Einführung in die Stochastik

Prof. Dr. H. R. Lerche

Institut für Mathematische Stochastik
Universität Freiburg

Wintersemester 2007/2008

Stand: 5. August 2009

Inhaltsverzeichnis

1 Einführung

1.1 Was ist Stochastik?

Stochastik ist der Oberbegriff von Wahrscheinlichkeitsrechnung und mathematischer Statistik. In der Stochastik werden mathematische Modelle von Zufallserscheinungen konstruiert, deren Gesetzmäßigkeiten studiert und ihre Anwendbarkeit auf reale Daten untersucht wird. Die Modelle basieren auf Zufallsbegriffen, wie z.B. dem der „Wahrscheinlichkeit“. Diese werden durch mathematische Axiome beschrieben. Die Axiome erklären jedoch nicht das Wesen des Zufalls. Dieses ist bis heute, trotz diverser mathematischer Ansätze durch von Miser und Kolmogorov, noch weitgehend ungeklärt.

1.2 Stochastische Tätigkeiten im Alltag

1. Raten
2. Entscheiden
3. Schätzen
4. Vergleichen / Testen
5. Vorhersagen
6. Versichern
7. Kontrollieren
8. Messen

1.–3. kennt man schon aus dem Kindesalter. Alle acht Typen von Tätigkeiten haben zum Ziel den Zufall unter Kontrolle zu bringen.

1.2.1 Beispiele

Raten

a) In welcher Hand ist der Gegenstand?

b) Welche Antwort ist richtig bei Unwissenheit, z. B. bei “Wer wird Millionär”?

c) Wer wird der nächste US-Präsident?

Entscheiden

a) Das Spiel: Stein–Schere–Papier

b) Wann und wo lege ich mein Geld an?

c) Zu welchem Arzt gehe ich?

d) Wenn ich im Zentrum einer Stadt parken will, ab welcher Entfernung lohnt es sich den nächsten freien Parkplatz zu nehmen?

Schätzen

a) Wieviel Sprit ist noch im Tank?

b) Wie hoch ist das Steueraufkommen in der BRD in Jahr 2007?

c) Wie häufig ist eine Krankheit in der Bevölkerung (Inzidenzrate)?

Vergleichen / Testen

a) ärztl. Untersuchung auf Krankheit

b) Vergleich von ärztlichen Behandlungen

c) Entwicklung von Medikamenten

Vorhersagen

a) Tippen: Toto, Lotto, Einzelspiele

b) Dollarkurs an Weihnachten

c) Das Wetter morgen in Freiburg

Versichern

Auto/ Haus/ Leben

Kontrollieren

a) Kontrollieren des Blutdrucks

b) Fehlerkontrolle in der Produktion eines industriellen Teils

c) Flugsicherung

Messen von physikalischen Größen in Experimenten wie Masse, Länge, Temperatur, Geschwindigkeit, Energie, Impuls

a) Die Kombination der Ergebnisse geschieht in der Regel mit der mit der sogenannten Fehlerausgleichsrechnung.

b) Will man sehr genau messen, kann man in Konflikt mit der Unschärferelation von Heisenberg geraten. Diese gibt eine untere Schranke für die Maßgenauigkeit zweier zueinander konjugierter physikalischer Größen, wie z.B. Ort und Impuls eines Teilchens. Die von Heisenberg, Schrödiger u.a. entwickelte Theorie ist stochastischer Natur. Aus heutiger Sicht funktioniert sie in der Praxis sehr gut, ist aber von ihren Grundlagen her noch immer unvollständig. Es scheint heute aber ziemlich klar, daß Einstein mit seinem Spruch: “Gott würfelt nicht” nur dann recht hat, wenn es Gott tatsächlich nicht gibt.

1.3 Bemerkungen zu Gewinnchancen und Gewinnquoten bei Casino-Spielen und anderen Wetten

Wir betrachen ein Zufallsexperiment mit zwei Ausgängen E und NE und nehmen an, daß $Ws(E) = p$ und $Ws(NE) = 1 - p$ ist. Was ist eine faire Gewinnquote G bei einer Wette auf Eintreten von E bei einem Einsatz von einer Geldeinheit? Die Antwort lautet

$$G = \frac{1-p}{p}$$

oder anschaulicher ausgedrückt

$$\frac{1}{G} = \frac{p}{1-p}.$$

Dies heißt Einsatz und Gewinn stehen im selben Verhältnis zueinander wie die zugehörigen Wahrscheinlichkeiten. Betrachten wir z.B. die Wette auf eine bestimmte Zahl, z.B. *13* bei einem Roulette mit 37 möglichen Ausgängen, d.h. mit nur einer *0*. Das Casino bietet dafür üblicherweise als Gewinnquoten *35 zu 1*. Fair wäre aber *36 zu 1*. Der Abschlag von $\frac{1}{37}$ ist die Gewinnmarge des Casinos (im Englischen *house percentage*). Allgemein läßt sich die Gewinnmarge pro Geldeinheit Einsatz als

$$1 - Ws(E)(G+1)$$

angeben. Dabei ist $(G + 1)$ die Zahlung des Casinos, wenn das Ereignis E eintritt. Das Casino zahlt *Gewinn plus Einsatz* aus.

Ähnlich ist es auch bei Fußballwetten wie jenen von *ODDSET* oder *bwin*. Bei diesen Wetten sind die Erfolgswahrscheinlichkeiten im Vorhinein aber nicht exakt bekannt, sondern werden durch den Wettanbieter meist aus Vergangenheitswerten geschätzt. Entsprechend höher sind auch deren Gewinnmargen (oder besser Sicherheitsspannen). Betrachten wir eine ODDSET-Wette aus dem Jahr 2004 für das 2. Ligaspiel *TSV 1860 gegen KSC*. Die Quoten lauteten:
bei Heimsieg(1) : 1,6
bei Unentschieden (0) : 3
bei Niederlage (2) : 4.
Dies sind im Grunde drei Wetten auf drei Ereignisse, die eng miteinander verbunden sind. Die Auszahlung ist z.B. bei Unentschieden 3, d.h. der ausgezahlte Gewinn ist 2. Bei diesen Wetten ist der Einsatz in den Quoten enthalten. Aus den Quoten, nennen wir sie Q_i für $i = 0, 1, 2$, läßt sich auf die zugrundeliegenden Wahrscheinlichkeiten schließen.

Sei $p_i = \frac{1}{Q_i}$. Dann ist

$$1 = Q_i \cdot p_i = (G_i + 1)p_i$$

oder anders geschrieben

$$\frac{1-p_i}{p_i} = G_i.$$

Das heißt, legt man eine faire Wette zugrunde, so erhält man aus Q_i (bzw. G_i) die von den Wettanbietern zugrundegelegten Wahrscheinlichkeiten p_i. Dies gilt für alle $i = 0, 1, 2$. Nun gilt aber in unserem Beispiel

$$p_1 = \frac{10}{16}, \quad p_0 = \frac{1}{3}, \quad p_2 = \frac{1}{4}$$

und damit

$$p_0 + p_1 + p_2 = \frac{58}{48} > 1.$$

Offensichtlich ist die Wette nicht fair. Da die Wahrscheinlichkeiten insgesamt zu groß sind, sind die Quoten zu niedrig.

Fair wäre z.B. 2,4 anstelle von 1,6 für Q_1. Denn dann wäre $p_1 = \frac{1}{2{,}4} = \frac{20}{48}$ und damit $p_0 + p_1 + p_2 = 1$.

Die Quoten sind offensichtlich so gestellt, daß eine beträchtliche Sicherheitsspanne (oder auch Gewinnmarge) zugunsten des Wettanbieters vorhanden ist. Wie groß ist diese Sicherheitsspanne?

Wir normieren die aus den Quoten bestimmten p_i so, daß sie zu Wahrscheinlichkeiten werden.

Seien

$$\tilde{p}_i = \frac{p_i}{\sum\limits_{i=0}^{2} p_i} \quad \text{für } i = 0, 1, 2.$$

Die erwarteten Auszahlungen unter diesen Wahrscheinlichkeiten sind

$$A_i = \frac{Q_i \tilde{p}_i}{\sum_j p_j} = \frac{1}{\sum_j p_j} < 1 \quad \text{für } i = 0, 1, 2.$$

Die Sicherheitsspanne ist für alle i dieselbe und sie beträgt $1 - \frac{1}{\sum_j p_j}$.

Für das oben betrachtete Beispiel ist sie $1 - \frac{48}{58} = \frac{10}{58} = 0,172$, das heißt 17,2 % behält der Wettanbieter in seiner Vorausschau im Mittel ein. Da die tatsächlichen Wahrscheinlichkeiten nicht exakt bekannt sind, ist es eine sinnvolle Vorsichtsmaßnahme eine genügend große Sicherheitsspanne zu wählen.

Betrachten wir nun ein Szenario, das von den Erwartungen des Wettanbieters abweicht. Was ist die Gewinnerwartung über alle drei Wetten im obigen Beispiel, wenn tatsächlich die $p_i = \frac{1}{3}$ sind? Angenommen für alle drei Wetten geht gleich viel Geld ein, sagen wir jeweils eine Geldeinheit. Angenommen (1) tritt ein, so muß der Wettanbieter 1,6 zahlen bei einer Einnahme von 3, d.h. er macht 1,4 Einheiten Gewinn, bei (0) zahlt er 3 und nimmt 3 ein, bei (2) zahlt er 4 und nimmt 3 ein. Er verliert also 1. Folglich hat er $\frac{1}{3} \cdot 1,4 - \frac{1}{3} \cdot 1 = 0,133$ als erwarteten Gewinn, dies sind immer noch 4,4% aller dieses Spiel betreffenden Umsätze.

1.4 Geschichtliches

- Astragalus (1,3,4,6) 2000-3000 v. Chr. (Art Vorform des Würfels)
- Würfel (von den Ägyptern eingeführt) 1500 v. Chr.

1.4.1 Geschichte der Wahrscheinlichkeitstheorie

1494 Pacioli: 1. Aufgabensammlung:
2 Spieler A & B spielen „balla".
Wer zuerst 6 Spiele gewinnt, bekommt den Einsatz. Beim Stand von 5:3 zugunsten von A wird das Spiel unterbrochen. Wie ist der Einsatz fair zu verteilen, wenn man davon ausgeht, daß beide Spieler gleich geschickt sind?
Erst 1656 gab Pascal die richtige Antwort.

1550 Cardano: Liber de Ludo Alea
erstmals die Formeln $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
A, B unabhängig $\Rightarrow P(A \cap B) = P(A) \cdot P(B)$

1650 Briefwechsel zwischen Fermat und Pascal

1656 Huygens: De Rationiciis in Alea Ludo
14. Problem aus dem Buch:
Wenn ein anderer und ich abwechselnd 2 Würfel werfen, gewinne ich wenn ich zuerst 7 Punkte geworfen habe. Und er gewinnt, wenn er zuerst 6 Punkte geworfen hat. Was ist das Chancenverhältnis, wenn ich ihn anfangen lasse? (Antwort $\frac{31}{60}$)

1713 Bernoulli: Ars Conjectandi (erstmals Gesetz der großen Zahlen)

1733 De Moivre: Doctrine of Chance (erstmals zentraler Grenzwertsatz aber ohne Kenntnis der Grenzverteilung)

1808 Adrain und

1809 C. F. Gauss entdecken die Normalverteilung

1812 Laplace: Théorie Analytique de Probabilités

1890 Tchebychev, Markov: Ungleichungen etc.

1900 Hilberts 6. Problem: Axiomatisierung der Wahrscheinlichkeitsrechnung

1914 Borel: Le hasard,

> „Weniger als eine Million Leute leben in Paris. Täglich berichten mir die Zeitungen über die merkwürdigen Ereignisse und Unfälle, die einigen von ihnen zustoßen. Unser Leben wäre unmöglich, wenn wir uns um all die Ereignisse sorgen würden, über die wir lesen. So kann man sagen, von einem prakitischen Standpunkt aus können wir alle Eregnisse ignorieren, die eine Wahrscheinlichkeit kleiner als 1/1 000 000 haben."

1933 Kolmogorov: Axiomatisierung der Wahrscheinlichkeitsrechnung.

1.4.2 Geschichte der Statistik („Wissenschaft des Staates“)

1 v. Chr. Augustus veranlasst Volkszählung

1538 Heinrich VIII veranlasst Volkszählung wegen Pest (Bill of Mortality)

1661 John Graunt: „Natural and Political Observations upon the Bill of Mortality“
Er schätzt die Größe der Bevölkerung von London auf 380.160.

1693 Halley: Sterbetafeln von Breslau

1800 Gauss und Legendre entdeckten die statistische Ausgleichsrechnung

1880 Lexis: Epidemiologie

1923 R. Fischer: Grundlegende Methoden der Statistik

1933 Neyman-Peason: Testtheorie

1948 A. Wald: Entscheidungstheorie

2 Wahrscheinlichkeiten bekannter Zufallsmechanismen

2.1 Elementare Mengenlehre und Sprechweisen

2.1.1 Elementare Mengenlehre

Wir benutzen die folgenden Bezeichnungen:

Ω	: Grundmenge (Menge aller möglichen Ereignisse)
$\emptyset$	: leere Menge
$\in$	: enthalten sein
$\notin$	: nicht enthalten sein
A^c	: Komplement von A (in Ω)
$\mathcal{P}(\Omega)$	: Potenzmenge von Ω

2.1.2 Sprechweisen

$A \subset \Omega$ heißt „Ereignis“
Ω heißt „sicheres Ereignis“
$\emptyset$ heißt „unmögliches Ereignis“
A^c heißt „Komplementärereignis von A“
$A \cup B$ heißt „A oder B tritt ein“ (kein ausschließendes oder)
$A \cap B$ heißt „A und B treten ein“
$\#A$ heißt „Anzahl der Elemente von A“

2.2 Beispiele mit dem Würfel

2.2.1 Beispiel 1: Der faire Würfel

Die Grundmenge (Menge aller möglichen Ereignisse) ist $\Omega = \{1, \ldots, 6\}$.

Ws(Ergebnis ist „i“) $= \frac{1}{6}$ $(i = 1, \ldots, 6)$

Ist $A \subset \Omega$, so ist $Ws(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{6}$.

Hier sind einige Aussagen:

a) Ergebnis ist „1“ oder „2“

b) Ergebnis ist gerade

c) Ergebnis ist ungerade

d) Ergebnis ist gerade und kleiner als „4“.

Die zu diesen Aussagen gehörenden Wahrscheinlichkeiten sind:

a) $\{1, 2\} \Rightarrow \text{Ws}(\{1, 2\}) = \frac{2}{6} = \frac{1}{3}$

b) $\{2,4,6\} \Rightarrow \mathrm{Ws}(\{2,4,6\}) = \frac{1}{2}$

c) $\{1,3,5\} \Rightarrow \mathrm{Ws}(\{1,3,5\}) = \frac{1}{2}$

d) $\{2\} = \{2,4,6\} \cap \{1,2,3\} \Rightarrow \mathrm{Ws}(\{2\}) = \frac{1}{6}$

2.2.2 Beispiel 2: Zwei Würfe

Betrachtet man zwei Würfe, so ist die Grundmenge Ω^2 gleich der Menge aller geordneten Paare mit Einträgen aus $\{1,\dots,6\}$: $\Omega^2 = \{(i,j)|1 \leq i,j \leq 6\} = \{(1,1),(1,2),\dots,(6,5),(6,6)\}$.

Für die Wahrscheinlichkeit $Ws(\{(i,j)\})$ das Paar (i,j) zu würfeln gilt:

$Ws(\{(i,j)\}) = \frac{1}{\#\Omega^2} = \frac{1}{36}$.

Für $A \subset \Omega^2$ ist $Ws(A) = \frac{\#A}{\#\Omega^2} = \frac{\#A}{36}$.

1. Was ist die Wahrscheinlichkeit, dass die Summe zweier Würfe gleich k ist?
   Sei $A_k := \{\text{Summe beider Würfe ist gleich } k\} = \{(i,j) \in \Omega^2 | i+j=k\}$.

$$\begin{aligned}
Ws(A_2) &= \frac{\#\{(1,1)\}}{36} = \frac{1}{36} \\
Ws(A_3) &= \frac{\#\{(1,2),(2,1)\}}{36} = \frac{2}{36} = \frac{1}{18} \\
Ws(A_4) &= \frac{\#\{(1,3),(2,2),(3,1)\}}{36} = \frac{3}{36} = \frac{1}{12} \\
Ws(A_5) &= \frac{\#\{(1,4),(2,3),(3,2),(4,1)\}}{36} = \frac{4}{36} = \frac{1}{9} \\
Ws(A_6) &= \frac{\#\{(1,5),(2,4),(3,3),(4,2),(5,1)\}}{36} = \frac{5}{36} \\
Ws(A_7) &= \frac{\#\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}}{36} = \frac{6}{36} = \frac{1}{6} \\
Ws(A_8) &= \frac{\#\{(2,6),(3,5),(4,4),(5,3),(6,2)\}}{36} = \frac{5}{36} \\
Ws(A_9) &= \frac{\#\{(3,6),(4,5),(5,4),(6,3)\}}{36} = \frac{4}{36} = \frac{1}{9} \\
Ws(A_{10}) &= \frac{\#\{(4,6),(5,5),(6,4)\}}{36} = \frac{3}{36} = \frac{1}{12} \\
Ws(A_{11}) &= \frac{\#\{(5,6),(6,5)\}}{36} = \frac{2}{36} = \frac{1}{18} \\
Ws(A_{12}) &= \frac{\#\{(6,6)\}}{36} = \frac{1}{36} \\
Ws(A_k) &= \frac{6-|k-7|}{36}
\end{aligned}$$

2. Wie hoch ist die Wahrscheinlichkeit, dass die Summe zweier Würfe kleiner gleich 10

ist?

$$\begin{aligned} Ws(\text{Summe} \leq 10) &= 1 - Ws(\text{Summe } > 10) \\ &= 1 - (Ws(\text{Summe} = 11) + Ws(\text{Summe} = 12)) \\ &= 1 - \frac{2}{36} - \frac{1}{36} = \frac{33}{36} = \frac{11}{12} \end{aligned}$$

2.2.3 Einschub: Produktmengen

Produkt von zwei Mengen
Seien A und B Mengen. Die Produktmenge von A und B ist definiert durch:
$A \times B := \{(a,b)|a \in A \text{ und } b \in B\}$.
Für die Mächtigkeit der Produktmenge gilt: $\#(A \times B) = \#A \cdot \#B$.

Produkt von endlich vielen Mengen
Seien $A_1, \ldots, A_n$ Mengen.
Man definiert: $A_1 \times \ldots \times A_n := \{(a_1, \ldots, a_n)|a_i \in A_i, i = 1, \ldots, n\}$.
Es gilt: $\#(A_1 \times \ldots \times A_n) = \prod\limits_{i=1}^{n} \#A_i$.

2.2.4 Beispiel 3: n Würfe eines Würfels

Bei n Würfen ist die Grundmenge:

$\Omega^n = \underbrace{\Omega \times \ldots \times \Omega}_{\text{n-mal}} = \{(a_1, \ldots, a_n)|1 \leq a_i \leq 6, i = 1, \ldots, n\}$.

Für $A \subset \Omega^n$ ist $Ws(A) = \frac{\#A}{\#\Omega^n}$

Oder in Worten: $Ws(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$.

Das folgende Beispiel zeigt, daß diese Formel nicht immer gilt.

2.2.5 Beispiel 4: Ein 2-Stufen-Experiment

1. Stufe: Ein Würfel wird einmal geworfen mit Ereignis „i“.

2. Stufe: Es wird i-mal gewürfelt und jedes Ereignis festgehalten.

Was ist die Wahrscheinlichkeit, daß die Summe der Würfe der 2. Stufe ≤ 6 ist?

Der **Grundraum** ist:

$\Omega = \bigcup\limits_{i=1}^{6} (\{i\} \times \Omega^i)$ mit $\Omega^i = \{(a_1, \ldots, a_i)|1 \leq a_j \leq 6, j = 1, \ldots, i\}$.

Das Ereignis A ist gegeben durch:
$A = \{\text{Summe} \leq 6\} = \bigcup\limits_{i=1}^{6} (\{i\} \times A_i)$ mit $A_i = \{(a_1, \ldots, a_i)|1 \leq a_j \leq 6, \sum\limits_{j=1}^{i} a_j \leq 6\}$

Damit ist die Wahrscheinlichkeit, daß die Summe der Würfe der 2. Stufe ≤ 6 ist:
$Ws(A) = Ws(\bigcup_{i=1}^{6}(\{i\} \times A_i)) = \sum_{i=1}^{6} Ws(\{i\} \times A_i) = \sum_{i=1}^{6} Ws(\{i\}) \cdot Ws(A_i) = \frac{1}{6}\sum_{i=1}^{6} Ws(A_i)$.
Bestimme die A_i und ihre Mächtigkeiten:

$$\begin{aligned}
A_1 &= \{1, .., 6\} \\
A_2 &= \{(1,1), (1,2), \ldots, (1,5), (2,1), \ldots, (2,4), \ldots, (5,1)\} \\
A_3 &= \{(1,1,1), \ldots, (4,1,1)\} \\
&\vdots \\
\#A_1 &= 6 \\
\#A_2 &= 15 \\
\#A_3 &= 20 \\
\#A_4 &= 15 \\
\#A_5 &= 6 \\
\#A_6 &= 1
\end{aligned}$$

Somit ist die gesuchte Wahrscheinlichkeit
$Ws(\text{Summme} \leq 6) = \frac{1}{6}(1 + \frac{15}{6^2} + \frac{20}{6^3} + \frac{15}{6^4} + \frac{6}{6^5} + \frac{1}{6^6}) = 0,254$.

2.3 Etwas anderere Würfel

Schere-Stein-Papier:
Zwei Spieler treten gegeneinander an. Auf ein Kommando formen beide ihre Hand entweder zu „Schere“, zu „Stein“ oder zu „Papier“. Es gilt: Stein$>$Schere$>$Papier$>$Stein.
Das bedeutet: Stein gewinnt gegen Schere, Schere gewinnt gegen Papier und Papier gewinnt gegen Stein.

Efron-Würfel:
Eine Variante zu Schere-Stein-Papier ist die zufällige Auswahl von je 2 Efron-Würfeln. Das sind drei unterschiedliche Würfel A, B, C, auf denen die folgenden Zahlen stehen:
$A \leftrightarrow (1,6,8), \quad B \leftrightarrow (2,4,9), \quad C \leftrightarrow (3,5,7)$.

Die Wahrscheinlichkeiten, daß mit einem Würfel eine höhere Zahl gewürfelt wird als mit einem anderen, betragen:
$Ws(A > B) = \frac{4}{9}, \quad Ws(B > C) = \frac{4}{9}, \quad Ws(C > A) = \frac{4}{9}$,
denn

$$\begin{aligned}
Ws(A < B) &= Ws(\{A = 1\}) + Ws(A = 8, B = 9) + Ws(A = 6, B = 9) \\
&= \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} \\
&= \frac{5}{9}.
\end{aligned}$$

Damit ist kein Würfel besser als die beiden anderen.

2.4 Das Münzwurfmodell

2.4.1 Die p-Münze

Die möglichen Ergebnisse bei einem Münzwurf sind „Zahl“ und „Wappen“ bzw. „1“ und „0“. Die Grungmenge ist also $\Omega = \{0,1\}$. Die Wahrscheinlichkeit für Zahl ist $Ws(1) = p$, die Wahrscheinlichkeit für Wappen ist $Ws(0) = 1-p$ (Modell einer p-Münze).
Ist $p = \frac{1}{2}$, so sprechen wir vom fairen Münzwurf.

Warum ist die Annahme $p \neq \frac{1}{2}$ sinnvoll?
Drei Beispiele dazu:

a) **Würfel**
1, falls Ergebnis beim Würfel $= 6$: $Ws(1) = \frac{1}{6}$,
0, falls Ergebnis beim Würfel $\neq 6$: $Ws(0) = \frac{5}{6}$.
In diesem Fall ist $p = \frac{1}{6}$.

b) **Wahlumfrage**
1, falls Ja zur CDU: $Ws(1) = 0,46$,
0, falls Nein zur CDU: $Ws(0) = 0,54$.
In diesem Fall ist $p = 0,46$.

c) **Roulette**
1, falls Ergebnis „Rot“: $Ws(1) = \frac{18}{37}$
0, falls Ergebnis „nicht Rot“ (also „Schwarz oder Zéro“): $Ws(0) = \frac{19}{37}$.
In diesem Fall ist $p = \frac{18}{37}$.

Wie einen Würfel kann man auch eine p-Münze mehrfach werfen:

2.4.2 Zwei Würfe

Hier ist die Grundmenge $\Omega^2 = \{(1,1),(1,0),(0,1),(0,0)\}$.
$Ws((1,1)) = p^2, \quad Ws((1,0)) = p(1-p)$,
$Ws((0,1)) = (1-p)p, \quad Ws((0,0)) = (1-p)^2$.
Die Wahrscheinlichkeit in zwei Würfen genau eine 1 zu werfen ist:
$Ws(\text{in zwei Würfen genau eine 1}) = Ws((1,0)) + Ws((0,1)) = 2p(1-p)$.

2.4.3 n Münzwürfe

Beispiel $n = 5$
$Ws(\{(1,1,0,0,1)\}) = Ws(1)Ws(1)Ws(0)Ws(0)Ws(1) = p^3(1-p)^2$

Allgemein: Die Grundmenge ist $\Omega^n = \{(e_1,\dots,e_n) | e_i \in \{0,1\}, i = 1,\dots,n\}$.
Sei $(\alpha_1,\dots,\alpha_n)$ eine gegebene 0-1-Folge der Länge n. Dann gilt für die Wahrscheinlichkeit, daß mit n Würfen genau die Folge $(\alpha_1,\dots,\alpha_n)$ geworfen wird:

$$Ws((\alpha_1,\dots,\alpha_n)) = p^{\sum_{i=1}^n \alpha_i}(1-p)^{n-\sum_{i=1}^n \alpha_i} = p^k(1-p)^{n-k} \text{ mit } k = \sum_{i=1}^n \alpha_i,$$

$k=$ Anzahl der „1“ in der Folge $(\alpha_1,\dots,\alpha_n)$.

2.4.4 Binomialverteilung

Frage: Was ist die Wahrscheinlichkeit p_k mit einer p-Münze in n Würfen k Einsen zu werfen?

$$\begin{aligned}
p_k &= Ws(\{(e_1,\ldots,e_n) \in \Omega^n | \sum_{i=1}^{n} e_i = k\}) \\
&= \sum_{(\alpha_1,\ldots,\alpha_n)\in\{(e_1,\ldots,e_n)\in\Omega^n | \sum\limits_{i=1}^{n} e_i = k\}} Ws((\alpha_1,\ldots,\alpha_n)) \\
&= \#\{(e_1,\ldots,e_n) \in \Omega^n | \sum_{i=1}^{n} e_i = k\} \cdot p^k(1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

$\binom{n}{k}$ heißt Binomialkoeffizient und beschreibt die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Aus der Analysisvorlesung istbekannt daß $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ mit $n! = n(n-1)(n-2)\ldots 1$. (Binomische Formel $(a+b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, n \in \mathbb{N}, a, b \in \mathbb{R}$)
Aufgrund der Binomischen Formel gilt $\sum\limits_{k} p_k = 1$.

3 Grundbegriffe

3.1 Diskreter Wahrscheinlichkeitsraum

3.1.1 Definition (Diskreter Wahrscheinlichkeitsraum)

Ein diskreter Wahrscheinlichkeitsraum (Ω, P) besteht aus einer höchstens abzählbaren Menge Ω und einer Abbildung $P : \mathcal{P}(\Omega) \to [0,1]$ mit den folgenden Eigenschaften:

a) $P(\Omega) = 1$

b) $P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$ für jede Folge A_i ($A_i \subset \Omega$, $i = 1, 2, \ldots$) mit $A_i \cap A_j = \emptyset$ für $i \neq j$.

Bemerkung
Die Abbildung $P : \mathcal{P}(\Omega) \to [0,1]$ heißt diskretes Wahrscheinlichkeitsmaß. Ω heißt Grundraum. $P(A)$ heißt Wahrscheinlichkeit von A. $\mathcal{P}(\Omega)$ bezeichnet die Potenzmenge von Ω.

3.1.2 Definition (Wahrscheinlichkeitsfunktion)

Sei Ω eine nichtleere höchstens abzählbar Menge. Eine Abbildung $p : \Omega \to [0,1]$ heißt Wahrscheinlichkeitsfunktion, falls $\sum\limits_{\omega \in \Omega} p(\omega) = 1$.

3.1.3 Satz 1 (Zusammenhang zwischen Ws-Maß und Ws-Funktion)

1. Sei p eine Wahrscheinlichkeitsfunktion auf Ω und $A \subset \Omega$.
Durch $P(A) := \sum\limits_{\omega \in A} p(\omega)$ wird ein Ws-Maß definiert.
2. Sei P ein Ws-Maß.
Durch $p : \Omega \to [0,1]$, $p(\omega) := P(\{\omega\})$ wird eine Wahrscheinlichkeitsfunktion erklärt.

Der Satz besagt: Wahrscheinlichkeitsmaß und Wahrscheinlichkeitsfunktion stehen in eineindeutiger Beziehung zueinander.

Beweis:
Zu (1): Es sind die Eigenschaften des Wahrscheinlichkeitsmaßes nachzuweisen.
a) $P(\Omega) = \sum\limits_{\omega \in \Omega} p(\omega) \overbrace{=}^{\text{Definition}} 1.$
b) Seien $A_i \subset \Omega$, $i \geq 1$ disjunkt. Dann ist:
$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{\omega \in \bigcup A_i} p(\omega) \overbrace{=}^{\text{Reihenumordnungssatz}} \sum_{i=1}^{\infty} \sum_{\omega \in A_i} p(\omega) = \sum_{i=1}^{\infty} P(A_i).$$

Zu (2): Es ist die Eigenschaft der Wahrscheinlichkeitsfunktion nachzuprüfen.
$\sum\limits_{\omega \in \Omega} p(\omega) = \sum\limits_{\omega \in \Omega} P(\{\omega\}) = P(\bigcup\limits_{\omega \in \Omega} \{\omega\}) = P(\Omega) = 1.$

3.1.4 Beispiele

1. **Würfel**: $\Omega = \{1, 2, \ldots, 6\}$, $p(\omega) = \frac{1}{6}$.
2. **Fairer Münzenwurf**: $\Omega^n = \{\omega = (\omega_1, \ldots, \omega_n) | \omega_i \in \{0, 1\}, i = 1, ..., n\}$, $p(\omega) = \frac{1}{2^n}$.
3. **Gleichverteilung**: Ω endlich, $p(\omega) = \frac{1}{\#\Omega}$.
4. **Binomialverteilung**: $\Omega = \{0, 1, \ldots, n\}$ und $0 < p < 1$,
   $p(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$.
5. **Poisson-Verteilung**: $\Omega = \{0, 1, 2, \ldots\} = \mathbb{N} \cup \{0\}, \lambda > 0$.
   $p(\omega) = \frac{\lambda^\omega}{\omega!} e^{-\lambda}$, $\sum\limits_{\omega \in \Omega} p(\omega) = 1$, denn $\sum\limits_{\omega} p(\omega) = e^{-\lambda} \sum\limits_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!} = e^{-\lambda} e^{\lambda} = 1$.
6. **Empirische Verteilung beim Würfel**: $\Omega = \{1, \ldots, 6\}$.
   Ein Würfel werde n-mal geworfen. n_ω sei die Anzahl der Würfe mit Ergebnis ω.
   Dann ist $\sum\limits_{\omega=1}^{6} n_\omega = n$. $p(\omega) = \frac{n_\omega}{n}$, $\sum\limits_{\omega=1}^{6} p(\omega) = 1$, p selbst ist hierbei zufällig.

3.1.5 Folgerungen aus der Definition:

(1) $P(\emptyset) = 0$,

(2) $P(\bigcup\limits_{i=1}^{n} A_i) = \sum\limits_{i=1}^{n} P(A_i)$ für A_i disjunkt,

(3) $P(\bigcup\limits_{i=1}^{\infty} A_i) \leq \sum\limits_{i=1}^{\infty} P(A_i)$ für beliebige $A_i \subset \Omega$,

(4) $P(\bigcup\limits_{i=1}^{n} A_i) \leq \sum\limits_{i=1}^{n} P(A_i)$ für beliebige $A_i \subset \Omega$,

(5) $P(A^c) = 1 - P(A)$,

(6) $P(A) \leq P(B)$ falls $A \subset B$,

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

(8) $P(A_1 \cup \ldots \cup A_n) = \sum\limits_{k=1}^{n} \sum\limits_{\{i_1, \ldots, i_k\} \in \mathcal{P}_k} (-1)^{k-1} P(A_{i_1} \cap \ldots \cap A_{i_k})$.
Dabei ist $\mathcal{P}_k$ die Menge aller k-elementigen Teilmengen der Menge $\{1, \ldots, n\}$.

(9) Sei $A_n \subset \Omega, A_n \subset A_{n+1}$ für $n \geq 1$ und $A = \bigcup\limits_{i=1}^{\infty} A_i$. Dann gilt $P(A) = \lim\limits_{n \to \infty} P(A_n)$.

(10) Sei $A_n \subset \Omega, A_n \supset A_{n+1}$ für $n \geq 1$ und $A = \bigcap\limits_{n=1}^{\infty} A_n$. Dann gilt $P(A) = \lim\limits_{n \to \infty} P(A_n)$.

Beweis:

Zu (1): $P(\emptyset) = P(\cup_{n\in\mathbb{N}}\emptyset) = \sum_{n\in\mathbb{N}} P(\emptyset)$. Da $P(\emptyset) \in [0,1]$, folgt $P(\emptyset) = 0$.

Zu (2): Seien $A_1, \ldots, A_n \subset \Omega$ disjunkt. Setze $A_i = \emptyset$ für $i > n$. Dann ist:

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i) \overbrace{=}^{\text{Definition}} \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \overbrace{=}^{(1)} \sum_{i=1}^{n} P(A_i).$$

Zu (4): (4) ist ein Spezialfall von (3). Setze in (3) $A_i = \emptyset$ für $i > n$. Dann ist:

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{n} A_i) + P(\emptyset) = P(\bigcup_{i=1}^{n} A_i) + P(\bigcup_{i=n+1}^{\infty} A_i) = P(\bigcup_{i=1}^{\infty} A_i) \overbrace{\leq}^{(3)} \sum_{i=1}^{\infty} P(A_i)$$

$$= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \overbrace{=}^{(1)} \sum_{i=1}^{n} P(A_i).$$

Zu (5): $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$. Subtraktion von $P(A)$ auf beiden Seiten ergibt die Behauptung $P(A^c) = 1 - P(A)$.

Zu (6): Sei $A \subset B$. Dann ist $B = A \cup (A^c \cap B)$, A und $A^c \cap B$ sind disjunkt. Also gilt nach (2): $\mathrm{P}(B) = P(A) + P(A^c \cap B) \geq P(A)$.

Zu (7): Für A und B gilt: $B = (A \cap B) \cup (A^c \cap B)$ und die Mengen $A \cap B$ und $A^c \cap B$ sind disjunkt. Also ist $P(B) = P(A \cap B) + P(A^c \cap B)$ und damit:
$(*)$ $P(A^c \cap B) = P(B) - P(A \cap B)$.

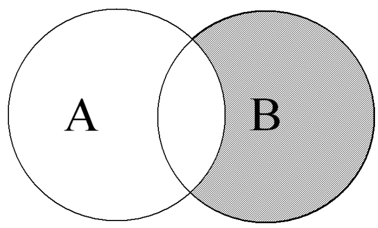


Abbildung 1: $A^c \cap B$

Daraus folgt: $P(A \cup B) = P(A) + P(A^c \cap B) \overbrace{=}^{(*)} P(A) + P(B) - P(A \cap B)$.

Etwas allgemeiner:

$P(A \cap B \cap C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Zu (8) :

Wir schreiben die Aussage wie folgt um:

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|-1} P\Big(A_I\Big)$$

Dabei ist $A_I := \bigcap_{i \in I} A_i$ und I bezeichnet eine nichtleere Teilmenge von $\{1, \ldots, n\}$.

Zeige die rechte Seite ist gleich der linken. Sei $J_w = \{i \mid w \in A_i\}$. Dann gilt $w \in A_I$ genau dann, wenn $I \subset J_w$. Die rechts Seite ist folglich gleich

$$\sum_{I\subset\{1,\ldots,n\}} (-1)^{|I|-1}\Big(\sum_{w\in A_I} p(w)\Big) = \sum_{w\in\bigcup_{i=1}^n A_i} p(w)\Big(\sum_{I\subset J_w}(-1)^{|I|-1}\Big).$$

Sei nun $|J_w| = j \geq 1$.

Dann gilt

$$\begin{aligned}\sum_{I\subset J_w}(-1)^{|I|-1} &= \sum_{i=1}^{j}(-1)^{i-1}\binom{j}{i}\\ &= 1-\sum_{i=1}^{j}(-1)^{i}\binom{j}{i}\\ &= 1\end{aligned}$$

Der 2. Term ist gleich null wegen der binomischen Formel.

Einsetzen liefert die Behauptung.

Zu (9): A ist die abzählbare Vereinigung der disjunkten Mengen $B_1 := A_1$, $B_2 := A_2\backslash A_1$, $B_3 := A_3\backslash A_2, \ldots$, und A_n ist die endliche disjunkte Vereinigung der Mengen $B_1, \ldots, B_n$.
Also gilt: $P(A) = P(\bigcup_{i=1}^{\infty} B_i) \overbrace{=}^{\text{Definition}} \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty}\sum_{i=1}^{n} P(B_i) \overbrace{=}^{(2)} \lim_{n\to\infty} P(\bigcup_{i=1}^{n} B_i) = \lim_{n\to\infty} P(A_n)$.

Zu (10): Mit $A_n \supset A_{n+1}$ für $n = 1,2,3,\ldots$ und $A = \bigcap_{i=1}^{\infty} A_i$ gilt auch $A_n^c \subset A_{n+1}^c$ für $n = 1,2,3,\ldots$ und $A^c = \bigcup_{i=i}^{\infty} A_i^c$.

Also gilt: $P(A) = 1 - P(A^c) \overbrace{=}^{(9)} 1 - \lim_{n\to\infty} P(A_n^c) = 1 - (\lim_{n\to\infty}(1 - P(A_n)) = \lim_{n\to\infty} P(A_n)$.

Beispiel zur Stetigkeit des Wahrscheinlichkeitsmaßes

Wir betrachten eine p-Münze mit $0 < p \leq 1$.

