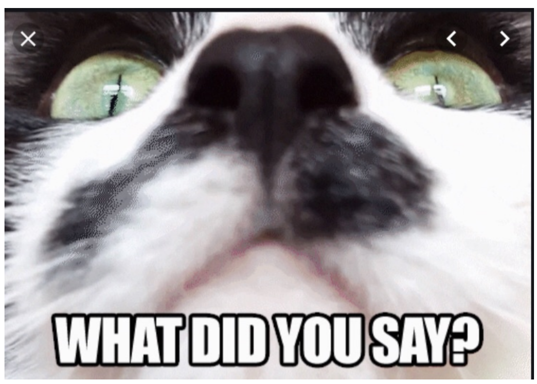
Mitschrift

Wahrscheinlichkeit

WS 2019



Vorbesprechung

GRILL 58801/10572

Sprechstunde: Mittwoch 13-14
Freitag 10-11

DA06B12

Zeiten:

Mittwoch $10^{00} - 12^{00}$ bzw

Freitag $8^{00} - 10^{00}$

Prüfung:

schriftlich & mündlich
praktisch – theoretisch

UE im 5. Sem

4 BSP / nicht 1 pro Kapitel !

https://institute.tuwien.ac.at/mstoch/mitarbeiterinnen/professoren/grill/karl_grill_1/

Inhalte & Übersicht

1) Kapitel → Wahrscheinlichkeit
2) Kapitel → Stochastische Prozesse
3) Kapitel → Informationstheorie (je unwahr. desto größer Informationswert)
4) Kapitel → Statistik (WSLK umgedreht, Bedeutung IRL)

Ab Jänner EI X $8 - 11^{15}$ (nur noch Freitags!)

Wahrscheinlichkeitstheorie

Empirisches Gesetz der großen Zahlen:

Ein "zufälliges" Experiment wird sehr oft wiederholt.

zB: faire Münze (Kopf, Zahl) $\frac{1}{2}$, Würfel $(1, \dots, 6)$ $\frac{1}{6}$, Urne (Kugeln verschiedenfarbig, zahlen) ...

Die relativen Häufigkeiten "konvergieren" gegen "Wahrscheinlichkeit" Physikalisches Gesetz

Wahrscheinlichkeiten sind etwas, mit dem man rechnet wie mit (relativen) Häufigkeiten.

BSP Würfel Ausgänge 1 ... 6 alle Elementarereignis

Elementarereignisse (ω) werden zusammengefasst zur Grundmenge $\Omega = \{1 \dots 6\}$ Würfel

$\omega \in \Omega$ heißt Elementarereignis

(komplexere) bestimmte Ereignisse definieren: Augenzahl ist gerade / [Augenzahl ist 2, 4, 6] $\in \{2,4,6\}$

Augenzahl ist mindestens 4 $[4,5,6]$ $\in \{4,5,6\}$

Ein Ereignis ist eine Teilmenge von Grundmenge $A \subseteq \Omega$

Elementarereignis $\neq$ Ereignis

ω (Zahl) $\neq$ Menge

Wahrscheinlichkeit = Wahrscheinlichkeitsmaß ($\approx$ Wahrscheinlichkeitsfunktion) ist eine Funktion, die in Ereignissen (= Teilmengen der Grundmenge Ω) Zahlenwerte zuordnet

Eigenschaften (Axiome von Kolmogorov)

- $0 \leq P(A) \leq 1$ WSLK liegt immer zw 0 und 1
- $\mathbb{P}(\emptyset) = 0$, $P(\Omega) = 1$ $\emptyset$ ... unmögliches Ereignis, Ω ... sicheres Ereignis

Additivität
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ Wahrscheinlichkeiten addierbar wenn $A \cap B = \emptyset$ Disjunkt

  $\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n)$ wenn $A_1 \dots A_n$ disjunkt sind.

Abzählbare / Sigma Additivität
- $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A)$ $\longrightarrow$ $A_i \cap A_j = \emptyset \quad \forall\, i,j \quad i \neq j \quad 1 \leq i,j \leq n$

BSP faircr Würfel

- $\Omega = \{1 \dots 6\}$ mögliche Versuche Ereignisse: Alle Mengen die aus 1-6 gebildet werden - 64 Stück
- Alle Elementarereignisse sollen gleiche Wahrscheinlichkeit haben: $\frac{1}{6}$
- $\mathbb{P}(\Omega) = \mathbb{P}(\{1,2,3,4,5,6\}) = \mathbb{P}(\{1\} \cup \{2\} \cup \dots \cup \{6\}) = \mathbb{P}(\{1\}) + \dots + \mathbb{P}(\{6\})$

BSP unfairer Würfel: $\mathbb{P}(\{6\}) = 0.5$

$\mathbb{P}(\{1\}) = \dots = \mathbb{P}(\{5\}) = 0.1$

BSP Münze wird geworfen bis Kopf rauskommt

Grill
" Türken haben liebere Söhne!"

- $\Omega = \{1, 2, 3 \dots\}$ Versuch Möglichkeiten bis Kopf kommt (Grundmenge ist ∞)
- $\mathbb{P}(\{1\}) = \frac{1}{2}$

  $\mathbb{P}(\{2\}) = \frac{1}{4}$ Z,K KZ K,K ZZ

  $\vdots$

  $\mathbb{P}(\{n\}) = \frac{1}{2^n}$

BSP zufälliger Zeitpunkt

zwischen 7 und 8: wie ist die Wahrscheinlichkeit für eintreffen zwischen 7^{15} und 7^{30}

7:00 und 8:00 $\rightarrow$ 61 Möglichkeiten

günstig 7:15:00 bis 7:30:00

$\Omega = [0,1]$

Zeit in Sekunden 7:00:00 - 8:00:00 $\rightarrow$ 3601

7:15:00 - 7:00:00 $\rightarrow$ 901

$\mathbb{P}(A) = \frac{16}{61}$

$\mathbb{P}(A) = \frac{90}{3601}$

$\mathbb{P}([a,b]) = b-a$

Wahrscheinlichkeit = Länge d. Intervalls

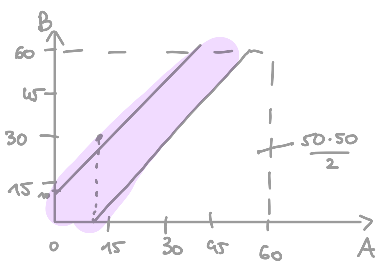
Zeit in hundertstel Sekunden $\mathbb{P}(A) = \frac{90001}{360001}$

$\rightarrow$ im Grenzwert $\mathbb{P}(A) = \frac{1}{4}$

Albert kommt zufällig zwischen 7 und 8. Bettina auch. Jeder bleibt 10 min. Wie groß ist die Wahrscheinlichkeit, dass sie sich treffen?

Ankunftszeiten voneinander unabhängig.

P(treffen)



WSLK ist Anteil an Gesamtfläche

$$P(\text{treffen}) = \frac{\text{Fläche } \square}{\text{Fläche } \square} = \frac{3600 - 2500}{3600} = \frac{11}{36}$$

Folgerungen (zusätzliche Eigenschaften)

•) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ weil $A \cap A^c = \emptyset \quad A \cup A^c = \Omega$

$1 = P(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$

•) $A \subseteq B \quad \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$

Sprechweise:

$A \cup B$: A oder B tritt ein. (Oder ist inklusiv!)

$A \cap B$: A und B tritt ein.

$A \triangle B$: entweder A oder B (exklusives oder - Symmetrische Differenz)

$A \triangle B : (A \cup B) \setminus (A \cap B)$ oder $(A \setminus B) \cup (B \setminus A)$

$A^C = \Omega \setminus A$: Gegenereignis, A tritt nicht ein

nur die Element weg, die auch in B liegen!

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

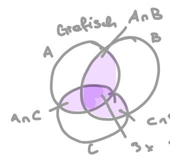
$= \mathbb{P}(B) - \mathbb{P}(A \cap B)$

•) Additionstheorem: $\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \setminus A)) =$
2 Wslk

$= \mathbb{P}(A + \mathbb{P}(B \setminus A)) =$

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

3 Wslk: $\mathbb{P}(A \cup B \cup C) = \ldots = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C)$



Jeweils 1x abziehen

3x zu
3x ab
1x fehlt → 1x dazu

n Wslk: $2^n - 1$ Summanden ☺

∞ Ada

allgemeiner Fall $P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \cdot S_k$ Siebformel

$$S_k = \sum_{1 \le i_1 \le \ldots \le i_n} P(A_{i_1} \cap \ldots \cap A_{i_k}) \Rightarrow \mathbb{P}\left(\bigcup_{i=1} A_i\right) = \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ J \neq \emptyset}} (-1)^{|J|-1} \, \mathbb{P}\left(\bigcap_{i \in J}^{k} A_i\right)$$

$\boxed{\text{BSP}}$ zufällige Vertauschung von n Elementen

$\mathbb{P}$(kein Fixpunkt) wslk wenn du wählst kein FP
Alle Permutationen die keinen FP hat!

$A_1 \ldots A_n$: $A_i = FP$

$S_1 - S_n + S_3 - \ldots \pm S_n$ ← letztes Glied Sieb

$\mathbb{P}(\text{kein FP}) = 1 - \mathbb{P}(FP) = 1 - \mathbb{P}(\cup A_n) = 1 - (\underbrace{\mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)} - \mathbb{P}(A_1 \cap A_2) - \ldots + \ldots)$

k facher Durchschnitt: $S_k = \frac{1}{k!}$

S_1: $\mathbb{P}(A_1) = \frac{(n-1)!}{n!} = \mathbb{P}(A_2) = \ldots = \mathbb{P}(A_n) \Rightarrow S_1 = n \frac{(n-1)!}{n!} = 1$

S_2: $\mathbb{P}(A_1 \cap A_2) = \frac{(n-2)!}{n!} = \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_n) \Rightarrow S_2 \binom{n}{2} \frac{(n-2)!}{n!} = \frac{1}{2!}$

S_n: $\Rightarrow \mathbb{P}(A_1 \cap \ldots \cap A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - + \ldots (-1)^{(n-1)} \frac{1}{n!}$

$\mathbb{P}(\text{kein FP}) = \frac{0!}{1} - \frac{1!}{1} + \frac{1}{2!} - \frac{1}{3!} \cdots + (-1)^n \frac{1}{n!}$

$= \sum_{k=0}^{n} (-1)^k \frac{1}{k!} \approx \frac{1}{e}$ Taylorreihen entwicklung

•) $A \subseteq B \Rightarrow \mathbb{P}(A) < \mathbb{P}(B)$ Monotonie

•) $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \ldots$

$(\lim A_n) = \bigcup_n A_n \qquad \mathbb{P}(\bigcup_n A_n) = \lim_{n \to \infty} \mathbb{P}(A_n)$

$B_1 = A_2 \qquad B_n = A_n \setminus A_{n-1}$

$A_n = B_1 \cup \ldots \cup B_n$

$\mathbb{P}(\cup A_n) = P(\cup B_n) = \sum \mathbb{P}(B_n) = \lim \sum \mathbb{P}(B_n) = \lim \sum \mathbb{P}(\cup B_n) = \lim_{N \to \infty} \sum \mathbb{P}(A_n)$

•) $\mathbb{P}(A \cup B) \le \mathbb{P}(A) + \mathbb{P}(B)$ Dreiecksungleichung

gilt auch für abzählbare Vereinigungen

Wahrscheinlichkeit für Durchschnitte – Bedingte Wahrscheinlichkeit

Wslk für A_1 wenn B eingetreten ist.

Frequentistisch: Experiment N mal wiederholen

B wurde bei ca $N \cdot \mathbb{P}(B)$ beobachtet, davon zirka $N \cdot \mathbb{P}(A \cap B)$ mit A

\# wie oft A auftritt $= N \cdot \mathbb{P}(A \cap B)$

bedingte Wslk

$$\mathbb{P}(A|B) = \frac{\cancel{N} \cdot \mathbb{P}(A \cap B)}{\cancel{N} \cdot \mathbb{P}(B)} \qquad \mathbb{P}(B) \neq 0$$

"Wahrscheinlichkeit für A unter der Bedingung B"

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

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

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

A und B heißen unabhängig, wenn $\mathbb{P}(A|B) = \mathbb{P}(B) \cdot \mathbb{P}(A)$

Multiplikationstheorem: $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B)$ oder $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$

$\boxed{\text{BSP}}$ Wir werfen Würfel, d.h. von Ergebnis geben wir weiße Kugeln in Urne; immer eine schwarze dabei

→ $\mathbb{P}$(schwarze Kugel)

$A_1 \dots A_6$: A_i $i =$ Augenzahl

B = schwarze Kugel gezogen

$$\mathbb{P}(B \cap \text{"}i=1\text{"}) = \mathbb{P}(i=1) \cdot \mathbb{P}(B \mid \text{"}i=1\text{"}) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

gleich Wslk für Würfel ergebnis

Bedingte Wslk ändert sich

$\vdots$

$\frac{1}{6} \cdot \frac{1}{3} \dots$

$\frac{1}{6} \cdot \frac{1}{4} \dots$

$\vdots$

$\frac{1}{7}$

Symmetrie: Alle Elementarereignisse haben die gleiche Wahrscheinlichkeit.

$|\Omega| < \infty$ wenn für $A \subseteq \Omega$: $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ gilt, dann heißt $(\Omega, \mathbb{P})$ (WSLK Raum) ein Laplacescher Wahrscheinlichkeitsraum

Unabhängig: Wenn (Genau dann, wenn) $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ heißen A und B voneinander unabhängig.

paarweise unabhängig: $A_1 \cap \dots \cap A_k$: Wenn für je 2 Ereignisse die Unabhängigkeit gilt $1 \leq i < j \leq n$ bzw $1 \leq i\ \ j \leq n$ $i \neq j$ dann heißen $A_1, \dots, A_k$ paarweise unabhängig.

$n = 2$

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2)$$
$$\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$$
$$\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_3)$$

vollständig unabhängig: $A_1 \dots A_k$ heißen vollständig unabhängig, wenn für $2 \leq k \leq n$

$$1 \leq i_1 \leq i_2 \leq \dots \leq i_k : \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$$

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3)$$

Bei 4: 6×2, 4×3, 1×4 $\varnothing$

$\binom{n}{2}$ paarweise

2^n Möglichkeiten

$\boxed{\text{BSP}}$ Münzwurf

$A_1 \to$ 1. Münze Kopf

$A_2 \to$ 2. Münze Kopf

$A_3 \to$ Beide Münzen verschieden

$\Omega = \{(K,K),(K,Z),(Z,K),(Z,Z)\}$

$A_1 = \{(K,K),(K,Z)\}$

$A_2 = \{(K,K),(Z,K)\}$

$A_3 = \{(K,Z),(Z,K)\}$

$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{2}{4} = \frac{1}{2}$

$A_1 \cap A_2 = \{(K,K)\}$

$A_1 \cap A_3 = \{(K,Z)\}$

$A_2 \cap A_3 = \{(Z,K)\}$

$A_1 \cap A_2 \cap A_3 = \{\}$

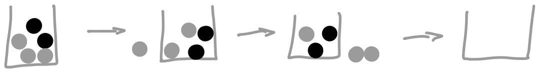
$\mathbb{P}(A_1 \cap A_2) = \frac{1}{4} = \mathbb{P}(A_1)\,\mathbb{P}(A_2)$

$\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{8} = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3)$

$\boxed{\text{BSP}}$ Urne 3 weiße, 2 schwarze Kugeln, 3x ohne zurücklegen, wie groß ist Wsk, das alle 3 Kugeln weiß sind.

A_1 ... 1 Kugel weiß

A_2 ... 2 Kugel weiß



A_3 ... 3. Kugel weiß (A_3^c ... 3 Kugel schwarz)

•) $\mathbb{P}(\text{alle weiß}) = \mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 | A_1) \cdot \mathbb{P}(A_3 | A_1 \cap A_2)$

$$\frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = \frac{6}{60} = \frac{1}{10}$$

•) $\mathbb{P}(2 \text{ weiße Kugeln}) =$

$$\mathbb{P}(wws \vee wsw \vee sww) = \mathbb{P}((A_1 \cap A_2 \cap A_3^c) \cup (A_1 \cap A_2^c \cap A_3) \cup (A_1^c \cap A_2 \cap A_3)) =$$

$$= \mathbb{P}(A_1 \cap A_2 \cap A_3^c) + \mathbb{P}(A_1 \cap A_2^c \cap A_3) + \mathbb{P}(A_1^c \cap A_2 \cap A_3) =$$

$$= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3^c | A_1 \cap A_2) + \mathbb{P}(A_1)\, \mathbb{P}(A_2^c|A_1)\, \mathbb{P}(A_3|A_1 \cap A_2) + \overset{s}{\mathbb{P}(A_1^c)}\, \overset{w}{\mathbb{P}(A_2|A_1^c)}\, \mathbb{P}(A_3|A_1^c \cap A_2)$$

$$= \frac{3}{5} \cdot \frac{2}{3} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} + \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{3}{5}$$

•) $\mathbb{P}(1 \text{ weiße})$

$$\left.\begin{matrix} wss \\ sws \\ ssw \end{matrix}\right\} \rightarrow$$

•) $\mathbb{P}(A_2) = \mathbb{P}(2. \text{ Kugel weiß})$ nur 2 weil A_2 (nach 2 Zügen!)

$$\mathbb{P}(ww) + \mathbb{P}(sw) = \mathbb{P}(A_1^c \cap A_2) + \mathbb{P}(A_1 \cap A_2)$$

$$= \mathbb{P}(A_1^c) \cdot \mathbb{P}(A_2 | A_1^c) + \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 | A_1) =$$

$$= \frac{2}{5} \cdot \frac{3}{4} + \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{5} \hat{=} \mathbb{P}(A_1)$$

$\mathbb{P}(A_3) = \mathbb{P}(3. \text{ Kugel weiß})$

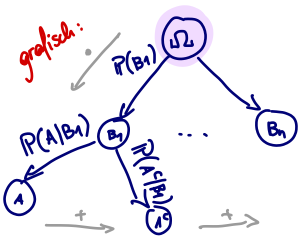
$$= \mathbb{P}(www) + \mathbb{P}(sww) + \dots$$

$n=2$
B_1 … 1 Kugel weiß
B_2 … 1. Kugel schwarz

Satz der (vollständigen) totalen Wahrscheinlichkeit

Ereignis A, Ereignis $B_1, B_2 \dots B_n$ von denen genau eines eintreten muss
(dh $B_1 \dots B_n$ disjunkt $i \neq j$ $B_i \cap B_j = \emptyset$ und $B_1 \dots B_n = \Omega$)

grafisch:

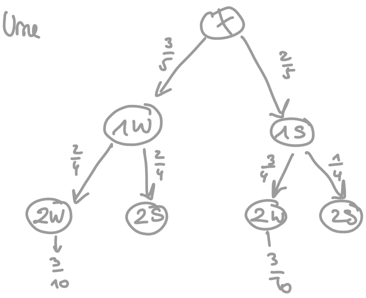


Bekannt: $\mathbb{P}(B_1) \dots \mathbb{P}(B_n)$ und $\mathbb{P}(A|B_1) \dots \mathbb{P}(A|B_n)$

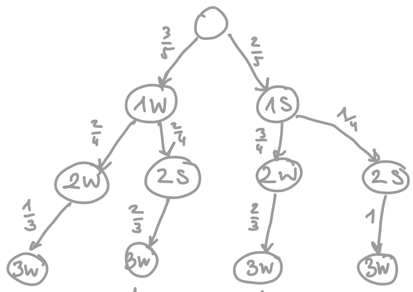
$$\Longrightarrow \mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(B_i) \cdot \mathbb{P}(A|B_i)$$

$$\mathbb{P}(A) = \mathbb{P}(\Omega \cap A) = \mathbb{P}((B_1 \cup \dots \cup B_n) \cap A) = \mathbb{P}(\overbrace{(B_1 \cap A) \cup (B_2 \cap A) \cup \dots (B_n \cap A)}^{\text{Bleiben disjunkt}}) = \mathbb{P}(A \cap B_1) + \dots + \mathbb{P}(B_n \cap A) =$$
$$= \mathbb{P}(B_1) \cdot \mathbb{P}(A|B_1) + \dots + \mathbb{P}(B_n) \cdot \mathbb{P}(A|B_n)$$

BSP Urne

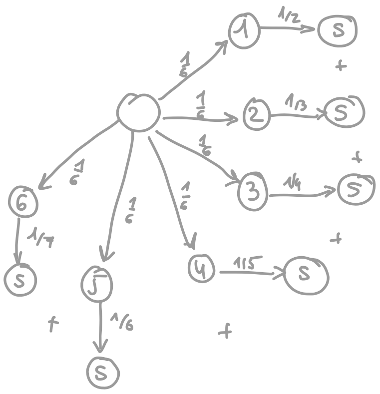


$$\mathbb{P}(2W) = \frac{3}{10} + \frac{3}{10} = \frac{3}{5}$$



$$\frac{1}{10} + \frac{1}{5} + \frac{1}{5} + \frac{1}{10} = \frac{3}{5}$$

BSP Würfel – je nach Augenzahl: so viele weiße wie Augenzahl & eine schwarze Kugel, wslk das schwarze Kugel gezogen



$$\frac{1}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{7} = \frac{223}{840}$$ totale WLK für 1 Schwarze Kugel

falls man nur A_2 kennt: (und A_1 wissen will …

$$\mathbb{P}(A_1|A_2) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)} = \frac{\mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1)}{\frac{3}{5}} = \frac{\frac{3}{5} \cdot \frac{2}{4}}{\frac{3}{5}} = \frac{2}{4}$$

Satz von Bayes

Annahmen wie vorher $(B_1 \dots B_n, A)$

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_i)\cdot\mathbb{P}(A|B_i)}{\sum_{j=0}^{n}\mathbb{P}(B_j)\,\mathbb{P}(A|B_j)}$$ } Summe der totalen Wahrscheinlichkeiten

Häufigkeiten $\mathbb{P}_a, \mathbb{P}_b, \mathbb{P}_o$

$A, 0^+$

BSP Blutgruppe

Frau A

Sohn θ

Pschyrembel:

2 Allele a b o

aa, ao, oa] A $\mathbb{P}(\text{"A"}) = p_a^2 + 2p_a p_o = 47\%$

bb, ob, bo] B $\mathbb{P}(\text{"B"}) = p_b^2 + 2p_b p_o = 9\%$

ab, ba] AB $\mathbb{P}(\text{"AB"}) = 2p_a p_b = 4\%$

oo] θ $\mathbb{P}(\text{"}\theta\text{"}) = p_o^2 = 40\%$

$p_o \approx \frac{19}{30} \simeq 0.633$

$p_a = \frac{9}{3} \approx 0.3$

$p_b = \frac{2}{30} \approx 0.067$

$(a,0) \rightarrow (0,0) \leftarrow (a,0), (b,0), (0,0)$

Formeln aufschreiben:

$\mathbb{P}(a,0) = 2p_a p_o$

$\mathbb{P}(b,0) = 2p_b p_o$

$\mathbb{P}(0,0) = p_o^2$

$\mathbb{P}(\text{sohn } 00 \mid \text{ich } a0) = 1$ $\quad\mathbb{P}(00|aa) = \theta$

$\mathbb{P}(\text{sohn } 00 \mid \text{ich } bo) = \frac{1}{2}$ $\quad\mathbb{P}(00|bb) = \theta$

$\mathbb{P}(\text{sohn } 00 \mid \text{ich } oo) = 1$ $\quad\mathbb{P}(00|ab) = \theta$

$$\Longrightarrow \mathbb{P}(\text{sohn "00"}) = 2p_a p_o \tfrac{1}{2} + 2p_b p_o \tfrac{1}{2} + p_o^2 \cdot 1$$
$$= p_o\underbrace{(p_a + p_b + p_o)}_{p_o} = p_o$$

$$\mathbb{P}(A|0) = \frac{\mathbb{P}(A \text{ und } 0)}{\mathbb{P}(0)} = \frac{8p_a p_o \frac{1}{2}}{p_o} = p_a = 0.3$$

Grill hat Blutgruppe A

$$\mathbb{P}(B|0) = \frac{\mathbb{P}(B \text{ und } 0)}{\mathbb{P}(0)} = p_b = 0.067$$

$$\mathbb{P}(0|0) = p_o = 0.63$$

Zufallsvariablen

Eine Zufallsvariable ist ein zufälliger Zahlenwert.

Mathematisch: Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$, Zufallsvariable ist Abbildung $X: \Omega \to \mathbb{R}$

Verteilung einer Zufallsvariablen ist das Wahrscheinlichkeitsmaß, was angibt, mit welcher Wahrscheinlichkeit X gewisse Werte annimmt.

$\mathbb{P}_X$ $\qquad X$

$$A \in \mathbb{R}: \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

BSP Werfen von 2 Würfeln

$\Omega = \{(1,1),(1,2)\dots(6,6)\} = \{(x,y) : 1 \le x \le 6,\ 1 \le y \le 6\}$

X ... Differenz der Beiden Augenzahlen (absolut)

$\mathbb{P}_X(\{5\}) = \mathbb{P}(X \in \{5\}) = \mathbb{P}(X=5) = \mathbb{P}(\{(x,y) : |y-x| = 5\}) = \mathbb{P}(\{1,6,\ 6,1\}) = \frac{2}{36}$

$\mathbb{P}_X(\{4\}) = \mathbb{P}(X \in \{4\}) = \frac{4}{36} = \frac{2}{18}$

$\mathbb{P}_X(\{3\}) = \mathbb{P}(X \in \{3\}) = \frac{4}{18}$

$\mathbb{P}_X(\{2\}) = \mathbb{P}(X \in \{2\}) = \frac{4}{18}$

$\mathbb{P}_X(\{1\}) = \mathbb{P}(X \in \{1\}) = \frac{5}{18}$

$\mathbb{P}_X(\{0\}) = \mathbb{P}(X \in \{0\}) = \frac{3}{18}$

$\mathbb{P}(X \in \overset{\text{gerade}}{\{0,2,4\}}) = \frac{3+4+2}{18} = \frac{1}{2}$

$\mathbb{P}(X = \pi) = \theta$

$\mathbb{P}(X = -10024) = \theta$

diskret : Eine Zufallsvariable heißt diskret, wenn sie nur endlich oder abzählbar viele Werte annimmt.

stetig: Wenn ZV alle reellen Werte / alle Werte in einem Intervall annimmt, heißt sie stetig.

Wahrscheinlichkeitsfunktion: $p_X(x) = \mathbb{P}(X=x) = \mathbb{P}_X(\{x\})$ $x \in \mathbb{R}$ wenn X diskret ist

$\mathbb{P}_X(A) = \sum_{x \in A} p_X(x)$

$F_X(x) = \sum_{y \le x} p_X(x)$

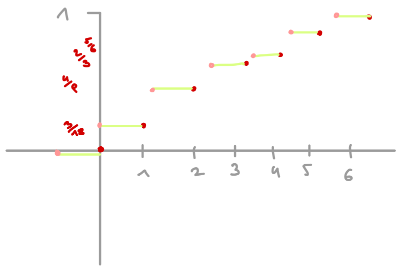
Verteilungsfunktion: $F_X(x) = \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \le x)$

- Stufenförmig, stückweise konstant mit Sprüngen (Sprünge: stelle mit der einzelne Punktwerte angenommen werden)
- F ist monoton nicht fallend
- auch von rechtsstetig

3 Arten von Verteilung an Beispielen

$\boxed{\text{BSP}}$ Würfel (von vorher) mit Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{3}{18} & 0 \le x < 1 \quad \mathbb{P}(X \le x) = \mathbb{P}(x=0) = \frac{3}{18} \\ \frac{8}{18} & 1 \le x < 2 \quad \mathbb{P}(X \le x) = \mathbb{P}(x \le 1) = \mathbb{P}(x=0) + \mathbb{P}(x=1) \\ \frac{12}{18} & 2 \le x < 3 \\ \frac{17}{18} & 4 \le x < 5 \\ 1 & 5 \le x \end{cases}$$

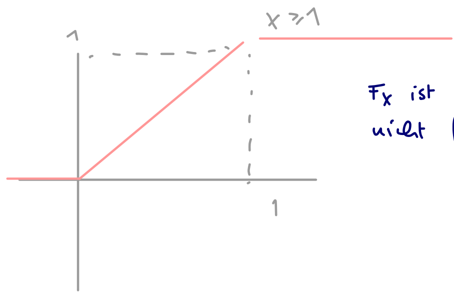


Linker Rand wird mitgezählt, Rechter Rand nicht!
daher $0 \le x < 1$!

$\boxed{\text{BSP}}$ X stetig : Ankunftszeit im Kaffeehausbeispiel Verteilung von X angeben

X Ankunftszeit (in Stunden nach 7 Uhr)

$$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} 0 : x < 0 \\ \mathbb{P}(X \text{ zwischen } 0 \text{ und } x) = x \quad 0 \le x < 1 \\ 1 : x \ge 1 \end{cases}$$



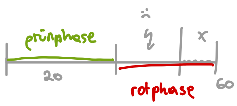
F_X ist stetig nicht fallend

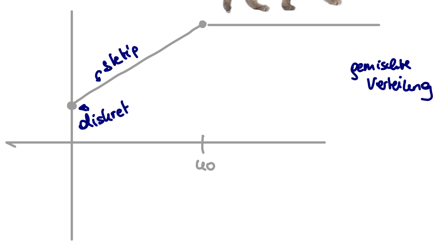
Sagt ein Autofahrer zum anderen am Schwarzenbergplatz: "Oh es ist rot, gemma auf an Kaffee!"

