Vorlesung: „Stochastik für Informatiker“

gehalten von *Egmar Rödel*

Beginn: 18. Oktober 1993

Zusammenfassung

Version vom *12. Oktober 1994*.

Dieses Script wurde aus Vorlesungsmitschriften von E. Rödels: „Stochastik für Informatiker“ von Torsten Hilbrich mittels LaTeX (Report-Style) erstellt. Für Fehler jeglicher Art kann keine Haftung übernommen werden.

Inhaltsverzeichnis

Abbildungsverzeichnis

Kapitel 0

Einführung

0.1 Geplanter Inhalt

1. Wahrscheinlichkeitstheorie
   - Ereignisse, Wahrscheinlichkeit
   - Zufallsgröße
   - Grenzwertsätze
   - Markoff'schen Ketten
2. Entropie, Information
   - optimale Kodierung
   - Suchbäume
3. Grundlagen der Simulation
   - Pseudozufallszahlen
     - Implementierung
     - Statistik, Güte
   - Simulated Annealing (Anwendung Markoff'sche Ketten, Simulation)

Kapitel 1

Wahrscheinlichkeitstheorie

1.1 Zufällige Ereignisse, Ereignisfelder

- Beschreibung von Experimenten
  - Kartenspiele
  - Münzwurf

Menge aller möglichen Ausgänge eines Experiments erfassen. Menge $= \Omega = \{\omega\}$

- diskret – endlich oder höchstens abzählbar viele Ausgänge
- überabzählbar viele Ausgänge

1.1 Beispiel (endliche Anzahl)
Nachfolgend zwei kleine Beispiele für Experimente mit einer endlichen Anzahl von Ereignissen

1. Münzwurf

$$\begin{aligned}
\omega_1 &= \text{z - Zahl} \\
\omega_2 &= \text{w - Wappen} \\
\Omega &= \{\omega_1, \omega_2\} \\
\mathcal{P}(\Omega) &= \{\{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \emptyset\} \\
\Omega &\text{ — } \text{sicheres Ereignis (interpretiert als } \omega_1 \text{ oder } \omega_2) \\
\Omega &= \{\omega_1\} \cup \{\omega_2\} \\
\emptyset &= \{\omega_1\} \cap \{\omega_2\} \text{ – unmögliches Ereignis}
\end{aligned}$$

2. Würfeln (1 mal) — $\Omega = 1, 2, 3, 4, 5, 6$

   1-er Kombination $\binom{6}{1} = 6$

   2-er Kombination $\binom{6}{2} = 15$

   3-er Kombination $\binom{6}{3} = 20$

   4-er Kombination $\binom{6}{4} = 15$

5-er Kombination $\binom{6}{5} = 6$

6-er Kombination $\binom{6}{6} = 1$ (sicheres Ereignis)

Menge aller Ereignisse φ: 64 Ereignisse
unmögliches Ereignis: 1

„Experiment“: Menge aller Elementarereignisse $\Omega = \{\omega_i\}$
$\mathcal{E}$ — Menge aller möglichen Ereignisse des Experiment

---

$$
\begin{array}{rcl}
\mathcal{E} & \in & \mathcal{P}(\Omega) \\
A, B & \in & \mathcal{E} \Rightarrow A \cup B = C \in \mathcal{E} \text{ (A oder B tritt ein)} \\
C & = & \{\omega \in \Omega : \omega \in A \text{ oder } \omega \in B\} \\
D & = & A \cap B = \{\omega \in \Omega : \omega \in A \text{ und } \omega \in B\} \\
\overline{A} & = & \{\omega \in \Omega : \omega \notin A\} \\
 & & \text{(A tritt nicht auf, das zu A komplementäre Ereignis)} \\
A \cup \overline{A} & = & \Omega \text{ (sicheres Ereignis)} \\
A \cap \overline{A} & = & \emptyset \text{ (unmögliches Ereignis)} \\
\overline{A} & = & \Omega \setminus A
\end{array}
$$

1.2 Beispiel
Hier nun Beispiele für Experimente mit abzählbar bzw. überabzählbar unendlich vielen Ereignissen

1. Abzählbar viele — Werfen einer Münze, bis zum ersten Mal die Zahl fällt

   $$\Omega = \{z, wz, wwz, wwz, wwwz, \ldots\}$$

2. — Ankünfte von Fahrzeugen an einer Kreuzung

   $$\Omega = \{0, 1, 2, \ldots\}$$

3. Überabzählbar viele Ausgänge
   stetige Modelle, - Experimente

   (a) Lebensdauer einer Glühbirne

   $$\Omega = [0, \infty] = \mathbb{R}^1_+$$

   Ereignisse: Intervalle, Punkte

   $$\emptyset = [0, 1] \cap [3, 5]$$

   (b) Messung einer physikalischen Konstante

   $$\underbrace{y}_{Messung} = \underbrace{m}_{Konstante} + \underbrace{\varepsilon}_{Me\ss fehler}$$

   $$\Omega = \mathbb{R}^1$$

   (c) Experimente, deren Ausgänge Funktionen sind, siehe Abbildung 1.1

   $$\Omega = \Omega' \times T \qquad \text{z.B. } T = [0, \infty]$$

   Ereignisse sind bestimmte Funktionsverläufe, siehe Abbildung 1.2
   $\Rightarrow$ stochastische Prozesse
   Sei $\omega = (\omega_0', t)$. Für festes $w' \in \Omega'$ erhalten wir eine Funktion in t, für festes $t \in T$ eine zufällige Variable (ω' variert). (siehe Abbildungen 1.1 und 1.2).

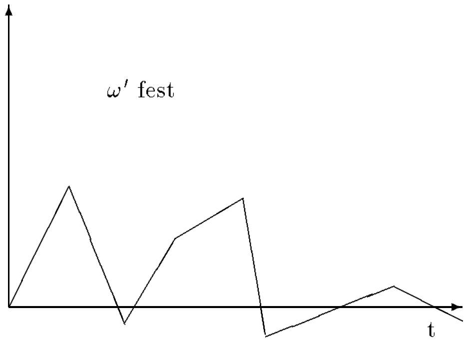

Abbildung 1.1: Fieberkurve, EKG, ...einer Person (ω')

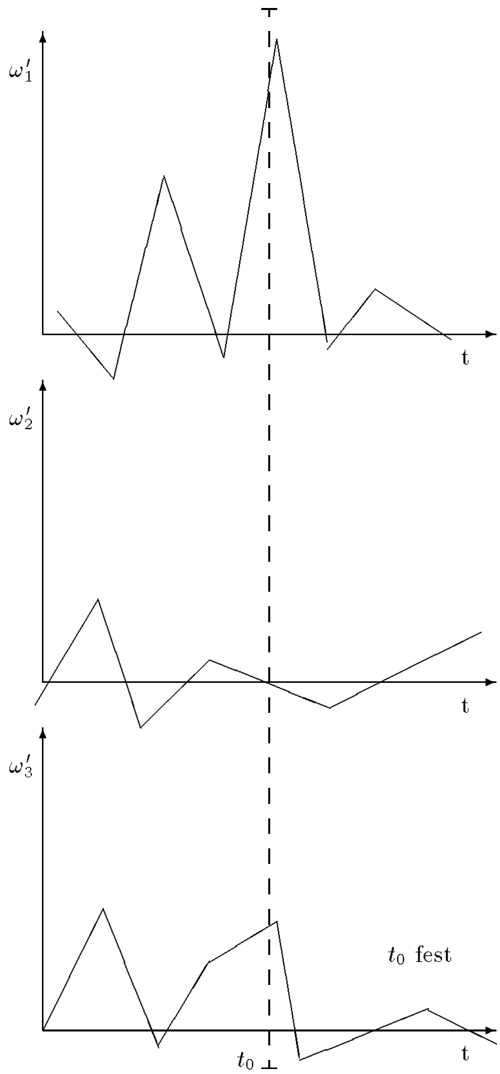

Abbildung 1.2: Die Daten mehrerer Personen

1.2 Ereignisfelder

1.1 Definition (Ereignisfelder)
Wenn Ω die Menge aller Elementarereignisse ist, so heißt $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ *Ereignisfeld* über Ω, falls

1. $\Omega \in \mathcal{E}$
2. $A_1, A_2, \ldots, A_n; \quad A_i \in \mathcal{E} \quad (i = 1, 2, \ldots, n, \ldots)$
   $\Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{E}$
3. $A \in \mathcal{E} \Rightarrow \overline{A} \in \mathcal{E}$

Folgerung aus 2.:

$$\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{E}$$

Beweis:

$$\begin{aligned} A_i \in \mathcal{E} \quad &\Rightarrow \quad \overline{A_i} \in \mathcal{E} \\ &\Rightarrow \quad \bigcap_{n=1}^{\infty} \overline{A_i} \in \mathcal{E} \\ &\Rightarrow \quad \overline{\bigcap_{n=1}^{\infty} \overline{A_i}} \in \mathcal{E} \\ &\Rightarrow \quad \bigcup_{n=1}^{\infty} A_i \in \mathcal{E} \end{aligned}$$

Folgerung aus 3.:

$$\begin{aligned} \overline{\Omega} \quad &= \quad \emptyset \in \mathcal{E} \\ &\quad\ \ \emptyset \text{ — unmögliche Ereignisse} \end{aligned}$$

Ein Ereignisfeld wird auch σ-Algebra genannt.

1.3 Kolmogoroff'sches Axiomensystem

$$\begin{aligned} \text{Abbildung} \qquad & P : \mathcal{E} \longrightarrow \mathbb{R}^1 \\ & A \in \mathcal{E} \Rightarrow P(A) \in \mathbb{R}^1 \end{aligned}$$

1.2 Definition (Wahrscheinlichkeit)
P heißt *Wahrscheinlichkeit*, falls

1. $0 \leq P(A) \leq 1 \qquad \forall A \in \mathcal{E}$
2. $P(\Omega) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, falls
   $A_i \cap A_j = \emptyset \quad \forall i \neq j$

P(A) — Wahrscheinlichkeit des Ereignisses A

$A, B \in \mathcal{E} : A \cap B = \emptyset$ — A,B unvereinbar

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

$\Rightarrow P(A \cup B) \leq P(A) + P(B)$

Folgerungen:

1. $A, B \in \mathcal{E}; A \subseteq B$

   (a) $P(B \setminus A) = P(B) - P(A)$ (Subtraktivität)

   Beweis:

$$\begin{aligned} B &= A \cup (B \setminus A) \\ P(B) &= P\left(A \cup (B \setminus A)\right) \\ & \text{weil } A \cap (B \setminus A) = \emptyset \\ &\Rightarrow \\ &= P(A) + P(B \setminus A) \end{aligned}$$

   (b) $A \subseteq B \Rightarrow P(A) \leq P(B)$ (Monotonie)

   $P(A) = P(B) - P(B \setminus A) \leq P(B)$

2. $P\left(\overline{A}\right) = 1 - P(A)$

   Beweis:

$$\begin{aligned} \Omega &= A \cup \overline{A} \\ \overline{A} &= \Omega \setminus A \\ & \text{weil } A \cup \overline{A} = \Omega \\ P(\Omega) &= 1 = P(A) + P\left(\overline{A}\right) \\ &\Rightarrow P\left(\overline{A}\right) = 1 - P(A) \end{aligned}$$

3. ist $\{A_n\}$ eine Folge von Ereignissen mit: $A_n \subseteq A_{n+1} \forall n$, dann gilt:

   $P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$ (Stetigkeit von unten)

Beweis:

$$\begin{aligned}
B_1 &:= A_1 \\
B_n &:= A_{n+1} \setminus A_n \qquad \text{(Ringe)} \\
\lim_{n\to\infty} A_n &= \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \\
& B_i \cap B_j = \emptyset \quad \forall i \neq j \qquad (*) \\
P\left(\lim_{n\to\infty} A_n\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\
& \text{wegen (*) und Axiom 3:} \\
&= \sum_{i=1}^{\infty} P(B_i) \\
&= P(A_1) + \sum_{i=1}^{\infty} P(A_{i+1} \setminus A_i) \\
&= P(A_1) + \sum_{i=1}^{\infty} (P(A_{i+1}) - P(A_i)) \\
&= \lim_{n\to\infty} \left(P(A_1) + \sum_{i=1}^{n-1} (P(A_{i+1}) - P(A_i)) \right) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

4. wenn $\{A_n\}, A_n \supseteq A_{n+1} \quad \forall n$

$$\Rightarrow P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} (P(A_n))$$

Beweis:

$$\begin{aligned}
\lim_{n\to\infty} A_n &= \bigcap_{i=1}^{\infty} A_i \\
\bigcap_{i=1}^{\infty} A_i &= \overline{\bigcup_{i=1}^{\infty} \overline{A_i}} \\
P\left(\lim_{n\to\infty} A_n\right) &= P\left(\overline{\bigcup_{i=1}^{\infty} \overline{A_i}}\right) \\
&= 1 - P\left(\bigcup_{i=1}^{\infty} \overline{A_i}\right) \\
& \text{da } A_n \supseteq A_{n+1} \\
&\Rightarrow \overline{A_n} \subseteq \overline{A_{n+1}} \\
&= 1 - P\left(\lim_{n\to\infty} \overline{A_n}\right) \\
&= 1 - \lim_{n\to\infty} P\left(\overline{A_n}\right) \\
&= 1 - \left(1 - \lim_{n\to\infty} P(A_n)\right) \\
&= \underline{\underline{\lim_{n\to\infty} P(A_n)}}
\end{aligned}$$

1.3 Definition (Wahrscheinlichkeitsraum)
Ein Tripel $(\Omega, \mathcal{E}, P)$ heißt *Wahrscheinlichkeitsraum*, wenn Ω die Menge der Elementarereignisse ist, $\mathcal{E}$ ein Ereignisfeld über Ω ist $(\subseteq \mathcal{P}(\Omega))$ und P den KOLMOGOROFF-*Axiomen* genügt.

$$
\begin{aligned}
\Omega &\quad \text{— Menge von Elementarereignissen} \\
\mathbb{a} &\subseteq \mathcal{P}(\Omega) \\
&\quad \text{— System von Teilmengen} \\
\mathcal{E}(\mathbb{a}) &= \{\mathcal{E} : \mathbb{a} \subseteq \mathcal{E}, \mathcal{E}\ \sigma\text{-Algebra}\} \\
\mathcal{E}_{\mathbb{a}} &= \bigcap_{\mathcal{E} \in \mathcal{E}_{\mathbb{a}}} \mathcal{E}
\end{aligned}
$$

$\mathcal{E}_{\mathbb{a}}$ ist die von $\mathbb{a}$ erzeugt σ-Algebra bzw. die kleinste σ-Algebra über Ω, die $\mathbb{a}$ enthält

1.3 Beispiel
Dies sind Beispiele für σ-Algebra und Wahrscheinlichkeiten.

1. $\Omega = \mathbb{R}$

$$
\begin{aligned}
(a, b) &= \{x \in \mathbb{R} : a < x < b\} \\
\mathbb{a} &= \{(a, b) : -\infty < a < b < \infty\} \\
\mathcal{E}_{\mathbb{a}} &= \mathbb{B}^1 \text{ — } \sigma\text{-Algebra der B}\textsc{orel}\text{mengen}
\end{aligned}
$$

2. $\Omega = [0, 1]$

$\mathcal{E} = \{A \subseteq [0, 1] : A \in \mathbb{B}^1\}$

$$P(A) := \int_A 1\, dx$$

$$P(\Omega) = \int_0^1 1\, dx = 1$$

3. $P(A) := \int_A 1,5(1 - x^2)\, dx$

$$P(\Omega) = 1,5 \left(x - \frac{x^3}{3}\right)\Bigg]_0^1 = 1$$

1.4 Die klassische Definition der Wahrscheinlichkeit

$$
\begin{aligned}
\Omega &= \{\omega_1, \ldots, \omega_N\} \\
P(\{\omega_i\}) &= \frac{1}{N} \\
\mathcal{E} &= \mathcal{P}(\Omega) \\
A &\in \mathcal{E} \Rightarrow P(A) = \frac{\#\{\omega : \omega \in A\}}{N}
\end{aligned}
$$

1.4 Beispiel (Paradoxon von De Méré)
Gleichzeitiges Würfeln mit 3 Würfeln

Folgende Ereignisse werden betrachtet:

A = Es fallen 11 Augen
B = Es fallen 12 Augen

Uns interessiert, wie wahrscheinlich jedes dieser Ereignisse ist.

$$P := P(\{i,j,k\}) = \frac{1}{N}$$

$$\Omega = \{(i,j,k) : 1 \leq i,j,k \leq 6\} \;, N = 6^3 = 216$$

$$P = \frac{1}{6^3} = \frac{1}{216}$$

$$\begin{array}{llcr|lllcr}
\underline{A} & 6-4-1 & +Perm & = 6 & \underline{B} & 6-5-1 & = & 6 \\
 & 6-3-2 & & = 6 & & 6-4-2 & = & 6 \\
 & 5-5-1 & & = 3 & & 6-3-3 & = & 3 \\
 & 5-4-2 & & = 6 & & 5-5-2 & = & 3 \\
 & 5-3-3 & & = 3 & & 5-4-3 & = & 6 \\
 & 4-4-3 & & = 3 & & 4-4-4 & = & 1 \\
\hline
 & & & = 27 & & & = & 25 \\
 & & P(A) = \frac{27}{216} & & & P(B) = \frac{25}{216} & &
\end{array}$$

Also ist das Werfen von 11 Augen wahrscheinlicher als das Werfen von 12 Augen.

1.5 Kombinatorik und Wahrscheinlichkeiten

1.5.1 Kombination von Elementen aus mehreren Mengen

Mengen: $S_1, \ldots, S_r$, wobei $S_i = \{s_{i_1}, \ldots, s_{i_{n_i}}\}$
Wieviele Kombinationen von Elementen $(s_{1_{i_1}}, \ldots, s_{r_{i_r}}) | s_{k_{i_k}} \in S_k$ gibt es?

$\boxed{N = n_1 \cdot \ldots \cdot n_r}$

1.5.2 Auswahl von Elementen aus einer Menge mit Zurücklegen

S enthalte n Elemente.
Wieviele Möglichkeiten gibt es, r Elemente zu entnehmen.

$$(s_{i_1}, \ldots, s_{i_r}) \qquad s_{i_1} \in S$$

Entnehme $s_{i_1} \rightarrow$ Notieren $\rightarrow$ zurücklegen

$\boxed{N = n^r}$

1.5.3 Aufteilung von r Elementen auf n Fächer ($n \geq r$)

In jedes Fach kommt höchstens eines der r Elemente

1. Element:	n	Fächer frei
2. Element:	n-1	Fächer frei
⋮		
r. Element:	n-r+1	Fächer frei

$\boxed{N = n \cdot (n-1) \cdot \ldots \cdot (n-r+1)}$

n=r $\boxed{N = n!}$ – ist die Anzahl der möglichen *Permutationen* von n-Elementen

1.5 Beispiel
r Studenten verteilen sich zufällig auf n Abteile ($n \geq r$)

A – in jedem Abteil sitzt höchstens ein Student

$$\begin{aligned} N &= n^r \\ n(A) &= \#\{\omega : \omega \in A\} \qquad \text{Anzahl der Möglichkeiten, A zu realisieren} \\ &= n \cdot (n-1) \cdot \ldots \cdot (n-r+1) \\ P(A) &= \frac{n(A)}{N} = \frac{n(n-1) \cdot \ldots \cdot (n-r+1)}{n^r} \end{aligned}$$

$$\begin{aligned} \text{S} \quad & n \text{ Elemente} \\ & r \text{ Elemente entnehmen ohne zurücklegen} \\ N \quad = \quad & n(n-1) \cdot \ldots \cdot (n-r+1) \end{aligned}$$

Reihenfolge soll nicht berücksichtigt werden, nur die Zusammensetzung.

$$N = \frac{n(n-1) \cdot \ldots \cdot (n-r+1)}{r!} = \binom{n}{r}$$

1.6 Beispiel
Urne mit n Kugeln, davon n_1 schwarze und $n - n_1$ weiße Kugeln

Wir entnehmen k Kugeln, ohne Zurücklegen

$$\Rightarrow N = \binom{n}{k}$$

Wie groß ist die Wahrscheinlichkeit, daß unter den k Kugeln k_1 schwarz sind?

$A = k_1$ schwarze Kugeln werden gezogen

$n(A) =?$

$\binom{n_1}{k_1}$ Möglichkeiten, k_1 schwarze Kugeln aus n_1 schwarzen Kugeln zu entnehmen

$\binom{n-n_1}{k-k_1}$ Möglichkeiten, $k - k_1$ weißen Kugeln aus $n - n_1$ weißen Kugeln zu entnehmen

$$\boxed{\begin{aligned} n(A) &= \tbinom{n_1}{k_1} \cdot \tbinom{n-n_1}{k-k_1} \\ P(A) &= \tfrac{n(A)}{N} \\ &= \frac{\binom{n_1}{k_1} \cdot \binom{n-n_1}{k-k_1}}{\binom{n}{k}} \\ &= P_{n_1,n}(k_1, k) \end{aligned}}$$

Der letzte Ausdruck heißt auch *Hypergeometrische Wahrscheinlichkeit.*

1.7 Beispiel (Lotto 6 aus 49)
Gezogene Kugeln werden im nachhinein als schwarze Kugeln angesehen, damit kann die Formel des Beispiels 1.6 auf Seite 10 angewendet werden.

$$\begin{aligned} A &= 4 \text{ Richtige} \\ n &= 49 \\ n_1 &= 6 \\ k &= 6 \\ k_1 &= 4 \\ P(A) &= \frac{\binom{6}{4} \cdot \binom{43}{2}}{\binom{49}{6}} \\ &= \frac{13.545}{13.983.816} \end{aligned}$$

1.6 Bedingte Wahrscheinlichkeiten, unabhängige Ereignisse

1.8 Beispiel (3-maliges Werfen einer Münze)
Definieren als erstes zwei Ereignisse und berechnen deren Wahrscheinlichkeit.

$$\begin{aligned} A &= \{\text{Wappen fällt genau einmal}\} \\ N &= 2^3 = 8 \\ A &= \{wzz, zwz, zzw\} \\ P(A) &= \frac{n(A)}{N} = \frac{3}{\underline{\underline{8}}} \\ B &= \{\text{Anzahl von Wappen ist ungerade}\} \\ &= \{wzz, zwz, zzw, www\} \\ P(B) &= \frac{4}{8} = \frac{1}{\underline{\underline{2}}} \end{aligned}$$

Angenommen, B ist schon eingetreten. Wie groß ist die Wahrscheinlichkeit, daß unter dieser Bedingung A eintritt.

$$P(A/B) = \frac{3}{4}$$

$$\Omega \Rightarrow \Omega' = B$$

1. $A \supseteq B$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

2. $A \subseteq B$

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

3. $A \cap B \neq \emptyset$ teilweise Überschneidung

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

1.4 Definition (bedingte Wahrscheinlichkeit)
Haben $A, B \in \mathcal{E}$ mit $P(B) > 0$

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

heißt *bedingte Wahrscheinlichkeit* von A unter B

$$P(A \cap B) = P(A/B) \cdot P(B)$$

1.5 Definition (Unabhängigkeit)
Zwei Ereignisse $A, B \in \mathcal{E}$ heißen *unabhängig*, wenn

$$P(A/B) = P(A)$$

$$\Rightarrow P(A \cap B) = P(A) \cdot P(B)$$

1.9 Beispiel (2-maliges Werfen einer Münze)
Unser Ereignis heißt

A = 2-mal Zahl

$$P(A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$$\Omega = \{zz, ww, zw, wz\}$$

1.10 Beispiel (Skatblatt)
Wir betrachten zwei voneinander unabhängige Ereignisse, wollen beweisen, daß sie es sind

$$\begin{aligned} A &= \text{Ziehen eines Königs} \\ P(A) &= \frac{4}{32} = \frac{1}{8} \\ B &= \text{Ziehen einer Herzkarte} \\ P(B) &= \frac{1}{4} \end{aligned}$$

Wahrscheinlichkeit eines Herz-Königs

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \underline{\underline{\frac{1}{8}}}$$

$\Rightarrow$ A und B sind voneinander unabhängig

1.1 Satz (Wahrscheinlichkeitssraum mit P_B)
Voraussetzungen

$$A \in \mathcal{E}, B \in \mathcal{E}, P(B) > 0$$

$\boxed{P_B(A) = P(A/B)}$

$\Rightarrow (\Omega, \mathcal{E}, P_B)$ ist Wahrscheinlichkeitsraum.

$$P_B(\Omega) = P(\Omega/B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

die anderen beiden Wahrscheinlichkeitsaxiome sind auch erfüllt.

1.2 Satz
Voraussetung: $A, B, C \in \mathcal{E}$

Behauptung:

$$P_B(A/C) = P(A/B \cap C)$$

Beweis:

$$\begin{aligned}
P_B(A/C) &= \frac{P_B(A\cap C)}{P_B(C)} \\
&= \frac{P(A\cap C/B)}{P(C/B)} \\
&= \frac{P(A\cap B\cap C)\cdot P(B)}{P(B)\cdot P(B\cap C)} \\
&= \underline{\underline{P(A/B\cap C)}}
\end{aligned}$$

1.6 Definition (Vollständigkeit)
Ein System von Ereignissen $\{A_i\}_{i=1}^n, A_i \in \mathcal{E}$ heißt *vollständig* (oder *ausschöpfend*), falls

1. $\bigcup\limits_{i=1}^{n} A_i = \Omega$
2. $A_i \cap A_j = \emptyset, i \neq j$

Haben $\{A_i\}_{i=1}^n$ vollständig, nehmen ein $A \in \mathcal{E}$

$$\begin{aligned}
A &= A\cap\Omega \\
&= A\cap\left(\bigcup_{i=1}^{n} A_i\right) \\
&= \bigcup_{i=1}^{n}(A\cap A_i) \\
P(A) &= P\left(\bigcup_{i=1}^{n}(A\cap A_i)\right) \\
&= \sum_{i=1}^{n} P(A\cap A_i) \\
&= \sum_{i=1}^{n}\left(P(A/A_i)\cdot P(A_i)\right) \qquad \text{(Formel der totalen Wahrscheinlichkeit)}
\end{aligned}$$

$$\begin{aligned}
P(A_i/A) &= \frac{P(A_i\cap A)}{P(A)} \\
&= \frac{P(A/A_i)\cdot P(A_i)}{\sum_{i=1}^{n}\left(P(A/A_i)\cdot P(A_i)\right)}
\end{aligned}$$

Diese Formel heißt *Formel von* Bayes (*Theorem von* Bayes).

1.6.1 Anwendung bedingter Wahrscheinlichkeiten in der KI

Es gibt verschiedene Arten, Wissen zu verarbeiten.

- logisches Schließen (boolesche Logik)
- Fuzzy-Logik
- Probabilistische Schlußweisen

- BAYES'sches Schließen
- DEMPSER-SHAFER-Kalkül

1.3 Lemma
Wenn A, B Ereignisse und A, B unabhängig sind, dann sind auch $A, \overline{B}$ und $\overline{A}, B$ sowie $\overline{A}, \overline{B}$ unabhängig.

Beweis: Am Beispiel von $A, \overline{B}$, andere äquivalent.

$$\begin{aligned} P(A/\overline{B}) &= \frac{P(A \cap \overline{B})}{P(\overline{B})} \\ &= \frac{P(A \setminus B)}{1 - P(B)} = \frac{P(A \setminus (A \cap B))}{1 - P(B)} \\ &= \frac{P(A) - P(A \cap B)}{1 - P(B)} \\ & A, B \text{ unabhängig} \iff P(A \cap B) = P(A) \cdot P(B) \\ &= \frac{P(A) - P(A) \cdot P(B)}{1 - P(B)} \\ &= \frac{P(A)(1 - P(B))}{1 - P(B)} = \underline{\underline{P(A)}} \end{aligned}$$

$$\boxed{P(A/B) = P(A/\overline{B}) = P(A), \text{ falls } A, B \text{ unabhängig}}$$

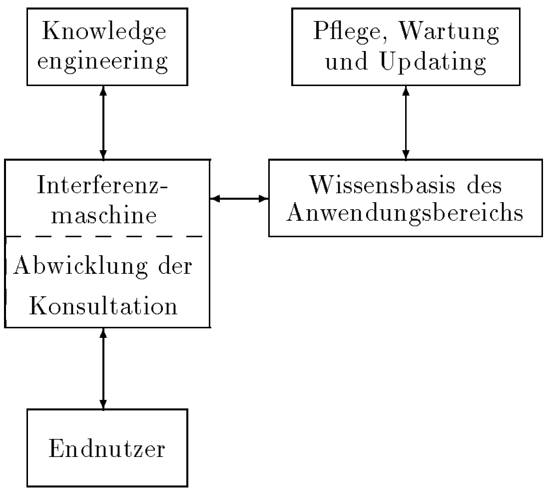

Abbildung 1.3: Aufbau eines Expertensystems

1.11 Beispiel (Beispiel MYCIN)
Aufbau der Wissensbasis:

H_i	$P_o(H_i)$	Symptom Nr 1	$P(E_1/H_i)$	$P(E_1/\overline{H_i})$
		Symptom Nr 2	$P(E_2/H_i)$	$P(E_2/\overline{H_i})$
		Symptom Nr 3	$P(E_3/H_i)$	$P(E_3/\overline{H_i})$

H_i — a-priori-Wahrscheinlichkeit für H_i
$P(E/H)$ — Wahrscheinlichkeit für das Auftreten des Symptoms E, falls H vorliegt
$P(E/\overline{H})$ — Wahrscheinlichkeit für das Auftreten des Symptoms E, falls H nicht vorliegt

1.12 Beispiel (Langzeitverhalten eines Ein-Prozessorsystem mit einer I/O-Einheit)
Wenn ein Programm beendet wird, so folgt mit Wahrscheinlichkeit p die Aktivierung von I/O-Einheit und mit Wahscheinlichkeit $q = 1 - p$ ein erneuter Programmstart. Nach Beendigung eines I/O-Vorgangs wird immer neues Programm gestartet.

