Vorlesung Stochastik für Studierende der Informatik[1]

Philipp Harms

18. Juli 2017

Inhaltsverzeichnis

[1] Diese Notizen sind zu großen Teilen Peter Pfaffelhubers gleichnamigem Skript [Pfa13] entnommen.

1 Grundlegendes

1.1 Häufige Modelle: Münzwurf, Würfeln, Urne

1.1.1 Beispiel (Münze).

- Kopf/Zahl mit Wahrscheinlichkeit p bzw. $1-p$.
- Frage: Was ist $\mathbb{P}[\text{Kopf}, \text{Kopf}]$?

1.1.2 Beispiel (Würfeln).

- Würfeln mit 2 Würfeln gleichzeitig, alle sechs Augenzahlen haben gleiche Wahrscheinlichkeit.
- Frage: Was ist $\mathbb{P}[\text{Mäxchen}]$, wenn ein Mäxchen aus einem Einser und einem Zweier besteht?

1.1.3 Beispiel (Urne).

- n Kugeln, r Farben, k_i Kugeln der Farbe i, $k_1 + \cdots + k_r = n$.
- Frage: Was ist $\mathbb{P}[2 \text{ Kugeln der Farbe } i]$, wenn die Kugeln ohne Zurücklegen gezogen werden?

1.2 Kombinatorik

1.2.1 Definition (Notation).

- Mengen: ungeordnet mit paarweise verschiedenen Elementen. Bsp.: $\{1, 2, 3\}$.
- Multimengen: ungeordnet. Bsp.: $\{\!|1, 1, 2|\!\} = \{\!|1, 2, 1|\!\}$.
- Tupel: geordnet. Bsp.: $(1, 2, 3) \neq (1, 3, 2)$, $(1, 1, 2) \neq (1, 2, 1)$.
- $\mathbb{N} = \{0, 1, 2, \dots\}$, $n! = n(n-1)\dots 1$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- Potenzmenge 2^M einer Menge M.

1.2.2 Definition (Stichprobe). Sei M eine Menge, $k \in \mathbb{N}$ und $x_1, \dots, x_k \in M$.

Stichprobe	geordnet (Variation)	ungeordnet (Kombination)
Ohne Wiederholung	$(x_1, \dots, x_k)$ mit $x_i \neq x_j$ für $i \neq j$	$\{x_1, \dots, x_k\}$ mit $x_i \neq x_j$ für $i \neq j$
Mit Wiederholung	$(x_1, \dots, x_k)$	$\{\!|x_1, \dots, x_k|\!\}$

Geordnete Stichproben ohne Wiederholung mit $k = n$ heißen Permutationen oder Anordnungen.

1.2.3 Lemma (Stichproben aus einer Menge). *Sei M eine Menge mit n Elementen und $k \in \mathbb{N}$. Dann gilt:*

Anzahl k-elementiger Stichprobem	*geordnet (Variation)*	*ungeordnet (Kombination)*
Ohne Wiederholung	$\frac{n!}{(n-k)!}$ *für* $0 \leq k \leq n$	$\binom{n}{k}$ *für* $0 \leq k \leq n$
Mit Wiederholung	n^k *für* $k \geq 0$	$\binom{n+k-1}{k}$ *für* $k \geq 0$

Beweis.

(i) Geordnete Stichprobe ohne Wiederholung: n Möglichkeiten für x_1, $(n-1)$ Möglichkeiten für x_2, ..., $(n-k+1)$ Möglichkeiten für x_k, ergibt $n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$ Möglichkeiten insgesamt.

(ii) $\frac{\#\{\text{geordnete Stichproben ohne Wiederholung mit } k \text{ Elementen}\}}{\#\{\text{Anordnungen von } k \text{ Elementen}\}} = \frac{n!/(n-k)!}{k!} = \binom{n}{k}$.

(iii) n Möglichkeiten für jedes x_i ergibt n^k Möglichkeiten insgesamt.

(iv) Beispiel: $M = \{1, 2, 3, 4, 5\}, k = 3$. Wir kodieren Stichproben mit Hilfe von $k = 3$ Platzhaltern $\bullet$ und $n - 1 = 4$ Trennstrichen $|$. Die Anzahl der Platzhalter zwischen dem $(i-1)$-ten und i-ten Trennstrich gibt an, wie oft die Zahl i in der Stichprobe vorkommt. So entspricht etwa die Stichprobe $\{\!|1, 1, 3|\!\}$ dem Code $\bullet \bullet || \bullet ||$. Jeder Code entspricht einer eindeutigen Stichprobe. Wegen (ii) gibt es $\binom{n+k-1}{k}$ solcher Codes.

□

1.2.4 Lemma (Stichproben aus einer Multimenge). *Seien $x_1, \dots, x_r$ paarweise verschieden, sei $M = \{\!|\underbrace{x_1, \dots, x_1}_{k_1 \text{ mal}}, \dots, \underbrace{x_r, \dots, x_r}_{k_r \text{ mal}}|\!\}$ und sei $n = k_1 + \dots + k_r$. Dann gibt es $\frac{n!}{k_1! \dots k_r!}$ verschiedene Permutationen von M, d.h. Tupeln $(y_1, \dots, y_n)$ mit $\{\!|y_1, \dots, y_n|\!\} = M$.*

Beweis. Könnte man alle Objekte unterscheiden, so gäbe es $n!$ Möglichkeiten. Dies dividiert man durch die Anzahl $k_1! \dots k_r!$ der möglichen Anordnungen der ununterscheidbaren Objekte. □

1.3 Wahrscheinlichkeitsräume und Zufallsvariablen

1.3.1 Definition (Axiome der Wahrscheinlichkeitstheorie).

(i) Ein meßbarer Raum $(\Omega, \mathcal{F})$ besteht aus einer Menge Ω mit einer σ-Algebra $\mathcal{F}$, d.h., $\mathcal{F}$ besteht aus Teilmengen von Ω und erfüllt

$$\emptyset \in \mathcal{F}, \qquad \forall F \in \mathcal{F} : F^c \in \mathcal{F}, \qquad \forall F_1, F_2, \dots \in \mathcal{F} : \bigcup_{i=1}^{\infty} F_i \in \mathcal{F}.$$

(ii) Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ besteht aus einem meßbaren Raum $(\Omega, \mathcal{F})$ mit einem Wahrscheinlichkeitsmaß $\mathbb{P}$, d.h., $\mathbb{P} \colon \mathcal{F} \to [0, 1]$ erfüllt

$$\mathbb{P}[\emptyset] = 0, \qquad \forall F \in \mathcal{F} : \mathbb{P}[F^c] = 1 - \mathbb{P}[F],$$

$$\forall F_1, F_2, \dots \in \mathcal{F} \text{ disjunkt} : \mathbb{P}\left[\bigcup_{i=1}^{\infty} F_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[F_i]$$

(iii) Eine Zufallsvariable X ist eine Funktion $X\colon \Omega \to E$ von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ in einen meßbaren Raum $(E, \mathcal{A})$, die meßbar ist, d.h.,

$$\forall A \in \mathcal{A} : X^{-1}(A) := \{\omega \in \Omega : Y(\omega) \in A\} \in \mathcal{F}.$$

Die Verteilung (= das Bildmaß) von X unter $\mathbb{P}$ ist das Wahrscheinlichkeitsmaß $\mathbb{Q}\colon \mathcal{A} \to [0, 1]$ mit $\mathbb{Q}[A] = \mathbb{P}[X^{-1}(A)] = \mathbb{P}[X \in A]$.

1.3.2 Bemerkung.

- $\{\emptyset, \Omega\}$ ist die kleinste σ-Algebra auf Ω und die Potenzmenge die größte. Mengen $F \in \mathcal{F}$ interpretieren wir als Ereignisse und $\mathbb{P}[F]$ als deren Auftrittswahrscheinlichkeit.
- Oft kommt es nur auf die Verteilung von X an, so dass wir $(\Omega, \mathcal{F}, \mathbb{P})$ nicht näher spezifizieren und nur mit Wahrscheinlichkeiten $\mathbb{Q}[A] = \mathbb{P}[X \in A]$ des Bildmaßes rechnen.
- Manchmal verwenden wir den kanonischen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P}) = (E, \mathcal{A}, \mathbb{Q})$ und die kanonische Zufallsvariable $X = \mathrm{Id}_E$.

1.3.3 Lemma (Einschluss-Ausschluss-Formel, Siebformel)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(E, \mathcal{A})$ ein meßbarer Raum, und $X\colon \Omega \to E$ eine Zufallsvariable. Dann gilt für alle $A_1, \dots, A_n \in \mathcal{A}$ dass*

$$\begin{aligned}&\mathbb{P}[X \in A_1 \cup \dots \cup A_n]\\ &\quad= \sum_{1\le i\le n}^{n} \mathbb{P}[X \in A_i] - \sum_{1\le i<j\le n} \mathbb{P}[X \in A_i \cap A_j] + \dots \pm \mathbb{P}[X \in A_1 \cap \dots \cap A_n].\end{aligned}$$

Beweis. Vollständige Induktion. Die Fälle $n = 0, 1$ sind trivial. Der Fall $n + 1$ folgt aus dem Fall n:

$$\begin{aligned}&\mathbb{P}[X \in A_1 \cup \dots \cup A_{n+1}]\\ &\quad= \mathbb{P}[X \in A_1 \cup \dots \cup A_n] + \mathbb{P}[X \in A_{n+1} \backslash (A_1 \cup \dots \cup A_n)]\\ &\quad= \mathbb{P}[X \in A_1 \cup \dots \cup A_n] + \mathbb{P}[X \in A_{n+1}] - \mathbb{P}[X \in (A_1 \cap A_{n+1}) \cup \dots \cup (A_n \cap A_{n+1})]\\ &\quad= \sum_{1\le i\le n}^{n} \mathbb{P}[X \in A_i] - \sum_{1\le i<j\le n} \mathbb{P}[X \in A_i \cap A_j] + \dots \pm \mathbb{P}[X \in A_1 \cap \dots \cap A_n]\\ &\qquad + \mathbb{P}[X \in A_{n+1}] - \sum_{1\le i\le n}^{n} \mathbb{P}[X \in A_i \cap A_{n+1}] + \sum_{1\le i<j\le n} \mathbb{P}[X \in A_i \cap A_j \cap A_{n+1}] - \dots \square\end{aligned}$$

1.3.4 Definition (Diskrete und kontinuierliche Zufallsvariablen)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(E, \mathcal{A})$ ein meßbarer Raum, und $X\colon \Omega \to E$ eine Zufallsvariable.

(i) X heißt *diskrete* Zufallsvariable, wenn für alle bis auf endlich oder abzählbar viele $x \in E$ gilt dass $\mathbb{P}[X = x] = 0$ und wenn für alle $A \in \mathcal{A}$ gilt dass $\mathbb{P}[X \in A] = \sum_{x\in A} \mathbb{P}[X = x]$. In dem Fall heißt die Funktion $f\colon E \to [0, \infty), f(x) = \mathbb{P}[X = x]$ Zähldichte von X oder $\mathbb{P}$.

(ii) X heißt *kontinuierliche* Zufallsvariable, wenn $E \subseteq \mathbb{R}^n$ und es eine Funktion $f\colon E \to [0, \infty)$ gibt, so dass $\mathbb{P}[X \in A] = \int_A f(x)dx$ für alle $A \in \mathcal{A}$. In dem Fall heißt f Dichte von X oder $\mathbb{P}$.

1.4 Unabhängigkeit

1.4.1 Definition (Unabhängigkeit von Zufallsvariablen)**.** Seien $X_1, \dots, X_n$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in meßbaren Räumen $(E_1, \mathcal{A}_1), \dots, (E_n, \mathcal{A}_n)$.

(i) $X_1, \dots, X_n$ heißen unabhängig, wenn

$$\forall A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n : \mathbb{P}[X_1 \in A_1, \dots, X_n \in A_n] = \mathbb{P}[X_1 \in A_1] \dots \mathbb{P}[X_n \in A_n].$$

(ii) $X_1, \dots, X_n$ heißen paarweise unabhängig, wenn

$$\forall i \neq j, A_i \in \mathcal{A}_i, A_j \in \mathcal{A}_j : \mathbb{P}[X_i \in A_i, X_j \in A_j] = \mathbb{P}[X_i \in A_i]\mathbb{P}[X_j \in A_j].$$

Ereignisse $F_1, \dots, F_n \in \mathcal{F}$ heißen (paarweise) unabhängig wenn die $\{0,1\}$-wertigen Zufallsvariablen $\mathbb{1}_{F_1}, \dots, \mathbb{1}_{F_n}$ (paarweise) unabhängig sind.

1.4.2 Bemerkung.

- Frage: gibt es Zufallsvariablen, die von sich selbst unabhängig sind?
- Bemerkung: Es gibt paarweise unabhängige Zufallsvariablen, die abhängig sind [Pfa13, Beispiel 1.24].

1.4.3 Lemma. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $F_1, \dots, F_n \in \mathcal{F}$ Ereignisse. Dann ist äquivalent:*

(i) $F_1, \dots, F_n$ sind unabhängig.

