussagen sind äquivalent:

a) $X_1, \ldots, X_n$ sind stochastisch unabhängig,

b) $\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) \; = \; \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j) \;\; \forall B_1, \ldots, B_n \subseteq \mathbb{R}$,

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

BEWEIS: „a) $\Longrightarrow$ b)“ folgt aus der Definition der Unabhängigkeit von Ereignissen.

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

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

„c) $\Longrightarrow$ b)“:

Memo: b) $\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j) \; \forall B_1, \ldots, B_n \subseteq \mathbb{R}$

Memo: c) $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j = x_j) \quad \forall x_1, \ldots, x_n \in \mathbb{R}.$

Sei $M_j := \{t \in \mathbb{R} : \mathbb{P}(X_j = t) > 0\}$ für $j = 1, \ldots, n$ (abzählbare Menge!).

Seien $B_1, \ldots, B_n \subseteq \mathbb{R}$ und $B_j^* := B_j \cap M_j$ $(\mathbb{P}(X_j \in B_j^*) = \mathbb{P}(X_j \in B_j))$.

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

11.6 Satz (Blockungslemma für Zufallsvariablen)

Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $k \in \{1, \ldots, n-1\}$.

Weiter seien $g : \mathbb{R}^k \to \mathbb{R}$ und $h : \mathbb{R}^{n-k} \to \mathbb{R}$ Funktionen.

Dann sind $g(X_1, \ldots, X_k)$ und $h(X_{k+1}, \ldots, X_n)$ stochastisch unabhängig.

Beweis: Sei $Y_1 := g(X_1, \ldots, X_k)$, $Y_2 := h(X_{k+1}, \ldots, X_n)$. Seien $y_1, y_2 \in \mathbb{R}$.

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

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

Typische Beispiele für Anwendungen des Blockungslemmas sind:

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

usw.

11.7 Satz (Die allgemeine Transformationsformel)

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum und $Z : \Omega \to \mathbb{R}^n$ ein n-dimensionaler Zufallsvektor. Sei $M_0 := \{z \in \mathbb{R}^n : \mathbb{P}(Z = z) > 0\}$.

Weiter sei $g : \mathbb{R}^n \to \mathbb{R}$ eine reelle Funktion. Dann gilt:

Der Erwartungswert von $g(Z)$ existiert $\iff \sum_{z \in M_0} |g(z)| \, \mathbb{P}(Z = z) < \infty$.

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

Beweis:

$$\begin{aligned}
\sum_{\omega \in \Omega_0} |g(Z)(\omega)| \, \mathbb{P}(\{\omega\}) &= \sum_{z \in \mathbb{R}^n : \mathbb{P}(Z=z)>0} \; \sum_{\omega \in \Omega_0 : Z(\omega) = z} |g(Z(\omega))| \, \mathbb{P}(\{\omega\}) \\
&= \sum_{z \in M_0} |g(z)| \sum_{\omega \in \Omega_0 : Z(\omega) = z} \mathbb{P}(\{\omega\}) \\
&= \sum_{z \in M_0} |g(z)| \, \mathbb{P}(Z = z)
\end{aligned}$$

$|\cdot|$ weglassen $\Longrightarrow$ Beh.

Memo: $\mathbb{E}g(Z) = \sum_{z:\mathbb{P}(Z=z)>0} g(z)\,\mathbb{P}(Z=z)$

11.8 Satz (Multiplikationsformel für Erwartungswerte)

Seien X und Y unabhängige Zufallsvariablen mit $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$.

Dann existiert auch der Erwartungswert des Produktes XY, und es gilt

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

BEWEIS: $Z = (X, Y)$, $g(X, Y) = X\,Y$, $\mathbb{P}(Z = z) = \mathbb{P}(X = x, Y = y)$

$$\begin{aligned}\sum_{\omega\in\Omega_0} |X(\omega)Y(\omega)|\,\mathbb{P}(\{\omega\}) &= \sum_{(x,y)\in\mathbb{R}^2:\mathbb{P}(X=x,Y=y)>0} |x|\,|y|\,\mathbb{P}(X = x, Y = y)\\ &= \sum_{(x,y)\in\mathbb{R}^2:\mathbb{P}(X=x,Y=y)>0} |x|\,|y|\,\mathbb{P}(X=x)\,\mathbb{P}(Y=y)\\ &= \left(\sum_{x:\mathbb{P}(X=x)>0} |x|\,\mathbb{P}(X=x)\right)\left(\sum_{y:\mathbb{P}(Y=y)>0} |y|\,\mathbb{P}(Y=y)\right)\end{aligned}$$

$|\cdot|$ weglassen $\Longrightarrow$ Beh.

X, Y unabhängig, $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty \implies \mathbb{E}\,|XY| < \infty$.

Gilt diese Implikation auch, falls X und Y nicht stochastisch unabhängig?

Antwort: Nein!

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

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

eine Wahrscheinlichkeitsverteilung auf $\mathbb{N}$.

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

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

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

11.9 Satz (diskrete Faltungsformel)

Es seien X und Y unabhängige Zufallsvariablen. Dann gilt:

$$\mathbb{P}(X+Y=t) \;=\; \sum_{x\in\mathbb{R}:\mathbb{P}(X=x)>0} \mathbb{P}(X=x)\,\mathbb{P}(Y=t-x), \qquad t\in\mathbb{R}.$$

Beweis:

$$\{X+Y=t\} = \sum_{x\in\mathbb{R}:\mathbb{P}(X=x)>0} \{X=x, Y=t-x\} + \sum_{x\in\mathbb{R}:\mathbb{P}(X=x)=0} \{X=x, Y=t-x\}$$

$$\begin{aligned}\mathbb{P}(X+Y=t) &= \sum_{x\in\mathbb{R}:\mathbb{P}(X=x)>0} \mathbb{P}(X=x, Y=t-x)\\ &= \sum_{x\in\mathbb{R}:\mathbb{P}(X=x)>0} \mathbb{P}(X=x)\,\mathbb{P}(Y=t-x)\end{aligned}$$

11.10 Beispiel (Faltung von Gleichverteilungen)

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

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

Die Faltungsformel liefert für $k \in \{2, 3, \ldots, 2n\}$

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

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

11.11 Satz (Additionsgesetz für die Binomialverteilung)

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: 1. Möglichkeit: Mit Faltungsformel (Übungsaufgabe !)

2. Möglichkeit: Betrachte Bernoulli-Kette der Länge $m+n$ mit Trefferw' p.

Sei $A_j :=$ {„Treffer im j-ten Versuch“}, $j = 1, \ldots, m+n$.

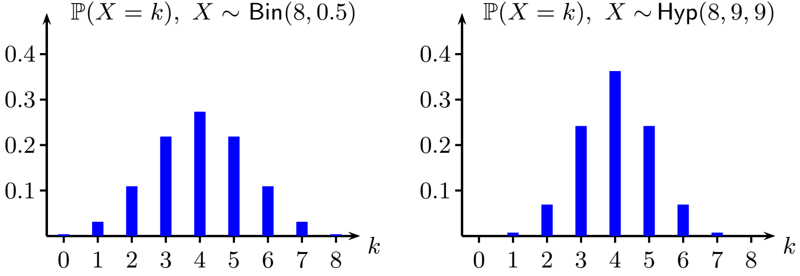
Nach Satz 10.5 gilt:

- $X \; := \; \sum_{j=1}^{m} \mathbf{1}\{A_j\} \sim \mathrm{Bin}(m,p)$
- $Y \; := \; \sum_{j=m+1}^{m+n} \mathbf{1}\{A_j\} \sim \mathrm{Bin}(n,p)$
- $X+Y \; = \; \sum_{j=1}^{m+n} \mathbf{1}\{A_j\} \sim \mathrm{Bin}(m+n,p)$

X und Y sind stochastisch unabhängig (Blockungslemma).

$\mathbb{P}^{X+Y}$ ist durch $\mathbb{P}^X$ und $\mathbb{P}^Y$ festgelegt $\Longrightarrow$ Behauptung.

12 Varianz, Kovarianz, Korrelation



Zwei Verteilungen mit gleichem Erwartungswert, aber unterschiedlicher „Streuung“

12.1 Definition (Varianz und Standardabweichung)

Es sei X eine Zufallsvariable mit $\mathbb{E}X^2 < \infty$. Dann heißen

$$\mathbb{V}(X) \;:=\; \mathbb{E}(X - \mathbb{E}X)^2 \quad \left(:= \mathbb{E}\left[(X - \mathbb{E}\,X)^2\right]\right)$$

die Varianz (*der Verteilung*) von X und

$$+\sqrt{\mathbb{V}(X)}$$

die Standardabweichung oder Streuung (*der Verteilung*) von X.

Beachte:

- $|X| \leq 1 + X^2 \implies \mathbb{E}|X| < \infty$,
- $(X-a)^2 \leq X^2 + 2|a||X| + a^2 \; \forall a \in \mathbb{R} \implies \mathbb{V}(X)$ existiert.

$\mathbb{V}(X)$ ist der Erwartungswert der Zufallsvariablen $g(X)$ mit $g(x) = (x - \mathbb{E}\,X)^2$.

Im Folgenden sei stillschweigend stets $\mathbb{E}\,X^2 < \infty$ vorausgesetzt.

Memo: $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}\,X)^2, \quad \mathbb{E}\,g(X) = \sum_{j\ge 1} g(x_j)\,\mathbb{P}(X = x_j)$

12.2 Satz (Darstellungsformeln für die Varianz)

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

b) Falls $\sum_{j\ge 1} \mathbb{P}(X = x_j) = 1$, so gilt

$$\begin{aligned}\mathbb{V}(X) &= \sum_{j\ge 1} (x_j - \mathbb{E}\,X)^2\,\mathbb{P}(X = x_j)\\ &= \sum_{j\ge 1} x_j^2\,\mathbb{P}(X = x_j) - (\mathbb{E}\,X)^2.\end{aligned}$$

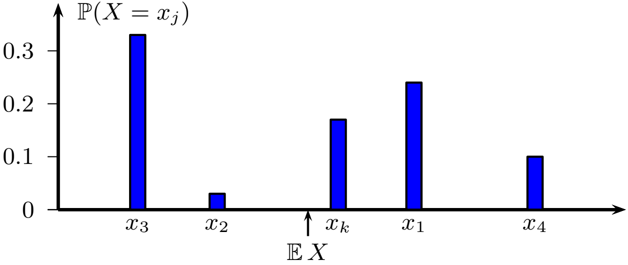
Beweis von a): $(X - \mathbb{E}\,X)^2 = X^2 - 2\cdot\mathbb{E}\,X\cdot X + (\mathbb{E}\,X)^2 \Longrightarrow$

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

12.3 Beispiel (Gleichverteilung auf $\{1, 2, \ldots, k\}$)

Sei $\mathbb{P}(X = j) = 1/k$ für $j = 1, \ldots, k$. Dann gilt

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



Stabdiagramm als Masseverteilung mit Schwerpunkt $\mathbb{E}\, X$.
Mit *Winkelgeschwindigkeit* v um $\mathbb{E}\, X$ rotieren lassen

$$v_j := |x_j - \mathbb{E}X|\, v \qquad \textit{Rotationsgeschwindigkeit} \text{ von } x_j,$$

$$E_j := \frac{1}{2}\,\mathbb{P}(X = x_j)\, v_j^2 \qquad \textit{Rotationsenergie} \text{ von } x_j$$

$$\sum_{j=1}^{k} E_j = \frac{v^2}{2} \underbrace{\sum_{j=1}^{k} (x_j - \mathbb{E}X)^2\, \mathbb{P}(X = x_j)}_{\textit{Trägheitsmoment}} \qquad \textit{gesamte Rotationsenergie}$$

Memo: $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}\,X)^2 = \sum_{\omega\in\Omega_0} (X(\omega) - \mathbb{E}\,X)^2\, \mathbb{P}(\{\omega\})$

Memo: $\mathbb{V}(X) = \mathbb{E}X^2 - (\mathbb{E}\,X)^2$

12.4 Satz (Eigenschaften der Varianz)

a) $\mathbb{V}(X) \geq 0$,

$\mathbb{V}(X) = 0 \iff \mathbb{P}(X = \mathbb{E}X) = 1$,

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

c) $\mathbb{V}(X) = \mathbb{E}(X-t)^2 - (\mathbb{E}X - t)^2, \quad t \in \mathbb{R}$ („*Verschiebungssatz*“),

d) $\mathbb{V}(\mathbf{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A)), \qquad A \subseteq \Omega$,

e) Sind $X_1, \ldots, X_n$ stochastisch unabhängig, so gilt (Beweis mit 12.12 d),e))

$$\mathbb{V}\,(X_1 + \ldots + X_n) = \mathbb{V}(X_1) + \ldots + \mathbb{V}(X_n).$$

BEWEIS: a) $\surd$

b) $$\begin{aligned}\mathbb{V}(aX+b) &= \mathbb{E}\,(aX + b - \mathbb{E}(aX+b))^{\,2} = \mathbb{E}\,(aX + b - a\mathbb{E}\,X - b)^{\,2} \\ &= \mathbb{E}\left(a^2(X - \mathbb{E}\,X)^2\right) = a^2\mathbb{V}(X) \quad \surd\end{aligned}$$

Memo: $\mathbb{E}(aX+bY) = a\mathbb{E}(X)+b\mathbb{E}(Y), \quad \mathbb{V}(X) = \mathbb{E}X^2-(\mathbb{E}\,X)^2$

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

BEWEIS:
$$\begin{aligned}(X-t)^2 &= (X-\mathbb{E}\,X+\mathbb{E}\,X-t)^2\\ &= (X-\mathbb{E}\,X)^2+2(X-\mathbb{E}\,X)(\mathbb{E}X-t)+(\mathbb{E}X-t)^2\end{aligned}$$

$$\Longrightarrow \mathbb{E}(X-t)^2 = \mathbb{V}(X)+2\cdot(\mathbb{E}\,X-t)\cdot 0+(\mathbb{E}\,X-t)^2 \ \surd$$

d) $\mathbb{V}(\mathbf{1}_A) = \mathbb{P}(A)(1-\mathbb{P}(A)), \qquad A\subseteq\Omega$

BEWEIS: $\mathbb{V}(\mathbf{1}_A) = \mathbb{E}\mathbf{1}_A^2-(\mathbb{E}\mathbf{1}_A)^2 = \mathbb{P}(A)-\mathbb{P}(A)^2 \quad \surd$

12.5 Folgerung (Minimaleigenschaft des Erwartungswertes)

Es gilt $\quad \mathbb{V}(X) = \min_{t\in\mathbb{R}} \mathbb{E}(X-t)^2.$

BEWEIS: c) $\Longrightarrow \ \mathbb{E}(X-t)^2 = \mathbb{V}(X)+(\mathbb{E}X-t)^2 \ \surd$

12.6 Satz (Varianz einer Indikatorsumme)

Es seien $A_1, \ldots, A_n \subseteq \Omega$ und $X := \mathbf{1}\{A_1\} + \ldots + \mathbf{1}\{A_n\}$. Dann gilt:

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

Speziell: $\mathbb{P}(A_j) = \mathbb{P}(A_1) \;\; \forall j,\; \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_1 \cap A_2) \;\; \forall i \neq j \;\; \Longrightarrow$

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

Beweis: Wir verwenden $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}\,X)^2$.

$$\begin{aligned} X^2 \;&=\; \left(\sum_{i=1}^{n} \mathbf{1}\{A_i\} \right)^2 \;=\; \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}\{A_i\}\mathbf{1}\{A_j\} \\ &=\; \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}\{A_i \cap A_j\} \\ &=\; \sum_{i=1}^{n} \mathbf{1}\{A_i\} + 2 \sum_{1 \le i < j \le n} \mathbf{1}\{A_i \cap A_j\} \end{aligned}$$

Memo: $X^2 = \sum_{i=1}^{n} \mathbf{1}\{A_i\} + 2 \sum_{1\le i<j\le n} \mathbf{1}\{A_i \cap A_j\}$

$$\begin{aligned}
\Longrightarrow \mathbb{E}\, X^2 &= \sum_{i=1}^{n} \mathbb{P}(A_i) + 2 \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j) \\
(\mathbb{E}\, X)^2 &= \left(\sum_{i=1}^{n} \mathbb{P}(A_i)\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}(A_i)\mathbb{P}(A_j) \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i)^2 + 2 \sum_{1\le i<j\le n} \mathbb{P}(A_i)\mathbb{P}(A_j) \\
\Longrightarrow \mathbb{V}(X) &= \sum_{i=1}^{n} \mathbb{P}(A_i)\left(1 - \mathbb{P}(A_i)\right) + 2 \sum_{1\le i<j\le n} \left(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\right)
\end{aligned}$$

12.7 Beispiel

a) Pólya-Verteilung:

Falls $X \sim \mathrm{Pol}(n,r,s,c)$, so gilt

$$\mathbb{V}(X) \;=\; np(1-p)\left(1+\frac{(n-1)c}{r+s+c}\right), \qquad p := \frac{r}{r+s}.$$

Beachte:

$$\mathbb{P}(A_j) \;=\; \frac{r}{r+s}, \qquad \mathbb{P}(A_i \cap A_j) \;=\; \frac{r(r+c)}{(r+s)(r+s+c)}$$

b) Speziell: Binomialverteilung $(c=0)$

Falls $X \sim \mathrm{Bin}(n,p)$, so gilt $\mathbb{V}(X) \;=\; np(1-p)$.

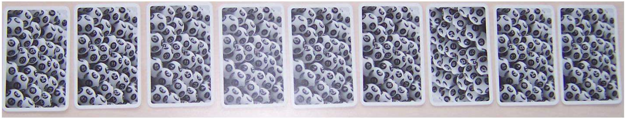
c) Speziell: Hypergeometrische Verteilung $(c=-1)$

Falls $X \sim \mathrm{Hyp}(n,r,s)$, so gilt

$$\mathbb{V}(X) \;=\; np(1-p)\left(1-\frac{n-1}{r+s-1}\right).$$

Mit Darstellungsformel?

12.8 Beispiel (Anzahl der Rekorde in zufälliger Permutation)



$\Omega := Per_n^n(oW)$, $\quad \mathbb{P} :=$ Gleichverteilung auf Ω.

$$\begin{aligned} A_j &:= \{(a_1,\ldots,a_n) \in \Omega : a_j = \max(a_1,a_2,\ldots,a_j)\}, \quad j = 1,\ldots,n, \\ X_n &:= \mathbf{1}\{A_1\} + \ldots + \mathbf{1}\{A_n\} \qquad \text{(Anzahl der Rekorde)} \end{aligned}$$

$A_1,\ldots,A_n$ unabhängig (Übungsaufgabe), $\quad \mathbb{P}(A_j) = \frac{1}{j}, \quad j = 1,\ldots,n.$

$$\begin{aligned} \mathbb{V}(X_n) &= \sum_{j=1}^n \mathbb{V}\left(\mathbf{1}\{A_j\}\right) = \sum_{j=1}^n \mathbb{P}(A_j)(1 - \mathbb{P}(A_j)) \\ &= \sum_{j=1}^n \frac{1}{j}\left(1 - \frac{1}{j}\right) = \sum_{j=1}^n \frac{1}{j} - \sum_{j=1}^n \frac{1}{j^2} \sim \log n \end{aligned}$$

Memo: $\mathbb{E}(aX+b) = a\mathbb{E}\,X + b, \quad \mathbb{V}(aX+b) = a^2\mathbb{V}(X)$

12.9 Satz und Definition (Standardisierung)

Sei X eine Zufallsvariable mit positiver Varianz. Dann heißt

$$\widetilde{X} \;:=\; \frac{X - \mathbb{E}\,X}{\sqrt{\mathbb{V}(X)}}$$

die zu X standardisierte Zufallsvariable oder Standardisierung von X.

Der Übergang von X zu $\dfrac{X - \mathbb{E}X}{\sqrt{\mathbb{V}(X)}}$ heißt Standardisierung.

Es gilt $\mathbb{E}\widetilde{X} \;=\; 0, \;\; \mathbb{V}(\widetilde{X}) \;=\; 1.$

Beim Standardisieren wird durch die Standardabweichung geteilt! Beispiel:

$$X \sim \mathsf{Bin}(n,p) \;\Longrightarrow\; \widetilde{X} = \frac{X - np}{\sqrt{np(1-p)}}.$$

Memo: $U \le V \Longrightarrow \mathbb{E}\,U \le \mathbb{E}\,V, \quad \mathbb{E}\,\mathbf{1}_A = \mathbb{P}(A)$

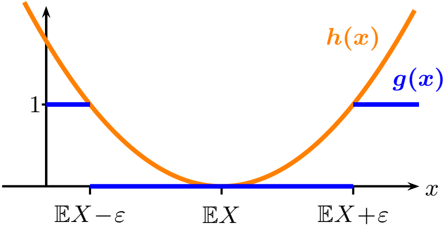
12.10 Satz (Tschebyschow-Ungleichung)

Für jedes (noch so große) $\varepsilon > 0$ gilt

$$\mathbb{P}(|X - \mathbb{E}X| \ge \varepsilon) \;\le\; \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

$$\underbrace{\mathbf{1}\{|X - \mathbb{E}X| \ge \varepsilon\}}_{= g(X)} \le \underbrace{\left(\frac{X - \mathbb{E}X}{\varepsilon}\right)^2}_{= h(X)}$$

$$\mathbb{E}\,g(X) \;\le\; \mathbb{E}\,h(X) \quad \surd$$



Beachte: Ist $\widetilde{X}$ standardisiert, so gilt $\mathbb{P}(|\widetilde{X}| \ge \varepsilon) \;\le\; \dfrac{1}{\varepsilon^2}$.

Es gilt $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$, aber im Allg. $\mathbb{V}(X+Y) \neq \mathbb{V}(X) + \mathbb{V}(Y)$.

$$\begin{aligned}
\mathbb{V}(X+Y) &= \mathbb{E}(X+Y-\mathbb{E}(X+Y))^2 \\
&= \mathbb{E}\big((X-\mathbb{E}X)+(Y-\mathbb{E}Y)\big)^2 \\
&= \mathbb{E}\big[(X-\mathbb{E}X)^2 + 2\,(X-\mathbb{E}X)\,(Y-\mathbb{E}Y) + (Y-\mathbb{E}Y)^2\big] \\
&= \mathbb{E}(X-\mathbb{E}X)^2 + 2\,\mathbb{E}\big[(X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\big] + \mathbb{E}(Y-\mathbb{E}Y)^2 \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{E}\big[(X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\big]
\end{aligned}$$

12.11 Definition (Kovarianz, Unkorreliertheit)

Seien X und Y Zufallsvariablen mit $\mathbb{E}\,X^2 < \infty$ und $\mathbb{E}\,Y^2 < \infty$. Dann heißt

$$\mathrm{C}(X,Y) \;:=\; \mathbb{E}\big[(X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\big]$$

die Kovarianz von (bzw. zwischen) X und Y.

X und Y heißen unkorreliert, falls $\mathrm{C}(X,Y) = 0$ gilt, andernfalls korrreliert.

Beachte: $|uv| \leq \frac{1}{2}(u^2+v^2),\ u,v \in \mathbb{R} \implies \mathrm{C}(X,Y)$ existiert.

Memo: $\mathrm{C}(X,Y) = \mathbb{E}(X-\mathbb{E}\,X)(Y-\mathbb{E}\,Y)$

12.12 Satz (Eigenschaften der Kovarianzbildung)

a) $\mathrm{C}(X,Y) \ = \ \mathbb{E}(X\,Y) - \mathbb{E}X\,\mathbb{E}Y,$ $\surd$

b) $\mathrm{C}(X,X) \ = \ \mathbb{V}(X)$ und $\mathrm{C}(Y,X) \ = \ \mathrm{C}(X,Y),$ $\surd$

c) $\mathrm{C}(X+a,Y+b) \ = \ \mathrm{C}(X,Y),$ $\surd$ (Verschiebungs-Invarianz)

d) X,Y unabhängig $\Longrightarrow \mathrm{C}(X,Y)=0,$ $\surd$ **Umkehrung im Allg. falsch!**

e) $$\mathbb{V}\left(\sum_{j=1}^n X_j\right) \ = \ \sum_{j=1}^n \mathbb{V}(X_j) + 2\sum_{1\le i<j\le n} \mathrm{C}(X_i,X_j),$$

f) $$\mathrm{C}\left(\sum_{i=1}^m a_iX_i, \sum_{j=1}^n b_jY_j\right) = \sum_{i=1}^m\sum_{j=1}^n a_ib_j\mathrm{C}(X_i,Y_j).$$ ($\mathrm{C}(\cdot,\cdot)$ **ist bilinear**)

Die Kovarianz $\mathrm{C}(\cdot,\cdot)$ ist ein bilineares Funktional auf dem Vektorraum $L^2 := \{X:\Omega\to\mathbb{R} : \mathbb{E}(X^2)<\infty\}$.

f):

$$\begin{aligned}
\mathrm{C}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{m} a_i X_i - \mathbb{E}\sum_{i=1}^{m} a_i X_i\right)\left(\sum_{j=1}^{n} b_j Y_j - \mathbb{E}\sum_{j=1}^{n} b_j Y_j\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{m} a_i (X_i - \mathbb{E}\, X_i)\right)\left(\sum_{j=1}^{n} b_j (Y_j - \mathbb{E}\, Y_j)\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j (X_i - \mathbb{E}\, X_i)(Y_j - \mathbb{E}\, Y_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \mathrm{C}(X_i, Y_j)
\end{aligned}$$

e):

$$\begin{aligned}
\mathbb{V}\left(\sum_{j=1}^{n} X_j\right) &= \mathrm{C}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{C}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathbb{V}(X_i) + \sum_{i \neq j} \mathrm{C}(X_i, X_j) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2\sum_{i<j} \mathrm{C}(X_i, X_j)
\end{aligned}$$

Memo: $\mathrm{C}(\cdot,\cdot)$ bilinear , $\mathrm{C}(X,Y)=0$, falls X,Y unabhängig

12.13 Beispiel (Unkorrelierte, nicht unabhängige Zufallsvariablen)

Seien X und Y unabhängige Zufallsvariablen mit identischer Verteilung. Es gilt

$$\begin{aligned} \mathrm{C}(X+Y,X-Y) &= \mathrm{C}(X,X)+\mathrm{C}(Y,X)-\mathrm{C}(X,Y)-\mathrm{C}(Y,Y) \\ &= \mathbb{V}(X)-\mathbb{V}(Y) \ = \ 0 \end{aligned}$$

$\Longrightarrow$ $X+Y$ und $X-Y$ unkorreliert.

Sei speziell $\mathbb{P}(X=j)=\mathbb{P}(Y=j)=1/6$, $j=1,\ldots,6$

$$\begin{aligned} \Longrightarrow \frac{1}{36} &= \mathbb{P}(X+Y=12, X-Y=0) \\ &\neq \mathbb{P}(X+Y=12)\cdot\mathbb{P}(X-Y=0) \\ &= \frac{1}{36}\cdot\frac{1}{6} \end{aligned}$$

$\Longrightarrow$ $X+Y$ und $X-Y$ nicht stochastisch unabhängig.

Vereinfachen Sie unter Verwendung der Bilinearität von $\mathrm{C}(\cdot,\cdot)$:

a)

$$\begin{aligned}\mathrm{C}(2X-3Y,3X-4Y) &= 2\cdot 3\cdot \mathrm{C}(X,X) - 3\cdot 3\cdot \mathrm{C}(Y,X)\\ &\quad - 2\cdot 4\cdot \mathrm{C}(X,Y) + 3\cdot 4\cdot \mathrm{C}(Y,Y)\\ &= 6\,\mathbb{V}(X) - 9\,\mathrm{C}(X,Y) - 8\,\mathrm{C}(X,Y) + 12\,\mathbb{V}(Y)\\ &= 6\,\mathbb{V}(X) - 17\,\mathrm{C}(X,Y) + 12\,\mathbb{V}(Y)\end{aligned}$$

b)

$$\mathrm{C}(aX,-aY+c) = -a^2\mathrm{C}(X,Y)$$

c) $X_1,\ldots,X_n$ unabhängig $\Longrightarrow$

$$\mathrm{C}\left(X_1,\frac{1}{n}\sum_{j=1}^{n}X_j\right) = \frac{1}{n}\cdot\sum_{j=1}^{n}\mathrm{C}(X_1,X_j) = \frac{1}{n}\cdot\mathbb{V}(X_1)$$

Memo: $\mathbb{E}\, g(Z) = \sum_{z:\mathbb{P}(Z=z)>0} g(z)\mathbb{P}(Z=z)$, $\mathrm{C}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\,\mathbb{E}Y$

12.14 Satz (Darstellungsformel für die Kovarianz)

Sei $\sum_{i\geq 1} \mathbb{P}(X = x_i) = 1 = \sum_{j \geq 1} \mathbb{P}(Y = y_j)$. Dann gilt

$$\mathrm{C}(X,Y) \; = \; \sum_{i\geq 1} \sum_{j \geq 1} x_i y_j \, \mathbb{P}(X = x_i, Y = y_j) \; - \; \mathbb{E}\, X \, \mathbb{E}\, Y.$$

- „Ich-weiß-es-nicht-besser-Methode“ zur Berechnung der Kovarianz
- nach Möglichkeit unter Ausnutzung von Satz 12.12 vermeiden!

Memo: $\mathrm{C}(X,Y) = \sum_{i\ge 1} \sum_{j \ge 1} x_i y_j \, \mathbb{P}(X = x_i, Y = y_j) \; - \; \mathbb{E}\,X \, \mathbb{E}\,Y.$

12.15 Beispiel

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

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

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

$$
\begin{aligned}
\mathbb{E}\,X &= \mathbb{E}\,Y = \frac{1}{2}(1+2) = \frac{3}{2}, \\
\mathrm{C}(X,Y) &= 1 \cdot 1 \cdot c + 1 \cdot 2 \cdot \left(\frac{1}{2} - c\right) + 2 \cdot 1 \cdot \left(\frac{1}{2} - c\right) + 2 \cdot 2 \cdot c - \left(\frac{3}{2}\right)^2 \\
&= c - \frac{1}{4}
\end{aligned}
$$

Seien X und Y Zufallsvariablen mit positiven Varianzen.

Mache eine Vorhersage für Y aufgrund von X.

Erlaubte Vorhersagefunktionen $g : \mathbb{R} \to \mathbb{R}$: $g(x) = a + bx$ mit $a, b \in \mathbb{R}$.

Gütekriterium: Mittlere quadratische Abweichung $\mathbb{E}\,(Y - a - bX)^2$.

12.16 Satz

Das Optimierungsproblem $\mathbb{E}(Y - a - bX)^2 = \min_{a,b}!$ hat die Lösung

$$b^* = \frac{\mathrm{C}(X,Y)}{\mathbb{V}(X)}, \quad a^* = \mathbb{E}(Y) - b^*\,\mathbb{E}(X)$$

mit dem Minimalwert $M^* = \mathbb{V}(Y)\,(1 - r^2(X,Y))$, wobei

$$r(X,Y) := \frac{\mathrm{C}(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}.$$

Memo: $\mathbb{E}(Y-a-bX)^2 = \min_{a,b}!$

(Verschiebungssatz!)

↓

$$Z := Y-bX \Longrightarrow \mathbb{E}(Y-a-bX)^2 = \mathbb{E}(Z-a)^2 = \mathbb{V}(Z) + (\mathbb{E}\,Z-a)^2 \geq \mathbb{V}(Z)$$

$$\Longrightarrow a := \mathbb{E}\,Z = \mathbb{E}\,Y - b\,\mathbb{E}\,X.$$

Einsetzen $\Longrightarrow \mathbb{E}(Y-\mathbb{E}\,Y-b(X-\mathbb{E}\,X))^2 = \min_{b\in\mathbb{R}}!$

Sei $\widetilde{Y} := Y - \mathbb{E}\,Y$, $\widetilde{X} := X - \mathbb{E}\,X$,

$$h(b) := \mathbb{E}(\widetilde{Y} - b\widetilde{X})^2 = \min_{b\in\mathbb{R}}!$$

$$\begin{aligned} 0 \leq h(b) &= \mathbb{E}(\widetilde{Y}^2) - 2b\,\mathbb{E}(\widetilde{X}\,\widetilde{Y}) + b^2\mathbb{E}(\widetilde{X}^2) \\ &= \mathbb{V}(Y) - 2b\,\mathrm{C}(X,Y) + b^2\,\mathbb{V}(X) \end{aligned}$$

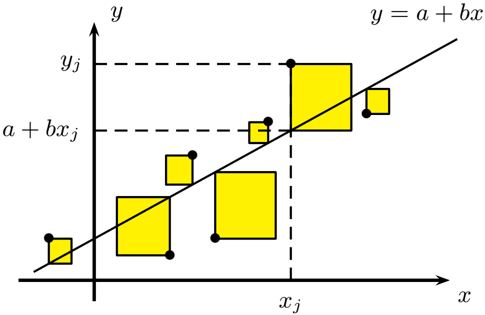
$h'(b) = 0 \Longrightarrow b^* = \mathrm{C}(X,Y)/\mathbb{V}(X)$. Einsetzen von b^* liefert Minimalwert M^*

$$\begin{aligned} 0 \leq M^* &= \mathbb{V}(Y) - 2\frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)} + \frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)} = \mathbb{V}(Y) - \frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)} \\ &= \mathbb{V}(Y)\left(1 - \frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)\,\mathbb{V}(Y)}\right) =: \mathbb{V}(Y)\,(1 - r^2(X,Y)). \end{aligned}$$

12.17 Bemerkung (Methode der kleinsten Quadrate)

Sei $\mathbb{P}(X = x_j, Y = y_j) = 1/n$, $j = 1\ldots, n$. Dann gilt

$$\mathbb{E}(Y - a - bX)^2 \ = \ \frac{1}{n}\sum_{j=1}^{n}(y_j - a - bx_j)^2.$$



Die optimale Gerade $x \mapsto a^*x + b^*$ heißt (empirische) Regressionsgerade von Y auf X.

Memo: $\min_{a,b} \mathbb{E}(Y - a - bX)^2 = \mathbb{V}(Y)\,(1 - r^2(X,Y))$

12.18 Definition ((Pearson-) Korrelationskoeffizient)

a) Es gelte $\mathbb{V}(X)\,\mathbb{V}(Y) > 0$. Der Ausdruck

$$r(X,Y) \;:=\; \frac{\mathrm{C}(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}$$

heißt (Pearson-) Korrelationskoeffizient von X und Y.

b) X und Y heißen $\left\{\begin{array}{l}\text{positiv}\\ \text{un-}\\ \text{negativ}\end{array}\right.$ korreliert, falls $r(X,Y) \left\{\begin{array}{l} > 0\\ = 0\\ < 0\end{array}\right.$ ist.

Memo: $\min_{a,b} \mathbb{E}(Y-a-bX)^2 = \mathbb{V}(Y)\,(1-r^2(X,Y))$

Memo: $r(X,Y) = \dfrac{\mathrm{C}(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}.$

12.19 Satz

Für Zufallsvariablen X, Y gelten:

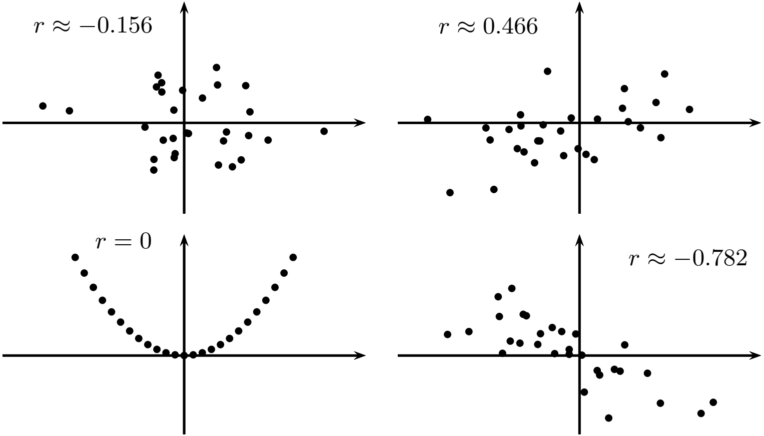
a) $\mathrm{C}(X,Y)^2 \ \le\ \mathbb{V}(X)\,\mathbb{V}(Y)$ (Cauchy-Schwarz-Ungleichung),

b) $-1 \ \le\ r(X,Y) \ \le\ 1$,

c) $|r(X,Y)| = 1 \iff \exists a, b \in \mathbb{R}$ mit $\mathbb{P}(Y = a + bX) = 1$.

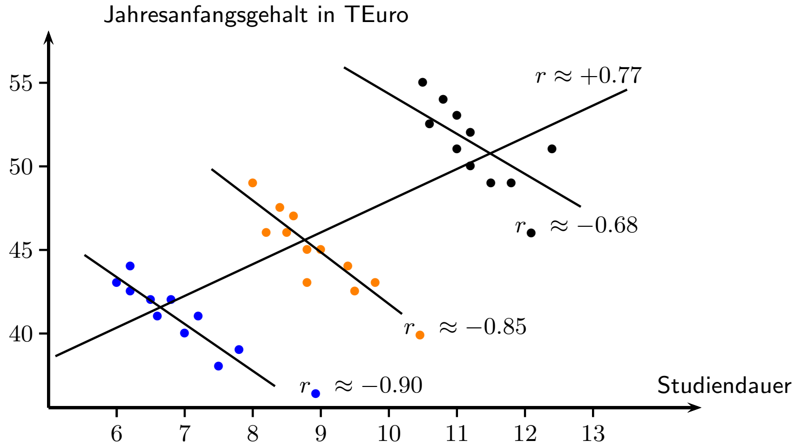
Dabei sei für b) und c) $\mathbb{V}(X) > 0$, $\mathbb{V}(Y) > 0$ vorausgesetzt.

$|r(X,Y)|$ ist Maß für die Vorhersagbarkeit von Y durch affine Funktion von X.



(X, Y) ist jeweils gleichverteilt auf (x_j, y_j), $j = 1, \ldots, n$. In diesem Fall ist

$$r(X,Y) = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \cdot \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2}}, \quad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j,\ \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j.$$



Innerhalb eines jeden Fachs führt schnellerer Abschluss zu höherem Startgehalt.

Global sind Studiendauer und Startgehalt positiv korreliert.

(Simpson-Paradoxon für Korrelationen)

13 Die Multinomialverteilung

Stochastischer Vorgang mit s möglichen Ausgängen, nummeriert von 1 bis s.

Ausgang j heiße Treffer j-ter Art.

Beispiele:

- Würfelwurf ($s = 6$)
- Augensumme beim zweifachen Würfelwurf ($s = 11$)
- Verteilen einer Kugel auf eines von s Fächern
- Drehen eines Glücksrades mit s Sektoren
- Ziehen einer Kugel aus einer Urne, die verschiedenfarbige Kugeln (insgesamt s verschiedene Farben) enthält

Der Ausgang wird notiert und der stochastische Vorgang $n - 1$ mal in unabhängiger Folge wiederholt (insgesamt n Experimente bzw. Versuche).

Modell: Produktexperiment mit

$$\Omega = \{1,\ldots,s\}^n = \{\omega = (a_1,\ldots,a_n) : a_i \in \{1,\ldots,s\} \text{ für } i = 1,\ldots,n\}$$

$a_i = j$ bedeutet „Treffer j-ter Art im i-ten Versuch“.

Sei p_j die Wahrscheinlichkeit für einen Treffer j-ter Art; $\sum_{j=1}^{s} p_j = 1$.

Wovon hängt $p(\omega) := \mathbb{P}(\{\omega\})$, $\omega = (a_1,\ldots,a_n)$, ab?

Nötige Info: Wie viele Treffer 1.Art, 2.Art, …, s-ter Art kommen vor?

z.B. $s = 4$, $n = 7$:

$$p(3,3,1,4,3,2,4) = p_3 \cdot p_3 \cdot p_1 \cdot p_4 \cdot p_3 \cdot p_2 \cdot p_4 = p_1^1 \cdot p_2^1 \cdot p_3^3 \cdot p_4^2$$

Allgemein: Mit $k_j := \sum_{i=1}^{n} \mathbf{1}\{a_i = j\}$ (Anzahl der Treffer j-ter Art) gilt

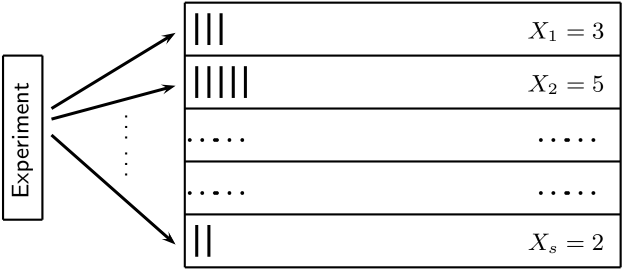
$$p(a_1,\ldots,a_n) = \prod_{j=1}^{s} p_j^{k_j}.$$

Sei $A_{i,j} := \{(a_1, \ldots, a_n) \in \Omega : a_i = j\}$ (Treffer j-ter Art im i-ten Versuch)

Es gilt:

- $\mathbb{P}(A_{i,j}) = p_j$
- Für jede Wahl von $j_1, \ldots, j_n \in \{1, \ldots, s\}$ sind $A_{1,j_1}, A_{2,j_2}, \ldots, A_{n,j_n}$ stochastisch unabhängig

Sei $X_j := \sum_{i=1}^{n} \mathbf{1}\{A_{i,j}\}$ (Anzahl der Treffer j-ter Art, $X_j \sim \text{Bin}(n, p_j)$)



Welche Verteilung besitzt der Zufallsvektor $(X_1, \ldots, X_s)$?

Seien $k_1, \ldots, k_s \in \mathbb{N}_0$ mit $k_1 + \ldots + k_s = n$ (mögliche Trefferanzahlen)

$\{X_1 = k_1, \ldots, X_s = k_s\}$ ist Menge der Tupel $\omega = (a_1, \ldots, a_n) \in \Omega$,
in denen k_1 mal 1, k_2 mal 2 und … und k_s mal s auftritt.

Jedes solche Tupel ω besitzt die W' $p(\omega) = p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$.

Anzahl solcher Tupel?

Von n Plätzen k_1 für 1, danach von den verbleibenden $n - k_1$ Plätzen k_2 für 2 usw. auswählen. Multiplikationsregel $\Longrightarrow$

$$\begin{aligned}\text{Anzahl} &= \binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \binom{n-k_1-k_2}{k_3} \cdot \ldots \cdot \binom{n-k_1-\ldots-k_{s-1}}{k_s} \\ &= \frac{n!}{k_1!(n-k_1)!} \cdot \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \cdot \ldots \cdot \frac{(n-k_1-\ldots-k_{s-1})!}{k_s!(n-k_1-\ldots-k_{s-1}-k_s)!} \\ &= \frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_s!}\end{aligned}$$

(sog. Multinomialkoeffizient)

Multinomialer Lehrsatz:

$$(x_1 + \ldots + x_s)^n \;=\; \sum_{(k_1,\ldots,k_s)\in\mathbb{N}_0^n : k_1+\ldots+k_s=n} \frac{n!}{k_1!\cdot k_2!\cdot\ldots\cdot k_s!} \prod_{j=1}^{s} x_j^{k_j}$$

$(X_1, \ldots, X_s)$ hat eine Multinomialverteilung in folgendem Sinn:

13.1 Definition (Multinomialverteilung)

Der Zufallsvektor $(X_1, \ldots, X_s)$ besitzt eine Multinomialverteilung mit Parametern n und $p_1, \ldots, p_s$ $(s \geq 2,\ n \geq 1,\ p_1 \geq 0, \ldots, p_s \geq 0,\ p_1 + \cdots + p_s = 1)$, falls gilt:

$$\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) \;=\; \frac{n!}{k_1!\,k_2!\ \ldots\ k_s!}\, p_1^{k_1}\, p_2^{k_2}\ \ldots\ p_s^{k_s}$$

$(k_1, \ldots, k_s \in \mathbb{N}_0,\ k_1 + \ldots + k_s = n$; sonst $\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) := 0)$.

Kurz:

$$(X_1, \ldots, X_s) \sim \mathrm{Mult}(n; p_1, \ldots, p_s)$$

13.2 Beispiele (Würfelwurf, $(X_1, \ldots, X_6) \sim$ Mult$(6; 1/6, \ldots, 1/6)$)

Ein echter Würfel wird sechs mal geworfen.

a) Mit welcher Wahrscheinlichkeit tritt jede Augenzahl genau einmal auf?

b) Mit welcher W' treten (irgendwelche) zwei Augenzahlen je dreimal auf?

c) Mit welcher W' treten eine Zahl zweimal und vier Zahlen einmal auf?

Zu a): $\mathbb{P}(X_1 = 1, \ldots, X_6 = 1) \; = \; \frac{6!}{1!^6} \cdot \left(\frac{1}{6}\right)^6 \; \approx \; 0.0154.$

Zu b):

$$\binom{6}{2} \cdot \mathbb{P}(X_1 = 3, X_2 = 3, X_3 = 0, \ldots, X_6 = 0) = 15 \cdot \frac{6!}{3!^2 \, 0!^4} \cdot \left(\frac{1}{6}\right)^6 \; \approx \; 0.00643.$$

Zu c):

$$6 \cdot 5 \cdot \mathbb{P}(X_1 = 2, X_2 = 0, X_3 = 1, \ldots, X_6 = 1) = 30 \cdot \frac{6!}{2!^1 \, 0!^1 \, 1!^4} \cdot \left(\frac{1}{6}\right)^6 \; \approx \; 0.231.$$

Nach Erzeugungsweise von $(X_1, \ldots, X_s) \sim \mathrm{Mult}(n; p_1, \ldots, p_s)$ gilt $X_j \sim \mathrm{Bin}(n, p_j)$.

Wie erhält man dieses Resultat mit Marginalverteilungsbildung?

Sei o.B.d.A. $j = 1$, $k_1 \in \{0, 1, \ldots, n\}$.

$$\begin{aligned}
\mathbb{P}(X_1 = k_1) &= \sum_{k_2+\ldots+k_s=n-k_1} \mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_s = k_s) \\
&= \sum_{k_2+\ldots+k_s=n-k_1} \frac{n!}{k_1!\, k_2! \cdots k_s!}\, p_1^{k_1}\, p_2^{k_2} \ldots p_s^{k_s} \\
&= \frac{n!}{k_1!\,(n-k_1)!}\, p_1^{k_1} \sum_{k_2+\ldots+k_s=n-k_1} \frac{(n-k_1)!}{k_2! \cdots k_s!}\, p_2^{k_2} \ldots p_s^{k_s} \\
&= \binom{n}{k_1} p_1^{k_1} (p_2 + \cdots + p_s)^{n-k_1} \qquad \text{(multinomialer Lehrsatz)} \\
&= \binom{n}{k_1} p_1^{k_1} (1 - p_1)^{n-k_1}
\end{aligned}$$

Trefferarten zu paarweise disjunkten Gruppen zusammenfassen

z.B. im Fall $s = 6$ (Würfelwurf):

- $1, 3, 4 \rightarrow$ Treffer 1. Art,
- $6 \rightarrow$ Treffer 2. Art,
- $2, 5 \rightarrow$ Treffer 3. Art

Gemeinsame Verteilung von „Gruppen-Trefferanzahlen“ ist wieder eine Multinomialverteilung. Formal:

13.3 Satz (Multinomialverteilung und Vergröberung)

Sei $(X_1, \ldots, X_s) \sim$ Mult$(n; p_1, \ldots, p_s)$. Weiter sei $T_1 + \ldots + T_l = \{1, \ldots, s\}$ mit $l \geq 2$ und $T_r \neq \emptyset, r = 1, \ldots, l$. Sei

$$Y_r := \sum_{k \in T_r} X_k, \qquad q_r := \sum_{k \in T_r} p_k, \qquad r = 1, \ldots, l.$$

Dann gilt $(Y_1, \ldots, Y_l) \sim$ Mult$(n; q_1, \ldots, q_l)$.

Insbesondere gilt also $X_i + X_j \sim$ Bin$(n, p_i + p_j)$ für $i \neq j$.

13.4 Folgerung (Kovarianz und Korrelation von Mult$(n; p_1, \ldots, p_s)$)

Sei $(X_1, \ldots, X_s) \sim$ Mult$(n; p_1, \ldots, p_s)$. Dann gelten:

a) $\mathrm{C}(X_i, X_j) \; = \; -np_ip_j, \qquad i \neq j,$

b) $r(X_i, X_j) \; = \; -\sqrt{\dfrac{p_ip_j}{(1-p_i)(1-p_j)}}, \qquad i \neq j;\; p_i\,p_j > 0.$

BEWEIS: $X_i \sim$ Bin(n, p_i), $X_j \sim$ Bin(n, p_j), $X_i + X_j \sim$ Bin$(n, p_i + p_j) \implies$

$$\mathbb{V}(X_i + X_j) \;=\; n\,(p_i + p_j)\,(1 - (p_i + p_j))$$

Andererseits:

$$\begin{aligned} \mathbb{V}(X_i + X_j) &= \mathbb{V}(X_i) + \mathbb{V}(X_j) + 2\,\mathrm{C}(X_i, X_j) \\ &= np_i(1-p_i) + np_j(1-p_j) + 2\,\mathrm{C}(X_i, X_j) \end{aligned}$$

Hieraus folgt a) (und damit b)) durch direktes Ausrechnen.

- Die Multinomialverteilung ist eine Verallgemeinerung der Binomialverteilung.
- Das zugehörige Produktexperiment ist eine Verallgemeinerung der Bernoulli-Kette.

13.5 Beispiel (Vererbung)

Für Vererbung eines Merkmals sei Gen mit den Ausprägungen A (dominant) und a (rezessiv) verantwortlich.

Annahmen:

- Zwei heterozygote Aa-Eltern bringen unabhängig voneinander mit je gleicher W' $1/2$ die Keimzellen A bzw. a hervor.
- Verschmelzung der Keimzellen zu einer (diploiden) Zelle erfolgt rein zufällig

$\Longrightarrow$ Möglichkeiten AA, Aa, aA und aa haben gleiche W' $1/4$.

Die Fälle Aa und aA sind nicht unterscheidbar $\Longrightarrow$

für den Genotyp eines Nachkommen gibt es die Möglichkeiten AA (W' 1/4), Aa (W' 1/2) und aa (W' 1/4).

n-malige unabhängige Paarung zweier Aa-Eltern $\Longrightarrow$ Genotyp-Anzahlen

$$\begin{aligned} X_{AA} &= \text{Anzahl aller Nachkommen mit Genotyp } AA, \\ X_{Aa} &= \text{Anzahl aller Nachkommen mit Genotyp } Aa, \\ X_{aa} &= \text{Anzahl aller Nachkommen mit Genotyp } aa \end{aligned}$$

Es gilt $(X_{AA}, X_{Aa}, X_{aa}) \sim \text{Mult}(n; 1/4, 1/2, 1/4)$, also

$$\mathbb{P}(X_{AA} = i, X_{Aa} = j, X_{aa} = k) \; = \; \frac{n!}{i!j!k!} \left(\frac{1}{4}\right)^i \left(\frac{1}{2}\right)^j \left(\frac{1}{4}\right)^k$$

für jede Wahl von $i, j, k \geq 0$ mit $i + j + k = n$.

13.6 Mehrdimensionale hypergeometrische Verteilung

Eine Urne enthalte r_j Kugeln der Farbe j, $j = 1, \ldots, s$, $s \geq 2$.

n mal rein zufällig Ziehen ohne Zurücklegen.

Sei X_j die Anzahl der gezogenen Kugeln der Farbe j, $j = 1, \ldots, s$.

Die Verteilung von $(X_1, \ldots, X_s)$ heißt mehrdimensionale hypergeometrische Verteilung. Es gilt (Übungsaufgabe)

$$\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s) \;=\; \frac{\binom{r_1}{k_1} \cdot \ldots \cdot \binom{r_s}{k_s}}{\binom{r_1 + \ldots + r_s}{n}}$$

für $(k_1, \ldots, k_s) \in \mathbb{N}_0^s$ mit $k_1 + \ldots + k_s = n$.

Falls das Ziehen mit Zurücklegen erfolgt, so

$$(X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s),$$

wobei

$$p_1 = \frac{r_1}{r_1 + \ldots + r_s}, \ldots, p_s = \frac{r_s}{r_1 + \ldots + r_s}.$$

14 Wartezeitverteilungen

Betrachte eine Bernoulli-Kette mit Trefferwahrscheinlichkeit p, $0 < p < 1$.

Sei X die Anzahl der Nieten (0) vor dem ersten Treffer (1).

Welche Verteilung besitzt X?

$$X = k \iff \underbrace{0\,0\,0 \ldots 0}_{k}\,1$$

Die Wahrscheinlichkeit hierfür ist $(1-p)^k\,p$ (Produktexperiment!)

Möglicher Grundraum: $\Omega = \{1, 01, 001, 0001, 00001, \ldots\}$

14.1 Definition und Satz (geometrische Verteilung)

Die Zufallsvariable X hat eine geometrische Verteilung mit Parameter p, $0 < p < 1$, kurz: $X \sim \mathsf{G}(p)$, falls gilt:

$$\mathbb{P}(X = k) \; = \; (1-p)^k \, p, \qquad k \in \mathbb{N}_0.$$

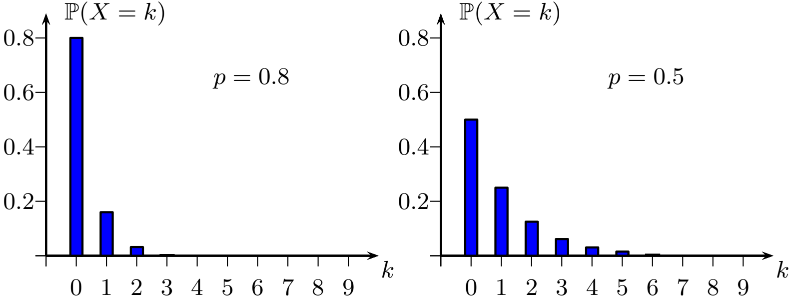
Falls $X \sim \mathsf{G}(p)$, so gelten:

a) $\mathbb{E}(X) \; = \; \dfrac{1-p}{p} \; = \; \dfrac{1}{p} \; - \; 1 \; ,$

b) $\mathbb{V}(X) \; = \; \dfrac{1-p}{p^2} \; .$

Beachte: $X+1$ modelliert die Zahl der Versuche bis zum ersten Treffer (einschließlich des letzten Versuchs, der den ersten Treffer ergibt).

Also: $p = 1/6 \implies \mathbb{E}(X+1) = 6.$



Stabdiagramme geometrischer Verteilungen

Memo: $\mathbb{P}(X = k) \; = \; (1-p)^k \, p, \qquad k \in \mathbb{N}_0.$

a) $\mathbb{E}\,X = \dfrac{1-p}{p}$

Beweis:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\,(1-p)^k\,p \\
&= p(1-p)\sum_{k=1}^{\infty} k\,(1-p)^{k-1} \\
&= p(1-p)\,\frac{\mathsf{d}}{\mathsf{d}t}\sum_{k=0}^{\infty} t^k \Big|_{t=1-p} \\
&= p(1-p)\,\frac{\mathsf{d}}{\mathsf{d}t}\,\frac{1}{1-t}\Big|_{t=1-p} \\
&= p(1-p)\,\frac{1}{(1-t)^2}\Big|_{t=1-p} \\
&= \frac{1-p}{p} \;\; \checkmark
\end{aligned}
$$

Memo: $\mathbb{P}(X = k) \; = \; (1-p)^k \, p, \qquad k \in \mathbb{N}_0.$

b) $\mathbb{V}(X) = \dfrac{1-p}{p^2}.$

Trick: Bestimme erst $\mathbb{E}X(X-1)$.

BEWEIS:

$$\begin{aligned}
\mathbb{E}\, X(X-1) \;&=\; \sum_{k=0}^{\infty} k(k-1)\,(1-p)^k\, p \\
&=\; p(1-p)^2 \sum_{k=2}^{\infty} k(k-1)\,(1-p)^{k-2} \\
&=\; p(1-p)^2 \, \frac{\mathrm{d}^2}{\mathrm{d}t^2} \sum_{k=0}^{\infty} t^k \Big|_{t=1-p} \\
&=\; p(1-p)^2 \, \frac{\mathrm{d}^2}{\mathrm{d}t^2} \, \frac{1}{1-t} \Big|_{t=1-p} \\
&=\; p(1-p)^2 \, \frac{2}{(1-t)^3} \Big|_{t=1-p} \\
&=\; p(1-p)^2 \, \frac{2}{p^3} \;=\; \frac{2(1-p)^2}{p^2}
\end{aligned}$$

$$\Longrightarrow \mathbb{V}(X) \;=\; \mathbb{E}\, X(X-1) + \mathbb{E}\, X - (\mathbb{E}\, X)^2 \;=\; \cdots \;=\; (1-p)/p^2 \;\checkmark$$

14.2 Beispiel

Ein Lottospieler gibt regelmäßig 10 verschiedene Tippreihen ab.

Pro Woche finden zwei Ausspielungen statt.

Wie groß ist der Erwartungswert (in Jahren) bis zum ersten „Sechser“?

Modellierung der Folge der Ausspielungen als Bernoulli-Kette

mit Trefferwahrscheinlichkeit $p = 10/\binom{49}{6}$.

Sei X die Anzahl der Ausspielungen bis zum ersten Sechser.

Es gilt $X - 1 \sim \mathsf{G}(p)$.

$$\mathbb{E}\,X \;=\; \frac{1}{p} \;=\; 1\ 398\ 381{,}6 \quad \text{(Ausspielungen)}$$

Ca. 104 Ausspielungen pro Jahr $\Longrightarrow$

$$\mathbb{E}\left(\frac{X}{104}\right) \;=\; \frac{1\ 398\ 381{,}6}{104} \;\approx\; 13\ 446 \quad \text{(Jahre)}$$

14.3 Satz (Gedächtnislosigkeit der geometrischen Verteilung)

a) Sei $X \sim$ G(p). Dann gilt

$$\mathbb{P}(X = m + k | X \geq k) \;=\; \mathbb{P}(X = m) \quad \text{für jede Wahl von } k, m \in \mathbb{N}_0 \; (\star)$$

b) Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable mit $(\star)$ und $0 < \mathbb{P}(X = 0) < 1$.

Dann gilt $X \sim$ G(p) für ein $p \in (0, 1)$.

Beweis: a) Übungsaufgabe!

b) Setze $k := 1$ in $(\star)$ $\Longrightarrow$

$$\mathbb{P}(X = m) = \mathbb{P}(X = m{+}1 | X \geq 1) = \frac{\mathbb{P}(X = m + 1, X \geq 1)}{\mathbb{P}(X \geq 1)} = \frac{\mathbb{P}(X = m + 1)}{\mathbb{P}(X \geq 1)}$$

für jedes $m \geq 0$. Sei $p_m := \mathbb{P}(X = m)$, $m \geq 0$ $\Longrightarrow$

$$\frac{p_{m+1}}{p_m} = 1 - p_0, \;\; m \in \mathbb{N}_0$$

$$\Longrightarrow \;\; \mathbb{P}(X = j) \;=\; p_j \;=\; \left(\prod_{m=0}^{j-1} \frac{p_{m+1}}{p_m} \right) p_0 \;=\; (1 - p_0)^j \, p_0, \quad j \in \mathbb{N}_0. \; \surd$$

Bernoulli-Kette: Verallgemeinerung der Fragestellung:

Sei $r \geq 1$ fest.

Welche Verteilung besitzt die Anzahl X der Nieten vor dem r-ten Treffer?

$$X = k \iff \underbrace{\star \star \star \star \star \star \star \star \star \star \star \star \star \star \star \cdots \star}_{\text{Hier } k \text{ Nullen und } r-1 \text{ Einsen}} 1$$

W' für eine konkrete Folge aus k Nullen, $r-1$ Einsen und einer Eins am Ende:

$$(1-p)^k \cdot p^{r-1} \cdot p.$$

Anzahl der Auswahlen von k Nullen aus $k+r-1$ Plätzen: $\binom{k+r-1}{k}$

14.4 Definition und Satz (negative Binomialverteilung)

Die Zufallsvariable X besitzt eine negative Binomialverteilung mit Parametern r und p $(r \in \mathbb{N},\ 0 < p < 1)$, kurz: $X \sim \mathrm{Nb}(r,p)$, falls gilt:

$$\mathbb{P}(X = k) \ = \ \binom{k+r-1}{k}\, p^r\, (1-p)^k, \qquad k \in \mathbb{N}_0.$$

Falls $X \sim \mathrm{Nb}(r,p)$, so gelten:

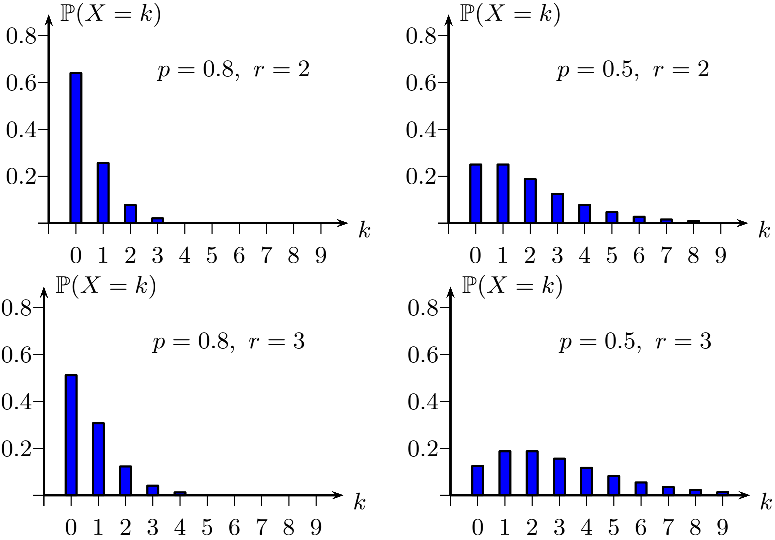
a) $\mathbb{E}(X) \ = \ r \cdot \dfrac{1-p}{p}$,

b) $\mathbb{V}(X) \ = \ r \cdot \dfrac{1-p}{p^2}$.

