Elemente der Stochastik

R. Vogt

SS 2009

Von studentischer Seite wird immer wieder die Frage gestellt, warum in der Fachmathematik nicht mehr oder weniger fertige fachliche Unterrichtseinheiten vermittelt werden, die dann ohne allzu große Mühe in schulischen Unterricht umgesetzt werden können.

Die Frage zeigt deutlich, daß viele Studierende nur ein beschränktes Verständnis des Wesens der Mathematik besitzen. Die Gefahr besteht, daß dieses Verständnis in den späteren Unterricht einfließt, mit katastrophalen Ergebnissen, wie die mangelnde Akzeptanz der Mathematik durch die Schüler und das falsche Bild der Mathematik in der Öffentlichkeit nur allzu deutlich belegen.

Mathematik ist eine universelle einsetzbare Denkstruktur, die man sich nur durch intensives Training und Nachdenken aneignen kann. In diesem Sinne wird die Mathematik von einer zunehmenden Zahl anderer universitärer Fächer als die zentrale Hilfswissenschaft anerkannt.

In der Wahrscheinlichkeitstheorie läßt sich sehr deutlich die Vorgangsweise der Mathematik erläutern: Aus einfachen, alltäglichen Problemen über Wahrscheinlichkeiten wird ein mathematisches Modell in Form eines Axiomensystems abstrahiert, auf das man sich leicht einigen kann. Mit zunehmender Komplexität der zu untersuchenden Probleme muß dieses Modell, noch immer basierend auf Axiomen, verfeinert werden. Die dafür benötigten mathematischen Werkzeuge müssen weiterentwickelt werden, das Modell muß auf seine Stimmigkeit überprüft werden. Da dies weit über die Vermittlung mathematischer Rezepte hinausgeht, fällt gerade dieser Teil vielen Studierenden schwer. Natürlich ist dies wieder ein Hinweis auf die mangelnde mathematische Ausbildung in vielen Schulen, in denen Rechenfertigkeiten vermittelt werden, deren Sinn viele nicht einsehen, aber auf zentrale Anliegen der Mathematik nicht eingegangen wird.

Inhaltsverzeichnis

Teil I

Endliche Wahrscheinlichkeitstheorie

1 Das mathematische Modell

Beim Werfen einer Münze sei W das *Ereignis*, daß „Wappen“ auftritt. Wir wiederholen den Versuch n-mal. Sei

$$h_n(W)$$

die Anzahl des Ergebnisses „Wappen“ bei n Würfen.

1.1 Definition: $h_n(W)$ ist die *absolute Häufigkeit* des Ereignisses W bei dem angegebenen Versuch. Der Quotient

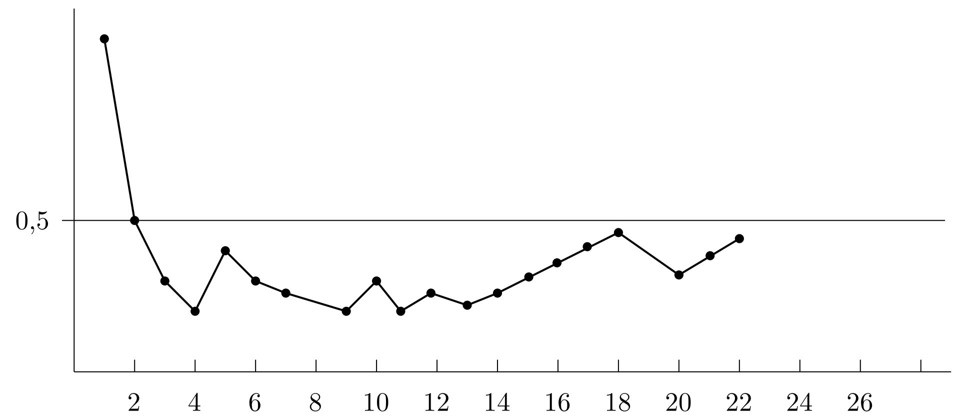
$$r_n(W) = \frac{h_n(W)}{n}$$

heißt *relative Häufigkeit* des Ereignisses W.

Trägt man $r_n(W)$ in einem Graphen auf (**Beachte:** Man macht nur **eine** Versuchsreihe mit sehr großem n), so wird der Graph für kleine n kräftig schwanken, sich aber für große n **nach allen Erfahrungen** bei einem festen Wert in der Nähe von 0,5 einpendeln. Dieser Wert hängt von der Beschaffenheit der Münze ab.

1.2 Beispielversuche

$Z, W, Z, Z, W, Z, Z, Z, Z, W, Z, W, Z, W, W, W, W, Z, Z, Z, W, W, Z$



Will man diesen Versuch mathematisch fassen, liegt es nahe, dem Ereignis W die Wahrscheinlichkeit

1.3 $$P(W) = \lim_{n\to\infty} r_n(W)$$

zuzuordnen. Ist das sinnvoll?

Schauen wir auf unser Experiment zurück. Skepsis muß die Bemerkung „nach allen Erfahrungen“ auslösen. So etwas läßt sich mathematisch nicht fassen. Weiter verlangt die Definition (1.3) eine unendliche Folge $r_n(W)$, und die liegt niemals vor. Selbst wenn sie vorläge, ist nicht klar, daß der Grenzwert existiert. Wir sehen also, (1.3) ist nur ein bedingt brauchbarer Ansatz.

Wie geht ein Mathematiker weiter vor? Er untersucht zunächst Eigenschaften der relativen Häufigkeit.

Betrachten wir wieder unser Experiment. Wir können uns für folgende Ereignisse interessieren:

W = Wappen tritt auf, Z = Zahl tritt auf.

Wir setzen

$$\Omega = \{W, Z\}$$

Ω ist die Menge der *Elementarereignisse*. Ihre speziellen Wahrscheinlichkeiten festzulegen, kann schwer sein, wie wir gesehen haben. Die Münze ist ja fast immer nicht „ideal“. Aber einige Dinge können wir sofort sehen: Dazu betrachten wir Teilmengen $A, B \subset \Omega$ und fragen uns, mit welcher Wahrscheinlichkeit der Ausgang unseres Experiments in A liegt.

1.4 Satz: Ein Versuch habe eine endliche Menge Ω von Ausgängen.

(1) $\forall\ A \subset \Omega$ gilt $0 \le r_n(A) \le 1$

(2) $r_n(\Omega) = 1$, man nennt Ω das *sichere Ereignis*.

(3) Seien $A, B \subset \Omega$ und $A \cap B = \emptyset$, dann gilt

$$r_n(A \cup B) = r_n(A) + r_n(B).$$

Beweis:

(1) Da $0 \le h_n(A) \le n$, folgt $0 \le r_n(A) \le 1$.

(2) Da jeder Versuchsausgang in Ω liegt, ist $h_n(\Omega) = n$. Also $r_n(\Omega) = 1$.

(3) Da $A \cap B = \emptyset$, ist $h_n(A \cup B) = h_n(A) + h_n(B)$. Es folgt

$$r_n(A \cup B) = r_n(A) + r_n(B).$$

$\square$

1.5 Bezeichnung: (1) Eine Teilmenge $A \subset \Omega$ nennen wir ein *Ereignis*, die Elemente von Ω heißen *Elementarereignisse*.

(2) Zwei Ereignisse A, B heißen *unverträglich*, falls $A \cap B = \emptyset$.

Satz 1.4 erheben wir nun zur Definition.

1.6 Definition: Ein *endlicher Wahrscheinlichkeitsraum* besteht aus einer endlichen Menge Ω und einer Abbildung

$$P : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$$

von der Potenzmenge $\mathcal{P}(\Omega)$ aller Teilmengen von Ω nach $\mathbb{R}$. Es gelten die Axiome

(1) $0 \leq P(A) \leq 1 \quad \forall\ A \subset \Omega$

(2) $P(\Omega) = 1$

(3) $P(A \cup B) = (P(A) + P(B)$ für alle $A, B \in \mathcal{P}(\Omega)$ mit $A \cap B = \emptyset$.

P heißt *Wahrscheinlichkeitsmaß*, kurz *W-Maß*, und (Ω, P) nennen wir einen endlichen W-Raum.

1.7 Bezeichnungen: (1) Seien $A, B \in \mathcal{P}(\Omega)$. Falls $A \cap B = \emptyset$, schreiben wir $A + B$ für $A \cup B$. Also beinhaltet $A + B$ u.a. stets, daß $A \cap B = \emptyset$.

(2) $A \backslash B = \{a \in A;\ a \notin B\}$

(3) $\mathcal{C}A = \Omega \backslash A$, das Komplement von A in Ω.

1.8 Folgerungen aus den Axiomen:

(1) $P(\mathcal{C}A) = 1 - P(A) \quad \forall\ A \in \mathcal{P}(\Omega)$

(2) Für das *unmögliche Ereignis* $\emptyset$ gilt $P(\emptyset) = 0$

(3) Für $A \subset B$ gilt $P(A) \leq P(B)$.

(4) $P(B\backslash A) = P(B \cap \mathcal{C}A) = P(B) - P(A \cap B) \quad \forall\, A, B \in \mathcal{P}(\Omega)$

(5) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall\, A, B \in \mathcal{P}(\Omega)$

(6) Sind $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$ paarweise unverträglich, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$, so gilt

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n).$$

Beweis:

(1) $\Omega = A + \mathcal{C}A$. Aus den Axiomen 2 und 3 folgt

$$1 = P(\Omega) = P(A) + P(\mathcal{C}A)$$

(2) $\emptyset = \mathcal{C}\Omega$. Also $P(\emptyset) = P(\mathcal{C}\Omega) = 1 - P(\Omega) = 1 - 1 = 0$

(4) $B = B \cap \Omega = B \cap (A + \mathcal{C}A) = B \cap A + B \cap \mathcal{C}(A) = B \cap A + B\backslash A$.
Es folgt

$$P(B) = P(B \cap A) + P(B\backslash A)$$

(3) Da $A \subset B$, folgt $B \cap A = A$. Aus (4) und Axiom 1 folgt:

$$P(B) = P(A) + P(B\backslash A) \geq P(A)$$

(5) $A \cup B = A \cap B + A\backslash B + B\backslash A$. Also folgt aus (4) und (6)

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

(6) durch Induktion nach $n \geq 2$. Für $n = 2$ ist dies Axiom 3.

Induktionsschritt:
Voraussetzung: Sind $A_1, \ldots, A_n, n \geq 2$, paarweise unverträglich, so gilt $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$.
Behauptung: Sind $A_1, \ldots, A_{n+1}$ paarweise unverträglich, so gilt

$$P(A_1 \cup \cdots \cup A_{n+1}) = P(A_1) + \cdots + P(A_n) + P(A_{n+1}).$$

Beweis: Sei $B = A_1 \cup \cdots \cup A_n$. Dann gilt

$$B \cap A_{n+1} = (A_1 \cup \cdots \cup A_n) \cap A_{n+1} = (A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1}) = \emptyset.$$

Aus Axiom 3 und der Induktionsvoraussetzung folgt

$$\begin{aligned} P(A_1 \cup \cdots \cup A_{n+1}) = P(B \cup A_{n+1}) &= P(B) + P(A_{n+1}) \\ &= P(A_1) + \cdots + P(A_n) + P(A_{n+1}). \end{aligned}$$

$\square$

In einem endlichen W-Raum (Ω, P) ist jedes Ereignis $A \subset \Omega$ eine endliche Menge $A = \{a_1, \ldots, a_r\}$. Es folgt aus (1.8) (6)

$$P(A) = P(\{a_1\} + \cdots + \{a_r\}) = P(\{a_1\}) + \cdots + P(\{a_r\}).$$

Damit erhalten wir sofort

1.9 Satz: Ein endlicher W-Raum (Ω, P) ist eindeutig gegeben durch die Menge Ω der Elementarereignisse und eine Funktion

$$f : \Omega \longrightarrow \mathbb{R}$$

genannt *Zähldichte* mit den Eigenschaften

(1) $f(\omega) \geq 0 \quad \forall\, \omega \in \Omega$

(2) $\sum\limits_{\omega \in \Omega} f(\omega) = 1.$

Der Zusammenhang zwischen Zähldichte f und W-Maß P ist gegeben durch:

$$\begin{aligned} f(\omega) &= P(\{\omega\}) \\ P(A) &= \sum_{\omega \in \Omega} f(\omega) \end{aligned}$$

(Dabei wird die leere Summe (d.h. $A = \emptyset$) 0 gesetzt.)

Der Beweis ist klar!

1.10 Ausblick: Mit (1.6) haben wir eine Definition für ein Wahrscheinlichkeitsmaß gefunden, die jede relative Häufigkeit offensichtlich erfüllt. Wir werden daraus eine Theorie entwickeln, die sich somit stets anwenden läßt. Bei konkreten Beispielen ist aber unser Anfangsproblem nicht gelöst. Wir müssen nach wie vor im konkreten Fall Zähldichten oder W-Maße zuordnen und dabei (vielleicht nicht verifizierbare) Annahmen machen. Aber wir werden später mit Hilfe unseres Axiomensystems zeigen können, daß der Grenzwert 1.3 trotz aller Bedenken sinnvoll ist.

2 Gleichverteilungen

Gleichverteilungen sind die einfachsten W-Räume. Beschäftigt man sich in der Schule mit Wahrscheinlichkeitstheorie, stehen in den unteren Klassen Gleichverteilungen im Vordergrund.

2.1 Definition: Ein endlicher W-Raum (Ω, P) heißt *Gleichverteilung* oder *Laplace-Raum*, wenn seine Zähldichte

$$f : \Omega \longrightarrow \mathbb{R}$$

konstant ist, d.h. es gibt ein $r \in \mathbb{R}$, so daß $f(\omega) = r$ für alle $\omega \in \Omega$.

Aus (1.9) folgt sofort

2.2 Für die Zähldichte f und das W-Maß P eines Laplace-Raumes gilt:

$$\begin{aligned} f(\omega) &= \frac{1}{|\Omega|} \quad \forall\, \omega \in \Omega \\ P(A) &= \frac{|A|}{|\Omega|} \quad \forall\, A \subset \Omega. \end{aligned}$$

Dabei bezeichnet $|A|$ die Anzahl der Elemente von A.

Bei einer Gleichverteilung besitzen die Elementarereignisse alle die gleiche Wahrscheinlichkeit. Die einfachsten Beispiele sind das Werfen eines idealen Würfels oder einer idealen Münze:

2.3 Beispiel: Ein idealer Würfel wird geworfen. Die Elementarereignisse sind die auftretenden Augenzahlen. Also

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Die Zähldichte ist

$$f(x) = \frac{1}{6} \quad \forall\, x \in \Omega.$$

Man kann nun nach den Ereignissen $G =$ „Augenzahl ist gerade“ oder $A =$ „Augenzahl ist prim“ oder $K_n =$ „Augenzahl ist $\leq n$“ fragen, also nach $G = \{2, 4, 6\},\ A = \{2, 3, 5\},\ K_5 = \{1, 2, 3, 4, 5\}$ usw.

Es folgt

$$P(G) = \frac{1}{2}, \quad P(A) = \frac{1}{2}, \quad P(K_n) = \frac{n}{6} \text{ für } 1 \leq n \leq 6.$$

2.4 Beispiel: Werfen einer idealen Münze. Die Elementarereignisse sind Z = „Zahl", W = „Wappen". Also

$$\Omega = \{Z, W\}.$$

Die Zähldichte ist

$$f(x) = \frac{1}{2} \quad \forall\, x \in \Omega.$$

Die damit verbundenen Fragestellungen sind meist von etwas künstlicher Natur. Interessanter und amüsanter sind Probleme folgender Art

2.5 Probleme:

(1) Wie groß ist die Wahrscheinlichkeit dafür, beim Skatspiel 3 Buben, darunter den Kreuzbuben zu erhalten?

(2) Wie groß ist die Wahrscheinlichkeit dafür, im Lotto den Hauptgewinn oder überhaupt nur einen Gewinn zu erzielen?

(3) **Geburtstagsproblem:** Wie groß ist die Wahrscheinlichkeit dafür, daß auf einer Party mit n Personen mindestens zwei der Personen am selben Tag Geburtstag haben?

Zur Lösung dieser Probleme benötigen wir kombinatorische Formeln, also mathematische Hilfsmittel, die wir noch nicht zur Verfügung haben.

3 Kombinatorische Formeln

In diesem Abschnitt leiten wir einige kombinatorische Formeln her, die wir für die Behandlung der in (2.5) angesprochenen Probleme benötigen. Diese Formeln wurden bereits im Grundkurs behandelt. Deshalb verzichten wir auf den ausführlichen induktiven Beweis. Wir formulieren die Probleme mit Hilfe des „Urnenmodels".
Auf n verschiedene Urnen $U_1, \ldots, U_n$ sind k Kugeln zu verteilen.

3.1 In jede Urne dürfen beliebig viele Kugeln gelegt werden. Die Kugeln sind unterscheidbar, etwa durch Nummern.

$$\text{Anzahl der möglichen Kombinationen: } n^k.$$

Denn jede der k Kugeln kann in jede der n Urnen gelegt werden.

3.2 In jede Urne darf höchstens eine Kugel gelegt werden, die Kugeln sind unterscheidbar; $n \geq k$

Anzahl der möglichen Kombinationen:

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-(k-1)) = \frac{n!}{(n-k)!}.$$

Im Sonderfall $n = k$ also $n!$

Denn für die erste Kugel haben wir alle n Urnen frei, für die zweite $(n-1)$, für die dritte $(n-2)$ usw.

3.3 In jede Urne darf höchstens eine Kugel gelegt werden. Die Kugeln sind **nicht unterscheidbar**.

Die Anzahl der Kombinationsmöglichkeiten ist

$$\binom{n}{k} = \frac{n(n-1) \cdot \ldots \cdot (n-k-1)}{k!} = \frac{n!}{k!(n-k)!}.$$

Beweis: Wenn die Kugeln unterscheidbar wären, hätten wir

$$\frac{n!}{(n-k)!}$$

Kombinationen. Da sie aber nicht unterscheidbar sind, fallen einige dieser Kombinationen zusammen, und zwar genau so viele, wie wir die Nummern $1, \ldots k$ auf die k Kugeln verteilen können. Nach (3.2) gibt es dafür $k!$ Möglichkeiten, so daß wir insgesamt

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

Möglichkeiten haben. □

Wir verallgemeinern das vorausgegangene Problem

3.4 In jede Urne darf höchstens eine Kugel gelegt werden. Es gibt r verschiedene Typen $T_1, \ldots, T_r$ von Kugeln (etwa verschiedene Farben), und zwar k_1 Kugeln vom Typ T_1, k_2 Kugeln vom Typ T_2 usw., so daß $k = k_1 + \cdots + k_r$.
Zeigen Sie: Die Anzahl der verschiedenen Kombinationen ist

$$\frac{n!}{k_1!k_2! \ldots k_r!(n-k)!}.$$

Beweis: Wenn alle Kugeln unterscheidbar wären, hätten wir

$$\frac{n!}{(n-k)!}$$

Kombinationen. Von diesen Kombinationen fallen einige zusammen: Wir können alle Kugeln vom Typ T_i untereinander austauschen, ohne die Kombination zu ändern. Numerieren wir die Kugeln vom Typ T_i mit $1, 2, \ldots, k_i$, gibt es $k_1!$ verschiedene Möglichkeiten der Anordnung der numerierten Kugeln auf die vorgegebenen Plätze. Da das für alle Typen zutrifft, erhalten wir die Formel. $\square$

Da offensichtlich $n!$ und die Binomialkoeffizienten $\binom{n}{k}$ eine große Rolle spielen, wiederholen wir ihre Definition.

3.5 Definition: Für $k, n \in \mathbb{N} : \{0, 1, 2, \ldots\}$ definieren wir

$$n! = \begin{cases} 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n & \text{für } n > 0 \\ 1 & \text{für } n = 0 \end{cases}$$

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{für } n \geq k \\ 0 & \text{für } n < k \end{cases}$$

3.6 Aufgaben: (1) $\binom{n}{k} = \binom{n}{n-k}$

(2) $\binom{n}{0} = \binom{n}{n} = 1$

(3) $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$

3.7 In jede Urne dürfen beliebig viele Kugeln gelegt werden. Die Kugeln sind nicht unterscheidbar.

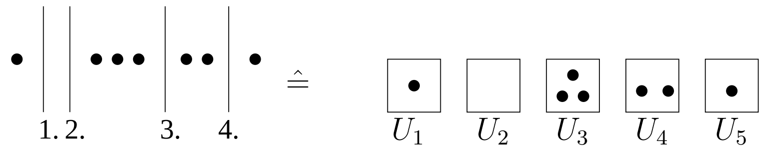
$$\text{Anzahl der möglichen Kombinationen } = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}.$$

Beweis: Wir denken uns $n + k - 1$ Positionen auf die wir $n - 1$ Leisten verteilen können. Auf die übrigen k Positionen legen wir die Kugeln. Nach (3.3) gibt es

$$\binom{n+k-1}{n-1} \stackrel{1.5}{=} \binom{n+k-1}{n+k-1-(n-1)} = \binom{n+k-1}{k}$$

verschiedene Möglichkeiten der Leistenverteilung. Die Kugeln von der ersten Leiste kommen in die erste Urne, die vor der zweiten in die zweite usw. und die nach der $(n-1)$-ten Leiste in die n-te Urne. So bestimmt jede Leistenverteilung eine der gesuchten Kombinationen und umgekehrt. $\square$

Beispiel: $n = 5, k = 7$



Anwendungsbeispiele:

3.8 Wie viele verschiedene Autonummern sind maximal im Landkreis Osnabrück möglich, wenn jede Nummer aus 2 Buchstaben und 3 Ziffern besteht und alle 26 Buchstaben und 10 Ziffern ausgenutzt werden?

Antwort: Wir haben $26^2 = 676$ Buchstabenkombinationen und $10^3 = 1000$ Ziffernkombinationen. Also gibt es maximal 676000 verschiedene Autonummern.

3.9 6 Paare feiern eine Tanzparty. Wie viele verschiedene Paarkombinationen auf der Tanzfläche sind möglich, wenn immer alle Teilnehmer tanzen?

Antwort: Die 6 Männer müssen auf die 6 Damen verteilt werden. Nach (3.2) gibt es dafür

$$6! = 720$$

verschiedene Möglichkeiten.

3.10 Wieviel verschiedene k-elementige Teilmengen hat eine Menge von n Elementen?

Antwort: Fasse die Elemente der Menge als n Urnen auf, auf die wir k Kugeln so verteilen, daß in jede Urne höchstens eine gelegt wird. Die k vollen Urnen bestimmen eine k-elementige Teilmenge. Nach (3.3) gibt es somit

$$\binom{n}{k}$$

verschiedene k-elementige Teilmengen.