(ii) Für alle $I \subseteq \{1, \dots, n\}$ gilt dass $\mathbb{P}\left[\bigcap_{i \in I} F_i\right] = \prod_{i \in I} \mathbb{P}[F_i]$.

1.4.4 Lemma. *Seien $(X_1, \dots, X_n)$ eine diskrete oder kontinuierliche Zufallsvariable mit Dichte oder Zähldichte f. Dann ist äquivalent:*

(i) $X_1, \dots, X_n$ sind unabhängig.

(ii) Es gibt Funktionen $f_1, \dots, f_n$ so dass $f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$.

2 Verteilungen und ihre Eigenschaften

2.1 Konstruktion von Verteilungen und Zufallsvariablen

2.1.1 Lemma (Konstruktion von Verteilungen und Zufallsvariablen)**.** *Sei E eine Menge, $(\Omega, \mathcal{F})$ ein meßbarer Raum, $X \colon \Omega \to E$, und $f \colon E \to [0, \infty)$.*

(i) Sei E endlich oder abzählbar und $\mathcal{A} = 2^E$ die Potenzmenge von E.

- *Dann ist $\mathcal{A}$ eine σ-Algebra.*
- *X ist meßbar wenn $\forall x \in E : X^{-1}(\{x\}) \in \mathcal{F}$.*
- *Wenn $\sum_{x \in E} f(x) = 1$ dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P} \colon \mathcal{A} \to [0,1]$ mit Zähldichte f.*

(ii) Sei E ein Quader, d.h. eine Menge der Form $I_1 \times \cdots \times I_n \subseteq \mathbb{R}^n$, wobei $I_1, \ldots, I_n$ Intervalle sind, und sei $\mathcal{A}$ die von Quadern erzeugte σ-Algebra:

$$\mathcal{A} := \bigcap \{\mathcal{B}; \mathcal{B} \text{ ist } \sigma\text{-Algebra auf } E, \text{ die alle Quader enthält}\}.$$

- *Dann ist $\mathcal{A}$ eine σ-Algebra, genannt die von Quadern erzeugte σ-Algbra.*
- *X ist meßbar wenn für alle Quader $I \subseteq E$ gilt dass $X^{-1}(I) \in \mathcal{F}$.*
- *Jede stetige Funktion ist meßbar bezüglich der von Quadern erzeugten σ-Algebren.*
- *Wenn f integrierbar ist mit $\int_E f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$, dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}\colon \mathcal{A} \to [0,1]$ mit Dichte f.*

2.1.2 Bemerkung.

- Standardmäßig betrachten wir Mengen $E \subseteq \mathbb{R}$ als meßbare Räume mit den σ-Algebren von Lemma 2.1.1.
- Die Existenz der in den folgenden Abschnitten definierten Verteilungen ist durch Lemma 2.1.1 gesichert.
- Frage: Gibt es eine 1:1 Beziehung zwischen Wahrscheinlichkeitsmaßen und Zähldichten auf E, wenn E endlich oder abzählbar ist? Das gleiche für Quader $E \subseteq \mathbb{R}^n$ und Dichten?
- Frage: Gibt es eine 1:1 Beziehung zwischen Verteilungen und Zufallsvariablen mit dieser Verteilung? Gibt es zu jeder Verteilung eine Zufallsvariable mit dieser Verteilung?
- Frage: Sei $E = \mathbb{R}$, $x \in E$, und $\mathbb{P}[A] = \mathbb{1}_A(x)$ für jedes A. Ist $\mathbb{P}$ ein Wahrscheinlichkeitmaß? Hat $\mathbb{P}$ eine Dichte? Bemerkung: $\mathbb{P}$ heißt Dirac-Maß.
- Frage: Ist die Summe von zwei Wahrscheinlichkeitsmaßen wieder ein Wahrscheinlichkeitsmaß? Ist die Summe von zwei Zufallsvariablen wieder eine Zufallsvariable?

2.2 Uniforme Verteilung

2.2.1 Definition (Uniforme Verteilung)**.**

(i) Die uniforme Verteilung $U(E)$ auf einer endlichen Menge E ist das eindeutige Wahrscheinlichkeitsmaß $\mathbb{P}$ auf E mit konstanter Zähldichte. Es erfüllt $\mathbb{P}[X \in A] = \frac{\#A}{\#E}$ und hat die Zähldichte $f(x) = \frac{1}{\#E}$.

(ii) Die uniforme Verteilung $U([a,b])$ auf einem Intervall $[a,b]$ ist das eindeutige Wahrscheinlichkeitsmaß $\mathbb{P}$ auf E mit konstanter Dichte. Es erfüllt $\mathbb{P}[X \in [c,d]] = \frac{d-c}{b-a}$ und hat die Dichte $f(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a}$.

2.2.2 Bemerkung.

- Ein zufälliges Experiment, bei dem alle Ausgänge gleich wahrscheinlich (also uniform verteilt) sind, wird auch Laplace-Experiment genannt.

2.2.3 Beispiel. Im folgenden Beispiel sind X und Y auf unterschiedlichen Wahrscheinlichkeitsräumen definiert, haben aber dieselbe uniforme Verteilung:

- $\Omega = [0,1]$, $\mathbb{P}[[c,d]] = d - c$, $Y\colon F \to E$, $Y = \mathbb{1}_{(1/2,1]}$.
- $E = \{0,1\}$, $\mathbb{Q}[A] = \frac{\#A}{\#E}$, $X = \mathrm{Id}_E$.

2.3 Bernoulli-Verteilung

2.3.1 Definition (Bernoulli-Verteilung)**.** Für jedes $p \in [0,1]$ ist die Bernoulli-Verteilung $B(p)$ das eindeutige Wahrscheinlichkeitsmaß $\mathbb{P}$ zauf $\{0,1\}$ mit Zähldichte $f(1) = p$, $f(0) = 1-p$.

2.3.2 Beispiel. Wurf einer Münze.

2.4 Binomialverteilung

2.4.1 Herleitung (Summen von Bernoulli-Verteilungen)**.**

- Ziel: Berechnen der Verteilung von $Y = X_1 + \dots + X_n$, wenn $X_1, \dots, X_n \sim B(p)$ unabhängig Bernoulli-verteilt sind.

- Schritt 1: Wir wählen einen passenden Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ für $X = (X_1, \dots, X_n)$ wie folgt. Jede Zufallsvariable X_i hat die Zähldichte
$$f_i\colon \{0,1\} \to [0,1], \qquad f_i(1) = p, \qquad f_i(0) = 1-p.$$
Weil $X_1, \dots, X_n$ unabhängig sind, hat X die Zähldichte
$$f\colon \{0,1\}^n \to [0,1], \qquad \begin{aligned} f(x_1, \dots, x_n) &= f_1(x_1) \dots f_n(x_n) \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n (1-x_i)}. \end{aligned}$$
Wir definieren also
$$\Omega = \{0,1\}^n, \qquad \mathcal{F} = 2^\Omega, \qquad \mathbb{P} = (\text{W-Maß mit Zähldichte } f), \qquad X = \mathrm{Id}_\Omega\,.$$

- Schritt 2: Wir wählen einen passenden meßbaren Raum $(E, \mathcal{A})$ als Bildraum für Y wie folgt,
$$E = \{0, \dots, n\}, \qquad \mathcal{A} = 2^E,$$
und definieren die Zufallsvariable $Y = h(X)$, wobei
$$h\colon \Omega \to E, \qquad h(x_1, \dots, x_n) = x_1 + \dots + x_n.$$

- Schritt 3: Wir berechnen die Verteilung von Y wie folgt: für jedes $k \in E$ gilt
$$\mathbb{P}[Y = k] = \mathbb{P}[X \in h^{-1}(k)] = \sum_{x \in h^{-1}(k)} \mathbb{P}[X = x].$$
Beispiel:

k	$h^{-1}(k)$	$\#h^{-1}(k)$
0	$\{(0,\dots,0)\}$	1
1	$\{(1,0,\dots,0),(0,1,\dots,0),\dots,(0,0,\dots,1)\}$	n
2	$\{(1,1,0,\dots,0),(1,0,1,\dots,0),\dots\}$	$n(n-1)/2$

Wir sehen: Die Menge $h^{-1}(k)$ hat $\binom{n}{k}$ Elemente, die jeweils mit Wahrscheinlichkeit $p^k(1-p)^{n-k}$ auftreten. Daher gilt

$$\mathbb{P}[Y=k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Wir fassen zusammen: Die Verteilung von Y ist das eindeutige Wahrscheinlichkeitsmaß auf $(E, \mathcal{A})$ mit Zähldichte

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

2.4.2 Definition (Binomialverteilung)**.** Für jedes $n \in \mathbb{N}$ und $p \in [0,1]$ ist die Binomial-Verteilung $B(n,p)$ die eindeutige Verteilung auf $\{0, \ldots, n\}$ mit Zähldichte

$$k \mapsto \binom{n}{k} p^k (1-p)^{n-k}.$$

2.4.3 Bemerkung. $B(1,p) = B(p)$.

2.5 Kenngrößen von Verteilungen

2.5.1 Definition (Kenngrößen von Verteilungen)**.** Seien X und Y Zufallsvariablen mit Werten in $\mathbb{R}^m$ und $\mathbb{R}^n$ und sei $\alpha \in (0,1)$.

(i) Sei $n=1$. Die *Verteilungsfunktion* von X (oder $\mathbb{P}$) ist die Funktion

$$F\colon \mathbb{R} \to [0,1], \qquad F(x) = \mathbb{P}[X \leq x].$$

Ein α-*Quantil* von X (oder $\mathbb{P}$) ist eine Zahl $q_\alpha \in \mathbb{R}$ so dass

$$\mathbb{P}[X < q_\alpha] \leq \alpha \leq \mathbb{P}[X \leq q_\alpha].$$

0.5-Quantile heißen *Median*.

(ii) Der *Erwartungswert* (= Mittelwert) von X ist definiert als

$$\mathbb{E}[X] = \sum_{x \in \mathbb{R}^n} x f(x) \in \mathbb{R}^n,$$

falls X eine diskrete Zufallsvariable mit Zähldichte f ist, und als

$$\mathbb{E}[X] = \int_{\mathbb{R}^n} x f(x) dx \in \mathbb{R}^n,$$

falls X eine kontinuierliche Zufallsvariable mit Dichte f ist.

(iii) Die *Kovarianz* und *Varianz* sind definiert als

$$\operatorname{Cov}[X,Y] = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^\top\big] \in \mathbb{R}^{m \times n}, \qquad \operatorname{Var}[X] = \operatorname{Cov}[X,X],$$

falls die Erwartungswerte existieren. Im Fall $n = m = 1$ definieren wir die *Standardabweichung* und den *Korrelationskoeffizienten* als

$$\sigma_X = \sqrt{\operatorname{Var}[X]}, \qquad \operatorname{Cor}[X,Y] = \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y}.$$

X und Y heißen unkorreliert wenn $\operatorname{Cor}[X,Y] = 0$.

2.5.2 Bemerkung.

- Der Erwartungswert existiert nur, wenn die Summe bzw. das Integral in Definition 2.5.1 wohldefiniert und endlich ist. Dasselbe gilt für die Kovarianz.
- Definition 2.5.1 kann auch auf Zufallsvariablen mit Werten in $E \subset \mathbb{R}$ angewandt werden, da diese auch als $\mathbb{R}$-wertige Zufallsvariablen aufgefasst werden können.
- Frage: Angenommen X hat eine bestimmte physikalische Dimension, z.B. Meter. Was ist die Dimension der Verteilungsfunktion, des Medians, des Erwartungswertes, und der Varianz von X?
- Frage: Angenommen X und Y haben dieselbe Verteilung. Haben sie dann auch dieselbe Verteilungsfunktion, denselben Median, denselben Erwartungswert, und dieselbe Varianz?
- Frage: Beispiele von Verteilungen wo Mittelwert und Median verschieden sind? Beispiele von Verteilungen wo der Median nicht eindeutig ist?
- Frage: Zeigen Sie $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$.

2.5.3 Proposition (Kenngrößen von Bernoulli und uniformen Verteilungen)**.**

(i) Für $X \sim B(p)$ gilt

$$\begin{aligned}\mathbb{E}[X] &= 1 \cdot f(1) + 0 \cdot f(0) = p,\\ \operatorname{Var}[X] &= (1-p)^2 f(1) + (0-p)^2 f(0) = (1-p)^2 p + p^2(1-p) = p(1-p).\end{aligned}$$

(ii) Für $X \sim U(\{1,\dots,n\})$ gilt

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^n i f(i) = \frac{1}{n}\sum_{i=1}^n i = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}\\ \operatorname{Var}[X] &= \frac{1}{n}\sum_{i=1}^n \left(i - \frac{n+1}{2}\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(i^2 - 2i\frac{n+1}{2} + \frac{(n+1)^2}{4}\right)\\ &= \frac{1}{n}\left(\frac{n(n+1)(2n+1)}{6} - 2\frac{n+1}{2}\frac{n(n+1)}{2} + \frac{n(n+1)^2}{4}\right)\\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)\big(2(2n+1) - 3(n+1)\big)}{12}\\ &= \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12},\end{aligned}$$

wobei wir die folgenden Summenformeln verwendet haben:

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}, \qquad \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}.$$

(iii) Für $X \sim U([a,b])$ gilt

$$\mathbb{E}[X] = \int_a^b x f(x)dx = \frac{1}{b-a}\int_a^b x dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{x=b} = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2},$$

$$\begin{aligned}\operatorname{Var}[X] &= \frac{1}{b-a}\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx = \frac{1}{b-a}\left[\frac{1}{3}\left(x-\frac{a+b}{2}\right)^3\right]_{x=a}^{x=b} \\ &= \frac{1}{b-a}\frac{2}{3}\left(\frac{b-a}{2}\right)^3 = \frac{(b-a)^2}{12}.\end{aligned}$$

2.5.4 Lemma (Eigenschaften des Erwartungswerts)**.** *Seien X und Y diskrete oder kontinuierliche $\mathbb{R}^n$-wertige Zufallsvariablen und sei $h\colon \mathbb{R}^n \to \mathbb{R}^m$ meßbar.*

(i) Transformationen: Ist Y diskret mit Zähldichte f, dann gilt $\mathbb{E}[h(Y)] = \sum_{y\in\mathbb{R}^n} h(y)f(y)$.

