1. Einführung und Motivation

Grobüberblick der Stochastik

Stochastik: Lehre von Zufallsprozessen

Wahrscheinlichkeitstheorie: Zufallsprozesse mit bekannten Wahrscheinlichkeiten

Beispiel: fairer Würfel

Statistik: Zufallsprozesse mit unbekannter Wahrscheinlichkeit

Beispiel: unfairer Würfel (Schätze die Wahrscheinlichkeit ab)

Randomisierte Algorithmen

Monte-Carlo-Algorithmen: geben immer eine Antwort, aber nicht jede Antwort ist richtig

Las-Vegas-Algorithmen: geben stets richtige Antworten, lassen jedoch ggf. auf sich warten

Freivalds-Algorithmus

Problemstellung: Zu drei gegebenen $n \times n$ Matrizen A, B, C entscheide, ob $A \cdot B = C$ ist.

Freivalds Algorithmus (1)

1. Wähle Zufallsvektor $x \in \{0,1\}^n$.
2. Berechne $A \cdot (B \cdot x)$ und $C \cdot x$.
3. Ist $A \cdot (B \cdot x) \neq C \cdot x$, antworte mit $AB \neq C$.
4. Ist $A \cdot (B \cdot x) = C \cdot x$, antworte mit $AB = C$.

Antwort $AB \neq C$ unproblematisch, da $ABx \neq Cx$.
Antwort $AB = C$ problematisch, da von $ABx = Cx$ auf $ABy = Cy$ für alle Zufallsvektoren y geschlossen wurde.

Als Urnenmodell

Eine Urne $U_{A,B,C}$ enthält zu jedem Zufallsvektor $x \in \{0,1\}^n$ eine gefärbte Kugel K_x:

- falls $ABx = Cx$, ist K_x blau,
- falls $ABx \neq Cx$, ist K_x rot.

Im Fall $AB = C$ sind **alle** Kugeln blau.

Freivalds Algorithmus (2)

1. Wähle eine positive ganze Zahl k.
2. Greife k-mal unabhängig in die Urne $U_{A,B,C}$.
3. Sind alle gezogenen Kugeln blau, so antworte mit $AB = C$.
4. Ist eine gezogene Kugel rot, so antworte mit $AB \neq C$.

Satz von Freivalds

Im Fall $AB \neq C$ ist höchstens die Hälfte aller Kugeln blau.

Färbungsproblem

F_j: Fachabteilung
A_i: Arbeitsgruppe

: Experte
: Expertin

	F_1	F_2	F_3	F_4	F_5	F_6	F_7
A_1	•		•	•		•	
A_2	•	•		•			•
A_3		•	•		•	•	
A_4	•			•	•		•
A_5	•	•	•			•	
A_6			•	•	•		•
A_7	•		•		•		•
A_8		•		•	•	•	

Frage: Gibt es eine Gesamtabordnung, so dass in den Arbeitsgruppen keine reinen Herrenrunden und keine reinen Damenkränzchen entstehen?

Zulässige Färbung:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7
A_1	• (blau)		• (rot)	• (blau)		• (blau)	
A_2	• (blau)	• (rot)		• (blau)			• (blau)
A_3		• (rot)	• (rot)		• (rot)	• (blau)	
A_4	• (blau)			• (blau)	• (rot)		• (blau)
A_5	• (blau)	• (rot)	• (rot)			• (blau)	
A_6			• (rot)	• (blau)	• (rot)		• (blau)
A_7	• (blau)		• (rot)		• (rot)		• (blau)
A_8		• (rot)		• (blau)	• (rot)	• (blau)	

Abstrakt:

$F = [1:7]$ identifiziert die Fachabteilungen $F_1, \ldots, F_7$.

konkret: S_i beschreibt die Zusammensetzung der i-ten Arbeitsgruppe. Im Beispiel sind etwa $S_1 = \{1,3,4,6\}$, $S_2 = \{1,2,4,7\}$ usw.

Aufgabe: Färbe jedes Element von F jeweils mit einer von zwei Farben (rot, blau) so ein, dass keine der Teilmengen $S_1, S_2, \ldots, S_m$ einfarbig ist.
[Eine solche Färbung heiße $(S_1, \ldots, S_m)$-zulässig.]
konkret: S_i einfarbig rot: reines Damenkränzchen,
S_i einfarbig blau: reine Herrenrunde.

Las-Vegas-Färbungsalgorithmus

1. **Eingabe**: n-elementige Menge F sowie r-elementige, paarweise verschiedene Teilmengen $S_1, \ldots, S_m$ von F.
2. Färbe die Elemente von F gemäß Gleichverteilung blau oder rot.
3. Teste, ob Gesamtfärbung $(S_1, \ldots, S_m)$-zulässig.
4. Falls zulässig, Färbung gefunden, sonst weiter mit 2.

Satz

Es sei $r \geq 2$ und $m \leq 2^{r-2}$. Dann gilt:

(1) Es gibt eine $(S_1, \ldots, S_m)$-zulässige Gesamtfärbung.
(2) Im Mittel ist der Las-Vegas-Algorithmus spätestens nach zwei (!) Färbungsversuchen erfolgreich.

2. Wahrscheinlichkeitsräume

Wahrscheinlichkeitsraum

→ modelliert ein Zufallsexperiment $(\Omega, \mathcal{A}, P)$.

Ω: **Ergebnisraum**
$\mathcal{A}$: **Ereignis-Algebra**
P: **W-Maß**

Ergebnisräume: Ausgang des Zufallsexperiment mengentheoretisch beschreiben

Einmaliges Würfeln: $\Omega = [1:6]$

n-maliges Würfeln:

1. Folge der gewürfelten Augenzahlen

$$\Omega_n = [1:6]^n = \{(\omega_1, \ldots, \omega_n) \mid \forall i : \omega_i \in [1:6]\}$$

2 Häufigkeiten der gewürfelten Augenzahlen

$$\Omega'_n = \{(h_1, \ldots, h_6) \in [0:n]^6 \mid h_1 + \ldots + h_6 = n\}.$$

Ω'_n ist dabei das Bild der Vergröberungsabbildung

$$\Omega_n \ni \omega \mapsto (h_j := |\{i : \omega_i = j\}|)_{j \in [1:6]},$$

die jeder Augenzahlfolge das zugehörige **Histogramm** zuordnet.

Münzwurffolgen: $\Omega = \{0,1\}^{\mathbb{N}} = \{\omega = (\omega_i)_{i \in \mathbb{N}} \mid \forall i : \omega_i \in \{0,1\}\}.$

Satz 1.1. $\Omega = \{0,1\}^{\mathbb{N}}$ ist überabzählbar unendlich.

Ereignisse: entsprechen Teilmengen A des Ergebnisraums

Ist $\omega \in \Omega$ das Ergebnis eines Zufallsexperiments, so sagt man im Fall

- $\omega \in A$: das Ereignis A ist **eingetreten**.
- $\omega \notin A$: das Ereignis A ist **nicht eingetreten**.

Beispiele:

- Beim einmaligen Würfeln beschreibt die Teilmenge $A := \{2,4,6\}$ des Ergebnisraums $\Omega = [1:6]$ das Ereignis

  Es wurde eine gerade Augenzahl gewürfelt.

- Das Ereignis *Bei n Münzwürfen fällt mindestens k-mal Zahl* wird beschrieben durch die Teilmenge

$$A = \{\omega \in \{0,1\}^n \mid \omega_1 + \ldots + \omega_n \geq k\}$$

  des Ergebnisraums $\Omega = \{0,1\}^n$. [$0 \cong$ Kopf, $1 \cong$ Zahl.]