Frage: Mit welcher Wahrscheinlichkeit befindet sich das System im n-ten Zyklus im Programmzustand?

$$
\begin{array}{rcl}
A_n & & \text{im n-ten Zyklus Programm} \\
\overline{A_n} & & \text{im n-ten Zyklus I/O} \\
P(A_n) & = & ? \\
\lim\limits_{n\to\infty} P(A_n) & = & ? \text{ (Langzeitverhalten)} \\
P(A_1) & = & 1 \text{ (d.h. es wird mit einem Programm begonnen)} \\
P(A_{n+1}/A_n) & = & q = 1 - p \\
P(\overline{A_{n+1}}/A_n) & = & p \\
P(\overline{A_{n+1}}/\overline{A_n}) & = & 0 \\
P(A_{n+1}/\overline{A_n}) & = & 1
\end{array}
$$

$$
\begin{array}{rcl}
q_n & = & P(A_n) \\
q_1 & = & P(A_1) = 1 \\
q_2 & = & P(A_2) = P(A_2 \cap A_1) + P(A_2 \cap \overline{A_1}) \\
 & = & P(A_2/A_1) \cdot P(A_1) + \underbrace{P(A_2/\overline{A_1}) \cdot P(\overline{A_1})}_{0} \\
 & = & q \\
q_3 & = & P(A_3/A_2) \cdot P(A_2) + P(A_3/\overline{A_2}) \cdot P(\overline{A_2}) \\
 & = & q \cdot q + 1 \cdot (1 - q) \\
 & = & (1-p)^2 + p = 1 - p + p^2 \\
q_4 & = & P(A_4/A_3) \cdot P(A_3) + P(A_4/\overline{A_3}) \cdot P(\overline{A_3}) \\
 & = & q \cdot q_3 + 1 - q_3 \\
 & = & (1-p) \cdot (1 - p + p^2) + 1 - (1 - p + p^2) \\
 & = & 1 - p + p^2 - p^3
\end{array}
$$

$$
\boxed{q_n = P(A_n) = \sum_{i=0}^{n-1} (-p)^i \qquad P(A_{n+1}) = q_{n+1} = q \cdot q_n + 1 - q_n}
$$

$$
\begin{array}{rcl}
\lim\limits_{n\to\infty} q_n & = & \underbrace{\sum_{i=0}^{\infty} (-p)^i}_{\text{geometrische Reihe}} \\
 & = & \lim\limits_{n\to\infty} \dfrac{1 - (-p)^n}{1 + p} = \underline{\underline{\dfrac{1}{1+p}}}
\end{array}
$$

1.7 Zufällige Variablen (Zufallsgrößen)

1.7 Definition (Meßbarkeit)
Wir haben zwei Tupel $(\Omega_1, \mathcal{E}_1), (\Omega_2, \mathcal{E}_2)$, bei denen Ω_i Mengen sind und $\mathcal{E}_i$ eine σ-Algebra über Ω_i. Eine Abbildung $T : \Omega_1 \to \Omega_2$ heißt $\mathcal{E}_1$- $\mathcal{E}_2$-*meßbar*, falls

$$T^{-1}(A) := \{\omega \in \Omega_1 : T(\omega) \in A\} \in \mathcal{E}_1 \quad \forall A \in \mathcal{E}_2$$

$\mathbb{R}^1, \mathbb{B}^1$ ist σ-Algebra der BORELmengen.

1.8 Definition (zufällige Variable)
ξ heißt (reellwertige) *zufällige Variable*, falls ξ eine $\mathcal{E}$-$\mathbb{B}^1$-meßbare Abbildung ist. Hierbei ist $(\Omega, \mathcal{E}, P)$ ein Wahrscheinlichkeitsraum.

$$\boxed{\begin{array}{c} \Omega_1 = \Omega \ , \ \mathcal{E}_1 = \mathcal{E} \\ \Omega_2 = \mathbb{R}^1 \ , \ \mathcal{E}_2 = \mathbb{B}^1 \end{array}}$$

1.13 Beispiel (Bernoulli-Schema)
Qualitätskontrolle

$$\begin{aligned} \Omega &= \{\text{„gut“}, \text{„schlecht“}\} \\ &= \{\omega_1, \omega_2\} \end{aligned}$$

Ausschußquote: $0 < p < 1$

$$\begin{aligned} P(\{\omega_2\}) &= p \\ P(\{\omega_1\}) &= 1 - p \\ \mathcal{E} &= \{\emptyset, \{\omega_1\}, \{\omega_2\}, \Omega\} \end{aligned}$$

$$\boxed{\xi(\omega) = \begin{cases} 1, & \text{falls } \omega = \omega_2 \\ 0, & \text{falls } \omega = \omega_1 \end{cases}}$$

$\Rightarrow \xi$ ist BOREL-meßbar

Wir entnehmen n-mal ein Produkt und prüfen jedesmal, ob die Qualität gut oder schlecht ist. Erhalten eine Menge zufälliger „Qualitätsausprägungen“ $\xi_1, \ldots, \xi_n$.

$$\xi_i = \begin{cases} 1, & \text{falls } \omega = \omega_2 \\ 0, & \text{falls } \omega = \omega_1 \end{cases}$$

Anzahl der fehlerhaften Stücke insgesamt: $z_n = \sum_{i=1}^{n} \xi_i$

$$\begin{aligned} P(\xi = 1) &= \ ? \\ P(z_n = j) &= \ ? \end{aligned}$$

ξ ist eine zufällige Variable

$$\begin{aligned} P_\xi(B) &= P(\xi \in B); B \in \mathbb{B}^1 \\ &= P(\{\omega \in \Omega : \xi(\omega) \in B\}) \\ &= P(\xi^{-1}(B)), \text{ siehe Abbildung 1.4} \end{aligned}$$

$$B = \{1\}$$

$$
\begin{aligned}
P_\xi(\{1\}) &= P(\{\omega : \xi(\omega) = 1\}) \\
&= P(\{\omega_2\}) = p \\
P_\xi(\{0\}) &= P(\{\omega : \xi(\omega) = 0\}) \\
&= P(\{\omega_1\}) \\
&= 1 - p
\end{aligned}
$$

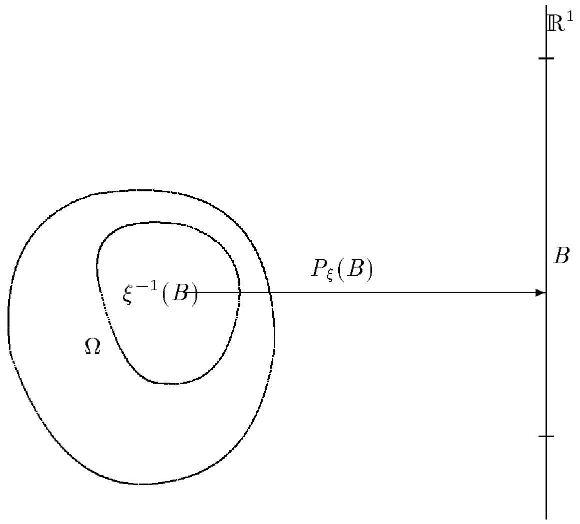

Abbildung 1.4: Die Wahrscheinlichkeitsfunktion

Das einfache Bernoulli-Schema

$$
\begin{aligned}
\xi : \Omega = \{\omega_1, \omega_2\} \to \{0, 1\} \\
P(\{\omega_2\}) = p
\end{aligned}
$$

Zweifache Bernoulli-Schema

$$
\begin{aligned}
\Omega &= \{(\omega_1,\omega_1),(\omega_1,\omega_2),(\omega_2,\omega_1),(\omega_2,\omega_2)\} \to \{0,1,1,2\} \\
z_2 &= \xi_1 + \xi_2 \\
P(\{\omega_1\}) &= 1 - p \\
P(\{\omega_2\}) &= p \\
P((\omega_1,\omega_1)) &= (1-p)^2 \\
P((\omega_1,\omega_2)) &= P((\omega_2,\omega_1)) = p(1-p) \\
P((\omega_2,\omega_2)) &= p^2
\end{aligned}
$$

$$
\begin{aligned}
P(z_2 = 0) &= P\left({z_2}^{-1}(\{0\})\right) \\
&= (1-p)^2 \\
P(z_2 = 2) &= p^2 \\
P(z_2 = 1) &= P((\omega_1,\omega_2),(\omega_2,\omega_1))
\end{aligned}
$$

$$\begin{aligned} &= P\left((\omega_1,\omega_2)\right) + P\left((\omega_2,\omega_1)\right) \\ &= 2p(1-p) \end{aligned}$$

$$\boxed{P(z_2 = i) = \binom{2}{i} p^i (1-p)^{2-i}} \quad ; i = 0, 1, 2, \ldots$$

Formel für n-faches BERNOULLI-Schema:

$$\boxed{\begin{array}{l} Z_n = \sum_{i=1}^{n} \xi_i \in \{0, 1, \ldots, n\} \\ P(Z_n = i) = \binom{n}{i} p^i (1-p)^{n-i} \end{array}}$$

$\rightarrow$ Binomialwahrscheinlichkeiten, Z_n heißt *binomialverteilte* Zufallsgröße, $Z_n \sim B(p, n)$

Bei Münzwurf gelten: Wappen = 0, Zahl = 1, p = $\frac{1}{2}$

1.9 Definition (Verteilungsfunktion)
Die Funktion $F_\xi(x) = P(\xi < x) = P(\{\omega : \xi(\omega) < x\})$, heißt *Verteilungsfunktion von* ξ, dabei ist $x \in \mathbb{R}^1$.

1.14 Beispiel

$$\begin{aligned} \xi &\sim B(\tfrac{1}{2}, 2) (\text{2-maliger Münzwurf}) \\ F_\xi(x) &= \sum_{i<x} \binom{n}{i} p^i (1-p)^{n-i}, \quad \text{wobei } p = \frac{1}{2} \text{ und } n = 2 \\ & \text{siehe Abbildung 1.5} \end{aligned}$$

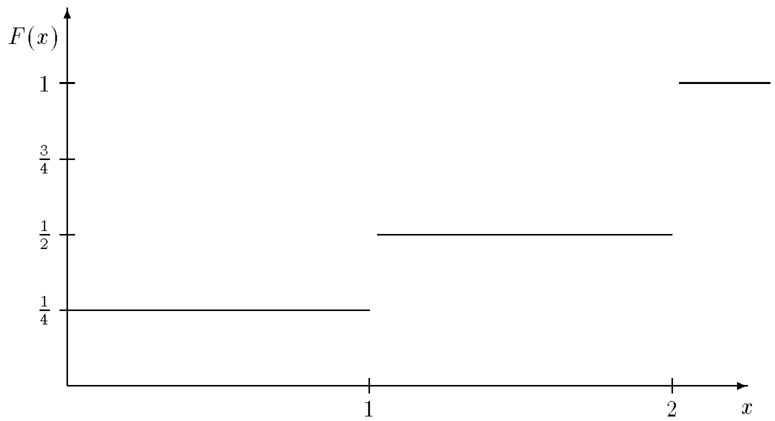

Abbildung 1.5: Einfache Verteilungsfunktion

Für $x \leq 0$ ist der Funktionswert 0, für $0 < x \leq 1$ ist er $\frac{1}{4}$, für $1 < x \leq 1$ ist er $\frac{1}{2}$. Darüberhinaus sind alle Werte 1.

1.15 Beispiel
Telefonzentrale: ω_i—i-Anrufe / Zeiteinheit.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$$

$$P(\{\omega_i\}) = \frac{\lambda^i}{i!} e^{-\lambda} \text{ — } \lambda \text{ ist ein Parameter} \qquad \Rightarrow \text{ Poisson-wahrscheinlichkeiten}$$

$$\xi : \omega_i \to i$$

$$F_\xi(x) = \sum_{i<x} \frac{\lambda^i}{i!} e^{-\lambda}$$

Daraus resultiert wieder eine Stufenfunktion, ähnlich dem vorherigem Beispiel.

1.16 Beispiel (stetige Verteilungsfunktion)
Wir haben mit ξ eine zufällige Variable und $A \subseteq [0,1]$.

$$\begin{aligned}
P(\xi \in A) &= \int_A 1\, dx \\
F_\xi(x) &= P(\xi < x) = \int_0^x 1\, dx = x \text{ , } 0 < x < 1 \\
P(\xi \notin [0,1]) &= 0 \\
F_\xi(x) &= \begin{cases} x \text{ , falls } 0 < x < 1 \\ 1 \text{ , falls } x \geq 1 \\ 0 \text{ , falls } x \leq 0 \end{cases}
\end{aligned}$$

siehe Abbildung 1.6

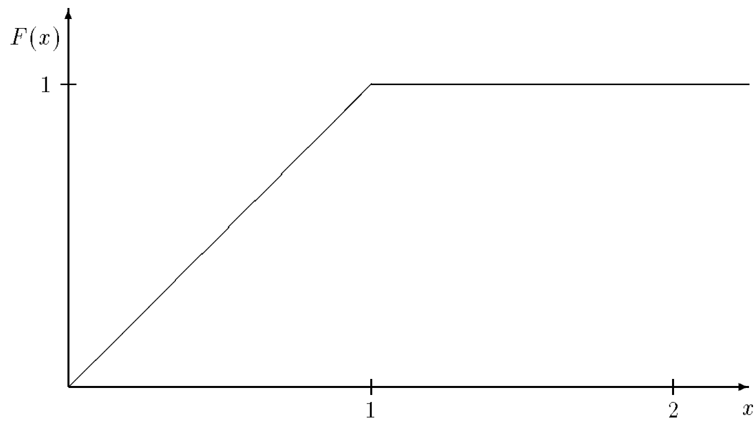

Abbildung 1.6: Eine Verteilungsfunktion im stetigem Fall

1.17 Beispiel
Messung einer physikalischen Konstante m:

$$\xi = m + \varepsilon$$

$$P(\xi < x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2}\, dt$$

Diese Verteilung stammt von Gauss und wird als Gauss-*verteilung*, *Normalverteilung* oder auch Laplace-*Verteilung* bezeichnet.

$$
\begin{array}{rcl}
 & & (\Omega, \mathcal{E}, P) \\
 & & \xi : \Omega \to \mathbb{R}^1 \\
 & & \xi : \mathcal{E}\text{-}\mathbb{B}^1\text{-meßbar, d.h.} \\
\forall B \in \mathbb{B}^1 & : & \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{E} \\
 & & P_\xi(B) = P(\xi \in B) = P\left(\{\omega : \xi(\omega) \in B\}\right) = P\left(\xi^{-1}(B)\right) \\
 & & (\Omega, \mathcal{E}, P) \xrightarrow{\xi} (\mathbb{R}^1, \mathbb{B}^1, P_\xi) \\
 & & P_\xi \text{ — die von } \xi \text{ erzeugte Wahrscheinlichkeit}
\end{array}
$$

$$
\begin{aligned}
F_\xi(x) &= F(x) = P_\xi(-\infty, x) = P(\xi \in (-\infty, x)) \\
&= P(\xi < x) \text{ — Verteilungsfunktion von } \xi
\end{aligned}
$$

1.4 Satz

Wir haben eine zufällige Variable ξ und definieren $F(x) := P(\xi < x)$. Dann gilt:

1. $F(x)$ ist monoton wachsend
2. $\lim\limits_{x\to-\infty} F(x) = 0$
   $\lim\limits_{x\to+\infty} F(x) = 1$
3. $F(x)$ ist linksseitig stetig, d.h.
   $$\lim_{x\nearrow x_0} F(x) = F(x_0)$$
4. $P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1)$

Beweis

1. sei $x_1 \leq x_2$
   definieren zwei Mengen:
   $$A_1 := \{\omega : \xi(\omega) < x_1\}$$
   $$A_2 := \{\omega : \xi(\omega) < x_2\}$$
   wegen der Meßbarkeit von ξ sind $A_i \in \mathcal{E}$. Da $A_1 \subseteq A_2$
   $$
   \begin{aligned}
   P(A_1) &= P(\xi < x_1) = F(x_1) \\
   &\leq P(A_2) = F(x_2) \\
   &\Rightarrow \underline{\underline{F(x_1) \leq F(x_2)}}
   \end{aligned}
   $$

2. $\{x_n\} \downarrow$ sei eine monoton fallende Folge mit $\lim\limits_{n\to\infty} x_n = -\infty$. Weiterhin sei $\{y_n\} \uparrow$ gegeben mit $\lim\limits_{n\to\infty} y_n = +\infty$.
   Definieren
   $$
   \begin{aligned}
   A_n &:= \{\omega : \xi(\omega) < x_n\} \downarrow \\
   & \quad\ \text{d.h. } A_i \supset A_{i+1} \\
   B_n &:= \{\omega : \xi(\omega) < y_n\} \uparrow \\
   A &= \bigcap_{n=1}^{\infty} A_n = \emptyset \\
   B &= \bigcup_{n=1}^{\infty} B_n = \left\{\omega : \xi(\omega) < \sup_i\{y_i\}\right\} = \Omega
   \end{aligned}
   $$

$$
\begin{aligned}
\lim_{n\to\infty} F(x_n) &= \lim_{n\to\infty} P(A_n) = P(A) = 0 \\
\lim_{n\to\infty} F(y_n) &= \lim_{n\to\infty} P(B_n) = P(B) = 1 \\
&\quad \text{Grenzwerte sind von der Wahl der Folgen } \{x_n\}, \{y_n\} \text{ unabhängig} \\
&\Rightarrow \underline{\underline{\lim_{x\searrow-\infty} F(x) = 0}} \\
&\quad \underline{\underline{\lim_{x\nearrow+\infty} F(x) = 1}}
\end{aligned}
$$

3. Definieren

$$
\begin{aligned}
A &= \{\omega : \xi(\omega) < x_0\} \\
A_n &= \{\omega : \xi(\omega) < x_n\} \\
&\quad x_n \nearrow x_0 \\
&\Rightarrow \{A_n\} \uparrow : A_n \subseteq A_{n+1} \qquad A_n \nearrow A \\
\lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty} F(x_n) \text{ ,da } P(\xi < x_n) = P(A_n) = F(x_n) \\
&= P\left(\lim_{n\to\infty} A_n\right) = P(A) = F(x_0) \\
&\Rightarrow \underline{\underline{\lim_{x\nearrow x_o} F(x) = F(x_0)}}
\end{aligned}
$$

4. $\{\omega : x_1 \le \xi(\omega) < x_2\} = \{\omega : \xi(\omega) < x_2\} \setminus \{\omega : \xi(\omega) < x_1\}$

   Daraus folgt:

$$
\begin{aligned}
P(\xi_1 \le \xi \le x_2) &= P\left(\{\omega : \xi(\omega) < x_2\} \setminus \{\omega : \xi(\omega) < x_1\}\right) \\
&= P\left(\{\omega : \xi(\omega) < x_2\}\right) - P\left(\{\omega : \xi(\omega) < x_1\}\right) \\
&\quad \text{wegen der Substraktivität von P} \\
&= F(x_2) - F(x_1)
\end{aligned}
$$

1.7.1 Typen von Zufallsgrößen

1. diskrete Zufallsgrößen

$$
\begin{aligned}
\xi : \Omega &\to \underbrace{\{x_1, x_2, \ldots, x_n, \ldots\}}_{\text{abzählbare Teilmenge des } \mathbb{R}^1} = W \subset \mathbb{R}^1 \\
p_i &= P\left(\{\omega : \xi(\omega) = x_i\}\right) = P(\xi = x_i) = P(A_i) \\
\xi &: \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_n & \ldots \\ p_1 & p_2 & p_3 & \ldots & p_n & \ldots \end{pmatrix} \\
A_i &:= \{\omega : \xi(\omega) = x_i\} \\
&\quad A_i \cap A_j = \emptyset \\
P(\xi = x) &= \begin{cases} p_i & x = x_i \\ 0 & x \ne x_i \quad \forall x_i \in W \end{cases} \\
\underline{F(x)} &= P(\xi < x) = P\left(\bigcup_{i:x_i<x} A_i\right) = \sum_{i:x_i<x} P(A_i) = \underline{\sum_{i:x_i<x} p_i} \\
F(x) &= \begin{cases} 0 & \text{, falls } x \le x_1 (x_1 < x_2 < x_3 \ldots) \\ \sum\limits_{i:x_i<x} p_i & \text{, falls } x_1 < x \end{cases}
\end{aligned}
$$

$$\boxed{\sum_i p_i = 1}$$

1.18 Beispiel
Für solche Funktionen $F(x)$ mit diskreten Zufallsgrößen.

$$\begin{aligned}
\xi \quad &: \quad \begin{pmatrix} 0 & 1 & 2 & \dots & n \\ p_0 & p_1 & p_2 & \dots & p_n \end{pmatrix} \sim B(p,n) \\
& \quad p_i = \binom{n}{i} \cdot p^i (1-p)^{n-i}, \text{ falls } 0 < p < 1 \\
\sum_{i=0}^{n} p_i \quad &= \quad \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p+1-p)^n = 1
\end{aligned}$$

1.19 Beispiel

$$\begin{aligned}
\xi \quad &: \quad \begin{pmatrix} 0 & 1 & 2 & \dots & n & \dots \\ p_0 & p_1 & p_2 & \dots & p_n & \dots \end{pmatrix} \sim \mathrm{PO}(\lambda) \text{ — Poisson-verteilt} \\
& \quad p_n = \frac{\lambda^n}{n!} e^{-\lambda} \\
\sum_{n=0}^{\infty} p_n \quad &= \quad \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}}_{e^{\lambda}} = 1
\end{aligned}$$

$$B \in \mathbb{B}^1 : P(\xi^{-1}B) = \sum_{i : x_i \in B} p_i \qquad \left(= \int_B d\,F(x) \right)$$

2. stetige Zufallsgrößen

$$P(\xi < x) \quad = \quad F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad (*)$$

1.10 Definition (stetig)
Eine zufällige Variable ξ heißt *stetig*, falls eine nichtnegative Funktion f existiert, so daß für $F(x)$ (*) (Seite 22) gilt.

Es gilt: $\lim\limits_{x \to \infty} F(x) = \int_{-\infty}^{+\infty} f(t)\,dt = 1$

1.11 Definition (Dichtefunktion)
Die Funktion f heißt *Dichtefunktion*, wenn

(a) $f(x) \geq 0 \quad \forall x$
(b) $\int_{\mathbb{R}} f(x)\,dx = 1$

$$\begin{aligned}
P(\xi = x) \quad &= \quad \int_x^x f(t)\,dt = 0, \text{ sogar wenn x möglich ist} \\
P(a \leq \xi \leq b) \quad &= \quad \int_a^b f(t)\,dt
\end{aligned}$$

1.20 Beispiel
Gleichverteilung über $[0, 1)$

$$\begin{aligned}
& \xi : \Omega \to [0,1) \\
F(x) &= P(\xi < x) = \begin{cases} 0 \text{ , falls } x \leq 0 \\ x \text{ , falls } 0 < x < 1 \\ 1 \text{ , falls } x \geq 1 \end{cases} \\
f(x) &= \begin{cases} 0 \text{ , falls } x \leq 0 \\ 1 \text{ , falls } 0 < x < 1 \\ 0 \text{ , falls } x \geq 1 \end{cases} \\
& \text{siehe Abbildung 1.7} \\
P(\xi < x) &= P(\xi \leq x) \text{ trifft im Gegensatz zu den diskreten Zufallsgößen zu}
\end{aligned}$$

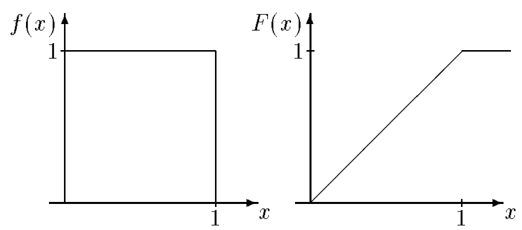

Abbildung 1.7: Die Funktionen f und F für Gleichverteilung über $[0, 1)$

$$\boxed{f(x) = \frac{d\,F(x)}{dx} \text{ , falls } f(x) \text{ stetig ist}}$$

1.7.1.1 Veranschaulichen der Dichte-Funktion

Teilen den Wertebereich von ξ in Intervalle I_j ein und beobachten für jeden der Versuche ξ_i, in welches I_j ξ_i $(i = 1, 2, \ldots, n)$ fällt.

$$\begin{aligned}
n_j &= \#\{\xi_i \in I_j\} \\
\Delta(I_j) &= \Delta_j \qquad \text{Länge von } I_j \\
& \Delta_0 = \max_j \{\Delta_j\} \\
f\text{emp.}(x) &= \frac{\frac{n_i}{n}}{\Delta_i} \text{ für } x \in I_i \\
\lim_{\substack{n \to \infty \\ \Delta_0 \to 0}} f\text{emp.}(x) &= \underline{\underline{f(x)}} \\
& \text{siehe Abbildung 1.8}
\end{aligned}$$

1.21 Beispiel

$$F(x) := F_\xi(x)$$

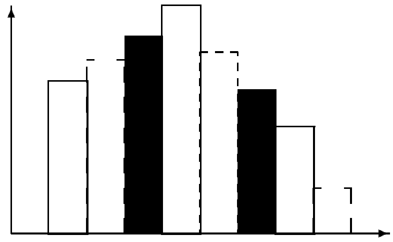

Abbildung 1.8: Durch Experiment ermittelte Dichtefunktion

$$= \int_{-\infty}^{x} \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}}_{f(t)\geq 0 (\sigma>0)} \, dt$$

zu zeigen:

$$I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx = 1$$

$$I^2 = \int\int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma^2}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \, dx\, dy$$

$$s = \frac{x-\mu}{\sigma} \qquad dx = \sigma\, ds$$

$$t = \frac{y-\mu}{\sigma} \qquad dy = \sigma\, dt$$

$$= \frac{\sigma^2}{2\pi\sigma^2} \int\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(s^2+t^2)}\, ds\, dt$$

$$s = r\cos\varphi, t = r\sin\varphi$$

$$0 < r < \infty$$

$$0 < \varphi < 2\pi$$

$$|J| = \begin{vmatrix} \frac{\partial s}{\partial r} & \frac{\partial s}{\partial \varphi} \\ \frac{\partial t}{\partial r} & \frac{\partial t}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix}$$

$$= r\cos^2\varphi + r\sin^2\varphi$$

$$= \underline{\underline{r}}$$

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} |J|\, dr\, d\varphi$$

$$(s^2 + t^2 = r^2)$$

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{\infty} r\, e^{-\frac{r^2}{2}}\, dr \right) d\varphi \qquad \left(e^{-\frac{r^2}{2}}\right)' = -r \cdot e^{-\frac{r^2}{2}}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left[-e^{-\frac{r^2}{2}} \right]_0^{\infty} d\varphi$$

$$= \underline{\underline{\frac{1}{2\pi} \int_0^{2\pi} d\varphi = 1}}$$

ξ	zufällige Variable
ξ	benutzt Dichte $f_{\mu,\sigma}(x)$

$$\begin{array}{rcl}
\xi & \sim & N(\mu,\sigma) \text{ — Normalverteilung mit den Parametern } \mu \text{ und } \sigma \\
 & & \sigma \text{ ist ein Maß für die Streung von } \xi \text{ um } \mu \\
 & & \text{siehe Abbildung 1.9} \\
\underbrace{\xi}_{\text{Messung}} & = & \underbrace{\mu}_{\text{Konstante}} + \underbrace{\varepsilon}_{\text{Meßfehler}} \\
 & & \epsilon \sim N(0,\sigma),\ \sigma \text{ beschreibt die Genauigkeit der Messung } \xi
\end{array}$$

ξ besitzt eine Standardnormalverteilung, wenn

$$\begin{array}{rcl}
 & & \xi \sim N(0,1) \\
F(x) & = & P(\xi < x) = \Phi(x) \\
 & = & \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \\
 & \Rightarrow & \text{GAUSS'sche Glockenkurve}
\end{array}$$

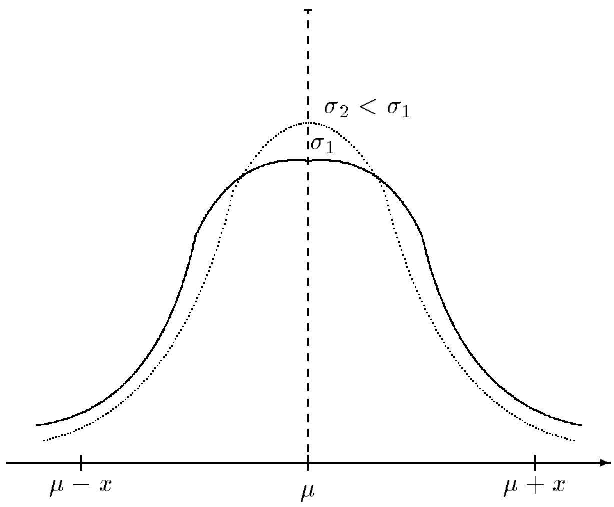

Abbildung 1.9: Die Normalverteilung $N(\mu,\sigma_1)$ sowie $N(\mu,\sigma_2)$

1.7.2 Transformation zufälliger Variablen

$$\Omega \overset{\xi}{\Rightarrow} \mathbb{R}^1 \overset{\eta=g(\xi)}{\Rightarrow} \mathbb{R}^1$$

$$\begin{array}{rcl}
 & & \text{g — reellwertige Funktion} \\
F_\xi(x) & = & P(\xi < x) \\
F_\eta(y) & = & P(\eta < y) \\
 & & (\Omega,\mathcal{E},P) \overset{\xi}{\Rightarrow} (\mathbb{R}^1,\mathbb{B}^1,P_\xi) \\
 & & \text{g — reellwertige Funktion} \\
F_\xi(x) & = & P(\xi < x)
\end{array}$$

$$F_\eta(y) = P(\eta < y) = P(g(\xi) < y) = P(\xi \in \underbrace{\{x : g(x) < y\}}_{\in \mathbb{B}^1})$$

1.5 Satz

ξ verteilt mit der stetigen Dichtefunktion $f(x)$.

