Einführung in die Stochastik für Informatiker

Dozent: Professor Mathar

Eine Vorlesungsmitschrift SS 2000

Thorsten Uthke

14. Juli 2000

Inhaltsverzeichnis

1 Einleitung

Betrachte "Zufallsexperimente", z.B.

- Münzwurf, Würfelwurf, Spiele, Roulette, Lotto
- Ankunft von Kunden an Schaltern, Pakete in Netzwerken
- Input für Algorithmen
- Signale, die von einer Quelle ausgesendet werden
- Positionierung von Mobilstationen in Zellnetzen

Gemeinsam in diesen Beispielen: Interessierende Größen können nicht genau vorhergesagt werden, sind zufallsabhängig
Stochastik: mathematische Behandlung von Zufallsphänomenen

<u>Teilgebiete:</u>

- Wahrscheinlichkeitstheorie
  - theoretisch (stochastische Prozesse, Grenzwertsätze, stochastische Diffentialgleichungen)
  - angewandt = stochastische Modellierung (Warteschlangen, Zuverlässigkeitstheorie, stochastische Signalerkennung)
- mathematische Statistik (mit Teilgebieten)

Hier: Wahrscheinlichkeitstheorie, angewandt, stochastische Modellierung, Betonung der Anwendung in der Informatik
<u>Ziel der Vorlesung:</u> Bereitstellung der Grundlagen, Basis für weiterführende Vorlesungen

> Warteschlangensysteme (I), Warteschlangennetze (II), Informationstheorie I,II, Kryptologie, stochastische Simulation, zufallsgesteuerte Optimierungsverfahren

Historiche Entwicklung

> Fermat (1601-65), Pascal (1623-62), Bernoulli (1654-1705)
> Laplace (1749-1827): Kombinatorischer Zugang, motiviert durch Spielprobleme, relative Häufigkeiten als Wahrscheinlichkeiten
> Kolmogoroff (1933): Axiomatische Entwicklung

2 σ-Algebren und W'keitsverteilungen

Mathematische Beschreibung von Zufallsexperimenten mit Mengen.
Betrachte relevante **Ergebnisse** und fasse sie zu einer Menge zusammen.
Ω: **Ergebnismenge** (oft $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}^n, \{0,1\}^{\mathbb{N}}$)
Ereignisse werden durch Teilmengen von Ω beschrieben.
Zunächst $\mathcal{P}(\Omega)$ als Ereignismenge.

W'keiten von Ereignissen durch Funktion $P : \mathcal{P}(\Omega) \to [0,1]$ mit den Eigenschaften (*):

- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $P(A^C) = 1 - P(A) \quad \forall A \in \mathcal{P}(\Omega)$
- $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in \mathcal{P}(\Omega)$ mit $A \cap B = \emptyset$

σ-Additivität:
$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad A_i \in \mathcal{P}(\Omega)$ mit A_i paarweise disjunkt [1]

Schrechweisen:
A^C: "A tritt nicht ein", $A \in \mathcal{P}(\Omega)$
$A \cup B$: "A oder B treten ein", $A, B \in \mathcal{P}(\Omega)$
$A \cap B$: "A und B treten ein", $A, B \in \mathcal{P}(\Omega)$

Wie erhält man W'keiten? Bei endlichem Ω durch Abzählen.

Definition 2.1 (Laplacescher W'begriff) Ω sei eine endliche Menge. $|A|$ bezeichnet die Mächtigkeit von $A \subseteq \Omega$. Durch $P(A) = \frac{|A|}{|\Omega|}$ für $A \in \mathcal{P}(\Omega)$ wird eine W'verteilung P auf $\mathcal{P}(\Omega)$ mit den Eigenschaften (*) definiert. P heißt **Laplace-Verteilung** oder **diskrete Gleichverteilung** über Ω.

Beispiel 2.2 (Binäre Suche) Geg. geordnetes Feld von $2^n - 1$ Elementen.
Schlüsselelement y vergleichbar mit den Elementen.
Problem: Ist y in dem Feld vorhanden und an welcher Stelle?
Algorithmus: Binary search
Stochastisches Modell:
$\Omega = \{0, 1, \ldots, 2^n - 1\}$, $w \in \Omega = \begin{cases} w \geq 1: & \text{gesuchtes Element an Stelle } w \\ w = 0: & y \text{ nicht vorhanden} \end{cases}$
Bestimme Ereignisse A_k: y wird im k-ten Schritt gefunden. $A_1 = \{2^{n-1}\}$,
$A_2 = \{2^{n-2}, 3 \cdot 2^{n-2}\}, \ldots, A_k = \{(2j-1)2^{n-k} | j = 1, \ldots, 2^{k-1}\}$ für $k = 1, \ldots, n$
Annahme: Jede Platznummer und 0 sind gleichwahrscheinlich. Dann $|A_k| = 2^{k-1}$,
$P(A_k) = \frac{2^{k-1}}{2^n}$ für $1 \leq k \leq n \implies$ W'keit, y in genau k Schritten zu finden

[1] d.h. $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Zusammengesetzte Ereignisse B_k: y wird in höchstens k Schritten gefunden
$B_k = \sum_{j=1}^{k} A_j$ mit $j = 1, \ldots, k$ und die A_j disjunkt
$P(B_k) = P(\bigcup_{j=1}^{k} A_j) = \sum_{j=1}^{k} P(A_j) = \sum_{j=1}^{k} \frac{2^{j-1}}{2^n} = \frac{1}{2^n}(2^k - 1) = \frac{2^k-1}{2^n}$

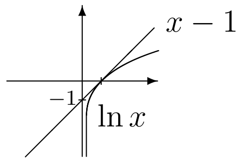
Komplizierte Anzahlbestimmungen durch kombinatorische Überlegungen.

Beispiel 2.3 (Hashing)
Universum U, $M \subseteq U$, $|M| = k$, M soll abgespeichert werden
Hashtafel: $a : array[0, \ldots, n-1]$ of ...
Hashfunktion: $h : U \to [0, \ldots, n-1]$ Einfache Funktion!
Speichere $x \in M$ in $a[h(x)]$
Kollision, falls $h(x) = h(y)$ für $x \neq y$ (Auflösung durch linerare Liste)

Stochastisches Modell:
(”Rein zufälliges” Ablegen von k Daten in Feld der Länge n)
$S = \{1, \ldots, n\}$ (Speicherplätze)
$\Omega = S^k$ ($k-n-$Permutationen mit Wiederholungen/Kollisionen)
$A_{k,n}$ ($k-n-$Permutationen ohne Wiederholungen/Kollisionen)

Wahrscheinlichkeit für kollisionsfreie Abspeicherung:
$P(A_{k,n}) = \frac{|A_{k,n}|}{|\Omega|} = \frac{n(n-1)\ldots(n-k+1)}{n^k} = \prod_{i=0}^{k-1}(1-\frac{i}{n}) = \exp(\ln(\prod_{i=0}^{k-1}(1-\frac{i}{n})))$
$= \exp(\sum_{i=0}^{k-1} \ln(1-\frac{i}{n}))$
$\overset{(*)}{\leq} \exp(-\sum_{i=0}^{k-1} \frac{i}{n})$
$= \exp(-\frac{(k-1)k}{2n})$

Benutzte Abschätzung $(*)$:
$\ln x \leq x - 1 \quad \forall x \geq 0$
$\Longrightarrow \ln(1-x) \leq -x \quad \forall x \leq 1$



Zahlenbeispiel: Für $n = 365$ und $k = 23$ ist $P(A_{k,n}) \leq 0,499998$

Der Laplacesche W'begriff reicht nicht aus, da

- Ω ist oft ∞ oder überabzählbar
- Viele Experimente sind nicht durch eine diskrete Gleichverteilung beschreibbar

Beispiel 2.4 (∞-Münzwurf) Kopf $\widehat{=}1$, Zahl $\widehat{=}0$
Bestimme ein Modell für nicht-abbrechende Folge von Münzwürfen.
(Wird gebraucht bei der Frage ”Wann tritt zum ersten Mal Kopf (Zahl) auf?”. Jeder beliebig ferne Wurf könnte der erste sein.).

Stochastisches Modell:
$\Omega = \{w = (x_1, x_2, \ldots) | x_i \in \{0,1\}\} = \{0,1\}^{\mathbb{N}}$ Menge der 0-1-Folgen
Bekannt: Ω ist überabzählbar.
Problem: Beschreibe Gleichverteilung auf Ω
Ansatz: $P(\{w\}) = \delta > 0$

Sei $A = \{w_1, w_2, \ldots\} \subset \Omega$ unendlich, aber abzählbar. Dann $P(A) = \sum_{i=1}^{\infty} P(\{w_i\})$
$= \sum_{i=1}^{\infty} \delta = \infty$ Wiederspruch zu $P(\Omega) = 1$.
Also: $P(\{w\}) = 0 \quad \forall w \in \Omega$ Wenig hilfreich!

Beispiel 2.5 (Gleichverteilung über $[0,1]$**)** (z.B. jede reelle Zahl zwischen 0 und 1 ist gleich wahrscheinlich)
$\Omega = [0,1]$ (Ergebnismenge)
Versuch: Ereignismenge $= \mathcal{P}(\Omega)$
Forderungen:

1) $P([a,b]) = b - a \quad \forall 0 \leq a < b \leq 1$
2) P ist σ-additiv

Es gilt: Eine Funktion P mit diesen Eigenschaften existiert nicht. Im $\mathbb{R}^3$ gibt es sogar keine, wenn nur endliche Additivität verlangt wird: Es gibt kein endlich additives drehinvariantes[2] Maß auf $\mathcal{P}(\mathbb{R}^3)$. (Hausdorff 1914)

Für einen allgemeinen W'begriff (ohne Existenzprobleme):
Ereignismenge nicht $\mathcal{P}(\Omega)$, sondern kleineres Mengensystem wählen, das noch alle interessanten Ereignisse enthält.

Definition 2.6 Sei $\Omega \neq \emptyset$, $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ ein System von Teilmengen von Ω. $\mathfrak{A}$ heißt **σ-Algebra** (von Ereignissen) über Ω, wenn

(i) $\Omega \in \mathfrak{A}$
(ii) $A \in \mathfrak{A} \implies A^C \in \mathfrak{A}$
(iii) $A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N} \implies \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$

$(\Omega, \mathfrak{A})$ heißt **Meßraum**.

Mit den De Morgan-Regeln: $A_n \in \mathfrak{A} \overset{ii)}{\implies} A_n^C \in \mathfrak{A} \overset{iii)}{\implies} \bigcup_{n=1}^{\infty} A_n^C \in \mathfrak{A}$
$\overset{ii)}{\implies} (\bigcup_{n=1}^{\infty} A_n^C)^C \in \mathfrak{A} \implies \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N}$
σ-Algebren enthalten alle Ereignisse, die durch Verknüpfungen (abzählbar vieler Ereignisse) mit "und", "oder" und "nicht" entstehen. Dies ist wichtig für die Festlegung von W'verteilungen.

Beispiel 2.7

a) $\mathcal{P}(\Omega)$ ist stets eine σ-Algebra (die feinste σ-Algebra)
b) $\{\emptyset, \Omega\}$ ist eine σ-Algebra (die gröbste σ-Algebra)
c) $\Omega = \mathbb{N}$, $G = \{2,4,6,\ldots\}$, $U = \{1,3,5,\ldots\}$, $\mathfrak{A} = \{\emptyset, G, U, \Omega\}$ ist σ-Algebra

[2]Wenn man im $\mathbb{R}^3$ einen Körper dreht, muss er derselbe bleiben.

d) $\Omega = \mathbb{R}$, $E = \{(a,b] | a < b \text{ mit } a, b \in \mathbb{R}\}$ ist keine σ-Algebra
Denn sei $a < b < c < d$. $(a,b] \in E$ und $(c,d] \in E$, aber $(a,b] \cup (c,d] \notin E$.

Problem: Gibt es eine kleinste (im Sinne der Mengeninklusion) σ-Algebra, die E enthält?

Lemma 2.8 $\Omega \neq \emptyset$, $\mathfrak{A}_i$ sei σ-Algebra über Ω für alle $i \in I$. Dann ist $\bigcap_{i \in I} \mathfrak{A}_i$ ebenfalls eine σ-Algebra. (Beweis in der Übung 2)

Sei $E \subseteq \mathcal{P}(\Omega)$. $\mathfrak{A}(E) = \bigcap\limits_{\substack{\mathfrak{A} \text{ ist } \sigma\text{-Algebra} \\ \text{und } E \subseteq \mathfrak{A}}} \mathfrak{A}$ ist die kleinste σ-Algebra, die E enthält, ist die **von E erzeugte σ-Algebra**.

Definition 2.9 Sei $\Omega = \mathbb{R}$, $E = \{(a,b] | a < b \text{ mit } a, b \in \mathbb{R}\}$.
$\mathfrak{B} = \mathfrak{A}(E)$ heißt **Borelsche σ-Algebra**.

Auf σ-Algebren können W'verteilungen mit den Eigenschaften (*) (Seite 4) definiert werden.

Definition 2.10 $\Omega \neq \emptyset$, $\mathfrak{A}$ eine σ-Algebra. Eine Abbildung $P : \mathfrak{A} \to [0,1]$ mit

(i) $P(\Omega) = 1$
(ii) $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \quad \forall A_n \in \mathfrak{A}$, paarweise disjunkt

heißt **Wahrscheinlichkeitsverteilung** oder **Wahrscheinlichkeitsmaß** auf $(\Omega, \mathfrak{A})$. $(\Omega, \mathfrak{A}, P)$ heißt **Wahrscheinlichkeitsraum**.

Der Laplacesche W'begriff (Def. 2.1) ist ein Spezialfall von Def. 2.10. Denn für $\mathfrak{A} = \mathcal{P}(\Omega)$ erfüllt $P(A) = \frac{|A|}{|\Omega|}$

(i) $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$
(ii) $P(\bigcup\limits_{n=1}^{\infty} A_n) = \frac{\left|\bigcup\limits_{n=1}^{\infty} A_n\right|}{|\Omega|} = \frac{\sum\limits_{n=1}^{\infty} |A_n|}{|\Omega|} = \sum\limits_{n=1}^{\infty} \frac{|A_n|}{|\Omega|} = \sum\limits_{n=1}^{\infty} P(A_n)$
$\forall A_n$ paarweise disjunkt

Sei $(\Omega, \mathfrak{A}, P)$ W'raum. Eine Folge von Ereignissen $\{A_n\}_{n \in \mathbb{N}}$ heißt **aufsteigend** (**absteigend**), wenn $A_n \subseteq A_{n+1}$ ($A_{n+1} \subseteq A_n$) für alle $n \in \mathbb{N}$.
$\bigcup\limits_{n=1}^{\infty} A_n = \lim\limits_{n \to \infty} A_n$ ($\bigcap\limits_{n=1}^{\infty} A_n = \lim\limits_{n \to \infty} A_n$) heißt **Limes** der Mengenfolge $\{A_n\}$.

Lemma 2.11 (Eigenschaften von W'verteilungen) $(\Omega, \mathfrak{A}, , P)$ sei W'raum. $A, B, A_n \in \mathfrak{A} \quad \forall n \in \mathbb{N}$.

a) $P(A^C) = 1 - P(A)$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

c) $P(A) \leq P(B)$ falls $A \subseteq B$

d) $P(B \setminus A) = P(B) - P(A)$ falls $A \subseteq B$

e) $\{A_n\}$ aufsteigend $\Longrightarrow$ $P(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ (Stetigkeit von unten)
$\{A_n\}$ absteigend $\Longrightarrow$ $P(\lim\limits_{n\to\infty} A_n) = \lim\limits_{n\to\infty} P(A_n)$ (Stetigkeit von oben)

Beweis: a)-d) in der 2. Übung
e) 1.) $\{A_n\}$ aufsteigend, d.h. $A_n \subseteq A_{n+1} \quad \forall n \in \mathbb{N}$
Setze $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $\ldots$, $B_{n+1} = A_{n+1} \setminus A_n$
Dann gilt B_n p.d. (paarweise disjunkt) und $\bigcup\limits_{n=1}^{\infty} A_n = \bigcup\limits_{n=1}^{\infty} B_n$. Setze $A_0 = \emptyset$,
dann ist $P(\bigcup\limits_{n=1}^{\infty} A_n) = P(\bigcup\limits_{n=1}^{\infty} B_n) \overset{ii)}{=} \sum\limits_{n=1}^{\infty} P(B_n) = \lim\limits_{k\to\infty} \sum\limits_{n=1}^{k} P(A_n \setminus A_{n-1})$
$= \lim\limits_{k\to\infty} \sum\limits_{n=1}^{k} [P(A_n) - P(A_{n-1})] = \lim\limits_{k\to\infty} P(A_k)$
2.) $\{A_n\}$ absteigend $\Longrightarrow$ $\{A_n^C\}$ aufsteigend. Es gilt:
$P(\bigcap\limits_{n=1}^{\infty} A_n) = 1 - P(\bigcup\limits_{n=1}^{\infty} A_n^C) \overset{1.)}{=} 1 - \lim\limits_{n\to\infty} P(A_n^C) = \lim\limits_{n\to\infty} (1 - P(A_n^C))$
$= \lim\limits_{n\to\infty} P(A_n)$

Lemma 2.12 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$.

a) **Siebformel** von Pomcare-Sylvester (inclusion-exclusion-formula)
$P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i) - \sum\limits_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum\limits_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \ldots + \ldots + (-1)^{n+1} P(\bigcup\limits_{i=1}^{n} A_i)$

b) **Bonferroni-Ungleichung** (Bonferroni inequality)
$\sum\limits_{i=1}^{n} P(A_i) - \sum\limits_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) \leq P(\bigcup\limits_{i=1}^{n} A_i) \leq \sum\limits_{i=1}^{n} P(A_i)$
Weitere obere bzw. untere Schranken durch Abbruch in a) nach + oder - Zeichen

Beispiel 2.13 (Sortieren) (Recourtre-Problem)
Betrachte Feld der Länge n von vergleichbaren, verschiedenen Elementen. Alle Anordnungen seien gleichwahrscheinlich.
Stoch. Modell: $\Omega = \{\text{Permutationen von } \{1, \ldots, n\}\}$, $\mathfrak{A} = \mathcal{P}(\Omega)$
$P(\{w\}) = \frac{1}{|\Omega|} = \frac{1}{n!} \quad \forall w \in \Omega$

a) Bestimme die W'keit, daß mindestens ein Element an der richtigen Stelle ist (vorsortiert).
$A_j = \{w = (w_1, \ldots, w_n) \in \Omega | w_j = j\}$ j-tes Element an der richtigen Stelle

Gesucht: $P(A_1 \cup \ldots \cup A_n) = P(\bigcup_{i=1}^n A_i)$
Berechnung mit Hilfe der Siebformel: $1 \leq i_1 < i_2 \leq n$, $l \leq n$
$\bigcap\limits_{j=1}^{l} A_{i_j} = \{w \in \Omega | w_{i_j} = i_j,\ j = 1, \ldots, l\}$, $|\bigcap\limits_{j=1}^{l} A_{i_j}| = (n-l)!$
Also $P(\bigcap\limits_{j=1}^{l} A_{i_j}) = \frac{(n-l)!}{n!} = \frac{1}{\binom{n}{l} l!}$ mit $l = 1, \ldots, n$
Außerdem: $|\{(i_1, \ldots, i_l) | 1 \leq i_1 < \ldots < i_l \leq n\}| = \binom{n}{l}$
Insgesamt: $P(\bigcup\limits_{j=1}^{n} A_j) = \sum\limits_{j=1}^{n} P(A_i) - \sum\limits_{i<j} P(A_i \cap A_j) + \ldots + (-1)^{n+1} P(\bigcap\limits_{j=1} A_j)$
$= \frac{n}{n} - \binom{n}{2} \frac{1}{\binom{n}{2} 2!} + \binom{n}{3} \frac{1}{\binom{n}{3} 3!} - \ldots + (-1)^{n+1} \binom{n}{n} \frac{1}{\binom{n}{n} n!}$
$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + \frac{(-1)^{n+1}}{n!}$
$= 1 - (1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!}) \overset{n \to \infty}{\to} 1 - e^{-1} \approx 0,6321$
n groß: W'keit für mindestens ein vorsortiertes Element ist $0,6321$ (fast unabhängig von n)

b) Bestimme die W'keit, daß mindestens k Elemente vorsortiert sind.
$P(\bigcup\limits_{1 \leq i_i < \ldots < i_k \leq n} \bigcap\limits_{l=1}^{k} A_{i_l}) \overset{Lemma\ 2.12b)}{\leq} \sum\limits_{1 \leq i_i < \ldots < i_k \leq n} P(\bigcap\limits_{l=1}^{k} A_{i_l}) = \binom{n}{k} \frac{1}{\binom{n}{k} k!} = \frac{1}{k!}$
fällt sehr schnell mit steigendem k

c) Bestimme die W'keit, daß genau k Elemente vorsortiert sind.
Nach a): W'keit, daß in einem Feld der Länge $n-k$ kein Element vorsortiert ist: $1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!}$
$\Longrightarrow$ Anzahl der Anordnungen, beim denen kein Element vorsortiert ist: $(n-k)!(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!})$
Anzahl der Möglichkeiten, ein Feld der Länge n in ein Teilfeld der Länge $n-k$ und eines der Länge k aufzuteilen: $\binom{n}{k}$
Insgesamt: W'keit, daß genau k Elemente vorsortiert sind:
$\frac{1}{n!}\binom{n}{k}(n-k)!(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!}) = \frac{1}{k!}(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + \frac{(-1)^{n-k}}{(n-k)!})$

Wie legt man W'keiten fest, wenn schon bekannt ist, daß das Ergebnis in einer bestimmten Teilmenge liegt?