(ii) Linearität: Für alle $a, b \in \mathbb{R}$ gilt, dass $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

(iii) Monotonie: Gilt $n = 1$ und $X \le Y$, dann gilt auch $\mathbb{E}[X] \le \mathbb{E}[Y]$.

Beweis. Wir zeigen nur den diskreten Fall.

(i): Sei $Z = h(Y)$. Dann gilt

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{z\in\mathbb{R}^m} z\mathbb{P}[Z = z] = \sum_{z\in h(\mathbb{R}^n)} z\mathbb{P}[Y \in h^{-1}(z)] \\ &= \sum_{z\in h(\mathbb{R}^n)} z \sum_{y\in h^{-1}(z)} \mathbb{P}[Y = y] = \sum_{z\in h(\mathbb{R}^n)} \sum_{y\in h^{-1}(z)} h(y)\mathbb{P}[Y = y] \\ &= \sum_{y\in\mathbb{R}^n} h(y)\mathbb{P}[Y = y] = \sum_{y\in\mathbb{R}^n} h(y)f(y),\end{aligned}$$

und die linke Seite ist endlich genau dann wenn die rechte Seite endlich ist.

(ii): Wir wenden (i) auf die Zufallsvariable $aX + bY$ an:

$$\begin{aligned}\mathbb{E}[aX + bY] &= \sum_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n} (ax + by)\mathbb{P}[(X, Y) = (x, y)] \\ &= a \sum_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n} x\mathbb{P}[(X, Y) = (x, y)] + b \sum_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n} y\mathbb{P}[(X, Y) = (x, y)] \\ &= a \sum_{x\in\mathbb{R}^n} x\mathbb{P}[X = x] + b \sum_{y\in\mathbb{R}^n} y\mathbb{P}[Y = y] = a\mathbb{E}[X] + b\mathbb{E}[Y].\end{aligned}$$

(iii): Sei $Z = Y - X \ge 0$. Dann gilt wegen (ii) dass

$$\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Z] = \sum_{z\in\mathbb{R}^n} z\mathbb{P}[Z = z] \ge 0. \qquad \square$$

2.5.5 Lemma (Eigenschaften der Varianz)**.** *Seien X eine Zufallsvariable mit endlicher Varianz und Werten in $\mathbb{R}^m$, und seien Y und Z Zufallsvariablen mit endlicher Varianz und Werten in $\mathbb{R}^n$. Dann gilt:*

(i) $\operatorname{Cov}[X, Y] = \mathbb{E}[XY^\top] - \mathbb{E}[X]\mathbb{E}[Y]^\top$.

(ii) Linearität: Für alle $a, b \in \mathbb{R}$ gilt

$$\begin{aligned}\operatorname{Cov}[X, aY + bZ] &= a \operatorname{Cov}[X, Y] + b \operatorname{Cov}[X, Z], \\ \operatorname{Var}[aY + bZ] &= a^2 \operatorname{Var}[Y] + ab \operatorname{Cov}[Y, Z] + ab \operatorname{Cov}[Z, Y] + b^2 \operatorname{Var}[Z].\end{aligned}$$

(iii) Unabhängigkeit/Unkorelliertheit: $\operatorname{Cov}[X,Y]=0$ *wenn* X *und* Y *unabhängig sind.*

(iv) Cauchy-Schwarz: Für $m,n=1$ *gilt*

$$\operatorname{Cov}[X,Y]^2 \le \operatorname{Var}[X]\operatorname{Var}[Y], \qquad -1 \le \operatorname{Cor}[X,Y] \le 1.$$

Beweis. Wir setzen $\mu_X := \mathbb{E}[X]$, $\mu_Y := \mathbb{E}[Y]$, $\tilde{X} = X-\mu_X$, $\tilde{Y} = Y-\mu_Y$.

(i): Wegen der Linearität des Erwartungswerts gilt dass

$$\begin{aligned}\operatorname{Cov}[X,Y] &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)^\top] = \mathbb{E}[XY^\top] - \mu_X\mathbb{E}[Y^\top] - \mathbb{E}[X]\mu_Y^\top + \mu_X\mu_Y^\top \\ &= \mathbb{E}[XY^\top] - \mathbb{E}[X]\mathbb{E}[Y^\top].\end{aligned}$$

(ii): Wegen der Linearität des Erwartungswerts gilt

$$\begin{aligned}\operatorname{Cov}[X,aY+bZ] &= \mathbb{E}[\tilde{X},a\tilde{Y}+b\tilde{Z}] = a\mathbb{E}[\tilde{X},\tilde{Y}] + b\mathbb{E}[\tilde{X},\tilde{Z}] = a\operatorname{Cov}[X,Y] + b\operatorname{Cov}[X,Z]. \\ \operatorname{Var}[aY+bZ] &= \mathbb{E}[(a\tilde{Y}+b\tilde{Z})^2] = \mathbb{E}[a^2\tilde{Y}^2 + 2ab\tilde{Y}\tilde{Z} + b^2\tilde{Z}^2] \\ &= a^2\operatorname{Var}[Y] + 2ab\operatorname{Cov}[Y,Z] + b^2\operatorname{Var}[Z].\end{aligned}$$

(iii): Wir zeigen die Behauptung nur für X,Y diskret. Wegen der Unabhängigkeit von X und Y und Lemma 1.4.4 gilt

$$f_{X,Y}(x,y) = \mathbb{P}[X=x,Y=y] = \mathbb{P}[X=x]\mathbb{P}[Y=y] = f_X(x)f_Y(y)$$

und es folgt aus der Transformationseigenschaft des Erwartungswerts dass

$$\begin{aligned}\operatorname{Cov}[X,Y] &= \mathbb{E}[\tilde{X}\tilde{Y}^\top] = \sum_{x,y}(x-\mu_X)(y-\mu_Y)^\top f_{X,Y}(x,y) \\ &= \sum_x (x-\mu_X)f_X(x)\sum_y (y-\mu_Y)f_Y(y) = \mathbb{E}[\tilde{X}]\mathbb{E}[\tilde{Y}] = 0.\end{aligned}$$

(iv): Falls $\mathbb{E}[\tilde{X}^2]=0$, dann gilt $\mathbb{P}[\tilde{X}=0]=1$ und beide Seiten der ersten Ungleichung sind Null. Falls $\mathbb{E}[\tilde{X}^2]>0$, dann gilt für $c = \mathbb{E}[|\tilde{X}\tilde{Y}|]/\mathbb{E}[\tilde{X}^2]$ wegen der Monotonie und Linearität des Erwartungswerts dass

$$\begin{aligned}0 &\le \mathbb{E}[(-c|\tilde{X}|+|\tilde{Y}|)^2] = c^2\mathbb{E}[|\tilde{X}|^2] - 2c\mathbb{E}[|\tilde{X}\tilde{Y}|] + \mathbb{E}[\tilde{Y}^2] \\ &= \mathbb{E}[|\tilde{X}\tilde{Y}|]^2/\mathbb{E}[|\tilde{X}|^2] - 2\mathbb{E}[|\tilde{X}\tilde{Y}|]^2/\mathbb{E}[|\tilde{X}|^2] + \mathbb{E}[\tilde{Y}^2] \\ &= \mathbb{E}[\tilde{Y}^2] - \mathbb{E}[|\tilde{X}\tilde{Y}|]^2/\mathbb{E}[|\tilde{X}|^2].\end{aligned}$$

Dies zeigt die erste Ungleichung. Die zweite Ungleichung folgt aus

$$\operatorname{Cor}[X,Y]^2 = \frac{\operatorname{Cov}[X,Y]^2}{\operatorname{Var}[X]\operatorname{Var}[Y]} \le 1. \qquad \square$$

2.5.6 Proposition (Kenngrößen der Binomialverteilung)**.** *Sei* $X \sim B(n,p)$*. Dann gibt es unabhängige* $X_1,\dots,X_n \sim B(p)$*, so dass* X *die gleiche Verteilung hat wie* $\sum_{i=1}^n X_i$*, und es gilt*

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^n \mathbb{E}[X_i] = np, \\ \operatorname{Var}[X] &= \sum_{i=1}^n \operatorname{Var}[X_i] = np(1-p).\end{aligned}$$

2.6 Numerische Simulation von Zufallsvariablen

Viele Zufallszahlengeneratoren liefern unabhängige $U([0,1])$ verteilte Zufallsvariablen, aus denen mit Hilfe des folgenden Lemmas Zufallsvariablen mit beliebiger Verteilung erzeugt werden können.

2.6.1 Lemma (Simulationslemma). *Sei $F\colon \mathbb{R} \to [0,1]$ die Verteilungsfunktion einer (diskreten oder stetigen) Verteilung $\mathbb{P}$ auf $\mathbb{R}$, F^{-1} die Pseudoinverse von F, d.h.,*

$$F^{-1}(y) = \inf\{x; F(x) \geq y\}, \qquad y \in \mathbb{R},$$

$U \sim U([0,1])$ und $X = F^{-1}(U)$. Dann gilt $X \sim \mathbb{P}$.

Beweis. Für alle $x \in \mathbb{R}$ gilt

$$\mathbb{P}[X \leq x] = \mathbb{P}[U \leq F(x)] = F(x). \qquad \square$$

3 Weitere wichtige Verteilungen

3.1 Multinomialverteilung

3.1.1 Herleitung (Ziehen mit Wiederholung aus Multimengen). Sei $n \in \mathbb{N}$, $k \leq n$, und

$$M = \{\!|m_1, \dots, m_n|\!\} = \{\!|\underbrace{1, \dots, 1}_{n_1 \text{ mal}}, \dots, \underbrace{r, \dots, r}_{n_r \text{mal}}|\!\}, \qquad p_i = n_i/n.$$

- Schritt 1: Geordnete Stichprobe X mit Wiederholung aus $\{1, \dots, n\}$. Wir setzen

$$\Omega = \{1, \dots, n\}^k, \qquad \mathbb{P} \sim U(\Omega), \qquad X = \mathrm{Id}_\Omega .$$

Dann gilt wegen Lemma 1.2.3 für alle $x \in \Omega$ dass

$$\mathbb{P}[X = x] = \frac{1}{\#\Omega} = \frac{1}{n^k}.$$

- Schritt 2: Geordnete Stichprobe Y mit Wiederholung aus M. Wir setzen

$$E = \{1, \dots, r\}^k, \qquad f\colon \Omega \to E, \qquad f(x) = (m_{x_1}, \dots, m_{x_k}), \qquad Y = f(X).$$

Sei $y \in E$ beliebig mit

$$\{\!|y_1, \dots, y_k|\!\} = \{\!|\underbrace{1, \dots, 1}_{k_1 \text{ mal}}, \dots, \underbrace{r, \dots, r}_{k_r \text{mal}}|\!\}.$$

Dann gilt wegen Lemma 1.2.3

$$\mathbb{P}[Y = y] = \sum_{x \in f^{-1}(y)} \mathbb{P}[X = x] = \frac{\#f^{-1}(y)}{n^k} = \frac{n_1^{k_1} \dots n_r^{k_r}}{n^k} = p_1^{k_1} \dots p_r^{k_r}.$$

- Schritt 3: Ungeordnete Stichprobe Z ohne Wiederholung aus M. Wir setzen

  $$F = \big\{\{\!|y_1, \dots, y_k|\!\}; y \in E\big\}, \qquad g\colon E \to F, \qquad g(y) = \{\!|y_1, \dots, y_k|\!\}, \qquad Z = g(Y).$$

  Jedes $z \in F$ entspricht einem eindeutigen Tupel $(k_1, \dots, k_r)$ so dass

  $$\{\!|z_1, \dots, z_k|\!\} = \{\!|\underbrace{1, \dots, 1}_{k_1 \text{ mal}}, \dots, \underbrace{r, \dots, r}_{k_r \text{mal}}|\!\},$$

  und es gilt wegen Lemma 1.2.4

  $$\begin{aligned}\mathbb{P}[Z = z] &= \sum_{y \in g^{-1}(z)} \mathbb{P}[Y = y] = \big(\#g^{-1}(z)\big)\mathbb{P}[Y = y]\\ &= \frac{k!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}.\end{aligned}$$

- Schritt 4: Dies kann auch als Verteilung $\mathbb{P}[(K_1, \dots, K_r) = (k_1, \dots, k_r)]$ aufgefasst werden, wobei

  $$K_i = \sum_{\ell=1}^{k} \mathbb{1}_{\{i\}}(Y_\ell), \qquad i \in \{1, \dots, r\}.$$

3.1.2 Definition (Multinomialverteilung)**.** Für alle $k \in \mathbb{N}$ und $p_1, \dots, p_r \in [0, 1]$ mit $p_1 + \dots + p_r = 1$ ist die Multinomialverteilung $\mathrm{B}(k, p_1, \dots, p_r)$ die eindeutige Verteilung auf der Menge

$$\{(k_1, \dots, k_r) \in \mathbb{N}^r; k_1 + \dots + k_r = k\}$$

mit Zähldichte

$$(k_1, \dots, k_r) \mapsto \frac{k!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}.$$

3.1.3 Beispiel (Anwendung in statistischen Sprachmodellen)**.**

- Texte werden als Mengen von Wörtern repräsentiert (bag of words).
- Worthäufigkeiten sind multinomialverteilt.
- Dies kann zur Klassifikation von Texten verwendet werden (text mining, spam recognition, etc.).