Behauptung: Wenn man eine p-Münze mehrfach wirft, kommt irgendwann eine „1“. Man muß nur oft genug werfen. In Formeln: $P(\text{irgendwann kommt „1“}) = 1$.

Beweis:
1. Möglichkeit: Der Grundraum ist hier die Menge aller unendlich langen 0-1-Folgen $\Omega = \{(\omega_1, \omega_2, \ldots)|\omega_i \in \{0,1\}\}$ und unser Ereignis ist $A = \{\text{irgendwann kommt „1“}\}$.
Wir definieren die Ereignisse $A_n := \{\text{unter den ersten n Würfen kommt eine „1“}\}$ $(n = 1,2,3,\ldots)$. Dann sind die Komplementärereignisse zu A und den A_n:
$A^c = \{(\omega_1, \omega_2, \ldots)|\omega_i = 0, i = 1,2,3,\ldots\} = \{(0,0,0,\ldots)\}$,
$A_n^c = \{(\underbrace{0,\ldots,0}_{n\text{-mal }'0'}, \omega_{n+1}, \ldots)|\omega_{n+i} \in \{0,1\} \text{ für } i \geq 1\}$.

Da $A_n^c \supset A_{n+1}^c$ für alle n und $A^c = \bigcap\limits_{i=1}^{\infty} A_n^c$, können wir Formel (10) aus den Folgerungen anwenden: $P(A^c) = \lim\limits_{n\to\infty} P(A_n^c) = \lim\limits_{n\to\infty}(1-p)^n = 0$. Damit ist $P(A) = 1 - P(A^c) = 1$.

2. Möglichkeit: Man kann dies auch so erhalten:

P(irgendwann kommt „1“) $= \sum\limits_{i=1}^{\infty} P$(„1“ kommt erstmals im i-ten Wurf)

$= \sum\limits_{i=1}^{\infty} P(\{(\underbrace{0,\ldots,0}_{(i-1)\text{-mal } '0'},1)\}) = \sum\limits_{i=1}^{\infty}(1-p)^{i-1}p = p\sum\limits_{j=0}^{\infty}(1-p)^j = p\frac{1}{1-(1-p)} = 1.$

Fixpunktfreie Permutationen

Das ist ein Beispiel für Poincarés Ein- und Ausschlußformel.

Aufgabe: n Personen kommen mit je einem Geschenk zu einer Party. Die Geschenke werden zufällig unter den Gästen verteilt. Wie groß ist die Wahrscheinlichkeit, dass mindestens eine Person ihr Geschenk zurückerhält?

Lösung: Nummeriere die Personen von $1,\ldots,n$. Die Permutation $\omega = (\omega_1,\omega_2,\ldots,\omega_n)$ bedeute, daß Person 1 das Geschenk von Person ω_1 erhält, Person 2 erhält das Geschenk von Person ω_2, ... und Person n erhält das Geschenk von Person ω_n. Die Menge all dieser Permutationen ergibt den Grundraum $\Omega = \{\omega = (\omega_1,\ldots,\omega_n) | 1 \le \omega_i \le n, \omega_i \ne \omega_j, i \ne j\}$.

Jede dieser Permutationen soll mit gleicher Wahrscheinlichkeit eintreten. Wir können also von einer Gleichverteilung auf Ω ausgehen. Da $\#\Omega = n!$, gilt für die Wahrscheinlichkeitsfunktion $p(\omega) = \frac{1}{n!}$ und für das Wahrscheinlichkeitsmaß $P(A) = \frac{\#A}{\#\Omega}$.

A_i sei das Ereignis, daß Person i ihr Geschenk zurückerhält ($i = 1,\ldots,n$). Also ist das Ereignis A, daß mindestens eine Person ihr Geschenk zurückerhält $A = A_1 \cup \ldots \cup A_n$. Mit der Ein- und Ausschlußformel läßt sich dieser Ausdruck umformen zu:

$P(A_1 \cup \ldots \cup A_n) = \sum\limits_{k=1}^{n} \sum\limits_{\{i_1,\ldots,i_k\}\in\mathcal{P}_k} (-1)^{k-1} P(A_{i_1} \cap \ldots \cap A_{i_k}).$

Berechnen wir nun $P(A_{i_1} \cap \ldots \cap A_{i_k})$:

$A_{i_1} \cap \ldots \cap A_{i_k}$ ist die Menge der Permutationen, die die Punkte $i_1,\ldots,i_k$ als Fixpunkte haben. O.B.d.A. seien dies die Punkte $1,\ldots,k$. Die Anzahl der Permutationen, die die ersten k Elemente als Fixpunkte haben, ist gleich der Anzahl der Permutationen der übrigen $n-k$ Elemente, also $(n-k)!$. Damit ist:

$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_1 \cap \ldots \cap A_k) = \frac{\#(A_1\cap\ldots\cap A_k)}{\#\Omega} = \frac{(n-k)!}{n!}.$

Zurück zur Aufgabe: Mit der gefundenen Beziehung ergibt sich weiter:

$P(A_1 \cup \ldots \cup A_n) = \sum\limits_{k=1}^{n} \sum\limits_{\{i_1,\ldots,i_n\}\in\mathcal{P}_k} (-1)^{k-1}\frac{(n-k)!}{n!} = \sum\limits_{k=1}^{n}(-1)^{k-1}\binom{n}{k}\frac{(n-k)!}{n!}$

$\overset{\binom{n}{k}=\frac{n!}{k!(n-k)!}}{\widehat{=}} \sum\limits_{k=1}^{n}(-1)^{k-1}\frac{1}{k!} = 1 - \sum\limits_{k=0}^{n}(-1)^k\frac{1}{k!}.$

Die Wahrscheinlichkeit, daß mindestens ein Gast sein Geschenk zurückerhält ist folglich:

$1 - \sum\limits_{k=0}^{n}\frac{(-1)^k}{k!}.$

Für sehr viele Gäste ($n \to \infty$) konvergiert diese Wahrscheinlichkeit gegen $1 - e^{-1}$.

Das Komplementärereignis hierzu ist, daß kein Gast sein mitgebrachtes Geschenk zurückerhält. Die Wahrscheinlichkeit dafür konvergiert gegen $e^{-1} \cong 0,37$. Folglich sind für große n etwa 37% aller Permutationen fixpunktfrei.

4 Gleichverteilungen und Kombinatorik

4.1 Die Gleichverteilung

4.1.1 Die Gleichverteilung

Bei der Gleichverteilung betrachten wir einen endlichen Grundraum Ω und gehen davon aus, daß alle Elementarereignisse aus Ω mit der gleichen Wahrscheinlichkeit eintreten. (Ein Element $\omega \in \Omega$ heißt Elementarereignis.)
Aus $1 = \sum\limits_{\omega \in \Omega} p(\omega) = \#\Omega \cdot p(\omega)$ folgt $p(\omega) = \frac{1}{\#\Omega}$ für alle $\omega \in \Omega$.
Für $A \subset \Omega$ gilt: $P(A) = \frac{\#A}{\#\Omega}$. In Worten: $P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$.

Ein Gleichverteilungsproblem besteht also in der Bestimmung der Anzahl der günstigen Fälle und der Anzahl der möglichen Fälle.

4.1.2 Einleitendes Beispiel

In einer Urne sind 5 weiße und 4 schwarze nicht nummerierte Kugeln. Es werden 3 Kugeln gezogen. Was ist die Wahrscheinlichkeit 2 weiße und eine schwarze Kugel zu ziehen?

Die Anzahl der möglichen Fälle ist $\binom{9}{3}$. Die Anzahl der günstigen Fälle ist $\binom{5}{2}\binom{4}{1}$.
Die gesuchte Wahrscheinlichkeit beträgt also $\frac{\binom{5}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{10}{21}$.

4.2 Das Kombinationsprinzip

4.2.1 Das Kombinationsprinzip in Worten

Sei Ω eine Menge von n-Tupeln $\omega = (\omega_1, \ldots, \omega_n)$, die man als Ergebnisse eines aus n Teilexperimenten bestehenden Zufallsexperiments auffassen kann, wobei ω_i das Ergebnis des i-ten Teilexperiments ist. Für das erste Teilexperiment gebe es k_1 mögliche Ausgänge. Für jedes i sei k_i die Zahl der möglichen Ausgänge des i-ten Teilexperimentes, unabhängig davon wie die früheren Teilexperimente ausgegangen sind. Dann ist $\#\Omega = k_1 \ldots k_n$.

4.2.2 Das Kombinationsprinzip in Formeln

Seien $A_1, A_2, \ldots, A_n$ endliche Mengen und $\Omega \subset A_1 \times \ldots \times A_n$.
Sei $k_j \in \mathbb{N}$ mit $k_1 = |A_1|, k_j \leq |A_j|$ für $j = 2, \ldots, n$.
Sei $\Omega^j_{\omega'_1,\ldots,\omega'_{j-1}} = \{(\omega_1, \ldots, \omega_j) | \omega_i = \omega'_i, i = 1, \ldots, j-1\}$ für $j = 2, \ldots, n$ und
sei $\Omega = \{(\omega_1, \ldots, \omega_n) | \omega_1 \in A_1, (\omega_1, \ldots, \omega_i) \in \Omega^i_{\omega_1,\ldots,\omega_{i-1}}$ für $i = 2, \ldots, n\}$.
Gilt für $j = 2, \ldots, n \quad |\Omega^j_{\omega_1,\ldots,\omega_{j-1}}| = k_j$ für alle $(\omega_1, \ldots, \omega_{j-1})$, so folgt $\#\Omega = \prod\limits_{j=1}^{n} k_j$.

4.3 Urnen- und Schachtelmodelle

4.3.1 Das Urnenmodell

In einer Urne seien N nummerierte Kugeln. Man zieht n-mal aus der Urne eine Kugel. Man kann mit oder ohne Zurücklegen und mit oder ohne Beachtung der Reihenfolge ziehen. Folgende vier Fälle werden unterschieden:

1) Ziehen mit Zurücklegen und mit Beachtung der Reihenfolge.
2) Ziehen mit Zurücklegen und ohne Beachtung der Reihenfolge.
3) Ziehen ohne Zurücklegen und mit Beachtung der Reihenfolge.
4) Ziehen ohne Zurücklegen und ohne Beachtung der Reihenfolge.

4.3.2 Das Schachtelmodell

Ein komplementäres Modell ist das Schachtelmodell.
Es werden n Kugeln auf N nummerierte Schachteln verteilt. Man kann nun Mehrfachbelegungen der Schachteln zulassen oder nicht und man kann die Kugeln nummerieren oder nicht. Auch hier gibt es vier Möglichkeiten:

1) Verteilung mit Mehrfachbesetzung und mit Nummerierung.
2) Verteilung mit Mehrfachbesetzung und ohne Nummerierung.
3) Verteilung ohne Mehrfachbesetzung und mit Nummerierung.
4) Verteilung ohne Mehrfachbesetzung und ohne Nummerierung.

4.3.3 Zusammenhang zwischen Urnenmodell und Schachtelmodell

Zwischen dem Urnenmodell und dem Schachtelmodell besteht ein ganz enger Zusammenhang. Die Fragen: „Wie viele Möglichkeiten gibt es n Kugeln aus einer Urne mit N Kugeln zu ziehen?“ und „Wie viele Möglichkeiten gibt es n Kugeln auf N Schachteln zu verteilen?“ sind äquivalent. Dabei ist das Zurücklegen in die Urne äquivalent zur Mehrfachbesetzung der Schachteln und das Beachten der Reihenfolge ist äquivalent zur Angabe der Schachtelnummern, in die die Kugeln fallen.
So gibt es zum Beispiel genau so viele Möglichkeiten 3 Kugeln aus einer Urne mit 8 Kugeln ohne Zurücklegen und mit Beachtung der Reihenfolge zu ziehen wie es Möglichkeiten gibt 3 nummerierte Kugeln auf 8 Schachteln zu verteilen.

4.3.4 Anzahl der Elemente der Grundräume

Hier ist eine Übersicht, wie man die Anzahl der möglichen Fälle berechnet. Dabei bedeutet $[N]^n := \frac{(n+N-1)!}{(N-1)!} = N(N+1)\ldots(N+n-1)$ die ober Faktorielle und
$[N]_n := \frac{N!}{(N-n)!} = N(N-1)\ldots(N-n+1)$ die unter Faktorielle.

Urnenmodell	mit Zurücklegen	ohne Zurücklegen	
mit Reihenfolge	N^n	$[N]_n$	nummeriert
ohne Reihenfolge	$\frac{[N]^n}{n!}$	$\frac{[N]_n}{n!} = \binom{N}{n}$	nicht nummeriert
	mit Mehrfachbesetzung	ohne Mehrfachbesetzung	**Schachtelmodell**

Tabelle 1: Urnenmodell und Schachtelmodell

4.3.5 Beispiel: Lotto „6 aus 49“

In einer Urne befinden sich 49 nummerierte Kugeln. Davon werden 6 ohne Zurücklegen gezogen. Es kommt nur auf die gezogenen Nummern an, nicht auf deren Reihenfolge.

Die Mächtigkeit des Grundraumes ist somit $\binom{49}{6}$. Für i richtig getippte Zahlen ist die Anzahl der günstigen Fälle $\binom{6}{i}\binom{43}{6-i}$. Somit ist die Wahrscheinlichkeit für „i Richtige“:
$P(\text{„i“ Richtige}) = \frac{\binom{6}{i}\binom{43}{6-i}}{\binom{49}{6}}$, $1 \leq i \leq 6$.

$P(\text{„i“ Richtige mit Zusatzzahl}) = \frac{\binom{1}{1}\binom{6}{i}\binom{42}{6-(i+1)}}{\binom{49}{6}}$, $0 \leq i \leq 5$.

Die Chance für „6-Richtige“ ist: $P(“i = 6'') = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816}$

Richtige	Günstige Fälle	Chance
6R	1	$1/13983816$
5R+Z	6	$1/2330636$
5R	258	$1/54200$
4R+Z	630	$1/22196$
4R	13545	$1/1032$
3R	246820	$1/57$

Tabelle 2: Lotto ’6 aus 49’

4.3.6 Eine kombinatorische Aufgabe

In einer Urne seien n weiße und n schwarze Kugeln. n Personen ziehen je zwei Kugel ohne Zurücklegen und ohne Beachtung der Reihenfolge.
Frage: Wie hoch ist die Wahrscheinlichkeit, daß jede Personen eine weiße und eine schwarze Kugel zieht?

Antwort: Sei $A = \{\underbrace{1, 2, \ldots, n}_{weiß}, \underbrace{n+1, \ldots, 2n}_{schwarz}\}$ die Menge der Kugeln in der Urne. Der Grundraum ist dann: $\Omega = \{(\{a_1, b_1\}, \ldots, \{a_n, b_n\}) | a_i, b_i \in A, \bigcup\limits_{i=1}^{n} \{a_i, b_i\} = A\}$.
Die erste Person zieht zwei Kugeln ohne Zurücklegen und ohne Beachtung der Reihenfolge. Dafür gibt es $\binom{2n}{2}$ Möglichkeiten.
Nun zieht die zweite Person. In der Urne befinden sich jetzt nur noch $2n - 2$ Kugeln. Für

die zweite Person gibt es also noch $\binom{2n-2}{2}$ Möglichkeiten.
Führt man diesen Gedanken fort, so gibt es für die k-te Person $\binom{2n+2-2k}{2}$ Möglichkeiten und für die letzte Person gibt es dann nur noch $\binom{2n+2-2n}{2} = \binom{2}{2} = 1$ Möglichkeit.

Nach dem Kombinationsprinzip ist somit die Mächtigkeit des Grundraumes:
$\#\Omega = \binom{2n}{2}\binom{2n-2}{2}\binom{2n-4}{2}\dots\binom{2}{2} = \frac{(2n)(2n-1)}{2}\frac{(2n-2)(2n-3)}{2}\dots 1 = \frac{(2n)!}{2^n}$.

Sei $B_n := \{(\{a_1,b_1\},\dots,\{a_n,b_n\}) \in \Omega | a_i \in \{1,\dots,n\}, b_i \in \{n+1,\dots,2n\}, \bigcup\limits_{i=1}^{n}\{a_i,b_i\} = A\} \subset \Omega$ das Ereignis, daß alle n Person zwei Kugeln mit unterschiedlichen Farben ziehen.
Dann ist die Mächtigkeit von $B_n:\ \#B_n = n \cdot n \cdot (n-1)\cdot(n-1)\cdot\ldots\cdot 1 = (n!)^2$.

Damit ist $P(B_n) = \frac{(n!)^2}{(2n)!/2^n}$.

Für große n läßt sich die Wahrscheinlichkeit näherungsweise berechnen:
$P(B_n) = \frac{(n!)^2 2^n}{(2n)!} \cong \frac{(2\pi n)n^{2n}e^{-2n}\cdot 2^n}{\sqrt{4\pi n}(2n)^{2n}e^{-2n}} = \sqrt{\pi n}\frac{1\cdot 2^n}{2^{2n}} = \sqrt{\pi n}\frac{1}{2^n}$.
Dabei haben wir die Stirling-Formel verwendet. Diese lautet $n! \sim \sqrt{2\pi n}n^n e^{-n}$, das heißt $n!/\sqrt{2\pi n}n^n e^{-n} \to 1$ für $n \to \infty$.

4.4 Verteilungen, die aus Gleichverteilungen entstehen

In einer Urne seien N Kugeln. Davon seien W weiß und S schwarz, $N = W + S$. Es werden n Kugeln gezogen. Was ist die Wahrscheinlichkeit r weiße Kugeln zu ziehen?

Die Menge der Kugeln in der Urne wird beschrieben durch $A = \{1,\dots,N\}$, oder genauer:
$A = \{\underbrace{1,\dots,W}_{weiße},\underbrace{W+1,\dots,W+S}_{schwarze}\}$

Wir definieren eine Funktion $\varphi : A \to \{0,1\}$ durch: $\varphi(i) := \begin{cases} 1 \text{ falls } i \in \{1,\dots,W\} \\ 0 \text{ falls } i \in \{W+1,\dots,W+S\} \end{cases}$.

4.4.1 Ziehen mit Zurücklegen und mit Reihenfolge

Sei $(\varepsilon_1,\dots,\varepsilon_n)$ eine 0-1-Folge der Länge n. Dann gilt:

$$P(\{(a_1,\dots,a_n)|a_i \in A, \varphi(a_i) = \varepsilon_i, i = 1,\dots,n\}) = \frac{W^r S^{n-r}}{N^n},$$

wobei $r = \sum\limits_{i=1}^{n}\varepsilon_i$ die Anzahl der weißen gezogenen Kugeln ist.

Die Wahrscheinlichkeit “r $(0 \le r \le n)$ weiße Kugeln zu ziehen” beträgt:

$$P(\text{“}r \text{ weiße”}) = P(\{(a_1,\dots,a_n)|\sum_{i=1}^{n}\varphi(a_i) = r\}) = \binom{n}{r}\left(\frac{W}{N}\right)^r\left(\frac{S}{N}\right)^{n-r}.$$

Das ist die Binomialverteilung mit n Beobachtungen und $p = \frac{W}{N}$, kurz $b(n;\frac{W}{N})$.

4.4.2 Ziehen ohne Zurücklegen und mit Reihenfolge

Sei $(a_1, \ldots, a_n)$, $1 \leq a_i \leq N$ für $i = 1, \ldots, n$ und $a_i \neq a_j$ für $i \neq j$ eine Folge der Länge n. Dann gilt: $P(\{(a_1, \ldots, a_n)\}) = \frac{1}{[N]_n}$.

Sei $(\varepsilon_1, \ldots, \varepsilon_n)$ eine 0-1-Folge der Länge n. Dann gilt:

$$P(\{(a_1, \ldots, a_n) | a_i \neq a_j \text{ für } i \neq j \text{ und } \varphi(a_i) = \varepsilon_i \text{ für } i = 1, \ldots, n, \}) = \frac{[W]_r [S]_{n-r}}{[N]_n}.$$

Die Wahrscheinlichkeit "r $(0 \leq r \leq n)$ weiße Kugeln zu ziehen" ist nun:

$$\begin{aligned} p(\text{“}r \text{ weiße”}) &= P(\{(a_1, \ldots, a_n) \mid a_i \neq a_j \text{ für } i \neq j \text{ und } \sum_{i=1}^{n} \varphi(a_i) = r\}) \\ &= \binom{n}{r} \frac{[W]_r [S]_{n-r}}{[N]_n} \\ &= \frac{[W]_r / r! \, [S]_{n-r} / (n-r)!}{[N]_n / n!} \\ &= \frac{\binom{W}{r} \binom{S}{n-r}}{\binom{N}{n}}, \quad 0 \leq r \leq n \end{aligned}$$

Wir nennen diese Wahrscheinlichkeit $p(r)$.

Es gilt $\sum\limits_{r=0}^{n} \binom{W}{r} \binom{S}{n-r} = \binom{W+S}{n} \Leftrightarrow \sum\limits_{r=0}^{n} p(r) = 1$. Damit ist p eine Wahrscheinlichkeitsfunktion auf $\{0, \ldots, n\}$. Das dazugehörige Wahrscheinlichkeitsmaß heißt die **hypergeometrische Verteilung** $h(n; W, S)$.

4.5 Hypergeometrische Verteilung und Binomialverteilung

Wir gehen wieder von einer Urne mit N Kugeln aus. Davon seien W Kugeln weiß und S Kugeln schwarz, $W + S = N$. Wir ziehen n-mal ohne Beachtung der Reihenfolge.

Grundlegende Tatsache: Ist N groß im Verhältnis zu n, so unterscheiden sich „Ziehen mit Zurücklegen“ (Binomialverteilung) und „Ziehen ohne Zurücklegen“ (hypergeometrische Verteilung) nur wenig.

Beispiel: „Wahlumfrage“:
Anzahl der Wahlberechtigten $N = 5 \cdot 10^7$, Anzahl der Befragten $n = 1200$.

Die genaue Formulierung dieser Tatsache lautet:

Für $N \to \infty$ erhöhe sich die Anzahl der weißen und der schwarzen Kugeln in der Urne. Das heißt: Die Anzahl W der weißen Kugeln und die Anzahl S der schwarzen Kugeln sollen beide von N abhängen. Um das zu verdeutlichen schreiben wir W_N statt W und S_N statt S. Natürlich gilt für alle N: $W_N + S_N = N$.

4.5.1 Satz 1

Es gelte außerdem $\lim\limits_{N\to\infty} \frac{W_N}{N} = p$ $(0 \leq p \leq 1)$.
Sei $n \in \mathbb{N}$ fest. Dann gilt für $N \to \infty$: $\frac{\binom{W_N}{r}\binom{S_N}{n-r}}{\binom{N}{n}} \to \binom{n}{r} p^r (1-p)^{n-r}$ für $0 \leq r \leq n$.

Bemerkung Der Satz besagt, dass die Wahrscheinlichkeitsfunktionen der hypergeometrischen Verteilungen gegen die Binomialverteilung $b(n;p)$ konvergieren.

Beweis: Schreibe W statt W_N

$$\frac{\binom{W}{r}\binom{N-W}{n-r}}{\binom{N}{n}}$$

$$= \binom{n}{r} \frac{[W]_r [N-W]_{n-r}}{[N]_n} \qquad \text{Es gilt: } [N]_n = [N]_{n-r}[N-n+r]_r.$$

$$= \binom{n}{r} \frac{[W]_r}{[N]_r} \frac{[N-W]_{n-r}}{[N]_{n-r}} \underbrace{\frac{[N]_r}{[N-n+r]_r}}_{=:R_N(r)}$$

$$= \binom{n}{r} \frac{W(W-1)\dots(W-r+1)}{N(N-1)\dots(N-r+1)} \frac{(N-W)(N-W-1)\dots(N-W-n+r+1)}{N(N-1)\dots(N-n+r+1)} \cdot R_N(r)$$

$$\to \binom{n}{r} p^r (1-p)^{n-r} \quad \text{, weil } R_N(r) \underset{N\to\infty}{\longrightarrow} 1 \text{ für } 0 \leq r \leq n \text{ und } n \text{ fest.}$$

4.5.2 Anwendung in der Qualitätskontrolle

Sowohl Binomial- als auch Hypergeomoetrische Verteilung treten in der Qualitätskontrolle auf.

In einer Warenlieferung oder Produktionseinheit (engl. batch) sei W die Anzahl der defekten Stücke und S die Anzahl der intakten Stücke.
Wird eine Produktionseinheit verkauft, so einigen sich Produzent und Abnehmer darauf, daß der Verkauf nur dann stattfindet, wenn die Lieferung gewisse Qualitätsstandards erfüllt. Der Qualitätsstandard gelte als erfüllt, wenn der Anteil der defekten Stücke in der Lieferung maximal c ist.

Es wäre ideal, wenn es eine Möglichkeit gäbe, den Anteil der defekten Stücke exakt zu bestimmen. Das ist jedoch nur möglich, wenn man jedes einzelne Stück prüft. Eine Untersuchung aller Stücke hat jedoch manchmal ungewollte Folgen. Wird z.B. bei einer Lieferung von Feuerwerkskörpern oder Einmalblitzlichtern jedes Stück untersucht, so bedeutet das die Zerstörung der ganzen Lieferung. Außerdem kostet die Kontrolle jedes Stücks Zeit und Geld.
Also bleibt nur die Möglichkeit aufgrund von Stichproben die im Allgemeinen unbekannten Größen W und S zu schätzen. Ein Testverfahren soll folgendes beachten:
1) Der Stichprobenumfang soll möglichst klein sein, damit die ungewollten Nebenwirkungen des Tests gering sind.
2) Es soll die unbekannten Größen W und S möglichst genau schätzen, und die Wahrscheinlichkeit eines Meßfehlers gering halten.

Wenn $A(p)$ die Abnahmewahrscheinlichkeit einer Lieferung bezeichnet, in der ein Anteil p von Stücken defekt ist, wäre ein Prüfverfahren ideal, für das gilt:

$$A(p) = \begin{cases} 1 \text{ falls } p \leq c \\ 0 \text{ falls } p > c \end{cases} .$$

Das geht jedoch nur mit einer Volluntersuchung aller Stücke, wobei wir beim oben genannten Problem wären. Man muß sich also für einen Mittelweg zwischen Umfang der Stichprobe und Genauigkeit des Testverfahrens entscheiden.

Bei einer Qualitätskontrolle werden n Teile aus einer Einheit gezogen und überprüft. Die Wahrscheinlichkeit bei dieser Stichprobe vom Umfang n genau r defekte Stücke zu finden beträgt

$$p(r) = \binom{W}{r}\binom{S}{n-r}/\binom{W+S}{n}.$$

Man legt eine Grenze c fest, bei der die Einheit gerade noch akzeptiert wird, d.h. sind höchstens c Teile in dieser Stichprobe defekt, wird die Produktionseinheit abgenommen.

Wir definieren zwei Größen:
Das **Produzentenrisiko** α ist die Wahrscheinlichkeit, daß eine Prokuktionseinheit, in der maximal ein Anteil von p_1 Stücken defekt ist, nicht abgenommen wird.
Das **Abnehmerrisiko** β ist die Wahrscheinlichkeit, daß eine Produktionseinheit, in der der Anteil der defekten Stücke größer oder gleich p_2 ist, abgenommen wird.

Um den Rechenaufwand zu verringern verwendet man die Binomialverteilung als Näherung der hypergeometrischen Verteilung. p ist dabei der Anteil der defekten Teile in der untersuchten Einheit. Damit ist

$$P_p(\text{höchstens } c \text{ defekte Teile in Stichprobe vom Umfang } n) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} =: A(p)$$

die Wahrscheinlichkeit, daß die Produktionseinheit abgenommen wird.
Trägt man für einen festen Stichprobenumfang die Abnahmewahrscheinlichkeit $A(p)$ gegen den Anteil p der defekten Stücke auf, so ergibt sich folgendes Schaubild, das man „operation characteristic“ (OC) nennt:

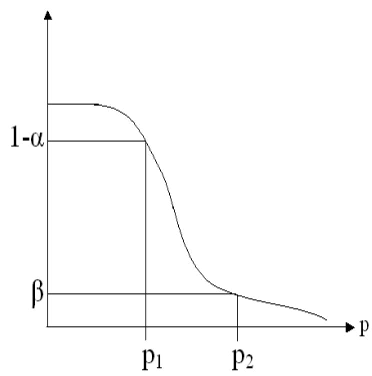


Abbildung 2: OC-Kurve

Wenn p_1, p_2, α und β festgelegt sind, stellt sich die Frage nach n und c. Im Allgemeinen wird die OC-Funktion steiler, wenn n größer wird. (Und damit wird es einfacher die Grenzen einzuhalten.)

4.5.3 Beispiel

Für $n = 100$ und $c = 5$ gilt:
a) $A(0,05) = 0,9601$,
b) $A(0,10) = 0,005$.

4.6 Kompliziertere Verteilungen, die aus Gleichverteilungen entstehen

In einer Urne seien N Kugeln. Davon seien N_i vom Typ i ($i = 1, \ldots, k$ und $k \geq 2$), wobei der Typ z.B. die „Farbe“ oder den „physikalischen Energiezustand“ bezeichnen soll. Natürlich gilt wieder $\sum\limits_{i=1}^{k} N_i = N$.
Es werden n Kugeln gezogen.
Frage: Was ist die Wahrscheinlichkeit jeweils n_i Kugeln vom Typ i zu ziehen ($i = 1, .., k$)? Dabei ist $\sum\limits_{i=1}^{k} n_i = n$.

Einschub: Multinomialkoeffizient: $\binom{n}{n1 \ldots n_k}$ ist Anzahl der Möglichkeiten eine n-elementige Menge in k Teilmengen vom Umfang $n_i, i = 1, \ldots, k$, zu zerlegen, wobei $\sum\limits_{i=1}^{k} n_i = n$.
Für $k = 2$ gilt: $\binom{n}{n_1 n_2} = \frac{n!}{n_1! n_2!} = \binom{n}{n_1}$.
Für $k \geq 2$ gilt:
$\binom{n}{n_1 \ldots n_k} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \cdots \binom{n_k}{n_k} = \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \cdots 1 = \frac{n!}{n_1! n_2! \ldots n_k!}$.

4.6.1 Ziehen mit Zurücklegen und mit Reihenfolge

Die Wahrscheinlichkeit jeweils n_i Kugeln vom Typ i zu ziehen (mit $\sum\limits_{i=1}^{k} n_i = n$) ist:
$P(\{(n_1, \ldots, n_k)\}) = \frac{\binom{n}{n_1 \ldots n_k} N_1^{n_1} \ldots N_l^{n_k}}{N^n} = \binom{n}{n_1 \ldots n_k} p_1^{n_1} \ldots p_k^{n_k}$ mit $p_i = \frac{N_i}{N}$.
Es gilt: $\sum\limits_{n_1+\ldots+n_k=n} P(\{(n_1, \ldots, n_k)\}) = \sum\limits_{n_1+\ldots+n_k=n} \binom{n}{n_1 \ldots n_k} p_1^{n_1} \ldots p_k^{n_k} = (p_1 + \ldots + p_k)^n = 1$.
Dabei haben wir die Multinomialformel
$(a_1 + a_2 + \ldots + a_k)^n = \sum\limits_{n_1+\ldots+n_k=n} \binom{n}{n_1 \ldots n_k} a_1^{n_1} \ldots a_k^{n_k}$ verwendet.

Damit wird durch P eine Wahrscheinlichkeitsfunktion erklärt. Sie heißt **Multinomialverteilung**.

In der Physik wird diese Verteilung auch Maxwell-Bolzmann-Verteilung genannt. Sie gibt an, wieviele Teilchen sich im Energieniveau „i“ befinden. Dabei wird angenommen, dass die Teilchen unterscheidbar (nummerierbar) sind; eine typische Annahme der klassischen

Physik!
In den folgenden Beispielen sind die Teilchen nicht unterscheidbar, was typisch für die Quantenmechanik ist.

4.6.2 Ziehen ohne Zurücklegen und ohne Reihenfolge

$$P(\{(n_1,\dots,n_k)\}) = \frac{\binom{N_1}{n_1}\cdots\binom{N_k}{n_k}}{\binom{N}{n}} \quad 0 \le n_i \le N_i, i = 1,\dots,k$$

In der Physik auch Fermi-Dirac-Verteilung genannt.
Im Schachtelbild: ein Teilchen pro Schachtel gemäß Pauli-Verbot.
Typische Teilchen: Elektronen.

4.6.3 Ziehen mit Zurücklegen und ohne Reihenfolge

$$P(\{(n_1,\dots,n_k)\}) = \frac{\binom{N_1+n_1-1}{n_1}\cdots\binom{N_k+n_k-1}{n_k}}{\binom{N+n-1}{n}} \quad 0 \le n_i \le N_i, i = 1,\dots,k$$

In der Physik auch Bose-Einstein-Verteilung genannt.
Im Schachtelbild: mehrere Teilchen pro Schachtel, kein Pauli-Verbot.
Typische Teilchen: Photonen

4.7 Die probalilistische Methode in der Kombinatorik

Wir untersuchen Ramsey-Zahlen. Dazu betrachten wir den vollständigne Graphen K_N mit N Ecken. Dieser Graph verbindet alle N Ecken miteinander.

4.7.1 Beispiele

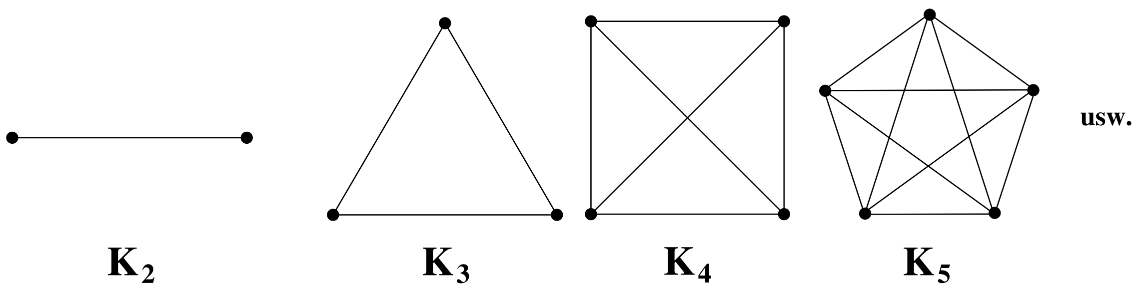


Abbildung 3: vollständige Graphen mit N Ecken

Wir sagen K_N hat die Eigenschaft (m, n), wenn, egal wie wir die Kanten von K_N rot oder blau färben, es immer einen vollständigen Untergraphen K_m gibt, dessen Kanten alle rot sind, oder es einen vollständigen Untergraphen K_n gibt, dessen Kanten alle blau sind. Ist $s \ge N$, so hat K_s auch diese Eigenschaft. Die kleinste Zahl N mit der Eigenschaft (m, n) heißt Ramsey-Zahl $R(m, n)$.

Bemerkung: Es gilt $R(m,2) = m$ ebenso $R(2,m) = m$. Denn, entweder sind alle Kanten von K_m rot oder es gibt eine blaue Kante, also ein blaues K_2.
Man kann zeigen: Für $m, n \geq 2$ ist

$$R(m,n) \leq \binom{m+n-2}{m-1},$$

und insbesondere $R(k,k) \leq 2^{2k-3}$ für $k \geq 2$. Wir leiten nun eine untere Schranke für $R(k,k)$ her. Dazu müssen wir zeigen, daß für ein möglichst großes $N < R(k,k)$ es keine Färbung von K_N gibt, für die ein roter oder blauer K_k auftritt.

4.7.2 Satz (Erdös, P.)

$$R(k,k) \geq 2^{k/2} \quad \text{für } k \geq 2$$

Beweis: Wir wissen $R(2,2) = 2$. Außerdem ist $R(3,3) \geq 6$, wegen der folgenden Fünfeckfärbung: den Rand außen *blau* und alle Diagonalen *rot*. Sei $k \geq 4$ und angenommen, daß $N < 2^{k/2}$. Wir betrachten alle rot-blau Färbungen von K_N, wobei jede Kante unabhängig mit Wahrscheinlichkeit $\frac{1}{2}$ *rot* oder *blau* gefärbt wird. Alle Färbungen, es gibt $2^{\binom{N}{2}}$, sind gleichwahrscheinlich. Sei A eine Eckenmenge der Größe k. Die Wahrscheinlichkeit des Ereignisses A_R, alle Kanten in A sind rot gefärbt, ist $2^{-\binom{k}{2}}$. Dann ist die Wahrscheinlichkeit, daß irgendeine k-Menge *rot* gefärbt ist,

$$P_R = P\left(\bigcup_{|A|=k} A_R\right) \leq \sum_{|A|=k} P(A_R) = \binom{N}{k} \cdot 2^{-\binom{k}{2}}.$$

Mit $N < 2^{\frac{k}{2}}$ und $k \geq 4$ und wegen $\binom{N}{k} \leq \frac{N^k}{2^{k-1}}$ für $k \geq 2$ folgt

$$P_R \leq \binom{N}{k} 2^{-\binom{k}{2}} \leq \frac{N^k}{2^{k-1}} 2^{-\binom{k}{2}} < 2^{\frac{k^2}{2} - \binom{k}{2} - k + 1} = 2^{-\frac{k}{2}+1} \leq \frac{1}{2} \Rightarrow P_R < \frac{1}{2}.$$

Ganz entsprechend folgt, daß die Wahrscheinlichkeit P_B, daß irgendeine k-Menge *blau* gefärbt ist $P_B < \frac{1}{2}$

$$\Rightarrow P_R + P_B < 1 \text{ für } N < 2^{k/2}.$$

D.h. es muß eine Färbung ohne rote oder blaue K_k geben, d.h. K_N hat <u>nicht</u> die Eigenschaft (k,k).

5 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

5.1 Bedingte Wahrscheinlichkeit: Definition und Folgerungen

5.1.1 Beispiel:

Eine faire Münze wird dreimal geworfen. Die Ergebnismenge ist $\Omega = \{0,1\}\times\{0,1\}\times\{0,1\}$. Dabei steht 1 für Zahl und 0 für Wappen. Sei B das Ergebnis "mindestens zweimal Zahl": $B = \{(0,1,1),(1,0,1),(1,1,0),(1,1,1)\}$. Für die faire Münze ist

$$P(B) = |B| \cdot 2^{-3} = \frac{4}{8} = \frac{1}{2}.$$

Angenommen wir wissen bereits, daß der erste Wurf "Zahl", d.h. "1" ergab, wie ändert sich die Wahrscheinlichkeit für das Ergebnis von B? Wir wissen also, daß das Ereignis $A = \{(1,0,0),(1,1,0),(1,0,1),(1,1,1)\}$ auf alle Fälle eintritt. Daher kann nur noch ein Elementarereignis von A eintreten, d.h. A ist der neue Grundraum. Da alle Elementarereignisse gleichwahrscheinlich sind, hat man

$$\frac{|A\cap B|}{|A|} = \frac{3}{4}.$$

Dabei ist

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

5.1.2 Definition (Bedingte Wahrscheinlichkeit)

Für $A, B \subset \Omega$ mit $P(A) > 0$ ist die bedingte Wahrscheinlichkeit von B gegeben A definiert als

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

Bemerkung: Die Sprechweise „gegeben A“ in der Definition bedeutet, man weiß, daß A eingetreten ist.

