Einführung in die Stochastik - Mitschrieb

Vorlesung im Wintersemester 2011/2012

Sarah Lutteropp

31. Januar 2012

Inhaltsverzeichnis

Vorwort

Dies ist ein Mitschrieb der Vorlesung “Einführung in die Stochastik” vom Wintersemester 2011/2012 am Karlsruher Institut für Technologie, die von

Herrn Prof. Dr. Günther Last gehalten wird.

Kapitel 1

Deskriptive Statistik

1.1 Der Grundraum

$\emptyset \neq \Omega =$ Grundraum (Grundgesamtheit, Merkmalsraum, Stichprobenraum)
Annahme: Ω ist diskret(endlich oder abzählbar unendlich) (Häufig $\Omega \subseteq \mathbb{R}$)

1.2 Absolute und relative Häufigkeit

$x_1, \ldots, x_n \in \Omega$ ("Daten")
$h(\omega) = \operatorname{card}\left\{j \in \{1, \ldots, n\} \colon x_j = \omega\right\}, \omega \in \Omega$, absolute Häufigkeit von ω

Bemerkung $\sum\limits_{\omega \in \Omega} h(\omega) = n$

Definition $\frac{1}{n} h(\omega) =$ relative Häufigkeit von ω
$h(A) = \operatorname{card}\left\{j \in \{1, \ldots, n\} \colon x_j \in A\right\}, A \subset \Omega =$ absolute Häufigkeit von A,
$\frac{1}{n} h(A) =$ relative Häufigkeit von A

1.3 Histogramm

$x_1, \ldots, x_n \in \mathbb{R}, b_1 < b_2 < \ldots < b_s$ mit $b_1 \leq \min\limits_{1 \leq i \leq n} x_i, b_s > \max\limits_{1 \leq i \leq n} x_i$
TODO: BILD
$d_j(b_{j+1} - b_j) = h([b_j, b_{j+1})) = \operatorname{card}\left\{i \in \{1, \ldots, n\} \colon b_j \leq x_i < b_{j+1}\right\}$

1.4 Lagemaße

Definition Ein **Lagemaß** ist eine Abbildung $l \colon \mathbb{R}^n \to \mathbb{R}$ mit

$$l(x_1 + a, \ldots, x_n + a) = l(x_1, \ldots, x_n) + a$$

"Verschiebungskovarianz". $x_1, \ldots, x_n, a \in \mathbb{R}$

1.4.1 Arithmetisches Mittel

$x_1, \ldots, x_n \in \mathbb{R}, \bar{x} := \frac{1}{n} \sum\limits_{j=1}^{n} x_j$ "Schwerpunkt der Daten"

Fakt $\sum\limits_{j=1}^{n} (x_i - t)^2 \xrightarrow{t} \text{Min}$
Lösung: $t = \bar{x}$
"Prinzip der kleinsten Quadrate"

Beweis $\frac{1}{n} \sum\limits_{j=1}^{n} (x_j - t)^2 = t^2 - 2\bar{x}t + \frac{1}{n} \sum\limits_{j=1}^{n} x_j^2 = (t - \bar{x})^2 + \frac{1}{n} \sum\limits_{j=1}^{n} x_j^2 - (\bar{x})^2$

1.4.2 Median, Quantile

$x_1, \ldots, x_n \in \mathbb{R} \Rightarrow x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ geordnete Stichprobe

Definition

$$x_{1/2} := \begin{cases} x_{(\frac{n+1}{2})} & \text{, falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{, falls } n \text{ gerade} \end{cases}$$

heißt **Median** von $x_1, \ldots, x_n$.

Fakt $\sum\limits_{j=1}^{n} |x_j - x_{1/2}| = \min\limits_{t} \sum\limits_{j=1}^{n} |x_j - t|$ Übungsaufgabe

Bemerkung Der Median ist "robust" gegenüber "Ausreißern". Ist etwa $x_1 = \ldots = x_9 = 1$ und $x_{10} = 1000 (n = 10)$, so gilt $\bar{x} = 100,9, x_{1/2} = 1$

Definition Für 0 < p < 1 heißt

$$x_p := \begin{cases} x_{(\lfloor n \cdot p + 1 \rfloor)} & \text{, falls } n \cdot p \notin \mathbb{N} \\ \frac{1}{2}(x_{(n \cdot p)} + x_{(n \cdot p + 1)}) & \text{, falls } n \cdot p \in \mathbb{N} \end{cases}$$

p-Quantil von $x_1, \ldots, x_n$.

Interpretation Mindestens $p \cdot 100\%$ der Daten liegen links von x_p und mindestens $(1 - p) \cdot 100\%$ liegen rechts von x_p.
$x_{1/4}$ = unteres Quartil, $x_{3/4}$ = oberes Quartil

1.5 Streuungsmaße

Definition Eine Abbildung $\sigma\colon \mathbb{R}^n \to \mathbb{R}$ mit

$$\sigma(x_1+a,\dots,x_n+a) = \sigma(x_1,\dots,x_n) \text{ (Translationsinvarianz)}$$

heißt **Streuungsmaß**.

1.5.1 Empirische Varianz

$s^2 := \frac{1}{n-1} \sum\limits_{j=1}^{n} (x_j - \bar{x})^2$ = **empirische Varianz** von $x_1,\dots,x_n$

1.5.2 Empirische Standardabweichung

$s := +\sqrt{s^2}$ = **empirische Standardabweichung** von $x_1,\dots,x_n$

1.5.3 Spannweite

$x_{(n)} - x_{(1)}$ = **Spannweite** von $x_1,\dots,x_n$

1.5.4 Quartilsabstand

$x_{(3/4)} - x_{(1/4)}$ = **Quartilsabstand** von $x_1,\dots,x_n$

1.6 Empirischer Korrelationskoeffizient

$(x_1,y_1),\dots,(x_n,y_n) \in \mathbb{R}^2$ TODO: BILD
Gesucht: Gerade $y = a + b\cdot x$ so, dass

$$(*) \sum_{j=1}^{n} (y_j - a - bx_j)^2 \overset{a,b}{\to} \text{Min}$$

Definition $\sigma_x^2 = \frac{1}{n}\sum\limits_{j=1}^{n}(x_j-\bar{x})^2$ $\sigma_y^2 = \frac{1}{n}\sum\limits_{j=1}^{n}(y_j-\bar{y})^2$
$\sigma_{xy} = \frac{1}{n}\sum\limits_{j=1}^{n}(x_j-\bar{x})(y_j-\bar{y})$ **empirische Kovarianz** $\sigma_x^2 > 0, \sigma_y^2 > 0$.
Lösung von (*): $b^* = \frac{\sigma_{xy}}{\sigma_{x^2}}, a^* = \bar{y} - b^*\cdot\bar{x}$
$\min\limits_{a,b}\sum\limits_{j=1}^{n}(y_j - a - bx_j)^2 \overset{!}{=} \min\limits_{b}\sum\limits_{j=1}^{n}(y_i - \bar{y} - b(x_j-\bar{x}))^2 = \dots$
"lineare Regression"

Einsetzen von a^* und b^* in die Zielfunktion:

$$0 \le \sum_{j=1}^{n}(y_j - a^* - b^*x_j)^2 = \dots = n\sigma_y^2(1 - (\frac{\sigma_{xy}}{\sigma_x\sigma_y})^2)$$

Definition $r_{xy} := \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ heißt **empirischer Korrelationskoeffizient** (*Pearson*).

Folgerung $|r_{xy}| \leq 1$
Es gilt $r_{xy} = \pm 1 \Leftrightarrow$ Punktewolke liegt exakt auf der Geraden $y = a^* + b^* x$.
Dabei ist $b^* > 0$, falls $r_{xy} = 1$ und $b^* < 0$, falls $r_{xy} = -1$.
Dieser empirische Korrelationskoeffizient ist ein Maß für die (affin) lineare Abhängigkeit zwischen den x_j und den y_j.

Kapitel 2

Ereignisse und Zufallsvariablen

2.1 Definition

Gegeben sei eine Grundmenge Ω. Die Elemente von Ω heißen **Elementarereignisse**. Teilmengen von Ω heißen **Ereignisse**. (Idee: $\omega \in \Omega$ ist Ausgang eines zufälligen Versuchs.)

Interpretation Ein Ereignis $A \subset \Omega$ "tritt ein", wenn $\omega \in A$.

2.2 Beispiele

- (i) (Münzwurf)
  $\Omega = \{0, 1\}$(oder $\Omega = \{W, Z\}$)
- (ii) (m Münzwürfe)
  $\Omega = \{0, 1\}^m (A = \{\omega = (\omega_1, \ldots, \omega_m) : \sum\limits_{j=1}^{m} \omega_j \geq k\}$ Ereignis)
- (iii) Werfen von 2 Würfeln
  $\Omega = \{1, \ldots, 6\}^2$
- (iv) Brownsche Bewegung
  (TODO: BILD) Bewegung eines Blütenpollens in einer Flüssigkeit
  $\Rightarrow$ Zukunftsmusik
  $\Omega = C([0, 1], \mathbb{R}^2)$

2.3 Bemerkung (Mengentheoretische Operationen)

Seien $A, B, A_1, A_2, \ldots \subset \Omega$.
$A \cap B = \{\omega \in \Omega : \omega \in A$ und $\omega \in B\} \mathrel{\hat{=}}$ "A und B treten ein"
$A \cup B \mathrel{\hat{=}}$ "A oder B treten ein"
$\bar{A} \equiv A^c := \Omega \backslash A = \{\omega \in \Omega : \omega \notin A\} \mathrel{\hat{=}}$ "A tritt nicht ein"

$A \backslash B = A \cap B^c \,\hat{=}\,$ "A tritt ein, aber nicht B"
$A \subset B \,\hat{=}\,$ "wenn A, dann B"
$\emptyset \,\hat{=}\,$ "unmögliches Ereignis"
$\Omega \,\hat{=}\,$ "sicheres Ereignis"

Abkürzung $AB = A \cap B$

2.4 Definition

Eine Abbildung $X\colon \Omega \to \mathbb{R}$ heißt (reelle) **Zufallsvariable**. Für $\omega \in \Omega$ heißt $X(\omega)$ **Realisierung** der Zufallsvariable zu ω.

Idee Mit $\omega \in \Omega$ bekommt auch $X(\omega)$ einen zufälligen Charakter.

Definition $X^{-1}\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\Omega) = \{A\colon A \in \Omega\}$ ist definiert durch

$$X^{-1}(A) = \{\omega \in \Omega\colon X(\omega) \in A\} \text{ ("Urbild von A unter X")}$$

Bemerkung

- $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B), A, B \subset \mathbb{R}$
- $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$
- $X^{-1}(\bigcup\limits_{j=1}^{\infty} A_j) = \bigcup\limits_{j=1}^{\infty} X^{-1}(A_j)$
- $X^{-1}(\bigcap\limits_{j=1}^{\infty} A_j) = \bigcap\limits_{j=1}^{\infty} X^{-1}(A_j)$

Vereinbarung Es sei X eine Zufallsvariable und $t \in \mathbb{R}$. Wir setzen

- $\{X = t\} := \{\omega\colon X(\omega) = t\} (= X^{-1}(t))$
- $\{X \geq t\} := \{\omega\colon X(\omega) \geq t\}$

2.5 Definition

Sind X, Y Zufallsvariablen, so definiert man

- $(X + Y)(\omega) = X(\omega) + Y(\omega)$
- $(X - Y)(\omega) = X(\omega) - Y(\omega)$
- $(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$

$\omega \in \Omega$, neue Zufallsvariablen $X+Y, X-Y, X \cdot Y$
analog für $a \in \mathbb{R}$

- $aX(\omega) = a \cdot (X(\omega))$
- $\min(X,Y) = (X \wedge Y)(\omega) := \min\{X(\omega), Y(\omega)\} \ldots$

2.6 Definition

Sei $A \subset \Omega$. Die Funktion $1_A \colon \Omega \to \mathbb{R}$ ist definiert durch

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

und heißt **Indikatorfunktion** von A.

2.7 Bemerkungen (Rechenregeln für Indikatorfunktionen)

- $1_\emptyset \equiv 0$
- $1_\Omega \equiv 1$
- $(1_A)^2 = 1_A$
- $1_{A^c} = 1 - 1_A$
- $1_{A \cap B} = 1_A \cdot 1_B$
- $1_{A \cup B} = 1_A + 1_B - 1_{A \cap B}$
- $A \subset B \Leftrightarrow 1_A \leq 1_B$
- $1_{A \triangle B} = |1_A - 1_B|$

2.8 Definition

Seien $A_1, \ldots, A_n \subset \Omega$. Die Zufallsvariable

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

heißt **Zählvariable** oder **Indikatorsumme**.

Bemerkung

- $\{X=0\} = \{\omega\colon X(\omega)=0\} = A_1^c \cap \dots A_n^c$
- $\{X=n\} = A_1 \cap \dots \cap A_n$
- $\{X = k\}$ = "genau k der Ereignisse $A_1, \dots, A_n$ treten ein" $= \bigcup\limits_{T\subset\{1,\dots,n\},|T|=k} \big(\bigcap\limits_{j\in T} A_j \cap \bigcap\limits_{j\notin T} A_j^c\big)$
  $(T \subset \{1,\dots,n\}, |T| = \text{card } T = k)$

Kapitel 3

Diskrete Wahrscheinlichkeitsräume

3.1 Motivation

Zufallsexperiment mit Ausgängen in Ω
n-malige, 'unabhängige' Wiederholung
$\Rightarrow$ Ergebnis $(a_1, \dots, a_n) \in \Omega^n$
$r_n(A) := \frac{1}{n} \sum\limits_{j=1}^{n} 1_A(a_j), A \subset \Omega$ relative Häufigkeit von A
$0 \leq r_n(A) \leq 1, r_n(\emptyset) = 0, r_n(\Omega) = 1$
$r_n(A \cup B) = r_n(A) + r_n(B), A \cap B = \emptyset$
empirisches Gesetz über Stabilisierung relativer Häufigkeiten:
$r_n(A) \underset{n\to\infty}{\rightsquigarrow} ?$

3.2 Definition

Ein Paar $(\Omega, \mathbb{P})$ bestehend aus einer diskreten Menge $\Omega \neq \emptyset$ und einer Funktion $\mathbb{P}\colon \mathcal{P} \to \mathbb{R}$ heißt **diskreter Wahrscheinlichkeitsraum**, falls:

- (P1) $\mathbb{P}(A) \geq 0, A \subset \Omega$
- (P2) $\mathbb{P}(\Omega) = 1$
- (P3) $\mathbb{P}(\bigcup\limits_{j=1}^{\infty} A_j) = \sum\limits_{j=1}^{\infty} \mathbb{P}(A_j), A_i \cap A_j = \emptyset, i \neq j$
  Diese Eigenschaft heißt σ-Additivität.

Man nennt $\mathbb{P}$ **Wahrscheinlichkeitsmaß (auf Ω)** (oder Wahrscheinlichkeitsverteilung) und $\mathbb{P}(A)$ heißt **Wahrscheinlichkeit von A**.

3.3 Folgerung

- a) $\mathbb{P}(\emptyset) = 0$
- b) $\mathbb{P}(\bigcup\limits_{j=1}^{n} A_j) = \sum\limits_{j=1}^{n} \mathbb{P}(A_j), A_i \cap A_j = \emptyset, i \neq j$
- c) $0 \leq \mathbb{P}(A) \leq 1, A \subset \Omega$
- d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), A, B \subset \Omega$
- e) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (Monotonie)
- f) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ (Komplementärwahrscheinlichkeit)
- g) $\mathbb{P}(\bigcup\limits_{j=1}^{\infty} A_j) \leq \sum\limits_{j=1}^{\infty} A_j$ (Subadditivität)
- h) $A_n \subset A_{n+1}, n \in \mathbb{N} \Rightarrow \mathbb{P}(\bigcup\limits_{n=1}^{\infty} A_n) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$ (Stetigkeit von unten)
- i) $A_n \supset A_{n+1}, n \in \mathbb{N} \Rightarrow \mathbb{P}(\bigcap\limits_{n=1}^{\infty} A_n) = \lim\limits_{n\to\infty} \mathbb{P}(A_n)$ (Stetigkeit von oben)

Beweis • a): $A_j = \emptyset, j \in \mathbb{N}$ (P3) [1] $\mathbb{P}(\emptyset) = 0$.
• b): $A_{n+1} = A_{n+2} = \ldots = \emptyset$ in P3!
• c) + f): Für $A \subset \Omega$ gilt nach b) (für n = 2):
$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \overset{(b)}{=} \mathbb{P}(A) + \mathbb{P}(A^c)$
• d): Nach b) gilt $\mathbb{P}(A) = \mathbb{P}(A \backslash B) + \mathbb{P}(A \cap B), \mathbb{P}(B) = \mathbb{P}(B \backslash A) + \mathbb{P}(A \cap B)$ und somit $\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A \backslash B) + \mathbb{P}(B \backslash A) + \mathbb{P}(A \cap B) \overset{(b)}{=} \mathbb{P}(A \cup B)$
• e): Wegen $B = A \cup (B \backslash A)$ folgt[2] $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geq \mathbb{P}(A)$
• g): $B_1 := A_1, B_2 := A_2 \backslash A_1, \ldots, B_n := A_n \backslash (\bigcup\limits_{j=1}^{n-1} A_j), n \geq 2$.
Dann gilt $B_n \subset A_n$ und $\bigcup\limits_{j=1}^{n} B_j = \bigcup\limits_{j=1}^{n} A_j$ sowie $B_i \cap B_j = \emptyset, i \neq j$.
Es folgt aus (P3):
$\mathbb{P}(\bigcup\limits_{j=1}^{\infty} A_j) \overset{!}{=} \mathbb{P}(\bigcup\limits_{j=1}^{n} B_j) \overset{(P3)}{=} \sum\limits_{j=1}^{n} \mathbb{P}(B_j) \overset{e)}{\leq} \sum\limits_{j=1}^{n} \mathbb{P}(A_j)$ (∞ ist zugelassen)
• h) + i): Übungsaufgabe

[1] $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) = 2 \cdot \mathbb{P}(\emptyset)$
[2] (aus der Additivität)

3.4 Satz

Seien $A_1, \dots, A_n \subset \Omega$. Setze

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

Dann gilt

- a) $\mathbb{P}(\bigcup\limits_{j=1}^{n} A_j) = \sum\limits_{k=1}^{n} (-1)^{k-1} S_k$ ‘Siebformel’
- b) $\mathbb{P}(\bigcup\limits_{j=1}^{n} A_j) \le \sum\limits_{k=1}^{2s+1} (-1)^{k-1} S_k, s = 0, \dots, \lfloor \frac{n-1}{2} \rfloor$
  $\mathbb{P}(\bigcup\limits_{j=1}^{n} A_j) \ge \sum\limits_{k=1}^{2s} (-1)^{k-1} S_k, s = 1, \dots, \lfloor \frac{n}{2} \rfloor$

Beweisidee für Siebformel vollständige Induktion nach n:
n=2: $\mathbb{P}(A_1 \cup A_2) \overset{(d)}{=} \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) = S_1 - S_2$
n=3: $\mathbb{P}(\underbrace{A_1 \cup A_2} \cup A_3) \overset{(d)}{=} \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}((A_1 \cup A_2) \cap A_3)$ [3]
$\overset{(d)}{=} \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3) = S_1 - S_2 + S_3$

3.5 Definition + Satz

a) Sei $(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum. Dann heißt $p \colon \Omega \to \mathbb{R}$ definiert durch $p(\omega) := \mathbb{P}(\{\omega\})$ **Wahrscheinlichkeitsfunktion** (von $\mathbb{P}$).
Es gilt $\mathbb{P}(A) = \sum\limits_{\omega \in A} p(\omega), A \subset \Omega$.
b) Sind Ω diskret und $p \colon \Omega \to \mathbb{R}$ eine Abbildung mit $p(\omega) \ge 0$ und $\sum\limits_{\omega \in \Omega} p(\omega) = 1$, so erhält man vermöge $\mathbb{P}(A) := \sum\limits_{\omega \in A} p(\omega)$ einen diskreten Wahrscheinlichkeitsraum.

Beweis • a) σ-Additivität $(A = \bigcup\limits_{\omega \in A} \{\omega\})$
• b) σ-Additivität: Großer Umordnungssatz! (Analysis)

3.6 Definition

$|\Omega| =: n < \infty$. Definiere $\mathbb{P}(A) = \frac{|A|}{n}$. Dann heißt $(\Omega, \mathbb{P})$ (ein diskreter Wahrscheinlichkeitsraum!) **Laplace-Raum**. Man nennt $\mathbb{P}$ **Gleichverteilung** auf Ω.
(‘homogene Münze’, ‘Würfeln’, …)

[3] $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$

3.7 Definition

Sei $\Omega \neq \emptyset$ beliebig! $(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum $\Leftrightarrow$ $\exists$ abzählbare Menge $\Omega_0 \subset \Omega$, $\exists p \colon \Omega \to [0, \infty)$ mit $p(\omega = 0)$ für alle $\omega \notin \Omega_0$, und $\sum\limits_{\omega \in \Omega_0} p(\omega) = 1$, und $\mathbb{P}(A) = \sum\limits_{\omega \in A \cap \Omega_0} p(\omega), A \subset \Omega$.