3.1.4 Proposition (Kenngrößen der Multinomialverteilung)**.** *Sei* $(K_1, \dots, K_r) \sim \mathrm{B}(k, p_1, \dots, p_r)$. *Dann gilt für* $Y_1, \dots, Y_k$ *wie in Herleitung 3.1.1,* $i, j \in \{1, \dots, r\}$, $i \neq j$, $q_i = 1 - p_i$, *dass*

$$\begin{aligned}\mathbb{E}[K_i] &= \sum_{\ell=1}^{k} \mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)] = kp_i,\\ \mathrm{Var}[K_i] &= \sum_{\ell=1}^{k} \mathrm{Var}[\mathbb{1}_{\{i\}}(Y_\ell)] + 2 \sum_{1 \le \ell < m \le k} \underbrace{\mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{i\}}(Y_m)]}_{=0} = kp_i q_i,\\ \mathrm{Cov}[K_i, K_j] &= \sum_{\ell=1}^{k} \mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{j\}}(Y_\ell)] + 2 \sum_{1 \le \ell < m \le k} \underbrace{\mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{j\}}(Y_m)]}_{=0}\\ &= \sum_{\ell=1}^{k} \Big(\underbrace{\mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)\mathbb{1}_{\{j\}}(Y_\ell)]}_{=0} - \mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)]\mathbb{E}[\mathbb{1}_{\{j\}}(Y_\ell)] \Big) = -kp_i p_j.\end{aligned}$$

3.2 Hypergeometrische Verteilung

3.2.1 Herleitung (Ziehen ohne Wiederholung aus Multimengen)**.** Sei $n \in \mathbb{N}$, $k \leq n$, und

$$M = \{\!|m_1, \ldots, m_n|\!\} = \{\!|\underbrace{1, \ldots, 1}_{n_1 \text{ mal}}, \ldots, \underbrace{r, \ldots, r}_{n_r \text{mal}}|\!\}.$$

- Schritt 1: Geordnete Stichprobe X ohne Wiederholung aus $\{1, \ldots, n\}$. Wir setzen

$$\Omega = \big\{x \in \{1, \ldots, n\}^k ; x_i \neq x_j \text{ für } i \neq j\big\}, \qquad \mathbb{P} \sim U(\Omega), \qquad X = \mathrm{Id}_\Omega\,.$$

  Dann gilt wegen Lemma 1.2.3 für alle $x \in \Omega$ dass

$$\mathbb{P}[X = x] = \frac{1}{\#\Omega} = \frac{1}{\frac{n!}{(n-k)!}}.$$

- Schritt 2: Geordnete Stichprobe Y ohne Wiederholung aus M. Wir setzen

$$\begin{gathered} E = \big\{(y_1, \ldots, y_k) \in \{1, \ldots, r\}^k ; \{\!|y_1, \ldots, y_k|\!\} \subseteq M\big\}, \\ f \colon \Omega \to E, \qquad f(x) = (m_{x_1}, \ldots, m_{x_k}), \qquad Y = f(X). \end{gathered}$$

  Sei $y \in E$ beliebig mit

$$\{\!|y_1, \ldots, y_k|\!\} = \{\!|\underbrace{1, \ldots, 1}_{k_1 \text{ mal}}, \ldots, \underbrace{r, \ldots, r}_{k_r \text{mal}}|\!\}.$$

  Dann gilt wegen Lemma 1.2.3

$$\mathbb{P}[Y = y] = \sum_{x \in f^{-1}(y)} \mathbb{P}[X = x] = \frac{\#f^{-1}(y)}{\frac{n!}{(n-k)!}} = \frac{\frac{n_1!}{(n_1-k_1)!} \cdots \frac{n_r!}{(n_r-k_r)!}}{\frac{n!}{(n-k)!}}.$$

- Schritt 3: Ungeordnete Stichprobe Z ohne Wiederholung aus M. Wir setzen

$$\begin{gathered} F = \big\{\{\!|y_1, \ldots, y_k|\!\} ; \{\!|y_1, \ldots, y_k|\!\} \subseteq M\big\}, \\ g \colon E \to F, \qquad g(y) = \{\!|y_1, \ldots, y_k|\!\}, \qquad Z = g(Y). \end{gathered}$$

  Jedes $z \in F$ entspricht einem eindeutigen Tupel $(k_1, \ldots, k_r)$ so dass

$$\{\!|z_1, \ldots, z_k|\!\} = \{\!|\underbrace{1, \ldots, 1}_{k_1 \text{ mal}}, \ldots, \underbrace{r, \ldots, r}_{k_r \text{mal}}|\!\},$$

  und es gilt wegen Lemma 1.2.4

$$\begin{aligned} \mathbb{P}[Z = z] &= \sum_{y \in g^{-1}(z)} \mathbb{P}[Y = y] = \big(\#g^{-1}(z)\big)\mathbb{P}[Y = y] \\ &= \frac{k!}{k_1! \ldots k_r!} \frac{\frac{n_1!}{(n_1-k_1)!} \cdots \frac{n_r!}{(n_r-k_r)!}}{\frac{n!}{(n-k)!}} = \frac{\binom{n_1}{k_1} \cdots \binom{n_r}{k_r}}{\binom{n}{k}}. \end{aligned}$$

- Schritt 4: Dies kann auch als Verteilung $\mathbb{P}[(K_1, \ldots, K_r) = (k_1, \ldots, k_r)]$ aufgefasst werden, wobei

$$K_i = \sum_{\ell=1}^{k} \mathbb{1}_{\{i\}}(Y_\ell), \qquad i \in \{1, \ldots, r\}.$$

3.2.2 Definition (Hypergeometrische Verteilung)**.** Für alle $k, n_1, \dots, n_r \in \mathbb{N}$ mit $k \le n := n_1 + \dots + n_r$ ist die hypergeometrische Verteilung $\mathrm{Hyp}(k, n_1, \dots, n_r)$ die eindeutige Verteilung auf der Menge

$$\{(k_1, \dots, k_r); 0 \le k_i \le n_i, k_1 + \dots + k_r = k\}$$

mit Zähldichte

$$(k_1, \dots, k_r) \mapsto \frac{\binom{n_1}{k_1} \dots \binom{n_r}{k_r}}{\binom{n}{k}}.$$

3.2.3 Beispiel (Anwendung in der Qualitätssicherung)**.**

- In einer Lieferung sind fehlerhafte Produkte enthalten.
- Eine Stichprobe wird entnommen und getestet.
- Die Anzahl der fehlerhaften Produkte in der Stichprobe ist hypergeometrisch verteilt.
- Bei zu vielen fehlerhaften Produkten in der Stichprobe wird die Lieferung abgewiesen.

3.2.4 Proposition (Kenngrößen der hypergeometrischen Verteilung)**.** *Sei* $(K_1, \dots, K_r) \sim \mathrm{Hyp}(k, n_1, \dots, n_r)$*. Dann gilt für* $Y_1, \dots, Y_k$ *wie in Herleitung 3.2.1,* $i, j \in \{1, \dots, r\}$*,* $i \neq j$*,* $p_i = n_i/n$*,* $q_i = 1 - p_i$*, dass*

$$\begin{aligned}
\mathbb{E}[K_i] &= \sum_{\ell=1}^{k} \mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)] = kp_i, \\
\mathrm{Var}[K_i] &= \sum_{\ell=1}^{k} \mathrm{Var}[\mathbb{1}_{\{i\}}(Y_\ell)] + 2 \sum_{1 \le \ell < m \le k} \mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{i\}}(Y_m)] \\
&= kp_iq_i + 2 \sum_{1 \le \ell < m \le k} \big(\mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)\mathbb{1}_{\{i\}}(Y_m)] - \mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)]\mathbb{E}[\mathbb{1}_{\{i\}}(Y_m)]\big) \\
&= kp_iq_i + 2 \sum_{1 \le i < j \le n} \left(\frac{n_i}{n}\frac{n_i - 1}{n - 1} - \frac{n_i}{n}\frac{n_i}{n}\right) = \dots = kp_iq_i \left(1 - \frac{k-1}{n-1}\right), \\
\mathrm{Cov}[K_i, K_j] &= \sum_{\ell=1}^{k} \mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{j\}}(Y_\ell)] + 2 \sum_{1 \le \ell < m \le k} \mathrm{Cov}[\mathbb{1}_{\{i\}}(Y_\ell), \mathbb{1}_{\{j\}}(Y_m)] \\
&= -kp_ip_j + 2 \sum_{1 \le \ell < m \le k} \big(\mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)\mathbb{1}_{\{j\}}(Y_m)] - \mathbb{E}[\mathbb{1}_{\{i\}}(Y_\ell)]\mathbb{E}[\mathbb{1}_{\{j\}}(Y_m)]\big) \\
&= -kp_ip_j + 2 \sum_{1 \le i < j \le n} \left(\frac{n_i}{n}\frac{n_j}{n - 1} - \frac{n_i}{n}\frac{n_j}{n}\right) = \dots = -kp_ip_j \frac{n-k}{n-1}.
\end{aligned}$$

3.2.5 Bemerkung (Binomial- bzw. Multinomialapproximation)**.**

- Für $(K_1, \dots, K_r) \sim \mathrm{Hyp}(k, np_1, \dots, np_r)$ und $(X_1, \dots, X_r) \sim B(k, p_1, \dots, p_r)$ gilt

$$\mathbb{E}[K_i] = \mathbb{E}[X_i] = kp_i,$$

$$\lim_{N\to\infty} \operatorname{Cov}[K_i,K_j] = \operatorname{Cov}[X_i,X_j] = \begin{cases} kp_iq_i, & i=j, \\ -kp_ip_j, & i\neq j \end{cases}.$$

In diesem (und auch einem stärkeren) Sinn konvergiert die hypergeometrische gegen die multinomiale Verteilung.

- Die Approximation ist gut für k/n klein und p_i nahe bei $1/2$.

3.3 Poisson-Verteilung und Gesetz der kleinen Zahlen

3.3.1 Lemma (Gesetz der kleinen Zahlen; Poisson-Approximation)**.** *Für jedes $n \in \mathbb{N}$ sei $X_n \sim B(n,p_n)$ so dass*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} np_n = \lambda > 0.$$

Dann gilt für jedes $k \in \mathbb{N}$ dass

$$\lim_{n\to\infty} \mathbb{P}[X_n = k] = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Beweis.

$$\begin{aligned} \mathbb{P}[X_n = k] &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\to 1} \underbrace{(np_n)^k}_{\to\lambda^k} \underbrace{\left(1-\frac{np_n}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1-\frac{np_n}{n}\right)^{-k}}_{\to 1}. \end{aligned}$$

□

3.3.2 Definition (Poisson-Verteilung)**.** Für jedes $\lambda \in [0,\infty)$ ist die Poisson-Verteilung $\operatorname{Poi}(\lambda)$ die eindeutige Verteilung auf $\mathbb{N}$ mit Zähldichte

$$k \mapsto \frac{\lambda^k e^{-\lambda}}{k!}.$$

3.3.3 Bemerkung.

- Eigentlich müsste es Gesetz der (vielen) unwahrscheinlichen Zahlen heißen.
- Mit $\operatorname{Poi}(np)$ lässt sich leichter rechnen als mit $B(n,p)$, vor allem für große n.
- Bemerkung: Die Summe unabhängiger Poisson-Verteilungen ist Poisson-verteilt; siehe 4. Übungsblatt.

3.3.4 Beispiel (Anwendung in der Netzwerktechnik)**.**

- Jeder Netzwerkclient sendet mit Wahrscheinlichkeit p eine Anfrage an den Server und mit Wahrscheinlichkeit $1-p$ keine Anfrage.
- Die Clients sind unabhängig, es gibt sehr viele Clients, und p ist klein.
- Dann ist die Anzahl der Anfragen approximativ Poisson-verteilt.

