Autor: Xuân Baldauf <xuan--2003.04.25--stochastik--sem02@studium.baldauf.org>

Datum: 09.04.2003

# Stochastik

gelesen von Manfred Wollenberg

### Organisatorisches

- Anmeldung zu den Übungen: http://www.informatik.uni-leipzig.de/Anmeldung/Sto03.html
- Übungsaufgaben alle 14 Tage.
- Übungsscheinerteilungsquote: 50%
- Klausur
  - keine Hilfsmittel
- Sprechzeit: Mittwoch, 16.00..18.00 Uhr, Zimmer HG 3-49.
- Vortragsweise
  - Folie
  - Tafel
  - mündlich

### Literatur

1. U. Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik; Vieweg Verlag; 1990 u.w.A [?] (sehr empfohlen)
2. K. Nawrotzki; Lehrbuch der Stochastik; H. Deutsch Verlag; 1994 (empfohlen)
3. K. Behren, G.Neuhau: Grundkurs Stochastik; Teubner; 1990 u.w.A (empfohlen)
4. N. Menzel: Stochastik für Einsteiger; Vieweg Verlag; 1997 (sehr gut, aber eher elementar)

## Inhalt der Vorlesung

0. Gegenstand der Stochastik
1. Wahrscheinlichkeitsraum
2. Bedingte Wahrscheinlichkeit
3. Diskrete Zufallsgrößen
4. stetige Zufallsgrößen
5. Grenzwertsätze
6. Statistik

# 0. Gegenstand der Stochastik

„Stochastik“ bezeichnet die Gebiete „Wahrscheinlichkeitstheorie und mathematische Statistik“.
Grobe Erklärung: „Stochastik ist die Mathematik des Zufalls“
Genauere Erklärung:

- **Stochastik** ist ein Teilgebiet der Mathematik, in dem Methoden zur Beschreibung und Beherrschung von zufälligen Vorgängen entwickelt werden.

- Dabei liefert die **Wahrscheinlichkeitstheorie** die mathematischen Modelle und die **mathematische Statistik** benutzt diese Modelle in der Statistik (das heißt: der Wissenschaft, die sich mit der Erfassung, Darstellung und Aufbereitung von empirischen Daten von Massenerscheinungen befasst).
- Insbesondere stellt die **mathematische Statistik** Verfahren bereit zur Ermittlung geeigneter mathematischer Modelle und ihrer Parameter aus den empirischen Daten.

**Definition: Zufall**

Ein Ereignis heißt **zufällig**, wenn es nicht vorhersehbar ist. (Im Alltagsleben und auch in der Stochastik.)

Die Stochastik beachtet dabei nicht, warum „der Einzelne“ oder „die Menschen“ das Ereignis nicht vorhersehen können, also die zwei Möglichkeiten:
1. Das Ereignis ist determiniert, aber wir kennen alle Einflüsse, Bedingungen, Daten oder|und es sind so viele, dass jede Rechenkunst versagt. (Solche Ereignisse sind in der Erkenntnistheorie nicht zufällig.)
2. Das Ereignis ist nicht determiniert. Selbst bei voller Kenntnis aller Einflüsse, Daten ist das Ereignis nicht berechenbar. (Solche Ereignisse sind in der Erkenntnistheorie zufällig.)

**Beispiel für Fall 1**

Chaotische Vorgänge (Wetter), Bewegung vieler Teilchen

**Beispiel für Fall 2**

Quantenmechanische Vorgänge (radioaktive Vorgänge, Mutationen, Billiardspiel)

## Geschichte der Stochastik

(gerafft)

| | |
|---|---|
| 16. Jahrhundert | Geronimo Cardano (1501..1576): Buch über das Würfelspiel |
| 17. Jahrhundert | Blaise Pascal (1623..1662) und Pierre de Fermats (1601..1655), Jacob Bernoulli (1654..1705): Gesetz der Großen Zahlen und erstes wissenschaftliches Buch: „Kunst des Vermutens“ |
| 18. Jahrhundert | Pierre S. de Laplace (1749..1827): Klassische Definition der Wahrscheinlichkeit und Buch „Analytische Theorie der Wahrscheinlichkeit“. |
| 19. Jahrhundert | Andei A Markov (1856..1922): Markov-Prozesse |
| 20. Jahrhundert | Karl Pearson (1857..1906): Methoden der Datenanalyse |
| | Andrej Kolmogorov (1903..1987): Axiomatisierung der Wahrscheinlichkeitstheorie |
| | John (=Janos=Johann) von Neumann (1903..1957): Monte Carlo Methode |

# 1. Wahrscheinlichkeitsraum

Ziel dieses Kapitels: Ein allgemeines mathematisches Modell zu entwickeln für beliebige zufällige Vorgänge.

## 1.1 Zufällige Vorgänge

- Erklärung: Ein **zufälliger Vorgang** ist ein Vorgang, dessen Ausgang (Ergebnis) im Rahmen verschiedener Möglichkeiten ungewiss ist und der sich, unter Einhaltung der den Vorgang kennzeichnenden äußeren Bedingungen, (zumindest gedanklich) beliebig oft wiederholen lässt.
- Bemerkung:
  - Die beliebige Wiederholbarkeit ist eine Idealisierung, die in der Praxis nur approximativ erfüllt ist.
  - Das mathematische Modell für zufällige Vorgänge, welches wir hier entwickeln, wird allerdings auch auf Vorgänge angwendet, welche im Sinne der Erklärung keine „richtig zufälligen Vorgänge" sind.
- Die möglichen Ausgänge (Ergebnisse) eines zufälligen Vorgangs fasst man zusammen zur sogenannten **Ergebnismenge**.
- Das bietet die Möglichkeit, Ereignise als Teilmengen dieser Ergebnisse (notiert als $\Omega$) aufzufassen und Operationen mit Ereignissen als Mengenoperationen betrachten.

#### Beispiel 1: Werfen eine Münze

Ergebnisse sind: Zahl oben oder Wappen oben.
$\Omega=\{Z, W\}$

#### Beispiel 2: Erfassung der Malariakranken

Mit hilfe einer Stichprobe sollen Malariakranke erfasst werden. Umfang der Stichprobe sei 100 Personen. Zufällige Ergebnisse sind die Zahlen 0,1,...,100.
$\Omega=\{0,1,\ldots,100\}$

### Bezeichnungen

#### Definition: Ergebnismenge

Gegeben sei ein zufälliger Vorgang mit den Ausgängen $w_1, w_2, \ldots, w_n$.
Die **Ergebnismenge** (auch: „Grundmenge", „Ergebnisraum", „Stichprobenraum") ist die Menge $\Omega=\{w_1, w_2, \ldots, w_n\}$.

#### Definition: Ereignis A

Gegeben sei ein zufälliger Vorgang mit den Ausgängen $w_1, w_2, \ldots, w_n$.
Ein **Ereignis** $A$ ist eine Teilmenge von $\Omega$, also $A=\{w_{j_1}, \ldots, w_{j_2}\} \subseteq \Omega$. Eine Ereignismenge wird interpretiert als das „Ereignis des zufälligen Vorgangs", welches als eingetreten gilt, wenn mindestens eines der Ereignisse $w_{j_1}, \ldots, w_{j_2}$ eingetreten ist. [Dies ist also eine Oder-Verknüpfung.]

**Definition: elementares Ereignis**

Ein **elementares Ereignis** (auch: „atomares Ereignis“) ist ein Ereignise $A=\{w\}$ mit $w\in\Omega$ . [Die Ereignis-Menge ist also eine Einer-Menge.]

**Definition: zusammengesetztes Ereignis**

Ein **zusammengesetztes Ereignis** A ist ein Ereignis, welches nicht elementar ist. Es gilt also $|A|\neq 1$ . [Ist das unmögliche Ereignis damit auch ein zusammengesetztes Ereignis? Nach Definition ja. Nach intuitiver Interpretation nicht.]

**Definition: sicheres Ereignis $\Omega$**

Ein **sicheres Ereignis** ist ein Ereignis, was immer eintritt. [Dieses Ereignis ist damit das betrachtete „Universum“ von Ereignissen (Ergebnismenge).]

**Definition: unmögliches Ereignis $\emptyset$**

Ein **unmögliches Ereignis** ist ein Ereignis, was nie eintritt. [Dieses Ereignis ist damit die leere Menge.]

**Definition: unvereinbar**

Zwei Ereignisse $A, B$ sind **unvereinbar**, wenn $A\cap B=\emptyset$ .

**Definition: komplementär**

Zwei Ereignisse $A, B$ sind **komplementär**, wenn $B=\overline{A}=\Omega\setminus A$ .

**Definition: vollständiger Satz von Ereignissen**

Ein **vollständiger Satz von Ereignissen** ist die Menge $\{A_j\}_{j=1}^{N}$ , für die gilt:

- $\bigcup_{j=1}^{N}(A_j)=\Omega$ und
- $\forall(i\neq j):((A_i\cap A_j)=\emptyset)$

[Ein vollständiger Satz von Ereignissen ist damit eine Partitionierung von $\Omega$ , oder? Bei der Partitionierung ist jedes Element der Partitionierung nicht leer. Dies ist hier offensichtlich nicht gefordert.]

**Beispiel**

Zwei nichtunterscheidbare Würfel werden gleichzeitig geworfen. Die Ergebnismenge ist $\Omega=\{(i,j)|(i\leqslant j)\wedge(j\in\{1,2,\ldots,6\})\}$ . Eine andere Darstellungsmöglichkeit wäre: $\Omega=\{(i+j,i)|i,j\in\{1,2,\ldots,6\}\}$

a) Ergebnis: Das Zahlenpaar $(1,3)$ kennzeichnet, dass ein Würfel die „1“ oben hat, der andere die „3“.

b) Ereignis: $A=\{(i,j)|j\in\{1,6,\ldots,6\}\}$ : Dies ist ein zusammengesetztes Ereignis

c)
$$A_1=\{(1,1)\}$$
$$\vdots$$
$$A_5=\{(i,5)|i\in\{1,2,\ldots,5\}\}$$
$$A_6=\{(i,6)|i\in\{1,2,\ldots,6\}\}$$
$$A_1\cup A_2\cup\ldots\cup A_6=\Omega$$
$$\forall(i\neq j):((A_i\cap A_j)=\emptyset)$$

Damit ist $\{A_j\}_{j=1}^6$ ist ein vollständiger Satz von Ereignissen.

Datum: 11.04.2003

Die relative Häufigkeit liefert ein empirisches Maß für den „Gewissheitsgrad“ des Eintretens des Ereignisses.

**Definition: absolute Häufigkeit**

Es sei $A$ ein Ereignis eines zufälligen Vorgangs. Der Vorgang werde $N$ mal wiederholt und $k$ -mal trete das Ereignis $A$ ein. Dann heißt:
$H_N(A):=k$ **absolute Häufigkeit** von $A$ in den $N$ Versuchen.

**Definition: relative Häufigkeit**

Es sei $A$ ein Ereignis eines zufälligen Vorgangs. Der Vorgang werde $N$ mal wiederholt und $k$ -mal trete das Ereignis $A$ ein. Dann heißt:
$h_N(A):=\frac{k}{N}$ **relative Häufigkeit** von $A$ in den $N$ Versuchen.

**Satz: Empirisches Gesetz der großen Zahlen**

Wenn es sich um einen zufälligen Vorgang handelt, gilt:
Die relativen Häufigkeiten $h_N(A)$ nähern sich im Mittel langer Versuchsreihen für $N\to\infty$ einem stabilen Wert $P(A)$ an, der als **Wahrscheinlichkeit des Ereignisses** $A$ bezeichnet wird.
Das ist natürlich nur approximativ nachprüfbar, Grenzüberänge sind Idealisierungen.

**Eigenschaften**

der relativen Häufigkeiten

1. $h_N(A)\in[0,1]$ Die relative Häufigkeit ist nicht negativ.
2. $h_N(\Omega)=1$ Die relative Häufigkeit ist normiert.
3. $h_N(A\cup B)=h_N(A)+h_N(B)$ Die relative Häufigkeit ist additiv.
4. $h_N(\emptyset)=0$ Die relative Häufigkeit des unmöglichen Ereignisses ist 0.
5. $h_N(A)=1-h_n(\overline{A})$ Die Summe der relativen Häufigkeit aller Elemente einer Partition ist die relative Häufigkeit des sicheren Ereignisses, nämlich 1.

## 1.2 Ereignisalgebra

Wir haben die möglichen Ergebnisse zusammengefasst zur **Ergebnismenge**. Um ein mathematisches Modell eines zufälligen Vorgangs zu entwerfen werden wir die Familie von Ereignissen mit Hilfe des Begriffs der **Ereignisalgebra** erfassen. Dazu brauchen wir aber erst einmal den Begriff der Mengenalgebra.

**Definition: Mengenalgebra**

Eine Familie $F$ von Mengen heißt **Mengenalgebra** über der nichtleeren Menge $\Omega$ , wenn gilt:

1. $\Omega\in F$
2. $F\subseteq\text{Potenzmenge}(\Omega)$
3. $(A\in F)\Rightarrow(\overline{A}\in F)$

4. $((A \in F) \wedge (B \in F)) \Rightarrow (A \cup B \in F)$

### Definition: σ-Algebra

Die Mengenalgebra $F$ über $\Omega$ heißt $\sigma$ **-Algebra**, wenn für jede Folge $(A_i)_i$ aus $F$ [also $\forall (i) : (A_i \in F)$ ] gilt: $\left( \bigcup_{\forall (i)} (A_i) \right) \in F$ .

#### Bemerkung

Wenn $\Omega$ endlich ist, dann ist jede Mengenalgebra $F$ über $\Omega$ sogar eine $\sigma$ -Algebra.
Wenn $\Omega$ unendlich ist, gilt das nicht immer.

#### Eigenschaften

1. Für eine Mengenalgebra $F$ gilt:
   - $\emptyset \in F$
   - $((A \in F) \wedge (B \in F)) \Rightarrow ((A \cap B) \in F)$
2. Für eine $\sigma$ -Algebra $F$ gilt:
   - $((A_i)_i \subset F) \Rightarrow \left( \left( \bigcap_{\forall (i)} (A_i) \right) \in F \right)$
3. In einer $\sigma$ -Algebra kann man abzählbar oft die Mengenoperationen ausführen, ohne die Algebra zu verlassen. [Die $\sigma$ -Algebra ist also abgeschlossen gegenüber den Mengenoperationen.]

#### Beispiel: Mengenalgebra ist im allgemeinen keine ?-Algebra

Sei $\Omega = \mathbb{N}$ und $F = \{ A \subset \mathbb{N} \,|\, (|A| < \infty) \vee (|\overline{A}| < \infty) \}$ .

- $F$ ist eine Mengenalgebra, denn es gilt:
  - $F \subseteq \text{Potenzmenge}(\Omega) = \text{Potenzmenge}(\mathbb{N})$
  - $\Omega \in F$ , da $\mathbb{N} \in F$
  - aus Symmetrie der Definition folgt: $(A \in F) \Rightarrow (\overline{A} \in F)$
  - Für die Abgeschlossenheit gegenüber Vereinigung gibt es vier mögliche Fälle: Seien $A, B \in F$ . Dann
    - 1. Fall: $((|A| < \infty) \wedge (|B| < \infty)) \Rightarrow (|A \cup B| < \infty)$
    - 2. Fall: $((|\overline{A}| < \infty) \wedge (|\overline{B}| < \infty)) \Rightarrow ((\overline{A \cup B} = (\overline{A} \cap \overline{B}) \subseteq \overline{B}) \Rightarrow (|\overline{A \cup B}| < \infty))$
    - Die anderen Fälle ( $((|A| < \infty) \wedge (|\overline{B}| < \infty))$ und $((|\overline{A}| < \infty) \wedge (|B| < \infty))$ ) lassen sich auf obige Fälle zurückführen.

  Damit hat $F$ alle Eigenschaften, die an eine Mengenalgebra gefordert werden.
- $F$ ist aber keine $\sigma$ -Algebra
  - Definieren wir eine Folge $(A_i)_i$ mit $\forall (i \in \mathbb{N}) : (A_i \in \{2 \cdot i\})$ . Es gilt damit: $A_i \in F$ .
  - Die Vereinigungsmenge über alle Folgenglieder ist: $A = \bigcup_{\forall (i)} (A_i) = 2 \cdot \mathbb{N}$
  - Das Komplement dieser Vereinigungsmenge ist: $\overline{A} = (2 \cdot \mathbb{N} + 1)$
    - Für das Komplement gilt aber: $\overline{A} \notin F$ , denn $\neg (|\overline{A}| < \infty)$ .

  Damit kann $F$ die Eigenschaften, die an eine $\sigma$ -Algebra gefordert werden, nicht haben.

### Definition 2: Ereignisalgebra

Es sei
- $\Omega$ eine Ergebnismenge und
- $F$ eine $\sigma$-Algebra über $\Omega$.

Dann heißt $F$ **Ereignisalgebra** über der Ergebnismenge $\Omega$.
[Jede Ereignis-Algebra ist also eine $\sigma$-Algebra.]

### Satz 1

Es sei $\Omega$ eine nichtleere Menge. Dann ist
- $F_1 = \{\emptyset, \Omega\}$ die kleinste Ereignisalgebra über $\Omega$ und
- $F_2 = Potenzmenge(\Omega)$ die größte Ereignisalgebra über $\Omega$.

#### Zweckmäßige Wahl von *F* für gegebenes ?

1. Man wählt eine Familie $E$ von Teilmengen von $\Omega$, die man als „Ereignisse“ haben will.
2. Dann definiert man sich $F(E) := \bigcap_{\forall(B):((B \supset E) \wedge B.istEreignisalgebraÜber(\Omega))} (B)$, wobei $B$ <u>alle</u> Ereignisalgebren über $\Omega$ durchläuft.

Damit erhält man: $F(E)$ ist wieder eine Familie von Teilmengen von $\Omega$ und es gilt:

### Satz 2

$F(E)$ ist die kleinste $E$ enthaltene Ereignisalgebra über $\Omega$. [Es gilt also: $F(E) \supset E$.]

#### Beweisidee

1. Es gibt wenigstens ein $B = Potenzmenge(\Omega) \supset E$. Also $\exists(F(E))$.
2. $(R \in F(E)) \Rightarrow \forall(B):(R \in B) \Rightarrow \forall(B):(\overline{R} \in B) \Rightarrow (\overline{R} \in F(E))$. Insbesondere ist $\Omega \in F(E)$.
3. $(R, S \in F(E)) \Rightarrow \forall(B):(R, S \in B) \Rightarrow \forall(B):((R \cup S) \in B) \Rightarrow ((R \cup S) \in F(E))$
4. Die $\sigma$-Algebra-Eigenschaft zeigt man genauso.
5. $\forall(B):(F(E) \subseteq B)$

#### Bemerkung

Wenn $\Omega$ diskret ist und $E \in \{\{\omega\} | \omega \in \Omega\}$ (wir also eine Familie von allen möglichen Elementarereignissen betrachten), dann folgt: $F(E) = Potenzmenge(\Omega)$. Dies ist die Standardwahl von $F$.

#### Bemerkung: Borelsche ?-Algebra

Im Fall $\forall(m \in \mathbb{N}):(\Omega = \mathbb{R}^m)$ wählt man als $E$ die Familie aller $m$-dimensionalen Intervalle | Rechtecke | Quader. Dann folgt $F(E) \subset Potenzmenge(\Omega)$. $F(E)$ heißt **Borelsche $\sigma$-Algebra** über $\mathbb{R}^m$ (geschrieben als „ $B(\mathbb{R}^m)$ “). Dies ist die Standardwahl von $F$.

[Die Borelsche $\sigma$-Algebra über dem Intervall $I$ („ $B(I)$ “) enthält also alle beliebigen Teil-Intervalle des gegebenen Intervalls $I$. Dies ist weniger als alle beliebigen Teilmengen, da Teil-Intervalle immer zusammenhängend sind (da sie Intervalle sind).]

[Tafelbild]
$\left(\Delta \in \mathbb{R}^m\right) \Rightarrow \left(F(\Delta) = \left(B(\mathbb{R}^m) \cap \Delta\right)\right)$.

**Bemerkung**

Es sei $\Omega$ nichtleere Menge und es seien $A_1, \ldots, A_N \subseteq \Omega$ Teilmengen von $\Omega$, die einen vollständigen Satz von Ereignissen bilden.
Dann ist $F(E)$ mit $E = \{A_1, \ldots, A_N\}$ die Familie von Teilmengen $C_i$ der Form

$$C_i = \bigcup_{j=1}^{M} (B_j) \text{ mit } B_j \in (E \cup \emptyset) \text{ und } M \in \{1, \ldots, N\}$$

## Mathematische Modellierung der Ereignisse eines zufälligen Vorgangs

- Die Menge der (interessierenden) Ergebnisse ist die Ergebnismenge $\Omega$.
- Die Menge der (interessierenden) Ereignisse ist eine Ereignisalgebra $F$ über $\Omega$.
- Wenn jedes der Ereignisse $A_1, \ldots, A_l$ eintritt, dann entspricht das der Menge (dem Ereignis) $A_1 \cap A_2 \cap \ldots \cap A_l$.
- Wenn mindestens eines der Ereignisse $A_1, \ldots, A_l$ eintritt, dann entspricht das der Menge (dem Ereigbnis) $A_1 \cup A_2 \cup \ldots \cup A_l$.

# 1.3 Wahrscheinlichkeitsraum

- Die Menge der Ereignisse eines zufälligen Vorgangs wird mathematisch erfasst durch die Ereignisalgebra $F$ über die Ereigbnismenge („Grundmenge“) $\Omega$
- Als mathematisches Äquivalent für die „experimentellen Wahrscheinlichkeiten“ (relative Häufigkeiten) von Ereignissen postuliert man einfach eine Wahrscheinlichkeitsverteilung, wobei man sich an den Eigenschaften von relativen Häufigkeiten orientiert.

### Definition 1: Wahrscheinlichkeitsverteilung, Wahrscheinlichkeitsraum

Es sei $F$ eine Ereignisalgebra über einer Ergebnismenge $\Omega$. Dann heißt die Abbildung $P(\cdot): F \to \mathbb{R}$ eine **Wahrscheinlichkeitsverteilung** (kurz Verteilung) über $F$ und das Tripel $(\Omega, F, P(\cdot))$ ein **Wahrscheinlichkeitsraum**, wenn die folgenden Eigenschaften erfüllt sind:
1. Positivität: $\forall (A \in F) : (P(A) \geqslant 0)$
2. Normiertheit: $P(\Omega) = 1$
3. $\sigma$-Additivität: Für jede Folge $(A_j)_j \subset F$ mit $\forall (j \neq k) : ((A_j \cap A_k) = \emptyset)$ (wo also die Folgenglieder disjunkt sind) gilt: $P\left(\bigcup_{\forall (j)} (A_j)\right) = \sum_{\forall (j)} (P(A_j))$

**Bemerkung**

Vom zufälligen Vorgang ist also nur das Tripel $(\Omega, F, P(\cdot))$ geblieben. Umgekehrt kann man auch jedes solche Tripel dann auffassen als Wahrscheinlichkeitsraum eines zufälligen Vorgangs.

Ein zufälliger Vorgang ist also gegeben durch:
- die Ergebnismenge $\Omega$,

- die Ereignisalgebra $F$,
  - Beispiele dafür sind:
    - $F(E)$
    - $Potenzmenge(\Omega)$
    - $B(\Omega)$ (Borelsche $\sigma$-Algebra über $\Omega$)
- die Verteilung $P(\cdot)$.
  - ermittelt aus
    - theoretischen Überlegungen oder
    - experimentell über die relativen Häufigkeiten
    - „Gemisch“ aus beidem obigem

**Beispiel 1: Münzwurf mit 2 unterschiedlichen (idealen) Münzen**

- $\Omega=\{(Z,Z),(Z,W),(W,Z),(W,W)\}$
- $Potenzmenge(\Omega)=F\supset A=\{(Z,Z),(W,W)\}$
- $|\Omega|=4$
- $|Potenzmenge(\Omega)|=2^{|\Omega|}=2^4=16$
- $P(\cdot)$ folgt aus Symmetrieüberlegung:
  - $P(\{(Z,Z)\})=P(\{(Z,W)\})=P(\{(W,Z)\})=P(\{(W,W)\})$
  - $1=P(\Omega)$
    $=P(\{(Z,Z)\}\cup\{(Z,W)\}\cup\{(W,Z)\}\cup\{(W,W)\})$
    $=P(\{(Z,Z)\})+P(\{(Z,W)\})+P(\{(W,Z)\})+P(\{(W,W)\})$
  - $1=P(\Omega)=4\cdot P(\{(Z,Z)\})$
    $\Rightarrow P(\{(Z,Z)\})=\frac{1}{4}$

**Beispiel 2: Test einer Person auf Krankheit**

- Der zufälliger Vorgang ist: Eine willkürlich herausgegriffene Person aus einer Gruppe von Menschen ist krank.
- $\Omega=\{K,G\}$ [(krank oder gesund)], $F=Potenzmenge(\Omega)$
- $P(\{K\})=\frac{\text{Anzahl der Kranken in der Gruppe}}{\text{Anzahl der Menschen in der Gruppe}}$

Experimentelle Ermittlung [...Tafel abgewischt...]

## Satz 1

Es sei $(\Omega,F,P(\cdot))$ ein Wahrscheinlichkeitsraum.
Dann gilt (mit $A\in F$ und $n\in\mathbb{N}$):

- $P(\emptyset)=0$
  - Beweis:
    - $P(\emptyset)=1-P(\overline{\emptyset})=1-P(\Omega)=1-1=0$
- $P(A)\leqslant 1$
  - Beweis:
    - $P(A)\leqslant P(A)+P(\overline{A})=P(\Omega)=1$

- $P(\overline{A})=1-P(A)$
  - Beweis:
    - $1=P(\Omega)=P(A\cup\overline{A})=P(A)+P(\overline{A})$
- Für jedes $n$-Tupel paarweiser disjunkter Ereignisse $(A_1, A_2, \ldots, A_n)\in F$ gilt: $P(A_1\cup A_2\cup\ldots\cup A_n)=P(A_1)+P(A_2)+\ldots+P(A_n)$

**Beispiel 1: Gleichverteilung (Laplace-Verteilung)**

Es sei $\Omega=\{w_1, w_2, \ldots, w_n\}$ eine Menge aus $n$ Elementen, also $|\Omega|=n$. Es sei $P(\cdot)$ gegeben durch $\forall(A\in F):\left(P(A)=\frac{|a|}{w}\right)$, insbesondere $\forall(\omega\in\Omega):\left((\{w\})=\frac{1}{n}\right)$.
Man sieht leicht, $P(\cdot)$ ist wirklich eine Verteilung (Man prüfe 1., 2., 3. aus Definition 1).
$P(\cdot)$ ist eine sogenannte **Gleichverteilung**, da die Wahrscheinlichkeiten für das Eintreten der elementaren Ereignisse $\{w_j\}$ alle gleich sind.

**Beispiel: Münzwurf**

**Beispiel: Würfelwurf**

**Beispiel 2: Zweiersystem (Bernoulli-Experiment)**

Es sei $\Omega=\{0,1\}$, $F=Potenzmenge(\Omega)=\{\emptyset, \{0\}, \{1\}, \Omega\}$. Dies ist also ein zufälliger Vorgang mit zwei Ergebnissen, bezeichnet mit „0“ und „1“.

- Wir geben eine Zahl $p\in[0,1]$ vor und definieren eine Verteilung über $(F, \Omega)$ durch:
  - $P(\emptyset)=0$, $P(\{0\})=1-p$, $P(\{1\})=p$, $P(\Omega)=1$

  Man sieht leicht, dass $P(\cdot)$ alle Eigenschaften einer Verteilung erfüllt.
- Umgekehrt sei $P(\cdot)$ eine Verteilung über diese Ereignisalgebra $(Potenzmenge(\Omega), \Omega)$, $\Omega=\{0,1\}$. Dann gibt es eine Zahl $p\in[0, 1]$, sodass
  - $P(\emptyset)=0$, $P(\{0\})=1-p$, $P(\{1\})=p$, $P(\Omega)=1$

  gilt.

### Satz

Eine Verteilung über ein Zweiersystem ist vollständig durch genau eine Zahl $p\in[0,1]$ festgelegt.