Ereignissysteme

System, was in konsistenter Weise alle Ereignisse fasst, sodass ihnen eine Wahrscheinlichkeit P(A) zugewiesen werden kann

$\mathcal{A} \subseteq 2^{\Omega} := \{X \mid X \subseteq \Omega\}$

σ-Algebren als Ereignissysteme

System $\mathcal{A} \subseteq 2^{\Omega}$ heißt σ-Algebra über Ω, wenn gilt:

1. $\Omega \in \mathcal{A}$.
2. Gegenereignisse sind in $\mathcal{A}$: $A \in \mathcal{A} \Rightarrow A^c := \Omega \setminus A \in \mathcal{A}$.
3. Ist unter der Vereinigung abgeschlossen: $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$

$(\Omega, \mathcal{A})$ bildet den Ereignisraum

Eigenschaften:

(a) $\emptyset \in \mathcal{A}$.
(b) $A, B \in \mathcal{A}$ impliziert $A \cup B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, $A \setminus B \in \mathcal{A}$.
(c) $\mathcal{A}$ ist unter abzählbar unendlicher Durchschnittsbildung abgeschlossen.

Beispiele:

Im Fall $\emptyset \neq A \subset \Omega$ gilt für die von $\mathcal{G} = \{A\}$ erzeugte σ-Algebra:

$$\sigma(\mathcal{G}) = \{\emptyset, A, A^c, \Omega\}.$$

Ist Ω höchstens abzählbar unendlich, so ist die von $\mathcal{G} := \{\{\omega\} \mid \omega \in \Omega\}$ erzeugte σ-Algebra die Potenzmenge von Ω:

$$\sigma(\mathcal{G}) = 2^{\Omega}.$$

Sei $\Omega = \mathbb{R}^n$ und $\mathcal{G} := \{\prod_{i=1}^n [a_i, b_i] : a_i, b_i \in \mathbb{Q} \wedge a_i < b_i\}$ das System aller achsenparallelen kompakten Quader in $\mathbb{R}^n$ mit rationalen Endpunkten. Dann heißt

$$\mathcal{B}^n := \sigma(\mathcal{G})$$

die **Borelsche σ-Algebra** auf $\mathbb{R}^n$ und jedes $A \in \mathcal{B}^n$ heißt eine **Borel-Menge**.

Typische σ-Algebren:

Diskreter Fall: Ω ist höchstens abzählbar unendlich. Hier setzt man $\mathcal{A} = 2^{\Omega}$.
Reeller Fall: $\emptyset \neq \Omega \subseteq \mathbb{R}^n$. Dann wählt man

$$\mathcal{A} = \mathcal{B}^n_{\Omega} := \{A \cap \Omega \mid A \in \mathcal{B}^n\}.$$

Übung: $\mathcal{B}^n_{\Omega}$ ist eine σ-Algebra.

System aller σ-Algebren über Ω

$\Sigma(\Omega) := \{\mathcal{A} \mid \mathcal{A}$ ist σ-Algebra über $\Omega\}$

Eigenschaften:

(1) $\Sigma(\Omega)$ ist bezüglich Mengeninklusion halbgeordnet.
(2) $\{\emptyset, \Omega\}$ ist die kleinste σ-Algebra über Ω.
(3) 2^Ω ist die größte σ-Algebra über Ω.
(4) Der Durchschnitt beliebig vieler σ-Algebren über Ω ist wieder eine σ-Algebra über Ω.
(5) Zu jedem System $\mathcal{G}$ von Teilmengen von Ω gibt es eine kleinste, $\mathcal{G}$ enthaltende σ-Algebra über Ω. Diese von $\mathcal{G}$ erzeugte σ-Algebra $\sigma(\mathcal{G})$ ist der Durchschnitt aller $\mathcal{G}$ enthaltenden σ-Algebren über Ω:

$$\sigma(\mathcal{G}) = \bigcap_{\mathcal{G}\subseteq\mathcal{A}\in\Sigma(\Omega)} \mathcal{A}.$$

Wahrscheinlichkeitsmaße

$P(A) \in [0,1]$ beschreibt die Wahrscheinlichkeit, dass Ereignis A eintritt

$P : \mathcal{A} \to [0,1]$ ist ein W-Maß / W-Verteilung, wenn gilt:

1. Normierung: $P(\Omega) = 1$
2. σ-Additivität: $P(\bigsqcup_{i\geq 1} A_i) = \sum_{i\geq 1} P(A_i)$

Eigenschaften:

(1) $P(\emptyset) = 0$.
(2) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$; insbesondere ist

$$P(A) + P(A^c) = 1.$$

(3) **Monotonie**: $A \subseteq B \Rightarrow P(A) \leq P(B)$.
(4) σ-**Subadditivität**: $P(\cup_{i\geq 1} A_i) \leq \sum_{i\geq 1} P(A_i)$.
(5) σ-**Stetigkeit**: Wenn $A_1 \subseteq A_2 \subseteq \ldots$ und $A = \cup_{i\geq 1} A_i$ bzw. $A_1 \supseteq A_2 \supseteq \ldots$ und $A = \cap_{i\geq 1} A_i$, so gilt

$$\lim_{n\to\infty} P(A_n) = P(A).$$

Wahrscheinlichkeitsfunktion / Zähldichte

$p(\omega)$ ordnet jedem Ergebnis $\omega \in \Omega$ eine Wahrscheinlichkeit zu

$p : \Omega \to [0,1]$ mit $\sum_{\omega\in\Omega} p(\omega) = 1$

Konstruktion von W-Maßen

diskreter Fall:

Für P auf $(\Omega, 2^\Omega)$ gilt:

1 $$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

2 $$P(A) := \sum_{\omega \in A} p(\omega)$$, mit $p : \Omega \to [0,1]$ mit $\sum_{\omega \in \Omega} p(\omega) = 1$

stetiger Fall:

Sei $\Omega \subseteq \mathbb{R}^n$ borelsch, und $\rho : \Omega \to [0, \infty)$ mit:

(a) $\{x \in \Omega \mid \rho(x) \leq c\} \in \mathcal{B}^n_\Omega$ für alle $c > 0$
(b) $\int_\Omega \rho(x)dx = 1$

Dann gilt für P auf $(\Omega, \mathcal{B}^n_\Omega)$:

$$P(A) = \int_A \rho(x)dx$$

Konvexität

Menge $K \subseteq \mathbb{R}^n$ ist konvex, wenn gilt:

$x, y \in K$ und $\lambda \in [0,1]$ impliziert $\lambda y + (1-\lambda)x \in K$.

konvexe Hülle: Sei $X \subseteq \mathbb{R}^n$, kleinste X enthaltende konvexe Hülle

$$\operatorname{conv}(X) = \bigcap_{X \subseteq K,\ K \text{ konvex}} K$$

Konvexkombination: Beschreibung einer konvexen Hülle durch

$\lambda_0 x_0 + \ldots + \lambda_r x_r$, mit $(\lambda_0, \ldots, \lambda_r)$ eine Zähldichte

Gleichverteilung

diskreter Fall:

W-Funktion: $\rho(\omega) := \dfrac{1}{|\Omega|}$

Verteilung: $\mathcal{U}_\Omega(A) = |A|/|\Omega|$ (# günstigen Fälle / # möglichen Fälle)

Stetiger Fall:

Sei $\Omega \subseteq \mathbb{R}^n$ Borel-Menge mit Volumen $0 < \lambda^n(\Omega) < \infty$

W-Funktion: $\rho(x) := \dfrac{1}{\lambda^n(\Omega)}$

Verteilung: $\mathcal{U}_\Omega(A) = \lambda^n(A)/\lambda^n(\Omega)$

3. Zufallsvariable

Messbare Abbildung / Zufallsvariable

Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ Messräume

Dann ist $X : \Omega \to \Omega'$ eine Zufallsvariable / $(\mathcal{A}, \mathcal{A}')$-messbar, wenn die Urbilder im Ereignisraum liegen: $\forall A' \in \mathcal{A}' : \quad X^{-1}[A'] := \{\omega \in \Omega \mid X(\omega) \in A'\} \in \mathcal{A}$.