3.11 Wie viele verschiedene Blätter kann man beim Skatspiel erhalten?

Antwort: Ein Skatspiel hat 32 Karten. Ein Blatt ist eine Kombination von 10 dieser Karten. Gefragt ist also nach der Anzahl der 10-elementigen Teilmengen einer 32-elementigen Menge. Nach (3.10) ist diese

$$\binom{32}{10} = 64.512.240.$$

3.12 Aufgaben: (1) Beim Spiel „Master Mind“ muß man eine Kombination farbiger Stifte auf 5 Plätzen $P_1, \dots, P_5$ erraten. Es stehen Stifte in 8 verschiedenen Farben zur Verfügung (von jeder Farbe mindestens 5). Wie viele verschiedene Kombinationen gibt es?

(2) Wie viele verschiedene Lotto-Tips gibt es im Spiel “6 aus 49”?

(3) Wie viele verschiedene Teilmengen hat eine n-elementige Menge? Folgern Sie aus der Antwort und aus (3.10), daß

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

(4) Beweisen Sie den Polynomialsatz

$$(x_1 + \dots + x_r)^n = \sum \frac{n!}{k_1!k_2! \dots k_r!} x_1^{k_1} x_2^{k_2} \dots x_r^{k_r}.$$

Die Summe wird gebildet über alle $(k_1, k_2, \dots, k_r)$ mit $0 \leq k_i \leq n$ und $k_1 + k_2 + \dots + k_r = n$.

4 Beispiele

In diesem Abschnitt wollen wir auf die Probleme 2.5 eingehen.

4.1 Frage: Wie groß ist die Wahrscheinlichkeit dafür

(1) schon beim Austeilen 3 Buben, darunter den Kreuzbuben zu erhalten?

(2) mindestens einen Buben im Skat zu finden? Dabei gehen wir davon aus, daß wir unser eigenes Blatt nicht kennen.

Antwort (1): Man erhält 10 Karten aus 32. Sei Ω die Menge der verschiedenen Skatblätter. Wir nehmen an, daß jedes Blatt gleichwahrscheinlich ist, d.h. daß die Karten ideal sind. Nach (3.11) ist

$$|\Omega| = \binom{32}{10}.$$

Sei $A \subset \Omega$ die Menge der Blätter, die 3 Buben, darunter den Kreuzbuben enthalten. Wir müssen $|A|$ bestimmen.

Es gibt nur 3 verschiedene Bubenkombinationen: Es fehlt Pik- oder Herz- oder Karobube. Die restlichen 7 Karten sind aus der Menge der 28 Nichtbuben ausgewählt. Es folgt

$$|A| = 3 \cdot \binom{28}{7}.$$

Die gesuchte Wahrscheinlichkeit ist

$$\frac{|A|}{|\Omega|} = \frac{3 \cdot \binom{28}{7}}{\binom{32}{10}} = \frac{3 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 24 \cdot 23 \cdot 22 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot \ldots \cdot 10}{32 \cdot 31 \cdot 30 \cdot 29 \cdot 28 \cdot 27 \cdot 26 \cdot 25 \cdot 24 \cdot 23 \cdot 2 \cdot 3 \cdot \ldots \cdot 7}$$
$$= \frac{\cancel{3} \cdot 22 \cdot \cancel{8} \cdot 9\ \cancel{10}}{4\ \cancel{32} \cdot 31 \cdot \cancel{30} \cdot 29} = \frac{99}{2 \cdot 31 \cdot 29} = \frac{99}{1798} = 0,055\ldots$$

Antwort (2): Der Skat besteht aus 2 Karten. Wir haben also $\binom{32}{2}$ verschiedene Möglichkeiten für den Skat. Sei Ω die Menge dieser Blätter, $A_1 \subset \Omega$ die Teilmenge mit einem Buben, A_2 die Teilmenge mit 2 Buben. Dann ist $A_1 \cap A_2 = \emptyset$. Gesucht wird $|A_1|$ und $|A_2|$.

Bestimmung von A_1: Wir haben 4 Möglichkeiten für den einen Buben. Die 2. Karte ist aus den 28 Nicht-Buben. Also

$$A_1 = 4 \cdot 28.$$

Bestimmung von A_2: Wir wählen 2 aus vier Buben. Also

$$A_2 = \binom{4}{2} = 6.$$

Die gesuchte Wahrscheinlichkeit ist

$$\frac{|A_1| \cup |A_2|}{|\Omega|} = \frac{|A_1| + |A_2|}{|\Omega|} = \frac{(4 \cdot 28 + 6)}{\binom{32}{2}} = \frac{(4 \cdot 28 + 6) \cdot 2}{32 \cdot 31}$$
$$= \frac{4 \cdot 59}{32 \cdot 31} = \frac{59}{248} = 0,2379\ldots$$

4.2 Frage: Wie groß ist die Wahrscheinlichkeit dafür, beim Lotto “6 aus 49”

(1) sechs richtige zu haben?

(2) 5 richtige mit Zusatzzahl zu haben?

(3) überhaupt einen Gewinn zu erzielen?

Sie M die Menge der 49 Ziffern. Wir gehen von einem idealen Lottoziehgerät aus, d.h. alle Zahlen werden mit derselben Wahrscheinlichkeit gezogen. Sei Ω die Menge der möglichen Tips, d.h. Ω ist die Menge der 6-elementigen Teilmengen von M. Sei $X \in \Omega$ der richtige Tip. Sei $A_k \subset \Omega, 0 \leq k \leq 6$, die Menge aller Tips mit genau k richtigen. Also

$$A_k = \{T \in \Omega; |T \cap X| = k\}$$

(beachte, daß T und X Teilmengen von M sind.)

Antwort 1: Da alle Tips gleichwahrscheinlich sind, ist die gesuchte Wahrscheinlichkeit

$$\frac{1}{|\Omega|} = \frac{1}{\binom{49}{6}} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{1}{13.983.816}.$$

Antwort 2: Der Tip enthält die Zusatzzahl. Die übrigen 5 Zahlen müssen aus den 6 richtigen kommen. Wir haben dafür $6 = \binom{6}{5}$ Möglichkeiten. Die gesuchte Wahrscheinlichkeit ist

$$\frac{6}{13.983.816}.$$

Antwort 3: Man hat gewonnen, wenn man mindestens 3 richtige Zahlen getippt hat. Gesucht wird also

$$P = \frac{|A_3 \cup A_4 \cup A_5 \cup A_6|}{|\Omega|} = \frac{|A_3| + |A_4| + |A_5| + |A_6|}{|\Omega|}.$$

Berechnung von $|A_k|$: Sei T ein Tip in A_k. Dann enthält T k Zahlen aus $X \subset M$ und $6-k$ Zahlen aus $Z = M \backslash X$. Für die k richtigen Zahlen haben wir $\binom{6}{k}$ Möglichkeiten, für die restlichen falschen $\binom{43}{6-k}$ Möglichkeiten, denn $|Z| = 43$. Es folgt

$$|A_k| = \binom{6}{k} \cdot \binom{43}{6-k}.$$

Unsere gesuchte Wahrscheinlichkeit ist somit

$$P = \frac{\binom{6}{3}\binom{43}{3} + \binom{6}{4}\binom{43}{2} + \binom{6}{5}\binom{43}{1} + \binom{6}{6}\binom{43}{0}}{\binom{49}{6}}$$

$$= \frac{\frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} \cdot \frac{43 \cdot 42 \cdot 41}{1 \cdot 2 \cdot 3} + \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{43 \cdot 42}{1 \cdot 2} + 6 \cdot 43 + 1}{\frac{49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}}.$$

Hierbei haben wir die Formel $\binom{n}{n-k} = \binom{n}{k}$ benutzt. Es folgt

$$\begin{aligned} P &= \frac{5 \cdot 4 \cdot 43 \cdot 7 \cdot 41 + 3 \cdot 5 \cdot 43 \cdot 21 + 6 \cdot 43 + 1}{49 \cdot 47 \cdot 46 \cdot 3 \cdot 44} \\ &= \frac{246820 + 13545 + 259}{13983816} = \frac{260624}{13983816} = \frac{32578}{1747977} \\ &= 0,018637\ldots \end{aligned}$$

Die Wahrscheinlichkeit, beim Lottospiel “6 aus 49” überhaupt einen Gewinn zu erzielen, ist kleiner als 1,9 Prozent.

4.3 Geburtstagsproblem: Wir machen die nicht ganz realistische Annahme, daß alle Tage als Geburtstage gleich wahrscheinlich sind. Weiter gehen wir von 365 Tagen im Jahr aus.

Frage: n Personen besuchen eine Party. Wie groß ist die Wahrscheinlichkeit dafür, daß mindestens zwei Personen am selben Tag Geburtstag haben?

Sei A_n dieses Ereignis. Ω ist die Menge aller n-Tupel von Zahlen $x_i,\ 1 \le i \le 365$. Es folgt

$$A_n = \{(x_1, \ldots, x_n) \in \Omega;\ \text{mindestens zwei der } x_i \text{ sind gleich}\}.$$

Für $n > 365$ ist $A_n = \Omega$, also

$$P(A_n) = 1.$$

Für $n \le 365$ berechnen wir $P(\mathcal{C}A_n) : x = (x_1, \ldots, x_n)$ ist genau dann in $\mathcal{C}A_n$, wenn alle x_i verschieden sind. Wir benutzen nun das Urnenmodel: Tage $\widehat{=}$ Urnen, Personen $\widehat{=}$ Kugeln.

Für x_1 haben wir 365 Möglichkeiten, für x_2 noch 364 usw. Also

$$|\mathcal{C}A_n| = 365 \cdot 364 \cdot \ldots (365 - n + 1) = \frac{365!}{(365-n)!}.$$

Da $|\Omega| = 365^n$, folgt

$$P(A_n) = 1 - P(\mathcal{C}A_n) = 1 - \frac{365!}{(365-n)! \cdot 365^n} \quad \text{für } n \le 365.$$

Auf der nachfolgenden Tabelle haben wir eine Liste der Wahrscheinlichkeiten für $1 \le n \le 113$.

Die Werte sind bis auf einen Fehler $< 10^{-8}$ berechnet.

Tabelle Wahrscheinlichkeiten beim Geburtstagproblem

n	$P(A_n)$	n	$P(A_n)$	n	$P(A_n)$
1	0.00000000	41	0.90315161	81	0.99993311
2	0.00273973	42	0.91403047	82	0.99994795
3	0.00820417	43	0.92392286	83	0.99995965
4	0.01635591	44	0.93288537	84	0.99996882
5	0.02713557	45	0.94097590	85	0.99997600
6	0.04046248	46	0.94825284	86	0.99998159
7	0.05623570	47	0.95477440	87	0.99998593
8	0.07433529	48	0.96059797	88	0.99998928
9	0.09462383	49	0.96577961	89	0.99999186
10	0.11694818	50	0.97037358	90	0.99999385
11	0.14114138	51	0.97443199	91	0.99999537
12	0.16702479	52	0.97800451	92	0.99999652
13	0.19441028	53	0.98113811	93	0.99999740
14	0.22310251	54	0.98387696	94	0.99999806
15	0.25290132	55	0.98626229	95	0.99999856
16	0.28360401	56	0.98833235	96	0.99999893
17	0.31500767	57	0.99012246	97	0.99999922
18	0.34691142	58	0.99166498	98	0.99999942
19	0.37911853	59	0.99298945	99	0.99999958
20	0.41143838	60	0.99412266	100	0.99999969
21	0.44368834	61	0.99508880	101	0.99999978
22	0.47569531	62	0.99590957	102	0.99999984
23	0.50729723	63	0.99660439	103	0.99999988
24	0.53834426	64	0.99719048	104	0.99999992
25	0.56869970	65	0.99768311	105	0.99999994
26	0.59824082	66	0.99809570	106	0.99999996
27	0.62685928	67	0.99844004	107	0.99999997
28	0.65446147	68	0.99872639	108	0.99999998
29	0.68096854	69	0.99896367	109	0.99999998

30	0.70631624	70	0.99915958	110	0.99999999
31	0.73045463	71	0.99932075	111	0.99999999
32	0.75334753	72	0.99945288	112	0.99999999
33	0.77497185	73	0.99956081	113	1.00000000
34	0.79531686	74	0.99964864		
35	0.81438324	75	0.99971988		
36	0.83218211	76	0.99977744		
37	0.84873401	77	0.99982378		
38	0.86406782	78	0.99986095		
39	0.87821966	79	0.99989067		
40	0.89123181	80	0.99991433		

5 Klassische Wahrscheinlichkeitsverteilungen

5.1 Beispiel: In einer Urne liegen r rote und s schwarze Kugeln. Wir ziehen n Kugeln, $0 \leq n \leq r+s$. Wie groß ist die Wahrscheinlichkeit dafür, daß wir genau k rote in dieser Kollektion von n Kugeln haben?

Antwort: Dieses Problem ist eine einfache Verallgemeinerung von Beispiel 4.2. Wir behandeln es analog:
Sei R die Menge der roten Kugeln, S die Menge der schwarzen Kugeln und $M = R \cup S$. Der Elementarereignisraum ist die Menge der Kollektionen von n Kugeln, d.h. die n-elementigen Teilmengen von M:

$$\Omega = \{Y \subset M;\ |Y| = n\}.$$

Aufgrund der Aufgabenstellung ist uns auf Ω eine Gleichverteilung gegeben

$$f : \Omega \longrightarrow \mathbb{R}, \qquad f(A) = \frac{1}{|\Omega|} \quad \forall\ A \in \Omega.$$

Sei nun analog zu (4.2) A_k die Menge der Kugelkollektionen mit genau k roten Kugeln, d.h.

$$A_k = \{Y \in \Omega;\ |Y \cap R| = k\}.$$

Die Wahrscheinlichkeit für einen Zug von genau k roten Kugeln unter n Kugeln ist somit

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

Die Berechnung von $|A_k|$ verläuft wie unter (4.2):

Es gibt $\binom{r}{k}$ verschiedene Möglichkeiten, k Elemente aus R auszuwählen, nämlich die Anzahl der k-elementigen Teilmengen von R. Diese Kollektion von k roten Kugeln müssen wir durch eine $(n-k)$-elementige Teilmenge aus S zu einer Kollektion von n Kugeln mit genau k roten Kugeln ergänzen. Es gibt $\binom{s}{n-k}$ verschiedene solche Teilmengen. Also besitzt A_k genau

$$\binom{r}{k} \cdot \binom{s}{n-k}$$

Elemente. Die Elemente von Ω sind die n-elementigen Teilmengen von M. Davon gibt es

$$\binom{r+s}{n}$$

verschiedene. Wir erhalten also

$$\boxed{P(A_k) = \frac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{r+s}{n}}}$$

Die Lösung des Beispiels 5.1 können wir in der Form eines W-Raumes angeben: Der Ereignisraum ist in diesem Fall

$$H = \{0, 1, \ldots, n\}.$$

Das Ereignis „k“ bedeutet: Bei der Ziehung von n Kugeln werden genau k rote Kugeln gezogen. Demzufolge ist die Zähldichte

$$g : H \longrightarrow \mathbb{R}, \quad k \longmapsto \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} =: h(k; n, r, s).$$

5.2 Satz und Definition: (H, g) ist ein W-Raum. Die Zähldichte g nennt man *hypergeometrische Verteilung*.

Der Beweis dieses Satzes ist ein Spezialfall von folgendem Satz, dessen Beweis wir als Aufgabe stellen.

5.3 Satz: Sei (Ω, f) ein endlicher W-Raum und P das zugehörige W-Maß. Seien $A_0, \ldots, A_n \subset \Omega$ paarweise disjunkte Teilmengen, so daß

$$\Omega = A_0 + A_1 + \cdots + A_n.$$

Dann ist (H, g) mit $H = \{0, 1, \ldots, n\}$ und der Zähldichte

$$g : H \longrightarrow \mathbb{R}, \quad g(k) = P(A_k)$$

ein endlicher W-Raum.

5.4 Aufgabe: Zeigen Sie: Für $r, s \in \mathbb{N}$ und $0 \leq n \leq r + s$ gilt

$$\sum_{k=0}^{n} \binom{s}{k} \cdot \binom{s}{n-k} = \binom{r+s}{n}$$

(Wenn Sie unsere Ergebnisse konsequent nutzen, erübrigt sich jede Rechnung.)

Das Zufallsexperiment 5.1 kann ich auch wie folgt formulieren. Man zieht nacheinander n Kugeln, notiert, ob sie rot oder schwarz sind, legt die Kugeln nach jeder Ziehung *nicht* zurück. Die Wahrscheinlichkeit dafür, daß die k-te Kugel rot ist, hängt natürlich davon ab, welche Kugeln in den vorausgegangenen Zügen gezogen wurden. D.h. die n Einzelexperimente sind voneinander *abhängig*.

Wir können nun das gleiche Experiment machen, legen aber nach jedem Einzelzug die Kugel wieder zurück. D.h. die Wahrscheinlichkeit dafür, daß beim k-ten Zug eine rote Kugel erscheint, ist (im Idealfall) dieselbe wie beim 1. Zug. Experimente dieser Art nennt man nach dem schweizer Mathematiker Jakob Bernoulli (1654–1705) Bernoulli-Experiment.

5.5 Definition: Ein Zufallsexperiment, bei dem das Ereignis A eintreten kann, wird n-mal wiederholt. Sei A_i das Ereignis, daß A bei i-ten Versuch eintritt. Wir nennen die Versuchsreihe ein *Bernoulli-Experiment* für A vom *Umfang* n, wenn gilt

(1) $P(A_1) = P(A_2) = \ldots = P(A_n) = p$

(2) die Einzelexperimente sind voneinander unabhängig.

Auf den Begriff der Unabhängigkeit werden wir später noch formal eingehen. Es ist aber schon ohne eine formal-mathematische Fassung intuitiv klar, was damit gemeint ist.

Wie im Fall der hypergeometrischen Verteilung interessieren wir uns jetzt für die Wahrscheinlichkeit dafür, daß bei einem Bernoulli-Experiment für A vom Umfang n das Ereignis A genau k-mal eintritt.

5.6 Bernoulli- oder Binomialverteilung: Wir machen ein Bernoulli-Experiment vom Umfang n für ein Ereignis A eines Zufallsexperimentes. Sei $P(A) = p$. Gesucht wird die Wahrscheinlichkeit

$$b(k;\, n, p)$$

dafür, daß A genau k-mal in der Versuchsreihe vorkommt. Sei $q = P(\mathcal{C}A) = 1 - p$. Da uns nur interessiert, ob A eintritt oder nicht, schreiben wir T für Treffer, falls A auftritt, und sonst N für Niete.

Die Elementarereignisse des Bernoulli-Experiments sind also n-Tupel von Symbolen T, N. Der Elementarereignisraum Ω hat 2^n Elemente. Die Wahrscheinlichkeit dafür, daß im i-ten Versuch T auftritt, ist p, und dafür, daß im i-ten Versuch N auftritt, ist q. Also ist die Wahrscheinlichkeit für ein fest gegebenes n-Tupel $x \in \Omega$

$$p^k \cdot q^{n-k}, \quad \text{falls } x \text{ genau } k \text{ Treffer enthält.}$$

Jede k-elementige Teilmenge S von $\{1, \ldots, n\}$ bestimmt ein solches x: man setzt Treffer an den Koordinaten ein, die in S liegen. Da es $\binom{n}{k}$ solche S gibt, erhalten wir

$$\boxed{b(k; n, p) = \binom{n}{k} p^k \cdot q^{n-k}}$$

Auch dieses Ergebnis können wir zu einem neuen W-Raum machen: Uns interessiert, ob wir bei einem Bernoulli-Experiment vom Umfang n genau k Treffer haben. Der Elementarereignisraum ist also wieder

$$B = \{0, 1, 2, \ldots n\}$$

und die Wahrscheinlichkeit für genau k Treffer, d.h. für das Elementarereignis k ist $b(k; n, p)$.

Aus Satz 5.3 folgt

5.7 Satz und Definition: $B = \{0, 1, \ldots, n\}$ mit der Zähldichte

$$k \longmapsto b(k;\, n, p),$$

wobei $0 \leq p \leq 1$ fest gegeben ist, ist ein endlicher W-Raum. Die Zähldichte nennt man *Bernoulli-* oder *Binomialverteilung*.

5.8 Beispiel: Beim „Mensch ärgere dich nicht“ braucht man bei dreimaligem Würfeln mindestens eine 6, um herauszukommen. Wie groß ist die Wahrscheinlichkeit dafür, in der ersten Runde herauszukommen?

Antwort: Sei A das Ereignis: Beim Würfeln tritt keine 6 auf. Dann ist $p(A) = \frac{5}{6}$. Daß in allen drei Würfen keine 6 auftritt ist (A tritt dreimal auf)

$$b(3; 3, \frac{5}{6}) = \binom{3}{3} \left(\frac{5}{6}\right)^3 \cdot \left(\frac{1}{6}\right)^0 = \left(\frac{5}{6}\right)^3 = \frac{125}{216}.$$

Die gesuchte Wahrscheinlichkeit ist somit

$$1 - \frac{125}{216} = \frac{91}{216} = 0,42\ldots$$

5.9 Polynomialverteilung: Bei einem Bernoulli-Experiment interessiert uns nur, ob bei einem Einzelversuch ein Treffer oder eine Niete eintritt. Es könnte aber sein, daß bei einem Einzelversuch mehrere uns interessierende Ausgänge möglich sind.

Sei also (Ω, P) ein endlicher W-Raum und $A_1, \ldots, A_r$ paarweise unverträgliche Ereignisse, so daß

$$\Omega = A_1 + \cdots + A_r.$$

Damit tritt bei jedem Versuch genau eines der Ereignisse $A_1, \ldots, A_r$ ein. Sei $p_i = P(A_i)$. Wie beim Bernoulli-Experiment wiederholen wir den Einzelversuch n-mal und gehen davon aus, daß die Einzelexperimente voneinander unabhängig sind. Gesucht wird die Wahrscheinlichkeit

$$p(k_1, \ldots, k_r;\, n, p_1, \ldots, p_r) \qquad k_1 + \cdots + k_r = n$$

dafür, daß bei einer Versuchsreihe vom Umfang n die Ereignisse $A_1, \ldots, A_r$ genau $k_1, \ldots, k_r$-mal auftreten.

Der zugehörige Elementarereignisraum $\hat{\Omega}$ besteht aus allen n-Tupeln in den Symbolen $A_1, \ldots, A_r$. Die i-te Koordinate eines Tupels ist A_k, falls beim i-ten Versuch das Ereignis A_k eintritt.