Wiederholung $(\Omega, \mathbb{P})$ Wahrscheinlichkeitsraum
$p \colon \Omega \to [0,1]$, $\sum\limits_{\omega \in \Omega} p(\omega) = 1$
$\mathbb{P}(A) := \sum\limits_{\omega \in A} p(\omega), A \subset \Omega$
$p(\omega) := \mathbb{P}(\{\omega\})$

Ω allgemeine Menge, $\Omega_0 \subset \Omega$ diskret
$p \colon \Omega \to [0,1]$, $\sum\limits_{\omega \in \Omega_0} p(\omega) = 1, p(\omega) = 0, \omega \notin \Omega_0$
$\mathbb{P}(A) := \sum\limits_{\omega \in A} p(\omega) := \sum\limits_{\omega \in A \cap \Omega_0} p(\omega)$
$\Omega_0 =$ Träger von $\mathbb{P}$

3.8 Definition

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum mit Träger Ω_0. Es sei $X \colon \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann heißt die Funktion $\mathbb{P}^X : \mathbb{P}(\mathbb{R}) \to \mathbb{R}$ definiert durch $\mathbb{P}^X(B) := \mathbb{P}\left(X^{-1}(B)\right), B \subset \mathbb{R}$ **Verteilung um** X.

3.9 Satz

In der Situation von Definition 3.8 ist $(\mathbb{R}, \mathbb{P}^X)$ ein diskreter Wahrscheinlichkeitsraum mit Träger $B_0 := X(\Omega_0) = \{X(\omega) \colon \omega \in \Omega_0\}$

Beweis. Für $B \subset \mathbb{R}$.

$$\begin{aligned} \mathbb{P}^X(B) &= \mathbb{P}(\{\omega \colon X(\omega) \in B\}) \\ &\overset{!}{=} \mathbb{P}(\{\omega \colon X(\omega) \in B \cap B_0\}) \end{aligned}$$

Definiert man für $t \in \mathbb{R}$

$$p_t = \mathbb{P}(\{\omega \colon X(\omega) = t\}) = \mathbb{P}(X = t)$$

so ergibt sich aus der σ-Additivität von $\mathbb{P}$

$$\mathbb{P}^X(B) = \sum_{t \in B \cap B_0} \mathbb{P}(\{\omega \colon X(\omega) = t\}) = \sum_{t \in B \cap B_0} p_t$$

$\square$

Kapitel 4

Kombinatorik

$|A| = card(A) =$ Anzahl der Elemente einer endlichen Menge A

4.1 Grundregeln

$A_1, \dots, A_k$ endliche Menge

(i) $A_i \cap A_j = \emptyset, i \neq j \Rightarrow \left| \bigcup\limits_{j=1}^{n} A_j \right| = \sum\limits_{j=1}^{n} A_j$

(ii) $|A_1 \times \ldots \times A_n| = \prod\limits_{j=1}^{k} |A_j|$

4.2 Satz

Es sollen k-Tupel $(a_1, \dots, a_k)$ durch sukzessives Festlegen von $a_1, a_2, \dots, a_k$ nach folgenden Regeln gebildet werden:

- es gibt j_1 Möglichkeiten für die Wahl von a_1
- es gibt (dann) j_2 Möglichkeiten für die Wahl von a_2
- $\dots$
- es gibt (dann) j_k Möglichkeiten für die Wahl von a_k

Dann gibt es genau $j_1 \cdot \ldots \cdot j_k$ solcher Tupel.

4.3 Beispiel (Urnenmodelle)

Betrachte Urne mit n durchnummerierten Kugeln.Es werden k Kugeln nach folgenden Regeln gezogen: $(M := \{1, \dots, n\})$

Zurücklegen (Wiederholung) \ Beachtung der Reihenfolge	ja	nein
ja	k-Permutationen aus M mit Wiederholung, Per_k^n	k-Kombinationen aus M mit Wiederholung, Kom_k^n
nein	k-Permutationen aus M ohne Wiederholung, $Per_{k,\neq}^n$	k-Kombinationen aus M ohne Wiederholung, $Kom_{k,\neq}^n$

4.4 Definition

$M = \{1, \ldots, n\}(n \in \mathbb{N})$

- $Per_k^n := M^k$
- $Per_{k,\neq}^n := \{(a_1, \ldots, a_k) \in M^k \colon a_i \neq a_j \text{ für } i \neq j\}$
- $Kom_k^n := \{(a_1, \ldots, a_k = \in M^k \colon a_1 \leq a_2 \leq \ldots \leq a_k\}$
- $(Kom_{k,\neq}^n := \{(a_1, \ldots, a_k) \in M^k \colon a_1 < a_2 < \ldots < a_k\}$

4.5 Satz

- (i) $|Per_k^n| = n^k$
- (ii) $|Per_{k,\neq}^n| := n^{\underline{k}} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$
- (iii) $|Kom_k^n| = \binom{n+k-1}{k}$
- (iv) $|Kom_{k,\neq}^n| = \binom{n}{k}$

Beweis. (i): 4.1.(ii)
(ii) Satz 4.2
(iv) Betrachte Äquivalenzrelation

$$(a_1, \ldots, a_k) \sim (b_1, \ldots, b_k) \Leftrightarrow \{a_1, \ldots, a_k\} = \{b_1, \ldots, b_k\}$$

auf $Per_{k,\neq}^n$. Jede Äquivalenzklasse hat $k!$ Elemente! Es folgt

$$|Kom_{k,\neq}^n| \cdot k! = |Per_{k,\neq}^n| = n^{\underline{k}}$$

(iii) Die Abbildung $g \colon Kom_k^n \to Kom_{k,\neq}^{n+k-1}$ definiert durch

$$(a_1, \ldots, a_k) \mapsto (a_1, a_2 + 1, a_3 + 2, \ldots, a_k + k - 1)$$

ist eine Bijektion! (Umkehrabbildung!) Es folgt(!)

$$|Kom_k^n| = |Kom_{k,\neq}^{n+k-1}| = \binom{n+k-1}{k}$$

□

4.6 Beispiel (Geburtstagsproblem)

Wie groß ist die Wahrscheinlichkeit, dass unter k rein zufällig ausgewählten Personen mindestens zwei am selben Tag Geburtstag haben?

Antwort Betrachte $\Sigma = Per_k^n$ mit $n = 365$, und der Laplace-Verteilung. Es sei $A := \{(a_1, \ldots, a_k) \in \Omega\colon$ es gibt $i, j \in \{1, \ldots, k\}$ mit $i \neq j, a_i = a_j\}$. Es gilt

$$\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}(A^c) \\
&= 1 - \mathbb{P}(Per_{k,\neq}^n) \\
&\stackrel{!}{=} 1 - \frac{|Per_{k,\neq}^n|}{card\Omega} \\
&= 1 - \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{n^k} \\
&= 1 - \frac{n}{n} \cdot \frac{(n-1)}{n} \cdot \ldots \cdot \frac{n-k+1}{n} \\
&= 1 - (1 - \frac{1}{n}) \cdot \ldots \cdot (1 - \frac{k-1}{n})
\end{aligned}$$

$\underline{k = 23:}$ $\mathbb{P}(A) \approx 0,507 > \frac{1}{2}$
$n = \binom{49}{6}, k = 4004, \mathbb{P}(A) = 0,5001 > \frac{1}{2}$

4.7 Beispiel

n Personen bringen (zu einer Feier) je ein Geschenk mit. Geschenke werden "rein zufällig" verteilt. Mit welcher Wahrscheinlichkeit bekommt mindestens eine Person ihr eigenes Geschenk?
$\rightsquigarrow$ Siebformel!

4.8 Beispiel (Besetzungsmodelle)

k Teilchen sollen auf n nummerierte Fächer verteilt werden. Analogie zu Urnenmodell: Nummer der Kugel $\hat{=}$ Nummer des Fachs, Nummer der Ziehung $\hat{=}$ Nummer des Teilchens

Unterscheidbare Teilchen \ Mehrfachbesetzungen	ja	nein
ja	Per_k^n Maxwell-Boltzmann	Kom_k^n Bose-Einstein-Statistik
nein	$Per_{k,\neq}^n$ Fermi-Dirak-Statistik	$Kom_{k,\neq}^n$

Statistische Physik

Kapitel 5

Der Erwartungswert

$p(\omega) = \mathbb{P}(\{\omega\})$, $(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum

5.1 Definition

- Der Erwartungswert einer Zufallsvariablen $X\colon \Omega \to \mathbb{R}$ existiert (genauer: X ist integrierbar bezüglich $\mathbb{P}$), falls

  $$\sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty \tag{5.1}$$

  In diesem Fall heißt

  $$\mathbb{E}X = \mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) p(\omega)$$

  (Physik: $< X >= \mathbb{E}[X]$) **Erwartungswert von** X.

- Ist $X \geq 0$ eine Zufallsvariable, so heißt

  $$\mathbb{E}X := \sum_{\omega \in \Omega} X(\omega) p(\omega) \in [0, \infty]$$

  ebenfalls Erwartungswert von X.

5.2 Satz

Sei $L^1 \equiv L^1(\mathbb{P}) := \{X\colon X \text{ erfüllt } 5.1\}$. Dann ist L^1 ein reeller Vektorraum. Genauer:

- (i) $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y, X, Y \in L^1$
- (ii) $\mathbb{E}[aX] = a\mathbb{E}X, X \in L^1, a \in \mathbb{R}$
- (iii) $\mathbb{E}1_A = \mathbb{P}(A), A \subset \Omega$

- (iv) $X \leq Y \Rightarrow \mathbb{E}X \leq \mathbb{E}Y$
- (v) $|\mathbb{E}X| \leq \mathbb{E}|X|$

Beweis. (i) $|(X+Y)(\omega)| \leq |X(\omega)| + |Y(\omega)|$.
Also $X+Y \in L^1(\mathbb{P})$ und

$$\sum_{\omega\in\Omega} (X(\omega)+Y(\omega))p(\omega) \stackrel{!}{=} \sum_{\omega\in\Omega} X(\omega)p(\omega) + \sum_{\omega\in\Omega} Y(\omega)p(\omega)$$

(ii) analog
(iii) $\mathbb{E}1_A = \sum\limits_{\omega\in\Omega} p(\omega) = \mathbb{P}(A)$
(iv) + (v) Übungsaufgabe □

5.3 Folgerung

Seien $A_1, \dots, A_n \subset \Omega$ und $X := \sum\limits_{j=1}^{n} 1_{A_j}$. Dann gilt $\mathbb{E}X = \sum\limits_{j=1}^{n} \mathbb{P}(A_j)$.
(Gilt auch für ∞ viele Ereignisse.)

5.4 Satz (Transformationsformel)

Seien $X\colon \Omega \Rightarrow \mathbb{R}$ und $g\colon \mathbb{R} \to \mathbb{R}$. Definiere $g(X)\colon \Omega \to \mathbb{R}$ durch

$$g(X)(\omega) = g(X(\omega)).$$

Dann ist $g(X) \in L^1(\mathbb{P})$ genau dann, wenn

$$\sum_{x\colon \mathbb{P}(X=x)>0} |g(x)|\mathbb{P}(X=x) < \infty$$

[1]

In diesem Fall gilt

$$\mathbb{E}g(x) = \sum_{x\colon \mathbb{P}(X=x)>0} g(x)\mathbb{P}(X=x)$$

Beweis. Es gilt

$$\sum_{\omega\in\Omega} |g(X(\omega))|p(\omega) = \sum_{x\in\mathbb{R}\colon \mathbb{P}(X=x)>0} |g(x)| \sum_{\omega\in\Omega\colon X(\omega)=x} p(\omega)$$

$$\Omega = \bigcup_{x\in\mathbb{R}\colon X(\omega)=x, \mathbb{P}(X=x)>0} \{\omega\colon X(\omega)=x\} \cup \Omega', \mathbb{P}(\Omega')=0$$

[1] $\mathbb{P}(X=x) = \mathbb{P}(\{\omega \in \Omega\colon X(\omega)=x\})$

$$= \sum |g(X(\omega))|p(\omega) = \sum_{\omega \notin \Omega'} |g(X(\omega))|p(\omega)$$

$$= \sum_{x \in \mathbb{R}:\, \mathbb{P}(X=x)>0} \sum_{\omega \in \{\omega \in \Omega:\, X(\omega)=x\}} |g(X(\omega))|p(\omega)$$

$$= \sum_{x\cdots} |g(x)|\mathbb{P}(X = x)$$

Ist das endlich, so gilt die Rechnung auch ohne Betragsstriche!
Insbesondere gilt

$$\mathbb{E}X = \sum_{x \in \mathbb{R}} x\mathbb{P}(X = x)\,(g(x) \equiv x)$$

$\square$

5.5 Beispiele

- Würfelwurf, X=Augenzahl, $\mathbb{P}(X = j) = \frac{1}{6}$.
  Also

$$\mathbb{E}X = \sum_{j=1}^{6} j \cdot \mathbb{P}(X = j) = \frac{6 \cdot 7}{2} \cdot \frac{1}{6} = \frac{7}{2} = 3,5$$

- Zweifacher Würfelwurf, X=Maximum der Augenzahlen ($\Omega = \{1, \dots, 6\}^2, \mathbb{P} =$ Gleichverteilung)

$$\mathbb{P}(X = 1) = \frac{1}{36}$$

$$\mathbb{P}(X = 2) = p((1, 2)) + p((2, 1)) + p((2, 2)) = \frac{3}{36}$$

  Allgemein: $\mathbb{P}(X = j) = \frac{2j-1}{36}, j = 1, \dots, 6$
  Es folgt

$$\mathbb{E}X = \sum_{j=1}^{6} j \cdot \frac{2j-1}{36} \overset{?}{\approx} 4,47$$

Kapitel 6

Die hypergeometrische Verteilung und die Binomialverteilung

Urne mit Kugeln $\underbrace{1, 2, \ldots, r}_{rot}, \underbrace{r+1, \ldots, r+s}_{schwarz}$
$r, s \in \mathbb{N}_0, r + s > 0$.

6.1 Definition

- n mal Ziehen ohne Zurücklegen
- $a_j :=$ Nummer der j-ten gezogenen Kugel
- $\Omega = Per^{r+s}_{n,\neq}$
- $\mathbb{P} =$ Gleichverteilung ("unabhängiges", "rein zufälliges" Ziehen)
- $A_j := \{(a_1, \ldots, a_n) \in \Omega \colon a_j \leq r\} \hat{=} \{$j-te gezogene Kugel ist rot$\}$
- $X := \sum\limits_{j=1}^{n} 1_{A_j} =$ Anzahl der gezogenen roten Kugeln

$\mathbb{P}^X$ (die Verteilung von X) heißt **hypergeometrische Verteilung** mit Parametern r, s, n, kurz:

$$X \sim Hyp(n, r, s), n \leq r + s$$

$$\mathbb{P}^X = Hyp(n, r, s)$$

6.2 Satz

Es gilt

- (i) $\mathbb{E}X = n \cdot \frac{r}{r+s}$
- (ii) $\mathbb{P}(X = k) = \frac{\binom{r}{s}\binom{s}{n-k}}{\binom{r+s}{n}}, k = 0, \ldots, r \wedge n$

Beweis. (i) Es gilt (Symmetrieargument!) $|A_j| = r \cdot (r+s-1)^{\underline{n-1}}$
$|\Omega| = (r+s)^{\underline{n}} \Rightarrow \mathbb{P}(A_j) = \frac{|A_j|}{|\Omega|} = \frac{r}{r+s}$
Aus 5.3 folgt $\mathbb{E}X = n \cdot \frac{r}{r+s}$
(ii) $|\{X = k\}| \stackrel{!}{=} \binom{n}{k} r^{\underline{k}} s^{\underline{n-k}}$
$\Rightarrow \mathbb{P}(X = k) = \frac{\binom{n}{k} r^{\underline{k}} s^{\underline{n-k}}}{(r+s)^{\underline{n}}} = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}$ □

6.3 Motivation

X Zufallsvariable, $\sum\limits_{k=1}^{r} \mathbb{P}(X = x_n) = 1$
$X_1, X_2, \ldots, X_n$ "unabhängige" Wiederholungen von X (= Ergebnis eines zufälligen Versuchs)
$\bar{X} := \frac{1}{n}(X_1 + \ldots + X_n)$ Zufallsvariable!
Mit $h_j := card\{i \in \{1, \ldots, n\} \colon X_i = x_j\}$ gilt $\bar{X} \stackrel{!}{=} \frac{1}{n}(h_1 x_1 + h_2 x_2 + \ldots + h_n x_n)$
empirisches Gesetz über Stabilität relativer Häufigkeiten
$\underset{n\to\infty}{\to} \mathbb{P}(X = x_1)x_1 + \ldots + \mathbb{P}(X = x_r)x_r \stackrel{!}{=} \mathbb{E}X$
$X \sim Hyp(n, r, s) = \mathbb{P}^X, n \leq r + s$
$\mathbb{P}(X = k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, k = 0, \ldots, n$
Wegen $\binom{m}{l} := 0$ für $m < l$ gilt: $\mathbb{P}(X = k) = 0$ für $k < r$ und für $n - k > s (k < n - s)$

6.4 Definition

Binomialverteilung:

- n maliges Ziehen aus einer Urne mit $r + s$ Kugeln mit Zurücklegen
- $\Omega = Per_n^{r+s} = \{(a_1, \ldots, a_n) \colon 1 \leq a_i \leq r + s, i = 1, \ldots, n\}$
- $\mathbb{P}$ Gleichverteilung

$X := \sum\limits_{j=1}^{n} 1_{A_j}, A_j := \{(a_1, \ldots, a_n) \in \Omega \colon a_j \leq r\}$
$\mathbb{P}^X$ heißt Binomialverteilung mit Parametern n und $p := \frac{r}{r+s}$. Man schreibt auch $Bin(n, p) := \mathbb{P}^X$.

6.5 Satz

Es gilt

1. $\mathbb{E}X = np$
2. $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, 0 \leq k \leq n$

Beweis.

1. $|A_j| = r \cdot (r+s)^{n-1}$
   $|\Omega| = (r+s)^n \rightsquigarrow \mathbb{P}(A_j) = \frac{|A_j|}{|\Omega|} = \frac{r}{r+s} = p$
   Folgerung 5.3 $\rightsquigarrow \mathbb{E}X = np$.
2. $card\{X = k\} = \binom{n}{k} r^k s^{n-k}$
   $\rightsquigarrow \mathbb{P}(X = k) = \frac{\binom{n}{k} r^k s^{n-k}}{(r+s)^k (r+s)^{n-k}}$

$\square$

Bemerkung $Bin(n,p)$ ist für jedes $p \in [0,1]$ definiert.

Kapitel 7

Mehrstufige Experimente

7.1 Beispiel

Urne mit einer roten und drei schwarzen Kugeln
<u>1. Experiment</u> Kugel ziehen, Farbe notieren, Kugel und eine weitere Kugel derselben Farbe zurücklegen
<u>2. Experiment</u> Erneut Kugel ziehen
<u>Modell:</u> $\Omega := \{0,1\} \times \{0,1\}, \quad (0 \hat{=} s, 1 \hat{=} r)$

Konstruktion von $\mathbb{P}$ $\quad p(\omega) := \mathbb{P}(\{\omega\})$

$$\left.\begin{aligned} p(1,1) &:= \tfrac{1}{4} \cdot \tfrac{2}{5} = \tfrac{2}{20} = \tfrac{1}{10} \\ p(1,0) &:= \tfrac{1}{4} \cdot \tfrac{3}{5} = \tfrac{3}{20} \\ p(0,1) &:= \tfrac{3}{4} \cdot \tfrac{1}{5} = \tfrac{3}{20} \\ p(0,0) &:= \tfrac{3}{4} \cdot \tfrac{4}{5} = \tfrac{12}{20} \end{aligned}\right\} \text{1. Pfadregel}$$

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

Betrachte $B := \{(1,1),(0,1)\}$. Dann gilt

$$\mathbb{P}(B) = p(1,1) + p(0,1) = \text{(2. Pfadregel)}$$

$$= \frac{2}{20} + \frac{3}{20} = \frac{1}{4} \overset{!}{=} \mathbb{P}(\text{erste Kugel ist rot})$$

(TODO: Bild(Baumdiagramm))

7.2 Definition

Mehrstufige Experimente $\quad \Omega = \Omega_1 \times \ldots \times \Omega_n$ ($\Omega_j \hat{=}$ Grundraum für j-tes Teilexperiment)
$\omega = (a_1, \ldots, a_n) \in \Omega$
Problem: Definiere $p(\omega) = \mathbb{P}(\{\omega\})$

1. Startverteilung $p_1\colon \Omega_1 \to [0,1] \quad \sum\limits_{\omega\in\Omega_1} p_1(\omega) = 1$

2. Übergangswahrscheinlichkeiten $p_2(a_2|a_1) \geq 0 \quad \sum\limits_{a_2\in\Omega_2} p_2(a_2|a_1) \overset{!}{=} 1$

   ($p_2(a_2|a_1)\hat{=}$ Wahrscheinlichkeit, dass 2. Versuch das Ergebnis a_2 liefert unter der Bedingung, dass 1. Versuch Ergebnis a_1 geliefert hat.)
   $p_3(a_3|a_1,a_2) \geq 0 \quad \sum\limits_{a_3\in\Omega_3} p_3(a_3|a_1,a_2) = 1$
   ...
   $p_n(a_n|a_1,\dots,a_{n-1}) \geq 0 \quad \sum\limits_{a_n\in\Omega_n} p_n(a_n|a_1,\dots,a_{n-1}) = 1$

Setze für $\omega = (a_1,\dots,a_n) \in \Omega$

$p(\omega) := p_1(a_1)\cdot p_2(a_2|a_1)\cdot p_3(a_3|a_1,a_2)\cdot\ldots\cdot p_n(a_n|a_1,\dots,a_{n-1})$ 1. Pfadregel

Schließlich sei

$\mathbb{P}(A) := \sum\limits_{\omega\in A} p(\omega), \quad A \subset \Omega$ Produkt von Übergangswahrscheinlichkeiten

7.3 Satz

$(\Omega,\mathbb{P})$ ist diskreter Wahrscheinlichkeitsraum.

