Wahrscheinlichkeits-Rechnung und Statistik

Eine Einführung

Karl Stroetmann

18. Dezember 2006

Inhaltsverzeichnis

Kapitel 1

Einführung

Die mathematische Statistik, die oft auch als induktive Statistik bezeichnet wird, beschäftigt sich mit dem Testen von Hypothesen und dem Schätzen von Parametern. Zusätzlich hat die mathematische Statistik die Aufgabe, die *Konfidenz* für eine solche Schätzung anzugeben. Die Konfidenz stellt dabei ein Maß für die Sicherheit der gemachten Voraussage dar. Statistische Methoden werden immer dann angewendet, wenn es entweder zu aufwendig oder sogar unmöglich ist alle Elemente einer gegebenen Menge auf eine Eigenschaft hin zu untersuchen. Wir geben hierfür ein einfaches Beispiel. Ein Biologe hat im Wald einen Ameisenhaufen entdeckt und möchte wissen, wieviele Ameisen in dieser Kolonie wohnen. Da es zu aufwendig ist, alle Ameisen zu zählen, wählt er ein statistisches Verfahren: Er fängt 1 000 Ameisen und markiert diese Ameisen mit einem Farbtupfer. Anschließend entlässt er die markierten Ameisen in die Freiheit. Am nächsten Tag kommt er wieder und fängt 200 Ameisen. Er stellt fest, dass sich unter diesen Ameisen 5 Tiere befinden, die einen Farbtuper haben. Unter der Annahme, dass sich die Ameisen in der Nacht gut durchmischt haben und dass folglich die Wahrscheinlichkeit für eine Ameise, bei der zweiten Zählung gefangen zu werden, unabhängig davon ist, ob die Ameise farblich markiert ist, kann der Biologe nun die Anzahl aller Ameisen schätzen:

1. Bezeichnet n die Gesamtzahl der Ameisen, so ist der Prozentsatz p der markierten Ameisen durch die Formel
$$p = \frac{1000}{n} \tag{1.1}$$
gegeben.

2. Aus der Voraussetzung, dass die Wahrscheinlichkeit dafür, dass eine Ameise bei der zweiten Zählung gefangen wird, unabhängig von der Markierung ist, folgt, dass für die Zahl der gefangenen Tiere, die farblich markiert sind, in etwa
$$5 = p \cdot 200 \tag{1.2}$$
gilt.

Setzen wir den Wert von p aus Gleichung (1.1) in Gleichung (1.2) ein, so finden wir
$$5 = \frac{1000}{n} \cdot 200 \quad \Leftrightarrow \quad n = \frac{1000 \cdot 200}{5} = 40\,000. \tag{1.3}$$
Der Biologe wird also vermuten, dass die Kolonie von etwa 40 000 Ameisen bewohnt wird. Die Statistik beantwortet zusätzlich die Frage, wie sicher diese Aussage ist. Wir werden im Laufe der Vorlesung ein Verfahren entwickeln, mit dessen Hilfe wir ein Intervall $[n_1, n_2]$ berechnen können, so dass für die Anzahl n der Ameisen die Aussage
$$n \in [n_1, n_2]$$

mit einer *Konfidenz* von $95,0\%$ gilt.

Das Gebäude der mathematischen Statistik ruht auf zwei Säulen:

1. Wahrscheinlichkeitsrechnung und
2. Maßtheorie

Mit der Wahrscheinlichkeitsrechnung werden wir uns im Laufe der Vorlesung ausführlich beschäftigen. Die Maßtheorie wird dann wichtig, wenn wir uns mit kontinuierlichen Größen beschäftigen. Die Behandlung der Maßtheorie würde allerdings den Rahmen dieser Vorlesung bei weitem sprengen, so dass wir uns im wesentlichen auf diskrete Größen beschränken werden.

Die mathematische Statistik wird in einer Vielzahl von Bereichen angewendet:

1. Versicherungs-Gesellschaften berechnen den Preis einer Versicherung mit Hilfe statistischer Methoden.
2. Mediziner verwenden statistische Methoden zur Einschätzung der Wirksamkeit von Medikamenten.
3. In der Informatik werden statistische Methoden unter anderem
   (a) bei der Spracherkennung,
   (b) zur Signalverarbeitung,
   (c) bei Expertensystemen,
   (d) beim *Data Mining*,

   und noch in zahlreichen anderen Gebieten eingesetzt. Eigentlich gibt es kaum mehr ein Gebiet der Informatik, in dem nicht auf die eine oder andere Weise Statistik oder wenigstens Wahrscheinlichkeitstheorie eingesetzt wird. Ein ganz modernes Beispiel für den Einsatz von Statistik ist übrigens *Google*™. Dieses Beispiel zeigt, dass mit Statistik auch richtig viel Geld verdient werden kann und sollte Sie dazu motivieren, in dieser Vorlesung gut aufzupassen!

Wir werden uns in dieser Vorlesung zunächst (und schwerpunktmäßig) mit der Wahrscheinlichkeitsrechnung beschäftigen, denn andernfalls müsste ich Ihnen statistische Verfahren in Form von Kochrezepten beibringen, die Sie nicht verstehen sondern nur anwenden könnten. Das kann aber kaum der Sinn einer Mathematik-Vorlesung sein.

Neben der mathematischen Statistik gibt es auch noch die *deskriptive Statistik*, deren Aufgabe es ist, eine gegebene Menge von Daten durch geeignete Kennzahlen (Mittelwerte, Streuung, etc.) zu beschreiben. Dieser Zweig der Statistik spielt vor allem in den Sozialwissenschaften und in der Politik eine wichtige Rolle. Aus Zeitgründen werden wir diesen Zweig der Statistik in dieser Vorlesung nur streifen können.

Kapitel 2

Einführung in die Wahrscheinlichkeits-Rechnung

Unter einem *Zufalls-Experiment* verstehen wir ein Experiment, dessen Ausgang nicht eindeutig vorbestimmt ist. Ein Beispiel für ein solches Experiment wäre der Wurf einer Münze, bei dem als Ergebnis entweder *Wappen* oder *Zahl* auftritt. Ein anderes Zufalls-Experiment wäre der Wurf eines Würfels. Hier können als Ergebnisse die natürlichen Zahlen 1 bis 6 auftreten. Mathematisch werden solche Zufalls-Experimente durch den Begriff des *Wahrscheinlichkeits-Raums* erfasst. Hier gibt es zwei Möglichkeiten. Falls die Ergebnisse durch reelle Zahlen dargestellt werden, so haben wir es mit einen *kontinuierlichen Wahrscheinlichkeits-Raum*, andernfalls sprechen wir von einem *diskreten Wahrscheinlichkeits-Raum*. Die Theorie der kontinuierlichen Wahrscheinlichkeits-Räume ist wesentlich komplexer als die der diskreten Wahrscheinlichkeits-Räume, daher beginnen wir mit letzteren.

2.1 Diskrete Wahrscheinlichkeits-Räume

Definition 1 (Wahrscheinlichkeits-Raum) Ein *diskreter Wahrscheinlichkeits-Raum* ist ein Tripel $\langle \Omega, 2^\Omega, P\rangle$, so dass gilt:

1. Ω ist eine Menge, die entweder endlich ist, dann gilt also

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

   oder Ω ist *abzählbar unendlich*, dann gilt

   $$\Omega = \{\omega_n \mid n \in \mathbb{N}\}.$$

   Die Elemente von Ω bezeichnen wir als die *Ergebnisse* eines Zufalls-Experiments und die Menge Ω nennen wir den *Ergebnis-Raum*.

2. 2^Ω ist die Potenzmenge von Ω, also die Menge aller Teilmengen der Menge Ω. Diese Teilmengen bezeichnen wir auch als *Ereignisse* und die Menge 2^Ω nennen wir den *Ereignis-Raum*. Mengen der Form $\{\omega_i\}$, die genau ein Element aus Ω enthalten, nennen wir *Elementar-Ereignisse*.

3. $P : 2^\Omega \to \mathbb{R}$ ist eine Abbildung, die jedem Ereignis $A \subseteq \Omega$ eine reelle Zahl $P(A)$ zuordnet. Die Zahl $P(A)$ bezeichnen wir als die *Wahrscheinlichkeit*, mit der das Ereignis A eintritt. Die Wahrscheinlichkeit P muss den folgenden *Kolmogorow-Axiomen* (Andrei Nikolajewitsch Kolmogorow; 1903 – 1987) genügen:

   (a) $0 \leq P(A) \leq 1$ für alle $A \subseteq \Omega$.

   (b) $P(\emptyset) = 0$.

   Die leere Menge bezeichnen wir als das *unmögliche Ereignis*.

(c) $P(\Omega) = 1$.
Die Menge Ω bezeichnen wir als das *sichere Ereignis*.

(d) $\forall A, B \in 2^\Omega : A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$
Schließen zwei Ereignisse A und B sich gegenseitig aus, gilt also $A \cap B = \emptyset$, so nennen wir diese Ereignisse *unvereinbar*. Sind A und B unvereinbare Ereignisse, so ergibt sich die Wahrscheinlichkeit des Ereignisses $A \cup B$ als die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse.

Die Funktion P bezeichnen wir als die *Wahrscheinlichkeits-Verteilung*.

Schreibweise: Das vierte Kolmogorow-Axiom bezeichnen wir als die *Additivität* der Wahrscheinlichkeits-Verteilung. Um dieses Axiom einfacher schreiben zu können, vereinbaren wir folgende Schreibweise: Sind A und B zwei <u>disjunkte</u> Mengen, so schreiben wir die Vereinigung von A und B als $A \uplus B$. Der Term $A \uplus B$ steht also für zweierlei:

1. Für die Vereinigungs-Menge $A \cup B$.
2. Für die Aussage $A \cap B = \emptyset$.

Mit dieser Schreibweise lautet das Axiom der Additivität

$$P(A \uplus B) = P(A) + P(B).$$

Diese Gleichung gilt genau dann, wenn $A \uplus B$ definiert ist und das ist genau dann der Fall, wenn $A \cap B = \emptyset$ gilt.

Beispiel: Ein möglicher Wahrscheinlichkeits-Raum für das Zufalls-Experiment "Würfeln mit einem kubischen Würfel" ist

$$\langle \Omega, 2^\Omega, P \rangle \quad \text{mit}$$

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- $P(A) = \dfrac{|A|}{|\Omega|} = \dfrac{1}{6} \cdot |A|$.

  Hier bezeichnet $|A|$ die Anzahl der Elemente der Menge A.

Das Ereignis "es wurde eine gerade Zahl gewürfelt" wird dann durch die Menge $G = \{2, 4, 6\}$ beschrieben. Für die Wahrscheinlichkeit von G gilt

$$P(G) = \frac{1}{6} \cdot |G| = \frac{1}{6} \cdot 3 = \frac{1}{2}.$$

Bei dem oben angegebenen Wahrscheinlichkeits-Raum sind wir davon ausgegangen, dass alle Elementar-Ereignisse die selbe Wahrscheinlichkeit haben. Diese Annahme ist aus Symmetrie-Gründen naheliegend. In diesem Fall nennen wir die Wahrscheinlichkeits-Verteilung P auch *gleichmäßig*, das zugehörige Zufalls-Experiment heißt dann ein *Laplace-Experiment* (Pierre Simon Laplace; 1749 – 1827). Falls die Wahrscheinlichkeit für alle Seiten eines Würfels den selben Wert hat, so sprechen wir von einem Laplace-Würfel.

Allgemein hat der Wahrscheinlichkeits-Raum bei einem Laplace-Experiment die Form

$$\langle \Omega, 2^\Omega, P \rangle \quad \text{mit}$$

1. Ω endlich und
2. $P(A) = \dfrac{|A|}{|\Omega|}$.

Die Gültigkeit der Kolmogorow-Axiome kann in diesem Fall sofort nachgerechnet werden.

Die Annahme der Gleichmäßigkeit einer Wahrscheinlichkeits-Verteilung ist logisch nicht zwingend. Wenn der Würfel beispielsweise in unsymmetrischer Weise mit Blei beschwert ist, so könnte die Wahrscheinlichkeits-Verteilung auch wie folgt gegeben sein:

$$P(\{1\}) = 0.5, \quad P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = 0.1.$$

Dann würden wir für das Ereignis "es wurde eine gerade Zahl gewürfelt" gelten:

$$P(G) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 0.1 + 0.1 + 0.1 = 0.3.$$

2.2 Additionssätze

Die Kolmogorow-Axiome geben an, wie sich die Wahrscheinlichkeit eines Ereignisses $A \cup B$ aus den Wahrscheinlichkeiten der Ereignisse A und B berechnen läßt, wenn die Ereignisse A und B unvereinbar sind. Wir wollen jetzt den Fall untersuchen, dass $A \cap B \neq \emptyset$ gilt. In diesem Fall zerlegen wir die Menge $A \cup B$ in drei Teilmengen:

$$A \cup B = (A\backslash B) \uplus (B\backslash A) \uplus (A \cap B)$$

Da die drei Mengen $A\backslash B$, $B\backslash A$ und $A \cap B$ paarweise disjunkt sind gilt

$$P(A \cup B) = P(A\backslash B) + P(B\backslash A) + P(A \cap B) \tag{2.1}$$

Außerdem gilt

$$A = (A\backslash B) \uplus (A \cap B) \quad \text{und} \quad B = (B\backslash A) \uplus (A \cap B).$$

Daraus folgt sofort

$$P(A) = P(A\backslash B) + P(A \cap B) \qquad \text{und} \qquad P(B) = P(B\backslash A) + P(A \cap B).$$

Subtrahieren wir auf beiden Seiten dieser Gleichungen den Term $P(A \cap B)$, so erhalten wir

$$P(A) - P(A \cap B) = P(A\backslash B) \qquad \text{und} \qquad P(B) - P(A \cap B) = P(B\backslash A).$$

Aus Gleichung (2.1) folgt nun

$$\begin{aligned} P(A \cup B) &= P(A\backslash B) + P(B\backslash A) + P(A \cap B) \\ &= \big(P(A) - P(A \cap B)\big) + \big(P(B) - P(A \cap B)\big) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned} \tag{2.2}$$

Es gibt eine ganz analoge Formel zur Berechnung der Anzahl der Elemente einer Menge. Bezeichen wir für eine Menge M die Anzahl ihrer Elemente mit $|M|$, so gilt

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Diese Formel läßt sich wie folgt interpretieren: Wenn wir die Elemente von A zählen und anschließend die Elemente von B zählen, so zählen wir die Elemente der Schnittmenge $A \cap B$ doppelt und müssen daher die Anzahl dieser Elemente abziehen.

Die Gleichung (2.2) läßt sich verallgemeinern. Betrachten wir die Vereinigung dreier Mengen A, B und C so finden wir

$$\begin{aligned}
& P(A \cup B \cup C) \\
= \;& P\big((A \cup B) \cup C\big) \\
= \;& P(A \cup B) + P(C) - P\big((A \cup B) \cap C\big) \\
& \text{Gleichung (2.2) auf } A \cup B \text{ anwenden} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - P\big((A \cup B) \cap C\big) \\
& \text{Distributiv-Gesetz } (A \cup B) \cap C = (A \cap C) \cup (B \cap C) \text{ berücksichtigen} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - P\big((A \cap C) \cup (B \cap C)\big) \\
& \text{Gleichung (2.2) auf } (A \cap C) \cup (B \cap C) \text{ anwenden} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - \big(P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C))\big) \\
& (A \cap C) \cap (B \cap C) = A \cap B \cap C \text{ berücksichtigen} \\
= \;& P(A) + P(B) - P(A \cap B) + P(C) - \big(P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)\big) \\
& \text{Klammer auflösen und umsortieren} \\
= \;& P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)
\end{aligned}$$

In ähnlicher Weise lassen sich Formeln herleiten, die die Wahrscheinlichkeit einer Vereinigung von n Ereignissen angeben.

Ist ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^\Omega, P\rangle$ gegeben, so definieren wir für ein Ereignis $A \in 2^\Omega$ das *Komplement* von A als

$$A^c = \Omega\backslash A.$$

Wegen $A \cup A^c = \Omega$ gilt dann

$$P(A^c) = 1 - P(A).$$

Aufgabe 1: An einer Schule können die Schüler Spanisch, Englisch und Französisch lernen.

- 40% der Schüler lernen Spanisch.
- 60% der Schüler lernen Englisch.
- 55% der Schüler lernen Französisch.
- 30% der Schüler lernen Spanisch und Englisch.
- 20% der Schüler lernen Spanisch und Französisch.
- 35% der Schüler lernen Französisch und Englisch.
- 10% der Schüler lernen Spanisch, Französisch und Englisch.

Wie groß ist der Prozentsatz der Schüler, die überhaupt keine Fremdsprache lernen.

Lösung: Wir führen folgende Bezeichnungen ein.

1. S: Menge der Schüler, die Spanisch lernt.
2. F: Menge der Schüler, die Französisch lernt.
3. E: Menge der Schüler, die Englisch lernt.
4. K: Menge der Schüler, die keine Fremdsprache lernt.

Dann gilt

$$\begin{array}{lcl}
P(K) & = & P\big((S \cup F \cup E)^c\big) \\
 & = & 1 - P(S \cup F \cup E) \\
 & = & 1 - P(S) - P(F) - P(E) + P(S \cap F) + P(S \cap E) + P(F \cap E) - P(S \cap E \cap F) \\
 & = & 1 - 0.4 - 0.55 - 0.6 + 0.2 + 0.3 + 0.35 - 0.1 \\
 & = & 0.2
\end{array}$$

Also lernen 20% der Schüler überhaupt keine Fremdsprache. □

Aufgabe 2: Zur Ermittelung der Noten würfelt ein Lehrer mit zwei Laplace-Würfeln. Die Note ergibt sich dann als das Minimum der gewürfelten Augenzahlen.

1. Geben Sie den Wahrscheinlichkeits-Raum für dieses Zufalls-Experiment an.
2. Geben Sie die für die Notenvergabe relevanten Ereignisse an.

Lösung: Wir definieren

$$\Omega = \big\{\langle i, j\rangle \mid i \in \{1, 2, 3, 4, 5, 6\},\ j \in \{1, 2, 3, 4, 5, 6\}\big\}$$

Da es sich um zwei Laplace-Würfel handelt, sind alle Ereignisse gleich wahrscheinlich. Da $|\Omega| = 36$ ist, hat die Wahrscheinlichkeits-Verteilung für eine beliebige Menge $A \in 2^\Omega$ den Wert

$$P(A) = \tfrac{1}{36} \cdot |A|.$$

Die für die Notenvergabe interessanten Ereignisse sind:

1. Vergabe einer 1:
$$A_1 = \big\{\langle 1,n\rangle \mid n \in \{1,\cdots,6\}\big\} \cup \big\{\langle n,1\rangle \mid n \in \{1,\cdots,6\}\big\}$$
Die Menge A_1 besteht aus $6+6-1$ Elementen, denn
$$\big\{\langle 1,n\rangle \mid n \in \{1,\cdots,6\}\big\} \cap \big\{\langle n,1\rangle \mid n \in \{1,\cdots,6\}\big\} = \big\{\langle 1,1\rangle\big\}.$$
Also hat die Wahrscheinlichkeit für die Note 1 den Wert $\frac{11}{36} = 0.30\overline{5}$.

2. Vergabe einer 2:
$$A_2 = \big\{\langle 2,n\rangle \mid n \in \{2,\cdots,6\}\big\} \cup \big\{\langle n,2\rangle \mid n \in \{2,\cdots,6\}\big\}$$
Die Menge A_2 besteht aus $5+5-1$ Elementen. Also hat die Wahrscheinlichkeit für die Note 2 den Wert $\frac{9}{36} = 0.25$.

3. Vergabe einer 3:
$$A_3 = \big\{\langle 3,n\rangle \mid n \in \{3,4,5,6\}\big\} \cup \big\{\langle n,3\rangle \mid n \in \{3,4,5,6\}\big\}$$
Die Menge A_3 besteht aus $4+4-1$ Elementen. Also hat die Wahrscheinlichkeit für die Note 3 den Wert $\frac{7}{36} = 0.19\overline{4}$.

4. Vergabe einer 4:
$$A_4 = \big\{\langle 4,n\rangle \mid n \in \{4,5,6\}\big\} \cup \big\{\langle n,4\rangle \mid n \in \{4,5,6\}\big\}$$
Die Menge A_4 besteht aus $3+3-1$ Elementen. Also hat die Wahrscheinlichkeit für die Note 4 den Wert $\frac{5}{36} = 0.13\overline{8}$.

5. Vergabe einer 5:
$$A_5 = \big\{\langle 5,n\rangle \mid n \in \{5,6\}\big\} \cup \big\{\langle n,5\rangle \mid n \in \{5,6\}\big\}$$
Die Menge A_5 besteht aus $2+2-1$ Elementen. Also hat die Wahrscheinlichkeit für die Note 5 den Wert $\frac{3}{36} = 0.08\overline{3}$.

6. Vergabe einer 6:
$$A_6 = \big\{\langle 6,6\rangle\big\}$$
Die Menge A_6 enthält nur ein Element. Also hat die Wahrscheinlichkeit für die Note 6 den Wert $\frac{1}{36} = 0.02\overline{7}$. □

Aufgabe 3: Der Lehrer aus der letzten Aufgabe stellt fest, dass sich bei seinem Verfahren ein zu guter Notendurchschnitt ergibt. Daher ändert er das Verfahren ab: Zur Ermittelung der Note wird jetzt die Summe der Augenzahlen durch zwei geteilt. Falls das Ergebnis keine ganze Zahl ist, wird aufgerundet. Geben Sie die für die Notenvergabe relevanten Ereignisse an und berechnen Sie die Wahrscheinlichkeiten für das Auftreten dieser Ereignisse. □

2.3 Kombinatorik

Die letzten beiden Aufgaben zeigen, dass die Berechnung von Wahrscheinlichkeiten oft auf die Berechnung der Anzahl der Elemente einer Menge zurück geführt werden kann. Wir werden daher jetzt diese Anzahl für bestimmte Mengen, die in der Praxis häufig vorkommen, berechnen. Die Anzahl der Elemente einer Menge bezeichnen wir auch als die *Mächtigkeit* oder auch *Kardinalität* der Menge.

Wir nehmen als erstes an, dass wir n Mengen A_1, A_2, $\cdots$, A_n gegeben haben. Wir untersuchen die Menge M aller Listen der Länge n, für die das i-te Element ein Element der Menge M_i ist. Die Menge M ist nichts anderes als das kartesische Produkt der Mengen A_1, A_2, $\cdots$, A_n.
$$M = A_1 \times A_2 \times \cdots \times A_n = \big\{[x_1,\cdots,x_n] \;\big|\; \forall i \in \{1,\cdots,n\} : x_i \in A_i\big\}$$
Für die Mächtigkeit dieser Menge gilt
$$|M| = |A_1| * |A_2| * \cdots * |A_n|, \tag{2.3}$$

denn für das erste Element einer Liste gibt es $|A_1|$ Möglichkeiten, für das zweite Element gibt es $|A_2|$ Möglichkeiten und für das letzte Element gibt es $|A_n|$ Möglichkeiten. Da die einzelnen Möglichkeiten beliebig kombiniert werden können, ist die Gesamtzahl aller Möglichkeiten durch das Produkt gegeben. Die Gleichung (2.3) bezeichnen wir daher als Produkt-Regel.

Anzahl der k-Tupel mit Wiederholung Ist M eine Menge der Mächtigkeit n, so gibt es nach der Produkt-Regel insgesamt n^k Möglichkeiten, ein Tupel der Länge k mit Elementen aus M zu bilden, denn es gilt

$$\big|M^k\big| = |M|^k = n^k. \tag{2.4}$$

Anzahl der k-Tupel ohne Wiederholung Oft sind nur solche k-Tupel interessant, die kein Element mehrfach enthalten. Solche k-Tupel bezeichnen wir als *Permutationen*. Ist M eine Menge und $k \in \mathbb{N}$, so definieren wir die *Menge der k-Permutationen aus M* als

$$P(M,k) = \big\{[x_1, \cdots, x_k] \in M^k \;\big|\; \forall i, j \in \{1, \cdots, k\} : i \neq j \rightarrow x_i \neq x_j\big\}.$$

Gilt $|M| = n$, so haben wir für das Element x_1 insgesamt n Möglichkeiten. Für das zweite Element x_2 einer Permutation haben wir eine Möglichkeit weniger, also nur noch $n-1$ Möglichkeiten der Auswahl. Allgemein haben wir für das i-te Element x_i nur noch $n-(i-1)$ Möglichkeiten, denn wir haben ja schon $i-1$ Elemente vor ausgewählt. Damit ergibt sich

$$\big|P(M,k)\big| = n * (n-1) * \cdots * \big(n-(k-1)\big) = \frac{n!}{(n-k)!}. \tag{2.5}$$

Als Speziallfall ergibt sich

$$\big|P(M,n)\big| = n!.$$

Damit gibt es insgesamt $n!$ Möglichkeiten um die Elemente einer Menge der Mächtigkeit n in einer Liste anzuordnen.

Anzahl der k-Kombinationen ohne Wiederholung Wir bestimmen jetzt die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Dazu definieren wir $C(M,k)$ als die Menge aller k-elementigen Teilmengen von M:

$$C(M,k) = \big\{N \in 2^M \;\big|\; |N| = k\big\}.$$

Diese Teilmengen bezeichnen wir auch als die k-elementigen *Kombinationen* der Menge M. Um die Mächtigkeit von $C(M,k)$ zu bestimmen überlegen wir uns, wie die Mengen $P(M,k)$ und $C(M,k)$ zusammenhängen. Ist eine Kombination

$$\{x_1, x_2, \cdots, x_k\} \in C(M,k)$$

gegeben, so gibt es $k!$ Möglichkeiten, um diese k Elemente in einem Tupel anzuordnen. Daher erhalten wir die Anzahl der k-Permutationen aus der Anzahl der k-Kombinationen durch Multiplikation mit $k!$:

$$|P(M,k)| = |C(M,k)| * k!.$$

Setzen wir hier den Wert ein, den wir in Gleichung (2.5) für die Anzahl der k-Permutationen gefunden haben, so erhalten wir

$$\frac{n!}{(n-k)!} = |C(M,k)| * k!.$$

Division dieser Gleichung durch $k!$ liefert

$$|C(M,k)| = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k} \tag{2.6}$$

Anzahl der k-Kombinationen mit Wiederholung Als letztes stellen wir uns die Frage, wieviele Möglichkeiten es gibt, aus einer Menge M der Mächtigkeit n solcher Multimenge auszuwählen, die aus genau k Elementen bestehen. Um Multimengen von Mengen zu unterscheiden setzen wir den Index $_m$ an die schließende Klammer, wir schreiben also beispielsweise

$$A = \{1, 2, 2, 2, 2, 3\}_m$$

um auszudrücken, dass A eine Multimenge ist, die die Elemente 1, 2 und 3 enthält und bei der außerdem das Element 2 mit der Vielfachheit 4 auftritt. Die Mächtigkeit von A ist dann 6. Analog benutzen wir für die Vereinigung von Multimengen den Operator $\cup_m$, es gilt also beispielsweise

$$\{1, 2, 2, 3\}_m \cup_m \{1, 2, 3, 4, 4\}_m = \{1, 1, 2, 2, 2, 3, 3, 4, 4\}_m$$

Mit dieser Definition der Vereinigung von Multimengen gilt offenbar

$$|A \cup_m B| = |A| + |B| \quad \text{für beliebige Multimengen } A \text{ und } B.$$

Die Menge aller Multimengen der Mächtigkeit k mit Elementen aus M bezeichnen wir mit

$$C_m(M, k).$$

Unsere Idee ist, die Auswahl der Multimengen auf die Auswahl von Mengen zurück zu führen. Hat die Menge M die Form $M = \{x_1, \cdots, x_n\}$ und wollen wir aus M eine Multimenge mit k Elementen auswählen, so definieren wir eine Menge M^+, die zusätzlich zu den Elementen von M noch $k - 1$ *Joker-Elemente* enthält:

$$M^+ := \{x_1, \cdots, x_n, j_1, \cdots, j_{k-1}\}.$$

Die Behauptung ist nun, dass die Teilmengen von M^+ genau den Multimengen von M entsprechen. Um diese Behauptung zu belegen müssen wir jeder Teilmenge von M^+ genau eine Multimenge mit k Elementen aus M zuordnen. Wir brauchen also eine Funktion f

$$f : C(M^+, k) \to C_m(M, k)$$

Diese Funktion f muss bijektiv sein:

1. f muss surjektiv sein, für jedes $L \in C_m(M, k)$ muss es also ein $L^+ \in C(M^+, k)$ mit der Eigenschaft

   $$f(L^+) = L$$

   geben.

2. f muss injektiv sein, es muss also für alle $L_1^+, L_2^+ \in C(M^+, k)$

   $$f(L_1^+) = f(L_2^+) \to L_1^+ = L_2^+$$

   gelten.

Um f definieren zu können ist es notwendig, die Mengen aus $C(M^+, k)$ in eindeutiger Weise zu repräsentieren. Zu diesem Zweck legen wir auf den Elementen aus M^+ wie folgt eine Reihenfolge fest:

$$x_1 < x_2 < \cdots < x_n < j_1 < j_2 < \cdots < j_{k-1}.$$

Wichtig ist dabei lediglich, dass die Elemente aus M in der Reihenfolge vor den Joker-Elementen kommen. Nachdem diese Reihenfolge festgelegt ist, können wir die Mengen aus $C(M^+, k)$ durch geordnete Listen der Länge k repräsentieren. Auch die Multimengen aus $C_m(M, k)$ repräsentieren wir durch Listen der Länge k. Der Unterschied zu den anderen Listen besteht darin, dass in den Listen, die Multimengen repräsentieren, Elemente mehrfach auftreten können. Dafür enthalten diese Listen aber keine Joker-Elemente. Die Funktion f definieren wir jetzt durch Induktion nach der Anzahl der Elemente k.

I.A.: $k = 0$. Wir setzen

$$f([]) = [].$$

I.S.: $k \mapsto k + 1$. Jetzt gibt es zwei Fälle. Bei der Berechnung von

$$f([y_1, \cdots, y_k, y_{k+1}])$$

müssen wir unterscheiden, ob das letzte Element y_{k+1} ein Joker-Element ist oder nicht.

(a) y_{k+1} ist ein Joker-Element, es gilt also $y_{k+1} = j_i$ für ein $i \in \{1, \cdots, k - 1\}$. Zunächst definieren wir nach IV

$$L := f\big([y_1, \cdots, y_k]\big).$$

Dies ist eine Liste der Länge k. Wir bezeichnen das i-te Element dieser Liste L mit $L(i)$. Das Joker-Element j_i interpretieren wir als Anweisung das i-te Element der Liste L an die Liste hinten anzuhängen, wir definieren also

$$f\big([y_1, \cdots, y_k, y_{k+1}]\big) := L + \big[L(i)\big].$$

(b) y_{k+1} ist kein Joker-Element, es gilt also $y_{k+1} \in M$. Wieder definieren wir nach IV

$$L := f\big([y_1, \cdots, y_k]\big).$$

und setzen dann

$$f\big([y_1, \cdots, y_k, y_{k+1}]\big) := L + [y_{k+1}].$$

Um die obige Definition zu erläutern, demonstrieren wir die Definition der Funktion f an einem Beispiel. Wir betrachten die Menge

$$M = \{a, b, c, d\},$$

die aus den ersten vier Buchstaben des Alphabets besteht. Wir wollen die Menge $C_m(M, 3)$ berechnen, die aus allen dreielementigen Multimengen der ersten vier Buchstaben besteht. Also ist $k = 3$ und wir benötigen $k - 1 = 2$ Joker-Elemente. Wir verwenden die natürlichen Zahlen 1 und 2 als diese Joker-Elemente. Im Prinzip ist es völlig egal, was wir als Joker-Elemente definieren, solange sich diese nur von den Elementen der Menge M unterscheiden. Jetzt gilt also

$$M^+ = \{a, b, c, d, 1, 2\}.$$

Wir ordnen die Elemente der Menge M wie folgt an:

$$a < b < c < d < 1 < 2.$$

Als erstes berechnen wir f für die Menge $\{a, 1, 2\} \in C(M^+, 3)$, die durch die Liste $[a, 1, 2]$ dargestellt wird. Wir haben

$$f([a]) = f([]) + [a] = [] + [a] = [a], \quad f([a, 1]) = [a] + [a](1) = [a, a] \quad \text{und}$$
$$f([a, 1, 2]) = [a, a] + [a, a](2) = [a, a, a].$$

Also entspricht die Menge $\{a, 1, 2\}$ der Multimenge $\{a, a, a\}$. Als zweites Beispiel betrachten wir die Menge $\{a, b, 2\} \in C(M^+, 3)$. Diesmal ergibt sich

$$f([a]) = [a], \quad f([a, b]) = [a, b] \quad \text{und} \quad f([a, b, 2]) = [a, b] + [a, b](2) = [a, b, b].$$

Diese Beispiele sollten Sie davon überzeugen, dass die Funktion f tatsächlich eine bijektive Abbildung der Menge $C(M^+, k)$ in die Menge $C_m(M, k)$ ist. Wenn es eine solche bijektive Abbildung gibt, dann müssen diese beiden Mengen aber die selbe Anzahl von Elementen haben. Besteht die Menge M aus n Elementen, so hat die Menge M^+ die Mächtigkeit $n + k - 1$. Damit gilt

$$|C_m(M, k)| = |C(M^+, k)| = \binom{n + k - 1}{k}. \tag{2.7}$$

Die Anzahl der k-elementigen Multimengen einer n-elementigen Menge hat also den Wert $\binom{n+k-1}{k}$. In der Wahrscheinlichkeits-Theorie werden solche Multimengen auch als *Kombinationen mit Wiederholung* bezeichnet.

Aufgabe 4: Es sei eine Gruppe von n Personen gegeben. Berechnen Sie die Wahrscheinlichkeit dafür, dass von diesen n Personen wenigstens zwei Personen am selben Tag Geburtstag haben. Zur Vereinfachung dürfen Sie folgendes annehmen:

1. Keine der Personen hat an einem Schalttag Geburtstag.
2. Für alle anderen Tage ist die Wahrscheinlichkeits-Verteilung gleichmäßig.

Wie groß muss die Zahl n sein, damit es sich lohnt darauf zu wetten, dass wenigstens zwei Personen am selben Tag Geburtstag haben.