$g(x)$ sei differenzierbar und $g'(x) \neq 0 \quad \forall x \in W_\xi$ (W_ξ ist der Wertevorrat von ξ).

Dann hat die zufällige Variable $\eta = g(\xi)$ die Dichtefunktion

$$h(y) = f\left(g^{-1}(y)\right) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

Beweis:

Fall 1: $g'(x) > 0$

$$\begin{aligned}
\eta &= \eta(\omega) = g(\xi(\omega)) \\
A &= \{\omega : \xi(\omega) < x\} = \{\omega : \eta(\omega) < y\} \\
P(A) &= F_\xi(x) = F_\eta(y) \\
F'_\xi(x) &= f(x) = \frac{dF_\xi(x)}{dx} = \frac{dF_\eta(y)}{dy} \cdot \frac{dy}{dx} \\
&= h(y)\frac{dy}{dx} \\
f(x) &= f\left(g^{-1}(y)\right) = h(y)\frac{dy}{dx} \\
h(y) &= f\left(g^{-1}(y)\right)\frac{dx}{dy} \\
&= f\left(g^{-1}(y)\right)\frac{dg^{-1}(y)}{dy} \\
&\quad \text{da } g \text{ streng monoton} \\
&= \underline{\underline{f\left(g^{-1}(y)\right) \cdot \left|\frac{dg^{-1}(y)}{dx}\right|}}
\end{aligned}$$

Fall 2: $g'(x) < 0$

$$\begin{aligned}
A &= \{\omega : \xi(\omega) < x\} = \{\omega : y < \eta(\omega) < \infty\} \\
F_\xi(x) &= P\left(\overline{\{\omega : \eta(\omega) \leq y\}}\right) \\
&= P\left(\overline{\{\omega : \eta(\omega) < y\}}\right) \\
&= 1 - P\left(\{\omega : \eta(\omega) < y\}\right) \\
&= 1 - F_\eta(y) \\
\frac{dF_\xi(x)}{dx} &= f(x) = \frac{d\left(1 - F_\eta(y)\right)}{dx} \\
&= -h(y) \cdot \underbrace{\frac{dg^{-1}(y)}{dy}}_{<0} \\
&= \underline{\underline{h(y) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|}}
\end{aligned}$$

1.22 Beispiel
ξ besitze Gleichverteilung über $[a,b)$, d.h.

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ , falls } x \in [a,b) \\ 0 \text{ , sonst} \end{cases}$$

siehe Abbildung 1.10

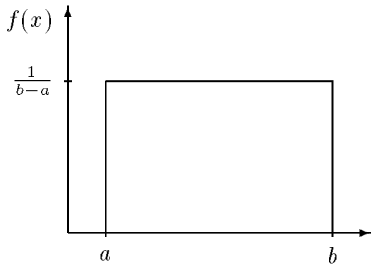

Abbildung 1.10: Die Dichtefunktion für Gleichverteilung in $[a,b)$

1.23 Beispiel
ξ sei gleichverteilt über $[0, \frac{\pi}{2}]$

$$f(x) = \begin{cases} \frac{2}{\pi} \text{ , falls } 0 \leq x < \frac{\pi}{2} \\ 0 \text{ , sonst} \end{cases}$$

$$\begin{aligned} y &= g(x) = \sin x \\ g(x) &> 0 \quad \forall x \in W_\xi = [0, \frac{\pi}{2}) \\ g^{-1}(y) &= \arcsin y \\ \eta &= \sin \xi \\ h(y) &= f(\arcsin y) \frac{d \arcsin y}{dy} \\ h(y) &= \begin{cases} \frac{2}{\pi} \frac{1}{\sqrt{1-y^2}} & \text{, falls } y \in [0,1) \\ 0 & , sonst \end{cases} \end{aligned}$$

1.24 Beispiel
ξ ist zufällige Variable

$$\left.\begin{aligned} F(x) &- \text{Verteilungsfunktion, differenzierbar} \\ F'(x) &= f(x) \text{ — Dichtefunktion} \\ & f(x) > 0 \end{aligned}\right\} \text{auf einem gewissen Intervall}$$

$$\begin{aligned} P(\xi < x) &= F(x) \\ \eta &= F(\xi) \\ h(y) &= f\left(F^{-1}(y)\right) \frac{dF^{-1}(y)}{dy} \equiv \begin{cases} 1 \text{ , falls } y \in [0,1] \\ 0 \text{ , sonst} \end{cases} \end{aligned}$$

$$\begin{aligned}
F\left(F^{-1}(y)\right) &= y \\
\frac{dF(F^{-1}(y)}{dy} &= 1 \\
F'(F^{-1}(y))F'(y) &= 1 \\
\Rightarrow & \quad f(F^{-1}(y))F'(y) = 1 \\
F'(y) &= \frac{1}{f(F^{-1}(y))} \\
\eta &= F(\xi) \text{ ist gleichverteilt über } [0,1)
\end{aligned}$$

$$\begin{aligned}
U & \quad \text{gleichverteilt über } [0,1) \\
G & - \text{beliebige Verteilungsfunktion} \\
G' & = g > 0
\end{aligned}$$

$$\Rightarrow \boxed{G^{-1}(U) \text{ besitzt die Verteilungsfunktion } G}$$

1.25 Beispiel
$\eta = a\xi + b$ ist gegeben,

$$\begin{aligned}
y &= g(x) = ax + b; a \neq 0 \\
& \quad f(x) \text{ Dichte von } \xi
\end{aligned}$$

Die Dichte von η berechnet sich dann wie folgt:

$$\begin{aligned}
h(y) &= f\left(g^{-1}(y)\right) \cdot \left|\frac{dg^{-1}(y)}{dy}\right| \\
&= f\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|} \\
x &= g^{-1}(y) = \frac{y-b}{a}
\end{aligned}$$

1.7.2.1 Anwendung

Wir haben eine zufällige Variable ξ mit:

$$\xi \sim N(\mu, \sigma), \sigma > 0$$

Daraus folgt, die Dichtefunktion f ist wie folgt definiert:

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

Wenn wir nun eine zweite zufällige Variable η haben, für die gilt:

$$\eta = \frac{\xi - \mu}{\sigma}; y = g(x) = \frac{x-\mu}{\sigma}$$

ergibt sich die Dichtefunktion h mit der Formel:

$$h(y) = f\left(g^{-1}(y)\right) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

wobei

$$\begin{aligned} x &= g^{-1}(y) = \sigma y + \mu \\ \frac{dx}{dy} &= \frac{dg^{-1}(y)}{dy} = \sigma \\ h(y) &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{\sigma y + \mu - \mu}{\sigma}\right)^2} \cdot \underbrace{|\sigma|}_{=\sigma,\text{ da } \sigma>0} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \\ &\Rightarrow \eta \sim N(0,1) \end{aligned}$$

Arbeit mit einer Tabelle von Wahrscheinlichkeiten Es ist $\xi \sim N(\mu, \sigma)$ gegeben. Wir suchen die Wahrscheinlichkeit

$$P(\xi < x) = P\left(\frac{\xi - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = P\left(\eta < \frac{x - \mu}{\sigma}\right)$$

Dabei ist $\eta \sim N(0,1)$, d.h. η ist eine standardisierte, normalverteilte Zufallsgröße

- $\xi \sim N(0,1)$

$$\begin{aligned} P(\xi < 1) &= 0,841345 \\ P(\xi < -1) &= \Phi(-1) = \underbrace{1 - \Phi(1)}_{\text{da Gesamtfläche} = 1} \approx 0,15 \\ P(-1 \leq x \leq 1) &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 \approx 0,68 \end{aligned}$$

- $\xi \sim N(1,1)$

$$P(\xi < 1) = P\left(\eta < \frac{1-1}{1}\right) = P(\eta < 0) = 0,5$$

Die Umkehrung — Quantile Gesucht ist zu einem Wahrscheinlichkeitswert $P(\xi < x_p)$ der entsprechende Wert von x_p.

z.B. für $\xi \sim N(0,1)$ und $p = 0,95$ ist der Wert von $x_p \approx 1,645$.

$$\begin{aligned} & \xi \sim N(\mu, \sigma) \\ x_p(\mu, \sigma) &= \sigma x_p(0,1) + \mu \\ P\left(\xi < x_p(\mu, \sigma)\right) &= p \\ &= P\left(\frac{\xi - \mu}{\sigma} < \underbrace{\frac{x_p(\mu, \sigma) - \mu}{\sigma}}_{=x_p(0,1)}\right) = p \end{aligned}$$

1.12 Definition ($k \cdot \sigma$-Intervall)
Für $\xi \sim N(\mu, \sigma)$ ist $k \cdot \sigma$-Intervall die Wahrscheinlichkeit:

$$P(\mu - k\sigma \leq \xi \leq \mu + k\sigma)$$

Für diese $k \cdot \sigma$-Intervalle gilt:

$$\begin{aligned}
&P(\mu - k\sigma \leq \xi \leq \mu + k\sigma) \\
&= P\left(\frac{\mu - k\sigma - \mu}{\sigma} \leq \frac{\xi - \mu}{\sigma} \leq \frac{\mu + k\sigma - \mu}{\sigma}\right) \\
&= P\Bigg(- k \leq \underbrace{\frac{\xi - \mu}{\sigma}}_{\sim N(0,1)} 1 \leq k\Bigg) \\
&= \Phi(k) - \underbrace{\Phi(-k)}_{1-\Phi(k)} \\
&= \underline{\underline{2\Phi(k) - 1}}
\end{aligned}$$

Zum Beispiel gilt für das 3σ-Intervall:

$$2\Phi(3) - 1 = 2 \cdot 0,999 - 1 = 0,998$$

$$\begin{aligned}
&\xi : (\Omega, \mathcal{E}, P) \Rightarrow (\mathbb{R}^p, \mathbb{B}^p) \\
&\quad \mathbb{B}^p \quad \text{— } \sigma\text{-Algebra der Borelmengen im } \mathbb{R}^p \\
&\quad \mathbb{B}^p \quad \text{— kleinste } \sigma\text{-Algebra über dem } \mathbb{R}^n \\
&\qquad\qquad \text{die alle offenen Intervalle enthält}
\end{aligned}$$

1.7.3 Zufällige Vektoren

1.13 Definition (Zufälliger Vektor)
Wenn $\xi_i, i \in [1, p]$ reellwertige, zufällige Variablen sind, dann ist ξ, definiert durch $\xi^T = (\xi_1, \ldots, \xi_p)$ ein *zufälliger Vektor*. Die ξ_i sind jeweils über dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$ definiert.

1.14 Definition (Verteilungsfunktion)

$$F_\xi(x_1, \ldots, x_p) := P(\xi_1 < x_1, \ldots, \xi_p < x_p)$$

heißt *Verteilungsfunktion* des zufälligen Vektors ξ. Sie wird auch mit $F_{\xi_1,\ldots,\xi_p}(x_1, \ldots, x_p)$ bezeichnet.

Es gilt:

$$\begin{aligned}
&F_{\xi_1,\ldots,\xi_p}(x_1, \ldots, x_p) \\
&= P\left(\omega : \xi_1(\omega) < x_1 \wedge \ldots \wedge \xi_p(\omega) < x_p\right) \\
&= \underline{P\left(\bigcap_{i=1}^{p} \{\omega \in \Omega : \xi_i(\omega) < x_i\}\right)}
\end{aligned}$$

1.7.3.1 Eigenschaften der Verteilungsfunktion

1. $F_\xi(x_1, \ldots, x_p) = F_{\xi_{i_1}, \ldots, \xi_{i_p}}(x_{i_1}, \ldots, x_{i_p})$
2. $\lim\limits_{x_p \to -\infty} F_\xi(x_1, \ldots, x_p) = 0$
3. $\lim\limits_{x_p \to \infty} F_\xi(x_1, \ldots, x_p) = F_{\xi_1, \ldots, \xi_{p-1}}(x_1, \ldots, x_{p-1})$
4. $\lim\limits_{\substack{x_1 \to \infty \\ \vdots \\ x_p \to \infty}} F_\xi(x_1, \ldots, x_p) = 1$
5. $F_\xi(x_1, \ldots, x_p)$ ist in jedem Argument nichtfallend
6. $F_\xi(x_1, \ldots, x_p)$ ist in jedem Argument linksseitig stetig

1.15 Definition
Der zufällige Vektor $\xi^T = (\xi_1, \ldots, \xi_p)$ heißt *stetig verteilt*, wenn

$$F_\xi(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_p} f(t_1, \ldots, t_p)\, dt_1 \ldots dt_p$$

wobei

1. $f(x_1, \ldots, x_p) \geq 0 \qquad \forall x_1, \ldots, x_p$
2. $\int_{\mathbb{R}^p} f(x_1, \ldots, x_p)\, dx_1 \ldots dx_p = 1$

$f(x_1, \ldots, x_p)$ heißt dann *Dichtefunktion* von ξ.

Falls $f(x_1, \ldots, x_p)$ stetig ist, so gilt:

$$f(x_1, \ldots, x_p) = \frac{\partial^p F_\xi(x_1, \ldots, x_p)}{\partial x_1 \ldots \partial x_p}$$

1.16 Definition (diskret)
Ein zufälliger Vektor ξ mit $\xi^T = (\xi_1, \ldots, \xi_p)$ heißt *diskret*, falls der Wertevorrat für jedes ξ_i höchstens abzählbar ist.

1.17 Definition (Verteilungsfunktion)
ξ_i habe den Wertevorrat $\{x_{i_1}, \ldots, x_{i_k}, \ldots\}$. Dann ist die Wahrscheinlichkeit p als

$$p_{x_{1_j} \cdots x_{p_k}} = P(\xi_1 = x_{1_j}, \ldots, \xi_p = x_{p_k})$$

und die *Verteilungsfunktion* als

$$F_\xi(x_1, \ldots, x_p) = \sum_{\substack{j: x_{1_j} < x_1 \\ \vdots \\ k: x_{p_k} < x_p}} p_{x_{1_j} \cdots x_{p_k}}$$

definiert.

1.26 Beispiel (Würfel)
Wir werfen zwei Würfel. $\xi = (\xi_1, \xi_2)$, wobei ξ_1 die Augenzahl des ersten Würfels ist und ξ_2 die des zweiten.

Wir wollen nun wissen, wie groß die Wahrscheinlichkeit ist, daß der 1. Würfel weniger als 4 Augen zeigt und der 2. Würfel weniger als 3. Wir wissen bereits, daß die beiden Würfel voneinander unabhängig sind.

$$P(\xi_1 < 4, \xi_2 < 3) = \sum_{i<4;j<3} p_{ij} = \frac{6}{36} = \frac{1}{6}$$

Dabei gilt (da die Würfel voneinander unabhängig sind)

$$P(\xi_1 = i, \xi_2 = j) = p_{ij} = P\left(\{\xi_1 = i\} \cap \{\xi_2 = j\}\right) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

und

$$(p_{ij}) = \begin{pmatrix} \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \\ \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \\ \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \\ \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \\ \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \\ \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} \end{pmatrix}$$

1.6 Satz (Randdichte)
Wir haben mit $f(x_1, x_2)$ eine beliebige Dichte für einen zufälligen Vektor $\xi = \binom{\xi_1}{\xi_2}$. Die Dichte von ξ_1 (*Randdichte*) berechnet sich dann mit:

$$f_{\xi_1} = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2$$

Beweis: Wenn

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2)\, dt_1\, dt_2$$

die Verteilungsfunktion von f ist und wir den Grenzwert für $x_2 \to \infty$ berechnen, dann erhalten wir:

$$\lim_{x_2 \to \infty} F(x_1, x_2) = P(\xi_1 < x_1) = \int_{-\infty}^{x_1} f_1(t_1)\, dt_1$$

Wir betrachten den zufälligen Vektor für den Fall: $p = 2$

1. $\xi = \binom{\xi_1}{\xi_2}$ diskret (endlich)

$$\begin{aligned} P\left(\xi_1 \in \{x_1, \ldots, x_k\}\right) &= 1 \\ P\left(\xi_2 \in \{y_1, \ldots, y_l\}\right) &= 1 \\ X &= \{x_1, \ldots, x_k\} \\ Y &= \{y_1, \ldots, y_l\} \\ P(\xi \in X \times Y) &= 1 \\ X \times Y &= \{(x_i, y_j) : x_i \in X \text{ und } y_j \in Y\} \\ P\left(\xi = (x_i, y_j)\right) &= P(\xi_1 = x_i \wedge \xi_2 = y_j) \\ &= P(\xi_1 = x_i, \xi_2 = y_j) = p_{ij} \end{aligned}$$

$P = (p_{ij})_{\substack{i=1,\dots,k \\ j=1,\dots,l}}$

$\xi_1 \backslash \xi_2$	y_1	y_2	$\dots$	y_l	
x_1	p_{11}	p_{12}	$\dots$	p_{1l}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\dots$	p_{2l}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_k	p_{k1}	p_{k2}	$\dots$	p_{kl}	$p_{k\cdot}$
$\sum$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\dots$	$p_{\cdot l}$	1

$$\begin{aligned}
p_{\cdot j} &= \sum_{i=1}^{k} p_{ij} \\
p_{i\cdot} &= \sum_{j=1}^{l} p_{ij}
\end{aligned}$$

Diese Werte sind die Randwahrscheinlichkeiten bezüglich ξ_1 bzw. ξ_2. Für ξ_1 und ξ_2 gelten folgende Wertetabellen:

$$\xi_1 : \begin{pmatrix} x_1 & x_2 & \dots & x_k \\ p_{1\cdot} & p_{2\cdot} & \dots & p_{k\cdot} \end{pmatrix}$$

$$\xi_2 : \begin{pmatrix} y_1 & y_2 & \dots & y_l \\ p_{\cdot 1} & p_{\cdot 2} & \dots & p_{\cdot l} \end{pmatrix}$$

$$\begin{aligned}
P_{\cdot j} &= \sum_{i=1}^{k} p_{ij} \\
&= \sum_{i=1}^{k} P(\xi_1 = x_i, \xi_2 = y_j) \\
&= \sum_{i=1}^{k} P(\xi_1 = x_i \wedge \xi_2 = y_j) \\
&= \sum_{i=1}^{k} P\left(\{\omega : \xi_1(\omega) = x_i\} \cap \{\omega : \xi_2(\omega) = y_j\}\right) \\
&= \underline{P\left(\{\omega : \xi_2(\omega) = y_j\}\right)}
\end{aligned}$$

Diese Formel folgt aus der *Formel der totalen Wahrscheinlichkeit* auf der Seite 13, denn es gilt:

$$\begin{aligned}
\{\omega : \xi_1(\omega) = x_i\} \cap \{\omega : \xi_1(\omega) = x_j\} &= \emptyset \quad \forall i \neq j \\
\bigcup_{i=1}^{k} \{\omega : \xi_1(\omega) = x_i\} &= \Omega
\end{aligned}$$

d.h. $\{\omega : \xi_1(\omega) = x_i\}_{i=1,\dots,n}$ ist vollständig

$$\Omega \stackrel{\xi_1}{\Rightarrow} \{x_1, \dots, x_k\}$$

$$\sum_{i,j} p_{ij} = 1$$

1.27 Beispiel
Zur Verdeutlichung ein konkretes Beispiel mit $k = l = 2$.

$$\begin{aligned}
\xi_1 &= \begin{cases} 1, \text{ falls weiblich} \\ 2, \text{ falls männlich} \end{cases} \\
\xi_2 &= \begin{cases} 1, \text{ falls Raucher} \\ 2, \text{ falls Nichtraucher} \end{cases}
\end{aligned}$$

${}_{\xi_1}\backslash^{\xi_2}$	1	2
1	p_{11}	p_{12}
2	p_{21}	p_{22}

Erstellen nun eine 2×2-Kontigenztafel. Dabei wird für jedes der 4 Felder anhand einer statistischen Erhebung die Anzahl der Personen ermittelt, für die die beiden Eigenschaften zutreffen.

${}_{\xi_1}\backslash^{\xi_2}$	1	2	
1	n_{11}	n_{12}	$n_{1\cdot}$
2	n_{21}	n_{22}	$n_{2\cdot}$
	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

Dabei bedeuten:

$\boldsymbol{n_{ij}}$ Anzahl der Personen mit dem Geschlecht $\boldsymbol{i}$ und dem Raucherverhalten $\boldsymbol{j}$

$\boldsymbol{n_{\cdot 1}}$ Anzahl der Raucher

$\boldsymbol{n_{\cdot 2}}$ Anzahl der Nichtraucher

$\boldsymbol{n_{1\cdot}}$ Anzahl der Frauen

$\boldsymbol{n_{2\cdot}}$ Anzahl der Männer

$\boldsymbol{n_{\cdot\cdot}}$ Gesamtzahl der Befragten

Mit $p'_{ij} = \frac{n_{ij}}{n_{\cdot\cdot}}$ ergibt sich eine Schätzung für p_{ij}.

2. $\xi = \binom{\xi_1}{\xi_2}$ als stetige zufällige Variable

   Wir erhalten eine 2-dimensionale Dichtefunktion $f(x, y)$, für die folgendes gilt:

   (a) $$\int\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

   (b) $$f(x,y) \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$$

   Die Verteilungsfunktion ergibt sich dann durch:

   $$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv = P(\xi_1 < x, \xi_2 < y)$$

   da f(x,y) stetig ist, gilt weiterhin:

   $$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\,\partial y}$$

Randverteilung

$$\lim_{y\to\infty} F(x,y) = \underbrace{F_\xi(x) = P(\xi < x)}_{\text{Randverteilung von } \xi}$$

$$\begin{aligned}
f_{\xi_1}(x) &= \frac{dF_{\xi_1}(x)}{dx} = \int_{-\infty}^{\infty} f(x,y)\,dy \\
f_{\xi_2}(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx \\
F_{\xi_1}(x) &= \int_{-\infty}^{x} f_{\xi_1}(t)\,dt \\
F_{\xi_2}(y) &= \int_{-\infty}^{y} f_{\xi_2}(t)\,dt
\end{aligned}$$

Dabei sind:

$\boldsymbol{f_{\xi_i}}$ Randdichte von $\boldsymbol{\xi_i}$

$\boldsymbol{F_{\xi_i}}$ Randverteilung von $\boldsymbol{\xi_i}$

1.7.4 Transformationsformel für zufällige Vektoren

Wir haben mit $\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_p \end{pmatrix}$ einen zufälligen Vektor mit einer Dichtefunktion: $f(x_1, \ldots, x_p)$.
Weiterhin eine Menge $B \in \mathbb{B}^1$ und eine Wahrscheinlichkeitsfunktion

$$P(\xi \in B) = \int_B f(x_1, \ldots, x_p)\,dx_1 \ldots dx_p$$

Eine Funktion $g : \mathbb{R}^p \to \mathbb{R}^p$ sei umkehrbar eindeutig und es soll für ein X gelten:

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \overset{g}{\longleftrightarrow} Y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$$

Möglich wäre z.B. eine Funktion $g(x) = A \cdot x$, wobei A eine reguläre (p,p)-Matrix ist.
Die Umkehrabbildung $g^{-1} : \mathbb{R}^p \to \mathbb{R}^p$ ist mittels Funktionen ψ_i definiert (diese existieren aufgrund der umkehrbar eindeutigen Funktion g), die $y_1, \ldots, y_p$ eine Zahl x_i zuordnet, d.h. $x_i = \psi_i(y_1, \ldots, y_p)$. Die Funktion g^{-1} ist dann folgendermaßen definiert:

$$g^{-1}(y) = \begin{pmatrix} \psi_1(y_1, \ldots, y_p) \\ \vdots \\ \psi_p(y_1, \ldots, y_p) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

Wir betrachten nun $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_p \end{pmatrix} = g(\xi)$ und suchen die Dichtefunktion h_η. Dabei gehen wir (ohne Beweis) formal wie bei den zufälligen Variablen vor.

$$
\begin{aligned}
h_\eta(y_1,\dots,y_n) &= f(\psi_1(y_1,\dots,y_p),\dots,\psi_p(y_1,\dots,y_p))\cdot|J| \\
&\ \text{wobei} \\
J &= \left|\frac{\partial\psi_i}{\partial y_j}\right|_{i,j:1,\dots,p} \\
&\ \text{obiger Ausdruck soll die Determinante dieser Matrix von partiellen} \\
&\ \text{Ableitungen sein}
\end{aligned}
$$

1.28 Beispiel
Summe unabhängiger Zufallsgrößen

$\xi = \binom{\xi_1}{\xi_2}$ ist ein zufälliger Vektor ($p = 2$), wobei ξ_1, ξ_2 voneinander unabhängig sind. Bei gegebener Dichte von ξ suchen wir die Dichte von $\eta = \binom{\xi_1+\xi_2}{\xi_2}$.

$$
\begin{aligned}
f(x_1,x_2) &= f_1(x_1)\cdot f_2(x_2) \\
&\ \text{gilt, da } \xi_1,\xi_2 \text{ unabhängig}
\end{aligned}
$$

$$
\begin{aligned}
g_1(\xi_1,\xi_2) &= \eta_1 = \xi_1+\xi_2 \\
g_2(\xi_1,\xi_2) &= \eta_2 = \xi_2 \\
\Rightarrow & \\
x_1 &= \psi_1(y_1,y_2) = y_1 - y_2 \\
x_2 &= \psi_2(y_1,y_2) = y_2 \\
J &= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} \\
&= 1 \\
&\ \text{Dichte von } \eta = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \\
h_\eta(y_1,y_2) &= f(\psi_1(y_1,y_2),\psi_2(y_1,y_2))\cdot|1| \\
&= f_1(\psi_1(y_1,y_2))\cdot f_2(\psi_2(y_1,y_2)) \\
&= \underline{f_1(y_1-y_2)\cdot f_2(y_2)}
\end{aligned}
$$

Wir haben nun mit $h_\eta(y_1, y_2)$ die gemeinsame Dichte von $\eta_1 = \xi_1 + \xi_2$ und $\eta_2 = \xi_2$ berechnet. Wie bekommt man nun die Randdichte für $\eta_1 = \xi_1 + \xi_2$ (Randdichte bzgl. der ersten Komponente)?

$$h_\eta(y_1) = \int_{-\infty}^{\infty} h(y_1,y_2)\,dy_2$$

$$\boxed{h_{\xi_1+\xi_2}(y_1) = \int_{-\infty}^{\infty} f_1(y_1-y_2)\cdot f_2(y_2)\,dy_2}$$

Diese Formel wird auch *Faltung* aus f_1 und f_2 genannt, geschrieben wird dies $f_1 * f_2$. Hinweis: Faltung der Dichten ergibt Dichte der Summen.