3.3.5 Proposition (Kenngrößen der Poisson-Verteilung)**.** *Sei* $X \sim \mathrm{Poi}(\lambda)$*. Dann gilt*

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

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = \left(\sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} \right) + \lambda - \lambda^2 \\ &= \lambda^2 e^{-\lambda} \underbrace{\left(\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \right)}_{=e^{\lambda}} + \lambda - \lambda^2 = \lambda. \end{aligned}$$

3.4 Geometrische und Exponential-Verteilung

3.4.1 Herleitung (Wartezeit auf den ersten Erfolg)**.**

- Ziel: Berechnen der Verteilung der Wartezeit Y bis zum ersten Auftreten einer 1 in einer Folge $X_1, X_2, \dots \sim B(p)$ unabhängiger Bernoulli-verteilter Zufallsvariablen.
- Schritt 1: Wahrscheinlichkeitsraum für $X_1, X_2, \dots$. Wir setzen für alle $k \in \mathbb{N}$

$$\Omega = \big\{(x_1, x_2, \dots); x_i \in \{0,1\}\big\}, \qquad X = \mathrm{Id}_\Omega,$$
$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_k] = \prod_{i=1}^{k} p^{x_i}(1-p)^{1-x_i}.$$

- Schritt 2: Wartezeit Y. Wir setzen

$$h(x_1, x_2, \dots) = \min\{i \in \mathbb{N}_{>0}; x_i = 1\}, \qquad Y = h(X).$$

  Dann gilt für jedes $k \in \mathbb{N}_{>0}$ dass

$$\mathbb{P}[Y = k] = \mathbb{P}[X_1 = \dots = X_{k-1} = 0, X_k = 1] = (1-p)^{k-1}p.$$

3.4.2 Definition (Geometrische Verteilung)**.** Für jedes $p \in [0,1]$ ist die geometrische Verteilung $\mathrm{geo}(p)$ die eindeutige Verteilung auf $\mathbb{N}_{>0}$ mit Zähldichte

$$k \mapsto (1-p)^{k-1}p.$$

3.4.3 Proposition (Kenngrößen der geometrischen Verteilung)**.** *Sei* $X \sim \mathrm{geo}(p)$*. Dann gilt für alle* $k \in \mathbb{N}_{>0}$ *dass*

$$\mathbb{P}[X > k] = \sum_{\ell=k+1}^{\infty} (1-p)^{\ell-1}p = p(1-p)^k \underbrace{\sum_{\ell=0}^{\infty} (1-p)^\ell}_{=\frac{1}{p}} = (1-p)^k,$$

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\mathbb{P}[X = k] = \sum_{k=1}^{\infty} \sum_{\ell=1}^{k} \mathbb{P}[X = k] = \sum_{\ell=1}^{\infty} \sum_{k=\ell}^{\infty} \mathbb{P}[X = k]$$

$$= \sum_{\ell=1}^{\infty} \mathbb{P}[X \geq \ell] = \sum_{\ell=0}^{\infty} \mathbb{P}[X > \ell] = \frac{1}{p},$$

$$\mathbb{E}[X(X-1)] = 2\sum_{k=1}^{\infty} \frac{k(k-1)}{2}\mathbb{P}[X=k] = 2\sum_{k=1}^{\infty}\sum_{\ell=0}^{k-1} \ell\, \mathbb{P}[X=k] = 2\sum_{\ell=0}^{\infty}\sum_{k=\ell+1}^{\infty} \ell\, \mathbb{P}[X=k]$$

$$= 2\sum_{\ell=0}^{\infty} \ell\, \mathbb{P}[X>\ell] = 2\frac{1-p}{p}\sum_{\ell=0}^{\infty} \ell\, \mathbb{P}[X=\ell] = 2\frac{1-p}{p}\mathbb{E}[X] = 2\frac{1-p}{p^2}.$$

$$\mathrm{Var}[X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2 = 2\frac{1-p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

3.4.4 Lemma (Exponentialapproximation). *Für jedes $n \in \mathbb{N}$ sei $X_n \sim \mathrm{geo}(p_n)$, so dass*

$$\lim_{n\to\infty} np_n = \lambda \in (0,\infty).$$

Dann gilt für $Y_n = X_n/n$ und jedes $x \in [0,\infty)$ dass

$$\lim_{n\to\infty} \mathbb{P}[Y_n > x] = e^{-\lambda x} = \int_x^{\infty} \lambda e^{-\lambda x} dx.$$

Beweis.

$$\mathbb{P}[Y_n > x] = \mathbb{P}[X_n > nx] = (1-p_n)^{\lfloor nx \rfloor} = \left((1-p_n)^{\frac{1}{p_n}}\right)^{p_n \lfloor nx \rfloor} \to e^{-\lambda x}. \qquad \square$$

3.4.5 Definition (Exponentialverteilung). Für jedes $\lambda \in (0,\infty)$ ist die Exponentialverteilung $\mathrm{Exp}(\lambda)$ die eindeutige Verteilung auf $[0,\infty)$ mit Dichte

$$x \mapsto \lambda e^{-\lambda x}.$$

3.4.6 Proposition (Kenngrößen der Exponentialverteilung). *Sei $X \sim \mathrm{Exp}(\lambda)$. Dann gilt*

$$\mathbb{E}[X] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \left[x \cdot (-1)e^{-\lambda x}\right]_{x=0}^{x=\infty} - \int_0^{\infty} 1 \cdot (-1)e^{-\lambda x} dx = \frac{1}{\lambda},$$

$$\mathbb{E}[X^2] = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx = \left[x^2 \cdot (-1)e^{-\lambda x}\right]_{x=0}^{x=\infty} - \int_0^{\infty} 2x \cdot (-1)e^{-\lambda x} dx = \frac{2}{\lambda}\mathbb{E}[X] = \frac{2}{\lambda^2},$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

3.5 Normalverteilung

3.5.1 Definition. Für jedes $\mu \in \mathbb{R}$ und $\sigma \in (0,\infty)$ ist die Normalverteilung $N(\mu,\sigma^2)$ die eindeutige Verteilung auf $\mathbb{R}$ mit Dichte

$$x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

3.5.2 Bemerkung.

- Die obige Funktion ist eine Dichte weil

$$\begin{aligned}
\left(\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\right)^2 &= \left(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\exp\left(-\frac{x^2}{2}\right)dx\right)^2 \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2}\exp\left(-\frac{x^2+y^2}{2}\right)dxdy = \frac{1}{2\pi}\int_0^\infty\exp\left(-\frac{r^2}{2}\right)2r\pi dr \\
&= \left[-\exp\left(-\frac{r^2}{2}\right)\right]_{r=0}^{r=\infty} = 1.
\end{aligned}$$

- Wir behandeln Grenzwertsätze zur Normalverteilung in Abschnitt 4.

3.5.3 Lemma (Transformation auf Standardnormalverteilung)**.** *Wenn* $X \sim N(\mu,\sigma^2)$ *und* $Y = (X-\mu)/\sigma$, *dann gilt* $Y \sim N(0,1)$.

Beweis. Für alle $y \in \mathbb{R}$ gilt

$$\begin{aligned}
\mathbb{P}[Y \le y] = \mathbb{P}[X \le \mu + \sigma y] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\mu+\sigma y}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}\exp\left(-\frac{x^2}{2}\right)dx.
\end{aligned}$$

□

3.5.4 Proposition (Kenngrößen der Normalverteilung)**.** *Sei* $X \sim N(\mu,\sigma^2)$ *und* $Y = (X-\mu)/\sigma \sim N(0,1)$. *Dann gilt*

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} y\exp\left(-\frac{y^2}{2}\right)dy = \frac{1}{\sqrt{2\pi}}\left[-\exp\left(-\frac{y^2}{2}\right)\right]_{y=-\infty}^{y=\infty} = 0, \\
\operatorname{Var}[Y] &= \mathbb{E}[Y^2] = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} y\cdot y\exp\left(-\frac{y^2}{2}\right)dy \\
&= \frac{1}{\sqrt{2\pi}}\left[y\cdot(-1)\exp\left(-\frac{y^2}{2}\right)\right]_{y=-\infty}^{y=\infty} - \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(-1)\exp\left(-\frac{y^2}{2}\right)dy = 1, \\
\mathbb{E}[X] &= \mu + \sigma\mathbb{E}[Y] = \mu, \\
\operatorname{Var}[X] &= \sigma^2\operatorname{Var}[Y] = \sigma^2.
\end{aligned}$$

4 Approximationssätze

4.0.1 Bemerkung (Beispiele)**.**

- Schon gesehen: Binomialapproximation der hypergeometrischen Verteilung, Poissonapproximation der Binomialverteilung, Exponentialapproximation der geometrischen Verteilung, etc.

- In diesem Kapitel: Gesetz der großen Zahlen und zentraler Grenzwertsatz.

4.1 Konvergenzbegriffe

4.1.1 Definition. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine Zufallsvariable.

(i) X_n konvergiert fast sicher gegen X, geschrieben $X_n \stackrel{f.s.}{\to} X$, wenn

$$\mathbb{P}[\lim_{n\to\infty} X_n = X] = 1.$$

(ii) X_n konvergiert in Wahrscheinlichkeit gegen X, geschrieben $X_n \stackrel{p}{\to} X$, wenn

$$\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}[|X_n - X| > \epsilon] = 0.$$

(iii) X_n konvergiert in Verteilung gegen X, geschrieben $X_n \stackrel{d}{\to} X$, wenn für jede stetige beschränkte Funktion f gilt, dass

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

4.1.2 Lemma. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen und X eine Zufallsvariable.*

(i) $X_n \stackrel{f.s.}{\to} X \implies X_n \stackrel{p}{\to} X \implies X_n \stackrel{d}{\to} X.$

(ii) Seien X_n und X rellwertige Zufallsvariablen mit Verteilungsfunktionen F_n bzw. F. Dann gilt $X_n \stackrel{d}{\to} X$ genau dann, wenn

$$\forall x \in \mathbb{R} : (F \textit{ ist stetig bei } x) \implies \lim_{n\to\infty} F_n(x) = F(x).$$

4.1.3 Bemerkung.

- Überblick, weitere Information: `en.wikipedia.org/wiki/Convergence_of_random_variables`.
- Beweis: Vorlesung Maß- und Integrationstheorie.

4.2 Ungleichungen

4.2.1 Lemma (Ungleichungen)**.** *Sei X eine $\mathbb{R}_+$-wertige Zufallsvariable und $\epsilon > 0$.*

(i) Markov:

$$\mathbb{P}[X \geq \epsilon] \leq \frac{1}{\epsilon}\mathbb{E}[X].$$

(ii) Tschebychev: Wenn $\mu := \mathbb{E}[X] < \infty$, dann gilt

$$\mathbb{P}[|X - \mu| \geq \epsilon] \leq \frac{\mathrm{Var}[X]}{\epsilon^2}.$$

Beweis. (i): Aus $X \geq \epsilon \mathbb{1}_{X\geq\epsilon}$ und der Monotonie des Erwartungswerts folgt

$$\epsilon\mathbb{P}[X \geq \epsilon] = \mathbb{E}[\epsilon\mathbb{1}_{X\geq\epsilon}] \leq \mathbb{E}[X].$$

(ii): Wende Markov auf $(X-\mu)^2$ an:

$$\mathbb{P}[|X - \mu| \geq \epsilon] = \mathbb{P}[(X-\mu)^2 \geq \epsilon^2] \leq \frac{\mathbb{E}[(X-\mu)^2}{\epsilon^2} = \frac{\mathrm{Var}[X]}{\epsilon^2}.$$

□

4.3 Gesetz der großen Zahlen

4.3.1 Theorem (Gesetz der großen Zahlen)**.** *Seien $X_1, X_2, \dots$ reellwertige, unabhängige, identisch verteilte Zufallsvariablen mit $\mu := \mathbb{E}[X_1] < \infty$ und sei*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

das empirische Mittel von $X_1, \dots, X_n$.

(i) *Schwaches Gesetz der großen Zahlen: $\overline{X}_n$ konvergiert in Wahrscheinlichkeit gegen μ, d.h.,*

$$\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left[\left|\overline{X}_n - \mu\right| \geq \epsilon\right] = 0.$$

(ii) *Starkes Gesetz der großen Zahlen: $\overline{X}_n$ konvergiert fast sicher gegen μ, d.h.,*

$$\mathbb{P}\left[\lim_{n\to\infty} \overline{X}_n = \mu\right] = 1.$$

4.3.2 Bemerkung.

- Kernaussage: Das empirisches Mittel konvergiert gegen den Erwartungswert.
- Dies rechtfertigt die Namensgebung des Mittelwerts gleich Erwartungswerts.
- Das schwache Gesetz der großen Zahlen sagt, dass für große n das Ereignis $|\overline{X}_n - \mu| > \epsilon$ sehr unwahrscheinlich ist. Es kann aber mit positiver Wahrscheinlichkeit passieren, dass $|\overline{X}_n - \mu| > \epsilon$ für unendlich viele n.
- Das starke Gesetz der großen Zahlen sagt, dass dies nicht der Fall sein kann. Es impliziert das schwache Gesetz der großen Zahlen.