Sei nun $S = S(k_1, \ldots, k_r) \subset \hat{\Omega}$ die Menge der Tupel, in denen A_i genau k_i-mal auftritt, $1 \leq i \leq r$. Gesucht wird die Wahrscheinlichkeit für S. Sei

$$x = \underbrace{A_1, \ldots, A_1}_{k_1}, \underbrace{A_2, \ldots, A_2}_{k_2}, \ldots, \underbrace{A_r, \ldots, A_r}_{k_r}$$

$P(x) = p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_r^{k_r}$. Jedes andere $y \in S$ erhält man aus x durch Permutation der Koordinaten. Folglich ist $P(y) = P(x)$, denn die Einzelexperimente sind voneinander unabhängig. Es bleibt also nur noch $|S|$ zu bestimmen. Wir müssen k_1 Kopien von A_1, k_2 Kopien von A_2 usw. auf n „Urnen“ verteilen. Nach (3.4) haben wir dafür

$$\frac{n!}{k_1! k_2! \ldots k_r!}$$

Möglichkeiten, so daß

$$\boxed{p(k_1,\dots,k_r;n,p_1,\dots,p_k)=\frac{n!}{k_1!k_2!\dots k_r!}p_1^{k_1}\dots p_r^{k_r}}$$

Wieder erhalten wir einen neuen W-Raum.

5.10 Sei Pol_n die Menge der r-Tupel $(k_1,\dots,k_r)$ mit $k_i\in\mathbb{N}$ und $k_1+\dots+k_n=n$. Seien $p_1,\dots,p_r\in\mathbb{R}_+$ mit $p_1+\dots+p_r=1$. Denn definiert

$$(k_1,\dots,k_r)\longmapsto p(k_1,\dots,k_r;n,p_1,\dots,p_r)$$

eine Zähldichte auf Pol_n. Das zugehörige W-Maß heißt *Polynomialverteilung.*

5.11 Beispiel: Ein idealer Würfel wird 6-mal geworfen. Wie groß ist die Wahrscheinlichkeit dafür, daß 3-mal die sechs und 3-mal die fünf auftritt?

Antwort: Sei A_i das Ereignis „beim Würfeln tritt i auf“. Also $p_i=P(A_i)=\frac{1}{6}$.
Gesucht wird

$$p(0,0,0,0,3,3)=\frac{6!}{3!\cdot 3!}\left(\frac{1}{6}\right)^3\cdot\left(\frac{1}{6}\right)^3=\frac{20}{6^6}=\frac{20}{46.656}=0,000428\dots$$

5.12 Geometrische Verteilung: Wir betrachten folgende Variante eines Bernoulli-Experiments. Wieder interessieren wir uns nur für Treffer (mit Wahrscheinlichkeit p) und Niete (mit Wahrscheinlichkeit $q=1-p$). Wir brechen die Versuchsreihe ab, sobald wir einen Treffer erzielen.

Gesucht wird die Wahrscheinlichkeit

$$g(k;p)$$

dafür, daß die Versuchsreihe nach dem k-ten Versuch abbricht.
Für $p=1$ tritt bereits beim ersten Versuch ein Treffer auf, so daß $g(1;1)=1$ und $g(k;1)=0$ für $k>1$. Sei nun $p<1$.
Nun ist

$$g(k;p)=P\big((\overbrace{N,\dots,N}^{k-1},T)\big)=q^{k-1}\cdot p.$$

Also

$$\boxed{g(k;p))=p(1-p)^{k-1}\text{ für }p<1}\quad\text{und}\quad\boxed{g(k;1)=\begin{cases}1 & k=1\\ 0 & k>1\end{cases}}$$

Wieder können wir das als neuen W-Raum deklarieren. Allerdings können unsere Versuchsreihen beliebige Länge haben, so daß unser Elementarereignisraum die Menge $\mathbb{N}\backslash\{0\}$ ist.

5.13 Satz und Definition: Sei $0 \leq p \leq 1$. Dann definiert

$$k \longmapsto g(k;p)$$

eine „Zähldichte“ auf der Menge $\mathbb{N}\backslash\{0\}$. Das zugehörige W-Maß heißt *geometrische Verteilung.*

5.14 Warnung: $\mathbb{N}\backslash\{0\}$ ist nicht mehr endlich. Also müssen wir mit dem Begriff der Zähldichte vorsichtig sein. In diesem Fall macht es aber Sinn:

(1) $g(k;p) \geq 0$

(2) $\sum\limits_{k \in \mathbb{N}\backslash\{0\}} g(k;p) = \sum\limits_{k=1}^{\infty} p(1-p)^{k-1} = p \cdot \sum\limits_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1$

5.15 Beispiel: Ein(e) Student(in) feiert eine bestandene Klausur sehr heftig und kommt dilirisch nach Hause. Die Person hat 7 ähnliche Schlüssel, von denen einer in die Haustür paßt. Sie wählt einen Schlüssel aus und versucht aufzuschließen. Gelingt das nicht, legt sie den Schlüssel zurück, wählt wiederum wahllos und versucht es nochmals. Mit welcher Wahrscheinlichkeit erwischt sie genau beim 4. Versuch den richtigen Schlüssel? Mit welcher Wahrscheinlichkeit benötigt sie höchstens 4 Versuche?

Antwort 1: $p = \dfrac{1}{7}$. Gefragt ist $g(4;p) = \dfrac{1}{7} \cdot \left(\dfrac{6}{7}\right)^3 = \dfrac{216}{2401} = 0,0899\ldots$

Antwort 2: Gesucht wird

$$\begin{aligned}\sum_{k=1}^{4} g\left(k;\frac{1}{7}\right) &= p \cdot \sum_{k=1}^{4}\left(\frac{6}{7}\right)^{k-1} = \frac{1}{7} \cdot \sum_{k=0}^{3}\left(\frac{6}{7}\right)^{k} \\ &= \frac{1}{7} \cdot \frac{1-(\frac{6}{7})^4}{1-\frac{6}{7}} = 1 - \left(\frac{6}{7}\right)^4 = 1 - \frac{1296}{2401} = \frac{1105}{2401} = 0,460\ldots\end{aligned}$$

Die geometrische Verteilung zeigt, daß unser mathematisches Modell sehr schnell an seine Grenzen stößt. Es gibt eine Reihen von Versuchsanordnungen mit unendlich vielen Ausgängen. Deshalb ist es wünschenswert, unser Modell zumindest auf **abzählbare** Mengen von Elementarereignissen zu erweitern. Wir könnten unser Axiomensystem 1.6 wörtlich übernehmen. Das genügt aber nicht, um Satz 1.9 zu beweisen, denn jetzt können Teilmengen

unendlich viele Elemente haben. Alternativ könnten wir uns von der geometrischen Verteilung leiten lassen, für die wir noch immer eine Zähldichte haben:

5.16 Ansatz: Ein **diskreter W-Raum** soll durch eine endliche oder abzählbare unendliche Menge Ω zusammen mit einer Zähldichte

$$f : \Omega \longrightarrow \mathbb{R}$$

bestimmt sein, die den Bedingungen (1.9) genügt, d.h.

(1) $f(\omega) \geq 0 \quad \forall\, \omega \in \Omega$

(2) $\sum\limits_{\omega \in \Omega} f(\omega) = 1$

Übersetzen wir das wie im Beweis von (1.9) in ein Axiomensystem wie in (1.6) erhalten wir

5.17 Definition: Ein *diskreter W-Raum* ist ein Paar (Ω, P) bestehend aus einer endlichen oder abzählbar unendlichen Menge Ω von sog. *Elementarereignissen* und einer Abbildung

$$P_i \mathcal{P}(\Omega) \longrightarrow \mathbb{R},$$

genannt *W-Mass*, so daß gilt

(1) $0 \leq P(A) \leq 1 \quad \forall\, A \subset \Omega$

(2) $P(\Omega) = 1$

(3) Ist $\mathfrak{M}$ eine endliche oder abzählbar unendliche Menge von Teilmengen von Ω und gilt $A \cap B = \emptyset$ für alle $A, B \in \mathfrak{M}$ mit $A \neq B$, dann gilt

$$P\left(\bigcup_{A \in \mathfrak{M}} A\right) = \sum_{A \in \mathfrak{M}} P(A)$$

Der **einzige** Unterschied zum Axiomensystem (1.6) ist, daß in Axiom 3 auch **abzählbare** Vereinigungen paarweise disjunkter Teilmengen zugelassen sind. Diese Erweiterung des Axiomensystems erlaubt die wörtliche Übertragung des Beweises von Satz 1.9 auf den abzählbaren Fall.

Anhang: Reihen

In (5.16) und (5.17) treten „unendliche Summen“, also, präziser ausgedrückt, Reihen auf. Wir erinnern daran, daß der Grenzwert einer Reihe von der Reihenfolge der Summation abhängen kann. In (5.16) und (5.17) ist diese Reihenfolge nicht spezifiziert. Da es sich um Reihen mit positiven Summanden handelt, ist dies auch nicht nötig. In diesem Anhang erläutern wir diese Problematik und wiederholen Ergebnisse aus der Theorie der Reihen.

5.18 Definition: Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$. Die zugehörige *Reihe* Σa_n ist die Folge $(s_n)_{n\in\mathbb{N}}$ der *Partialsummen*

$$s_n = \Sigma_{k=0}^n a_k = a_0 + a_1 + \cdots + a_k.$$

Wir schreiben

$$\Sigma_{n=0}^\infty a_n \text{ für } \lim_{n\to\infty} s_n$$

und nennen diesen Grenzwert (der nicht zu existieren braucht) die Summe der Reihe Σa_n.

5.19 Beispiel: Sei $(a_n)_{n\geq 1}$ die Folge mit

$$a_{2n-1} = \frac{1}{n} \qquad a_{2n} = -\frac{1}{n} \qquad n \geq 1$$

also die Folge

$$1,\ -1,\ \frac{1}{2},\ -\frac{1}{2},\ \frac{1}{3},\ -\frac{1}{3},\ \ldots$$

Es folgt: $s_{2n-1} = \frac{1}{n},\ s_{2n} = 0$. Damit konvergiert die Reihe, und

$$\Sigma_{n=1}^\infty a_n = 0.$$

Wir ordnen die Reihe wie folgt um: Wir sortieren die positiven Terme in Blöcken

$$1,\ \left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{2^2}\right\},\ \left\{\frac{1}{2^2+1}, \ldots, \frac{1}{2^4}\right\},\ \left\{\frac{1}{2^4+1}, \ldots, \frac{1}{2^6}\right\}, \ldots$$

Hinter den k-ten Block ordnen wir die k-te negative Zahl ein.

Summieren wir die Blöcke, erhalten wir Werte ≥ 1, denn

$$\underbrace{\frac{1}{2^{2k}+1} + \cdots + \frac{1}{2^{2k+1}}}_{2^{2k}\text{Summanden}} + \underbrace{\frac{1}{2^{2k+1}+1} + \cdots + \frac{1}{2^{2(k+1)}}}_{2^{2k+1}\text{Summanden}}$$

$$\geq \frac{2^{2k}}{2^{2k+1}} + \frac{2^{2k+1}}{2^{2(k+1)}} = \frac{1}{2} + \frac{1}{2} = 1.$$

Betrachten wir die Partialsumme bis zur negativen Zahl $-\frac{1}{k}$, erhalten wir einen Wert

$$t_k \geq 1 - 1 + \left(1 - \frac{1}{2}\right) + \left(1 - \frac{1}{3}\right) + \cdots + \left(1 - \frac{1}{k}\right) \geq \frac{k}{2}.$$

Damit hat die zugehörige Reihe den Grenzwert $+\infty$.

Analog können wir den Grenzwert $-\infty$ durch geeignetes Umordnen erreichen. Mit etwas mehr Sorgfalt, kann man durch geeignetes Umordnen **jeden** vorgegebenen Grenzwert erzielen. □

Vor dem Hintergrund dieses Beispiels wird klar, daß (5.16) nicht unkommentiert bleiben kann: Die Reihe (5.16)(2) könnte von der Reihenfolge der $\omega \in \Omega$ abhängen, und wir haben keine vorgegebene Reihenfolge.

Wir erinnern:

5.20

$$\sum_{k=0}^{\infty} a_k = A \Longleftrightarrow \lim_{n\to\infty} s_n = A, \text{ wobei } s_n = \sum_{k=0}^{n} a_k$$

$$\Longleftrightarrow \forall\ s > 0\ \exists\ n_0, \text{so das } \left| A - \sum_{k=0}^{n} a_k \right| < s \quad \forall\ n > n_0.$$

5.21 Definition: Eine Reihe Σa_n heißt *absolut konvergent*, wenn $\Sigma |a_n|$ konvergiert.

5.22 Eine absolut konvergente Reihe ist konvergent (folgt aus dem Cauchy-Konvergenzkriterium).

Wichtig für uns ist der folgende Satz.

5.23 Satz: Sei Σa_n eine absolut konvergente Reihe mit Grenzwert A. Dann konvergiert auch jede Umordnung der Reihe gegen A.

Beweis: Sei $\tau : \mathbb{N} \to \mathbb{N}$ eine bijektive Abbildung. Wir müssen zeigen, daß

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_{\tau(k)} = A.$$

Sei $s > 0$ vorgegeben. Nach Voraussetzung existiert $B = \sum\limits_{k=0}^{\infty} |a_k| = \lim\limits_{n\to\infty} \sum\limits_{k=0}^{n} |a_k|$.

Also existiert n_0, so daß

$$\left| B - \sum_{k=0}^{n} |a_k| \right| = \sum_{k=n+1}^{\infty} |a_k| < \frac{s}{2} \quad \forall\, n \geq n_0.$$

Es folgt

$$\left| A - \sum_{k=0}^{n_0} a_k \right| = \left| \sum_{k=n_0+1}^{\infty} a_k \right| \leq \sum_{k=n_0+1}^{\infty} |a_k| < \frac{s}{2}.$$

Sei nun $N = \max\{\tau^{-1}(0), \dots, \tau^{-1}(n_0)\}$. Dann gilt

$$\{0, 1, 2, \dots, n_0\} \subset \{\tau(0), \tau(1), \dots, \tau(N)\}.$$

Für $m \geq N$ gilt dann

$$\left| A - \sum_{k=0}^{m} a_{\tau(k)} \right| \leq \underbrace{\left| \sum_{k=0}^{n_0} a_k - \sum_{k=0}^{m} a_{\tau(k)} \right|}_{\leq \sum_{k=n_0+1}^{\infty} |a_k| < \frac{s}{2}} + \underbrace{\left| A - \sum_{k=0}^{n_0} a_k \right|}_{< \frac{s}{2}} < s$$

$\square$

Es folgt, daß eine Reihe mit nur positiven Gliedern beliebig umgeordnet werden kann, ohne daß sich der Grenzwert ändert. Damit macht (5.16) (2) und (5.17) (2) Sinn.

6 Bedingtheit und Unabhängigkeit

Wir haben bei der Behandlung von Bernoulli-Experimenten bereits intuitiv vom Begriff von voneinander unabhängigen Experimenten Gebrauch gemacht. Wir wollen diesen Begriff axiomatisch behandeln, aber beginnen mit Beispielen, bei denen die Unabhängigkeit ausgeschlossen werden kann.

6.1 Beispiel: (Hypergeometrische Verteilung)
Wir haben eine Urne mit r roten und s schwarzen Kugeln. Sei $M = r + s$. Wir ziehen nacheinander zwei Kugeln ohne sie zurückzulegen. Wir fragen nach folgenden Ereignissen:

R: die erste Kugel ist rot
S: die erste Kugel ist schwarz
E: die zweite Kugel ist rot

Mit den Wahrscheinlichkeiten für solche Ereignisse haben wir uns schon beschäftigt (s. Beispiel 4.1(2)),

$$P(E) = \frac{r}{M},$$

denn die erste Kugel hat gegenüber der zweiten keine Besonderheiten. Nun ist die Wahrscheinlichkeit von E aber natürlich abhängig vom Ergebnis des ersten Zuges:

Wenn man **weiß**, daß im ersten Zug eine rote Kugel gezogen wird, ist die Wahrscheinlichkeit für eine rote Kugel im zweiten Zug nicht mehr $\frac{r}{M}$. Wir wollen das näher untersuchen: Ω ist die Menge der geordneten Paare von Kugeln aus der Urne, geordnet, weil wir zwischen erster und zweiter Kugel unterscheiden. Wir haben eine Gleichverteilung

$$|\Omega| = M \cdot (M-1)$$
$$|R| = r \cdot (M-1)$$
$$|S| = s \cdot (M-1)$$
$$|E| = (M-1) \cdot r. \qquad \text{Also } P(E) = \frac{(M-1)\cdot r}{M \cdot (M-1)} = \frac{r}{M}.$$

Wenn wir aber wissen, daß die erste Kugel rot ist, liegt bereits das Ereignis R vor. **Unter dieser Voraussetzung** ist die Wahrscheinlichkeit für das Eintreten von E dann

$$\frac{|R \cap E|}{|R|} = \frac{\frac{|R\cap E|}{|\Omega|}}{\frac{|R|}{|\Omega|}} = \frac{P(R \cap E)}{P(R)} = \frac{\frac{r(r-1)}{M(M-1)}}{\frac{r}{M}} = \frac{r-1}{M-1}.$$

6.2 Definition: Seien E und F Ereignisse in einem W-Raum (Ω, P), so daß $P(F) > 0$. Dann heißt

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

die *bedingte Wahrscheinlichkeit* von E unter (der Voraussetzung) F.

6.3 Beispiel: Drei Münzen werden nacheinander geworfen. E sei das Ereignis „wenigstens zweimal Kopf“ und F das Ereignis „die erste Münze zeigt Kopf“. Der Elementarereignisraum ist

$$\Omega = \{KKK, KKZ, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ\}.$$

Hier steht K für Kopf und Z für Zahl. Wir nehmen wieder eine Gleichverteilung an. E und F sind die Teilmengen

$$E = \{KKK, KKZ, KZK, ZKK\}$$
$$F = \{KKK, KKZ, KZK, KZZ\}.$$

Wir erhalten

$$P(E) = \frac{|E|}{|\Omega|} = \frac{1}{2} \qquad P(F) = \frac{|F|}{|\Omega|} = \frac{1}{2}$$
$$P(E \cap F) = \frac{|E \cap F|}{|\Omega|} = \frac{3}{8}.$$

Wir fragen nun nach der Wahrscheinlichkeit von E, falls wir bereits wissen, daß die erste Münze Kopf zeigt, also unter der Voraussetzung F

$$P(E|F) = \frac{\frac{3}{8}}{\frac{1}{2}} = \frac{3}{4}.$$

In vielen praktischen Situationen sind uns bedingte Wahrscheinlichkeiten gegeben oder lassen sich leicht berechnen. Die folgenden Sätze ermöglichen es uns, aus solchen Daten andere (bedingte) Wahrscheinlichkeiten zu ermitteln.

6.4 Multiplikationssatz: Sei (Ω, P) ein W-Raum.
Für beliebige Ereignisse $A_1, A_2, \dots, A_n \subset \Omega$, so daß $P(A_1 \cap \dots, \cap A_{n-1}) \neq 0$, gilt

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Beweis durch Induktion nach n: Da $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$ und

$$(A_1 \cap \dots \cap A_{n-1}) \subset (A_1 \cap \dots \cap A_{n-2}) \subset \dots \subset (A_1 \cap A_2) \subset A_1$$

folgt aus (1.8.3), daß

$$0 < P(A_1 \cap \dots \cap A_{n-1}) \leq P(A_1 \cap \dots \cap A_{n-2}) \leq \dots \leq P(A_1 \cap A_2) \leq P(A_1),$$

so daß die Faktoren der rechten Seite der Formel alle definiert sind. Für $n = 1$ ist die Formel offensichtlich richtig.
Induktionsschluß:
Voraussetzung: Die Formel gelte für n.
Behauptung: Sie gilt auch für $n + 1$.
Beweis: Nach (6.2) mit $E = A_{n+1}$ und $F = A_1 \cap A_2 \cap \dots \cap A_n$ gilt

$$P(A_{n+1}|A_1 \cap A_2 \cap \dots \cap A_n) = \frac{P(A_1 \cap A_2 \cap \dots \cap A_n \cap A_{n+1})}{P(A_1 \cap A_2 \cap \dots \cap A_n)}.$$

Es folgt

$$\begin{aligned}
&P(A_1 \cap A_2 \cap \ldots \cap A_n \cap A_{n+1}) \\
&\quad = P(A_1 \cap A_2 \cap \ldots \cap A_n) \cdot P(A_{n+1}|A_1 \cap \ldots \cap A_n) \\
&\quad = P(A_1) \cdot P(A_2|A_1) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1}) \cdot P(A_{n+1}|A_1 \cap \ldots \cap A_n),
\end{aligned}$$

letzteres nach Induktionsvoraussetzung. □

6.5 Beispiel: Eine Sendung von 60 Glühbirnen enthält 6 defekte. Man wählt willkürlich 5 Glühbirnen aus. Wie groß ist die Wahrscheinlichkeit, daß keine defekt ist.

Antwort: Sei A_k das Ereignis: In einer Kollektion von 5 Glühbirnen ist die k-te Birne nicht defekt, $1 \le k \le 5$. Gesucht ist

$$P(A_1 \cap A_2 \cap \ldots \cap A_5) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_5|A_1 \cap \ldots \cap A_4).$$

Offensichtlich ist $P(A_1) = \frac{54}{60} = \frac{9}{10}$. Die Wahrscheinlichkeit, daß auch die zweite nicht defekt ist, ist

$$P(A_2|A_1) = \frac{53}{59},$$

denn unter den verbliebenen 59 Birnen sind die 6 defekten. Genauso erhalten wir unter der Hypothese, daß die ersten beiden Birnen in Ordnung sind, die Wahrscheinlichekit

$$P(A_3|A_1 \cap A_2) = \frac{52}{58} = \frac{26}{29}.$$

Analog folgt

$$P(A_4|A_1 \cap A_2 \cap A_3) = \frac{51}{57} = \frac{17}{19} \text{ und } P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{50}{56} = \frac{25}{28}.$$

Aus (6.4) folgt somit

$$P(A_1 \cap A_2 \cap \ldots \cap A_5) = \frac{9}{10} \cdot \frac{53}{59} \cdot \frac{26}{29} \cdot \frac{17}{19} \cdot \frac{25}{28} = \frac{527085}{910252} = 0,579\ldots$$

Die gefragte Wahrscheinlichkeit ist also knapp 58 %.

Wir wollen jetzt eine Bezeichnung für eine Situation einführen, der wir in unseren Untersuchungen schon oft begegnet sind.

6.6 Definition: Eine *Partition* einer Menge Ω ist eine Familie $\mathfrak{M}$ von Teilmengen von Ω, so daß

(1) $M \cap N = \emptyset \quad \forall\, M \neq N$ in $\mathfrak{M}$

(2) $\cup\, \mathfrak{M} = \Omega$

In anderen Worten: Eine Partition von Ω ist eine Zerlegung von Ω in disjunkte Teilmengen.