Beachte: (Warum negative Binomialverteilung?)

$$\binom{k+r-1}{k} = \frac{(k+r-1)!}{k!(r-1)!} = \frac{(k+r-1)(k+r-2)\ldots(r+1)r}{k!}$$

$$= \frac{(-1)^k}{k!} \cdot (-r)(-r-1)\ldots(-r-(k-2))(-r-(k-1)) = (-1)^k \binom{-r}{k}$$



Stabdiagramme von negativen Binomialverteilungen

Memo: $\binom{k+r-1}{k} = (-1)^k \binom{-r}{k}$

Memo: $(1+x)^\alpha \; = \; \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \qquad \alpha \in \mathbb{R}, \; |x| < 1$ (Binomialreihe)

Nachweis von $\sum_{k=0}^{\infty} \mathbb{P}(X=k) \; = \; 1$:

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbb{P}(X=k) \quad &= \quad \sum_{k=0}^{\infty} \binom{k+r-1}{k} (1-p)^k \, p^r \\
&= \quad \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} (1-p)^k p^r \\
&= \quad p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-(1-p))^k \\
&= \quad p^r (1 + (-(1-p))^{-r} \quad = \quad 1 \quad \surd
\end{aligned}$$

Memo: $(1+x)^\alpha \; = \; \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \qquad \alpha \in \mathbb{R}, \; |x| < 1$ (Binomialreihe)

$$(1+x)^{-r} \; = \; \sum_{m=0}^{\infty} \binom{-r}{m} x^m, \qquad (1+x)^{-s} \; = \; \sum_{n=0}^{\infty} \binom{-s}{n} x^n$$

$$(1+x)^{-(r+s)} \; = \; \sum_{k=0}^{\infty} \binom{-(r+s)}{k} x^k$$

$$\begin{aligned}
(1+x)^{-(r+s)} \; &= \; (1+x)^{-r} \cdot (1+x)^{-s} \\
&= \; \sum_{m,n=0}^{\infty} \binom{-r}{m} \binom{-s}{n} \, x^{m+n} \\
&= \; \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} \binom{-r}{j} \binom{-s}{k-j} \right) x^k \qquad \text{(Cauchy-Produkt)}
\end{aligned}$$

$$\Longrightarrow \; \binom{-(r+s)}{k} \; = \; \sum_{j=0}^{k} \binom{-r}{j} \binom{-s}{k-j}$$

Offenbar gilt G(p) $=$ Nb$(1,p)$.

Besteht ein Zusammenhang zwischen Nb(r,p) und G(p) für $r \geq 2$?

$$\underbrace{0\ \ 0}_{X_1}\ \ 1\ \ \underbrace{0\ \ 0\ \ 0\ \ 0}_{X_2}\ \ 1\ \ \underbrace{0\ \ 0}_{X_3}\ \ 1\ \ \underbrace{0\ \ 0\ \ 0\ \ 0\ \ 0}_{X_4}\ \ 1\ \ \underbrace{0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0}_{X_5}\ \ 1$$

Vermutung: Ist $X \sim$ Nb(r,p), so sollte $X \sim X_1 + \ldots + X_r$ gelten.

Dabei: $X_1, \ldots, X_r$ stochastisch unabhängig und je G(p)-verteilt.

14.5 Satz (Struktur- und Additionsgesetz für Nb(r,p))

a) Sind $X_1, \ldots, X_r$ unabhängig und $X_j \sim$ G(p) $\forall j$, so gilt $\displaystyle\sum_{j=1}^{r} X_j \sim$ Nb(r,p).

b) X, Y unabhängig, $X \sim$ Nb(r,p), $Y \sim$ Nb(s,p)

$\Longrightarrow$ $X + Y \sim$ Nb$(r+s,p)$.

Beachte: b) $\Longrightarrow$ a).

Memo: $\binom{-(r+s)}{k} = \sum_{j=0}^{k} \binom{-r}{j}\binom{-s}{k-j}, \quad \binom{k+r-1}{k} = (-1)^k \binom{-r}{k}$

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{j=0}^{k} \mathbb{P}(X=j)\,\mathbb{P}(Y=k-j) \qquad \text{(Faltungsformel)} \\
&= \sum_{j=0}^{k} \binom{j+r-1}{j}(1-p)^j p^r \binom{k-j+s-1}{k-j}(1-p)^{k-j}p^s \\
&= (1-p)^k p^{r+s} \sum_{j=0}^{k} \binom{j+r-1}{j}\binom{k-j+s-1}{k-j} \\
&= (1-p)^k p^{r+s} \sum_{j=0}^{k} (-1)^j \binom{-r}{j} (-1)^{k-j} \binom{-s}{k-j} \\
&= (1-p)^k p^{r+s} (-1)^k \sum_{j=0}^{k} \binom{-r}{j}\binom{-s}{k-j} \\
&= (1-p)^k p^{r+s} (-1)^k \binom{-(r+s)}{k} = \binom{k+r+s-1}{k}(1-p)^k p^{r+s} \checkmark
\end{aligned}$$

Mit dem Additionsgesetz erhält man unmittelbar:

$$X \sim \mathsf{Nb}(r,p) \implies \mathbb{E}(X) \; = \; r \cdot \frac{1-p}{p}, \quad \mathbb{V}(X) \; = \; r \cdot \frac{1-p}{p^2}.$$

Begründung:

Seien $X_1, \ldots, X_r$ unabhängig und je $\mathsf{G}(p)$-verteilt. Es gilt

$$\mathbb{E}(X_j) \; = \; \frac{1-p}{p}, \qquad \mathbb{V}(X_j) = \frac{1-p}{p^2}, \qquad j = 1, \ldots, r.$$

Wegen $X \sim X_1 + \ldots + X_r$ folgt

$$\begin{aligned} \mathbb{E}(X) \; &= \; \mathbb{E}(X_1 + \ldots + X_r) \; = \; \sum_{j=1}^{r} \mathbb{E}(X_j) \; = \; r \cdot \frac{1-p}{p}, \\ \mathbb{V}(X) \; &= \; \mathbb{V}(X_1 + \ldots + X_r) \; = \; \sum_{j=1}^{r} \mathbb{V}(X_j) \; = \; r \cdot \frac{1-p}{p^2}. \end{aligned}$$

Selbsttest:

Ein echter Würfel wird in unabhängiger Folge geworfen.

Mit welcher W' tritt die dritte Sechs im 16. Wurf auf?

Antwort:

$$\begin{aligned}
\text{W'} &= \binom{15}{2} \cdot \left(\frac{5}{6}\right)^{13} \cdot \left(\frac{1}{6}\right)^{2} \cdot \frac{1}{6} \\
&= \binom{13+3-1}{13} (1-p)^{13} p^{3}, \quad p = \frac{1}{6} \\
&\approx 0.054
\end{aligned}$$

14.6 Sammelbilder-Probleme (Coupon-Collector-Problem)

n leere Fächer. Rein zufällig s Fächer auswählen, in jedes je ein Teilchen legen. Diesen Besetzungsvorgang („s-Auswahl“) in unabhängiger Folge wiederholen.

Sei X_n die Anzahl der Besetzungsvorgänge, bis jedes Fach belegt ist.

Welche Verteilung besitzt X_n?

$n = 6$, $s = 1$: Wie oft muss man würfeln, bis jede Augenzahl aufgetreten ist?

$n = 49$, $s = 6$: Wie viele Lottoausspielungen sind nötig, bis jede Zahl Gewinnzahl ist?

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42
43	44	45	46	47	48	49

$n = 640$, $s = 5$: Sammelalbum zur Fußball-WM 2014
Überraschungseier, Fußball-Bundesliga, usw.

Memo: $X_n =$ Anzahl der Besetzungsvorgänge, bis jedes Fach belegt ist

Mindestanzahl an Besetzungsvorgängen bis zu einer „vollständigen Serie“:

$$a \ := \ \min\left\{m \in \mathbb{N} : \frac{n}{s} \le m\right\} \ = \ \left\lceil \frac{n}{s} \right\rceil$$

Zufallsvariable X_n nimmt Werte $k \in \{a, a+1, a+2, \ldots\}$ an.

Sei für jedes $j \in \{1, 2, \ldots, n\}$

$W_j :=$ Anz. der Besetzungsvorgänge, bis Fach j mindestens ein Teilchen enthält

Entscheidende Erkenntnis:

$$\{X_n > k\} \ = \ \bigcup_{j=1}^{n} \{W_j > k\}, \quad k \ge a-1.$$

$\mathbb{P}(X_n > k)$ mit Formel des Ein- und Ausschließens bestimmen ?!

$\mathbb{P}(X_n = k) \ = \ \mathbb{P}(X_n > k-1) - \mathbb{P}(X_n > k)$

Memo: $\{X_n > k\} \;=\; \bigcup_{j=1}^{n} \{W_j > k\}, \quad k \geq a-1.$

Sei k fest, $A_j := \{W_j > k\}$. Formel des Ein- und Ausschließens $\Longrightarrow$

$$\mathbb{P}(X_n > k) \;=\; \sum_{r=1}^{n} (-1)^{r-1} S_r, \quad S_r = \sum_{1 \leq i_1 < \ldots < i_r \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r})$$

$A_{i_1} \cap \ldots \cap A_{i_r} \quad \Longleftrightarrow \quad$ bei jedem der ersten k Besetzungsvorgänge gelangt kein Teilchen in die Fächer mit den Nummern $i_1, \ldots, i_r$ (nur möglich bei $r \leq n-s$)

$q_r \;:=\; \frac{\binom{n-r}{s}}{\binom{n}{s}}$ (W', dass bei *einer* s-Auswahl r bestimmte Fächer frei bleiben)

Unabhängigkeit der k Besetzungsvorgänge $\Longrightarrow$

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r}) \;=\; \mathbb{P}(A_1 \cap \ldots \cap A_r) \;=\; q_r^k \Longrightarrow$$

$$\mathbb{P}(X_n > k) \;=\; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} q_r^k$$

Memo: $a = \left\lceil \frac{n}{s} \right\rceil \; = \; \min\left\{m \in \mathbb{N} : \frac{n}{s} \leq m\right\}, \quad q_r = \binom{n-r}{s} \Big/ \binom{n}{s}$

14.7 Satz (Verteilung der Anzahl der Besetzungsvorgänge)

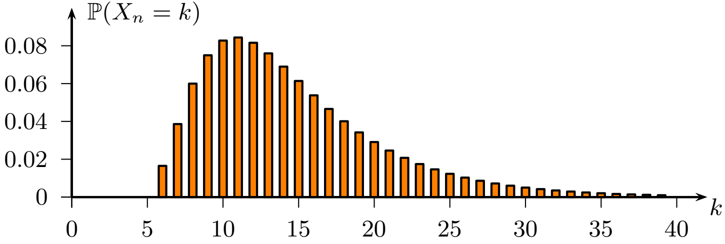
Für die Anzahl X_n der Besetzungsvorgänge bis zur Besetzung aller Fächer im Sammelbilder-Problem mit n Fächern und s-Auswahl gilt:

a) $$\mathbb{P}(X_n = k) \; = \; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} q_r^{k-1} (1-q_r), \qquad k \geq a,$$

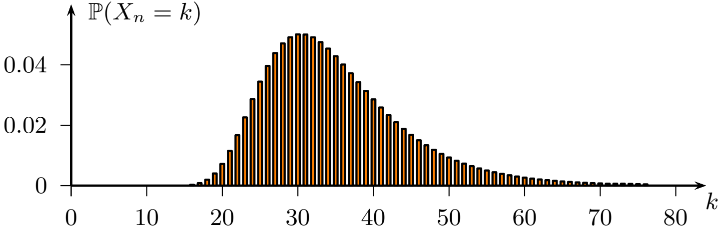
b) $$\mathbb{E}\, X_n \; = \; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} \frac{q_r^{a-1}\,(q_r + a\,(1-q_r))}{1-q_r}.$$

Beweis von b)

$$\begin{aligned} \mathbb{E}\, X_n \quad &= \quad \sum_{k=a}^{\infty} k\, \mathbb{P}(X_n = k) = \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} (1-q_r) \sum_{k=a}^{\infty} k q_r^{k-1} \\ &= \quad \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} (1-q_r)\, \frac{\mathrm{d}}{\mathrm{d}x} \frac{x^a}{1-x}\Big|_{x=q_r} \quad \Longrightarrow \text{ Beh.} \end{aligned}$$



Verteilung der Wartezeit beim Sammelbilder-Problem mit $n = 6, \ s = 1$



Verteilung der Wartezeit beim Sammelbilder-Problem mit $n = 49, \ s = 6$

14.8 Satz (Grenzverteilung für die Zahl der Besetzungsvorgänge)

Im Fall $s = 1$ gilt für jedes $x \in \mathbb{R}$:

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \le n \cdot (x + \log n)\right) \; = \; \exp\left(-\mathrm{e}^{-x}\right)$$

Beweis: Seien $x \in \mathbb{R}$ beliebig, n so groß, dass $x+\log n \ge 1$; $k_n := \lfloor n(x+\log n) \rfloor$.

$$\mathbb{P}(X_n > n(x + \log n)) \; = \; \mathbb{P}(X_n > k_n) \; = \; \sum_{r=1}^{n-1} (-1)^{r-1} \binom{n}{r} q_r^{k_n}$$

Bonferroni-Ungleichungen $\Longrightarrow$

$$\mathbb{P}(X_n > n(x + \log n)) \quad \le \quad \sum_{r=1}^{2u+1} (-1)^{r-1} \binom{n}{r} q_r^{k_n}, \qquad u \ge 0 \text{ fest},$$

$$\mathbb{P}(X_n > n(x + \log n)) \quad \ge \quad \sum_{r=1}^{2u} (-1)^{r-1} \binom{n}{r} q_r^{k_n}, \qquad u \ge 1 \text{ fest}.$$

Zeige: $\displaystyle\lim_{n\to\infty} \binom{n}{r} q_r^{k_n} \; = \; \frac{\mathrm{e}^{-xr}}{r!} \; \Longrightarrow$

$$\limsup_{n\to\infty} \mathbb{P}(X_n > n(x+\log n)) \quad \le \quad \sum_{r=1}^{2u+1} (-1)^{r-1} \frac{\mathsf{e}^{-xr}}{r!},$$

$$\liminf_{n\to\infty} \mathbb{P}(X_n > n(x+\log n)) \quad \ge \quad \sum_{r=1}^{2u} (-1)^{r-1} \frac{\mathsf{e}^{-xr}}{r!}.$$

Jetzt $u \to \infty$:

$$\begin{aligned} \sum_{r=1}^{\infty} (-1)^{r-1} \frac{\mathsf{e}^{-xr}}{r!} &= -\sum_{r=1}^{\infty} \frac{(-\mathsf{e}^{-x})^r}{r!} \\ &= -\left(\exp(-\mathsf{e}^{-x}) - 1\right) \\ &= 1 - \exp(-\mathsf{e}^{-x}) \end{aligned}$$

$$\Longrightarrow \quad \lim_{n\to\infty} \mathbb{P}(X_n \le n(x+\log n))) \;=\; \exp(-\mathsf{e}^{-x}) \;\surd$$

Memo: z.z.: $\lim\limits_{n\to\infty}\binom{n}{r} q_r^{k_n} \;=\; \dfrac{\mathrm{e}^{-xr}}{r!}$

$$q_n \;=\; \frac{n-r}{n} \;=\; 1-\frac{r}{n}, \quad k_n = \lfloor n(x+\log n)\rfloor, \quad n^{\underline{r}} = n(n-1)\ldots(n-r+1)$$

Sei $\varepsilon_n \;:=\; \lfloor n(x+\log n)\rfloor - n(x+\log n) \;\Longrightarrow\; -1 \le \varepsilon_n \le 0.$

$$\begin{aligned}
\binom{n}{r} q_r^{k_n} \;&=\; \frac{1}{r!}\cdot\frac{n^{\underline{r}}}{n^r}\cdot n^r\cdot\left(1-\frac{r}{n}\right)^{k_n}\\
&=\; \frac{1}{r!}\cdot\frac{n^{\underline{r}}}{n^r}\cdot n^r\cdot\left(1-\frac{r}{n}\right)^{n(x+\log n)}\cdot\left(1-\frac{r}{n}\right)^{\varepsilon_n}\\
&=\; \frac{1}{r!}\cdot \mathrm{e}^{r\log n}\cdot\exp\left(n(x+\log n)\log\left(1-\frac{r}{n}\right)\right)\cdot\frac{n^{\underline{r}}}{n^r}\cdot\left(1-\frac{r}{n}\right)^{\varepsilon_n}\\
&=\; \frac{1}{r!}\cdot\exp\left[r\log n + n(x+\log n)\log\left(1-\frac{r}{n}\right)\right]\cdot\frac{n^{\underline{r}}}{n^r}\cdot\left(1-\frac{r}{n}\right)^{\varepsilon_n}
\end{aligned}$$

$\Longrightarrow$ zu zeigen: $r\log n + n(x+\log n)\log\left(1-\dfrac{r}{n}\right) \to -rx$

Memo: Sei $a_n := r \log n + n(x + \log n) \log \left(1 - \frac{r}{n}\right)$

Memo: Zu zeigen: $\lim\limits_{n \to \infty} a_n \; = \; -rx$

Beachte: $\log t \; \leq \; t - 1 \quad \Longrightarrow - \log t \; = \; \log \frac{1}{t} \; \leq \; \frac{1}{t} - 1 \; \Longrightarrow \log t \; \geq \; 1 - \frac{1}{t}$

$$\Longrightarrow \; -\frac{r}{n-r} \; \leq \log \left(1 - \frac{r}{n}\right) \; \leq \; -\frac{r}{n}$$

Hiermit folgt

$$\begin{aligned}
a_n \quad & \leq \quad r \log n - \frac{r}{n} \cdot n(x + \log n) \; = \; r \log n - rx - r \log n \; = \; -rx \\
a_n \quad & \geq \quad r \log n - \frac{r}{n-r} \cdot n(x + \log n) \\
& = \quad r \log n - \frac{n}{n-r} rx - \frac{rn}{n-r} \log n \\
& = \quad -\frac{n}{n-r} rx + r \log n \left(1 - \frac{n}{n-r}\right) \; = \; -\frac{n}{n-r} rx - r^2 \frac{\log n}{n-r} \\
& \to \quad -rx \; \surd
\end{aligned}$$

Konsequenzen: $\mathbb{P}(X_n \;\le\; n(x + \log n)) \quad \approx \quad \exp(-\mathrm{e}^{-x})$ für großes n

$$\mathbb{P}\left(\frac{X_n}{n} \;\le\; x + \log n\right) \quad \approx \quad \exp(-\mathrm{e}^{-x}) \qquad \text{für großes } n$$

$\dfrac{X_n}{n}$ ist die Anzahl der Teilchen pro Fach bis zur vollständigen Serie

Zahl der Teilchen pro Fach bis zur vollständigen Serie wächst logarithmisch!

Wähle p mit $0 < p < 1$. Setze $p = \exp(-\mathrm{e}^{-x}) \implies$

$$\log p = -\mathrm{e}^{-x} \quad \implies \log(-\log p) = -x \implies \quad x_p = -\log(-\log p))$$

$$x_{0.5} \approx 0.367, \quad x_{0.9} \approx 2.25, \quad x_{0.95} \approx 2.97$$

$$n = 640 \implies \log 640 \approx 6.461 \implies x_{0.5} + \log n \approx 6.828$$

$$\begin{aligned} n(x_{0.5} + \log n) &\approx 4370 \implies \mathbb{P}(X_{640} \le 4370) \approx 0.5 \\ n(x_{0.9} + \log n) &\approx 5575 \implies \mathbb{P}(X_{640} \le 5575) \approx 0.9 \\ n(x_{0.95} + \log n) &\approx 6036 \implies \mathbb{P}(X_{640} \le 6036) \approx 0.95 \end{aligned}$$

Im Fall $s = 1$ kann X_n als Summe unabhängiger Wartezeiten modelliert werden.

Sind $j < n$ Fächer belegt, so „Treffer“, wenn das nächste Teilchen in eines der $n - j$ noch freien Fächer fällt.

Die Wahrscheinlichkeit hierfür ist $p_j = \dfrac{n-j}{n} = 1 - \dfrac{j}{n}$.

Folgerung: $X_n \;\sim\; 1 + Y_1 + Y_2 + \ldots + Y_{n-1}$.

Dabei: $Y_1, \ldots, Y_{n-1}$ unabhängig und $Y_j - 1 \sim \mathsf{G}(p_j)$, $j = 1, \ldots, n-1$.

$$\mathbb{E}\, Y_j \;=\; \frac{1}{p_j} \;=\; \frac{n}{n-j}, \qquad \mathbb{V}(Y_j) \;=\; \frac{1-p_j}{p_j^2} \;=\; \frac{j\,n}{(n-j)^2}$$

$$\mathbb{E}\, X_n \;=\; \sum_{j=0}^{n-1} \frac{n}{n-j} \;=\; n \cdot \sum_{k=1}^{n} \frac{1}{k} \;\approx n \;\cdot (\log n + 0.57721)$$

$$\mathbb{V}(X_n) \;=\; \sum_{j=1}^{n-1} \frac{j\,n}{(n-j)^2} \;=\; n \cdot \sum_{k=1}^{n-1} \frac{n-k}{k^2} \;=\; n^2 \sum_{k=1}^{n-1} \frac{1}{k^2} - n \sum_{k=1}^{n-1} \frac{1}{k}$$

$$\approx \;\; n^2 \cdot \left(\frac{\pi^2}{6} - \frac{\log(n-1) + 0.57721}{n} \right) \;\sim\; \frac{\pi^2}{6} \cdot n^2$$

Verständnisfragen:

Ein echter Würfel wird geworfen, bis die erste Sechs auftritt.

- Wie wahrscheinlich ist es, vorher mindestens eine Drei zu werfen?
- Wie wahrscheinlich ist es, vorher genau zwei Dreien zu werfen?
- Welche Verteilung besitzt die Anzahl der vor der ersten Sechs geworfenen Dreien?

15 Die Poisson-Verteilung

- eine grundlegende diskrete Verteilung,
- ungerechtfertigterweise nach Poisson benannt,
- entsteht u.a. als Approximation der Binomialvertelung Bin(n, p) bei großem n und kleinem p
- tritt approximativ bei Zählvorgängen mit seltenen Ereignissen auf (Unfälle, starke Erdbeben, Gewitter, Selbstmorde ...)

Betrachte Folge von Binomialverteilungen $(\text{Bin}(n, p_n))_{n\geq 1}$ mit

$$n \cdot p_n \ = \ \lambda, \quad n \geq 1, \quad \text{wobei } \lambda \in (0, \infty).$$

Die erwartete Trefferzahl np_n bleibt also konstant.

Sei $k \in \mathbb{N}_0$ fest. Für $n \in \mathbb{N}$ mit $n \geq k$ gilt

$$\begin{aligned}
\binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} \quad &= \quad \frac{(n \cdot p_n)^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^{-k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^{n} \\
&= \quad \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n^{\underline{k}}}{n^k}}_{\substack{\downarrow \\ 1}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\substack{\downarrow \\ 1}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n}}_{\substack{\downarrow \\ \mathrm{e}^{-\lambda}}}
\end{aligned}$$

$$\to \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda} \quad \text{für} \quad n \to \infty.$$

15.1 Satz (Gesetz seltener Ereignisse)

Sei (p_n) eine Folge in $[0,1]$ mit $\lim_{n\to\infty} np_n = \lambda$, wobei $0 < \lambda < \infty$. Dann gilt

$$\lim_{n\to\infty} \binom{n}{k} p_n^k \, (1-p_n)^{n-k} \;=\; \mathrm{e}^{-\lambda}\, \frac{\lambda^k}{k!}\,, \qquad k = 0,\, 1,\, 2, \ldots$$

15.2 Definition (Poisson-Verteilung)

Die Zufallsvariable X besitzt eine Poisson-Verteilung *mit Parameter* λ, $\lambda > 0$, kurz: $X \sim \mathrm{Po}(\lambda)$, falls gilt:

$$\mathbb{P}(X = k) \; = \; \mathrm{e}^{-\lambda} \, \frac{\lambda^k}{k!} \, , \qquad k = 0,\, 1,\, 2, \ldots$$

15.3 Satz (Eigenschaften der Poisson-Verteilung)

a) Falls $X \sim \mathrm{Po}(\lambda)$, so gilt $\mathbb{E}(X) \; = \; \mathbb{V}(X) \; = \; \lambda$.

b) Sind X, Y unabhängig, $X \sim \mathrm{Po}(\lambda)$, $Y \sim \mathrm{Po}(\mu)$, so gilt

$$X + Y \; \sim \; \mathrm{Po}(\lambda + \mu) \qquad \text{(Additionsgesetz)}$$

BEWEIS: a)

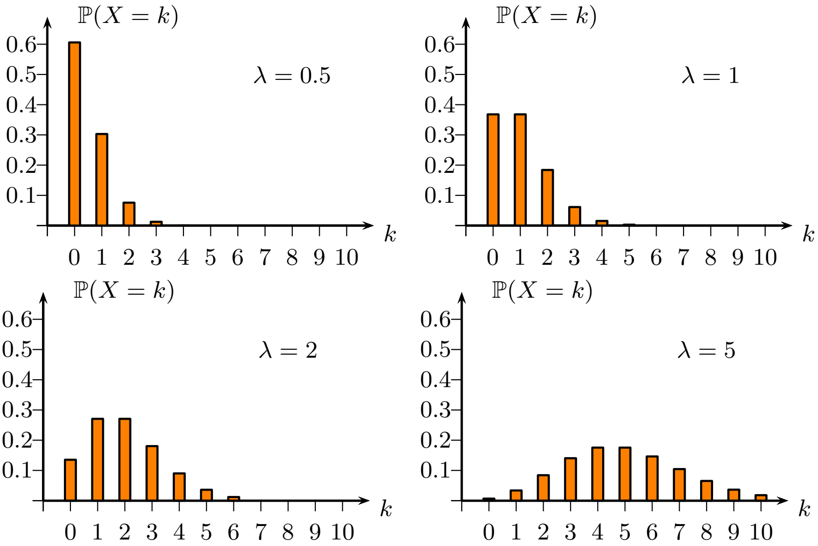
$$\mathbb{E}\, X \; = \; \sum_{k=0}^{\infty} k\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \; = \; \mathrm{e}^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \; = \; \lambda\, \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda} \; = \; \lambda$$

$$\mathbb{E}\, X(X-1) \; = \; \sum_{k=0}^{\infty} k(k-1)\mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \; = \; \lambda^2 \mathrm{e}^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \; = \; \lambda^2 \; \checkmark$$

b) z.z.: X, Y unabhängig, $X \sim \mathsf{Po}(\lambda), Y \sim \mathsf{Po}(\mu) \Longrightarrow X + Y \sim \mathsf{Po}(\lambda + \mu)$

Beweis: Sei $k \in \mathbb{N}_0$ beliebig. Faltungsformel $\Longrightarrow$

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{j=0}^{k} \mathbb{P}(X=j)\,\mathbb{P}(Y=k-j) \\
&= \sum_{j=0}^{k} \mathrm{e}^{-\lambda}\,\frac{\lambda^j}{j!}\,\mathrm{e}^{-\mu}\,\frac{\mu^{k-j}}{(k-j)!} \\
&= \mathrm{e}^{-(\lambda+\mu)}\,\frac{(\lambda+\mu)^k}{k!}\sum_{j=0}^{k}\binom{k}{j}\left(\frac{\lambda}{\lambda+\mu}\right)^{j}\left(\frac{\mu}{\lambda+\mu}\right)^{k-j} \\
&= \mathrm{e}^{-(\lambda+\mu)}\,\frac{(\lambda+\mu)^k}{k!}\cdot 1 \;\checkmark
\end{aligned}$$



Stabdiagramme von Poisson-Verteilungen

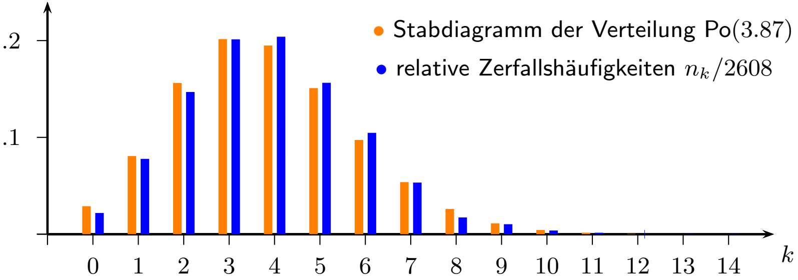
15.4 Beispiel (Das Rutherford-Geiger-Experiment)

Experiment: Radioaktives Präparat über 2608 Zeitintervalle von je 7.5 Sekunden Länge untersuchen. Insgesamt 10097 Zerfälle gezählt, also im Durchschnitt 3.87 Zerfälle innerhalb von 7.5 Sekunden.

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

Werte zum Rutherford-Geiger-Versuch

Dabei: $n_k =$ Anzahl der Zeitintervalle, in denen k Zerfälle beobachtet wurden.



Erklärungsversuch: O.B.d.A. Untersuchungszeitraum $I := (0, 1]$.

Sei X zufällige Anzahl der Zerfälle in I, $\quad \lambda := \mathbb{E}\,X$.

Zerlege I: $\quad I = \sum_{j=1}^{n} I_j$, $I_j := \left(\frac{j-1}{n}, \frac{j}{n}\right]$.

Sei $X_{n,j} =$ Anzahl der Zerfälle in I_j, also $X = X_{n,1} + X_{n,2} + \ldots + X_{n,n}$.

Ann.: $X_{n,1}, \ldots, X_{n,n}$ unabhängig und identisch verteilt $\Longrightarrow \mathbb{E}(X_{n,j}) = \lambda/n$.

Regularitätsbedingung: $\quad \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{j=1}^{n} \{X_{n,j} \geq 2\}\right) \;=\; 0.$

Idee: $X_{n,j} \approx \mathbf{1}\{X_{n,j} \geq 1\} \;\Longrightarrow\; X \approx S_n := \sum_{j=1}^{n} \; \mathbf{1}\{X_{n,j} \geq 1\}$.

$$\sim \; \mathrm{Bin}(n, p_n),\; p_n := \mathbb{P}(X_{n,1} \geq 1)$$

$p_n = \mathbb{E}\mathbf{1}\{X_{n,1} \geq 1\} \leq \mathbb{E}\,X_{n,1} = \frac{\lambda}{n}$.

Fordern wir noch $\lim_{n\to\infty} np_n = \lambda$, so folgt nach dem Gesetz seltener Ereignisse

$$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}$$

Memo: $X = \sum_{j=1}^{n} X_{n,j}, \quad S_n = \sum_{j=1}^{n} \mathbf{1}\{X_{n,j} \ge 1\}, \quad \lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$

Beachte:

$$\begin{aligned} \mathbb{P}(X = k) &= \mathbb{P}(X = k, X = S_n) + \mathbb{P}(X = k, X \ne S_n) \\ &= \mathbb{P}(S_n = k, X = S_n) + \mathbb{P}(X = k, X \ne S_n) \\ &= \mathbb{P}(S_n = k) - \mathbb{P}(S_n = k, X \ne S_n) + \mathbb{P}(X = k, X \ne S_n) . \end{aligned}$$

$$\{X \ne S_n\} \subseteq \underbrace{\cup_{j=1}^{n} \{X_{n,j} \ge 2\}}_{\mathbb{P}(\cdot) \to 0} \text{ (!)}$$

Regularitätsbedingung $\Longrightarrow \lim_{n\to\infty} \mathbb{P}(X \ne S_n) = 0 \Longrightarrow$

$\lim_{n\to\infty} \mathbb{P}(S_n = k, X \ne S_n) = 0 = \lim_{n\to\infty} \mathbb{P}(X = k, X \ne S_n) \Longrightarrow$

$$\mathbb{P}(X = k) = \lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} .$$

16 Bedingte Erwartungswerte und bedingte Verteilungen

Soll sie stoppen? Eine Sechs wirft sie auf 0 Punkte zurück.

16.1 Definition (bedingter Erwartungswert)

Es seien

- $(\Omega, \mathbb{P})$ ein diskreter W-Raum,
- $\mathbb{P}(\Omega_0) = 1$ für eine abzählbare Teilmenge Ω_0 von Ω,
- $A \subseteq \Omega$ ein Ereignis mit $\mathbb{P}(A) > 0$,
- $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}\,|X| < \infty$.

Dann heißt

$$\mathbb{E}(X|A) \ := \ \frac{1}{\mathbb{P}(A)} \sum_{\omega \in A \cap \Omega_0} X(\omega)\, \mathbb{P}(\{\omega\})$$

bedingter Erwartungswert von X unter der Bedingung A
bzw. unter der Hypothese A.

Beachte: $\mathbb{E}(X|\Omega) \ = \ \mathbb{E}(X) \ = \ \mathbb{E}(X|\Omega_0)$.

Memo: $\mathbb{E}(X|A) \;=\; \dfrac{1}{\mathbb{P}(A)} \displaystyle\sum_{\omega \in A \cap \Omega_0} X(\omega)\,\mathbb{P}(\{\omega\})$

Seien

- $Z : \Omega \to \mathbb{R}^k$ ein Zufallsvektor,
- $A \;=\; \{Z = z\} \;=\; \{\omega \in \Omega : Z(\omega) = z\},\; z \in \mathbb{R}^k$.

Dann heißt

$$\mathbb{E}(X|Z = z) \;:=\; \mathbb{E}(X|\{Z = z\})$$

bedingter Erwartungswert von X unter der Bedingung $Z = z$.

Falls $Z =: (Z_1, \ldots, Z_k)$, so auch

$$\mathbb{E}(X|Z_1 = z_1, \ldots, Z_k = z_k) \;:=\; \mathbb{E}(X|Z = z), \qquad z = (z_1, \ldots, z_k).$$

Memo: $\mathbb{E}(X|A) \; = \; \dfrac{1}{\mathbb{P}(A)} \displaystyle\sum_{\omega \in A \cap \Omega_0} X(\omega)\, \mathbb{P}(\{\omega\})$

Memo: $\mathbb{P}_A(B) \; = \; \mathbb{P}(B|A) \; = \; \dfrac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}, \quad B \subseteq \Omega.$

Memo: $\mathbb{P}_A$ ist die bedingte Verteilung von $\mathbb{P}$ unter A.

$$\begin{aligned}
\mathbb{E}(X|A) \quad &= \quad \sum_{\omega \in A \cap \Omega_0} X(\omega)\, \frac{\mathbb{P}(\{\omega\} \cap A)}{\mathbb{P}(A)} \\
&= \quad \sum_{\omega \in A \cap \Omega_0} X(\omega)\, \mathbb{P}_A(\{\omega\}) \\
&= \quad \sum_{\omega \in \Omega_0} X(\omega)\, \mathbb{P}_A(\{\omega\}) \qquad (\mathbb{P}_A(\{\omega\}) = 0, \text{ falls } \omega \notin A) \\
&= \quad \mathbb{E}_{\mathbb{P}_A}(X)
\end{aligned}$$

$\mathbb{E}(X|A)$ ist Erwartungswert von X bezüglich der bedingten Verteilung $\mathbb{P}_A$.

Memo: $\mathbb{E}(X|A) \;=\; \dfrac{1}{\mathbb{P}(A)} \displaystyle\sum_{\omega\in A\cap\Omega_0} X(\omega)\,\mathbb{P}(\{\omega\})$

16.2 Beispiel

Zweifacher Würfelwurf, Laplace-Modell. X_j sei Augenzahl des j-ten Wurfs.

$\mathbb{E}(X_1|X_1+X_2\geq 9) =?$

$\Omega = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$

$$\begin{aligned} A &:= \{X_1+X_2\geq 9\} \\ &= \{(3,6),(4,5),(5,4),(6,3),(4,6),(5,5),(6,4),(5,6),(6,5),(6,6)\} \end{aligned}$$

$\mathbb{P}(A) = 10/36, \quad \mathbb{P}(\{\omega\}) = 1/36,\ \omega\in\Omega \implies$

$$\begin{aligned} \mathbb{E}(X_1|A) &= \mathbb{E}(X_1|X_1+X_2\geq 9) \\ &= \frac{1}{10/36}\cdot\frac{1}{36}\cdot(3+4+5+6+4+5+6+5+6+6) \\ &= 5. \end{aligned}$$

Memo: $\mathbb{E}(X|A) \;=\; \mathbb{E}_{\mathbb{P}_A}(X) \;=\; \sum_{\omega\in A\cap\Omega_0} X(\omega)\,\mathbb{P}_A(\{\omega\})$

16.3 Satz (Eigenschaften des bedingten Erwartungswertes)

Seien

- X, Y Zufallsvariablen mit $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$,
- $A \subseteq \Omega$ mit $\mathbb{P}(A) > 0$,
- $Z : \Omega \to \mathbb{R}^k$ ein Zufallsvektor und $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z) > 0$.

Dann gelten:

a) $\mathbb{E}(X + Y|A) \;=\; \mathbb{E}(X|A) \;+\; \mathbb{E}(Y|A)$, $\quad\surd$

b) $\mathbb{E}(a\,X|A) \;=\; a\,\mathbb{E}(X|A)$, $\quad a \in \mathbb{R}$, $\quad\surd$

c) $\mathbb{E}(\mathbf{1}_B|A) \;=\; \mathbb{P}(B|A)$, $\quad B \subseteq \Omega$, $\quad\surd$

d) $\mathbb{E}(X|A) = \sum_{j\geq 1} x_j\,\mathbb{P}(X = x_j|A)$, falls $\sum_{j\geq 1} \mathbb{P}(X = x_j) = 1$, $\quad\surd$

e) $\mathbb{E}(X|Z = z) = \sum_{j\geq 1} x_j\,\mathbb{P}(X = x_j|Z{=}z)$, falls $\sum_{j\geq 1} \mathbb{P}(X{=}x_j) = 1$, $\quad\surd$

f) X, Z stochastisch unabhängig $\Longrightarrow \mathbb{E}(X|Z = z) \;=\; \mathbb{E}(X)$. $\quad\surd$

16.4 Beste Prognose im Sinne der mittleren quadratischen Abweichung

Wunsch: Realisierungen $X(\omega)$ von X mit Hilfe der Realisierungen $Z(\omega)$ eines k-dimensionalen Zufallsvektors Z vorhersagen.

Vorhersage erfolgt über eine Funktion $h : \mathbb{R}^k \to \mathbb{R}$.

Dabei: $h(Z(\omega)) =$ Prognosewert für $X(\omega)$ bei Kenntnis von $Z(\omega)$.

Gütekriterium: Mittlere quadratische Abweichung (MQA)

$$\mathbb{E}\left(X - h(Z)\right)^2 \;=\; \sum_{\omega \in \Omega_0} \left(X(\omega) - h(Z(\omega))\right)^2 \mathbb{P}(\{\omega\})$$

Welche Prognose-Funktion h liefert die kleinstmögliche MQA?

Beachte: $\min_{a \in \mathbb{R}} \mathbb{E}(X-a)^2 \;=\; \mathbb{E}(X - \mathbb{E}X)^2$.

Vgl. Kapitel 12: Dort Aufgabe $\mathbb{E}(X-(a+bY))^2 = \min\limits_{a,b \in \mathbb{R}}!$

Memo: $\mathbb{E}\,(X-h(Z))^2 \;=\; \sum_{\omega\in\Omega_0}(X(\omega)-h(Z(\omega)))^2\,\mathbb{P}(\{\omega\})\;(*)$

16.5 Satz (bedingter Erwartungswert als beste Vorhersage)

Es gelte $\mathbb{P}(Z=z_j)>0$, $j\ge 1$, und $\sum_{j\ge1}\mathbb{P}(Z=z_j)=1$.

Dann wird die MQA $\mathbb{E}(X-h(Z))^2$ minimal für

$$h(z):=\begin{cases}\mathbb{E}(X|Z=z_j)\;, & \text{falls } z=z_j \text{ für ein } j\ge 1\\ 0\;, & \text{falls } z\in\mathbb{R}^k\setminus\{z_1,z_2,z_3,\ldots\}\end{cases}.$$

Beweis: Sei $A_j:=\{Z=z_j\}$. Sortiere in $(*)$ nach gleichen Werten z_j für $Z(\omega)$. Mit $\mathbb{P}_{A_j}(\{\omega\})=\frac{\mathbb{P}(\{\omega\})}{\mathbb{P}(Z=z_j)}$, $\omega\in A_j$, und $\mathbb{P}_{A_j}(\{\omega\})=0$ für $\omega\in\Omega\setminus A_j$ folgt

$$\begin{aligned}\mathbb{E}\,(X-h(Z))^2 &= \sum_{j\ge1}\sum_{\omega\in A_j}(X(\omega)-h(z_j))^2\,\mathbb{P}(\{\omega\})\\ &= \sum_{j\ge1}\mathbb{P}(Z=z_j)\sum_{\omega\in A_j}(X(\omega)-h(z_j))^2\,\mathbb{P}_{A_j}(\{\omega\})\\ &= \sum_{j\ge1}\mathbb{P}(Z=z_j)\sum_{\omega\in \Omega_0}(X(\omega)-h(z_j))^2\,\mathbb{P}_{A_j}(\{\omega\})\\ &= \sum_{j\ge1}\mathbb{P}(Z=z_j)\,\mathbb{E}_{\mathbb{P}_{A_j}}(X-h(z_j))^2.\end{aligned}$$

Memo: $\mathbb{E}\,(X-h(Z))^2 = \sum_{j\geq 1} \mathbb{P}(Z=z_j)\,\mathbb{E}_{\mathbb{P}_{A_j}}(X-h(z_j))^2$

Memo: Allgemein gilt $\mathbb{E}(Y-\mathbb{E}\,Y)^2 \;=\; \min_{a\in\mathbb{R}} \mathbb{E}(Y-a)^2$

Folgerung: Die MQA $\mathbb{E}_{\mathbb{P}_{A_j}}(X-h(z_j))^2$ wird minimal für

$$h(z_j) \;:=\; \mathbb{E}_{\mathbb{P}_{A_j}}(X) \;=\; \mathbb{E}(X|A_j) \;=\; \mathbb{E}(X|Z=z_j), \qquad j\geq 1.\ \surd$$

Beachte: Der zweite Teil der Definition

$$h(z) \;:=\; \begin{cases} \mathbb{E}(X|Z=z_j)\ , & \text{falls } z=z_j \text{ für ein } \ j\geq 1 \\ 0\ , & \text{falls } z\in\mathbb{R}^k\setminus\{z_1,z_2,z_3,\ldots\} \end{cases} \;.$$

von $h(z)$ ist willkürlich (dient nur dazu, die Funktion h auf ganz $\mathbb{R}^k$ zu definieren.)

Memo: $h(z) = \mathbb{E}(X|Z = z_j)$, falls $z = z_j$ für ein $j \geq 1$.

Memo: $h(z) := 0$, falls $z \in \mathbb{R}^k \setminus \{z_1, z_2, z_3, \ldots\}$

16.6 Definition (bedingte Erwartung)

Die durch

$$\mathbb{E}(X|Z)(\omega) := h(Z(\omega)) = \begin{cases} \mathbb{E}(X|Z = Z(\omega)) \ , & \text{falls } Z(\omega) \in \{z_1, z_2, \ldots\} \\ 0 \ , & \text{sonst} \ , \end{cases}$$

($\omega \in \Omega$) definierte Zufallsvariable $\mathbb{E}(X|Z)$ heißt bedingte Erwartung von X bei gegebenem Z.

Beachte:

- $\mathbb{E}(X|Z)(\omega)$, $\omega \in \Omega$, hängt nur von $Z(\omega)$ ab,
- $\mathbb{E}(X|Z)$ ist als Funktion auf Ω konstant auf den Mengen $\{Z = z_j\}$, $j \geq 1$.

Memo: $\mathbb{E}(X|Z)(\omega) = \mathbb{E}(X|Z = Z(\omega))$, falls $Z(\omega) \in \{z_1, z_2, \ldots\}$

16.7 Beispiel (Zweifacher Würfelwurf, $\mathbb{E}\left(\max(X_1, X_2)|X_1\right)$)

Sei X_j = die Augenzahl des j-ten Wurfs, $M := \max(X_1, X_2)$.

$$\mathbb{P}(M = k|X_1 = j) \;=\; \begin{cases} j/6, & \text{falls } k = j, \\ 1/6, & \text{falls } k > j. \end{cases}$$

$$\begin{aligned} \mathbb{E}(M|X_1 = j) &= j \cdot \frac{j}{6} + \sum_{k=j+1}^{6} k \cdot \frac{1}{6} = \frac{1}{6} \cdot \left(j^2 + 21 - \frac{j\,(j+1)}{2} \right) \\ &= 3.5 + \frac{j(j-1)}{12}, \qquad j = 1, \ldots, 6 \end{aligned}$$

$$\mathbb{E}(M|X_1) = 3.5 + \frac{X_1\,(X_1 - 1)}{12}.$$

j	1	2	3	4	5	6
$\mathbb{E}(M\|X_1 = j)$	3.5	3.67	4	4.5	5.17	6

Bemerkungen:

- Falls anderes Gütekriterium, z.B.

  $$\mathbb{E}|M - h(X_1)| \ = \ \min_h !,$$

  so andere optimale Vorhersagefunktion.

- Falls $h(j) \in \{1, 2, 3, 4, 5, 6\}$ gefordert, so ebenfalls andere optimale Vorhersagefunktion.

Memo: $\Omega = \sum_{j\geq 1} A_j,\ \mathbb{P}(A_j) > 0\ \forall j \Longrightarrow \mathbb{P}(B) = \sum_{j\geq 1} \mathbb{P}(B|A_j)\,\mathbb{P}(A_j)$

Memo: $\mathbb{E}(X|A_j) = \dfrac{1}{\mathbb{P}(A_j)} \displaystyle\sum_{\omega\in A_j\cap\Omega_0} X(\omega)\,\mathbb{P}(\{\omega\})$

16.8 Satz (Formel vom totalen Erwartungswert)

Seien

- $(\Omega, \mathbb{P})$ ein diskreter W-Raum,
- $A_1, A_2, \ldots$ paarweise disjunkte Ereignisse mit $\mathbb{P}(A_j) > 0$, $j \geq 1$, und $\sum_{j\geq 1}\mathbb{P}(A_j) = 1$.

Dann gilt für jede Zufallsvariable $X : \Omega \to \mathbb{R}$ mit $\mathbb{E}|X| < \infty$:

$$\mathbb{E}(X) = \sum_{j\geq 1} \mathbb{E}(X|A_j)\,\mathbb{P}(A_j).$$

Speziell: $X = \mathbf{1}_B$, $B \subseteq \Omega \Longrightarrow$ Formel von der totalen Wahrscheinlichkeit.

16.9 Folgerung (Iterierte Erwartungswertbildung)

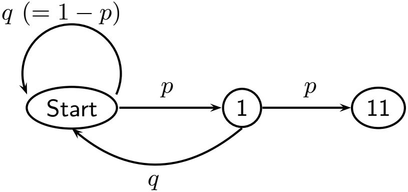
Gilt speziell $A_j = \{Z = z_j\}$ für einen Zufallsvektor $Z : \Omega \to \mathbb{R}^k$, der die Werte $z_1, z_2, \ldots$ mit positiver Wahrscheinlichkeit annimmt, so folgt

$$\begin{aligned} \mathbb{E}(X) &= \sum_{j \geq 1} \mathbb{E}(X | Z = z_j)\, \mathbb{P}(Z = z_j) \\ &= \mathbb{E}(\mathbb{E}(X|Z)). \end{aligned}$$

16.10 Beispiel (Warten auf den ersten Doppeltreffer)

Sei X die Anzahl der Versuche in einer Bernoulli-Kette mit Trefferw' p, bis erstmals zwei direkt aufeinander folgende Treffer auftreten.

$\mathbb{E}\,X =$?



Zustandsgraph beim Warten auf den ersten Doppeltreffer

$$\begin{aligned}\Omega \quad &:= \quad \{11, 011, 0011, 1011, 00011, 01011, 10011, 000011, \ldots\} \\ &= \quad \{a_1 \ldots a_n :\ n \geq 2,\ a_j \in \{0,1\},\ a_j a_{j+1} = 0\ (j \leq n-2),\ a_{n-1}a_n = 1\}\end{aligned}$$

$$\begin{aligned} A_1 &:= \{\omega = a_1 \ldots a_n \in \Omega : a_1 = 0\}, \\ A_2 &:= \{\omega = a_1 \ldots a_n \in \Omega : a_1 = 1,\ a_2 = 0\}, \\ A_3 &:= \{\omega = a_1 \ldots a_n \in \Omega : a_1 = 1,\ a_2 = 1\}. \end{aligned}$$

Es gelten $A_1 + A_2 + A_3 = \Omega$, $\mathbb{P}(A_1) = q$, $\mathbb{P}(A_2) = pq$, $\mathbb{P}(A_3) = p^2$.

$$\begin{aligned} \mathbb{E}(X|A_1) &= 1 + \mathbb{E}\,X, \\ \mathbb{E}(X|A_2) &= 2 + \mathbb{E}\,X, \\ \mathbb{E}(X|A_3) &= 2. \end{aligned}$$

Aus der Formel vom totalen Erwartungswert folgt

$$\begin{aligned} \mathbb{E}\,X &= (1 + \mathbb{E}\,X)\cdot q \ + \ (2 + \mathbb{E}\,X)\cdot pq \ + \ 2p^2 \\ \Longrightarrow \mathbb{E}\,X &= \frac{1+p}{p^2}. \end{aligned}$$

Insbesondere gilt $\mathbb{E}(X) = 6$ im Fall $p = 1/2$.

16.11 Beispiel (Zwischen Angst und Gier: Die Sechs verliert)

Wiederholter Würfelwurf. Addition der Augenzahlen auf Punktekonto, solange keine Sechs auftritt. Man kann ohne Sechs jederzeit stoppen, und erzielter Punktestand ist Gewinn. Bei Auftreten einer Sechs ist der Gewinn gleich 0.

Welche Stoppstrategie, wenn Spiel oftmals wiederholt gespielt wird?

Bei Punktestand k sei X_k Punktestand nach gedanklichem weiteren Wurf.

Welche Werte nimmt X_k an? Antwort: $k+1, k+2, k+3, k+4, k+5, 0$, jeweils mit W' $1/6$.

$$\mathbb{E}(X_k) \;=\; \frac{1}{6}\sum_{j=1}^{5}(k+j) \;=\; \frac{5\,k+15}{6} \;>k \iff k<15.$$

Also: Spiele weiter, falls Punktestand <15, sonst stoppe!

Sei G der zufällige Spielgewinn unter dieser Strategie. $\mathbb{E}(G)=?$

Sei $\mathbb{E}_k(G)$ der Erwartungswert von G unter der Bedingung, dass ein momentaner Punktestand von k vorliegt

$$\Longrightarrow\; \mathbb{E}(G) \;=\; \mathbb{E}_0(G).$$

Nach der Stoppregel gilt: $\mathbb{E}_k(G) = k, \quad$ falls $\; k \in \{15, 16, 17, 18, 19\}$.

Für $k \leq 14$ betrachten wir das zufällige Ergebnis X des nächsten Wurfs.

Formel vom totalen Erwartungswert für $\mathbb{E}_k(G) \implies$

$$\mathbb{E}_k(G) \;=\; \sum_{j=1}^{6} \mathbb{E}_k(G|X=j)\,\mathbb{P}(X=j).$$

Es gilt $\mathbb{E}_k(G|X=6) = 0$ und $\mathbb{E}_k(G|X=j) = \mathbb{E}_{k+j}(G)$ für $j \leq 5$, also

$$\mathbb{E}_k(G) \;=\; \frac{1}{6}\sum_{j=1}^{5} \mathbb{E}_{k+j}(G).$$

Jetzt $\mathbb{E}_0(G)$ durch sog. Rückwärts-Induktion berechenbar:

$$\begin{aligned}
\mathbb{E}_{14}(G) \;&=\; \frac{1}{6}\cdot(15+16+17+18+19) \;=\; \frac{85}{6} \;\approx\; 14.167,\\
\mathbb{E}_{13}(G) \;&=\; \frac{1}{6}\cdot\left(\frac{85}{6}+15+16+17+18\right) \;=\; \frac{481}{36} \;\approx\; 13.361
\end{aligned}$$

usw. Schließlich ergibt sich $\mathbb{E}(G) \;=\; \mathbb{E}_0(G) \;\approx\; 6.154$.

16.12 Satz (Die Substitutionsregel)

Es seien

- $(\Omega, \mathbb{P})$ ein diskreter W-Raum,
- $X : \Omega \to \mathbb{R}^n$, $Z : \Omega \to \mathbb{R}^k$ Zufallsvektoren,
- $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ eine Funktion mit $\mathbb{E}|g(X, Z)| < \infty$.

Dann gilt für jedes $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z) > 0$:

$$\mathbb{E}(g(X,Z)|Z = z) \; = \; \mathbb{E}(g(X,z)|Z = z).$$

BEWEIS:

$$\begin{aligned}
\mathbb{E}(g(X,Z)|Z = z) &= \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega \in \Omega : Z(\omega) = z} g(X(\omega), Z(\omega)) \, \mathbb{P}(\{\omega\}) \\
&= \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega \in \Omega : Z(\omega) = z} g(X(\omega), z) \, \mathbb{P}(\{\omega\}) \\
&= \mathbb{E}(g(X,z)|Z = z) \; \surd
\end{aligned}$$

16.13 Beispiel (Augensumme bei zufälliger Wurfanzahl)

- Echten Würfel werfen
- Fällt Augenzahl j, so danach j echte Würfel werfen
- Sei X die *insgesamt* gewürfelte Augensumme
- $\mathbb{E}\, X =$?

Modellierung:

- $\Omega := \{1, 2, \ldots, 6\}^7 = \{\omega = (\omega_0, \omega_1, \ldots, \omega_6) : 1 \leq \omega_i \leq 6$ für $i = 0, \ldots, 6\}$
- $\mathbb{P}$ sei Gleichverteilung auf Ω
- $X_i(\omega) := \omega_i, \ \ 0 \leq i \leq 6.$
- $X_0, X_1, \ldots, X_6$ sind stochastisch unabhängig
- $$X(\omega) := X_0(\omega) + \sum_{i=1}^{X_0(\omega)} X_i(\omega), \qquad \omega \in \Omega.$$

Memo: $X = X_0 + \sum_{i=1}^{X_0} X_i$

$$
\begin{aligned}
\mathbb{E}(X|X_0 = j) &= \mathbb{E}\left(X_0 + \sum_{i=1}^{X_0} X_i \middle| X_0 = j\right) \\
&= \mathbb{E}\left(j + \sum_{i=1}^{j} X_i \middle| X_0 = j\right) \qquad (\text{Substitutionsregel}) \\
&= \mathbb{E}(j|X_0 = j) + \sum_{i=1}^{j} \mathbb{E}(X_i|X_0 = j) \quad (\text{Additivität}) \\
&= j + \sum_{i=1}^{j} \mathbb{E}(X_i) \qquad (X_i,\ X_0 \text{ unabhängig}) \\
&= j + j \cdot 3.5
\end{aligned}
$$

Iterierte Erwartungswertbildung (Formel vom totalen Erwartungswert) $\Longrightarrow$

$$
\mathbb{E}\,X \;=\; \sum_{j=1}^{6} \mathbb{E}(X|X_0 = j)\,\mathbb{P}(X_0 = j) \;=\; \frac{1}{6} \cdot 4.5 \cdot \sum_{j=1}^{6} j \;=\; 15.75.
$$

16.14 Definition (bedingte Verteilung)

Es seien

- $(\Omega, \mathbb{P})$ ein diskreter W-Raum,
- $X : \Omega \to \mathbb{R}^n$, $Z : \Omega \to \mathbb{R}^k$ Zufallsvektoren,
- $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z) > 0$.

Dann heißt das durch

$$\mathbb{P}^X_{Z=z}(B) \ := \ \mathbb{P}(X \in B | Z = z), \qquad B \subseteq \mathbb{R}^n,$$

definierte Wahrscheinlichkeitsmaß $\mathbb{P}^X_{Z=z}$ auf $\mathcal{P}(\mathbb{R}^n)$

bedingte Verteilung von X unter der Bedingung $Z = z$.

Falls $\sum_{j \geq 1} \mathbb{P}(X = x_j) = 1$, so gilt

$$\mathbb{P}(X \in B | Z = z) \ = \ \sum_{j : x_j \in B} \mathbb{P}(X = x_j | Z = z).$$

$\mathbb{P}^X_{Z=z}$ ist durch das System $\mathbb{P}(X = x_j | Z = z)$, $j \geq 1$, eindeutig bestimmt.

16.15 Beispiel (Hypergeometrische Verteilung als bedingte Verteilung)

Seien X, Y unabhängig, wobei $X \sim \mathrm{Bin}(m,p)$, $Y \sim \mathrm{Bin}(n,p)$, $0 < p < 1$. Dann gilt:

$$\mathbb{P}^X_{X+Y=k} \;=\; \mathrm{Hyp}(k,m,n), \qquad k \in \{1,2,\ldots,m+n\}.$$

Beweis:

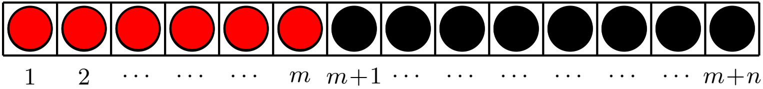
$$\begin{aligned}
\mathbb{P}(X=j|X+Y=k) &= \frac{\mathbb{P}(X=j, X+Y=k)}{\mathbb{P}(X+Y=k)} = \frac{\mathbb{P}(X=j, Y=k-j)}{\mathbb{P}(X+Y=k)} \\
&= \frac{\mathbb{P}(X=j)\,\mathbb{P}(Y=k-j)}{\mathbb{P}(X+Y=k)} \\
&= \frac{\binom{m}{j}p^j(1-p)^{m-j}\,\binom{n}{k-j}p^{k-j}(1-p)^{n-(k-j)}}{\binom{m+n}{k}p^k(1-p)^{m+n-k}} \\
&= \frac{\displaystyle\binom{m}{j}\binom{n}{k-j}}{\displaystyle\binom{m+n}{k}}. \;\surd
\end{aligned}$$

Memo: X, Y unabhängig, $X \sim \mathrm{Bin}(m,p)$, $Y \sim \mathrm{Bin}(n,p)$, $0 < p < 1$.

Memo: $\mathbb{P}^X_{X+Y=k} = \mathrm{Hyp}(k,m,n), \qquad k \in \{1,2,\ldots,m+n\}$

Begriffliche Einsicht?

Bernoulli-Kette der Länge $m+n$ mit Trefferwahrscheinlichkeit p.



X = Trefferanzahl in den ersten m Versuchen, Y = Trefferanzahl in den letzten n Versuchen; $X+Y=k \Longleftrightarrow$ es treten insgesamt k Treffer auf.

Jede Auswahl von k der $m+n$ Versuche für die Treffer ist gleich wahrsch.'

Interpretiere Versuche als von 1 bis $m+n$ nummerierte Kugeln in Urne. Die Kugeln $1,\ldots,m$ seien rot, die anderen schwarz.

k Kugeln rein zufällig ohne Zurücklegen ziehen. Kugelnummern entspr. Versuchsnummern mit Treffer. X = Anz. der gezogenen roten Kugeln.

16.16 Beispiel (Multinomialverteilung bei gegebener Anzahl von Treffern)

Sei $(X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s)$ mit $p_1 > 0, \ldots, p_s > 0$.

Sei $l \in \{0, 1, \ldots, n-1\}$. $\mathbb{P}_{X_s=l}^{(X_1,\ldots,X_{s-1})} = ?$

Seien $k_1, \ldots, k_{s-1} \in \mathbb{N}_0$ mit $k_1 + \ldots + k_{s-1} = n - l$.

$$\begin{aligned}
\mathbb{P}(X_1 = k_1, \ldots, X_{s-1} = k_{s-1} | X_s = l) &= \frac{\mathbb{P}(X_1 = k_1, \ldots, X_{s-1} = k_{s-1}, X_s = l)}{\mathbb{P}(X_s = l)} \\
&= \frac{\dfrac{n!}{k_1! \ldots k_{s-1}!\, l!} p_1^{k_1} \ldots p_{s-1}^{k_{s-1}} p_s^l}{\dfrac{n!}{l!(n-l)!} p_s^l (1-p_s)^{n-l}} \\
&= \frac{(n-l)!}{k_1! \ldots k_{s-1}!} \left(\frac{p_1}{1-p_s}\right)^{k_1} \ldots \left(\frac{p_{s-1}}{1-p_s}\right)^{k_{s-1}}
\end{aligned}$$

$$\Longrightarrow \quad \mathbb{P}_{X_s=l}^{(X_1,\ldots,X_{s-1})} = \text{Mult}\left(n-l; \frac{p_1}{1-p_s}, \ldots, \frac{p_{s-1}}{1-p_s}\right).$$

Begriffliche Einsicht?

16.17 Beispiel (Gleichverteilung und geometrische Verteilung)

Es seien X, Y stochastisch unabhängig und je G(p)-verteilt, wobei $0 < p < 1$.

Sei allgemein U(M) die Gleichverteilung auf einer endlichen Menge M.

Dann gilt für jedes $k \in \mathbb{N}_0$:

$$\mathbb{P}^X_{X+Y=k} \;=\; \mathsf{U}(\{0, 1, \ldots, k\}).$$

Begriffliche Einsicht?

Exkurs: Das Briefumschlag-Paradoxon

In zwei Umschlägen A und B ist jeweils ein unbekannter Geldbetrag, wobei der eine doppelt so hoch ist wie der andere.

Ich darf einen Umschlag wählen und den darin enthaltenen Betrag behalten.

Ich wähle Umschlag A und überlege:

Befindet sich der Betrag x in A, so befindet sich entweder der Betrag $2x$ oder der Betrag $x/2$ in B.

Wenn ich tausche, ist der Erwartungswert meines Geldbetrages gleich

$$\frac{1}{2} \cdot 2x + \frac{1}{2} \cdot \frac{x}{2} \;=\; \frac{5}{4} \cdot x \;>\; x$$

Also sollte ich tauschen, oder?

Aber: Die gleiche Überlegung hätte ich auch bei Umschlag B angestellt!

17 Erzeugende Funktionen

17.1 Definition (Erzeugende Funktion einer Zahlenfolge)

Sei $a := (a_k)_{k\geq 0}$ eine reelle Zahlenfolge. Die durch

$$g_a(t) \ := \sum_{k=0}^{\infty} a_k \, t^k, \qquad -\rho < t < \rho,$$

definierte Potenzreihe g_a heißt erzeugende Funktion von a.

Dabei gelte

$$\rho \ := \ \frac{1}{\limsup_{k\to\infty} |a_k|^{1/k}} \in (0,\infty] \qquad \text{(Konvergenzradius positiv, evtl. } = \infty \text{)}$$

Beispiel: Sei $q > 0$, $a_k := q^k \quad \Longrightarrow$

$$g_a(t) \ = \ \sum_{k=0}^{\infty} q^k \, t^k \ = \ \sum_{k=0}^{\infty} (qt)^k \ = \ \frac{1}{1-qt}, \quad \text{falls } |t| < \frac{1}{q}$$

Memo: $g_a(t) \;=\; \sum_{k=0}^{\infty} a_k\, t^k, \quad -\rho < t < \rho$

- g_a ist in $(-\rho, \rho)$ beliebig oft differenzierbar, wobei

  $g_a'(t) = \sum_{k=1}^{\infty} k\, a_k\, t^{k-1}, \quad g_a''(t) = \sum_{k=2}^{\infty} k(k-1)\, a_k\, t^{k-2}$, allgemein:

  $$\frac{\mathrm{d}^r g_a(t)}{\mathrm{d}t^r} \;=\; \sum_{k=r}^{\infty} k(k-1) \cdot \ldots \cdot (k-r+1)\, a_k\, t^{k-r} \;, \qquad |t| < \rho.$$

- g_a legt a fest: $\left.\frac{\mathrm{d}^r g_a(t)}{\mathrm{d}t^r}\right|_{t=0} \;=\; r! \cdot a_r$ (Identitätssatz!)