1.29 Beispiel (Faltung)
Es seien ξ_1 und ξ_2 über $[0,1)$ gleichverteilt und unabhängig. Die ξ_i besitzen jeweils die Dichte

$$f_i(x) = \begin{cases} 1 \text{ , falls } 0 \le x < 1 \\ 0 \text{ , sonst} \end{cases}$$

$$\begin{aligned}
& & & 0 \le \xi_i \le 1 \\
& & & 0 \le \xi_1 + \xi_2 \le 2 \\
\eta_1 &= & & \xi_1 + \xi_2 \\
h_{\eta_1}(y) &= & & \int_{-\infty}^{\infty} f_1(y-x) f_2(x)\,dx \\
&= & & \int_0^1 f_1(y-x)\,dx \quad \text{da } f_2(x) = 1 \quad \forall x : 0 \le x \le 1 \\
f_1(y-x) &= & & \begin{cases} 1 \text{ , falls } 0 \le y - x \le 1 \\ 0 \text{ , sonst} \end{cases} \\
& & (a) & \; 0 \le y \le 1 \\
& & & f_1(y-x) = 1, \text{ falls } 0 \le x \le y \\
& & (b) & \; 1 < y \le 2 \\
& & & f_1(y-x) = 1, \text{ falls } y - 1 \le x \le 1 \\
& & & \text{siehe Abbildung 1.11}
\end{aligned}$$

$$\begin{aligned}
h_{\eta_1}(y) &= \int_0^1 f_1(y-x)\,dx = \begin{cases} \int_0^y \,dx & \text{, falls } 0 \le y \le 1 \\ \int_{y-1}^1 \,dx & \text{, falls } 1 < y \le 2 \\ 0 & \text{, sonst} \end{cases} \\
&= \begin{cases} y & \text{, falls } 0 \le y \le 1 \\ 2 - y & \text{, falls } 1 < y \le 2 \\ 0 & \text{, falls } y \notin [0,2] \end{cases}
\end{aligned}$$

siehe Abbildung 1.12

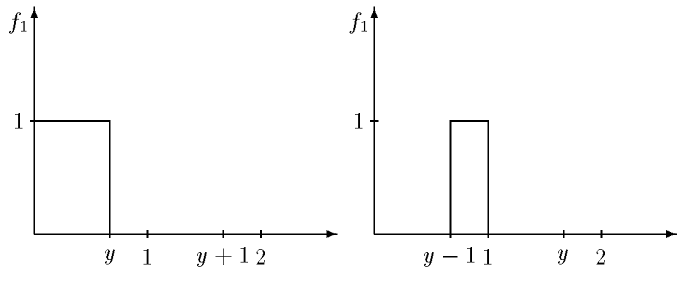

Abbildung 1.11: Die Dichte f_1 für $0 \le y \le 1$ und $1 < y \le 2$

Wir betrachten nun die Addition von 3 solcher Werte, d.h. $\eta_3 = (\xi_1 + \xi_2) + \xi_3$. Die Dichtefunktion für ξ_3 entspricht der von ξ_i.

$$h_{\eta_3}(z) = \int_{-\infty}^{+\infty} h_{\xi_1+\xi_2}(z-x) \cdot f_3(x)\,dx$$

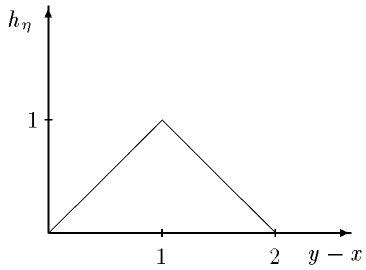

Abbildung 1.12: Die Dichtefunktion von $\xi_1 + \xi_2$

$$= h_{\xi_1+\xi_2} * f_3(x)$$

$$h_{\xi_1+\xi_2}(z-x), \quad 0 \le z \le 3$$

(a) $0 \le z \le 1$

(b) $1 < z \le 2$
siehe Abbildung 1.13

(c) $2 < z \le 3$
siehe Abbildung 1.14

$$h_{\xi_1+\xi_2+\xi_3}(z) = \begin{cases} 0 & \text{, falls } z \notin [0,3] \\ \frac{z^2}{2} & \text{, falls } 0 \le z \le 1 \\ 1 - \frac{(z-1)^2}{2} - \frac{(2-z)^2}{2} & \text{, falls } 1 < z \le 2 \\ \frac{(3-z)^2}{2} & \text{, falls } 2 < z \le 3 \end{cases}$$

siehe Abbildung 1.15

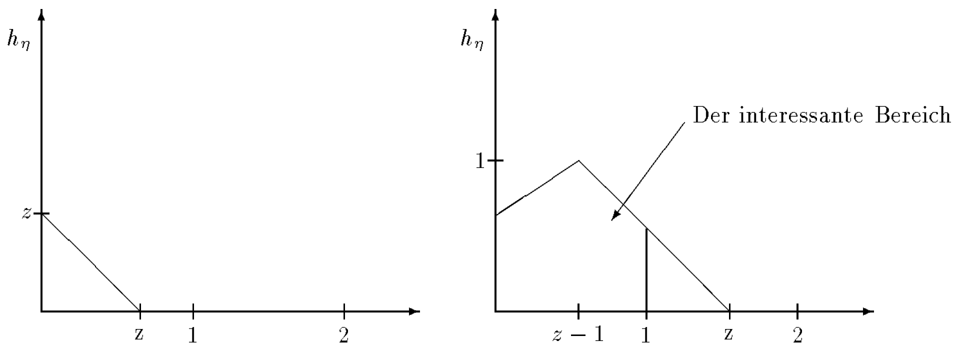

Abbildung 1.13: Die Dichte h_η für $0 \le z \le 1$ und $1 < z \le 2$

1.8 Unabhängigkeit von Zufallsgrößen

1.18 Definition (unabhängig)
Wir haben mit ξ, η zwei zufällige Variablen auf $(\Omega, \mathcal{E}, P)$. ξ und η heißen *stochastik unabhängig* genau dann, wenn für alle $A, B \in \mathbb{B}^1$ gilt:

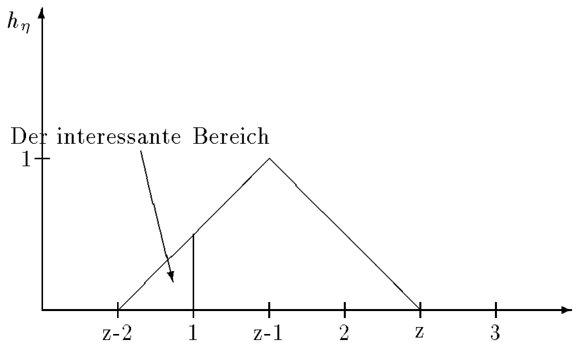

Abbildung 1.14: Die Dichte h_η für f_1 für $2 < z \leq 3$

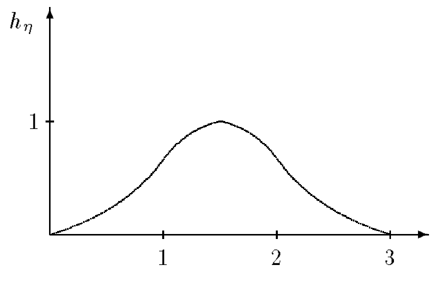

Abbildung 1.15: Die Gesamtdichte für die Summe $\xi_1 + \xi_2 + \xi_3$

- $P(\xi \in A \wedge \eta \in B) = P(\xi \in A, \eta \in B) = P(\xi \in A) \cdot P(\eta \in B)$
- $P(A \cap B) = P(A) \cdot P(B)$
  äquivalent dazu ist
  $$P(\xi \in A \wedge \eta \in B) = P\left(\{\omega : \xi(\omega) \in A\} \cap \{\omega : \eta(\omega) \in B\}\right)$$

Haben mit $\binom{\xi}{\eta}$ einen zufälligen Vektor, dabei sind ξ und η unabhängig.

$$\begin{aligned}
F(x,y) &= P(\xi < x, \eta < y) \\
&= P(\xi \in \underbrace{(-\infty, x)}_{A \in \mathbb{B}^1} \wedge \eta \in \underbrace{(-\infty, y)}_{B \in \mathbb{B}^1}) \\
&= P(\xi \in (-\infty, x)) \cdot P(\eta \in (-\infty, y)) \\
&= F_\xi(x) \cdot F_\eta(y) \\
\Longleftrightarrow & \quad f(x,y) = f_\xi(x) \cdot f_\eta(y)
\end{aligned}$$

1.7 Satz
Die zufälligen Variablen ξ und η sind unabhängig genau dann, wenn für alle $\mathbb{B}^1$-meßbaren Funktionen φ und ψ gilt: $\varphi(\xi)$ und $\psi(\eta)$ sind unabhängig.

Beweis: φ, ψ sind auf $\mathbb{R}^1$ definiert und reellwertig.

$$\begin{aligned}
\varphi^{-1}(A) &= \{x : \varphi(x) \in A\} \in \mathbb{B}^1 \quad \forall A \in \mathbb{B}^1 \\
\varphi^{-1}(B) &= \{y : \psi(y) \in B\} \in \mathbb{B}^1 \quad \forall B \in \mathbb{B}^1 \\
\Longleftrightarrow & \quad \varphi \text{ und } \psi \ \mathbb{B}^1\text{-meßbar}
\end{aligned}$$

$(\rightarrow)$ $\quad \xi, \eta$ unabhängig

$$\begin{aligned}
&P(\varphi(\xi) \in A, \psi(\eta) \in B) \\
&= P(\xi \in \varphi^{-1}(A), \eta \in \psi^{-1}(B)) \\
&\Omega \begin{array}{l} \rightarrow^\xi \mathbb{R}^1 \rightarrow^\varphi \mathbb{R}^1 \\ \rightarrow^\eta \mathbb{R}^1 \rightarrow^\psi \mathbb{R}^1 \end{array} \\
&= P(\xi \in \varphi^{-1}(A)) \cdot P(\eta \in \psi^{-1}(B)) \\
&= P(\varphi(\xi) \in A) \cdot P(\psi(\eta) \in B)
\end{aligned}$$

$(\leftarrow)$ $\quad \forall \varphi, \psi$ ($\mathbb{B}^1$-meßbar)

$\varphi(\xi), \psi(\eta)$ sind unabhängig

$\Rightarrow \quad \varphi(x) \equiv \psi(x) \equiv x$

d.h. $\xi = f(\xi)$ und $\eta = \psi(\eta)$ sind unabhängig

1.9 Charakteristika von Verteilungsfunktionen

1.9.1 Erwartungswert

1.30 Beispiel (Mittelwert)
Wir betrachten einmal den Würfel.

$$\xi : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Mittelwert: $\dfrac{1+2+3+4+5+6}{6} = 1 \cdot \dfrac{1}{6} + 2 \cdot \dfrac{1}{6} + \cdots = 3,5$

Allgemeiner:

Mittelwert: $x_1 \cdot p_1 + \cdots + x_n \cdot p_n$

$$p_i = P(\xi = x_i); \quad \sum_{i=1}^{k} p_i = 1; \quad p_i \geq 0$$

1.19 Definition (Erwartungswert, diskret)
Falls ξ eine diskrete zufällige Variable ist und $\sum_i |x_i| \cdot p_i < \infty$, so heißt $E\xi = \sum_i x_i \cdot p_i$ der *Erwartungswert* von ξ.

1.31 Beispiel (Binomialverteilt)
Wir betrachten $B(p,1)$:

$$\begin{aligned} \xi &\sim B(p,1) \qquad \xi : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix} \\ E\xi &= 0 \cdot (1-p) + 1 \cdot p = p \end{aligned}$$

1.32 Beispiel (Binomialverteilt)
Wir betrachten $B(p,n)$:

$$\begin{aligned} \xi &\sim B(p,n) \qquad \xi = \sum_{i=0}^{n} \xi_i; \quad \xi_i \sim B(p,1) \text{ und unabhängig} \\ & \qquad \xi : \begin{pmatrix} 0 & 1 & \dots & n \\ p_0 & p_1 & \dots & p_n \end{pmatrix} \\ p_i &= \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

Ist $E\xi = n \cdot p$?

$$\begin{aligned} E\xi &= \sum_{i=0}^{n} i \cdot p_i = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\ &= n \cdot p \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ & \quad \text{da } i \binom{n}{i} = n \binom{n-1}{i-1} \\ &= n \cdot p \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-(i+1)} \\ &= n \cdot p \underbrace{\sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{(n-1)-i}}_{(p+1-p)^{n-1}=1} \\ &= n \cdot p \end{aligned}$$

1.33 Beispiel (Poissonverteilt)
Wir betrachten eine POISSON-verteilte zufällige Variable ξ, d.h.

$$p_i = P(\xi = i) = \frac{\lambda^i}{i!} e^{-\lambda} \geq 0; \quad \sum_i p_i = 1$$

$$\begin{aligned} E\xi &= \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \lambda \cdot e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda\, e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!}}_{e^{\lambda}} \\ &= \lambda \end{aligned}$$

Wir realisieren eine Folge $\xi_k \sim B(p, 1)$ und definieren

$$\eta = \inf_k \{k : \xi_k = 1\}$$

Wie ist hier $E\eta$?

$$\begin{aligned} \eta &= \begin{pmatrix} 1 & 2 & 3 & \dots & k & \dots \\ p & p \cdot q & q^2 p & \dots & q^{k-1} p & \dots \end{pmatrix} \\ P(\eta = 2) &= P(\xi_1 = 0 \wedge \xi_2 = 1) \\ &= P(\xi_1 = 0) \cdot P(\xi_2 = 1) \\ &= \underbrace{(1-p)}_{q} \cdot p \\ &= p \cdot q \\ P(\eta = k) &= p \cdot q^{k-1} \\ & \quad \text{Geometrische Verteilung} \end{aligned}$$

$$\sum_{k=1}^{\infty} p \cdot q^{k-1} = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1$$

$$E\eta = \sum_{k=1}^{\infty} k \cdot p q^{k-1} = p \sum_{k=1}^{\infty} k \cdot q^{k-1} = \frac{p}{(1-q)^2}$$

Beweis:

$$\begin{aligned} s_n &= \sum_{k=1}^{n} k q^{k-1} = \frac{1-(n+1)q^n + n q^{n+1}}{(1-q)^2} \\ & \quad \text{vollständige Induktion} \\ \lim_{n\to\infty} s_n &= \frac{1}{(1-q)^2} \end{aligned}$$

1.20 Definition (Erwartungswert, stetig)
Falls $f(x)$ die Dichtefunktion der zufälligen Variablen ξ ist und

$$\int_{-\infty}^{+\infty} |x| f(x)\, dx < \infty$$

dann heißt

$$E\xi = \int_{-\infty}^{+\infty} x f(x)\, dx$$

Erwartungswert von ξ.

1.34 Beispiel (Gleichverteilung)
Haben ein gleichverteiltes ξ, d.h. $\xi \sim R(0,1)$. Daraus folgt für f:

$$f(x) = \begin{cases} 1 \text{ , falls } 0 \leq x < 1 \\ 0 \text{ , sonst} \end{cases}$$

Die Dichtefunktion ist in Abbildung 1.7 auf Seite 23 abgebildet.

$$E\xi = \int_0^1 x\, dx = \left.\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

Dieser Wert entspricht voll dem Erwartetem.

1.35 Beispiel (Normalverteilt)
Die zufällige Variable ist normalverteilt, d.h. $\xi \sim N(\mu, \sigma)$, der Kurvenverlauf der Normalverteilung ist in Abbildung 1.9 auf Seite 25 zu sehen.

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \\
E\xi &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx \\
& \quad z = \frac{x-\mu}{\sigma} \\
& \quad dx = \sigma\, dz \\
& \quad x = z \cdot \sigma + \mu \\
E\xi &= \int_{-\infty}^{+\infty} \frac{1 \cdot \sigma}{\sqrt{2\pi}\,\sigma} \cdot (z\sigma + \mu) \cdot e^{-\frac{z^2}{2}}\, dz \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot e^{-\frac{z^2}{2}}\, dz + \underbrace{\frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\, dz}_{\mu} \\
&= \frac{\sigma}{\sqrt{2\pi}} \underbrace{\left[-e^{-\frac{z^2}{2}} \right]_{-\infty}^{+\infty}}_{0} + \mu = \underline{\underline{\mu}}
\end{aligned}$$

1.21 Definition (Modalwert)
Der *Modalwert* ist der Punkt größter Dichte, d.h. $\max_x f(x)$

1.36 Beispiel
Bei der Normalverteilung ist μ der Modalwert und der Erwartungswert. Zur Veranschaulichung siehe Abbildung 1.9 auf Seite 25.

Eigenschaften des Erwartungswertes Nachfolgend sind ξ, ξ_1, ξ_2 zufällige Variablen und $a, b, c \in \mathbb{R}^1$.

- $E(c \cdot \xi) = c \cdot E\xi$
- $E(\xi_1 + \xi_2) = E\xi_1 + E\xi_2$
- $E(a\xi_1 + b\xi_2) = aE\xi_1 + bE\xi_2$
- $P(\xi = c) = 1 \Rightarrow E\xi = Ec = c$
- $P(\xi \geq c) = 1 \Rightarrow E\xi \geq c$

1.37 Beispiel
ξ habe die Dichte $f(x)$, d.h.

$$\begin{aligned} P(\xi \geq c) &= \int_c^\infty f(x)\,dx = 1 \\ P(\xi < c) &= \int_{-\infty}^c f(x)\,dx = 0 \\ E\xi &= \int_c^\infty x f(x)\,dx \geq c \underbrace{\int_c^\infty f(x)\,dx}_{1} \end{aligned}$$

1.9.2 Momente, Streuung und Varianz

1.22 Definition (Moment)
$E\xi^p$ heißt *p-tes Moment* der zufälligen Variablen ξ, falls $E|\xi|^p$ existiert.

$$E\xi^p = \begin{cases} \int_{-\infty}^{+\infty} x^p f(x)\,dx, & \text{falls } \xi \text{ stetig} \\ \sum_i x_i^p p_i & \text{, falls } \xi \text{ diskret} \end{cases}$$

1.23 Definition (zentrales Moment)
$E(\xi - E\xi)^p$ heißt *p-tes zentrales Moment* von ξ.

$E(\xi - E\xi)^2$ heißt auch *Streuung* bzw. *Varianz* von ξ: σ_ξ^2.

1.38 Beispiel (Varianz)
$\xi \sim B(p, 1)$

$$\Rightarrow \xi : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}; \qquad E\xi = p$$

$$\begin{aligned} \sigma_\xi^2 &= p^2(1-p) + (1-p)^2 \cdot p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \\ &= \underline{p(1-p)} \end{aligned}$$

Es gilt allgemein:

$$\begin{aligned} E(\xi - E\xi)^2 &= E(\xi^2 - 2\xi E\xi + (E\xi)^2) \\ &= E\xi^2 - E(2\xi E\xi) + E(E\xi)^2 \\ &= E\xi^2 - 2E\xi \cdot E\xi + (E\xi)^2 \\ &= \underline{E\xi^2 - (E\xi)^2} \end{aligned}$$

1.39 Beispiel (Varianz)
$\xi \sim PO(\lambda)$ — Poisson-verteilt

$$
\begin{aligned}
E\xi &= \lambda \\
\sigma_\xi^2 &= E\xi^2 - \lambda^2 \\
E\xi^2 &= \sum_{i=0}^{\infty} i^2 \frac{\lambda^i}{i!}\, e^{-\lambda} \\
& \quad \xi^2 : \begin{pmatrix} 0 & 1 & 4 & 9 & \dots \\ p_0 & p_1 & p_2 & p_3 & \dots \end{pmatrix} \\
i^2 &= i(i-1) + i \\
E\xi^2 &= \sum_{i=0}^{\infty} i(i-1)\frac{\lambda^i}{i!}\, e^{-\lambda} + \sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}\, e^{-\lambda} \\
&= \sum_{i=2}^{\infty} i(i-1)\frac{\lambda^i}{i!}\, e^{-\lambda} + \lambda \\
&= \lambda^2 \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!}\, e^{-\lambda} + \lambda \\
&= \lambda^2 \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}\, e^{-\lambda}}_{1} + \lambda \\
&= \underline{\lambda^2 + \lambda} \\
\sigma_{\xi^2} &= E\xi^2 - (E\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}
$$

Wir betrachten ein $\xi \sim R(0,1)$, d.h.

$$
f(x) = \begin{cases} 1, \text{ falls } 0 \le x < 1 \\ 0, \text{ sonst} \end{cases}
$$

Dann ist:

$$
\begin{aligned}
E\xi &= \int_0^1 x\, dx = \frac{1}{2} \\
E\xi^2 &= \int_0^1 x^2\, dx = \left.\frac{x^3}{3}\right]_0^1 = \frac{1}{3} \\
\sigma_\xi^2 &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}
\end{aligned}
$$

Haben nun ein normalverteiltes ξ, d.h. $\xi \sim N(\mu, \sigma)$.

$$
\begin{aligned}
E\xi &= \mu \\
E\xi^2 &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} x^2\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx \\
& \quad t = \frac{x-\mu}{\sigma} \\
& \quad x = \sigma t + \mu \\
& \quad dx = \sigma\, dt \\
E\xi^2 &= \frac{1\sigma}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} (\sigma t + \mu)^2\, e^{-\frac{1}{2}t^2}\, dt
\end{aligned}
$$

$$\begin{aligned}
&= \underbrace{\frac{1\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2\, e^{-\frac{1}{2}t^2}\, dt}_{\sigma^2} + \underbrace{\frac{2\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\, e^{-\frac{t^2}{2}}\, dt}_{0} + \underbrace{\frac{1\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\, dt}_{\mu^2} \\
&= \underline{\sigma^2 + \mu^2} \\
\sigma_\xi^2 &= E\xi^2 - (E\xi)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2
\end{aligned}$$

Dabei sind $\sigma_\xi^2 = \sigma^2$ die *Streuung* oder *Varianz* und $\sigma_\xi = \sqrt[+]{\sigma_\xi^2}$ die *Standardabweichung.*

1.9.2.1 Eigenschaften der Streuung

1. $P(\xi = c) = 1 \Longleftrightarrow \sigma_\xi^2 = 0$

Beweis:

$$\begin{aligned}
P(\xi = c) = 1 &\Rightarrow E\xi = c \\
\sigma_\xi^2 &= E(\xi - c)^2 \\
&= E(c-c)^2 = 0 \\
\sigma_\xi^2 = 0 &\Rightarrow \underbrace{E(\xi - c)^2}_{\geq 0} = 0 \\
&\Rightarrow P(\xi - c = 0) = 1
\end{aligned}$$

2. $\sigma_{\xi+b}^2 = \sigma_\xi^2; \quad b \in \mathbb{R}^1$

Beweis:

$$\begin{aligned}
\sigma_{\xi+b}^2 &= E(\xi + b - E(\xi + b))^2 = E(\xi + b - E\xi - Eb)^2 \\
&= E(\xi + b - E\xi - b)^2 = E(\xi - E\xi)^2 \\
&= \sigma_\xi^2
\end{aligned}$$

3. $\min\limits_{c \in \mathbb{R}^1} E(\xi - c)^2 = E(\xi - E\xi)^2 = \sigma_\xi^2$

$$\min_c \sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i \quad \text{(analog)}$$

Beweis:

$$\begin{aligned}
E(\xi - c)^2 &= E(\xi - E\xi + E\xi - c)^2 \\
&= \underbrace{E(\xi - E\xi)^2}_{\sigma_\xi^2} + 2E((E\xi - c)(\xi - E\xi)) \\
&= \sigma_\xi^2 + 2(E\xi - c)E(\xi - E\xi) + (E\xi - c)^2 \\
& E(\xi - E\xi) = E\xi - E(E\xi) = E\xi - E\xi = 0 \\
&= \sigma_\xi^2 + (E\xi - c)^2 \geq \sigma_\xi^2
\end{aligned}$$

4. ξ_1, ξ_2 unabhängig $\Rightarrow \sigma_{\xi_1+\xi_2}^2 = \sigma_{\xi_1}^2 + \sigma_{\xi_2}^2$

Beweis:

$$\begin{aligned}
\sigma^2_{\xi_1+\xi_2} &= E((\xi_1+\xi_2)-E(\xi_1+\xi_2))^2 \\
&= E((\xi_1-E\xi_1)+(\xi_2-E\xi_2))^2 \\
&= E((\xi_1-E\xi_1)^2+2(\xi_1-E\xi_1)(\xi_2-E\xi_2)+(\xi_2-E\xi_2)^2) \\
&= E(\xi_1-E\xi_1)^2+2E((\xi_1-E\xi_1)(\xi_2-E\xi_2))+E(\xi_2-E\xi_2)^2 \\
&= \sigma^2_{\xi_1}+\underbrace{2m_{11}}_{0}+\sigma^2_{\xi_2} \\
&= \sigma^2_{\xi_1}+\sigma^2_{\xi_2} \\
&\quad \text{da } m_1 1 = 0 \text{ wegen der Unabhängigkeit von } \xi_1 \text{ und } \xi_2\text{, siehe Lemma 1.8}
\end{aligned}$$

5. $$\sigma^2_{a\xi} = a^2 \cdot \sigma^2_\xi$$

1.24 Definition (Kovarianz)
Für einen zufälligen Vektor $\xi = \binom{\xi_1}{\xi_2}$ ist

$$m_{11} = E((\xi_1 - E\xi_1)(\xi - E\xi_2))$$

die *Kovarianz* zwischen ξ_1 und ξ_2.

1.8 Lemma
Wenn ξ_1 und ξ_2 unabhängig sind, dann gilt: $m_{11} = 0$.

Beweis: $f(x_1, x_2)$ sei die Dichte von ξ.

$$\begin{aligned}
m_{11} &= \iint_{-\infty}^{+\infty} (x_1-\mu_1)(x_2-\mu_2) f(x_1,x_2)\,dx_1\,dx_2 \\
&\quad \text{wobei } \mu_i = E\xi_i \\
\xi_1, \xi_2 &\quad \text{unabhängig} \iff \\
&\quad f(x_1,x_2) = f_1(x_1)\cdot f_2(x_2) \\
&\quad f_i(x_i) \text{ — Randdichte von } \xi_i \\
m_{11} &= \int_{-\infty}^{+\infty} (x_1-\mu_1) f_1(x_1)\,dx_1 \cdot \int_{-\infty}^{+\infty} (x_2-\mu_2) f_2(x_2)\,dx_2 \\
&= E(\xi_1 - E\xi_1)\cdot E(\xi_2 - E\xi_2) \\
&\quad E(\xi_i - E\xi_i) = 0 \\
&\Rightarrow m_{11} = 0
\end{aligned}$$

Man kann sich nun fragen, ob $m_{11} = 0 \Rightarrow \xi_1, \xi_2$ unabhängig gilt?

Gegenbeispiel:

$$\xi_1 \sim R(0, \pi)$$

$$f(x) = \begin{cases} \frac{1}{\pi},\ 0 \le x \le \pi \\ 0,\ \text{sonst} \end{cases}$$

Setzen $\xi_2 = \sin \xi_1$, d.h. ξ_1, ξ_2 sind streng abhängig.

$$\begin{aligned}
m_{11} &= E((\xi_1 - E\xi_1)(\xi_2 - E\xi_2)) \\
&= E(\xi_1\xi_2 - \xi_1 E\xi_2 - \xi_2 E\xi_1 + E\xi_1 E\xi_2) \\
&= E(\xi_1\xi_2) - E(\xi_1 E\xi_2) - E(\xi_2 E\xi_1) + E\xi_1 E\xi_2 \\
&= E(\xi_1\xi_2) - E\xi_1 E\xi_2 \\
E\xi_1 &= \int_0^\pi x\frac{1}{\pi}\,dx = \frac{1}{\pi}\frac{\pi^2}{2} = \frac{\pi}{2} \\
E\xi_2 &= \int_0^\pi \sin x\frac{1}{\pi}\,dx = \frac{1}{\pi}[-\cos x]_0^\pi = \frac{1}{\pi}(1+1) = \frac{2}{\pi} \\
\xi_2 &= g(\xi_1) = \sin \xi_1 \\
Eg(\xi_1) &= \int g(x)f(x)\,dx \\
& \quad x_1x_2 = x\sin x = g(x) \\
E(\xi_1\xi_2) &= \frac{1}{\pi}\int_0^\pi x\sin x\,dx \\
&= \left[-\frac{1}{\pi}x\cos x\right]_0^\pi + \underbrace{\frac{1}{\pi}\int_0^\pi \cos x\,dx}_{0} \\
&= -\frac{1}{\pi}(-\pi) = 1 \\
m_{11} &= E(\xi_1\xi_2) - \underbrace{\frac{\pi}{2}\frac{2}{\pi}}_{1} = 0
\end{aligned}$$

1.9.3 Korrelation

1.25 Definition (Korrelationskoeffizient)
Wenn ξ_1, ξ_2 zufällige Variablen sind, heißt

$$\varrho(\xi_1, \xi_2) = \frac{m_{11}}{\sigma_{\xi_1}\sigma_{\xi_2}}$$

Korrelationskoeffizient zwischen ξ_1 und ξ_2.

$0 < \sigma_{\xi_1}, \sigma_{\xi_2} < \infty$

ξ_1, ξ_2 unabhängig $\Rightarrow \varrho(\xi_1, \xi_2) = 0$. Die Umkehrung gilt im allgemeinen nicht.