5.1.3 Folgerungen

(1) $P(\cdot|A)$ ist Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$ mit $P(B|A) = 1$ für $B \supset A$ und $P(C|A) = 0$ für $C \subset A^c$.

(2) $P(B)P(A|B) = P(A\cap B) = P(A)P(B|A)$.

(3) Seien $A_1, A_2, \ldots, A_k \subset \Omega$ mit $P(A_1 \cap A_2 \cap \ldots \cap A_k) > 0$. Dann gilt:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\ldots P(A_k|A_1\cap\ldots\cap A_{k-1}).$$

Beweis:
Zu (1):
a) $P(\Omega|A) = \frac{P(A\cap\Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1.$

b) Seien $B_i \subset \Omega$ disjunkt. Dann gilt:

$$P(\bigcup_{i=1}^{\infty} B_i|A) = \frac{P([\bigcup_{i=1}^{\infty} B_i]\cap A)}{P(A)} = \frac{P(\bigcup_{i=1}^{\infty}[B_i\cap A])}{P(A)} = \frac{\sum_{i=1}^{\infty} P(B_i\cap A)}{P(A)} = \sum_{i=1}^{\infty} P(B_i|A).$$

Damit ist $P(\cdot|A)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$.

Sei $B \supset A$ und $C \subset A^c$.

Dann gilt: $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$ und $P(C|A) = \frac{P(A\cap C)}{P(A)} = \frac{P(\emptyset)}{P(A)} = 0.$

Zu (3): Beweis mit Induktion. Richtig für $k = 2$ per Definition. Gelte die Formel für $k-1$, d.h.

$$P(A_1 \cap \ldots \cap A_{k-1}) = P(A_1)P(A_2|A_1)\ldots P(A_{k-1}|A_1 \cap \ldots \cap A_{k-2})$$

so schreibe

$$P(A_1 \cap \ldots \cap A_k) = P(A_k|A_1 \cap \ldots A_{k-1}) \cdot P(A_1 \cap \ldots A_{k-1})$$

und setze die Formel für $k-1$ ein.

5.1.4 Beispiel: (Das Geburtstagsproblem)

k Personen befinden sich in einem Raum. Mit welcher Wahrscheinlichkeit haben mindestens zwei von ihnen am selben Tag Geburtstag?
Wir setzen voraus

- das Jahr hat 365 Tage,
- jeder Tag kommt mit gleicher Wahrscheinlichkeit als Geburtstag in Frage,
- es besteht keine Abhängigkeit zwischen den Geburtstagen verschiedener Personen (also keine Zwillinge!).

Der Einfachheit halber denken wir uns die Personen von 1 bis k nummeriert und stellen uns vor, daß wir sie der Reihe nach befragen. Sei

$$\Omega = \{1, \ldots, 365\}^k = \{\omega|\omega = (\omega_1, \ldots, \omega_k); \omega_i \in \{1, \ldots, 365\}\}.$$

Dabei ist ω_i der Geburtstag der i-ten Person.
Sei

$$\begin{aligned} D_j &= \{(j+1)\text{-te Person hat an einem anderen Tag Geburtstag als die Personen} \\ &\qquad 1 \text{ bis } j\} \\ &= \{\omega \in \Omega|\omega_{j+1} \neq \omega_i \text{ für } 1 \leq i \leq j\}. \end{aligned}$$

Dann gilt

$$P(D_1) = P(\omega_1 \neq \omega_2) = \frac{365 \cdot 364}{365 \cdot 365} = \frac{364}{365}.$$

Sei nun $j \geq 2$. Auf dem Ereignis $D_1 \cap \ldots \cap D_{j-1}$ haben die Personen $1, \ldots, j$ an j verschiedenen Tagen Geburtstag. Damit ergibt sich die bedingte Wahrscheinlichkeit des Ereignisses D_j gegeben $D_1 \cap \ldots \cap D_{j-1}$ zu

$$P(D_j | D_1 \cap \ldots \cap D_{j-1}) = \frac{365 - j}{365} = 1 - \frac{j}{365}.$$

(Man beachte, daß dies bedingte Wahrscheinlichkeiten sind. So ist etwa $P(D_3 | D_2^c) = 364/365$.) Wir erhalten nun

$$\begin{aligned}
&P(D_1 \cap D_2 \cap \ldots \cap D_{k-1}) \\
&= P(D_1) P(D_2|D_1) P(D_3|D_1 \cap D_2) \ldots P(D_{k-1}|D_1 \cap D_2 \cap \ldots \cap D_{k-2}) \\
&= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right) \\
&= \prod_{j=1}^{k-1} \left(1 - \frac{j}{365}\right).
\end{aligned}$$

Für größere Werte von k bietet sich folgende Näherung an. Es gilt $\log(1-h) \approx -h$ und somit

$$\begin{aligned}
\log(P(D_1 \cap D_2 \cap \ldots \cap D_{k-1})) &= \sum_{j=1}^{k-1} \log(1 - j/365) \approx -(1/365) \sum_{j=1}^{k-1} j \\
&= -\frac{k(k-1)}{2 \cdot 365}.
\end{aligned}$$

Die Wahrscheinlichkeit, daß mindestens zwei Personen am selben Tag Geburtstag haben, ist daher näherungsweise

$$1 - e^{-\frac{k(k-1)}{2 \cdot 365}}.$$

Die Näherung ist sehr gut. Für $k = 23$ liefert sie 0.500 im Vergleich zu dem exakten Wert 0.506.

5.1.5 Beispiel: Sterbetafeln

Wir betrachten eine Bevölkerungsgruppe, z.B. die Einwohner einer Stadt oder eines Landes und wollen die Lebensdauern ihrer Einwohner erfassen. Dazu ordnen wir jedem Individuum sein ganzzahliges Lebensalter zu. Wir nennen diese Größe T. T ist eine ganzzahlige Größe, die vom Zufall abhängt. Sei $p(k)$ die Wahrscheinlichkeit im Alter k zu sterben. Diese ist dann $p(k) = P(T = k)$. Im Versicherungswesen, insbesondere bei Lebensversicherungen interessiert die Sterberate. Diese wird wie folgt erklärt.

Sei $S(l) := P(T \geq l)$ die Wahrscheinlichkeit mindestens l Jahre alt zu werden (Überlebenswahrscheinlichkeit) und sei $h(l) = P(T = l | T \geq l)$ die Wahrscheinlichkeit im Alter von l Jahren zu sterben, wenn man bereits dieses Lebensjahr erreicht hat (Sterberate). Es gilt:

$$h(l) = P(T = l | T \geq l) = \frac{P(T = l, T \geq l)}{P(T \geq l)} = \frac{P(T = l)}{P(T \geq l)} = \frac{p(l)}{S(l)} = \frac{S(l) - S(l+1)}{S(l)}.$$

Darstellung von $S(l)$ durch $h(l)$:

Es gilt: $S(l) = \prod_{i=1}^{l-1}(1-h(i))$. Das ergibt sich aus der Anwendung der Folgerung (3):

$S(l) = P(T \geq l) = \prod_{i=1}^{l-1} P(T \geq i+1 | T \geq i) = \prod_{i=1}^{l-1} \frac{S(i+1)}{S(i)} = \prod_{i=1}^{l-1}(1-h(i))$.

Beispiel:

Wir nehmen folgendes an: Die Wahrscheinlichkeit im Alter i zu sterben gegeben man hat das Alter i bereits erreicht, sei für alle i gleich p. In Formeln: $h(i) = p$ für alle i.
Dann ist $S(l) = (1-p)^{l-1}$ und
$p(k) = P(T=k) = \frac{P(T=k)}{P(T \geq k)} \cdot P(T \geq k) = h(k)S(k) = p(1-p)^{k-1}$ für $k \in \mathbb{N}$.
Dies ist die Wahrscheinlichkeitsfunktion der geometrischen Verteilung.

Beispiel: Sterbetafel von Breslau nach Halley (1693)

Überlebenswahrscheinlichkeit berechnet nach der Halleyschen Tafel:

$$
\begin{array}{lllllllll}
S(1) & = & P(T \geq 1) = 1 & p(1) & = & 0 & & & \\
S(2) & = & P(T \geq 2) = 855/1000 & p(2) & = & 145/1000 & h(2) & = & 145/855 \\
S(3) & = & P(T \geq 3) = 798/1000 & p(3) & = & 57/1000 & h(3) & = & 57/798 \\
S(4) & = & P(T \geq 4) = 760/1000 & p(4) & = & 38/1000 & h(4) & = & 38/760 \\
 & \vdots & & & \vdots & & & \vdots & \\
S(82) & = & 28/1000 & p(82) & = & 5/1000 & h(82) & = & 5/28 \\
S(83) & = & 23/1000 & p(83) & = & 4/1000 & h(83) & = & 4/23 \\
S(84) & = & 19/1000 & p(84) & = & 19/1000 & h(84) & = & 1 \\
S(85) & = & 0 & p(85) & = & 0 & h(85) & = & 0
\end{array}
$$

5.2 Satz von der vollständigen Wahrscheinlichkeit

5.2.1 Satz (Satz von der vollständigen Wahrscheinlichkeit)

Es seien $A_1, A_2, ...$ paarweise disjunkte Mengen mit $\bigcup_{i=1}^{\infty} A_i = \Omega$. Weiter sei $B \subset \Omega$. Dann gilt: $P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$. Dabei setzt man $P(B|A_k)P(A_k) = 0$, falls $P(A_k) = 0$.

Beweis:
$A_1 \cap B, A_2 \cap B, ...$ sind paarweise disjunkt und $\bigcup_{i=1}^{\infty}(A_i \cap B) = B$.

Damit gilt: $P(B) = P(\bigcup_{i=1}^{\infty}(A_i \cap B)) = \sum_{i=1}^{\infty} P(A_i \cap B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$.

5.2.2 Beispiel: Das Ziegenproblem (Aufgabe 4 auf dem ersten Übungsblatt)

Es gibt zwei mögliche Strategien. Die eine ist ohne Wechseln, die andere ist mit Wechseln. Welche ist besser?

1. Strategie: Ohne Zusatzinformationen wird man mit Gleichverteilung raten. Die Wahrscheinlichkeit beim ersten Raten den Hauptgewinn zu treffen ist $\frac{1}{3}$. Da man nicht wechselt, ändert sich durch den zweiten Rateversuch nichts an der Gewinnwahrscheinlichkeit. Die Wahrscheinlichkeit zu gewinnen ist also $\frac{1}{3}$.

2. Strategie: Das erste Rateergebniss wird unter Benutzung der Information des Showmasters korrigiert. $R_i, i = 1, 2$ seien die Rateergebnisse in Stufe i.
Sei P_1 das Wahrscheinlichkeitsmaß, das vorliegt, falls der Hauptgewinn hinter Tür 1 steht.
Es gilt $P_1(R_1 = j) = \frac{1}{3}, j = 1, 2, 3$.
Bei Methode 2 gilt auch noch:
a) $P_1(R_2 = 1|R_1 = 1) = 0$
b) $P_1(R_2 = 1|R_1 = 2) = 1$
c) $P_1(R_2 = 1|R_1 = 3) = 1$
Dies folgt, da der Quizmaster Tür 3 oder Tür 2 als nicht besetzt zeigen muß.

Nach Satz 5.2.1 gilt:

$$P_1(R_2 = 1) = \sum_{i=1}^{3} P_1(R_2 = 1|R_1 = i)P_1(R_1 = i) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}.$$

5.3 Bayessche Formel

5.3.1 Satz (Bayessche Formel)

Sei (Ω, P) ein Wahrscheinlichkeitsraum, $\{B_1, ..., B_l\}$ disjunkte Zerlegung von Ω und $A \subset \Omega$. Dann gilt die Bayessche Formel:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum\limits_{m=1}^{l} P(A|B_m)P(B_m)}.$$

Beweis:

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} \overbrace{=}^{5.1.3,(2)} \frac{P(A|B_j)P(B_j)}{P(A)} \overbrace{=}^{5.2.1} \frac{P(A|B_j)P(B_j)}{\sum\limits_{m=1}^{l} P(A|B_m)P(B_m)}.$$

5.3.2 Beispiel: Farbenblindheit

M männlich, W weiblich, fb farbenblind.
$P(M) = P(W) = \frac{1}{2}, \quad P(fb|M) = \frac{1}{12}, \quad P(fb|W) = \frac{1}{288}$.
Was ist die Wahrscheinlichkeit männlich zu sein, wenn man farbenblind ist?

$$P(M|fb) = \frac{P(fb|M)P(M)}{P(fb|M)P(M) + P(fb|W)P(W)} = \frac{\frac{1}{12} \cdot \frac{1}{2}}{\frac{1}{12} \cdot \frac{1}{2} + \frac{1}{288} \cdot \frac{1}{2}} = \frac{24}{25}.$$

5.3.3 Beispiel: Diagnostischer Test

Wir betrachten einen Test auf Vorhandensein einer Krankheit, z.B. den PSA-Test auf Prostata-Karzinom. Der Test hat die Ausgänge "positiv" und "negativ". Aus "positiv" schließt man auf Vorhandensein der Krankheit, aus "negativ" auf Nichtvorhandensein. Doch der Test kann ein falsches Ergebnis liefern. Man unterscheidet zwischen zwei Fehlern:

Fehler 1. Art: falsch negativ (Die Krankheit wurde nicht entdeckt; kein Alarm trotz Gefahr)
Fehler 2. Art: falsch positiv (Es liegt keine Krankheit vor; falscher Alarm)

Wir betrachen das folgende Diagnose-Beispiel:
Krankheitsrate: 1%; Testfehler: 10%.
Die Anwendung der Bayesschen Formel ergibt folgendes:

$$P(k|+) = \frac{P(+|k)P(k)}{P(+|k)P(k) + P(+|g)P(g)} = \frac{\frac{9}{10}\cdot\frac{1}{100}}{\frac{9}{10}\cdot\frac{1}{100} + \frac{1}{10}\cdot\frac{99}{100}} = \frac{1}{12}.$$

Dabei steht k für krank, $+$ steht für "der Test war positiv" und $-$ steht für "der Test war negativ".

Ein Weg ohne die Bayessche Formel und ohne Wahrscheinlichkeitsrechung das Resultat zu erhalten, geht so: Man stellt sich die Größen in einer Vierfelder-Tafel bezogen auf 1000 Probanden dar und liest das Ergebnis daraus ab.

	gesamt	Test positiv	Test negativ
krank	10	9	1
gesund	990	99	891
gesamt	1000	108	892

Wie groß ist die Wahrscheinlichkeit krank zu sein, wenn der Test positiv ist?
Antwort: $\frac{9}{108} = \frac{1}{12}$.

5.3.4 Beispiel

Eine Urne enthalte 3 Kugeln, von denen jede weiß oder schwarz sein kann. Es werden zwei Kugeln nacheinander ohne Zurücklegen gezogen. Die erste Kugel ist „schwarz", die zweite „weiß".
Welches sind die bedingten Wahrscheinlichkeiten für die Urnenbelegungen, falls vor der Ziehung alle Urnenbelegungen als gleich wahrscheinlich gelten (Siehe Abbildung 3).
Mögliche Urnenbelegungen:

1. Ziehung: „schwarz"

$$P(\underbrace{3}_{\text{Urne}}|\underbrace{S}_{\text{Farbe}}) = \frac{P(S|3)P(3)}{\sum\limits_{i=1}^{4} P(S|i)P(i)} = \frac{\frac{2}{3}\cdot\frac{1}{4}}{\frac{1}{4}(0+\frac{1}{3}+\frac{2}{3}+1)} = \frac{1}{3},$$

$$P(2|S) = \frac{1}{6}, \quad P(4|S) = \frac{1}{2}, \quad P(1|S) = 0.$$

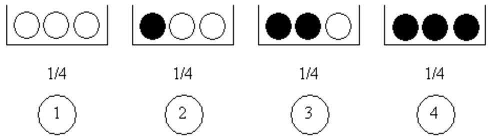


Abbildung 4: Urnenbelegung

2. Ziehung: „weiß“

$$P(3|S,W) = \frac{P(S,W|3)P(3)}{\sum_{i=1}^{4} P(S,W|i)P(i)} = \frac{\frac{2}{3}\cdot\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{4}(0+\frac{1}{3}+\frac{1}{3}+0)} = \frac{1}{2},$$
$$P(2|S,W) = \frac{1}{2}, \quad P(1|S,W) = P(4|S,W) = 0.$$

Probe: Wenn unsere Rechnung richtig ist, und bei einer dritten Ziehung wieder eine weiße Kugel gezogen wird, so muß die Wahrscheinlichkeit, daß Urne 2 vorlag, 1 sein und die Wahrscheinlichkeit, daß eine andere Urne vorlag 0.

3. Ziehung: „weiß“

$$P(2|S,W,W) = \frac{P(S,W,W|2)P(2)}{\sum_{i=1}^{4} P(S,W,W|i)P(i)} = \frac{\frac{1}{3}\cdot\frac{1}{4}}{\frac{1}{4}(0+\frac{1}{3}+0+0)} = 1.$$

Für jede andere Urne steht im Zähler eine Null und der Nenner ist verschieden von Null. Somit ist die Wahrscheinlichkeit für alle anderen Urnen 0.

5.4 Unabhängigkeit von Ereignissen

Beim Münzenwurfproblem sind wir schon davon ausgegangen, daß die Wahrscheinlichkeit für die Ergebnisse verschiedener Würfe miteinander multipliziert werden können. Hier nun der formale Hintergrund.

5.4.1 Definition

Zwei Ereignisse $A, B \subset \Omega$ heißen (stochastisch) unabhängig, falls

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

gilt.

Sind A und B unabhängig und gilt $P(B) > 0$, so ist

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

Sind A und B unabhängig und gilt $P(A) > 0$, so ist

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

Stochastische Unabhängigkeit zweier Ereignisse bedeutet, daß Kenntnis über das Eintreten des einen Ereignisses keine Information hinsichtlich des Eintretens des anderen Ereignisses liefert.

Sind A und B unabhängig, so sind auch A und B^c, A^c und B sowie A^c und B^c unabhängig. Dann ist nämlich

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) \\ &= P(A)P(B^c). \end{aligned}$$

Betrachtet man mehr als zwei Ereignisse gleichzeitig, so muß man mit der Definition vorsichtig sein.

5.4.2 Definition

Drei Ereignisse A, B und C heißen unabhängig, falls die folgenden vier Gleichungen gelten:

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \\ P(A \cap B \cap C) &= P(A)P(B)P(C). \end{aligned}$$

Paarweise Unabhängigkeit impliziert nicht Unabhängigkeit von drei Ereignissen. (Dies soll heißen, die ersten drei Gleichungen implizieren nicht die dritte.) Wir zeigen das in folgendem Beispiel.

5.4.3 Beispiel

Sei $\Omega = \{0, 1\} \times \{0, 1\}$ und $P(\{\omega\}) = 1/4$ für $\omega \in \Omega$, d.h. wir betrachten zweimaliges Werfen einer fairen Münze.

Sei $A = \{(0, 1), (0, 0)\}$, $B = \{(0, 1), (1, 1)\}$ und $C = \{(0, 0), (1, 1)\}$.

Dann gilt $|A| = |B| = |C| = 2$ und $|A \cap B| = |A \cap C| = |B \cap C| = 1$. Folglich gelten $P(A) = P(B) = P(C) = 1/2$ und $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$. Aber $P(A \cap B \cap C) = 0$, da $A \cap B \cap C = \emptyset$. Somit gilt $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$.

5.4.4 Definition (Unabhängigkeit von Ereignissen)

Die Ereignisse $A_1, ..., A_n$ heißen unabhängig, falls für jede Teilmenge $I \subset \{1, ..., n\}$ gilt

$$P\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in I} P(A_j).$$

Bemerkung: Dies sind $2^n - n - 1$ nichttriviale Gleichungen bei n Ereignissen. Für $n = 4$ also 11 Gleichungen.

5.4.5 Satz

Seien $A_1, ..., A_n$ unabhängig und $C_i \in \{\emptyset, A_i, A_i^C, \Omega\}$ für $i = 1, ..., n$.
Dann sind $C_1, ..., C_n$ unabhängig. In Formeln:

$$P\left(\bigcap_{i\in I} C_i\right) = \prod_{i\in I} P\left(C_i\right) \text{ für alle } I \subset \{1, \ldots, n\}.$$

Insbesondere sind $A_1^c, ..., A_n^c$ unabhängig.

Beweis:
Es genügt den Fall $I = \{1, \ldots, n\}$ zu betrachten.

Ist eines der $C_i = \emptyset$, so steht auf der linken Seite $P(\emptyset)$ und auf der rechten Seite steht ein Produkt, in dem ein Faktor 0 ist.

Ist eines der $C_i = \Omega$, so können wir o.B.d.A. annehmen, daß das C_n ist. In diesem Fall steht auf der linken Seite: $P\left(\bigcap\limits_{i=1}^{n} C_i\right) = P\left(\left[\bigcap\limits_{i=1}^{n-1} C_i\right] \cap \Omega\right) = P\left(\bigcap\limits_{i=1}^{n-1} C_i\right)$. Auf der rechten Seite steht: $\prod\limits_{i=1}^{n} P\left(C_i\right) = \left[\prod\limits_{i=1}^{n-1} P\left(C_i\right)\right] P(\Omega) = \prod\limits_{i=1}^{n-1} P(C_i)$. Somit können wir die $C_i = \Omega$ ignorieren.

O.B.d.A. nehmen wir an, daß $C_i \in \{A_i, A_i^c\}$ und nach eventueller Umnummerierung $C_i = A_i^c$ für $i = 1, ..., m$ und $C_i = A_i$ für $i = m+1, ..., n$ ist.

1. Fall: $m + 1 \leq n$
Mit $\overline{A_m} = \bigcap\limits_{j=m+1}^{n} A_j$ gilt dann:

$$\begin{aligned}
P\left(\bigcap_{i=1}^{n} C_i\right) &= P\left(\bigcap_{j=1}^{m} A_j^c \cap \bigcap_{j=m+1}^{n} A_j\right) \\
&= P\left(\bigcap_{j=1}^{m} A_j^c \cap \overline{A_m}\right) \\
&= P\left(\left(\bigcup_{j=1}^{m} A_j\right)^c \cap \overline{A_m}\right) \\
&= P\left(\overline{A_m}\right) - P\left(\overline{A_m} \cap \left(\bigcup_{j=1}^{m} A_j\right)\right) \\
&= P\left(\overline{A_m}\right) - P\left(\bigcup_{j=1}^{m} \left(\overline{A_m} \cap A_j\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= P\left(\overline{A_m}\right) - \sum_{k=1}^{m} \sum_{\{i_1,\dots,i_k\}\subset\{1,\dots,n\}} (-1)^{k-1} P\left(\overline{A_m} \cap A_{i_1} \cap \ldots \cap A_{i_k}\right) \\
&= P\left(\overline{A_m}\right) + \sum_{k=1}^{m} \sum_{\{i_1,\dots,i_k\}} (-1)^{k} P\left(\overline{A_m}\right) P\left(A_{i_1}\right) \ldots P\left(A_{i_k}\right) \\
&= P\left(\overline{A_m}\right) \left[1 + \sum_{k=1}^{m} \sum_{\{i_1,\dots,i_k\}} (-1)^{k} P\left(A_{i_1}\right) \ldots P\left(A_{i_k}\right)\right] \\
&= P\left(\overline{A_m}\right) \prod_{j=1}^{m} \left(1 - P\left(A_j\right)\right) \\
&= P\left(\overline{A_m}\right) \prod_{j=1}^{m} P\left(A_j^c\right) \\
&= \prod_{j=m+1}^{n} P\left(A_j\right) \prod_{j=1}^{m} P\left(A_j^c\right). \\
&= \prod_{i=1}^{n} P\left(C_i\right).
\end{aligned}
$$

2. Fall: $m = n$
Hier ist $C_i = A_i^c$ für $i = 1, ..., n$. Nun setzt man für $\overline{A_m} = \Omega$ und argumentiert bis zur drittletzten Zeile genauso!

5.4.6 Korollar

Seien $A_1, ..., A_n$ unabhängig. Dann gilt:

$$
P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \left(1 - P\left(A_i\right)\right) \geq 1 - \exp\left(-\sum_{i=1}^{n} P\left(A_i\right)\right).
$$

Beweis:
Es gilt

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= 1 - P\left(\left(\bigcup_{i=1}^{n} A_i\right)^c\right) \\
&= 1 - P\left(\bigcap_{i=1}^{n} A_i^c\right) \\
&= 1 - \prod_{i=1}^{n} P\left(A_i^c\right) \\
&= 1 - \prod_{i=1}^{n} \left(1 - P\left(A_i\right)\right).
\end{aligned}
$$

Wegen $\exp(-x) \geq 1 - x$ gilt weiter

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &\geq 1 - \prod_{i=1}^{n} \exp\left(-P\left(A_i\right)\right) \\ &= 1 - \exp\left[-\sum_{i=1}^{n} P\left(A_i\right)\right]. \end{aligned}$$

5.4.7 Definition

Eine Folge von Ereignissen $(A_n; n \geq 1)$ heißt unabhängig, falls $A_1, \ldots, A_k$ unabhängig sind für alle $k \geq 1$.

5.4.8 Korollar (Borel-Cantelli Lemma)

Sei $(A_n)_{n\geq 1}$ eine Folge von Ereignissen und $A = \limsup A_n := \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{m=n}^{\infty} A_m$.
Dann gilt:

(1) $\sum\limits_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A) = 0$.

(2) Sind $A_1, A_2, ...$ unabhängig und gilt $\sum\limits_{n=1}^{\infty} P(A_k) = \infty \Rightarrow P(A) = 1$.

Beweis:
Zu (1): Sei $\varepsilon > 0$. Dann gilt wegen $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$ und $A \subset \bigcup\limits_{m=n}^{\infty} A_m$ für $n \geq 1$:
$P(A) \leq P\left(\bigcup\limits_{m=n}^{\infty} A_m\right) \leq \sum\limits_{m=n}^{\infty} P(A_m) < \varepsilon$ für ein hinreichend großes n.
$P(A)$ ist also kleiner als jede positive Zahl. Damit folgt $P(A) = 0$.
Zu (2): Es gilt

$$\begin{aligned} P(A) &= \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} A_m\right) \\ &= \lim_{n\to\infty} \lim_{p\to\infty} P\left(\bigcup_{m=n}^{n+p} A_m\right) \\ &\geq \lim_{n\to\infty} \lim_{p\to\infty} \left(1 - \underbrace{\exp\left(-\sum_{m=n}^{n+p} P(A_m)\right)}_{\to 0 \text{ für } p\to\infty}\right) \\ &= 1 \end{aligned}$$

Folglich ist $P(A) = 1$.

Was ist die Wahrscheinlichkeit, daß im k-ten Wurf erstmals eine “6” gewürfelt wird.

Erstmals im k-ten Wurf eine "6" zu werfen, bedeutet, in den vorangegangenen $k-1$ Würfen keine "6" , dann aber eine "6" zu werfen. Folglich ist

$$P(\text{erstmals "6" im } k\text{-ten Wurf}) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} =: p_k, \quad k \geq 1$$

Es muß natürlich $\sum\limits_{k\geq 1} p_k = 1$ gelten, was man leicht mit Hilfe der geometrische Reihe $\sum\limits_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} = 6$ einsieht.

Gamblers Rules
Angenommen man spielt ein Spiel sehr oft hintereinander und dessen Gewinnchance ist $1/N$. Wie oft muß man spielen, damit man mit mindestens 50% Wahrscheinlichkeit wenigstens einmal gewinnt?

$$\begin{aligned} P(\text{kein Gewinn in } n \text{ Spielen}) &= \left(1 - \frac{1}{N}\right)^n \\ P(\text{Gewinn in } n \text{ Spielen}) &= 1 - \left(1 - \frac{1}{N}\right)^n \geq \frac{1}{2} \\ &\Leftrightarrow \left(1 - \frac{1}{N}\right)^n \leq \frac{1}{2} \\ &\Leftrightarrow n \log\left(1 - \frac{1}{N}\right) \leq \log\frac{1}{2}. \end{aligned}$$

Sei $n^* = [\log(\frac{1}{2}) \,/\, \log(1 - \frac{1}{N})]$; dabei ist $[x]$ die kleinste ganze Zahl größer als x für $x \in \mathbb{R}$. Da $\log(1+z) \sim z$ für $z \to 0$ gilt, ist die rechte Seite von n^* asymptotisch gleich

$$\log\left(\frac{1}{2}\right) \Big/ \left(\frac{-1}{N}\right) = N \log 2, \quad \text{wobei} \quad \log(2) \approx 0,69 \approx 2/3.$$

Im Fall des Würfels benötigt man also

$$n^* = \left[\log\left(\frac{1}{2}\right) \Big/ \left(\frac{5}{6}\right)\right] = [3,8] = 4$$

Würfe um mit mindestens 50% Wahrscheinlichkeit eine "6" zu werfen.

Beispiel: Fluß in einem Leiter
Angenommen für jeden von den Schaltern in dem folgenden Schaltkreis ist die Wahrscheinlichkeit, daß der Schalter geschlossen ist p_i und daß er offen ist $q_i = 1 - p_i$, $i = 1, \ldots, 5$. Man berechne die Wahrscheinlichkeit, daß ein Strom durch den Schaltkreis fließt unter der Annahme, daß die Zustände der Schalter unabhängig sind.

$$\begin{aligned} P(\text{Strom fließt}) = {} & P(\text{Strom fließt oben entlang}) + P(\text{Strom fließt unten entlang}) \\ & -P(\text{Strom fließt sowohl oben als auch unten entlang }) \end{aligned}$$

Dabei ist wegen der Unabhängigkeit $P(\text{Strom fließt oben}) = p_1 \cdot p_2$, $P(\text{Strom fließt unten}) = p_3 \cdot p_4 \cdot p_5$ und $P(\text{Strom fließt oben und unten}) = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5$ und damit

$$P(\text{Strom fließt }) = p_1 \cdot p_2 + p_3 \cdot p_4 \cdot p_5 - p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5.$$

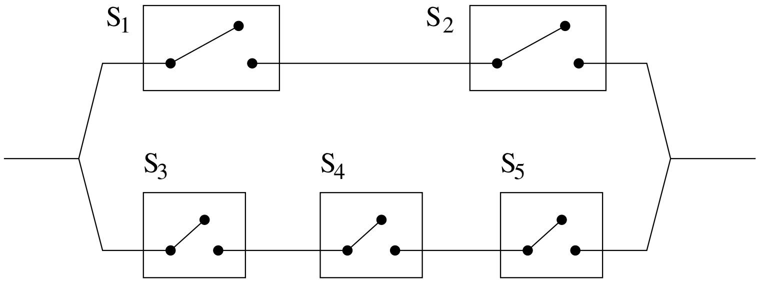


Abbildung 5: Schaltkreis

5.5 Anwendung der Unabhängigkeit in der Zahlentheorie

5.5.1 Primzahlen und Unabhängigkeit

Sei $N \in \mathbb{N}$. Dann gilt $N = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ für geeignete Primzahlen p_i und $\alpha_i \in \mathbb{N}$.
Sei $\varphi(N) = \#\{i \in \mathbb{N} | i < N \text{ mit } \mathrm{GGT}(i, N) = 1\}$ die Anzahl der natürlichen Zahlen, die kleiner N und zu N teilerfremd sind.

Beispiel:

N	2	3	4	5	6	7	8	9	10
$\varphi(N)$	1	2	2	4	2	6	4	6	4

Die „1“ zählt stets mit!

Behauptung: $\varphi(N) = N \cdot \prod_{j=1}^{k} (1 - \frac{1}{p_j})$ (Eulersche Funktion).

Wir übersetzen die Aufgabenstellung in die Sprache der Wahrscheinlichkeitsrechnung:
Definiere $\Omega_N := \{i \in \mathbb{N} | 1 \leq i \leq N\} = \{1, \ldots, N\}$. Für $A \subset \Omega_N$ sei $P(A) := \frac{\#A}{\#\Omega_N}$.
Sei $A_i = \{m \in \Omega_N | p_i \text{ teilt } m\}$ die Teilmenge der Zahlen aus Ω_N, die durch p_i teilbar sind.

Dann ist $\#A_i = \frac{N}{p_i}$ (Bemerkung: $\frac{N}{p_i}$ ist eine natürliche Zahl, da p_i ein Faktor in der Primzahlzerlegung von N ist.) und $P(A_i) = \frac{N/p_i}{N} = \frac{1}{p_i}$.
Weiter gilt: $P(A_{i_1} \cap ... \cap A_{i_l}) = \frac{1}{p_{i_1} \cdots p_{i_l}} = \prod_{j=1}^{l} \frac{1}{p_{i_j}} = \prod_{j=1}^{l} P(A_{i_j})$, denn $|A_{i_1} \cap \ldots \cap A_{i_l}| = \frac{N}{p_{i_1} \cdots p_{i_l}}$.
Die Ereignisse $A_1, \ldots, A_k$ sind also unabhängig. Folglich gilt
$P(\text{Zahl} \leq N \text{ ist teilerfremd zu } N) = P(A_1^c \cap \ldots \cap A_k^c) = \prod_{j=1}^{k} P(A_j^c) = \prod_{j=1}^{k} (1 - \frac{1}{p_j})$.

Und damit folgt die Behauptung $\varphi(N) = N \cdot \prod_{j=1}^{k} (1 - \frac{1}{p_j})$.

5.5.2 2-dimensionaler Fall

Wähle zufällig 2 Zahlen $\leq N$. Dabei soll „zufällig“ heißen, daß jedes Paar (i, j) mit derselben Wahrscheinlichkeit $\frac{1}{N^2}$ gewählt wird. Wie groß ist die Wahrscheinlichkeit, daß diese beiden Zahlen teilerfremd sind?

Der Grundraum ist $\Omega_N^2 = \{(i,j) | i,j \in \mathbb{N}, \max(i,j) \leq N\}$. Für $A \subset \Omega_N^2$ sei $P(A) = \frac{\#A}{\#N}$. Da $N \in \mathbb{N}$ ist, gibt es eindeutig bestimmte Primzahlen p_i und $\alpha_i \in \mathbb{N}$ mit $N = p_1^{\alpha_1} \dots p_k^{\alpha_k}$. Sei $A_m := \{(i,j) | p_m \text{ teilt } i \text{ und } p_m \text{ teilt } j\} \cap \Omega_N$. Dann ist $\#A_m = (\frac{N}{p_m})^2$ und $P(A_m) = \frac{1}{p_m^2}$. Ebenso wie im 1-dimensionalen Fall folgt: Die $A_1, \dots, A_k$ sind unabhängig und

$$P(\text{Zahlenpaar} \leq N \text{ ist teilerfremd}) = P(A_1^c \cap ... \cap A_k^c) = \prod_{j=1}^{k} P(A_j^c) = \prod_{j=1}^{k} (1 - \frac{1}{p_j^2}).$$

Betrachten wir nun den Grenzübergang $N \to \infty$:

$\lim\limits_{N\to\infty} P(\text{Zahlenpaar} \leq N \text{ ist teilerfremd}) = \lim\limits_{k\to\infty} \prod\limits_{j=1}^{k} (1 - \frac{1}{p_j^2}) = \frac{6}{\pi^2}$.

Dabei folgt das Ergebnis aus dem folgenden Lemma:

5.5.3 Lemma

Für $s > 1$ sei $\zeta(s) = \sum\limits_{n \geq 1} n^{-s}$. Dann gilt

$$\zeta(s) = \lim_{N\to\infty} \prod_{\substack{p \leq N \\ p \text{ prim}}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Insbesondere ist $\zeta(2) = \lim\limits_{N\to\infty} \prod\limits_{\substack{p \leq N \\ p \text{ prim}}} (1 - \frac{1}{p^2})^{-1} = \sum\limits_{n \geq 1} n^{-2} = \frac{\pi^2}{6}$.

Beweis:
Das folgende Produkt ist ein Produkt geometrischer Reihen.

$$\prod_{\substack{p \leq N \\ p \text{ prim}}} \left(1 - \frac{1}{p^s}\right)^{-1} = \prod_{\substack{p \leq N \\ p \text{ prim}}} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots\right) = \sum_{(\alpha_1,\dots,\alpha_l)} \frac{1}{p_1^{s\alpha_1} \cdots p_l^{s\alpha_l}}$$

Dabei sind p_i die Primzahlen mit $p_1 < p_2 < \dots < p_l \leq N < p_{l+1}$. Nun gilt aber weiter, da sich jedes $n \leq N$ als Produkt von Potenzen von p_i angeben läßt, daß obige Summe größer gleich $\sum\limits_{n \leq N} \frac{1}{n^s}$ ist. Andererseits ist diese Summe kleiner gleich $\zeta(s)$. Folglich gilt

$$0 \leq \zeta(s) - \prod_{\substack{p \leq N \\ p \text{ prim}}} \left(1 - \frac{1}{p^s}\right) \leq \sum_{n > N} \frac{1}{n^s}.$$

Da für $s > 1$ die rechte Seite für $N \to \infty$ gegen 0 konvergiert, folgt die Behauptung.

5.5.4 Bemerkung

Läßt sich das oben definierte P im Fall $N \to \infty$ als Wahrscheinlichkeitsmaß auf $\mathbb{N} \times \mathbb{N}$ interpretieren?

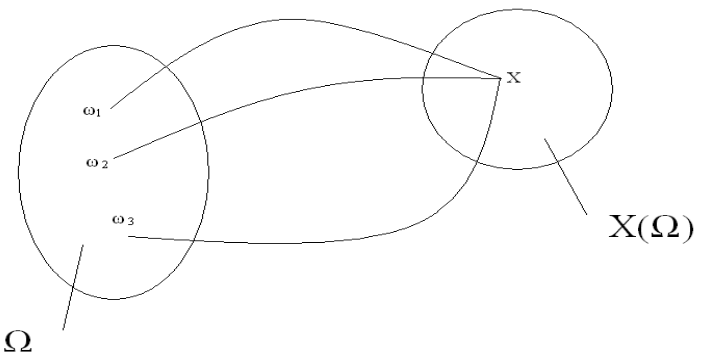
Antwort: Nein! Denn sei für $A \subset \mathbb{N} \times \mathbb{N}$
$Q(A) = \lim\limits_{N\to\infty} P(A \cap \Omega_N)$, so ist Q kein Wahrscheinlichkeitsmaß, da für jedes Paar $(i,j) \in \mathbb{N}^2$ gilt: $Q(\{(i,j)\}) = \lim_{N\to\infty} \frac{1}{N^2} = 0$ und damit $\sum\limits_{(i,j)} Q(\{(i,j)\}) = 0$ im Widerspruch zu $Q(\mathbb{N} \times \mathbb{N}) = 1$.