Beweis. zu zeigen: $\sum\limits_{\omega\in\Omega} p(\omega) = 1$
Induktion (oder direkt)! Zum Beispiel gilt für $n = 2$

$$\sum_{\omega\in\Omega} p(\omega) = \sum_{(a_1,a_2)\in\Omega_1\times\Omega_2} p_1(a_1)p_2(a_2|a_1) = \sum_{a_1\in\Omega_1}\sum_{a_2\in\Omega_2} p_1(a_1)p_2(a_2|a_1)$$

$$\sum_{a_1\in\Omega_1} p_1(a_1)\cdot 1 = 1.$$

□

7.4 Beispiel

Unabhängige Experimente $(\Omega_j,\mathbb{P}_j), j = 1,\dots,n$, diskrete Wahrscheinlichkeitsräume, $p_i(a_i) = \mathbb{P}_i(\{a_i\})$
Idee: "Unabhängiges" Durchführen der zugehörigen Experimente

$$\Omega := \Omega_1 \times \ldots \times \Omega_n, p(\omega) := p_1(a_1)\cdot\ldots\cdot p_n(a_n), \omega = (a_1,\dots,a_n) \in \Omega$$

$\big(p_2(a_2|a_1) = p_2(a_2),\dots,p_n(a_n|a_1,\dots,a_{n-1}) = p_n(a_n)\big)$

$$\mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$$

Man nennt $\mathbb{P}$ das **Produkt** von $\mathbb{P}_1, \ldots, \mathbb{P}_n$ und schreibt

$$\mathbb{P} := \bigotimes_{i=1}^{n} \mathbb{P}_i.$$

z.B. kann $\Omega = \Omega_1 \times \Omega_2, \quad \Omega_1 = \Omega_2 = \{1, \ldots, 6\}$
$p_1(a_1) = p_2(a_2) = \frac{1}{6}$
Dann ist

$$p(a_1, a_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

und $\mathbb{P}$ ist die Laplace-Verteilung auf Ω.

Kapitel 8

Bedingte Wahrscheinlichkeiten

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum.

8.1 Definition

Sei $B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von $A \subset \Omega$ unter der Bedingung B.
Alternativ: $P_B(A) := \mathbb{P}(A|B)$

8.2 Satz

P_B ist ein Wahrscheinlichkeitsmaß auf Ω. Dabei ist $P_B(A) = 1$ falls $B \subset A$ und $P_B(A) = 0$ falls $A \cap B = \emptyset$. Es gilt:

$$p_B(\omega) := \begin{cases} \frac{p(\omega)}{\mathbb{P}(B)} & \text{, falls } \omega \in B \\ 0 & \text{, sonst} \end{cases} \qquad \text{mit } p_B(\omega) := \mathbb{P}_B(\{\omega\})$$

Beweis ist klar! ($\sum\limits_{\omega \in \Omega} p_B(\omega) = \frac{1}{\mathbb{P}(B)} \sum\limits_{\omega \in B} p(\omega) = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.)

Motivation Für $A \subset B$

$$\frac{h_n(A)}{h_n(B)} = \frac{\frac{1}{n} h_n(A)}{\frac{1}{n} h_n(B)} \rightsquigarrow \frac{\mathbb{P}(A)}{\mathbb{P}(B)}.$$

8.3 Bemerkung (Zusammenhang zu Übergangswahrscheinlichkeiten)

$\Omega = \Omega_1 \times \Omega_2, \quad p(\omega) = p_1(a_1)p_2(a_2|a_1), \quad \omega = (a_1, a_2)$
Für $a_1 \in \Omega_1$ sei

$$B := \{a_1\} \times \Omega_2.$$

Für $a_2 \in \Omega_2$ sei

$$A := \Omega_1 \times \{a_2\}.$$

Es gilt $A \cap B = \{(a_1, a_2)\}$,

$$\mathbb{P}(A \cap B) = \sum_{\omega \in A \cap B} p(\omega) = \sum_{(b_1,b_2) \in A \cap B} p_1(b_1)p_2(b_2|b_1) = p_1(a_1)p_2(a_2|a_1),$$

$$\mathbb{P}(B) = \sum_{b_2 \in \Omega_2} p(a_1|b_2) = \sum_{b_2 \in \Omega_2} p_1(a_1)p_2(b_2|a_1) = p_1(a_1)$$

Es folgt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \overset{p_1(a_1)>0}{=} p_2(a_2|a_1)$$

8.4 Satz (Multiplikationsformel)

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

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

Beweis. Für $n = 2$:

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

Allgemein: Ausschreiben der Definitionen + kürzen
$n = 3$: rechte Seite: $\mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} = \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} = \mathbb{P}(A_1 \cap A_2 \cap A_3)$ □

8.5 Satz

Sei $A_1, A_2, \ldots$ Zerlegung von $\Omega(\bigcup A_i = \Omega, A_i \cap A_j = \emptyset, i \neq j)$.
Sei $B \subset \Omega$. Dann gilt

1. $\mathbb{P}(B) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)\mathbb{P}(B|A_j)$ Formel der totalen Wahrscheinlichkeit

2. [1] Für $\mathbb{P}(B) > 0$, so gilt

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\mathbb{P}(B|A_k)}{\sum_{j=1}^{\infty} \mathbb{P}(A_j)\mathbb{P}(B|A_j)}, \quad k = 1, 2, \ldots$$

[1] Formel von Bayes

(Man vereinbart $\mathbb{P}(B|A_j)\mathbb{P}(A_j) := 0$, falls $\mathbb{P}(A_j) = 0$)

Beweis. 1. $B = B \cap \Omega = \bigcup_{j=1}^{\infty} \underbrace{B \cap A_j}_{\text{paarweise disjunkt}}$ Aus der σ-Additivität von $\mathbb{P}$ folgt

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

2. rechte Seite der Behauptung: $\frac{\mathbb{P}(B \cap A_k)}{\mathbb{P}(B)} \overset{!}{=} \mathbb{P}(A_k|B)$

□

8.6 Beispiel

Eine Krankheit komme bei 4% der Bevölkerung vor[2]. Ein Test spreche bei 90% der Kranken an und bei 20% der Gesunden!

Modell

- Ω : Menge der Personen in Deutschland
- $K \subset \Omega$: Menge der kranken Personen
- $A \subset \Omega$: Menge der (hypothetisch) positiv getesteten Personen
- $\mathbb{P}$ = Gleichverteilung auf Ω

Dann

$$\begin{aligned}
\mathbb{P}(K|A) = {} & \text{Wahrscheinlichkeit, dass eine positiv getestete Person krank ist} \\
& \overset{Bayes}{=} \frac{\mathbb{P}(K)\mathbb{P}(A|K)}{\mathbb{P}(K)\mathbb{P}(A|K) + \mathbb{P}(K^c)\mathbb{P}(A|K^c)} \quad (K = A_j, K^c = A_k) \\
& = \frac{0,04 \cdot 0,9}{0,04 \cdot 0,9 + 0,96 \cdot 0,2} = \frac{0,036}{0,036 + 0,192} = \frac{0,036}{0,228} = 0,158
\end{aligned}$$

8.7 Beispiel (Ziegenproblem)

Ausgelassen.

[2] Die Mediziner sprechen von "Prävalenz".

8.8 Beispiel (Simpson-Paradoxon)

Zulassung von Studenten in Berkeley (1973)

- Zulassungsrate Männer: 44%
- Zulassungsrate Frauen: 35%

Aber: Zulassungsraten der Männer in den einzelnen Fächern kleiner als die der Frauen

Erklärung

- $A \widehat{=}$ Zulassung [3]
- $B \widehat{=}$ Frau [4]
- $K_j \widehat{=}$ Bewerbung für Fach j

Dann kann gelten

$$\mathbb{P}(A|B) < \mathbb{P}(A|B^c)$$

aber

$$\mathbb{P}(A|B \cap K_j) > \mathbb{P}(A|B^c \cap K_j), \quad j = 1, 2, \dots$$

Denn:

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

$$= \sum_j \underbrace{\mathbb{P}(K_j|B)}_{\text{Bewerbungsrate der Frauen im j-ten Fach}} \underbrace{\mathbb{P}(A|B \cap K_j)}_{\text{siehe oben}}$$

analog

$$\mathbb{P}(A|B^c) = \sum \mathbb{P}(K_j|B^c)\mathbb{P}(A|B^c \cap K_j)$$

Die absolute Erfolgsquote ist eine gewichtete Summe der relativen Erfolgsquoten.

[3] Ereignis, dass rein zufällig ausgewählter Bewerber erfolgreich ist mit seiner Bewerbung.

[4] Ereignis, dass zufällig ausgewählte weibliche Bewerberin erfolgreich ist.

Kapitel 9

Stochastische Unabhängigkeit

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum.

9.1 Definition

$A_1, \ldots, A_n \subset \Omega$ heißen **stochastisch unabhängig**, falls

$$\mathbb{P}(\bigcap_{j \in T} A_j) = \prod_{j \in T} \mathbb{P}(A_j), \quad T \subseteq \{1, \ldots, n\}, |T| \geq 2.$$

($2^n - n - 1$ Gleichungen.)

9.2 Bemerkung

1. A, B stochastisch unabhängig
   $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
   $\overset{\mathbb{P}(B)>0}{\Leftrightarrow} \mathbb{P}(A|B) = \mathbb{P}(A)$ (Interpretation!) [1]
   $\Leftrightarrow \mathbb{P}(B|A) \overset{\mathbb{P}(B)>0}{=} \mathbb{P}(B)$

2. $\mathbb{P}(B) = 0 \rightsquigarrow A$ und B sind stochastisch unabhängig
   $\mathbb{P}(B) = 1 \rightsquigarrow A$ und B sind stochastisch unabhängig

3. A, B, C unabhängig $\Leftrightarrow$

$$\left.\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B) \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C) \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C) \end{aligned}\right\} \text{paarweise stochastische Unabhängigkeit}$$

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

[1] Wenn die Kenntnis des Eintretens von B keinerlei Rückschlüsse auf das Eintreten von A zulässt.

Wiederholung $A, B \subset \Omega$ stochastisch unabhängig

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

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

$A_1, \ldots, A_n$ stochastisch unabhängig $\Leftrightarrow$

$$\mathbb{P}(\bigcap_{j \in T} A_j) = \prod_{j \in T} \mathbb{P}(A_j), \quad T \subset \{1, \ldots, n\}, |T| \geq 2$$

9.3 Bemerkungen

(iv) A, B stochastisch unabhängig. Dann:

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

Also sind A und B^c (also auch A^c und B^c bzw. A^c und B) stochastisch unabhängig.

(v) Seien $A_1, \ldots, A_n$ unabhängig und $1 \leq i_1 < \ldots < i_k \leq n$. Dann sind $A_{i_1}, \ldots, A_{i_k}$ stochastisch unabhängig.

(vi) Ist A von A unabhängig, so ist

$$\mathbb{P}(A) = \mathbb{P}(A)^2$$

d.h. $\mathbb{P}(A) \in \{0, 1\}$.

(vii) Man nennt $A_1, A_2, \ldots \subset \Omega$ stochastisch unabhängig

$$\overset{d}{\Leftrightarrow} A_1, \ldots, A_n \text{ stochastisch unabhängig für jedes } n \geq 2.$$

9.4 Beispiel (Produkträume)

Sei $(\Omega, \mathbb{P}) := (\bigotimes_{j=1}^{n} \Omega_j, \bigotimes_{j=1}^{n} \mathbb{P}_j)$, d.h.

$$\begin{aligned} \mathbb{P}(\{(a_1, \ldots, a_n)\}) &= p(\omega) \quad \omega = (a_1, \ldots, a_n) \\ &= p_1(a_1) \cdot \ldots \cdot p_n(a_n) \quad (p_i(a_i) = \mathbb{P}_i(\{a_i\})) \end{aligned}$$

Sei $B = B_1^* \times \ldots \times B_n^*, \quad B_i^* \in \Omega_i$. Dann $\mathbb{P}(B_1^* \times \ldots \times B_n^*)$

$$= \sum_{(a_1, \ldots, a_n) \in B_1^* \times \ldots \times B_n^*} p_1(a_1) \cdot \ldots \cdot p_n(a_n)$$

$$= \sum_{a_1 \in B_1^*} \dots \sum_{a_n \in B_n^*} p_1(a_1) \cdot \ldots \cdot p_n(a_n)$$

$$= \prod_{j=1}^{n} \sum_{a \in B_j^*} p_j(a) = \prod_{j=1}^{n} \mathbb{P}_j(B_j^*) \qquad (*)$$

Sei jetzt $A_j^* \subset \Omega_j, j = 1, \dots, n$.

Behauptung $A_j = \Omega_1 \times \ldots \times \Omega_{j-1} \times A_j^* \times \Omega_{j+1} \times \ldots \times \Omega_n, \quad j = 1, \dots, n$ stochastisch unabhängig.

Beweis. Sei $T \subset \{1, \dots, n\}$ mit $|T| \geq 2$.
Definiere

$$B_j^* := \begin{cases} A_j^*, & j \in T, \\ \Omega_j, & j \notin T. \end{cases}$$

Dann

$$\mathbb{P}(\bigcap_{j \in T} A_j) = \mathbb{P}(B_1^* \times \ldots \times B_n^*)$$

[2]

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

$$\overset{(*)}{=} \prod_{j \in T} \mathbb{P}_j(A_j)$$

[3] □

9.5 Satz

$A_1, \dots, A_n$ stochastisch unabhängig $\Leftrightarrow$

$$\mathbb{P}(\bigcap_{j \in I} A_j \cap \bigcap_{j \in J} A_j^c) = \prod_{j \in I} \mathbb{P}(A_j) \prod_{j \in J} \mathbb{P}(A_j^c) \quad I, J \subset \{1, \dots, n\}, I \cap J = \emptyset$$

(Hierbei $\bigcap_{j \in \emptyset} B_j := \Omega$, $\prod_{j \in \emptyset} a_j := 1$)

Beweis. Induktion über Anzahl der Elemente von J (vergleiche auch Bemerkung 9.3 (iv)) □

[2] $(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2)$
[3] mit $B_i^* = \Omega_i$ bis auf ein i

Definition Für $A \subset \Omega$ sei $A^0 := A^c, A^1 := A$.
Für $B_1, \ldots, B_n \subset \Omega$ sei

$$\sigma(B_1, \ldots, B_k) := \{B \subset \Omega \colon \exists U \subset \{0,1\}^k \text{ mit } B = \bigcup_{(\epsilon_1,\ldots,\epsilon_n)\in U} B_1^{\epsilon_1} \cap \ldots \cap B_k^{\epsilon_k}\}.$$

(Die von $B_1, \ldots, B_k$ erzeugte <u>Algebra</u>).

Beispiel 9.1. *(TODO: Bild)*

Bemerkung 9.1. *Eine Menge der Form*

$$B_1^{\epsilon_1} \cap \ldots \cap B_k^{\epsilon_k} \text{ für } (\epsilon_1, \ldots, \epsilon_k) \in \{0,1\}^k$$

heißt <u>Atom</u> von $\sigma(B_1, \ldots, B_k)$. Jede Menge in $\sigma(B_1, \ldots, B_k)$ ist Vereinigung von Atomen. Insbesondere gilt

$$B_1, \ldots, B_k \in \sigma(B_1, \ldots, B_k), \emptyset \in \sigma(B_1, \ldots, B_k), \Omega \in \sigma(B_1, \ldots, B_k).$$

9.6 Satz (Blockungslemma)

Seien $A_1, \ldots, A_k, A_{k+1}, \ldots, A_n$ stochastisch unabhängig und
$B \in \sigma(A_1, \ldots, A_k), C \in \sigma(A_{k+1}, \ldots, A_n)$. Dann sind B und C stochastisch unabhängig.

Beweis. Es gilt

$$\mathbb{P}(B \cap C) = \mathbb{P}\left(\underbrace{(\bigcup_{(\epsilon_1,\ldots,\epsilon_k)\in U} A_1^{\epsilon_1} \cap \ldots A_k^{\epsilon_k})}_{B} \cap \underbrace{(\bigcup_{(\epsilon_{k+1},\ldots,\epsilon_n)\in V} A_{k+1}^{\epsilon_{k+1}} \cap \ldots \cap A_n^{\epsilon_n})}_{C} \right)$$

disjunkte Vereinigung

$$= \sum_U \mathbb{P}\left((A_1^{\epsilon_1} \cap \ldots \cap A_k^{\epsilon_k}) \cap \bigcup_V (A_{k+1}^{\epsilon_{k+1}} \cap \ldots \cap A_n^{\epsilon_n}) \right)$$

$$= \sum_{U,V} \mathbb{P}(A_1^{\epsilon_1} \cap \ldots \cap A_k^{\epsilon_k} \cap A_{k+1}^{\epsilon_{k+1}} \cap \ldots A_n^{\epsilon_n})$$

$$= \sum_{U,V} \left(\underbrace{\prod_{j=1}^{k} \mathbb{P}(A_j^{\epsilon_j})}_{\mathbb{P}(A_1^{\epsilon_1}\cap\ldots\cap A_k^{\epsilon_k})} \underbrace{\prod_{j=k+1}^{n} \mathbb{P}(A_j^{\epsilon_j})}_{\mathbb{P}(A_{k+1}^{\epsilon_{k+1}}\cap\ldots\cap A_n^{\epsilon_n})} \right)$$

$$= \sum_{(\epsilon_1,\dots,\epsilon_k)\in U} \mathbb{P}(A_1^{\epsilon_1} \cap \ldots \cap A_k^{\epsilon_k}) \sum_{(\epsilon_{k+1},\dots,\epsilon_n)\in V} \mathbb{P}(A_{k+1}^{\epsilon_{k+1}} \cap \ldots \cap A_n^{\epsilon_n})$$

$$\stackrel{!}{=} \mathbb{P}(B)\mathbb{P}(C)$$

□

9.7 Satz

$A_1, \ldots, A_n \subset \Omega$ stochastisch unabhängig. Ferner gelte $\mathbb{P}(A_i) = p, \quad i = 1, \ldots, n$. Dann ist

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

$Bin(n,p)$-verteilt.

Beweis. Es gilt

$$\{X = k\} = \bigcup_{T\subseteq\{1,\dots,n\},|T|=k} \left(\bigcap_{j\in T} A_j \cap \bigcap_{j\notin T} A_j^c \right)$$

disjunkte Vereinigung, also

$$\mathbb{P}(X = k) = \mathbb{P}(\{X = k\}) = \sum_{T\subseteq\{1,\dots,n\},|T|=k} \mathbb{P}(\bigcap_{j\in T} A_j \cap \bigcap_{j\notin T} A_j^c)$$

$$\stackrel{\text{Voraussetzung}}{=} \sum_{T\subseteq\{1,\dots,n\},|T|=k} p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

□

9.8 Beispiel (Bernoulli-Kette der Länge n)

$(\Omega, \mathbb{P}) := \bigotimes_{j=1}^{n} (\Omega_j, \mathbb{P}_j)$
$\Omega_1 = \ldots = \Omega_n = \{0, 1\}$
$\mathbb{P}_j(\{1\}) = 1 - \mathbb{P}_j(\{0\}) = p, \quad j = 1, \ldots, n$
Die Ereignisse

$$A_j := \{(a_1, \ldots, a_n) \in \Omega \colon a_j = 1\}, \quad j = 1, \ldots, n$$

sind stochastisch unabhängig.
Ferner $\mathbb{P}(A_j) = p$.

Kapitel 10

Gemeinsame Verteilung

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum (mit Träger Ω_0).

10.1 Definition

Seien $X_j\colon \Omega \to \mathbb{R} \quad (j = 1, \ldots, k)$ Zufallsvariablen. Definiere $X\colon \Omega \to \mathbb{R}^k$ vermöge $X(\omega) = (X_1(\omega), \ldots, X_k(\omega))\,, \quad \omega \in \Omega$. Dann heißt X k-**dimensionaler Zufallsvektor**.
Für $B \subset \mathbb{R}^k$ sei $X^{-1}(B) \equiv \{X \in B\} := \{\omega \in \Omega\colon X(\omega) \in B\}$. Die Abbildung

$$\mathbb{P}^X\colon \mathcal{P}(\mathbb{R}) \to [0,1]$$

definiert durch

$$\mathbb{P}^X(B) := \mathbb{P}(\{X \in B\})(= \mathbb{P}(X \in B))$$

heißt Verteilung von X oder auch gemeinsame Verteilung von $X_1, \ldots, X_k$.
Die Verteilung von X_j heißt j**-te Marginalverteilung (von** X_j**)**.