Lösung: Wir definieren $T := \{1, \cdots, 365\}$. Unser Zufalls-Experiment besteht darin, dass wir eine Liste der Länge n mit Elementen aus T auswählen. Die Menge Ω ist also durch

$$\Omega = \big\{[x_1, \cdots, x_n] \bigm| \forall i \in \{1, \cdots, n\} : x_i \in T\big\} = T^n$$

gegeben. Nach dem oben gezeigten gilt also

$$|\Omega| = |T|^n = 365^n.$$

Das uns interessierende Ereignis A_n, dass wenigstens zwei Personen am selben Tag Geburtstag haben, läßt sich nicht unmittelbar angeben. Wir betrachten daher zunächst das komplementäre Ereignis A_n^c, das ausdrückt, dass alle n Personen an verschiedenen Tagen Geburtstag haben. Es gilt

$$A_n^c = \big\{[x_1, \cdots, x_n] \bigm| \big(\forall i \in \{1, \cdots, n\} : x_i \in T\big) \ \wedge\ \big(\forall i, j \in \{1, \cdots, n\} : i \neq j \to x_i \neq x_j \in T\big)\big\}$$

Bei der Menge A_n^c handelt es sich gerade um die Menge aller n-Permutationen der Menge $T = \{1, \cdots, 365\}$. Daher gilt

$$\big|A_n^c\big| = \frac{365!}{(365-n)!}.$$

Da wir von einer gleichmäßigen Wahrscheinlichkeits-Verteilung ausgehen, gilt für die Wahrscheinlichkeit des gesuchten Ereignisses A_n

$$P(A_n) = 1 - P(A_n^c) = 1 - \frac{|A_n^c|}{|\Omega|} = 1 - \frac{365!}{(365-n)! \cdot 365^n}.$$

Tragen wir für $n = 1, \cdots, 60$ die Wahrscheinlichkeiten $P(A_n)$, so erhalten wir die in Tabelle 2.1 gezeigten Werte. Abbildung 2.1 zeigt diese Werte graphisch. Wir erkennen, dass bereits bei einer Gruppe von 23 Personen die Wahrscheinlichkeit dafür, dass zwei Personen am selben Tag Geburtstag haben, größer als 0.5 ist. Die Wette lohnt sich also ab einer Gruppengröße von 23 Personen. $\square$

n	$P(A_n)$	n	$P(A_n)$	n	$P(A_n)$
1	0.0	2	0.0027397260	3	0.0082041659
4	0.0163559124	5	0.0271355737	6	0.0404624837
7	0.0562357031	8	0.0743352924	9	0.0946238339
10	0.1169481777	11	0.1411413783	12	0.1670247888
13	0.1944102752	14	0.2231025120	15	0.2529013198
16	0.2836040053	17	0.3150076653	18	0.3469114179
19	0.3791185260	20	0.4114383836	21	0.4436883352
22	0.4756953077	23	0.5072972343	24	0.5383442579
25	0.5686997040	26	0.5982408201	27	0.6268592823
28	0.6544614723	29	0.6809685375	30	0.7063162427
31	0.7304546337	32	0.7533475278	33	0.7749718542
34	0.7953168646	35	0.8143832389	36	0.8321821064
37	0.8487340082	38	0.8640678211	39	0.8782196644
40	0.8912318098	41	0.9031516115	42	0.9140304716
43	0.9239228557	44	0.9328853686	45	0.9409758995
46	0.9482528434	47	0.9547744028	48	0.9605979729
49	0.9657796093	50	0.9703735796	51	0.9744319933
52	0.9780045093	53	0.9811381135	54	0.9838769628
55	0.9862622888	56	0.9883323549	57	0.9901224593
58	0.9916649794	59	0.9929894484	60	0.9941226609

Tabelle 2.1: Die ersten 60 Werte für das Geburtstags-Problem.

Aufgabe 5: In einer Ameisen-Kolonie, in der 40 000 Ameisen leben, sind 1 000 Ameisen mit Farbe markiert worden. Ein Forscher fängt nun zufällig 200 dieser Ameisen. Berechnen Sie die Wahrscheinlichkeit, dass er genau k markierte Tiere fängt. Gehen Sie dabei von einer gleichmäßigen Wahrscheinlichkeits-Verteilung aus, nehmen Sie also an, dass die Wahrscheinlichkeit dafür, dass

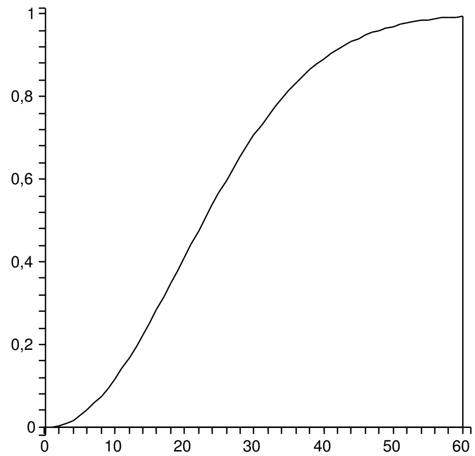

Abbildung 2.1: Das Geburtstags-Problem.

eine Ameise gefangen wird, unabhängig von der farblichen Markierung ist. Tabellieren Sie die Werte für $k = 0, \cdots, 20$ und tragen Sie mit Hilfe von *Maple* die gefundenen Werte in einem Diagramm auf.

Lösung: Wir modellieren die Ameisen-Kolonie durch die Menge $M := \{1, \cdots, 40\,000\}$. Die Menge der farblich markierten Ameisen bezeichnen wir mit F. Wir legen willkürlich fest, dass die ersten Tausend Ameisen diejenigen Ameisen sind, die markiert sind. Also gilt

$$F = \{1, \cdots, 1\,000\}$$

Das der Aufgabe zugrunde liegende Zufalls-Experiment besteht darin, dass wir aus der Menge M eine Teilmenge der Größe 200 auswählen. Damit läßt sich der Ergebnis-Raum als

$$\Omega = \big\{A \in 2^M \;\big|\; |A| = 200\big\}$$

definieren. Ω ist also die Menge aller Teilmengen von M, die genau 200 Elemente enthalten. Nach Gleichung (2.6) gilt für die Mächtigkeit dieser Menge

$$|\Omega| = |C(M, 200)| = \binom{|M|}{200} = \binom{40\,000}{200}$$

Das uns interessierende Ereignis Λ_k wird dann durch

$$\Lambda_k = \big\{B \in 2^M \;\big|\; |B| = 200 \wedge |B \cap F| = k\big\}$$

gegeben: Λ_k besteht aus genau den Teilmengen von M, die einerseits 200 Elemente enthalten und die andererseits genau k Elemente aus der Menge $\{1, \cdots, 1\,000\}$ enthalten. Die Berechnung der Wahrscheinlichkeit des Ereignisses Λ_k läuft auf die Berechnung von $|\Lambda_k|$ hinaus. Dazu formen wir die Definition von Λ_k etwas um. Eine Menge B liegt genau dann in Λ_k, wenn B insgesamt k Elemente aus der Menge F und $200 - k$ Elemente aus der Menge $M \backslash F$ enthält. Folglich gilt

$$\begin{aligned}
|\Lambda_k| &= \Big|\big\{C \cup D \mid C \subseteq F \wedge D \subseteq M\backslash F \wedge |C| = k \wedge |D| = 200 - k\big\}\Big| \\
&= \Big|\big\{C \cup D \mid C \in \{C' \in 2^F \mid |C'| = k\} \;\wedge\; D \in \{D' \in 2^{M\backslash F} \mid |D'| = 200 - k\}\big\}\Big| \\
&= \Big|\{C' \in 2^F \mid |C'| = k\}\Big| \cdot \Big|\{D' \in 2^{M\backslash F} \mid |D'| = 200 - k\}\Big| \\
&= \binom{|F|}{k} \cdot \binom{|M\backslash F|}{200 - k} \qquad \text{nach Gleichung (2.6)} \\
&= \binom{1000}{k} \cdot \binom{39\,000}{200 - k}.
\end{aligned}$$

Damit haben wir für die Wahrscheinlichkeit die Formel

$$P(\Lambda_k) = \frac{|\Lambda_k|}{|\Omega|} = \frac{\binom{1000}{k} \cdot \binom{39\,000}{200-k}}{\binom{40\,000}{200}}$$

gefunden. In Tabelle 2.2 sind die Wahrscheinlichkeiten aufgelistet. Für $k \geq 17$ sinkt die Wahrscheinlichkeit unter 10^{-5} und ist damit vernachlässigbar.

k	$P(\Lambda_k)$	n	$P(\Lambda_k)$	k	$P(\Lambda_k)$
0	0.0062425837	1	0.0321774371	2	0.0824301154
3	0.1399249245	4	0.1770598037	5	0.1781466648
6	0.1484517284	7	0.1053817150	8	0.0650519877
9	0.0354730483	10	0.0173006288	11	0.0076226006
12	0.0030592431	13	0.0011261811	14	0.0003825168
15	0.0001204896	16	0.0000353530	17	0.0000096999

Tabelle 2.2: Wie viele bunte Ameisen werden gefangen?

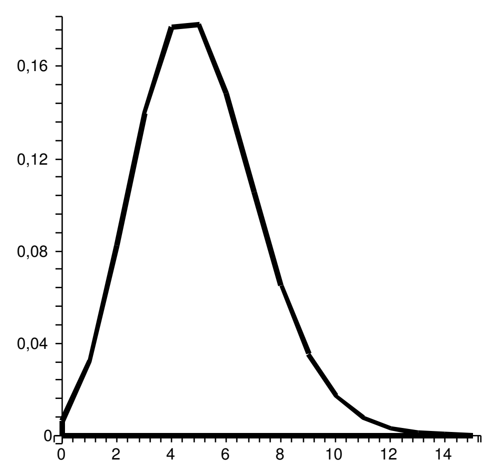

Abbildung 2.2: Wahrscheinlichkeits-Verteilung der Anzahl der gefärbten Ameisen

2.4 Die hypergeometrische Verteilung

Die in der letzten Aufgabe behandelte Situation kommt in der Praxis häufig vor. Ein Beispiel liefert die Überwachung der Qualität der Produktion von elektronischen Bauteilen. Eine Fertigungsanlage produziert am Tag eine bestimmte Zahl N solcher Bauteile. Davon sind K Bauteile defekt, während die restlichen $N - K$ Bauteile funktionieren. Zur Überprüfung der Produktions-Qualität wird eine Stichprobe vom Umfang n genommen. In dieser Stichprobe findet man dann k defekte Bauteile, während die restlichen $n - k$ Bauteile einwandfrei sind. Falls die Zahlen N, K und n gegeben sind, können wir uns fragen wie wahrscheinlich es ist, dass unter den n Bauteilen der Stichprobe insgesamt k Bauteile defekt sind. Die Situation ist dann die selbe wie in der letzten Aufgabe:

1. Die Gesamtzahl N entspricht der Anzahl aller Ameisen.

2. Die Anzahl K der defekten Bauteile entspricht der Zahl der gefärbten Ameisen.

3. Die Umfang n der Stichprobe entspricht der Zahl der am zweiten Tag gefangenen Ameisen.

4. Die Anzahl k der defekten Bauteile in der Stichprobe entspricht der Anzahl der am zweiten Tag gefangenen Ameisen, die gefärbt sind.

Zur Berechnung der Wahrscheinlichkeits-Verteilung gehen wir daher wie bei der Lösung der Aufgabe vor und definieren die Menge $\mathcal{M} := \{1, \cdots, N\}$. Jeder Zahl dieser Menge entspricht also genau ein Bauteil. Die Menge der defekten Bauteile bezeichnen wir mit $\mathcal{F}$. Wir gehen ohne Beschränkung der Allgemeinheit davon aus, dass die defekten Bauteile in der Aufzählung aller Bauteile am Anfang stehen, die Bauteile mit den Nummern $1, \cdots, K$ sind also defekt. Also gilt

$$\mathcal{F} = \{1, \cdots, K\}$$

Das der Aufgabe zugrunde liegende Zufalls-Experiment besteht darin, dass wir aus der Menge M eine Teilmenge der Größe n auswählen. Damit läßt sich der Ergebnis-Raum als

$$\Omega = \left\{A \in 2^{\mathcal{M}} \;\middle|\; |A| = n\right\}$$

Nach Gleichung (2.6) gilt für die Mächtigkeit dieser Menge

$$|\Omega| = |C(\mathcal{M}, n)| = \binom{|\mathcal{M}|}{n} = \binom{N}{n}$$

Das uns interessierende Ereignis Λ_k wird dann durch

$$\Lambda_k = \left\{B \in 2^{\mathcal{M}} \;\middle|\; |B| = n \wedge |B \cap \mathcal{F}| = k\right\}$$

gegeben: Λ_k besteht aus genau den Teilmengen von $\mathcal{M}$, die n Elemente enthalten von denen k defekt sind. enthalten. Die Berechnung der Wahrscheinlichkeit des Ereignisses Λ_k läuft auf die Berechnung von $|\Lambda_k|$ hinaus. Dazu formen wir die Definition von Λ_k etwas um. Eine Menge B liegt genau dann in Λ_k, wenn B insgesamt k Elemente aus der Menge $\mathcal{F}$ und $n-k$ Elemente aus der Menge $\mathcal{M}\backslash\mathcal{F}$ enthält. Folglich gilt

$$\begin{aligned}
|\Lambda_k| &= \Big|\big\{C \cup D \mid C \subseteq \mathcal{F} \wedge D \subseteq \mathcal{M}\backslash\mathcal{F} \wedge |C| = k \wedge |D| = n-k\big\}\Big| \\
&= \Big|\big\{C \cup D \mid C \in \{C' \in 2^{\mathcal{F}} \mid |C'| = k\} \;\wedge\; D \in \{D' \in 2^{\mathcal{M}\backslash\mathcal{F}} \mid |D'| = n-k\}\big\}\Big| \\
&= \Big|\{C' \in 2^{\mathcal{F}} \mid |C'| = k\}\Big| \cdot \Big|\{D' \in 2^{\mathcal{M}\backslash\mathcal{F}} \mid |D'| = n-k\}\Big| \\
&= \binom{|\mathcal{F}|}{k} \cdot \binom{|\mathcal{M}\backslash\mathcal{F}|}{n-k} \qquad \text{nach Gleichung (2.6)} \\
&= \binom{n}{k} \cdot \binom{N-K}{n-k}.
\end{aligned}$$

Damit haben wir für die Wahrscheinlichkeits-Verteilung die Formel

$$P(\Lambda_k) = \frac{|\Lambda_k|}{|\Omega|} = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}} \tag{2.8}$$

gefunden. In der Statistik wird das obige Beispiel abstrahiert. Statt von Bauteilen sprechen wir hier von Kugeln in einer Urne. Von diesen Kugeln sind dann K Kugeln schwarz, was den defekten Bauteilen entspricht, die restlichen Kugeln sind weiß. Dann gibt die obige Formel die Wahrscheinlichkeit dafür an, dass sich bei einer Entnahme von n Kugeln k Kugeln schwarz sind. Die in Gleichung 2.8 angegebene Wahrscheinlichkeits-Verteilung wird in der Literatur als hypergeometrische Verteilung bezeichnet.

Aufgabe 6: Eine Firma erhält eine Lieferung von 100 Geräten. Der zuständige Prüfer wählt zufällig 10 Geräte aus. Die Lieferung wird genau dann akzeptiert, wenn dabei kein defektes Gerät gefunden wird. Nehmen Sie an, dass von den gelieferten 100 Geräten 10 Geräte defekt sind. Wie hoch ist die Wahrscheinlichkeit dafür, dass die Lieferung trotzdem akzeptiert wird?

2.5 Die Binomial-Verteilung

Wir greifen das Beispiel aus der Einleitung zur Ameisen-Zählung wieder auf und modifizieren es mit dem Ziel, die Durchführung zu vereinfachen:

1. Der erste Tag ist unproblematisch, denn es ist nicht erforderlich, 1 000 Ameisen zu fangen. Es reicht ja aus, wenn wir insgesamt 1 000 Ameisen farblich markieren. Direkt nach der Markierung kann die Ameise wieder freigelassen werden.

2. Der zweite Tag ist schwieriger, denn hier müssten wir tatsächlich erst 200 Ameisen einsammeln bevor wir mit dem Zählen der markierten Ameisen beginnen. Also ändern wir das Experiment so ab, dass wir nacheinander 200 Ameisen untersuchen und jedesmal überprüfen, ob die Ameise markiert ist.

Am zweiten Tag kann es durchaus passieren, dass wir die selbe Ameise mehrmal zählen. Dadurch ändert sich natürlich auch die Wahrscheinlichkeits-Verteilung. Wenn wir berechnen wollen, mit welcher Wahrscheinlichkeit wir nun k gefärbte Ameisen finden, werden die Dinge erfreulicherweise einfacher. Wir behandeln gleich den allgemeinen Fall und nehmen folgendes an:

1. Für jede einzelne Ameise hat die Wahrscheinlichkeit, dass die Ameise gefärbt ist, den Wert
$$p = \frac{K}{N}.$$
Hier bezeichnet N die Gesamtzahl der Ameisen und K ist die Anzahl der insgesamt gefärbten Ameisen. Die Menge $\mathcal{M}$ der Ameisen hat also die Form
$$\{a_1, a_2, \cdots, a_N\}$$
Ohne Beschränkung der Allgemeinheit gehen wir davon aus, dass die Ameisen a_1 bis a_K gefärbt sind.

2. Das Zufalls-Experiment besteht darin, dass wir n Ameisen auf ihre Färbung untersuchen. Dabei erhalten wir als Ergebnis eine Liste $[x_1, x_2, \cdots, x_n]$ von Ameisen. Stellen wir die Ameisen durch natürliche Zahlen von 1 bis N dar, so gilt also $x_i \in \{a_1, \cdots, a_N\}$.

Da Listen der Länge n nichts anderes als n-Tupel sind, ist der Ergebnis-Raum unseres Zufalls-Experiments also das n-fache kartesische Produkt der Menge $\mathcal{M}$:
$$\Omega = \mathcal{M}^n.$$
Nach Gleichung (2.4) folgt daraus
$$|\Omega| = |\mathcal{M}|^n = N^n.$$
Wir definieren nun Λ_k als die Menge aller n-Tupel, die aus k gefärbten Ameisen bestehen. Wir denken uns ein solches n-Tupel aus zwei Teilen bestehend: einem k-Tupel von gefärbten Ameisen und einem $(n-k)$-Tupel von ungefärbten Ameisen. Es gibt insgesamt K^k solcher k-Tupel und $(N-K)^{n-k}$ solcher $(n-k)$-Tupel. Als nächstes überlegen wir uns, wieviele Möglichkeiten es gibt, aus einem k-Tupel und einem $(n-k)$-Tupel ein n-Tupel zu erzeugen. Betrachten wir zunächst ein konkretes Beispiel: Um aus dem 3-Tupel $[x_1, x_2, x_3]$ und dem 2-Tupel $[y_1, y_2]$ ein 5-Tupel zu erstellen müssen wir die Menge I der Indizes festlegen, an denen wir die Elemente x_i einfügen. Setzen wir beispielsweise $I := \{1, 2, 3\}$, so würden die Elemente x_i am Anfang stehen und wir hätten
$$I = \{1, 2, 3\}: \qquad [x_1, x_2, x_3, y_1, y_2].$$
Für $I = \{1, 3, 4\}$ würde sich
$$I = \{1, 3, 4\}: \qquad [x_1, y_1, x_2, x_3, y_2].$$
ergeben. Jede solche Index-Menge $I \subseteq \{1, \cdots, 5\}$ legt also eindeutig fest, wie wir aus den beiden Tupeln ein 5-Tupel bilden können.

Im allgemeinen Fall gilt einerseits $I \subseteq \{1, \cdots, n\}$ und andererseits $|I| = k$. Damit ist I dann eine k-elementige Teilmenge einer n-elementigen Menge. Nach Gleichung (2.6) gilt also
$$|I| = \binom{n}{k}.$$
Damit ergibt sich für die Mächtigkeit des Ereignisses Λ_k der folgende Ausdruck
$$|\Lambda_k| = |I| \cdot K^k \cdot (N-K)^{n-k} = \binom{n}{k} \cdot K^k \cdot (N-K)^{n-k}.$$

Die Wahrscheinlichkeit, dass wir k gefärbte Ameisen zählen, ergibt sich jetzt zu

$$
\begin{aligned}
P(k) := P(\Lambda_k) &= \frac{|\Lambda_k|}{|\Omega|} \\
&= \frac{\binom{n}{k} \cdot K^k \cdot (N-K)^{n-k}}{N^n} \\
&= \binom{n}{k} \cdot \frac{K^k}{N^k} \cdot \frac{(N-K)^{n-k}}{N^{n-k}} \\
&= \binom{n}{k} \cdot \left(\frac{K}{N}\right)^k \cdot \left(1 - \frac{K}{N}\right)^{n-k} \\
&= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad\qquad \text{mit} \quad p := \frac{K}{N}
\end{aligned}
$$

Wir abstrahieren nun von den Ameisen und fassen unsere Ergebnisse wie folgt zusammen. Ist ein Zufalls-Experiment dadurch gekennzeichnet, dass n mal ein Experiment durchgeführt wird, bei dem es nur zwei mögliche Ergebnisse gibt, die wir jetzt mit a und b bezeichnen und hat die Wahrscheinlichkeit für das Auftreten von a bei jedem solchen Experiment den selben Wert p, dann ist die Wahrscheinlichkeit dafür, dass bei einer n-fachen Durchführung dieses Experiments k mal das Ergebnis a auftritt, durch

$$
P(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \tag{2.9}
$$

gegeben. Diese Wahrscheinlichkeits-Funktion bezeichnen wir als *Binomial-Verteilung* und definieren

$$
B(n,p;k) := \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.
$$

2.6 Praktische Berechnung von Fakultät und Binomial-Koeffizienten

In der Wahrscheinlichkeits-Rechnung verfolgen uns die Fakultät und die Binomial-Koeffizienten auf Schritt und Tritt. Für kleine Werte von n können wir $n!$ und $\binom{n}{k}$ problemlos mit dem Taschenrechner ausrechnen. Aber schon bei der Lösung von Aufgabe 5 stößt der Taschenrechner an seine Grenzen, denn der Ausdruck

$$\binom{40\,000}{200},$$

der bei der Lösung dieser Aufgabe im Nenner auftritt, liefert eine ganze Zahl mit 545 Stellen. Solche Zahlen sind mit einem gewöhnlichen Taschenrechner nicht mehr darstellbar. Wir stellen daher in diesem Abschnitt eine Näherungs-Formel für die Fakultät und für den Binomial-Koeffizienten vor.

Wir beginnen mit einer Approximation der Fakultät. Die klassische Näherung von Stirling (James Stirling; 1692 - 1770) lautet

$$n! \approx \sqrt{2 \cdot \pi \cdot n} \cdot \left(\frac{n}{e}\right)^n.$$

Genauer ist die Formel von Lanzcos (Cornelius Lanzcos; 1893 - 1974), sie lautet:

$$n! \approx \sqrt{2 \cdot \pi} \cdot \left(\frac{n + \frac{1}{2}}{e}\right)^{n+\frac{1}{2}}.$$

Die Formel auf der rechten Seite approximiert die Fakultät in dem folgenden Sinne: Es gilt

$$\lim_{n \to \infty} \frac{\sqrt{2 \cdot \pi} \cdot \left(\frac{n+\frac{1}{2}}{e}\right)^{n+\frac{1}{2}}}{n!} = 1.$$

Für eine Herleitung dieser Formeln bleibt uns leider nicht die Zeit. Oft sind die Fakultäten so groß, dass Sie nicht mehr auf dem Taschenrechner dargestellt werden können. Sie treten dann meist in Brüchen auf und der zu berechnende Bruch ist durchaus noch darstellbar. Dann kann es hilfreich sein, zum Logarithmus überzugehen. Um beispielsweise $\binom{n}{k}$ zu berechnen, gehen wir wie folgt vor:

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k! \cdot (n-k)!} \\
&= \exp\left(\ln\left(\frac{n!}{k! \cdot (n-k)!}\right)\right) \\
&= \exp\Bigl(\ln\bigl(n!\bigr) - \ln\bigl(k!\bigr) - \ln\bigl((n-k)!\bigr)\Bigr) \qquad (2.10)
\end{aligned}$$

Zu Berechnung der verschiedenen Fakultäten in dieser Formel wenden wir auf die Näherungsformel von Lanzcos den Logarithmus an und erhalten

$$\ln\bigl(n!\bigr) \approx \frac{1}{2} \cdot \ln(2 \cdot \pi) + \left(n + \frac{1}{2}\right) \cdot \left(\ln\left(n + \frac{1}{2}\right) - 1\right).$$

Setzen wir diesen Wert in Gleichung (2.10) ein, so können wir den Binomial-Koeffizienten approximieren. Berechnen wir $\binom{100}{50}$ auf diese Weise, so erhalten wir

$$\binom{100}{50} \approx 1.007667751 \cdot 10^{29}.$$

Der exakte Wert ist

$$\binom{100}{50} = 1.008913445 \cdot 10^{29}$$

und der relative Fehler liegt bei 1.2‰. Bei der Berechnung von Wahrscheinlichkeiten sind solche Genauigkeiten meistens ausreichend.

Noch einfacher können Binomial-Koeffizienten über die Formel

$$\binom{n}{k} \approx \sqrt{\frac{2}{\pi \cdot n}} \cdot 2^n \cdot \exp\left(-\frac{\left(k - \frac{1}{2}n\right)^2}{\frac{1}{2}n}\right).$$

berechnet werden. Diese Formel geht auf de Moivre (Abraham de Moivre; 1667 - 1754) zurück. Diese Näherung liefert brauchbare Werte sobald die Bedingung $n > 36$ erfüllt ist. Die Werte sind am genauesten für $k \approx \frac{n}{2}$. Berechnen wir mit dieser Formel eine Näherung für $\binom{100}{50}$, so erhalten wir

$$\binom{100}{50} \approx 1.008913445 \cdot 10^{30}.$$

Diesmal liegt der relative Fehler bei 2.5‰ und ist damit doppelt so hoch wie bei der Verwendung der Formel von Lanzcos.

Der Binomial-Koeffizient tritt oft in Termen der Form

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

auf. Für diesen Fall gibt es noch eine Approximations-Formel, die einfacher zu handhaben ist. Die Formel lautet

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \frac{1}{\sqrt{2 \cdot \pi \cdot n \cdot p \cdot (1-p)}} \cdot \exp\left(-\frac{(k - n \cdot p)^2}{2 \cdot n \cdot p \cdot (1-p)}\right)$$

und geht auf Laplace (Pierre Simon Laplace; 1749 - 1827) zurück. Diese Näherung liefert brauchbare Werte sobald die Bedingung

$$n \cdot p \cdot (1-p) > 9$$

erfüllt ist. Für $p = \frac{1}{2}$ geht diese Formel in die von de Moivre angegebene Formel über. Die Werte, die mit dieser Formel berechnet werden, sind am brauchbarsten für die Werte von k, die in der Nähe von $n \cdot p$ liegen.

Am einfachsten ist es, den Binomial-Koeffizienten mit Hilfe von *Maple* zu berechnen, denn *Maple* ist in der Lage, auch sehr große Binomial-Koeffizienten fehlerfrei zu berechnen.

Aufgabe 7: Berechnen Sie den kleinsten Wert von n, für den die Ungleichung

$$\binom{n}{\frac{2}{3} \cdot n} \leq \frac{1}{100} \cdot \binom{n}{\frac{1}{2} \cdot n}$$

erfüllt ist. Benutzen Sie die Formel von de Moivre zur Approximation der Binomial-Koeffizienten.

Lösung: Approximieren wir die Binomial-Koeffizienten durch die Formel von de Moivre, so erhalten wir die Ungleichung

$$\sqrt{\frac{2}{\pi \cdot n}} \cdot 2^n \cdot \exp\left(-\frac{\left(\frac{2}{3} - \frac{1}{2}\right)^2 n^2}{\frac{1}{2}n}\right) \leq \frac{1}{100} \cdot \sqrt{\frac{2}{\pi \cdot n}} \cdot 2^n \cdot \exp(0)$$

Das ist offenbar äquivalent zu

$$\exp\left(-\frac{1}{18}n\right) \leq \frac{1}{100}.$$

Logarithmieren wir diese Ungleichung, so erhalten wir

$$-\frac{1}{18}n \leq -\ln(100) \quad \text{bzw.} \quad n \geq 18 \cdot \ln(100) \approx 82.89306335.$$

Der Binomial-Koeffizient $\binom{n}{k}$ ist nur für ganzzahlige Werte von n und k definiert. Damit der Ausdruck

$$\binom{n}{\frac{2}{3}n}$$

überhaupt existiert muss folglich n durch 3 teilbar sein. Die kleinste natürliche Zahl, die größer als 82.89306335 ist und durch 3 geteilt werden kann, ist 84. Folglich lautet die Lösung $n = 84$.

Die letzte Aufgabe ist relativ künstlich. Normalerweise interessiert uns nicht eine einzelne Wahrscheinlichkeit der Form

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k},$$

sondern wir wollen vielmehr wissen, welchen Wert ein Ausdruck der Form

$$\sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

annimmt, denn dieser Ausdruck gibt die Wahrscheinlichkeit dafür an, dass bei einer Binomial-Verteilung bei einer n-fachen Durchführung eines Experiments ein Ergeignis höchstens k-mal auftritt. Wir wollen daher jetzt eine Näherung für diese Summe herleiten. Dazu definieren wir zunächst

$$F_p^n(k) := \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}.$$

Die Funktion $F_p^n(k)$ bezeichnen wir auch als die *Verteilungsfunktion*. Um die nachfolgende Rechnung zu vereinfachen, definieren wir zur Abkürzung

$$\mu := n \cdot p, \quad q := 1-p \quad \text{und} \quad \sigma := \sqrt{n \cdot p \cdot q}.$$

Allgemein läßt sich eine Summe der Form $\sum_{i=0}^{k} f(i)$ wie folgt durch ein Integral approximieren:

$$\sum_{i=a}^{b} f(i) \approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} f(t)\,\mathrm{d}t$$

Wir approximieren nun in der Definition der Verteilungsfunktion $F_p^n(k)$ einerseits den Binomial-Koeffizienten durch die Formel von Laplace und andererseits nähern wir die Summe durch ein Integral an. Dann erhalten wir

$$\begin{aligned}
F_p^n(k) &= \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^i \\
&\approx \sum_{i=0}^{k} \frac{1}{\sqrt{2\,\pi\,n\,p\,(1-p)}} \cdot \exp\left(-\frac{(i-n\cdot p)^2}{2\,n\,p\,(1-p)}\right) \\
&= \sum_{i=0}^{k} \frac{1}{\sqrt{2\,\pi}\cdot\sigma} \cdot \exp\left(-\frac{(i-\mu)^2}{2\,\sigma^2}\right) \\
&\approx \int_{-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\,\pi}\cdot\sigma} \cdot \exp\left(-\frac{(u-\mu)^2}{2\,\sigma^2}\right)\,\mathrm{d}u \\
&= \frac{1}{\sqrt{2\,\pi}\cdot\sigma} \cdot \int_{-\frac{1}{2}}^{k+\frac{1}{2}} \exp\left(-\frac{(u-\mu)^2}{2\,\sigma^2}\right)\,\mathrm{d}u
\end{aligned}$$

Das hier auftretende Integral vereinfachen wir, indem wir die Variablen-Substitution

$$t(u) = \frac{u-\mu}{\sigma}, \quad \text{also } \mathrm{d}t = \tfrac{1}{\sigma}\,\mathrm{d}u \quad \sigma\,\mathrm{d}t = \mathrm{d}u$$

durchführen. Setzen wir hier für u die Grenzen $-\frac{1}{2}$ bzw. $k+\frac{1}{2}$ ein, so erhalten wir

$$\begin{aligned}
F_p^n(k) &= \frac{1}{\sqrt{2\,\pi}\cdot\sigma} \cdot \int_{(-\frac{1}{2}-\mu)/\sigma}^{(k+\frac{1}{2}-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right)\sigma\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\,\pi}} \cdot \int_{(-\frac{1}{2}-\mu)/\sigma}^{(k+\frac{1}{2}-\mu)/\sigma} \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t
\end{aligned}$$

Um an dieser Stelle weiter zu kommen, benötigen wir die Gauß'sche Integralfunktion $\Phi(x)$. Diese Funktion wird wie folgt definiert:

$$\Phi(x) = \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t$$

Damit haben für die Verteilungsfunktion $F_p^n(k)$ die Näherung

$$F_p^n(k) = \Phi\left(\frac{k+\frac{1}{2}-\mu}{\sigma}\right) - \Phi\left(\frac{-\frac{1}{2}-\mu}{\sigma}\right)$$

gefunden. Berücksichtigen wir noch, dass $\mu = n \cdot p$ und $\sigma = \sqrt{n\,p\,q}$ gilt und bedenken, dass wir eine Näherung für große Werte von n suchen, so gilt

$$\Phi\left(\frac{-\frac{1}{2}-\mu}{\sigma}\right) = \Phi\left(\frac{-\frac{1}{2}-n\,p}{\sqrt{n\,p\,q}}\right) \underset{n\to\infty}{\longrightarrow} \Phi\left(-\infty\right) = 0$$

Damit lautet unsere Näherung für die Verteilungsfunktion $F_p^n(k)$

$$F_p^n(k) \approx \Phi\left(\frac{k+\frac{1}{2}-n\,p}{\sqrt{n\,p\,q}}\right).$$

In dieser Formel wird häufig die sogenannte *Stetigkeits-Korrektur* $\frac{1}{2}$ weggelassen, denn gegenüber dem Term $n\,p$ fällt diese kaum ins Gewicht. Wir werden also im folgenden die Näherung

$$\boxed{F_p^n(k) \approx \Phi\left(\frac{k-n\,p}{\sqrt{n\,p\,q}}\right)}$$

verwenden. Die Gauß'sche Integralfunktion liegt tabelliert vor. In *Maple* steht diese Funktion nicht unmittelbar zu Verfügung. Stattdessen gibt es dort die Gauß'sche *Fehlerfunktion erf*(x), die wie folgt definiert ist:

$$\mathit{erf}(x) := \frac{2}{\sqrt{\pi}} \cdot \int_0^x \exp\bigl(-u^2\bigr)\,\mathrm{d}u.$$

Wir können die Gauß'sche Integralfunktion auf die Gauß'sche Fehlerfunktion zurückführen, indem wir in dem Integral die Variablen-Substitution $u = \frac{1}{\sqrt{2}}t$, also $t^2 = 2\,u^2$ durchführen. Wegen $\mathrm{d}u = \frac{1}{\sqrt{2}}\,\mathrm{d}t$, also $\mathrm{d}t = \sqrt{2}\,\mathrm{d}u$ erhalten wir

$$\begin{aligned}
\Phi(x) &= \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{0} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t \;+\; \frac{1}{\sqrt{2\,\pi}} \cdot \int_{0}^{x} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t \\
&= \frac{1}{\sqrt{2\,\pi}} \cdot \sqrt{\frac{\pi}{2}} \;+\; \frac{1}{\sqrt{2\,\pi}} \cdot \sqrt{2} \cdot \int_0^{x/\sqrt{2}} \exp\left(-u^2\right)\,\mathrm{d}u \\
&= \frac{1}{2} \;+\; \frac{1}{2} \cdot \mathit{erf}\left(\frac{x}{\sqrt{2}}\right)
\end{aligned}$$

Dabei haben wir ausgenutzt, dass

$$\int_{-\infty}^{0} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t = \sqrt{\frac{\pi}{2}}$$

gilt. Damit sind wir jetzt in der Lage, auch interessantere Aufgaben aus der Wahrscheinlichkeits-Rechnung numerisch zu lösen.

Aufgabe 8: Eine Spielbank bietet das folgende Spiel an: Falls am Roulette-Tisch eine Zahl aus der Menge $\{19, \cdots, 36\}$ kommt, erhalten Sie einen Euro, andernfalls, wenn also eine Zahl aus der Menge $\{0, \cdots, 18\}$ kommt, müssen Sie einen Euro zahlen. Angenommen, Sie spielen dieses Spiel n-mal. Wie groß muss n sein damit die Wahrscheinlichkeit, dass die Bank einen Verlust macht, höchstens 1% beträgt?

Hinweis: Es gilt $\Phi(2.326347874) \approx 0.99$, also $\Phi^{-1}(0.99) \approx 2.326347874$. (Hier steht Φ^{-1} für die

Umkehrfunktion von Φ und <u>nicht</u> für die Funktion $\frac{1}{\Phi}$!