$\boxed{\text{BSP}}$ Ankunftszeit (bei Ampel) zeigt 20 sek grün

40 sek rot

X ist Wartezeit





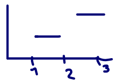
$\mathbb{P}(X=0) = \frac{20}{60}$

$0 \le X < 40$: $F_X(x) = \mathbb{P}(X \le x) = \frac{20+x}{60}$

stetig: alle reellen
diskret: endlich abzählbar
gemischt:

Verteilungsfunktion $F_X(x) = \mathbb{P}(X < x)$

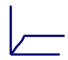
diskret: stückweise konstant, dazwischen Sprünge



nicht fallend
rechtsstetig

stetig: stückweise differenzierbar, Gleichverteilung auf $[0,1]$

$$F_X = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x \ge 1 \end{cases} \qquad f_X = F_X'(x) = \begin{cases} 0 & \text{sonst} \\ 1 & 0 \le x \le 1 \end{cases}$$



für 0 und 1 gelogen!

Eigenschaften:

- rechtsstetig
- nicht fallend
- $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$ "linkes Ende"
- $F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$ "rechtes Ende"

Dichtefunktion $f_X : \mathbb{R} \to \mathbb{R}^+$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Analogie zur Wahrscheinlichkeitsfunktion $p_X(x) = \mathbb{P}(X = x)$

$$F_X(x) = \sum_{t \le x} p_X(t)$$

Eigenschaften

- $f_X(x) \ge 0$
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Eigenschaften Wahrscheinl.funktion

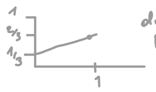
- $p_X(x) \ge 0$
- $\sum p_X(x) = 1$

gemischte Verteilungen

Hat sowohl eine "Dichtefunktion" als auch eine "Wahrscheinlichkeitsfunktion"

BSP Wartezeit F_X

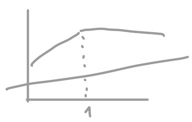
BSP komplizierte Wartezeit $\mathbb{P}(X=0) = \mathbb{P}(X=1) = \frac{1}{3}$ dazwischen gleichmäßig



Eigenschaften

- $F_X(x) = \sum p_X(t) + \int_{-\infty}^{\infty} f_X(t)\, dt$
- $p_X(0) = \frac{1}{2}$

$$f_X(x) = \begin{cases} \frac{1}{2} & 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$$



um auf nur WSL Fkt zu kommen $1 \to * \frac{1}{2}$

Dichte: $0 \to * \frac{1}{2}$

$$F_X(x) = \frac{1}{2} \sum_{t \le x} p_X^*(t) + \frac{1}{2} \int_{-\infty}^{\infty} f_X^*(t)\, dt$$

richtige WSLK

$$f_X^*(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$$

Mischung von F_1 und F_2 mit den Gewichten α und $1-\alpha$

$0 \le \alpha \le 1$

$$F(x) = \alpha \cdot F_1(x) + (1-\alpha) F_2(x)$$

Jede gem. Vert. kann als Mischung von stetig + diskreter Verteilung angesehen werden

Spezielle Verteilungen

1) Diskrete Verteilungen

endlich höchstens abzählbar viele endliche Werte

$D(a)$ deterministisch $\mathbb{P}(X=a)=1$ $\quad p_X(x) = \begin{cases} 1 & x=a \\ 0 & \text{sonst} \end{cases}$ aka degeneriert (verfallen)

$A(p)$ Alternativverteilung $\mathbb{P}(X=0)=1-p$, $\mathbb{P}(X=1)=p$ $\quad 0 \le p \le 1$ $\quad p_X(x) = \begin{cases} p: & x=1 \\ 1-p: & x=0 \\ 0 & \text{sonst} \end{cases}$

$D(a,b)$ Diskrete Gleichverteilung $\mathbb{P}(X=a) = \mathbb{P}(X=a+1) = \dots = \mathbb{P}(X=b) = \frac{1}{b-a+1}$ $\quad p_X(x) = \begin{cases} \frac{1}{b-a+1} & a \le x \le b, x \in \mathbb{Z} \\ 0 & \text{sonst} \end{cases}$

aka Laplacescher Wahr. Raum

$a,b \in \mathbb{Z}$ $\quad a \le b$

BSP Würfel $a=1$, $b=6$

$B(n,p)$ Binomialverteilung $p_X(x) = \begin{cases} \mathbb{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x} & 0 \le x \le n \\ 0 & \text{sonst} \end{cases}$

n unabhängige Versuche, Jeder Versuch Erfolg oder Misserfolg

X = Anzahl der Erfolge $\quad p = \mathbb{P}(\text{Erfolg})$

BSP $n=5$, $X=2$, $(p=\frac{1}{3})$ Möglichkeiten 2 Erfolge auf 5 Plätze aufzuteilen:

EEMMM, EMEMM, EMMEM, EMMME, MEEMM, MEMEM, MEMME, MMEEM, MMEME, MMMEE $\rightarrow 10$

$p_X(2) = \mathbb{P}(X=2) = 10\,p^2(1-p)^3 = \binom{5}{2} p^2 (1-p)^3$

$H(N,A,n)$ Hypergeometrische Verteilung $\quad p_X(x) = \begin{cases} \dfrac{\binom{A}{x}\cdot\binom{N-A}{n-x}}{\binom{N}{n}} \end{cases}$

$(n, A \le N)$

Fallunterscheidungen (fallen eigentlich weg:)

$[\max(0, n+A-N) \le x \le \min(n,A)$

BSP Urne: N Kugeln, A weiß, n Kugeln ohne Zurücklegen ziehen

$[N=5, A=3$, 3 bereits gezogen$]$

$G^*(p)$, $G(p)$ Geometrische Verteilung

$\mathbb{P}(X=x) = (1-p)^{x-1}\cdot p$ $\quad x=1,2\dots$ unabhängige Versuche bis zum ersten Erfolg, X = Anzahl der Versuche

$\mathbb{P}(X=x) = (1-p)^x \cdot p$ $\quad x=0,1\dots$ $\quad X$ = Anzahl der Misserfolge

$NB^*(n,p)$, $NB(n,p)$ Negative Binomialverteilung

$\mathbb{P}(X=x) = \binom{x-1}{n-1}^{x-n} p^n$ $\quad$ bis zum n-ten Erfolg $x=n,n+1$

$\mathbb{P}(X=x) = \binom{x+n-1}{n-1}(1-p)^x p^n = \binom{x+n-1}{x}(1-p)^x p^n$ $\quad$ zählen der Misserfolge $x=0,1\dots$

$\binom{x+n-1}{x} = \frac{(x+n-1)\dots(n+1)n}{x!}$ $\quad$ geht für jedes positive n, n muss nicht ganzzahlig sein

$P(\lambda)$ Poisson Verteilung (verteilung für seltene Ereignisse $\downarrow$ sehr viele Versuche mit kl. erfolgswahrsch.)

$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$ $\quad \lambda > 0$, $x=0,1\dots$

aka Grenzfall der Binomialverteilung $n\to\infty$, $n\cdot p \to \lambda$

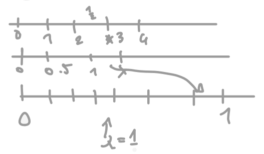
Spezielle Verteilungen (2)

stetige Verteilungen

stetige Gleichverteilung: $U(a,b)$ $\quad a<b$ $\quad$ Dichte $\downarrow$ $f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases}$

Standardgleichverteilung $U(0,1)$ $\quad a=0,\ b=1$

BSP Position von x des ersten markierten Intervalls



$\mathbb{P}(X \le x) = \mathbb{P}(\text{es gibt Markierung links von } x)$

$1-\mathbb{P}(\text{es gibt keine Markierung links von } x) = 1-\left(1-\frac{1}{2^{n+1}}\right)^{\lfloor 2^n \cdot x \rfloor} \longrightarrow 1-e^{-\frac{x}{2}}$

Exponentialverteilung $E(\lambda)$ $\quad \lambda > 0$ $\quad F(x) = \begin{cases} 1-e^{-\lambda x} & x \ge 0 \\ 0 & x<0 \end{cases}$ $\quad$ Dichte $f(x)$: $f(x) = \begin{cases} \lambda e^{-\lambda x} : & x \ge 0 \\ 0 : & x<0 \end{cases}$

Gammaverteilung: $\Gamma(\alpha, \lambda)$ $\quad f(x) = \frac{1}{\Gamma(\alpha)} \cdot \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}$ $\quad x>0$

$F(x)$ = möglich wenn α Ganzzahlig
→ partielle Integration
Grenzwert & Poissonvert. Mix

Vorkenntnisse:

- Gammafunktion $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\, dx$ $\quad \alpha > 0$
- Funktionalgleichung: $\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$ $\quad$ Interpolation der Fakultätsfunktion
  - $\Gamma(1) = 1$, $\Gamma(n) = (n-1)!$ $\quad n \in \mathbb{N}$
  - $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- Betafunktion: $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx$ $\quad \alpha, \beta > 0$
  - $B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$

Verteilungsfunktion
$F(x) = \int_{-\infty}^{\infty} f(u)\, du$
nicht geschlossen integrierbar

Betaverteilung (erster Art) $B_1(\alpha, \beta)$ $\quad f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ $\quad [0<x<1]$

Betaverteilung (zweiter Art) $B_2(\alpha, \beta)$ $\quad f(x) = \frac{1}{B(\alpha,\beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}$ $\quad 0<x$

Normalverteilung $N(\mu, \sigma^2)$ $\quad f(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\qquad N(0,1)$ $\quad \mu = 0$ $\quad \sigma^2 = 1$

Standardnormalverteilung $\quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\Phi(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$ $\quad$ Verteilungsfunktion der StNV

$N(\mu, \sigma^2) \rightarrow F_{\mu,\sigma^2}(x) = \mathbb{P}(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Errorfunction: $\operatorname{errfun}(x) = \int_0^x e^{-u^2}\, du$ $\quad \Phi(x) \ldots = \ldots \operatorname{errfun}\left(\frac{x}{\sqrt{2}}\right)$

$\boxed{\text{BSP}}$ $X \sim N(0,1)$

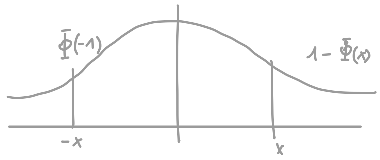
Die Verteilungsfunktion der Standardnormalverteilung:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

	0	1	2	3	4	5	6	7	8	9
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.567	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.691	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.736	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9	.816	.819	.821	.824	.826	.829	.831	.834	.836	.839
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6	.945	.946	.947	.948	.949	.951	.952	.953	.954	.954
1.7	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2	.986	.986	.987	.987	.987	.988	.988	.988	.989	.989
2.3	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995
2.6	.995	.995	.996	.996	.996	.996	.996	.996	.996	.996
2.7	.997	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9	.998	.998	.998	.998	.998	.998	.998	.999	.999	.999

$\mathbb{P}(X \leq 1{,}96) = 0.975$

$\Phi(1{,}96)$



gleich weil stetige Verteilung!

$\mathbb{P}(X \leq -1{,}96) = \mathbb{P}(X < -1{,}96) = 1 - 0.975 = 0.025$

$X \sim N(3, 16)$

$\mathbb{P}(-2{,}12 \leq X \leq 9{,}58) = \Phi\left(\frac{9.58-3}{4}\right) - \Phi\left(\frac{-2.12-3}{4}\right) = \Phi(1.645) - \Phi(-1{,}28) = 0.95 - (1-0.9) =$
$= 0.85$

R ... Statistikpaket pnorm (1.96)

pnorm (9.58; 3; sqrt (16))

$X \sim F_X$ suche x so, dass $F_X(x) = p$ $(0 \leq p \leq 1)$

solches x heißt p- Quantil $p = \frac{1}{2}$ Median

$p = \frac{1}{4}, \frac{3}{4}$ Quartile

Chi - Quadrat Verteilung χ_f $\chi_f^2 = \Gamma\left(\frac{f}{2}, \frac{1}{2}\right)$ f ... Anzahl der Freiheitsgrade

$\chi_2^2 = E\left(\frac{1}{2}\right)$

t Verteilung
F Verteilung
$\Big\}$ mehr als eine Zufallsvariable, X und Y diskret

... Mehr als eine Zufallsvariable ...

Zufallsvariable: gemeinsame Verteilung von zwei oder mehr Zufallsvariablen (mehrdimensional = Zufallsvektor)

$$\mathbb{P}_{X,Y}(x,y) = \mathbb{P}(X=x,\ Y=y)$$ ... gemeinsame Wahrscheinlichkeitsfunktion von X und Y

$\boxed{\text{BSP}}$ 2mal würfeln

X kleinere Augenzahl

Y größere Augenzahl

$\mathbb{P}_X(x) = \mathbb{P}(X=x) = \mathbb{P}(X=1) =$

$\mathbb{P}(X=1, Y=1) + \dots + \mathbb{P}(X=1, Y=6) = \frac{11}{36}$

mehrdimensionale Zufallsvariable $\quad Z(X,Y) \quad$ "Zufallsvektor"

Wahrscheinlichkeitsfunktion: $p_Z(z) = p_Z(x,y) - p_{X,Y}(x,y)$

stetige Verteilung

Verteilungsfunktion: $F_Z(z) = \mathbb{P}(Z=z) \qquad z(x,y) = \underbrace{\mathbb{P}(X,Y) \leq (x,y)}_{\text{soll}}$ heißen $X \leq x$ und $Y \leq y$

im diskreten Fall: $F_{X,Y}(x,y) = \sum\limits_{\substack{u \leq x \\ v \leq y}} \mathbb{P}(X=u, Y=v) = \sum\limits_{\substack{u \leq x \\ v \leq y}} P_{X,Y}(u,v)$ (2 Variablen)

$F_X(x) = \sum P_X(y) = \sum\sum \dots \sum P_{X_1,\dots X_n}(y_1 \dots y_n)$ (n Variablen)

$\boxed{\text{BSP}}$ $F_{X,Y}(x,y) = \begin{cases} 0 & x \leq 1,\ y < 1 \\ \vdots \end{cases} \qquad u = \begin{array}{c|c} 1 & v=1 \\ 2 & v=2 \\ 3 & v=3 \\ & \vdots \end{array} \qquad F_{XY}(3{,}5;2{,}8) = P_{XY}(1,1) + \dots + P(3,2) = 1{,}P$

mehrdimensionale Verteilungsfunktion $\quad F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x_1, \dots X \leq x_n) \qquad x = (x_1 \dots x_n)$, $x = (x_1 \dots x_n)$

Grenzen vertauscht

stetig: $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x_1, \dots X \leq x_n) = \int\limits_{-\infty}^{x_1} \dots \int\limits_{-\infty}^{x_n} f_{X_1 \dots X_n}(y_1 \dots y_n)\, dy_1 \dots y_n = \int\limits_{-\infty}^{x_n} \dots \int\limits_{-\infty}^{x_1} \dots$

Eigenschaften einer Wahrscheinlichkeitsfunktion in n Dimensionen

$P_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

$\sum\limits_{x \in \mathbb{R}^n} P_X(x) = 1 = \sum\limits_{x_1} \dots \sum\limits_{x_n} P_X(x_1 \dots x_n)$

Eigenschaften einer Dichtefunktion:

$f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

$\int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{+\infty} f_X(x)\, dx_1 \dots dx_n = f_x(x_1 \dots x_n) = f_{X_1 \dots X_n}(x_1 \dots x_n) = 1$

Eigenschaften einer Verteilungsfunktion: $F(x_1 \dots x_n)$ $\quad n=2$

(monoton, rechtsstetig, $x \to 0$ / $x \to \infty$ 1)

1) $F(x_1, x_2)$ rechtsstetig (wie in $\mathbb{R}$)

2) $F(x_1, x_2)$ nichtfallend

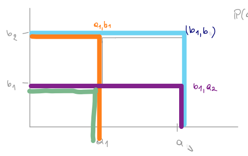
3a) $\lim\limits_{x_1 \to -\infty} F(x_1, x_2) = \lim\limits_{x_2 \to -\infty} F(x_1, x_2) = 0$

3b) $\lim\limits_{x_1, x_2 \to \infty} F_{X_1, X_2}(x_1, x_2) = 1$

3c) $\lim\limits_{x_2 \to +\infty} F_{X_1 X_2}(x_1, x_2) = F_{X_1}(x_1)$ Randverteilungsfunktion

4) $a_1 < b_1 \quad a_2 < b_2$

$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0$



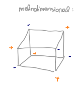
$\mathbb{P}(a_1 < X_1 \leq b_1,\ a_2 < X_2 \leq b_2)$

$a_1 < b_1$

$a_2 < b_2$

$F(b_1, b_2) - F(b_1, a_2) - F(a_1, a_2) + F(a_1, a_2)$

mehrdimensional:

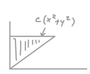


Additionstheorem

2^n Summanden

$\boxed{\text{BSP}}$ zum Punkt 4)

$f_{X,Y}(x,y) = f(x,y) = c(x^2+y^2): \quad 0 \leq x \leq y \leq 1$



1) c?

Randdichten $f_X, f_Y \qquad \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 0$

$1 = \int\limits_0^1 \int\limits_0^y (x^2+y^2)\,dx\,dy = c\int\limits_0^1 \left(\frac{x^3}{3} + y^2 x \Big|_0^y dy\right) = \int\limits_0^1 \frac{4y^3}{3}\,dy = c\,\frac{y^4}{3}\Big|_0^1 = \frac{c}{3} \implies c = 3$

$f_X(x) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \begin{cases} 0 & \text{sonst} \\ \int\limits_x^1 3(x^2+y^2)\,dy = 3x^2y + y^3\Big|_x^1 = 1 + 3x^2 - 4x^3 & 0 \leq x \leq 1 \end{cases}$

Verteilung: $F_X(x) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{+\infty} f_{X,Y}(u,y)\,dy\,du$

$f_Y(y) = \int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int\limits_0^y \dots = 4y^3 \qquad 0 \leq y \leq 1$

•) stetig, zweidimensional: $f_{X,Y}(x,y) = f(x,y) = c(x^2+y^2) \quad 0 \leq x \leq y \leq 1$

Erwartungswert

Gewinne $x_1 \dots x_n$ mit Wahrscheinlichkeiten $p(x_1), \dots p(x_n)$

N Versuche : $N \cdot p(x_1)$ mal $x_1 \to N \cdot p(x_1) \cdot x_1$

$\vdots$

$N\, p(x_n)$ mal $x_n \to N \cdot p(x_n)\, x_n$

$= N(p(x_1)x_1 + \dots p(x_n)x_n)$

X reelle Zufallsvariable

$\mathbb{E}(X) = \sum x\, p_X(x) = \int_{-\infty}^{\infty} x f_X(x)\, dx$ diskret

Annäherung:

x^1 auf ganze Zahlen gerundet $x^1 = \lfloor x \rfloor \cdot \mathbb{I}(x^{-1}) = \sum n\, \mathbb{P}(n \le X < n+1) = \sum_n n \int_n^{n+1} f_X(x) dx = \int_{-\infty}^{\infty} \lfloor x \rfloor f_X\, dx$

x^2 auf Zehntel gerundet

diskret $E(x) = \sum x\, p_X(x)$ gewichtetes Mittel

stetig $E(x) = \int_{-\infty}^{\infty} x f_X(x)\, dx \left(= \int_0^{\infty} x f(x) dx + \int_{-\infty}^{0} x f(x) dx\right)$

gemischt: $E(x) = \sum x\, p_X(x) + \int_{-\infty}^{+\infty} x f_X(x)\, dx$

alles was in diskret existiert, kann in stetig übersetzt werden

$\sum \to \int$ WSLKFKT → DICHTEFKT

Abhängigkeiten und Unabhängigkeiten von Zufallsvariablen

Zwei Ereignisse A, B heißen unabhängig, wenn $\mathbb{P}(A \wedge B) = \mathbb{P}(A)\,\mathbb{P}(B)$

Zwei Zufallsvariablen X und Y heißen unabhängig, wenn für alle $A, B \in \mathbb{R}$ gilt: $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$

oder $F_{XY}(x,y) = F_X(x) \cdot F_Y(y)$ oder (X diskret:) $P_{XY}(x,y) = p_X(x) \cdot p_Y(y) \to \mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y)$

oder (X stetig): $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

mehr als 2 Zufallsvariablen $x_1 \dots x_n$

"wenn jede endliche Auswahl unabhängig ist"

- paarweise unabhängig x_i, x_j sind unabhängig $\forall\ 1 \le i < j \le n$
- vollständig unabhängig $F_{x_1, \dots x_n}(x_1 \dots x_n) = F_{x_1}(x_1) \dots F_{x_n}(x_n)$
  - diskret : $P_{x_1, \dots x_n}(x_1 \dots x_n) = p_{x_1}(x_1) \dots p_{x_n}(x_n)$
  - stetig : $f_{x_1, \dots x_n}(x_1 \dots x_n) = f_{x_1}(x_1) \dots f_{x_n}(x_n)$

X und Y sind nicht unabhängig: $P_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y) = \underbrace{\mathbb{P}(X=x)}_{P_X(x)}\, \underbrace{\mathbb{P}(Y=y \mid X=x)}_{P_{Y|X}(Y|X)}$

$\left.\begin{matrix} P_{Y|X}(Y|X) \\ P_Y(y \mid X=x) \end{matrix}\right\}\ \mathbb{P}(Y=y \mid X=x) = \dfrac{P_{XY}(x,y)}{P_X(x)}$

Beispiele für Anwendung des Erwartungswertes

BSP Würfel, Augenzahl als Belohnung wird ca $6 \cdot 10^6$ mal wiederholt

(1)

$$\begin{aligned} 10^6 \cdot 1 &= 1 \cdot 10^6 \\ 10^6 \cdot 2 &= 2 \cdot 10^6 \\ 10^6 \cdot 3 &= 3 \cdot 10^6 \\ \vdots \quad & \quad 4 \cdot 10^6 \\ & \quad 5 \cdot 10^6 \\ & \quad 6 \cdot 10^6 \\ \hline & \quad 21 \cdot 10^6 \end{aligned}$$

Stundenlohn:

$$\frac{21 \cdot 10^6}{6 \cdot 10^6} = 3.5 \text{ pro Wurf}$$

BSP $P(X = 2^n) = \frac{1}{2^n}$ (St Petersburg Spiel)

(2) $$\mathbb{E}(X) = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum 1 = \infty$$

BSP $P(X = (-1)^n (n+1)) = \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} = 1$

(3) $\mathbb{E}(X) = \sum \left(\cancel{\frac{1}{n(n+1)}} \right)$ existiert nicht

BSP faires Würfel

(4) $$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

(5)

BSP $B(n,p)$

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

$$\mathbb{E}(X) = \sum x \binom{n}{x} p^x (1-p)^{n-x} = \sum x \frac{n!}{(n-x)!} p^x (1-p)^{n-x} =$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \underset{k=x-1}{=} \sum \frac{n! = n(n-1)!}{k!(n-1-k!)} p^{k+1} (1-p)^{n-1-k}$$

$$= \sum n \binom{n-1}{k-1} p^{k+1} (1-p)^{n-1k} = np \sum \underbrace{\binom{n-1}{k} p^k (1-p)^{n-1-k}}_{\text{Binomische Formel}}$$

$$\rightarrow \mathbb{E} = n \cdot p$$

BSP (6)

$$D(a,b): f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases}$$

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

(7)

BSP $N(0,1) \cdot \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{-x e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \Bigg|_{-\infty}^{\infty}$$

(8)

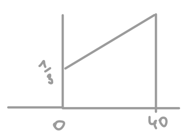
BSP Cauchyverteilung $f(x) = \frac{1}{\pi(1+x^2)}$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx$$

positiv: $$\int_0^{\infty} \frac{2x}{\pi(1+x^2)} dx = \frac{\log(1+x^2)}{2\pi} \Bigg|_0^{\infty} = \infty$$

negativ: $$\int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} d = -\infty$$

BSP (9) Ampel



$$F_X = \begin{cases} 0! & x < 0 \\ \frac{1}{3} + \frac{x}{60} & 0 < x < 40 \\ 1: & x \ge 40 \end{cases}$$

$$p_X(0) = \frac{1}{3}$$

$$f_X(x) = \frac{1}{60}$$

$$\mathbb{E}(X) = \underbrace{0 \cdot \frac{1}{3}}_{\sum x \, p_X(x)} + \int_0^{40} x \frac{1}{60} dx = 0 + \frac{40}{3}$$

Magie:

$$\left.\begin{array}{l} f_Y(y \mid X=x) \\ f_{Y|X}(y|x) \end{array}\right\} \frac{f_{X,Y}(x,y)}{f_X(x)}$$ bedingte Dichte

$$\mathbb{P}(Y \leq y \mid X=x) = \int_{-\infty}^{\infty} f_Y(t \mid X=x)\,dt$$

Weg um bedingte Wahrscheinlichkeit mit einem bedingten Ereignis mit WSLK 0 zu berechnen (Funktioniert nur wenn Bedingung ≥ 0)

kann als Grenzwert von $\mathbb{P}(Y \leq y \mid x-\varepsilon < X < x+\varepsilon)$ verstanden werden.

Bedingter Erwartungswert

Y, X diskret $\quad \mathbb{E}(Y \mid X=x) = \sum y \cdot p_Y(y \mid X=x) = \sum y\, \mathbb{P}_Y(Y=y \mid X=x) = \sum y \frac{p_{XY}(x,y)}{\mathbb{P}_X(x)}$

Y, X stetig $\quad \mathbb{E}(Y \mid X=x) = \int y \cdot f_Y(y \mid X=x)\, dy$

Eigenschaften des Erwartungswertes

1) Satz vom unachtsamen Statistiker

$X, g: \mathbb{R} \to \mathbb{R}, \quad Y = g(X) \qquad \mathbb{E}(Y) = \mathbb{E}(g(X)) = \begin{cases} \sum g(x)\, p_X(x) \\ \int g(x)\, f(x)\, dx \end{cases}$ geht auch mehrdimensional

2) Linearität

$\mathbb{E}(aX+b) = a\,\mathbb{E}(X) + b$

$\mathbb{E}(aX+b) = \sum (ax+b)\, p_X(x) = a \underbrace{\sum x\, p_X(x)}_{\mathbb{E}(X)} + b \cdot \underbrace{\sum p_X(x)}_{1} = \sum y\, \mathbb{P}(aX+b=y)$

$\mathbb{P}(g(X=y)) = \sum_{x: p(x)=y} \mathbb{P}(X=x)$

$\mathbb{E}(g(X)) = \sum y\, \mathbb{P}(g(X=y)) = \sum_y y \cdot \sum \mathbb{P}(X=x) = \sum_y \sum_{x: g(x)=y} y\, \mathbb{P}(X=x) = \sum \sum g(x)\, \mathbb{P}(X=x) = \sum_x g(x) \cdot \mathbb{P}_X(x)$

$\boxed{\text{BSP}}$ Augenzahl, $p = (X-3)^2$

$\mathbb{P}(g(X)=0) = \mathbb{P}((X-3)^2=0) = \mathbb{P}(X=3) = \frac{1}{6}$

$\mathbb{P}(g(X)=1) = \mathbb{P}((X-3)^2=1) = \mathbb{P}(X=2 \text{ or } X=4) =$

$= \mathbb{P}(X=2) + \mathbb{P}(X=4)$

3) Additivität

$X, Y \quad \mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) \qquad z = (X,Y)$ zweidimensional

$g(x,y) = x+y \qquad g(z) = X+Y$

$g(x,y) = x \qquad g(z) = X$

$g(x,y) = y \qquad g(z) = Y$

$\mathbb{E}(X+Y) = \mathbb{E}(g(z)) = \sum_{z \in \mathbb{R}^2} g(z) \cdot \mathbb{P}(Z=z) = \sum_{x,y} g(x,y) \cdot \mathbb{P}(X=x, Y=y) = \sum_{x,y} (x+y)\, \mathbb{P}(X=x, Y=y) = \sum_x \sum_y x\, \mathbb{P}(X=x, Y=y) + \sum_y \sum_x y\, \mathbb{P}(X=x, Y=y) =$

$= \underbrace{\sum x\, \mathbb{P}(X=x)}_{\mathbb{E}(X)} + \underbrace{\sum y\, \mathbb{P}(Y=y)}_{\mathbb{E}(Y)} = \mathbb{E}(X) + \mathbb{E}(Y)$

4) Monotonie

$X \leq Y \to \mathbb{E}(X) = \mathbb{E}(Y) \qquad$ Spezialfall $\quad X \geq 0 \to \mathbb{E}(X) \geq 0$

Wenn X und Y unabhängig sind: $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$

$\mathbb{E}(X \cdot Y) = \sum_{x,y} xy\, \mathbb{P}(X=x) \cdot \mathbb{P}(Y=y) = \sum_x \sum_y xy\, \mathbb{P}(X=x)\, \mathbb{P}(Y=y) = \sum_x x\, \mathbb{P}(X=x) \cdot \sum y\, \mathbb{P}(Y=y) = \mathbb{E}(Y) \cdot \mathbb{E}(X)$

$\boxed{\text{BSP}}$ $\mathbb{P}(X=-1) = \mathbb{P}(X=0) = \mathbb{P}(X=1) = \frac{1}{3}$

$Y = X^2 \qquad \mathbb{P}(Y=0) = \frac{1}{3}$

$\mathbb{P}(Y=1) = \frac{2}{3}$

$X, Y = X^3 - X \qquad \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$

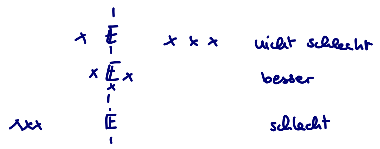
$\mathbb{E}(Y) = \frac{2}{3}$

$\mathbb{P}(X=0, Y=1) \neq \mathbb{P}(X=0)\, \mathbb{P}(Y=1)$

Philosophisch: Erwartungswert ist ein "Mittelwert" (vom Frequentistischen Standpunkt aus)

Lageparameter - typischer Wert

Verteilung von x um den $\mathbb{E}$



Wie genau gibt ein Lageparameter die Verteilung wieder?