6.7 Satz von der totalen Wahrscheinlichkeit: Sei (Ω, P) ein W-Raum mit W-Maß P und $\mathfrak{A}$ eine endliche oder abzählbare Partition von Ω mit $P(A) > 0 \quad \forall\, A \in \mathfrak{A}$. Dann gilt für jedes Ereignis $E \subset \Omega$:

$$P(E) = \sum_{A \in \mathfrak{A}} P(E|A) \cdot P(A).$$

Beweis: $E = \Omega \cap E = \left(\coprod_{A \in \mathfrak{A}} A \right) \cap E = \coprod_{A \in \mathfrak{A}} (A \cap E)$

Also
$$P(E) = \sum_{A \in \mathfrak{A}} P(A \cap E) \overset{6.2}{=} \sum_{A \in \mathfrak{A}} P(E|A) \cdot P(A)$$

□

Dieser einfache Satz findet insbesondere Anwendung in der Theorie der Markoff-Ketten oder der Zufallsprozesse. Wir wollen das an einem Beispiel erläutern, ohne weiter in die Theorie dieser Prozesse einzudringen.

6.8 Beispiel: Gegeben sind 3 Urnen A, B, C, die weiße und rote Kugeln nach folgender Tabelle enthalten:

Urne	A	B	C
weiße Kugeln	3	3	2
rote Kugeln	2	4	5

Wir wählen zufällig eine Urne U, entnehmen ihr eine Kugel und legen sie zufällig in eine der beiden anderen V. Dieser entnehmen wir wiederum eine Kugel.

Frage 1: Wie groß ist die Wahrscheinlichkeit dafür, daß beide Kugeln rot sind?

Zur Theorie: Hier haben wir es mit einer Folge von 4 Experimenten zu tun:

(1) Wir wählen eine von drei Urnen.

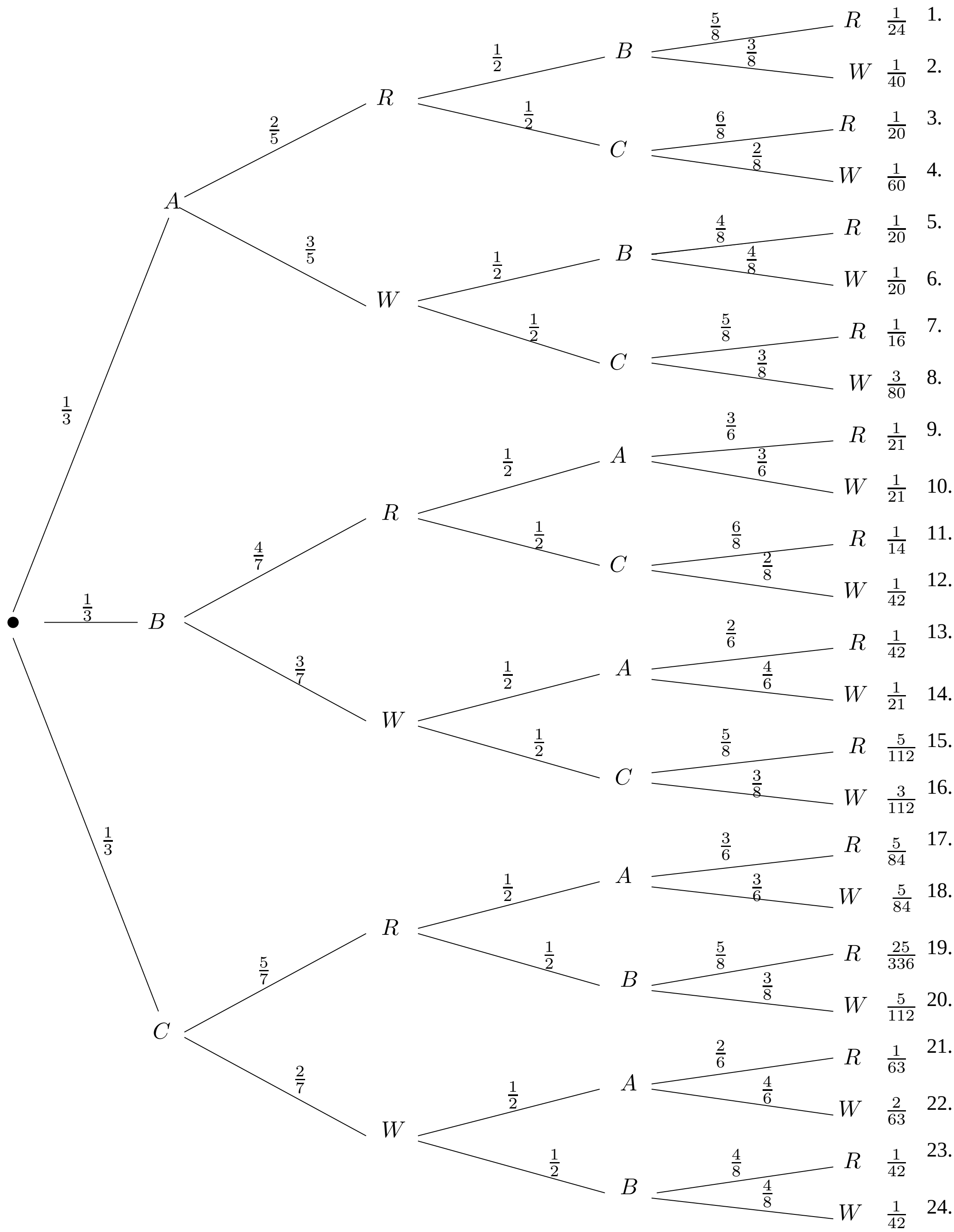
(2) Wir wählen daraus eine Kugel.

(3) Wir wählen eine der beiden anderen Urnen und legen die Kugel hinein.

(4) Wir wählen aus dieser eine Kugel.

Eine solche Folge nennt man einen *endlichen Zufallsprozess*. Die Wahrscheinlichkeit eines Ereignisses in einem solchen endlichen Zufallsprozess ermittelt man am übersichtlichsten mit Hilfe eines *Ereignisbaumes*. Wir demonstrieren das an unserem Beispiel.

Antwort: Wir stellen den Ereignisbaum mit den zugehörigen Wahrscheinlichkeiten auf:



Jeder Weg von links nach rechts gibt eine Ereignisfolge an, und jede mögliche Ereignisfolge ist in diesem Graph enthalten.

Betrachten wir den ersten Weg: Sei A_1 das Ereignis „erste Urne ist A“, A_2 das Ereignis „erste Kugel ist rot“, A_3 das Ereignis „zweite Urne ist B“, und A_4 das Ereignis „zweite Kugel ist rot“. Dann fragen wir nach

$$P(A_1 \cap A_2 \cap A_3 \cap A_4).$$

Nach (6.4) ist dies das Produkt der vorausgehenden bedingten Wahrscheinlichkeiten, die wir entlang der Kanten des Baumes eintragen: Bei der dritten Kante in jedem Weg ist zu beachten, daß man nur unter zwei Urnen wählen kann, und bei der vierten, daß sich durch das Hineinlegen einer Kugel die Wahrscheinlichkeiten für die abschließende Wahl geändert haben.

Insgesamt haben wir 24 verschiedene Ereignisfolgen mit folgender Zähldichte, die wir aus den jeweiligen Produkten erhalten

Ereignis	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Zähldichte	$\frac{1}{24}$	$\frac{1}{40}$	$\frac{1}{20}$	$\frac{1}{60}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{16}$	$\frac{1}{80}$	$\frac{1}{21}$	$\frac{1}{21}$	$\frac{1}{14}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{21}$

Ereignis	15	16	17	18	19	20	21	22	23	24
Zähldichte	$\frac{5}{112}$	$\frac{3}{112}$	$\frac{5}{84}$	$\frac{5}{84}$	$\frac{25}{336}$	$\frac{5}{112}$	$\frac{1}{63}$	$\frac{2}{63}$	$\frac{1}{42}$	$\frac{1}{42}$

Für unser Gesamtproblem besteht der Elementarereignisraum aus den 24 verschiedenen Wegen, die wir mit Nummern 1 bis 24 belegen, also

$$\Omega = \{1, 2, 3, \ldots, 24\}$$

mit der angegebenen Zähldichte. Uns interessiert das Ereignis „beide gezogenen Kugeln sind rot“, d.h. die Teilmenge

$$A = \{1, 3, 9, 11, 17, 19\}.$$

Wir erhalten

$$\begin{aligned} P(A) &= \frac{1}{24} + \frac{1}{20} + \frac{1}{21} + \frac{1}{14} + \frac{5}{84} + \frac{25}{336} \\ &= \frac{70 + 84 + 80 + 120 + 100 + 125}{1680} = \underline{\underline{\frac{579}{1680}}} = \underline{\underline{0,3446}}\ldots \end{aligned}$$

Bis jetzt haben wir (6.7) nicht benutzt. Stellen wir aber die

Frage 2: Wie groß ist die Wahrscheinlichkeit, daß die erste Kugel rot ist?

so können wir (6.7) benutzen. Da wir aber 11 rote und 8 weiße Kugeln haben, könnte man vermuten, daß die gesuchte Wahrscheinlichkeit

$$\frac{11}{19} = 0,5789\ldots$$

ist. Zweifler könnten nun aber behaupten, daß die ungleiche Verteilung der Kugeln auf die Urnen Einfluß auf den Ausgang haben könnte. Wir wollen das nachprüfen: Sei Ω die Menge aller Kugeln, seien A, B, C die Ereignisse „Kugel ist in A, B, C“ und seien R und W die Ereignisse „Kugel ist rot bzw. weiß“.

Offensichtlich gilt, daß alle Urnen gleichwahrscheinlich gewählt werden, und daß innerhalb einer Urne die Kugelwahlen gleichwahrscheinlich sind. Es folgt

$$P(A) = P(B) = P(C) = \frac{1}{3}$$
$$P(R|A) = \frac{2}{5}, \qquad P(R|B) = \frac{4}{7}, \qquad P(R|C) = \frac{5}{7}.$$

Aus (5.7) erhalten wir nun, da $\Omega = A \sqcup B \sqcup C$

$$\begin{aligned} P(R) &= P(R|A) \cdot P(A) + P(R|B) \cdot P(B) + P(R|C) \cdot P(C) \\ &= \frac{2}{5} \cdot \frac{1}{3} + \frac{4}{7} \cdot \frac{1}{3} + \frac{5}{7} \cdot \frac{1}{3} = \frac{1}{3} \cdot \frac{14 + 20 + 25}{35} \\ &= \frac{1}{3} \cdot \frac{59}{35} = \frac{59}{105} = 0,5619\ldots \end{aligned}$$

Problem 3: Wir ändern die Kugelverteilungen in den Urnen, indem wir wahllos eine Urne auswählen, ihr eine Kugel entnehmen und diese in eine andere Urne legen. Dieses Verfahren wiederholen wir. Beim Studium der Markoff-Ketten fragt man nach dem Zustand der 3 Urnen (also nach der Kugelverteilung) nach n Schritten.

$$\Omega = \text{Menge aller möglichen Verteilungen der Kugeln auf die 3 Urnen}$$
$$|\Omega| = \binom{r+2}{2} \cdot \binom{s+2}{2} \quad \text{nach (3.7)} \quad \begin{aligned} r &= \text{ Anzahl der roten Kugeln} \\ s &= \text{ Anzahl der schwarzen Kugeln} \end{aligned}$$

Wir haben also $\binom{13}{2} \cdot \binom{10}{2} = 3510$ mögliche Zustände Z_i.
Für die Markoff-Kette ist der Elementarereignisraum

$$\Omega^{n+1},$$

wobei die k-te Komponente den Zustand nach $k-1$ Schritten angibt. Sei $A_j(k)$ das Ereignis, daß das System nach k Schritten im Zustand Z_j ist. Da $A_1(k) \sqcup \ldots \sqcup A_{3510}(k) = \Omega^{n+1}$, folgt aus (6.7)

$$P(A_j(k)) = \sum_{r=1}^{3510} P(A_j(k)|A_r(k-1)) \cdot P(A_r(k-1)).$$

Damit läßt sich $P(A_j(k))$ aus den Zustandswahrscheinlichkeiten der Ereignisse $A_r(k-1)$ und den sog. *Übergangswahrscheinlichkeiten*

$$P(A_j|A_r)$$

ermitteln.

Da in unserem Beispiel $P(A_j|A_r)$ bereits eine 3510×3510-Matrix füllt, wollen wir von einer Berechnung dieser Übergangswahrscheinlichkeiten absehen.

Der nächste Satz erlaubt es uns, aus der Wahrscheinlichkeit

$$P(E|A_k)$$

des Eintretens von E unter den Voraussetzungen A_k auf die Wahrscheinlichkeit der Voraussetzungen A_k zu schließen, falls man weiß, daß E eingetreten ist.

6.9 Bayes'sche Regel: Sei (Ω, P) ein W-Raum und $E \subset \Omega$ ein Ereignis. Sei $\mathfrak{A} = \{A_k\}$ eine endliche oder abzählbar unendliche Partition von Ω. Es gelte $P(E) > 0$ und $P(A) > 0$ für alle $A \in \mathfrak{A}$. Dann gilt für $A_k \in \mathfrak{A}$

$$P(A_k|E) = \frac{P(A_k) \cdot P(E|A_k)}{\sum\limits_{A \in \mathfrak{A}} P(A) \cdot P(E|A)}.$$

Beweis: Aus (6.7) und dem Multiplikationssatz folgt

$$P(A_k|E) = \frac{P(E \cap A_k)}{P(E)} = \frac{P(A_k) \cdot P(E|A_k)}{P(E)} = \frac{P(A_k) \cdot P(E|A_k)}{\sum\limits_{A \in \mathfrak{A}} P(A) \cdot P(E|A)}.$$

□

6.10 Beispiel: In einem Betrieb werden in drei Anlagen Glühbirnen hergestellt, jede arbeitet mit Ausschuß. Der prozentuale Anteil an der Produktion und der prozentuale Ausschuß ist durch folgende empirisch ermittelte Tabelle gegeben

Anlage	A	B	C
Produktionsanteil	50%	40%	10%
Ausschuß	3%	4%	6%

(a) Wie groß ist die Wahrscheinlichkeit, daß eine zufällig gewählte Birne defekt ist?

(b) Wie groß ist dann die Wahrscheinlichkeit, daß diese Birne in A produziert wurde?

Antwort: Sei E das Ereignis „Glühbirne defekt“ und seien A, B, C die Ereignisse „die Glühbirne wurde in A, B, C hergestellt“.
Nach (6.7) gilt

$$\begin{aligned} P(E) &= P(A) \cdot P(E|A) + P(B) \cdot P(E|B) + P(C) \cdot P(E|C) \\ &= 0,5 \cdot 0,03 + 0,4 \cdot 0,04 + 0,1 \cdot 0,06 \\ &= 0,037 \end{aligned}$$

Das beantwortet (a). Bei (b) wird nach $P(A|E)$ gefragt. Wir benutzen (5.10):

$$\begin{aligned} P(A|E) &= \frac{P(A) \cdot P(E|A)}{P(A) \cdot P(E|A) + P(B) \cdot P(E|B) + P(C) \cdot P(E|C)} \\ &= \frac{0,5 \cdot 0,03}{0,037} = \frac{0,015}{0,037} \\ &= 0,4054\ldots \end{aligned}$$

6.11 Bemerkung: Fragen wie Beispiel 6.10 und viele der vorausgehenden Beispiele sind Fragen zum Stichprobenproblem, die in vielen Fertigungszweigen von Bedeutung sind.

Wir wenden uns nun einem neuen Begriff zu, der Unabhängigkeit. Zwei Ereignisse E und F heißen *unabhängig*, wenn das Eintreten von E die Wahrscheinlichkeit für das Eintreten von F nicht beeinflußt, wenn also

$$P(F) = P(F|E).$$

Da nun

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

falls $P(E) > 0$, ergeben diese Gleichungen

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

Letztere Gleichung benutzen wir zur formalen Definition der Unabhängigkeit, die auch im Falle $P(E) = 0$ noch sinnvoll ist.

6.12 Definition: Ereignisse $E_1, \ldots, E_n$ in einem W-Raum Ω heißen *paarweise unabhängig*, falls

$$P(E_i \cap E_j) = P(E_i) \cdot P(E_j) \quad \forall\, i \neq j$$

und (insgesamt) *unabhängig*, falls

$$P(E_{i_1} \cap \ldots \cap E_{i_k}) = P(E_{i_1}) \cdot \ldots \cdot P(E_{i_k}) \quad \forall\, i_1 < i_2 < \cdots < i_k$$

6.13 Beispiele: (1) Zwei ideale Münzen werden geworfen. Sei E das Ereignis „höchstens einmal Kopf" und F das Ereignis „mindestens einmal Kopf und mindestens einmal Zahl". Der W-Raum ist

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

mit konstanter Zähldichte $1/4$. Die Ereignisse E und F sind

$$E = \{KZ, ZK, ZZ\} \qquad F = \{KZ, ZK\} \qquad E \cap F = F.$$

Es folgt

$$P(E) = \frac{3}{4}, \qquad P(F) = P(E \cap F) = \frac{2}{4} = \frac{1}{2}.$$

Also

$$P(E \cap F \neq P(E) \cdot P(F).$$

Die Ereignisse E und F sind abhängig.

(2) Drei ideale Münzen werden geworfen. E und F seien wie in Beispiel (1)

$$\begin{aligned} \Omega &= \{KKK, KKZ, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ\} \\ E &= \{KZZ, ZKZ, ZZK, ZZZ\}, \\ F &= \{KKZ, KZK, KZZ, ZKK, ZKZ, ZZK\} \\ E \cap F &= \{KZZ, ZKZ, ZZK\}. \end{aligned}$$

Es folgt

$$P(E \cap F) = \frac{3}{8} = \frac{4}{8} \cdot \frac{6}{8} = P(E) \cdot P(F).$$

Also sind E und F hier unabhängig.

Diese Beispiele lassen erkennen, daß man in jedem Einzelfall genau nachprüfen muß, ob zwei Ereignisse unabhängig sind oder nicht. Es genügt oft nicht, sich auf seine Intuition zu verlassen.

Teil II

7 Diskrete Zufallsvariable

Im § 5 haben wir aus gegebenen W-Räumen neue konstruierte (vergl. die Sätze (5.2), (5.3), (5.7), (5.10) und (5.13)). Mit Ausnahme der Polynomialverteilung lassen sich diese Konstruktionen über ein allgemeines Verfahren gewinnen, das viele praktische Anwendungen hat.

7.1 Definition: Eine *Zufallsvariable* auf einem diskreten Wahrscheinlichkeitsraum (Ω, P) ist eine Funktion

$$X : \Omega \longrightarrow \mathbb{R}.$$

7.2 Beispiele: (1) Beim Werfen zweier Würfel besteht der zugehörige W-Raum aus allen Zahlenpaaren

$$\Omega = \{(1,1), (1,2), \ldots, (6,6)\}.$$

Die Zähldichte ist konstant $\frac{1}{36}$. Jedem Zahlenpaar ordnen wir die Summe zu und erhalten die Zufallsvariable

$$X : \Omega \longrightarrow \mathbb{R}, \qquad (i,j) \longmapsto i+j.$$

(2) Bei der Herleitung der hypergeometrischen Verteilung zogen wir n Kugeln aus einer Urne mit r roten und s schwarzen Kugeln. Der zugehörige W-Raum besteht aus allen n-elementigen Teilmengen dieser Kugeln mit der konstanten Zähldichte $\binom{r+s}{n}$. Die hypergeometrische Verteilung gehört zur Zufallsvariablen

$$X : \Omega \longrightarrow \mathbb{R},$$

die einer solchen Teilmenge die Anzahl ihrer roten Kugeln zuordnet.

(3) Bei der Herleitung der Binomialverteilung besteht der W-Raum Ω aus allen n-Tupeln von Symbolen T (für Treffer) und N (für Niete). Die Zahldichte ordnet jedem n-Tupel den Wert

$$p^k \cdot q^{n-k}$$

zu, falls das Tupel genau k Treffer hat. Die Binomialverteilung gehört zur Zufallsvariablen

$$X : \Omega \longrightarrow \mathbb{R},$$

die jedem Tupel die Anzahl der Treffer zuordnet.

Wie im Paragraphen 5 definiert eine diskrete Zufallsvariable einen neuen W-Raum.

7.3 Satz: Sei (Ω, P) ein diskreter W-Raum und $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable. Sei $\hat{\Omega} = \operatorname{Bild} X = \{y \in \mathbb{R};\ \exists\, \omega \in \Omega \text{ mit } X(\omega) = y\}$. Dann ist

$$v : \hat{\Omega} \longrightarrow \mathbb{R}, \qquad y \longmapsto P(X^{-1}(y))$$

eine Zähldichte auf $\hat{\Omega}$.

