

W. Oevel

Einführung in die Stochastik

Veranstaltungsnr: 171050

Skript zur Vorlesung, Universität Paderborn, Sommersemester 2007

Homepage: <http://math-www.uni-paderborn.de/~walter>

→ Lehrveranstaltungen SS 07

V3 Mo 10.15 – 11.00 D1
Mi 11.00 – 12.30 D1

Ü2 Mo 7.30 – 9.00 D1.328 Gruppe 1 (Peter Brune)
Di 11.00 – 12.30 D1.320 Gruppe 2 (Peter Brune)

Inhalt

1	Grundstrukturen	1
1.1	Motivation	1
1.2	Grundlegende Definitionen (Modelle)	3
1.2.1	Kombinatorische Modelle	5
1.2.2	Diskrete (nicht-kombinatorische) Modelle	10
1.2.3	Kontinuierliche Modelle	11
1.3	Bedingte Wahrscheinlichkeiten	13
1.3.1	Definitionen	13
1.3.2	Folgerungen, „totale W'keit“ und der „Satz von Bayes“	16
1.4	Irrfahrten auf Graphen	20
1.4.1	Wahrscheinlichkeitsbäume	20
1.4.2	Allgemeinere Wahrscheinlichkeitsgraphen	25
1.5	Unabhängigkeit von Ereignissen	30
1.5.1	Definition	30
1.5.2	Modellierung unabhängiger Experimente: Produktmodelle	33
1.6	Einige Beispiele	37
2	Zufallsvariablen	43
2.1	Definitionen	44
2.2	Das Riemann-Stieltjes-Integral	55
2.3	Erwartungswert und Streuung	58
2.4	Einige Standardverteilungen	71
2.4.1	Die Binomial-Verteilung	71
2.4.2	Die Poisson-Verteilung	72
2.4.3	Die geometrische Verteilung	75
2.4.4	Die hypergeometrische Verteilung	75
2.4.5	Die Exponentialverteilung	76
2.4.6	Die Gleichverteilung	77
2.4.7	Die Normal-(Gauß-)Verteilung	78
2.5	Unabhängigkeit von Zufallsvariablen	80
2.5.1	Definition und Folgerungen	80

2.5.2	Kovarianz	87
2.6	Bedingte Erwartungswerte	88
3	Anwendungen: Laufzeitanalyse	91
3.1	MERGESORT	92
3.2	QUICKSORT	96
4	Grenzwertsätze	101
4.1	Das (schwache) Gesetz der großen Zahl	101
4.2	Die Moivre-Laplace-Näherung	107
4.3	Der Zentrale Grenzwertsatz	116

Literatur

Die wichtigste Referenz für diese Vorlesung:

[Kre91] ULRICH KRENGEL: Einführung in die Wahrscheinlichkeitstheorie und Statistik. Braunschweig/Wiesbaden: Vieweg 1991.

Einige sehr elementare, leicht lesbare Einführungen:

[Beh94] ERHARD BEHREND: Überall Zufall: eine Einführung in die Wahrscheinlichkeitsrechnung. Mannheim: BI 1994.

[DM87] WALTER DÜRR UND HORST MEYER: Wahrscheinlichkeitsrechnung und schließende Statistik. Wien: Hanser 1987.

[Pap99] LOTHAR PAPULA: Mathematik fuer Ingenieure und Naturwissenschaftler Band III. Braunschweig: Vieweg 1999. (P41 TBG2788)

Enthält viele einfache Aufgaben mit Musterlösungen:

[Spi75] MURRAY R. SPIEGEL: Probability and Statistics, (Schaum's Outline) McGraw-Hill, 1975 (P41 TKA 2611)

Enthält nette Aufgaben:

[Eng87] A. ENGEL: Stochastik. Stuttgart: Klett 1987. (P40 TQH1915)

Enthält nette Aufgaben:

[Isa95] RICHARD ISAAC: The pleasures of probability. New Yor: Springer 1995. (P41 TKA 4469)

Recht elementar:

[BHPT95] O. BEYER, H. HACKEL, V. PIEPER, J. TIEDGE: Wahrscheinlichkeitsrechnung und mathematische Statistik. Frankfurt: Harri Deutsch, 1995. (P41 TLK3417)

Übungsaufgaben mit Lösungen:

[GiN90] HEINZ GILLER UND VOLKER NOLLAU: Übungsaufgaben zur Wahrscheinlichkeitsrechnung und mathematischen Statistik. Leipzig: Teubner, 1990. (P40 TLK 3417)

Etwas anspruchsvoller. Enthält stochastische Analysen von Sortier-Algorithmen:

[MaP90] R. MATHAR UND D. PFEIFER: Stochastik für Informatiker. Stuttgart: Teubner, 1990. (P41 TKK6966)

Etwas anspruchsvoller.

[Pfa91] JOHANN PFANZAGL: Elementare Wahrscheinlichkeitsrechnung. Berlin: De-Gruyter 1991.

Interaktive Materialien im Internet:

Für diejenigen, die es 'multimedial' mögen. Sehr elementar und nicht in die Tiefe gehend, aber hübsch und liebevoll gemacht (ein Besuch ist zu empfehlen). Es gibt einige Module zur Stochastik:

www.matheprisma.uni-wuppertal.de

Kapitel 1

Grundstrukturen

1.1 Motivation

Ziel: Lösung folgender Problemtypen:

↓2.4.07

- 1) Einfache (kombinatorische) Modelle: Mit welcher W'keit (Wahrscheinlichkeit) ist die Augenzahl beim Wurf von drei Würfeln größer als 10? Was ist die W'keit für 6 Richtige im Lotto?

Bedingte W'keiten: Von zwei Münzen ist eine manipuliert. Man nimmt eine und wirft sie N mal, wobei k mal „Kopf“ auftritt. Mit welcher W'keit handelt es sich um die manipulierte Münze?

- 2) Zufallsvariablen, Erwartungswerte etc: Ein Gesamtsystem besteht aus N unabhängigen Teilsystemen, die pro Arbeitszyklus mit den W'keiten p_1, \dots, p_N ausfallen. Das Gesamtsystem ist funktionsfähig, wenn mindestens 90% seiner Teile arbeiten. Wieviele Zyklen wird das System „im Durchschnitt“ funktionsfähig bleiben?
- 3) Grenzwertsätze: Am Wahltag haben die Parteien bislang 42%, 40%, 7%, 6% der Stimmen erhalten (Hochrechnung vor Ende der Auszählung), der Nachrichtensprecher behauptet, die Stimmverteilung würden mit dem Endergebnis bis auf 1% übereinstimmen. Macht so eine Aussage Sinn (kritisches Bewusstsein)? Können die Grünen noch die absolute Mehrheit erlangen? Wenn ja, mit welcher W'keit? (Das Problem hierbei ist, dass das Wahlverhalten des Einzelwählers unbekannt ist. Durch Grenzwertsätze kann man dennoch Aussagen machen.)

Motivation zur Definition eines „stochastischen Modells“:

Beispiel 1.1: Bei einem „fairen“ Würfel sind die Ergebnisse $1, \dots, 6$ eines Wurfs „gleichwahrscheinlich“. Betrachte die Ereignisse:

- 1) Ich würfle eine gerade Zahl.
- 2) Ich würfle mindestens eine 3.
- 3) Ich würfle 2 mal, die Augensumme ist 10.

Wie wahrscheinlich sind diese Ereignisse? Intuitiv (bei gleichwahrscheinlichen „Elementarereignissen“):

$$P(\text{Ereignis}) = \frac{\text{Anzahl aller positiven Fälle}}{\text{Anzahl aller möglichen Fälle}}$$

Das Symbol P (engl: probability) steht in dieser Veranstaltung immer für „W'keit“. Mathematisierung:

- 1) Menge aller möglichen Fälle $\Omega_1 = \{1, 2, \dots, 6\}$.

Ereignis $E_1 =$ Menge aller positiven Fälle $= \{2, 4, 6\}$:

$$\Rightarrow P(E_1) = \frac{|E_1|}{|\Omega_1|} = \frac{3}{6} = \frac{1}{2} .$$

- 2) Menge aller möglichen Fälle $\Omega_2 = \{1, 2, \dots, 6\}$.

Ereignis $E_2 =$ Menge aller positiven Fälle $= \{3, 4, 5, 6\}$:

$$\Rightarrow P(E_2) = \frac{|E_2|}{|\Omega_2|} = \frac{4}{6} = \frac{2}{3} .$$

- 3) Menge aller möglichen Fälle $\Omega_3 = \{(1, 1), (1, 2), \dots, (6, 6)\}$.

Ereignis $E_3 =$ Menge aller positiven Fälle $= \{(5, 5), (6, 4), (4, 6)\}$:

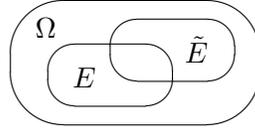
$$\Rightarrow P(E_3) = \frac{|E_3|}{|\Omega_3|} = \frac{3}{36} = \frac{1}{12} .$$

Strukturen, die diesen Beispielen gemeinsam sind:

- i) Eine Menge Ω , die „alle möglichen Fälle“ beschreibt.
- ii) Ein Ereignis E ist eine Teilmenge der Menge Ω aller möglichen Fälle.
- iii) Es ist ein „W'keitsmaß“ P gegeben, welches jedem Ereignis E eine W'keit $P(E) \in [0, 1]$ zuordnet.

In den Beispielen 1.1 war das W'keitsmaß jeweils $P(E) = |E|/|\Omega|$ (das „kombinatorische W'keitsmaß“), es hat die Eigenschaft

$$P(E \cup \tilde{E}) = P(E) + P(\tilde{E}) - P(E \cap \tilde{E}) .$$



1.2 Grundlegende Definitionen (Modelle)

↓4.4.07

Zunächst eine ganz formale Definition der mathematischen Situation, mit der W'keitsaufgaben modelliert werden. Diese abstrakte Definition wird dann in den folgenden Teilabschnitten („kombinatorische Modelle“ etc.) mit konkreten Inhalten und Beispielen gefüllt.

Definition 1.2: (Stochastisches Modell)

Ein **stochastisches Modell** (oder auch „**W'keitsraum**“ oder auch „**Experiment**“) ist ein Tripel (Ω, \mathcal{E}, P) bestehend aus

- i) einer beliebigen Menge $\Omega \neq \emptyset$ (der **Stichprobenraum**),
- ii) einer Menge $\mathcal{E} \subset \mathcal{P}(\Omega)$ von Teilmengen von Ω (die **Menge aller möglichen Ereignisse**). Die Elemente von \mathcal{E} (Teilmengen von Ω) heißen **Ereignisse**. \mathcal{E} muss die folgenden Eigenschaften haben:

- 1) Zu jedem Ereignis in \mathcal{E} liegt auch das **Komplementereignis** in \mathcal{E} :

$$E \in \mathcal{E} \quad \Rightarrow \quad \Omega \setminus E \in \mathcal{E}.$$

- 2) Die leere Menge \emptyset (und wegen 1. auch Ω) sind Ereignisse: $\emptyset =$ „das unmögliche Ereignis“, $\Omega =$ „das sichere Ereignis“.
- 3) Die Vereinigung endlich vieler oder auch abzählbar ∞ vieler Ereignisse ist wieder ein Ereignis:

$$E_1, E_2, \dots \in \mathcal{E} \quad \Rightarrow \quad \cup_i E_i \in \mathcal{E}.$$

- iii) einer Abbildung $P : \mathcal{E} \mapsto [0, 1]$ (das **W'keitsmaß**) mit den Eigenschaften

- 1) $P(\emptyset) = 0, P(\Omega) = 1$
- 2) Seien $E_1, E_2, \dots \in \mathcal{E}$ disjunkt, d.h. $E_i \cap E_j = \emptyset$ für $i \neq j$. Dann gilt (auch für abzählbar ∞ viele Ereignisse):

$$P(\cup_i E_i) = \sum_i P(E_i) \quad (\text{„}\sigma\text{-Additivität“}).$$

Bemerkung 1.3: In der mathematischen Literatur und in weiterführenden Vorlesungen zur Stochastik wird eine Menge $\mathcal{E} \subset \mathcal{P}(\Omega)$ mit den Eigenschaften 1.2.ii.1 - 1.2.ii.3 eine **Sigma-Algebra** in der Menge Ω genannt.

Bemerkung 1.4: In dieser Vorlesung reicht es meist aus, sich unter \mathcal{E} die Potenzmenge $\mathcal{P}(\Omega)$ (also die Menge aller Teilmengen) von Ω vorzustellen, d.h., jede Teilmenge von Ω ist ein zulässiges Ereignis. Trivialerweise erfüllt $\mathcal{E} = \mathcal{P}(\Omega)$ die in Definition 1.2.ii geforderten Eigenschaften.

In einigen Modellen ist die Einschränkung auf gewisse Teilmengen aus rein technischen Gründen nötig: wenn W'keiten durch Integrale gegeben sind, dürfen nur Teilmengen betrachtet werden, über denen Integrale definiert werden können.

Mit den in Definition 1.2.ii geforderten Eigenschaften von \mathcal{E} folgen automatisch zahlreiche weitere Eigenschaften, z.B.:

$$\begin{aligned} E_1, E_2 \in \mathcal{E} &\Rightarrow \Omega \setminus E_1 \text{ und } \Omega \setminus E_2 \in \mathcal{E} \Rightarrow (\Omega \setminus E_1) \cup (\Omega \setminus E_2) \in \mathcal{E} \\ &\Rightarrow \Omega \setminus \left((\Omega \setminus E_1) \cup (\Omega \setminus E_2) \right) = \Omega \setminus \left(\Omega \setminus (E_1 \cap E_2) \right) = E_1 \cap E_2 \in \mathcal{E} . \end{aligned}$$

Damit sind nicht nur Vereinigungen, sondern auch Durchschnitte von (endlich vielen oder abzählbar unendlich vielen) Ereignissen wieder Ereignisse.

Bemerkung 1.5: Das W'keitsmaß P werden wir (im ersten Teil der Vorlesung) stets explizit vorgeben (z.B. durch kombinatorisches Abzählen oder als Modellvorgabe). Später werden wir über Grenzwertsätze auch Aussagen bekommen, ohne das Maß genau zu kennen.

Folgerung (einige allgemeingültige Rechenregeln) 1.6:

Aus den in Definition 1.2.iii gegebenen Eigenschaften von P folgen unmittelbar einige allgemeingültige Regeln für das Rechnen mit W'keiten:

a) Für disjunkte Ereignisse ($E_1 \cap E_2 = \emptyset$) gilt in jedem Modell:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) .$$

b) Allgemein gilt in jedem Modell:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) .$$

c) Für komplementäre Ereignisse E und $\Omega \setminus E$ gilt in jedem Modell:

$$P(\Omega \setminus E) = 1 - P(E) .$$

d) Für $E_1 \subset E_2$ gilt¹ in jedem Modell $P(E_1) \leq P(E_2)$.

¹Achtung: es gilt nicht $P(E_1) < P(E_2)$ für $E_1 \subset E_2$, $E_1 \neq E_2$. Gegenbeispiel: Ich entscheide

Beweis: a) Folgt unmittelbar aus der Definition 1.2.

b) Benutze die disjunkten Zerlegungen

$$\begin{aligned} E_1 \cup E_2 &= (E_1 \setminus E_2) \cup (E_1 \cap E_2) \cup (E_2 \setminus E_1), \\ E_1 &= (E_1 \cap E_2) \cup (E_1 \setminus E_2), \\ E_2 &= (E_1 \cap E_2) \cup (E_2 \setminus E_1). \end{aligned}$$

Mit a) folgt

$$\begin{aligned} P(E_1 \cup E_2) &= P(E_1 \setminus E_2) + P(E_1 \cap E_2) + P(E_2 \setminus E_1), \\ P(E_1) &= P(E_1 \cap E_2) + P(E_1 \setminus E_2), \\ P(E_2) &= P(E_1 \cap E_2) + P(E_2 \setminus E_1). \end{aligned}$$

Subtrahiert man die zweite und dritte Gleichung von der ersten, erhält man

$$P(E_1 \cup E_2) - P(E_1) - P(E_2) = -P(E_1 \cap E_2).$$

c) folgt mittels

$$\Omega = E \cup (\Omega \setminus E) \stackrel{a)}{\Rightarrow} 1 = P(\Omega) = P(E) + P(\Omega \setminus E).$$

d) folgt mittels

$$E_1 \subset E_2 \Rightarrow E_2 = E_1 \cup (E_2 \setminus E_1) \stackrel{a)}{\Rightarrow} P(E_2) = P(E_1) + P(E_2 \setminus E_1) \geq P(E_1).$$

Q.E.D.

1.2.1 Kombinatorische Modelle

Eine kombinatorische Betrachtungsweise bietet sich für Probleme an, die aus endlich vielen *gleichwahrscheinlichen* Elementarereignissen bestehen (Beispiele 1.1):

Definition 1.7: (Kombinatorisches Modell)

Ein **kombinatorisches Modell** (Ω, \mathcal{E}, P) , auch „**Laplace-Experiment**“ genannt, besteht aus

- i) einer endlichen Menge $\Omega = \{\omega_1, \dots, \omega_n\}$,
- ii) den Ereignissen $\mathcal{E} = \mathcal{P}(\Omega) = \text{Potenzmenge von } \Omega$,
- iii) dem W'keitsmaß $P(E) = |E|/|\Omega|$ (das „**Zählmaß**“).

Die Elemente ω_i des Stichprobenraums Ω (genauer: die einelementigen Teilmengen $\{\omega_i\}$) heißen „**Elementarereignisse**“.

mich, einen Würfel durch $\Omega = \{1, \dots, 7\}$ zu modellieren (das ist meine legitime Entscheidung: ist zwar unnötig kompliziert, aber zulässig). Damit dieses Modell realistisch ist, muss ich $P(\{7\}) = 0$ setzen, es gilt $P(\{1, \dots, 6\}) = P(\{1, \dots, 7\}) = 1$.

Man kann sich unschwer überlegen, dass dies in der Tat ein stochastisches Modell im Sinne der Definition 1.2 ist: die Ereignismenge $\mathcal{E} = \mathcal{P}(\Omega)$ und das Maß P erfüllen die in 1.2.ii und 1.2.iii geforderten Eigenschaften.

Die Definition 1.7.iii) liefert den sauberen mathematischen Kontext für das intuitive Konzept

$$P(\text{Ereignis}) = \frac{\text{Anzahl aller positiven Fälle}}{\text{Anzahl aller möglichen Fälle}} .$$

In konkreten Aufgabenstellungen gilt es, in der „Modellierungsphase“ ein geeignetes Ω und ein Ereignis $E \subset \Omega$ zu konstruieren, das der Aufgabe entspricht. Die „mathematische“ Herausforderung ist dann, die Anzahl der Elemente $|E|$ und $|\Omega|$ zu bestimmen, um die W'keit des Ereignisses zu berechnen.

Beispiel 1.8: Lottoziehung: Ich mache einen Tip $T = \{t_1, \dots, t_6\}$ von 6 unterschiedlichen Zahlen t_i aus $\{1, \dots, 49\}$. Eine Ziehung ist die Auswahl einer 6-elementigen Teilmenge der Menge $\{1, \dots, 49\}$, also:

$$\Omega = \text{Menge aller Ziehungen „6 aus 49“} = \{Z \subset \{1, \dots, 49\}; |Z| = 6\} .$$

Alle Ziehungen sind intuitiv gleichwahrscheinlich. Das Ereignis $E =$ „6 Richtige“ bedeutet, dass die Ziehung mit meinem Tip T übereinstimmt, also: $E = \{T\}$. Mit $|\Omega| = \binom{49}{6} = \frac{49!}{6! \cdot 43!} = 13\,983\,816$ folgt $P(E) = 1/13\,983\,816$.

Hierbei ist die Aussage $|\Omega| = \binom{49}{6}$ direkt dem folgenden Hilfssatz zu entnehmen:

Hilfssatz 1.9: (Hilfsmittel für kombinatorische Aufgaben)

- a) Es gibt $n!$ unterschiedliche Anordnungen („Permutationen“) von n unterscheidbaren Objekten.
- b) Es gibt $\binom{n}{m} = \frac{n!}{m! (n-m)!}$ verschiedene m -elementige Teilmengen einer Menge mit n Elementen.
- c) (Verallgemeinerung von b)) Es gibt

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_m!}$$

verschiedene Möglichkeiten, eine Menge mit $n = n_1 + \dots + n_m$ Elementen in m durchnummerierte Teilmengen zu jeweils n_1, \dots, n_m Elementen zu zerlegen (die n_i dürfen dabei auch 0 sein, wobei $0! = 1$ gesetzt wird).

- d) Es gibt $\binom{n+m-1}{n}$ Möglichkeiten, n nicht unterscheidbare Objekte auf m durchnummerierte Boxen aufzuteilen (wobei Boxen leer bleiben können).

Beweisskizze: a) Es gibt n Möglichkeiten, das „erste Element“ der Anordnung zu wählen. Dann gibt es $n - 1$ Möglichkeiten, das „zweite Element“ der Anordnung aus den verbleibenden Objekten zu wählen. Danach $n - 2$ Möglichkeiten für das nächste Element usw. Macht zusammen $n \cdot (n - 1) \cdot (n - 2) \cdot \dots = n!$ Möglichkeiten.

c) Die n Elemente werden in eine Reihe gelegt, die an den ersten n_1 Stellen liegenden Elemente kommen in die erste Teilmenge, die nächsten n_2 Elemente in die zweite Teilmenge usw. Es gibt nach a) insgesamt $n!$ Möglichkeiten, eine solche Reihe zu erzeugen. Werden die ersten n_1 Elemente als Menge interpretiert, so ist ihre Reihenfolge egal, d.h., nach a) sind $n_1!$ der Anordnungen zu identifizieren. Werden auch die nächsten n_2 Elemente als Menge interpretiert, so sind jeweils $n_1! n_2!$ der Anordnung zu identifizieren. Usw.

b) Spezialfall von c): zerlege die Menge in zwei Teilmengen mit jeweils $n_1 = m$ und $n_2 = n - m$ Elementen. Nach c) gibt es $n!/m!(n - m)!$ Möglichkeiten.

d) Ich lege die Objekte in eine Reihe, lege ganz links und ganz rechts einen „Trennbalken“ und sortiere $m - 1$ weitere Trennbalken zwischen den Objekten ein. Der Raum zwischen zwei Trennbalken wird als Box angesehen (es liegen nun m Boxen vor). Da ich die Objekte nicht unterscheiden kann, ist die Aufteilung auf die Boxen eindeutig durch die Lage der Trennbalken gegeben. Es gibt nach a) $\binom{n+m-1}{m-1}$ Möglichkeiten, die Lage der „inneren“ $m - 1$ Trennbalken aus den insgesamt $n + m - 1$ möglichen Positionen in der Reihe von n Objekten und $m - 1$ Trennbalken zu wählen. Damit ist die Anzahl der Boxenaufteilungen

$$\binom{n+m-1}{m-1} \equiv \binom{n+m-1}{n}.$$

Q.E.D.

Beispiel 1.10:

Eine Urne enthält N Kugeln, davon S schwarz und $W = N - S$ weiß. Es werden n ($\leq N$) Kugeln *ohne Zurücklegen* gezogen. Mit welcher W'keit sind s der gezogenen Kugeln schwarz und $w = n - s$ weiß?

Lösung: Seien

$$\mathcal{W} = \{w_1, \dots, w_W\}, \quad \mathcal{S} = \{s_1, \dots, s_S\}$$

die weißen bzw. schwarzen Kugeln. Der Stichprobenraum sei die Menge aller Ziehungen einer n -elementigen Teilmenge aus der Urne $\mathcal{U} = \mathcal{S} \cup \mathcal{W}$:

$$\Omega = \{Z \subset \mathcal{S} \cup \mathcal{W}; |Z| = n\}.$$

Das Ereignis $E =$ „ s Kugeln sind schwarz, $w = n - s$ Kugeln sind weiß“ besteht aus den folgenden Ziehungen:

$$E = \{Z \subset \mathcal{S} \cup \mathcal{W}; Z = \tilde{\mathcal{S}} \cup \tilde{\mathcal{W}}; \tilde{\mathcal{S}} \subset \mathcal{S}; |\tilde{\mathcal{S}}| = s; \tilde{\mathcal{W}} \subset \mathcal{W}; |\tilde{\mathcal{W}}| = w\}.$$

↓11.4.07

Nach 1.9.b) gilt $|\Omega| = \binom{N}{n}$. Es gibt es $\binom{S}{s}$ Möglichkeiten, die s -elementige Teilmenge \tilde{S} aus der S -elementigen Menge S zu wählen. Es gibt es $\binom{W}{w}$ Möglichkeiten, die w -elementige Teilmenge \tilde{W} aus der W -elementigen Menge W zu wählen. Damit folgt $|E| = \binom{S}{s} \binom{W}{w}$, also

$$P(E) = \frac{\binom{S}{s} \binom{W}{w}}{\binom{N}{n}}.$$

Statt von einer „Urne“ redet man auch von einer „**Gesamtpopulation**“ von N Elementen, die sich in eine „**Erfolgs(teil)population**“ von S Elementen und eine komplementäre „**Misserfolgs(teil)population**“ von $W = N - S$ Elementen zerlegt. Wählt man n Elemente ohne Zurücklegen aus, so ist die W 'keit, dabei genau s Erfolge und $w = n - s$ Misserfolge zu ziehen, gegeben durch

$$P(\text{„genau } s \text{ Erfolge bei } n \text{ Ziehungen“}) = \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}.$$

Man nennt die Abbildung $s \rightarrow P(\text{„genau } s \text{ Erfolge bei } n \text{ Ziehungen“})$ die „**hypergeometrische Verteilung**“ mit den vorgegebenen Parametern N (Populationsgröße), S (Erfolgspopulationsgröße), n (Anzahl der Ziehungen).

Z.B.: Mit welcher W 'keit erhält man beim Skat s Asse? Es werden $n = 10$ Karten aus $N = 32$ Karten an mich verteilt. Unter den 32 Karten befinden sich $S = 4$ Asse. Die W 'keit, s Asse zu erhalten, ist durch die hypergeometrische Verteilung gegeben:

$$P(\text{„}s \text{ Asse“}) = \frac{\binom{4}{s} \binom{32-4}{10-s}}{\binom{32}{10}}.$$

Intuitiv: es gibt $\binom{4}{s}$ Möglichkeiten, von den 4 Assen s zu erhalten, es gibt $\binom{28}{10-s}$ Möglichkeiten, die restlichen $10 - s$ Karten meiner Hand aus den verbleibenden 28 Karten zu erhalten, es gibt $\binom{32}{10}$ Möglichkeiten, insgesamt 10 Karten aus 32 zu erhalten.

Beispiel 1.11: Mit mir haben sich insgesamt 22 Leute für ein Turnier gemeldet. Es werden willkürlich 4 Gruppen G_1, \dots, G_4 mit $|G_1| = |G_2| = 5$, $|G_3| = |G_4| = 6$ gebildet, die nacheinander die 1-te Runde austragen sollen. Es beginnt G_1 um 8^{oo}. Mit welcher W 'keit muss ich früh aufstehen?

Lösung: a) Intuitiv und einfach: ich stelle mir vor, ich ziehe eine Kugel aus einer Urne mit 22 Kugeln, auf denen die Ziffern 1 bis 22 stehen. Die Ziffern 1 bis 5 bedeuten G_1 , die Ziffern 6 bis 10 bedeuten G_2 usw.:

$$\Omega = \{1, 2, \dots, 22\}, \quad E = \{1, 2, \dots, 5\} \quad \Rightarrow \quad P(E) = \frac{|E|}{|\Omega|} = \frac{5}{22}.$$

b) Alternative (formalere) Lösung. Der Stichprobenraum sei die Menge aller möglichen Gruppeneinteilungen

$$\Omega = \{(G_1, G_2, G_3, G_4); G_1 \cup \dots \cup G_4 = \{1, \dots, 22\}, |G_1| = |G_2| = 5, |G_3| = |G_4| = 6\}.$$

Nach 1.9.c) gilt $|\Omega| = \frac{22!}{5! 5! 6! 6!} = 150\,570\,227\,808$. In wievielen dieser Möglichkeiten ist eine bestimmte Person (ich, mit der Nummer 22, sagen wir) in der Gruppe G_1 ? Dies entspricht den Möglichkeiten, die anderen 21 Leute in Gruppen $\tilde{G}_1, G_2, G_3, G_4$ mit $|\tilde{G}_1| = 4, |G_2| = 5, |G_3| = |G_4| = 6$ aufzuteilen. Das Ereignis „ich werde in G_1 eingeteilt“ entspricht damit

$$E = \{(\tilde{G}_1 \cup \{22\}, G_2, G_3, G_4); \tilde{G}_1 \cup \dots \cup G_4 = \{1, \dots, 21\}, \\ |\tilde{G}_1| = 4, |G_2| = 5, |G_3| = |G_4| = 6\}.$$

Nach 1.9.c) gilt $|E| = \frac{21!}{4! 5! 6! 6!} = 34\,220\,506\,320$. Die W'keit, dass ich früh aufstehen muss, ist damit

$$P(E) = \frac{|E|}{|\Omega|} = \frac{21!}{4! 5! 6! 6!} \cdot \frac{5! 5! 6! 6!}{22!} = \frac{21!}{4!} \cdot \frac{5!}{22!} = \frac{5}{22} \approx 0.227\dots$$

Beispiel 1.12: Eine Abzählübung: Wieviele Möglichkeiten gibt es, eine ganze Zahl n in genau m verschiedene Summanden $n = n_1 + \dots + n_m$ mit $n_i \in \{0, 1, 2, \dots\}$ zu zerlegen? Die Reihenfolge der Summanden soll dabei berücksichtigt werden.

Z.B. $n = 3, m = 2$: 4 Möglichkeiten $3 = 0 + 3 = 1 + 2 = 2 + 1 = 3 + 0$.

Z.B. $n = 4, m = 2$: 5 Möglichkeiten $4 = 0 + 4 = 1 + 3 = 2 + 2 = 3 + 1 = 4 + 0$.

Lösung: Dies lässt sich unmittelbar auf den Hilfssatz 1.9.d) zurückführen. Wir wiederholen das Argument:

Symbolisiere jeden Summanden n_i durch n_i aufgereihete Einsen. Diese „Einser-Reihen“ werden dann mit Pluszeichen verbunden. Es bleibt damit das Problem, alle Möglichkeiten zu finden, eine Kette aus n Einsen und $m - 1$ Pluszeichen anzuordnen. Da weder die Einsen noch die Pluszeichen unterschieden werden können, ist die Aufteilung durch die Position der $m - 1$ Pluszeichen in der Gesamtkette von $n + m - 1$ Objekten gegeben. Die Position der Pluszeichen entspricht einer Auswahl von $m - 1$ Platznummern aus den möglichen Plätzen $1, 2, \dots, n + m - 1$. Nach 1.9.b) ist die Anzahl der Möglichkeiten

$$\binom{n + m - 1}{m - 1} = \binom{n + m - 1}{n}.$$

Merke (Zusammenfassung) 1.13:

Ein kombinatorische Modell bietet sich in Situationen an, in denen alle Elementarereignisse des Zufallsexperimentes gleichwahrscheinlich sind: das Modell besteht aus einer Menge Ω endlich vieler Elemente, denen jeweils die gleiche W'keit $1/|\Omega|$ zugeschrieben wird. Für ein Ereignis $E \subset \Omega$ gilt

$$P(E) = \frac{|E|}{|\Omega|}.$$

Sobald Ω und E vom Modellierer festgelegt sind, besteht die Herausforderung darin, durch geschicktes Abzählen $|E|$ und $|\Omega|$ zu ermitteln.

1.2.2 Diskrete (nicht-kombinatorische) Modelle

Bei endlich vielen (oder abzählbar unendlich vielen) Elementarereignissen, die *nicht gleichwahrscheinlich* sind, bietet sich folgendes diskrete Modell an:

Definition 1.14: (Diskretes Modell)

Ein **diskretes Modell** (Ω, \mathcal{E}, P) besteht aus

- i) einer endlichen oder abzählbar unendlichen Menge Menge $\Omega = \{\omega_1, \omega_2, \dots\}$ (die ω_i heißen wieder „**Elementarereignisse**“),
- ii) den Ereignissen $\mathcal{E} = \mathcal{P}(\Omega) =$ Potenzmenge von Ω ,
- iii) dem folgenden *W'keitsmaß*: Es sind *W'keitswerte* $P(\{\omega_i\})$ für die Elementarereignisse vorgegeben, für die $\sum_{\omega \in \Omega} P(\omega) = 1$ gilt. Dann wird für ein Ereignis $E \subset \Omega$ die folgende *W'keit* definiert:

$$P(E) = \sum_{\omega \in E} P(\omega) .$$

Die kombinatorischen Modelle 1.7 sind der Spezialfall, wo für alle $\omega \in \Omega$ dieselbe *W'keit* $P(\{\omega\}) = 1/|\Omega|$ gewählt wird.

Beispiel 1.15: Ein Würfel ist manipuliert, die *W'keiten* p_i , die Zahl i zu würfeln, seien

i	1	2	3	4	5	6
p_i	0.1	0.1	0.1	0.2	0.2	0.3

Mit welcher *W'keit* würfle ich mindestens eine 4? Antwort:

$$\Omega = \{1, \dots, 6\} ; \quad E = \{4, 5, 6\} ; \quad P(E) = p_4 + p_5 + p_6 = 0.7 .$$

Beispiel 1.16: Ich werfe zweimal mit einem Würfel. Mit welcher *W'keit* erhalte ich die Augensumme 10?

a) Ich behandle dies kombinatorisch:

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}, \quad E = \{(4, 6), (5, 5), (6, 4)\}$$

$$\Rightarrow P(\text{„Augensumme 10“}) = \frac{|E|}{|\Omega|} = \frac{3}{36} = \frac{1}{12} .$$

b) Alternativ kann ich die Augensumme (eine Zahl zwischen 2 und 12) direkt als Ergebnis des Experiments „Doppelwurf“ ansehen, also

$$\Omega = \{2, 3, \dots, 12\} .$$

Dieser Stichprobenraum ist mit 11 Elementen wesentlich kleiner als der in a) gewählte Stichprobenraum mit 36 Elementen. Allerdings habe ich hier nun das Problem, dass die

Elementarereignisse nicht mehr gleichwahrscheinlich sind. Ich muss mir die W'keiten p_2, \dots, p_{12} aus der kombinatorischen Sicht a) startend konstruieren, damit das Modell realistisch ist:

$$\begin{aligned} p_2 &= P(\text{„Augensumme 2“}) = P(\{(1, 1)\}) = \frac{1}{36}, \\ p_3 &= P(\text{„Augensumme 3“}) = P(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18}, \\ p_4 &= \dots, \\ &\vdots \\ p_{10} &= P(\text{„Augensumme 10“}) = P(\{(4, 6), (5, 5), (6, 4)\}) = \frac{3}{36} = \frac{1}{12}, \\ &\vdots \end{aligned}$$

Es folgt $P(\text{„Augensumme 10“}) = P(\{10\}) = p_{10} = \frac{1}{12}$.

Merke (Zusammenfassung) 1.17:

Ein **diskretes Modell** entspricht einem Experiment mit endlich vielen oder abzählbar unendlich vielen Ausgängen, die unterschiedliche W'keiten haben können. Kennt man die W'keiten der Elementarereignisse (die möglichen Ausgänge des Experiments), so ergibt sich die W'keit eines Ereignisses als die Summe der W'keiten aller Elementarereignisse, aus denen das Ereignis zusammengesetzt ist:

$$P(E) = \sum_{\omega \in E} P(\{\omega\}).$$

1.2.3 Kontinuierliche Modelle

Bei nicht diskreten Situationen werden W'keiten typischerweise über Integrale definiert:

Definition 1.18: (Kontinuierliches Modell)

Ein **kontinuierliches Modell** (Ω, \mathcal{E}, P) besteht aus

- i) einer Teilmenge Ω von \mathbb{R} oder \mathbb{R}^2 oder ...,
- ii) den Ereignissen $\mathcal{E} = \{E \subset \Omega; \text{„man kann über } E \text{ integrieren“}\}$,
- iii) dem folgenden W'keitsmaß: es ist eine Funktion $\rho : \Omega \mapsto [0, \infty)$ gegeben: die **W'keitsdichte**. Sie muss $\int_{x \in \Omega} \rho(x) dx = 1$ erfüllen. Dann wird für ein Ereignis $E \in \mathcal{E}$ die folgende W'keit definiert:

$$P(E) = \int_{x \in E} \rho(x) dx = \int_{x \in \Omega} \chi_E(x) \cdot \rho(x) dx,$$

wobei

$$\chi_E(x) = \begin{cases} 1 & \text{für } x \in E \\ 0 & \text{für } x \notin E \end{cases}$$

die **charakteristische Funktion** der Teilmenge $E \subset \Omega$ ist.

16.4.07↓

Bemerkung 1.19: In dieser Situation darf man in der Tat nicht alle Teilmengen von Ω als Ereignisse zulassen, da es sonst mathematische Schwierigkeiten beim Integrieren gibt. Hierauf werden wir uns aber nicht einlassen. Man stelle sich für $\Omega = \mathbb{R}$ als zulässige Ereignisse die Intervalle sowie Vereinigungen bzw. Schnitte von Intervallen vor. Bei uns werden nur Vereinigungen bzw. Schnitte endlich vieler disjunkter Intervalle vorkommen. Dann gilt für $E = [a_1, b_1] \cup [a_2, b_2] \cup \dots$:

$$\int_{x \in E} \rho(x) dx = \int_{a_1}^{b_1} \rho(x) dx + \int_{a_2}^{b_2} \rho(x) dx + \dots$$

Bemerkung 1.20: Damit über die Dichte ρ integriert werden kann, muss sie gewisse Glattheitseigenschaften erfüllen, worauf hier aber nicht eingegangen wird. Man beachte aber, dass in Anwendungen unstetige Dichten mit Sprüngen eine wichtige Rolle spielen. Es reicht hier, sich ρ als „stückweise stetig“ mit endlich vielen Sprungstellen vorzustellen.

Beispiel 1.21: Die W'keit, eine Maus im Abstand $r \in \Omega = [0, \infty)$ von ihrem Mausloch zu finden, sei durch die Dichte

$$\rho(r) = 2c r e^{-cr^2}$$

mit einem positiven Parameter $c > 0$ gegeben. Der konstante Faktor $2c$ sorgt dafür, dass die Normierung $\int_{r \in \Omega} \rho(r) dr = \int_0^\infty \rho(r) dr = 1$ gilt. Für das Ereignis $[r_1, r_2]$ („die Maus befindet sich in einem Abstand zwischen r_1 und r_2 von ihrem Loch“) ergibt sich die W'keit

$$P([r_1, r_2]) = \int_{r_1}^{r_2} \rho(r) dr = e^{-cr_1^2} - e^{-cr_2^2}.$$

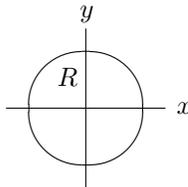
Kleine Werte von c entsprechen einer „mutigen Maus“, die mit großer W'keit weit weg von ihrem Loch zu finden ist: die W'keit

$$P([r, \infty)) = e^{-cr^2},$$

dass sie sich im Abstand $\geq r$ vom Loch aufhält, ist monoton fallend in c .

Beispiel 1.22: (Hier muss man wissen, wie man im \mathbb{R}^2 integriert).

Ein Schütze schießt auf eine im Ursprung von $\Omega = \mathbb{R}^2$ zentrierte Schießscheibe vom Radius R :



Die W'keit, dass ein kleines Flächenstück an der Stelle (x, y) getroffen wird, sei durch die Dichte

$$\rho(x, y) = \frac{c}{\pi} e^{-c(x^2+y^2)}$$

gegeben. Er trifft die Zielscheibe mit der W'keit

$$P(\text{Treffer}) = \int_{|(x,y)| \leq R} \rho(x, y) dx dy$$

$$\stackrel{\text{(Polarkoordinaten)}}{=} \int_0^R \int_0^{2\pi} \frac{c}{\pi} e^{-c r^2} r dr d\phi = 1 - e^{-c R^2}.$$

Merke (Zusammenfassung) 1.23:

Ein **kontinuierliches Modell** besteht aus einer Teilmenge Ω des \mathbb{R}^n (in Zukunft bei uns i.W. $n = 1$), auf dem eine **W'keitsdichte** $\rho(x) \geq 0$ gegeben ist, die

$$\int_{x \in \Omega} \rho(x) dx = 1$$

erfüllen muss. Die W'keit eines Ereignisses E ist dann

$$P(E) = \int_{x \in E} \rho(x) dx.$$

1.3 Bedingte Wahrscheinlichkeiten

1.3.1 Definitionen

Zur Motivation:

Beispiel 1.24: Ich werfe ein Münze: das Ergebnis ist „Kopf“ oder „Zahl“ mit den W'keiten $P(K) = P(Z) = 1/2$.

- Mit welcher W'keit habe ich nach n Würfeln n mal das Ergebnis „Kopf“?
- Ich habe schon $n - 1$ mal „Kopf“ geworfen. Mit welcher W'keit werfe ich bei nächsten mal „Kopf“?

Frage a) hat offensichtlich die kombinatorische Antwort $P(n \text{ mal „Kopf“}) = 1/2^n$: es gibt bei n Würfeln 2^n mögliche verschiedene Ergebnisse

$$\Omega = \{(\omega_1, \dots, \omega_n) ; \omega_i \in \{K, Z\}\},$$

davon beschreibt nur ein einziges mein Ereignis $E = \{(K, \dots, K)\}$. Diese W'keit wird für wachsendes n sehr schnell sehr klein.

Frage b) ist offensichtlich intuitiv unabhängig von n mit $P(K) = 1/2$ zu beantworten, denn die ersten $n - 1$ Würfe haben nichts mit dem n -ten Wurf zu tun (Unabhängigkeit).