Wiederholung $(\Omega, \mathbb{P})$ Wahrscheinlichkeitsraum
$\Omega_0 := \{\omega \in \Omega\colon \mathbb{P}(\{\omega\}) > 0\}$ diskret
$X = (X_1, \ldots, X_k)\colon \Omega \to \mathbb{R}^k$
Verteilung: $\mathbb{P}^X$

$$\mathbb{P}^X(A) := \mathbb{P}(X \in A) \equiv \mathbb{P}(\{\omega\colon X(\omega) \in A\}), \quad A \subset \mathbb{R}^k$$

Bemerkung Die gemeinsame Verteilung legt Randverteilungen fest. ($k = 2$)

$$\begin{aligned} &\mathbb{P}(X_1 = x_1) \quad \left(= \mathbb{P}^{X_1}(\{x_1\})\right) \\ &= \mathbb{P}\left(\{X_1 = x_1\} \cap \bigcup_{x_2 \in X_2(\Omega_0)} \{X_2 = x_2\}\right) \end{aligned}$$

$$\overset{\sigma\text{-Additivität}}{=} \sum_{x_2 \in X_2(\Omega_0)} \underbrace{\mathbb{P}(\{X_1 = x_1\} \cap \{X_2 = x_2\})}_{\mathbb{P}(X_1 = x_1, X_2 = x_2)}$$
$$= \sum_{x_2 \in X_2(\Omega_0)} \mathbb{P}^{(X_1, X_2)}(\{(x_1, x_2)\})$$

10.2 Beispiel

$\Omega := \{1, \ldots, 6\}^2$
$\mathbb{P} =$ Gleichverteilung (2-maliger Würfelwurf)

$$X_1((k,l)) := \min(k,l), X_2((k,l)) := \max(k,l)$$

(TODO: Tabelle)

$$\mathbb{P}(X_1 = i, X_2 = j), \quad i, j = 1, \ldots, 6$$

Es gilt

$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = 7 - i), \quad i = 1, \ldots, 6$$
$$\rightsquigarrow \mathbb{P}^{X_1} = \mathbb{P}^{7 - X_2} \quad (X_1 \neq 7 - X_2)$$
$$X_1 \overset{d}{=} 7 - X_2 \quad \text{Verteilungsgleichheit}$$

10.3 Beispiel

$\mathbb{P}(X_1 = i, X_2 = j)$
($c \in [0, \frac{1}{2}]$ fest)

i \ j	1	2	
1	c	$\frac{1}{2} - c$	$\frac{1}{2}$
2	$\frac{1}{2} - c$	c	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) + \mathbb{P}(X_1 = 1, X_2 = 2)$$

$$\rightsquigarrow \mathbb{P}^{X_1} = \mathbb{P}^{X_2} \mathrel{\hat{=}} \text{fairer Münzwurf!}$$

Also legen die Randverteilungen $\mathbb{P}^{X_1}, \mathbb{P}^{X_2}$ die gemeinsame Verteilung $\mathbb{P}^{(X_1, X_2)}$ <u>nicht</u> fest.

10.4 Definition

$X_1, \ldots, X_k \colon \Omega \to \mathbb{R}$ heißen **stochastisch unabhängig**

$$\Leftrightarrow \{X_1 \in B_1\}, \ldots, \{X_k \in B_k\} \text{ stochastisch unabhängig } \forall B_1, \ldots, B_k \subset \mathbb{R}$$

10.5 Satz

Die folgenden Aussagen sind äquivalent:

1. $X_1, \dots, X_k$ sind stochastisch unabhängig
2. $\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \prod_{i=1}^{k} \mathbb{P}(X_i \in B_i) \quad B_1, \dots, B_k \subset \mathbb{R}$
3. $\mathbb{P}(X_1 = x_1, \dots, X_k = x_k) = \prod_{i=1}^{k} \mathbb{P}(X_i = x_i) \quad x_1, \dots, x_k \in \mathbb{R}$

Beweis. (1) $\Rightarrow$ (2) Klar nach Definition.
(2) $\Rightarrow$ (1): Wähle in der Definition $B_j = \mathbb{R}$ für $j \notin T(\{X_j \in B_j\} = \Omega)$
(2) $\Rightarrow$ (3): Setze $B_i = \{x_i\}$
(3) $\Rightarrow$ (2): Für $B_1, \dots, B_k \subset \mathbb{R}$ gilt

$$\begin{aligned}\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) \quad (= \mathbb{P}^{(X_1,\dots,X_k)}(B_1 \times \dots \times B_k)) \quad (B_i \subset \underbrace{X_i(\Omega_0)}_{\text{diskret}})\\ = \sum_{x_1 \in B_1, \dots, x_k \in B_k} \mathbb{P}(X_1 = x_1, \dots, X_k = x_k)\\ \overset{(3)}{=} \sum \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_k = x_k)\\ = \mathbb{P}(X_1 \in B_1) \cdot \ldots \cdot \mathbb{P}(X_k \in B_k)\end{aligned}$$

[1] □

Bemerkung Im Falle stochastischer Unabhängigkeit legen die Randverteilungen die gemeinsame Verteilung fest.

10.6 Satz (Blockungslemma)

Es seien $X_1, \dots, X_k$ stochastisch unabhängige, eindimensionale Zufallsvariablen und $1 \leq l \leq k-1$, $g\colon \mathbb{R}^l \to \mathbb{R}$, $h\colon \mathbb{R}^{k-l} \to \mathbb{R}$. Dann sind $g(X_1, \dots, X_l)$ und $h(X_{l+1}, \dots, X_k)$ stochastisch unabhängig.

Beweis. Übung! □

[1] $\sum_{i,j} a_i b_j = (\sum a_i)(\sum b_j)$ "Fubini"

10.7 Satz (allgemeine Transformations-Formel)

Seien $Z\colon \Omega \to \mathbb{R}^k$ Zufallsvektor, $g\colon \mathbb{R}^k \to \mathbb{R}$. Setze

$$M := \{z \in \mathbb{R}^k \colon \mathbb{P}(Z = z) > 0\}.$$

Dann ist $g(Z)$ integrierbar[2] genau dann, wenn

$$\sum_{z \in M} |g(z)| \cdot \mathbb{P}(Z = z) < \infty$$

In diesem Fall ist der Erwartungswert

$$\mathbb{E}g(Z) = \sum_{z \in M} g(z) \cdot \mathbb{P}(Z = z)$$

Beweis. vgl. eindimensionalen Spezialfall. □

10.8 Satz

Seien $X, Y\colon \Omega \to \mathbb{R}$ stochastisch unabhängig. Sind X, Y integrierbar, so ist auch $X \cdot Y$ integrierbar und $\mathbb{E}(X \cdot Y) = (\mathbb{E}X) \cdot (\mathbb{E}Y)$.

Beweis. Setze $M := \{(x, y) \in \mathbb{R}^2 \colon \mathbb{P}(X = x, Y = y) > 0\} = \{(x, y)\colon \mathbb{P}(X = x) > 0, \mathbb{P}(Y = y) > 0\}$. Dann

$$\begin{aligned}
\mathbb{E}|X \cdot Y| &= \sum_{\omega \in \Omega_0} |X(\omega)||Y(\omega)|\mathbb{P}(\{\omega\}) \\
&\overset{!}{=} \sum_{(x,y) \in M} |x||y| \underbrace{\mathbb{P}(X = x, Y = y)}_{\mathbb{P}(X=x)\cdot\mathbb{P}(Y=y)} \\
&= \underbrace{\left(\sum_{x\colon \mathbb{P}(X=x)>0} |x|\mathbb{P}(X = x)\right)}_{<\infty} \cdot \underbrace{\left(\sum_{y\colon \mathbb{P}(Y=y)>0} |y|\mathbb{P}(Y = y)\right)}_{<\infty}
\end{aligned}$$

Dieselbe Rechnung "ohne Betragsstriche" liefert die behauptete Formel. □

10.9 Satz (Faltungsformel)

Sind X, Y unabhängige reelle Zufallsvariablen, so gilt

$$\mathbb{P}(X + Y = z) = \sum_{x \in X(\Omega_0)} \mathbb{P}(X = x)\mathbb{P}(Y = z - x), \quad z \in \mathbb{R}$$

[2] d.h. der Erwartungswert existiert.

Beweis. Ohne Unabhängigkeit gilt

$$\mathbb{P}(X+Y=z) \overset{!}{=} \sum_{X \in X(\Omega_0)} \mathbb{P}(X=x, Y=z-x)$$

□

10.10 Satz (Additionsgesetz für Binomialverteilungen)

Seien $X \sim Bin(m,p), Y \sim Bin(n,p)$ stochastisch unabhängig.
Dann ist $X+Y \sim Bin(m+n,p) \quad \forall m,n \geq 1, p \in [0,1]$

Beweis. Es seien $A_1, \ldots, A_m, A_{m+1}, \ldots, A_{m+n}$ unabhängige Ereignisse mit $\mathbb{P}(A_i) = p$. Dann sind

$$X' := \sum_{i=1}^{m} 1_{A_i}, \qquad Y' := \sum_{i=m+1}^{m+n} 1_{A_i}$$

$Bin(m,p)$ bzw. $Bin(n,p)$ Binomialverteilt. (Bernoulli-Kette)
Nach Blockungslemma sind X', Y' stochastisch unabhängig! Außerdem

$$X' + Y' = \sum_{i=1}^{m+n} 1_{A_i} \sim Bin(m+n,p)$$

Aber aus $(X', Y') \overset{d}{=} (X, Y)$ folgt

$$X' + Y' \overset{d}{=} X + Y, \text{ d.h. } X + Y \sim Bin(m+n,p)$$
$$\mathbb{P}^{X'+Y'} = \mathbb{P}^{X+Y}$$

□

Kapitel 11

Varianz, Kovarianz, Korrelation

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum.

11.1 Definition

Falls X Zufallsvariable und $\mathbb{E}X^2 < \infty$, so heißt

$$V(X) \equiv Var(X) := \mathbb{E}(X - \mathbb{E}X)^2$$

Varianz von X.

11.2 Bemerkungen

1. Wegen
$$|X| \leq 1 + X^2$$
$$(X - a)^2 \leq X^2 + 2|a||X| + a^2$$
ist $V(X)$ wohldefiniert.

2. Gilt $\sum\limits_{i=1}^{\infty} \mathbb{P}(X = x_i) = 1$, so ist
$$V(X) = \sum (x_i - \mathbb{E}X)^2 \mathbb{P}(X = x_i) \quad \text{(Transformationsformel)}$$

3. Es gilt
$$V(X) = \mathbb{E}(X^2 - 2\underbrace{(\mathbb{E}X)}_{\mu} X + \underbrace{(\mathbb{E}X)^2}_{\mu^2})$$
$$= \mathbb{E}X^2 - 2\mu\mathbb{E}X + \mu^2 \quad \text{(Linearität des Erwartungswertes)}$$
$$= \mathbb{E}X^2 - \mu^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

4. Varianz kann als Trägheitsmoment interpretiert werden.

11.3 Satz

Sei $\mathbb{E}X^2 < \infty$.

1. $V(X) = \mathbb{E}(X-c)^2 - (\mathbb{E}X - c)^2, \quad c \in \mathbb{R}$ (Steiner-Formel)
2. $V(X) = \min\limits_c \mathbb{E}(X-c)^2$
3. $V(aX+b) = a^2 V(X)$
4. $V(X) = 0 \Leftrightarrow \exists a \in \mathbb{R}\colon \mathbb{P}(X=a) = 1.$

Beweis. (1) $V(X) =$

$$= \mathbb{E}(\underbrace{X-c}+\underbrace{c-\mathbb{E}X})^2$$

$$= \mathbb{E}(X-c)^2 + 2\underbrace{\mathbb{E}(X-c)(c-\mathbb{E}X)}+(c-\mathbb{E}X)^2$$

[1]

$$= \mathbb{E}(X-c)^2 - 2\mathbb{E}(c-\mathbb{E}X)^2 + (c-\mathbb{E}X)^2$$

(2) (1) $\rightsquigarrow \mathbb{E}(X-c)^2 = V(X) + (\mathbb{E}X-c)^2$
(3) $\mathbb{E}(aX+b-\mathbb{E}(aX+b))^2$

$$= \mathbb{E}(aX+b-a\mathbb{E}X-b)^2$$

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

(4) Bemerkung 11.2.(2) $\rightsquigarrow$

$$V(X) = 0 \Leftrightarrow (x_i - \mathbb{E}X)^2\mathbb{P}(X=x_i) = 0$$

$$\Leftrightarrow \forall i \text{ mit } \mathbb{P}(X=x_i) \text{ gilt } x_i = \mathbb{E}X$$

$$\Leftrightarrow \text{ Es gibt nur ein } i_0 \text{ mit } \mathbb{P}(X=x_{i_0}) > 0$$

$$\text{Dann ist } \mathbb{P}(X=x_{i_0}) = 1, \text{ und } x_{i_0} = \mathbb{E}X.$$

□

11.4 Beispiel

1. $X = 1_A, \quad A \subset \Omega.$

$$VarX = \mathbb{E}1_A^2 - (\mathbb{E}1_A)^2 = \mathbb{E}1_A - (\mathbb{E}1_A)^2$$

$$= \mathbb{P}(A) - \mathbb{P}(A)^2 = \mathbb{P}(A)(1-\mathbb{P}(A))$$

[1] $\mathbb{E}c = c$

2. $X = \sum_{i=1}^{n} 1_{A_i}, \quad A_i \subset \Omega_i, i = 1, \ldots, n$

$$V(X) = \mathbb{E}\left(\sum_{i=1}^{n} 1_{A_i}\right) \cdot \left(\sum_{j=1}^{n} 1_{A_j}\right) - (\mathbb{E}\sum 1_{A_i})^2$$

$$= \sum_{i=1}^{n} \mathbb{P}(A_i) + \sum_{i \neq j} \mathbb{P}(A_i \cap A_j) - (\sum_{i=1}^{n} \mathbb{P}(A_i))^2$$

Es gelte etwa (für ein $c \geq -r, r, s \in \mathbb{N}$)

$$\mathbb{P}(A_i) = \frac{r}{r+s} =: p$$

$$\mathbb{P}(A_i \cap A_j) = \frac{r}{r+s}\frac{r+c}{r+s+c}, \quad i \neq j$$

($c = -1$: Ziehen ohne Zurücklegen, $c = 0 \hat{=}$ Ziehen mit Zurücklegen, $c > 0$: Polyasches Urnenschema). Dann

$$V\left(\sum_{i=1}^{n} X_i\right) = np(1-p) \cdot \left(1 + \frac{(n-1) \cdot c}{r+s+c}\right)$$

(3) $X \sim Bin(n,p) : V(x) = np(1-p) \quad (c = 0)$
(4) $X \sim Hyp(n,r,s) : V(X) = np(1-p)\left(1 - \frac{(n-1)}{r+s-1}\right) \quad (c = -1)$

11.5 Definition

X heißt <u>standardisiert</u>, wenn $\mathbb{E}X = 0$ und $V(X) = 1$. Ist X eine beliebige Zufallsvariable ($\mathbb{E}X^2 < \infty$), so heißt (falls $V(X) > 0$)

$$X^* = \frac{X - \mathbb{E}X}{\sqrt{V(X)}}$$

Standardisierung von X. (Es gilt $\mathbb{E}X^* = 0, V(X^*) = 1$)

Bemerkung $X \sim Bin(n,p), \quad X^* = \frac{X-np}{\sqrt{np(1-p)}}$

11.6 Satz (Tschebyschov-Ungleichung)

Für jede Zufallsvariable X mit $\mathbb{E}X^2 < \infty$ gilt

$$\mathbb{P}(|X - \mathbb{E}X| \geq c) \leq \frac{V(X)}{c^2}), \quad c > 0.$$

Beweis. Es sei (für gegebenes $c > 0$)

$$g(t) = \begin{cases} 1, & \text{falls } |t - \mathbb{E}X| \geq c \\ 0, & \text{sonst} \end{cases}, \quad t \in \mathbb{R}$$

Ferner sei

$$h(t) = \frac{(t - \mathbb{E}X)^2}{c^2}$$

(TODO: Bild)
Wegen $g \leq h$ ist $g(X) \leq h(X)$, also

$$\underbrace{\mathbb{E}g(X)}_{=\mathbb{P}(|X-\mathbb{E}X|\geq c)} \leq \underbrace{\mathbb{E}h(X)}_{=\underbrace{\mathbb{E}\frac{(X-\mathbb{E}X)^2}{c^2}}_{=\frac{1}{c^2}V(X)}}$$

□

11.7 Definition

Es gelte $\mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty$. Die Zahl

$$(Cov(X,Y) =)C(X,Y) := \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

heißt **Kovarianz zwischen** X **und** Y. Gilt $C(X,Y) = 0$, so heißen X und Y **unkorreliert**. Gilt $V(X) > 0, V(Y) > 0$, so heißt

$$\rho(X,Y) := \frac{C(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Korrelationskoeffizient zwischen X **und** Y. (Wegen $|a \cdot b| \leq \frac{1}{2}(a^2 + b^2)$ ist $(X - \mathbb{E}X)(Y - \mathbb{E}Y) \in L^1(\mathbb{P}))$[2].

11.8 Satz

1. $C(X,Y) = \mathbb{E}XY - (\mathbb{E}X\mathbb{E}Y)$
2. $C(X,Y) = C(Y,X), \quad C(X,X) = V(X)$
3. $C(aX + b, cY + d) = a \cdot c \cdot C(X,Y)$
4. X, Y unabhängig $\Rightarrow C(X,Y) = 0$
5. $V(X,Y) = V(X) + V(Y) + 2C(X,Y)$

[2] $L^1(\mathbb{P}) \rightsquigarrow$ integrierbar

6. $V(\sum\limits_{j=1}^{n} X_j) = \sum V(X_j) + \sum\limits_{i \neq j} C(X,Y)$

7. $C(1_A, 1_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$

8. $C(\sum\limits_{i=1}^{n} X_i, \sum\limits_{j=1}^{m} Y_j) = \sum\limits_{i,j} C(X_i, Y_j)$

9. $\rho(aX + b, cY + d) = sgn(a \cdot c)\rho(X,Y)$

Beweis. a),b),c) stimmt.
d) Satz 10.8.

$$C(X,Y) = \mathbb{E}X \cdot Y - \mathbb{E}X\mathbb{E}Y = 0.$$

e) folgt aus f, f folgt aus h
h) linke Seite

$$\begin{aligned} &\mathbb{E}(\sum_i X_i - \sum_i \mathbb{E}X_i, \sum_j X_j - \sum_j \mathbb{E}X_j) \\ &= \mathbb{E}(\sum_i (X_i - \mathbb{E}X_i))(\sum_j (X_j - \mathbb{E}X_j)) \\ &= \sum_{i,j} \mathbb{E}(X_i - \mathbb{E}X_i) \cdot (Y_j - \mathbb{E}Y_j) \end{aligned}$$

i) für $a, b \in \mathbb{R}, c, d \in \mathbb{R}$

$$\begin{aligned} &\rho(aX + b, cY + d) \\ &\stackrel{Def}{=} \frac{cov(aX + b, cY + d)}{\sqrt{Var(aX + b) \cdot Var(cY + d)}} \\ &\stackrel{11.8.c}{=} \frac{acCov(X,Y)}{\sqrt{a^2c^2}\sqrt{Var(X) \cdot Var(Y)}} \\ &= sgn(ac) \cdot \rho(X,Y) \end{aligned}$$

□

11.9 Folgerung

Sind $X_1, \ldots, X_n$ unabhängig, so folgt

$$V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i) \quad \text{(aus iv+vi(d+f))}$$

11.10 Beispiel

Für $X \sim Bin(n,p)$ gilt (nach Satz 9.6)

$$X \stackrel{d}{=} X_1 + \ldots + X_n$$

mit $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit

$$\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p \quad \text{für } i = 1, \ldots, n$$

Also ist nach 11.9 und 11.4.i

$$V(X) = \sum_{i=1}^{n} V(X_i)$$

$$= n \cdot V(X_1) = np(1-p)$$

in Übereinstimmmung mit 11.4.iii (siehe auch Übung). Das folgende Resultat ist analog zu 1.6:

11.11 Satz

Gilt $\mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty$ und $V(X)V(Y) > 0$, so folgt

$$\min_{a,b} \mathbb{E}(Y - a - bX)^2 = V(Y)(1 - \rho^2(X,Y))$$

Die Minimalstelle (a^*, b^*) ist gegeben durch

$$b^* = \frac{C(X,Y)}{V(X)}, \quad a^* = \mathbb{E}Y - b^*\mathbb{E}X$$

Beweis. $\left(\text{allg. } \inf_{x,y} f(x,y) = \inf_x \inf_y f(x,y) = \inf_y \inf_x f(x,y)\right)$
Es seien $a, b \in \mathbb{R}$ und $Z := Y - bX$. Dann ist

$$\mathbb{E}(Y - bX - a)^2 = \mathbb{E}(Z - a)^2$$

$$\stackrel{11.3.i,Steiner}{=} V(Z) + \underbrace{(\mathbb{E}Z - a)^2}_{\geq 0}$$

Also ist $a^* = \mathbb{E}Z = \mathbb{E}Y - b^*\mathbb{E}X$.
Es verbleibt die Aufgabe

$$\min_b \underbrace{\mathbb{E}(Y - \mathbb{E}Y - b(X - \mathbb{E}X))^2}_{=:f(b)}$$

$$f(b) = Var(Y) - 2bC(X,Y) + b^2 Var(X)$$

$$= {}^3V(X)\underbrace{\left(b - \frac{C(X,Y)}{V(X)}\right)^2}_{\geq 0} + V(Y) - \frac{C(X,Y)^2}{V(X)}$$

$$\Rightarrow b^* = \frac{C(X,Y)}{V(X)}$$

□

11.12 Folgerung

1. $C(X,Y)^2 \leq V(X) \cdot V(Y)$ (Cauchy-Schwarz-Ungleichung)
2. $|\rho(X,Y)| \leq 1$
3. $|\rho(X,Y)| = 1$

$$\Leftrightarrow \exists a, b \in \mathbb{R} \colon \mathbb{E}(Y - a - bX)^2 = 0$$
$$\Leftrightarrow \exists a, b \in \mathbb{R} \colon \mathbb{P}(Y - a - bX = 0) = 1$$
$$(\Leftrightarrow Y = a + bX \quad \mathbb{P} - \text{fast sicher})$$

insbesondere $\rho(X,Y) = 1 \Rightarrow b > 0$

$$(b^* = \frac{C(X,Y)}{V(X}) = \rho(X,Y) \cdot \underbrace{\sqrt{\frac{V(Y)}{V(X)}}}_{>0}$$

und $\rho(X,Y) = -1 \Rightarrow b < 0$.