6 Zufallsvariable und ihre Verteilung

6.1 Zufallsvariable, Verteilung einer Zufallsvariable

6.1.1 Definition (Verteilung)

Sei (Ω, P) ein Wahrscheinlichkeitsraum. Eine Zufallsvariable ist eine Abbildung $X : \Omega \to \mathbb{R}$. Durch $q(x) := P\{\omega | X(\omega) = x\}$ wird eine Wahrscheinlichkeitsfunktion auf $X(\Omega) \subset \mathbb{R}$ definiert. Das zugehörige Wahrscheinlichkeitsmaß $Q : \mathcal{P}(X(\Omega)) \to [0,1]$, $Q(A) = \sum_{x \in A} q(x)$ heißt **Verteilung von** X. Man schreibt auch P^X für Q.



Sei $p(x) = P(\{x\})$ die Wahrscheinlichkeitsfunktion von P. Für $x \in X(\Omega)$ ist im obigen Bild $q(x) = p(w_1) + p(w_2) + p(w_3)$.

q ist Wahrscheinlichkeitsfunktion, da

$$\begin{aligned}
\sum_{x \in X(\Omega)} q(x) &= \sum_{x \in X(\Omega)} P\left(\{\omega | X(\omega) = x\}\right) \\
&= \sum_{x \in X(\Omega)} \sum_{\omega | X(\omega) = x} p(\omega) \\
&= \sum_{\omega \in \Omega} p(\omega) = 1.
\end{aligned}$$

Für $A \subset X(\Omega)$ ist

$$\begin{aligned}
Q(A) &= \sum_{x \in A} q(x) \\
&= \sum_{x \in A} P\left(\{\omega | X(\omega) = x\}\right) \\
&= P\left(\{\omega | X(\omega) \in A\}\right) \\
&= P\left(X^{-1}(A)\right).
\end{aligned}$$

6.1.2 Beispiel

Wir betrachten die Summe der Augenzahlen zweier Würfelwürfe. Bei zwei Würfelwürfen ist der Grundraum $\Omega_2 = \{\omega = (\omega_1, \omega_2) | \omega_i \in \{1, \ldots, 6\}, i = 1, 2\}$. Wir können nun jedem $\omega = (\omega_1, \omega_2) \in \Omega_2$ eine reelle Zahl $X(\omega) = \omega_1 + \omega_2$ zuordnen. Dann ist X eine Abbildung von $\Omega_2 \to \mathbb{R}$ und damit eine Zufallsvariable.

Für die Wahrscheinlichkeitsfunktion und das Wahrscheinlichkeitsmaß auf Ω_2 gilt: $p(\omega) = \frac{1}{\#\Omega_2}$ und $P(A) = \frac{\#A}{\#\Omega_2}$.

Die zu X gehörende Wahrscheinlichkeitsfunktion $q : \mathbb{R} \to [0, 1]$ hat folgende Gestalt: $q(k) = P(\{\omega \in \Omega_2 | X(\omega) = k\} = P(\{\omega = (\omega_1, \omega_2) \in \Omega_2 | \omega_1 + \omega_2 = k\} = \frac{\#\{\omega|\omega_1+\omega_2=k\}}{\#\Omega_2}$.

$$
\begin{aligned}
q(2) &= \frac{\#\{\omega|\omega_1 + \omega_2 = 2\}}{\#\Omega_2} = \frac{\#\{(1,1)\}}{36} = \frac{1}{36} \\
q(3) &= \frac{\#\{\omega|\omega_1 + \omega_2 = 3\}}{\#\Omega_2} = \frac{\#\{(1,2),(2,1)\}}{36} = \frac{2}{36} \\
&\vdots \\
q(7) &= \ldots = \frac{6}{36} \\
&\vdots \\
q(11) &= \frac{\#\{\omega|\omega_1 + \omega_2 = 11\}}{\#\Omega_2} = \frac{\#\{(5,6),(6,5)\}}{36} = \frac{2}{36} \\
q(12) &= \frac{\#\{\omega|\omega_1 + \omega_2 = 12\}}{\#\Omega_2} = \frac{\#\{(6,6)\}}{36} = \frac{1}{36}
\end{aligned}
$$

Sei $A \subset \mathbb{R}$. Dann gilt für die Verteilung von X: $Q(A) = \sum_{k \in A} q(k)$. Sei beispielsweise $A = \{k \in \mathbb{R} | k \leq 3\}$. So ist $Q(A) = \sum_{k \leq 3} q(k) = \sum_{k=2}^{3} q(k) = \frac{1}{12}$.

Bemerkungen

1. Die Verteilung von X kann auch als Maß auf $\mathbb{R}$ angesehen werden. Man setzt

$$P^X(A) := P^X(A \cap X(\Omega)) = P\left(X^{-1}(A)\right) \quad \text{für } A \subset \mathbb{R}.$$

   P^X ist aber ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$. Denn

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

   und

$$P^X\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P^X(A_i)$$

für $A_i \subset \mathbb{R}$, paarweise disjunkt. Denn:

$$\begin{aligned} P^X\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(X^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{\infty} X^{-1}(A_i)\right) \\ &= \sum_{i=1}^{\infty} P(X^{-1}(A_i)) \\ &= \sum_{i=1}^{\infty} P^X(A_i). \end{aligned}$$

2. Sei eine Wahrscheinlichkeitsfunktion p auf einer diskreten Teilmenge $W = \{x_1, x_2, \ldots\} \subset \mathbb{R}$ gegeben. Durch $X : W \to \mathbb{R}$ mit $X(x_i) := x_i$ wird eine Zufallsvariable auf W erklärt, deren Verteilung P^X die Wahrscheinlichkeitsfunktion p hat. Denn:
$$q(x_i) = P^X(\{x_i\}) = P(\{z \in W | X(z) = x_i\}) = p(x_i).$$

3. Wegen 2) spricht man von einer Verteilung, wenn eine Wahrscheinlichkeitsfunktion p auf einer diskreten Teilmenge von $\mathbb{R}$ gegeben wird.

6.1.3 Beispiele von Verteilungen

1) **Binomial-Verteilung** $b(n, p)$
$b(n, p; k) = \binom{n}{k} p^k (1-p)^{n-k}$ definiert eine Wahrscheinlichkeitsfunktion auf $\{0, 1, ..., n\}$.

2) **Bernoulliverteilung** $b(1, p)$
$b(1, p; k) = p^k (1-p)^{1-k}$ für $k \in \{0, 1\}$.
Die Bernoulliverteilung ist ein Spezialfall der Binomialverteilung für $n = 1$.

3) **Poisson-Verteilung** $\mathrm{pois}(\lambda)$
$\mathrm{pois}(\lambda; k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ist eine Wahrscheinlichkeitsfunktion auf $\mathbb{N} \cup \{0\}$.

4) **Pascal-Verteilung** $\mathrm{pasc}(r, p)$
$\mathrm{pasc}(r, p; n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$ für $n \in \{r, r+1, r+2, ...\}$, $r \in \mathbb{N}$.

5) **Geometrische Verteilung**
Speziell: Für $r = 1$ ergibt sich $\mathrm{pasc}(1, p; n) = p(1-p)^{n-1}$.

6.2 Unabhängigkeit von Zufallsvariablen

6.2.1 Definition (Unabhängigkeit von Zufallsvariablen)

$X_1, X_2, ..., X_n$ seien Zufallsvariablen auf (Ω, P) und
$X_i(\Omega)$ sei der Wertebereich von X_i für $i = 1, ..., n$.

$X_1, ..., X_n$ heißen unabhängig, falls für alle $z_i \in X_i(\Omega)$ für $i = 1, \ldots, n$ gilt:

$$P\left(\bigcap_{i=1}^{n}\{X_i = z_i\}\right) = \prod_{i=1}^{n} P\left(\{X_i = z_i\}\right).$$

Dabei haben wir die **Kurzschreibweise**: $\{X_i = z_i\} = \{\omega \in \Omega | X_i(\omega) = z_i\}$ verwendet.

6.2.2 Satz

$X_1, .., X_n$ seien Zufallsvariablen auf (Ω, P). Dann sind folgende Aussagen äquivalent:

(1) $X_1, ..., X_n$ sind unabhängig.
(2) Für alle $A_1, \ldots, A_n \subset \mathbb{R}$ gilt: $P\left(\bigcap_{i=1}^{n}\{X_i \in A_i\}\right) = \prod_{i=1}^{n} P\left(\{X_i \in A_i\}\right)$.

Dabei bedeutet $\{X_i \in A_i\} := \{\omega \in \Omega | X_i(\omega) \in A_i\}$.

Beweis:
(2) $\Rightarrow$ (1) trivial.
(1) $\Rightarrow$ (2) Seien $A_1, ..., A_n$ gegeben.
Da nur die Elemente aus $X_i(\Omega) \cap A_i$ eine Wahrscheinlichkeit ≥ 0 haben und $X_i(\Omega)$ abzählbar ist, können wir o.B.d.A. annehmen, daß die A_i abzählbar sind und die Form $A_i = \{y_{i1}, y_{i2}, \ldots\}$ haben. Damit können wir die Mengen $\{X_i \in A_i\}$ folgendermaßen zerlegen: $\{X_i \in A_i\} = \bigcup_j \{X_i = y_{ij}\}$.

Somit gilt

$$\bigcap_{i=1}^{n}\{X_i \in A_i\} = \bigcap_{i=1}^{n}\bigcup_{j_i}\{X_i = y_{ij_i}\} = \bigcup_{j_1,\ldots,j_n} \{X_1 = y_{1j_1}, X_2 = y_{2j_2}, ..., X_n = y_{nj_n}\}$$

und damit

$$\begin{aligned}
P\left(\bigcap_{i}\{X_i \in A_i\}\right) &= \sum_{j_1,\ldots,j_n} P\left(\{X_1 = y_{1j_1}, X_2 = y_{2j_2}, ..., X_n = y_{nj_n}\}\right) \\
&= \sum_{j_1,\ldots,j_n} \prod_{i=1}^{n} P\left(X_i = y_{ij_i}\right) \\
&= \prod_{i=1}^{n}\left(\sum_{j_i} P\left(\{X_i = y_{ij_i}\}\right)\right) \\
&= \prod_{i=1}^{n} P\left(\{X_i \in A_i\}\right).
\end{aligned}$$

6.2.3 Beispiel (n-facher Münzenwurf einer p-Münze)

Beim n-fachen Münzwurf ist der Grundraum $\Omega = \{\omega = (\omega_1, ..., \omega_n) | \omega_i \in \{0,1\}\}$ und das dazugehörige Wahrscheinlichkeitsmaß $P(\{(\omega_1, ..., \omega_n)\}) = p^{\sum \omega_i}(1-p)^{n-\sum \omega_i}$.
Wir definieren die Zufallsvariablen $X_i : \Omega \to \{0,1\},\ X_i(\omega) = \omega_i$ für $i = 1, \ldots, n$.

Behauptung: $X_1, ..., X_n$ sind unabhängig.

Beweis: Sei $(\omega_1', ..., \omega_n') \in \Omega$ beliebig. Dann gilt:

$$P\left(\bigcap_{i=1}^{n} \{X_i = \omega_i'\}\right) = P\left(\{(\omega_1', ..., \omega_n')\}\right) = p^{\sum \omega_i'}(1-p)^{n-\sum \omega_i'} = \prod_{i=1}^{n} P(\{X_i = \omega_i'\})$$

und damit sind $X_1, ..., X_n$ unabhängig.

Sei nun $S_n = \sum\limits_{i=1}^{n} X_i$. Dann gilt:

$$\begin{aligned}
P(S_n = k) &= P\left(\left\{\sum_{i=1}^{n} X_i = k\right\}\right) \\
&= P\left(\left\{\omega \,\middle|\, \sum_{i=1}^{n} X_i(\omega) = k\right\}\right) \\
&= \sum_{\omega | \sum \omega_i = k} P\left(\{(\omega_1, ..., \omega_n)\}\right) \\
&= \sum_{\omega | \sum \omega_i = k} p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

Sind also die X_i Bernoulli-verteilt, d.h. $P(X_i = \omega_i) = p^{\omega_i}(1-p)^{1-\omega_i},\ \omega_i \in \{0,1\}$, so ist die Summe S_n der X_i nach $b(n,p)$-verteilt.

Allgemeiner gilt: Seien $X_1, ..., X_n$ unabhängig und $b(1,p)$ verteilt und $Y_1, ..., Y_m$ ebenfalls unabhängig und $b(1,p)$ verteilt. Dann ist $\sum\limits_{i=1}^{n} X_i + \sum\limits_{j=1}^{m} Y_i$ $b(n+m,p)$-verteilt.

6.2.4 Satz (Konstruktion eines Wahrscheinlichkeitsraumes aus n unabhängigen Zufallsgrößen zu vorgegebenen Verteilungen)

Seien (Ω_i, P_i) Wahrscheinlichkeitsräume mit $\Omega_i \subset \mathbb{R}$ für $i = 1, \ldots, n$ und p_i, die zu den Wahrscheinlichkeitsmaßen P_i gehörigen Wahrscheinlichkeitsfunktionen.
Weiter sei $\Omega = \Omega_1 \times \ldots \times \Omega_n = \{\omega = (\omega_1, ..., \omega_n) | \omega_i \in \Omega_i, i = 1, ..., n\}$ und $p(\omega) = \prod\limits_{i=1}^{n} p_i(\omega_i)$.
Seien $X_i(\omega) = \omega_i$ für $i = 1, ..., n$ Zufallsvariablen und sei P das zu p gehörige Wahrscheinlichkeitsmaß. Dann gilt:

a) $p(\omega)$ ist eine Wahrscheinlichkeitsfunktion auf Ω.

b) Die Zufallsvariable X_i ist nach P_i verteilt.

c) $X_1, ..., X_n$ sind unabhängig.

Beweis:

zu (a):
$$\sum_{\omega\in\Omega} p(\omega) = \sum_{(\omega_1,\dots,\omega_n)} \prod_{i=1}^{n} p_i(\omega_i) = \sum_{\omega_1} \dots \sum_{\omega_n} \prod_{i=1}^{n} p_i(\omega_i)$$
$$= \left(\sum_{\omega_1} p_1(\omega_1)\right) \dots \left(\sum_{\omega_n} p_n(\omega_n)\right) = 1$$

zu (b):
$$\begin{aligned}
\{X_i \in A_i\} &= \{\omega | X_i(\omega) \in A_i\} \\
&= \{\omega | \omega_i \in A_i\} \\
&= \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n \\
P(\{X_i \in A_i\}) &= P(\Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n) \\
&= P_1(\Omega_1) \cdots P_{i-1}(\Omega_{i-1}) \cdot P_i(A_i) \cdot P_{i+1}(\Omega_{i+1}) \cdots P_n(\Omega_n) \\
&= P_i(A_i)
\end{aligned}$$

zu (c):
$$\begin{aligned}
\bigcap_{i=1}^{n} \{X_i \in A_i\} &= \{\omega | X_i(\omega) \in A_i, i = 1, \dots, n\} \\
&= \{\omega | \omega_i \in A_i, i = 1, \dots, n\} \\
&= A_1 \times A_2 \times \dots \times A_n \\
P(\bigcap_{i=1}^{n} \{X_i \in A_i\}) &= P(A_1 \times ... \times A_n) \\
&= \sum_{\omega \in A_1 \times ... \times A_n} p(\omega) \\
&= \sum_{\omega \in A_1 \times ... \times A_n} \prod_{i=1}^{n} p_i(\omega_i) \\
&= \left(\sum_{\omega_1 \in A_1} p_1(\omega_1)\right) \dots \left(\sum_{\omega_n \in A_n} p_n(\omega_n)\right) \\
&= P_1(A_1) \dots P_n(A_n) \\
&= \prod_{i=1}^{n} P(\{X_i \in A_i\})
\end{aligned}$$

6.2.5 Satz (Summe von Binomialverteilungen, Poissonverteilungen, Pascalverteilungen)

Seien X_1 und X_2 unabhängige Zufallsvariablen.

a) Sind die X_i Binomial-verteilt nach $b(n_i, p)$ für $i = 1, 2$, dann ist $X_1 + X_2$ Binomial-verteilt nach $b(n_1 + n_2, p)$.

b) Sind die X_i Poisson-verteilt mit Parameter λ_i für $i = 1, 2$, dann ist $X_1 + X_2$ Poisson-verteilt mit Parameter $\lambda_1 + \lambda_2$.

Beweis:

zu (a):
$$\begin{aligned}
P(\{X_1 + X_2 = l\}) &= P\Big(\bigcup_{i=0}^{l} \{X_1 = i, X_2 = l - i\}\Big)\\
&= \sum_{i=0}^{l} P(\{X_1 = i, X_2 = l - i\})\\
&= \sum_{i=0}^{l} P(\{X_1 = i\})P(\{X_2 = l - i\})\\
&= \sum_{i=0}^{l} \binom{n_1}{i} p^i (1-p)^{n_1 - i} \binom{n_2}{l-i} p^{l-i}(1-p)^{n_2 - l + i}\\
&= \sum_{i=0}^{l} p^l (1-p)^{n_1 + n_2 - l} \binom{n_1}{i}\binom{n_2}{l-i}\\
&= p^l (1-p)^{n_1 + n_2 - l} \sum_{i=0}^{l} \binom{n_1}{i}\binom{n_2}{l-i}\\
&= p^l (1-p)^{n_1 + n_2 - l} \binom{n_1 + n_2}{l}
\end{aligned}$$

zu (b):
$$\begin{aligned}
P(\{X_1 + X_2 = l\}) &= P\Big(\bigcup_{i=0}^{l} \{X_1 = i, X_2 = l - i\}\Big)\\
&= \sum_{i=0}^{l} P(\{X_1 = i\})P(\{X_2 = l - i\})\\
&= \sum_{i=0}^{l} \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{l-i}}{(l-i)!} e^{-\lambda_2}\\
&= e^{-(\lambda_1 + \lambda_2)} \sum_{i=0}^{l} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{l-i}}{(l-i)!}\\
&= e^{-(\lambda_1 + \lambda_2)} \frac{1}{l!} \sum_{i=0}^{l} \binom{l}{i} \lambda_1^i \lambda_2^{l-i}, \quad \text{da } \frac{l!}{i!(l-i)!} = \binom{l}{i}\\
&= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^l}{l!} \qquad \text{(binomische Formel)}
\end{aligned}$$

7 Erwartungswert und Varianz von Verteilungen

7.1 Der Erwartungswert

7.1.1 Beispiele

1) Spiel mit einem Würfel: Der Gewinn sei „i“, falls der Würfel „i“ ergibt. Was ist ein fairer Einsatz?
   Antwort: $\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + ... + \frac{1}{6} \cdot 6 = 3,5$.

2) Sei n_i die Anzahl der Familien mit „i“ Kindern.
   Dann ist $n = n_0 + n_1 + ... + n_{18}$ die Anzahl der Familien,
   und $m = n_1 + 2n_2 + ... + 18n_{18}$ die Anzahl der Kinder.
   Mittlere Anzahl der Kinder pro Familie ist $\frac{m}{n}$.

7.1.2 Definition (Erwartungswert)

Sei (Ω, P) ein Wahrscheinlichkeitsraum, p die zu P gehörige Wahrscheinlichkeitsfunktion und X eine Zufallsgröße auf Ω.

$$E(X) := \sum_{\omega \in \Omega} X(\omega)p(\omega)$$

heißt Erwartungswert von X, falls $X \geq 0$ oder $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$ gilt.

7.1.3 Bemerkungen

(1) Der Erwartungswert ist der mittlere Wert einer Zufallsvariablen bzw. ihrer Verteilung.

(2) Sei $f : \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann ist $f \circ X$ eine Zufallsgröße und wenn ihr Erwartungswert existiert, so ist dieser $E(f \circ X) = \sum_{\omega \in \Omega} f(X(\omega))p(\omega)$.

(3) $EX = \infty$ ist möglich, falls $X \geq 0$ ist! Siehe Beispiel 7.2.4.

(4) Falls $\sum_{\omega \in \Omega} |X(\omega)|p(\omega) < \infty$, so gilt $E|X| = \sum_{\omega \in \Omega} |X(\omega)|p(\omega)$.
Dies sieht man so: Da auf der rechten Seite eine absolut konvergente Reihe steht, kann man die Reihe $\sum_{\omega \in \Omega} X(\omega)p(\omega)$ umordnen. Es gilt
$E(X) = \sum_{\omega} X^+(\omega)p(\omega) - \sum_{\omega} X^-(\omega)p(\omega) = E(X^+) - E(X^-)$,
wobei $X^+(\omega) := X(\omega) \vee 0$ und $X^-(\omega) := (-X(\omega) \vee 0)$ sind. Es gilt außerdem $X^{\pm}(\omega) \geq 0$. Da $|X(\omega)| = X^+(\omega) + X^-(\omega)$ ist, gilt auch $E|X| = E(X^+) + E(X^-)$.

7.1.4 Eigenschaften des Erwartungswertes

(1) $E(\alpha \cdot X) = \alpha \cdot E(X)$.

(2) Für $A \subseteq \Omega$ gilt: $E(1_A) = P(A)$, wobei $1_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$.

(3) $|EX| \leq E|X|$

(4) Sind $E|X|$ und $E|Y|$ kleiner unendlich, so gilt $E(X+Y) = E(X) + E(Y)$.

(5) Sind X, Y Zufallsvariablen mit $P(\{\omega | X(\omega) \leq Y(\omega)\}) = 1$, so gilt: $E(X) \leq E(Y)$.

Beweis:
Zu (1): $E(\alpha X) = \sum\limits_{\omega\in\Omega} \alpha X(\omega)p(\omega) = \alpha \sum\limits_{\omega\in\Omega} X(\omega)p(\omega) = \alpha E(X)$.
Zu (2): $E(1_A) = \sum\limits_{\omega\in\Omega} 1_A(\omega)p(\omega) = \sum\limits_{\omega\in A} p(\omega) = P(A)$.
Zu (3): $|EX| = |E(X^+) - E(X^-)| \leq |EX^+| + |EX^-| = E(X^+ + X^-) = E|X|$.
Zu (4):

$$\begin{aligned} E(X+Y) &= \sum_{\omega\in\Omega} (X(\omega) + Y(\omega))p(\omega) \\ &= \sum_{\omega\in\Omega} X(\omega)p(\omega) + \sum_{\omega\in\Omega} Y(\omega)p(\omega) \\ &= E(X) + E(Y) \end{aligned}$$

Zu (5):

$$\begin{aligned} E(X) &= \sum_{\omega\in\Omega} X(\omega)p(\omega) \\ &= \sum_{\omega\in\{\omega|X(\omega)\leq Y(\omega)\}} X(\omega)p(\omega) + \sum_{\omega\in\{\omega|X(\omega)>Y(\omega)\}} X(\omega)p(\omega) \\ &\overset{(*)}{\leq} \sum_{\omega\in\{\omega|X(\omega)\leq Y(\omega)\}} Y(\omega)p(\omega) + \sum_{\omega\in\{\omega|X(\omega)>Y(\omega)\}} X(\omega)\cdot 0 \\ &= \sum_{\omega\in\{\omega|X(\omega)\leq Y(\omega)\}} Y(\omega)p(\omega) + \sum_{\omega\in\{\omega|X(\omega)>Y(\omega)\}} Y(\omega)p(\omega) \\ &= \sum_{\omega\in\Omega} Y(\omega)p(\omega) \\ &= E(Y) \end{aligned}$$

(*) $P(X \leq Y) = 1$, und damit $P(X > Y) = 0$
und schließlich $p(\omega) = 0$ auf $\{\omega \mid X(\omega) > Y(\omega)\}$.

7.1.5 Folgerung (Poincarés Ein- und Ausschlußformel)

Unter Verwendung des Erwartungswertes läßt sich Poincarés Ein- und Ausschlußformel $P(\bigcup\limits_{k=1}^{n} A_k) = \sum\limits_{k=1}^{n}(-1)^{k-1} \sum\limits_{i_1,\dots,i_k} P(A_{i_1} \cap \dots \cap A_{i_k})$ ganz einfach beweisen.

Beweis:

$$\begin{aligned}
1 - P(\bigcup_{k=1}^{n} A_k) &= P((\bigcup_{k=1}^{n} A_k)^c) \\
&= P(\bigcap_{k=1}^{n} A_k^c) \\
&= E(1_{\bigcap\limits_{k=1}^{n} A_k^c}) \\
&= E(\prod_{k=1}^{n} 1_{A_k^c}) \\
&= E(\prod_{k=1}^{n}(1 - 1_{A_k})) \\
&= E(1 + \sum_{k=1}^{n} \sum_{i_1,\dots,i_k} (-1)^k \prod_{j=1}^{k} 1_{A_{i_j}}) \\
&= 1 + \sum_{k=1}^{n} \sum_{i_1,\dots,i_k} (-1)^k E(1_{A_{i_1} \cap \dots \cap A_{i_k}}) \\
&= 1 + \sum_{k=1}^{n} \sum_{i_1,\dots,i_k} (-1)^k P(A_{i_1} \cap \dots \cap A_{i_k}) \\
&= 1 - \sum_{k=1}^{n} \sum_{i_1,\dots,i_k} (-1)^{k-1} P(A_{i_1} \cap \dots \cap A_{i_k})
\end{aligned}$$

7.1.6 Satz (Transformationssatz)

Sei (Ω, P) ein Wahrscheinlichkeitsraum und X eine Zufallsvariable mit der Wertemenge $X(\Omega) = \{x_1, x_2, \dots\}$. Für $x \in X(\Omega)$ sei $q(x) = P(\{\omega | X(\omega) = x\})$. Weiter sei $f : \mathbb{R} \to \mathbb{R}$ und es gelte $f \geq 0$ oder $E|f(X)| < \infty$. Dann gilt:

a) $E(f(X)) = \sum\limits_{i=1}^{\infty} f(x_i) q(x_i)$.

b) Ist Y eine weitere Zufallsvariable auf (Ω, P) mit $Y(\Omega) = \{y_1, y_2, \dots\}$. Sei $g : \mathbb{R}^2 \to \mathbb{R}$ und gelte $g \geq 0$ oder $E|g(X, Y)| < \infty$. Dann gilt

$$Eg(X,Y) = \sum_{i,j} g(x_i, y_j) q(x_i, y_j) \qquad \text{mit } q(x_i, y_j) = P(X = x_i, Y = y_j).$$

Beweis:
zu a) $$\begin{aligned} E(f(X)) &= \sum_{\omega\in\Omega} f(X(\omega))p(\omega) \\ &= \sum_{i=1}^{\infty} \sum_{\{\omega|X(\omega)=x_i\}} f(X(\omega))p(\omega) \\ &= \sum_{i=1}^{\infty} f(x_i) \sum_{\{\omega|X(\omega)=x_i\}} p(\omega) \\ &= \sum_{i=1}^{\infty} f(x_i)P(\{\omega|X(\omega)=x_i\}) \\ &= \sum_{i=1}^{\infty} f(x_i)q(x_i) \end{aligned}$$

zu b) Der Beweis geht entsprechend wie a).

7.2 Beispiele von Erwartungswerten

7.2.1 Binomial-Verteilung

Seien $X_1, ..., X_n$ Zufallsvariablen mit $P(X_i = 1) = p_i = 1 - P(X_i = 0)$ für $i = 1, ..., n$.
Dann gilt : $E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = p_i$.
Außerdem gilt: $E(X_1 + X_2 + ... + X_n) = E(X_1) + ... + E(X_n) = \sum_{i=1}^{n} p_i$.

Gilt $p_i = p$ für $i = 1, ..., n$, so liegt die Binomial-Verteilung vor. Für den Erwartungswert der Binomial-Verteilung gilt: $E(X_1 + ... + X_n) = np$.

Auf dasselbe Ergebnis kommt man auch mit Hilfe von Satz 7.1.6:
$P(X = k) = q(k) = \binom{n}{k} p^k (1-p)^{n-k}$. Also gilt:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} k\binom{n}{k} p^k(1-p)^{n-k} \\ &\overbrace{=}^{(*)} \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \\ &= np\underbrace{\sum_{l=0}^{n-1} \binom{n-1}{l} p^{l}(1-p)^{(n-1)-l}}_{=1} \quad \text{(binomische Formel)} \\ &= np. \end{aligned}$$

(*) $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$

7.2.2 Poisson-Verteilung

$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} = q(k)$

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot P(X = k) \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda} \\ &= \lambda \sum_{l=0}^{\infty} q(l) \\ &= \lambda, \end{aligned}$$

da q Wahrscheinlichkeitsfunktion auf $\mathbb{N}_0$ ist.

7.2.3 Geometrische Verteilung

$P(X = k) = p(1-p)^{k-1}$ für $k = 1, 2, \ldots$.
$E(X) = \sum\limits_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum\limits_{k=1}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$.
Dabei haben wir folgende Identität verwendet: $\sum\limits_{k=1}^{\infty} kx^{k-1} = \frac{1}{(x-1)^2}$ für $|x| \leq 1$.

Ein anderer Weg: Ohne Beweis sei folgender Satz angegeben: Ist X eine Zufallsvariable mit Werten in $\mathbb{Z}^+$, so gilt: $E(X) = \sum\limits_{n=1}^{\infty} P(X \geq n)$.
$E(X) = \sum\limits_{n=1}^{\infty} P(X \geq n) = \sum\limits_{n=1}^{\infty} (1-p)^{n-1} = \sum\limits_{m=0}^{\infty} (1-p)^m = \frac{1}{1-(1-p)} = \frac{1}{p}$.

7.2.4 Beispiel für $E(T) = \infty$

Seien X_i für $i = 1, 2, ...$ unabhängige Zufallsvariablen mit $P(X_i = 1) = \frac{1}{2} = P(X_i = -1)$.
Die Gewinnsumme nach n Spielen beträgt $S_n = \sum\limits_{i=1}^{n} X_i$.
Für den Erwartungswert der X_i gilt: $EX_i = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$.
Der Erwartungswert der Gewinnsumme nach n Spielen ist: $ES_n = \sum\limits_{i=1}^{n} EX_i = 0$.

Sei $T = \min\{n \geq 1 | S_n = 1\}$ die Anzahl der Spiele bis, die Gewinnsumme zum ersten Mal den Wert 1 annimmt.

Fragen:
1) Wie groß ist die Wahrscheinlichkeit, daß man irgendwann die Gewinnsumme 1 hat?
2) Wie lang muß man im Mittel werfen, bis die Gewinnsumme 1 erreicht wird?

Antworten:
1) $P(T < \infty) = 1$.
2) $E(T) = \infty$.

Beweis: später !!

7.3 Varianz und Kovarianz

7.3.1 Definition (Varianz, Standardabweichung)

X sei eine Zufallsvariable auf (Ω, P) mit $E(X^2) < \infty$.
Dann heißt $\mathrm{Var}(X) = E(X - E(X))^2$ **Varianz von** X
und $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ **Standardabweichung von** X.

7.3.2 Bemerkungen:

(1) Beide Größen sind Maßzahlen für die Streubreite der Verteilung von X um $E(X)$ herum.

(2) Durch die Forderung $E(X^2) < \infty$ ist die Varianz wohldefiniert, denn es gilt:
$|X| \leq 1 + X^2 \Rightarrow E|X| \leq 1 + E(X^2) < \infty \Rightarrow |E(X)| < \infty$.

(3) $\mathrm{Var}(X) = E(X^2) - (E(X))^2$, denn:
Setze $\mu = E(X)$. Dann ist $\mathrm{Var}(X) = E(X - \mu)^2 = E(X^2 - 2X\mu + \mu^2)$
$= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$.

(4) Es gilt $0 \leq \mathrm{Var}(X) < \infty$, denn wegen (3) ist
$0 \leq \mathrm{Var}(X) = E(X - \mu)^2 = EX^2 - (EX)^2 \leq EX^2 \leq \infty$

(5) $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, denn:

$$\begin{aligned}
\mathrm{Var}(aX + b) &= E(aX + b - E(aX + b))^2 \\
&= E(aX - E(aX))^2 \\
&= E(a(X - E(X)))^2 \\
&= a^2 E(X - E(X))^2 \\
&= a^2 \mathrm{Var}(X).
\end{aligned}$$

7.3.3 Definition (Kovarianz, Korrelationskoeffizient)

X und Y seien Zufallsvariablen mit $\mathrm{Var}(X) < \infty$ und $\mathrm{Var}(Y) < \infty$.
Dann heißt

$$\mathrm{Kov}(X, Y) := E(XY - E(X)E(Y))$$

Kovarianz von X und Y
und

$$\mathrm{Kor}(X, Y) := \frac{\mathrm{Kov}(X, Y)}{\sigma(X)\sigma(Y)}$$

heißt **Korrelationskoeffizient** von X und Y.

7.3.4 Bemerkungen

1) Es gilt $-\infty < \mathrm{Kov}(X,Y) < \infty$
   Für $a,b \in \mathbb{R}$ gilt $2|a \cdot b| \le a^2 + b^2$. Setze $a = |X - EX|$ und $b = |Y - EY|$, so folgt $|(X-EX)(Y-EY)| \le (X-EX)^2 + (Y-EY)^2$ und damit $|\mathrm{Kov}(X,Y)| \le E|(X - EX)(Y-EY)| \le \frac{1}{2}(\mathrm{Var}(X) + \mathrm{Var}(Y)) < \infty$

2) Es gilt $\mathrm{Kov}(X,Y) = E(X \cdot Y) - EX \cdot EY$.

$$\begin{aligned}\mathrm{Kov}(X,Y) &= E\left((X-EX)(Y-EY)\right)\\ &= E\left(X \cdot Y - XEY - YEX + EX \cdot EY\right)\\ &= E(X \cdot Y) - EX \cdot EY.\end{aligned}$$

3) Sind X und Y unabhängig, so gilt $E(XY) = E(X)E(Y)$ und damit $\mathrm{Kov}(X,Y) = 0$ und $\mathrm{Kor}(X,Y) = 0$. Die Umkehrung gilt im Allgemeinen nicht.

Beweis:

$$\begin{aligned}E(XY) &= \sum_{\omega\in\Omega} X(\omega)Y(\omega)p(\omega) \overset{\text{Satz\_7.1.6}}{=} \sum_{i,j} x_i y_j P(\{\omega \mid X(\omega) = x_i,\ Y(\omega) = y_j\})\\ &= \sum_{i,j} x_i y_j P(\{\omega \mid X(\omega) = x_i\})P(\{\omega \mid Y(\omega) = y_j\})\\ &= (\sum_i x_i P(\{\omega \mid X(\omega) = x_i\}))(\sum_j y_j P(\{\omega \mid Y(\omega) = y_j\}))\\ &\overset{\text{Satz\_7.1.6}}{=} E(X)E(Y)\end{aligned}$$

7.3.5 Satz

Seien $X_i,\ i = 1, \dots, n$ Zufallsvariablen für $i = 1, \dots, n$ mit $E(X_i^2) < \infty$ und $S_n = \sum_{i=1}^{n} X_i$.
Dann gilt:

(1) $\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Kov}(X_i, X_j)$.

(2) Falls $X_1, X_2, \dots, X_n$ unabhängig sind, ist
$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$.
Diese Gleichung heißt **Gleichung von Bienaymé**.

Beweis:
Zu (1): Sei $\mu_k = E(X_k)$.

$$\begin{aligned}(S_n - E(S_n))^2 &= (\sum_{i=1}^{n}(X_i - \mu_i))^2 = (\sum_{i=1}^{n}(X_i - \mu_i))(\sum_{j=1}^{n}(X_j - \mu_j))\\ &= \sum_{i=1}^{n}(X_i - \mu_i)^2 + \sum_{i\neq j}(X_i - \mu_i)(X_j - \mu_j)\end{aligned}$$

Bildet man auf beiden Seiten den Erwartungswert, dann folgt (1).

Zu (2): Sind die X_i unabhängig, so gilt $\mathrm{Kov}(X_i, X_j) = 0$ für $i \neq j$. Daraus folgt die Behauptung.

7.4 Varianzen einiger Verteilungen

7.4.1 Gleichverteilung auf einer endlichen Menge

Sei $\Omega = \{1, ..., n\}$ und $p(i) = \frac{1}{n}$ für alle $i \in \Omega$. Weiter sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $X(\Omega) = \{x_1, ..., x_n\}$ ihr Wertebereich. Dann gilt:
$E(X) = \frac{1}{n} \sum\limits_{i=1}^{n} x_i =: \overline{x}_n, \quad \mathrm{Var}(X) = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2.$
Da $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$, gilt: $\frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2 = \frac{1}{n} \sum\limits_{i=1}^{n} x_i^2 - \overline{x}_n^2$ und damit
$\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2 + n\overline{x}_n^2.$

7.4.2 Bernoulli-Variablen und ihre Summen

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit $P(X_i = 1) = p_i$ und $P(X_i = 0) = 1 - p_i$. Dann gilt: $E(X_i) = p_i$ und $\mathrm{Var}(x_i) = p_i - p_i^2 = p_i(1 - p_i)$.
Sei $S_n = \sum\limits_{i=1}^{n} X_i$. Dann ist wegen $E(S_n) = \sum\limits_{i=1}^{n} p_i$ und $\mathrm{Var}(S_n) = \sum\limits_{i=1}^{n} p_i - \sum\limits_{i=1}^{n} p_i^2$.
Sei $\overline{p} := \frac{1}{n} \sum\limits_{i=1}^{n} p_i$. Dann gilt
$\mathrm{Var}(S_n)$ wird maximal, wenn $p_i = \overline{p}$ für alle i. In diesem Fall ist dann $\mathrm{Var}(S_n) = n\overline{p}(1-\overline{p})$.

Beweis: $\mathrm{Var}(S_n) = n\overline{p} - \sum\limits_{i=1}^{n} p_i^2 \overbrace{=}^{7.4.1} n\overline{p} - (n\overline{p}^2 + \sum\limits_{i=1}^{n} (p_i - \overline{p})^2)$
Die rechte Seite wird minimal, falls $\sum\limits_{i=1}^{n} (p_i - \overline{p})^2 = 0$. Dies gilt genau dann, wenn $p_i = \overline{p}$ für alle i ist.