Es wird innerhalb desselben Experiments nach unterschiedlichen Dingen gefragt:

- a) $P(\text{Ereignis: } n \text{ mal „Kopf“})$
- b) $P(\text{Ereignis: } n \text{ mal „Kopf“; Zusatzinformation: „Kopf“ ist schon } n - 1 \text{ mal aufgetreten})$

Frage: wie kann ich b) im Rahmen des für a) benutzten Modells Ω beschreiben? Das Experiment ist ja dasselbe, also sollte die Modellierung (Ω, \mathcal{E}, P) für beide Fälle benutzbar sein.

Wir betrachten die allgemeine Situation eines Modells (Ω, \mathcal{E}, P) . Es geht darum: ich weiß, dass ein Ereignis eingetreten ist. Was nützt diese Information, wenn ich nach der W'keit für ein anderes Ereignis frage?

Definition 1.25: (bedingte W'keit)

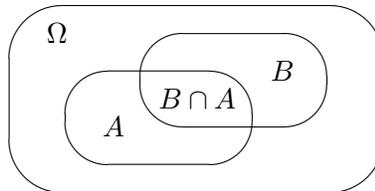
Seien $A, B \in \mathcal{E}$. Setze

$$P(B|A) = \begin{cases} \frac{P(A \cap B)}{P(A)} & \text{für } P(A) \neq 0, \\ 0 & \text{für } P(A) = 0 \end{cases}$$

und nenne es die **bedingte W'keit für das Ereignis B bei gegebenem Ereignis A** .

Interpretation (wichtige Modellierungshilfe) 1.26:

Dieser Begriff modelliert in der Tat die Situation, dass das Ereignis A eingetreten ist. Wenn ich nun nach der W'keit von B frage, so ist die Antwort das soeben definierte $P(B|A)$. Der Gedanke ist:



Ist A eingetreten, so befinden wir uns nicht mehr irgendwo in Ω , sondern in der Teilmenge A . Wenn B eintritt, so wissen wir, dass in Wirklichkeit das Ereignis $B \cap A$ eingetreten ist. Daher können wir A als neuen W'keitsraum interpretieren, welches jedem $B \subset \Omega$ die „reduzierte“ W'keit $P(B \cap A)$ zuschreibt. Dabei müssen wir diese W'keit aber noch mit einem Faktor f multiplizieren, damit das neue W'keitsmaß $P_A(B) = f \cdot P(B \cap A)$ auf A wieder normiert ist. Da $P_A(A) = 1$ gelten muss, folgt $f = 1/P(A)$. Also:

$$P_A(B) = \frac{1}{P(A)} \cdot P(B \cap A) = P(B|A) .$$

Merke:

Die bedingte W'keit $P(B|A)$ entspricht einer „normalen“ W'keit $P_A(B)$ auf einem geänderten (reduzierten) Modell $\Omega_A = A$, das nur aus Teilereignissen des „als eingetreten vorausgesetzten“ Ereignisses A besteht.

Bemerkung 1.27: Diese Interpretation ist auch mathematisch stichhaltig. Es gilt der leicht beweisbare Satz (Übungsaufgabe):

Für einen W'keitsraum (Ω, \mathcal{E}, P) und gegebenes $A \in \mathcal{E}$ mit $P(A) > 0$ ist $(\Omega_A, \mathcal{E}_A, P_A)$ mit $\Omega_A = A$, $\mathcal{E}_A = \{E \cap A; E \in \mathcal{E}\}$ und $P_A(B) = P(B|A)$ wieder ein W'keitsraum nach Definition 1.2.

Beispiel 1.28: Zurück zu Beispiel 1.24: n -facher Wurf einer Münze, $\Omega = \{(\omega_1, \dots, \omega_n); \omega_i \in \{K, Z\}\}$ mit dem Zählmaß.

a) Das Ereignis $B =$ „es wurde n mal K geworfen“ = $\{(K, \dots, K)\}$ hat die W'keit

$$P(B) = |B|/|\Omega| = 1/2^n .$$

b) Das Ereignis $A =$ „in den ersten $n - 1$ Würfeln wurde K geworfen“ ist $A = \{(K, \dots, K, \omega_n); \omega_n \in \{K, Z\}\}$. In b) wird nach $P(B|A)$ gefragt:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{|B|/|\Omega|}{|A|/|\Omega|} = \frac{|B|}{|A|} = \frac{1}{2} .$$

Hier habe ich A als Teilmenge von Ω aufgefasst. Ich kann mich auch auf den Standpunkt stellen, dass ich, wenn A eingetreten ist, A als neuen Stichprobenraum auffasse, den ich vereinfachen würde:

$$\Omega_A = A = \{(K, \dots, K, \omega_n); \omega_n \in \{K, Z\}\} \equiv \tilde{\Omega} = \{\omega_n; \omega_n \in \{K, Z\}\} .$$

Hier wird nun auch formal deutlich, dass unter der Zusatzinformation „ A ist eingetreten“ das Gesamtexperiment „ n -facher Münzwurf“ äquivalent zu einem einzelnen Münzwurf $\tilde{\Omega}$ wird.

1.3.2 Folgerungen, „totale W'keit“ und der „Satz von Bayes“

Es gibt einige einfache Rechenregeln für bedingte W'keiten:

18.4.07↓

Satz 1.29: (Eine Rechenregel für bedingte W'keiten)

Seien E_1, E_2, \dots Ereignisse in einem Modell (Ω, \mathcal{E}, P) . Dann gilt

$$P(E_1 \cap \dots \cap E_n) = P(E_1) P(E_2|E_1) P(E_3|E_2 \cap E_1) \cdots P(E_n|E_{n-1} \cap \dots \cap E_1)$$

Beweis: Einsetzen der Definition 1.25 liefert

$$P(E_1) P(E_2|E_1) P(E_3|E_2 \cap E_1) \cdots P(E_n|E_{n-1} \cap \dots \cap E_1) = P(E_1) \frac{P(E_2 \cap E_1)}{P(E_1)} \frac{P(E_3 \cap E_2 \cap E_1)}{P(E_2 \cap E_1)} \cdots \frac{P(E_n \cap E_{n-1} \cap \dots \cap E_1)}{P(E_{n-1} \cap \dots \cap E_1)}.$$

Fast alles kürzt sich heraus, es verbleibt $P(E_n \cap \dots \cap E_1)$.

Q.E.D.

Beispiel 1.30: Mit welcher W'keit hat jeder der 3 Spieler beim Skat genau ein As?

Lösung: Beim Skat werden von 32 Karten jeweils 10 an die drei Spieler und 2 an den „Skat“ verteilt. Sei E_i das Ereignis „der i -te Spieler hat genau ein As“. Gefragt ist nach $E_1 \cap E_2 \cap E_3$. Die W'keit, dass der 1-te Spieler genau ein As hat, ist

$$P(E_1) = \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}}$$

(wähle 10 Karten aus den 32, davon eine aus den 4 Assen und 9 aus den restlichen 28 Karten). Die W'keit, dass der 2-te Spieler **ebenfalls** genau ein As hat, ist

$$P(E_2|E_1) = P(E_1 \cap E_2|E_1) = \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}}$$

(wähle 10 Karten aus den 22, die nicht an den ersten Spieler verteilt wurden. Davon eine aus den verbleibenden 3 Assen und 9 aus den restlichen 19 Karten). Die W'keit, dass der 3-te Spieler **ebenfalls** genau ein As hat, ist

$$P(E_3|E_2 \cap E_1) = P(E_1 \cap E_2 \cap E_3|E_2 \cap E_1) = \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}}$$

(wähle 10 Karten aus den 12, die nicht an die beiden ersten Spieler verteilt wurden. Davon eine aus den verbleibenden 2 Assen und 9 aus den restlichen 10 Karten). Damit folgt für $E_1 \cap E_2 \cap E_3 =$ „jeder der 3 Spieler hat genau ein As“:

$$\begin{aligned} P(E_1 \cap E_2 \cap E_3) &= P(E_1) P(E_2|E_1) P(E_3|E_2 \cap E_1) \\ &= \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}} \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}} \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}} = \frac{50}{899} \approx 0.0556. \end{aligned}$$

Satz 1.31: („Formel der totalen W'keit“)

Seien $U_1, \dots, U_n \in \mathcal{E}$ disjunkte Ereignisse in einem Modell (Ω, \mathcal{E}, P) , also $U_i \cap U_j = \emptyset$ für $i \neq j$. Dann gilt für jedes Ereignis $E \subset U_1 \cup \dots \cup U_n$:

$$P(E) = \sum_{i=1}^n P(E | U_i) P(U_i).$$

Beweis:

$$\begin{aligned} P(E) &= P\left(E \cap \left(\bigcup_{i=1}^n U_i\right)\right) = P\left(\bigcup_{i=1}^n (E \cap U_i)\right) \\ &\stackrel{\text{(disjunkt)}}{=} \sum_{i=1}^n P(E \cap U_i) = \sum_{i=1}^n P(E | U_i) P(U_i). \end{aligned}$$

Hierbei wird benutzt, dass sich nach Folgerung 1.6.a) die W'keiten der disjunkten Mengen $E \cap U_i$ addieren.

Q.E.D.

Interpretation 1.32:

Für $E \subset U_1 \cup \dots \cup U_n$ können die U_i als **vollständiges System disjunkter Ursachen** interpretiert werden: das Ereignis E kann aus unterschiedlichen Gründen U_i eintreten. Ist bekannt, mit welcher W'keit das Ereignis E bei gegebenen Ursachen jeweils eintritt, so liefert die Formel der totalen W'keit die W'keit für E , wenn die W'keiten $P(U_i)$ bekannt sind, mit denen die Ursachen eintreten.

Eine typische Anwendung dieses Satzes ist die Situation, wo mehrere Experimente hintereinandergeschaltet werden, wobei die möglichen Ausgänge U_i des ersten Experiments einen Einfluss auf das Ergebnis des zweiten Experiments haben.

Beispiel 1.33: Ich habe einen Topf mit 7 fairen Münzen und 3 Münzen mit 2 Köpfen.

Erstes Experiment: Ich entnehme dem Topf eine Münze.

Zweites Experiment: Ich nehme die gezogene Münze und werfe sie 3 Mal.

Frage: Mit welcher W'keit werfe ich im Gesamtexperiment 3 mal „Kopf“?

Lösung: Ich zerlege den Stichprobenraum Ω des zweiten Experiments in die beiden Teile U_1 und U_2 , wobei U_1 die Situation beschreibt, dass ich im ersten Experiment eine faire Münze gezogen habe, während U_2 einer „Doppelkopf“-Münze entspricht. Offensichtlich gilt $P(U_1) = 7/10$, $P(U_2) = 3/10$.

Die W'keit, bei 3 Würfeln 3 Köpfe zu werfen, ist für eine faire Münze $1/8$ ($E = \{(K, K, K)\}$ bei insgesamt 8 Möglichkeiten $(K, K, K), (K, K, Z), \dots, (Z, Z, Z)$), also

$$P(\text{„3 Köpfe in 3 Würfeln“} | U_1) = \frac{1}{8}.$$

Für eine „Doppelkopf“-Münze ist die W'keit für 3 Köpfe bei 3 Würfeln natürlich 1:

$$P(\text{„3 Köpfe in 3 Würfeln“} \mid U_2) = 1.$$

Die Formel der totalen W'keit liefert für das Gesamtexperiment

$$\begin{aligned} & P(\text{„3 Köpfe in 3 Würfeln“}) \\ &= P(\text{„3 Köpfe in 3 Würfeln“} \mid U_1) P(U_1) + P(\text{„3 Köpfe in 3 Würfeln“} \mid U_2) P(U_2) \\ &= \frac{1}{8} \cdot \frac{7}{10} + 1 \cdot \frac{3}{10} = \frac{31}{80} = 0.3875 . \end{aligned}$$

Der folgende Satz beantwortet „die Schuldfrage“. Ein Ereignis E kann aus unterschiedlichen Ursachen U_1, U_2, \dots heraus eintreten. Das Ereignis ist eingetreten. Mit welcher W'keit ist die Ursache U_i verantwortlich gewesen?

Satz 1.34: (Thomas Bayes, 1702–1761)

Seien $U_1, \dots, U_n \in \mathcal{E}$ disjunkte Ereignisse in einem Modell (Ω, \mathcal{E}, P) , also $U_i \cap U_j = \emptyset$ für $i \neq j$. Dann gilt für jedes Ereignis $E \subset U_1 \cup \dots \cup U_n$:

$$P(U_i \mid E) = \frac{P(E \mid U_i) P(U_i)}{P(E)} \stackrel{(1.31)}{=} \frac{P(E \mid U_i) P(U_i)}{\sum_{j=1}^n P(E \mid U_j) P(U_j)} .$$

Beweis: Es gilt

$$P(U_i \mid E) = \frac{P(U_i \cap E)}{P(E)} = \frac{P(U_i) P(U_i \cap E) / P(U_i)}{P(E)} = \frac{P(U_i) P(E \mid U_i)}{P(E)} .$$

$P(E)$ ist über die Formel der totalen W'keit 1.31 durch $P(E \mid U_i)$ und $P(U_i)$ ausdrückbar.

Q.E.D.

Interpretation 1.35:

Bayes beantwortet folgende Frage: das Ereignis E ist eingetreten. Mit welcher W'keit war die Ursache U_i für das Eintreten von E verantwortlich?

Bemerkung 1.36: Es gilt:

$$\begin{aligned} \sum_{i=1}^n P(U_i \mid E) &= \sum_{i=1}^n \frac{P(U_i) P(E \mid U_i)}{P(E)} = \frac{1}{P(E)} \sum_{i=1}^n P(U_i) P(E \mid U_i) \\ &\stackrel{(1.31)}{=} \frac{P(E)}{P(E)} = 1 . \end{aligned}$$

Intuitiv: eine der Ursachen war mit Sicherheit für das Eintreten von E verantwortlich. Obige Formel ist oft eine nette Kontrollmöglichkeit, ob man in konkreten Aufgaben richtig gerechnet hat.

Beispiel 1.37: Ein Test auf eine Krankheit ist mit der W'keit 0.9 positiv, wenn ein Kranker getestet wird. Bei Gesunden fällt der Test mit der W'keit 0.1 positiv aus. Es ist bekannt, dass jeder 1450-te der Bevölkerung diese Krankheit hat. Eine Person wird getestet.

- Mit welcher W'keit ist der Test positiv?
- Der Test ist positiv. Mit welcher W'keit ist die Person krank? Mit welcher W'keit ist sie gesund?
- Wird die Krankenkasse für diesen Test zahlen?

Lösung: Sei $+$ das Ereignis „der Test ist positiv“. Das Ursachensystem für $+$ ist k = „die Person ist krank“ und g = „die Person ist gesund“. Es gilt $P(k) = 1/1450$, $P(g) = 1449/1450$. Weiterhin ist bekannt $P(+|k) = 0.9$ und $P(+|g) = 0.1$.

- a) Nach Satz 1.31 gilt

$$P(+) = P(+|k) P(k) + P(+|g) P(g) = 0.9 \cdot \frac{1}{1450} + 0.1 \cdot \frac{1449}{1450} \approx 0.10055\dots$$

- b) Nach Satz 1.34 gilt

$$P(k|+) = \frac{P(k) P(+|k)}{P(+)} = \frac{1}{1450} \frac{0.9}{0.10055\dots} \approx 0.0062\dots$$

$$P(g|+) = \frac{P(g) P(+|g)}{P(+)} = \frac{1449}{1450} \frac{0.1}{0.10055\dots} \approx \frac{0.9938\dots}{1.0000\dots}$$

Das wirkt überraschend. Die Tatsache, dass die Wahr'scheinlichkeit einer positiv getesteten Person, gesund zu sein, ungleich viel höher ist, als krank zu sein, liegt daran, dass ohne Test die W'keit, gesund zu sein, fast 1 ist.

- c) Der Test erhöht die W'keit $P(k) = 1/1450 \approx 0.00069$ nur um den recht kleinen Faktor $P(+|k)/P(+)$ = $0.9/0.10055\dots \approx 9.0$ auf die W'keit $P(k|+) \approx 0.0062$.

Die W'keit $P(g) = 1449/1450 \approx 0.99931\dots$ wird bei einem positiven Tests nur um den dicht bei 1 liegenden Faktor $P(+|g)/P(+)$ = $0.1/0.10055 \approx 0.9945\dots$ auf die W'keit $P(g|+) \approx 0.9938$ erniedrigt.

Der Test ist damit völlig wertlos.

Beispiel 1.38: Folgende Tabelle ist eine fiktive Aufschlüsselung der 1980er Bundestagswahl nach Konfessionen:

Bevölkerungsanteil		CDU	SPD	FDP	Grüne	Sonstige
46%	katholisch	54%	41%	3%	1%	1%
44%	evangelisch	45%	45%	6%	3%	1%
10%	sonstig	30%	38%	15%	10%	7%

Welcher Anteil der CDU-Wähler ist katholisch?

Lösung: Die gegebenen Informationen sind

$$P(\text{kath}) = 0.46, \quad P(\text{ev}) = 0.44, \quad P(\text{sonst}) = 0.10$$

sowie

$$P(\text{CDU} | \text{kath}) = 0.54, \quad P(\text{CDU} | \text{ev}) = 0.45, \quad \text{etc.}$$

Nach Satz 1.34 gilt

$$\begin{aligned} P(\text{kath} | \text{CDU}) &= \frac{P(\text{kath}) P(\text{CDU} | \text{kath})}{P(\text{CDU})} = \\ &= \frac{P(\text{kath}) P(\text{CDU} | \text{kath})}{P(\text{kath}) P(\text{CDU} | \text{kath}) + P(\text{ev}) P(\text{CDU} | \text{ev}) + P(\text{sonst}) P(\text{CDU} | \text{sonst})} \\ &= \frac{0.46 \cdot 0.54}{0.46 \cdot 0.54 + 0.44 \cdot 0.45 + 0.1 \cdot 0.3} = \frac{0.248\dots}{0.476\dots} \approx 52\% . \end{aligned}$$

Der Nenner $P(\text{CDU}) = 0.476\dots$ entspricht dem damaligen Wahlergebnis der CDU von 47.6%.

Merke (Zusammenfassung) 1.39:

Die **Formel der totalen W'keit** und der **Satz von Bayes** führen Ereignisse E auf unterschiedliche (disjunkte) Ursachen U_i zurück. Kennt man $P(E | U_i)$ sowie die W'keiten $P(U_i)$ der einzelnen Ursachen, so kann mit der Formel der totalen W'keit $P(E)$ berechnet werden. Bayes beantwortet die „Schuldfrage“: welche Ursache war mit welcher W'keit für das Eintreten eines beobachteten Ereignisses verantwortlich?

Typischer Einsatz: Modellierung mehrerer nicht unabhängiger Einzelexperimente.

1.4 Irrfahrten auf Graphen

23.4.07↓

Man kann viele W'keitsexperimente, die aus der Hintereinanderschaltung von Einzelexperimenten bestehen, als Irrfahrt auf einem (gerichteten gewichteten) Graphen interpretieren. Dies liefert eine nützliche Visualisierungsmöglichkeit, mit der W'keiten oft leicht graphisch zusammengesetzt werden können.

1.4.1 Wahrscheinlichkeitsbäume

Zunächst zur formalen Definition der speziellen Graphen, um die es hier geht:

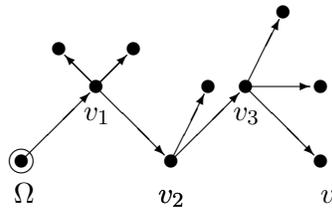
Definition 1.40: (Wurzelbäume)

Ein **numerierter Wurzelbaum** (Ω, V, K) mit **Wurzel** Ω besteht aus einer **Knotenmenge** $V = \{\Omega, E_1, E_2, \dots\}$ und einer **Kantenmenge** $K \subset V \times V = \{(v, s); v, s \in V\}$, so dass zu jedem $v \in V \setminus \{\Omega\}$ genau eine **Kantenfolge** $k_1, k_2, \dots, k_k \in K$ der Form

$$k_1 = (\Omega, v_1), \dots, k_i = (v_{i-1}, v_i), \dots, k_k = (v_{k-1}, v)$$

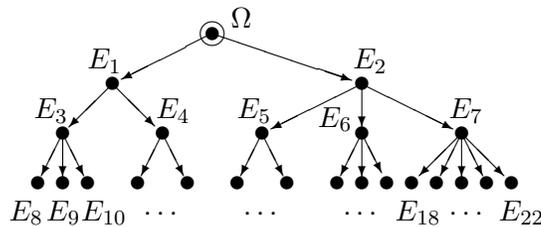
existiert (ein **eindeutiger Pfad der Länge k** von der Wurzel Ω nach v).

Beispiel:



Bemerkung 1.41: Gewurzelte Bäume sind gerichtete (Kanten sind geordnete Paare) zusammenhängende (... es existiert ein Pfad ...) Graphen ohne geschlossene Wege (... genau ein Pfad ...). In einer Kante $(v, s) \in K$ heißt v „Vater“ und s „Sohn“ (der Abstand des Vaters zur Wurzel ist um 1 kleiner als der Abstand des Sohns zur Wurzel).

Es bietet sich an, einen Wurzelbaum graphisch so darzustellen, dass alle Knoten mit dem selben Abstand zur Wurzel Ω jeweils eine „Schicht“ bilden. Die Wurzeln von Bäumen sind bei den Informatikern typischerweise oben ;-(:



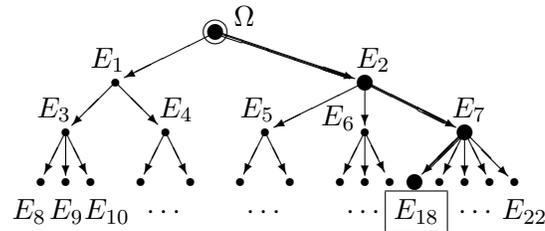
Eine wichtige Tatsache ist:

Per Definition der Wurzelbäume gehört zu jedem Knoten genau ein Pfad von der Wurzel zum Knoten. Jeder Knoten kann also auch als Pfad auf dem Graphen interpretiert werden!

Beispiel: Der Knoten E_{18} im Baum oben kann als der Pfad

$$((\Omega, E_2), (E_2, E_7), (E_7, E_{18}))$$

mit den dicker eingezeichneten Kanten interpretiert werden:



Nun die Interpretation eines aus Teilerperimenten zusammengesetzten Gesamtexperiments durch einen solchen Baum:

Modellierungsregeln 1.42:

- a) Ω ist der Stichprobenraum des Gesamtexperiments.
- b) Alle weiteren Knoten sind Ereignisse des Gesamtexperiments, also Teilmengen von Ω .
- c) Eine Kante von einem Knoten v („Vater“) zum Knoten s („Sohn“) wird als Durchführung eines Einzelerperiments interpretiert, das auf dem Stichprobenraum $v \subset \Omega$ ausgeführt wird und das Ergebnis (Ereignis) $s \subset v \subset \Omega$ liefert:

Jeder Sohn ist Teilmenge seines Vaters.

- d) Die Kanten werden gewichtet mit den **Übergangsw'keiten**

$$P(s|v) = \frac{P(s \cap v)}{P(v)} \stackrel{(s \subset v)}{=} \frac{P(s)}{P(v)}.$$

wobei P das W'keitsmaß auf dem Stichprobenraum Ω des Gesamtexperiments ist. Nach Interpretation 1.26 dürfen wir diese bedingten W'keiten (in Ω) in der Tat als normale W'keiten im Stichprobenraum $v \subset \Omega$ ansehen, d.h., zur Berechnung der Übergangsw'keiten $P(s|v)$ können wir ein (in der Regel einfacher zu beschreibendes Einzelerperiment) v benutzen.

Die Durchführungen des Gesamtexperiments (Hintereinanderschaltung von n Teilerperimenten) sind die von der Wurzel des W'keitsbaums ausgehenden Pfade. Der Pfad $(\Omega, v_1, v_2, \dots, v_{n-1}, v_n)$ über die Kanten (Ω, v_1) , (v_1, v_2) , \dots , (v_{n-1}, v_n) hat die Interpretation:

„Das auf dem Stichprobenraum Ω durchgeführte erste Einzelerperiment liefert das Ergebnis $v_1 \subset \Omega$. Das dann auf dem Stichprobenraum v_1 durchgeführte zweite Einzelerperiment liefert das Ergebnis $v_2 \subset v_1 \subset \Omega$. Usw.“

In der Interpretation der Knoten des W'keitsbaums als Ereignisse im Gesamtexperiment gelten für jeden Pfad $(\Omega, v_1, v_2, \dots, v_n)$ von der Wurzel Ω zum Knoten v_n die Inklusionen

$$v_n \subset v_{n-1} \subset \dots \subset v_1 \subset \Omega,$$

also

$$v_n = v_n \cap v_{n-1} = v_n \cap v_{n-1} \cap v_{n-2} = \dots = v_n \cap v_{n-1} \cap \dots \cap v_1.$$

Damit berechnet sich nach Satz 1.29 die W'keit des Ereignisses $v_n \subset \Omega$ durch

$$\begin{aligned} P(v_n) &= P(v_n \cap v_{n-1} \cap \dots \cap v_1) = \\ &= \underbrace{P(v_1)}_{=P(v_1|\Omega)} \cdot P(v_2|v_1) \cdot P(v_3|\underbrace{v_2 \cap v_1}_{=v_2}) \cdot \dots \cdot P(v_n|\underbrace{v_{n-1} \cap \dots \cap v_1}_{=v_{n-1}}), \end{aligned}$$

also:

$$P(v_n) = \underbrace{P(v_1|\Omega)}_{=P(v_1)} P(v_2|v_1) P(v_3|v_2) \dots P(v_n|v_{n-1}).$$

Die Berechnung der W'keiten $P(v_n)$ geschieht also rein über die Übergangsw'keiten aller Kanten längs des Pfads von Ω nach v_n . Dies ergibt die erste „Pfadregel“:

Satz 1.43: (Pfadregel 1)

Die W'keit eines Knoten in einem W'keitsbaum, der den Modellierungsregeln 1.42 genügt, ist das Produkt aller Übergangsw'keiten der Kanten des Pfades, der von der Wurzel zu dem Knoten führt.

Die Erleichterung in der Modellierung von W'keitsexperimenten ist, dass die W'keit eines Ereignisses im Gesamtexperiment vollständig durch die W'keiten von Ereignissen in den (einfacher zu beschreibenden) Telexperimenten berechenbar ist.

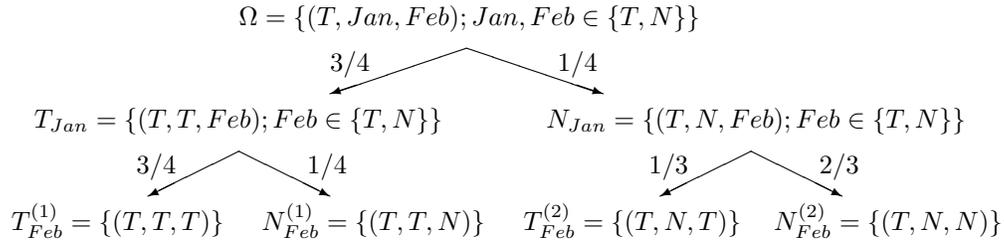
Beispiel 1.44: *Gabriel und Neumann* betrachteten im *Quarterly Journal of the Royal Meteorological Society*, vol. **83**, pp 375–380 (1957), folgendes einfache stochastische Modell für das Wetter in Israel in den Wintermonaten Dezember, Januar, Februar:

↓23.4.07

- a) Ist es in einem Monat überwiegend trocken (T), so ist der Folgemonat mit der W'keit 3/4 überwiegend trocken und mit der W'keit 1/4 überwiegend nass (N).
- b) Ist es in einem Monat nass, so ist der Folgemonat mit der W'keit 1/3 überwiegend trocken und mit der W'keit 2/3 überwiegend nass.

Wir haben einen trockenen Dezember. Mit welcher W'keit ist das Wetter im Februar trocken?

Der Stichprobenraum sei $\Omega = \{(Dez, Jan, Feb); Dez = T; Jan, Feb \in \{T, N\}\}$. Er beschreibt einen trockenen Dezember und die vier unterschiedlichen Wetterkombinationen für Januar und Februar. Wir betrachten den Baum



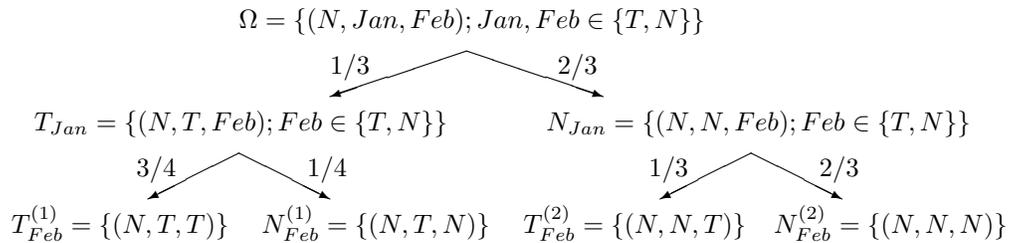
Die eingetragenen Übergangswahrscheinlichkeiten sind dabei die vorgegebenen Daten. Aus der Pfadregel 1 ergeben sich sofort die Wahrscheinlichkeiten für die Februar-Knoten, die die Pfade kodieren:

$$\begin{aligned}
 P(\{(T, T, T)\}) &= \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}, & P(\{(T, T, N)\}) &= \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}, \\
 P(\{(T, N, T)\}) &= \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}, & P(\{(T, N, N)\}) &= \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}.
 \end{aligned}$$

Das Ereignis „der Februar ist trocken“ besteht aus den Ereignissen $\{(T, T, T)\}$ und $\{(T, N, T)\}$. Da diese disjunkt sind, sind die Wahrscheinlichkeiten für die Vereinigungsmenge zu addieren:

$$\begin{aligned}
 P(\text{„der Februar ist trocken“}) &= P(\{(T, T, T)\} \cup \{(T, N, T)\}) \\
 &= P(\{(T, T, T)\}) + P(\{(T, N, T)\}) = \frac{9}{16} + \frac{1}{12} = \frac{31}{48}.
 \end{aligned}$$

Bei einem nassen Dezember würde sich ergeben:



Hiermit folgt

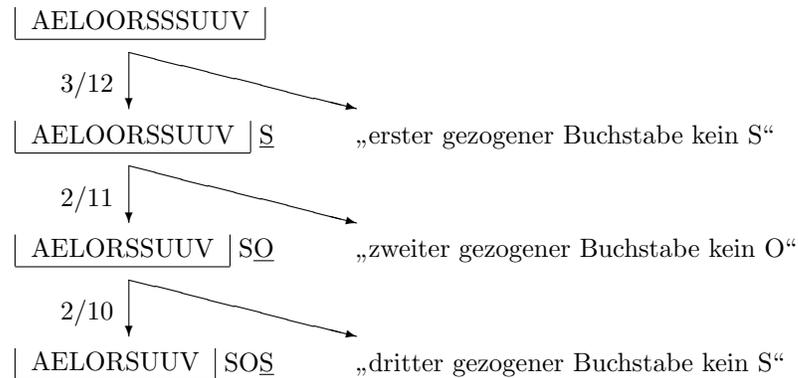
$$\begin{aligned}
 P(\{(N, T, T)\}) &= \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4}, & P(\{(N, T, N)\}) &= \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}, \\
 P(\{(N, N, T)\}) &= \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}, & P(\{(N, N, N)\}) &= \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}
 \end{aligned}$$

und somit:

$$\begin{aligned} P(\text{„der Februar ist trocken“}) &= P(\{(N, T, T) \cup \{N, N, T\}\}) \\ &= P(\{(N, T, T)\}) + P(\{(N, N, T)\}) = \frac{1}{4} + \frac{2}{9} = \frac{17}{36}. \end{aligned}$$

Beispiel 1.45: Aus der Urne $\boxed{\text{SAVE OUR SOULS}}$ (= $\boxed{\text{AELOORSSSUUV}}$) werden ohne Zurücklegen drei Buchstaben gezogen. Mit welcher W'keit bilden diese Buchstaben in der gezogenen Reihenfolge das Wort SOS?

Wir betrachten als Einzelexperimente die drei einzelnen Züge. Der entsprechende W'keitsbaum ist:



Die Übergangsw'keiten ergeben sich dabei leicht folgendermaßen aus den Einzelexperimenten:

Zug 1: Die Wahrscheinlichkeit ist $3/12$, aus einer Urne mit 12 Buchstaben, von denen 3 ein S sind, ein S zu ziehen.

Zug 2: Die Wahrscheinlichkeit ist $2/11$, aus einer Urne mit 11 Buchstaben, von denen 2 ein O sind, ein O zu ziehen.

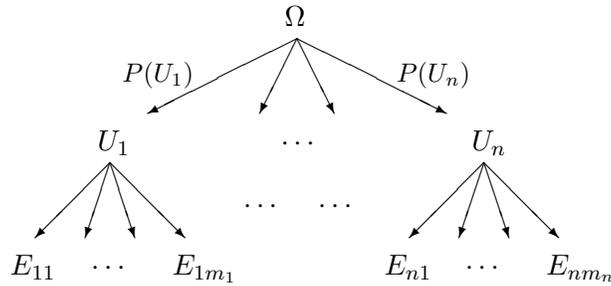
Zug 3: Die Wahrscheinlichkeit ist $2/10$, aus einer Urne mit 10 Buchstaben, von denen 2 ein S sind, ein S zu ziehen.

Als Endergebnis erhält man über die Pfadregel 1:

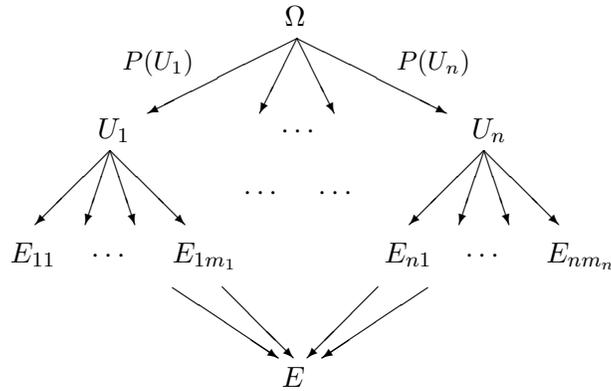
$$P(\text{„SOS“}) = \frac{3}{12} \cdot \frac{2}{11} \cdot \frac{2}{10} = \frac{1}{110}.$$

1.4.2 Allgemeinere Wahrscheinlichkeitsgraphen

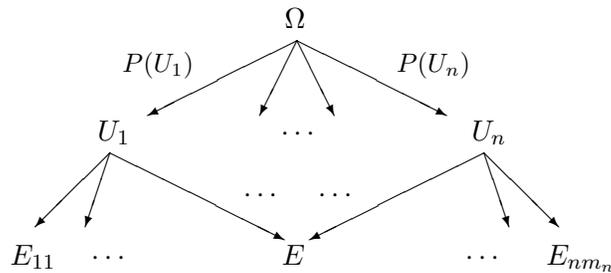
Wir verallgemeinern nun das graphische Konzept des letzten Abschnitts, indem wir allgemeinere Graphen zulassen, die nicht unbedingt Bäume zu sein brauchen. Als Motivation betrachten wir zunächst ein zweistufiges Experiment:



Das interessierende Ereignis E , dessen W'keit berechnet werden soll, sei die Vereinigung von mehreren der Ereignisse E_{ij} der unteren Schicht des Baums:



Wir wollen das Ereignis E direkt von den U_i anfahren, ohne über die „Teilergebnisse“ E_{ij} laufen zu müssen:



Wir können nun nicht mehr verlangen, dass ein Ereignis auf nur einem Pfad von der Wurzel aus erreicht werden kann:

Definition 1.46: (W'keitsgraph)

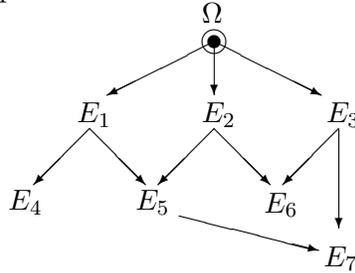
Ein **W'keitsgraph** (Ω, V, K) mit **Wurzel** Ω besteht aus einer **Knotenmenge** $V = \{\Omega, E_1, E_2, \dots\}$ und einer **Kantenmenge** $K \subset V \times V = \{(v, s); v, s \in V\}$, so dass zu jedem $s \in V \setminus \{\Omega\}$ mindestens eine und maximal endlich viele Kantenfolgen $k_1, k_2, \dots, k_k \in K$ der Form

$$k_1 = (\Omega, v_1), \dots, k_i = (v_{i-1}, v_i), \dots, k_k = (v_{k-1}, s)$$

existiert (Pfade von der Wurzel Ω nach s).

Die Bedingung „maximal endlich viele“ Pfade von Ω nach s impliziert, dass es keine Zyklen (geschlossene Wege) von einem Punkt s zurück zu sich selbst geben darf. Wäre ein solcher Zyklus vorhanden, könnte man auf beliebig vielen Pfaden von Ω nach s laufen, indem man unterwegs den Zyklus beliebig oft durchläuft.

Beispiel eines W'keitsgraphen:



Hier die **Modellierungsanleitung** zur Darstellung eines aus der Hintereinanderausführung von Telexperimenten zusammengesetzten Gesamtexperiments:

Modellierungsregeln 1.47:

- a) Ω ist der Stichprobenraum des Gesamtexperiments.
- b) Alle weiteren Knoten sind Ereignisse des Gesamtexperiments, also Teilmengen von Ω .
- c) Eine Kante von einem Knoten v („Vater“) zum Knoten s („Sohn“) wird als Durchführung eines Einzalexperiments interpretiert, das auf dem Stichprobenraum $v \subset \Omega$ ausgeführt wird und das Ergebnis (Ereignis) $s \subset \Omega$ liefert:

Ein Sohn ist dabei Teilmenge der Vereinigung aller seiner Väter.

- d) Die Kanten werden gewichtet mit den **Übergangsw'keiten**

$$P(s|v) = \frac{P(s \cap v)}{P(v)},$$

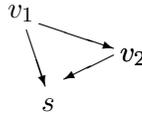
wobei P das W'keitsmaß auf dem Stichprobenraum Ω des Gesamtexperiments ist.

- e) Wir setzen weiterhin voraus:

Väter mit einem gemeinsamen Sohn sind disjunkt.

Hierbei ist 1.47.c) die Verallgemeinerung der Modellierungsregel 1.42.c) („Jeder Sohn ist Teilmenge seines Vaters“) für Bäume. Beim Punkt d) ist zu beachten, dass wir (bei Söhnen mit mehreren Vätern) nicht mehr die Inklusion $s \subset v$ haben, also nicht $P(s \cap v)$ zu $P(s)$ vereinfachen können. Regel 1.47.e) ist neu hinzugekommen (für Bäume, wo jeder Sohn nur genau einen Vater besitzt, ist dieser Punkt inhaltslos).

Die Regeln c) und e) implizieren gewisse Einschränkungen an die Graphen bzw. an die Übergangsw'keiten. Beispielsweise kann es keine „Dreiecksbeziehungen“ im Graphen geben (ein Sohn eines Vaters kann nicht gleichzeitig ein Enkel dieses Vaters sein):



Im obigen Beispiel müssen v_1 und v_2 als gemeinsame Väter von s disjunkte Teilmengen von Ω sein. Die Übergangsw'keit längs der Kante von v_1 nach v_2 wäre damit

$$P(v_2|v_1) = \frac{P(v_1 \cap v_2)}{P(v_1)} = \frac{P(\emptyset)}{P(v_1)} = 0.$$

Die Kante (v_1, v_2) ist damit überflüssig und kann entfallen (man kann sie natürlich auch formal mit der Übergangsw'keit 0 stehen lassen).

Definition 1.48: (Pfadw'keit)

Für einen Pfad $(\Omega, v_1, v_2, \dots, v_n)$ von der Wurzel zu einem Knoten definieren wir die **Pfadw'keit**:

$$P_{(\Omega, v_1, \dots, v_n)} = P(v_1) \cdot P(v_2|v_1) \cdot P(v_3|v_2) \cdot \dots \cdot P(v_n|v_{n-1})$$

als das Produkt der Übergangsw'keiten längs der Kanten des Pfades.

Bemerkung 1.49: Liegen auf einem Pfad $(\Omega, v_1, \dots, v_n)$ nur Knoten, die nur jeweils einen Vater haben, so gelten wie bei Bäumen die Inklusionen

$$v_n \subset v_{n-1} \subset \dots \subset v_1 \subset \Omega$$

und damit

$$P_{(\Omega, v_1, \dots, v_n)} = P(v_1 \cap v_2 \cap \dots \cap v_n) = P(v_n).$$

Hat aber mindestens einer der Knoten im Pfad mehrere Väter, so gilt diese Interpretation der Pfadw'keit nicht mehr! Es gilt weder $P_{(\Omega, v_1, \dots, v_n)} = P(v_n)$ noch $P_{(\Omega, v_1, \dots, v_n)} = P(v_1 \cap v_2 \cap \dots \cap v_n)$.

Satz 1.50: (Pfadregel 2)

In einem W'keitsgraphen, der den Modellierungsregeln 1.47 genügt, gilt für jeden Knoten v :

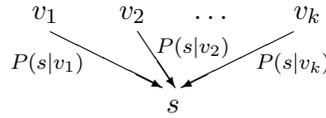
$$P(v) = \sum_{\text{Pfad} \in \text{Pfade}(v)} P_{\text{Pfad}},$$

wobei $\text{Pfade}(v)$ die Menge aller Pfade von Ω nach v ist und P_{Pfad} die Pfadw'keit nach Definition 1.48.

Beweis: Induktion nach der Länge der Pfade.

Induktionsstart: Für einen direkt an der Wurzel Ω hängenden Knoten v kann es neben Ω keinen weiteren Vater geben, denn dieser müsste wegen 1.47.e) disjunkt von Ω sein. Die Kante (Ω, v) (mit der Übergangsw'keit $(P(v))$) ist damit der einzige Pfad von Ω nach v und liefert die korrekte W'keit $P(v)$.