11.13 Bemerkung

Falls $\mathbb{P}(X = x_j, Y = y_j) = \frac{1}{n} \quad (1 \leq j \leq n)$ so

$$\mathbb{E}(Y - a - bX)^2 = \sum_{j=1}^{n} \frac{1}{n} \cdot (y_j - a - bx_j)^2$$

$\rightsquigarrow$ Methode der kleinsten Quadrate $\rightsquigarrow$ Kapitel 1 (empirischer Korrelationskoeffizient)

[3] quadratische Ergänzung

Kapitel 12

Wichtige diskrete Verteilungen

1. Binomialverteilung $Bin(n,p) \rightsquigarrow$ Kapitel 6, 9.6
2. Hypergeometrische Verteilung $Hyp(n,r,s) \rightsquigarrow$ Kapitel 6
3. Poisson-Verteilung $Po(\lambda)$
4. Geometrische Verteilung $G(p)$
5. Negative Binomialverteilung $Nb(r,p)$
6. Multinomialverteilung $Mult(n,p_1,\dots,p_s)$

12.1 Satz (Gesetz seltener Ereignisse)

Sei $(p_n)_{n\geq 1}, 0 \leq p_n \leq 1$ eine Folge mit

$$np_n \overset{n\to\infty}{\to} \lambda \quad (0 < \lambda < \infty)$$

Dann gilt

$$\underbrace{\binom{n}{k} p_n^k (1-p_n)^{n-k}}_{=\mathbb{P}(X_n=k) \text{ für } X_n \sim Bin(n,p_n)} \overset{n\to\infty}{\to} e^{-\lambda}\frac{\lambda^k}{k!}$$

Beweis. linke Seite:

$$\frac{n!}{k!(n-k)!}\frac{1}{n^k} \cdot (n\cdot p_n)^k (1-\frac{n\cdot p_n}{n})^{n-k}$$

$$\frac{1}{k!}\underbrace{\frac{n^k}{n^k}}_{\to 1}\underbrace{(n,p_n)^k}_{\to\lambda^k}\underbrace{(1-\frac{n\cdot p_n}{n})^{n-k}}_{\to e^{-\lambda}}$$

$$\frac{n^k}{n^k} = \frac{n}{n}\cdot\frac{(n-1)}{n}\cdot\frac{(n-2)}{n}\cdot\ldots\cdot\frac{(n-k+1)}{n}$$

und allgemein für $a_n \to 1$ gilt

$$(1 + \frac{a_n}{n})^n \to e^a$$

□

12.2 Definition

$$X \sim Po(\lambda) \Leftrightarrow \mathbb{P}(X = k) = e^k \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

(Poisson-Verteilung mit Parameter $\lambda, \lambda \in (0, \infty)$) Also, falls $n \cdot p_n \to \lambda$,

$$Bin(n, p_n) \to Po(\lambda) \text{ im Sinne von 12.1}$$

12.3 Satz

1. $X \sim Po(\lambda) \Rightarrow \mathbb{E}X = V(X) = \lambda$
2. X, Y unabhängig, $X \sim Po(\lambda), Y \sim Po(\mu)$

$$\Rightarrow X + Y \sim Po(\lambda + \mu)(\text{Additivgesetz})$$

Beweis. 1.

$$\mathbb{E}X = \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \lambda \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{= \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}} = \lambda$$

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda^2 e^{\lambda} = \lambda^2$$

$$\begin{aligned} Var(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 \\ &= \mathbb{E}(X(X-1)) + \mathbb{E}X - \mathbb{E}X^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

2. Faltungsformel (Übung!)

□

12.4 Definition und Satz

Sei $0 < p < 1$.

$$X \sim G(p) :\Leftrightarrow \mathbb{P}(X = k) = (1-p)^k \cdot p, \quad k \in \mathbb{N}_0$$

(geometrische Verteilung mit Parameter p)
Es gilt

1. $\mathbb{E}X = \frac{1}{p} - 1$
2. $V(X) = \frac{1-p}{p^2}$

X modelliert <u>Anzahl der Nieten vor dem ersten Treffer</u> in einer unendlichen Bernoulli-Kette mit Trefferwahrscheinlichkeit p.

Beweis.

$$\sum_{k=0}^{\infty} X^k \equiv \frac{1}{1-x} \text{ auf } (-1,1)$$

$$\Rightarrow \sum_{k=1}^{\infty} k \cdot x^{k-1} = \frac{1}{(1-x)^2} \text{ für } |x| < 1$$

$$\Rightarrow \sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

$$\mathbb{E}X = \sum_{k=1}^{\infty} k(1-p)^k \cdot p = p(1-p) \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$p(1-p)\frac{1}{(1-(1-p))^2}$$

$$= \frac{1-p}{p}$$

$$\mathbb{E}X(X-1) = p(1-p)^2 \underbrace{\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2}}_{=\frac{2}{p^3}}$$

$$= 2\left(\frac{(1-p)}{p}\right)^2$$

$$\Rightarrow Var(X) = 2\left(\frac{1-p}{p}\right)^2 + \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2$$

$$= \left(\frac{1-p}{p}\right)^2 + \frac{1-p}{p} = \left(\frac{1-p}{p}\right)\underbrace{\left(\frac{1-p}{p} + 1\right)}_{=\frac{1}{p}} = \frac{1-p}{p^2}$$

$\square$

12.5 Definition und Satz

X hat eine negative Binomialverteilung mit Parametern r und p ($r \in \mathbb{N}$ und $0 < p < 1$), falls gilt:

$$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \quad k \in \mathbb{N}_0$$

Es gilt

$$\mathbb{E}X = r \cdot \frac{1-p}{p}$$
$$V(X) = r \cdot \frac{1-p}{p^2},$$

Notation: $X \sim Nb(r, p)$

12.6 Satz

$X_1, \ldots, X_r \overset{u.i.v.[1]}{\sim} G(p)$

$$\Rightarrow X_1 + \ldots + X_r \sim Nb(r, p)$$

Beweis. Faltungsformel (Übung) □

Bemerkung

1. $X \sim Nb(r, p)$

$$\mathbb{P}(X = k) = \frac{(k+r-1)^{\underline{k}}}{k!} p^r (1-p)^k$$
$$= \frac{-r \cdot (-r-1) \cdot \ldots \cdot (-r-k+1)}{k!} (-1)^k p^r (1-p)^k$$
$$= \binom{-r}{k} \cdot p^r (-(1-p))^k$$

wobei $\binom{-r}{k} = \frac{-r^{\underline{k}}}{k!} \quad \forall r \in \mathbb{R}$

$X \sim Nb(r, p)$, falls

$$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r (1-p)^k$$
$$= \boxed{\binom{-r}{k} p^r (-(1-p))^k}$$

$k = 0, 1, 2, \ldots$ ($\binom{x}{k} := \frac{x^{\underline{k}}}{k!} = \frac{x(x-1)\cdots(x-k+1)}{k!}$) Wkt. für k Fehlversuche bis zum r-ten Erfolg in einer ("unendlichen" Bernoulli-Kette) $r = 1 : \mathbb{P}(X = k) = p(1-p)^k$

[1] unabhängig identisch verteilt

12.7 Bemerkungen

1. Siehe Box oben.
2. $X \sim Nb(r,p), Y \sim Nb(r,p), X$ und Y unabhängig

$$\Leftarrow X + Y \sim Nb(r+s,p)$$

Beweis. Faltungsformel. Inhaltlich folgt das aus der Interpretation der negativen Binomialverteilung. □

12.8 Beispiel (Multinomiales Versuchsschema)

n unabhängige Experimente mit Ausgängen in $\{1,\ldots,s\}$ mit $s \geq 2$. (Ausgang $k \hat{=}$ Treffer der k-ten Art)
p_k = Wkt. für Treffer der k-ten Art
$p_1 + \ldots + p_s = 1$

Modell: $(\Omega, \mathbb{P}) = (\bigotimes_{j=1}^{n} \Omega_j, \bigotimes_{j=1}^{n} \mathbb{P}_j)$
$\Omega = \{1,\ldots,s\}, \mathbb{P}_j(\{k\}) = p_k$

$$A_j^{(k)} := \{\omega = (\omega_1,\ldots,\omega_n) \in \Omega \mid \omega_j = k\}$$

$$X_k := \sum_{j=1}^{n} 1_{A_j^{(k)}} \quad \text{Anzahl der Treffer } k\text{-ter Art}$$

$$X_1 + \ldots + X_s = n$$

Es gilt für $i_1,\ldots,l_s \in \mathbb{N}_0, i_1 + \ldots + i_s = n$

$$\begin{aligned} &|\{X_1 = i_1,\ldots,X_s = i_s\}| \\ &= \binom{n}{i_1}\binom{n-i_1}{i_2}\cdots\binom{n-i_1\ldots-i_{s-1}}{i_s} \\ &= \frac{n!}{i_1!i_2!\ldots i_s!} =: \binom{n}{i_1\ldots i_s} \end{aligned}$$

$(p_1^{i_1} \cdot p_2^{i_2} \cdots p_s^{i_s})$

12.9 Definition

$(X_1,\ldots,X_s) \sim Mult(n,p_1,\ldots,p_s)$, falls

$$\mathbb{P}(X_1 = i_1,\ldots,X_s = i_s) = \binom{n}{i_1 \cdots i_s} p_1^{i_1} \cdots p_s^{i_s}$$

$i_1,\ldots,i_s \in \mathbb{N}_0, i_1 + \ldots + i_s = n$.

12.10 Folgerung

Es gelte $(X_1, \ldots, X_s) \sim Mult(n, p_1, \ldots, p_s)$

1. $X_k \sim Bin(n, p_k)$
2. $X_{i_1} + \ldots + X_{i_0} \sim Bin(n, p_{i_1} + \ldots + p_{i_\nu}), \quad (\{i_1, \ldots, i_\nu\} \subset \{1, \ldots, s\})$
3. $C(X_i, X_j) = -np_ip_j, \quad i \neq j.$
4. $\rho(X_i, X_j) = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}, \quad i \neq j, p_i < 1, p_j < 1.$

Beweis. 1. Ist Spezialfall von (2) $(\nu = 1)$.
Bildung der Randverteilung von $(X_1, \ldots, X_s)$
Benutze Multinomialformel

$$(X_1 + \ldots + X_s)^n = \sum_{j_1, \ldots, j_s} \binom{n}{j_1 \cdots j_s} X_1^{j_1} X_2^{j_2} \cdots X_s^{i_s}$$

$\underline{n = 2}$ $\binom{n}{j_1 j_2} = \binom{n}{j_1} = \binom{n}{j_2}, \quad j_1 + j_2 = n$

2. $B_j := \bigcup_{k=1}^{\nu} A_j^{(i,k)}$ = im j-ten Versuch liegt Erfolg mit Typ in $\{i_1, \ldots, i_\nu\}$ vor
$B_1, \ldots, B_n$ sind unabhängig (!) und es gilt

$$\mathbb{P}(B_j) = p_{i_1} + \ldots + p_{i_\nu} =: q$$

Satz 9.6 $\rightsquigarrow$

$$X_{i_1} + \ldots + X_{i_\nu} \stackrel{d}{=} \sum_{j=1}^{n} 1_{B_j} \sim Bin(n, q)$$

3. $$\begin{aligned} Var(X_i + X_j) &\stackrel{(2)}{=} n(p_i + p_j)(1 - p_i - p_j) \\ &= Var(X_i) + Var(X_j) + 2Cov(X_i, X_j) \\ &= np_i(1 - p) + np_j(1 - p_j) \\ &\rightsquigarrow 2Cov(X_i, X_j) = -2np_ip_j \end{aligned}$$

4. $$\begin{aligned} \rho(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{var(X_i)}\sqrt{Var(X_j)}} &= \frac{-np_ip_j}{\sqrt{np_i(1 - p_i)}\sqrt{np_j(1 - p_j)}} \\ &= -\frac{\sqrt{p_i}\sqrt{p_j}}{\sqrt{1 - p_i}\sqrt{1 - p_j}} \end{aligned}$$

□

Kapitel 13

Bedingte Erwartungswerte und bedingte Verteilungen

$(\Omega, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum, $X\colon \Omega \to \mathbb{R}$ mit $\mathbb{E}(X) < \infty$.

13.1 Definition

1. Für $A \subset \Omega$ mit $\mathbb{P}(A) > 0$ heißt

$$\mathbb{E}[X|A] := \frac{1}{\mathbb{P}(A)} \cdot \sum_{\omega \in A} X(\omega)\mathbb{P}(\{\omega\})$$

   bedingter Erwartungswert von X unter (der Bedingung) A.

2. Ist speziell $A = \{Z = z\}, z \in \mathbb{R}^k$, für einen Zufallsvektor $Z\colon \Omega \to \mathbb{R}^k$, so heißt

$$\mathbb{E}[X|Z = z] := \mathbb{E}[X|\{Z = z\}] = \frac{1}{\mathbb{P}(Z = z)} \cdot \sum_{\omega\colon Z(\omega)=z} X(\omega)\mathbb{P}(\omega)$$

   bedingter Erwartungswert von X unter der Bedingung $Z = z$. (Annahme: $\mathbb{P}(Z = z) > 0$)

13.2 Bemerkungen

1. Sei $\mathbb{P}_A(B) = \mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$. Dann ist $\mathbb{P}_A$ ein Wahrscheinlichkeits-Maß.
   Denn: $\mathbb{E}[X|A] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}_A(\{\omega\})$

2. $\mathbb{E}[X|A]\mathbb{P}(A) = \mathbb{E}X1_A \quad \Big(X = 1_B, \mathbb{E}[X|A] = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \mathbb{P}(B|A)\Big)$

13.3 Beispiel

X_1, X_2 unabhängig, $\mathbb{P}(X_i = k) = \frac{1}{6}, k = 1, \dots, 6$ $(\Omega = \{1, \dots, 6\}^2, \mathbb{P} =$ Gleichverteilung)
$\mathbb{E}[X_1 | X_1 + X_2 = 10] =?$
$A = \{X_1 + X_2 = 10\} = \{(4,6), (5,5), (6,4)\}$
$\mathbb{P}(A) = \frac{3}{36} = \frac{1}{12}$
$\mathbb{E}[X_1 | A] = 12(4 \cdot \frac{1}{36} + 5 \cdot \frac{1}{36} + 6 \cdot \frac{1}{36})$
$= \frac{1}{3}(15) = 5$
Übung: $\mathbb{E}[X_1 | X_1 + X_2 \geq 10]$

13.4 Satz (Formel vom totalen Erwartungswert)

Es gelte $\Omega = \bigcup_j A_j$ mit $A_i \cap A_j = \emptyset$ Für $i \neq j$ und $\mathbb{P}(A_j) > 0, j = 1, 2, \dots$.
Dann

$$\mathbb{E}X = \sum_j \mathbb{E}[X | A_j] \mathbb{P}(A_j)$$

$(X = 1_B \rightsquigarrow$ Formel der totalen Wahrscheinlichkeit)

Beweis. Analog zur Formel der totalen Wahrscheinlichkeit! □

13.5 Beispiel

"Unendliche" Bernoulli-Kette.
Beobachte bis erstmalig "1 1" auftritt. Sei X die Anzahl der Versuche.

$$\mathbb{E}X =?$$

Sei $A_1 = \{$Kette beginnt mit 0$\}$.

$$\mathbb{E}[X | A_1] = 1 + \mathbb{E}X$$

$A_2 := \{$Kette beginnt "1,0"

$$\mathbb{E}[X | A_2] = 2 + \mathbb{E}X$$

$A_3 = \{$Kette beginnt mit "1,1"$\}$

$$\mathbb{E}[X | A_3] = 2$$

13.4 $\rightsquigarrow \mathbb{E}X = (1-p)(\mathbb{E}X + 1) + p(1-p)(2 + \mathbb{E}X) + p^2 \cdot 2$

$$\mathbb{E}X\big(1 - (1-p) - p(1-p)\big) = 1 - p + 2p(1-p) + 2p^2$$

$$(\mathbb{E}X)p^2 = 1 + p$$

Erinnerung $\mathbb{E}[X|A] = \frac{1}{\mathbb{P}(A)} \sum\limits_{\omega \in A} X(\omega)\mathbb{P}(\{\omega\})$
$\mathbb{E}X = \sum\limits_{j} \mathbb{E}[A|A_j]\mathbb{P}(A_j)$

13.6 Satz (Eigenschaften)

$\mathbb{E}|X| < \infty, \mathbb{E}|Y| < \infty, Z\colon \Omega \to \mathbb{R}^n, z \in \mathbb{R}^n$ mit $\mathbb{P}(Z = z) > 0$.

1. $\mathbb{E}[X + Y|A] = \mathbb{E}[X|A] + \mathbb{E}[Y|A]$
2. $\mathbb{E}[aX|A] = a\mathbb{E}[X|A]$
3. $X \leq Y \Rightarrow \mathbb{E}[X|A] < \mathbb{E}[Y|A]$
4. $\mathbb{E}[1_A|B] = \mathbb{P}(A|B)$
5. $\mathbb{E}[X|A] = \sum\limits_{j} x_j \mathbb{P}(X = x_j|A) \qquad \sum\limits_{j} \mathbb{P}(X = x_j) = 1$
6. $\mathbb{E}[X|Z = z] = \sum\limits_{j} x_j \mathbb{P}(X = x_j|Z = z)$
7. $\mathbb{E}[X|Z = z] = \mathbb{E}X$ falls X und Z stochastisch unabhängig

Beweis. 13.2 und Satz 5.4 □

13.7 Satz (Substitutionsformel)

Sei $X\colon \Omega \to \mathbb{R}^n, Z\colon \Omega \to \mathbb{R}^k, g\colon \mathbb{R}^n \to \mathbb{R}^k$ mit $\mathbb{E}|g(X, Z)| < \infty$. Dann gilt für $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z > 0)$

$$\mathbb{E}[g(X, Z)|Z = z] = \mathbb{E}[g(X, z)|Z = z]$$

Beweis.

$$\mathbb{E}[g(X, Z)|Z = z] = \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega \in \Omega\colon\, Z(\omega) = z} g(X(\omega), Z(\omega))\mathbb{P}(\{\omega\})$$

$$= \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega \in \Omega\colon\, Z(\omega) = z} g(X(\omega), z)\mathbb{P}(\{\omega\}) = \text{ rechte Seite der Behauptung.}$$

□

13.8 Beispiel (Würfelwurf)

Falls k auftritt, wird noch k mal gewürfelt.
Sei X die Gesamtaugensumme.
$\mathbb{E}X = ?$

Modell: $\Omega = \{1, \dots, 6\}^7, \mathbb{P} =$ Gleichverteilung
$X_j(\omega) = \omega_j \quad \omega = (\omega_0, \dots, \omega_6) \in \Omega$
$X_0, \dots, X_6$ stochastisch unabhängig. Dann

$$X = X_0 + \sum_{j=1}^{X_0} X_j$$

$$\begin{aligned}
\mathbb{E}[X|X_0 = k] &= \mathbb{E}[X_0 + \sum_{j=1}^{X_0} X_j | X_0 = k] \\
&= \mathbb{E}[k + \sum_{j=1}^{k} X_j | X_0 = k] \\
&= k + \sum_{j=1}^{k} \mathbb{E}[X_j | X_0 = k] \\
&\overset{g}{=} k + \sum_{j=1}^{k} \mathbb{E}X_j = k + k \cdot 3,5 = 4,5k \\
\mathbb{E}X &= \sum_{j=1}^{6} \mathbb{E}[X|X_0 = k]\mathbb{P}(X_0 = k) \\
&= 4,5 \cdot \frac{1}{6} \cdot \sum_{k=1}^{6} k = \frac{9}{2} + \frac{1}{6} \cdot \frac{42}{2} = \frac{63}{4}
\end{aligned}$$

13.9 Definition

Es sei $Z\colon \Omega \to \mathbb{R}^k$ mit $\mathbb{P}(Z = z_j) > 0, j \geq 1, \sum_j \mathbb{P}(Z = z_j) = 1$.
Es sei X Zufallsvariable mit $\mathbb{E}(X) < \infty$. Dann heißt die durch

$$\mathbb{E}[X|Z](\omega) := \begin{cases} \mathbb{E}[X|Z = z_j], & \text{falls } Z(\omega) = z_j \\ 0, & \text{sonst.} \end{cases}$$

definierte <u>Zufallsvariable</u> $\mathbb{E}[X|Z]$ die **bedingte Erwartung** von X bei gegebenem Z.

13.10 Beispiel

In Bsp. 13.8 gilt
$$\mathbb{E}[X|X_0] = \frac{9}{2}X_0$$

13.11 Satz

Sei $\mathbb{E}X^2 < \infty$, $Z\colon \Omega \to \mathbb{R}^k$ (wie in 13.9), $h\colon \mathbb{R}^k \to \mathbb{R}$ mit $\mathbb{E}h(Z)^2 < \infty$. Dann wird
$$\mathbb{E}(X - h(Z))^2$$
minimal für $h(Z) = \mathbb{E}[X|Z]$.