Maßzahl: aka Mean Absolute Derivation → MAD mittlere absolute Abweichung von m

$MAD(m) = E(|X-m|)$

" in 0 hat man Ableitung 0"

natura non facis saltus

$m = \mathbb{E}(X)$ oder ein anderer Lageparameter

Betrag ist Pfui → viel beliebter: Quadrieren , MSD, MSQ , $MSE(m) = \mathbb{E}((X-m)^2)$

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

verschiebung ändert nichts an Varianz!

$\mathbb{V}(aX+b) = \mathbb{E}((aX+b-\mathbb{E}(aX+b))^2) = \mathbb{E}[(aX+b-a\mathbb{E}(aX+b)^2] = \mathbb{E}(a(X-\mathbb{E}X)^2) = \mathbb{E}(a^2(X-\mathbb{E}(X))^2) = a^2\,\mathbb{E}((X-\mathbb{E}(X)^2) = a^2\cdot\mathbb{V}(X)$

Standardabweichung / Streuung: $\sqrt{\mathbb{V}(X)}$ RMSE

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

Wenn X und Y unabhängig sind, gilt $Cov(X,Y) = 0$

Beweis: $\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) =$

$\mathbb{E}(XY - X\mathbb{E}(X) - Y\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)) =$

$\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) =$

$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Herleitung:

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

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

2x Produkt der

$= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,Cov(X,Y)$

unkorelliert: $\mathbb{E}(X)\cdot\mathbb{E}(Y) = \mathbb{E}(XY)$

$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$

Korrelation (Korrelationskoeffizient)

$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathbb{V}(X)\cdot\mathbb{V}(Y)}}$ $\quad -1 \leq \rho \leq 1$

Stoner - Movement - Senkeuce

Steinersche Verschiebungssatz:

$$\mathbb{V}(X) + \underset{a)}{0} + (\mathbb{E}(X)-m)^2$$

Herleitung

$$\mathbb{E}((X-m)^2) = \mathbb{E}((X-\mathbb{E}(X)+\mathbb{E}+m)^2) = \mathbb{E}\Big((X-\mathbb{E}(X))^2 + \underbrace{2(X-\mathbb{E}(X))(\mathbb{E}(x)-m)}_{2\cdot(\mathbb{E}(X)-m)\underbrace{(\mathbb{E}(x-\mathbb{E}(x)))}_{a)\ \mathbb{E}(x)-\mathbb{E}(x)=0}} + (\mathbb{E}(X)-m)^2\Big)$$

WIR SEHEN: $m=0 \Rightarrow \mathbb{E}(X^2) = \mathbb{V}(x) + (\mathbb{E}(x))^2$

1) $\mathbb{V}(x) = \mathbb{E}(x^2) - (\mathbb{E}(x))^2$

2) $\mathbb{E}(X^2) \geq \mathbb{E}(x)^2$

3) $\mathbb{E}(x)$ ist bester Lageparameter (der mit der kleinsten Abweichung)

$\boxed{\text{BSP}}$ $f(x,y) = 3(x^2+y^2)$: $0 < x < y < 1$

$$\rho(X,Y) = \frac{Conv(X,Y)}{\sqrt{\mathbb{V}(X)\ \mathbb{V}(Y)}}$$

$\mathbb{E}(X)$, $\mathbb{E}(y)$, $\mathbb{E}(X^2)$, $\mathbb{E}(XY)$, $\mathbb{E}(y^2)$

$$\mathbb{E}(XY) = \int_0^1\int_0^y 3xy(x^2+y^2)\,dx\,dy = 3\int_0^1\int_0^y x^{1+2}y^1 + y^{1+2}\,dx\,dy = 3\int_0^1 \frac{y^{1+2+1}}{1+2+1}y^1 + \frac{y^{1+1}}{1+1}y^{1+2}\,dy = 3\left[\frac{1}{1+2+1}+\frac{1}{1+1}\right]\frac{1}{1+1+1+2+1} = \frac{3}{8}$$

$= 3\cdot\left(\frac{1}{4}+\frac{1}{2}\right)\cdot\frac{1}{6} =$

Durch a,b ersetzen

$$3\int_0^1\int_0^y x^{a+2}y^b + x^a y^{b+2}\,dx\,dy = 3\int_0^1 \frac{y^{a+2+1}}{a+2+1}y^b + \frac{y^{a+1}}{a+1}y^{b+2}\,dy = \frac{3(2a+4)}{(a+3)(a+1)(a+b+4)}$$

Formel für Erwartungswertberechnungen !!

$a=0, b=1 \rightarrow \mathbb{E}(Y) = \frac{3\cdot4}{3\cdot1\cdot5} = \frac{4}{5}$

$a=1, b=0 \rightarrow \mathbb{E}(X) = \frac{3\cdot6}{4\cdot2\cdot5} = \frac{9}{20}$

$a=0, b=2 \rightarrow \mathbb{E}(Y^2) = \frac{3\cdot4}{3\cdot6} = \frac{2}{3}$

$a=2, b=0 \rightarrow \mathbb{E}(X^2) = \frac{4}{15}$

$\mathbb{V}(X) = \frac{4}{15} - \left(\frac{9}{20}\right)^2 = \frac{320-243}{1200} = \frac{72}{1200}$

$\mathbb{V}(Y) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{50-48}{75} = \frac{2}{75}$

$Cov(X,Y) = \frac{3}{8} - \frac{9}{20}\cdot\frac{4}{5} = \frac{75-72}{300} = \frac{3}{200}$

$\rho = \frac{\frac{3}{200}}{\sqrt{\frac{77}{1200}\cdot\frac{2}{75}}} = \frac{9}{2\sqrt{154}} \approx 0.36$

$\boxed{\text{BSP}}$ $f(x,y) = c\cdot\frac{x}{y}$ $\quad 0 < x < y < 1$

mit zweidimensionaler Dichte arbeiten

$X \sim B(n,p)$, $\mathbb{E}(X) = np \rightarrow \mathbb{V}(X)$?

Uff ?? *

$$\mathbb{E}(x^2) = \sum_{k=0}^n k^2\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^n k^2\binom{n}{k}p^k(1-p)^{n-k} = \dots = np\sum k\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k} =$$

$$= np\sum_{l=0}^{n-1}(l+1)\binom{n-1}{l}p^l(1-p)^{n-1-l} =$$

$$= np\,[(n-1)p+1] = n^2p^2 - np^2 + np$$

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

* Trick für den Kauderwelsch: Indikatorvariable 0 oder 1

$$Y_i = \begin{cases} 1 & i\text{ter Versuch} = \text{Erfolg} \\ 0 & \text{Misserfolg} \end{cases}$$

$$X = \sum_{i=1}^n Y_i$$

$$\mathbb{E}(X) = \sum \mathbb{E}(Y_i) = np$$

weil unabhängig !

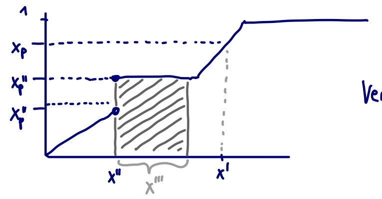
$$\mathbb{V}(X) = \sum \mathbb{V}(Y_i) = n(p-p^2)$$

Quantil:

(also p Fraktil)

x_p heißt p Quantil wenn $F_X(x_p - 0) \leq p \leq F_X(x_p)$

$$\mathbb{P}(X < x_p) \leq p \leq \mathbb{P}(X < x_p)$$



Versuchen F_X umzukehren

Wahl des Praktikers: Mittelpunkt (Mittelwert) von kleinstem und größten Quantil

Wahl des Mathematikers: linker Endpunkt (kleinster Wert)

verallgemeinerte Inverse: $F^{-1}(y) = \inf\{x : F_{(x)} \geq y\} =$

- x_p wenn $\exists! x_p : F_X(x_p) = p$
- x wenn $F(x)$ über y springt $(\nexists x : F_X(x) = y)$
- kleinstes x wenn es mehr als ein x gibt für $F_X(x) = y$

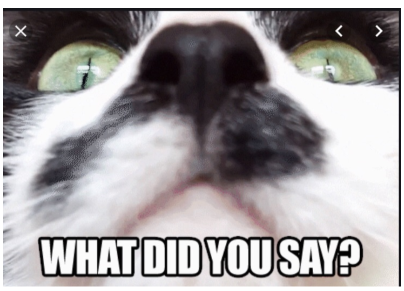
Für welches m wird $f(m) = \mathbb{E}(|X-m|)$ minimal?

solange: $f'(m) = 0 \quad \frac{d}{dm}\mathbb{E}(|X-m|) = 0$

$$\mathbb{E}\left(\frac{d}{dm}|X-m|\right) = \mathbb{E}\left(\operatorname{sgn}(X-m)(-1)\right) = 0$$

$$= (-1)\left[\mathbb{P}(X>m) - \mathbb{P}(X<m)\right] \qquad \mathbb{P}(X=m) \cong 0$$

$$\Rightarrow \mathbb{P}(X \leq m) = \frac{1}{2} \Rightarrow F_X(m) = \frac{1}{2}$$



Transformationssätze:

X Zufallsvariable (ein oder mehrdimensional)
Funktion $g: \mathbb{R} \to \mathbb{R}$ oder $\mathbb{R}^n \to \mathbb{R}$ oder $\mathbb{R}^n \to \mathbb{R}^n$

X diskret: mit möglichen Werten $x_1, \dots, x_k, \dots$
möglichen Werte für $g(x)$ sind $g(x_1), g(x_2), \dots$ müssen nicht alle verschieden sein!

$$P_Y(y) = \mathbb{P}(Y=y) = \mathbb{P}(g(X)=y) = \sum_{x: g(x)=y} \mathbb{P}(X=x) = \sum_{x: g(x)=y} P_X(x)$$

BSP $U(0,1)$ Gleichverteilung
$g(x) = x^3$ $\quad g(x) = x^a$ $\quad a > 0$
$Y = g(X) = X^3 \; (X^\alpha)$

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X^3 \le y) = \mathbb{P}(X \le \sqrt[3]{y}) = y^{\frac{1}{3}}$$
$$\mathbb{P}(X^\alpha \le y) = \mathbb{P}(X \le y^{\frac{1}{\alpha}}) = y^{\frac{1}{a}}$$
$$\begin{cases} 0: & y < 0 \\ & 0 \le y < 1 \\ 1: & y \ge 1 \end{cases}$$

$$f_Y(y) = F_Y'(y) = \begin{cases} \frac{1}{a} y^{\frac{1}{a}} & \to 0 \le y \le 1 \\ 0 & \to \text{sonst} \end{cases}$$

Allgemein: g umkehrbar $\quad \overset{1.\text{ Fall}}{g' > 0}$ oder $\overset{2.\text{ Fall}}{g' < 0}$

2. Fall:

$g(\infty) = \lim_{x \to \infty} g(x)$
$g(-\infty) = \lim_{x \to -\infty} g(x)$

BSP $g(x) = e^{-x}$ $\quad g(\infty) = 0$, $g(-\infty) = \infty$

$Y = g(X)$ $\quad X \sim f_X = f$ $\quad$ mögliche Werte für y: $\{g(x) : x \in \mathbb{R}\} = \,]g(\infty), g(-\infty)[$ $\quad$ Transformierte Zufallsvariable hat nur positive Werte

$g(\infty) < y < g(-\infty)$

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \mathbb{P}(X > g^{-1}(y)) = 1 - \mathbb{P}(X < g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

kleiner als $g(x) \to 0$
größer als $g(x) \to 1$

$$f_Y(y) = F_Y'(y) = -(g^{-1})'(y) \cdot f_X(g^{-1}(y)) = |(g^{-1})'(y)| f_X(g^{-1}(y)) = \frac{1}{|g'(g^{-1}(y))|} f_X(g^{-1}(y)) \qquad y \in g(\mathbb{R}) = \{g(x) : x \in \mathbb{R}\}$$

$$f_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f_X(g^{-1}(y)) \cdot [y \in g\mathbb{R}] = \frac{1}{\left|\frac{d\, g(x)}{dy}\right|} \cdot f_X(g^{-1}(y)) [y \in g\mathbb{R}]$$

BSP $g(x) = e^{-x}$ $\quad X \sim E(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

$$g^{-1}(y) = -\log(y) \quad \frac{dg^{-1}(y)}{dy} = -\frac{1}{y} \qquad \frac{dg}{dx}(x) = -e^{-x}$$

$$f_Y(y) = \left|-\frac{1}{y}\right| \lambda e^{-\lambda(-\log y)} : \begin{cases} y > 0 \,/\, -\log y > 0 & \log y < 0 \Leftrightarrow y < e^0 = 1 \quad = \lambda \cdot \frac{1}{y} = \lambda y^{\lambda - 1} \\ y \le 1 \quad 0 & \text{weil } f_X(g^{-1}(y)) = 0 \\ y \le 0 \quad 0 & \text{weil } g^{-1}(y) \text{ nicht definiert} \end{cases}$$

BSP $X \sim N(0,1)$ $\quad g(x) = x^2$ $\quad Y = X^2$

$$y \ge 0 \quad F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(|X| \le \sqrt{y}) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{2\sqrt{y}} \varphi(\sqrt{y}) + \frac{1}{2\sqrt{y}} \varphi(-\sqrt{y}) = \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} + \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{\frac{(-\sqrt{y}^2)}{2}} = \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} + \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} : y > 0 \; = \frac{y^{\frac{1}{2}-1} (\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} e^{-\frac{y}{2}}$$

↳ Gammaverteilung mit $a = \frac{1}{2}$ $\; \gamma = \frac{1}{2}$

$$f_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f_X(g^{-1}(y)) \cdot [y \in g\mathbb{R}] = \frac{1}{\left|\frac{d\, g(x)}{dy}\right|} \cdot f_X(g^{-1}(y)) [y \in g\mathbb{R}]$$

(Funktioniert genauso wenn $g: \mathbb{R}^n \to \mathbb{R}^n$, g diff, umkehrbar)

$$g(x) = \begin{pmatrix} g_1(x_1, \dots x_n) \\ g_n(x_1 \dots x_n) \end{pmatrix} \to \frac{dg}{dx} := \begin{pmatrix} \frac{dg_1}{x_1} & \dots & \frac{dg_1}{x_n} \\ \frac{dg_n}{x_1} & \dots & \frac{dg_n}{x_n} \end{pmatrix}$$

Anwendung: Verteilung einer Summe $\quad X, Y$ gemeinsame Verteilung mit Dichte $f_{X,Y}$ $\qquad S = X+Y \quad T = X$

$$(S, T) = g(X, Y) = (X+Y, X)$$

$g_1(x,y) = x+y$ $\qquad g^{-1}(s,t) = (t, s-t)$

$g_2(x,y) = x$

$g(x,y) = (\underset{s}{x+y}, \underset{t}{x})$

$$\frac{dg^{-1}(s,t)}{d(s,t)} = \begin{vmatrix} \frac{dg_1^{-1}}{ds} & \frac{dg_1^{-1}}{dt} \\ \frac{dg_2^{-1}}{ds} & \frac{dg_2^{-1}}{dt} \end{vmatrix} = \begin{vmatrix} \frac{dt}{ds} & \frac{dt}{dt} \\ \frac{d(s-t)}{ds} & \frac{d(s-t)}{dt} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

$$f_{S,T} = |-1| \cdot f_{XY}(g^{-1}(s,t)) = f_{XY}(t, s-t)$$

Dichte von S: Randdichte $f_S(s) = \int_{-\infty}^{\infty} f_{S,T}(s,t)\,dt = \int f_{X,Y}(t, s-t)\,dt$

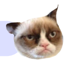
X und Y unabhängig: $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$ $\qquad f_S(s) = \int_{-\infty}^{\infty} f_X(t) f_Y(s-t)\,dt = \overbrace{f_X * f_Y}^{\text{Faltung!}}(s)$

im Diskreten: X, Y sind diskret $\quad S = X+Y$

$$P_S(s) = \mathbb{P}(S=s) = \mathbb{P}(S = X+Y = s) = \sum_x \mathbb{P}(X=x, X+Y=S) = \sum \mathbb{P}(X=x, Y=s-x) = \sum P_X(x, s-x)$$ statt $x = t$ → von stetig in Diskret

wenn X, Y unabhängig

$$P_S(s) = \sum_x p_X(x)\, p_Y(s-y) = P_X * P_Y(s)$$ diskrete Faltung!

Transformationsformel 

1) $g: \mathbb{R}^n \to \mathbb{R}^n$

$$f_{g(x)}(y) = \begin{cases} f_x(g^{-1}(y)) \left| \frac{\partial g^{-1}}{\partial y}(y) \right| : y \in g(\mathbb{R}^n) \\ 0 \quad \text{sonst} \end{cases}$$ (heißt: $\exists\, x \in \mathbb{R}^n$ mit $y = g(x)$)

X und Y unabhängig und stetig

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx = f_X * f_Y$$ Faltung

$$f_{\frac{X}{Y}}(t) = \int_{-\infty}^{\infty} |y|\, f_Y(y) \cdot f_X(z \cdot y)\, dy =$$

Hinweis:
$\Gamma(\alpha, \lambda) * \Gamma(\beta, \lambda) = \Gamma(\alpha+\beta, \lambda)$

$\boxed{\text{BSP}}$ $X, Y \sim E(\lambda)$ unabhängig $\qquad f_X(x) = f_Y(x) = \lambda e^{-\lambda x} \quad [x \geq 0]$

$S = X+Y$
$Q = X/Y$

integrand null wenn x negativ
$x > z \to x$ negativ

$$f_S(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)}\, dx = \lambda^2 e^{-\lambda z} \int_0^z dx = \begin{cases} z \geq 0 : \lambda^2 z e^{-\lambda z} \\ z < 0 : 0 \end{cases}$$

$$f_{\frac{X}{Y}}(z) = \begin{cases} z \geq 0 : \int_{-\infty}^{\infty} |y|\, f_Y(y) \cdot f_X(zy)\, dy = \int_0^{\infty} |y|\, \lambda^2 e^{-\lambda y(1+z)}\, dy = \int_0^{\infty} \frac{u e^{-\lambda u}}{(1+z)^2} = \frac{1}{(1+z)^2} \\ \\ z < 0 : 0 \end{cases}$$

man kann Verteilungen transformieren.

$g(x,y) = (x+y, \frac{x}{y})$, $\quad g^{-1}(s,q)$, $\quad s = x+y = y(1+q)$

$$y = \frac{s}{1+q}$$
$$x = \frac{sq}{1+q}$$

Funktionaldeterminante: $$\begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial q} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial q} \end{vmatrix} = \begin{vmatrix} \frac{q}{1+q} & \frac{1}{1+q} \\ \frac{s}{(1+p)^2} & -\frac{s}{(1+q)^2} \end{vmatrix} = -\frac{sq}{(1+q)^3} - \frac{s}{(1+p)^3} = -\frac{s}{(1+p)^2}$$

$$f_{S,Q}(s,q) = f_X\left(\frac{sq}{1+q}\right) f_Y\left(\frac{s}{1+p}\right) \left| -\frac{s}{(1+p)^2} \right| = \lambda e^{-\lambda \frac{sq}{1+p}} \cdot \lambda e^{\lambda \frac{s}{1+q}} \frac{s}{(1+p)^2}$$

$$= s\lambda^2 e^{-\lambda s} \frac{1}{(1+p)^2} \quad s \geq 0 \quad p \geq 0$$ ⇒ S, Q unabhängig

Multivariate Verteilung mit Namen: (Anwendung der Transformationsformel)

1) Multinomialverteilung (diskret, verallg. von Binomial): n Versuche mit Ausgängen $1, \dots k$ $\qquad$ $k=2$ Binomialverteilung

$(X_1 \dots X_k)$ Ergebnis $\quad X_i \dots$ # Ausgänge i $\quad$ Wahrscheinlichkeiten $p_1 \dots p_k$

$$\mathbb{P}_X(x) = \mathbb{P}_{X_1 \dots X_k}(x_1 \dots x_k) = \begin{cases} \frac{n!}{x_1! x_i! \dots x_k!} p_1 \cdot p_2 \cdot \dots p_k & x_i \geq 0, \; x_1 + \dots + x_k = n \\ 0 & \text{sonst} \end{cases}$$

2) Normalverteilung (stetig) $\quad N(0,1)$

$Y_1 \dots Y_n$ unabhängig $\rightarrow$ (aufschreiben als Zeilenvektoren

$$f_Y(y) = f_{Y_1 \dots Y_n}(y_1 \dots y_n) = \prod \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum y_i^2}{2}} = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{y^T y}{2}}$$

$X = AY + b \quad A \in \mathbb{R}^{n \times n}, \; b \in \mathbb{R}^n$

A regulär: $g(x) = Ax + b$

umkehrbar: $g^{-1}(y) = A^{-1}(y-b)$

statt y jetzt x $\rightarrow$ Spaltenvektoren

$$f_X(x) = f_{X_1 \dots X_n}(x_1 \dots x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{\sum x_i^2}{2}} = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{x^T x}{2}}$$

$Y = AX + b \quad A \in \mathbb{R}^{n \times n}, \; b \in \mathbb{R}^n$

A regulär $g(x) = Ax + b$

umkehrbar $g^{-1}(y) = A^{-1}(y-b)$

$N(b, \Sigma)$

Dichte einer mehrdim. Norm v.

$$f_Y(y) = f_X(A^{-1}(y-b)) \, |\det(A^{-1})| = \frac{1}{\sqrt{(2\pi)^n}} e^{-(y-b)^T (A^T)^{-1} A^{-1} (y-b)} \frac{1}{|\det(A)|} = \frac{1}{\sqrt{(2\pi)^n}} e^{-(y-b)^T (AA^T)^{-1}(y-b)} \frac{1}{|\det A|} = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{(y-b)^T \Sigma^{-1} (y-b)}{2}}$$

$\Sigma = AA^T$

$\det \Sigma = \det(AA^T) = \det(A) \det(A^T) = (\det(A))^2$

$b \dots$ Mittelwert, Σ Kovarianzmatrix

gemeinsame Normalverteilung, unabhängig wenn Kovarianz $= 0$!

Momente

n-tes Moment von X $\quad M_n = \mathbb{E}(X^n)$

n-tes zentriertes Moment $\quad m_n = \left(\left(\mathbb{E}(X - \mathbb{E}(X)\right)^n \right)$

n-tes absolutes Moment $\quad \mathbb{E}(|X^n|)$

n-te absolute zentrale Moment

3. Moment: $m_3 \qquad \frac{m_3}{\sqrt{(\mathbb{V}(x))^3}} = \frac{\mathbb{E}((x-\mathbb{E}(x))^3)}{\sqrt{(\mathbb{E}((x-\mathbb{E}(x))^2)^3}}$

m

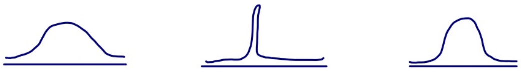
Schiefe $= 0$

rechtsschief

Schiefe > 0

4. Moment: $m_4 \qquad K = \frac{m_4}{\mathbb{V}(x)^2} - 3$

Kurtosis/ Wölbung $\rightarrow$



$K = 0$ schöner Rand

$K > 0$ spitz

$K < 0$ flach

Sidenote:

(a_n): $\sum_{n=0}^{\infty} a_n z^n$ Erzeugende Funktion, z Transformierte

$a_n = \frac{g^{(n)}(0)}{n!}$

Wahrscheinlichkeitserzeugende Funktion $\rightarrow$ X diskret (nicht negativ, ganzzahlig)

$$\mathbb{E}(z^X) = \sum_{x=0}^{\infty} z^x \, \mathbb{P}(x = X) = \sum_{x=0}^{\infty} z^x p_X(x) \qquad (\text{geht für } |x| \leq 1)$$

$$g'(1) = \sum_{x=0} x \, z^{x-1} p_X(x) \Big|_{z=1} = \sum x \, p_X(x) = \mathbb{E}(X)$$

Faktorielles Moment $\quad g''(1) = \mathbb{E}(X(X-1))$

n Faktorielles Moment: $g^n(1) = \mathbb{E}(X!)$?

$X \sim p_X$, $Y \sim p_Y$ $\quad X, Y$ unabhängig $\Rightarrow X + Y \sim p_X * p_Y$

$$g_X(z) = \sum_{n=0}^{\infty} p_X(n) z^n = \mathbb{E}(z^X)$$

$$g_Y(z) = \sum_{n=0}^{\infty} p_Y(n) \cdot z^n = \mathbb{E}(z^Y)$$

$$g_{X+Y}(z) = \mathbb{E}(z^{X+Y}) = \mathbb{E}(\underbrace{z^X \cdot z^Y}_{\text{unabhängig}}) = \mathbb{E}(z^X) \cdot \mathbb{E}(z^Y) = g_X(z) \cdot g_Y(z)$$

exponentielles Moment: $M_X(t) = \mathbb{E}(e^{tX})$ geht für stetige und diskrete Zufallsvariablen (WENN es geht!) $\quad$ Cauchy Verteilung :(

für $X \geq 0$, $t < 0$

$$L_X(t) = M_X(-t) = \mathbb{E}(e^{-tX}) = \underbrace{\int_0^{\infty} f_X(x) e^{-xt} \, dt}_{\text{Laplacetransformierte}} \quad \text{wenn } x \text{ stetig}$$

Momentenerzeugende Funktion:

$$M_X(t) = \mathbb{E}\left(\sum \frac{x^n t^n}{n!}\right) = \sum \frac{t^n \mathbb{E}(x^n)}{n!} = \sum_{n=0}^{\infty} \frac{M_n t^n}{n!} \quad \text{erzeugende Funktion für Momente}$$

$$e^{xt} = \sum \frac{x^n t^n}{n!}$$

$$M_n = \mathbb{E}(X^n) = M_X{}^{n}(0)$$ "nulltes Moment" $= 1$

Charakteristische Funktion (der Verteilung von x)

$\hookrightarrow$ legt Verteilung eindeutig fest!

$$|e^{ixt}| = |\cos(Xt) + i \sin(Xt)| = 1$$

$$\varphi_X(t) = \mathbb{E}(e^{iXt}) = \mathbb{E}(\cos(xt)) + i \, \mathbb{E}(\sin(xt))$$

BSP Poissonverteilung $p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$

$$g_X(t) = \sum \frac{z^x \lambda^x}{x!} e^{-\lambda} = e^{\lambda(z-1)}$$

$$M_X(e^t) = e^{\lambda(e^t - 1)} \qquad \text{char fkt:} \quad \varphi_X(t) = M_X(e^{it}) = e^{\lambda(e^{it}-1)}$$

BSP $N(0,1)$

$$M_X(t) = \int_{-\infty}^{\infty} e^{xt} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2-2xt}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-t)^2-t^2}{2}} dx = e^{\frac{t^2}{2}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx}_{1} = e^{+\frac{t^2}{2}}$$

als Momentenerzeugende

$$\varphi_X(t) = M_X(it) = e^{\frac{(it)^2}{2}} = e^{-\frac{t^2}{2}}$$

Inversenmethode:

Möglichkeit, Zufallszahlen mit beliebigen Verteilungen zu erzeugen.