- Im Fall $\rho < \infty$ gilt der Abelsche Grenzwertsatz:

  Konvergiert $g_a(t)$ im rechten Randpunkt ρ des Konvergenzintervalls, so ist die auf $(-\rho, \rho]$ erklärte Funktion g_a im Punkt ρ stetig, d.h. es gilt

  $$\sum_{k=0}^{\infty} a_k\, \rho^k \;=\; \lim_{t \uparrow \rho} g_a(t).$$

17.2 Beispiel (Fibonacci-Zahlen)

Sei $a_0 := a_1 := 1,\ \ a_k := a_{k-1} + a_{k-2}$ für $k \geq 2$ $\qquad (\Longrightarrow a_k \leq 2^k)$

$$
\begin{aligned}
g_a(t) &= \sum_{k=0}^{\infty} a_k\, t^k \ = \ a_0 + a_1\, t + \sum_{k=2}^{\infty} (a_{k-1} + a_{k-2})\, t^k \\
&= 1 + t + \sum_{k=2}^{\infty} a_{k-1}\, t^k + \sum_{k=2}^{\infty} a_{k-2}\, t^k \\
&= 1 + t + t \sum_{k=2}^{\infty} a_{k-1}\, t^{k-1} + t^2 \sum_{k=2}^{\infty} a_{k-2}\, t^{k-2} \\
&= 1 + t + t\ (g_a(t) - 1)\ + t^2\, g_a(t) \\
\Longrightarrow g_a(t) &= \frac{1}{1 - t - t^2}, \qquad |t| < \frac{1}{2} \\
\Longrightarrow g_a(t) &= \sum_{k=0}^{\infty} \underbrace{\left[\frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{k+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{k+1} \right) \right]}_{=a_k} t^k
\end{aligned}
$$

17.3 Definition (Erzeugende Funktion einer $\mathbb{N}_0$-wertigen Zufallsvariablen)

Es sei X eine $\mathbb{N}_0$-wertige Zufallsvariable, d.h. es gelte $\mathbb{P}(X \in \mathbb{N}_0) = 1$.

Die durch

$$g_X(t) \; := \; \sum_{k=0}^{\infty} \mathbb{P}(X = k)\, t^k, \qquad |t| \le 1,$$

definierte Potenzreihe heißt erzeugende Funktion (der Verteilung) von X.

Beachte:

- g_X ist die erzeugende Funktion der Folge $(\mathbb{P}(X = k))_{k \ge 0}$
- $g_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1 \Longrightarrow$ Konvergenzradius ist mindestens 1
- Die Funktion g_X legt die Verteilung von X fest
- Der Definitionsbereich Ω von X (Grundraum) ist irrelevant
- Es gilt $g_X(t) \; = \; \mathbb{E}\left(t^X\right), \qquad |t| \le 1$

17.4 Beispiele (Binomial-, Poisson- und negative Binomialverteilung)

a) Im Fall $X \sim \mathsf{Bin}(n,p)$ gilt

$$g_X(t) \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k \, (1-p)^{n-k} \, t^k \;=\; (1-p+pt)^n.$$

b) Im Fall $X \sim \mathsf{Po}(\lambda)$ ergibt sich

$$g_X(t) \;=\; \sum_{k=0}^{\infty} \mathsf{e}^{-\lambda} \, \frac{\lambda^k}{k!} \, t^k \;=\; \mathsf{e}^{-\lambda} \, \mathsf{e}^{\lambda t} \;=\; \mathsf{e}^{\lambda(t-1)} \; .$$

c) Gilt $X \sim \mathsf{Nb}(r,p)$, so folgt

$$g_X(t) \;=\; \left(\frac{p}{1-(1-p)t} \right)^r, \qquad |t| < \frac{1}{1-p} \qquad \text{(Übungsaufgabe !).}$$

Memo: X, Y unabhängig $\Longrightarrow u(X), v(Y)$ unabhängig, $g_X(t) = \mathbb{E}\left(t^X\right)$.

Memo: S, T unabhängig $\Longrightarrow \mathbb{E}(ST) = \mathbb{E}\,S\,\mathbb{E}\,T$

17.5 Satz (Eindeutigkeitssatz)

Es seien X und Y $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugenden Funktionen g_X und g_Y. Dann gilt:

$$g_X \;=\; g_Y \iff X \sim Y, \quad \text{also } \mathbb{P}(X = k) \;=\; \mathbb{P}(Y = k), \quad k \geq 0.$$

17.6 Satz (Multiplikationsformel)

Es seien X und Y unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugenden Funktionen g_X und g_Y. Dann gilt:

$$g_{X+Y}(t) \;=\; g_X(t)\, g_Y(t), \qquad |t| \leq 1.$$

BEWEIS: $|t| \leq 1 \;\Longrightarrow$

$$g_{X+Y}(t) \;=\; \mathbb{E}\left(t^{X+Y}\right) \;=\; \mathbb{E}\left(t^X\, t^Y\right) \;=\; \mathbb{E}\left(t^X\right) \mathbb{E}\left(t^Y\right) \;=\; g_X(t)\, g_Y(t)$$

$$\begin{array}{rcl}
X \sim \text{Bin}(\mathbf{n}, p) & \Longleftrightarrow & g_X(t) = (1-p+pt)^{\mathbf{n}} \quad \forall t \quad (\Leftarrow \ \text{mit Satz } 17.5) \\
Y \sim \text{Bin}(\mathbf{m}, p) & \Longleftrightarrow & g_Y(t) = (1-p+pt)^{\mathbf{m}} \quad \forall t \quad (\Leftarrow \ \text{mit Satz } 17.5) \\
X, Y \text{ unabhängig} & \Longrightarrow & g_{X+Y}(t) = g_X(t)\, g_Y(t) = (1-p+pt)^{\mathbf{m+n}} \ \forall t \\
 & & (\text{mit Satz } 17.6) \\
\text{Satz } 17.5 & \Longrightarrow & X + Y \sim \text{Bin}(\mathbf{m} + \mathbf{n}, p) \\
X \sim \text{Po}(\lambda) & \Longleftrightarrow & g_X(t) = \mathrm{e}^{\lambda(t-1)} \\
X \sim \text{Nb}(\mathbf{r}, p) & \Longleftrightarrow & g_X(t) = \left(\dfrac{p}{1-(1-p)t} \right)^{\mathbf{r}}
\end{array}$$

17.7 Folgerungen (Additionsgesetze für Bin(n, p), Po(λ) und Nb(r, p))

Sind X und Y unabhängig, so gelten:

a) $X \sim \text{Bin}(m, p),\ Y \sim \text{Bin}(n, p) \implies X + Y \sim \text{Bin}(m+n, p)$,

b) $X \sim \text{Po}(\lambda),\ Y \sim \text{Po}(\mu) \implies X + Y \sim \text{Po}(\lambda + \mu)$,

c) $X \sim \text{Nb}(r, p),\ Y \sim \text{Nb}(s, p) \implies X + Y \sim \text{Nb}(r+s, p)$.

17.8 Beispiel (Augensummenverteilung beim mehrfachen Würfelwurf)

Es seien $X_1, \ldots, X_n$ unabhängig und je gleichverteilt auf $\{1, \ldots, s\}$.

Welche Verteilung besitzt $S_n \;:=\; X_1 + \ldots + X_n$?

Sei

$$g(t) \;:=\; \mathbb{E}t^{X_1} \;=\; \frac{1}{s}\left(t + t^2 + \ldots + t^s\right)$$

Multiplikationsformel 17.6 $\Longrightarrow$

$$\begin{aligned}
g_{S_n}(t) \;&=\; g(t)^n \;=\; \frac{1}{s^n}\left(t + t^2 + \ldots + t^s\right)^n \\
&=\; \frac{1}{s^n}\, t^n \left(\sum_{j=0}^{s-1} t^j\right)^n \;=\; \frac{t^n}{s^n}\left(\frac{t^s - 1}{t - 1}\right)^n \qquad (t \neq 1) \\
&=\; \frac{t^n}{s^n}\,(t-1)^{-n}\,(t^s - 1)^n \\
&=\; \frac{t^n}{s^n}\,(1-t)^{-n}\,(-1)^n\,(1-t^s)^n\,(-1)^n \\
&=\; \frac{t^n}{s^n} \sum_{j=0}^{\infty} \binom{n+j-1}{j} t^j \sum_{i=0}^{n} (-1)^i \binom{n}{i} t^{i\,s}
\end{aligned}$$

$$\begin{aligned} g_{S_n}(t) &= \frac{t^n}{s^n}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\,t^j\,\sum_{i=0}^{n}(-1)^i\binom{n}{i}t^{i\,s} \\ &= \frac{1}{s^n}\sum_{j=0}^{\infty}\binom{n+j-1}{j}\sum_{i=0}^{n}(-1)^i\binom{n}{i}t^{n+i\,s+j} \\ &\overset{!}{=} \sum_{k=0}^{\infty}\mathbb{P}(S_n=k)\,t^k \end{aligned}$$

Setze $k := n+i\,s+j \iff j = k-n-i\,s \Longrightarrow$

$$\binom{n+j-1}{j} = \binom{k-i\,s-1}{k-i\,s-n} = \binom{k-i\,s-1}{n-1}$$

Beachte: $k-i\,s-1 \geq n-1 \iff i\,s \leq k-n \iff i \leq \lfloor\frac{k-n}{s}\rfloor$.

Weiter gilt $n \leq k \leq n\,s$, da andernfalls $\mathbb{P}(S_n=k)=0$. Es folgt

$$g_{S_n}(t) = \sum_{k=n}^{n\,s}\sum_{i=0}^{\lfloor\frac{k-n}{s}\rfloor}\binom{k-i\,s-1}{n-1}(-1)^i\binom{n}{i}\frac{1}{s^n}\,t^k$$

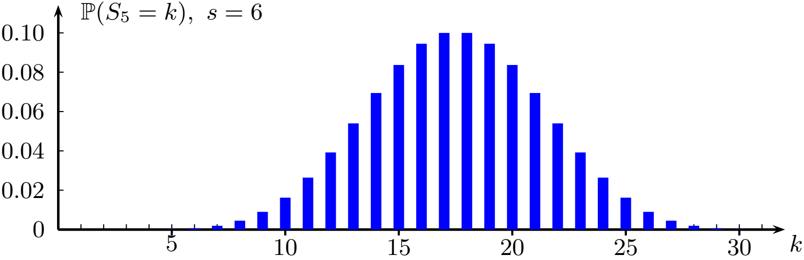
Koeffizientenvergleich $\Longrightarrow$

17.9 Satz (Augensummen-Verteilung beim n-fachen Würfelwurf)

Es seien $X_1,\ldots,X_n$ stochastisch unabhängig und je gleichverteilt auf $\{1,2,\ldots,s\}$. Sei $S_n := X_1+\ldots+X_n$. Dann gilt:

$$\mathbb{P}(S_n=k) \;=\; \sum_{i=0}^{\lfloor\frac{k-n}{s}\rfloor} \binom{k-i\,s-1}{n-1}(-1)^i\binom{n}{i}\frac{1}{s^n},$$

falls $k\in\{n,n+1,\ldots,n\,s\}$ und $\mathbb{P}(S_n=k)=0$, sonst.



Stabdiagramm der Verteilung der Augensumme beim fünffachen Würfelwurf

17.10 Satz (Erzeugende Funktionen und Momente)

Seien X eine $\mathbb{N}_0$-wertige Zufallsvariable mit erzeugender Funktion g und $r \in \mathbb{N}$. Dann sind folgende Aussagen äquivalent:

a) $\mathbb{E}\left[X\,(X-1)\cdot\ldots\cdot(X-r+1)\right]$ existiert

b) $g^{(r)}(1-) \;:=\; \lim\limits_{t\uparrow 1} \dfrac{\mathrm{d}^r}{\mathrm{d}t^r} g(t)$ existiert.

Im Fall der Existenz gilt

$$\mathbb{E}\left[X(X-1)\cdot\ldots\cdot(X-r+1)\right] \;=\; g^{(r)}(1-).$$

Beweis: a) $\iff \displaystyle\sum_{k=r}^{\infty} k(k-1)\cdot\ldots\cdot(k-r+1)\,\mathbb{P}(X=k) < \infty$

$$\iff \underbrace{\sum_{k=r}^{\infty} k(k-1)\ldots(k-r+1)\,\mathbb{P}(X=k)\,t^{k-r}}_{=\,g^{(r)}(t)} \text{ konv. für } t=1$$

$\iff$ b) (mit Abelschem Grenzwertsatz)

Memo: $\mathbb{E}\left[X(X-1)\cdot\ldots\cdot(X-r+1)\right] \;=\; g^{(r)}(1-)$.

$$\begin{aligned}
\mathbb{E}\,X &= g'(1-) \\
\mathbb{E}\,X(X-1) &= g''(1-) \qquad \Longrightarrow \\
\mathbb{V}(X) &= \mathbb{E}X^2 - (\mathbb{E}\,X)^2 \;=\; \mathbb{E}\,X(X-1) + \mathbb{E}\,X - (\mathbb{E}\,X)^2 \\
&= g''(1-) + g'(1-) - g'(1-)^2 \\
\mathbb{E}\,X(X-1)(X-2) &= g'''(1-) \qquad \text{usw.}
\end{aligned}$$

17.11 Beispiel (Poisson-Verteilung)

Sei $X \sim \mathrm{Po}(\lambda)$. Es gilt $g(t) = \exp(\lambda(t-1)), \quad t \in \mathbb{R}$.

$$g^{(r)}(t) \;=\; \frac{\mathrm{d}^r}{\mathrm{d}t^r} g(t) \;=\; \lambda^r \exp(\lambda(t-1)), \quad t \in \mathbb{R}.$$

$$\Longrightarrow \mathbb{E}\left[X(X-1)\ldots(X-r+1)\right] \;=\; g^{(r)}(1) \;=\; \lambda^r, \quad r \in \mathbb{N}.$$

$$\mathbb{E}\,X = \lambda, \quad \mathbb{V}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda, \quad \mathbb{E}\left[X(X-1)(X-2)\right] = \lambda^3$$

Memo: $g_X(t) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)\, t^k$

17.12 Beispiel (Ein Ding der Unmöglichkeit)

2 echte Würfel werfen $\Longrightarrow$ Augensumme hat Dreiecksverteilung auf $2, \ldots, 12$.

2 gefälschte Würfel werfen.

Dabei seien $p_1, \ldots, p_6,\ q_1, \ldots, q_6$ die W'en für die einzelnen Augenzahlen.

Kann für die Augensumme eine Gleichverteilung auf $2, \ldots, 12$ entstehen?

Sei X (bzw. Y) das Ergebnis des ersten (bzw. zweiten) Würfels. Es soll gelten:

$$\begin{aligned} g_{X+Y}(t) &= \frac{1}{11}\left(t^2 + t^3 + \cdots + t^{12}\right) = \frac{t^2}{11} \sum_{k=0}^{10} t^k \\ &= \frac{t^2}{11} \cdot \frac{t^{11} - 1}{t - 1}, \qquad t \neq 1. \end{aligned}$$

Memo: $g_{X+Y}(t) \;=\; \frac{t^2}{11} \cdot \frac{t^{11}-1}{t-1}, \qquad t \neq 1.$

Weiter gelten $\quad g_X(t) = \sum_{j=1}^{6} p_j\, t^j, \qquad g_Y(t) = \sum_{j=1}^{6} q_j\, t^j,$

wobei

$p_1 > 0,\; q_1 > 0,\; p_6 > 0,\; q_6 > 0$ (sonst Augensummen 2 und 12 unmöglich)

$$g_X(t) = t \sum_{k=0}^{5} p_{k+1} t^k =: t \cdot Q(t), \quad g_Y(t) = t \sum_{k=0}^{5} q_{k+1} t^k =: t \cdot R(t), \quad t \in \mathbb{R},$$

wobei Q, R Polynome vom Grad 5.

Beachte: $Q(0) = p_1 \neq 0, \quad Q(1) = 1 \neq 0, \quad R(0) = q_1 \neq 0, \quad R(1) = 1 \neq 0,$

$$\Longrightarrow\; Q(t)\, R(t) \;=\; \frac{1}{11} \cdot \frac{t^{11}-1}{t-1}, \qquad t \in \mathbb{R}\setminus\{0,1\}.$$

Also haben weder Q noch R eine reelle Nullstelle. Widerspruch!

17.13 Bemerkung (Unendlich viele unabhängige Zufallsvariablen)

Im Folgenden betrachten wir häufig unendliche viele stochastisch unabhängige Zufallsvariablen $Y_1, Y_2, \ldots$ auf einem W-Raum.

$$Y_1, Y_2, \ldots \text{ unabhängig } \quad :\Longleftrightarrow \quad Y_1, \ldots, Y_n \text{ unabhängig } \; \forall \, n \geq 2$$

Unendlich viele unabhängige Zufallsvariablen existieren im Allgemeinen – insbesondere, wenn sie die gleiche Verteilung besitzen sollen – nicht über *diskreten* Wahrscheinlichkeitsräumen (genaueres in Kapitel 24).

17.14 Randomisierte Summen

Seien $N, X_1, X_2, \ldots$ unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen.

Sei $S_0 := 0$, $S_n := X_1 + \ldots + X_n$, $n \geq 1$. Sei

$$S_N \;:=\; \sum_{j=1}^{N} X_j, \text{ also } S_N(\omega) \;:=\; \sum_{j=1}^{N(\omega)} X_j(\omega), \qquad \omega \in \Omega.$$

S_N heißt randomisierte Summe.

Annahmen:

- $X_1, X_2, \ldots$ sind identisch verteilt mit erzeugender Funktion $g(t) = \mathbb{E}t^{X_1}$.
- N besitzt die erzeugende Funktion $\varphi(t) = \mathbb{E}t^N$.

17.15 Satz (Erzeugende Funktion von S_N)

Die erzeugende Funktion von S_N ist

$$g_{S_N}(t) \;=\; \varphi(g(t)), \qquad |t| \leq 1.$$

Memo: $\varphi(t) = \mathbb{E}t^N, \; g(t) = \mathbb{E}t^{X_1}, \; S_N = \sum_{j=1}^{N} X_j, \; S_n = \sum_{j=1}^{n} X_j$

Memo: z.z.: $g_{S_N}(t) \; \left(= \sum_{k=0}^{\infty} \mathbb{P}(S_N = k)\, t^k \right) \; = \; \varphi(g(t))$

Beachte: N und $X_1 + \ldots + X_n$ sind unabhängig!

$$\begin{aligned}
\mathbb{P}(S_N = k) \; &= \; \sum_{n=0}^{\infty} \mathbb{P}(S_N = k, N = n) \; = \; \sum_{n=0}^{\infty} \mathbb{P}(S_n = k, N = n) \\
&= \; \sum_{n=0}^{\infty} \mathbb{P}(S_n = k)\, \mathbb{P}(N = n) \quad \text{(Blockungslemma)} \\
\Longrightarrow \; g_{S_N}(t) \; &= \; \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \mathbb{P}(S_n = k)\, \mathbb{P}(N = n) \right) t^k \\
&= \; \sum_{n=0}^{\infty} \mathbb{P}(N = n) \left(\sum_{k=0}^{\infty} \mathbb{P}(S_n = k)\, t^k \right) \\
&= \; \sum_{n=0}^{\infty} \mathbb{P}(N = n)\, g_{S_n}(t) \; = \; \sum_{n=0}^{\infty} \mathbb{P}(N = n)\, g(t)^n \; = \; \varphi(g(t))
\end{aligned}$$

Memo: $\varphi(t) = \mathbb{E}t^N = \exp(\lambda(t-1))$, $g(t) = \mathbb{E}t^{X_1} = 1-p+pt$

17.16 Beispiel (Radioaktives Präparat)

Sei N die Anzahl der in einem Zeitintervall emittierten Teilchen.

Modellannahme: $N \sim \mathrm{Po}(\lambda)$

Messgerät registriert jedes Teilchen (unabhängig voneinander) mit W' p.

Sei $X_j := \mathbf{1}\{„j\text{-tes Teilchen registriert“}\}, \ \ j \geq 1.$

$$S_N := \sum_{j=1}^{N} X_j \qquad \text{(Anzahl der registrierten Teilchen)}$$

Ann.: $N, X_1, X_2, \ldots$ unabhängig, $X_j \sim \mathrm{Bin}(1,p)$. Satz 17.15 $\Longrightarrow$

$$\begin{aligned} g_{S_N}(t) &= \varphi(g(t)) = \exp(\lambda(g(t)-1)) = \exp(\lambda(1-p+pt-1)) \\ &= \exp(\lambda p(t-1)). \end{aligned}$$

Satz 17.5 $\Longrightarrow$ $S_N \sim \mathrm{Po}(\lambda p)$.

17.17 Beispiel (Der Galton–Watson-Verzweigungsprozess)

- Population von Individuen,
- Lebensdauer jeweils eine Zeiteinheit,
- ungeschlechtliche Vermehrung,
- Individuen einer Generation kommen simultan zur Welt,
- Individuen einer Generation sterben gleichzeitig,
- Vermehrung unabhängig voneinander nach gleicher Reproduktionsverteilung.

Reproduktionsverteilung $(p_k)_{k\geq 0}$: $p_k :=$ Wahrsch.' für k Nachkommen.

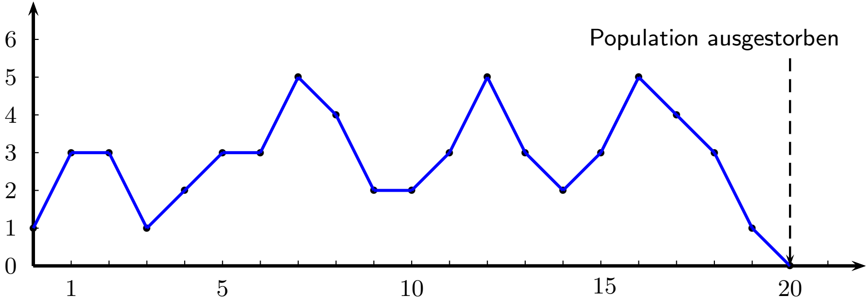
Sei M_n der Umfang der Population zur Zeit $n \geq 1$, $M_0 := 1$.

Reproduktionsgleichung: $$M_{n+1} = \sum_{j=1}^{M_n} X_{n+1}^{(j)}, \quad n \geq 0.$$ Dabei:

- $X_{n+1}^{(j)}$: Nachkommen-Anzahl des j-ten Individuums in n-ter Generation
- $\{X_n^{(j)} : n, j \in \mathbb{N}\}$ unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen mit erzeugender Funktion g

Memo: $M_{n+1} = \sum_{j=1}^{M_n} X_{n+1}^{(j)}$

Die durch die Reproduktionsgleichung rekursiv definierte Folge $(M_n)_{n\geq 0}$ heißt (einfacher) Galton–Watson-Prozess (kurz: GW-Prozess).



Realisierung eines GW-Prozesses zur Reproduktionsverteilung Bin$(3, 1/3)$

Memo: $M_{n+1} = \sum_{j=1}^{M_n} X_{n+1}^{(j)}, \quad n \geq 0; \quad g(t) = \mathbb{E}\left(t^{X_{n+1}^{(j)}}\right)$

Sei

$$\varphi_n(t) := \mathbb{E}t^{M_n}, \quad |t| \leq 1 \qquad \text{(erzeugende Funktion von } M_n\text{)}$$

Satz 17.15 $\Longrightarrow$

$$\varphi_{n+1}(t) = \varphi_n(g(t)); \qquad \varphi_1 = g \Longrightarrow$$

$$\varphi_n(t) = (g \circ \cdots \circ g)(t) \qquad (n\text{-fach Iterierte von } g)$$

Sei $w := \mathbb{P}\left(\bigcup_{n=1}^{\infty} \{M_n = 0\}\right)$ (Aussterbewahrscheinlichkeit)

w hängt nur von der Reproduktionsverteilung $(p_k)_{k \geq 0}$ ab.

Es gilt $\{M_k = 0\} \subseteq \{M_{k+1} = 0\}$ für jedes $k \geq 1$. $\mathbb{P}$ stetig von unten $\Longrightarrow$

$$w = \lim_{n \to \infty} \mathbb{P}(M_n = 0) = \lim_{n \to \infty} \varphi_n(0).$$

Sei $\mu := \sum_{k=1}^{\infty} k\, p_k = g'(1) < \infty$ (Erwartungswert der Reproduktionsvert.)

17.18 Satz (Aussterbewahrscheinlichkeit eines GW-Prozesses)

a) w ist die kleinste nichtnegative Lösung der Fixpunktgleichung $g(t) = t$.

b) Es gelte $p_1 < 1$ (sonst trivialerweise $w = 0$). Dann gilt: $\mu \le 1 \Rightarrow w = 1$.

c) Im Fall $\mu > 1$ gilt $w < 1$.

BEWEIS: a)

$$\begin{aligned} g(w) &= g\left(\lim_{n\to\infty} \varphi_n(0)\right) \\ &= \lim_{n\to\infty} g(\varphi_n(0)) && (g \text{ ist stetig}) \\ &= \lim_{n\to\infty} \varphi_{n+1}(0) && (\varphi_n = g \circ \cdots \circ g \ \ (n\text{-fach Iterierte})) \\ &= w. && (w \text{ ist Fixpunkt von } g) \end{aligned}$$

Sei $x \ge 0$ und $x = g(x) \implies$

$$x = g(x) \ge g(0) = \varphi_1(0) \implies x = g(x) = g(g(x)) \ge g(g(0)) = \varphi_2(0)$$

Induktiv: $x \ge \varphi_n(0)$, $n \in \mathbb{N}$, also $x \ge \lim_{n\to\infty} \varphi_n(0) = w$. $\surd$

Memo: a) w ist kleinste nichtnegative Lösung der Gleichung $g(t) = t$.

Memo: b) Es gelte $p_1 < 1$ (sonst $w = 0$). Dann gilt: $\mu \leq 1 \Rightarrow w = 1$.

Memo: c) Im Fall $\mu > 1$ gilt $w < 1$.

Zu b): Fallunterscheidung: 1) $p_0 + p_1 = 1$ 2) $p_0 + p_1 < 1$.

1) $\mathbb{P}(M_n = 0) = 1 - p_1^n \implies w = \lim_{n\to\infty} \mathbb{P}(M_n = 0) = 1 \ \surd$.

2) Wird in c) behandelt.

Zu c:) Falls $\mu = g'(1) = \sum_{k=1}^{\infty} k\, p_k > 1$, so gilt $p_0 + p_1 < 1$

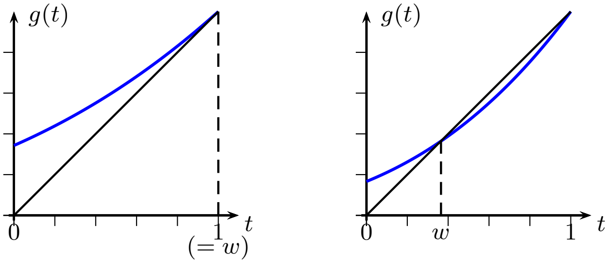
$\implies g'(t) = \sum_{k=1}^{\infty} kp_k t^{k-1}$ auf $[0,1]$ streng monoton wachsend

d.h. g strikt konvex auf $[0,1]$.

g kann dann (neben $t = 1$) höchstens einen weiteren Fixpunkt auf $[0,1]$ haben.

Falls $\mu = g'(1) \leq 1$ (Fall 2)), so $w = 1$.

Falls $\mu = g'(1) > 1$, so $w < 1$.



17.19 Beispiel (Geometrische Reproduktionsverteilung)

Sei für $\mu > 0$

$$p_k \ := \ \frac{1}{\mu+1} \cdot \left(\frac{\mu}{\mu+1} \right)^k, \qquad k \in \mathbb{N}_0.$$

Der Erwartungswert dieser Reproduktionsverteilung ist μ.

Erzeugende Funktion: $g(t) = \dfrac{1}{\mu+1-\mu t}, \qquad |t| \leq 1.$

$g(t) = t \iff \mu t^2 - (\mu+1)t + 1 = 0$

Für $\mu > 1$ gibt es neben $t = 1$ die zweite Lösung $1/\mu < 1$ in $[0,1]$.

Es gilt $w = \frac{1}{\max(1,\mu)}$.

18 Grenzwertsätze

In diesem Kapitel behandeln wir

- das Schwache Gesetz großer Zahlen

sowie

- Zentrale Grenzwertsätze

im Zusammenhang mit diskreten Verteilungen.

Diese Resultate sind grundlegende Ergebnisse der klassischen Wahrscheinlichkeitstheorie mit vielfältigen Anwendungen, auch in der Statistik.

Beide Resultate beziehen sich auf Summen unabhängiger Zufallsvariablen.

In der Folge schreiben wir kurz $\overline{X}_n := \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} X_j$.

18.1 Satz (Schwaches Gesetz der großen Zahlen)

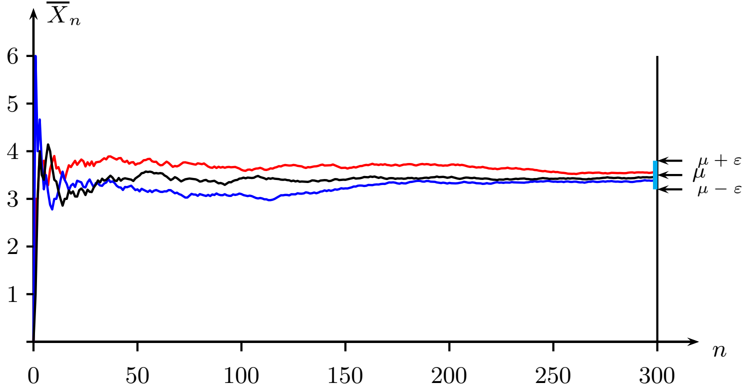
Seien $X_1, X_2, \ldots, X_n, \ldots$ unabhängige Zufallsvariablen mit gleichem Erwartungswert $\mu := \mathbb{E}X_1$ und gleicher Varianz $\sigma^2 := \mathbb{V}(X_1) < \infty$. Dann gilt:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) = 0 \quad \text{für jedes} \quad \varepsilon > 0.$$

Beweis: Sei $\varepsilon > 0$ beliebig. Es gilt $\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{j=1}^n \mathbb{E}(X_j) = \mu$.

$$\begin{aligned}
\mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) &\leq \frac{\mathbb{V}(\overline{X}_n)}{\varepsilon^2} && \text{(Tschebyschow-Ungleichung)} \\
&= \frac{1}{\varepsilon^2}\,\mathbb{V}\left(\frac{1}{n}\sum_{j=1}^n X_j\right) \\
&= \frac{1}{\varepsilon^2}\,\frac{1}{n^2}\,\mathbb{V}\left(\sum_{j=1}^n X_j\right) && (\mathbb{V}(aY) = a^2\mathbb{V}(Y)) \\
&= \frac{1}{\varepsilon^2 n^2}\sum_{j=1}^n \mathbb{V}(X_j) && (X_1, \ldots, X_n \text{ unabh.}) \\
&= \frac{1}{\varepsilon^2 n^2}\, n\,\sigma^2 \to 0 \text{ bei } n \to \infty.
\end{aligned}$$

Das schwache Gesetz großer Zahlen stellt einen Zusammenhang zwischen einem zufälligen arithmetischen Mittel und dem Erwartungswert her, vgl. das empirische Gesetz über die Stabilisierung relativer Häufigkeiten



Simulierte arithmetische Mittel der Augensumme beim Würfelwurf

18.2 Folgerung (Schwaches Gesetz großer Zahlen von Jakob Bernoulli)

Es seien $A_1, \ldots, A_n$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p. Dann gilt:

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{A_j\} - p \right| \geq \varepsilon \right) \; = \; 0 \qquad \text{für jedes } \varepsilon > 0 \; .$$

(Hauptergebnis der Ars conjectandi (1713) von Jacob Bernoulli)

$$T_n \; := \; \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\{A_j\} \text{ ist zufällige relative Trefferhäufigkeit.}$$

Zu jedem $\varepsilon > 0$ und zu jedem $\eta > 0$ existiert ein $n_0 = n_0(\varepsilon, \eta) \in \mathbb{N}$, so dass für jedes feste $n \geq n_0$ gilt:

$$\mathbb{P}(|T_n - p| < \varepsilon) \; \geq \; 1 - \eta.$$

In einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit p ist die zufällige relative Trefferhäufgkeit ein sinnvoller Schätzer für p.

18.3 Definition (stochastische Konvergenz)

Es seien $Y, Y_1, Y_2, \ldots$ Zufallsvariablen auf einem gemeinsamen W-Raum.

$$Y_n \xrightarrow{\mathbb{P}} Y \;:\Longleftrightarrow\; \lim_{n\to\infty} \mathbb{P}\left(|Y_n - Y| \geq \varepsilon\right) \;=\; 0 \;\text{ für jedes } \varepsilon > 0$$

($(Y_n)_{n\geq 1}$ konvergiert stochastisch gegen Y)

Gilt $\mathbb{P}(Y = a) = 1$ für ein $a \in \mathbb{R}$, so schreibt man hierfür auch

$$Y_n \xrightarrow{\mathbb{P}} a \;\; (\text{bei } n \to \infty).$$

18.1 und 18.2 besagen also, dass unter den jeweiligen Voraussetzungen

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu \qquad \text{und} \qquad T_n \xrightarrow{\mathbb{P}} p$$

gelten.

Für die stochastische Konvergenz gelten analoge Rechenregeln wie bei der Konvergenz von Zahlenfolgen, z.B.:

18.4 Satz (Rechenregel für stochastische Konvergenz)

Es gelten $X_n \xrightarrow{\mathbb{P}} X$ und $Y_n \xrightarrow{\mathbb{P}} Y$. Dann folgt

$$aX_n + bY_n \xrightarrow{\mathbb{P}} aX + bY, \qquad a, b \in \mathbb{R}.$$

Beweis: Seien o.B.d.A. $a \neq 0$ und $b \neq 0$. Sei $\varepsilon > 0$ beliebig. Zu zeigen ist

$$\lim_{n\to\infty} \mathbb{P}\left(|aX_n + bY_n - (aX + bY)| \geq \varepsilon\right) \;=\; 0.$$

Δ-Ungl. $\Longrightarrow |aX_n + bY_n - (aX + bY)| \leq |a| \cdot |X_n - X| + |b| \cdot |Y_n - Y|$.

Es folgt

$$\{|aX_n + bY_n - (aX + bY)| \geq \varepsilon\} \subseteq \left\{|a| \cdot |X_n - X| \geq \frac{\varepsilon}{2}\right\} \cup \left\{|b| \cdot |Y_n - Y| \geq \frac{\varepsilon}{2}\right\}$$

$\Longrightarrow$

$$\mathbb{P}(|aX_n + bY_n - (aX + bY)| \geq \varepsilon) \leq \underbrace{\mathbb{P}\left(|X_n - X| \geq \frac{\varepsilon}{2|a|}\right)}_{\to 0} + \underbrace{\mathbb{P}\left(|Y_n - Y| \geq \frac{\varepsilon}{2|b|}\right)}_{\to 0}$$

18.5 Bemerkungen (Stochastische Konvergenz und Erwartungswerte)

a) Falls $\mathbb{E}\, Y_n \to a$ und $\mathbb{V}(Y_n) \to 0$, so folgt $Y_n \xrightarrow{\mathbb{P}} a$ (Übungsaufgabe)

b) Aus $Y_n \xrightarrow{\mathbb{P}} a$ folgt nicht unbedingt $\mathbb{E}\, Y_n \to a$.

Gegenbeispiel: $\mathbb{P}(Y_n = a) := 1 - \dfrac{1}{n}, \quad \mathbb{P}(Y_n = n^2) := \dfrac{1}{n} \Longrightarrow$

$$\mathbb{P}(|Y_n - a| \geq \varepsilon) \ \leq\ \mathbb{P}(Y_n = n^2) \to 0 \ \Longrightarrow Y_n \xrightarrow{\mathbb{P}} a$$

$$\mathbb{E}\, Y_n = a \cdot \left(1 - \frac{1}{n}\right) + n^2 \cdot \frac{1}{n} \to \infty$$

c) Es gebe ein $K \in [0, \infty)$ mit $|Y_n| \leq K \ \ \forall\, n$. Dann gilt (Übungsaufgabe)

$$Y_n \xrightarrow{\mathbb{P}} a \ \Longrightarrow\ \mathbb{E}\, Y_n \to a.$$

Memo: $Y_n \xrightarrow{\mathbb{P}} a \;:\Longleftrightarrow\; \lim_{n\to\infty} \mathbb{P}(|Y_n - a| \geq \varepsilon) = 0 \;\forall\, \varepsilon > 0$

18.6 Satz (Stochastische Konvergenz und stetige Abbildungen)

Es gelte $Y_n \xrightarrow{\mathbb{P}} a$ bei $n \to \infty$. Die Funktion $g : \mathbb{R} \to \mathbb{R}$ sei an der Stelle a stetig. Dann folgt

$$g(Y_n) \;\xrightarrow{\mathbb{P}}\; g(a).$$

BEWEIS: Sei $\varepsilon > 0$ beliebig. Es gibt es ein von a und ε abhängendes $\delta > 0$ mit

$$\forall\, y \in \mathbb{R} : \; |y - a| < \delta \implies |g(y) - g(a)| < \varepsilon. \qquad (18.2)$$

Zu zeigen ist $\lim_{n\to\infty} \mathbb{P}(|g(Y_n) - g(a)| \geq \varepsilon) = 0$. Es ist

$$\begin{aligned} \mathbb{P}(|g(Y_n) - g(a)| \geq \varepsilon) \;&=\; \mathbb{P}(|g(Y_n) - g(a)| \geq \varepsilon, |Y_n - a| \geq \delta) \\ &\quad + \mathbb{P}(|g(Y_n) - g(a)| \geq \varepsilon, |Y_n - a| < \delta) \\ &\leq\; \mathbb{P}(|Y_n - a| \geq \delta) \,+\, 0 \end{aligned}$$

(nach (18.2) gilt $\{|Y_n - a| < \delta\} \cap \{|g(Y_n) - g(a)| \geq \varepsilon\} = \emptyset$) $\checkmark$

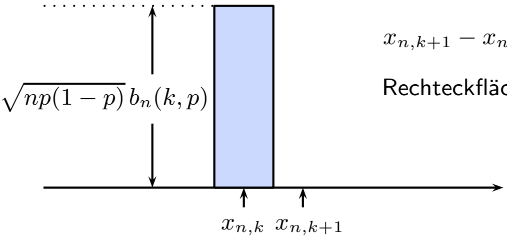
Sei $S_n \sim \mathrm{Bin}(n,p)$, $0 < p < 1$.

$$\widetilde{S}_n := \frac{S_n - np}{\sqrt{np(1-p)}}, \qquad x_{n,k} := \frac{k - np}{\sqrt{np(1-p)}}, \qquad k \in \mathbb{Z}.$$

$\widetilde{S}_n$ nimmt den Wert $x_{n,k}$ mit der Wahrscheinlichkeit

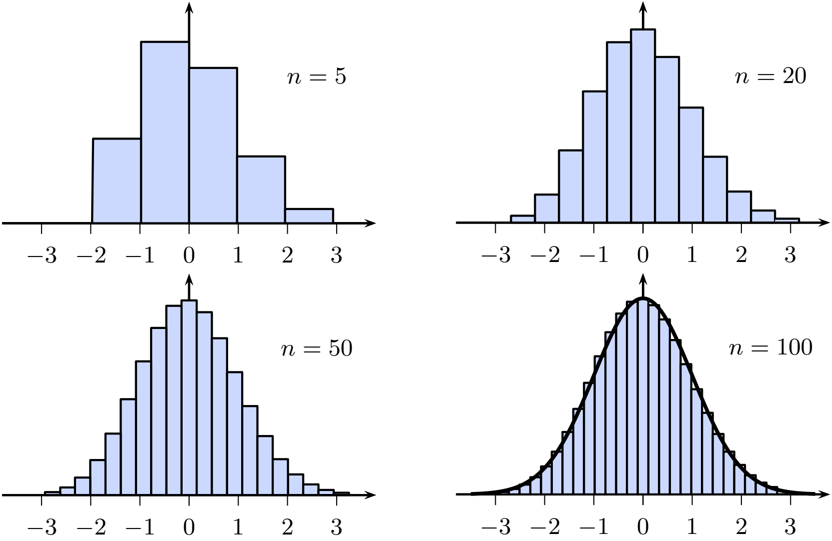
$$b_n(k,p) \;:=\; \binom{n}{k} p^k \, (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

an.



$$x_{n,k+1} - x_{n,k} \;=\; \frac{1}{\sqrt{np(1-p)}}$$

Rechteckfläche $= b_n(k,p)$



Histogramme standardisierter Binomialverteilungen für $p = 0.3$

18.7 Definition (Dichte der Standard-Normalverteilung)

Die durch

$$\varphi(x) \;:=\; \frac{1}{\sqrt{2\pi}}\,\exp\left(-\frac{x^2}{2}\right)$$

definierte Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ heißt Dichte der Standard-Normalverteilung oder auch Gauß'sche Glockenkurve. Es gilt

$$\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x \;=\; 1.$$

Nachweis von $\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$:

Beachte:

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = 1 \\ \Longleftrightarrow\quad & \int_{0}^{\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \sqrt{\frac{\pi}{2}} \\ \Longleftrightarrow\quad & \frac{\pi}{2} \;=\; \left(\int_{0}^{\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x\right)^2 = \int_{0}^{\infty}\int_{0}^{\infty} \exp\left(-\frac{x^2+y^2}{2}\right) \mathrm{d}x\mathrm{d}y \end{aligned}$$

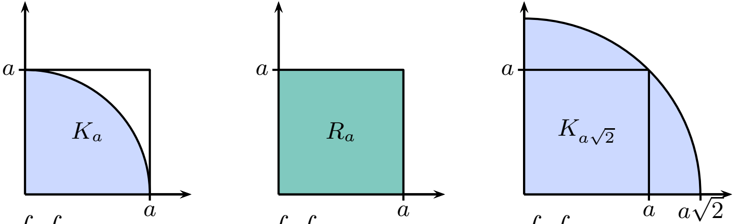
Sei

$$g(x,y) := \exp\left(-\frac{x^2+y^2}{2}\right)$$

Zu zeigen:

$$\lim_{a\to\infty} \int_{0}^{a}\int_{0}^{a} g(x,y)\,\mathrm{d}x\mathrm{d}y \;=\; \frac{\pi}{2}.$$

Memo: $g(x,y) := \exp\left(-\frac{x^2+y^2}{2}\right)$, z.z.: $\lim_{a\to\infty} \int_0^a \int_0^a g(x,y)\,\mathrm{d}x\mathrm{d}y \;=\; \frac{\pi}{2}$.

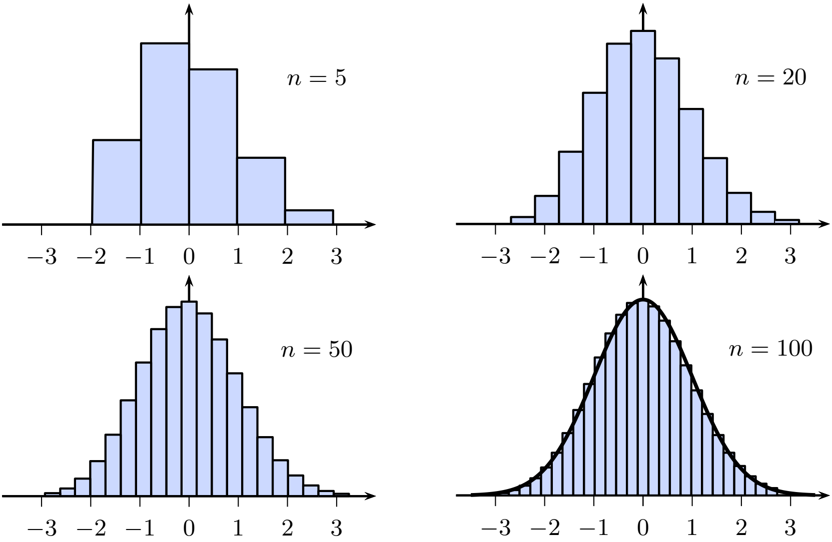


$$\int\int_{K_a} g(x,y)\,\mathrm{d}x\mathrm{d}y \;\le\; \int\int_{R_a} g(x,y)\,\mathrm{d}x\mathrm{d}y \;\le\; \int\int_{K_{a\sqrt{2}}} g(x,y)\,\mathrm{d}x\mathrm{d}y$$

Polarkoordinaten: $x = r\cos\alpha,\ y = r\sin\alpha \implies x^2+y^2 = r^2,\ \mathrm{d}x\mathrm{d}y = r\,\mathrm{d}r\mathrm{d}\alpha$

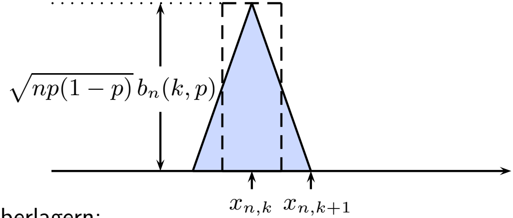
$$\begin{aligned}\int\int_{K_a} g(x,y)\,\mathrm{d}x\mathrm{d}y &= \int_0^{\pi/2}\left(\int_0^a r\mathrm{e}^{-r^2/2}\,\mathrm{d}r\right)\mathrm{d}\alpha \\ &= \frac{\pi}{2}\left(-\mathrm{e}^{-r^2/2}\right)\Big|_0^a = \frac{\pi}{2}\left(1-\mathrm{e}^{-a^2/2}\right) \overset{a\to\infty}{\to} \frac{\pi}{2}\end{aligned}$$

$$\int\int_{K_{a\sqrt{2}}} g(x,y)\,\mathrm{d}x\mathrm{d}y = \cdots = \frac{\pi}{2}\left(1-\mathrm{e}^{-a^2}\right) \overset{a\to\infty}{\to} \frac{\pi}{2} \;\checkmark$$

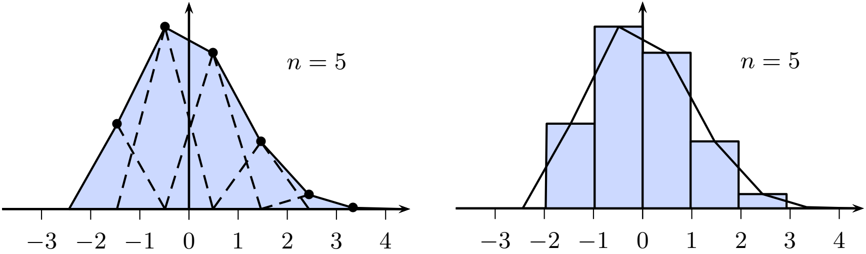


Histogramme standardisierter Binomialverteilungen für $p = 0.3$

Man könnte die Punktmasse $b_n(k,p)$ in $x_{n,k}$ auch mit Hilfe einer „Dach-Funktion“ zu einer „Dreiecksfläche verschmieren“



und überlagern:



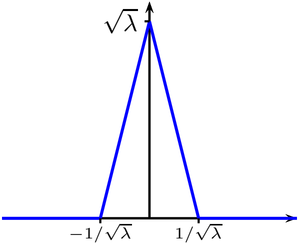
Wir versuchen diese Überlagerung auch bei der Poisson-Verteilung Po(λ):

$$\text{Sei } X_\lambda \sim \mathrm{Po}(\lambda), \quad \widetilde{X}_\lambda := \frac{X_\lambda - \lambda}{\sqrt{\lambda}}, \quad x_{\lambda,k} := \frac{k-\lambda}{\sqrt{\lambda}}, \quad p_\lambda(k) := \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}.$$

$\widetilde{X}_\lambda$ nimmt den Wert $x_{\lambda,k}$ mit Wahrscheinlichkeit $p_\lambda(k)$ an, $k \in \mathbb{N}_0$.

Betrachte *Dach-Funktion*

$$g_\lambda(x) \; := \; \sqrt{\lambda}\left(1 - \sqrt{\lambda}|x|\right) \mathbf{1}_{[-1/\sqrt{\lambda},1/\sqrt{\lambda}]}(x)$$



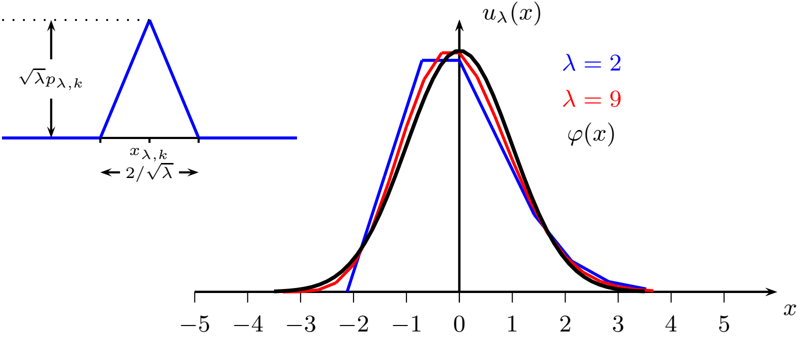
$$\int_{-\infty}^{\infty} g_\lambda(x)\,\mathrm{d}x = 1, \quad \int_{-\infty}^{\infty} x\, g_\lambda(x)\,\mathrm{d}x = 0,$$

$$\int_{-\infty}^{\infty} x^2\, g_\lambda(x)\,\mathrm{d}x = \frac{1}{6\lambda},$$

$$\int_{-\infty}^{\infty} x^2\, g_\lambda(x - x_{\lambda,k})\,\mathrm{d}x = \frac{1}{6\lambda} + x_{\lambda,k}^2.$$

Wir spannen die Funktion g_λ an den Stellen $x_{\lambda,k}$, $k \geq 0$, auf, gewichten mit $p_\lambda(k)$ und überlagern:

$$u_\lambda(x) \;:=\; \sum_{k=0}^{\infty} p_\lambda(k)\, g_\lambda(x - x_{\lambda,k}), \qquad x \in \mathbb{R}.$$



Konvergiert u_λ für $\lambda \to \infty$ gegen φ?

Memo: $u_\lambda(x) = \sum_{k=0}^{\infty} p_\lambda(k)\, g_\lambda(x - x_{\lambda,k})$

Memo: $\int g_\lambda(x)\,\mathrm{d}x = 1, \qquad \int x^2 g_\lambda(x - x_{\lambda,k})\,\mathrm{d}x = \frac{1}{6\lambda} + x_{\lambda,k}^2$

Beachte: $\lambda \to \infty$ heißt $\lambda_n \to \infty$ bei $n \to \infty$ für eine beliebige Folge (λ_n).

18.8 Hilfssatz Für die Funktion u_λ gelten:

a) $\int_{-\infty}^{\infty} u_\lambda(x)\,\mathrm{d}x \;=\; 1,$

b) $\int_{-\infty}^{\infty} x^2\, u_\lambda(x)\,\mathrm{d}x \;=\; 1 + \frac{1}{6\lambda},$

c) $\lim_{\lambda \to \infty} u_\lambda(x) \;=\; \varphi(x), \quad x \in \mathbb{R}.$

Dabei ist die Konvergenz gleichmäßig auf jedem kompakten Intervall.

BEWEIS: a) und b) folgen durch direkte Rechnung aus obigen Memos.

c): siehe Handout.

Beweis von Hilfssatz 18.8 c:
Die Funktion u_λ ist auf $\mathbb{R} \setminus \{x_{\lambda,k} : k \geq -1\}$ differenzierbar.

Ergänze Ableitung u'_λ an den „Knick-Stellen“ $x_{\lambda,k}$ durch rechtsseitig stetige Fortsetzung sowie auf $(-\infty, x_{\lambda,-1})$ durch $\frac{0}{0} := 0$. Zeige zunächst

$$\lim_{\lambda\to\infty} \frac{u'_\lambda(x)}{u_\lambda(x)} \;=\; -x, \qquad x \in \mathbb{R}, \tag{18.3}$$

wobei die Konvergenz auf jedem kompakten Intervall gleichmäßig ist.

Damit folgt für $t > 0$ ($t < 0$ analog)

$$\lim_{\lambda\to\infty} \int_0^t \frac{u'_\lambda(x)}{u_\lambda(x)}\,\mathrm{d}x \;=\; \int_0^t (-x)\,\mathrm{d}x \;=\; -\frac{1}{2}t^2.$$

Andererseits: $$\int_0^t \frac{u'_\lambda(x)}{u_\lambda(x)}\,\mathrm{d}x \;=\; \int_0^t \frac{d}{\mathrm{d}x} \log u_\lambda(x)\,\mathrm{d}x \;=\; \log\frac{u_\lambda(t)}{u_\lambda(0)}$$

$$\log(\cdot) \text{ stetig } \Longrightarrow \lim_{\lambda\to\infty} \frac{u_\lambda(t)}{u_\lambda(0)} \;=\; \exp\left(-\frac{1}{2}t^2\right), \tag{18.4}$$

wobei die Konvergenz auf kompakten Intervallen gleichmäßig ist.

Beachte: Wegen (18.4), gleichmäßig auf kompakten Intervallen, gilt

$$\begin{aligned}\lim_{a\to\infty}\left(\lim_{\lambda\to\infty}\int_{-a}^{a}\frac{u_\lambda(x)}{u_\lambda(0)}\,\mathrm{d}x\right) &= \lim_{a\to\infty}\int_{-a}^{a}\left(\lim_{\lambda\to\infty}\frac{u_\lambda(x)}{u_\lambda(0)}\right)\mathrm{d}x\\ &= \lim_{a\to\infty}\int_{-a}^{a}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x \;=\; \sqrt{2\pi}.\end{aligned}$$

Weiter gilt

$$\begin{aligned}1 \;&\geq\; \int_{-a}^{a}u_\lambda(x)\,\mathrm{d}x \;=\; 1-\int_{|x|>a}u_\lambda(x)\,\mathrm{d}x\\ &\geq\; 1-\frac{1}{a^2}\int_{|x|>a}x^2u_\lambda(x)\,\mathrm{d}x \;\geq\; 1-\frac{1}{a^2}\int_{-\infty}^{\infty}x^2u_\lambda(x)\,\mathrm{d}x\\ &=\; 1-\frac{1}{a^2}\left(1+\frac{1}{6\lambda}\right).\end{aligned}$$

Wegen $\lim_{a\to\infty}\left(\lim_{\lambda\to\infty}\frac{1}{u_\lambda(0)}\int_{-a}^{a}u_\lambda(x)\,\mathrm{d}x\right)=\sqrt{2\pi}$

folgt hieraus $\lim_{\lambda\to\infty}u_\lambda(0)=\frac{1}{\sqrt{2\pi}}$, also mit (18.4) Teil c) von Hilfssatz 18.8 .

Memo: Bleibt zu zeigen: $\lim_{\lambda\to\infty} \frac{u'_\lambda(x)}{u_\lambda(x)} = -x, \qquad x \in \mathbb{R}$.

Memo: z.z.: Konvergenz gleichmäßig auf kompakten Intervallen.

Sei $x \in \mathbb{R}$ beliebig. λ genügend groß $\Longrightarrow$ $\exists k = k(x,\lambda)$ mit

$$x_{\lambda,k} \le x < x_{\lambda,k+1} \qquad \left(\Longleftrightarrow \lambda + x\sqrt{\lambda} - 1 < k \le \lambda + x\sqrt{\lambda}\right). \qquad (18.5)$$

Sei $u_{\lambda,k} := u_\lambda(x_{\lambda,k})$, $D_\lambda(x) := u'_\lambda(x)/u_{\lambda,k} \Longrightarrow$

$$\begin{aligned} D_\lambda(x) &= \frac{u_{\lambda,k+1} - u_{\lambda,k}}{x_{\lambda,k+1} - x_{\lambda,k}} \cdot \frac{1}{u_{\lambda,k}} \\ &= \frac{\sqrt{\lambda}\,\mathrm{e}^{-\lambda}}{\frac{1}{\sqrt{\lambda}}\,\sqrt{\lambda}\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}} \cdot \left[\frac{\lambda^{k+1}}{(k+1)!} - \frac{\lambda^k}{k!}\right] = \sqrt{\lambda} \cdot \frac{\lambda - (k+1)}{k+1} \end{aligned}$$

$$\begin{aligned} (18.5) \Longrightarrow D_\lambda(x) &\le \sqrt{\lambda} \cdot \frac{\lambda - (\lambda + x\sqrt{\lambda})}{\lambda + x\sqrt{\lambda}} &&= -x + \frac{x^2}{x + \sqrt{\lambda}}, \\ D_\lambda(x) &\ge \sqrt{\lambda} \cdot \frac{\lambda - (\lambda + x\sqrt{\lambda} + 1)}{\lambda + x\sqrt{\lambda} + 1} &&= -x + \frac{x^2 + \frac{x}{\sqrt{\lambda}} - 1}{\sqrt{\lambda} + x + \frac{1}{\sqrt{\lambda}}}. \end{aligned}$$

Memo: $D_\lambda(x) \le \sqrt{\lambda} \cdot \dfrac{\lambda - (\lambda + x\sqrt{\lambda})}{\lambda + x\sqrt{\lambda}} = -x + \dfrac{x^2}{x + \sqrt{\lambda}}$

Memo: $D_\lambda(x) \ge \sqrt{\lambda} \cdot \dfrac{\lambda - (\lambda + x\sqrt{\lambda} + 1)}{\lambda + x\sqrt{\lambda} + 1} = -x + \dfrac{x^2 + \frac{x}{\sqrt{\lambda}} - 1}{\sqrt{\lambda} + x + \frac{1}{\sqrt{\lambda}}}$

Die zu $-x$ addierten Terme konvergieren auf kompakten Intervallen gleichmäßig gegen Null. Wegen

$$\frac{u'_\lambda(x)}{u_\lambda(x)} = D_\lambda(x) \cdot \frac{u_{\lambda,k}}{u_\lambda(x)}$$

und

$$\begin{aligned}\frac{u_\lambda(x)}{u_{\lambda,k}} &= \frac{1}{u_{\lambda,k}} \left[u_{\lambda,k} + \frac{x - x_{\lambda,k}}{x_{\lambda,k+1} - x_{\lambda,k}} \cdot (u_{\lambda,k+1} - u_{\lambda,k}) \right] \\ &= 1 + D_\lambda(x) \cdot (x - x_{\lambda,k})\end{aligned}$$

ergibt sich nun (18.3) unter Beachtung der Ungleichung $|x - x_{\lambda,k}| \le \frac{1}{\sqrt{\lambda}}$.

18.9 Satz (ZGWS für die Poissonverteilung, lokale Form)

Sei
$$p_\lambda(k) \;:=\; \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},$$
$$f_\lambda(k) \;:=\; \frac{1}{\sqrt{2\pi\lambda}}\exp\left[-\frac{(k-\lambda)^2}{2\lambda}\right] = \frac{1}{\sqrt{\lambda}}\cdot\varphi\left(\frac{k-\lambda}{\sqrt{\lambda}}\right), \; k\in\mathbb{N}_0, \; \lambda>0.$$

Dann gilt: $\forall C>0\,\forall\varepsilon>0\;\exists\lambda_0=\lambda_0(C,\varepsilon)$ mit

$$\forall\lambda\geq\lambda_0\;\forall k\in\mathbb{N}: \quad \left|\frac{k-\lambda}{\sqrt{\lambda}}\right|\leq C \implies \left|\frac{p_\lambda(k)}{f_\lambda(k)}-1\right|<\varepsilon.$$

BEWEIS:

Seien $\varepsilon > 0$, $C > 0$ gegeben, $M := \min_{|x| \le C} \varphi(x) = \varphi(C)$.

Nach Hilfssatz 18.8 c) existiert ein $\lambda_0 = \lambda_0(\varepsilon, C)$ mit

$$\sup_{|x| \le C} |u_\lambda(x) - \varphi(x)| < \varepsilon \varphi(C) \qquad \lambda \ge \lambda_0, \implies$$

$$\sup_{|x| \le C} \left| \frac{u_\lambda(x)}{\varphi(x)} - 1 \right| < \varepsilon \qquad \lambda \ge \lambda_0.$$

Wegen $x_{\lambda,k} = (k - \lambda)/\sqrt{\lambda}$ und

$$\frac{u_\lambda(x_{\lambda,k})}{\varphi(x_{\lambda,k})} = \frac{p_\lambda(k)}{f_\lambda(k)}$$

folgt die Behauptung.