7.4 Bezeichnung: v heißt *Verteilung* der Zufallsvariablen X.

Beweis: Da Ω endlich oder abzählbar unendlich ist, gilt dasselbe für $\hat{\Omega} = \operatorname{Bild} X$. Da P nur Werte ≥ 0 annimmt, ist $v(y) \geq 0 \quad \forall\, y \in \hat{\Omega}$. Weiter gilt

$$\Omega = \cup_{y\in\hat{\Omega}} X^{-1}(y) \quad \text{und} \quad X^{-1}(y_1) \cap X^{-1}(y_2) = \emptyset$$

für $y_1 \neq y_2$. Es folgt mit (1.6(3)) bzw. (5.17(3))

$$\sum_{y\in\hat{\Omega}} v(y) = \sum_{y\in\hat{\Omega}} P(X^{-1}(y)) = P(\cup_{y\in\hat{\Omega}} X^{-1}(y)) = P(\Omega) = 1.$$

□

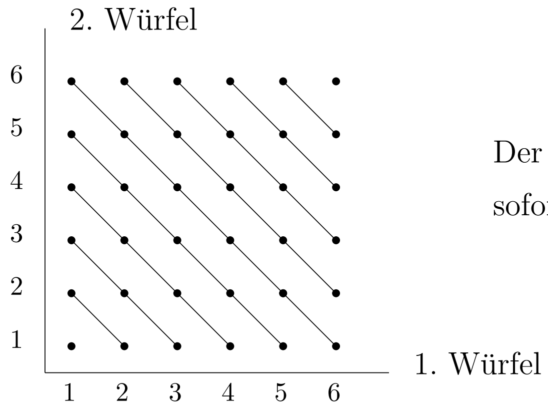
7.5 Beispiele: (1) Die hypergeometrische Verteilung und die Binomialverteilung sind die Verteilungen der Zufallsvariablen aus den Beispielen (7.2) (2) bzw. (7.2) (3). □

(2) Auf dem Laplace-Raum des Beispiels (7.2) (1) definieren wir zwei verschiedene Zufallsvariable.

$$X : \Omega \longrightarrow \mathbb{R}, \qquad (i,j) \longmapsto i + j$$
$$Y : \Omega \longrightarrow \mathbb{R}, \qquad (i,j) \longmapsto \max(i,j)$$

Verteilung von X:

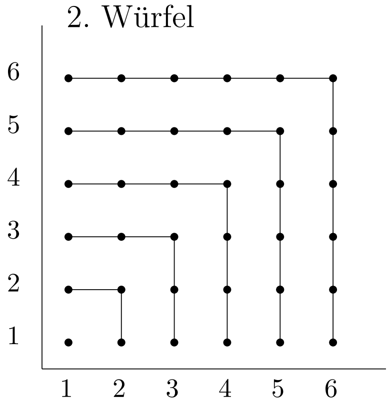
$$X(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$



Der Skizze links entnehmen wir sofort die Verteilungen von X.

y	2	3	4	5	6	7	8	9	10	11	12
$v(y)$	$\frac{1}{36}$	$\frac{2}{36}=\frac{1}{18}$	$\frac{3}{36}=\frac{1}{12}$	$\frac{4}{36}=\frac{1}{9}$	$\frac{5}{36}$	$\frac{6}{36}=\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Verteilung von Y: $Y(\Omega) = \{1, 2, 3, 4, 5, 6\}$



Der nebenstehenden Skizze entnimmt man sofort die Verteilungen von Y.

y_i	1	2	3	4	5	6
$\omega(y_i)$	$\frac{1}{36}$	$\frac{3}{36}=\frac{1}{12}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}=\frac{1}{4}$	$\frac{11}{36}$

Wir führen folgende Bezeichnungen ein, die im Weiteren immer wieder benutzt werden:

7.6 Bezeichnungen: Sei (Ω, P) ein diskreter W-Raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Seien $a \leq b$ aus $\mathbb{R}$ und $B \subset \mathbb{R}$. Wir setzen

$$\begin{aligned} P(X = a) &= P(\{\omega \in \Omega; X(\omega) = a\}) = P(X^{-1}(a)) \\ P(a \leq X \leq b) &= P(\{\omega \in \Omega; a \leq X(\omega) \leq b\}) = P(X^{-1}([a, b])) \\ P(X \in B) &= P(X^{-1}(B)) \end{aligned}$$

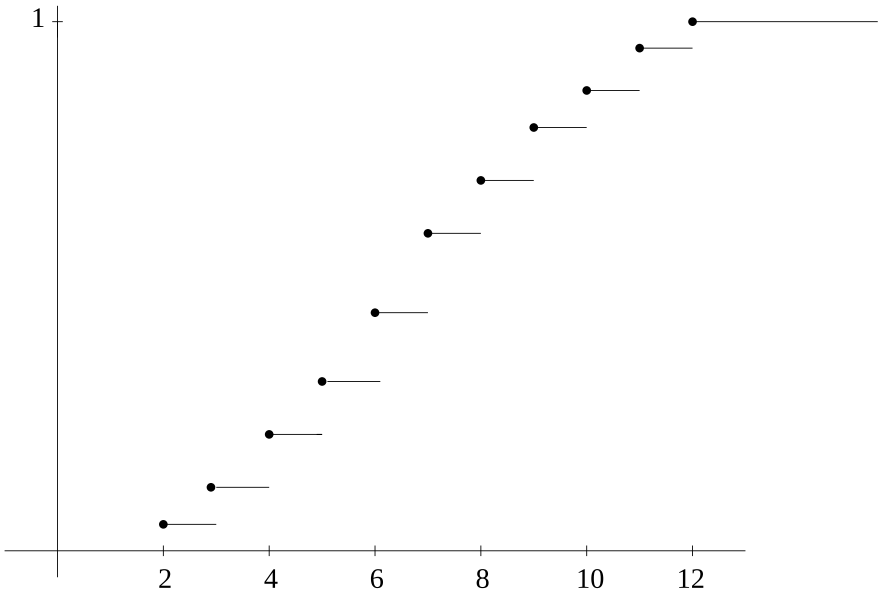
Entsprechend wird $P(a < X \leq b)$ definiert usw.

7.7 Definition: Sei $X : \Omega \to \mathbb{R}$ Zufallsvariable auf dem diskreten W-Raum (Ω, P). Dann heißt die Funktion

$$F : \mathbb{R} \longrightarrow \mathbb{R}, \qquad x \longmapsto P(X \leq x)$$

Verteilungsfunktion von X.

7.8 Beispiel: Sei Ω der Raum aller Würfe mit 2 Würfeln und X die Augensumme.



Die Verteilungsfunktion für eine Zufallsvariable auf einem endlichen W-Raum ist stets eine Treppenfunktion. Sprungstellen sind die Werte in $\mathbb{R}$, die im Bild von X auftauchen.

Folgende weitere Eigenschaften der Verteilungsfunktion macht man sich leicht klar:

Aufgabe: Für die Verteilungsfunktion $F : \mathbb{R} \to \mathbb{R}$ einer Zufallsvariablen auf einem endlichen W-Raum (Ω, P) gilt

(1) $F(x) = \sum\limits_{x_i \leq x} P(X = x_i), \quad \text{falls } \operatorname{Bild} X = \{x_1, \ldots, x_n\}$

(2) $P(a < X \leq b) = F(b) - F(a)$
$P(a \leq X \leq b) = F(b) - F(a - 0)$
$P(a < X) = 1 - F(a)$

Dabei ist $F(a - 0) = \lim\limits_{\substack{x \to x \\ x < a}} F(x)$. Falls $a \notin \operatorname{Bild} X$, ist $F(a - 0) = F(a)$.

8 Der Erwartungswert

Oft ist es schwierig, die Verteilung einer Zufallsvariablen explizit anzugeben. Daher interessiert man sich für charakteristische Größen, die leichter zu ermitteln sind und Auskünfte über die Zufallsvariable geben.

8.1 Definition: Sei (Ω, P) ein diskreter W-Raum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable auf Ω. Sei wieder $\hat{\Omega} = \operatorname{Bild} X \subset \mathbb{R}$. Konvergiert die Reihe

$$\sum_{x \in \hat{\Omega}} x \cdot P(X = x)$$

absolut, heißt ihre Summe $E(X)$ der *Erwartungswert* oder *Mittelwert* von X (und wird oft auch mit μ_X bezeichnet). Also

$$\mu_X = E(X) = \sum_{x \in \hat{\Omega}} x \cdot P(X = x).$$

Bemerkung: Da wir die Reihenfolge der Summation in (8.1) nicht festlegen und x negativ sein kann, müssen wir für eine sinnvolle Definition des Erwartungswertes die absolute Konvergenz fordern. Ist $\hat{\Omega}$ endlich, haben wir eine endliche Summe, also ist $E(X)$ dann immer definiert. Ist $\hat{\Omega}$ unendlich, braucht $E(X)$ nicht zu existieren.

Bei einer Zufallsvariablen interessieren wir uns für spezielle Ausgänge eines Experiments, z.B. die Augensumme beim Würfeln mit 2 Würfeln. Wiederholen wir das Experiment oft genug, liegt die relative Häufigkeit erfahrungsgemäß in der Nähe des Erwartungswertes.

8.2 Beispiele: (1) Würfeln mit 2 Würfeln: Wir ermitteln die Erwartungswerte der Zufallsvariablen X und Y aus Beispiel (7.5) (2).

$$\begin{aligned}
E(X) &= 2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{3}{36} + 5 \cdot \tfrac{4}{36} + 6 \cdot \tfrac{5}{36} + 7 \cdot \tfrac{6}{36} + 8 \cdot \tfrac{5}{36} + 9 \cdot \tfrac{4}{36} \\
&\qquad + 10 \cdot \tfrac{3}{36} + 11 \cdot \tfrac{2}{36} + 12 \cdot \tfrac{1}{36} \\
&= \tfrac{1}{36}(2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) \\
&\qquad = \tfrac{252}{36} = \underline{\underline{7}} \\
E(Y) &= 1 \cdot \tfrac{1}{36} + 2 \cdot \tfrac{3}{36} + 3 \cdot \tfrac{5}{36} + 4 \cdot \tfrac{7}{36} + 5 \cdot \tfrac{9}{36} + 6 \cdot \tfrac{11}{36} \\
&= \tfrac{1+6+15+28+45+66}{36} = \tfrac{161}{36} = \underline{\underline{4,472\ldots}}
\end{aligned}$$

(2) Ein Spieler würfelt. Erscheint eine Primzahl, erhält er diesen Betrag. Erscheint eine andere Zahl, muß er diesen Betrag bezahlen.

Wir definieren auf dem Ereignisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$ eine Zufallsvariable X:

Ω	1	2	3	4	5	6
X	-1	2	3	-4	5	-6

Sie gibt an, welche Beträge der Spieler gewinnt bzw. verliert. Die Verteilung von X ist

x_i	-6	-4	-1	2	3	5
$X(x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$.

Der Erwartungswert von X ist

$$\begin{aligned} E(x) &= -6 \cdot \tfrac{1}{6} - 4 \cdot \tfrac{1}{6} - 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} \\ &= \tfrac{-11+10}{6} = \underline{\underline{-\tfrac{1}{6}}}. \end{aligned}$$

8.3 Bemerkung: In der Spieltheorie wird der Erwartungswert E eines Spiels als der Wert des Spieles für den Spieler betrachtet. Ist E positiv, nennt man das Spiel *günstig* für den Spieler, denn er gewinnt mit größerer Wahrscheinlichkeit, als er verliert. Ist $E = 0$, nennt man das Spiel *fair*, im Falle $E < 0$ nennt man es *ungünstig*.

8.4 Bernoulli-Verteilung: Wir ermitteln den Erwartungswert der Bernoulli-Verteilung (3.10). Es gilt

$$E(X) = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot q^{n-k} = \sum_{k=1}^{n} \frac{k \cdot n!}{k!(n-k)!} p^k \cdot q^{n-k}.$$

Wir kürzen k und klammern $n \cdot p$ aus:

$$E(X) = n \cdot p \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}.$$

Wir setzen nun $\ell = k - 1$. Da k von 1 bis n läuft, läuft nun ℓ von 0 bis $n - 1$

$$E(X) = np \sum_{\ell=0}^{n-1} \frac{(n-1)!}{\ell!(n-1-\ell)!} p^\ell q^{n-1-\ell} = \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^\ell q^{n-1-\ell} = np(p+q)^{n-1} = np,$$

da $p + q = 1$. Wir fassen zusammen:

$$\boxed{\text{E(X)=np}}$$

Dieses Beispiel und erst recht der Versuch der Ermittlung des Erwartungswertes der hypergeometrischen Verteilung läßt erkennen, daß allgemeine Verfahren zur Berechnung des Erwartungswertes gefordert sind.

Wir beginnen mit einem einfachen Resultat:

8.5 Definition: Eine Zufallsvariable X heißt *symmetrisch verteilt* mit Symmetriepunkt a, wenn für alle $x \in \mathbb{R}$ gilt

(1) $a - x \in \operatorname{Bild} X \Longleftrightarrow a + x \in \operatorname{Bild} X$

(2) $P(X = a - x) = P(X = a + x)$.

8.6 Satz: Ist X eine symmetrisch verteilte Zufallsvariable mit Symmetriepunkt a, so gilt, falls der Erwartungswert existiert,

$$E(X) = a.$$

Beweis: Nach (8.5) (1) liegt das Bild von X symmetrisch um a. Es gibt also reelle Zahlen $r_1, r_2, r_3, \ldots$, so daß

$$\operatorname{Bild} X \subset \{a, a \pm r_1, a \pm r_2, \ldots\}.$$

Beachte hier, daß a nicht im Bils von X zu liegen braucht. Es folgt (wir behandeln den unendlichen Fall, der endliche ist analog; da die Reihe (8.1) absolut konvergiert, dürfen wir sie beliebig anordnen)

$$E(X) = a \cdot P(X = a) + \sum_{i=1}^{\infty} [(a + r_i) \cdot P(X = a + r_i) + (a - r_i) \cdot P(X = a - r_i)]$$

(wenn $a \notin \operatorname{Bild} X$, ist $P(X = a) = 0$). Also, wegen 8.5 (2)

$$\begin{aligned}
E(X) &= a \cdot P(X = a) + \sum_{i=1}^{\infty} (a + r_i + a - r_i) \cdot P(X = a + r_i) \\
&= a \cdot P(X = a) + \sum_{i=1}^{\infty} 2a \cdot P(X = a + r_i) \\
&= a \cdot \left[P(X = a) + \sum_{i=1}^{\infty} (P(X = a + r_i) + P(X = a - r_i)) \right] \\
&= a \cdot P(X^{-1}\{a, a \pm r_1, a \pm r_2, \ldots\}) = a \cdot P(\Omega) = a.
\end{aligned}$$

□

8.7 Beispiel: Die Augensumme beim Würfeln mit 2 Würfeln ist symmetrisch verteilt mit Symmetriepunkt 7.

Jetzt wollen wir zu einem komplizierteren Sachverhalt mit zahlreichen Anwendungsmöglichkeiten übergehen.

Bekanntlich lassen sich Abbildungen nach $\mathbb{R}$ punktweise addieren und multiplizieren. Explizit definiert man

8.8 Sind $X, Y : \Omega \longrightarrow \mathbb{R}$ Abbildungen und c eine reelle Zahl, dann setzen wir

$$\begin{aligned} X+Y: &\quad \Omega \longrightarrow \mathbb{R} &\qquad a \longmapsto X(a)+Y(a) \\ X+c: &\quad \Omega \longrightarrow \mathbb{R} &\qquad a \longmapsto X(a)+c \\ X\cdot Y: &\quad \Omega \longrightarrow \mathbb{R} &\qquad a \longmapsto X(a)\cdot Y(a) \\ c\cdot X: &\quad \Omega \longrightarrow \mathbb{R} &\qquad a \longmapsto c\cdot X(a) \end{aligned}$$

Auf diese Art erhalten wir neue Zufallsvariable.

Weiter erhalten wir neue Zufallsvariable offensichtlich wie folgt:

8.9 Sind $X, Y : \Omega \longrightarrow \mathbb{R}$ Zufallsvariable und sind

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \qquad g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

beliebige Abbildungen, so definieren wir neue Zufallsvariable

$$f(X) = f \circ X : \ \Omega \xrightarrow{X} \mathbb{R} \xrightarrow{f} \mathbb{R}\,, \quad \omega \longmapsto f(X(\omega))$$

$$g(X,Y) = g \circ (X,Y) : \ \Omega \xrightarrow{(X,Y)} \mathbb{R} \times \mathbb{R} \xrightarrow{g} \mathbb{R}\,, \ \omega \longmapsto g(X(\omega), Y(\omega)).$$

Bevor wir die Erwartungswerte dieser neuen Zufallsvariablen berechnen, wollen wir uns mit dem zweiten Fall näher beschäftigen: Haben wir zwei Zufallsvariable $X, Y : \Omega \longrightarrow \mathbb{R}$, können wir wie in § 7 einen neuen W-Raum bilden. Sei

$$\Omega_X = \operatorname{Bild} X \subset \mathbb{R}, \qquad \Omega_Y = \operatorname{Bild} Y \subset \mathbb{R}.$$

8.10 Satz: Sei (Ω, P) ein diskreter W-Raum und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ Zufallsvariable. Dann definiert

$$v : \Omega_X \times \Omega_Y \longrightarrow \mathbb{R}, \quad (x,y) \longmapsto P(X = x, Y = y) := P(X^{-1}(x) \cap Y^{-1}(y))$$

eine Zähldichte auf $\Omega_X \times \Omega_Y$.

Beweis: Da P nur positive Werte hat, ist $v(x,y) \geq 0$ für alle $(x,y) \in \Omega_X \times \Omega_Y$.
Weiter gilt

$$\Omega_X \times \Omega_Y = \coprod_{x \in \Omega_X} \{x\} \times \Omega_Y = \coprod_{y \in \Omega_Y} \Omega_X \times \{y\} \qquad (\textstyle\coprod = \text{disjunkte Vereinigung.})$$

Es folgt mit (5.17) (3)

$$\sum_{(x,y)\in\Omega_X\times\Omega_Y} v(x,y)$$

$$= \sum_{(x,y)\in\Omega_X\times\Omega_Y} P(X^{-1}(x)\cap Y^{-1}(y)) = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} P(X^{-1}(x)\cap Y^{-1}(y))$$

$$= \sum_{x\in\Omega_X} P\Big(\coprod_{y\in\Omega_Y} X^{-1}(x)\cap Y^{-1}(y)\Big) = \sum_{x\in\Omega_X} P\Big(X^{-1}(x)\cap \coprod_{y\in\Omega_Y} Y^{-1}(y)\Big)$$

$$= \sum_{x\in\Omega_X} P(X^{-1}(x)\cap\Omega) = \sum_{x\in\Omega_X} P(X^{-1}(x)) = P\Big(\coprod_{x\in\Omega_X} X^{-1}(x)\Big) = P(\Omega) = 1$$

$\square$

8.11 Bezeichnung: $v : \Omega_X \times \Omega_Y \longrightarrow \mathbb{R},\ (x,y) \longmapsto P(X = x, Y = y)$ heißt *gemeinsame Verteilung* von X und Y. Die Verteilung $v_X : \Omega_X \longrightarrow \mathbb{R}$ und $v_Y : \Omega_Y \longrightarrow \mathbb{R}$ von X bzw. Y heißen *Randverteilungen* von v.

Im Beweis von (8.10) haben wir gezeigt, daß sich die Randverteilungen aus der gemeinsamen Verteilung berechnen lassen:

8.12 Für $(x,y) \in \Omega_X \times \Omega_Y$ gilt:

$$\begin{aligned} v_X(x) = P(X = x) &= \sum_{y\in\Omega_Y} P(X = x, Y = y) = \sum_{y\in\Omega_Y} v(x,y) \\ v_Y(y) = P(Y = y) &= \sum_{x\in\Omega_X} P(X = x, Y = y) = \sum_{x\in\Omega_X} v(x,y). \end{aligned}$$

8.13 Beispiel: Wir werfen eine ideale Münze 3 mal. Also hat Ω acht Elemente

$$\Omega = \{(Z,Z,Z),(Z,Z,W),(Z,W,Z),(Z,W,W),(W,Z,Z),(W,Z,W),\\ (W,W,Z),(W,W,W)\}.$$

X sei die Anzahl der Wappen, Y die Position, an der zum erstenmal W auftritt, wobei wir $Y(Z,Z,Z) = 4$ setzen.

$$\Omega_X = \operatorname{Bild} X = \{0,1,2,3\}, \qquad \Omega_Y = \operatorname{Bild} Y = \{1,2,3,4\}$$

Bild $Y\backslash$Bild X	0	1	2	3	v_Y
1	0	$1/8$	$1/4$	$1/8$	$1/2$
2	0	$1/8$	$1/8$	0	$1/4$
3	0	$1/8$	0	0	$1/8$
4	$1/8$	0	0	0	$1/8$
v_X	$1/8$	$3/8$	$3/8$	$1/8$	1

z.B. $P(X=0, Y=3) = 0$

8.14 Beispiel: Sei $g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ die Addition, d.h. $g(x,y) = x+y$. Sei weiter Ω, X, Y wie im Beispiel (8.13). Beachte:

$$|\,\text{Bild}\,X \times \text{Bild}\,Y| = 16. \quad \text{aber} \quad |\Omega| = 8$$

Da $g(X,Y)$ eine Abbildung von Ω nach $\mathbb{R}$ ist, kann $g(X,Y)$ höchstens 8 Werte haben.

Ω	(Z,Z,Z)	(Z,Z,W)	(Z,W,Z)	(Z,W,W)	(W,Z,Z)	(W,Z,W)	(W,W,Z)	(W,W,W)
(X,Y)	$(0,4)$	$(1,3)$	$(1,2)$	$(2,2)$	$(1,1)$	$(2,1)$	$(2,1)$	$(3,1)$
$g(X,Y)$	4	4	3	4	2	3	3	4

Also: $\hat{\Omega} = \text{Bild}(g(X,Y)) = \{2,3,4\}$. Es folgt

$$E(g(X,Y)) = 2 \cdot \frac{1}{8} + 3 \cdot \frac{3}{8} + 4 \cdot \frac{4}{8} = \frac{1}{8}(2+9+16) = \frac{27}{8}.$$

Es ist nicht von ungefähr, daß im Schema (8.13) achtmal die 0 auftritt:

8.15 Sei (Ω, P) ein diskreter W-Raum und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ zwei Zufallsvariable. Dann gilt für die gemeinsame Verteilung

$$(x,y) \notin \text{Bild}(X,Y) = \{(X(\omega), Y(\omega)); \omega \in \Omega\} \Longrightarrow P(X=x, Y=y) = 0$$

denn: $P(X=x, Y=y) = P(X^{-1}(x) \cap Y^{-1}(y))$ und $X^{-1}(x) \cap Y^{-1}(y) = \emptyset$, falls $(x,y) \notin \text{Bild}(X,Y)$.