Motivierendes Beispiel 5000 Chips von zwei Firmen. Ω mit $|\Omega| = 5000$.
A: Chips der Firma A, mit $|A| = 1000$
B: Chips der Firma B, mit $|B| = 4000$
D: Defekte Chips, mit $|D| = 300$
Gelte: $|A \cap D| = 100$ (10% defekt) und $|B \cap D| = 200$ (5% defekt)
Ziehe zufällig einen Chip (Laplace-Modell).
W'keit, daß der Chip defekt ist, wenn er von Firma A stammt:
$P(D|A) = \frac{|D \cap A|}{|A|} = \frac{|D \cap A|/\Omega}{|A|/\Omega} = \frac{P(D \cap A)}{P(A)} = \frac{100/5000}{1000/5000} = 0,1$
Auch: $P(A|D) = \frac{|A \cap D|}{|D|} = \frac{|A \cap D|/\Omega}{|D|/\Omega} = \frac{P(A \cap D)}{P(D)} = \frac{100/5000}{300/5000} = \frac{1}{3}$

Definition 2.14 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A, B \in \mathfrak{A}$, $P(B) > 0$. $P(A|B) = \frac{P(A \cap B)}{P(B)}$ heißt **(elementare) bedingte Wahrscheinlichkeit von** A **unter (der Hypothese)** B. Durch $P(A^3|B) : \mathfrak{A} \to [0,1]$ mit $A \mapsto P(A|B)$ wird eine W'verteilung auf $\mathfrak{A}$ definiert, die (elementare) **bedingte Verteilung unter** B.

Satz 2.15 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $B_n \in \mathfrak{A}$ für $n \in \mathbb{N}$ sei Partition von Ω, d.h. $\bigcup_{n=1}^{\infty} B_n = \Omega$ und B_n paarweise disjunkt.

a) **Satz von der totalen W'keit** Für alle $A \in \mathfrak{A}$ gilt:
$P(A) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$ (Konvention: "Nicht definiert"$\cdot 0 = 0$, z.B. $\frac{x}{0} \cdot 0 = 0$)

b) **Bayes-Formel** Falls $P(A) > 0$, gilt für alle $n \in \mathbb{N}$
$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}$

Beweis:

a) $P(A) = P(\dot{\bigcup}_{n=1}^{\infty} A \cap B_n) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$

b) $P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}$

Wichtiger Spezialfall: $P(A|B) = P(A|B^C)$, wobei $P(B) > 0$, $P(B^C) > 0$ $(*)$
(W'keit für das Eintreten von A hängt nicht vom Eintreten von B ab.)
Dann gilt: $P(A) = P(A \cap B) + P(A \cap B^C) = P(A|B)P(B) + P(A|B^C)P(B^C) \overset{(*)}{=} P(A|B)(P(B) + P(B^C)) = P(A|B) = \frac{P(A \cap B)}{P(B)}$
Also: $P(A) = \frac{P(A \cap B)}{P(B)} \iff P(A)P(B) = P(A \cap B)$
Diese Definition wird erweitert auf n Ereignisse $A_1, \ldots, A_n \in \mathfrak{A}$, bzw. auf Folgen von Ereignissen.

Definition 2.16 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$ Ereignisse. $A_1, \ldots, A_n$ heißen **(gemeinsam) stochastisch unabhängig**, wenn $P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$ für alle $1 \leq i_1 < \ldots < i_k \leq n$ und für alle k.
Eine Folge $\{A_n\}_{n \in \mathbb{N}}$ von Ereignissen heißt stochastisch unabhängig, wenn die $A_1, \ldots, A_n$ für alle $n \in \mathbb{N}$ stochastisch unabhängig sind.

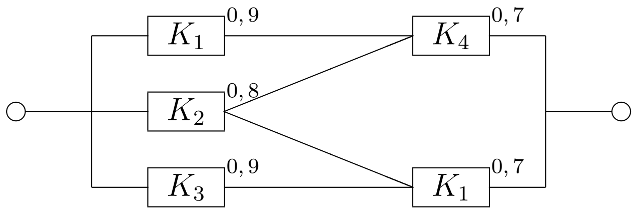
Beachte: Aus paarweiser stochastischer Unabhängigkeit folgt nicht die (gemeinsame) stochastische Unabhängigkeit!

Lemma 2.17 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_1, \ldots, A_n \in \mathfrak{A}$.
Es gilt: $A_1, \ldots, A_n \in \mathfrak{A}$ stochastisch unabhängig
$\iff \forall B_i \in \{A_i, A_i^C\}$, $i = 1, \ldots, n$: $P(\bigcap_{i=1}^{n} B_i) = \prod_{i=1}^{n} P(B_i)$
Es gilt: $A_1, \ldots, A_n \in \mathfrak{A}$ stochastisch unabhängig
$\iff \forall B_i \in \{A_i, A_i^C\}$, $i = 1, \ldots, n$: $B_1, \ldots, B_n$ stochastisch unabhängig

[3] A variabel in $\mathfrak{A}$

Beweis: Stochastik für Informatiker (Mathar & Pfeiffer)

Beispiel 2.18 System bestehe aus 5 Komponenten:



System ist intakt, wenn mindestens ein "Pfad" intakt ist.
W'keit, daß Komponete i intakt ist: p_i für $i = 1, \ldots, 5$ (Werte in Skizze)
$\{$Komponente K_i intakt$\}$ seien stochastisch unabhängige Ereignisse.

Modellbildung: $\Omega = \{(x_1, \ldots, x_5) | x_i \in \{0, 1\}\}$ ($x_i = 1$: Komponente i intakt)
$A_i = \{(x_1, \ldots, x_5)|x_i = 1\}$, $P(A_i) = p_i$ (Komponente i intakt)
$A_1, \ldots, A_5$ stochastisch unabhängig
$S = (A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5)$ (System intakt)
$P(S) = P((A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_2 \cap A_5) \cup (A_3 \cap A_5)$ Vereinigung nicht disjunkter Mengen $\implies$ Berechnung mit Sylvesterformel möglich, aber sehr aufwendig!

Mit Satz 2.15 a):
$P(S) = \underbrace{P(S|A_2)P(A_2)}_{(*1)} + \underbrace{P(S|A_2^C)}_{(*2)} P(A_2^C)$

$(*1)$: $P(S|A_2) = P(A_4 \cup A_5) = 1 - P(A_4^C \cap A_5^C) = 1 - P(A_4^C)P(A_5^C)$
$= 1 - (1 - p_4)(1 - p_5)$
$(*2)$: $P(S|A_2^C) = P((A_1 \cap A_4) \cup (A_3 \cap A_5)) = 1 - P((A_1 \cap A_4)^C \cap (A_3 \cap A_5)^C)$
$= 1 - (1 - p_1p_4)(1 - p_3p_5)$

Insgesamt: $P(S) = [1 - (1 - p_4)(1 - p_5)]p_2 + [1 - (1 - p_1p_4)(1 - p_3p_5)](1 - p_2)$
$= \ldots = 1 - p_2(1 - p_4)(1 - p_5) - (1 - p_2)(1 - p_1p_4)(1 - p_3p_5) = 0,90062$

Mit Satz 2.15 b): (W'keit, daß Komponente 2 intakt, wenn System intakt)
$P(A_2|S) = \frac{P(S|A_2)P(A_2)}{P(S)} = \frac{0,91 \cdot 0,8}{0,90062} = 0,80833$

Betrachte Limites von Mengenfolgen die nicht notwendig auf- oder absteigend sind.

Beispiel 2.19 (unendlicher Münzwurf) $\Omega = \{w = (x_1, x_2, \ldots)|x_i \in \{0, 1\}\}$
$A_n = \{w = (x_1, x_2, \ldots)|x_n = 1\}$ (im n-ten Wurf fällt Kopf)
Beschreibe das Ereignis A: es fällt ∞-oft Kopf = es treten ∞ viele der A_n ein
$A = \{w|w \in A_n$ für ∞ viele $n\} = \{w|\forall k \exists n \geq k : w \in A_n\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$
Analog: B: Fast alle der A_n treten ein (alle bis auf endlich viele Ausnahmen)
$B = \{w|w \in A_n$ für fast alle $n\} = \{w|\exists k \forall n \geq k : w \in A_n\} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$

Definition 2.20 $(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_n \in \mathfrak{A}$, $n \in \mathbb{N}$

$\limsup\limits_{n\to\infty} A_n = \lim\limits_{n\to\infty} \left(\bigcup\limits_{k=n}^{\infty} A_k \right) = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k$ heißt **Limes superior** der $\{A_n\}$

Der Limes superior einer Ereignisfolge $\{A_n\}$ ist die Menge der $w \in \Omega$, die in unendlich vielen der A_n liegen; d.h., daß unendlich viele der A_n eintreten.

$\liminf\limits_{n\to\infty} A_n = \lim\limits_{n\to\infty} \left(\bigcap\limits_{k=n}^{\infty} A_k \right) = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{k=n}^{\infty} A_k$ heißt **Limes inferior** der $\{A_n\}$

Der Limes inferior einer Ereignisfolge $\{A_n\}$ ist die Menge der $w \in \Omega$, die in fast allen der A_n liegen; d.h., daß fast alle der A_n eintreten.

Satz 2.21 (Borel-Cantelli-Lemma, auch 0-1-Gesetz)
$(\Omega, \mathfrak{A}, P)$ sei W'raum. $A_n \in \mathfrak{A}$, $n \in \mathbb{N}$

a) $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \quad \Longrightarrow \quad P(\limsup\limits_{n\to\infty} A_n) = 0$

b) $\sum\limits_{n=1}^{\infty} P(A_n) = \infty \quad \Longrightarrow \quad P(\limsup\limits_{n\to\infty} A_n) = 1$, falls $\{A_n\}_{n\in\mathbb{N}}$ stoch. unabh.

Beweis:

a) $P(\sum\limits_{n=k}^{\infty} A_n) \leq \sum\limits_{n=k}^{\infty} P(A_n) \to 0$ (für $k \to \infty$)

$$P(\limsup_{n\to\infty} A_n) = P(\bigcap_{k=1}^{\infty} \underbrace{\bigcup_{n=k}^{\infty} A_n}_{B_k,\ absteigend}) \overset{Lemma\ 2.11e}{=} \lim_{k\to\infty} P(\sum_{n=k}^{\infty} A_n) = 0$$

b) Rückseite von Übung 3

Beispiel 2.22 ∞-Würfelwurf, unabhängige Würfe,
$\Omega = \{w = (w_1, w_2, \ldots) | w_i \in \{1, \ldots, 6\}\}$
W'keit, daß ∞-oft die Sequenz $1, 2, \ldots, 6$ fällt?
Setze $A_n = \{w | w_n = 1, w_{n+1} = 2, \ldots, w_{n+5} = 6\}$ (ab dem n-ten Wurf fällt die Sequenz $1, 2, \ldots, 6$) Beachte: $\{A_n\}_{n\in\mathbb{N}}$ ist nicht stoch. unabhängig!
Die Folge $\{A_{6n}\}_{n\in\mathbb{N}}$ ist stoch. unabhängig. Mit Borel-Cantelli gilt:
$P(\limsup\limits_{n\to\infty} A_n) \geq P(\limsup\limits_{n\to\infty} A_{6n}) = 1$, da $\sum\limits_{n=1}^{\infty} P(A_{6n}) = \sum\limits_{n=1}^{\infty} \frac{1}{6^6} = \infty$.

3 Zufallsvariablen und ihre Verteilung

Zur Motivation: $(\Omega, \mathfrak{A}, P)$ W'raum: Beschreibung von Zufallsexperimenten, Modellierung von Zufallseinflüßen

- Oft interessiert nicht das gesamte Modell, sondern gewisse Teilgrößen, z.B. 5000 Stichproben mit Ausgang "gut/schlecht". Von Interesse ist nur die Anzahl der schlechten Stücke.
  $\Omega = \{(x_1, \dots, x_{5000}) | x_i \in \{g, s\}\}$ wäre die Ereignismenge
  $T = \{0, \dots, 5000\}$ von Interesse
- Ergebnisse von Zufallsexperimenten sind oft Zahlen oder Vektoren. Oft sind arithmetische Operationen nützlich, z.B. 5000 Münzwürfe mit Ergebnis 0 oder 1 (s.o.). $x_1, \dots, x_{5000} \in \{0, 1\}$
  Von Interesse: Anzahl Einsen: $\sum_{i=1}^{5000} x_i$
  oder $x_1, \dots, x_n \in \mathbb{R}^+$ (z.B. Verzögerungszeiten in einem Switch)
  $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ (mittlere Verzögerungszeit, Mittelwert)

Modellierung allgemein mit Zufallsvariablen

Definition 3.1 (Zufallsvariable) $(\Omega, \mathfrak{A}, P)$ sei W'raum. Eine Abbildung $X : \Omega \to \mathbb{R}$ heißt **Zufallsvariable** (ZV) oder **Zufallsgröße** (random variable (r.v.)), wenn $X^{-1}(B) = \{w \in \Omega | X(w) \in B\} \overset{kurz}{=} \{X \in B\} \in \mathfrak{A}$ für alle $B \in \mathfrak{B}^1$. Diese Bedingung heißt **Meßbarkeit** von X.
Schreibweise: $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ (X ist Abbildung und meßbar)

Bemerkung: Der Begriff "Zufallsvariable" hat sich eingebürgert, obwohl es sich eigentlich um "Funktionen" handelt. Betont wird die Modellierung des Zufalls, wichtig ist die Verteilung von ZV.

Lemma 3.2 $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ sei Zufallsvariable. Durch
$P^X(B) := P(X^{-1}(B)) = P(\{w | X(w) \in B\}) \overset{kurz}{=} P(X \in B)$ für $B \in \mathfrak{B}^1$ wird eine W'verteilung auf $(\mathbb{R}, \mathfrak{B}^1)$ definiert. P^X heißt **Verteilung der ZV** X.

Beweis: Überprüfe Definition 2.10:

(i) $P^X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$

(ii) B_n paarw. disj., $B_n \in \mathfrak{B}^1$, $n \in \mathbb{N}$

$$\begin{aligned} P^X\Big(\bigcup_{n=1}^{\infty} B_n\Big) &= P\Big(\underbrace{X^{-1}\big(\underbrace{\bigcup_{n=1}^{\infty} B_n}_{\in \mathfrak{B}^1}\big)}_{\in \mathfrak{A}}\Big) = P\Big(\bigcup_{n=1}^{\infty} \underbrace{X^{-1}(B_n)}_{paarw.\ disj.}\Big) \\ &= \sum_{n=1}^{\infty} P(X^{-1}(B_n)) = \sum_{n=1}^{\infty} P^X(B_n) \end{aligned}$$

Wesentlich in Definition 3.1: "meßbar", d.h. ZV X induzieren auf $(\mathbb{R}, \mathfrak{B}^1)$ die Verteilung P^X. Math. Modell für Zufallsexperimente häufig:

$$(\Omega, \mathfrak{A}, P) \xrightarrow{\quad\quad\quad\quad X \quad\quad\quad\quad} (\mathbb{R}, \mathfrak{B}^1, P^X)$$

Modelliert alle Zufallseinflüße, genaue Kentniss oft nicht erforderlich, lediglich Existenz nötig	Modelliert die interessierenden beobachteten Größen, P^X oft als Parameter bekannt. Beachte: $(\mathbb{R}, \mathfrak{B}^1, P^X)$ ist ein W'raum, bisherige Regeln gelten auch hier

Beispiel 3.3 (Binomialverteilung) n-facher Münzwurf, Kopf $\widehat{=}1$, Zahl $\widehat{=}0$, Würfe unabhängig. Kopf mit W'keit p, Zahl mit W'keit $1-p$ in jedem Wurf.
Modell: $\Omega = \{w = (x_1, \dots, x_n) | x_i \in \{0,1\}\}$ $\quad \mathfrak{A} = \mathcal{P}(\Omega)$,
$P(\{(x_1, \dots, x_n)\}) = \underbrace{p \dots p}_{Anz.\ Einsen} \underbrace{(1-p)\dots(1-p)}_{Anz.\ Nullen} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$

ZV X = Anz. der Einsen $\quad X(w) = X((x_1, \dots, x_n)) \sum\limits_{i=1}^n x_i$
Wertebereich von X: $T = \{0, 1, \dots, n\}$
Verteilung von X: $P^X(\{k\}) = P(X^{-1}(\{k\})) = P(\{w | X(w) = k\})$

$$\overset{kurz}{=} P(X = k) = P(\{(x_1, \dots, x_n) | \sum_{i=1}^n x_i = k\})$$
$$= \sum_{(x_1,\dots,x_n): \sum_{i=1}^n x_i = k} p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$
$$= \sum_{(x_1,\dots,x_n): \sum_{i=1}^n x_i = k} p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{für } k = 0, 1, \dots, n$$

Eine Zufallsvariable heißt **binomialverteilt** mit den Parametern $n \in \mathbb{N}$, $p \in [0,1]$, wenn $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, 1, \dots, n$.

Bezeichnung: $X \sim Bin(n, p)$
Interpretation: $Bin(n,p)$ ist die W'verteilung der Anzahl der "Treffer" in einer Bernoulli-Serie[4] der Länge n mit Trefferw'keit p.

Im weiteren: Allgemeine Methoden zur Beschreibung von W'verteilungen, direkt formuliert für ZV.
Beachte: Mit X = Identität lassen sich die folgenden Überlegungen auch direkt auf W'maße P anwenden.

3.1 Diskrete Verteilungen/Zufallsvariablen

Definition 3.4 Eine ZV X (auf einem W'raum $(\Omega, \mathfrak{A}, P)$), bzw. deren Verteilung P^X, heißt **diskret**, wenn eine höchstens abzählbare Menge $T = \{t_1, t_2, \dots\}$ mit $P^X(T) = P(X \in T) = 1$ existiert. T heißt **Träger** (support) von X bzw. P^X.

[4] Münzwurf-Serie der Länge n, unabhängige Würfe, Trefferw'keit p

X sei diskrete ZV mit Träger $T = \{t_1, t_2, \ldots\}$, $A \in \mathfrak{B}^1$. Dann gilt:
$P^X(A \cap T) \overset{Lemma\ 2.11b}{=} P^X(A) + \underbrace{P^X(T)}_{=1} - \underbrace{P^X(A \cup T)}_{=1} = P^X(A)$

Also: $P(X \in A) = P^X(A) = P^X(A \cap T) = P^X(\bigcup\limits_{i=1}^{\infty}(A \cap \{t_i\})) = \sum\limits_{i:t_i \in A}^{\infty} P^X(\{t_i\})$
$= \sum\limits_{i:t_i \in A}^{\infty} P(X = t_i)$

D.h. P^X, die Verteilung von X, ist eindeutig festgelegt durch $P(X = t_i)$ für $i = 1, 2, \ldots$

Definition 3.5 X sei diskrete ZV mit Träger $T = \{t_1, t_2, \ldots\}$. $f_X : T \to [0,1]$ mit $f_X(t_i) = P(X = t_i)$ für $i = 1, 2, \ldots$ heißt **Zähldichte** (discrete density function) der ZV X.

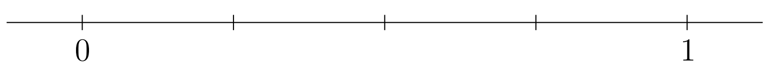
$X \sim Bin(n,p)$ (siehe Beispiel 3.3) ist ein Beispiel für eine diskrete ZV mit $f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, 1, \ldots, n$. ($0 \leq p \leq 1$ ist ein Parameter)

Beispiel 3.6 (Geometrische Verteilung)
Unendlicher unabhängiger Münzwurf (siehe Beispiel 3.3).
$\Omega = \{w = (x_1, x_2, \ldots) | x_i \in \{0,1\}\}$
X = Wartezeit bis zum ersten Treffer ("1"), Träger $T = \mathbb{N}_0$.
$P(X = k) = P(\{w = (x_1, x_2, \ldots) | x_1 = x_2 = \ldots = x_k = 0, X_{k+1} = 1\}) = (1-p)^k p$ für $k = 0, 1, \ldots$
Die Verteilung mit der Zähldichte $f_X(k) = (1-p)^k p$ für $k = 0, 1, \ldots$ ($0 < p \leq 1$ ein Parameter) heißt **geometrische Verteilung** . Bezeichnung: $X \sim Geo(p)$, $0 < p \leq 1$.

Beispiel 3.7 (Poissonverteilung, Gesetz seltener Ereignisse)
Sei $p_n \in (0,1)$ mit $n \cdot p_n \to \lambda$ für $n \to \infty$, $\lambda > 0$. Dann gilt:
$\binom{n}{k} p_n^k (1-p_n)^{n-k} \overset{n \to \infty}{\longrightarrow} e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}$
Es gilt $e^{-\lambda} \frac{\lambda^k}{k!} \geq 0$ und $e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^\lambda} = 1$.

Eine diskrete ZV X mit Zähldichte $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k = 0, 1, \ldots$ heißt **Poisson-verteilt** . ($\lambda > 0$ ein Parameter) Bezeichnung: $X \sim Poi(\lambda)$, $\lambda > 0$.

Interpretation: n Stücke gleicher Länge



Wahrscheinlichkeit für das Auftreten eines Ereignisses in jedem Teilstück ist $\frac{\lambda}{n}$, Auftreten unabhängig

3.2 Verteilungsfunktionen

Definition 3.8 Eine Funktion $F : \mathbb{R} \to [0, 1]$ mit den Eigenschaften

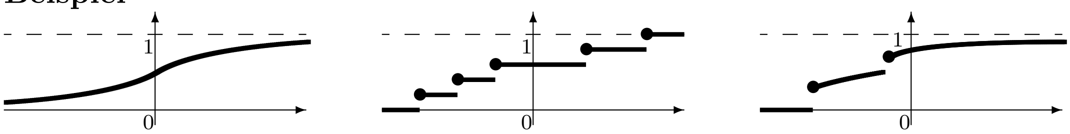
(i) F monoton steigend (nicht notwenigerweise streng monoton)

(ii) $\lim\limits_{x\to+\infty} F(x) = 1, \ \lim\limits_{x\to-\infty} F(x) = 0$

(iii) F ist rechtsseitig stetig, d.h. $\forall x_0,\ x_n \downarrow x_0 : F(x_n) \to F(x_0)$

heißt **Verteilungsfunktion** (VF) ([cumulative] distribution function, cdf).