Induktionsschritt: Die Behauptung gelte für alle Knoten, die von Ω aus mit einem Pfade der Länge $\leq n$ erreichbar sind. Sei nun s ein Knoten, der über einen Pfad der maximalen Länge $n + 1$ von Ω aus erreichbar ist. Seien die Knoten v_1, \dots, v_k die Väter von s :



Gemäß der Regel 1.47.e) sind die Ereignisse v_1, \dots, v_k disjunkt und überdecken gemäß Regel 1.47.c) das Ereignis s . Damit ist der Satz von der totalen W'keit 1.31 anwendbar, es gilt:

$$P(s) = \sum_{i=1}^k P(s | v_i) P(v_i).$$

Da die Väter v_i alle über Pfade der maximalen Länge n mit der Wurzel verbunden sind, gilt nach Induktionsvoraussetzung

$$P(v_i) = \sum_{\text{Pfad} \in \text{Pfade}(v_i)} P_{\text{Pfad}},$$

also

$$P(s) = \sum_{i=1}^k P(s | v_i) \cdot \sum_{\text{Pfad} \in \text{Pfade}(v_i)} P_{\text{Pfad}} = \sum_{i=1}^k \sum_{\text{Pfad} \in \text{Pfade}(v_i)} P(s | v_i) \cdot P_{\text{Pfad}}.$$

Für jeden Pfad in $\text{Pfade}(v_i)$ ist das Produkt $P(s | v_i) \cdot P_{\text{Pfad}}$ die Pfadw'keit des erweiterterten Pfades, der von Ω über v_i nach s führt. Da jeder der Pfade

von Ω nach s über einen der Väter v_i führt, erstreckt sich die Doppelsumme $\sum_i \sum_{\text{Pfad} \in \text{Pfade}(v_i)}$ in der Tat über alle Pfade von Ω nach s , womit die Pfadregel 2 für den Knoten s bewiesen ist.

Q.E.D.

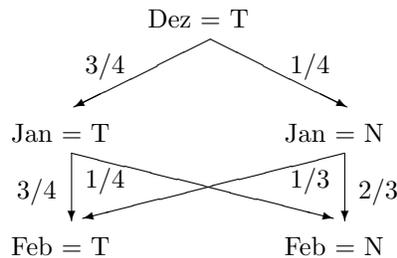
Beispiel 1.51: Wir betrachten erneut das Wetter in Israel nach Beispiel 1.44. Wir benutzen wieder das schon vorher verwendete Modell

$$\Omega = \{(Dez, Jan, Feb); Dez = T; Jan, Feb \in \{T, N\}\} = \text{„Dez} = T\text{“},$$

betrachten aber nun die Ereignisse

$$\begin{aligned} \text{„Jan} = T\text{“} &= \{(T, T, Feb); Feb \in \{T, N\}\}, \\ \text{„Jan} = N\text{“} &= \{(T, N, Feb); Feb \in \{T, N\}\}, \\ \text{„Feb} = T\text{“} &= \{(T, Jan, T); Jan \in \{T, N\}\}, \\ \text{„Feb} = N\text{“} &= \{(T, Jan, N); Jan \in \{T, N\}\} \end{aligned}$$

im folgenden Graphen:



Dieser Graph erfüllt in der Tat die Modellierungsregeln 1.47. Die W'keit für einen trockenen Februar ergibt sich über die Pfadregel 2 unmittelbar zu

$$P(\text{„Feb} = T\text{“}) = \frac{3}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{3} = \frac{31}{48}$$

als Summe der Pfadw'keiten der beiden Pfade (Dez = T, Jan = T, Feb = T) und (Dez = T, Jan = N, Feb = T). Man beachte, dass sich der W'keitsgraph hier gegenüber dem in Beispiel 1.44 verwendeten W'keitsbaum deutlich vereinfacht hat.

1.5 Unabhängigkeit von Ereignissen

30.4.07↓

Nun kommt die Formalisierung eines ganz zentralen Begriffs: Unabhängigkeit.

1.5.1 Definition

Intuitiv sind zwei Ereignisse A, B voneinander unabhängig, wenn das Eintreten von A keinerlei Informationen über B liefert, d.h., wenn $P(B|A) = P(B)$ gilt, also $P(A \cap B)/P(A) = P(B)$, also $P(A \cap B) = P(A)P(B)$:

Definition 1.52: (Unabhängigkeit von Ereignissen)

Zwei Ereignisse $A, B \in \mathcal{E}$ in einem Modell (Ω, \mathcal{E}, P) heißen **unabhängig**, wenn gilt

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

also $P(B|A) = P(B)$ bzw. $P(A|B) = P(A)$.

Die Definition ist offensichtlich symmetrisch, wie es intuitiv auch sein sollte: A ist genau dann unabhängig von B , wenn B unabhängig von A ist. Daher spricht man von der Unabhängigkeit eines Paares A, B .

Beispiel 1.53: Ich werfe 2 mal mit einem fairen Würfel.

- Sind die Ereignisse $A =$ „ich habe im ersten Wurf eine 6“ und $B_1 =$ „ich habe im zweiten Wurf eine 4“ unabhängig?
- Sind die Ereignisse $A =$ „ich habe im ersten Wurf eine 6“ und $B_2 =$ „meine Augensumme ist 10“ unabhängig?

Lösung: Seien

$$\begin{aligned} \Omega &= \text{„alle Wurfkombinationen“} = \{(1, 1), \dots, (6, 6)\} , \\ A &= \text{„im ersten Wurf eine 6“} = \{(6, 1), (6, 2), \dots, (6, 6)\} , \\ B_1 &= \text{„im zweiten Wurf eine 4“} = \{(1, 4), (2, 4), \dots, (6, 4)\} , \\ B_2 &= \text{„Augensumme ist 10“} = \{(4, 6), (5, 5), (6, 4)\} . \end{aligned}$$

- a) Mit $|\Omega| = 36$, $P(A) = |A|/|\Omega| = 6/36 = 1/6$, $P(B_1) = |B_1|/|\Omega| = 6/36 = 1/6$ und

$$A \cap B_1 = \{(6, 4)\} \Rightarrow P(A \cap B_1) = \frac{|A \cap B_1|}{|\Omega|} = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(A) P(B_1)$$

sind A und B_1 unabhängig (was intuitiv klar war).

- b) Mit $P(B_2) = |B_2|/|\Omega| = 3/36 = 1/12$ und

$$A \cap B_2 = \{(6, 4)\} \Rightarrow P(A \cap B_2) = \frac{|A \cap B_2|}{|\Omega|} = \frac{1}{36} \neq P(A) P(B_2) = \frac{1}{72}$$

sind A und B_2 nicht unabhängig. Intuitiv: die Chance, eine hohe Augensumme zu erhalten, hat sich durch die Tatsache erhöht (hier: verdoppelt), dass schon der erste Wurf ein hohes Ergebnis brachte.

Aus technischen Gründen brauchen wir noch den Begriff der Unabhängigkeit von mehr als zwei Ereignissen, der später in die Definition der Unabhängigkeit von Zufallsvariablen eingeht:

Definition 1.54: (Technische Definition: Unabhängigkeit von Ereignisfamilien)

Sei $\mathcal{A} = \{A_i; i \in I\} \subset \mathcal{E}$ eine Familie von Ereignissen (I eine beliebige Indexmenge) in einem Modell (Ω, \mathcal{E}, P) .

- a) Ein Ereignis $B \in \mathcal{E}$ heißt **unabhängig von der Familie \mathcal{A}** , wenn B unabhängig von jedem Schnitt $A_{i_1} \cap \cdots \cap A_{i_k}$ beliebig ausgewählter endlich vieler verschiedener Elemente dieser Familie ist:

$$P(B \cap A_{i_1} \cap \cdots \cap A_{i_k}) = P(B) \cdot P(A_{i_1} \cap \cdots \cap A_{i_k}) .$$

- b) Die Familie \mathcal{A} selbst heißt **unabhängig**, wenn jedes Ereignis $A_i \in \mathcal{A}$ unabhängig von der Restfamilie $\mathcal{A} \setminus \{A_i\}$ ist. Dies ist der Fall, wenn für jede beliebige Auswahl von endlich vielen verschiedenen Elementen $A_{i_1}, \dots, A_{i_k} \in \mathcal{A}$ gilt:

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot \cdots \cdot P(A_{i_k}) .$$

Bemerkung 1.55: (Warnung) Ist B unabhängig von jedem $A_i \in \mathcal{A}$, so ist B damit nicht automatisch unabhängig von der Familie \mathcal{A} . Entsprechend ist eine Familie \mathcal{A} nicht notwendigerweise unabhängig, wenn jedes Paar $A_i, A_j \in \mathcal{A}$ (mit $i \neq j$) unabhängig ist. Gegenbeispiel:

$$\Omega = \{1, 2, 3, 4\} \text{ mit dem Zählmaß, } A_1 = \{1, 3\}, A_2 = \{2, 3\}, A_3 = \{3, 4\} .$$

Es gilt die paarweise Unabhängigkeit der Ereignisse in $\mathcal{A} = \{A_1, A_2, A_3\}$:

$$P(A_1 \cap A_2) = \frac{1}{4} = P(A_1) P(A_2) , P(A_1 \cap A_3) = \frac{1}{4} = P(A_1) P(A_3) ,$$

$$P(A_2 \cap A_3) = \frac{1}{4} = P(A_2) P(A_3) ,$$

aber

$$P(A_1 \cap A_2 \cap A_3) = P(\{3\}) = \frac{1}{4} \neq P(A_1) P(A_2) P(A_3) = \frac{1}{8} .$$

Merke (Zusammenfassung) 1.56:

Unabhängigkeit zweier Ereignisse A, B wird modelliert durch

$$P(\text{„sowohl } A \text{ als auch } B \text{ treten ein“}) \equiv P(A \cap B) = P(A) \cdot P(B) .$$

1.5.2 Modellierung unabhängiger Experimente: Produktmodelle

Ein typisches Beispiel für unabhängige Ereignisse ist die „unabhängige Wiederholung eines Experiments“ (mehrere Male würfeln, mehrere Roulette-Versuche etc). Wie baut man sich systematisch ein Modell für *unabhängige* Durchführungen einzelner Zufallsexperimente?

Wir haben schon mehrfach in den Beispielen unabhängige Wiederholungen von Würfelexperimenten, Münzwürfen etc. durch Modelle der Form

$$\Omega = \{(\text{erstes Ergebnis}, \text{zweites Ergebnis}, \dots)\}$$

behandelt. Dies war mehr intuitiv. Die Tatsache, dass Unabhängigkeit in der Tat durch geordnete Tupel von Einzelexperimenten beschrieben werden kann, soll nun formalisiert werden. Zur Erinnerung: das kartesische Produkt zweier Mengen M_1, M_2 ist die Menge aller geordneten Tupel

$$M_1 \times M_2 = \{(m_1, m_2); m_1 \in M_1, m_2 \in M_2\} .$$

Im Sinne von Definition 1.14 wird ein W'keitsmaß für ein diskretes Produktmodell festgelegt:

Definition 1.57: (Produktmodelle²)

Seien $(\Omega_1, \mathcal{E}_1, P_1), \dots, (\Omega_n, \mathcal{E}_n, P_n)$ diskrete Modelle („Einzelexperimente“). Das **Produktmodell** ist das diskrete Modell (Ω, \mathcal{E}, P) mit

a) $\Omega = \Omega_1 \times \dots \times \Omega_n = \{(\omega_1, \dots, \omega_n); \omega_1 \in \Omega_1; \dots; \omega_n \in \Omega_n\} .$

b) $\mathcal{E} = \mathcal{P}(\Omega) =$ Potenzmenge von Ω .

c) Für die Elementarereignisse $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ wird definiert:

$$P(\underbrace{\{(\omega_1, \dots, \omega_n)\}}_{\omega}) = P_1(\{\omega_1\}) \cdot \dots \cdot P_n(\{\omega_n\})$$

(das „**Produktw'keitsmaß**“). Hiermit wird für beliebige Ereignisse $E \subset \Omega$ definiert:

$$P(E) = \sum_{\omega \in E} P(\{\omega\}) .$$

²Da die Einzelexperimente hier als diskret vorausgesetzt sind (also: die Ω_i sind abzählbar und $\mathcal{E}_i = \mathcal{P}(\Omega_i)$), ist der Produktraum wieder diskret. Man kann auch Produkte allgemeiner (nicht notwendigerweise diskreter) Einzelexperimente $(\Omega_i, \mathcal{E}_i, P_i)$ definieren, was aber etwas technischer ist. Dann setzt man z.B.

$$\mathcal{E} = \text{die von } \{E_1 \times \dots \times E_n; E_1 \in \mathcal{E}_1, \dots, E_n \in \mathcal{E}_n\} \text{ erzeugte Sigma-Algebra.}$$

Hierbei wird von einem Mengensystem eine Sigma-Algebra dadurch erzeugt, dass man zum Mengensystem alle Komplemente und abzählbaren Vereinigungen hinzunimmt, dann wiederum alle Komplemente und abzählbaren Vereinigungen hinzunimmt, usw.

2.5.07↓

Interpretation und Folgerungen 1.58:

Die Elementarereignisse $(\omega_1, \dots, \omega_n)$ sind folgendermaßen zu interpretieren: Das Experiment $(\Omega_1, \mathcal{E}_1, P_1)$ wird durchgeführt und liefert den Wert ω_1 . Unabhängig davon wird das Experiment $(\Omega_2, \mathcal{E}_2, P_2)$ durchgeführt und liefert den Wert ω_2 . Usw.

Das Ereignis $E =$ „im i -ten Schritt tritt das Ereignis $E_i \subset \Omega_i$ ein“ (und es ist egal, was in den anderen Schritten passiert) ist dann als das Ereignis

$$E = \Omega_1 \times \dots \times \Omega_{i-1} \times E_i \times \Omega_{i+1} \times \dots \times \Omega_n \subset \Omega$$

innerhalb des Gesamtexperiments zu interpretieren. Es gilt

$$\begin{aligned} P(E) &= \sum_{\omega_1 \in \Omega_1} \dots \sum_{\omega_i \in E_i} \dots \sum_{\omega_n \in \Omega_n} P_1(\omega_1) \cdot \dots \cdot P_j(\omega_j) \cdot \dots \cdot P_n(\omega_n) \\ &= \underbrace{\left(\sum_{\omega_1 \in \Omega_1} P_1(\{\omega_1\}) \right)}_{P_1(\Omega_1)=1} \dots \underbrace{\left(\sum_{\omega_i \in E_i} P_i(\{\omega_i\}) \right)}_{=P_i(E_i)} \dots \underbrace{\left(\sum_{\omega_n \in \Omega_n} P_n(\{\omega_n\}) \right)}_{P_n(\Omega_n)=1} = P_i(E_i). \end{aligned}$$

Mit einem weiteren Ereignis $\tilde{E} =$ „im j -ten Schritt tritt das Ereignis $\tilde{E}_j \subset \Omega_j$ ein“

$$\tilde{E} = \Omega_1 \times \dots \times \Omega_{j-1} \times \tilde{E}_j \times \Omega_{j+1} \times \dots \times \Omega_n \subset \Omega$$

folgt dann

$$E \cap \tilde{E} = \begin{cases} \Omega_1 \times \dots \times E_i \times \dots \times \tilde{E}_j \times \dots \times \Omega_n & \text{für } i \neq j \\ \Omega_1 \times \dots \times (E_i \cap \tilde{E}_j) \times \dots \times \Omega_n & \text{für } i = j \end{cases}$$

und

$$\begin{aligned} P(E \cap \tilde{E}) &= \begin{cases} \left(\sum_{\omega_i \in E_i} P_i(\{\omega_i\}) \right) \left(\sum_{\omega_j \in \tilde{E}_j} P_j(\{\omega_j\}) \right) & \text{für } i \neq j \\ \sum_{\omega \in E_i \cap \tilde{E}_j} P_i(\{\omega\}) & \text{für } i = j \end{cases} \\ &= \begin{cases} P_i(E_i) P_j(\tilde{E}_j) & \text{für } i \neq j \\ P_i(E_i \cap \tilde{E}_j) & \text{für } i = j \end{cases} = \begin{cases} P(E) P(\tilde{E}) & \text{für } i \neq j, \\ P_i(E_i \cap \tilde{E}_j) & \text{für } i = j. \end{cases} \end{aligned}$$

Resultat: zwei Ereignisse E und \tilde{E} sind automatisch unabhängig, wenn sie Ereignisse in unterschiedlichen Einzelexperimenten beschreiben. Das Produktmodell modelliert damit in der Tat unabhängige Ausführungen von Einzelexperimenten.

Bemerkung 1.59: Eine n -fache unabhängige Wiederholung eines Experiments $(\Omega_1, \mathcal{E}_1, P_1)$ ist der Spezialfall $(\Omega_1, \mathcal{E}_1, P_1) = \dots = (\Omega_n, \mathcal{E}_n, P_n)$.

Beispiel 1.60: Eine manipulierte Münze $\Omega_1 = \{K, Z\}$ mit $P(K) = 0.6$, $P(Z) = 0.4$ wird 3 mal geworfen. Mit welcher W'keit wird höchstens 1 mal Kopf geworfen?

Lösung:

$$\Omega = \{(\omega_1, \omega_2, \omega_3); \omega_i \in \{K, Z\}\},$$

$$E = \{(K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\}.$$

Die Elementarw'keiten der Elemente in E sind:

$$P(\{(K, Z, Z)\}) = P(\{(Z, K, Z)\}) = P(\{(Z, Z, K)\}) = 0.6 \cdot 0.4 \cdot 0.4,$$

$$P(\{(Z, Z, Z)\}) = 0.4 \cdot 0.4 \cdot 0.4.$$

Damit folgt:

$$P(E) = 3 \cdot 0.6 \cdot 0.4 \cdot 0.4 + 0.4 \cdot 0.4 \cdot 0.4 = 0.352.$$

Bemerkung 1.61: Ein Produktmodell $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$ kann natürlich auch dann verwendet werden, wenn die Einzelexperimente $\Omega_1, \dots, \Omega_n$ nicht unabhängig sind. Allerdings ist das Produktmodell dann nicht mit dem Produkt-W'keitsmaß

$$P(E_1 \times \dots \times E_n) = P_1(E_1) \cdot \dots \cdot P_n(E_n)$$

mit Einzelereignissen $E_i \subset \Omega_i$ auszustatten.

Eine sehr einfache, in Anwendungen aber sehr wichtige Situation ist das Bernoulli-Experiment bzw. unabhängige Wiederholungen eines solchen Experiments:

Definition 1.62:

Ein „**Bernoulli-Experiment**“ ist ein diskretes Experiment (= Modell) mit nur 2 möglichen Ausgängen 1 („Erfolg“) und 0 („Misserfolg“). Meist werden die Elementarw'keiten mit p („**Erfolgsw'keit**“) und $q = 1 - p$ („**Misserfolgsw'keit**“) bezeichnet:

$$\Omega = \{0, 1\}, \quad P(\{1\}) = p, \quad P(\{0\}) = q = 1 - p.$$

In praktischen Aufgaben tauchen oft unabhängige Wiederholungen von Bernoulli-Experimenten auf:

Satz 1.63: (Erfolgsw'keiten in wiederholten Bernoulli-Experimenten)

Ein Bernoulli-Experiment mit Erfolgsw'keit p wird n mal wiederholt. Die W'keit, dabei insgesamt genau k mal Erfolg zu haben ($k \in \{0, 1, \dots, n\}$), betragt

$$P(\text{„genau } k \text{ Erfolge“}) = \binom{n}{k} p^k q^{n-k} .$$

Beweis: Benutze den Stichprobenraum

$$\Omega = \{(\omega_1, \dots, \omega_n); \omega_i \in \{0, 1\}\}$$

(ein Produktmodell) zur Modellierung der n unabhangigen Wiederholungen. Jedes Elementarereignis $(\omega_1, \dots, \omega_n)$ mit genau k Erfolgen hat die W'keit $p^k q^{n-k}$ (dies ist die Unabhangigkeit). Ein solches Elementarereignis unterscheidet jedoch die Reihenfolgen, in denen die Erfolge und Misserfolge auftreten. Gefragt ist nach dem Ereignis

$$\text{„genau } k \text{ Erfolge“} \equiv \left\{ (\omega_1, \dots, \omega_n); \sum_{i=1}^n \omega_i = k \right\} ,$$

in dem die Reihenfolge der Erfolge nicht eingeht. Nach Hilfssatz 1.9.b) gibt es $\binom{n}{k}$ verschiedene Reihenfolgen mit genau k Erfolgen und $n - k$ Misserfolgen. Damit ist die W'keit fur genau k Erfolge ohne Berucksichtigung der Reihenfolge $\binom{n}{k} p^k q^{n-k}$.

Q.E.D.

Bezeichnung 1.64:

Die Abbildung

$$k \in \{0, 1, \dots, n\} \rightarrow P(\text{„genau } k \text{ Erfolge“}) = \binom{n}{k} p^k q^{n-k}$$

heißt **Binomialverteilung** zu den gegebenen Parametern n (Anzahl der Wiederholungen) und p (Erfolgsw'keit).

Beispiel 1.65: Zuruck zur manipulierten Munze mit $P(K) = p = 0.6$, $P(Z) = q = 0.4$ aus Beispiel 1.60. Ein Munzwurf kann als Bernoulli-Experiment aufgefasst werden. Bei n Wurfen ist die W'keit, hochsten 1 mal „Kopf“ (= „Erfolg“) zu werfen, gegeben durch

$$\begin{aligned} P(\text{„hochstens 1 Erfolg“}) &= P(\text{„genau 0 Erfolge“} \cup \text{„genau 1 Erfolg“}) \\ &\stackrel{(*)}{=} P(\text{„genau 0 Erfolge“}) + P(\text{„genau 1 Erfolg“}) \\ &= \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} \\ &= q^n + n p q^{n-1} . \end{aligned}$$

Hierbei wird in (*) die Additivität von W'keiten bei disjunkten Ereignissen benutzt (Folgerung 1.6.a). Für $n = 3$ ergibt sich (vergleiche mit Beispiel 1.60):

$$P(\text{„höchstens 1 Erfolg“}) = 0.4^3 + 3 \cdot 0.6 \cdot 0.4^2 = 0.352 .$$

Merke (Zusammenfassung) 1.66:

Ein **Produktmodell** bietet sich in Situationen an, in denen das Gesamtexperiment aus mehreren Einzelexperimenten besteht, die unabhängig voneinander durchgeführt werden. Es besteht aus geordneten Tupeln der Ergebnisse der Einzelexperimente, das W'keitsmaß ist auf den Elementarereignissen durch

$$P(\{(\omega_1, \dots, \omega_n)\}) = P(\{\omega_1\}) \cdot P(\{\omega_2\}) \cdot \dots \cdot P(\{\omega_n\})$$

gegeben, was die Unabhängigkeit der Einzelexperimente widerspiegelt. In diesem Modell sind Ereignisse automatisch unabhängig, wenn sie sich auf unterschiedliche Teilexperimente beziehen.

1.6 Einige Beispiele

In diesem Abschnitt werden einige etwas komplexere Beispiele diskutiert, in denen die meisten der bisherigen Begriffsbildungen und Aussagen zum Einsatz kommen:

Beispiel 1.67: (Verknüpfung mehrerer Experimente) Man hat 2 Münzen, eine faire ($P(K) = P(Z) = 1/2$) und eine manipulierte ($P(K) = 0.3$, $P(Z) = 0.7$), die man allerdings nicht unterscheiden kann. Man nimmt eine und möchte herausbekommen, ob es die manipulierte ist. Man wirft sie hierzu n mal und hat k mal „Zahl“. Mit welcher W'keit handelt es sich um die manipulierte Münze? (Die Idee ist: wenn bei n W'rfen etwa $n/2$ mal „Zahl“ eintritt, so dürfte es sich wohl um die faire Münzen handeln. Wird hingegen etwa $0.7 \cdot n$ mal „Zahl“ beobachtet, so ist die W'keit groß, dass es sich um die manipulierte Münze handelt.)

Lösung: Zerlege das Experiment des n -fachen Münzwurfs in das Ursachensystem F = „die Münze ist fair“ und U = „die Münze ist unfair“. Eine der beiden Münzen wird in einem ersten Bernoulli-Experiment mit der W'keit $1/2$ gezogen:

$$\Omega = F \cup U, \quad P(F) = P(U) = \frac{1}{2} .$$

Für gegebenes F bzw. U betrachten wir den einfachen Münzwurf jeweils als Bernoulli-Experiment mit den Erfolgsw'keiten $P(Z) = P(K) = 1/2$ bzw. $p = P(Z) = 0.7$, $q = P(K) = 0.3$. Sei beim n -fachen Wurf E_k das Ereignis „es wird k mal „Zahl“ geworfen“. Mit Satz 1.63 gilt:

$$P(E_k | F) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} \frac{1}{2^n}, \quad P(E_k | U) = \binom{n}{k} p^k q^{n-k} .$$

Nach Bayes 1.34 gilt:

$$\begin{aligned} P(U | E_k) &= \frac{P(E_k | U) P(U)}{P(E_k | U) P(U) + P(E_k | F) P(F)} = \frac{1}{1 + \frac{P(E_k | F) P(F)}{P(E_k | U) P(U)}} \\ &= \frac{1}{1 + \frac{1/2^n \cdot 1/2}{p^k q^{n-k} \cdot 1/2}} = \frac{1}{1 + \frac{1}{(2p)^k (2q)^{n-k}}} . \end{aligned}$$

Für $p = 0.7$, $q = 0.3$, $n = 1000$ ergeben sich beispielsweise folgende Werte. Mit Bemerkung 1.36 gilt $P(F | E_k) = 1 - P(U | E_k)$:

k	596	598	600	602	603	604	606	608
$P(U E_k)$	0.00291..	0.0156..	0.0796..	0.320..	0.523..	0.719..	0.939..	0.987..
$P(F E_k)$	0.997..	0.984..	0.920..	0.679..	0.476..	0.208..	0.0668..	0.0129..

Also: wird 600 mal „Zahl“ geworfen, handelt es sich mit großer W'keit um die faire Münze. Ab 603 mal „Zahl“ ist es wahrscheinlicher, dass es sich um die manipulierte Münze handelt. Ab 606 mal „Zahl“ kann man fast sicher davon ausgehen, dass es sich um die manipulierte Münze handelt. Bayes liefert damit ein extrem scharfes Kriterium, auf „manipuliert“ versus „fair“ schließen zu können. (Die extreme Schärfe liegt in diesem Beispiel daran, dass wir die Münzen sehr genau kennen: für die unfaire Münze ist $P(K) = 0.7$, $P(Z) = 0.3$ bekannt).

7.5.07↓

Beispiel 1.68: Ein System besteht aus 4 Komponenten vom Typ A und 16 Komponenten vom Typ B .

Typ A ist unzuverlässig: eine Komponente fällt mit der W'keit 0.1 pro Arbeitszyklus aus. Typ B ist zuverlässiger: eine Komponente fällt mit der W'keit 0.01 pro Arbeitszyklus aus.

Das System fällt sicher aus, wenn mindestens 2 Teile vom Typ A ausfallen oder mindestens 1 Teil vom Typ B ausfällt. Wenn genau 1 Teil vom Typ A ausfällt und kein Teil vom Typ B ausfällt, so fällt das Gesamtsystem mit der W'keit $1/2$ aus.

- Mit welcher W'keit fällt das Gesamtsystem in einem Arbeitszyklus aus?
- Ich will das Gesamtsystem sicherer machen. Ist es sinnvoller, Typ A oder Typ B zu verbessern, oder müssen beide Typen verbessert werden?

Lösung: Der Stichprobenraum für einen Arbeitszyklus sei

$$\begin{aligned} \Omega &= A_1 \times \dots \times A_4 \times B_1 \times \dots \times B_{16} \\ &= \{(a_1, \dots, a_4, b_1, \dots, b_{16}); a_1, \dots, b_{16} \in \{\text{defekt, ok}\}\} , \end{aligned}$$

wobei A_1, \dots, B_{16} jeweils Bernoulli-Experimente sind mit

$$\begin{aligned} p_A &= P_{A_i}(\text{defekt}) = 0.1 , & q_A &= P_{A_i}(\text{ok}) = 0.9 , \\ p_B &= P_{B_i}(\text{defekt}) = 0.01 , & q_B &= P_{B_i}(\text{ok}) = 0.99 . \end{aligned}$$

Frage a): Mit welcher W'keit fällt das Gesamtsystem in einem Arbeitszyklus aus?

Bilde zunächst ein disjunktes System von Ursachen für das Ereignis $E =$ „das Gesamtsystem fällt aus“:

$$\begin{aligned} U_1 &: \text{„kein } B \text{ fällt aus, kein } A \text{ fällt aus“}, \\ U_2 &: \text{„kein } B \text{ fällt aus, genau ein } A \text{ fällt aus“}, \\ U_3 &: \text{„kein } B \text{ fällt aus, mindestens zwei } A \text{ fallen aus“}, \\ U_4 &: \text{„mindestens ein } B \text{ fällt aus“}. \end{aligned}$$

Das Kriterium für die Wahl dieser Zerlegung ist:

- i) alle U_i müssen disjunkt sein,
- ii) alle Ausfallmöglichkeiten müssen abgedeckt sein: $E \subset \cup_i U_i$,
- iii) wir müssen $P(E | U_i)$ kennen,
- iv) $P(U_i)$ sollte ohne allzu großen kombinatorischen Aufwand berechenbar sein.

Die obige Zerlegung erfüllt diese Kriterien:

- i): Ist offensichtlich erfüllt.
- ii): Es gilt sogar $\cup_i U_i = \Omega$.
- iii):

$$P(E | U_1) = 0, \quad P(E | U_2) = \frac{1}{2}, \quad P(E | U_3) = 1, \quad P(E | U_4) = 1.$$

Fragen wir nach der W'keit des Ausfalls des Gesamtsystems, so gilt nach der Formel der totalen W'keit 1.31:

$$P(E) = \sum_i P(E | U_i) \cdot P(U_i) = \frac{1}{2} \cdot P(U_2) + P(U_3) + P(U_4). \quad (\#)$$

iv) Es verbleibt, $P(U_2)$, $P(U_3)$, $P(U_4)$ zu bestimmen. Dies ist nicht schwierig, da es sich um unabhängige Wiederholungen von 2 verschiedenen Bernoulli-Experimenten $A_1 = \dots = A_4$, $B_1 = \dots = B_{16}$ handelt. Betrachten wir zunächst die beiden Typen getrennt. Mit Satz 1.63 gilt:

$$P(\text{„kein } A \text{ defekt“}) = \binom{4}{0} \cdot p_A^0 q_A^4 = q_A^4,$$

$$P(\text{„genau ein } A \text{ defekt“}) = \binom{4}{1} \cdot p_A^1 q_A^3 = 4 p_A q_A^3,$$

$$\begin{aligned} P(\text{„mindestens zwei } A \text{ defekt“}) &= 1 - P(\text{„kein } A \text{ defekt“}) - P(\text{„genau ein } A \text{ defekt“}) \\ &= 1 - q_A^4 - 4 p_A q_A^3, \end{aligned}$$

$$P(\text{„kein } B \text{ defekt“}) = \binom{16}{0} \cdot p_B^0 q_B^{16} = q_B^{16},$$

$$P(\text{„mindestens ein } B \text{ defekt“}) = 1 - P(\text{„kein } B \text{ defekt“}) = 1 - q_B^{16}.$$

Da die A - und B -Teile unabhängig voneinander ausfallen können, werden die Bernoulli-Experimentwiederholungen für den Ausfall der A - und B -Teile unabhängig durch-

geführt. Damit gilt

$$\begin{aligned}
 P(U_2) &= P(\text{„kein } B \text{ defekt, genau ein } A \text{ defekt“}) \\
 &= P(\text{„kein } B \text{ defekt}) P(\text{genau ein } A \text{ defekt“}) \\
 &= q_B^{16} 4 p_A q_A^3 \approx 0.2483, \\
 P(U_3) &= P(\text{„kein } B \text{ defekt, mindestens zwei } A \text{ defekt“}) \\
 &= P(\text{„kein } B \text{ defekt“}) P(\text{„mindestens zwei } A \text{ defekt“}) \\
 &= q_B^{16} (1 - q_A^4 - 4 p_A q_A^3) \approx 0.04453, \\
 P(U_4) &= P(\text{„mindestens ein } B \text{ defekt“}) = 1 - q_B^{16} \approx 0.1485.
 \end{aligned}$$

Mit Gleichung (#) folgt die Ausfallw'keit des Gesamtsystems:

$$\begin{aligned}
 P(E) &= \frac{1}{2} q_B^{16} 4 p_A q_A^3 + q_B^{16} (1 - q_A^4 - 4 p_A q_A^3) + 1 - q_B^{16} \\
 &= 1 - q_A^4 q_B^{16} - 2 p_A q_A^3 q_B^{16}.
 \end{aligned}$$

Für $p_A = 0.1$, $q_A = 0.9$, $p_B = 0.01$, $q_B = 0.99$ ergibt sich der konkrete Wert

$$P(E) = 0.3172..$$

Frage b): Ich will das Gesamtsystem sicherer machen. Ist es sinnvoller, Typ A oder Typ B zu verbessern, oder müssen beide Typen verbessert werden?

Wir berechnen die W'keiten der Ereignisse

$$\begin{aligned}
 U_A &= \text{„mindestens ein } A\text{-Teil fällt aus, kein } B\text{-Teil fällt aus“}, \\
 U_B &= \text{„mindestens ein } B\text{-Teil fällt aus, kein } A\text{-Teil fällt aus“}, \\
 U_{AB} &= \text{„mindestens ein } A\text{-Teil fällt aus und mindestes ein } B\text{-Teil fällt aus“}
 \end{aligned}$$

unter dem vorgegebenen Ereignis $E = \text{„das Gesamtsystem ist ausgefallen“}$. Dann interpretieren wir

$$\begin{aligned}
 P(U_A|E) &= P(\text{„nur die } A\text{-Teile waren Schuld am Ausfall“}), \\
 P(U_B|E) &= P(\text{„nur die } B\text{-Teile waren Schuld am Ausfall“}), \\
 P(U_{AB}|E) &= P(\text{„sowohl } A\text{- als auch } B\text{-Teile waren Schuld am Ausfall“}).
 \end{aligned}$$

Hierbei bilden U_A, U_B, U_{AB} ein neues disjunktes Ursachensystem für das Ereignis E . In Analogie zu den Rechnungen in a) folgt:

$$\begin{aligned}
 P(U_A) &= P(\text{„mindestens ein } A \text{ defekt“}) P(\text{„kein } B \text{ defekt“}) \\
 &= \left(1 - P(\text{„kein } A \text{ defekt“})\right) P(\text{„kein } B \text{ defekt“}) \\
 &= (1 - q_A^4) q_B^{16} \approx 0.2928, \\
 P(U_B) &= P(\text{„mindestens ein } B \text{ defekt“}) P(\text{„kein } A \text{ defekt“}) \\
 &= \left(1 - P(\text{„kein } B \text{ defekt“})\right) P(\text{„kein } A \text{ defekt“}) \\
 &= (1 - q_B^{16}) q_A^4 \approx 0.09746, \\
 P(U_{AB}) &= P(\text{„mindestens ein } B \text{ defekt“}) P(\text{„mindestens ein } A \text{ defekt“}) \\
 &= \left(1 - P(\text{„kein } B \text{ defekt“})\right) \left(1 - P(\text{„kein } A \text{ defekt“})\right) \\
 &= (1 - q_B^{16}) (1 - q_A^4) \approx 0.05108.
 \end{aligned}$$

Wir benutzen den Satz von Bayes 1.34:

$$P(U_i | E) = \frac{P(E | U_i) P(U_i)}{P(E)} .$$

Die bedingten W'keiten $P(E | U_B) = P(E | U_{AB}) = 1$ sind bekannt (siehe Formulierung der Aufgabe). Damit folgt

$$P(U_B | E) = \frac{1 \cdot P(U_B)}{P(E)} \approx 0.3072 , \quad P(U_{AB} | E) = \frac{1 \cdot P(U_{AB})}{P(E)} \approx 0.1610 .$$

Es fehlt noch $P(U_A | E)$. Nach Bemerkung 1.36 gilt

$$P(U_A | E) = 1 - P(U_B | E) - P(U_{AB} | E) \approx 0.5318 .$$

Zu Übungszwecken berechnen wir diesen Wert auch noch anders. Da $P(E | U_A)$ nicht direkt gegeben ist, können wir den Satz von Bayes nicht unmittelbar heranziehen. Zur Berechnung benutzen wir $U_A = U_2 \cup U_3$ mit den Ursachen U_2, U_3 aus a). Es gilt

$$\begin{aligned} P(U_A | E) &= P(U_2 \cup U_3 | E) = \frac{P(E \cap (U_2 \cup U_3))}{P(E)} \\ &= \frac{P((E \cap U_2) \cup (E \cap U_3))}{P(E)} \quad (\text{disjunkt}) \quad \frac{P(E \cap U_2) + P(E \cap U_3)}{P(E)} \\ &= P(U_2 | E) + P(U_3 | E) . \end{aligned}$$

Die bedingten W'keiten $P(E | U_2) = 1/2$, $P(E | U_3) = 1$ sind aus der Aufgabenstellung bekannt. Damit folgt per Bayes

$$P(U_2 | E) = \frac{1/2 \cdot P(U_2)}{P(E)} , \quad P(U_3 | E) = \frac{1 \cdot P(U_3)}{P(E)}$$

und mit den Werten aus a):

$$\begin{aligned} P(U_A | E) &= P(U_2 | E) + P(U_3 | E) = \frac{1/2 \cdot P(U_2) + P(U_3)}{P(E)} \\ &\approx \frac{0.5 \cdot 0.2483 + 0.04453}{0.3172} \approx 0.5318 . \end{aligned}$$

Zusammenfassung:

$$\begin{array}{lll} P(\text{„nur A-Teile waren schuld“}) & = & P(U_A | E) \approx 0.5318 , \\ P(\text{„nur B-Teile waren schuld“}) & = & P(U_B | E) \approx 0.3072 , \\ P(\text{„sowohl A- als auch B-Teile waren schuld“}) & = & P(U_{AB} | E) \approx 0.1610 . \\ & & \hline & & 1.0000 \end{array}$$

Wenn ich es mir nur leisten kann, einen Typ zu verbessern, so sollte dies der Typ A sein. Dies wird aber keine dramatische Verbesserung erbringen: man muss beide Typen verbessern, um das Gesamtsystem deutlich sicherer zu machen.

Kapitel 2

Zufallsvariablen

Motivation: Oft interessiert man sich nicht für den kompletten Stichprobenraum (den man häufig gar nicht genau kennt), sondern nur für Teilaspekte. Beispiel: Wurf mehrerer Würfel. Die genaue Aufteilung der Ergebnisse auf die einzelnen Würfel ist irrelevant, interessant ist nur die Augensumme. Dieser Wert ist reell und verhält sich zufällig. Die Idee ist, statt des Gesamtexperiments (bei n Würfeln)

↓9.5.07

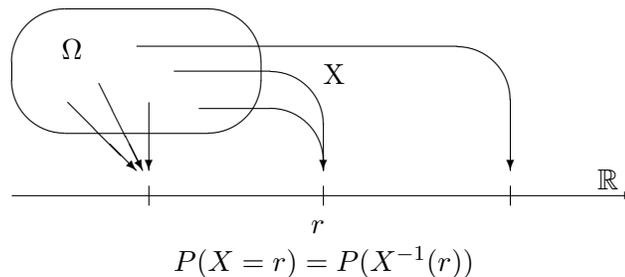
$$\Omega = \{\omega = (\omega_1, \dots, \omega_n); \omega_i \in \{1, \dots, 6\}\}$$

nur den Wert $X(\omega) = \sum_{i=1}^n \omega_i$ zu betrachten. Wir betrachten die Menge aller möglichen Werte

$$\tilde{\Omega} = X(\Omega) = \{n, n+1, \dots, 6n\}$$

als neues (aber nicht mehr kombinatorisches) Zufallsexperiment, für das wir die Elementarw'keiten $P(X = r)$ für $r = n, \dots, 6n$ durch Abzählen bestimmen können.

In diesem Fall ist $\tilde{\Omega}$ eine endliche Teilmenge von \mathbb{R} (notwendigerweise endlich, da Ω schon endlich war). Wir wollen nun eine allgemeine Theorie aufbauen, in der wir beliebige Abbildungen $X : \Omega \mapsto \mathbb{R}$ über beliebigen Modellen betrachten und allgemein \mathbb{R} mit einem „Bild-W'keitsmaß“ ausstatten, das von den W'keiten des ursprünglichen Experiments stammt:



Wir haben mit dem Konzept kontinuierlicher W'räume 1.18 bereits eine Möglichkeit vorgestellt, den speziellen Stichprobenraum \mathbb{R} über eine W'keitsdichte

$\rho(r)$ mit einem W'keitsma zu versehen:

$$P(r_1 < X \leq r_2) = \int_{r_1}^{r_2} \rho(r) dr .$$

Ist Ω diskret, so nimmt X nur diskrete Werte an. Leider knnen wir mit Hilfe einer Dichte keine W'keiten $P(X = r)$ f'ur einzelne Punkte definieren, da ein Integral 'uber ein Intervall der L'ange 0 immer 0 ist. Hier hilft ein einfacher Trick: wir geben die W'keiten statt mit einer Dichte mit der Stammfunktion

$$F_X(r) = P(-\infty < X \leq r) \quad \left(= \int_{-\infty}^r \rho(r) dr \right)$$

der Dichte an. Also: wir versehen den Stichprobenraum \mathbb{R} mit einem W'keitsma, indem wir die W'keiten $P(-\infty < X \leq r)$ vorgeben. Damit lassen sich diskrete Modelle (ohne Dichten) und kontinuierliche Modelle (mit Dichten) einheitlich behandeln.