Beweis von Theorem 4.3.1. Wir zeigen nur (i) unter der zusätzlichen Annahme, dass $\sigma^2 := \mathrm{Var}[X_1] < \infty$. Wegen Chebychev gilt

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\overline{X}_n]}{\epsilon^2} = \frac{\mathrm{Var}[X_1]}{n\epsilon^2} \to 0. \qquad \square$$

4.3.3 Bemerkung (Anwendung in Monte-Carlo Methoden)**.**

- Ziel: Berechnen von $\mathbb{E}[g(X)]$.
- Lösung: Man erzeugt unabhängige Zufallsvariablen X_n, $n \in \mathbb{N}$, mit der gleichen Verteilung wie X und setzt $Y_n = g(X_n)$. Dann konvergiert $\overline{Y}_n$ gegen $\mathbb{E}[f(X)]$.
- Bewertung: Das Verfahren ist attraktiv, wenn es einfach ist, X_n zu erzeugen, jedoch schwierig, $\sum_x g(x)f(x)$ bzw. $\int g(x)f(x)dx$ auszuwerten. Beispielsweise ist X hochdimensional oder die (Zähl-)dichte f komplex zu berechnen.
- Konvergenzrate: $n^{-1/2}$ (zentraler Grenzwertsatz, siehe Abschnitt 4.4).

4.3.4 Beispiel (Münzwurf)**.**

- Ziel: Bestimmen der Wahrscheinlichkeit p, dass eine Münze auf Kopf fällt.
- Lösung: Die Münze wird wiederholt geworfen. Für jedes $n \in \mathbb{N}$ sei $X_n = 1$ bei Kopf und $X_n = 0$ bei Zahl. Dann konvergiert $\overline{X}_n$ gegen $\mathbb{E}[X_1] = \mathbb{P}[X_1 = 1] = p$. Also ist $\overline{X}_n$ für großes n ein guter Schätzer für p.

4.4 Zentraler Grenzwertsatz

4.4.1 Theorem (Zentraler Grenzwertsatz)**.** *Seien X_n und $\overline{X}_n$ wie in Theorem 4.3.1, $\sigma^2 := \mathrm{Var}[X_1] < \infty$ und*

$$Z_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}, \qquad\qquad Z \sim N(0,1).$$

Dann konvergiert Z_n gegen Z in Verteilung, d.h.,

$$\forall -\infty \le c < d \le \infty : \lim_{n\to\infty} \mathbb{P}\big[Z_n \in [c,d]\big] = \mathbb{P}\big[Z \in [c,d]\big].$$

4.4.2 Bemerkung.

- Kernaussage: Die stochastischen Fluktuationen von $\overline{X}_n$ um μ sind asymptotisch normalverteilt (Konvergenz in Verteilung). Die Konvergenzrate ist $n^{-1/2}$.
- Dies ist einer der Gründe für die Wichtigkeit und häufige Verwendung der Normalverteilung.

Beweis von Theorem 4.4.1. Wir beweisen die Aussage nur für Bernoulli Zufallsvariablen $X_1, X_2, \dots \sim B(p)$. In dem Fall gilt $\mu = p$ und $\sigma^2 = p(1-p) =: pq$. Für alle $k \in \{0, \dots, n\}$ gilt

$$(*) := \mathbb{P}[X_1 + \cdots + X_n = k] = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}.$$

Die Stirling'sche Formel

$$\lim_{n\to\infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1$$

gibt uns die Approximation

$$\begin{aligned}(*) &\approx \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} \cdot \left(\frac{n-k}{e}\right)^{n-k} \sqrt{2\pi(n-k)}} p^k q^{n-k} = \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \\ &= \sqrt{\frac{n}{2\pi k(n-k)}} \left(\left(\frac{k}{np}\right)^{\frac{k}{n}} \left(\frac{n-k}{nq}\right)^{\frac{n-k}{n}}\right)^{-n} = \sqrt{\frac{n}{2\pi k(n-k)}} \exp\left(-n\eta\left(\frac{k}{n}\right)\right),\end{aligned}$$

wobei

$$\eta(t) = t \log\left(\frac{t}{p}\right) + (1-t)\log\left(\frac{1-t}{q}\right).$$

Die Taylor-Approximation

$$\eta(t) \approx \eta(p) + \eta'(p)(t-p) + \frac{1}{2}\eta''(p)(t-p)^2 = \frac{1}{2pq}(t-p)^2$$

gibt

$$(*) \approx \sqrt{\frac{n}{2\pi k(n-k)}} \exp\left(-n\frac{1}{2pq}\left(\frac{k}{n} - p\right)^2\right) = \sqrt{\frac{n}{2\pi k(n-k)}} \exp\left(-\frac{z_{k,n}^2}{2}\right),$$

wobei

$$z_{k,n} = \sqrt{n}\frac{k/n-\mu}{\sigma} = \sqrt{n}\frac{k/n-p}{\sqrt{pq}}.$$

Für beschränktes $z_{k,n} \in [c,d]$ gilt

$$\frac{k}{n} = p + z_{k,n}\sqrt{\frac{pq}{n}} \approx p, \qquad \frac{n-k}{n} = 1 - \frac{k}{n} \approx q,$$

und daher

$$(*) \approx \sqrt{\frac{1}{2\pi npq}} \exp\left(-\frac{z_{k,n}^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_{k,n}^2}{2}\right)(z_{k+1,n} - z_{k,n}).$$

Daher gilt (Konvergenz einer Riemannsumme gegen das Integral)

$$\begin{aligned}\mathbb{P}\big[Z_n \in [c,d]\big] &= \sum_{k:\, z_{k,n}\in[c,d]} \mathbb{P}[X_1 + \cdots + X_n = k] \\ &\approx \frac{1}{\sqrt{2\pi}} \int_c^d \exp\left(-\frac{z^2}{2}\right) dz = \mathbb{P}\big[Z \in [c,d]\big].\end{aligned}$$

□

5 Markov-Ketten

Siehe [Pfa13, Kapitel 5].

6 Kausale Inferenz

Dieses Kapitel basiert auf [Pea09; PGJ16].

6.1 Kausalität versus bedingte Wahrscheinlichkeiten

6.1.1 Beispiel (Simpsons Paradoxon)**.** Eine Umfrage unter zuckerkranken und nicht zuckerkranken Patienten zur Wirkung von homöopathischen Kopfwehtabletten ergibt folgendes Ergebnis:

Weniger Kopfweh ($Y=1$)	Tabletten ($X=1$)	Keine Tabletten ($X=0$)
Zuckerkrank ($Z=1$)	$192/263 = 73\%$	$55/80 = 69\%$
Nicht zuckerkrank ($Z=0$)	$81/87 = 93\%$	$234/270 = 87\%$
Gesamt	$273/350 = 78\%$	$289/350 = 83\%$

In jeder Untergruppe (zuckerkrank vs. nicht zuckerkrank) scheinen die Pillen positiv zu wirken, insgesamt jedoch negativ. Wie ist das möglich? Wir werden sehen, dass der kausale Effekt von X auf Y nicht alleine durch die obigen Wahrscheinlichkeiten bestimmt ist, sondern von zusätzlicher Information abhängt, die sich durch kausale Modelle beschreiben lässt.

6.2 Kausale Modelle

Kausale Modelle werden als gerichtete Graphen dargestellt.

6.2.1 Beispiel (Simpson's Paradoxon, 1. Modell)**.** Das kausale Modell

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X \arrow[rr] & & Y \end{tikzcd}$$

bedeutet, dass es Funktionen f_X, f_Y, f_Z und unabhängige Zufallsvariablen U_X, U_Y, U_Z gibt, so dass

$$X = f_X(Z, U_X), \qquad Y = f_Y(X, Z, U_Y), \qquad Z = f_Z(U_Z).$$

Die Interpretation ist, dass Zuckerkranke im Vergleich zu nicht Zuckerkranken unterschiedlich stark (laut Daten: stärker) nach den zuckerhaltigen Tabletten verlangen. Sowohl Zuckerkrankheit als auch Tabletten haben einen Einfluss auf die Heilung von Kopfschmerzen.

6.2.2 Beispiel (Simpson's Paradoxon, 2. Modell)**.** Das kausale Modell

$$\begin{tikzcd} & Z \arrow[dr] & \\ X \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

bedeutet, dass es Funktionen f_X, f_Y, f_Z und unabhängige Zufallsvariablen U_X, U_Y, U_Z gibt, so dass

$$X = f_X(U_X), \qquad Y = f_Y(X, Z, U_Y), \qquad Z = f_Z(X, U_Z).$$

Die Interpretation ist, dass die Tabletten einen (laut Daten negativen) Einfluss auf Diabetes haben, was indirekt Einfluss auf die Heilung von Kopfschmerzen hat. Zusätzlich gibt es noch den direkten Effekt der Tabletten auf Kopfschmerzen.

6.2.3 Bemerkung.

- X, Y, Z heißen endogene und U_X, U_Y, U_Z exogene Variablen.
- Jede Menge von Zufallsvariablen hat eine Beschreibung als ein kausales Modell mit einem vollständigen Graphen. Das Fehlen von Kanten im Graph entspricht bestimmten Unabhängigkeitseigenschaften.

6.3 Interventionen

Interventionen sind fiktive Veränderungen am kausalen Modell, die bestimmte Variablen zu einer Konstante fixieren und alle anderen Variablen belassen wie sie sind.

6.3.1 Beispiel (Intervention $\mathrm{do}(X = x)$ im 1. Modell)**.** Die Intervention $\mathrm{do}(X = x)$ löscht im kausalen Modell alle zu X führenden Kanten und ersetzt X durch die Konstante x. Das Ergebnis ist das kausale Modell

$$\begin{tikzcd} & Z \arrow[dr] & \\ x \arrow[rr] & & Y \end{tikzcd}$$

mit der Bedeutung, dass f_Y, f_Z, U_Y, U_Z gleich sind wie vor der Intervention und

$$X = x, \qquad Y = f_Y(x, Z, U_Y), \qquad Z = f_Z(U_Z).$$

Die Wahrscheinlichkeiten in dem kausalen Modell nach der Intervention werden mit $\mathbb{P}[\cdot|\, \mathrm{do}(X = x)]$ bezeichnet.

6.3.2 Beispiel (Intervention $\mathrm{do}(X = x)$ im 2. Modell)**.** Bei der Intervention $\mathrm{do}(X = x)$ müssen im 2. Modell keine Kanten gelöscht werden, und man erhält das kausale Modell

$$\begin{tikzcd} & Z \arrow[dr] & \\ x \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

mit der Bedeutung, dass f_Y, f_Z, U_Y, U_Z gleich sind wie vor der Intervention und

$$X = x, \qquad Y = f_Y(x, Z, U_Y), \qquad Z = f_Z(x, U_Z).$$

Die Wahrscheinlichkeiten in dem kausalen Modell nach der Intervention werden mit $\mathbb{P}[\cdot|\, \mathrm{do}(X = x)]$ bezeichnet.

6.4 Kausale Effekte

Kausale Effekte beschreiben Unterschiede zwischen fiktiven Interventionen. Sie werden auch *ceteris paribus Analysen* genannt.

6.4.1 Beispiel (Durschnittlicher Behandlungseffekt im 1. Modell)**.** Der *durchschnittliche Behandlungseffekt* von X auf Y ist definiert als

$$\mathbb{P}[Y = 1|\, \mathrm{do}(X = 1)] - \mathbb{P}[Y = 1|\, \mathrm{do}(X = 0)].$$

Dieser Effekt kann mit Hilfe von bedingten Wahrscheinlichkeiten berechnet werden,

$$\begin{aligned} \mathbb{P}[Y = 1|\, \mathrm{do}(X = x)] &= \sum_{z \in \{0,1\}} \mathbb{P}[Y = 1|\, \mathrm{do}(X = x), Z = z]\mathbb{P}[Z = z|\, \mathrm{do}(X = x)] \\ &= \sum_{z \in \{0,1\}} \mathbb{P}[Y = 1|X = x, Z = z]\mathbb{P}[Z = z], \end{aligned}$$

wie folgende Nebenrechnung mit der Abkürzung $\mathbb{P}_x = \mathbb{P}[\cdot|\, \mathrm{do}(X = x)]$ und unter Verwendung von [Pfa13, Lemma 5.2.2] zeigt:

$$\begin{aligned} &\mathbb{P}[Y = 1|\, \mathrm{do}(X = x), Z = z] \\ &\qquad = \mathbb{P}_x[f_Y(x, Z, U_Y) = 1|Z = z] && \text{(Def. von } Y \text{ und } \mathbb{P}_x) \\ &\qquad = \mathbb{P}_x[f_Y(x, z, U_Y) = 1] && (U_Y \perp\!\!\!\perp Z \text{ unter } \mathbb{P}_x \\ &\qquad = \mathbb{P}[f_Y(x, z, U_Y) = 1] && (U_Y \text{ hat dieselbe Vert. unter } \mathbb{P} \text{ und } \mathbb{P}_x) \\ &\qquad = \mathbb{P}[f_Y(x, z, U_Y) = 1|X = x, Z = z] && (U_Y \perp\!\!\!\perp (X, Z) \text{ unter } \mathbb{P}) \\ &\qquad = \mathbb{P}[Y = 1|X = x, Z = z] && \text{(Def. von } Y\text{),} \\ &\mathbb{P}[Z = z|\, \mathrm{do}(X = x)] \\ &\qquad = \mathbb{P}_x[f_Z(U_Z) = z] && \text{(Def. von } Z \text{ und } \mathbb{P}_x) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}[f_Z(U_Z) = z] && (U_Z \text{ hat dieselbe Vert. unter } \mathbb{P} \text{ und } \mathbb{P}_x) \\
&= \mathbb{P}[Z = z] && (\text{Def. von } Z).
\end{aligned}
$$

Es werden also die Wahrscheinlichkeiten aus den *Untergruppen* zuckerkrank/nicht zuckerkrank herangezogen, um den Behandlungseffekt zu bestimmen. Gemäß der Angaben in den ersten beiden Zeilen der Tabelle in Beispiel 6.2.1 haben die Tabletten also eine positive Wirkung.