18.10 Satz (ZGWS für die Poisson-Verteilung)

Sei $X_\lambda \sim \mathrm{Po}(\lambda)$. Dann gelten für die standardisierte Zufallsvariable

$$\widetilde{X}_\lambda \; := \; \frac{X_\lambda - \lambda}{\sqrt{\lambda}} :$$

a) $\lim\limits_{\lambda\to\infty} \mathbb{P}\Big(a \le \widetilde{X}_\lambda \le b\Big) \; = \; \int_a^b \varphi(t)\,\mathrm{d}t, \qquad -\infty < a < b < \infty,$

b) $\lim\limits_{\lambda\to\infty} \mathbb{P}\Big(\widetilde{X}_\lambda \le x\Big) \; = \; \int_{-\infty}^x \varphi(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$

BEWEIS: a) Sei $P_\lambda(a,b) := \mathbb{P}\left(a \le \widetilde{X}_\lambda \le b\right)$ sowie $x_{\lambda,k} := (k-\lambda)/\sqrt{\lambda} \Longrightarrow$

$$\begin{aligned} P_\lambda(a,b) \;&=\; \sum_{k\in\mathbb{N}_0 : a \le x_{\lambda,k} \le b} p_\lambda(k)\sqrt{\lambda}(x_{\lambda,k+1} - x_{\lambda,k}) \\ &=\; \sum_{k\in\mathbb{N}_0 : a \le x_{\lambda,k} \le b} u_\lambda(x_{\lambda,k})(x_{\lambda,k+1} - x_{\lambda,k}) \end{aligned}$$

Dreiecksungleichung $\Longrightarrow$

$$\left|P_\lambda(a,b) - \int_a^b \varphi(x)\,\mathsf{d}x\right| \le \left|\sum_{k:a\le x_{\lambda,k}\le b} \left[u_\lambda(x_{\lambda,k}) - \varphi(x_{\lambda,k})\right](x_{\lambda,k+1} - x_{\lambda,k})\right| + \left|\sum_{k:a\le x_{\lambda,k}\le b} \varphi(x_{\lambda,k})(x_{\lambda,k+1} - x_{\lambda,k}) - \int_a^b \varphi(x)\,\mathsf{d}x\right|.$$

2. Term konvergiert bei $\lambda \to \infty$ gegen Null (Riemannsche Näherungssumme!)

1. Term ist höchstens gleich

$$\sup_{a\le t\le b} |u_\lambda(t) - \varphi(t)| \cdot \frac{1}{\sqrt{\lambda}} \cdot |\{k \in \mathbb{N}_0 : a \le x_{\lambda,k} \le b\}|.$$

Beachte:

$|\{k \in \mathbb{N}_0 : a \le x_{\lambda,k} \le b\}| \le 1 + (b-a)\sqrt{\lambda}$,

Hilfssatz 18.8 c) $\Longrightarrow$ 1. Faktor konvergiert gegen Null, q.e.d. a)

b) Für $a < 0$ gilt aufgrund der Tschebyschow-Ungleichung

$$\mathbb{P}(\widetilde{X}_\lambda < a) \ \leq \ \mathbb{P}(|\widetilde{X}_\lambda| > |a|) \ \leq \ \frac{1}{a^2}.$$

Es folgt für $a < x$ und $a < 0$

$$\begin{aligned} \mathbb{P}(a \leq \widetilde{X}_\lambda \leq x) \quad & \leq \quad \mathbb{P}(\widetilde{X}_\lambda \leq x) \ = \ \mathbb{P}(\widetilde{X}_\lambda < a) + \mathbb{P}(a \leq \widetilde{X}_\lambda \leq x) \\ & \leq \quad \frac{1}{a^2} + \mathbb{P}(a \leq \widetilde{X}_\lambda \leq x). \end{aligned}$$

Mit Teil a) ergibt sich

$$\begin{aligned} \int_a^x \varphi(t)\,\mathrm{d}t \quad & \leq \quad \liminf_{\lambda\to\infty} \mathbb{P}(\widetilde{X}_\lambda \leq x) \ \leq \ \limsup_{\lambda\to\infty} \mathbb{P}(\widetilde{X}_\lambda \leq x) \\ & \leq \quad \frac{1}{a^2} + \int_a^x \varphi(t)\,\mathrm{d}t. \end{aligned}$$

Die Behauptung folgt nun beim Grenzübergang $a \to -\infty$ wegen $\lim_{a\to-\infty} \int_{-\infty}^a \varphi(t)\,\mathrm{d}t = 0$.

18.11 Satz (ZGWS für die Binomialverteilung, lokale Form)

Für $n \in \mathbb{N}$, $k \in \mathbb{N}_0$, $0 < p < 1$ sei $x_{n,k} := (k - np)/\sqrt{np(1-p)}$,

$$\begin{aligned} b_n(k,p) &:= \binom{n}{k} p^k (1-p)^{n-k}, \\ f_n(k,p) &:= \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \exp\left(-\frac{x_{n,k}^2}{2}\right) \\ &= \frac{1}{\sqrt{np(1-p)}} \cdot \varphi\left(x_{n,k}\right). \end{aligned}$$

Dann gilt: $\forall C > 0 \; \forall \varepsilon > 0 \; \exists \; R = R(C, \varepsilon)$ mit:

$$\forall\, n \in \mathbb{N} \;\; \forall\, k \in \mathbb{N}_0 : \;\; n \geq R \text{ und } |x_{n,k}| \leq C \Longrightarrow \left| \frac{b_n(k,p)}{f_n(k,p)} - 1 \right| < \varepsilon.$$

BEWEIS: Es sei $\lambda_n := np$, $\mu_n := n(1-p)$. Mit $p_\lambda(k)$ und $f_\lambda(k)$ wie in Satz 18.9 gilt dann

$$b_n(k,p) = \frac{p_{\lambda_n}(k) \cdot p_{\mu_n}(n-k)}{p_{\lambda_n+\mu_n}(n)}, \quad f_n(k,p) = \frac{f_{\lambda_n}(k) \cdot f_{\mu_n}(n-k)}{f_{\lambda_n+\mu_n}(n)}$$

und somit

$$\begin{aligned} \frac{b_n(k,p)}{f_n(k,p)} &= \frac{p_{\lambda_n}(k)}{f_{\lambda_n}(k)} \cdot \frac{p_{\mu_n}(n-k)}{f_{\mu_n}(n-k)} \cdot \left\{ \frac{p_{\lambda_n+\mu_n}(n)}{f_{\lambda_n+\mu_n}(n)} \right\}^{-1} \\ &=: a_n(k) \cdot b_n(k) \cdot c_n^{-1}. \end{aligned}$$

Zu vorgegebenem $\varepsilon > 0$ und $C > 0$ sei $\eta > 0$ so gewählt, dass $2\eta^2 + 6\eta \le \varepsilon$. Nach Satz 18.9 existiert ein $\lambda_0 = \lambda_0(\varepsilon, C)$, so dass gilt:

$$\begin{aligned} |a_n(k) - 1| < \eta, &\quad \text{falls} \quad \lambda_n \ge \lambda_0 \text{ und } \left| \frac{k-np}{\sqrt{np}} \right| \le C\ , \\ |b_n(k) - 1| < \eta, &\quad \text{falls} \quad \mu_n \ge \lambda_0 \text{ und } \left| \frac{n-k-n(1-p)}{\sqrt{n(1-p)}} \right| \le C\ , \\ |c_n - 1| < \eta, &\quad \text{falls} \quad \lambda_n + \mu_n\ (= n) \ge \lambda_0. \end{aligned}$$

O.B.d.A. sei dabei λ_0 so groß gewählt, dass $c_n \geq 1/2$ für $n \geq \lambda_0$.

Setze $R := \lambda_0/(p(1-p))$. Dann folgt aus $n \geq R$

$$\lambda_n \geq \lambda_0, \quad \mu_n \geq \lambda_0, \quad n \geq \lambda_0.$$

Außerdem impliziert $|k - np| \leq C\sqrt{np(1-p)}$ jede der Bedingungen

$$\left|\frac{k-np}{\sqrt{np}}\right| \leq C \text{ und } \left|\frac{n-k-n(1-p)}{\sqrt{n(1-p)}}\right| \leq C.$$

Es folgt somit

$$\begin{aligned}\left|\frac{a_n(k)b_n(k)}{c_n} - 1\right| \;&\leq\; \frac{|a_n(k)-1|\cdot|b_n(k)-1|+|a_n(k)-1|+|b_n(k)-1|+|c_n-1|}{c_n} \\ &\leq\; 2(\eta^2+3\eta) \;\leq\; \varepsilon \;\checkmark\end{aligned}$$

Völlig analog zum Beweis von Satz 18.10 erhält man nun das folgende Resultat:

18.12 Satz (Zentraler Grenzwertsatz von de Moivre–Laplace)

Sei $S_n \sim \text{Bin}(n,p)$, $0 < p < 1$. Dann gelten für die standardisierte Zuf.variable

$$\widetilde{S}_n \; := \; \frac{S_n - np}{\sqrt{np(1-p)}} :$$

a) $$\lim_{n\to\infty} \mathbb{P}\left(a \le \widetilde{S}_n \le b\right) \; = \int_a^b \varphi(t)\,\mathrm{d}t, \qquad -\infty < a < b < \infty,$$

b) $$\lim_{n\to\infty} \mathbb{P}\left(\widetilde{S}_n \le x\right) \; = \int_{-\infty}^x \varphi(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

Bemerkungen:

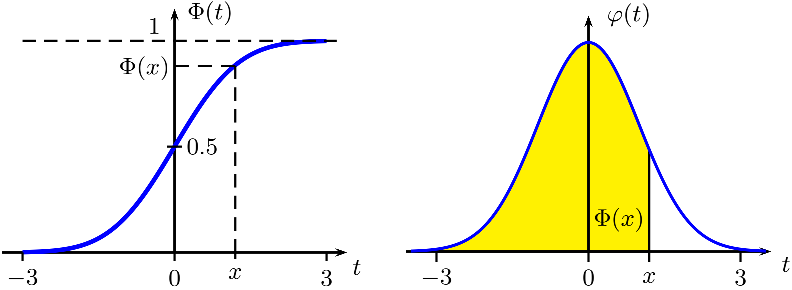
- in a) und b) kann jedes der $\le$-Zeichen durch das $<$-Zeichen ersetzt werden, ohne den jeweiligen Grenzwert zu ändern.
- a) folgt aus b): $\mathbb{P}(a < \widetilde{S}_n \le b) = \mathbb{P}(\widetilde{S}_n \le b) - \mathbb{P}(\widetilde{S}_n \le a)$, $\mathbb{P}(\widetilde{S}_n = a) \to 0$
- Das in b) rechts stehende Integral erhält die folgende eigene Bezeichnung:

18.13 Definition (Verteilungsfunktion der Standard-Normalverteilung)

Die durch

$$\Phi(x) \;:=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t, \qquad x \in \mathbb{R},$$

definierte Funktion $\Phi : \mathbb{R} \to \mathbb{R}$ heißt Verteilungsfunktion der Standard-Normalverteilung.



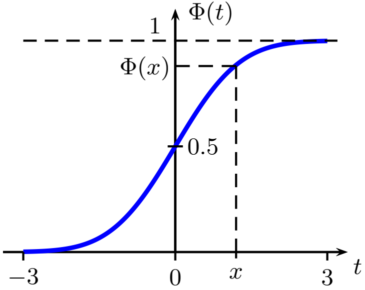
$\Phi(x)$ als Fläche unter der Gaußschen Glockenkurve

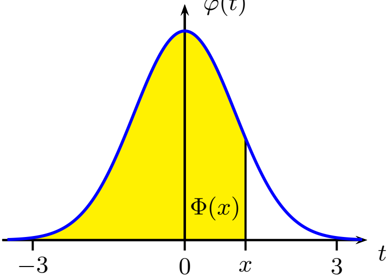
- Es gilt

$$\int_a^b \varphi(t)\,\mathrm{d}t \;=\; \Phi(b) - \Phi(a), \qquad -\infty < a < b < \infty.$$

- Wegen der Symmetrie von φ um 0 ist der Graph von Φ punktsymmetrisch zu $(0, 1/2)$, d.h. es gilt

$$\Phi(-x) \;=\; 1 - \Phi(x), \qquad x \in \mathbb{R}.$$





x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.76	0.7764	1.52	0.9357	2.28	0.9887
0.02	0.5080	0.78	0.7823	1.54	0.9382	2.30	0.9893
0.04	0.5160	0.80	0.7881	1.56	0.9406	2.32	0.9898
0.06	0.5239	0.82	0.7939	1.58	0.9429	2.34	0.9904
0.08	0.5319	0.84	0.7995	1.60	0.9452	2.36	0.9909
0.10	0.5398	0.86	0.8051	1.62	0.9474	2.38	0.9913
0.12	0.5478	0.88	0.8106	1.64	0.9495	2.40	0.9918
0.14	0.5557	0.90	0.8159	1.66	0.9515	2.42	0.9922
0.16	0.5636	0.92	0.8212	1.68	0.9535	2.44	0.9927
0.18	0.5714	0.94	0.8264	1.70	0.9554	2.46	0.9931
0.20	0.5793	0.96	0.8315	1.72	0.9573	2.48	0.9934
0.22	0.5871	0.98	0.8365	1.74	0.9591	2.50	0.9938
0.24	0.5948	1.00	0.8413	1.76	0.9608	2.52	0.9941
0.26	0.6026	1.02	0.8461	1.78	0.9625	2.54	0.9945
0.28	0.6103	1.04	0.8508	1.80	0.9641	2.56	0.9948
0.30	0.6179	1.06	0.8554	1.82	0.9656	2.58	0.9951

Verteilungsfunktion Φ der Standard-Normalverteilung (Auszug)
Für $x < 0$ verwende man die Beziehung $\Phi(-x) = 1 - \Phi(x)$

Praktische Anwendung des ZGWS von de Moivre-Laplace:

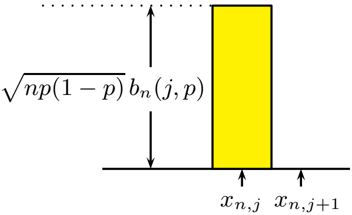
Sei $S_n \sim \mathrm{Bin}(n,p)$, n groß. Seien $k, l \in \mathbb{N}_0$, $0 \le k < l \le n$.

$$\begin{aligned}
\mathbb{P}(k \le S_n \le l) &= \sum_{j=k}^{l} \binom{n}{j} p^j (1-p)^{n-j} \\
&= \mathbb{P}\left(\underbrace{\frac{k-np}{\sqrt{np(1-p)}}}_{=:\, a} \le \underbrace{\frac{S_n-np}{\sqrt{np(1-p)}}}_{=\, \widetilde{S}_n} \le \underbrace{\frac{l-np}{\sqrt{np(1-p)}}}_{=:\, b} \right) \\
&\approx \Phi(b) - \Phi(a)
\end{aligned}$$

Vielfach bessere Näherung (sog. Stetigkeitskorrektur):

$$\begin{aligned}
\mathbb{P}(k \le S_n \le l) &\approx \Phi\left(\frac{l-np+\frac{1}{2}}{\sqrt{np(1-p)}} \right) - \Phi\left(\frac{k-np-\frac{1}{2}}{\sqrt{np(1-p)}} \right) \\
&= \Phi\left(b + \frac{1}{2\sqrt{np(1-p)}} \right) - \Phi\left(a - \frac{1}{2\sqrt{np(1-p)}} \right).
\end{aligned}$$

$$x_{n,j} := \frac{j-np}{\sqrt{np(1-p)}}, \quad b_n(j,p) \;:=\; \binom{n}{j} p^j \, (1-p)^{n-j}$$

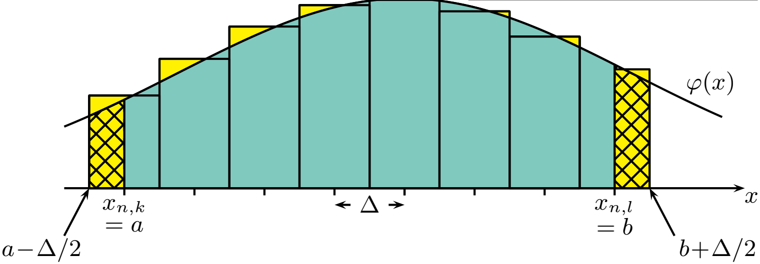


$x_{n,j+1} - x_{n,j} \;=\; \frac{1}{\sqrt{np(1-p)}} =: \Delta$

Rechteckfläche $= b_n(j,p)$

$\mathbb{P}(k \le S_n \le l) = \sum_{j=k}^{l} b_n(j,p)$

$\boxed{\int_a^b \varphi(x)\, \mathrm{d}x = \Phi(b) - \Phi(a)}$



18.14 Beispiel (Würfelwurf)

Echten Würfel $n = 600$ mal werfen; $S_n :=$ Anzahl der geworfenen Sechsen.

$\mathbb{P}(90 \le S_n \le 110) =$?

Es ist $p = 1/6$, $S_n \sim \mathrm{Bin}(n,p)$, $np = 100$, $\sigma_n := \sqrt{np(1-p)} \approx 9.13$.

$$\begin{aligned}
\mathbb{P}(90 \le S_n \le 110) &= \mathbb{P}\left(\frac{90-100}{\sigma_n} \le \frac{S_n - 100}{\sigma_n} \le \frac{110-100}{\sigma_n}\right) \\
&\approx \Phi\left(\frac{10}{9.13}\right) - \Phi\left(\frac{-10}{9.13}\right) \\
&\approx \Phi(1.1) - (1 - \Phi(1.1)) = 2 \cdot \Phi(1.1) - 1 \\
&\approx 2 \cdot 0.864 - 1 = 0.728.
\end{aligned}$$

Die verbesserte Approximation (Stetigkeitskorrektur) ergibt

$$\begin{aligned}
\mathbb{P}(90 \le S_n \le 110) &\approx \Phi\left(\frac{10+0.5}{9.13}\right) - \Phi\left(\frac{-10-0.5}{9.13}\right) \\
&\approx 2 \cdot \Phi(1.15) - 1 \approx 0.75.
\end{aligned}$$

Der exakte Wert (Maple) ist 0.753.

Beachte:

- Seien $Y_1, \ldots, Y_n$ unabhängig, je Po(λ)-verteilt $\Longrightarrow$

  $X_{n\lambda} := Y_1 + \ldots + Y_n \sim$ Po$(n\lambda)$ (Additionsgesetz für Po-Verteilung).

  Es gilt

$$\widetilde{X}_{n\lambda} \;=\; \frac{X_{n\lambda} - n\lambda}{\sqrt{n\lambda}} \;=\; \frac{\sum_{j=1}^{n} Y_j - n\mathbb{E}\, Y_1}{\sqrt{n\mathbb{V}(Y_1)}}$$

- Seien $A_1, \ldots, A_n$ unabhängig, $\mathbb{P}(A_j) = p \;\forall j$, $0 < p < 1 \Longrightarrow$

  $Y_1 := \mathbf{1}\{A_1\}, \ldots, Y_n := \mathbf{1}\{A_n\}$ unabhängig, $\mathbb{E}\, Y_j = p$, $\mathbb{V}(Y_j) = p(1-p)$.

  Weiter gilt $S_n := \sum_{j=1}^{n} Y_j \sim$ Bin(n,p) sowie

$$\widetilde{S}_n \;=\; \frac{S_n - np}{\sqrt{np(1-p)}} \;=\; \frac{\sum_{j=1}^{n} Y_j - n\mathbb{E}\, Y_1}{\sqrt{n\mathbb{V}(Y_1)}}$$

In jedem dieser beiden Fälle gilt (ZGWS für Po(λ) bzw. ZGWS für Bin(n,p)):

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{j=1}^{n} Y_j - n\mathbb{E}\, Y_1}{\sqrt{n\mathbb{V}(Y_1)}} \leq x \right) \;=\; \Phi(x), \quad x \in \mathbb{R}.$$

Hier liegt ein allg. Sachverhalt vor (→ Vorlesung „Wahrscheinlichkeitstheorie“)

18.15 Satz (Zentraler Grenzwertsatz von Lindeberg-Lévy)

Es seien $Y_1, \ldots, Y_n, \ldots$ unabhängige, identisch verteilte (u.i.v.) Zufallsvariablen mit existierender, positiver Varianz. Dann gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{j=1}^n Y_j - n\,\mathbb{E}(Y_1)}{\sqrt{n\mathbb{V}(Y_1)}} \le x\right) = \Phi(x), \quad x \in \mathbb{R}.$$

18.16 Beispiel (Negative Binomialverteilung)

Seien $Y_1, \ldots, Y_n, \ldots$ stochastisch unabhängig und je G(p)-verteilt, $0 < p < 1$.

Additionsgesetz für die negative Binomialverteilung $\Longrightarrow$

$$S_n := Y_1 + \ldots + Y_n \ \sim \text{Nb}(n,p). \text{ Mit } \mathbb{E}\,Y_1 = \frac{1-p}{p}, \ \ \mathbb{V}(Y_1) = \frac{1-p}{p^2} \text{ folgt}$$

$$\lim_{n\to\infty} \mathbb{P}\left(S_n - n \cdot \frac{1-p}{p} \le x \cdot \sqrt{n\,\frac{1-p}{p^2}}\right) = \Phi(x), \qquad x \in \mathbb{R}.$$

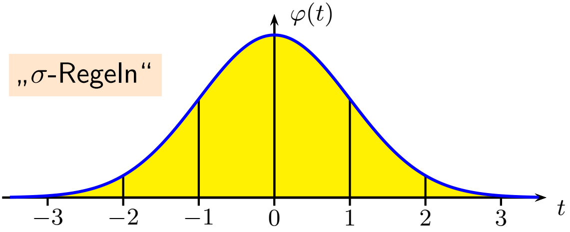
Memo: $Y_1, \ldots, Y_n, \ldots$ u.i.v., $\mu := \mathbb{E}\,Y_1, \;\; 0 < \sigma^2 := \mathbb{V}(Y_1) < \infty \;\; \Longrightarrow$

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{\sum_{j=1}^n Y_j - n\,\mu}{\sigma\sqrt{n}} \le b\right) \;=\; \int_a^b \varphi(x)\,\mathrm{d}x, \quad \sigma\sqrt{n} = \sqrt{\mathbb{V}\left(\sum_{j=1}^n Y_j\right)}$$

Speziell: $b = k \in \mathbb{N}$, $a = -b \Longrightarrow$

$$\lim_{n\to\infty} \mathbb{P}\left(n\mu - k\,\sigma\sqrt{n} \le \sum_{j=1}^n Y_j \le n\mu + k\,\sigma\sqrt{n}\right) \;=\; \int_{-k}^k \varphi(x)\,\mathrm{d}x$$

„σ-Regeln“



$$\int_{-1}^1 \varphi(x)\,\mathrm{d}x \approx 0.6826, \qquad \int_{-2}^2 \varphi(x)\,\mathrm{d}x \approx 0.9544, \qquad \int_{-3}^3 \varphi(x)\,\mathrm{d}x \approx 0.9974$$

18.17 Beispiel (Bernoulli-Kette mit $p = 1/2$)

Seien $A_1, A_2, \ldots$ unabhängig, $\mathbb{P}(A_j) = 1/2$, $j \geq 1$.

$$Y_j := \mathbf{1}\{A_j\}, \quad \mu = \mathbb{E}\, Y_j = \frac{1}{2}, \quad \sigma^2 = \mathbb{V}(Y_j) = \frac{1}{4}$$

$$S_n := \sum_{j=1}^{n} Y_j \sim \text{Bin}\left(n, \frac{1}{2}\right).$$

Für großes n gilt

$$\begin{aligned} \mathbb{P}\left(\frac{n}{2} - 1 \cdot \frac{\sqrt{n}}{2} \leq S_n \leq \frac{n}{2} + 1 \cdot \frac{\sqrt{n}}{2}\right) &\approx 0.6826, \\ \mathbb{P}\left(\frac{n}{2} - 2 \cdot \frac{\sqrt{n}}{2} \leq S_n \leq \frac{n}{2} + 2 \cdot \frac{\sqrt{n}}{2}\right) &\approx 0.9544, \\ \mathbb{P}\left(\frac{n}{2} - 3 \cdot \frac{\sqrt{n}}{2} \leq S_n \leq \frac{n}{2} + 3 \cdot \frac{\sqrt{n}}{2}\right) &\approx 0.9974. \end{aligned}$$

Also für $n = 10000$:

$$\begin{aligned} &\mathbb{P}(4950 \leq S_n \leq 5050) \approx 0.6826 \\ &\mathbb{P}(4900 \leq S_n \leq 5100) \approx 0.9544 \\ &\mathbb{P}(4850 \leq S_n \leq 5150) \approx 0.9974 \end{aligned}$$

19 Pseudozufallszahlen und Simulation

- Zufallsvorgänge werden häufig mit dem Computer simuliert.
- Beispiel: Nachfolgende Zahlen als Ergebnisse von 25 Würfelwürfen:

$$4\ 3\ 3\ 4\ 4\ 6\ 1\ 2\ 3\ 4\ 5\ 4\ 5\ 6\ 3\ 3\ 4\ 1\ 3\ 6\ 2\ 6\ 3\ 6\ 5$$

- Bausteine für die Simulation sind gleichverteilte Pseudozufallszahlen.
- Hierfür existieren (Pseudo-)Zufallszahlengeneratoren.
- Wunsch: Nachbildung einer „stetigen Gleichverteilung“ auf dem Intervall $[0,1]$ (→ Kap. 24) und der stochastischen Unabhängigkeit.
- Die stetige Gleichverteilung auf $[0,1]$ ordnet Teilintervallen von $[0,1]$ deren Länge als Wahrscheinlichkeit zu.

Hinter jedem Zufallszahlengenerator verbirgt sich ein Algorithmus.

Dieser erzeugt eine

- deterministische,
- jederzeit reproduzierbare
- Folge $x_0, x_1, x_2, \ldots$ im Intervall $[0, 1]$.

Wunsch: $x_0, x_1, x_2, \ldots$ sollen

- „unabhängig voneinander und gleichverteilt in $[0, 1]$“

wirken.

Zufallsgeneratoren versuchen, die diskrete Gleichverteilung auf

$$\Omega_m \ := \ \left\{ \frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m} \right\}$$

zu simulieren, wobei m groß, z.B. $m = 2^{32}$.

„n-maliger unabhängiger rein zufälliger Auswahl einer Zahl aus Ω_m“ entspricht dann die Gleichverteilung auf dem n-fachen kartesischen Produkt Ω_m^n.

Gute Generatoren müssen Tests hinsichtlich der *statistischen Qualität* der produzierten Zahlen bestehen.

Häufig verwendet: der lineare Kongruenzgenerator.

Er basiert auf nichtnegativen ganzen Zahlen

- m (Modul),
- a (Faktor),
- b (Inkrement),
- z_0 (Anfangsglied) mit $z_0 \leq m-1$

und verwendet das iterative Kongruenzschema

$$z_{j+1} \equiv a\, z_j + b \pmod{m}, \quad j \geq 0.$$

Durch die Normierungsvorschrift

$$x_j := \frac{z_j}{m}, \qquad j \geq 0,$$

entsteht dann eine Folge $x_0, x_1, \ldots$ in $[0,1]$.

Memo: $z_{j+1} \equiv a\, z_j + b \pmod{m}, \quad x_j := \dfrac{z_j}{m}, \quad j \geq 0.$

Zahlenbeispiel:

$$m = 100, \qquad a = 18, \quad b = 11, \quad z_0 = 40 \implies$$

$$\begin{aligned}
z_1 &\equiv 18 \cdot 40 + 11 \equiv 731 \equiv 31 \pmod{100},\\
z_2 &\equiv 18 \cdot 31 + 11 \equiv 569 \equiv 69 \pmod{100},\\
z_3 &\equiv 18 \cdot 69 + 11 \equiv 1253 \equiv 53 \pmod{100},\\
z_4 &\equiv 18 \cdot 53 + 11 \equiv 965 \equiv 65 \pmod{100},\\
z_5 &\equiv 18 \cdot 65 + 11 \equiv 1181 \equiv 81 \pmod{100},\\
z_6 &\equiv 18 \cdot 81 + 11 \equiv 1469 \equiv 69 \pmod{100}.
\end{aligned}$$

Also:

$$x_0 = 0.4, \; x_1 = 0.31, \; x_2 = 0.69, \; x_3 = 0.53, \; x_4 = 0.65, \; x_5 = 0.81$$
$$x_6 = 0.69 = x_2$$

Generator läuft schon nach 6 Schritten in die Periode 4.

Memo: $z_{j+1} \equiv a\, z_j + b \pmod{m}, \quad x_j := \dfrac{z_j}{m}, \quad j \geq 0.$

Beachte:

- Wegen $z_j \in \{0, \ldots, m-1\}$ sind höchstens m Zufallszahlen möglich,
- also sollte m sehr groß sein.

Die maximale Periodenlänge m wird bei $b \geq 1$ genau dann erreicht, wenn gilt:

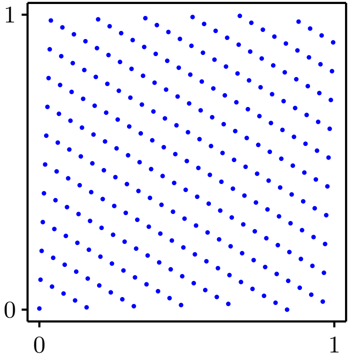
- b ist teilerfremd zu m,
- jede Primzahl, die m teilt, teilt auch $a-1$,
- Ist m durch 4 teilbar, so ist auch $a-1$ durch 4 teilbar.

Literatur: Knuth, D.E.: The art of computer programming, Vol. 2, 3. Auflage (1997), S.16.

Grundsätzlich sollte m wesentlich größer als die Zahl der benötigten Zufallszahlen sein.

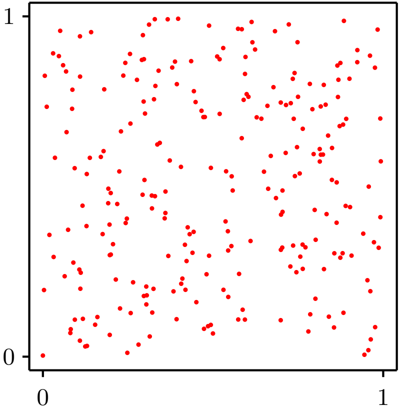
Prinzipielle Schwäche linearer Kongruenzgeneratoren: Gitterstruktur

Für jedes $d \geq 2$ liegen die Vektoren $\mathbf{x_j} := (x_j, x_{j+1}, \ldots, x_{j+d-1}),\ j \geq 0$, auf einem Gitter im $\mathbb{R}^d$ (d.h. $\mathbf{x_j} = \mathbf{x_0} + \frac{1}{m} \cdot$ ganzzahlige Linearkombination von d linear unabhängigen Vektoren $\in \mathbb{N}_0^d$)



Die 256 Pseudozufalls-Paare $(x_0, x_1), \ldots, (x_{255}, x_{256})$ des Generators mit $m = 256,\ a = 25,\ b = 1,\ z_0 = 1.$

Gittereffekt wird kaum sichtbar, wenn bei großem m relativ wenige Punktepaare (x_j, x_{j+1}) geplottet werden.



Die ersten 250 Pseudozufalls-Paare (x_0, x_1), (x_1, x_2), $\ldots, (x_{249},\ x_{250})$ des Generators mit $m = 2^{24},\ a = 54677,\ b = 1,\ z_0 = 1$

19.1 Simulation von Zufallsexperimenten

Simulation eines Experimentes, das mit W' p_j den Ausgang j ergibt $(j = 1, \ldots, s,\ p_1 + \ldots + p_s = 1)$:

- Erzeuge Pseudozufallszahl x.
- Stelle fest, in welchem der Intervalle

  $$[0, p_1),\ [p_1, p_1 + p_2), \ldots, [p_1 + p_2 + \ldots + p_{s-1}, 1)$$

  x liegt.
- Liegt x im Intervall mit rechtem Endpunkt $p_1 + \ldots + p_j$, so „Ausgang j“

Speziell: Laplace-Experiment mit den möglichen Ausgängen $1, \ldots, s$.

Hier $p_1 = \ldots = p_s = \frac{1}{s}$.

Beachte:

$$\frac{j-1}{s} \le x < \frac{j}{s} \iff j = \lfloor x\, s \rfloor + 1.$$

19.2 Beispiel (Würfeln mit MAPLE)

Der lineare Kongruenzgenerator von MAPLE ist gegeben durch

$$m = 10^{12} - 11, \quad a = 427419669081, \quad b = 0, \quad z_0 = 1.$$

Die ersten 12 hiermit erzeugten Pseudozufallszahlen sind

$$\begin{array}{cccc}
0.4274196691, & 0.3211106933, & 0.3436330737, & 0.4742561436, \\
0.5584587190, & 0.7467538305, & 0.0320622221, & 0.7229741218, \\
0.6043056139, & 0.7455800374, & 0.2598119527, & 0.3100754872.
\end{array}$$

Hiermit erhält man über die Transformation $x \mapsto \lfloor 6\,x \rfloor + 1$ die simulierten Augenzahlen

$$3,\ 2,\ 3,\ 3,\ 4,\ 5,\ 1,\ 5,\ 4,\ 5,\ 2,\ 2$$

von 12 Würfelwürfen.

20 Deskriptive Statistik

Mit dem Wort Statistik (von ital. *Statista* = Staatsmann) assoziiert man im Allgemeinen Begriffe wie

- *Außenhandelsstatistiken*,
- *Bevölkerungsstatistiken*,
- *Wahlstatistiken*,
- *Arbeitslosenstatistiken*,
- *Insolvenzstatistiken*,
- *Betriebsstatistiken*,
- *Schadensstatistiken*,
- *Krebsstatistiken*,
- *Einkommensstatistiken* usw.

Gemeinsam: Aufbereitete Daten aus verschiedenen Bereichen.

Ursprung der beschreibenden (deskriptiven) Statistik:

- Universitätsstatistik (Herrmann Conring)
  (Wissenschaft und Lehre von den Staatsmerkwürdigkeiten)
- Staatskunde (Gottfried Achenwall)
- politische Arithmetik (John Graunt, William Petty)
- amtliche Statistik (schon im Altertum)

Grobe Einteilung der Statistik:

- beschreibende Statistik
- schließende (induktive) Statistik

Beachte:

Die schließende Statistik verwendet Modelle der Wahrscheinlichkeitstheorie (entstand als Mathematische Statistik ab ca. 1900)

20.1 Untersuchungseinheiten und Merkmale

Bei

- statistischen Untersuchungen (Erhebungen)

werden an geeignet ausgewählten

- Untersuchungseinheiten (Beobachtungseinheiten, Versuchseinheiten)

jeweils die Werte eines oder mehrerer

- Merkmale

festgestellt.

Ein Merkmal ist eine zu untersuchende Größe der Beobachtungseinheit.

Werte, die von Merkmalen angenommen werden können, heißen Merkmalsausprägungen.

Untersuchungseinheit	**Merkmal**	**Ausprägungen**
Baum	Baumart	Eiche, Buche, ...
Baum	Schadstufe	$0, 1, 2, 3, 4$
Neugeborenes	Größe (in cm)	$\ldots, 49.5, 50, 50.5, \ldots$
arbeitslose Person	Schulabschluss	keiner, Sonderschule, Hauptschule, Realschule, Gymnasium
vollzeiterwerbstätige Person	Bruttoeinkommen im Jahr 2014 (in Euro)	$\ldots, 39999,$ $40000,$ $40001, \ldots$
Betonwürfel	Druckfestigkeit (in 0.1 N/mm^2)	$\ldots, 399, 400, 401, \ldots$

Grobunterscheidung:

- quantitatives MM (in natürlicher Weise zahlenmäßig erfassbar)
  - diskretes MM (Ausprägungen sind isolierte Zahlenwerte)
  - stetiges MM (Prinzipiell jeder Wert in einem Intervall möglich)
- qualitatives MM (artmäßig erfassbar)
  - nominales MM (Klassifizierung nach rein qualitativen Gesichtspunkten)
  - ordinales MM (Ausprägungen weisen Rangfolge auf)

20.2 Grundgesamtheit und Stichprobe

Grundgesamtheit (Population):

Diejenige Menge der Untersuchungseinheiten, über die hinsichtlich eines oder mehrerer interessierender Merkmale eine Aussage gemacht werden soll.

Die Grundgesamtheit (GG) ist die Menge aller *denkbaren* Untersuchungseinheiten (endlich oder unendlich groß, evtl. fiktiv).

Festlegung einer GG nicht immer einfach (z.B.: was ist ein Arbeitsloser?)

Stichprobe:

Zufällig gewonnene, endliche Teilmenge aus einer GG.

Hat diese Teilmenge n Elemente, so spricht man von einer *Stichprobe vom Umfang* n.

Vorsicht beim Auftreten des Begriffs repräsentative Stichprobe!

Im Folgenden sei $x_1, \ldots, x_n$ eine Stichprobe vom Umfang n eines Merkmals X.

20.3 Empirische Häufigkeitsverteilung, Stab- und Kreisdiagramm

Das Merkmal X habe s mögliche Ausprägungen $a_1, a_2, \ldots, a_s$.

Absolute Häufigkeiten in der Stichprobe $x_1, \ldots, x_n$:

$$h_j \;:=\; \sum_{i=1}^{n} \mathbf{1}\{x_i = a_j\} \qquad (j = 1, \ldots, s, \quad h_1 + \ldots + h_s = n)$$

Relative Häufigkeiten:

$$r_j \;:=\; \frac{h_j}{n} \;=\; \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{x_i = a_j\} \qquad (j = 1, \ldots, s, \quad r_1 + \ldots + r_s = 1).$$

Auch: Prozentanteile $100 \cdot r_j$%

Ohne Kenntnis von n können aus den relativen Häufigkeiten die absoluten Häufigkeiten nicht rekonstruiert werden!

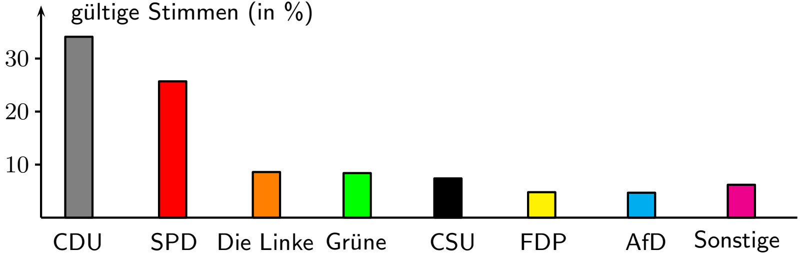
Partei	Zweitstimmen	in Prozent
CDU	14 921 877	34.1
SPD	11 252 215	25.7
Die Linke	3 755 699	8.6
Grüne	3 694 057	8.4
CSU	3 243 569	7.4
FDP	2 083 533	4.8
AfD	2 056 985	4.7
Sonstige	2 718 951	6.2

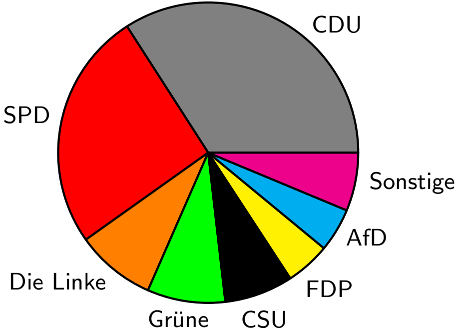
Stimmverteilung bei der Bundestagswahl 2013

Grafische Darstellungen von empirischen Häufigkeitsverteilungen erfolgen als Stabdiagramme oder Kreisdiagramme.

Stabdiagramm: Absolute bzw. relative Häufigkeiten werden als Funktion der Merkmalsausprägungen angezeigt.

Kreisdiagramm: Kreisfläche in Sektoren aufteilen, deren Flächen proportional zu den Häufigkeiten (absolut oder relativ) der Ausprägungen sind.


gültige Stimmen (in %)
30
20
10
CDU
SPD
Die Linke
Grüne
CSU
FDP
AfD
Sonstige


CDU
SPD
Sonstige
AfD
FDP
Die Linke
Grüne
CSU

20.4 Histogramme

- Zweck: übersichtliche Darstellung großer Datenmengen bei stetigem und/oder diskreten MM mit vielen Ausprägungen
- Mittel: Klasseneinteilung
- 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

$$\begin{aligned} & a_1 < a_2 < \ldots < a_s < a_{s+1}, \\ & a_1 \le \min_{1\le j\le n} x_j, \qquad \max_{1\le j\le n} x_j < a_{s+1} : \end{aligned}$$

Bilde über $[a_j, a_{j+1})$ ein Rechteck, dessen Fläche gleich der zugehörigen

relativen Klassenhäufigkeit

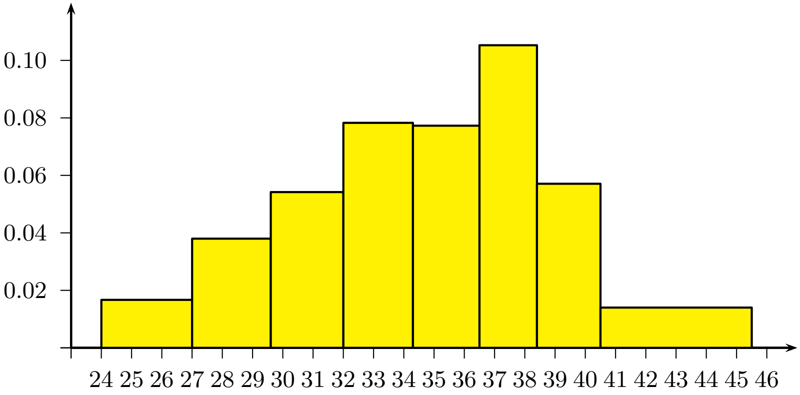
$$k_j \ := \ \frac{1}{n} \sum_{i=1}^{n} \ \mathbf{1}\{a_j \le x_i < a_{j+1}\}$$

ist. Also: Rechteckhöhe d_j durch $d_j\,(a_{j+1} - a_j) = k_j$ definiert.

Daten ($n = 100$ Werte)

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

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:



20.5 Stamm- und Blatt-Darstellung

Vorkommastellen der Größe als Stamm vorsortieren.

Nachkommastellen nach rechts als Blatt antragen.

24	7																		
25	8	1																	
26	4	1																	
27	4																		
28	4	5	0	7	8														
29	0	2	9																
30	4	9	4	8	7	6	6	7	4										
31	3	1	4																
32	5	9	7	9	2	5	6	9											
33	7	8	0	6	7	2	8	3	0	7									
34	5	3	8	7	4														
35	7	0	1	5	3	1	3	1	0										
36	2	0	3																
37	4	5	1	8	3	1	4	5	4	0	7	1	4	0	2	7	2	8	4
38	0	4	5	9															
39	1	7	2	2	4	3	3	0	0										
40																			
41	8	6																	
42	4																		
43	4	6																	
44	2																		
45	3																		
↓ *Stamm*	⟶ *Blatt*																		

Stamm- und Blatt-Darstellung ($n = 100$, Einheit = 1)

20.6 Lagemaße

Gegeben: Stichprobe $x_1, \ldots, x_n \in \mathbb{R}$.

Ziel: Zahl $l(x_1, \ldots, x_n)$ angeben, die *grobe Lage* der Stichprobe beschreibt.

Forderung: Translationsäquivarianz

$$l(x_1 + a, \ldots, x_n + a) \; = \; l(x_1, \ldots, x_n) + a \qquad \forall x_1, \ldots, x_n, a \in \mathbb{R}$$

Gebräuchlichstes Lagemaß: Arithmetisches Mittel

$$\overline{x}_n \; := \; \frac{1}{n}\,(x_1 + \cdots + x_n) \; = \; \frac{1}{n} \sum_{j=1}^{n} \; x_j$$

(auch: Mittelwert oder Durchschnitt von $x_1, \ldots, x_n$)

- $\overline{x}_n$ ist der physikalische Schwerpunkt von $x_1, \ldots, x_n$
- $\overline{x}_n$ ist empfindlich gegenüber „Ausreißern“
- $\overline{x}_n$ minimiert $\sum_{j=1}^{n}(x_j - t)^2$ als Funktion von t

Weitere Lagemaße sind über die geordnete Stichprobe definiert.

Sei $x_{(j)}$ der j-kleinste Wert von $x_1, \ldots, x_n$, also insbesondere

$$x_{(1)} = \min_{1 \le j \le n} x_j, \qquad x_{(n)} = \max_{1 \le j \le n} x_j.$$

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) \ \text{mit} \ x_{(1)} \le x_{(2)} \le \cdots \le x_{(n-1)} \le x_{(n)}$$

heißt geordnete Stichprobe von $x_1, \ldots, x_n$.

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

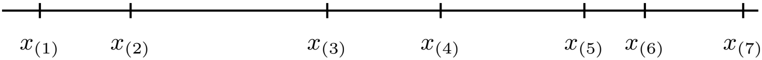
Empirischer Median (Zentralwert) von $x_1, \ldots, x_n$:

$$x_{1/2} := \begin{cases} x_{\left(\frac{n+1}{2}\right)} & , \quad \text{falls } n \text{ ungerade} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & , \quad \text{falls } n \text{ gerade} \end{cases}$$

- Median ist „Hälftigkeitswert“ (mindestens 50% aller x_j sind kleiner oder gleich $x_{1/2}$ und mindestens 50% aller x_j größer oder gleich $x_{1/2}$)
- Für die Daten obiger Tabelle ist $x_{1/2} = \dfrac{x_{(5)} + x_{(6)}}{2} = \dfrac{4.5 + 6.0}{2} = 5.25.$
- Der Median $x_{1/2}$ minimiert

$$\mathbb{R} \ni t \mapsto s(t) \ := \sum_{j=1}^{n} |x_j - t|$$

(Abstands-Summe) als Funktion von t.



Zur Minimaleigenschaft des Medians

- Der Median ist robust gegenüber Ausreißern!

Weitere Lage-Maße: p-Quantil und α-getrimmtes Mittel

Für $0 < p < 1$ heißt

$$x_p \; := \; \begin{cases} x_{(\lfloor n\,p+1 \rfloor)} & , \quad \text{falls } n\,p \notin \mathbb{N}, \\ \frac{1}{2}\left(x_{(n\,p)} + x_{(n\,p+1)}\right) & , \quad \text{falls } n\,p \in \mathbb{N}, \end{cases}$$

empirisches p-Quantil von $x_1, \ldots, x_n$.

Dabei ist $\lfloor y \rfloor := \max\{k \in \mathbb{Z} : k \leq y\}$.

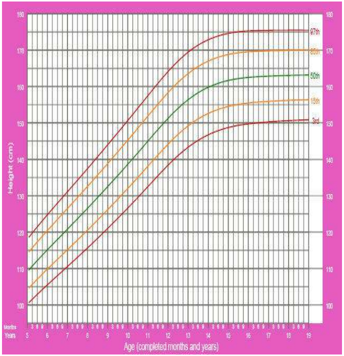
- mindestens $p\,100\%$ aller x_j sind kleiner oder gleich x_p
- mindestens $(1-p)\,100\%$ aller x_j sind größer oder gleich x_p
- „x_p teilt die geordnete Stichprobe im Verhältnis p zu $1-p$ auf“.

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

Beispiel: $x_{0.25} = x_{(\lfloor 3.5 \rfloor)} = x_{(3)} = 3.0$

Für spezielle Werte von p gibt es für x_p spezielle Begriffsbildungen:

- $p = 0.5$: empirischer Median,
- $p = 0.25$: unteres Quartil,
- $p = 0.75$: oberes Quartil,
- $p = j/10$: j-tes Dezil,
- $p = j/5$: j-tes Quintil.



Es seien $0 < \alpha < 1/2$ und $k := \lfloor n\,\alpha \rfloor$.

Dann heißt

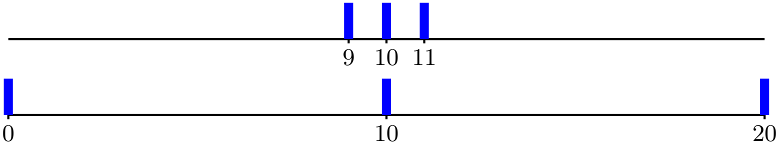
$$\begin{aligned} x_{t,\alpha} \quad &:= \quad \frac{1}{n-2\,k}\,\left(x_{(k+1)} + x_{(k+2)} + \cdots + x_{(n-k-1)} + x_{(n-k)}\right) \\ &= \quad \frac{1}{n-2\,k}\,\sum_{j=k+1}^{n-k} x_{(j)} \end{aligned}$$

α-getrimmtes Mittel oder $\alpha\,100\%$-getrimmtes Mittel von $x_1, \ldots, x_n$.

- $x_{t,\alpha}$ ist „Kompromiss“ zwischen arithmetischem Mittel ($\alpha \downarrow 0$) und Median ($\alpha \uparrow 1/2$)
- $x_{t,\alpha}$ ignoriert je $\alpha\,100\%$ der Daten in den Enden der geordneten Stichprobe
- $x_{t,\alpha}$ ist flexibles Instrument gegenüber Ausreißern

20.7 Streuungsmaße

Die Stichproben 9, 10, 11 und 0, 10, 20 haben den gleichen Mittelwert, aber unterschiedliche „Streuungen“.



Im Folgenden betrachten wir Streuungsmaße.

Ein Streuungsmaß $\sigma(x_1, \ldots, x_n)$ ist translationsinvariant, d.h. es gilt

$$\sigma(x_1 + a, x_2 + a, \ldots, x_n + a) \;=\; \sigma(x_1, \ldots, x_n) \qquad \forall x_1, \ldots, x_n, a \in \mathbb{R}.$$

Die gebräuchlichsten Streuungsmaße sind

- die empirische Varianz,
- die empirische Standardabweichung.

Für eine Stichprobe $x_1, \ldots, x_n$ heißen

$$s_n^2 \; := \; \frac{1}{n-1} \sum_{j=1}^{n} \left(x_j - \overline{x}_n\right)^2$$

die empirische Varianz oder Stichprobenvarianz von $x_1, \ldots, x_n$ und

$$s_n \; := \; +\sqrt{s_n^2} \; = \; \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} \left(x_j - \overline{x}_n\right)^2}$$

die empirische Standardabweichung oder Stichprobenstandardabweichung von $x_1, \ldots, x_n$.

- Warum teilt man durch $n-1$ und nicht durch n?
  Antwort: Beides möglich; es gibt jeweils Optimalitätsgesichtspunkte unter stochastischen Modellen (später).
- Alternative Darstellung: $s_n^2 \; = \; \dfrac{1}{n-1} \left(\displaystyle\sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2 \right)$
- s_n^2 ist empfindlich gegenüber Ausreißern

Memo: $s_n^2 \; := \; \dfrac{1}{n-1} \displaystyle\sum_{j=1}^{n} (x_j - \overline{x}_n)^2$

Beachte:

- $s_n^2(ax_1, \ldots, ax_n) \; = \; a^2 \, s_n^2(x_1, \ldots, x_n)$
- $s_n(ax_1, \ldots, ax_n) \; = \; |a| \, s_n(x_1, \ldots, x_n)$

Weitere Streuungsmaße:

- $\dfrac{1}{n} \displaystyle\sum_{j=1}^{n} |x_j - \overline{x}_n|$ (mittlere absolute Abweichung),
- $x_{(n)} - x_{(1)} \; = \; \max\limits_{1 \leq j \leq n} x_j - \min\limits_{1 \leq j \leq n} x_j,$ (Stichprobenspannweite),
- $x_{3/4} - x_{1/4}$ (Quartilsabstand),
- Median$(|x_1 - x_{1/2}|, \, |x_2 - x_{1/2}|, \ldots, |x_n - x_{1/2}|)$ (Median-Abweichung)

Quartilsabstand und Medianabweichung sind robust gegenüber Ausreißern.

20.8 Der Variationskoeffizient

Im Fall $x_1 > 0, \ldots, \ x_n > 0$ heißt der Quotient

$$V \ := \ V(x_1, \ldots, x_n) \ := \ \frac{s_n}{\overline{x}_n}$$

aus Standardabweichung und arithmetischem Mittel (empirischer) Variationskoeffizient von $x_1, ..., x_n$.

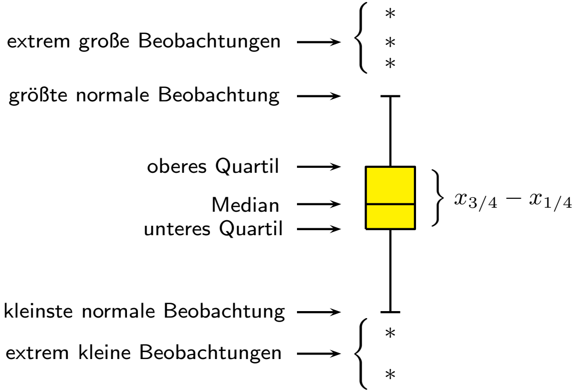
- V beschreibt die Stärke der *relativen* Streuung
- $V(ax_1, \ldots, ax_n) = V(x_1, \ldots, x_n) \ \forall a > 0$ (Maßstabsinvarianz).

20.9 Der Box-Plot

Zweck: Schneller visueller Vergleich verschiedener Stichproben.

Box vom unteren zum oberen Quartil, beim Median unterteilt.

Stäbe bis zum größten $x_j \leq x_{3/4} + 1.5 \cdot (x_{3/4} - x_{1/4})$ (größte normale Beob.)
und zum kleinsten $x_j \geq x_{1/4} - 1.5 \cdot (x_{3/4} - x_{1/4})$ (kleinste normale Beob.)



extrem große Beobachtungen

größte normale Beobachtung

oberes Quartil

Median

unteres Quartil

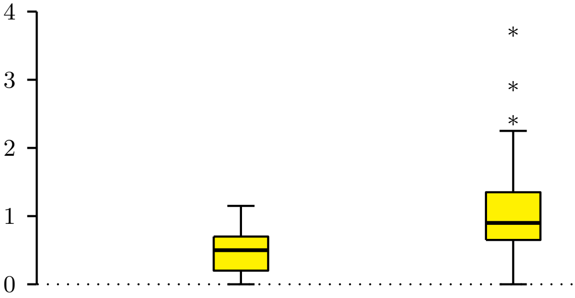
kleinste normale Beobachtung

extrem kleine Beobachtungen

$x_{3/4} - x_{1/4}$

Beispiel:

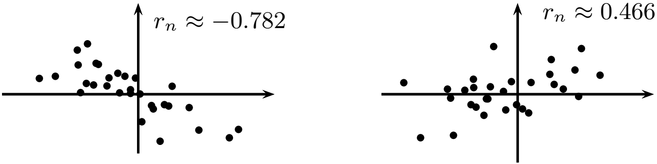
Untersuchung mit 140 Studierenden am KIT, in der u.a. der Cadmium-Gehalt im Blut festgestellt wurde.



Box-Plots zum Cadmiumgehalt (in μg pro Liter) im Blut von Studierenden bei Nichtrauchern (links) und Rauchern (rechts)

20.10 Streudiagramm, empirische Regressionsgerade

Ein Streudiagramm oder Scatterplot entsteht, wenn Punktepaare (x_j, y_j) $(j = 1, \ldots, n)$ in der Ebene geplottet werden.



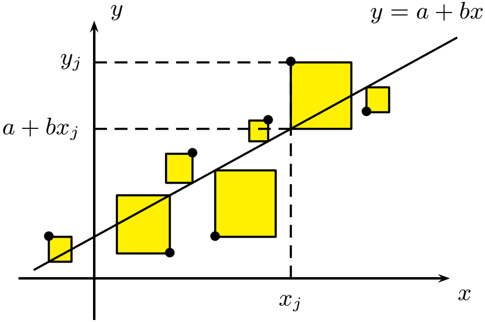
$$\text{Sei} \quad \bar{x} := \frac{1}{n}\sum_{j=1}^{n} x_j, \quad \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j,$$

$$s_{xx} := \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2, \quad s_{yy} := \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2, \quad s_{xy} := \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}),$$

$$r_n := \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \cdot \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2}}$$

r_n heißt empirischer Korrelationskoeffizient von $(x_j, y_j), 1 \leq j \leq n$, s. Kap.12.

Methode der kleinsten Quadrate:



$$\text{Aufgabe: } \sum_{j=1}^{n} (y_j - a - bx_j)^2 \overset{!}{=} \min_{a,b}$$

Die optimale Gerade $x \mapsto a^* x + b^*$ heißt (empirische) Regressionsgerade von y auf x.

$$\text{Es gilt:} \quad b^* = \frac{s_{xy}}{s_{xx}}, \quad a^* = \bar{y} - b^* \bar{x}.$$

20.11 Definition (geometrisches Mittel, kein Lagemaß!)

Es seien $x_1, \ldots, x_n > 0$. Dann heißt

$$\overline{x}_g \; := \; \left(\prod_{j=1}^{n} x_j \right)^{1/n}$$

geometrisches Mittel von $x_1, \ldots, x_n$.

20.12 Beispiel (Kapitalverzinsung)

Ein Kapital K wird im j-ten Jahr mit p_j Prozent verzinst, $j = 1, \ldots, n$.

$$\text{Kapitalstand nach } n \text{ Jahren: } K \prod_{j=1}^{n} x_j, \quad x_j = 1 + \frac{p_j}{100}.$$

Bei gleicher jährlicher Verzinsung um p Prozent:

$$\text{Kapitalstand nach } n \text{ Jahren: } K \left(1 + \frac{p}{100} \right)^n .$$

Gleichsetzen liefert

$$1 + \frac{p}{100} \; = \; \overline{x}_g \quad \Longrightarrow \quad p = 100 \cdot (\overline{x}_g - 1) \quad \text{(Durchschnittszinssatz)}$$

20.13 Definition (harmonisches Mittel, kein Lagemaß!)

Es seien $x_1, \ldots, x_n > 0$. Dann heißt

$$\overline{x}_h \ := \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}}$$

harmonisches Mittel von $x_1, \ldots, x_n$.

20.14 Beispiel (Durchschnittsgeschwindigkeit)

Ein Pkw durchfahre den j-ten Teil einer in n gleich lange Teilstrecken unterteilten Gesamtstrecke mit der *konstanten* Geschwindigkeit x_j km/h $(j = 1, \ldots, n)$.

Sei s die Länge jeder Teilstrecke (in km).

Benötigte Zeit für j-te Teilstrecke: $\frac{s}{x_j}$

Erzielte Durchschnittsgeschwindigkeit:

$$\frac{ns}{\dfrac{s}{x_1} + \dfrac{s}{x_2} + \ldots + \dfrac{s}{x_n}} = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}} = \overline{x}_h$$

21 Induktive Statistik: Punktschätzung

Bislang:

- $(\Omega, \mathbb{P})$ diskreter W-Raum
- $\mathbb{P}$ bekannt
- $\mathbb{P}$ „steuert das Auftreten von Daten“ $\omega \in \Omega$
  $\omega \;\leftrightarrow\;$ Ergebnis eines stochastischen Vorgangs
- Falls $X = (X_1, \ldots, X_n)$ n-dimensionaler Zufallsvektor auf Ω,
  so Daten in Form von Realisierungen $x = (x_1, \ldots, x_n)$ von X:
  $\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$
- Von Interesse: Die Wahrscheinlichkeiten $\mathbb{P}(X = x)$ in Abhängigkeit von x

Im Folgenden:

- $\mathbb{P}$ nicht vollständig bekannt
- Aufgrund von Daten $x = (x_1, \ldots, x_n)$ ist eine begründete Aussage über $\mathbb{P}$ zu treffen.

21.1 Beispiel (Schätzung einer Wahrscheinlichkeit, I)

- Bernoulli-Kette der Länge n mit unbekannter Trefferw' p, $0 < p < 1$
- In n Versuchen seien k Treffer aufgetreten; $p = ?$
- Modell: Für zufällige Trefferzahl S_n gilt $S_n \sim \mathrm{Bin}(n, p)$, $p =?$
- Das Ereignis $\{S_n = k\}$ ist eingetreten (gegebene Daten!)
- Es gilt $\quad \mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$
- p muss spezifiziert werden, um Wahrscheinlichkeiten zu berechnen (Indizierung von $\mathbb{P}$ mit p)
- Neue Sichtweise: k gegeben.

  Welches Modell $\mathbb{P}_p$ (welches p) passt am besten zu k?
- Für jedes $p \in (0,1)$ gilt $\mathbb{P}_p(S_n = k) > 0$.

  Jedes p kann Daten k erzeugt haben.
- Nur Antwort „$0 < p < 1$“ mit Sicherheit richtig;

  jede genauere Antwort wie z.B. „$0.32 \le p \le 0.91$“ ist potenziell falsch!

Memo: $\mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ (k bekannt, p unbekannt)

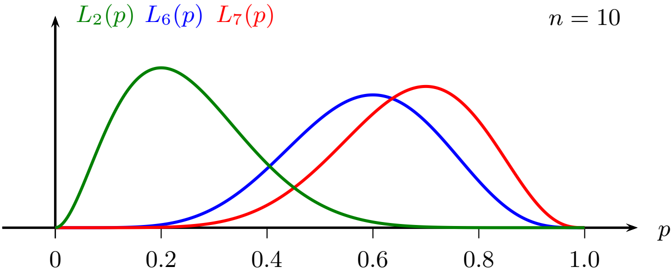
Unter welchem Wert von p hat das Ereignis $\{S_n = k\}$ die größte W'keit?

Die Funktion

$$L_k : \begin{cases} [0,1] \to \mathbb{R}, \\ \quad p \quad \mapsto L_k(p) \ := \ \mathbb{P}_p(S_n = k) \end{cases}$$

heißt Likelihood-Funktion für p zur Beobachtung $S_n = k$.

(Sichtweise: Daten k fest, variiere Modelle)

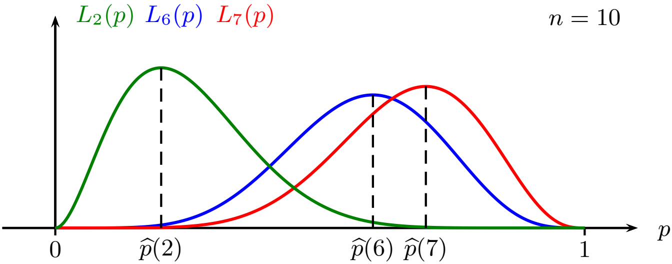


Memo: $L_k(p) \; = \; \mathbb{P}_p(S_n = k) \; = \; \binom{n}{k} p^k (1-p)^{n-k}$

Ein Wert $\widehat{p}(k) \in [0,1]$ mit

$$L_k\left(\widehat{p}(k)\right) \; = \; \max_{0 \le p \le 1} L_k(p)$$

heißt Maximum-Likelihood-Schätzwert (kurz: ML-Schätzwert) für p zur Beobachtung $S_n = k$.



Memo: $L_k(p) \ = \ \mathbb{P}_p(S_n = k) \ = \ \binom{n}{k} p^k (1-p)^{n-k}$

Behauptung: $L_k(p)$ wird maximal für $\widehat{p} = \widehat{p}(k) = \dfrac{k}{n}$.

Beweis: 1. Fall: $k = 0 \Longrightarrow L_0(p) = (1-p)^n$, maximal für $\widehat{p} = 0 = 0/n$.

2. Fall: $k = n \Longrightarrow L_n(p) = p^n$, maximal für $\widehat{p} = 1 = n/n$.

3. Fall: $1 \le k \le n-1$

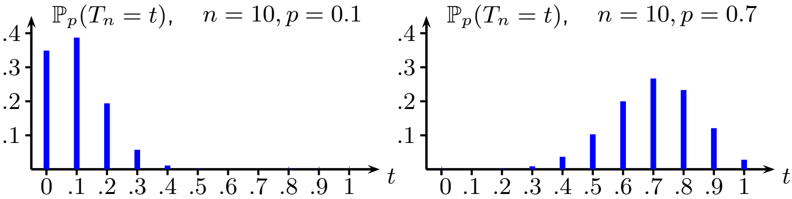
$$\begin{aligned} L_k^{'}(p) \ &= \ \binom{n}{k} p^{k-1}(1-p)^{n-k-1} \cdot (k(1-p) - (n-k)p) \\ &= \ 0 \iff p = \frac{k}{n} \end{aligned}$$

$$\Longrightarrow \ \widehat{p} = \frac{k}{n} \quad \text{(es liegt ein Maximum vor)}$$

D.h.: Relative Trefferhäufigkeit $\frac{k}{n}$ ist der ML-Schätzwert für p.

- $\widehat{p}$ ist Realisierung der Zufallsvariablen $T_n := \frac{1}{n} S_n$.
- Die Zufallsvariable T_n heißt Schätzer für p
- Die Realisierungen von T_n sind die konkreten Schätzwerte.
- Die Verteilung des Schätzers T_n hängt vom unbekannten p ab:

$$\mathbb{P}_p\left(T_n = \frac{k}{n}\right) \;=\; \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$



Es gilt $\quad \mathbb{E}_p(T_n) \;=\; \frac{1}{n}\,\mathbb{E}_p(S_n) \;=\; p \;\; \forall p \in (0,1)$

Dieses Verhalten werden wir „Erwartungstreue“ von T_n nennen.