$X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$

Für $\mathcal{A} = 2^{\Omega}$ ist jede Abbildung eine Zufallsvariable

Sparversion: Wird die σ-Algebra $\mathcal{A}'$ von $\mathcal{G}'$ erzeugt, $\sigma(\mathcal{G}') = \mathcal{A}'$, so ist $X : \Omega \to \Omega'$ bereits dann eine ZV, wenn gilt (Übung):

$$\forall A' \in \mathcal{G}' : \quad X^{-1}[A'] \in \mathcal{A}.$$

Bildmaße

Sei $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$

W-Maß P' zu P bzgl. X: $P'(A') := P(X^{-1}[A']) = P(\{X \in A'\}) =: P(X \in A')$ für $A' \in \mathcal{A}'$

Verteilungen

Bildmaß P bzgl. X ist Verteilung von X

Familie von W-Räumen: $((\Omega_i, \mathcal{A}_i, P_i))_{i \in I}$

Familie von ZVs. $X_i : (\Omega_i, \mathcal{A}_i) \to (\Omega, \mathcal{A})$

Die Familie $(X_i)_{i \in I}$ heißt **identisch verteilt**, wenn alle Verteilungen $P_i \circ X_i^{-1}$ übereinstimmen.

Beispiel. n-maliger Münzwurf mit fairer Münze.

- Hier ist $I = [1 : n]$, $(\Omega_i = \{0,1\}^n, \mathcal{A}_i = 2^{\Omega_i}, P_i \equiv$ Gleichverteilung) für alle i. $(\Omega, \mathcal{A}) = (\{0,1\}, 2^{\{0,1\}})$.
- $X_i : \{0,1\}^n \to \{0,1\}$ sei die Projektion $(\omega_1, \ldots, \omega_n) \mapsto \omega_i$.
- Dann sind $X_1, \ldots, X_n$ identisch verteilte ZVs, denn die Verteilung von jedem X_i ist die Gleichverteilung auf $\Omega = \{0,1\}$.

Gemeinsame Verteilungen

Ereignisse, die von mehreren ZVs abhängen, können nicht durch die einzelnen Verteilungen beschrieben werden

Seien $X_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$

Produktabbildung: $X := X_1 \otimes \ldots \otimes X_n : \Omega \to \Omega_1 \times \ldots \times \Omega_n$

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

→ ist ebenfalls eine Zufallsvariable

P_X : gemeinsame Verteilung der $X_1, \ldots, X_n$

Produkt-σ-Algebra: $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n := \sigma\Big(\bigcup_{j=1}^{n} \pi_j^{-1}[\mathcal{A}_j]\Big) \subseteq 2^{\Omega_1 \times \ldots \times \Omega_n}$

Beispiel:

Für P_X zugehörige Zähldichte p_X auf $[0:n] \times [1:n+1]$ gilt für $(k,h) \in [1:n]^2$ gilt:

$$p_X(k,h) = \binom{n-h}{k-1} \cdot 2^{-n}$$

Es gilt: $p_X(0,h) = 0$, $p_X(k, n+1) = 0$, $p_X(0, n+1) = 2^{-n}$

Randverteilung

Marginalverteilung: gemeinsame Verteilung des Teilproduktes

$X_{i_1} \otimes \ldots \otimes X_{i_k} : \Omega \to \Omega_{i_1} \times \ldots \times \Omega_{i_k}$

Marginalisierung: Eliminieren der irrelevanten Komponenten durch komplettes Aufsummieren

$$p_{X_1 \otimes \ldots \otimes X_k}(\omega_1, \ldots, \omega_k) = \sum p_{X_1 \otimes \ldots \otimes X_n}(\omega_1, \ldots, \omega_k, \omega_{k+1}, \ldots, \omega_n),$$ mit $(i_1, \ldots, i_k) = (1, \ldots, k)$

dabei wird – bei festem $(\omega_1, \ldots, \omega_k)$ – über alle $(\omega_{k+1}, \ldots, \omega_n) \in \Omega_{k+1} \times \ldots \times \Omega_n$ summiert.

4. Standardmodelle

Laplace Räume:

W-Räume mit Gleichverteilung $(\Omega, 2^{\Omega}, \mathcal{U}_{\Omega})$

Verteilung: $\boxed{\mathcal{U}_{\Omega}(A) := \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.}$

W-Funktion: $\omega \mapsto \frac{1}{|\Omega|}.$

Beispiele:

- n-maliger Wurf einer fairen Münze: $\Omega = \{0,1\}^n$, $|\Omega| = 2^n$
- n-maliger Wurf eines fairen Würfels: $\Omega = [1:6]^n$, $|\Omega| = 6^n$
- Zahlenlotto: $\Omega = \{Z \mid Z \subset [1:49] \wedge |Z| = 6\}$, $|\Omega| = \binom{49}{6} = \frac{49!}{43! \cdot 6!}$
- Reihenfolge eines gut gemischten Skatblatts: $\Omega = \{\pi \mid \pi \text{ ist Permutation von } [1:32]\}$, $|\Omega| = 32!$

Kontinuierlicher Fall: $\boxed{\mathcal{U}_{\Omega}(A) := \frac{\lambda^n(A)}{\lambda^n(\Omega)} = \frac{\text{günstiges Volumen}}{\text{mögliches Volumen}}.}$

Urnenmodell

Zufallsereignis: Mehrfaches Ziehen von teilweise verschiedenfarbiger Kugeln

Varianten:

Ω_{ZR}: Ziehen mit Zurücklegen unter Berücksichtigung der Reihenfolge

Ω_{Zr}: Ziehen mit Zurücklegen ohne Berücksichtigung der Reihenfolge

Ω_{zR}: Ziehen ohne Zurücklegen unter Berücksichtigung der Reihenfolge

Ω_{zr}: Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge

Modellparameter:

N: Anzahl der Kugeln in der Urne
F: Menge der verschiedenen Farben der Kugeln in der Urne
N_f: Anzahl der f-farbigen Kugeln in der Urne ($f \in F$)
Beachte: $N = \sum_{f \in F} N_f$.
n: Anzahl der Ziehungen

Fiktiver Laplace Raum: Kugeln werden zusätzlich noch durchnummeriert

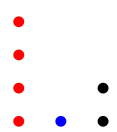
$\longrightarrow$ W-Raum bekannt

Vergröbern um Information zu komprimieren

$$[1:N]^n \xrightarrow{X_{ZR}} \Omega_{ZR} \xrightarrow{X_{Zr}} \Omega_{Zr}.$$

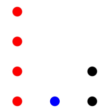
$$(\textcolor{red}{2}, \textcolor{blue}{4}, \textcolor{red}{5}, \textcolor{red}{2}, 1, \textcolor{red}{2}, 3) \xrightarrow{X_{ZR}} (\textcolor{red}{\bullet}, \textcolor{blue}{\bullet}, \textcolor{red}{\bullet}, \textcolor{red}{\bullet}, \bullet, \textcolor{red}{\bullet}, \bullet)$$

$$(\textcolor{red}{\bullet}, \textcolor{blue}{\bullet}, \textcolor{red}{\bullet}, \textcolor{red}{\bullet}, \bullet, \textcolor{red}{\bullet}, \bullet) \xrightarrow{X_{Zr}}$$



Farbenhistogramme:

$$H = \begin{pmatrix} \bullet & \bullet & \bullet \\ 4 & 1 & 2 \end{pmatrix} \equiv$$