### Definition 2: diskret

Ein Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ heißt **diskret**, wenn $\Omega$ aus abzählbar vielen Elementen besteht und $F=Potenzmenge(\Omega)$ ist.

### Definition 2: endlich

Ein Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ heißt **endlich**, wenn $\Omega$ aus endlich vielen Elementen besteht und $F=Potenzmenge(\Omega)$ ist.

### Definition 3: diskrete Wahrscheinlichkeitsdichte

Es sei $(\Omega, F, P(\cdot))$ ein diskreter Wahrscheinlichkeitsraum. Dann heißt die Funktion $p(\cdot): \Omega \to \mathbb{R}$ definiert durch $\forall(\omega \in \Omega): (p(\omega) = P(\{\omega\}))$
**diskrete Wahrscheinlichkeitsdichte**.

### Definition 3: Wahrscheinlichkeitsfunktion

Eine diskrete Wahrscheinlichkeitsdichte heißt auch **Wahrscheinlichkeitsfunktion**.

#### Bemerkung

Die Wahrscheinlichkeitsdichte ist also die Einschränkung der Verteilung auf die elementaren Ereignisse ( $\approx$ Ergebnisse). Sie wird auch Zählmaß genannt.

### Satz 2

Es sei $(\Omega, F, P(\cdot))$ ein diskreter Wahrscheinlichkeitsraum.
Dann
- ist die Verteilung $P(\cdot)$ vollständig durch die Wahrscheinlichkeiten $\forall(\omega \in \Omega): (p(\omega) = P(\{\omega\}))$ bestimmt und es
- gilt: $P(A) = \sum_{\forall(\omega \in A)} (p(\omega))$

#### Beweis

Es ist nur die Formel zu zeigen. Es gilt:

$$P(A) = P\left(\bigcup_{\forall(w \in A)} (\{w\})\right) = \sum_{\forall(w \in A)} (P(\{\omega\})) = \sum_{\forall(w \in A)} (p(\omega))$$

#### Bemerkung

Dieser Satz ist sehr einfach, aber wichtig. Es genügt also die Wahrscheinlichkeiten der elementaren Ereignisse zu kennen.

### Satz 3

Es sei $\Omega$ eine diskrete Menge und es sei $p(\cdot): \Omega \to \mathbb{R}$ eine Abbildung mit den Eigenschaften:
1. $\forall(\omega \in \Omega): (p(\omega) \geqslant 0)$
2. $\left(\sum_{\forall(\omega \in \Omega)} (p(w))\right) = 1$

Wenn wir eine Abbildung $P(\cdot): Potenzmenge(\Omega) \to \mathbb{R}$ durch $\forall(A \in Potenzmenge(\Omega)): \left(P(A) := \sum_{\forall(\omega \in A)} (p(\omega))\right)$ definieren, so ist $(\Omega, Potenzmenge(\Omega), P(\cdot))$ ein diskreter Wahrscheinlichkeitsraum.

#### Bemerkung 1

Durch Vorgabe einer Funktion $p: \Omega \to \mathbb{R}$ mit den Eigenschaften 1. und 2. kann man sich Verteilungen über $(\Omega, Potenzmenge(\Omega))$ verschaffen.

#### Bemerkung 2

Bei diskreten Wahrscheinlichkeitsräumen schreibt man oft nur $(\Omega, P(\cdot))$, lässt also $F = Potenzmenge(\Omega)$ weg.

## 1.4 Eigenschaften von Verteilungen

### Definition: fast unmöglich

Ein Ereignis $A \in F$ eines Wahrscheinlichkeitsraumes $(\Omega, F, P(\cdot))$ heißt **fast unmöglich**, wenn gilt: $P(A)=0$.

### Definition: fast sicher

Ein Ereignis $A \in F$ eines Wahrscheinlichkeitsraumes $(\Omega, F, P(\cdot))$ heißt **fast sicher**, wenn gilt: $P(A)=1$.

#### Eigenschaften

- Das unmögliche Ereignis $\emptyset$ ist auch fast unmöglich, da $P(\emptyset)=0$.
- Das sichere Ereignis $\Omega$ ist auch fast sicher, da $P(\Omega)=1$.

[Fast sichere Ereignisse unterscheiden sich also vom sicheren Ereignis nur dadurch, dass ihre Elemente anders sein können, die „Wirkung“ (nämlich ihre Wahrscheinlichkeit) jedoch gleich ist. Damit ist ein fast sicheres Ereignis genauso sicher wie das sichere Ereignis. Das Wort „fast“ ist in dieser Hinsicht eher irreführend.]

#### Beispiel

Sei $\Omega=\{1,2,3,4,5\}$, $F=Potenzmenge(\Omega)$, $\forall(j \in \{1,2,3,4\}):\left(P(\{j\})=\frac{1}{4}\right)$, $P(\{5\})=0$.

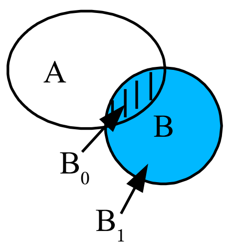
- $\{5\}$ ist ein fast unmögliches Ereignis, aber eben nicht das unmögliche Ereignis selbst.
- $\{1,2,3,4\}$ ist ein fast sicheres Ereignis, denn $P(\{1,2,3,4\})=1$, aber eben nicht das sichere Ereignis selbst.
- $\forall(A \subseteq \Omega):\left(P(A)=\sum_{\forall(\omega \in A)}\left(P(\{\omega\})\right)\right)$

[...]

### Satz 1

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum. Dann gilt:

1. $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
   - Beweis:
     $B$ wird in $B_0$ und $B_1$ zerlegt:
     - $B_0 := A \cap B$, $B_1 := \overline{A} \cap B$
     - $\Rightarrow (A \cup B)=(A \cup (\overline{A} \cap B))=(A \cup B_1)$
     - Das ergibt $P(A \cup B)=P(A \cup B_1)=P(A)+P(B_1)$ (*)
     - Weiter gilt:



$P(B)=P(B_0 \cup B_1)=P(B_0)+P(B_1)$
$\Rightarrow P(B_1)=P(B)-P(B_0)$

- Eingesetzt in (*) ergibt dies: $P(A \cup B)=P(A)+P(B)-P(A \cap B)$

2. $(A \subseteq B) \Rightarrow (P(A) \leqslant P(B))$ (Monotonie)
   - Beweis:
     - $A \subseteq B$. Das ergibt $B=A \cup (B \cap \overline{A})$, also:
       $P(B)=P(A \cup (B \cap \overline{A}))=P(A)+P(B \cap \overline{A}) \geqslant P(A)$ (und das ist Monotonie)
3. Es seien $X, Y \in F$ mit $P(X)=0$, $P(Y)=1$.
   Dann gilt:
   - $\forall (A \in F):(P(A \cup X)=P(A))$
   - $\forall (A \in F):(P(A \cap X)=0)$
     - Beweis:
       - Es gilt: $(A \cap X) \subset X$. Wir benutzen die Aussage 2. (Monotonie) und erhalten $0 \leqslant P(A \cap X) \leqslant P(X)=0$
         $\Rightarrow P(A \cap X)=0$
   - $\forall (A \in F):(P(A \cup Y)=1)$
     - Beweis:
       - Es gilt:
         $1 \geqslant P(A \cup Y)=P(A)+P(Y)-P(A \cap Y)=P(A)+1-P(A \cap Y) \geqslant P(A)+1-P(A)=1$
         $\Rightarrow P(A \cup Y)=1$
   - $\forall (A \in F):(P(A \cap Y)=P(A))$

Datum: 23.04.2003

[...zu spät 30 Minuten...]
<import from=„Manfred Wollenberg“>

**Satz 2**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $(A_j)_{j=1}^{\infty}\subseteq F$ eine Folge. Dann gilt:

1. $P\left(\bigcup_{\forall(j)}(A_j)\right)=\lim_{n\to\infty}\left(P\left(A_1\cup A_2\cup\ldots\cup A_n\right)\right)$
2. $P\left(\bigcap_{\forall(j)}(A_j)\right)=\lim_{n\to\infty}\left(P\left(A_1\cap A_2\cap\ldots\cap A_n\right)\right)$

**Satz 2**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $(A_j)_{j=1}^{\infty}\subseteq F$ eine monoton wachsende Folge (also $A_1\subseteq A_2\subseteq\ldots\subseteq A_i\ldots$ ). Dann gilt:

3. $P\left(\bigcup_{\forall(j)}(A_j)\right)=\lim_{n\to\infty}\left(P(A_n)\right)$

**Satz 2**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $(A_j)_{j=1}^{\infty}\subseteq F$ eine monoton fallende Folge (also $A_1\supseteq A_2\supseteq\ldots\supseteq A_i\ldots$ ). Dann gilt:

4. $P\left(\bigcap_{\forall(j)}(A_j)\right)=\lim_{n\to\infty}\left(P(A_n)\right)$

**Bemerkung**

Diese Sätze reduzieren also die Berechnungen von Wahrscheinlichkeiten mit unendlichen Folgen auf Berechnungen von Wahrscheinlichkeiten mit endlichen Folgen.

## 1.5 Beispiele für Wahrscheinlichkeitsräume

### Diskreter (unendlicher) Wahrscheinlichkeitsraum

- Der zufällige Vorgang bestehe darin, dass wir eine (ideale) Münze so oft werfen, bis die Seite „Zahl“ oben liegt. Ergebnisse sind die Menge der Zahlen der Würfe, bis erstmals die Seite „Zahl“ oben liegt. Also kann jede natürliche Zahl Ergebnis sein. Es besteht aber auch die Möglichkeit, dass nie die Seite „Zahl“ oben liegt. Dies wird notiert durch $\{\infty\}$.
- Die Ergebnismenge dieses Wahrscheinlichkeitsraums ist $\Omega=\left(\mathbb{N}\cup\{\infty\}\right)$
- Berechnung der Wahrscheinlichkeit der elementaren Ereignisse:
  - Fall $n=1$
    $P(\{1\})=\frac{1}{2}$
  - Fall $n=2$
    $P(\{1\})=\frac{1}{4}=\left(\frac{1}{2}\right)^2$, denn es gibt 4 gleichberechtigte Möglichkeiten bei zwei aufeinanderfolgenden Würfen (die Möglichkeiten $(W,W)$, $(W,Z)$, $(Z,W)$, $(Z,Z)$), von denen genau eine (nämlich $(W,Z)$) die günstige ist.

- Fall $n \in \mathbb{N}$ : (allgemein)
  $P(\{n\}) = \left(\frac{1}{2}\right)^n$
  Damit ist $\lim_{n \to \infty} (P(\{n\})) = 0$

</import>

- Damit folgt $\sum_{n=1}^{\infty} (P(\{n\})) = \sum_{n=1}^{\infty} \left(\left(\frac{1}{2}\right)^n\right) = 1$ . Man wird also für das elementare Ereignis $\{\infty\}$ die Wahrscheinlichkeit $P(\{\infty\}) = 0$ ansetzen. Im Prinzip kann man also auch $\Omega = \mathbb{N}$ wählen.
- Nach Satz 3 (1.4) haben wir damit eine Verteilung $P(\cdot)$ auf $F = Potenzmenge(\Omega)$ gegeben: $P(A) = \sum_{\forall (w \in A)} (P(\{\omega\}))$ .

**Beispiel**

- $A$ sei das Ereignis: „Zahl" liegt zuerst oben bei einer geraden Zahl von Würfen der Münze.
- Berechnung von $P(A)$ : $P(A) = P(\{2 \cdot n | n \in \mathbb{N}\})$ ,
  also $P(A) = \sum_{n=1}^{\infty} (P(\{2 \cdot n\})) = \sum_{n=1}^{\infty} \left(\left(\frac{1}{2}\right)^{2 \cdot n}\right) = \sum_{n=1}^{\infty} \left(\left(\frac{1}{4}\right)^n\right) = \frac{1}{4} \cdot \frac{1}{1 - \frac{1}{4}} = \sum_{n=0}^{\infty} \left(\left(\frac{1}{4}\right)^n\right) - 1 = \frac{1}{1 - \frac{1}{4}} - 1 = \frac{1}{3}$
- Das Resultat ist etwas überraschend, denn damit ist $P(\overline{A}) = \frac{2}{3}$ , mit dem Ereignis $\overline{A}$ , dass „Zahl" zuerst oben liegt bei einer ungeraden Anzahl von Würfen.
- Hier ist
  - $X = \mathbb{N}$ ein fast sicheres Ereignis und
  - $Y = \{\infty\}$ ein fast unmögliches Ereignis.
- Es gibt hier zu jeder Zahl $q \in [0,1]$ ein Ereignis $E_q \in Potenzmenge(\Omega) = F$ , sodass gilt: $P(E_q) = q$ : $\forall (q \in [0,1]) : \exists (E_q \in Potenzmenge(\Omega) = F) : (P(E_q) = q)$
  - Begründung:
    - $q$ hat die Darstellung: $\forall (a_l \in \{0,1\}) : \left(q = \sum_{l=1}^{\infty} (a_l \cdot 2^{-l})\right)$ (Dual[zahl]darstellung)
    - Setze $E_q := \{l \in \mathbb{N} | a_l = 1\} \in \mathbb{N}$ . Wir betrachten $E_q \in Potenzmenge(\mathbb{N})$ als Ereignis und erhalten: $P(E_q) = \sum_{\forall (v) : (a_v = 1)} (P(\{v\})) = \sum_{l=1}^{\infty} (a_l \cdot P(\{l\})) = \sum_{l=1}^{\infty} \left(a_l \cdot \left(\frac{1}{2}\right)^l\right) = \sum_{l=1}^{\infty} (a_l \cdot 2^{-l}) = q$
  - Es gibt also ein <u>Kontinu</u>um von Ereignissen in dem <u>diskret</u>en Wahrscheinlichkeitsraum.

**Zu den Übungsaufgaben**

- Die Mindestanzahl der Würfe, damit man wenigstens einmal das Ergebnis „Zahl oben" mit einer Wahrscheinlichkeit von 0.99 hat, ist:
  $$A_n = \{\text{Zahl oben in } n \text{ Würfen mindestens einmal}\}$$
  $$= \{1, 2, \ldots, n\}$$
  $$= \bigcup_{j=1}^{n} (\{j\})$$

$$P(A_n)=\sum_{j=1}^{n}\left(P(\{j\})\right)=\sum_{j=1}^{n}\left(\left(\frac{1}{2}\right)^{j}\right)=1-\left(\frac{1}{2}\right)^{n}=1-P(\overline{A_n})$$

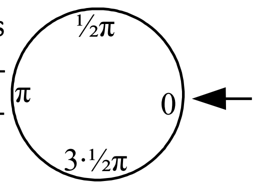
- Abschätzung von $n$ (Zahl der Würfe): $P(A_n)\geqslant 0.99$
  Das ergibt:
  $$1-\left(\frac{1}{2}\right)^{n}\geqslant 0.99$$
  $$\Leftrightarrow 0.01\geqslant\left(\frac{1}{2}\right)^{n}$$
  $$\Leftrightarrow 2^{n}\geqslant 100$$
  $$\Leftrightarrow n\cdot\ln(2)\geqslant\ln(100)$$
  $$\Leftrightarrow n\geqslant\frac{\ln(100)}{\ln(2)}\approx 6.65$$
  $\Rightarrow$ Die Mindestanzahl ist $n=7$

**Bemerkung**

Ähnliche Aufgaben treten oft auf. Sie interessiert ein Ereignis $A$ eines zufälligen Vorgangs. Insbesondere: Wie oft muss ich den zufälligen Vorgang wiederholen, damit wenigstens einmal $A$ in dieser Reihe von Versuchen mit vorgegebener Wahrscheinlichkeit $q$ eintritt. $P(A)=p$ (Ersatz für $\frac{1}{2}$, $q=0.99$ ). Der Wahrscheinlichkeitsraum ist: $(\Omega=\mathbb{N}, \mathit{Potenzmenge}(\mathbb{N}), P(\cdot))$ .

## Stetiger Wahrscheinlichkeitsraum

Betrachten wir einen zufälliger Vorgang: Das Drehen eines Kreisrouletts um seine Achse. Das Ergebnis ist eine Zahl (Winkel), die dem Pfeil gegenüber steht.




- Die Ergebnismenge $\Omega=[0,2\cdot\pi[\subset\mathbb{R}$ ist ein Intervall der reellen Achse. Als Ereignisalgebra wählt man hier die kleinste Ereignisalgebra, die alle Teil-Intervalle $]\alpha,\beta[\subseteq[0,2\cdot\pi[$ enthält. Das ist die sogenannte Borelsche $\sigma$-Algebra $B([0,2\cdot\pi[)$ . $B([0,2\cdot\pi[)$ ist echt kleiner als $\mathit{Potenzmenge}([0,2\cdot\pi[)$ .
- Als Verteilung $P(\cdot)$ auf $F=B([0,2\cdot\pi[)$ wählt man hier eine „stetige Gleichverteilung“ auf dem Intervall $[0,2\cdot\pi[$ : $P(\Delta):=\frac{|\Delta|}{2\cdot\pi}$ (wobei $|\Delta|$ das Maß von $\Delta\in F$ ist, also für $\Delta=]\alpha,\beta[$ ist $|\Delta|=\beta-\alpha$ , also ist $|\Delta|$ die Länge des Intervalls). Man sieht sofort: $P(\cdot)$ ist normiert: $P(\Omega)=1$ und $P(\Delta)\geqslant 0$ . Die $\sigma$-Additivität von $P(\cdot)$ ist schwieriger zu zeigen, gilt auf $F$ aber nicht auf $\mathit{Potenzmenge}([0,2\cdot\pi[)$
- Diese Verteilung $P(\cdot)$ über $B([0,2\cdot\pi[)$ ist eine spezielle **geometrische Gleichverteilung**. Allgemein heißt die Verteilung $P(\cdot)$ über ein Gebiet $D\subseteq\mathbb{R}^n$ **geometrisch gleichverteilt**, wenn $\forall(\Delta\in B(D)):\left(P(\Delta)=\frac{|\Delta|}{|D|}\right)$ .

## 1.6 Endliche Wahrscheinlichkeitsräume und Kombinatorik

In den Anwendungen hat man es meist mit endlichen Wahrscheinlichkeitsräumen $(\Omega, Potenzmenge(\Omega), P(\cdot))$, $\Omega=\{\omega_1,\ldots,w_n\}$ zu tun. Oft ist die Verteilung $P(\cdot)$ eine Laplaceverteilung, das heißt: $\forall(j):\left(P(\{\omega_i\})=\frac{1}{n}\right)$, also $\forall(A\in\Omega):\left(P(A)=\frac{|A|}{n}\right)$. Zum Beispiel Würfeln, Lose-Ziehen usw. Dann lassen sich viele Berechnungen auf Kombinatorische Betrachtungen zurückführen.

### Kombinatorik

- Die Kombinatorik beschäftigt sich mit der Auswahl und Anordnung von Teilmengen aus einer gegebenen endlichen Menge.

**Satz: Grundprinzip des Zählens (Multiplikationsregel):**

Es sollen $k$-Tupel $(\alpha_1,\alpha_2,\ldots,\alpha_n)$ gebildet werden, indem man die $k$-Plätze nacheinander von links nach rechts besetzt. Dabei gibt es

- $j_1$ Möglichkeiten für die Wahl von $\alpha_1$
- $j_2$ Möglichkeiten für die Wahl von $\alpha_2$
- ...
- $j_n$ Möglichkeiten für die Wahl von $\alpha_n$

Dann gibt es insgesamt $j_1\cdot j_2\cdot\ldots\cdot j_k$ verschiedene solche $k$-Tupel.

### Permutationen und Kombinationen

Es sei eine $n$-elementige Menge $M$ gegeben. Da die Natur der Elemente von $M$ unwichtig ist, kann man oft $M=\{1,2,\ldots,n\}$ setzen, was wier hier tun.

**Definition: Permutation mit Wiederholung**

Eine $k$-**Permutation** aus $M=\{1,2,\ldots,n\}$ **mit Wiederholung** ist ein $k$-Tupel $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ mit Komponenten $\alpha_1,\alpha_2,\ldots,\alpha_k\in M$. Es können Elemente aus $M$ mehrfach auftreten.

**Definition: Permutation ohne Wiederholung**

Eine $k$-**Permutation** aus $M=\{1,2,\ldots,n\}$ **ohne Wiederholung** ist ein $k$-Tupel $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ mit lauter verschiedenen Komponenten $\alpha_1,\alpha_2,\ldots,\alpha_k\in M$. „Lauter verschieden“ heißt: $\forall(i\neq j):(\alpha_i\neq\alpha_j)$. Folglich gilt $k\leqslant n=|M|$.

**Definition: Kombination mit Wiederholung**

Eine $k$-**Kombinationen** aus $M=\{1,2,\ldots,n\}$ **mit Wiederholung** ist eine $k$-Permutationen aus $M$ mit Wiederholung und Anordnungseigenschaft $\alpha_1\leqslant\alpha_2\leqslant\ldots\leqslant\alpha_k$.

**Definition: Kombination ohne Wiederholung**

Eine $k$-**Kombination** aus $M=\{1,2,\ldots,n\}$ **ohne Wiederholung** ist eine $k$-Permutationen aus $M$ ohne Wiederholung und der Anordnungseigenschaft $\alpha_1 < \alpha_2 < \ldots < \alpha_k$.

**Bemerkung**

Man schreibt für die entsprechenden Mengen von $k$-Tupeln:
$Per_k^n(mW)$, $Per_k^n(oW)$, $Kom_k^n(mW)$, $Kom_k^n(oW)$

**Bemerkung**

Früher bezeichnete man nur $Per_k^n(oW)$ als Permutationen und andere als Variationen. Also „Vorsicht bei Bezeichnungen".
$k$-Kombinationen bedeutet für beliebige Mengen $M$, man interessiert sich nicht für die „zeitliche Reihenfolge" der Auswahl, wie zum Beispiel beim Lotto.

### Satz

Es gilt:

- $\left|Per_k^n(mW)\right| = n^k$
- $\left|Per_k^n(oW)\right| = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$
- $\left|Kom_k^n(mW)\right| = \binom{n+k-1}{k}$
- $\left|Kom_k^n(oW)\right| = \binom{n}{k}$

**Beispiel**

Sei $M = \{1,2,3\}$, $n=3$, $k=2$.
Dann ist:

- $Per_2^3(mW) = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}$ $3^2 = 9$ Elemente
- $Per_2^3(oW) = \{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}$ $3 \cdot 2 = 6$ Elemente
- $Kom_2^3(mW) = \{(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)\}$ 6 Elemente
- $Kom_2^3(oW) = \{(1,2),(1,3),(2,3)\}$ 3 Elemente

**Beispiel**

$Per_4^3(mW) = \{(1,1,1,1), \ldots\}$

**Beweisansatz**

Benutzung des Grundprinzips des Zählens:
$\left|Per_k^n(mW)\right| = n \cdot n \cdot n \cdot \ldots \cdot n = n^k$
[...]

**Beispiel**

[...]
Da zeitliche Reihenfolge keine Rolle spielt, sind die Ergebnisse $\Omega = Kom_6^{49}(oW)$ (Laplace-Experiment) alle Ergebnisse gleich wahrscheinlich.

$P(6 \text{ richtige}) = \frac{1}{|\Omega|} = \frac{1}{\binom{49}{6}} = \frac{1}{13983816}$ Dies ist eine sehr kleine Zahl.

$P(4 \text{ richtige}) = \frac{\text{Anzahl der günstigen Fälle}}{|\Omega|}$

**Ermittlung dieser Anzahl**

4 Richtige können sich auf die 6 Richtige verteilen auf folgende Weise: $|Kom_4^6(oW)| = \binom{6}{4}$

2 falsche Zahlen verteilen sich auf die restlichen 43 Zahlen: $|Kom_2^{49}(oW)| = \binom{49}{2}$

Die Anzahl der günstigen Fälle ist also $\binom{6}{4} \cdot \binom{49}{2}$. Damit ist $P(4 \text{ richtige}) = \frac{\binom{4}{4} \cdot \binom{49}{2}}{\binom{49}{0}}$.

### Zufalls-Modelle

Verschiedene zufällige Vorgänge lassen sich mit Hilfe von sogenannten Urnenmodellen und Fächermodellen beschreiben. Die in diesen Modellen mit Hilfe der Kombinatorik erhaltenen Aussagen kann man sofort übertragen.

## Urnenmodell

Ein undurchsichtiges Gefäß (Urne) ist mit $n$ gleichartigen, von $1$ bis $n$ durchnummerierten Kugeln gefüllt. Der zufällige Vorgang besteht im Ziehen von $k$ Kugeln, wobei man 4 Arten des Ziehens unterscheidet:

1. Ziehen unter Beachtung der Reihenfolge mit Zurücklegen:
   Nach jedem Zug wird die Nummer der Kugel notiert und dann die Kugel zurückgelegt. Man erhält nach dem Ziehen ein $k$-Tupel von Zahlen, ein Element aus $Per_k^n(mW)$
2. Ziehen unter Beachtung der Reihenfolge ohne Zurücklegen:
   Nach dem Ziehen hat man ein $k$-Tupel von Zahlen, ein Element aus $Per_k^n(oW)$
3. Ziehen ohne Beachtung der Reihenfolge mit Zurücklegen
   Wie in 1., nur die Reihenfolge interessiert nicht. Man erhält ein Element von $Kom_k^n(mW)$
4. Ziehen ohne Beachtung der Reihenfolge ohne Zurücklegen
   Wie in 2., nur die Reihenfolge interessiert nicht. Man erhält ein Element von $Kom_k^n(oW)$

## Fächermodell

Man hat $n$ Fächer, durchnummeriert von $1$ bis $n$, und $k$ Objekte, die man in die Fächer verteilen will. Es gibt wieder 4 Arten:

1. Unterscheidbare Objekte und Mehrfachbelegung
   $Per_k^n(mW)$: Man erhält ein $k$-Tupel $(\alpha_1, \alpha_2, \dots, \alpha_n)$, wobei $a_j$ die Nummer des Faches

angibt. Indem das $j$-te Objekt unterbracht wird. Die $\alpha_j$ können gleiche Zahlen sein (Mehrfachbelegung)

2. Unterscheidbare Objekte und keine Mehrfachbelegung
   $Per_k^n(oW)$: Man hat ein k-Tupel wie in 1., aber alle Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_k$ sind verschiedene, jedes Fach tritt höchstens einmal auf ( $k \leqslant n$ ).
3. Nichtunterscheidbare Objekte und Mehrfachbelegung
   $Kom_k^n(mW)$: Man kann nach Belegung nur noch feststellen, wieviel Objekte in den Fächern sind. Die Zahlen $\alpha_1, \ldots, \alpha_k$ sind dann nach Größe geordnet: $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ mit $\alpha_1 \leqslant \alpha_2 \leqslant \ldots \leqslant \alpha_k$ (die Nummern der Fächer wo ein Objekt ist)
4. Nichtunterscheidbare Objekte und keine Mehrfachbelegung:
   $Kom_k^n(oW)$ : Wie in 3., aber $\alpha_1 < \alpha_2 < \ldots < \alpha_k$ ( $k \leqslant n$ )

Urnenmodelle und Fächermodelle sind Laplaceexperimente mit den jeweils angegebenen Ergebnismengen (für zufälliges Ziehen bzw. Verteilen).

## Abschätzungen

Für die Berechnung von Wahrscheinlichkeiten im Urnen- und Fächermodell sind zwei Abschätzungen nützlich.

### Abschätzung

Es gilt die Abschätzung $\frac{n!}{(n-r)! \cdot n^r} \approx \exp\left(-\frac{r \cdot (r-1)}{2 \cdot n}\right)$ für $r \ll n$. Der Fehler beträgt weniger als 3% bei $10 \cdot r \leqslant n$

Die Zahl $\frac{n!}{(n-r)! \cdot n^r}$ gibt die Wahrscheinlichkeit an, bei der zufälligen Ziehung von $r$-Kugeln aus einer Urne mit $n$ Kugeln (mit Zurücklegen und Reihenfolge) $r$ verschiedene zu ziehen.

#### Herleitung

$$\frac{|Per_k^n(oW)|}{|Per_k^n(mW)|} = \frac{n!}{(n-r)! \cdot n^r}$$

#### Beispiel

Sei $n=20$, $r=2$. Dann ist: $\frac{20!}{18! \cdot 20^2} = \frac{19}{20} = 0.95$ und $\exp\left(-\frac{2 \cdot 1}{2 \cdot 20}\right) \approx 0.951229$.