7.4.3 Poisson-Verteilung

Sei X Poisson-verteilt mit Parameter λ. Dann gilt:
$E(X) = \lambda$ und $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \lambda$, denn:

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\
&= \lambda^2 \underbrace{\sum_{l=0}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda}}_{=1} + \lambda \underbrace{\sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda}}_{=1} \\
&= \lambda^2 + \lambda.
\end{aligned}$$

Und damit ist $\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

7.4.4 Hypergeometrische Verteilung

In einer Urne seien r rote und s schwarze Kugeln. Davon werden n Kugeln ohne Zurücklegen gezogen ($n \leq r+s$). Um die Varianz der Anzahl der schwarzen Kugeln in der Ziehung zu bestimmen definieren wir für $i = 1, \ldots, n$ die Zufallsvariablen

$X_i = \begin{cases} 1 & \text{falls } i\text{-te Kugel Schwarz ist} \\ 0 & \text{sonst} \end{cases}$.

$S_n = \sum\limits_{i=1}^{n} X_i$ sei die Anzahl der schwarzen Kugeln in der Ziehung.

Mit $p := P(X_1 = 1) = \frac{s}{r+s}$ gilt $E(X_1) = E(X_1^2) = p$ und damit $\mathrm{Var}(X_1) = E(X_1^2) - (E(X_1))^2 = p - p^2 = p(1-p)$.

Es läßt sich zeigen, daß $P(X_i = 1) = P(X_1 = 1) = p$ für alle i und $P(X_j = 1, X_k = 1) = \frac{s}{r+s}$ für alle $j \neq k$ ist. (Siehe Krengel, Kap 2.7: Austauschbare Verteilungen.)

Damit ist: $\mathrm{Var}(X_i) = E(X_i^2) - (E(X_i))^2 = p - p^2 = p(1-p) = \mathrm{Var}(X_1)$ und

$$\begin{aligned}
\mathrm{Kov}(X_j, X_k) &= E(X_j X_k) - E(X_j)E(X_k) = P(X_j = 1, X_k = 1) - p^2 \\
&= p\frac{s-1}{r+s-1} - p^2 \\
&= -p(1-p)\frac{1}{r+s-1} \qquad \text{für } j \neq k.
\end{aligned}$$

Unter Verwendung von Satz 7.3.5 folgt daraus:

$$\begin{aligned}
\mathrm{Var}(S_n) &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{j \neq k} \mathrm{Kov}(X_j, X_k) \\
&= n\mathrm{Var}(X_1) - n(n-1)p(1-p)\frac{1}{r+s-1} \\
&= np(1-p) - np(1-p)(n-1)\frac{1}{r+s-1} \\
&= np(1-p)\left[1 - \frac{n-1}{r+s-1}\right].
\end{aligned}$$

Wenn wir alle $r+s$ Kugeln aus der Urne ziehen, erwarten wir natürlich, daß wir alle schwarzen Kugeln ziehen und somit S_n auf jeden Fall den Wert s annimmt. Für $n = r+s$ erwarten wir also $\mathrm{Var}(S_n) = 0$. Ein Vergleich mit der Formel bestätigt diese.

7.5 Die Tschebychew-Ungleichung und das Gesetz der Großen Zahlen

Seien $X_1, X_2, ..., X_n$ unabhängig und $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Dann gilt:

$$\mathrm{Var}\left(\overline{X}_n\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right) \overbrace{=}^{\text{Satz 7.3.5}} \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Haben die $X_1, ..., X_n$ alle dieselbe Verteilung, so gilt:

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n}\mathrm{Var}(X_1) \text{ und } E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X_1).$$

Da die Varianz von $\overline{X}_n$ gegen 0 geht für $n \to \infty$, so vermutet man: $\overline{X}_n \to E(X_1)$. Es ist aber nicht ganz klar, was Konvergenz hier bedeutet.

7.5.1 Satz (Tschebyschew-Ungleichung)

Sei X eine Zufallsvariable mit $E(X^2) < \infty$. Dann gilt für jedes $\varepsilon > 0$:

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

Beweis:
Sei $X'(\omega) := X(\omega) - E(X)$.
Dann gilt: $E(X') = 0, \quad \mathrm{Var}(X) = \mathrm{Var}(X') = E(X'^2)$ und damit

$$\begin{aligned}
P(|X - E(X)| > \varepsilon) &= P(|X'| > \varepsilon) \\
&= \sum_{\{x \in X'(\Omega):\ |x| > \varepsilon\}} P(X' = x) \\
&\leq \sum_{\{x \in X'(\Omega):\ |x| > \varepsilon\}} \frac{x^2}{\varepsilon^2} P(X' = x) \\
&\leq \frac{1}{\varepsilon^2} \sum_{x \in X'(\Omega)} x^2 P(X' = x) \\
&= \frac{1}{\varepsilon^2} E(X'^2) \\
&= \frac{1}{\varepsilon^2} \mathrm{Var}(X') \\
&= \frac{1}{\varepsilon^2} \mathrm{Var}(X).
\end{aligned}$$

7.5.2 Satz (Gesetz der Großen Zahlen)

Für jedes n seien X_i, $i = 1, \dots, n$ unabhängige und identisch verteilte Zufallsvariablen mit $E(X_1^2) < \infty$. Sei $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} P(\{\omega :\ |\overline{X}_n(\omega) - E(X_1)| > \varepsilon\}) = 0.$$

Schreibweise: $\overline{X}_n \xrightarrow{P} E(X_1)$ (**stochastische Konvergenz**).

Beweis:
Setze in die Tschebychew-Ungleichung (Satz 7.5.1) $X := \overline{X}_n$ ein. Dann gilt:

$$P\left(|\overline{X}_n - E(\overline{X}_n)| > \varepsilon\right) \leq \frac{\mathrm{Var}\left(\overline{X}_n\right)}{\varepsilon^2} = \frac{\mathrm{Var}\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\sum\limits_{i=1}^{n} \mathrm{Var}(X_1)}{\varepsilon^2} = \frac{1}{n}\frac{\mathrm{Var}(X_1)}{\varepsilon^2}.$$

Damit gilt:

$$\lim_{n\to\infty} P\left(|\overline{X}_n - E(\overline{X}_n)| > \varepsilon\right) = \lim_{n\to\infty} P\left(|\overline{X}_n - E(X_1)| > \varepsilon\right) \leq \lim_{n\to\infty} \frac{1}{n}\frac{\mathrm{Var}(X_1)}{\varepsilon^2} = 0.$$

7.5.3 Beispiel: Die p-Münze

Wir betrachten eine p-Münze, d. h. $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$. Damit ist $E(X_i) = p$.
Die X_i seien unabhängig und $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.
Die relative Häufigkeit der „1“ bei n Würfen konvergiert gegen p, $\overline{X}_n \xrightarrow{P} E(X_1) = p$. Das ist aber genau die Aussage des Gesetzes der großen Zahlen.

7.5.4 Anwendung (Wahlumfrage)

Vor einer Wahl werden n Personen befragt, ob sie die Partei A wählen werden oder nicht (Ja-Nein-Antworten). Dabei gelte $X_i = \begin{cases} 1 & \text{falls Person i A wählt} \\ 0 & \text{sonst} \end{cases}$.
und $P(X_i = 1) = p = 1 - P(X_i = 0)$.
Schätze p mit Hilfe von $\overline{X}_n$.
Es gilt: $E(\overline{X}_n) = p, \mathrm{Var}(\overline{X}_n) = \frac{p(1-p)}{n}$ und $\overline{X}_n \to p$.

7.6 Approximation stetiger Funktionen auf dem Intervall $[0,1]$ durch Polynome

Auch hierfür kann man das Gesetz der großen Zahlen anwenden.

7.6.1 Satz

Sei $f : [0,1] \to \mathbb{R}$, f stetig.
Sei $B_n(p) := \sum\limits_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k}$. Dann gilt:
Die Folge von Funktionen $(B_n)_{n\in\mathbb{N}}$ konvergiert gleichmäßig gegen f für $n \to \infty$.
In Formeln: $\lim\limits_{n\to\infty} \max\limits_{0\leq p\leq 1} |B_n(p) - f(p)| = 0$.

Beweis:
Seien $X_1, ..., X_n$ unabhängige, identisch verteilte Bernoulli-Variablen mit
$P(X_i = 1) = p = 1 - P(X_i = 0)$ und $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Dann ist $Ef(S_n) = \sum_{k=0}^{n} f(\frac{k}{n})b(n,p;k)$, wobei $b(n,p;k) = \binom{n}{k}p^k(1-p)^{n-k}$.
Da f auf dem kompakten Intervall $[0,1]$ stetig ist, ist f gleichmäßig stetig auf $[0,1]$. Damit gilt: $\forall\varepsilon > 0\ \exists\delta > 0$ mit $|f(x) - f(y)| \leq \varepsilon$ falls, $|x-y| \leq \delta$.
Außerdem ist f beschränkt, also $\max\limits_{0\leq x\leq 1} |f(x)| \leq M$ für ein geeignetes $M \in \mathbb{R}$.

Sei also $\varepsilon \geq 0$. Dann gibt es ein $\delta \geq 0$ mit $|f(x) - f(y)| \leq \varepsilon$ falls $|x-y| \leq \delta$ und es gilt:

$$
\begin{aligned}
|f(p) - B_n(p)| &= |\sum_{k=0}^{n} f(p)b(n,p;k) - \sum_{k=0}^{n} f(\frac{k}{n})b(n,p;k)| \\
&= |\sum_{k=0}^{n}(f(p) - f(\frac{k}{n}))b(n,p;k)| \\
&\leq \sum_{\{k:|\frac{k}{n}-p|\leq\delta\}} |f(p) - f(\frac{k}{n})|b(n,p;k) + \sum_{\{k:|\frac{k}{n}-p|>\delta\}} |f(p) - f(\frac{k}{n})|b(n,p;k) \\
&\leq \varepsilon \sum_{\{k:|\frac{k}{n}-p|\leq\delta\}} b(n,p;k) + 2M \sum_{\{k:|\frac{k}{n}-p|>\delta\}} b(n,p;k) \\
&\leq \varepsilon + 2M \cdot P(\{|S_n - p| > \delta\}) \\
&\overbrace{\leq}^{Satz 7.5.1} \varepsilon + 2M\frac{p(1-p)}{n\delta^2} \\
&\leq \varepsilon + 2M\frac{1}{4n\delta^2} \qquad \text{da } \max_{0\leq p\leq 1} p(1-p) = 1/4 \text{ ist,} \\
&= \varepsilon + \frac{M}{2n\delta^2} \\
&\leq 2\varepsilon, \qquad \text{falls } n \text{ hinreichend groß ist.}
\end{aligned}
$$

Wir haben also ein n gefunden, das nur von ε und nicht von $p \in [0,1]$ abhängt. Damit ist die Konvergenz gleichmäßig.

8 Satz von de Moivre-Laplace und seine Verallgemeinerung

8.1 Der Satz von de Moivre-Laplace

Seien $X_i, \quad i = 1, \ldots, n$ unabhängige Zufallsvariablen mit $P(X_i = 1) = p = 1 - P(X_i = 0)$ und $S_n = \sum\limits_{i=0}^{n} X_i$. Dann ist S_n binomialverteilt nach $b(n, p)$.

8.1.1 Satz (de Moivre-Laplace)

Für alle $-\infty \leq a \leq b \leq \infty$ gilt:

$$\lim_{n\to\infty} P(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) = \Phi(b) - \Phi(a)$$

Dabei ist $\Phi(y) := \int\limits_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

Bemerkungen:

1) $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$ ist eine standardisierte Zufallsvariable, d.h. $E(S_n^*) = 0$ und $\mathrm{Var}(S_n^*) = 1$. Der Satz besagt demnach, daß die Verteilungen der standardisierten Variablen konvergieren.

2) Der Satz beschreibt die Konvergenz von Flächen: $\lim\limits_{n\to\infty} \sum\limits_{k \leq \alpha_n} b(n, p; k) \to \Phi(\alpha)$, wobei $\alpha_n = np + \alpha\sqrt{np(1-p)}$ ist.

3) Eigenschaften von Φ: $\Phi(-\infty) = 0, \Phi(+\infty) = 1$, Φ ist monoton wachsend und symmetrisch um "0".

4) Der Satz wird oft folgendermaßen verwendet:
$P(a \leq S_n \leq b) \approx \Phi(\frac{b-np}{\sqrt{np(1-p)}}) - \Phi(\frac{a-np}{\sqrt{np(1-p)}})$.

5) Sei $\hat{p}_n := \frac{S_n}{n} = \overline{X}_n$. Dann besagt Satz 8.1.1: $\lim\limits_{n\to\infty} P(a \leq \frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}} \leq b) = \Phi(b) - \Phi(a)$.
Denn $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} = \frac{\frac{S_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \sqrt{n}\frac{\frac{S_n}{n} - p}{\sqrt{p(1-p)}} = \frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}}$.
Diese Formulierung drückt die statistische Bedeutung von Satz 8.1.1 aus:
Nach dem Gesetz der großen Zahlen gilt: $\hat{p}_n \xrightarrow{P} p$ für $n \to \infty$, das heißt:
Für alle $\varepsilon > 0$ gilt $\lim\limits_{n\to\infty} P_p(|\hat{p}_n - p| > \varepsilon) = 0$.
Der Satz von de Moivre-Laplace gibt die Schwankung von $\hat{p}_n$ um p an.

6) Satz 8.1.1 bleibt richtig, wenn wir im Argument von P die $\leq$ durch $<$ ersetzen. Also $\lim\limits_{n\to\infty} P(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b) = \Phi(b) - \Phi(a)$.

8.1.2 Korollar

Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit $P(X_i = 1) = p = 1 - P(X_i = 0)$ für $i = 1, \ldots, n$. Dann gilt für alle $\varepsilon > 0$: $\lim\limits_{n\to\infty} P(|\frac{S_n}{n} - p| > \varepsilon) = 0$.

8.1.3 Beispiel

Wie groß ist die Wahrscheinlichkeit, dass in 6000 Würfen mit einem Würfel die "6".

a) mehr als 1100 auftritt?

b) höchstens 950 mal oder mindestens 1050 mal auftritt?

Lösung:

a) $P\left(1100 < S_{6000} < \infty\right) = P\left(100 < S_{6000} - 6000 \cdot \frac{1}{6} < \infty\right)$
$= P\left(\frac{100}{\sqrt{6000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} < S^*_{6000} < \infty\right) = P\left(\sqrt{12} < S^*_{6000} < \infty\right)$
$\overbrace{\cong}^{\text{Satz 8.1.1}} \quad \Phi\left(\infty\right) - \Phi\left(\sqrt{12}\right) = 1 - \Phi\left(\sqrt{12}\right) = 0{,}00028.$

b) $P\left(S_{6000} \leq 950 \text{ oder } S_{6000} \geq 1050\right) = 1 - P\left(950 < S_{6000} < 1050\right)$
$= 1 - P\left(\frac{-50}{\sqrt{6000 \frac{1}{6} \frac{5}{6}}} < S^*_{6000} < \frac{50}{\sqrt{6000 \frac{1}{6} \frac{5}{6}}}\right) = 1 - P\left(-\sqrt{3} < S^*_{6000} < \sqrt{3}\right)$
$\overbrace{\cong}^{\text{Satz 8.1.1}} \quad 1 - \left[\Phi\left(\sqrt{3}\right) - \Phi\left(-\sqrt{3}\right)\right] = 0{,}0832$

8.1.4 Vergleich der normierten Binomialverteilung mit Φ

Näherungsweise gilt:

$$P(S_n \leq b') = P(S^*_n \leq \frac{b' - np}{\sqrt{np(1-p)}}) \approx \Phi(\frac{b' - np}{\sqrt{np(1-p)}})$$

wobei $S^*_n = \frac{S_n - np}{\sqrt{np(1-p)}}$.

Wie gut ist diese Näherung für verschiedene n?
Für große n ist die Näherung gut, vor allem dann, wenn b' in der Nähe von np liegt. Man erhält aber noch bessere Werte, wenn man eine so genannte Stetigkeitskorrektur vornimmt, indem man b' durch $b' + \frac{1}{2}$ ersetzt. Das macht sich besonders für kleine n deutlich bemerkbar.

Numerisch:
$p = 0{,}4 \quad b' = np \cdot 0{,}8$

n	b'	$\sum_{k=0}^{b} b(n,p;k)$	$\Phi(\frac{b'-np}{\sqrt{np(1-p)}})$	$\Phi(\frac{b'+\frac{1}{2}-np}{\sqrt{np(1-p)}})$
12	4	.4382	.3187	.4298
25	8	.2735	.2071	.2701
50	16	.1561	.1241	.1562
100	32	.0615	.0512	.0629
200	64	.0119	.0105	.0126

Man beachte, daß gilt: $P(S_n < b') \to 0$ für $n \to \infty$

8.1.5 Beispiel: Unsicherheit bei Wahlumfragen

Es werden 2000 Wähler befragt, ob sie die Partei A wählen werden. Die Wahrscheinlichkeit, daß eine befragte Person „Ja“ antwortet, sei $p_0 = 0,55$.
Frage: Was ist die Wahrscheinlichkeit, daß $55 \pm 2,5\%$ der Befragten mit „Ja“ antworten?

Behauptung: Sei S_{2000} die Anzahl der Personen, die „Ja“ antworten. Dann gilt:

$$P_{p_0}\left\{\frac{S_{2000}}{2000} \in [p_0 \pm 0,025]\right\} \cong 0,9754.$$

Beweis:
$55 \pm 2,5\%$ der befragten Personen sind 1100 ± 50 Personen. Mit Satz 8.1.1 folgt:

$$\begin{aligned}
&P_{p_0}(1050 \leq S_{2000} \leq 1150)\\
&= P_{p_0}(-50 \leq S_{2000} - 1100 \leq 50)\\
&= P_{p_0}\left(\frac{-50}{\sqrt{2000 \cdot 0,55(1-0,55)}} \leq \frac{S_{2000} - 1100}{\sqrt{2000 \cdot 0,55(1-0,55)}} \leq \frac{50}{\sqrt{2000 \cdot 0,55(1-0,55)}}\right)\\
&= P_{p_0}(-2,2473 \leq S^*_{2000} \leq 2,2473)\\
&\approx \Phi(2,2473) - \Phi(-2,2473) = 0,9754.
\end{aligned}$$

8.2 Die Landauschen Symbole und die Stirling Formel

8.2.1 Definition (Landausche Symbole)

Seien $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ Folgen mit Werten in $\mathbb{R}$. Sei $b_n \neq 0$ für alle n.

1. Man schreibt $a_n = o(b_n)$ für $n \to \infty$, wenn $\frac{a_n}{b_n} \to 0$ für $n \to \infty$.
2. Seien $a_n > 0$ und $b_n > 0$ für alle n. Man schreibt $a_n = O(b_n)$ für $n \to \infty$, falls eine Konstante $K > 0$ und ein $n_0 \in \mathbb{N}$ existieren, so daß $a_n \leq K \cdot b_n$ für $n \geq n_0$ gibt.

3. Sind a_n, $b_n \neq 0$ für alle n, so schreibt man:

$$a_n \sim b_n \text{ für } n \to \infty \text{ falls } \frac{a_n}{b_n} \to 1 \text{ für } n \to \infty.$$

   Mann sagt a_n is asymptotisch äquivalent zu b_n.

8.2.2 Beispiele

1. Die folgenden zwei Aussagen sind äquivalent:
   $a_n = o(1)$ und $a_n \to 0$ für $n \to \infty$.

2. $n = o(n^2)$ für $n \to \infty$.

3. $\ln n = o(n)$ für $n \to \infty$. Denn setze $y = \ln n$, dann ist:
   $\frac{\ln n}{n} = \frac{y}{e^y} = \frac{y}{1+y+\frac{y^2}{2!}+\frac{y^3}{3!}+\ldots} \leq \frac{1}{1+\frac{y}{2!}+\frac{y^2}{3!}+\ldots} \to 0$ für $y \to \infty$
   und $y \to \infty$.

4. Es gilt: $a_n \sim b_n$ für $n \to \infty \quad \Leftrightarrow \quad \frac{a_n}{b_n} = 1 + o(1)$ für $n \to \infty$.
   Dabei bedeutet $a_n = b_n + o(c_n)$ für $n \to \infty$: Es existiert eine Folge $(d_n)_{n\in\mathbb{N}}$ mit $a_n = b_n + d_n$ und $d_n = o(c_n)$ für $n \to \infty$.

8.2.3 Stirling Formel: $n! \sim \sqrt{2\pi n}n^n e^{-n}$.

Sei $a_n = \sqrt{2\pi n}n^n e^{-n}$. Dann wird $n!$ für große n gut durch a_n approximiert. Es gibt jedoch zwei Folgen a_n' und a_n'', die $n!$ noch etwas besser approximieren und besonders für kleinere n sehr nützlich sind: $a_n' = a_n \cdot e^{\frac{1}{12n}}$ und $a_n'' = a_n \cdot e^{\frac{1}{12n+1}}$.

Hier ist ein Vergleich der Formeln:

n	$n!$	a_n	a_n'	a_n''
1	1	0,922	1,002	0,996
2	2	1,919	2,006	1,997
3	6	5,836	6,003	5,995
4	24	23,506	24,001	23,990
5	120	118,019	120,002	119,969

n	1	2	3	4
$a_n/n!$	0,922	0,9595	0,9771	0,9794
$a_n'/n!$	1,002	1,0003	1,00005	1,00004
$a_n''/n!$	0,996	0,9985	0,99917	0,99958

Man sieht, daß insbesondere für kleine n die Formeln a_n' und a_n'' besser sind. Da aber für alle drei Formeln asymptotische Äquivalenz zu $n!$ gilt, ist es für mathematische Beweise nicht wesentlich, welche man nimmt. Deshalb werden wir stets die Formel für a_n verwenden.

8.3 Approximation der Binomialverteilung

Die Wahrscheinlichkeit bei n Würfen mit einer p-Münze k_n Einsen zu werfen beträgt $P(X = k_n) = \binom{n}{k_n} p^{k_n}(1-p)^{n-k_n}$. Für $n \to \infty$ und $k_n \to \infty$ so, daß auch $n - k_n \to \infty$, gilt näherungsweise unter Verwendung der Stirling-Formel (Dabei schreiben wir k für k_n):

$$\begin{aligned}
\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} &\sim \frac{\sqrt{2\pi n} n^n e^{-n} p^k (1-p)^{n-k}}{\sqrt{2\pi k} k^k e^{-k} \sqrt{2\pi(n-k)}(n-k)^{n-k} e^{-(n-k)}} \\
&= \frac{1}{\sqrt{2\pi \frac{k}{n}(1-\frac{k}{n})n}} (\frac{p}{k/n})^k (\frac{1-p}{1-k/n})^{n-k} \\
&= \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} (\frac{p}{p_n})^k (\frac{1-p}{1-p_n})^{n-k} \quad \text{mit } p_n = \frac{k}{n} \\
&= \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} \left[(\frac{p}{p_n})^{p_n} (\frac{1-p}{1-p_n})^{1-p_n} \right]^n \\
&= \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} e^{-nI(p_n,p)}
\end{aligned}$$

mit $I(q,p) = q\ln(\frac{q}{p}) + (1-q)\ln(\frac{1-q}{1-p})$, wobei $0 < p < 1$ und $0 < q < 1$. $I(q,p)$ heißt relative Entropie von q bezüglich p. Zusammenfassend haben wir folgendes bewiesen.

8.3.1 Satz

Sei $0 < p < 1$ und sei $\delta < \min(p, 1-p)$. Dann gilt

$$P_p(S_n = np_n) = \frac{1}{\sqrt{2\pi n p_n(1-p_n)}} e^{-nI(p_n,p)}(1 + o(1))$$

gleichmäßig für alle Folgen $(p_n; n \geq 1)$ mit $np_n \in \{0, 1, \ldots, n\}$ und $\min(p_n, 1-p_n) > \delta/2$ für $n \to \infty$.

Bemerkung: Gleichmäßig bedeutet hier, daß der $o(1)$-Term lediglich von δ und n abhängt.

Nun wollen wir den Exponenten $nI(p_n, p)$ entwickeln und einen lokalen Grenzwertsatz herleiten. Dazu brauchen wir einige Eigenschaften der relativen Entropie.

8.3.2 Eigenschaften von $I(q,p)$

Es gilt:

1) $I(p,p) = 0$,

2) $I(q,p) > 0$ für $q \neq p$,

3) $I(q,p)$ ist strikt konvex und zweimal stetig differenzierbar in beiden Argumenten.

Folglich ist eine quadratische Approximation nahe p möglich.

8.3.3 Lemma

Sei $0 < p < 1$. Dann gilt $I(p_n, p) = \frac{1}{2}\frac{(p_n-p)^2}{p_n(1-p_n)} + o((p_n-p)^2)$ für $p_n \to p$.
Beweis:

$$\begin{aligned}
I(p_n,p) &= -p_n \ln(\frac{p}{p_n}) - (1-p_n)\ln(\frac{1-p}{1-p_n}) \\
&= -p_n \ln(1-\frac{p_n-p}{p_n}) - (1-p_n)\ln(1-\frac{p-p_n}{1-p_n}) \\
&\overbrace{=}^{Taylor} -p_n[-(\frac{p_n-p}{p_n}) - \frac{1}{2}(\frac{p_n-p}{p_n})^2 + o((\frac{p_n-p}{p_n})^2)] \\
&\quad -(1-p_n)[-(\frac{p-p_n}{1-p_n}) - \frac{1}{2}(\frac{p-p_n}{1-p_n})^2 + o((\frac{p-p_n}{1-p_n})^2)] \\
&\overbrace{=}^{8.2.5,2} [(p_n-p) + \frac{1}{2}\frac{(p_n-p)^2}{p_n} + o(\frac{(p_n-p)^2}{p_n})] \\
&\quad +[(p-p_n) + \frac{1}{2}\frac{(p_n-p)^2}{1-p_n} + o(\frac{(p_n-p)^2}{1-p_n})] \\
&\overbrace{=}^{8.2.5,3} \frac{1}{2}\frac{(p_n-p)^2}{p_n(1-p_n)} + o((p_n-p)^2) + o((p_n-p)^2), \\
&\overbrace{=}^{8.2.5,4} \frac{1}{2}\frac{(p_n-p)^2}{p_n(1-p_n)} + o((p_n-p)^2) \qquad \text{für } p_n \to p,
\end{aligned}$$

da $\ln(1-x) = -x - \frac{1}{2}x^2 + o(x^2)$ für $x \to 1$.

Damit ergibt sich folgender Satz:

8.3.4 Satz (Lokaler zentraler Grenzwertsatz)

Sei $0 < p < 1$. Dann gilt für jedes $K > 0$ und alle Folgen $(p_n)_{n\in\mathbb{N}}$ mit $np_n \in \{0,1,2,...,n\}$ und $|p_n - p| \le \frac{K}{\sqrt{n}}$ für $n \to \infty$:

(1) $P_p(S_n = np_n) = \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} e^{-\frac{n(p_n-p)^2}{2p_n(1-p_n)}}(1+o(1))$,

(2) $P_p(S_n = np_n) = \frac{1}{\sqrt{2\pi p(1-p)n}} e^{-\frac{n(p_n-p)^2}{2p(1-p)}}(1+o(1))$.

Diese Konvergenz ist gleichmäßig für alle Folgen $(p_n)_{n\in\mathbb{N}}$ mit den oben genannten Bedingungen.
Beweis:
Zu (1): Sei $|p_n - p| \le \frac{K}{\sqrt{n}}$. Dann existiert eine Folge $(a_n)_{n\in\mathbb{N}}$ mit $p_n = p + \frac{a_n}{\sqrt{n}}$ und $|a_n| \le K$ für alle n. Nach Lemma 8.3.3 ist $nI(p_n,p) = n[\frac{(p_n-p)^2}{2p_n(1-p_n)} + o((p_n-p)^2)]$. Da $n(p_n-p)^2 = \frac{n\cdot a_n^2}{n} \le K^2$ ist, folgt sofort $nI(p_n,p) = \frac{n(p_n-p)^2}{2p_n(1-p_n)} + o(K^2)$ und weil $o(K^2) = o(1)$ ist, ergibt sich $nI(p_n,p) = \frac{n(p_n-p)^2}{2p_n(1-p_n)} + o(1)$. Damit ist

$$e^{-nI(p_n,p)} = e^{-\frac{n(p_n-p)^2}{2p_n(1-p_n)}+o(1)} = e^{-\frac{n(p_n-p)^2}{2p_n(1-p_n)}} e^{o(1)} = e^{-\frac{n(p_n-p)^2}{2p_n(1-p_n)}}(1+o(1)).$$

Also gilt mit 8.3.1

$$P(S_n = np_n) = \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} e^{-nI(p_n,p)}(1+o(1)) = \frac{1}{\sqrt{2\pi p_n(1-p_n)n}} e^{-\frac{n(p_n-p)^2}{2p_n(1-p_n)}}(1+o(1))^2,$$

woraus Behauptung (1) folgt.

Zu (2): Führe diesen Fall auf (1) zurück: Zeige

$$\frac{n(p_n-p)^2}{2}[\frac{1}{p(1-p)} - \frac{1}{p_n(1-p_n)}] = o(1) \quad \text{gleichmäßig für } |a_n| \leq K.$$

Es gilt: $\frac{n(p_n-p)^2}{2}[\frac{1}{p(1-p)} - \frac{1}{p_n(1-p_n)}] = \frac{n(p_n-p)^2}{2}[\frac{1}{p} - \frac{1}{p_n} + \frac{1}{1-p} - \frac{1}{1-p_n}]$.
Weiter gilt: $\frac{1}{p} - \frac{1}{p_n} = \frac{a_n/\sqrt{n}}{p(p+a_n/\sqrt{n})} = \frac{a_n/\sqrt{n}}{p^2(1+\frac{a_n}{p\sqrt{n}})} \sim \frac{a_n/\sqrt{n}}{p^2} \leq \frac{|K|}{\sqrt{n}p^2} = o(1)$.
Entsprechend zeigt man: $\frac{1}{1-p} - \frac{1}{1-p_n} = o(1)$.
Zusammen mit $n(p_n-p)^2 \leq K^2$ folgt daraus: $\frac{n(p_n-p)^2}{2}[\frac{1}{p(1-p)} - \frac{1}{p_n(1-p_n)}] = o(1)$.
Außerdem ist $\frac{1}{\sqrt{p_n(1-p_n)}} = \frac{1}{\sqrt{p(1-p)}}(1+o(1))$ gleichmäßig für $|p_n - p| \leq \frac{K}{\sqrt{n}}$ für $n \to \infty$.
Damit ist der Satz bewiesen.

Jetzt können wir den Satz von de Moivre-Laplace beweisen:

Beweis von Satz 8.1.1

1. Fall: Sei $-\infty < a < b < \infty$. Setze nun $a_n = np + a\sqrt{p(1-p)n}$ und $b_n = np + b\sqrt{p(1-p)n}$. Dann gilt aufgrund des lokalen Grenzwertsatzes

$$\begin{aligned}
P(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b) &= P(a_n \leq S_n \leq b_n) \\
&= \sum_{a_n \leq k \leq b_n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{a_n \leq k \leq b_n} \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\frac{(k-np)^2}{np(1-p)}}(1+o(1)) \\
\text{Mit } l = k - np \qquad &= \sum_{a\sqrt{np(1-p)} \leq l \leq b\sqrt{np(1-p)}} \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\frac{l^2}{np(1-p)}}(1+o(1)) \\
\text{Setze nun } h = \frac{1}{\sqrt{n}} \qquad &= \sum_{a\sqrt{p(1-p)} \leq l\cdot h \leq b\sqrt{p(1-p)}} \frac{h}{\sqrt{2\pi p(1-p)}} e^{-\frac{1}{2}\frac{(lh)^2}{p(1-p)}}(1+o(1))
\end{aligned}$$

Dies sind aber Riemann-Summen einer stetigen Funktion auf einem kompakten Intervall. Folglich hat mal Konvergenz für $h \to 0$

$$\longrightarrow \int_{a\sqrt{p(1-p)}}^{b\sqrt{p(1-p)}} \frac{1}{\sqrt{2\pi p(1-p)}} e^{-\frac{z^2}{2p(1-p)}} dz.$$

Setze $y = \frac{z}{\sqrt{p(1-p)}}$. Dann folgt:

$$\int\limits_{a\sqrt{p(1-p)}}^{b\sqrt{p(1-p)}} \frac{1}{\sqrt{2\pi p(1-p)}} e^{-\frac{z^2}{2p(1-p)}} dz = \int\limits_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

und damit die Behauptung.

2. Fall: Sei $a = -\infty, b \in \mathbb{R}$. Sei $\varepsilon > 0$ vorgegeben. Wegen $\int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1$ existiert ein $a_\varepsilon > 0$ mit $\Phi(-a_\varepsilon) + 1 - \Phi(a_\varepsilon) = \frac{\varepsilon}{2}$ und $-a_\varepsilon < b < a_\varepsilon$. Nach Teil 1 gilt $\lim\limits_{n\to\infty} P(-a_\varepsilon \leq S_n^* \leq a_\varepsilon) = \Phi(a_\varepsilon) - \Phi(-a_\varepsilon)$ und damit

$$P(S_n^* < -a_\varepsilon) \leq 1 - P(-a_\varepsilon \leq S_n^* \leq a_\varepsilon) \to 1 - \Phi(a_\varepsilon) + \Phi(-a_\varepsilon) + \frac{\varepsilon}{2} \leq \varepsilon$$

für n hinreichend groß.
Nun gilt für $-a_\varepsilon < b < a_\varepsilon$

$$\begin{aligned}
&\left| \int\limits_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dx - P(S_n^* \leq b) \right| \\
\leq &\left| \int\limits_{-\infty}^{-a_\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy - P(S_n^* \leq -a_\varepsilon) + \int\limits_{-a_\varepsilon}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy - P(-a_\varepsilon \leq S_n^* \leq b) \right| \\
\leq &\left| \int\limits_{-\infty}^{-a_\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right| + |P(S_n^* \leq -a_\varepsilon)| + \left| \int\limits_{-a_\varepsilon}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy - P(-a_\varepsilon \leq S_n^* \leq b) \right| \\
\leq &\frac{\varepsilon}{2} + \varepsilon + \frac{\varepsilon}{2} = 2\varepsilon \qquad \text{für } n \text{ hinreichend groß!}
\end{aligned}$$

8.4 Der Zentrale Grenzwertsatz

Eine wesentliche Verallgemeinerung des Satzes von de Moivre-Laplace ist möglich.

8.4.1 Satz

Seien $X_1, ..., X_n, ...$ unabhängige und identisch verteilte Zufallsvariablen
mit $EX_1^2 < \infty$.
Dann gilt für $-\infty \leq a < b \leq \infty$

$$\lim_{n\to\infty} P(a \leq \frac{S_n - nE(X_1)}{\sqrt{n\mathrm{Var}(X_1)}} \leq b) = \Phi(b) - \Phi(a).$$

Der Beweis wird hier nicht gegeben. Es soll aber daraufhingewiesen werden, daß z.B. für unabhängige Poission-Variable $X_1, \ldots, X_n$ mit $E(X_1) = \lambda$ gilt

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq b\right) = \Phi(b) - \Phi(a).$$

9 Erzeugende Funktionen

9.1 Definition und Eigenschaften erzeugender Funktionen

9.1.1 Definition (erzeugende Funktion)

Sei X Zufallsgröße mit Werten in $\mathbb{N} \cup \{0\}$. Dann heißt
$f_X(z) = \sum\limits_{n=0}^{\infty} P(X = n) z^n, \quad -1 \leq z \leq 1,$
die erzeugende Funktion der Verteilung von X.

Bemerkung:

1) $f_X(z)$ ist für $|z| \leq 1$ wohldefiniert, da
$P(X = n)|z^n| \leq P(X = n)$ für $|z| \leq 1$ und $\sum\limits_{n=0}^{\infty} P(X = n) = 1$ gilt.

2) Es gilt: $Ez^X = \sum\limits_{\omega} p(\omega) z^{X(\omega)} \overbrace{=}^{\text{Trafosatz}} \sum\limits_{n=0}^{\infty} P(X = n) z^n = f_X(z)$.

3) f_X legt die Verteilung von X eindeutig fest (siehe Satz 9.1.2),
die Verteilung von X legt f_X eindeutig fest.

9.1.2 Satz

Es gilt:

1) a) $f_X(0) = P(X = 0)$,
b) $f_X(1) = 1$,
c) $P(X = n) = \frac{f_X^{(n)}(0)}{n!}$.

2) $f_X(z)$ ist für $0 \leq z \leq 1$ monoton wachsend ($f'(z) \geq 0$) und konvex ($f''(z) \geq 0$).

3) Falls $E(X) < \infty$ ist, gilt $f'_X(1) = \lim\limits_{z \nearrow 1} f'_X(z) = E(X)$.

4) Falls $E(X^2) < \infty$ ist, gilt $f''_X(1) = \lim\limits_{z \nearrow 1} f''_X(z) = E(X^2) - E(X)$

Beweis:
Zu 1): Nachrechnen.

Zu 2): Gliedweises differenzieren liefert für $0 \leq z < 1 \quad f'(z) \geq 0, \quad f''(z) \geq 0$.

Zu 3):

Lemma von Abel[a]
Sei $\sum\limits_{n=0}^{\infty} c_n$ konvergente Reihe mit $c_n \in \mathbb{R}$ dann konvergiert $f(x) = \sum\limits_{n=0}^{\infty} c_n x^n$ für $x \in [0,1]$ und f ist im Intervall $[0,1]$ stetig.
Insbesondere: $\lim\limits_{x \nearrow 1} f(x) = f(1) = \sum\limits_{n=0}^{\infty} c_n$.

[a] s. Forster Band 1,S.180

Da $EX = \sum\limits_{k=1}^{\infty} kP(X=k) < \infty$, liefert das Lemma von Abel, angewandt auf $f_X'(z) = \sum\limits_{k=1}^{\infty} kP(X=k)z^{k-1}$, die Behauptung:
$f_X'(1) = \lim\limits_{z \nearrow 1} f_X'(z) = \lim\limits_{z \nearrow 1}(\sum\limits_{k=1}^{\infty} kP(X=k)z^{k-1}) = \sum\limits_{k=1}^{\infty} kP(X=k) = E(X)$.

Zu 4): $EX^2 < \infty \Rightarrow E(X) < \infty$. Wende das Lemma von Abel auf f_X'' an:
$f_X''(z) = \frac{d}{dz}(\sum\limits_{k=1}^{\infty} P(X=k)kz^{k-1}) = \sum\limits_{k=2}^{\infty} P(X=k)k(k-1)z^{k-2}$
(Terme mit $k=0$ und $k=1$ verschwinden).
$f_X''(1) = \lim\limits_{z \nearrow 1} f_X''(z) = \lim\limits_{z \nearrow 1} \sum\limits_{k=2}^{\infty} P(X=k)k(k-1)z^{k-2}$
$\overbrace{=}^{Abel} \sum\limits_{k=2}^{\infty} P(X=k)(k^2-k) = E(X^2) - E(X)$.