2.1 Definitionen

Definition 2.1: (Zufallsvariable)

Sei (Ω, \mathcal{E}, P) ein beliebiges stochastisches Modell.

- i) Eine (reelle) **Zufallsvariable** ist eine Abbildung $X : \Omega \mapsto \mathbb{R}$, f'ur welche die Urbilder aller Intervalle der Form $(-\infty, r]$ Ereignisse sind:

$$X^{-1}((-\infty, r]) = \{\omega \in \Omega; X(\omega) \leq r\} \in \mathcal{E} .$$

- ii) Die Abbildung

$$r \mapsto F_X(r) = P(X^{-1}((-\infty, r])) \equiv P(X \leq r)$$

heißt (**kumulative**) **Verteilungsfunktion** von X (engl: cumulative distribution function = CDF)

- iii) X heißt **diskret**, wenn die Bildmenge $X(\Omega)$ abz'ahlbar ist. X heißt **kontinuierlich**, wenn $F_X(r)$ nach r differenzierbar ist. Die Ableitung $\rho(r) = F'_X(r)$ nennt man die **Dichte** der Verteilungsfunktion F_X (engl: probability density function = PDF)

Bemerkung 2.2: Ist das Modell (Ω, \mathcal{E}, P) diskret (speziell, kombinatorisch), so ist wegen $\mathcal{E} = \mathcal{P}(\Omega) =$ Potenzmenge von Ω trivialerweise jede Abbildung $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable.

Beispiel 2.3: Betrachten wir noch einmal Beispiel 1.16: werfe 2 mal mit einem fairen Würfel, betrachte die Augensumme X :

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}, \quad X : (\omega_1, \omega_2) \in \Omega \rightarrow \omega_1 + \omega_2.$$

Das Bild $X(\Omega) = \{2, 3, \dots, 12\}$ ist eine diskrete Teilmenge von \mathbb{R} . Wir hatten in Beispiel 1.16 bereits einige der W'keiten $p_k = P(X = k) \equiv P(X^{-1}(\{k\}))$ berechnet, wir vervollständigen hier die Rechnung:

$$\begin{aligned} p_2 &= P(X = 2) = P(\{(1, 1)\}) = \frac{1}{36}, \\ p_3 &= P(X = 3) = P(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18}, \\ p_4 &= P(X = 4) = P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36} = \frac{1}{12}, \\ p_5 &= P(X = 5) = P(\{(1, 4), (2, 3), \dots\}) = \frac{1}{9}, \\ p_6 &= P(X = 6) = \dots = \frac{5}{36}, \\ p_7 &= P(X = 7) = \dots = \frac{1}{6}, \\ p_8 &= p_6 = \frac{5}{36}, \\ p_9 &= p_5 = \frac{1}{9}, \\ p_{10} &= p_4 = \frac{1}{12}, \\ p_{11} &= p_3 = \frac{1}{18}, \\ p_{12} &= p_2 = \frac{1}{36}. \end{aligned}$$

Offensichtlich gilt für beliebiges $r \in \mathbb{R}$:

$$\begin{aligned} \boxed{F_X(r)} &= P(X \leq r) \equiv P(X^{-1}((-\infty, r])) = P(X^{-1}((-\infty, r] \cap \{2, 3, \dots, 12\})) \\ &= \sum_{\substack{k \in \{2, \dots, 12\} \\ k \leq r}} P(X^{-1}(\{k\})) = \boxed{\sum_{\substack{k \in \{2, \dots, 12\} \\ k \leq r}} P(X = k)}. \end{aligned}$$

Beispielsweise ergibt sich damit für jedes $-\infty < r < 2$ derselbe Wert

$$F_X(r) = P(X \leq r) = 0.$$

Für jedes $2 \leq r < 3$ ergibt sich derselbe Wert

$$F_X(r) = P(X \leq r) = P(X^{-1}(\{2\})) = p_2 = \frac{1}{36}.$$

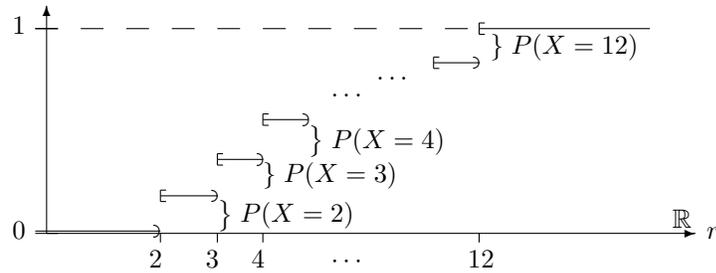
Für jedes $3 \leq r < 4$ ergibt sich derselbe Wert

$$F_X(r) = P(X \leq r) = P(X^{-1}(\{2, 3\})) = p_2 + p_3 = \frac{1}{36} + \frac{1}{18} = \frac{1}{12}.$$

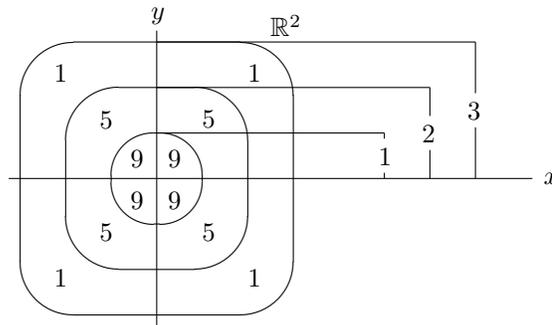
Usw. Für jedes $12 \leq r < \infty$ ergibt sich derselbe Wert

$$F_X(r) = P(X \leq r) = P(X^{-1}(\{2, \dots, 12\})) = p_2 + \dots + p_{12} = 1.$$

Ergebnis: $F_X(r)$ ist eine Treppenfunktion, die bei $r = -\infty$ mit 0 startet und monoton bis 1 anwächst. Die Sprungstellen sind dort, wo X Werte annimmt. Die Sprunghöhen (der Zuwachs, wenn man eine Stelle $X = k$ überschreitet) sind jeweils $p_k = P(X = k)$:



Beispiel 2.4: (Vergleiche auch mit Beispiel 1.22) Ein Schütze schießt auf folgende im Ursprung des \mathbb{R}^2 zentrierte Zielscheibe vom Radius 3 mit 3 kreisförmigen Ringen mit den Punktzahlen 1, 5, 9:



Wir geben vor, dass ein Treffer mit der W'keitsdichte

$$\rho(t) = 2te^{-t^2}$$

im Abstand t vom Ursprung landet. Betrachte $X =$ „Abstand des Treffers vom Zentrum“

$$X : t \in \Omega = [0, \infty) \rightarrow t \in \mathbb{R}$$

mit der Verteilungsfunktion

$$F_X(r) = P(X \leq r) = \begin{cases} 0 & \text{für } r < 0, \\ \int_0^r 2te^{-t^2} dt & \text{für } 0 \leq r \end{cases} = \begin{cases} 0 & \text{für } r < 0, \\ 1 - e^{-r^2} & \text{für } 0 \leq r. \end{cases}$$

Ergebnis: Die Verteilungsfunktion $F_X(r)$ ist monoton steigend: sie beginnt bei $r = -\infty$ mit 0, bleibt konstant 0 bis $r = 0$, dann steigt sie monoton in der Form $1 - e^{-r^2}$ bis 1 an.

Beispiel 2.5: Betrachte erneut die Zielscheibe des letzten Beispiels mit den Zufallsvariablen $X =$ „Abstand des Treffers vom Zentrum“ und $Y =$ „die geschossene Punktzahl“:

$$Y : t \in \Omega = [0, \infty) \rightarrow \begin{cases} 0 & \text{für } 3 < t, \\ 1 & \text{für } 2 < t \leq 3, \\ 5 & \text{für } 1 < t \leq 2, \\ 9 & \text{für } 0 \leq t \leq 1. \end{cases}$$

Obwohl das unterliegende Modell $\Omega = [0, \infty)$ kontinuierlich ist, ist die Zufallsvariable diskret. Die Variable nimmt nur 4 mögliche Wert an: $Y(\Omega) = \{0, 1, 5, 9\}$. Die Wahrscheinlichkeiten für die geschossenen Punktzahlen sind

$$\begin{aligned} p_0 &= P(Y = 0) = P(3 < X) = \int_3^\infty \rho(t) dt = \left[-e^{-t^2} \right]_{t=3}^{t=\infty} = e^{-9} - 0 \approx 0.000123\dots, \\ p_1 &= P(Y = 1) = P(2 < X \leq 3) = \int_2^3 \rho(t) dt = \left[-e^{-t^2} \right]_{t=2}^{t=3} = e^{-4} - e^{-9} \approx 0.01819\dots, \\ p_5 &= P(Y = 5) = P(1 < X \leq 2) = \int_1^2 \rho(t) dt = \left[-e^{-t^2} \right]_{t=1}^{t=2} = e^{-1} - e^{-4} \approx 0.3495\dots, \\ p_9 &= P(Y = 9) = P(0 \leq X \leq 1) = \int_0^1 \rho(t) dt = \left[-e^{-t^2} \right]_{t=0}^{t=1} = 1 - e^{-1} \approx 0.6321\dots \end{aligned}$$

1.0000

Die Verteilungsfunktion von Y ist wieder eine Treppenfunktion:

$$F_Y(r) = \begin{cases} 0 & \text{für } r < 0, \\ p_0 & \text{für } 0 \leq r < 1, \\ p_0 + p_1 & \text{für } 1 \leq r < 5, \\ p_0 + p_1 + p_5 & \text{für } 5 \leq r < 9, \\ p_0 + p_1 + p_5 + p_9 = 1 & \text{für } 9 \leq r. \end{cases}$$

Notation 2.6:

Betrachte eine Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ über einem beliebigen Modell (Ω, \mathcal{E}, P) . Für $A \subset \mathbb{R}$ benutzen wir die vereinfachte Notation

$$P(X \in A) \equiv P(X^{-1}(A)) = P(\{\omega \in \Omega; X(\omega) \in A\}) .$$

Beispiele:

$$\begin{aligned} P(X \leq r) &\equiv P(X \in (-\infty, r]) \equiv P(X^{-1}((-\infty, r])) , \\ P(r_1 < X \leq r_2) &\equiv P(X \in (r_1, r_2]) \equiv P(X^{-1}((r_1, r_2])) . \end{aligned}$$

Rechenregeln 2.7:

Einige wichtige Rechenregeln in dieser Notation sind:

- a) für disjunkte Mengen $A_1, A_2 \subset \mathbb{R}$, $A_1 \cap A_2 = \emptyset$, gilt

$$P(X \in A_1 \cup A_2) = P(X \in A_1) + P(X \in A_2) .$$

Z.B., für $r_1 \leq r_2$, $[r_1, r_2] = \{r_1\} \cup (r_1, r_2] = A_1 \cup A_2$:

$$P(r_1 \leq X \leq r_2) = P(X = r_1) + P(r_1 < X \leq r_2) .$$

(beachte Folgerung 1.6: die Urbilder $X^{-1}(A_1)$ und $X^{-1}(A_2)$ sind wieder disjunkt und es gilt $X^{-1}(A_1) \cup X^{-1}(A_2) = X^{-1}(A_1 \cup A_2)$.)

- b) Für $A_1 \subset A_2 \subset \mathbb{R}$ gilt die disjunkte Zerlegung $A_2 = A_1 \cup (A_2 \setminus A_1)$ und damit

$$P(A_2 \setminus A_1) = P(A_2) - P(A_1) .$$

Z.B., für $r_1 < r_2$, $A_1 = (-\infty, r_1] \subset A_2 = (-\infty, r_2]$:

$$P(r_1 < X \leq r_2) = P(X \leq r_2) - P(X \leq r_1) .$$

Folgerung 2.8:

Für $r_1 < r_2$ gilt: $P(r_1 < X \leq r_2) = F_X(r_2) - F_X(r_1) .$

Also: W'keiten auf Intervallen (zunächst nur halboffene Intervalle) sind aus F_X konstruierbar! In der Tat kann aus F_X ein W'keitsmaß auf allen („vernünftigen“) Teilmengen in \mathbb{R} konstruiert werden.

14.5.07↓

Der nächste Satz beschreibt einige Eigenschaften, die jede Verteilungsfunktion über einem beliebigen Modell hat:

Satz 2.9: (einige technische Eigenschaften von Verteilungsfunktionen)

- 1) Es gilt $0 \leq F_X(r) \leq 1$ für alle $r \in \mathbb{R}$.
- 2) $F_X(r)$ ist monoton wachsend: $F_X(r_1) \leq F_X(r_2)$ für $r_1 < r_2$.
- 3) $\lim_{r \rightarrow -\infty} F_X(r) = 0$, $\lim_{r \rightarrow \infty} F_X(r) = 1$.
- 4) $F_X(r)$ ist rechtsseitig stetig: $F_X(r) = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r + \epsilon)$.
- 5) $F_X(r)$ hat höchstens abzählbar viele Unstetigkeitsstellen, diese sind Sprungstellen. An einer Sprungstelle r gilt für die Sprunghöhe:

$$F_X(r) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r - \epsilon) = P(X = r) .$$

Beweis:

1) Klar, da $F_X(r)$ eine W'keit ist.

2) Klar, da $\{\omega \in \Omega; X(\omega) \leq r_1\} \subset \{\omega \in \Omega; X(\omega) \leq r_2\}$ für $r_1 < r_2$ und die W'keiten nach Folgerung 1.6.c) monoton wachsen, wenn die Mengen größer werden.

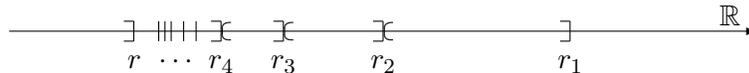
3) Betrachte die disjunkten Ereignisse $E_0 = \{\omega \in \Omega; -\infty < X(\omega) \leq 0\}$ und $E_i = \{\omega \in \Omega; i - 1 < X(\omega) \leq i\}$ für $i = 1, 2, \dots$. Es gilt $F_X(n) = P(\bigcup_{i=0}^n E_i) = \sum_{i=0}^n P(E_i)$. Für unendliche abzählbare Vereinigungen disjunkter Ereignisse gilt mit der „ σ -Additivität“ in Definition 1.2.iii.2):

$$P\left(\bigcup_{i=0}^{\infty} E_i\right) = \sum_{i=0}^{\infty} P(E_i) = \lim_{n \rightarrow \infty} \sum_{i=0}^n P(E_i) = \lim_{n \rightarrow \infty} F_X(n).$$

Es gilt $\bigcup_{i=0}^{\infty} E_i = \Omega$ und $P(\Omega) = 1$. Da F_X monoton ist, folgt der Grenzwert auch für beliebiges nichtganzzahliges $n \in \mathbb{R}$. Mit einem ähnlichen Argument zeigt man $\lim_{r \rightarrow -\infty} F_X(r) = 0$, indem man die leere Menge als Schnitt der Mengen $X^{-1}((-\infty, i])$ (mit negativem i) darstellt.

4) Sei (r_i) eine beliebige streng monoton fallende, gegen r konvergierende Folge. Es gilt:

$$X^{-1}((-\infty, r_n]) = X^{-1}((-\infty, r]) \cup \bigcup_{i=n}^{\infty} X^{-1}((r_{i+1}, r_i]).$$



Mit der σ -Additivität folgt hieraus

$$\underbrace{P(X \leq r_n)}_{F_X(r_n)} = \underbrace{P(X \leq r)}_{F_X(r)} + \sum_{i=n}^{\infty} P(X^{-1}((r_{i+1}, r_i])).$$

Für jede konvergierende Reihe $\sum_{i=n}^{\infty} p_i$ gilt aber $\lim_{n \rightarrow \infty} \sum_{i=n}^{\infty} p_i = 0$. Damit folgt $\lim_{n \rightarrow \infty} F_X(r_n) = F_X(r)$.

5) Sei (r_i) eine beliebige streng monoton steigende, gegen r konvergierende Folge. Es gilt:

$$X^{-1}((-\infty, r)) = X^{-1}((-\infty, r_1]) \cup \bigcup_{i=1}^{\infty} X^{-1}((r_i, r_{i+1}])).$$



Mit der σ -Additivität folgt hieraus

$$P(X < r) = \underbrace{P(X \leq r_1)}_{F_X(r_1)} + \sum_{i=1}^{\infty} \underbrace{P(r_i < X \leq r_{i+1})}_{F_X(r_{i+1}) - F_X(r_i)}.$$

Mit $\sum_{i=1}^{\infty} \dots = \lim_{n \rightarrow \infty} \sum_{i=1}^n \dots$ folgt

$$\lim_{n \rightarrow \infty} \underbrace{F_X(r_1) + \sum_{i=1}^{n-1} (F_X(r_{i+1}) - F_X(r_i))}_{F_X(r_n)} = P(X < r),$$

also $\lim_{n \rightarrow \infty} F_X(r_n) = P(X < r)$. Es folgt die Behauptung, dass die Sprunghöhe einer Unstetigkeitsstelle r die Interpretation $P(X = r)$ hat:

$$P(X = r) = P(X \leq r) - P(X < r) = F_X(r) - \lim_{n \rightarrow \infty} F_X(r_n).$$

Es kann höchstens abzählbar viele solcher Sprünge geben. Sei dazu $S \subset \mathbb{R}$ die Menge aller Unstetigkeitspunkte von F_X . Sei

$$S_n = \left\{ r \in \mathbb{R}; F_X(r) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r - \epsilon) \geq \frac{1}{n} \right\} \subset S$$

die Menge aller Unstetigkeitsstellen mit einer Sprunghöhe $\geq 1/n$. Wegen der Monotonie von F_X kann diese Menge maximal n Elemente enthalten, da $F_X(r) \in [0, 1]$. Mit $S = S_1 \cup S_2 \cup \dots$ ist S eine abzählbare Vereinigung endlicher Mengen und damit abzählbar.

Q.E.D.

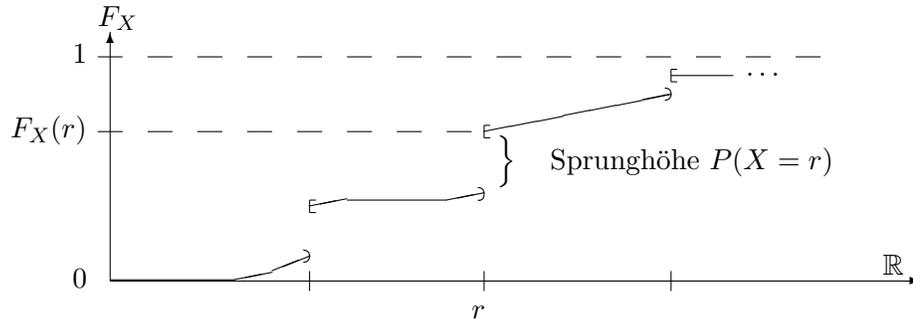
15.5.07↓

Interpretation 2.10:

Anstatt sich die technischen Formulierungen des letzten Satzes einzuprägen, dürfte es einfacher sein, sich das folgende Bild zu merken. Der Satz besagt nur, dass dieses Bild typisch ist:

- 1-3) F_X ist monoton (aber nicht unbedingt streng monoton) von 0 bis 1 wachsend.
- 4) Jeder Funktionswert stimmt mit dem Grenzwert von rechts überein (F_X ist „rechtsseitig stetig“).
- 5) Der Graph von F_X setzt sich aus Stetigkeitsintervallen und höchstens abzählbar vielen Sprungstellen zusammen.

An Sprungstellen r ist die Sprunghöhe gleich $P(X = r)$.



In Verallgemeinerung von Folgerung 2.8 gilt:

Folgerung 2.11:

a) Für $r_1 < r_2$ gilt:
$$P(r_1 < X \leq r_2) = F_X(r_2) - F_X(r_1) .$$

b) Für $r_1 \leq r_2$ gilt:
$$P(r_1 \leq X \leq r_2) = F_X(r_2) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon) .$$

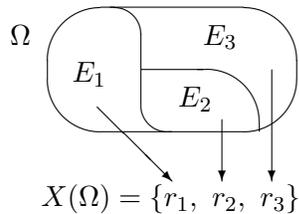
Beweis: a) war schon in Folgerung 2.8 gezeigt worden. b) folgt mit

$$P(r_1 \leq X \leq r_2) = P(X = r_1) + P(r_1 < X \leq r_2)$$

aus Teil 5) des Satzes 2.9 und a).

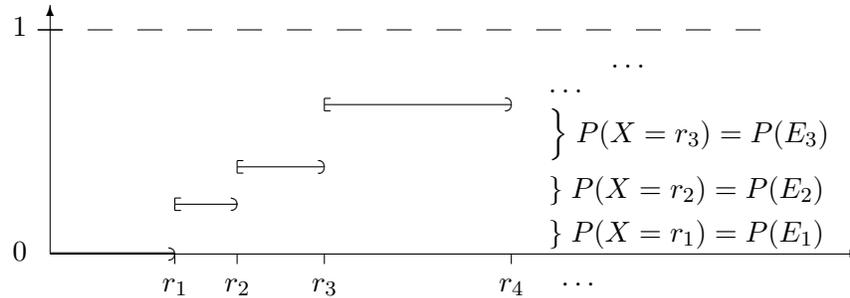
Q.E.D.

Bemerkung 2.12: Für eine diskrete Variable $X(\Omega) = \{r_1, r_2, \dots\}$ zerlegt sich der Stichprobenraum $\Omega = \cup_i E_i$ in disjunkte Urbilder $E_i = X^{-1}(\{r_i\})$ der Werte, die X annimmt, z.B.:

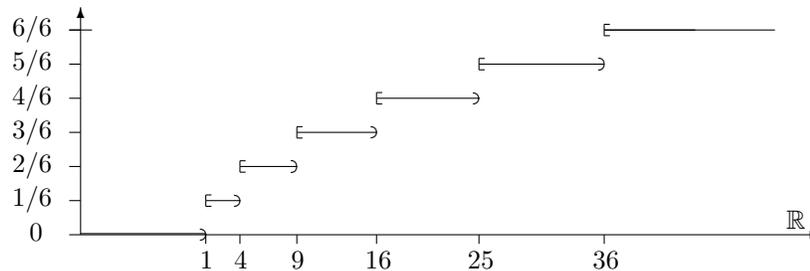


Damit ist F_X für diskrete Variable immer eine Treppenfunktion:

$$F_X(r) = P(X \leq r) = \sum_{r_i \leq r} P(E_i) = \sum_{r_i \leq r} P(X = r_i) .$$



Beispiel 2.13: Ein Wurf mit einem fairen Würfel: $\Omega = \{1, \dots, 6\}$. Betrachte $X : \omega \mapsto \omega^2$, also $X(\Omega) = \{1, 4, 9, 16, 25, 36\}$:



Zusammenfassung 2.14:

Es gibt mehrere Möglichkeiten, eine Zufallsvariable zu interpretieren:

i) Als Abbildung $X : \Omega \rightarrow \mathbb{R}$. Wenn man das unterliegende Experiment Ω kennt, hat man vollständiges Wissen über das zufällige Verhalten der Werte, die X annimmt.

ii) Man kann sich eine Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ mit Verteilungsfunktion F_X auch selbst als ein stochastisches Modell $(\mathbb{R}, \mathcal{B}, P)$ vorstellen. Hierbei ist \mathcal{B} die Menge aller Teilmengen von \mathbb{R} , über denen man eine Integrationstheorie aufbauen kann (die sogenannten „Borel“-Mengen). Speziell sind Intervalle $(r_1, r_2], [r_1, r_2], (-\infty, r]$, abzählbare Vereinigungen und Schnitte solcher Intervalle, die Komplemente etc. Ereignisse in \mathcal{B} . Die Angabe der Verteilungsfunktion F_X ist genauso gut wie die Vorgabe der W 'keiten auf den Ereignissen: das W 'keitsmaß P auf den Teilmengen von \mathbb{R} ist aus F_X konstruierbar! Z.B. gilt für die „Basistypen“ von Ereignissen in \mathcal{B} :

$$\begin{aligned}
\text{a)} \quad & P((-\infty, r]) = F_X(r), \\
\text{b)} \quad & P((r_1, r_2]) = F_X(r_2) - F_X(r_1), \\
\text{c)} \quad & P([r_1, r_2]) = F_X(r_2) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon), \\
\text{d)} \quad & P(\{r\}) = P([r, r]) = F_X(r) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r - \epsilon).
\end{aligned}$$

Hierbei ist d) der Spezialfall $r = r_1 = r_2$ von c). Für disjunkte Vereinigungen, Komplemente etc. von Mengen $E, E_i \in \mathcal{B}$ ergeben sich die W'keiten dann durch $P(\cup_i E_i) = \sum_i P(E_i)$, $P(\mathbb{R} \setminus E) = 1 - P(E)$ etc.

Mit den obigen Regeln kann $P(E)$ für alle „interessanten“ Teilmengen E von \mathbb{R} aus F_X konstruiert werden.

Beispielsweise:

$$\begin{aligned}
P((r_1, r_2)) &= \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r_2 - \epsilon) - F_X(r_1), \\
P([r_1, r_2)) &= \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r_2 - \epsilon) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F_X(r_1 - \epsilon),
\end{aligned}$$

In der Sichtweise ii) kann man das unterliegende Experiment Ω eigentlich vergessen: es diene nur dazu, die W'keiten vom Raum Ω auf die reelle Achse hinüberzuziehen und als Verteilungsfunktion F_X zu kodieren. Dies ergibt ein W'maß auf \mathbb{R} , welches sich – wie oben beschrieben – auf den geschlossenen/offenen/halboffenen Intervallen einfach ergibt. Man hat in dieser Sichtweise aber gewisse Informationen verloren, wenn man das unterliegende Ω vergisst. Zwar kann man alle gewünschten Aussagen über X machen, aber man kann nicht mehr mehrere Zufallsvariablen X_1, X_2 über demselben Ω vergleichen (z.B., weiss man nicht, wie sich die Summe $X_1 + X_2$ verhält, wenn man nur F_{X_1} und F_{X_2} kennt).

Bemerkung 2.15: Im wichtigen Spezialfall diskreter Zufallsvariabler mit $X(\Omega) = \{r_1, r_2, \dots\} \subset \mathbb{R}$ ist die Interpretation als Zufallsexperiment auf dem Stichprobenraum \mathbb{R} mit einem durch F_X (Treppenfunktion) gegebenen W'maß ziemlich gekünstelt und unnötig kompliziert. Alle Informationen stecken in den Sprunghöhen der Unstetigkeitsstellen, und es ist technisch wesentlich bequemer, sich die Zufallsvariable direkt als diskretes Modell $(\Omega_X, \mathcal{E}_X, P_X)$ mit $\Omega_X = X(\Omega) = \{r_1, r_2, \dots\}$ und $\mathcal{E}_X = \mathcal{P}(\Omega_X)$ vorzustellen, wobei das W'maß durch die Elementarw'keiten

$$P_X(\{r_i\}) = P(X = r_i) = P(X^{-1}(\{r_i\}))$$

eindeutig festgelegt ist.

Der Vollständigkeit halber soll noch ein in der Literatur vielfach verwendeter Begriff erwähnt werden, der speziell in der Statistik z.B. beim Testen von Hypothesen eine wichtige Rolle spielt: die Quantilfunktion. Sie ist die Inverse der (kumulativen) Verteilungsfunktion F_X . Für nicht streng monotone Verteilungsfunktionen, die im mathematischen Sinne nicht invertierbar sind, vereinbart man:

Definition 2.16: (Die Quantilfunktion einer Verteilung)

Sei F_X die Verteilungsfunktion einer Zufallsvariable X . Die **Quantilfunktion** der Verteilung ist:

$$Q_X : p \in (0, 1] \rightarrow \min \{r; p \leq F_X(r)\} \in \mathbb{R} .$$

Da F_X rechtsseitig stetig ist, existiert das Minimum für jedes $p \in (0, 1)$ (für $p = 1$ setzt man $Q_X(1) = \infty$, falls $F_X(r) < 1$ für jedes endliche $r \in \mathbb{R}$ gilt).

Quantilfunktionen haben allgemein folgende Eigenschaften (Übungsaufgabe):

- a) Sie sind monoton steigend in p .
- b) Sie sind linksseitig stetig.
- c) An allen Stellen gilt: $F_X(Q_X(p)) \geq p$ und $Q_X(F_X(r)) \leq r$.
- d) An allen Stellen $r = Q_X(p)$, wo $F_X(r)$ beidseitig stetig ist, gilt $F_X(Q_X(p)) = p$.
- e) An allen Stellen, wo $F_X(r)$ streng monoton ist, gilt $Q_X(F_X(r)) = r$.

In diesem Sinne ist Q_X die Inverse von F_X .

Bemerkung 2.17: Im Computeralgebrasystem MuPAD existiert seit Version 2.5 eine umfangreiche Bibliothek `stats`, in der etliche Standardverteilungen mit ihren (kumulativen) Verteilungsfunktionen `stats::nameCDF` und den Quantilen `stats::nameQuantile` installiert sind. Weiterhin gibt es die (kontinuierlichen) Dichten `stats::namePDF` bzw. die diskreten W'keitswerte `stats::namePF` sowie Zufallsgeneratoren `stats::nameRandom`.

Zahlenwerte der Binomialverteilung (Beispiel 1.64 auf Seite 36 sowie Seite 71) mit den Parametern n („Anzahl der Wiederholungen“) und p („Erfolgsw'keit“) sind beispielsweise folgendermaßen zu berechnen:

```

>> n:= 10: p:= 1/2:
>> F:= stats::binomialCDF(n, p):
>> F(0), F(1), F(2.5), F(3)

1/1024, 11/1024, 0.0546875, 11/64

>> Q:= stats::binomialQuantile(n, p):
>> Q(0), Q(1/4), Q(0.3), Q(9/10), Q(1)

0, 4, 4, 7, 10

```

Weiterhin stehen z.B. die hypergeometrische Verteilung (Beispiel 1.10 auf Seite 7 und Seite 75) als

```

stats::hypergeometricCDF, stats::hypergeometricQuantile,
stats::hypergeometricPF, stats::hypergeometricRandom
oder die später ab Seite 78 diskutierte Normalverteilung als
stats::normalCDF, stats::normalQuantile,
stats::normalPDF, stats::normalRandom

```

etc. zur Verfügung.

2.2 Das Riemann-Stieltjes-Integral

Wir verallgemeinern den bekannten Begriff der Riemann-Integration, um kontinuierliche Zufallsvariablen (mit Dichten) und diskrete Zufallsvariablen (ohne Dichten, die Verteilungsfunktion ist eine Treppenfunktion) gemeinsam behandeln zu können:

Definition 2.18:

Sei $F : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ eine monoton wachsende, rechtsseitig stetige, beschränkte Funktion. Zerlege das Intervall $[a, b]$ in $n + 1$ Zwischenpunkte $r_i = a + i \frac{b-a}{n}$, $i = 0, \dots, n$, und definiere für eine Funktion $f : [a, b] \rightarrow \mathbb{R}$ die **Riemann-Stieltjes-Summe** (bzgl. F)

$$R_n(f) = \sum_{i=0}^{n-1} f(r_i) \left(F(r_{i+1}) - F(r_i) \right).$$

Wenn der Grenzwert existiert, definiert man als **Riemann-Stieltjes-Integral** (bzgl. F)

$$\int_{(a,b]} f(r) dF(r) = \lim_{n \rightarrow \infty} R_n(f).$$

In Analogie zur Riemann-Integration kann man sich unschwer überlegen, dass zumindestens für (stückweise) stetiges f dieses Integral existiert.

Bemerkung 2.19: Für $F(r) = r$ ergibt sich das übliche Riemann-Integral.

Bemerkung 2.20: Wie beim üblichen Riemann-Integral definiert man uneigentliche Stieltjes-Integrale (über unendlichen Intervallen) über

$$\int_{(-\infty, b]} f(r) dF(r) = \lim_{a \rightarrow -\infty} \int_{(a, b]} f(r) dF(r) ,$$

$$\int_{\mathbb{R}} f(r) dF(r) = \lim_{b \rightarrow \infty} \int_{(-\infty, b]} f(r) dF(r) .$$

Satz 2.21: (einige Eigenschaften von Stieltjes-Integralen)

Es gilt:

1)

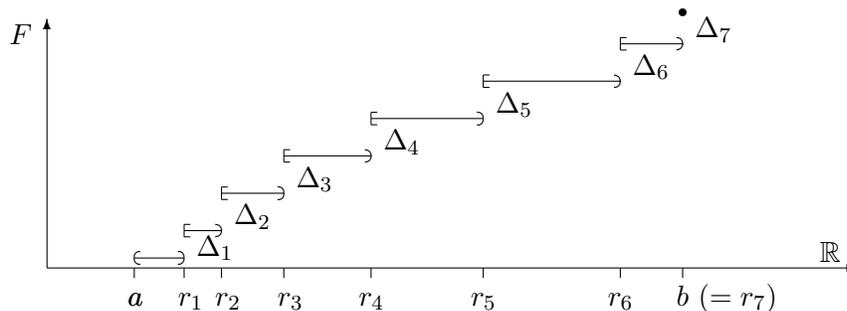
$$\int_{(a, b]} 1 dF(r) = F(b) - F(a).$$

2) Falls F stetig differenzierbar ist:

$$\int_{(a, b]} f(r) dF(r) = \int_a^b f(r) F'(r) dr .$$

3) Ist F eine Treppenfunktion mit Sprungstellen $r_1, r_2, \dots \in (a, b]$ und Sprunghöhen

$$\Delta_i = F(r_i) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} F(r_i - \epsilon) ,$$



so gilt für stetiges f :

$$\int_{(a, b]} f(r) dF(r) = \sum_i f(r_i) \Delta_i .$$

4) Ist f stetig differenzierbar, so gilt („partielle Integration“):

$$\int_{(a,b]} f(r) dF(r) = f(b) F(b) - f(a) F(a) - \int_a^b f'(r) F(r) dr .$$

Beweisskizze:

↓21.5.07

i) Für $f(r) \equiv 1$ sind alle Stieltjes-Summen „Teleskopsummen“:

$$R_n(f) = \sum_{i=0}^{n-1} \left(F(r_{i+1}) - F(r_i) \right) = F(r_n) - F(r_0) = F(b) - F(a) .$$

2) Für stetig differenzierbares F gilt der Mittelwertsatz $F(r_{i+1}) - F(r_i) = F'(\xi_i) (r_{i+1} - r_i)$ mit einem Zwischenwert $\xi_i \in (r_i, r_{i+1})$. Die Stieltjes-Summe wird damit zu einer Riemann-Summe für den Integranden $f(r) F'(r)$:

$$R_n(f) = \sum_{i=0}^{n-1} f(r_i) \left(F(r_{i+1}) - F(r_i) \right) = \sum_{i=0}^{n-1} f(r_i) F'(\xi_i) (r_{i+1} - r_i) .$$

3) Analog zu Riemann-Integralen kann man sich leicht überlegen, dass für Stieltjes-Integrale

$$\int_{(a,b]} f(r) dF(r) = \int_{(a,c]} f(r) dF(r) + \int_{(c,b]} f(r) dF(r)$$

mit beliebigem $c \in (a, b)$ gilt. Seien $r_1 < r_2 < \dots$ die Sprungstellen von F . Mit der Setzung $r_0 = a$ folgt

$$\int_{(a,b]} f(r) dF(r) = \sum_i \int_{(r_{i-1}, r_i]} f(r) dF(r) .$$

Da F auf $(r_{i-1}, r_i]$ konstant ist bis auf den Sprung am rechten Intervallende, besteht eine Stieltjes-Summe für $\int_{(r_{i-1}, r_i]} f(r) dF(r)$ mit den Stützstellen $r_i^{(j)} = r_{i-1} + j(r_i - r_{i-1})/n$ nur aus einem einzigen Term:

$$\begin{aligned} R_n(f) &= \sum_{j=0}^{n-1} f(r_i^{(j)}) \left(F(r_i^{(j+1)}) - F(r_i^{(j)}) \right) = f(r_i^{(n-1)}) \left(\underbrace{F(r_i^{(n)})}_{F(r_i)} - \underbrace{F(r_i^{(n-1)})}_{F(r_{i-1})} \right) \\ &= f(r_i^{(n-1)}) \Delta_i . \end{aligned}$$

Mit $n \rightarrow \infty$, $r_i^{(n-1)} = r_i - (b-a)/n \rightarrow r_i$ und der Stetigkeit von f folgt

$$\int_{(r_{i-1}, r_i]} f(r) dF(r) = f(r_i) \Delta_i .$$

4) Durch Umsummation ergibt sich die Darstellung

$$\begin{aligned} R_n(f) &= \sum_{i=0}^{n-1} f(r_i) \left(F(r_{i+1}) - F(r_i) \right) \\ &= f(r_n) F(r_n) - f(r_0) F(r_0) - \sum_{i=1}^n \left(f(r_i) - f(r_{i-1}) \right) F(r_i), \end{aligned}$$

wobei $f(r_n) F(r_n) = f(b) F(b)$, $f(r_0) F(r_0) = f(a) F(a)$. Analog zu 2) lässt sich die verbleibende Summe als übliche Riemann-Summe für den Integranden $f'(r) F(r)$ interpretieren, die gegen das übliche Riemann-Integral $\int_a^b F(r) f'(r) dr$ konvergiert.

Q.E.D.

2.3 Erwartungswert und Streuung

Der Zusammenfassung 2.14 folgend können wir uns nun mit Hilfe des Stieltjes-Integrals eine Zufallsvariable X mit Verteilung F_X als ein Modell $(\mathbb{R}, \mathcal{E}, P)$ vorstellen, wo die Ereignisse („Borel-Mengen“) aus Intervallen zusammengesetzt sind und das W'keitsmaß P für ein Ereignis $E \subset \mathbb{R}$ durch

$$P(X \in E) = \int_{r \in E} dF_X(r)$$

definiert ist und hierdurch über die kumulative Verteilungsfunktion $F_X(r)$ kodiert ist. Speziell gilt nämlich für $E = (r_1, r_2]$ nach Satz 2.21.1) $P(r_1 < X \leq r_2) = F_X(r_2) - F_X(r_1)$, was nach Folgerung 2.8 genau der Definition von W'keiten für Zufallsvariable entspricht:

- Für eine diskrete Zufallsvariable $X(\Omega) = \{r_1, r_2, \dots\}$ ergibt sich das W'maß

$$P(X \in E) = \sum_{r_i \in E} P(X = r_i)$$

gemäß der Definition 1.14 diskreter Modelle, sobald man $P(X = r_i)$ kennt.

- Für eine kontinuierliche Verteilungsfunktion mit einer Dichte $\rho(r) = F_X'(r)$ ergibt sich die Vorgabe von W'keiten

$$P(X \in (r_1, r_2]) = \int_{r_1}^{r_2} F_X'(r) dr = \int_{r_1}^{r_2} \rho(r) dr$$

als Integrale über Dichten gemäß der Definition 1.18 kontinuierlicher Modelle.

Was bringt uns nun die Tatsache, dass wir es hier mit Zufallswerten in \mathbb{R} zu tun haben? Wir können Arithmetik betreiben! Speziell heißt dies, dass man Erwartungswerte definieren kann.

Die Stieltjes-Integrale fassen diskrete und kontinuierliche Modelle zusammen. Wir können damit für beliebige Modelle einheitlich definieren:

Definition 2.22: (Erwartungswert)

Sei $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F_X . Dann nennt man

$$E(X) = \int_{-\infty}^{\infty} r \, dF_X(r)$$

den **Erwartungswert** von X .

Sprechweise 2.23:

↓23.5.07

Fassen wir X als Zufallsexperiment $(\mathbb{R}, \mathcal{E}, P)$ auf, so nimmt X zufällige Werte in \mathbb{R} an. Umgangssprachlich wird das Konzept des Erwartungswerts so formuliert:

„ X nimmt im Mittel den Wert $E(X)$ an.“

Im Rahmen dieser Vorlesung sind nur die folgenden Spezialfälle interessant (d.h., man braucht Stieltjes-Integrale nicht wirklich auszuwerten, sondern kann entweder einfach summieren oder ein übliches Riemann-Integral berechnen):

Spezialfälle 2.24:

- a) Ist X diskret, also ist die Bildmenge $X(\Omega) = \{r_1, r_2, \dots\}$ abzählbar (F_X ist eine Treppenfunktion), so gilt mit den Sätzen 2.21.3) und 2.9.5):

$$E(X) = \sum_{r \in X(\Omega)} r \cdot P(X = r) .$$

- b) Ist X kontinuierlich mit der Dichte $\rho(r) = \frac{d}{dr} F_X(r)$, so gilt mit Satz 2.21.2):

$$E(X) = \int_{-\infty}^{\infty} r \cdot \rho(r) \, dr .$$

Bemerkung 2.25: Nach (Gegen-)Beispiel 2.5 muss für eine diskrete Zufallsvariable das unterliegende Modell Ω nicht notwendigerweise diskret sein. Umgekehrt gilt jedoch, dass jede Zufallsvariable über einem diskreten $\Omega = \{\omega_1, \omega_2, \dots\}$ diskret ist. In diesem Fall ist die obige Definition äquivalent zu

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) .$$

Dies ist eine alternative Berechnungsformel für den Erwartungswert, die sich anbietet, wenn man das unterliegende Modell Ω kennt und die Verteilungsdaten $P(X = r)$ nicht explizit ausgerechnet hat.

Beweisskizze: Siehe Übungsaufgabe 45.