### Abschätzung: Häufigkeiten von Fixpunken

Die Zahl der $n$-Permutationen von $n$ Elementen (eigentliche Permutationen) mit genau $m$ Fixpunkten

$F_{n,m} = \frac{n!}{m!} \cdot \left(1 - \frac{1}{1!} + \ldots + \frac{(-1)^{n-m}}{(n-m)!}\right)$. Für $(n-m) \geqslant 5$ gilt: $F_{n,m} \approx \frac{n!}{m!} \cdot \frac{1}{e}$

**Beispiel**

Die Permutationen $(1,2,\ldots,n)$ und $(3,2,\ldots,n-1)$ haben genau einen Fixpunkt. Es kann aber sein, dass kein Fixpunkt existiert

$\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}$, $\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$ [Hier fehlt viel Text auf der Folie.]

**Beispiel**

$$e^x=\sum_{n=1}^{\infty}\left(\frac{x^k}{n!}\right),\quad x=-1$$

$$e^{-1}=\frac{1}{e}=\sum_{n=0}^{\infty}\left(\frac{(-1)^n}{n!}\right)$$

$$F_{3,1}=\frac{3!}{1!}\cdot\left(1-1+\frac{1}{2}\right)=3$$

**Beispiel**

Sei $n=3$, $m=1$. Hier gibt es 3 Permutationen mit einem Fixpunkt:

$\begin{pmatrix}1 & 2 & 3\\ 1 & 3 & 2\end{pmatrix}$, $\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{pmatrix}$, $\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix}$

**Beispiel: Test eines Weinkenners**

- Ein Mann behauptet, die verschiedenen Rotweinsorten eines Gebietes allein durch Trinken eines kleinen Schlucks „blind" zu erkennnen.
- Frage: Man lasse den Mann $n$ verschiedene Weinsorten blind testen. Wieviele davon muss er richtig erkennen, damit man sagen kann, er sei ein Weinkenner?
- Ein mathematisches Modell ist: Wir wählen ein Fächermodell mit $n$ Fächern, durchnummeriert von $1$ bis $n$ und $n$ Objekte, ebenfalls durchnummeriert von $1$ bis $n$. (Fächer entsprechen den $n$ Weinsorten, die er testen soll und die $n$ Objekte den Gläsern mit dem entsprechenden Wein.)

  Die möglichen Ergebnisse (einer Zuordnung) sind Elemente von $\Omega=Per_k^n(oW)$, mit $|\Omega|=n!$. Die richtige Zuordnung von $m$ Objekten bedeutet, es ist eine Permutation mit $m$ Fixpunkten. Die richtige Zuordnung von $m$ Objekten beduetet, es ist eine Permutation mit $m$ Fixpunkten. Die Wahrscheinlichkeit, $m$ Fixpunkte bei zufälliger Zuordnung zu erhalten, ist $P(m \text{ richtige})=\frac{F_{n,m}}{n!}$
- Konkret: Der Fall $n=10$, $m=4$ ergibt:

  $P(4 \text{ richtige})=\frac{F_{10,4}}{10!}\approx\frac{0.39}{4!}\approx 0.015=1.5\,\%$
  - Das bedeutet, dass ein Test mit 10 Weinsorten und 4 „Treffern" mit 98.5% Wahrscheinlichkeit ergibt, dass der Probant ein Weinkenner ist.

## 1.7 Reale zufällige Vorgänge und das mathematische Modell

- Als mathematisches Modell eines zufälligen Vorgangs wählen wir einen

Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ (Kolmogorov 1933). Zuerst hat man also die Ergebnismenge anzugeben, die Ereignisalgeba ist dann dadurch (meist) fixiert.

- Die Verteilung $P(\cdot)$ erhält man entweder aus theoretischen Überlegungen oder aus den „experimentellen Wahrscheinlichkeiten“ (den relativen Häufigkeiten)
- In dieses mathematische Modell lassen sich auch zufällige Vorgänge unterbringen, die nicht wiederholbar sind und|oder subjektive Wertungen zulassen. So wenn es um Risikoabschätzungen geht oder man im privaten Bereich Voraussagen über das Eintreffen von möglichen Ereignissen machen will.
  - Beispiele
    - Funktionsfähigkeit eines
      - chemischen Werkes,
      - einer neuen Rakete
      - eines Atomreaktors
    - Besuch einer Person
      - Tante Erna
      - Schwiegermutter
      - ...
- Das Problem jeweils ist die Ermittlung der Wahrscheinlichkeit.
- Die Wichtigste Aufgabe beim Aufstellen des mathematischen Modells ist die Festlegung der Ergebnismenge $\Omega$. Diese Festlegung ist aber nicht eindeutig. Die Wahl von $\Omega$ hängt von der „Beobachtungstiefe“ ab, die die bei der Beschreibung des zufälligen Vorgangs benötigt wird. [Beobachtungstiefe ist hier die Antwort auf die Frage, wieviele Merkmale in die Beobachtung mit eingehen sollen.]

**Beispiel: Ausfall eines Gerätes**

Ergebnisse seien die Zeiten, wann das Gerät ausfällt.

- Die Ergebnisse können zum Beispiel so dargestellt werden
  - diskrete Beschreibung in Tagen:
    - unendlich: $\Omega_1 = \mathbb{N}_0^+ = \{0, 1, 2, \ldots\}$
    - endlich: $\Omega_2 = \{0, 1, 2, \ldots, 10000\}$
  - stetige Beschreibung in Stunden
    - unendlich: $\Omega_3 = \mathbb{R}^+ = [0, \infty[$

**Beispiel: Krebserkrankungen**

Die Anzahl der Krebskranken in einer Region soll genauer erfasst werden in Abhängigkeit von einigen Parametern.
Der zufällige Vorgang sei hier: Es wird zufällig ein Kranker ausgewählt und dann ermittelt man seine Parameter-Werte und definiert dies als das Ergebnis.

- 1. Möglichkeit: Ein Parameter, nämlich das Geschlecht der Person.
  - In diesem Fall ist $\Omega_1 = \{M, W\}$
- 2. Möglichkeit: Zwei Parameter, nämlich Geschlecht und Altersgruppe der Person.
  - Man nehme etwa folgende Gruppen:
    - Gruppe 1: 0..20 Jahre
    - Gruppe 2: 20..20 Jahre

    - Gruppe 3: 40..20 Jahre
    - Gruppe 4: 60..20 Jahre
    - Gruppe 5: 80..$\infty$ Jahre
- In diesem Fall ist die Ergebnismenge gegeben durch 10 geordnete Paare: $\Omega_2=\{(M,k),(W,k) \mid k\in\{1,2,3,4,5\}\}$
- Das Ergebnis $(W,3)$ bedeutet, die zufällig herausgegriffene kranke Person ist weiblich und im Alter zwischen 40 Jahre und 60 Jahre.
- Hier liegt eine größere Beobachtungstiefe vor. Ob sie notwendig ist, hängt vom Ziel der Erfassung ab.

### Achtung

Die Anwendung von Wahrscheinlichkeitsaussagen (aus statistischen Ermittlungen) auf den Einzelfall (zum Beispiel auf eine einzelne Person) ist mit Vorsicht zu betrachten. „Die Statistik entfaltet ihre Kraft bei Anwendung auf „Viele“.“

**Beispiel Unglücksraben**

(aus Paul Witzlawik: „Anleitung zum Unglücklichsein“)

1. Bobby Joe Keesy
    - 1962: desertierte, stahl Flugzeug nach Kuba
    - 1964: zurück in die USA, Gefängnis
    - 1970: Geiselname in Amman
    - 1973: Kriegsgefangener in Vietnam
    - 1975: ermordet Vizekonsul
2. Mike Marga
    - 83 mal überfallen
    - 4 mal Auto gestohlen

# 2. Bedingte Wahrscheinlichkeit und Unabhängigkeit

## 2.1 Bedingte Wahrscheinlichkeit

- Oft ist es von Interesse zu wissen, wie sich das Eintreten eines bestimmten Ereignisses $B$ auf die Möglichkeit eines anderen Ereignisses $A$ auswirkt. Man sucht also nach der Wahrscheinlichkeit dafür, dass A eintritt, wenn die Teilinformation vorliegt, dass $B$ bereits eingetreten ist. Diese Wahrscheinlichkeit nennt man dann **bedingte Wahrscheinlichkeit**.
- In einfachen Fällen lässt sich die bedingte Wahrscheinlichkeit unmittelbar angeben. Im Rahmen des mathematischen Modells eines zufälligen Vorgangs müssen wir eine Definition geben, die sich an den „einfachen Fällen“ (zum Beispiel dem Laplaceexperiment) orientiert.

### Definition 1: bedingte Wahrscheinlichkeit

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $A, B\in F$ mit $P(B)>0$. Dann heißt $P(A|B):=\frac{P(A\cap B)}{P(B)}$ die **bedingte Wahrscheinlichkeit** des Ereignisses $A$ unter der Bedingung $B$.

**Bemerkung**

Anstatt „unter der Bedingung $B$ “ sagt man auch
- „unter der Voraussetzung $B$ “ oder
- „unter der Hypothese $B$ “.

**Bemerkung**

Man sagt:
- B **begünstigt** das Eintreten von $A$, wenn $P(A|B)>P(A)$.
- B **behindert** das Eintreten von $A$, wenn $P(A|B)<P(A)$.
- B **beeinflusst** das Eintreten von $A$ **nicht**, wenn $P(A|B)=P(A)$.

Alle drei Fälle können auftreten:
- $A\subset B$
- $(A\cap B)=\emptyset$
- [...]

## 2.2 Eigenschaften bedingter Wahrscheinlichkeiten

**Satz: Multiplikationsformel**

Eine unmittelbare Folgerung der Definition der bedingten Wahrscheinlichkeit ist die sogenannte **Multiplikationsfomel**: $P(A\cap B)=P(A|B)\cdot P(B)=P(B|A)\cdot P(A)$.
Sie ist von Nutzen, da es oft einfacher ist, $P(A|B)$ beziehungsweise $P(B|A)$ zu berechnen als $P(A\cap B)$. Dies geht aber auch allgemeiner:

**Satz 1: allgemeiner Multiplikationssatz**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum. Dann gilt für $n\in(\mathbb{N}\setminus\{0,1\})$ und $n$ Ereignisse $A_1, A_2, \ldots, A_n$ aus $F$ mit $P(A_1\cap A_2\cap\ldots\cap A_{n-1})>0$:
$P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cdot A_2)\cdot\ldots\cdot P(A_n|A_1\cap A_2\cap\ldots A_{n-1})$

Wenn man für ein Ereignis $B$ und ein vollständiges System von Ereignissen $(A_j)_j$ die bedingte Wahrscheinlichkeit $P(B|A_j)$ kennt, ist der folgende Satz nützlich:

Datum: 07.05.2003

**Satz 2: Formel der totalen Wahrscheinlichkeit**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $(A_j)_{j=1}^n$ ein vollständiger Satz von Ereignissen aus $F$ mit $P(A_j)>0$. Dann gilt:

$$\forall (B \in F): \left( P(B) = \sum_{j=1}^{n} \left( P(B|A_j) \cdot P(A_j) \right) \right)$$

**Beweis**

$B = \bigcup_{j=1}^{n} (B \cap A_j)$. $(B \cap A)_{j=1}^n$ sind paarweise disjunkt. Wir benutzen die Additivität für die Wahrscheinlichkeit: $P(B) = P\left( \bigcup_j (B \cap A_j) \right) = \sum_{j=1}^{n} \left( P(B \cap A_j) \right) = \sum_{j=1}^{n} \left( P(B|A_j) \cdot P(A_j) \right)$

Eine einfache Konsequenz dieser Formel ist das folgende Korollar. Es ist nützlich, wenn man die bedingte Wahrscheinlichkeiten $P(B|A_j)$ kennt, aber die „umgekehrten" Wahrscheinlichkeiten $P(A_j|B)$ sucht:

**Korollar: Satz von Bayes**

Es seien $(\Omega, F, P(\cdot))$, $B$, $(A_i)_{i=1}^n$ wie in Satz 2 und $P(B)>0$. Dann gilt:

$$\forall (A_i): \left( P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{k=1}^{n} \left( P(B|A_k) \cdot P(A_k) \right)} = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} \right)$$

**Beweis**

Es gilt: $P(A_j|B) \cdot P(B) = P(A \cap B) = P(B|A_j) \cdot P(A_j)$.

Daraus folgt: $P(A_j|B) = \dfrac{P(B|A_j) \cdot P(A_j)}{P(B)}$.

Das Einsetzen der Multiplikationsformel liefert die Behauptung.

## 2.2 Eigenschaften bedingter Wahrscheinlichkeiten

- Eine unmittelbare Folgerung der Definition der bedingten Wahrscheinlichkeit ist die sogenannte Multiplikationsformel:
  $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$
  Sie ist von Nutzen, da es oft einfacher ist, $P(A|B)$ beziehungsweise $P(B|A)$ zu berechnen als $P(A \cap B)$

**Beispiel: Defekte Sicherungen**

(Beispiel für die Anwendung des Multiplikationssatzes)

Es sei gegeben eine Schachtel mit 10 Sicherungen. Davon sind 4 defekt.

Sie entnehmen nacheinander zweimal eine Sicherung.
Wie groß ist die Wahrscheinlichkeit dafür, dass beide entnommenen Sicherungen ganz sind?

$A=\{\text{beide entnommenen Sicherungen sind ganz}\}$

$A_1=\{\text{1. entnommene Sicherung ist ganz}\}$

$A_2=\{\text{2. entnommene Sicherung ist ganz}\}$

Ergebnismenge $\Omega=\{(0,0),(0,1),(1,0),(1,1)\}$

- Elementarereignis 0: Sicherung ist ganz
- Elementarereignis 1: Sicherung ist ganz
- Der erste Eintrag des Tupels repräsentiert die erste Entnahme,
- der zweite Eintrag des Tupels repräsentiert die zweite Entnahme.

$A_1=\{(1,1),(1,0)\}$

$A_2=\{(1,1),(0,1)\}$

$P(A)=P(A_1\cap A_2)=P(A_2|A_1)\cdot P(A_1)$

- Wir berechnen $P(A_1)$ und $P(A_2|A_1)$ durch Rückführung auf das Laplace-Experiment:

$$P(A_1)=\frac{6}{10}\,,\quad P(A_2|A_1)=\frac{5}{9}$$

$$\Rightarrow P(A)=\frac{6}{10}\cdot\frac{5}{9}=\frac{1}{3}$$

**Beispiel:** **Test auf Krankheit**

Über den Test ist folgendes bekannt:
1. $p\cdot 100\,\%$ der Kranken werden durch den Test ermittelt.
2. $q\cdot 100\,\%$ der Gesunden werden durch den Test erkannt.
3. Die Krankheit ist mit $r\cdot 100\,\%$ in der Bevölkerung verbreitet.

**Frage**

Eine zufällig herausgegriffene Person wird dem Test unterworfen und als krank eingestuft. Wie groß ist die Wahrscheinlichkeit, dass die Person wirklich krank ist?

**Mathematisches Modell**

Als Ergebnismenge wählen wir $\Omega=\{(0,0),(0,1),(1,0),(1,1)\}$ , wobei
- die 1. Komponente angibt:
  - 0: nicht krank,
  - 1: krank,
- die 2. Komponente angibt:
  - 0: Test sagt „nicht krank“,
  - 1: Test sagt „krank“.

**Spezielle Ereignisse**

- $K=\{(1,0),(1,1)\}$ (Person ist krank)
- $N=\{(1,0),(0,0)\}$ (Negativer Test, sagt: „nicht krank“)
- Bekannt ist:

  - $P(K)=r$ ,
  - $P(\overline{N}|K)=p$ ,
  - $P(N|\overline{K})=q$
- Wir wollen wissen: $P(K|\overline{N})$ , also die Wahrscheinlichkeit dafür, dass die Person auch wirklich krank ist, wenn der Test positiv ist. Wir benutzen die Bayes-Formel: $P(K|\overline{N})=\frac{P(\overline{N}|K)\cdot P(K)}{P(\overline{N}|K)\cdot P(K)+P(\overline{N}|\overline{K})\cdot P(\overline{K})}=\frac{p\cdot r}{p\cdot r+(1-q)\cdot(1-r)}$
  mit
  - $P(\overline{N}|\overline{K})=1-P(N|\overline{K})=1-q$
  - $P(\overline{K})=1-P(K)=1-r$
- Mit $p=0.9$ , $q=0.99$ , $r=0.05$ folgt: $P(K|\overline{N})=0.83$
- Mit $p=0.9$ , $q=0.99$ , $r=0.01$ folgt: $P(K|\overline{N})=0.47$

**Satz**

Eine nützliche Formel ist:
Man habe zwei Ereignisse $A$ , $C$ und einen vollständigen Satz von Ereignissen $(K_j)_{j=1}^{n}$ . Dann gilt: $P(A|C)=\sum_{j=1}^{n}\left(P(A|C\cap K_j)\cdot P(K_j|C)\right)$ .
(Dieser Satz ist lediglich eine Ausdehnung der Formel der totalen Wahrscheinlichkeit.)

**Satz: Simpson-Paradoxon**

(nach E. H. Simpson, 1951)
Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $(K_j)_{j=1}^{n}$ ein vollständiger Satz von Ereignissen. Dann sagt man, dass Simpson Paradoxon liege vor, wenn gilt: $\left(\forall(j):\left(P(A|B\cap K_j)>P(A|\overline{B}\cap K_j)\right)\right)\wedge\left(P(A|B)<P(A|\overline{B})\right)$

Das heißt, das Zusammenlegen von Teilmengen $(K_j)$ kehrt die Präferenzen um.
- Die Möglichkeit dazu folgt aus der Formel oben, wegen des Einflusses der „Gewichtsfaktoren“ $P(K_j|C)$
  - Es gibt reichhaltige Literatur dazu, zum Beispiel: Krümer: „Denkste“.
  - Das Simpson-Paradoxon ist Ursache für viele Falschmelduungen in den Medien:
    - „Scheinbare Bevorzugung von Männern“
    - „Scheinbare Zunahme von Krebs“
    - ...

**Das Ziegenproblem**

Ein anderes „wichtiges Problem“ bei der bedingten Wahrscheinlichkeit ist das sogenannte Ziegenproblem. Man hat eine Show mit zwei Ziegen und einem Auto. (Buch dazu von [...]) [...]

## 2.3 Stochastische Unabhängigkeit

Die Unabhängigkeit von Ereignissen ist ein zentraler Begriff. Erfasen lässt sich die Nichtbeeinflussbarkeit von Ereignissen durch folgende Definition:

**Definition: stochastische unabhängig**

Es sei $(\Omega, F, P(\cdot))$ ein Wahrscheinlichkeitsraum und $A, B \in F$.
Dann heißt $A$ (**stochastisch**) **unabhängig** von $B$, wenn $P(A \cap B) = P(A) \cdot P(B)$.

Mit dem Multiplikationssatz folgt damit eine andere Möglichkeit, Unabhängigkeit zu definieren: $A$ heißt **stochastisch unabhängig** von $B$ dann, wenn im Fall $P(B) > 0$ gilt: $P(A|B) = P(A)$.

**Bemerkung**

Es gilt nicht zu verwechseln „Unabhängigkeit“ mit „Unvereinbarkeit“, also $A \cap B = \emptyset$

**Bemerkung**

Im speziellen Fall $B = \Omega$ ist $P(A \cap \Omega) = P(A) = P(A) \cdot P(\Omega)$. Also ist ein Ereignis $A$ immer unabhängig vom sicheren Ereignis $\Omega$.

**Beispiel: Mehrmaliger Würfelwurf**

Ein Würfel hat kein Gedächtnis. Also hängt das Ergebnis des $n$-ten Wurfes nicht ab von den Ergebnissen der $n-1$ Würfe vorher.

- $P(\text{„6“ im } n \text{ ten Wurf} | \text{keine „6“ in allen vorhergehenden Würfen}) = \frac{1}{6}$

Dies ist nicht zu verwechseln mit der Wahrscheinlichkeit, in $n$ Würfen wenigstens eine „6“ zu erhalten:

- $P(\text{wenigstens eine „6“ in } n \text{ Würfen}) = 1 - P(\text{keine „6“ in } n \text{ Würfen}) = 1 - \left(\frac{5}{6}\right)^n$
- $\lim_{n \to \infty} \left(P(\text{wenigstens eine „6“ in } n \text{ Würfen})\right) = 1$

**Beispiel: Zweimaliger Würfelwurf**

$A = \{\text{Augensumme ist gerade}\}$
$B = \{\text{1 Wurf ist ungerade Zahl}\}$
Sind $A$, $B$ unabhängig? Ja, da

- $P(A) = \frac{3 \cdot 3 + 3 \cdot 3}{36} = \frac{1}{2}$
- $P(B) = \frac{3 \cdot 6}{36} = \frac{18}{26} = \frac{1}{2}$

Daraus folgt: $P(A \cap B) = \frac{3 \cdot 3}{36} = \frac{1}{4} = P(A) \cdot P(B)$

**Satz: Satz 1**

Sind die Ereignisse $A$ und $B$ stochastisch unabhängig, so sind es auch die Ereignisse
- $\overline{B}$ und $B$
- $A$ und $\overline{B}$
- $\overline{A}$ und $\overline{B}$

**Beweis**

$B=(A\cap B)\cup(\overline{A}\cap B)$. $A\cap B$ ist disjunkt zu $\overline{A}\cap B$.
Damit folgt $P(B)=P(A\cap B)+P(\overline{A}\cap B)$ und weiter
$P(\overline{A}\cap B)=P(B)-P(A)\cdot P(B)=P(B)\cdot(1-P(A))=P(B)\cdot P(\overline{A})$.
Es folgt: $\overline{A}$ und $B$ sind unabhängig.
Und so weiter auch für die anderen Fälle.

Datum: 09.05.2003

[...abgestürzt...]
[...Seite 28!...]

## 2.4 Mehrstufige Zufallsexperimente

Oft besteht ein zufälliger Vorgang (Experiment) aus Teilvorgängen (Stufen), welche der Reihe nach ablaufen. Manchmal ist es zweckmäßig, diesen Vorgang als ein Ganzes zu betrachten und behandeln, und machmal ist es günstiger, ihn zu zerlegen.

**Beispiel:** ***n*-maliger Münzwurf**

**Beispiel:** **6 aus 49**

### A) Diskrete mehrstufige Experimente

Der zufällige Vorgang bestehe aus $n$ Stufen. Die Ergebnisse sind dann $n$-Tupel $\omega=(a_1,a_2,\ldots,a_n)$, wobei $a_j$ den Ausgang des $j$-ten Teilexperiments notiert.
Wenn $\Omega_i$ die Ergebnismenge des $j$-ten Teilexperiments ist, so ist das kartesische Produkt $\Omega:=\Omega_1\times\Omega_2\times\ldots\times\Omega_n=\{\omega=(a_1,a_2,\ldots,a_n)|\forall(a_j\in\Omega_j)\}$ die passende Ergebnismenge für das Gesamtexperiment.
Die Verteilung $P(\cdot)$ über $Potenzmenge(\Omega)$ ergibt sich im diskreten Fall aus der Wahrscheinlichkeitsdichte $\forall(\omega\in\Omega):\left(p(\omega)=P(\{\omega\})\right)$.

#### Satz

Betrachten wir die speziellen Ereignisse

- $A_1=\{a_1\}\times\Omega_2\times\Omega_3\ldots\times\Omega_{n-1}\times\Omega_n$,
- $A_2=\Omega_1\times\{a_2\}\times\Omega_3\ldots\times\Omega_{n-1}\times\Omega_n$,
- ...,
- $A_n=\Omega_1\times\Omega_2\times\Omega_3\ldots\times\Omega_{n-1}\times\{a_n\}$.

Dann ist $\{\omega\}=\{(a_1,\ldots,a_n)\}=A_1\cap A_2\cap A_3\cap\ldots\cap A_{n-1}\cap A_n$.
Nach dem Multiplikationssatz gilt dann
$p(\omega)=P(\{\omega\})=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cap A_2)\cdot\ldots\cdot P(A_n|A_1\cap A_2\cap\ldots\cap A_{n-1})$

#### Bemerkung

In dieser Formel treten die bedingten Wahrscheinlichkeiten dafür auf, dass im $j$-ten Teilversuch $a_j$ eintritt, wenn wir wissen, dass $A_1\cap A_2\cap\ldots\cap A_{j-1}=\{a_1\}\times\{a_2\}\times\ldots\times\{a_{j-1}\}\times\Omega_j\times\Omega_{j+1}\times\ldots\times\Omega_n$ bereits eingetreten ist. Anders gesagt: Wir wissen schon, dass die $j-1$ Teilversuche die Ergebnisse $a_1,a_2,\ldots,a_{j-1}$ liefern.

**Bemerkung: Übergangswahrscheinlichkeiten**

Die bedingten Wahrscheinlichkeiten $P(A_j | A_1 \cap A_2 \cap \ldots A_{j-1})$ nennt man dann auch Übergangswahrscheinlichkeiten von $(a_1, a_2, \ldots, a_{j-1})$ nach $a_j$, notiert als „ $p(a_j | a_1, a_2, \ldots, a_{j-1})$ “.

Mehrstufige Zufallsexperimente werden oft mit Hilfe von Diagrammen notiert und das Ergebnis $\omega = (a_1, a_2, \ldots, a_n)$ als Pfad in einem solchen Diagramm.

**Bemerkung: 1. Pfadregel**

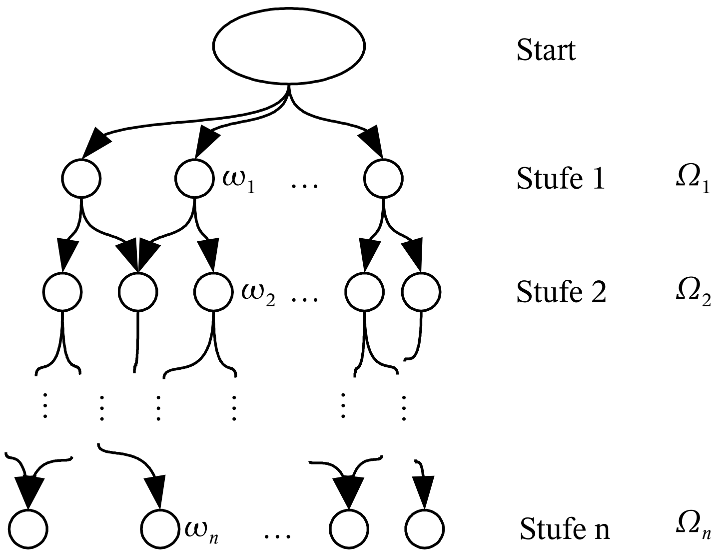
Der Multiplikationssatz entpuppt sich dann als **1. Pfadregel**: Die Wahrscheinlichkeit des Ereignisses $\omega$ ist das Produkt der Übergangswahrscheinlichkeiten längs des Pfades $(a_1, a_2, \ldots, a_n)$, also $p(\omega) = p(a_1) \cdot p(a_2 | a_1) \cdot p(a_3 | a_1, a_2) \cdot \ldots \cdot p(a_n | a_1, a_2, \ldots, a_{n-1})$.

**Bemerkung: 2. Pfadregel**

Die Addititvität der Wahrscheinlichkeiten ist dann die **2. Pfadregel**: Die Wahrscheinlichkeit des Ereignisses $A$ errechnet sich als Summe der Wahrscheinlichkeiten aller zu $A$ gehörenden pfade $P(A) = \sum_{\forall(\omega \in A)} (p(\omega))$.

**Diagramm für *n*-stufiges Zufallsexperiment**

- Ein Pfad über $\omega_1, \omega_2, \ldots, \omega_n$ ist das Ergebnis $(\omega_1, \omega_2, \ldots, \omega_n)$.




Die Wahrscheinlichkeit für den Pfad (also für ein solches Ergebnis) $(\omega_1, \omega_2, \ldots, \omega_n)$ ist $P(\{(\omega_1, \omega_2, \ldots, \omega_n)\}) = p(\omega_1) \cdot p(\omega_2 | \omega_1) \cdot p(\omega_3 | \omega_1, \omega_2) \cdot \ldots \cdot p(\omega_n | \omega_1, \omega_2, \ldots, \omega_{n-1})$.