1.9 Satz
Für zufällige Variablen ξ_1 und ξ_2 gilt:

$$-1 \le \varrho(\xi_1, \xi_2) \le 1$$

Beweis: Definieren hierzu eine Funktion A:

$$A(t,u) := E[\underbrace{t(\xi_1 - E\xi_1) + u(\xi_2 - E\xi_2)}_{\ge 0}]^2 \ge 0$$

$$
\begin{aligned}
A(t,u) &= E(t^2(\xi_1 - E\xi_1)^2 + 2tu(\xi_1 - E\xi_1)\cdot(\xi_2 - E\xi_2) + u^2(\xi_2 - E\xi_2)^2) \\
&= t^2\sigma_{\xi_1}^2 + 2tum_{11} + u^2\sigma_{\xi_2}^2 \geq 0 \qquad \forall t,u \in \mathbb{R}^1 \\
\frac{A(\sigma_{\xi_2},\sigma_{\xi_1})}{\sigma_{\xi_1}\sigma_{\xi_2}} &= \frac{\sigma_{\xi_2}^2\cdot\sigma_{\xi_1}^2 + 2\sigma_{\xi_1}\sigma_{\xi_2}m_{11} + \sigma_{\xi_1}^2\sigma_{\xi_2}^2}{\sigma_{\xi_1}\sigma_{\xi_2}} \\
&= \sigma_{\xi_1}\sigma\xi_2 + 2m_{11} + \sigma_{\xi_1}\sigma_{\xi_2} \\
&\Rightarrow \boxed{\sigma_{\xi_1}\sigma_{\xi_2} + m_{11} \geq 0} \\
\frac{A(-\sigma_{\xi_2},\sigma_{\xi_1})}{\sigma_{\xi_1}\sigma_{\xi_2}} &\geq 0 \\
&\Rightarrow \boxed{\sigma_{\xi_1}\sigma_{\xi_2} - m_{11} \geq 0} \\
&\Rightarrow -\sigma_{\xi_1}\sigma_{\xi_2} \leq m_{11} \leq \sigma_{\xi_1}\sigma_{\xi_2} \\
&\Rightarrow -1 \leq \underbrace{\frac{m_{11}}{\sigma_{\xi_1}\sigma_{\xi_2}}}_{\varrho(\xi_1,\xi_2)} \leq +1
\end{aligned}
$$

1.10 Satz
Es gilt:

$$|\varrho(\xi_1,\xi_2)| = 1 \Longleftrightarrow P(\xi_1 = a\xi_2 + b) = 1 \qquad a,b \in \mathbb{R}^1$$

Beweis: $(\Rightarrow) P(\xi_1 = a\xi_2 + b) = 1$
Es gelten für $E\xi_1$ und σ_{ξ_1}:

$$E\xi_1 = E(a\xi_2 + b) = aE\xi_2 + b$$

$$\sigma_{\xi_1}^2 = \sigma_{a\xi_2+b}^2 = \sigma_{a\xi_2}^2 = a^2\sigma_{\xi_2}^2$$

$$
\begin{aligned}
\varrho(\xi_1,\xi_2) &= \frac{m_{11}}{\sigma_{\xi_1}\sigma_{\xi_2}} = \frac{E((\xi_1 - E\xi_1)(\xi_2 - E\xi_2))}{a\sigma_{\xi_2}\cdot\sigma_{\xi_2}} \\
&= \frac{E(((a\xi_2 + b) - (aE\xi_2 + b))(\xi_2 - E\xi_2)}{|a|\cdot\sigma_{\xi_2}^2} \\
&= \frac{E(a(\xi_2 - E\xi_2)(\xi_2 - E\xi_2))}{|a|\sigma_{\xi_2}^2} \\
&= \frac{a\cdot\sigma_{\xi_2}^2}{|a|\sigma_{\xi_2}^2} = \underline{sgn(a)}, \text{ d.h. } |\sigma| = 1
\end{aligned}
$$

$(\longleftarrow)\varrho^2(\xi_1,\xi_2) = 1$

$$
\begin{aligned}
\varrho(\xi_1,\xi_2) &= \frac{E(\xi_1 - E\xi_1)(\xi_2 - E\xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}} \\
&= E\Bigg(\underbrace{\frac{\xi_1 - E\xi_1}{\sigma_{\xi_1}}}_{\xi_1^*}\underbrace{\frac{\xi_2 - E\xi_2}{\sigma_{\xi_2}}}_{\xi_2^*}\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{\xi_i^*}^2 &= \sigma\left(\frac{\xi - E\xi_i}{\sigma_{\xi_i}}\right) \\
&= \frac{1}{\sigma_{\xi_i}^2}\sigma_{(\xi_i - E\xi_i)}^2 \\
&= \frac{1}{\sigma_{\xi_i}^2}\sigma_{\xi_i}^2 = 1
\end{aligned}
$$

$$\begin{aligned} E\left(\frac{1}{\sigma_{\xi_i}}(\xi_i - E\xi_i)\right) &= \frac{1}{\sigma_{\xi_i}}(E\xi_i - E(E\xi_i)) \\ &= \frac{1}{\sigma_{\xi_i}}(E\xi_1 - E\xi_i) = 0 \end{aligned}$$

$E\xi_i^* = 0$; $\quad \sigma_{\xi_i^*} = 1$ — solche Zufallsgrößen heißen *standardisiert*

$$\varrho(\xi_1, \xi_2) = E(\xi_1^* \xi_2^*)$$

1. $\varrho(\xi_1, \xi_2) = 1$

$$\begin{aligned} \sigma^2_{\xi_1^* - \xi_2^*} &= E((\xi_1^* - \xi_2^*) - \underbrace{E(\xi_1^* - \xi_2^*)}_{0})^2 \\ &= E(\xi_1^* - \xi_2^*)^2 \\ &= E\xi_1^{*2} - 2E\xi_1^*\xi_2^* + E\xi_2^{*2} \\ E\xi_i^{*2} &= E(\xi_i^* - E\xi_i^*)^2 \\ &= \sigma^2_{i^*} = 1 \\ \Rightarrow & \\ \sigma^2_{\xi_1^* - \xi_2^*} &= 2 - 2\varrho(\xi_1, \xi_2) = 0 \end{aligned}$$

$$\sigma^2_{\xi_1^* - \xi_2^*} = 0 \Longleftrightarrow P(\xi_1^* - \xi_2^* = const) = 1$$

Da $E\xi_1^* = E\xi_2^* \Rightarrow const = 0 \Rightarrow P(\xi_1^* = \xi_2^*) = 1$.

$$\begin{aligned} P(\xi_1^* = \xi_2^*) &= P\left(\frac{\xi_1 - E\xi_1}{\sigma_{\xi_1}} = \frac{\xi_2 - E\xi_2}{\sigma_{\xi_2}}\right) \\ &= P\left(\xi_1 = \frac{\sigma_{\xi_1}\xi_2 - \sigma_{\xi_1}E\xi_2}{\sigma_{\xi_2}} + E\xi_1\right) \\ &= P\left(\xi_1 = \underbrace{\frac{\sigma_{\xi_1}}{\sigma_{\xi_2}}}_{a>0}\xi_2 - \underbrace{\frac{\sigma_{\xi_1}}{\sigma_{\xi_2}}E\xi_2 + E\xi_1}_{-b}\right) \\ &= \underline{1} \end{aligned}$$

2. $\varrho(\xi_1, \xi_2) = -1$

$$\Rightarrow \sigma_{\xi_1^* + \xi_2^*} = 0$$

$$\Rightarrow P(\xi_1 = a\xi_2 + b) = 1, \qquad a < 0$$

ansonsten wie oben zu beweisen.

1.10 Einfache Lineare Regression

Es sei

$$V(a, b) = E(\xi_1 - (a\xi_2 + b))^2$$

Gesucht sind a^*, b^*, so daß

$$\min_{a,b} V(a, b) = V(a^*, b^*).$$

a^* - heißt *Regressionskoeffizient* von ξ_2 bzgl. ξ_1
b^* - heißt *Regressionskonstante*

Bestimmung von a^*, b^*:

$$\frac{\partial V}{\partial a} = 0 \qquad \frac{\partial V}{\partial b} = 0$$

$$\begin{aligned}
\frac{\partial V}{\partial a} &= 2E((\xi_1 - a\xi_2 - b)\xi_2) = 0 \\
\frac{\partial V}{\partial b} &= 2E(\xi_1 - a\xi_2 - b) = 0 \\
&= E\xi_1 - a^*E\xi_2 - b^* = 0 \\
& \boxed{b^* = E\xi_1 - a^*E\xi_2}
\end{aligned}$$

$$E((\xi_1 - a^*\xi_2 - b^*)\xi_2) = 0$$

$$\begin{aligned}
& E(\xi_1, \xi_2) - a^*E\xi_2^2 - b^*E\xi_2 \\
&= E(\xi_1, \xi_2) - a^*E\xi_2^2 - (E\xi_1 - a^*E\xi_2)E\xi_2 \\
&= E(\xi_1, \xi_2) - E\xi_1 E\xi_2 - a^*(E\xi_2^2 - (E\xi_2)^2) \\
&= m_{11} - a^*\sigma_{\xi_2}^2 = 0
\end{aligned}$$

$$a^* = \frac{m_{11}}{\sigma_{\xi_2}^2} = \frac{m_{11}}{\sigma_{\xi_1}\sigma_{\xi_2}} \frac{\sigma_{\xi_1}}{\sigma_{\xi_2}} = \underline{\varrho \frac{\sigma_{\xi_1}}{\sigma_{\xi_2}}}$$

1.10.1 Anwendung in der Statistik — Methode der kleinsten Quadrate

Gegeben ist eine Punktmenge (x_i, y_i). Hierzu ist eine Gerade $y = ax + b$ gesucht, die

$$\sum_{i=1}^{n} (y_i - ax_i - b)^2$$

minimiert.

(x_i, y_i) können wir als Beobachtungen zweier zufälliger Variablen ξ und η auffassen, wobei wir in Übereinstimmung mit der Bestimmung von a^*, b^* vereinbaren: $\eta = \xi_1; \xi := \xi_2$..

Lineare Regression

$$\begin{aligned}
\hat{a}_0 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\hat{b}_0 &= \bar{y} - \hat{a}_0\bar{x} \\
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \quad (\approx E\xi) \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (\approx E\eta) \\
\varrho(\xi, \eta) &= \frac{cov(\xi, \eta)}{\sigma_\xi \sigma_\eta} \approx r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})}}
\end{aligned}$$

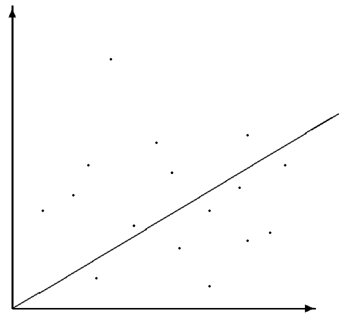

Abbildung 1.16: Darstellung der Punktmenge und durchgehende Gerade

Multiple Lineare Regression

$$\min \sum_{i=1}^{p} \left(y_i - a_0 - \sum_{j=1}^{n} a_j x_{ij} \right)^2$$

$$\boxed{x_{ij} = x_i^j}$$

$$\min \sum_{i=1}^{p} \left(y_i - a_0 - \sum_{j=1}^{n} a_j x_i^j \right)^2$$

Polynomialregression als Spezialfall der multiplen Regression

Spezialfall einer nichtlinearen Regression:

$$\begin{aligned} f(x) &= (\ln(ax+b))^c \\ \sum_{i=1}^{n}(y_i - f(x_i))^2 &= \sum_{i=1}^{n}(y_i - (\ln(ax_i+b))^c)^2 \\ &\quad \text{ist zu minimieren über } a, b \text{ und } c \end{aligned}$$

Hierbei muß im Gegensatz zur linearen Regression ein nichtlinerares Gleichungssystem zur Bestimmung von a, b, c gelöst werden.

1.11 Ungleichungen

1.11.1 Ungleichung von Jensen

1.11 Satz
Wenn g eine konvexe Funktion ist ($g''(x) > 0$) und ξ eine zufällige Variable, dann gilt:

$$Eg(\xi) \geq g(E\xi)$$

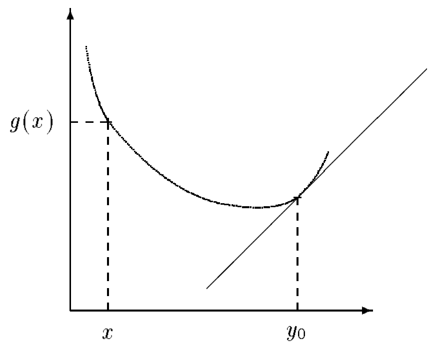

Abbildung 1.17: Verdeutlichung der Ungleichung von JENSEN

Beweis:

$$\begin{aligned}
g(x) &\geq g(y_0) + (x - y_0)m_{y_0} \qquad m_{y_0} = g'(y_0) \\
& \quad y_0 = E\xi, \quad x = \xi \\
g(\xi) &\geq g(E\xi) + (\xi - E\xi)m_{E\xi} \\
Eg(\xi) &\geq E(g(E\xi) + (\xi - E\xi)m_{E\xi}) = g(E\xi) + m_{E\xi}E(\xi - E\xi) \\
&= g(E\xi) + m_{E\xi}\underbrace{(E\xi - E(E\xi))}_{0}
\end{aligned}$$

1.40 Beispiel

1. $g(x) = x^2$

   $E\xi^2 \geq (E\xi)^2$

   $\sigma_\xi^2 = E(\xi - E\xi)^2 = E\xi^2 - (E\xi)^2 \geq 0$

2. $g(x) = |x|$

   $E|\xi| \geq |E\xi|$

3. $g(x) = \ln x$ — eine konkave Funktion

   $E(\ln \xi) \leq \ln(E\xi)$

1.11.2 Markoff'sche Ungleichung

1.12 Satz
Es gilt:

$$P(|\xi| > c) \leq \frac{E|\xi|}{c} \qquad c > 0$$

Beweis:

$$\begin{aligned}
\xi &: (\Omega, \mathcal{E}, P) \to (\mathbb{R}^1, \mathbb{B}^1, P_\xi) \\
& \quad P_\xi(B) = P(\xi^{-1}(B)) \quad \forall B \in \mathbb{B}^1 \\
\eta(\omega) &= cI_{\{\omega_i : |\xi(\omega)| > c\}} = \begin{cases} c, \text{ falls } |\xi(\omega)| > c \\ 0 \text{ sonst} \end{cases} \\
0 &\leq \eta(\omega) \leq |\xi(\omega)|
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \quad E\eta \leq E|\xi| \\
\eta \quad &: \quad \begin{pmatrix} 0 & c \\ \underbrace{P(|\xi| \leq c)}_{p_0} & \underbrace{P(|\xi| > c)}_{p_1} \end{pmatrix} \\
P(\eta = 0) \quad &= \quad p_0, \quad P(\eta = c) = p_1 \\
E\eta \quad &= \quad 0 \cdot p_0 + c \cdot p_1 \\
&= \quad c \cdot P(|\xi| > 0) \leq E|\xi|
\end{aligned}$$

$$\boxed{P(|\xi| > c) \leq \frac{E|\xi|}{c}}$$

1.11.3 Tschebyscheff'sche Ungleichung

Ersetzen ξ durch $(\xi - E\xi)^{2i}$, $\quad i = 1, 2, 3, \ldots$
Ersetzen c durch ε^{2i}; $\quad \varepsilon > 0$

$$P((\xi - E\xi)^{2i} > \varepsilon^{2i}) \leq \frac{E(\xi - E\xi)^{2i}}{\varepsilon^{2i}}$$

$$\Longleftrightarrow P(|\xi - E\xi| > \varepsilon) \leq \frac{E(\xi - E\xi)^{2i}}{\varepsilon^{2i}}$$

$\boxed{i = 1}$

$$P(|\xi - E\xi| > \varepsilon) \leq \frac{\sigma_\xi^2}{\varepsilon^2}, \qquad 0 < \sigma_\xi^2 < \infty$$

1.13 Satz (Das schwache Gesetz der großen Zahlen)
Es seien $\xi_1, \xi_2, \ldots, \xi_n$ unabhängige zufällige Variable mit $E\xi_i = \mu$ und $var\xi_i = \sigma^2$.

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

Dann gilt:

$$\lim_{n \to \infty} P(|\bar{\xi} - \mu| > \varepsilon) = 0 \qquad \forall \varepsilon > 0$$

Beweis: Unter Verwendung der TSCHEBYSCHEFF'schen Ungleichung:

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

$$\begin{aligned}
E\bar{\xi} \quad &= \quad E\left(\frac{1}{n} \sum_{i=1}^{n} \xi_i\right) = \frac{1}{n} \sum_{i=1}^{n} E\xi_i \\
&= \quad \frac{1}{n} \sum_{i=1}^{n} \mu = \underline{\mu} \\
\sigma_{\bar{\xi}^2} \quad &= \quad E(\bar{\xi} - E\bar{\xi})^2 = E(\bar{\xi} - \mu)^2 \\
&= \quad var\left(\frac{1}{n} \sum_{i=1}^{n} \xi_i\right) = (\frac{1}{n})^2 \cdot var\left(\sum_{i=1}^{n} \xi_i\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i=1}^{n} var(\xi_i), \quad \text{da } \xi_i \text{ unabhängig} \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \\
P(|\bar{\xi} - \mu| > \epsilon) &\leq \frac{\sigma_{\bar{\xi}^2}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \overset{n\to\infty}{\Rightarrow} 0
\end{aligned}
$$

Die Vorraussetzung: „$\xi_1, \ldots, \xi_n$ unabhängig“ kann zu „$\xi_1, \ldots, \xi_n$ unkorrelliert“ abgeschwächt werden.
unkorrelliert: $cov(\xi_i, \xi_j) = 0 \quad i \neq j$

1.41 Beispiel
Ereignis A, $p = P(A)$ ist unbekannt. Zur Schätzung von $P(A)$ führen wir eine Reihe von unabhängigen Experimenten durch. Es tritt entweder A oder $\bar{A}$ ein.

$$
\begin{aligned}
n &\quad - \quad \text{Anzahl der Experimente} \\
n(A) &\quad - \quad \text{Häufigkeit des Auftretens von } A \\
\hat{p}_n = \frac{n(A)}{n} &\quad - \quad \text{relative Häufigkeit von } A \text{ bei } n \text{ Versuchen}
\end{aligned}
$$

Frage: $\hat{p}_n \longrightarrow p$

$$
\text{Definieren } \xi_i = \begin{cases} 1, \text{ falls } A \text{ im } i\text{-ten Experiment eintritt} \\ 0, \text{ falls } A \text{ im } i\text{-ten Experiment nicht eintritt} \end{cases}
$$

$$
\begin{aligned}
&\Rightarrow \quad \xi_i \sim B(1,p); \quad P(\xi_i = 1) = p, P(\xi_i = 0) = 1 - p \\
&\qquad Var\xi_i = p(1-p) \\
&\qquad E\xi_i = p \\
\bar{\xi}_n &= \frac{1}{n}\sum_{i=1}^{n} \xi_i = \frac{1}{n} n(A) = \hat{p}_n \\
\lim_{n\to\infty} P(|\hat{p}_n - p| > \varepsilon) &= 0 \qquad \forall \varepsilon > 0 \\
P(|\hat{p}_n - p| > \varepsilon) &\leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \\
&\qquad \text{da } p(1-p) \leq \frac{1}{4}, \text{ siehe Abbildung 1.18}
\end{aligned}
$$

1.42 Beispiel

1. Wir haben $\xi_1, \ldots, \xi_n$, die unkorreliert sind. Die Verteilungsfunktion $F(x)$ mit
$$F(x) = P(\xi_i < x) \quad \forall i = 1, 2, \ldots, n \quad \forall x \in \mathbb{R}$$
soll durch
$$F_n(x) := \frac{\#\{\xi_i : \xi_i < x\}}{n}$$
approximiert werden. $F_n(x)$ ist die *empirische* Verteilungsfunktion. $\xi_{(1)}$ sei der kleinste Wert der ξ_i, $\xi_{(2)} \geq \xi_{(1)}$ der zweitkleinste Wert usw., d.h. $\xi_{(1)} \leq \xi_{(2)} \leq \ldots \leq \xi_{(n)}$
Es gilt:
$$\lim_{n\to\infty} P(|F_n(x) - F(x)| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

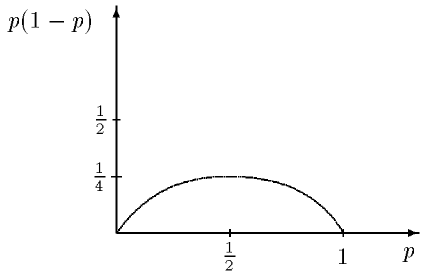

Abbildung 1.18: Veranschaulichung von p(1-p)

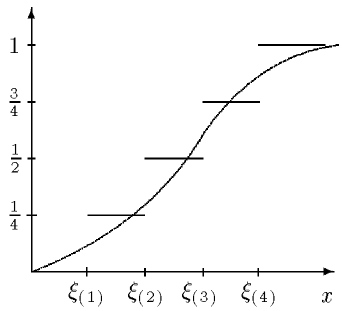

Abbildung 1.19: Vergleich von $F_n(x)$ und $F(x)$ für $n = 4$

2. $\xi_1, \ldots, \xi_n$ sind unabhängig

$$F(x) = P(\xi_i < x) \Rightarrow P\left(\lim_{n\to\infty} F_n(x) = F(x)\right) = 1$$

3. weitere Verschärfung, auch mit $\xi_1, \ldots, \xi_n$ sind unabhängig

$$\boxed{P\left(\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0\right) = 1}$$

Dies ist der *Satz von* Glivenko-Cantelli, auch genannt *Hauptsatz der Stochastik*

1.12 Grenzwerte

1.26 Definition (stochastische Konvergenz)
Eine Folge $\{\xi_n\}$ zufälliger Variablen *konvergiert stochastisch* gegen eine zufällige Variable ξ_0, falls

$$\lim_{n\to\infty} P(|\xi_n - \xi_0| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

Dies wird symbolisch mit

$$\text{p-lim } \xi_n = \xi_0$$

bezeichnet.

Das Schwache Gesetz der großen Zahlen $\{\xi_n\}$ ist eine Folge zufälliger Variablen mit

$$cov(\xi_i, \xi_j) = \delta_{ij} \cdot \sigma^2 = \begin{cases} \sigma^2, \text{ falls } i = j \\ 0 \text{ , sonst} \end{cases}$$

$$E\xi_i = \mu$$

$$\Rightarrow \text{p-lim } \bar{\xi}_n = \mu$$

1.27 Definition (fast sichere Konvergenz)
Eine Folge $\{\xi_n\}$ heißt *fast sicher konvergent gegen* ξ_0, falls

$$P\left(\left\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi_0(\omega)\right\}\right) = 1$$

Dies wird symbolisch mit

$$\lim \xi_n = \xi_0 \text{ f.s.}$$

bezeichnet (für f.s. – *fast sicher* kann auch a.e. – *almost everywhere* stehen).

1.43 Beispiel
Wir konstruieren eine Folge mit p-lim $\xi_n = 0$ aber $P(\omega : \lim \xi_n = 0) = 0$.

$$\begin{aligned}
\Omega &= [0,1] \\
\mathcal{E} &= [0,1] \cap \mathbb{B}^1 \\
0 &\le P(A) = \int_A 1\,dx \le 1 \qquad A \subseteq [0,1] \\
A_n &\in \mathcal{E} \\
A_n &= [k \cdot 2^{-h}, (k+1) \cdot 2^{-h}] \qquad h, k \in Z^+ \cup \{0\} \\
& \quad n = 2^h + k \\
& \quad 0 \le k < 2^h \\
\xi_n(\omega) &= \begin{cases} 1, \text{ falls } \omega \in A_n \\ 0, \text{ sonst} \end{cases} \\
\text{p-lim } \xi_n &= 0 \\
&\Leftrightarrow \lim_{n\to\infty} P(|\xi_n - 0| > \varepsilon) = 0 \quad \forall 0 < \varepsilon \le 1 \\
&\Leftrightarrow \lim_{n\to\infty} P(|\xi_n| > \varepsilon) = 0 \\
P(|\xi_n| > \varepsilon) &= P(|\xi_n| = 1) = P(A_n) = 2^{-h} < \frac{2}{n} \overset{n\to\infty}{\Rightarrow} 0 \\
& \quad \text{d.h. p-lim } \xi_n = 0 \\
& \quad n < 2^h + 2^h = 2 \cdot 2^h \\
& \quad 2^h > \frac{n}{2} \\
& \quad 2^{-h} < \frac{2}{n}
\end{aligned}$$

$$\begin{aligned}
&1 = 2^0 + 0 \to h = k = 0 \quad , A_1 = [0,1] \\
&2 = 2^1 + 0 \to h = 1, k = 0, \; A_2 = [0, \tfrac{1}{2}] \\
&3 = 2^1 + 1 \to h = k = 1 \quad , A_3 = [\tfrac{1}{2}, 1] \\
&4 = 2^2 + 0 \to h = 2, k = 0, \; A_4 = [0, \tfrac{1}{4}] \\
&5 = 2^2 + 1 \to h = 2, k = 1, \; A_5 = [\tfrac{1}{4}, \tfrac{1}{2}] \\
&6 = 2^2 + 2 \to h = k = 2 \quad , A_6 = [\tfrac{1}{2}, \tfrac{3}{4}] \\
&7 = 2^2 + 3 \to h = 2, k = 3, \; A_7 = [\tfrac{3}{4}, 1] \\
&8 = 2^3 + 0 \to h = 3, k = 0, \; A_8 = [0, \tfrac{1}{8}]
\end{aligned}$$

Der Grenzwert $\lim_{n\to\infty} A_n$ existiert nicht, ξ_n konvergiert nirgends.

1.14 Satz

Wir haben $\xi_0 = \text{p-lim } \xi_n$ und $\eta_0 = \text{p-lim } \xi_n$. Daraus folgt:

$$P(\xi_0 = \eta_0) = 1$$

Beweis: Wir nehmen ein $\varepsilon > 0$ beliebig.

$$\begin{aligned}
&P(\omega : |\xi_0(\omega) - \eta_0(\omega)| > \varepsilon) \\
&= P(\omega : |\xi_0(\omega) - \xi_n(\omega) + \xi_n(\omega) - \eta_0(\omega)| > \varepsilon) \\
&\leq P(\omega : \underbrace{|\xi_0 - \xi_n|}_{a} + \underbrace{|\xi_n - \eta_0|}_{b} > \varepsilon) \\
&\quad \text{siehe Abbildung 1.20} \\
&\leq P(\{\omega : |\xi_n - \xi_0| > \frac{\varepsilon}{2}\} \cup \{\omega | \xi_n - \eta_0| > \frac{\varepsilon}{2}\}) \\
&\leq P(\omega : |\xi_n - \xi_0| > \frac{\varepsilon}{2}) + P(\omega : |\xi_n - \eta_0| > \frac{\varepsilon}{2}) \stackrel{n \to \infty}{\longrightarrow} 0 \\
&\Rightarrow \underline{P(\omega : \xi_0(\omega) = \eta_0(\omega)) = 1}
\end{aligned}$$

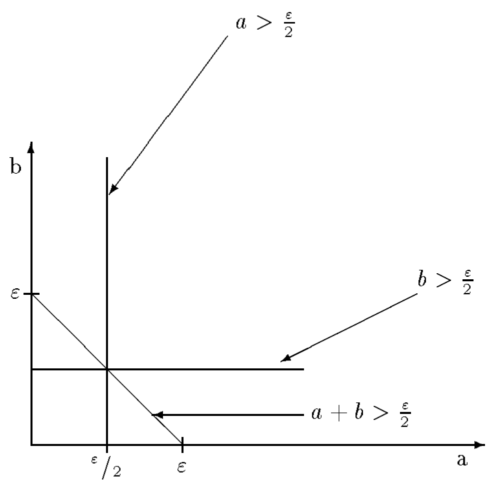

Abbildung 1.20: Geometrischer Beweis für $\{a + b > \varepsilon\} \subset \{a > \frac{\varepsilon}{2}\} \cup \{b > \frac{\varepsilon}{2}\}$

1.44 Beispiel (Fast sichere Konvergenz)

$\{A_n\}$ ist eine Folge von Ereignissen, wir definieren

$$\lim_{n \to \infty} \sup A_n := \lim_{n \to \infty} \Bigg(\underbrace{\bigcup_{k=n}^{\infty} A_k}_{B_n} \Bigg) = \bigcap_{n=1}^{\infty} \Bigg(\bigcup_{k=n}^{\infty} A_k \Bigg)$$

$$B_{n+1} \subseteq B_n, \quad B_n \downarrow$$

1.15 Lemma (von Borel-Cantelli)

$\{A_n\}$ ist eine Folge von Ereignissen aus einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$. Dann gilt:

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P\left(\lim_{n \to \infty} \sup A_n\right) = 0$$

Beweis:

$$
\begin{aligned}
P\left(\bigcup_{k=n}^{\infty} A_k\right) &\leq \sum_{k=n}^{\infty} P(A_n) \stackrel{n\to\infty}{\longrightarrow} 0 \\
\left(\sum_{k=1}^{\infty} a_n\right. &< \left.\infty \Rightarrow \sum_{k=n}^{\infty} a_n < \varepsilon; \quad \text{sofern } n \text{ hinreichend groß}\right) \\
P\left(\lim_{n\to\infty} \sup A_n\right) &= P\left(\lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) = 0
\end{aligned}
$$

1.12.1 Übersicht über die Konvergenzen

Es gilt

$$
\begin{aligned}
&\lim_{n\to\infty} \xi(\omega) = \xi(\omega) \text{ f.s.} \\
\Longleftrightarrow\quad & P\left(\{\omega : \lim_{n\to\infty} \xi_n(\omega) = \xi_0(\omega)\}\right) = 1 \\
& \text{fast sichere Konvergenz} \\
\Longleftrightarrow\quad & P\left(\bigcap_{k=n}^{\infty} \{\omega : |\xi_k(\omega) - \xi_0(\omega)| < \varepsilon\}\right) \stackrel{n\to\infty}{\longrightarrow} 1 \\
& \text{(Konvergenz (modifiziert) der Analysis, Cauchy-Kriterium)} \\
\Longleftrightarrow\quad & P\left(\bigcup_{k=n}^{\infty} \{\omega : |\xi_k(\omega) - \xi_0(\omega)| \geq \varepsilon\}\right) \stackrel{n\to\infty}{\longrightarrow} 0 \\
& \text{Morgansche Gesetze} \\
\Longleftrightarrow\quad & P\left(\lim_{n\to\infty} \sup B_n\right) = \lim_{n\to\infty} P(B_n) = 0, \text{ wobei} \\
& B_n = \bigcup_{k=n}^{\infty} \{\omega : |\xi_k(\omega) - \xi_0(\omega)| \geq \varepsilon\} \downarrow
\end{aligned}
$$

1.16 Satz
Wenn $\lim\limits_{n\to\infty} \xi_n = \xi_0$ f.s., so gilt auch

$$\text{p-}\lim_{n\to\infty} \xi_n = \xi_0$$

Beweis:

$$
\begin{aligned}
0 &\leq \lim_{n\to\infty} P(\{\omega : |\xi_n(\omega) - \xi_0(\omega)| > \varepsilon\}) \\
&\leq \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} \{\omega : |\xi_k(\omega) - \xi_0(\omega)| > \varepsilon\}\right) = 0 \\
&\quad \text{da } \lim \xi_n = \xi_0 \text{ f.s.}
\end{aligned}
$$

1.12.2 Verteilungskonvergenz

1.28 Definition (Verteilungskonvergenz)
$\{\xi_n\}$ ist eine Folge von zufälligen Variablen,
ξ_0 ist eine zufällige Variable mit $P(\xi_0 < x) = F(x)$.
ξ_n *konvergiert der Verteilung nach gegen* ξ_0, wenn

$$\lim_{n\to\infty} P(\xi_n < x) = F(x), \text{ für alle } x, \text{ in denen } F(x) \text{ stetig ist}$$

symbolisch: $\xi_n \longrightarrow^{D} \xi_0$ (D-Distribution)

Es gilt:

$$\boxed{\lim_{n\to\infty} \xi_n = \xi_0 \text{ f.s.} \Rightarrow \text{p-}\lim \xi_n = \xi_0 \Rightarrow \xi_n \xrightarrow{D} \xi_0}$$

Da die Verteilungskonvergenz in dieser Kette die schwächste ist, heißt sie auch *schwache Konvergenz.*

1.17 Satz (Der zentrale Grenzwertsatz)
$\xi_1, \ldots, \xi_n, \ldots$ sind unabhängige Variablen mit:

$$P(\xi_i < x) = F(x)$$

(d.h. insbesondere: $E\xi_i = \mu, \quad var\xi_i = \sigma^2$)
mit $z_n := \sum_{i=1}^n \xi_i$ und $\bar{z}_n = \frac{z_n}{n} = \bar{\xi}_n$ gilt:

$$\lim_{n\to\infty} P\left(\frac{z_n - n\mu}{\sqrt{n}\,\sigma} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

Eine äquivalente Formulierung für diesen Satz ist:

$$\eta_n := \frac{z_n - n\mu}{\sqrt{n}\,\sigma}, \quad \xi_0 \sim N(0,1)$$

$$\eta_n \xrightarrow{D} \xi_0$$

$$\boxed{\begin{aligned} \frac{z_n - n\mu}{\sqrt{n}\,\sigma} &= \frac{n(\bar{z}_n - \mu)}{\sqrt{n}\,\sigma} \\ &= \sqrt{n}\,\frac{\bar{z}_n - \mu}{\sigma} \xrightarrow{D} \xi_0 \sim N(0,1) \end{aligned}}$$

1.18 Satz (von Berry-Esséen)
Es gilt:

$$\left| P\left(\frac{z_n - n\mu}{\sqrt{n}\,\sigma}\right) - \Phi(x) \right| < K, \text{ mit } K = \frac{0,8 \cdot M}{\sigma^3 \sqrt{n}}$$

$0 < M < \infty$ muß jeweils konkret bestimmt werden:

$$M = E|\xi_i - \mu|^3$$

1.45 Beispiel (gleichverteilt)
Wir haben über $[0,1)$ gleichverteilte ξ_i, d.h. $\xi_i \sim R(0,1)$.

$$\begin{aligned} \mu &= E\xi_i = \int_0^1 x\, dx = \frac{1}{2} \\ \sigma^2 &= E\xi_i^2 - \mu^2 = \int_0^1 x^2\, dx - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = -\frac{1}{12} \\ M &= \int_0^1 |x - \frac{1}{2}|^3\, dx = 2\int_{\frac{1}{2}}^1 (x - \frac{1}{2})^3\, dx = \frac{1}{32} \end{aligned}$$

Berry-Esséen-Schranke

n	K
12	0,3
100	0,104
1000	0,033

1.46 Beispiel (binomialverteilt)
Wir haben binomialverteilte $\xi_i \sim B(1,p)$.

$$\begin{aligned} z_n &= \sum_{i=1}^{n} \xi_i \sim B(n,p) \\ E\xi_i &= \mu = p \\ \sigma^2 &= p(1-p) \end{aligned}$$

Wenden nun auf z_n den Grenzwertsatz an:

$$\begin{aligned} P\left(\frac{z_n - np}{\sqrt{np(1-p)}} < x\right) &\approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \\ P(z_n < x) &= \sum_{i<x} \binom{n}{i} p^2 (1-p)^{n-i} \\ &= P\left(\frac{z_n - np}{\sqrt{np(1-p)}} < \frac{x - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Der Satz von De Moivre-Laplace sagt aus, daß für große n die Binomialverteilung durch die (einfachere) Normalverteilung ersetzt werden kann.