Beweis. Beweis durch Verwirrung und Klammerchaos. (Ohne Gewähr)

$$\mathbb{E}(X - h(Z))^2$$
$$\overset{13.4}{=} \sum_j \mathbb{E}[(X - h(Z))^2|Z = z_j]\mathbb{P}(Z = z_j)$$
$$= \sum_j \mathbb{E}[(X - h(z_j))^2|Z = z_j]\mathbb{P}(Z = z_j)$$
$$= \sum_j (\mathbb{E}[(X - \mathbb{E}[X|Z = z_j]) + (\mathbb{E}[X|Z = z_j] - h(z_j))^2|Z = z_j]\mathbb{P}(Z = z_j)$$

TODO: Rauskriegen, was das heißen soll

$$\sum_j \mathbb{E}[(X - \mathbb{E}[X|Z = z_j]))^2|Z = z_j]\mathbb{P}(Z = z_j)$$
$$2\sum_j \mathbb{E}[(\underbrace{\mathbb{E}[X|Z = z_j] - h(z_j))]}_{\text{Vor den EW ziehen}}(X - \mathbb{E}[X|Z = z_j])|Z = z_j)\mathbb{P}(Z = z_j)$$
$$+\sum_j (\mathbb{E}[X|Z = z_j] - h(z_j))^2\mathbb{P}(Z = z_j)$$

Nebenrechnung: $\mathbb{E}[X - \mathbb{E}[X|Z = z_j]|Z = z_j] = 0$

$$\rightsquigarrow \text{Minimierer: } h(z_j) = \mathbb{E}[X|Z = z_j]$$

$\square$

Kapitel 14

Grenzwertsätze

14.1 Satz (Schwache Gesetz der großen Zahlen, SGGZ)

Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen mit gleicher Verteilung und $\mathbb{E}X_1^2 < \infty$. Sei

$$\bar{X_n} := \frac{1}{n}\sum_{j=1}^{n} X_j$$

Dann gilt

$$\mathbb{P}(|\bar{X_n} - \mathbb{E}X_1| > \epsilon) \to 0$$

$\lim n \to \infty, \epsilon > 0$

Beweis.

$$\mathbb{E}\bar{X_n} = \frac{1}{n}(\mathbb{E}X_1 + \ldots + \mathbb{E}X_n) = \mathbb{E}X_1$$

$$V(\bar{X_n}) = \frac{1}{n^2}V(X_1 + \ldots + X_n) = \frac{1}{n^2}(V(X_1) + \ldots + V(X_n)) = \frac{1}{n}V(X_1)$$

Tschebyschevsche Ungleichung (Satz 11.6)

$$\mathbb{P}(|\bar{X_n} - \mathbb{E}\bar{X_n}| > \epsilon) \leq \frac{V(\bar{X_n})}{\epsilon^2} = \frac{V(X_1)}{n\epsilon^2}$$

$\square$

14.2 Definition

Sind $Y_1, Y_2, Y_3, \ldots$ Zufallsvariablen, so schreibt man

$$Y_n \xrightarrow{\mathbb{P}} Y \text{ für } n \to \infty$$

$$\mathbb{P}(|Y_n - Y| > \epsilon) \to 0 \text{ für } n \to \infty \quad \forall \epsilon > 0.$$

(SGGZ: $\bar{X_n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1$)

Bemerkung Auf einem diskreten Wahrscheinlichkeitsraum gibt es keine unendlichen Folgen von unabhängigen Zufallsvariablen. Lösung:

$$\mathbb{P} \to \mathbb{P}_n$$

$\rightsquigarrow$ Wahrscheinlichkeitstheorie-Vorlesung

14.3 Folgerung (SGGZ von Jakob Bernoulli)

Seien $A_1, A_2, \ldots$ unabhängige Ereignisse mit $\mathbb{P}(A_j) = p, j \geq 1$. Dann

$$R_n = \frac{1}{n} \sum_{j=1}^{n} 1_{A_j} \xrightarrow{\mathbb{P}} p$$

Bemerkung $\forall \epsilon > 0 \forall \delta \in (0,1) \exists n \in \mathbb{N} \colon \mathbb{P}(|R_n - p| \leq \epsilon) \geq 1 - \delta$
Das bedeutet nicht

$$\mathbb{P}(\bigcap_{n>=n_0} \{|R_n - p| \leq \epsilon\}) \geq 1 - \delta!$$

$\rightsquigarrow$ Starkes GGZ (Wahrscheinlichkeitstheorie)

[1] Es seien $X_1, X_2, \ldots$ unabhängige $Bern(p)$-verteilte Zufallsvariablen. Setze

$$S_n := X_1 + \ldots + X_n, n > 1 \quad \text{Irrfahrt, } X_i \in \{-1, 1\} \text{ besser!}$$

Frage Wie stark schwankt S_n um seinen Erwartungswert np? (TODO: Bild)

14.4 Satz

Es sei (a_n) eine reelle Zahlenfolge. Dann gilt

$$\lim_{n \to \infty} \mathbb{P}(|S_n - np| \leq a_n) = \begin{cases} 1, & \text{falls } \frac{a_n}{\sqrt{n}} \to \infty \\ 0 & \text{falls } \frac{a_n}{\sqrt{n}} \to 0 \end{cases}$$

Wiederholung $S_n := X_1 + \ldots + X_n, X_i$ unabhänbh. $Bern(p) = \underbrace{Bin(1,p)}_{\text{entweder } p \text{ oder } 1-p}$
(TODO: Bild)
absolute Häufigkeiten!

[1] Random Walk

Satz

$\lim_{n\to\infty} \mathbb{P}(|S_n - np| \leq a_n) = \begin{cases} 1 & \frac{a_n}{\sqrt{n}} \to \infty \\ 0 & \frac{a_n}{\sqrt{n}} \to 0 \end{cases}$

Beweis. Sterlingsche Formel (Ausblick/Exkurs)
$n! = \sqrt{2\pi n} n^n e^{-n}(1 + R(n))$ mit $R(n) \to 0$ für $n \to \infty, n \in \mathbb{N}$
Genauer $R(n) = e^{\eta(n)} - 1, 0 \leq \eta(n) \leq \frac{1}{12n}$
Insbesondere $n! \sim \sqrt{2\pi n} n^n e^{-n}$ für $n \to \infty$
$((c_n) \sim (d_n) \Leftrightarrow \frac{c_n}{d_n} \to 1)$

Nun:

- Tschebyschevsche Ungleichung
  $\mathbb{P}(|S_n - np| \leq a_n) = 1 - \mathbb{P}(|S_n - np| > a_n) \geq 1 - \frac{V(S_n)}{a_n^2} = 1 - \frac{np(1-p)}{a_n^2}$
  Im Fall $a_n/\sqrt{n} \to \infty$ strebt das gegen 1.

- Für den anderen Fall setzen wir $p = \frac{1}{2}$ (der Einfachheut halber!)
  $\mathbb{P}(|S_n - np| \leq a_n) = \sum_{n:\, |K-\frac{n}{2}| \leq a_n} \underbrace{\binom{n}{k}}_{\text{max. wenn } k \text{ halb so groß wie } n} \underbrace{2^{-n}}_{\text{Konstante}} \leq$
  $\underbrace{(2a_n + 1)\binom{n}{\lfloor \frac{n}{2} \rfloor}}_{\text{nach unten abgeschätzt}} 2^{-n}$
  Für $n = 2m$ gilt $\binom{2m}{m} \cdot 2^{-n} = \frac{(2m)!}{m!m!} \cdot 2^{-2m} \overset{\text{Asymm., äquiv.}}{\approx}$
  $\frac{\sqrt{2\pi \cdot 2m}(2m)^{2m}e^{-2m}}{2\pi m \cdot m^n \cdot m^m e^{-m} e^{-m}} \cdot 2^{-2n} \approx \frac{1}{\sqrt{\pi m}}$
  Es folgt

$$\limsup_{m\to\infty} \mathbb{P}(|S_{2m} - 2mp| \leq a_{2m}) \leq \limsup_{m\to\infty} \frac{2a_{2m} + 1}{\sqrt{\pi} \cdot \sqrt{m}}$$

$$= \frac{\sqrt{2}2}{\sqrt{\pi}} \lim_{m\to\infty} \frac{a_{2m}}{\sqrt{2m}} = 0, \text{ falls } \frac{a_n}{\sqrt{n}} \to 0$$

  Analog für $m = 2n + 1$!

□

Fazit: Die Abweichungen von S_n vom Erwartungswert weichen "typischerweise" wie $\sqrt{n}$.

Beachte: S_n^{*} [2] $:= \frac{S_n - np}{\sqrt{np(1-p)}}, \mathbb{E}S_n^* = 0, V(S_n^*) = 1$

[2] standardisiert

Fragen: $\mathbb{P}(|S_n^*| \leq a) \underset{n\to\infty}{\to} ? \quad (a \in \mathbb{R})$

14.5 Definition

1. Sei $l(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, x \in \mathbb{R}$ ($l(x)$ = Gaußsche Glockenkurve) heißt **Dichte der standardisierten Normalverteilung**.
2. $\phi(x) := \int_{-\infty}^{\infty} l(t)dt, x \in \mathbb{R}$ [3] heißt **Verteilungsfunktion der standardisierten Normalverteilung**.

Bemerkung

$\int_{-\infty}^{\infty} l(t)dt = 1$ (Analysis 3).
Dann ist $\lim_{x\to\infty} \phi(x) = 1$
(TODO: Bild)

14.6 Satz

Immernoch in der Bernoullikette gilt

1. $\lim_{n\to\infty} \mathbb{P}(a \leq S_n^* \leq b) = \underbrace{\phi(b)}_{=\int_a^b l(t)dt} - \phi(a)$ und $\lim_{n\to-\infty} \mathbb{P}(\ldots) = 0$
2. $\lim_{n\to\infty} \mathbb{P}(S_n^* \leq b) = \phi(b) \quad \to$ (Moivre-Laplace)

Beweis. Sterlingsche Formel! (2 Seiten Rechnung) $\rightsquigarrow$ Wahrscheinlichkeitstheorie-Vorlesung im Sommersemester
(TODO: Bild) □

14.7 Satz (ZGWS Lindeberg-Levy)

Seien $X_1, X_2, \ldots$ unabhängig auf gleicher Verteilung mit $\mathbb{E}X_1^2 < \infty$. Setze

$$S_n := X_1 + \ldots + X_n, \quad n \in \mathbb{N}$$

$$S_n^* = \frac{S_n - n\mathbb{E}X_1}{\sqrt{n \cdot V(X_1)}}, \quad n \in \mathbb{N}$$

Dann gilt $lim_{n\to\infty}\mathbb{P}(S_n^y \leq h) = \phi(h), \quad h \in \mathbb{R}$

Beweis. Wahrscheinlichkeitstheorie im Sommersemester. □

[3] uneigentliches Riemann-Integral

Anwendung: Sei $S_n \sim Bin(n,p)$.

Nun sei $\mathbb{P}(a \leq S_n \leq b) = \frac{a-np^{+\frac{1}{2}}}{\sqrt{np(1-p)}} \leq \underbrace{\frac{S_n-np}{\sqrt{np(1-p)}}}_{S_n^*} \leq \frac{b-np^{+\frac{1}{2}}}{\sqrt{np(1-p)}} =$
$\phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$

erst standardisieren, dann approximieren

Stetigkeitskorrekturen

$\limsup: \quad n = 600, p = \frac{1}{6} \quad a = 90, b = 110$

Exakter Wert: 0.7531

Approx. Wert: 0.7264 mit Stetigkeitskorrektur 0.7498

Kapitel 15

Statistik - Schätzprobleme

15.1 Schätzung der Erfolgswahrscheinlichkeit einer Bernoulli-Kette

(als Motivation)

k-Treffer in einer B-Kette im Umfang n mit <u>unbekannter</u> Erfolgswahrscheinlichkeit p.
Wie groß ist p?

Modell Annahme einer zufälligen Trefferzahl mit $S_n \sim Bin(n, p)$ [1].
Ereignis $S_n = k$ sei eingetreten.

$$\mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

[2]

Plausibler "Schätzer" für p:

$$\hat{p} := \frac{k}{n} \text{ (relative Häufigkeit)}$$

$\hat{p}$ ist Realisierung einer Zufallsvariable $R_n := \frac{1}{n} S_n$

1.

2.

(TODO: Rest texen)

---

[1] auch ohne es zu kennen
[2] abhängig vom Parameter p

Erinnerung $S_n \sim Bin(n,p) \quad (S_n := X_1 + \ldots + X_n)$

$$\mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$S_n = k \text{ Realisierung}$$

$$p = ?$$

natürliche Idee: $\hat{p} = \frac{k}{n}$
$\hat{p}$ ist Realisierung der Zufallsvariable $R_n = \frac{1}{n} S_n$
Die Funktion

$$p \mapsto L_k(p) = \mathbb{P}(S_n = k)$$

heißt **Likelyhood-Funktion**.
$\hat{p}$ maximiert L_k!

15.2 Allgemeiner Modellrahmen

Daten $x \in \mathcal{X}$ = "Stichprobenraum" gegeben.
Oft: $x = (x_1, \ldots, x_n) \in \mathbb{R}^n \quad (\mathcal{X} \subset \mathbb{R}^n)$

Annahme: $x = X(\omega)$ Realisierung einer $\mathcal{X}$-wertigen Zufallsvariable $X : \Omega \to \mathcal{X}$

Gesucht: Verteilung $\mathbb{P}^X$ von X!
Ω spielt "keine Rolle".
Wichtig ist das (unbekannte) $\mathbb{P}^X$.

Oft: $\Omega = \mathcal{X}, X = id_\Omega, \mathbb{P} = \mathbb{P}^X$

Annahme: $\mathbb{P} \in \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ wobei $\Theta \neq \emptyset$ Parameterraum (A priori-Informationen!)
$(\mathcal{X}, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ heißt **statistisches Modell**.
Falls $\mathbb{P} = \mathbb{P}_\vartheta$, so schreiben wir $\mathbb{P}_\vartheta(X = x), \mathbb{E}_\vartheta X, V_\vartheta(X), \ldots$

15.3 Beispiel (Binomialverteilung)

$\mathcal{X} = \{0,1\}^n, \quad X = (X_1, \ldots, X_n), \quad X_1, \ldots, X_n$ unabhängig, $Bern(p)$[3]
$\vartheta \in \Theta = (0,1)$
$\mathbb{P}_\vartheta(X = (x_1, \ldots, x_n)) = \vartheta^k (1-\vartheta)^{n-k}, \quad k = x_1 + \ldots + x_n.$
[4]
$\mathbb{P}_\vartheta(X_1 + \ldots + X_n = k) = \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}$

[3] $= Bin(1,p)$
[4] Unabhängig verteilt: $\mathbb{P}(X_1 = i_1, \ldots, X_n = i_n) = \mathbb{P}(X_1 = i_1) \cdot \ldots \cdot \mathbb{P}(X_n = i_n)$

15.4 Beispiel

Warensendung vom Umfang N mit ϑ defekten und $N-\vartheta$ intakten Exemplaren. n-maliges Ziehen ohne Zurücklegen.

$$X_j := \begin{cases} 1 & j-\text{te entnommene Exemplar ist defekt} \\ 0 & \text{sonst} \end{cases}$$

$$X = (X_1, \dots, X_n) \quad (n \leq N)$$

$$\Theta = \{0, \dots, N\}, \quad \vartheta \hat{=} \text{ Anzahl der defekten Exemplare}$$

$$\mathbb{P}_\vartheta(X = (x_1, \dots, x_n)) = (*) \quad k = \sum_{i=1}^{n} x_i$$

(*) $\frac{\vartheta}{N} \cdot \frac{\vartheta-1}{N-1} \cdot \ldots \cdot \frac{\vartheta-k+1}{N-k+1} \cdot \frac{N-\vartheta}{(N-k)} \cdots \frac{N-\vartheta-(n-k)+1}{(N-k+1)} = \frac{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}}{N^{\underline{n}}}$

$X_1 + \ldots + X_n \sim Hyp(\underbrace{n, N}_{\text{fix}} - \vartheta, \underbrace{\vartheta}_{=?})$ Qualitätskontrolle

15.5 Definition

Sei $(\mathcal{X}, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ ein statistisches Modell.
Ein (Punkt)**Schätzer** (für den unbekannten Parameter ϑ) ist eine Abbildung

$$T\colon \mathcal{X} \to \widetilde{\Theta} \supset \Theta$$

(oft $\widetilde{\Theta} = \Theta$)
$T(x)$ heißt **konkreter Schätzwert**[5].

15.6 Bemerkungen

1. Allgemein heißt eine Abbildung auf $\mathcal{X}$ **Stichprobenfunktion**.

2. $\mathbb{P}_\vartheta$ "steuert" Auftreten von x und damit $T(x)$. T ist eine auf $\mathcal{X}$ definierte Zufallsvariable! Ist T eine Schätzfunktion, so definiert

$$\mathbb{P}_\vartheta(T = t) := \mathbb{P}_\vartheta(\{x \in \mathcal{X} : T(x) = t\})$$

Wünschenswert: $\mathbb{P}_\vartheta(T = \vartheta) \approx 1$[6]$, \vartheta \in \Theta$
$\mathbb{P}_\vartheta(|T - \vartheta| \leq 1) \approx 1, \vartheta \in \Theta$

[5] auf der Grundlage der Daten
[6] nicht erreichbar

15.7 Definition

Sei $T\colon \mathcal{X} \to \widetilde{\Theta} \supset \Theta \quad (\widetilde{\Theta} \subset \mathbb{R})$. Dann

$$\mathbb{E}_\vartheta T^2 = \sum_{x \in \underbrace{\mathcal{X}}_{\text{diskret!}}} T(x)^2 \mathbb{P}_\vartheta(\{x\}) < \infty, \quad \vartheta \in \widetilde{\Theta} \subset \mathbb{R}$$

1. $MQA_T(\vartheta) := \mathbb{E}_\vartheta(T - \vartheta)^2 = \sum_{x \in \mathcal{X}} (T(x) - \vartheta)^2 \mathbb{P}_\vartheta(\{x\})$ heißt **mittlere quadratische Abweichung** von T an der Stelle ϑ.

2. $b_T(\vartheta) = \mathbb{E}_\vartheta T - \vartheta$ heißt **Verzerrung** (Bias) von T an der Stelle ϑ. Gilt $b_T(\vartheta) = 0$, so heißt T **erwartungstreu**.[7]

15.8 Bemerkung

Es gilt $MQA_T(\vartheta) = V_\vartheta(T) + b_T(\vartheta)^2$ (Steiner-Formel)
Wünschenswert: $MQA_T(\vartheta) \leq MQA_{T^*}(\vartheta), \quad \forall \vartheta \forall T^*$
(*) ist nicht erfüllbar, da $T^*(x) = \vartheta^*, \quad x \in \mathcal{X} \quad (\vartheta^* \in \Theta)$
Es gilt $V_\vartheta(T^*) = 0$ [8] und deswegen:

$$b_T(\vartheta)^2 = (\vartheta^* - \vartheta)^2$$

(TODO: Bild: Parabel)

15.9 Definition

Gegeben sei eine Folge $(\mathcal{X}_n, \{\mathbb{P}_\vartheta^{(n)} : \vartheta \in \Theta\})$ von statistischen Modellen. Die Zufallsvariablen seinen von der Form $X^{(n)}$ ($n \hat{=}$ Stichprobenumfang).
Eine Folge $T_n\colon \mathcal{X}_n \to \widetilde{\Theta}$ heißt **Schätzfolge**.

1. (T_n) heißt **konsistent**, wenn $\mathbb{P}_\vartheta^{(n)}(|T_n - \vartheta| \geq \epsilon) \underset{n \to \infty}{\to} 0$.

2. (T_n) heißt **asymptotisch erwartungstreu**, wenn $\lim_{n \to \infty} \mathbb{E}_\vartheta T_n = \vartheta, \quad \vartheta \in \Theta$.

15.10 Beispiel

$X_1, \ldots, X_n$ unabhängig, $Bern(p)$. $T_n := (x_1, \ldots, x_n) = \frac{1}{n}(x_1 + \ldots + x_n)$ $(\mathbb{E}_\vartheta T_n = \vartheta)$ ist konsistent!

[7] In dem Fall ist die mittlere quadratische Abweichung nichts anderes als die Varianz.
[8] Zufallsvariable konstant, Varianz 0

Wiederholung $(\mathcal{X}, \{\mathbb{P}_\vartheta \colon \vartheta \in \Theta\})$ statistisches Modell
$x = X(\omega)$ Beobachtung (Daten)
$T \colon \mathcal{X} \to \widetilde{\Theta} \supset \Theta$ Schätzfunktion
Idee: $T(x)$ schätzt den unbekannten Parameter
$MQA_T(\vartheta) := \mathbb{E}_\vartheta(T - \vartheta)^2 \quad (\Theta \subset \mathbb{R})$

Bemerkung 15.8:

$$MQA_T(\vartheta) = V_\vartheta(T) + b_T(\vartheta)^2$$

nicht erfüllbarer Wunsch für T: $MQA_T(\vartheta) \leq MQA_{T^*}(\vartheta), \quad \forall \vartheta \forall T^*$

15.11 Definition

$(\mathcal{X}, \{\mathbb{P}_\vartheta \colon \vartheta \in \Theta\})$ statistisches Modell
$x \in \mathcal{X}$. Definiere $L_x \colon \Theta \to [0,1]$

$$L_x(\vartheta) = \mathbb{P}_\vartheta(X = x), \quad \vartheta \in \Theta.$$

Diese Funktion heißt **Likelihood-Funktion** zu gegebenen x. Existiert ein $\hat{\vartheta}(x) \in \Theta$ (manchmal: $\widetilde{\vartheta} \in \widetilde{\Theta}$) mit

$$L_x(\hat{\vartheta}(x)) = \sup_{\vartheta \in \Theta} L_x(\vartheta) \quad (*)$$

so heißt $\hat{\vartheta}(x)$ ML-Schätzwert (von ϑ) zu x.
Ein Schätzer $\hat{\vartheta} \colon \mathcal{X} \to \Theta$ mit (*) für jedes $x \in \mathcal{X}$ heißt ML-Schätzer. [9]

15.12 Beispiel

$L_\vartheta(x) = \vartheta^k (1-\vartheta)^{n-k}$
$X = (x_1, \ldots, x_n), k = \sum_{i=1}^{n} x_i$
$\hat{\vartheta}(x) = \frac{k}{n}$

15.13 Beispiel (vgl. 15.4)

$\mathbb{P}_\vartheta(X = x) = L_x(\vartheta)$

$$= \frac{\vartheta^{\underline{k}} (N - \vartheta)^{\underline{n-k}}}{N^{\underline{n}}}$$

ML-Schätzer? $x_1 + \ldots + x_n = k$

1. $k = 0 \quad L_x(\vartheta) = \frac{(N-\vartheta)^{\underline{n}}}{N^{\underline{n}}} \rightsquigarrow \hat{\vartheta}(x) = 0.$

[9] Zu geg. Datum wähle mir den Parameterwert, welcher das Datum am wahrscheinlichsten macht.