Raum: $\Omega_{Zr} = \{H \in [0:n]^F \mid \sum_{f\in F} H(f) = n\}$

ZV: $X_{Zr} : F^n \ni (f_1, \ldots, f_n) \mapsto (F \ni f \mapsto |\{i \mid f_i = f\}|)$

Standardmodelle

Bernoulli-Verteilung:

$\Omega = \{0,1\}^n$ 1: Erfolg, 0: Misserfolg

Erfolgswahrscheinlichkeit: p

Misserfolgswahrscheinlichkeit: $q = 1 - p$

Wahrscheinlichkeit: $p^k \cdot q^{n-k}$, mit $\omega = (\omega_1, \ldots, \omega_n)$ und in ω genau k-mal eine Eins

Multinomialverteilung:

N-Kugeln mit Farbe $f \in F$

n-maliges Ziehen mit Zurücklegen

Wahrscheinlichkeit eines Farbenhistogramms:

$$\boxed{P_{Zr}(H) = \binom{n}{H} \cdot \prod_{f\in F} \left(\frac{N_f}{N}\right)^{H(f)}}$$

Multinomialkoeffizient: $\binom{n}{H} := \frac{n!}{\prod_{f\in F} H(f)!}$

$\longrightarrow$ Anzahl der Farbenfolgen der Länge n, in denen die Farben $1, \ldots, c$ mit Vielfachheiten $n_1, \ldots, n_c$ vorkommen

Binomialverteilung

N-Kugeln mit zwei Farben $F = \{0,1\}$, wobei 1 als Erfolg und 0 als Misserfolg

n-maliges Ziehen mit Zurücklegen

Wahrscheinlichkeit für k-Erfolge:

$$\boxed{B_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}}$$

Hypergeometrische Verteilung

N gefärbte Kugeln, wobei Farbe $f \in F$ mit Vielfachheit N_f drankommt

n-Ziehungen ohne Rücklegen

Wahrscheinlichkeit eines Farbenhistogramms:

$$\boxed{P_Y(H) = \frac{\prod_{f\in F} \binom{N_f}{h_f}}{\binom{N}{n}}.}$$

Geometrische Verteilung

Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p

Wahrscheinlichkeit erstmals im k-ten Schritt Erfolg zu haben:

$$\boxed{P(k) = p \cdot (1-p)^{k-1}}$$

Poisson-Verteilung

Motivation: Wieviele E-Mails werden in einem festen Zeitintervall $(0, t]$, $t > 0$, über einen Mail-Server geleitet?

Ziehe von n Kugeln mit Zurücklegen mit 2 Farben:

Sorte 1: eine E-Mail ist im aktuellen Teilintervall weiterzuleiten;
Sorte 0: keine E-Mail ist im aktuellen Teilintervall weiterzuleiten.

Kugelanteil Sorte 1: $\alpha \cdot t/n$

Poisson-Verteilung zum Parameter λ

$$\boxed{\lim_{n\to\infty} B_{n,p_n}(k) = e^{-\lambda}\frac{\lambda^k}{k!} =: P_\lambda(k)}$$, mit $\lambda := \alpha \cdot t$ und $p_n := \lambda/n$

Gauß-Verteilung

Dichte-Funktion:

$$\boxed{\phi_{m,v}(x) := \frac{1}{\sqrt{2\pi v}} e^{-(x-m)^2/(2v)}}$$

m: Erwartungswert
v: Varianz

W-Maß: $N_{m,v}$

Standard-Normalverteilung: $N_{0,1}$

5. Bedingte Wahrscheinlichkeit und Unabhängigkeit

revidiertes W-Maß:

W-Maß P_A unter der Bedingung, dass A bereits eingetreten ist

(a) $P_A(A) = 1$, denn A ist jetzt ein sicheres Ereignis!
(b) Die Neubewertung der Teilereignisse B von A ist proportional zur ursprünglichen Bewertung:

$$\exists c_A > 0 \, \forall B \in \mathcal{A}, B \subseteq A : \; P_A(B) = c_A P(B).$$

Neubewertung:

$(\Omega, \mathcal{A}, P)$ ein W-Raum und $A \in \mathcal{A}$ mit $P(A) > 0$

$$P_A(B) := \frac{P(A \cap B)}{P(A)} \quad \text{für } B \in \mathcal{A}.$$

$$\frac{\text{Anzahl der Fälle, in denen } A \textbf{ und } B \text{ eingetreten sind}}{\text{Anzahl der Fälle, in denen } A \text{ eingetreten ist}}$$

Bedingte Wahrscheinlichkeit

$$\boxed{P(B|A) := \frac{P(A \cap B)}{P(A)}} \quad \text{für } B \in \mathcal{A}$$

Satz. Durch $B \mapsto P(B|A)$ wird ein W-Maß P_A auf $(\Omega, \mathcal{A})$ definiert.

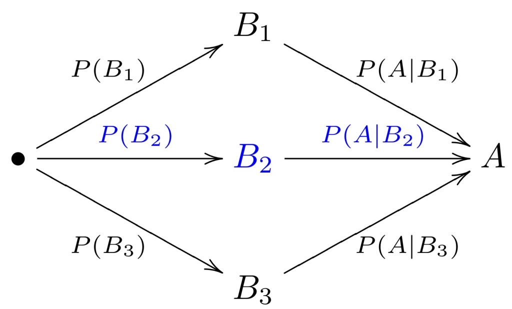
Formel von Bayes

Satz. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $\Omega = \sqcup_{i \in I} B_i$ eine disjunkte Zerlegung von Ω in höchstens abzählbar unendlich viele Ereignisse $B_i \in \mathcal{A}$ mit $P(B_i) > 0$. Dann gilt:

Fallunterscheidungsformel:
$$\boxed{P(A) = \sum_{i \in I} P(B_i) P(A|B_i).}$$

Formel von Bayes:
$$\boxed{P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i \in I} P(B_i)P(A|B_i)}.}$$

Wegemodell



Wahrscheinlichkeit A über B_k zu erreichen:

$$P(B_k) \cdot P(A|B_k) = P(B_k) \cdot \frac{P(A \cap B_k)}{P(B_k)} = P(A \cap B_k).$$

Satz von Bayes:
$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i \in I} P(B_i)P(A|B_i)}. \quad = \frac{\text{Wahrscheinlichkeit des Wegs über } B_k}{\text{Gesamtwahrscheinlichkeit aller Wege}}$$

Multiplikationsformel:

Sei $A_1, \ldots, A_n \in \mathcal{A}$ und $A := A_1 \cap \ldots \cap A_n$

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

Konstruktion mehrstufiger Modelle

Ereignisräume: $\Omega_1, \ldots, \Omega_n,\ n \geq 2.$

W-Funktionen: p_1, $\omega_i \in \Omega_i$ für $i < k$ sei $p_{k|\omega_1,\ldots,\omega_{k-1}}$

ZVs: $X_i : \Omega \to \Omega_i$ die i-te Projektion

Konstruktion des W-Maßes:

(a) Für alle $\omega_1 \in \Omega_1$ gilt $P(X_1 = \omega_1) = p_1(\omega_1)$.