**Beispiel: *n*-facher Münzwurf**

Sei $\forall(j \in \{1, 2 \ldots, n\}) : (\Omega_j = \{0,1\})$, mit:
- 0 bedeutet: „Wappen oben“,
- 1 bedeutet: „Zahl oben“.

Es sei $\forall(j \in \{1,2\ldots,n\}):\left(P_j(\{x\})=\frac{1}{2}\right)$.

Dann ist hier $P(\{(x_1,\ldots,x_n)\})=\left(\frac{1}{2}\right)^n$ und $\Omega=\Omega_1 \cdot \Omega_2 \cdot \ldots \cdot \Omega_n$.

**Beispiel: Anlage aus *n* Baugruppen**

Sei $\forall(j \in \{1,2\ldots,n\}):(\Omega_j=\{0,1\})$, mit:

- 0 bedeutet: „Das $j$-te Gerät ist ausgefallen.“,
- 1 bedeutet: „Das $j$-te Gerät funktioniert.“.

Es sei $\forall(j \in \{1,2\ldots,n\}):(P_j(\{1\})=p_j \in [0,1[)$.

Hier ist $P(\{(x_1,x_2,\ldots,x_n)\})=p_1^{x_1} \cdot (1-p_1)^{1-x_1} \cdot p_2^{x_2} \cdot (1-p_2)^{1-x_2} \cdot \ldots p_n^{x_n} \cdot (1-p_n)^{1-x_n}$.

- Ein Spezialfall ist, wenn $p_1=p_2=\ldots=p_n=p$ und $A=\{(1,1,\ldots,1)\}$, wenn also alle Baugruppen arbeiten. In diesem Fall arbeitet das ganze Gerät.
  - Für diesen Fall gilt: $P(A)=p^n$.
  - Mit $p=0.99$ folgt
    - für $n=10$: $P(A)\approx 0.90$,
    - für $n=100$: $P(A)\approx 0.35$,
    - für $n=1000$: $P(A)\approx 0.00003$.

[...Seite 30'' und 31 ausgelassen...]

Datum: 14.05.2003

[...abgestürzt...]
<import from=„Manfred Wollenberg">

## 2.5 Bernoulli-Schema

- Es wird ein mathematisches Modell zur Beschreibung einer Reihe unabhängiger Realisierungen eines zufälligen Vorgangs mit der Registrierung des Eintretens des interessierenden Ereignisses $A$. Dies soll erlauben, die Frage zu beantworten, wieviele Realisierungen notwendig sind, um die Wahrscheinlichkeit des Ereignisses mit vorgegebener Genauigkeit aus der relativen Häufigkeit zu bestimmen (Abschnitt 3.3)
- Dazu betrachten wir das **Bernoulli-Experiment**:
  $\Omega_0=\{0,1\}$, $P_0(\{1\})=p$, $P_0=(\{0\})=1$, wobei
  - $1$ steht für das Eintreten des interessierenden Ereignisses
  - $0$ steht für das Nichteintreten.

  $p:=P(A)$
- Jetzt betrachten wir eine Versuchsreihe von $n$ solchen Bernoulli-Experimenten. Der zugehörige Wahrscheinlichkeitsraum ergibt sich zu $(\Omega_n, F_n, P_n(\cdot))$ („Produktraum") mit
  - $\Omega_n:=\Omega_0\times\Omega_0\times\ldots\times\Omega_0=\{(x_1,x_2,\ldots,x_n)|\forall(j\in\{1,2,\ldots,n\}):(x_j\in\{0,1\})\}$ (Die Mege aller geordneten $n$-Tupel mit Werten aus $\{0,1\}$),
  - $F_n=Potenzmenge(\Omega_n)$ und
  - $$\begin{aligned}P_n(\{(x_1,x_2,\ldots,x_n)\})&=\prod_{j=1}^{n}(P_0(\{x_j\}))\\&=(p^{x_1}\cdot(1-p)^{1-x_1})\cdot(p^{x_2}\cdot(1-p)^{1-x_2})\cdot\ldots\cdot(p^{x_n}\cdot(1-p)^{1-x_n})\\&=p^k\cdot(1-p)^{n-k}\end{aligned}$$

    wenn $k$-mal die $1$ vorliegt (also $k$ mal das Ereignis $A$ in den $n$ Versuchen eintritt).

**Definition: Bernoulli-Schema**

Der Wahrscheinlichkeitsraum $(\Omega_n, Potenzmenge(\Omega_n), P_n(\cdot))$ heißt **Bernoulli-Schema** der Länge $n$ (mit dem Parameter $p$), auch genannt **Bernoulli-Kette**.

Interessiert man sich nicht für die Reihenfolge der n Einzelergebnisse $(x_1,x_2,\ldots,x_n)$, sondern nur, wie oft die $1$ dabei ist, also interessiert man sich nur für die Zahl $k$, so kann man eine gröbere Beschreibung einführen:

- Man wählt als Ergebnimenge jetzt $\Omega=\{0,1,2,\ldots,n\}$, also die „absoluten Häufigkeiten" von $A$ in den $n$ Versuchen.
- Zu jeder Zahl $k\in\Omega$ gehören nun alle $n$-Tupel $(x_1,x_2,\ldots,x_n)$, die alle mit der gleichen Wahrscheinlichkeit $p^k\cdot(1-p)^{n-k}$ auftreten. Die Anzahl dieser $n$-Tupel ist $|Kom_k^n(oW)|=\binom{n}{k}$.
  (Verteilung von $k$ Elementen auf $n$ Fächer)
- Folglich ist die Wahrscheinlichkeit für das Ereignis $\{k\}$:

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

**Definition: Binomialverteilung**

Die Verteilung $B_{n,p}$ heißt **Binomialverteilung**.
Diese Verteilung gibt uns also die Wahrscheinlichkeiten für die zufällige Anzahl der „Erfolge“ des Eintretens in $n$ unabhängigen 0,1-Vorgängen mit Erfolgswahrscheinlichkeit $p$ an.

**Beispiel: *n*-maliger Würfelwurf**

- Anzahl der „Sechsen“

**Beispiel: Familie mit vier Kindern**

- Zufällige Verteilung nach Geschlechter

# 3. Diskrete Zufallsgrößen

## 3.1 Definition von Zufallsgrößen

- Ist ein zufälliger Vorgang gegeben, so interessieren oft nicht die Ergebnisse selbst, sondern daraus gewonnene (oft einfachere) Größen.

**Beispiel: zweimaliger Würfelwurf**

Hier ist $\Omega=\{(i,j)|i,j\in\{1,2,3,4,5,6\}\}$. Oft ist nur die Augensumme $i+j$ von Interesse, also die Abbildung $X:\Omega\to\{2,3,\ldots,12\}$, $X((i,j)):=i+j$. $X$ nimmt die Werte $2,3,\ldots,12$ zufällig an.

**Beispiel: Bernoulli-Schema**

Hier wird n mal das gleiche Zufallsexperiment durchgeführt und registriert, ob das interessierende Ereignis $A$ in den $n$ Versuchen eingetreten ist oder nicht. Ergebnismenge ist $\Omega_n=\{\omega=(\omega_1,\omega_2,\ldots,\omega_n)|\forall(i\in\{1,2,\ldots,n\}):(w_j\in\{0,1\})\}$. $\omega=(\omega_1,\omega_2,\ldots,\omega_n)$ gibt also genau an, in welchen Versuchen $A$ eingetreten ist. Es interessiert aber oft nur die Zahl $\omega_1+\omega_2+\ldots+\omega_n$, das heißt: die Abbildung $X:\Omega_n\to\{0,1,\ldots,n\}$, $X((\omega_1,\omega_2,\ldots,\omega_n)):=\omega_1+\omega_2+\ldots+\omega_n$. $X$ repräsentiert die absolute Häufigkeit von $A$ in den $n$ Versuchen.

**Definition 1: relle Zufallsgröße**

Es sei $(\Omega,F,P(\cdot))$ ein Wahrscheinlichkeitsraum und es sei $X:\Omega\to\mathbb{R}$ eine reellwertige Abbildung. Dann heißt $X$ eine (**relle**) **Zufallsgröße**, wenn $\forall(t\in\mathbb{R}):(X^{-1}(]-\infty,t[)\in F)$.

**Bemerkung: Zufallsvariablen**

Manchmal spricht man auch von **Zufallsvariablen**.

**Bemerkung**

Die Forderung an $X$, dass die Urbilder von Intervallen $]-\infty,t[$ aus $\mathbb{R}$ Elemente von $F$ sind, impliziert dann auch, dass die Urbilde von allen offenen, abgeschlossenen und halboffenen Intervallen aus $\mathbb{R}$ auch Elemente von $F$ sind. Das bedeutet, dass $X$ eine „messbare Abbildung ist“.

**Bemerkung**

Für diskrete Wahrscheinlichkeitsräume mit $F = Potenzmenge(\Omega)$ ist jede Abbildung $X: \Omega \to \mathbb{R}$ eine Zufallsgröße. Denn die Forderung $X^{-1}(]-\infty, t[) \in F = Potenzmenge(\Omega)$ ist automatisch erfüllt. Zum Beispiel, wenn $\Omega \subseteq \mathbb{R}$ und $X$ stückweise stetig ist.

### Definition 2: Wahrscheinlichkeitsverteilung

Es sei $X$ eine Zufallsgröße über $(\Omega, F, P(\cdot))$. Dann heißt die Abbildung $B(\mathbb{R}) \ni \Delta \to P_X(\Delta) = P(X^{-1}(\Delta)) = P(\{\omega \in \Omega | X(\omega) \in \Delta\})$ Wahrscheinlichkeitsverteilung von X.

### Definition 2: Verteilungsfunktion der Zufallsgröße

Es sei $X$ eine Zufallsgröße über $(\Omega, F, P(\cdot))$. Dann heißt die Funktion $F_X: \mathbb{R} \to \mathbb{R}$, $F_x(t) := P(X < t) := P_X(]-\infty, t[)$ Verteilungsfunktion der Zufallsgröße $X$.

**Bemerkung**

Die Wahrscheinlichkeitsverteilung $P_X(\cdot)$ ordnet jeder Borelmenge $\Delta$ (Punkt, Intervall, Rechteck, Quader,...) von $\mathbb{R}$ die Wahrscheinlichkeit zu, dass $X$ Werte in $\Delta$ annimmt.

**Bemerkung**

Die Verteilungsfunktion $F_X(\cdot)$ gibt die Wahrscheinlichkeit dafür, dass $X$ Werte im Intervall $]-\infty, t[$ annimmt.
</import>

**Bemerkung**

Es gilt:

- $\forall(t\in\mathbb{R}):\left(F_x(t)\in[0,1]\right)$
- $\lim\limits_{t\to-\infty}\left(F_x(t)=0\right)$
- $\lim\limits_{t\to\infty}\left(F_x(t)=1\right)$
- $F_x(\cdot)$ ist
  - monoton wachsend und
  - linksseitig stetig

**Beweis**

für die linksseitige Stetigkeit.

**Behauptung**

$\forall(x\in\mathbb{R}):\left(\lim\limits_{y\to x-0}\left(F_x(y)\right)=F_x(x)\right)$

**Beweis**

Es genügt die Folgenstetigkeit zu zeigen, also die Folge $x_n\to x$ mit $x_1<x_2<\ldots<x$ und $\lim\limits_{n\to\infty}\left(F_x(x_n)=F(x)\right)$.

Wir haben
$$\begin{aligned}F_x(x)&=P_x\left(\left]-\infty,x\right[\right)\\&=P_x\left(\bigcup_{\forall(n)}\left(\left]-\infty,x_n\right[\right)\right)\\&=P\left(X^{-1}\left(\bigcup_{\forall(n)}\left(\left]-\infty,x_n\right[\right)\right)\right)\\&=P\left(\bigcup_{\forall(n)}\left(\underbrace{X^{-1}\left(\left]-\infty,x_n\right[\right)}_{A_n}\right)\right)\\&=P\left(\bigcup_{\forall(n)}\left(A_n\right)\right)\\&=\lim_{N\to\infty}\left(P\left(A_1\cup A_2\cup\ldots\cup A_n\right)\right)\\&=\lim_{N\to\infty}\left(P\left(A_N\right)\right)\\&=\lim_{N\to\infty}\left(P\left(X^{-1}\left(\left]-\infty,x_N\right[\right)\right)\right)\\&=\lim_{N\to\infty}\left(P_X\left(\left]-\infty,x_N\right[\right)\right)\\&=\lim_{N\to\infty}\left(F_X\left(x_n\right)\right)\end{aligned}$$

Daraus folgt die Stetigkeit von links.

**Beispiel**

Würfelwurf mit $\Omega=\{1,2,\ldots,6\}$, $F=Potenzmenge(\Omega)$, $P(\{h\})=\frac{1}{6}$

$X:\Omega\to\mathbb{R}$, $k\mapsto k$ (Identische Abbildung)

$k\in\Omega$, $P_X(\{k\})=P(\{k\})=\frac{1}{6}$, $\forall\left(y\in(\mathbb{R}\setminus\Omega)\right):\left(P_X(\{y\})=0\right)$

$F_X(x)=P\left(X^{-1}(]-\infty,x[)\right)=P(]-\infty,x[)\cap(1,2,\ldots,6)$

[Grafik]

**Beispiel:** **Gleichverteilung auf Intervall** $[0,1]=\Omega$

$F=B([0,1])$

$\forall(\Delta\in F):\left(P(\Delta)=|\Delta|\right)$

Zufallsgröße: $X:\Omega\to\mathbb{R}$, $\omega\to\omega^2\in\mathbb{R}$

Bild: $X(\Omega)=\Omega$

Verteilte Funktion $F_X(\cdot)$:

- 1. Fall: $x<0$
  Dann ist $X^{-1}(]-\infty,x[)=\emptyset$
  $\Rightarrow F_X(x)=P\left(X^{-1}(]-\infty,x[)\right)=P(\emptyset)=0$
- 2. Fall: $x\geqslant 1$
  Dann ist $X^{-1}(]-\infty,x[)=\Omega$
  $\Rightarrow F_X(x)=P\left(X^{-1}(]-\infty,x[)\right)=P(\Omega)=1$
- 3. Fall: $x\in[0,1]$.
  Dann ist $x=X(\sqrt{x})$
  $\Rightarrow F_X(x)=P\left(X^{-1}(]-\infty,x[)\right)=P([0,\sqrt{x}])=\sqrt{x}$

## Satz 1

Sei $X$ eine Zufallsgröße über dem Wahrscheinlichkeitsraum $(\Omega,F,P(\cdot))$. Dann ist $(\mathbb{R},B(\mathbb{R}),P_X(\cdot))$ ein Wahrscheinlichkeitsraum.

**Beweis**

Zu zeigen ist, dass $P_X(\cdot)$ eine Verteilung über $B(\mathbb{R})$ ist.

- Normierung:
  Die Normierung folgt aus $P_X(\mathbb{R})=P\left(X^{-1}(\mathbb{R})\right)=P(\Omega)=1$
- Positivität:
  Die Positivität folgt aus $\forall\left(\Delta\in B(\mathbb{R})\right):\left(P_X(\Delta)=P\left(X^{-1}(\Delta)\right)\geqslant 0\right)$
- $\sigma$-Additivität:
  Es sei $(\Delta_i)_i$ eine Folge paarweise disjunkter Elemente von $B(\mathbb{R})$. Dann ist
  $P_X\left(\bigcup_{\forall(j)}(\Delta_j)\right)=P\left(X^{-1}\left(\bigcup_{\forall(j)}(\Delta_j)\right)\right)=P\left(\bigcup_{\forall(j)}\left(X^{-1}(\Delta_j)\right)\right)=\sum_{\forall(j)}\left(P\left(X^{-1}(\Delta_j)\right)\right)=\sum_{\forall(j)}\left(P_X(\Delta_j)\right)$

#### Bemerkung

Wegen dieses Satzes kann man gleich mit der Verteilung $P_X(\cdot)$ über $(\mathbb{R}, B(\mathbb{R}))$ rechnen, und den ursprünglichen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ mit seiner vollen Information kann man vergessen.

#### Satz 2

Es sei $X$ eine Zufallsgröße über $(\Omega, F, P(\cdot))$. Es sei $g(\cdot)$ eine reellwertige stetige Funktion auf $\mathbb{R}$. Dann ist die Abbildung $g(X): \Omega \to \mathbb{R}$, $g(X)(\omega) := g(X(\omega))$ wieder eine Zufallsgröße über $(\Omega, F, P(\cdot))$.

#### Zufallsgrößen über den gleichen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$

Diese Zufallsgrößen können wie Funktionen betrachtet werden. Deshalb können sie
- addiert,
- multipliziert und
- skalar multipliziert

werden:
Seien $X$, $Y$ Zufallsgröße und $a, b \in \mathbb{R}$. Dann sind
- $a \cdot X + b \cdot Y$ mit $(a \cdot X + b \cdot Y)(\omega) := a \cdot X(\omega) + b \cdot Y(\omega)$
- $X \cdot Y$ mit $(X \cdot y)(\omega) := X(\omega) \cdot Y(\omega)$

wieder Zufallsgrößen über $(\Omega, F, P(\cdot))$.

## 3.2 Charakteristiken diskreter Zufallsgrößen

#### Definition 1: diskret

Eine Zufallsgröße $X$ über den Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ heißt **diskret**, wenn ihr Wertebereich $X(\Omega)$ in $\mathbb{R}$ **diskret** ist.

#### Bemerkung

Wenn die Ergebnismenge $\Omega$ des Wahrscheinlichkeitsraumes diskret ist, ist natürlich jede Abbildung $X: \Omega \to \mathbb{R}$ eine diskrete Zufallsgröße. [Sicher? Wie ist das mit der Abzählbarkeit rationaler Zahlen?]

#### Bemerkung

Eine diskrete Zufallsgröße $X$ ist vollständig charakterisiert durch ihre Werte $\forall (i \in I): (X_i)$ (wobei $I$ eine endliche oder abzählbar unendliche Indexmenge ist) und die dazugehörigen Wahrscheinlichkeiten $p_i := P_X(\{x_i\}) = P(X^{-1}(\{x_i\}))$. Also wird durch eine Verteilungstabelle

| | | | | | |
|---|---|---|---|---|---|
| $X_k$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ | $\dots$ |
| $p_k$ | $p_1$ | $p_2$ | $\dots$ | $p_k$ | $\dots$ |

die diskrete Zufalllsgröße vollständig charakterisiert

**Bemerkung**

Jede solche Tabelle mit reellen Zahlen $x_k$ und Zahlen $p_k \geqslant 0$ mit $\sum_{\forall(k)} : (p_k = 1)$ ist dann eine Verteilungstabelle, wenn man $\Omega := \{x_1, x_2, \ldots\}$ und $F = Potenzmenge(\Omega)$ wählt, $P(\cdot)$ ist gegeben duch $\forall(k) : (P(\{x_k\}) = p_k)$ und $X : \Omega \rightarrow \mathbb{R}$, $X(x_k) := x_k$ (identische Abbildung).

**Beispiel: Zweimaliger Würfelwurf**

$X$ repräsentiere die Augensumme.

**Beispiel: Münzwurf (beliebig oft)**

$X$ nimmt den Wert $k$ an, wenn erstmals $k$-ten Wurf [... ja was?..]

### Definition 2: Erwartungswert

Es sei $X$ eine diskrete Zufallsgröße mit Werten $(x_i)_i$ und Wahrscheinlichkeiten $(p_i)_i$. Dann heißt die Zahl $E(X)$ (wenn sie existiert, also $\sum_{\forall(i)} (|x_i| \cdot p_i) < \infty$ ) mit $E(X) := \sum_{\forall(i)} (x_i \cdot p_i)$ **Erwartungswert** von $X$.

**Beispiel**

[Grafik]

**Bemerkung**

$E(X)$ muss nicht in jedem Fall existieren, wenn es unendlich viele Werte $x_i$ gibt. Zum Beispiel sei $\forall(i \in \mathbb{N}) : \left( (x_i = 2^i) \wedge \left( p_i = \left( \frac{1}{2} \right)^i \right) \right)$.

**Bemerkung**

$E(X)$ ist so etwas wie der Mittelwert der Werte $x_i$ versehen mit den „Gewichten“ $p_i$.

**Bemerkung**

$E(X)$ muss nicht selbst zu den Werten $x_i$ von $X$ gehören. Beispielsweise ist beim Würfelwurf $E(X) = 3.5$.

**Bemerkung**

Wenn $X : \Omega \rightarrow \mathbb{R}$ und $(\Omega, Potenzmenge(\Omega), P(\cdot)).istDiskret()$, dann gilt auch $E(X) = \sum_{\forall(\omega \in \Omega)} (X(\omega) \cdot P(\{\omega\}))$

Datum: 23.05.2003

**Beweis**

von $E(x) = \sum_{\forall(\omega)} \left( X(\omega) \cdot P(\{\omega\}) \right)$

Es ist

$$\sum_{\forall(\omega\in\Omega)} \left( X(\omega) \cdot P(\{\omega\}) \right) = \sum_{\forall(i)} \left( \sum_{\forall(w):(X(\omega)=x_i)} \left( x_i \cdot P(\{\omega\}) \right) \right) = \sum_{\forall(i)} \left( x_i \cdot P(\{\omega\in\Omega | X(\omega)=x_i\}) \right) = \sum_{\forall(i)} (x_i \cdot p_i)$$

**Beispiel**

Die Zufalllsgröße $X$ sei zugeordnet dem Merkmal „Körpergröße bei der Geburt“. Hier diskretisieren wir die möglichen Werte in 2cm-Schritten. Die $p_i$ sind „experimentell“ als relative Häufigkeiten ermittelt.

| X in cm | ⩽ 40 | 42 | 44 | 46 | 48 | 50 | 52 | 54 | 56 | 58 | ⩾ 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 0,01 | 0,04 | 0,07 | 0,10 | 0,18 | 0,23 | 0,18 | 0,09 | 0,06 | 0,03 | 0,01 |

$\sum_{\forall(k)} (p_k) = 1$ Und $E(X) = \sum_{\forall(k)} (x_k \cdot p_k) \approx 49.82\,\text{cm}$ .

Hier ist $E(X)$ ein guter Wert für mittlere Größe [...]

## Satz 1

Es sei $X$ eine diskrete Zufallsgröße mit den Werten $(X_i)_i$ und den Wahrscheinlichkeiten $(p_i)_i$ . Es sei $g(\cdot)$ e ine stetige reellwertige Funktion auf ℝ . Konvergiert die Reihe $\sum_{\forall(i)} (g(x_i) \cdot p_i)$ absolut, so gilt:

$$E(g(X)) = \sum_{\forall(k)} (g(x_k) \cdot p_k)$$

**Beweis**

Sei der Spezialfall gegeben: $g(\cdot)$ ist streng monoton.

Dann gilt: $(g(x_k) = g(x_i)) \Leftrightarrow (x_k = x_i)$ .

Also folgt:

$$E(g(X)) = \sum_{\forall(k)} \left( g(x_k) \cdot P(g(X) = g(X_k)) \right) = \sum_{\forall(k)} \left( g(x_k) \cdot P(X = x_k) \right) = \sum_{\forall(k)} (g(x_k) \cdot p_k)$$

[Grafik]

**Bemerkung**

Weitere Informationen über Zufallsgrößen erhält man aus der „Streuung“ ihrer Werte.

## Definition: Varianz

Es sei $X$ eine diskrete Zufallsgröße

- mit den Werten $(x_i)_i$ und
- den Wahrscheinlichkeiten $(p_i)_i$ sowie
- dem Erwartungswert $E(X)$ .

Dann heißt die Zahl $V(X)$ (wenn sie existiert) mit
$V(X):=E\left(\left(X-E(X)\right)^2\right)=\sum_{\forall(k)}\left(\left(x_k-E(X)\right)^2\cdot p_k\right)$ die **Varianz**.

Die Varianz von $X$ ist der Erwartungswert der Zufallsgröße $g(X)$ mit $g(x)=(x-E(X))^{2.}$

**Definition: Standardabweichung**

Die Zahl $\sigma(x)=\sqrt{V(x)}$ heißt **Standardabweichung**.

**Satz**

Die Varianz von $X$ ist der Erwartungswert der Zufallsgröße $g(X)$ mit $g(X)=(X-E(X))^2$.

**Bemerkung**

Die Varianz muss nicht existieren für eine Zufallsgröße $X$ mit unendlich vielen Werten.

**Bemerkung**

Die Standardabweichung ist ein Maß dafür, wie die mittlere Abweichung vom Erwartungswert ist. Ist $\sigma(X)$ klein, so ist es vernünftig von $E(X)$ als mittleren Wert von $X$ zu sprechen.

**Bemerkung: Standardisierung der Zufallsgröße**

Für eine Zufallsgröße $X$ mit bekannten $E(X)$ und $V(X)$ führt man oft eine andere „standardisierte“ Zufallsgröße $\widetilde{X}$ ein:

$$X\rightarrow\widetilde{X}:=\frac{1}{\sigma\cdot(X)}\cdot(X-E(X))$$

Dann gilt: $(E(\widetilde{X})=0)\wedge(V(\widetilde{X})=1)$
Diese Prozedur wird **Standardisierung der Zufallsgröße** genannt.

**Vorgehensweise**

Wie erhält man $E(\widetilde{X})=0$ ?
$\widetilde{X}$ hat die Werte $\forall(i):\left(\widetilde{X}_i=\frac{1}{\sigma(X)}\cdot(X_i-E(X))\right)$.

Also ist

$$\begin{aligned}
E(\widetilde{X}) &= \sum_{\forall(i)}(\widetilde{X}_i\cdot p_i)\\
&= \sum_{\forall(i)}\left(\frac{1}{\sigma(X)}\cdot(X_i-E(X))\cdot p_i\right)\\
&= \frac{1}{\sigma(X)}\cdot\left(\sum_{\forall(i)}(x_i\cdot p_i)-\sum_{\forall(i)}(E(x)\cdot p_i)\right)\\
&= \frac{1}{\sigma(X)}\cdot(E(X)-E(X))\\
&= 0
\end{aligned}$$

## Satz 2

Es sei $X$ eine Zufallsgröße mit Erwartungswert $E(X)$ und Varianz $V(X)$. Dann

- $\exists(E(X^2))$ und
- es gilt $V(X)=E(X^2)-(E(X))^2$

### Beweis

$$
\begin{aligned}
V(X) &= \sum_{\forall(k)} \left( (X_k - E(X))^2 \cdot p_k \right) \\
&= \sum_{\forall(k)} \left( \left( {x_k}^2 - 2\cdot x_k \cdot E(X) + (E(X))^2 \right) \cdot p_k \right) \\
&= \sum_{\forall(k)} \left( {x_k}^2 \cdot p_k \right) - 2\cdot E(X) \cdot \sum_{\forall(k)} \left( x_k \cdot p_k \right) + (E(X))^2 \cdot \sum_{\forall(k)} (p_k) \\
&= E(X^2) - 2\cdot E(X) \cdot E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

## Satz 3

Dia Varianz der Zufallsgröße $X$ ist genau dann gleich $0$, wenn $X$ nur <u>einen</u> Wert mit der Wahrscheinlichkeit $1$ annimmt.

### Beweis

$V(X)=\sum_{\forall(k)} \left( (X_k - E(X))^2 \cdot p_k \right) \geqslant 0$ und alle Summanden sind nicht negativ. Es muss wenigstens ein $X_j$ auftreten mit der Wahrscheinlichkeit $p_j > 0$. Das ergibt $(X_j - E(X)) = 0$.. Nun sind aber alle möglichen anderen $x_k \neq x_j$, also $(X_k - E(X))^2 > 0$, also $p_k = 0$. Das bedeutet aber, es gibt nur einen Wert $x_j$ mit Wahrscheinlichkeit $p_j = 1$.