Beispiel 2.26: (diskret) Wurf mit zwei fairen Würfeln:

$$\Omega = \{(\omega_1, \omega_2); \omega_1, \omega_2 \in \{1, \dots, 6\}\}.$$

Sei X die Augensumme $X : (\omega_1, \omega_2) \mapsto \omega_1 + \omega_2$. Nach Definition 2.22 braucht man zur Berechnung des Erwartungswerts die (nichtkombinatorischen) W'keiten $P(X = r)$ mit $r = 2, 3, \dots, 12$. Diese waren in Beispiel 2.3 angegeben worden:

$$\begin{aligned} E(X) &= \sum_{r \in X(\Omega)} r P(X = r) = \sum_{r=2}^{12} r P(X = r) \\ &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + \dots + 12 \cdot \frac{1}{36} = 7. \end{aligned}$$

Nach Bemerkung 2.25 kann der Erwartungswert auch direkt durch Summation über das unterliegende kombinatorische Modell berechnet werden:

$$\begin{aligned} E(X) &= \sum_{\omega_1=1}^6 \sum_{\omega_2=1}^6 (\omega_1 + \omega_2) \underbrace{P(\{(\omega_1, \omega_2)\})}_{1/36} \\ &= \frac{1}{36} \left(\left(\sum_{\omega_1=1}^6 \omega_1 \right) \cdot \left(\sum_{\omega_2=1}^6 1 \right) + \left(\sum_{\omega_2=1}^6 1 \right) \cdot \left(\sum_{\omega_1=1}^6 \omega_1 \right) \right) \\ &= \frac{1}{36} \cdot 2 \cdot 6 \cdot \sum_{\omega=1}^6 \omega = \frac{1}{36} \cdot 2 \cdot 6 \cdot \frac{6 \cdot 7}{2} = 7. \end{aligned}$$

Beispiel 2.27: (kontinuierlich) Ein zum Zeitpunkt $t_0 = 0$ existierendes radioaktives Atom zerfällt mit der W'keit $\lambda e^{-\lambda t} dt$ im Zeitintervall $(t, t + dt)$, $t > 0$. D.h., die Dichte der Zerfallsw'keit ist

$$\rho(t) = \begin{cases} 0 & \text{für } t < 0, \\ \lambda e^{-\lambda t} & \text{für } t \geq 0. \end{cases}$$

Sei $T \in [0, \infty]$ der Zerfallszeitpunkt. Die „mittlere Lebensdauer“ ist der Erwartungswert

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} t \rho(t) dt = \int_0^{\infty} \underbrace{t}_u \underbrace{\lambda e^{-\lambda t}}_{v'} dt = \\ &= \left[\underbrace{t}_u \cdot \underbrace{(-e^{-\lambda t})}_v \right]_{t=0}^{t=\infty} - \int_0^{\infty} \underbrace{1}_{u'} \underbrace{(-e^{-\lambda t})}_v dt = \int_0^{\infty} e^{-\lambda t} dt = \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_{t=0}^{t=\infty} = \frac{1}{\lambda}. \end{aligned}$$

Bemerkung 2.28: (Warnung) Nicht jede Zufallsvariable hat einen Erwartungswert. Betrachte z.B. $X : \Omega \mapsto [0, \infty)$ mit der Dichte $\rho(r) = 0$ für $r < 0$ und $\rho(r) = \frac{2}{\pi} \frac{1}{1+r^2}$ für $r \geq 0$. Das Integral

$$E(X) = \int_0^\infty \frac{2}{\pi} \frac{r}{1+r^2} dr = \lim_{R \rightarrow \infty} \int_0^R \frac{2}{\pi} \frac{r}{1+r^2} dr = \lim_{R \rightarrow \infty} \frac{1}{\pi} \ln(1+R^2) = \infty$$

existiert nicht.

Wir betrachten nun eine (glatte) Funktion $f : \mathbb{R} \mapsto \mathbb{R}$. Für eine gegebene Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ kann man

$$\tilde{X} = f(X) : \omega \in \Omega \mapsto f(X(\omega)) \in \mathbb{R}$$

als neue Zufallsvariable auffassen, zu der sich eine neue Verteilungsfunktion $F_{\tilde{X}}(r) = P(\tilde{X} \leq r) = P(f(X) \leq r) = P(X \in f^{-1}((-\infty, r]))$ bestimmen lässt, mit der dann der Erwartungswert $E(\tilde{X})$ berechnet werden kann. Aber: wir brauchen $F_{\tilde{X}}$ gar nicht! Man kann den Erwartungswert von \tilde{X} direkt mittels der Verteilung $F_X(r)$ bestimmen. Im folgenden Satz ist F_X im diskreten Fall durch die Sprunghöhen $P(X = r)$ bzw. im kontinuierlichen Fall durch die Ableitung $\rho(r) = F'_X(r)$ kodiert:

Satz 2.29: (Transformation von Zufallsvariablen)

Sei $f : \mathbb{R} \mapsto \mathbb{R}$ eine glatte Funktion.

a) Für eine diskrete Variable X mit $X(\Omega) = \{r_1, r_2, \dots\}$ gilt

$$E(f(X)) = \sum_{r \in X(\Omega)} f(r) P(X = r).$$

Ist das unterliegende Experiment $\Omega = \{\omega_1, \omega_2, \dots\}$ selbst schon diskret, so gilt mit Bemerkung 2.25 auch die alternative Formel:

$$E(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) P(\{\omega\}).$$

b) Für eine kontinuierliche Variable X mit der Dichte $\rho(r) = F'_X(r)$ gilt

$$E(f(X)) = \int_{r \in X(\Omega)} f(r) \rho(r) dr.$$

Beweisskizze:

Sei \tilde{X} die Zufallsvariable $f(X)$. Im diskreten Fall gilt mit $X(\Omega) = \{r_1, r_2, \dots\}$ für $\tilde{r} \in \tilde{X}(\Omega) = \{f(r_1), f(r_2), \dots\}$:

$$P(\tilde{X} = \tilde{r}) = P(X \in f^{-1}(\{\tilde{r}\})) = \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} P(X = r)$$

und damit

$$\begin{aligned} E(\tilde{X}) &= \sum_{\tilde{r} \in \tilde{X}(\Omega)} \tilde{r} P(\tilde{X} = \tilde{r}) = \sum_{\tilde{r} \in \tilde{X}(\Omega)} \tilde{r} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} P(X = r) \\ &= \sum_{\tilde{r} \in \tilde{X}(\Omega)} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} \tilde{r} P(X = r) = \sum_{\tilde{r} \in \tilde{X}(\Omega)} \sum_{\substack{r \in X(\Omega) \\ f(r) = \tilde{r}}} f(r) P(X = r) \\ &= \sum_{r \in X(\Omega)} f(r) P(X = r) = E(f(X)) . \end{aligned}$$

Im kontinuierlichen Fall führen wir den Beweis nur für *streng monoton wachsendes* f (der allgemeine Fall ist mit den hier zur Verfügung stehenden Mitteln technisch zu aufwendig):

$$\begin{aligned} E(\tilde{X}) &= \int_{-\infty}^{\infty} \tilde{r} dF_{\tilde{X}}(\tilde{r}) = \int_{-\infty}^{\infty} \tilde{r} \frac{dF_{\tilde{X}}(\tilde{r})}{d\tilde{r}} d\tilde{r} \\ &\stackrel{\text{(Substitution: } \tilde{r}=f(r))}{=} \int_{-\infty}^{\infty} f(r) \frac{dF_{\tilde{X}}(f(r))}{d\tilde{r}} f'(r) dr \\ &\stackrel{\text{(Kettenregel)}}{=} \int_{-\infty}^{\infty} f(r) \frac{dF_{\tilde{X}}(f(r))}{dr} dr . \end{aligned}$$

Mit

$$F_{\tilde{X}}(f(r)) = P(\tilde{X} \leq f(r)) = P(f(X) \leq f(r)) = P(X \leq r) = F_X(r)$$

folgt

$$E(\tilde{X}) = \int_{-\infty}^{\infty} f(r) \frac{dF_X(r)}{dr} dr = \int_{-\infty}^{\infty} f(r) dF_X(r) = E(f(X)) .$$

Q.E.D.

30.5.07↓

Beispiel 2.30: Betrachte einen Zufallszahlengenerator X , der (Gleitpunkt-)Zahlen zwischen -1 und 1 auswirft. Er sei gleichverteilt, d.h., für $-1 \leq r_1 \leq r_2 \leq 1$ gelte

$$P(r_1 < X \leq r_2) = \int_{r_1}^{r_2} \rho(r) dr = \frac{r_2 - r_1}{2}$$

mit der Dichte

$$\rho(r) = \begin{cases} \frac{1}{2} & \text{für } r \in [-1, 1], \\ 0 & \text{für } r \notin [-1, 1]. \end{cases}$$

Die Verteilungsfunktion ist

$$F_X(r) = \int_{-\infty}^r \rho(t) dt = \begin{cases} 0 & \text{für } r < -1, \\ \frac{r+1}{2} & \text{für } -1 \leq r \leq 1, \\ 1 & \text{für } r > 1. \end{cases}$$

Der Erwartungswert ist 0:

$$E(X) = \int_{-\infty}^{\infty} r \rho(r) dr = \int_{-1}^1 \frac{r}{2} dr = 0.$$

Die Variable $\tilde{X} = X^2$ hat folgende Verteilung (für $\tilde{r} \geq 0$):

$$F_{\tilde{X}}(\tilde{r}) = P(X^2 \leq \tilde{r}) = P(-\sqrt{\tilde{r}} \leq X \leq \sqrt{\tilde{r}}) = P(X \leq \sqrt{\tilde{r}}) - P(X < -\sqrt{\tilde{r}}).$$

Da $\tilde{X} = X^2$ keine Werte $\tilde{r} < 0$ annehmen kann, ergibt sich insgesamt:

$$F_{\tilde{X}}(\tilde{r}) = F_X(\sqrt{\tilde{r}}) - F_X(-\sqrt{\tilde{r}}) = \begin{cases} 0 & \text{für } \tilde{r} < 0, \\ \sqrt{\tilde{r}} & \text{für } 0 \leq \tilde{r} \leq 1, \\ 1 & \text{für } \tilde{r} > 1. \end{cases}$$

Durch Ableiten erhalten wir die Dichtefunktion $\tilde{\rho}(\tilde{r})$ für \tilde{X} , mit der sich dann der Erwartungswert ergibt:

$$\tilde{\rho}(\tilde{r}) = \frac{d}{d\tilde{r}} F_{\tilde{X}}(\tilde{r}) = \begin{cases} \frac{1}{2} \frac{1}{\sqrt{\tilde{r}}} & \text{für } \tilde{r} \in [0, 1], \\ 0 & \text{für } \tilde{r} \notin [0, 1] \end{cases}$$

$$\Rightarrow E(\tilde{X}) = \int_{-\infty}^{\infty} \tilde{r} \tilde{\rho}(\tilde{r}) d\tilde{r} = \int_0^1 \tilde{r} \frac{1}{2\sqrt{\tilde{r}}} d\tilde{r} = \int_0^1 \frac{\sqrt{\tilde{r}}}{2} d\tilde{r} = \frac{1}{3}.$$

Das mühselige Bestimmen der Verteilung von \tilde{X} war völlig unnötig. Nach Satz 2.29 geht's einfacher direkt über die Verteilung von X :

$$E(X^2) = \int_{-\infty}^{\infty} r^2 \rho(r) dr = \int_{-1}^1 \frac{r^2}{2} dr = \frac{1}{3}.$$

Bemerkung 2.31: Mit Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$, $Y : \Omega \rightarrow \mathbb{R}$ kann man Arithmetik betreiben. Man kann z.B. Zufallsvariablen addieren oder multiplizieren, d.h., man betrachtet die neuen Zufallsvariablen $Z = X + Y : \Omega \mapsto \mathbb{R}$ oder $Z = X \cdot Y : \Omega \mapsto \mathbb{R}$, die punktweise durch

$$\begin{aligned} X + Y & : \omega \in \Omega \mapsto X(\omega) + Y(\omega), \\ X \cdot Y & : \omega \in \Omega \mapsto X(\omega) \cdot Y(\omega) \end{aligned}$$

definiert sind. Es gilt die Faustregel:

Die Verteilung einer Funktion einer einzigen Zufallsvariablen X (z.B., $Z = X^2$) ist eindeutig durch die Verteilung F_X bestimmt.

Die Verteilung einer Funktion mehrerer Zufallsvariablen X, Y (z.B., $Z = X + Y$ oder $Z = X \cdot Y$) lässt sich i.A. nicht allein aus den Verteilungen F_X, F_Y bestimmen (außer, wenn die Variablen X, Y unabhängig sind, siehe Abschnitt 2.5).

Beispiel 2.32: Betrachte den zweifachen Wurf eines fairen Würfels

$$\Omega = \{(\omega_1, \omega_2); \omega_1, \omega_2 \in \{1, \dots, 6\}\}.$$

Sei

$$\begin{aligned} X_1 &= \text{Ergebnis des ersten Wurfs} : (\omega_1, \omega_2) \rightarrow \omega_1, \\ X_2 &= \text{Ergebnis des zweiten Wurfs} : (\omega_1, \omega_2) \rightarrow \omega_2, \\ X &= X_1 + X_2 = \text{Augensumme} : (\omega_1, \omega_2) \rightarrow \omega_1 + \omega_2. \end{aligned}$$

Offensichtlich sind X_1, X_2 gleichverteilt:

$$P(X_1 = k_1) = P(X_2 = k_2) = \frac{1}{6} \quad \text{für alle } k_1, k_2 \in \{1, \dots, 6\}.$$

Die Verteilung der Augensumme X war in Beispiel 2.3 berechnet worden:

$$P(X = 2) = \frac{1}{36}, \quad P(X = 3) = \frac{2}{36}, \quad P(X = 4) = \frac{3}{36}, \quad \dots, \quad P(X = 12) = \frac{1}{36}.$$

Es besteht keine offensichtliche Möglichkeit, aus den Verteilungswerten $P(X_1 = k_1), P(X_2 = k_2)$ der Summanden direkt auf die Verteilung der Summe $X = X_1 + X_2$ zu schließen!

Trotzdem kann der Erwartungswert von X sehr leicht aus den Erwartungswerten der Summanden berechnet werden! Offensichtlich gilt

$$E(X_1) = E(X_2) = \frac{1 + 2 + \dots + 6}{6} = \frac{7}{2},$$

der Erwartungswert von $X = X_1 + X_2$ war in Beispiel 2.26 als $E(X) = 7$ bestimmt worden, also $E(X_1 + X_2) = E(X_1) + E(X_2)$. Zufall? Nein! Siehe Satz 2.33.1).

Die wohl Wichtigste der folgenden Aussagen ist die Linearität 1) von Erwartungswerten:

Satz 2.33: (Rechenregeln für Erwartungswerte)

Seien X, Y reelle Zufallsvariable über demselben (Ω, \mathcal{E}, P) , sei $\alpha \in \mathbb{R}$.

1) Der Erwartungswert ist linear:

$$E(\alpha X) = \alpha E(X), \quad E(X + Y) = E(X) + E(Y).$$

2) Sei $X = \alpha$ konstant. Dann gilt $E(X) = \alpha$.

3) Sei $X(\omega) \geq Y(\omega)$ für alle $\omega \in \Omega$ (kurz: $X \geq Y$). Dann gilt

$$E(X) \geq E(Y).$$

Speziell gilt $E(X) \geq 0$ für $X \geq 0$. Für beliebiges X gilt:

$$E(|X|) \geq |E(X)|.$$

4) Sei $A \subset \Omega$ ein Ereignis. Für die „Indikatorfunktion“

$$1_A : \omega \in \Omega \longrightarrow \begin{cases} 1 & \text{für } \omega \in A, \\ 0 & \text{für } \omega \in \Omega \setminus A \end{cases}$$

gilt $E(1_A) = P(A)$. Für unabhängige Ereignisse A, B gilt:

$$E(1_A 1_B) = E(1_{A \cap B}) = P(A \cap B) = P(A) P(B) = E(1_A) E(1_B).$$

5) Sei $X \geq 0$. Für jedes $\alpha \in \mathbb{R}$ gilt:

$$E(X) \geq \alpha \cdot P(X \geq \alpha).$$

Beweisskizze:

Für den Spezialfall eines diskreten Stichprobenraums $\Omega = \{\omega_1, \omega_2, \dots\}$ folgend die Eigenschaften 1) – 4) unmittelbar aus der Darstellung 2.24.a):

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Für den allgemeinen Fall:

2) Aus der Definition der Integrale über Stieltjes-Summen folgt sofort

$$E(X) = \int_{\mathbb{R}} \alpha dF_X(r) = \alpha \int_{\mathbb{R}} dF_X(r) = \alpha F_X(\infty) = \alpha.$$

4) $E(1_A) = \int_{\mathbb{R}} 1_A dF_X(r) = \int_A dF_X(r) = P(A)$.

1) Sei $\alpha > 0$. Eine $\int_{(a,b]} r dF_{\alpha X}(r)$ approximierende Stieltjes-Summe hat die Form

$$\begin{aligned} \sum_{i=0}^{n-1} r_i \left(F_{\alpha X}(r_{i+1}) - F_{\alpha X}(r_i) \right) &= \alpha \sum_{i=0}^{n-1} r_i P(r_i < \alpha X \leq r_{i+1}) . \\ &= \alpha \sum_{i=0}^{n-1} \frac{r_i}{\alpha} P\left(\frac{r_i}{\alpha} < X \leq \frac{r_{i+1}}{\alpha}\right) . \end{aligned}$$

Dies ist eine Stieltjes-Summe für $\alpha \int_{(a/\alpha, b/\alpha]} r dF_X(r)$, also gilt mit $n \rightarrow \infty$:

$$\int_{(a,b]} r dF_{\alpha X}(r) = \alpha \int_{(a/\alpha, b/\alpha]} r dF_X(r) .$$

Im Grenzwert $a \rightarrow -\infty, b \rightarrow -\infty$ folgt $E(\alpha X) = \alpha E(X)$.

Mit kleinen Variationen gelten dieselben Argumente auch für $\alpha < 0$. Für $\alpha = 0$ ist die Behauptung trivial.

Nun wird $E(X + Y) = E(X) + E(Y)$ gezeigt. Seien zunächst X und Y und damit auch $Z = X + Y$ diskret: $X(\Omega) = \{x_1, x_2, \dots\}$, $Y(\Omega) = \{y_1, y_2, \dots\}$,

$$Z(\Omega) = \{z_1, z_2, \dots\} = \{x + y; x \in X(\Omega); y \in Y(\Omega)\} .$$

Mit der Notation

$$\begin{aligned} P(Z = z_k) &= \sum_{\substack{x_i, y_j \\ x_i + y_j = z_k}} P(\{\omega \in \Omega; X(\omega) = x_i; Y(\omega) = y_j\}) \\ &\equiv \sum_{x_i} \sum_{y_j} P(x_i + y_j = z_k) \end{aligned}$$

folgt

$$\begin{aligned} E(Z) &= \sum_{z_k} z_k P(Z = z_k) = \sum_{z_k} \sum_{x_i} \sum_{y_j} (x_i + y_j) P(x_i + y_j = z_k) \\ &= \sum_{z_k} \sum_{x_i} \sum_{y_j} x_i P(x_i + y_j = z_k) + \sum_{z_k} \sum_{x_i} \sum_{y_j} y_j P(x_i + y_j = z_k) \\ &= \sum_{x_i} x_i \sum_{z_k} \sum_{y_j} P(x_i + y_j = z_k) + \sum_{y_j} y_j \sum_{z_k} \sum_{x_i} P(x_i + y_j = z_k) \\ &= \sum_{x_i} x_i P(X = x_i) + \sum_{y_j} y_j P(Y = y_j) = E(X) + E(Y) . \end{aligned}$$

Für den allgemeinen Fall deuten wir nur die Beweisidee an: zu gegebenem $n \in \mathbb{N}$ betrachte die Approximation von X durch die „Treppenfunktionen“

$$X_n = \sum_{k=-\infty}^{\infty} \frac{k}{n} 1_{A_{kn}}$$

mit der disjunkten Zerlegung

$$A_{kn} = \left\{ \omega \in \Omega; \frac{k}{n} < X(\omega) \leq \frac{k+1}{n} \right\}, \quad \bigcup_{k=-\infty}^{\infty} A_{kn} = \Omega.$$

Beachte: jedes $\omega \in \Omega$ liegt in genau einem der A_{kn} , d.h., die Summe $X_n(\omega) = \sum_k (k/n) 1_{A_{kn}}(\omega)$ besteht aus genau einem Term. X_n ist eine diskrete Variable, die nur Werte der Form „ganze Zahl“/n annimmt. Es gilt $X_n < X \leq X_n + 1/n$. Es folgt mit den entsprechenden Approximation Y_n für Y , Z_n für Z :

$$Z_n \leq Z = X + Y \leq X_n + Y_n + \frac{2}{n} \leq X + Y + \frac{2}{n} = Z + \frac{2}{n} \leq Z_n + \frac{3}{n},$$

also

$$-\frac{2}{n} \leq Z_n - X_n - Y_n \leq \frac{1}{n}.$$

Die Rechenregel 3) gilt sicherlich für diskrete Variable (siehe die Argumente unten für 3), die Linearität des Erwartungswerts ist für die diskreten Variablen X_n, Y_n, Z_n sowie die Konstanten gezeigt). Es folgt

$$-\frac{2}{n} \leq E(Z_n - X_n - Y_n) = E(Z_n) - E(X_n) - E(Y_n) \leq \frac{1}{n}. \quad (\#)$$

Mit wachsendem n folgt $E(X_n) \rightarrow E(X)$, $E(Y_n) \rightarrow E(Y)$, $E(Z_n) \rightarrow E(Z)$ (diesen wichtigen, aber sehr technischen Schritt unterschlagen wir hier). Für die Grenzwerte folgt mit (#) dann $0 \leq E(Z) - E(X) - E(Y) \leq 0$, also $E(Z) = E(X) + E(Y)$.

3) Sei $Z = X - Y \geq 0$. Alle den Erwartungswert $E(Z)$ approximierenden endlichen Stieltjes-Summen sind ≥ 0 , damit gilt auch im Grenzwert $E(Z) \geq 0$. Mit der Linearität 1) folgt $E(X) - E(Y) \geq 0$. Mit $|X| \geq \pm X$ folgt $E(|X|) \geq \pm E(X)$.

5) Setze $A_\alpha = \{\omega \in \Omega; X(\omega) \geq \alpha\}$. Für $X \geq 0$ gilt dann $X \geq \alpha 1_{A_\alpha}$. Mit 1) bis 4) folgt sofort die Behauptung 5).

Q.E.D.

Beispiel 2.34: Gegeben sei eine Variable X mit der Verteilungsfunktion F_X . Betrachte die neue Zufallsvariable $Y = \alpha X + \beta$ mit konstanten $\alpha, \beta \in \mathbb{R}$. Der Erwartungswert von Y ist

$$E(Y) = E(\alpha X + \beta) = E(\alpha X) + E(\beta \cdot 1) = \alpha \cdot E(X) + \beta.$$

In diesem einfachen Fall kann man sogar die Verteilung von Y explizit über die Verteilung von X ausdrücken. Für $\alpha > 0$ gilt

$$F_Y(r) = P(Y \leq r) = P(\alpha X + \beta \leq r) = P\left(X \leq \frac{r - \beta}{\alpha}\right) = F_X\left(\frac{r - \beta}{\alpha}\right).$$

Für $\alpha < 0$ gilt:

$$\begin{aligned} F_Y(r) &= P(Y \leq r) = P(\alpha X + \beta \leq r) = P\left(X \geq \frac{r - \beta}{\alpha}\right) = 1 - P\left(X < \frac{r - \beta}{\alpha}\right) \\ &= 1 + P\left(X = \frac{r - \beta}{\alpha}\right) - P\left(X \leq \frac{r - \beta}{\alpha}\right) = 1 + P\left(X = \frac{r - \beta}{\alpha}\right) - F_X\left(\frac{r - \beta}{\alpha}\right). \end{aligned}$$

Für die Berechnung von $E(Y)$ wird dies aber nicht benötigt: wegen der Linearität weiß man unmittelbar, dass $E(Y) = \alpha \cdot E(X) + \beta$ gilt.

Der folgende Begriff „Varianz“ ist genauso fundamental wie der Begriff „Erwartungswert“:

Definition 2.35: (Varianz und Streuung)

Die **Varianz** einer Zufallsvariable X ist

$$\text{Var}(X) = E\left((X - E(X))^2\right) \stackrel{(*)}{=} E(X^2) - E(X)^2 \geq 0.$$

Die **Streuung** oder auch **Standardabweichung** von X ist

$$\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{E\left((X - E(X))^2\right)}.$$

Hierbei gilt die Identität (*) wegen der Linearität des Erwartungswerts:

$$\begin{aligned} E\left((X - E(X))^2\right) &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2. \end{aligned}$$

Die Interpretation der Varianz wird durch den folgenden Satz geliefert, dessen Stärke seine Allgemeinheit ist (keinerlei Voraussetzungen über X). In praktischen Anwendungen ist die Abschätzung allerdings recht grob:

Satz 2.36: (Chebyshevsche Ungleichung)

Sei X eine beliebige Zufallsvariable über einem beliebigen Modell (Ω, \mathcal{E}, P) mit Erwartungswert $E(X)$ und Varianz $\text{Var}(X)$. Dann gilt für jedes $\epsilon > 0$:

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

Äquivalenterweise gilt für die Komplementaussage:

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{\text{Var}(X)}{\epsilon^2} .$$

Beweis: Wende Satz 2.33.5) an auf $Y = (X - E(X))^2$ mit $\alpha = \epsilon^2$. Es ergibt sich sofort

$$\begin{aligned} \text{Var}(X) = E(Y) &\geq \alpha P(Y \geq \alpha) = \epsilon^2 P((X - E(X))^2 \geq \epsilon^2) \\ &= \epsilon^2 P(|X - E(X)| \geq \epsilon) . \end{aligned}$$

Q.E.D.

Bemerkung 2.37: Für $\epsilon \leq \sigma(X)$ ergibt sich keinerlei Information, da dann $\text{Var}(X)/\epsilon^2 \geq 1$ gilt und W'keiten trivialerweise immer zwischen 0 und 1 liegen. Wählt man $\epsilon = n \sigma(X)$ als kleines Vielfaches von $\sigma(X)$, so ergibt sich mit

$$P\left(E(X) - n \sigma(X) < X < E(X) + n \sigma(X)\right) \geq 1 - \frac{1}{n^2}$$

die Interpretation:

Mit großer W'keit nimmt eine Zufallsvariable Werte an, die höchstens um ein kleines Vielfaches von σ vom Erwartungswert abweichen.

Beispielsweise gilt für $\epsilon = 3 \sigma(X)$, dass X mit mindestens $8/9 \hat{=} 88.9\%$ -iger W'keit Werte innerhalb des Intervalls $(E(X) - 3 \sigma(X), E(X) + 3 \sigma(X))$ annimmt.

Merke: Die Standardabweichung σ gibt die Größenordnung an, um welche die Zufallswerte vom Erwartungswert abweichen („streuen“). Für kleines σ werden nur Werte dicht beim Erwartungswert beobachtet, für großes σ kommen mit großer W'keit auch Werte „weit weg“ vom Erwartungswert vor.

Beispiel 2.38: Betrachte die Zufallsvariable $X : \Omega \mapsto \{0, 1, \dots, n\} \subset \mathbb{R}$ mit

↓4.6.07

$$P(X = k) = \binom{n}{k} p^k q^{n-k},$$

wobei $p \in [0, 1]$ und $q = 1 - p$. Nach Satz 1.63 beschreibt X die Anzahl der Erfolge bei n -facher Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p . Es gilt

$$\begin{aligned} E(X) &= \sum_{k=0}^n k P(X = k) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = p \sum_{k=0}^n \binom{n}{k} \underbrace{k p^{k-1}}_{d p^k / d p} q^{n-k} \\ &= p \frac{d}{d p} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \stackrel{\text{(Binomi)}}{=} p \frac{d}{d p} (p + q)^n = p n (p + q)^{n-1} \stackrel{(p+q=1)}{=} p n. \end{aligned}$$

Dies ist intuitiv: im Mittel f'uhrt der Bruchteil p aller n Versuche zum Erfolg, also sollte der Erwartungswert f'ur die Gesamtanzahl der Erfolge $p n$ sein. Wie stark streuen die Erfolge um diesen Mittelwert?

$$\text{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - (p n)^2.$$

Es gilt

$$\begin{aligned} E(X^2) - E(X) &= E(X^2 - X) = \sum_{k=0}^n (k^2 - k) P(X = k) = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k} \\ &= p^2 \sum_{k=0}^n \binom{n}{k} \underbrace{k(k-1) p^{k-2}}_{d^2 p^k / d p^2} q^{n-k} = p^2 \frac{d^2}{d p^2} \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \\ &\stackrel{\text{(Binomi)}}{=} p^2 \frac{d^2}{d p^2} (p + q)^n = p^2 n(n-1) (p + q)^{n-2} \stackrel{(p+q=1)}{=} p^2 n(n-1). \end{aligned}$$

Es folgt

$$E(X^2) = p^2 n(n-1) + E(X) = p^2 n(n-1) + p n$$

und damit

$$\text{Var}(X) = E(X^2) - (p n)^2 = p^2 (n^2 - n) + p n - p^2 n^2 = p(1-p)n = p q n.$$

Beispiel 2.39: Frage: mit welcher W'keit liegt bei 1000 W'rfen einer fairen M'unze die Anzahl der „K'opfe“ zwischen 400 und 600?

Antwort: dies ist eine $n = 1000$ -fache Wiederholung des Bernoulli-Experiments „M'unzwurf“ mit $p = P(K) = q = P(Z) = 1/2$. Nach Beispiel 2.38 hat die „Anzahl der K'opfe“ X den Erwartungswert $E(X) = p n = 500$, die Varianz ist $\text{Var}(X) = p q n = n/4 = 250$, die Streuung ist $\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{250} \approx 15.8$. Nach Chebyshev gilt:

$$P(400 \leq X \leq 600) = P(|X - 500| \leq 100) = P(|X - E(X)| < 101)$$

$$\geq 1 - \frac{\text{Var}(X)}{101^2} = 1 - \frac{250}{101^2} \approx 0.975 .$$

Für die W'keit, dass X zwischen 480 und 520 liegt, folgt analog

$$\begin{aligned} P(480 \leq X \leq 520) &= P(|X - 500| \leq 20) = P(|X - E(X)| < 21) \\ &\geq 1 - \frac{\text{Var}(X)}{21^2} = 1 - \frac{250}{21^2} \approx 0.433 . \end{aligned}$$

Für die W'keit, dass X zwischen 490 und 510 liegt, folgt analog

$$\begin{aligned} P(490 \leq X \leq 510) &= P(|X - 500| \leq 10) = P(|X - E(X)| < 11) \\ &\geq 1 - \frac{\text{Var}(X)}{11^2} = 1 - \frac{250}{11^2} \approx -1.07 \end{aligned}$$

(hier hat Chebyshev keine Aussagekraft mehr). Die exakte Formel

$$P(a \leq X \leq b) = \sum_{k=a}^b \binom{n}{k} p^k q^{n-k}$$

liefert

$$\begin{aligned} P(400 \leq X \leq 600) &\approx 0.99999999820 , \\ P(480 \leq X \leq 520) &\approx 0.805 , \\ P(490 \leq X \leq 510) &\approx 0.493 . \end{aligned}$$

Man sieht an diesen Werten: die von Chebyshev gelieferten Abschätzungen für die W'keiten sind zwar korrekt, aber arg grob.

2.4 Einige Standardverteilungen

Hier werden einige wichtige Verteilungen mit den für Anwendungen wichtigen Interpretationen vorgestellt.

2.4.1 Die Binomial-Verteilung

Zu gegebenem $n \in \mathbb{N}$ und $p \in [0, 1]$, $q = 1 - p$, betrachte $X : \Omega \rightarrow \{0, 1, \dots, n\}$ mit

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n .$$

Nach Beispiel 2.38 gilt

$$E(X) = pn, \quad \text{Var}(X) = pqn, \quad \sigma(X) = \sqrt{pqn} .$$

Kürzel: Man sagt, X ist „Bi(n, p)-verteilt“.

MuPAD-Routinen: `stats::binomialCDF`, `stats::binomialQuantile`,
`stats::binomialPF`, `stats::binomialRandom`.

Interpretation: Nach Satz 1.63 beschreibt X die Anzahl der Erfolge in einer n -fachen Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p .

Beispiel: Anzahl der „Köpfe“ bei n -fachem Münzwurf (auch mit einer unfairen Münze), $p = P(\text{„Kopf“})$.

6.6.07↓

2.4.2 Die Poisson-Verteilung

Zu gegebenem $\lambda > 0$ betrachte $X : \Omega \rightarrow \{0, 1, 2, \dots\}$ mit

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}; \quad k = 0, 1, 2, \dots$$

Man rechnet leicht nach:

$$E(X) = \lambda; \quad \text{Var}(X) = \lambda; \quad \sigma(X) = \sqrt{\lambda}.$$

Kürzel: Man sagt, X ist „Pois(λ)-verteilt“.

MuPAD-Routinen: `stats::poissonCDF`, `stats::poissonQuantile`,
`stats::poissonPF`, `stats::poissonRandom`.

Interpretation: Im folgenden Sinne beschreibt die Poisson-Verteilung „seltene Bernoulli-Ereignisse“.

Betrachte eine binomial-verteilte Variable X mit großem n und kleinem $p \ll 1/\sqrt{n} \ll 1$. Mit $\lambda = pn$ gilt dann

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad (\text{„Poisson-Näherung“})$$

für alle ganzzahligen Werte $k \ll \sqrt{n}$.

Etwas genauer:

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} \left(1 + \frac{k}{2n} - \frac{k^2}{2n} - \frac{p^2 n}{2} + kp + \dots \right).$$

Beweis:

Mit den Forderungen $n \gg 1$, $p \ll 1/\sqrt{n}$ und $k \ll \sqrt{n}$ gelten folgende Bedingungen:

$$n \gg 1, \quad n - k \gg 1, \quad \frac{k}{n} \ll \frac{1}{\sqrt{n}} \ll 1, \quad \frac{k^2}{n} \ll 1, \quad \frac{p^2}{n} \ll 1, \quad kp \ll 1.$$

Es gilt die Stirling-Formel

$$m! \approx \sqrt{2\pi m} \ m^m e^{-m} \left(1 + \frac{1}{12m} + \frac{1}{288m^2} + \dots\right).$$

Schon ab $m = 5$ ist die Näherung $m! \approx \sqrt{2\pi m} \ m^m e^{-m}$ auf besser als 2% genau. Es folgt für $n \gg 1$, $n - k \gg 1$:

$$\begin{aligned} \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\frac{\lambda^k}{k!} e^{-\lambda}} &= \frac{n!}{(n-k)!} \frac{1}{n^k} e^\lambda e^{(n-k) \ln(1-p)} \\ &= \frac{n!}{(n-k)!} \frac{1}{n^k} e^{np} e^{n \ln(1-p)} e^{-k \ln(1-p)} \\ &\stackrel{\text{(Stirling)}}{\approx} \frac{\sqrt{n} n^n e^{-n}}{\sqrt{n-k} (n-k)^{n-k} e^{-(n-k)}} \frac{1}{n^k} e^{np+n \ln(1-p)} e^{-k \ln(1-p)} \\ &= \frac{1}{\sqrt{1-\frac{k}{n}}} \frac{e^{-k}}{\left(1-\frac{k}{n}\right)^{n-k}} e^{np+n \ln(1-p)} e^{-k \ln(1-p)} \\ &= \frac{1}{\sqrt{1-\frac{k}{n}}} e^{-k-(n-k) \ln\left(1-\frac{k}{n}\right)} e^{n(p+\ln(1-p))} e^{-k \ln(1-p)} \\ &= \frac{1}{\sqrt{1-\frac{k}{n}}} e^{-k\left(1+\left(\frac{n}{k}-1\right) \ln\left(1-\frac{k}{n}\right)\right)} e^{n(p+\ln(1-p))} e^{-k \ln(1-p)}. \end{aligned}$$

Beachte $\ln(1-\epsilon) = -\epsilon - \epsilon^2/2 - \epsilon^3/3 - \dots$ für $\epsilon < 1$. Für $k/n \ll 1$ gelten damit die Approximationen

$$\frac{1}{\sqrt{1-\frac{k}{n}}} = 1 + \frac{1}{2} \frac{k}{n} + \frac{3}{8} \frac{k^2}{n^2} + \dots \approx 1 + \frac{k}{2n},$$

$$-k \left(1 + \left(\frac{n}{k} - 1\right) \ln\left(1 - \frac{k}{n}\right)\right) = -\frac{k^2}{2n} \left(1 + \frac{1}{3} \frac{k}{n} + \frac{1}{6} \frac{k^2}{n^2} + \dots\right) \approx -\frac{k^2}{2n}.$$

Analog mit $p \ll 1$:

$$n(p + \ln(1-p)) = -\frac{np^2}{2} \left(1 + \frac{2p}{3} + \frac{p^2}{2} + \dots\right) \approx -\frac{np^2}{2},$$

$$-k \ln(1-p) = kp \left(1 + \frac{p}{2} + \frac{p^2}{3} + \dots\right) \approx kp.$$

Mit $k^2/n \ll 1$, $p^2 n \ll 1$, $kp \ll 1$ folgt dann

$$\frac{\binom{n}{k} p^k (1-p)^{n-k}}{\frac{\lambda^k}{k!} e^{-\lambda}} \approx \left(1 + \frac{k}{2n}\right) e^{-\frac{k^2}{2n}} e^{-\frac{p^2 n}{2}} e^{kp}$$

$$\approx 1 + \frac{k}{2n} - \frac{k^2}{2n} - \frac{p^2 n}{2} + kp \approx 1 .$$

Q.E.D.

Merke: Die Poisson-Verteilung ist zuständig für häufige ($n \gg 1$) Wiederholungen von Bernoulli-Experimenten mit kleiner Erfolgsw'keit $p \ll 1/\sqrt{n} \ll 1$, deren Erwartungswert $\lambda = np$ bekannt ist.

Beispiel 2.40: Man stellt fest, dass eine Sekretärin im Durchschnitt 1.5 Tippfehler pro Seite macht. Mit welcher W'keit hat eine betrachtete Seite keinen Tippfehler?

Lösung: Intuitiv ist das Auftauchen eines Tippfehlers ein „seltenes Ereignis“, also verwende man die Poisson-Verteilung

$$P(\text{„genau } k \text{ Tippfehler auf einer Seite“}) = \frac{\lambda^k e^{-\lambda}}{k!}$$

mit dem Erwartungswert $\lambda = 1.5$. Damit folgt

$$P(\text{„kein Tippfehler“}) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-1.5} \approx 0.223 .$$

Genaueres Modell: Sei p die W'keit, beim Tippen eines Buchstabens die falsche Taste zu drücken. Voraussetzung: p sei für jeden Buchstaben gleich groß. Das Tippen einer Seite wird damit zu einer häufigen Wiederholung ($n =$ Anzahl der Buchstaben auf einer Seite, Größenordnung 1000) eines Bernoulli-Experiments mit geringer „Erfolgs“-W'keit. Ein „Erfolg“ = „Tippfehler“ ist in der Tat unwahrscheinlich: der Erwartungswert $\lambda = np$ ist empirisch als 1.5 bestimmt, also ist $p = P(\text{„Tippfehler“})$ von der Größenordnung $1.5/1000$. Mit $p \approx 0.0015 \ll 1/\sqrt{n} \approx 0.03$ lässt sich die Poisson-Näherung anwenden. Für $X =$ „Anzahl der Tippfehler auf einer Seite“ ergibt sich

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \lambda = 1.5 .$$

Weitere Frage: Mit welcher W'keit wird ein 20 Seiten langes Skript weniger als 5 Tippfehler enthalten?

Lösung: Nun ist $\tilde{n} = 20n$, wobei n wie oben die Anzahl der Buchstaben auf einer Seite sei. Für p gilt (halbwegs) $p \approx 0.0015 \ll 1/\sqrt{\tilde{n}} \approx 0.007$, die Poisson-Näherung ist damit noch einigermaßen akzeptabel. Mit dem Erwartungswert von

$$\tilde{\lambda} = \tilde{n}p = 20np = 20\lambda = 20 \cdot 1.5 = 30$$

Tippfehlern für 20 Seiten erhält man für die Variable $\tilde{X} =$ „Anzahl der Tippfehler im gesamten Manuskript“:

$$P(\tilde{X} < 5) = \sum_{k=0}^4 \frac{30^k e^{-30}}{k!} \approx 3.6 \cdot 10^{-9} .$$

Das ist realistisch: wie jeder passionierte Leser weiß, gibt es keine Bücher ohne Tippfehler.

2.4.3 Die geometrische Verteilung

Zu gegebenem $p \in (0, 1]$, $q = 1 - p$, betrachte $X : \Omega \rightarrow \{1, 2, 3, \dots\}$ mit

$$P(X = k) = p \cdot q^{k-1}, \quad k = 1, 2, 3, \dots$$

Es gilt

$$E(X) = \frac{1}{p}, \quad \text{Var}(X) = \frac{q}{p^2}, \quad \sigma(X) = \frac{\sqrt{q}}{p}.$$

Kürzel: Man sagt, X ist „Geom(p)-verteilt“ (oder auch Geo(p)).

MuPAD-Routinen: `stats::geometricCDF`, `stats::geometricQuantile`,
`stats::geometricPF`, `stats::geometricRandom`.

Interpretation: Nach Blatt 6, Aufgabe 34, beschreibt X die Anzahl der benötigten Versuche, bis in der Wiederholung eines Bernoulli-Experiments mit Erfolgsw'keit p zum ersten Mal „Erfolg“ eintritt.

Beispiel: Man muss im Mittel 13 983 816 Mal Lotto spielen, bis man „6 Richtige“ erzielt (nach Beispiel 1.8 ist die Erfolgsw'keit des Bernoulli-Experiments „6 Richtige bei einem Lotto-Spiel“ $p = \frac{1}{13\,983\,816}$).