(b) Für alle $k \in [2:n]$ und $\omega_i \in \Omega_i$ gilt

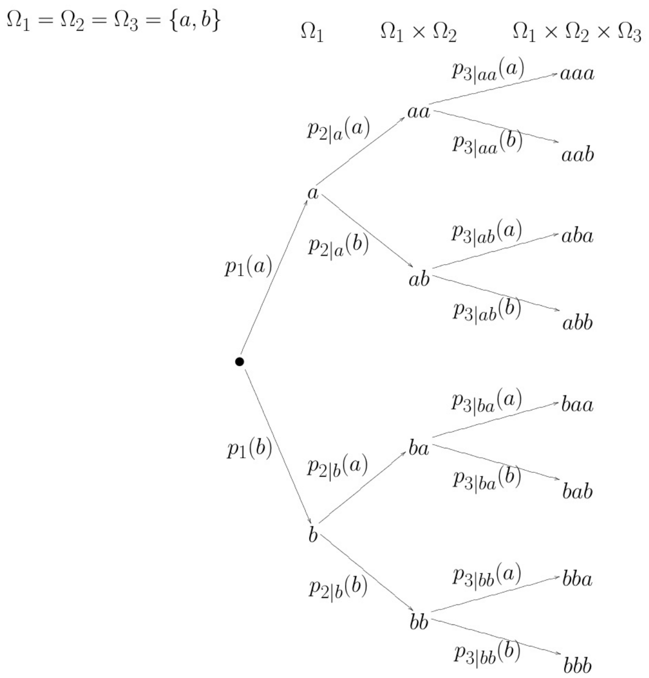
$$P(X_k = \omega_k | X_1 = \omega_1, \ldots, X_{k-1} = \omega_{k-1}) = p_{k|\omega_1,\ldots,\omega_{k-1}}(\omega_k),$$

sofern $P(X_1 = \omega_1, \ldots, X_{k-1} = \omega_{k-1}) > 0$.

Das W-Maß P ist gegeben durch

$$P(\{\omega\}) = p_1(\omega_1) \cdot p_{2|\omega_1}(\omega_2) \cdot p_{3|\omega_1,\omega_2}(\omega_3) \cdot \ldots \cdot p_{n|\omega_1,\ldots,\omega_{n-1}}(\omega_n)$$

für $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$.



Stochastische Unabhängigkeit

Wahrscheinlichkeit, dass A eintritt, wird nicht durch das Eintreten von B beeinflusst

$A, B \in \mathcal{A}$ heißen stochastisch unabhängig bzgl. P

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

Unabhängigkeit von Ereignisfamilien

Familie $(A_i)_{i \in I}$ heißen stochastisch unabhängig bzgl. P, wenn für alle $J \subseteq I \neq \emptyset$ gilt:

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

Unabhängigkeit von ZV-Familien

Familie $(Y_i)_{i \in I}$, mit $Y_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$ heißen stochastisch unabhängig bzgl P, wenn für alle $J \subseteq I \neq \emptyset$ und alle $A_j \in \mathcal{A}_j$ mit $j \in J$ gilt:

$$P\left(\bigcap_{j \in J} \{Y_j \in A_j\}\right) = \prod_{j \in J} P(Y_j \in A_j).$$

Unabhängigkeitskriterien

Satz. Es sei $(\Omega, \mathcal{A}, P)$ ein W-Raum und $Y_1, \ldots, Y_n$ eine endliche Familie von ZVs $Y_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$. Dann gilt:

(a) **Diskreter Fall**: Hat jedes Y_i einen endlichen Wertebereich Ω_i und ist $\mathcal{A}_i = 2^{\Omega_i}$, so ist $(Y_i)_{i \in [1:n]}$ genau dann unabhängig, wenn für beliebige $\omega_i \in \Omega_i$ gilt:

$$P(Y_1 = \omega_1, \ldots, Y_n = \omega_n) = \prod_{i=1}^{n} P(Y_i = \omega_i).$$

(b) **Reeller Fall**: Ist jedes Y_i reellwertig und ist $\mathcal{A}_i = \mathcal{B}^1_{\Omega_i}$, so ist $(Y_i)_{i \in [1:n]}$ genau dann unabhängig, wenn für beliebige $c_i \in \mathbb{R}$ gilt:

$$P(Y_1 \leq c_1, \ldots, Y_n \leq c_n) = \prod_{i=1}^{n} P(Y_i \leq c_i).$$

6 Erwartungswert und Varianz

Kenngrößen reellwertiger Zufallsvariablen

Erwartungswert $E_P(X)$: P-gewichteter Mittelwert von X
Varianz $V_P(X)$: Abweichung von X von seinem Erwartungswert

Erwartungswert:

Allgemein: **Definition**. Die ZV $X : \Omega \to \mathbb{R}$ **besitzt** einen Erwartungswert, kurz: $X \in \mathscr{L}^1(P)$, wenn $X_{(n)} \in \mathscr{L}^1(P)$ für ein n (und damit für alle n) gilt. In dem Fall heißt

$$\mathbf{E}_P(X) := \lim_{n\to\infty} \mathbf{E}_P(X_{(n)})$$

$$\mathbf{E}_P(X) = \int X \, dP.$$

diskreter Fall:

X besitzt Erwartungswert $\Leftrightarrow$ $\sum_{x \in X[\Omega]} |x| P(X = x) < \infty.$

Erwartungswert: $\boxed{\mathbf{E}(X) = \mathbf{E}_P(X) := \sum_{x \in X[\Omega]} x P(X = x)}$, $\mathbf{E}_P(X) = \sum_{\omega\in\Omega} X(\omega)P(\omega).$

Menge aller ZVs mit Erwartungswert: $\mathscr{L}^1(P) := \{X : \Omega \to \mathbb{R} \mid \mathbf{E}_P(|X|) < \infty\} = \mathscr{L}^1$

Stetiger Fall

Diskretisierung durch die ZV X durch Partitionierung der reellen Achse:

$$X_{(n)}(\omega) := \frac{k}{n}, \quad \text{wobei } k \in \mathbb{Z} \text{ und } \quad \frac{k}{n} \le X(\omega) < \frac{k+1}{n}.$$

Satz.Für die $1/n$-Diskretisierungen $X_{(n)}$ von X gilt:

(1) Für alle $n \in \mathbb{N}$ ist $X_{(n)} \le X < X_{(n)} + \frac{1}{n}$.
(2) Ist $X_{(m)} \in \mathscr{L}^1(P)$ für ein m, so ist $X_{(n)} \in \mathscr{L}^1(P)$ für alle n und in diesem Fall ist $(\mathbf{E}_P(X_{(n)}))_{n\ge1}$ eine Cauchy-Folge.

Rechenregeln:

1. Monotonie: Aus $X \le Y$, d.h. $X(\omega) \le Y(\omega)$ für alle $\omega \in \Omega$, folgt $\mathbf{E}_P(X) \le \mathbf{E}_P(Y)$.
2. Linearität: $\mathbf{E}_P(c_1X_1 + c_2X_2) = c_1\mathbf{E}_P(X_1) + c_2\mathbf{E}_P(X_2)$
3. σ-Additivität: Sind alle $X_n \ge 0$ und ist $X = \sum_{n\ge1} X_n$, so gilt $\mathbf{E}_P(X) = \sum_{n\ge1} \mathbf{E}_P(X_n)$.
4. Monotone Konvergenz: $\mathbf{E}_P(Y) = \lim_{n\to\infty} \mathbf{E}_P(Y_n)$.
5. Produktregel: X, Y sind unabhängig $\mathbf{E}_P(X \cdot Y) = \mathbf{E}_P(X)\mathbf{E}_P(Y)$.
6. Cauchy-Schwarzsche Ungleichung: $\mathbf{E}_P(XY)^2 \le \mathbf{E}_P(X^2) \cdot \mathbf{E}_P(Y^2)$.

Bernoulli-Experiment

Sei $X_1, \dots, X_n$ eine endliche Bernoulli-Folge zur Erfolgswahrscheinlichkeit p.
Dann beschreibt die ZV

$$S := \sum_{i=1}^{n} X_i$$

die Anzahl der Erfolge.

Erwartungswert: $$\mathbf{E}_P(S) = \sum_{i=1}^{n} \mathbf{E}_P(X_i) = \sum_{i=1}^{n} P(X_i = 1) = np.$$

Erwartungswert bei W-Maßen mit Dichtefunktion

Es sei $\Omega \subset \mathbb{R}^d$ Borelsch und P das W-Maß auf $(\Omega, \mathcal{B}^d_\Omega)$ zur Dichtefunktion ρ, d.h. $P(A) = \int_A \rho(\omega)d\omega$ für alle $A \in \mathcal{B}^d_\Omega$.