9.1.3 Satz

Seien X und Y unabhängige Zufallsgrößen mit Werten in $\mathbb{N} \cup \{0\}$.
Dann gilt: $f_{X+Y}(z) = f_X(z)f_Y(z)$.

Beweis: Zu Satz 9.1.3:

$$\begin{aligned}
f_{X+Y}(z) = E(z^{(X+Y)}) &= E(z^X z^Y) \\
&= \sum_{\omega\in\Omega} z^{X(\omega)} z^{Y(\omega)} p(\omega) \\
&\overbrace{=}^{\text{Trafosatz}} \sum_{k,l} z^k z^l P(X=k, Y=l) \\
&= \sum_{k,l} z^k z^l P(X=k)P(Y=l) \\
&= (\sum_k z^k P(X=k))(\sum_l z^l P(Y=l)) \\
&= f_X(z) f_Y(z)
\end{aligned}$$

9.1.4 Folgerung

Seien $X_1, X_2, ..., X_n$ unabhängige Zufallsgrößen. Dann gilt: $f_{\sum\limits_{i=1}^{n} X_i}(z) = \prod\limits_{i=1}^{n} f_{X_i}(z)$.

9.1.5 Beispiele

1) Seien $X_1, ..., X_n$ unabhängig und Bernoulli-verteilt mit $P(X_i = 1) = p, P(X_i = 0) = q$. Dann gilt: $f_{X_i}(z) = q + pz$. Setzt man $S_n = \sum\limits_{i=1}^{n} X_i$, so gilt für die erzeugende Funktion:
   $f_{S_n}(z) = (q + pz)^n = \sum\limits_{k=0}^{n} \binom{n}{k} p^k q^{n-k} z^k$.
   Wie zu erwarten war, ist das die erzeugende Funktion der Binomialverteilung.

2) Seien $X_1, X_2, ..., X_n$ unabhängig und geometrisch verteilt mit
   $P(X_i = k) = pq^k, \quad (k = 0, 1, 2, ...)$, so ergibt sich als erzeugende Funktion:
   $f_{X_i}(z) = \sum\limits_{k=0}^{\infty} pq^k z^k = \frac{p}{1-qz}$ und mit Satz 9.1.2, 3 folgt:

   $$E(X_i) = f'_{X_i}(1) = \frac{pq}{(1-qz)^2}|_{z=1} = \frac{q}{p}.$$

   Setzen wir wieder $S_n = \sum\limits_{i=1}^{n} X_i$, so folgt weiter:

   $$f_{S_n}(z) = (\frac{p}{1-qz})^n = p^n(1-qz)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} p^n (-q)^k z^k$$

   Dabei haben wir dieallgemeine binomische Formel angewandt:

   $$(1+s)^a = 1 + \binom{a}{1} s + \binom{a}{2} s^2 + \binom{a}{3} s^3 + ... \text{ mit } |s| < 1, a \in \mathbb{R}.$$

   Hierbei ist $\binom{a}{k} := \frac{a(a-1)...(a-k+1)}{k!}, a \in \mathbb{R}, k \in \mathbb{N} \cup \{0\}$,

   und es gilt: $\binom{-n}{k} = \frac{(-n)(-n-1)...(-n-k+1)}{k!} = (-1)^k \frac{n(n+1)...(n+k-1)}{k!} = (-1)^k \binom{n+k-1}{k}$.

   Damit ist: $f_{S_n}(z) = \sum\limits_{k=0}^{\infty} \binom{n+k-1}{k} p^n q^k z^k$

   und $P(S_n = k) = \binom{n+k-1}{k} p^n q^k$.

3) Seien $X_1, X_2, ..., X_n$ unabhängig und Poisson-verteilt mit
   $P(X_i = k) = \frac{\lambda_i^k}{k!} e^{-\lambda_i}, \quad (k = 0, 1, 2, ...)$, so ist die erzeugende Funktion:
   $f_{X_i}(z) = \sum\limits_{k=0}^{\infty} \frac{\lambda_i^k z^k}{k!} e^{-\lambda_i} = e^{\lambda_i z - \lambda_i} = e^{\lambda_i(z-1)}$.

   Für die erzeugende Funktion von $S_n = \sum\limits_{i=1}^{n} X_i$ gilt dann: $f_{S_n}(z) = e^{\sum\limits_{i=1}^{n} \lambda_i(z-1)}$.

   Wegen Satz 9.1.2 gilt: S_n ist Poisson-verteilt mit Parameter $\sum\limits_{i=1}^{n} \lambda_i$.

9.2 Poisson-Prozesse

9.2.1 Herleitung des Poisson-Prozesses durch erzeugende Funktionen

Sei X_t die Anzahl der „Anrufe“ im Zeitintervall $[0, t]$ und $X_0 = 0$, $0 \leq t \leq K$.

Annahme:

1) Sei $0 < t_1 < t_2 < ... < t_k < K$. Dann sind $X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_k} - X_{t_{k-1}}$ unabhängig.

2) $X_{t+a} - X_{s+a}$ ist für alle $a > 0$ genau so verteilt wie $X_t - X_s$.

3) a) $P(X_{t+\delta} - X_t = 1) = \lambda\delta + o(\delta)$ für $\delta \to 0$,
   b) $P(X_{t+\delta} - X_t = 0) = 1 - \lambda\delta + o(\delta)$ für $\delta \to 0$,

Bemerkung: (a) & (b)$\Rightarrow$ $P(X_{t+\delta} - X_t > 1) = o(\delta)$ für $\delta \to 0$,
denn

$$\begin{aligned} P(X_{t+\delta} - X_t > 1) &= 1 - P(X_{t+\delta} - X_t = 0) - P(X_{t+\delta} - X_t = 1) \\ &= 1 - [1 - \lambda\delta + o(\delta)] - [\lambda\delta + o(\delta)] \\ &= 2o(\delta) \\ &= o(\delta) \text{ für } \delta \to 0. \end{aligned}$$

Aufgabe: Berechne unter den Annahmen (1)-(3) die erzeugende Funktion von X_t.

Behauptung: $f_{X_t}(z) = e^{\lambda t(z-1)}$. X_t ist Poisson-verteilt mit Parameter λt.

Beweisidee:
Zerlege das Intervall $[0, t]$ in N Stücke der Länge $\frac{t}{N}$ und berechne die erzeugende Funktion von $X_{t/N}$. Berechne anschließend X_t.

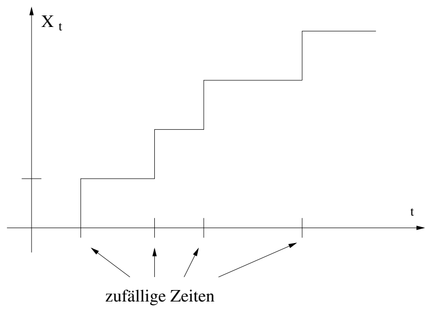
Aus 3)a), 3)b) und 3)c) folgt: $f_{X_{\frac{t}{N}}}(z) = 1 - \lambda\frac{t}{N} + \lambda z\frac{t}{N} + o(\frac{t}{N})$ für $N \to \infty$.
Wegen 2) ist dies auch eine erzeugende Funktion von $X_{\frac{2t}{N}} - X_{\frac{t}{N}}, X_{\frac{3t}{N}} - X_{\frac{2t}{N}}, ..., X_t - X_{\frac{(N-1)t}{N}}$.
Da $X_t = \sum\limits_{k=0}^{N}\left(X_{\frac{jt}{N}} - X_{\frac{(j-1)t}{N}}\right)$ ist, folgt mit 1) und Satz 9.1.3

$$\begin{aligned} f_{X_t}(z) &= (f_{X_{\frac{t}{N}}}(z))^N \\ &= (1 - \frac{\lambda t}{N} + \frac{\lambda z t}{N} + o(\frac{t}{N}))^N \\ &= (1 + \frac{\lambda t(z-1) + o(\frac{1}{N})}{N})^N \text{ für } N \to \infty \\ &= e^{\lambda t(z-1)} \text{ für } N \to \infty. \end{aligned}$$

Dabei haben wir verwendet: $(1 + \frac{x_n}{n})^n \to e^x$, falls $x_n \to x$ und $n \to \infty$.

9.2.2 Definition (Poisson-Prozess)

Die Menge der Zufallsvariablen $\{X_t, 0 \leq t \leq K\}$ mit den Eigenschaften (1)–(3) heißt Poisson-Prozess zum Parameter λ.



Bemerkung: Eigenschaft (3) lässt sich ersetzen durch (3'): Für $0 \leq s \leq t$ gilt: $X_t - X_s$ ist Poisson-verteilt mit Parameter $\lambda(t-s)$. Wir haben in 7.1.9, 3) gezeigt: (1)-(3)$\Rightarrow$(3'). Umgekehrt läßt sich leicht zeigen: (1)+(2)+(3')$\Rightarrow$(3).

9.3 Ausgedünnte Poisson-Prozesse

9.3.1 Satz

Sei $\{X_t, t \in [0,K]\}$ Poisson-Prozess mit Parameter λ. Seien $Z_1, Z_2, ...$ Bernoulli-Variable mit $P(Z_i = 1) = p$. Weiter seien $X_t, Z_1, Z_2, ...$ für unabhängig (d.h. für jedes $n \geq 1$ und sind $X_t, Z_1, Z_2, ..., Z_n$ unabhängig).

Dann ist $Y_t = \sum\limits_{i=1}^{X_t} Z_i$ Poisson-verteilt mit Parameter λpt.

Weitergehend ist $\{Y_t, t \in [0,K]\}$ Poisson-Prozess mit Parameter λp, wenn die geforderte Unabhängigkeitseigenschaft für alle $t > 0$ gilt.

Beweis:

Zeige, dass $f_{Y_t}(z) = e^{\lambda pt(z-1)}$ gilt. Sei $S_n = \sum\limits_{i=1}^{n} Z_i$. Dann ist:

$$P(Y_t = k) = P(S_{X_t} = k) = \sum_{n=0}^{\infty} P(X_t = n, S_n = k) = \sum_{n=0}^{\infty} P(X_t = n) P(S_n = k)$$

$$\begin{aligned} f_{Y_t}(z) &= \sum_{k=0}^{\infty} P(Y_t = k) z^k \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} P(Y_t = n) P(S_n = k) z^k \\ &= \sum_{n=0}^{\infty} P(X_t = n) (\sum_{k=0}^{\infty} P(S_n = k) z^k) \end{aligned}$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} P(X_t = n) f_{S_n}(z) \\
&\overbrace{=}^{\text{Satz 9.1.3}} \sum_{n=0}^{\infty} P(X_t = n)(f_{Z_1}(z))^n \\
&= f_{X_t}(f_{Z_1}(z)) \\
&= f_{X_t}(1 + p(z-1)) \\
&= e^{\lambda t p(z-1)}.
\end{aligned}$$

9.3.2 Korollar

Sei $Y_t := \sum_{i=1}^{X_t} Z_i$ und $W_t := \sum_{i=1}^{X_t} (1 - Z_i)$. Dann sind Y_t und W_t unabhängig.

Beweis:
Aus Satz 9.3.1 folgt, daß auch W_t Poisson-verteilt ist mit Parameter $\lambda(1-p)$. Die Aussage über die Unabhängigkeit folgt aus dem folgenden Lemma, wenn man $N = X_t$ setzt.

9.3.3 Lemma

Sei N Poisson-verteilte Zufallsvariable mit Parameter λ. Seien $Z_1, Z_2, \ldots$ Bernoulli-Variable mit $P(Z_i = 1) = p$. Für jedes $n \geq 1$ seien $N, Z_1, Z_2, \ldots, Z_n$ unabhängig. Dann sind die Zufallsvariablen

$$Y = \sum_{i=1}^{N} Z_i \ , \qquad W = \sum_{i=1}^{N} (1 - Z_i)$$

unabhängig.

Beweis:
Es gilt wegen der Unabhängigkeitsvoraussetzung und wegen $N = Y + W$

$$\begin{aligned}
P(Y = k, W = n - k) &= P(Y = k, W = n - k, N = n) \\
&= P(Y = k, W = n - k | N = n) \ P(N = n) \\
&= P\left(\sum_{i=1}^{n} Z_i = k, \sum_{i=1}^{n} (1 - Z_i) = n - k \right) P(N = n) \\
&= P\left(\sum_{i=1}^{n} Z_i = k \right) P(N = n) \\
&= \frac{n!}{k!(n-k)!} \, p^k (1-p)^{n-k} \ \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \frac{(p\lambda)^k}{k!} \, e^{p\lambda} \ \frac{[(1-p)\lambda]^{n-k}}{(n-k)!} \ e^{-(1-p)\lambda} \\
&= P(Y = k) \ P(W = n - k).
\end{aligned}$$

Die letzte Gleichung benutzt von Satz 9.3.1 die erste Aussage.

Die Kükenaufgabe:
Auf einer Farm legen die Hühner innerhalb eines Jahres N Eier, N sei Poisson-verteilt mit Parameter λ. Aus jedem Ei schlüpft unabhängig von den anderen Eier ein Küken mit Wahrscheinlichkeit p. Wenn Y die Anzahl der geschlüpften Küken ist, was ist $E(N|Y=k)$?

Nun ist $Y = \sum\limits_{i=1}^{N} Z_i$, mit Z_i unabhängig und Bernoulli-verteilt mit Parameter p. Sei $W = \sum\limits_{i=1}^{N}(1-Z_i)$. Dann ist $N = Y + W$ und Y und W sind unabhängig nach dem Lemma. Somit gilt

$$
\begin{aligned}
E(N|Y=k) &= \frac{1}{P(Y=k)}\, E\left(N I_{\{Y=k\}}\right) \\
&= \sum_{l=0}^{\infty} l\, \frac{P(N=l,\ Y=k)}{P(Y=k)} \\
&= \sum_{l=k}^{\infty} l\, \frac{P(W=l-k,\ Y=k)}{P(Y=k)} \\
\text{mit } m=l-k \qquad &= k + \sum_{m=0}^{\infty} m\, P(W=m) \\
&= k + \lambda(1-p).
\end{aligned}
$$

Dies folgt, da W Poisson-verteilt ist mit Parameter $\lambda(1-p)$.

9.4 Poisson-Prozess über dem Einheitsquadrat

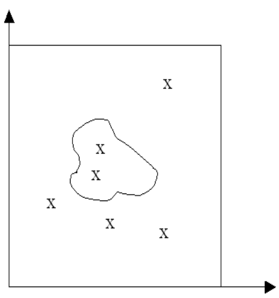
Sei $E = [0,1]^2$ und $B \subset E$. X_B sei die Anzahl der „Treffer“ auf B (z.B. Raindrops).

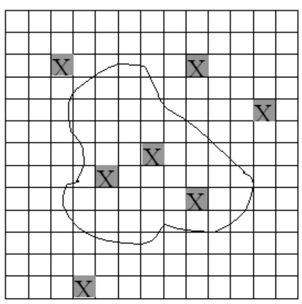
Annahmen:

1) Gilt $A, B \subset E$ und $A \cap B = \emptyset$, dann sind X_A, X_B unabhängig.

2) Sei $s \in E$, $A \subset E$ sowie $A + s \subset E$. Dann ist X_{A+s} genauso verteilt wie X_A.

3) $F(A)$ sei die Fläche von A. Dann gilt für $F(A) \to 0$:
   a) $P(X_A = 1) = \lambda F(A) + o(F(A))$,
   b) $P(X_A = 0) = 1 - \lambda F(A) + o(F(A))$.





Aufgabe: Leite unter den Annahmen 1)-3) die erzeugende Funktion für die Anzahl der „Treffer“ her.

Zerlege E in $n \times n$ kleine Quadrate $Q_{i,n}$, $(i \in \{1, \dots n^2\})$ und nummeriere diese durch. K_n bzw. G_n seien die Indexmengen der einbeschriebenen bzw. der einbeschreibenden Quadrate der Menge A. Es gilt dann: $\bigcup\limits_{i \in K_n} Q_{i,n} \subset A \subset \bigcup\limits_{i \in G_n} Q_{i,n}$.
Definiere: $\tilde{X}_{K_n} := X_{\bigcup\limits_{K_n} Q_{i,n}}$ und $\tilde{X}_{G_n} := X_{\bigcup\limits_{G_n} Q_{i,n}}$.
Dann gilt für alle n $\tilde{X}_{K_n} \leq X_A \leq \tilde{X}_{G_n}$ und damit folgt: $\lim\limits_{n\to\infty} \uparrow \tilde{X}_{K_n} = X_A = \lim\limits_{n\to\infty} \downarrow \tilde{X}_{G_n}$.
Für die erzeugenden Funktionen folgt:

$$
\begin{aligned}
f_{\tilde{X}_{K_n}}(z) &\overbrace{=}^{(1)} \prod_{i \in K_n} f_{X_{Q_{i,n}}}(z) \\
&\overbrace{=}^{(3)} \prod_{i \in K_n} (1 - \lambda F(Q_{i,n}) + \lambda F(Q_{i,n}) z + o(F(Q_{i,n}))) \\
&\overbrace{=}^{(2)} (1 - \frac{\lambda}{n^2} + \frac{\lambda}{n^2} z + o(\frac{1}{n^2}))^{F(A) n^2 (1 + o(1))} \text{ für } n \to \infty \\
&\to e^{\lambda F(A)(z-1)} \text{ für } n \to \infty.
\end{aligned}
$$

Ebenso gilt: $f_{\tilde{X}_{G_n}}(z) \to e^{\lambda F(A)(z-1)}$ und damit folgt: $f_{X_A}(z) = e^{\lambda F(A)(z-1)}$ d.h. X_A ist Poisson-verteilt mit Parameter $\lambda F(A)$.

Bemerkung :

1) Diese Konstruktion funktioniert auch für $[0,1]^k$.

2) Satz 9.3.1 und das Korollar gelten entsprechend.

10 Markov-Ketten

10.1 Die Kain und Abel-Aufgabe (nach A. Engel)

Abel schlägt seinem Bruder Kain folgendes Spiel vor. Sie werfen abwechselnd eine faire Münze, bis erstmals eine der Ziffernfolgen

a) 1111 oder

b) 0011 auftritt.

Kain gewinnt bei (a) und Abel bei (b). Wie groß ist die Wahrscheinlichkeit, daß Kain gewinnt? Ein möglicher Pfad im Verlauf des Spiels ist 01100101110011.

Wir stellen nun die Struktur der Aufgabe mit Hilfe eines Graphen dar, wobei die "Zustände" die möglichen Ergebnismuster sind und die Pfeile die möglichen Übergänge angeben. Alle Übergänge geschehen mit Wahrscheinlichkeit $\frac{1}{2}$.

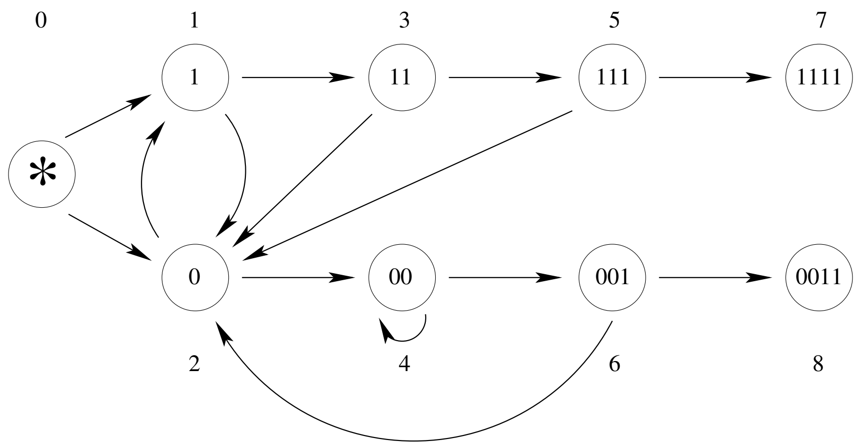


Abbildung 10.1: Der Graph der möglichen Übergänge

Die Zustände sind mit den Ziffern 0 bis 8 bezeichnet, wobei "0" den Zustand benennt, bei dem noch kein Ergebnis vorliegt. Der angegebene Spielverlauf 01100101110011 übersetzt sich nun (in eineindeutiger Zuordnung) in 21324621352468. Wie läßt sich nun die Aufgabe lösen? Wir bezeichnen mit p_i die Wahrscheinlichkeit bei Start im Zustand "i" den Zustand "7" zu erreichen bevor man den Zustand "8" erreicht hat. Dann gilt natürlich sofort $p_7 = 1$ und $p_8 = 0$. Außerdem sind die folgenden Gleichungen intuitiv plausibel.

$p_1 = \frac{1}{2}p_2 + \frac{1}{2}p_3$

$p_2 = \frac{1}{2}p_1 + \frac{1}{2}p_4$

$p_3 = \frac{1}{2}p_2 + \frac{1}{2}p_5$

$p_4 = \frac{1}{2}p_4 + \frac{1}{2}p_6$

$p_5 = \frac{1}{2} + \frac{1}{2}p_2$

$p_6 = \frac{1}{2}p_2.$

Das Schema ist ein lineares Gleichungssystem mit sechs Unbekannten und Gleichungen. Durch Einsetzen erhält man sofort

$p_4 = p_6 = \frac{1}{2}p_2$

$p_1 = \frac{3}{2}p_2$

$p_3 = \frac{1}{4} + \frac{3}{4}p_2$

sowie $\quad p_1 = \frac{1}{8} + \frac{7}{8}p_2.$

Damit ist $p_1 = \frac{3}{10}$, $p_2 = \frac{1}{5}$ und folglich $p_0 = \frac{1}{4}$. Das heißt Kain gewinnt nur mit Wahrscheinlichkeit $\frac{1}{4}$.

10.2 Definition von Markov-Ketten und erste Folgerungen

10.2.1 Definition (Markov-Kette)

(Ω, P) sei Wahrscheinlichkeitsraum, E sei endliche Menge. Seien $X_i : \Omega \to E$, $i = 0, 1, 2, 3, \ldots$ Zufallsvariablen. Die Menge der Zufallsvariablen $\mathcal{X} = \{X_0, X_1, X_2, \ldots, X_n, \ldots\}$ heißt Markov-Kette (der Länge n), falls

$$P(X_i = x_i \mid X_0 = x_0, \ldots, X_{i-1} = x_{i-1}) = P(X_i = x_i \mid X_{i-1} = x_{i-1})$$

für $i = 1, 2, \ldots, n, \ldots$ gilt. Eine Markov-Kette $\mathcal{X}$ heißt stationär, falls

$$P(X_i = x_1 \mid X_{i-1} = x_0) = P(X_1 = x_1 \mid X_0 = x_0)$$

für $i = 1, 2, \ldots$ gilt.

Bezeichnungsweisen

1. E heißt Zustandsraum,
2. $q(x, y), x, y \in E$ heißt stochastische Matrix, falls $q(x, y) \geq 0$ und $\sum\limits_{y \in E} q(x, y) = 1$, $\forall\, x \in E$ gilt,
3. $\pi(x_0) = P(X_0 = x_0)$ mit $x_0 \in E$ heißt Startverteilung.

Konstruktion von stationären Markov-Ketten

Gegeben seien:

1. E endlich oder abzählbar,
2. $q(x, y), x, y \in E$ eine stochastische Matrix,
3. $\pi(x), x \in E$ eine Wahrscheinlichkeitsfunktion auf E, d.h. $\pi(x) \geq 0$ für alle $x \in E$ und $\sum\limits_{x \in E} \pi(x) = 1$.

Zur Existenz:
Sei $\Omega_{n+1} := \{\omega = (x_0, \ldots, x_n) \mid x_i \in E, i = 0, \ldots, n\}$ und $X_i(\omega) := x_i$, für $i = 0, 1, \ldots, n$.

10.2.2 Satz:

Sei $n \in \mathbb{N}$. Durch $p(\omega) := \pi(x_0)q(x_0, x_1) \ldots q(x_{n-1}, x_n)$ wird eine Wahrscheinlichkeitsfunktion auf Ω_{n+1} gegeben, so dass gilt:

1. $P(X_0 = x_0) = \pi(x_0)$,
2. $P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = p(\omega)$,
3. $P(X_{i+1} = x_{i+1} | X_0 = x_0, \ldots, X_i = x_i) = q(x_i, x_{i+1}) = P(X_{i+1} = x_{i+1} | X_i = x_i)$ für $i = 0, \ldots, n-1$.

Beweis:
Zu (2):
Sei $\omega = (x_0, \ldots, x_n)$. Mit $p(\omega) = \pi(x_0)q(x_0, x_1) \ldots q(x_{n-1}, x_n)$ gilt: $p(\omega) \geq 0$ und $\sum\limits_{\omega \in \Omega_{n+1}} p(\omega) = 1$. Da $X_i(\omega) = x_i$ ist, folgt $P(\{\omega | X_i(\omega) = x_i, i = 0, \ldots, n\}) = P(\{\omega\}) = p(\omega)$.
Zu (1):
(+) $P(\{X_0 = x_0, \ldots, X_i = x_i\}) = \sum\limits_{x_{i+1}, \ldots, x_n} \pi(x_0)q(x_0, x_1)q(x_{i-1}, x_i) \ldots q(x_{n-1}, x_n)$
$= \pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i)$
für $i = 0, 1, 2, \ldots, n-1$.
Zu (3):
Es gilt

$$\begin{aligned} P(X_{i+1} = x_{i+1} \mid X_0 = x_0, \ldots, X_i = x_i) &= \frac{\pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i), q(x_i, x_{i+1})}{\pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i)} \\ &= q(x_i, x_{i+1}) \end{aligned}$$

sowie

$$\begin{aligned} P(X_{i+1} = x_{i+1} | X_i = x_i) &= \frac{P(X_{i+1} = x_{i+1}, X_i = x_i)}{P(X_i = x_i)} \\ &= \frac{\sum\limits_{x_0, \ldots, x_{i-1}} \pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i)q(x_i, x_{i+1})}{\sum\limits_{x_0, \ldots, x_{i-1}} \pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i)} \\ &= q(x_i, x_{i+1}). \end{aligned}$$

Bemerkungen:

1. $P(X_i = x_i) = \sum\limits_{x_0, \ldots, x_{i-1}} \pi(x_0)q(x_0, x_1) \ldots q(x_{i-1}, x_i) = \sum\limits_{x_0} \pi(x_0)q^i(x_0, x_i)$.
   Dabei ist $q^{n+1}(x, y) = \sum\limits_{z \in E} q^n(x, z)\, q(z, y)$ das $n+1$-fache Matrixprodukt.

2. Mit Satz 10.2.2 ist eine Markov-Kette der Länge n konstruiert!

3. Wegen $(+)$ im Beweis von Satz 10.2.2 ist die Markov-Kette der Länge i eingebettet in eine Markov-Kette der Länge n für $i \leq n$.

4. Mit der Bedingung $(+)$ lässt sich eine Markov-Kette beliebiger Länge konstruieren.

10.2.3 Beispiele für Markov-Ketten

1) Kain und Abel-Aufgabe:
Der Zustandsraum ist $E = \{0, 1, 2, \ldots, 8\}$. Für π gilt $\pi(0) = 1$. Weiter ist

$$(q(x,y))_{x,y\in E} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

2) Sei $E = \{1, 2, 3\}$,

$$q(x,y) = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} \text{ und } q^2(x,y) = \begin{pmatrix} \frac{5}{8} & 0 & \frac{3}{8} \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ 0 & \frac{3}{4} & \frac{1}{4} \end{pmatrix}.$$

Für jede Startverteilung π mit $\pi(x) > 0$ für $x \in E$ gilt $\lim\limits_{n\to\infty} \sum\limits_{x\in E} \pi(x) q^n(x,y) = \alpha(y)$ mit $\alpha(1) = \frac{8}{19}$, $\alpha(2) = \frac{6}{19}$, $\alpha(3) = \frac{5}{19}$.

3) Seien $Y_1, Y_2, \ldots$ unabhängig, $X_i = \sum\limits_{j=1}^{i} Y_j$ und $X_0 = x_0$. Dann ist $\{X_0, X_1, X_2, \ldots\}$ eine Markov-Kette.

Beweis:
Sei $m \in \mathbb{N}$ und sei $y_i = x_i - x_{i-1}$ für $i \geq 1$ und $y_0 = x_0$. Dann ist:

$$\begin{aligned} & P(X_{m+1} = x | X_0 = x_0, \ldots, X_m = x_m) \\ &= P(X_{m+1} = x | X_0 = y_0, X_1 = y_0 + y_1, X_2 = y_0 + y_1 + y_2, \ldots, X_m = y_0 + \cdots + y_m) \\ &= P(Y_{m+1} = x - (y_0 + y_1 + \cdots + y_m) | Y_0 = y_0, Y_1 = y_1, \ldots, Y_m = y_m) \\ &= \frac{P(Y_{m+1} = x - (y_0 + y_1 + \cdots + y_m)) P(Y_0 = y_0) P(Y_1 = y_1) \cdots P(Y_m = y_m)}{P(Y_0 = y_0) P(Y_1 = y_1) \cdots P(Y_m = y_m)} \\ &= P(Y_{m+1} = x - (y_0 + y_1 + \cdots + y_m)) \\ &= P(X_{m+1} = x | X_m = x_m). \end{aligned}$$

4) **Ehrenfestsches Urnenmodell**

N Teilchen befinden sich in den Behältern 1 und 2.
i Teilchen seien in Behälter 1, $N - i$ Teilchen in Behälter 2.
In jeder Zeiteinheit springt ein Teilchen entweder von $1 \to 2$ oder von $2 \to 1$.

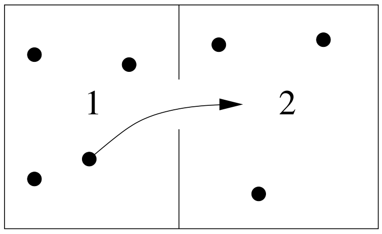


Abbildung 10.2: Ehrenfestsches Urnenmodell

Die stochastische Matrix wird gegeben durch $q(i, i+1) = \frac{N-i}{N}, q(i, i-1) = \frac{i}{N}$ und $q(i, j) = 0$ für $j \neq i \pm 1$.

10.3 Absorbierende Zustände

Wir können eine Markov-Kette als Bewegung eines Teilchens durch die Zustände auffassen. In manchen Markov-Ketten gibt es Zustände, die nicht verlassen werden können. Solche Zustände heißen absorbierend.

10.3.1 Definition (absorbierender Zustand, absorbierender Rand)

Sei $\mathcal{X} = \{X_0, X_1, \ldots\}$ eine stationäre Markov-Kette mit Zustandsraum E.
Ein Zustand $x \in E$ heißt absorbierend, wenn gilt $q(x, x) = 1$.
Die Menge $A \subset E$ aller absorbierender Zustände heißt absorbierender Rand von $\mathcal{X}$.

Bei der "Kain und Abel"-Aufgabe ist $\{7, 8\}$ der absorbierende Rand. Der folgende Satz liefert eine Begründung für die Kain und Abel-Aufgabe.

10.3.2 Satz

Sei $\mathcal{X}$ eine stationäre Markov-Kette und $A \subset E$ ihr absorbierender Rand. Wir nehmen an, dass für jedes $x \in E$ die Wahrscheinlichkeit nach A zu kommen gleich 1 ist.
Sei $A = A_1 \cup A_2$ und $A_1 \cap A_2 = \emptyset$. Sei $P_{A_1}(x)$ die Wahrscheinlichkeit bei Start in x beim ersten Eintritt in A nach A_1 zu gelangen. Dann gilt:

(a) $P_{A_1}(x) = 1$ für $x \in A_1$ und $P_{A_1}(x) = 0$ für $x \in A_2$,

(b) Für $x \notin A$ ist $P_{A_1}(x) = \sum\limits_{z \in A_1} q(x, z) + \sum\limits_{n \geq 2} \sum\limits_{\substack{x_i \notin A \text{ für} \\ i=1,\ldots,n-1, \\ z \in A_1}} q(x, x_1) q(x_1, x_2) \ldots q(x_{n-1}, z)$,

(c) Es gilt stets $P_{A_1}(x) = \sum\limits_{y \in E} q(x, y) P_{A_1}(y)$.

Beweis:
Zu (a): Dies folgt direkt aus der Definition.

Zu (b): Sei $x \notin A$. Dann ist

$$\begin{aligned}
P_{A_1}(x) &= P(X_0 = x, X_1 \in A_1) + P\Big(\bigcup_{n\geq 2}\{X_0 = x, X_i \notin A \text{ für } i < n, X_n \in A_1\}\Big) \\
&= \sum_{z\in A_1} q(x,z) + \sum_{n\geq 2} P(\{X_0 = x, X_i \notin A \text{ für } i < n, X_n \in A_1\}) \\
&= \sum_{z\in A_1} q(x,z) + \sum_{n\geq 2} \sum_{\substack{x_1,x_2,\dots,x_{n-1}\notin A,\\ z\in A_1}} q(x,x_1)\dots q(x_{n-1},z).
\end{aligned}$$

Zu (c): Sei $x \notin A$. So gilt

$$\begin{aligned}
P_{A_1}(x) &= P(X_0 = x, X_1 \in A_1) + P(X_0 = x, X_1 \notin A, X_2 \in A_1) \\
&\quad + P(X_0 = x, \exists n \geq 3 \text{ mit } X_n \in A_1 \text{ und } X_i \notin A \text{ für } i < n) \\
&= \sum_{y\in A_1} q(x,y) + \sum_{y\notin A, z\in A_1} q(x,y)q(y,z) \\
&\quad + \sum_{n\geq 3}\sum_{y\notin A}\sum_{\substack{x_i\notin A,\\ i=2,\dots,n-1,\\ z\in A_1}} q(x,y)q(y,x_2)\dots q(x_{n-1},z) \\
&\overset{(*)}{=} \sum_{y\in A} q(x,y)P_{A_1}(y) + \sum_{y\notin A} q(x,y)\Big(\sum_{z\in A_1} q(y,z)\Big) \\
&\quad + \sum_{y\notin A} q(x,y)\Big(\sum_{n\geq 3}\sum_{\substack{x_i\notin A,\\ i=2,\dots,n-1,\\ z\in A_1}} q(y,x_2)\dots q(x_{n-1},z)\Big) \\
&= \sum_{y\in A} q(x,y)P_{A_1}(y) + \sum_{y\notin A} q(x,y)P_{A_1}(y) \\
&= \sum_{y} q(x,y)P_{A_1}(y).
\end{aligned}$$

(*) $P_{A_1}(x) = 1$ für $x \in A_1$ und $P_{A_1}(x) = 0$ für $x \in A\backslash A_1$.

Bemerkung: Die Gleichung c) aus Satz 10.3.2 lautet in Vektorschreibweise $P_{A_1} = qP_{A_1}$. Dies bedeutet P_{A_1} ist rechter Eigenvektor von q zum Eigenwert 1. Man sagt auch, P_{A_1} ist harmonisch oder P_{A_1} erfüllt die Mittelwerteigenschaft.

10.3.3 Berechnung von Ruin-Wahrscheinlichkeiten

Hans und Rudolf spielen ein Spiel. In jeder Runde gewinnt Hans mit der Wahrscheinlichkeit p und Rudolf gewinnt mit der Wahrscheinlichkeit $q = 1 - p$. Der Gewinner einer Runde erhält von seinem Gegner einen Euro. Es wird so lange gespielt bis einer der Spieler kein Geld mehr hat. Wie hoch ist die Ruin-Wahrscheinlichkeit $P(x)$ von Hans, wenn Hans zu Beginn x Euro hat und Hans und Rudolf zusammen b Euro haben?

Seien $x, b \in \mathbb{N}$ mit $0 < x < b$. Weiter seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen mit $P(X_i = 1) = p = 1 - P(X_i = -1)$ für alle i und es sei $S_0 = x$ und $S_n = x + \sum_{i=1}^{n} X_i$ für $n \geq 1$. Damit ist $S_0, S_1, S_2, \ldots$ eine Markov-Kette und S_n ist das Kapital, das Hans nach n Runden besitzt. Wir frage nach der Ruin-Wahrscheinlichkeit:

$$P(x) = P\left(\exists n \text{ mit } S_n = 0 \text{ und } 0 < S_i < b \text{ für } i < n \mid S_0 = x\right).$$

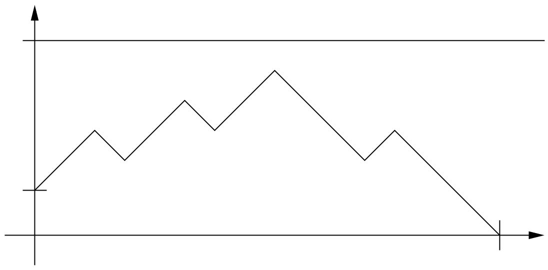


Abbildung 10.3: Ruin-Wahrscheinlichkeiten

10.3.4 Satz

Sei $q = 1 - p$.

Für $p = \dfrac{1}{2}$ gilt $P(x) = 1 - \dfrac{x}{b}$ für $0 \leq x \leq b$.

Für $p \neq \dfrac{1}{2}$ gilt $P(x) = \dfrac{(\frac{q}{p})^b - (\frac{q}{p})^x}{(\frac{q}{p})^b - 1}$ für $0 \leq x \leq b$.

Beweis:
Fall 1: Sei $p = \frac{1}{2}$.
Satz 10.3.2 liefert: $P(x) = \frac{1}{2}P(x-1) + \frac{1}{2}P(x+1)$ für $0 < x < b$, $P(0) = 1$, $P(b) = 0$.
Durch Lösen dieses linearen Gleichungssystems ergibt sich: $P(x) = 1 - \frac{x}{b}$.
Fall 2: Sei $p \neq \frac{1}{2}$.
Satz 10.3.2 liefert: $P(x) = pP(x+1) + qP(x-1)$ für $0 < x < b$, $P(b) = 0$, $P(0) = 1$.
Dann ist

$$\begin{aligned} P(x) &= pP(x+1) + qP(x-1) \\ pP(x) + qP(x) &= pP(x+1) + qP(x-1) \\ P(x+1) - P(x) &= \tfrac{q}{p}(P(x) - P(x-1)). \end{aligned}$$

Wiederholtes Anwenden liefert

$$P(x+1) - P(x) = (\tfrac{q}{p})^x (P(1) - P(0)).$$

Sei $r := \frac{q}{p}$.
Aufaddieren liefert $P(x) - P(0) = \frac{r^x - 1}{r - 1}(P(1) - P(0))$.
Wegen $P(0) = 1$ gilt $P(x) = 1 + \frac{r^x - 1}{r - 1}(P(1) - 1)$.
Wegen $P(b) = 0$ gilt $P(1) - 1 = -\frac{r-1}{r^b - 1}$.
Einsetzen in die vorangegangene Gleichung liefert die Behauptung.