## Eigenschaften

von $E(X)$ und $V(X)$:
Seien $X, Y, X_1, \ldots, X_N$ Zufallsgröße über $(\Omega, F, P(\cdot))$ und $a, b \in \mathbb{R}$.
Dann gilt:

1. $E(X_1 + X_2 + \ldots + X_N) = E(X_1) + E(X_2) + \ldots + E(X_n)$ (Additivität der Erwartungswerte)

### Beweis

Es genügt zu zeigen für den Fall $N=2$:

$$
\begin{aligned}
E(X_1 + X_2) &= \sum_{\forall(\omega \in \Omega)} \left( (X_1(\omega) + X_2(\omega)) \cdot P(\{\omega\}) \right) \\
&= \sum_{\forall(\omega \in \Omega)} \left( X_1(\omega) \cdot P(\{\omega\}) \right) + \sum_{\forall(\omega \in \Omega)} \left( X_2(\omega) \cdot P(¿) \right) \\
&= E(X_1) + E(X_2)
\end{aligned}
$$

2. $E(a \cdot X + b) = a \cdot E(X) + b$ (Linearität der Erwartungswerte)
3. Wenn $X \leqslant Y$ (das heißt: $\forall(\omega \in \mathbb{R}) : (X(\omega) \leqslant Y(\omega))$ ) gilt, dann gilt auch $E(X) \leqslant E(Y)$ (Monotonie der Erwartungswerte)

4. $V(a\cdot X)=a^2\cdot V(X)$
5. $V(X+b)=V(X)$

## 3.3 Spezielle diskrete Verteilungen

**Definition: diskrete Verteilung**

Verteilungen $P_X(\cdot)$, die zu diskreten Zufallsgrößen $X$ gehören, heißen **diskrete Verteilung**en. Sie sind gegeben durch die Wahrscheinlichkeiten $p_k=p(x_k)=P(X=x_k)$ zu den Werten $x_k$ von $X$.

**Satz**

Verschiedene Zufallsgrößen können gleiche Verteilungen besitzen.

**Satz**

Eine diskrete Verteilung ist durch eine Folge von Zahlen $(P_k)_{k\in J}$ mit $(\forall(k\in J):(p_k\geqslant 0))\wedge\left(\sum_{\forall(k\in J)}(p_k)=1\right)$ gegeben.

**Definition: gleichverteilt, gleichmäßige Verteilung, Laplace-Verteilung**

Eine diskrete Zufallsgröße $X$, die die Werte $\forall(k\in\{1,2,\ldots,n\}):(X_k)$ mit den Wahrschenlichkeiten $\forall(k):\left(p_k=\frac{1}{n}\right)$ annimmt, heißt **gleichverteilt**. Man sagt, $X$ besitzt eine **gleichmäßige Verteilung** (**Laplace-Verteilung**).

**Satz**

Wenn $X$ gleichverteilt ist, dann ist $E(X)=\frac{1}{n}\cdot\sum_{k=1}^{n}(x_k)$ das arithmetische Mittel der Werte von $X$.

**Definition: binomialverteilt, Binomialverteilung**

Eine diskrete Zufallsgröße $X$, die die Werte $\forall(k\in\{1,2,\ldots,n\}):(X_k)$ mit den Wahrscheinlichkeiten $\forall(k):\left(p_k=B_{n,p}(k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}\right)$ annimmt, heißt **binomialverteilt** mit den Parametern $n$ und $p$. Man sagt, $X$ besitzt eine **Binomialverteilung**.

**Satz**

Für den Fall der Binomialverteilung gilt: $E(X)=n\cdot p$.

**Beweis**

Wir definieren $\varphi(x):=(x\cdot p+(1-p))^n=\sum_{m=0}^{n}\left(x^m\cdot\binom{n}{m}\cdot p^m\cdot(1-p)^{n-m}\right)$ mit dem Binomischen Satz. Das ergibt:

$$\varphi'(x)=n\cdot p\cdot(x\cdot p+(1-p))^{n-1}=\sum_{m=0}^{n}\left(m\cdot x^{m-1}\cdot\binom{n}{m}\cdot p^m\cdot(1-p)^{n-m}\right)$$

Für $x=1$ folgt aber:

$$n\cdot p=\sum_{m=0}^{n}\left(m\cdot\binom{n}{m}\cdot p^m\cdot(1-p)^{n-m}\right)=\sum_{m=0}^{n}\left(m\cdot B_{n,p}(\{m\})\right)=E(X)$$

### Satz

Für den Fall der Binomialverteilung gilt: $V(X)=n\cdot p\cdot(1-p)$

**Beispiel: Bernoulli-Schema**

(mit absoluten und relativen Häufigkeiten)
Wir haben eine $n$-malige Wiederholung eines zufälligen Vorgangs. Beobachtet wird, wie oft das Ereignis $A$ eintritt. $P(A)=p$. Dann ist die Wahrscheinlichkeit für das $k$-malige Eintreten in den $n$ Versuchen gegeben durch $B_{n,p}(k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}$. $k$ ist die absolute Häufigkeit.

$Y_{(n)}$ sei die zugehörige Zufallsgröße mit den Werten $0,1,\ldots,n$. Es ist also $P(Y_{(n)}=k)=p_k=B_{n,p}(k)$, das heißt: $Y_{(n)}$ ist binomialverteilt.

$X_{(n)}:=\frac{1}{n}\cdot Y_{(n)}$ beschreibt dann die relative Häufigkeit. Damit gilt: $X_{(n)}$ hat die Werte $\frac{k}{n}$ mit Wahrscheinlichkeit $p_k$.

$E(Y_{(n)})=n\cdot p$ und $V(Y_{(n)})=n\cdot p\cdot(1-p)$, sowie $E(X_{(n)})=E\left(\frac{1}{n}\cdot Y_{(n)}\right)=p$,

$V(X_{(n)})=V\left(\frac{1}{n}\cdot Y_{(n)}\right)=\frac{1}{n^2}\cdot V(Y_{(n)})=\frac{p\cdot(1-p)}{n}$ oder $\sigma(X_{(n)})=\sqrt{\frac{p\cdot(1-p)}{n}}$

### Deutung

Die relative Häufigkeit strebt mit wachsender Zahl $n$ gegen die Wahrscheinlichkeit $p=P(A)$ des beobachteten Ereignisses. Die Abweichungen davon werden immer kleiner.

**Beispiel: Ausschussanteil in Stichprobe aus der laufenden Produktion**

Einer laufenden Produktion wird eine Stichprobe von $n$ Geräten entnommen. Es sei $p$ die Wahrscheinlichkeit, dass ein entnommenes Gerät defekt ist. Die nacheinander erfolgende Entnahme der $n$ Geräte wird gedeutet als Bernouilli-Schema von $n$ Experimenten mit interessierendem Ereignis $A=\{\text{Entnommenes Gerät ist defekt}\}$ und $P(A)=p$.

- Die Zufallsgröße $X$ sei hier diejenige, die die Anzahl der defekten Geräte in der Stichprobe angibt, also $x_k=k\in\{0,1,\ldots,n\}$.
  Dann ist $P(X=k)=p_k=B_{n,p}(k)$.
  Die Anzahl der Ausschussgeräte ist binomialverteilt.

**Beispiel: Stichprobe aus der laufenden Produktion**

- Der Hersteller gibt an, 5% der produzierten Ware seien Ausschuss. Sie entnehmen der laufenden Produktion eine Stichprobe vom Umfang $n=10$. Zwei der Geräte seien defekt: $k=2$.
  Kann man der Angabe des Herstellers trauen?
- Berechnung der Wahrscheinlichkeit für 2 defekte Geräte in einr Stichprobe mit $n=10$. $X$ ist binomialverteilt, also $P(X=2)=B_{10,0.05}(2)=\binom{10}{2}\cdot(0.05)^2\cdot(0.95)^2=45\cdot(0.05)^2\cdot(0.95)^2=0.075$
- Also ist 7.5% die Wahrscheinlichkeit dafür, dass 2 defekte Geräte in der Stichprobe sind. Vermutlich ist die Aussage des Herstellers falsch.

**Beispiel: Ausschussanteil in Stichprobe, die Gesamtanzahl der Geräte ist fest**

Wir haben $N$ Geräte, davon seien $M$ defekt. Wir nehmen eine Stichprobe vom Umfang $n\leqslant N$.
Die Zufallsgröße $X$ beschreibt die Zahl der defekten Geräte in der Stichprobe, $\forall(k\in\{0,1,\ldots,n\}):(x_k=k)$, wobei $(k\leqslant M)\wedge(n-k\leqslant N-M)$

- Die Gesamtzahl der Stichproben vom Umfang $n$ ist $\binom{N}{n}$
- Die Anzahl der Möglichkeiten, $k$ defekte Geräte in der Stichprobe zu haben, ergibt sich aus $\binom{M}{k}\cdot\binom{N-M}{n-k}$
- Also ist die Wahrscheinlichkeit dafür, $k$ defekte Geräte in der Stichprobe zu haben, der Quotient $P_X(\{k\})=P(X=k)=\dfrac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$

**Definition: hypergeometrisch verteilt, hypergeometrische Verteilung**

Es sei $X$ eine diskrete Zufallsgröße. Für $X$ existieren natürliche Zahlen $N,M,n$ mit $(M\leqslant N)\wedge(n\leqslant N)$, sodass die Werte von $X$ die ganze Zahlen $k$ sind mit $(0\leqslant k\leqslant n)\wedge(k\leqslant M)\wedge(n-k\leqslant N-m)$.
Dann heißt X **hypergeometrisch verteilt**, wenn gilt:

$$p_k=P(X=k)=H_{N,M,n}(k)=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$$

$H_{N,M,n}(\cdot)$ heißt **hypergeometrische Verteilung**.

Datum: 28.05.2003

**Eigenschaften**

Wenn X hypergeometrisch verteilt ist, dann gilt:
- $E(X)=n\cdot p$
- $V(X)=n\cdot p\cdot(1-p)\cdot\frac{N-n}{N-1}$ mit $p:=\frac{M}{N}$

**Bemerkung**

Hypergeometrische Verteilung und Binomial-Verteilung sind bei statistischen Qualitätskontrollen, Umfrageauswertungen von Bedeutung

## Urnenmodell

Gegeben seien
- $N$ Kugeln
- davon $M$ schwarze Kugeln,
- also $N-M$ weiße Kugeln.

Die Wahrscheinlichkeit, beim Herausgreifen eine schwarze Kugel herauszugreifen, ist $p=\frac{M}{N}$.

Betrachten wir nun den Fall, dass $n$ Kugeln gezogen werden (es gilt $n\leqslant N$) (Wir machen eine „Stichprobe“ vom Umfang $n$)
- 1. Fall mit jeweiligem Zurücklegen:
  Dann liegt Binomialverteilung vor für die Zufallsgröße $X$, die die Anzahl der schwarzen Kugeln ( $k$ ) in der Stichprobe angibt.
- 2. Fall ohne Zurücklegen:
  Dann liegt hypergeometrische Verteilung vor.

## Beispiel

Eine Warenlieferung enthalte 10 Geräte. Davon seien 2 Geräte defekt. Sie entnehmen dieser Lieferung 3 Geräte. Wie wahrscheinlich ist es, dass Sie
- genau ein beziehungsweise
- höchstens ein

defektes Gerät unter diesen drei haben?

- Wir haben also eine Stichprobe vom Umfang $n=3$. Die Zufalllsgröße $X$ nehme als Werte die Zahl der defekten Geräte in der Stichprobe an.
- $X$ ist hypergeometrisch verteilt. Es interessieren
  - $P(X=1)$ und
  - $P(X\leqslant 1)=P(X=0)+P(X=1)$
- $P(X=1)=\frac{\binom{2}{1}\cdot\binom{8}{2}}{\binom{10}{3}}=\frac{2\cdot 28}{120}=\frac{7}{15}\approx 0.4\overline{6}$

- $P(X=0)=\dfrac{\binom{2}{0}\cdot\binom{8}{3}}{\binom{10}{3}}=\dfrac{1\cdot 56}{120}=\dfrac{7}{15}\approx 0.4\overline{6}$
- Damit ist $P(X\leqslant 1)=\dfrac{14}{15}\approx 0.9\overline{3}$. Dies ist also ein sehr wahrscheinliches Resultat.

**Definition 4: Poisson-verteilt, Poisson-Verteilung**

Es sei $X$ eine diskrete Zufalllsgröße mit den Werten $\forall(k\in\mathbb{N}_0^+):(x_k=k)$. Es sei ein $\lambda\in\mathbb{R}^+$ gegeben. Die Wahrscheinlichkeiten $\forall(k):(p_k=P(X=x_k))$ seien gegeben durch

$$\forall(k):\left(p_k=\frac{\lambda^k}{k!}\cdot e^{-\lambda}\right)$$

Dann heißt $X$ **Poisson-verteilt** mit dem Parameter $\lambda$. Die Verteilung heißt **Poisson-Verteilung**.

**Eigenschaften: Erwartungswert**

Wenn $X$ Poisson-verteilt mit dem Parameter $\lambda$ ist, dann gilt:
- $E(X)=\lambda$

**Beweis**

$$\begin{aligned}E(X)&=\sum_{k=0}^{\infty}(x_k\cdot p_k)\\&=\sum_{k=0}^{\infty}\left(k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}\right)\\&=e^{-\lambda}\cdot\sum_{k=1}^{\infty}\left(\frac{\lambda^k}{(k-1)!}\right)\\&=\lambda\cdot e^{-\lambda}\cdot\sum_{k=1}^{\infty}\left(\frac{\lambda^{k-1}}{(k-1)!}\right)\\&=\lambda\cdot e^{-\lambda}\cdot\sum_{j=1}^{\infty}\left(\frac{\lambda^j}{j!}\right)\\&=\lambda\cdot e^{-\lambda}\cdot e^{\lambda}\\&=\lambda\end{aligned}$$

**Eigenschaften: Varianz**

Wenn $X$ Poisson-verteilt mit dem Parameter $\lambda$ ist, dann gilt:
- $V(X)=\lambda$

**Bemerkung**

Die Poisson-Verteilung heißt auch „Verteilung der seltenen Ereignisse“. Sie beschreibt oft die Anzahl $k$ des Auftretens eines Ereignisses bei großer Zahl von Versuchen mit sehr kleiner Erfolgswahrscheinlichkeit $p$ des einzelnen Ereignisses.

**Bemerkung**

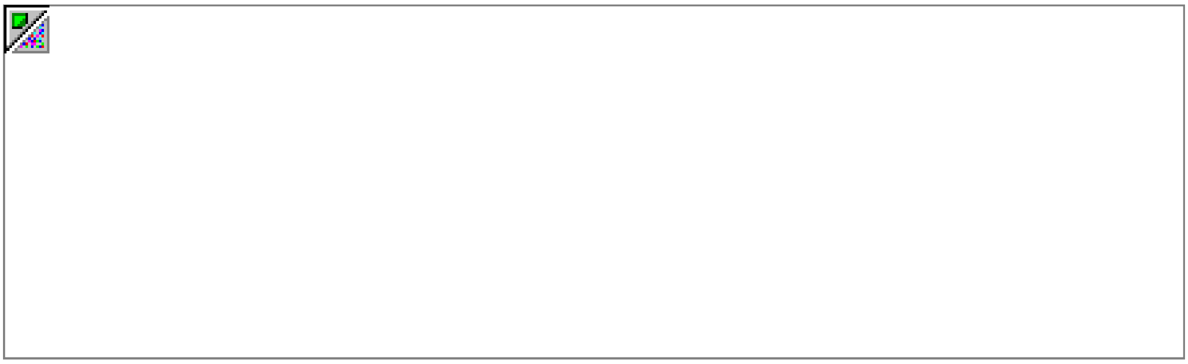
Die Poisson-Verteilung ist der „Grenzwert“ der Binomialverteilung und hypergeometrischen Verteilung.

**Bild der Poissonverteilung**

Siehe rechts.

### Zusammenhänge zwischen Binomial-Verteilung, hypergeometrischer Verteilung und Poisson-Verteilung

- Die hypergeometrische Verteilung $H_{N,M,n}(\cdot)$ ist oft recht schwierig zu berechnen. Für $N \gg n$ und $M \gg k$ kann man sie durch die Binomialverteilung approximieren:
  $H_{N,M,n}(\{k\}) \approx B_{n,p}(\{k\})$ mit $p := \frac{M}{N}$
  - Genauer: Es gilt: $\lim\limits_{N\to\infty} \left(H_{N,p\cdot N,n}(\{k\})\right) = B_{n,p}(\{k\})$
    - Beweis:



$$H_{N,M,n}(\{k\})=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$$

$$=\frac{M!\cdot(N-M)!\cdot n!\cdot(N-n)!}{k!\cdot(M-k)\cdot(n-k)!\cdot(N-M-n+k)!\cdot N!}$$

$$=\frac{\overbrace{\Big(\underbrace{M\cdot(M-1)\cdot\ldots\cdot(M-k+1)}_{k\text{ Faktoren}}\Big)\cdot\Big(\underbrace{(N-m)\cdot(N-m-1)\cdot\ldots\cdot(N-M-n+k-1)}_{n-k\text{ Faktoren}}\Big)}^{n+(n-k)=n\text{ Faktoren}}}{\underbrace{(N\cdot(N-1)\cdot\ldots\cdot(N-n+1))}_{n\text{ Faktoren}}}\cdot\binom{n}{k}$$

$$=\binom{n}{k}\cdot\frac{\frac{M}{N}\cdot\left(\frac{M}{N}-\frac{1}{N}\right)\cdot\ldots\cdot\left(\frac{M}{N}-\frac{k+1}{N}\right)\cdot\left(1-\frac{M}{N}\right)\cdot\ldots\cdot\left(1-\frac{M+n-k+1}{N}\right)}{1\cdot\left(1-\frac{1}{N}\right)\cdot\ldots\cdot\left(1-\frac{n-1}{N}\right)}$$

$$=\binom{n}{k}\cdot\frac{p\cdot\left(p-\frac{1}{N}\right)\cdot\ldots\cdot\left(p-\frac{k+1}{N}\right)\cdot(1-p)\cdot\ldots\cdot\left(1-p+\frac{n-k+1}{N}\right)}{1\cdot\left(1-\frac{1}{N}\right)\cdot\ldots\cdot\left(1-\frac{n-1}{N}\right)}$$

$$\lim_{N\to\infty}\left(H_{N,M,n}(\{k\})\right)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k}=B_{n,p}(\{k\})$$

- Zahlenbeispiel:
  - Sei $p=\frac{M}{N}=0.25$ , $k=5$ , $n=20$ .
  - Dann ist $B_{n,p}(\{k\})=0.20233$

| $N$ | 100 | 1000 | 2000 | 10000 | 100000 |
|---|---|---|---|---|---|
| $H_{n,p\cdot N,n}(\{k\})$ | $\underline{0.2}2601$ | $\underline{0.20}438$ | $\underline{0.20}335$ | $\underline{0.202}53$ | $\underline{0.2023}5$ |

**Approximation der Binomialverteilung durch Poissonverteilung**

- Die Binomialverteilung $B_{n,p}(\cdot)$ kann für kleine $p$ und große $n$ durch die Poissonverteilung $P_{0,\lambda}(\cdot)$ approximiert werden:
  - $B_{n,p}(\{k\})=P_{0\lambda}(\{k\})$ mit $\lambda=n\cdot p$ .
    Diese Approximation ist für $n\cdot p<10$ und $n>1500\cdot p$ recht gut.
  - Präziser gilt: $\lim_{n\to\infty}\left(B_{n,\frac{\lambda}{n}}(\{k\})\right)=P_{0\lambda}(\{k\})$ .

**Beispiel: Platzreservierung**

(zum Beispiel bei Flugzeugen)
Im Mittel werden etwa 94% der gebuchten Plätze auch genutzt.

- Vorschlag zur Umsatzsteigerung: Für 100 Plätze werden 103 Buchungen vorgenommen.
- Frage: Mit welcher Wahrscheinlichkeit erhalten alle Fahrgäste mit Platzkarten, die die Reise antreten, auch einen gebuchten Platz?
- Die Zufallsgröße $X$ sei die Zahl derjenigen, die Platzkarten haben, aber nicht kommen. $X$

ist eigentlich binomialverteilt.

- Die Binomialverteilung ist aber unhandlich zu berechnen. Ist die Poisson-Verteilung nutzbar?
  - 1. Bedingung: $n \cdot p = 103 \cdot 0.06 < 10$ ist erfüllt.
  - 2. Bedingung: $n = 103 > 1500 \cdot p = 90$ ist auch erfüllt.

  Also liefert die Poisson-Verteilung eine gute Näherung für diesen Binomialverteilung.
- $E(X) = \lambda = p \cdot n = 6.18$
- Gesucht ist: $P(X \geqslant 3) = 1 - P(X=0) - P(X=1) - P(X=2)$
  (denn wenn mindestens 3 Fahrgäste mit Buchung nicht kommen, erhalten alle (anderen) ihre gebuchten Plätze)
  - $P(X=0) = e^{-\lambda} \approx 0.00207$
  - $P(X=1) = \lambda \cdot e^{-\lambda} = 6.18 \cdot e^{-\lambda} \approx 6.18 \cdot 0.00207$
  - $P(X=2) = \frac{6.18^2}{2} \cdot e^{-\lambda}$
  - Damit ist $P(X \geqslant 3) \approx 94.56\,\%$
- Bei 102 Buchungen auf 100 Plätze wird das Wahrscheinlichkeit noch besser, nämlich $\approx 98.5\,\%$

## 3.4 Zufallsvektoren

Mehrere Zufallsgrößen über einen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$ fasst man oft zusammen zu einem Zufallsvektor.

### Definition 1: Zufallsvektor

Es seien für $X_1, X_2, \ldots, X_n$ Zufallsgrößen über den Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$. Dann heißt die Abbildung
$X := (X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$, $\omega \mapsto X(\omega) := (X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ **Zufallsvektor**.

### Definition: gemeinsame Verteilung, Verteilung des Zufallsvektors

Die Verteilung $\forall (\Delta \in B(\mathbb{R}^n)) : (P_X(\Delta) := P(X^{-1}(\Delta)))$ heißt dann

- die **gemeinsame Verteilung** der $X_1, \ldots, X_n$ oder
- die **Verteilung des Zufallsvektors** $X$.

### Bemerkungen zu den Übungsaufgaben

- 4. Aufgabe: ist nur fakultativ, bringt Zusatzpunkte

Datum: 04.06.2003

- Meistens werden wir der Einfachheit halber nur den Fall $n=2$ (also $\boldsymbol{X}=(X_1, X_2)$) betrachten und oft die Schreibweise $\boldsymbol{X}=(X, Y)$ benutzen.

### Definition 2: Randverteilung, Marginalverteilung

Es seien $X, Y$ zwei Zufallsgrößen über $(\Omega, F, P(\cdot))$ und $(X, Y)$ der zugehörige Zufallsvektor, sowie $P_{(X,Y)}(\cdot)$ die zugehörige gemeinsame Verteilung.
Dann heißen die Verteilungen
- $P_1(\cdot): B(\mathbb{R}) \ni \Delta_1 \to P_1(\Delta_1) := P_{(X,Y)}(\Delta_1 \times \mathbb{R})$
- $P_2(\cdot): B(\mathbb{R}) \ni \Delta_2 \to P_2(\Delta_2) := P_{(X,Y)}(\mathbb{R} \times \Delta_1)$

die **Randverteilung**en (**Marginalverteilung**en) der gemeinsamen Verteilung.

#### Bemerkung

Man sieht sofort, dass
- $\forall(\Delta_1 \in B(\mathbb{R})) : (P_1(\Delta_1) = P_X(\Delta_1))$
- $\forall(\Delta_2 \in B(\mathbb{R})) : (P_2(\Delta_2) = P_X(\Delta_2))$

#### Beispiel: Geburt eines Kindes

Sei
- die Zufallsgröße $X$ die Körperlänge $X \in \{25\,cm, 26\,cm, 27\,cm, \ldots, 75\,cm\}$
- die Zufallsgröße $Y$ die Masse $Y \in \{1.0\,kg, 1.1\,kg, 1.2\,kg, \ldots, 6.0\,kg\}$

(Die Zufallsgrößen seien diskret.)
Aus den relativen Häufigkeiten erhält man dann die gemeinsame Verteilung:
- $P_{X,Y}(x_i, y_i) = P((X=x_i) \wedge (Y=y_i)) = p_{i,j}$ (Doppelfolge)

Die Randverteilungen ergeben sich hier zu:
- $\forall(k) : \left(P(X=x_k) := \sum_{j=1}^{M} (p_{k,j}) =: p_k\right)$
- $\forall(j) : \left(P(Y=y_j) = \sum_{k=1}^{N} (p_{k,j}) =: q_j\right)$

Die Vermutung ist: Die Werte $x_i$ und $y_j$ sind in irgendeiner Form miteinander gekoppelt oder voneinander abhängig. Ein längeres Kind wird wahrscheinlich auch eine größere Masse haben.

### Beispiel: Produktion von rechteckigen Metallblechen

Diese Bleche sollen die Maße $l_0$ (für die Länge) und $h_0$ (für die Höhe) haben.
Die Zufallsgröße $X$ beschreibt die zufällige Abweichung der tatsächlichen Länge von $l_0$.
Die Zufallsgröße $Y$ beschreibt die zufällige Abweichung der tatsächlichen Höhe von $h_0$.
Wir beschreiben die Abweichung stetig:
$P(X \in \Delta_1, Y \in \Delta_2) = P_{(X,Y)}(\Delta_1 \times \Delta_2)$
Zum Beispiel:
- $\Delta_1 = [-1, 1] \cdot mm$

- $\Delta_2=[-2,3]\cdot mm$

$P_{X,Y}$ wird experimentell ermittelt über relative Häufigkeiten.
Die Vermutung ist: $X$ und $Y$ sind nicht gekoppelt.

### Definition 3: stochastisch unabhängig

Es sei $(X,Y)$ ein Zufallsvektor. Dann heißen die Zufallsgrößen $X$ und $Y$ **stochastisch unabhängig** voneinander, wenn $\forall(\Delta_1,\Delta_2\in B(\mathbb{R})):(P_{(X,Y)}(\Delta_1\times\Delta_2)=P_1(\Delta_1)\cdot P_2(\Delta_2))$ mit den Randverteilungen $P_1(\cdot)=P_X(\cdot)$ und $P_2(\cdot)=P_Y(\cdot)$.

Wenn die Zufallsgrößen $X$ und $Y$ (voneinander) stochastisch unabhängig sind, dann ist die gemeinsame Verteilung $P_{(X,Y)}(\cdot)$ also durch die Verteilungen $P_X(\cdot)$ und $P_Y(\cdot)$ der beiden Zufallsgrößen $X$ und $Y$ bereits festgelegt.

### Definition: diskreter Zufallsvektor

Wenn die Zufallsgrößen $X$ und $Y$ des Zufallsvektors $(X,Y)$ diskret sind, dann heißt $(X,Y)$ **diskreter Zufallsvektor**, wenn [wirklich?]

- $(X,Y)$ hat die Werte $(x_l,y_j)$ mit den Wahrscheinlichkeiten $p_{l,j}$,
- $X$ hat die Werte $x_l$ mit den Wahrscheinlichkeiten $p_l$,
- $Y$ hat die Werte $y_j$ mit den Wahrscheinlichkeiten $q_j$ und
- es gilt stochastische Unabhängigkeit, wenn $\forall(l):\forall(j):(p_{l,j}=p_k\cdot p_j)$

Aus $\Delta_1\cap\left(\bigcup_k(\{x_h\})\right)=\{x_l\}$ und $\Delta_2\cap\left(\bigcup_r(\{y_r\})\right)=\{y_j\}$ folgt
$P_{(X,Y)}(\Delta_1\times\Delta_2)=P(X=x_l,Y=y_j)=P_{(X,Y)}(\{x_l\}\times\{y_j\})=:p_{l,j}$

#### Verteilungstabelle für $(X,Y)$

(wenn $(X,Y)$ diskret ist)

| | $X_1$ | $X_2$ | $\dots$ | $X_N$ | $Y$ |
|---|---|---|---|---|---|
| $Y_1$ | $p_{11}$ | $p_{21}$ | $\dots$ | $p_{N1}$ | $q_1$ |
| $Y_2$ | $p_{12}$ | $p_{22}$ | $\dots$ | $p_{N2}$ | $q_1$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $Y_M$ | $p_{1M}$ | $p_{2M}$ | $\dots$ | $p_{NM}$ | $q_M$ |
| $X$ | $p_1$ | $p_2$ | $\dots$ | $p_N$ | |

Es gilt: $q_j=\sum_j(p_{lj})$
Die untere Zeile der Tabelle stellt die Randverteilung von $(X,Y)$ für $X$ dar.
Die rechte Spalte der Tabelle stellt die Randverteilung von $(X,Y)$ für $Y$ dar.