2. $k = n \quad L_x(\vartheta) = \frac{\vartheta^n}{N^{\underline{n}}} \rightsquigarrow \hat{\vartheta}(x) = N$

3. $0 < k < n$:

$$\frac{L_x(\vartheta+1)}{L_x(\vartheta)} \overset{!}{=} \frac{\vartheta+1}{\vartheta-k+1}\frac{N-\vartheta+k-n}{N-\vartheta} > 1$$

$$\Leftrightarrow \vartheta < \frac{k(N+1)}{n} - 1$$

$$\Rightarrow \hat{\vartheta}(x) : \begin{cases} \lfloor \frac{k(N+1)}{n} \rfloor & \text{falls } \frac{k(N+1)}{n} \notin \mathbb{N} \\ \in \{\frac{k(N+1)}{n} - 1, \frac{k(N+1)}{n}\} & \text{sonst} \end{cases}$$

Vergleich mit heur. Schätzer:

$$\frac{k}{n} = \frac{\vartheta}{N} \Rightarrow \hat{\vartheta} = \frac{Nk}{n}$$

Kapitel 16

Konfidenzbereiche

16.1 Beispiel

Sei $S_n \sim Bin(n,p), \quad p \in [0,1]$ unbekannt.

$$\mathbb{E}_p S_n = np, \quad V_p(S_n) = np(1-p)$$

Für $R_n = \frac{1}{n} S_n$

$$\mathbb{E}_p R_n = p, \quad V_p(R_n) = \frac{p(1-p)}{n} \overset{!}{\leq} \frac{1}{4n}$$

Sei $\alpha \in (0,1) \quad$ (z.B. $\alpha = 0.05$). Dann:

$$\mathbb{P}_p(R_n - \frac{1}{2\sqrt{\alpha n}} \leq p \leq R_n + \frac{1}{2\sqrt{\alpha n}})$$

$$= \mathbb{P}_p(|R_n - p| \leq \frac{1}{2\sqrt{\alpha n}}) \geq 1 - \frac{V_p(R_n)}{(\frac{1}{2\sqrt{\alpha n}})^2} = 1 - \frac{p(1-p)4\alpha n}{n} \geq 1 - \alpha$$

[1]

Mit

$$I_n := [\max\{0, R_n - \frac{1}{2\sqrt{\alpha n}}, \min\{1, R_n + \frac{1}{2\sqrt{\alpha n}}\}]$$

gilt

$$\mathbb{P}_p(I_n \ni p) \geq 1 - \alpha \quad (\text{z.B. } \geq 0.95)$$

16.2 Definition

Sei $(\mathcal{X}, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ statistisches Modell
$\alpha \in (0,1)$. Eine Abbildung

$$C : \mathcal{X} \to \mathcal{P}(\Theta)$$

[1] Hier wurde die Tschebyschew- Ungl. benutzt.

heißt **Konfidenzbereich** (für ϑ) zum Konfidenzwert $1-\alpha$, falls

$$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon C(x) \ni \vartheta\}) \geq 1-\alpha \quad \vartheta \in \Theta$$

Falls $\Theta \subset \mathbb{R}$ und $C(x)$ Intervall ($\forall x$), so heißt C **Konfidenzintervall**.

16.3 Bemerkungen

1. C ist $\mathcal{P}(\Theta)$-wertige Zufallsvariable.
2. $C(x) = \Theta, x \in \mathcal{X}$, ist ein Konfidenzbereich. Sinnlos! (Ziel: $C(x)$ "klein")
3. $C(x)$ konkreter Konfidenzbereich. Die Tatsache, dass $\vartheta \in C(x)$ gilt, hat keine Wahrscheinlichkeit.
4. Es sei $\Theta \subset \mathbb{R}$ und $\beta \in (0,1)$. Dann heißt
   - $$O \colon \mathcal{X} \to \Theta$$
     obere Konfidenzschranke (für ϑ) zum Konfidenzniveau $1-\beta$, falls
     $$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon \vartheta \leq O(x)\}) \geq 1-\beta$$
   - $$U \colon \mathcal{X} \to \Theta$$
     untere Konfidenzschranke (für ϑ) zum Konfidenzniveau $1-\beta$, falls
     $$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon U(x) \leq \vartheta\}) \geq 1-\beta$$

16.4 Beispiel (Konfidenzschranken für p in $Bin(n,p)$)

Sei $S_n \sim Bin(n,p)$. Gesucht $p_u, p_o \colon \{0,\dots,n\} \to [0,1]$ mit

$$\mathbb{P}_p(p \leq p_o(S_n) \geq 1-\beta), \quad p \in [0,1] \quad (2)$$

$$\mathbb{P}_p(p_u(S_n) \leq p) \geq 1-\beta), \quad p \in [0,1] \quad (3)$$

1. $k=0: \quad p_u(0) := 0 \quad p_o(0) := 1-\beta^{\frac{1}{n}}$
   Dann gilt $\beta = (1-p_o(0))^n \geq (1-p)^n$, falls $p \geq p_o(0)$
2. $k=n: \quad p_o(n) := 1, \quad p_u(n) := \beta^{\frac{1}{n}}$
   Dann gilt $p_u(n)^n = \beta \geq p^n, \quad p \leq p_u(n)$
3. $0<k<n: \quad p_o(k)$ ist Lösung von $P(S_n \leq k) = \beta$

Behauptung Dann gilt (2)!

Wiederholung $(\mathcal{X}, \{\mathbb{P}_\vartheta \colon \vartheta \in \Theta\}), \quad x \mapsto C(x) \subset \Theta$
$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon C(x) \ni \vartheta\}) \geq 1 - \alpha$
$\Theta \subset \mathbb{R}:$
$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon \vartheta \leq O(x)\}) \geq 1 - \beta$ obere Konfidenzschranke
$\mathbb{P}_\vartheta(\{x \in \mathcal{X} \colon U(x) \leq \vartheta\}) \geq 1 - \beta$ untere Konfidenzschranke

Beispiel

$S_n \sim Bin(n,p)$ [2], $S_n = k$

$$(*) \quad \mathbb{P}_p(S_n \leq k) = \beta$$

(stetige Funktion von p) [3]
$p_o(k)$ Lösung von (*)
$p_u(k)$ Lösung von

$$\mathbb{P}_p(S_n \geq k) = \beta \quad (**)$$

Es gelte $p \in (0,1)$. Setze

$$M := \{j \in \{0, \dots, n\} \colon p > p_o(j)\}$$

zu zeigen: $\mathbb{P}_p(p > p_o(S_n)) \leq \beta$,! d.h.

$$\mathbb{P}_p(S_n \in M) \leq \beta.$$

Nebenrechnung $M \neq \emptyset$
$k_0 := \max M$
$M \subset \{0, \dots, k_0\}$

Es folgt

$$\mathbb{P}_p(S_n \in M) \leq \mathbb{P}_p(S_n \leq k_0)$$

[4] Fällt in p! (Übung)
$p_o(k_0) < p$

$$\leq \mathbb{P}_{p_o(k_0)}(S_n \leq k_0) \underset{*}{=} \beta$$

[5]

Analog folgt

$$\mathbb{P}_p(p < p_u(S_n)) \leq \beta.$$

[2] n bekannt, p unbekannt
[3] β ist vorgegeben
[4] Wegen der Monotonie der Wahrscheinlichkeitsmaße
[5] benutzt: Monotonie und Tatsache, dass k in M ist

16.5 Allgemeines Konstruktionsprinzip

Wähle zu jedem $\vartheta \in \Theta$ eine Menge $A(\vartheta) \subset \mathcal{X}$ mit

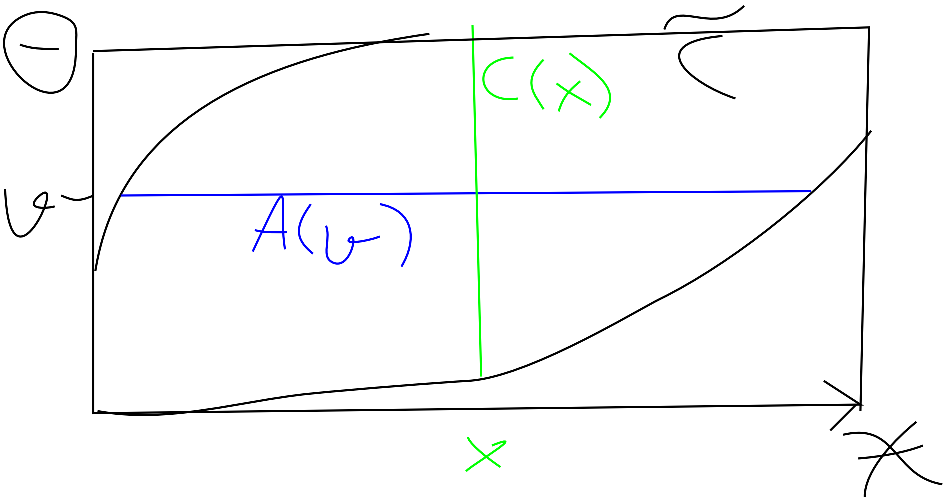
$$\mathbb{P}_\vartheta(A(\vartheta)) \geq 1 - \alpha$$

Setze

$$C(x) := \{\vartheta \in \Theta \colon x \in A(\vartheta)\}$$

[6] Dann gilt

$$x \in A(\vartheta) \Leftrightarrow \vartheta \in C(x)$$



$$\widetilde{C} := \{(x, \vartheta) \colon x \in A(\vartheta)\}$$

Ziel: $C(x)$ "möglichst klein", d.h. $A(\vartheta)$ möglichst klein

16.6 Bemerkung

Setzt man in 16.4

$$\mathcal{X} = \{0, \ldots, n\}, \Theta = [0, 1]$$
$$A(\vartheta) := \{x \in \mathcal{X} \colon u(\vartheta) \leq x \leq o(\vartheta =\},$$

wobei

$$u(\vartheta) := \max \left\{ k \in \mathcal{X} \colon \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \leq \frac{\alpha}{2} \right\},$$

$$o(\vartheta) := \min \left\{ k \in \mathcal{X} \colon \sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \leq \frac{\alpha}{2} \right\}$$

Dann folgt 16.5 auf den in 16.4 angegebenen Konfidenzbereich $[p_u(k), p_o(k)]$.

[6] Konfidenzbereich

16.7 Definition

Betrachte Folge $(\mathcal{X}_n, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ von stochastischen Modellen. Eine Folge $C_n : \mathcal{X}_n$ heißt **asymptotischer Konfidenzbereich zum Niveau** $1-\alpha$, falls

$$\lim_{n\to\infty} \mathbb{P}_\vartheta(\{x \in \mathcal{X}_n \ : C_n(x) \ni \vartheta\}) \geq 1-\alpha$$

16.8 $(Bin(n,p), S_n := X_1 + \ldots + X_n)$

$X_1 + \ldots + X_n$ unabhängig, $Bin(1,p) = Bern(p)$ verteilt.

$$R_n := \frac{1}{n} S_n.$$

Dann gilt für $h > 0$

$$\mathbb{P}_\vartheta\left(-h \leq \frac{S_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \leq h\right) \underset{n\to\infty}{\overset{ZGWS}{\to}} \phi(h) - \phi(-h) = 2\phi(h) - 1$$

[7]

$$\begin{aligned} A_n(\vartheta)\text{"} &\Leftrightarrow \text{"}\frac{|S-n-n\vartheta|}{\sqrt{n\vartheta(1-\vartheta)}} \leq h \\ &\Leftrightarrow \sqrt{n}|R_n - n\vartheta| \leq h\sqrt{\vartheta(1-\vartheta)} \\ &\Leftrightarrow (n+h^2)\vartheta^2 - (2nR_n + h^2)\vartheta + nR_n^2 \leq 0 \\ &\Leftrightarrow U_n \leq \vartheta \leq O_n, \end{aligned}$$

wobei

$$O_n = \frac{R_n + \frac{h^2}{2n} + \frac{h}{\sqrt{n}}\sqrt{R_n(1-R_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}$$

$$U_n = \frac{R_n + \frac{h^2}{2n} - \frac{h}{\sqrt{n}}\sqrt{R_n(1-R_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}$$

Damit ist

$$\mathbb{P}_\vartheta(U_n \leq \vartheta \leq O_n) \underset{n\to\infty}{\to} 2\phi(h) - 1$$

$h : 2\phi(h) - 1 = \alpha$

[7] nach dem Zentralen Grenzwertsatz

Kapitel 17

Testtheorie: fällt weg

Kapitel 18

Allgemeine Modelle

Bisher: Ω diskret, $\mathbb{P}\colon \mathcal{P}(\Omega) \to [0,1]$

Aber: Im Fall $\Omega = \mathbb{R}^k$ ist die Wahl $\mathcal{P}(\Omega)$ keine gute Idee! Zum Beispiel gibt es dann[1] kein Wahrscheinlichkeitsmaß (vgl. unten folgende Definition) mit

$$\mathbb{P}([a,b]) = b - a \quad 0 \le a \le b \le 1$$

18.1 Definition

Sei $\Omega \neq \emptyset$. Eine Menge $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt **σ-Algebra (über Ω)**, falls

1. $\emptyset \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c = \Omega \backslash A \in \mathcal{A}$ [2]
3. $A_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

18.2 Bemerkungen

$\mathcal{A}$ sei σ-Algebra

1. $\Omega \in \mathcal{A}$ $(\emptyset = \Omega^c \in \mathcal{A})$
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
   $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$ [3]
3. $A_n, B_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \bigcap_{n=1}^{\infty} A_n = (\bigcup A_n^c)^c \in \mathcal{A}$

[1] im Fall $k = 1$

[2] Abgeschlossenheit unter Komplementbildung

[3] Eine σ-Algebra ist abgeschlossen unter abzählbaren Vereinigungen und Durchschnitten.

4. $\mathcal{P}(\Omega)$ ist eine σ-Algebra.
   Wie auch $\{\emptyset, \Omega\}$.

18.3 Lemma

Seien $I \neq \emptyset, \mathcal{A}_i$ σ-Algebren über $\Omega, i \in I$.
Dann ist

$$\bigcap_{i \in I} \mathcal{A}_i$$

eine σ-Algebra.

Beweis. klar! □

18.4 Satz und Definition

Sei $\mathcal{M} \subset \mathcal{P}(\Omega)$. Dann existiert genau eine σ-Algebra $\sigma(\mathcal{M})$ über Ω mit

1. $\mathcal{M} \subset \sigma(\mathcal{M})$
2. Für alle σ-Algebren $\mathcal{A}$ mit $\mathcal{M} \subset \mathcal{A}$, dann ist $\sigma(\mathcal{M}) \subset \mathcal{A}$.

Man nennt $\sigma(\mathcal{M})$ die **von $\mathcal{M}$ erzeugte σ-Algebra** und $\mathcal{M}$ **Erzeuger von** $\sigma(\mathcal{M})$.

Beweis.

$$\bigcap_{\mathcal{A}\ \sigma\text{-Algebren}, \mathcal{M} \subset \mathcal{A}} \mathcal{A}$$

q.e.d. (Lemma 18.3) □

18.5 Folgerung

1. $\mathcal{M}_1 \subset \mathcal{M}_2 \Rightarrow \sigma(\mathcal{M}_1 \subset \sigma(\mathcal{M}_2))$
2. $\sigma(\sigma(\mathcal{M})) \subset \sigma(\mathcal{M})$
3. $\mathcal{M}_1 \subset \sigma(\mathcal{M}_2)$ und $M_2 \subset \sigma(\mathcal{M}_1) \Rightarrow \sigma(\mathcal{M}_1) = \sigma(\mathcal{M}_2)$ Übungsaufgabe

18.6 Folgerung und Definition

Seien $k \in \mathbb{N}$ und

$$\mathcal{O}^k := \{A \subset \mathbb{R}^n \colon A \text{ offen } \},$$

$$\mathcal{A}^k := \{A \subset \mathbb{R}^n \colon A \text{ abgeschlossen } \},$$

$$\mathcal{K}^k := \{A \subset \mathbb{R}^n \colon A \text{ kompakt } \},$$

$$\mathcal{I}^k := \{(x, y]: x, y \in \mathbb{R}^k, x < y\}$$

[4] wobei $x \leq y := x_i \leq y_i, \quad i = 1, \ldots, k,$
$x < y := x_i < y_i, \quad i = 1, \ldots, k$

$$\mathcal{B}^k := \sigma(\mathcal{O}^k)$$

heißt **σ-Algebra der Borelmengen** [5] Es gilt

1. $\mathcal{B}^k = \sigma(\mathcal{A}^k)$
2. $\mathcal{B}^k = \sigma(\mathcal{K}^k)$
3. $\mathcal{B}^k = \sigma(\mathcal{I}^k)$

Wiederholung $\Omega \neq \emptyset, \mathcal{A} \subset \mathcal{P}(\Omega)$ σ-Algebren

$$\mathcal{A} = \sigma(\mathcal{M}) := \bigcap_{\dashv' \sigma - Algebra, \dashv' \supset \mathcal{M}} \dashv'$$

$$\mathcal{B}^k := \sigma(\{U \subset \mathbb{R}^k : U \text{ offen})$$

Borelsche σ-Algebra

1. $\mathcal{B}^k = \sigma(\mathcal{A}^k)$

   Beweis. Sei $A \subset \mathbb{R}^k$ abgeschlossen. Dann ist $\mathbb{R}^k \backslash A$ offen, also in B^k. Also ist $A \in \mathcal{B}^k$, d.h. $\mathcal{A}^k \subset \mathcal{B}^k$. Also ist $\sigma(\mathcal{A}^k) \subset \mathcal{B}^k$. Umgekehrt analog. □

2. $\mathcal{B}^k = \sigma(\mathcal{K}^k)$

   Beweis. $(\subset)$ klar nach (1).
   Sei $A \in \mathcal{A}^k$. Dann gilt

   $$A \cap [-n, n]^k \uparrow_{n \to \infty} A,$$

   d.h. $\bigcup_{n=1}^{\infty} A \cap [-n, n]^k = A$.
   Damit ist $A \in \sigma(\mathcal{K}^k)$, d.h.

   $$\mathcal{B}^k = \sigma(\mathcal{A}^k) \subset \sigma(\mathcal{K}^k)$$

   □

3. $\mathcal{B}^k = \sigma(\mathcal{I}^k)$

   Beweis. Übungsaufgabe □

[4] halboffene Quader
[5] Die kleinste σ-Algebra, die alle offenen Mengen enthält.

18.7 Definition (Axiomensystem von Kolmogrov)

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\Omega \neq \emptyset, \mathcal{A} \subset \mathcal{P}(\Omega)$ ist σ-Algebra und einem Wahrscheinlichkeitsmaß (auf $(\Omega, \mathcal{A})$, auf $\mathcal{A}$, auf Ω) $\mathbb{P}\colon \mathcal{A} \to \mathbb{R}$ mit:

1. $\mathbb{P}(A) \geq 0, \quad A \in \mathcal{A}$
2. $\mathbb{P}(\Omega) = 1$
3. $\mathbb{P}(\bigcup\limits_{j=1}^{\infty} A_j) = \sum\limits_{j=1}^{\infty} \mathbb{P}(A_j), \quad A_j \in \mathcal{A}, \quad A_i \cap A_j = \emptyset, i \neq j.$ (σ-Additivität)

Die Elemente von $\mathcal{A}$ heißen **Ereignisse**.

18.8 Bemerkung

Alle bisherigen Definitionen und Resultate (Stetigkeit von oben bzw. unten, Bayes' Formel, Unabhängigkeit) bleiben erhalten.

18.9 Definition und Satz

$\mathbb{P}$ sei Wahrscheinlichkeitsmaß auf $\mathcal{B}^1$.
Die Funktion $F\colon \mathbb{R} \to [0,1]$ definiert durch

$$F(x) := \mathbb{P}((-\infty, x]), x \in \mathbb{R},$$

heißt **Verteilungsfunktion** (VF) von $\mathbb{P}$. Es gilt

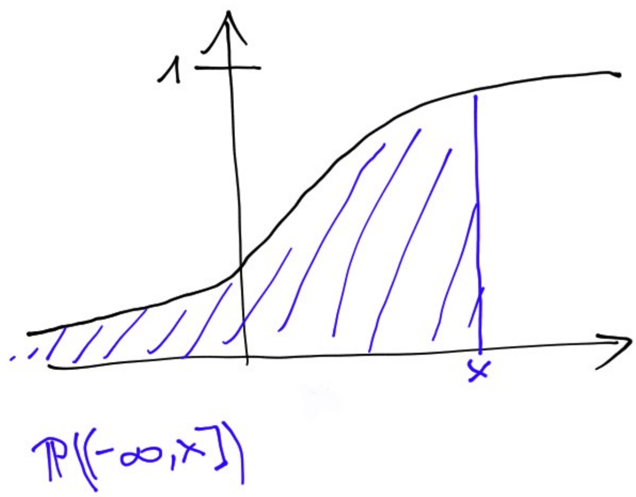
1. $x \leq y \Rightarrow F(x) \leq F(y)$
2. F ist rechtsseitig stetig.
3. $F(-\infty) := \lim\limits_{x \to -\infty} F(x) = 0,$
4. $F(\infty) := \lim\limits_{x \to \infty} F(x) = 1.$

Beweis. 1. $(-\infty, x] \subset (-\infty, y] \quad (x \leq y)$. $\mathbb{P}$ monoton $\rightsquigarrow F(x) \leq F(y)$

2. Es gelte $x_n \downarrow x$. Dann ist $\bigcap\limits_{n=1}^{\infty} (-\infty, x_n] = (-\infty, x]$
   $\mathbb{P}$ ist stetig von oben $\rightsquigarrow$

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

3.
$$\emptyset = \bigcap_{n=1}^{\infty} (-\infty, -n],$$
$$\mathbb{R} = \bigcup_{n=1}^{\infty} (-\infty, n]$$



□

18.10 Existenz- und Eindeutigkeitssatz

Sei $F\colon \mathbb{R} \to [0,1]$ eine Verteilungsfunktion, d.h. es gelte (1)-(3). Dann existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{B}^1$ mit

$$F(x) = \mathbb{P}((-\infty, x]), \quad x \in \mathbb{R}.$$

Beweis. WT. □

18.11 Satz

Sei F Verteilungsfunktion eines Wahrscheinlichkeitsmaßes $\mathbb{P}$ auf $\mathcal{B}^1$.