$X \in \mathcal{L}^1(P)$: $$\boxed{\mathbf{E}_P(X) = \int_\Omega X(\omega)\rho(\omega)d\omega.}$$

Spezialfall: $\Omega = \mathbb{R}$, $X = \mathrm{id}_\mathbb{R}$ $$\mathbf{E}_P(\mathrm{id}_\mathbb{R}) = \int_\mathbb{R} x\rho(x)dx.$$

Momente von Zufallsvariablen

Gesamtheit aller ZVs, für die das r-te Moment existiert:

$$\mathcal{L}^r(P) := \{X : \Omega \to \mathbb{R} \mid X^r \in \mathcal{L}^1(P)\}.$$

1. Moment: Erwartungswert
2. Moment: Varianz

Es gilt:
(1) Für $r \in \mathbb{N}$ ist $\mathcal{L}^r(P)$ ein reeller Untervektorraum von $\mathcal{L}^1(P)$.
(2) Die Räume $\mathcal{L}^r(P)$ bilden eine Kette:

$$\mathcal{L}^1(P) \supseteq \mathcal{L}^2(P) \supseteq \mathcal{L}^3(P) \supseteq \dots$$

Varianz und Kovarianz

Varianz: $\boxed{\mathbf{V}(X) = \mathbf{V}_P(X) := \mathbf{E}_P([X - \mathbf{E}_P(X)]^2) = \mathbf{E}_P(X^2) - \mathbf{E}_P(X)^2}$

Standardabweichung: $\sqrt{\mathbf{V}(X)}$

Kovarianz: $\boxed{\mathrm{Cov}_P(X,Y) := \mathbf{E}_P([X - \mathbf{E}_P(X)][Y - \mathbf{E}_P(Y)]) = \mathbf{E}_P(XY) - \mathbf{E}_P(X)\mathbf{E}_P(Y)}$

Ist $\mathrm{Cov}_P(X,Y) = 0$, so heißen X und Y **unkorreliert**.

Korrelationskoeffizient: $$\boxed{\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbf{V}(X)\mathbf{V}(Y)}}}$$

Rechenregeln:

1. $\mathrm{Cov}_P(aX + b, cY + d) = ac\mathrm{Cov}_P(X,Y) \longrightarrow \mathbf{V}_P(aX + b) = a^2\mathbf{V}_P(X).$
2. $\mathrm{Cov}_P(X,Y)^2 \leq \mathbf{V}_P(X)\mathbf{V}_P(Y)$

3. $\mathbf{V}_P(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbf{V}_P(X_i) + \sum_{j \neq i} \mathrm{Cov}_P(X_i, X_j).$

Sind $X_1, \dots, X_n$ paarweise unkorreliert:

$$\mathbf{V}_P(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbf{V}_P(X_i).$$

4. Sind X und Y unabhängig, so sind X und Y auch unkorreliert.

Spezialfall: Gleichverteilung

$$\boxed{\mathbf{E}(X) = \overline{x} := \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{und} \quad \mathbf{V}(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2,}$$

Spezialfall: Poisson-Verteilung

$\mathbf{E}_P(X) = \lambda.$, $V_P(X) = \lambda$

Standardisierung:

Ist $X \in \mathscr{L}^2(P)$ mit $\mathbf{V}_P(X) > 0$, so folgt aus der Regel (1), dass die ZV

$$X^* := \frac{X - \mathbf{E}_P(X)}{\sqrt{\mathbf{V}_P(X)}}$$

standardisiert ist, d.h. es gilt $\mathbf{E}_P(X^*) = 0$ und $\mathbf{V}_P(X^*) = 1$.

Normalverteilung

$\mathcal{N}_{m,v}$ mit Dichtefunktion $\boxed{\phi_{m,v}(x) := \frac{1}{\sqrt{2\pi v}} e^{-(x-m)^2/(2v)}}$

Erwartungswert: $\mathbf{E}(\mathcal{N}_{m,v}) \quad = \quad \int_{\mathbb{R}} x \phi_{m,v}(x) dx = m$

Varianz: $\mathbf{V}(\mathcal{N}_{m,v}) \quad = \quad \int_{\mathbb{R}} (x - m)^2 \phi_{m,v}(x) dx = v.$

7. Gesetze der großen Zahl

Stochastische Konvergenz

Motivation:

$$P\left(\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\right) = \binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}} \xrightarrow{n\to\infty} 0,$$

→ relative Häufigkeit liegt bei großem n nur zu einer sehr geringen Wahrscheinlichkeit beim Erwartungswert

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbf{E}_P(X_1)\right| < \epsilon\right) \xrightarrow{n\to\infty} 1,$$

→ liegt aber zu hoher Wahrscheinlichkeit nahe beim Erwartungswert

Folge $(Y_n)_n$ heißt stochastisch konvergent gegen Y ($Y_n \xrightarrow{P} Y$), wenn gilt:

$\forall_{\varepsilon > 0}$: $\lim_{n\to\infty} P(|Y_n - Y| < \epsilon) = 1$

Schwaches Gesetz der großen Zahl

Seien $(X_i)_{i\geq 1}$ paarweise unkorrelierter ZVe in $\mathcal{L}^2(P)$

Wahrscheinlichkeit, dass das arithmetische Mittel der ersten n zentralisierten ZVs mindestens um ε vom Mittel abweicht, ist beschränkt durch:

$$\boxed{P\left(\left|\frac{1}{n}\sum_{i=1}^n (X_i - \mathbf{E}_P(X_i))\right| \geq \epsilon\right) \leq \frac{v}{n\epsilon^2} \xrightarrow{n\to\infty} 0}$$

Sind alle Erwartungswerte gleich:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbf{E}_P(X_1)$$

Für Bernoulli-Versuche:

ZV $Y_n := (X_1 + \ldots + X_n)/n$ die relative Häufigkeit der Erfolge

$$P(|Y_n - p| \geq \epsilon) \leq \frac{\mathbf{V}_P(X_1)}{n\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Tschebyscheff-Ungleichung

Sei $X \in \mathcal{L}^2(P)$

$$\boxed{P(|X - \mathbf{E}_P(X)| \geq \epsilon) \leq \frac{\mathbf{V}_P(X)}{\epsilon^2}}$$

Fast sichere Konvergenz

Seien Y, Y_1, Y_2 reelle ZVs

Folge $(Y_n)_{n \geq 1}$ konvergiert P-fast sicher gegen Y, wenn gilt:

Für alle Stellen $\omega \in \Omega$, an denen punktweise Konvergenz herrscht, gilt:

$$A := \{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}$$

$$P(A) = 1$$

Fast sichere Konvergenz => stochastische Konvergenz

Sigma-Stetigkeit von W-Maßen

Seien $(A_n)_{n \geq 1}$ eine aufsteigende Kette von Ereignissen, mit $A = \bigcup_{n \geq 1} A_n$

$$P(A) = \lim_{n\to\infty} P(A_n)$$

Starkes Gesetz der großen Zahl

Seien $(X_i)_{i \geq 1}$ ein Folge paarweise unkorrelierter ZVs in $\mathcal{L}^2(P)$

$$\boxed{\frac{1}{n}\sum_{i=1}^n (X_i - \mathbf{E}_P(X_i)) \to 0 \quad P\text{-fast sicher.}}$$

Borel-Cantelli Lemma

Seien $A_n \in \mathcal{A}$ und $A \equiv \limsup_{k\to\infty} A_k := \{\omega \in \Omega \mid \omega \in A_k \text{ für unendlich viele } k\}$

Dann gilt:

(0) Ist $\sum_{k\geq 1} P(A_k) < \infty$, so ist $P(A) = 0$.