Beispiel



Satz 3.9 $(\Omega.\mathfrak{A}, P)$ sei W'raum. $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}, \mathfrak{B}^1)$ eine ZV.
Durch $F_X(x) = P(X \leq x) = P(\{w \in \Omega | X(w) \leq x\}) = P^X((-\infty, x])$ für $x \in \mathbb{R}$ wird eine Verteilungsfunktion definiert, die **Verteilungsfunktion der ZV** X, bzw. der Verteilung P^X.

Beweis: Übung 5

Verteilungen auf $(\mathbb{R}, \mathfrak{B}^1)$ werden eindeutig durch VF beschrieben.

Satz 3.10 (Eindeutigkeitssatz für VF) 2 ZV X und Y besitzen dieselbe Verteilung (auf $(\mathbb{R}, \mathfrak{B}^1)$) $\iff$ $F_X(x) = F_Y(x)$ für alle $x \in \mathbb{R}$.

Beweis: "$\Longrightarrow$" einfach, in Übung 5
"$\Longleftarrow$" sehr schwierig! Benutzt Fortsetzungs- und Eind.-satz der Maßtheorie

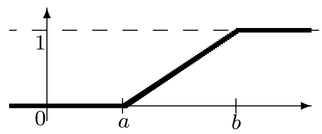
Darstellungsmethoden für Verteilungen/ZV

- Diskrete ZV $\to$ Zähldichte $f_x(t_i)$
- Auf $(\mathbb{R}, \mathfrak{B}^1)$ Verteilungsfunktionen $F_X(x) = P(X \leq x)$ mit $x \in \mathbb{R}$

Beispiel 3.11 (Verteilungsfunktionen)

a) X heißt **gleichverteilt** (**rechteckverteilt**) (uniform, rectangle distributed) auf $[0, 1]$, Bezeichnung $X \sim R(0, 1)$, wenn $F_X(x) = \begin{cases} 0 & , x \leq 0 \\ x & , 0 < x < 1 \\ 1 & , x \geq 1 \end{cases}$
Seien $0 \leq a < b \leq 1$.
$P(a < x \leq b) = P(\{w | a < X(w) \leq b\}) = P(\{w | X(w) \leq b\} \setminus \{w | X(w) \leq a\}) = P(\{w | X(w) \leq b\}) - P(\{w | X(w) \leq a\}) = P(X \leq b) - P(X \leq a)$
$= F_x(b) - F_X(a) = b - a$

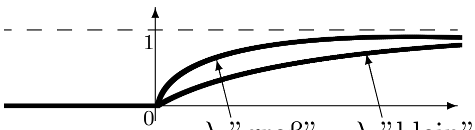
(W'keit für zufälligen Wert im Intervall $(a,b]$ ist gleich der Länge von $(a,b]$, nämlich $b-a$.)
Analog: Rechteckverteilung auf $[a,b]$, Bezeichnung $X \sim R(a,b)$ für $a < b \in \mathbb{R}$ mit $F_X(x) = \begin{cases} 0 & ,x \leq a \\ \frac{1}{b-a}(x-a) & ,a < x < b \\ 1 & ,x \geq b \end{cases}$



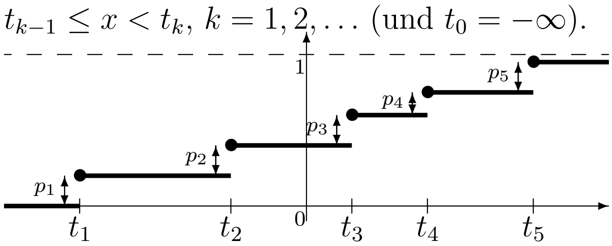
Seien $a \leq c < d \leq b$.
$P(c < x \leq d) = F_x(d) - F_X(c) = \frac{1}{b-a}(d-a) - \frac{1}{b-a}(c-a) = \frac{1}{b-a}(d-c)$

b) X heißt **exonentialverteilt**, $X \sim Exp(\lambda)$, $\lambda > 0$ ein Parameter, wenn
$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & ,x \geq 0 \\ 0 & ,x < 0 \end{cases} = (1 - e^{-\lambda x}) \cdot \mathbb{1}_{(0,\infty]}(x)$

Bei dieser Darstellung benutzt: **Indikatorfunktion** $\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$



c) Diskrete ZV X mit geordnetem Träger $T = \{t_1, t_2, \ldots\} \subseteq \mathbb{R}$, Zähldichte $f_X(t_i) = p_i$, $p_i \geq 0$ und $\sum_{i=1}^{\infty} p_i = 1$.
Dann $F_X(x) = P(X \leq x) = \sum\limits_{k:t_k \leq x} P(X = t_k) = \sum\limits_{k:t_k \leq x} p_k = \sum_{j=1}^{k-1} p_j$, falls $t_{k-1} \leq x < t_k$, $k = 1, 2, \ldots$ (und $t_0 = -\infty$).



"Treppenfunktion"

Zum Beispiel: $X \sim Geo(p)$ mit $f_X(k) = (1-p)^k p$, $k \in \mathbb{N}_0$ und $(0 < p \leq 1)$.
Es gilt: $\sum\limits_{j=0}^{k} (1-p)^j p = p\frac{1-(1-p)^{k+1}}{1-(1-p)} = 1 - (1-p)^{k+1}$
Also: $F_X(x) = \begin{cases} 0 & ,x < 0 \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & ,x \geq 0 \end{cases}$

Berechnung von Wahrscheinlichkeiten mit Verteilungsfunktionen

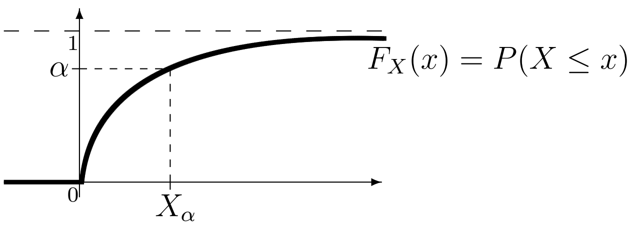
X sei ZV mit Verteilungsfunktion $F_X(x) = P(X \leq x)$. $a < b \in \mathbb{R}$

- $P(a < X \leq b) = P^X((-\infty, b] \setminus (-\infty, a]) = P^X((-\infty, b]) - P^X((-\infty, a])$
  $= F_X(b) - F_X(a)$
- Sei $a \in \mathbb{R}$, $a_n < a_{n+1} < a$ mit $a_n \to a$ für $n \to \infty$. Dann ist $(a_n, a]$ absteigend mit $\lim\limits_{n\to\infty} (a_n, a] = \{a\}$.
  $P(x = a) = P^X(\{a\}) = P^X(\lim\limits_{n\to\infty} (a_n, a]) \overset{Lemma\ 2.11b}{=} \lim\limits_{n\to\infty} P^X((a_n, a])$

$= \lim_{n\to\infty} (F_X(a) - F_X(a_n)) = F_X(a) - \lim_{n\to\infty} F_X(a_n) = F_X(a) - F_X(a-)$[5]
Insbesondere: Ist F_X stetig, so gilt $P(X = a) = 0$ für alle $a \in \mathbb{R}$.

- $P(a \le X \le b) = P(X = a) + P(a < X \le b) = F_X(a) - F_X(a-) + F_X(b) - F_X(a) = F_X(b) - F_X(a-)$

Bedienzeit X von Anforderungen an einen Server seien $Exp(\lambda)$-verteilt, $X \sim Exp(\lambda)$, $\lambda > 0$.
Problem: Bestimme die Zeit X_α, unterhalb derer die Bedienzeit mit vorgegebener Wahrscheinlichkeit α liegt, z.B. $\alpha = 0.99$. Bestimme also X_α mit $P(X \le X_\alpha) = \alpha$.
Interpretation: In 99% der Fälle ist die Bedienzeit kürzer als X_α.
Graphisch für $X \sim Exp(\lambda)$:



Definition 3.12 X sei ZV mit Verteilungsfunktion $F_X(x)$, $0 \le \alpha \le 1$.
$X_\alpha = \min\{x | F_X(x) \ge \alpha\}$ heißt α**-Quantil** (α-Percentil, $(1-\alpha)$-Fraktil) von F_X.
$F_X^-(t) = \min\{x | F_X(x) \ge t\}$, $t \in (0, 1)$, heißt **Pseudoinverse** von F_X. (Nötig, weil X_α genau in einem Sprung von F_X liegen könnte.)

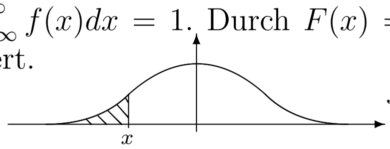
- Graphisch bei nicht-stetigen Vfkt.
- Ist F invertierbar, so gilt $F^- = F^{-1}$ (Inverse).

Betrachte: $\alpha = \frac{1}{2}$. $X_{\frac{1}{2}}$ heißt **Median** von F_X. Es gilt $P(X \le X_{\frac{1}{2}}) \le \frac{1}{2}$ und $P(X < X_{\frac{1}{2}} \le \frac{1}{2}) \le \frac{1}{2}$. Der Median "halbiert" die Verteilung, ist in diesem Sinn der "mittlere Wert". Der Median ist ein Schätzer für den mittleren Wert, der robust gegen Ausreißer ist.

3.3 Dichten

Methode zur Konstruktion von Verteilungsfunktionen, Verteilungen.

Definition 3.13 $F : \mathbb{R} \to \mathbb{R}^+$ sei (uneigentlich Riemann-)integrierbar mit $\int_{-\infty}^{\infty} f(x)dx = 1$. Durch $F(x) = \int_{-\infty}^{\infty} f(t)dt$ wird eine Verteilungsfunktion definiert.



$\int_{-\infty}^{x} f(t)dt$

[5]linksseitiger Grenzwert von $F_X(a)$

Gilt für eine ZV X, daß $F_X(x) = F(x)$, so heißt f **(Verteilungs-) Dichte** (probability density function (pdf)) von X bzw. P^X. X bzw. P^X heißt dann **absolut-stetig**.

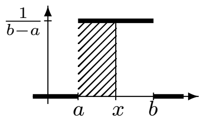
Mit dem Hauptsatz der Differential- und Integralrechnung: $f(x) = F'(X)$ für alle Stetigkeitspunkte x von f.

Beispiel 3.14

a) Rechtecksverteilung auf $[a, b]$, $a < b \in \mathbb{R}$, $X \sim R(a, b)$:

$$f(x) = \begin{cases} \frac{1}{b-a} & , a \leq x \leq b \\ 0 & , \text{sonst} \end{cases} = \tfrac{1}{b-a}\mathbb{1}_{[a,b]}(x) \text{ für alle } x \in \mathbb{R}$$

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \begin{cases} \frac{1}{b-a}(x-a) & , a \leq x \leq b \\ 0 & , x \leq a \\ 1 & , x \geq b \end{cases}$$



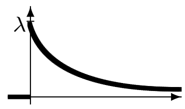
Beachte: Die Dichte $\widehat{f}(x) = \frac{1}{b-a}\mathbb{1}_{(a,b)}(x)$ führt zu derselben Verteilungsfunktion F. Dichten sind nicht eindeutig, sie sind "fast sicher" eindeutig.

b) Exponentialverteilung, $X \sim Exp(\lambda)$ mit $\lambda > 0$:

$$f(x) = \lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x) \text{ für alle } x \in \mathbb{R}$$

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{0}^{x} \lambda e^{-\lambda t}dt = -e^{-\lambda t}\Big|_0^x$$

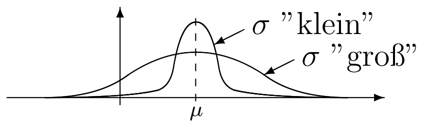
$$= 1 - e^{-\lambda x} \text{ mit } x \geq 0$$



c) Normalverteilung, $X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma > 0$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ für alle } x \in \mathbb{R}$$

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$



hat keine "geschlossene" Darstellung. Berechnung numerisch, approximativ oder mit Tabellen.

Berechnung von Wahrscheinlichkeiten mit Dichten

X sei ZV mit Verteilungsfunktion F_X und f_X, $a < b \in \mathbb{R}$.

- $P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(t)dt - \int_{-\infty}^{a} f_X(t)dt = \int_a^b f_X(t)dt$
- $P(X = a) = F_X(a) - F_X(a-) = 0$, da $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ stetig
- $P(a < X \leq b) = \int_a^b f_X(t)dt$
- Allgemein gilt bei absolut-stetigen ZV (also solchen mit Dichte) $P(X \in \langle a, b\rangle) = \int_a^b f_X(t)dt$, wobei $\langle$ bzw. $\rangle$ beliebig für "abgeschlossen" oder " offen" steht
- Schreibe allgemein für Mengen $B \in \mathfrak{B}^1$: $P(X \in B) = \int_B f_X(t)dt$ (auch wenn B kein Intervall)

3.4 Erzeugende Funktionen und Laplace-Transformierte

Eine weitere Methode zur eindeutigen Beschreibung von Verteilungen.

Definition 3.15

a) X sei diskrete ZV mit Träger $T = \{t_1, t_2, \ldots\}$ und Zähldichte $f_X(t_k) = p_k$ mit $k \in \mathbb{N}_0$. Dann heißt $G_X(z) = \sum\limits_{k=0}^{\infty} p_k z^k$ mit $|z| < 1$ **erzeugende Funktion** von X bzw. P^X (probability generating function).
b) X sei absolut-stetig mit Dichte f_X, wobei $f_X(x) = 0$, falls $x < 0$ (d.h. $P(X < 0) = 0$). Dann heißt $L_X(s) = \int\limits_0^{\infty} e^{-sx} f_X(x)dx$ mit $s \geq 0$ die **Laplace-Transformierte** von X bzw. P^X.

Analog zu Satz 3.10 für Verteilungsfunktionen gilt auch hier Eindeutigkeit.

Satz 3.16

a) X und Y seien diskrete Zufallsvariablen mit denselbem Träger T. X und Y besitzen dieselbe Verteilung $\iff$ $G_X(z) = G_Y(z)$ für alle $|z| \leq 1$.
b) X, Y seien absolut-stetig mit $f_X(x) = f_Y(x) = 0$ für alle $x < 0$. X und Y haben dieselbe Verteilung $\iff$ $L_X(s) = L_Y(s)$ für alle $s \geq 0$.

Beweis:

a) Eindeutigkeitssatz für Potenzreihen
b) Feller II, Seite 403, Kapitel XIII oder Satz von Stone-Weierstraß

Satz 3.17 (Inversionsformeln)

a) $G_X(z)$ sei erzeugende Funktion einer diskreten ZV X mit Träger $T = \{t_1, t_2, \ldots\}$. Dann gilt $P(X = t_k) = \frac{1}{k!} G_X^{(k)}(0)$.
b) $L_X(s)$ sei Laplace-Transformierte einer absolut-stetigen ZV X. Dann gilt:
$f_X(x) = \lim\limits_{y\to\infty} \frac{1}{2\pi i} \int\limits_{c-iy}^{c+iy} e^{sx} L_X(s)ds$ für alle $c \in \mathbb{R}$
(Integration über den Weg $c - iy \to c + iy$)

Beispiel 3.18

a) Geometrische Verteilung, $X \sim Geo(p)$ mit $0 < p \leq 1$
$G_X(z) = \sum\limits_{k=0}^{\infty} (1-p)^k p z^k = p \sum\limits_{k=0}^{\infty} ((1-p)z)^k = \frac{p}{1-z+pz}$ mit $|z| \leq 1$
b) Poissonverteilung, $X \sim Poi(\lambda)$ mit $\lambda > 0$
$G_X(z) = \sum\limits_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda} \sum\limits_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} e^{\lambda z} = e^{-\lambda(1-z)}$ für alle $z \in \mathbb{R}$

c) Rechtecksverteilung, $X \sim R(0,1)$

$L_X(s) = \int\limits_0^1 e^{-sx}dx = \frac{1-e^{-s}}{s}$ für alle $s \geq 0$

d) Exponentialverteilung, $X \sim Exp(\lambda)$

$L_X(s) = \int\limits_0^\infty e^{-sx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda+s} \underbrace{\int_0^\infty (\lambda+s)e^{-(s+\lambda)x}dx}_{\text{1, da Integral über die Dichte}} = \frac{\lambda}{\lambda+s}$

e) Gammaverteilung, $X \sim \Gamma(\alpha,\lambda)$

$L_X(s) = \int\limits_0^\infty e^{-sx}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx = \frac{\lambda^\alpha}{(\lambda+s)^\alpha} \underbrace{\int\limits_0^\infty \frac{(\lambda+s)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\lambda+s)x}dx}_{1} = \left(\frac{\lambda}{\lambda+s}\right)^\alpha$

f) Erlangverteilung, $X \sim Erl(n,\lambda) = \Gamma(n,\lambda)$ für $n \in \mathbb{N}$

L wie in e) oder alternativ mit Satz 6.10 für X als Faltung der Exp-Vert.:

$L_{X1+\ldots+X_n}(s) = L_{X_1}(s) \cdot \ldots \cdot L_{X_n}(s) = L^n_{X_1}(s)$ für stid X_i,

hier: $X_i \sim Exp(\lambda)$, $L_{X_i} = \frac{\lambda}{\lambda+s}$, also $L^n_{X_1}(s) = \left(\frac{\lambda}{\lambda+s}\right)^n$

Nice to know...

$\sum\limits_{k=0}^\infty x^k = \frac{1}{1-x}$

$\sum\limits_{k=0}^\infty \frac{x^k}{k!} = e^x$

4 Produkträume und Zufallsvektoren

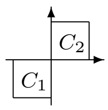
Zufallsexperimente mit mehreren Ausgängen (gleichzeitig oder dasselbe Experiment mehrfach wiederholt).
Modell:

$$(\Omega, \mathfrak{A}, P) \begin{array}{l} \nearrow (\mathbb{R}, \mathfrak{B}^1) \\ \longrightarrow (\mathbb{R}, \mathfrak{B}^1) \\ \qquad \vdots \\ \searrow (\mathbb{R}, \mathfrak{B}^1) \end{array}$$

Ausgänge gemeinsam betrachten: $(\Omega, \mathfrak{A}, P) \xrightarrow{X=(X_1,\ldots,X_n)} (\mathbb{R}^n, ?, ?)$

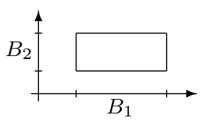
4.1 Produkträume

Gegeben Meßräume $(\Omega_i, \mathfrak{A}_i)$ für $i = 1, \ldots, n$. Oft $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B}^1)$.
Setze $\Omega = \bigtimes_{i=1}^n \Omega_i = \{w = (w_1, \ldots, w_n) | w_i \in \Omega_i\}$.
Frage: Welche σ-Algebra über Ω wählen?
Ansatz: $\varepsilon = \{A_1 \times \ldots \times A_n | A_i \in \mathfrak{A}_i,\ i = 1, \ldots, n\}$
Aber: ε ist im Allgemeinen keine σ-Algebra über $\Omega = \bigtimes_{i=1}^n \Omega_i$.
Gegenbeispiel: $\Omega_1 = \Omega_2 = \mathbb{R}$, $\mathfrak{A}_1 = \mathfrak{A}_2 = \mathfrak{B}^1$. Dann $C_1 = [-1, 0] \times [-1, 0] \in \varepsilon$, $C_2 = [0, 1] \times [0, 1] \in \varepsilon$, aber $C_1 \cup C_2 \notin \varepsilon$, ist nicht als kartesisches Produkt zweier Mengen darstellbar.



Definition 4.1 (Produkt-σ-Algebra)
$(\Omega_i, \mathfrak{A}_i)$ seien Meßräume, $i = 1, \ldots, n$. $\varepsilon = \{A_1 \times \ldots \times A_n | A_i \in \mathfrak{A}_i,\ i = 1, \ldots, n\}$.
$\bigotimes_{i=1}^n \mathfrak{A}_i := \mathfrak{A}(\varepsilon)$[6] heißt **Produkt-$\sigma$-Algebra** von $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$ über $\Omega = \bigtimes_{i=1}^n \Omega_i$.
$(\bigtimes_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathfrak{A}_i)$ heißt **Produkt-Meßraum** der $(\Omega_i, \mathfrak{A}_i)$ für $i = 1, \ldots, n$.

Beispiel 4.2 $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B}^1)$. $\varepsilon = \{B_1 \times \ldots \times B_n | B_i \in \mathfrak{B}^1\}$.
$\mathfrak{A}(\varepsilon)$ heißt **n-dimensionale Borelsche σ-Algebra**.
Graphisch für $n = 2$:



"$\mathfrak{B}^n$ enthält alle n-dimensionalen Rechtecke"

4.2 Zufallsvektoren und Folgen von Zufallsvariablen

Definition 4.3 $(\Omega, \mathfrak{A}, P)$ sei W'raum. Eine Abb. $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ heißt **Zufallsvektor**, wenn $X^{-1}(B) \in \mathfrak{A}$ für alle $B \in \mathfrak{B}^n$ (Meßbarkeit). Bezeichnung: $X : (\Omega, \mathfrak{A}) \to (\mathbb{R}^n, \mathfrak{B}^n)$.
Bemerkung: Es gilt: $X = (X_1, \ldots, X_n)$ ist Zufallsvektor $\iff$ $X_1, \ldots, X_n$ sind ZV.

[6] D.h. wir definieren $\bigotimes_{i=1}^n \mathfrak{A}_i$ als die von ε erzeugte σ-Algebra

Lemma 4.4 X sei Zufallsvektor.
Durch $P^X(B) = P(X^{-1}(B)) = P(\{w|X(w) \in B\}) \overset{kurz}{=} P(X \in B)$ für alle $B \in \mathfrak{B}^n$ wird eine Wahrscheinlichkeitsverteilung auf $(\mathbb{R}^n, \mathfrak{B}^n)$ definiert.
$P^X = P^{(X_1,\ldots,X_n)}$ heißt **gemeinsame Verteilung** von $(X_1, \ldots, X_n)$.