Lösung: Die Wahrscheinlichkeit, dass die Bank bei einem einzelnen dieser Spiele verliert, ist

$$p = \frac{18}{37}$$

Die Bank macht keinen Verlust, wenn bei den n Spielen mindestens $\frac{n}{2}$-mal eine Zahl kleiner als 19 kommt. Wir nehmen zunächst an, dass n gerade ist. Die Bank macht dann keinen Verlust, wenn Sie höchstens $\frac{n}{2}$ Spiele verliert. Die Wahrscheinlichkeit dafür ist durch die Summe

$$\sum_{k=0}^{\frac{n}{2}} \binom{n}{k} p^k \cdot (1-p)^{n-k} = F_p^n\left(\frac{n}{2}\right) \approx \Phi\left(\frac{\frac{n}{2} - n\,p}{\sqrt{n\,p\,q}}\right) = \Phi\left(\frac{\sqrt{n}\left(\frac{1}{2} - p\right)}{\sqrt{p\,q}}\right).$$

gegeben, wobei $q = 1 - p$ gilt. Die Forderung, dass diese Wahrscheinlichkeit mindestens 99% betragen soll, führt auf die Ungleichung

$$\Phi\left(\frac{\sqrt{n}\left(\frac{1}{2} - p\right)}{\sqrt{p\,q}}\right) \geq 0.99.$$

Damit erhalten wir

$$\frac{\sqrt{n}\left(\frac{1}{2} - p\right)}{\sqrt{p\,q}} \geq \Phi^{-1}(0.99) = \alpha.$$

Daraus erhalten wir sofort

$$n \geq \frac{p \cdot q}{\left(\frac{1}{2} - p\right)^2} \cdot \alpha^2 \approx 7403.471598$$

Um richtig sicher zu gehen, dass die Bank am Schluss gewinnt, sollten Sie also mindesten 7404 mal spielen.

Aufgabe 9: Angenommen, wir würfeln n mal mit einem Laplace-Würfel. Wie gross muss n gewählt werden, damit wir mit einer Wahrscheinlichkeit von 99% davon ausgehen können, dass mindestens $\frac{n}{7}$ mal eine Sechs gewürfelt wird?

2.7 Zufallsgrößen

Es wird Zeit, dass wir einen Begriff formal definieren, der uns in verschiedenen Beispielen schon mehrfach begegnet ist.

Definition 2 (Zufallsgröße) Es sei ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^\Omega, P\rangle$ gegeben. Eine Funktion

$$X : \Omega \to \mathbb{R}$$

bezeichnen wir als *Zufallsgröße*. □

Beispiel: Ein einfaches Beispiel für eine Zufallsgröße wäre die Summe der Augenzahlen, wenn mit zwei Würfeln gewürfelt wird. Der Ergebnis-Raum Ω ist in diesem Fall

$$\Omega = \big\{\langle i,j\rangle \big| i,j \in \{1,\cdots,6\}\big\}$$

und wenn wir davon ausgehen, dass es sich bei den Würfeln um Laplace-Würfel handelt, dann ist die Wahrscheinlichkeits-Verteilung $P : 2^\Omega \to \mathbb{R}$ durch die Formel

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{36} \cdot |A|$$

gegeben. Die Zufallsgröße $S : \Omega \to \mathbb{R}$ definieren wir durch

$$S\big(\langle i,j\rangle\big) := i + j.$$

Wenn uns die Wahrscheinlichkeit für das Auftreten einer bestimmten Augensumme interessiert, dann müssen wir zunächst die diesbezüglichen Ereignisse definieren. Bei diesen Ereignissen handelt es sich um die Mengen

$$\big\{\langle i,j\rangle \in \Omega \;\big|\; i+j = s\big\} \qquad \text{für } s = 2,\cdots,12.$$

Ist X eine Zufallsgröße, die auf einem Ergebnis-Raum Ω definiert ist, so vereinbaren wir zur Abkürzung die folgende Schreibweise:

$$P(X = x) := P\big(\{\omega \in \Omega \mid X(\omega) = x\}\big).$$

In unserem konkreten Beispiel schreiben wir also

$$P(S = s) = P\big(\{\langle i,j\rangle \in \Omega \mid i+j = s\}\big).$$

Für $s \leq 7$ gilt

$$\big\{\langle i,j\rangle \in \Omega \;\big|\; i+j = s\big\} = \big\{\langle k, s-k\rangle \mid k \in \{1,\cdots,s-1\}\big\}$$

Für $s > 7$ finden wir

$$\big\{\langle i,j\rangle \in \Omega \;\big|\; i+j = s\big\} = \big\{\langle 7-k, s-7+k\rangle \mid k \in \{1,\cdots,13-s\}\big\}$$

(Die Bedingung $k \leq 13 - s$ folgt dabei aus der Forderung $s - 7 + k \leq 6$.) Daraus ergibt sich

$$\Big|\big\{\langle i,j\rangle \in \Omega \;\big|\; i+j = s\big\}\Big| = \begin{cases} s-1 & \text{falls } s \leq 7; \\ 13-s & \text{sonst.} \end{cases}$$

Also haben wir

$$P(S = s) = \begin{cases} \dfrac{s-1}{36} & \text{falls } s \leq 7; \\[2ex] \dfrac{13-s}{36} & \text{sonst.} \end{cases}$$

Wir sehen hier, dass die Wahrscheinlichkeit für die verschiedenen möglichen Werte der Summe S nicht mehr gleichmäßig verteilt sind, obwohl die Wahrscheinlichkeits-Verteilung P auf dem zugrunde liegenden Wahrscheinlichkeits-Raum sehr wohl gleichmäßig ist. Die Ursache hierfür ist einfach einzusehen: Es gibt beispielsweise 6 Möglichkeiten, in der Summe eine 6 zu würfeln, aber es gibt nur eine einzige Möglichkeite um in der Summe ein 12 zu würfeln.

Aufgabe 10: Betrachten Sie allgemein den Fall, dass mit n Würfeln gewürfelt wird. Der Ergebnis-Raum Ω ist dann das n-fache kartesische Produkt der Menge $\{1,\cdots,6\}$:

$$\Omega = \{1,\cdots,6\}^n$$

Berechnen Sie mit Hilfe eines geeigneten Programms (in einer Programmiersprache ihrer Wahl) die Wahrscheinlichkeit dafür, dass die Augensumme aller Würfel den Wert s hat und erstellen Sie (beispielsweise mit *Maple*) den Graphen der Funktion

$$P_n : \{n, \cdots, 6 * n\} \to \mathbb{R}$$

die durch

$$P_n(s) = P\Big(\big\{\langle x_1, \cdots, x_n\rangle \in \Omega \mid s = \sum_{i=1}^{n} x_i\big\}\Big).$$

gegeben ist. Zeichnen Sie diesen Graphen für die Werte $n = 2$, $n = 3$, $n = 5$ und $n = 10$.
Hinweis: $P_n(s)$ können Sie am besten rekursiv berechnen, indem Sie P_n auf P_{n-1} zurück führen.

2.7.1 Erwartungswert einer Zufallsgröße

Definition 3 (Erwartungswert) Ist $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und ist

$$X : \Omega \to \mathbb{R}$$

eine Zufallsgröße, so definieren wir den *Erwartungswert* $E[X]$ als

$$E[X] := \sum_{\omega\in\Omega} P(\{\omega\}) \cdot X(\omega).$$

Hat der Wertebereich von X die Form

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} = \{x_n \mid n \in \mathbb{N}\},$$

so können wir den Erwartungswert auch durch die Formel

$$E[X] = \sum_{n=0}^{\infty} P(X = x_n) \cdot x_n$$

berechnen. Eine analoge Formel gilt, wenn die Menge $X(\Omega)$ endlich ist. □

Der Erwartungswert einer Zufallsgröße gibt den durchschnittlichen Wert an, der sich bei einer großen Zahl von Versuchen ergeben würde. Diese Aussage werden wir später noch näher präzisieren und dann auch beweisen können.

Beispiel: Wir berechnen den Erwartungswert der Augenzahl beim würfeln mit einem Würfel. Es gilt

$$\begin{aligned}
E[S] &= \sum_{s=1}^{6} P(S = s) \cdot s \\
&= \sum_{s=1}^{6} \frac{1}{6} \cdot s \;=\; \frac{1}{6} \cdot \sum_{s=1}^{6} s \\
&= \frac{1}{6} \cdot \frac{1}{2} \cdot 6 \cdot (6 + 1) \;=\; \frac{7}{2}
\end{aligned}$$

Als nächstes berechnen wir den Erwartungswert der Augensumme für das würfeln mit zwei Laplace-Würfeln. Hier gilt

$$\begin{aligned}
E[S] &= \sum_{s=2}^{12} P(S = s) \cdot s \\
&= \frac{1}{36} \sum_{s=2}^{7} (s - 1) \cdot s + \frac{1}{36} \sum_{s=8}^{12} (13 - s) \cdot s \\
&= \frac{252}{36} \;=\; 7
\end{aligned}$$

Dieses Ergebnis war auch zu erwarten, denn beim würfeln mit einem Würfel ist der Erwartungswert $\frac{7}{2}$ und der Erwartungswert der Augensumme beim würfeln mit zwei Würfeln sollte doppelt so groß sein. $\square$

Aufgabe 11: Beim *Mensch-ärger-dich-nicht* müssen Sie am Anfang eine 6 würfeln um das Spiel beginnen zu können. Berechnen Sie den Erwartungswert für die Anzahl der Würfe, die benötigt werden um eine 6 zu würfeln.

Lösung: Die Wahrscheinlichkeit, dass die 6 sofort beim ersten Wurf kommt, beträgt $\frac{1}{6}$, während die Wahrscheinlichkeit, dass beim ersten Wurf keine 6 gewürfelt wird, offenbar den Wert $\frac{5}{6}$ hat. Die Wahrscheinlichkeit dafür, dass im zweiten Wurf die 6 fällt, ist das Produkt aus der Wahrscheinlichkeit, dass im ersten Wurf keine 6 fällt und der Wahrscheinlichkeit, dass im zweiten Wurf eine sechs fällt und hat daher den Wert $\frac{5}{6} \cdot \frac{1}{6}$. Allgemein hat die Wahrscheinlichkeit dafür, dass im $n+1$-ten Wurf eine 6 fällt, den Wert

$$\left(\frac{5}{6}\right)^n \cdot \frac{1}{6}.$$

Bezeichen wir die Zufallsgröße, die die Anzahl der benötigten Würfe angibt, mit N, so erhalten wir für den Erwartungswert von N die Formel

$$E[N] = \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^n \cdot \frac{1}{6} \cdot (n+1) = \frac{1}{6} \cdot \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^n \cdot (n+1).$$

Diese Summe können wir mit einem Trick auf die geometrische Reihe zurück führen. Wir haben im letzten Semester gesehen, dass für alle $q \in]-1,1[$

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$$

gilt. Differenzieren wir diese Formel nach q, so erhalten wir die Gleichung

$$\sum_{n=1}^{\infty} n \cdot q^{n-1} = \frac{1}{(1-q)^2}.$$

Die Summe auf der linken Seite geht erst bei $n = 1$ los, denn das konstante Glied fällt beim Differenzieren weg. Ersetzen wir in dieser Summe n durch $n+1$, so haben wir

$$\sum_{n=0}^{\infty} (n+1) \cdot q^n = \frac{1}{(1-q)^2}.$$

Diese Summe hat aber genau die Form, die oben bei der Berechnung des Erwartungswerts auftritt. Damit gilt

$$E[N] = \frac{1}{6} \cdot \sum_{n=0}^{\infty} \left(\frac{5}{6}\right)^n \cdot (n+1) = \frac{1}{6} \cdot \frac{1}{\left(1-\frac{5}{6}\right)^2} = \frac{1}{6} \cdot \frac{1}{\left(\frac{1}{6}\right)^2} = 6.$$

Das Ergebnis ist intuitiv einleuchtend. $\square$

Der Erwartungswert gibt den mittleren Wert einer Zufallsgröße wieder. Damit wissen wir aber noch nichts darüber, wie weit die einzelnen Werte der Zufallsgröße um diesen Mittelwert streuen. Darüber gibt die *Varianz* Aufschluss.

Definition 4 (Varianz, Standard-Abweichung)
Ist $\langle \Omega, 2^{\Omega}, P \rangle$ ein Wahrscheinlichkeits-Raum und ist

$$X : \Omega \to \mathbb{R}$$

eine Zufallsgröße, so definieren wir die *Varianz* $\mathrm{Var}[X]$ als den Erwartungswert der Zufallsgröße $w \mapsto (X - E[X])^2$, also gilt

$$\mathrm{Var}[X] := E\big[(X - E[X])^2\big].$$

Hat der Wertebereich von X die Form

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\} = \{x_n \mid n \in \mathbb{N}\}$$

und setzen wir zur Abkürzung $\mu = E[X]$, so können wir die Varianz auch durch die Formel

$$\mathrm{Var}[X] = \sum_{n=0}^{\infty} P(X = x_n) \cdot (x_n - \mu)^2$$

berechnen. Eine analoge Formel gilt, wenn die Menge $X(\Omega)$ endlich ist.

Die *Standard-Abweichung* ist als die Quadrat-Wurzel aus der Varianz definiert

$$\sigma(X) := \sqrt{\mathrm{Var}[X]}.$$

Die Standard-Abweichung hat die selbe Einheit wie die Zufallsgröße X. $\Box$

Beispiel: Wir berechnen die Varianz der Zufallsgröße S, die die Augenzahl beim würfeln mit einem Würfel wiedergibt. Wegen $E[S] = \frac{7}{2}$ gilt

$$\begin{aligned} \mathrm{Var}[S] &= \sum_{s=1}^{6} P(S = s) \cdot \left(s - \frac{7}{2}\right)^2 \\ &= \sum_{s=1}^{6} \frac{1}{6} \cdot \left(s - \frac{7}{2}\right)^2 = \frac{1}{6} \cdot \sum_{s=1}^{6} \left(s - \frac{7}{2}\right)^2 \\ &= \frac{35}{12}. \end{aligned}$$

Damit gilt für die Standard-Abweichung

$$\sigma(X) = \sqrt{\frac{35}{12}} \approx 1.707825129.$$

Führen wir die selbe Rechnung für das Experiment "Würfeln mit zwei Würfeln" durch, so erhalten wir für die Varianz

$$\begin{aligned} \mathit{Var}[S] &= \sum_{s=2}^{12} P(S = s) \cdot (s - 7)^2 \\ &= \frac{1}{36} \sum_{s=2}^{7} (s - 1) \cdot (s - 7)^2 + \frac{1}{36} \sum_{s=8}^{12} (13 - s) \cdot (s - 7)^2 \\ &= \frac{35}{6} \end{aligned}$$

Wir sehen, dass die Varianz jetzt genau doppelt so groß ist wie beim würfeln mit einem Würfel. Diese Beobachtung werden wir später verallgemeinern und beweisen. $\Box$

2.8 Bedingte Wahrscheinlichkeiten

In diesem Abschnitt beantworten wir die folgende Frage: Es sei ein Wahrscheinlichkeits-Raum $\langle\Omega, 2^\Omega, P\rangle$ gegeben. Wir betrachten ein Ereignis $A \in 2^\Omega$. Angenommen wir erfahren nun, dass ein Ereignis B eingetreten ist. Wie verändert sich durch diese Information die Wahrscheinlichkeit für das Eintreten des Ereignisses A? Die Wahrscheinlichkeit für das Eintreten von A unter der Voraussetzung, dass B bereits eingetreten ist, bezeichnen wir mit

$$P(A|B).$$

Beispiel: Wir betrachten das Zufalls-Experiment "*Wurf eines Laplace-Würfels*" mit dem Ergebnis-Raum $\Omega = \{1, \cdots, 6\}$. Es sei A das Ereignis, dass eine 6 gewürfelt wird, also $A = \{6\}$. Da wir vorausgesetzt haben, dass es sich um einen Laplace-Würfel handelt, gilt

$$P(A) = \frac{1}{6}.$$

Es sei weiterhin B das Ereignis, dass die Augenzahl größer als 4 ist, also $B = \{5, 6\}$. Wir nehmen nun an, dass der Würfel geworfen wird und wir gesagt bekommen, dass das Ereignis B eingetreten ist. Das Ergebnis des Zufalls-Experiments ist uns allerdings nicht bekannt. Da dann nur noch zwei Möglichkeiten für das Ergebnis bleiben, nämlich die Zahlen 5 und 6, würden wir in dieser neuen Situation dem Ereignis A die Wahrscheinlichkeit $\frac{1}{2}$ zuordnen, also gilt

$$P(A|B) = \frac{1}{2}. \qquad \Box$$

Beispiel: Aus einer Sterbetafel, in der das Lebensalter von 100 000 Frauen verzeichnet ist, entnehmen wir die Information, dass 89 835 aller Frauen das Alter von mindestens 60 Jahren erreichen, während 57 062 ein Alter von mehr als 80 Jahren erreichen. Angenommen, eine Frau wird 60 Jahre alt. Mit welcher Wahrscheinlichkeit erreicht Sie dann ein Alter von 80 Jahren?

Wir bezeichnen das Ereignis, dass eine Frau das Alter von 60 Jahren erreicht, mit A, während das Ereignis, dass eine Frau das Alter von 80 Jahren erreicht, mit B bezeichnet wird. Offenbar ist B eine Teilmenge von A und es ist klar, dass der Anteil der sechzigjährigen Frauen, die auch noch das Alter von 80 Jahren erreichen, durch den Bruch $\frac{|B|}{|A|}$ gegeben wird. Also gilt

$$P(B|A) = \frac{|B|}{|A|} = \frac{57\,062}{89\,835} \approx 0.63519.$$

Damit beträgt also die Wahrscheinlichkeit dafür, dass eine sechzigjährige Frau das Alter von 80 Jahren erreicht, 63.5%. $\Box$

Wir verallgemeinern die obigen Beispiele. Wir gehen davon aus, dass ein Wahrscheinlichkeits-Raum

$$\langle\Omega, 2^\Omega, P\rangle$$

mit einer gleichmäßigen Wahrscheinlichkeits-Verteilung P gegeben ist, es gilt also

$$P(E) = \frac{|E|}{|\Omega|} \quad \text{für alle } E \in 2^\Omega.$$

Wir betrachten zwei Ereignisse A und B und berechnen die bedingte Wahrscheinlichkeit für das Eintreten von B, wenn wir bereits wissen, dass A eingetreten ist. Wenn A eingetreten ist, kommen für das Eintreten von B nur noch die Ergebnisse in Frage, die in $B \cap A$ liegen. Nehmen wir an, dass diese Ergebnisse nach wie vor die selbe Wahrscheinlichkeit haben, dann haben wir einen neuen Wahrscheinlichkeits-Raum, dessen Ereignis-Raum die Menge A ist. Folglich gilt

$$P(B|A) = \frac{|B \cap A|}{|A|} = \frac{\frac{|B\cap A|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{P(B \cap A)}{P(A)}. \tag{2.11}$$

Diese Gleichung für die bedingte Wahrscheinlichkeit gilt auch im allgemeinen Fall. Sind zwei Ereignisse A und B gegeben und führen wir das zugrunde liegende Zufalls-Experiment n-mal aus, so erwarten wir, dass für große n ein Ereignis E etwa $n \cdot P(E)$ mal eintritt. Ist das Ereignis A bereits eingetreten, so tritt das Ereignis B genau dann ein, wenn das Ereignis $B \cap A$ eintritt. Also ist die relative Häufigkeit für das Eintreten des Ereignisses B unter der Annahme, dass A bereits eingetreten ist, durch den Quotienten

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

gegeben und daher definieren wir die bedingte Wahrscheinlichkeit immer als diesen Quotienten.

Aufgabe 12: Eine Lieferung von Glühbirnen enthalte erfahrungsgemäß drei Arten von Glühbirnen:

1. Glühbirnen, die bereits defekt sind. Der Anteil dieser Glühbirnen betrage 10%.
2. Glühbirnen, die zwar funktionieren, aber nur eine Lebensdauer von ein Paar Tagen haben. Hier beträgt der Anteil 20%.
3. Glühbirnen, die voll funktionsfähig sind.

Angenommen, Sie testen eine Glühbirne und stellen fest, dass diese Birne noch nicht defekt ist. Wie hoch ist dann die Wahrscheinlichkeit dafür, dass diese Glühbirne voll funktionsfähig ist?

Lösung: Wir bezeichnen das Ereignis "*Glühbirne defekt*" mit D, das Ereignis "*Glühbirne hat kurze Lebensdauer*" mit K und das Ereignis "*Glühbirne voll funktionsfähig*" mit F. Gefragt ist dann nach der Wahrscheinlichkeit, dass die Glühbirne voll funktionsfähig ist unter der Bedingung, dass die Glühbirne nicht defekt ist. Das ist genau die bedingte Wahrscheinlichkeit $P(F|D^c)$:

$$\begin{aligned}
P(F|D^c) &= \frac{P(F \cap D^c)}{P(D^c)} \\
&= \frac{P\big(F \cap (K \cup F)\big)}{P(K \cup F)} && \text{wegen } D^c = K \cup F \\
&= \frac{P\big((F \cap K) \cup (F \cap F)\big)}{P(K \cup F)} \\
&= \frac{P\big(\emptyset \cup F\big)}{P(K \cup F)} && \text{wegen } F \cap K = \emptyset \\
&= \frac{P(F)}{P(K) + P(F)} && \text{wegen } K \cap F = \emptyset \\
&= \frac{0.7}{0.2 + 0.7} && \text{wegen } P(F) = 1 - P(D) - P(K) \\
&= 0.\overline{7}
\end{aligned}$$

Damit hat die gesuchte Wahrscheinlichkeit den Wert $0.\overline{7}$.

Die Gleichung (2.11) zur Berechnung der bedingten Wahrscheinlichkeit kann wie folgt umgestellt werden:

$$P(B \cap A) = P(B|A) \cdot P(A) \tag{2.12}$$

Aufgabe 13: Wir nehmen an, dass bei der Produktion von Glühbirnen die Wahrscheinlichkeit dafür, dass eine Glühbirne defekt ist, den Wert 0.1 hat. Diese Glühbirnen werden anschließend in Kisten verpackt und ausgeliefert. Die Wahrscheinlichkeit, dass eine solche Kiste beim Transport hinfällt, betrage 5%. Außerdem gehen wir davon aus, dass beim Hinfallen einer Kiste 30% der intakten Glühbirnen zerstört werden. Berechnen Sie die Wahrscheinlichkeit dafür, dass eine gelieferte Glühbirne funktionsfähig ist. □

2.8.1 Die totale Wahrscheinlichkeit und die Formel von Bayes

Es sei ein Wahrscheinlichkeits-Raum $\langle \Omega, 2^\Omega, P \rangle$ gegeben. Wir betrachten zwei Ereignisse A und B. Nach Definition des komplementären Ereignisses gilt

$$B \cup B^c = \Omega.$$

Wegen $A \cap \Omega = A$ folgt daraus

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Da die beiden Mengen $A \cap B$ und $A \cap B^c$ disjunkt sind, können wir damit die Wahrscheinlichkeit des Ereignisses A durch die folgende Formel berechnen:

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c) \quad \text{nach Gleichung (2.12)} \end{aligned}$$

Dieses Ergebnis läßt sich verallgemeinern. Ist eine Familie $B_1, \cdots, B_n$ von Ereignissen gegeben, so dass

1. $B_1 \cup \cdots \cup B_n = \Omega$
2. $B_i \cap B_j = \emptyset$ für $i \neq j$

gilt, so haben wir

$$\begin{aligned} P(A) &= P(A \cap B_1) + \cdots + P(A \cap B_n) \\ &= P(A|B_1) \cdot P(B_1) + \cdots + P(A|B_n) \cdot P(B_n) \\ &= \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i). \end{aligned} \tag{2.13}$$

Dies ist die Formel von der *totalen Wahrscheinlichkeit*. Eine Familie $B_1, \cdots, B_n$ von Ereignissen mit den oben angegebenen Eigenschaften bezeichnen wir als *Zerlegung* von Ω.

Aufgabe 14: Erfahrungsgemäß sind etwa 8% aller Männer farbenblind, während nur 0.6% aller Frauen farbenblind sind. Nehmen Sie an, dass der Anteil der Männer in der Gesamtbevölkerung 47% beträgt und berechnen Sie die Wahrscheinlichkeit dafür, dass eine beliebige Person farbenblind ist.

Lösung: Es sei B das Ereignis, das eine Person farbenblind ist. Das Ereignis, dass eine Person männlich ist bezeichnen wir mit M und das dazu komplementäre Ereignis bezeichnen wir mit F. Dann gilt

$$P(B) = P(B|M) \cdot P(M) + P(B|F) \cdot P(F) = 0.08 \cdot 0.47 + 0.006 \cdot 0.53 \approx 0.04078.$$

Also sind etwa 4.1% aller Personen farbenblind. □

Aus der Gleichung $P(B \cap A) = P(B|A) \cdot P(A)$ können wir eine weitere Formel ableiten:

$$\begin{aligned} P(B|A) \cdot P(A) &= P(B \cap A) \\ &= P(A \cap B) \\ &= P(A|B) \cdot P(B) \end{aligned}$$

Teilen wir diese Formel durch $P(A)$, so ergibt sich

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

Nehmen wir nun an, dass $B_1, \cdots, B_n$ eine Zerlegung von Ω ist, so gilt nach dem Satz von der totalen Wahrscheinlichkeit

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i).$$

Gilt nun $B = B_k$ für ein $k \in \{1, \cdots, n\}$, so haben wir insgesamt

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum\limits_{i=1}^{n} P(A|B_i) \cdot P(B_i)}. \tag{2.14}$$

Dies ist die Formel von *Bayes*. Diese Formel führt die Wahrscheinlichkeit dafür, dass ein Ereignis B_k unter der Bedingung A eintritt auf die Wahrscheinlichkeiten, dass das Ereignis A unter der

Bedingung B_i eintritt, zurück. Die Formel von Bayes ist sehr wichtig im Bereich der Medizin und der Rechtssprechung. Wir betrachten entsprechende Beispiele.

Beispiel: Ein (hypothetischer) Bluttest erkennt das Vorliegen von Malaria mit einer Wahrscheinlichkeit von 99.9%. In einem halben Promille der Fälle liefert der Test aber auch ein positives Ergebnis wenn keine Malaria vorliegt. Die Wahrscheinlichkeit, dass eine Person aus dem Großraum Stuttgart an Malaria erkrankt ist, liegt bei etwa 10^{-6}. Angenommen, Sie machen diesen Bluttest und erhalten ein positives Ergebnis. Wie hoch ist dann die Wahrscheinlichkeit, dass Sie Malaria haben?

Lösung: Wir bezeichnen das Ereignis, dass eine im Großraum Stuttgart lebende Person Malaria hat, mit B_1 und das dazu komplementäre Ereignis mit B_2. Weiterhin sei A das Ereignis, dass der Bluttest ein positives Result ergibt. Nach der Aufgabenstellung haben wir

$$P(B_1) = 10^{-6}, \quad P(B_2) = 1 - 10^{-6}, \quad P(A|B_1) = 0.999 \quad \text{und} \quad P(A|B_2) = 0.0005.$$

Gesucht ist die bedingte Wahrscheinlichkeit $P(B_1|A)$. Nach der Formel von Bayes gilt

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)} \\ &= \frac{0.999 \cdot 10^{-6}}{0.999 \cdot 10^{-6} + 0.0005 \cdot (1 - 10^{-6})} \\ &\approx 0.001994017946 \end{aligned}$$

Damit liegt die Wahrscheinlichkeit dafür, dass Sie tatsächlich an Malaria erkrankt sind, unter 2‰. Dieses Beispiel zeigt, dass die bedingte Wahrscheinlichkeit für das Eintreten eines Ereignisses A unter einer Bedingung B völlig verschieden ist von der Wahrscheinlichkeit des Eintretens des Ereignisses B unter der Bedingung A! □

Das bei der letzten Aufgabe erhaltene Ergebnis erscheint zunächst kontraintuitiv. Wir beleuchten den Sachverhalt daher noch aus einer anderen Perspektive. Wir gehen davon aus, dass im Großraum Stuttgart etwa 1 000 000 Personen leben. Von diesen Personen hat dann genau eine Person Malaria. Würden wir nun alle Personen aus dem Großraum Stuttgart testen, so würden wir zwei Gruppen von Personen positiv testen:

1. Die an Malaria erkrankte Person wird (mit hoher Wahrscheinlichkeit) positiv getestet.

2. Bei den gesunden Personen schlägt der Test mit einer Wahrscheinlichkeit von einem halbe Promille an. Daher erhalten wir also zusätzlich etwa $0.0005 \cdot 1\,000\,000 = 500$ weitere positive Testergebnisse.

Da es insgesamt 501 positive Tests gibt, ist das Verhältnis von positiven Tests, die korrekt das Vorliegen von Malaria erkennen, zu den positiven Tests, die fälschlicherweise das Vorliegen von Malaria behaupten, durch den Bruch $\frac{1}{501}$ geben und das sind etwa 2 Promille.

Aufgabe 15: Auf einer Insel ist die Tochter des dort amtierenden Königs einem besonders gruseligen Verbrechen zum Opfer gefallen. Es gibt zunächst keinen Verdächtigen, aber dafür kann am Tatort eine DNA-Probe des Täters sichergestellt werden. Da es zu teuer ist, alle 100 000 auf der Insel lebenden Männer zu testen, werden zufällig 100 Männer für einen DNA-Test ausgewählt. Bei einem dieser Männer ist der Test in der Tat positiv. Bei dem besagten Test wird ein spezielles Gen verglichen. Die Chance, dass dieses Gen bei zwei zufällig ausgewählten Männern identisch ist, liegt bei 1 zu 50 000. Wie groß ist die Wahrscheinlichkeit, dass es sich bei dem positiv getesteten Mann um den Täter handelt? Beantworten Sie die selbe Frage für den Fall, dass auf der Insel 1 000 000 Männer leben.

2.8.2 Das Ziegen-Problem

Aufgabe 16: Das folgende Problem wird in der Literatur als das *Ziegen-Problem* bezeichnet.

> *Bei einer Fernseh-Show steht der Kandidat vor der Auswahl, eine von drei Türen zu öffnen. Hinter einer der Türen befindet sich ein Auto, das der Kandidat gewinnt, wenn er diese Tür öffnet. Hinter den anderen beiden Türen befindet sich jeweils eine Ziege. Der Kandidat hat keinerlei Informationen, wo sich das Auto befindet und wählt zufällig eine der Türen aus. Nachdem der Kandidat seine Wahl getroffen hat, gibt er diese Wahl bekannt und der Show-Master tritt in Aktion. Dieser weiß, hinter welcher Tür sich das Auto befindet und öffnet eine Tür, die einerseits verschieden ist von der Tür, die der Kandidat gewählt hat und hinter der andererseits eine Ziege steht. Falls es hier zwei Möglichkeiten gibt, trifft der Show-Master die Wahl zufällig, wobei er beide Türen mit der selben Wahrscheinlichkeit wählt. Der Kandidat erhält jetzt die Möglichkeit, seine ursprüngliche Wahl zu revidieren und die andere, noch verbleibende Tür auszuwählen. Angenommen, der Kandidat hat die erste Tür gewählt und der Show-Master hat die zweite Tür geöffnet um dem Kandidaten die dort verborgende Ziege zu zeigen. Sollte der Kandidat nun seine Wahl revidieren?*

Lösung: Wir überlegen uns zunächst, wie der Ergebnisraum Ω aussieht und definieren

$$M = \{1, 2, 3\} \quad \text{und} \quad \Omega = \bigl\{\langle a, w, o\rangle \mid a \in M \wedge w \in M \wedge o \in M\backslash\{a, w\}\bigr\}$$

Hier gibt a die Tür an, hinter der das <u>A</u>uto steht, w gibt die Tür an, die der Kandidat ge<u>w</u>ählt hat und o gibt die Tür an, die vom Show-Master ge<u>ö</u>ffnet wurde. Die Wahrscheinlichkeits-Verteilung P, die dem Problem zugrunde liegt, ist **keine** gleichmäßige Verteilung. Der Grund dafür ist die dritte Komponente o der Trippel $\langle a, w, o\rangle$. Zunächst sind alle Werte von a und w tatsächlich gleichwahrscheinlich. Für gegebene Werte $a, w \in M$ definieren wir das Ereignis $E(a, w)$ als

$$E(a, w) = \bigl\{\langle a, w, o\rangle \in M^3 \bigm| o \neq a \wedge o \neq w\bigr\}.$$

Das Ereignis $E(a, w)$ fasst alle die Ergebnisse aus Ω zusammen, bei denen die Werte von a und w fest sind. Wir nehmen an, dass diese Ereignisse alle die selbe Wahrscheinlichkeit haben. Das es insgesamt 9 Paare $\langle a, w\rangle$ gibt, gilt

$$\forall a, w \in M : P\bigl(E(a, w)\bigr) = \frac{1}{9}.$$

Wenn $a \neq w$ ist, dann enthält die Menge $E(a, w)$ genau ein Element, aber wenn $a = w$ ist, was dem Fall entspricht, dass der Kandidat die Tür gewählt hat, hinter der tatsächlich das Auto steht, dann enthält die Menge $E(a, w)$ zwei Elemente, denn dann hat der Show-Master zwei Türen zur Verfügung, die er öffnen kann. In einem solchen Fall erachten wir beide Möglichkeiten als gleichwahrscheinlich. Daher gilt insgesamt

$$P\Bigl(\bigl\{\langle a, w, o\rangle\bigr\}\Bigr) = \begin{cases} \dfrac{1}{9} & \text{falls } a \neq w; \\[2ex] \dfrac{1}{18} & \text{sonst.} \end{cases}$$

Um das Problem zu lösen, definieren wir nun eine Reihe von Ereignissen.

1. $A_n := \bigl\{\langle a, w, o\rangle \in \Omega \bigm| a = n\bigr\}$ für $n = 1, 2, 3$.

   Das Ereignis A_n gibt an, dass das Auto hinter der Tür mit der Nummer n steht.

2. $W_n := \bigl\{\langle a, w, o\rangle \in \Omega \bigm| w = n\bigr\}$ für $n = 1, 2, 3$.

   Das Ereignis W_n gibt an, dass der Kandidat die Tür mit der Nummer n gewählt hat.

3. $O_n := \bigl\{\langle a, w, o\rangle \in \Omega \bigm| o = n\bigr\}$ für $n = 1, 2, 3$.

   Das Ereignis O_n gibt an, dass der Show-Master die Tür mit der Nummer n geöffnet hat.

Um die Aufgabe zu lösen müssen wir die bedingten Wahrscheinlichkeiten

$$P(A_1|W_1 \cap O_2) \quad \text{und} \quad P(A_3|W_1 \cap O_2)$$

berechnen, denn dies sind die Wahrscheinlichkeiten dafür, dass das Auto hinter der ersten bzw. hinter der zweiten Tür steht. Die Bedingung $W_1 \cap O_2$ drückt dabei unser bisheriges Wissen aus: Der Kandiat hat die Tür 1 gewählt und der Show-Master hat die Tür 2 geöffnet.

Wir beginnen mit der Berechnung von $P(A_3|W_1 \cap O_2)$. Nach Gleichung (2.11) gilt

$$P(A_3|W_1 \cap O_2) = \frac{P(A_3 \cap W_1 \cap O_2)}{P(W_1 \cap O_2)}. \tag{2.15}$$

Nun haben wir $A_3 \cap W_1 \cap O_2 = \bigl\{\langle 3, 1, 2\rangle\bigr\}$, also gilt

$$P(A_3 \cap W_1 \cap O_2) = P\Bigl(\bigl\{\langle 3, 1, 2\rangle\bigr\}\Bigr) = \frac{1}{9}. \tag{2.16}$$

Die Wahrscheinlichkeit $P(W_1 \cap O_2)$ berechnen wir mit Hilfe des Satzes von der totalen Wahrscheinlichkeit, wobei wir als Zerlegung von Ω die Mengen A_1, A_2 und A_3 wählen:

$$P(W_1 \cap O_2) = \sum_{i=1}^{3} P(W_1 \cap O_2|A_i) \cdot P(A_i) \tag{2.17}$$

Die Ausdrücke $P(W_1 \cap O_2|A_i) \cdot P(A_i)$ berechnen wir für $i = 1, 2, 3$ mit Hilfe von Gleichung (2.12).