(1) Ist $\sum_{k\geq 1} P(A_k) = \infty$ und sind die Mitglieder der Ereignisfamilie $(A_k)_{k\geq 1}$ paarweise stochastisch unabhängig, so gilt $P(A) = 1$.

Verteilungsfunktion der Gauß-Glocke

Verteilungsfunktion: $F_P := (\mathbb{R} \ni c \mapsto P((-\infty, c]) \in [0,1]$

Dichtefunktion der Standardnormalverteilung $\mathcal{N}_{0,1}$:

$$\boxed{\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}}$$

Verteilungsfunktion der Standardnormalverteilung:

$$\Phi(c) := \mathcal{N}_{0,1}((-\infty, c]) = \int_{-\infty}^{c} \phi(x)dx.$$

Zentraler Grenzwertsatz

Supremumsabstand zweier Verteilungsfunktion F und G:

$$d(F, G) := \sup_{c \in \mathbb{R}} |F(c) - G(c)|.$$

Seien $(X_i)_{i \geq 1}$ eine Folge unabhängiger, identisch verteilter reellwertiger ZVs in $\mathcal{L}^2(P)$

$\mathbf{E}_P(X_i) = m$, $\mathbf{V}_P(X_i) = v > 0.$, $S_n := X_1 + \ldots + X_n$ die n-te Partialsumme

Standardisierung von S_n: $\boxed{S_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sqrt{v}}}$

Sei F_n die Verteilungsfunktion von S_n^*

Dann konvergiert diese im Supremumsabstand gegen die Verteilungsfunktion der Standardnormalverteilung

$$\lim_{n \to \infty} d(F_n, \Phi) = 0.$$

→ die Verteilungsfunktion der Standardisierung ($E_P = 0$, $V_P = 1$) ist die Standardnormalverteilung

8. Markov Ketten

Stochastische Matrizen

$V \neq \emptyset$ und $\Pi = (\Pi(x,y))_{x,y\in V}$ ist eine reellwertige Matrix

Π heißt Zeilenstochastisch, wenn in jeder Zeile der Matrix eine W-Funktion steht:

1. alle Einträge liegen im Intervall $[0,1]$
2. Die Summe der Einträge jeder Zeile beträgt 1

Betrachtet werden Zufallsprozesse in V, die in jedem Schritt mit der Wahrscheinlichkeit $\Pi(x,y)$ von Zustand x in y wechselt.

$$x \xrightarrow{\Pi(x,y)} y.$$

Markov-Ketten

Seien $X_0, X_1, \ldots$ sind ZVs auf dem W-Raum $(\Omega, \mathcal{A}, P)$

$X_i: (\Omega, \mathcal{A}) \longrightarrow (V, 2^V)$ heißt Markov-Kette mit Zustandsraum V und Übergangsmatrix Π

Markov-Eigenschaft muss gelten:

$$P(X_0 = x_0, \ldots, X_n = x_n) > 0 \qquad \begin{aligned} &P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) \\ = \ &P(X_{n+1} = x_{n+1} | X_n = x_n), \end{aligned}$$

→ darf nur vom aktuellen Zustand abhängen (Gegenwart)
→ stationäre Übergangswahrscheinlichkeit bleibt gleich

Startverteilung: $\alpha := P \circ X_0^{-1}$

Mehrstufige Modelle

Konstruktion mehrstufiger Modelle umfasst Markov-Ketten als Spezialfall

Setze die Wahrscheinlichkeiten. $p_{k|\omega_0,\ldots,\omega_{k-1}}(\omega_k) := \Pi(\omega_{k-1}, \omega_k),$

Dann existiert zu jeder Startverteilung α auf V genau ein W-Maß P^α auf:

$$(\Omega, \mathcal{A}) := (\prod_{k\geq 0} V, \bigotimes_{k\geq 0} 2^V)$$

sodass $X_n: \Omega \longrightarrow V$ eine Markov-Kette zu Π und α bildet:

$$X_n : (\omega_k)_{k\geq 0} \mapsto \omega_n,$$

Ist $\alpha = \delta_x$ für ein $x \in V$ (*sicherer Start* in x), so schreibt man kurz P^x statt P^α.

Matrixpotenzen

Die n-te Potenz: Π^n enthält an der Position (x,y) die Wahrscheinlichkeit genau in n Schritten von Zustand x in y zu wechseln

$$P^x(X_n = y) = \Pi^n(x,y).$$

Übergangsgraphen

Visualisierung der Markov-Kette zu Π durch einen gerichteten, kantengewichteten Übergangsgraphen (V, E, Π)

V: endliche Knotenmenge
E: Kantenmenge $E := \{(x,y) \in V \times V | \Pi(x,y) > 0\}$
Π: Kantengewichte

Websurfen als Markov-Prozess

gerichteter Graph:

$G = (V, E)$ V: Menge der Webseiten
E: Menge der Links

Menge der von x erreichbaren Webseiten:

$$G[x] := \{y \in V | (x,y) \in E\} \cup \{x\}$$

Modellierung:

$p \in (0,1)$: gleichverteilte Wahrscheinlichkeit Surfer gibt eine neue URL ein

$1-p$: Wahrscheinlichkeit Surfer folgt einem Link von x oder bleibt auf Seite x

Übergangswahrscheinlichkeit:

$$\Pi(x,y) := \begin{cases} \frac{p}{N} & \text{falls } (x,y) \notin E \\ \frac{p}{N} + \frac{1-p}{|G[x]|} & \text{falls } (x,y) \in E. \end{cases}$$

Es gilt: $\sum_{y \in V} \Pi(x,y) = \underbrace{\sum_{y=1}^{N} \frac{p}{N}}_{=p} + \underbrace{\sum_{y \in G[x]} \frac{1-p}{|G[x]|}}_{=1-p} = 1.$

Ergodensatz

Bedingung: Es existiert ein $L \geq 1$, sodass alle Einträge Π^L positiv sind

Dann existiert die Grenzverteilung: W-Funktion $\rho = (\rho_1, \ldots, \rho_N)$

Eigenschaften:

1. $\lim_{m\to\infty} \Pi^m =: \Pi^\infty$: $\quad \Pi^\infty = \begin{pmatrix} \rho_1 & \rho_2 & \ldots & \rho_N \\ \rho_1 & \rho_2 & \ldots & \rho_N \\ \vdots & \vdots & \ldots & \vdots \\ \rho_1 & \rho_2 & \ldots & \rho_N \end{pmatrix}$
2. $(\Pi^m)_{m\geq 1}$ konvergiert exponentiell schnell gegen Π^∞
3. Für die Grenzverteilung gilt: $\boxed{\rho\Pi = \rho.}$

9. Steilkurs Statistik

Statistische Modelle

$\mathcal{M} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$

$\mathcal{X}$: Stichprobenraum
$\mathcal{A}$: σ-Algebra auf $\mathcal{X}$
$(P_\theta)_{\theta\in\Theta}$: Familie von W-Maßen $(\mathcal{X}, \mathcal{A})$, $|\Theta| \geq 2$

Bemerkung: $\mathcal{X}$ statt Ω, da ZV $X: \Omega \to \mathcal{X}$ die detaillierte Beschreibung Ω auf $\mathcal{X}$, die tatsächliche beobachtbaren Ergebnisse, abbildet