8.16 Satz: Sei (Ω, P) ein diskreter W-Raum, $X, Y : \Omega \longrightarrow \mathbb{R}$ Zufallsvariable und $f : \mathbb{R} \longrightarrow \mathbb{R}$, $g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ Funktionen. Dann gilt

$$E(f(X)) = \sum_{x \in \text{Bild}\,X} f(x) \cdot P(X=x)$$

$$E(g(X,Y)) = \sum_{x \in \text{Bild}\,X} \sum_{y \in \text{Bild}\,Y} g(x,y) \cdot P(X=x, Y=y).$$

Beweis: Die erste Formel folgt aus der zweiten: Setze $Y = 0$ und

$$g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}, \qquad g(x, y) = f(x).$$

Dann gilt: $g(X, Y)(\omega) = g(X(\omega), Y(\omega)) = f(X(\omega)) = f(X)(\omega)$. Also

$$\begin{aligned} E(f(X)) = E(g(X, Y)) &= \sum_{x \in \operatorname{Bild} X} g(x, 0) \cdot P(X = x, Y = 0) \\ &= \sum_{x \in \operatorname{Bild} X} f(x) \cdot P(X = x, Y = 0). \end{aligned}$$

Da $P(X = x, Y = 0) = P(X^{-1}(x) \cap Y^{-1}(0)) = P(X^{-1}(x) \cap \Omega) = P(X^{-1}(x)) = P(X = x)$, erhalten wir die erste Formel.
Wir setzen $Z = g(X, Y)$. Dann ist Z die Abbildung

$$\begin{aligned} Z : \Omega &\xrightarrow{(X,Y)} \mathbb{R} \times \mathbb{R} \xrightarrow{g} \mathbb{R} \\ \omega &\longmapsto (X(\omega), Y(\omega)) \longmapsto g(X(\omega), Y(\omega)) \end{aligned}$$

$$(*) \quad E(Z) = \sum_{z \in \operatorname{Bild} Z} z \cdot P(Z^{-1}(z)) = \sum_{z \in \operatorname{Bild} Z} z \cdot \sum_{(x,y) \in g^{-1}(z)} P(X = x, Y = y),$$

denn

$$\begin{aligned} Z^{-1}(z) = (X, Y)^{-1}(g^{-1}(z)) &= \coprod_{(x,y) \in g^{-1}(z)} (X, Y)^{-1}(x, y) \\ &= \coprod_{(x,y) \in g^{-1}(z)} X^{-1}(x) \cap Y^{-1}(y). \end{aligned}$$

Da jedes Paar $(x, y) \in \operatorname{Bild}(X, Y)$ in einem $g^{-1}(z)$mit $z \in \operatorname{Bild} Z$ liegt, können wir (*) umschreiben:

$$\begin{aligned} E(Z) &= \sum_{(x,y) \in \operatorname{Bild}(X,Y)} g(x, y) \cdot P(X = x, Y = y) \\ &= \sum_{x \in \operatorname{Bild} X} \sum_{y \in \operatorname{Bild} Y} g(x, y) \cdot P(X = x, Y = y) \end{aligned}$$

Die letzte Summation geht über **alle** Paare $(x, y) \in \operatorname{Bild} X \times \operatorname{Bild} Y$. Ist aber $(x, y) \notin \operatorname{Bild}(X, Y)$, so ist $P(X = x, Y = y)$ nach (8.15) null. $\square$

8.17 Folgerung: Sind X, Y Zufallsvariable auf einem diskreten W-Raum (Ω, P) und $a, b, c \in \mathbb{R}$, dann gilt

$$E(aX + bY + c) = a \cdot E(X) + b \cdot E(Y) + c.$$

Beweis: Benutze (8.16) mit

$$g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}, \qquad (x, y) \longmapsto ax + by + c.$$

Dann gilt, wir setzen $A = \operatorname{Bild} X$ und $B = \operatorname{Bild} Y$,

$$\begin{aligned}
&E(aX + bY + c) \\
&= \sum_{x \in \operatorname{Bild} X} \sum_{y \in \operatorname{Bild} Y} (ax + by + c) \cdot P(X = x, Y = y) \\
&= \sum_{x \in A} \sum_{y \in B} ax \cdot P(X = x, Y = y) + \sum_{x \in A} \sum_{y \in B} byP(X = x, Y = y) \\
&\qquad + \sum_{x \in A} \sum_{y \in B} c \cdot P(X = x, Y = y) \\
&= a \cdot \sum_{x \in A} x \cdot \sum_{y \in B} P(X = x, Y = y) + b \sum_{y \in B} y \sum_{x \in A} P(X = x, Y = y) + c \cdot P(\Omega) \\
&= a \cdot \sum_{x \in A} x \cdot P(X = x) + b \sum_{y \in B} y \cdot P(Y = y) + c \quad \text{nach (8.12)} \\
&= a \cdot E(X) + b \cdot E(Y) + c
\end{aligned}$$

□

Diese kleine Formel vereinfacht die Berechnung des Erwartungswertes in vielen Fällen.

8.18 Erwartungswert der Bernoulli-Verteilung: Sei (Ω, P) der W-Raum eines Bernoulli-Experiments der Länge n, d.h. die Elemente von Ω sind n-Tupel von Symbolen T und N (vergl. Bspl. 7.2 (3)). Sei

$$X_k : \Omega \longrightarrow \mathbb{R}$$

die Zufallsvariable, gegeben durch

$$X_k(a) = \begin{cases} 1 & k\text{-te Koordinate von } a \text{ ist } T \\ 0 & k\text{-te Koordinate von } a \text{ ist } N. \end{cases}$$

Zu berechnen ist der Erwartungswert von

$$X = X_1 + \cdots + X_n.$$

Nach (8.17) gilt

$$E(X) = \sum_{k=1}^{n} E(X_k).$$

Die Wahrscheinlichkeit dafür, daß an der k-ten Stelle ein Treffer vorliegt, ist aber p, so daß $E(X_k) = p$ für alle k. Also

$$E(X) = n \cdot p.$$

Als **Folgerung** erhalten wir die kombinatorische Formel, die wir im Beispiel 8.4 explizit nachgerechnet haben.

8.19 $\sum\limits_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = n \cdot p.$

Insbesondere erhalten wir für $p = q = \frac{1}{2}$:

$\sum\limits_{k=0}^{n} k \cdot \binom{n}{k} = 2^{n-1} \cdot n$

8.20 Hypergeometrische Verteilung: Sei (Ω, P) der W-Raum, der zur hypergeometrischen Verteilung führt, genauer: Ω ist die Menge aller n-Tupel ω von roten bzw. schwarzen Kugeln. Sei $X_k : \Omega \to \mathbb{R}$ die Zufallsvariable

$$X_k(\omega) = \begin{cases} 1 & k\text{-te Kugel von } \omega \text{ ist rot} \\ 0 & k\text{-te Kugel von } \omega \text{ ist schwarz.} \end{cases}$$

Da an der k-ten Stelle eine schwarze oder rote Kugel auftreten muß und es r rote und s schwarze Kugeln gibt, ist

$$E(X_k) = \frac{r}{r+s} \quad \forall\ k.$$

Gesucht wird der Erwartungswert von

$$X = X_1 + X_2 + \cdots + X_n.$$

Es folgt:

$$\underline{\underline{E(X) = n \cdot \frac{r}{r+s}}}.$$

Wieder erhalten wir eine kombinatorische Formel, denn nach (5.1) gilt

$$E(X) = \sum_{k=0}^{n} k \cdot h(k; n, r, s) = \sum_{k=0}^{n} k \cdot \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}} = n \cdot \frac{r}{r+s}.$$

Es folgt

$$\sum_{k=0}^{n} k \cdot \binom{r}{k}\binom{s}{n-k} = \frac{n \cdot r \cdot (r+s)!}{(r+s)n!(r+s-n)!} = \frac{r \cdot (r+s-1)!}{(n-1)!(r+s-1-(n-1))!}.$$

Also

8.21

$$\underline{\underline{\sum_{k=0}^{n} k \cdot \binom{r}{k}\binom{s}{n-k} = r \cdot \binom{r+s-1}{n-1}}} \qquad n \leq r+s$$

8.22 Geometrische Verteilung für einen Bernoulli-Versuch mit Wahrscheinlichkeit p. Der Ausgangs-W-Raum, d.h. Ω besteht aus allen Tupeln der Form $(N, N, \ldots, N, T)$ und

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

ordnet jedem Tupel seine Länge zu. Sei zunächst $p \neq 1$

$$\underline{\underline{E(X)}} = \sum_{k=0}^{n} k \cdot g(k; p) = \sum_{k=0}^{\infty} k \cdot q^{k-1} \cdot p = \frac{p}{(1-q)^2} = \frac{1}{\underline{\underline{p}}}$$

Denn für $|x| < 1$ gilt, $\frac{1}{1-x} = \Sigma_{k=0}^{\infty} x^k$. Aus der Theorie der Potenzreihen, weiß man, daß

$$\sum_{k=0}^{\infty} k \cdot x^{k-1} = \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2}.$$

Für $p = 1$ gilt dasselbe Ergebnis. □

9 Unabhängigkeit von Zufallsvariablen

Wir wollen den Begriff der Unabhängigkeit auf Zufallsvariable erweitern.

9.1 Definition: Sei (Ω, P) ein diskreter W-Raum. Dann heißen die Zufallsvariablen $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ *unabhängig*, wenn für jede Teilmenge $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ und jedes k-Tupel reeller Zahlen $(x_1, \ldots, x_k) \in \mathbb{R}^k$ gilt:

$$P(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) := P\left(\bigcap_{r=1}^{k} X_{i_r}^{-1}(x_r)\right) = \prod_{r=1}^{k} P(X_{i_r} = x_r).$$

D.h. für jedes Tupel $(x_1, \ldots, x_k) \in \mathbb{R}^k$ sind die Ereignisse $X_{i_1}^{-1}(x_1), \ldots, X_{i_k}^{-1}(x_k)$ unabhängig.

9.2 Satz: Sei $X : \Omega \longrightarrow \mathbb{R}$ konstant und $Y : \Omega \longrightarrow \mathbb{R}$ eine beliebige Zufallsvariable auf einem diskreten W-Raum (Ω, P). Dann sind X und Y unabhängig.

Beweis: Nach Voraussetzung gibt es ein $c \in \mathbb{R}$, so daß $X(\omega) = c$ für alle $\omega \in \Omega$. Also

$$X^{-1}(c) = \Omega \quad \text{und} \quad X^{-1}(x) = \emptyset \text{ für } x \neq c,\ x \in \mathbb{R}.$$

Es folgt:

$$P(X = c, Y = y) = P(X^{-1}(c) \cap Y^{-1}(y)) = P(\Omega \cap Y^{-1}(y)) = P(Y = y)$$

und für $x \neq c$

$$P(X = x, Y = y) = P(X^{-1}(x) \cap Y^{-1}(y)) = P(\emptyset) = 0.$$

Da $P(X = c) = P(\Omega) = 1$ und für $x \neq c$ gilt $P(X = x) = 0$, folgt

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y).$$

□

9.3 Satz: Sind $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ unabhängige Zufallsvariable auf einem diskreten W-Raum (Ω, P) und $f_i : \operatorname{Bild} X_i \longrightarrow \mathbb{R},\ i = 1, \ldots, n$ beliebige Funktionen, dann sind auch $f_1(X_1), \ldots, f_n(X_n)$ unabhängig.

Der Satz folgt aus einem allgemeineren Resultat. Wir führen zunächst eine Bezeichnung ein.

9.4 Bezeichnung: Für Zufallsvariable $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ und Teilmengen $A_1, \ldots, A_n \subset \mathbb{R}$ setzen wir

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right), \quad (X_i \in A_i) := X_i^{-1}(A_i).$$

9.5 Satz: Sind $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ unabhängige Zufallsvariable auf einem diskreten W-Raum (Ω, P) und $A_1, \ldots, A_n \subset \mathbb{R}$ beliebige Teilmengen, dann gilt für alle Teilmengen $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$

$$P(X_{i_1} \in A_{i_1}, \ldots, X_{i_k} \in A_{i_k}) = \prod\nolimits_{r=1}^{k} P(X_{i_r} \in A_{i_r}) \qquad (*)$$

d.h. die Ereignisse $(X_1 \in A_1), \ldots, (X_n \in A_n)$ sind unabhängig.

Beweis: Induktion nach n. Für $n = 1$ ist nichts zu zeigen.

Induktionsschritt von n nach $n + 1$.

Gegeben seien also unabhängige Zufallsvariable $X_1, \dots, X_{n+1}$ und Teilmengen $A_1, \dots, A_{n+1} \subset \mathbb{R}$. Für jede Teilmenge $\{i_1, \dots, i_k\} \subset \{n+1\}$ mit $k \leq n$ gilt dann $(*)$ nach Induktionsvoraussetzung. Wir prüfen

$$
\begin{aligned}
&P(X_1 \in A_1, \dots, X_{n+1} \in A_{n+1}) \\
&= P\Big(\bigcap\nolimits_{i=1}^{n+1} X_i^{-1}(A_i)\Big) = P\Bigg(\bigcap\nolimits_{i=1}^{n+1}\Big(\coprod_{x_i \in A_i} X_i^{-1}(x_i)\Big)\Bigg) \\
&= P\Bigg(\coprod_{\underbrace{(x_1, \dots, x_{n+1})}_{=\underline{x}} \in \underbrace{A_1 \times \dots \times A_{n+1}}_{=\underline{A}}} \bigcap\nolimits_{i=1}^{n+1} X_i^{-1}(x_i)\Bigg) \\
&= \sum_{\underline{x} \in \underline{A}} P\Big(\bigcap\nolimits_{i=1}^{n+1} X_i^{-1}(x_i)\Big) = \sum_{\underline{x} \in \underline{A}} \prod\nolimits_{i=1}^{n+1} P(X_i^{-1}(x_i)) \\
&= \prod_{i=1}^{n+1}\Big(\sum_{x_i \in A_i} P(X_i^{-1}(x_i))\Big) = \prod\nolimits_{i=1}^{n+1} P\Big(\coprod_{x_i \in A_i} X_i^{-1}(x_i)\Big) = \prod\nolimits_{i=1}^{n+1} P(X_i^{-1}(A_i)).
\end{aligned}
$$

□

Satz 9.3 folgt nun, indem wir für ein gegebenes Tupel $(x_1, \dots, x_n) \in \mathbb{R}^n$ setzen: $A_1 = f_1^{-1}(x_1), \dots, A_n = f_n^{-1}(x_n)$.

10 Varianz und Covarianz

10.1 Beispiel: Beim Roulette-Spiel setzt Spieler A immer auf die 1 und Spieler B auf die Kolonne $\{1, 2, \dots, 12\}$. Was ist der Erwartungswert ihres Gewinns?

Hier: $\Omega = \{0, 1, 2, \dots, 36\}$ mit Zähldichte $f(x) = \frac{1}{37}$.

Spieler A**:** $X : \Omega \longrightarrow \mathbb{R}$, $X(1) = 35$, $X(x) = -1$ für $x \neq 1$

Spieler B**:** $Y : \Omega \longrightarrow \mathbb{R}$, $Y(x) = \begin{cases} 2 & 1 \leq x \leq 12 \\ -1 & \text{sonst} \end{cases}$

$$E(X) = 35 \cdot \frac{1}{37} - 1 \cdot \frac{36}{37} = -\frac{1}{37} \qquad E(Y) = 2 \cdot \frac{12}{37} - 1 \cdot \frac{25}{37} = -\frac{1}{37}$$

Die Erwartungswerte sind dieselben, obwohl die Werte von X und Y sehr verschieden über $\mathbb{R}$ verteilt sind.

Das Beispiel zeigt, daß die mittlere Abweichung vom Mittelwert $E(X)$ von Interesse sein kann. Sei $\mu = E(X)$. Wollen wir die mittlere Abweichung einschließlich ihres Vorzeichens finden, fragen wir nach

$$E(X-\mu) = E(X) - \mu = E(X) - E(X) = 0$$

(wegen (8.17)), ein wenig informatives Ergebnis. Wir könnten nach dem mittleren Abstand, also nach

$$E(|X-\mu|)$$

fragen. Das Arbeiten mit Absolutbeträgen kann aber beschwerlich sein. Daher fragt man nach dem Quadrat der Abweichung, also nach

$$E((X-\mu)^2).$$

Dieses hat z.Tl. auch andere innermathematische Gründe. Für uns wichtig ist zu erkennen, daß kleine Abweichungen (quadratisch) weniger berücksichtigt werden als große Abweichungen. Da wir diskrete Zufallsvariable über i.a. nicht endlichen W-Räumen betrachten, ist die Existenz von $E((X-\mu)^2)$ nicht immer gegeben. Wir definieren

10.2 Definition: Ist X eine Zufallsvariable auf einem diskreten W-Raum mit Erwartunswert μ, so daß $E(X^2)$ existiert, dann heißt

$$\mathrm{Var}(X) := E((X-\mu)^2)$$

die *Varianz* von X und

$$\sigma(X) := {}_{+}\sqrt{\mathrm{Var}(X)}$$

die *Streuung* oder *Standardabweichung* von X.
Ist $Y : \Omega \longrightarrow \mathbb{R}$ eine weitere Zufallsvariable, für die $E(Y^2)$ existiert, dann heißt

$$\mathrm{Cov}(X,Y) = E((X-\mu_X)\cdot(Y-\mu_Y))$$

mit $\mu_X = E(X)$ und $\mu_Y = E(Y)$ die *Kovarianz* und

$$\rho(X,Y) = \mathrm{Cov}(X,Y)/\sigma(X)\cdot\sigma(Y)$$

der *Korrelationskoeffizient* von X und Y.

10.3 Bemerkung: Die Voraussetzung „$E(X^2)$“ existiert, ist mit Bedacht gewählt: Für $x \in \mathbb{R}$ gilt: $|x| \leq 1 + x^2$, wie man sich leicht überlegt. Weiter gilt:

$$\begin{aligned}
E(X) \text{ existiert } &\iff \sum_{x\in\Omega_X} x\cdot P(X=x) \text{ konvergiert absolut}, \quad (\Omega_X = \text{Bild}\, X)\\
&\overset{def}{\iff} \sum_{x\in\Omega_X} |x|\cdot P(X=x) \text{ konvergiert}\\
&\overset{(8.16)}{\iff} E(|X|) \text{ existiert}
\end{aligned}$$

und

$$\begin{aligned} E(X^2) &\overset{8.16}{=} \sum_{x\in\Omega_X} x^2 \cdot P(X=x) \\ E(1+X^2) &\overset{8.16}{=} \sum_{x\in\Omega_X} (1+x^2)\cdot P(X=x) \overset{8.17}{=} 1+E(X^2). \end{aligned}$$

Damit existiert $E(1+X^2)$, und die zugehörige Reihe ist eine Majorante zur Reihe von $E(|X|)$. Also existieren auch $E(|X|)$ und $\mu_X = E(X)$ und damit auch $E(X-\mu_X) = EX - \mu_X$ (8.17).
Weiter gilt für $x, y \in \mathbb{R}$

$$|x\cdot y| \leq x^2 + y^2$$

denn $|x\cdot y| \leq 2|x|\cdot|y|$ und $x^2 - 2|x|\cdot|y| + y^2 = (|x|-|y|)^2 \geq 0$. Also existiert auch $E(X\cdot Y)$, da $E(|X|\cdot|Y|)$ existiert. Es hat $E(X^2+Y^2) = E(X^2)+E(Y^2)$ als Majorante. Damit existiert

$$\operatorname{Cov}(X,Y) = E(X\cdot Y - \mu_X\cdot Y - \mu_Y\cdot X + \mu_X\cdot\mu_Y)$$

und $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$.
Umgekehrt folgt aus der Existenz von $\operatorname{Var} X$, daß $E(X^2)$ existiert.

Folgende Formeln helfen bei der Berechnung der Varianz und Kovarianz. Einige davon haben wir bereits in der Bemerkung hergeleitet.

10.4 Rechenregeln: Sind $X, Y, X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ Zufallsvariable auf einem diskreten W-Raum, so daß $E(X^2), E(Y^2)$ und $E(X_i^2)$ existieren, dann gilt:

(1)

$$\begin{aligned} \operatorname{Var}(X) &= \textstyle\sum_{x\in\Omega_X} (x-E(X))^2 \cdot P(X=x) \qquad\qquad (\Omega_X = \operatorname{Bild} X) \\ &= E(X^2) - (E(X))^2 \\ &= \textstyle\sum_{x\in\Omega_X} x^2\cdot P(X=x) - (E(X))^2 \end{aligned}$$

(2) Für alle $a, b, c \in \mathbb{R}$ gilt

$$\operatorname{Var}(aX + bX + c) = a^2\cdot\operatorname{Var}(X) + b^2\operatorname{Var}(Y) + 2ab\cdot\operatorname{Cov}(X,Y)$$

(3)

$$\begin{aligned}\operatorname{Var}(X_1 + \cdots + X_n) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j)\\ &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{1\le i<j\le n} \operatorname{Cov}(X_i, X_j)\end{aligned}$$

(4)

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} (x - E(x))\cdot(y - E(Y))\cdot P(X = x, Y = y)\\ &= E(X\cdot Y) - E(X)\cdot E(Y) \qquad\qquad (\Omega_Y = \operatorname{Bild} Y)\\ &= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} x\cdot y\cdot P(X = x, Y = y) - E(X)\cdot E(Y)\end{aligned}$$

(5) $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$

(6)

$$\operatorname{Cov}(aX_1 + bX_2 + c, dY) = ad\operatorname{Cov}(X_1,Y) + bd\operatorname{Cov}(X_2,Y) \qquad \forall\ a,b,c,d\in\mathbb{R}$$

(7) $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$.

Beweis: (7) folgt sofort aus der Definition. Jetzt folgen (1) und (2) aus (4), (5) und (6), wie wir gleich sehen werden. Zur Erinnerung: $E(X) = \mu_X$.
(4) Sei $g : \mathbb{R}\times\mathbb{R} \longrightarrow \mathbb{R}$ die Abbildung $g(x,y) = (x-\mu_X)\cdot(y-\mu_Y)$. Aus (8.16) erhalten wir:

$$\begin{aligned}\operatorname{Cov}(X,Y) &= E((X-\mu_X)\cdot(Y-\mu_Y)) = E(g(x,y))\\ &= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} g(x,y)\cdot P(X = x, Y = y)\\ &= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} (x-\mu_X)\cdot(y = \mu_Y)\cdot P(X = x, Y = y).\end{aligned}$$

Weiter gilt: $E((X-\mu_X)\cdot(Y-\mu_Y)) = E(X\cdot Y - \mu_X\cdot Y - \mu_X\cdot Y + \mu_X\cdot\mu_Y)$

$\overset{8.17}{=} E(X\cdot Y) - \mu_X\cdot E(Y) - \mu_Y\cdot E(X) + E(X)\cdot E(Y) = E(X\cdot Y) - E(X)\cdot E(Y)$.

Wenden wir wieder (8.16) mit $g(x, y) = x \cdot y$ an, erhalten wir

$$E(X \cdot Y) = E(g(x, y)) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} x \cdot y \cdot P(X = x, Y = y)$$

(5) $\operatorname{Cov}(X, Y) = E((X - \mu_X) \cdot (Y - \mu_Y)) = E((Y - \mu_Y) \cdot (X - \mu_X)) = \operatorname{Cov}(Y, X)$

(6)
$$\begin{aligned}
&\operatorname{Cov}(aX_1 + b(X_2 + c, dY) \\
&= E((aX_1 + bX_2 + c) \cdot (dY)) - E(aX_1 + bX_2 + c) \cdot E(dY) \\
&= E(adX_1 \cdot Y + bdX_2 \cdot Y + cdY) - (aE(X_1) + bE(X_2) + c) \cdot d \cdot E(Y) \\
&= adE(X_1 \cdot Y) + bd \cdot E(X_2 \cdot Y) + cdE(Y) - adE(X_1)E(Y) \\
&\qquad\qquad\qquad\qquad\qquad\qquad - bdE(X_2) \cdot E(Y) - cdE(Y) \\
&= ad(E(X_1 \cdot Y) - E(X_1) \cdot E(Y)) + bd(E(X_2 \cdot Y) - E(X_2) \cdot E(Y)) \\
&= ad \cdot \operatorname{Cov}(X_1, Y) + bd \cdot \operatorname{Cov}(X_2, Y) \quad \text{(wir benutzten (8.17))}
\end{aligned}$$

(1) wird wie (4) bewiesen, wobei wir g durch die Abbildung

$$\begin{aligned}
&f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto (x - \mu_X)^2 && \text{für die erste Gleichung und} \\
&f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^2 && \text{für die zweite Gleichung setzen.}
\end{aligned}$$

(2) Setze $Z = aX + bY + c$. Dann gilt

$$\operatorname{Var}(aX + bY + c) = \operatorname{Cov}(aX + bY + c, Z) \overset{(6)}{=} a \cdot \operatorname{Cov}(X, Z) + b \cdot \operatorname{Cov}(Y, Z) \quad (*)$$

Nun gilt

$$\begin{aligned}
\operatorname{Cov}(X, Z) &\overset{(5)}{=} \operatorname{Cov}(Z, X) = \operatorname{Cov}(aX + bY + c, X) \\
&= a \cdot \operatorname{Cov}(X, X) + b \cdot \operatorname{Cov}(Y, X) = a \cdot \operatorname{Var}(X) + b \cdot \operatorname{Cov}(Y, X)
\end{aligned}$$

analog

$$\operatorname{Cov}(Y, Z) = a \cdot \operatorname{Cov}(X, Y) + b \cdot \operatorname{Cov}(Y, Y) = a \cdot \operatorname{Cov}(X, Y) + b \cdot \operatorname{Var}(Y).$$

Wir setzen in $(*)$ ein und erhalten das gewünschte Ergebnis.