1. $P(W_1 \cap O_2|A_1) \cdot P(A_1) = P(W_1 \cap O_2 \cap A_1) = P\Bigl(\bigl\{\langle 1, 1, 2\rangle\bigr\}\Bigr) = \frac{1}{18}.$
2. $P(W_1 \cap O_2|A_2) \cdot P(A_2) = P(W_1 \cap O_2 \cap A_2) = P\Bigl(\bigl\{\bigr\}\Bigr) = 0.$
3. $P(W_1 \cap O_2|A_3) \cdot P(A_3) = P(W_1 \cap O_2 \cap A_3) = P\Bigl(\bigl\{\langle 3, 1, 2\rangle\bigr\}\Bigr) = \frac{1}{9}.$

Damit haben wir nun nach Gleichung (2.17)

$$P(W_1 \cap O_2) = \frac{1}{18} + 0 + \frac{1}{9} = \frac{3}{18} = \frac{1}{6}. \tag{2.18}$$

Setzen wir die Ergebnisse der Gleichungen (2.16) und (2.18) in Gleichung (2.15) ein, so erhalten wir

$$P(A_3|W_1 \cap O_2) = \frac{\frac{1}{9}}{\frac{1}{6}} = \frac{2}{3}. \tag{2.19}$$

Eine analoge Rechnung liefert $P(A_1|W_1 \cap O_2) = \frac{1}{3}$. Wir brauchen die Rechnung nicht durchzuführen, denn es muss

$$P(A_1|W_1 \cap O_2) + P(A_2|W_1 \cap O_2) + P(A_3|W_1 \cap O_2) = 1$$

gelten und wegen $P(A_2|W_1 \cap O_2) = 0$ wissen wir

$$P(A_1|W_1 \cap O_2) = 1 - P(A_3|W_1 \cap O_2) = 1 - \tfrac{2}{3} = \tfrac{1}{3}.$$

Folglich ist es für den Kandiaten vorteilhaft, seine Entscheidung zu revidieren. $\square$

2.9 Unabhängige Ereignisse

Es gibt viele Situationen in denen das Wissen, dass ein Ereignis E eingetreten ist, die Wahrscheinlichkeit für das Auftreten eines anderen Ereignisses F nicht beeinflußt. In diesem Fall gilt

$$P(F|E) = P(F). \tag{2.20}$$

Ein einfaches Beispiel dafür wäre das Würfeln mit zwei Würfeln. Wenn F das Ereignis ist, dass im ersten Wurf eine Sechs gewürfelt wird und E das Ereignis ist, das der im zweiten Wurf eine Drei gewürfelt wird, also

$$F = \big\{\langle 6, i\rangle \;\big|\; i \in \{1, \cdots, 6\}\big\} \quad \text{und} \quad E = \big\{\langle i, 3\rangle \;\big|\; i \in \{1, \cdots, 6\}\big\},$$

dann ist offensichtlich, dass das Ereignis E das Ereignis F nicht beeinflussen kann, denn die beiden Ereignisse betreffen ja verschiedene Würfel.

Ersetzen wir in Gleichung (2.20) die bedingte Wahrscheinlichkeit durch den in Gleichung (2.11) gegebenen Wert, so haben wir

$$\frac{P(F \cap E)}{P(E)} = P(F).$$

Multiplikation dieser Gleichung mit $P(E)$ liefert

$$P(F \cap E) = P(E) \cdot P(F). \tag{2.21}$$

Wir bezeichnen zwei Ereignisse E und F als *unabhängig*, wenn Gleichung (2.21) erfüllt ist. Wegen $F \cap E = E \cap F$ ist diese Definition symmetrisch: Die Ereignisse E und F sind genau dann unabhängig, wenn die Ereignisse F und E unabhängig sind.

Unabhängige Ereignisse treten dann auf, wenn ein Zufalls-Experiment durchgeführt wird, dass aus zwei unabhängigen Zufalls-Experimenten besteht. Ein einfaches Beispiel dafür ist das Würfeln mit zwei Würfeln. Um diese Situation formal beschreiben zu können, führen wir das Produkt zweier Wahrscheinlichkeits-Räume ein.

Definition 5 (Produkt-Raum) Sind $\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle$ und $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$ zwei Wahrscheinlichkeits-Räume, so definieren wir

$$\mathcal{W}_1 \times \mathcal{W}_2 := \langle \Omega_1 \times \Omega_2, 2^{\Omega_1 \times \Omega_2}, P\rangle,$$

wobei die neue Wahrscheinlichkeits-Verteilung P dadurch definiert wird, dass wir die Wahrscheinlichkeiten der Elementar-Ereignisse angeben:

$$P\big(\{\langle x, y\rangle\}\big) := P_1(\{x\}) \cdot P_2(\{y\}) \quad \text{für alle } x \in \Omega_1 \text{ und alle } y \in \Omega_2.$$

Für beliebige Ereignisse E ist die Wahrscheinlichkeits-Verteilung dann durch

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

gegeben. □

Definition 6 Ist

$$\langle \Omega, 2^{\Omega}, P\rangle = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle \times \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$$

ein Produkt-Raum, und ist E ein Ereignis aus diesem Raum, so sagen wir, dass E *durch die erste Komponente bestimmt* ist, falls es eine Menge $\widehat{E}$ gibt, so dass

$$E = \big\{\langle x, y\rangle \;\big|\; x \in \widehat{E} \wedge y \in \Omega_2\big\}$$

gilt. Analog sagen wir, dass E *durch die zweite Komponente bestimmt* ist, wenn es eine Menge $\widehat{E}$ gibt, so dass

$$E = \big\{\langle x, y\rangle \;\big|\; x \in \Omega_1 \wedge y \in \widehat{E}\big\}$$

gilt. □

Beispiel: Wir betrachten das Zufalls-Experiment "Würfeln mit zwei Würfeln" und definieren zunächst $\Omega_1 := \{1, \cdots, 6\}$,

$$P_1(A) := \frac{1}{6} \cdot |A| \quad \text{und} \quad \mathcal{W}_1 := \langle \Omega_1, 2^{\Omega_1}, P_1\rangle.$$

Weiter sei $\mathcal{W} := \mathcal{W}_1 \times \mathcal{W}_1$. Dann ist das Ereignis

$$\big\{\langle 6, n\rangle \;\big|\; n \in \{1, \cdots, 6\}\big\}$$

durch die erste Komponente bestimmt, während das Ereignis

$$\big\{\langle m, n\rangle \;\big|\; m \in \{1, \cdots, 6\} \wedge n \in \{2, 4, 6\}\big\}$$

durch die zweite Komponente bestimmt wird. □

Satz 7 Ist $\mathcal{W} = \langle \Omega_1, 2^{\Omega_1}, P_1\rangle \times \langle \Omega_2, 2^{\Omega_2}, P_2\rangle$ ein Produkt-Raum und sind E und F Ereignisse, so dass E durch die erste und F durch die zweite Komponente bestimmt ist, so sind die Ereignisse E und F unabhängig.

Beweis: Nach Voraussetzung gibt es Mengen $\widehat{E}$ und $\widehat{F}$, so dass

$$E = \bigl\{\langle x, y\rangle \in \Omega_1 \times \Omega_2 \mid x \in \widehat{E} \wedge y \in \Omega_2\bigr\} \quad \text{und} \quad F = \bigl\{\langle x, y\rangle \in \Omega_1 \times \Omega_2 \mid x \in \Omega_1 \wedge y \in \widehat{F}\bigr\}$$

gilt. Es ist zu zeigen, dass

$$P(E \cap F) = P(E) \cdot P(F)$$

gilt. Nach Definition der Wahrscheinlichkeits-Verteilung P gilt

$$\begin{aligned} P(E \cap F) &= \sum_{\omega \in E \cap F} P\bigl(\{\omega\}\bigr) \\ &= \sum_{\langle x,y\rangle \in E \cap F} P\bigl(\{\langle x, y\rangle\}\bigr) \end{aligned}$$

Die Bedingung $\langle x, y\rangle \in E \cap F$ formen wir um:

$$\begin{aligned} & \langle x, y\rangle \in E \cap F \\ \leftrightarrow\ & \langle x, y\rangle \in E \wedge \langle x, y\rangle \in F \\ \leftrightarrow\ & \Bigl(x \in \widehat{E} \wedge y \in \Omega_2\Bigr) \wedge \Bigl(x \in \Omega_1 \wedge y \in \widehat{F}\Bigr) \\ \leftrightarrow\ & x \in \widehat{E} \wedge y \in \widehat{F} \end{aligned}$$

Damit können wir die Summe in der Gleichung für $P(E \cap F)$ umschreiben:

$$P(E \cap F) = \sum_{x \in \widehat{E},\ y \in \widehat{F}} P\bigl(\{\langle x, y\rangle\}\bigr).$$

Nach Definition des Produkt-Raums gilt $P\bigl(\{\langle x, y\rangle\}\bigr) = P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr)$. Damit haben wir

$$P(E \cap F) = \sum_{x \in \widehat{E},\ y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr).$$

Jetzt können wir die Summe, die oben über x und y läuft, in zwei getrennte Summen aufspalten und finden

$$\begin{aligned} P(E \cap F) &= \sum_{x \in \widehat{E},\ y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr) \\ &= \sum_{x \in \widehat{E}} \sum_{y \in \widehat{F}} P_1\bigl(\{x\}\bigr) \cdot P_2\bigl(\{y\}\bigr) \\ &= \sum_{x \in \widehat{E}} P_1\bigl(\{x\}\bigr) \sum_{y \in \widehat{F}} P_2\bigl(\{y\}\bigr) \\ &= \left(\sum_{x \in \widehat{E}} P_1\bigl(\{x\}\bigr)\right) \cdot \left(\sum_{y \in \widehat{F}} P_2\bigl(\{y\}\bigr)\right) \\ &= P_1\bigl(\widehat{E}\bigr) \cdot P_2\bigl(\widehat{F}\bigr) \end{aligned}$$

Der Beweis ist abgeschlossen, wenn wir zeigen können, dass $P_1(\widehat{E}) = P(E)$ und $P_2(\widehat{F}) = P(F)$ gilt. Dies folgt aus

$$
\begin{aligned}
P(E) &= P\big(\{\langle x,y\rangle \in \Omega_1\times\Omega_2 \mid x\in \widehat{E} \wedge y \in \Omega_2\}\big) \\
&= \sum\limits_{\langle x,y\rangle\in\{\langle x,y\rangle\in\Omega_1\times\Omega_2 \,\mid\, x\in\widehat{E}\,\wedge\, y\in\Omega_2\}} P\big(\{\langle x,y\rangle\}\big) \\
&= \sum\limits_{x\in\widehat{E},\, y\in\Omega_2} P\big(\{\langle x,y\rangle\}\big) \\
&= \sum\limits_{x\in\widehat{E},\, y\in\Omega_2} P_1\big(\{x\}\big)\cdot P_2\big(\{y\}\big) \\
&= \sum\limits_{x\in\widehat{E}} \sum\limits_{y\in\Omega_2} P_1\big(\{x\}\big)\cdot P_2\big(\{y\}\big) \\
&= \sum\limits_{x\in\widehat{E}} P_1\big(\{x\}\big) \sum\limits_{y\in\Omega_2} P_2\big(\{y\}\big) \\
&= \left(\sum\limits_{x\in\widehat{E}} P_1\big(\{x\}\big)\right)\cdot\left(\sum\limits_{y\in\Omega_2} P_2\big(\{y\}\big)\right) \\
&= P_1\big(\widehat{E}\big)\cdot P_2\big(\Omega_2\big) \\
&= P_1\big(\widehat{E}\big)\cdot 1 \\
&= P_1\big(\widehat{E}\big)
\end{aligned}
$$

Damit haben wir $P_1(\widehat{E}) = P(E)$ gezeigt. Der Nachweis von $P_2(\widehat{F}) = P(F)$ verläuft völlig analog, so dass wir auf die Details verzichten können. $\square$

Es ist offensichtlich, dass der Begriff des Produkt-Raums auch auf Produkte von mehr als zwei Faktoren erweitert werden kann. Als Anwendung der bisher präsentierten Theorie zeigen wir eine alternative Lösung des Ziegen-Problems, bei der wir den zugrunde liegenden Wahrscheinlichkeits-Raum nicht explizit angeben. Stattdessen können wir uns auf die Betrachtung von Ereignissen beschränken.

Beispiel: Diesmal definieren wir die Ereignisse A_i, W_i und O_i unmittelbar ohne Rückgriff auf einen zugrunde liegenden Ergebnisraum Ω wie folgt:

1. A_n: "*Das Auto steht hinter der Tür mit der Nummer* n".

   Da die Ereignisse A_1, A_2 und A_3 alle die selbe Wahrscheinlichkeit haben, gilt
   $$P(A_n) = \frac{1}{3} \quad \text{für alle } n \in \{1,2,3\}.$$

2. W_n: "*Der Kandidat hat die Tür mit der Nummer* n *gewählt*".

   Da die Ereignisse W_1, W_2 und W_3 alle die selbe Wahrscheinlichkeit haben, gilt
   $$P(W_n) = \frac{1}{3} \quad \text{für alle } n \in \{1,2,3\}.$$

3. O_n: "*Der Showmaster hat die Tür mit der Nummer* n *geöffnet*".

   Die Wahrscheinlichkeit $P(O_n)$ läßt sich nicht unmittelbar angeben, denn das Ereignis O_n wird offenbar von den anderen Ereignissen beeinflußt.

Wir berechnen wieder die bedingte Wahrscheinlichkeit $P(A_3|W_1\cap O_2)$:
$$P(A_3|W_1\cap O_2) = \frac{P(A_3\cap W_1\cap O_2)}{P(W_1\cap O_2)}$$
Um an dieser Stelle weitermachen zu können, müssen wir $P(A_3\cap W_1\cap O_2)$ berechnen. Wir stellen diese Wahrscheinlichkeit als bedingte Wahrscheinlichkeit dar:
$$P(A_3\cap W_1\cap O_2) = P(O_2\cap A_3\cap W_1) = P(O_2|A_3\cap W_1)\cdot P(A_3\cap W_1)$$
Die Ereignisse A_3 und W_1 sind offenbar unabhängig, daher gilt

$$P(A_3 \cap W_1) = P(A_3) \cdot P(W_1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Die bedingte Wahrscheinlichkeit $P(O_2|A_3 \cap W_1)$ hat den Wert 1, denn wenn das Auto hinter der dritten Tür steht und der Kandidat die erste Tür wählt, dann hat der Showmaster keine Wahlmöglichkeit und öffnet immer die zweite Tür. Damit haben wir also

$$P(A_3 \cap W_1 \cap O_2) = P(O_2|A_3 \cap W_1) \cdot P(A_3 \cap W_1) = \frac{1}{9}$$

Jetzt müssen wir noch die Wahrscheinlichkeit $P(W_1 \cap O_2)$ bestimmen. Das geht so:

$$\begin{aligned} P(W_1 \cap O_2) &= P\big(W_1 \cap O_2 \cap (A_1 \cup A_2 \cup A_3)\big) \\ &= P\big((W_1 \cap O_2 \cap A_1) \cup (W_1 \cap O_2 \cap A_2) \cup (W_1 \cap O_2 \cap A_3)\big) \\ &= P(W_1 \cap O_2 \cap A_1) + P(W_1 \cap O_2 \cap A_2) + P(W_1 \cap O_2 \cap A_3) \end{aligned}$$

Es bleibt die Aufgabe, die Wahrscheinlichkeiten $P(W_1 \cap O_2 \cap A_i)$ für $i = 1, 2, 3$ zu berechnen. Damit das möglich ist, müssen wir diese Wahrscheinlichkeiten als bedingte Wahrscheinlichkeiten schreiben. Wir beginnen mit $P(W_1 \cap O_2 \cap A_1)$:

$$P(W_1 \cap O_2 \cap A_1) = P(O_2 \cap A_1 \cap W_1) = P(O_2|A_1 \cap W_1) \cdot P(A_1 \cap W_1)$$

Die Ereignisse A_1 und W_1 sind offenbar unabhängig, daher gilt

$$P(A_1 \cap W_1) = P(A_1) \cdot P(W_1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

Die bedingte Wahrscheinlichkeit $P(O_2|A_1 \cap W_1)$ hat aus Symmetrie-Gründen den selben Wert wie die Wahrscheinlichkeit $P(O_3|A_1 \cap W_1)$, denn wenn das Auto hinter der ersten Tür steht und der Kandidat diese Tür wählt, dann kann der Showmaster entweder die zweite oder die dritte Tür öffnen. Da diese beiden Wahrscheinlichkeiten zusammen den Wert 1 ergeben müssen, folgt

$$P(O_2|A_1 \cap W_1) = \tfrac{1}{2}.$$

Damit haben wir also

$$P(W_1 \cap O_2 \cap A_1) = P(O_2|A_1 \cap W_1) \cdot P(A_1 \cap W_1) = \frac{1}{2} \cdot \frac{1}{9} = \frac{1}{18}.$$

Jetzt berechnen wir $P(W_1 \cap O_2 \cap A_2)$. Hier müssen wir nicht lange überlegen, denn wenn das Auto hinter der zweiten Tür steht, dann wird der Showmaster die zweite Tür nicht öffnen, also gilt

$$W_1 \cap O_2 \cap A_2 = \emptyset, \quad \text{folglich ist} \quad P(W_1 \cap O_2 \cap A_2) = 0.$$

Jetzt benötigen wir $P(W_1 \cap O_2 \cap A_3)$. Es gilt aber $W_1 \cap O_2 \cap A_3 = A_3 \cap W_1 \cap O_2$ und für dieses Ereignis haben wir oben bereits die Wahrscheinlichkeit $\frac{1}{9}$ gefunden. Damit haben wir insgesamt

$$\begin{aligned} P(A_3|W_1 \cap O_2) &= \frac{P(A_3 \cap W_1 \cap O_2)}{P(W_1 \cap O_2)} \\ &= \frac{\frac{1}{9}}{P(W_1 \cap O_2 \cap A_1) + P(W_1 \cap O_2 \cap A_2) + P(W_1 \cap O_2 \cap A_3)} \\ &= \frac{\frac{1}{9}}{\frac{1}{18} + 0 + \frac{1}{9}} \\ &= \frac{2}{1+2} \\ &= \frac{2}{3} \end{aligned}$$

Das ist das selbe Ergebnis, was wir auch schon früher gefunden haben. □

Aufgabe 17: Anton, Bruno und Charlie haben sich unsterblich in die selbe Frau verliebt und beschließen, ein Dreier-Duell durchzuführen. Dieses wird folgendermaßen durchgeführt: Die drei Duellanten stellen sich so auf, dass zwischen jedem Paar ein Abstand von 25 Metern besteht. Es wird ausgelost, wer als erstes einen Schuß abgeben darf, danach wird immer in der Reihenfolge Anton, Bruno, Charlie geschossen, mit der kleinen Einschränkung, dass jemand, der tot ist, nicht mehr mitspielen darf. Wenn jemand einen Schuß abgeben darf, so steht es ihm frei, auf welchen

Kontrahenten er schießt. Außerdem hat er auch die Möglichkeit, in die Luft zu schießen. Dies gilt allerdings nicht für Anton, der darf nicht in die Luft schießen. Die Bewaffnung der drei Duellanten ist unterschiedlich.

1. Anton verfügt über ein Maschinen-Gewehr. Seine Trefferwahrscheinlichkeit liegt daher bei 100%.
2. Bruno setzt eine Pumpgun ein und hat eine Trefferwahrscheinlichkeit von 80%.
3. Charlie besitzt einen klassischen Fünfundvierziger Peacemaker mit dem er eine Trefferwahrscheinlichkeit von 50% erzielt.

Lösen Sie die folgenden Aufgaben:

1. Überlegen Sie, welche Strategie für die einzelnen Kontrahenten optimal ist.
2. Berechnen Sie die Überlebenswahrscheinlichkeit für jeden der Duellanten.
3. Wie groß ist der Erwartungswert für die Dauer des Duells.

Hinweis: Die am schwierigsten zu berechnende Situation ist die, wenn Bruno den Anton erschossen hat und Charlie und Bruno sich nun abwechselnd beschießen. Um die bei diesem Duell auftretenden Wahrscheinlichkeiten berechnen zu können, ist es zweckmäßig, für dieses Duell die folgenden Abkürzungen einzuführen.

1. $c(n)$ ist die Wahrscheinlichkeit dafür, dass Charlie nach der n-ten Runde noch lebt.

   Eine Runde besteht dabei aus dem Schuss von Charlie und dem Schuss von Bruno, falls dieser noch dazu in der Lage ist.
2. $b(n)$ ist die Wahrscheinlichkeit dafür, dass Bruno nach der n-ten Runde noch lebt.
3. $g(n)$ ist die Wahrscheinlichkeit dafür, dass das Spiel nach der n-ten Runde beendet ist.

   Die Wahrscheinlichkeit g kann auf die beiden Wahrscheinlichkeiten $c(n)$ und $b(n)$ zurückgeführt werden.

Es ist zweckmäßig, für $c(n)$ und $b(n)$ ein System von Rekurrenz-Gleichungen aufzustellen und dieses System mit Hilfe von *Maple* zu lösen. Anschließend können die Grenzwerte dieser Folgen für n gegen Unendlich berechnet werden. Es ergibt sich

$$\lim_{n\to\infty} c(n) = \frac{5}{9} \quad \text{und} \quad \lim_{n\to\infty} b(n) = \frac{4}{9}.$$

Der Rest der Aufgabe ist dann nicht mehr schwer.

Der Begriff der Unabhängigkeit von Ereignissen kann auf mehr als zwei Ereignisse ausgedehnt werden. Wir sagen, dass $\{A_1, \cdots, A_n\}$ eine Menge unabhängiger Ereignisse ist, falls gilt:

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) * \cdots * P(A_n).$$

2.10 Unabhängige Zufallsgrößen

Wir verallgemeinern die Begriffsbildung des letzten Abschnitts und betrachten nun die Unabhängigkeit von Zufallsgrößen.

Definition 8 (Unabhängige Zufallsgrößen) Es sei

1. $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ seien zwei Zufallsgrößen.

Dann sind die Zufallsgrößen X und Y unabhängig, wenn für alle $x, y \in \mathbb{R}$ die Ereignisse

$$\big\{\omega \in \Omega \bigm| X(\omega) = x\big\} \quad \text{und} \quad \big\{\omega \in \Omega \bigm| Y(\omega) = y\big\}.$$

unabhängig sind. Nach Definition der Unabhängigkeit zweier Ereignisse ist das äquivalent zu der Forderung, dass für alle $x, y \in \mathbb{R}$ die Gleichung

$$\begin{array}{ll} & P\Big(\big\{\omega \in \Omega \bigm| X(\omega) = x\big\} \cap \big\{\omega \in \Omega \bigm| Y(\omega) = y\big\}\Big) \\ = & P\Big(\big\{\omega \in \Omega \bigm| X(\omega) = x\big\}\Big) \cdot P\Big(\big\{\omega \in \Omega \bigm| Y(\omega) = y\big\}\Big) \end{array}$$

gilt. Diese Gleichung können wir kürzer in der Form

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$$

schreiben. □

Beispiel: Angenommen, wir würfeln mit zwei Laplace-Würfel. Der Ergebnis-Raum hat dann die Form

$$\Omega = \big\{\langle i, j\rangle \bigm| i, j \in \{1, \cdots, 6\}\big\}.$$

Definieren wir die Zufallsgrößen X und Y durch

$$X\big(\langle i, j\rangle\big) = i \quad \text{und } Y\big(\langle i, j\rangle\big) = j,$$

so sind diese Zufallsgrößen unabhängig, denn wählen wir beispielsweise $x = 2$ und $y = 5$, so müssen wir zeigen, dass

$$P(X = 2 \wedge Y = 5) = P(X = 2) \cdot P(Y = 5)$$

gilt. Dies rechnen wir sofort nach, denn einerseits gilt

$$P(X = 2 \wedge Y = 5) = P\big(\{\langle 2, 5\rangle\}\big) = \frac{1}{36},$$

andererseits haben wir

$$\begin{array}{lcl} P(X = 2) & = & P\big(\{\langle i, j\rangle \in \Omega \mid X(\langle i, j\rangle) = 2\}\big) \\ & = & P\big(\{\langle i, j\rangle \in \Omega \mid i = 2\}\big) \\ & = & P\big(\{\langle 2, j\rangle \mid j \in \{1, \cdots, 6\}\big) \\ & = & \dfrac{\big|\{\langle 2, j\rangle \mid j \in \{1, \cdots, 6\}\big|}{36} \\ & = & \dfrac{6}{36} \\ & = & \dfrac{1}{6}. \end{array}$$

Genau so sehen wir $P(Y = 5) = \frac{1}{6}$ und daher gilt insgesamt

$$P(X = 2 \wedge Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(X = 2) \cdot P(Y = 5).$$

Die Tatsache, dass wir hier mit $x = 2$ und $y = 5$ gearbeitet haben, ist völlig unerheblich, denn wir hätten bei jedem anderen Wert das selbe Ergebnis bekommen. Daher sind die beiden Zufallsgrößen X und Y unabhängig. □

Um das letzte Beispiel verallgemeinern zu können, fehlt noch eine Definition.

Definition 9 Es seien

1. $\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1 \rangle$ und $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2 \rangle$ Wahrscheinlichkeits-Räume.
2. $\mathcal{W} = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega, 2^{\Omega}, P \rangle$ der aus $\mathcal{W}_1$ und $\mathcal{W}_2$ gebildete Produkt-Raum.
3. $X : \Omega \to \mathbb{R}$ eine Zufallsgröße auf $\mathcal{W}$.

Dann *wird X durch die erste Komponente bestimmt*, falls

$$\forall \omega_1 \in \Omega_1 : \forall \omega_2, \omega_3 \in \Omega_2 : X\big(\langle \omega_1, \omega_2 \rangle\big) = X\big(\langle \omega_1, \omega_3 \rangle\big)$$

gilt. Bei der Auswertung der Zufallsgröße X spielt also die zweite Komponente keine Rolle. Analog sagen wir, dass X *durch die zweite Komponente bestimmt wird*, wenn die erste Komponente bei der Auswertung von X keine Rolle spielt, wenn also gilt:

$$\forall \omega_2 \in \Omega_2 : \forall \omega_1, \omega_3 \in \Omega_1 : X\big(\langle \omega_1, \omega_2 \rangle\big) = X\big(\langle \omega_3, \omega_2 \rangle\big).$$ □

Satz 10 Es seien

1. $\mathcal{W}_1 = \langle \Omega_1, 2^{\Omega_1}, P_1 \rangle$ und $\mathcal{W}_2 = \langle \Omega_2, 2^{\Omega_2}, P_2 \rangle$ Wahrscheinlichkeits-Räume.
2. $\mathcal{W} = \mathcal{W}_1 \times \mathcal{W}_2 = \langle \Omega, 2^{\Omega}, P \rangle$ der aus $\mathcal{W}_1$ und $\mathcal{W}_2$ gebildete Produkt-Raum.
3. $X_1 : \Omega \to \mathbb{R}$ und $X_2 : \Omega \to \mathbb{R}$ Zufallsgrößen auf $\mathcal{W}$.

Falls nun X_1 durch die erste Komponente bestimmt wird und X_2 durch die zweite Komponente bestimmt wird, dann sind die Zufallsgrößen X_1 und X_2 unabhängig.

Beweis: Wir müssen zeigen, dass für beliebige Zahlen $x_1, x_2 \in \mathbb{R}$ die Gleichung

$$P(X_1 = x_1 \wedge X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$

gilt. Dies ist gleichbedeutend damit, dass die beiden Ereignisse

$$E_1 := \big\{\omega \in \Omega \bigm| X_1(\omega) = x_1\big\} \quad \text{und} \quad E_2 := \big\{\omega \in \Omega \bigm| X_2(\omega) = x_2\big\}$$

unabhängig sind. Dies folgt aus Satz 7, falls wir zeigen können, dass E_1 auf die erste und E_2 auf zweite Komponente beschränkt ist. Wir führen den Nachweis für das Ereignis E_1, der Nachweis für das Ereignis E_2 verläuft völlig analog. Dazu definieren wir eine Menge $\widehat{E}_1$ durch

$$\widehat{E}_1 := \big\{\omega_1 \in \Omega_1 \bigm| \exists \omega_2 \in \Omega_2 : X_1(\langle \omega_1, \omega_2 \rangle) = x_1\big\}$$

und zeigen, dass $E_1 = \big\{\langle \omega_1, \omega_2 \rangle \in \Omega \bigm| \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2\big\}$ gilt. Den Nachweis, dass diese beiden Mengen gleich sind führen wir, indem wir zeigen, dass

$$\langle \omega_1, \omega_2 \rangle \in \big\{\omega \in \Omega \bigm| X_1(\omega) = x_1\big\} \;\Leftrightarrow\; \langle \omega_1, \omega_2 \rangle \in \big\{\langle \omega_1, \omega_2 \rangle \in \Omega \bigm| \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2\big\}$$

gilt. Wir betrachten zunächst die linke Seite:

$$\begin{aligned} & \langle \omega_1, \omega_2 \rangle \in \big\{\omega \in \Omega \bigm| X_1(\omega) = x_1\big\} \\ \Leftrightarrow\; & X_1(\langle \omega_1, \omega_2 \rangle) = x_1 \end{aligned} \tag{2.22}$$

Jetzt formen wir die rechte Seite um:

$$\begin{aligned} & \langle \omega_1, \omega_2 \rangle \in \big\{\langle \omega_1, \omega_2 \rangle \in \Omega \bigm| \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2\big\} \\ \Leftrightarrow\; & \omega_1 \in \widehat{E}_1 \wedge \omega_2 \in \Omega_2 \\ \Leftrightarrow\; & \omega_1 \in \widehat{E}_1 \\ \Leftrightarrow\; & \omega_1 \in \big\{\omega_1 \in \Omega_1 \bigm| \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1\big\} \\ \Leftrightarrow\; & \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1 \end{aligned} \tag{2.23}$$

Um zu zeigen, dass die Bedingungen (2.23) und (2.22) äquivalent sind, bemerken wir zunächst, dass die Richtung

$$X_1(\langle \omega_1, \omega_2 \rangle) = x_1 \;\Rightarrow\; \exists \omega_3 \in \Omega_2 : X_1(\langle \omega_1, \omega_3 \rangle) = x_1$$

offensichtlich ist, denn wir können für das ω_3, dessen Existenz auf der rechten Seite gefordert wird, ja ω_2 einsetzen. Um die umgekehrte Richtung

$$\exists\omega_3 \in \Omega_2 : X_1(\langle\omega_1,\omega_3\rangle) = x_1 \;\Rightarrow\; X_1(\langle\omega_1,\omega_2\rangle) = x_1$$

zu zeigen, nehmen wir also an, dass für ein $\omega_3 \in \Omega_2$ die Gleichung

$$X_1(\langle\omega_1,\omega_3\rangle) = x_1$$

gilt. Nun folgt aus der Voraussetzung, dass X_1 durch die erste Komponente bestimmt wird,

$$X_1(\langle\omega_1,\omega_2\rangle) = X_1(\langle\omega_1,\omega_3\rangle) = x_1$$

und damit ist der Beweis abgeschlossen. □

Der nächsten Satz ist eine unmittelbare Konsequenz aus der Definition des Begriffs der unabhängigen Zufallsgrößen.

Satz 11 Es sei

1. $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ seien zwei unabhängige Zufallsgrößen.
3. $A, B \subseteq \mathbb{R}$.

Dann sind die Ereignisse

$$\bigl\{\omega \in \Omega \bigm| X(\omega) \in A\bigr\} \quad \text{und} \quad \bigl\{\omega \in \Omega \bigm| Y(\omega) \in B\bigr\}$$

unabhängig, es gilt also

$$P(X \in A \wedge Y \in B) = P(X \in A) \cdot P(Y \in B).$$

Beweis: Zunächst eine Vorbemerkung. Ist eine Menge

$$C = \{c_n \mid n \in \mathbb{N}\}$$

gegeben, wobei wir stillschweigend voraussetzen, dass die Elemente c_n voneinander verschieden sind, und ist weiter Z eine Zufallsgröße auf Ω, so kann die Wahrscheinlichkeit $P(Z \in C)$ wie folgt berechnet werden:

$$\begin{aligned} P(Z \in C) &= P\bigl(\{\omega \in \Omega \mid Z(\omega) \in C\}\bigr) \\ &= P\left(\bigcup_{n=0}^{\infty}\{\omega \in \Omega \mid Z(\omega) = c_n\}\right) \\ &= \sum_{n=0}^{\infty} P\bigl(\{\omega \in \Omega \mid Z(\omega) = c_n\}\bigr) \\ &= \sum_{n=0}^{\infty} P\bigl(Z = c_n\bigr). \end{aligned}$$

Diese Identität werden wir weiter unten benötigen.

Wir nehmen nun an, dass wir die Mengen A und B in der Form

$$A = \{a_n \mid n \in \mathbb{N}\} \quad \text{und} \quad B = \{b_n \mid n \in \mathbb{N}\},$$

schreiben können, wobei wir stillschweigend $a_m \neq a_n$ für $n \neq m$ und $b_m \neq b_n$ für $n \neq m$ voraussetzen, wir betrachten also nur den Fall, dass die beiden Mengen A und B unendlich viele Elemente enthalten. Dann gilt in Analogie zu der oben gezeigten Identität:

$$
\begin{aligned}
P(X \in A \wedge Y \in B) &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = a_m \wedge Y = b_n) \\
&= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = a_m) \cdot P(Y = b_n) && X \text{ und } Y \text{ sind unabhängig} \\
&= \sum_{m=0}^{\infty} P(X = a_m) \cdot \sum_{n=0}^{\infty} P(Y = b_n) && \text{Distributiv-Gesetz} \\
&= \left(\sum_{m=0}^{\infty} P(X = a_m)\right) \cdot \left(\sum_{n=0}^{\infty} P(Y = b_n)\right) && \text{Distributiv-Gesetz} \\
&= P(X \in A) \cdot P(Y \in B)
\end{aligned}
$$

$\square$

Genau wie wir den Begriff der Unabhängigkeit auch für mehr als zwei Ereignisse definieren konnten, so können wir auch den Begriff der Unabhängigkeit von Zufallsgrößen für mehr als zwei Zufallsgrößen definieren. Eine Menge $\{X_1, \cdots, X_n\}$ von Zufallsgrößen ist unabhängig falls für beliebige Werte $x_1, \cdots, x_n$ gilt:

$$P(X_1 = x_1 \wedge \cdots \wedge X_n = x_n) \;=\; P(X_1 = x_1) * \cdots * P(X_n = x_n).$$

2.11 Eigenschaften von Erwartungswert und Varianz

Der Erwartungswert hat die folgende Linearitäts-Eigenschaft.