2.4.4 Die hypergeometrische Verteilung

Zu gegebenem $N, S, n \in \{0, 1, 2, \dots\}$ mit $0 \leq S \leq N$ und $0 \leq n \leq S$ betrachte $X : \Omega \rightarrow \{0, 1, \dots, \min(n, S)\}$ mit

$$P(X = s) = \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}, \quad s = 0, 1, \dots, \min(n, S).$$

Es gilt

$$E(X) = \frac{S n}{N}, \quad \text{Var}(X) = \sigma^2(X) = \frac{(N-S)(N-n)S n}{(N-1)N^2}.$$

Kürzel: Man sagt, X ist „Hypergeom(N, S, n)-verteilt“ (oder auch HyG(N, S, n)).

MuPAD-Routinen: `stats::hypergeometricCDF`,
`stats::hypergeometricQuantile`, `stats::hypergeometricPF`,
`stats::hypergeometricRandom`.

Interpretation: Nach Beispiel 1.10 beschreibt X die Anzahl der „Erfolge“, wenn man n Mal ohne Zurücklegen aus einer Urne mit N Objekten zieht, von

denen S als „Erfolg“ gelten.

Beispiel: Die Anzahl der Asse pro Hand beim Skat sind HyperGeom(32, 4, 10)-verteilt. Man erhält im Mittel $E(X) = 1.25$ Asse.

2.4.5 Die Exponentialverteilung

Zu gegebenem $\lambda > 0$ betrachte $X : \Omega \rightarrow [0, \infty)$ mit der Dichte

$$\rho(r) = \lambda e^{-\lambda r}, \quad r \geq 0$$

(und $\rho(r) = 0$ für $r < 0$), also für $0 \leq t$:

$$1 - F_X(t) = P(t < X) = \int_t^\infty \lambda e^{-\lambda r} dr = e^{-\lambda t}.$$

Man rechnet leicht nach:

$$E(X) = \frac{1}{\lambda}; \quad \text{Var}(X) = \frac{1}{\lambda^2}; \quad \sigma(X) = \frac{1}{\lambda}.$$

Kürzel: Man sagt, X ist „Exp(λ)-verteilt“.

MuPAD-Routinen: `stats::exponentialCDF`,
`stats::exponentialQuantile`, `stats::exponentialPDF`,
`stats::exponentialRandom`.

Interpretation: Ein Exp(λ)-verteiltes X beschreibt die Wartezeit bis zum ersten Eintreten eines Ereignisses, wenn die beobachteten Ereignisse unabhängig voneinander mit zeitlich konstanter Rate $\lambda =$ „mittlere Anzahl von Ereignissen pro Zeiteinheit“ eintreten. Diese Verteilung hat „kein Gedächtnis“: für alle $t, t_0 \in [0, \infty)$ ergibt sich die bedingte W'keit

$$P(t_0 + t < X \mid t_0 < X) = \frac{P(t + t_0 < X)}{P(t_0 < X)} = \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} = P(t < X).$$

Im Klartext: zum Zeitpunkt 0 habe ich eine Beobachtung begonnen. $P(t < X)$ ist die W'keit, dass das Ereignis nicht innerhalb der ersten t Zeiteinheiten eintritt. Ich habe nun t_0 Zeiteinheiten gewartet, das Ereignis ist nicht eingetreten. Die bedingte W'keit, dass das Ereignis nicht in den nächsten t Zeiteinheiten eintritt, stimmt mit $P(t < X)$ überein. Also: es ist egal, zu welchem Zeitpunkt t_0 ich mit der Beobachtung beginne.

Beispiel: radioaktiver Zerfall, siehe Beispiel 2.27.

Beispiel 2.41: Ich beobachte im Mittel jede halbe Stunde einen Meteoriten am Himmel. Mit welcher W'keit beobachte ich einen Meteoriten innerhalb der ersten 15 Minuten meiner Beobachtung?

Lösung: Sei $X =$ „der Zeitpunkt, an dem ich zum ersten Mal seit Beginn meiner Beobachtung einen Meteoriten sehe“. Der Erwartungswert ist als $E(X) = 1/\lambda = 1/2$ vorgegeben, also $\lambda = 2$ (Meteoriten pro Stunde). Damit ist die W'keit, in den ersten 15 Minuten (mindestens) einen Meteoriten zu sehen

$$P\left(X \leq \frac{15}{60}\right) = \int_0^{1/4} \lambda e^{-\lambda r} dr = 1 - e^{-\lambda/4} = 1 - e^{-1/2} \approx 0.39.$$

2.4.6 Die Gleichverteilung

↓11.6.07

Zu gegebenem Intervall $[a, b] \subset \mathbb{R}$ betrachte $X : \Omega \rightarrow [a, b]$ mit der konstanten Dichte

$$\rho(r) = \frac{1}{b-a}, \quad r \in [a, b]$$

(und $\rho(r) = 0$ für $r \notin [a, b]$). Man rechnet leicht nach:

$$E(X) = \frac{a+b}{2}; \quad \text{Var}(X) = \frac{(b-a)^2}{12}; \quad \sigma(X) = \frac{b-a}{\sqrt{12}}.$$

Kürzel: Man sagt, X ist „UC(a, b)-verteilt“ (uniform, continuous).

MuPAD-Routinen: `stats::uniformCDF`, `stats::uniformQuantile`,
`stats::uniformPDF`, `stats::uniformRandom`.

Interpretation: Dies ist das kontinuierliche Analogon zur diskreten „Gleichwahrscheinlichkeit“ der kombinatorischen Modelle. Jeder Punkt aus $[a, b]$ wird mit der gleichen W'keit gewählt.

Beispiel: Drehen eines Glücksrads. Die durch den Winkel $\in [0, 2\pi)$ gegen eine fixierte Markierung beschriebene Endstellung ist gleichverteilt.

Bemerkung 2.42: In Softwaresystemen steht typischerweise ein Zufallszahlengenerator für gleichverteilte Gleitpunktzahlen zwischen 0 und 1 zur Verfügung. Sei $Y : \Omega \mapsto [0, 1]$ dieser UC(0,1)-Generator. Hiermit kann man sich leicht einen Zufallszahlengenerator bauen, dessen Werte X einer beliebigen vorgegebenen Verteilungsfunktion F genügen, die natürlich die in Satz 2.9 beschriebenen Eigenschaften haben muss. Man erzeugt dazu eine UC(0,1)-Zahl Y und berechnet

dann (für kontinuierliches F) den Wert $X = F^{-1}(Y)$, indem man die Gleichung $F(X) = Y$ (typischerweise numerisch) löst. Damit hat X in der Tat die Verteilungsfunktion $F_X = F$:

$$F_X(r) = P(X \leq r) = P(F^{-1}(Y) \leq r) \stackrel{(F \text{ streng monoton})}{=} P(Y \leq F(r))$$

$$\stackrel{UC(0,1)}{=} \int_0^{F(r)} d\tilde{r} = F(r) .$$

Anmerkung: Die Vorschrift „löse $F(X) = Y$ “ gilt, wenn die Verteilung kontinuierlich ist, genauer, wenn F streng monoton wachsend ist und keine Sprünge macht. Dann hat $F(X) = Y$ genau eine Lösung. Für diskrete Verteilungen (F ist eine Treppenfunktion) hat $F(X) = Y$ i.A. keine Lösung, da Y nur mit W'keit 0 einen der diskreten Werte trifft. Die genaue Vorschrift für den allgemeinen Fall ist, X zu bestimmen als:

$$X = \text{Quantil}_F(Y) = \min \{x; Y \leq F(x)\},$$

wobei Quantil_F die in Definition 2.16 eingeführte Quantilfunktion der Verteilungsfunktion F ist. Das Minimum existiert (für $Y \in (0, 1]$), da F rechtsseitig stetig ist. Wir zeigen nun allgemein, dass X die Verteilungsfunktion F hat. Für $r \in \mathbb{R}$ definiere:

$$A = \{\omega \in \Omega; X(\omega) \leq r\} = \{\omega \in \Omega; \min \{x; Y(\omega) \leq F(x)\} \leq r\} ,$$

$$B = \{\omega \in \Omega; Y(\omega) \leq F(r)\} .$$

Es gilt zu zeigen, dass $A = B$ gilt.

i) Sei $\omega \in A$. Damit gilt $x_0 \leq r$ für $x_0 = \min \{x; Y(\omega) \leq F(x)\}$ sowie $Y(\omega) \leq F(x_0)$. Da F monoton ist, gilt mit $x_0 \leq r$ auch $Y(\omega) \leq F(x_0) \leq F(r)$. Damit gilt $\omega \in B$. Wir haben also $A \subset B$ gezeigt.

ii) Sei $\omega \in B$, also $Y(\omega) \leq F(r)$. Damit liegt r in der Menge $\{x; Y(\omega) \leq F(x)\}$ und es folgt $\min \{x; Y(\omega) \leq F(x)\} \leq r$. Damit gilt $\omega \in A$. Wir haben also $B \subset A$ gezeigt.

Aus $A \subset B$ und $B \subset A$ folgt $A = B$ und damit

$$P(A) = P(X \leq r) = F_X(r) = P(B) = P(Y \leq F(r)) \stackrel{UC(0,1)}{=} F(r) .$$

2.4.7 Die Normal-(Gauß-)Verteilung

13.6.07↓

Dies ist die wichtigste aller Verteilungen, weil nach den Grenzwertsätzen des Kapitels 4 die Mittelwerte von unabhängigen Wiederholungen *beliebiger* Zufallsexperimente gegen eine Normalverteilung konvergieren. Diese Verteilung kommt daher immer ins Spiel, wenn man das genaue Verhalten (das unterliegende Modell bzw. die Verteilung der Zufallsvariable) gar nicht kennt und durch häufige Wiederholungen statistische Daten ermittelt.

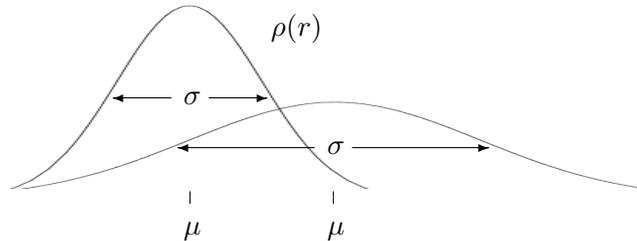
Zu gegebenem $\mu \in \mathbb{R}$ und $\sigma > 0$ betrachte $X : \Omega \rightarrow \mathbb{R}$ mit der Dichte

$$\rho(r) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}.$$

Man rechnet nach:

$$E(X) = \mu; \quad \text{Var}(X) = \sigma^2; \quad \sigma(X) = \sigma.$$

Der Parameter μ ist die Stelle, wo die W'keitsdichte ihr Maximum annimmt. Höhe und Γ



Kürzel: Man sagt, X ist „ $N(\mu, \sigma^2)$ -verteilt“.

MuPAD-Routinen: `stats::normalCDF`, `stats::normalQuantile`, `stats::normalPDF`, `stats::normalRandom`.

Interpretation und Beispiele: Die Mittelwerte unabhängiger Wiederholungen beliebiger Zufallsexperimente sind approximativ normalverteilt, wenn man nur genügend oft wiederholt. Details und Beispiele sind mit dem „Zentralen Grenzwertsatz“ (Abschnitt 4.3) beschrieben.

Bemerkung 2.43: Die Funktion

$$\text{erf}(r) = \frac{2}{\sqrt{\pi}} \int_0^r e^{-x^2} dx$$

heißt „**Fehlerfunktion**“ (engl.: „**error function**“). Sie ist eine Standardfunktion der Mathematik und in vielen Softwarepaketen (z.B. MuPAD) verfügbar. Es gilt der folgende Zusammenhang mit der $N(\mu, \sigma^2)$ -Verteilungsfunktion:

$$\begin{aligned} \boxed{F_X(r)} &= \int_{-\infty}^r \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \stackrel{\substack{y=\frac{x-\mu}{\sigma\sqrt{2}} \\ dx = \sigma\sqrt{2} dy}}{=} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{r-\mu}{\sigma\sqrt{2}}} e^{-y^2} dy \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^0 e^{-y^2} dy + \frac{1}{2} \text{erf}\left(\frac{r-\mu}{\sigma\sqrt{2}}\right) \quad \boxed{= \frac{1}{2} \left(1 + \text{erf}\left(\frac{r-\mu}{\sigma\sqrt{2}}\right)\right)}. \end{aligned}$$

Eine weitere interessante Funktion ist

$$\Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-r}^r e^{-\frac{x^2}{2}} dx = \operatorname{erf}\left(\frac{r}{\sqrt{2}}\right).$$

Für eine $N(0, 1)$ -verteilte Variable X gilt die Interpretation:

$$\Phi(r) = P(-r \leq X \leq r).$$

Eine Wertetabelle für Φ ist auf Seite 81 beigefügt.

2.5 Unabhängigkeit von Zufallsvariablen

Für die Praxis ist die Frage nach der Unabhängigkeit von Zufallsvariablen sehr wichtig: erhöhen AKWs das Risiko, an Krebs zu erkranken? Besteht ein Zusammenhang zwischen der beruflichen Tätigkeit und den Vorlieben, wo und wie der Urlaub verbracht wird?

2.5.1 Definition und Folgerungen

Das Problem bei solchen Fragestellungen ist meist, dass man den unterliegenden Stichprobenraum nicht wirklich kennt und bestenfalls die Verteilung der interessierenden Zufallsvariablen einzeln durch empirische Messungen approximieren kann (Statistik). Leider ist aus den Verteilungen der Zufallsvariablen die Unabhängigkeit nicht abzulesen: man braucht zusätzliche Informationen, die nur dem den beobachteten Größen unterliegenden Modell (Ω, \mathcal{E}, P) zu entnehmen sind, die „gemeinsame Verteilung“:

Definition 2.44:

Seien $X_1, \dots, X_n : \Omega \mapsto \mathbb{R}$ Zufallsvariablen über einem gemeinsamen Modell (Ω, \mathcal{E}, P) . Die Funktion

$$\begin{aligned} F_{X_1, \dots, X_n}(r_1, \dots, r_n) &\equiv P(X_1 \leq r_1 ; \dots ; X_n \leq r_n) \\ &= P\left(X_1^{-1}((-\infty, r_1]) \cap \dots \cap X_n^{-1}((-\infty, r_n])\right) \end{aligned}$$

heißt **gemeinsame Verteilungsfunktion** der Variablen X_1, \dots, X_n .

Satz 2.45:

Gemeinsame Verteilungsfunktionen haben allgemein folgende Eigenschaften:

- a) $F_{X_1, \dots, X_n}(r_1, \dots, r_n)$ ist monoton steigend und rechtsseitig stetig in jedem r_i (bei fixierten $r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_n$).

Wertetabelle für die $N(0, 1)$ -W'keiten $P(|X| \leq r) = \Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-r}^r e^{-\frac{x^2}{2}} dx$:

r	$\Phi(r)$								
0.00	0.000	0.60	0.451	1.20	0.770	1.80	0.928	2.40	0.9836
0.01	0.008	0.61	0.458	1.21	0.774	1.81	0.930	2.41	0.9840
0.02	0.016	0.62	0.465	1.22	0.778	1.82	0.931	2.42	0.9845
0.03	0.024	0.63	0.471	1.23	0.781	1.83	0.933	2.43	0.9849
0.04	0.032	0.64	0.478	1.24	0.785	1.84	0.934	2.44	0.9853
0.05	0.040	0.65	0.484	1.25	0.789	1.85	0.936	2.45	0.9857
0.06	0.048	0.66	0.491	1.26	0.792	1.86	0.937	2.46	0.9861
0.07	0.056	0.67	0.497	1.27	0.796	1.87	0.939	2.47	0.9865
0.08	0.064	0.68	0.503	1.28	0.799	1.88	0.940	2.48	0.9869
0.09	0.072	0.69	0.510	1.29	0.803	1.89	0.941	2.49	0.9872
0.10	0.080	0.70	0.516	1.30	0.806	1.90	0.943	2.50	0.9876
0.11	0.088	0.71	0.522	1.31	0.810	1.91	0.944	2.51	0.9879
0.12	0.096	0.72	0.528	1.32	0.813	1.92	0.945	2.52	0.9883
0.13	0.103	0.73	0.535	1.33	0.816	1.93	0.946	2.53	0.9886
0.14	0.111	0.74	0.541	1.34	0.820	1.94	0.948	2.54	0.9889
0.15	0.119	0.75	0.547	1.35	0.823	1.95	0.949	2.55	0.9892
0.16	0.127	0.76	0.553	1.36	0.826	1.96	0.950	2.56	0.9895
0.17	0.135	0.77	0.559	1.37	0.829	1.97	0.951	2.57	0.9898
0.18	0.143	0.78	0.565	1.38	0.832	1.98	0.952	2.58	0.9901
0.19	0.151	0.79	0.570	1.39	0.835	1.99	0.953	2.59	0.9904
0.20	0.159	0.80	0.576	1.40	0.838	2.00	0.9545	2.60	0.9907
0.21	0.166	0.81	0.582	1.41	0.841	2.01	0.9556	2.61	0.9909
0.22	0.174	0.82	0.588	1.42	0.844	2.02	0.9566	2.62	0.9912
0.23	0.182	0.83	0.593	1.43	0.847	2.03	0.9576	2.63	0.9915
0.24	0.190	0.84	0.599	1.44	0.850	2.04	0.9586	2.64	0.9917
0.25	0.197	0.85	0.605	1.45	0.853	2.05	0.9596	2.65	0.9920
0.26	0.205	0.86	0.610	1.46	0.856	2.06	0.9606	2.66	0.9922
0.27	0.213	0.87	0.616	1.47	0.858	2.07	0.9615	2.67	0.9924
0.28	0.221	0.88	0.621	1.48	0.861	2.08	0.9625	2.68	0.9926
0.29	0.228	0.89	0.627	1.49	0.864	2.09	0.9634	2.69	0.9929
0.30	0.236	0.90	0.632	1.50	0.866	2.10	0.9643	2.70	0.9931
0.31	0.243	0.91	0.637	1.51	0.869	2.11	0.9651	2.72	0.9935
0.32	0.251	0.92	0.642	1.52	0.871	2.12	0.9660	2.74	0.9939
0.33	0.259	0.93	0.648	1.53	0.874	2.13	0.9668	2.76	0.9942
0.34	0.266	0.94	0.653	1.54	0.876	2.14	0.9676	2.78	0.9946
0.35	0.274	0.95	0.658	1.55	0.879	2.15	0.9684	2.80	0.9949
0.36	0.281	0.96	0.663	1.56	0.881	2.16	0.9692	2.82	0.9952
0.37	0.289	0.97	0.668	1.57	0.884	2.17	0.9700	2.84	0.9955
0.38	0.296	0.98	0.673	1.58	0.886	2.18	0.9707	2.86	0.9958
0.39	0.303	0.99	0.678	1.59	0.888	2.19	0.9715	2.88	0.9960
0.40	0.311	1.00	0.683	1.60	0.890	2.20	0.9722	2.90	0.9963
0.41	0.318	1.01	0.688	1.61	0.893	2.21	0.9729	2.92	0.9965
0.42	0.326	1.02	0.692	1.62	0.895	2.22	0.9736	2.94	0.9967
0.43	0.333	1.03	0.697	1.63	0.897	2.23	0.9743	2.96	0.9969
0.44	0.340	1.04	0.702	1.64	0.899	2.24	0.9749	2.98	0.9971
0.45	0.347	1.05	0.706	1.65	0.901	2.25	0.9756	3.00	0.99730
0.46	0.354	1.06	0.711	1.66	0.903	2.26	0.9762	3.10	0.99806
0.47	0.362	1.07	0.715	1.67	0.905	2.27	0.9768	3.20	0.99863
0.48	0.369	1.08	0.720	1.68	0.907	2.28	0.9774	3.30	0.99903
0.49	0.376	1.09	0.724	1.69	0.909	2.29	0.9780	3.40	0.99933
0.50	0.383	1.10	0.729	1.70	0.911	2.30	0.9786	3.60	0.99968
0.51	0.390	1.11	0.733	1.71	0.913	2.31	0.9791	3.50	0.99953
0.52	0.397	1.12	0.737	1.72	0.915	2.32	0.9797	3.70	0.99978
0.53	0.404	1.13	0.742	1.73	0.916	2.33	0.9802	3.80	0.99986
0.54	0.411	1.14	0.746	1.74	0.918	2.34	0.9807	3.90	0.99990
0.55	0.418	1.15	0.750	1.75	0.920	2.35	0.9812	4.00	0.999937
0.56	0.425	1.16	0.754	1.76	0.922	2.36	0.9817	5.00	0.999999
0.57	0.431	1.17	0.758	1.77	0.923	2.37	0.9822		
0.58	0.438	1.18	0.762	1.78	0.925	2.38	0.9827		
0.59	0.445	1.19	0.766	1.79	0.927	2.39	0.9832		

$$b) F_{X_1, \dots, X_n}(r_1, \dots, \underbrace{-\infty}_{r_i}, \dots, r_n) = 0 .$$

$$c) F_{X_1, \dots, X_n}(r_1, \dots, \underbrace{\infty}_{r_i}, \dots, r_n) = \\ F_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_n) .$$

Beweisskizze:

a) folgt analog zur Beweisskizze zu Satz 2.9.

b) ergibt sich aus

$$\lim_{r_i \rightarrow -\infty} P(X_1 \leq r_1 ; \dots ; X_i \leq r_i ; \dots ; X_n \leq r_n) \\ = P\left(\bigcap_{r_i \in \mathbb{R}} \left\{\omega \in \Omega; X_1(\omega) \leq r_1 ; \dots, X_i(\omega) \leq r_i ; \dots, X_n(\omega) \leq r_n\right\}\right) \\ = P(\emptyset) = 0 .$$

c) ergibt sich aus

$$\lim_{r_i \rightarrow \infty} P(X_1 \leq r_1 ; \dots ; X_i \leq r_i ; \dots ; X_n \leq r_n) \\ = P\left(\bigcup_{r_i \in \mathbb{R}} \left\{\omega \in \Omega; X_1(\omega) \leq r_1 ; \dots ; X_i(\omega) \leq r_i ; \dots ; X_n(\omega) \leq r_n\right\}\right) \\ = P\left(\{\omega \in \Omega; X_1(\omega) \leq r_1 ; \dots ; \cancel{X_i(\omega) \in \mathbb{R}} ; \dots ; X_n(\omega) \leq r_n\}\right) .$$

Q.E.D.

Definition 2.46:

Zufallsvariablen $X_1, \dots, X_n : \Omega \mapsto \mathbb{R}$ über einem gemeinsamen Modell (Ω, \mathcal{E}, P) heißen **unabhängig**, wenn für jede Wahl von r_1, \dots, r_n die Ereignisse

$$X_1^{-1}((-\infty, r_1]) , \dots , X_n^{-1}((-\infty, r_n])$$

eine im Sinne von Definition 1.54 unabhängige Ereignisfamilie bilden.

Bemerkung 2.47: (Warnung) Gemäß der Warnung 1.55 gilt wiederum: ist eine Familie von Variablen X_1, \dots, X_n unabhängig, so sind auch jeweils Paare von Variablen unabhängig. Es gilt aber nicht die Umkehrung: auch wenn die Zufallsvariablen paarweise unabhängig sind, so braucht die Gesamtfamilie nicht unabhängig zu sein.

Intuitiv hätte man erwartet, dass Unabhängigkeit von Zufallsvariablen dadurch definiert wird, dass die Ereignisse $X_1 \in E_1, X_2 \in E_2$ etc. für beliebige $E_1, \dots, E_n \subset \mathbb{R}$ unabhängig sein sollen. Die obige Definition beschränkt sich auf die speziellen Ereignisse $E_i = (-\infty, r_i]$. Dies genügt, denn daraus folgt die Unabhängigkeit für beliebige E_i :

Satz 2.48:

Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn die gemeinsame Verteilung das Produkt der einzelnen Verteilungen ist:

$$F_{X_1, \dots, X_n}(r_1, \dots, r_n) = F_{X_1}(r_1) \cdot \dots \cdot F_{X_n}(r_n) .$$

Dies ist genau dann der Fall, wenn für beliebige Ereignisse $E_1, \dots, E_n \subset \mathbb{R}$ gilt:

$$P(X_1 \in E_1 ; \dots ; X_n \in E_n) = P(X_1 \in E_1) \cdot \dots \cdot P(X_n \in E_n) .$$

Beweisskizze:

Aus der Definition 1.54 folgt unmittelbar, dass die gemeinsame Verteilung für unabhängige Variablen faktorisieren muss. Andererseits, wenn sie faktorisiert, so gilt mit Satz 2.45.c) und $F_{X_i}(\infty) = 1$ auch, dass die gemeinsame Verteilung jeder Teilauswahl von Variablen aus X_1, \dots, X_n faktorisiert, womit nach Definition auch für jede Teilfamilie der Ereignisse

$$\mathcal{A} = \left\{ X_1^{-1}((-\infty, r_1]) , \dots , X_n^{-1}((-\infty, r_n]) \right\}$$

die Unabhängigkeit

$$P(X_{i_1} \leq r_{i_1} ; \dots ; X_{i_k} \leq r_{i_k}) = P(X_{i_1} \leq r_{i_1}) \cdot \dots \cdot P(X_{i_k} \leq r_{i_k})$$

gilt. Die Unabhängigkeit der Variablen sei nun vorausgesetzt, d.h., die Produktformel gilt für die speziellen Ereignisse „ $X_1 \leq r_1 ; \dots ; X_n \leq r_n$ “. Wir zeigen hier nur, dass sich dies verallgemeinert auf die „typischen“ Ereignisse $E_i = (a_i, b_i]$ (und Schnitte und Vereinigungen solcher Intervalle). Sei dazu zunächst

$$E_1 = (a_1, b_1] , E_2 = (-\infty, r_1] , \dots , E_n = (-\infty, r_n] .$$

Aus der Definition der gemeinsamen Verteilung folgt unmittelbar:

$$\begin{aligned} & P(X_1 \in E_1 ; \dots ; X_n \in E_n) \\ &= F_{X_1, \dots, X_n}(b_1, r_2, \dots, r_n) - F_{X_1, \dots, X_n}(a_1, r_2, \dots, r_n) \\ &= \left(F_{X_1}(b_1) - F_{X_1}(a_1) \right) \cdot F_{X_2}(r_2) \cdot \dots \cdot F_{X_n}(r_n) \\ &= P(a_1 < X_1 \leq b_1) \cdot P(X_2 \leq r_2 ; \dots ; X_n \leq r_n) . \end{aligned}$$

Es ist ziemlich offensichtlich, wie man mit analogen Argumenten die Produktformel auch für allgemeineres E_1 zeigen kann, das sich durch Schnitte oder Vereinigungen solcher Intervalle ergibt. Hat man erst einmal die Produktformel für allgemeine Ereignisse $E_1 \subset \mathbb{R}$ und die speziellen Ereignisse $E_2 = (-\infty, r_2], \dots, E_n = (-\infty, r_n]$ gezeigt, also

$$\begin{aligned} P(X \in E_1; X_2 \leq r_2; \dots; X_n \leq r_n) \\ = P(X \in E_1) \cdot P(X_2 \leq r_2; \dots; X_n \leq r_n), \end{aligned}$$

so folgt rekursiv dann auch sofort die vollständige Produktzerlegung für beliebige Ereignisse $X_2 \in E_2 \subset \mathbb{R}_2, \dots, X_n \in E_n \subset \mathbb{R}_n$.

Q.E.D.

Die folgende Bemerkung liefert ein leicht zu handhabendes Kriterium, mit dem bei diskreten Zufallsvariablen die Unabhängigkeit nachgerechnet werden kann:

18.6.07↓

Bemerkung 2.49: Nach Satz 2.48 bedeutet Unabhängigkeit bei diskreten Zufallsvariablen $X : \Omega \mapsto \{x_1, x_2, \dots\}, Y : \Omega \mapsto \{y_1, y_2, \dots\}$, dass

$$P(X = x_i; Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

für alle i, j gilt. In der Tat ist dies ein notwendiges und hinreichendes Kriterium für die Unabhängigkeit diskreter Variabler.

Beispiel 2.50: Betrachte den 2-fachen Wurf $\Omega = \{(\omega_1, \omega_2); \omega_1, \omega_2 \in \{1, \dots, 6\}\}$ eines fairen Würfels, sei $X =$ Ergebnis des ersten Wurfs, $Y =$ Ergebnis des zweiten Wurfs. Es gilt für alle Paare $i, j \in \{1, \dots, 6\}$:

$$\begin{aligned} P(X = i; Y = j) &= P(\{(i, j)\}) = \frac{1}{36}, \\ P(X = i) &= P(\{(i, \omega_2); \omega_2 \in \{1, \dots, 6\}\}) = \frac{1}{6}, \\ P(Y = j) &= P(\{(\omega_1, j); \omega_1 \in \{1, \dots, 6\}\}) = \frac{1}{6}. \end{aligned}$$

Für alle Paare i, j gilt damit $P(X = i; Y = j) = 1/36 = P(X = i) \cdot P(Y = j)$, womit X und Y unabhängig sind.

Anmerkung: dies ist ein einfaches Beispiel für die in Bemerkung 2.54 formalisierte Situation, wo die Zufallsvariablen sich auf unabhängige Durchführungen/Wiederholungen eines Zufallsexperiments (hier: einfacher Wurf des Würfels) beziehen. Hätte sich hier Abhängigkeit zwischen X und Y ergeben, wäre die angegebene Definition von Unabhängigkeit von Zufallsvariablen wenig sinnvoll gewesen.

Beispiel 2.51:

Man zieht zufällig ein Wort aus der nebenstehenden Urne. Sei X die Anzahl der Buchstaben im gezogenen Wort, sei Y die Anzahl der Vokale im gezogenen Wort. Sind X und Y unabhängig?

Der	Zufall
regiert	die Welt

Antwort: Für

$$\begin{aligned} X &: \{\text{'Der'}, \text{'Zufall'}, \text{'regiert'}, \text{'die'}, \text{'Welt'}\} \mapsto \{3, 4, 6, 7\}, \\ Y &: \{\text{'Der'}, \text{'Zufall'}, \text{'regiert'}, \text{'die'}, \text{'Welt'}\} \mapsto \{1, 2, 3\} \end{aligned}$$

ergibt sich folgende gemeinsame Verteilung:

	$Y = 1$	$Y = 2$	$Y = 3$	Σ
$X = 3$	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{2}{5} = P(X = 3)$
$X = 4$	$\frac{1}{5}$	0	0	$\frac{1}{5} = P(X = 4)$
$X = 6$	0	$\frac{1}{5}$	0	$\frac{1}{5} = P(X = 6)$
$X = 7$	0	0	$\frac{1}{5}$	$\frac{1}{5} = P(X = 7)$
Σ	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	1
	$\underbrace{\hspace{1.5cm}}_{P(Y=1)}$	$\underbrace{\hspace{1.5cm}}_{P(Y=2)}$	$\underbrace{\hspace{1.5cm}}_{P(Y=3)}$	

In dieser Tabelle sind die W'keiten $P(X = i; Y = j)$ aufgelistet, z. B.:

$$P(X = 3; Y = 1) = P(\{\text{'Der'}\}) = \frac{1}{5}, \quad P(X = 3; Y = 2) = P(\{\text{'die'}\}) = \frac{1}{5}$$

usw. Es gilt z.B.

$$P(X = 3; Y = 3) = 0 \neq P(X = 3) \cdot P(Y = 3) = \frac{2}{5} \cdot \frac{1}{5},$$

damit sind die Variablen abhängig.

Der folgende Sachverhalt liefert zwar kein handhabbares Kriterium, um Unabhängigkeit zu testen, ist aber wegen des folgenden Spezialfalls 2.53 interessant:

Bemerkung 2.52: Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn für alle (glatten) Funktion $f_1, \dots, f_n : \mathbb{R} \mapsto \mathbb{R}$ gilt:

$$\mathbb{E} \left(f_1(X_1) \cdot \dots \cdot f_n(X_n) \right) = \mathbb{E} (f_1(X_1)) \cdot \dots \cdot \mathbb{E} (f_n(X_n)) .$$

Beweisidee:

Man braucht hierzu das Konzept mehrdimensionaler Stieltjes-Integration

$$E(f(X_1, \dots, X_n)) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(r_1, \dots, r_n) dF_{X_1, \dots, X_n}(r_1, \dots, r_n)$$

für Funktionen $f : \mathbb{R}^n \mapsto \mathbb{R}$. Wenn F_{X_1, \dots, X_n} und f faktorisieren, so faktorisiert auch das Integral über den „Satz von Fubini“.

Q.E.D.

Als Spezialfall folgt:

Satz 2.53: (Unabhängigkeit \Rightarrow Erwartungswerte faktorisieren)

Für unabhängige Zufallsvariable $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ gilt:

$$E(X_1 \cdot \dots \cdot X_n) = E(X_1) \cdot \dots \cdot E(X_n) .$$

Beweisskizze:

Mangels Technik (mehrdimensionale Integration) führen wir den Beweis nur für diskrete Variable. Seien dazu $X_i(\Omega) = \{x_{i1}, x_{i2}, \dots\}$. Die Produktvariable

$$Y = X_1 \cdot \dots \cdot X_n : \Omega \rightarrow \{x_{1i_1} \cdot \dots \cdot x_{ni_n}; i_1, \dots, i_n \in \{1, 2, \dots\}\}$$

ist wieder diskret. Für den diskreten Spezialfall 2.25 gilt

$$\begin{aligned} E(Y) &= \sum_{y \in Y(\Omega)} y P(Y = y) \\ &= \sum_{i_1} \dots \sum_{i_n} x_{1i_1} \cdot \dots \cdot x_{ni_n} P(X_1 = x_{1i_1}; \dots; X_n = x_{ni_n}) \\ &= \sum_{i_1} \dots \sum_{i_n} x_{1i_1} \cdot \dots \cdot x_{ni_n} P(X_1 = x_{1i_1}) \cdot \dots \cdot P(X_n = x_{ni_n}) \\ &= \left(\sum_{i_1} x_{1i_1} P(X_1 = x_{1i_1}) \right) \cdot \dots \cdot \left(\sum_{i_n} x_{ni_n} P(X_n = x_{ni_n}) \right) \\ &= E(X_1) \cdot \dots \cdot E(X_n) . \end{aligned}$$

Q.E.D.

20.6.07↓

Es gibt eine natürliche Situation, in der unabhängige Variablen auftreten, nämlich bei unabhängig durchgeführten Zufallsexperimenten, denen jeweils getrennt eine Zufallsvariable zugeordnet wird. Im entsprechenden Produktmodell 1.57 ergibt sich folgendes Bild:

Bemerkung 2.54: Seien $X_i : \Omega_i \rightarrow \mathbb{R}$ ($i = 1, \dots, n$) Zufallsvariablen über (eventuell verschiedenen) Modellen $(\Omega_i, \mathcal{E}_i, P_i)$. Betrachte das Produktmodell

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{(\omega_1, \dots, \omega_n); \omega_i \in \Omega_i\}$$

gemäß Definition 1.57 mit dem W'keitsmaß

$$P(E_1 \times \dots \times E_n) = P_1(E_1) \cdot \dots \cdot P_n(E_n)$$

für Ereignisse $E_i \in \mathcal{E}_i$. Die durch

$$\hat{X}_i : (\omega_1, \dots, \omega_n) \rightarrow X_i(\omega_i)$$

definierten Zufallsvariablen sind unabhängig.

Beweis: Mit

$$\begin{aligned} & P(\hat{X}_1 \leq r_1 ; \dots ; \hat{X}_n \leq r_n) \\ &= P\left(\{(\omega_1, \dots, \omega_n); X_1(\omega_1) \leq r_1 ; \dots ; X_n(\omega_n) \leq r_n\}\right) \\ &= P\left(X_1^{-1}((-\infty, r_1]) \times \dots \times X_n^{-1}((-\infty, r_n])\right) \\ &= P_1(X_1 \leq r_1) \cdot \dots \cdot P_n(X_n \leq r_n) \\ &= P(\hat{X}_1 \leq r_1) \cdot \dots \cdot P(\hat{X}_n \leq r_n) \end{aligned}$$

faktoriert die Verteilungsfunktion.

Q.E.D.

Beispiel 2.55: Betrachte unabhängige Würfe eines Würfels. Sei X_1 das Ergebnis des ersten Wurfs, X_2 das Quadrat des Ergebnisses des zweiten Wurfs, X_3 irgendeine Funktion des Ergebnisses des dritten Wurfs usw. In diesem Fall kann man sofort sagen, dass die X_i unabhängig sind.

2.5.2 Kovarianz

Fassen wir zusammen: für beliebige Zufallsvariable X_1, \dots, X_n gilt die Linearität des Erwartungswerts

$$\mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)$$

(egal, ob die X_i unabhängig sind oder nicht). Für die Varianz gilt jedoch

$$\begin{aligned} \sigma^2(X_1 + \dots + X_n) &= \mathbb{E}\left(\left(\sum_i X_i\right)^2\right) - \left(\mathbb{E}\left(\sum_i X_i\right)\right)^2 \\ &= \mathbb{E}\left(\left(\sum_i X_i\right)\left(\sum_j X_j\right)\right) - \left(\sum_i \mathbb{E}(X_i)\right)\left(\sum_j \mathbb{E}(X_j)\right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i,j} \left(\mathbf{E}(X_i X_j) - \mathbf{E}(X_i) \mathbf{E}(X_j) \right) \\
&= \sum_i \left(\mathbf{E}(X_i^2) - \mathbf{E}(X_i)^2 \right) + \sum_{\substack{i,j \\ i \neq j}} \left(\mathbf{E}(X_i X_j) - \mathbf{E}(X_i) \mathbf{E}(X_j) \right) \\
&= \sum_i \sigma^2(X_i) + \sum_{\substack{i,j \\ i \neq j}} \left(\mathbf{E}(X_i X_j) - \mathbf{E}(X_i) \mathbf{E}(X_j) \right).
\end{aligned}$$

Varianzen addieren sich also im allgemeinen nicht. Die obige Betrachtung führt zu:

Definition 2.56:

Für die Zufallsvariablen X_1, \dots, X_n nennt man die Matrix mit den Einträgen

$$C_{ij} = \mathbf{E}(X_i X_j) - \mathbf{E}(X_i) \mathbf{E}(X_j) = \mathbf{E} \left((X_i - \mathbf{E}(X_i)) (X_j - \mathbf{E}(X_j)) \right)$$

„**Kovarianzmatrix**“ der Variablen.

Die Einträge $i \neq j$ der Kovarianzmatrix C_{ij} verschwinden für *paarweise unabhängige* Variablen. In diesem Fall gilt

$$\begin{aligned}
\mathbf{E} \left((X_i - \mathbf{E}(X_i)) (X_j - \mathbf{E}(X_j)) \right) &= \mathbf{E} \left(X_i - \mathbf{E}(X_i) \right) \cdot \mathbf{E} \left(X_j - \mathbf{E}(X_j) \right) \\
&= \left(\mathbf{E}(X_i) - \mathbf{E}(\mathbf{E}(X_i)) \right) \cdot \left(\mathbf{E}(X_j) - \mathbf{E}(\mathbf{E}(X_j)) \right) \\
&= \left(\mathbf{E}(X_i) - \mathbf{E}(X_i) \right) \cdot \left(\mathbf{E}(X_j) - \mathbf{E}(X_j) \right) = 0.
\end{aligned}$$

Merke 2.57:

Für paarweise unabhängige Variablen addieren sich neben den Erwartungswerten auch die Varianzen:

$$\sigma^2(X_1 + \dots + X_n) = \sigma^2(X_1) + \dots + \sigma^2(X_n).$$

2.6 Bedingte Erwartungswerte

Für Verteilungsfunktionen und Erwartungswerte gibt es das Analogon zu bedingten W'keiten:

Definition 2.58: (Bedingte Erwartung)

Sei $A \subset \mathcal{E}$ ein Ereignis in einem Modell (Ω, \mathcal{E}, P) , sei $X : \Omega \mapsto \mathbb{R}$ eine Zufallsvariable. Dann heißt

$$F_X(r | A) = P(X^{-1}((-\infty, r]) | A) \equiv P(X \leq r | A)$$

die **bedingte Verteilung** von X bei gegebenem Ereignis A . Sei $f : \mathbb{R} \mapsto \mathbb{R}$ eine glatte Funktion. Man nennt

$$E(f(X) | A) = \int_{\mathbb{R}} f(r) dF_X(r | A)$$

den **bedingten Erwartungswert** von $f(X)$ bei gegebenem A .

Im diskreten Fall $X(\Omega) = \{r_1, r_2, \dots\}$ ist dies:

$$E(f(X) | A) = \sum_{r \in X(\Omega)} f(r) P(X = r | A) .$$

Im kontinuierlichen Fall ist dies:

$$E(f(X) | A) = \int_{r \in X(\Omega)} f(r) \rho_A(r) dr,$$

wo $\rho_A(r) = \frac{d}{dr} F_X(r | A)$ die Dichte der bedingten Verteilung ist.

In Analogie zum Satz von der totalen W'keit 1.31 gilt:

Satz 2.59:

Sei U_1, \dots, U_n eine disjunkte Zerlegung des Stichprobenraums $\Omega = \cup_i U_i$ eines beliebigen Modells (Ω, \mathcal{E}, P) . Dann gilt für jede Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ und jede glatte Funktion $f : \mathbb{R} \mapsto \mathbb{R}$:

$$E(f(X)) = \sum_{i=1}^n E(f(X) | U_i) P(U_i) .$$

Beweis: Die Formel von der totalen W'keit 1.31 liefert:

$$F_X(r) = P(X \leq r) = \sum_{i=1}^n P(X \leq r | U_i) P(U_i) = \sum_{i=1}^n F_X(r | U_i) P(U_i) .$$

Aus der Definition über Stieltjes-Summen 2.18 ist unmittelbar klar, dass Stieltjes-Integrale linear in der Maßfunktion sind. Mit $\alpha, \beta \in \mathbb{R}$ gilt für $F = \alpha F_1 + \beta F_2$:

$$\int_{(a,b]} f(r) dF(r) = \alpha \int_{(a,b]} f(r) dF_1(r) + \beta \int_{(a,b]} f(r) dF_2(r) .$$

Mit der Summendarstellung von $F_X(r)$ folgt unmittelbar die Behauptung.