#### Beispiel: Zweimaliger Würfelwurf

Sei

- $X$ die Zufallsgröße über Augenzahl des ersten Wurfes
- $Y$ die Zufallsgröße über Augenzahl des zweiten Wurfes

- $(X, Y)$ hat die Werte $(l, j)$ mit den Wahrscheinlichkeiten $p_{l,j}=\frac{1}{36}$,
- $X$ hat die Werte $l \in \{1,2,3,4,5,6\}$ mit den Wahrscheinlichkeiten $p_l=\frac{1}{6}$,
- $Y$ hat die Werte $j \in \{1,2,3,4,5,6\}$ mit den Wahrscheinlichkeiten $q_j=\frac{1}{6}$.
- Da $\forall(j):\forall(l):(p_{l,j}=p_l \cdot q_j)$, gilt stochastische Unabhängigkeit.

**Beispiel:** **Summe von 2 Würfelwürfen**

Sei $Z$ die Summe der beiden Augenzahlen: $Z=X+Y$. $Z$ hat die Werte $Z_l \in \{2,3,\ldots,12\}$ mit den Wahrscheinlichkeiten

- $\forall(l \in \{1,2,3,4,5,6\}):\left(r_l=\frac{l}{36}\right)$
- $\forall(l \in \{7,8,9,10,11\}):\left(r_l=\frac{12-l}{36}\right)$

| Z \ X | 1 | 2 | 3 | 4 | 5 | 6 | $q$ |
|---|---|---|---|---|---|---|---|
| 2 | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | 0 | $\frac{1}{36}$ |
| 3 | $\frac{1}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | 0 | $\frac{2}{36}$ |
| 4 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | 0 | 0 | 0 | $\frac{3}{36}$ |
| 5 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | 0 | 0 | $\frac{4}{36}$ |
| 6 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | 0 | $\frac{5}{36}$ |
| 7 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{6}{36}$ |
| 8 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{5}{36}$ |
| 9 | 0 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{4}{36}$ |
| 10 | 0 | 0 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{3}{36}$ |
| 11 | 0 | 0 | 0 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{2}{36}$ |
| 12 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ |

$X$ und $Z$ sind nicht stochastisch unabhängig, denn $p_{31}=0$ , aber $p_3 \cdot q_1 = \frac{1}{6} \cdot \frac{1}{36} \neq 0$

### Satz 1

Es seien $X$ und $Y$ zwei Zufallsgrößen, die stochastisch unabhängig sind.
Dann gilt: $E(X \cdot Y) = E(X) \cdot E(Y)$

#### Beweis

$$E(X \cdot Y) = \sum_{i} , j(x_i \cdot y_j \cdot p_{i,j}) = \sum_{i} , j(x_i \cdot y_j \cdot p_i \cdot q_j) = \left(\sum_{i} (x_i \cdot p_i)\right) \cdot \left(\sum_{j} (y_j \cdot q_j)\right) = E(X) \cdot E(Y)$$

### Satz 1

Es seien $X$ und $Y$ zwei Zufallsgrößen, die stochastisch unabhängig sind.
Dann gilt: $V(X+Y) = V(X) + V(Y)$

#### Beweis

$$\begin{aligned} V(X+Y) &= E\left((X+Y)^2\right) - \left(E(X+Y)\right)^2 \\ &= E\left(X^2\right) + E\left(Y^2\right) + 2 \cdot E(X \cdot Y) - \left(E(X) + E(Y)\right)^2 \\ &= E\left(X^2\right) + E\left(Y^2\right) + 2 \cdot E(Y) \cdot E(Y) - \left(E(X)\right)^2 - \left(E(Y)\right)^2 - 2 \cdot E(X) \cdot E(Y) \\ &= \left(E\left(X^2\right) + \left(E(X X)\right)^2\right) + \left(E\left(Y^2\right) - \left(E(Y)\right)^2\right) \end{aligned}$$

### Definition 3: Kovarianz

Es sei $(X, Y)$ ein Zufallsvektor. Dann heißt $C(X, Y) := E\left((X - E(X)) \cdot (Y - E(Y))\right)$ die **Kovarianz** von $(X, Y)$ .

### Definition 3: Korrelationskoeffizient

Es sei $(X, Y)$ ein Zufallsvektor. Dann heißt $\rho(X, Y) := \frac{C(X, Y)}{\sigma(X) \cdot \sigma(Y)}$ der **Korrelationskoeffizient** von $(X, Y)$ .

### Satz 2

Es sei $(X, Y)$ ein Zufallsvektor. Wenn $X$ und $Y$ stochastisch unabhängig sind, dann sind sie auch unkorreliert, das heißt: $C(X, Y) = 0$ .

#### Beweis

Mit $X$ und $Y$ sind auch $\tilde{X} = X - E(X)$ und $\tilde{Y} = Y - E(Y)$ stochastisch unabhängig, also
$C(X, Y) = \underbrace{E(\tilde{X} \cdot \tilde{Y}) = E(\tilde{X}) \cdot E(\tilde{Y})}_{\text{Satz 1}} = 0$

**Interpretation von $C(X,Y)$ und $\rho(X,Y)$**

- Positive Kovarianz bedeutet, dass eine Tendenz besteht, dass $X$ für diejenige $\omega$ größere Werte annimt, für die auch $Y$ das tut.
- Negative Kovarianz bedeutet, dass eine Tendenz besteht, dass $X$ für diejenige $\omega$ größere Werte annimt, für die auch $Y$ kleinere Werte annimt das tut.
- Sprechweise:
  - Bei negativer Kovarianz ist die Zufallsgröße **negativ korreliert**.
  - Bei positiver Kovarianz ist die Zufallsgröße **positiv korreliert**.
  - Bei einer Kovarianz von 0 ist die Zufallsgröße **nicht korreliert**.

$$C(X,Y)=E((X-E(X))\cdot(Y-E(Y)))=\sum_{\omega\in\Omega}((X(\omega)-E(X))\cdot(Y(\omega)-E(Y))\cdot P(\{\omega\}))>0$$

**Bemerkung**

Korrelation hat nichts mit Kausalität zu tun. Also $\rho(X,Y)\neq 0$ bedeutet nicht, $X$ impliziert $Y$ oder $Y$ impliziert $X$.
Korrelation bedeutet etwas wie eine lineare Abhängigkeit zwischen $X$ und $Y$.
Wenn man ein Diagramm macht, auf dem $X$ über $Y$ aufgetragen ist und jeder Ort die Helligkeit entsprechend der Wahrscheinlichkeit des zugehörigen Zufallsvektors hat, dann entspricht Korrelation, dass es eine „Punktewolke“ gibt, die sich etwa entlang einer Gerade erstreckt.

Datum: 03.07.2003

<import from= „Manfred Wollenberg“>

### Satz 3

Es sei $(X, Y)$ ein Zufallsvektor. Dann gilt:

1. $|\rho(X, Y)| \leqslant 1$
2. $(\exists(a):\exists(b):(Y = a \cdot X + b)) \Leftrightarrow (|\rho(X, Y)| = 1)$
   [Der Korrelationskoeffizient ist genau dann bei 1, wenn $X$ und $Y$ voneinander linear abhängig sind.]

#### Beweis

Es sei $Y = a \cdot X + b$. Dann ist

$$\begin{aligned} C(X, Y) &= C(X, a \cdot X + b) \\ &= E((X - E(X)) \cdot (a \cdot X + b - a \cdot E(X) - b)) \\ &= E(a \cdot (X - E(X)) \cdot (X - E(X))) \\ &= a \cdot V(x) \end{aligned}$$

Folglich ist $|\rho(X, Y)| = \dfrac{|C(X, Y)|}{\sigma(X) \cdot \sigma(Y)} = \dfrac{|a| \cdot V(X)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}} = \dfrac{|a| \cdot V(X)}{\sqrt{V(X)} \cdot \sqrt{a^2 \cdot V(X)}} = 1$

### Satz 4

Es seien $X$, $Y$ Zufallsgrößen über einen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$. Dann gilt:

$C(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$

### Satz 4

Es seien $X$, $Y$ Zufallsgrößen über einen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$. Dann gilt:

$\forall(a):\forall(b):\forall(c):\forall(d):(C(a \cdot X + b, c \cdot Y + d) = a \cdot c \cdot C(X, Y))$

### Satz 4

Es seien $X$, $Y$, $X_1, X_2, \ldots, X_n$ Zufallsgrößen über einen Wahrscheinlichkeitsraum $(\Omega, F, P(\cdot))$. Dann gilt:

$$V(X_1 + X_2 + \ldots + X_n) = \sum_{j=1}^{n} (V(X_i)) + \sum_{\forall(i):\forall(j):(i \neq j)} (C(X_i, X_j))$$

#### Beispiel

Sei $X = Y$. Es gilt $C(X, X) = E((X - E(X)) \cdot (Y - E(Y))) = V(X)$.

$X$ ist mit sich selbst maximal korreliert.

#### Beispiel

Es seien $X$ und $Y$ zwei Zufallsgrößen. Wir bilden neue Zufallsgrößen $X_1 = X + Y$ und

$X_2 = X - Y$. Dann ist
$$\begin{aligned} C(X_1, X_2) &= C(X+Y, X-Y) \\ &= C(X,X) + C(Y,X) - C(X,Y) - C(Y,Y) \\ &= C(X,X) - C(Y,Y) \\ &= V(X) - V(Y) \end{aligned}$$

Es gibt also genau dann <u>keine</u> Korrelation, wenn die Varianzen gleich sind.

**Beispiel**

Die Zufallsgröße X habe die Werte $-1$, $0$, $1$ mit den Wahrscheinlichkeiten $p_1 = p_2 = p_3 = \frac{1}{3}$.

Dann hat die Zufallsgröße $X^2$ die Werte $0$, $1$ mit den Wahrscheinlichkeiten $q_1 = \frac{1}{3}$, $q_2 = \frac{2}{3}$.

$X$ und $X^2$ sind natürlich stochastisch abhängig. Aber

$C(X, X^2) = E(X \cdot X^2) - E(X) \cdot E(X^2) = E(X^3) - \sum_{i=1}^{3} \left( (x_i)^3 \cdot p_i \right) = 0$.

| $X^2 \setminus X$ | -1 | 0 | 1 | $X^2$ |
|---|---|---|---|---|
| 0 | 0 | $\frac{1}{3}$ | 0 | $\frac{1}{3}$ |
| 1 | $\frac{1}{3}$ | 0 | $\frac{1}{3}$ | $\frac{2}{3}$ |
| $X$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 1 |

</import>
[...nachträglich abgeschrieben von Christian Wagner...]

# 4. Stetige Zufallsgrößen

## 4.1 Stetige Zufallsgrößen und Verteilungsdichten

**Definition: stetige Zufallsgröße**

Eine Zufallsgröße $X$ über $(\Omega, F, P(\cdot))$ heißt **stetig** (oder stetig verteilt), wenn es eine Funktion $f_X : \mathbb{R} \to \mathbb{R}$ gibt, für die gilt:

1. $f_X$ ist nicht negativ
2. $f_X$ ist integrabel
3. $\forall (a<b) : \left( P(a \leqslant X < b) = \int_a^b (f_X(y) dy) \right)$
4. $\int_{-\infty}^{\infty} (f_X(y) dy) = 1$

**Bemerkung**

Die 4. Forderung folgt (fast) aus der 3. Forderung. In der Praxis ist $f_X(\cdot)$ stetig oder stückweise stetig.

**Bemerkung: Verteilungsdichte, Dichte**

$f_X(\cdot)$ heißt **Verteilungsdichte** (oder **Wahrscheinlichkeitsdichte**) der Zufallsgröße $X$. Anstelle der diskrete Zufallsgröße treten hier die „stetigen" Wahrscheinlichkeiten $f_X(x)$ auf. Es ist hier $\forall(a\in\mathbb{R}):(P(X=a)=0)$.

**Bemerkung**

Die Vorgabe der Verteilungsdichte $f_X(\cdot)$ bestimmt vollständig die stetige Zufallsgröße. Sie bestimmt die Wahrscheinlichkeiten, mit denen $X$ Werte in Intervallen annimmt. Jede nichtnegative integrable Funktion $f:\mathbb{R}\rightarrow\mathbb{R}$ mit der 4. Forderung definiert eine Dichte.

**Bemerkung**

Für eine stetige Zufallsgröße $X$ ist natürlich $\forall(a\in\mathbb{R}):(P(X=a)=0)$ (Die Wahrscheinlichkeit, dass die Zufallsgröße einen Wert in einem unendlich kleinen Intervall annimmt, ist 0), im Gegensatz zu diskreten Zufalllsgrößen.

**Beispiel**

Es sei $f_X(y)=\begin{cases} 0 & \text{für } y\notin[0,1] \\ \frac{2}{3}\cdot(1+y) & \text{für } y\in[0,1] \end{cases}$

[Grafik]
Es gilt:

- $\forall(y):(f_X(y)\geqslant 0)$
- $\int\limits_{-\infty}^{\infty}(f_X(y)\cdot dy)=\int\limits_0^1\left(\frac{2}{3}\cdot(1+y)\cdot dy\right)=1$

**Beispiel**

Es sei $a\in\mathbb{R}^+$ und $f_X(y)=\begin{cases} 0 & \text{für } y\in]-\infty,0] \\ \alpha\cdot e^{-\alpha\cdot y} & \text{für } y\in]0,\infty[ \end{cases}$. Dies ist tatsächlich eine Dichte, denn

$$\int\limits_{-\infty}^{\infty}(f_x(y)dy)=\int\limits_0^{\infty}(\alpha\cdot e^{-a\cdot y}dy)=\alpha\cdot\int\limits_0^{\infty}(e^{-a\cdot y}dy)=\alpha\cdot\left[\frac{1}{-\alpha}\cdot e^{-\alpha\cdot y}\right]_0^{\infty}=1$$

## 4.1.1 Approximation durch diskrete Zufallsgrößen

Gegeben sei eine stetige Zufallsgröße $X$ mit Dichte $f_X(\cdot)$ auf einem Intervall $[a,b]\subseteq\mathbb{R}$.
Gesucht ist die diskrete Zufallsgröße $X_{diskret}$, die $X$ approximiert.
Die Vorgehensweise ist folgendermaßen:

- Partitionierung von $[a,b]$ in $N$ Abschnitte $\Delta_1,\Delta_2,\ldots,\Delta_N$:
  - $\bigcup\limits_{j=\{1,2,\ldots,N\}}(\Delta_j)=[a,b]$
  - $\forall(i,j\in\{1,2,\ldots,N\}):((i\neq j)\Rightarrow((\Delta_i\cap\Delta_j)=\emptyset))$
- Zuordnung der Abschnitte zu den sie identifizierenden Stellen:
  - $\forall(j\in\{1,2,\ldots,N\}):(x_j\in\Delta_j)$

- Berechnung der Wahrscheinlichkeiten: $\forall\left(j\in\{1,2,\dots,N\}\right):\left(p_j=\int\limits_{y\in\Delta_j}\left(f_X(y)dy\right)\right)$
- $X_{diskret}$ hat dann die Werte $x_1,x_2,\dots,x_N$ mit den Wahrscheinlichkeiten $p_1,p_2,\dots,p_N$
- Es gilt dann: $\sum_{j=1}^{N}\left(p_j\right)=1$

## 4.2 charakteristische Größen für stetige Zufallsgrößen

**Definition: Erwartungswert**

Sei $X$ eine stetige Zufallsgröße mit der Dichte $f_X(\cdot)$. Dann heißt die Zahl (falls sie existiert) $E(X):=\int_{-\infty}^{\infty}\left(y\cdot f_X(y)dy\right)$ **Erwartungswert** von $X$.

**Definition: Varianz**

Sei $X$ eine stetige Zufallsgröße mit der Dichte $f_X(\cdot)$. Dann heißt die Zahl (falls sie existiert) $V(X):=\int_{-\infty}^{\infty}\left(\left(y-E(X)\right)^2\cdot f_X(y)dy\right)$ **Varianz** von $X$.

**Bemerkung: Existenz**

Die Größen Erwartungswert und Varianz müssen nicht existieren.

**Beispiel**

Es sei zum Beispiel $f_X(y)=\begin{cases}0 & \text{wenn } y\leqslant 1\\ y-2 & \text{wenn } y>1\end{cases}$. Dann folgt:

$$E(X)=\int_{-\infty}^{\infty}\left(y\cdot f_X(y)dy\right)=\int_{1}^{\infty}\left(y\cdot y^{-2}dy\right)=\int_{1}^{\infty}\left(\frac{1}{y}dy\right)=\infty$$

**Bemerkung: Kompatibilität**

Es gelten alle Formeln wie für diskrete Zufallsgröße, so insbesondere:
- $E(X_1+\dots+X_n)=E(X_1)+\dots+E(X_n)$
- $V(X)=E(X^2)-\left(E(X)\right)^2$
- $E(a\cdot X+b)=a\cdot E(X)+b$
- $V(a\cdot X+b)=a^2\cdot V(X)$

Auch die Definitionen für
- $C(X,Y)$,
- $\rho(X,Y)$,
- stochastische Unabhängigkeit
- ...

bleiben richtig.

**Bemerkung: Verteilungsfunktion**

Die **Verteilungsfunktion** $F_X(x):=P(X<x)$ ist gegeben durch $F_X(x)=\int_{-\infty}^{x}\left(f_X(y)dy\right)$.

### 4.2.1 Eigenschaften von $F_X(\cdot)$

Die **Verteilungsfunktion** einer stetigen Zufallsgröße hat die folgenden Eigenschaften:

- Sie ist stetig auf $\mathbb{R}$.
- Sie ist monoton wachsend.
- Sie gibt die Wahrscheinlichkeit dafür an, dass $X$ Werte im Intervall $]-\infty, x[$ annimmt.

## 4.3 spezielle stetige Verteilungen

### 4.3.1 Gleichverteilung

**Definition: Gleichverteilung**

Eine Zufallsgröße $X$ hat eine **gleichmäßig stetige Verteilung** („**Gleichverteilung**") auf dem Intervall $[a,b]$, wenn für die Dichte gilt: $f_X(Y)=\begin{cases} 0 & \text{für } x\notin[a,b] \\ \frac{1}{b-a} & \text{für } x\notin[a,b] \end{cases}$.

**Eigenschaften: Erwartungswert**

Für die Gleichverteilung gilt dann:

$$E(X)=\int\left(y\cdot f_X(y)\,dy\right)=\int_a^b\left(\frac{1}{b-a}\cdot y\,dy\right)=\frac{1}{b-a}\cdot\left[\frac{y^2}{2}\right]_a^b=\frac{b^2-a^2}{2}\cdot\frac{1}{b-a}=\frac{b+a}{2}$$

**Eigenschaften: Varianz, Standardabweichung**

Für die Gleichverteilung gilt dann: $V(X)\frac{(b-a)^2}{12}$

sowie: $\sigma(X)=\frac{b-a}{\sqrt{12}}=\frac{b-a}{2\cdot\sqrt{3}}\approx\frac{b-a}{2}\cdot 0.58$

**Beispiel: Kreisroulett**

siehe Abschnitt 1.5.2

### 4.3.2 Exponentialverteilung mit Parameter $\alpha$

**Definition: exponentiell verteilt**

Eine Zufallsgröße $X$ heißt **exponentiell verteilt**, wenn für die Dichte gilt: $f_X(y)=\begin{cases} 0 & \text{wenn } y<0 \\ \alpha\cdot e^{-\alpha\cdot y} & \text{wenn } y\geqslant 0 \end{cases}$ mit $\alpha>0$.

**Eigenschaften: Verteilungsfunktion**

Es gilt dann $F_X(x)=\int_{-\infty}^{x}\left(f_X(y)\,dy\right)=\alpha\cdot\int_0^x\left(e^{-\alpha\cdot y}\,dy\right)=1-e^{\alpha\cdot x}$

**Eigenschaften: Erwartungswert**

und $E(X)=\int\left(y\cdot f_X(y)dy\right)=\alpha\cdot\int\left(y\cdot e^{-\alpha\cdot y}dy\right)=\alpha\cdot\frac{1}{\alpha}\cdot\left[\frac{1}{-\alpha}\cdot e^{-\alpha\cdot y}\right]_0^\infty=\alpha\cdot\frac{1}{\alpha^2}=\alpha\cdot\alpha^{-2}=\alpha^{-1}$

**Eigenschaften: Varianz, Standardabweichung**

sowie $E(X)=\alpha^{-2}$
und $\sigma(X)=E(X)=\alpha^{-1}$

**Beispiel: Radioaktiver Zerfall von Atomkernen**

Die Zufallsgröße $X$ sei die Zerfallszeit $x$, wobei die Beobachtung bei $x=0$ beginnt. Die Wahrscheinlichkeit, dass der Atomkern im Zeitraum $[0,x]$ zerfällt, ist dann gegeben durch

$$P_x([0,x])=P(0\leqslant X\leqslant x)=\lambda\cdot\int_0^x\left(e^{-\lambda\cdot y}dy\right)$$

mit der für diese Atomkerne charakteristischen mittleren Lebensdauer $\lambda^{-1}=E(X)$.

Die Halbwertszeit $T_{\frac{1}{2}}$ ist der Zeitraum, in der der Atomkern mit Wahrscheinlichkeit $\frac{1}{2}$ zerfallen ist. $T_{\frac{1}{2}}$ ist also gegeben durch $P(0\leqslant X\leqslant T)=\frac{1}{2}=\lambda\cdot\int_0^{T_{\frac{1}{2}}}\left(e^{-\lambda\cdot y}dy\right)=1-e^{-\lambda\cdot T_{\frac{1}{2}}}$.

Es soll nun gelten $e^{-\lambda\cdot T_{\frac{1}{2}}}=\frac{1}{2}$. Damit folgt $T_{\frac{1}{2}}=\frac{\ln(2)}{\lambda}$. In der Physik gilt das sogenannte Zerfallsgesetz $N(t)=N_0\cdot e^{-\lambda\cdot t}$, wobei $N_0$ die Anzahl der Kerne am Anfang und $N(t)$ die Anzahl der Kerne zum Zeitpunkt $t$ ist. Für $T_{\frac{1}{2}}$ gilt dann: $N\left(T_{\frac{1}{2}}\right)=N_0\cdot e^{-T_{\frac{1}{2}}\cdot\lambda}=N_0\cdot\frac{1}{2}$.

Dieser Zusammenhang lässt sich auch anwenden auf „Absterben“ von Sprachen, Völkern usw.

## 4.3.3. Normalverteilung mit Parametern $\mu$ und $\sigma$

**Definition: normalverteilt**

Eine Zufallsgröße $X$ heißt **normal verteilt** mit den Parametern $\mu$ und $\sigma$, wenn für die Dichte gilt:

$$f_X(y)=\frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot e^{-\frac{(y-\mu)^2}{2\cdot\sigma^2}}$$

**Bemerkung**

Für die Dichte schreibt man oft $\varphi(X,\mu,\sigma^2)$ und sagt, $X$ ist $N(\mu,\sigma^2)$ verteilt.

**Bemerkung: Gaußsche Glockenkurve**

Der Graph der Funktion beschreibt die **Gaußsche Fehlerkurve** [achja?] (oder **Glockenkurve**) ist dem Maximum bei $\mu$ und den Wendepunkten $\mu-\sigma$ und $\mu+\sigma$.

**Bemerkung: Normalverteilung**

Die Funktion $\varphi\left(X,\mu,\sigma^2\right)$ heißt auch **Normalverteilung**.

**Eigenschaften**

Es gilt dann:

- $\varphi\left(X,\mu,\sigma^2\right)\geqslant 0$
- $\int_{-\infty}^{\infty}\left(\varphi\left(y,\mu,\sigma^2\right)dy\right)=1$
- $E(X)=\mu$
- $V(X)=\sigma^2$ und damit
- $\sigma(X)=\sigma$

**Definition: standardisierte Zufallsgröße**

Die **standardisierte Zufallsgröße** $\widetilde{X}:=\frac{X-E(X)}{\sigma(X)}=\frac{X-\mu}{\sigma}$ ist dann $N(0,1)$-verteilt.

**Bezeichnungen**

- $\varphi(X)=\varphi(X,0,1)$
- $\Phi(X):=\Phi(X,0,1)$ mit $\Phi\left(X,\mu,\sigma^2\right)=\int_{-\infty}^{x}\left(\varphi\left(y,\mu,\sigma^2\right)dy\right)=F_x(x)$ (Verteilungsfunktion)

**Bemerkung: tabelliert**

$\Phi(X)$ ist tabelliert. Man erhält daraus $\Phi\left(X,\mu,\sigma^2\right)$ mit Hilfe von $\Phi\left(X,\mu,\sigma^2\right)=\Phi\left(\frac{X-\mu}{\sigma}\right)$.

Es gilt:

- $\Phi(-x)=1-\Phi(x)$ (dies folgt aus $\forall(x):\left(\varphi(x)=\varphi(-x)\right)$)

Datum: 18.06.2003

[...5 Minuten zu spät...]

- Standardisierte Zufallsgröße $\widetilde{X}:=\frac{X-E(X)}{\sigma(X)}=\frac{X-\mu}{\sigma}$ ist dann $N(0,1)$ -verteilt
- Bezeichnungen:
  - $\varphi(x):=\varphi(x,0,1)$
  - $\Phi(x):=\Phi(x,0,1)$ mit $\Phi(x,\mu,\sigma^2)=\int_{-\infty}^{x}\left(\varphi(y,\mu,\sigma^2)\,dy\right)=F_X(x)$ (Verteilungsfunktion)
  - $\Phi(x)$ ist tabelliert. Man erhält daraus $\Phi(x,\mu,\sigma^2)$ vermittels $\Phi(x,\mu,\sigma^2)=\Phi(\frac{x-\mu}{\sigma})$

**Begründung**

für die Formel $\Phi\left(x,\mu,\sigma^2\right)=\Phi\left(\frac{x-\mu}{\sigma}\right)$ :

$$\Phi\left(\frac{x-\mu}{\sigma}\right)=\int_{-\infty}^{\frac{x-\mu}{\sigma}}\left(\frac{1}{\sqrt{2\cdot\pi}}\cdot e^{-\frac{y^2}{2}}dy\right)$$

Substituieren wir nun $y=\frac{z-\mu}{\sigma}$ . Damit ist $dy=\frac{1}{\sigma}\cdot dz$

$$\Rightarrow\Phi\left(\frac{x-\mu}{\sigma}\right)=\int_{-\infty}^{x}\left(\frac{1}{\sqrt{2\cdot\pi}\cdot\sigma}\cdot e^{-\frac{(z-\mu)^2}{2\cdot\sigma^2}}dz\right)=\Phi\left(x,\mu,\sigma^2\right)$$

**Satz:** $3\cdot\sigma$ **-Regel**

Eine Zufallsgröße $X$, die $N\left(\mu,\sigma^2\right)$-verteilt ist, nimmt mit 99.7% Sicherheit nur Werte im Intervall $\left]\mu-3\cdot\sigma,\mu+3\cdot\sigma\right[$ an.

**Begründung**

(wir haben eine stetige Zufallsgröße)

$$\begin{aligned}P(a\leqslant X\leqslant b)&=P(a<X<b)\\&=P\left(\frac{a-\mu}{\sigma}\leqslant\frac{X-\mu}{\sigma}\leqslant\frac{b-\mu}{\sigma}\right)\\&=F_X\left(\frac{b-\mu}{\sigma}\right)-F_X\left(\frac{a-\mu}{\sigma}\right)\\&=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\end{aligned}$$

Man benutzt $\Phi_X(x)=1-\Phi(-x)$ und $a=\mu-n\cdot\sigma$ , $b=\mu+n\cdot\sigma$ . Dann folgt:

$$P\left(\mu-n\cdot\sigma\leqslant X\leqslant\mu+n\cdot\sigma\right)=\Phi(n)-\Phi(-n)=2\cdot\Phi(n)-1$$

**Beispiel**

$n=1:$ $P(\mu-\sigma\leqslant X\leqslant\mu+\sigma)\approx 0.683$
$n=2:$ $P(\mu-2\cdot\sigma\leqslant X\leqslant\mu+2\cdot\sigma)\approx 0.955$
$n=3:$ $P(\mu-3\cdot\sigma\leqslant X\leqslant\mu+3\cdot\sigma)\approx 0.997$

**Bemerkung**

Normalverteilungen treten bei einer Vielzahl von Zufälligen Vorgängen auf:
- Beobachtungsfehler bei Messungen,
- Abweichungen von der Normgröße,
- allgemein bei zufälligen Vorgängen, die sich aus einer großen Zahl von einzelnen zufälligen Teilvorgängen ergeben, welche relativ unabhängig voneinander sind.