Memo: $S_n \sim \mathrm{Bin}(n,p), \qquad T_n = \frac{S_n}{n}, \qquad \mathbb{E}_p(T_n) \;=\; p \;\; \forall\, p$

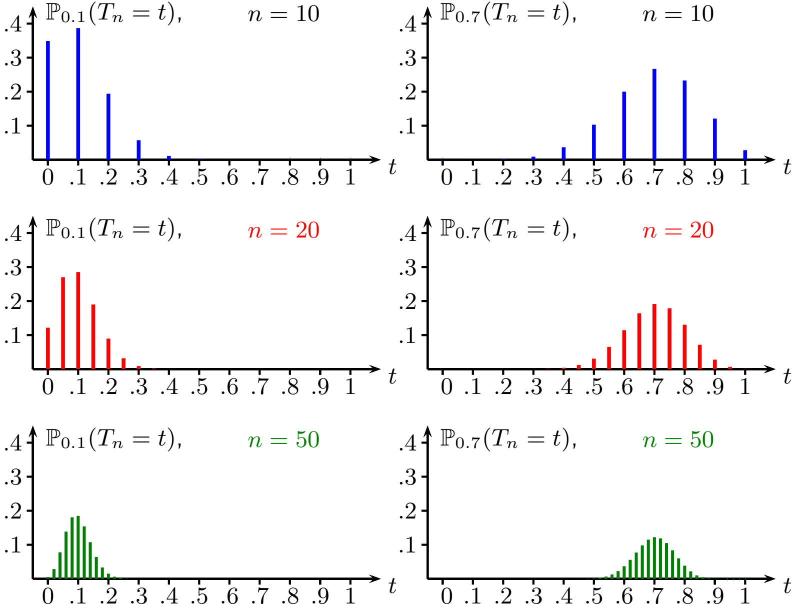
Was passiert bei Vergrößerung von n? Schätzung wird genauer!

Betrachte hierzu die Varianz von T_n:

$$\begin{aligned} \mathbb{V}_p(T_n) &= \frac{1}{n^2}\,\mathbb{V}_p(S_n) \\ &= \frac{np(1-p)}{n^2} \\ &= \frac{p(1-p)}{n} \;\to 0 \text{ bei } n \to \infty \;\forall p \end{aligned}$$

Tschebyschow-Ungleichung $\Longrightarrow$

$$\begin{aligned} \mathbb{P}_p\left(|T_n - p| \geq \varepsilon\right) &\leq \frac{\mathbb{V}_p(T_n)}{\varepsilon^2} \\ &= \frac{p(1-p)}{n\varepsilon^2} \;\to 0 \;\; \text{bei } n \to \infty \;\forall\, \varepsilon > 0. \end{aligned}$$



$\mathbb{P}_{0.1}(T_n = t)$, $n = 10$

$\mathbb{P}_{0.7}(T_n = t)$, $n = 10$

$\mathbb{P}_{0.1}(T_n = t)$, $n = 20$

$\mathbb{P}_{0.7}(T_n = t)$, $n = 20$

$\mathbb{P}_{0.1}(T_n = t)$, $n = 50$

$\mathbb{P}_{0.7}(T_n = t)$, $n = 50$

21.2 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 $\mathcal{P}(\mathcal{X})$
- Die Zuordnung $\Theta \ni \vartheta \mapsto \mathbb{P}_\vartheta$ sei injektiv
- $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ 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 $X = (X_1, \ldots, X_n)$

  dabei kanonische Konstruktion: $\Omega := \mathcal{X}$, $X = \mathrm{id}_\Omega$
- damit: $\mathbb{P}_\vartheta(\{x\}) = \mathbb{P}_\vartheta(X = x)$

21.3 Beispiel (Bernoulli-Schema)

- $\Theta \;:=\; (0,1)$,
- $\mathcal{X} \;:=\; \{0,1\}^n$,
- $X = (X_1,\ldots,X_n)$, wobei $X_1,\ldots,X_n$ unabhängig, je $\sim$ Bin$(1,\vartheta)$
- $x = (x_1 \ldots, x_n) \in \mathcal{X} \Longrightarrow$

$$\mathbb{P}_\vartheta(X = x) \;=\; \vartheta^k (1-\vartheta)^{n-k}, \quad k = x_1 + \ldots + x_n$$

- Unter $\mathbb{P}_\vartheta$ gilt: $S_n := X_1 + \ldots + X_n \sim$ Bin(n,ϑ), d.h.

$$\mathbb{P}_\vartheta(S_n = k) \;=\; \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}, \quad k = 0,1,\ldots,n.$$

21.4 Beispiel (Qualitätskontrolle)

Situation: Warensendung vom Umfang N mit ϑ defekten und $N-\vartheta$ intakten Exemplaren. N ist bekannt, ϑ unbekannt.

n mal rein zufällig Ziehen ohne Zurücklegen

- $\Theta \ := \ \{0, 1, \ldots, N\}$,
- $\mathcal{X} \ := \ \{0, 1\}^n$,
- $X := (X_1, \ldots, X_n)$, $\ X_j := \mathbf{1}\{$„j-tes entnommenes Exemplar defekt“$\}$
- Für $x = (x_1, \ldots, x_n) \in \mathcal{X}$ mit $x_1 + \ldots + x_n = k$ ist

$$\begin{aligned}\mathbb{P}_\vartheta(X = x) &= \frac{\vartheta(\vartheta-1)\ldots(\vartheta-k+1)(N-\vartheta)(N-\vartheta-1)\ldots(N-\vartheta-(n-k)+1)}{N(N-1)\ldots(N-n+1)} \\ &= \frac{\vartheta^{\underline{k}}\,(N-\vartheta)^{\underline{n-k}}}{N^{\underline{n}}}\end{aligned}$$

- Unter $\mathbb{P}_\vartheta$ gilt: $S_n := X_1 + \ldots + X_n \sim \mathrm{Hyp}(n, \vartheta, N-\vartheta)$

21.5 Definition ((Punkt-)Schätzer)

Es sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell.

Ein (Punkt)-Schätzer für ϑ ist eine Abbildung $T : \mathcal{X} \rightarrow \widetilde{\Theta}$, wobei $\widetilde{\Theta} \supseteq \Theta$.

Der Wert $T(x)$ heißt konkreter Schätzwert für ϑ (zu $x \in \mathcal{X}$).

Beachte:

- Im Allgemeinen ist $\Theta = \widetilde{\Theta}$.
- Aus mathematischen Gründen manchmal echte Obermenge nötig, z.B.:

21.6 Beispiel (Binomialfall)

In der Situation von Beispiel 21.3 ist $T : \mathcal{X} \rightarrow \widetilde{\Theta} := [0, 1]$,

$$T(x) \; := \frac{1}{n} \sum_{j=1}^{n} x_j, \qquad x = (x_1, \ldots, x_n) \in \{0, 1\}^n = \mathcal{X}$$

ein Schätzer für ϑ. Hier wird meist $\Theta := (0, 1)$ angenommen.

Memo: $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ statistisches Modell, $T : \mathcal{X} \to \tilde{\Theta}$ Schätzer

21.7 Bemerkungen

- 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 eine auf $\mathcal{X}$ definierte Zufallsvariable mit Werten in $\widetilde{\Theta}$.
-
$$\mathbb{P}_\vartheta(T = t) \;=\; \mathbb{P}_\vartheta(\{x \in \mathcal{X} : T(x) = t\})$$
definiert die Verteilung von T unter $\mathbb{P}_\vartheta$.
- Ideal wäre $\mathbb{P}_\vartheta(T = \vartheta) = 1$ für jedes $\vartheta \in \Theta$.
- 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.

Im Folgenden gelte stets $\mathbb{P}_\vartheta(\mathcal{X}_0) = 1$, $\vartheta \in \Theta$, für eine (von ϑ unabhängige) abzählbare Teilmenge $\mathcal{X}_0$ von $\mathcal{X}$ (vgl. Ω_0 und Ω).

21.8 Definition (Mittlere quadratische Abweichung, Verzerrung)

Sei $T : \mathcal{X} \to \widetilde{\Theta}$ mit $\widetilde{\Theta} \subseteq \mathbb{R}$. Es gelte

$$\mathbb{E}_\vartheta(T^2) \;=\; \sum_{x \in \mathcal{X}_0} T^2(x)\, \mathbb{P}_\vartheta(\{x\}) \;<\; \infty, \qquad \vartheta \in \Theta.$$

a) $\mathrm{MQA}_T(\vartheta) \;:=\; \mathbb{E}_\vartheta\,(T - \vartheta)^2 \;=\; \sum_{x \in \mathcal{X}_0} (T(x) - \vartheta)^2\, \mathbb{P}_\vartheta(\{x\})$

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.

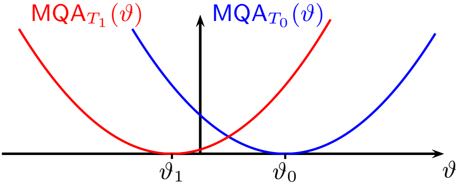
c) T heißt erwartungstreu, falls $\mathbb{E}_\vartheta(T) = \vartheta, \quad \vartheta \in \Theta$.

Memo: $\mathrm{MQA}_T(\vartheta) := \mathbb{E}_\vartheta\,(T-\vartheta)^2$

Beachte:

- Nach dem Verschiebungssatz gilt $\mathrm{MQA}_T(\vartheta) \;=\; \mathbb{V}_\vartheta(T) + (\mathrm{b}_T(\vartheta))^2$.
- Sei $\vartheta_0 \in \Theta$ fest. Für den unsinnigen Schätzer $T_0(x) := \vartheta_0 \;\forall\; x \in \mathcal{X}$ (ignoriert die Daten!) gilt $\mathbb{V}_\vartheta(T_0) = 0$, $\vartheta \in \Theta$, sowie $\mathbb{E}_\vartheta(T_0) = \vartheta_0$, $\vartheta \in \Theta$,

$$\Longrightarrow \quad \mathrm{MQA}_{T_0}(\vartheta) \;=\; (\vartheta_0 - \vartheta)^2$$



Folgerung: Man muss solche extremen Schätzer ausschließen,

z.B. durch Einschränkung auf erwartungstreue Schätzer.

Memo: $\mathsf{MQA}_T(\vartheta) := \mathbb{E}_\vartheta\,(T-\vartheta)^2, \quad \mathsf{b}_T(\vartheta) = \mathbb{E}_\vartheta(T) - \vartheta$

21.9 Beispiel (Binomialfall)

Seien $X_1, \ldots, X_n$ unabhängig und je Bin$(1,\vartheta)$-verteilt, wobei $\vartheta \in \Theta := [0,1]$, $\mathcal{X} = \{0,1\}^n$,

$$T_n(x_1,\ldots,x_n) \; := \; \frac{1}{n}\sum_{j=1}^n x_j,$$

$$T_n := T_n(X_1,\ldots,X_n)$$

$$\mathbb{E}_\vartheta(T_n) \; = \; \vartheta \;\forall\, \vartheta \in \Theta \implies \mathsf{b}_{T_n}(\vartheta) \; = \; 0 \;\forall\, \vartheta \in \Theta$$

$$\mathbb{V}_\vartheta(T_n) \; = \; \frac{\vartheta(1-\vartheta)}{n} \;\forall\, \vartheta \in \Theta \implies \mathsf{MQA}_{T_n}(\vartheta) \; = \; \frac{\vartheta(1-\vartheta)}{n}.$$

21.10 Definition (Maximum-Likelihood-Schätzung)

Es seien $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $x \in \mathcal{X}$.

- Die Funktion

$$L_x : \begin{cases} \Theta & \to [0,1], \\ \vartheta & \mapsto L_x(\vartheta) := \mathbb{P}_\vartheta(X = x) \end{cases}$$

  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

$$L_x\left(\widehat{\vartheta}(x)\right) = \sup_{\vartheta \in \Theta} L_x(\vartheta), \qquad (21.6)$$

  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 (21.6) für jedes $x \in \mathcal{X}$ heißt Maximum-Likelihood-Schätzer (kurz: ML-Schätzer) für ϑ.

21.11 Beispiel (Fortsetzung von Beispiel 21.4)

Seien $\mathcal{X} = \{0,1\}^n$, $\Theta = \{0,1,\ldots,N\}$, $X = (X_1,\ldots,X_n)$, $x = (x_1,\ldots,x_n)$,

$$L_x(\vartheta) \;=\; \mathbb{P}_\vartheta(X = x) \;=\; \frac{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}}{N^{\underline{n}}}, \qquad k = x_1 + \ldots + x_n.$$

Welche Gestalt besitzt der ML-Schätzer für ϑ?

Naive Hochrechnung („gesunder Menschenverstand"):

$$\frac{k}{n} \approx \frac{\vartheta}{N} \implies \text{Schätzwert } \vartheta^*(k) = \left\lfloor \frac{N \cdot k}{n} \right\rfloor$$

a) $k = 0 \implies L_x(\vartheta) = \dfrac{(N-\vartheta)^{\underline{n}}}{N^{\underline{n}}} \implies \widehat{\vartheta}(k) = 0 = \vartheta^*(k)$,

b) $k = n \implies L_x(\vartheta) = \dfrac{\vartheta^{\underline{n}}}{N^{\underline{n}}} \implies \widehat{\vartheta}(k) = N = \vartheta^*(k)$,

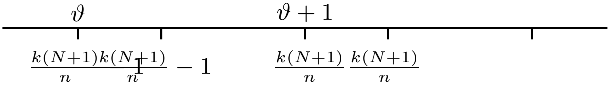
c) $1 \le k \le n-1 \implies$

$$\begin{aligned}\frac{L_x(\vartheta+1)}{L_x(\vartheta)} &= \frac{(\vartheta+1)^{\underline{k}}(N-\vartheta-1)^{\underline{n-k}}}{N^{\underline{n}}} \cdot \frac{N^{\underline{n}}}{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}} \\ &= \frac{\vartheta+1}{\vartheta-k+1} \cdot \frac{N-\vartheta+k-n}{N-\vartheta}\end{aligned}$$

Direkte Rechnung liefert:

$$\frac{L_x(\vartheta+1)}{L_x(\vartheta)} > 1 \quad \Longleftrightarrow \quad \vartheta < \frac{k(N+1)}{n} - 1$$

$$\frac{L_x(\vartheta+1)}{L_x(\vartheta)} = 1 \quad \Longleftrightarrow \quad \vartheta = \frac{k(N+1)}{n} - 1$$



Also: $\widehat{\vartheta}(k) \begin{cases} = \left\lfloor \frac{k(N+1)}{n} \right\rfloor, & \text{falls } \frac{k(N+1)}{n} \notin \mathbb{N} \\ \in \left\{ \frac{k(N+1)}{n}, \frac{k(N+1)}{n} - 1 \right\}, & \text{sonst ,} \end{cases}$

$$\Longrightarrow \ \widehat{\vartheta}(k) := \left\lfloor \frac{k(N+1)}{n} \right\rfloor \quad \text{ist ein ML-Schätzwert}$$

Beachte: $\vartheta^*(k) \ = \ \left\lfloor \frac{N\,k}{n} \right\rfloor \ \leq \ \left\lfloor \frac{k(N+1)}{n} \right\rfloor \ = \ \widehat{\vartheta}(k) \ \leq \ \vartheta^*(k) + 1$

21.12 Bemerkung (Loglikelihood-Funktion)

Für $x \in \mathcal{X}$ heißt die durch

$$\Theta \ni \vartheta \to \log L_x(\vartheta) = \log \mathbb{P}_\vartheta(X = x)$$

definierte Funktion $\log L_x$ die Loglikelihood-Funktion zu x.

Beachte: $\log L_x(\cdot)$ 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

$$\begin{aligned} L_x(\vartheta) &= \mathbb{P}_\vartheta(X = x) = \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{j=1}^n \mathbb{P}_\vartheta(X_j = x_j) \\ \Longrightarrow \log L_x(\vartheta) &= \sum_{j=1}^n \log \mathbb{P}_\vartheta(X_j = x_j) \end{aligned}$$

Summe meist leichter zu differenzieren als Produkt!

21.13 Beispiel (ML-Schätzung bei geometrischer Verteilung)

Seien $X_1, \ldots, X_n$ unabhängig und je G(ϑ)-verteilt, wobei $\vartheta \in \Theta = (0,1)$.

Aufgabe: ML-Schätzung von ϑ aufgrund von $X_1, \ldots, X_n$.

Seien $x_1, \ldots, x_n \in \mathbb{N}_0$, $x := (x_1, \ldots, x_n)$, $X := (X_1, \ldots, X_n)$.

$$\begin{aligned} L_x(\vartheta) &= \mathbb{P}_\vartheta(X = x) = \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{j=1}^n \mathbb{P}_\vartheta(X_j = x_j) = \prod_{j=1}^n \left\{(1-\vartheta)^{x_j}\vartheta\right\} = \vartheta^n (1-\vartheta)^{x_1+\ldots+x_n} \\ \Longrightarrow \log L_x(\vartheta) &= n \log \vartheta + \left(\sum_{j=1}^n x_j\right) \log(1-\vartheta) \end{aligned}$$

$$\frac{\mathsf{d}}{\mathsf{d}\vartheta} \log L_x(\vartheta) = \frac{n}{\vartheta} - \frac{1}{1-\vartheta} \sum_{j=1}^n x_j = 0 \iff \widehat{\vartheta}(x) = \frac{1}{1 + \frac{1}{n}\sum_{j=1}^n x_j}$$

Der ML-Schätzer für ϑ ist also

$$\widehat{\vartheta}_n := \frac{1}{1 + \overline{X}_n}, \qquad \overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j.$$

21.14 Definition (Schätzfolge, asymptotische E-Treue, Konsistenz)

Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit Verteilung $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$.

Ist $\mathcal{X} \subseteq \mathbb{R}$ der Stichprobenraum für jedes einzelne X_j, so ist $\mathcal{X}^n \subseteq \mathbb{R}^n$ der Stichprobenraum für $(X_1, \ldots, X_n)$.

Ist für jedes $n \in \mathbb{N}$ $T_n : \mathcal{X}^n \to \widetilde{\Theta} \supseteq \Theta$ ein Schätzer für ϑ, so heißt $(T_n)_{n \geq 1}$ eine Schätzfolge (Beachte: $T_n = T_n(X_1, \ldots, X_n)$ 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

$$\lim_{n \to \infty} \mathbb{P}_\vartheta\left(|T_n - \vartheta| \geq \varepsilon\right) \;=\; 0 \quad \forall \varepsilon > 0,\ \forall\, \vartheta \in \Theta.$$

Beachte: (T_n) konsistent $\iff$ $T_n \xrightarrow{\mathbb{P}_\vartheta} \vartheta \quad \forall\, \vartheta \in \Theta$

21.15 Beispiel (Fortsetzung von Beispiel 21.13)

Seien $X_1, X_2, \ldots$ unabhängig und je G(ϑ)-verteilt.

Der auf $X_1, \ldots, X_n$ basierende ML-Schätzer ist

$$\widehat{\vartheta}_n \; := \; \frac{1}{1 + \overline{X}_n}, \qquad \overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j.$$

Nach dem Gesetz großer Zahlen gilt

$$\overline{X}_n \; \xrightarrow{\mathbb{P}_\vartheta} \; \mathbb{E}_\vartheta(X_1) \; = \; \frac{1-\vartheta}{\vartheta} \; =: \; a.$$

Da die Funktion $g(t) := \frac{1}{1+t}$, $t \geq 0$, stetig ist, folgt nach Satz 18.6

$$\widehat{\vartheta}_n \; = \; \frac{1}{1 + \overline{X}_n} \; = \; g(\overline{X}_n) \; \xrightarrow{\mathbb{P}_\vartheta} \; g(a) \; = \; \frac{1}{1+a} \; = \; \vartheta.$$

Somit ist die Schätzfolge $(\widehat{\vartheta}_n)$ konsistent für ϑ.

Der Schätzer $\widehat{\vartheta}_n$ ist nicht erwartungstreu, die Schätzfolge $(\widehat{\vartheta}_n)$ ist asymptotisch erwartungstreu (Übungsaufgabe!)

22 Induktive Statistik: Konfidenzbereiche

- Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell.
- Sei $T : \mathcal{X} \longrightarrow \widetilde{\Theta}$ ein Punktschätzer für ϑ, wobei $\widetilde{\Theta} \supseteq \Theta$.
- T liefert für Daten $x \in \mathcal{X}$ einen konkreten Schätzwert $T(x)$ für ϑ.
- $T(x)$ macht keine Aussage über die Größe des Fehlers $T(x) - \vartheta$.
- Kann man $T(x)$ mit einer Genauigkeitsangabe versehen, z.B. bei $\widetilde{\Theta} \subseteq \mathbb{R}$ in Form eines Intervalls

$$\mathcal{C}(x) \ = \ [T(x) - a(x), T(x) + b(x)]\,?$$

- Welchen Wahrheitsanspruch besitzt die Aussage

„das Intervall $\mathcal{C}(x)$ enthält den unbekannten Parameter ϑ“?

- Beachte: ϑ ist unbekannt, aber nicht zufällig, und $\mathcal{C}(x)$ ist festes Intervall.

22.1 Beispiel (Binomialverteilung)

Sei $X = (X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ unabhängig, je $\sim$ Bin$(1, p)$.

$\mathcal{X} := \{0,1\}^n$.

Sei $T_n \; := \; T_n(X_1, \ldots, X_n) \; := \; \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} X_j$.

$$\mathbb{E}_p(T_n) \; = \; p, \quad \mathbb{V}_p(T_n) \; = \; \frac{p(1-p)}{n} \; \leq \; \frac{1}{4n}.$$

Sei $0 < \alpha < 1$, z.B. $\alpha = 0.05$. Es gilt für jedes $p \in [0,1]$ (Tschebyschow-Ungl.)

$$\begin{aligned} \mathbb{P}_p\left(T_n - \frac{1}{2\sqrt{\alpha n}} \leq p \leq T_n + \frac{1}{2\sqrt{\alpha n}}\right) \quad &= \quad \mathbb{P}_p\left(|T_n - p| \leq \frac{1}{2\sqrt{\alpha n}}\right) \\ &\geq \quad 1 - \frac{\mathbb{V}_p(T_n)}{\left(1/(2\sqrt{\alpha n})\right)^2} \\ &= \quad 1 - \frac{p(1-p)4\alpha n}{n} \\ &\geq \quad 1 - \alpha. \end{aligned}$$

Memo: $\mathbb{P}_p\left(T_n - \frac{1}{2\sqrt{\alpha n}} \leq p \leq T_n + \frac{1}{2\sqrt{\alpha n}}\right) \;\geq\; 1-\alpha$

Für das zufällige Intervall $I_n = I_n(X_1, \ldots, X_n)$ mit

$$I_n \;:=\; \left[\max\left(0, T_n - \frac{1}{2\sqrt{\alpha n}}\right), \min\left(1, T_n + \frac{1}{2\sqrt{\alpha n}}\right)\right] \;\subseteq [0,1]$$

gilt somit $\mathbb{P}_p(I_n \ni p) \;\geq\; 1-\alpha \;\; \forall\, p \in [0,1]$.

22.2 Definition (Konfidenzbereich)

Seien $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $\alpha \in (0,1)$, z.B. $\alpha = 0.05$.

Eine Abbildung

$$\mathcal{C} : \mathcal{X} \to \mathcal{P}(\Theta)$$

heißt Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$, falls gilt:

$$\mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \;\geq\; 1-\alpha \qquad \forall\, \vartheta \in \Theta. \tag{22.7}$$

Memo: $\mathcal{C} : \mathcal{X} \to \mathcal{P}(\Theta)$ Konfidenzbereich für ϑ zum Niveau $1-\alpha$

$$:\Longleftrightarrow \mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \geq 1-\alpha \ \forall \vartheta \in \Theta \qquad (22.3)$$

- Auch synonym: Vertrauensbereich und Vertrauenswahrscheinlichkeit
- Sind $\Theta \subseteq \mathbb{R}$ und für jedes $x \in \mathcal{X}$ die Menge $\mathcal{C}(x)$ ein Intervall, so spricht man auch von einem Konfidenzintervall oder Vertrauensintervall.

Beachte:

- $\mathcal{C}$ ist eine $\mathcal{P}(\Theta)$-wertige Zufallsvariable auf $\mathcal{X}$
  $\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\} \ = \ \{$„$\mathcal{C}$ überdeckt ϑ“$\}$
- ϑ ist nicht zufällig, sondern – vor Durchführung des Experiments – die Realisierung x und damit $\mathcal{C}(x)$
- $\mathbb{P}_\vartheta(\mathcal{C}(X) \ni \vartheta) \geq 1-\alpha \quad \forall \vartheta \in \Theta$.
- $\mathcal{C}(X)$ wird bei wiederholter unabhängiger Durchführung unter gleichen Bedingungen in mindestens $(1-\alpha)\cdot 100\%$ aller Fälle ϑ enthalten
- $\mathcal{C}(x) := \Theta \ \forall x \in \mathcal{X}$ erfüllt (22.7), ist aber sinnlos! Ziel: „Kleines $\mathcal{C}(x)$“

22.3 Allgemeines Konstruktionsprinzip

Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell.

- Wähle zu jedem $\vartheta \in \Theta$ eine Menge $A(\vartheta) \subseteq \mathcal{X}$ mit
$$\mathbb{P}_\vartheta(A(\vartheta)) \ \geq \ 1-\alpha. \tag{22.8}$$
- Setze $\mathcal{C}(x) := \{\vartheta \in \Theta : x \in A(\vartheta)\}, \quad x \in \mathcal{X}.$

Dann gilt $x \in A(\vartheta) \iff \vartheta \in \mathcal{C}(x) \quad \forall\, x \in \mathcal{X}\ \forall\, \vartheta \in \Theta.$

Aus (22.8) folgt $\mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \ \geq \ 1-\alpha \qquad \forall\, \vartheta \in \Theta$

D.h.: $\mathcal{C}$ ist Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$.

Ziel: $A(\vartheta)$ „klein“ $\implies$ $\mathcal{C}(x)$ „klein“.

22.4 Beispiel (Binomialverteilung)

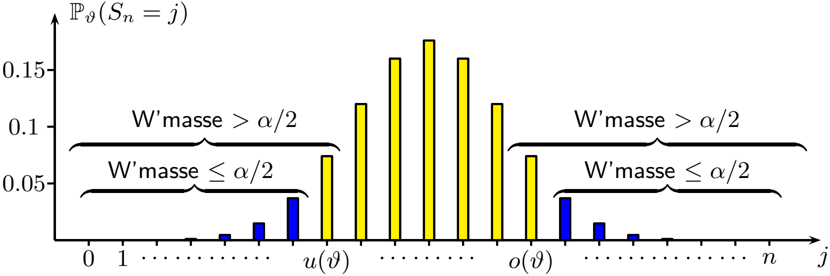
Sei $S_n \sim \mathrm{Bin}(n,p)$, $\vartheta := p \in \Theta = (0,1)$, $\mathbb{P}_\vartheta(S_n = j) = \binom{n}{j}\vartheta^j(1-\vartheta)^{n-j}$.

$\mathcal{X} = \{0, 1, \ldots, n\}$

$A(\vartheta) := \{x \in \mathcal{X} : u(\vartheta) \le x \le o(\vartheta)\}$, wobei für $0 < \alpha < 1$

$$u(\vartheta) \quad := \quad \max\left\{k \in \mathcal{X} : \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$$

$$o(\vartheta) \quad := \quad \min\left\{k \in \mathcal{X} : \sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$$



Beachte: Es gilt $u(\vartheta) \le o(\vartheta)$ sowie $\mathbb{P}_\vartheta(A(\vartheta)) \ge 1 - \alpha$.

Memo: $u(\vartheta) = \max\left\{k \in \mathcal{X} : \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$

Es gilt für $k \in \{1, \ldots, n\}$ (differenzieren!)

$$\sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \;=\; 1 - \binom{n}{k} k \int_0^{\vartheta} t^{k-1} (1-t)^{n-k} \, \mathrm{d}t \qquad (22.9)$$

Rechte Seite von (22.9) stetig und streng monoton fallend in ϑ

$\Longrightarrow\;$ $u : [0,1] \to \{0, 1, \ldots, n\}$ monoton wachsend,

$u(\cdot)$ rechtsseitig stetig.

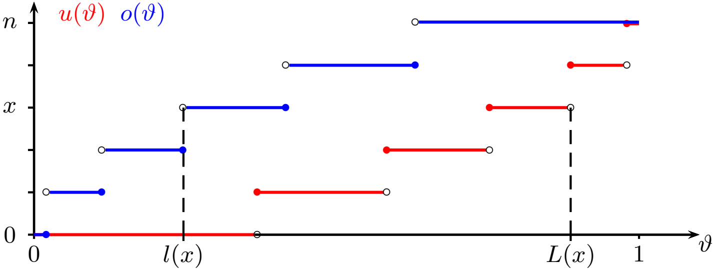
Memo: $o(\vartheta) = \min\left\{k \in \mathcal{X} : \sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$

Es gilt für $k \in \{0, \ldots, n-1\}$

$$\sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \;=\; \binom{n}{k} (n-k) \int_0^{\vartheta} t^k (1-t)^{n-k-1} \, \mathrm{d}t$$

$\Longrightarrow$ $o : [0,1] \to \{0, 1, \ldots, n\}$ monoton wachsend,

$o(\cdot)$ ist linksseitig stetig.



Memo: $A(\vartheta) := \{x \in \mathcal{X} : u(\vartheta) \leq x \leq o(\vartheta)\}$

Sei $l(x) := \inf\{\vartheta \in \Theta : o(\vartheta) = x\}, \quad L(x) := \sup\{\vartheta \in \Theta : u(\vartheta) = x\}$

Es gilt $x \in A(\vartheta) \iff l(x) < \vartheta < L(x)$ (!)

Somit ist

$$\mathcal{C}(x) \; := (l(x), L(x)), \qquad x \in \mathcal{X},$$

ein Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$.

$l(x)$ ($L(x)$) heißt untere (obere) Konfidenzgrenze für ϑ (zur Konf.w' $1-\alpha$).

Memo: $l(x) \;=\; \inf\{\vartheta \in \Theta : o(\vartheta) = x\}, \quad L(x) \;=\; \sup\{\vartheta \in \Theta : u(\vartheta) = x\}$

Memo: $u(\vartheta) \;=\; \max\left\{k \in \mathcal{X} : \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$

Memo: $o(\vartheta) \;=\; \min\left\{k \in \mathcal{X} : \sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \frac{\alpha}{2}\right\}$

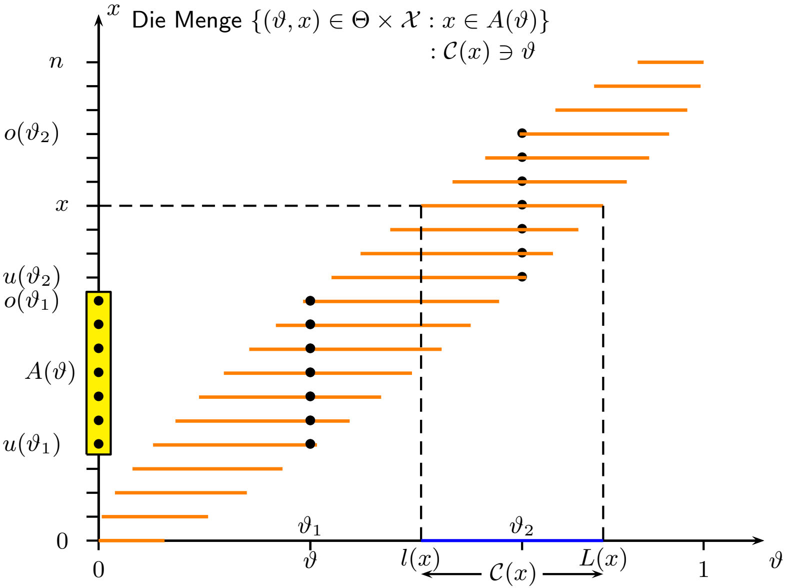
Für die Funktionen $l(\cdot)$ und $L(\cdot)$ gilt: (Übungsaufgabe!)

a) $l(0) = 0, \; L(0) = 1 - \left(\frac{\alpha}{2}\right)^{1/n}, \qquad l(n) = \left(\frac{\alpha}{2}\right)^{1/n}, \; L(n) = 1.$

b) Für $x \in \{1, 2, \ldots, n-1\}$ ist

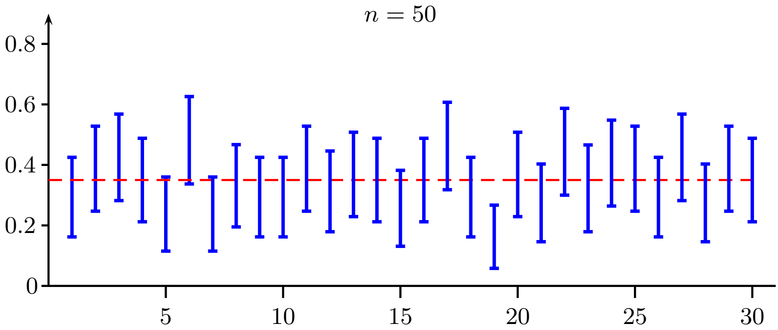
1) $l(x)$ die Lösung ϑ der Gleichung $\sum_{j=x}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \frac{\alpha}{2}$

2) $L(x)$ die Lösung ϑ der Gleichung $\sum_{j=0}^{x} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \frac{\alpha}{2}$.


x
Die Menge $\{(\vartheta, x) \in \Theta \times \mathcal{X} : x \in A(\vartheta)\}$
$: \mathcal{C}(x) \ni \vartheta$
n
$o(\vartheta_2)$
x
$u(\vartheta_2)$
$o(\vartheta_1)$
$A(\vartheta)$
$u(\vartheta_1)$
0
ϑ_1
ϑ_2
0
ϑ
$l(x)$
$\mathcal{C}(x)$
$L(x)$
1
ϑ

x	$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	$l(x)$	$L(x)$	$l(x)$	$L(x)$	$l(x)$	$L(x)$	$l(x)$	$L(x)$
0	0.000	0.168	0.000	0.116	0.000	0.088	0.000	0.071
1	0.001	0.249	0.001	0.172	0.001	0.132	0.001	0.106
2	0.012	0.317	0.008	0.221	0.006	0.169	0.005	0.137
3	0.032	0.379	0.021	0.265	0.016	0.204	0.013	0.165
4	0.057	0.437	0.038	0.307	0.028	0.237	0.022	0.192
5	0.087	0.491	0.056	0.347	0.042	0.268	0.033	0.218
6	0.119	0.543	0.077	0.386	0.057	0.298	0.045	0.243
7	0.154	0.592	0.099	0.423	0.073	0.328	0.058	0.267
8	0.191	0.639	0.123	0.459	0.091	0.356	0.072	0.291
9	0.231	0.685	0.147	0.494	0.108	0.385	0.086	0.314
10	0.272	0.728	0.173	0.528	0.127	0.412	0.100	0.337
11	0.315	0.769	0.199	0.561	0.146	0.439	0.115	0.360
12	0.361	0.809	0.227	0.594	0.166	0.465	0.131	0.382
13	0.408	0.846	0.255	0.626	0.186	0.491	0.146	0.403
14	0.457	0.881	0.283	0.657	0.206	0.517	0.162	0.425
15	0.509	0.913	0.313	0.687	0.227	0.542	0.179	0.446
16	0.563	0.943	0.343	0.717	0.249	0.567	0.195	0.467
17	0.621	0.968	0.374	0.745	0.270	0.591	0.212	0.488
18	0.683	0.988	0.406	0.773	0.293	0.615	0.229	0.508
19	0.751	0.999	0.439	0.801	0.315	0.639	0.247	0.528
20	0.832	1.000	0.472	0.827	0.338	0.662	0.264	0.548

Binomialverteilung: Konfidenzgrenzen für p, $1 - \alpha = 0.95$



Konkrete Konfidenzintervalle für p $(1 - \alpha = 0.95)$

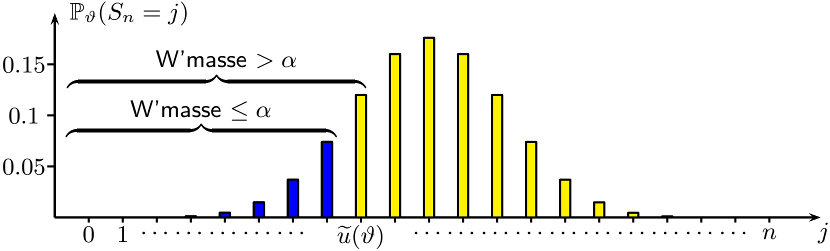
22.5 Beispiel (Binomialverteilung, einseitige Konfidenzbereiche)

Sei $S_n \sim \mathrm{Bin}(n,p)$, $\vartheta := p \in \Theta = (0,1)$, $\mathbb{P}_\vartheta(S_n = j) = \binom{n}{j}\vartheta^j(1-\vartheta)^{n-j}$.

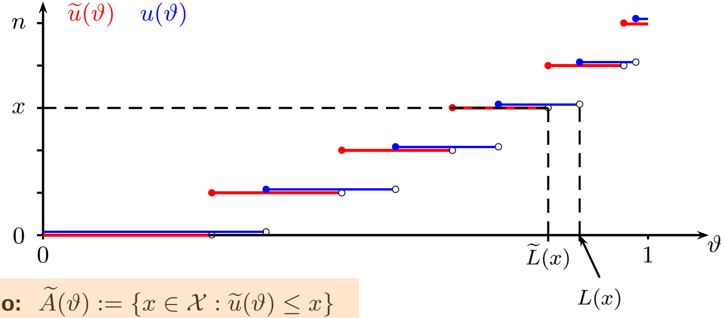
$\mathcal{X} = \{0,1,\ldots,n\}$. Jetzt nur Interesse an oberer Konfidenzgrenze für p.

$\widetilde{A}(\vartheta) := \{x \in \mathcal{X} : \widetilde{u}(\vartheta) \le x\}$, wobei für $0 < \alpha < 1$

$$\widetilde{u}(\vartheta) \; := \; \max\left\{ k \in \mathcal{X} : \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \le \alpha \right\}$$



Beachte: $\widetilde{u}(\vartheta) \ge u(\vartheta)$



Memo: $\widetilde{A}(\vartheta) := \{x \in \mathcal{X} : \widetilde{u}(\vartheta) \le x\}$

$$\text{Sei } \widetilde{L}(x) := \sup\{\vartheta \in \Theta : \widetilde{u}(\vartheta) = x\}$$

Es gilt $x \in \widetilde{A}(\vartheta) \iff \vartheta < \widetilde{L}(x)$ (!)

Somit ist

$$\widetilde{\mathcal{C}}(x) := [0, \widetilde{L}(x)), \qquad x \in \mathcal{X},$$

ein Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1 - \alpha$.

Beachte: Im Vgl. zu $\mathcal{C}(x) = (l(x), L(x))$ wird Aussage „nach oben genauer.“

Memo: $L(0) = 1 - \left(\frac{\alpha}{2}\right)^{1/n}, \; L(n) = 1$

Memo: $1 \leq x \leq n-1 \Longrightarrow L(x)$ ist Lsg. von $\sum_{j=0}^{x} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \frac{\alpha}{2}$.

In gleicher Weise gilt:

- $\widetilde{L}(0) = 1 - \alpha^{1/n}, \; \widetilde{L}(n) = 1$
- Für $x \in \{1, \ldots, n-1\}$ ist $\widetilde{L}(x)$ die Lösung ϑ von

$$\sum_{j=0}^{x} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \alpha.$$

22.6 Beispiel (Poisson-Verteilung)

$X_1, \ldots, X_n$ unabhängig, je $\sim$ Po(λ), $\vartheta := \lambda \in \Theta = (0, \infty)$.

$Y_n := \sum_{j=1}^n X_j \sim$ Po$(n\lambda)$, $\quad \mathcal{X} := \mathbb{N}_0$,

$$p_\mu(k) \ := \ \mathrm{e}^{-\mu} \, \frac{\mu^k}{k!}, \quad \mu \in (0, \infty), \ k \in \mathbb{N}_0$$

$$u(\vartheta) := \max\left\{ k \in \mathcal{X} : \sum_{j=0}^{k-1} p_{n\lambda}(j) \le \frac{\alpha}{2} \right\}, \quad o(\vartheta) := \min\left\{ k \in \mathcal{X} : \sum_{j=k+1}^{\infty} p_{n\lambda}(j) \le \frac{\alpha}{2} \right\}$$

Sei $l(y) := \inf\{\vartheta > 0 : o(\vartheta) = y\}, \quad L(y) := \sup\{\vartheta > 0 : u(\vartheta) = y\}, \quad y \in \mathcal{X}$.

$$\Longrightarrow \ l(y) < \vartheta < L(y) \ \Longleftrightarrow \ u(\vartheta) \le y \le o(\vartheta), \quad y \in \mathcal{X}, \ \vartheta > 0.$$

Also: $\mathbb{P}_\vartheta \left(l(Y_n) < \vartheta < L(Y_n) \right) \ \ge \ 1 - \alpha \qquad \forall \vartheta > 0.$

Alles analog zum Binomialfall, mit Integraldarstellung

$$\sum_{j=k}^{\infty} \mathrm{e}^{-u} \, \frac{u^j}{j!} \ = \ \frac{1}{(k-1)!} \int_0^u \mathrm{e}^{-t} t^{k-1} \, \mathrm{d}t, \qquad u \ge 0, \ k \ge 1.$$

Im Folgenden seien $X_1, X_2, \ldots$ stochastisch unabhängig und identisch verteilt mit Verteilung $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$.

Ist $\mathcal{X} \subseteq \mathbb{R}$ der Stichprobenraum für jedes einzelne X_j, so ist $\mathcal{X}^n \subseteq \mathbb{R}^n$ der Stichprobenraum für $(X_1, \ldots, X_n)$.

22.7 Definition (Asymptotischer Konfidenzbereich)

Für jedes $n \in \mathbb{N}$ sei $\mathcal{C}_n : \mathcal{X}^n \to \mathcal{P}(\Theta)$ eine Abbildung.

Die Folge $(\mathcal{C}_n)$ heißt asymptotischer Konfidenzbereich für ϑ zum Niveau $1-\alpha$, falls gilt:

$$\lim_{n\to\infty} \mathbb{P}_\vartheta\left(\{x \in \mathcal{X}^n : \mathcal{C}_n(x) \ni \vartheta\}\right) \;=\; 1-\alpha \quad \forall\, \vartheta \in \Theta.$$

Äquivalent:

$$\lim_{n\to\infty} \mathbb{P}_\vartheta\left(\{\mathcal{C}_n(X_1, \ldots, X_n) \ni \vartheta\}\right) \;=\; 1-\alpha \quad \forall\, \vartheta \in \Theta.$$

22.8 Asymptotische Konfidenzintervalle für p bei Bin(n,p)

Sei $S_n \sim$ Bin(n,p). ZGWS von de Moivre-Laplace $\Longrightarrow$

$$\lim_{n\to\infty} \mathbb{P}_p\left(-h \le \frac{S_n - np}{\sqrt{np(1-p)}} \le h\right) \;=\; \Phi(h) - \Phi(-h), \qquad h > 0.$$

Beachte: Mit $T_n := S_n/n$ folgt

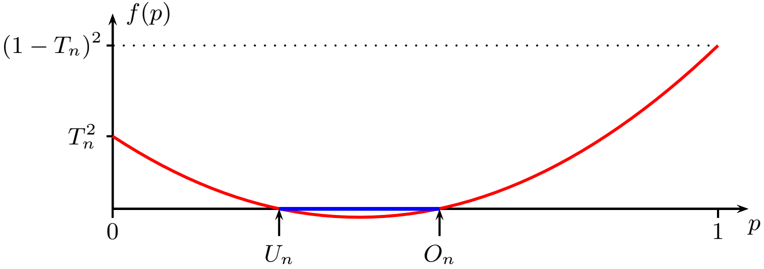
$$\begin{aligned}
-h \le \frac{S_n - np}{\sqrt{np(1-p)}} \le h \quad & \Longleftrightarrow \quad \left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \le h \\
& \Longleftrightarrow \quad \left(\frac{S_n - np}{\sqrt{np(1-p)}}\right)^2 \le h^2 \\
& \Longleftrightarrow \quad (S_n - np)^2 \le h^2 np(1-p) \\
& \Longleftrightarrow \quad p^2(n^2 + h^2 n) - p(2nS_n + h^2 n) + S_n^2 \le 0 \\
& \Longleftrightarrow \quad p^2\left(1 + \frac{h^2}{n}\right) - p\left(2T_n + \frac{h^2}{n}\right) + T_n^2 \le 0
\end{aligned}$$

Memo: $-h \le \dfrac{S_n - np}{\sqrt{np(1-p)}} \le h \iff p^2\left(1+\dfrac{h^2}{n}\right) - p\left(2T_n + \dfrac{h^2}{n}\right) + T_n^2 \le 0$

Zum Auflösen nach p Nullstellen von

$$f(p) \; := \; p^2\left(1+\frac{h^2}{n}\right) - p\left(2T_n + \frac{h^2}{n}\right) + T_n^2$$

bestimmen!



Also: $\quad -h \le \dfrac{S_n - np}{\sqrt{np(1-p)}} \le h \iff U_n \le p \le O_n$

Es gilt

$$U_n = \frac{T_n + \frac{h^2}{2n} - \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}$$

$$O_n = \frac{T_n + \frac{h^2}{2n} + \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}$$

und damit

$$\lim_{n\to\infty} \mathbb{P}_p(U_n \le p \le O_n) = \Phi(h) - \Phi(-h) = 2\Phi(h) - 1.$$

Setze

$$2\Phi(h) - 1 \overset{!}{=} 1 - \alpha \implies \Phi(h) = 1 - \frac{\alpha}{2} \implies h = h_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

$1-\alpha$	0.9	0.95	0.975	0.99
$\Phi^{-1}(1-\alpha/2)$	1.645	1.960	2.326	2.576

Memo: $O_n(U_n) = \dfrac{T_n + \frac{h^2}{2n} \pm \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}$

Unter Vernachlässigung aller Terme der Größenordnung $1/n$ erhält man:

$$U_n^* := T_n - \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)}, \quad O_n^* := T_n + \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)}$$

Mit $h := h_\alpha := \Phi^{-1}(1-\alpha/2)$ gilt (ohne Beweis)

$$\lim_{n\to\infty} \mathbb{P}_p\left(U_n^* \leq p \leq O_n^*\right) = 1-\alpha \quad \forall p \in (0,1)$$

$$\Longrightarrow \quad \mathcal{C}_n(x_1,\ldots,x_n) := [U_n^*(x_1,\ldots,x_n), O_n^*(x_1,\ldots,x_n)]$$

ist asymptotischer Konfidenzbereich für p zum Niveau $1-\alpha$.

Beachte: $O_n^* - U_n^* = \dfrac{2h}{\sqrt{n}}\sqrt{T_n(1-T_n)} \leq \dfrac{h}{\sqrt{n}} \Longrightarrow$

Faustregel: Länge des Konfidenzintervalls halbieren $\Longrightarrow$ n vervierfachen!

Planung von n, um Länge des Konfidenzintervalls zu begrenzen!

Beachte: Aus

$$\lim_{n\to\infty} \mathbb{P}_p\left(-\widetilde{h} \le \frac{S_n - np}{\sqrt{np(1-p)}} \right) \;=\; 1 - \Phi(-\widetilde{h}) \;=\; \Phi(\widetilde{h})$$

folgt wegen

$$-\widetilde{h} \le \frac{S_n - np}{\sqrt{np(1-p)}} \iff p \le \widetilde{O}_n$$

mit

$$\widetilde{O}_n := \frac{T_n + \frac{\widetilde{h}^2}{2n} + \frac{\widetilde{h}}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{\widetilde{h}^2}{4n}}}{1 + \frac{\widetilde{h}^2}{n}}$$

(größere der beiden Nullstellen einer quadratischen Gleichung für p)

die asymptotische einseitige Konfidenzaussage

$$\lim_{n\to\infty} \mathbb{P}_p(p \le \widetilde{O}_n) \;=\; \Phi(\widetilde{h}).$$

Es ist $\Phi(1.645) = 0.95$.

Der Wert h in O_n ist bei gleicher Konfidenzwahrscheinlichkeit 0.95 gleich 1.96.

$$\Longrightarrow \; \widetilde{O}_n < O_n.$$

22.9 Beispiel (Genauigkeit der Aussagen beim "ZDF-Politbarometer")

http://www.forschungsgruppe.de/Rund_um_die_Meinungsforschung/
Methodik_Politbarometer/methodik_1.pdf

> *.... ergeben sich bei einem Stichprobenumfang von n = 1.250 folgende Vertrauensbereiche: Der Fehlerbereich beträgt bei einem Parteianteil von 40 Prozent rund +/- drei Prozentpunkte und bei einem Parteianteil von 10 Prozent rund +/- zwei Prozentpunkte.*

Vereinfachendes Binomial-Urnenmodell:

In einer Urne sei für jeden von N Wahlberechtigten eine Kugel, davon r rote (Präferenz für Partei A). Von Interesse ist $p := r/N$.

Aus Urne rein zufällige Stichprobe vom Umfang n ziehen.

$X_j \ := \ \mathbf{1}\{j\text{-ter Befragter präferiert Partei A } (j\text{-te Kugel rot})\}, \quad j = 1, \ldots, n.$

Obwohl das Ziehen ohne Zurücklegen erfolgt, arbeiten wir mit dem Modell

$$X_1, \ldots, X_n \text{ unabhängig und je } \mathrm{Bin}(1, p)\text{-verteilt},$$

da N im Vergleich zu n sehr groß ist.

Memo: $U_n^* := T_n - \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)}, \quad O_n^* := T_n + \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)}$

Sei $1-\alpha = 0.95 \implies h = h_\alpha = \Phi^{-1}(0.975) = 1.96.$

Ein approximatives Konfidenzintervall für p aufgrund der zufälligen relativen Trefferhäufigkeit T_n (Anteil der Partei-A-Anhänger unter den Befragten) ist

$$\left[T_n - \frac{1.96}{\sqrt{n}}\,\sqrt{T_n(1-T_n)},\ T_n + \frac{1.96}{\sqrt{n}}\,\sqrt{T_n(1-T_n)}\right]$$

Die *halbe* Länge dieses Intervalls ist bei $n = 1250$

$$\frac{1.96}{\sqrt{1250}}\,\sqrt{T_n\,(1-T_n)} \;=\; \begin{cases} 0.027\ldots \text{ bei } T_n = 0.4 \\ 0.017\ldots \text{ bei } T_n = 0.1 \end{cases}$$

.... ergeben sich bei einem Stichprobenumfang von n = 1.250 folgende Vertrauensbereiche: Der Fehlerbereich beträgt bei einem Parteianteil von 40 Prozent rund +/- drei Prozentpunkte und bei einem Parteianteil von 10 Prozent rund +/- zwei Prozentpunkte.

22.10 Beispiel (Die Randomized-Response-Technik)

Wie gewinnt man Antworten auf heikle Fragen?

Haben Sie schon einmal Rauschgift genommen?

Ist auf dieser Karte eine Eins?

1

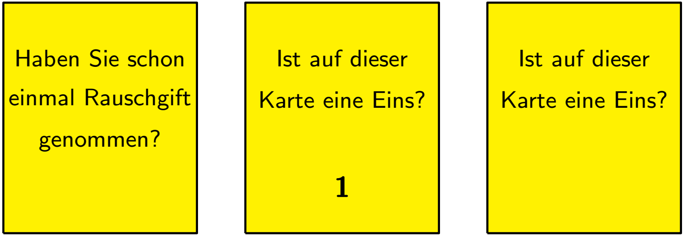
Ist auf dieser Karte eine Eins?

- Jede Karte ist gleich oft in einem Kartenstapel vertreten.
- Man zieht rein zufällig eine Karte.
- Man beantwortet die darauf stehende Frage wahrheitsgemäß.
- Man legt die Karte zurück und mischt gut durch.
- Anonymität ist gewährleistet!

Modell: $X_j = 1(0)$, falls j-ter Befragter mit Ja (Nein) antwortet.

Annahme: $X_1, \ldots, X_n$ unabhängig und identisch verteilt.

Sei p die (unbekannte) W', dass eine rein zufällig gewählte Person schon einmal Rauschgift genommen hat.



Haben Sie schon einmal Rauschgift genommen?

Ist auf dieser Karte eine Eins?

1

Ist auf dieser Karte eine Eins?

Sei $K_i := \{\text{von links gesehen } i\text{-te Karte gezogen}\} \Longrightarrow$

$$\begin{aligned} \mathbb{P}(K_i) &= 1/3, \quad i = 1,2,3, \\ \mathbb{P}(X_j = 1|K_1) &= p, \quad \mathbb{P}(X_j = 1|K_2) \ = \ 1, \quad \mathbb{P}(X_j = 1|K_3) = 0. \end{aligned}$$

Formel von der totalen Wahrscheinlichkeit $\Longrightarrow$

$$\mathbb{P}(X_j = 1) \;=\; \sum_{i=1}^{3} \mathbb{P}(X_j = 1 | K_i) \cdot \mathbb{P}(K_i) \;=\; \frac{p+1}{3}.$$

D.h.: $X_1, \ldots, X_n$ stochastisch unabhängig, je $\sim$ Bin $\left(1, \frac{p+1}{3}\right)$

Sei

$$T_n \;:=\; \overline{X}_n \;=\; \frac{1}{n} \sum_{j=1}^{n} X_j$$

(relativer Anteil der Ja-Antworten unter n Befragten)

T_n ist ein sinnvoller Schätzer für $\frac{p+1}{3}$:

$$\mathbb{E}_p(T_n) \;=\; \frac{p+1}{3}, \quad \mathbb{V}_p(T_n) = \frac{1}{n} \cdot \frac{p+1}{3} \cdot \left(1 - \frac{p+1}{3}\right)$$

Memo: $T_n := \overline{X}_n, \quad \mathbb{E}_p(T_n) = \frac{p+1}{3}, \quad \mathbb{V}_p(T_n) = \frac{1}{n}\,\frac{p+1}{3}\,\left(1 - \frac{p+1}{3}\right)$

Setze $\widehat{p}_n := 3T_n - 1 \Longrightarrow$

$$\mathbb{E}_p(\widehat{p}_n) = 3\,\mathbb{E}_p(T_n) - 1 = 3 \cdot \frac{p+1}{3} - 1 = p$$

$\Longrightarrow$ $\widehat{p}_n$ ist ein erwartungstreuer Schätzer für p.

$$\begin{aligned}\mathbb{V}_p(\widehat{p}_n) &= 9\mathbb{V}_p(T_n) = \frac{9}{n}\,\frac{p+1}{3}\left(1 - \frac{p+1}{3}\right)\\ &= \frac{p(1-p)+2}{n}\end{aligned}$$

$\Longrightarrow$ $(\widehat{p}_n)$ ist konsistente Schätzfolge für p.

Memo: $X_1, \ldots, X_n$ stochastisch unabhängig, je $\sim$ Bin $\left(1, \frac{p+1}{3}\right)$

Folg.: Für jedes p mit $0 < p < 1$ gilt

$$\lim_{n\to\infty} \mathbb{P}_p \left(T_n - \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)} \le \frac{p+1}{3} \le T_n + \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)} \right) = 0.95$$

Nach p auflösen ergibt

$$\mathbb{P}_p \left[3\left(T_n - \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)}\right) - 1 \le p \le 3\left(T_n + \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)}\right) - 1 \right] \to 0.95$$

Wegen $\widehat{p}_n = 3T_n - 1$ folgt

$$\mathbb{P}_p \left[\widehat{p}_n - 3 \cdot \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)} \le p \le \widehat{p}_n + 3 \cdot \frac{1.96}{\sqrt{n}} \sqrt{T_n(1-T_n)} \right] \to 0.95$$

Zahlenbeispiel: $n = 2500$, $T_n = \frac{912}{2500} \Longrightarrow \widehat{p}_n = \frac{236}{2500} = 0.0944 \Longrightarrow$

konkretes Konfidenzintervall für p: $[0.0378, 0.1510]$

! Konfidenzintervall $\widehat{p}_n \pm \frac{1.96}{\sqrt{n}} \sqrt{\widehat{p}_n(1-\widehat{p}_n)}$ führt auf $[0.0829, 0.1059]$

23 Induktive Statistik: Statistische Tests

23.1 Beispiel (Die „tea tasting lady“)

- Eine englische Lady trinkt regelmäßig Tee mit Milch.
- Sie behauptet, geschmacklich unterscheiden zu können, ob zuerst Milch oder zuerst Tee eingegossen wurde.
- Es werden ihr n mal zwei Tassen Tee gereicht. Davon ist eine vom Typ „erst Tee, dann Milch“, die andere vom Typ „erst Milch, dann Tee“.
- Sei ϑ die Wahrscheinlichkeit, dass die Lady die Eingießreihenfolge erkennt.
- $\vartheta = 1/2$: Lady rät blind.
- $\vartheta > 1/2$: Lady hat (mehr oder weniger große) geschmackliche Fähigkeiten.
- Sei $X_j = \mathbf{1}\{\text{Lady erkennt Reihenfolge beim } j\text{-ten Tassenpaar richtig}\}$.
- Annahme: $S_n := \sum_{j=1}^{n} X_j \sim \mathsf{Bin}(n, \vartheta)$.
- Sei $n = 20$. Ab wie vielen richtig erkannten Paaren würden Sie ihr geschmackliche Fähigkeiten attestieren?

Plausibles Verfahren: Sprich der Lady nur dann besondere geschmackliche Fähigkeiten zu, wenn sie von 20 Tassenpaaren mindestens k richtig zuordnet.

Wie wahrscheinlich ist das bei blindem Raten?

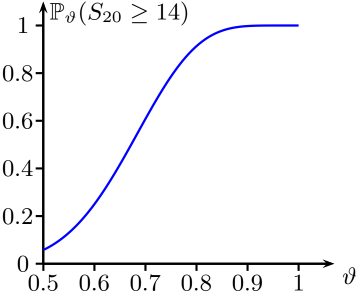
$$\mathbb{P}_{1/2}(S_{20} \geq k) \ = \sum_{j=k}^{20} \binom{20}{j} \left(\frac{1}{2}\right)^{j} \left(1 - \frac{1}{2}\right)^{20-j}.$$

$\mathbb{P}_{1/2}(S_{20} \geq 14) = 0.0576, \quad \mathbb{P}_{1/2}(S_{20} \geq 15) = 0.0207, \quad \mathbb{P}_{1/2}(S_{20} \geq 16) = 0.0059.$

Wähle z.B. $k = 14$:

W', dass besondere geschmackliche Fähigkeiten attestiert werden, hängt von ϑ ab:

$$\mathbb{P}_{\vartheta}(S_{20} \geq 14) = \sum_{j=14}^{20} \binom{20}{j} \vartheta^j (1-\vartheta)^{20-j}$$



23.2 Grundbegriffe (Hypothese und Alternative)

- Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell.
- Es gelte $\Theta = \Theta_0 + \Theta_1$, wobei $\Theta_0 \neq \emptyset$, $\Theta_1 \neq \emptyset$.
- Treffe aufgrund von $x \in \mathcal{X}$ eine „begründete Entscheidung“ zwischen

$$H_0 : \vartheta \in \Theta_0 \qquad \text{und} \qquad H_1 : \vartheta \in \Theta_1.$$

- **Sprechweise:** „Zu testen ist die Hypothese H_0 gegen die Alternative H_1“.

Asymmetrische Sichtweise, da die beiden möglichen Entscheidungen d_1 für H_1 und d_0 für H_0 unterschiedliche Auswirkungen haben können.

Im Fall der tea tasting lady ist $\Theta = [1/2, 1]$, $\Theta_0 = \{1/2\}$, $\Theta_1 = (1/2, 1]$ und

$$H_0 : \vartheta = \frac{1}{2}, \qquad H_1 : \vartheta > \frac{1}{2}.$$

d_1 attestiert der Lady besondere geschmackliche Fähigkeiten,

d_0 meint, dass sie nur blind rät.

23.3 Definition (Nichtrandomisierter Test)

Es seien $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $\Theta = \Theta_0 + \Theta_1$.

Ein nichtrandomisierter Test für das Testproblem $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ ist (formal) eine Indikatorfunktion $\mathbf{1}_{\mathcal{K}} : \mathcal{X} \to \mathbb{R}$.

Diese definiert die Entscheidungsregel

$$\text{Falls } \begin{cases} x \in \mathcal{K}, & \text{so Entscheidung } d_1, \\ x \in \mathcal{X} \setminus \mathcal{K}, & \text{so Entscheidung } d_0. \end{cases}$$

Die Menge $\mathcal{K} \subseteq \mathcal{X}$ heißt kritischer Bereich (des Tests).

Die Menge $\mathcal{X} \setminus \mathcal{K}$ heißt Annahmebereich (des Tests).

Sprechweisen:

- Im Fall $x \in \mathcal{K}$: die Hypothese H_0 wird verworfen

  bzw.: die Stichprobe x steht im Widerspruch zu H_0.

- Im Fall $x \in \mathcal{X} \setminus \mathcal{K}$: die Hypothese H_0 wird nicht verworfen

  bzw.: die Stichprobe x steht nicht im Widerspruch zu H_0.

23.4 Definition (Prüfgröße, kritischer Wert)

Der kritische Bereich $\mathcal{K} \subseteq \mathcal{X}$ ist oft von folgender Gestalt:

$$\mathcal{K} \;=\; \{x \in \mathcal{X} : T(x) \geq c\} \;=\; \{T \geq c\}$$

Dabei sind $T : \mathcal{X} \to \mathbb{R}$ eine Funktion und $c \in \mathbb{R}$.

T heißt Testgröße oder Prüfgröße. c heißt kritischer Wert.

- Also: H_0 ablehnen, falls die Prüfgröße einen kritischen Wert erreicht.
- Auch kritische Bereiche der Form $\{T > c\}$, $\{T < c\}$, $\{T \leq c\}$, $\{|T| > c\}$, $\{|T| \geq c\}$ usw. möglich.

23.5 Beispiel (Tea tasting lady)

$\mathcal{X} = \{0,1\}^n$, $x = (x_1, \ldots, x_n)$, $T(x) = x_1 + \ldots + x_n$

Prüfgröße: Anzahl der richtig zugeordneten Tassenpaare.

Der kritische Wert c ist noch festzulegen.

Bei jedem Test gibt es zwei Fehlermöglichkeiten:

23.6 Definition (Fehler erster und zweiter Art)

Ein Fehler 1. Art ist die fälschliche Ablehnung von H_0.

Ein Fehler 2. Art ist die fälschliche Ablehnung von H_1.