Q.E.D.

Diese Formel eignet sich bei Hintereinanderschaltung von Experimenten. Ein erstes Experiment liefert das Ursachensystem U_i für ein zweites Experiment:

Beispiel 2.60: Gegeben sei ein Topf mit 70 fairen Münzen und 30 Münzen mit 2 Köpfen. Man entnimmt dem Topf eine Münze und wirft 1000 mal. Wieviele Köpfe wird man im Mittel werfen?

Lösung: Seien $U_1 =$ „ich habe eine faire Münze gezogen“, $U_2 =$ „ich habe eine unfaire Münze gezogen“. Nach der Auswahl erfolgt die $n = 1000$ -fache Wiederholung eines Bernoulli-Experiments mit dem Erwartungswert $n \cdot P(\text{„Kopf“})$. Für $X =$ „die Anzahl der geworfenen Köpfe bei 1000 Würfeln“ gilt

$$E(X | U_1) = 500, \quad E(X | U_2) = 1000, \quad P(U_1) = \frac{70}{100}, \quad P(U_2) = \frac{30}{100}$$

und damit

$$E(X) = E(X | U_1) P(U_1) + E(X | U_2) P(U_2) = \frac{70}{100} \cdot 500 + \frac{30}{100} \cdot 1000 = 650.$$

Kapitel 3

Anwendungen: Laufzeitanalyse von Sortieralgorithmen

(siehe Kapitel I, § 9 in U. Krengel: *Einführung in die W'keitstheorie und Statistik*, 5. Auflage, Vieweg, 2000).

Als typische Anwendung des Konzepts von Zufallsvariablen in der Informatik werden in diesem Kapitel die Laufzeiten zweier wohlbekannter Sortieralgorithmen diskutiert: MERGESORT und QUICKSORT. Beides sind Standardalgorithmen, die auch in der Praxis verwendet werden. Es geht um die Frage der "mittleren Laufzeit", d.h., wie groß ist der Erwartungswert der Laufzeit, wenn eine stochastisch gewählte Liste zu sortieren ist?

Bei MERGESORT ergibt sich für Listen der Länge n eine Laufzeit von $\approx cn \log_2(n)$, wobei sich die Laufzeiten im günstigsten/ungünstigsten/mittleren Fall nur durch den Wert der "Konstanten" c unterscheiden. Hier liefert die stochastische Analyse ein nur geringfügig realistischeres Bild als die "best case"- und "worst case"-Analyse: im Mittel gilt $c = 1$.

Bei QUICKSORT stellt sich heraus, dass die stochastische Analyse absolut notwendig ist, um die praktische Brauchbarkeit des Algorithmus zu beurteilen: der "worst case" ist mit der Laufzeit $O(n^2)$ unakzeptabel langsam. Trotzdem ist QUICKSORT ein in der Praxis beliebter und guter Algorithmus: im stochastischen Mittel ist die Laufzeit nur $O(n \ln(n))$.

Gegeben ist das Problem, eine ungeordnete Liste von Objekten "aufsteigend" anzuordnen. Auf den Objekten sei eine Ordnung gegeben, d.h., für jedes Paar a, b kann $a < b$ entschieden werden (z.B., es handelt sich um Zahlen oder um Zeichenketten, die lexikographisch anzuordnen sind). Ohne Beschränkung der Allgemeinheit kann man sich vorstellen, dass die willkürlich angeordneten Zahlen $1, 2, \dots, n$ zu sortieren sind.

Als "Laufzeit" wird hier die Anzahl der Vergleichsoperationen $a < b$ gezählt, die im Laufe der Sortierung durchzuführen sind!

3.1 MERGESORT

Zur Vereinfachung der Analyse wird hier vorausgesetzt, dass die Länge n der zu sortierenden Liste eine Zweierpotenz $n = 2^m$ ist. Das allgemeine Ergebnis stimmt i.W. mit diesem Spezialfall überein.

Die Grundidee ist “divide and conquer”: zerlege eine ungeordnete Liste in 2 gleichgroße Teillisten, sortiere diese getrennt, dann erfolgt ein Verschmelzen (engl.: “merge”) der beiden geordneten Teillisten zu einer geordneten Gesamtliste. Das Verschmelzen geschieht durch folgenden rekursiven Hilfsalgorithmus MERGE, der als Vorbereitung zu diskutieren ist:

Es seien zwei sortierte Listen $L_1 = [a_1, a_2, \dots, a_{k_1}]$ und $L_2 = [b_1, b_2, \dots, b_{k_2}]$ der Längen k_1 bzw. k_2 mit $a_1 < a_2 < \dots$ bzw. $b_1 < b_2 < \dots$ gegeben.

Sei L eine (globale) Ergebnisliste, die die sortierten Elemente aufnehmen soll. Diese sei zunächst leer:

MERGE($L_1 : [a_1, a_2, \dots]$, $L_2 : [b_1, b_2, \dots]$)

- 0) Ist L_1 bzw. L_2 leer, dann füge L_2 bzw. L_1 hinten an L an und RETURN(L).
- 1) Ist $a_1 < b_1$, so entferne a_1 aus L_1 und füge a_1 hinten in die Ergebnisliste L ein. Sonst entferne b_1 aus L_2 und füge b_1 hinten in die Ergebnisliste L ein.
- 2) rekursiver Aufruf $L = \text{MERGE}(L_1, L_2)$ mit den verkürzten Listen L_1 bzw. L_2 und RETURN(L).

Im rekursiven Aufruf MERGE(L_1, L_2) enthält L die bereits angeordneten kleinsten Elemente, L_1 und L_2 enthalten die noch an L anzufügenden verbleibenden größeren Elemente. In jedem MERGE-Schritt wächst L um ein Element, damit ruft sich MERGE maximal $k_1 + k_2$ mal rekursiv auf. Falls es nicht zu einem vorzeitigen Abbruch 0) kommt (falls L_1 oder L_2 leer ist), wird pro Aufruf genau eine Vergleichsoperation $a_1 < b_1$ durchgeführt. Der letzte Aufruf terminiert sicherlich durch 0) und braucht dabei keine Vergleichsoperation. Damit gilt folgende Abschätzung für das Verschmelzen zweier geordneter Listen der Länge k_1, k_2 :

Es werden maximal $k_1 + k_2 - 1$ Vergleichsoperationen gebraucht. Andererseits werden mindestens $\min(k_1, k_2)$ Vergleichsoperationen gebraucht, denn die ersten Elemente müssen jeweils verglichen werden, bis eine Liste leer ist.

Allgemeiner gilt: seien $c_1 < c_2 < \dots < c_z$ die z größten Elemente aller beteiligten Objekte $\{a_1, \dots, a_{k_1}, b_1, \dots, b_{k_2}\}$. Enthält $L_1 = [a_1, a_2, \dots, a_{k_1-z}, c_1, \dots, c_z]$ diese Elemente, so ist nach höchstens $k_1 - z + k_2$ Schritten die Liste L_2 leer, denn die Elemente c_1, c_2 etc. werden mit Sicherheit nicht aus L_1 entfernt, solange L_2 noch nicht leer ist. Ist L_2 leer, so terminiert MERGE sofort. Analog ist L_1 nach höchstens $k_1 + k_2 - z$ Schritten leer, wenn L_2 die Elemente c_1, \dots, c_z enthält.

Im Folgenden interessiert nur der Spezialfall $k_1 = k_2 =: k$. Zusammenfassend gilt:

MERGE braucht zum Verschmelzen zweier geordneter Listen mit jeweils k Elementen zu einer geordneten Liste mit $2k$ Elementen minimal k und maximal $2k - 1$ Vergleichsoperationen. Enthält eine der Listen die z größten aller beteiligten Elemente, so braucht MERGE maximal $2k - z$ Vergleichsoperationen.

Wir brauchen später folgende Umformulierung (ersetze z durch $2k - z$):

MERGE braucht beim Verschmelzen zweier geordneter Listen der Länge k maximal z Vergleichsoperationen, wenn die $2k - z$ größten aller beteiligten Elemente alle in einer der beiden Listen sitzen.

Nun zur eigentlichen (rekursiven) Sortierung MERGESORT, in der MERGE als Hilfsfunktion eingesetzt wird. Gegeben sei eine ungeordnete Liste L der Länge $n = 2^m$, die rekursiv immer weiter halbiert werden kann, bis Teillisten mit jeweils einem Element entstehen:

MERGESORT($L : [a_1, \dots, a_n]$)

- 0) Hat L nur ein Element, dann RETURN(L).
- 1) Zerlege L in zwei gleichgroße Teillisten:

$$L_1 = [a_1, \dots, a_{\frac{n}{2}}], \quad L_2 = [a_{\frac{n}{2}+1}, \dots, a_n].$$
- 2) Sortiere die Teillisten durch einen rekursiven Aufruf:

$$S_1 = \text{MERGESORT}(L_1), \quad S_2 = \text{MERGESORT}(L_2).$$
- 3) Verschmelze die sortierten Teillisten: $S = \text{MERGE}(S_1, S_2)$.
- 4) RETURN(S).

Sei M_k die Anzahl der Vergleichsoperationen, die MERGESORT beim Sortieren einer Liste der Länge k benötigt. Sei Z_{2k} die Anzahl der Vergleichsoperationen, die MERGE beim Verschmelzen zweier geordneter Listen der Länge k zu einer geordneten Liste der Länge $2k$ benötigt. Der rekursiven Arbeitsweise von

MERGESORT entsprechend gilt:

$$\begin{aligned}
 M_n &= 2 M_{\frac{n}{2}} + Z_n = 2(2 M_{\frac{n}{4}} + Z_{\frac{n}{2}}) + Z_n \\
 &= 4 M_{\frac{n}{4}} + 2 Z_{\frac{n}{2}} + Z_n = 4(2 M_{\frac{n}{8}} + Z_{\frac{n}{4}}) + 2 Z_{\frac{n}{2}} + Z_n \\
 &= 8 M_{\frac{n}{8}} + 4 Z_{\frac{n}{4}} + 2 Z_{\frac{n}{2}} + Z_n = \dots \\
 &\dots \\
 &= n M_1 + \frac{n}{2} Z_2 + \frac{n}{4} Z_4 + \dots + 2 Z_{\frac{n}{2}} + Z_n .
 \end{aligned}$$

Wird die Eingabeliste stochastisch gewählt, sind M_k und Z_k Zufallsvariable. Mit $M_1 = 0$ folgt der fragte Erwartungswert

$$E(M_n) = \frac{n}{2} E(Z_2) + \frac{n}{4} E(Z_4) + \dots + 2 E(Z_{\frac{n}{2}}) + E(Z_n) . \quad (\#)$$

Es gilt damit, den Erwartungswert der Z_k ($k = 2, 4, 8, \dots, n$) zu bestimmen.

Betrachte die Situation, wenn MERGESORT mit einer ungeordneten Teilliste $L = [a_1, \dots, a_{2k}]$ der Länge $2k$ aufgerufen wird. Diese Liste wird in zwei ungeordnete Teillisten $L_1 = [a_1, \dots, a_k]$ und $L_2 = [a_{k+1}, \dots, a_{2k}]$ aufgespalten. Der rekursive Aufruf $S_1 = \text{MERGESORT}(L_1)$, $S_2 = \text{MERGESORT}(L_2)$ liefert geordnete Listen S_1, S_2 der Länge k . Es geht um die Zufallsvariable

$$Z_{2k} = \text{Kosten von MERGE}(S_1, S_2) .$$

Die ungeordnete Liste $L = [a_1, \dots, a_{2k}]$ kann als eine Teilmenge $\{a_1, \dots, a_{2k}\}$ aller Elemente der Eingangsliste angesehen werden. Der Zufallsmechanismus kann darauf reduziert werden, L zufällig in zwei gleichgroße Teilmengen L_1, L_2 mit jeweils k Elementen aufzuspalten. Beachte, dass in der Tat die Reihenfolge der Elemente in den Teillisten L_1, L_2 irrelevant für die anfallenden Kosten Z_{2k} sind, denn jede Umordnung innerhalb von L_1 bzw. L_2 führt zu den selben sortierten Listen S_1 bzw. S_2 . Damit entspricht eine unsortierte Liste einer Menge.

Nach Hilfssatz 1.9.c) gibt es $\binom{2k}{k} = \frac{(2k)!}{k! k!}$ Möglichkeiten, L (als Menge mit $2k$ Elementen) in zwei Teilmengen L_1, L_2 mit jeweils k Elementen aufzuteilen.

Nach den Vorüberlegungen zu MERGE nimmt die mögliche Anzahl der Vergleichsoperationen Z_{2k} beim MERGE zweier Listen der Länge k mögliche Werte zwischen k und $2k - 1$ an. Sei z eine ganze Zahl zwischen k und $2k - 1$. Wir fragen: wieviele der $\binom{2k}{k}$ Möglichkeiten der Aufteilung von L in L_1 und L_2 sind so, dass das Ereignis " $Z_{2k} \leq z$ " eintritt?

Nach den Vorüberlegungen zu MERGE zerlegt sich das Ereignis " $Z_{2k} \leq z$ " in die beiden disjunkten Fälle

- 1) die $2k - z$ größten Elemente von L wandern alle nach L_1 ,
- 2) die $2k - z$ größten Elemente von L wandern alle nach L_2 .

In den $\binom{2k}{k}$ Aufteilung der Menge L in die Teilmengen L_1 und L_2 gibt es $\binom{z}{k}$ Möglichkeiten, die 1) entsprechen. (Die $2k - z$ größten Elemente gehen in L_1 . Es verbleiben $2k - (2k - z) = z$ Elemente von L aufzuteilen in die verbleibenden $k - (2k - z) = z - k$ Elemente von L_1 und k Elemente, die in L_2 wandern.) Analog gibt es $\binom{z}{k}$ Möglichkeiten, die 2) entsprechen. Damit folgt

$$P(Z_{2k} \leq z) = 2 \frac{\binom{z}{k}}{\binom{2k}{k}} = \frac{2 z!}{k! (z-k)!} \frac{k! k!}{(2k)!} = \frac{2 z! k!}{(z-k)! (2k)!}$$

und hieraus

$$\begin{aligned} P(Z_{2k} = z) &= P(Z_{2k} \leq z) - P(Z_{2k} \leq z-1) \\ &= \frac{2 z! k!}{(z-k)! (2k)!} - \frac{2 (z-1)! k!}{(z-1-k)! (2k)!} \\ &= \frac{2 (z-1)! k!}{(z-1-k)! (2k)!} \left(\frac{z}{z-k} - 1 \right) \\ &= \frac{2 (z-1)! k!}{(z-1-k)! (2k)!} \frac{k}{z-k} \\ &= \frac{2 (z-1)! k! k}{(z-k)! (2k)!} . \end{aligned}$$

Mit den möglichen Werten $z = k, k+1, \dots, 2k-1$ ergibt sich als Erwartungswert

$$\begin{aligned} E(Z_{2k}) &= \sum_{z=k}^{2k-1} z P(Z_{2k} = z) = \sum_{z=k}^{2k-1} \frac{2 z! k! k}{(z-k)! (2k)!} \\ &= \frac{2 k! k}{(2k)!} \sum_{z=k}^{2k-1} \frac{z!}{(z-k)!} = \frac{2 k! k}{(2k)!} \sum_{z=0}^{k-1} \frac{(z+k)!}{z!} \stackrel{(*)}{=} \boxed{\frac{2 k^2}{k+1}} . \end{aligned}$$

Hierbei wird in (*) die Identität $\sum_{z=0}^{k-1} (z+k)!/z! = k(2k)!/(k+1)!$ verwendet (z.B. durch Induktion nach k zu beweisen). Die Formel (#) liefert nun als Endergebnis den gefragten Erwartungswert für die Kosten von MERGESORT einer Liste mit $n = 2^m$ Elementen:

$$\begin{aligned} E(M_n) &= \frac{n}{2} E(Z_2) + \frac{n}{4} E(Z_4) + \dots + 2 E(Z_{\frac{n}{2}}) + E(Z_n) \\ &= \sum_{i=1}^m \frac{2^m}{2^i} E(Z_{2^i}) = 2^m \sum_{i=1}^m 2^{-i} E(Z_{2 \cdot 2^{i-1}}) = 2^m \sum_{i=1}^m 2^{-i} \frac{2(2^{i-1})^2}{2^{i-1} + 1} \end{aligned}$$

$$\begin{aligned}
&= 2^m \sum_{i=1}^m \frac{2^{-i} 2^{2i-2}}{2^{i-1} + 1} = 2^m \sum_{i=1}^m \frac{2^{i-1}}{2^{i-1} + 1} = 2^m \sum_{i=1}^m \frac{2^{i-1} + 1 - 1}{2^{i-1} + 1} \\
&= 2^m \left(\sum_{i=1}^m \frac{2^{i-1} + 1}{2^{i-1} + 1} - \sum_{i=1}^m \frac{1}{2^{i-1} + 1} \right) = 2^m \left(m - \sum_{i=1}^m \frac{1}{2^{i-1} + 1} \right).
\end{aligned}$$

Die Summe $\sum_{i=1}^m 1/(2^{i-1}+1)$ konvergiert mit wachsendem m schnell gegen einen Grenzwert mit dem numerischen Wert 1.264.... Mit $2^m = n$ und $m = \log_2(n)$ bekommt man als mittlere Laufzeit von MERGESORT auf Listen der Länge n :

$$E(M_n) \approx n (\log_2(n) - 1.264\dots).$$

Zum Vergleich noch die Ergebnisse für den “best case” und den “worst case”. Nach den Vorüberlegungen zu MERGE gilt $k \leq Z_{2k} \leq 2k - 1$. Setzt man jeweils die unteren und oberen Grenzen in

$$M_n = \frac{n}{2} Z_2 + \frac{n}{4} Z_4 + \dots + 2 Z_{\frac{n}{2}} + Z_n = \sum_{i=1}^m \frac{2^m}{2^i} Z_{2^i}$$

ein, so ergibt sich bestenfalls die Laufzeit

$$M_n \geq \sum_{i=1}^m \frac{2^m}{2^i} 2^{i-1} = \frac{2^m m}{2} = \frac{1}{2} n \log_2(n)$$

und schlechtestenfalls

$$M_n \leq \sum_{i=1}^m \frac{2^m}{2^i} (2^i - 1) = 2^m \left(m - 1 + \frac{1}{2^m} \right) = n \left(\log_2(n) - 1 + \frac{1}{n} \right).$$

Die mittlere Laufzeit liegt (für großes n) recht dicht an der schlechtesten Laufzeit und ist damit etwa doppelt so hoch wie die bestmögliche Laufzeit. All dies ist jedoch $O(n \log_2(n))$ und damit in jedem Fall eine sehr schnelle Sortierung.

3.2 QUICKSORT

Eine Variante von QUICKSORT wird z.B. im `sort`-Kommando des UNIX-Betriebssystems verwendet. Es handelt sich um einen leicht zu implementierenden und “im Mittel” schnellen $O(n \ln(n))$ -Algorithmus, der jedoch sehr seltene “worst case”-Ausreißer mit der Laufzeit $O(n^2)$ aufweist.

Der QUICKSORT-Algorithmus sortiert ein Liste $L = [a_1, \dots, a_n]$ der Länge n rekursiv:

QUICKSORT($L : [a_1, \dots, a_n]$)

- 0) Ist L leer oder hat L nur ein Element, dann RETURN(L).
- 1) Wähle zufällig einen (gleichverteilten) Index $k \in \{1, \dots, n\}$.
- 2) Vergleiche alle Elemente in L mit dem k -ten Element $L[k]$. Sei $L^<$ die Menge aller Listenelemente in L , die kleiner sind als $L[k]$. Sei $L^>$ die Menge aller Listenelemente in L , die größer sind als $L[k]$.
- 3) Stelle die Mengen $L^<$ und $L^>$ als unsortierte Listen dar und sortiere sie durch einen rekursiven Aufruf:

$$S^< = \text{QUICKSORT}(L^<) , \quad S^> = \text{QUICKSORT}(L^>) .$$

- 4) Verkette die sortierten Teillisten:

$$S = [\text{Elemente von } S^<, L[k], \text{Elemente von } S^>].$$

- 5) RETURN(S).

Sei Q_n die Anzahl der Vergleichsoperationen, die QUICKSORT zum Sortieren eine Liste der Länge n benötigt. Dieser Wert wird wiederum als die Laufzeit interpretiert. Die Kosten fallen im Schritt 2) an, wo $n - 1$ Vergleichsoperationen gebraucht werden, um $L[k]$ mit allen anderen Listenelementen zu vergleichen.

Der “worst case” tritt ein, wenn in jedem rekursiven Schritt entweder $L^<$ oder $L^>$ leer sind, d.h., es wird zufällig das jeweils kleinste oder größte Element $L[k]$ gewählt. In diesem Fall addieren sich die in 2) anfallenden Vergleiche zu

$$Q_n = (n - 1) + (n - 2) + \dots + 1 = \frac{n(n - 1)}{2}$$

auf, die “worst case”-Laufzeit ist also $O(n^2)$. Dies sieht nicht sehr attraktiv aus. Glücklicherweise ist das Eintreten des “worst case”-Falls sehr unwahrscheinlich. Ein realistischeres (und wesentlich positiveres) Bild liefert eine genauere stochastische Analyse. Wir betrachten dazu eine fixierte Eingangsliste und berücksichtigen nur den in Schritt 1) eingebauten Zufallsmechanismus. Das Ergebnis der Analyse gilt unverändert auch dann, wenn man als weiteren Zufallsmechanismus die Eingabeliste zufällig variiert (siehe Krenzel für Details).

Sei $Z_n - 1$ die Anzahl der Elemente von $L^<$, d.h., Z_n ist die Position, die das zufällig gewählte Element $L[k]$ in der fertig sortierten Liste hat (also: die Zufallsvariable Z_n kann Werte zwischen 1 und n annehmen). Die Liste $L^>$ hat

dann $n - Z_n$ Elemente. Mit den $n - 1$ Vergleichen in 2) und den rekursiven Aufrufen in 3) gilt:

$$Q_n = n - 1 + Q_{Z_n-1} + Q_{n-Z_n} ,$$

wobei Z_n zufällige Zahlen zwischen 1 und n darstellt. Für die Erwartungswerte folgt

$$E(Q_n) = n - 1 + E(Q_{Z_n-1}) + E(Q_{n-Z_n}) .$$

Wir müssen zunächst die Verteilung von Z_n bestimmen. Sei dazu K die Zufallsvariablen, die den in 1) zufällig gewählten Index k liefert. Beachte, dass bei einer gegebenen unsortierten Liste $L = [a_1, \dots, a_n]$ jeder Wert k eineindeutig zu einem Wert z von Z_n gehört. Damit ist Z_n genauso wie K auf $\{1, \dots, n\}$ gleichverteilt¹. Mit Satz 2.59 und dem Ursachensystem $U_z = "Z_n = z"$, $z = 1, \dots, n$, ergibt sich

$$\begin{aligned} E(Q_n) &= n - 1 + \sum_{z=1}^n E(Q_{Z_n-1} \mid Z_n = z) P(Z_n = z) \\ &\quad + \sum_{z=1}^n E(Q_{n-Z_n} \mid Z_n = z) P(Z_n = z) . \end{aligned}$$

Für gegebenes $Z_n = z$ stimmen die Zufallsvariablen Q_{Z_n-1} bzw. Q_{n-Z_n} aber mit Q_{z-1} bzw. Q_{n-z} überein, und mit $P(Z_n = z) = 1/n$ folgt die Rekursion

$$E(Q_n) = n - 1 + \frac{1}{n} \sum_{z=1}^n E(Q_{z-1}) + \frac{1}{n} \sum_{z=1}^n E(Q_{n-z}) .$$

Nach Umsummierung $z \rightarrow n - z + 1$ der zweiten Summe sieht man, dass sie mit der ersten Summe übereinstimmt. Es folgt

$$E(Q_n) = n - 1 + \frac{2}{n} \sum_{z=1}^n E(Q_{z-1}) = n - 1 + \frac{2}{n} \sum_{i=1}^{n-1} E(Q_i)$$

(beachte $E(Q_0) = 0$ bei der letzten Umsummation). Die Modellierung ist hiermit abgeschlossen, bis auf verbleibende mathematische Arbeit ist $q_n := E(Q_n)$ eindeutig durch diese Rekursion bestimmt: wir kennen $q_0 = q_1 = 0$, hieraus folgt q_2 , daraus q_3 etc. Es verbleibt lediglich, die obige Rekursion in eine geschlossene Form zu bringen. Zunächst folgt aus

$$n q_n = n(n - 1) + 2 \sum_{i=1}^{n-1} q_i$$

¹Formal: sei $z = \pi(k)$ der Wert von Z_n , der dem Wert k von K entspricht (π ist eine Permutation von $\{1, \dots, n\}$). Damit stimmen die Ereignisse " $K = k$ " und " $Z_n = \pi(k)$ " überein und haben für alle Werte von k (und damit für alle Werte von $z = \pi(k)$) dieselbe W'keit $1/n$.

und ($n \rightarrow n - 1$)

$$(n - 1) q_{n-1} = (n - 1)(n - 2) + 2 \sum_{i=1}^{n-2} q_i$$

durch Differenzbildung

$$n q_n - (n - 1) q_{n-1} = 2(n - 1) + 2 q_{n-1} ,$$

also

$$n q_n = 2(n - 1) + (n + 1) q_{n-1} .$$

Wir stellen q_n in der Form $q_n = 2 + (n + 1) b_n$ dar. Einsetzen liefert eine einfachere Rekursion für b_n :

$$\begin{aligned} n(2 + (n + 1) b_n) &= 2(n - 1) + (n + 1)(2 + n b_{n-1}) \\ \Rightarrow 2n + n(n + 1) b_n &= 2n - 2 + 2n + 2 + n(n + 1) b_{n-1} \\ \Rightarrow n(n + 1) b_n &= 2n + n(n + 1) b_{n-1} \\ \Rightarrow b_n &= b_{n-1} + \frac{2}{n + 1} . \end{aligned}$$

Diese Rekursion ist sofort gelöst:

$$\begin{aligned} b_n &= \frac{2}{n + 1} + b_{n-1} = \frac{2}{n + 1} + \frac{2}{n} + b_{n-2} = \frac{2}{n + 1} + \frac{2}{n} + \frac{2}{n - 1} + b_{n-3} \\ &= \dots = \frac{2}{n + 1} + \frac{2}{n} + \dots + \frac{2}{2} + b_0 = 2 \sum_{i=1}^{n+1} \frac{1}{i} - 2 + b_0 . \end{aligned}$$

Mit $q_0 = 0$ und dementsprechend $b_0 = -2$ folgt $b_n = -4 + 2 \sum_{i=1}^{n+1} \frac{1}{i}$ und hiermit

$$\begin{aligned} q_n = 2 + (n + 1) b_n &= 2 - 4(n + 1) + 2(n + 1) \sum_{i=1}^{n+1} \frac{1}{i} \\ &= 2 + 2(n + 1) \left(-2 + \sum_{i=1}^{n+1} \frac{1}{i} \right) . \end{aligned}$$

Die verbleibende Summe ist wohlbekannt, es gilt

$$\lim_{n \rightarrow \infty} \left(\sum_{i=1}^{n+1} \frac{1}{i} - \ln(n + 1) \right) = C = 0.5772\dots$$

(C ist die sogenannte Euler-Konstante).

Ergebnis: der Erwartungswert der Laufzeit von QUICKSORT für (große) Listen der Länge n ist

$$q_n = \mathbb{E}(Q_n) \approx 2(n+1) \left(\ln(n+1) + C - 2 \right) + 2,$$

also von der Größenordnung $2n \ln(n) \approx 1.39 n \log_2(n)$. Die mittlere Laufzeit ist damit dramatisch kleiner als die “worst case”-Laufzeit $n(n-1)/2$.

Kapitel 4

Grenzwertsätze

Dieses Kapitel ist gewissermaßen der Höhepunkt der bisher entwickelten Theorie und bietet gleichzeitig einen Übergang zum Einsatz der W'keitstheorie in der Praxis. Bislang waren wir immer von Modellvorgaben ausgegangen: "Sei X so-und-so-verteilt mit dem/der/den Erwartungswert/Streuung/Eigenschaften etc." In der praktischen Anwendung ist es kaum realistisch, von konkreten Verteilungen mit bekannten Parametern auszugehen. Stattdessen wird man eher einen Satz von Messdaten vorliegen haben, die stochastisch zu interpretieren sind (Statistik). Die folgenden Grenzwertsätze sind hilfreich, statistische Daten in das Gerüst der W'keitstheorie einzuordnen.

4.1 Das (schwache) Gesetz der großen Zahl

Gegeben sei eine unfaire Münze. Um $P(\text{„Kopf“})$ zu bestimmen, wird man sie häufig werfen und dann

$$P(\text{„Kopf“}) \approx \frac{\text{Anzahl der geworfenen Köpfe}}{\text{Anzahl aller Würfe}}$$

setzen. Wieso (bzw., in welchem Sinne) liefert dies eine vernünftige Schätzung für $P(\text{„Kopf“})$?

Satz 4.1: Das (schwache) Gesetz der großen Zahl

Seien X_1, \dots, X_n unabhängige Zufallsvariable, alle mit dem selben Erwartungswert $\mu = E(X_1) = \dots = E(X_n)$ und der selben Streuung $\sigma = \sigma(X_1) = \dots = \sigma(X_n)$. Sei

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

die "Mittelwert"-Variable. Für jedes $\epsilon > 0$ gilt

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n},$$

bzw., äquivalenterweise

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2 n} .$$

Bew: Wegen der Unabhängigkeit ist nicht nur der Erwartungswert, sondern auch die Varianz linear (siehe Abschnitt 2.5.2):

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu ,$$

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n} .$$

Die Chebyshev-Ungleichung 2.36 liefert sofort:

$$P(|\bar{X}_n - E(\bar{X}_n)| \geq \epsilon) = P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n} .$$

Q.E.D.

25.6.57↓

Interpretation 1, 4.2:

Man betrachte den Spezialfall, dass alle Variablen X_1, \dots, X_n die selbe Verteilung haben, also das selbe Zufallsexperiment darstellen: es geht also um n unabhängige Wiederholungen des selben Experiments.

Die Variablen X_1, \dots, X_n werden als unabhängige Wiederholungen ein und des selben Zufallsexperiments X interpretiert. Die Mittelwertvariable \bar{X}_n entspricht dann dem Zufallsexperiment: „führe n unabhängige Einzelmessungen von X durch und betrachte den Mittelwert“.

Zu jedem noch so kleinen vorgegebenen $\epsilon > 0$ wird die Wahrscheinlichkeit, dass der Mittelwert um mehr als ϵ vom (in der Regel unbekanntem) Erwartungswert des Einzelexperiments X abweicht, mit wachsendem n immer kleiner (also, je öfter man das Einzelexperiment wiederholt). Komplementär: die Wahrscheinlichkeit, dass der Mittelwert bis auf ein ϵ den Erwartungswert liefert, liegt bei genügend großem n praktisch bei 1.

In dieser Interpretation liefert das Gesetz der großen Zahl überhaupt erst die Grundlage, die mathematische W'keitstheorie auf praktische Fragestellungen anwenden zu können (Statistik). Die Parameter einer Verteilung wie z.B. der Erwartungswert sind in praktischen Experimenten unbekannt, man muss sie durch Messungen abschätzen:

Man kann den Erwartungswert $\mu = E(X)$ eines Einzelexperiments X durch empirische Mittelwerte \bar{X}_n messen! Wenn die “Stichprobengröße” n nur groß genug ist, d.h., wird nur genügend oft wiederholt, so approximiert der Mittelwert mit großer W'keit den Erwartungswert. Für jedes feste $\epsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1 .$$

Umgekehrt liefert diese Aussage die Interpretation des Erwartungswertes als Kennzeichen einer Zufallsvariable, der ja in Definition 2.22 als rein mathematisches Objekt eingeführt wurde:

Der Erwartungswert einer Zufallsvariable entspricht dem Wert, den man nach häufiger unabhängiger Wiederholung des Experiments durch Mittelung erhält.

Daher der Sprachgebrauch 2.23:

„ X nimmt im Mittel den Erwartungswert $E(X)$ an“.

Interpretation 2, 4.3:

Wir lassen nun einen gewissen Paradigmen-Wechsel zu. Wie interpretiert man eine W'keitsaussage wie

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2 n}$$

in einer praktischen Anwendung? Bislang war unsere Interpretation: sei das Modell (der Stichprobenraum, die Verteilung, die Parameter wie z.B. der Erwartungswert μ etc.) vorgegeben. Dann wurde die obige W'keit folgendermaßen interpretiert:

Führen wir (in Zukunft) ein reales Zufallsexperiment durch, so werden wir dabei mit der berechneten bzw. abgeschätzten W'keit konkrete Werte $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ ermitteln, deren Mittelwert $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ bis auf eine absolute Genauigkeit ϵ Werte in der Nähe des (vorgegebenen) Erwartungswerts μ annimmt.

D.h., eine praktische Interpretation beinhaltet, Zufallsvariable X_i (mathematisch ein “Abstraktum”: Funktionen vom Stichprobenraum nach \mathbb{R}) durch konkrete Zahlen x_i (“Messwerte eines tatsächlich durchgeführten Zufallsexperiments”) zu ersetzen.

Von nun an erlauben wir auch folgende Interpretation: Das tatsächliche Zufallsexperiment sei bereits durchgeführt, es liegen Zahlenwerte (“Messungen”) $X_1 = x_1, \dots, X_n = x_n$ mit dem Mittelwert $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ vor. Wo bleibt nun in der Aussage

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2 n}$$

das intuitive Konzept von “W’keit”, wenn \bar{X}_n einen gemessenen Zahlenwert \bar{x}_n angenommen hat? Nun, man mache sich klar, was ein Modell (im Sinne der philosophischen Erkenntnistheorie) prinzipiell ist: es kann nicht (und soll nicht) ein mathematisches Abbild einer “absoluten Wahrheit” sein. Konkret heißt dies hier, weder konkrete Verteilungen noch Parameter sind in praktischen Anwendungen wirklich bekannt.

Ersetzen wir in $P(|\bar{X}_n - \mu| < \epsilon)$ die Zufallsvariable \bar{X}_n durch einen Messwert \bar{x}_n , so interpretieren wir

$$P(|\bar{x}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2 n}$$

folgendermaßen:

$P(|\bar{x}_n - \mu| < \epsilon)$ gibt die „**Sicherheit**“ an, dass der (unbekannte) Erwartungswert μ bis auf eine absolute Genauigkeit ϵ in der Nähe des gemessenen Mittelwerts \bar{x}_n liegt.

Diese Interpretation lässt sich nicht sauber in das strenge Konzept der W’keitstheorie einordnen (dazu müssten wir μ in irgendeinem Sinne als eine Zufallsvariable interpretieren). Daher sprechen wir bei $P(|\bar{x}_n - \mu| < \epsilon)$ nicht von einer W’keit, sondern von einer (nur intuitiv zu interpretierenden) Sicherheit.

In der Tat ist die obige Interpretation ein arg einfaches Weltbild, das zu Proteststürmen ausgemachter Stochastiker führen dürfte, die in der „Schätztheorie“ eine saubere Interpretation im Sinne des W’keitsmaßes der unterliegenden Modelle benutzen.

Das Intervall $[\bar{x}_n - \epsilon, \bar{x}_n + \epsilon]$ wird auch **Konfidenzintervall** für den Parameter μ zur Sicherheit $P(|\bar{X}_n - \mu| < \epsilon)$ genannt.

Bemerkung 4.4: Es gibt auch ein sogenanntes starkes Gesetz der großen Zahl. Dies ist die (stärkere) Aussage

$$P(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| = 0) = 1,$$

wobei mit $P(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| = 0)$ gemeint ist $P(\{\omega \in \Omega; \lim_{n \rightarrow \infty} \bar{X}_n(\omega) = \mu\})$. Siehe Krengel, Kapitel §12.

Beispiel 4.5: Zurück zum Problem, für eine unfaire Münze $P(\text{„Kopf“})$ durch empirische Messung zu bestimmen. Allgemein, sei ein Bernoulli-Experiment $X : \Omega \mapsto \{0, 1\}$ mit unbekannter Erfolgsw'keit $p = P(X = 1)$ (und $q = P(X = 0) = 1 - p$) gegeben. Nach Beispiel 2.38 (mit $n = 1$) gilt

$$E(X) = p, \quad \text{Var}(X) = pq.$$

Seien X_1, \dots, X_n unabhängige Wiederholungen von X . Damit hat der Mittelwert die Interpretation

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{\text{Anzahl der Erfolge}}{\text{Anzahl der Versuche}}.$$

Das Gesetz der großen Zahl liefert für jedes $\epsilon > 0$:

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon) \geq 1 - \frac{pq}{n\epsilon^2} \geq 1 - \frac{1}{4n\epsilon^2}.$$

(Beachte: für jedes Bernoulli-Experiment gilt $\text{Var}(X) = pq = p(1-p) \leq 1/4$. Dies ist hier hilfreich, da p ja nicht bekannt ist, sondern erst noch abgeschätzt werden soll.)

Angenommen, eine unfaire Münze wurde $n = 1000$ mal geworfen, worden, es sei dabei 600 mal „Kopf“ eingetreten, also: das Ereignis $\bar{X}_n = \bar{x}_n = 0.6$ ist eingetreten. Mit verschiedenen Werten von ϵ ergeben sich die Sicherheiten

$$\begin{aligned} \epsilon = \frac{1}{10} : P(0.6 - 0.1 < p < 0.6 + 0.1) &\geq 1 - \frac{1}{4 \cdot 1000 \cdot (1/10)^2} = 0.975, \\ \epsilon = \frac{1}{20} : P(0.6 - 0.05 < p < 0.6 + 0.05) &\geq 1 - \frac{1}{4 \cdot 1000 \cdot (1/20)^2} = 0.9, \\ \epsilon = \frac{1}{50} : P(0.6 - 0.02 < p < 0.6 + 0.02) &\geq 1 - \frac{1}{4 \cdot 1000 \cdot (1/50)^2} = 0.375. \end{aligned}$$

Also: nach dem Experiment kann ich mit mindestens 97.5%-iger Sicherheit sagen, dass p im Intervall $(0.5, 0.7)$ liegt. Mit mindestens 90%-iger Sicherheit liegt p im Intervall $(0.55, 0.65)$. Mit mindesten 37.5%-iger Sicherheit liegt p im Intervall $(0.58, 0.62)$.

Anmerkung: eine sehr ähnliche Diskussion haben wir schon einmal in Beispiel 2.39 geführt. Die Variable $S_n = n\bar{X}_n = \sum_i X_i$ = „Anzahl der Erfolge eines Bernoulli-Experiments“ ist bekannterweise binomial-verteilt, also

$$\begin{aligned} P(|\bar{X}_n - p| < \epsilon) &= P(|n\bar{X}_n - np| < n\epsilon) = P(|S_n - E(S_n)| < n\epsilon) \\ &= \sum_{\substack{k=0, \dots, n \\ |k - np| < n\epsilon}} P(S_n = k) = \sum_{\substack{k > n(p-\epsilon) \\ k < n(p+\epsilon)}} \binom{n}{k} p^k q^{n-k}. \end{aligned}$$

Da wir p nicht kennen, können wir diese W'keiten zwar nicht exakt berechnen, aber

mit $p \approx 0.6$, $n = 1000$ grob abschätzen:

$$P\left(|\bar{X}_n - p| < \frac{1}{10}\right) \approx \sum_{k=n(0.6-0.1)+1}^{n(0.6+0.1)-1} \binom{n}{k} 0.6^k 0.4^{n-k} \approx 1 - \frac{1.33}{10^{10}},$$

$$P\left(|\bar{X}_n - p| < \frac{1}{20}\right) \approx \sum_{k=n(0.6-0.05)+1}^{n(0.6+0.05)-1} \binom{n}{k} 0.6^k 0.4^{n-k} \approx 0.9986,$$

$$P\left(|\bar{X}_n - p| < \frac{1}{50}\right) \approx \sum_{k=n(0.6-0.02)+1}^{n(0.6+0.02)-1} \binom{n}{k} 0.6^k 0.4^{n-k} \approx 0.792.$$

Die vom Gesetz der großen Zahl gelieferten Sicherheiten sind also arg untertrieben (kein Wunder, denn das Gesetz der großen Zahl beruht auf Chebyshev 2.36).

Beispiel 4.6: Bei Würfeln mit einem evtl. manipulierten Würfel soll die W'keit p , eine „Sechs“ zu würfeln, empirisch ermittelt werden. Wie oft muss man werfen, um p mit einer Sicherheit von mindestens 97.5% auf eine absolute Genauigkeit von $\epsilon = 1/1000$ festlegen zu können?

Sei \bar{X}_n wieder die Mittelwertvariable $\frac{1}{n} \times$ „Anzahl der geworfenen Sechsen bei n Würfeln“ (Erwartungswert $E(\bar{X}_n) = p$):

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon) \geq 0.975 =: S.$$

Gemäß Beispiel 4.5 gilt

$$P(\bar{X}_n - \epsilon < p < \bar{X}_n + \epsilon) \geq 1 - \frac{pq}{n\epsilon^2} \geq 1 - \frac{1}{4n\epsilon^2}.$$

Also muss n so groß sein, dass $1 - 1/(4n\epsilon^2) \geq S$ gilt, also

$$n \geq \frac{1}{4(1-S)\epsilon^2} = \frac{10^6}{4 \cdot 0.025} = 10\,000\,000.$$

Dies ist wiederum eine zwar sichere, aber ziemlich grobe Abschätzung. Setzt man analog zu Beispiel 4.5 die Approximation $p \approx 1/6$ in die exakte Binomial-Verteilung der Variable $S_n =$ „Anzahl der Sechsen“ ein, so ergibt sich mit $n = 10\,000\,000$ für eine gemessene relative Häufigkeit \bar{x}_n der Sechsen die wesentlich näher bei 1 liegende Sicherheit

$$P\left(|\bar{x}_n - p| < \frac{1}{1000}\right) \approx \sum_{k=\lceil n(1/6-1/1000) \rceil}^{\lfloor n(1/6+1/1000) \rfloor} \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k} \approx 1 - \frac{2.155\dots}{10^{17}}.$$

Probiert man einige Werte von n für $p = 1/6$ mit der obigen exakten Formel aus, so zeigt sich, dass bereits $n \approx 700\,000$ Versuche für die Sicherheit 0.975 reichen:

$$P\left(|\bar{x}_n - p| < \frac{1}{1000}\right) \approx \begin{cases} 0.9624 & \text{für } n = 600\,000, \\ 0.9752 & \text{für } n = 700\,000, \\ 0.9836 & \text{für } n = 800\,000. \end{cases}$$

Wir werden diese Rechnung in Beispiel 4.13 noch einmal aufgreifen.