6.4.2 Beispiel (Durschnittlicher Behandlungseffekt im 2. Modell)**.** Da die Intervention im 2. Modell keine Kanten löscht, gilt

$$\mathbb{P}[Y = 1 | \operatorname{do}(X = x)] = \mathbb{P}[Y = 1 | X = x].$$

Dies kann auch aus folgender Nebenrechnung gesehen werden, wobei wir wieder $\mathbb{P}_x = \mathbb{P}[\cdot | \operatorname{do}(X = x)]$ setzen und [Pfa13, Lemma 5.2.2] verwenden:

$$
\begin{aligned}
&\mathbb{P}[Y = 1 | \operatorname{do}(X = x)] \\
&\quad = \mathbb{P}_x[f_Y(x, f_Z(x, U_Z), U_Y) = 1] && (\text{Def. von } Y, Z \text{ und } \mathbb{P}_x) \\
&\quad = \mathbb{P}[f_Y(x, f_Z(x, U_Z), U_Y) = 1] && ((U_Y, U_Z) \text{ hat dieselbe Vert. unter } \mathbb{P} \text{ und } \mathbb{P}_x) \\
&\quad = \mathbb{P}[f_Y(x, f_Z(x, U_Z), U_Y) = 1 | X = x] && (X \perp\!\!\!\perp (U_Y, U_Z) \text{ under } \mathbb{P}). \\
&\quad = \mathbb{P}[Y = 1 | X = x] && (\text{Def. von } Y, Z).
\end{aligned}
$$

Es werden also die *aggregierten* Wahrscheinlichkeiten herangezogen, um den Behandlungseffekt zu bestimmen. Gemäß der Angaben in der dritten Zeile der Tabelle in Beispiel 6.2.1 haben die Tabletten also eine negative Wirkung.

6.4.3 Bemerkung.

- Die Beispiele zeigen, dass statistische Information (d.h. die Daten aus Tabelle Beispiel 6.2.1) durch kausale Annahmen ergänzt werden muss, um einen kausalen Effekt zu identifizieren.
- Diese Identifizierungsannahmen sind oft der am schwierigsten lesbare Teil ökonometrischer Arbeiten, lassen sich jedoch mit gerichteten Graphen einfach und präzise darstellen.

7 Statistik

7.1 Schätzprobleme

Die folgende Definition sollte parallel zu den darauffolgenden Beispielen gelesen werden.

7.1.1 Definition.

(i) Ein *statistisches Modell* ist eine Zufallsvariable X, deren Verteilung $\mathbb{P}_\theta$ von einem Parameter $\theta \in \Theta$ abhängt, wobei Θ eine Menge ist.

(ii) Eine *Statistik* ist eine (messbare) Funktion von X. Ein *Schätzer* für eine Funktion $m : \Theta \to \Theta'$ ist eine Θ'-wertige Statistik.

(iii) Ein Schätzer $h(X)$ für $m : \Theta \to \Theta'$ heißt *unverzerrt* (=erwartungstreu) falls

$$\forall \theta \in \Theta : \mathbb{P}_\theta[h(X)] = m(\theta).$$

(iv) Eine Folge $h_n(X)$ von Schätzern für $m : \Theta \to \Theta'$ heißt *konsistent* falls

$$\forall \theta \in \Theta, \forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}_\theta[|h_n(X) - m(\theta)| > \epsilon] = 0.$$

(v) Sei $x \mapsto f_\theta(x)$ die Dichte bzw. Zähldichte von X. Dann heißt die Funktion $(\theta, x) \mapsto f_\theta(x)$ *Likelihood*. Ein Schätzer $h(X)$ heißt *Maximum-Likelihood-Schätzer* für θ falls

$$\forall x \in E : h(x) \in \arg\max_{\theta\in\Theta} f_\theta(x).$$

7.1.2 Beispiel (Schätzen des Parameters p der Bernoulli-Verteilung)**.**

(i) Statistisches Modell: $X = (X_1, X_2, \dots)$, $X_i \sim B(p)$ sind identisch und unabhängig verteilt unter $\mathbb{P}_p$, $\Theta = [0, 1]$, $\theta = p$.

(ii) Der empirische Mittelwert $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ ist ein Schätzer für p, bzw. genauer gesagt für $m = \mathrm{Id}_\Theta$.

(iii) $\overline{X}_n$ ist unverzerrt für p, weil $\mathbb{E}[\overline{X}_n] = p$.

(iv) $\overline{X}_n$ ist konsistent für p, weil $\lim_{n\to\infty} \mathbb{E}[|\overline{X}_n - p| > \epsilon] = 0$ dank des Gesetzes der großen Zahlen.

(v) $\overline{X}_n$ ist ein Maximum-Likelihood-Schätzer für p: Maximierung über Likelihood und Log-Likelihood ist äquivalent und

$$\begin{aligned}
\log f_p(x_1, \dots, x_n) &= \log\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right) \\
&= \sum_{i=1}^n x_i \log(p) + \sum_{i=1}^n (1 - x_i)\log(1-p), \\
\frac{d}{dp}\log\big(f_p(x_1, \dots, x_n)\big) &= \frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1-x_i)}{1-p} = \frac{\sum_{i=1}^n x_i - pn}{(1-p)}.
\end{aligned}$$

Ableitung Null setzen liefert $\hat{p}_{ML} = \overline{X}_n$. Das Maximum ist eindeutig (Kurvendiskussion).

7.1.3 Beispiel (Schätzen der Parameter μ und σ^2 der Normalverteilung)**.**

(i) Statistisches Modell: $X = (X_1, X_2, \dots)$, $X_i \sim N(\mu, \sigma^2)$ sind identisch und unabhängig verteilt under $\mathbb{P}_{\mu,\sigma^2}$, $\Theta = \mathbb{R} \times (0, \infty)$, $\theta = (\mu, \sigma^2)$.

(ii) Die empirische Varianz $s_n(X) := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ ist ein Schätzer für σ^2.

(iii) $s_n(X)$ ist unverzerrt für σ^2 weil

$$\begin{aligned}
\mathbb{E}_\theta[s_n^2(X)] &= \frac{n}{n-1}\Big(\mathbb{E}[(X_1 - \mu)^2] - 2\mathbb{E}[(X_1 - \mu)(\overline{X}_n - \mu)] + \mathbb{E}[(\overline{X}_n - \mu)^2]\Big) \\
&= \frac{n}{n-1}\Big(\sigma^2 - \frac{2}{n}\sigma^2 - \sigma^2\Big) = \sigma^2.
\end{aligned}$$

(iv) $s_n^2(X)$ ist konsistent für σ^2, weil

$$\begin{aligned}\frac{n-1}{n}s_n^2(X) &= \frac{1}{n}\sum_{i=1}^n \big((X_i-\mu)-(\overline{X}_n-\mu)\big)^2\\ &= \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2 - \frac{2}{n}\sum_{i=1}^n (X_i-\mu)(\overline{X}_n-\mu) + \frac{1}{n}\sum_{i=1}^n (\overline{X}_n-\mu)\\ &= \underbrace{\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2}_{\xrightarrow{p}\sigma^2} - \underbrace{(\overline{X}_n-\mu)^2}_{\xrightarrow{p}0} \xrightarrow{p} \sigma^2.\end{aligned}$$

(v) Das Paar $\overline{X}_n$ und $\frac{n-1}{n}s_n^2(X)$ ist ein Maximum-Likelihood-Schätzer für (μ,σ^2):

$$\begin{aligned}\log\big(f_{\mu,\sigma^2}(x_1,\dots,x_n)\big) &= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}\right)\right)\\ &= -n\log(\sigma) - \sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2} + C,\end{aligned}$$

wobei die Konstante nicht von μ und σ abhängt. Daher gilt

$$\begin{aligned}\frac{\partial}{\partial\mu}\log\big(f_{\mu,\sigma^2}(x_1,\dots,x_n)\big) &= \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2},\\ \frac{\partial}{\partial\sigma}\log\big(f_{\mu,\sigma^2}(x_1,\dots,x_n)\big) &= -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^3}.\end{aligned}$$

Ableitung Null setzen liefert

$$\hat{\mu}_{ML} = \overline{X}_n, \qquad \hat{\sigma}_{ML} = \frac{1}{n}\sum_{i=1}^n (X_i-\hat{\mu}_{ML})^2 = \frac{n-1}{n}s_n^2(X).$$

7.1.4 Bemerkung.

- Frage: Ist X_1 ein Schätzer für den Erwartungswert? Warum verwenden wir $\overline{X}_n$ und nicht X_1?
- Frage: Fallen Ihnen neben $s_n^2(X)$ noch weitere Schätzer für die Varianz ein? Welche Eigenschaften haben sie.

7.2 Testprobleme

Die folgende Definition sollte parallel zu den darauffolgenden Beispielen gelesen werden.

7.2.1 Definition. Sei X und $(\mathbb{P}_\theta)_{\theta\in\Theta}$ ein statistisches Modell.

(i) Ein *statistischer Test* ist ein Schätzer für eine Funktion $m:\Theta\to\{0,1\}$. Wir setzen $\Theta_0 = m^{-1}(\{0\})$, $\Theta_1 = m^{-1}(\{1\})$ und führen folgende Begriffe ein:

$$\begin{array}{r|l} \theta \in \Theta_0 & \text{Nullhypothese } H_0 \\ \theta \in \Theta_1 & \text{Alternativhypothese } H_1 \\ T = 0 & H_0 \text{ angenommen, } H_1 \text{ verworfen} \\ T = 1 & H_1 \text{ angenommen, } H_0 \text{ verworfen} \\ \theta \in \Theta_0 \wedge T = 1 & \text{Fehler 1. Art} \\ \theta \in \Theta_1 \wedge T = 0 & \text{Fehler 2. Art} \end{array}$$

(ii) Der Test hat *Signifikanzniveau* (=Signifikanz, Niveau) $\alpha \in [0, 1]$ falls

$$\underbrace{\sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T = 1]}_{\text{Fehler 1. Art}} \leq \alpha$$

und *Teststärke* (=Power, Sensitivität) $1 - \beta$ falls

$$\underbrace{\sup_{\theta \in \Theta_1} \mathbb{P}_\theta[T = 0]}_{\text{Fehler 2. Art}} \leq \beta.$$

(iii) Der Test heißt *unverfälscht* falls

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T = 1] \leq \inf_{\theta \in \Theta_1} \mathbb{P}_\theta[T = 1].$$

(iv) Für jedes $\alpha \in [0, 1]$ sei T_α ein Test zum Signifikanzniveau α, so dass $T_\alpha \leq T_\beta$ für alle $\alpha \leq \beta$. Dann ist der *p-Wert* die Statistik

$$p = \inf\{\alpha \in [0, 1]; T_\alpha = 1\}.$$

7.2.2 Bemerkung.

- Frage: Wünscht man sich kleines oder großes Signifikanzniveau und kleine oder große Teststärke, um die H_0 zu widerlegen?
- Signifikanz von 5% heißt, dass H_0 in maximal 5% der Fälle zu unrecht verworfen wird. Salopp ausgedrückt: $\mathbb{P}[T = 1 | H_0] \leq 5\%$.
- Signifikanz von 5% heißt *nicht*, dass in 95% der signifikanten Testergebnisse H_1 gilt; in vielen Situationen sind es unter 60% (vgl. Aufgabe 2 auf Übungsblatt 10.) Salopp ausgedrückt: $\mathbb{P}[H_0 | T = 1] \not\leq 5\%$.
- Der p-Wert ist eine Statistik, also eine Funktionen des Datensatzes X. Er ist unter relativ allgemeinen Voraussetzungen $U([0, 1])$-verteilt unter H_0. Interpretation: Wenn wir Daten mit niedrigem p-Wert extrem nennen, gibt es unter H_0 für jeden Datensatz mit p-Wert λ einen Anteil λ von mindestens so extremen Daten.
- Der Test $\mathbb{1}_{p \leq \alpha}$ ist (bis auf Rundungsfehler) gleich T_α. Genauer gesagt gilt, dass $\mathbb{1}_{p \leq \alpha}$ gleich der rechtsstetigen Modifikation $\lim_{\beta \downarrow \alpha} T_\beta$ von T_α ist. Interpretation: für jeden gegebenen p-Wert kann H_0 mit Signifikanzniveau p verworfen werden.

7.2.3 Beispiel (Einseitiger Test für den Parameter p der Binomialverteilung)**.** Seien X und $(\mathbb{P}_p)_{p \in [0,1]}$ wie in Beispiel 7.1.2.