Im folgenden werden gemeinsame Verteilungen mit Verteilungsfunktionen und Dichten beschrieben. Klar: $X_1, \ldots, X_n$ diskret $\Longrightarrow$ $(X_1, \ldots, X_n)$ diskret. Verteilung mit Zähldichten, wie in Kapitel 3 beschreiben.

Satz 4.5 X sei Zufallsvektor. P^X wird eindeutig beschrieben durch die (n-dimensionale) Verteilungsfunktion $F_X(x_1, \ldots, x_n) = P^X((-\infty, x_1] \times \ldots \times (-\infty, x_n]) = P^X(\{w|X_1(w) \in (-\infty, x_1], \ldots, X_n(w) \in (-\infty, x_n]\})$
$= P(\{w|X_1(w) \leq x_1\} \cap \ldots \cap \{w|X_n(w) \leq x_n\}) \overset{kurz}{=} P(X_1 \leq x_1, \ldots, X_n \leq x_n)$.
Bezeichnung: $X \sim F_X$

Beschreibung von Verteilungen mit n-dimensionalen Dichten ist oft einfacher.

Definition 4.6 $F_X(x_1, \ldots, x_n)$ sei die Verteilungsfunktion des Zufallsvektors X. Eine (uneigentlich Riemann-)integrierbare Funktion $f_X : \mathbb{R}^n \to \mathbb{R}^+$ heißt **Dichte** von X, wenn $F_X(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \ldots dt_n$ für alle $x_1, \ldots, x_n \in \mathbb{R}$. X heißt dann **absolut-stetig** mit Dichte f_X.
Bezeichnung: $X \sim f_X$.

Sprechweisen: X hat/besitzt/ist verteilt nach Vfkt. F_X/Dichte f_X.

Berechnung von Wahrscheinlichkeiten mit Dichten:
$X = (X_1, \ldots, X_n)$ sei absolut-stetig mit Verteilung P^X, Verteilungsfunktion F_X und Dichte f_X. Dann ist $P^X(\langle a_1, b_1\rangle \times \ldots \times \langle a_1, b_1\rangle) = P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) = \int_{a_n}^{b_n} \ldots \int_{a_1}^{b_1} f_X(t_1, \ldots, t_n) dt_1 \ldots dt_n$ für alle $a_i \leq b_i \in \mathbb{R}$, wobei "$\langle$" bzw. "$\rangle$" beliebig für "offen" oder "abgeschlossen" stehen.
Allgemein: $P^X(B) = \int \underset{B}{\ldots} \int f_X(t_1, \ldots, t_n) dt_1 \ldots dt_n$ für alle $B \in \mathfrak{B}^n$.

Mit dem Hauptsatz der Differential- und Integralrechnung gilt $f_X(x_1, \ldots, x_n) = \frac{\partial}{\partial x_1 \ldots \partial x_n} F_X(x_1, \ldots, x_n)$ in allen Stetigkeitspunkten $(x_1, \ldots, x_n)$ von f_X.

Beispiel 4.7 (Indizierung mit X zur Vereinfachung weggelassen)

a) Gleichverteilung auf $T \in \mathfrak{B}^n$ mit $c = \int \underset{T}{\ldots} \int 1 dx_1 \ldots dx_n \quad < \infty$

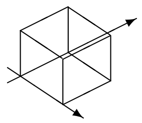
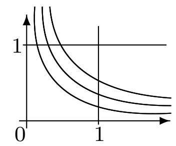
$F(x_1, \ldots, x_n) = \frac{1}{c} \mathbb{1}_T(x_1, \ldots, x_n) = \begin{cases} \frac{1}{c} & , (x_1, \ldots, x_n) \in T \\ 0 & , \text{ sonst} \end{cases}$

<u>*Speziell:*</u>

- Auf dem Einheitswürfel des $\mathbb{R}^n$
  $T = \{(x_1, \ldots, x_n) | 0 \leq x_i \leq 1,\ i = 1, \ldots, n\} = [0, 1]^n$
  $f(x_1, \ldots, x_n) = \mathbb{1}_T(x_1, \ldots, x_n)$

$$F(x_1,\dots,x_n)=\begin{cases}0 & ,\exists i: x_i<0\\ \min\{x_1,1\}\cdot\ldots\cdot\min\{x_n,1\} & ,\text{ sonst}\end{cases}$$

Graphisch für $n=2$:

f  F 

Isohöhenlinien

- Auf der Einheitskugel des $\mathbb{R}^n$
  $T=\{(x_1,\dots,x_n)|\sum_{i=1}^n x_i^2\le 1\}$, $\quad c=\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$ Volumen der Einheitskugel
  $\boxed{\Gamma(n)=(n-1)!\text{ für alle } n\in\mathbb{N},\ \Gamma(\tfrac{1}{2})=\sqrt{\pi},\ \Gamma(x+1)=x\Gamma(x)}$
  $f(x_1,\dots,x_n)=\frac{\Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}}\mathbb{1}_T(x_1,\dots,x_n)$ mit $(x_1,\dots,x_n)\in\mathbb{R}^n$
  Die Verteilungsfunktion ist schwierig zu berechnen $\rightarrow$ Dirichlet-Integral

b) n-dimensionale Normalverteilung, $X\sim N(\mu,\Sigma)$, $\mu\in\mathbb{R}^n$, $\Sigma\in\mathbb{R}^{n\times n}$ positiv definit. Für $x=(x_1,\dots,x_n)^T\in\mathbb{R}^n$ ist
$f(x_1,\dots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\Sigma)^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

c) $f_1,\dots,f_k:\mathbb{R}^n\to\mathbb{R}^+$ seien n-dim. Dichten (auch $n=1$), $\alpha_1,\dots,\alpha_k\ge 0$, $\sum_{i=1}^k\alpha_i=1$. Dann ist $f(x)=\sum_{i=1}^k\alpha_i f_i(x)$, $x=(x_1,\dots,x_n)\in\mathbb{R}^n$, eine n-dimensionale Dichte. $f(x)$ heißt **Mischung** der Dichten $f_1,\dots,f_k$.

Stochastische Unabhängigkeit

Definition 4.8 $X_1,\dots,X_n$ seien ZV. (Dann ist $X=(X_1,\dots,X_n)$ ein Zufallsvektor.) $X_1,\dots,X_n$ heißen **stochastisch unabhängig**, wenn $F_X(x_1,\dots,x_n)=F_{X_1}(x_1)\cdot\ldots\cdot F_{X_n}(x_n)$ für alle $(x_1,\dots,x_n)\in\mathbb{R}^n$.
Also: $X_1,\dots,X_n$ stoch. unabh., wenn $P(X_1\le x_1,\dots,X_n\le x_n)=P(X_1\le x_1)\cdot\ldots\cdot P(X_n\le x_n)$ für alle $(x_1,\dots,x_n)\in\mathbb{R}^n$.
$P(\{w|X_1(w)\le x_1\}\cap\ldots\cap P(\{w|X_n(w)\le x_n\}=P(\{w|X_1(w)\le x_1\}\cdot\ldots\cdot P(\{w|X_n(w)\le x_n\}$, d.h. $X_1,\dots,X_n$ sind stoch. unabh. genau dann, wenn die Ereignisse $\{w|X_i(w)\le x_i\}$ für $1=1,\dots,n$ und alle $x_1,\dots,x_n$ stoch. unabh. im Sinn der Definition 2.16 sind.
(Beachte: Auswahlen von $i_1,\dots,i_k$ durch Setzen der anderen $x_j=\infty$.)

Lemma 4.9 $(X_1,\dots,X_n)$ sei Zufallsvektor, absolut-stetig mit Dichte $f_{(X_1,\dots,X_n)}$. Dann hat jedes der X_i eine Dichte $f_{X_i}(x)$
$=\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}f_{(X_1,\dots,X_n)}(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n)dx_1\dots dx_{i-1}dx_{i+1}\dots dx_n$.
Es gilt: $f_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\prod_{i=1}^n f_{X_i}(x_i)$ für alle $x_1,\dots,x_n\in\mathbb{R}$
$\Longrightarrow\quad X_1,\dots,X_n$ stoch. unabh.

Umgekehrt: Sind $X_1,\dots,X_n$ stoch. unabh., absolut-stetig mit Dichten f_{X_i}, so ist $f_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\prod_{i=1}^n f_{X_i}(x_i)$ für alle $(x_1,\dots,x_n)\in\mathbb{R}^n$ eine Dichte des Zufallvektors $(X_1,\dots,X_n)$.

Beweis:
$f_{X_i}(x) = \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} f_{(X_1,\dots,X_n)}(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_n)dx_1\dots dx_{i-1}dx_{i+1}\dots dx_n$
ist eine Dichte von x_i, da $F_{X_i}(x) = P(X_i \le x) = \int\limits_{-\infty}^{x} f_{X_i}(t_i)dt_i$.

Es gilt $F_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \int\limits_{-\infty}^{x_n} \dots \int\limits_{-\infty}^{x_1} \prod\limits_{i=1}^{n} f_{X_i}(t_i)dt_i \dots dt_n$
$= \prod\limits_{i=1}^{n} \int\limits_{-\infty}^{x_i} f_{X_i}(t_i)dt_i = \prod\limits_{i=1}^{n} F_{X_i}(x_i)$ für alle $x_1,\dots,x_n \in \mathbb{R}$.
Also $X_1,\dots,X_n$ stochastisch unabhängig nach Definition 4.8.

Umgekehrt: $X_1,\dots,X_n$ stoch. unabh. $\implies$ $F_{(X_1,\dots,X_n)}(x_1,\dots,x_n) = \prod\limits_{i=1}^{n} F_{X_i}(x_i)$
$= \prod\limits_{i=1}^{n} \int\limits_{-\infty}^{x_i} f_{X_i}(t_i)dt_i = \int\limits_{-\infty}^{x_n} \dots \int\limits_{-\infty}^{x_1} \prod\limits_{i=1}^{n} f_{X_i}(t_i)dt_i \dots dt_n \implies \prod\limits_{i=1}^{n} f_{X_i}(x_i)$ ist eine Dichte von $(X_1,\dots,X_n)$.

Im Fall von Zähldichten vereinfacht sich Lemma 4.9 zu:

Lemma 4.10 $X_1,\dots,X_n$ seien diskrete ZV mit Trägern $T_1,\dots,T_n$. Dann gilt: $X_1,\dots,X_n$ stoch. unabh. $\iff$ $P(X_1 = t_1,\dots,X_n = t_n) = \prod\limits_{i=1}^{n} P(X_i = x_i)$ für alle $t_i \in T_i$.

Beweis: Übung 8 Aufgabe 29

Definition 4.11 Eine Folge von ZV $\{X_n\}_{n\in\mathbb{N}}$ auf einem W'raum $(\Omega,\mathfrak{A},P)$ heißt **stochastisch unabhängig**, wenn $X_1,\dots,X_n$ stochastisch unabhängig für alle $n \in \mathbb{N}$. Besitzen stochastisch unabhängige ZV alle dieselbe Verteilung, so heißen sie **stochastisch unabhängig identisch verteilt (stid)** (i.i.d., independent identically distributed).

Beispiel 4.12

a) X_1, X_2 stid $\sim Bin(n,p)$
$P(X_1 = X_2) = P^{(X_1,X_2)}(\{(i,j)|0 \le i,j \le n,\ i = j\})$
$= \sum\limits_{i=0}^{n} P^{(X_1,X_2)}(\{(i,i)\}) = \sum\limits_{i=0}^{n} P(X_1 = i, X_2 = i)$
$= \sum\limits_{i=0}^{n} P(X_1 = i)P(X_2 = i) = \sum\limits_{i=0}^{n} [\binom{n}{i} p^i (1-p)^{n-i}]^2$

b) X_1, X_2 stochastisch unabhängig, absolut-stetig mit Dichten f_1, f_2
$P(X_1 = X_2) = \int\limits_{X_1=X_2}\int f(x_1,x_2)dx_1dx_2 = \int\limits_{-\infty}^{\infty} \underbrace{\int_{x_2}^{x_2} f_1(x_1)dx_1}_{0} f_2(x_2)dx_2 = 0$
$\boxed{f(x_1,x_2) = f_1(x_1)f_2(x_2)}$

Beispiel 4.13 $X_1, \ldots, X_n$ stid $\sim F$ (Verteilungsfunktion)

a) $Y = \max{}^{7}\{X_1, \ldots, X_n\}$
Interpretation z.B. Maximum der Laufzeiten $(X_1, \ldots, X_n)$
$P(Y \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = \prod\limits_{i=1}^{n} P(X_i \leq x) = F^n(x)$ für alle $x \in \mathbb{R}$
Also: $Y = \max\{X_1, \ldots, X_n\}$ hat die Verteilungsfunktion $F^n(x)$.

b) $Y = \min\{X_1, \ldots, X_n\}$
$P(Y > x) = P(X_1 > x, \ldots, X_n > x) = (1 - F(x))^n$
Also: $Y = \min\{X_1, \ldots, X_n\}$ hat die Verteilungsfunktion $1 - (1 - F(x))^n$.
Z.B. X_i stid $\sim Exp(\lambda)$, dann $P(Y \leq x) = 1 - e^{-n\lambda x} \quad \Longrightarrow \quad Y \sim Exp(n\lambda)$

Beispiel 4.14 X_1, X_2 stid $\sim Exp(\lambda)$ mit $\lambda > 0$
$P(X_1 + X_2 \leq z) = \int\limits_{0 \leq X_1 + X_2 \leq z} \int \lambda e^{-\lambda x_1} \mathbb{1}_{[0,\infty)}(x_1) \lambda e^{-\lambda x_2} \mathbb{1}_{[0,\infty)}(x_2) dx_1 dx_2$
$= \int\limits_0^z \int\limits_0^{z - x_2} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2} dx_1 dx_2 = \int\limits_0^z \lambda e^{-\lambda x_2} [1 - e^{-\lambda(z - x_2)}] dx_2$
$= \int\limits_0^z [\lambda e^{-\lambda x_2} - \lambda e^{-\lambda z}] dx_2 = 1 - e^{-\lambda z} - \lambda z e^{-\lambda z} = F_{X_1 + X_2}(z)$ für $z \geq 0$
Durch Differenzieren:
$f_{X_1 + X_2}(z) = \lambda e^{-\lambda z} - \lambda(e^{-\lambda z} - \lambda z e^{-\lambda z}) = \lambda^2 z e^{-\lambda z}$ für $z \geq 0$
Also: Die Verteilung der Summe von stochastisch unabhängig identisch $Exp(\lambda)$-verteilten ZV ist absolut stetig mit Dichte $f(x) = \lambda^2 x e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$.
Eine allgemeine Klasse von Verteilungen, die diese Dichte als Spezialfall enthält, sind die **Γ-Verteilungen** mit den Dichten $f_{\alpha,\lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$ mit $\alpha, \lambda > 0$ als Parameter, wobei $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ das Γ-Integral (insbesondere $\Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}$) ist. Eine ZV mit Dichte $f_{\alpha,\lambda}(x)$ heißt **$\Gamma(\alpha, \lambda)$-verteilt**.
Spezialfälle:

(i) $\alpha = 1 \quad \rightarrow \quad Exp(\lambda)$-Verteilung
(ii) $\alpha = 2 \quad \rightarrow$ s.o., Verteilung der Summe von 2 stid $Exp(\lambda)$-verteilten ZV
(iii) $\alpha = n \in \mathbb{N} \quad \rightarrow$ Erlang-Verteilung mit Parametern n, λ
$f_{n,\lambda}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$

Beispiel 4.15 $\{X_n\}_{n \in \mathbb{N}}$ Folge von stid ZV, $B \in \mathfrak{B}^1$.
$S = \min\{n \in \mathbb{N} | X_n \in B\}$ heißt **erste Eintrittszeit in** B.
Verteilung von S: $P(S = k) = P(X_1 \notin B, \ldots, X_{k-1} \notin B, X_k \in B)$
$= P(X_1 \notin B) \cdot \ldots \cdot P(X_{k-1} \notin B) P(X_k \in B) = (1-p)^k p$ für $k \in \mathbb{N}$,
wobei $p = P(X_1 \in B) = \ldots = P(X_n \in B)$.
Also: S geometrisch verteilt mit Träger $\mathbb{N}$. $S - 1 \sim Geo(p)$
Konkret: unabhängige Folge von Münzwürfen, X_n stid $\sim Bin(1, p)$

[7] punktweises Maximum der X_i an den Trägerpunkten

Setze $B = \{1\}$. $P(X_n \in \{1\}) = P(X_n = 1) = p$.
$\{S = k\}$: Beim k-ten Wurf tritt erstmalig eine 1 ("Kopf, Treffer") auf
Setze $X = S - 1 =$ "Wartezeit bis zum ersten Treffer". Dann $X \sim Geo(p)$

Satz 4.16 $X_1, \ldots, X_n$ seien stochastisch unabhängige ZV und $I, J \subseteq \{1, \ldots, n\}$, $I \cap J = \emptyset$. Dann sind $(X_i)_{i \in I}$ und $(X_j)_{j \in J}$ stochastisch unabhängige ZV. Sind f, g meßbare Abbildungen (mit entsprechenden Bild-Meßräumen), so sind $f((X_i)_{i \in I})$ und $g((X_j)_{j \in J})$ stochastisch unabhängig.

Beweis: Mathar & Pfeiffer, Lemma 2.1.6 und Lemma 2.1.7, Seite 74

Beispiel 4.17 X_1, X_2, X_3, X_4 stochastisch unabhängig $\Longrightarrow$ $X_1 + X_2, X_3 + X_4$ stochastisch unabhängig

5 Transformationen von ZV/Verteilungen

Im folgenden: allgemeine Hilfsmittel zur Berechnung der Verteilung von Funktionen von Zufallsvariablen

Satz 5.1 (Transformationen für Dichten)
$X = (X_1, \ldots, X_n)$ sei absolut-stetiger Zufallsvektor auf dem W'raum $(\Omega, \mathfrak{A}, P)$ mit Dichte f_X. Gelte $f_X(x_1, \ldots, x_n) = 0$ für alle $(x_1, \ldots, x_n) \in M^C$ für eine offene Menge $M \subseteq \mathbb{R}^n$. $T : (\mathbb{R}^n, \mathfrak{B}^n) \to (\mathbb{R}^n, \mathfrak{B}^n)$ sei eine meßbare Abbildung (Zufallsvektor, $T^{-1}(B) \in \mathfrak{B}^n \quad \forall B \in \mathfrak{B}^n$) mit

(i) $\widetilde{T} = T|_M$ ist injektiv (Restriktion von T auf M)

(ii) $\widetilde{T}$ ist stetig differenzierbar auf M

(iii) $\det\left(\left(\frac{\partial \widetilde{T}_i}{\partial X_j}\right)_{1 \leq i,j \leq n}\right) \neq 0$ auf M (Funktionsdeterminante)

Dann ist $Y = \widetilde{T}(X)$ absolut-stetig mit Dichte

$$f_Y(y_1, \ldots, y_n) = \frac{f_X(\widetilde{T}^{-1}(y_1, \ldots, y_n))}{\left|\det\left(\left(\frac{\partial \widetilde{T}_i}{\partial x_j}\right)_{1 \leq i,j \leq n}\Big|_{\widetilde{T}^{-1}(y_1,\ldots,y_n)}\right)\right|} \cdot \mathbb{1}_{\widetilde{T}(M)}(y_1, \ldots, y_n)$$

$$= \left|\det\left(\left(\frac{\partial \widetilde{T}_i^{-1}}{\partial y_j}\right)_{1 \leq i,j \leq n}\right)\right| \cdot f_X(\widetilde{T}^{-1}(y_1, \ldots, y_n)) \cdot \mathbb{1}_{\widetilde{T}(M)}(y_1, \ldots, y_n)$$

Beweis: Krickeberg (63)

Beispiel 5.2
$X = (X_1, X_2)$ gleichverteilt auf $Q = (0,1)^2$ mit Dichte $f_X(x_1, x_2) = \mathbb{1}_Q(x_1, x_2)$, d.h. X_1, X_2 stochastisch unabhängig, $X_i \sim R(0,1)$, $i = 1, 2$.
$T(X_1, X_2) = (\sqrt{x_1}\cos(2\pi x_2), \sqrt{x_1}\sin(2\pi x_2)$ mit $(x_1, x_2) \in Q$

$$\det\left(\left(\frac{\partial \widetilde{T}_i}{\partial y_j}\right)_{1 \leq i,j \leq n}\right) = \det\begin{pmatrix} \frac{1}{2\sqrt{x_1}}\cos(2\pi x_2) & -2\pi\sqrt{x_1}\sin(2\pi x_2) \\ \frac{1}{2\sqrt{x_1}}\sin(2\pi x_2) & 2\pi\sqrt{x_1}\cos(2\pi x_2) \end{pmatrix}$$

$$= \pi\cos^2(2\pi x_2) + \pi\sin^2(2\pi x_2) = \pi(\underbrace{\cos^2(2\pi x_2) + \sin^2(2\pi x_2)}_{=1}) = \pi$$

$Y = T(X)$ besitzt nach Satz 5.1 eine Dichte $f_Y(y_1, y_2) = \frac{1}{\pi}\mathbb{1}_{T(Q)}(y_1, y_2)$[8] mit $\widehat{K} = T(Q) =$"Einheitskreis ohne die positive X-Achse"
Also: $Y = T(X)$ ist gleichverteilt auf dem Einheitskreis. (Vergleiche Beispiel 4.7)

[8] $f_X(T^{-1}(y_1, y_2)) = \mathbb{1}_Q(T^{-1}(y_1, y_2)) = \mathbb{1}_{T(Q)}(y_1, y_2)$

Beispiel 5.3 (Rayleigh-Verteilung) X, Y stid $\sim N(0, \sigma^2)$, also gemeinsame Dichte $f_{(X,Y)}(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{\sigma^2}}$ mit $(x,y) \in \mathbb{R}^2$
Transformation in Polarkoordinaten
$(r, \varphi) = T(X,Y) : \mathbb{R}^2 \setminus \{0\} \to (0,\infty) \times (0, 2\pi)$ mit $r = \sqrt{x^2+y^2}$ und
$\varphi = \begin{cases} \arctan(\frac{y}{x}) & , y > 0 \\ \pi + \arctan(\frac{y}{x}) & , y < 0 \end{cases}$, wobei $\arctan(\pm\infty) = \frac{\pi}{2}$.
$(x,y) = T^{-1}(r,\varphi) = (r\cos\varphi, r\sin\varphi)$
$\det\left(\left(\frac{\partial T_i^{-1}}{\partial z_j}\right)_{1 \le i,j \le 2}\right) = \det\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = r(\cos^2\varphi + \sin^2\varphi) = r$
$(R, \Phi) = T(X,Y)$ besitzt nach Satz 5.1 eine Dichte $f_{(R,\Phi)}(r,\varphi)$
$= r\frac{1}{2\pi\sigma^2} e^{-\frac{r^2\cos^2\varphi + r^2\sin^2\varphi}{2\sigma^2}} \mathbb{1}_{(0,\infty)}(r)\mathbb{1}_{(0,2\pi)}(\varphi) = \underbrace{\frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbb{1}_{(0,\infty)}(r)}_{\text{Rayleigh}-\text{Verteilung}} \cdot \underbrace{\frac{1}{2\pi}\mathbb{1}_{(0,2\pi)}(\varphi)}_{\text{Gleichverteilung}}$
Also: R, Φ sind stochastisch unabhängig mit $f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \mathbb{1}_{(0,\infty)}(r)$,
$R \sim Ray(\sigma^2)$ **Rayleigh-Verteilung** und $f_\Phi(\varphi) = \frac{1}{2\pi}\mathbb{1}_{(0,2\pi)}(\varphi)$, $\Phi \sim R(0, 2\pi)$.