(3) Die erste Gleichung folgt aus der zweiten: Wir fassen alle Summanden $\operatorname{Cov}(X_i X_i) = \operatorname{Var}(X_i)$ der linken Summe zur ersten Summe der rechten Seite zusammen. Es bleiben die Summanden $\operatorname{Cov}(X_i, X_j)$ mit $i \neq j$, etwa $i < j$. Dann gilt

$$\operatorname{Cov}(X_i, X_j) + \operatorname{Cov}(X_j, X_i) \underset{(5)}{=} 2 \operatorname{Cov}(X_i, X_j).$$

Also bilden diese Summanden die zweite Summe der rechten Seite. Die erste Gleichung beweisen wir induktiv.

Für $n = 1$ ist nichts zu zeigen.
Induktionsschritt von n nach $n+1$: Nach (10.4(2)) und Induktion gilt

$$\begin{aligned}
\operatorname{Var}(X_1 + \cdots + X_n + X_{n+1}) &= \operatorname{Var}(X_1 + \cdots + X_n) + \operatorname{Var}(X_{n+1}) \\
&\qquad + 2\operatorname{Cov}(X_1 + \cdots + X_n, X_{n+1}) \\
&= \sum_{i=1}^{n+1} \operatorname{Var}(X_i) + \underbrace{2 \sum_{1 \le i < j \le n} \operatorname{Cov}(X_i, X_j)}_{A} + \underbrace{2\operatorname{Cov}(X_1 + \cdots + X_n, X_{n+1})}_{B}.
\end{aligned}$$

Nun gilt

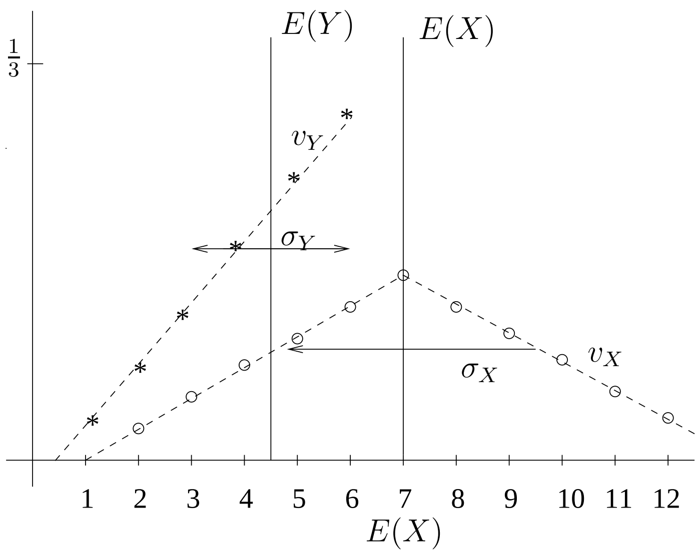
$$\begin{aligned}
&\operatorname{Cov}(X_1 + \cdots + X_n, X_{n+1}) \\
&= E((X_1 + \cdots + X_n) \cdot X_{n+1}) - E(X_1 + \cdots + X_n) \cdot E(X_{n+1}) \\
&= E(X_1 \cdot X_{n+1}) + \cdots + E(X_n \cdot X_{n+1}) - E(X_1) \cdot E(X_{n+1}) \\
&\qquad - \cdots - E(X_n)E(X_{n+1}) \\
&= \operatorname{Cov}(X_1, X_{n+1}) + \cdots + \operatorname{Cov}(X_n, X_{n+1}),
\end{aligned}$$

so daß

$$A + B = 2 \cdot \sum_{1 \le i < j \le n+1} \operatorname{Cov}(X_i, X_j).$$

□

10.5 Beispiel: Wir würfeln mit zwei idealen Würfeln und behandeln wieder die Zufallsvariablen X = „Augensumme“ und Y = „maximale Augenzahl“. Zunächst machen wir uns ein Bild über die Verteilungen



$$E(X) = 7 \qquad E(Y) = \frac{161}{36}$$

$$\begin{aligned}
\mathrm{Var}(X) &= 4 \cdot \frac{1}{36} + 9 \cdot \frac{2}{36} + 16 \cdot \frac{3}{36} + 25 \cdot \frac{4}{36} + 36 \cdot \frac{5}{36} + 49 \cdot \frac{6}{36} + 64 \cdot \frac{5}{36} \\
&\quad + 81 \cdot \frac{4}{36} + 100 \cdot \frac{3}{36} + 121 \cdot \frac{2}{36} + 144 \cdot \frac{1}{36} - 49 \\
&= \frac{4 + 18 + 48 + 100 + 180 + 294 + 320 + 324 + 300 + 242 + 144 - 1764}{36} \\
&= \frac{210}{36} = \frac{35}{6} = 5{,}833\ldots
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(Y) &= 1 \cdot \frac{1}{36} + 4 \cdot \frac{3}{36} + 9 \cdot \frac{5}{36} + 16 \cdot \frac{7}{36} + 25 \cdot \frac{9}{36} + 36 \cdot \frac{11}{36} - \left(\frac{161}{36}\right)^2 \\
&= \frac{1 + 12 + 45 + 112 + 225 + 396}{36} - \frac{161^2}{36^2} \\
&= \frac{791}{36} - \frac{161^2}{36^2} = \frac{2555}{1296} = 1{,}971\ldots
\end{aligned}$$

10.6 Bernoulli-Verteilung: Sei Ω der Elementarereignisraum eines Bernoulli-Experiments der Länge n und Wahrscheinlichkeit p für Treffer, d.h. $\Omega = \{N, T\}^n$. Sei X_k die Zufallsvariable aus Beispiel (8.18), so daß

$$X = X_1 + \cdots + X_n$$

die Zufallsvariable ist, deren Verteilung die Bernoulli-Verteilung ist. Es gilt

$$X_i \cdot X_j = \begin{cases} 1\ , & \text{falls } X_i = 1 \text{ und } X_j = 1 \\ 0\ , & \text{sonst} \end{cases} .$$

Also

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) = E(X_i \cdot X_j) - E(X_i) \cdot E(X_j) &= P(X_i = 1, X_j = 1) - p^2 \\
&= \begin{cases} p^2 - p^2\ , & i \neq j \\ p - p^2\ , & i = j \end{cases} ,
\end{aligned}$$

da X_i und X_j für $i \neq j$ voneinander unabhängig sind. Es folgt

$$\underline{\underline{\mathrm{Var}(X)}} = \mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = n(p - p^2) = n \cdot p(1-p) = \underline{\underline{n \cdot p \cdot q}}.$$

10.7 Hypergeometrische Verteilung: Wir gehen wie im Beispiel (8.20) vor und benutzen dieselben Bezeichnungen: Sei Ω die Menge aller denkbaren n Züge von roten und schwarzen Kugeln ohne Zurücklegen. Wir setzen

$$X_k(\omega) = \begin{cases} 1 & k\text{-te Kugel ist rot} \\ 0 & \text{sonst} \end{cases}.$$

und suchen $\mathrm{Var}(X)$, wobei $X = X_1 + \cdots + X_n$. Wir wissen

$$P(X_k = 1) = \frac{r}{M}, \qquad \text{wobei } M = r + s$$

$$X_i \cdot X_j = \begin{cases} 1 & i\text{-te und } j\text{-te Kugel ist rot} \\ 0 & \text{sonst} \end{cases}.$$

Sei nun $i \leq j$. Dann gilt

$$P(X_i = 1, X_j = 1) = \begin{cases} r/_M & i = j \\ \frac{r}{M} \cdot \frac{r-1}{M-1} & i \neq j \end{cases}$$

Es folgt nach (10.4) (3) und (8.20)

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \cdots + X_n) = \sum\nolimits_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j) \quad (*)$$

$$\mathrm{Var}(X_i) = 1 \cdot P(X_i = 1) - E(X_i)^2 = \frac{r}{M} - \left(\frac{r}{M}\right)^2$$

$$\mathrm{Cov}(X_i, X_j) = 1 \cdot P(X_i = 1, X_j = 1) - E(X_i) \cdot E(X_j) = \frac{r}{M} \cdot \frac{r-1}{M-1} - \left(\frac{r}{M}\right)^2 \text{ für } i \neq j.$$

Also

$$\begin{aligned} \underline{\underline{\mathrm{Var}\, X}} &= n\left(\frac{Mr - r^2}{M^2}\right) + 2\frac{n^2 - n}{2} \cdot \frac{Mr^2 - Mr - (M-1)r^2}{M^2 \cdot (M-1)} \\ &= \frac{n \cdot (M-1) \cdot r(M-r) + n(n-1) \cdot r(Mr - M - Mr + r)}{M^2 \cdot (M-1)} \\ &= \frac{n \cdot r \cdot (M-r) \cdot (M-1) + n \cdot r(1-n)(M-r)}{M^2(M-1)} \\ &= \underline{\underline{\frac{n \cdot r \cdot (M-r) \cdot (M-n)}{M^2 \cdot (M-1)}}}. \end{aligned}$$

Um den Unterschied zur Bernoulli-Verteilung deutlicher zu machen, setzen wir

$$p = \frac{r}{M},$$

d.h. p ist die Wahrscheinlichkeit dafür, eine rote Kugel zu ziehen. Sei wieder $q = 1 - p$. Dann gilt

$$\frac{r \cdot (M-r)}{M^2} = \frac{r}{M} \cdot \frac{M-r}{M} = \frac{r}{M} \cdot \left(1 - \frac{r}{M}\right) = p \cdot q.$$

Mit diesen Bezeichnungen erhalten wir

$$\mathrm{Var}(X) = n \cdot p \cdot q \left(\frac{M-n}{M-1}\right).$$

10.8 Geometrische Verteilung: Sei X die Zufallsvariable aus Beispiel 8.22. Wir suchen $\mathrm{Var}(X)$. In (8.22) haben wir benutzt, daß

$$\frac{1}{(1-x)^2} = \left(1 - \frac{1}{1-x}\right)' = \sum\nolimits_{k=1}^{\infty} k \cdot x^{k-1}.$$

Leiten wir nochmals ab, erhalten wir

$$\sum_{k=1}^{\infty} k \cdot (k-1) \cdot x^{k-2} = \left(\frac{1}{(1-x)^2}\right)' = \frac{2}{(1-x)^3}$$

$$\begin{aligned}\underline{\underline{\mathrm{Var}\, X}} &= E(X^2) - (E(X))^2 = \sum_{k=1}^{\infty} k^2 \cdot P(X = k) - \frac{1}{p^2} = \sum_{k=1}^{\infty} k^2 \cdot q^{k-1} \cdot p - \frac{1}{p^2} \\ &= \sum_{k=1}^{\infty} k(k-1) \cdot q^{k-1} \cdot p + \sum_{k=1}^{\infty} kq^{k-1}p - \frac{1}{p^2} = p \cdot q \frac{2}{(1-q)^3} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2pq}{p^3} + \frac{p-1}{p^2} = \frac{2q-q}{p^2} = \underline{\underline{\frac{q}{p^2}}}.\end{aligned}$$

Hier haben wir die Ergebnisse von (8.22) mit benutzt.

Bei der Berechnung der Bernoulli-Varianz erhielten wir $\mathrm{Cov}(X_i, X_j) = 0$ für $i \neq j$, im hypergeometrischen Fall dagegen war $\mathrm{Cov}(X_i, X_j) \neq 0$. Das kommt nicht von ungefähr:

10.9 Satz: (1) Sind X und Y unabhängige Zufallsvariable auf (Ω, P), so gilt

$$\mathrm{Cov}(X, Y) = 0, \text{ also } E(X \cdot Y) = E(X) \cdot E(Y).$$

(2) Sind $X_1, \ldots, X_n : \Omega \longrightarrow \mathbb{R}$ paarweise unabhängig, so gilt

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Beweis:

$$\begin{aligned}
E(X \cdot Y) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} x \cdot y \cdot P(X = x, Y = y) \\
&= \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} x \cdot y \cdot P(X = x) \cdot P(Y = y) \\
&= \sum_{x \in \Omega_X} x \cdot P(X = x) \cdot \sum_{y \in \Omega_Y} y \cdot P(Y = y) = E(X) \cdot E(Y).
\end{aligned}$$

Es folgt, $\operatorname{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0$.

Teil (2) folgt nun aus (10.4) (3). □

Die Umkehrung von (10.9) braucht nicht zu gelten: Aus $\operatorname{Cov}(X, Y) = 0$ folgt nicht, daß X und Y unabhängig sind.

10.10 Beispiel: Ein Spieler würfelt: Bei 5 oder 6 erhält er 1 DM, bei 1 oder 2 muß er 1 DM bezahlen. Hier ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ mit konstanter Zähldichte, d.h. der Würfel ist ideal. Das Spiel wird durch die Zufallsvariable $X : \Omega \longrightarrow \mathbb{R}$ beschrieben mit der Wertetabelle

ω	1	2	3	4	5	6
$X(\omega)$	-1	-1	0	0	1	1

und Verteilung

$X(\Omega)$	-1	0	1
v_X	$1/3$	$1/3$	$1/3$

.

Sei Y die Zufallsvariable $Y = X^2$. Dann ist $Y(\Omega) = \{0, 1\}$ und $X \cdot Y = X^3$. Aber $X^3 = X$ in unserem Fall. Die Verteilung von Y ist

$Y(\Omega)$	0	1
v_Y	$1/3$	$2/3$

.

Es folgt

$$\begin{aligned}
E(X) = E(X \cdot Y) = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0 \\
E(Y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}.
\end{aligned}$$

Also

$$\operatorname{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 0 - 0 = 0.$$

Da aber

$$P(X=1,Y=1) = P(\{5,6\} \cap \{1,2,5,6\}) = P(\{5,6\}) = \frac{1}{3}$$

und

$$P(X=1) \cdot P(Y=1) = P(\{5,6\}) \cdot P(\{1,2,5,6\}) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9},$$

sind X und Y abhängig. $\square$

Die Kovarianz ist in gewisser Hinsicht ein Maß für die Abhängigkeit, aber $\mathrm{Cov}(X,Y) = 0$ bedeutet noch nicht „Unabhängigkeit“. Statt dessen definiert man

10.11 Definition: Zwei Zufallsvariable heißen *nicht-korreliert*, wenn

$$\mathrm{Cov}(X,Y) = 0.$$

Der Korrelationskoeffizient war als $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$ eingeführt. Also ist nicht-korreliert gleichbedeutend mit

$$\rho(X,Y) = 0,$$

es sei denn σ_X oder σ_Y ist 0, so daß $\rho(X,Y)$ nicht definiert ist. Warum betrachtet man überhaupt $\rho(X,Y)$ und nicht nur $\mathrm{Cov}(X,Y)$?

Da $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X,Y)$, kann man durch einfache Manipulationen für die Kovarianz jeden Wert erzeugen, falls $\mathrm{Cov}(X,Y) \neq 0$. Damit ist sie als Maß für die Bindung zwischen X und Y ungeeignet. Dagegen gilt

10.12 Für $a, b \in \mathbb{R} \backslash \{0\}$ und $c, d \in \mathbb{R}$ gilt

$$|\rho(aX+c, bY+d)| = |\rho(X,Y)|.$$

Beweis: $\mathrm{Var}(aX+b) = \mathrm{Cov}(aX+b, aX+b) = a^e \cdot \mathrm{Cov}(X,X) = a^2 \cdot \mathrm{Var}(X)$. Also $\sigma_{aX+c} = |a| \cdot \sigma_X$. Analog gilt $\sigma_{bY+d} = |b| \cdot \sigma_Y$

$$\mathrm{Cov}(aX+c, bY+d) = ab \cdot \mathrm{Cov}(X,Y).$$

Es folgt

$$|\rho(aX+c, bY+d)| = \left| \frac{ab \cdot \mathrm{Cov}(X,Y)}{|a| \cdot |b| \cdot \sigma_X \cdot \sigma_Y} \right| = |\rho(X,Y)|.$$

$\square$

Als nächstes untersuchen wir, was passiert, falls $\sigma_X = 0$.

10.13 Bezeichnung: Sei (Ω, P) ein diskreter W-Raum mit Zähldichte $f : \Omega \longrightarrow \mathbb{R}$. Dann bezeichne

$$\Omega_w := \{\omega \in \Omega;\ f(\omega) \neq 0\} \subset \Omega$$

die *wesentliche* Teilmenge aller der Ereignisse, die tatsächlich auftreten können.

10.14 Da $P(\Omega \backslash \Omega_w) = 0$, genügt es für die Berechnungen von $E(X), \mathrm{Var}(X)$ und $\mathrm{Cov}(X, Y)$ die Einschränkungen der Zufallsvariable auf Ω_w zu betrachten.

10.15 Satz: Sei (Ω, P) ein diskreter W-Raum, und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ Zufallsvariable. Dann gilt

(1) $\sigma_X = 0 \Longleftrightarrow X|\Omega_w : \Omega_w \longrightarrow \mathbb{R}$ ist konstant.

(2) $\sigma_X = 0 \Longrightarrow X$ und Y sind unabhängig.

Beweis:

(1) Wir setzen wieder $E(X) = \mu$. Dann gilt mit $A := \mathrm{Bild}(X : \Omega_w \longrightarrow \mathbb{R})$

$$\sigma_X = 0 \Longleftrightarrow \mathrm{Var}(X) = \sum_{x \in \Omega_X} (x-\mu)^2 \cdot P(X = x) = \sum_{x \in A} (x-\mu)^2 \cdot P(X = x) = 0.$$

Nun gilt aber nach Definition von $\Omega_w : P(X = x) \neq 0 \Longleftrightarrow x \in A$. Also ist die rechte Summe genau dann 0, wenn $x = \mu$ für alle $x \in A$, d.h. ist $\omega \in \Omega_w$, muß gelten $X(\omega) = \mu$, so daß $X : \Omega_w \longrightarrow \mathbb{R}$ konstant ist.

(2) folgt aus (1) und (9.2).

□

Wir wollen jetzt genauer untersuchen, wofür $\rho(X, Y)$ ein Maß ist. Dazu zeigen wir zunächst

10.16 Satz: Seien X und Y Zufallsvariable, deren Varianz existiert. Dann gilt

(1) $\quad -\sqrt{E(X^2) \cdot E(Y^2)} \leq E(X \cdot Y) \leq \sqrt{E(X^2) \cdot E(Y^2)}$

(2) $\quad -1 \leq \rho(X, Y) \leq 1$

Beweis: Zur Abkürzung setzen wir $A := \sqrt{E(X^2)}$, $B := \sqrt{E(Y^2)}$. Es gilt

$$\begin{aligned}
0 &\leq E\left(\left(\frac{X}{A} \pm \frac{Y}{B}\right)^2\right) \\
&= E\left(\frac{X^2}{A^2} \pm 2\frac{XY}{AB} + \frac{Y^2}{B^2}\right) = \frac{1}{A^2} \cdot E(X^2) \pm 2\frac{E(XY)}{AB} + \frac{1}{B^2} \cdot E(Y^2) \\
&= 2 \pm 2\frac{E(XY)}{AB}
\end{aligned}$$

Es folgt

$$-1 \leq \pm\frac{E(XY)}{AB} \quad \text{also} \quad -AB \leq \pm E(XY)$$

und damit

$$-AB \leq E(XY) \leq AB$$

Das beweist (1). Setzen wir nun statt X die Zufallsvariable $X - E(X)$ und statt Y die Variable $Y - E(Y)$ ein, geht (1) über in

$$-\sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)} \leq \operatorname{Cov}(X,Y) \leq \sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)},$$

und (2) folgt. □

10.17 Satz: Sei (Ω, P) diskreter W-Raum, und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ zwei Zufallsvariable, so daß $\rho(X,Y) = \pm 1$. Dann gilt auf Ω_w

$$Y = g(X) : \Omega_w \longrightarrow \mathbb{R},$$

wobei g eine **lineare** Funktion ist: $g(x) = a\rho x + b$ mit

$$a = \frac{\sigma_Y}{\sigma_X} \text{ und } b = E(Y) - \rho\frac{\sigma_Y}{\sigma_X} \cdot E(X) = E(Y) - a\rho E(X).$$

Beweis: Können wir zeigen, daß

$$E((Y - g(X))^2) = 0,$$

so folgt wie im Beweis (10.15), daß $Y = g(X)$ auf Ω_w.
Nun gilt

$$\begin{aligned}
E((Y - g(x))^2) &= E((Y - \rho a X - E(Y) + \rho a E(X))^2) \\
&= E[((Y - E(Y)) - \rho a(X - E(X))^2] \\
&= E[(Y - E(Y))^2 - 2\rho a(Y - E(Y)) \cdot (X - E(X)) \\
&\qquad + \rho^2 a^2 (X - E(X))^2] \\
&= E((Y - E(Y))^2) - 2\rho a E((Y - E(Y)) \cdot (X - E(X)) \\
&\qquad + \rho^2 a^2 E((X - E(X))^2) \\
&= \operatorname{Var}(Y) - 2\rho a \operatorname{Cov}(X,Y) + \rho^2 a^2 \operatorname{Var}(X).
\end{aligned}$$

Da $\pm 1 = \rho = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}$ und $a = \frac{\sigma_Y}{\sigma_X}$, folgt

$$\begin{aligned} E((Y - g(X))^2) &= \sigma_Y^2 - 2\rho \frac{\sigma_Y}{\sigma_X} \cdot \rho \cdot \sigma_X \cdot \sigma_Y + \frac{\sigma_y^2}{\sigma_X^2} \cdot \sigma_X^2 \\ &= \sigma_Y^2 - 2\sigma_Y^2 + \sigma_Y^2 = 0. \end{aligned}$$

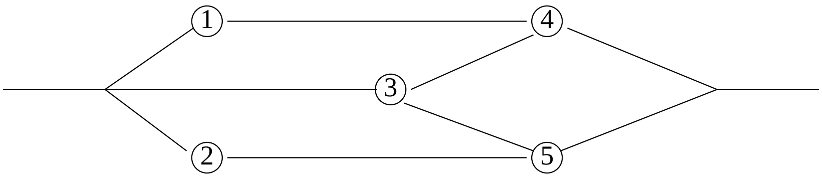
□

10.18 Bemerkung: Y ist vollkommen durch X bestimmt, falls Y eine Funktion von X ist, also $Y = g(X)$ für eine Funktion $g : \mathbb{R} \longrightarrow \mathbb{R}$. Die Korrelation mißt, in wie weit sich Y als **lineare** Funktion in X darstellen läßt.