**Bemerkung**

Die Normalverteilung heißt „normal", weil sie als die normale Verteilung angesehen wird.

**Beispiel**

Die Körpergröße der Frauen in den USA im Alter zwischen 18 und 30 Jahren ist nahezu perfekt normalverteilt mit $\mu=162.5\,cm$, $\sigma=6.4\,cm$. Das bedeutet, dass 68.3% dieser Frauen eine Körpergröße zwischen $\mu-\sigma=156.1\,cm$ und $\mu+\sigma=168.9\,cm$ haben.

## Stetige Zufallsgröße $X$ mit Dichte $f_X(x)$

Wir können den Fall haben, dass uns die Zufallsgröße selbst gar nicht so sehr interessiert, sondern eher eine Funktion $g(x)$ dieser Zufallsgröße $X$.
Wir haben also die Funktionen
- $X:\Omega\rightarrow\mathbb{R}$, $\omega\rightarrow X(\omega)$
- $g(X):\Omega\rightarrow\mathbb{R}$, $\omega\rightarrow g(X(\omega))$

Dann gilt:
- $E(g(X))=\int_{-\infty}^{\infty}(g(y)\cdot f_X(y)dy)$, wie bei diskreter Zufallsgröße $E(g(X))=\sum_{\forall(i)}(g(x_i)\cdot p_i)$
- Dichte von $g(X)$:
  - Zuerst ermittelt man $F_{g(X)}(x):=P(g(x)<x)$. Dann folgt:
  - $f_{g(X)}(x)=F_{g(X)}'(x)$

**Beispiel**

Sei $g(X)=X^2$, $X$ selbst eine stetige Zufallsgröße.
$F_{X^2}(x)=P(X^2<x)=P(|X|<\sqrt{x})=P(-\sqrt{x}<X<\sqrt{x})=F_X(\sqrt{x})-F_X(-\sqrt{x})$

konkret: $f_X(y)=\begin{cases}\frac{1}{2}\cdot(y+1) & \text{wenn } y\in[-1,1]\\ 0 & \text{sonst}\end{cases}$

[Grafik, Foto auf Digicam]

Test: $f_X(y)\geqslant 0$, $\int_{-\infty}^{\infty}(f_X(y)dy)=1$

$$F_X(x)=\int_{-\infty}^{\infty}\left(f_X(y)\,dy\right)=\int_{-1}^{x}\left(\frac{1}{2}\cdot(y+1)\,dy\right)=\frac{1}{2}\cdot\left[\frac{y^2}{2}+y\right]_{-1}^{x}=\frac{1}{4}\cdot x^2+\frac{1}{2}\cdot x+\frac{1}{4}$$

$$E(X)=\int_{-1}^{1}\left(x\cdot\frac{1}{2}\cdot(x+1)\cdot dx\right)=1$$

Für $0\leqslant x\leqslant 1$ gilt:

$$F_{X^2}(x)=F_X(\sqrt{x})-F_X(-\sqrt{x})=\frac{1}{4}\cdot(\sqrt{x})^2+\frac{1}{2}\cdot\sqrt{x}+\frac{1}{4}-\left(\frac{1}{4}\cdot(-\sqrt{x})^2+\frac{1}{2}\cdot(-\sqrt{x})+\frac{1}{4}\right)=\sqrt{x}$$

$$f_{X^2}(x)=F_{X^2}{}'(x)=\begin{cases}0 & \text{wenn } x\leqslant 0\\ \dfrac{1}{2\cdot\sqrt{x}} & \text{wenn } x>0\end{cases}$$

$$E(X^2)=\int_{-\infty}^{\infty}\left(x\cdot f_{X^2}(x)\,dx\right)=\int_{0}^{1}\left(\frac{1}{2}\cdot\sqrt{x}\,dx\right)=\frac{1}{2}\cdot\left[\frac{2}{3}\cdot x^{\frac{3}{2}}\right]_0^1=\frac{1}{3}$$ Dies können wir überprüfen mit

$$E(X^2)=\int_{-\infty}^{\infty}\left(x^2\cdot f_X(x)\,dx\right)$$

„Sie sollten wissen, dass man das machen kann, aber die Rechnung ist recht aufwändig. Deswegen werde ich dies wohl nicht in der Klausur dran nehmen.“

## 4.4 Stetige Zufallsvektoren

### Definition 1: stetiger Zufallsvektor

Es sei $(X,Y)$ ein Zufallvektor. Er heißt stetig, wenn es eine Funktion $f_{(X,Y)}:\mathbb{R}^2\to\mathbb{R}$ gibt, die

- die nicht negativ ist,
- die integrabel ist,
- deren Integral 1 ist und
- $\forall(a_1,a_2,b_1,b_2\in\mathbb{R}):$

$$\left(\left((a_1\leqslant b_1)\wedge(a_2\leqslant b_2)\right)\Rightarrow\left(P\left((a_1\leqslant X\leqslant b_1)\wedge(a_2\leqslant Y\leqslant b_2)\right)=\int_{a_1}^{b_1}\left(dx\int_{a_2}^{b_2}\left(dy\cdot f_{(X,Y)}(x,y)\right)\right)\right)\right)$$

#### Beispiel

$$f(x,y)=x^2\cdot|y|+3$$
$$f(1,3)=1\cdot 3+3=6$$

#### Bemerkung

Was ist mit „integrabel“ gemeint?

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left(f_{(x,y)}(x,y)\,dx\right)dy\right)=\int_{-\infty}^{\infty}\left(\underbrace{\left(\int_{-\infty}^{\infty}\left(f_{(X,Y)}(x,y)\,dx\right)\right)}_{F(y)}dy\right)$$

### Zu den Übungsaufgaben

- Es wird noch eine 6. Serie geben.

- Es wird keine 7. Serie geben. „Wenn einer ein Bedürfnis hat, kann er aber trotzdem privat was kriegen.“
- Ich werde aber trotzdem eine 7. Serie in's Netz stellen, damit Sie sich auf die Klausur vorbereiten können.
- 2. Aufgabe: Poisson-Verteilung mit einem Parameter, der schon da steht.
- 3. Aufgabe: Fleißaufgabe
- 4. Aufgabe: ist mit der 3. Aufgabe verwandt, Sie müssen sich das Schema aber selbst verschaffen. Dies ist auch eine Fleißaufgabe, sie müssen viele Wahrscheinlichkeiten herausfinden und aufpassen, wenn bei beiden Würfen die selbe Augenzahl gewürfelt wurde.
- 6. Aufgabe: Ablaufplan: Verbale Beschreibung der Vorgehensweise

Datum: 20.06.2003

### Exkurs: Doppelintegrale oder Flächenintegral

- Bestimmtes Integral („Einfachintegral“) von $f$ über $[a,b]$
  - [Grafik]
  - $\int_a^b (f(x)dx)$ existiert für stückweise stetige $f$ mit $-\infty<a<b<\infty$
  - Geometrische Deutung im Fall $f(x)\geqslant 0$ :
    - Das Integral ist der Inhalt der Fläche $F$, $|F|=\int_a^b (f(x)dx)$
  - Definition und numerische Berechnung über Riemannsche Summen $\sum_{\int i=1}^{n} (f(\eta_i)\cdot|\Delta_i|)$
  - Analytische Berechung über die Stammfunktion von $f$
- Doppelintegral („Flächenintegral“) von $f$ über ein Gebiet $D$ in der $(x,y)$-Ebene
  - [Grafik]
  - $\iint_{(x,y)\in D} (f(x,y)dx\,dy)$ existiert für „vernünftige“ $D$ und $f$
    - zum Beispiel ist $D$ ein beschränktes Gebiet mit glatten Rändern und $f$ stetig auf $\overline{D}$
  - Geometrische Deutung im Fall $f(x,y)\geqslant 0$ :
    - Das Integral ist der Rauminhalt (Volumen) des Körpers im $(x,y,z)$-Raum mit „Höhe“ $z=f(x,y)$ für $(x,y)\in D$
    - $|V|=\iint_{(x,y)\in D} (f(x,y)dx\,dy)$
    - $V$ ist der (allgemeine Zylinder) über $D$ mit „Deckfläche“ $F=|(x,y,z)|(z=f(x,y))\wedge((x,y)\in D)\}$
  - Definition und numerische Berechnung über Riemannsche Summen: $\int_{i=1}^{n} (f(x_i,y_i)\cdot|D_i|)$ mit $(x_i,y_i)\in D_i$ und Zerlegung $\bigcup_{i=1}^{n}(D_i)=D$
  - Ist $D=[a,b]\times[c,d]$, dann ist

$$\iint_{(x,y)\in D} (f(x,y)dx\,dy)=\int_c^d\left(\int_a^b (f(x,y)dx)dy\right)=\int_a^b\left(\int_c^d (f(x,y)dy)dx\right)$$

**Beispiel**

Sei $\forall((x,y)\in D):(f(x,y)=c>0)$. Dann ist $\iint_{(x,y)\in D} (f(x,y)dx\,dy)=\iint_{(x,y)\in D} (c\,dx\,dy)=c\cdot\iint_{(x,y)\in D} (dx\,dy)=c\cdot|D|$, also das Volumen $V$ des Zylinders über $D$ mit Höhe $c$.

**Beispiel**

Sei $D=[a,b]\times[c,d]$, $\forall((x,y)\in D):(f(x,y)=f_1(x)\cdot f_2(x))$. Dann ist

$$\iint_{(x,y)\in D} \left(f_1(x)\cdot f_2(y)\,dx\,dy\right) = \int_a^b \left(\int_c^d \left(f_1(x)\cdot f_2(y)\,dy\right)dx\right)$$
$$= \int_a^b \left(f_1(x)\cdot\left(\int_c^d \left(f_2(y)\,dy\right)\right)dx\right)$$
$$= \left(\int_a^b \left(f_1(x)\,dx\right)\right)\cdot\left(\int_c^d \left(f_2(y)\,dy\right)\right)$$

**Bemerkung**

Wenn der Zufallsvektor $(X,Y)$ stetig ist, dann sind auch die Zufallsgrößen $X$ und $Y$ stetig. Ihre Dichten ergeben sich durch die Integration zu

- $f_X(x) = \int_{-\infty}^{\infty} \left(dy\, f_{(X,Y)}(x,y)\right)$ ,
- $f_Y(y) = \int_{-\infty}^{\infty} \left(dx\, f_{(X,Y)}(x,y)\right)$

**Bemerkung:** **Kovarianz**

Die Kovarianz $C(X,Y)$ erhält man aus
$$C(X,Y) = E(X\cdot Y) - E(X)\cdot E(Y) = \int \left(dx \int \left(dy\left(x\cdot y\cdot f_{(X,Y)}(x,y)\right)\right)\right) - E(X)\cdot E(Y)$$

## Stochastische Unabhängigkeit

Es sei $(X,Y)$ ein stetiger Zufallsvektor mit der Dichte $f_{(X,Y)}(x,y) = f_1(x)\cdot f_2(y)$ . Dann ist:

- $f_X(x) = \int_{-\infty}^{\infty} \left(f_{(X,Y)}(x,y)\,dy\right) = f_1(x)\cdot\left(\underbrace{\int_{-\infty}^{\infty} \left(f_2(y)\,dy\right)}_{\alpha}\right)$
- $f_Y(y) = \int_{-\infty}^{\infty} \left(f_{(X,Y)}(x,y)\,dx\right) = f_2(y)\cdot\left(\underbrace{\int_{-\infty}^{\infty} \left(f_1(x)\,dx\right)}_{\beta}\right)$

Mit $\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left(f_{(X,Y)}(x,y)\,dy\right)dx\right) = 1$ folgt aber sofort $\alpha\cdot\beta = 1$ . Wir können also gleich $\alpha = \beta = 1$ wählen und $f_1(x) = f_X(x)$ , $f_2(y) = f_Y(y)$ setzen.

Es folgt damit stochastische Unabhängigkeit, denn
$$P_{(X,Y)}\left(\Delta_1\times\Delta_2\right) = \iint_{(x,y)\in(\Delta_1\times\Delta_2)} \left(f_{(X,Y)}(x,y)\,dx\,dy\right)$$
$$= \left(\int_{x\in\Delta_1} \left(f_X(x)\,dx\right)\right)\cdot\left(\int_{y\in\Delta_2} \left(f_Y(y)\,dy\right)\right)$$
$$= P_X\left(\Delta_1\right)\cdot P_Y\left(\Delta_2\right)$$

## Beispiel

Zufallsvektor $(X,Y)$ mit Dichte $f_{(X,Y)}(x,y) = \begin{cases} \frac{3}{8}\cdot\left(x^2+y^2\right) & \text{für } (x,y)\in\left([-1,1]\times[-1,1]\right) \\ 0 & \text{sonst} \end{cases}$

- Randverteilungen
  - Randverteilung für $X$:

$$f_X(x)=\begin{cases}\frac{3}{4}\cdot\left(x^2+\frac{1}{3}\right) & \text{für } x\in[-1,1]\\ 0 & \text{sonst}\end{cases}$$

    - Warum?:
    - $f_X(x)=\int\limits_{-\infty}^{\infty}\left(f_{(X,Y)}(x,y)\,dy\right)=\int\limits_{-1}^{1}\left(\frac{3}{8}\cdot(x^2+y^2)\,dy\right)=$[...abgewischt...]
  - Randverteilung für $Y$:

$$f_Y(y)=\begin{cases}\frac{3}{4}\cdot\left(\frac{1}{3}+y^2\right) & \text{für } y\in[-1,1]\\ 0 & \text{sonst}\end{cases}$$

- Erwartungswerte: $E(X)=E(Y)=0$
- Varianzen: $V(X)=V(Y)=\frac{7}{15}$, $\sigma(X)=\sigma(Y)\approx 0.68$
- Kovarianz: $C(X,Y)=0$ (also ist $X$ mit $Y$ nicht korrelliert)

$$\begin{aligned}C(X,Y)&=E(X\cdot Y)-E(X)\cdot E(Y)\\&=E(X\cdot Y)\\&=\iint\left(x\cdot y\cdot f_{(X,Y)}(x,y)\,dx\,dy\right)\\&=\frac{3}{8}\cdot\int\limits_{-1}^{1}\left(\int\limits_{-1}^{1}\left((x^2+y^2)\cdot x\cdot y\,dy\right)dx\right)\\&=\frac{3}{8}\cdot\int\limits_{-1}^{1}\left(\int\limits_{-1}^{1}\left((x^3\cdot y+y^3\cdot x)\,dy\right)dx\right)\end{aligned}$$

[...abgedeckt...]

**Beispiel**

Gegeben sei ein stetiger Zufallsvektor $(X,Y)$ mit Dichte

$$f_{(X,Y)}(x,y)=\begin{cases}(\lambda\cdot\mu)\cdot e^{-\lambda\cdot x}\cdot e^{-\mu\cdot y} & \text{wenn } (x,y)\in[0,\infty[^2\\ 0 & \text{sonst}\end{cases}$$

Hier ist $f_{(X,Y)}(x,y)=f_X(x)\cdot f_Y(y)$ und $X$, $Y$ sind exponentiell verteilt mit den Parametern $\lambda$ und $\mu$.

**Beispiel**

Gegeben sei ein stetiger Zufallsvektor $(X,Y)$ mit Dichte $f_{(X,Y)}(x,y)$. Wir interessieren uns für die Zufallsgröße $X+Y$. Dann gilt:

$F_{X+Y}(t)=P(X+Y<t)=\iint\limits_{(x,y)\in K_t}\left(f_{(X,Y)}(x,y)\,dx\,dy\right)$, mit $K_t\subset\mathbb{R}^2$ bestimmt durch $x+y<t$.

[Grafik]

$$f_{(X,Y)}(x,y)=\begin{cases}\frac{1}{\pi} & \text{wenn } x^2+y^2<1\\ 0 & \text{sonst}\end{cases}$$

Hier ist:

- $\forall(t<-\sqrt{2}):(F_{X+Y}(t)=0)$
- $\forall(t>\sqrt{2}):(F_{X+Y}(t)=1)$
- $\forall(t\in[-\sqrt{2},\sqrt{2}]):\left(F_{X+Y}(t)=\frac{1}{\pi}\cdot|K_t\cap E|\right)$
  - Dies ist aus der Formel für das Kreissegment ermittelbar. Das geht zwar elementar, ist aber länglich.

Wir können also wie folgt zusammenfassen:

- Zufallsgröße
  - diskret:
    - Beschreibung durch $\forall(i):((x_i,p_i))$
    - $E(X)=\sum_{\forall(j)}(x_j\cdot p_j)$
    - $V(X)=\sum_{\forall(j)}\left((x_j-E(X))^2\cdot p_j\right)$
  - stetig:
    - Beschreibung durch $x$, $f_X(x)$
    - $E(X)=\int(x\cdot f_X(x)dx)$
    - $V(X)=\int\left((x-E(X))^2\cdot f_X(x)dx\right)$
- Zufallsvektor $(X,Y)$
  - diskret:
    - Beschreibung durch $\forall(i):\forall(j):((x_i,y_j),p_{ij})$
    - Randverteilung:
      - Randverteilung für $X$: $p_i=\sum_{\forall(i)}(p_{ij})$
      - Randverteilung für $Y$: $q_y=\sum_{\forall(i)}(p_{ij})$
    - Kovarianz: $C(X,Y)=E((X-E(X))\cdot(Y-E(Y)))=E(X\cdot Y)-E(X)\cdot E(Y)$
      $E(X\cdot Y)=\sum_{\forall(i):\forall(j)}(x_i\cdot y_j\cdot p_{ij})$
  - stetig:
    - Beschreibung durch $f_{(X,Y)}(x,y)$
    - Randverteilung:
      - Randverteilung für $X$: $f_X(x)=\int(f_{(X,Y)}(x,y)\cdot dy)$
      - Randverteilung für $Y$: $f_Y(y)=\int(f_{(X,Y)}(x,y)dx)$
    - Kovarianz: $C(X,Y)=E((X-E(X))\cdot(Y-E(Y)))=E(X\cdot Y)-E(X)\cdot E(Y)$
      $E(X\cdot Y)=\int\left(dx\cdot\int\left(dx\cdot x\cdot y\cdot f_{(X,Y)}(x,y)\right)\right)$

# 5. Grenzwertsätze

## 5.1 Schwaches Gesetz der großen Zahlen

Ziel ist, den Zusammenhang zwischen Wahrscheinlichkeit und relativer Häufigkeit genauer aufzuzeigen, das empirische Gesetz der großen Zahl hier im mathematischen Modell wiederzufinden.

**Satz 1: Tschebyscheffsche Ungleichung**

Es sei $X$ eine Zufallsgröße mit Erwartungswerte $E(X)$ und Varianz $V(X)$. Es sei $\epsilon>0$ gegeben. Dann gilt:

- $P\left(-\epsilon\leqslant X-E(X)\leqslant\epsilon\right)\geqslant 1-\frac{V(X)}{\epsilon^2}$ beziehungsweise
- $P\left(\left|X-E(X)\right|\geqslant\epsilon\right)\leqslant\frac{V(X)}{\epsilon^2}$

(Die Wahrscheinlichkeit dafür, dass $X$ Werte annimmt, die um mehr als $\epsilon$ von $E(X)$ abweichen, ist kleiner als die Varianz geteilt durch $\epsilon^2$ )

**Beweis: stetiger Fall**

Wir haben, für $\epsilon>0$ ,

$$\begin{aligned}
V(X) &= \int_{-\infty}^{\infty}\left(\left(x-E(X)\right)^2\cdot f_X(x)\,dx\right)\\
&\geqslant \int_{\left|x-E(X)\right|\geqslant\epsilon}\left(\left(x-E(X)\right)^2\cdot f_X(x)\,dx\right)\\
&\geqslant \epsilon^2\cdot\int_{\left|x-E(X)\right|\geqslant\epsilon}\left(f_X(x)\,dx\right)\\
&= \epsilon^2\cdot P\left(\left|X-E(x)\right|\geqslant\epsilon\right)
\end{aligned}$$

Das ergibt $P\left(\left|X-E(X)\right|\geqslant\epsilon\right)\leqslant\frac{V(X)}{\epsilon^2}$ .

Datum: 25.06.2003

**Bemerkung**

Die Tschebyscheffsche Ungleichung ist eine sehr grobe Abschätzung. Dafür gilt sie ultimativ für alle Verteilungen.

**Beispiel**

Man wähle $\epsilon=3\cdot\sigma(X)$. Dann folgt: $P(|X-E(X)|\leqslant 3\cdot\sigma(X))\geqslant\frac{8}{9}=0.89$

Wenn $X$ normalverteilt ist, folgt aus der $3\cdot\sigma$-Regel: $P(|X-E(X)|\leqslant 3\cdot\sigma)\geqslant 0.996$

Die Abschätzung ist also sehr „konservativ“.

### Satz 2: Bernoullis Gesetz der großen Zahl

Es sei $Y_m$ die Zufallsgröße, die die relative Häufigkeit eines Ereignisses $A$ in $n$ Versuchen beschreibt. Es sei $p=P(A)$. Dann gilt :

- $\forall(\epsilon\in\mathbb{R}^+):\left(\lim\limits_{n\to\infty}(P(|Y_n-p|<\epsilon))=1\right)$ beziehungsweise
- $\forall(\epsilon\in\mathbb{R}^+):\left(\lim\limits_{n\to\infty}(P(|Y_n-p|>\epsilon))=0\right)$

**Beweis**

Wir wissen bereits, dass $(E(Y_n)=p)\wedge\left(V(Y_n)=\frac{p\cdot(1-p)}{n}\right)$. Damit folgt aus der Tschebyscheffschen Ungleichung:

$$P(|Y_n-p|\leqslant\epsilon)\geqslant 1-\frac{V(Y_n)}{\epsilon^2}=1-\frac{p\cdot(1-p)}{n\cdot\epsilon^2}$$

Daraus folgt:

$$\lim_{n\to\infty}(P(|Y_n-p|\leqslant\epsilon))\geqslant\lim_{n\to\infty}\left(1-\frac{V(Y_n)}{\epsilon^2}\right)=\lim_{n\to\infty}\left(1-\frac{p\cdot(1-p)}{n\cdot\epsilon^2}\right)=1$$

[prüfungsrelevant:] „Also das ist so simpel, das könnte man Sie sogar abfragen, wenn wir wollen.“

## Konvergenzarten in der Stochastik

### Definition: stochastische Konvergenz

Sei eine Folge von Zufallsgrößen $X_n$ und eine Zufallsgröße $X$ gegeben.
Gilt $\forall(\epsilon\in\mathbb{R}^+):\left(\lim\limits_{n\to\infty}(P(|X_n-X|<\epsilon)=1)\right)$, so liegt **stochastische Konvergenz** ( $X_n \underset{\text{stochastische Konvergenz}}{\rightarrow} X$ ) vor.

### Definition: fast sichere Konvergenz

Sei eine Folge von Zufallsgrößen $X_n$ und eine Zufallsgröße $X$ gegeben.

Gilt $P\left(\left\{\omega\in\Omega\middle|\lim_{n\to\infty}\left(X_n(\omega)\right)=X(\omega)\right\}\right)=1$, so liegt **fast sichere Konvergenz** ( $X_n \underset{\text{fast sicher}}{\rightarrow} X$ ) vor.

**Definition: Konvergenz im Sinne der Verteilung**

Sei eine Folge von Zufallsgrößen $X_n$ und eine Zufallsgröße $X$ gegeben.
Gilt $\forall(x\in\mathbb{R}):\left(\lim_{n\to\infty}\left(F_{X_n}(x)\right)=F_X(x)\right)$, so liegt **Konvergenz im Sinne der Verteilung** vor.

„Das Gesetz der großen Zahlen ist das Murphy'sche Gesetz: Alles was schief gehen kann, geht irgendwann schief."

**Bemerkung**

Die relative Häufigkeit konvergiert stochastisch gegen die Wahrscheinlichkeit $P(A)$.

**Bemerkung**

Jedes Ereignis $A$ mit $P(A)>0$ tritt in einer langen Kette von Versuchen auf. Deshalb heißt das Gesetz „Gesetz der großen Zahlen".

**Bemerkung**

Das Gesetz ist von Jacob Bernoulli (um etwa 1685) entdeckt worden. Er sagte dazu: „Diese Entdeckung gilt mir mehr, als wenn ich die Quadratur des Kreises geliefert hätte. Denn wenn diese auch gänzlich gefunden wäre, so wäre ie doch sehr wenig nütz."

**Bemerkung**

Das Gesetz heißt auch „schwaches Gesetz der großen Zahl" oder „Gesetz der großen Zahl(en) von Bernoulli".

## 5.2 Approximation der Binomialverteilung

**Definition: standardisierte Größe**

Es sei $X$ binomialverteilt mit Parametern $n$, $p$. Das heißt:

$$\forall\left(k_1,k_2\in\mathbb{N}_0^+\right):\left(\left(k_1\leqslant k_2\right)\Rightarrow\left(P\left(k_1\leqslant X\leqslant k_2\right)=\sum_{l=k_1}^{k_2}\left(B_{n,p}(l)\right)\right)\right)$$

Dann ist die **standardisierte Größe** $\widetilde{X}:=\frac{X-E(X)}{\sigma(X)}=\frac{X-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}$.

Es gilt:

- $\widetilde{X}$ hat die Werte $\forall\left(k\in\{0,1,\ldots,n\}\right):\left(x_k=\frac{k-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}\right)$
- $E(\widetilde{X})=0$
- $V(\widetilde{X})=1$

Man vergleiche dies mit der Zufallsgröße $Y$, die normalverteilt ist mit $(\mu=0)\wedge(\sigma=1)$

**Beispiel**

Sei eine Zufallsgröße $X$ gegeben mit $(p=0.3) \wedge (n=5)$. Daraus folgt:

- $E(X)=1.5$
- $\sigma(X) \approx 1.025$

Dann ist die zugehörige standardisierte Zufallsgröße etwa wie folgt

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $x_k$ | -1,46 | -0,49 | 0,49 | 1,46 | 2,44 | 3,42 |
| $p_k$ | 0,17 | 0,36 | 0,31 | 0,13 | 0,07 | 0,00 |

## Histogramme und standardisierte Normalverteilung

[Grafik, die viele Verteilungen standardisierter Zufallsgröße mit wachsendem $n$ bis zum Grenzübergang der Gauß-Glockenkurve zeigt]
Für große $n$ sind die standardisierten Binomialverteilungen nicht von Normalverteilungen zu unterscheiden.

**Satz: de Moire - Laplace**

Es seien die Zufallsgrößen $X_n$ binomialverteilt mit den Parametern $n$, $p$. Dann gilt:
$\forall (a,b): \left( (a<b) \Rightarrow \left( \lim_{n\to\infty} \left( P\left(a \leqslant \widetilde{X}_n \leqslant b\right)\right) = \Phi(b) - \Phi(a) \right)\right)$.
Hierbei ist $\Phi(\cdot)$ die Normalverteilungsfunktion mit $(\mu=0) \wedge (\sigma=1)$.

**Beweis**

Der Beweis benutzt die Stirlingsche Formel: $n! \approx \sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n$. Aber ich stelle ihn hier nicht vor.

**Bemerkung**

Es gilt die Faustregel: für $n \cdot p \cdot (1-p) \geqslant 10$ ist die Approximation der Binomialverteilung durch die Normalverteilung gut.

**Beispiel 1: Würfelexperiment**

Wir haben 1200 Würfe. Erwartet werden 200 mal „Sechsen“. Wie groß ist die Wahrscheinlichkeit dafür, dass die Zahl der „Sechsen“ zwischen 180 und 220 liegt?

Die Zufallsgröße $X$ messe die absolute Häufigkeit der „Sechsen“.

$$P(180 \leqslant X \leqslant 220) = \sum_{l=180}^{220} \left( B_{1200,\frac{1}{2}}(l) \right)$$

Dies zu berechnen ist sehr langwierig. „Damit haben Sie eine Weile, zum Beispiel ein Semester lang, zu tun.“ Dafür ist die Idee der Approximation.