Satz 12 Ist $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei Zufallsgrößen, so können wir für beliebige $\alpha, \beta \in \mathbb{R}$ eine Zufallsgröße $Z : \Omega \to \mathbb{R}$ durch

$$Z(\omega) = \alpha \cdot X(\omega) + \beta \cdot Y(\omega)$$

definieren. Für den Erwartungswert dieser Zufallsgröße gilt:

$$E\big[Z\big] = \alpha \cdot E\big[X\big] + \beta \cdot E\big[Y\big].$$

Beweis: Dieser Satz wird bewiesen, indem wir die Definition des Erwartungswerts expandieren:

$$
\begin{aligned}
E[Z] &= \sum_{\omega\in\Omega} P(\{\omega\}) \cdot Z(\omega) \\
&= \sum_{\omega\in\Omega} P(\{\omega\}) \cdot \big(\alpha \cdot X(\omega) + \beta \cdot Y(\omega)\big) \\
&= \sum_{\omega\in\Omega} P(\{\omega\}) \cdot \alpha \cdot X(\omega) + \sum_{\omega\in\Omega} P(\{\omega\}) \cdot \beta \cdot Y(\omega) \\
&= \alpha \cdot \sum_{\omega\in\Omega} P(\{\omega\}) \cdot X(\omega) + \beta \cdot \sum_{\omega\in\Omega} P(\{\omega\}) \cdot Y(\omega) \\
&= \alpha \cdot E[X] + \beta \cdot E[Y]
\end{aligned}
$$

$\square$

Aufgabe 18: Beweisen Sie: Ist $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und definieren wir für ein beliebiges $\beta \in \mathbb{R}$ die Zufallsgröße $X : \Omega \to \mathbb{R}$ als

$$X(\omega) = \beta,$$

so gilt $E[X] = \beta$. $\square$

Im Gegensatz zu dem Erwartungswert ist die Varianz kein linearer Operator. Es gilt aber der folgende Satz.

Satz 13 (Verschiebungs-Satz) Ist $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und ist $X : \Omega \to \mathbb{R}$ ein Zufallsgröße, so können wir für beliebige $\alpha, \beta \in \mathbb{R}$ eine Zufallsgröße $Z : \Omega \to \mathbb{R}$ durch

$$Z(\omega) = \alpha \cdot X(\omega) + \beta$$

definieren. Für die Varianz dieser Zufallsgröße gilt:

$$\mathrm{Var}\big[Z\big] = \alpha^2 \cdot \mathrm{Var}\big[X\big].$$

Beweis: Zur Abkürzung definieren wir zunächst $\mu := E[X]$. Nach dem gerade bewiesenen Satz gilt

$$E[\alpha \cdot X + \beta] = \alpha \cdot E[X] + \beta \cdot E[1] = \alpha \cdot \mu + \beta.$$

Nun gilt für die Varianz von Z:

$$\begin{aligned}
\mathrm{Var}[Z] &= E\big[(Z - E[Z])^2\big] \\
&= E\Big[\big(\alpha \cdot X + \beta - (\alpha \cdot \mu + \beta)\big)^2\Big] \\
&= E\Big[\big(\alpha \cdot X - \alpha \cdot \mu\big)^2\Big] \\
&= E\Big[\big(\alpha \cdot (X - \mu)\big)^2\Big] \\
&= E\big[\alpha^2 \cdot (X - \mu)^2\big] \\
&= \alpha^2 \cdot E\big[(X - \mu)^2\big] \\
&= \alpha^2 \cdot \mathrm{Var}[X]
\end{aligned}$$

□

Um den später folgenden Satz beweisen zu können, benötigen wir einen Satz über den Erwartungswerts des Produktes zweier unabhängiger Zufallsgrößen.

Lemma 14 Ist $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei **unabhängige** Zufallsgrößen, so gilt

$$E[X \cdot Y\big] = E\big[X\big] \cdot E\big[Y\big].$$

Beweis: Der Wertebereich der Zufallsgrößen X und Y sei durch die Mengen

$$X(\Omega) = \{x_n \mid n \in \mathbb{N}\} \quad \text{und} \quad Y(\Omega) = \{y_n \mid n \in \mathbb{N}\}$$

gegeben. Dann ist der Wertebereich der Zufallsgröße $Z : \Omega \to \mathbb{R}$, die durch

$$Z(\omega) = X(\omega) \cdot Y(\omega)$$

definiert ist, wie folgt gegeben:

$$Z(\Omega) = \{x_m \cdot y_n \mid m, n \in \mathbb{N}\}.$$

Damit finden wir für den Erwartungswert von Z

$$\begin{aligned}
E[Z] &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = x_m \wedge Y = y_n) \cdot x_m \cdot y_n \\
&= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P(X = x_m) \cdot P(Y = y_n) \cdot x_m \cdot y_n && X \text{ und } Y \text{ sind unabhängig} \\
&= \sum_{m=0}^{\infty} P(X = x_m) \cdot x_m \cdot \sum_{n=0}^{\infty} P(Y = y_n) \cdot y_n && \text{Distributiv-Gesetz} \\
&= \left(\sum_{m=0}^{\infty} P(X = x_m) \cdot x_m\right) \cdot \left(\sum_{n=0}^{\infty} P(Y = y_n) \cdot y_n\right) && \text{Distributiv-Gesetz} \\
&= E[X] \cdot E[Y]
\end{aligned}$$

□

Aufgabe 19: Zeigen Sie: Sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufallsgrößen, ist $\alpha \in \mathbb{R}$ und ist die Zufallsgröße $Z : \Omega \to \mathbb{R}$ definiert durch

$$Z(\omega) = X(\omega) + \alpha,$$

so sind auch die Zufallsgrößen Z und Y unabhängig. □

Der nächste Satz ist für die weitere Theorie von fundamentaler Bedeutung.

Satz 15 Ist $\langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum und sind $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei **unabhängige** Zufallsgrößen, so gilt

$$\mathrm{Var}\big[X + Y\big] = \mathrm{Var}\big[X\big] + \mathrm{Var}\big[Y\big].$$

Beweis: Zur Abkürzung setzen wir $\mu_X := E[X]$ und $\mu_Y := E[Y]$. Wegen $E[X + Y] = \mu_X + \mu_Y$ gilt dann

$$\begin{array}{rcl}
\mathrm{Var}[X+Y] & = & E\big[(X + Y - \mu_X - \mu_Y)^2\big] \\
& = & E\Big[\big((X - \mu_X) + (Y - \mu_Y)\big)^2\Big] \\
& = & E\big[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2\cdot(X - \mu_X)\cdot(Y - \mu_Y)\big] \\
& = & E\big[(X - \mu_X)^2\big] + E\big[(Y - \mu_Y)^2\big] + 2\cdot E\big[(X - \mu_X)\cdot(Y - \mu_Y)\big] \\
& = & \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot E\big[(X - \mu_X)\cdot(Y - \mu_Y)\big] \\
& = & \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot E\big[(X - \mu_X)\big]\cdot E\big[Y - \mu_Y)\big] \\
& & \text{nach Lemma 14, denn } X - \mu_X \text{ und } Y - \mu_Y \text{ sind unabhängig nach Aufgabe 19} \\
& = & \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot\big(E[X] - \mu_X\big)\cdot\big(E[Y] - \mu_Y\big) \\
& = & \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\cdot\big(\mu_X - \mu_X\big)\cdot\big(\mu_Y - \mu_Y\big) \\
& = & \mathrm{Var}[X] + \mathrm{Var}[Y].
\end{array}$$

$\square$

Der letzte Satz läßt sich auf Mengen von unabhängigen Zufallsgrößen verallgemeinern. Ist $\{X_1, \cdots, X_n\}$ eine Menge unabhängiger Zufallsgrößen, so gilt

$$\mathrm{Var}\left(\sum_{i=0}^{n} X_i\right) = \sum_{i=0}^{n} \mathrm{Var}(X_i).$$

In dieser Form hat der Satz eine äußerst wichtige Konsequenz. Wir gehen aus von einem Zufalls-Experiment, dass durch einen Wahrscheinlichkeits-Raum der Form

$$\mathcal{W} = \langle\Omega, 2^\Omega, P\rangle$$

beschrieben wird. Weiter sei X eine Zufallsgröße, die bei diesem Experiment eine Rolle spielt. Wiederholen wir das Zufalls-Experiment n-mal, so haben wir eine Menge $\{X_1, \cdots, X_n\}$ von n Zufallsgrößen, die offenbar unabhägig sind. Wir definieren jetzt die Zufallsgröße

$$\overline{X} := \frac{1}{n}\sum_{k=1}^{n} X_k.$$

$\overline{X}$ ist offenbar das arithmetische Mittel der Zufallsgrößen $X_1, \cdots, X_n$. Für den Erwartungswert von $\overline{X}$ gilt dann

$$\begin{array}{rcl}
E\Big[\overline{X}\Big] & = & E\left[\frac{1}{n}\cdot\sum\limits_{k=1}^{n} X_k\right] \\
& = & \frac{1}{n}\cdot\sum\limits_{k=1}^{n} E\left[X_k\right] \\
& = & \frac{1}{n}\cdot\sum\limits_{k=1}^{n} E[X] \\
& = & \frac{1}{n}\cdot E[X]\cdot\sum\limits_{k=1}^{n} 1 \\
& = & \frac{1}{n}\cdot E[X]\cdot n \\
& = & E[X]
\end{array}$$

Dabei haben wir ausgenutzt, dass die Zufallsgrößen X_i alle den selben Erwartungswert $E[X]$ haben. Natürlich ist es nicht weiter verwunderlich, dass das arithmetische Mittel dann ebenfalls den Erwartungswert $E[X]$ hat. Wesentlich interessanter ist die Frage nach der Varianz der Zufallsgröße $\overline{X}$. Es gilt

$$
\begin{aligned}
\operatorname{Var}\left[\overline{X}\right] &= \operatorname{Var}\left[\frac{1}{n}\cdot\sum_{k=1}^{n} X_i\right] \\
&= \left(\frac{1}{n}\right)^2 \cdot \operatorname{Var}\left[\sum_{k=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\cdot\sum_{k=1}^{n}\operatorname{Var}\left[X_i\right] \\
&= \frac{1}{n^2}\cdot\sum_{k=1}^{n}\operatorname{Var}\left[X\right] \\
&= \frac{1}{n^2}\cdot\operatorname{Var}\left[X\right]\cdot\sum_{k=1}^{n} 1 \\
&= \frac{1}{n^2}\cdot\operatorname{Var}\left[X\right]\cdot n \\
&= \frac{1}{n}\cdot\operatorname{Var}\left[X\right]
\end{aligned}
$$

Die Standard-Abweichung σ ist die Wurzel der Varianz, es gilt also

$$\boxed{\sigma\left[\overline{X}\right] = \frac{1}{\sqrt{n}}\cdot\sigma[X].}$$

Dieser Zusammenhang wird in der Literatur als das $\sqrt{n}$-Gesetz bezeichnet:

Satz 16 ($\sqrt{n}$-Gesetz) Es sei

1. $\mathcal{W} = \langle\Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to \mathbb{R}$ eine Zufallsgröße auf $\mathcal{W}$,
3. $\mathcal{W}^n = \underbrace{\mathcal{W}\times\cdots\times\mathcal{W}}_{n}$ sei das n-fache kartesisches Produkt von $\mathcal{W}$ mit sich selbst,
4. $X_i : \Omega^n \to \mathbb{R}$ sei für $i = 1,\cdots,n$ definiert durch
   $$X_i\big(\langle\omega_1,\cdots,\omega_n\rangle\big) : X(\omega_i)$$
5. $\overline{X} : \Omega^n \to \mathbb{R}$ sei definiert durch
   $$\overline{X}(\omega) := \frac{1}{n}\cdot\sum_{k=1}^{n} X_i(\omega)$$

Dann gilt:

1. $E\left[\overline{X}\right] = E[X]$
2. $\operatorname{Var}\left[\overline{X}\right] = \frac{1}{n}\cdot\operatorname{Var}[X] \quad$ und $\quad \sigma\left[\overline{X}\right] = \frac{1}{\sqrt{n}}\cdot\sigma[X]$ □

Das $\sqrt{n}$-Gesetz hat in der Praxis eine sehr wichtige Anwendung bei der Messung. Wird eine physikalische Größe gemessen, so können wir diese Messung als Zufalls-Experiment ansehen. Der Erwartungswert des Zufalls-Experiments ist dann der tatsächliche Wert, während die Standard-Abweichung als die ungefähre Größe des Meßfehlers interpretiert werden kann. Das $\sqrt{n}$-Gesetz zeigt uns eine Möglichkeit um die Genauigkeit unserer Messung zu verbessern: Führen wir die selbe Messung n-mal durch und bilden den Mittelwert unserer Ergebnisse, so wird die Meß-Genauigkeit um den Faktor $\sqrt{n}$ gesteigert. Ein anderes Beispiel ist das Würfeln: Gibt die Zufallsgröße X die

Augenzahl beim einmaligen Würfeln an, so haben wir für Varianz und Standard-Abweichung die Werte

$$\mathrm{Var}[X] = \tfrac{35}{12} \quad \text{und} \quad \sigma[X] = \sqrt{\tfrac{35}{12}} \approx 1.707825128.$$

Würfeln wir hingegen 100-mal und bilden das arithmetische Mittel $\overline{X}$ aller gewürfelten Augenzahlen, so finden wir für die Standard-Abweichung den Wert

$$\sigma\Big[\overline{X}\Big] = 0.1707825128$$

Im nächsten Abschnitt werden wir sehen, welche konkrete Aussagen über die Wahrscheinlichkeit aus der Standard-Abweichung abgeleitet werden können.

2.12 Der Satz von Tschebyschow und das Gesetz der großen Zahlen

Der folgende Satz, der nach Pafnuti Lwowitsch Tschebyschow (1821 - 1894) benannt ist, zeigt uns, welche konkrete Bedeutung die Standard-Abweichung $\sigma[X]$ hat.

Satz 17 (Ungleichung von Tschebyschow) Es sei

1. $\langle \Omega, 2^\Omega, P\rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to \mathbb{R}$ eine Zufallsgröße mit Erwartungswert $\mu = E[X]$ und Standard-Abweichung $\sigma^2 = \mathrm{Var}[X]$.

Dann gilt für alle $r \in \mathbb{R}_+$

$$P\big(|X-\mu| \geq r \cdot \sigma\big) \leq \frac{1}{r^2}.$$

Beweis: Zunächst formen wir die Bedingung $|X-\mu| \geq r\cdot\sigma$ um. Es gilt offenbar für beliebige $\omega \in \Omega$

$$|X(\omega)-\mu| \geq r\cdot\sigma \;\Leftrightarrow\; \big(X(\omega)-\mu\big)^2 \geq r^2\cdot\sigma^2 \;\Leftrightarrow\; \left(\frac{X(\omega)-\mu}{\sigma}\right)^2 \geq r^2$$

Wir zeigen jetzt zunächst die Ungleichung

$$r^2 \cdot P\big(|X-\mu| \geq r\cdot\sigma\big) \leq 1.$$

Wenn wir diese Ungleichung durch r^2 dividieren, folgt die Behauptung.

$$\begin{array}{lcl}
r^2 \cdot P\big(|X-\mu| \geq r\cdot\sigma\big) & = & r^2 \cdot P\big(\,(X-\mu)^2 \geq r^2\cdot\sigma^2\big) \\
& = & r^2 \cdot \displaystyle\sum_{\left\{\omega\in\Omega \,\middle|\, (\frac{X-\mu}{\sigma})^2 \geq r^2\right\}} P\big(\{\omega\}\big) \\
& = & \displaystyle\sum_{\left\{\omega\in\Omega \,\middle|\, (\frac{X-\mu}{\sigma})^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot r^2 \\
& \leq & \displaystyle\sum_{\left\{\omega\in\Omega \,\middle|\, (\frac{X-\mu}{\sigma})^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot \left(\frac{X-\mu}{\sigma}\right)^2
\end{array}$$

Im letzten Schritt haben wir ausgenutzt, dass die Summe nur über die ω läuft, für die $\left(\frac{X-\mu}{\sigma}\right)^2 \geq r^2$ gilt, so dass die Summe größer wird, wenn wir r^2 durch $\left(\frac{X-\mu}{\sigma}\right)^2$ ersetzen. Im nächsten Schritt vergrößern wir die Summe, indem wir über alle $\omega \in \Omega$ summieren:

$$
\begin{aligned}
r^2 \cdot P\big(|X-\mu| \geq r\cdot\sigma\big) &\leq \sum_{\left\{\omega\in\Omega \,\middle|\, \left(\frac{X-\mu}{\sigma}\right)^2 \geq r^2\right\}} P\big(\{\omega\}\big)\cdot\left(\frac{X-\mu}{\sigma}\right)^2 \\
&\leq \sum_{\omega\in\Omega} P\big(\{\omega\}\big)\cdot\left(\frac{X-\mu}{\sigma}\right)^2 \\
&= \frac{1}{\sigma^2}\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\cdot(X-\mu)^2 \\
&= \frac{1}{\sigma^2}\cdot \mathrm{Var}[X] \\
&= 1
\end{aligned}
$$
□

Aufgabe 20: Schätzen Sie mit Hilfe der Ungleichung von Tschebyschow ab, wie oft Sie würfeln müssen, damit die Wahrscheinlichkeit, dass der arithmetische Mittelwert der Augenzahlen um mehr als 0.1 vom Erwartungswert abweicht, kleiner als 1% ist?

Lösung: Zunächst formalisieren wir die Aufgabenstellung. Gesucht ist ein n, so dass für den arithmetischen Mittelwert

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_i$$

die Beziehung

$$P\big(|\overline{X}-\mu| \geq 0.1\big) \leq 0.01$$

gilt. Um diesen Ausdruck auf die Ungleichung von Tschebyschow zurück führen zu können, formen wir den Wert 0.1 um:

$$0.1 = 0.1\cdot\frac{1}{\sigma}\cdot\sigma \quad \text{und definieren } r = 0.1\cdot\frac{1}{\sigma}$$

Damit können wir die obige Ungleichung auch in der Form

$$P\big(|\overline{X}-\mu| \geq r\cdot\sigma\big) \leq 0.01$$

schreiben. Damit diese Ungleichung aus der Ungleichung von Tschebyschow folgt, muss

$$\frac{1}{r^2} \leq 0.01$$

gelten. Setzen wir hier den Wert für r ein, so haben wir

$$10^2\cdot\sigma^2 \leq 0.01, \quad \text{also} \quad 10\cdot\sigma \leq 0.1, \quad \text{also} \quad \sigma \leq \tfrac{1}{100}.$$

Die Standard-Abweichung σ können wir aus dem $\sqrt{n}$-Gesetz bestimmen, denn da die Augenzahl beim Würfeln mit einem Würfel die Varianz $\frac{35}{12}$ hat, gilt für die Varianz beim n-maligen Würfeln

$$\sigma = \sqrt{\frac{1}{n}\cdot\frac{35}{12}}$$

Also haben wir die Ungleichung

$$
\begin{aligned}
& \sqrt{\frac{1}{n}\cdot\frac{35}{12}} \leq \frac{1}{100} \\
\Leftrightarrow\quad & \frac{1}{n}\cdot\frac{35}{12} \leq \frac{1}{10\,000} \\
\Leftrightarrow\quad & 10\,000\cdot\frac{35}{12} \leq n \\
\Leftrightarrow\quad & 29166.\overline{6}\cdots \leq n
\end{aligned}
$$

Also müssen wir 29167 mal würfeln. □

Aufgabe 21: Angenommen, wir würfeln n mal mit einem Laplace-Würfel. Bestimmen Sie mit Hilfe der Ungleichung von Tschebyschow ein n so, dass Sie mit einer Wahrscheinlichkeit von 99%

davon ausgehen können, dass mindestens $\frac{n}{7}$ mal eine Sechs gewürfelt wird?
Bemerkung: Sie können diese Aufgabe auch mit Hilfe der Näherung von Laplace für die (kumulative) Verteilungsfunktion lösen. Dann ergibt sich ein kleinerer Wert. Das ist nicht weiter verwunderlich, denn die Tschebyschow-Ungleichung gilt für beliebige Verteilungen, während die Näherung von Laplace nur gilt, wenn die Verteilungsfunktion die Form

$$\sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

hat.

Wenn wir ein Zufalls-Experiment nur oft genug durchführen und dabei eine Zufallsgröße X ermitteln, so wird die Wahrscheinlichkeit, dass der arithmetische Mittelwert, den wir bei n Messungen finden, beliebig klein, wenn wir nur n groß genug wählen. Genauer gilt:

Satz 18 (Schwaches Gesetz der großen Zahlen) Es sei

1. $\mathcal{W} = \langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to \mathbb{R}$ eine Zufallsgröße auf $\mathcal{W}$ mit dem Erwartungswert $\mu = E[X]$ und Standard-Abweichung $\sigma := \sqrt{\mathrm{Var}[X]}$,
3. $\mathcal{W}^n = \underbrace{\mathcal{W} \times \cdots \times \mathcal{W}}_{n}$ sei das n-fache kartesisches Produkt von $\mathcal{W}$ mit sich selbst,
4. $X_i : \Omega^n \to \mathbb{R}$ sei für $i = 1, \cdots, n$ definiert durch
$$X_i\big(\langle \omega_1, \cdots, \omega_n \rangle\big) := X(\omega_i)$$
5. $\overline{X} : \Omega^n \to \mathbb{R}$ sei definiert durch
$$\overline{X}(\omega) := \frac{1}{n} \cdot \sum_{i=1}^{n} X_i(\omega)$$

Dann gilt für alle $\varepsilon > 0$:

$$\lim_{n\to\infty} P\big(|\overline{X} - \mu| \geq \varepsilon\big) = 0.$$

Beweis: Nach dem $\sqrt{n}$-Gesetz haben wir für die Zufallsgröße $\overline{X}$:

$$E\Big[\overline{X}\Big] = \mu \quad \text{und} \quad \mathrm{Var}\Big[\overline{X}\Big] = \frac{\sigma^2}{n}, \quad \text{also } \sigma\Big[\overline{X}\Big] = \frac{\sigma}{\sqrt{n}}.$$

Definieren wir nun

$$r := \frac{\varepsilon}{\sigma} \cdot \sqrt{n}, \quad \text{so gilt } \varepsilon = r \cdot \frac{\sigma}{\sqrt{n}}$$

und mit nach der Ungleichung von Tschebyschow folgt

$$P\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{1}{\frac{\varepsilon^2}{\sigma^2} \cdot n}, \quad \text{also } P\Big(|\overline{X} - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Wegen $\lim\limits_{n\to\infty} \dfrac{\sigma^2}{n \cdot \varepsilon^2} = 0$ folgt die Behauptung. □

2.13 Erwartungswert und Varianz der Binomial-Verteilung

In diesem Abschnitt wollen wir Erwartungswert und Varianz einer binomial-verteilten Zufalls-Größe berechnen. Damit dies einfach möglich ist, präsentieren wir zunächst einen alternativen Zugang zur Binomial-Verteilung. Dazu betrachten wir das folgende Beispiel.

Beispiel: Ein Betrunkener befindet sich in einer Stadt, in der die Straßen ein quadratisches Muster bilden. Dementsprechend lassen sich alle Kreuzungen durch Angabe zweier natürlicher Zahlen k und l spezifizieren, die die Koordinaten dieser Kreuzung angeben. Wir nehmen nun an, dass der Betrunkene zunächst an der Kreuzung steht, die durch das Paar $\langle 0, 0\rangle$ spezifiziert wird. Der Betrunkene geht jetzt zufällig los, wobei er mit der Wahrscheinlichkeit p nach Osten und mit der Wahrscheinlichkeit $1 - p$ nach Norden geht. Jedesmal, wenn der Betrunkene wieder an eine Kreuzung kommt, geht er wieder mit der Wahrscheinlichkeit p nach Osten und $(1 - p)$ nach Norden. Beträgt der Abstand zwischen zwei Kreuzungen eine Längeneinheit, so legt der Betrunkene insgesamt einen Weg von n Längeneinheiten zurück. Wir stellen uns die Frage, mit welcher Wahrscheinlichkeit der Betrunkene dann an der Kreuzung mit den Koordinaten $\langle k, n-k\rangle$ ankommt.

Um die Kreuzung $\langle k, n-k\rangle$ zu erreichen muss der Betrunkene k-mal nach Osten und $n-k$-mal nach Norden gegangen sein. Da wir davon ausgehen, dass die einzelnen Entscheidungen, die der Betrunkene an den Kreuzungen trifft, voneinander unabhängig sind, hat ein bestimmter Weg zur Kreuzung $\langle k, n - k\rangle$ die Wahrscheinlichkeit

$$p^k \cdot (1 - p)^{n-k}.$$

Das Problem ist, dass es im Normalfall mehrere Wege gibt, die von der Kreuzung $\langle 0, 0\rangle$ starten, bei der Kreuzung $\langle k, n - k\rangle$ enden und insgesamt eine Länge von n haben. Um die Gesamt-Wahrscheinlichkeit zu berechnen, müssen wir die Wahrscheinlichkeit aller Wege aufsummieren. Da die Wahrscheinlichkeit für jeden Weg die selbe ist, reicht es aus, wenn wir die Anzahl der Wege der Länge n bestimmten, die von $\langle 0, 0\rangle$ nach $\langle k, n - k\rangle$ führen. Wir definieren

$$s(n, k) := \text{Anzahl der Wege der Länge } n \text{ von } \langle 0, 0\rangle \text{ nach } \langle k, n - k\rangle.$$

Wir berechnen die Funktion $s(n, k)$ durch Induktion über n.

I.A.: $n = 0$. Es gibt genau einen Weg (der Länge 0) von $\langle 0, 0\rangle$ nach $\langle 0, 0\rangle$. Also gilt

$$s(0, 0) := 1.$$

I.S.: $n \mapsto n + 1$. Die Kreuzung $\langle k, n + 1 - k\rangle$ kann der Betrunkene auf zwei Arten erreichen. Entweder er kommt von der Kreuzung $\langle k, n - k\rangle$ und geht von da aus nach Norden, oder er kommt von der Kreuzung $\langle k-1, n+1-k\rangle$ und geht von da aus nach Osten. Die Gesamtzahl aller Wege ergibt sich, wenn wir zu der Anzahl der Wege, die nach $\langle k, n\rangle$ führen, die Anzahl der Wege, die nach $\langle k - 1, n + 1 - k\rangle$ führen, hinzu addieren. Wegen $\langle k - 1, n + 1 - k\rangle = \langle k - 1, n - (k - 1)\rangle$ gilt also

$$s(n + 1, k) = s(n, k) + s(n, k - 1).$$

Wir werden zeigen, dass für alle $k, n \in \mathbb{N}$

$$s(n, k) = \binom{n}{k}$$

gilt. Dazu benötigen wir den folgenden Hilfssatz.

Lemma 19 (Pascal'sche Regel) Die durch

$$\binom{n}{k} := \frac{n!}{k! \cdot (n - k)!}$$

definierten Binomial-Koeffizienten genügen der Rekurrenz-Gleichung

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

Beweis: Der Beweis ergibt sich durch eine einfache Expansion der Definition.

$$
\begin{aligned}
\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)! \cdot (n-(k-1))!} + \frac{n!}{k! \cdot (n-k)!} \\
&= \frac{n!}{(k-1)! \cdot (n+1-k))!} + \frac{n!}{k! \cdot (n-k)!} \\
&= n! \cdot \left(\frac{k}{k! \cdot (n+1-k)!} + \frac{n+1-k}{k! \cdot (n+1-k)!} \right) \\
&= n! \cdot \frac{k+n+1-k}{k! \cdot (n+1-k)!} \\
&= n! \cdot \frac{n+1}{k! \cdot (n+1-k)!} \\
&= \frac{(n+1)!}{k! \cdot (n+1-k)!} \\
&= \binom{n+1}{k} \qquad \square
\end{aligned}
$$

Damit können wir jetzt das obige Beispiel zu Ende führen und zeigen, dass für die Anzahl der Wege

$$
s(n,k) = \frac{n!}{k! \cdot (n-k)!}
$$

gilt. Wir führen den Nachweis durch Induktion nach n.

I.A.: $n = 0$. Der einzige mögliche Wert für k ist $k = 0$, folglich gilt

$$
s(0,0) = 1.
$$

Andererseits haben wir

$$
\binom{0}{0} = \frac{0!}{0! \cdot 0!} = \frac{1}{1 \cdot 1} = 1.
$$

I.S.: $n \mapsto n+1$. Es gilt

$$
\begin{aligned}
s(n+1,k) &= s(n,k) + s(n,k-1) \\
&\stackrel{IV}{=} \binom{n}{k} + \binom{n}{k-1} \\
&= \binom{n+1}{k} \qquad \text{nach der Pascal'schen Regel} \qquad \square
\end{aligned}
$$

Damit gilt für die Wahrscheinlichkeit $P\big(\{\langle k, n-k\rangle\}\big)$, dass der Betrunkene nach einem Weg der Länge n an der Kreuzung $\langle k, n-k\rangle$ landet, die Formel

$$
P\big(\{\langle k, n-k\rangle\}\big) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = B(n,p;k).
$$

Wir abstrahieren jetzt von dem Alkohol und destilieren die mathematische Essenz des Beispiels. Dazu definieren wir zunächst den Begriff des *Bernoulli-Experiments* (Jakob Bernoulli; 1654 - 1705).

Definition 20 (Bernoulli-Experiment) Wir nennen ein Zufalls-Experiment ein *Bernoulli-Experiment* wenn es nur zwei mögliche Ergebnisse des Experiments gibt. Der Wahrscheinlichkeits-Raum kann dann in der Form

$$
\mathcal{W} = \big\langle \{0,1\}, 2^{\{0,1\}}, P \big\rangle
$$

geschrieben werden. Dann setzen wir

$$
p := P\big(\{1\}\big)
$$

und nennen p den *Parameter* des Bernoulli-Experiments. $\square$

Wird ein Bernoulli-Experiment n-mal durchgeführt, so sprechen wir von einer *Bernoulli-Kette*. Ist $\mathcal{W}_B$ der Wahrscheinlichkeits-Raum des einzelnen Bernoulli-Experiments, so ist das n-fache Produkt

$$\mathcal{W} = \mathcal{W}_B^n$$

der Wahrscheinlichkeits-Raum der Bernoulli-Kette. Definieren wir für das i-te Bernoulli-Experiment eine Zufalls-Größe

$$X_i : \{0,1\} \to \mathbb{R} \quad \text{durch} \quad X_i(\omega) := \omega,$$

so können wir den Erwartungswert und die Varianz von X_i sehr einfach berechnen. Es gilt

$$\begin{aligned} E[X_i] &= \sum_{\omega\in\{0,1\}} X_i(\omega) \cdot P\bigl(\{\omega\}\bigr) \\ &= X_i(0) \cdot P\bigl(\{0\}\bigr) + X_i(1) \cdot P\bigl(\{1\}\bigr) \\ &= 0 \cdot (1-p) + 1 \cdot p \\ &= p \end{aligned}$$

Für die Varianz finden wir

$$\begin{aligned} \mathrm{Var}[X_i] &= E\Bigl[\bigl(X_i - E[X_i]\bigr)^2\Bigr] \\ &= E\Bigl[\bigl(X_i - p\bigr)^2\Bigr] \\ &= \sum_{\omega\in\{0,1\}} (X_i(\omega) - p)^2 \cdot P\bigl(\{\omega\}\bigr) \\ &= (X_i(0) - p)^2 \cdot P\bigl(\{0\}\bigr) + (X_i(1) - p)^2 \cdot P\bigl(\{1\}\bigr) \\ &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= \bigl(p + (1-p)\bigr) \cdot p \cdot (1-p) \\ &= p \cdot (1-p) \end{aligned}$$

Als nächstes definieren wir für die Bernoulli-Kette die Zufalls-Größe

$$X : \{0,1\}^n \to \mathbb{R} \quad \text{durch} \quad X\bigl([\omega_1, \cdots, \omega_n]\bigr) = \sum_{i=1}^{n} X_i(\omega_i).$$

Das Beispiel oben zeigt, dass die Zufalls-Größe X binomial-verteilt ist, es gilt

$$P(X = k) = B(n, p; k).$$

Da X die Summe von n Zufalls-Größen ist, gilt für den Erwartungswert

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = n \cdot p.$$

Da diese Zufalls-Größen unabhängig sind, haben wir im letzten Abschnitt gesehen, dass wir auch die Varianz als Summe der Varianzen der einzelnen Zufalls-Größen berechnen können. Also gilt

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} p \cdot (1-p) = n \cdot p \cdot (1-p).$$

Aufgabe 22: Die Augenfarbe wird durch ein einzelnes Paar von Genen vererbt. Jeder Mensch hat zwei dieser Gene. Die Augenfarbe braun ist dominant, d.h. wenn Sie auch nur ein Gen für braune Augen haben, dann sind ihre Augen braun. Nur wenn Sie zwei Gene für blaue Augen haben, haben Sie blaue Augen. Nehmen Sie nun an, dass eine Familie 6 Kinder hat und dass der älteste Sohn blaue Augen hat, während die Eltern beide braune Augen haben. Wir groß ist die Wahrscheinlichkeit, dass insgesamt genau drei Kinder blaue Augen haben?

2.14 Die Poisson-Verteilung

In diesem Abschnitt betrachten wir einen Spezialfall der Binomial-Verteilung, der in der Praxis häufig auftritt. Dieser Spezialfall ist dadurch gekennzeichnet, dass in der Formel

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

einerseits der Wert von n sehr groß wird, andererseits aber der Wert von p sehr klein ist. Wir illustrieren diesen Spezialfall durch ein Beispiel.

Beispiel: Die (hypothetische) Firma *Diamond-Connections* habe 10 Millionen Kunden. In einem vorgegebenen Zeitintervall von, sagen wir mal, drei Minuten ruft jeder dieser Kunden mit einer Wahrscheinlichkeit p im Call-Center der Firma an. Da die meisten Kunden etwas anderes zu tun haben als im Call-Center anzurufen, kommen im Schnitt pro Zeitintervall fünf Anrufe an, die Wahrscheinlichkeit p hat also den Wert

$$p = \frac{5}{10\,000\,000} = \frac{1}{2\,000\,000}$$

Unser Ziel ist es zu berechnen, wieviele Mitarbeiter die Firma *Diamond-Connections* einstellen muss, damit sicher gestellt ist, dass die Wahrscheinlichkeit, dass kein Platz des Call-Centers mehr frei ist, kleiner als 1% ist. Wir wollen zur Vereinfachung weiter voraussetzen, dass alle Gespräche genau ein Zeitintervall andauern und dass zusätzlich alle Gespräche am Anfang eines Zeitintervalls beginnen.

Wir berechnen zunächst die Wahrscheinlichkeit, dass in einem gegebenen Zeitintervalls k Teilnehmer anrufen. Setzen wir voraus, dass verschiedene Teilnehmer unabhängig sind, so ist die Zufalls-Größe K, die die Anzahl der Teilnehmer angibt, binomial-verteilt, es gilt also

$$\begin{aligned}
P(K = k) &= B(n, p; k) \\
&= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k}
\end{aligned}$$

Für die Werte von n, die in der Praxis in Frage kommen, können wir die Formel in der oben angegebenen Form nicht auswerten. Auch die Approximations-Formel von Laplace bringt uns an dieser Stelle nicht weiter, denn diese Formel liefert nur dann brauchbare Ergebnisse, wenn die Bedingung $n \cdot p \cdot (1-p) > 9$ erfüllt ist. Um hier weiterzukommen, führen wir die Abkürzung $\lambda = n \cdot p$ ein, λ ist also der Erwartungswert der Zufalls-Größe K. Dann gilt $p = \frac{\lambda}{n}$ und wenn wir diesen Wert in die obere Formel einsetzen, erhalten wir

$$\begin{aligned}
P(K = k) &= \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \\
&= \frac{n!}{k! \cdot (n-k)!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{1}{k!} \cdot \frac{n * (n-1) * \cdots * \big(n-(k-1)\big)}{n^k} \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{1}{k!} \cdot 1 * \left(1 - \frac{1}{n}\right) * \cdots * \left(1 - \frac{k-1}{n}\right) \cdot \lambda^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}
\end{aligned}$$

Wir überlegen uns nun, wie sich dieser Ausdruck verhält, wenn n sehr große Werte annimmt. Zunächst haben wir

$$\lim_{n\to\infty} 1 - \frac{i}{n} = 1 \qquad \text{für alle } i = 1, \cdots, k-1, \text{ also auch}$$

$$\lim_{n\to\infty} 1 * \left(1 - \frac{1}{n}\right) * \cdots * \left(1 - \frac{k-1}{n}\right) = 1.$$

Weiter gilt

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Diese Gleichung haben wir im zweiten Semester gezeigt, indem wir die Umformungen

$$\left(1 - \frac{\lambda}{n}\right)^n = \exp\left(\ln\left(1 - \frac{\lambda}{n}\right)^n\right) = \exp\left(n \cdot \ln\left(1 - \frac{\lambda}{n}\right)\right) = \exp\left(\frac{\ln\left(1 - \frac{\lambda}{n}\right)}{\frac{1}{n}}\right)$$

benutzt haben. Anschließend konnten wir den Grenzwert mit Hilfe der Regel von L'Hospital berechnen. Schließlich gilt auch

$$\lim_{n\to\infty} 1 - \frac{\lambda}{n} = 1 \quad \text{und also auch} \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1.$$

Damit haben wir für große n (und kleine p) jetzt die Näherung

$$P(K = k) = B(n, p; k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda} \tag{2.24}$$

gefunden. Die durch Gleichung (2.24) definierte Wahrscheinlichkeits-Verteilung heißt Poisson-Verteilung nach Simón Denis Poisson (1781 - 1840).