Lemma 5.4 $X = (X_1, X_2)$ Zufallsvektor, absolut-stetig mit Dichte $f_X(x_1, x_2)$. Dann ist $Y = X_1 + X_2$ absolut-stetig mit Dichte $f_Y(y) = \int_{-\infty}^{\infty} f_X(t, y-t)dt$, $y \in \mathbb{R}$.

Beweis: Mit Satz 5.1: $T(X_1, X_2) = (X_1, X_1 + X_2)$[9], $T^{-1}(Y_1, Y_2) = (Y_1, Y_2 - Y_1)$.
$\det\left(\frac{\partial T_i}{\partial x_j}\right) = \det\begin{pmatrix} \frac{\partial T_1}{\partial x_1} & \frac{\partial T_1}{\partial x_2} \\ \frac{\partial T_2}{\partial x_1} & \frac{\partial T_2}{\partial x_2} \end{pmatrix} = \det\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = 1$
$f_{T(X)}(y_1, y_2) \overset{Satz\ 5.1}{=} f_X(y_1, y_2 - y_1)$.
Es folgt: $f_Y(y) = \int_{-\infty}^{\infty} f_X(y_1, y - y_1)dy_1$. (2te Randverteilung)

Speziell in Lemma 5.4: Wenn X_1, X_2 stochastisch unabhängig
$\implies$ $f_{(X_1,X_2)} = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$. Dann ist

$$\boxed{f_{(X_1+X_2)}(y) = \int\limits_{-\infty}^{\infty} f_{X_1}(t) \cdot f_{X_2}(y-t)dt, y \in \mathbb{R}}$$

Bei stochastisch unabhängigen ZV X_1, X_2 heißt die Verteilung von $X_1 + X_2$ **Faltung** (convolution) der Verteilung von X_1 und X_2.
Bezeichnung: $P^{X_1+X_2} = P^{X_1} * P^{X_2}$ (Nur möglich, wenn X_1, X_2 s.u.!!)

Beispiel 5.5

a) X_1, X_2 stochastisch unabhängig, $X_1 \sim \Gamma(\alpha, \lambda)$, $X_2 \sim \Gamma(\beta, \lambda)$, $\alpha, \beta, \lambda > 0$.
Es gilt: $X_1 + X_2 \sim \Gamma(\alpha + \beta, \lambda)$, d.h. $\Gamma(\alpha, \lambda) * \Gamma(\beta, \lambda) = \Gamma(\alpha + \beta, \lambda)$

Beweis: Übung 8 Aufgabe 31 a

[9] Trick: Transformation T auf zwei Dimensionen "aufblasen", damit Satz 5.1 anwendbar wird

b) X_1, X_2 stochastisch unabhängig, $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$. Dann gilt: $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, d.h. $N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Beweis:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-t-\mu_2)^2}{2\sigma_2^2}} dt = \ldots = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_1^2}} e^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}}$$

c) $X \sim N(0,1)$. Dann gilt: $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

Beweis: Sei $X > 0$. $P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = \int\limits_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

$\boxed{\text{Substituiere: } u = t^2, \text{ also } t = \sqrt{u} \implies dt = \frac{1}{2\sqrt{u}} du \text{ (Ableitung nach } u)}$

$$\stackrel{Symmetrie}{=} 2\int\limits_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \frac{1}{2\sqrt{u}} du$$

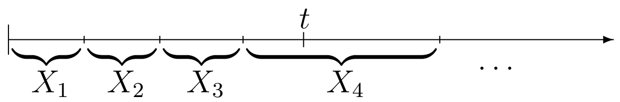
Dichte von X^2 also: $f_{X^2} = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} = \underbrace{\frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} x^{-\frac{1}{2}} e^{-\frac{x}{2}}}_{\text{Dichte einer } \Gamma(\frac{1}{2},\frac{1}{2})-\text{Vert.}} & , x \geq 0 \\ 0 & , \text{sonst} \end{cases}$

Seien $X_1, \ldots, X_n$ stid $\sim N(0,1)$. Da $X_i^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ ist, gilt nach a):
$X_1^2 + \ldots + X_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$.
$\Gamma(\frac{n}{2}, \frac{1}{2})$ heißt **χ^2-Verteilung mit n Freiheitsgraden**. Bezeichnung: χ_n^2
Dichte: $f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{(0,\infty)}(x)$

Mit Beispiel 5.3: Sei $Y \sim \chi_2^2 = \Gamma(1, \frac{1}{2}) \implies X = \sqrt{Y} \sim Ray(1)$

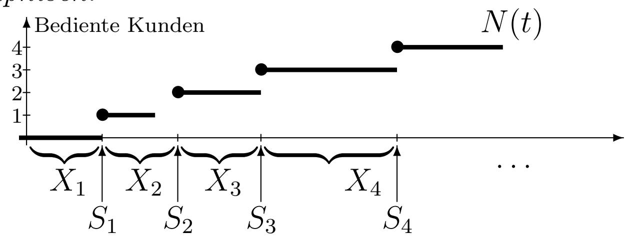
d) (siehe Beispiel 4.14) $X_1, \ldots, X_n$ stid $\sim Exp(\lambda) = \Gamma(1, \lambda)$, $\lambda > 0$.
Mit a): $S_n = X_1 + \ldots + X_n \sim \Gamma(n, \lambda)$ (Erlang-Verteilung mit Par. n, λ)
Interpretation:
X_i =”Bedienzeiten eines Servers”, stid $\sim Exp(\lambda)$
S_n =”Gesamtbedienzeit für n Anforderungen”
Umgekehrte Fragestellung: $t \geq 0$. Mit welcher Wahrscheinlichkeit werden im Intervall $[0, t]$ genau n Anforderungen bedient?

Definition 5.6 $\{X_n\}_{n \in \mathbb{N}}$ sei Folge von stochastisch unabhängigen ZV, $X_i \sim Exp(\lambda)$, $i \in \mathbb{N}$, $\lambda > 0$.



$N(t) = \max\{n \in \mathbb{N}_0 | \sum\limits_{i=1}^n X_i \leq t\} = |\{n \in \mathbb{N} | \sum\limits_{i=1}^n X_i \leq t\}|$, heißt **Poisson-Prozeß** mit Parameter $\lambda > 0$. Bezeichnung: $PP(\lambda)$

Graphisch:



Sprechweisen:
Die X_n heißen **Zwischenankunftszeiten** (interarrival times)/**Verweilzeiten** (sojourn times). Die $S_n = \sum_{i=1}^n X_i$ heißen **Ankunftszeiten** (arrival times).

Lemma 5.7 Für alle $t \geq 0$ besitzt $N(t)$ eine Poissonverteilung mit Parameter λt, d.h. $N(t) \sim Poi(\lambda t)$, $P(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k \in \mathbb{N}$.
Interpretation: Wie ist die W'keit, daß genau k Kunden bis zum Zeitpunkt t komplett bedient werden?

Beweis:

$$k = 0: \; P(N(t) = 0) = P(X_1 > t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t} = e^{-\lambda t} \cdot \frac{(\lambda t)^0}{0!}$$

$$\begin{aligned}
k \geq 1: \; P(N(t) = k) &= P(\sum_{i=1}^{k} X_i \leq t, \sum_{i=1}^{k+1} X_i > t) = P(\{S_k \leq t\} \cap \{S_{k+1} > t\} \\
&= P(\{S_k \leq t\} \setminus \{S_{k+1} \leq t\}) = P(\{S_k \leq t\}) - P(\{S_{k+1} \leq t\}) \\
&\quad \text{(wegen der Monotonie von } P) \quad \boxed{S_k \sim Erl(k,\lambda), S_{k+1} \sim Erl(k+1,\lambda)} \\
&= \frac{\lambda^k}{(k-1)!}\int_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda^{k+1}}{k!}\int_0^t x^k e^{-\lambda x}dx \\
&= \frac{\lambda^k}{(k-1)!}\left(\int_0^t x^{k-1}e^{-\lambda x}dx - \frac{\lambda}{k}\int_0^t x^k e^{-\lambda x}dx\right) \\
&= \frac{\lambda^k}{(k-1)!}(\int_0^t \underbrace{x^{k-1}}_{u'}\underbrace{e^{-\lambda x}}_{v}dx + \int_0^t \underbrace{\frac{x^k}{k}}_{u}\underbrace{(-\lambda e^{-\lambda x}}_{v'}dx) \qquad \boxed{\int_a^b uv' = uv\Big|_a^b - \int_a^b u'v} \\
&= \frac{\lambda^k}{(k-1)!}\frac{x^k}{k}e^{-\lambda x}\Big|_0^t = \frac{\lambda^k}{k!}t^k e^{-\lambda t} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}
\end{aligned}$$

Es gilt sogar für $N_{(s,t]} = N(t) - N(s)$ mit $0 \leq s \leq t$, gennant **Zuwachs in** $(s,t]$:

- $N_{(s,t]} \sim Poi(\lambda(t-s))$
- $N_{(s_i,t_i]}$ mit $i \in \mathbb{N}$ sind stochastisch unabhängig $Poi(\lambda(t_i - s_i))$-verteilt, falls die Intervalle $(s_i, t_i]$ paarweise disjunkt sind, d.h. die Zuwächse eines Poissonprozesses sind stochastisch unabhängige poissonverteilte ZV.

Im folgenden: Summen von diskreten ZV.

Lemma 5.8 X_1, X_2 stochastisch unabhängig, diskret mit Träger $\mathbb{N}_0$, Zähldichten f_{X_1}, f_{X_2}. Dann besitzt $X_1 + X_2$ die Zähldichte $f_{X_1+X_2}(k) = \sum\limits_{i=0}^{k} f_{X_1}(i) \cdot f_{X_2}(k-i)$, $k \in \mathbb{N}_0$.

Beweis: $P(X_1 + X_2 = k) = P(\bigcup\limits_{i=0}^{k}(\{X_1 = i\} \cap \{X_2 = k-i\}))$
$= \sum\limits_{i=0}^{k} P(X_1 = i, X_2 = k-i) = \sum\limits_{i=0}^{k} \underbrace{P(X_1 = i)}_{f_{X_1}(i)} \underbrace{P(X_2 = k-i)}_{f_{X_2}(k-i)}$

Beispiel 5.9 $X_1, \dots, X_n$ stid $\sim$ *Geo*(p), $0 < p < 1$.
Dann gilt $P(\sum\limits_{i=1}^{n} X_i = k) = \binom{n+k-1}{k}(1-p)^k p^n = \binom{n+k-1}{n-1}(1-p)^k p^n$, $k \in \mathbb{N}_0$.
Diese Verteilung heißt **negative Binomialverteilung**, Bezeichnung: $\overline{Bin}(n,p)$.
Also: $\underbrace{Geo(p) * \ldots * Geo(p)}_{n \text{ mal}} = \overline{Bin}(n,p)$.

Beweis: Übung 8, Aufgabe 31 d)

$X_1 + \ldots + X_n =$ Wartezeit bis zum n-ten Treffer (ohne die Treffer mitzuzählen).

Lemma 5.10

a) $Bin(n_1, p) * Bin(n_2, p) = Bin(n_1 + n_2, p)$, $n_1, n_2 \in \mathbb{N}$, $0 \le p \le 1$, insbesondere $\underbrace{Bin(1,p) * \ldots * Bin(1,p)}_{n \text{ mal}} = Bin(n,p)$
b) $\overline{Bin}(n_1, p) * \overline{Bin}(n_2, p) = \overline{Bin}(n_1 + n_2, p)$, $n_1, n_2 \in \mathbb{N}$, $0 < p < 1$
c) $Poi(\lambda_1) * Poi(\lambda_2) = Poi(\lambda_1 + \lambda_2)$

Diese drei Verteilungen sind **faltungsstabil**.

Beweis: Übung 8, Aufgabe 31

Lemma 5.11 X_1, X_2 stochastisch unabhängig, absolut-stetig mit Dichten f_{X_1}, f_{X_2}, $f_{X_i}(x) = 0$, falls $x \le 0$ für $i = 1, 2$. Dann gilt:

a) $Y = X_1 \cdot X_2$ ist absolut-stetig mit Dichte
$$f_Y(y) = \int\limits_0^\infty \tfrac{1}{t} f_{X_1}(\tfrac{y}{t}) f_{X_2}(t)\,dt\,\mathbb{1}_{(0,\infty)}(y)$$
b) $Z = \frac{X_1}{X_2}$ ist absolut-stetig mit Dichte
$$f_Z(y) = \int\limits_0^\infty t f_{X_1}(yt) f_{X_2}(t)\,dt\,\mathbb{1}_{(0,\infty)}(y)$$

Beweis: Transformationssatz für Dichten mit $T(X,Y) = (Y, X \cdot Y)$ bzw. $T(X,Y) = (Y, \frac{X}{Y})$ (siehe Übung 9 Aufgabe 34).

6 Erwartungswerte und Momente von ZV

Motivierendes Beispiel

a) Spiel: Symmetrischer Würfel
Modell: ZV X auf W'raum $(\Omega, \mathfrak{A}, P)$, $P(X = i) = \frac{1}{6}$ für $i = 1, \ldots, 6$.

- Auszahlung = Augenzahl in Euro
  Frage: "Mittlere", erwartete Auszahlung (= Einsatz beim Spiel)
  $E(X) = \frac{1}{6} \cdot 1 + \ldots + \frac{1}{6} \cdot 6 = \frac{1}{6}(1 + \ldots + 6) = \frac{21}{6} = 3,5$
- Auszahlung = Augenzahl zum Quadrat (in Euro)
  $g(X) = X^2 \qquad E(X) = \frac{1}{6} \cdot 1^2 + \ldots + \frac{1}{6} \cdot 6^2 = \frac{1}{6}(1^2 + \ldots + 6^2) = 15\frac{1}{6}$

b) Roulette (Petersburger Paradoxon) (Nikolaus Bernoulli, 1695-1726)
Strategie: Beim n-ten Spiel setze 2^n DM auf "Rot", $n \in \mathbb{N}_0$.
Bei Gewinn $\Rightarrow$ Stop Bei Verlust $\Rightarrow n \to n + 1$
Erstmalig "Rot" beim n-ten Spiel bedeutet:

$$\begin{array}{rcl} \text{Gesamteinsatz} : 1 + 2 + \ldots + 2^n & = & 2^{n+1} - 1 \\ \text{Auszahlung} : & & \underline{2^{n+1} \qquad} \\ \text{Gewinn} : & & 1 \end{array}$$

Sichere Gewinnstrategie?
Modell: $\{X_n\}_{n \in \mathbb{N}_0}$ stid, $X_n \sim Bin(1, \frac{1}{2})$, wobei $\{X_n = 1\}$: "Rot" im Spiel n.
$S = \min\{n \in \mathbb{N}_0 | X_n = 1\}$: Zeitpunkt von erstmalig "Rot"
$S \sim Geo(\frac{1}{2})$, $P(S = k) = (1 - p)^k p$, $k \in \mathbb{N}_0$.
Auszahlung: $A = 1$, falls $S < \infty$,
$P(A = 1) = P(S < \infty) = \sum\limits_{k=0}^{\infty} P(S = k) = 1$
Erwartete Auszahlung: $E(A) = 1 \cdot P(A = 1) = 1$
Ist eine sichere Gewinnstrategie, erfordert aber unendliches Kapital und unendliche Zeit.
Annahme: Maximales Kapital $= 2^L - 1$ DM $\Longrightarrow$ höchstens $L - 1$ Spiele spielbar
Auszahlung: $A = 1 \iff S \leq L - 1$ und $A = -(2^L - 1) \iff S \geq L$
$P(A = 1) = P(S \leq L - 1) = \sum\limits_{k=0}^{L-1} \frac{1}{2^{k+1}} = 1 - \frac{1}{2^L}$
$P(A = -(2^L - 1)) = P(S \geq L) = 1 - P(A = 1) = \frac{1}{2^L}$
Erwartete Auszahlung: $E(A) = 1 \cdot (1 - \frac{1}{2^L}) - (2^L - 1)\frac{1}{2^L} = 0$

In beiden Beispielen: $E(X) = \sum\limits_i i \cdot P(X = i)$, bzw. $E(g(X)) = \sum\limits_i g(i) \cdot P(X = i)$ für Trägerpunkte i.
Beachte bei der Erweiterung von E auf ∞ viele Trägerpunkte oder absolut-stetige ZV, daß $\sum$ bzw. $\int$ wohldefiniert sind.

Definition 6.1 (Erwartungswert von ZV) g sei eine reellwertige Funktion.

a) X sei eine diskrete ZV mit Träger $T = \{x_1, x_2, \dots, \} \subset \mathbb{R}$ und Zähldichte f. Falls $\sum_{i=1}^{\infty} |g(x_i)| f(x_i) < \infty$, so heißt $E(g(X)) = \sum_{i=1}^{\infty} g(x_i) f(x_i) = \sum_{i=1}^{\infty} g(x_i) P(X = x_i)$ **Erwartungswert** von $g(X)$.
b) X sei absolut-stetig mit Dichte f. Falls $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$, so heißt $E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$ **Erwartungswert** von $g(X)$.

Insbesondere $g(X) = X$:

$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$ bei diskreten ZV	$E(X^2) = \sum_{i=1}^{\infty} x_i^2 P(X = x_i)$
$E(X) = \int_{-\infty}^{\infty} x P(X = x)$ bei abs.-stetigen ZV	$E(X^2) = \int_{-\infty}^{\infty} x^2 P(X = x)$

Für beliebige ZV kann der Erwartungswert mit Hilfe der Verteilungsfunktion wie folgt berechnet werden.

Lemma 6.2 X sei ZV mit Verteilungsfunktion F. Falls $\int_{-\infty}^{0} F(x) dx < \infty$ und $\int_{0}^{\infty} 1 - F(x) dx < \infty$, so gilt:

$$E(X) = -\int_{-\infty}^{0} F(x) dx + \int_{0}^{\infty} 1 - F(x) dx$$

Beweis: Nur für den Fall differenzierbarer Verteilungsfunktion F. ($F'(x) = f(x)$ ist dann eine Dichte.)

$\int_{-\infty}^{0} \underbrace{F(x)}_{u} \cdot \underbrace{1}_{v'} dx = F(x) \cdot x\Big|_{-\infty}^{0} - \int_{-\infty}^{0} x \cdot f(x) dx = -\int_{-\infty}^{0} x \cdot f(x) dx$

$\int_{0}^{\infty} \underbrace{(1 - F(x))}_{u} \cdot \underbrace{1}_{v'} dx = (1 - F(x)) \cdot x\Big|_{0}^{\infty} - \int_{0}^{\infty} x \cdot f(x) dx = \int_{0}^{\infty} x \cdot f(x) dx$

Also: $E(X) = \int_{-\infty}^{0} x \cdot f(x) dx + \int_{0}^{\infty} x \cdot f(x) dx = -\int_{-\infty}^{0} F(x) dx + \int_{0}^{\infty} (1 - F(x)) dx$

Beispiel 6.3

a) Geometrische Verteilung, $X \sim Geo(p)$, $P(X = k) = f(k) = (1-p)^k p$, $k \in \mathbb{N}_0$, $0 < p \leq 1$.

$$E(X) = \sum_{k=0}^{\infty} k(1-p)^k p = \sum_{k=1}^{\infty} k(1-p)^k p = \sum_{k=0}^{\infty} (k+1)(1-p)^{k+1} p$$

$$= (1-p) \underbrace{\sum_{k=0}^{\infty} k(1-p)^k p}_{E(X)} + (1-p) \underbrace{\sum_{k=0}^{\infty} (1-p)^k p}_{1} = (1-p)E(X) + 1 - p$$

Also: $p \underbrace{\sum_{k=0}^{\infty} k(1-p)^k p}_{E(X)} = 1 - p \quad \Longrightarrow \quad E(X) = \frac{1-p}{p}$, $0 < p \leq 1$.

b) Exponentialverteilung, $X \sim Exp(\lambda)$, $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, $\lambda > 0$.

$$E(X) = \int_0^\infty \underbrace{x}_{u} \underbrace{\lambda e^{-\lambda x}}_{v'} dx = \underbrace{-xe^{-\lambda x}\Big|_0^\infty}_{0,\text{ da } \lim\limits_{x\to\infty} \frac{x}{e^x}=0} + \int_0^\infty e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}$$

Also: $E(X) = \frac{1}{\lambda}$, falls $X \sim Exp(\lambda)$.

c) Normalverteilung, $X \sim N(\mu, \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$. $\boxed{\text{Ersetze im 2. Schritt } x \text{ durch } x+\mu}$

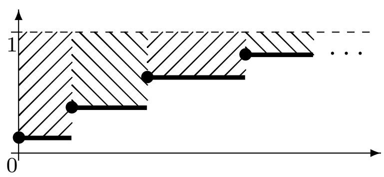
$$\begin{aligned} E(X) &= \int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty (x+\mu) e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \underbrace{\int_{-\infty}^\infty x e^{-\frac{x^2}{2\sigma^2}} dx}_{0,\text{ wegen Sym. zum Ursprung}} + \mu \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty e^{-\frac{x^2}{2\sigma^2}} dx}_{1} = \mu \end{aligned}$$

Also: $E(X) = \mu$, falls $X \sim N(\mu, \sigma^2)$.