- Bei einem Fehler 1. Art gilt $\vartheta \in \Theta_0$, und man trifft die Entscheidung d_1,
- Bei einem Fehler 2. Art gilt $\vartheta \in \Theta_1$, und man trifft die Entscheidung d_0
- Beachte: Im Allg. ist $\mathbb{P}_\vartheta(\{x\}) > 0 \ \forall \vartheta \in \Theta \Longrightarrow$ Fehler unvermeidlich!

Im Fall der *tea tasting lady* bedeutet

- ein Fehler 1. Art das fälschliche Attestieren besonderer geschmacklicher Fähigkeiten
- ein Fehler 2. Art das fälschliche Attestieren von blindem Raten

		„Wirklichkeit“	
		$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
Entscheidung	d_0	richtige Entscheidung	Fehler 2. Art
	d_1	Fehler 1. Art	richtige Entscheidung

Wirkungstabelle eines Tests

- x kann als Realisierung von $X \ \ (= \mathrm{id}_{\mathcal{X}})$ sowohl von $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_0$ als auch von $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_1$ erzeugt worden sein
- $\Longrightarrow$ Fehlentscheidungen unvermeidbar!
- Ziel: Wahrscheinlichkeiten für Fehlentscheidungen durch „vernünftige“ Wahl eines Tests (kritischen Bereichs $\mathcal{K}$) klein halten!

23.7 Definition (Gütefunktion eines Tests)

Die durch $g := g_{\mathcal{K}} : \begin{cases} \Theta \longrightarrow [0,1] \\ \vartheta \longmapsto g(\vartheta) := \mathbb{P}_\vartheta(X \in \mathcal{K}) \end{cases}$

definierte Funktion heißt Gütefunktion des durch $\mathcal{K}$ gegebenen Tests.

- g ordnet jedem ϑ die Verwerfungswahrscheinlichkeit von H_0 unter $\mathbb{P}_\vartheta$ zu.
- Die Funktion $\Theta \ni \vartheta \mapsto 1 - g(\vartheta)$ heißt Operationscharakteristik.
- Ideal wäre: $g(\vartheta) = 0$ für jedes $\vartheta \in \Theta_0$ (d.h. nie Fehler 1. Art),

  $g(\vartheta) = 1$ für jedes $\vartheta \in \Theta_1$ (d.h. nie Fehler 2 Art)

Zwei triviale Tests erfüllen jeweils die Hälfte dieses Idealfalls:

- $\mathcal{K} = \mathcal{X}$ (ohne Ansehen der Daten stets ablehnen)

  $\Longrightarrow g(\vartheta) = 1$ für jedes $\vartheta \in \Theta$.
- $\mathcal{K} = \emptyset$ (ohne Ansehen der Daten nie ablehnen)

  $\Longrightarrow g(\vartheta) = 0$ für jedes $\vartheta \in \Theta$.

Vernünftige Tests lassen Fehler erster und zweiter Art zu, beschränken aber die Wahrscheinlichkeit für *einen* der beiden Fehler.

Konvention (und damit Festlegung von Hypothese H_0 und Alternative H_1):

Kontrolle der Wahrscheinlichkeit für einen Fehler *erster* Art

(dieser wird als schwerwiegender betrachtet als ein Fehler 2. Art)

23.8 Definition (Test zum Niveau α)

Es sei $\mathbf{1}_{\mathcal{K}}$ ein Test für $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ mit kritischem Bereich $\mathcal{K}$.

Weiter sei $\alpha \in (0, 1)$ (üblich: $\alpha \le 0.1$).

Der Test $\mathbf{1}_{\mathcal{K}}$ heißt Test zum Niveau α oder Niveau-α-Test, falls gilt:

$$g(\vartheta) \ = \ \mathbb{P}_\vartheta(X \in \mathcal{K}) \ \le \ \alpha \qquad \forall\, \vartheta \in \Theta_0.$$

Hierdurch wird erreicht, dass die Hypothese H_0 im Fall ihrer Gültigkeit auf die Dauer in höchstens $100 \cdot \alpha\%$ aller Fälle verworfen wird (Gesetz großer Zahlen).

Memo: Bei Niveau-α-Test: $g(\vartheta) \ = \ \mathbb{P}_\vartheta(X \in \mathcal{K}) \ \leq \ \alpha \qquad \forall\, \vartheta \in \Theta_0.$

Beachte:

- Lehnt ein Niveau α-Test H_0 ab, so kann man „praktisch sicher sein“, dass H_0 nicht gilt
  (sonst wäre man zur Entscheidung d_1 nur mit einer W' $\leq \ \alpha$ gelangt).
- Wird H_0 abgelehnt, so sagt man auch:

  die Ablehnung von H_0 ist signifikant zum Niveau α

  bzw. die Daten stehen auf dem $\alpha \cdot 100\%$-Niveau im Widerspruch zu H_0
- Wird H_0 aufgrund von $x \in \mathcal{X}$ nicht verworfen, so heißt dies nur, dass x bei einer zugelassenen W' α für einen Fehler 1. Art nicht im Widerspruch zu H_0 steht.
- Formulierungen wie „H_0 ist verifiziert“ oder „H_0 ist validiert“ sind hier völlig fehl am Platze.
- Sie suggerieren, dass man im Falle des Nicht-Verwerfens von H_0 die Gültigkeit von H_0 „bewiesen“ hätte, was jedoch blanker Unsinn ist!

Memo: Bei Niveau-α-Test: $g(\vartheta) = \mathbb{P}_\vartheta(X \in \mathcal{K}) \le \alpha \quad \forall\, \vartheta \in \Theta_0$.

Beachte:

- Intuitiv naheliegend: Bilde $\mathcal{K}$ aus denjenigen $x \in \mathcal{X}$, die unter H_0 am unwahrscheinlichsten sind.
- Die Wahl von α hängt davon ab, welcher Prozentsatz fälschlicher Ablehnungen von H_0 toleriert werden soll.
- Je kleiner α, umso *bedeutungsvoller* (*signifikanter*) ist im Fall einer Ablehnung von H_0 der erhaltene Widerspruch zu H_0.
- Ein kleiner Wert von α dient also der Sicherung der Alternative.
- Die W' für einen Fehler 2. Art eines Tests zum Niveau α hängt immer von der zugrunde liegenden Verteilung $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_1$ ab.
- Diese W' ist umso kleiner, je „verschiedener“ diese Verteilung von den Verteilungen $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_0$ ist.

23.9 Beispiel (Einseitiger Binomialtest)

- Eine Standardtherapie zur Behandlung einer bestimmten Krankheit habe eine Erfolgswahrscheinlichkeit von 0.5.
- Eine Forschergruppe hat eine neue Therapie entwickelt.
- Diese soll an einer Zufallsstichprobe von n Patienten aus der großen Population aller an dieser Krankheit Leidenden erprobt werden.

Stark vereinfachendes Modell (Bernoulli-Kette):

- $\mathcal{X} := \{0,1\}^n, \ \ x = (x_1, \ldots, x_n), \quad \Theta \ := \ [0,1]$
- $\mathbb{P}_\vartheta(\{x\}) := \vartheta^k(1-\vartheta^{n-k}, \ \ k := x_1 + \ldots + x_n$
- $T(x) := x_1 + \ldots + x_n, \ \ x = (x_1, \ldots, x_n) \in \mathcal{X}$
- $\Longrightarrow \mathbb{P}_\vartheta(T = k) \ = \ \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$
- $X \ := \ (X_1, \ldots, X_n) \ := \ \mathrm{id}_{\mathcal{X}}$
- Interpretation: $X_j = \ \mathbf{1}\{j$-ter Patient der Studie wird geheilt$\}$
- $T(X) = \sum_{j=1}^n X_j \sim \mathrm{Bin}(n, \vartheta)$ ($T(X) =$ zufällige Anz. der Heilerfolge)

Mögliche Fehlentscheidungen:

a) Behauptung der Überlegenheit der neuen Therapie ($\vartheta > 1/2$), obwohl diese in Wirklichkeit nicht besser ist als die Standard-Therapie ($\vartheta \leq 1/2$),

b) Nichtvertreten einer wahren Forschungshypotheses („Nichterkennen“ eines Wertes ϑ mit $\vartheta > 1/2$)

Fehler in a) schwerwiegender! (Warum?)

Also:

$$H_0 : \vartheta \leq \frac{1}{2} \qquad \left(\text{d.h.: } \Theta_0 = \left[0, \frac{1}{2}\right]\right)$$

$$H_1 : \vartheta > \frac{1}{2} \qquad \left(\text{d.h.: } \Theta_1 = \left(\frac{1}{2}, 1\right]\right).$$

Der kritische Bereich $\mathcal{K}$ ist plausiblerweise von der Gestalt

$$\mathcal{K} \ = \ \left\{x = (x_1, \ldots, x_n) \in \mathcal{X} : T(x) = \sum_{j=1}^{n} x_j \geq c\right\}$$

Beachte: Die folgenden Überlegungen sind auch auf das Beispiel der tea tasting lady anwendbar.

Festlegung des kritischen Wertes c mit Hilfe des vorgegebenen Testniveaus α:

Die Gütefunktion dieses Tests ist (mit $T_n := \sum_{j=1}^n X_j$)

$$\begin{aligned} g(\vartheta) &= g_{\mathcal{K}}(\vartheta) = \mathbb{P}_\vartheta\left(T_n \geq c\right) \\ &= \sum_{j=c}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \\ &= \frac{n!}{(c-1)!(n-c)!} \int_0^\vartheta x^{c-1}(1-x)^{n-c}\,\mathrm{d}x. \end{aligned}$$

$g(\cdot)$ ist streng monoton wachsend.

Also liegt genau dann ein Niveau α-Test vor, wenn gilt:

$$g\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^n \sum_{j=c}^{n} \binom{n}{j} \leq \alpha.$$

c sollte unter dieser Nebenbedingung möglichst klein sein! (Warum?)

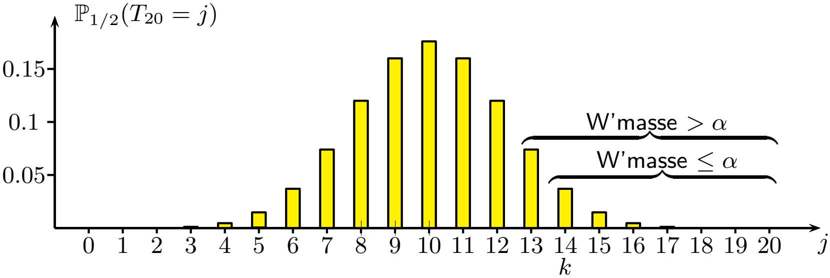
Damit die W' für einen Fehler 2. Art kleiner wird!

Memo: $g\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^n \sum_{j=c}^{n} \binom{n}{j} \le \alpha.$

$$c = c(n,\alpha) := \min\left\{k \in \{0,1,\ldots,n\} : \left(\frac{1}{2}\right)^n \sum_{j=k}^{n} \binom{n}{j} \le \alpha\right\}.$$

Zahlenbeispiel: $n = 20$, $\alpha = 0.1$.

$$\left(\frac{1}{2}\right)^{20} \sum_{j=14}^{20} \binom{20}{j} = 0.0577 \le 0.1, \qquad \left(\frac{1}{2}\right)^{20} \sum_{j=13}^{20} \binom{20}{j} = 0.1316 > 0.1$$



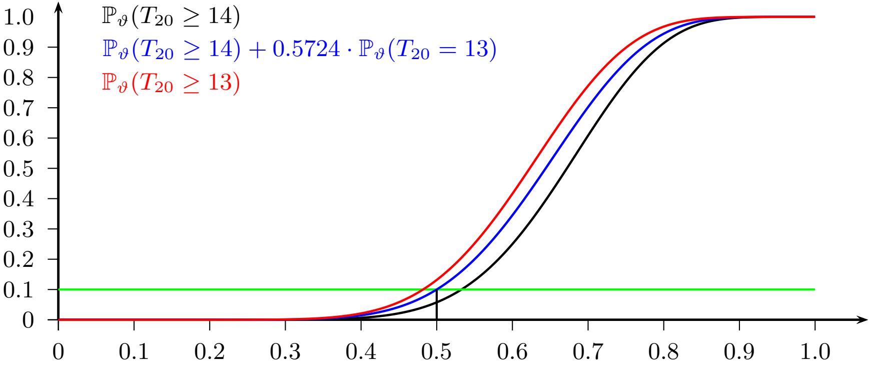
Also: H_0 genau dann zum Niveau 0.1 ablehnen, falls $T_{20} \geq 14$

Gütefunktion: $g_{20,14}(\vartheta) = \mathbb{P}_\vartheta(T_{20} \geq 14)$

Test schöpft erlaubte Fehlerwahrscheinlichkeit $\alpha = 0.1$ nicht voll aus!

Verbesserung durch Randomisierung möglich.

Zusätzlich mit W' 0.5724 ablehnen, falls $T_{20} = 13$ (mit Zufallszahlengenerator)



Wie kommt man auf die Randomisierungswahrscheinlichkeit 0.5724?

Antwort: Aus der Forderung $\mathbb{P}_{1/2}(T_{20} \geq 14) + \gamma \cdot \mathbb{P}_{1/2}(T_{20} = 13) \overset{!}{=} 0.1$

23.10 Definition (Randomisierter Test)

Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell mit $\Theta = \Theta_0 + \Theta_1$ wie in 23.2.

Ein randomisierter Test zum Testen von $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ ist eine Abbildung

$$\varphi : \mathcal{X} \to [0, 1],$$

wobei $\varphi(x) =$ Wahrscheinlichkeit für Ablehnung von H_0 aufgrund von $x \in \mathcal{X}$.

- Nichtrandomisierte Tests sind Spezialfälle randomisierter Tests ($\varphi = \mathbf{1}_\mathcal{K}$).
- Im Fall $0 < \varphi(x) < 1$ erfolgt die Testentscheidung mit Hilfe einer in $[0, 1]$ gleichverteilten Pseudozufallszahl y. Ablehnung von $H_0 \Longleftrightarrow y \leq \varphi(x)$.
- Nachteil der Randomisierung: Testergebnis hängt nicht nur von x ab, sondern von „externem Zufall“.
- Vorteil der Randomisierung: Volle Niveau-Ausschöpfung und damit kleinere W' für einen Fehler 2.Art.

Ein randomisierter Test ist oft von der Gestalt

$$\varphi(x) = \begin{cases} 1, & \text{falls } T(x) > c, \\ \gamma, & \text{falls } T(x) = c, \\ 0, & \text{falls } T(x) < c. \end{cases}$$

Dabei sind $T : \mathcal{X} \to \mathbb{R}$ eine Prüfgröße und $c \in \mathbb{R}$ ein kritischer Wert.

Der Wert $\gamma \in [0, 1]$ heißt Randomisierungswahrscheinlichkeit.

Es wird also nur dann randomisiert, wenn das Testergebnis sozusagen „auf der Kippe steht“.

23.11 Beispiel (Allgemeiner einseitiger Binomialtest)

$X_1, \ldots, X_n$ unabhängig und je Bin$(1, \vartheta)$-verteilt, $\Theta = [0,1]$, $\mathcal{X} = \{0,1\}^n$.

Sei $\vartheta_0 \in (0,1)$ gegeben, $\Theta_0 := [0, \vartheta_0]$, $\Theta_1 := (\vartheta_0, 1]$.

Also: Hypothese $H_0 : \vartheta \leq \vartheta_0$, Alternative $H_1 : \vartheta > \vartheta_0$.

Prüfgröße (Anzahl der Treffer):

$$T_n(x_1, \ldots, x_n) := \sum_{j=1}^{n} x_j, \quad (x_1, \ldots, x_n) \in \mathcal{X}, \quad T_n := \sum_{j=1}^{n} X_j$$

Sei $\alpha \in (0,1)$ die zugelassene W' für einen Fehler 1. Art.

H_0 ablehnen, falls $T_n > c$; eventuell noch randomisieren.

$k \in \{0, 1, \ldots, n-1\} \Longrightarrow \mathbb{P}_\vartheta(T_n > k)$ streng monoton wachsend in $\vartheta \Longrightarrow$

$$\sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta\left(T_n > k\right) = \mathbb{P}_{\vartheta_0}\left(T_n > k\right)$$

$$\begin{aligned} \text{Setze } c &:= \min\{k \in \{0, 1, \ldots, n\} : \mathbb{P}_{\vartheta_0}(T_n > k) \leq \alpha\} \\ \gamma &:= \frac{\alpha - \mathbb{P}_{\vartheta_0}(T_n > c)}{\mathbb{P}_{\vartheta_0}(T_n = c)} \quad (\Longrightarrow 0 \leq \gamma < 1) \end{aligned}$$

Memo: $\varphi(x) =$ W' für Ablehnung von H_0 aufgrund von $x \in \mathcal{X}$.

23.12 Definition (Gütefunktion eines randomisierten Tests)

Sei $\varphi : \mathcal{X} \to [0,1]$ ein randomisierter Test. Dann heißt die durch

$$g_\varphi(\vartheta) := \sum_{x \in \mathcal{X}} \varphi(x)\, \mathbb{P}_\vartheta(\{x\})$$

definierte Funktion $g_\varphi : \Theta \to [0,1]$ die Gütefunktion von φ.

Beachte:

- $\mathcal{X}$ sei abzählbar, sonst Summe über abzählbares $\mathcal{X}_0$ mit $\mathbb{P}_\vartheta(\mathcal{X}_0) = 1$
- Es ist $g_\varphi(\vartheta) = \mathbb{P}_\vartheta($„$H_0$ ablehnen“$)$
- Spezialfall nichtrandomisierter Test: $\varphi = \mathbf{1}_\mathcal{K}$
- Mit der Konvention $X := \mathrm{id}_\mathcal{X}$ ist

$$g_\varphi(\vartheta) = \sum_{x \in \mathcal{X}} \varphi(x)\, \mathbb{P}_\vartheta(X = x)$$

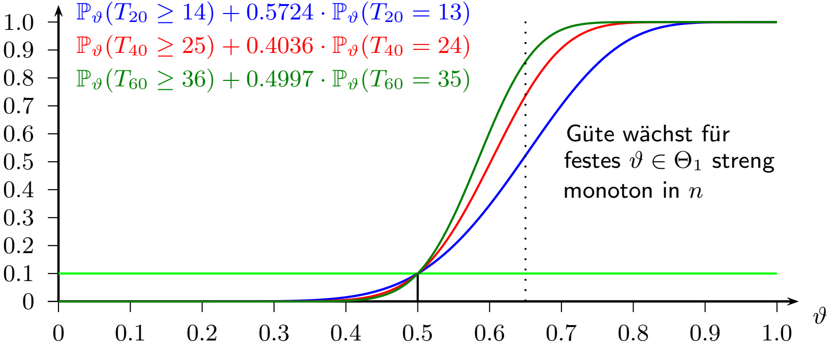
- Es ist $g_\varphi(\vartheta) = \mathbb{E}_\vartheta\, \varphi$

23.13 Beispiel (Gütefunktionen einseitiger Binomialtests)

In der Situation von 23.11 sei $\vartheta_0 = 0.5$, $\alpha = 0.1$.

Betrachte Test (zum Niveau α) für $n = 20$, $n = 40$ und $n = 60$:

$$\begin{aligned}\varphi_{20} &= \mathbf{1}\{T_{20} \geq 14\} + 0.5724 \cdot \mathbf{1}\{T_{20} = 13\},\\ \varphi_{40} &= \mathbf{1}\{T_{40} \geq 25\} + 0.4036 \cdot \mathbf{1}\{T_{40} = 24\},\\ \varphi_{60} &= \mathbf{1}\{T_{60} \geq 36\} + 0.4997 \cdot \mathbf{1}\{T_{60} = 35\}.\end{aligned}$$



Güte wächst für festes $\vartheta \in \Theta_1$ streng monoton in n

Zur Untersuchung der Eigenschaften von Testverfahren bei wachsendem Stichprobenumfang sind folgende Begriffsbildungen grundlegend:

23.14 Definition (asymptotisches Niveau, Konsistenz)

Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilt mit Verteilung $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$.

Sei $\mathcal{X}^n \subseteq \mathbb{R}^n$ der Stichprobenraum für $(X_1, \ldots, X_n)$.

Sei $\varphi_n : \mathcal{X}^n \to [0,1]$ ein Test für $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$, der auf den Realisierungen von $(X_1, \ldots, X_n)$ basiert. Dann heißt $(\varphi_n)_{n \geq 1}$ eine Testfolge.

a) Die Testfolge $(\varphi_n)_{n \in \mathbb{N}}$ besitzt asymptotisch das Niveau α, falls gilt:

$$\overline{\lim_{n\to\infty}} \sup_{\vartheta \in \Theta_0} g_{\varphi_n}(\vartheta) \leq \alpha.$$

b) Die Testfolge $(\varphi_n)_{n \in \mathbb{N}}$ heißt konsistent, falls gilt:

$$\lim_{n\to\infty} g_{\varphi_n}(\vartheta) = 1 \qquad \forall\, \vartheta \in \Theta_1.$$

Die Konsistenzeigenschaft ist eine Minimalforderung an eine Testfolge.

23.15 Beispiel (Einseitiger Binomialtest)

Seien $X_1, \ldots, X_n, \ldots$ unabhängig und je Bin$(1, \vartheta)$-verteilt, $\Theta = (0,1)$, $\mathcal{X}^n = \{0,1\}^n$, Hypothese $H_0 : \vartheta \le \vartheta_0$, Alternative $H_1 : \vartheta > \vartheta_0$.

Sei $\alpha \in (0,1)$ die zugelassene W' für einen Fehler 1. Art.

Betrachte Folge (φ_n) nichtrandomisierter Tests φ_n mit

$$\varphi_n(x_1, \ldots, x_n) := \mathbf{1}\left\{ \sum_{j=1}^n x_j > c_n \right\},$$

wobei $c_n \; = \; c_n(\vartheta_0, n, \alpha) \; := \; n\vartheta_0 + \sqrt{n\vartheta_0(1-\vartheta_0)} \cdot \Phi^{-1}(1-\alpha)$.

ZGWS von de Moivre-Laplace $\Longrightarrow$

$$\begin{aligned}
\lim_{n\to\infty} \sup_{\vartheta \in \Theta_0} g_{\varphi_n}(\vartheta) \quad &= \quad \lim_{n\to\infty} g_{\varphi_n}(\vartheta_0) \; = \; \lim_{n\to\infty} \mathbb{P}_{\vartheta_0}\left(\sum_{j=1}^n X_j > c_n \right) \\
&= \quad \lim_{n\to\infty} \mathbb{P}_{\vartheta_0}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_0}{\sqrt{n\vartheta_0(1-\vartheta_0)}} > \Phi^{-1}(1-\alpha) \right) \\
&= \quad 1 - \Phi\left(\Phi^{-1}(1-\alpha)\right) \; = \; \alpha
\end{aligned}$$

$\Longrightarrow$ Testfolge (φ_n) hat asymptotisches Niveau α.

Memo: $c_n = n\vartheta_0 + \sqrt{n\vartheta_0(1-\vartheta_0)} \cdot \Phi^{-1}(1-\alpha).$

Memo: H_0 ablehnen, falls $\sum_{j=1}^{n} X_j > c_n \left(\Longleftrightarrow \overline{X}_n > \frac{c_n}{n}\right)$

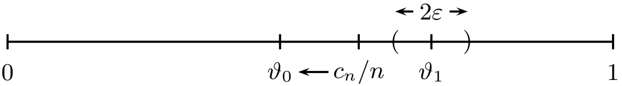
Konsistenz? Sei ϑ_1 mit $\vartheta_0 < \vartheta_1 < 1$ beliebig. Zu zeigen:

$$1 = \lim_{n\to\infty} g_{\varphi_n}(\vartheta_1) = \lim_{n\to\infty} \mathbb{P}_{\vartheta_1}\left(\sum_{j=1}^{n} X_j > c_n\right) = \lim_{n\to\infty} \mathbb{P}_{\vartheta_1}\left(\overline{X}_n > \frac{c_n}{n}\right)$$

Beachte: $c_n/n \downarrow \vartheta_0$.

Sei $\varepsilon > 0$ beliebig. Schwaches Gesetz großer Zahlen $\Longrightarrow$

$$\mathbb{P}_{\vartheta_1}\left(\left|\overline{X}_n - \vartheta_1\right| < \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 1. \tag{23.10}$$



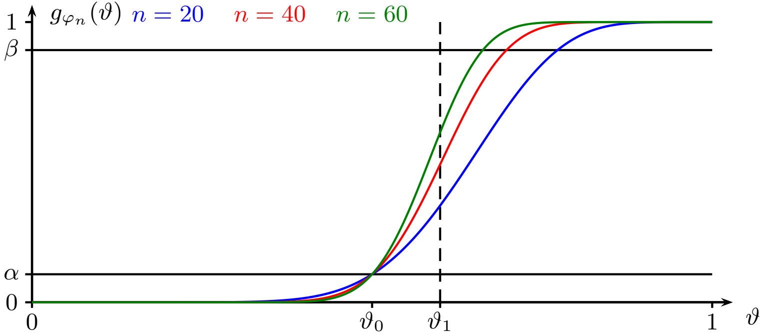
n hinreichend groß $\Longrightarrow \{\left|\overline{X}_n - \vartheta_1\right| < \varepsilon\} \subseteq \{\overline{X}_n > c_n/n\}$. (23.10) $\Longrightarrow$ Beh.

23.16 Beispiel (Planung des Stichprobenumfangs)

In der Situation von Beispiel 23.15 (einseitiger Binomialtest, $H_0 : \vartheta \leq \vartheta_0$, $H_1 : \vartheta > \vartheta_0$) sei ϑ_1 mit $\vartheta_1 > \vartheta_0$ fest gewählt.

Weiter sei β mit $\alpha < \beta < 1$ vorgegeben. Wie groß muss n mindestens sein, damit $g_{\varphi_n}(\vartheta_1) \geq \beta$ gilt?

Damit wäre die W' für einen Fehler 2. Art unter $\mathbb{P}_{\vartheta_1}$ höchstens $1 - \beta$.



Memo: $\varphi_n(x_1,\ldots,x_n) = \mathbf{1}\{x_1+\ldots+x_n > c_n\}$

Memo: $c_n = n\vartheta_0 + \Phi^{-1}(1-\alpha)\sqrt{n\vartheta_0(1-\vartheta_0)}$

Forderung: $\mathbb{P}_{\vartheta_1}\left(\sum_{j=1}^n X_j > c_n\right) \overset{!}{=} \beta \implies$

$$\begin{aligned}
\beta &\overset{!}{=} \mathbb{P}_{\vartheta_1}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}} > \frac{c_n - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}}\right)\\
&= \mathbb{P}_{\vartheta_1}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}} > \frac{\sqrt{n}(\vartheta_0-\vartheta_1) + \Phi^{-1}(1-\alpha)\sqrt{\vartheta_0(1-\vartheta_0)}}{\sqrt{\vartheta_1(1-\vartheta_1)}}\right)\\
&\approx 1 - \Phi\left(\sqrt{n}\frac{\vartheta_0-\vartheta_1}{\sqrt{\vartheta_1(1-\vartheta_1)}} + \Phi^{-1}(1-\alpha)\sqrt{\frac{\vartheta_0(1-\vartheta_0)}{\vartheta_1(1-\vartheta_1)}}\right)
\end{aligned}$$

Hiermit ergibt sich durch Auflösen nach n:

$$n \approx \frac{\vartheta_1(1-\vartheta_1)}{(\vartheta_0-\vartheta_1)^2}\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(1-\alpha)\sqrt{\frac{\vartheta_0(1-\vartheta_0)}{\vartheta_1(1-\vartheta_1]}}\right]^2$$

$$n \;\approx\; \frac{\vartheta_1(1-\vartheta_1)}{(\vartheta_0-\vartheta_1)^2}\left[\Phi^{-1}(1-\beta)-\Phi^{-1}(1-\alpha)\sqrt{\frac{\vartheta_0(1-\vartheta_0)}{\vartheta_1(1-\vartheta_1]}}\right]^2$$

Zahlenbeispiel:

$\vartheta_0 = 1/2,\; \vartheta_1 = 0.6,\; \alpha = 0.1,\; \beta = 0.9,\; \Phi^{-1}(0.1) = -\Phi^{-1}(0.9) = -1.282$

$\Longrightarrow n \approx 161$. Exakter Wert (mit MAPLE): $n = 163$.

Bemerkung: Falls $H_0 : \vartheta \geq \vartheta_0$ gegen $H_1 : \vartheta < \vartheta_0$ getestet wird, so:

$$H_0 \text{ ablehnen, falls } \sum_{j=1}^{n} X_j < c_n := n\vartheta_0 + \Phi^{-1}(\alpha)\sqrt{n\vartheta_0(1-\vartheta_0)}.$$

Die Forderung $\beta \overset{!}{=} \mathbb{P}_{\vartheta_1}\left(\sum_{j=1}^{n} X_j < c_n\right)$ für ein $\vartheta_1 < \vartheta_0$ führt wegen

$$\Phi^{-1}(\alpha) \;=\; -\Phi^{-1}(1-\alpha), \qquad \Phi^{-1}(\beta) \;=\; -\Phi^{-1}(1-\beta)$$

auf die gleiche Näherungsformel für den Mindeststichprobenumfang wie oben.

23.17 Beispiel (Zweiseitiger Binomialtest)

Seien $X_1, \ldots, X_n$ unabhängig und je Bin$(1, \vartheta)$-verteilt, $\vartheta \in \Theta = (0, 1)$.

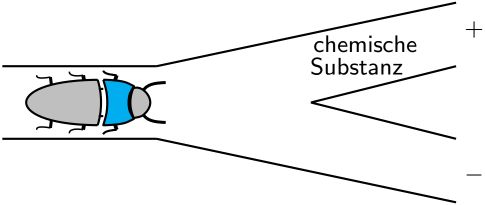
Sei $\vartheta_0 \in \Theta$ gegeben, $\Theta_0 := \{\vartheta_0\}$, $\Theta_1 := \Theta \setminus \{\vartheta_0\}$.

Hypothese $H_0 : \vartheta = \vartheta_0$, Alternative $H_1 : \vartheta \neq \vartheta_0$.

(sogenannter zweiseitiger Binomialtest)

(weil Alternative *zweiseitig*; ϑ kann größer oder kleiner als ϑ_0 sein)

Wichtigster Spezialfall: $\vartheta_0 = 1/2$ (im Folgenden behandelt).



Zweifach-Wahlapparat: Hat chemische Substanz anziehende oder abstoßende Wirkung?

$H_0 : \vartheta = 1/2$, $H_1 : \vartheta \neq 1/2$ (Sind zwei Fälle gleich wahrscheinlich?)

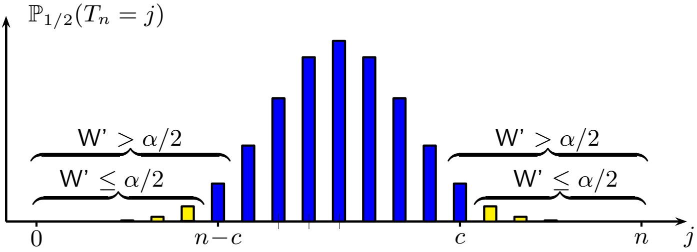
Prüfgröße (Anzahl der Treffer):

$$T_n(x_1, \ldots, x_n) := \sum_{j=1}^{n} x_j, \quad (x_1, \ldots, x_n) \in \mathcal{X} = \{0,1\}^n, \quad T_n := \sum_{j=1}^{n} X_j$$

Sei $\alpha \in (0,1)$ die zugelassene W' für einen Fehler 1. Art.

H_0 ablehnen, falls $T_n > c$ oder $T_n < n - c$; eventuell noch randomisieren.

$$c \ := \ \min \left\{ k \in \{0,1,\ldots,n\} : \mathbb{P}_{1/2}(T_n > k) \leq \frac{\alpha}{2} \right\}$$



Damit: $\mathbb{P}_{1/2}(T_n > c) + \mathbb{P}_{1/2}(T_n < n - c) \leq \alpha$

Eventuell steht hier „$<$“.

Dann in den Fällen $T_n = c$ oder $T_n = n - c$ die Hypothese H_0 jeweils mit Randomisierungswahrscheinlichkeit γ ablehnen.

Wegen $\mathbb{P}_{1/2}(T_n = k) = \mathbb{P}_{1/2}(T_n = n - k)$ Bestimmung von γ aus

$$2\,\mathbb{P}_{1/2}(T_n > c) + \gamma\, 2\,\mathbb{P}_{1/2}(T_n = c) \overset{!}{=} \alpha \implies \gamma = \frac{\frac{\alpha}{2} - \mathbb{P}_{1/2}(T_n > c)}{\mathbb{P}_{1/2}(T_n = c)}.$$

Damit ergibt sich der randomisierte Test ψ_n mit

$$\psi_n(x_1, \ldots, x_n) = \begin{cases} 1, & \text{falls} \quad T_n(x) > c \\ \gamma, & \text{falls} \quad T_n(x) = c \\ 0, & \text{falls} \quad n - c < T_n(x) < c \\ \gamma, & \text{falls} \quad T_n(x) = n - c \\ 1, & \text{falls} \quad T_n(x) < n - c. \end{cases}$$

Für diesen gilt $\mathbb{E}_{1/2}\, \psi_n = 2\mathbb{P}_{1/2}(T_n > c) + 2\mathbb{P}_{1/2}(T_n = c) = \alpha$.

Zahlenbeispiel: $n = 20$, $\alpha = 0.1$.

$$\left(\frac{1}{2}\right)^{20} \sum_{j=15}^{20} \binom{20}{j} = 0.0207, \quad \left(\frac{1}{2}\right)^{20} \sum_{j=14}^{20} \binom{20}{j} = 0.0577 \implies c = 14.$$

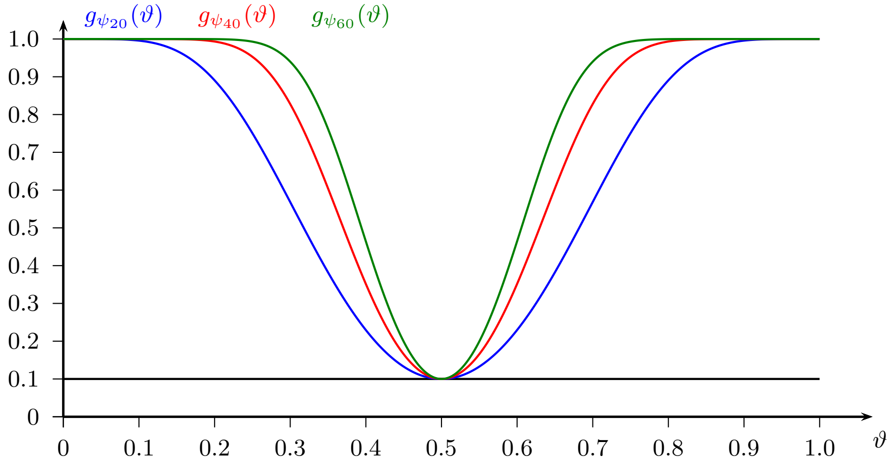
Weiter ist $\gamma = \dfrac{0.05 - 0.0207}{0.0577 - 0.0207} = 0.7919.$

Damit ergibt sich der Test zu

$$\psi_{20}(x_1, \ldots, x_{20}) = \begin{cases} 1, & \text{falls} \quad T_{20} > 14, \\ 0.7919, & \text{falls} \quad T_{20} = 14, \\ 0, & \text{falls} \quad 7 \le T_{20} \le 13, \\ 0.7919, & \text{falls} \quad T_{20} = 6, \\ 1, & \text{falls} \quad T_{20} < 6. \end{cases}$$

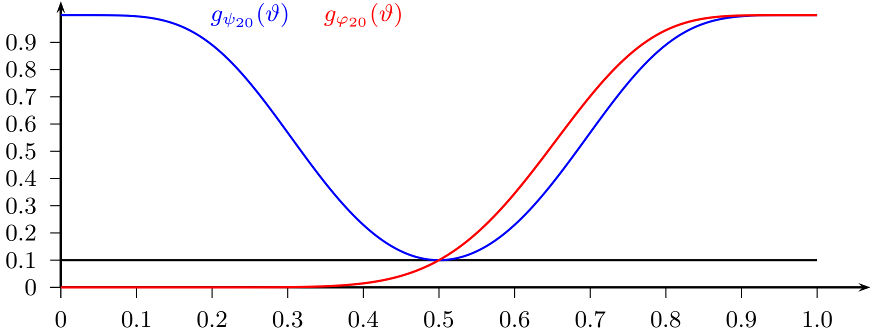
Gütefunktion:

$$\begin{aligned} g_{\psi_{20}}(\vartheta) &= \sum_{j=0}^{5} \binom{20}{j} \vartheta^j (1-\vartheta)^{20-j} + \sum_{j=15}^{20} \binom{20}{j} \vartheta^j (1-\vartheta)^{20-j} \\ &\quad + 0.7919 \cdot \binom{20}{6} \left(\vartheta^6 (1-\vartheta)^{14} + \vartheta^{14} (1-\vartheta)^6\right). \end{aligned}$$



Gütefunktionen zweiseitiger Binomialtests ($H_0 : \vartheta = 1/2$, $\alpha = 0.1$) für die Stichprobenumfänge $n = 20$, $n = 40$ und $n = 60$

Vergleich der Gütefunktionen von ein- und zweiseitigem Binomialtest:



Beachte:

- Der einseitige Test erkennt Alternativen ϑ mit $\vartheta > 1/2$ mit größerer W' als der zweiseitige Test.
- Die Wahl der Alternative (ein- oder zweiseitig) muss vor Datenerhebung festgelegt werden, sonst Erschleichung von Signifikanz!

23.18 Der p-Wert

Folgende Vorgehensweise ist gängige Praxis:

Ist $x \in \mathcal{X}$ beobachtet, so liefern Statistik-Programme den p-Wert $p^*(x)$ zur Beobachtung x.

Hierbei sind das Testproblem $(H_0,\ H_1)$ und die Prüfgröße sowie die prinzipielle Gestalt des kritischen Bereiches gegeben, z.B.

$$\mathcal{K} = \{x \in \mathcal{X} : T(x) \geq c\}.$$

$p^*(x)$ ist die kleinste Zahl α, für die die Wahl von α als Testniveau („gerade noch“) zur Ablehnung von H_0 führt.

Im Fall des einseitigen Binomialtests ($H_0 : \vartheta \leq 1/2$ gegen $H_1 : \vartheta > 1/2$) ist der p-Wert von $x = (x_1, \ldots, x_n)$ mit $t := x_1 + \ldots + x_n$ gleich

$$p^*(x) \;=\; \mathbb{P}_{1/2}\left(\sum_{j=1}^{n} X_j \geq t\right).$$

Gilt $p^*(x) \leq \alpha$, so lehne H_0 zum Niveau α ab.

Andernfalls erhebe keinen Einwand gegen H_0.

23.19 Zusammenhang zwischen Konfidenzbereichen und Tests

Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell.

Sei $\mathcal{C} : \mathcal{X} \to \mathcal{P}(\Theta)$ ein Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1 - \alpha$, also:

$$\mathbb{P}_\vartheta(\mathcal{C}(X) \ni \vartheta) \;\geq\; 1 - \alpha \quad \forall \vartheta \in \Theta.$$

Sei $\Theta = \Theta_0 + \Theta_1$.

Das folgende Verfahren ist ein Test für $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ zum Niveau α:

Kritischer Bereich:

$$\mathcal{K} := \{x \in \mathcal{X} : \mathcal{C}(x) \cap \Theta_0 = \emptyset\}.$$

Es folgt für jedes $\vartheta \in \Theta_0$:

$$\begin{aligned} \mathbb{P}_\vartheta(X \in \mathcal{K}) \quad &= \quad \mathbb{P}_\vartheta(\mathcal{C}(X) \cap \Theta_0 = \emptyset) \\ &\leq \quad \mathbb{P}_\vartheta(\mathcal{C}(X) \not\ni \vartheta) \;\leq\; \alpha \end{aligned}$$

Somit ist $\mathbf{1}_\mathcal{K}$ in der Tat ein Test zum Niveau α.

23.20 Der Chi–Quadrat–Anpassungstest

- eines der ältesten Testverfahren (1900)
- Zweck: Prüfung der Verträglichkeit von relativen Häufigkeiten mit hypothetischen W'keiten im multinomialen Versuchsschema

Wiederholung: Multinomiales Versuchsschema (vgl. Kap. 13).

n unabhängige gleichartige Versuche mit jeweils s möglichen Ausgängen $1, 2, \ldots, s$.

Ausgang j heißt Treffer j-ter Art, $j = 1, \ldots, s$.

Sei p_j die W' für einen Treffer j-ter Art (in jedem Versuch gleich).

Sei X_j die Anzahl der Treffer j-ter Art, $j = 1, \ldots, s$.

Es gilt $X := (X_1, \ldots, X_s) \sim \text{Mult}(n; p_1, \ldots, p_s)$.

$$\mathbb{P}(X = \boldsymbol{k}) \;=\; \frac{n!}{k_1! k_2! \cdot \ldots \cdot k_s!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_s^{k_s}$$

Dabei: $\boldsymbol{k} = (k_1, \ldots, k_s) \in \mathcal{X}$,

$$\mathcal{X} \;:=\; \{\boldsymbol{k} = (k_1, \ldots, k_s) : k_1, \ldots, k_s \in \mathbb{N}_0 \text{ mit } k_1 + \ldots + k_s = n\}.$$

Jetzt: $\vartheta := (p_1, \ldots, p_s)$ unbekannt,

$$\Theta \ := \ \left\{ (p_1, \ldots, p_s) \colon p_1, \ldots, p_s \geq 0, \ \sum_{j=1}^{s} p_j = 1 \right\}.$$

Sei $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_s) \in \Theta$ vorgegeben.

Zu testen ist die Hypothese $H_0 : p_j \ = \ \pi_j$ für jedes $j = 1, \ldots, s$
gegen die Alternative $H_1 : p_j \ \neq \ \pi_j$ für mindestens ein $j \in \{1, \ldots, s\}$.

Beispiel: $s = 6$, $\pi_1 = \pi_2 = \ldots = \pi_6 = 1/6$ (ist der Würfel echt?)

Die H_0-Wahrscheinlichkeiten $\mathbb{P}_{\boldsymbol{\pi}}(X = \boldsymbol{k})$ sind

$$m_n(\boldsymbol{k}) \ := \ \frac{n!}{k_1! \cdot \ldots \cdot k_s!} \cdot \prod_{j=1}^{s} \pi_j^{k_j} \ , \qquad \boldsymbol{k} \in \mathcal{X}.$$

Idee: Bilde einen kritischen Bereich $\mathcal{K} \subseteq \mathcal{X}$ aus den unter H_0 unwahrscheinlichsten $\boldsymbol{k}$ (bis vorgegebenes Testniveau α erreicht).

Zahlenbeispiel: $n = 4$, $s = 3$ und $\pi_1 = \pi_2 = 1/4$, $\pi_3 = 1/2$.

Hier besteht $\mathcal{X}$ aus 15 Tripeln.

(k_1, k_2, k_3)	$\frac{4!}{k_1!k_2!k_3!}$	$\prod_{j=1}^{3} \pi_j^{k_j}$	$m_4(\boldsymbol{k})$	$\chi_4^2(\boldsymbol{k})$
$(4,0,0)$	1	$1/256$	$1/256$	12
$(0,4,0)$	1	$1/256$	$1/256$	12
$(3,1,0)$	4	$1/256$	$4/256$	6
$(1,3,0)$	4	$1/256$	$4/256$	6
$(2,2,0)$	6	$1/256$	$6/256$	4
$(3,0,1)$	4	$1/128$	$8/256$	5.5
$(0,3,1)$	4	$1/128$	$8/256$	5.5
$(0,0,4)$	1	$1/16$	$16/256$	4
$(2,1,1)$	12	$1/128$	$24/256$	1.5
$(1,2,1)$	12	$1/128$	$24/256$	1.5
$(2,0,2)$	6	$1/64$	$24/256$	2
$(0,2,2)$	6	$1/64$	$24/256$	2
$(0,1,3)$	4	$1/32$	$32/256$	1.5
$(1,0,3)$	4	$1/32$	$32/256$	1.5
$(1,1,2)$	12	$1/64$	$48/256$	0

Mit $\mathcal{K} \ := \ \{(k_1, k_2, k_3) \in \mathcal{X} : k_3 = 0\}$ gilt

$$\mathbb{P}_\pi(X \in \mathcal{K}) \ = \ \frac{1+1+4+4+6}{256} \ = \ 0.0625.$$

Folglich besitzt dieser Test die W' 0.0625 für einen Fehler erster Art.

Prinzipiell kann man so auch für größere Werte von n und s vorgehen.

Aber: Rechenaufwand steigt mit wachsendem n und s rapide an.

Praktikablere Möglichkeit?

Karl Pearson: Es gilt (ohne Beweis)

$$\lim_{n\to\infty} \frac{m_n(\boldsymbol{k})}{f_n(\boldsymbol{k})} \ = \ 1,$$

wobei

$$f_n(\boldsymbol{k}) \ := \ \left[(2\pi)^{s-1} n^{s-1} \prod_{j=1}^{s} \pi_j\right]^{-1/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{s} \frac{(k_j - n\pi_j)^2}{n\pi_j}\right).$$

Somit entsprechen bei großem n kleinen Werten von $m_n(\boldsymbol{k})$ große Werte von

$$\chi_n^2(k_1,\ldots,k_s) := \sum_{j=1}^{s} \frac{(k_j - n\pi_j)^2}{n\pi_j}.$$

(k_1,k_2,k_3)	$\frac{4!}{k_1!k_2!k_3!}$	$\prod_{j=1}^{3} \pi_j^{k_j}$	$m_4(\boldsymbol{k})$	$\chi_4^2(\boldsymbol{k})$
$(4,0,0)$	1	$1/256$	$1/256$	12
$(0,4,0)$	1	$1/256$	$1/256$	12
$(3,1,0)$	4	$1/256$	$4/256$	6
$(1,3,0)$	4	$1/256$	$4/256$	6
$(2,2,0)$	6	$1/256$	$6/256$	4
$(3,0,1)$	4	$1/128$	$8/256$	5.5
$(0,3,1)$	4	$1/128$	$8/256$	5.5
$(0,0,4)$	1	$1/16$	$16/256$	4
$(2,1,1)$	12	$1/128$	$24/256$	1.5
$(1,2,1)$	12	$1/128$	$24/256$	1.5
$(2,0,2)$	6	$1/64$	$24/256$	2
$(0,2,2)$	6	$1/64$	$24/256$	2
$(0,1,3)$	4	$1/32$	$32/256$	1.5
$(1,0,3)$	4	$1/32$	$32/256$	1.5
$(1,1,2)$	12	$1/64$	$48/256$	0

Memo: $\chi_n^2(k_1, \ldots, k_s) := \sum_{j=1}^{s} \frac{(k_j - n\pi_j)^2}{n\pi_j}$

$\chi_n^2 : \mathcal{X} \to \mathbb{R}$ heißt χ^2-Testgröße (sprich: *Chi–Quadrat*).

Folgerung: Sinnvoller kritischer Bereich ist von der Gestalt

$$\mathcal{K} := \left\{ \boldsymbol{k} \in \mathcal{X} : \chi_n^2(k_1, \ldots, k_s) \geq c \right\}.$$

c wird aus der vorgegebenen W' α für einen Fehler 1. Art bestimmt.

Hierzu: Welche Verteilung besitzt

$$T_n \ := \ \sum_{j=1}^{s} \frac{(X_j - n\pi_j)^2}{n\pi_j}$$

unter der Hypothese H_0?

Wegen $X_j \sim \mathrm{Bin}(n, \pi_j)$ gilt $\mathbb{E}_\pi (X_j - n\pi_j)^2 = n\pi_j(1 - \pi_j)$ und somit

$$\mathbb{E}_\pi(T_n) \ = \ \sum_{j=1}^{s} (1 - \pi_j) \ = \ s - 1$$ (weder von n noch von π abhängig!)

Karl Pearsons entscheidende Entdeckung: Es gilt

$$\lim_{n\to\infty} \mathbb{P}_\pi(T_n \geq c) \;=\; \int_c^\infty f_{s-1}(t)\,\mathrm{d}t. \tag{23.11}$$

(unabhängig von π!!) Dabei ist allgemein für $r \in \mathbb{N}$:

$$f_r(t) \;:=\; \frac{1}{2^{r/2}\,\Gamma(r/2)}\,\exp\left(-\frac{t}{2}\right)\,t^{r/2-1}, \quad t > 0,$$

und $f_r(t) := 0$, sonst.

Die durch

$$\Gamma(x) \;:=\; \int_0^\infty e^{-u}u^{x-1}\,\mathrm{d}u$$

definierte Funktion $\Gamma : (0,\infty) \to \mathbb{R}$ heißt Gamma-Funktion. Es gelten:

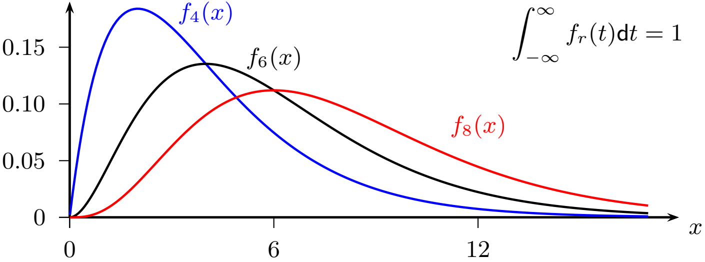
$$\begin{aligned} \Gamma(x+1) &= x\,\Gamma(x), && \text{(partielle Integration)} \\ \Gamma\left(\frac{1}{2}\right) &= \sqrt{\pi} && \text{(äquiv. zu } \int_{-\infty}^\infty \exp(-t^2/2)\mathrm{d}t = \sqrt{2\pi}\text{)} \\ \Gamma(n) &= (n-1)!, \quad n \in \mathbb{N} && \text{(mit } \Gamma(1) = 1\text{)} \end{aligned}$$

23.21 Definition (Dichte der χ^2_r-Verteilung)

Die durch

$$f_r(t) \;:=\; \frac{1}{2^{r/2}\cdot\Gamma(r/2)}\cdot\exp\left(-\frac{t}{2}\right)\cdot t^{r/2-1}, \quad t>0,$$

und $f_r(t) := 0$, sonst, definierte Funktion $f_r : \mathbb{R}\to\mathbb{R}$ heißt Dichte der Chi-Quadrat-Verteilung mit r Freiheitsgraden. (kurz: χ^2_r-Verteilung).



Dichten von Chi-Quadrat-Verteilungen

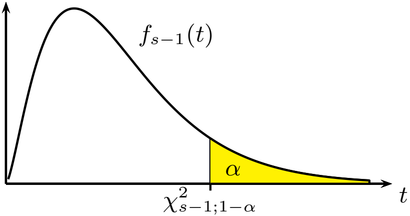
Memo: $\lim_{n\to\infty} \mathbb{P}_\pi(T_n \geq c) \; = \; \int_c^\infty f_{s-1}(t)\; \mathrm{d}t$

Memo: $T_n = \sum_{j=1}^{s} \frac{(X_j - n\pi_j)^2}{n\pi_j}$; $\quad H_0$ ablehnen, falls $T_n \geq c$

Wähle c so, dass $\int_c^\infty f_{s-1}(t)\,\mathrm{d}t = \alpha$ ($\Longrightarrow$ Test hat asymptotisches Niveau α).

Der Wert c mit dieser Eigenschaft heißt $(1-\alpha)$-Quantil der χ^2_{s-1}-Verteilung.

Er wird mit $\chi^2_{s-1;1-\alpha}$ bezeichnet.



	α					
r	0.1	0.05	0.025	0.01	0.005	0.001
1	2.71	3.84	5.02	6.63	7.88	10.83
2	4.61	5.99	7.38	9.21	10.60	13.82
3	6.25	7.81	9.35	11.34	12.84	16.27
4	7.78	9.49	11.14	13.28	14.86	18.47
5	9.24	11.07	12.83	15.09	16.75	20.51
6	10.64	12.59	14.45	16.81	18.55	22.46
7	12.02	14.07	16.01	18.48	20.28	24.32
8	13.36	15.51	17.53	20.09	21.95	26.12

$(1-\alpha)$-Quantile $\chi^2_{r;1-\alpha}$ der χ^2_r-Verteilung

Durchführung des χ^2-Tests zum asymptotischen Niveau α:

- Bestimme Wert $c = \chi^2_{s-1;1-\alpha}$ aus Tabelle.
- Berechne zu Trefferanzahlen $k_1, \ldots, k_s$ den Wert der χ^2-Testgröße

$$\chi^2_n(k_1, \ldots, k_s) \;=\; \sum_{j=1}^{s} \frac{(k_j - n\pi_j)^2}{n\pi_j}.$$

- Lehne $H_0 : p_j = \pi_j \;\forall j = 1, \ldots, s$ ab, falls $\chi^2_n(k_1, \ldots, k_s) \geq c$.
- Im Fall $\chi^2_n(k_1, \ldots, k_s) < c$ erhebe keinen Einwand gegen H_0.

23.22 Beispiel (Vererbung, Mendelsche Gesetze)

Gregor Mendel beobachtete 1865 simultan die Form (rund, kantig) und Farbe (gelb, grün) von Erbsen.

Theorie: W'en für die Merkmalausprägungen (r, ge), (r, gr), (k, ge) und (k, gr) verhalten sich wie 9:3:3:1. Mendel zählte unter $n = 556$ Erbsen

$$315 \text{ mal (r, ge)}, 108 \text{ mal (r, gr)}, 101 \text{ mal (k, ge)}, 32 \text{ mal (k, gr)}.$$

Sind diese Ergebnisse mit der Theorie verträglich?

$$\text{Hier: } s = 4,\ \pi_1 = \frac{9}{16},\ \pi_2 = \frac{3}{16},\ \pi_3 = \frac{3}{16},\ \pi_4 = \frac{1}{16}.$$

$n = 556,\ k_1 = 315,\ k_2 = 108,\ k_3 = 101,\ k_4 = 32.$

$$\chi_n^2 = \sum_{j=1}^{4} \frac{(k_j - n\pi_j)^2}{n\pi_j} \ = \ \cdots \ = \ 0.470.$$

Wegen $\chi^2_{3;0.95} = 7.81$ (Tabelle) wird H_0 nicht verworfen.

Der Wert 0.470 ist verdächtig klein. Hat Mendel seine Daten frisiert?

Für Anwendung des χ^2-Tests wie oben sollte $n\,\min(\pi_1,\ldots,\pi_s) \geq 5$ gelten.
Sonst: Mit Hilfe von Pseudozufallszahlen Monte-Carlo-Test wie folgt:

- Wähle große Zahl M, z.B. $M = 10\,000$, und setze $Z := 0$.
- Führe für $m = 1, 2, \ldots, M$ folgenden Algorithmus durch:
  1) Simuliere n mal ein Experiment, das mit W' π_j einen Treffer j-ter Art ergibt $(j = 1, \ldots, s)$. Die so simulierten Trefferanzahlen seien

$$k_{1,m},\ k_{2,m}, \ldots, k_{s,m}.$$

  2) Mit Hilfe von $k_{1,m},\ k_{2,m}, \ldots, k_{s,m}$ berechne man

$$\chi^2_{n,m} := \sum_{j=1}^{s} \frac{(k_{j,m} - n\pi_j)^2}{n\pi_j}.$$

  3) Gilt $\chi^2_{n,m} \geq \chi^2_n(\boldsymbol{k})$, so $Z := Z + 1$.

- $$\frac{Z}{M} \text{ ist Schätzwert für } \mathbb{P}_\pi(T_n \geq \chi^2_n(\boldsymbol{k})).$$

- Lehne H_0 zum Niveau α ab, falls $Z/M \leq \alpha$. Andernfalls erhebe keine Einwände gegen H_0.

23.23 Der exakte Test von Fisher

Ein neues Medikament wurde an einer kleinen Zahl von Patienten überprüft. Dazu wurden die an dem Test teilnehmenden Patienten per Losentscheid in eine Versuchs- und in eine Kontrollgruppe aufgeteilt. Die Patienten der Versuchsgruppe erhielten das neue Medikament, die in der Kontrollgruppe ein Placebo. Die Untersuchung wurde als Doppelblindstudie durchgeführt. Die Ergebnisse des Tests wurden in einer Vierfeldertafel festgehalten:

	Linderung	Keine Linderung	Summe
Versuchsgruppe	7	4	11
Kontrollgruppe	1	11	12
Summe	8	15	23

- Sprechen die Ergebnisse für die Wirkung des neuen Medikaments?
- Wie wahrscheinlich ist dieses oder ein noch extremeres Ergebnis, wenn das Medikament keine Wirkung hat?

(aus: Stochastik: Arbeitsbuch mit CD-ROM, Braunschweig; Schroedel 2012)

Allgemeine Situation:

- Es liegt Grundgesamtheit (GG) von N Individuen vor.
- GG wird anhand zweier Merkmale A, B mit Ausprägungen 1,0 zufällig unterteilt.
- Im obigen Beispiel:

$$\begin{array}{lcl} A_1 & \simeq & \text{Versuchsgruppe,} \\ A_0 & \simeq & \text{Kontrollgruppe,} \\ B_1 & \simeq & \text{Linderung (Effekt),} \\ B_0 & \simeq & \text{keine Linderung (kein Effekt).} \end{array}$$

- Man beobachtet folgende Häufigkeiten:

	B_1	B_0	Summe
A_1	N_{11}	N_{10}	n
A_0	N_{01}	N_{00}	$N-n$
Summe	r	s	N

- Bei gegebenen Randsummen ist nur N_{11} zufällig.

	B_1	B_0	Summe
A_1	N_{11}	N_{10}	n
A_0	N_{01}	N_{00}	$N-n$
Summe	r	s	N

Sei $\mathbb{P}$ die Gleichverteilung auf der Grundgesamtheit.

Wir interpretieren A_1, A_0, B_1, B_0 als Teilmengen der GG.

$$\begin{aligned} & \text{Medikament hat keine Wirkung (Behandlung hat keinen Effekt)} \\ \Leftrightarrow \quad & \mathbb{P}(B_1|A_1) = \mathbb{P}(B_1|A_0) \\ \Leftrightarrow \quad & A_1 \text{ und } B_1 \text{ stochastisch unabhängig} \quad (\text{Hypothese } H_0) \end{aligned}$$

- Interpretiere die r „Erfolgsfälle“ (B_1, Linderung, Effekt) als rote Kugeln.
- Interpretiere die anderen Fälle (B_0) als schwarze Kugeln.
- Unter H_0 besitzt N_{11} bei gegebenem r, s, n die Verteilung Hyp(n, r, s).
- Der exakte Test von Fisher verwendet als Testgröße N_{11}.
- Der Test bewertet die Signifikanz eines erhaltenen Wertes k für N_{11} mit Hilfe obiger hypergeometrischer Verteilung.

	B_1	B_0	Summe
A_1	N_{11}	N_{10}	n
A_0	N_{01}	N_{00}	$N-n$
Summe	r	s	N

Der p-Wert $p^*(k)$ der Realisierung k von N_{11} ist je nachdem, was im Vergleich zu k als „noch extremeres Ereignis“ zu betrachten ist, entweder

$$p^*(k) = \sum_{j=k}^{r} \frac{\binom{r}{j}\binom{s}{n-j}}{\binom{r+s}{n}}$$

oder

$$p^*(k) = \sum_{j=0}^{k} \frac{\binom{r}{j}\binom{s}{n-j}}{\binom{r+s}{n}}.$$

	Linderung	Keine Linderung	Summe
Versuchsgruppe	7	4	11
Kontrollgruppe	1	11	12
Summe	8	15	23

- Sprechen die Ergebnisse für die Wirkung des neuen Medikaments?
- Wie wahrscheinlich ist dieses oder ein noch extremeres Ergebnis, wenn das Medikament keine Wirkung hat?

Der p-Wert ist für den Fall obiger Daten gleich

$$p^*(7) = \frac{\binom{8}{7}\binom{15}{4}}{\binom{23}{11}} + \frac{\binom{8}{8}\binom{15}{3}}{\binom{23}{11}} \approx 0.0084$$

Die Nullhypothese der Wirkungslosigkeit des Medikaments würde sogar auf dem 1%-Niveau abgelehnt.

23.24 Trugschlüsse beim Umgang mit Tests

- Der Leiter der Abteilung für Materialbeschaffung hat eine Sendung von Schaltern stichprobenartig auf deren Funktionsfähigkeit hin überprüft.
- Er stellt fest, dass bei dieser Stichprobe der Anteil defekter Schalter signifikant über dem vom Hersteller behaupteten Ausschussanteil liegt.
- Dabei überprüft er die vom Hersteller aufgestellte Behauptung mit einem statistischen Test zum Niveau 0.05.
- Er sagt: „Die Chance, dass eine genaue Überprüfung zeigt, dass die Sendung den Herstellerangaben entspricht, ist höchstens 5%“.
- Er empfiehlt, die Lieferung zu reklamieren und zurückgehen zu lassen.
- Ist seine Aussage richtig?

Nein! Es handelt sich um den häufig anzutreffenden Trugschluss, es existiere eine „bedingte Wahrscheinlichkeit

$$\mathbb{P}(H_0 \text{ gilt} | \text{Test führt zur Ablehnung von } H_0)\text{“},$$

und diese „Wahrscheinlichkeit“ sei höchstens α $(= 0.05)$.

23.25 Wie erschleicht man sich Signifikanz?

Antwort: Durch „Herauspicken“ und Verschweigen!

Unter der Überschrift *Kernige Statistik* berichtete DIE ZEIT am 19.12.1997, dass das Ergebnis eines vom Bundesumweltministerium in Auftrag gegebenen Gutachtens über Leukämie bei Kindern, die in der Nähe von Kernkraftwerken leben, von einer Expertenkommission in Zweifel gezogen wurde.

Gutachter konnten kein erhöhtes Krankheitsrisiko feststellen. Amtliche Kommission kam zu dem Schluss, dass die Häufung von Leukämiefällen bei Kindern rund um das AKW Krümmel „mit großer Wahrscheinlichkeit auf dieses AKW zurückzuführen sei“; ein „nennenswerter Beitrag durch andere denkbare Verursacher sei unwahrscheinlich“.

Ein von der Kommission festgestelltes fünffach höheres Krankheitsrisiko für Leukämie bei Kindern bei allen sechs seit 1980 ans Netz gegangenen AKWs ergab sich nur dann, wenn bestimmte ausgewählte Vergleichsregionen herangezogen werden; bis auf das AKW Krümmel lagen die Erkrankungsraten bei den übrigen AKWs sogar *unter* dem Bundesdurchschnitt.