X F_X stetig, streng Monoton ($f_X > 0$), F_X umkehrbar

$$U = F_X(X) \in [0,1]$$

$$0 < u < 1: \quad \mathbb{P}(U \le u) = \mathbb{P}(F_X(X) \le u) = \mathbb{P}(X \le F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$$

$U = F_X(X) \sim U(0,1)$ "U ist gleichverteilt auf 0 und 1"

$$X = F_X^{-1}(U)$$

Wenn F eine stetige, streng monotone Verteilungsfunktion ist dann

1) X nach F verteilt, dann X gleichverteilt auf $0,1$ $\quad F(X) \sim U(0,1)$ geht noch allgemeiner: nur stetig ✓

2) wenn U auf $0,1$ gleichverteilt, dann $\quad U \sim U(0,1) \Rightarrow \underbrace{F^{-1}(U)}_{X} \sim F$

weil: $\mathbb{P}(X \le x) = \mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F(x)$

Normalverteilte Zufallsvariablen → in inv. Norm. Vert. einsetzen

Binomial geht auch (obwohl nicht stetig, diskret)
↳ verallgemeinerte Inverse

BSP Exponentialverteilung

$$F(x) = 1 - e^{-\lambda x}$$

$$F^{-1}(y) = -\frac{1}{\lambda} \log(1-y)$$

$$X = F^{-1}(U) = -\frac{1}{\lambda} \log(1-U) \sim E(\lambda)$$

leichte Verbesserung $X = -\frac{1}{\lambda} \log(U)$

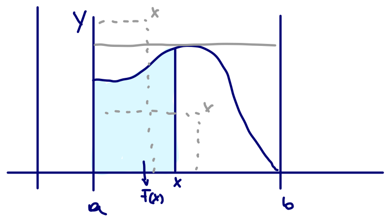
$$U = \frac{\text{random()} + 0.5}{\text{RAND\_MAX} + 1.01}$$

U_1, U_2

$$X_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2)$$

$$X_2 = \sqrt{-2 \log U_2} \cos(2\pi U_1)$$

Annahme – Verwerfungsmethode



$a \le X \le b$

Dichte $f_X = 0$ für $x < a,\ x > b$

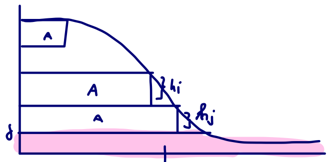
$F(x) = \int_a^x f(u)\,du$

$U \sim U(a,b)$ $\quad Y > f(X) \rightarrow$ war nix weg werfen neu versuchen

$X \sim U(0,M)$ $\quad Y \le f(X) \rightarrow X$ Zufallszahl → passiert mit Wahrscheinlichkeit $\frac{\int f(x)\,dx}{M(b-a)} = \frac{1}{M(b-a)}$

Ziggurat Methode

↓ Babylonische Türme



A gibt an wie groß Fläche ist

Dichte auf $x \ge 0$ fallend

Wähle h so, dass möglichst jedes □ Fläche A hat; stelle δ so ein, dass sich oberstes Rechteck genau ausgeht und 2^n Stufen entstehen

(auch für Normalverteilung)

Skript S. 30 : Folgen von Zufallsvariablen

Gesetz der großen Zahlen

$(X_1, \dots X_n)$ Folge von Zufallsvariablen, unabhängig, identisch verteilt $\qquad F_X$ Verteilungsfunktion

Erwartungswert: $\mathbb{E}(X) = \mathbb{E}(X_n) = \mu$

Varianz: $\mathbb{V}(X) = \sigma^2$

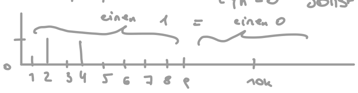
Stichprobenmittel: $\bar{X}_n = \dfrac{X_1 + \dots + X_n}{n}$ $\qquad \bar{X}_n \to \mu$ wenn $n \to \infty$, wie soll diese Konvergenz verstanden werden?

Möglichkeiten der Konvergenzdefinition: Y_n, Y sind Zufallsvariablen

1) $Y_n \to Y$ in Wahrscheinlichkeit, wenn $\forall \varepsilon > 0 : \lim\limits_{n\to\infty} \mathbb{P}(|Y_n - Y| > \varepsilon) = 0$

2) $Y_n \to Y$ mit Wahrscheinlichkeit 1 wenn $\mathbb{P}(Y_n \overset{\text{konvergiert nicht gegen}}{\not\to} Y) = \mathbb{P}(|Y_n - Y| > \varepsilon) = 0$ $\quad \infty$ oft

$z_k \sim D(1, 2^k)$ $\qquad k = 0, 1, \dots$ $\qquad \begin{cases} Y = 1 & \text{wenn } 2^k + z_k \\ Y_n = 0 & \text{sonst} \end{cases}$ $\qquad 2 < n \le 2^{k+1}$: $\mathbb{P}(Y_n = 1) = \frac{1}{2^k}$



nur für 1 Fall daneben ???

Wenn $\sum \mathbb{P}(|Y_n - Y| > \varepsilon) < \infty$ für alle $\varepsilon > 0$, dann konvergiert es mit WSLK 1.

3) $Y_n \to Y$ konvergieren im Quadratmittel, wenn $\mathbb{E}((Y_n - Y)^2) \to 0$

Schwaches Gesetz der Großen Zahlen

X_n unabhängig, identisch verteilt, $\bar{X}_n \to \mu$ in WSLK, $\mathbb{E}(X_n) = \mu$, $\mathbb{V}(X_n) = \sigma^2$

1) Markov: $Y \ge 0 \quad c > 0$

$$\mathbb{P}(Y \ge c) \le \frac{\mathbb{E}(Y)}{c}$$

Bew: $Z = \begin{cases} c & Y \ge c \\ 0 & \text{sonst} \end{cases}$ $\qquad Z \le Y$: $\overbrace{\mathbb{E}(Z)}^{c \cdot \mathbb{P}(Y \ge c)} \le \mathbb{E}(Y)$

2) Chebychev:

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \ge c) \le \frac{\mathbb{V}(Y)}{c^2}$$

$$\mathbb{P}(\underbrace{(Y - \mathbb{E}(Y))^2}_{Z} \ge c) \le \frac{\mathbb{E}(Z)}{c^2} = \frac{\mathbb{E}((Y - \mathbb{E}(Y))^2)}{c^2} = \frac{\mathbb{V}(Y)}{c^2}$$

- $\mathbb{E}(\bar{X}_n) = \mathbb{E}(\frac{1}{n}(X_1 + \dots + X_n)) = \frac{1}{n}\mathbb{E}(X_1 + \dots + X_n) = \frac{1}{n}[\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)] = \frac{n \cdot \mu}{n} = \mu$ $\qquad$ The mean of the mean is the mean (?)
- $\mathbb{V}(\bar{X}_n) = \mathbb{V}(\frac{1}{n}(X_1 + \dots + X_n)) = \frac{1}{n^2}\mathbb{V}(X_1 + \dots + X_n) = \frac{1}{n^2}[\mathbb{V}(X_1) + \dots + \mathbb{V}(X_n)] = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$

Man wendet Chebychev an

$$\mathbb{P}(|\underbrace{\bar{X}_n}_{\mathbb{E}(\bar{X}_n)} - \mu| \ge \varepsilon) \le \frac{\mathbb{V}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2} \to 0$$

Es wurde gezeigt: 1) $\bar{X}_n \to \mu$ im Quadratmittel (immer stärkere Versionen)

2) es genügt (X_n) unkorreliert (Vorraussetzung aufweichen)

Starkes Gesetz der großen Zahlen

selbe Voraussetzungen wie vorher, dann $\bar{X}_n \to \mu$ mit Wslk 1

Simulation ad Stichprobenmittel:

möchte $\mathbb{E}(X)$ berechnen, erzeuge n unabhängige Zufallszahlen $X_1 \dots X_n$ mit selber Verteilung wie X

$\bar{X}_n$ als Näherungswert für $\mathbb{E}(X)$

Etwas allgemeiner $\frac{g(X_1)+\dots+g(X_n)}{n} \to \mathbb{E}(g(x))$

UE: wenn $X \sim f$ $\quad \mathbb{E}(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx \approx \frac{1}{n}\sum_{i=1}^{n} g(x_i)$

$X \sim U[0,1]$: $\int_0^1 g(x)dx \approx \frac{1}{n}\sum g(x)$

Problem: Genauigkeit?

Möchte $\frac{1}{n}\sum_{i=1}^{n} \underbrace{g(X_i)}_{Y_i}$ $\qquad \left| \frac{1}{n}\sum_{i=1}^{n} Y_i - \mathbb{E}(Y) \right| < \delta$ $\qquad \delta = \begin{matrix} 0.01 \\ 0.005 \end{matrix}$ für 2 Dezimalstellen

Fehlerschranke kann nur mit gewisser Wahrscheinlichkeit garantiert werden!

$\mathbb{P}\left(\left|\frac{1}{n}\sum Y - \mathbb{E}(Y)\right| \le \delta\right) \ge 1-\varepsilon$

$\delta \dots$ Fehlerschranke

$\varepsilon \dots$ (kleine) Wahrscheinlichkeit für einen großen Fehler

$\mathbb{P}(|\bar{Y}_n - \mathbb{E}(Y)| \le \delta) \le \varepsilon$ Wie kann das erreicht werden?

1) Chebyschev: $\mathbb{P}(|\bar{Y}_n - \mathbb{E}(Y)| \ge \delta) = \mathbb{P}(|\bar{Y}_n - \mathbb{E}(\bar{Y}_n)| \ge \delta) = \frac{\mathbb{V}(\bar{Y}_n)}{\delta^2} = \frac{\mathbb{V}(Y)}{n \cdot \delta^2}$

$n = \frac{\mathbb{V}(Y)}{\varepsilon\, \delta^2}$

$\delta = \sqrt{\frac{\mathbb{V}(Y)}{n\varepsilon}}$ $\qquad \begin{cases} O\left(\frac{1}{\sqrt{n}}\right) \\ \Theta\left(\frac{1}{\sqrt{n}}\right) \end{cases}$ Wie groß ist Fehler wenn ich gewisse # von Versuchen mache? ↳ nach δ umformen

$$\int_0^1\int_0^1 \dots \int_0^1 (x_1 \dots x_{20})\, dx_1 \dots dx_{20} = \mathbb{E}(f(U_1 \dots U_{20})) \approx \frac{1}{n}\sum^{n} (20_{i+1}, \dots, U_{20\,i+20})$$

ziemlich pessimistisch: Abhängig von Irrtumswahrscheinlichkeit ε

daher

Bessere Näherung

2) Zentraler Grenzwertsatz (ZGWS) $\qquad \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le x\right) \to \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$

$(X_1 \dots X_n)$ unabhängig, identisch verteilt

$\left.\begin{matrix} S_n = X_1 + \dots + X_n \\ \bar{X}_n = \frac{X_1+\dots+X_n}{n} \end{matrix}\right\}$ ist näherungsweise normal verteilt $\begin{cases} \mathcal{N}(n\mu, n\sigma^2) \\ \mathcal{N}(\mu, \frac{\sigma^2}{n}) \end{cases}$ $\qquad \begin{matrix} \mathbb{E}(X_n) = \mu \\ \mathbb{V}(X_n) = \sigma^2 \end{matrix}$

allgemein: standardisieren: $\mathbb{E}(X) = \mu \quad \mathbb{V}(X) = \sigma^2$

zentrieren: $\mathbb{E}(X) = 0 \quad$ i.e. $X - \mathbb{E}(X)$

normalisiert $\mathbb{E}(X) = 1 \quad \mathbb{V}(X) = 1$ $\qquad z = \frac{X-\mu}{\sigma} = \frac{X - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}}$

Grenzwertsatz - speziell

De Moivre - Laplace: $X \sim B(n,p)$

$\mathbb{P}(X=x) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(x-np)^2}{2np(1-p)}}$

wenn $|x - np| \le C \cdot \sqrt{n}$

$\frac{\text{linke Seite}}{\text{rechte Seite}} \to 1$

Grenzwertsatz (X_n) unabh., identisch verteilt

$\mathbb{E}(X_n) = \mu$
$\mathbb{V}(X_n) = \sigma^2$ $\quad (>0, <\infty)$

$S_n = (X_1 + \dots + X_n) \approx N(n\mu, n\sigma^2)$

BSP Simulation

$$I = \int_0^1 g(x)\,dx \rightarrow \int_0^1 (1-x^2)\,dx$$

Integral durch Näherung ersetzen: $I_n = \frac{1}{n} \sum_{i=1}^{n} g(u_i) = \frac{1}{n} \sum (1-u_i^2)$

Genauigkeit gegeben? → gebe WSK α (die sehr klein ist) und eine Genauigkeit δ (auch klein) vor, sodass die WSK einen Fehler $> \delta$ zu erhalten, soll $\geq \alpha$ sein. Wie groß muss n sein? # der Versuche

Chebychev: Sicher, aber zu pessimistisch): $n = \frac{\mathbb{V}(g(u)) \} \sigma^2}{\alpha \cdot \delta^2}$ $\quad \alpha = 0.01$, $\delta = 0.01$ $\quad \mathbb{V}(g(u))$ ist aber nicht so leicht zu berechnen. → Abschätzen, ZGWS

ZGWS: $I_n = \frac{1}{n} \sum g(u_i) \approx N(I, \frac{1}{n} \mathbb{V}(g(u))) = \mathbb{E}(g(u)) = \int_0^1 g(x)\,dx$

$$\mathbb{P}(|I_n - I| \leq \delta) = 1-\alpha \rightarrow \mathbb{P}(-\delta \leq I_n - I \leq \delta) = \mathbb{P}\left(\underbrace{-\frac{\delta}{\sqrt{\frac{\sigma^2}{n}}}}_{a} \leq \frac{I_n - I}{\sqrt{\frac{\sigma^2}{n}}} \leq \overbrace{\frac{\delta}{\sqrt{\frac{\sigma^2}{n}}}}^{b}\right) = \Phi(b) - \Phi(a) = 2\Phi(b) - 1$$

$\Phi(b) = 1 - \frac{\alpha}{2}$ $\quad b = \Phi^{-1}(1-\frac{\alpha}{2}) = z_{1-\frac{\alpha}{2}}$ → $1-\frac{\alpha}{2}$ Quantil der Standard Normal Verteilung

$$> n = \frac{z^2_{1-\frac{\alpha}{2}} \cdot \sigma}{\delta^2}$$

$\alpha = 0.01$ $\quad \frac{1}{\alpha} = 100$ $\quad z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.576$ $\quad z^2_{1-\frac{\alpha}{2}} \approx 6.64$

Spezialfall: de Moivre-Laplace: $X \sim B(n,p)$

$$\mathbb{P}(X=x) \approx \underbrace{\frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(x-np)^2}{2np(1-p)}}}_{\binom{n}{x} p^x (1-p)^{n-x}}$$

] Dichte einer Normalverteilung mit $\mu = np$ $\quad \sigma^2 = np(1-p)$
($x - np$ in Größenordnung $\sqrt{n}$)

S

Beweis: 1) Stirlingformel $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

2) $\frac{\mathbb{P}(X=x+1)}{\mathbb{P}(X=x)} = \ldots$ $\frac{\mathbb{P}(X=np+x+1)}{\mathbb{P}(X=np+x)}$ $\frac{\frac{n!}{(np+x+1)!\,(n(1-p)-x-1)!}}{\frac{n!}{(np+x)!\,(n(1-p)-x)!}} \cdot \frac{p^{np+x+1}(1-p)^{n(1-p)-x-1}}{p^{np+x}(1-p)^{n(1-p)-x}} = \frac{(n(1-p)-x)p}{(np+x+1)(1-p)} = \frac{p(x+1)}{p(x)} = \frac{1-\frac{x}{n(1-p)}}{1+\frac{x+1}{np}}$ $\quad g(x+1) = \frac{1-\frac{x}{n(1-p)}}{1+\frac{x+1}{np}} g(x)$ $\quad g(x+1) - g(x) = \frac{1-\frac{x}{n(1-p)} - 1 - \frac{x+1}{np}}{1+\frac{x+1}{np}} g(x) = \frac{-\frac{x}{np(1-p)} - \frac{1}{np}}{1+\frac{x+1}{np}} g(x)$ $\quad g(x) = \exp\left(C - \frac{x^2}{2np(1-p)}\right)$

$\frac{g'(x)}{g(x)} = \log(g'(x)) = -\frac{x}{np(1-p)}$ $\quad 1 = \sum_{x=-np}^{n(1-p)} g(x) = e^C \sum^{n(1-p)} e^{-\frac{x^2}{2np(1-p)}} \approx e^C \int e^{-\frac{x^2}{2np(1-p)}dx}$ $\quad$ schlampig $\quad g'(x) = -\frac{x}{np(1-p)} \cdot g(x)$

Beweis für Zentralen Grenzwertsatz: Momentenerzeugende für $N(0,1)$ $\quad M(t) = \mathbb{E}(e^{\overset{\sim N(0,1)}{x}t}) = e^{\frac{t^2}{2}}$

Charakteristische Funktion $\quad \varphi(t) = \mathbb{E}(e^{ixt}) = M(it) = e^{-\frac{t^2}{2}}$

$$\frac{S_n - n\mu}{\sqrt{n \cdot \sigma^2}} = \frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\,\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underbrace{\frac{x_i - \mu}{\sigma}}_{y_i}$$

$\mathbb{E}(y_i) = 0$
$\mathbb{V}(y_i) = 1$

$\mathbb{E}(e^{\frac{iS_n}{\sqrt{n}}t}) = \mathbb{E}(e^{\frac{it}{\sqrt{n}}(x_1 + \ldots + x_n)}) = \mathbb{E}(e^{\frac{it}{\sqrt{n}}x_1} \ldots e^{\frac{it}{\sqrt{n}}x_n}) \xrightarrow{\text{weil alle gleich, nur einen Ausrechnen}} \mathbb{E}(e^{\frac{itx_1}{\sqrt{n}}})^n = $
$= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n =$

$e^{\frac{it}{\sqrt{n}}X} = 1 + \frac{itX}{\sqrt{n}} - \frac{t^2X^2}{2} + o\left(\frac{1}{n}\right) \Rightarrow \mathbb{E}(e^{\frac{it}{\sqrt{n}}X}) = 1 + \frac{itx}{\sqrt{n}}\underbrace{\mathbb{E}(X)}_{\mu=0} - \frac{t^2}{2n}\underbrace{\mathbb{E}(x^2)}_{\mathbb{V}(x)=\sigma^2=1} + o\left(\frac{1}{n}\right) \approx 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$

BSP Würfel wird 2000 x geworfen

$$\mathbb{P}(X \leq 350) \approx \Phi\left(\frac{350 - \frac{2000}{6}}{\sqrt{2000 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = \Phi\left(\frac{2100 - 2000}{\sqrt{10\,000}}\right) = \Phi(1) = 0.841$$

$$\mathbb{P}(X < 351) \approx \Phi(1.06) = 0.855$$

Stetigkeitskorrektur: X diskret (ganzzahlig) durch Normalverteilung approximiert.

verwende $\mathbb{P}(X \leq x) = \mathbb{P}(X \leq x + 0.5)$

$\mathbb{P}(X < x) = \mathbb{P}(X \overset{\leq}{<} x - 0.5)$

gute Schätzung für Varianz: Beginne mit n (moderat groß, zw 100 und 1000)

$I_n = \frac{1}{n} \sum g(u_i)$ $\quad I_n \quad$ Schätzwert für Integral / Erwartungswert

$Q_n = \frac{1}{n} \sum g(u_i)^2$ $\quad Q_n - I_n^2 \quad$ Schätzwert für Varianz

genau: In jedem Schritt prüfen

WE DONE YET?

NO!

Statistik

Wir beschäftigen uns mit **mathematischer Statistik**

Gegensatz: Beschreibende Statistik (fasst große Datenmengen übersichtlich zusammen)
Histogramme, Torten, ...

$\boxed{\text{BSP}}$ Wahlumfrage — mauve Partei : Wie groß ist Wähleranteil?

einen von N Wahlberechtigten auswählen (zufällig)

$$X = \begin{cases} 1 & \text{ja, es wählt diese Partei} \\ 0 & \text{nein} \end{cases}$$

$\mathbb{P}(X = 1) = p$

$\mathbb{P}(X = 0) = 1 - p$

n Versuche $\quad \hat{p}_n = \bar{X}_4 = \frac{\#\text{ja-Antworten}}{n}$

unabhängig (mit zurücklegen) : mathematisch besser zu berechnen $\rightarrow$ Binomialverteilung: $\mathbb{V} = \frac{p(1-p)}{n}$ $\rightarrow$ pessimistisch

bzw ohne zurücklegen · praktisch besser (bei destruktiver Prüfung einziger Weg) $\rightarrow$ Hypergeometrische Verteilung: $\sigma^2 = \frac{p(1-p)}{n} \underbrace{\frac{N-n}{N-1}}_{<1}$ $\rightarrow$ genauer!

Statistisches Modell: Menge P von Verteilungen, welche das Problem beschreiben können

parametrisches Modell: Es gibt endlich viele reelle Zahlen, die die Verteilung festlegen (Parameter)

Aussagen einfacher, schärfer zu erreichen, aber ACHTUNG: komplett falsch, wenn Verteilung falsch

BSP Alternativ (Wahl), Normal (Größe, Gewicht), oder keine Einschränkungen

nicht parametrisches Modell: unendlich viele Dimensionen, Chancen für falsches Modell gering, Aussagen unscharf

Stichprobe: Folge $(X_1 \dots X_n)$ von unabhängig identisch verteilten Zufallsvariablen mit einer unbekannten Verteilung $p \in P$
(ziehen mit Zurücklegen)

$\downarrow$

Stichprobenumfang : n

Aufgabe der schließenden Statistik: Aus der Stichprobe Aussagen über die unbekannte Verteilung P erhalten.

Parametrisches Modell: kann durch endlich viele reelle Parameter festgelegt werden.

$(P_\vartheta, \vartheta \in \Theta)$ für $\Theta \subseteq \mathbb{R}^k$ mit $\Theta = (\vartheta_1, \dots \vartheta_k)$ wobei k ... reeller Parameter

Wahlumfrage / Alternativverteilung : $k = 1$ $\quad \Theta = [0, 1]$

Normalverteilung : $k = 2$ $\quad (\mu = \vartheta_1, \sigma^2 = \vartheta_2)$ $\quad \Theta = \mathbb{R} \times \mathbb{R}^k$

Gleichverteilung : $U([0, \vartheta])$: Θ : $\mathbb{R}^+$

Gleichverteilung : $U(\vartheta_1, \vartheta_2)$

Statistik $T = T(x_1 \dots x_n)$ Zufallsvariable, die Funktion (der Werte) in der Stichprobe ist.

GRUNDAUFGABEN der Statistik

1) Schätzproblem: einen Schätzwert für Parameter oder Funktion des Parameters aus der Stichprobe
2) Testproblem: gut/schlecht bzw ja/nein Entscheidungen aus Stichprobe — Bier $500\,cm^3$/Flasche — $497\,cm^3$ Mittel als Stichprobe
3) Prognoseproblem: untersuche Vergangenheit (Stichprobe) suche Parameter der Verteilung in der Zukunft

1) Schätzung

Schätzwert $\hat{\vartheta}$ für unbekannten Parameter ϑ

mit Index versehen, weil von Stelle die Verteilung abhängt!

Schätzer ist eine Folge $(\hat{\vartheta}_n)$ von Statistiken $\hat{\vartheta}_n = \vartheta_n(x_1, \dots, x_n)$

Eigenschaften von Schätzern:

konsistenz: $\hat{\vartheta}_n \to \vartheta_n \quad (n \to \infty)$ $\hat{\vartheta}_n$ ist konsistent für ϑ

- Schwache Konsistenz: Ein Schätzer heißt schwach konsistent, wenn $\hat{\vartheta}_n \to \vartheta_n$ Wahrscheinlichkeit $\mathbb{P}(|\hat{\vartheta}_n - \vartheta| > \varepsilon) \to 0$ für $n \to \infty$
- starke Konsistenz: Ein Schätzer heißt stark konsistent, wenn $\hat{\vartheta}_n \to \vartheta$ mit Wahrscheinlichkeit 1

Erwartungswert $\mathbb{E}_\vartheta(\hat{\vartheta}_n) = \vartheta$ (muss $\forall \vartheta$ gelten!)

- unverzerrt: Ein Schätzer heißt erwartungstreu, wenn $\mathbb{E}_\vartheta(\hat{\vartheta}_n) = \vartheta$

nach Steiner: Varianz d Fehlers + Quadrat der Verzerrung

- asymptotisch unverzerrt: $\lim_{n \to \infty} \hat{\vartheta}_n = \vartheta$
- Verzerrung von $\hat{\vartheta}_n$: $\mathbb{E}_\vartheta(\hat{\vartheta}_n) - \vartheta$

Varianz ϑ_n ist erwartungstreu, dann $\mathbb{V}_\vartheta(\hat{\vartheta}_n) = \mathbb{E}_\vartheta\left(\left(\hat{\vartheta}_n - \mathbb{E}_\vartheta(\hat{\vartheta}_n)\right)^2\right) = \mathbb{E}_\vartheta\left((\hat{\vartheta}_n - \vartheta)^2\right)$

MSE (mean square error), mittlerer quadratischer Fehler: $\mathbb{E}_\vartheta\left((\hat{\vartheta}_n - \vartheta)^2\right)$

Ein Schätzer $\hat{\vartheta}_n$ heißt effizient, wenn er erwartungstreu ist und die kleinste Varianz von allen erwartungstreuen Schätzern hat.

$\hat{\vartheta}_n$ ist effizient wenn
1) $\mathbb{E}_\vartheta(\hat{\vartheta}_n) = \vartheta \quad \forall \vartheta \in \Theta$
2) für jeden erwartungstreuen Schätzer $\tilde{\vartheta}_n$ gilt: $\mathbb{V}_\vartheta(\hat{\vartheta}_n) \leq \mathbb{V}(\tilde{\vartheta}_n) \quad \forall \vartheta$

Wenn $\hat{\vartheta}_n, \tilde{\vartheta}_n$ beide erwartungstreu sind und $\mathbb{V}(\hat{\vartheta}_n) \leq \mathbb{V}(\tilde{\vartheta}_n)$ dann heißt $\hat{\vartheta}_n$ effizienter als $\tilde{\vartheta}_n$

wenn man lineare Schätzer betrachtet $\hat{\vartheta}_n = \sum^n a_i x_i$

wenn man lineare Schätzer betrachtet: $\hat{\vartheta}_n = \sum a_i x_i \; (+\beta)$

BLUE: Best Linear Unbiased Estimator

Stichprobenvarianz $S_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$ (aka $\hat{\sigma}_n$ ...)

$\boxed{\text{BSP}}$ $\bar{X}_n$ ist erwartungstreuer, stark konsistenter Schätzer für $\mathbb{E}(X)$

$(X, X_1, \dots X_n)$ sind unabhängig, erwartungstreu

Frage nach der Effizienz nicht beantwortbar, da es immer vom Modell abhängig ist! Norm, Bin, exp, γ, Poisson → Stichprobenmittel ist effizient

$\bar{X}_n$ ist effizientester Schätzer

$\boxed{\text{BSP}_2}$ Varianz

$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ Schätzer für $\mathbb{V}(X)$: $\hat{\mathbb{V}}(X) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

Schätzen durch: $\frac{1}{n} \sum_{i=1}^{n} X_i^2$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$

stark konsistent: Ja, mit Wsk 1 $\subseteq$ schwach konsistent

erwartungstreu? $\mathbb{E}\left(\frac{1}{n}\left(\sum x_i^2 - \bar{x}_n^2\right)\right) = \frac{1}{n} \sum \mathbb{E}(x_i^2) - \mathbb{E}(\bar{x}_n^2) = \frac{1}{n} \sum \left(\mathbb{V}(X_i) + (\mathbb{E}(X_i))^2\right) - \left(\mathbb{V}(\bar{X}_n) + (\mathbb{E}(\bar{X}_n))^2\right) = \mathbb{V}(X) + (\mathbb{E}(X))^2 - \left(\frac{\mathbb{V}(X)}{n} + (\mathbb{E}(X))^2\right) = \frac{n-1}{n} \mathbb{V}(X)$

nein, um Faktor $\frac{n-1}{n}$ zu klein!

aus Schätzer erwartungstreue Version machen: $\hat{\hat{V}} := \frac{n}{n-1} \cdot \hat{V} = \frac{1}{n-1} \sum (x_i - \bar{x}_n)^2$ S_n oder S_n^2 → Stichprobenvarianz

↳ geht nur wenn n mind 2!