Lemma 6.4 X sei diskrete ZV mit Träger $\mathbb{N}_0$. Dann gilt:

$$E(X) = \sum_{k=1}^\infty P(X \geq k) = \sum_{k=0}^\infty P(X > k)$$

Beweis: Mit Lemma 6.2:
$E(X) = \int_0^\infty (1 - F(x))dx$
$= \sum_{k=0}^\infty P(X > k)$
$= \sum_{k=1}^\infty P(X \geq k)$.



Satz 6.5 (Eigenschaften des Erwartungswerts)
(auftretende Erwartungswerte sollen existieren)

a) $E(aX + bY) = aE(X) + bE(Y) \qquad \forall a, b \in \mathbb{R}$
b) $X \leq Y \implies E(X) \leq E(Y)$ (Monotonie)
c) Für $X = \mathbb{1}_A$, $A \in \mathfrak{A}$, gilt: $E(X) = E(\mathbb{1}_A) = P(A)$
d) $P(|X| > c) \leq \frac{E(|X|)}{c} \qquad \forall c > 0$ (**Markoff-Ungleichung**)
e) X, Y stochastisch unabhängig $\implies E(X \cdot Y) = E(X) \cdot E(Y)$

Erwartungswerte von Funktionen von Zufallsvektoren (transformierte Zufallsvektoren) werden wie folgt berechnet.

Satz 6.6 $(X_1, \ldots, X_n)$ sei Zufallsvektor, $g : \mathbb{R}^n \to \mathbb{R}$ meßbare Funktion.

a) $(X_1,\ldots,X_n)$ sei diskret mit Träger $T=\{t_1,t_2,\ldots\}\subset\mathbb{R}^n$ und Zähldichte f. Falls $\sum\limits_{i=1}^{\infty}|g(t_i)|f(t_i)<\infty$, so gilt $E(g(X_1,\ldots,X_n))=\sum\limits_{i=1}^{\infty}g(t_i)f(t_i)$.

$\boxed{\text{Also: } E(X\cdot Y)=\sum\limits_{i=1}^{\infty}\sum\limits_{j=1}^{\infty}x_i\cdot y_i\cdot f_{(X,Y)}(x_i,y_i)}$

b) $(X_1,\ldots,X_n)$ sei absolut-stetig mit Dichte f.
Falls $\int\ldots\int|g(x_1,\ldots,x_n)|f(x_1,\ldots,x_n)dx_1\ldots dx_n<\infty$, so gilt
$E(g(X_1,\ldots,X_n))=\int\ldots\int g(x_1,\ldots,x_n)f(x_1,\ldots,x_n)dx_1\ldots dx_n$.

$\boxed{\text{Also: } E(X\cdot Y)=\int\int\ x\cdot y\cdot f_{(X,Y)}(x,y)dxdy}$

Beweis: von Satz 6.5

a) Für absolut-stetige ZV: $E(aX+bY)\overset{6.6b}{=}\int\int(ax+by)f_{(X,Y)}(x,y)dxdy$
$=a\int\int xf_{(X,Y)}(x,y)dxdy+b\int\int yf_{(X,Y)}(x,y)dxdy$
$=a\underbrace{\int xf_X(x)dx}_{E(X)}+b\underbrace{\int yf_Y(y)dy}_{E(Y)}=aE(X)+bE(Y)$

c) $\mathbb{1}_A(w)=\begin{cases}1 & ,w\in A\\ 0 & ,w\notin A\end{cases}$
$E(\mathbb{1}_A)=1\cdot P(A)+0\cdot P(A^C)=P(A)$

d) $\forall c>0$: $c\cdot\mathbb{1}_{\{|x|>c\}}(x)=\left\{\begin{array}{ll}c & ,|x|>c\\ 0 & ,|x|\le c\end{array}\right\}\le|x|$
$E(c\cdot\mathbb{1}_{\{|x|>c\}})=c\cdot P(|x|>c)\le E(|X|)$
Also $P(|x|>c)\le\frac{E(|X|)}{c}\qquad\forall c>0$.

e) Für diskrete ZV: X,Y stochastisch unabhängig
$E(X\cdot Y)=\sum\limits_{i,j}x_iy_jf_{(X,Y)}(x_i,y_i)=\sum\limits_{i,j}x_iy_jf_X(x_i)f_Y(y_j)$
$=\underbrace{(\sum\limits_i x_if_X(x_i))}_{E(X)}\underbrace{(\sum\limits_j y_jf_Y(y_j))}_{E(Y)}=E(X)\cdot E(Y)$

Definition 6.7 X,Y seien ZV. X und Y heißen **unkorreliert** (uncorrelated), wenn $E(X\cdot Y)=E(X)\cdot E(Y)$.

Also mit Satz 6.5: X,Y stochastisch unabhängig $\implies$ X,Y unkorreliert.
Die Umkehrung hiervon ist im Allgemeinen falsch, siehe Übung.[10]

Transformationen von besonderer Bedeutung sind $g(X)=X^k$ mit $k\in\mathbb{N}_0$ und $g(X,Y)=(X-E(X))(Y-E(Y))$.

Definition 6.8 X,Y seien ZV. Alle auftretenden Erwartungswerte sollen existieren.

[10] Gilt aber bei der Normalverteilung.

a) $E(X^k)$ mit $k \in \mathbb{N}_0$ heißt **k-tes Moment** (moment) von X.

b) $E((X - E(X))^k)$ heißt **k-tes zentrales Moment** (central moment).
Insbesondere $k = 2$:
$E((X - E(X))^2) = V(X) = Var(X)$ heißt **Varianz von X** (variance).
$\sqrt{V(X)}$ heißt **Standardabweichung von X** (standard deviation).

c) $Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right)$ heißt **Kovarianz von X und Y** (covariance).
$Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$ heißt **Korrelation von X und Y** (correlation).

Lemma 6.9 $X, Y, X_1, \ldots, X_n$ seien ZV. Auftretende Momente (Erwartungswerte) sollen existieren.

a) $Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$
$Cov(X, X) = Var(X)$
$Var(X) = E(X^2) - (E(X))^2$

b) $Var(aX + b) = a^2 Var(X) \qquad \forall a, b \in \mathbb{R}$

c) $Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i) + 2 \sum\limits_{i<j} Cov(X_i, Y_j)$
$Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i)$, falls $X_1, \ldots, X_n$ paarweise unkorreliert

d) $|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}$ **Cauchy-Schwarz-Ungleichung**
Insbesondere folgt: $|Corr(X, Y)| \leq 1$.

Beweis:

1. $Cov(X, X) = E[(X - E(X))(X - E(X))] = E[(X - E(X))^2] = Var(X)$
$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E[X \cdot Y - X \cdot E(Y) - Y \cdot E(X)$
$+ E(X)E(Y)] = E(X \cdot Y) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y)$
$= E(X \cdot Y) - E(X)E(Y)$
$Var(X) = Cov(X, X) = E(X^2) - E(X)^2$

2. $Var(aX + b) = E[((aX + b) - E(aX + b))^2] = E[(aX + b - aE(X) - b)^2]$
$= a^2 E[(X - E(X))^2] = a^2 Var(X)$

3. $Var(\sum\limits_{i=1}^{n} X_i) = E\left[(\sum\limits_{i=1}^{n} X_i)^2\right] - E\left[\sum\limits_{i=1}^{n} X_i\right]^2 = \sum\limits_{i,j} E(X_i X_j) - \sum\limits_{i,j} E(X_i)E(X_j)$
$= \sum\limits_{i} Var(X_i) + 2 \sum\limits_{i,j} [E(X_i X_j) - E(X_i)E(X_j)]$
$= \sum\limits_{i} Var(X_i) + 2 \sum\limits_{i<j} Cov(X_i, X_j)$

4. Hier bei Existenz von Dichten:
$|E(X, Y)| = |\int \int X \cdot Y f(x, y) dx dy| \leq \int \int |x||y| f(x, y) dx dy$

$$\overset{Cauchy-Schwarz}{\leq} \sqrt{(\int\int x^2 f(x,y)dxdy)(\int\int y^2 f(x,y)dxdy)}$$
$$= \sqrt{(\int x^2 f_X(x)dx)(\int y^2 f_Y(y)dy)} = \sqrt{E(X^2)E(Y^2)}$$

Behauptung durch Ersetzen von X durch $X-E(X)$ und Y durch $Y-E(Y)$.

Interpretation der Größen aus Lemma 6.9

$E(X)$:	mittlerer Wert
$Var(X)$:	mittlere quadratische Abweichung vom Erwartungswert. Maßzahl für die Streuung
$Cov(X,Y)$:	Korrekturterm bei der Varianzberechn. von Summen von ZV. $Var(X,Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
$Corr(X,Y)$:	Maßzahl für den linearen Zusammenhang zwischen ZV. Es gilt: $\exists a,b \in \mathbb{R} : P(X = aY + b) = 1 \quad \Leftrightarrow \quad \lvert Corr(X,Y)\rvert = 1$.

Satz 6.10 $G_{X_1}(z), G_{X_2}(z)$ seien erzeugende Funktionen von stochastisch unabhängigen diskreten ZV X_1, X_2, bzw. $L_{X_1}(s), L_{X_2}(s)$ die Laplace-Transformierten von stochastisch unabhängigen absolut-stetigen ZV $X_1, X_2 \geq 0$. Dann gilt:
$G_{X_1+X_2}(z) = G_{X_1}(z)G_{X_2}(z)$ für $|z| \leq 1$, bzw.
$L_{X_1+X_2}(s) = L_{X_1}(s)L_{X_2}(s)$ für $s \geq 0$.

Beweis:
Es gilt: $G_X(z) = \sum\limits_{k=0}^{\infty} p_k z^k = E(z^X)$ und $L_X(s) = \int\limits_0^{\infty} e^{sx} f(x)dx = E(e^{-sX})$.
Also: $G_{X_1+X_2}(z) = E(z^{X_1+X_2}) = E(\underbrace{z^{X_1}}_{s.u.}\underbrace{z^{X_2}}_{s.u.}) = E(z^{X_1})E(z^{X_2})$
$= G_{X_1}(z)G_{X_2}(z)$ für $|z| \leq 1$.
$L_{X_1+X_2}(s) = E(e^{-s(X_1+X_2)}) = E(\underbrace{e^{-sX_1}}_{s.u.}\underbrace{e^{-sX_2}}_{s.u.}) = E(e^{-sX_1})E(e^{-sX_2})$
$= L_{X_1}(s)L_{X_2}(s)$ fürs ≥ 0.

Beispiel 6.11

a) $X \sim Bin(n,p)$. Dann $G_X(z) = (1-p+pz)^n$ für $|z| \leq 1$.
$X_1 \sim Bin(n_1,p)$, $X_2 \sim Bin(n_2,p)$ stochastisch unabhängig.
$\Longrightarrow \quad G_{X_1+X_2}(z) = (1-p+pz)^{n_1}(1-p+pz)^{n_2} = (1-p+pz)^{n_1+n_2}$ ist die erzeugende Funktion einer $Bin(n_1+n_2,p)$-Verteilung.

b) $X_1 \sim Exp(\lambda)$, $L_{X_1}(s) = \frac{s}{s+\lambda}$ für $s \geq 0$.
$X_2 \sim Exp(\mu)$, $L_{X_2}(s) = \frac{s}{s+\mu}$ für $s \geq 0$.
$L_{X_1+X_2}(s) = \frac{s}{s+\lambda}\frac{s}{s+\mu} = \frac{\mu}{\mu-\lambda}\frac{\lambda}{s+\lambda}\frac{\lambda}{\lambda-\mu}\frac{\mu}{s+\mu}$ (*)
Die Dichte $f(x) = (\frac{\mu}{\mu-\lambda}\lambda e^{-\lambda x} + \frac{\lambda}{\lambda-\mu}\mu e^{-\mu x})\mathbb{1}_{0,\infty}(x)$ (**)
hat als Laplace-Transformierte (*). $X_1 + X_2$ hat die Dichte (**).

Satz 6.12 (Transformierte und Momente)

a) $G(z) = \sum\limits_{k=0}^{\infty} P(X = k)z^k$ sei erzeugende Funktion einer diskreten ZV X mit Träger $\mathbb{N}_0$. Dann gilt:
$E(X) = G'(1)$ und $\underbrace{E[X(X-1)\dots(X-k+1)]}_{k-\text{tes faktorielles Moment}} = G^{(k)}(1)$.
Also: $E(X^2) = G'(1) + G''(1)$. $\boxed{\text{Damit: } Var(X) = G'(1) + G''(1) - G'(1)^2}$
Beachte: $G^{(i)}(1) = \lim\limits_{z \to 1^-} G^{(i)}(z)$, falls $G(z)$ nicht ex. für $z > 1$.

b) $L(s) = \int_0^{\infty} e^{-sx} f(x)dx$ sei Laplace-Transformierte einer absolut-stetigen ZV $X \geq 0$ mit Dichte f. Dann gilt: $E(X) = -L'(0)$ und $E(X^k) = (-1)^k L^{(k)}(0)$.
$\boxed{\text{Damit: } Var(X) = L''(0) - L'(0)^2}$

Beweis: Nur eine intuitive "Idee", kein echter Beweis. Nur für a) $E(X)$
$G'(z) = \sum\limits_{k=0}^{n} p_k k z^{k-1}$ für $|z| < 1$ (durch gliedweises Differenzieren)
Also: $G'(1) = \sum\limits_{k=0}^{n} p_k k = E(X)$.

Satz 6.13 (Berechnung von Momenten)

a) $X \sim Bin(n, p)$ mit $0 \leq p \leq 1$.
Berechnung des Erwartungswertes:

1) $E(X) = \sum\limits_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \ldots = np$

2) $X \sim \sum\limits_{i=1}^{n} X_i$ mit X_i stid $\sim Bin(1, p)$ und $E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$
$E(X) = E(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} E(X_i) = \sum\limits_{i=1}^{n} p = np$
$\boxed{\text{Dieser Trick läßt sich bei faltungsstabilen Verteilungen anwenden!}}$

3) Erzeugende Funktion $G(z) = (1 - p + pz)^n$ und $G'(z) = n(1 - p + pz)^{n-1}p$. Also: $G'(1) = np = E(X)$.

$Var(X)$ wie in 2):
$Var(X) = Var(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} Var(X_i) = \sum\limits_{i=1}^{n} p(1-p) = np(1-p)$
$$\boxed{\begin{array}{l} Var(X_i) = E(X_i^2) - E(X_i)^2 = E(X_1^2) - p^2 = p - p^2 = p(1-p) \\ E(X_i^2) = 1^2 p + 0^2(1-p) = p \end{array}}$$

b) $X \sim Exp(\lambda)$ mit $\lambda > 0$. $L_X(s) = \frac{\lambda}{s+\lambda}$ und $L_X^{(k)}(s) = (-1)^k k! \frac{\lambda}{(s+\lambda)^{k+1}}$
$L_X^{(k)}(0) = (-1)^k k! \lambda^{-k}$
Also: $E(X^k) = k!\lambda^{-k}$, $E(X) = \frac{1}{\lambda}$ und $E(X^2) = \frac{2}{\lambda^2}$
$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

c) $X \sim N(0,1)$, d.h. normalverteilt mit $\mu = 0$ und $\sigma^2 = 1$. Dann $E(X) = 0$.
$E(X^2) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx \overset{Symmetrie}{=} \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{\infty} x^2 e^{-\frac{x^2}{2}} dx$

$\boxed{\text{Substitution: } \frac{x^2}{2} = y \text{ und } dx = \frac{dy}{\sqrt{2y}}}$

$= \frac{2}{\sqrt{2\pi}} \int\limits_0^{\infty} 2y e^{-y} \frac{1}{\sqrt{2y}} dy = \frac{2}{\sqrt{\pi}} \int\limits_0^{\infty} \frac{\frac{2}{\sqrt{2}}}{\sqrt{2}} \frac{y}{\sqrt{y}} e^{-y} dy = \underbrace{\frac{2}{\sqrt{\pi}} \int\limits_0^{\infty} y^{\frac{1}{2}} e^{-y} dy}_{\text{Integral über die Dichte einer } \Gamma(\frac{3}{2},1)-\text{Verteilung}} = 1$

Es folgt: $Var(X) = E(X^2) - E(X)^2 = 1 - 0 = 1$ (Varianz einer Standard-Normalverteilung ist 1)
$Y = \sigma X + \mu$ besitzt die Dichte $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$, d.h. $Y \sim N(\mu, \sigma^2)$.
Also $E(Y) = E(\sigma X + \mu) = \sigma^2 \underbrace{Var(X)}_{1} = \sigma^2$
Also: Bei $N(\mu, \sigma^2)$ ist μ der Erwartungswert und σ^2 die Varianz.

7 Bedingte Verteilungen und Erwartungswerte

a) X, Y diskrete ZV mit Träger T_X, T_Y, gemeinsamer Zähldichte $f_{(X,Y)}(x,y) = P(X=x, Y=y)$ für $(x,y) \in T_X \times T_Y$. Definiere die **bedingte Zähldichte von** X **unter** $Y = y$ (analog zu Definition 2.14) durch

$$f_{X|Y}(x|y) = \begin{cases} \frac{P(X=x,Y=y)}{P(Y=y)} = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{, falls } f_Y(y) > 0 \\ f_X(x) \text{ (oder bel. andere Zähldichte)} & \text{, falls } f_Y(y) = 0 \end{cases}$$

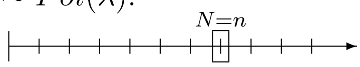
Beachte: $f_{X|Y}(x|y)$ ist eine Zähldichte für alle $y \in T_Y$ mit y fest.

Die zugehörige Vert. heißt **bedingte Verteilung von** X **unter** $Y = y$, also $P(X \in A | Y=y) = P^{X|Y=y}(A) = \sum\limits_{x \in A} f_{X|Y}(x|y)$ mit $A \in \mathfrak{A}$ und $y \in T_Y$.

Der Satz von der totalen Warscheinlichkeit (Satz 2.15) liefert

$$f_X(x) = P(X=x) = \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y) \text{ mit } x \in T_X. \qquad (*)$$

Beispiel 7.1 Sei $f_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, \ldots, n$, also $P^{X|N=n} = Bin(n,p)$ und $N \sim Poi(\lambda)$.

Zweistufiges Experiment: 

$P(X=k) = \sum\limits_{n=0}^{\infty} P(X=k|N=n)P(N=n)$ $\qquad \boxed{\text{Es ist } \binom{n}{k} = 0 \text{ für } k > n}$

$$= \sum_{n=k}^{\infty} P(X=k|N=n)P(N=n) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} \frac{\lambda^k p^k}{k!} \underbrace{\sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!}}_{e^{\lambda(1-p)}} = e^{-\lambda} \frac{\lambda^k p^k}{k!} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p} \sim Poi(\lambda p),$$

d.h. $X \sim Poi(\lambda p)$.

Betrachte den Erwartungswert der bedingten Verteilung:

$$E(g(X)|Y=y) = \sum_{x \in T_X} g(x) f_{X|Y}(x|y) \text{ für } y \in T_Y$$

heißt **bedingter Erwartungswert von** $g(X)$ **unter** $Y = y$ (falls existent).

Berechnung des Erwartungswertes mit den bedingten E-Werten wie folgt:

$$E(g(X)) = \sum_{x \in T_X} g(X) f_X(x) = \sum_{x \in T_X} g(X) \sum_{y \in T_Y} f_{X|Y}(x|y) f_Y(y)$$

$$= \sum_{y \in T_Y} \underbrace{\sum_{x \in T_X} g(X) f_{X|Y}(x|y)}_{E(g(X)|Y=y)} f_Y(y) = \sum_{y \in T_Y} E(g(X)|Y=y) f_Y(y)$$

b) X, Y absolut-stetige ZV mit gemeinsamer Dichte $f_{(X,Y)}(x,y)$.
Definiere die **bedingte Dichte von** X **unter** $Y = y$ durch

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{, falls } f_Y(y) > 0 \\ f_X(x) \text{ (oder bel. andere Zähldichte)} & \text{, falls } f_Y(y) = 0 \end{cases}$$

Beachte: $f_{X|Y}(x|y)$ ist eine Dichte für alle $y \in \mathbb{R}$.

Die zugehörige Vert. heißt **bedingte Verteilung von** X **unter** $Y = y$, also $P(X \in A|Y = y) = P^{X|Y=y}(A) = \int\limits_A f_{X|Y}(x|y)dx$ mit $A \in \mathfrak{A}$ und $y \in \mathbb{R}$.
Speziell heißt $F_{X|Y=y}(x) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(z|y)dz$ **bedingte Verteilungsfunktion von** X **unter** $Y = y$.

Wie bei a) (*): $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$.