Jetzt können wir die eingangs gestellte Frage nach der Anzahl der Mitarbeiter des Call-Centers beantworten. Hat das Call-Center m Mitarbeiter, so können auch nur m Anrufer bedient werden. Wir müssen daher m so groß wählen, dass die Wahrscheinlichkeit, dass mehr als m Kunden in einem Zeitintervall anrufen, kleiner als 1% ist. Es gilt

$$P(K > m) = 1 - P(K \leq m) = 1 - \sum_{k=0}^{m} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = 1 - e^{-\lambda} \cdot \sum_{k=0}^{m} \frac{\lambda^k}{k!}$$

Für $\lambda = n \cdot p = 10\,000\,000 \cdot \frac{5}{10\,000\,000} = 5$ finden wir mit *Maple* folgende Zahlen:

$$P(K > 10) \approx 0.0136952688, \quad \text{aber } P(K > 11) \approx 0.0054530920.$$

Die Wahrscheinlichkeit, dass mehr als 11 Kunden in einem Zeitintervall anrufen, liegt also bei 0.5% und damit reichen 11 Mitarbeiter aus. □

Die obigen Betrachtungen des Call-Centers sind aus mehreren Gründen unrealistisch. Das größte Problem in unserer Modellierung des Call-Centers ist die Annahme, dass die einzelnen Kunden ihre Anrufe unabhängig von einander tätigen. Das ist zum Beispiel dann sicher nicht mehr der Fall, wenn beispielsweise durch einen Fehler in der Software zur Erstellung von Rechnungen ein Fehler auftritt, denn in einer solchen Situation werden schlagartig sehr viele Kunden im Call-Center anrufen. Phänomene wie die oben beschriebenen werden in der *Theorie der Warteschlangen* genauer untersucht, siehe z.B. [GH85] oder *Wikipedia*, Stichwort *Warteschlangentheorie*.

Aufgabe 23: In Singapur gibt es pro Monat durchschnittlich drei Exekutionen. Aus tiefer Reue über ihre Missetaten stellen alle Delinquenten ihre Organe zur Verfügung. Bestimmen Sie die Wahrscheinlichkeit, dass der örtlichen Organhändler seinem amerikanischen Geschäftspartner im nächsten Monat mindestens drei Sätze Innereien anbieten kann.

2.14.1 Erwartungswert und Varianz einer Poisson-verteilten Zufalls-Größe

Es sei $\langle \Omega, 2^\Omega, P \rangle$ ein Wahrscheinlichkeits-Raum $K : \Omega \to \mathbb{N}$ eine Zufalls-Größe, die Poisson-verteilt ist, für die also

$$P(K = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

gilt. Wir wollen nun den Erwartungswert und die Varianz von K berechnen. Am einfachsten geht das, wenn wir uns erinnern, wie wir die Poisson-Verteilung hergeleitet haben: Die Poisson-Verteilung ist aus der Binomial-Verteilung hergeleitet worden, indem wir dort $\lambda = n \cdot p$ gesetzt und dann n gegen Unendlich laufen gelassen haben. Dann gilt $p = \frac{\lambda}{n}$ und wir haben

$$\lim_{n\to\infty} B\Big(n, \frac{\lambda}{n}; k\Big) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

Definieren wir also binomial-verteilte Zufalls-Größen X_n so, dass

$$P(X_n = k) = B\Big(n, \frac{\lambda}{n}; k\Big),$$

dann sollte gelten

$$E[K] = \lim_{n\to\infty} E[X_n] \quad \text{und} \quad \mathrm{Var}[K] = \lim_{n\to\infty} \mathrm{Var}[X_n].$$

Erwartungswert und Varianz einer binomial-verteilten Zufalls-Größe X haben wir im letzen Abschnitt berechnet und $E[X] = n \cdot p$ und $\mathrm{Var}[X] = n \cdot p \cdot (1 - q)$ gefunden. Also haben wir

$$E[K] = \lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} n \cdot \frac{\lambda}{n} = \lambda.$$

Für die Varianz finden wir

$$\mathrm{Var}[K] = \lim_{n\to\infty} \mathrm{Var}[X_n] = \lim_{n\to\infty} n \cdot \frac{\lambda}{n} \cdot \left(1 - \frac{\lambda}{n}\right) = \lambda \cdot \left(1 - \lim_{n\to\infty} \frac{\lambda}{n}\right) = \lambda$$

Also haben sowohl der Erwartungswert als auch die Varianz einer Poisson-verteilten Zufalls-Größe den Wert λ.

Aufgabe 24: Berechnen Sie den Erwartungswert und die Varianz einer Poisson-verteilten Zufalls-Größe durch direkte Anwendung der Definition von Erwartungswert und Varianz ohne auf die Binomial-Verteilung zurückzugreifen.

2.14.2 Die Summe Poisson-verteilter Zufalls-Größen

Wir betrachten zwei unabhängige Zufalls-Größen X_1 und X_2, die beide Poisson-verteilt sind mit den Parametern λ_1 und λ_2, es gilt also

$$P(X_1 = k) = \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_1} \quad \text{und} \quad P(X_2 = k) = \frac{\lambda_2^k}{k!} \cdot e^{-\lambda_2}.$$

Wir berechnen die Verteilung der Zufalls-Größe $Z := X_1 + X_2$. Es gilt

$$\begin{aligned}
P(Z = k) &= P(X_1 + X_2 = k) \\
&= \sum_{i=0}^{k} P(X_1 = i \wedge X_2 = k - i) \\
&= \sum_{i=0}^{k} P(X_1 = i) \cdot P(X_2 = k - i) \\
&\ \text{wegen der Unabhängigkeit von } X_1 \text{ und } X_2 \\
&= \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} \cdot e^{-\lambda_1} \cdot \frac{\lambda_2^{k-i}}{(k-i)!} \cdot e^{-\lambda_2} \\
&= e^{-\lambda_1 - \lambda_2} \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} \cdot \frac{\lambda_2^{k-i}}{(k-i)!} \\
&= e^{-\lambda_1 - \lambda_2} \cdot \frac{1}{k!} \cdot \sum_{i=0}^{k} \frac{k!}{i! \cdot (k-i)!} \cdot \lambda_1^i \cdot \lambda_2^{k-i} \\
&= e^{-\lambda_1 - \lambda_2} \cdot \frac{1}{k!} \cdot \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda_1^i \cdot \lambda_2^{k-i} \\
&= e^{-\lambda_1 - \lambda_2} \cdot \frac{1}{k!} \cdot (\lambda_1 + \lambda_2)^k
\end{aligned}$$

Dabei haben wir im letzten Schritt den im ersten Semester in der Form

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} \cdot x^i \cdot y^{n-i}$$

durch Induktion über n bewiesenen Binomischen Lehrsatz verwendet. Unsere Herleitung zeigt, dass für unabhängige Poisson-verteilte Zufalls-Größen mit dem Erwartungswerten λ_1 und λ_2 auch die Summe $X_1 + X_2$ Poisson-verteilt ist mit dem Erwartungswert $\lambda_1 + \lambda_2$.

Kapitel 3

Stetige Zufalls-Größen

Bisher haben wir nur mit diskreten Wahrscheinlichkeits-Räumen $\langle \Omega, 2^\Omega, P \rangle$ gearbeitet, Ω war dabei eine höchstens abzählbare Menge diskreter Ergebnisse. Bei vielen Zufalls-Experimenten ist das Ergebnis aber nicht eine natürliche Zahl, sondern eine beliebige reelle Zahl oder sogar ein Tupel reeller Zahlen. In diesen Fällen hat der Wahrscheinlichkeits-Raum die Form

$$\langle \Omega, \mathfrak{A}, P \rangle.$$

Die Komponenten dieses Tripels haben dabei die folgenden Eigenschaften:

1. Die erste Komponente Ω bezeichnen wir als den *Ergebnisraum*. Es gilt
$$\Omega \subseteq \mathbb{R} \text{ oder } \Omega \subseteq \mathbb{R}^n \text{ für ein } n \in \mathbb{N} .$$

2. Die zweite Komponente $\mathfrak{A}$ bezeichnen wir als den *Ereignisraum*. Es gilt
$$\mathfrak{A} \subseteq 2^\Omega.$$

   Der Ereignisraum $\mathfrak{A}$ ist jetzt nicht mehr mit der Potenzmenge von Ω identisch, sondern ist nur noch eine Teilmenge dieser Potenzmenge. Der Grund dafür ist, vereinfacht gesagt, dass die Potenzmenge von Ω bestimmte Mengen enthält, die so kompliziert sind, dass wir diesen Mengen keine Wahrscheinlichkeit zuordnen können. Daher können diese Mengen auch nicht Elemente unseres Ereignisraums sein. In der Mathematik nennt man diese komplizierten Mengen "*nicht-meßbare Mengen*[1]".

   Um mit dem Ereignisraum arbeiten zu können müssen wir fordern, dass der Ereignisraum bestimmten Abschuss-Eigenschaften genügt. Genauer muss gelten:

   (a) $\Omega \in \mathfrak{A}$.
   (b) Aus $A \in \mathfrak{A}$ folgt $A^c \in \mathfrak{A}$.
   (c) Gilt für alle $n \in \mathbb{N}$ dass $A_n \in \mathfrak{A}$ ist, dann gilt auch
   $$\bigcup_{n \in \mathbb{N}} A_n \in \mathfrak{A}.$$

   Ein Mengensystem, dass diesen Abschuss-Eigenschaften genügt, bezeichnen wir auch als σ-Algebra.

[1]Den Teilmengen von $\mathbb{R}$, mit denen wir bisher gearbeitet haben, kann eine Länge zugeordnet werden. Es lassen sich aber mit Hilfe des sogenannten Auswahl-Axioms Mengen konstruieren, von denen gezeigt werden kann, dass diesen Mengen keine Länge mehr zugeordnet werden kann. Solche Mengen müssen wir aus unserem Ereignisraum ausschließen.

Im $\mathbb{R}^3$ führen die nicht-meßbaren Mengen zu dem sogenannten *Banach-Tarski-Paradoxon*: Es ist möglich, eine Kugel mit dem Radius 1 so in eine endliche Anzahl von $m + n$ Stücken zu zerlegen, dass einerseits die ersten m Stücke wieder zu einer Kugel vom Radius 1 zusammengefügt werden können und dass andererseits auch die restlichen n Stücke zu einer Kugel vom Radius 1 zusammengefügt werden können. Dadurch ist es möglich, die Kugel zu verdoppeln. Nähere Information hierzu finden Sie in der Wikipedia unter dem Stichwort *Banach-Tarski-Paradoxon*.

Wenn $\Omega \subseteq \mathbb{R}$ ist, dann reicht es für unsere Zwecke aus, wenn $\mathfrak{A}$ alle Intervalle $[a, b]$ enthält, für die $[a, b] \subseteq \Omega$ ist.

3. Die Wahrscheinlichkeits-Verteilung ist eine Abbildung, die den selben Axiomen wie im diskreten Fall genügt:

   (a) $0 \leq P(A) \leq 1 \quad$ für alle $A \subseteq \Omega$.

   (b) $P(\emptyset) = 0$.

   (c) $P(\Omega) = 1$.

   (d) Ist $(A_n)_{n\in\mathbb{N}}$ eine Folge paarweise disjunkter von Mengen aus $\mathfrak{A}$, gilt also
   $$\forall n \in \mathbb{N} : A_n \in \mathfrak{A} \quad \text{und} \quad \forall m, n \in \mathbb{N} : m \neq n \rightarrow A_m \cap A_n = \emptyset$$
   $$\text{so folgt} \quad P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} P(A_n).$$

Um Zufalls-Experimente beschreiben zu können, führen wir den Begriff der *stetigen Zufalls-Größe* ein, wobei wir uns auf den eindimensionalen Fall beschränken. Eine *stetige Zufalls-Größe* X ist eine Abbildung
$$X : \Omega \rightarrow \mathbb{R}.$$
Hat Ω die Form $\Omega = [a, b]$ mit

1. $a \in \mathbb{R} \cup \{-\infty\}$,
2. $b \in \mathbb{R} \cup \{+\infty\}$,
3. $a < b$,

so ist die *(kumulative) Verteilungs-Funktion*[2] F_X der Zufalls-Größe X definiert als
$$F_X : [a, b] \rightarrow [0, 1] \quad \text{mit } F_X(x) := P(X \leq x).$$
Aus der Verteilungs-Funktion läßt sich sofort die Wahrscheinlichkeit eines Ereignisses der Form
$$\big\{\omega \in \Omega \;\big|\; u < X(\omega) \leq v\big\}$$
berechnen, denn es gilt
$$P\big(\big\{\omega \in \Omega \;\big|\; u < X(\omega) \leq v\big\}\big) = F_X(v) - F_X(u).$$
Anstelle der Verteilungs-Funktion werden wir oft mit der Wahrscheinlichkeits-Dichte p_X arbeiten. Die Wahrscheinlichkeits-Dichte
$$p_X : \Omega \rightarrow \mathbb{R}$$
wird als Grenzwert definiert:
$$p_X(x) := \lim_{h\rightarrow 0} \frac{F_X(x + h) - F_X(x)}{h} = \frac{\mathrm{d}F_X}{\mathrm{d}x}.$$
Nach dem Hauptsatz der Differenzial- und Integral-Rechnung kann die Verteilungs-Funktion auf die Wahrscheinlichkeits-Dichte zurückgeführt werden. Hat Ω die Form $[a, b]$, so gilt:
$$F_X(x) = \int_a^x p_X(t)\,\mathrm{d}t.$$
Beispiel: Der einfachste Fall einer stetigen Zufalls-Größe ist eine *gleichverteilte* Zufalls-Größe. Es seien $a, b \in \mathbb{R}$ und der Wahrscheinlichkeits-Raum habe die Form
$$\langle [a, b], \mathfrak{A}, P\rangle.$$
Die Wahrscheinlichkeits-Verteilung $P(A)$ sei für ein Intervall $[u, v] \subseteq [a, b]$ durch das Verhältnis der Längen definiert:
$$P\big([u, v]\big) = \frac{v - u}{b - a}$$

[2] Das Attribut *kumulativ* lassen wir im folgenden weg.

Weiter sei die Zufalls-Größe $X : [a, b] \to \mathbb{R}$ trivial definiert durch $X(x) := x$. Die Verteilungs-Funktion F_X der Zufalls-Größe X hat dann die Form

$$F_X(x) = P(X \leq x) = P\big([a, x]\big) = \frac{x-a}{b-a}.$$

Damit ergibt sich die Wahrscheinlichkeits-Dichte p_X der Zufalls-Größe X als

$$p_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x-a}{b-a}\right) = \frac{1}{b-a}.$$

Aufgabe 25: In einer fremden Stadt kommen Sie zu einem zufälligen Zeitpunkt an eine Bushaltestelle, an der kein Fahrplan hängt. Sie wissen allerdings, dass Busse im Intervall von 30 Minuten ankommen. Berechnen Sie die Wahrscheinlichkeit, dass Sie länger als 20 Minuten auf den Bus warten müssen.

Lösung: Der Wahrscheinlichkeits-Raum hat die Form

$$\langle [0, 30], \mathfrak{A}, P \rangle$$

und die Zufalls-Größe X, die die Zahl in Minuten angibt, bis der nächste Bus kommt, ist gleichverteilt. Also ist die Verteilungs-Funktion F_X durch

$$F_X(x) = \frac{x-0}{30-0} = \frac{x}{30}$$

gegeben. Damit haben wir

$$\begin{aligned} P(X \geq 20) &= P(20 \leq X \leq 30) \\ &= F_X(30) - F_X(20) \\ &= \frac{30}{30} - \frac{20}{30} \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

Die Wahrscheinlichkeit, dass Sie länger als 20 Minuten warten müssen, beträgt also etwa 33%. □

Wir betrachten nun ein Beispiel, bei dem zwar die Wahrscheinlichkeits-Verteilung P gleichmäßig ist, in dem aber die Zufalls-Größe selber nicht mehr gleichverteilt ist.

Beispiel: Der Feuerleitrechner eines Artillerie-Geschützes ist durch Feindeinwirkung beschädigt und produziert für den Abschusswinkel φ nur noch gleichverteilte Zufallszahlen aus dem Intervall $\left[0, \frac{\pi}{4}\right]$. Unser Wahrscheinlichkeits-Raum hat also die Form

$$\left\langle \left[0, \frac{\pi}{4}\right], \mathfrak{A}, P \right\rangle \quad \text{mit } P\big([u, v]\big) = \frac{v-u}{\frac{\pi}{4}} = \frac{4}{\pi} \cdot (v-u)$$

Die Zufalls-Größe, an der wir interssiert sind, ist in diesem Fall die Schussweite $R(\varphi)$, die sich (unter Vernachlässigung des Luftwiderstands) nach der Formel

$$R(\varphi) = \frac{v_0^2}{g} \cdot \sin(2 \cdot \varphi) \tag{3.1}$$

berechnet. Dabei bezeichnet v_0 die Geschwindigkeit, die das Geschoss beim Austritt aus der Mündung des Geschützes besitzt. Für die weitere Rechnung nehmen wir $v_0 = 300\,\mathrm{m/s}$ an. Die Konstante g steht für die Erdbeschleunigung, die in unseren Breiten etwa den Wert $9.8\,\mathrm{m/s^2}$ besitzt. Wir wollen die Verteilungs-Funktion der Zufalls-Größe "Schussweite" R berechnen. Nach Gleichung (3.1) nimmt die Schussweite ihren Maximalwert

$$R_{\max} = \frac{v_0^2}{g}$$

bei $\varphi = \frac{\pi}{4}$ an. Für die zugehörige Verteilungs-Funktion F_R finden wir

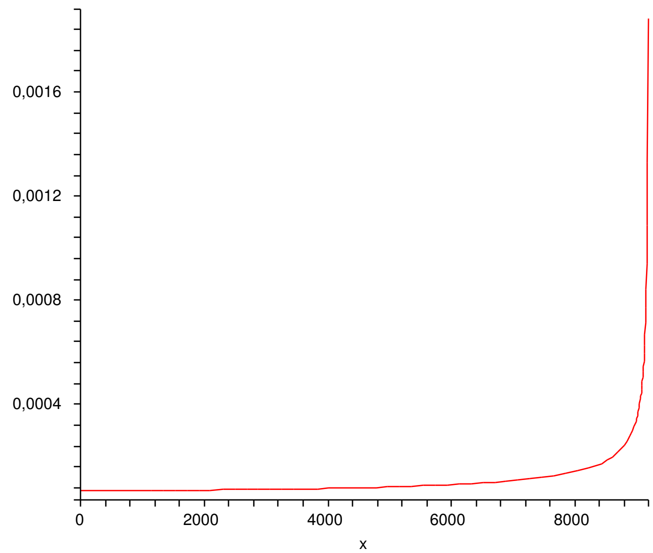

Abbildung 3.1: Wahrscheinlichkeits-Dichte der Schussweite.

$$\begin{aligned}
F_R(x) &= P(R \leq x) \\
&= P\left(\frac{v_0^2}{g} \cdot \sin(2 \cdot \varphi) \leq x\right) \\
&= P\left(\sin(2 \cdot \varphi) \leq x \cdot \frac{g}{v_0^2}\right) \\
&= P\left(\varphi \leq \frac{1}{2} \cdot \arcsin\Bigl(x \cdot \frac{g}{v_0^2}\Bigr)\right) \\
&= \frac{4}{\pi} \cdot \frac{1}{2} \cdot \arcsin\Bigl(x \cdot \frac{g}{v_0^2}\Bigr) \\
&= \frac{2}{\pi} \cdot \arcsin\Bigl(\frac{x}{R_{\max}}\Bigr)
\end{aligned}$$

Die zugehörige Wahrscheinlichkeits-Dichte finden wir, indem wir die Verteilungs-Funktion nach x differenzieren. Es ergibt sich

$$p_R(x) = \frac{\mathrm{d}F_f}{\mathrm{d}x} = \frac{2}{\pi \cdot R_{\max}} \cdot \frac{1}{\sqrt{1 - \left(\frac{x}{R_{\max}}\right)^2}}.$$

Diese Wahrscheinlichkeits-Dichte ist keinesweg gleichverteilt. Abbildung 3.1 zeigt den Graphen der Wahrscheinlichkeits-Dichte unter der Annahme $v_0 = 300\,\mathrm{m/s}$. An der rechten Intervall-Grenze divergiert die Wahrscheinlichkeits-Dichte, es gilt

$$\lim_{x \to R_{\max}} p_R(x) = \infty.$$

Der nächste Satz verallgemeinert das obige Beispiel.

Satz 21 (Variablen-Transformation) Es sei

1. $\mathcal{W} = \langle \Omega, \mathfrak{A}, P \rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to [a, b]$ eine stetige Zufalls-Größe,
3. $\varphi : [a, b] \to [c, d]$ eine streng monoton wachsende Funktion

4. $\psi : [c,d] \to [a,b]$ die Umkehrfunktion von φ, also
$$\psi\bigl(\varphi(t)\bigr) = t \quad \text{für alle } t \in [a,b],$$

5. $Y : \Omega \to [c,d]$ eine Zufalls-Größe, die definiert ist durch
$$Y(\omega) := \varphi\bigl(X(\omega)\bigr).$$

Dann gilt für die Wahrscheinlichkeits-Dichten p_X und p_Y die Beziehung
$$p_Y(t) = p_X\bigl(\psi(t)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}$$

Beweis: Wir betrachten die Verteilungs-Funktion der Zufalls-Größe Y. Es gilt
$$\begin{array}{lcll} F_Y(t) & = & P(Y \leq t) & \\ & = & P(\varphi \circ X \leq t) & \text{nach Definition von } Y \\ & = & P\bigl(X \leq \psi(t)\bigr) & \text{denn } \psi \text{ ist Umkehrfunktion von } \varphi \\ & = & F_X\bigl(\psi(t)\bigr) & \text{nach Definition der Verteilungs-Funktion von } X \end{array}$$
Wir haben also $F_Y(t) = F_X\bigl(\psi(t)\bigr)$. Wenn wir diese Gleichung nach t differenzieren, erhalten wir die Behauptung mit Hilfe der Ketten-Regel:
$$p_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_Y(t) = \frac{\mathrm{d}}{\mathrm{d}t}F_X\bigl(\psi(t)\bigr) = \frac{\mathrm{d}F_X}{\mathrm{d}t}\bigl(\psi(t)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t} = p_X\bigl(\psi(t)\bigr) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}.$$ □

3.1 Erwartungswert und Varianz stetiger Zufalls-Größen

Es sei $\langle [a,b], \mathfrak{A}, P\rangle$ ein Wahrscheinlichkeits-Raum und
$$X : [a,b] \to [c,d]$$
sei eine stetige Zufalls-Größe mit Wahrscheinlichkeits-Dichte p_X. Wir überlegen uns, wie wir die Definition des Erwartungswerts auf den kontinuierlichen Fall übertragen können. Für eine diskrete Zufalls-Größe Y, deren Wertebereich sich als Menge der Form
$$Y(\Omega) = \bigl\{ y_k \bigm| k \in \{1, \cdots, n\} \bigr\}$$
schreiben läßt, hatten wir seinerzeit die Formel
$$E[Y] = \sum_{k=0}^{n} y_k \cdot P(Y = y_k)$$
gefunden. Da der Wertebereich der Zufalls-Größe X kontinuierlich ist, können wir versuchen, die möglichen Werte von X in viele kleine Intervalle der Form
$$[x_k, x_{k+1}] \quad \text{mit } x_k = c + \frac{k}{n} \cdot (d-c) \text{ für } k = 0, \cdots, n-1$$
aufzuteilen. Jedes dieser Intervalle hat die Länge
$$h = x_{k+1} - x_k = \frac{k+1}{n} \cdot (d-c) - \frac{k}{n} \cdot (d-c) = \frac{d-c}{n}.$$
Wenn n groß ist, wird die Länge h diese Intervalle sehr klein, so dass die Zahlen, die in dem selben Intervall liegen, nahezu gleich sind. Dann können wir den Erwartungswert der Zufalls-Größe X näherungsweise durch
$$E[X] \approx \sum_{k=1}^{n} x_k \cdot P(x_{k-1} \leq X \leq x_k)$$
berechnen. Nun gilt
$$P(x_{k-1} \leq X \leq x_k) = \int_{x_{k-1}}^{x_k} p_X(x)\; \mathrm{d}x.$$
Also haben wir für den Erwartungswert
$$E[X] \approx \sum_{k=1}^{n} x_k \cdot \int_{x_{k-1}}^{x_k} p_X(x)\; \mathrm{d}x$$

Hier können wir die Konstante x_k in das Intergral hineinziehen und durch die Variable x ersetzen, denn innerhalb eines Intervalls $[x_{k-1}, x_k]$ ändert sich x kaum, wenn n groß ist. Dann haben wir

$$E[X] \approx \sum_{k=1}^{n} \int_{x_{k-1}}^{x_k} x \cdot p_X(x)\ \mathrm{d}x = \int_c^d x \cdot p_X(x)\ \mathrm{d}x$$

Diese Überlegungen motivieren im Fall stetiger Zufalls-Größen die Definition des *Erwartungswerts* einer stetigen Zufalls-Größe X durch die Formel

$$E[X] := \int_c^d x \cdot p_X(x)\,\mathrm{d}x.$$

Hier bezeichnen c und d den minimalen bzw. maximalen Wert, den die Zufalls-Größe X annehmen kann. Die Varianz einer stetigen Zufalls-Größe wird nun wie früher durch

$$\mathrm{Var}[X] := E\Big[(X - E[X])^2\Big]$$

definiert. Setzen wir $\mu := E[X]$, so heißt dass

$$\mathrm{Var}[X] = \int_c^d (x-\mu)^2 \cdot p_X(x)\,\mathrm{d}x. \qquad \square$$

Aufgabe 26: Berechnen Sie Varianz und Erwartungswert der Zufalls-Größe X, falls für die Verteilungs-Funktion F_X gilt

$$F_X(t) = \begin{cases} 1 - e^{-\lambda\cdot t} & \text{falls } t \geq 0; \\ 0 & \text{sonst.} \end{cases} \qquad \square$$

Lösung: Zunächst müssen wir die Wahrscheinlichkeits-Dichte p_X berechnen. Es gilt

$$p_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}(1 - e^{-\lambda\cdot t}) = \lambda \cdot e^{-\lambda\cdot t} \qquad \text{für } x \geq 0.$$

Damit erhalten wir für den Erwartungswert

$$\begin{array}{lcl} E[X] & = & \displaystyle\int_0^\infty x \cdot \lambda \cdot e^{-\lambda\cdot x}\,\mathrm{d}x \\ & & \text{partielle Integration: } u(x) = x \text{ und } v'(x) = \lambda \cdot e^{-\lambda\cdot x} \\ & & \text{also: } \qquad\qquad\quad u'(x) = 1 \text{ und } v(x) = -e^{-\lambda\cdot x} \\ & = & \underbrace{-x \cdot e^{-\lambda\cdot x}\Big|_0^\infty}_{0} + \displaystyle\int_0^\infty e^{-\lambda\cdot x}\,\mathrm{d}x \\ & & \text{denn } \lim\limits_{x\to\infty} x \cdot e^{-\lambda\cdot x} = 0 \\ & = & \dfrac{-1}{\lambda} \cdot e^{-\lambda\cdot x}\Big|_0^\infty \\ & = & \dfrac{1}{\lambda} \end{array}$$

Für die Varianz finden wir entsprechend

$$
\begin{aligned}
\mathrm{Var}[X] &= \int_0^\infty \Big(x - \frac{1}{\lambda}\Big)^2 \cdot \lambda \cdot e^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{partielle Integration: } u(x) = (x - \tfrac{1}{\lambda})^2 \text{ und } v'(x) = \lambda \cdot e^{-\lambda\cdot x} \\
&\quad \text{also} \qquad\qquad\qquad\quad u'(x) = 2 \cdot (x - \tfrac{1}{\lambda}) \text{ und } v(x) = -e^{-\lambda\cdot x} \\
&= -\Big(x - \frac{1}{\lambda}\Big)^2 \cdot e^{-\lambda\cdot x}\Big|_0^\infty + 2 \cdot \int_0^\infty \Big(x - \frac{1}{\lambda}\Big) \cdot e^{-\lambda\cdot x}\,\mathrm{d}x \\
&= \frac{1}{\lambda^2} + 2 \cdot \int_0^\infty \Big(x - \frac{1}{\lambda}\Big) \cdot e^{-\lambda\cdot x}\,\mathrm{d}x \\
&\quad \text{partielle Integration: } u(x) = x - \tfrac{1}{\lambda} \text{ und } v'(x) = e^{-\lambda\cdot x} \\
&\quad \text{also} \qquad\qquad\qquad\quad u'(x) = 1 \text{ und } v(x) = -\tfrac{1}{\lambda} \cdot e^{-\lambda\cdot x} \\
&= \frac{1}{\lambda^2} - \left(2 \cdot \Big(x - \frac{1}{\lambda}\Big) \cdot \frac{1}{\lambda} \cdot e^{-\lambda\cdot x}\Big|_0^\infty\right) + \frac{2}{\lambda} \cdot \int_0^\infty e^{-\lambda\cdot x}\,\mathrm{d}x \\
&= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} - \left(\frac{2}{\lambda^2} \cdot e^{-\lambda\cdot x}\Big|_0^\infty\right) \\
&= \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} \\
&= \frac{1}{\lambda^2}.
\end{aligned}
$$

□

Satz 22 Es sei $X : \Omega \to [a,b]$ eine Zufalls-Größe mit Wahrscheinlichkeits-Dichte $p_X(t)$. Dann kann die Varianz $\mathrm{Var}[X]$ wie folgt berechnet werden:

$$\mathrm{Var}[X] = E\big[X^2\big] - \big(E[X]\big)^2.$$

Beweis: Es sei $\mu := E[X] = \int_a^b x \cdot p_X(x)\,\mathrm{d}x$. Dann gilt

$$
\begin{aligned}
\mathrm{Var}[X] &= E\Big[\big(X - E[X]\big)^2\Big] \\
&= \int_a^b (x-\mu)^2 \cdot p_X(x)\,\mathrm{d}x \\
&= \int_a^b \big(x^2 - 2\cdot\mu\cdot x + \mu^2\big) \cdot p_X(x)\,\mathrm{d}x \\
&= \int_a^b x^2 \cdot p_X(x)\,\mathrm{d}x \;-\; 2\cdot\mu\cdot \underbrace{\int_a^b x \cdot p_X(x)\,\mathrm{d}x}_{E[X]} + \mu^2 \cdot \underbrace{\int_a^b p_X(x)\,\mathrm{d}x}_{1} \\
&= E\big[X^2\big] - 2\cdot\mu^2 + \mu^2 \\
&= E\big[X^2\big] - \big(E[X]\big)^2
\end{aligned}
$$

□

Satz 23 (Verhalten des Erwartungswerts bei Variablen-Transformation) Es sei

1. $\mathcal{W} = \langle \Omega, \mathfrak{A}, P\rangle$ ein Wahrscheinlichkeits-Raum,
2. $X : \Omega \to [a,b]$ eine stetige Zufalls-Größe,
3. $\varphi : [a,b] \to [c,d]$ eine stetige Funktion
4. $Y : \Omega \to [c,d]$ eine Zufalls-Größe, die definiert ist durch

   $$Y := \varphi \circ X, \quad \text{also} \quad Y(\omega) := \varphi\big(X(\omega)\big) \quad \text{für alle } \omega \in \Omega.$$

Dann gilt für den Erwartungswert von Y

$$E[Y] = E[\varphi \circ X] = \int_a^b \varphi(x) \cdot p_X(x)\,\mathrm{d}x$$

Beweis: Wir setzen zusätzlich voraus, dass die Funktion φ streng monoton steigend ist und dass $\psi : [c,d] \to [a,b]$ die Umkehrfunktion von φ ist. Nach dem Satz über Variablen-Transformation gilt dann

$$p_Y(t) = p_X\big(\psi(t)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}t}.$$

Nach Definition des Erwartungswerts gilt

$$\begin{array}{rcll}
E[Y] & = & \displaystyle\int_c^d y \cdot p_Y(y)\ \mathrm{d}y & \text{nach Definition von } E[Y] \\
& = & \displaystyle\int_c^d y \cdot p_X\big(\psi(y)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}y}\ \mathrm{d}y & \text{nach Satz 21 über Variablen-Transformationen} \\
& = & \displaystyle\int_c^d \varphi(\psi(y)) \cdot p_X\big(\psi(y)\big) \cdot \frac{\mathrm{d}\psi}{\mathrm{d}y}\ \mathrm{d}y & \text{denn } \psi \text{ ist die Umkehrfunktion von } \varphi \\
& = & \displaystyle\int_a^b \varphi(\psi) \cdot p_X(\psi)\,\mathrm{d}\psi & \text{Substitutionsregel der Analysis} \\
& = & \displaystyle\int_a^b \varphi(x) \cdot p_X(x)\,\mathrm{d}x & \text{ob die Integrations-Variable } x \text{ oder } \psi \text{ heißt, ist egal}
\end{array}$$

Im allgemeinen Fall kann der Beweis dadurch geführt werden, dass wir das Intervall $[a,b]$ in Intervalle zerlegen, in denen die Funktion φ streng monoton ist. □

3.2 Moment-erzeugende Funktion

In diesem Abschnitt führen wir ein Hilfsmittel ein, mit dem wir später die Wahrscheinlichkeits-Dichte für eine Reihe wichtiger Zufalls-Größen berechnen können. Dieses Hilfsmittel ist die moment-erzeugende Funktion einer Zufalls-Größe X. Der Begriff der moment-erzeugenden Funktion wird außerdem zum Beweis des *zentralen Grenzwert-Satzes* benötigt.

Definition 24 (Moment-erzeugende Funktion) Es sei $X : [a,b] \to [c,d]$ eine stetige Zufalls-Größe. Dann definieren wir die moment-erzeugende Funktion

$$M_X : \mathbb{R} \to \mathbb{R}$$

für alle $t \in \mathbb{R}$ als den folgenden Erwartungswert:

$$M_X(t) := E\big[\exp(t \cdot X)\big] = \int_c^d \exp\big(t \cdot x\big) \cdot p_X(x)\ \mathrm{d}x.$$

Dabei haben wir bei der letzten Gleichung Gebrauch von Satz 23 gemacht. □

Satz 25 (Verschiebungs-Satz für moment-erzeugende Funktionen) Es sei $X : \Omega \to \mathbb{R}$ eine Zufalls-Größe, $a \in \mathbb{R}$ und $b \in \mathbb{R}\backslash\{0\}$. Dann gilt

$$M_{(X+a)/b}(t) = M_X\left(\frac{t}{b}\right) \cdot \exp\left(\frac{a \cdot t}{b}\right).$$

Aufgabe 27: Beweisen Sie den Verschiebungs-Satz für moment-erzeugende Funktionen.
Hinweis: Verwenden Sie Satz 21.