Schätzer konstruieren

Momentenmethode: $E_\vartheta(X) = g(\vartheta)$ eine Funktion von ϑ, wir setzen $g(\hat{\vartheta}_n) = \bar{X}_n$

wenn g umkehrbar, gilt $\hat{\vartheta}_n = g^{-1}(\bar{X}_n)$

wenn g stetig ist, ist $\hat{\vartheta}_n$ konsistent

$\boxed{\text{BSP}}$ Alternativverteilung $\mathbb{P}(X=1) = p$ $\mathbb{P}(X=0) = 1-p$ $\bar{X}_n$ ... erwartungstreuer Schätzer für $E_p(X) = 1p + 0(1-p) = p$

$\hat{p}_n = \bar{X}_n$ ... erwartungstreu, konsistent

$\boxed{\text{BSP}}$ Exponentialverteilung $f_\lambda(x) = \lambda e^{-\lambda x}$ $[x \geq 0]$ λ ... Parameter

$\underbrace{E_\lambda(X)}_{\bar{X}_n} = \frac{1}{\lambda}$ $\hat{\lambda} = \frac{1}{\bar{X}_n}$ ist nicht erwartungstreu, aber konsistent; für $n > 1$ kann man erwartungstreu machen

Gibt es mehr als einen Parameter, dann zusätzliche Gleichung aus $E(X^k) \approx \frac{1}{n} \sum x_i^k$ $k = 2, 3, \dots$

Bsp: $N(\mu, \sigma^2)$ $\bar{X}_n = E_{(\mu,\sigma^2)}(X) = \mu$ $\hat{\mu} = \bar{X}_n$

$\frac{1}{n} \sum X_i^2 = E_{(\mu,\sigma^2)}(X^2) = \hat{\mu}^2 + \hat{\sigma}^2$ $\hat{\sigma}^2 = \underbrace{\frac{1}{n} \sum x_i^2 - \bar{x}_n^2}_{\text{unkorrigiertes Stichprobenmittel}}$

Wodurch soll man bei der Varianz dividieren? n oder $n-1$

Grundgesamtheit : $\frac{1}{n}$

Stichprobe : $\frac{1}{n-1}$

Wenn aus Grundgesamtheit mit Umfang N ohne Zurücklegen n mal ziehen: erwartungstreuer Schätzer für σ^2: $\hat{\sigma}_2^2 = \frac{N-1}{N} \cdot \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$

Likelyhoodfunktion

Stichprobe ist aus diskreter Verteilung mit Punktwahrscheinlichkeit

$p_\vartheta(x_1) \dots p_\vartheta(x_n)$ WSLK, das genau diese Stichprobe $(x_1, \dots x_n)$ gezogen wird, wenn ϑ der Parameter

Bester Schätzwert ist der Wert von ϑ, welcher diese (also $p_\vartheta(x_1)\dots$) WSLK approximiert

$$L(x_1 \dots x_n, \vartheta) = \begin{cases} p_\vartheta(x_1) \dots p_\vartheta(x_n) & \text{wenn } f \text{ diskret} \quad \prod_{i=1}^{n} p(x_i, \vartheta) \\ f_\vartheta(x_1) \dots f_\vartheta(x_n) & \text{wenn } f \text{ stetig} \quad \prod_{i=1}^{n} f(x, \vartheta) \end{cases}$$

Notation: statt $p_\vartheta(x) = p(x, \vartheta)$

statt $f_\vartheta(x) = f(x, \vartheta)$

ML (Maximum Likelihood Schätzer) Wert von ϑ, der $L(x_1 \dots x_n, \vartheta)$ maximiert. $\max L(x_1 \dots x_n, \vartheta)$

$(\mu)^2$

BSP Normalverteilung $f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x-}{2\sigma^2}}$

$$L(x_1, \dots, x_n, \mu, \sigma^2) = \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}^{\,n}} e^{-\sum \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$\log L \to \max$

Maximum finden mittels ableiten, null setzen :)

$$\log L(x_1 \dots x_n, \mu, \sigma^2) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \sum \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \mu} := 0 \Leftrightarrow \sum \frac{2(x_i-\mu)}{2\sigma^2} = \sum (x_i - \mu) = 0$$

$$\frac{\partial}{\partial \sigma} := 0 = -\frac{n}{\sigma} + \sum \frac{2x_i - \mu}{2\sigma^2} \qquad \sum (x_i-\mu)^2 - n - \sigma^2 = 0$$

$$\sum x_i - n\mu = 0 \qquad \mu = \frac{\sum x_i}{n} = \bar{x}_n$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \mu)^2 = \frac{1}{n} \sum (x_i - \bar{x}_n)^2$$

BSP Alternativverteilung $A(\vartheta)$

$p(1, \vartheta) = \vartheta$

$p(0, \vartheta) = 1 - \vartheta$

$$p(x_i, \vartheta) = \begin{cases} \vartheta & x_i = 1 \\ 1-\vartheta & x_i = 0 \end{cases} \to \vartheta^{x_i} (1-\vartheta)^{1-x_i}$$

$$L(x_1 .. x_n, \vartheta) = \prod_{i=1}^{n} \vartheta^{x_i} (1-\vartheta)^{1-x_i} = \vartheta^{\sum x_i} (1-\vartheta)$$

$$\log L(x_1 \dots \vartheta) = \left(\sum x_i\right) \cdot \log \vartheta + \left(n - \sum x_i\right) \log(1-\vartheta)$$

Ableiten: $\frac{\partial}{\partial \vartheta} = 0$ $\quad 0 = \frac{\sum x_i}{\vartheta} - \frac{n \sum x_i}{1-\vartheta} = \left(\sum x_i\right) \left[\left(\frac{1}{\vartheta}\right) + \frac{1}{(1-\vartheta)} \right] - \frac{n}{1-\vartheta} =$

$$= \sum x_i \frac{1}{\vartheta(1-\vartheta)} = \frac{n}{1-\vartheta} \qquad \hat{\vartheta} = \frac{\sum x_i}{n}$$

BSP $U(0, \vartheta)$ $\quad f(x) = \begin{cases} \frac{1}{\vartheta} & 0 \le x \le \vartheta \\ 0 & \text{sonst} \end{cases}$

Momentenmethode: $\mathbb{E}(X) = \frac{\vartheta}{2}$ $\quad \vartheta = 2 \cdot E(X)$

Momentenschätzer: $2 \cdot \bar{X}_n = \hat{\vartheta}_n$

$$L(X_1, \dots X_n) = f(X_1, \vartheta) \dots f(X, \vartheta) = \begin{cases} \frac{1}{\vartheta^n} & \text{wenn } X_i \le \vartheta \quad \forall i = 1 \dots n \\ 0 & \text{sonst} \end{cases}$$

$$L(X_1 \dots X_n) = \begin{cases} \frac{1}{\vartheta^n} & \max(X_1 \dots X_n) \le \vartheta \\ 0 & \text{sonst} \end{cases}$$

UE: $E_\vartheta(\hat{\vartheta}_n) = \frac{n}{n+1} \vartheta$

$\tilde{\vartheta}_n = \frac{n+1}{n} \hat{\vartheta}_n^{ML}$

$\underbrace{\mathbb{V}(\tilde{\vartheta}_n)}_{\frac{1}{n(n+2)}} \le \underbrace{\mathbb{V}(\hat{\vartheta}_n^{(moment)})}_{\frac{1}{3n}}$

$\Rightarrow$ maximum Likelyhood Schätzer $\hat{\vartheta}_n^{(ML} = \max(X_1 \dots X_n)$ $\nearrow$ nicht erwartungstreu

Wie klein kann die Varianz eines erwartungstreuen Schätzers werden?

Cramér-Raó: Unter gegebenen Voraussetzungen ($p(x,\vartheta)$, $f(x,\vartheta)$ ist brav) ist $\hat{\vartheta}_n$ erwartungstreu für ϑ, wobei ϑ eindimensional ist, dann gilt:

$$\mathbb{E}\left(\left(\frac{\partial}{\partial\vartheta}\log L(x_1 \dots x_n, \vartheta)\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial^2\vartheta}\log L(x_1 \dots x_n, \vartheta)\right) = I_n(\vartheta) \quad \text{"Fisherinformation"}$$

Beweis: $\hat{\vartheta}_n = T(X_1 \dots X_n)$

$$\vartheta = \mathbb{E}_\vartheta(\hat{\vartheta}_n) = \sum_{x_1 \dots x_n} T(x_1 \dots x_n) \cdot \mathbb{P}(X_1 = x_1 \dots X_n = x_n) = \sum_{x_1 \dots x_n} T(x_1 \dots x_n) \cdot L(x_1 \dots x_n, \vartheta) \qquad \sum_{x_1 \dots x_n} L(x_1 \dots x_n; \vartheta) = 1$$

$$= \sum T(x_1 \dots x_n) \frac{\partial}{\partial\vartheta} L(x_1 \dots x_n, \vartheta) = 1$$

Cauchy Schwarz! $\mathbb{E}(X \cdot Y) \leq \sqrt{\mathbb{E}(X^2) \cdot \mathbb{E}(Y^2)}$

$\mathbb{E}(X^2) = a^2 > 0$ $\quad \left(\frac{X}{a} - \frac{Y}{b}\right)^2 \geq 0$

$\mathbb{E}(Y^2) = b^2 > 0$

$0 \leq \mathbb{E}\left(\left(\frac{X}{a} - \frac{Y}{b}\right)^2\right) = \mathbb{E}\left(\frac{X^2}{a^2} + \frac{Y^2}{b^2} - \frac{XY}{ab} \cdot 2\right) = 2 - 2\frac{\mathbb{E}(XY)}{ab}$

$$= \sum \frac{\partial}{\partial\vartheta} L(x_1 \dots x_n, \vartheta) = 0$$

$$1 = \sum \left[T(x_1 \dots x_n) - \vartheta\right] \frac{\frac{\partial}{\partial\vartheta} L(x_1 \dots x_n, \vartheta)}{L(x_1 \dots x_n, \vartheta)} L(x_1 \dots x_n, \vartheta)$$

$$1 = \mathbb{E}_\vartheta\left(T(x_1 \dots x_n, \vartheta) - \vartheta\right) \frac{\partial}{\partial\vartheta} \log L(x_1 \dots x_n \vartheta) \leq \sqrt{\underbrace{\mathbb{E}_\vartheta\left((\hat{\vartheta}_n - \vartheta)^2\right)}_{\mathbb{V}(\hat{\vartheta}_n)} \cdot \underbrace{\mathbb{E}\left(\left(\frac{\partial}{\partial\vartheta} \log L(x_1 \dots x_n, \vartheta)\right)\right)}_{\text{Fisherinformation } I_n(\vartheta)}}$$

$\boxed{\text{BSP}}$ Normalverteilung $N(\mu, \sigma^2)$ $\quad \sigma^2$ bekannt $\rightarrow$ nur 1 Parameter μ

$$\sigma^2 = 1 \implies L(x_1 \dots x_n, \mu) = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{\sum(x_i - \mu)^2}{2}}$$

$$\log L(x_1 \dots x_n, \mu) = -\frac{n}{2}\log(2\pi) - \frac{\sum(x_i - \mu)^2}{2}$$

$$\left.\begin{aligned} \frac{\partial}{\partial\mu} \log L &= \sum x_i - \mu = \sum x_i - n\mu \\ \frac{\partial^2}{\partial^2\mu} \log L &= -n \end{aligned}\right] \quad I_n(\mu) = -\mathbb{E}_\mu(-n) = n$$

Cramer Rao Schranke: $\hat{\mu}_n$ erwartungstreu $\rightarrow \mathbb{V}_\mu(\hat{\mu}_n) \geq \frac{1}{n}$

$\hookrightarrow \hat{\mu}_n = \bar{x}_n \quad \mathbb{V}_\mu(\hat{\mu}_n) = \frac{1}{n} \mathbb{V}_\mu(x) = \frac{1}{n}$ $\rightarrow$ kleinste Varianz die es gibt

$\bar{X}_n$ ist effizient

$\boxed{\text{BSP}}$ Alternativverteilung

$$\log L(x_1 \dots x_n, \vartheta) = \sum x_i \log\vartheta + \left(n - \sum x_i\right)\log(1 - \vartheta)$$

$$\frac{\partial}{\partial\vartheta} \log L(x_1, \dots x_n, \vartheta) = \frac{\sum x_i}{\vartheta} - \frac{n - \sum x_i}{1 - \vartheta}$$

$$\frac{\partial^2}{\partial^2\vartheta} \log L(\dots) = -\frac{\sum x_i}{\vartheta^2} - \frac{n - \sum x_i}{(1 - \vartheta)^2}$$

$$\mathbb{E}\left(\frac{\partial^2}{\partial^2\vartheta} \log L\right) = -\frac{n\vartheta}{\vartheta^2} - \frac{n(1-\vartheta)}{(1-\vartheta)^2} = -\frac{n(\vartheta + 1\vartheta)}{\vartheta(1\vartheta)} = -\frac{n}{\vartheta(1 - \vartheta)} =$$

$\hat{E}_n \dots$ erwartungstreu $\rightarrow \mathbb{V}(\hat{\vartheta}_n) \geq \frac{\vartheta(1-\vartheta)}{n} \quad \mathbb{V}(\bar{X}_n)$

Konfidenzintervall

dua Schätzen mit Genauigkeitsangabe

ein Paar $[a(x_1 \dots x_n) \le b(x_1 \dots x_n)]$ von Statistiken mit $P_\vartheta(a(x_1 \dots x_n) \le \vartheta \le b(x_1 \dots x_n)) \ge \gamma$

Konfidenzintervall mit Überdeckungswahrscheinlichkeit γ $\quad 0 < \gamma < 1$

$\boxed{\text{BSP}}$ $N(\mu, \sigma^2)$, $\sigma^2 = 1$

Konfidenzintervall Ausgangspunkt: Schätzer $\hat{\mu}_n = \bar{x}_n$

Ansatz: Konfidenzintervall symmetrisch um $\bar{x}_n$ $\Rightarrow$ $a = \bar{x}_n - c$, $b = \bar{x}_n + c$

$$P_\mu(\bar{x}_n - c \le \mu \le \bar{x}_n + c) = \gamma = P_\mu(\mu \le \bar{x}_n \le \mu + c) = P_\mu(-c \le \bar{x}_n - \mu \le +c) =$$

$$= P_\mu\left(-\frac{c}{\sqrt{\frac{\sigma^2}{n}}} \le \underbrace{\frac{\bar{x}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}}_{N(0,1)} \le \frac{c}{\sqrt{\frac{\sigma^2}{n}}}\right) = \Phi\left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}}\right) - \Phi\left(-\frac{c}{\sqrt{\frac{\sigma^2}{n}}}\right) =$$

$$= 2\Phi\left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}}\right) - 1$$

$$2\Phi\left(\frac{c}{\sqrt{\frac{\sigma^2}{n}}}\right) = 1 + \gamma \rightarrow \frac{c}{\sqrt{\frac{\sigma^2}{n}}} = \Phi^{-1}\left(\frac{1+\gamma}{2}\right) = z_{\frac{1+\gamma}{2}}$$

$$c = z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{n}} \rightarrow \text{Konfidenzintervall } \left[\bar{x}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\sigma^2}{2}}, \bar{x}_n + z_{\frac{1+\gamma}{n}}\sqrt{\frac{\sigma^2}{2}}\right]$$

Wenn σ^2 unbekannt $\rightarrow$ Schätzen durch Stichproben $\quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$

Möglichkeiten

1) billig: n hinreichend groß, dann $S_n^2 \approx \sigma^2$ $\rightarrow$ einfach einsetzen $\Rightarrow$

↳ näherungsweises Konfidenzintervall $\left[\bar{x}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{n}}, \bar{x}_n + z_{\frac{1+\gamma}{2}}\sqrt{\frac{S_n^2}{2}}\right]$

$\boxed{\text{BSP}}$ $n = 20$ $\quad \bar{x}_n = 1{,}32$ $\quad S_n^2 = 0.36$ $\quad \gamma = 0.95$

1) näherungsweises Konfidenzintervall

$$\left[1{,}32 - \underbrace{1.96\sqrt{\frac{0.36}{20}}}_{0.263}, 1{,}32 + 1.96\sqrt{\frac{0.36}{20}}\right]$$

$[1.05, 1.59]$

2) exakter Weg $\quad$ Satz: $x_1 \dots x_n$ unabhängig und $\sim N(\mu, \sigma^2)$ $\rightarrow$ $\bar{x}_n$ und S_n^2 unabhängig

$$\bar{x}_n \sim N\left(\mu, \sqrt{\frac{\sigma^2}{n}}\right) \qquad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1} \; \left(= \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)\right)$$

$$\frac{\bar{x}_n - \mu}{\sqrt{\frac{S_n^2}{n}}} \sim t_{n-1} \qquad \text{Dichte} \quad f(x) = C_n \cdot \frac{1}{\left(1 + \frac{x^2}{n-1}\right)^{\frac{n}{2}}}$$

Ansatz für Konfidenzintervall $\left[\bar{x}_n - c\sqrt{\frac{S_n^2}{n}}, \bar{x}_n + c\sqrt{\frac{S_n^2}{n}}\right]$ $\quad c = t_{n-1, \frac{1+\gamma}{2}}$

2) exakt: statt $z_{0.975} = 1.96$ $\quad t_{n-1, \frac{1+\gamma}{2}}$

$t_{20-1,}$ ⇣

$t_{19, 0.975} = 2.093$

$$\left[1.32 - \underbrace{2.093\sqrt{\frac{0.36}{20}}}_{0.28}, 1.32 + 2.093\sqrt{\frac{0.36}{20}}\right] = [1.04, 1.6]$$

$$\left[\frac{19 \cdot 0.36}{32{,}852}, \frac{6.84}{8.907}\right] = [0.2, \dots 0.7]$$

Konfidenzintervall σ^2 $\rightarrow$ basiert auf S_n^2

im Raum nicht symmetrisch

↓

definiert als $[a \cdot S_n^2, b \cdot S_n^2]$

$$P(\sigma^2 < a \cdot S_n^2) = P(\sigma^2 > b \cdot S_n^2) = \frac{1-\gamma}{2}$$

$$P\left(\frac{n-1}{a} < \underbrace{\frac{S_n^2}{\sigma^2}(n-1)}_{\chi^2_{n-1}}\right) = \frac{1-\gamma}{2} \rightarrow P\left(\frac{n-1}{a} \ge \dots\right) = 1 - \frac{1-\gamma}{2} \qquad \frac{n-1}{a} = \chi^2_{n-1, \frac{1+\gamma}{2}}$$

$$P\left(\frac{n-1}{b} \ge \dots\right) = \frac{1-\gamma}{2} \qquad \frac{n-1}{b} = \chi^2_{n-1, \frac{1-\gamma}{2}}$$

$$\text{KI für } \sigma^2 \left[\frac{(n-1)S_n^2}{\chi^2_{n-1, \frac{1+\gamma}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1, \frac{1-\gamma}{2}}}\right]$$

Anteilswert/allgemein: $\hat{\vartheta} = \bar{x}_n$ $\qquad n \cdot \bar{x}_n \sim B(n, \vartheta) \approx N(n\vartheta, n\vartheta(1-\vartheta))$

$$\bar{x}_n \approx N\left(\vartheta, \frac{\vartheta(1-\vartheta)}{n}\right)$$

$\hat{\vartheta} = \bar{x}_n$ für ϑ einsetzen

Mittelpunkt des Konfidenzintervalls

näherungsweises Konfidenzintervall: $\left[\bar{x}_n - z_{\frac{1+\gamma}{2}}\sqrt{\frac{\vartheta(1-\vartheta)}{n}}, \bar{x}_n + z_{\frac{1+\gamma}{2}}\sqrt{\frac{\vartheta(1-\vartheta)}{n}}\right]$

Statistische Tests (Signifikanz - Hypothesetests)

entscheidung zwischen zwei Möglichkeiten (gut oder schlecht)

BSP Bier, 500ml Inhalt → haben wir genug? oder Ist die Münze fair? ($p = \frac{1}{2}$) oder Ist der Würfel fair? (alle WSK der Augenzahlen $p = \frac{1}{6}$)
↳ $\mu = 500$ Inhalt: $N(\mu, \sigma^2)$

Hypothese: Teilmenge des Parameterraums, mögliche Entscheidungen

- Nullhypothesen H_0 (wie es sein soll)
- Alternative / Gegenhypothesen (Abweichungen von der Norm, was bewiesen werden soll)
- einfache Hypothese $|H| = 1$ (nur 1 Element)
- Zusammengesetzte Hypothese $|H| > 1$

BSP. $p = \frac{1}{2}$ $(\{\frac{1}{2}\})$ einfach $p < \frac{1}{2}$, $p \leq \frac{1}{2}$, $p \neq \frac{1}{2}$ zusammengesetzt
als Menge

$\mu = 500$ ist zusammengesetzt, $H = \{(\mu, \sigma^2) : \mu = 500, \sigma^2 > 0\}$ wenn σ^2 unbekannt

Bei zusammengesetzten Hypothesen:

einseitig. $p < \frac{1}{2}$, $p \leq \frac{1}{2}$, $p > \frac{1}{2}$ $\mu > 500$, $\mu < 500$

zweiseitig: $p \neq \frac{1}{2}$ $\mu \neq 500$

Vorgehen: H_0, H_1 formulieren (Wann entscheiden wir uns für H_0, wann für H_1?)

Annahmebereich: alle möglichen Stichprobenergebnisse für die H_0 angenommen wird

Verwerfungsbereich: Stichproben, für die H_1 genommen wird.

$$\left.\begin{matrix} H_0 = \frac{1}{2} \\ H_1 \neq \frac{1}{2} \end{matrix}\right] n = 4 \qquad A = \begin{Bmatrix} 0011 & 0101 \\ 0110 & 1010 \\ 1100 & 1001 \end{Bmatrix} \qquad V = \begin{Bmatrix} 0000 & 1000 & 1110 \\ 1111 & \vdots & \vdots \\ & 0001 & 0111 \end{Bmatrix}$$

Schöner:

a) **Teststatistik** (Statistik, die vom Parameter beeinflusst wird.)

BSP Münze: Anzahl / Erfolge / Köpfe

Vorzeichentest: Anzahl der Beobachtungen $> \mu$

Normalverteilung: $\bar{X}_n$ Stichprobenmittel (generell: Schätzer für Parameter)

b) **Asymetrische Sprechweise:** H_0 wird angenommen oder verworfen → stärker

Teststatistik T, H_0 wird verworfen, wenn T größer / kleiner als "kritische Wert" t_c

BSP Münze $T = \#$ Köpfe, verworfen von H_0 wenn $T \neq 2$ $|T-2| > 0$

Fehler 1. Art: Nullhypothese H_0 trifft zu, wird aber verworfen.

Fehler 2. Art: Nullhypothese trifft nicht zu, wird aber angenommen.

↳ Dilemma: kann WSLK für Fehler 1. Art nicht verkleinern, ohne Wahrscheinlichkeit für Fehler 2. Art zu vergrößern (und umgekehrt)

Lösung: Eine Schranke für die WSLK des Fehlers 1. Art an $\mathbb{P}(\text{Fehler 1. Art}) \leq \alpha$

Signifikanzniveau α, wobei $\alpha = 0.05$ (wenn nicht anders gesagt)

Ablauf:
1) Hypothesen formulieren
2) Wahl der Teststatistik (Kochrezept), α festlegen
3) Verteilung der Teststatistik T unter H_0 → kritischer Wert
4) Stichprobe, Teststatistik berechnen, mit kritischem Wert vergleichen → Entscheidung

Am Computer nach Niveau wird nicht gefragt, es wird "p-Wert" gefragt, dh: WSLK, dass ein Ergebnis auftritt, welches mindestens so schlecht (für die Nullhypothese H_0) ist wie das aus der Stichprobe

BSP $H_0 = \frac{1}{2}$ $H_1 > \frac{1}{2}$

Teststatistik: $T =$ Anzahl "Kopf", $\alpha = 0.05$, $n = 12$ (Stichprobengröße)

verwerfe H_0, wenn $T > t_c$

binomial → (theoretisch gut)

$$\mathbb{P}(T > t_c) = \sum_{k=t_c+1}^{12} \binom{12}{k} \frac{1}{12} \leq \alpha$$

$\sum \binom{12}{k} \leq 2^{12} \cdot 0.05 = 204{,}8$

$\binom{12}{12} = 1$

$\binom{12}{11} = 12$

$\binom{12}{10} = \binom{12}{2} = 66$

$\binom{12}{9} = 220$

$= t_c =$

irgend was mit 125 und 7,e % ... ?

praktisch: Approximation mittels Normalverteilung

H_0 wird verworfen, $T = \#\text{Kopf} = \sum^{n} X_i$

$t_c = np + z_{1-\alpha}\sqrt{np(1-p)}$

Kochrezept für Anteilswerte $\alpha = 0.05$ wobei $T = \dfrac{\text{Anzahl Erfolge} - np_0}{\sqrt{n \cdot p_0(1-p_0)}}$

(1) $H_0: p = p_0$ / $H_0: p > p_0$ ($H_0: p \leq p_0$) } H_0 verworfen wenn $T > z_{1-\alpha}$

(2) $H_0: p = p_0$ ($p \leq p_0$) / $H_0: p < p_0$ } H_0 verworfen wenn $T < z_\alpha = -z_{1-\alpha}$

(3) $H_0: p = p_0$ / $H_0: p \neq p_0$ } H_0 verworfen wenn $|T| > z_{1-\frac{\alpha}{2}}$

BSP $N(\mu, \sigma^2)$, $H_0: \mu = \mu_0$ / $H_1: \mu > \mu_0$, Annahme: σ^2 bekannt (ist einfacher owo)

$T =$ Schätzer für $\mu = \overline{X}_n$, $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ unter H_0

$T = \dfrac{\overline{X}_n - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$

H_0 $\mu = \mu_0$ ($\mu \leq \mu_0$) / H_1 $\mu > \mu_0$] verwerfen, wenn $T > z_{1-\alpha}$

H_0 $\mu \geq \mu_0$ / H_1 $\mu < \mu_0$] verwerfen, wenn $T < z_\alpha = -z_{1-\alpha}$

H_0 $\mu = \mu_0$ / H_1 $\mu = \mu_0$] verwerfen, wenn $|T| > z_{1-\frac{\alpha}{2}}$

Wenn σ^2 (Varianz) nicht bekannt ist, muss es mit S_n^2 geschätzt werden

$$T = \frac{\overline{X}_n - \mu_0}{\sqrt{\frac{S_n^2}{n}}}$$

Computer: H_0 wird verworfen, wenn p-Wert $< \alpha$

Verteilung von T mit t_{n-1} ($n-1$ Freiheitsgrade - exakt) oder $N(0,1)$ näherungsweise

H_0 $\mu \leq \mu_0$ / H_1 $\mu > \mu_0$ verworfen wenn $T > t_{n-1, 1-\alpha}$

H_0 $\mu \geq \mu_0$ / H_1 $\mu < \mu_0$ verworfen wenn $T < t_{n-1, \alpha} = -t_{n-1, 1-\alpha}$

H_0 $\mu = \mu_0$ / H_1 $\mu \neq \mu_0$ verworfen wenn $|T| > t_{n-1, 1-\frac{\alpha}{2}}$

BSP Bier $\mu_0 = 500$ $n = 25$ $\overline{x}_n = 497$ $S_n^2 = 25$ $t_{n-1;\alpha} = -t_{n-1;1-\alpha} = -t_{24;0,95} = -1,711 < -2,797$

$T = \frac{497 - 500}{\sqrt{\frac{25}{25}}} = -3$ $\longrightarrow$ H_0 wird verworfen

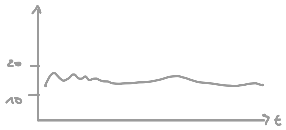
Stochastik

Stochastischer Prozess: Familie von Zufallsvariablen mit X_t, $t \in T$ $\quad$ $t \in \mathbb{R}$ (Parameterraum, Indexraum)

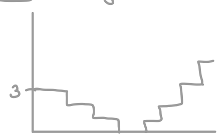
diskreter Stochastischer Prozess: T ist endlich / abzählbar $\quad$ Ω_X – Wertebereich (Zustands) Phasenraum)

stetiger Stochastischer Prozess: T ist endliches / unendliches Intervall

BSP Temperatur – stetig per Definition



BSP Fahrzeuge an der Kreuzung

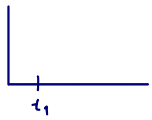


diskreter Zustandsraum

Sei X_n, $n \geq 1$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen

$S_n = X_1 + \dots + X_n$ $\quad (S_0 = 0)$

unabhängig identisch verteilt aka idependent and identically distributed iid



X_{t_1} $\quad t_1 \geq 0$

(X_{t_1}, X_{t_2}), $\; t_1, t_2 \geq 0$, $\; t_2 > t_1$

Charakterisierung (eines stochastischen Prozesses): Die Zufallsvariablen X_n $(n \geq 1)$ können voneinander abhängig sein. Beschrieben wird das durch die gemeinsamen Verteilungen $(X_{t_1}, X_{t_2}, \dots X_{t_n})$ wobei $t_1 < t_2 < \dots < t_n$ $\; n \geq 1$

Kolmogow: Ein stochastischer Prozess wird durch gemeinsame Verteilungen $(X_{t_1}, X_{t_2}, \dots X_{t_n})$ mit $t_1 < t_2 < \dots < t_n$ $(n \geq 1)$ beschrieben

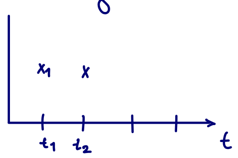
Wichtige Klassen der stochastischen Prozesse: stationäre Prozesse, prozesse mit unabhängigen Zuwächsen, Markovprozesse, Zählprozesse, Erneuerungsprozess

räumliche Stochastische Prozesse: gibts auch

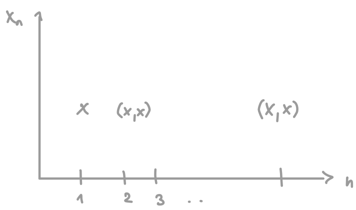
stationärer Stochastischer Prozess

stationär: $(X_t, t \in T)$ ist stationär, wenn die gemeinsamen Verteilungen $(X_{t_1}, \dots X_{t_n})$ mit denen von $(X_{t_1+h}, \dots X_{t_n+h})$ für alle $t_1 < \dots < t_n$ und $h > 0$ übereinstimmen. (Verteilungen sind so gleich, dass h beliebig sein kann)

Veranschaulichung



BSP



Folge von Zufallsvariablen, wobei X_n unabhängig und gleichverteilt sind.