Analog zum diskreten Fall heißt $E(g(X)|Y = y) = \int g(X)f_{X|Y}(x|y)dx$ für $y \in \mathbb{R}$ **bedingter Erwartungswert von** $g(X)$ **unter** $Y = y$.
Wie oben:

$$E(g(X)) = \int g(X)f_X(x)dx = \int g(X)\int f_{X|Y}(x|y)f_Y(y)dydx$$
$$= \int \underbrace{\left(\int g(X)f_{X|Y}(x|y)dx\right)}_{E(g(X)|Y=y)} f_Y(y)dy = \int E(g(X)|Y = y)f_Y(y)dy$$

c) Gemischte Fälle: absolut-stetig/diskret werden analog behandelt

Beispiel 7.2 (Wartezeit in einem Wartesystem)
N: Anzahl der Kunden in der Warteschlange, $N \sim Geo(p)$
X_n mit $n \in \mathbb{N}$: Bedienzeit des Kunden n, X_n stid $\sim Exp(\lambda)$
Gesucht: $W = \sum_{i=1}^{N+1} X_i$ Gesamtwartezeit für neu ankommende Kunden
Bekannt: $P^{W|N=n} = Erl(n+1, \lambda)$
Es gilt: $P(W \leq t) = \sum\limits_{n=0}^{\infty} P(W \leq t|N = n)P(N = n)$

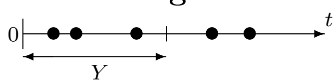
$$= \sum_{n=0}^{\infty} \int_0^t \frac{\lambda^{n+1}}{n!} y^{n-1+1} e^{-\lambda y} dy(1-p)^n p = \int_0^t \lambda p e^{-\lambda y} \underbrace{\sum_{n=0}^{\infty} \frac{(\lambda y(1-p))^n}{n!}}_{e^{\lambda y(1-p)}} dy$$

$$= \int_0^t \lambda p e^{-\lambda y} e^{\lambda y(1-p)} dy = \int_0^t \underbrace{\lambda p e^{-\lambda p y}}_{\substack{\text{Dichte einer} \\ \text{Exp-Vert.}}} dy = 1 - e^{-\lambda p t} \text{ für } t \geq 0$$

Also: $W \sim Exp(\lambda p)$ mit $E(W) = \frac{1}{\lambda p}$

Beispiel 7.3 (Ankünfte eines PP in zufälligen Intervallen)

$$\left.\begin{array}{l} N(t)\text{: } PP(\lambda) \text{ mit } \lambda > 0 \\ Y \sim Exp(\mu) \text{ mit } \mu > 0 \end{array}\right\} \text{s.u.}$$



Gesucht: $N(Y) = \max\{n \in \mathbb{N}_0 | \sum\limits_{i=1}^{n} X_i \leq Y\}$, wobei X_i stid $\sim Exp(\lambda)$.

$N(Y)$ ist Anzahl der Ankünfte im zufälligen Intervall $[0, Y]$.
Bekannt: $P^{N(Y)|y=t} = Poi(\lambda t)$ für alle $t > 0$.
$P(N(Y)=k) = \int_0^\infty P(N(Y)=k|Y=t)\mu e^{-\mu t}dt = \int_0^\infty \frac{(\lambda t)^k}{k!}e^{-\lambda t}\mu e^{-\mu t}dt$
$= \frac{\lambda^k \mu}{k!}\int_0^\infty t^k e^{-(\lambda+\mu)t}dt = \frac{\lambda^k\mu}{k!}\frac{k!}{(\lambda+\mu)^{k+1}}\underbrace{\int_0^\infty \frac{(\lambda+\mu)^{k+1}}{k!}t^k e^{-(\lambda+\mu)t}dt}_{1,\text{ da }\Gamma(k+1,\lambda+\mu)-\text{Verteilung}}$
$= \frac{\mu}{\lambda+\mu}\left(\frac{\lambda}{\lambda+\mu}\right)^k$ (für $k \in \mathbb{N}_0$) $= \left(1 - \frac{\mu}{\lambda+\mu}\right)^k \frac{\mu}{\lambda+\mu}$
Also: $N(Y) \sim Geo(\frac{\mu}{\lambda+\mu})$

8 Grenzwertsätze

Zur Motivation

1) Unendlicher Münzwurf, Wahrscheinlichkeit für Kopf $= p$ (unbekannt)
   Frage: Ist die Münze fair?
   Idee: Werfe die Münze sehr oft in unabhängigen Versuchen.
   $h_n = \frac{\text{Anzahl der Würfe mit Kopf}}{n = \text{Anzahl aller Würfe}}$
   Frage: Konvergiert h_n gegen p für $n \to \infty$?
2) Wartesystem: Anzahl der in einem Zeitraum ankommenden Kunden durch $X_i \sim Poi(\lambda)$ (Zeitintervall der Länge 1)
   Anzahl der bis zum Zeitpunkt n ankommenden Kunden durch $Y_n = \sum_{i=1}^n X_i \sim Poi(n\lambda)$
   Frage: Welche Verteilung ergibt sich asymptotisch (n groß)?
   Es existieren Konstanten: a_n, b_n: $\frac{Y_n - a_n}{b_n} \overset{asymptotisch}{\sim} N(0,1)$
   Gilt die gleiche Aussage, wenn $X_i \sim Bin(k,p)$? $\to$ Ja!
   Gilt immer, auch wenn die Verteilung der X_i unbekannt.

Zunächst geeignete Konvergenzbegriffe:

Definition 8.1 $\{X_n\}_{n\in\mathbb{N}}$ sei eine Folge von ZV und X sei ZV, alle auf W'raum $(\Omega, \mathfrak{A}, p)$. Die Folge heißt

a) **P-fast sicher konvergent** (almost surely, almost everywhere) gegen X, wenn $P(\{w \mid \lim\limits_{n\to\infty} X_n(w) = X(w)\}) = P(\lim\limits_{n\to\infty} X_n = X) = 1$.
   Bez.: $X_n \overset{n\to\infty}{\to} X$ P-fast sicher konvergent (P-f.s.), $\lim\limits_{n\to\infty} X_n = X$ P-f.s.
b) **P-stochastisch konvergent** (stochastically convergent) gegen X, wenn $\lim\limits_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad$ für alle $\varepsilon > 0$.
   Bez.: $X_n \overset{n\to\infty}{\to} X$ P-stoch., $\lim\limits_{n\to\infty} X_n = X$ P-stoch.
c) **schwach konvergent** oder **verteilungskonvergent** (convergent in distribution) gegen X, wenn $\lim\limits_{n\to\infty} F_n(x) = F(x)$ für alle Stetigkeitspunkte von F, wobei $X_n \sim F_n$ und $X \sim F$. *Bez.:* $X_n \overset{asym.}{\sim} X$, $X_n \overset{\mathfrak{D}}{\to} X$.

Lemma 8.2 $P(\limsup\limits_{n\to\infty}\{|X_n - X| > \varepsilon\}) = 0 \iff X_n \overset{n\to\infty}{\to} X$ P-f.s.
Also $X_n \overset{n\to\infty}{\to} X$ P-f.s. $\iff P(|X_n - X| > \varepsilon$ für ∞ viele $n) = 0$ für alle $\varepsilon > 0$.

Beweis: "$\Leftarrow$": ✓
"$\Rightarrow$": Für alle $\varepsilon = \frac{1}{k}$: $P(|X_n - X| > \frac{1}{k}$ für ∞ viele $n) = 0$
Es folgt: $P\Big(\underbrace{\bigcup_k \Big\{w \mid |X_n(w) - X(w)| > \frac{1}{k} \text{ für } \infty \text{ viele } n\Big\}}_{=\{w \mid \exists k \forall n_0 \exists n \geq n_0 : |X_n(w) - X(w)| > \frac{1}{k}\}}\Big) \leq \sum_k P(\ldots) = 0$

Aus dem Komplement: $P(\{w|\forall k \exists n_0 \forall n \geq n_0 : |X_n(w) - X(w)| \leq \frac{1}{k}\}) = 1$, d.h. $X_n \to X$ P-f.s.

Satz 8.3 (Zusammenhänge zwischen Konvergenzarten)

a) $X_n \to X$ P-f.s. $\overset{(i)}{\Longrightarrow}$ $X_n \to X$ P-stoch. $\overset{(ii)}{\Longrightarrow}$ $X_n \overset{\mathfrak{D}}{\to} X$ (Umkehrung i.A. falsch)

b) $\sum\limits_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$ für alle $\varepsilon > 0 \quad \Longrightarrow \quad X_n \to X$ P-f.s.

Beweis:

a) i) $\forall \varepsilon > 0$ gilt: $0 \leq \lim\limits_{n\to\infty} P(|X_n - X| > \varepsilon) \leq \lim\limits_{n\to\infty} P(\bigcup\limits_{m\geq n} \{|X_n - X| > \varepsilon\})$

$\overset{Stetigkeit}{=} P(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{m\geq n} \{|X_n - X| > \varepsilon\}) = P(\limsup\limits_{n\to\infty} \{|X_n - X| > \varepsilon\})$

$= P(|X_n - X| > \varepsilon$ für ∞ viele $n) = 0$, da $X_n \to X$ P-f.s. (Vergl. Lem. 8.2)

ii) Mathar & Pfeiffer, Lemma 2.3.2, Seite 137

b) Sei $\varepsilon > 0$. Mit dem Borel-Cantelli-Lemma (BCL) und Lemma 8.2:

$\sum\limits_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \overset{BCL}{\Longrightarrow} P(\limsup\limits_{n\to\infty} \{|X_n - X| > \varepsilon\}) = 0 \overset{L\ 8.2}{\Longrightarrow}$ Beh.

Lemma 8.4 (Tschebyscheff-Ungleichung) (Chebyshev inequality)
X sei ZV mit $Var(X) < \infty$. Dann gilt für alle $\varepsilon > 0$:

$$P(|X - E(X)| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Beweis: Mit Markoff-Ungleichung: $P(|X| > c) \leq \frac{E(|X|)}{c}$ für alle $c > 0$

$P(|X - E(X)| > \varepsilon) = P((X - E(X))^2 > \varepsilon^2) \leq \frac{E(|X - E(X)|^2)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$

Satz 8.5 (Starkes Gesetz großer Zahlen, SGGZ)[11]
$\{X_n\}_{n\in\mathbb{N}}$ sei eine Folge von paarweise unkorrelierten ZV auf einem W'raum $(\Omega, \mathfrak{A}, P)$ mit $Var(X_n) \leq M < \infty$ für alle $n \in \mathbb{N}$. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i)) \overset{n\to\infty}{\to} 0 \ P\text{-f.s.}$$

Beweis: O.B.d.A. $E(X_i) = 0$ für alle i, X_i paarweise unkorreliert.
Setze $\overline{X_n} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ (**arithmetisches Mittel**).

$\boxed{\text{Damit: } Var(\overline{X_n}) = Var(\frac{1}{n} \sum\limits_{i=1}^{n} X_i) = \frac{1}{n^2} \sum\limits_{i=1}^{n} Var(X_i) \leq \frac{1}{n^2} n \cdot M}$

[11] strong law of large numbers, LLN

Zeige: $\forall \varepsilon > 0$: $\sum\limits_{i=1}^{\infty} P(|\overline{X_n}|>\varepsilon) < \infty$. Dann $\overline{X_n} \to 0$ P-f.s. (Satz 8.3 b))

Betrachte zunächst die Teilfolge $\overline{X_{n^2}}$. Mit der Tschbyscheff-Ungleichung

$P(|\overline{X_{n^2}}|>\varepsilon) \leq \frac{Var(\overline{X_{n^2}})}{\varepsilon^2} \leq \frac{1}{n^4\varepsilon^2}n^2M = \frac{M}{n^2\varepsilon^2}$ für alle $n \in \mathbb{N}$, $\varepsilon > 0$.

Also: $\sum\limits_{i=1}^{\infty} P(|\overline{X_{n^2}}|>\varepsilon) \leq \sum\limits_{i=1}^{\infty} \frac{M}{n^2\varepsilon^2} = \frac{M}{\varepsilon^2}\sum\limits_{i=1}^{\infty}\frac{1}{n^2} < \infty \quad \Longrightarrow \quad \overline{X_{n^2}} \to 0$ P-f.s.

Für die nicht-quadratischen Terme:

Sei $n = n(k)$ definiert durch $n^2 \leq k < (n+1)^2$ für alle $k \in \mathbb{N}$.

$$Var(k\overline{X_k} - n^2\overline{X_{n^2}}) = Var(k\frac{1}{k}\sum_{i=1}^{k} X_i - n^2\frac{1}{n^2}\sum_{i=1}^{n^2} X_i) = Var(\sum_{i=n^2+1}^{k} X_i) \leq (k-n^2)M$$

Es folgt (Tsch.-Ungl.): $P(|k\overline{X_k} - n^2\overline{X_{n^2}}|\geq\varepsilon n^2) \leq \frac{(k-n^2)M}{\varepsilon^2 n^4}$

Also mit Satz 8.3 b): $\frac{k}{n(k)^2}\overline{X_k} \overset{k\to\infty}{\to} 0$ P-f.s.

Insgesamt: $\exists F_1 \in \mathfrak{A}$ mit $P(F_1) = 1$ und für alle $w \in F_1 : \overline{X_{n^2}} \to 0$ (für $n \to \infty$)

$\exists F_2 \in \mathfrak{A}$ mit $P(F_2) = 1$ und für alle $w \in F_2 : \frac{k}{n(k)^2}\overline{X_k} - \overline{X_{n(k)^2}} \to 0$ (für $k \to \infty$)

$\frac{k}{n(k)^2} \to 1$ für $k \to \infty$

Es folgt: $\overline{X_n} \to 0$ $\forall w \in F_1 \cap F_2$, wobei

$P(F_1 \cap F_2) = \underbrace{P(F_1)}_{1} + \underbrace{P(F_2)}_{1} - \underbrace{P(F_1 \cup F_2)}_{1} = 1$, d.h. $\overline{X_n} \to 0$ P-f.s.

Bemerkung 8.6

a) Wegen Satz 8.3 (fast sichere Konvergenz impliziert stochastische):

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)) \overset{n\to\infty}{\to} 0 \; P\text{-f.s.} \quad \Longrightarrow \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)) \overset{n\to\infty}{\to} 0 \; P\text{-stoch.}$$

Die rechte Aussage heißt **schwaches Gesetz großer Zahlen**

b) Das SGGZ gilt auch, wenn $\{X_n\}_{n\in\mathbb{N}}$ stid und $E(X_1)$ existiert (ohne die Existenz der Varianz zu fordern).

Dann gilt mit $\mu = E(X_1) : \frac{1}{n}\sum\limits_{i=1}^{n} \overset{n\to\infty}{\to} \mu$ P-f.s.

Beweis: sehr aufwendig..., siehe Shiryayev

Beispiel 8.7 (Anwendung des SGGZ)

$\{X_n\}_{n\in\mathbb{N}}$ sei Folge von stid ZV, $A \in \mathfrak{B}^1$, $P(X_n \in A) = p$ für alle $i \in \mathbb{N}$.

$\sum\limits_{i=1}^{n} \mathbb{1}_A(X_i)$: Anzahl des Auftretens $\{X_i \in A\}$ bis zum n-ten Versuch.

$Y_i := \mathbb{1}_A(X_i)$ sind stid mit $E(Y_i) = P(X_i \in A) = p$ für alle $i \in \mathbb{N}$

$$\boxed{\text{Da: } \frac{1}{n}\sum_{i=1}^{n}(\mathbb{1}_A(X_i) - p) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_A(X_i) - p \overset{n\to\infty}{\to} 0 \; P\text{-f.s.} \quad p \text{ nach rechts bringen:}}$$

Mit SGGZ: $\frac{1}{n}\sum\limits_{i=1}^{n} Y_i = \frac{1}{n}\underbrace{\sum_{i=1}^{n}\mathbb{1}_A(X_i)}_{\text{rel. Häufigkeit des Auftretens von } \{X_i \in A\}} \overset{n\to\infty}{\to} p$ P-f.s.

Beispiel 8.8 $\{X_n\}_{n\in\mathbb{N}}$ sei stid Folge von ZV. $\mu = EX_i$, $\sigma^2 = Var(X_i)$. Dann
$\frac{1}{n}\sum\limits_{i=1}^{n} X_i = \overline{X_n} \overset{n\to\infty}{\to} \mu$ P-f.s. (d.h. $\overline{X_n}$ ist **stark konsistenter Schätzer für** μ)
$\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \overset{n\to\infty}{\to} E(X_1^2)$ P-f.s. (wegen SGGZ; ersetze X_i durch X_i^2)
$\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X_n})^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i^2 - 2X_i \underbrace{\overline{X_n}}_{\text{fest bzgl. } \sum} + \overline{X_n}^2) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \underbrace{\frac{1}{n} 2\overline{X_n} \sum_{i=1}^{n} X_i}_{2\overline{X_n}^2} + \frac{1}{n}\sum\limits_{i=1}^{n} \overline{X_n}^2$

$= \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - (\overline{X_n})^2 \quad \to \quad E(X_1^2) - \mu^2 = Var(X_1) = \sigma^2$ (wegen SGGZ)

(D.h. $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X_n})^2$ ist ein **stark konsistenter Schätzer für** σ^2)

Satz 8.9 (Zentraler Grenzwertsatz, ZGWS, CLT)
$\{X_n\}_{n\in\mathbb{N}}$ sei stid Folge von ZV, $\mu = E(X_n)$, $\sigma^2 = Var(X_n)$ existiere. Dann gilt

$$\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \overset{asym.}{\sim} N(0,1) \qquad (n \to \infty),\ \sigma^2 > 0$$

d.h. $P(\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i - \mu) \le z) \quad \to \quad \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx \quad \forall z \in \mathbb{R}$

Eine Anwendung hiervon (siehe auch das folgende Beispiel):
$$P\left(\sum_{i=1}^{n} X_i \le z\right) = P\left(\frac{\sum\limits_{i=1}^{n} X_i - n\cdot E(X_n)}{\sqrt{n\cdot Var(X_n)}} \le \frac{z - n\cdot E(X_n)}{\sqrt{n\cdot Var(X_n)}}\right) \overset{ZGWS}{\approx} \Phi\left(\frac{z - n\cdot E(X_n)}{\sqrt{n\cdot Var(X_n)}}\right)$$

Beweis: z.B. Casella & Berger 216 ff.

Für $S_n = \sum\limits_{i=1}^{n} X_i$ gilt $E(S_n) = n\mu$, $Var(S_n) = n\sigma^2$.
$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n}\sigma}\sum\limits_{i=1}^{n}(X_i - \mu) \overset{asym.}{\sim} N(0,1) \qquad (n \to \infty)$
Dies bezeichnet man als die **standardisierte Summe**, und diese konvergiert nach Verteilung gegen die Standardnormalverteilung $N(0,1)$.

Beispiel 8.10 $\{X_n\}_{n\in\mathbb{N}}$ stid $\sim Bin(1,p)$, $0 < p < 1$.
Es gilt: $E(X_n) = p$, $Var(X_n) = p(1-p)$
Mit ZGWS: $\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \overset{asym.}{\sim} N(0,1) \qquad (n \to \infty)$
Also: $P\left(\sum\limits_{i=1}^{n} X_i \le z\right) \quad \left[= \sum\limits_{k=0}^{\lfloor z \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \quad \text{(sehr aufwendig für große } n\text{)}\right]$

$$= P\left(\underbrace{\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}}_{\overset{asym.}{\sim} N(0,1)} \leq \frac{z-np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{z-np}{\sqrt{np(1-p)}}\right), \quad \text{wobei } \Phi(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$$

(Approximation ist "brauchbar", falls $n \geq 30$)

Anwendung: Lieferung von n Teilen, W'keit, für Defekt: $p = 0,08$.
Gesucht: W'keit, daß Lieferung mehr als 9% defekte Teile enthält!

$$P\left(\sum_{i=1}^{n} X_i > 0,09n\right) = 1 - P\left(\sum_{i=1}^{n} X_i \leq 0,09n\right) \approx 1 - \Phi\left(\frac{0,09n - 0,08n}{\sqrt{n0,08(1-0,08)}}\right)$$
$$= 1 - \Phi\left(\frac{0,01}{\sqrt{0,08 \cdot 0,92}}\sqrt{n}\right) = 1 - \Phi\left(0,03686\sqrt{n}\right) = ...$$

9 Schätzfunktionen und Konfidenzintervalle

Zur Motivation:

a) Werfe Münze in unabhängigen Versuchen, W'keit für Kopf=p (unbekannt). Schätze p. Modell: $X_1, \dots, X_n$ stid $\sim Bin(1, p)$
$\widehat{p} = \widehat{p}(X_1, \dots, X_n) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i \to p$ P-f.s. (bekannt, SGGZ) ist ein vernünftiger Schätzer für p.
Beachte: $\widehat{p}(X_1, \dots, X_n)$ ist eine ZV. Setzt man **Realisationen** $x_1, \dots, x_n$ ein, so erhält man den **Schätzwert** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

b) $X_1, \dots, X_n$ stid $\sim Exp(\lambda)$. Familie der Exp-Vert., parametrische Familie
$\widehat{\vartheta} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i \to \frac{1}{\lambda} = E(X_1)$ P-f.s. (SGGZ)
$\widehat{\vartheta}$ ist ein Schätzer für $g(\lambda) = \frac{1}{\lambda} = E(X_1)$

c) In a), b) wird der Erwartungswert durch das arithmetische Mittel geschätzt. Es gibt auch schwierigere Situationen, z.B. M: Anzahl der regelmäßigen Besucher einer Webseite
M unbekannt, schätze M. Wie? Intuitives Verfahren:

- Merke die Adressen von n Besuchern (im Logfile), "markierte Besucher"
- Zu einem späteren Zeitpunkt: Merke m Adressen, bestimme die Anzahl x von bereits markierten Besuchern.

Unter entsprechenden Vorraussetzungen sollte $\frac{n}{M} \approx \frac{x}{m}$ (ungefähr gleich) sein. Ein vernünftiger Schätzer für M ist $\left\lfloor \frac{n \cdot m}{x} \right\rfloor$.
Kann dieses Vorgehen exakt begründet werden?