Kapitel 2

Information und Entropie

2.1 Entropie

Entropie bedeutet Unbestimmtheit. Zur Verdeutlichung betrachten wir ein Experiment G mit den Ausgängen $A_1, A_2, \ldots, A_m$. Das Eintreten eines Ereignisses (A) liefert eine Information, dabei gilt:

$$\begin{aligned} P(A) \text{ groß} &\quad \rightarrow \quad \text{Information klein} \\ P(A) \text{ klein} &\quad \rightarrow \quad \text{Information groß} \end{aligned}$$

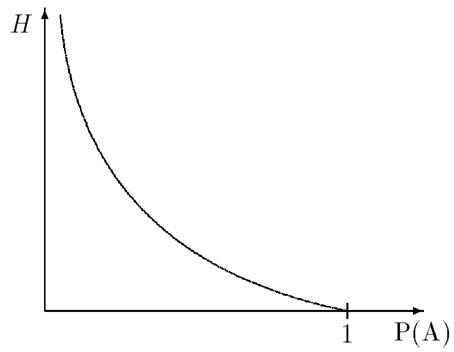

Abbildung 2.1: Das Verhältnis zwischen Entropie H und Wahrscheinlichkeit $P(A)$

Ein Ausgang, der eine große Information zu liefern in der Lage ist, hat eine große Unbestimmtheit (bzw. *Entropie*). Das Verhältnis zwischen Information $I(A)$ und Wahrscheinlichkeit $P(A)$ ist durch die Formel:

$$I(A) = -\log_c P(A)$$

beschrieben (siehe Abbildung 2.1).

Für die Wahrscheinlicheiten p_i gilt:

$$p_i = P(A_i) \geq 0, \qquad \sum_{i=1}^{m} p_i = 1$$

Die Information I ist eine Abbildung:

$$A_i \longrightarrow I(A_i) := -\log p_i$$

Es gilt:

$$I(\omega) = -\log p_i, \qquad \text{falls } \omega \in A_i$$

ist eine zufällige Variable
Für die Menge Ω aller Elementarereignisse ω gilt:

$$\Omega = A_1 \cup A_2 \cup \ldots \cup A_m$$

für die A_i soll gelten: $A_i \cap A_j = \emptyset \quad i \neq j$.

$$\boxed{EI = -\sum_{i=1}^{m} p_i \log p_i}$$

2.1 Definition (Entropie)
Unter obigen Vorraussetzungen heißt

$$H(G) = EI = -\sum_{i=1}^{m} p_i \log p_i$$

Entropie des Experiments G.
Wenn ξ eine zufällige Variable ist:

$$\xi : \begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ p_1 & p_2 & \ldots & p_m \end{pmatrix},$$

so definieren wir

$$H(\xi) = H(G),$$

sofern das Experiment G die Ausgänge

$$A_i = \{\omega : \xi(\omega) = x_i\} \quad (i = 1, \ldots, m)$$

mit $p_i = P(A_i)$ hat.

Die Entropie wird durch die Wahrscheinlichkeiten p_i bestimmt, d.h.: $H(G) = H(p_1, \ldots, p_m)$

2.1 Beispiel (Münzwurf)
Wenn wir als Basis des Logarithmus 2 annehmen, gilt: $\log p_i = \log_2 p_i$.
Beim Münzwurf sind die Wahrscheinlichkeiten: $p_1 = p_2 = \frac{1}{2}$.

$$H(p_1, p_2) = -2\frac{1}{2} \cdot \log_2 \frac{1}{2} = -\log_2 \frac{1}{2} = \log_2 2 = \underline{1}$$

Dieses Ergebnis entspricht den Erwartungen, daß mit zwei Zuständen ein *Bit* kodiert[1] werden kann.

[1] Das „kodiert“ ist nicht so fehl am Platze, wie wenig später noch zu sehen sein wird

Wir betrachten jetzt die Entropie für zwei zufällige Variablen.

$$\xi : \begin{pmatrix} x_1 & \dots & x_m \\ p_{1\cdot} & \dots & p_{m\cdot} \end{pmatrix} \qquad \eta : \begin{pmatrix} y_1 & \dots & y_n \\ p_{\cdot 1} & \dots & p_{\cdot n} \end{pmatrix}$$

$$\begin{aligned} (\xi, \eta) &\in \{x_1, \dots, x_m\} \times \{y_1, \dots, y_n\} \\ P(\xi = x_i, \eta = y_j) &= p_{ij} \\ p_{i\cdot} &= \sum_{j=1}^{n} p_{ij} \\ p_{\cdot j} &= \sum_{i=1}^{m} p_{ij} \\ H(\xi, \eta) &= -\sum_{i,j} p_{ij} \log p_{ij} \end{aligned}$$

2.1.1 Eigenschaften der Entropie

Für die Entropie nehmen wir folgende Rechenregeln an:

- $0 \cdot \log 0 = 0$
- $\frac{0}{0} = 0$

Diese Regeln lassen sich mit den Regeln von Bernoulli-l´Hospital beweisen.

1. $0 \leq H(\xi) \leq \log m$

   (a) $H(\xi) = 0 \Longleftrightarrow p_{i\cdot} \in \{0, 1\} \quad \forall i$

   (b) $H(\xi) = \log m \Longleftrightarrow p_{i\cdot} = \dfrac{1}{m} \quad \forall i \in \{1, \dots, m\}$

   Beweis:

   (a) $0 \leq x \leq 1 \Rightarrow -x \log x \geq 0 \Rightarrow H(\xi) = -\sum_{i=1}^{m} p_{i\cdot} \log p_{i\cdot} \geq 0$

   Falls $H(\xi) = 0$, so muß $p_{i\cdot} \log p_{i\cdot} = 0 \quad \forall i$

   $p_{i\cdot} \in \{0, 1\}$ und da $\sum_{i=1}^{n} p_{i\cdot} = 1$, muß genau ein $p_{i\cdot}$ den Wert 1 haben.

   (b) $\varphi(x) = x \cdot \log x; \quad x \in [0, 1]$

$$\begin{aligned} \varphi''(x) &= \frac{c}{x} > 0 \quad (c > 0) \\ & \quad q_i \geq 0, \sum_{i=1}^{m} q_i = 1 \\ \varphi\left(\sum_{i=1}^{m} q_i x_i\right) &\leq \sum_{i=1}^{m} q_i \varphi(x_i) \\ & \quad q_i = \frac{1}{m}, \quad x_i = p_{i\cdot} \\ \varphi\left(\sum_{i=1}^{m} q_i p_{i\cdot}\right) &= \varphi\left(\frac{1}{m}\sum_{i=1}^{m} p_{i\cdot}\right) = \sum_{i=1}^{m} \frac{1}{m} p_i \log\left(\frac{1}{m}\underbrace{\sum p_{i\cdot}}_{1}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{m}\sum_{i=1}^{m} p_{i\cdot} \log \frac{1}{m} \\
&\leq \sum_{i=1}^{m} q_i \varphi(p_{i\cdot}) = \frac{1}{m}\sum_{i=1}^{m} p_{i\cdot} \log p_{i\cdot}
\end{aligned}
$$

Damit gilt:

$$
\begin{aligned}
-\sum_{i=1}^{m} p_{i\cdot} \log \frac{1}{m} &= -\log \frac{1}{m} = \log m \\
&\geq -\sum_{i=1}^{m} p_{i\cdot} \log p_{i\cdot} = H(\xi)
\end{aligned}
$$

Zu zeigen ist nun:

$$H(\xi) = \log m \Longleftrightarrow p_{i\cdot} = \frac{1}{m} \quad \forall i$$

$$
\begin{aligned}
&H(\xi) - \log m \\
&\quad= -\sum_{i=1}^{m} p_{i\cdot} \log p_{i\cdot} - \sum_{i=1}^{m} p_{i\cdot} \log m \\
&\quad= \sum_{i=1}^{m} p_{i\cdot} \log \frac{1}{p_{i\cdot} m} = \log e \sum_{i=1}^{m} p_{i\cdot} \ln \frac{1}{p_{i\cdot} m} \\
&\quad\leq \log e \sum_{i=1}^{m} p_{i\cdot} \left(\frac{1}{p_{i\cdot} m} - 1 \right)
\end{aligned}
$$

$$\ln z \leq z - 1 \quad \forall z > 0$$

$$\ln z = z - 1 \Longleftrightarrow z = 1$$

$$\Rightarrow H(\xi) - \log m \leq \log e \left(\sum_{i=1}^{m} \frac{1}{m} - 1 \right) = 0$$

„=“ gilt genau dann, wenn $\dfrac{1}{p_{i\cdot} m} = 1 \quad \forall i = 1, \ldots, m$

$$\Longleftrightarrow p_{i\cdot} = \frac{1}{m}$$

$$
\begin{aligned}
H(\xi/\eta = y_j) &= -\sum_{i=1}^{m} p_{i/j} \log p_{i/j}, \\
&\qquad \text{wobei } p_{i/j} = \frac{p_{ij}}{p_{\cdot j}} = P(\xi = x_i/\eta = y_j) \\
H(\xi/\eta) &= \sum_{j=1}^{n} p_{\cdot j} H(\xi/\eta = y_j) \\
&= -\sum_{i}\sum_{j} p_{i/j} p_{\cdot j} \log p_{i/j} \\
&= -\sum_{i/j} p_{ij} \log p_{i/j}
\end{aligned}
$$

$H(\xi/\eta)$ heißt *bedingte Entropie von ξ unter η*. Analog gilt:

$$H(\eta/\xi) = -\sum_i \sum_j p_{ij} \log p_{j/i},$$

wobei $p_{j/i} = \dfrac{p_{ij}}{p_{i\cdot}}$ ist.

2. $$H(\xi, \eta) = H(\xi) + H(\eta/\xi) = H(\eta) + H(\xi/\eta)$$

Beweis:

$$\begin{aligned} H(\xi, \eta) &= -\sum_{i,j} p_{ij} \log p_{ij} \\ &= -\sum_i \left(\sum_j p_{ij} \log p_{i\cdot} + \sum_j p_{ij} \log \frac{p_{ij}}{p_{i\cdot}} \right) \\ &= -\sum_i p_{i\cdot} \log p_{i\cdot} - \sum_i \sum_j p_{ij} \log p_{j/i} \\ &= H(\xi) + H(\eta/\xi) \end{aligned}$$

3. $$0 \leq H(\xi/\eta) \leq H(\xi)$$

(a) $H(\xi/\eta) = 0 \Longleftrightarrow p_{i/j} \in \{0, 1\} \quad \forall i, j$

(b) $H(\xi/\eta) = H(\xi) \Longleftrightarrow$ ξ und η stochastisch unabhängig

$\Leftrightarrow p_{ij} = p_{i\cdot} p_{\cdot j} \quad \forall i, j$

Beweis:

(a) $p_{ij} \geq 0$

$$\begin{aligned} H(\xi/\eta) &= -\sum_{i,j} p_{ij} \log p_{i/j} \\ &= -\sum_j p_{\cdot j} \underbrace{\sum_i p_{i/j} \log p_{i/j}}_{\leq 0 \forall j} \end{aligned}$$

$$\sum_{i=1}^m p_{i/j} = \sum_{i=1}^m \frac{p_{ij}}{p_{\cdot j}} = \frac{1}{p_{\cdot j}} \underbrace{\sum_i p_{ij}}_{p_{\cdot j}} = 1$$

„$\overset{\xi}{=}$“ $\Leftrightarrow$ $p_{i/j} \in \{0, 1\}$

(b) Zur Erinnerung: $\ln z \leq z - 1 \quad \forall z > 0$

$$\begin{aligned} &H(\xi/\eta) - H(\xi) \\ &= -\sum_{i,j} p_{ij} \log p_{i/j} + \sum_i p_{i\cdot} \log p_{i\cdot} \\ &= \sum_{i,j} p_{ij} \log \frac{1}{p_{i/j}} + \sum_{i,j} p_{ij} \log p_{i\cdot} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i,j} p_{ij} \log \frac{p_{i\cdot}}{p_{i/j}} = \log e \sum_{i,j} p_{ij} \ln \frac{p_{i\cdot}}{p_{i/j}} \\
&\leq \sum_{i,j} p_{ij} \left(\frac{p_{i\cdot}}{p_{i/j}} - 1 \right) \\
&= \log e \sum_{i,j} p_{ij} \left(\frac{p_{i\cdot} p_{\cdot j}}{p_{ij}} - 1 \right) \\
&= \log e \sum_{i,j} p_{i\cdot} p_{\cdot j} - 1 \\
&= 0 \qquad \boxed{\sum_{i,j} p_{ij} = 1} \\
&\quad \sum p_{i\cdot} = \sum p_{\cdot k} = 1
\end{aligned}$$

$$\begin{aligned}
H(\xi/\eta) &\leq H(\xi) \\
\text{„=“} &\iff \forall i,j : \frac{p_{i\cdot}}{p_{i/j}} = 1 \\
& \quad \text{wegen } \ln z = z - 1 \\
&\iff z = 1 \\
\frac{p_{i\cdot}}{p_{i/j}} &= \frac{p_{i\cdot} p_{\cdot j}}{p_{ij}} = 1 \\
&\iff p_{ij} = p_{i\cdot} p_{\cdot j} \\
&\iff \xi, \eta \text{ unabhängig}
\end{aligned}$$

Bemerkung: Für $\xi_1, \ldots, \xi_n$ gilt:

$$H(\xi_1) \geq H(\xi_1/\xi_2) \geq H(\xi_1/\xi_2, \xi_3) \geq \ldots \geq H(\xi_1/\xi_2, \ldots, \xi_n)$$

4. Beziehungen zwischen $H(\xi, \eta)$ und $H(\xi)$ sowie $H(\eta)$

   (a) $H(\xi, \eta) \leq H(\xi) + H(\eta)$

   Beweis:

$$\begin{aligned}
H(\xi, \eta) &= H(\xi) + H(\xi/\eta) = H(\eta) + H(\xi/\eta) \\
&\leq H(\xi) + H(\eta) \\
\text{„=“} &\iff \xi \text{ und } \eta \text{ unabhängig}
\end{aligned}$$

   (b) $H(\xi, \eta) \geq \max(H(\xi), H(\eta))$

   (c) $H(\xi, \eta) = H(\xi) + H(\eta) \iff \xi$ und η unabhängig

2.2 Beispiel
Russisches Alphabet hat $m = 32$ Zeichen (inkl. ␣)

$\Rightarrow H(\xi) = 4,35$ bit $\leq \log_2 32 = 5$

Englisches Alphabet ($m = 27$)

$\Rightarrow H(\xi) = 4,03$ bit $\leq \log_2 27 = 4,75$

Wir betrachten noch einmal das englische Alphabet. Wie ist die Entropie, wenn die Vorgänger (η_i) eines Zeichens (ξ) betrachtet werden.

$$\begin{aligned} H(\xi/\eta) &= 3,3 \text{ bit} \\ H(\xi/\eta_1,\eta_2) &= 3,1 \text{ bit} \\ &\vdots \\ H(\xi/\eta_1,\ldots,\eta_7) &\approx 2 \text{ bit} \end{aligned}$$

2.1.2 Der Eindeutigkeitssatz von Chinchin

Definieren

$$\begin{aligned} \mathbb{P}_m &= \left\{(p_1,\ldots,p_m) : p_i \geq 0 \quad \forall i \wedge \sum_{i=1}^{m} p_i = 1\right\} \\ & 1 \leq m \leq \infty \\ \mathbb{P} &= \bigcup_{t=1}^{\infty} \mathbb{P}_t \\ H &: \mathbb{P} \to \mathbb{R} \\ & (p_1,\ldots,p_m) \longrightarrow H(p_1,\ldots,p_m) \quad \forall m \geq 1 \end{aligned}$$

Wir fordern von H die folgenden Eigenschaften:

A1: $$H(p_1,\ldots,p_m) \leq H\left(\frac{1}{m},\ldots,\frac{1}{m}\right)$$

A2: $$(p_{11},p_{12},\ldots,p_{mn}) \in P_{mn}$$

$$\begin{aligned} & H(p_{11},p_{12},\ldots,p_{mn}) \\ &\quad = H(p_{1\cdot},\ldots,p_{m\cdot}) + \sum_{i=1}^{m} p_{i\cdot} H(p_{1/i},\ldots,p_{n/i}) \\ &\qquad p_{i\cdot} = \sum_{j=1}^{n} p_{ij} \qquad i = 1,\ldots,m \\ &\qquad p_{j/i} = \frac{p_{ij}}{p_{i\cdot}} \end{aligned}$$

A3: $$H(p_1,\ldots,p_i,0,p_{i+1},\ldots,p_m) = H(p_1,\ldots,p_i,p_{i+1},\ldots,p_m)$$

$$\forall (p_1,\ldots,p_m) \in \mathbb{P}_m \quad (m \geq 1)$$

2.1 Satz (Eindeutigkeitssatz)
Haben eine „Entropie“ $H : \mathbb{P} \to \mathbb{R}; H \not\equiv 0$. Wenn H den Bedingungen *A1,A2,A3* genügt und überdies auf allen $\mathbb{P}_n$ stetig ist, so gilt:

$$H(p_1,\ldots,p_m) = -c \cdot \sum_{i=1}^{m} p_i \log p_i \qquad \forall m, \quad c > 0 \text{ konstant}$$

2.2 Optimale Kodierung

Haben eine Funktion

$$\xi : \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,3 & 0,2 & 0,2 & 0,15 & 0,15 \end{pmatrix}$$

Problem: Mit möglichst wenigen ja-nein-Fragen herausfinden, welches x_i realisiert wurde.

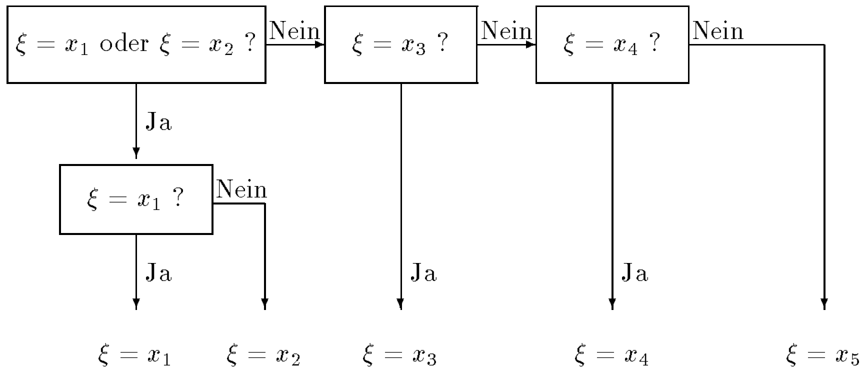

Abbildung 2.2: Mögliche Fragestrategie

Wenn wir „ja“ mit 1 bezeichnen und „nein“ mit 0, dann ergibt sich folgender Entscheidungsweg:

$$\begin{aligned} (11) &\longleftrightarrow \xi = x_1 \\ (10) &\longleftrightarrow \xi = x_2 \\ (01) &\longleftrightarrow \xi = x_3 \\ (001) &\longleftrightarrow \xi = x_4 \\ (000) &\longleftrightarrow \xi = x_5 \end{aligned}$$

Wir bezeichnen die Länge der Rechnungen (bzw. die Anzahl der Fragen) mit N,

$$N : \{x_1, \ldots, x_m\} \longrightarrow \Gamma^+ \qquad \Gamma^+ = \{g \in \Gamma : g > 0\}$$

2.3 Beispiel
Wir schauen uns die Entropie und den Erwartungswert an.

$$\begin{aligned} N(x_1) &= N(x_2) = N(x_3) = 2 \\ N(x_4) = N(x_5) = 3 & \\ & N : \begin{pmatrix} 2 & 3 \\ 0,7 & 0,3 \end{pmatrix} \\ EN(\xi) &= 2 \cdot 0,7 + 3 \cdot 0,3 = 2,3 \\ \log x &= \log_2 x \\ H(\xi) &= -0,3 \cdot \log_2 0,3 - 2 \cdot 0,2 \cdot \log_2 0,2 - 2 \cdot 0,15 \cdot \log_2 0,15 \\ &= 2,271 \\ & \boxed{EN(\xi) \geq H(\xi)?} \end{aligned}$$

Ziel: Konstruktion einer Kodierung aller Buchstaben, die „möglichst kurz“ ist.

2.2 Definition (Quelle)
$\mathbb{X} = \{x_1, \ldots, x_m\}$ ist ein Alphabet. Wenn $\{\xi_n\}_{n\in\Gamma^+}$ eine Folge unabhängiger und identisch verteilter zufälliger Variablen mit Wert in $\mathbb{X}$ ist, $P(\xi_n = x_j) = p_j \forall n$, so heißt $\{\xi_n\}_{n\in\Gamma^+}$ *diskrete gedächtnislose Quelle* mit dem Quellalphabet $\mathbb{X}$.

$$\xi_n : \begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ p_1 & p_2 & \ldots & p_m \end{pmatrix}$$

Ziel: Optimale Kodierung des Quellalphabets $\mathbb{X}$ durch ein Kodealphabet $\mathbb{Y} = \{y_1, \ldots, y_d\}, d \in \Gamma^+$, z.B. $\mathbb{Y} = \{0, 1\}$.

2.3 Definition (Kodierung)
Eine *Kodierung* g von $\mathbb{X}$ durch $\mathbb{Y}$ ist eine umkehrbar eindeutige Abbildung von $\mathbb{X}$ in $\bigcup_{l=1}^{\infty} \mathbb{Y}^l$, wobei

$$\mathbb{Y}^l = \mathop{\times}_{i=1}^{l} \mathbb{Y}$$

$$g : \mathbb{X} \longrightarrow \bigcup_{l=1}^{\infty} \mathbb{Y}^l$$

$$x_j \in \mathbb{X} \xrightarrow{g} g(x_j) = (y_{j_1}, \ldots, y_{j_{n_j}})$$

$g(x_j)$ heißt *Kodewort* der Buchstaben $x_j \in \mathbb{X}$. Die Kodewortmenge von g ist:

$$K = \{g(x_1), g(x_2), \ldots, g(x_m)\}$$

die Länge des Kodewortes $g(x_j)$ bezeichnen wir mit n_j.

Mittlere Kodewortlänge

$$EN_g = \sum_{j=1}^{m} p_j n_j$$

$$N_g : \begin{pmatrix} n_1 & \ldots & n_m \\ p_1 & \ldots & p_m \end{pmatrix} \text{ – Kodewortlänge des Kodes } g$$

Unsere Forderungen

1. $\bar{n}_g := EN_g$ möglichst klein machen
2. Eindeutige Dekodierbarkeit!

2.4 Definition (eindeutige Dekodierung)
Eine Kodierung g heißt *eindeutig dekodierbar (e.d.)*, wenn die Abbildung G mit:

$$G : \bigcup_{l=1}^{\infty} \mathbb{X}^l \longrightarrow \bigcup_{l=1}^{\infty} \mathbb{Y}^l$$

injektiv (umkehrbar eindeutig) ist.

$$(u_1, \ldots, u_r) \in \mathbb{X}^r \xleftrightarrow{G} (g(u_1), \ldots, g(u_r))$$

2.5 Definition (Präfix)
$a = (a_1, \ldots, a_r)$
$b = (b_1, \ldots, b_s) \qquad r \leq s, a, b \in K$
a heißt *Präfix* von b, wenn ein

$$(c_1, \ldots, c_{s-r}) \in \bigcup_{l=1}^{\infty} \mathbb{Y}^l$$

existiert, so daß $b = (a_1, \ldots, a_r, c_1, \ldots, c_{s-r})$.

g heißt *präfixfreier Kode* (PF-Kode), falls kein Kodewort Präfix eines *anderen* Kodewortes ist.

2.2 Satz
Jede PF-Kodierung ist e.d. Kodierung.

Gedanken zum Beweis: Wenn wir einen Text von Kodebuchstaben haben, kann für jedes Stück (beginnend vom Textanfang) entschieden werden, ob es ein Kodebuchstabe ist oder nicht. Wenn wir einen solchen Kodebuchstaben gefunden haben (eindeutig aufgrund der Präfixfreiheit), wird er vom Text abgetrennt und der Vorgang wiederholt.

Beispiel:

$$\underbrace{010111}_{\text{Kodeb. 1}}\ \underbrace{1}_{\text{Kodeb. 2}}\ \underbrace{00}_{\text{Kodeb. 3}}$$

Haben nun $\mathbb{X} = \{x_1, \ldots, x_n\}; \mathbb{Y} = \{y_1, \ldots, y_d\}$, wobei $n_1, \ldots, n_m$ gegeben seien. Wann existiert ein PF-Kode?