$\acute{\varepsilon}\rho\gamma o\nu$ – $\acute{o}\delta o\varsigma$

Ergodensatz von Birkhoff

Wenn eine Folge von Zufallsvariablen $(X_n, n \geq 1)$ stationär ist und $E[X_1] < \infty$ (muss endlich sein) dann gilt

1) die Zufallsvariable $X_\infty =: \lim\limits_{x \to \infty} \overline{X}_n$ (Folge) existiert mit Wahrscheinlichkeit 1

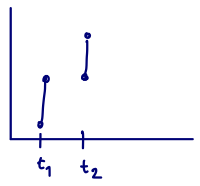
2) $E[X_\infty] = E[X_1]$ wobei $\overline{X}_n = \dfrac{\sum X_i}{n}$

$i=1 \rightarrow X_1$
$i=2 \rightarrow \dfrac{X_1 + X_2}{2}$
$i=3 \rightarrow \dfrac{X_1 + X_2 + X_3}{3}$

Wenn $\lim\limits_{x \to \infty} f(X_n, X_{n+1}, \dots)$ ein Wert für die beschränkte Funktion $f()$ ist, dann heißt $(X_n, n \geq 1)$ ergodisch

Prozesse mit unabhängigen Zuwächsen

unabhängiger Zuwachs: $(X_t, t \in T)$ unabh. Zuw. wenn $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ für $t_1 < \dots < t_n$
voneinander unabhängig



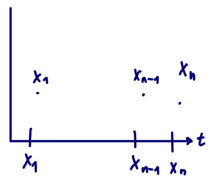
$\boxed{\text{BSP}}$ Sei $(X_n, n \geq 1)$ unabh, identisch Verteilt, dann ist $S_n = X_1 \dots X_n$ ein Prozess mit unabhängigen Zuwächsen

$S_n, \; S_m - S_n, \; S_i - S_j, \; m \leq j < i$

$S_n = X_1 + \dots + X_n$ $\qquad S_m - S_n = X_{n+1} + X_{n+2} + \dots + X_m$ $\qquad S_i - S_j = X_{j+1} + \dots + X_i$

Markovprozesse ~~Malakovprozess~~

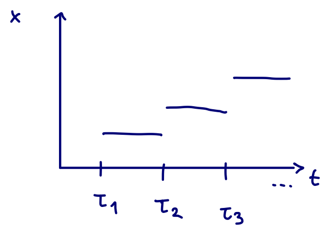
Folgende Eigenschaft muss gelten: $P(X_{t_n} = x_n \mid X_{t_1} = x_1, \dots X_{t_{n-1}} = x_{n-1}) = P(X_{t_n} = x_n \mid X_{t_{n-1}} = x_{n-1})$ $\forall \; t_1 < \dots < t_n, \; n \geq 1$



Markoveigenschaft:

Zählprozess

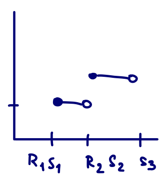
es gilt: $X_t = \# \{ i : 0 < T_i \leq t \}$ mit $(T_i > 0, \; i \geq 1)$ ist Zufallsfolge von Zeitpunkten/Ereignissen



Erneuerungsprozess

es gilt: $(R_n, n \geq 1)$ ist Folge von unabhängig und identisch verteilten Zufallsvariablen und $S_0 = 0$ und $S_n = R_1 + \dots + R_n$

$(t_n, t \in T)$ ist Erneuerungsprozess $\iff$ $X_t = n$ für $S_n \leq t < S_{n+1}$



$X_t = \# \{ n : 0 < S_n \leq t \}$

Erneuerungsprozess ist Spezialfall von Zählprozess

Poissonprozess

Ein Poissonprozess wird mit Hilfe von Annäherung von diskreter Zeit auf stetige Zeit durch schrittweise Erhöhung der Auflösung erzeugt.

Wir teilen $[0,\infty]$ auf kleine Teile mit der Länge 2^{-N} auf. Einige von diesen kleinen Teilen werden markiert bzw. gewisse Ereignisse treten in einigen, zufällig ausgewählten kleinen Intervallen auf. Die Markierungen sind voneinander unabhängig. Die Wahrscheinlichkeit des Auftretens eines gewissen Ereignisses ist proportional mit der Länge des kleinen Intervalls, dh $P(\text{Markierung}) = \lambda 2^{-N}$. N ist groß genug gewählt, sodass $\lambda 2^{-N} < 1$ ist.

→ Sei $X(t)$ die Anzahl der Markierungen im Intervall $[0,t]$

1) $[X(t+s) - X(s)]$ und $X(s)$ sind unabhängig

2) $[X(t+s) - X(s)]$ hat Binomialverteilung mit $p = \lambda 2^{-N}$ und $n = 2^N t$ $\sim B(\lambda 2^{-N}, 2^N t)$

$\Rightarrow E[X(t+s) - X(s)] = np = \lambda 2^{-N} 2^N t = \lambda t$ → unabhängig von N

Wir setzen den Grenzübergang $N \to \infty$ mit

1) $[X(t+s) - X(s)]$ und $X(s)$ sind unabhängig $\qquad X \sim B(p,n) \xrightarrow[np = \text{konstant}]{n\to\infty} P(\text{konstant})$

2) $[X(t+s) - X(s)]$ hat Poissonverteilung $\sim P(\lambda, t)$

$\Rightarrow P([X(t+s) - X(s)] = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad p \geq 0$

Man nennt den Zählprozess $(N(t), n \geq 1)$ Poissonprozess mit Parameter λ, wenn gilt:

1) Prozess mit unabhängigen Zuwächsen, dh $[N(s+t) - N(s)]$ und $N(t)$ für $0 \leq n \leq s$ sind unabhängig

a 2) Prozess mit stationären Zuwächsen, dh $[N(s+t) - N(s)]$ ist $P(\lambda t)$ also $P(N(s+t) - N(s) = k) = \frac{\lambda t}{k!} e^{\lambda t}$ also nur von t abhängig für alle $s \geq 0$

b •) 1. Alternative Form: Kurzzeitverhalten

$$P(N(t+h) - N(t) = 1) = \lambda h + o(h)$$

$$P(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h)$$

$$P(N(t+h) - N(t) \geq 2) = o(h)$$

c •) 2. Alternative Form: Exponentiell verteiltes Zwischenzuwachsintervall

Sei $(T_i > 0, i \geq 1)$ durch $N(t) = \#\{i : 0 < T_i \leq t\}$ definiert und $T_0 = 0$

Dann sind die Intervalle $T_{n+1} - T_n$, $n \geq 0$ mit $\exp(\lambda)$ verteilt.

Kochrezept:

Zuerst a dann b dann c beweisen, von c auf a folgern

ex: Kleinordnung:

Fkt $f(h)$ ist $o(h)$ wenn $\lim_{h \to 0} \frac{f(h)}{h} = 0$
geht gegen Null schneller als h
Alle Fkt die $> h^1$ sind, sind Kleinordnungen.

$2h \notin o(h)$

$h^2 \in o(h)$

Kochrezept : Zuerst a dann b dann c beweisen, von c auf a folgern

$a \to b$ (stationärer Prozess $\to$ Kurzzeitverhalten)

$$P(N(t+h)-N(t)=0) = e^{-\lambda h} \overset{h \ll 1}{=} 1-\lambda h + o(h)$$

$$P(N(t+h)-N(t)=1) = \lambda h e^{-\lambda h} = \lambda h - \underbrace{\lambda^2 \lambda^2 o(h)}_{o(h)} = \lambda h + o(h)$$

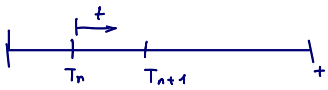
$$P(N(t+h)-N(t)\geq 2) = 1-P(N(t+h)-N(t)=0)-P(N(t+h)-N(t)=1) = 1-(1-\lambda h+o(h)) - \lambda h - o(h) = o(h)$$

$b \to c$ (Kurzzeitverhalten $\to$ exp vert. Zw. Zuwachs)

$$P(t) = P(\text{kein Zuwachs findet statt in } [0,t])$$

$$P(t+\Delta t) \underset{\text{1. Eigenschaft}}{=} \overset{\text{kein Zuwachs}}{P(t)} \underset{\text{b. Eigenschaft}}{=} P(\Delta t) = P(t)(1-\lambda\Delta t + o(\lambda\Delta t)) = P(t) - \lambda P(t)\Delta t + P(t)\, o(\Delta t)$$

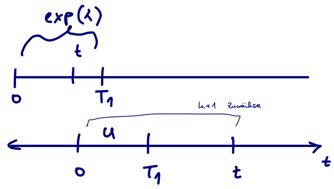
$$\lim_{\Delta t \to \infty} \frac{P(t+\Delta t)-P(t)}{\Delta t} = -\lambda P(t) + \lim_{\Delta t\to\infty} P(t)\frac{o(\Delta t)}{\Delta t}$$



$$P'(t) = -\lambda P(t) \qquad P(t) = 0$$

$$P(t) = e^{-\lambda t} \qquad P(t) = P(T_{n+1}-T_n > t) = e^{-\lambda t} \to P(T_{n+1}-T_n \leq t) = 1-e^{-\lambda t} \qquad n \geq 0$$

$c \to a$ (exp .... $\to$ stationäre Zuwächse)

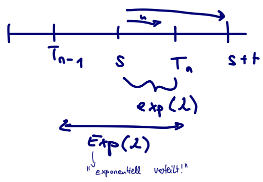


$$n = \emptyset \quad P(T_n > t) = e^{-\lambda t} = P(N(t)=0)$$

$n \geq 1$ mit mathematischem Indikator (Induktion?)

$$P(N(t)=k+1) = \int_{u=0}^{t} P(N(t)=p+1 \mid T_1 = u)\cdot f(u)\,du = \int_{u=0}^{t} P(N(t+u)=k) f_u\,du = \int_{u=0}^{t} \frac{[\lambda(t-u)]^2}{k!} e^{-\lambda(t-k)} \lambda e^{-\lambda u} du =$$

$$= \frac{\lambda^{k+1}}{k!} e^{-\lambda t} \int_{u=0}^{t} (t-u)^2 du = \underbrace{\frac{\lambda^{k+1}}{k!} e^{-\lambda t}}_{\text{konstant}} \left. \frac{(t-u)^{k+1}}{k+1} \right|_{u=0}^{t} = \underbrace{\frac{\lambda^{k+1}}{k!} e^{-\lambda t}}_{\text{konstant}} \frac{t^{k+1}}{k+1} = \frac{(\lambda t)^{k+1}}{(k+1)!} e^{-\lambda t}$$

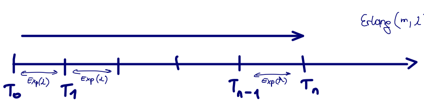


ist ja schön, aber wir können es noch anders beweisen, und zwar so:

$$P(T_n \leq t) = P(T_n \leq t, T_{n+1} \leq t) + P(T_n \leq t, T_{n+1} > t)$$

Erlang(n,λ) = $\Gamma(n, \lambda)$

$$\Rightarrow P(T_n \leq t, T_{n+1} > t) = P(T_n \leq t) - P(T_{n+1} \leq t)$$



$$P(N(t)=n) = P(T_n \leq t, T_{n+1} > t) = P(T_n \leq t) - P(T_{n+1} \leq t) = \int_{x=0}^{t} \underbrace{\frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda k} - \frac{\lambda^{n+1} x^n}{n!} e^{-\lambda x}}_{\frac{1}{\lambda} f'_{\text{Erlang}(n+1,\lambda)}(x)} dx = \frac{1}{\lambda} \frac{\lambda^{n+1} t^n}{n!} e^{-\lambda t} =$$

$$= \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$\Gamma_{(n,\lambda)} = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda t}$$

Verbundene Eigenschaften der Exp(λ) Verteilung

1) Erlang (n, λ) Verteilung:

Sei $(x_n, n \geq 1)$ eine Folge von unabhängig verteilten Exp(λ) Zufallsvariablen, dann ist $y_n = x_1 + \dots + x_n$ Erlang(n, λ) verteilt.

$$\text{Erlang}\ (n,\lambda) = \Gamma(n,\lambda)$$

$$n = 1: f_{\text{erlang}\ (1,\lambda)}(x) = \lambda e^{-\lambda x} = f_{\Gamma(1,\lambda)}$$

$$n > 1 \quad f_{\text{erlang}\ (n,\lambda)}(x) = \int_{x=0}^{t} f_{\text{erlang}\ (n,\lambda)}(x)\, f_{\text{Exp}(}(z-x)\, dx = \int \frac{\lambda^n x^{n+1}}{(n-1)!} \lambda e^{-\lambda(z-x)} = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \underbrace{\int_{x=0}^{t} x^{n-1}\, dx}_{\frac{z^n}{n}} =$$

$$= \frac{\lambda^{n+1} z^n}{n!} e^{-\lambda z} = \Gamma(n+1, \lambda)$$

2) Gedächtnislosigkeit der Exponentialverteilung

$$P(x \leq t + \tau \mid x > t) = 1 - \frac{P(x > t+\tau, x > t)}{P(x > t)} = 1 - \frac{P(x > t + \tau)}{P(x > t} = 1 - \frac{e^{-\lambda(t+\tau)}}{\underbrace{e^{-\lambda t}}_{e^{-\lambda\tau}}} = 1 - e^{-\lambda\tau} \sim \text{Exp}(\lambda)$$

$x \sim \text{Exp}(\lambda)$

$$\exp(\lambda) \cdot \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Hazard Rate (Verteilung der verbleibenden Zeit)

$$G_t(x) = P(X - t \leq x \mid x > t) = \frac{P(x \leq x + t, x > t)}{P(x > t)} = \frac{F(t+x) - F(t)}{1 - F(t)}$$

$$\text{Rate: } \lambda(t) = \lim_{\Delta x \to 0} \frac{G_t(\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(t+\Delta x) - F(t)}{\Delta x} \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)}$$

3 Möglichkeiten für $f(t) = \lim_{\Delta x \to 0} \frac{F(t+x) - F(t)}{\Delta x}$

aging (wird älter): $\lambda(t)$ steigt

deaging (wird jünger): $\lambda(t)$ fällt

ageless (zeitlos): $\lambda(t) = c$

Intuitives Denken mit Poissonprozess

1) Kurzzeitverhalten

$$P(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h)$$

$$P(N(t+h) - N(t) = 1) = \lambda h + o(h)$$

$$P(N(t+h) - N(t) \geq 2) = o(h)$$

2) Hazardrate ist λ

inhomogener Prozess: nicht stationär, unabhängige Zuwächse

$$\lambda t \to \int_0^t \lambda(x)\, dx$$

$$P(N(t) = k) = \left(\int_0^t \lambda(x)\, dx\right)^k \frac{1}{k!} e^{-\int_0^t \lambda(x) dx}$$

Anwendung unter anderem bei Modellierung von Telefonanrufen (Warteschlangentheorie), bei Zerfall des Teilchens, genetische Änderung durch radioaktive Einstrahlung

$\boxed{\text{BSP}}$ $X \sim Poi(\lambda)$ $\quad Y \sim \Gamma(n,1)$ $\Rightarrow P(Y \leq \lambda) = P(X \geq n)$, $n \geq 1$

1. Lösung: (mittels Induktion)

$$\int_0^\lambda \frac{1^n x^{n-1}}{n-1} e^{-nx} = \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

bei $n = 1$ $\quad \int_0^\lambda e^{-x} dx = -e^{-x}\Big|_0^\lambda = 1 - e^{-\lambda}$

bei $k \to n+1$ $\quad \int_0^\lambda \frac{x^n}{n!} e^{-x} dx = \frac{1}{n!} \int_0^\lambda x^n e^{-x} dx = \frac{1}{n!}\left(-x^n e^{-x}\right)\Big|_0^\lambda + \int_0^1 n x^{n-1} e^{-x} dx = \underbrace{\int_0^\lambda \frac{x^{n-1}}{(n-1)!} e^{-x} dx}_{\sum \frac{\lambda^k}{k!} e^{-x} \ldots} - \frac{\lambda^n e^{-\lambda}}{n!} =$

$$= \sum_{k=n}^{\infty} \frac{\lambda^k}{n!} e^{-\lambda} - \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{k=n}^{\infty} \frac{\lambda^k}{h!} e^{-\lambda}$$

2. Lösung (eleganter) ... $\frac{d}{d\lambda}$

$$\frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \sum_{k=n}^{n} \frac{k \lambda^{k-1}}{k!} e^{-\lambda} - \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=n}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} - \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} \quad , n \geq 1$$

Wenn $\lambda \to 0$, sodass $\int_0^k \frac{\lambda^{n-1}}{(n-1)!} e^{-x} dx$ und auch $\sum^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$ also man 'sieht' das Konstante beim Integral 0 wird

3. Lösung: Interpretation

$X \sim$ Anzahl der Zuwächse im Intervalle $[0, \lambda]$

$Y \sim$ Zeitpunkt des ersten Zuwachs

$P(Y \leq \lambda) = P($n-ter Zuwachs, der vor λ stattfindet$) = P($# Zuwächse $\geq n) = P(X \geq n)$

Markovprozesse

Ein stochastischer Prozess heißt Markov, wenn $\mathbb{P}(X_t \leq x_n \mid X_{t+1})$ wobei T = Indexraum, S = Zustandsraum

Wir kategorisieren nach Intervallen

mit diskreter Zeit (Indexmenge $\in \mathbb{Z}, \mathbb{N}$ or $\mathbb{N}^+$)

mit stetiger Zeit (Indexmenge $\mathbb{R}$ or $\mathbb{R}^+$)

nach Zustandsraum

diskret oder stetig

Zufallsraum kann sowohl diskret als auch stetig sein.

DTMC Markovkette in diskreter Zeit

Ein stochastischer Prozess heißt DTMC genau dann, wenn gilt: $\mathbb{P}(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \dots X_{n+1} = x_{n+1}) = \mathbb{P}(X_n = x_n \mid X_{n+1} = x_{n+1})$ für $n \geq 1$

Übergangswahrscheinlichkeit: $\mathbb{P}(X_n = i \mid X_{n+1} = i)$ für $n \geq 1$ und $i, j \in S$

homogen: DTMC ist homogen, wenn Übergangswahrscheinlichkeit $\mathbb{P}(X_n = j \mid X_{n+1} = i)$ $\forall n \geq 1$ verschiebungsinvariant ist (gleich bleibt?)

also: $\mathbb{P}(X_n = j \mid X_{n-1} = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \dots \Rightarrow \mathbb{P}_{ij} = \mathbb{P}(X_n = j \mid X_{n-1} = i)$

Ab jetzt werden homogene DTMCs mit Indexwerten $\in \mathbb{N}$ ausgeschrieben!

Begriffe:

1 stufige Übergangswahrscheinlichkeit: $\mathbb{P}_{ij} = \mathbb{P}(X_1 = j \mid X_{n-1} = i)$

n stufige Übergangswahrscheinlichkeit: $\mathbb{P}_{ij}^{(n)} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$

Anfängliche Zustandswahrscheinlichkeit: $\mathbb{P}_i = \mathbb{P}(X_0 = i)$

n Stufige Zustandswahrscheinlichkeit: $\mathbb{P}_i^{(n)} = \mathbb{P}(X_n = i)$

CTMC Markovkette in stetiger Zeit

Markovkette – Markovprozess mit diskreten S/T

Mehrdimensionaler Markovprozess: S = mehrdimensional

1 stufige Übergangsmatrix: $\underline{P} = [\mathbb{P}_{ij}]$

n stufige Übergangsmatrix: $\underline{P}^{(n)} = [\mathbb{P}_{ij}^{(n)}]$

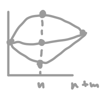
Anfänglicher Zustandsvektormatrix: $\underline{p} = [\mathbb{P}_i]$

n – Stufige Zustandsvektormatrix: $\underline{p}^{(n)} = [\mathbb{P}_i^{(n)}]$

Beschreibung mehrstufiger Übergänge

Chapman Kolmogorov Gleichung:

$$\underline{P}^{(n+m)} = \underline{P}^m \cdot \underline{P}^n \qquad n, m \geq 1$$



Beweis: Alle (disjunkten) Wege von einem Zustand zum anderen addieren (entspricht Matrizenmultiplikation)

Grundlegende Zusammenhänge

- $\underline{P}^{(n)} = \underline{P}^n$ $\qquad \underline{P}^{(n)} = \underline{P}^{(n-1)} \cdot \underline{P}$ (n mal machbar)
- $\underline{P}^{(n)}$ ist stochastisch
  $\downarrow$
  alle Elemente von $\underline{P}^{(n)}$ sind WSK $\to 0 \leq a_{ij} \leq 1$
  $\downarrow$
  Zeile i von $\underline{P}^{(n)}$ beschreibt WSK für Zustandswechsel $\Rightarrow \Sigma = 1$
- $\underline{p}^{(n)} = \underline{p}^{(0)} \cdot \underline{P}^{(n)}$ $\quad$ ($\underline{p}^{(0)}$ ist Anfangszustandsvektor!)

ex: stochastische Matrix:

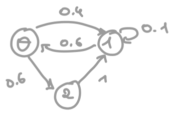
1) $0 \leq a_{ij} \leq 1$

2) Zeilensummen $= 1 \Leftrightarrow \sum_{j \in S} a_{ij} = 1 \quad \forall i \in S$

Festlegungsmöglichkeiten einer DTMC

a) Übergangsmatrix

b) Zustandsdiagramm



$$\Longleftrightarrow \begin{bmatrix} 0 & 0.4 & 0.6 \\ 0.9 & 0.1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \underline{\underline{P}}$$

Summe aller Pfeile weg von X muss 1 ergeben!

Charakterisierung des Verhaltens einer DTMC mittels $\mathbb{P}(\text{Aufenthaltsdauer in Zustand } i = n) = P_{ii}^{(n)}(1 - P_{ii}) \sim \text{Geometrisch}(1 - P_{ii})$

immer nach einer Zeit von $\sim$ Geometrisch wird Zustand gewechselt

Simulation einer DTMC

mit der Übergangsmatrix oder mit der (geometrisch verteilten) Aufenthaltsdauer

"Erzeugen" der Aufenthaltsdauer $\Rightarrow$ Bestimmen des nächsten Zustands (sofern es Übergang gibt) mit $\frac{P_{ij}}{1 - P_{ii}}$

$\boxed{\text{BSP}}$ $\underline{\underline{P}} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$

$$|\lambda \underline{\underline{I}} - \underline{\underline{P}}| = [\lambda - (1-a)][\lambda - (1-b)] - ab = \lambda^2 - \lambda(1-a+1-b) + (1-a)(1-b) - ab \overset{!}{=} 0$$

$$\lambda_{1,2} \overset{\text{binom. Formel}}{=} \dots = \frac{2-(a+b)+(a+b)}{2} = 1$$

$$\frac{2-(a+b)-(a+b)}{2} = 1-a-b$$

$$(\underline{\underline{P}} - \lambda \underline{\underline{I}})\, \underline{\underline{u}} = 0$$

$$\lambda = 1 \rightarrow (\underline{\underline{P}} - \underline{\underline{I}})\, \underline{u} = \emptyset \quad \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \emptyset \quad \begin{matrix} -au_1 + au_2 = 0 \\ bu_1 - bu_2 = 0 \end{matrix} \quad \left. \begin{matrix} u_1 = u_2 \\ u_1 = u_2 \end{matrix} \right\} \Rightarrow u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\lambda = 1-a-b \rightarrow (\underline{\underline{P}} - \lambda \underline{\underline{I}})\, \underline{u} = \emptyset \quad \begin{bmatrix} b & a \\ b & a \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \emptyset \rightarrow u_2 = \begin{pmatrix} a \\ -b \end{pmatrix} \longrightarrow \underline{\underline{U}} = \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix}$$

$$\underline{\underline{P}}^n = \underbrace{\underline{\underline{U}}\, \underline{\underline{\Lambda}}^n\, \underline{\underline{U}}^{-1}}_{\text{Spaltenzerlegung}} = \begin{bmatrix} 1 & a \\ 1 & -b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-b-a)^n \end{bmatrix} \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{1}{a+b} & \frac{-1}{a+b} \end{bmatrix} = \frac{1}{a+b} \begin{pmatrix} b & a \\ b & a \end{pmatrix} + 1-a+b)^n \frac{1}{a+b} \begin{pmatrix} a & -a \\ -b & b \end{pmatrix}$$

Wiederholung:

Ein stochastischer Prozess heißt Markovprozess in diskreter Zeit $\Leftrightarrow$ Markoveigenschaft $\mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, \dots X_0 = x_0) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1})$

Homogene Markovkette in diskrete Zeit: homogen $\Leftrightarrow$ $\mathbb{P}(X_n = j \mid X_{n-1} = i)$ $\forall i \in S \wedge n \geq 0$ in n Verschiebungsmoment ist.

$\hookrightarrow$ also $\mathbb{P}(X_{n+m} = j \mid X_{n+m-1} = i) = \mathbb{P}(X_n = j \mid X_{n-1} = i) = \mathbb{P}(X_1 = j \mid X_0 = i) \Rightarrow P_{ij} =$ Übergangswslk, $\underline{\underline{P}}$ Übergangsmatrix

Chapman-Kolmogorov Gleichung: $\underline{\underline{P}}^{n+m} = P^n \cdot P^m \quad \forall n, m \geq 1$

$P_{ij}^{(m)}$ n Stufige Übergangswslk $\Rightarrow$ $\underline{\underline{P}}^n$ n Stufige Üwslk

$P_i^{(0)}$ Anfängliche Zustandswslk $\Rightarrow$ $\underline{P}$ Anfängliche WSLK

$P_i^{(n)}$ n stufige Zustandswslk $\Rightarrow$ n stufige WSLK

Vorkenntnisse aus LinAG:

1) Bilden der Potenz der $\underline{\underline{P}}$

$$\underline{\underline{P}} = \underline{\underline{U}}\,\underline{\underline{\Lambda}}\,\underline{\underline{U}}^{-1}$$ Λ ist diagonale Matrix $\begin{bmatrix} 2&0&0\\0&1&0\\0&0&2\end{bmatrix}$

$$\underline{\underline{P}}^n = \underbrace{(\underline{\underline{U}}\,\Lambda\,\underline{\underline{U}}^{-1})\cdot\ldots\cdot(\underline{\underline{U}}\,\Lambda\,\underline{\underline{U}}^{-1})}_{n\text{ mal}} = \underline{\underline{U}}\,\Lambda^n\,\underline{\underline{U}}^{-1}$$ In Spalten zerlegen

2) Spektralzerlegung (von quadr. Matrix)

$\underline{\underline{A}}\,\underline{u} = \lambda\cdot\underline{u} \rightarrow$ $\underline{u}$ ist Eigenvektor, λ ist Eigenwert $\underline{\underline{A}}\cdot\underline{s} = \underline{s}$ ↵ bei stochastischem $\underline{\underline{A}}$ muss ein $\lambda = 1$ sein!

$(\overbrace{\lambda\,\underline{\underline{I}} - \underline{\underline{A}}}^{\text{singulär}})\,\underline{u} = \underline{\theta} \rightarrow |\lambda\,\underline{\underline{I}} - \underline{\underline{A}}| = \theta$ ... charakteristische Gleichung (n dimensional, ordnung)

$\Lambda = \begin{bmatrix}\lambda_1 & \cdots & 0\\ & \ddots & \\ 0 & \cdots & \lambda_n\end{bmatrix}$ $\underline{\underline{U}} = [\underline{u}_1, \underline{u}_2 \ldots] \rightarrow \underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Lambda}}\,\underline{\underline{U}}^{-1}$ (A ist diagonalisierbar) $\Leftrightarrow$ Matrix A genau n unabhängige Eigenvektoren hat

3) Inverses einer Matrix 2. Ordnung

$$\underline{\underline{A}} = \begin{bmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix} \quad \underline{\underline{A}}^{-1} = \frac{\begin{vmatrix} a_{22} & -a_{21}\\ -a_{12} & a_{11}\end{vmatrix}}{\det \underline{\underline{A}}}$$

4)

Berechnung von $\underline{\underline{U}}\,\underline{\underline{A}}\,\underline{\underline{U}}^{-1}$ mit Hilfe der Diaden

$\underline{\underline{U}} = [v_1\ v_2 \ldots]$ $[\underline{v}]^{(\underline{v})}$

$\underline{\underline{U}}^{-1} = \begin{bmatrix} w_1\\ \vdots\end{bmatrix}$ $\begin{bmatrix}u_1\\u_2\end{bmatrix}^{[v_1\,v_2]} = \begin{bmatrix} u_1 v_1 & u_1 v_2\\ u_2 v_1 & u_2 v_2\end{bmatrix}$

$$\left[\sum_i u_{i\lambda}\lambda_i w_{\lambda j}\right] = \sum \lambda_j\,[v_{i\kappa}]\,[w_{\kappa j}]$$ $\downarrow v_\kappa$, $\downarrow w_\kappa$ $\underline{\underline{U}} = [u_1\ u_2 \ldots]$

$$\rightarrow \underline{\underline{U}}\,\underline{\underline{\Lambda}}\,\underline{\underline{U}}^{-1} = \sum \lambda_\kappa \underbrace{v_\kappa}_{\text{spalte}} \underbrace{w_\kappa}_{\text{zeile}}$$

1) $p_{ij}^{(m)} \geq 0$

2) $\sum p_{ij}^{(m)} = 1 \quad \forall i \in S$

$$\underline{\underline{P}}^{(n)} = \underline{\underline{P}}^{(0)}\cdot\underline{\underline{P}}^{(n)} = \underline{\underline{P}}^{(0)}\cdot\underline{\underline{P}}^n$$

Aufenthaltsdauer im Zustand i ~ Geometrische Verteilung $(1-p_{ii})$

Simulationsmöglichkeiten:

Laut Übergangsmatrix

Mit Ausnutzung der Eigenschaft von Aufenthaltsdauer ~ $Geo(1-p_{ii})$

1) Erzeugen einer Zufallsvariable

2) Sprung in Zustand j $\frac{p_{ij}}{1-p_{ij}}$

WH: Z-Transformation / Erzeugende Funktion

$\underline{P(z)} = \sum P_t z^t \qquad |z| \leq 1$

↳ eindeutig verteilt für $p_i = \frac{1}{i!} \frac{d^i P(z)}{dz^i}\Big|_{z=0}$

→ Bestimmung von Faktorialmomenten

$\frac{d P(z)}{dz} = \sum_{i=1}^{\infty} i p_i = E[X]$

$\frac{d^2 P(z)}{d^2 z} = E[x(x-1)]$

ist.