1. $\mathbb{P}((a,b]) = F(b) - F(a), \quad a \leq b$
2. $\mathbb{P}(\{x\}) = F(x) - F(\underbrace{x-}_{\text{linksseitiger Grenzwert}})$
3. $\mathbb{P}(\{x\}) = 0 \Leftrightarrow F$ ist stetig in x

4. F besitzt höchstens abzählbar unendlich viele Unstetigkeitsstellen.

Beweis. 1. Für $a \leq b$ gilt $(-\infty, a] \cup (a, b] = (-\infty, b]$ also

$$F(a) + \mathbb{P}((a, b]) = F(b).$$

2. Sei $x_n \uparrow x$ mit $x_n < x$. Dann

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

also

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

Damit

$$\begin{aligned} F(x) - F(x-) &= \mathbb{P}((-\infty, x]) - \mathbb{P}((-\infty, x)) \\ &= \mathbb{P}((-\infty, x] \setminus (-\infty, x)) = \mathbb{P}(\{x\}). \end{aligned}$$

3. klar.

4. $\{x \colon F(x) \neq F(x-)\} \subset \bigcup_{n=1}^{\infty} \underbrace{\{x \colon \mathbb{P}(\{x\}) \geq \frac{1}{n}\}}_{\text{hat höchstens } n \text{ Punkte}}$

□

18.12 Beispiel

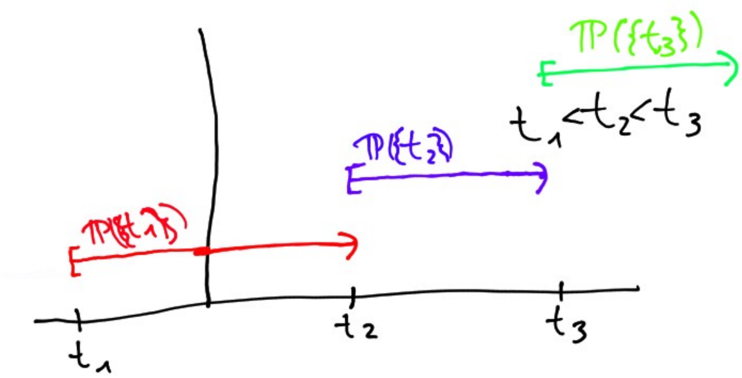
Sei $\mathbb{P}$ diskretes Wahrscheinlichkeitsmaß auf $\mathcal{B}^1$ mit Träger

$$T = \{t_1, t_2, t_3, \ldots\},$$

d.h. $\mathbb{P}(\{t_j\}) > 0$, $\sum_{j=1}^{\infty} \mathbb{P}(\{t_j\}) = 1$

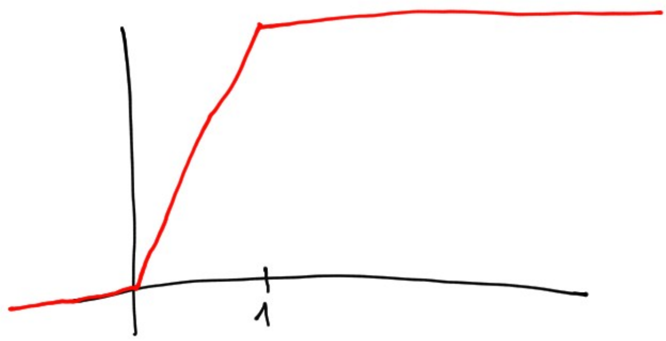
Dann gilt (nach Definition)

$$F(x) = \sum_{j \colon t_j \leq x} \mathbb{P}(\{t_j\})$$



18.13 Beispiel

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



Das zugehörige Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathcal{B}^1$ heißt **Gleichverteilung** auf $[0,1]$ ($\mathbb{P} = U([0,1])$). Der Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}^1, \mathbb{P})$ modelliert einen "rein zufälligen" Punkt in $[0,1]$.

18.14 Definition

$f\colon \mathbb{R} \to \mathbb{R}$ heißt **(Lebesque)Dichte** (über $\mathbb{R}$), falls

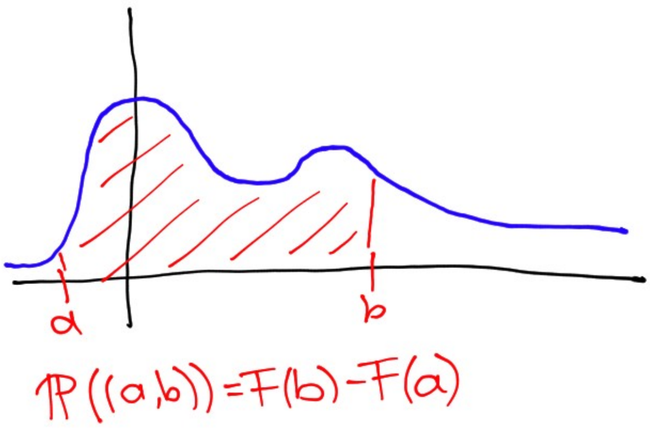
1. $f(x) \geq 0, x \in \mathbb{R}$
2. $\{x\colon f(x) \geq c\} \in \mathcal{B}^1, c \in \mathbb{R}$. ($f$ ist Borel-messbar)
3. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ (Lebesque-Integral)

Die durch

$$F(x) := \int\limits_{-\infty}^{x} f(y)dy, x \in \mathbb{R}$$

definiterte Funktion $F\colon \mathbb{R} \to [0,1]$ ist stetig und ist eine Verteilungsfunktion im Sinne von 18.9(1),(2),(3). Das nach 18.10 zu F gehörende Wahrscheinlichkeitsmapß $\mathbb{P}$ auf $\mathcal{B}^1$ erfüllt

$$\mathbb{P}(a,b]) = \int\limits_{a}^{b} f(y)dy$$



Ist f in x stetig, dann ist F in x stetig differenzierbar und $F'(x) = (x)$.

Das bedeutet

$$\text{"}\mathbb{P}((x, x + dx]) \approx f(x)dx\text{"}$$

(TODO: Rest texen)

18.15 Beispiel

Seien $f_1, \ldots, f_k$ Dichten auf $\mathbb{R}$. Setze

$$f(x_1, \ldots, x_k) = \prod_{j=1}^{k} f_j(x_j), \quad x_1, \ldots, x_k \in \mathbb{R}$$

Diese Funktion ist messbar und

$$\int f(x)dx = \int \ldots \int f_1(x_1) \ldots f_k(x_k)dx_1 \ldots dx_k = \prod_{j=1}^{k} \int f_j(x_j)dx_j = 1$$

18.16 Beispiel

$f_j(t) = f(t) = \frac{1}{\sqrt{2\pi}} exp[-\frac{t^2}{2}]$
Dann ist

$$f(x_1, \ldots, x_k) = \prod_{j=1}^{k} f_j(x_j) = \frac{1}{(2\pi)^{\frac{k}{2}}} exp[-\frac{1}{2} \sum_{j=1}^{k} x_j^2]$$

die Dichte der k-dimensionalen Normalverteilung.

18.17 Definition und Bemerkung

Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}^k$. Für $j \in \{1, \ldots, k\}$ sei

$$\mathbb{P}_j(B) := \mathbb{P}(\mathbb{R}^{j-1} \times B \times \mathbb{R}^{k-j}), B \in \mathcal{B}^1$$

Das ist ein Wahrscheinlichkeitsmaß. Man nennt es j-te Randverteilung von $\mathbb{P}$. Hat $\mathbb{P}$ die Dichte f, so hat $\mathbb{P}_j$ die Dichte

$$f_j(t) = \int \ldots \int f(x_1, \ldots, x_{i-1}, t, x_{i+q}, \ldots, x_k) dx_1 \ldots dx_{j-1} \ldots dx_k$$

(Fubini)

Kapitel 19

Zufallsvariablen

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ Wahrscheinlichkeitsraum.
Früher: $X\colon \Omega \to \mathbb{R}, \mathbb{P}^X(B) := \mathbb{P}(X^{-1}(B))$

19.1 Definition und Satz

Sei $\Omega' \neq \emptyset, \mathcal{A}' \subset \mathcal{P}(\Omega')$ σ-Algebra.

Eine Abbildung $X\colon \Omega \to \Omega'$ heißt **$(\mathcal{A}, \mathcal{A}')$-messbar**, falls

$$X^{-1}(B) \in \mathcal{A}, \quad B \in \mathcal{A}'.$$

Dann heißt X **Ω'-wertige Zufallsvariable**.

Die Verteilung $\mathbb{P}^X(\mathbb{P}(X_{\dots}), \mathcal{L}(x), \mathbb{P}X^{-1}, \dots)$ von X ist definiert als

$$\mathbb{P}^X(B) := \mathbb{P}(X^{-1}(B)), B \in \mathcal{A}'$$

Das ist ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{A}')$. (ÜA). Ist $(\Omega', \mathcal{A}') = (\mathbb{R}^k, \mathcal{B}^k)$, nennt man X einen k-dimensionalen Zufallsvektor. Im Fall $k = 1$ spricht man von einer Zufallsvariable.

19.2 Bemerkung

1. Sei $\mathcal{M}' \subset \mathcal{A}'$ mit $\sigma(\mathcal{M}') = \mathcal{A}'$. Dann gilt

$$X^{-1}(A') \in \mathcal{A}', A' \in \mathcal{M}' \Leftrightarrow X^{-1}(A') \in \mathcal{A}', A' \in \mathcal{U}'.$$

Zu zeigen ist $(\Rightarrow)$. Setze

$$\mathcal{G} := \{A' \subset \Omega' \colon X^{-1}(A') \in \mathcal{A}\}$$

Das ist eine σ-Algebra (ÜA).

Also gilt (wegen $\mathcal{M}' \subset \mathcal{G}$)

$$\mathcal{A}' = \sigma(\mathcal{M}') \subset \sigma(\mathcal{G}) = \mathcal{G}$$

2. Sei $g\colon \mathbb{R}^k \to \mathbb{R}^l$ stetig. Dann gilt
$$g^{-1}(A') \in \mathcal{O}^k \subset \mathcal{B}^k, A' \in \mathcal{O}^l$$
Wegen $\sigma(\mathcal{O}^l) = \mathcal{B}^l$ folgt aus (i)
$$g^{-1}(A') \in \mathcal{B}^k, A' \in \mathcal{B}^l.$$
Also ist g messbar.

3. $g\colon \mathbb{R}^k \to \mathbb{R}$ messbar (d.h. $(\mathcal{B}^k, \mathcal{B}^1)$-messbar) $\Leftrightarrow$
$$\{x\colon g(x) \geq c\} \in \mathcal{B}^k.$$
Beweis. Das System
$$\{[c, \infty)\colon c \in \mathbb{R}\}$$
ist ein Erzeuger von $\mathcal{B}^1$.
Man benutze jetzt (i). □

4. Sei $X\colon \Omega \to \Omega'$ $(\mathcal{A}, \mathcal{A}')$-messbar.
Sei $h\colon \Omega' \to \Omega''$ $(\mathcal{A}', \mathcal{A}'')$-messbar, wobei $\mathcal{A}'' \subset \mathcal{P}(\Omega'')$ eine σ-Algebra ist.
Dann ist die Abbildung $h \circ X\colon \Omega \to \Omega''$ $(\mathcal{A}, \mathcal{A}'')$-messbar. Für $A'' \in \mathcal{A}''$
$$(h \circ X)^{-1}(A'') = X^{-1}(\underbrace{h^{-1}(A'')}_{\in \mathcal{A}'}) \in \mathcal{A}.$$

5. Sei $\mathcal{D}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^k, \mathcal{B}^k)$. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{P}^X = \mathcal{D}$.
Dazu: $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^k, \mathcal{B}^k, \mathcal{D})$,
$$X = id_{\mathbb{R}^k}$$

Erinnerung $(\Omega, \mathcal{A}, \mathbb{P})$ $X\colon \Omega \to \Omega'$ $\mathcal{A}'$

Messbarkeit: $X^{-1}(B) \in \mathcal{A}, B \in \mathcal{A}'$
$\mathbb{P}^X$ Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{A}')$
$$\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)) =: \mathbb{P}(X \in B)$$
$(= \mathbb{P}(\{X \in B\}) = \mathbb{P}(\{\omega \in \Omega\colon X(\omega) \in B\}))$

$\Omega' = \mathbb{R}^k$ $\mathcal{B}^k$ Zufallsvektor
$k = 1$ Zufallsvariable

19.3 Bemerkung

Sei $\mathbb{Q}$ Wahrscheinlichkeitsmaß auf $\mathbb{R}^k$.
$X\colon \Omega \to \mathbb{R}^k$ mit Verteilung $\mathbb{Q}$. Man schreibt

$$X \sim \mathbb{Q}$$

("X ist nach $\mathbb{Q}$ verteilt.")

<u>z.B.:</u> $X \sim U[a,b]$ [1]
$X \sim Exp(\lambda), X \sim N(\mu, \sigma^2)$
Hat $\mathbb{Q}$ die Dichte f, so sagt man "X hat Dichte f".
Man sagt X ist eine (absolut) stetige Zufallsvariable. [2]

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

$$\mathbb{P}(X = t) = 0!$$

19.4 Bemerkung

Sei $\mathcal{Q}^k := \{[a,b]\colon a \leq b\}$ [3]
Für $\lambda^k([a,b]) := \prod_{j=1}^{k} (b_j - a_j)$ [4]

$N \in \mathcal{B}^k$ heißt **Nullmenge**

$$\Leftrightarrow \forall \epsilon > 0 \quad \exists k_1, k_2, \ldots \in \mathcal{Q}^k$$

mit $N \subset \bigcup_{i=1}^{\infty} k_i$ und $\sum_{j=1}^{\infty} \lambda^k(k_j) \leq \epsilon$.

(Hyperebenen $\{x\colon <x,n>= c\}$ sind Nullmengen)

Sind $f, g\colon \mathbb{R}^k \to \mathbb{R}$ Dichten eines Wahrscheinlichkeitsmaßes $\mathbb{Q}$, so gilt

$$f(x) = g(x), \quad x \notin N,$$

mit einer Nullmenge N.

---

[1] X auf dem Intervall $[a,b]$ gleichverteilt.
[2] Das bedeutet: X hat eine Dichte.
[3] $a, b \in \mathbb{R}^k$
[4] Lebesque-Maß

19.5 Bemerkung

Hat ein k-dimensionaler Zufallsvektor die Dichte f, so hat die X_j die Dichte

$$f_j(t) = \int \dots \int f(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_k) dx_1 \dots dx_k$$

19.6 Beispiel

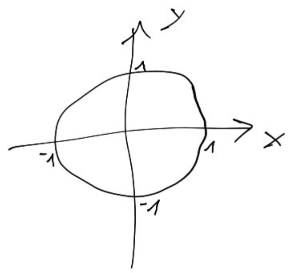
(X, Y) habe Dichte

$$f(x, y) := \frac{1}{\pi} 1_{\{} x^2 + y^2 \leq 1\} \equiv 1_K(x, y), K = \{(x, y) \colon x^2 + y^2 \leq 1\})$$

Dann

$$f_1(x) = \int f(x, y) dy = \int 1_{\{x^2+y^2 \leq 1\}} dy = \int 1_{\{|y| \leq \sqrt{1-x^2}\}} dy$$

$$= \begin{cases} 0, & \text{falls } |x| > 1 \\ \frac{2}{\pi}\sqrt{1 - x^2}, & \text{falls } |x| \leq 1 \end{cases}$$

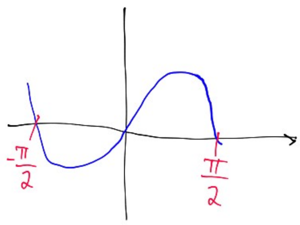


Ferner

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_1(x) dx = \int_{-1}^{t} \frac{2}{\pi} \sqrt{1 - x^2} dx, \quad -1 \leq t \leq 1$$

$$\stackrel{!}{=} \frac{1}{2} + \frac{1}{\pi} \arcsin(t) + \frac{1}{\pi} t \sqrt{1 - t^2}$$

(Differentiation und $\frac{\partial}{\partial t} \arcsin(t) = \frac{1}{\sqrt{1-t^2}}$)



19.7 Definition

1. [5] $A_1, \ldots, A_n \in \mathcal{A}$ heißen **stochastisch unabhängig**, falls

$$\mathbb{P}(\bigcap_{j\in T} A_j) = \prod_{j\in T} \mathbb{P}(A_j), \quad T \subset \{1,\ldots,n\}, T \neq \emptyset$$

2. Zufallsvariablen $X_1, \ldots, X_n$ heißen **stochastisch unabhängig**, falls

   $X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n)$ stochastisch unabhängig, für alle $B_1, \ldots, B_n \in \mathcal{B}^1$,

   bzw. falls

$$\mathbb{P}(X_1 \in B_1, \ldots X_n \in B_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j) \text{ für alle } B_1, \ldots, B_n \in \mathcal{B}^1.$$

19.8 Bemerkung

Hat $(X_1, \ldots, X_n)$ die Dichte f. Dann gilt:

1. $f(x) = \prod_{j=1}^{n} f_j(x_j) \Rightarrow X_1, \ldots, X_n$ stochastisch unabhängig
2. $X_1, \ldots, X_n$ stochastisch unabhängig $\Rightarrow f(x) = \prod_{j=1}^{n} f_j(x_j)$ außerhalb einer Nullmenge

Beweis. (1) :

$$\begin{aligned}
&\mathbb{P}((X_1, \ldots, X_n) \in B_1 \times \ldots \times B_n) \\
&= \mathbb{P}^X(B_1 \times \ldots \times B_n) = \int\limits_{B_1\times\ldots\times B_n} f(x)dx \\
&\overset{(1),Fubuni}{=} \int\limits_{B_1}\int\limits_{B_2} \ldots \int\limits_{B_n} f(x_1)\ldots f(x_n)dx_n \ldots dx_1 \\
&= \prod_{j=1}^{n} \int\limits_{B_j} f_j(x)dx = \prod_{j=1}^{n} \mathbb{P}^{X_j}(B_j) = \prod_{j=1}^{n} \mathbb{P}(X_j \in B_j)
\end{aligned}$$

(2) : WT □

[5] Die Ereignisse

Kapitel 20

Rechnen mit Dichten

$X = (X_1, \dots, X_k)$ k-dimensionaler Zufallsvektor mit Dichte f. Sei

$$T \colon \mathbb{R}^k \to \mathbb{R}^s \quad (s \leq k)$$

Sei $Y := T(X)$. (Y ist s-dimensionaler Zufallsvektor, falls Z messbar ist.)

20.1 Satz

Sei $K = 1$, d.h. X sei reelle Zufallsvariable. Die Dichte f sei stückweise stetig und es gelte

$$\mathbb{P}(X \in I) = 1$$

für ein offenes Intervall $I \subset \mathbb{R}$. Die Einschränkung $T\big|_I$ (von T auf I) sei stetig differenzierbar, streng monoton und $T'(x) \neq 0, x \in I$. Dann hat $Y = T(X)$ die Verteilungsfunktion

$$G(y) = \begin{cases} F(T^{-1}(y)), y \in T(I), & \text{falls } T \uparrow \\ 1 - F(T(y)), y \in T(I), & \text{falls } T \downarrow \end{cases}$$

wobei F die Verteilungsfunktion von X ist. Ferner hat G die Dichte

$$g(y) = \frac{f(T^{-1}(y))}{|T'(T^{-1}(y))|}, \quad y \in T(I)$$

und $g(y) = 0$ für $y \notin T(I)$.

Beweis. Sei T wachsend. Dann

$$\begin{aligned} G(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(T(X) \leq y) \\ &= \mathbb{P}(X \leq T^{-1}(y)) \quad y \in T(I) \end{aligned}$$

Die Funktion G ist stetig differenzierbar mit Ableitung

$$g(y) = G'(y) = \frac{f(T^{-1}(y))}{T'(T^{-1}(y))}$$

□

20.2 Beispiel

$T(x) = ax + b, \quad a \neq 0, b \in \mathbb{R}$
$T'(x) = a, T^{-1}(y) = \frac{y-b}{a}$
$g(y) = \frac{1}{|a|} f(\frac{y-b}{a})$

20.3 Beispiel

$X \sim N(0,1)$ [1], $T(x) = \sigma x + \mu$ $(\sigma > 0, \mu \in \mathbb{R})$, $Y := T(X) = \sigma X + \mu$

$$g(y) = \frac{1}{\sigma} \varphi(\frac{y-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{(y-\mu)^2}{\sigma^2}]$$

[2]

20.4 Beispiel

$T(x) = x^2, Y = X^2$

$$\begin{aligned} G(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(X^2 \leq y) \\ &\overset{y \geqq 0}{=} \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}) \end{aligned}$$

[3] Dichte $(y > 0)$:

$$g(y) = f(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

[1] Normalverteilung mit Mittelwert 0 und Standardabweichung 1
[2] φ ist die Dichte der Standardnormalverteilung
[3] F ist die Verteilungsfunktion von X.