10.4 Rekurrente und transiente Zustände

10.4.1 Bezeichnungen

Sei $f^n(x,y) := P(X_n = y, X_i \neq y \text{ für } 1 \leq i < n | X_0 = x)$. Dann bezeichnet $f^n(x,y)$ die Wahrscheinlichkeit bei Start in x nach n Schritten erstmals den Zustand y zu erreichen.

$f^*(x,y) := P(X_n = y \text{ für ein } n \geq 1 | X_0 = x) = \sum_{n\geq 1} f^n(x,y)$ ist die Wahrscheinlichkeit bei Start in x irgendwann nach y zu gelangen.

$q^*(x,y) := \sum_{n=0}^{\infty} q^n(x,y)$ ist die erwartete Anzahl der Besuche in y bei Start in x, denn:

$$\begin{aligned} q^*(x,y) &= \sum_{n=0}^{\infty} q^n(x,y) = \sum_{n=0}^{\infty} P(X_n = y | X_0 = x) \\ &= \sum_{n=0}^{\infty} E(1_{\{X_n = y\}} | X_0 = x) = E\left(\sum_{n=0}^{\infty} 1_{\{X_n=y\}} \middle| X_0 = x \right). \end{aligned}$$

10.4.2 Definition (rekurrenter Zustand)

$x \in E$ heißt rekurrent, falls $f^*(x,x) = 1$,
$x \in E$ heißt transient, falls $f^*(x,x) < 1$.

10.4.3 Satz

1. $x \in E$ ist rekurrent $\Leftrightarrow q^*(x,x) = \infty$,
2. $x \in E$ ist transient $\Leftrightarrow q^*(x,x) < \infty$.

Beweis: Dieser folgt aus nachfolgendem Lemma.

10.4.4 Lemma

Es gilt

a) $q^*(x,y) = f^*(x,y)\, q^*(y,y) + \delta(x,y)$,

b) $f^*(x,x) = \frac{q^*(x,x)-1}{q^*(x,x)}$,

c) $q^*(x,x) = \frac{1}{1-f^*(x,x)}$.

Beweis:
Für alle $x, y \in E$ gilt: $q^n(x,y) = \sum_{i=1}^{n} f^i(x,y) q^{n-i}(y,y)$. Denn es gilt mit
$T_y = \min\{m \geq 1 | X_m = y\}$

$$q^n(x,y) = P(X_n = y | X_0 = x)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} P(T_y = i | X_0 = x) P(X_n = y | T_y = i, X_0 = x) \\
&= \sum_{i=1}^{n} f^i(x,y) P(X_n = y | X_i = y) \\
&= \sum_{i=1}^{n} f^i(x,y) P(X_{n-i} = y | X_0 = y) \\
&= \sum_{i=1}^{n} f^i(x,y) q^{n-i}(y,y).
\end{aligned}
$$

Weiter gilt nach Vertauschung der Summation

$$
\begin{aligned}
q^*(x,y) &= \sum_{i=1}^{\infty} q^n(x,y) + \delta(x,y) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{n} f^i(x,y) q^{n-i}(y,y) + \delta(x,y) \\
&= \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} f^i(x,y) q^{n-i}(y,y) + \delta(x,y) \\
&= \sum_{i=1}^{\infty} f^i(x,y) \sum_{m=0}^{\infty} q^m(y,y) + \delta(x,y) \\
&= f^*(x,y) \cdot q^*(y,y) + \delta(x,y).
\end{aligned}
$$

Damit ist a) gezeigt. Für $x = y$ folgt $q^*(x,x) - f^*(x,x)q^*(x,x) = 1$.
Auflösen nach $q^*(x,x)$ liefert $q^*(x,x)(1 - f^*(x,x)) = 1$ und damit $q^*(x,x) = \frac{1}{1-f^*(x,x)}$.
Auflösen nach $f^*(x,x)$ liefert: $f^*(x,x) = \frac{q^*(x,x)-1}{q^*(x,x)}$.

10.4.5 Beispiel auf $\mathbb{Z}$ (Irrfahrt auf $\mathbb{Z}$)

Seien $Y_1, Y_2, \ldots$ unabhängige Zufallsvariable mit $P(Y_i = 1) = p$, $P(Y_i = -1) = q$ für alle i. Weiter sei $X_0 = 0$ und $X_n = Y_1 + Y_2 + \cdots + Y_n$ für $n \geq 1$.
Behauptung: Für $p = q = \frac{1}{2}$ ist 0 rekurrent, für $p \neq q$ ist 0 transient.

Beweis:
Wende Satz 10.4.3 an! Sei $n \in \mathbb{N}$. Nach $2n + 1$ Schritten liegt ein ungerader Zustand vor, also kann X_{2n+1} nicht 0 sein und damit gilt: $q^{2n+1}(0,0) = 0$. Fällt in $2n$ Experimenten insgesamt n-mal die 1 und n-mal die -1, so ist $X_{2n} = 0$ und es gilt: $q^{2n}(0,0) = \binom{2n}{n} p^n q^n$.
Für große n können wir die Stirling'schen Formel anwenden:

$$
q^{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}} (4pq)^n \text{ , da } \binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} 2^{2n}.
$$

Fall 1: $p \neq q \Rightarrow (4pq) < 1 \Rightarrow \sum_{n \geq 1} q^{2n}(0,0) = \sum\limits_{n \geq 1} \frac{1}{\sqrt{\pi n}} (4pq)^n (1 + \mathrm{o}(1)) < \infty \Rightarrow$ „0“ ist transient.
Fall 2: $p = q = \frac{1}{2} \Rightarrow (4pq) = 1$

$\Rightarrow q^{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}}$
$\Rightarrow \sum\limits_n q^{2n}(0,0) \sim \sum\limits_n \frac{1}{\sqrt{\pi n}} = \infty \Rightarrow$ „0“ ist rekurrent.

10.4.6 Definition (irreduzibel)

Die stochastische Matrix q heißt irreduzibel, falls für alle $x, y \in E$ ein $m \in \mathbb{N}$ existiert mit $q^m(x,y) > 0$.

10.4.7 Satz

Sei q irreduzibel. Dann gilt: Existiert ein $z \in E$, das rekurrent ist, so sind alle $x \in E$ rekurrent.

Beweis:
Sei $x \in E$ beliebig. Dann existieren $k, l \in \mathbb{N}$ mit $q^k(x,z) > 0$ und $q^l(z,x) > 0$.
Dann ist

$$q^n(x,x) \geq q^k(x,z) q^m(z,z) q^l(z,x),$$

falls $n = k + m + l$ ist. Es folgt

$$q^*(x,x) = \sum_{n \geq 0} q^n(x,x) = q^k(x,z) \sum_{m \geq 0} q^m(z,z) q^l(z,x).$$

Da die rechte Seite nach Voraussetzung gleich unendlich ist, ist auch $q^*(x,x) = \infty$ und damit $f^*(x,x) = 1$.

10.5 Stationäre Verteilungen

10.5.1 Definition (stationäre Verteilung)

Eine Wahrscheinlichkeitsfunktion π auf E heißt stationär bezüglich q, falls für alle $y \in E$ gilt:

$$\sum_x \pi(x) q(x,y) = \pi(y).$$

In Vektorschreibweise: $\pi^T q = \pi^T$, d. h. π ist linker Eigenvektor von q mit Eigenwert 1.

10.5.2 Bemerkung

Ist π stationär und Startverteilung von $\mathcal{X}$, d.h. $P(X_0 = x) = \pi(x)$, so gilt:

$$P(X_n = y) = \pi(y).$$

Beweis: Es gilt

$$P(X_n = y) = \sum_x P(X_n = y \mid X_0 = x) \pi(x)$$

$$\begin{aligned}
&= \sum_x \pi(x) q^n(x,y) \\
&= \sum_x \pi(x) \sum_z q(x,z) q^{n-1}(z,y) \\
&= \sum_z \sum_x \pi(x) q(x,z) q^{n-1}(z,y) \\
&= \sum_z q^{n-1}(z,y)\pi(z) \\
&\vdots \\
&= \pi(y).
\end{aligned}$$

10.5.3 Beispiel (Ehrenfestsches Urnenmodell)

Behauptung:
$\pi(i) = \binom{N}{i} 2^{-N}$ ist stationäre Verteilung des Ehrenfestschen Urnenmodells.

Beweis:
Zu zeigen: $\sum\limits_i \pi(i) q(i,j) = \pi(j)$.

$$\begin{aligned}
&\sum_i \pi(i) q(i,j) \\
&= \sum_i \binom{N}{i} 2^{-N} q(i,j) \\
&= \binom{N}{j-1} 2^{-N} \frac{N-(j-1)}{N} + \binom{N}{j+1} 2^{-N} \frac{j+1}{N} \\
&= \left[\frac{N!}{(j-1)!(N-(j-1))!} \, \frac{N-(j-1)}{N} + \frac{N!}{(j+1)!(N-(j+1))!} \, \frac{j+1}{N} \right] 2^{-N} \\
&= \left[\frac{(N-1)!}{(j-1)!(N-j)!} \, \frac{j}{j} + \frac{(N-1)!}{j!(N-j-1)!} \, \frac{N-j}{N-j} \right] 2^{-N} \\
&= \frac{(N-1)!j + (N-1)!(N-j)}{j!(N-j)!} 2^{-N} \\
&= \binom{N}{j} 2^{-N} = \pi(j).
\end{aligned}$$

Es gibt noch einen einfacheren Weg die Stationarität von π zu zeigen. Man verwendet dazu das folgende Lemma.

10.5.4 Lemma

Gilt für alle $x, y \in E$ $\pi(x)q(x,y) = \pi(y)q(y,x)$, so ist π stationär.

Beweis:

$$\begin{aligned}\sum_x \pi(x)q(x,y) &= \sum_x \pi(y)q(y,x) \\ &= \pi(y)\sum_x q(y,x) \\ &= \pi(y).\end{aligned}$$

Bemerkung: Für das Ehrenfestsche Urnenmodell lautet die in 10.5.4 vorausgesetzte Gleichung $\binom{N}{i}\frac{N-i}{N} = \binom{N}{i+1}\frac{i+1}{N}$. Es ist offensichtlich, daß diese gilt.

Der folgende Satz liefert die Existenz einer stationären Verteilung.

10.5.5 Satz

Sei q irreduzibel. Dann sind folgende Aussagen äquivalent:

a) Es existiert ein $z \in E$ mit $E(T_z|X_0 = z) < \infty$,

b) Für alle $x, y \in E$ gilt $E(T_y|X_0 = x) < \infty$,

c) Es existiert eine stationäre Wahrscheinlichkeitsfunktion π bezüglich q.

Bemerkung: Bedingung (a) ist stärker als Rekurrenz, denn es gibt Markov-Ketten mit einem rekurrenten Zustand x für den $E(T_x|X_0 = x) = \infty$ gilt. Siehe z.B. die Irrfahrt von oben.

Beweisskizze: Zeige (a) $\Rightarrow$ (c) Setze: $P_z(\cdot) = P(\cdot|X_0 = z)$. Dann ist P_z das Wahrscheinlichkeitsmaß bei Start in z.
Sei $H(x) := \sum\limits_{n=1}^{\infty} P_z(X_n = x, T_z \geq n)$. Dann gilt $H(z) = 1$ und weiter

$$\begin{aligned}\sum_{x\in E} H(x) &= \sum_{n=1}^{\infty}\sum_{x\in E} P_z(X_n = x, T_z \geq n) \\ &= \sum_{n=1}^{\infty} P_z(T_z \geq n) = E_z T_z.\end{aligned}$$

Wir behaupten (+): $H(y) = \sum\limits_{x\in E} H(x)q(x,y)$.
Setzt man nun $\pi(x) = \frac{H(x)}{E_z T_z}$, so ist $\pi(y) = \sum\limits_x \pi(x)q(x,y)$ und $\sum\limits_y \pi(y) = 1$, d.h. π ist stationäre Verteilung.
Zeige Behauptung (+):

$$
\begin{aligned}
H(y) &= \sum_{n=1}^{\infty} P_z(X_n = y, T_z \geq n) \\
&= P_z(X_1 = y) + \sum_{n\geq 2} P_z(X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\
&= q(z,y) + \sum_{n\geq 2} \sum_{x\in E\setminus\{z\}} P_z(X_{n-1} = x, X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\
&= q(z,y) + \sum_{x\in E\setminus\{z\}} \sum_{n\geq 2} P_z(X_{n-1} = x, X_n = y, X_j \neq z \text{ für } 1 \leq j \leq n-1) \\
&= q(z,y) + \sum_{x\in E\setminus\{z\}} \sum_{n\geq 2} P_z(X_{n-1} = x, X_j \neq z \text{ für } 1 \leq j \leq n-1) \cdot P(X_n = y | X_{n-1} = x) \\
&= q(z,y) + \sum_{x\in E\setminus\{z\}} \sum_{l\geq 1} P_z(X_l = x, X_j \neq z \text{ für } 1 \leq j \leq l) \cdot P(X_n = y | X_{n-1} = x) \\
&= q(z,y) + \sum_{x\in E\setminus\{z\}} H(x) q(x,y) \\
&= \sum_{x} H(x) q(x,y).
\end{aligned}
$$

10.5.6 Korollar

Es gilt

$$\pi(x) = \frac{H(x)}{E_z T_z}$$

und insbesondere

$$\pi(z) = \frac{1}{E_z T_z},$$

wobei

$$H(x) = \sum_{n=1}^{\infty} P_z(X_n = x,\ T_z \geq n) \text{ ist für } z \in E.$$

Bemerkung: Beim Ehrenfestschen Urnenmodell sei $\sigma(i) = \min\{n > 0 \mid X_n = i\}$. Dann ist nach dem Korollar: $E_0\sigma(0) = 2^{2k}$ und $E_k\sigma(k) = \binom{2k}{k}^{-1} 2^{2k}$, wobei $k = N/2$ ist. Setzt man $N = 6 \cdot 10^{23}$ und ist ein Zeitschritt 1 Sekunde lang, so ist $E_k\sigma(k) \cong 10^{12}$ Sekunden, aber $E_0\sigma(0) \cong 10^{1,8 \cdot 10^{23}}$ Sekunden.

10.6 Konvergenz gegen die stationäre Verteilung

10.6.1 Definition (Periode, aperiodisch)

Sei q eine irreduzible stochastische Matrix. Weiter sei $N(x,y) := \{n \in \mathbb{N} | q^n(x,y) > 0\}$. Wir definieren die Periode von $x \in E$ durch $d(x) := ggT(N(x,x))$, wobei $ggT(B) := sup\{k \in \mathbb{N} | B \subset k \cdot \mathbb{N}\}$ der größte gemeinsame Teiler der Menge B ist. q heißt aperiodisch, falls $d(x) = 1$ ist für alle $x \in E$. q hat Periode k, falls $d(x) = k$ ist für alle $x \in E$.

10.6.2 Beispiel

Für die stochastische Matrix beim Ehrenfestschen Urnenmodell gilt $d = 2$.

10.6.3 Satz

Sei q irreduzibel und aperiodisch mit stationärer Verteilung π. Dann gilt für alle $x \in E$

$$\lim_{n\to\infty} \sum_{y\in E} |q^n(x,y) - \pi(x)| = 0.$$

Sei q irreduzibel mit Periode k und mit stationärer Verteilung π, so gilt

$$\lim_{n\to\infty} \sum_{y\in E} \left|q^{nk}(x,y) - \pi(x)\right| = 0.$$

Beweis: Zum Beweis siehe Dümbgen, S. 129.

10.6.4 Satz

Sei q irreduzibel und aperiodisch mit stationärer Verteilung π. Sei $f : E \to \mathbb{R}$, so daß $\sum\limits_{x\in E} |f(x)|\pi(x) < \infty$ ist. Dann gilt für jede Startverteilung μ mit P_μ-Wahrscheinlichkeit 1:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} f(x_k) = \sum_{x\in E} f(x)\pi(x).$$

Bemerkung: Dies ist ein grundlegender Satz für Markov-Ketten, der dem Gesetz der Großen Zahlen für unabhängige Beobachtungen entspricht. Es ist ein Basisresultat für die Simulation bei Markov-Ketten, der sogenannten MCMC-Methode.

10.6.5 Bemerkungen zum Ehrenfestschen Urnenmodell

Dieses Modell wurde entwickelt, um eine Streitfrage in der statistischen Physik zu klären. Boltzmann behauptete, ein großes Teilchensystem tendiere sehr schnell zu seinem Gleichgewicht, Zermelo entgegnete, auch dann müßten aus physikalischen Gründen sehr unwahrscheinliche Zustände angenommen werden können. Dies erschien paradox. Das Ehrenfestsche Modell gestattet die Details zu berechnen.

Sei $N = 2k$. Sei $\tau(k) = \inf\{n \geq 0 \mid X_n = k\}$. Dann gilt:

a) $E_0(\tau(k)) = k \ln k + k + O(1)$

b) $E_k(\tau(0)) = \frac{1}{2k} 2^{2k}(1 + o(\frac{1}{k}))$ für $k \to \infty$.

11 Wahrscheinlichkeitsmaße mit Dichten

11.1 Wahrscheinlichkeits-Dichte, Verteilungsfunktion, σ-Algebren

11.1.1 Definition (Wahrscheinlichkeits-Dichte)

Eine Funktion $f : \mathbb{R} \to [0, \infty)$ mit $\int\limits_{-\infty}^{\infty} f(x)dx = 1$ heißt Wahrscheinlichkeits-Dichte auf $\mathbb{R}$.

11.1.2 Beispiele

a) Exponentialverteilung mit Parameter $\lambda > 0$:
$$f_\lambda(x) = \lambda \cdot e^{-\lambda x} \cdot 1_{[0,\infty)}(x), \quad 1_A(x) = \begin{cases} 1 \text{ falls } x \in A \\ 0 \text{ falls } x \notin A \end{cases},$$

b) Gleichverteilung auf $[a, b]$: $f(x) = \frac{1}{b-a} \cdot 1_{[a,b]}(x)$,

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

d) Gamma-Verteilung $G(\alpha, \beta)$: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{[0,\infty)}(x)$.

11.1.3 Definition (Verteilungsfunktion)

Sei f eine Wahrscheinlichkeits-Dichte auf $\mathbb{R}$ und $F(x) := \int\limits_{-\infty}^{x} f(x)dy$.
Dann heißt F Verteilungsfunktion zur Dichte f.

Eigenschaften von F

1) $F(-\infty) = 0, F(\infty) = 1$,

2) F ist monoton wachsend, d.h. $x \leq y \Rightarrow F(x) \leq F(y)$,

3) F ist stetig.

11.1.4 Beispiele

1) $F(x) = (1 - e^{-\lambda x}) = \int\limits_{0}^{x} \lambda e^{-\lambda y} dy \quad$ für $x \geq 0$,

2) $F(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \Phi(x)$,

3) $F(x) = \int\limits_{0}^{x} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy \quad$ für $x \geq 0$.

Bemerkung: f legt F fest (Integrieren von $-\infty$ bis x), F legt f fest (Differenzieren nach der oberen Grenze). F und f stehen also in eineindeutiger Beziehung zueinander. Zu f und F gehört genau ein Wahrscheinlichkeitsmaß $Q : \mathcal{B} \to [0,1]$. (Dabei ist $\mathcal{B}$ die σ-Algebra der Borelschen Mengen. Die Definition folgt weiter unten.) Dies kann auf verschiedene Weisen angegeben werden:

1) $Q((-\infty, x]) = \int\limits_{-\infty}^{x} f(y)dy = F(x)$,

2) $Q(A) = \int\limits_{-\infty}^{\infty} 1_A(x)f(x)dx$, falls A Borelmenge ist.

Eigenschaften von Q

Für A_i disjunkt gilt: Q ist σ-additiv, denn

$$\begin{aligned}
Q\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int 1_{\bigcup\limits_{i=1}^{\infty} A_i}(x)f(x)dy \\
&= \int \sum_{i=1}^{\infty} 1_{A_i}(x)f(x)dx \\
&= \sum_{i=1}^{\infty} \int 1_{A_i}(x)f(x)dx \\
&= \sum_{i=1}^{\infty} Q(A_i).
\end{aligned}$$

Nun zum Definitionsbereich von Wahrscheinlichkeitsmaßen. Potenzmengen sind im Allgemeinen zu groß. Die passenden Mengensysteme sind σ-Algebren.

11.1.5 Definition (σ-Algebra)

Ein System $\mathfrak{A}$ von Teilmengen von Ω heißt σ-Algebra, falls gilt:

a) $\Omega \in \mathfrak{A}$,

b) $A \in \mathfrak{A} \Rightarrow A^C \in \mathfrak{A}$,

c) $A_i \in \mathfrak{A}$ für $i = 1, 2, 3, \ldots \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$.

11.1.6 Folgerungen

Es gilt

1) $\emptyset \in \mathfrak{A}$,

2) $A_i \in \mathfrak{A}$ für $i = 1, 2, 3, \ldots \Rightarrow \bigcap\limits_{i=1}^{\infty} A_i \in \mathfrak{A}$.

Demnach ist eine σ-Algebra ein Mengensystem, das unter abzählbaren Durchschnitten, abzählbaren Vereinigungen und der Komplementbildung abgeschlossen ist.

11.1.7 Beispiele

1) $\mathcal{P}(\Omega)$ ist σ-Algebra.

2) Sei $I \neq \emptyset$ und für jedes $i \in I$ sei $\mathfrak{A}_i$ eine σ-Algebra.
   Dann ist $\mathfrak{A} = \bigcap\limits_{i \in I} \mathfrak{A}_i$ eine σ-Algebra.

3) Sei $\mathcal{S} \subset P(\Omega)$. Dann existiert $\sigma(\mathcal{S}) := \bigcap\limits_{\substack{\mathfrak{A} \text{ ist } \sigma\text{-Algebra}, \\ \mathcal{S} \subset \mathfrak{A}}} \mathfrak{A}$.
   Dies ist die von $\mathcal{S}$ erzeugte σ-Algebra.

Die Beweise für die Folgerungen und die Beispiele sind in einem Buch über Maßtheorie (z.B. Elstrodt) nachzulesen.

11.1.8 Beispiel und Definition (Borelsche σ-Algebra)

Sei $\mathcal{S} = \{[a,b) | a, b \in \mathbb{R}^k\}$.
Dann heißt $\mathcal{B}^k := \sigma(\mathcal{S})$ die Borelsche σ-Algebra auf $\mathbb{R}^k$. Wir schreiben $\mathcal{B}$ für $\mathcal{B}^1$.
$A \in \mathcal{B}$ heißt Borelsche Menge oder kurz borelsch.

11.2 Wahrscheinlichkeitsmaße und Zufallsvariablen auf $\mathbb{R}^k$

11.2.1 Definition (Wahrscheinlichkeitsdichte / Wahrscheinlichkeitsmaß auf $\mathbb{R}^k$)

Eine Funktion $f : \mathbb{R}^k \to [0, \infty)$ mit $\int\limits_{\mathbb{R}^k} f(x)dx = 1$ heißt Wahrscheinlichkeitsdichte auf $\mathbb{R}^k$.
Das zugehörige Wahrscheinlichkeitsmaß Q ist erklärt als $Q(A) = \int\limits_{\mathbb{R}^k} 1_A(x) f(x) dx, \quad A \in \mathcal{B}^k$.

11.2.2 Definition (Wahrscheinlichkeitsraum)

Ein Wahrscheinlichkeitsraum ist ein Tripel $(\Omega, \mathfrak{A}, P)$, bestehend aus einer nicht leeren Menge Ω, einer σ-Algebra $\mathfrak{A}$ und einem Wahrscheinlichkeitsmaß P auf $(\Omega, \mathfrak{A})$.
Dabei heißt P Wahrscheinlichkeitsmaß auf $(\Omega, \mathfrak{A})$, falls $P : \mathfrak{A} \to [0,1]$ mit P σ-additiv und $P(\Omega) = 1$ ist.
P heißt σ-additiv auf $\mathfrak{A}$, falls $P(\bigcup_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$ für $A_i \in \mathfrak{A}$, $i = 1, 2, ...$, mit A_i paarweise disjunkt.

11.2.3 Definition (Zufallsvariable)

$(\Omega, \mathfrak{A}, P)$ sei ein Wahrscheinlichkeitsraum.
Eine Funktion $X : \Omega \to \mathbb{R}$ heißt Zufallsvariable, falls $X^{-1}(B) \in \mathfrak{A}$ für alle $B \in \mathcal{B}$ ist.

Bemerkung zum Sinn dieser Definition

1) Man will $P(\{\omega | X(\omega) \in B\}) = P(\{X \in B\}) = P(X^{-1}(B))$ wohldefiniert haben für alle $B \in \mathcal{B}$.
2) Damit X Zufallsvariable ist genügt es für alle $\alpha \in \mathbb{R}$ zu fordern: $\{X \leq \alpha\} = X^{-1}((-\infty, \alpha]) \in \mathfrak{A}$.

11.2.4 Definition (Wahrscheinlichkeitsverteilung einer Zufallsvariable)

Sei X eine Zufallsvariable. Die Verteilung von X ist definiert durch $Q(B) := P(X^{-1}(B))$ für alle $B \in \mathcal{B}$. Damit ist Q ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$.

Bemerkung: Zur Dichte f existiert stets eine Zufallsvariable X so, dass die Verteilung von X die Dichte f hat. Wähle $X(z) = z$ für $z \in \mathbb{R}$.

11.3 Erwartungswert und Varianz

11.3.1 Definition (Erwartungswert)

Sei X Zufallsvariable und sei f Wahrscheinlichkeits-Dichte und $g : \mathbb{R} \to \mathbb{R}$ messbar (d.h. $g^{-1}(B) \in \mathcal{B}$ für $B \in \mathcal{B}$). Dann heißt

$$Eg(X) := \int_{\mathbb{R}} g(x) f(x) dx$$

der Erwartungswert von $g(X)$.

Bemerkung: Der Erwartungswert einer Zufallsvariable X mit Dichte f ist

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad \text{und} \quad E(X^p) = \int_{-\infty}^{\infty} x^p f(x) dx$$

ist das p-te Moment. Diese Erwartungswerte existieren genau dann, wenn die uneigentlichen Integrale existieren. Existiert $E(X^2)$, so gilt:

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = E((X - E(X))^2) = E(X^2) - (EX)^2.$$

Im Übrigen gelten die gleichen Gesetze wie im diskreten Fall.

11.3.2 Beispiele

Ist X normalverteilt nach $N(\mu, \sigma^2)$, d.h. $F_X(\alpha) = P(X \leq \alpha) = \int\limits_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx$, so gilt $E(X) = \mu$ und $\mathrm{Var}(X) = \sigma^2$, denn

a)
$$\begin{aligned}
E(X) &= \int\limits_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \int\limits_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx + \mu \\
&= \int\limits_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} dy + \mu, \quad \text{wenn } y = x - \mu \text{ gesetzt wird,} \\
&= \mu.
\end{aligned}$$

Das Integral ist Null, da über eine antisymmetrische Funktion integriert wird.

b)
$$\begin{aligned}
\mathrm{Var}(X) &= \int\limits_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \sigma^2 \int\limits_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \quad \text{wenn } z = \frac{x-\mu}{\sigma} \text{ gesetzt wird,} \\
&= \sigma^2 \left[\frac{-z}{\sqrt{2\pi}} e^{-z^2/2} \Bigg|_{-\infty}^{\infty} + \int\limits_{-\infty}^{\infty} e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} \right] \\
&= \sigma^2.
\end{aligned}$$

Hier wurde partielle Integration angewandt.

11.3.3 Weiteres Beispiel

Was ist die Dichte von X^2, wenn X nach $N(0,1)$ verteilt ist? Es gilt

$$P(X^2 < \beta) = P(|X| < \sqrt{\beta}) = \int\limits_{-\sqrt{\beta}}^{\sqrt{\beta}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\beta} y^{-\frac{1}{2}} e^{-y/2} dy.$$

Hierbei wurde $z^2 = y$ gesetzt.

11.4 Unabhängigkeit von Zufallsvariablen

11.4.1 Definition (Unabhängigkeit von Zufallsvariablen)

Die Zufallsvariablen $X_1, ..., X_k$ heißen unabhängig, falls gilt:

$$P(X_1 \le \alpha_1, X_2 \le \alpha_2, ..., X_k \le \alpha_k) = \prod_{i=1}^{k} P(X_i \le \alpha_i)$$

für $\alpha_i \in \mathbb{R}, i = 1, ..., k$.

11.4.2 Folgerung

Sind $X_1, ..., X_k$ unabhängig und hat die Verteilung von X_i die Dichte f_i, so hat die gemeinsame Verteilung von $(X_1, \dots, X_k)$ die Wahrscheinlichkeitsdichte $f = \prod\limits_{i=1}^{k} f_i$.

Beweis:
Es gilt

$$\begin{aligned}
P(X_1 \leq \alpha_1, X_2 \leq \alpha_2, ..., X_k \leq \alpha_k) &= \prod_{i=1}^{k} P(X_i \leq \alpha_i) \\
&= \prod_{i=1}^{k} \int_{-\infty}^{\alpha_i} f_i(x_i) dx_i \\
&= \int_{-\infty}^{\alpha_1} ... \int_{-\infty}^{\alpha_k} \prod_{i=1}^{k} f_i(x_i) dx_1 \dots dx_n \\
&= \int_{-\infty}^{\alpha_1} ... \int_{-\infty}^{\alpha_k} f(x) dx.
\end{aligned}$$

11.4.3 Folgerung

Sind X und Y unabhängig mit Dichten f und g. Dann hat die Verteilung von $Z = X + Y$ die Dichte $(f * g)(t) := \int\limits_{-\infty}^{\infty} f(t - s) g(s) ds$. $f * g$ heißt Faltung von f und g.

Beweis:
Es gilt

$$\begin{aligned}
P(Z \leq \alpha) &= P(X + Y \leq \alpha) \\
&= \int\limits_{\{(x,y) | x+y \leq \alpha\}} f(x) g(y) dx dy \\
&= \int\limits_{\{(x,y) | t \leq \alpha\}} f(t - s) g(s) ds dt \\
&= \int_{-\infty}^{\alpha} (\int_{-\infty}^{\infty} f(t - s) g(s) ds) dt
\end{aligned}$$

mit $t = x + y$ und $s = y$. Damit ist $\int\limits_{-\infty}^{\infty} f(t - s) g(s) ds$ die Dichte von $Z = X + Y$.

11.4.4 Beispiel

Die Exponentialverteilung hat die Dichte $f(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$.
Für $n \geq 1$ gilt: $(f^*)^n(z) = \frac{\lambda^n z^{n-1}}{(n-1)!} e^{-\lambda z} 1_{[0.\infty)}(z)$.
Dabei ist $(f^*)^n := f * f * f * f * ... * f$ das n-fache Faltungsprodukt.

Beweis durch Induktion:
Induktionsanfang: Richtig für $n = 1$.
Induktionsvorraussetzung: $(f^*)^n(z) = \frac{\lambda^n z^{n-1}}{(n-1)!} e^{-\lambda z} 1_{[0,\infty)}(z)$.
Induktionsschritt: $(n \to n+1)$

$$\begin{aligned}
(f^*)^{n+1}(z) &= f * (f^*)^n(z) \\
&= \int\limits_{-\infty}^{\infty} f(z-x)(f^*)^n(x)dx \\
&= \int\limits_{-\infty}^{\infty} \lambda e^{-\lambda(z-x)} 1_{[0,\infty)}(z-x) \; \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} 1_{[0,\infty)}(x)dx \\
&= \lambda^{n+1} e^{-\lambda z} \int\limits_{-\infty}^{\infty} \frac{x^{n-1}}{(n-1)!} 1_{(-\infty,z]}(x) 1_{[0,\infty)}(x)dx \\
&= \lambda^{n+1} e^{-\lambda z} \int\limits_{0}^{z} \frac{x^{n-1}}{(n-1)!} dx \quad \text{für } z \geq 0 \text{ und } \; = 0 \text{ für } z < 0 \\
&= \lambda^{n+1} \frac{z^n}{n!} e^{-\lambda z} 1_{[0,\infty)}(z).
\end{aligned}$$

Bemerkung: Die Gammaverteilung $G(\alpha, \beta)$ mit den Parametern(α, β) hat die Dichte $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{[0,\infty)}(x)$. Folglich ist $(f^*)^n$ die Dichte einer $G(n, \lambda)$-Verteilung.

11.4.5 Beispiel: Herleitung des Poisson-Prozesses über exponential verteilte Zufallsvariablen

X_i sei exponentialverteilt mit Parameter λ. X_i sei die Wartezeit zwischen dem $(i-1)$-ten und dem i-ten Anruf. $S_n = \sum\limits_{i=1}^{n} X_i$ sei die Gesamtwartezeit bis zum n-ten Anruf, mit $S_0 = 0$. Die Anzahl der Anrufe bis zur Zeit t ist $N_t = \max\{k \geq 0 | S_k \leq t\}$. Was ist die Verteilung von N_t? Es gilt

$$\begin{aligned}
P(N_t = n) &= P(S_n \leq t, S_{n+1} > t) \\
&= P(S_{n+1} > t) - P(S_n > t) \\
&= \int\limits_{t}^{\infty} \lambda^{n+1} \frac{s^n}{n!} e^{-\lambda s} ds - \int\limits_{t}^{\infty} \lambda^n \frac{s^{n-1}}{(n-1)!} e^{-\lambda s} ds \\
&= -\lambda^n \frac{s^n}{n!} e^{-\lambda s} \Big|_t^{\infty}
\end{aligned}$$

$$= \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Folglich ist $P(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$.

11.5 Momenterzeugende Funktionen

In Analogie zu Kapitel 9 definiert man für Verteilungen mit Dichten die sogenannte momenterzeugenden Funktionen (MEF). Ist X Zufallsvariable mit Dichte f, so ist $M_X(t) = Ee^{tX} = \int e^{tx} f(x)dx$ die MEF der Verteilung von X.
Für die MEFs gelten entsprechende Ausssagen wie für die erzeugende Funktionen von Kaptitel 9:

a) Eindeutigkeit,

b) Faltungseigenschaft ($M_{X+Y}(t) = M_X(t)M_Y(t)$) bei Unabhängigkeit,

c) Berechnung von Momenten.

11.5.1 Beispiel: Die MEF von $N(\mu, \sigma^2)$

Es gilt

$$M(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2/2\sigma^2} dx = e^{t\mu + \frac{1}{2}\sigma^2 t^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-(\mu+\sigma^2 t))^2/2\sigma^2} dx = e^{t\mu+\frac{1}{2}\sigma^2 t^2}.$$

11.5.2 Beispiel zur Faltung

Sei X nach $N(\mu_1, \sigma_1^2)$ und Y nach $N(\mu_2, \sigma_2^2)$ verteilt. Außerdem seien X und Y unabhängig. Dann gilt für die MEF von $X + Y$:

$$M_{X+Y}(t) = Ee^{tX} Ee^{tY} = e^{t\mu_1 + \frac{1}{2}\sigma_1^2 t^2} e^{t\mu_2 + \frac{1}{2}\sigma_2^2 t^2} = e^{t(\mu_1+\mu_2)+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}.$$

Dies ist die MEF von $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, d.h. $X + Y$ ist nach $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ verteilt.

11.5.3 Die MEF der Gammaverteilung

Es gilt $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ und damit

$$\begin{aligned} M(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{tx} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(\beta-t)x} x^{\alpha-1} dx \end{aligned}$$

$$= \frac{\beta^\alpha}{(\beta - t)^\alpha}$$
$$= (1 - \frac{t}{\beta})^{-\alpha}.$$

Weiterhin gilt $E(X) = M'(0) = \frac{\alpha}{\beta}$, $\mathrm{Var}(X) = M''(0) - [M'(0)]^2 = \frac{\alpha}{\beta^2}$.

11.6 χ^2_k-Verteilung

11.6.1 Definition (χ^2_k-verteilt)

Seien $Y_1, ..., Y_k$ unabhängige Zufallsvariablen und alle nach $N(0,1)$-verteilt. Dann ist $V := Y_1^2 + Y_2^2 + ... + Y_k^2$ nach χ^2_k-verteilt, d.h. Chiquadrat-verteilt mit k Freiheitsgraden.

Berechnung der Dichte von V

Was ist $P\left(\sum\limits_{i=1}^k Y_i^2 \leq \beta\right)$?
Es gilt: $P(Y_i \leq \alpha) = \int\limits_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} dx_i$.
Wir definieren $K(\beta) := \{(y_1, ..., y_k) | \sum\limits_{i=1}^k y_i^2 \leq \beta\}$. Damit ist:

$$\begin{aligned} P((Y_1, ..., Y_k) \in K(\beta)) &\overset{(*)}{=} \int\limits_{K(\beta)} \prod_{i=1}^k \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y_i^2}{2}}\right) dy_1...dy_k \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} \int\limits_{K(\beta)} e^{-\frac{1}{2}\sum\limits_{i=1}^k y_i^2} dy_1...dy_k \\ &\overset{(**)}{=} B_1 \cdot \int\limits_{S^{k-1}} d\sigma \int\limits_0^{\sqrt{\beta}} e^{-\frac{r^2}{2}} r^{k-1} dr \\ &\overset{(***)}{=} B_2 \int\limits_0^{\beta} e^{-\frac{z}{2}} z^{\frac{k}{2}-1} dz. \end{aligned}$$

(*) wegen Unabhängigkeit,
(**) setze $r^2 = \sum\limits_{i=1}^k y_i^2$,
(***) $z = r^2 \Rightarrow dz = 2rdr$. Dabei ist $B_2 = \frac{2^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})}$.

Zusammenfassend: $P(V \leq \beta) = \dfrac{2^{-\frac{k}{2}}}{\Gamma(\frac{k}{2})} \int\limits_0^{\beta} e^{-\frac{z}{2}} z^{\frac{k}{2}-1} dz.$

Dies ist die Verteilungsfunktion einer χ^2_k-Verteilung. Insbesondere ist diese eine $G(\frac{k}{2}, \frac{1}{2})$-Verteilung. Ihre MEF ist daher $M(t) = (1 - 2t)^{-k/2}$.

11.6.2 Empirischer Mittelwert und empirische Varianzen bei der Normalverteilung

Seien $X_1, ..., X_n$ unabhängig nach $N(\mu, \sigma^2)$ verteilt und $T_n := \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \mu)^2$.
Dann ist $ET_n = \sigma^2$ und $\frac{n}{\sigma^2}T_n$ ist χ_n^2-verteilt.
Denn: $ET_n = E\left(\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \mu)^2\right) = \frac{1}{n}\sum\limits_{i=1}^{n}E((X_i - \mu)^2) = \frac{1}{n}n\text{Var}(X_1) = \sigma^2$
und $\frac{n}{\sigma^2}T_n = \frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i - \mu)^2 = \sum\limits_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2$.