→ $P(z) = \sum p_i = 1$

Eigenschaften der Zustände

Nachfolger: Ein Zustand ist ein Nachfolger von einem Zustand, wenn es ein n gibt, sodass $P_{ij}^{(n)} > 0$ (Berechnung $i \to j$)

Kommunizierende Zustände: Die Zustände i und j sind kommunizierend, wenn i Nachfolger von j und j Nachfolger von i

dh. $i \to j$ und $j \to i$ ($\Rightarrow$ ist Äquivalenzrelation (RST))

Periode eines Zustandes: Die Periode eines Zustandes ist durch $d(i) = ggT\{n \geq 0, P_{ii}^{(m)} > 0\}$ gegeben

Aperiodischer Zustand: Zustand i ist aperiodisch, wenn $\exists\, n_0$ sodass $P_{ii}^{(m)} > 0 \quad \forall\, n \geq n_0$

Periodischer Zustand: Zustand i ist periodisch mit Periode $d(i)$, wenn $d(i) \geq 2$ (sonst aperiod.)

Rekurrenter und Transienter Zustand

$\tau_i^* = \inf\{n \geq 1,\ X_n = i \mid X_0 = i\}$ – Rückkehrzeit

$\tau_i = \inf\{n \geq 1,\ X_n = i\}$ – Rückkehr oder Übergangszeit

Der Zustand i einer DTMC ist rekurrent wenn •) $P(\tau^* < \infty) = 1$, sonst (bei $P(\tau^* < \infty) < 1$) ist es transient

ODER Ein Zustand ist rekurrent wenn $\sum_{n=0}^{\infty} p_{ii}^{(m)} = \infty$, sonst (bei $\sum p_{ii} < \infty$) ist es transient.

äquivalente Definition

Beweis für 1&2

$f_{ii}^{n} = P(\tau_i^* = n) = P(\text{erste Rückkehr in den nten Übergang})$

$f_{ii} = \sum f_{ii}^{n} = P(\tau_i^* < \infty)$

$P_{ii}^{(n)} = \sum f_{ii}^{k} \cdot P_{ii}^{(n-k)} \qquad \Big| \sum_{n=1}^{\infty} z^n$

$\underbrace{\sum_{n=1} p_{ii}^{(n)} z^n}_{P_{ii}(z)-1} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{ii}^{(k)} \cdot P_{ii}^{(n-k)} \cdot z^n = \sum \sum f_{ii}^{(k)} p_{ii}^{(n-k)} z^k z^{n-k} = \underbrace{\sum f_{ii}^{(k)} \cdot z^k}_{F_{ii}(z)} \underbrace{\sum p_{ii}^{n-k} z^{n-k}}_{P_{ii}(z)}$

$P_{ii}(z) - 1 = F_{ii}(z) P_{ii}(z)$

$P_{ii}(z)[1 - F_{ii}(z)] = 1$

$P_{ii}(z) = \frac{1}{1 - F_{ii}(z)}$

$z = 1 \qquad P_{ii}(z) = \sum p_{ii}^{(m)}$

$F_{ii}(z) = \sum f_{ii} = f_{ii} = P(\tau_i^* < \infty)$

$P(\tau_i^* < \infty) = 1 \Rightarrow \sum p_{ii}^{(n)} < \infty$

$P(\tau_i^* < \infty) < 1 \to \sum p_{ii}^{(n)} < \infty$

Alle vier Definitionen sind äquivalent!

ODER Ein Zustand ist rekurrent wenn $P(r_i^* = \infty) = 1$ sonst $(P(r_i^* = \infty) = 0)$ ist es transient

[Nv]

$r_i^* = \#\{n > 0,\ X_n = i \mid X_0 = i\}$ Anzahl der Rückkehrer

$r_i^* = \#\{n > 0,\ X_n = i\}$ Anzahl der Rückkehrer oder Übergang

Beweis für 3

$Q_{ii}^{(m)} = P(v_i^* = \infty)$

$Q_{ii} = \lim Q_{ii} = P(r_i^* = \infty)$

$Q_{ii}^{(m)} = \sum f_{ii}\, Q_{ii}^{(m-1)} = f_{ii} Q_{ii}^{n-1}$

$Q_{ii}^{(1)} = P(\tau_i^* < \infty) = f_{ii} \qquad \Rightarrow Q_{ii}^{(n)} = f_{ii}^{n}$

Wenn $f_{ii} = P(\tau_i^* < \infty) = 1 \to \lim Q_{ii} = P(v_i^* = \infty) = 1$

$f_{ii} = P(\tau_i^* < \infty) < 1 \to \lim Q_{ii} = P(v_i^* = \infty) = 0$

Beweis für 4 (und alle)

Ereignis A_n = DTMC befindet sich nach ntem Übergang im Zustand i $X_0 = i$ $(X_n = i \mid X_0 = i)$

$v_i^* = \sum 1$

$E[v_i^*] = E[\sum 1_{\{A_n\}}] = \sum^{\infty} E[1_{\{A_n\}}] = \sum p_{ii}^{(m)}$

ODER Ein Zustand ist rekurrent wenn $E[v_i^* = \infty]$ sonst (bei $E[v_i^*] < \infty$) ist es transient

Positiv rekurrente: Ein rekurrenter Zustand ist positiv rekurrent, wenn $E[\tau_i^* < \infty]$

nullrekurrenter Zustand: Ein rekurrenter Zustand ist positiv rekurrent, wenn ($E[\tau_i^* < \infty]$) ist er Nullrekurrent

Absorbierender Zustand: Ein Zustand i ist absorbierend, wenn er keine Übergänge in einen anderen Zustand hat.

Folgerungen: aperiodisch, rekurrent $\sum p_{ii}^{(m)} = \sum_{n=0}^{\infty} 1 = \infty$

Klassen: Die von kommunizierend als Äquivalenzrelation partitionierte Menge einer DTMC heißen Klassen.

irreduzible Markovkette in Diskreter Zeit: Eine DTMC ist irreduzibel, wenn sie aus genau einer Klasse besteht.

Klasseneigenschaft: Eine Eigenschaft heißt Klasseneigenschaft, wenn sie entweder für alle Zustände der Klasse oder für keine gilt.

Eigenschaften der Klassen: Periode, Rekurrenz | Transienz, Positive Rekurrenz / Nullrekurrenz

1) Periode: Zustände k, j sind kommunizierend

$P_{jk}^{(n)} > 0$

$P_{kj}^{(s)} > 0$

$P_{jj}^{(n+s)} \geq P_{jk}^{(n)} \cdot P_{kj}^{(s)} > 0 \Rightarrow d(j) \mid n+s$

Existiert ein n sodass $P_{kk}^{(n)} > 0$

$P_{jj}^{(k+n+s)} \geq P_{jk}^{(s)} \cdot P_{kk}^{(n)} \cdot P_{kj}^{(s)} > 0 \rightarrow d(j) \mid k+n+s$?

$d(\ ) \mid n \Rightarrow d(j) \mid d(p)$

$\downarrow$ für alle n auf $P_{kk}^{(n)} > 0$

umgekehrt gilts auch: $d(k) \mid d(j)$

$\Rightarrow d(j) = d(k)$

2) Rekurrent, Transient: Zustände j, k sind kommunizierend

$a = P_{jk}^{(r)} > 0 \qquad b = P_{kj}^{(s)} > 0$

\ | Kann man nicht lesen

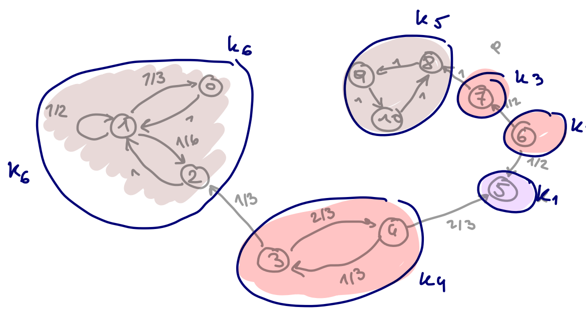
$P^{(r+m+s)} \geq P_{jk}^{(r)} P_{kk}^{(m)} P_{kj}^{(s)}$ für $n > 1$

$\sum_{n=1} P^{(r+m+s)} \geq \sum P_{jk}^{(r)} P_{kk}^{(m)} P_{kj}^{(s)} = ab \sum P_{rr}^{(m)}$

Wenn r rekurrent, $\sum P_{rj}^{(r)} = \infty \Rightarrow \sum P_{ii}^{(r+j+k)} = \infty \rightarrow$ Zustand ist auch rekurrent

Wenn j transient, $\sum P^{(r+m+s)} < \infty \Rightarrow \sum P_{rr}^{(\dots)} \leq \infty$ Zustand P ist auch transient

BSP



Klasse 1 — absorbierend

Klasse 2, 3, 4 — transient

Klasse 5, 6 — rekurrent

Knoten = Zustände

Kanten = Wahrscheinlichkeit

Zustand 8 $\sum p_{ii}^{(3n)} = \sum 1 = \infty \rightarrow$ rekurrent

Zustand 1
$f_{11}^{(1)} = \frac{1}{2}$
$f_{11}^{(2)} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$
$f_{11}^{(n)} = 0 \quad n \geq 3$

$\Rightarrow P(\tau_1^* < \infty) < f_{11} = \sum f_{11} = f_{11}^{(1)} + f_{11}^{(2)} = \frac{1}{2} + \frac{1}{2} = 1$

$E[\tau_1^*] = \sum n f_{11}^{(m)} = 1 \cdot \frac{1}{2} + 2 \frac{1}{2} = 1{,}5 < \infty$

$\Rightarrow$ Zustand 1 ist positiv rekurrent

Klasse 5 Periode 3

Klasse 4 Periode 2

Zustand 0:

$$T_{\{1,2\}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} \\ 1 & 0 \end{bmatrix} \rightarrow \text{Substochastisch}$$

$\sum f_{00} = P_{01} [T_{\{1,2\}}] P_{10} = 1 \cdot 3 \cdot \frac{1}{3} = 1 \rightarrow$ Zustand 0 ist rekurrent

$\underline{\underline{T}}$ ist substochastisch & $\sum T_{\{1,2\}} < \infty$ wenn alle $|\lambda_i| < 1$ $= \sum_{n=0}^{\infty} T_{\{1,2\}} < \infty$

$$\sum_{n}^{\infty} T_{\{1,2\}} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -1 & 1 \end{bmatrix}^{-1} = \frac{1}{\frac{1}{2} - \frac{1}{6}} \begin{bmatrix} 1 & 1/6 \\ 1 & 1/2 \end{bmatrix} = 3 \begin{bmatrix} 1 & 1/6 \\ 1 & 1/2 \end{bmatrix}$$

Stationäre Verteilung einer DTMC und Grenzverteilung einer DTMC

Def: Die stationäre Verteilung einer DTMC ist durch die Lösung des folgenden linearen Gleichungssystems gegeben:

$$\underline{p} = \underline{p}\,\underline{\underline{P}} \qquad \underline{p}\cdot\underline{e} = 1 \qquad (\text{wobei } \underline{e} \text{ ist Spaltenvektor bestehend aus 1em})$$

Folgerungen:

$$\underline{p}^{(0)} = \underline{p} \Rightarrow \underline{p}^{(1)} = \underline{p}^{(0)}\underline{\underline{P}} = \underline{p}\,\underline{\underline{P}} = \underline{p} \Rightarrow \text{rekursiv } \underline{p}^{(n)} = \underline{p} \quad \forall n \geq 0$$

(1D Randverteilung von Kette ist $\underline{p}$) eindimensionale

aka stationäre Version

ndimensionale Randverteilung $\underline{p}^{(0)} = \underline{p}$

$$P(X_n = i_0, \dots, X_{n+m} = i_m) = P(X_n = i_0)\,P(X_{n+1} = i_1 \mid X_n = i_0)\,P(X_{n+2} = i_2 \mid X_n = i_0, X_{n+1} = i_1) \dots P(X_{n+m} = i_n \mid X_n = i_0, \dots, X_{n+m-1} = i_m) =$$

'kürzen' durch Markoveigenschaft

$$= P(X_n = i_0)\,P(X_{n+1} = i_1 \mid X_n = i_0)\,P(X_{n+2} = i_2 \mid X_{n+1} = i_1) \dots P(X_{n+m} = i_n \mid X_{n+m-1} = i_m) =$$

$$= p_{i_0}^{(n)}\, p_{i_0,i_1} \dots p_{i_{m+1},i_n}$$

gleich!

$$= P(X_{n+1} = i_0, X_{n+h+1} = i_1 \dots = p_{i_0}^{(n)}\, p_{i_0,i_1} \dots p_{i_{m+1},i_n}$$

Die Markovkette in diskreter Zeit $\underline{p}^k = p$ ist ein stationärer Prozess (und somit eine stationäre Version der DTMC)

Diese Version der DTMC (also wenn 1D + nD) ist stationär

Fundamentalsatz Grenzwertsatz der Markovkette in diskreter Zeit

Wenn DTMC irreduzibel, aperiodisch und rekurrent ist, dann sind die Grenzwerte

1) $$\lim_{n\to\infty} p_{jj}^{(n)} = \frac{1}{\underset{\uparrow \text{Erwartungswert}}{\mu_j}} = \frac{1}{\sum n f_{jj}^{(n)}} = \begin{cases} < \infty & \text{wenn Zustand } j \text{ positiv rekurrent ist.} \\ = 0 & \text{wenn Zustand } j \text{ der Nullwert ist.} \end{cases}$$

2) vom Anfangszustand unabhängig

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} p_{jj}^{(n)} \qquad \forall i \in \underset{\text{Zustandsraum}}{S}$$

Folgerungen:

a) $$\underset{\downarrow\ \underline{p}^{(\infty)}}{p_j^{(\infty)}} = \lim_{n\to\infty} p_j = \lim_{n\to\infty} \sum_{i\in S} p_i^{(0)} \cdot p_{ij}^{(n)} = \underbrace{\sum_{i\in S} p_i^{(n)}}_{1} \underbrace{\lim_{n\to\infty} p_{ij}^{(n)}}_{\text{unabh. von } i} = \lim_{n\to\infty} p_{ij}^{(n)} = \underbrace{p_{ij}^{(\infty)}}_{\underline{\underline{P}}^{(\infty)}}$$

nach ∞ vielen Schritten ist verteilung genau gleich mit so einer Matrix (?)

aka

$$\begin{array}{cc} & \text{Spalte } j \\ \text{Zeile } i \rightarrow & \left[\begin{matrix} x \\ x \\ x \end{matrix}\right] \end{array}$$

Zeilen der Matrix sind gleich!

(Zeilen von $\underline{\underline{P}}^{(\infty)}$ sind gleich)

b) $$\underline{p}^{(\infty)} = \underline{p}^{(0)} \cdot \underline{\underline{P}}^{(\infty)}$$

$$\underline{\underline{P}}^{(\infty)} = \begin{pmatrix} \underline{p}^\infty \\ \vdots \\ \underline{p}^\infty \end{pmatrix} = \begin{pmatrix} p_0^\infty & p_1^\infty & \dots & p_j^\infty \\ p_0^\infty & p_1^\infty & \dots & p_j^\infty \\ \vdots \\ p_0^\infty & p_1^\infty & \dots & p_j^\infty \end{pmatrix}$$

$$(a_0\ a_1 \dots a_j \dots) \begin{pmatrix} p_0^{(\infty)} & p_1^{(\infty)} & \dots & p_j^{(\infty)} \\ p_0^{(\infty)} & \vdots & & \vdots \\ \vdots \\ p_0^{(\infty)} & p_1^{(\infty)} & & p_j^{(\infty)} \end{pmatrix} = \begin{pmatrix} \sum a_i\, p_0^{(\infty)} \\ \vdots \\ \sum a_i\, p_0^{(\infty)} \end{pmatrix}^T$$

Wenn eine DTMC irreduzibel, aperiodisch und positiv rekurrent ist, dann sind die Grenzwerte $\lim_{n\to\infty} p_{ij} = \lim_{n\to\infty} p_j$ eine Wahrscheinlichkeitsverteilung $\underline{p}^\infty \cdot \underline{e} = 1$ und diese eindeutig mit dem folgenden GLS bestimmt werden können:

$$\underline{p}^\infty = \underline{p}^\infty \cdot \underline{\underline{P}} \qquad \underline{p}^\infty \cdot \underline{e} = 1$$

Folgerung: $\underset{\text{Grenzwertverteilung}}{\underline{p}^{(\infty)}} = \underset{\text{Stationäre Verteilung}}{p}$

Beweis: Aus zweiter Aussage folgt 1. Aussage → nur 2. Aussage muss bewiesen werden.

x_k = Lösung des Gleichungssystems

$$x_k = \sum_{i\in S} x_i\, p_j \quad \Big|\ \sum_k p_{kl}$$

$$\sum_k x_k\, p_{kl} = \sum_k \sum x_i\, \underbrace{p_j \cdot p_{kl}}_{p_{il}^{(2)}} = \sum x_i\, p_{ik}^{(2)} \Rightarrow x_k = \sum_{i\in S} x_i\, p_{ik}^{(2)} \dots \text{ Rekursiv weiter machen} \Rightarrow$$

ok, das sieht ja fast so aus wie am Anfang!

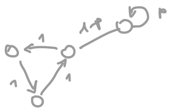
$$x_k = \sum_{i\in S} x_i\, p_{ik}^{(n)} \geq 1$$

$$\lim_{n\to\infty} x_k = x_k = \lim_{n\to\infty} \sum x_i\, p_{ik}^{(n)} = \sum x_i \overbrace{\lim_{n\to\infty} p_{ik}^{(n)}}^{\text{unabh. von } i}$$

$$= \lim_{n\to\infty} p_{jj}^{(n)} = \lim_{n\to\infty} p_j^{(n)} = \underline{p}^{(\infty)}$$

Eine DTMC ist irreduzibel, periodisch und rekurrent, dann

- Existieren die Grenzwerte $\lim_{n\to\infty} p_j^{(n)}$ nicht.
- Existieren die Grenzwerte $\lim_{n\to\infty} \frac{1}{n}\sum_{k}^{n} p_j^{(k)}$ schon.



$p_1^{(m)} = 1$ $\quad$ $p_1^{(3m)} = 1$

$p_1^{(m+1)} = 0$ $\quad$ $p_1^{(3m+1)} = 0$

$p_1^{(m+2)} = 0$ $\quad$ $p_1^{(3m+1)} = 0$

$p_1^{(m+3)} = 1$

$\lim \frac{m}{3n}$ $\quad$ $\lim \frac{n}{3m+1}$ $\quad$ $\lim \frac{m}{3m+1}$ $\rightarrow 0$

Relationen in den verschiedenen Verteilungen für irreduzible, positiv rekurrente DTMC

Zustand i	stationäre Verteilung p_i	$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} p_i^{(k)}$	Grenzwertverteilung $p_i^{(\infty)} = \lim_{n\to\infty} p_i^{(n)}$
aperiodisch	Alle 3 Grenzwerte existieren und sind gleich!		
periodisch	Grenzwerte existieren und sind gleich		$\nexists$ $\rightarrow$ existiert nicht!

Absorption

Eine DTMC hat eine Menge von rekurrenten Zuständen (C) und eine transiente Klasse (T)

Absorptionswahrscheinlichkeit: $a_i = P($ Absorption in C findet in endlicher Zeit statt $| X_0 = i)$ $\qquad (i \in T)$

↳ berechnen: Wenn eine DTMC eine Menge von rekurrenten Zuständen und eine transiente Klasse hat, dann sind die Absorptionswslk die kleinsten nicht negativen Lösungen folgenden LGS:

$a_i = 1 \qquad i \in C$

$a_i = \sum p_{ij} \cdot a_j \qquad i \in T$

Beweis: $a_i^{(n)} \overset{?}{=} P($Die Kette befindet sich nach n Übergängen in C$| X_0 = i)$, $i \in T$ (Übergang C kann nach k Stufen passieren $0 \le k \le n$)

1) $a_i^{m+1} = \sum_{i\in T} p_{ij}\, a_j^{(m)} + \sum_{j\in C} p_{ij} = 1$

2) $a_i^{(m+1)} \geq a_i^{(m)}$ (mit Induktion)

$a_i^{(m)} = \theta$

$a_i^{(m+1)} = \sum p_{ij}\, a_j^{(m+1)} + \sum_{j\in C} p_{ij} \cdot 1$

$a_i^{(n+1)} = \sum p_{ij}\, a_j^{(m)} + \sum_{j\in C} p_{ij} \cdot 1$

3) $\underbrace{\lim_{n\to\infty} a^{m+1}}_{a_i} = \lim \sum p_{ij}\, a_j^{(m)} + \sum_{j\in C} p_{ij} \cdot 1$

$a_i = \sum_{j\in T} p_{ij}\, a_j + \sum_{j\in C} p_{ij} \cdot 1 = \sum_{j\in C} p_{ij}\, a_j$

4) $y_i \geq a_i^{(0)} \underset{n=0}{=} 0$ (mit mathematische Induktion)

$y_i \geq a_i^{(m)} \Rightarrow y_i \geq a_i^{(m+1)}$

$a_i^{(m+1)} = \sum_{j\in T} p_{ij}\, a_j^{(m)} + \sum_{j\in C} p_{ij} \cdot 1 \geq y_i = \sum_{j\in T} p_{ij}\, y_j + \sum_{j\in C} p_{ij} \cdot 1$

$\Rightarrow y_i \geq a_i^{(m)} \quad \forall m \geq 1$

$y_i \geq a_i$ $\quad$ ($a_i^{(0)}$ monoton steigend und $a_i = \lim_{n\to\infty} a_i^{(n)}$)

Relation für Absorptionswahrscheinlichkeit in Matrix / Vektor form

$\underline{a} = [a_i] \quad i \in T \qquad \underline{a}$ ist Spaltenvektor

$\underline{T} = [p_{ij}] \quad ij \in T$

$\underline{C} = [p_{ij}] \quad i \in T,\ j \in C$

$\underline{\underline{a}} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1}\, \underline{\underline{C}}\, \underline{e}$

Relation für Absorptionswahrscheinlichkeit in Matrix/Vektorform

$\underline{a} = [a_i] \quad i \in T \qquad \underline{a}$ ist Spaltenvektor

$\underline{\underline{T}} = [p_{ij}] \quad ij \in T$

$\underline{\underline{C}} = [p_{ij}] \quad i \in T,\ j \in C$

$$\underline{a} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1}\ \underline{\underline{C}}\ \underline{e}$$

Beweis: $a_i = \sum_{j \in C} p_i a_i \quad i \in T$

$a_i = \sum_{l \in T} p_{lj} a_l + \sum_{j \in C} p_{ij} \cdot 1 \quad i \in T$

$\underline{a} = \underline{\underline{T}}\ \underline{a} + \underline{\underline{C}}\ \underline{e}$

$(\underline{\underline{I}} - \underline{\underline{T}})\underline{a} = \underline{\underline{C}}\ \underline{e}$

$\underline{a} = \underbrace{(\underline{\underline{I}} - \underline{\underline{T}})^{-1}}_{\sum_{n=0}^{\infty} \underline{\underline{T}}^n}\ \underline{\underline{C}}\ \underline{e}$

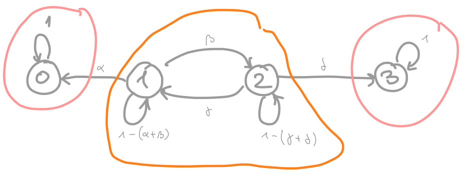
$\sum_{n=0}^{\infty} \underline{\underline{T}}^n \to |\lambda_i| < 1 \Rightarrow T$ ist substochastisch

Alternativer Beweis

$\sum P$(Absorption der Menge C, der Zustand j erreicht) $= \sum_{j \in C} \sum_{n=0}^{\infty}$ (Absorption nach n Übergängen der Klasse T im Zustand $j \in C$ stattfindet) $=$

$= \sum\sum \underline{\underline{T}}^n [\underline{\underline{C}}] = \sum \underline{\underline{T}}^n \underline{\underline{C}}\ \underline{e} = (\underline{\underline{I}} - \underline{\underline{T}})^{-1} \underline{\underline{C}}\ \underline{e}$

BSP



absorbierender Zustand

transient

$$\underline{\underline{P}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1-(\alpha+\beta) & \beta & 0 \\ 0 & \gamma & 1-(\gamma+\delta) & \delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow$$

$= \underline{\underline{T}}$

$\underline{\underline{C}} = \begin{bmatrix} \alpha & 0 \\ 0 & \delta \end{bmatrix}$

$\underline{\underline{C}} \cdot \underline{e} = \begin{bmatrix} \alpha \\ \delta \end{bmatrix}$

$$[\underline{\underline{I}} - \underline{\underline{T}}]^{-1} = \begin{bmatrix} \alpha+\beta & -\beta \\ -\gamma & \gamma+\delta \end{bmatrix}^{-1} = \frac{1}{(\alpha+\beta)(\gamma+\delta) - \beta\gamma} \begin{bmatrix} \gamma+\delta & \beta \\ \gamma & \alpha+\beta \end{bmatrix}$$

$$\underline{a} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1}\ \underline{\underline{C}}\ \underline{e} = \begin{bmatrix} \alpha(\gamma+\delta) + \beta\delta \\ \gamma\alpha + (\alpha+\beta)\delta \end{bmatrix} \frac{1}{\alpha\gamma + \alpha\delta + \beta\delta} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

WSLK absorbiert zu werden, wenn bei ① angefangen wird

WSLK bei ② anzufangen

Wenn man nur Zustand 0 oder 3 betrachtet:

$\underline{\underline{C}}_0 = \binom{\alpha}{0} \Rightarrow \underline{\underline{C}}_0 \cdot \underline{e} = \binom{\alpha}{0} \Rightarrow \underline{a}_0 = \frac{1}{\alpha\gamma + \alpha\delta + \beta\delta} \begin{bmatrix} \gamma+\delta & \beta \\ \gamma & \alpha+\beta \end{bmatrix} \binom{\alpha}{0} = \begin{bmatrix} \alpha(\gamma+\delta) \\ \gamma\alpha \end{bmatrix} \frac{1}{\alpha\gamma + \alpha\delta + \beta\delta}$

$\underline{\underline{C}}_3 = \binom{0}{\delta} \Rightarrow \underline{\underline{C}}_3 \cdot \underline{e} = \binom{0}{\delta} \Rightarrow \underline{a}_3 = \frac{1}{\alpha\gamma + \alpha\delta + \beta\delta} \begin{bmatrix} \gamma+\delta & \beta \\ \gamma & \alpha+\beta \end{bmatrix} \binom{0}{\delta} = \begin{bmatrix} \beta\delta \\ (\alpha+\beta)\delta \end{bmatrix} \frac{1}{\alpha\gamma + \alpha\delta + \beta\delta}$

Mittlere Zeit bis zur Absorption: Wenn eine DTMC wieder aus den Mengen T und C besteht, können die mittleren Absorptionszeiten mit folgendem GLS bestimmt werden:

$t_i = 0 \quad i \in C$

$t_i = 1 + \sum p_{ij} \cdot t_j \quad i \in T$

Relation für mittlere Absorptionszeiten in Matrix/Vektorform:

$\underline{t} = [t_i] \quad i \in T$ Spaltenvektor

$\underline{t} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1} \underline{e}$

Rechnung $t = 1 + \sum p_{ij} t_j \quad i \in T$

$\underline{t} = \underline{e} + \underline{\underline{T}}\,\underline{t} \Rightarrow (\underline{\underline{I}} - \underline{\underline{T}})\underline{t} = \underline{e}$

$\underline{t} = [\underline{\underline{I}} - \underline{\underline{T}}]^{-1} \underline{e}$

Varianz der Absorptionszeit:

$v_i = 0 \quad i \in C$

$v_i = \sum_{j \in S} p_{ij} (1 + \underline{t}_j - \underline{t}_i)^2 + \sum p_i v_i \quad i \in T$

Voraussetzungen für Zustände und Klasseneigenschaften

Irreduzibilität – hinreichende Kriterien

Übergangsmatrix: Die WSLK in den Nebendiagonalen der $\underline{\underline{P}}$ über und unter der Hauptdiagonale sind positiv.

Zustandsdiagramm: Es gibt einen Übergang von allen Zuständen, ein Zustand hinauf und hinunter.

diag(A)

aperiodischer Zustand (hinreichende Kriterien)

Übergangsmatrix: Die WSLK in Hauptdiagonale sind positiv.

Zustandsdiagramm: Es gibt ein „loop" in jedem Zustand zu sich selbst mit positiven Nebendiagonalen
↳ Schleife

rekurrenter Zustand

Für ein irreduzible DTMC, sind der Zustand $\emptyset$ und durch die Irreduzibilität alle Zustände genau dann rekurrent, wenn die kleinste nichtnegative Lösung (für $a_i = 1$ für $i \geq 1$) durch folgende Gleichung gegeben ist:

$a_0 = 1$

$a_i = \sum p_{ij} a_j$ für $i \geq 1$

$\{n > 0 \quad X_n = i \mid X_0 = j\} \longrightarrow P(\tau_i^* < \infty) \leq 1$

$X_0 = j \qquad i \geq 1$

positiv rekurrenter Zustand (hinreichende Bedingung)

Foster Kriterium: Sei eine DTMC irreduzibel. Wenn die Zahlen $I \geq 1$, $c > 0$, $d > 0$ existieren und der Erwartungswert ist

$$E[X_{n+1} \mid X_n = k] = \begin{cases} \leq c & \text{für } k \leq I \\ \leq k - d & \text{für } k > I \end{cases}$$

dann ist die DTMC positiv rekurrent.