Man geht wie folgt vor: Man berechnet die standardisierte Zufallsgröße $\widetilde{X}=\frac{X-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}$. Hier ist $n\cdot p=200$, $\sigma(X)=\sqrt{n\cdot p\cdot(1-p)}\approx 12.91$. Folglich nimmt dann $\widetilde{X}$ Werte an zwischen $\frac{180-200}{12.91}\approx -1.55$ und $\frac{220-200}{12.91}\approx 1.55$.

Das ergibt $P(180\leqslant X\leqslant 220)\approx P(-1.55\leqslant\widetilde{X}\leqslant 1.55)\approx\Phi(1.55)-\Phi(-1.55)=2\cdot\Phi(1.55)-1\approx 0.878$.

Die Wahrscheinlichkeit dafür, in den 1200 Versuchen zwischen 180 und 220 mal „Sechsen“ zu erhalten, ist 87.8%.

**Beispiel 2: Wahlvorhersage**

Ziel ist das Schätzung der Prozentzahl der Partei $A$ mit Hilfe einer Stichprobe. Die Gesamtzahl der Wähler sei $N$, die Anzahl der Wähler für $A$ sei $N_A$. $p=\frac{N_A}{N}$ ist dann die Wahrscheinlichkeit dafür, das ein zufällig herausgegriffener Wähler Wähler für $A$ ist. Eine Stichprobe vom Umfang $n$ liefere $n_A$ Wähler für die Partei $A$. Als Schätzwert für $p$ wird $p_0:=\frac{n_A}{n}$ genommen.

Wie geht man vor?

$X$ sei die Zufallsgröße, die die Werte $n_A\in\{0,1,\dots,n\}$ in den Stichproben annimmt.

- $X$ ist hypergeometrisch verteilt (der Wähler wird ja nicht „zurückgelegt“). Für $N\gg n$ können wir annehmen, $X$ ist binomialverteilt, also $P(X=n_a)=B_{n,p}(n_A)$ mit $p=\frac{N_A}{N}$.

  (Dies ist die erste Approximation: Ersetzung von der hypergeometrischer Verteilung durch eine entsprechende Binomialverteilung.)
- Wir wollen nun wissen, wie gut $p_0=\frac{n_A}{n}$ den gesuchten Wert $p=\frac{N_A}{N}$ approximiert.

  Dazu gibt man sich zwei Zahlen vor:
  - den **Vertrauensbereich** („Confidence-Bereich“), der die zulässige Abweichung von $p_0$ zu $p$ angibt. Dieser sei hier 1%. Also $|p-p_0|\leqslant 0.01$.
  - die **Garantiewahrscheinlichkeit**, die angibt, wie wahrscheinlich es ist, dass wirklich $|p-p_0|\leqslant 0.01$ ist. Wir wählen 95% (und damit die Irrtumswahrscheinlichkeit von 5%). Das heißt: $P(-0.01\leqslant p_0-p\leqslant 0.01)\geqslant 0.95$
- zweite Approximation: Ersetzung der Binomialverteilung durch eine Normalverteilung

  Wir benutzen die standardisierte Größe $\widetilde{X}=\frac{X-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}$. Nun ist $\frac{(p_0-p)\cdot n}{\sigma_n}=\frac{n_A-p\cdot n}{\sigma_n}$ und

  $$P(-0.01\leqslant p_0-p\leqslant 0.01)\approx P\left(\frac{-0.01\cdot n}{\sigma_n}\leqslant\widetilde{X}\leqslant\frac{0.01\cdot n}{\sigma_n}\right)\approx 2\cdot\Phi\left(\frac{0.01\cdot n}{\sigma_n}\right)-1$$

  Diese Zahl soll größer oder gleich $0.95$ sein. Das ist eine Forderung an $n$ (Umfang der Stichprobe). Nun folgt aus $2\cdot\Phi(z)-1=0.95$: $z\approx 1.96$. Mit $z=\frac{0.01\cdot n}{\sqrt{n\cdot p\cdot(1-p)}}$ dann

  $n=p\cdot(1-p)\cdot(1.96)^2\cdot 10000$

  Leider kennen wir $p$ nicht. Also wählen wir den ungünstigsten Fall, für den $n$ am größten wird. Das ist der Fall $p=0.5$. Das ergibt $n\approx 9600$. Die Stichprobe vom Umfang $n\geqslant 9600$ liefert also mit 95% Sicherheit $p_0=\frac{n_A}{n}\in\left]p-0.01\,,\,p+0.01\right[$

**Zu den Übungsaufgaben**

- Die Serie 6 ist die letzte Serie.

Datum: 02.07.2003

[...28 Minuten zu spät...]
<import from=„Bettina Biehl“>

## 5.3 Zentraler Grenzwertsatz

Wir haben gesehen, dass eine Folge von Zufallsgrößen $(X_n)$, die binomialverteilt sind mit Parametern $p$ und $n$, gegen eine normalverteilte Zufallsgröße $X$ mit $\mu=0$ und $\alpha=1$ strebt [achja?] (zumindest nach Standardisierung [na also]).

Dies ist ein Spezialfall des folgenden allgemeinen Satzes:

### Satz: zentraler Grenzwertsatz

(von Lindeberg|Levy um 1925)
Es sei

- $\forall(j):(X_j)$ eine Folge identisch verteilter Zufallsgrößen mit
  - $\forall(j):(E(X_j)=\mu)$ und
  - $\forall(j):(V(X_j)=\sigma)$. Es seien
- diese Zufallsgrößen stochastisch unabhängig und
- $Y_n:=X_1+X_2+\ldots+X_n$,
- $\widetilde{Y}_n=\dfrac{Y_n-\mu\cdot n}{\sqrt{n}\cdot\sigma}$ sowie
- $\forall(n):(F_n)$ die Verteilungsfunktionen dieser standardisierten Zufallsgrößen $\widetilde{Y}_n$.

Dann gilt:

- $\lim\limits_{n\to\infty}\left(\sup\left(\left|F_n(x)-\Phi(x)\right|\forall(x)\right)\right)=0$.

#### Bemerkung

Dieser Satz gilt auch unter schwächeren Annahmen an die Zufallsgröße $X_j$. Sie müssen nicht unbedingt gleiche Verteilungen besitzen.

#### Deutung des zentralen Grenzwertsatzes

Bei zufälligen Vorgängen, die sich durch Überlagerung einer Vielzahl weitgehend unabhängiger, zufälliger Einflüsse ergeben, die alle nur einen <u>klein</u>en Einfluss auf diesen zufälligen Vorgang haben, können wir von einer <u>Normalverteilung</u> [deswegen „normal“] der Zufallsgröße dieses Vorgangs ausgehen.

Je mehr zufällige Effekte einen Vorgang in verschiedenster, regelloser Weise beeinflussen, desto größer ist die Harmonie und Regelmäßigkeit dieses Vorgangs.

#### Beispiel

Um die Wahrscheinlichkeit $P_X(\{k,k+1,\ldots,l\})$ zu berechnen, kann man für $(n\cdot p)\geqslant 10$ näherungsweise folgende Rechnung nutzen:

$P(k \leqslant X \leqslant l) \approx \Phi\left(\frac{l-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}\right)-\Phi\left(\frac{k-n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}}\right)$ (wobei $\Phi(\cdot)$ die Normaverteilung ist)

Mit Stetigkeitskorrektur gilt:

$P(k \leqslant X \leqslant l) \approx \Phi\left(\frac{l-n\cdot p+\frac{1}{2}}{\sqrt{n\cdot p\cdot(1-p)}}\right)-\Phi\left(\frac{k-n\cdot p-\frac{1}{2}}{\sqrt{n\cdot p\cdot(1-p)}}\right)$ (Normalverteilung um $\frac{1}{2}$ verschoben.) Die Approximation wird dadurch verbessert.

</import>

**Beispiel: Würfelwürfe**

$p=\frac{1}{6}$, $n=600$

Gesucht ist $P(90 \leqslant X \leqslant 110)$.

- Exakter Wert: 0.7501
- Näherung ohne Stetigkeitskorrektur: 0.7266
- Näherung mit Stetigkeitskorrektur: 0.7498

**Bemerkung**

„Wenn $p$ klein ist, muss man große Werte $n$ haben.“ Diese Aussage ist ein Spezialfall des Zentralen Grenzwertsatzes.

# 6. Statistik

## 6.1 Beschreibende Statistik

- Die Aufgabe der beschreibenden Statistik ist, große Datenmengen in übersichtlicher und anschaulicher Form darzustellen und zusammenzufassen, sodass das Wesentliche dabei klar zum Ausdruck kommt. [Data-Mining?]
- Hilfsmittel sind:
  - Tabellen
  - Strichlisten
  - grafische Darstellungen sowie
  - (durch Berechnung) statistische Maßzahlen

**Definition: Grundgesamtheit, Population**

Die **Grundgesamtheit** (oder **Population**) bezeichnet die Menge der Untersuchungsobjekte, die ein oder mehrere interessierende Merkmale besitzen.

**Definition: Stichprobe**

Eine **Stichprobe** ist eine endliche Teilmenge der Grundgesamtheit, die man „zufällig“ auswählt.

### Definition: Umfang der Stichprobe

Der **Umfang der Stichprobe** ist die Anzahl ihrer Elemente.

### Definition: Merkmale

Merkmale $X$, $Y$ sind dann Zufallsgrößen mit möglichen Werten (Merkmalsausprägungen) $x_1, \ldots, x_i$ , $y_1, \ldots, y_k$ (meist sind dies reelle Zahlen, angeordnet der Größe nach).

Bei einer Stichprobe vom Umfang $n$ ermittelt man die Werte des Merkmals $X$ (bzw. $X$, $Y$), die die Elemente der Stichprobe haben. Meist ordnet man gleich die $n$ Elemente der Stichprobe nach den Werten $a_j$, die $X$ annimmt: $\forall(j):(a_j \leqslant a_{j+1})$

#### Beispiel: Körpergröße von Geschwisterpaaren

- Grundgesamtheit: (erwachsene) Geschwisterpaare (mit jeweils einer Schwester und einem Bruder)
- Merkmal $X$: Größe des Bruders
- Merkmal $Y$: Größe der Schwester
- Stichprobenumfang: $n=11$

Stichprobe (bereits geordnet nach $X$) (Angaben in cm):

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | 165 | 167 | 167 | 170 | 173 | 178 | 178 | 180 | 180 | 182 | 185 |
| $Y$ | 150 | 157 | 165 | 160 | 162 | 165 | 165 | 157 | 175 | 166 | 161 |

[Grafik der Verteilungswolke]

## Statistische Maßzahlen

### Definition: Median, Zentralwert

Der **Median** (oder **Zentralwert**) für eine Zufallsgröße $X$ mit den Ergebnissen $\forall(j):(a_j)$ ist:

$$x_{\frac{1}{2}} = \begin{cases} a_{\frac{n+1}{2}} & \text{wenn } n.istUngerade() \\ \frac{1}{2} \cdot \left(a_{\frac{n}{2}} + a_{\frac{n}{2}+1}\right) & \text{wenn } n.istGerade() \end{cases}$$

#### Eigenschaften

Für den Median gilt:

- Die Hälfte aller $a_j$ sind kleiner oder gleich $x_{\frac{1}{2}}$.
- Die Hälfte aller $a_j$ sind größer oder gleich $x_{\frac{1}{2}}$.

#### Beispiel: Körpergrößen von Geschwisterpaaren

$n=11$
$x_{\frac{1}{2}}=178$
$y_{\frac{1}{2}}=162$

**Definition: Variationsbreite**

Die Variationsbreite ist $\delta_X := a_{max} - a_{min}$ .

**Beispiel**

$$\delta_X = 185\,cm - 165\,cm = 20\,cm$$
$$\delta_Y = 175\,cm - 150\,cm = 25\,cm$$

**Definition: arithmetisches Mittel, Mittelwert**

Das arithmetische Mittel (oder: der Mittelwert) ist: $\overline{x} = \frac{1}{n} \cdot \sum_{j=1}^{n} (a_j)$ .

**Beispiel**

$$\overline{x} = 175\,cm$$
$$\overline{y} = 162.1\,cm$$

**Definition: empirische Streuung**

$$S_X{}^2 := \frac{1}{n-1} \cdot \sum_{j=1}^{n} \left( (a_k - \overline{x})^2 \right)$$

**Definition: empirische Standardabweichung**

$$S_X := \sqrt{S_X{}^2}$$

**Beispiel**

$$S_X \approx 6.88\,cm$$
$$S_Y \approx 6.41\,cm$$

**Definition: oberes Quartil**

$x_{0.75}$ ist das $x$, für das 75% aller $a_j$ kleiner oder gleich $x_{0.75}$ sind.

**Definition: unteres Quartil**

$x_{0.25}$ ist das $x$, für das 25% aller $a_j$ kleiner oder gleich $x_{0.25}$ sind.

**Beispiel**

$x_{0.75} = 180\,cm$ $y_{0.75} = 165\,cm$
$x_{0.25} = 167\,cm$ $y_{0.25} = 157\,cm$

**Definition: empirische Kovarianz**

$$C_e(X, Y) := \frac{1}{n-1} \cdot \sum_{j=1}^{n} \left( (x_j - \overline{x}) \cdot (y_j - \overline{y}) \right)$$

**Definition: empirischer Korrelationskoeffizient**

Entsprechend heißt $r(X,Y)=\frac{C_e(X,Y)}{S_X \cdot S_Y}$ der **empirische Korrelationskoeffizient**.

**Beispiel**

$\nu(X,Y)\approx 0.503$. Der Korrelationskoeffizient ist recht groß. Dies deutet also auf einen Zusammenhang zwischen den Körpergrößen der Geschwister hin.

**Definition: Häufigkeitstabelle, Kontingenztafel**

Stichproben mit zwei Merkmalen werden oft durch **Häufigkeitstabelle**n erfasst.

| $X \backslash Y$ | $a_1$ | $a_2$ | $\dots$ | $a_k$ | $X$ |
|---|---|---|---|---|---|
| $b_1$ | $H_{11}$ | $H_{12}$ | $\dots$ | $H_{1k}$ | $\tilde{H}_1$ |
| $b_2$ | $H_{21}$ | $H_{22}$ | $\dots$ | $H_{2k}$ | $\tilde{H}_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $b_l$ | $H_{l1}$ | $H_{l2}$ | $\dots$ | $H_{lk}$ | $\tilde{H}_l$ |
| $Y$ | $H_1$ | $H_2$ | $\dots$ | $H_k$ | $\Sigma$ |

- $\forall(j):(a_j)$ sind die Merkmalsausprägungen von $Y$
- $\forall(i):(b_i)$ sind die Merkmalsausprägungen von $X$
- $H_{ij}$ ist die Anzahl der Beobachtungen von $(b_i, a_k)$
- $L \times K$ heißt Kontingenztafel

**Beispiel: Arbeitsunfälle in einem Betrieb**

- Die Zufallsgröße $X$ sei die Tätigkeit mit den Werten
  - P: Produktion
  - V: Verwaltung
  - T: Transport
- Die Zufallsgröße $Y$ sei die Verletzung mit den Werten
  - 1: obere Extremitäten
  - 2: untere Extremitäten
  - 3: sonstige

| $X \backslash Y$ | 1 | 2 | 3 | $X$ |
|---|---|---|---|---|
| P | 32 | 23 | 5 | 60 |
| T | 6 | 19 | 5 | 30 |
| V | 2 | 8 | 20 | 30 |
| $Y$ | 40 | 50 | 30 | 120 |

Um die relativen Häufigkeiten zu erhalten, teile man durch die Summe.

**Beispiel:** **Heiratswahrscheinlichkeit von 1885**

- über die Zufallsgrößen
  - Alter und
  - Herkunft.

| Gruppen \ Alter in Jahren | 20..25 | 25..30 | 30..35 | 35..40 |
|---|---|---|---|---|
| englischer Adel | 0,05 | 0,08 | 0,07 | 0,06 |
| Belgien | 0,04 | 0,08 | 0,08 | 0,06 |
| Frankreich | 0,06 | 0,12 | 0,12 | 0,08 |
| Sachsen | 0,13 | 0,18 | 0,14 | 0,11 |
| Belgien (verwitwet) | 0,50 | 0,46 | 0,36 | 0,27 |

Dieses Beobachtungsmaterial ist mit den relativen Häufigkeiten als Wahrscheinlichkeiten gedeutet.

### zur Klausur

- am 21.07.2003, 15:00 Uhr
  - Hörsaal 13: für diejenigen, deren 1. Buchstabe im Bereich 'A'..'K' ist.
  - Hörsaal 19: für alle anderen
- Zeit: 60 Minuten
- Mitzubringen
  - Schreibzeug
  - Papier
  - Studentenausweis
  - Taschenrechner (für einfaches Rechnen)
  - sonstige Hilfsmittel (Aufzeichnungen, Bücher, ...) sind nicht erlaubt
- Inhalt der Klausur:
  - Fragen zum Stoff der Vorlesung und Aufgaben
  - keine Statistik
  - keine stetigen Zufallsvetoren
  - Schwerpunkte
    - Alles über Zufallsgrößen
      - Definition
      - diskret, stetig
      - $E(X)$, $V(X)$, $C(X,Y)$
      - spezielle Verteilungen
        - Binomialverteilung
        - Normalverteilung
      - Verteilungsfunktion
      - Dichte
      - Randverteilung
    - Grenzwertsätze
      - Bernoullis Gesetze der großen Zahl(en)
      - Zentraler Grenzwertsatz

- Bedingte Wahrscheinlichkeit
    - Definition
    - stochastische Unabhängigkeit
    - Bayes-Formel
    - Satz von der totale Wahrscheinlichkeit
- Aufgaben sind verhältnismäßig einfach, sodass sie in kurzer Zeit lösbar sind.

Datum: 09.07.2003

[..28 Minuten zu spät...]
<import from= „Manfred Wollenberg“>

## 6.2 Schätztheorie

Schätztheorie befasst sich allgemein mit der Schätzung von unbekannten Größen eines zufälligen Vorgangs, beziehungsweise von großen Datenmengen. Das betrifft:

- Wahrscheinlichkeiten
- Erwartungswerte
- Varianzen
- Art der Verteilung des Zufälligen Vorgangs
- ...

Man unterscheidet dabei in der Schätztheorie:

- **Punktschätzung**:
  Hier geht es darum, die unbekannte Größe möglichst genau zu ermitteln, gute Schätzwerte zu finden.
- **Intervallschätzung**:
  Hier versucht man den gesuchten Wert einzugrenzen durch obere und untere Schranken, die den Wert mit großer Sicherheit einschließen.

Meist sind beide Schätzungen miteinander verbunden.

**Beispiel: Punktschätzung**

Punktschätzung einer unbekannten Wahrscheinlichkeit $p=P(A)$ vermittels des Schätzwertes $p_0=\frac{n_A}{n}$, der relativen Häufigkeit aus der Stichprobe.

**Beispiel: Punktschätzung**

Punktschätzung des Erwartungswertes $E(X)$ (sowie der Varianz $V(X)$) vermittels des Schätzwertes $\mu_0=\bar{x}$ (sowie $\sigma_0^2=S_X^2$) aus der Stichprobe mit dem Ergebnis $x=(x_1, x_2, \ldots, x_n)$
.
[...]

</import>
[Tafelbild]
Extremwert von $L_X(\lambda)$
[...mutwillig abgewischt...]

[...]
Gesucht ist der Wert $\lambda=p$ (hier ist $\lambda=g(\lambda)$). Dazu benötigen wir Schätzer $G(X): X_{schätz} \to \mathbb{R}, x \mapsto [\ldots]$

[...]

Die Differentialrechnung liefert $\hat{\lambda}=\frac{x}{n}$, also $G(X)=\frac{1}{n}\cdot X$ (relative Häufigkeit) und Schätzung $\forall(p\in P(A)):\left(G(X)=\frac{x}{n}\right)$

## Intervallschätzung

Hier sucht man Intervalle zu bestimmen, in denen der gesuchte Parameter widehat S mit vorgegebener Sicherheit bei gegebenem Stichprobenumfang n liegt. Wichtige Größen sind:

- **Statistische Sicherheit** (Konfidenzniveau) $\kappa$
- **Vertrauensintervall** (Konfidenzintervall) $I_\alpha$
- **Irrtumswahrscheinlichkeit** $\epsilon=1-\alpha$

Dabei ist $I_\alpha=I_\alpha(x)$ mit $x\in X_{schätz}$ und es muss gelten:
$\forall(\delta\in\Theta):\left(P_\delta\left(\left\{x\in X_{schätz}\middle|\delta\in I_\alpha(x)\right\}\right)\geqslant\alpha\right)$

Für Werte des Parameters $\delta$ im Vertrauensbereich wird also die Wahrscheinlichkeit größer als $\alpha$. Deshalb sagt man: Der wahre Wert $\hat{\delta}$ liegt mit Mindestwahrscheinlichkeit $\alpha\in I_\alpha(x)$, wenn die Stichprobe für $X$ den Wert $x$ ergibt. Die Konstruktion des $I_\alpha(x)$ ist nicht so einfach. Sie findet man oft in Tabellen.

Eine andere typische Fragestellung ist: Man gebe sich ein $\alpha$ und ein Intervall $I_\eta=]\hat{\delta}-\eta,\hat{\delta}+\eta[$ vor. Man ermittelt dann aus $\alpha$ und $I_\eta$ den Umfang der Stichprobe $n$, so das der Schätzwert $\delta(x)$ für $\hat{\delta}$ mit Sicherheit $\alpha$ in $I_\eta$ liegt.

#### Beispiel

Intervallschätzung für die unbekannte Wahrscheinlichkeit $p=P(A)$. Ermittlung des Stichprobenumfangs.

#### Beispiel

Als Schätzwert für $p$ wird gewählt die Zahl $\frac{x}{n}$ (relative Häufigkeit).
$(X_{schätz}, B(X_{schätz}), P)$, $X_{schätz}=\{0,1,\ldots,n\}$, $P=\{B_{n,\lambda}|y\in[0,1]\}$.
Die Sicherheit $\alpha$ (Garantiewahrscheinlichkeit) und die Länge $\eta$ vom Intervall $I_\eta=]p-\eta,p+\eta[$ seien vorgegeben.
Das bedeutet: $P_\lambda(|X_n-\lambda|\geqslant\eta)=P_\lambda(X_n\notin I_\eta)\leqslant\epsilon=1-\alpha$ für $\lambda=p$ ($x_n=\frac{1}{n}\cdot x$ (relative Häufigkeit))

Wir kennen $\lambda=p$ nicht, also müssen wir diesen Ausdruck unabhängig vom Wert von % lambda abschätzen. Wir benutzen die Tschebyscheffsche Ungleichung und erhalten $P_\lambda(|X_n-\lambda|\geqslant\eta)\leqslant\frac{\lambda\cdot(1-\lambda)}{n\cdot\eta^2}\leqslant\frac{1}{4\cdot n\cdot\eta^2}$. Obige Abschätzung gilt also dann immer, wenn

$$\frac{1}{4 \cdot n \cdot \eta^2} \leqslant \epsilon \rightarrow n \geqslant \frac{1}{4 \cdot \epsilon \cdot \eta^2}$$

Zuordnung von Irrtumswahrscheinlichkeiten und Intervallgrößen zu notwendigem Stichprobenumfang, um Stichproben mit diesen Eigenschaften zu erhalten:

| $\eta \setminus \epsilon$ | 0,10 | 0,05 | 0,01 |
|---|---|---|---|
| 0,10 | 250 | 500 | 2500 |
| 0,05 | 1000 | 2000 | 10000 |
| 0,01 | 25000 | 50000 | 250000 |

Dies ist aber eine sehr grobe Abschätzung, da die Tschebyscheffsche Ungleichung sehr konservativ ist, weil sie für alle Verteilungen gilt. Kennen wir die Verteilung, so lassen sich feinere Abschätzungen finden. Beispielsweise gilt für $P_\lambda = B_{n,\lambda}$

| $\eta \setminus \epsilon$ | 0,10 | 0,05 | 0,01 |
|---|---|---|---|
| 0,10 | 150 | 185 | 265 |
| 0,05 | 600 | 750 | 1100 |
| 0,01 | 1500 | 1900 | 2700 |

Wenn $n$ gegeben ist und auch $\alpha$, so gibt es für große $n$ einfache grobe Abschätzungen für die Vertrauensintervalle $I_\alpha$, in denen die gesuchte Wahrscheinlichkeit $p = P(A)$ liegt. Wenn zum Beispiel $\alpha = 0.95$ gilt, so folgt $I_\alpha(x) = \left] p_0 - \frac{1}{\sqrt{n}}, p_0 + \frac{1}{\sqrt{n}} \right[$ mit $p_0 = \frac{x}{n}$ (relative Häufigkeit).

## 6.3 Testtheorie

In der Testtheorie geht es nun nicht mehr darum, den wahren Parameterwert $\delta_0$ aus dem statistischen Modell $(X_{schätz}, B(X_{schätz}), P)$ mit $P = \{P_\delta | \delta \in \Theta\}$ mit Hilfe der Beobachtungsdaten (Stichprobe) zu ermitteln, sondern nur noch festzustellen: Gehört $\delta$ zu einer bestimmten Klasse oder nicht? $\delta \in \Theta_0 \subset \Theta$ oder $\delta_0 \in \Theta_1 = \Theta \setminus \Theta_0$ ?

Viele Test (Hypothesenprüfungen) lassen ich in diesem Rahmen einbetten.

Es gibt also zwei Hypothesen:

- Hypothese $H_0$: $\delta_0 \in \Theta_0$
- Hypothese $H_1$: $\delta_0 \in \Theta_1$

Die Prüfung (der Test) der Hypothese $H_0$ erfolgt, indem man ein Ereignis $K \in B(X_{schätz})$ auswählt und zwar so, dass das Ergebnis $x \in K$ der Stichprobe sehr unwahrscheinlich ist, falls $\delta_0 \in \Theta_0$ richtig wäre. Diese Teilmenge $K$ von $X_{schätz}$ heißt **kritischer Bereich**.

Die Regel ist: Wenn das Ergebnis $x \in K$ vorliegt, wird die Hypothese $H_0$ a bgelehnt. Im anderen Fall ( $x \notin K$ ) ist erst einmal nichts gegen $H_0$ einzuwenden.

Es können zwei Fehlentscheidungen auftreten:

- Fehler 1. Art:

Die Hypothese $H_0$ wird abgelehnt, obwohl sie richtig ist. Wir beobachten also $x \in K$, obwohl $\delta_0 \in \Theta_0$. Die Wahrscheinlichkeit für diesen Fehler ist: $\forall(\delta_0 \in \Theta_0):¿$

- Fehler 2. Art:
  Die Hypothese $H_0$ wird nicht abgelehnt, obwohl sie falsch ist. Wir beobachten also $x \notin K$, obwohl $\delta_0 \notin \Theta_0$ ist. Die Wahrscheinlichkeit für diesen Fehler ist $\forall(\delta_0 \notin \Theta_0):\left(P_{\delta_0}(\{x \notin K\})=1-P_{\delta_0}(K)\right)$

Datum: 18.07.2003

**Beispiel: Test eines Würfels**

- 6 Ausgänge: Zahlen $1,2,\ldots,6$, $s=6$, $r=5$
- Anzahl der Würfe: $n=120$
- Die Hypothese ist: $\forall(i\in\{1,2,3,4,5,6\}):\left(\pi_j=\frac{1}{6}\right)$ (idealer Würfel)

| $a_j$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $H_j$ | 20 | 14 | 14 | 16 | 16 | 40 |

- $X_{120}(X)=\sum_{j=1}^{6}\left(\frac{1}{20}(H_j-20)^2\right)=25.2$
- Festlegung von $\alpha$: $\alpha=0.01$
  - also ist $C(6,0.01)=15.09<25.2$
  - Also ist der Würfel mit Wahrscheinlichkeit von 99% kein idealer Würfel.

# Was ist Zufall?

Betrachten wir die Zahlenfolge 8,9,7,9,3,2,3,8,4,6,2,6,4,3,3,8,3,2,7,9. Ist diese Zahlenfolge zufällig? Sie ist die 11. bis 30. Stelle der Dezimaldarstellung von $\pi$.

- Buchempfehlung: Paul Watzlawick: „Wie wirklich ist die Wirklichkeit?“

# Table of Contents