Modellklassen

Sei $\mathcal{M} = (\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$

parametrisches Modell: $\Theta \subset \mathbb{R}^d$ für $d \in \mathbb{N}$
einparametrisches Modell: $d = 1$

diskretes Modell: $\mathcal{X}$ ist diskret ($|\mathcal{X}| \leq |\mathbb{N}|$, $\mathcal{A} = 2^{\mathcal{X}}$)
Jedes P_θ ist gegeben durch: W-Funktion $p_\theta : x \mapsto p_\theta(x) := P_\theta(\{x\})$

stetiges Modell: $\mathcal{X}$ ist eine Borel-Teilmenge eines $\mathbb{R}^n$
$\mathcal{A} = 2^{\mathcal{X}}$ Borel-σ-Algebra
P_θ besitzt eine Dichtefunktion ρ_θ

Standardmodell: $\mathcal{M}$ ist diskret oder stetig

Produktmodelle

Sei $(E, \mathcal{E}, (Q_\theta)_{\theta\in\Theta})$ ein statistisches Modell

n-faches Produktmodell: $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta\in\Theta}) := (E^n, \mathcal{E}^{\otimes n}, (Q_\theta^{\otimes n})_{\theta\in\Theta})$

$X_i: \mathcal{X} \to E$ beschreibt den Ausgang des i-ten Teilexperiments
$\to X_i$ sind bzgl. $P_\theta = Q_\theta^{\,n}$ unabhängig und identisch verteilt

Zufallszahlen: Erwartungswert und Varianz

Seien $\Omega = [0, \theta]^n$, P_θ Gleichverteilung auf Ω

Dichtefunktion: $\rho_\theta(\omega) := \theta^{-n}$ $(\omega \in \Omega)$ $= \left(\frac{1}{\theta}\right)^n$

$X := \mathrm{Id}_\Omega$: modelliert das Gesamtprotokoll der n Stichproben
X_i: Ergebnis der i-ten Stichprobe

Dann gilt: $\mathbf{E}_\theta(X_i) = \theta/2$ und $\mathbf{V}_\theta(X_i) = \theta^2/12$.

Statistiken und Schätzer

Sei $(\mathcal{X}, \mathcal{A}, (P_\theta)_{\theta\in\Theta})$ ein statistisches Modell und $(\Sigma, \mathcal{S})$ ein Messraum.

Statistik: ZV $S : (\mathcal{X}, \mathcal{A}) \to (\Sigma, \mathcal{S})$

Schätzer: $\tau : \Theta \to \Sigma$ ordnet jedem $\theta \in \Theta$ eine Kenngröße $\tau(\theta) \in \Sigma$ zu
Statistik $T : \mathcal{X} \to \Sigma$ ist Schätzer für τ

→ für $\tau = id_\Theta$ ist T ein Schätzer für Θ

erwartungstreu: Erwartungswert des Schätzers ist gleich dem zu schätzenden Parameters

verzerrt: Erwartungswert weicht vom tatsächlichen Wert ab, liegt aber in der Nähe (fast erwartungstreu)

Maximum-Likelihood-Schätzer

Idee: $p_{T(x)}(x) = \max_{\theta\in\Theta} p_\theta(x)$.

→ Bestimme den Schätzwert $T(x)$ so, dass der Parameter für $P_{T(x)}(x)$ die maximale Auftrittwahrscheinlichkeit für x erzeugt

Likelihood-Funktion: $p : \mathcal{X} \times \Theta \to [0, \infty)$ $\quad p(x,\theta) := p_\theta(x)$

Likelihood-Funktion zum Beobachtungswert x: $p(x,\cdot) : \Theta \to [0,\infty)$

Maximum-Likelihood-Schätzer (MLE)

$T : \mathcal{X} \to \Theta$ für θ

Für jedes $x \in \mathcal{X}$ muss $T(x)$ eine Maximalstelle von $p(x,\cdot)$ sein:

$$p(x, T(x)) = \max_{\theta\in\Theta} p(x,\theta).$$

Log-Likelihood-Funktion: $\log p(x,\cdot)$

→ besitzt die selben Maximalstellen, wie die Likelihood-Funktion
→ oft leichter zu berechnen

Schätzung der Erfolgswahrscheinlichkeit

Reißzwecke fällt mit unbekannter Wahrscheinlichkeit Θ auf die Spitze

Statistisches Modell: Binomialmodell $\quad ([0:n], 2^{[0:n]}, (B_{n,\theta})_{\theta\in[0,1]})$

Likelihood-Funktion: $p(k,\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}.$

MLE-Bestimmung: Betrachtung der Nullstellen der Ableitung der log-Likelihood-Funktion. $0 = \frac{d}{d\theta}\log p(k,\theta)$

$$0 = \frac{d}{d\theta}(k\log\theta + (n-k)\log(1-\theta)) = \frac{k}{\theta} - \frac{n-k}{1-\theta}.$$

$\Rightarrow$ Nullstelle: $\theta = k/n.$ $\rightarrow$ Maximum

$\Rightarrow$ $T(k) := k/n$ einziger MLE für θ.

Schätzung des Zufallszahlenbereichs

- In einer Fernsehshow führt der Moderator einen Apparat vor, der Zufallszahlen im Intervall $[0,\theta]$ gemäß Gleichverteilung ausspuckt, wenn er vom Moderator auf den Wert $\theta > 0$ eingestellt wurde.
- Zwei Spieler dürfen den Apparat $n = 10$ mal bedienen und sollen dann θ möglichst gut schätzen.
- Wer besser rät, hat gewonnen.

Statistisches Modell: Produktmodell der skalierten Gleichverteilung

$$([0,\infty)^n, \mathcal{B}_{[0,\infty)}^{\otimes n}, (\mathcal{U}_{[0,\theta]}^{\otimes n})_{\theta>0})$$

Likelihood-Funktion: $p((x_1,\dots,x_n),\theta) = \begin{cases} \theta^{-n} & \text{falls } x_1,\dots,x_n \le \theta, \\ 0 & \text{sonst.} \end{cases}$

MLE-Bestimmung:

Bei festem $x = (x_1,\dots,x_n)$ ist $p(x,\theta)$ genau für $\theta = \max(x_1,\dots,x_n)$ maximal, denn $(0,\infty) \ni \theta \mapsto \theta^{-n}$ ist streng monoton fallend. Wegen $x_1,\dots,x_n \le \theta$ muss $\theta \ge \max(x_1,\dots,x_n)$ sein. Also ist

$$\tilde{T}_n(x) = \max(x_1,\dots,x_n)$$

der MLE.

MLE im Gauß'schen Produktmodell

Bei vielen Experimenten kann angenommen werden, dass die Meßwerte durch eine normalverteilte Zufallsvariable mit Erwartungswert m und Varianz v>0 modelliert werden können. n unabhängige Experimente

Statistisches Modell: n-fache Gauß'sche Produktmodell

$$(\mathcal{X},\mathcal{A},(P_\theta)_{\theta\in\Theta}) := (\mathbb{R}^n, \mathcal{B}^n, (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R},v>0})$$

Likelihood-Funktion: $p((x_1,\dots,x_n),\theta) = \prod_{i=1}^n N(m,v)(x_i) = (2\pi v)^{-n/2} exp\left[-\sum_{i=1}^n \frac{(x_i-m)^2}{2v}\right]$

MLE-Bestimmung:

Likelihood-Funktion wird maximal, wenn m so gewählt wird, dass die quadratische Abweichung minimal wird

Erwartungswert: $$m = M((x_1, \cdots, x_n)) := \frac{1}{n}\sum_{i=1}^{n} x_i$$

Varianz: $$v = V((x_1, \cdots, x_n)) := \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2$$

Korrigierte Stichprobenvarianz

Sei $n \geq 2$ und $(\mathbb{R}^n, \mathcal{B}^n, Q_\theta^{\otimes n} : \theta \in \Theta)$ ein reelles n-faches Produktmodell.

Stichprobenmittelwert: $$M = \frac{1}{n}\sum_{i=1}^{n} X_i$$

korrigierte Stichprobenvarianz: $$V^* = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2$$

=> m=M, v=V sind erwartungstreue Schätzer