9.1 Methoden zur Bestimmung von Schätzern

Definition 9.1 $X_1, \dots, X_n$ seien ZV mit gemeinsamer Verteilung $P_\vartheta^{(X_1,\dots,X_n)}$, $\vartheta \in \Theta$ (**Parameterraum**). $g(\vartheta) : \Theta \to \gamma$ sei eine Funktion.
Jede Abbildung $h(X_1, \dots, X_n)$ mit Werten in γ heißt **statistische Schätzfunktion** oder **(Punkt-) Schätzer** ((point) estimator).

Beispiel 9.2

a) $X_1, \dots, X_n$ stid $\sim N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times \mathbb{R}^+$

$$g_1(\vartheta) = g_1(\mu, \sigma^2) = \mu, \quad h_1(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{OK}\checkmark$$

$$g_2(\vartheta) = g_2(\mu, \sigma^2) = \sigma^2, \quad h_2(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad \text{OK}\checkmark$$

b) $X_1, \ldots, X_n$ stid $\sim Exp(\lambda)$, $\vartheta = \lambda$, $\Theta = \mathbb{R}^+$

$g_1(\lambda) = \frac{1}{\lambda} = E(X_1)$, $\quad h_1(X_1, \ldots, X_n) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ OK✓

$g_2(\lambda) = \lambda$, $\quad h_2(X_1, \ldots, X_n) = \left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right)^{-1}$ Vernünftig?

Im Folgenden: Methoden um geeignete Schätzer zu finden.

Definition 9.3 (Maximum-Likelihood-Schätzer)
$f(x_1, \ldots, x_n|\vartheta)$ für $\vartheta \in \Theta$ sei eine Zähldichte oder Dichte.
$L(\vartheta|x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\vartheta)$ (als Funktion von ϑ bei gegebenen $x_1, \ldots, x_n$) heißt **Likelihood-Funktion**, $\widehat{\vartheta}(x_1, \ldots, x_n)$ heißt **Maximum-Likelihood-Schätzer** (MLS) (maximum likelihood estimator, MLE), falls

$$L(\widehat{\vartheta}(x_1, \ldots, x_n)|x_1, \ldots, x_n) = \sup_{\vartheta \in \Theta} L(\vartheta|x_1, \ldots, x_n) \text{ für alle } x_1, \ldots, x_n.$$

$g(\widehat{\vartheta})$ heißt MLS von $g(\vartheta)$.

Konzept: $\widehat{\vartheta}(x_1, \ldots, x_n)$ ist bei gegebenen Beobachtungen $x_1, \ldots, x_n$ der "passendste" Parameter; der, der die Beobachtungen "am wahrscheinlichsten" macht.

Bemerkungen:

a) $\widehat{\vartheta}$ zu bestimmen, ist eine Maximierungsaufgabe, die oft durch Differenzieren (und Nullsetzen der 1. Ableitung) gelöst werden kann.

b) Oft ist es günstiger, statt $L(\vartheta|x)$ die **Log-Likelihood-Funktion** $\log L(\vartheta|x)$ zu betrachten (siehe Beisp. 9.4). Das Maximum von $L(\vartheta|x)$ und $\log L(\vartheta|x)$ liegt an derselben Stelle.

Beispiel 9.4

a) $X_1, \ldots, X_n$ stid $\sim N(\mu, \sigma^2)$. Schätze $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

$$\begin{aligned} f(x_1, \ldots, x_n|\vartheta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2} \\ &= L(\vartheta|x_1, \ldots, x_n) \end{aligned}$$

$\log L(\vartheta|x_1, \ldots, x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i - \mu)^2$

Setze $\tau = \sigma^2$, partielle Abbleitung $= 0$ setzen, auflösen nach μ, τ:

(i) $\frac{\partial \log L(\vartheta|x_1, \ldots, x_n)}{\partial \mu} = \frac{2}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i - \mu) \stackrel{!}{=} 0$

(ii) $\frac{\partial \log L(\vartheta|x_1, \ldots, x_n)}{\partial \tau} = -\frac{n}{2}\frac{2\pi}{2\pi\tau} + \frac{1}{2\tau^2}\sum\limits_{i=1}^{n}(x_i - \mu)^2 \stackrel{!}{=} 0$

Aus i) $\widehat{\mu} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \overline{x}$

ii) $-\frac{n}{2}\frac{1}{\tau} + \frac{1}{2\tau^2}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = 0 \iff \tau = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$

$\widehat{\tau} = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$

Insgesamt: $(\widehat{\mu}, \widehat{\sigma^2}) = (\overline{x}, \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2)$ ist der MLS für (μ, σ^2).

(Noch zu prüfen, daß das Maximum bei $(\widehat{\mu}, \widehat{\sigma^2})$ wirklich angenommen wird.)

b) $X_1, \ldots, X_n$ stid $\sim Exp(\lambda)$. Schätze $\vartheta = \lambda \in \mathbb{R}^+$.

$f(x_1, \ldots, x_n|\lambda) = \prod\limits_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = L(\lambda|x_1, \ldots, x_n)$ für $x_i > 0$

$\log L(\lambda|x_1, \ldots, x_n) = n \log \lambda - \lambda \sum\limits_{i=1}^{n} x_i$ (Aufgabe: $\max\limits_{\lambda} \log L(\ldots)$)

$\frac{\partial \log L(\lambda|x_1,\ldots,x_n)}{\partial \lambda} = \frac{n}{\lambda} - \sum\limits_{i=1}^{n} x_i \stackrel{!}{=} 0 \iff \frac{n}{\lambda} = \sum\limits_{i=1}^{n} x_i \iff \lambda = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$

Also $\widehat{\lambda} = \frac{1}{\overline{x}}$ ist ein ML-Schätzer für λ.

(Noch zu zeigen: $\widehat{\lambda}$ ist ein Maximum von $L(\lambda|x_1, \ldots, x_n)$.)

c) $X_1, \ldots, X_n$ stid $\sim Bin(1, p)$. Schätze $p \in [0, 1]$.

$f(x_1, \ldots, x_n|p) = \prod\limits_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$

$= L(p|x_1, \ldots, x_n)$ für $x_i \in \{0, 1\}$

$\log L(p|x_1, \ldots, x_n) = (\sum\limits_{i=1}^{n} x_i) \log p + (n - \sum\limits_{i=1}^{n} x_i) \log(1-p)$

(Aufgabe: Maximum über $p \in [0, 1]$!)

$\frac{\partial \log L(p|x_1,\ldots,x_n)}{\partial p} = \frac{1}{p}\sum\limits_{i=1}^{n} x_i - \frac{1}{1-p}(n - \sum\limits_{i=1}^{n} x_i) \stackrel{!}{=} 0$

$\iff \frac{1}{p}\overline{x} - \frac{1}{1-p}(1-\overline{x}) = 0 \iff p = \overline{x}$

Also $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$ ist ein ML-Schätzer.

Eine weitere Methode zur Konstruktion von Schätzern: **Bayes-Methode**.

- Modelliere Vorkenntnisse über den Parameter ϑ durch eine Wahrscheinlichkeitsverteilung über Θ, die **a-priori**-Verteilung, beschrieben durch die Zähldichte oder Dichte $\pi(\vartheta)$.

- $f(x|\vartheta)$ für $x = (x_1, \ldots, x_n)$ sei die Zähldichte oder Dichte der Verteilung von $X_1, \ldots, X_n$ bei Vorliegen von ϑ, aufgefaßt als bedingte Verteilung von $(X_1, \ldots, X_n)$, gegeben ϑ.

- Gegeben die Beobachtungen $x = (x_1, \ldots, x_n)$. Die zu $f(\vartheta|x)$ gehörige Verteilung heißt **a-posteriori**-Verteilung von ϑ. $f(\vartheta|x)$ reflektiert den Kenntnisstand über ϑ <u>nach</u> Beobachten von $x_1, \ldots, x_n$.

Gegeben: $f(x,\vartheta)$ und die a-priori-Verteilung $\pi(\vartheta)$
Berechnung von $f(\vartheta|x)$ wie folgt (hier für Dichten):

$f(x) = \int f(x|\vartheta)\pi(\vartheta)d\vartheta$ (siehe Kapitel 7)

Also:

$$f(\vartheta|x) = \frac{f(x,\vartheta)}{f(x)} = \frac{f(x|\vartheta)\pi(\vartheta)}{\int f(x|\vartheta)\pi(\vartheta)d\vartheta} \qquad \text{(Nenner ist normierende Konstante)}$$

Ein natürlicher Schätzer für ϑ in der folgenden Definition:

Definition 9.5 Bez. wie oben. Die ZV $\widehat{\vartheta}(x)$ besitze die (Zähl-)Dichte $f(\vartheta|x)$. Dann heißt $E(\widehat{\vartheta}(x))$ **Bayes-Schätzer** von ϑ.

Bemerkung: Bei Dichten ist $\int \vartheta f(\vartheta|x)d\vartheta$ der Bayes-Schätzer.

Beispiel 9.6 Sei $X \sim Bin(n,p)$, n fest.
Die a-priori-Verteilung für p sei $Beta(\alpha,\beta)$ für $\alpha,\beta > 0$ mit Dichte
$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$ für $0 \le p \le 1$.
Es gilt: $X \sim Beta(\alpha,\beta) \implies E(X) = \frac{\alpha}{\alpha+\beta}$

Sei $x \in \{0,1,\ldots,n\}$.

$$\begin{aligned} f(p|x) &= \frac{f(x|p)\pi(p)}{\int f(x|p)\pi(p)dp} = \frac{\binom{n}{x}p^x(1-p)^{n-x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}}{\int \ldots dp} \\ &= \frac{\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{x+\alpha-1}(1-p)^{n-x+\beta-1}}{\int \ldots dp} \end{aligned}$$

Begründung: Da das Ganze eine Dichte ist, und $p^{x-\alpha-1}(1-p)^{n-x+\beta-1}$ Teil einer $Beta(x+\alpha, n-x+\beta)$-Verteilung ist, heben sich die Vorfaktoren und der Korrekturterm genau weg, so daß der passende Vorfaktor übrigbleibt.

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(n-x+\beta)}p^{x+\alpha-1}(1-p)^{n-x+\beta-1}$$

ist die Dichte einer $Beta(x+\alpha, n-x+\beta)$-Vert.

Also: $\widehat{p}(x) = \frac{x+\alpha}{n+\alpha+\beta}$ ist der Bayes-Schätzer zur a-priori-Verteilung π.
Speziell für $\alpha = \beta = 1$, wobei $Beta(1,1) = R(0,1)$: $\widehat{p}(x) = \frac{x+1}{n+2}$
Der ML-Schätzer für p lautet: $\widehat{p}_{ML}(x) = \frac{x}{n}$

9.2 Gütekriterien für Schätzer

Definition 9.7 $H = h(X_1,\ldots,X_n)$ sei ein Schätzer für $g(\vartheta) \in \mathbb{R}$.
$E_\vartheta(H - g(\vartheta))^2 = E_\vartheta(h(X_1,\ldots,X_n) - g(\vartheta))^2$ heißt **mittlerer quadratischer Fehler** (mean squared error, MSE).

MSE mißt die mittlere quadratische Abweichung vom zu schätzenden Wert $g(\vartheta)$. MSE ist eine Funktion von ϑ. Es gilt:

$$\begin{aligned}
&E_\vartheta\left[(H-g(\vartheta))^2\right] = E_\vartheta\left[(H-E_\vartheta(H)+E_\vartheta(H)-g(\vartheta))^2\right] \\
&= E_\vartheta\left[(H-E_\vartheta(H))^2 + 2\big(H-E_\vartheta(H)\big)\underbrace{\big(E_\vartheta(H)-g(\vartheta)\big)}_{konst.\ bzgl.\ \vartheta} + \underbrace{\big(E_\vartheta(H)-g(\vartheta)\big)^2}_{konst.\ bzgl.\ \vartheta}\right] \\
&= E_\vartheta\left[(H-E_\vartheta(H))^2\right] + 2\big(E_\vartheta(H)-g(\vartheta)\big)\underbrace{E_\vartheta\left[H-E_\vartheta(H)\right]}_{0} + \big(E_\vartheta(H)-g(\vartheta)\big)^2 \\
&= E_\vartheta\left[(H-E_\vartheta(H))^2\right] + \big(E_\vartheta(H)-g(\vartheta)\big)^2 \\
&= \underbrace{Var_\vartheta(H)}_{\text{Präzision}} + \underbrace{\big(Bias_\vartheta(H)\big)^2}_{\text{Schiefe}}
\end{aligned}$$

Wichtig ist der Fall, daß $Bias_\vartheta(H) = 0$.

Definition 9.8 Ein Schätzer heißt **erwartungstreu** (unbiased) für $g(\vartheta)$, wenn $Bias_\vartheta(H) = 0$ für alle $\vartheta \in \Theta$, falls also $E_\vartheta(H) = g(\vartheta)$ für alle $\vartheta \in \Theta$.

Offensichtlich gilt für erwartungstreue Schätzer:
$\mathrm{MSE}(H) = E_\vartheta(H-g(\vartheta))^2 = Var_\vartheta(H)$.

Beispiel 9.9 $X_1,\dots,X_n \sim N(\mu,\sigma^2)$, $\vartheta = (\mu,\sigma^2)$.
Sind $\overline{X}$ und $\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ erwartungstreu?

a) $\widehat{\mu} = H(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

$$E_\vartheta(\overline{X}) = E_\vartheta\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n} E_\vartheta(X_i) = \frac{1}{n}n\mu = \mu \qquad \forall\, \vartheta = (\mu,\sigma^2).$$

Also: $\overline{X}$ ist erwartungstreu für μ.

b) $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

Vorüberlegungen:

$\overline{X} \sim N(\mu,\frac{\sigma^2}{n}) \implies E(\overline{X}^2) = Var(\overline{X}) + E(\overline{X})^2 = \frac{\sigma^2}{n} + \mu^2$

$Y \sim N(0,1) \implies E(Y^2) = 1$ (siehe Beispiel 6.13 c)

$Y \sim N(0,1) \implies Z = \sigma Y + \mu \sim N(\mu,\sigma^2)$

$$E(Z^2) = E((\sigma Y+\mu)^2) = \sigma^2\underbrace{E(Y^2)}_{1} + 2\sigma\mu\underbrace{E(Y)}_{0} + \mu^2 = \sigma^2 + \mu^2$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

$$E_\vartheta(\widehat{\sigma^2}) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{E(X_i^2)}_{\sigma^2+\mu^2} - \underbrace{E(\overline{X}^2)}_{\frac{\sigma^2}{n}+\mu^2} = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} + \mu^2 = \frac{n-1}{n}\sigma^2$$

Also: $\widehat{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ ist nicht erwartungstreu für σ^2.

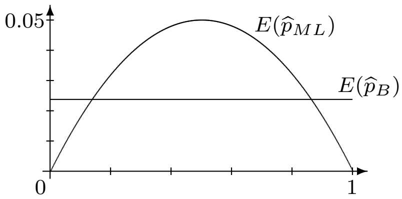
Aber: $\widehat{\sigma^2}$ ist **asymptotisch erwartungstreu**, d.h. $\lim\limits_{n\to\infty} E(\sigma^2) = \sigma^2$.

Allerdings: $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ ist erwartungstreu für σ^2.

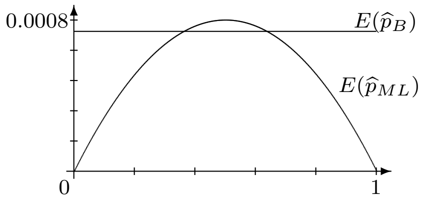
Denn: $E(S^2) = E(\frac{n}{n}S^2) = E(\frac{n}{n-1}\frac{1}{n}\sum\limits_{i=1}^{n-1}(X_i - \overline{X})^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$.

Beispiel 9.10 (Gütevergleich von 2 Schätzern) $X \sim Bin(n,p)$. (n fest)
ML-Schätzer: $\widehat{p}_{ML} = \frac{x}{n}$
Bayes-Schätzer zur a-priori-Verteilung $Beta(\alpha,\beta)$: $\widehat{p}_B = \frac{x+\alpha}{n+\alpha+\beta}$ (s. Bsp. 9.6)
Welcher ist besser? Vergleich mit MSE:
$E_p((\widehat{p}_{ML} - p)^2) = Var(\widehat{p}_{ML}) = Var(\frac{1}{n}X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$
$E_p((\widehat{p}_B - p)^2) = Var(\widehat{p}_B) + (Bias(\widehat{p}_B))^2 = Var\left(\frac{X+\alpha}{n+\alpha+\beta}\right) + \left(E_p\left(\frac{X+\alpha}{n+\alpha+\beta}\right) - p\right)^2$
$= \frac{Var(X)}{n+\alpha+\beta} + \left(\frac{E(X)+\alpha}{n+\alpha+\beta} - p\right)^2 = \frac{np(1-p)}{n+\alpha+\beta} + \left(\frac{np+\alpha}{n+\alpha+\beta} - p\right)^2$
Wähle α, β so, daß MSE($\widehat{p}_B$) konstant, d.h. kein Wert von p wird bei der Schätzung bevorzugt.
Lösung: $\alpha = \beta = \frac{1}{n}\sqrt{n}$
Dann gilt: $\widehat{p}_B = \frac{X+\frac{1}{n}\sqrt{n}}{n+\sqrt{n}}$ und $E_p((\widehat{p}_B - p)^2) = \frac{n}{4(n+\sqrt{n})^2}$

$\underline{n = 5:}$



$\underline{n = 300:}$



9.3 Konfidenzintervalle

Punktschätzer: Bestimme das $\widehat{\vartheta}$, das möglichst nahe am Parameter ϑ liegt.
Ziel jetzt: Angabe von Schranken, so daß der Parameter ϑ mit vorgegebener W'keit innerhalb der Schranken liegt.
Idee: $X_1, \ldots, X_n$ stid $\sim Bin(1,p)$. Mit schwachem GGZ:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > \varepsilon\right) \overset{n\to\infty}{\to} 0 \iff P\left(\overline{X} - \varepsilon \le p \le \overline{X} + \varepsilon\right) \overset{n\to\infty}{\to} 1 \qquad \forall\, \varepsilon > 0$$

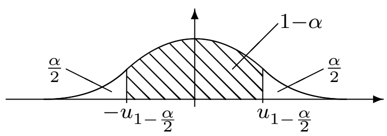
Für große n wird der Parameter p durch das zufällige Intervall $[\overline{X} - \varepsilon, \overline{X} + \varepsilon]$ mit hoher Wahrscheinlichkeit überdeckt. Nutze die Kenntnisse über die Verteilung von $\overline{X}$, um ε und $P(\ldots)$ zu quantifizieren.

Definition 9.11 $X_1, \ldots, X_n$ seien ZV mit gemeinsamer Verteilung $P^{(X_1,\ldots,X_n)}$. Ein Intervall der Form $[L(X_1,\ldots,X_n), U(X_1,\ldots,X_n)]$ heißt **Konfidenzintervall zum Niveau** $1-\alpha$ für $g(\vartheta)$, falls

$P_\vartheta\left(g(\vartheta) \in [L(X_1,\ldots,X_n), U(X_1,\ldots,X_n)]\right) \ge 1 - \alpha$ für alle $\vartheta \in \Theta$.

Beispiel 9.12 $X_1, \dots, X_n$ stid $\sim N(\mu, \sigma^2)$, σ^2 fest. Dann $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ und $\frac{n\overline{X}-n\mu}{\sqrt{\sigma^2 n}} = \frac{\overline{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$. Bestimme μ so, daß

$$P\left(\left|\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \leq u\right) = 1-\alpha \quad (*)$$



Wähle $u = u_{1-\frac{\alpha}{2}}$ als $(1-\alpha)$-Fraktil der $N(0,1)$-Verteilung. Löse (*) nach μ auf:

$$P(|\overline{X}-\mu| \leq \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}) = P(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} = 1-\alpha$$

Also: $[\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}]$ ist ein $(1-\alpha)$-Konfidenzintervall für μ.

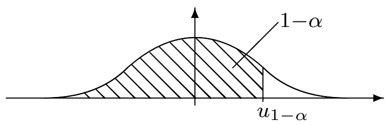
Beachte: Die Länge des Konfidenzintervalls

- fällt mit wachsendem n
- wächst mit wachsendem Niveau $1-\alpha$

Auch einseitige Konfidenzintervalle durch: $P(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq u) = 1-\alpha$

Wähle $u = u_{1-\alpha}$, dann

$$P(\overline{X} - \mu \leq \frac{\sigma}{\sqrt{n}}u_{1-\alpha})$$
$$= P(\mu \geq \overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha} = 1-\alpha$$



Also $[\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha}, \infty)$ ist einseitiges $(1-\alpha)$-Konfidenzintervall für μ.

$\boxed{\text{Das andere einseitige } (1-\alpha)\text{-Konf.-intervall durch Ansetzen des } u_\alpha\text{-Fraktils.}}$

Tabelle:

α	$0,1$	$0,05$	$0,025$	$0,01$	$0,005$	$0,0025$
$1-\alpha$	$0,9$	$0,95$	$0,975$	$0,99$	$0,995$	$0,9975$
$\mu_{1-\alpha}$	$1,282$	$1,645$	$1,960$	$2,326$	$2,576$	$2,807$

(ausführliche Tabelle in Casella & Berger Seite 608 ff)

Beispiel 9.13 (Approximatives Konfidenzintervall für W'keiten)
$X_1, \dots, X_n \sim Bin(1,p)$. Dann $\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{asym.}{\sim} N(0,1)$. (s. Bsp. 8.10)

Es gilt: $P\left(\frac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq u_{1-\frac{\alpha}{2}}\right) \approx 1-\alpha$, wobei $u_{1-\frac{\alpha}{2}}$ Fraktil der $N(0,1)$-Verteilung ist.

$|\overline{X}-p| \leq u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \iff (\overline{X}-p)^2 = u_{1-\frac{\alpha}{2}}\frac{p(1-p)}{n}$ ist eine quadratische Gleichung in p mit den Lösungen $p_L(\overline{X}) \leq p_U(\overline{X})$.

$[p_L(\overline{X}), p_U(\overline{X})]$ ist ein **approximatives** $(1-\alpha)$**-Konfidenzintervall** für p.

Index