(i) Sei $m = \mathbb{1}_{(1/2,1]}$, $\Theta_0 = [0,1/2]$, $\Theta_1 = (1/2,1]$, $\alpha = 5\%$, $n = 100$, $Y = \sum_{i=1}^n X_i$ und $q_{n,1-\alpha} = 58$ das $(1-\alpha)$-Quantil von $B(n,1/2)$. Dann ist

$$T_\alpha = \mathbb{1}_{\{Y > q_{n,1-\alpha}\}}$$

ein Test von $H_0 : p \in [0,1/2]$ gegen $H_1 : p \in (1/2,1]$. Der Test verwirft die Nullhypothese falls $Y > 58$.

(ii) Bis auf Rundungsfehler hat der Test Signifikanzniveau 5% und Stärke 5%, weil

$$\begin{aligned}\sup_{p\le 1/2} \mathbb{P}_\theta[T_\alpha = 1] = \mathbb{P}_{1/2}[T_\alpha = 1] \approx 5\%,\\ \sup_{p>1/2} \mathbb{P}_\theta[T_\alpha = 0] = \mathbb{P}_{1/2}[T_\alpha = 0] \approx 95\%.\end{aligned}$$

(iii) Bis auf Rundungsfehler ist der Test unverfälscht, weil

$$\sup_{p\le 1/2} \mathbb{P}_p[T_\alpha = 1] \approx 5\% \approx \inf_{p>1/2} \mathbb{P}_p[T_\alpha = 1].$$

(iv) Sei F die Verteilungsfunktion von $B(n,1/2)$. Dann gilt

$$\begin{aligned} p &= \inf\{\alpha \in [0,1]; T_\alpha = 1\} = \inf\{\alpha \in [0,1]; Y > q_{n,1-\alpha}\}\\ &\approx \inf\{\alpha \in [0,1]; F(Y) > 1-\alpha\} = 1 - F(Y).\end{aligned}$$

Der p-Wert ist also, so wie die Tests $(T_\alpha)_{\alpha\in[0,1]}$, eine Statistik. Unter $\mathbb{P}_{1/2}$ ist der p-Wert uniform $U([0,1])$-verteilt. Der Test T_α verwirft die Nullhypothese genau dann wenn $p < \alpha$.

7.2.4 Bemerkung.

- Normalapproximation des Binomialtests: Nach dem zentralen Grenzwertsatz gilt für große n approximativ $B(n,1/2) \approx nN(1/2,1/4) = N(n/2,n^2/4)$. Demgemäß kann in Beispiel 7.2.3 das $(1-\alpha)$-Quantil von $B(n,1/2)$ durch das $(1-\alpha)$-Quantil von $N(n/2,n^2/4)$ approximiert werden.

7.2.5 Definition. Für jedes $n \in \mathbb{N}_{>0}$ ist die *Student* t-*Verteilung* $\mathrm{t}(n-1)$ die Verteilung der $\mathbb{R}$-wertigen Zufallsvariable

$$\frac{\overline{X}_n}{s_n(X)/\sqrt{n}},$$

wobei $X_1,\dots,X_n \sim N(0,1)$ unabhängig identisch verteilt sind, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ und $s_n^2(X) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

7.2.6 Lemma. *Sei $n \in \mathbb{N}_{>0}$ und seien $X_1,\dots,X_n \sim N(\mu,\sigma^2)$ unabhängig identisch verteilt.*

(i) Es gilt

$$\frac{\overline{X}_n - \mu}{s_n(X)/\sqrt{n}} \sim \mathrm{t}(n-1).$$

(ii) Im obigen Bruch sind Zähler und Nenner unabhängig.

Beweis. (i) $Y_i = (X_i - \mu)/\sigma \sim N(0,1)$ sind unabhängig identisch verteilt und

$$\frac{\overline{X}_n - \mu}{s_n(X)/\sqrt{n}} = \frac{\overline{Y}_n}{s_n(Y)/\sqrt{n}} \sim \mathrm{t}(n-1).$$

(ii) Die Zufallsvariablen $\overline{X}_n, X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n$ sind unabhängig, da sie normalverteilt und paarweise unkorreliert sind, wie man leicht nachrechnet.

□

7.2.7 Beispiel (Zweiseitiger Test für den Parameter μ der Normalverteilung)**.** Seien X und $(\mathbb{P}_{\mu,\sigma^2})_{\mu\in\mathbb{R},\sigma^2\in(0,\infty)}$ wie in Beispiel 7.1.3.

(i) Sei $\mu^* \in \mathbb{R}$, $m = \mathbb{1}_{\mathbb{R}\setminus\{\mu^*\}}$, $\Theta_0 = \{\mu^*\}$, $\Theta_1 = \mathbb{R}\setminus\{\mu^*\}$, $\alpha = 5\%$, $n = 100$, $Y = \frac{\overline{X}_n - \mu^*}{s_n(X)/\sqrt{n}}$ und $t_{n-1,\alpha/2} \approx -1.98$ bzw. $t_{n-1,1-\alpha/2} \approx 1.98$ das $(\alpha/2)$-Quantil bzw. $(1-\alpha/2)$-Quantil der Student $\mathrm{t}(n-1)$-Verteilung. Dann ist

$$T_\alpha = \mathbb{1}_{\{(Y<t_{n-1,\alpha/2})\vee(Y>t_{n-1,1-\alpha/2})\}}$$

ein Test von $H_0 : \mu = \mu^*$ gegen $H_1 : \mu \neq \mu^*$. Der Test verwirft die Nullhypothese falls die Teststatistik Y außerhalb des Intervalls $[-1.98, 1.98]$ liegt.

(ii) Der Test hat Signifikanzniveau 5% und Stärke 5%, weil

$$\sup_{\sigma^2>0} \underbrace{\mathbb{P}_{\mu^*,\sigma^2}[T_\alpha = 1]}_{=5\%} = 5\%, \qquad \sup_{\mu\neq\mu^*,\sigma^2>0} \mathbb{P}_{\mu,\sigma^2}[T_\alpha = 0] = \sup_{\sigma^2>0} \underbrace{\mathbb{P}_{\mu^*,\sigma^2}[T_\alpha = 0]}_{=95\%} = 95\%.$$

Werte von μ, die strikt von μ^* weg bleiben, werden mit größerer Stärke erkannt.

(iii) Der Test ist unverfälscht, weil

$$\sup_{\mu=\mu^*,\sigma^2>0} \mathbb{P}_{\mu,\sigma^2}[T_\alpha = 1] = 5\% = \inf_{\mu\neq\mu^*,\sigma^2>0} \mathbb{P}_{\mu,\sigma^2}[T_\alpha = 1].$$

(iv) Sei F die Verteilungsfunktion von $\mathrm{t}(n-1)$. Dann gilt

$$\begin{aligned} p &= \inf\{\alpha \in [0,1]; T_\alpha = 1\} = \inf\{\alpha \in [0,1]; (Y < t_{n-1,\alpha/2}) \vee (Y > t_{n-1,1-\alpha/2})\} \\ &= \inf\{\alpha \in [0,1]; (F(Y) < \alpha/2) \vee (F(Y) > 1-\alpha/2)\} \\ &= \inf\{\alpha \in [0,1]; |F(Y) - 1/2| > 1/2 - \alpha/2\} = 1 - 2|F(Y) - 1/2|. \end{aligned}$$

Der p-Wert ist also, so wie die Tests $(T_\alpha)_{\alpha\in[0,1]}$, eine Statistik. Unter H_0 ist der p-Wert uniform $U([0,1])$-verteilt. Der Test T_α verwirft die Nullhypothese genau dann wenn $p < \alpha$.

7.2.8 Bemerkung. Bei großen Datensätzen ist die empirische Varianz $s_n^2(X)$ sehr nahe an der wahren Varianz σ^2. Demgemäß kann die Verteilung $\mathrm{t}(n-1)$ dann durch die Normalverteilung $N(0,1)$ approximiert werden.

7.2.9 Definition. Für alle $n_1, n_2 \in \mathbb{N}$ und $k \in \{0, \dots, n_1\}$ sei $\mathrm{hyp}(k, n_1, n_2)$ die Verteilung der ersten Koordinate von $\mathrm{Hyp}(k, n_1, n_2)$, das heißt, die Verteilung auf $\{0, \dots, n_1\}$ mit Zähldichte

$$k_1 \mapsto \frac{\binom{n_1}{k_1}\binom{n_1}{n_1+n_2-k_1}}{\binom{n_1+n_2}{k}}.$$

7.2.10 Bemerkung. Interpretation: $\mathrm{hyp}(k, n_1, n_2)$ ist die Verteilung der Anzahl weißer Kugeln in einer ungeordneten Stichprobe ohne Zurücklegen der Größe k aus einer Grundmenge von n_1 weißen und n_2 schwarzen Kugeln.

7.2.11 Lemma. *Seien X und Y unabhängige Bernoulli-verteilte Zufallsvariablen, sei $n \in \mathbb{N}_{>0}$, und sei $(X_1, Y_1), \ldots, (X_n, Y_n)$ eine Stichprobe*[1] *von (X, Y). Dann ist in der* Kontingenztabelle

Anzahl	$Y = 0$	$Y = 1$	*Gesamt*
$X = 0$	$\sum_{i=1}^n (1 - X_i)(1 - Y_i)$	$\sum_{i=1}^n (1 - X_i) Y_i$	$\sum_{i=1}^n (1 - X_i)$
$X = 1$	$\sum_{i=1}^n X_i (1 - Y_i)$	$\sum_{i=1}^n X_i Y_i$	$\sum_{i=1}^n X_i$
Gesamt	$\sum_{i=1}^n (1 - Y_i)$	$\sum_{i=1}^n Y_i$	

die bedingte Verteilung von $\sum_{i=1}^n X_i Y_i$ gegeben $\sum_{i=1}^n X_i$ und $\sum_{i=1}^n Y_i$ gleich

$$\mathrm{hyp}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n (1 - Y_i)\right).$$

Beweis. Sei $k, n_1 \in \{0, \ldots, n\}$, $k_1 \in \{0, \ldots, k\}$, $k_2 = k - k_1$ und $n_2 = n - n_1$. Dann gilt

$$\begin{aligned}\mathbb{P}\left[\sum_{i=1}^n X_i Y_i = k_1 \middle| \sum_{i=1}^n X_i = k, \sum_{i=1}^n Y_i = n_1\right] &= \mathbb{P}\left[\sum_{i=1}^k Y_i = k_1 \middle| \sum_{i=1}^n X_i = k, \sum_{i=1}^n Y_i = n_1\right] \\ &= \mathbb{P}\left[\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} \middle| \sum_{i=1}^n X_i = k, \sum_{i=1}^n Y_i = n_1\right].\end{aligned}$$

□

7.2.12 Beispiel (Zweiseitiger Test für die Unabhängigkeit von Merkmalen)**.** Sei Θ die Menge der Wahrscheinlichkeitsverteilungen auf $\{0, 1\} \times \{0, 1\}$, sei $n = 100$, und für jedes $\theta \in \Theta$ sei $\mathbb{P}_\theta$ die Verteilung einer Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$ von $(X, Y) \sim \theta$.

(i) Sei Θ_0 die Menge aller $\theta \in \Theta$ so dass $X \perp\!\!\!\perp Y$ unter θ, sei $\Theta_1 = \Theta \backslash \Theta_0$, sei $m = \mathbb{1}_{\Theta_1}$, sei $Z = \sum_{i=1}^n X_i Y_i$, sei $\alpha = 5\%$, und sei $h_{\alpha/2}$ bzw. $h_{1-\alpha/2}$ das $(\alpha/2)$ bzw. $(1 - \alpha/2)$-Quantil der hypergeometrischen Verteilung

$$\mathrm{hyp}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n (1 - Y_i)\right). \tag{1}$$

Dann ist

$$T_\alpha = \mathbb{1}_{\{(Z < h_{\alpha/2}) \vee (Y > h_{1-\alpha/2})\}}$$

ein Test von $H_0 : X \perp\!\!\!\perp Y$ gegen $H_1 : X \not\perp\!\!\!\perp Y$.

(ii) Der Test hat (bis auf Rundungsfehler) Signifikanzniveau 5% weil

$$\sup_{\theta \in \Theta_0} \underbrace{\mathbb{P}_\theta[T_\alpha = 1]}_{\approx 5\%} \approx 5\%.$$

[1] Das heißt, (X_i, Y_i) sind unabhängig und haben die gleiche Verteilung wie (X, Y).

(iii) Der Test ist unverfälscht (ohne Beweis).

(iv) Sei F die Verteilungsfunktion von (1). Dann zeigt die gleiche Rechnung wie in Beispiel 7.2.7.(iv), dass

$$p = 1 - 2|F(Z) - 1/2| \sim U([0, 1]).$$

Der Test T_α verwirft die Nullhypothese genau dann, wenn $p < \alpha$.

Literatur

[Pea09] J. Pearl. *Causality. Models, reasoning, and inference*. Titleaddon. Cambridge University Press, 2009.

[Pfa13] P. Pfaffelhuber. *Stochastik für Studierende der Informatik*. Lecture Notes. 2013.

[PGJ16] J. Pearl, M. Glymour und N. P. Jewell. *Causal inference in statistics: a primer*. John Wiley & Sons, 2016.