Allgemein gilt, dass das Gesetz der großen Zahl (also i.W. die Chebyshevsche Ungleichung) zwar sehr wichtig für die Philosophie der W'keitsrechnung ist, für die Praxis aber nur schlechte Abschätzungen liefert (siehe die letzten Beispiele). Der folgende Abschnitt liefert wesentlich genauere Abschätzungen für die Binomial-Verteilung.

4.2 Die Normalverteilung als Grenzwert der Binomial-Verteilung: der Satz von Moivre-Laplace

In Abschnitt 2.4.2 wurde mit der Poisson-Verteilung bereits eine Grenzverteilung der Binomial-Verteilung vorgestellt. Zur Erinnerung: betrachtet man $n \rightarrow \infty$ und gleichzeitig $p \rightarrow 0$, so dass $\lambda = np$ konstant ist, so gilt

$$\lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Wir betrachten hier nun einen anderen Grenzübergang: den Grenzfall $n \rightarrow \infty$, wobei p konstant gehalten wird.

Für binomial-verteiltes S_n (= Anzahl der Erfolge bei bei n Wiederholungen eines Bernoulli-Experiments mit Erfolgsw'keit p) betrachte die relative Erfolgshäufigkeit $\bar{X}_n = S_n/n$ und

$$Y_n = \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} (\bar{X}_n - p).$$

Es gilt¹

$$\begin{aligned} \mathbb{E}(Y_n) &= \frac{1}{\sqrt{npq}} \mathbb{E}(S_n - np) \stackrel{(*)}{=} 0, \\ \text{Var}(Y_n) &\stackrel{(**)}{=} \frac{1}{npq} \text{Var}(S_n - np) \stackrel{(**)}{=} \frac{1}{npq} \text{Var}(S_n) \stackrel{(*)}{=} 1. \end{aligned}$$

Man nennt dieses Y_n auch „zentrierte“ (Erwartungswert 0) und „normierte“ (Streuung 1) relative Erfolgshäufigkeit. Diese ist für großes n normalverteilt:

¹Zu (*) beachte man Beispiel 2.38. Zu (**) beachte man:

$$\begin{aligned} \text{Var}(\alpha X + \beta) &= \mathbb{E}((\alpha X + \beta)^2) - (\mathbb{E}(\alpha X + \beta))^2 \\ &= \alpha^2 \mathbb{E}(X^2) + 2\alpha\beta \mathbb{E}(X) + \beta^2 - (\alpha \mathbb{E}(X) + \beta)^2 = \alpha^2 \mathbb{E}(X^2) - \alpha^2 \mathbb{E}(X)^2 = \alpha^2 \text{Var}(X). \end{aligned}$$

Satz 4.7: (Moivre-Laplace)

Sei $S_n : \Omega \rightarrow \{0, 1, \dots, n\}$ binomial-verteilt:

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p.$$

Für $p \cdot q \neq 0$ gilt:

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

In Worten: für großes n genügt die zentrierte und normierte Erfolgshäufigkeit einer Normalverteilung (mit Erwartungswert 0 und Streuung 1).

Der Approximationsfehler für großes n ist $O(1/\sqrt{n})$:

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Beweisskizze: Elementares Abschätzen der Binomialkoeffizienten $\binom{n}{k}$ über die Stirling-Formel lässt die Exponentialfunktion auftauchen. Die exakte W'keit $P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right)$ ist eine Summe über Binomial-Terme, die nach der Approximation als Riemann-Summe für das Integral interpretiert werden kann. Die Riemann-Summe konvergiert für $n \rightarrow \infty$ gegen das Integral.

Hier die technischen Details:

Es gilt

$$\begin{aligned} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) &= P\left(np + a \sqrt{npq} \leq S_n \leq np + b \sqrt{npq}\right) \\ &= \sum_{k=k_a}^{k_b} \binom{n}{k} p^k q^{n-k}, \quad k_a = \lceil np + a \sqrt{npq} \rceil, \quad k_b = \lfloor np + b \sqrt{npq} \rfloor \end{aligned}$$

abzuschätzen. Die Terme sind damit für Werte von k zu approximieren, für die $|k - np| \leq \text{const} \sqrt{n}$ gilt, also $1/k = O(1/n)$ und $1/(n-k) = O(1/n)$. Mit der Stirling-Formel für Fakultäten (siehe Abschnitt 2.4.2) folgt

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} &= \frac{n! p^k q^{n-k}}{k! (n-k)!} \\ &= \frac{p^k q^{n-k}}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{e}\right)^n \left(\frac{e}{k}\right)^k \left(\frac{e}{n-k}\right)^{n-k} \frac{1 + O\left(\frac{1}{n}\right)}{\left(1 + O\left(\frac{1}{k}\right)\right) \left(1 + O\left(\frac{1}{n-k}\right)\right)} \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Ersetze k durch die Abweichungen vom Erwartungswert $y = y(k) = k - np$. Mit den relevanten Werten von k gilt $|y| \leq \text{const} \sqrt{n}$, also $y = O(\sqrt{n})$. Mit $k = np + y$, $n - k = n - np - y = nq - y$ gilt:

$$\begin{aligned} \sqrt{\frac{n}{k(n-k)}} &= \sqrt{\frac{n}{(np+y)(nq-y)}} = \sqrt{\frac{n}{n^2 pq + ny(q-p) - y^2}} \\ &= \frac{1}{\sqrt{npq \left(1 + \frac{y(q-p)}{npq} - \frac{y^2}{n^2 pq}\right)}} = \frac{1}{\sqrt{npq}} \left(1 + O\left(\frac{y}{n}\right)\right) = \frac{1}{\sqrt{npq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \end{aligned}$$

Für den nächsten Faktor gilt:

$$\begin{aligned} \left(\frac{np}{k}\right)^k &= \left(\frac{np}{np+y}\right)^{np+y} = \left(\frac{1}{1+\frac{y}{np}}\right)^{np+y} = e^{-np(1+\frac{y}{np}) \ln(1+\frac{y}{np})} \\ &= e^{-y - \frac{y^2}{2np} + O\left(\frac{y^3}{n^2}\right)} = e^{-y - \frac{y^2}{2np}} \left(1 + O\left(\frac{y^3}{n^2}\right)\right) = e^{-y - \frac{y^2}{2np}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \end{aligned}$$

Für den nächsten Faktor gilt analog:

$$\begin{aligned} \left(\frac{nq}{n-k}\right)^{n-k} &= \left(\frac{nq}{nq-y}\right)^{nq-y} = \left(\frac{1}{1-\frac{y}{nq}}\right)^{nq-y} = e^{-nq(1-\frac{y}{nq}) \ln(1-\frac{y}{nq})} \\ &= e^{y - \frac{y^2}{2nq} + O\left(\frac{y^3}{n^2}\right)} = e^{y - \frac{y^2}{2nq}} \left(1 + O\left(\frac{y^3}{n^2}\right)\right) = e^{y - \frac{y^2}{2nq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right). \end{aligned}$$

Insgesamt folgt (beachte $\frac{1}{p} + \frac{1}{q} = \frac{p+q}{pq} = \frac{1}{pq}$):

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-y - \frac{y^2}{2np}} e^{y - \frac{y^2}{2nq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{y^2}{2npq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{y^2}{2npq}} + O\left(\frac{1}{n}\right). \end{aligned}$$

Die gesuchte W'keit lässt sich nach diesen Vorbereitungen abschätzen. Es wird y wieder durch $y = k - np$ ersetzt:

$$\begin{aligned} P &:= P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \sum_{k=k_a}^{k_b} \binom{n}{k} p^k q^{n-k} \\ &= \underbrace{\sum_{k=k_a}^{k_b} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{(k-np)^2}{2npq}}}_I + \underbrace{\sum_{k=k_a}^{k_b} O\left(\frac{1}{n}\right)}_F. \end{aligned}$$

Setzt man $x_k = (k - np)/\sqrt{npq}$, so lässt sich der erste Term I als Riemann-Summe mit einer äquidistanten Zerlegung des Intervalls $[x_{k_a}, x_{k_b+1}]$ mit konstanter Schrittweite $x_{k+1} - x_k = 1/\sqrt{npq}$ interpretieren. Hierbei ist

$$x_{k_a} = a + O\left(\frac{1}{\sqrt{n}}\right), \quad x_{k_b+1} = b + O\left(\frac{1}{\sqrt{n}}\right)$$

und es folgt

$$I = \sum_{k=k_a}^{k_b} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} (x_{k+1} - x_k) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

(Die Approximation eines Integrals durch eine äquidistante Riemann-Summe hat den Fehler $O(1/\text{Schrittweite})$.) Im Fehlerterm F erstreckt sich die Summe nur über $k_b - k_a + 1 \approx 1 + (b - a) \sqrt{npq} = O(\sqrt{n})$ Terme, also gilt $|F| \leq O(\sqrt{n}) O(1/n) = O(1/\sqrt{n})$. Insgesamt ergibt sich damit die Behauptung

$$P = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Für die in Bemerkung 4.8 behauptete gleichmäßige Konvergenz bezüglich a und b müsste man die Abschätzungen verfeinern. Dies ersparen wir uns hier.

Q.E.D.

Bemerkung 4.8: Man kann zeigen, dass diese Approximation durch die Normalverteilung gleichmäßig in den Grenzen a und b ist, d.h., der Fehlerterm $O(1/\sqrt{n})$ kann unabhängig von a und b nach oben abgeschätzt werden durch einen nur von n abhängenden Fehlerterm, der für $n \rightarrow \infty$ verschwindet.

Visualisierung 4.9:

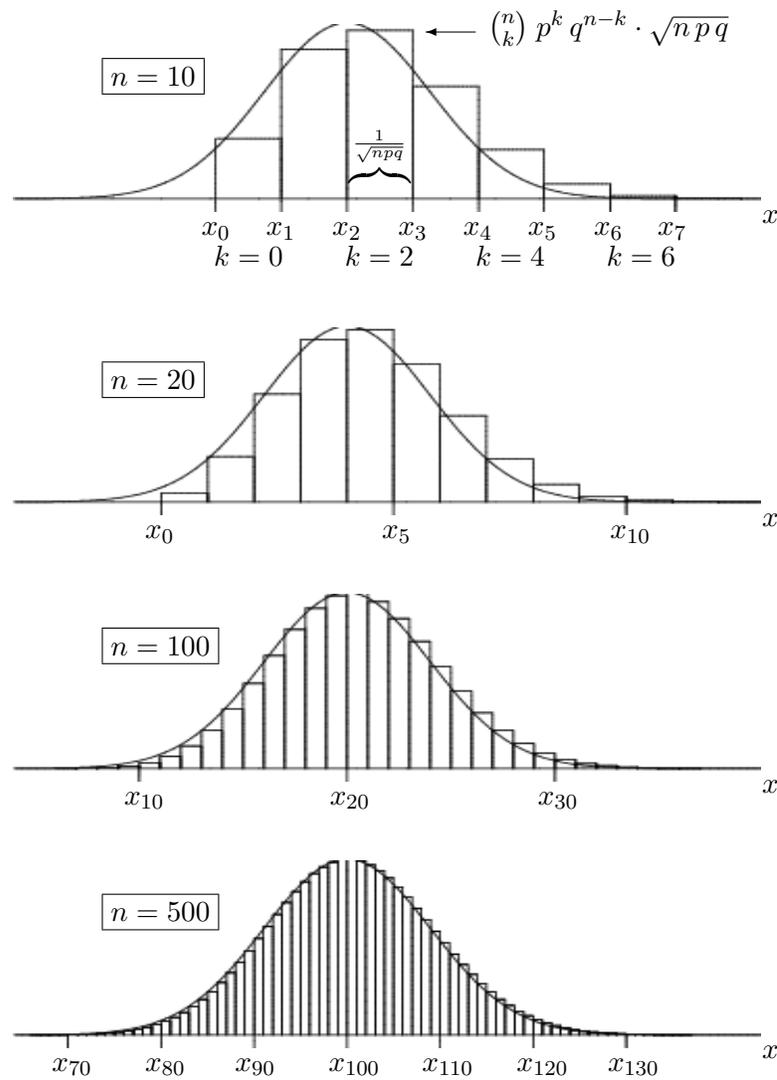
Die folgenden Bilder visualisieren den Grenzübergang. Über dem Intervall $[a, b] = [-4, 5]$ wird die Dichte $\rho(x) = e^{-x^2/2}/\sqrt{2\pi}$ der $N(0, 1)$ -Verteilung gezeichnet. Weiterhin wird für $p = 0.2$ und verschiedene Werte von n das Analogon einer "Dichte" für die diskrete Variable $Y_n = (S_n - np)/\sqrt{npq}$ als Balkendiagramm folgendermaßen dargestellt:

Die $\text{Bi}(n, p)$ -verteilte Variable S_n nimmt die Werte $k = 0, \dots, n$ an. Diesen k -Werten werden auf der x -Achse die Intervalle $[x_k, x_{k+1})$ der Länge $dx = 1/\sqrt{npq}$ von jeweils

$$x_k = \frac{k - np}{\sqrt{npq}}$$

bis $x_{k+1} = x_k + dx$ zugeordnet. Über diesen Intervallen sind jeweils die konstanten W'keitswerte $y_k = \binom{n}{k} p^k q^{n-k} / dx$ der Binomial-Verteilung dargestellt. Der Skalierungsfaktor $1/dx$ sorgt dabei dafür, dass die Fläche $y_k \cdot dx$ eines Balkens (mit Breite dx) genau die W'keit $P(S_n = k) = \binom{n}{k} p^k q^{n-k}$ ist.

Mit wachsendem n konvergiert die diskrete $\text{Bi}(n, p)$ -Verteilung offensichtlich gegen die kontinuierliche $N(0, 1)$ -Verteilung:



Folgerungen 4.10:

Die spezielle Form der Moivre-Laplace-Näherung in Satz 4.7 wird in äquivalente Aussagen umgeschrieben. Mit

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = P\left(\underbrace{np + a\sqrt{npq}}_{a'} \leq S_n \leq \underbrace{np + b\sqrt{npq}}_{b'}\right)$$

oder auch (mit der mittleren Häufigkeit $\bar{X}_n = S_n/n$)

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = P\left(a \leq \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{pq}} \leq b\right) =$$

$$P\left(\underbrace{a\sqrt{\frac{pq}{n}}}_A \leq \bar{X}_n - p \leq \underbrace{b\sqrt{\frac{pq}{n}}}_B\right) = P\left(\underbrace{p + a\sqrt{\frac{pq}{n}}}_{A'} \leq \bar{X}_n \leq \underbrace{p + b\sqrt{\frac{pq}{n}}}_{B'}\right)$$

folgen unmittelbar allgemeine Approximationen für binomial-verteilte Variablen S_n und die verwandten mittleren Häufigkeiten $\bar{X}_n = S_n/n$. Aus der Moivre-Laplace-Näherung

$$P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right)$$

folgt mit $a' = np + a\sqrt{npq}$, $b' = np + b\sqrt{npq}$, also $a = \frac{a' - np}{\sqrt{npq}}$, $b = \frac{b' - np}{\sqrt{npq}}$:

$$P(a' \leq S_n \leq b') = \frac{1}{\sqrt{2\pi}} \int_{\frac{a' - np}{\sqrt{npq}}}^{\frac{b' - np}{\sqrt{npq}}} e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Mit $A = a\sqrt{\frac{pq}{n}}$, $B = b\sqrt{\frac{pq}{n}}$, also $a = A\sqrt{\frac{n}{pq}}$, $b = B\sqrt{\frac{n}{pq}}$:

$$P(A \leq \bar{X}_n - p \leq B) = \frac{1}{\sqrt{2\pi}} \int_A^B \frac{\sqrt{\frac{n}{pq}}}{\sqrt{\frac{n}{pq}}} e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Mit $A = A' - p$, $B = B' - p$:

$$P(A' \leq \bar{X}_n \leq B') = \frac{1}{\sqrt{2\pi}} \int_{(A'-p)\sqrt{\frac{n}{pq}}}^{(B'-p)\sqrt{\frac{n}{pq}}} e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Die Integrale können z.B. mit der Wertetabelle der Funktion Φ auf Seite 81 numerisch ermittelt werden:

$$\int_a^b e^{-x^2/2} dx = \int_0^b e^{-x^2/2} dx - \int_0^a e^{-x^2/2} dx$$

$$\begin{aligned}
&= \frac{1}{2} \left(\underbrace{\int_{-b}^b e^{-x^2/2} dx}_{\Phi(b) \text{ für } b \geq 0} - \underbrace{\int_{-a}^a e^{-x^2/2} dx}_{\Phi(a) \text{ für } a \geq 0} \right) \\
\Rightarrow &\boxed{\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = \frac{1}{2} (\text{sign}(b) \Phi(|b|) - \text{sign}(a) \Phi(|a|))}.}
\end{aligned}$$

Bemerkung und Warnung 4.11:

↓2.7.07

Für großes n liefert Moivre-Laplace gute Näherungen, falls die zu berechnenden W 'keiten nicht allzu klein sind. Man sollte aber nicht erwarten, dass die Näherungen (außer bei riesigem n) auf mehr als einige wenige Dezimalstellen genau sind. Sind die zu berechnenden W 'keiten klein, so macht sich der additive Fehler $O(1/\sqrt{n})$ stark bemerkbar.

Beispiel: Betrachte eine unfaire Münze mit $p = P(\text{„Kopf“}) = 0.6$. Die Variable $S_n = \text{„Anzahl der Köpfe bei } n \text{ Würfen“}$ ist $\text{Bi}(n, p)$ -verteilt. Wir vergleichen die exakte Formel

$$P(a' \leq S_n \leq b') = \sum_{k=a'}^{b'} \binom{n}{k} p^k q^{n-k}$$

mit der Moivre-Laplace-Näherung

$$P(a' \leq S_n \leq b') = \frac{1}{\sqrt{2\pi}} \int_{\frac{a'-np}{\sqrt{npq}}}^{\frac{b'-np}{\sqrt{npq}}} e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Für $n = 100$:

$$P(55 \leq S_n \leq 70) = \underbrace{0.8541\dots}_{\text{exakt}} \approx \underbrace{0.8256\dots}_{\text{Moivre}}, \text{ Fehler} \approx 0.0285.$$

Für $n = 10\,000$:

$$P(5950 \leq S_n \leq 6100) = \underbrace{0.8286\dots}_{\text{exakt}} \approx \underbrace{0.8256\dots}_{\text{Moivre}}, \text{ Fehler} \approx 0.0030.$$

Die Fehler verhalten sich in der Tat wie $O(1/\sqrt{n})$: sie fallen etwa um den Faktor 1/10, wenn n um den Faktor 100 wächst.

Achtung: werden kleine W 'keiten $\ll 1$ approximiert, macht sich der additive Fehler der Moivre-Laplace-Näherung bemerkbar. Z.B., für $n = 10\,000$:

$$P(5500 \leq S_n \leq 5700) = \underbrace{5.62 \cdot 10^{-10}}_{\text{exakt}} \approx \underbrace{4.57 \cdot 10^{-10}}_{\text{Moivre}}.$$

Bemerkung 4.12: Wählt man in den Folgerungen 4.10 speziell $B = -A = \epsilon > 0$, so erhält man

$$\begin{aligned} P(-\epsilon \leq \bar{X}_n - p \leq \epsilon) &= \frac{1}{\sqrt{2\pi}} \int_{-\epsilon \sqrt{\frac{n}{pq}}}{\epsilon \sqrt{\frac{n}{pq}}} e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right) \\ &= \Phi\left(\epsilon \sqrt{\frac{n}{pq}}\right) + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Damit folgt

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - p| \leq \epsilon) = 1,$$

wie bereits durch das Gesetz der großen Zahl vorausgesagt wurde: der Erwartungswert $E(X) = p$ der Bernoulli-Variablen $X : \Omega \mapsto \{0, 1\}$ kann als empirischer Mittelwert \bar{X}_n häufiger Wiederholungen X_i mit großer W'keit („Sicherheit“) S bestimmt werden. Das Gesetz der großen Zahl 4.1 liefert für das Bernoulli-Experiment X mit Erwartungswert $\mu = p$ und Varianz $\sigma^2 = pq$

$$S = P(|\bar{X}_n - p| < \epsilon) \geq 1 - \frac{pq}{\epsilon^2 n},$$

dies wird durch Moivre-Laplace verfeinert zu

$$S = P(|\bar{X}_n - p| \leq \epsilon) = \Phi\left(\epsilon \sqrt{\frac{n}{pq}}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

Damit haben wir nun eine bessere Antwort auf die Frage

„Wie oft muss das Experiment X wiederholt werden, um den Erwartungswert $E(X) = p$ durch den Mittelwert $\bar{X}_n = \frac{1}{n} \sum_i X_i$ unabhängiger Messungen X_i mit Sicherheit $S = P(|\bar{X}_n - p| < \epsilon)$ auf eine absolute Genauigkeit ϵ festlegen zu können?“

Das Gesetz der großen Zahl liefert die Abschätzung (siehe auch die Beispiele 4.5 und 4.6)

$$n \geq \frac{1}{4\epsilon^2} \frac{1}{1-S} \geq \frac{pq}{\epsilon^2} \frac{1}{1-S},$$

über Moivre-Laplace erhalten wir aus $S \approx \Phi\left(\epsilon \sqrt{\frac{n}{pq}}\right)$ mit der Umkehrfunktion Φ^{-1} nun die realistischeren kleineren Approximationen

$$n \approx \frac{pq}{\epsilon^2} (\Phi^{-1}(S))^2$$

bzw.

$$n \geq \frac{1}{4\epsilon^2} (\Phi^{-1}(S))^2 \geq \frac{pq}{\epsilon^2} (\Phi^{-1}(S))^2$$

(wir können immer $p q \leq 1/4$ benutzen, falls die Größenordnung von p , q nicht bekannt ist). Zum Vergleich für verschiedene vorgegebene Sicherheiten S :

S	$\frac{1}{1-S}$	$(\Phi^{-1}(S))^2$
0.5	2	0.455
0.683	3.2	1.002
0.9	10	2.706
0.954	21.7	4.019
0.99	100	6.635
0.999	1 000	10.828

Fazit: Moivre-Laplace erleichtert das Leben (liefert realistische Abschätzungen), wenn man mit großer Sicherheit den Parameter p empirisch bestimmen möchte.

Beispiel 4.13: Wir betrachten noch einmal Beispiel 4.6: wie oft muss ein Würfel geworfen werden, um $p = P(\text{„Sechs“})$ mit einer Sicherheit von 97.5% auf eine absolute Genauigkeit $\epsilon = 1/1000$ festlegen zu können? Mit der relativen Häufigkeit \bar{X}_n der bei n Würfeln beobachteten „Sechsen“ ist n so zu bestimmen, dass

$$P\left(|\bar{X}_n - p| < \frac{1}{1000}\right) \geq 0.975 =: S$$

gilt. Nach der vorigen Bemerkung 4.12 gilt

$$S \leq P\left(|\bar{X}_n - p| < \epsilon\right) \approx \Phi\left(\epsilon \sqrt{\frac{n}{pq}}\right) \Rightarrow n \geq \frac{pq}{\epsilon^2} (\Phi^{-1}(S))^2.$$

Die Wertetabelle für Φ auf Seite 81 kann zur Bestimmung von Werten der Umkehrfunktion Φ^{-1} benutzt werden. Man findet $\Phi(2.241) \approx 0.975 = S$, also muss in guter Näherung

$$n \geq \frac{pq}{\epsilon^2} (2.241)^2$$

gelten. Für $\epsilon = 1/1000$ folgt mit der zusätzlichen Annahme eines annähernd fairen Würfels $p \approx 1/6$, $q \approx 5/6$, dass

$$n \geq 697\,512.25$$

Würfe durchzuführen sind (vergleiche mit den in Beispiel 4.6 gefundenen Werten). Ohne die Zusatzannahme $p \approx 1/6$, $q \approx 5/6$ kann man wegen $p q = p(1-p) \leq 1/4$ die Genauigkeit ϵ mit der Sicherheit $S = 0.975$ in jedem Fall für

$$n \geq \frac{1}{4\epsilon^2} (2.241)^2 \approx 1\,255\,520.25$$

Würfe garantieren.

4.3 Der Zentrale Grenzwertsatz

↓4.7.07

Der letzte Abschnitt zeigte, dass eine $\text{Bi}(n, p)$ -verteilte Variable S_n sich für großes n durch eine normalverteilte Variable annähern lässt. Die Binomial-Verteilung ist die Verteilung einer Summenvariable: $S_n = \sum_{i=1}^n X_i$ ist die Anzahl der Erfolge bei unabhängigen Wiederholungen X_i eines Bernoulli-Experiments $X : \Omega \mapsto \{0, 1\}$. Es stellt sich nun heraus, dass dies nur ein Spezialfall eines viel allgemeineren Gesetzes ist: jede Summe von unabhängigen Wiederholungen X_i eines *beliebigen* Experiments X nähert sich für großes n einer Normalverteilung an, egal ob das Experiment ein Bernoulli-Experiment ist oder einer (eventuell beliebig komplizierten) Verteilung gehorcht!

Diese bemerkenswerte Eigenschaft erlaubt in vielen Fällen, praktische Anwendungen auch dann quantitativ zu erfassen, wenn nichts über den unterliegenden Zufallsmechanismus bekannt ist. Nach häufigen Wiederholungen weiß man, dass Summen und Mittelwerte sich approximativ normalverteilt verhalten (wenn man nur genügend oft wiederholt hat).

Satz 4.14: (Zentraler Grenzwertsatz)

Sei X_1, X_2, \dots ein Familie unabhängiger Zufallsvariablen über einem gemeinsamen Stichprobenraum (Ω, \mathcal{E}, P) . Alle Variablen haben die selbe Verteilung, der gemeinsame Erwartungswert sei $\mu = E(X_1) = E(X_2) = \dots$, die gemeinsame Streuung sei $\sigma = \sigma(X_1) = \sigma(X_2) = \dots$.

Für $S_n = \sum_{k=1}^n X_k$ gilt:

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Die Konvergenz ist gleichmäßig in a und b . Der Approximationsfehler für großes n ist $O(1/\sqrt{n})$:

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx + O\left(\frac{1}{\sqrt{n}}\right).$$

Beweisskizze: (nur eine sehr grobe Andeutung)

Zu einer Zufallsvariable $X : \Omega \mapsto \mathbb{R}$ definiert man die sogenannte „Fourier-Transformierte“ $\hat{X}(\alpha) = E(e^{i\alpha X})$ (mit $i = \sqrt{-1}$, $\alpha \in \mathbb{R}$). Die Verteilungsfunktion F_X von X ist durch die Funktion $\hat{X}(\alpha)$ eindeutig bestimmt. Beispielsweise lässt sich für eine kontinuierliche Variable die Dichte $\rho(x) = F'_X(x)$ durch sogenannte „Fourier-Rücktransformation“

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\alpha x} \hat{X}(\alpha) d\alpha$$

aus $\hat{X}(\alpha)$ rekonstruieren. Die Fourier-Transformation der $N(0, 1)$ -Verteilung berechnet sich zu

$$\hat{X}(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-\alpha^2/2}.$$

Es wird gezeigt, dass sich die Fourier-Transformation der Variable

$$Y_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

für großes n der Fourier-Transformation $e^{-\alpha^2/2}$ der $N(0, 1)$ -Verteilung annähert. Hieraus folgt dann, dass sich auch die Verteilung von Y_n der $N(0, 1)$ -Verteilung annähert.

Es gilt

$$\mathbb{E}(e^{i\alpha Y_n}) = \mathbb{E}\left(e^{i\alpha \frac{1}{\sqrt{n}\sigma} \sum_k (X_k - \mu)}\right) = \mathbb{E}\left(\prod_{k=1}^n e^{i\alpha \frac{X_k - \mu}{\sqrt{n}\sigma}}\right).$$

Da mit den X_k auch die Variablen $e^{i\alpha(X_k - \mu)/(\sqrt{n}\sigma)}$ unabhängig sind, gilt nach Satz 2.53 die Produktzerlegung des Erwartungswerts:

$$\mathbb{E}(e^{i\alpha Y_n}) = \prod_{k=1}^n \mathbb{E}\left(e^{i\alpha \frac{X_k - \mu}{\sqrt{n}\sigma}}\right) = \left(\mathbb{E}\left(e^{i\alpha \frac{X_1 - \mu}{\sqrt{n}\sigma}}\right)\right)^n.$$

Hierbei wurde verwendet, dass alle Variablen die selbe Verteilung haben. Durch Reihenentwicklung von \exp und \ln folgt:

$$\begin{aligned} \ln\left(\mathbb{E}(e^{i\alpha Y_n})\right) &= n \ln\left(\mathbb{E}\left(e^{i\alpha \frac{X_1 - \mu}{\sqrt{n}\sigma}}\right)\right) \\ &= n \ln\left(\mathbb{E}\left(1 + \frac{1}{1!} \frac{i\alpha}{\sqrt{n}\sigma} (X_1 - \mu) + \frac{1}{2!} \left(\frac{i\alpha}{\sqrt{n}\sigma} (X_1 - \mu)\right)^2 + O\left(\frac{1}{\sqrt{n}^3}\right)\right)\right) \\ &= n \ln\left(\mathbb{E}(1) + \frac{1}{1!} \frac{i\alpha}{\sqrt{n}\sigma} \underbrace{\mathbb{E}(X_1 - \mu)}_0 - \frac{1}{2!} \frac{\alpha^2}{n\sigma^2} \underbrace{\mathbb{E}((X_1 - \mu)^2)}_{\sigma^2} + O\left(\frac{1}{\sqrt{n}^3}\right)\right) \\ &= n \ln\left(1 - \frac{\alpha^2}{2n} + O\left(\frac{1}{\sqrt{n}^3}\right)\right) = n \left(-\frac{\alpha^2}{2n} + O\left(\frac{1}{\sqrt{n}^3}\right)\right) = -\frac{\alpha^2}{2} + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Es folgt, dass die Fourier-Transformierte von Y_n gegen die Fourier-Transformierte $e^{-\alpha^2/2}$ der $N(0, 1)$ -Verteilung konvergiert:

$$\mathbb{E}(e^{i\alpha Y_n}) = e^{-\alpha^2/2} \left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right) \xrightarrow{n \rightarrow \infty} e^{-\alpha^2/2}.$$

Q.E.D.

In Analogie zu den Folgerungen 4.10 ergeben sich folgende Umformulierungen. Hierbei braucht in 4.10 lediglich der Erwartungswert p und die Varianz pq des Bernoulli-Experiments durch einen allgemeinen Erwartungswert μ bzw. eine allgemeine Varianz σ^2 ersetzt zu werden:

Folgerungen 4.15:

Mit

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) = P\left(\underbrace{n\mu + a\sqrt{n}\sigma}_{a'} \leq S_n \leq \underbrace{n\mu + b\sqrt{n}\sigma}_{b'}\right)$$

oder auch (mit den Mittelwerten $\bar{X}_n = S_n/n$)

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) = P\left(a \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq b\right) =$$

$$P\left(\underbrace{a \frac{\sigma}{\sqrt{n}}}_A \leq \bar{X}_n - \mu \leq \underbrace{b \frac{\sigma}{\sqrt{n}}}_B\right) = P\left(\underbrace{\mu + a \frac{\sigma}{\sqrt{n}}}_{A'} \leq \bar{X}_n \leq \underbrace{\mu + b \frac{\sigma}{\sqrt{n}}}_{B'}\right)$$

folgen unmittelbar allgemeine Approximationen für Summenvariablen S_n und Mittelwerte $\bar{X}_n = S_n/n$ für großes n . Aus dem Zentralen Grenzwertsatz

$$P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

folgt mit $a' = n\mu + a\sqrt{n}\sigma$, $b' = n\mu + b\sqrt{n}\sigma$, also $a = \frac{a' - n\mu}{\sqrt{n}\sigma}$, $b = \frac{b' - n\mu}{\sqrt{n}\sigma}$:

$$P(a' \leq S_n \leq b') \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{a' - n\mu}{\sqrt{n}\sigma}}^{\frac{b' - n\mu}{\sqrt{n}\sigma}} e^{-x^2/2} dx .$$

Mit $A = a \frac{\sigma}{\sqrt{n}}$, $B = b \frac{\sigma}{\sqrt{n}}$, also $a = A \frac{\sqrt{n}}{\sigma}$, $b = B \frac{\sqrt{n}}{\sigma}$:

$$P(A \leq \bar{X}_n - \mu \leq B) \approx \frac{1}{\sqrt{2\pi}} \int_{A \frac{\sqrt{n}}{\sigma}}^{B \frac{\sqrt{n}}{\sigma}} e^{-x^2/2} dx .$$

Mit $A = A' - \mu$, $B = B' - \mu$:

$$P(A' \leq \bar{X}_n \leq B') \approx \frac{1}{\sqrt{2\pi}} \int_{(A' - \mu) \frac{\sqrt{n}}{\sigma}}^{(B' - \mu) \frac{\sqrt{n}}{\sigma}} e^{-x^2/2} dx .$$

Die Integrale können z.B. mit der Wertetabelle der Funktion Φ auf Seite 81 numerisch ermittelt werden:

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = \frac{1}{2} \left(\text{sign}(b) \Phi(|b|) - \text{sign}(a) \Phi(|a|) \right).$$

In Analogie zur Bemerkung 4.12 folgt

Bemerkung 4.16: Wählt man in den Folgerungen 4.15 $B = -A = \epsilon > 0$, so erhält man für großes n

$$P(-\epsilon \leq \bar{X}_n - \mu \leq \epsilon) \approx \frac{1}{\sqrt{2\pi}} \int_{-\epsilon}^{\epsilon} \frac{\sqrt{n}}{\sigma} e^{-x^2/2} dx = \Phi\left(\epsilon \frac{\sqrt{n}}{\sigma}\right)$$

und damit

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq \epsilon) = 1,$$

wie bereits durch das Gesetz der großen Zahl vorausgesagt wurde: der Erwartungswert $E(X) = \mu$ einer beliebigen Variable X kann als empirischer Mittelwert \bar{X}_n häufiger Wiederholungen X_i mit großer W'keit („Sicherheit“) S bestimmt werden. Das Gesetz der großen Zahl 4.1 liefert für das Experiment X mit Erwartungswert μ und Varianz σ^2

$$P(|\bar{X}_n - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2 n},$$

dies wird durch den Zentralen Grenzwertsatz verfeinert zu

$$P(|\bar{X}_n - \mu| \leq \epsilon) \approx \Phi\left(\epsilon \frac{\sqrt{n}}{\sigma}\right).$$

Bemerkung 4.17: Führt man in den Folgerungen 4.15

$$P(A' \leq \bar{X}_n \leq B') \approx \frac{1}{\sqrt{2\pi}} \int_{(A'-\mu) \frac{\sqrt{n}}{\sigma}}^{(B'-\mu) \frac{\sqrt{n}}{\sigma}} e^{-x^2/2} dx$$

die Substitution $x = (r - \mu) \frac{\sqrt{n}}{\sigma}$ durch, so ergibt sich

$$P(A' \leq \bar{X}_n \leq B') \approx \frac{1}{\sqrt{2\pi} (\sigma/\sqrt{n})} \int_{A'}^{B'} e^{-\frac{(r-\mu)^2}{2\sigma^2/n}} dr.$$

Dieses Integral entspricht den W'keiten einer $N(\mu, \sigma^2/n)$ -verteilten Variable (siehe Abschnitt 2.4.7), d.h., es gilt die Aussage:

Der Mittelwert von n unabhängigen Messungen einer beliebigen Zufallsvariable mit Erwartungswert μ und Streuung σ genügt für großes n approximativ einer Normalverteilung mit Erwartungswert μ und Streuung σ/\sqrt{n} .

Bemerkung 4.18: In der Praxis ergibt sich allerdings das Problem: wie groß muss n denn nun wirklich sein, damit die Annahme einer Normalverteilung realistisch ist? Es gilt folgende Abschätzung nach Berry und Esseen, falls man neben dem Erwartungswert μ und der Varianz $\sigma^2 = E((X - \mu)^2)$ der Variablen auch noch die Größe $\kappa = E(|X - \mu|^3)$ kennt. Für beliebiges $b \in \mathbb{R}$ gilt:

$$\left| P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-x^2/2} dx \right| \leq \frac{0.8 \kappa}{\sigma^3 \sqrt{n}}.$$

Mit $\int_a^b = \int_{-\infty}^b - \int_{-\infty}^a$ folgt durch die Dreiecksungleichung

$$\left| P\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \leq \frac{1.6 \kappa}{\sigma^3 \sqrt{n}}.$$

Hieran sieht man deutlich den Approximationsfehler $O(1/\sqrt{n})$ des Zentralen Grenzwertsatzes.

Bemerkung 4.19: In der Praxis kennt man in der Regel weder den Erwartungswert μ , noch die Streuung σ , noch die Größe κ . Hier setzt nun die Statistik ein: es seien n unabhängige Messungen x_1, \dots, x_n einer Zufallsvariablen X gegeben. Man schätzt dann μ , σ , κ ab durch

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{v}_n = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2,$$

$$\bar{k}_n = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n |x_i - \bar{x}_n|^3.$$

Für großes n liegt \bar{x}_n mit großer W'keit in der Nähe von μ (das haben wir in diesem Kapitel gezeigt). Man nennt \bar{x}_n einen „Schätzwert“ für μ . Analog kann man zeigen, dass \bar{v}_n eine Schätzung für σ^2 liefert (siehe auch Übungsaufgabe 48). Heuristisch liefert \bar{k}_n eine Näherung für κ .

Beispiel 4.20: (Eine typische „Fehlerrechnung“ der Physik)

Eine physikalische Größe X wird mehrfach gemessen, die Größe (oder der Messprozess) unterliegt einer stochastischen Störung. Es ergeben sich die $n = 10$ Werte

$$99, 100, 103, 99, 102, 100, 98, 101, 102, 101.$$

Der „wahre“ Wert (der Erwartungswert μ von X) soll mit einer Sicherheit von $S = 0.954$ abgeschätzt werden.

Als Mittelwert und Schätzung der Varianz berechnet man

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = 100.5, \quad \bar{v}_n = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = 2.5.$$

Wir schätzen die unbekannte Streuung σ von X durch die empirische Streuung \bar{v}_n ab: mit $\sigma^2 \approx \bar{v}_n$ ergibt sich $\sigma \approx \sqrt{\bar{v}_n} \approx 1.6$. Setzt man der Messung \bar{x}_n entsprechende Mittelwertvariable als normalverteilt voraus, so liefert Bemerkung 4.16 die Sicherheit

$$S = P(|\bar{x}_n - \mu| \leq \epsilon) = \Phi\left(\epsilon \frac{\sqrt{n}}{\sigma}\right).$$

Es ist die absolute Genauigkeit ϵ gefragt, mit der \bar{x}_n den gesuchten Erwartungswert μ approximiert:

$$\epsilon = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(S) \approx \frac{1.6}{\sqrt{10}} \Phi^{-1}(S) \approx 0.51 \cdot \Phi^{-1}(S).$$

Für $S = 0.954$ gilt $\Phi^{-1}(S) \approx 2.00$; es folgt $\epsilon \approx 1.0$. Also gilt mit 95.4%-iger Sicherheit (in Physikernotation):

$$\mu = 100.5 \pm 1.0.$$

Für $S = 0.683$ gilt $\Phi^{-1}(S) \approx 1.00$; es folgt $\epsilon \approx 0.51$. Also gilt mit 68.3%-iger Sicherheit:

$$\mu = 100.5 \pm 0.51.$$

Diese Aussagen sind natürlich dadurch etwas verfälscht, dass man σ nicht wirklich kennt, sondern empirisch über $\sigma^2 \approx \bar{v}_n$ approximiert hat.

Man findet in der Tat bei Naturwissenschaftlern und Ingenieuren folgendes Rezept für „Fehlerrechnungen“:

Bei n Messungen x_1, \dots, x_n einer Größe X gilt $X = \bar{x}_n \pm \frac{2\sigma_n}{\sqrt{n}}$
mit

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Unter der Approximation des Zentralen Grenzwertsatzes gilt diese Aussage mit 95.4%-iger Sicherheit (der Faktor 2 in $2\sigma_n/\sqrt{n}$ entspricht $\Phi^{-1}(0.954) \approx 2.00$).

Zusatzfrage: wieviele Messungen muss man machen, um μ mit 95.4%iger Sicherheit auf die Genauigkeit $\epsilon = 0.1$ festzulegen? Nun ist

$$S = P(|\bar{x}_n - \mu| \leq \epsilon) = \Phi\left(\epsilon \frac{\sqrt{n}}{\sigma}\right)$$

nach n aufzulösen:

$$n = \frac{\sigma^2}{\epsilon^2} (\Phi^{-1}(S))^2 \approx \frac{\sigma_n^2}{\epsilon^2} (\Phi^{-1}(S))^2 \approx \frac{2.5}{0.1^2} 2.0^2 = 1000.$$