Seien $\overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n}X_i$ und $S_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$. Dann gilt: $E\overline{X}_n = \mu$ und $ES_n^2 = \sigma^2$.

Denn: $E\overline{X}_n = E\left(\frac{1}{n}\sum\limits_{i=1}^{n}X_i\right) = \frac{1}{n}\sum\limits_{i=1}^{n}E(X_i) = \frac{1}{n}nE(X_1) = \mu$
und $ES_n^2 = E\left(\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2\right) = \frac{1}{n-1}\sum\limits_{i=1}^{n}E[((X_i - \mu) - (\overline{X}_n - \mu))^2]$
$= \frac{1}{n-1}\sum\limits_{i=1}^{n}(\sigma^2 - \frac{2}{n}\sum\limits_{j=1}^{n}E[(X_i - \mu)(X_j - \mu)] + \frac{\sigma^2}{n}) = \frac{1}{n-1}n\sigma^2\frac{n-1}{n} = \sigma^2$.

11.6.3 Wie ist $S_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$ verteilt?

Um die Verteilung von S_n^2 zu berechnen brauchen wir folgende Lemmata:

Lemma 1

Die Zufallsvariable $\overline{X}_n$ ist unabhängig von $(X_1 - \overline{X}_n, ..., X_n - \overline{X}_n)$.

Lemma 2

$\overline{X}_n$ und S_n^2 sind unabhängig.

Beweis:
S_n^2 ist eine Funktion von $(X_1 - \overline{X}_n, ..., X_n - \overline{X}_n)$ und folglich unabhängig von $\overline{X}_n$.

Lemma 3

$\frac{n}{\sigma^2}(\overline{X}_n - \mu)^2$ ist χ_1^2-verteilt.

11.6.4 Satz

$\frac{n-1}{\sigma^2}S_n^2 = \frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i - \overline{X}_n)^2$ ist χ^2_{n-1}-verteilt.

Beweis:
Nach Bemerkung (11.6.2) gilt: $W := \frac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2 = \frac{n}{\sigma^2}T_n$ ist χ^2_n-verteilt. Weiter gilt

$$\begin{aligned} W &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(X_i-\overline{X}_n)+(\overline{X}_n-\mu)\right]^2 \\ &= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\overline{X}_n)^2 + 2\sum_{i=1}^{n}(X_i-\overline{X}_n)(\overline{X}_n-\mu) + n(\overline{X}_n-\mu)^2\right] \\ &= \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2}_{U} + \underbrace{\frac{n}{\sigma^2}(\overline{X}_n-\mu)^2}_{V}, \end{aligned}$$

denn $\sum\limits_{i=1}^{n}(X_i-\overline{X}_n) = 0$ und damit der gemeinsame Term gleich Null.

Die linke Seite ist χ^2_n-verteilt.
Nach Lemma 1 ist $Ee^{tW} = Ee^{t(U+V)} = Ee^{tU}Ee^{tV}$.
Dann folgt mit Lemma 3: $Ee^{tU} = Ee^{tW}/Ee^{tV} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-\frac{n-1}{2}}$.
Damit ist U χ^2_{n-1}-verteilt.

11.7 t_n-Verteilung

11.7.1 Definition (t_n-Verteilung)

Ist X nach $N(0,1)$-verteilt und V nach χ^2_n-verteilt und sind X und V unabhängig, so heißt die Verteilung von $\frac{X}{\sqrt{V/n}}$ t-Verteilung mit n Freiheitsgraden (kurz: t_n-Verteilung).

11.7.2 Korollar

Sind $\overline{X}_n$ und S_n^2 wie oben, so ist $\frac{\overline{X}_n-\mu}{S_n/\sqrt{n}}$ t_{n-1}-verteilt.

Beweis:
Es ist

$$\frac{\overline{X}_n-\mu}{S_n/\sqrt{n}} = \frac{\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}}{\sqrt{S_n^2/\sigma^2}} = \frac{\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X}_n)^2/\sigma^2}}.$$

$X = \sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}$ ist $N(0,1)$-verteilt, der Term unter der Wurzel ist χ^2_{n-1}-verteilt.

11.7.3 Satz

Die t-Verteilung mit n Freiheitsgraden hat die Dichte

$$f_n(z) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\frac{z^2}{n})^{-\frac{n+1}{2}}.$$

Beweis:
Sei $U = X\sqrt{n}$. Dann ist U nach $N(0,n)$ verteilt und V χ_n^2-verteilt. Weiter gilt wegen der Unabhängigkeit von U und V

$$P(U/\sqrt{V} \leq \alpha) = \int\limits_{\{(u,v)|u/\sqrt{v}\leq\alpha\}} f(u)g(v)dudv = \int\limits_0^\alpha \underbrace{\left(\int f(z\sqrt{v})g(v)\sqrt{v}dv\right)}_{h(z)} dz$$

mit $z := \frac{u}{\sqrt{v}}$ und folglich

$$\begin{aligned}
h(z) &= \frac{1}{\sqrt{2\pi n}}\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\int\limits_0^\infty e^{-\frac{vz^2}{2n}}\, v^{\frac{n}{2}-1}\, e^{-v/2}\sqrt{v}\, dv \\
&= \frac{1}{\sqrt{2\pi n}2^{n/2}\Gamma(\frac{n}{2})}\int\limits_0^\infty v^{\frac{n+1}{2}-1}e^{-\frac{v}{2}(1+\frac{z^2}{n})}dv \\
&= \frac{2^{\frac{n+1}{2}}(1+\frac{z^2}{n})^{-\frac{n+1}{2}}\,\Gamma(\frac{n+1}{2})}{\sqrt{2\pi n}2^{n/2}\Gamma(\frac{n}{2})} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}}.
\end{aligned}$$

Bemerkung: Sei F_n die Verteilungsfunktion der t-Verteilung. Dann gilt: $F_n(z) \to \Phi(z)$ für $n \to \infty$ und $f_n(z) \to \varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ für $n \to \infty$, denn $(1+\frac{z^2}{n})^{-\frac{n+1}{2}} \to e^{-z^2/2}$, da $(1+\frac{x}{n})^n \to e^x$, falls $n \to \infty$.

12 Schließende Statistik

12.1 Einführendes Beispiel: Füllt die Brauerei zu wenig ab?

In einer Brauerei wird eine Maschine eingesetzt um Flaschen mit 0,5l Inhalt abzufüllen. Die Maschine kann in der Abfüllmenge variiert werden. Für die abgefüllte Menge in 8 verschiedenen Flaschen werden die folgenden Werte gemessen (in Litern):

$x_1 = 0{,}489$	$x_5 = 0{,}496$
$x_2 = 0{,}502$	$x_6 = 0{,}482$
$x_3 = 0{,}508$	$x_7 = 0{,}503$
$x_4 = 0{,}497$	$x_8 = 0{,}483$

Tabelle 3: Messwerte Flaschenabfüllung

Annahme: Die Daten $x_1, ..., x_8$ sind Realisierungen von Zufallsvariablen $X_1, ..., X_8$, wobei die X_i unabhängig und nach $N(\mu, \sigma^2)$ verteilt sind. μ und σ werden als unbekannt angenommen.

Es stellen sich folgende Aufgaben bzw. Fragen:

1) Schätze μ.
2) Schätze σ^2.
3) Wie stark schwankt der Schätzer von μ?
4) Wie stark schwankt der Schätzer von σ^2?
5) Füllt die Brauerei zu wenig ab?

Antworten:

1) Wir wählen den Mittelwert als Schätzer für μ:
$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n,$
$\hat{\mu}_8 = \frac{1}{8}(0{,}489 + 0{,}502 + ... + 0{,}483) = \frac{1}{8} \cdot 3{,}96 = 0{,}495 =: \overline{x}_8.$

2) Ein Schätzer für die Varianz σ^2:
$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2,$
$\hat{\sigma}_8^2 = \frac{1}{7}\sum_{i=1}^{8}(x_i - \overline{x}_8)^2 = [0{,}006^2 + ... + 0{,}008^2 + 0{,}012]^2 = 9,086 \cdot 10^{-5}.$
Damit ergibt sich als Schätzung für die Standardabweichung:
$\hat{\sigma}_8 = \sqrt{\hat{\sigma}_8^2} = 9{,}532 \cdot 10^{-3} \approx 0{,}01.$

Die Fragen 3)–5) beantworten wir weiter unten.

12.2 Punktschätzer

Gegeben sei ein statistisches Modell $(\Omega, \mathfrak{A}, P_\theta, \theta \in \Theta)$. Dabei ist Ω die Beobachtungsmenge, $\mathfrak{A}$ die σ-Algebra über Ω und $\{P_\theta; \theta \in \Theta\}$ eine durch $\Theta \subset \mathbb{R}^k$ indizierte Schar von Wahrscheinlichkeitsmaßen.

12.2.1 Definition (Schätzer)

Ein Vektor $T = (T_1, \ldots, T_k)$, bestehend aus Zufallsvariablen, heißt Schätzer, falls $T(\Omega) \subset \Theta$ gilt.

12.2.2 Definition (Maximum-Likelihood-Schätzer)

Sei $\{f_\theta(\omega) | \theta \in \Theta\}$ eine durch θ indizierte Schar von Wahrscheinlichkeitsfunktionen oder Wahrscheinlichkeitsdichten auf Ω. Dann heißt $\hat{\theta}(\omega) := \arg \max\limits_{\theta \in \Theta} f_\theta(\omega)$ Maximum-Likelihood-Schätzer von θ. Es gilt $f_{\hat{\theta}(\omega)}(\omega) = \max\limits_{\theta} f_\theta(\omega)$.
Somit stellt sich die Aufgabe: Maximiere für das beobachtete ω die Funktion $\theta \mapsto f_\theta(\omega)$.

12.2.3 Beispiele:

1) Binomialverteilung:
$f_\theta(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \quad 0 \le k \le n, \quad 0 \le \theta \le 1.$
Hier ist der Maximum-Likelihood-Schätzer: $\hat{\theta}(k) = \frac{k}{n}$. Denn:

$$\begin{aligned}
&\frac{\partial}{\partial \theta} f_\theta(k) \stackrel{!}{=} 0 \\
\Leftrightarrow\ &\binom{n}{k} [k\theta^{k-1}(1-\theta)^{n-k} - (n-k)(1-\theta)^{n-k-1}\theta^k] \stackrel{!}{=} 0 \\
\Rightarrow\ &\binom{n}{k} \theta^{k-1}(1-\theta)^{n-k-1}(k(1-\theta) - (n-k)\theta) = 0 \\
\Rightarrow\ &k - k\theta - n\theta + k\theta = 0 \\
\Rightarrow\ &k = n\theta \\
\Rightarrow\ &\theta = \frac{k}{n} \\
\Rightarrow\ &\hat{\theta} = \frac{k}{n}.
\end{aligned}$$

2) Poisson-Verteilung
$f_\theta(k) = \frac{\theta^k}{k!} e^{-\theta}, \quad \theta \in (0,\infty), k \in \mathbb{N} \cup \{0\}.$
Hier ist $\hat{\theta} = k$, da

$$\begin{aligned}
0 \stackrel{!}{=} \frac{\partial}{\partial \theta} f_\theta(k) &= \frac{k\theta^{k-1}}{k!} e^{-\theta} - \frac{\theta^k}{k!} e^{-\theta} \\
\Rightarrow 0 \stackrel{!}{=} k - \theta &\Rightarrow \hat{\theta} = k.
\end{aligned}$$

3) Exponentialverteilung

$f_\theta(x) = \theta e^{-\theta x} 1_{[0,\infty)}(x), \quad \theta \in (0,\infty)$

Man beobachtet in unabhängigen Experimenten die Messwerte $x_1, \ldots, x_n$:

$f_\theta(x_1, ..., x_n) = \prod\limits_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i}$

$0 \stackrel{!}{=} \frac{\partial}{\partial\theta} f_\theta(x_1, ..., x_n) = n\theta^{n-1} e^{-\theta \sum x_i} - \theta^n (\sum\limits_{i=1}^{n} x_i) e^{-\theta \sum x_i}$

$\Rightarrow n - \theta \sum\limits_{i=1}^{n} x_i \stackrel{!}{=} 0 \quad \Rightarrow \hat{\theta}_n = \frac{n}{\sum\limits_{i=1}^{n} x_i}.$

Nebenbei: $\sum\limits_{i=1}^{n} x_i/n$ ist auch Maximum-Likelihood-Schätzer von θ^{-1}.

4) Normalverteilung

Man beobachtet in unabhängigen Experimenten die Messwerte $x_1, \ldots, x_n$:

$\theta = (\mu, \sigma^2), \mu \in \mathbb{R}, \sigma \in (0,\infty)$.

$f_\theta(x_1, ..., x_n) = \prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}$

Weil die logarithmische Funktion streng monoton wachsend ist, hat eine Funktion genau dann ein Maximum, wenn ihr Logarithmus ein Maximum hat. Also erhalten wir dasselbe Ergebnis, wenn wir die Funktion $l_\theta(x_1, ..., x_n) := \log f_\theta(x_1, ..., x_n)$ maximieren:

$$\begin{aligned} l_\theta(x_1, ..., x_n) &= \log f_\theta(x_1, ..., x_n) \\ &= -n \log\sigma - n \log(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

a) $\frac{\partial l_\theta}{\partial\mu} \stackrel{!}{=} 0 \Rightarrow \frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} 2(x_i - \mu) = 0 \Rightarrow \hat{\mu}_n = \frac{\sum\limits_{i=1}^{n} x_i}{n}.$

b) $\frac{\partial l_\theta}{\partial\sigma} \stackrel{!}{=} 0 \Rightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum\limits_{i=1}^{n} (x_i - \mu)^2 = 0 \Rightarrow \hat{\sigma}_n^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \mu)^2.$

Setze nun $\hat{\mu}_n$ ein, so erhält man $\hat{\sigma}_n^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2$.

5) Hypergeometrische Verteilung

$f_\theta(k) = \frac{\binom{\theta}{k}\binom{N-\theta}{n-k}}{\binom{N}{k}}, \quad 0 \le \theta \le N$

Eine kurze Rechnung ergibt: $\hat{\theta}(k) = [\frac{N \cdot k}{n}]$, wobei $[\cdot]$ die Gauss-Klammer ist.

Das Ergebnis ist auch intuitiv plausibel. „Stochastischer Dreisatz“: $\frac{k}{n} \sim \frac{\theta}{N} \Rightarrow \theta \sim \frac{N \cdot k}{n}$.

12.2.4 Definition (stochastische Konvergenz)

Sei P ein Wahrscheinlichkeitsmaß und für jedes $n \in \mathbb{N}$ sei X_n eine Zufallvariable. Die Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert P-stochastisch gegen die Zufallsvariable X, in Formeln $X_n \xrightarrow{P} X$, wenn für alle $\varepsilon > 0$ gilt: $\lim\limits_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$.

Bemerkung: In den oben aufgeführten Beispielen 1), 3) und 4) kann man mit Hilfe des Gesetzes der großen Zahlen zeigen, dass für alle θ_0 gilt: $\hat{\theta}_n \stackrel{P_{\theta_0}}{\rightarrow} \theta_0$.
Diese Konvergenzeigenschaft eines Schätzers heißt Konsistenz!

12.2.5 Satz (Konsistenz des M-L-Schätzers)

Für unabhängige Beobachtungen gilt unter geeigneten Glattheitsannahmen an $f_\theta(x)$, daß $\hat{\theta}_n \stackrel{P_{\theta_0}}{\rightarrow} \theta_0$.

Beweisskizze:
Wegen dem Gesetz der großen Zahlen gilt:
$\frac{1}{n} l_\theta(x_1, ..., x_n) = \frac{1}{n} \sum\limits_{i=1}^{n} \log f_\theta(x_i) \rightarrow E_{\theta_0} \log f_\theta(x_1)$.
$\theta \mapsto E_{\theta_0} \log f_\theta(X_1)$ wird maximal in $\theta = \theta_0$, da wegen der Jensen-Ungleichung $E_{\theta_0} \log \dfrac{f_\theta(X_1)}{f_{\theta_0}(X_1)} \leq \log E_{\theta_0} \dfrac{f_\theta(X_1)}{f_{\theta_0}(X_1)} = 0$ ist. Man beachte, $E_{\theta_0} \dfrac{f_\theta(X_1)}{f_{\theta_0}(X_1)} = \displaystyle\int f_\theta(x) dx = 1$.
Daraus folgt $E_{\theta_0} \log f_\theta(X_1) \leq E_{\theta_0} \log f_{\theta_0}(X_1)$. Wegen Stetigkeit folgt
$\arg\max\limits_\theta \frac{1}{n} l_\theta \rightarrow \arg\max\limits_\theta E_{\theta_0} \log f_\theta = \theta_0$.

12.3 Bewertung von Schätzern: Die Risikofunktion bei Bernoulli-Beobachtungen

Seien $X_1, ..., X_n$ unabhängige Bernoulli-verteilte Zufallsvariablen, die alle dieselbe unbekannte Erfolgswahrscheinlichkeit θ haben. $T(X_1, ..., X_n)$ sei ein Schätzer von θ.
Als ein Maß für die Abweichung von T definieren wir die Risikofunktion (auch mittlerer quadratischen Fehler genannt) durch: $R(\theta, T) := E_\theta((T(X_1, ..., X_n) - \theta)^2)$.

Aufgabe: Suche einen Schätzer mit möglichst kleiner Risikofunktion.

12.3.1 Beispiele:

1) $T_1(x) = \sum\limits_{i=1}^{n} \frac{x_i}{n}$ 2) $T_2(x) = \theta_0 \quad 0 < \theta_0 < 1$ 3) $T_3(x) = \frac{\sum\limits_{i=1}^{n} x_i + \sqrt{n}/2}{n + \sqrt{n}}$

Dann gilt:

1) $R(\theta, T_1) = \frac{\theta(1-\theta)}{n}$ 2) $R(\theta, T_2) = (\theta_0 - \theta)^2$ 3) $R(\theta, T_3) = \frac{1}{4(\sqrt{n}+1)^2}$

12.3.2 Folgerung

Ein Schätzer T^*, der besser als alle Schätzer wäre, müsste eine Risikofunktion haben mit $R(\theta, T^*) \equiv 0$, weil alle Schätzer vom Typ T_2 minorisiert werden müßten.
Tatsächlich ist T_3 der Schätzer mit der kleinsten konstanten Risikofunktion (Siehe Dinges-Rost: Minimax-Schätzer).

Eine schwächere Eigenschaft ist folgende:

12.3.3 Definition (zulässig)

Ein Schätzer T' heißt zulässig, falls für jeden Schätzer T mit $R(\theta, T) \leq R(\theta, T')$ für alle $\theta \in \Theta$ gilt: $R(\theta, T) = R(\theta, T')$ für alle θ.

Bemerkung: Sowohl T_1 als auch T_3 sind zulässig.

12.3.4 Satz (Die Cramer–Rao-Ungleichung)

Seien $X_1, \ldots, X_n$ unabhängig identisch verteilte Zufallsvariablen mit Ws-Dichte oder Ws-Funktion $f_\theta(x)$.
Sei $T_n(X_1, ..., X_n)$ ein Schätzer von θ mit Werten in $\Theta \subset \mathbb{R}$, sei $E_\theta T_n$ differenzierbar in θ und gelte $\frac{\partial}{\partial\theta} E_\theta T_n = \int T_n(x) \frac{\partial}{\partial\theta} f_\theta^n(x) dx$ mit $f_\theta^n(x) = \prod_{i=1}^{n} f_\theta(x_n)$. Sei $b_n(\theta) := E_\theta T_n - \theta$ und $I(\theta) = E_\theta((\frac{\partial}{\partial\xi} \log f_\xi(X_1)|_{\xi=\theta})^2)$. Dann gilt:

$$E_\theta(T_n - \theta)^2 \geq \frac{(1 + b_n'(\theta))^2}{nI(\theta)}.$$

Beweis: Siehe Krengel, Paragraph 4.5: Die Informationsungleichung.

Bemerkung: Die Cramer–Rao-Ungleichung gibt eine Abschätzung nach unten für die Risikofunktion $R(\theta, T)$. $I(\theta)$ heißt Fischer-Information (nach Sir Ronald Fischer).

Für die Beispiele (1)–(4) des ML-Schätzers gilt, dass die untere Schranke der Cramer–Rao-Ungleichung angenommen wird. Dabei ist stets $E_\theta T_n = \theta$.
Was ist $I(\theta)$?

1) Binomialverteilung: $I(\theta) = \frac{n}{\theta(1-\theta)}$

2) Poisson-Verteilung: $I(\theta) = \frac{1}{\theta}$

3) Exponentialverteilung: $I(\theta) = \frac{1}{\theta^2}$

4) Normalverteilung (μ unbekannt): $I(\theta) = \frac{1}{\sigma^2}$

12.4 Konfidenzintervalle

Hier wird nun behandelt, wie stark die Schätzer schwanken.

Wir beginnen mit einem Beispiel zu Aussagen mit Irrtumswahrscheinlichkeiten.

12.4.1 Beispiel: Fischteichgröße

Aus einem Fischteich werden 50 Fische gefangen, markiert und wieder ausgesetzt. Bei einem erneuten Fang von 50 Fischen ist kein markierter Fisch dabei. Was läßt sich über die Größe des Fischteichs N sagen?
Rein logisch, daß mindestens 100 Fische im Teich sind. Aber hätte man nur einen markierten Fisch gefangen, so wäre der Maximum Likelihood-Schätzer (berechnet im Hypergeometrischen Modell) $\hat{N} = 2500$. Folglich ist N wohl viel größer.

$$\begin{aligned} \text{Sei} \quad p_N &= P_N(\text{ kein markierter Fisch im Fang }), \\ \text{dann ist} \quad p_N &= \frac{\binom{N-50}{50}}{\binom{N}{50}} \quad \text{im Hypergeometrischen Modell.} \end{aligned}$$

Sei $\alpha > 0$ vorgegeben und $N(\alpha)$ so gewählt, daß $p_{N(\alpha)} \le \alpha$ aber $p_{N(\alpha)+1} > \alpha$ ist. Dann gilt auch $\max\limits_{N \le N(\alpha)} p_N \le \alpha$. Nun kann man, falls kein markierter Fisch im Fang ist, mit einer Irrtumswahrscheinlichkeit α folgern, daß $N \ge N(\alpha)$ ist.

Die folgende Tabelle liefert α in Abhängigkeit von N, einmal exakt und einmal als Binomialapproximation $p_N \approx \left(\frac{N-50}{N}\right)^{50}$. Für große N ist die Approximation erstaunlich gut.

N	α	α_{Bin}
300	$4,33 \cdot 10^{-5}$	$1,09 \cdot 10^{-4}$
600	0,0106	0,0128
1200	0,1137	0,1191
2500	0,3605	0,3642
5000	0,6035	0,6050

12.4.2 Definition (Konfidenzintervall)

Sei $\Omega \subset \mathbb{R}^n$ und P_θ eine Schar von Ws-Maßen, $\theta \in \Theta \subset \mathbb{R}$. Sei $\mathcal{R}$ die Menge der abgeschlossenen Intervalle auf $\mathbb{R}$. $I : \Omega \to \mathcal{R}$ heißt Konfidenzintervall (K.I.) mit Sicherheitswahrscheinlichkeit $1-\alpha$ mit $0 < \alpha < 1$, wenn für alle $\theta \in \Theta$ gilt: $P_\theta(\{\omega | \theta \in I(\omega)\}) \ge 1-\alpha$.

Bemerkungen:
1) $I(\omega) = \Theta$ ist Konfidenzintervall mit Sicherheitswahrscheinlichkeit (SWS) 1.

2) Ein Konfidenzintervall sollte aber möglichst klein sein.

12.4.3 Beispiele zu Konfidenzintervallen bei Normalverteilung

Seien $X_1, ..., X_n$ unabhängig und alle nach $N(\mu, \sigma^2)$-verteilt.

a) σ^2 sei bekannt, μ sei unbekannt. Sei k_α so, dass $\Phi(k_\alpha) = \alpha$ ist. Es gilt $-k_\alpha = k_{1-\alpha}$. Sei $x = (x_1, ..., x_n)$. $I_n(x) = [\hat{\mu}_n(x) - \frac{k_{1-\alpha}}{\sqrt{n}}\sigma, \hat{\mu}_n(x) + \frac{k_{1-\alpha}}{\sqrt{n}}\sigma]$ ist zweiseitiges Konfidenzintervall zur Sicherheitswahrscheinlichkeit $1 - 2\alpha$. Denn:

$$P_{\mu,\sigma^2}\left(\left\{x \middle| \, \mu \in \left[\hat{\mu}_n(x) \pm \frac{k_{1-\alpha}}{\sqrt{n}}\sigma\right]\right\}\right) = P_{\mu,\sigma^2}\left(\left\{x \middle| \left|\frac{\sqrt{n}(\hat{\mu}_n(x) - \mu)}{\sigma}\right| \le k_{1-\alpha}\right\}\right)$$

$$= 1 - 2\Phi(k_\alpha)$$
$$= 1 - 2\alpha.$$

b) σ^2 und μ seien unbekannt. Sei $t_{\beta;n}$ definiert durch $P(T_n \leq t_{\beta;n}) = \beta$, wobei T_n eine nach t_n-verteilte Zufallsvariable ist. Es gilt $-t_{\alpha;n} = t_{1-\alpha;n}$. Ersetze in I_n σ^2 durch seinen Schätzer $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x}_n)^2$. $\tilde{I}_n(x) = [\hat{\mu}_n(x) \pm \frac{t_{1-\alpha;n-1}}{\sqrt{n}}\hat{\sigma}_n(x)]$ ist $(1-2\alpha)$-Konfidenzintervall, denn:

$$\begin{aligned}
P_{\mu,\sigma^2}(\{x|\mu \in \tilde{I}_n(x)\}) &= P_{\mu,\sigma^2}\left(\left\{x \middle| \mu \in \left[\hat{\mu}_n(x) \pm \frac{t_{1-\alpha;n-1}}{\sqrt{n}}\hat{\sigma}_n(x)\right]\right\}\right) \\
&= P_{\mu,\sigma^2}\left(\left\{x \middle| \left|\frac{\sqrt{n}(\hat{\mu}_n(x) - \mu)}{\hat{\sigma}_n(x)}\right| \leq t_{1-\alpha;n-1}\right\}\right) \\
&= 1 - 2\alpha.
\end{aligned}$$

Wir kommen nun zurück zum Bierabfüllproblem und beantworten die anfangs gestellten Fragen 3)–5).

3. Ein $(1-\alpha)$-zweiseitiges Konfidenz-Intervall für μ lautet:

$$\left[\overline{x}_n - \frac{\hat{\sigma}_n}{\sqrt{n}}t_{1-\frac{\alpha}{2};n-1}, \overline{x}_n + \frac{\hat{\sigma}_n}{\sqrt{n}}t_{1-\frac{\alpha}{2};n-1}\right].$$

Für $\alpha = 0{,}10$ ergibt sich $t_{0,95;7} = 1{,}895$. Dies ist das obere 5 %-Quantil der t_7 Verteilung. Das Konfidenzintervall für μ ist $[0{,}4886; 0{,}5014]$, denn:

$$\begin{aligned}
&P_{\mu;\sigma^2}\left(\mu \in \left[\overline{X}_n \pm \frac{\hat{\sigma}_n}{\sqrt{n}}t_{1-\alpha/2;n-1}\right]\right) \\
&= P_{\mu;\sigma^2}\left(-t_{\alpha/2;n-1} \leq \underbrace{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\hat{\sigma}_n}}_{t_{n-1}\text{-verteilt}} \leq t_{1-\alpha/2;n-1}\right) \\
&= 1 - 2\frac{\alpha}{2} = 1 - \alpha
\end{aligned}$$

4. Zweiseitiges 95 %-Konfidenzintervall für σ^2. Aus Abschnitt 11.6 wissen wir, daß $\frac{n-1}{\sigma^2}\hat{\sigma}_n^2$ nach χ^2_{n-1}-verteilt ist. Ein $(1-\alpha)$-Konfidenzintervall für σ^2 ist

$$\left[\frac{(n-1)\hat{\sigma}_n^2}{\chi^2_{1-\alpha/2;n-1}}; \frac{(n-1)\hat{\sigma}_n^2}{\chi^2_{\alpha/2;n-1}}\right].$$

Für $\alpha = 0{,}05$ ist $\chi^2_{0,975;7} = 16,01$ und $\chi^2_{0,025;7} = 1{,}69$.
Das 95 %-Konfidenzintervall für σ^2 lautet $[3{,}97 \cdot 10^{-3}; 37{,}64 \cdot 10^{-5}]$.
Begründung: $P(\sigma^2 \in KI) = P_{\mu,\sigma^2}[\chi^2_{\alpha/2;n-1} \leq \frac{(n-1)}{\sigma^2}\hat{\sigma}_n^2 \leq \chi^2_{1-\alpha/2;n-1}] = 1 - \alpha$.

5. Präzisierung der Frage: Kann man mit einer Irrtumswahrscheinlichkeit α (z.B. 1 %) sagen, dass die Brauerei zu wenig abfüllt?
   Sei $\mu_0 = 0,5$. Wir geben nun ein einseitiges Konfidenzintervall I_n an, für das gilt:
   $$P_\mu(\mu \in I_n) \geq 1-\alpha \quad \text{für alle } \mu. \tag{*}$$
   Liegt aber μ_0 nicht in I_n, so schließen wir mit einer Irrtumswahrscheinlichkeit α, daß die Brauerei zu wenig abfüllt. Das Konfidenzintervall lautet:
   $$I_n = \left(-\infty, \overline{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}}\, t_{1-\alpha;n-1}\right).$$
   Dies erfüllt (*), denn
   $$\begin{aligned} P_\mu\left(\mu \in I_n\right) &= P_\mu\left(\mu \leq \overline{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}}\, t_{1-\alpha;n-1}\right) \\ &= P_\mu\left(\mu \leq \overline{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}}\, t_{\alpha;n-1}\right) \\ &= P_\mu\left(\sqrt{n}\frac{\overline{X}_n - \mu}{\hat{\sigma}_n} \geq t_{\alpha;n-1}\right) \\ &= 1-\alpha. \end{aligned}$$
   Setze nun $\alpha = 0,01$. Dann ist: $\overline{x}_8 + \frac{\hat{\sigma}_8}{\sqrt{8}} t_{0,99;7} = 0,495 + \frac{0,01}{\sqrt{8}} \cdot 3 = 0,505$ weil $t_{0,99;7} = 3$ ist. Weil $\mu_0 < 0,505$ ist, lässt sich mit einer Irrtumswahrscheinlichkeit von 1% nicht behaupten, die Brauerei fülle zu wenig ab.

Oft läßt sich das Konfidenzintervall nur näherungsweise angeben. Dabei hilft dann der folgende Satz über die asymptotische Verteilung des ML-Schätzers.

12.4.4 Satz

Sei Φ die Verteilungsfunktion von $N(0,1)$. $\hat{\theta}_n$ sei ML-Schätzer. Unter Glattheitsvoraussetzungen an $f_\theta(x)$ gilt für $\beta \in \mathbb{R}$:
$$\lim_{n\to\infty} P_\theta(\sqrt{nI(\theta)}(\hat{\theta}_n - \theta) \leq \beta) = \Phi(\beta)$$
mit $I(\theta)$ der Fisher-Information. Ist $I(\theta)$ stetig, so gilt auch
$$\lim_{n\to\infty} P_\theta\Big(\sqrt{nI(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \leq \beta\Big) = \Phi(\beta).$$

Bemerkung: Setzt man $\Phi(-\beta_\alpha) = \alpha$, so gilt $P_\theta\Big(\hat{\theta}_n - \frac{\beta_\alpha}{\sqrt{nI(\hat{\theta}_n)}} \leq \theta \leq \hat{\theta}_n + \frac{\beta_\alpha}{\sqrt{nI(\hat{\theta}_n)}}\Big) \to 1 - 2\Phi(-\beta_\alpha) = 1-2\alpha$. Das heißt, wir haben ein näherungsweises Konfidenzintervall zur Sicherheitswahrscheinlichkeit $(1-2\alpha)$ vorliegen.

12.4.5 Beispiel: Hufschlagtote in der Preußischen Armee

Man hat als Datensatz die Anzahl der Hufschlagtoten in 14 Kavallerie-Corps der Preußischen Armee über 20 Jahre von 1875 bis 1894.

Zusammengefaßt ergibt sich folgende Tabelle:

Anzahl Tote in Corps p.a.	Häufigkeit in Corps p.a.
0	144
1	91
2	32
3	11
4	2
≥ 5	0

Abbildung 6: Hufschlagtote

Die Daten werden sehr gut durch eine Poisson-verteilte Zufallsvariable X beschrieben. Dies ergibt sich aus einem χ^2-Anpassungstest, den wir hier nicht weiter diskutieren wollen. Es stellt sich die Frage nach Parameterschätzung und Konfidenzintervall von θ, dem Poisson-Parameter.
14 Corps über 20 Jahre entspricht 280 Corpsjahren. Der ML-Schätzer lautet:

$\hat{\theta}_{280} = \frac{196}{280} = 0{,}7$ Todesfälle pro Corpsjahr.

Ein asymptotisches 95 %-Konfidenzintervall ergibt sich mit
$n = 280$, $\hat{\theta}_n = 0{,}7$ und $I(\hat{\theta}_n) = \frac{1}{0{,}7}$ als
$\hat{\theta}_n \pm \frac{1{,}96}{\sqrt{nI(\hat{\theta}_n)}} = 0{,}7 \pm 0{,}098$ und damit als $[0{,}602; 0{,}798]$.

Anhang A

Mit Hilfe der Bayesschen Formel erfolgreich klassifizieren

Angenommen es gäbe k mögliche Zustände. Man entscheide möglichst optimal, welcher der Zustände vorliegt.
Gegeben seien die Wahrscheinlichkeiten $\pi(i)$ des i-ten Zustandes vor der Ziehung. Dabei gelte $\pi(i) \geq 0$ für $i = 1, ..., k$ und $\sum\limits_{i=1}^{k} \pi(i) = 1$.
$p(x|i), x \in \mathfrak{X}$ sei die Wahrscheinlichkeit x zu beobachten, wenn der Zustand i vorliegt.
Nach der Bayesschen Formel gilt:

$$p(j|x) = \frac{p(x|j)\pi(j)}{\sum\limits_i p(x|i)\pi(i)}.$$

Eine Abbildung $d : \mathfrak{X} \to \{1, ..., k\}$ heißt Klassifikationsvariable.

Beispiel:
$d^*(x) = j$, falls $p(j|x) \geq p(i|x) \forall i$.

$\pi(i)$ heißt a priori-Wahrscheinlichkeit,
$p(i|x)$ nennt man a posteriori-Wahrscheinlichkeit.

Ist d eine Klassifikationsvariable, so ist das Risiko eine falsche Entscheidung zu treffen $\mathcal{R}(d) = \sum\limits_{i=1}^{k} \pi(i) P_i(\{x \in \mathfrak{X} | d(x) \neq i\})$. Dabei ist P_i das Wahrscheinlichkeitsmaß auf $\mathfrak{X}$ bei Vorliegen des i-ten Zustandes.

Satz

Für alle Klassifikationsvariablen d gilt $\mathcal{R}(d^*) \leq \mathcal{R}(d)$.

Beweis:

$$\begin{aligned} d^*(x) = j &\Leftrightarrow p(j|x) = \max_i p(i|x) \\ &\Leftrightarrow p(x|j)\pi(j) = \max_i p(x|i)\pi(i) \end{aligned}$$

$$
\begin{aligned}
1-\mathcal{R}(d) &= 1-\sum_{i=1}^{k}\pi(i)P_i(\{x\in\mathfrak{X}|d(x)\neq i\})\\
&= \sum_{i=1}^{k}\pi(i)-\sum_{i=1}^{k}\pi(i)P_i(\{x\in\mathfrak{X}|d(x)\neq i\})\\
&= \sum_{i=1}^{k}\pi(i)(1-P_i(\{x\in\mathfrak{X}|d(x)\neq i\}))\\
&= \sum_{i=1}^{k}\pi(i)P_i(\{x\in\mathfrak{X}|d(x)= i\})\\
&= \sum_{i=1}^{k}\sum_{\substack{x\in\mathfrak{X}\\ d(x)=i}} p(x|i)\pi(i)\\
&= \sum_{x\in\mathfrak{X}}\sum_{i=1}^{k} 1_{\{x'\in\mathfrak{X}|d(x')=i\}}(x)p(x|i)\pi(i)\\
&\leq \sum_{x\in\mathfrak{X}}\max_j p(x|j)\pi(j)\\
&= \sum_{x\in\mathfrak{X}}\sum_{i=1}^{k} 1_{\{x'|d^*(x')=i\}}(x)p(x|i)\pi(i)\\
&= \sum_{i=1}^{k}\sum_{\substack{x\in\mathfrak{X}\\ d^*(x)=i}} p(x|i)\pi(i)\\
&= \sum_{i=1}^{k}\pi(i)P_i(\{x\in\mathfrak{X}|d^*(x)= i\})\\
&= \sum_{i=1}^{k}\pi(i)(1-P_i(\{x\in\mathfrak{X}|d^*(x)\neq i\}))\\
&= 1-\sum_{i=1}^{k}\pi(i)P_i(\{x\in\mathfrak{X}|d^*(x)\neq i\})\\
&= 1-\mathcal{R}(d^*),
\end{aligned}
$$

wobei $1_A(x) = \begin{cases} 1 \text{ für } x\in A \\ 0 \text{ für } x\notin A \end{cases}$.

Literaturverzeichnis

Literatur:

Dinges-Rost : Prinzipien der Stochastik, Teubner Studienbücher, 1982

Dümbgen, L. : Stochastik für Informatiker, Springer, 2003

Engel, A. : Wahrscheinlichkeitsrechnung und Statistik I - II, Klett, 1973

Feller, W. : An Introduction to Probability and its Applications, Vol I, Wiley, 1957

Georgii, H.-O. : Stochastik, Walter de Gruyter, 2002

Krengel, U. : Einführung in die Wahrscheinlichkeitstheorie und die Statistik, Vieweg, 2002, 3. Aufl.

Krickeberg-Ziezold : Stochastische Methoden, Springer, 4. Aufl., 1995

Pfanzagl, J. : Elementare Wahrscheinlichkeitsrechnung, Walter de Gruyter, 1991

Shiryaev, A. : Probability, Springer, 1984

Zur Historie:

Hoffmann-Jørgensen, J. : Probability with a View towards Statistics, Chapman & Hall, 1994

Einführend:

Pitman, J. : Probability, Springer 1994