Ergänzungen

- Allgemeine Struktur einer DTMC

$$\underline{\underline{P}} = \begin{pmatrix} \boxed{K_1} & \cdot & \cdot \\ \cdot & \boxed{K_1} & \cdot \\ \boxed{\quad T \quad} & & \end{pmatrix}$$

T… Menge der Transienten Zustände (Bestehend aus einer oder mehreren Klassen)

- Endliche DTMC: Hat endliche Anzahl von Zuständen

Eine endliche DTMC muss mindestens einen rekurrentes Zustand haben.

Beweis: $\lim_{n\to\infty} \sum_{j=0}^{m-1} p_{ij}^{(n)} = 1 \Rightarrow \sum \lim p_{ij}^{(n)} = \left(\lim_{n\to\infty} p_i^{(n)} = p_j^{(n)}\right) = 1$

Existiert ein $j \in (0 \ldots M_S)$ wofür $\lim p_j^{(n)} > 0$ so ist es positiv rekurrent.

Folgerung: Eine irreduzible DTMC ist immer auch positiv rekurrent.

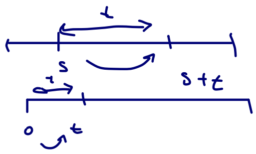
Markovkette in stetiger Zeit CTMC

Ein paar Begriffe ...

... Ein stochastischer Prozess $(X(t), t \geq 0)$ heißt Markovprozess in stetiger Zeit, wenn er die Markoveigenschaften erfüllt.

$$P(X(t_n)=x_n \mid X(t_{n-1})=x_{n-1}, \dots X(t_1)=x_1, X(t_0)=x_0) = P(X(t_n)=x_n \mid X(t_{n-1})=x_{n-1}) \quad \forall\, t_0 < t_1 < \dots < t_n \text{ und } x_i \in S \text{ für } i=0,1,2,\dots$$

... homogene CTMC: Ist homogen, wenn $P(X(s+t)=j \mid X(s)=i) = P(X(t)=j \mid X(0)=i) \quad \forall\, t,s \geq 0$ und $i,j \in S$



... Chapman Komolgorov: $\underline{\underline{P}}(s+t) = \underline{\underline{P}}(s) \cdot \underline{\underline{P}}(t)$

$P_i(0) \Rightarrow \underline{P}(0)$
$P_i(t) \Rightarrow \underline{P}(t)$

Infinitsemale Parameter

- $\lim_{t \to \infty} P_{ij}(t) = \delta_{ij} = \begin{cases} 1 & \text{wenn } i=j \\ 0 & \text{wenn } i \neq j \end{cases} \Rightarrow P_{ij}(t)$ ist stetig zum Zeitpunkt $t=0$

  ↳ notwendige Bedingung dass Ableitung existiert

- $$q_{ij} = \lim_{t \to 0^+} \frac{P_{ij}(t) - P_{ij}(0)}{t} = \lim_{t \to 0^+} \frac{P_{ij}(t) - \delta_{ij}(t)}{t} = \frac{dP_{ij}(t)}{dt}\Big|_{t=0} = \frac{d\, P(X(t)=j \mid X(0)=i)}{dt}\Big|_{t=0} = \frac{d\, P(X(t+T)=j \mid X(T)=i)}{dt}\Big|_{t=0}$$

  Folge immer Positiv weil es kommt von θ — Homogenität — "zum Zeitpunkt $t=0$"

  $q_{ij} \geq 0$ für $i \neq j$

  $q_{ii} \leq 0$ für $i=j$ ($q_{ii} = -\infty$ kann auch vorkommen!)

- Erzeuger / Generatormatrix

  $$\underline{\underline{Q}} = [q_{ij}] = \lim_{t \to 0^-} \frac{\underline{\underline{P}}(t) - \underline{\underline{P}}(0)}{t} = \lim_{t \to 0^-} \frac{\underline{\underline{P}}(t) - \underline{\underline{I}}}{t}$$

Kriterien für Infinitsemalparameter

1) CTMC ist konservativ wenn $q_{ii} > -\infty \quad \forall i \in S$, $\sum_{j \in S} q_{ij} = 0 \quad \forall i \in S$

2) $\inf q_{ii} > -\infty \quad \forall i \in S \Rightarrow \sum_{j \in S} q_{ij} = \sum \lim_{t \to 0} \frac{P_{ij}(t) - \delta_{ij}}{t} = \lim_{t \to 0^-} \sum \frac{[P_{ij}(t) - \delta_{ij}(t)]}{t} = \lim_{t \to 0^-} 0 = 0$

Kolmogorov Gleichungen

Rückwärtsgleichung: Wenn eine CTMC 1) erfüllt, dann gilt

$$\underline{\underline{P}}'(t) = \underline{\underline{Q}}\,\underline{\underline{P}}(t) = \lim_{t \to 0^+} \frac{\underline{\underline{P}}(t+\Delta t) - \underline{\underline{P}}(t)}{t} = \lim_{t \to 0^+} \frac{\underline{\underline{P}}(t)\underline{\underline{P}}(\Delta t) - \underline{\underline{P}}(t)}{t} = \lim_{t \to 0^+} \frac{\underline{\underline{P}}(\Delta t)\underline{\underline{P}}(t) - \underline{\underline{P}}(t)}{t} = \underbrace{\lim_{t \to 0^+} \frac{\underline{\underline{P}}(\Delta t) - \underline{\underline{I}}}{t}}_{\underline{\underline{Q}}} \cdot \underline{\underline{P}}(t)$$

Vorwärtsgleichung: Wenn eine CTMC 2) erfüllt, dann gilt

$$\underline{\underline{P}}'(t) = \underline{\underline{P}}(t)\,\underline{\underline{Q}} = \lim_{t \to 0^+} \frac{\underline{\underline{P}}(t+\Delta t) - \underline{\underline{P}}(t)}{t} = \lim_{t \to 0^+} \frac{\underline{\underline{P}}(t)\underline{\underline{P}}(\Delta t) - \underline{\underline{P}}(t)}{t} = \lim_{t \to 0^+} \underline{\underline{P}}(t) \frac{\underline{\underline{P}}(\Delta t) - \underline{\underline{I}}}{t} = \underline{\underline{P}}(t) \lim_{t \to 0^+} \frac{\underline{\underline{P}}(\Delta t) - \underline{\underline{I}}}{t}$$

Lösung: $\underline{\underline{P}}(t) = e^{\underline{\underline{Q}}t}$

Interpretation des Verhaltens einer CTMC

- Interpretation der Infinitsemalen Parameter

  q_{ij} Die Parameter der exponentielle verteilten Aufenthaltsdauer im Zustand i vor einem Übergang in den Zustand j

  $X_i = \min X_{ij}, \quad X_{ij} \sim \text{Exp}(q_{ij})$

  $$P(X_i < t) = 1 - P(X_i > t) = 1 - \prod_{j \neq i} \underbrace{P(X_{ij} > t)}_{e^{-q_{ij}t}} = 1 - \prod e^{-q_{ij}t} = 1 - e^{-(\sum_{j \neq i} q_{ij})t} = 1 - e^{q_{ii}t} = 1 - e^{-(-q_{ii})t}$$

  $$\sum_j q_{ij} = 0 \Leftrightarrow q_{ii} + \sum_{j \neq i} q_{ij} = 0 \Rightarrow q_{ii} = -\sum_{j \neq i} q_{ij}$$

  q_{ii} (-1) mal mal die Parameter der exp. verteilen Aufenthaltsdauer im Zustand i

Kurzzeitverhalten

- $P(X(t+\Delta t)=j \mid X(t)=i) = P_{ij}(\Delta t) = q_{ij}\Delta t + o(\Delta t)$ „klein oh"
- $P(X(t+\Delta t)=i \mid X(t)=i) = P_{ii}(\Delta t) = 1 + q_{ii}\Delta t + o(\Delta t)$ $\quad \| \; 1-\sum_{j\neq i} q_{ii}\Delta t = 1+q_{ii}\Delta t$
- $P(X(t+\Delta t)=j, X(t+\frac{\Delta t}{2})=k \mid X(t)=i) = o(\Delta t)$ Poisson ?
- $P(X(t+\Delta t)\neq i \mid X(t)=i) = \sum_{j\neq i} q_{ij}\Delta t + o(\Delta t) = -q_{ii}\Delta t + o(\Delta t)$ von i irgendwo hin

Simulation einer CTMC (2 Arten)

1) Laut Interpretation von Q

Paralell verlaufende, exponential verteilte Zufallsvariablen simuliert $\sim Exp(q_{ij})$

Übergang in Zustand, welcher von $X_i \sim \min X_{ij}$ bestimmt wird.

2) Mit Hilfe von Eingebetteter DTMC

Aufenthaltsdauer in Zustand $i \sim Exp(-q_{ii})$

Übergang nach $j\neq i$ mit WSLK $\frac{q_{ij}}{-q_{ii}}$

Eingebettete DTMC (in Übergangszeitpunkten)

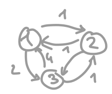
$P_{ii} = 0$

$P_{ij} = \frac{q_{ij}}{\sum_{j\neq i} q_{ij}} = \frac{q_{ij}}{-q_{ii}} \quad j\neq i$

Zustands und Klasseneigenschaften

$\underset{\text{DTMC}}{Y(n)} = \underset{\text{CTMC}}{X(a_n)} \qquad a>0 \qquad t=a_n$

$\underline{\underline{Q}} = \begin{bmatrix} -3 & 1 & 2 \\ 1 & -2 & 1 \\ 4 & 1 & -5 \end{bmatrix}$



Irreduzibilität: DTMC – CTMC mapping (mit Zustandsdiagramm darstellen)

Periode: Alle Zustände sind aperiodisch ($p_{ii}(t)>0$ für $t>0$)

Für ein Zustandspaar i, j $\begin{cases} \text{entweder } p_{ij}(t)>0 & \forall t>0 \\ \text{oder } p_{ij}(t)=0 & \forall t>0 \end{cases}$

$\underline{P}(t) = e^{\underline{Q}t} = \sum_{n=0}^{\infty} \frac{\underline{Q}^n t^n}{n!} \qquad q_{ij}>0$

Rekurrenz | Transienz mittels DTMC-CTMC „mapping" – für alle $a>0$ muss b gezeigt werden (mit Zustandsdiagramm darstellen)

Absorbierender Zustand mittels DTMC-CTMC „mapping" – für alle $a>0$ muss b gezeigt werden (mit Zustandsdiagramm darstellen)

$q_{ii}=0$ in Matrix hat man volle Nullzeile

Absorption:

gegeben: CTMC mit einer Menge von rekurrenten Zuständen C und einer transienten Klasse T

AbsorptionsWSLK: $a_i = P(\text{Absorption in } C \mid X_0 = i)$

Berechnung: Wenn eine CTMC eine Menge von rekurrenten Zuständen C und eine Transiente Klasse T hat, dann sind die AbsorptionsWSLK Lösungen von dem folgenden LGS:

$a_i = 1 \quad i\in C$

$0 = \sum q_{ij} a_j \quad i \in T$

$a_i = (1+a_{ii}dt)a_i + \sum_{j\neq i} q_{ij}\,dt\,a_j \quad \| \cdot dt, \lim_{dt\to 0}$

$0 = a_{ii}a_i + \sum_j a_{ij}a_j \Rightarrow 0 = \sum q_{ij}a_j$

Alternative Form der Berechnung der AbsorptionsWSLK in Matrix-Vektor-Form

$\underline{\underline{S}} = [q_{ij}] \quad i,j\in T$

$\underline{\underline{R}} = [q_{ij}] \quad i\in T, j\in C$

$\underline{a} = [a_i] \quad i\in T$

$\underline{a} = -\underline{\underline{S}}^{-1}\underline{\underline{R}}\,\underline{e}$

mittlere Zeit bis zur Absorption →

Alternative Matrix/Vektor Form

$t_{ii} = 0 \quad i\in C$

$0 = 1 + \sum q_{ij} t_j$

$\underline{t} = -\underline{\underline{S}}^{-1}\underline{e}$

Ergänzungen

- stationäre Verteilung $\underline{p}\,\underline{\underline{Q}} = \underline{0}$ und $\underline{p}\,\underline{e} = 1$
- Eine irreduzible und positiv rekurrente CTMC hat immer eine Grenzverteilung $(\underline{p}^{\infty})$, welche genau gleich wie $\underline{p}$ ist.
- Eine irreduzible und endliche CTMC hat immer eine Grenzverteilung $(\underline{p}^{\infty})$, bei der $\underline{p}^{\infty} = \underline{p}$ gilt.

⇝ Bei der Prüfung ist es nicht nötig Beweise zu können

Informationstheorie

Wie kann man Information messen?

Nachricht: Kette von "Buchstaben" die zufällig erzeugt wird

Je kleiner die WSLK, desto größer der Informationsgehalt.

Wahrscheinlichkeit p

Informationsgehalt $\log_2 \frac{1}{p}$ (Basiswechsel indem man $\log$ mit Basis multipliziert)

Napier (nepbits) $\ln \frac{1}{p}$

BSP: "Draußen schneit es"

Nordpol → wenig Information

Surinam → viel Information

Informationsgehalt abh. von WSLK

BSP: Frage – Antwort "Bar-Kochba" (Ich-Seh-ich-Seh-was-du-nicht-siehst)

m Objekte in Menge M zusammenfassen $M = \{1, \dots m\} / \{0 \dots m-1\}$

darf ja/nein Fragen stellen, mit möglichst wenig Fragen ein Objekt identifizieren.

l Fragen: $\begin{cases} \text{jedes ja} = 1 \\ \text{jedes nein} = 1 \end{cases}$ höchstens 2^l versch. Folgen von Antworten. Eindeutigkeit $2^l \geq m$

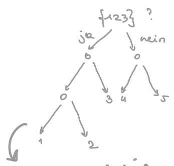
⇒ Optimale Anzahl von Fragen: $\lceil \log_2 m \rceil$

BSP: Element von M wird zufällig bestimmt.

$M = \{1, 2, 3, 4, 5\}$

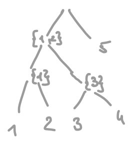
Fragen sind Teilmengen von M

Fragebaum 1



Wahrscheinlichkeiten halbieren

Fragebaum 2



"je nachdem" ist nicht notwendig

2 in Fragestrategie verwandeln

1: 111
2: 110
3: 101
4: 100
5: 0

Fragestrategie mit Bäumen identifizieren → Code: für $\begin{cases} \text{ja} = 1 \\ \text{nein} = 0 \end{cases}$ Zahlen aneinanderhängen

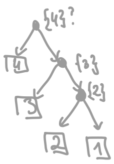
Blattlänge: Anzahl der Fragen für ein Element = Codewortlänge

Mittlere Anzahl an Fragen: $\bar{l} = \sum p_i \cdot l_i$ (Für Element i gibt es l_i Fragen)

BSP: $m = 4$, $P = (0.1, 0.2, 0.3, 0.4)$

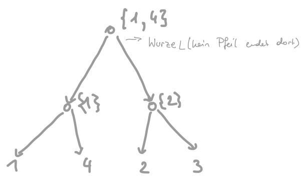
Element m wird zufällig gewählt mit Verteilung $P = (p_1 \dots p_n)$

Optimaler Wert für $\bar{l}$?



$\bar{l} = 0.4 \cdot 1 + 0.3 \cdot 2 + 3(0.2 + 0.1) = 1.9$

→ nicht soooo toll



WH Baum: gerichteter Graph ohne Zyklus

Def: jeder Knoten höchstens 2 Kanten

Wurzel: keine zuführende Kante

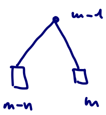
Vollständig: aus jedem Knoten im inneren 2 Kanten

Blatt: keine neue Kante

Blattlänge: Σ Weg

Das geht besser: Baum mit Blattlänge $l_1 \dots l_m$ → WSLK darauf aufteilen

Tiefste Ebene



alter Baum:

$p_1 \geq \dots \geq p_m$

$p_1 \dots p_m$

$l_1 \dots l_m$

neuer Baum:

$p_1^* \dots p_{m-1}^*$ $\quad p_{m-1}^* = p_{m-1} + p_m \quad p_1^* = p_1, \dots p_{m-2}^* = p_{m-2}$

$l_1^* \dots l_{m-1}^* \quad l_{m-1} = l_m$

$l_1^* = l_1 \dots l_{m-2}^* = l_{m-2} \quad l_{m-1}^* = l_{m-1} - 1 = l_m - 1$

mittlere Blattlänge $\bar{l} = p_1 l_1 + \dots + p_{m-1} l_{m-1} + p_m l_m = p_1 l_1^* + \dots + p_{m-2} l_{m-2}^* + p_{m-1}(l_{m-1}^* + 1) + p_m(l_{m-1}^* + 1) = l^* + p_m + p_{m-1}$

Huffman-Algorithmus: Optimaler Code / Fragestrategie für ($P = p_1 \dots p_m$)

(0) Wenn $m = 1$: $\boxed{\alpha}$ break

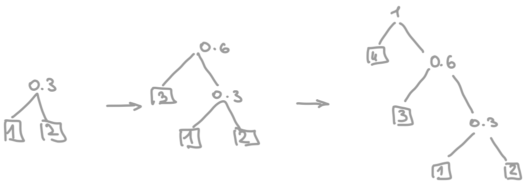
rekursiv: (1) ordne die Wahrscheinlichkeiten absteigend $p_1 \geq \dots \geq p_m$

(2) Fasse die beiden kleinsten WSLK zsm $P^* = (p_1^* \dots p_{m-1}^*) = (p_1 \dots p_{m-2}, p_{m-1} + p_m)$

(3) konstruiere den optimalen Baum (optimale Fragestrategie für P^*)

(4) Ersetze in kürzeren optimalen Baum aus 3 $m-1 \to$ $\boxed{m-1}$ $\boxed{m}$

BSP $(0.1, 0.2, 0.3, 0.4)$
$\quad 1 \quad 2 \quad 3 \quad 4$



$H^*(P) = (0.1, 0.2, 0.3, 0.4) = 1.9$

mittlere Unbestimmtheit: $H^*(P)$ Optimaler Wert für die mittlere Anzahl von Fragen

Abschätzung: Minimierung von $\bar{L} = \sum_{i=1}^{m} p_i l_i$ unter der Nebenbedingung "$\exists$ Baum mit Blattlängen $l_1 \dots l_n$"

Ungleichung von Kraft:

$\exists$ Binärbaum mit Blattlängen $l_1 \dots l_m$ genau dann, wenn $\sum_{i=1}^{m} 2^{-l_i} \leq 1$

Gleichheit gilt, wenn der Baum vollständig ist.

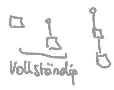
minimiere $\sum p_i l_i$ unter NB $\sum 2^{-l_i} = 1 \Rightarrow \sum p_i l_i - \lambda \sum 2^{-l_i}$

$$\frac{\partial}{\partial l_i} = 0 \Rightarrow p_i - \lambda \cdot 2^{-l_i} \cdot (-\log 2) = 0 \qquad c = \frac{1}{(-\lambda \cdot \log 2)}$$

$$2^{-l_i} = c \cdot p_i$$

$$\underbrace{\sum 2^{-l_i}}_{1} = \underbrace{c}_{\downarrow\, 1 \leftarrow} \underbrace{\sum p_i}_{1} \qquad p_i = 2^{-l_i} \qquad l_i = \log\left(\frac{1}{p_i}\right)$$

Beweis $m = 1$



Vollständig $\to$ nicht vollständig

$l_1 = 0 \quad l_m = 1 \quad l_2 = 2$

Höhe h des Baumes $\max_{i \leq m}(l_i)$

$h+1$ Höhe $h+1$ Höhe h

$l_1 \leq l_2 \leq \dots l_m$

entweder ist $\sum 2^{-l_i} \leq \frac{1}{2}$ $\quad l_i^* = l_i - 1 \to \sum 2^{l_i^*} \leq 1$

oder es gibt $k < m$ mit $l_i^* = l_i - 1$: $\sum_{k=1}^{h} 2^{-l_i} = \frac{1}{2}$

$2^{-l_1} + \dots 2^{-l_k} \geq 1$ mit kleinstem k

$2^{l_k - l_1} + \dots - 2^{l_k - l_{k-1}} + 1 \geq 2^{l_k - 1}$

Entropie in P^λ: $H(p_1 \dots p_m) = \sum_{i=1}^{m} p_i \cdot \log_2\left(\frac{1}{p_i}\right) \leq H^*(p_1 \dots p_m)$

weiters: $H^*(p_1 \dots p_n) \leq H(p_1 \dots p_n) + 1$

weiters: $l_i = \log_2 \left\lceil \frac{1}{p_i} \right\rceil \geq \log_2 \frac{1}{p_i} \Rightarrow \sum 2^{-l_i} \leq \sum 2^{-\log_2 \frac{1}{p_i}} = \sum p_i = 1$

$\hookrightarrow H^*(p_1 \dots p_n) \leq \sum p_i \left\lceil \log_2\left(\frac{1}{p_i}\right) \right\rceil \leq \sum p_i \left(\log_2 \frac{1}{p_i} + 1\right) = H(p_1 \dots p_m) + 1$

$H(x \dots x_n) = n \cdot H(x) \Leftrightarrow x_1 \dots x_n$ sind unabhängig

$n \cdot H(x) \leq H^*(x_1 \dots x_n) \leq n H(x_1) + 1$

pro Element: $H(x) \leq \frac{H^*(\vec{x})}{n} \leq H(x) + \frac{1}{n}$

Vorstellung (nicht ganz korrekt aber funktioniert) $H(P)$ ist die optimale Anzahl von Fragen die optimale Strategie hat $l_i = \log_2\left(\frac{1}{p_i}\right)$

"Dumme Geschichte" $\quad x \sim P$ in Wirklichkeit aber wir glauben $x \sim \underbrace{Q - (q_1 \dots q_m)}_{\text{Fairer Würfel}}$

$\downarrow$ gezinkter Würfel

optimal : $\sum p_i \log_2\left(\frac{1}{p_i}\right) \Rightarrow H(P)$

wirklich: $\sum p_i \log_2\left(\frac{1}{q_i}\right) \Rightarrow$ nicht optimal $> H(P)$

Strafe des Irrtums / Relative Entropie / Diskrepanz / Kullback-Leibler Distanz / I-Divergenz $\quad D(P, Q) = \sum p_i \log_2\left(\frac{p_i}{q_i}\right)$

$D(P, Q) \geq 0$ mit Gleichheit für $P = Q$

Beweis log konkav: $\log_2(\lambda x + (1-\lambda) y) \geq \lambda \log_2(x) + (1-\lambda) \log_2(y)$

$\log_2\left(\sum \lambda_i x_i\right) \geq \sum \lambda_i \log(x_i) \qquad \sum \lambda_i = 1, \lambda_i \geq 0$

$-D(PQ) = \sum p_i \log\left(\frac{q_i}{p_i}\right) \leq \log\left(\sum \underbrace{x_i \frac{q_i}{p_i}}_{=1}\right) = 0$

Eigenschaften von $H(p)$

$0 \leq H(P) \leq \log_2(m)$

$p \cdot \log_2\left(\frac{1}{2}\right)$ für $p = 0$ soll 0 sein

$H(x) = 0$ wenn X deterministisch ist

$H(P) \leq \log_2(m)$ $Q\left(\frac{1}{m}, \dots, \frac{1}{m}\right)$

$H(P) \leq \sum_{i=1}^{m} p_i \log_2\left(\frac{1}{p_i}\right) = \sum_{i=1}^{m} p_i \log_2(m)$

$\boxed{\text{BSP}}$ 2x Würfeln, X kleinere Augenzahl, Y größere Augenzahl

$X \backslash Y$	1	2	3	4	5	6	
1	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{11}{36}$
2	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{9}{36}$
3	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{7}{36}$
4	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{5}{36}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$
	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	

Frage nach X ($H(x)$) und Y ($H(y)$) getrennt

insgesamt $H(x) + H(y)$

$\frac{1}{36} \cdot \log_2(36) + \frac{3}{36} \cdot \log_2\left(\frac{36}{3}\right) + \dots = 2.31[illegible]5$

gemeinsam: $H(x,y) = \frac{5}{6} \underbrace{\log_2(18)}_{15 \cdot \frac{2}{36}} + \frac{1}{6} \log(36) = 4.33659$

Information zwischen X und Y: $I(x,y) = H(x) + H(y) = H(x,y)$

$$H(x,y) = -\sum_{i=1}\sum_{j=1} \mathbb{P}(X=x_i, Y=y_j) \cdot \log(\mathbb{P}(X=x_i, Y=y_j)) = -\sum_{ij} \mathbb{P}(X=x_i, Y=y_i) \log\left[\mathbb{P}(X=x_i)\,\mathbb{P}(Y=y_j \mid X=x)\right] =$$

$$= \underbrace{-\sum_i\sum_j \mathbb{P}(X=x_i, Y=y_j) \log \mathbb{P}(X=x_i)}_{H(X)} + \underbrace{\left(\sum_i \mathbb{P}(X_i=x_i) \sum_j -\mathbb{P}(Y=y_j \mid X=x_i) \log \mathbb{P}(Y=y_j \mid X=x_i)\right)}_{H(Y \mid X=x_i)}$$ Entropie der bedingten

$$= H(X) + \sum \mathbb{P}(X=x_i)\, H(Y \mid X=x_i) = H(x) + H(Y|X) = H(Y) + H(X|Y)$$

Prüfung: Entropien berechnen

sei $\mathbb{P}(Y = f(x_i)) = 1 \Rightarrow$ mit WSLK $= 1$ gilt $y = f(x)$

$I(x,y)$ im Beispiel $4.6\ldots - 4.3\ldots = 0.30331$

$H(x|y) = H(x,y) - H(y) = 2.01664$

$0 \leq I(x,y) \leq \min H(x,y)$

$0 = I(x,y)$ wenn x, y unabhängig sind

$I(x,y) = \min H(x,y)$ wenn $I(x,y) = H(x) \Leftrightarrow H(Y|X) = 0 \Leftrightarrow H(Y \mid X=x_i) = 0 \quad \forall i$

Vert. von Y unter $X = x_i$ ist deterministisch

Codes

Begriffe:

Quellalphabet $M = \{1 \dots m\}$ oder $\{0 \dots m-1\}$ oder $\{x_1 \dots x_m\}$

Codealphabet $C = \{0 \dots r-1\}$

Binärcode: $r = 2$

Alphabet A

Nachrichten der Länge n: $A^n = \{a_1 \dots a_n, a_i \in A, i = 1 \dots n\}$

$A^* = \bigcup_{n \geq 0}^{\infty} A^n \qquad A^0 = \{''\}$

$A^{**} = A^* \cup A^{\mathbb{N}}$

$(a_1, a_2, \dots a_n)$ wird zu $a_1 a_2 \dots a_n$

Code: Abbildung $c: M^{**} \to C^{**}$ Einschränken: für jeden Quellbuchstaben gibt es ein Codewort

Nachricht wird durch aneinanderhängen codiert

$c(x_1 x_2 \dots) = c(x_1) c(x_2) \dots$

$\boxed{\text{BSP}}$ $m = 3$, $r = 2$ $M = \{1, 2, 3\}$

c: $c(1) = 0$, $c(2) = 10$, $c(3) = 11$

$3121 \mapsto 110100 \xmapsto{-1} 3121$

Präfixfreier Code $\Leftrightarrow$ c ist fortlaufend entzifferbar

(beim sequentiellen lesen ist klar, wo ein Code endet)

(kein Codewort darf Anfang eines anderen Codeworts sein)

$c \dots$ eindeutig entzifferbar

$c \dots$ endlich eindeutig entzifferbar

•) $m = r = 2$ $c(1) = 0$, $c(2) = 01$

kann eindeutig entziffert werden, wenn man den nächsten Buchstaben auch kennt.

(geht mit entzifferten und unentzifferten Angaben)

•) $m = 3$ $r = 2$

$1 \to 0$
$2 \to 01$
$3 \to 11$

110010 rückwärts lesen
$3\ 1\ 2\ 1$

a) $1333 \to 01111$
b) $2333 \to 01111$ $\Big\}$ Oh shit, same!

Optimaler Code $\to$ Huffman Code aus Huffman Baum erzeugen

kanonischer Huffman Code

1) bestimme Codewortlänge aus Huffman Baum: $l_1 \dots l_n$ mit $l_1 \leq \dots \leq l_n$

2) erstes Codewort: nur 0er Binärzahl mit l_i stellen $C_{i+1} = (C_i + 1)\, 2^{e_{i+1} - e_i} = (C_i + 1) \overset{\text{shiften}}{\ll} (l_{i+1} - l_i)$

Aus Beweis der Entropie geht Ungleichung $l_i = \lceil \log_2 \frac{1}{p_i} \rceil$ hervor

Shannon Code $P = (0.1, 0.2, 0.3, 0.4)$ ordne WSK absteigend

x_i	p_i	$f_{i-1} = p_1 + \dots + p_{i+1}$	$l_i = \lceil \log_2 \frac{1}{p_i} \rceil$	$c_i = f_{i-1}$
4	0.4	0	2	00
3	0.3	0.4	2	01
2	0.2	0.7	3	101
1	0.1	0.9 $\hookrightarrow (0.1110\dots)_2$	4	1110

Fano Code mit $l_i = \lceil \log_2 \frac{1}{p_i} \rceil + 1$

	f_i	$\frac{f_i + f_{i-1}}{2}$	c_i
0.1	0.1	0.05	00001
0.2	0.3	0.2	0011
0.3	0.6	0.45	011
0.4	1.0	0.8	110