- Statistiken können (etwa durch Weglassen unliebsamer Daten) „frisiert werden“
- Die Formulierung, die Häufung der Leukämiefälle rund um das AKW Krümmel sei *mit großer Wahrscheinlichkeit* auf das AKW zurückzuführen, bringt den Wahrscheinlichkeitsbegriff ins Spiel
- Frage: Ist die beobachtete Leukämierate unter einem stochastischen Modell für die *normalerweise beobachteten zufälligen Schwankungen* der Anzahl der Leukämieerkrankungen, bezogen auf alle Mitglieder einer bestimmten Gruppe innerhalb der Gesamtbevölkerung, signifikant erhöht?
- Modell: Anzahl der Leukämieerkrankungen, bezogen auf 10000 Kinder in der Gesamtbevölkerung (und festen Zeitraum), ist Po(4)-verteilt.

Ort Nr.	1	2	3	4	5	6	7	8	9	10
Leukämiefälle	5	3	1	6	4	9	4	2	3	2

Fiktive Leukämiefälle an 10 Orten (mit Pseudozufallszahlen)

Zusammen sind 39 Fälle aufgetreten, im Durchschnitt $3.9 = 39/10$.

Erschleichung von Signifikanz:

Alle Orte bis auf denjenigen mit der *höchsten Leukämierate*, also Ort Nr. 6, weglassen und die dort beobachteten neun Leukämiefälle als Realisierung *nur einer* Po(4)-verteilten Zufallsvariablen ansehen.

Beachte: $\mathbb{P}(X_6 \geq 9) \;=\; \sum_{j=9}^{\infty} \mathrm{e}^{-4} \cdot \frac{4^j}{j!} \;\approx\; 0.021 \quad (\ll 0.05!!)$

Vermeintliche statistische Signifikanz!

Herausgesucht hat man aber den größten Wert

$$M_{10} = \max(X_1, X_2, \ldots, X_{10})$$

$$\begin{aligned}\mathbb{P}(M_{10} \geq 9) \;&=\; \mathbb{P}\left(\bigcup_{j=1}^{10}\{X_j \geq 9\}\right) \;=\; 1 - \mathbb{P}\left(\bigcap_{j=1}^{10}\{X_j \leq 8\}\right)\\ &=\; 1 - \mathbb{P}(X_1 \leq 8)^{10} \;\approx\; 0.194\end{aligned}$$

Der Wert 0.194 ist leider unspektakulär!

24 Allgemeine Modelle

- Bislang: $(\Omega, \mathbb{P})$ diskreter W-Raum, d.h.:
- $\Omega \neq \emptyset$ beliebig,
- $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ mit $\mathbb{P}(\Omega) = 1$ und

$$\mathbb{P}\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j) \quad \text{für paarweise disjunkte } A_1, A_2, \ldots$$

- es ex. abzählbare Teilmenge $\Omega_0 =: \{\omega_1, \omega_2, \ldots\}$ von Ω mit $\mathbb{P}(\Omega_0) = 1$,
- $\mathbb{P}$ ist durch Punktmassen $\mathbb{P}(\{\omega_j\})$, $j \geq 1$, festgelegt.

Jetzt: Gibt es ein W-Maß $\mathbb{P} : \mathcal{P}(\mathbb{R}) \to [0,1]$ mit

- $\mathbb{P}([0,1]) = 1$,
- $\mathbb{P}([a,b]) = b - a$ für alle $a, b \in [0,1]$ mit $a \leq b$?

  $(\Longrightarrow \mathbb{P}(\{x\}) = 0 \ \forall x \in \mathbb{R})$
- Antwort: Nein!

Problem berührt folgende Grundfrage der Maßtheorie:

Gibt es eine „Längen-Funktion“ $L : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ mit:

a) $L(\emptyset) \ = \ 0$,

b) $L([a,b]) \ = \ b - a$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$,

c) $L(A + x) \ = \ L(A)$ für jedes $A \subseteq \mathbb{R}$ und jedes $x \in \mathbb{R}$,

d) $L\left(\sum_{j=1}^{\infty} A_j\right) \ = \ \sum_{j=1}^{\infty} L(A_j)$, falls $A_1, A_2, \ldots$ paarweise disjunkt?

Antwort: Nein! Die Potenzmenge als Definitionsbereich ist zu groß!

Beachte:

- c) ist die Translationsinvarianz von L.
- Aus der σ-Additivität in d) und a) folgt $L(A + B) \ = \ L(A) + L(B)$.
- Aus $A \subseteq B$ folgt $L(A) \ \leq \ L(B)$.

24.1 Satz Es gibt keine Funktion $L : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ mit a) - d).

BEWEIS: $x \sim y :\Longleftrightarrow x - y \in \mathbb{Q}$ definiert Äquivalenzrelation auf $\mathbb{R}$

$\Longrightarrow \ \exists I \neq \emptyset, \quad \exists K_i \subseteq \mathbb{R}$ für $i \in I$ mit $\mathbb{R} = \sum_{i \in I} K_i$

und $\forall\, x, y \in \mathbb{R} : \ x \sim y \iff \exists\, i \in I$ mit $x, y \in K_i$.

Auswahlaxiom $\Longrightarrow \exists\, A \subseteq \mathbb{R}$ mit $|A \cap K_i| = 1 \ \forall\, i \in I$

O.B.d.A. $A \subseteq [0, 1)$, da $x \sim x - \lfloor x \rfloor \in [0, 1)$.

Es gilt $\mathbb{R} = \displaystyle\sum_{r \in \mathbb{Q}} (r + A)$

Beweis: (i) Es gilt $\subseteq$, weil:

$x \in \mathbb{R} \Longrightarrow \exists a \in A$ mit $x \sim a \Longrightarrow x = r + a$ mit $r \in \mathbb{Q} \ \ \surd$

(ii) Sei $r_1 \neq r_2$. Beh.: Es gilt $(r_1 + A) \cap (r_2 + A) = \emptyset$. Beweis:

Ann.: $x \in (r_1 + A) \cap (r_2 + A) \Longrightarrow \exists a_1, a_2 \in A$ mit $x = r_1 + a_1 = r_2 + a_2$

$\Longrightarrow a_1 - a_2 = r_2 - r_1 \ \Longrightarrow a_1 \sim a_2 \Longrightarrow a_1 = a_2 \Longrightarrow r_1 = r_2$ Widerspruch!

Memo: $\mathbb{R} = \sum_{r\in\mathbb{Q}} (r+A), \quad A \subseteq [0,1)$

Mit der σ-Additivität und Translationsinvarianz von L folgt

$$\begin{aligned} \infty \; &= \; L(\mathbb{R}) \; = \; L\left(\sum_{r\in\mathbb{Q}} (r+A)\right) \; = \; \sum_{r\in\mathbb{Q}} L(r+A) \\ &= \; \sum_{r\in\mathbb{Q}} L(A) \qquad \Longrightarrow \; L(A) > 0. \end{aligned}$$

Andererseits: $\quad A \subseteq [0,1) \Longrightarrow \sum_{r\in\mathbb{Q}\cap[0,1)} (r+A) \subseteq [0,2) \Longrightarrow$

$$\begin{aligned} 2 \; &= \; L([0,2)) \; \geq \; L\left(\sum_{r\in\mathbb{Q}\cap[0,1)} (r+A)\right) \; = \; \sum_{r\in\mathbb{Q}\cap[0,1)} L(r+A) \\ &= \; \sum_{r\in\mathbb{Q}\cap[0,1)} L(A) \qquad \Longrightarrow L(A) = 0. \quad \text{Widerspruch!} \end{aligned}$$

- In gleicher Weise lässt sich zeigen:

  (mit $x \sim y :\Longleftrightarrow x - y \in \mathbb{Q}^k,$ wobei $x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k)$)

  Flächenmessung $(k = 2)$ und Volumenmessung $(k = 3)$ sowie die Definition eines höherdimensionalen geometrischen Elementarvolumens $(k \geq 4)$ ist nicht auf der vollen Potenzmenge des $\mathbb{R}^k$ möglich.

- Konsequenz: Definitionsbereich von L (sowie von Flächen- und (höherdimensionalen) Volumen-Maßen) einschränken,
  und zwar auf ein geeignetes System von Teilmengen des $\mathbb{R}^k$.

- Diese Notwendigkeit ergibt sich auch für viele W-Maße, die nur noch auf einer geeigneten Teilmenge von $\mathcal{P}(\Omega)$ definiert werden können.

- Ein solches Mengensystem (System von Teilmengen von Ω) sollte abgeschlossen sein gegenüber der Bildung wichtiger mengentheoretischer Operationen.

24.2 Definition (σ-Algebra)

Es sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra (über Ω), falls gilt:

a) $\emptyset \in \mathcal{A}$,

b) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$,

c) $A_n \in \mathcal{A}\ (n \in \mathbb{N}) \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

24.3 Folgerungen Falls $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ σ-Algebra, so gelten:

a) $\Omega \in \mathcal{A}$, $\surd$

b) $A_1, \ldots, A_n \in \mathcal{A} \Longrightarrow A_1 \cup \ldots \cup A_n \in \mathcal{A}$, $\surd$

c) $A_n \in \mathcal{A}\ (n \in \mathbb{N}) \Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ $\quad \left[\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c\right]$ $\surd$

d) $A_1, \ldots, A_n \in \mathcal{A} \Longrightarrow A_1 \cap \ldots \cap A_n \in \mathcal{A}$. $\surd$

24.4 Beispiele

a) $\{\Omega, \emptyset\}$ ist die kleinste (gröbste) σ-Algebra über Ω.

b) $\mathcal{P}(\Omega)$ ist die größte (feinste) σ-Algebra über Ω.

c) $\{A \subseteq \Omega : A$ abzählbar oder A^c abzählbar$\}$ ist eine σ-Algebra (Übung!)

d) $\{\emptyset, A, A^c, \Omega\}$ ist eine σ-Algebra.

Für die Konstruktion von σ-Algebren ist folgendes Resultat wichtig:

24.5 Satz (Schnitte von σ-Algebren sind σ-Algebren)

Es sei $I \neq \emptyset$. Sind $\mathcal{A}_i$, $i \in I$, σ-Algebren über Ω, so ist auch deren Durchschnitt

$$\bigcap_{i \in I} \mathcal{A}_i := \{A \subseteq \Omega : A \in \mathcal{A}_i \quad \forall\, i \in I\}$$

eine σ-Algebra über Ω.

BEWEIS: Nachweis der definierenden Eigenschaften einer σ-Algebra.

$$\emptyset \in \bigcap_{i \in I} \mathcal{A}_i \checkmark, \qquad A \in \bigcap_{i \in I} \mathcal{A}_i \Longrightarrow A^c \in \bigcap_{i \in I} \mathcal{A}_i \checkmark \quad \text{usw.}$$

24.6 Satz und Definition (Erzeugte σ-Algebra, Erzeugendensystem)

Zu beliebigem $\mathcal{M} \subseteq \mathcal{P}(\Omega)$ existiert genau eine σ-Algebra $\sigma(\mathcal{M})$ über Ω mit:

a) $\mathcal{M} \subseteq \sigma(\mathcal{M})$,

b) Ist $\mathcal{A}$ eine beliebige σ-Algebra über Ω mit $\mathcal{M} \subseteq \mathcal{A}$, so gilt $\sigma(\mathcal{M}) \subseteq \mathcal{A}$.

$\sigma(\mathcal{M})$ heißt die von $\mathcal{M}$ erzeugte σ-Algebra.

$\mathcal{M}$ heißt Erzeugendensystem (kurz: Erzeuger) von $\sigma(\mathcal{M})$.

Beweis: Sei $I := \{\mathcal{F} : \mathcal{F}\ \sigma\text{-Algebra über } \Omega \text{ mit } \mathcal{M} \subseteq \mathcal{F}\}$.

Es gilt $I \neq \emptyset$, da $\mathcal{P}(\Omega) \in I$. Satz 24.5 $\Longrightarrow$

$$\sigma(\mathcal{M}) := \bigcap_{\mathcal{F} \in I} \mathcal{F} \quad \text{ist } \sigma\text{-Algebra über } \Omega \text{ mit } \mathcal{M} \subseteq \sigma(\mathcal{M}) \text{ und b).}$$

Eindeutigkeit: Seien $\mathcal{A}_1, \mathcal{A}_2$ σ-Algebren über Ω mit a) und b) anstelle von $\sigma(\mathcal{M})$. Dann folgt $\mathcal{A}_1 \subseteq \mathcal{A}_2$ und $\mathcal{A}_2 \subseteq \mathcal{A}_1$, also $\mathcal{A}_1 = \mathcal{A}_2$.

$\sigma(\mathcal{M})$ ist die kleinste, $\mathcal{M}$ enthaltende σ-Algebra über Ω.

Memo: $\emptyset \in \mathcal{A}, \quad A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}, \quad A_n \in \mathcal{A}\ (n \in \mathbb{N}) \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$

24.7 Beispiel

Sei $\mathcal{M} := \{\{\omega\} : \omega \in \Omega\}$. $\sigma(\mathcal{M}) = ?$

Sei $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine beliebige σ-Algebra mit $\mathcal{M} \subseteq \mathcal{A}$.

Nach der dritten Eigenschaft einer σ-Algebra muss $\mathcal{A}$ jede abzählbare Teilmenge von Ω enthalten.

Nach der zweiten Eigenschaft einer σ-Algebra muss $\mathcal{A}$ die Komplemente der abzählbaren Teilmengen von Ω enthalten.

Folg.: $\mathcal{F} := \{A \subseteq \Omega : A \text{ abzählbar oder } A^c \text{ abzählbar}\} \subseteq \mathcal{A}$.

Es gilt: $\mathcal{F}$ ist eine σ-Algebra über Ω (Übungsaufgabe!)

Weiter gilt $\mathcal{M} \subseteq \mathcal{F}$.

Somit folgt $\sigma(\mathcal{M}) = \mathcal{F}$.

24.8 Folgerung (Arbeitstechniken für Erzeuger)

Für Mengensysteme $\mathcal{M}, \mathcal{M}_1, \mathcal{M}_2 \subseteq \mathcal{P}(\Omega)$ gelten:

a) $\mathcal{M}_1 \subseteq \mathcal{M}_2 \Longrightarrow \sigma(\mathcal{M}_1) \subseteq \sigma(\mathcal{M}_2)$,

b) $\sigma(\sigma(\mathcal{M})) = \sigma(\mathcal{M})$,

c) $\mathcal{M}_1 \subseteq \sigma(\mathcal{M}_2)$ und $\mathcal{M}_2 \subseteq \sigma(\mathcal{M}_1) \Longrightarrow \sigma(\mathcal{M}_1) = \sigma(\mathcal{M}_2)$.

BEWEIS: a) $\mathcal{M}_1 \subseteq \mathcal{M}_2 \Longrightarrow \mathcal{M}_1 \subseteq \sigma(\mathcal{M}_2) \Longrightarrow \sigma(\mathcal{M}_1) \subseteq \sigma(\mathcal{M}_2)$

b) $\sigma(\mathcal{M}) \subseteq \sigma(\sigma(\mathcal{M}))$ $\surd$

„$=$“ gilt, da $\sigma(\mathcal{M})$ eine σ-Algebra ist.

c) folgt aus a) und b). $\surd$ Beachte:

- Eine σ-Algebra kann verschiedene Erzeuger besitzen.
- Im Allgemeinen kann man $\sigma(\mathcal{M})$ nicht konstruktiv „von innen heraus“ aus $\mathcal{M}$ konstruieren!

24.9 Definition (Borelsche σ-Algebra)

Es bezeichne $\mathcal{O}^k$ das System der offenen Mengen des $\mathbb{R}^k$.

Das System $\mathcal{B}^k := \sigma(\mathcal{O}^k)$ heißt σ-Algebra der Borelmengen des $\mathbb{R}^k$.

24.10 Bemerkungen

a) $O \subseteq \mathbb{R}^k$ offen $\Longleftrightarrow \forall\, x \in O\, \exists\, \varepsilon > 0 : \{y : \|x-y\| < \varepsilon\} \subseteq O$

b) Jede abgeschlossene Menge ist eine Borelmenge.

c) Jede abzählbare Menge ist eine Borelmenge.

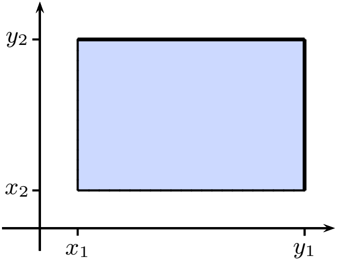
d) Sei $(x,y] := \{z \in \mathbb{R}^k : x < z \leq y\}$ („$<$“ und „$\leq$“ komponentenweise!) und analog (x,y) sowie $[x,y]$. Mit $a_n := \left(\frac{1}{n}, \cdots, \frac{1}{n}\right) \in \mathbb{R}^k$ gilt

$$(x,y] = \bigcap_{n=1}^{\infty} (x, y + a_n) \quad \Longrightarrow (x,y] \in \mathcal{B}^k$$

e) Es gilt $\mathcal{B}^k \neq \mathcal{P}(\mathbb{R}^k)$, aber alle praktisch wichtigen Mengen gehören zu $\mathcal{B}^k$.

Seien

- $\mathcal{A}^k$ das System der abgeschlossenen Mengen des $\mathbb{R}^k$,
- $\mathcal{K}^k$ das System der kompakten Mengen des $\mathbb{R}^k$,
- $\mathcal{I}^k := \{(x,y] : x, y \in \mathbb{R}^k, x < y\}$ das System der halboffenen (verallgemeinerten) Intervalle des $\mathbb{R}^k$,



- $\mathcal{J}^k := \{(-\infty, x] = \{y \in \mathbb{R}^k : y \leq x\} : x \in \mathbb{R}^k\}$.

Memo: $\mathcal{M}_1 \subseteq \sigma(\mathcal{M}_2)$ und $\mathcal{M}_2 \subseteq \sigma(\mathcal{M}_1) \Longrightarrow \sigma(\mathcal{M}_1) = \sigma(\mathcal{M}_2)$

24.11 Satz (Erzeuger der Borelmengen)

Für die Borelsche σ-Algebra $\mathcal{B}^k$ gilt

$$\mathcal{B}^k \;=\; \sigma(\mathcal{A}^k) \;=\; \sigma(\mathcal{K}^k) \;=\; \sigma(\mathcal{I}^k) \;=\; \sigma(\mathcal{J}^k)$$

Beweis: Wir zeigen die Gültigkeit der beiden ersten Gleichheitszeichen (Weiteres in den Übungen).

Es gilt $\mathcal{A}^k \subseteq \sigma(\mathcal{O}^k)$ und $\mathcal{O}^k \subseteq \sigma(\mathcal{A}^k) \Longrightarrow \sigma(\mathcal{A}^k) = \sigma(\mathcal{O}^k) = \mathcal{B}^k$.

Sei $A \in \mathcal{A}^k$, $A_n := A \cap [-n,n]^k \in \mathcal{K}^k \Longrightarrow A = \cup_{n=1}^{\infty}(A \cap [-n,n]^k) \in \sigma(\mathcal{K}^k)$

$\Longrightarrow \mathcal{A}^k \subseteq \sigma(\mathcal{K}^k) \;\Longrightarrow \sigma(\mathcal{A}^k) \subseteq \sigma(\mathcal{K}^k)$. Wegen $\mathcal{K}^k \subseteq \mathcal{A}^k$ gilt $\sigma(\mathcal{K}^k) \subseteq \sigma(\mathcal{A}^k)$

24.12 Satz (Existenz und Eindeutigkeit des Borel-Lebesgue-Maßes)

Es gibt genau eine Funktion $\lambda^k : \mathcal{B}^k \to [0, \infty]$ mit folgenden Eigenschaften:

a) $\lambda^k(\emptyset) = 0$,

b) $\lambda^k\left(\sum_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \lambda^k(B_j)$, falls $B_1, B_2, \ldots \in \mathcal{B}^k$ paarweise disjunkt,

c) $\lambda^k([x, y]) = \prod_{j=1}^{k} (y_j - x_j)$ für alle $x, y \in \mathbb{R}^k$ mit $x \le y$.

λ^k heißt Borel-Lebesgue–Maß im $\mathbb{R}^k$.

- λ^k löst in zufriedenstellender Weise das Problem, Längen, Flächen, Volumina usw. zu messen.
- λ^k ist bewegungsinvariant, d.h. es gilt

$$\lambda^k(T(B)) \ = \ \lambda^k(B), \qquad B \in \mathcal{B}^k,$$

für jede Bewegung (abstandserhaltende Abbildung) $T : \mathbb{R}^k \to \mathbb{R}^k$, d.h. $T(x) = Ux + b$, wobei U orthogonale $(k \times k)$-Matrix und $b \in \mathbb{R}^k$.

24.13 Definition (Axiomensystem von Kolmogorow, 1933)

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$.

Dabei sind $\Omega \neq \emptyset$ und $\mathcal{A}$ eine σ-Algebra über Ω.

Weiter ist $\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:

a) $\mathbb{P}(A) \geq 0, \qquad A \in \mathcal{A}$, (Nichtnegativität)

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

c) $\mathbb{P}(\sum_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ (σ-Additivität)
für jede Folge $(A_j)_{j \geq 1}$ paarweise disjunkter Mengen aus $\mathcal{A}$.

$\mathbb{P}$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{A}$ (kurz: W-Maß).

Jede Menge A aus $\mathcal{A}$ heißt Ereignis.

- Speziell: Diskreter W-Raum, ($\mathcal{A} = \mathcal{P}(\Omega)$, $\Omega_0 \subseteq \Omega$ abzählbar, $\mathbb{P}(\Omega_0) = 1$).
- Alle früher abgeleiteten Folgerungen und einschlägigen Definitionen bleiben erhalten (nur stets $A \in \mathcal{A}$)!

Wir notieren für spätere Zwecke eine wichtige Folgerung (Übungsaufgabe):

Ist (A_n) eine aufsteigende Folge aus $\mathcal{A}$, gilt also $A_n \subseteq A_{n+1}$, $n \in \mathbb{N}$, so folgt

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

Kurz: $A_n \uparrow A := \bigcup_{n=1}^{\infty} A_n \implies \mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Durch Komplementbildung ergibt sich:

Ist (A_n) eine absteigende Folge aus $\mathcal{A}$, gilt also $A_n \supseteq A_{n+1}$, $n \in \mathbb{N}$, so folgt

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

Kurz: $A_n \downarrow A := \bigcap_{n=1}^{\infty} A_n \implies \mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Memo: $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ σ-Algebra, $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ W-Maß

Memo: $\mathcal{B}^k \subseteq \mathcal{P}(\mathbb{R}^k)$ σ-Algebra der Borelmengen im $\mathbb{R}^k$

Bislang:

- Ist $(\Omega, \mathbb{P})$ ein diskreter W-Raum, so heißt jede Abbildung $X : \Omega \to \mathbb{R}$ (reelle) Zufallsvariable.
- $\mathbb{P}^X(B) \; := \; \mathbb{P}(X^{-1}(B))$ ist für jede Teilmenge $B \subseteq \mathbb{R}$ definiert.
- $\mathbb{P}^X$ ist die Verteilung von X (W-Maß auf $\mathcal{P}(\mathbb{R})$).

Jetzt:

- Viele interessante W-Maße über $\mathbb{R}$ sind nur noch auf dem System $\mathcal{B}^1$ der Borelmengen definiert.
- Des Weiteren muss $X^{-1}(B) \in \mathcal{A}$ gefordert werden, da $\mathbb{P}$ nur auf einer σ-Algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ definiert ist.
- Gleiches gilt, wenn $X : \Omega \to \mathbb{R}^k$ ein k-dimensionaler Zufallsvektor ist.

Somit ergibt sich zwangsläufig die folgende Definition:

24.14 Definition (Zufallsvektor, Messbarkeit, Verteilung)

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $X : \Omega \to \mathbb{R}^k$ eine Abbildung.

X heißt k-dimensionaler Zufallsvektor, falls gilt:

$$X^{-1}(B) \in \mathcal{A} \qquad \forall\, B \in \mathcal{B}^k. \tag{24.12}$$

Im Fall $k = 1$ heißt X (reelle) Zufallsvariable.

Bedingung (24.12) heißt (Borel-) Messbarkeit von X. Das durch

$$\mathbb{P}^X(B) \; := \; \mathbb{P}(X^{-1}(B)), \qquad B \in \mathcal{B}^k,$$

definierte Wahrscheinlichkeitsmaß $\mathbb{P}^X$ auf $\mathcal{B}^k$ heißt Verteilung von X.

- $\mathbb{P}^X$ ist W-Maß (!) $\Longrightarrow (\mathbb{R}^k, \mathcal{B}^k, \mathbb{P}^X)$ ist W-Raum
- Ist Q ein W-Maß auf $\mathcal{B}^k$, so liefert die kanonische Konstruktion

  $$\Omega := \mathbb{R}^k, \quad \mathcal{A} := \mathcal{B}^k, \quad \mathbb{P} := Q, \quad X := \mathrm{id}_\Omega$$

  einen W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$ und einen Zufallsvektor $X : \Omega \to \mathbb{R}^k$ mit Verteilung $\mathbb{P}^X = Q$.

Wie konstruiert man (insbesondere nicht diskrete) W-Maße auf $\mathcal{B}^k$?

Sei $f : \mathbb{R}^k \to \mathbb{R}_{\geq 0}$ eine Lebesgue-integrierbare Funktion (→ Analysis 3) mit

$$\int_{\mathbb{R}^k} f(x)\,\mathrm{d}x \;=\; 1.$$

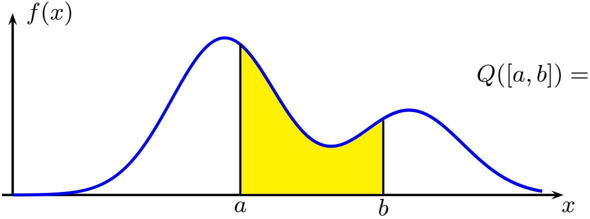
Dann wird durch

$$Q(B) \;:=\; \int_B f(x)\,\mathrm{d}x \;:=\; \int_{\mathbb{R}^k} \mathbf{1}_B(x)\,f(x)\,\mathrm{d}x, \quad B \in \mathcal{B}^k,$$

ein W-Maß auf $\mathcal{B}^k$ definiert. Hier: $x = (x_1, \ldots, x_k)$ und $\mathrm{d}x = \mathrm{d}x_1 \ldots \mathrm{d}x_k$.

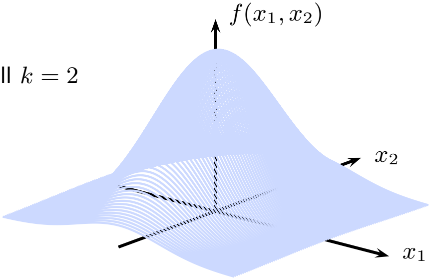
- Damit Q wohldefiniert und σ-additiv ist, sind obige Integrale grundsätzlich als Lebesgue-Integrale zu verstehen.
- Die σ-Additivität von Q folgt aus dem Satz von der monotonen Konvergenz (Satz von Beppo Levi der Maß- und Integrationstheorie).
- Im Folgenden werden f und B so beschaffen sein, dass für konkrete Berechnungen auch mit dem Riemann-Integral gearbeitet werden kann.
- f heißt (Wahrscheinlichkeits-)Dichte von Q.

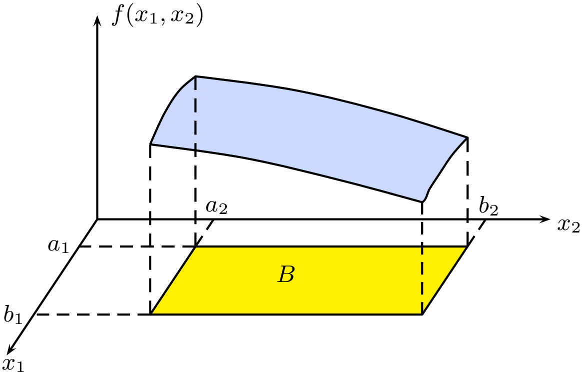
Fall $k = 1$ (im Folgenden überwiegend):



$Q([a,b]) = \int_a^b f(x)\,\mathrm{d}x$

Fall $k = 2$





$$\int_B f(x)\,\mathsf{d}x \;=\; \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x_1,x_2)\,\mathsf{d}x_2 \right)\,\mathsf{d}x_1$$

Die folgende Verteilung ist konzeptionell die einfachste stetige Verteilung.

Hier sind Wahrscheinlichkeiten direkt mit Längen, Flächen oder (höheren) Volumina verknüpft.

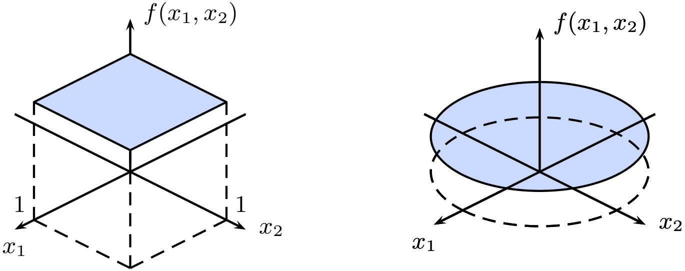
24.15 Definition (Gleichverteilung auf einer Menge $B \subset \mathbb{R}^k$)

Sei $B \subset \mathbb{R}^k$ eine beschränkte Borelmenge mit $\lambda^k(B) > 0$. Dann wird durch

$$f(x) \ := \ \frac{1}{\lambda^k(B)}, \text{ falls } x \in B, \quad \text{und } f(x) := 0, \qquad \text{sonst},$$

eine Wahrscheinlichkeitsdichte auf $\mathbb{R}^k$ definiert.

Ein Zufallsvektor X mit der Dichte f heißt gleichverteilt auf B, kurz: $X \sim \mathrm{U}(B)$.



Dichte der Gleichverteilung auf dem Einheitsquadrat und auf dem Einheitskreis

24.16 Definition und Satz (Verteilungsfunktion)

Ist X eine Zufallsvariable auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$, so heißt die durch

$$F(x) \ := \ \mathbb{P}(X \leq x), \qquad x \in \mathbb{R},$$

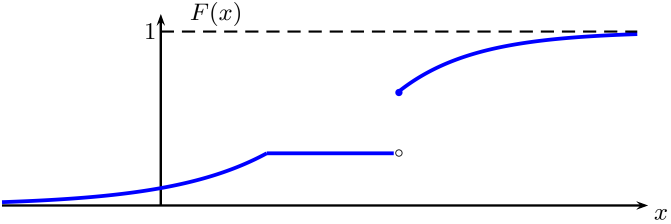
definierte Funktion $F : \mathbb{R} \rightarrow [0, 1]$ die Verteilungsfunktion von X.

F besitzt folgende Eigenschaften:

a) F ist monoton wachsend,

b) F ist rechtsseitig stetig,

c) es gilt $\lim_{n\to\infty} F(-n) \ = \ 0, \quad \lim_{n\to\infty} F(n) \ = \ 1.$



Memo: $F(x) = \mathbb{P}(X \le x) = \mathbb{P}^X((-\infty, x])$

a) F ist monoton: $x \le y \Longrightarrow (-\infty, x] \subseteq (-\infty, y] \Longrightarrow F(x) \le F(y)$ $\surd$

b) F ist rechtsseitig stetig:

Sei $x \in \mathbb{R}$ beliebig, (x_n) eine beliebige Folge mit $x_n \downarrow x$.

Sei $B_n := (-\infty, x_n]$, $n \in \mathbb{N}$, $B := (-\infty, x]$.

Wegen $x_{n+1} \le x_n$ gilt $B_{n+1} \subseteq B_n$, $n \in \mathbb{N}$, sowie $B = \cap_{n=1}^{\infty} B_n$, d.h. $B_n \downarrow B$.

$\mathbb{P}^X$ stetig von oben $\Longrightarrow$

$$F(x) = \mathbb{P}^X((-\infty, x]) = \mathbb{P}^X(B) = \lim_{n\to\infty} \mathbb{P}^X(B_n) = \lim_{n\to\infty} F(x_n) \ \surd$$

c) Es gilt $(-\infty, -n] \downarrow \emptyset$ sowie $(-\infty, n] \uparrow \mathbb{R}$.

$\mathbb{P}^X$ stetig von oben $\Longrightarrow$

$$0 = \mathbb{P}^X(\emptyset) = \lim_{n\to\infty} \mathbb{P}^X((-\infty, -n]) = \lim_{n\to\infty} F(-n).$$

$\mathbb{P}^X$ stetig von unten $\Longrightarrow$

$$1 = \mathbb{P}^X(\mathbb{R}) = \lim_{n\to\infty} \mathbb{P}^X((-\infty, n]) = \lim_{n\to\infty} F(n). \ \surd$$

24.17 Satz (Weitere Eigenschaften von Verteilungsfunktionen)

Es sei F die Verteilungsfunktion einer Zufallsvariablen X. Dann gelten:

a) $\mathbb{P}(a < X \leq b) \; = \; F(b) - F(a), \qquad a, b \in \mathbb{R}, \; a < b,$

b) $\mathbb{P}(X = x) \; = \; F(x) - F(x-), \qquad x \in \mathbb{R}.$

Dabei bezeichne

$$F(x-) := \lim_{y \to x, y < x} F(y)$$

den linksseitigen Grenzwert von F an der Stelle x.

c) F besitzt höchstens abzählbar viele Unstetigkeitsstellen.

BEWEIS: Übungsaufgabe!

24.18 Satz (Existenz- und Eindeutigkeitssatz)

Sei $F : \mathbb{R} \to [0,1]$ eine beliebige monoton wachsende rechtsseitig stetige Funktion mit $\lim_{n\to\infty} F(-n) = 0$ und $\lim_{n\to\infty} F(n) = 1$.

Dann existiert genau ein W-Maß Q auf der Borelschen σ-Algebra $\mathcal{B}^1$ mit

$$Q((-\infty, x]) \ = \ F(x), \qquad x \in \mathbb{R}. \tag{24.13}$$

Beweis: mit dem Maßfortsetzungssatz (→ Maßtheorie).

24.19 Folgerung

Zu jeder Funktion $F : \mathbb{R} \to [0,1]$ mit obigen Eigenschaften existieren ein W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$ und eine Zufallsvariable $X : \Omega \to \mathbb{R}$, so dass X die Verteilungsfunktion F besitzt, so dass also gilt:

$$F(x) \ = \ \mathbb{P}(X \le x), \qquad x \in \mathbb{R}.$$

Beweis: Satz 24.18 $\Longrightarrow$ $\exists$ W-Maß Q auf $\mathcal{B}^1$ mit (24.13). Setze (kanonische Konstruktion!) $\Omega := \mathbb{R}$, $\mathcal{A} := \mathcal{B}^1$, $\mathbb{P} := Q$ und $X := \mathrm{id}_\Omega$. $\surd$

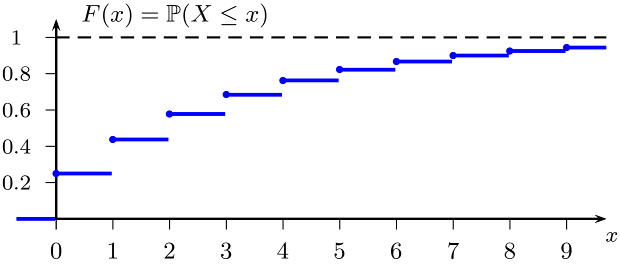
24.20 Definition (Diskrete Zufallsvariable/Verteilungsfunktion)

Eine Zufallsvariable X heißt diskret (verteilt)

$$:\Longleftrightarrow \ \exists B \subseteq \mathbb{R}, \ B \text{ abzählbar } \ \text{und } \mathbb{P}(X \in B) = 1.$$

Man sagt auch, dass X eine diskrete Verteilung besitzt.

Die Verteilungsfunktion einer diskreten Zufallsvariablen heißt diskrete Verteilungsfunktion.



Verteilungsfunktion einer diskreten Verteilung (G(1/4))

24.21 Definition (Stetige Zufallsvariable, Dichte)

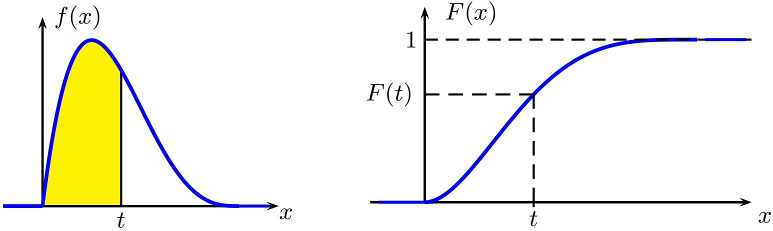
Eine Zufallsvariable X heißt (absolut) stetig (verteilt)

$$:\Longleftrightarrow \quad \exists f : \mathbb{R} \to \mathbb{R}_{\geq 0} \text{ mit } \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1 \text{ und}$$

$$F(x) \;=\; \mathbb{P}(X \leq x) \;=\; \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

In diesem Fall sagt man auch, X habe eine (absolut) stetige Verteilung.

f heißt Dichte von X bzw. Dichte der Verteilung(sfunktion) von X.



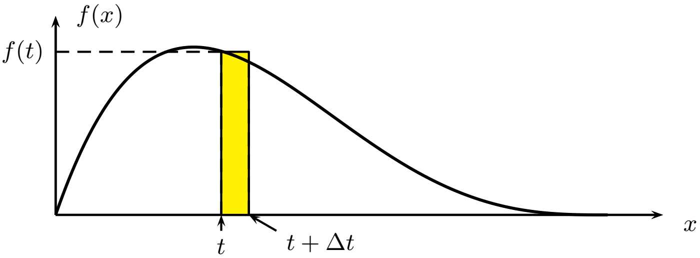
Dichte und Verteilungsfunktion einer stetigen Zufallsvariablen

Sei X eine stetige Zufallsvariable mit Dichte f.

Sei t eine Stetigkeitsstelle von f. Für kleines $\Delta t > 0$ gilt

$$\mathbb{P}(t \le X \le t + \Delta t) \;=\; \int_t^{t+\Delta t} f(x)\,\mathsf{d}x \;\approx\; \Delta t\, f(t) \quad \Longrightarrow$$

$$f(t) \;\approx\; \frac{1}{\Delta t}\,\mathbb{P}(t \le X \le t + \Delta t)$$



Vgl. Massendichte (Grenzwert von Masse pro Volumeneinheit)

Memo: Stetige Zufallsvariable: $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$.

Beachte:

- f kann an abzählbar vielen Stellen beliebig abgeändert werden, ohne dass sich F und damit die Verteilung $\mathbb{P}^X$ von X ändert.
- Im Gegensatz zur Verteilungsfunktion ist also die Dichte einer stetigen Zufallsvariablen nicht eindeutig bestimmt.
- In allen auftretenden Beispielen wird f bis auf höchstens endlich viele Ausnahmestellen stetig sein.
- Nach dem Ersten Hauptsatz der Differential- und Integralrechnung gilt $F'(x) = f(x)$ in jeder Stetigkeitsstelle x von f.
- Sei andererseits F eine Verteilungsfunktion, die außerhalb einer endlichen (evtl. leeren) Menge M stetig differenzierbar ist. Dann wird durch

  $$f(x) \;:=\; F'(x), \qquad x \in \mathbb{R} \setminus M,$$

  und $f(x) := 0$, falls $x \in M$, eine Dichte definiert, und es gilt

  $$F(x) \;=\; \mathbb{P}(X \leq x) \;=\; \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

- Jede Verteilungsfunktion F ist fast überall differenzierbar (d.h. mit Ausnahme einer Menge N mit $\lambda^1(N) = 0$, tiefliegender Satz der Analysis)
- Für die Ableitung F' gilt

$$F(x) \;\le\; \int_{-\infty}^{x} F'(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}.$$

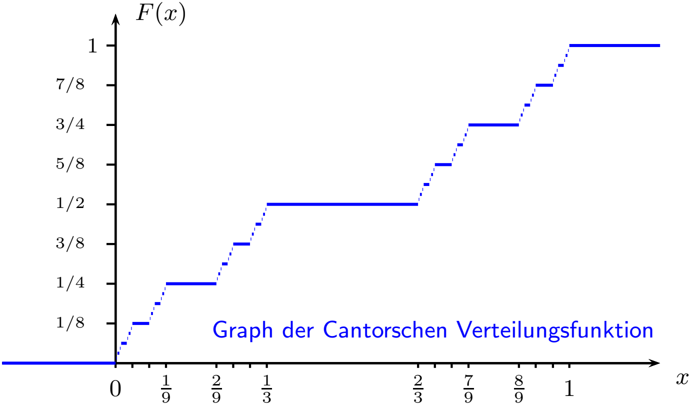
Für eine stetige Zufallsvariable steht hier stets das Gleichheitszeichen.

Wie das folgende prominente Beispiel zeigt, kann

$$\int_{-\infty}^{x} F'(t)\,\mathrm{d}t \;=\; 0, \qquad x \in \mathbb{R},$$

gelten.

Es gibt eine stetige Verteilungsfunktion, die außerhalb einer Menge $C \in \mathcal{B}^1$ mit $\lambda^1(C) = 0$ (sog. Nullmenge) differenzierbar ist und dort die Ableitung 0 hat.



$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \ldots = \sum_{k=1}^{\infty} \frac{2^{k-1}}{3^k} = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \cdot \frac{1}{1 - \frac{2}{3}} = 1$$

25 Grundlegende stetige Verteilungen

25.1 Definition (Gleichverteilung auf einem Intervall)

Seien $a, b \in \mathbb{R}$ mit $a < b$. Die Zufallsvariable X hat eine Gleichverteilung auf dem Intervall (a, b), falls X die Dichte

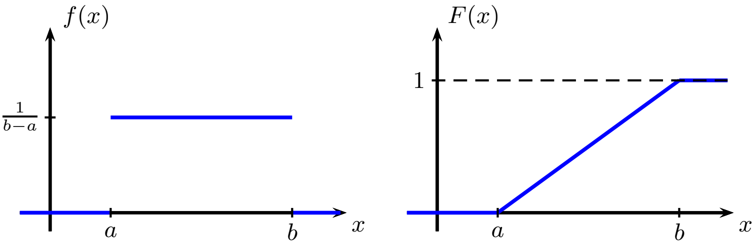
$$f(x) := \begin{cases} \frac{1}{b-a}, & \text{falls} \quad a < x < b, \\ 0, & \text{sonst} \end{cases}$$

besitzt. Wir schreiben hierfür kurz $X \sim \mathsf{U}(a, b)$.

Die Verteilungsfunktion von X hat die Darstellung

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

Beachte: Pseudozufallszahlengeneratoren simulieren die Verteilung $\mathsf{U}(0, 1)$.



Dichte und Verteilungsfunktion der Gleichverteilung $\mathsf{U}(a,b)$

25.2 Satz Aus $X \sim \mathsf{U}(0,1)$ folgt $a + (b-a)X \sim \mathsf{U}(a,b)$.

Beweis: Sei $Y := a + (b-a)X$. Wegen $\mathbb{P}(0 \leq X \leq 1) = 1$ gilt $\mathbb{P}(a \leq Y \leq b) = 1$. Für x mit $a < x < b$ gilt

$$\mathbb{P}(Y \leq x) \ = \ \mathbb{P}(a + (b-a)X \leq x) \ = \ \mathbb{P}\left(X \leq \frac{x-a}{b-a}\right) \ = \ \frac{x-a}{b-a}. \ \surd$$

25.3 Definition (Exponentialverteilung)

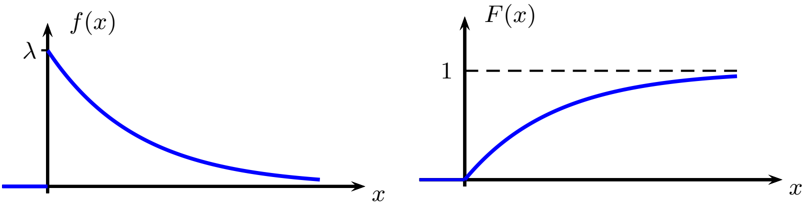
Die Zufallsvariable X hat eine Exponentialverteilung mit Parameter $\lambda > 0$, falls X die Dichte

$$f(x) := \begin{cases} \lambda\, \mathbf{e}^{-\lambda x}, & \text{falls } x > 0, \\ 0, & \text{sonst,} \end{cases}$$

besitzt. Wir schreiben hierfür kurz $X \sim \mathrm{Exp}(\lambda)$.

Die Verteilungsfunktion von X ist

$$F(x) = \begin{cases} 1 - \mathbf{e}^{-\lambda x}, & \text{falls } x > 0, \\ 0, & \text{sonst.} \end{cases}$$



Dichte und Verteilungsfunktion der Exponentialverteilung

- Die Verteilung Exp(λ) ist das „gedächtnislose Analogon“ der geometrischen Verteilung bei kontinuierlicher Zeitmessung.

  Für $t, h > 0$ gilt

$$\begin{aligned}\mathbb{P}(X \geq t+h | X \geq t) &= \frac{\mathbb{P}(X \geq t+h,\ X \geq t)}{\mathbb{P}(X \geq t)} = \frac{\mathbb{P}(X \geq t+h)}{\mathbb{P}(X \geq t)} \\ &= \frac{1-F(t+h)}{1-F(t)} = \frac{\exp(-\lambda(t+h))}{\exp(-\lambda t)} = \mathbf{e}^{-\lambda h} \\ &= \mathbb{P}(X \geq h).\end{aligned}$$

- Der Parameter λ bewirkt nur eine Skalenänderung: Es gilt

$$X \sim \mathsf{Exp}(1) \Longrightarrow \frac{1}{\lambda} X \sim \mathsf{Exp}(\lambda) \qquad \text{(Übungsaufgabe!)}$$

- Es besteht ein direkter Zusammenhang mit der Gleichverteilung:

$$X \sim \mathsf{U}(0,1) \Longrightarrow -\frac{1}{\lambda} \log(1-X) \sim \mathsf{Exp}(\lambda) \qquad \text{(Übungsaufgabe!)}$$

25.4 Definition (Normalverteilung)

Die Zufallsvariable X hat eine Normalverteilung mit Parametern μ und σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$), falls X die Dichte

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

besitzt. Wir schreiben hierfür kurz $X \sim \mathrm{N}(\mu, \sigma^2)$.

Die Verteilung $\mathrm{N}(0,1)$ heißt Standard-Normalverteilung oder standardisierte Normalverteilung (vgl. Zentraler Grenzwertsatz).

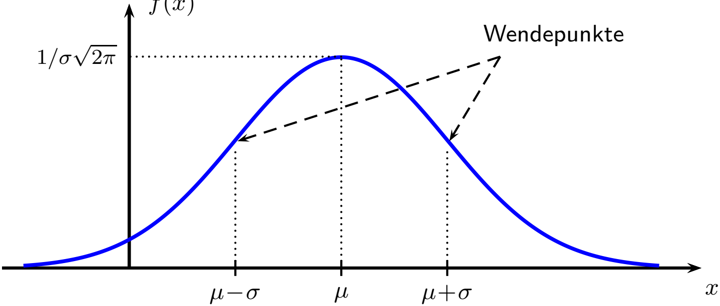
Beachte: Mit $\varphi(x) := \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{x^2}{2}\right)$ (Gaußsche Glockenkurve) gilt

$$f(x) = \frac{1}{\sigma}\,\varphi\left(\frac{x-\mu}{\sigma}\right).$$

Mit der Substitution $t := (x-\mu)/\sigma$ folgt

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \;=\; \frac{1}{\sigma}\int_{-\infty}^{\infty} \varphi\left(\frac{x-\mu}{\sigma}\right)\mathrm{d}x \;=\; \int_{-\infty}^{\infty} \varphi(t)\,\mathrm{d}t \;=\; 1.$$

Memo: $f(x) \;=\; \dfrac{1}{\sigma\sqrt{2\pi}}\,\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$



Dichte der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$

Verteilungsfunktion F der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$?

Memo: $\Phi(x) \;=\; \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t, \quad \varphi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$

Memo: Dichte von $\mathrm{N}(\mu, \sigma^2) : f(t) \;=\; \frac{1}{\sigma}\,\varphi\left(\frac{t-\mu}{\sigma}\right)$

Also:

$$\begin{aligned} F(x) \;&=\; \int_{-\infty}^{x} f(t)\,\mathrm{d}t \;=\; \int_{-\infty}^{x} \frac{1}{\sigma}\,\varphi\left(\frac{t-\mu}{\sigma}\right)\mathrm{d}t \\ &=\; \int_{-\infty}^{(x-\mu)/\sigma} \varphi(z)\,\mathrm{d}z \quad \left(\text{mit } z := \frac{x-\mu}{\sigma} \right) \\ &=\; \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

25.5 Satz Aus $X \sim \mathrm{N}(0,1)$ folgt $\mu + \sigma X \sim \mathrm{N}(\mu, \sigma^2)$.

BEWEIS: $\mathbb{P}(\mu + \sigma X \le x) \;=\; \mathbb{P}\left(X \le \frac{x-\mu}{\sigma}\right) \;=\; \Phi\left(\frac{x-\mu}{\sigma}\right).$

Memo: Verteilungsfunktion von $\mathrm{N}(\mu, \sigma^2)$: $F(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$

25.6 Satz Aus $X \sim \mathrm{N}(\mu, \sigma^2)$ folgt $\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$.

BEWEIS:

$$\begin{aligned} \mathbb{P}\left(\frac{X-\mu}{\sigma} \le t\right) &= \mathbb{P}(X \le \mu + \sigma t) \\ &= \Phi\left(\frac{\mu + \sigma t - \mu}{\sigma}\right) = \Phi(t). \surd \end{aligned}$$

Fazit: Um Wahrscheinlichkeiten für eine beliebige Normalverteilung auszurechnen, reicht die (Tabelle der) Verteilungsfunktion Φ der Standard-Normalverteilung aus.

- Die Normalverteilung ist eine der wichtigsten stetigen Verteilungen.
- Ihre Bedeutung beruht vor allem auf dem Zentralen Grenzwertsatz. Zur Erinnerung: $X_1, X_2, , \ldots$ unabhängig, identisch verteilt, $\mu := \mathbb{E}\, X_1$, $0 < \sigma^2 := \mathbb{V}(X_1) < \infty$, $S_n := \sum_{j=1}^n X_j \Longrightarrow$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\,\mu}{\sigma\sqrt{n}} \le x\right) \;=\; \Phi(x) \;=\; \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t.$$

(Zentraler Grenzwertsatz von Lindeberg-Lévy).

- $X \mapsto \mu + \sigma X$ und $X \mapsto a + (b-a)X$ sind Transformationen der Zufallsvariablen X.
- Sei allgemein X eine ZV mit Verteilungsfunktion F und Dichte f.
- Sei $T : \mathbb{R} \to \mathbb{R}$ eine messbare Funktion ($T^{-1}(B) \in \mathcal{B}^1 \ \forall B \in \mathcal{B}^1$).
- 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 VF 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 $\mathbb{P}(X = -y^{1/4}) = 0$.
- $T(X) = \mathbf{e}^X$: $\{\mathbf{e}^X \leq y\} = \{X \leq \log y\}$ für $y > 0$,
  also $G(y) \ = \ \mathbb{P}(Y \leq y) \ = \ F(\log y)$, $y > 0$; $G(y) = 0$, sonst.

25.7 Satz (Streng monotone Transformationen)

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 Restriktion 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 jedem dieser beiden Fälle 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:

Sei T streng monoton wachsend. Für $y \in T(O)$ gilt

$$\begin{aligned} G(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(T(X) \le y) = \mathbb{P}(X \le T^{-1}(y)) \\ &= F(T^{-1}(y)). \end{aligned}$$

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.

25.8 Beispiel (Lognormalverteilung)

- Sei $X \sim \mathsf{N}(\mu, \sigma^2)$,
- $T(x) := \mathbf{e}^x$, $x \in \mathbb{R}$,
- $Y := T(X) = \mathbf{e}^X$.
- Sei $G(y) := \mathbb{P}(Y \leq y)$, $y \in \mathbb{R}$.
- Für $y \leq 0$ gilt $G(y) = 0$
- Für $y > 0$ ist

$$G(y) \;=\; \mathbb{P}(Y \leq y) \;=\; \mathbb{P}(\mathbf{e}^X \leq y) \;=\; \mathbb{P}(X \leq \log y) \;=\; \Phi\left(\frac{\log y - \mu}{\sigma}\right)$$

- 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).

25.9 Definition (Lognormalverteilung)

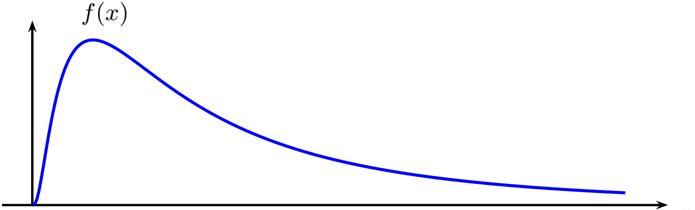
Die positive Zufallsvariable X besitzt eine Lognormalverteilung mit Parametern μ und σ^2 ($\mu \in \mathbb{R}, \sigma^2 > 0$), falls X die Dichte

$$f(x) \;=\; \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$$

für $x > 0$ und $f(x) = 0$, sonst, besitzt.

Wir schreiben hierfür kurz $X \sim \mathrm{LN}(\mu, \sigma^2)$.

Beachte: $X \sim \mathrm{LN}(\mu, \sigma^2) \Longrightarrow \log X \sim \mathrm{N}(\mu, \sigma^2)$.



Dichte der Lognormalverteilung

26 Kenngrößen von Verteilungen

Sei $(\Omega, \mathbb{P})$ ein diskreter W-Raum, $X : \Omega \to \mathbb{R}$ Zufallsvariable.

Sei $\Omega_0 \subseteq \Omega$ abzählbar, $\mathbb{P}(\Omega_0) = 1$.

$$\text{Der Erwartungswert von } X \text{ existiert } :\Longleftrightarrow \sum_{\omega \in \Omega_0} |X(\omega)| \, \mathbb{P}(\{\omega\}) < \infty.$$

In diesem Fall heißt

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

der Erwartungswert von X.

Falls $\sum_{j \geq 1} \mathbb{P}(X = x_j) = 1$, so gilt:

$$\text{Der Erwartungswert von } X \text{ existiert } \Longleftrightarrow \sum_{j \geq 1} |x_j| \, \mathbb{P}(X = x_j) < \infty.$$

In diesem Fall gilt

$$\mathbb{E}\, X \ = \sum_{j \geq 1} x_j \, \mathbb{P}(X = x_j) \qquad \text{(Darstellungsformel)}$$

Jetzt: $(\Omega, \mathcal{A}, \mathbb{P})$ allgemeiner W-Raum, $X : \Omega \to \mathbb{R}$ Zufallsvariable.

Ziel: $\mathbb{E}\, X$ definieren.

Beachte: $\mathbb{P}(X = x) = 0$ für jedes $x \in \mathbb{R}$ möglich (bei stetigem X).

Sei zunächst $X \geq 0$.

Ansatz: Zerlege den Wertebereich $[0, \infty)$ von X und approximiere X durch eine Zufallsvariable X_n mit endlichem Wertebereich:

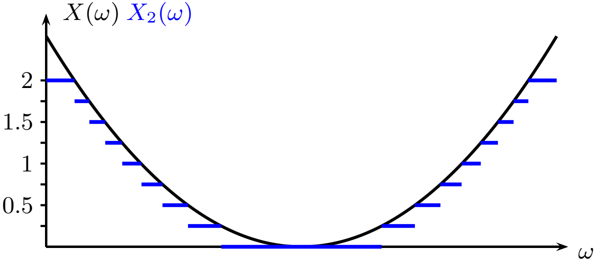
Sei für $\omega \in \Omega$

$$X_n(\omega) := \begin{cases} 0, & \text{falls } X(\omega) \;=\; 0, \\ \dfrac{j-1}{2^n}, & \text{falls } \dfrac{j-1}{2^n} < X(\omega) \leq \dfrac{j}{2^n}, \qquad j \in \{1, 2, \ldots, n2^n\} \\ n, & \text{falls } X(\omega) > n. \end{cases}$$

Es gilt

- $X_n \leq X_{n+1}$, $n \in \mathbb{N}$,
- $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ für jedes $\omega \in \Omega$.

$$X_n(\omega) \; = \; \begin{cases} 0, & \text{falls } X(\omega) \; = \; 0, \\ \dfrac{j-1}{2^n}, & \text{falls } \dfrac{j-1}{2^n} < X(\omega) \leq \dfrac{j}{2^n}, \qquad j \in \{1, 2, \ldots, n2^n\} \\ n, & \text{falls } X(\omega) > n. \end{cases}$$



Sei F die Verteilungsfunktion von X. Es gilt

$$\mathbb{P}\left(X_n = \frac{j-1}{2^n}\right) \; = \; F\left(\frac{j}{2^n}\right) - F\left(\frac{j-1}{2^n}\right), \quad \mathbb{P}(X_n = n) = 1 - F(n)$$

Motiviert durch die Darstellungsformel setzt man

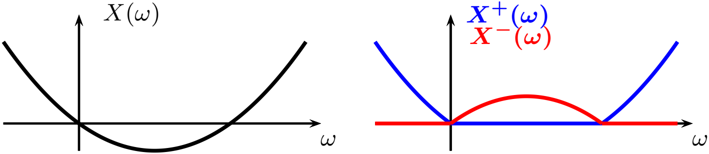
$$\mathbb{E}(X_n) \; := \; \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \left(F\left(\frac{j}{2^n}\right) - F\left(\frac{j-1}{2^n}\right)\right) + n(1-F(n))$$

sowie $$\mathbb{E}(X) \; := \; \lim_{n\to\infty} \mathbb{E}(X_n) \qquad (\leq \infty!)$$

Für beliebiges X setzt man

$$X^+(\omega) \; := \; \max(X(\omega),0)$$ (sog. Positivteil von X),

$$X^-(\omega) \; := \; -\min(X(\omega),0)$$ (sog. Negativteil von X).



Beachte: $X = X^+ - X^-, \quad |X| = X^+ + X^-$.

26.1 Definition (Erwartungswert)

Der Erwartungswert von X existiert, falls gilt: $\mathbb{E}(X^+) < \infty$, $\mathbb{E}(X^-) < \infty$.

In diesem Fall heißt

$$\mathbb{E}(X) \ := \ \mathbb{E}(X^+) - \mathbb{E}(X^-) \ =: \ \int_\Omega X \, \mathrm{d}\mathbb{P}$$

der Erwartungswert von X.

- Die Darstellung $\int_\Omega X \mathrm{d}\mathbb{P}$ entspricht $\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$
- Alle strukturellen Rechenregeln (Linearität, Monotonie) bleiben erhalten.
- Weiterhin gilt $\mathbb{E}\,\mathbf{1}_A \ = \ \mathbb{P}(A)$.

Memo: $\mathbb{E}\,X = \mathbb{E}\,X^+ - \mathbb{E}\,X^-,\ X^+ = \max(X,0),\ X^- = -\min(X,0).$

26.2 Satz (Darstellungsformel für stetige Zufallsvariablen)

Es sei X eine stetige Zufallsvariable mit Dichte f. Dann gilt:

Der Erwartungswert von X existiert $\iff \int_{-\infty}^{\infty} |x|\, f(x)\, \mathrm{d}x < \infty.$

In diesem Fall gilt $\qquad \mathbb{E}\,X = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x.$

Beweis:

$$X_n^+(\omega) = \begin{cases} 0, & \text{falls } X^+(\omega) = 0, \\ \dfrac{j-1}{2^n}, & \text{falls } \dfrac{j-1}{2^n} < X^+(\omega) \le \dfrac{j}{2^n} \qquad j \in \{1,2,\ldots,n2^n\} \\ n, & \text{falls } X^+(\omega) > n. \end{cases}$$

$$\mathbb{E}\,X_n^+ = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \cdot \left(F\left(\frac{j}{2^n}\right) - F\left(\frac{j-1}{2^n}\right)\right) + n(1-F(n))$$

$$\begin{aligned}
\mathbb{E}\,X_n^+ &= \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \left(F\left(\frac{j}{2^n}\right) - F\left(\frac{j-1}{2^n}\right)\right) + n(1-F(n)) \\
&= \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \int_{(j-1)/2^n}^{j/2^n} f(x)\,\mathrm{d}x \;+\; n \int_n^\infty f(x)\,\mathrm{d}x \\
&\le \sum_{j=1}^{n2^n} \int_{(j-1)/2^n}^{j/2^n} x\,f(x)\,\mathrm{d}x \;+ \int_n^\infty x\,f(x)\,\mathrm{d}x \;=\; \int_0^\infty x\,f(x)\,\mathrm{d}x.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\,X_n^+ &\ge \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \int_{(j-1)/2^n}^{j/2^n} f(x)\,\mathrm{d}x \\
&= \sum_{j=1}^{n2^n} \int_{(j-1)/2^n}^{j/2^n} x\,f(x)\,\mathrm{d}x \;- \sum_{j=1}^{n2^n} \int_{(j-1)/2^n}^{j/2^n} \left(x - \frac{j-1}{2^n}\right) f(x)\,\mathrm{d}x \\
&\ge \int_0^n x\,f(x)\,\mathrm{d}x - \frac{1}{2^n} \int_0^n f(x)\,\mathrm{d}x \;\to\; \int_0^\infty x\,f(x)\,\mathrm{d}x
\end{aligned}$$

Also: $\mathbb{E}\,X^+ \;=\; \lim_{n\to\infty} \mathbb{E}\,X_n^+ \;=\; \int_0^\infty x\,f(x)\,\mathrm{d}x$

Analog:

$$\mathbb{E}\,X^- \;=\; -\int_{-\infty}^{0} x\,f(x)\,\mathrm{d}x \Longrightarrow \;\mathbb{E}\,X \;=\; \mathbb{E}\,X^+ - \mathbb{E}\,X^- \;=\; \int_{-\infty}^{\infty} x\,f(x)\,\mathrm{d}x.$$

Beachte: $\displaystyle\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$ ist „stetiges Analogon“ von $\displaystyle\sum_{j\geq 1} x_j \mathbb{P}(X = x_j)$

Allgemeiner gilt (ohne Beweis → Wahrscheinlichkeitstheorie):

26.3 Satz (Darstellungsformel für $\mathbb{E}g(X)$ (stetige Zufallsvariable))

Es seien X eine stetige Zufallsvariable mit Dichte f und $g : \mathbb{R} \to \mathbb{R}$ eine messbare Funktion. Dann gilt:

Der Erwartungswert von $g(X)$ existiert $\iff \displaystyle\int_{-\infty}^{\infty} |g(x)|\,f(x)\,\mathrm{d}x < \infty.$

In diesem Fall folgt $\qquad \displaystyle\mathbb{E}\,g(X) \;=\; \int_{-\infty}^{\infty} g(x)\,f(x)\,\mathrm{d}x.$

vgl. „diskretes Analogon“ $\qquad \displaystyle\mathbb{E}\,g(X) \;=\; \sum_{j\geq 1} g(x_j)\,\mathbb{P}(X = x_j)$

26.4 Definition (Erwartungswert, Varianz, Momente)

Sei X eine Zufallsvariable auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$. Dann heißen (stets im Fall der Existenz!)