2.3 Satz (Kraft'sche Ungleichung)
Für jede e.d. Kodierung mit Kodewortlängen $n_1, \ldots, n_m$ gilt:

$$\sum_{j=1}^{m} d^{-n_j} \quad \leq \quad 1 \tag{2.1}$$

Beweis: g ist eine e.d. Kodierung mit Wortlängen $n_1, \ldots, n_m$. Bilden nun die Mengen β_e mit

$$\beta_l = \#\{j : n_j = l\}, \quad \forall l \in \Gamma^+$$

$$\begin{aligned}
r &= \max_{1 \leq j \leq m} n_j \\
k &\in \Gamma^+ \text{ beliebig} \\
\left(\sum_{j=1}^{k} d^{-n_j}\right)^k &= \left(\sum_{l=1}^{r} \beta_l d^{-l}\right)^k = \sum_{l=k}^{k \cdot r} \gamma_l d^{-l} \\
& \gamma_l = \sum_i \beta_{i_1} \cdot \beta_{i_2} \cdot \ldots \cdot \beta_{i_k} \\
& i_1, \ldots, i_k \in \{1, \ldots, r\} \\
& i_1 + i_2 + \ldots + i_k = l \\
& l = k, k+1, \ldots, k * r
\end{aligned}$$

γ_l ist die Anzahl der Quellwörter aus $\bigcup_{s=1}^{\infty} \mathbb{X}^s$, deren Kodewort die Länge l hat und sich aus k einzelnen Kodewörtern zusammensetzt.

g e.d. Kode

$$\Rightarrow \text{ Kodewort } \overset{g}{\longleftrightarrow} \text{ Quellwort}$$

d^l ist maximale Anzahl der möglichen Kodewörter der Länge l.

$$\Rightarrow \gamma_l \leq d^l$$

$$\left(\sum_{j=1}^{m} d^{-n_j}\right)^k \leq \sum_{l=k}^{k\cdot r} \gamma_l d^{-l} \leq \sum_{l=k}^{k\cdot r} \left(\underbrace{d^l d^{-l}}_{1}\right) \leq k\cdot r \quad \forall k \in \Gamma^+$$

$$\sum_{j=1}^{m} d^{-n_j} < (k\cdot r)^{\frac{1}{k}} \overset{k\to\infty}{\longrightarrow} 1$$

Umgekehrt gelte die Behauptung (2.1)

o.B.d.A.: $n_1 \leq n_2 \leq \ldots \leq n_m$

$$\begin{array}{rcl}
 & & a \in \mathbb{Y}^l, l \leq n_m \\
N(a) & = & \{b \in \mathbb{Y}^{n_m} : a \text{ ist Präfix von } b\} \\
 & & g_1 \in \mathbb{Y}^{n_1} \text{ beliebig} \\
 & \Rightarrow & \#N(g_1) = d^{n_m - n_1} \\
\text{aus 2.1} & \Rightarrow & \sum_{j=1}^{m} d^{-n_j} \leq 1 \\
 & & \sum_{j=1}^{m} d^{n_m - n_j} \leq d^{n_m} \\
 & \Rightarrow & d^{n_m - n_1} < d^{n_m} \quad ; m \geq 2 \\
\mathbb{Y}^{n_m} \setminus N(g_1) & \neq & \emptyset \\
 & \Rightarrow & \exists g_2 \in \mathbb{Y}^{n_2} : g_1 \text{ ist nicht Präfix von } g_2 \\
 & \Rightarrow & \#N(g_2) = d^{n_m - n_2} \quad ; m \geq 3 \\
\text{aus 2.1} & \Rightarrow & d^{n_m - n_1} + d^{n_m - n_2} < d_{n_m} \\
 & \Rightarrow & \mathbb{Y}^{n_m} \setminus (N(g_1) \cup N(g_2)) \neq \emptyset \\
 & \Rightarrow & \exists g_3 \in \mathbb{Y}^{n_3} : g_3 \text{ hat weder } g_1 \text{ noch } g_2 \text{ als Präfix} \\
 & \vdots & \\
ENg & = & \bar{n}_g = \sum_{j=1}^{n} p_j n_j
\end{array}$$

2.4 Satz (von Shannon)
$\{\xi_n\}_{n\in\Gamma^+}$ ist eine diskrete, gedächtnislose Quelle mit Alphabet $\mathbb{X} = \{x_1, \ldots, x_n\}$ und der Entropie $H(\xi)$. $\mathbb{Y} = \{y_1, \ldots, y_d\}$, wobei $d \geq 2$.

1. g ist ein e.d. Kode mit Wortlängen $n_1, \ldots, n_m$:

   $$\frac{H(\xi)}{\log d} \leq \bar{n}_g$$

2. $\exists$ PF-Kode g mit Wortlängen $n_1, \ldots, n_m$ derart, daß

   $$\bar{n}_g < \frac{H(\xi)}{\log d} + 1$$

Beweis:

$$\begin{aligned} P(\xi = x_j) &= p_j \\ P(\xi_n = y_j) &= P(\xi = x_j) = p_j \end{aligned}$$

1. $H(\xi) - \bar{n}_g \log d$

$$\begin{aligned} &= -\sum_{j=1}^{m} p_j \log p_j - \log d \sum_{j=1}^{m} p_j n_j \\ &= \sum_{j=1}^{m} p_j \log \underbrace{\frac{d^{-n_j}}{p_j}}_{t>0}, \text{ da } \ln z \leq z - 1 \\ &\leq \log e \sum_{j=1}^{m} p_j \left(\frac{d^{-n_j}}{p_j} - 1 \right) \\ &= \log e \left(\sum_{j=1}^{m} d^{-n_j} - 1 \right) \leq 0 \end{aligned}$$

2. $d^{-n_j} \neq p_j$

$$\frac{d^{-n_j}}{p_j} \leq 1$$

Wählen $n_1, \ldots, n_m$ mit:

$d^{-n_j} \leq p_j < d^{-n_j + 1}$

$\Rightarrow \sum_{j=1}^{m} d^{-n_j} \leq \sum_{j=1}^{m} p_j = 1$

$\Rightarrow \exists$ PF-Kode mit den Kodewortlängen $n_1, \ldots, n_m$

$$\begin{aligned} -n_j \log d &\leq \log p_j < (-n_j + 1) \log d \\ -n_j &\leq \frac{\log p_j}{\log d} < -n_j + 1 \\ -n_j - 1 &< -\frac{\log p_j}{\log d} \leq n_j \\ & \boxed{n_j < -\frac{\log p_j}{\log d} + 1} \leq n_j + 1 \\ \bar{n}_g &= \sum_{j=1}^{m} p_j n_j \\ &< -\sum_{j=1}^{m} p_j \frac{\log p_j}{\log d} + 1 \\ &= -\frac{1}{\log d} \sum_{j=1}^{m} p_j \log p_j + 1 \\ &= \frac{H(\xi)}{\log d} + 1 \end{aligned}$$

Zusammengefaßt gilt:

$$\boxed{\frac{H(\xi)}{\log d} \leq \bar{n}_g < \frac{H(\xi)}{\log d} + 1}$$

Wahl von n_j **(**$j = 1, \ldots, m$**)**

$$n_j = \left\lceil \frac{-\log p_j}{\log d} \right\rceil$$

$$\lceil x \rceil = \min\{g \in \Gamma : g \geq x\}$$

$$\bar{n}_g = \frac{H(\xi)}{\log d} \Longleftrightarrow p_j = d^{-n_j}$$

Nur möglich, wenn p_j rational!

2.2.1 Blockkodierung

Haben ein Quellalphabet $\mathbb{X} = \{x_1, \ldots, x_m\}$. Wir betrachten die Menge $\mathbb{X}^L$, für die gilt:

$$\mathbb{X}^L = \{u = (u_1, \ldots, u_L) : u_i \in \mathbb{X}\}$$

Für die Wahrscheinlichkeit gilt:

$$p_{1 \ldots L} = P(\xi_1 = u_1, \ldots, \xi_L = u_L)$$

Wenn $\{\xi_n\}_{n \in \Gamma^+}$ eine gedächtnislose Quelle auf $\mathbb{X}$ ist, dann ist $\xi^{(L)} = (\xi_1, \ldots, \xi_L)$ eine diskrete, gedächtnislose Quelle auf $\mathbb{X}^L$. Für die Entropie gilt hierbei:

$$\boxed{H(\xi^{(L)}) = L \cdot H(\xi)}$$

$$H(\xi_i) = H(\xi) \qquad i = 1, 2, \ldots, L$$

$$H(\xi_1, \xi_2) = H(\xi_1) + H(\xi_2) = 2 \cdot H(\xi)$$

Wenn g eine Abbildung

$$\mathbb{X} = \{x_1, \ldots, x_m\} \xrightarrow{g} \bigcup_{l=1}^{\infty} \mathbb{Y}^l$$

ist, heißt $g^{(L)}$ *Blockkodierung* von $\mathbb{X}^{(L)}$, wobei gilt:

$$\bar{n}^{(L)} = \sum_{(u_1, \ldots, u_L) \in \mathbb{X}^{(L)}} P(\xi_1 = u_1, \ldots, \xi_L = u_L) \cdot n(u_1, \ldots, u_L)$$

und $n(u_1, \ldots, u_L)$ ist die Kodewortlänge des Blocks $u = (u_1, \ldots, u_L)$.

2.5 Korrolar (zum Satz von Shannon)
$\{\xi_n\}_{n \in \Gamma^+}$ ist eine gedächtnislose, diskrete Quelle über $\mathbb{X}$.

$$L \in \Gamma^+, \quad \mathbb{Y} = \{y_1, \ldots, y_d\}$$

Für die Kodierung $g^{(L)}$ von Wörtern der Länge L gilt:

1. $$\frac{H(\xi^{(L)})}{\log d} \leq \bar{n}^{(L)}$$

   d.h.

   $$\frac{L \cdot H(\xi)}{\log d} \leq \bar{n}^{(L)}$$

2. $$\exists \text{ PF-Kode } g^{(L)} : \bar{n}^{(L)} < \frac{L \cdot H(\xi)}{\log d} + 1$$

Zusammengefaßt:

$$\boxed{\frac{H(\xi)}{\log d} \leq \frac{\bar{n}^{(L)}}{L} < \frac{H(\xi)}{\log d} + \frac{1}{L}}$$

Ein e.d. Kode g, der die untere Grenze in der SHANNON-Ungleichung erreicht, heißt *absolut optimal.*

$$p_j \quad (i = 1, \ldots, m) \text{ irrational } \Rightarrow \exists \text{ kein absolut optimaler Kode}$$

2.6 Definition (optimaler Kode)
$\{\xi_n\}_{n\in\Gamma^+}$ ist eine gedächtnislose, diskrete Quelle. Ein e.d. Kode g^* mit Kodewortlängen $n_1^*, \ldots, n_m^*$ heißt *optimal*, wenn

$$\bar{n}_{g^*} = \sum_{j=1}^{n} n_j^* p_j \leq \sum_{j=1}^{m} n_j p_j = \bar{n}_g$$

für alle e.d. Kodes g mit Kodewortlängen von $n_1, \ldots, n_m$ gilt.

2.6 Lemma
Es existiert stets ein optimaler PF-Kode.

Beweis: Für jeden PF-Kode gilt:

$$\sum_{j=1}^{m} d^{-n_j} \leq 1$$

Es ist zu zeigen:

$$\inf_{n_1,\ldots,n_m} \left\{ \sum_{j=1}^{m} n_j p_j : (n_1, \ldots, n_m) \in Z \right\}$$

existiert, wobei

$$Z = \left\{ n \in \Gamma_+^m : \sum_{j=1}^{m} d^{-n_j} \leq 1 \right\}$$

$$n \in \Gamma_+^m \Rightarrow n = (n_1, \ldots, n_m), \quad n_j \in \Gamma^+$$

Wir definieren eine Menge M:

$$M = \{n \in Z : \neg\exists n' \in Z : n' \neq n, n' \leq n\}^2$$

Wir wollen zeigen: M ist nichtleer und endlich

[2] d.h.: $n_i' \leq n_i \forall i, \quad \exists i_0 : n_{i_0}' < n_{i_0}$

Annahme: $M = \emptyset$

$\Rightarrow \forall n \in Z \quad \exists n' \in Z : n' \neq n \wedge n' \leq n$

$\Rightarrow$ zu n' gibt es ein $n'' \in Z : n'' \neq n', n'' \leq n'$

$\Rightarrow M$ ist unendlich

Für jedes $n \in Z$:

$$card\{n' : n' \neq n, n' \leq n\} < \infty, \qquad (1 \leq n_j)$$

$\Rightarrow$ Widerspruch, $M \neq \emptyset$
Es sei $\{n^{(k)}\}$ eine Folge unendlich vieler verschiedener Elemente aus M:

$$n^{(k)} = (n_{k_1}, \ldots, n_{k_m})$$

zu jedem k gibt es eine Komponente $l(k)$ mit:

$$n_{k+1,l(k)} < n_{k,l(k)}$$

$\Rightarrow$ (Folge enthält unendlich viele Elemente)

$\exists$ eine Komponente $l_0 : n_{k+1,l_0)} < n_{k,l_0}$ für unendlich viele k

Widerspruch, d.h. M ist endlich, da $n_j \geq 1$

$$\Rightarrow \inf_{n_1,\ldots,n_m} \left\{ \sum_{j=1}^{m} p_j n_j : (n_1, \ldots, n_m) \in Z \right\} = \inf_{n_1,\ldots,n_m} \left\{ \sum_{j=1}^{m} p_j n_j : (n_1, \ldots, n_m) \in M \right\}$$

$\forall (n_1, \ldots, n_m) \in Z \setminus M \quad \exists (n'_1, \ldots, n'_m) \in M$ mit:

$$\sum_{i=1}^{m} n'_i p_i \leq \sum_{i=1}^{m} n_i p_i$$

$$\Rightarrow \text{ gesucht: } \min_{n_1,\ldots,n_m} \left\{ \sum_{j=1}^{m} n_j p_j : (n_1, \ldots, n_m) \in M \right\}$$

$$\exists (n_1^*, \ldots, n_m^*) \in M : \bar{n}_{g^*} = \sum_{j=1}^{m} n_j^* p_j \leq \sum_{j=1}^{m} n_j p_j, \quad \forall (n_1, \ldots, n_m) \in Z$$

$\bar{n}_{g*}$ ist die reale Entropie der Quelle ξ.

2.4 Beispiel
Es seien $m = 3$ und $d = 2$, d.h. $\mathbb{X} = \{x_1, x_2, x_3\}$ und $\mathbb{Y} = \{0, 1\}$. Wir wollen ein Alphabet bestehend aus 3 Buchstaben kodieren, wobei das Zielalphabet 2 Buchstaben hat. Die Werte n_i sagen uns hierbei, wie viele Buchstaben des Kodealphabets hierbei verwendet werden, um den i-ten Buchstaben des Quellalphabets zu kodieren.

x_1	x_2	x_3	
n_1	n_2	n_3	
1	1	1	$d^{-1} + d^{-1} + d^{-1} = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2} > 1$
2	1	1	$d^{-2} + d^{-1} + d^{-1} = \frac{5}{4} > 1$
2	2	1	$\sum = 1$
2	1	2	$\sum = 1$
1	2	2	$\sum = 1$

Nur bei den letzten 3 Zeilen ist die Anzahl der n_i ausreichend für eine Kodierung. Die Menge M der möglichen n_i ist nachfolgend angegeben:

$$M = \{(2, 2, 1), (2, 1, 2), (1, 2, 2)\}$$

Für die optimale Kodierung muß dasjenige Element aus M bestimmt werden, für das die $\sum_j n_j p_j$ minimal wird.

2.2.2 Konstruktion optimaler PF-Kodes

Wir gehen von folgenden Prämissen aus:

- $\mathbb{X} = \{x_1, \dots, x_m\}$ ist das Quellalphabet
- $\mathbb{Y} = \{0, 1\}$ ist das Kodealphabet
- $\xi : \begin{pmatrix} x_1 & \dots & x_m \\ p_1 & \dots & p_m \end{pmatrix}$ ist die Quelle
- $p_1 \geq p_2 \geq \dots \geq p_m, \qquad p_i > 0$

2.7 Lemma (Huffman-Kodierung)

$$\xi' : \begin{pmatrix} x_1' & \dots & x_{m-2}' & x_{m-1}' \\ p_1 & \dots & p_{m-2} & p_{m-1}' \end{pmatrix}$$

$$p_{m-1}' = p_{m-1} + p_m$$

Wenn g' optimaler Kode für ξ':

$$\begin{aligned} \Rightarrow \quad & \text{Funktion } g \text{, definiert wie nachfolgend} \\ & g(x_j) = g'(x_j') \qquad j = 1, \dots, m-2 \\ & g(x_{m-1}) = g'(x_{m-1}', 0) \\ & g(x_m) = g'(x_{m-1}', 1) \end{aligned}$$

ist optimaler Kode für ξ.

2.5 Beispiel (Huffman 1)

x_j	p_j	$g^*(x_i)$
x_1	$0,3$	00
x_2	$0,2$	10
x_3	$0,2$	11
x_4	$0,15$	010
x_5	$0,15$	011

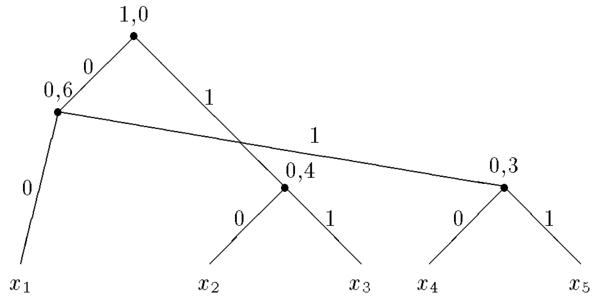

Abbildung 2.3: Beispiel 1 für die Huffman-Kodierung

2.6 Beispiel (Huffman 2)

x_j	p_j	$g^*(x_i)$
x_1	$0,3$	00
x_2	$0,25$	01
x_3	$0,25$	10
x_4	$0,1$	110
x_5	$0,1$	111

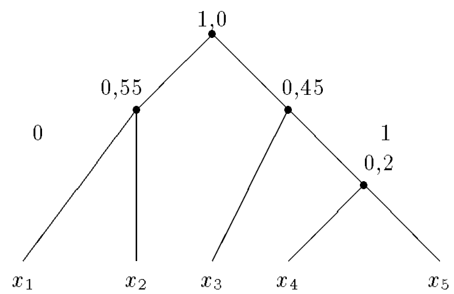

Abbildung 2.4: Beispiel 2 für die Huffman-Kodierung

2.2.3 Der Shannon-Fano-Algorithmus

$$\mathbb{Y} = \{0,1\} \qquad p_1 \geq p_2 \geq \ldots \geq p_m > 0$$

1. Gesucht wird ein Index j_1 mit:

$$\sum_{i=1}^{j_1} p_i \approx \frac{1}{2}$$

$$1 \leq i \leq j_1 : x_i \to 0$$

$$m \geq i > j_1 : x_i \to 1$$

2. Gesucht werden Indizes j_{11}, j_{22} mit:

$$\sum_{i=1}^{j_{11}} p_i \approx \frac{1}{2} \sum_{i=1}^{j_1} p_i$$

$$\sum_{i=j_1+1}^{j_{22}} p_i \approx \frac{1}{2} \sum_{i=j_1+1}^{m} p_i$$

$$1 \leq i \leq j_{11} : x_i \to 00$$

$$j_{11} < i \leq j_1 : x_i \to 01$$

$$j_1 < i \leq j_{22} : x_i \to 10$$

$$m \geq i > j_{22} : x_i \to 11$$

3. usw.

2.7 Beispiel (Shannon-Fano)

x_j	p_j			
x_1	$0,3$	0	0	
x_2	$0,2$	0	1	
x_3	$0,2$	1	0	
x_4	$0,15$	1	1	0
x_5	$0,15$	1	1	1

Nachteil: Es kommt nicht immer ein optimaler Kode heraus.

2.3 Binäre Suchbäume

2.7 Definition (binärer Suchbaum)

$S = \{x_1, x_2, \ldots, x_m\}, \qquad m \geq 2$

S sei bezüglich $\prec$ linear geordnet, d.h.

$$x_1 \prec x_2 \prec \ldots \prec x_m$$

Ein binärer Baum heißt *binärer Suchbaum* über S, wenn für alle Knoten $x_j \in S$ gilt:

$$x_i \prec x_j \prec x_k$$

wobei x_k ein beliebiger Knoten im rechten Teilbaum von x_j und x_i ein beliebiger Knoten im linken Teilbaum von x_j ist.

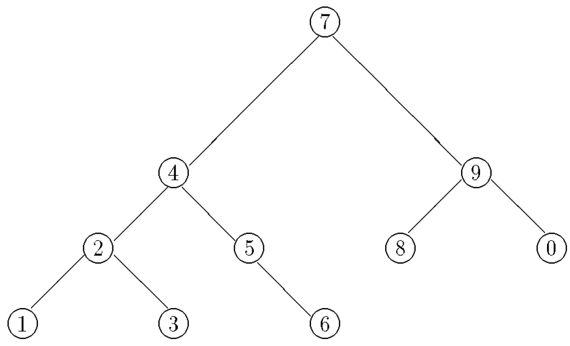

Abbildung 2.5: Beispiel für einen binären Suchbaum

2.8 Beispiel

$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$

Knoten x_i habe die Zugriffswahrscheinlichkeit p_i

t_i Anzahl der Vorgängerknoten von x_i

$t_i + 1$ Anzahl der nötigen Vergleiche, um x_i zu finden

$Z_T(x_i) = t_i + 1$ – Zugriffszeit

$$Z_T : S \longrightarrow \Gamma^+$$

$$Z_T : \begin{pmatrix} t_1 + 1 & t_2 + 1 & \dots & t_m + 1 \\ p_1 & p_2 & \dots & p_m \end{pmatrix}$$

Mittlere Zugriffszeit:

$$EZ_T = \sum_{i=1}^{m} p_i(t_i + 1)$$

Ziel: Baum so konstruieren, so daß EZ_T minimal wird!

2.8 Lemma

T sei ein binärer Suchbaum über $S = \{x_1, \dots, x_m\}$. t_i sei die Tiefe (Anzahl der Vorgänger) des Knotens x_i.

Dann gilt für alle $c : 0 \leq c \leq 1$:

$$c \cdot \sum_{i=1}^{m} \left(\frac{1-c}{2}\right)^{t_i} \leq 1$$

Beweis: Wir definieren $m(k)$ – die Anzahl der Knoten, deren Tiefe k ist

$$m(k) := \#\{x_i : t_i = k\} \leq 2^k$$

$$\begin{aligned}
& c\sum_{i=1}^{m}\left(\frac{1-c}{2}\right)^{t_i} \\
= \quad & c\sum_{k=0}^{m} m(k)\left(\frac{1-c}{2}\right)^{k} \\
\leq \quad & c\sum_{k=0}^{m}(1-c)^k = 1-(1-c)^{m+1} \leq 1 \\
& \text{da } \sum_{i=0}^{m} q^i = \frac{q^{m+1}-1}{q-1}
\end{aligned}$$

$S = \{x_1, \ldots, x_m\}$ sei die Knotenmenge eines binären Suchbaumes, $(p_1, \ldots, p_m)$ sei die Zugriffsverteilung über S.

$$\sum_{i=1}^{m} p_i = 1; \quad p_i \geq 0$$

2.9 Satz (Entropie der Zugriffsverteilung)
Es sei $H = H(p_1, \ldots, p_m)$ die Entropie der Zugriffsverteilung der Knoten eines binären Suchbaumes T über S. Dann gilt:

$$\max_{y\in\mathbb{R}}\left\{\frac{H-y}{\log_b(2+b^{-y})}\right\} \leq EZ_T$$

wobei, $b > 1$ die Basis des Logarithmus ist.

Beweis:

$$EZ_T = \sum_{i=1}^{m} p_i(t_i+1), \quad p_i \text{ Zugriffswahrscheinlichkeit}$$

$$\begin{aligned}
& 0 \leq c < 1 \\
& c' := \frac{1-c}{2} \\
& q_i = c \cdot {c'}^{t_i} \qquad t_i \text{ – Tiefen des Suchbaumes } T
\end{aligned}$$

$$\begin{aligned}
& H(p_1, \ldots, p_m) + \sum_{i=1}^{m} p_i \log q_i \\
= \quad & -\sum_{i=1}^{m} p_i \log p_i + \sum_{i=1}^{m} p_i \log q_i \\
= \quad & \sum_{i=1}^{m} p_i \log \frac{q_i}{p_i} \\
= \quad & \log e \sum_{i=1}^{m} p_i \ln \frac{q_i}{p_i} \\
\leq \quad & \log e \sum_{i=1}^{m} p_i\left(\frac{q_i}{p_i}-1\right)
\end{aligned}$$

$$
\begin{aligned}
& \text{da } \sum_{i=1}^{m} p_i = 1 \\
= \quad & \log e \left(\sum_{i=1}^{m} q_i - 1 \right) \leq 0, \text{ d.h. } \sum_{i=1}^{m} \log q_i \leq -H
\end{aligned}
$$

$$
\begin{aligned}
& & & \text{da } q_i = c \cdot {c'}^{t_i} \\
\log q_i & = & & \log c + t_i \log c' \\
t_i & = & & \frac{\log q_i - \log c}{\log c'} \\
EZ_T & = & & \sum_{i=1}^{m} p_i \left(\frac{\log q_i - \log c}{\log c'} + 1 \right) \\
& = & & \underbrace{\frac{1}{\log c'}}_{\leq 0} \underbrace{\sum_{i=1}^{m} p_i \log q_i}_{\leq -H} - \frac{\log c}{\log c'} + 1 \\
& \geq & & -\frac{1}{\log c'} H - \frac{\log c - \log c'}{\log c'} \\
& = & & -\frac{1}{\log c'} \left(H - \log \frac{c'}{c} \right) \\
& & & \text{da } -\log c' = \log \frac{1}{c'} \\
& = & & \frac{H - \log \frac{c'}{c}}{\log \frac{1}{c'}} \\
& & & y = \log \frac{c'}{c} = \log \frac{1-c}{2c} \\
& & & y = y(c) : (0,1) \longrightarrow (-\infty, +\infty) \\
& & & c' = \frac{1-c}{2} \Rightarrow c = 1 - 2c' \\
& & & \frac{1}{c'} = b^{-y} + 2 \\
& & & b^{-y} = \frac{c}{c'} = \frac{1-2c'}{c'} = \frac{1}{c'} - 2 \\
& = & & \underline{\underline{\frac{H - y}{\log(2 + b^{-y})}}} \qquad \forall y \in \mathbb{R}^1
\end{aligned}
$$

Wenn wir als Beispiel $y = 0$ festsetzen, erhalten wir

$$
\boxed{\frac{H}{\log 3} \leq EZ_T}
$$

2.10 Satz (Suchbaum)
Es seien die Knoten $p_1, \ldots, p_m$ gegeben. Es existiert nun ein Suchbaum T^*, für den gilt:

$$
EZ_{T^*} \leq \frac{H}{\log 2} + 1
$$

Beweis: Kann wie im vorherigen Satz geführt werden.

2.3.1 Konstruktion eines binären Suchbaumes

Wir haben eine Menge von Knoten $x_1, \ldots, x_m$ und die zugehörigen Wahrscheinlichkeiten:

$$\begin{array}{ccccccccc} x_1 & \prec & x_2 & \prec & x_3 & \prec & \ldots & \prec & x_m \\ p_1 & & p_2 & & p_3 & & \ldots & & p_m \end{array}$$

Aus dieser linear geordneten Menge von Knoten suchen wir denjenigen Knoten x_k, für den gilt:

$$\sum_{i=1}^{k-1} p_i < \frac{1}{2} \text{ und } \sum_{i=1}^{k} p_i \geq \frac{1}{2}$$

Dieser Knoten x_k wird nun die Wurzel des Baumes T^*.

Nachfolgend die Vorgehensweise für die Teilbäume. Wir haben Knoten $x_l, \ldots, x_r \in S$ mit $1 \leq l \leq r \leq m$. x_k wird dann als Wurzel des Teilbaumes, der $x_l, \ldots, x_r$ enthält, ausgewählt, wenn gilt:

1. $$\sum_{i=l}^{k-1} p_i < \frac{1}{2} \sum_{i=l}^{r} p_i$$

2. $$\sum_{i=l}^{k} p_i \geq \frac{1}{2} \sum_{i=l}^{r} p_i$$

2.9 Beispiel

$S = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$

1	2	3	4	5	6	7	8	9	10
0,2	0,1	0,05	0,3	0,05	0,03	0,08	0,03	0,1	0,06

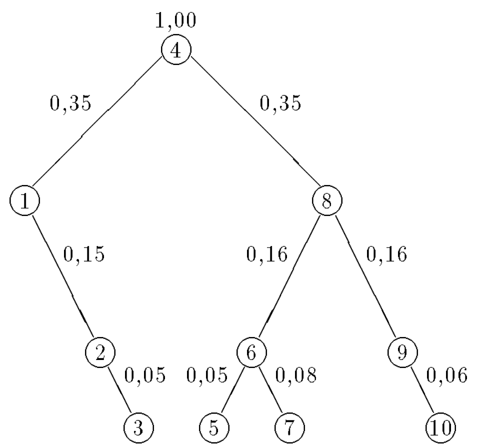

Abbildung 2.6: Beispiel 1 von Suchbäumen

Für diesen Baum gilt:

$$\begin{aligned} EZ_T &= 2,41 \\ H &= 2,0244 \qquad \text{auf Basis des ln} \\ \frac{H}{\ln 3} &= 1,843 \leq 3,921 = \frac{H}{\ln 2} + 1 \end{aligned}$$

2.10 Beispiel

Wir untersuchen nun den Fall, daß alle Knoten die gleiche Zugriffswahrscheinlichkeit haben.

$$S = \{x_1, \ldots, x_m\}$$

$$p_i = \frac{1}{m}$$

Für $b = 2$ gilt:

$$\boxed{\frac{\log_2 m}{\log_2 3} \leq EZ_T \leq \log_2 m + 1}$$

weil $H(\frac{1}{m}, \ldots, \frac{1}{m}) = \log_2 m$.

Beispiel für $m = 10$.

1	2	3	4	5	6	7	8	9	10
0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

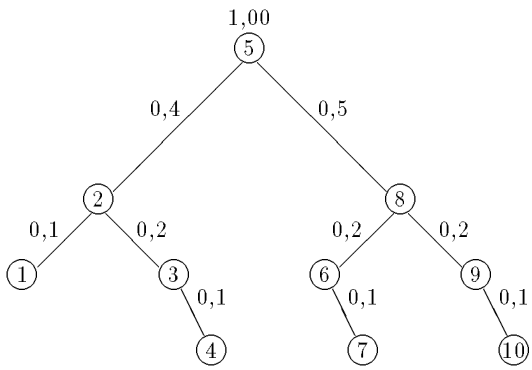

Abbildung 2.7: Beispiel 2 von Suchbäumen

Kapitel 3

Grundlagen der Simulation

Zur Einführung in dieses Thema ein kleines Beispiel.