Beispiel: Es sei $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_+$. Die Zufalls-Größe $N_{\mu,\sigma} : \Omega \to \mathbb{R}$ habe die Wahrscheinlichkeits-Dichte

$$p_{N_{\mu,\sigma}}(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2 \cdot \sigma^2}\right) \qquad \text{mit } \sigma > 0.$$

Eine solche Zufalls-Größe heißt *normal-verteilt* mit Parametern μ und σ. Wir werden später sehen, dass μ der Erwartungswert von $N_{\mu,\sigma}$ ist. Den Parameter σ werden wir als die Standard-Abweichung von $N_{\mu,\sigma}$ erkennen. Wir definieren eine lineare Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ durch

$$\varphi(x) = \frac{x-\mu}{\sigma}$$

Die Umkehrfunktion ψ von φ ist dann

$$\psi(y) = y \cdot \sigma + \mu.$$

Definieren wir die Zufallsgröße $N_{0,1}$ durch $N_{0,1} = \varphi \circ N_{\mu,\sigma}$ so folgt nach Satz 21 für die Wahrscheinlichkeits-Dichte p_N:

$$p_{N_{0,1}}(y) = p_{N_{\mu,\,\sigma}}(y \cdot \sigma + \mu) \cdot \sigma = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{y^2}{2}\right).$$

Die moment-erzeugende Funktion von $N_{0,1}$ erhalten wir wie folgt:

$$\begin{aligned}
M_{N_{0,1}}(t) &= \int_{-\infty}^{\infty} \exp(t \cdot x) \cdot \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{x^2}{2}\right)\, \mathrm{d}x \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\infty} \exp(t \cdot x) \cdot \exp\left(-\frac{x^2}{2}\right)\, \mathrm{d}x \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \cdot (x^2 - 2 \cdot t \cdot x)\right)\, \mathrm{d}x \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \cdot (x^2 - 2 \cdot t \cdot x + t^2) + \frac{1}{2} \cdot t^2\right)\, \mathrm{d}x \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{1}{2} \cdot t^2\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \cdot (x-t)^2\right)\, \mathrm{d}x \\
&= \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{1}{2} \cdot t^2\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \cdot y^2\right)\, \mathrm{d}y \\
&\ \text{mit der Variablen-Transformation } x \mapsto y + t \\
&= \exp\left(\frac{1}{2} \cdot t^2\right) \\
&\ \text{denn } \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \cdot y^2\right)\, \mathrm{d}y = \sqrt{2 \cdot \pi}\,.
\end{aligned}$$

Zwischen der Zufallsgröße $N_{\mu,\sigma}$ und der Zufallsgröße $N_{0,1}$ besteht der Zusammenhang

$$N_{\mu,\sigma} = \psi \circ N_{0,1} = N_{0,1} \cdot \sigma + \mu = \left(N_{0,1} + \frac{\mu}{\sigma}\right) \cdot \sigma = \left(N_{0,1} + \frac{\mu}{\sigma}\right) \cdot \frac{1}{\frac{1}{\sigma}}$$

Nach dem Verschiebungs-Satz für moment-erzeugende Funktionen gilt daher

$$\begin{aligned}
M_{N_{\mu,\,\sigma}}(t) &= \exp(\mu \cdot t) \cdot M_{N_{0,1}}\big(t \cdot \sigma\big) \\
&= \exp(\mu \cdot t) \cdot \exp\left(\frac{1}{2} \cdot t^2 \cdot \sigma^2\right) \\
&= \exp\left(\frac{1}{2} \cdot t^2 \cdot \sigma^2 + \mu \cdot t\right)
\end{aligned}$$

□

Aufgabe 28: Überzeugen Sie sich mit Hilfe von *Maple* davon, dass gilt:

$$E\big[N_{\mu,\sigma}\big] = \mu \quad \text{und} \quad \mathrm{Var}\big[N_{\mu,\sigma}\big] = \sigma^2.$$

Satz 26 (Eindeutigkeits-Satz) Es seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei Zufallsgrößen und für die moment-erzeugenden Funktionen M_X und M_Y gelte

$$M_X(t) = M_Y(t) \quad \text{für alle } t \in \mathbb{R}.$$

Dann sind die Wahrscheinlichkeits-Dichten p_X und p_Y gleich:

$$p_X(t) = p_Y(t) \quad \text{für alle } t \in \mathbb{R}.$$

□

Der Beweis dieses Satzes gelingt, indem die moment-erzeugende Funktion einer Zufallsgröße auf die Fourier-Transformierte zurückgeführt wird. Da wir die Fourier-Transformation in der Analysis nicht mehr behandelt haben, können wir hier die Details dieses Beweises nicht diskutieren. Trotzdem ist der Satz für unsere weitere Arbeit ein wichtiges Hilfsmittel, denn er gestattet uns, die Wahrscheinlichkeits-Dichten bestimmter Zufallsgrößen durch Berechnung der moment-erzeugenden Funktion zu berechnen.

Satz 27 Es seien $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ zwei unabhängige Zufallsgrößen. Dann läßt sich die moment-erzeugende Funktion der Zufallsgröße $X + Y$ als Produkt der moment-erzeugenden Funktionen von X und Y schreiben:

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) \quad \text{für alle } t \in \mathbb{R}.$$

Beweis: Es gilt

$$\begin{aligned}
M_{X+Y}(t) &= E\big[\exp\big(t \cdot (X+Y)\big)\big] \\
&= E\left[\exp(t \cdot X) \cdot \exp(t \cdot Y)\right] \\
&= E\left[\exp(t \cdot X)\right] \cdot E\left[\exp(t \cdot Y)\right] \\
& \quad\ \text{denn wenn } X \text{ und } Y \text{ unabhängig sind,} \\
& \quad\ \text{dann sind auch } \exp(t \cdot X) \text{ und } \exp(t \cdot Y) \text{ unabhängig} \\
&= M_X(t) \cdot M_Y(t)
\end{aligned}$$

Hier haben wir im vorletzten Schritt ausgenutzt, dass mit X und Y auch die Zufallsgrößen $\exp(t \cdot X)$ und $\exp(t \cdot Y)$ unabhängig sind. Genau wir im diskreten Fall auch gilt für den Erwartungswert des Produkts unabhängiger Zufallsgrößen X und Y die Gleichung

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

□

Satz 28 Ist $f(t) = M_X(t)$ die moment-erzeugende Funktion der Zufallsgröße X, und ist die Funktion f zweimal differenzierbar, so lassen sich Erwartungswert und Varianz der Zufallsgröße X wie folgt berechnen:

$$E[X] = \frac{\mathrm{d}f}{\mathrm{d}x}(0) \quad \text{und} \quad \mathrm{Var}[X] = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}(0) - \big(E[X]\big)^2.$$

Beweis: Es gilt

$$\begin{aligned}
M_X(t) &= \int_a^b \exp(t \cdot x) \cdot p_X(t)\, \mathrm{d}x \\
&= \int_a^b \sum_{n=0}^{\infty} \frac{(t \cdot x)^n}{n!} \cdot p_X(t)\, \mathrm{d}x \quad \text{Taylor-Reihe: } \exp(x) = \textstyle\sum\limits_{n=0}^{\infty} \frac{x^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot \int_a^b x^n \cdot p_X(t)\, \mathrm{d}x \\
&= \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot E\big[X^n\big]
\end{aligned}$$

Auf der anderen Seite hat die Taylor-Reihe von $f(t)$ die Form

$$f(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot f^{(n)}(0)$$

Durch einen Koeffizienten-Vergleich der beiden Reihen erkennen wir, dass

$$E[X^n] = f^{(n)}(0) \quad \text{für alle } n \in \mathbb{N} \text{ gilt.}$$

Wir haben in Satz 22 gezeigt, dass für die Varianz

$$\mathrm{Var}[X] = E\big[X^2\big] - \big(E[X]\big)^2$$

gilt. Daraus folgt nun die Behauptung. $\square$

Bemerkung: Der Ausdruck $E\big[X^n\big]$ wird auch als das n-te *Moment* der Zufallsgröße X bezeichnet. Der Beweis des letzten Satzes zeigt, dass das n-te Moment gerade die n-te Ableitung der moment-erzeugenden Funktion an der Stelle $t = 0$ ist. Dieser Umstand erklärt den Namen der moment-erzeugenden Funktion.

3.3 Der zentrale Grenzwert-Satz

Wir kommen nun zu dem wichtigsten Ergebnis der Wahrscheinlichkeits-Rechnung.

Satz 29 (Zentraler Grenzwertsatz) Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger stetiger Zufallsgrößen, die alle die selbe Wahrscheinlichkeits-Dichte haben, es gilt also

$$\forall m, n \in \mathbb{N} : p_{X_m}(t) = p_{X_n}(t).$$

Weiter gelte

$$\mu = E[X_n] \quad \text{und} \quad \sigma^2 = \mathrm{Var}[X_n].$$

Definieren wir die Zufallsgrößen S_n durch

$$S_n = \frac{1}{\sigma\cdot\sqrt{n}}\cdot\sum_{k=1}^{n}(X_k-\mu),$$

so konvergieren die Wahrscheinlichkeits-Dichten p_{S_n} der Zufallsgrößen S_n gegen die Normal-Verteilung, es gilt

$$\lim_{n\to\infty} p_{S_n}(t) = \frac{1}{\sqrt{2\cdot\pi}}\cdot\exp\left(-\frac{1}{2}\cdot t^2\right).$$

Beweis: Wir führen den Beweis, indem wir zeigen, dass die moment-erzeugenden Funktionen M_{S_n} der Zufallsgrößen S_n gegen die moment-erzeugende Funktion einer normal-verteilten Zufallsgröße mit Mittelwert 0 und Varianz 1 konvergieren. Die Behauptung folgt dann aus dem Eindeutigkeits-Satz. Es sei f_k die moment-erzeugende Funktion der Zufallsgröße

$$Y_k := \frac{X_k-\mu}{\sigma\cdot\sqrt{n}}, \quad \text{also}$$

$$f_k(t) = M_{Y_k}(t) = E\left[\exp\left(t\cdot\frac{1}{\sigma\cdot\sqrt{n}}\cdot(X_k-\mu)\right)\right].$$

Da die Zufallsgrößen X_n alle die selbe Wahrscheinlichkeits-Dichte haben, sind auch die moment-erzeugenden Funktionen M_{Y_k} für alle k gleich, so dass wir den Index k bei der Funktion f_k weglassen können. Für die moment-erzeugende Funktion M_{S_n} finden wir jetzt

$$\begin{aligned}
M_{S_n}(t) &= E\left[\exp\left(t\cdot\frac{1}{\sigma\cdot\sqrt{n}}\cdot\sum_{k=1}^{n}(X_k-\mu)\right)\right] \\
&= E\left[\prod_{k=1}^{n}\exp\left(t\cdot\frac{1}{\sigma\cdot\sqrt{n}}\cdot(X_k-\mu)\right)\right] \\
&= \prod_{k=1}^{n}E\left[\exp\left(t\cdot\frac{1}{\sigma\cdot\sqrt{n}}\cdot(X_k-\mu)\right)\right] \\
&\quad \text{denn die Zufallsgrößen } X_k \text{ sind unabhängig} \\
&= \prod_{k=1}^{n} f(t) \\
&= f^n(t)
\end{aligned}$$

In der Definition der Funktion f ersetzen wir die Exponential-Funktion durch die Taylor-Reihe

$$\exp(x) = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

und erhalten

$$\begin{aligned}
f(t) &= E\left[\sum_{i=0}^{\infty} \frac{1}{i!} \cdot \left(\frac{t}{\sigma \cdot \sqrt{n}}\right)^i \cdot (X_k - \mu)^i\right] \\
&= \sum_{i=0}^{\infty} \frac{1}{i!} \cdot \left(\frac{t}{\sigma \cdot \sqrt{n}}\right)^i \cdot E\big[(X_k - \mu)^i\big] \\
&= 1 + \frac{t}{\sigma \cdot \sqrt{n}} \cdot \underbrace{E[X_k - \mu]}_{0} + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot E\big[(X_k - \mu)^2\big] \\
&\quad + \frac{1}{6} \cdot \frac{t^3}{\sigma^3 \cdot n \cdot \sqrt{n}} \cdot E\big[(X_k - \mu)^3\big] + \cdots \\
&\approx 1 + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot E\big[(X_k - \mu)^2\big] \\
&= 1 + \frac{1}{2} \cdot \frac{t^2}{\sigma^2 \cdot n} \cdot \sigma^2 \\
&= 1 + \frac{1}{2} \cdot \frac{t^2}{n}
\end{aligned}$$

Dabei haben wir im letzten Schritt ausgenutzt, dass für $n \to \infty$ die Terme, bei denen $n^{i/2}$ mit $i > 2$ im Nenner steht, gegenüber dem Term mit $\frac{1}{n}$ vernachläßigt werden können. Daher haben wir

$$\begin{aligned}
\lim_{n\to\infty} M_{S_n}(t) &= \lim_{n\to\infty} \left(1 + \frac{1}{2} \cdot \frac{t^2}{n}\right)^n \\
&= \exp\left(\frac{1}{2} \cdot t^2\right)
\end{aligned}$$

Das ist aber gerade die moment-erzeugende Funktion einer Zufallsgröße mit der Wahrscheinlichkeits-Dichte $N(0,1)$. Aus dem Eindeutigkeits-Satz für die moment-erzeugenden Funktionen folgt nun

$$\lim_{n\to\infty} p_{S_n}(x) = p_{N(0,1)}(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{1}{2} \cdot x^2\right). \qquad \square$$

Die Voraussetzungen des zentralen Grenzwertsatzes können noch abgeschwächt werden. Es ist nicht erforderlich, dass die einzelnen Zufallsgrößen X_k alle die selbe Wahrscheinlichkeits-Dichte haben. Im wesentlichen reicht es aus, wenn die Erwartungswerte und die Varianzen dieser Zufallsgrößen beschränkt sind. Darüber hinaus muss dann noch eine weitere recht technische Bedingung erfüllt sein. Der zentrale Grenzwertsatz ist der Grund dafür, dass in der Praxis viele Zufallsgrößen normal-verteilt sind. Diese liegt einfach daran, dass viele Größen, die wir beobachten, durch eine große Zahl von Faktoren bestimmt werden. Auch wenn die einzelnen Faktoren beliebig verteilt sind, so ist die Summe dieser Faktoren doch annähernd normal-verteilt.

Aufgabe 29: Die Zufallsgrößen X_1 und X_2 seien normal-verteilt mit den Erwartungswerten μ_1 und μ_2 und den Varianzen σ_1^2 und σ_2^2. Außerdem seien X_1 und X_2 unabhängig. Berechnen Sie die Wahrscheinlichkeits-Dichte der Zufallsgröße $X_1 + X_2$.

Hinweis: Orientieren Sie sich am Beweis des zentralen Grenzwertsatzes.

3.4 Die χ^2-Verteilung

In diesem Abschnitt gehen wir davon aus, dass X_1, X_2, $\cdots$, X_n unabhängige Zufalls-Größen sind, die alle einer standardisierten Normalverteilung genügen, für die einzelnen X_i gilt also

$$P(X_i \leq x) = \frac{1}{\sqrt{2\,\pi}} \cdot \int_{-\infty}^{x} \exp\biggl(-\frac{t^2}{2}\biggr)\,\mathrm{d}t$$

Die Zufalls-Größe Z ist dann als

$$Z(\omega) = \sum_{i=1}^{n} X_i^2(\omega)$$

definiert. Die Zahl n wird auch als der *Freiheitsgrad* der χ^2-Verteilung bezeichnet. Unser Ziel ist es, die Wahrscheinlichkeits-Dichte p_Z für die Zufalls-Größe Z zu berechnen. Wir beginnen mit dem Spezialfall $n = 1$.

3.4.1 Die χ^2-Verteilung mit einem Freiheitsgrad

Es sei X eine Zufalls-Größe, die der Standard-Normal-Verteilung genügt, für die Wahrscheinlichkeits-Dichte gilt also

$$p(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\bigl(-\frac{x^2}{2}\bigr).$$

Wir stellen uns die Frage, wie die Zufalls-Größe X^2 verteilt ist. Wir berechnen zunächst die kumulative Verteilungs-Funktion von X^2, die wir mit F_{X^2} bezeichnen. Ist Ω der zu Grunde liegende Ergebnis-Raum, so haben wir

$$\begin{aligned}
F_{X^2}(x) &= P(X^2 \leq x) \\
&= P\Bigl(\bigl\{\omega \in \Omega \bigm| X^2(\omega) \leq x\bigr\}\Bigr) \\
&= P\Bigl(\bigl\{\omega \in \Omega \bigm| -\sqrt{x} \leq X(\omega) \leq \sqrt{x}\bigr\}\Bigr) \\
&= P\Bigl(\bigl\{\omega \in \Omega \bigm| X(\omega) \leq \sqrt{x}\bigr\} \backslash \bigl\{\omega \in \Omega \bigm| X(\omega) < -\sqrt{x}\bigr\}\Bigr) \\
&= \Phi\bigl(\sqrt{x}\bigr) - \Phi\bigl(-\sqrt{x}\bigr),
\end{aligned}$$

denn die Verteilungs-Funktion einer Zufalls-Größe mit Normalverteilung ist die Gauß'sche Integral-Funktion $x \mapsto \Phi(x)$. Berücksichtigen wir noch, dass

$$\Phi(x) + \Phi(-x) = 1, \quad \text{also } \Phi(-x) = 1 - \Phi(x)$$

gilt, so haben wir

$$F_{X^2}(x) = 2 \cdot \Phi\bigl(\sqrt{x}\bigr) - 1.$$

Hier ist es zweckmäßig, die Gauß'sche Integral-Funktion Φ durch die Gauß'sche Fehler-Funktion *erf* auszudrücken. Es gilt

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \cdot \mathit{erf}\biggl(\frac{x}{\sqrt{2}}\biggr).$$

Damit erhalten wir das Ergebnis

$$\boxed{F_{X^2}(x) = \mathit{erf}\biggl(\sqrt{\frac{x}{2}}\biggr)}$$

Die Wahrscheinlichkeits-Dichte p_{X^2} der Zufalls-Größe X^2 erhalten wir, indem wir die Verteilungs-Funktion nach x differenzieren. Wegen

$$\mathit{erf}(x) = \frac{2}{\sqrt{\pi}} \cdot \int_{0}^{x} \exp(-u^2)\,\mathrm{d}u$$

erhalten wir

$$p_{X^2}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{erf}\biggl(\sqrt{\frac{x}{2}}\biggr) = \frac{2}{\sqrt{\pi}} \cdot \frac{1}{2\sqrt{2\,x}} \cdot \exp\Bigl(-\frac{x}{2}\Bigr) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{x}} \cdot \exp\Bigl(-\frac{x}{2}\Bigr) \tag{3.2}$$

Aufgabe 30: Berechnen Sie die moment-erzeugende Funktion der Zufalls-Größe X^2.
Lösung: Die Wahrscheinlichkeits-Dichte von X^2 ist durch Gleichung (3.2) gegeben. Also gilt

$$\begin{aligned}
M_{X^2}(t) &= \int_0^\infty \exp(t\cdot x)\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{x}}\cdot\exp\left(-\frac{x}{2}\right)\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\cdot\int_0^\infty \frac{1}{\sqrt{x}}\cdot\exp\left(t\cdot x-\frac{x}{2}\right)\,\mathrm{d}x \\
&\quad \text{mit der Variablen-Transformation } x=y^2, \text{ also } \mathrm{d}x = 2\cdot y\ \mathrm{d}y \text{ erhalten wir} \\
&= \frac{1}{\sqrt{2\pi}}\cdot\int_0^\infty \frac{1}{y}\cdot\exp\left(\left(t-\frac{1}{2}\right)\cdot y^2\right)\cdot 2\cdot y\ \mathrm{d}y \\
&= \frac{2}{\sqrt{2\pi}}\cdot\int_0^\infty \exp\left(-(1-2\cdot t)\cdot\frac{y^2}{2}\right)\,\mathrm{d}y \\
&\quad \text{mit der Variablen-Transformation } y=\frac{1}{\sqrt{1-2\cdot t}}\cdot z \text{ wird daraus} \\
&= \frac{2}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1-2\cdot t}}\cdot\int_0^\infty \exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z \\
&= \frac{2}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{1-2\cdot t}}\cdot\frac{1}{2}\cdot\sqrt{2\pi} \\
&\quad \text{denn } \int_0^\infty \exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z = \frac{1}{2}\cdot\int_{-\infty}^\infty \exp\left(-\frac{z^2}{2}\right)\,\mathrm{d}z = \frac{1}{2}\cdot\sqrt{2\pi} \\
&= \frac{1}{\sqrt{1-2\cdot t}}
\end{aligned}$$

Wir halten das Ergebnis fest:

$$M_{X^2}(t) = \frac{1}{\sqrt{1-2\cdot t}} \tag{3.3}$$

3.4.2 Der allgemeine Fall

Um unser eigentliches Problem lösen zu können, benötigen wir zwei Definitionen.

Definition 30 (Gamma-Funktion) Die Gamma-Funktion ist für positive reelle Zahlen x als das folgende Integral definiert:

$$\Gamma(x) := \int_0^\infty t^{x-1}\cdot e^{-t}\ \mathrm{d}t.$$

□

Aufgabe 31: Weisen Sie die folgenden Eigenschaften der Gamma-Funktion nach:

(a) $\Gamma(1) = 1$.

(b) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Hinweis: Führen Sie das Integral durch eine geeignete Substitution auf das Integral

$$\int_0^\infty \exp\left(-\frac{1}{2}\cdot t^2\right)\,\mathrm{d}t$$

zurück.

(c) $\Gamma(x+1) = x\cdot\Gamma(x)$.

Hinweis: Partielle Integration.

(d) $\Gamma(n+1) = n!$ für alle $n \in \mathbb{N}$.

Hinweis: Vollständige Induktion.

Die letzte Eigenschaft zeigt, dass die Gamma-Funktion als eine Erweiterung der Fakultäts-Funktion auf die positiven reellen Zahlen aufgefaßt werden kann. □

Definition 31 (Gamma-Verteilung) Eine Zufalls-Größe Y genügt einer *Gamma-Verteilung mit Parametern α und β*, falls für die Wahrscheinlichkeits-Dichte p_Y gilt

$$p_Y(x) = \begin{cases} \dfrac{x^{\alpha-1} \cdot \exp\left(-\dfrac{x}{\beta}\right)}{\beta^\alpha \cdot \Gamma(\alpha)} & \text{für } x > 0; \\ 0 & \text{für } x \leq 0 \end{cases}$$

Die Parameter α und β sind dabei positive reelle Zahlen. □

Aufgabe 32: Berechnen Sie die moment-erzeugende Funktion für eine Gamma-verteilte Zufalls-Größe Y.

Lösung: Nach Definition der moment-erzeugenden Funktion p_Y gilt:

$$\begin{aligned}
M_Y(t) &= \int_0^\infty \exp(t \cdot x) \cdot p_Y(x) \, \mathrm{d}x \\
&= \int_0^\infty \exp(t \cdot x) \cdot \frac{x^{\alpha-1} \cdot \exp(-x/\beta)}{\beta^\alpha \cdot \Gamma(\alpha)} \, \mathrm{d}x \\
&= \frac{1}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \int_0^\infty x^{\alpha-1} \cdot \exp\left(-\frac{x}{\beta} + t \cdot x\right) \, \mathrm{d}x \\
&= \frac{1}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \int_0^\infty x^{\alpha-1} \cdot \exp\left(-\left(\frac{1}{\beta} - t\right) \cdot x\right) \, \mathrm{d}x \\
& \quad \text{mit der Variablen-Transformation } x = \frac{1}{\left(\frac{1}{\beta} - t\right)} \cdot y \text{ wird daraus} \\
&= \frac{1}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \frac{1}{\left(\frac{1}{\beta} - t\right)} \cdot \frac{1}{\left(\frac{1}{\beta} - t\right)^{\alpha-1}} \cdot \int_0^\infty y^{\alpha-1} \cdot \exp(-y) \, \mathrm{d}y \\
&= \frac{1}{\beta^\alpha \cdot \Gamma(\alpha)} \cdot \frac{1}{\left(\frac{1}{\beta} - t\right)^\alpha} \cdot \Gamma(\alpha) \\
&= \frac{1}{\beta^\alpha} \cdot \frac{1}{\left(\frac{1}{\beta} - t\right)^\alpha} \\
&= \frac{1}{\beta^\alpha} \cdot \left(\frac{\beta}{1 - \beta \cdot t}\right)^\alpha \\
&= \big(1 - \beta \cdot t\big)^{-\alpha}
\end{aligned}$$

Also hat die moment-erzeugende Funktion für eine Zufalls-Größe Y, die einer Gamma-Verteilung mit den Parametern α und β genügt, die Form

$$M_Y(t) = (1 - \beta \cdot t)^{-\alpha} \tag{3.4}$$

□

Wir lösen nun unser eigentliches Problem, nämlich die Berechnung der Wahrscheinlichkeits-Dichte einer Zufalls-Größe Z, die aus n unabhängigen, standard-normal-verteilten Zufalls-Größen X_i nach der Formel

$$Z = \sum_{i=1}^n X_i^2$$

berechnet wird. In diesem Fall sagen wir dass Z eine χ^2*-Verteilung mit n Freiheitsgraden* hat. Wir berechnen zunächst die moment-erzeugende Funktion M_Z von Z. Nach Satz 27 und Gleichung (3.3) gilt

$$M_Z(t) = \Big(M_{X^2}(t)\Big)^n = (1 - 2 \cdot t)^{-\frac{n}{2}}$$

Nach Gleichung (3.4) hat eine Zufalls-Größe, die einer Gamma-Verteilung mit den Parametern $\alpha = \frac{n}{2}$ und $\beta = 2$ genügt, die selbe moment-erzeugende Funktion. Aus dem Eindeutigkeits-Satz 26 für die moment-erzeugende Funktion folgt daher, dass die Wahrscheinlichkeits-Dichte der Zufalls-Größe $Z = \sum\limits_{i=1}^{n} X_i$ die Form

$$p_Z(x) = \frac{1}{A_n} \cdot x^{n/2-1} \cdot \exp\left(-\frac{x}{2}\right) \quad \text{mit } A_n = 2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right) \quad \text{für } x > 0 \text{ hat.}$$

Wir formulieren dieses Ergebnis als Satz. Vorher benötigen wir noch eine Definition.

Satz 32 Es seien $X_1, X_2, \cdots, X_n$ unabhängige Zufalls-Größen, die standard-normal-verteilt sind. Die Zufalls-Größe Z sei definiert als

$$Z := \sum_{i=1}^{n} X_i^2$$

Dann hat die Zufalls-Größe Z die Wahrscheinlichkeits-Dichte

$$p_Z(x) = \frac{1}{A_n} \cdot x^{n/2-1} \cdot \exp\left(-\frac{x}{2}\right) \quad \text{mit } A_n = 2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right).$$

Eine Zufalls-Größe mit dieser Wahrscheinlichkeits-Dichte heißt χ^2*-verteilt mit n Freiheitsgraden.*

□

Bei der Diskussion des χ^2-Tests im nächsten Kapitel benötigen wir die beiden folgenden Sätze.

Satz 33 Es seien U, V und W Zufalls-Größen mit folgenden Eigenschaften:

1. U ist χ^2-verteilt mit m Freiheitsgraden,
2. V ist χ^2-verteilt mit n Freiheitsgraden,
3. U und V sind unabhängig,
4. $W = U + V$.

Dann genügt die Zufalls-Größe W einer χ^2-Verteilung mit $m + n$ Freiheitsgraden.

Beweis: Aus dem Beweis des letzten Satzes können wir ablesen, wie die moment-erzeugenden Funktionen von U und V aussehen müssen. Es gilt

$$M_U(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \quad \text{und} \quad M_V(t) = (1 - 2 \cdot t)^{-\frac{n}{2}}.$$

Nach Satz 27 gilt daher für die moment-erzeugende Funktion M_W:

$$M_W(t) = M_U(t) \cdot M_V(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \cdot (1 - 2 \cdot t)^{-\frac{n}{2}} = (1 - 2 \cdot t)^{-\frac{m+n}{2}}.$$

Dies ist aber genau die moment-erzeugende Funktion einer Zufalls-Größe, die einer χ^2-Verteilung mit $m + n$ Freiheitsgraden genügt. Daher folgt die Behauptung nun aus dem Eindeutigkeits-Satz 26. □

Bemerkung: Der obige Satz läßt sich leicht auf die Summe von mehr als zwei χ^2-verteilten Zufalls-Größen verallgemeinern. Sind $U_1, \cdots, U_k$ unabhängige Zufalls-Größen, so dass für alle $i = 1, \cdots, k$ die Zufalls-Größe U_i einer χ^2-Verteilung mit m_i Freiheitsgraden genügt, dann genügt die Summe

$$W := \sum_{i=1}^{k} U_i$$

einer χ^2-Verteilung mit $m_1 + \cdots + m_k$ Freiheitsgraden. Diese Aussage kann mit Hilfe des letzten Satzes durch eine triviale Induktion nach k gezeigt werden. $\square$

Der nächste Satz zeigt, dass sich der letzte Satz auch umkehren läßt.

Satz 34 Es seien U, V und W Zufalls-Größen mit folgenden Eigenschaften:

1. U ist χ^2-verteilt mit m Freiheitsgraden,
2. W ist χ^2-verteilt mit $m+n$ Freiheitsgraden,
3. U und V sind unabhängig,
4. $W = U + V$.

Dann genügt die Zufalls-Größe V einer χ^2-Verteilung mit n Freiheitsgraden.

Beweis: Wir führen den Beweis indem wir zeigen, dass die moment-erzeugende Funktion M_V mit der moment-erzeugenden Funktion einer χ^2-verteilten Zufalls-Größe mit n Freiheitsgraden übereinstimmt.

Aus den ersten beiden Voraussetzungen folgt, dass die moment-erzeugenden Funktionen der Zufalls-Größen U und W die folgende Gestalt haben:

$$M_U(t) = (1 - 2 \cdot t)^{-\frac{m}{2}} \quad \text{und} \quad M_W(t) = (1 - 2 \cdot t)^{-\frac{m+n}{2}}.$$

Nach Satz 27 gilt für die moment-erzeugenden Funktionen

$$M_W(t) = M_U(t) \cdot M_V(t)$$

Setzen wir hier die obigen Werte ein, so folgt

$$(1 - 2 \cdot t)^{-\frac{m+n}{2}} = (1 - 2 \cdot t)^{-\frac{m}{2}} \cdot M_V(t)$$

und daraus folgt sofort

$$M_V(t) = (1 - 2 \cdot t)^{-\frac{n}{2}}.$$

Dies ist aber genau die moment-erzeugende Funktion einer χ^2-verteilten Zufalls-Größe. Nach dem Eindeutigkeits-Satz 26 hat die Zufalls-Größe V also eine χ^2-Verteilung mit n Freiheitsgraden. $\square$

Aufgabe 33: Nehmen Sie an, dass die Zufalls-Größe X einer χ^2-Verteilung mit n Freiheitsgraden genügt. Berechnen Sie Erwartungswert und Varianz von X.

Aufgabe 34: Es sei X eine Zufalls-Größe, die einer χ^2-Verteilung mit 4 Freiheitsgraden genügt. Berechnen Sie die Verteilungs-Funktion F_X.

Kapitel 4

Induktive Statistik

Zwei wesentliche Aufgaben der Statistik sind das *Schätzen von Parametern* und das *Testen von Hypothesen*. Wir besprechen diese beiden Aspekte in den beiden folgenden Abschnitten.

4.1 Parameter-Schätzung

Bisher haben wir uns im wesentlichen mit der Wahrscheinlichkeits-Rechnung beschäftigt. Bei der Wahrscheinlichkeits-Rechnung haben wir ein Model eines Prozesses in Form eines Wahrscheinlichkeits-Raums. Mit Hilfe dieses Models berechnen wir dann die Wahrscheinlichkeit für das Auftreten bestimmter Ereignisse. Damit das möglich ist, müssen uns die Parameter des Models bekannt sein. Ein solcher Parameter wäre beispielsweise der Erwartungswert λ einer Poisson-Verteilung oder Erwartungswert μ und Varianz σ^2 einer normal-verteilten Zufalls-Größe.

Bei der Statistik ist die Situation anders herum. Wir haben zwar meistens ebenfalls ein Model in Form eines Wahrscheinlichkeits-Raums, allerdings sind diesmal die Parameter unbekannt. Um diese Parameter bestimmen zu können, beobachten wir Ereignisse und versuchen dann, mit Hilfe der Beobachtungen die Parameter zu erschließen. Wir haben diesen Prozess bereits in der Einführung am Beispiel der Ameisenzählung demonstriert. Wir wollen diesen Prozess jetzt formalisieren. Dazu benötigen wir einige Definitionen.

Definition 35 (Stichprobe) Gegeben sei eine endliche oder unendliche Menge Ω, die wir im folgenden als *Population* (oder auch Grundgesamtheit) bezeichnen wollen. Ein *Merkmal* X dieser Population ist eine Funktion

$$X : \Omega \to \mathbb{R}.$$

Eine *Stichprobe* vom Umfang n ist ist eine Teilmenge

$$\{\omega_1, \cdots, \omega_n\} \subseteq \Omega$$

die genau n Elemente enthält. Setzen wir $x_i := X(\omega_i)$ für $i = 1, \cdots, n$, so werden die x_i auch als *Stichproben-Werte* bezeichnet. □

4.1.1 Schätzung des Erwartungswertes einer Zufalls-Größe

Ist eine Stichprobe gegeben, so können wir versuchen, von der Stichprobe auf die Grundgesamtheit zu schließen. Ist beispielsweise $\{\omega_1, \cdots, \omega_n\}$ eine Stichprobe und X eine Zufalls-Größe, so wird der *arithmetische Mittelwert* von X mit $\overline{X}$ bezeichnet und durch

$$\overline{X} := \frac{1}{n} \cdot \sum_{i=1}^{n} X(\omega_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

definiert. Um das Verhalten von $\overline{X}$ mit den Mitteln der Wahrscheinlichkeits-Rechnung untersuchen zu können, gehen wir aus von dem Wahrscheinlichkeits-Raum

$$\mathcal{V} = \langle \Omega, \mathfrak{A}, P \rangle$$

und definieren das n-fache kartesische Produkt

$$\mathcal{W} = \mathcal{V}^n = \langle \Omega^n, \mathfrak{A}^n, \widehat{P} \rangle.$$

Für ein Tupel von Ereignissen $\langle A_1, \cdots, A_n \rangle$ ist die Wahrscheinlichkeits-Verteilung $\widehat{P}$ dann wie folgt definiert

$$\widehat{P}\big(\langle A_1, \cdots, A_n \rangle\big) := P(A_1) * \cdots * P(A_n).$$

Dann können wir auf Ω^n die Zufalls-Größen X_i definieren als

$$X_i\big(\langle \omega_1, \cdots, \omega_i, \cdots, \omega_n \rangle\big) := X(\omega_i),$$

X_i ist also gerade die Anwendung von X auf die i-te Komponente der Stichprobe. Die Wahrscheinlichkeits-Verteilungen der Zufalls-Größen X_i sind dann alle gleich der Wahrscheinlichkeits-Verteilung der Zufalls-Größe X auf dem ursprünglichen Raum $\mathcal{V}$.