- $\mathbb{E}\,X$ der Erwartungswert von X,
- $\mathbb{E}\,X^k$ das k-te Moment von X ($k \in \mathbb{N}$),
- $\mathbb{V}(X) := \mathbb{E}(X - \mathbb{E}\,X)^2$ die Varianz von X,
- $\mathbb{E}(X - \mathbb{E}\,X)^k$ das k-te zentrale Moment von X ($k \in \mathbb{N}$).

Beachte:

- Besitzt X eine Dichte f, so folgt aus der Darstellungsformel für $\mathbb{E}g(X)$:

$$\mathbb{E}(X^k) \;=\; \int_{-\infty}^{\infty} x^k \, f(x) \, \mathrm{d}x.$$

- Für $r \in \{1, 2, \ldots, k-1\}$ gilt $|X|^r \leq 1 + |X|^k$ (punktweise auf Ω).

  Folgerung: $\mathbb{E}|X|^k < \infty \Longrightarrow \mathbb{E}\,|X|^r \leq 1 + \mathbb{E}\,|X|^k < \infty$

26.5 Beispiel (Normalverteilung $\mathsf{N}(0,1)$)

Es sei $X \sim \mathsf{N}(0,1) \implies f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$.

Für jedes $k \in \mathbb{N}$ gilt

$$\mathbb{E}\,|X|^k = \int_{-\infty}^{\infty} |x|^k \, \varphi(x)\, \mathrm{d}x < \infty \; (!)$$

Wegen $\varphi(x) = \varphi(-x)$, $x \in \mathbb{R}$, gilt $\mathbb{E}\,X^{2n+1} = 0$, $n \in \mathbb{N}_0$.

Sei $n \in \mathbb{N}$. Wegen $\varphi'(x) = -x\varphi(x)$ folgt mit partieller Integration

$$\begin{aligned} I_{2n} \quad &:= \quad \mathbb{E}\,X^{2n} = \int_{-\infty}^{\infty} x^{2n}\varphi(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^{2n-1}\, x\varphi(x)\,\mathrm{d}x \\ &= \quad -x^{2n-1}\varphi(x)\Big|_{-\infty}^{\infty} + (2n-1)\int_{-\infty}^{\infty} x^{2n-2}\varphi(x)\,\mathrm{d}x = (2n-1)I_{2n-2} \end{aligned}$$

Wegen $I_0 = 1$ folgt $\qquad \mathbb{E}\,X^{2n} = \prod_{j=1}^{n}(2j-1) = \frac{(2n)!}{2^n\, n!}$

Insbesondere: $X \sim \mathsf{N}(0,1) \implies \mathbb{E}\,X = 0, \quad \mathbb{V}(X) = 1 \quad (= \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2)$

Memo: $Y \sim \mathrm{N}(0,1) \implies \mathbb{E}(Y) = 0, \quad \mathbb{V}(Y) = 1.$

Diese Beziehungen rechtfertigen die Bezeichnung Standardnormalverteilung.

26.6 Beispiel (Normalverteilung $\mathrm{N}(\mu, \sigma^2)$)

Sei $X \sim \mathrm{N}(\mu, \sigma^2)$. Es gilt

$$X \sim \sigma Y + \mu, \quad \text{wobei} \quad Y \sim \mathrm{N}(0,1).$$

Es folgt

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\sigma Y + \mu) = \sigma\, \mathbb{E}\, Y + \mu = \mu, \\ \mathbb{V}(X) &= \mathbb{V}(\sigma Y + \mu) = \sigma^2\, \mathbb{V}(Y) = \sigma^2 \end{aligned}$$

26.7 Definition (Quantile, Quantilfunktion)

Seien X eine Zufallsvariable mit Verteilungsfunktion F und $0 < p < 1$. Die Zahl

$$Q_p \ := \ Q_p(F) \ := \ Q_p(X) \ := \ F^{-1}(p) \ := \ \inf\{x \in \mathbb{R} : F(x) \geq p\} \quad (26.14)$$

heißt p-Quantil (der Verteilung) von F (bzw. von X).

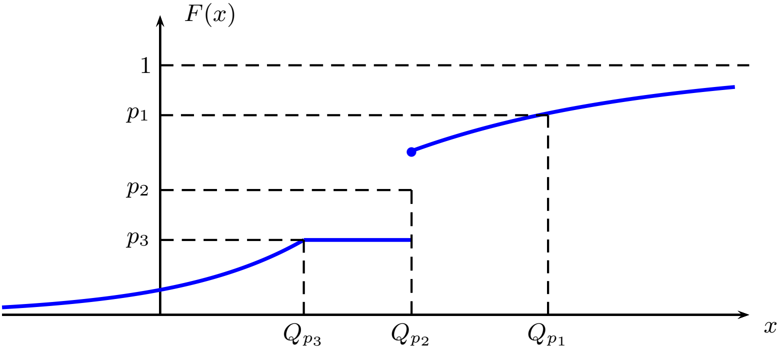
Die durch (26.14) definierte Funktion $F^{-1} : (0,1) \to \mathbb{R}$ heißt Quantilfunktion zu F.

- F^{-1} wohldefiniert, da $F(x) \to 1$ für $x \to \infty$, $F(x) \to 0$ für $x \to -\infty$.
- F^{-1} ist nicht (unbedingt) die Umkehrfunktion zu F.
- Analog zu empirischen Quantilen heißen

$Q_{1/2}$	der Median,
$Q_{1/4}$	das untere Quartil,
$Q_{3/4}$	das obere Quartil,
$Q_{3/4} - Q_{1/4}$	der Quartilsabstand

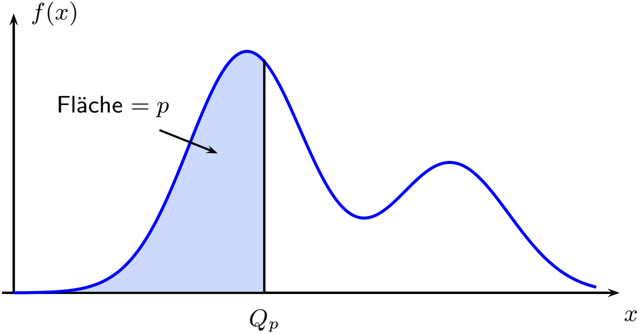
(der Verteilung) von F (bzw. von X).

Memo: $F^{-1}(p) \;=\; \inf\{x \in \mathbb{R} : F(x) \geq p\}$



Zur Definition des p-Quantils

Sei X eine stetige Zufallsvariable mit Dichte f. Dann teilt Q_p anschaulich die Fläche zwischen der x-Achse und dem Graphen von f im Verhältnis p zu $1 - p$ auf:



p-Quantil als „Flächen-Teiler“

Wiederholung vom 10. Juli:

$X \sim \mathsf{U}(a,b)$, falls $f(x) = \dfrac{1}{b-a}$ für $a < x < b$.

$X \sim \mathsf{Exp}(\lambda)$, falls $f(x) = \lambda \exp(-\lambda x)$ für $x > 0$,

$X \sim \mathsf{N}(\mu, \sigma^2)$, falls $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$.

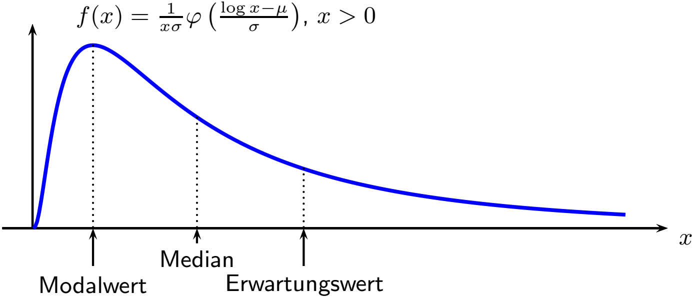
$$X \sim \mathsf{N}(\mu,\sigma^2) \Longleftrightarrow \frac{X-\mu}{\sigma} \sim \mathsf{N}(0,1). \quad \mathbb{P}(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Falls X Dichte f besitzt, so

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathsf{d}x, \quad \text{falls} \int_{-\infty}^{\infty} |x| f(x)\,\mathsf{d}x < \infty.$$

$$Q_p \;:= F^{-1}(p) \;:=\; \inf\{x \in \mathbb{R} : F(x) \ge p\} \quad (p\text{-Quantil})$$

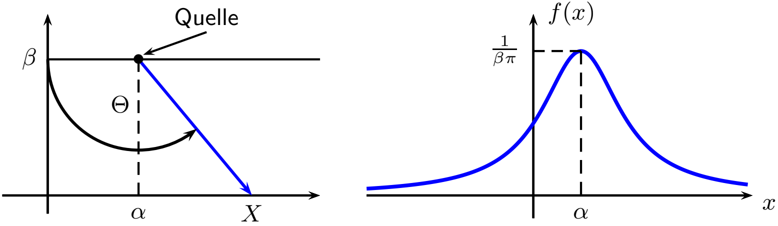
Memo: $Y \sim \mathrm{N}(\mu, \sigma^2) \implies X := \mathbf{e}^Y \sim \mathrm{LN}(\mu, \sigma^2)$



Die Lognormalverteilung $\mathrm{LN}(\mu, \sigma^2)$ hat eine rechtsschiefe Dichte.

$$\begin{aligned} \mathrm{Mod}(X) &= \exp(\mu - \sigma^2) && \text{(Modalwert = argmax } f) \\ \mathrm{Q}_{1/2}(X) &= \exp(\mu) && \text{(Median)} \\ \mathbb{E}(X) &= \exp(\mu + \sigma^2/2) && \text{(Übungsaufgabe!)} \end{aligned}$$

26.8 Beispiel (Cauchy-Verteilung $C(\alpha, \beta)$)



- Im Punkt (α, β) ist eine Quelle angebracht.
- Quelle sendet unter zufälligem Winkel $\Theta \sim U(0, \pi)$ Partikel in Richtung der x-Achse.
- Der zufällige Auftrittspunkt X hat die Dichte

$$f(x) \;=\; \frac{\beta}{\pi(\beta^2 + (x-\alpha)^2)}, \qquad x \in \mathbb{R}.$$

- Der Erwartungswert von X existiert nicht.
- $Q_{1/2}(X) = \alpha$, $Q_{3/4}(X) - Q_{1/4}(X) = 2\beta$ (Übungsaufgabe!)

26.9 Definition (Symmetrische Verteilung)

Die Zufallsvariable X heißt symmetrisch verteilt um den Wert a, falls gilt:

$$X - a \;\sim\; -(X - a)$$

(d.h.: $X - a$ ist symmetrisch um 0 verteilt).

Beispiele:

- $X \;\sim\;$ Bin$(n, 1/2) \quad \Longrightarrow \quad a = \dfrac{n}{2}$
- $X \;\sim\;$ U$(c, d) \quad \Longrightarrow \quad a = \dfrac{c+d}{2}$
- $X \;\sim\;$ N$(\mu, \sigma^2) \quad \Longrightarrow \quad a = \mu$
- $X \;\sim\;$ C$(\alpha, \beta) \quad \Longrightarrow \quad a = \alpha$

Hinreichende Bedingung für Verteilungssymmetrie um a bei stetigem X mit Dichte f:

$$f(a + x) \;=\; f(a - x) \quad \text{für jedes } x \in \mathbb{R}.$$

26.10 Satz (Erwartungswert = Median bei symmetrischen Verteilungen)

Die Zufallsvariable X sei symmetrisch verteilt um den Wert a. Dann gilt:

a) Existiert der Erwartungswert von X, so folgt $\mathbb{E}\,X = a$.

b) Ist X stetig mit Verteilungsfunktion F, so folgt $F(a) = \frac{1}{2}$.

c) Gilt in b) zusätzlich $|\{x \in \mathbb{R} : F(x) = 1/2\}| = 1$, so folgt $a = Q_{1/2}(X)$.

BEWEIS: a) Wegen $X - a \;\sim\; a - X \;= -(X - a)$ gilt

$$\mathbb{E}\,X - a \;=\; \mathbb{E}(X - a) \;=\; \mathbb{E}(a - X) \;=\; a - \mathbb{E}\,X. \;\surd$$

b) Wegen $X - a \;\sim\; a - X$ gilt

$$\mathbb{P}(X - a \le 0) \;=\; \mathbb{P}(a - X \le 0) \;\Longrightarrow$$

$$\begin{aligned} F(a) \;&=\; \mathbb{P}(X \le a) \;=\; \mathbb{P}(X - a \le 0) \\ &=\; \mathbb{P}(a - X \le 0) \;=\; \mathbb{P}(X \ge a) \;=\; 1 - F(a). \;\surd \end{aligned}$$

c) folgt aus b). Hinreichende Bedingung: $F'(a) > 0$.

26.11 Definition (Quantiltransformation)

Es sei X eine Zufallsvariable mit der Verteilungsfunktion F. Die durch

$$F^{-1}(p) \ := \ \inf\{x : F(x) \geq p\}$$

definierte Transformation $F^{-1} : (0, 1) \to \mathbb{R}$ heißt Quantiltransformation (zu F).

26.12 Beispiel (Exponentialverteilung)

Sei $X \sim \text{Exp}(\lambda)$, also

$$F(x) \ = \ \begin{cases} 1 - \exp(-\lambda x), & \text{falls } x \geq 0, \\ \quad 0, & \text{sonst.} \end{cases}$$

F auf $\{x : 0 < F(x) < 1\}$ stetig und streng monoton wachsend $\Longrightarrow$

$$p \ \overset{!}{=} F(F^{-1}(p)) \ = \ 1 - \exp(-\lambda F^{-1}(p))$$

$\Longrightarrow$

$$F^{-1}(p) \ = \ -\frac{1}{\lambda} \log(1 - p).$$

Memo: $F^{-1}(p) \;=\; \inf\{x : F(x) \geq p\}$

26.13 Satz (Quantiltransformation)

Es sei F eine Verteilungsfunktion. Dann gilt: Besitzt die Zufallsvariable U die Gleichverteilung $\mathrm{U}(0,1)$, so besitzt die Zufallsvariable

$$X \;:=\; F^{-1}(U)$$

die Verteilungsfunktion F.

Beweis: Es gilt $F(x) \geq p \Longleftrightarrow x \geq F^{-1}(p) \;\; (x \in \mathbb{R}, 0 < p < 1) \Longrightarrow$

$\mathbb{P}(X \leq x) \;=\; \mathbb{P}(F^{-1}(U) \leq x) \;=\; \mathbb{P}(U \leq F(x)) \;=\; F(x)$, $x \in \mathbb{R}$ $\surd$

Beachte:

- Die Quantiltransformation dient zur Erzeugung einer Pseudozufallszahl x mit Verteilungsfunktion F aus einer Pseudozufallszahl u mit Gleichverteilung $\mathrm{U}(0,1)$.
- Beispiel: $x = -\dfrac{1}{\lambda}\log(1-u)$ erzeugt $\mathrm{Exp}(\lambda)$-Zufallszahl.

27 Mehrdimensionale stetige Verteilungen

Memo: $(\Omega, \mathcal{A}, \mathbb{P})$ W-Raum, $X : \Omega \to \mathbb{R}^k$ Zufallsvektor

Memo: $\mathbb{P}^X : \mathcal{B}^k \to [0,1]$, $\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$

- Gilt $X =: (X_1, \ldots, X_k)$, so nennt man $\mathbb{P}^X$ auch gemeinsame Verteilung von $X_1, \ldots, X_k$.
- In diesem Fall heißt die Verteilung von X_j auch die j-te Marginalverteilung von X.
- Fast unverändert heißen $X_1, \ldots, X_k$ (stochastisch) unabhängig, falls gilt:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) \;=\; \prod_{j=1}^{k} \mathbb{P}(X_j \in B_j) \;\; \forall B_1, \ldots, B_k \in \mathcal{B}^1.$$

Wichtig für das Verständnis mancher Sachverhalte im Zusammenhang mit Wahrscheinlichkeitsverteilungen auf $\mathcal{B}^k$ ist folgendes Resultat.

27.1 Satz (Eindeutigkeitssatz)

Es sei $\mathcal{M} \subseteq \mathcal{B}^k$ ein durchschnittsstabiler Erzeuger von $\mathcal{B}^k$.

Es gelte also $A, B \in \mathcal{M} \Longrightarrow A \cap B \in \mathcal{M}$ sowie $\sigma(\mathcal{M}) = \mathcal{B}^k$.

Sind P und Q W-Maße auf $\mathcal{B}^k$ mit $P(B) = Q(B) \; \forall \, B \in \mathcal{M}$, so folgt $P = Q$.

BEWEIS: $\longrightarrow$ Wahrscheinlichkeitstheorie, Maßtheorie.

27.2 Folgerungen Sei $X = (X_1, \ldots, X_k)$ ein Zufallsvektor. Dann gilt:

Die Verteilung $\mathbb{P}^X$ von X (als W-Maß auf $\mathcal{B}^k$) ist festgelegt durch

- $\mathbb{P}(a \leq X \leq b), \;\; a, b \in \mathbb{R}^k, \;\; a \leq b$

  (setze $\mathcal{M} := \{[a, b] : a, b \in \mathbb{R}^k, \;\; a \leq b\} \cup \emptyset$) oder durch
- $F(x) := \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \ldots, X_k \leq x_k)$, $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$

  (setze $\mathcal{M} := \{(-\infty, x] : x \in \mathbb{R}^k\}$).

Die Funktion $\mathbb{R}^k \ni x \mapsto F(x) := \mathbb{P}(X \leq x)$ heißt Verteilungsfunktion von X.

27.3 Definition (Stetiger Zufallsvektor)

Der Zufallsvektor $X = (X_1, \ldots, X_k)$ heißt (absolut) stetig (verteilt), falls gilt: Es gibt eine Lebesgue-integrierbare Funktion $f : \mathbb{R}^k \to \mathbb{R}_{\geq 0}$ mit

$$\int_{\mathbb{R}^k} f(x)\,\mathrm{d}x \;=\; 1,$$

$$\mathbb{P}^X(B) \;=\; \mathbb{P}(X \in B) \;=\; \int_B f(x)\,\mathrm{d}x, \qquad B \in \mathcal{B}^k.$$

f heißt Dichte (der Verteilung) von X oder gemeinsame Dichte von $X_1, \ldots, X_k$.

27.4 Beispiel (Produkt-Ansatz)

Seien $f_1, \ldots, f_k : \mathbb{R} \to \mathbb{R}_{\geq 0}$ Dichten auf $\mathbb{R}^1$. Dann definiert
$f(x) := \prod_{j=1}^k f_j(x_j)$, $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$,
die Dichte eines k-dimensionalen Zufallsvektors $X = (X_1, \ldots, X_k)$.

Beweis: Nur $f \geq 0$ (klar) und $\int_{\mathbb{R}^k} f(x)\mathrm{d}x = 1$ zu zeigen (kanon. Konstruktion)

$$\int_{\mathbb{R}^k} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(x_1)\ldots f_k(x_k)\,\mathrm{d}x_1 \ldots \mathrm{d}x_k = \prod_{j=1}^k \int_{-\infty}^{\infty} f_j(x_j)\,\mathrm{d}x_j = 1$$

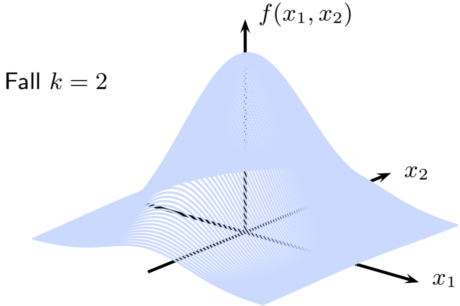
27.5 Beispiel (Standard-Normalverteilung im $\mathbb{R}^k$)

Ein Zufallsvektor X heißt standard-normalverteilt im $\mathbb{R}^k$, falls X die Dichte

$$f(x) \ := \ \prod_{j=1}^{k} \varphi(x_j) \ = \ \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{k} x_j^2 \right), \qquad x \in \mathbb{R}^k,$$

besitzt (spezieller Produkt-Ansatz mit $f_j(t) = \varphi(t) \ = \ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$).

Die Dichte f ist konstant auf Kreisen um den Nullpunkt ($k = 2$).



27.6 Satz (Gewinnung der marginalen Dichten)

Sei $X = (X_1, \ldots, X_k)$ ein Zufallsvektor mit Dichte f. Dann besitzt jede Komponente X_j eine mit f_j bezeichnete Dichte. Diese erhält man durch Integration aus f über die nicht interessierenden Variablen gemäß

$$f_j(t) = \int \int \ldots \int f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_k) \mathrm{d}x_1 \ldots \mathrm{d}x_{j-1} \mathrm{d}x_{j+1} \ldots \mathrm{d}x_k,$$

$t \in \mathbb{R}$. Dabei erstreckt sich jedes der $k-1$ Integrale über $\mathbb{R}$.

Vgl. Summation im diskreten Fall (Kapitel 11)

Der Einfachheit halber betrachten wir im Folgenden den Fall $k = 2$.

Um Indizes zu vermeiden, setzen wir

- (X, Y) für den zweidimensionalen Zufallsvektor,
- $h(x, y)$ für die gemeinsame Dichte von X und Y,
- $f(x)$ für die marginale Dichte von X,
- $g(y)$ für die marginale Dichte von Y.

Behauptung: Ist $h(x,y)$ die gemeinsame Dichte von (X,Y), so ist

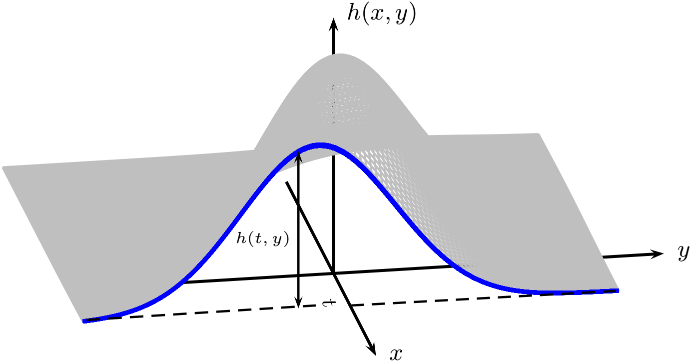
$$f(x) \ := \ \int_{-\infty}^{\infty} h(x,y)\,\mathrm{d}y, \qquad x \in \mathbb{R},$$

eine marginale Dichte von X.

Beweis: Zu zeigen: Für jede Borelmenge $B \in \mathcal{B}^1$ gilt $\mathbb{P}^X(B) = \int_B f(x)\,\mathrm{d}x$.

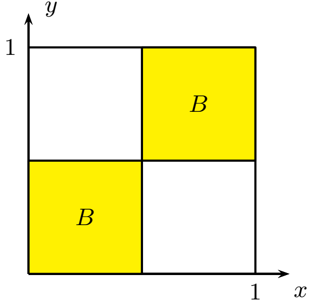
$$\begin{aligned}\mathbb{P}^X(B) &= \mathbb{P}(X \in B) \ = \ \mathbb{P}(X \in B, Y \in \mathbb{R}) \\ &= \mathbb{P}^{(X,Y)}(B \times \mathbb{R}) \ = \ \int_{B\times\mathbb{R}} h(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_B \left(\int_{-\infty}^{\infty} h(x,y)\mathrm{d}y \right) \mathrm{d}x \\ &= \int_B f(x)\,\mathrm{d}x. \ \surd\end{aligned}$$

Analog: $g(y) = \int_{-\infty}^{\infty} h(x,y)\,\mathrm{d}x, \quad y \in \mathbb{R},$ ist marginale Dichte von Y.



Bildung der marginalen Dichte $f(t) = \int h(t,y)\,\mathrm{d}y$ von X

27.7 Beispiel Sei $(X, Y) \sim \mathsf{U}(B)$, wobei $B := [0, 1/2]^2 \cup [1/2, 1]^2$.



$h(x, y) = 2 \cdot \mathbf{1}_B(x, y)$

$f(x) = 0$, falls $x < 0$ oder $x > 1$

$f(x) = \int_0^1 h(x, y) \mathrm{d}y = 1$, falls $0 \leq x \leq 1$

Ebenso: $g(y) = \mathbf{1}_{[0,1]}(y)$

Also: $X \sim \mathsf{U}(0, 1)$, $Y \sim \mathsf{U}(0, 1)$.

X und Y sind nicht stochastisch unabhängig, da

$$\frac{1}{2} = \mathbb{P}(X \leq 0.5, Y \leq 0.5) \neq \frac{1}{4} = \mathbb{P}(X \leq 0.5) \cdot \mathbb{P}(Y \leq 0.5).$$

Memo: X, Y unabhängig $:\Longleftrightarrow \mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B)\,\mathbb{P}(Y \in C)\,\forall B, C \in \mathcal{B}^1$

27.8 Satz (Unabhängigkeit und Dichten)

a) Seien X und Y unabhängige Zufallsvariablen mit Dichten f bzw. g. Dann hat der Zufallsvektor (X, Y) die „Produkt-Dichte"

$$h(x,y) \;=\; f(x) \cdot g(y), \qquad x,\ y \in \mathbb{R}. \tag{27.15}$$

b) Besitzt umgekehrt (X, Y) eine Dichte h der Gestalt (27.15) mit Dichten f und g, so sind X und Y unabhängig mit Dichten f bzw. g.

BEWEIS: a) Seien $[a, b]$, $[c, d]$ beliebige Intervalle, $B := [a, b] \times [c, d]$. Es gilt

$$\begin{aligned}
\mathbb{P}^{(X,Y)}(B) &= \mathbb{P}(X \in [a,b], Y \in [c,d]) \;=\; \mathbb{P}(X \in [a,b])\,\mathbb{P}(Y \in [c,d]) \\
&= \int_a^b f(x)\,\mathrm{d}x \int_c^d g(y)\,\mathrm{d}y \;=\; \int_a^b \int_c^d f(x)\,g(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_a^b \int_c^d h(x,y)\,\mathrm{d}x\,\mathrm{d}y \;=\; \int_B h(x,y)\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

Eindeutigkeitssatz 27.1 $\Longrightarrow \mathbb{P}^{(X,Y)}(B) = \int_B h(x,y)\,\mathrm{d}x\,\mathrm{d}y \;\forall\, B \in \mathcal{B}^1$. $\quad\surd$

b): Seien $[a, b]$, $[c, d]$ beliebige intervalle. Es gilt

$$\begin{aligned}\mathbb{P}(X \in [a,b], Y \in [c,d]) &= \int\int_{[a,b]\times[c,d]} f(x)\cdot g(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_a^b f(x)\,\mathrm{d}x \cdot \int_c^d g(y)\,\mathrm{d}y. \qquad (27.16)\end{aligned}$$

$c := -n$, $d := n$, $n \to \infty \Longrightarrow$

$$\mathbb{P}(X \in [a,b]) = \int_a^b f(x)\,\mathrm{d}x \cdot \int_{-\infty}^{\infty} g(y)\,\mathrm{d}y = \int_a^b f(x)\,\mathrm{d}x.$$

$a \to -\infty \Longrightarrow X$ hat Dichte f.

Ebenso: Y hat Dichte g.

Gleichung (27.16) besagt:

$$\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B)\cdot\mathbb{P}(Y \in C)\ \forall B, C \in \mathcal{M},$$

wobei $\mathcal{M} := \{[a,b] : a, b \in \mathbb{R},\ a \le b\} \cup \emptyset$.

Maßtheorie $\Longrightarrow \mathcal{M}$ kann durch $\sigma(\mathcal{M}) = \mathcal{B}^1$ ersetzt werden. $\surd$

27.9 Satz (Faltungsformel für stetige Zufallsvariablen)

Seien X, Y unabhängig mit Dichten f_X bzw. f_Y. Dann hat $X+Y$ die Dichte

$$f_{X+Y}(t) \;=\int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\, \mathrm{d}s, \qquad t\in\mathbb{R}.$$

Beweis: Für $z\in\mathbb{R}$ sei $B_z := \{(x,y) : x+y\le z\}$. Mit $t := y+s$ folgt

$$\begin{aligned}\mathbb{P}(X+Y\le z) &= \mathbb{P}((X,Y)\in B_z) \;=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-s} f_Y(y)\,\mathrm{d}y\right) f_X(s)\,\mathrm{d}s\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f_Y(t-s)\,\mathrm{d}t\right) f_X(s)\,\mathrm{d}s\\ &= \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f_X(s)\cdot f_Y(t-s)\,\mathrm{d}s\right)\mathrm{d}t.\ \checkmark\end{aligned}$$

- 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 von mehr als zwei unabhängigen Zufallsvariablen ergibt sich induktiv (X,Y,Z unabh. $\Longrightarrow$ (Blockungslemma) $X+Y,Z$ unabh.)

Memo: $f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\, \mathrm{d}s$

27.10 Satz (Additionsgesetz für die Normalverteilung)

Es seien X und Y unabhängige Zufallsvariablen mit den Normalverteilungen $\mathrm{N}(\mu, \sigma^2)$ bzw. $\mathrm{N}(\nu, \tau^2)$, wobei $\mu, \nu \in \mathbb{R}$, $\sigma^2, \tau^2 > 0$. Dann folgt:

$$X + Y \ \sim\ \mathrm{N}(\mu + \nu, \sigma^2 + \tau^2).$$

Beweis: Beachte: $X \sim \mu + \sigma U$, $Y \sim \nu + \tau V$, wobei U, V unabhängig und je $\mathrm{N}(0,1)$-verteilt. $\Longrightarrow$

$$X + Y \sim \mu + \nu + \sigma U + \tau V \implies \text{o.B.d.A. } \mu = \nu = 0.$$

$$f_X(s) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma^2}\right), \quad f_Y(t-s) = \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{(t-s)^2}{2\tau^2}\right) \implies$$

$$f_{X+Y}(t) = \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left\{\frac{s^2}{\sigma^2} + \frac{(t-s)^2}{\tau^2}\right\}\right) \mathrm{d}s.$$

Mit der Substitution

$$z = s \cdot \frac{\sqrt{\sigma^2+\tau^2}}{\sigma\tau} - \frac{t\sigma}{\tau\sqrt{\sigma^2+\tau^2}}$$

folgt

$$\begin{aligned} f_{X+Y}(t) &= \frac{1}{2\pi\sigma\tau} \cdot \frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}} \cdot \exp\left(-\frac{t^2}{2(\sigma^2+\tau^2)}\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) \mathsf{d}z \\ &= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \cdot \exp\left(-\frac{t^2}{2(\sigma^2+\tau^2)}\right) \end{aligned}$$

d.h. $X + Y \sim \mathsf{N}(0, \sigma^2 + \tau^2) \quad \surd$.

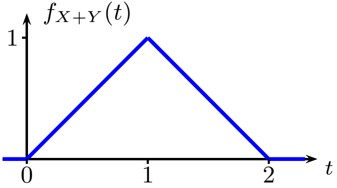
27.11 Beispiel (Faltung von Gleichverteilungen)

Seien X und Y unabhängig und je gleichverteilt in $(0,1)$,

$f_X(t) = f_Y(t) = \mathbf{1}_{[0,1]}(t)$, $t \in \mathbb{R}$. Faltungsformel $\Longrightarrow$

$$\begin{aligned} f_{X+Y}(t) &= \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(s) \cdot \mathbf{1}_{[0,1]}(t-s)\,\mathrm{d}s \\ &= \begin{cases} \int_0^t 1\,\mathrm{d}s = t, & \text{falls } 0 < t \leq 1, \\ \int_{t-1}^1 1\,\mathrm{d}s = 2-t, & \text{falls } 1 \leq t < 2 \end{cases} \end{aligned}$$

sowie $f_{X+Y}(t) = 0$, sonst (Integrationsgrenzen beachten!).



Die Gestalt von f_{X+Y} erklärt den Namen Faltungsformel.

27.12 Definition (Gammaverteilung)

Die Zufallsvariable X hat eine Gammaverteilung mit Parametern $\alpha > 0$ und $\lambda > 0$, falls X die Dichte

$$f(x) \ := \ \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, \mathrm{e}^{-\lambda x}, \quad \text{falls } x > 0,$$

und $f(x) := 0$, sonst, besitzt (kurz: $X \sim \Gamma(\alpha, \lambda)$).

Dabei ist $\Gamma : (0, \infty) \to \mathbb{R}$, $\Gamma(x) := \int_0^\infty \mathrm{e}^{-t} t^{x-1}\, \mathrm{d}t$, die Gamma-Funktion.

27.13 Satz (Momente der Gammaverteilung)

Es sei $X \sim \Gamma(\alpha, \lambda)$. Dann gelten (Übungsaufgabe):

a) $$\mathbb{E}(X^k) \ = \ \frac{\Gamma(k+\alpha)}{\lambda^k\, \Gamma(\alpha)} \ = \ \frac{1}{\lambda^k} \prod_{j=1}^{k} (\alpha + k - j), \qquad k \in \mathbb{N},$$

b) $$\mathbb{E}(X) \ = \ \frac{\alpha}{\lambda}, \qquad \mathbb{V}(X) \ = \ \frac{\alpha}{\lambda^2}.$$

27.14 Satz (Additionsgesetz für die Gammaverteilung)

Seien X, Y unabhängig, wobei $X \sim \Gamma(\alpha, \lambda)$, $Y \sim \Gamma(\beta, \lambda)$. Dann gilt

$$X + Y \ \sim\ \Gamma(\alpha + \beta, \lambda).$$

BEWEIS: Mit der Faltungsformel (Übungsaufgabe).

27.15 Bemerkung

Der Beweis dieses Gesetzes liefert als Nebenprodukt die wichtige Identität

$$\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u \ =\ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \alpha, \beta > 0. \tag{27.17}$$

27.16 Folgerung Wegen $\Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$ folgt:

Sind $X_1, \ldots, X_n$ unabhängige, je $\mathrm{Exp}(\lambda)$-verteilte Zufallsvariablen, so besitzt $X_1 + \ldots + X_n$ die Verteilung $\Gamma(n, \lambda)$, also die Dichte

$$f(x) = \frac{\lambda^n}{(n-1)!}\, x^{n-1}\, \mathrm{e}^{-\lambda x}, \quad x > 0 \qquad (f(x) = 0, \text{ sonst.})$$

27.17 Definition (Chi-Quadrat-Verteilung)

Es seien $Z_1, \ldots, Z_k$ unabhängige und je N$(0, 1)$-normalverteilte Zufallsvariablen.

Die Verteilung von

$$X \ := \ Z_1^2 + Z_2^2 + \ldots + Z_k^2$$

heißt Chi-Quadrat-Verteilung mit k Freiheitsgraden, kurz: $X \sim \chi_k^2$.

27.18 Folgerung Falls $X \sim \chi_k^2$, so gelten:

a) $\mathbb{E}(X) \ = \ k$,

b) $\mathbb{V}(X) \ = \ 2k$.

Beweis: Es gilt

$$\begin{aligned} \mathbb{E}(Z_1^2) &\ = \ \mathbb{V}(Z_1) \ = \ 1, \\ \mathbb{V}(Z_1^2) &\ = \ \mathbb{E}(Z_1^4) - \left(\mathbb{E}Z_1^2\right)^2 \ = \ 3 - 1 \ = \ 2 \end{aligned}$$

Additivität von $\mathbb{E}(\cdot)$ und auch $\mathbb{V}(\cdot)$ (da Unabhängigkeit!) $\Longrightarrow$ Beh.

27.19 Satz (Dichte der χ^2_k-Verteilung)

Eine Zufallsvariable X mit der χ^2_k-Verteilung besitzt die Dichte

$$f_k(t) \;=\; \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)}\,\exp\left(-\frac{t}{2}\right)\,t^{\frac{k}{2}-1}, \qquad t > 0,$$

und $f_k(t) := 0$, sonst.

Beachte: $\chi^2_k = \Gamma(k/2, 1/2)$.

Beweis: Es ist nur der Fall $k = 1$ zu zeigen. Der Rest ergibt sich mit dem Additionsgesetz für die Gammaverteilung.

Sei $F(t) := \mathbb{P}(Z_1^2 \le t)$. Für $t > 0$ ist

$$F(t) \;=\; \mathbb{P}(-\sqrt{t} \le Z_1 \le \sqrt{t}) \;=\; 2\Phi(\sqrt{t}) - 1 \;\Longrightarrow$$

$$F'(t) \;=\; 2\,\varphi(\sqrt{t})\,\frac{1}{2\sqrt{t}} \;=\; \frac{1}{\sqrt{2\pi}\sqrt{t}}\,\mathrm{e}^{-t/2} \;=\; \frac{1}{2^{1/2}\Gamma(1/2)}\,\mathrm{e}^{-t/2}\,t^{1/2-1} \;\checkmark$$

Wiederholung vom 14. Juli:

$X = (X_1, \ldots, X_k)$ stetig verteilt, falls

$$\mathbb{P}(X \in B) \;=\; \int_B f(x)\,\mathsf{d}x, \quad B \in \mathcal{B}^k,$$

für eine nichtnegative messbare Funktion f (sog. Dichte).

Die Dichte von X_j gewinnt man durch Marginalverteilungsbildung (Integration von f über die nicht interessierenden Variablen x_i mit $i \neq j$).

Zerfällt f in das Produkt der Dichten von $X_1, \ldots, X_k$, so sind $X_1, \ldots, X_k$ stochastisch unabhängig,

$$X, Y \text{ unabhängig } \Longrightarrow f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(s) f_Y(t-s)\,\mathsf{d}s$$

$$X, Y \text{ unabhängig }, X \sim \mathsf{N}(\mu, \sigma^2), Y \sim \mathsf{N}(\nu, \tau^2) \;\Longrightarrow\; X+Y \sim \mathsf{N}(\mu+\nu, \sigma^2+\tau^2)$$

$$Z_1, \ldots, Z_k \text{ unabhängig, je } \mathrm{N}(0,1)\text{-verteilt, so } Z_1^2 + \ldots + Z_k^2 \sim \chi_k^2$$

27.20 Satz (Allgemeine Darstellungsformel für Erwartungswerte)

Es seien Z ein k-dimensionaler Zufallsvektor mit Dichte f und $g : \mathbb{R}^k \to \mathbb{R}$ eine messbare Funktion. Dann existiert der Erwartungswert von $g(Z)$ genau dann, wenn gilt:

$$\int_{\mathbb{R}^k} |g(z)| \, f(z) \, \mathsf{d}z < \infty.$$

In diesem Fall folgt

$$\mathbb{E}\, g(Z) \; = \; \int_{\mathbb{R}^k} g(z) \, f(z) \, \mathsf{d}z.$$

BEWEIS: $\longrightarrow$ Maßtheorie, Wahrscheinlichkeitstheorie.

Beachte: Resultat ist „stetiges Analogon" von Satz 11.7. Dort: Diskreter Zufallsvektor Z, und

$$\mathbb{E}\, g(Z) \; = \sum_{z \in \mathbb{R}^k : \mathbb{P}(Z=z)>0} g(z) \, \mathbb{P}(Z = z).$$

Seien $\mathbb{E}\,X^2 < \infty$ und $\mathbb{E}\,Y^2 < \infty$. Wie früher heißen

$$C(X,Y) \;:=\; \mathbb{E}(X-\mathbb{E}\,X)(Y-\mathbb{E}\,Y) \;= \mathbb{E}(XY) - \mathbb{E}\,X\,\mathbb{E}\,Y$$

die Kovarianz zwischen X und Y und (bei $\mathbb{V}(X)\mathbb{V}(Y) > 0$)

$$r(X,Y) \;:=\; \frac{C(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}$$

der (Pearson-) Korrelationskoeffizient zwischen X und Y.

Besitzt (X,Y) eine Dichte h, so gilt

$$\mathbb{E}(XY) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,h(x,y)\,\mathsf{d}x\,\mathsf{d}y \quad \text{(allgemeine Darstellungsformel)}$$

Allgemein gilt (wie früher) die Multiplikationsformel für Erwartungswerte:

$$X, Y \text{ unabhängig} \;\Longrightarrow\; \mathbb{E}(XY) \;=\; \mathbb{E}\,X\,\mathbb{E}\,Y.$$

Beweis im Fall von stetigen Zufallsvariablen:

$$\mathbb{E}(XY) = \int\!\!\int xyf(x)g(y)\,\mathsf{d}x\mathsf{d}y = \left(\int xf(x)\,\mathsf{d}x\right)\left(\int yg(y)\,\mathsf{d}y\right) = \mathbb{E}X\,\mathbb{E}Y$$

27.21 Beispiel (Fortsetzung von Beispiel 27.7)

Sei $(X,Y) \sim \mathsf{U}(B)$, wobei $B = [0,1/2]^2 \cup [1/2,1]^2$.

Es ist $h = 2\,\mathbf{1}_B$ (Dichte), $X \sim Y \sim \mathsf{U}(0,1)$ (Marginalverteilungen).

$$\Longrightarrow \quad \mathbb{E}\,X \;=\; \mathbb{E}\,Y \;=\; \frac{1}{2}, \qquad \mathbb{V}(X) \;=\; \mathbb{V}(Y) \;=\; \frac{1}{12}.$$

Weiter ist

$$\begin{aligned}
\mathbb{E}(XY) \;&=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,h(x,y)\,\mathsf{d}x\,\mathsf{d}y \;=\; 2\int_0^1\int_0^1 xy\,\mathbf{1}_B(x,y)\,\mathsf{d}x\,\mathsf{d}y \\
&=\; 2\left(\int_0^{1/2}\int_0^{1/2} xy\,\mathsf{d}x\,\mathsf{d}y + \int_{1/2}^{1}\int_{1/2}^{1} xy\;\mathsf{d}x\,\mathsf{d}y\right) \\
&=\; 2\left(\frac{1}{64} + \frac{1}{4}\cdot\frac{9}{16}\right) \;=\; \frac{5}{16} \;\Longrightarrow
\end{aligned}$$

$$C(X,Y) \;=\; \mathbb{E}(XY) - \mathbb{E}\,X\cdot\mathbb{E}\,Y \;=\; \frac{5}{16} - \frac{1}{4} \;=\; \frac{1}{16}$$

$$r(X,Y) \;=\; \frac{C(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}} \;=\; \frac{1/16}{1/12} \;=\; \frac{3}{4}.$$

- Im Folgenden sind alle Vektoren als Spaltenvektoren zu vestehen.
- Für einen Spaltenvektor x sei $x^\top$ der zu x transponierte Zeilenvektor.
- Für jede auftretende Zufallsvariable Y gelte $\mathbb{E}\,Y^2 < \infty$.

27.22 Definition (Erwartungswertvektor, Kovarianzmatrix)

Es sei $X = (X_1, \ldots, X_k)^\top$ ein k-dimensionaler Zufallsvektor. Dann heißen

$$\mathbb{E}(X) \;:=\; (\mathbb{E}\,X_1, \ldots, \mathbb{E}\,X_k)^\top$$

der Erwartungswertvektor und

$$\Sigma(X) \;:=\; (C(X_i, X_j))_{1 \le i,j \le k}$$

die Kovarianzmatrix von X.

Memo: $\Sigma(X) \;=\; (C(X_i,X_j))_{1\le i,j\le k},\; C(X_i,X_j) = \mathbb{E}(X_i-\mathbb{E}\,X_i)(X_j-\mathbb{E}\,X_j)$

27.23 Bemerkung

Sei $Z = (Z_{i,j})_{1\le i\le m, 1\le j\le n}$ ein in Form einer $(m\times n)$-dimensionalen Matrix geschriebener Zufallsvektor. Mit

$$\mathbb{E}\,Z \;:=\; (\mathbb{E}\,Z_{i,j})_{1\le i\le m, 1\le j\le n}$$

gilt dann

$$\begin{aligned}\Sigma(X) &= \mathbb{E}\big[(X-\mathbb{E}\,X)(X-\mathbb{E}\,X)^\top\big]\\ &= \mathbb{E}\left[\begin{pmatrix} X_1-\mathbb{E}\,X_1\\ X_2-\mathbb{E}\,X_2\\ \vdots\\ X_k-\mathbb{E}\,X_k\end{pmatrix}\cdot\begin{pmatrix} X_1-\mathbb{E}\,X_1 & X_2-\mathbb{E}\,X_2 & \cdots & X_k-\mathbb{E}\,X_k\end{pmatrix}\right]\end{aligned}$$

27.24 (Rechenregeln (Übungsaufgabe))

Seien X ein k-dimensionaler Zufallsvektor, $b\in\mathbb{R}^n$ und $A\in\mathbb{R}^{n\times k}$. Dann gelten:

a) $\mathbb{E}(AX+b) \;=\; A\,\mathbb{E}\,X + b$,

b) $\Sigma(AX+b) \;=\; A\,\Sigma(X)\,A^\top$.

Memo: $\Sigma(X) \;=\; (C(X_i, X_j))_{1\le i,j\le k}$

27.25 Satz (Eine Kovarianzmatrix ist positiv-semidefinit)

Die Kovarianzmatrix $\Sigma(X)$ ist symmetrisch und positiv-semidefinit.

$$\Sigma(X) \text{ ist singulär } \iff \exists c \in \mathbb{R}^k,\; c \neq 0 \;\exists \gamma \in \mathbb{R} : \mathbb{P}(c^\top X = \gamma) = 1.$$

Also: $\Sigma(X)$ nicht invertierbar $\iff \mathbb{P}(X \in \mathcal{H}) = 1$ für eine $(k-1)$-dimensionale Hyperebene $\mathcal{H}$ des $\mathbb{R}^k$.

Beweis: Wegen $C(U,V) = C(V,U)$ ist $\Sigma(X)$ symmetrisch.

Sei $c = (c_1, \ldots, c_k)^\top \in \mathbb{R}^k$ beliebig. Es gilt

$$\sum_{i=1}^k \sum_{j=1}^k c_i c_j C(X_i, X_j) \;=\; C\left(\sum_{i=1}^k c_i X_i, \sum_{j=1}^k c_j X_j\right)$$

$$=\; \mathbb{V}\left(\sum_{j=1}^k c_j X_j\right) \;=\; \mathbb{V}(c^\top X) \;\geq\; 0. \;\checkmark$$

$$\Sigma(X) \text{ singulär } \iff \exists c \in \mathbb{R}^k,\; c \neq 0 \text{ mit } \mathbb{V}(c^\top X) = 0$$

$$\iff \exists c \in \mathbb{R}^k,\; c \neq 0 \;\exists \gamma \in \mathbb{R} : \mathbb{P}(c^\top X = \gamma) = 1.\checkmark$$

27.26 Beispiel (Multinomialverteilung)

Sei $X = (X_1, \ldots, X_s) \sim \mathsf{Mult}(n; p_1, \ldots, p_s)$.

Es gilt (vgl. Kapitel 13)

$$\mathbb{V}(X_j) = n\, p_j\, (1 - p_j), \quad C(X_i, X_j) = -n\, p_i\, p_j, \quad 1 \leq i \neq j \leq s,$$

also

$$\Sigma(X) \;=\; n\, (p_i \delta_{i,j} - p_i\, p_j)_{i \leq i,j \leq s}.$$

$\Sigma(X)$ ist singulär, da $\mathbb{P}(X_1 + \ldots + X_s = n) = 1$, also

$$\begin{aligned} 0 &= \mathbb{V}\left(\sum_{j=1}^{s} X_j \right) \\ &= C\left(\sum_{i=1}^{s} X_i, \sum_{j=1}^{s} X_j \right) \\ &= \sum_{i=1}^{s} \sum_{j=1}^{s} C(X_i, X_j) \\ &= (1 \cdots 1) \cdot \Sigma(X) \cdot (1 \cdots 1)^{\top} \end{aligned}$$

Seien X ein k-dimensionaler Zufallsvektor mit Dichte f und $T : \mathbb{R}^k \to \mathbb{R}^k$.

Unter Umständen besitzt der Zufallsvektor $Y := T(X)$ auch eine Dichte.

- Es gelte $\mathbb{P}(X \in O) = 1$ für eine offene Menge $O \subseteq \mathbb{R}^k$,
- Die Restriktion von T auf O sei stetig differenzierbar und injektiv,
- Für die Funktionaldeterminante $\det T'(x)$ auf O gelte $\det T'(x) \neq 0$, $x \in O$.

27.27 Satz (Transformationsssatz für Dichten)

Es bezeichne $T^{-1} : T(O) \to O$ die Umkehrabbildung von T auf $T(O)$.

Unter obigen Voraussetzungen besitzt $Y = T(X)$ die Dichte

$$g(y) \;=\; \frac{f(T^{-1}(y))}{|\det T'(T^{-1}(y))|}, \quad \text{falls } y \in T(O),$$

und $g(y) := 0$, sonst.

Beweis: Analysis 3 und Maßtheorie.

27.28 Beispiel (Erzeugung normalverteilter Pseudozufallszahlen)

Seien X_1, X_2 unabhängig und je $\mathrm{U}(0,1)$-verteilt.

$$\Longrightarrow \; f(x_1,x_2) \;=\; \begin{cases} 1, & \text{falls } 0 < x_1, x_2 < 1 \\ 0, & \text{sonst} \end{cases}$$

$\mathbb{P}((X_1,X_2) \in O) = 1$, wobei $O = (0,1)^2$.

Sei

$$T(x_1,x_2) \;:=\; \left(\sqrt{-2\log x_1}\,\cos(2\pi x_2), \sqrt{-2\log x_1}\,\sin(2\pi x_2)\right)$$

für $x_1, x_2 \in (0,1)$ und $T(x_1,x_2) := 0$, sonst.

T ist auf O stetig differenzierbar und injektiv.

Weiter gilt (nachrechnen!)

$$\det T'(x_1,x_2) \;=\; \frac{2\pi}{x_1}, \quad (x_1,x_2) \in O,$$

sowie $T(O) = \mathbb{R}^2 \setminus \{(y_1,y_2) \in \mathbb{R}^2 : y_1 \geq 0, y_2 = 0\}$.

Memo: $(y_1, y_2) := T(x_1, x_2) = \left(\sqrt{-2\log x_1}\,\cos(2\pi x_2), \sqrt{-2\log x_1}\,\sin(2\pi x_2)\right)$

$$y_1^2 + y_2^2 \;=\; -2\log x_1 \quad\Longrightarrow x_1 \;=\; \exp\Big[-\frac{1}{2}(y_1^2+y_2^2)\Big].$$

Memo: $\det T'(x_1, x_2) \;=\; \frac{2\pi}{x_1}, \quad (x_1, x_2) \in O.$

Memo: $g(y_1, y_2) \;=\; \dfrac{f(T^{-1}(y_1, y_2))}{|\det T'(T^{-1}(y_1, y_2))|}, \quad$ falls $(y_1, y_2) \in T(O)$.

Transformationssatz $\Longrightarrow$

$$g(y_1, y_2) \;=\; \left|\frac{2\pi}{\exp[-\frac{1}{2}(y_1^2+y_2^2)]}\right|^{-1} \;=\; \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y_1^2/2}\,\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y_2^2/2}$$

($(y_1, y_2) \in T(O)$, $g(y_1, y_2) = 0$ sonst) ist Dichte von $(Y_1, Y_2) := T(X_1, X_2)$.

Sei $N := \{(y_1, y_2) \in \mathbb{R}^2 : y_1 \geq 0, y_2 = 0\}$. Es gilt $\lambda^2(N) = 0$.

Folg: $g(y_1, y_2) = \varphi(y_1)\varphi(y_2)$, $(y_1, y_2) \in \mathbb{R}^2$, ist Dichte von (Y_1, Y_2).

Satz 27.8 b) $\Longrightarrow$ Y_1, Y_2 unabhängig und je $\mathrm{N}(0, 1)$-normalverteilt.

27.29 Beispiel (Affine Abbildung)

In der Situation des Transformationssatzes betrachten wir die Abbildung

$$T(x) \;:=\; Ax + b$$

mit einer invertierbaren $(k \times k)$-Matrix A und $b \in \mathbb{R}^k$.

Transformationssatz $\Longrightarrow Y := AX + b$ hat die Dichte

$$g(y) = \frac{f(A^{-1}(y-b))}{|\det A|} \cdot \mathbf{1}_{T(O)}(y).$$

27.30 Satz (Affine Transformation der Standardnormalverteilung im $\mathbb{R}^k$)

Seien $Y_1, \ldots, Y_k$ unabhängig und je $\mathrm{N}(0,1)$-verteilt sowie $Y := (Y_1, \ldots, Y_k)^\top$. Seien A eine reguläre $(k \times k)$-Matrix und $\mu \in \mathbb{R}^k$. Dann besitzt der Zufallsvektor

$$X := A\,Y + \mu$$

die Dichte

$$f(x) \;=\; \frac{1}{(2\pi)^{k/2}(\det\Sigma)^{1/2}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \qquad x \in \mathbb{R}^k,$$

wobei $\Sigma := A\,A^\top$.

Beweis: Y besitzt nach Beispiel 27.5 und Satz 27.8 b) die Dichte

$$f_Y(y) \;=\; \prod_{j=1}^{k} \varphi(y_j) \;=\; \frac{1}{(2\pi)^{k/2}} \cdot \exp\left(-\frac{1}{2} y^\top y\right), \qquad y = (y_1, \ldots, y_k)^\top \in \mathbb{R}^k.$$

X besitzt nach Beispiel 27.29 die Dichte

$$f(x) = \frac{f_Y\left(A^{-1}(x-\mu)\right)}{|\det A|}, \quad x \in \mathbb{R}^k. \quad \text{Direkte Rechnung} \Longrightarrow \text{Beh.}$$

27.31 Definition (Nichtausgeartete k-dimensionale Normalverteilung)

Es seien Σ eine symmetrische positiv definite $k \times k$-Matrix und $\mu \in \mathbb{R}^k$.

Der Zufallsvektor X besitzt eine nichtausgeartete k-dimensionale Normalverteilung mit Erwartungswertvektor μ und Kovarianzmatrix Σ, falls X die Dichte

$$f(x) \;=\; \frac{1}{(2\pi)^{k/2}(\det\Sigma)^{1/2}}\,\exp\left(-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)\right), \qquad x \in \mathbb{R}^k,$$

besitzt, kurz $X \sim \mathsf{N}_k(\mu, \Sigma)$.

27.32 Satz (Existenzsatz)

Zu jedem $\mu \in \mathbb{R}^k$ und jeder symmetrischen positiv definiten $k \times k$-Matrix Σ existiert ein Zufallsvektor X mit $X \sim \mathsf{N}_k(\mu, \Sigma)$.

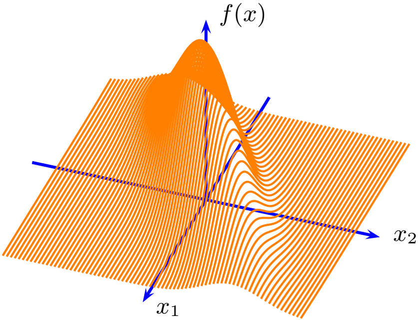
Beweis: Cholesky-Zerlegung $\Longrightarrow \exists\, A$ mit $\Sigma = A\,A^{\top}$.

Seien $Y_1, \ldots, Y_k$ unabhängig, je $\mathsf{N}(0,1)$-verteilt, $Y := (Y_1, \ldots, Y_k)^{\top}$.

Satz 27.30 $\Longrightarrow X := A\,Y + \mu \;\sim\; \mathsf{N}_k(\mu, \Sigma)$. $\surd$

Memo: $f(x) \;=\; \dfrac{1}{(2\pi)^{k/2}(\det\Sigma)^{1/2}}\,\exp\left(-\dfrac{1}{2}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right), \quad x\in\mathbb{R}^k$

Die Dichte f ist konstant auf Ellipsoiden mit Zentrum μ.



Dichte der zweidimensionalen Normalverteilung mit

$$\mu = \begin{pmatrix}0\\0\end{pmatrix} \quad \text{und} \quad \Sigma = \begin{pmatrix}2.25 & 1.2\\ 1.2 & 1\end{pmatrix}$$

27.33 Folgerung Falls $X \sim \mathrm{N}_k(\mu, \Sigma)$, so gilt:

a) $\mathbb{E}(X) \;=\; \mu$,

b) $\Sigma(X) \;=\; \Sigma$.

(Diese Beziehungen rechtfertigen die Sprechweise Normalverteilung mit Erwartungswertvektor μ und Kovarianzmatrix Σ).

Beweis: Wegen $X \sim AY + \mu$ mit Y wie im Beweis von Satz 27.32 und $\mathbb{E}(Y) = 0$ sowie $\Sigma(Y) = \mathrm{I}_k$ (k-reihige Einheitsmatrix) gilt

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(AY + \mu) \;=\; A\mathbb{E}(Y) + \mu \;=\; \mu, \\ \Sigma(X) &= \Sigma(AY + \mu) \;=\; A\,\Sigma(Y)\,A^\top \;=\; A\,A^\top \;=\; \Sigma. \;\checkmark \end{aligned}$$

27.34 Folgerung (Marginalverteilungen)

Falls $X = (X_1, \ldots, X_k) \sim \mathrm{N}_k(\mu, \Sigma)$, wobei $\mu = (\mu_1, \ldots, \mu_k)^\top$ und $\Sigma = (\sigma_{i,j})_{1 \le i,j \le k}$, so gilt

$$X_j \sim \mathrm{N}(\mu_j, \sigma_{j,j}), \quad 1 \le j \le k.$$

Beweis: Mit $A = (a_{i,j}), Y, \mu$ wie im Beweis von Satz 27.32 gilt $X \sim AY + \mu$ und somit

$$X_j \sim \sum_{l=1}^{k} a_{j,l} Y_l + \mu_j \sim \mathrm{N}\left(\mu_j, \sum_{l=1}^{k} a_{j,l}^2\right) \qquad \text{(Additionsgesetz 27.10)}$$

Wegen $\Sigma = A\,A^\top$ gilt

$$\sigma_{j,j} \;=\; \sum_{l=1}^{k} a_{j,l}^2 \;\checkmark$$

Alternativer Beweis: Direkte Rechnung (Marginalverteilungsbildung).

27.35 Satz (Unabhängigkeit und Unkorreliertheit)

Es sei $X = (X_1, \ldots, X_k) \sim \mathrm{N}_k(\mu, \Sigma)$. Dann gilt:

$$X_1, \ldots, X_k \text{ stochastisch unabhängig } \iff C(X_i, X_j) = 0 \ \forall \ i \neq j.$$

Beweis: „$\Longrightarrow$“ gilt wegen der Multiplikationsregel für Erwartungswerte.

„$\Longleftarrow$“:

Memo: $f(x) \ = \ \dfrac{1}{(2\pi)^{k/2}(\det \Sigma)^{1/2}} \exp\left(-\dfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)$

Voraussetzung $\Longrightarrow \Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_k^2)$, wobei $\sigma_j^2 = \mathbb{V}(X_j)$, $j = 1, \ldots, k$

$$\Longrightarrow \ \Sigma^{-1} = \mathrm{diag}(\sigma_1^{-2}, \ldots, \sigma_k^{-2}), \quad \det \Sigma = \prod_{j=1}^{k} \sigma_j^2.$$

$$\Longrightarrow \ f(x) \ = \ \prod_{j=1}^{k} \left[\frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right)\right], \quad x = (x_1, \ldots, x_k).$$

Satz 27.8 b) $\Longrightarrow$ Behauptung.

Lineare Algebra $\Longrightarrow$ Σ besitzt Orthonormalsystem von Eigenvektoren $v_1, \ldots, v_k$ und zugehörigen positiven Eigenwerten $\lambda_1, \ldots, \lambda_k$, also

$$\Sigma\, v_j \;=\; \lambda_j\, v_j, \quad \langle v_i, v_j \rangle \;=\; v_i^\top v_j \;=\; \delta_{i,j}, \quad 1 \le i, j \le k.$$

Sei

$$V \;:=\; (v_1 \; \cdots \; v_k), \qquad D \;:= \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

Dann gilt

$$V^\top \;=\; V^{-1}, \qquad \Sigma\, V \;=\; V\, D.$$

Sei

$$D^{1/2} := \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k}), \qquad A \;:=\; V\, D^{1/2}.$$

Dann folgt

$$A\, A^\top \;=\; V D^{1/2} D^{1/2} V^\top \;=\; V D V^{-1} \;=\; \Sigma V V^{-1} \;=\; \Sigma.$$

Satz 27.30 $\Longrightarrow$

$$\begin{aligned} X \;&\sim\; V D^{1/2} Y + \mu \\ &=\; \sum_{j=1}^{k} \sqrt{\lambda_j}\, Y_j\, v_j \;+\; \mu \end{aligned}$$

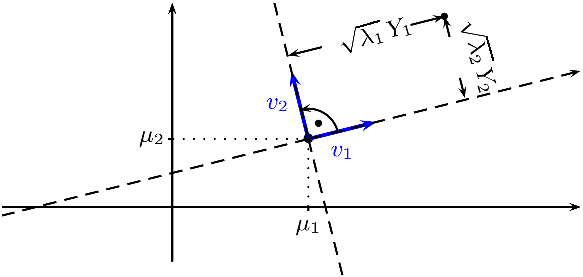
27.36 Satz (Hauptkomponentendarstellung von $\mathrm{N}(\mu, \Sigma)$)

Es gelte $X \sim \mathrm{N}_k(\mu, \Sigma)$. Die normierten Eigenvektoren von Σ und zugehörigen Eigenwerte seien $v_1, \ldots, v_k$ und $\lambda_1, \ldots, \lambda_k$.

Sind $Y_1, \ldots, Y_k$ stochastisch unabhängig und je $\mathrm{N}(0,1)$-verteilt, so gilt

$$X \sim \sqrt{\lambda_1}\, Y_1\, v_1 + \ldots + \sqrt{\lambda_k}\, Y_k\, v_k + \mu$$

(sog. Hauptkomponentendarstellung von X).



27.37 (Der Spezialfall $k = 2$)

Seien $\sigma, \tau > 0$, $|\rho| < 1$, Y_1, Y_2 unabhängig, je $\sim \mathrm{N}(0,1)$.

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} := \underbrace{\begin{pmatrix} \sigma & 0 \\ \sigma\tau & \tau\sqrt{1-\rho^2} \end{pmatrix}}_{=:A} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

Es gilt

$$A\,A^\top =: \Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix},$$

d.h. $\quad \rho\sigma\tau = C(X_1, X_2) \Longrightarrow \rho = r(X_1, X_2)$.

$$\begin{aligned} X_1 &= \sigma Y_1 + \mu_1 && \sim \mathrm{N}(\mu_1, \sigma^2), \\ X_2 &= \rho\tau Y_1 + \sqrt{1-\rho^2}\tau Y_2 + \mu_2 && \sim \mathrm{N}(\mu_2, \tau^2) \\ &= \underbrace{\rho\tau \frac{X_1 - \mu_1}{\sigma} + \mu_2}_{=:h(X_1)} + \sqrt{1-\rho^2}\,\tau\, Y_2. \end{aligned}$$

Die Gerade $x_1 \mapsto h(x_1)$ heißt Regressionsgerade von X_2 auf X_1.