Wir sollen einen Handlungsreisenden bei der Planung einer Rundreise durch 100 europäische Städte, die möglichst kurz ist, unterstützen. Wir könnten jetzt mit purer Gewalt alle 99! Möglichkeiten durchrechnen, dies ist jedoch zu aufwendig. Eine bessere Möglichkeit ist es, von einer Stadt auszugehen und eine Stadt zu suchen, die möglichst nah ist und in nachfolgenden Durchgängen zu keiner schlechteren Tour führt. Hierzu ist es nötig, manchmal (zufällig) von einem (lokal) optimalen Kurs abzuweichen, um eine eventuell bessere globale Lösung zu finden. Dieses Verfahren heißt "simulated annealing" und leitet sich von den Vorgängen bei der Erkaltung von Metallen ab (die Atome nehmen bei langsamer Abkühlung einen Zustand geringster Entropie ein).

Vorteil Lösung kann mit polynomialen Aufwand erreicht werden

Nachteil dieses Verfahren führt nicht notwendig zu einer optimalen Lösung, erreicht aber mit hoher Wahrscheinlichkeit eine Lösung, die dicht an der optimalen Lösung liegt

Die Grundlage aller Simulationverfahren sind gleichverteilte Zufallszahlen aus $[0, 1)$, unser Ziel besteht nun darin, solche Zufallszahlen zu bestimmen.

$$\xi \sim R(0,1) \qquad f(x) = \begin{cases} 1 & , \text{ falls } 0 \leq x < 1 \\ 0 & , \text{ sonst} \end{cases}$$

3.1 Erzeugung von Zufallszahlen

3.1.1 Erzeugung von Hand

1. Ziffern $0, 1, \ldots, 9$ werden zufällig erzeugt, wobei jede Ziffer die gleiche Wahrscheinlichkeit $p_i = \frac{1}{10}$ hat.

   Als Methoden wären z.B. die Urnenmethode oder die Glücksradmethode anzuwenden.

   Wir erhalten eine Folge von Ziffern, davon entnehmen wir eine vorgegebene Anzahl von Ziffern (z.B. 5) und betrachten diese als Dezimalstellen

   $$\ldots 14337645\underbrace{25243}879023$$

   $$\Rightarrow 0,25243$$

   Die so entstandenen Zahlen heißen *Zufallszahlen.*

2. Statt den Ziffern $0, 1, \ldots, 9$ können wir auch $0, 1$ wählen, erhalten dann eine binäre Zahl $u = (d_1, \ldots, d_n)$, für die gilt:

$$0 \le u \le \sum_{i=1}^{n} d_i 2^{-i} < 1$$

3. *4-Würfel-Prinzip*

   Beschriftung nach folgender Vorschrift:

   1. Würfel beschriften mit: 0–5
   2. Würfel beschriften mit: 0, 6, 12, 18, 24, 30
   3. Würfel beschriften mit: 0, 36, 72, 108, 144, 180
   4. Würfel beschriften mit: 0, 216, 432, 648, 864, 1080

   Bei Wurf erhalten wir ein k mit: $0 \le k \le 1295$

$$u = \frac{k}{1295}$$

   ist nun angenähert gleichverteilt über $[0, 1)$.

3.1.2 Elektronische Erzeugung

In elektronischen Geräten (Bauelementen) fließen auch im Ruhezustand Ströme (weißes Rauschen), deren Spannungen zufällig schwanken. Man zählt innerhalb von Zeitintervallen gleicher Länge, wie oft eine kritische Spannung U_0 überschritten wird.

Im Falle einer geraden Anzahl realisieren wir als Zufallsziffer eine 1, andernfalls eine 0. Aus der resultierenden 0-1-Folge erhalten wir nach obigem Muster eine Zufallszahl.

3.1.3 Mathematische Erzeugung

3.1.3.1 Die multiplikative Kongruenzmethode

Wir geben folgende Startwerte vor:

$$m, a, Z_0 \in \Gamma^+$$

mittels

$$Z_{i+1} = a \cdot Z_i \pmod{m}$$

kann nun eine Folge von Werten berechnet werden, dabei gilt:

$$a \cdot Z_i = (k \cdot m) + Z_{i+1}; \qquad 0 \le Z_{i+1} < m$$

$$u_i = \frac{Z_i}{m}$$

ist eine Folge von *Pseudozufallszahlen*.

Frage: Ist $u_1, \ldots$ angenähert eine Folge unabhängiger, auf $[0, 1)$ gleichverteilter Zufallszahlen?

3.1.3.2 Die lineare Kongruenzmethode

Startwerte:

$$m, a, r, Z_0 \in \Gamma^+$$

Berechnung:

$$Z_{i+1} = (a \cdot Z_i + r) \pmod m$$

3.1.3.3 Die mehrfache Kongruenzmethode

Startwerte:

$$m, a_1, \ldots, a_k, r, Z_0, \ldots, Z_{k-1} \in \Gamma^+$$

Berechnung:

$$Z_n = \left(\sum_{l=1}^{k} a_l \cdot Z_{n-l} + r \right) \pmod m$$

3.1.4 Eigenschaften von Pseudozufallszahlen

Der lineare Kongruenzoperator erzeugt höchstens m verschiedene Zufallszahlen

$\Longrightarrow m$ möglichst groß wählen

3.1 Beispiel (multiplikative Kongruenzmethode)
$m = 2^k, k = 4, a = 11, Z_0 = 3$

$$\begin{aligned} Z_1 &= 1 \\ Z_2 &= 11 \\ Z_3 &= 9 \\ Z_4 &= 3 = Z_0 \end{aligned}$$

In diesem Beispiel ist die Periodenlänge statt theoretisch möglichen 16 nur 4.

Es gilt: Wenn $m = 2^k$, $a \bmod 8 \in \{3, 5\}$, Z_0 ungerade und $r = 0$

$\Longrightarrow$ Länge der Periode $= 2^{k-2}$

3.1 Satz (lineare Kongruenzoperator)
Der lineare Kongruenzoperator

$$Z_{i+1} = a \cdot Z_i + r \quad (\bmod\ m)$$

besitzt die volle Periodenlänge m genau dann, wenn

1. $\operatorname{ggt}(r, m) = 1, \quad$ wobei $\operatorname{ggt}(0, m) = m$
2. $a (\bmod\ p) = 1 \quad \forall$ Primfaktoren von $m, \qquad m = \prod_i p_i^{t_i}$

3. $a (\mathrm{mod}\ 4) = 1,$ falls m Vielfaches von 4 ist

3.2 Beispiel (Werte für bestimmte Rechner bzw. Sprachen)

Rechner bzw. Sprache	m	a	r
IBM (RANDU)	2^{31}	65539	0
PRIME (RANDA)	$2^{31}-1$	16807	0
CDC (SIMULA)	2^{59}	5^{11}	0

3.1.5 Gitter von Zufallzahlen

$u_1, \ldots, u_n, \ldots$ sei eine Folge von Pseudozufallszahlen. Wir betrachten nun alle Paare (u_i, u_{i+1}), wenn wir diese Werte graphisch darstellen, entsteht eine (zweidimensionale) Gitterstruktur von Punkten, die je nach Werten der Zufallszahlen auch zu Linien ausarten können.

Diese Darstellung kann auch für k-Tupel definiert werden. Wenn die Periodenlänge k ist, entsteht *ein* Punkt im k-dimensionalen Raum.

3.3 Beispiel

$$\begin{aligned}
Z_{i+1} &= aZ_i \pmod{2^t} \\
5 &= a \pmod 8 \\
1 &= Z_0 \pmod 4 \\
Z_{(k)} &= (Z_l, Z_{l+1}, \ldots, Z_{l+k-1}) \\
u_{(k)} &= \left(\frac{Z_l}{2^t}, \ldots, \frac{Z_{l+k-1}}{2^t}\right) \\
&= (u_l, \ldots, u_{l+k-1}) \\
u_{(k)} &\in \left(\frac{1}{4} b_1 + G\right) \cap [0,1]^k \\
b_1 &= \frac{1}{2^{t-2}} \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{k-1} \end{pmatrix} \\
& \quad b_i = i\text{-te Einheitsvektor},\ i = 2, \ldots, k \\
G &= \left\{ \sum_{i=1}^{k} q_i b_i : q_1, \ldots, q_k \in \Gamma \right\}
\end{aligned}$$

3.2 Tests von Zufallszahlen

Es wird getestet:

1. Gleichverteilung von Pseudezufallszahlen über $[0, 1)$
2. Unabhängigkeit von Pseudozufallszahlen

3.2.1 Test auf Gleichverteilung

Wir haben Pseudozufallszahlen $u_1, \ldots, u_n$.

Annahme:

$$u_1, \ldots, u_n \sim R(0,1), \text{ unabhängig}$$

hierzu muß

$$\frac{\#\{u_i : a \leq u_i \leq b\}}{n} \overset{n\to\infty}{\longrightarrow} b-a \qquad (0 \leq a < b \leq 1)$$

$0 \leq x \leq 1$ sei fest, betrachten dann

$$\frac{\#\{u_i : u_i < x\}}{n} \approx x = F_n(x)$$

Hypothese H:

$$u_i \sim R(0,1)$$

Alternative H_a:

$$u_i \not\sim R(0,1)$$

3.1 Definition (Test)
Ein *Test* ist eine Entscheidungsvorschrift, die über die Akzeptanz genau einer von zwei alternativen Hypothesen entscheidet.

3.4 Beispiel (Qualitätskontrolle)
Betrachten eine stichpunktartige Qualitätskontrolle, bei der sich ein Käufer entscheiden muß, ob er den gesamten Warenbestand nimmt oder nicht.

Es gibt hierbei folgende Hypothesen:

H: Ware ist gut, d.h. Ausschußanteil $p \leq 2\%$.
H_a: Ware ist schlecht, d.h. $p > 2\%$.

Der Kunde führt bei n Produkten eine Kontrolle durch und ermittelt die „Zufallswerte“ $x_1, \ldots, x_n$. Dabei gilt:

$$x_i = \begin{cases} 0 & \text{gut} \\ 1 & \text{schlecht} \end{cases}$$

$$Z = \sum x_i = \text{ Anzahl fehlerhafter Produkte}$$

Es gibt einen kritischen Wert Z_α:

- $z > Z_\alpha \Longrightarrow$ Ablehnung von H
- $z \leq Z_\alpha \Longrightarrow$ Annahme von H

$$P(z > Z_\alpha : H \text{ wahr}) \leq \alpha$$

α ist das Risiko des Käufers, daß er den Erhalt der Ware ablehnt, obwohl diese gut ist. Übliche Werte für α sind: 0,01 oder 0,05

3.2.1.1 Der χ^2-Test

$\xi_1, \ldots, \xi_k$ seien unabhängig, identisch verteilte Zufallszahlen mit $\xi_i \sim N(0,1)$

$$\Longrightarrow \sum_{i=1}^{k} \xi_i^2 \sim \chi_k^2 \text{ – die } \chi\text{-Quadrat-Verteilung mit } k \text{ Freiheitsgraden}$$

Wenn ξ_i beliebige zufällige Variablen, unabhängig und identisch verteilt für $i = 1, \ldots, n$ und A_j zufällige Ereignisse (Intervalle in $\mathbb{R}$).

$$P(\xi_i \in B) = 1, \quad B = \bigcup_{j=1}^{k} A_j, \quad A_j \cap A_i = \emptyset \quad \forall i \neq j$$

$$P(\xi_i \in A_j) = p_j$$

Definieren

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - n \cdot p_j)^2}{np_j} \qquad n_j = \#\{\xi_i : \xi_i \in A_j\}$$

Für große n gilt dann: $\chi^2 \sim \chi_{k-1}^2$
Folgende sind unsere Hypothesen:

H:	$P(\xi_i \in A_j) = p_j \quad j = 1, \ldots, k$
H_A:	$P(\xi_i \in A_j) \neq p_j$

wenn H richtig ist, gilt wegen dem schwachen Gesetz großer Zahlen (siehe Seite 54): $n_j \approx n \cdot p_j$
Man sieht leicht, daß $0 \leq \chi^2$. Wie sieht es mit der oberen Grenze an? Wenn $\chi^2 \leq C_\alpha$ wollen wir Hypothese H annehmen, wenn $\chi^2 > C_\alpha$ lehnen wir diese ab.

$$P(\chi^2 > C_\alpha : H \text{ richtig}) = \alpha$$

ist die Wahrscheinlichkeit (bzw. das Risiko) dafür, das trotz guter Verteilung der Zufallszahlen wir die Hypothese H ablehnen, d.h. die Nicht-Gleichverteilung annehmen.

3.2.1.2 Anwendung auf Zufallszahlen

$$\begin{aligned}
B &= [0,1) \\
A_j &= \left[\frac{j-1}{k}, \frac{j}{k}\right) \\
p_j &= \frac{1}{k} \\
&\boxed{n \geq 5k} \\
\chi^2 &= \sum_{j=1}^{k} \frac{(n_j - \frac{n}{k})^2}{\frac{n}{k}} \\
&\sim \chi_{k-1}^2, \quad \text{falls } p_j = \frac{1}{k} \quad \forall j = 1, 2, \ldots, k
\end{aligned}$$

3.2.1.3 Der Kolmogoroff-Smirnoff-Test

$\xi_1, \xi_2, \ldots, \xi_n$ sind unabhängige und identisch verteilte Zufallsgrößen mit der Verteilungsfunktion:

$$P(\xi_i < x) = F(x) \quad \forall i$$

Die empirische Verteilungsfunktion ist:

$$F_n(x) = \frac{\#\{\xi_i : \xi_i < x\}}{n}$$

$$P\left(\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0\right) = 1$$

$$D_n = \sup_x |F_n(x) - F(x)|$$

3.2 Satz (von Glivenko-Cantelli)

$$\lim_{n\to\infty} P(\sqrt{n}\, D_n < x) = Q(x) = \begin{cases} 1 + 2\sum_{i=1}^{\infty}(-1)^i\, e^{-2i^2x^2} & \text{, falls x>0} \\ 0 & \text{, sonst} \end{cases}$$

Anpassungstest – Test of goodness of fit Wir haben Pseudozufallszahlen $u_1, \ldots, u_n$ mit $0 \le u_i < 1$. Wir nehmen die Gleichverteilung an, d.h. $F(x) = x, \quad 0 \le x < 1$.

$$D_n = \sup_x |F_n(x) - x|$$

siehe Abbildung

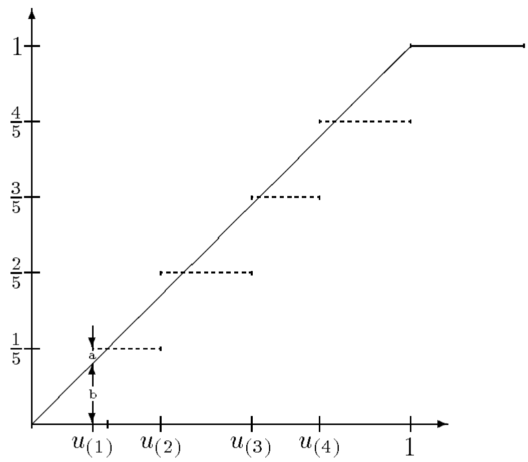

Abbildung 3.1: Vergleich zwischen der Gleichverteilung und einer empirischen Verteilungsfunktion für $n = 5$

- $u_1, \ldots, u_n$ der Größe nach ordnen
  erhalten Permutation mit $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$

- $$F_n(x) = \frac{\#\{u_i : u_i < x\}}{n}$$
- Bestimmen die Abweichung zwischen $F_n(x)$ und $F(x)$

1. $$D_n = \max\left\{\max_{1\le i\le n} a_i, \max_{1\le i\le n} b_i\right\}$$

   $$a_i = \left|u_{(i)} - \frac{i}{n}\right| \qquad b_i = \left|u_{(i)} - \frac{i-1}{n}\right|$$

2. $$\sqrt{n}\,D_n > C_\alpha \longrightarrow H \text{ ablehnen}$$

   $$\sqrt{n}\,D_n \le C_\alpha \longrightarrow H \text{ annehmen}$$

Die Hypothese lautet:

$$u_1, \ldots, u_n \sim R(0,1)$$

$$\begin{aligned} Q(C_\alpha) &= 1-\alpha \\ &= P(\sqrt{n}\,D_n < C_\alpha) = 1-\alpha \\ & P(\sqrt{n}\,D_n > C_\alpha) = \alpha \end{aligned}$$

α	$C_\alpha \approx$
0,001	1,63
0,05	1,36
0,01	1,23

3.2.2 Test auf Unabhängigkeit von Zufallszahlen

3.2.2.1 Run-Test

Betrachten Folge unabhängiger und identisch verteilter zufälliger Variable. Jede Teilfolge aufsteigender Zufallszahlen heißt *Run*.

2	1	2	3	2	4	1	7	8	9	0	Zufallszahlen
1.	2.			3.		4.				5.	Run
1	3			2		4				1	Länge des Runs

$u_1, \ldots, u_n$ sei Folge unabhängig, $R(0,1)$ verteilter zufälliger Variablen.

$$\begin{aligned} R &\; - \; \text{Länge eines Runs} \\ P(R=r) &= \frac{r}{(r+1)!} \\ R &: \begin{pmatrix} 1 & 2 & 3 & \ldots & r & \ldots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{8} & \ldots & \frac{r}{(r+1)!} & \ldots \end{pmatrix} \end{aligned}$$

Dehnen nun die Runs auf die Pseudozufallszahlen aus. $u_1, \ldots, u_n$ seien solche Pseudozufallszahlen und $R_1, \ldots, R_k$ ist die Folge der Längen der k Runs in der Folge.

Da Pseudozufallszahlen durch einen deterministischen Alogrithmus berechnet werden, sind diese auf keinen Fall unabhängig. Um nun unseren Test überhaupt sinnvoll zu machen, streichen wir nach jedem Run die nächste Zufallszahl.

$$\begin{aligned}
S_1 &= \inf\{n \in \Gamma^+ : u_n > u_{n+1}\} \\
S_2 &= \inf\{n > S_1 + 1 : u_n > u_{n+1}\} \\
&\vdots \\
S_{k+1} &= \inf\{n > S_k + 1 : u_n > u_{n+1}\} \\
& R_1^* = S_1; \quad R_2^* = S_2 - S_1 - 1; \quad \ldots; \quad R_{k+1}^* = S_{k+1} - S_k - 1
\end{aligned}$$

Wenn H gilt, also $u_1, \ldots, u_n$ einer $R(0,1)$-Verteilung entsprechen und unabhängig sind, dann gilt:

$$P(R^* = r) = \frac{r}{(r+1)!}$$

und die R_i^* $(i = 1, \ldots, m)$ bilden eine Folge von unabhängigen Runs.
Teilen Γ^+ in disjunkte Teilintervalle aufteilen:

$$[i_1 + 1, i_2], [i_2 + 1, i_3], \ldots, [i_k + 1, \infty)$$

$$0 = i_1 < i_2 < \ldots < i_k < \infty \qquad k \text{ Intervalle}$$

$$p_j^* = \sum_{l=i_j+1}^{i_{j+1}} P(R^* = l) = P(i_j + 1 \leq R^* \leq i_{j+1})$$

$$n_j = \#_{i=1,\ldots,m}\{R_i^* : i_j + 1 \leq R_i^* \leq i_{j+1}\}$$

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - mp_j^*)^2}{mp_j^*} \sim \chi_{k-1}^2$$

Falls H wahr ist, d.h. unabhängige, gleichverteilte Zufallszahlen liegen vor, gilt:
$\boxed{n_j \approx mp_j^*}$
Falls $\chi^2 >$ kritischer Wert, lehnen wir dir Unabhängigkeit ab.

3.2.2.2 Autokorrelation

Sei $u_1, \ldots, u_n$ eine Folge von zufälligen Variablen. Für alle m können wir nun bilden:

$$\rho_m(k) = \frac{cov(u_m, u_{m+k})}{\sigma_{u_m}\sigma_{u_{m+k}}}$$

wobei $1 \leq k \leq \frac{n}{2}$ und

$$u_{m+k} := u_{m+k-n}, \text{ falls } m + k > n$$

Wenn $u_1, \ldots, u_n$ identisch verteilt sind

$$\Longrightarrow \sigma_{u_j} = \sigma \quad \forall j$$

$$cov(u_m, u_{m+k}) = cov(u_1, u_{k+1})$$

$$\sigma_m(k) = \rho(k) = \frac{E(u_m \cdot u_{m+k}) - (Eu_m)^2}{\sigma^2} \quad \forall m, \quad k = 1, \ldots, \left[\frac{n}{2}\right]$$

Wir ersetzen jetzt den Erwartungswert durch den Mittelwert, dabei ist $u_1, \ldots, u_n$ eine Folge von Zufallszahlen:

$$\boxed{\hat{\rho}(k) = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i \cdot u_{i+k} - \left(\sum_{i=1}^{n} u_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} u_i^2 - \left(\sum_{i=1}^{n} u_i\right)^2}}$$ ist die *Autokorrelation* k-ter Ordnung.

Falls nun $u_1, \ldots, u_n$ echte Zufallszahlen, dann gilt:

$$\rho(k) = 0 \quad \forall k \geq 1 \qquad\qquad \hat{\rho}(k) \approx 0$$

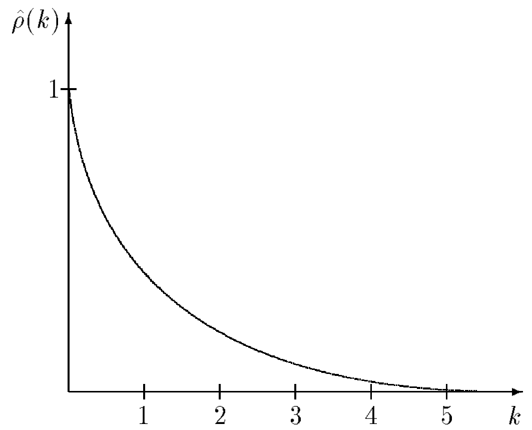

Abbildung 3.2: Darstellung der Autokorrelation k-ter Ordnung

3.3 Erzeugung speziell verteilter Zufallszahlen

3.3.1 Erzeugung einer diskreten zufälligen Variablen

Das Ziel ist es, folgende zufällige Variable zu erzeugen:

$$\xi : \begin{pmatrix} x_1 & x_2 & \ldots & x_m \\ p_1 & p_2 & \ldots & p_m \end{pmatrix}$$

Hierzu unterteilen wir das Intervall $[0, 1]$ folgendermaßen:

$u_1, \ldots, u_n$ ist eine Folge von erzeugten Pseudozufallszahlen. Wenn

$$\sum_{k=0}^{j-1} p_k \leq u_i < \sum_{k=0}^{j} p_k, \qquad \text{wobei } p_0 = 0$$

ist, definieren wir $\boxed{\xi := x_j}$.

3.3.2 Erzeugung stetig verteilter zufälliger Variablen

$$u_i \sim R(0,1) \quad i = 1, 2, \ldots, \quad \text{unabhängig}$$

$$h(u) := \begin{cases} 1, \text{ falls } 0 \leq u < 1 \\ 0, \text{ sonst} \end{cases}$$

$$\xi_i := \varphi(u_i)$$

Die so gebildeten ξ_i sind ebenfalls unabhängig und stetig verteilt. φ besitzt Umkehrfunktion φ^{-1}.

$$h_{\xi_i}(x) = h\left(\varphi^{-1}(x)\right) \cdot \left|\frac{d\varphi^{-1}}{dx}\right|$$

dabei sind $x = \varphi(u) \quad u = \varphi^{-1}(x)$.

3.3.2.1 Erzeugung einer Normalverteilung

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

$u_1, \ldots$ sind Folge von Pseudozufallszahlen.

Behauptung:

$$\xi_i := \Phi^{-1}(u_i) \Longrightarrow \xi_i \sim N(0,1)$$

Beweis:

$$\begin{aligned} h_{\xi_i}(x) &= h\left(\Phi(x)\right) \cdot \frac{d\Phi(x)}{dx} \\ & \quad\; 0 \leq \Phi(x) \leq 1 \Rightarrow h\left(\Phi(x)\right) \equiv 1 \\ &= \frac{d\Phi(x)}{dx} = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \end{aligned}$$

Allgemein

- Vorgeben einer Verteilungsfunktion $F(x) := P(\xi < x)$
  $\Rightarrow$ Dichtefunktion $f(x) = \frac{dF(x)}{dx}$
- Erzeugen Pseudozufallszahlen $u_1, u_2, \ldots$
- Erzeugen zufällige Variablen ξ_i mit

$$\xi_i = F^{-1}(u_i)$$

$$h_{\xi_i}(x) = \underbrace{h\left(F(x)\right)}_{\equiv 1} \cdot \frac{dF}{dx} = f(x)$$

Andere Möglichkeit für $N(0,1)$ erzeugen Pseudozufallszahlen $u_1, \ldots, u_m$.

$$\xi = \frac{\sum_{i=1}^{n} u_i - \frac{m}{2}}{\sqrt{\frac{m}{12}}} \sim N(0,1)$$

3.5 Beispiel $(m = 12)$

$$\xi = \sum_{i=1}^{12} u_i - 6$$

wenn $u_1, \ldots, u_m$ Zufallszahlen

$$\begin{aligned} Eu_i &= \frac{1}{2} = \mu \\ \sigma_{u_i}^2 &= \frac{1}{12} = \sigma^2 \\ \xi &= \frac{\sum_{i=1}^{m} u_i - m\mu}{\sqrt{m}\sigma} \end{aligned}$$

$\lim_{m \to \infty} P(\xi < x) = \Phi(x)$nach zentralen Grenzwertsatz, siehe Seite 60

3.3 Satz (Methode von Box-Muller)
Wenn $u, v \sim R(0,1)$ und unabhängig, dann gilt.

$$\xi = \sqrt{-2 \ln u} \cdot \cos(2\pi v)$$

und

$$\eta = \sqrt{-2 \ln u} \cdot \sin(2\pi v)$$

sind unabhängig standard-normalverteilt, d.h. $\xi, \eta \sim N(0,1)$.

Erzeugung von zwei Zufallszahlen, die beliebig abhängig sind $\xi, \eta \sim N(0,1)$ und unabhängig

$$\begin{aligned} \xi^* &:= \xi \\ \eta^* &:= \rho\xi + \sqrt{1-\rho^2}\eta \qquad \rho \in [0,1] \end{aligned}$$

Korrelation zwischen ξ^*, η^*:

$$\rho(\xi^*, \eta^*) = \rho$$

Wenn $\rho = 1$, dann gilt $\eta^* = \xi^* = \xi$, d.h. beide Werte sind gleich, bei $\rho = 0$ sind beide Werte unabhängig.

3.3.2.2 Erzeugung einer Exponentialverteilung

Wir haben $u_1, \ldots, u_n$ und wollen ein exponentialverteiltes $\xi \sim EX(\lambda)$ erzeugen, d.h.

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

ξ_i berechnet sich folgendermaßen:

$$\xi_i = F^{-1}(u_i) = -\frac{1}{\lambda} \ln(1 - u_i) \geq 0$$

3.3.2.3 Erzeugung einer Poissonverteilung

Es sind POISSONverteilte Zufallszahlen zu erzeugen, d.h.:

$$P(\xi = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda} \qquad i = 0, 1, 2, \ldots$$

Ausgangspunkt sind unabhängige, exponentialverteilte Zufallszahlen $\eta_1, \ldots, \eta_k \sim EX(\lambda)$.

3.4 Satz

$$\eta^{(k)} = \sum_{i=1}^{k} \eta_i$$

$\Longrightarrow$ Dichte von $\eta^{(k)}$:

$$f_\xi(y) = \begin{cases} \frac{\lambda^k}{(k-1)!} y^{k-1}\, e^{-\lambda y}, & \text{falls } y \geq 0 \\ 0 & \text{, sonst} \end{cases} \quad \text{– Dichte der sogenannten ERLANG-Verteilung}$$

Beweis: mittels vollständiger Induktion

k=1:

$$\begin{aligned} f_1(y) &= \lambda \cdot e^{-\lambda y} \\ F_1(y) &= 1 - e^{-\lambda y} \\ \frac{dF_1(y)}{dy} &= f_1(y) \\ & \quad \text{d.h. } \eta^{(1)} \sim EX(\lambda) \end{aligned}$$

Induktionsschritt:

$$\begin{aligned} f_{k+1}(y) &= \int_0^\infty f_k(x) \cdot f_1(y-x)\, dx \\ & \quad \text{Faltung der Dichten } f_k \text{ und } f_1 \\ &= \int_0^y \underbrace{\frac{\lambda^k}{(k-1)!} x^{k-1}\, e^{-\lambda x}}_{f_k(x)} \cdot \underbrace{\lambda \cdot e^{-\lambda(y-x)}}_{f_1(y-x)}\, dx \\ &= \int_0^y \frac{\lambda^{k+1}}{(k-1)!} x^{k-1}\, e^{-\lambda y}\, dx = \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda y} \cdot \int_0^y x^{k-1}\, dx \\ &= \underline{\frac{\lambda^{k+1}}{k!} \cdot y^k\, e^{-\lambda y}} \end{aligned}$$

3.5 Satz
Wenn $\eta_1, \ldots, \eta_n, \ldots \sim EX(\lambda)$ und unabhängig

$$\eta = \inf\left\{ n : \sum_{i=1}^{n+1} \eta_i > 1 \right\} \sim PO(\lambda)$$

d.h.:

$$P(\eta = i) = \frac{\lambda^i}{i!} e^{-\lambda} \qquad i = 1, 2, 3, \ldots$$

Index