Mit den X_i ist auch der arithmetische Mittelwert $\overline{X}$ eine Zufalls-Größe auf dem Stichproben-Raum Ω^n. Wir können den Erwartungswert von $\overline{X}$ berechnen und finden

$$\begin{aligned}
E\Big[\overline{X}\Big] &= E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} E\left[X_i\right] \\
&= \frac{1}{n} \cdot \sum_{i=1}^{n} E\left[X\right] \\
&= \frac{1}{n} \cdot E\left[X\right] \cdot \sum_{i=1}^{n} 1 \\
&= \frac{1}{n} \cdot E\left[X\right] \cdot n \\
&= E\left[X\right].
\end{aligned}$$

Diese Rechnung zeigt, dass wir den arithmetischen Mittelwert einer Stichprobe benutzen können um den Erwartungswert der Zufalls-Größe X zu *schätzen*. Ist

$$\langle x_1, \cdots, x_n \rangle = \langle X(\omega_1), \cdots, X(\omega_n) \rangle$$

ein *Stichproben-Ergebnis*, so liefert uns dieses Ergebnis eine Schätzung des Mittelwerts $E[X]$:

$$E[X] \approx \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \overline{X}\big(\langle \omega_1, \cdots, \omega_n \rangle\big).$$

Wir untersuchen noch die Varianz der Zufalls-Größe $\overline{X}$, denn diese gibt uns Aufschluss über die Genauigkeit unserer Schätzung. Mit einer Rechnung, die analog zur Herleitung des $\sqrt{n}$-Gesetzes ist, kann gezeigt werden, dass

$$\mathrm{Var}\Big[\overline{X}\Big] = \frac{1}{n} \cdot \mathrm{Var}[X] \quad \text{und damit} \quad \sigma\Big[\overline{X}\Big] = \frac{1}{\sqrt{n}} \cdot \sigma[X]$$

gilt. Dies zeigt, wie die Genauigkeit mit wachsender Größe n der Stichprobe zunimmt. Der zentrale Grenzwertsatz zeigt, dass $\overline{X}$ für große Werte von n annähernd normal verteilt ist. In der Praxis zeigt sich, dass die Wahrscheinlichkeits-Verteilung von $\overline{X}$ für $n \geq 30$ hinreichend gut durch eine Normal-Verteilung approximiert wird.

4.1.2 Schätzung der Varianz einer Zufalls-Größe

Als nächstes suchen wir eine Zufalls-Größe, mit der wir die Varianz einer Zufalls-Größe X abschätzen können. Dazu definieren wir die Stichproben-Varianz S^2 einer Stichprobe vom Umfang n als

$$S^2 := \frac{1}{n} \cdot \sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2$$

Die Zufalls-Größe S^2 ist genau wie die Zufalls-Größe $\overline{X}$ auf dem Stichproben-Raum Ω^n definiert. Wir werden im folgenden die Wahrscheinlichkeits-Verteilung von S^2 berechnen. In der Literatur wird die Stichproben-Varianz gelegentlich mit dem Faktor $\frac{1}{n-1}$ anstelle des Faktors $\frac{1}{n}$ definiert. Wir folgen bei der obigen Definition der Darstellung von Spiegel [SSS00].

Wir berechnen jetzt den Erwartungswert der Zufalls-Größe S^2. Dazu definieren wir

$$\mu := E[X] \quad \text{und} \quad \sigma^2 := \mathrm{Var}[X].$$

Zunächst formen wir die Summe $\sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2$ um:

$$\begin{aligned}
& \sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2 \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu + \mu - \overline{X}\big)^2 \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 + \big(\mu - \overline{X}\big)^2 + 2 \cdot (\mu - \overline{X}) \cdot (X_i - \mu) \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 + \sum_{i=1}^{n} \big(\mu - \overline{X}\big)^2 + \sum_{i=1}^{n} 2 \cdot (\mu - \overline{X}) \cdot (X_i - \mu) \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 + \sum_{i=1}^{n} \big(\mu - \overline{X}\big)^2 + 2 \cdot (\mu - \overline{X}) \cdot \sum_{i=1}^{n} (X_i - \mu) \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 + n \cdot \big(\mu - \overline{X}\big)^2 + 2 \cdot (\mu - \overline{X}) \cdot n \cdot (\overline{X} - \mu) \\
& \text{denn } \sum_{i=1}^{n} X_i = n \cdot \overline{X} \text{ und } \sum_{i=1}^{n} \mu = n \cdot \mu \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 + n \cdot \big(\mu - \overline{X}\big)^2 - 2 \cdot n \cdot \big(\mu - \overline{X}\big)^2 \\
= \;& \sum_{i=1}^{n} \big(X_i - \mu\big)^2 - n \cdot \big(\mu - \overline{X}\big)^2
\end{aligned}$$

Wir haben also

$$\sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2 = \sum_{i=1}^{n} \big(X_i - \mu\big)^2 - n \cdot \big(\mu - \overline{X}\big)^2 \tag{4.1}$$

gezeigt. Damit können wir $E\big[S^2\big]$ berechnen:

$$\begin{aligned}
E\big[S^2\big] &= \frac{1}{n} \cdot E\left[\sum_{i=1}^{n} \big(X_i - \overline{X}\big)^2\right] \\
&= \frac{1}{n} \cdot E\left[\left(\sum_{i=1}^{n} \big(X_i - \mu\big)^2\right) - n \cdot \big(\mu - \overline{X}\big)^2\right] \\
&= \frac{1}{n} \cdot \left(\left(\sum_{i=1}^{n} E\left[\big(X_i - \mu\big)^2\right]\right) - n \cdot E\left[\big(\mu - \overline{X}\big)^2\right]\right) \\
&= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} E\left[\big(X_i - \mu\big)^2\right] \; - \; \sum_{i=1}^{n} E\left[\big(\mu - \overline{X}\big)^2\right]\right)
\end{aligned}$$

Nun gilt einerseits

$$E\left[\big(X_i - \mu\big)^2\right] = E\left[\big(X - \mu\big)^2\right] = \sigma^2,$$

denn die Zufalls-Größen X_i haben ja alle die selbe Verteilung wie die Zufalls-Größe X, andererseits haben wir wegen $E\left[\overline{X}\right] = E[X] = \mu$

$$E\left[\left(\mu - \overline{X}\right)^2\right] = \operatorname{Var}\left[\overline{X}\right] = \frac{1}{n} \cdot \sigma^2.$$

Das ergibt insgesamt

$$\begin{aligned} E\left[S^2\right] &= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} E\left[\left(X_i - \mu\right)^2\right] \; - \; \sum_{i=1}^{n} E\left[\left(\mu - \overline{X}\right)^2\right] \right) \\ &= \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \sigma^2 \; - \; \sum_{i=1}^{n} \frac{1}{n} \cdot \sigma^2 \right) \\ &= \frac{1}{n} \cdot \left(n \cdot \sigma^2 - \sigma^2\right) \\ &= \frac{n-1}{n} \cdot \sigma^2 \end{aligned}$$

Auf den ersten Blick mag dieses Ergebnis verblüffen, denn es zeigt, dass die Stichproben-Varianz S^2 nicht dazu geeignet ist, die Varianz der Zufalls-Größe S zu schätzen. Es gilt aber

$$E\left[\frac{n}{n-1} \cdot S^2\right] = \sigma^2,$$

so dass wir die Zufalls-Größe

$$\widehat{S}^2 = \frac{n}{n-1} \cdot S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

zur Schätzung der Varianz verwenden können.

Der nächste Satz gibt uns Aufschluss über die Verteilung der Zufalls-Größe S^2 für den Fall, dass die zugrundeliegende Zufalls-Größe X normal verteilt ist und zeigt gleichzeitig die Bedeutung der χ^2-Verteilung.

Satz 36 Ist die Zufalls-Größe X normal verteilt mit Erwartungswert μ und Varianz σ^2, und wird für eine n-elementige Stichprobe die Zufalls-Größe S^2 durch

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

definiert, so hat die Zufalls-Größe

$$\frac{n}{\sigma^2} \cdot S^2 = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

eine χ^2-Verteilung mit $n-1$ Freiheitsgraden.

Beweis: Wir definieren drei Zufalls-Größen U, V und W durch

$$U := \frac{n}{\sigma^2} \cdot \left(\overline{X} - \mu\right)^2, \; V := \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \quad \text{und} \quad W := \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} \left(X_i - \mu\right)^2.$$

Um zu sehen, wie diese Zufalls-Größen zusammenhängen, schreiben wir Gleichung (4.1) um:

$$\sum_{i=1}^{n} \left(X_i - \mu\right)^2 = n \cdot \left(\mu - \overline{X}\right)^2 \; + \; \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

Wenn wir diese Gleichung durch σ^2 dividieren und die Gleichung $(\mu - \overline{X})^2 = (\overline{X} - \mu)^2$ berücksichtigen, dann sehen wir, dass

$$W = U + V$$

gilt. Wir untersuchen jetzt, wie die Zufalls-Größen U, V und W verteilt sind. Wir beginnen mit der Zufalls-Größe W. Zunächst ist mit X_i auch $X_i - \mu$ normal verteilt und die Zufalls-Größe $\frac{1}{\sigma} \cdot (X_i - \mu)$ ist normal verteilt mit Erwartungswert 0 und Varianz 1, also standard-normal-verteilt. Nach Satz 32 hat die Zufalls-Größe U daher eine χ^2-Verteilung mit n Freiheitsgraden.

Als nächstes untersuchen wir die Verteilung der Zufalls-Größe U. Nach Aufgabe 29 ist die Summe zweier normal verteilter Zufalls-Größen normal verteilt. Damit ist dann auch die Summe von n Zufalls-Größen normal verteilt. Also ist die Zufalls-Größe $\overline{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$ normal verteilt. Also ist auch $\overline{X} - \mu$ normal verteilt. Der Erwartungswert von $\overline{X} - \mu$ ist 0 und die Varianz ist nach dem $\sqrt{n}$-Gesetz $\frac{1}{n} \cdot \sigma^2$. Also hat die Zufalls-Größe $\frac{\sqrt{n}}{\sigma} \cdot (\overline{X} - \mu)$ den Erwartungswert 0 und die Varianz 1. Wenn wir nun Satz 32 für den Fall $n = 1$ anwenden, sehen wir, dass die Zufalls-Größe U, die ja das Quadrat der Zufalls-Größe $\frac{\sqrt{n}}{\sigma} \cdot (\overline{X} - \mu)$ ist, einer χ^2-Verteilung mit Freiheitsgrad $n = 1$ genügt.

Damit können wir jetzt Satz 34 auf die Zufalls-Größen U, V und W anwenden und folgern, dass die Zufalls-Größe V einer χ^2-Verteilung mit $n - 1$ Freiheitsgraden genügt. □

4.2 Testen von Hypothesen

Das Testen von Hypothesen ist eine der beiden Hauptaufgaben der Statistik, die besonders in der Medizin angewendet wird. Wir erläutern das Testen von Hypothesen an einem einführenden Beispiel. Wir wollen untersuchen, ob sich beim Werfen einer gegeben Münze die Ergebnisse "Kopf" und "Zahl" mit der selben Wahrscheinlichkeit einstellen. Wir kodieren das Ergebnis "Kopf" als 0 und das Ergebnis "Zahl" als 1. Unsere Hypothese ist, dass bei einem Wurf für die Wahrscheinlichkeit

$$P(\{0\}) = P(\{1\}) = \frac{1}{2}$$

gilt. Diese Hypothese bezeichnen wir als *Null-Hypothese*, denn es gibt Null Unterschied zwischen den beiden Wahrscheinlichkeiten. Um diese Hypothese zu überprüfen, werfen wir die Münze 100 mal. Angenommen, wir erhalten dabei 75 mal Kopf und 25 mal Zahl. Dann können wir uns fragen, wie wahrscheinlich ein solches Ereignis ist, wenn wir die Null-Hypothese annehmen. Wenn diese Wahrscheinlichkeit kleiner ist als ein vorgegebener Wert α, dann würden wir die Null-Hypothese ablehnen und sagen, dass der Test auf dem *Niveau α signifikant* war. In der Praxis wird für α oft ein Wert von 5% oder 1% gewählt.

Bezeichnen wir die Häufigkeit, mit der das Ereignis "Kopf" eintritt ist, mit X_0 und die Häufigkeit, mit der das Ereignis "Zahl" eintritt, mit X_1, und ist n die Anzahl der Münzwürfe, so misst die Größe

$$C := \frac{(X_0 - n \cdot \frac{1}{2})^2}{n \cdot \frac{1}{2}} + \frac{(X_1 - n \cdot \frac{1}{2})^2}{n \cdot \frac{1}{2}}$$

wie stark X_0 und X_1 von dem Erwartungswert $n \cdot \frac{1}{2}$ abweichen. Eine solche Zufalls-Größe wird auch als eine *Statistik* bezeichnet. Für die oben genannten Werte von $X_0 = 75$ und $X_1 = 25$ ergibt sich für die Statistik C

$$C = \frac{(75 - 50)^2}{50} + \frac{(25 - 50)^2}{50} = 2 \cdot \frac{25^2}{50} = 25$$

Um diesen Wert interpretieren zu können, berechnen wir die Wahrscheinlichkeit, dass die Zufalls-Größe C einen Wert größer oder gleich 25 annimmt, unter der Voraussetzung, dass die Null-Hypothese wahr ist. Dazu formen wir den Ausdruck für C um:

$$
\begin{aligned}
C &= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} + \frac{\left(X_1 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} \\
&= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} + \frac{\left(n - X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}} && \text{denn } X_1 = n - X_0 \\
&= \frac{1}{n\cdot\frac{1}{2}}\cdot\left(\left(X_0 - n\cdot\frac{1}{2}\right)^2 + \left(n - X_0 - n\cdot\frac{1}{2}\right)^2\right) \\
&= \frac{1}{n\cdot\frac{1}{2}}\cdot\left(\left(X_0 - n\cdot\frac{1}{2}\right)^2 + \left(n\cdot\frac{1}{2} - X_0\right)^2\right) \\
&= \frac{2}{n\cdot\frac{1}{2}}\cdot\left(X_0 - n\cdot\frac{1}{2}\right)^2 \\
&= \frac{\left(X_0 - n\cdot\frac{1}{2}\right)^2}{n\cdot\frac{1}{2}\cdot\left(1-\frac{1}{2}\right)}
\end{aligned}
$$

Wenn die Null-Hypothese zutrifft, dann ist die Zufalls-Größe X_0 binomial verteilt mit dem Erwartungswert $\mu = n\cdot\frac{1}{2}$ und der Varianz $\sigma^2 = n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})$. Für große Werte von n kann diese Verteilung durch eine Normal-Verteilung $N_{\mu,\sigma}$ mit Mittelwert μ und Varianz σ approximiert werden. Brauchbare Werte erhalten wir, sobald $\sigma^2 > 9$ gilt. Für $n = 100$ haben wir $\sigma^2 = 25$, so dass diese Bedingung erfüllt ist. Wenn X_0 aber ein Normal-Verteilung hat, so hat die Zufalls-Größe

$$
Y := \frac{X_0 - n\cdot\frac{1}{2}}{\sqrt{n\cdot\frac{1}{2}\cdot(1-\frac{1}{2})}}
$$

eine Standard-Normal-Verteilung $N_{0,1}$. Im letzten Kapitel haben wir gesehen, dass dann Y^2 eine χ^2-Verteilung mit einem Freiheitsgrad genügt. Die Zufalls-Größe Y^2 ist aber gleich C, so dass wir damit die Verteilung von C gefunden haben. Daher können wir jetzt die Wahrscheinlichkeit, dass C einen Wert größer oder gleich 25 annimmt, berechnen. Für die Wahrscheinlichkeits-Dichte einer χ^2-Verteilung mit einem Freiheitsgrad hatten wir im letzten Kapitel den Ausdruck

$$
p_{\chi^2,1}(x) = \frac{1}{A_1}\cdot\frac{1}{\sqrt{x}}\cdot\exp\left(-\frac{x}{2}\right) \qquad \text{mit } A_1 = \sqrt{2}\cdot\Gamma\left(\frac{1}{2}\right) = \sqrt{2\pi}
$$

gefunden. Wir setzen zur Abkürzung $z = 25$ und rechnen wie folgt:

$$
\begin{aligned}
P(C \geq 25) &= 1 - P(C < 25) \\
&\approx 1 - \int_0^z p_{\chi^2,1}(x)\,\mathrm{d}x \\
&= 1 - \int_0^z \frac{1}{\sqrt{2\pi\cdot x}}\cdot\exp\left(-\frac{x}{2}\right)\,\mathrm{d}x \\
&\ \ \text{Substitution } x = 2\,y^2,\text{ also } \mathrm{d}x = 4\cdot y\,\mathrm{d}y \\
&= 1 - \int_0^{\sqrt{z/2}} \frac{1}{\sqrt{4\pi}\cdot y}\cdot\exp\left(-y^2\right)\cdot 4\cdot y\,\mathrm{d}y \\
&= 1 - \frac{2}{\sqrt{\pi}}\cdot\int_0^{\sqrt{z/2}} \cdot\exp\left(-y^2\right)\,\mathrm{d}y \\
&= 1 - \mathit{erf}\left(\sqrt{\frac{z}{2}}\right) = 1 - \mathit{erf}\left(\frac{5}{\sqrt{2}}\right) \approx 5.7\cdot 10^{-7}
\end{aligned}
$$

Wenn die Null-Hypothese zutrifft, dann ist die Wahrscheinlichkeit, dass die Zufalls-Größe C einen Wert größergleich 25 annimmt, verschwindend gering. Daher verwenden wir die Null-Hypothese verwerfen. □

Wir verallgemeinern das letzte Beispiel und betrachten nun eine Anwendung der Statistik aus dem wirklichen Leben.

Beispiel: Die Würfel, die an der Berufsakademie zur Notenfindung eingesetzt werden, werden jedes Jahr in einem aufwendigen Prozess[1] geeicht. Dazu wird mit dem zu eichenden Würfel 100 mal gewürfelt und die Häufigkeit der einzelnen Ziffern wird notiert. Da bei der BA nur Noten von 1 bis 5 vergeben werden, handelt es sich bei den verwendeten Würfeln um kostspielige Spezialanfertigungen, mit denen sich nur Zahlen aus der Menge $\{1, \cdots, 5\}$ würfeln lassen. Bei meinem Würfel ergab sich die dabei die folgenden Verteilung der Häufigkeiten:

Zahl	1	2	3	4	5
Häufigkeit	10	16	14	35	25

Wir würden erwarten, dass jede Zahl mit einer durschnittlichen Häufigkeit von $100 \cdot \frac{1}{5} = 20$ in der Tabelle erscheint, aber es ist auch klar, dass es statistische Schwankungen dieser Häufigkeiten geben wird. Für $i = 1, \cdots, 5$ bezeichnen wir die Häufigkeit des Auftretens der Zahl i mit X_i. Als ein Mass für die Abweichung der oben gezeigten Häufigkeiten von dem Erwartungswert verwenden wir die Zufalls-Größe C, die wir in Analogie zu dem letzten Beispiel als

$$C = \sum_{i=1}^{5} \frac{(X_i - 20)^2}{100 \cdot \frac{1}{5}} = \frac{1}{20} \cdot \sum_{i=1}^{5} (X_i - 20)^2$$

definieren. Für meinen Würfel ergibt sich dabei ganz konkret der Wert

$$C = \frac{1}{20} \cdot (10^2 + 4^2 + 6^2 + 15^2 + 5^2) = \frac{402}{20} = 20.1$$

Wir fassen C nun als Zufalls-Größe auf und überlegen uns, wie wahrscheinlich es ist, dass die Zufalls-Größe C einen Wert größer oder gleich 20.1 annimmt, wir berechnen also die Wahrscheinlichkeit

$$P(C \geq 20.1).$$

Um diese Wahrscheinlichkeit berechnen zu können, müssen wir zunächst eine Annahme machen, die uns zeigt, wie die Zufalls-Größen X_i verteilt sind. Wir machen die Annahme, dass es sich bei dem Würfel um einen Laplace-Würfel handelt, dass also die Wahrscheinlichkeit für das Auftreten jeder einzelnen Ziffer den Wert $\frac{1}{5}$ hat. Diese Annahme ist jetzt die zu prüfende *Null-Hypothese*.

Die Zufalls-Größen X_i sind dann binomial verteilt und es gilt

$$P\big(X_i = k\big) = \binom{100}{k} \cdot \left(\frac{1}{5}\right)^k \cdot \left(\frac{4}{5}\right)^{100-k}.$$

Da $100 \cdot \frac{1}{5} \cdot \frac{4}{5} = 16 > 9$ ist, können wir diese Verteilung durch eine Normalverteilung approximiert. Wir setzen also

$$\mu = 100 \cdot \frac{1}{5} = 20 \quad \text{und} \quad \sigma^2 = 100 \cdot \frac{1}{5} \cdot \frac{4}{5} = 16.$$

und haben dann für die Verteilung der Zufalls-Größen X_i die Approximation

$$P\big(X_i \leq k\big) \approx \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{k} \exp\left(-\frac{(k-\mu)^2}{2 \cdot \sigma^2}\right) \,\mathrm{d}x$$

In Analogie zur letzten Aufgabe kann gezeigt werden, dass die Zufalls-Größe C einer χ^2-Verteilung mit $5-1$ Freiheitsgraden genügt. Die 1, die wir von der 5 abziehen, hat ihre Ursache in der Tatsache, dass die fünf Zufalls-Größen X_1, $\cdots$, X_5 nicht unabhängig sind, denn wenn wir X_1, X_2, X_3 und X_4 kennen, dann können wir X_5 über die Formel $X_1 + X_2 + X_3 + X_4 + X_5 = 100$ berechnen. Dies erklärt, dass die χ^2-Verteilung nicht fünf, sondern nur vier Freiheitsgrade hat. Wir wollen dieses Ergebnis hier nicht beweisen sondern nur darauf hinweisen, dass die Situation ganz ähnlich ist wie im Satz 36. Also gilt

[1] Unter den Studenten ist dieser Prozess als die sogenannte *Assistenten-Prüfung* bekannt.

$$
\begin{aligned}
P(C \geq 20.1) &= 1 - P(C < 20.1) \\
&= 1 - P\left(C \leq z\right) \quad \text{mit } z := 20.1 \\
&= 1 - \int_0^z \frac{1}{A_4} \cdot x \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x \\
&\quad \text{mit} \quad A_4 = 2^2 \cdot \Gamma(2) = 4
\end{aligned}
$$

Wir berechnen das Integral separat:

$$
\begin{aligned}
& \int_0^z \frac{1}{4} \cdot x \cdot \exp\left(-\frac{x}{2}\right)\, \mathrm{d}x \quad \text{mit der Substitution } y = \tfrac{x}{2} \text{ wird daraus} \\
= \; & \int_0^{z/2} y \cdot \exp\left(-y\right)\, \mathrm{d}y \\
& \text{partielle Integration: } u(y) = y,\ v'(y) = \exp(-y) \\
& \text{also:} \qquad\qquad\quad u'(y) = 1,\ v(y) = -\exp(-y) \\
= \; & -y \cdot \exp(-y)\Big|_0^{z/2} + \int_0^{z/2} \exp\left(-y\right)\, \mathrm{d}y \\
= \; & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) + \int_0^{z/2} \exp\left(-y\right)\, \mathrm{d}y \\
= \; & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) - \left(-\exp(-y)\Big|_0^{z/2}\right) \\
= \; & -\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) - \exp\left(-\frac{z}{2}\right) + 1
\end{aligned}
$$

Damit finden wir für die gesuchte Wahrscheinlichkeit $P(C \geq 20.1)$

$$
\begin{aligned}
P(C \geq 20.1) &= 1 - \left(-\frac{z}{2} \cdot \exp\left(-\frac{z}{2}\right) - \exp\left(-\frac{z}{2}\right) + 1\right) \\
&= \left(\frac{z}{2} + 1\right) \cdot \exp\left(-\frac{z}{2}\right) \\
&\approx 4.8 \cdot 10^{-4}
\end{aligned}
$$

Diese Wahrscheinlichkeit ist sehr klein. Dies deutet mit hoher Sicherheit darauf hin, dass die Null-Hypothese falsch ist. Wir müssen daher die Null-Hypothese verwerfen und davon ausgehen, dass sich die Wahrscheinlichkeit der einzelnen Zahlen bei meinem Würfel deutlich von dem Wert $\frac{1}{5}$ unterscheidet. Also sollte ich mir von den Studiengebühren einen neuen Würfel beschaffen. $\square$

Bemerkung: Das in dem obigen Beispiel skizzierte Verfahren wird als χ^2-Test bezeichnet. Die nächste Aufgabe zeigt, wie dieser Test in der Medizin angewendet wird.

Aufgabe 35: Zur Untersuchung der Frage, ob ein Zusammenhang zwischen Zigaretten-Rauchen und dem Auftreten von Lungenkrebs besteht, wurde eine Gruppe von 30 000 Rauchern über einen Zeitraum von 10 Jahren beobachtet. Am Ende des Tests wurde für jede einzelne Person überprüft, ob die Person während des Tests an Lungenkrebs erkrankt ist. Dabei ergaben sich die folgenden Ergebnisse:

	Raucher	Nichtraucher	Gesamt
Lungenkrebs	62	14	76
kein Lungenkrebs	9 938	19 986	29 924
Gesamt	10 000	20 000	30 000

Die zu testende Null-Hypothese lautet in diesem Fall, dass es zwischen dem Rauchen und dem Auftreten von Lungenkrebs keinen Zusammenhang gibt. Folglich hätte die Wahrscheinlichkeit p, dass eine Person an Lungenkrebs erkrankt, in beiden Fällen den selben Wert. Im Gegensatz zu dem letzten Beispiel kennen wir den Wert von p nicht und müssen ihn daher schätzen. Als Schätzwert wählen wir das Verhältnis der Gesamtzahl der Lungenkrebs-Erkrankungen zu der Zahl aller Test-Personen und erhalten

$$p = \frac{76}{30\,000} = 0.25\overline{3}\,\%.$$

Überprüfen Sie die Null-Hypothese mit Hilfe des χ^2-Tests.

Bemerkung: Selbstverständlich handelt es sich bei allen Zahlen, die ich den Beispielen verwende um Fakten. Bei den obigen Zahlen handelt es sich sogar um Fakten, die ich mir nicht selber ausgedacht habe. Sie können diese Zahlen in dem Buch von Sheldon M. Ross [Ros04] wiederfinden.

Lösung: Wir bezeichnen die Anzahl der an Lungenkrebs erkrankten Rauchern mit $X_{1,1}$, die Anzahl der Raucher, die nicht erkrankt sind mit $X_{1,2}$, die Anzahl der erkrankten Nichtraucher mit $X_{2,1}$ und die Anzahl der Nichtraucher, die nicht erkrankt sind, mit $X_{2,2}$. Weiter sei N_1 die Anzahl der Raucher und N_2 sei die Anzahl der Nichtraucher, es gilt also

$$N_1 = 10\,000 \quad \text{und} \quad N_2 = 20\,000.$$

In Analogie zu den vorigen Beispielen definieren wir die Zufalls-Größe C als

$$C = \frac{\big(X_{1,1} - N_1 \cdot p\big)^2}{N_1 \cdot p} + \frac{\big(X_{1,2} - N_1 \cdot (1-p)\big)^2}{N_1 \cdot (1-p)} + \frac{\big(X_{2,1} - N_2 \cdot p\big)^2}{N_2 \cdot p} + \frac{\big(X_{2,2} - N_2 \cdot (1-p)\big)^2}{N_2 \cdot (1-p)}$$

Setzen wir die konkreten Werte ein, so erhalten wir

$$C \approx 79.81$$

Wie in den vorherigen Beispielen auch hat C eine χ^2-Verteilung, wir müssen aber noch die Anzahl der Freiheitsgrade bestimmen.

1. Wir haben vier Zufalls-Größen $X_{1,1}$, $X_{1,2}$, $X_{2,1}$ und $X_{2,2}$.
2. Zwischen diesen vier Zufalls-Größen gibt es drei verschiedene Beziehungen:
   (a) $X_{1,1} + X_{1,2} = N_1$
   (b) $X_{2,1} + X_{2,2} = N_2$
   (c) $\dfrac{X_{1,1} + X_{2,1}}{N_1 + N_2} = p$

Daher hat die χ^2-Verteilung nur $4 - 3 = 1$ Freiheitsgrad! Setzen wir $z := 79.81$ so gilt also

$$\begin{aligned} P(C \geq 79.81) &= P(C \geq z) \\ &= 1 - P(C < z) \\ &\approx 1 - \int_0^z p_{\chi^2,1}(x)\;\mathrm{d}x \\ &= 1 - \int_0^z \frac{1}{\sqrt{2\pi \cdot x}} \cdot \exp\left(-\frac{x}{2}\right)\;\mathrm{d}x \end{aligned}$$

Das Integral, das hier auftritt, haben wir bereits zu Beginn dieses Abschnitts berechnet. Es gilt also

$$P(C \geq z) \approx 1 - \mathit{erf}\left(\sqrt{\frac{z}{2}}\right) \approx 0.41 \cdot 10^{-18}$$

Dieser Wert ist so klein, dass der Zufall als Ursache für die höhere Zahl der Lungenkrebs-Erkrankungen bei den Rauchern praktisch ausgeschlossen werden kann. Das heißt natürlich noch lange nicht, dass zwischen dem Rauchen und dem Auftreten von Lungenkrebs eine kausale Beziehung besteht, denn die beobachtete Korrelation könnte auch andere Ursachen haben:

1. Nehmen wir einmal an, dass Rauchen vor Aids schützt[2]. Wenn dann die Nichtraucher zu einem nennenswerten Prozentsatz an Aids wegsterben würden, noch bevor Sie ihre Ansprüche auf einen vollwertigen Lungenkrebs realisieren könnten, so würde dies die beobachteten Unterschiede erklären.

2. Es wäre theoretisch möglich, dass es ein Gen gibt, dass einerseits das Auftreten von Lungenkrebs begünstigt, andererseits aber auch den gesunden Menschenverstand dahingehend beeinträchtigt, dass Personen, die dieses Gen besitzen, verstärkt zu Rauchern werden. Wenn dies der Fall wäre, so würde es einem Raucher wenig nützen, wenn er das Rauchen aufgibt, denn einerseits würde sich dadurch sein Risiko, an Lungenkrebs zu erkanken, nicht ändern und andererseits würde sich auch sein Intelligenz-Quotient nicht erhöhen. Das einzige was sich erhöhen würde, wäre das Risiko an Aids zu erkanken.

   Einer der größten Statistiker des letzten Jahrhunderts, Sir Ronald Fisher (1890 - 1962), war tatsächlich der Ansicht, dass die Korrelation zwischen dem Rauchen und dem Auftreten von Lungenkrebs genetisch bedingt ist. Der Fairness halber soll erwähnt werden, dass zu dem Zeitraum, an dem Sir Ronald diese Hypothese aufstellte, der medizinische Wissensstand noch keine klare Aussage zuließ. Der Vollständigkeit halber soll allerdings auch erwähnt werden, dass Sir Ronald bei der Tabak-Industrie unter Vertrag stand und selber ein leidenschaftlicher Zigarren-Raucher war. Und um die Geschichte abzuschließen, bemerken wir noch, dass Sir Ronald dann auch nicht an Lungenkrebs gestorben ist, sondern an Kehlkopfkrebs. Dies ist die bevorzugte Krebsart bei Zigarren-Rauchern und hat in etwa den selben Spaßfaktor wie Lungenkrebs. □

Wir fassen die bisherigen Beispiele in einem Satz zusammen, den wir allerdings nicht beweisen können.

Satz 37 (χ^2-Test) Es seien n verschiedene Zufalls-Größen X_1, $\cdots$, X_n gegeben. Weiter sei eine Null-Hypothese H_0 gegeben. Wenn diese Hypothese zutrifft, dann gelte:

1. Die Zufalls-Größen genügen (zumindest näherungsweise) einer Normalverteilung.

2. Zwischen den Zufalls-Größen bestehen k *unabhängige* Beziehungen.

   Mit *unabhängig* ist hier gemeint, dass keine der Beziehungen aus den restlichen $k-1$ Beziehungen gefolgert werden kann.

Dann genügt die Zufalls-Größe

$$C = \sum_{i=1}^{n} \frac{(X_i - E[X_i])^2}{E[X_i]}$$

einer χ^2-Verteilung mit $n-k$ Freiheitsgraden. Nimmt die Zufalls-Größe also einen bestimmen Wert z an, so gilt für die Wahrscheinlichkeit, dass $C \geq z$ ist

$$P(C \geq z) = 1 - \int_0^z p_{\chi^2,n-k}(x)\ \mathrm{d}x.$$

Ist diese Wahrscheinlichkeit kleiner als ein vorgegebener Wert α (in der Praxis nimmt man oft $\alpha = 0.05$, $\alpha = 0.01$ oder $\alpha = 0.001$), so sagen wir, dass die Null-Hypothese auf dem Signifikanz-Niveau α verworfen werden kann. □

Aufgabe 36: Besuchen Sie im Internet die Seite

`http://www.cemetery.org/alphaindex.html`

Bei dieser Seite handelt es sich um einen virtuellen Friedhof. Notieren Sie die Wochentage des Ablebens der Verblichenen. Überprüfen Sie die Hypothese, dass für den Todestag eines Menschen alle Wochentage die selbe Wahrscheinlichkeit haben mit Hilfe des χ^2-Tests und interpretieren Sie das Ergebnis. □

---

[2] Diese Annahme ist keineswegs so absurd wie sie zunächst erscheinen mag, denn bekanntlich macht Rauchen häßlich.

4.3 Ausblick

Leider konnten wir in der zur Verfügung stehenden Zeit nur einen kleinen Teil der Statistik behandeln. Dabei sind viele Aspekte zu kurz gekommen, denn die Vorlesung hat sich ausschließlich mit dem mathematischen Teil der Statistik befaßt. Was die fehlenden Aspekte angeht, kann ich hier nur auf die am Ende des Skriptes angegebene Literatur verweisen. Die nicht-mathematischen Fragen der Statistik werden besonders anschaulich in dem Buch von Freedman et. al. [FPP98] behandelt. Der in der Vorlesung behandelte Stoff ist hingegen am klarsten in dem (sehr preiswerten) Buch von Spiegel [SSS00] beschrieben.

Literaturverzeichnis

[BH98] Friederich Barth and Rudolf Haller. *Stochastik Leistungskurs.* Oldenbourg Verlag, München, 1998.

[FPP98] David Freedman, Robert Pisani, and Roger Purves. *Statistics.* W. W. Norton & Company, New York, 3rd edition, 1998.

[GH85] Donald Gross and Carl M. Harris. *Fundamentals of Queueing Theory.* Wiley, 1985.

[GS97] Charles M. Grinstead and Laurie J. Snell. *Introduction to Probability.* American Mathematical Society, 1997. Available at `http://www.dartmouth.edu/~prob/prob/prob.pdf`.

[NIS06] NIST. *Engineering Statistics.* National Institute of Standards and Technology, 2006. http://www.itl.nist.gov/div898/handbook/index.htm.

[Pap99] Lothar Papula. *Mathematik für Ingenieure und Naturwissenschaftler.* Vieweg, 1999.

[Ros04] Sheldon M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists.* Elsevier Academic Press, 2004.

[SSS00] Murray R. Spiegel, John Schiller, and R. Alu Srinivasan. *Probability and Statistics.* McGraw-Hill; New York, 2000.