11 Indikatorfunktionen

Probleme folgender Art treten im täglichen Leben oft auf. Obwohl wir sie mit unseren Mitteln prinzipiell lösen könnten, laufen wir schnell in Probleme:

11.1 Beispiel: Gegeben sei ein System von Generatoren 1,2,3 und Pumpen 4, 5, die für die Kühlung eines Atomkraftwerkes verantwortlich sind.



Generator 1 betreibt Pumpe 4 und Generator 2 betreibt Pumpe 5. Wir haben einen Notstromgenerator 3, der 4 oder 5 betreiben kann, falls 1 oder 2 ausfällt. Es genügt eine Pumpe, um den GAU zu verhindern.

Jedes der 5 Module kann unabhängig von den anderen defekt sein und den Durchfluß der Kühlflüssigkeit verhindern. Sei q_i die Wahrscheinlichkeit dafür, daß der i-te Modul defekt ist.

Gesucht ist die Wahrscheinlichkeit für den GAU.

Solche Probleme lassen sich leicht mit Indikatorfunktionen behandeln.

11.2 Definition: Sei (Ω, P) ein diskreter W-Raum und sei $A \subset \Omega$. Die *Indikatorfunktion* von A ist die Abbildung

$$1_A : \Omega \longrightarrow \mathbb{R}, \quad \omega \longmapsto \begin{cases} 1\,, & \text{falls } \omega \in A \\ 0\,, & \text{falls } \omega \notin A \end{cases}$$

Wir setzen außerdem $\mathbb{1} = 1_\Omega$.

11.3 Aufgaben: (1) $1_{\Omega\setminus A} = \mathbb{1} - 1_A$

(2) $1_{A\cap B} = 1_A \cdot 1_B$

(3) $(1_A)^2 = 1_A$

(4) $A \subset B \Longrightarrow 1_A \leq 1_B$

(5) $A \cap B = \emptyset \Longrightarrow 1_{A\cup B} = 1_A + 1_B$

(6) $A, B \subset \Omega \Longrightarrow 1_{A\cup B} = 1_A + 1_B - 1_A \cdot 1_B$

(7) $E(1_A) = P(A)$

Aus diesen elementaren Eigenschaften folgt leicht

11.4 Lemma: Sind $A_1, \ldots, A_n$ Teilmengen von Ω, so gilt

(1) $1_{A_1\cap A_2\cap\ldots\cap A_n} = 1_{A_1} \cdot 1_{A_2} \cdot \ldots \cdot 1_{A_n}$

(2) $1_{A_1\cup\ldots\cup A_n} = \mathbb{1} - (\mathbb{1} - 1_{A_1}) \cdot (\mathbb{1} - 1_{A_2}) \cdot \ldots \cdot (\mathbb{1}a1_{-A_n})$

Beweis:

(1) Induktion nach n.

Für $n = 1$ ist die Aussage richtig.

Induktionsschritt von n nach $n+1$: Sei $B = A_1 \cap \ldots \cap A_n$. Dann gilt

$$1_{A_1\cap\ldots\cap A_{n+1}} = 1_{B\cap A_{n+1}} = 1_B \cdot 1_{A_{n+1}} \underset{\substack{\uparrow\\ \text{Induktion}}}{=} (1_{A_1} \cdot \ldots \cdot 1_{A_n}) \cdot 1_{A_{n+1}}.$$

(2) Aus dem Grundkurs sollte die de Morgansche Formel bekannt sein

$$\Omega\setminus(A_1 \cup \ldots \cup A_n) = (\Omega\setminus A_1) \cap (\Omega\setminus A_2) \cap \ldots (\Omega\setminus A_n).$$

Es folgt

$$\begin{aligned} 1_{A_1\cup\ldots\cup A_n} = \mathbb{1} - (\mathbb{1} - 1_{A_1\cup\ldots\cup A_n}) &\overset{11.3}{=} \mathbb{1} - (1_{\Omega\setminus(A_1\cup\ldots\cup A_n)}) \\ &= \mathbb{1} - (1_{(\Omega\setminus A_1)\cap\ldots\cap A_n)}) \\ \overset{(1)}{=} \mathbb{1} - (1_{\Omega\setminus A_1} \cdot \ldots \cdot 1_{\Omega\setminus A_n}) &= \mathbb{1} - (\mathbb{1} - 1_{A_1}) \cdot \ldots \cdot (\mathbb{1} - 1_{A_n}). \end{aligned}$$

□

11.5 Lösung des Problems 11.1: Wir haben 4 Wege:

W_1 über	①,④	Sei A_i das Ereignis „W_i offen“
W_2 über	②,⑤	$A = A_1 \cup \ldots \cup A_4$
W_3 über	③,④	B_j das Ereignis, j hat keine Störung,
W_4 über	③,⑤	$P(B_j) = p_j = 1 - q_j$

Wir haben Probleme, wenn keiner der Wege offen ist. Gesucht wird also

$$q = P(\Omega \backslash A),$$

denn A ist das Ereignis „mindestens einer der Wege ist offen“. Beachte Ω ist die Menge der Zustnde des Systems. Nun gilt

$$1_{\Omega\backslash A} = \mathbb{1} - 1_A = (\mathbb{1} - 1_{A_1}) \cdot (\mathbb{1} - 1_{A_2}) \cdot (\mathbb{1} - 1_{A_3}) \cdot (\mathbb{1} - 1_{A_4}).$$

Weiter ist

$$A_1 = B_1 \cap B_4, \quad A_2 = B_2 \cap B_5, \quad A_3 = B_3 \cap B_4, \quad A_4 = B_3 \cap B_5.$$

Also $1_{A_1} = 1_{B_1} \cdot 1_{B_2},\ 1_{A_2} = 1_{B_1} \cdot 1_{B_5}$ usw., so daß

$$\begin{aligned}
&1_{\Omega\backslash A}\\
&= (\mathbb{1} - 1_{B_1} \cdot 1_{B_2}) \cdot (\mathbb{1} - 1_{B_2} \cdot 1_{B_5}) \cdot (\mathbb{1} - 1_{B_3} \cdot 1_{B_4}) \cdot (\mathbb{1} - 1_{B_3} \cdot 1_{B_5})\\
&= \mathbb{1} - 1_{B_1} \cdot 1_{B_2} - 1_{B_2} \cdot 1_{B_5} - 1_{B_3} \cdot 1_{B_4} - 1_{B_3} \cdot 1_{B_5} + 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_5}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_4}\\
&\quad + 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_5} + 1_{B_2} \cdot 1_{B_5} \cdot 1_{B_3} \cdot 1_{B_4} + 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_5} + 1_{B_3} \cdot 1_{B_4} \cdot 1_{B_5}\\
&\quad - 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_4} \cdot 1_{B_5} - 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_4} \cdot 1_{B_5} - 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_4} \cdot 1_{B_5}\\
&\quad + 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_4} \cdot 1_{B_5} - 1_{B_1} \cdot 1_{B_2} \cdot 1_{B_3} \cdot 1_{B_5}.
\end{aligned}$$

Wir wenden nun (11.3) (7) an und benutzen (8.17). Da die B_i voneinander unabhängig sind, gilt außerdem $E(B_i \cdot B_j) = E(B_i) \cdot E(B_j)$ für $i \neq j$ nach (10.9) (1)

$$\begin{aligned}
q = E(1_{\Omega\backslash A}) = 1 &- p_1p_2 - p_2p_5 - p_3p_4 - p_3p_5 + p_1p_2p_5 + p_1p_2p_3p_4 + p_2p_3p_5\\
&+ p_3p_4p_5 - p_1p_2p_3p_4p_5
\end{aligned}$$

12 Das schwache Gesetz der großen Zahl

Zu Beginn von § 10 hatten wir erläutert, warum wir die mittlere quadratische Abweichung vom Mittelwert statt die absolute betrachten. Das nächste Ergebnis stellt einen Zusammenhang zwischen diesen beiden her.

12.1 Tschebyscheff'sche Ungleichung
(Tschebyscheff 1821 – 1894, Petersburg):
Sei X eine diskrete Zufallsvariable, deren Erwartungswert μ und Varianz existieren. Sei $c > 0$ reelle Zahl. Dann gilt

$$P(|X-\mu| \geq c) \leq \frac{\operatorname{Var}(X)}{c^2}.$$

Beweis: Sei $B = \operatorname{Bild} X$ und $A = \{x \in B; |x-\mu| \geq c\}$. Wir wollen

$$P(|X-\mu| \geq c) = P(X^{-1}(A)) = P\left(\coprod_{x\in A} X^{-1}(x)\right) = \sum_{x\in A} P(X=x)$$

abschätzen. Nach (10.4) gilt

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{x\in B}(x-\mu)^2 \cdot P(X=x) = \sum_{x\in A}(x-\mu)^2 P(X=x) \\ &\qquad + \underbrace{\sum_{x\in B\setminus A}(x-\mu)^2 \cdot P(X=x)}_{\geq 0} \\ &\geq \sum_{x\in A}(x-\mu)^2 \cdot P(X=x) \\ &\underset{(*)}{\geq} \sum_{x\in A} c^2 \cdot P(X=x) = c^2 \cdot \sum_{x\in A} P(X=x) = c^2 \cdot P(|X-\mu| \geq c).\end{aligned}$$

Dabei gilt $(*)$, weil $(x-\mu)^2 \geq c \quad \forall\, x \in A$. □

In § 1 haben wir problematisiert, warum man den Grenzwert

$$\lim_{n\to\infty} r_n(A)$$

der relativen Häufigkeiten eines Ereignisses A nicht für den Aufbau der Wahrscheinlichkeitstheorie nehmen sollte. Die nächsten Ergebnisse zeigen, daß unser Ansatz der intuitiven Idee des Grenzwertes nicht widerspricht.

12.2 Bernoullis schwaches Gesetz der großen Zahl: Ein Bernoulli-Experiment für ein Ereignis A mit der Wahrscheinlichkeit $P(A)$ wird n-mal wiederholt. Sei $r_n(A)$ die Zufallsvariable, die die relative Häufigkeit für A angibt. Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n \to \infty} P(|r_n(A) - P(A)| \leq \varepsilon) = 1.$$

Erläuterung: Wir setzen wieder $P(A) = p$ und $q = 1 - p$. Unser Elementarereignisraum Ω_n ist die Menge aller n-Tupel von Symbolen T = Treffer und N = Niete. Sei $S_n(A)$ die Zufallsvariable, die jedem n-Tupel die Trefferzahl zuordnet. Dann ist

$$r_n(A) = \frac{S_n(A)}{n}.$$

Bernoullis Gesetz besagt nun, daß für große n die Wahrscheinlichkeit dafür sehr klein ist, daß die relative Häufigkeit von der Wahrscheinlichkeit p um mehr als ε abweicht. Mit anderen Worten: Mit **Wahrscheinlichkeit 1** geht $r_n(A)$ mit wachsendem n gegen p. (Ausrutscher können vorkommen, sind aber **sehr** selten.)

Wie bei der Untersuchung der Bernoulli- oder Binomialverteilung setzen wir jetzt

$$X_k : \Omega_n \longrightarrow \mathbb{R}, \quad X_k(n\text{-Tupel}) = \begin{cases} 1 & k\text{-te Koordinate ist } T \\ 0 & k\text{-te Koordinate ist } N \end{cases}$$

Dann gilt $S_n(A) = X_1 + \cdots + X_n;\ E(X_i) = p,\ \mathrm{Var}(S_n(A)) = n \cdot p \cdot q$ (vergl. (8.18), (10.6)). Also

$$\begin{aligned} r_n(A) &= \frac{1}{n}(X_1 + \cdots + X_n) \\ P(A) = p &= \frac{1}{n}(E(X_1) + \cdots + E(X_n)) \\ \mathrm{Var}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) &= \frac{1}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) = \frac{p \cdot q}{n} = \mathrm{Var}(r_n(A)). \end{aligned}$$

Damit folgt (12.2) aus folgendem allgemeineren Ergebnis:

12.3 Markoffs Gesetz der großen Zahlen: Sei $X_1, X_2, \ldots$ eine Folge beliebiger Zufallsvariabel $\Omega \longrightarrow \mathbb{R}$, deren Varianzen die Bedingung

$$\lim_{n \to \infty} \mathrm{Var}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = 0$$

erfüllen. Dann gilt für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - \frac{1}{n}(E(X_1) + \cdots + E(X_n))\right| \leq \varepsilon\right) = 1.$$

Beweis: Sei Y_n die Zufallsvariable $\frac{1}{n}(X_1 + \cdots + X_n)$. Dann gilt $E(Y_n) = \frac{1}{n}(E(X_1) + \cdots + E(X_n))$. Aus der Tschebycheff'schen Ungleichung erhalten wir

$$P\big(|Y_n - E(Y_n)| > \varepsilon\big) \leq \frac{\mathrm{Var}(Y_n)}{\varepsilon^2}$$

oder

$$P(|Y_n - E(Y_n)| \leq \varepsilon) > 1 - \frac{\mathrm{Var}(Y_n)}{\varepsilon^2}.$$

Es folgt

$$1 \geq \lim_{n\to\infty} P(|Y_n - E(Y_n)| \leq \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}\lim_{n\to\infty} \mathrm{Var}(Y_n) = 1.$$

$\square$

Teil III

Approximationen der Bernoulli-Verteilung

13 Die Poisson-Verteilung

Die Bernoulli-Verteilung hat zahlreiche Anwendungen, weil Versuchswiederholungen oft auftreten, ihr Nachteil ist aber, daß die Berechnung der Binomialkoeffizienten $\binom{n}{k}$ für große n und k sehr aufwendig ist. Für kleine Erfolgswahrscheinlichkeiten gibt es aber eine gute Approximation der Bernoulli-Verteilung, die sich außerdem noch gut berechnen läßt.

13.1 Definition: Seien $u, v : \mathbb{Z} \to \mathbb{R}$ zwei Zähldichten auf $\mathbb{Z}$. Dann heißt

$$u * v : \mathbb{Z} \to \mathbb{R}, \quad k \mapsto \sum_{i \in \mathbb{Z}} u(i) \cdot v(k-i)$$

die *Faltung* von u und v.

13.2 Lemma: $u * v$ ist eine Zähldichte auf $\mathbb{Z}$.

Beweis: $u * v(k) \geq 0$, weil $u(i) \geq 0$ und $v(k-i) \geq 0$.

$$\sum_{k \in \mathbb{Z}} (u * v)(k) = \sum_k \sum_i u(i) \cdot v(k-i) = \sum_i u(i) \cdot \underbrace{\sum_k v(k-i)}_{=1} = \sum_i (i) = 1$$

$\square$

13.3 Beispiel: Sei (Ω, P) ein diskreter W-Raum, und seien

$$X, Y : \Omega \to \mathbb{Z}$$

unabhängige Zufallsvariable. Sind v_X, v_Y, v_{X+Y} die Verteilungen von X, Y, bzw. $X + Y$, so gilt

$$v_X * v_Y = v_{X+Y}$$

Beweis:

$$\begin{aligned}
v_{X+Y}(k) &= P(X+Y=k) \\
&= P\left(\coprod_i (X^{-1}(i) \cap Y^{-1}(k-i))\right) && = \sum_i P(X=i, Y=k-i) \\
&= \sum_i P(X=i)P(Y=k-i) && = \sum_i v_X(i) \cdot v_Y(k-i) \\
& && = (v_X * v_Y)(k)
\end{aligned}$$

Wegen der Unabhängigkeit gilt $P(X=i, Y=k-i) = P(X=i) \cdot P(Y = k-i)$. □

13.4 Definition: Eine Zufallsvariable $X : \Omega \to \mathbb{N}$ heißt *Poisson-verteilt* mit Parameter $\lambda \leq 0$ (kurz: $P(\lambda)$-verteilt), wenn gilt

$$P(X=k) = e^{-\lambda} \cdot \lambda^k / k!$$

13.5 Lemma: Sind X_1, X_2 unabhängig und ist X_i $P(\lambda_i)$-verteilt, $i = 1, 2$, dann ist $X_1 + X_2$ $P(\lambda_1 + \lambda_2)$-verteilt

Beweis:

$$\begin{aligned}
P(X_1 + X_2 = k) &\overset{\text{Bspl.13.3}}{=} \sum_{i=0}^{k} e^{-\lambda_1} \cdot \frac{\lambda_1^i}{i!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_2^{k-i}}{(k-i)!} \cdot \frac{k!}{k!} \\
&= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \cdot \lambda_1^i \cdot \lambda_2^{k-i} \\
&= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{k!} \cdot \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda_1^i \cdot \lambda_2^{k-i} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!}
\end{aligned}$$

□

Der folgende Satz zeigt nun, daß die Poissonverteilung eine gute Approximation der Trefferzahl in einer Reihe von Experimenten mit kleiner Erfolgswahrscheinlichkeit ist.

13.6 Satz: Seien $X_1, \ldots, X_n$ paarweise unabhängige Zufallsvariable, die nur die Werte 0 (für Niete) oder 1 (für Treffer) annehmen. Sei $P(X_i = 1) = p_i$ und $P(X_i = 0) = q_i = 1 - p_i$, und sei $\lambda = p_1 + \ldots + p_n$. Sei $S = X_1 + X_2 + \ldots + X_n$. Dann gilt

$$\sum_{k=0}^{\infty} \left| P(S=k) - e^{-\lambda} \cdot \frac{\lambda^k}{k!} \right| \leq 2 \sum_{i=1}^{n} p_i^2$$

Beweis: Um die linke Seite abzuschätzen, benutzen wir einen kleinen Trick: Wir konstruieren einen neuen W-Raum:

$$\Omega = \{(w_1, \ldots, w_n) \in \mathbb{Z}^n; \quad w_i \geq -1 \quad \forall\, i = 1, \ldots, n\}$$

Die Zähldichte ist gegeben durch

$$P(w_1, \ldots, w_n) = P_1(w_1) \cdot P_2(w_2) \cdot \ldots \cdot P_n(w_n)$$

wobei

$$P_i(w_i) = \begin{cases} q_i = 1 - p_i & \text{falls } w_i = 0 \\ e^{-p_i} \cdot \frac{p_i^k}{k!} & \text{falls } w_i = k \geq 1 \\ e^{-p_i} - q_i & \text{falls } w_i = -1 \end{cases}$$

Dann gilt:

$$\sum_{k=-1}^{\infty} P_i(k) = e^{-p_i} - q_i + q_i + e^{-p_i} \sum_{k=1}^{\infty} \frac{p_i^k}{k!} = e^{-p_i} + e^{-p_i}(e^{p_i} - 1) = 1,$$

da $e^x = \sum\limits_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \sum\limits_{k=1}^{\infty} \frac{x^k}{k!}$. Es folgt

$$\begin{aligned} \sum_{w \in \Omega} P(w) &= \sum_{(w_1, \ldots, w_n) \in \Omega} P_1(w_1) \cdot \ldots \cdot P_n(w_n) \\ &= \left(\sum_{k=-1}^{\infty} P_1(w_1) \right) \cdot \left(\sum_{k=-1}^{\infty} P_2(w_2) \right) \cdot \ldots \cdot \left(\sum_{k=-1}^{\infty} P_n(w_n) \right) = 1. \end{aligned}$$

Also ist (Ω, P) tatsächlich ein W-Raum. Wir definieren Zufallsvariable

$$X_i, Y_i : \Omega \to \mathbb{R}, \quad i = 1, \ldots, n$$

durch

$$X_i(w_1, \ldots, w_n) = \begin{cases} 0 \ \text{falls } w_i = 0 \\ 1 \ \text{falls } w_i \neq 0 \end{cases} \quad Y_i(w_1, \ldots, w_n) = \begin{cases} k \ \text{falls } w_i = k \geq 1 \\ 0 \ \text{sonst} \end{cases}$$

Da

$$\begin{aligned} &P(Y_i = k) = P(w_i = k) = e^{-p_i} \cdot \tfrac{p_i^k}{k!} \quad \text{für } k \geq 1 \\ &P(Y_i = 0) = P(w_i = 0) + P(w_i = -1) = q_i + e^{-p_i} - q_i = e^{-p_i} \end{aligned}$$

ist Y_i $P(p_i)$-verteilt.

Weiter gilt

$$\begin{aligned} P(X_i = 1) &= P(w_i \neq 0) = e^{-p_i} - q_i + e^{-p_i} \sum_{k=1}^{\infty} \frac{p_i^k}{k!} = -q_i + e^{-p_i} \cdot \sum_{k=0}^{\infty} \frac{p_i^k}{k!} \\ &= 1 - q_i = p_i \end{aligned}$$

Also haben diese neuen X_i dieselbe Verteilung wie die im Satz gegebenen. Damit kann $S = X_1 + \ldots + X_n$ mit den neuen X_i berechnet werden. Nach Lemma 13.5 ist $T = Y_1 + \ldots + Y_n$ $P(p_1 + \ldots + p_n) = P(\lambda)$-verteilt. Es folgt

$$\begin{aligned} \sum_{k=0}^{\infty} \left| P(S=k) - e^{-\lambda} \frac{\lambda^k}{k!} \right| &= \sum_{k=0}^{\infty} \Big| P(S=k) - P(T=k) \Big| = (*) \\ &= \sum_{k=0}^{\infty} \Big| P(S=k, T=k) + P(S=k, T \neq k) - \\ &\qquad P(S=k, T=k) - P(S \neq k, T=k) \Big| \end{aligned}$$

Beachte hierbei: $P(S = k) = P(S = k, T = k) + P(S = k, T \neq k)$, weil $P(T = k) + P(T \neq k) = P(\Omega)$. Es folgt

$$\begin{aligned} (*) &= \sum_{k=0}^{\infty} \Big| P(S=k, T\neq k) - P(S \neq k, T=k) \Big| \\ &\leq \sum_{k=0}^{\infty} \Big(\Big| P(S=k, T \neq k) \Big| + \Big| P(S \neq k, T = k) \Big| \Big) = 2P(S \neq T) \end{aligned}$$

Wenn $S \neq T$ ist, ist sicherlich mindestens ein $X_i \neq Y_i$. Also folgt

$$2P(S \neq T) \leq 2 \sum_{i=1}^{\infty} P(X_i \neq Y_i) \leq 2 \sum_{i=1}^{\infty} p_i^2.$$

Letzteres, weil $P(X_i = Y_i) = P(X_i = Y_i = 0) + P(X_i = Y_i = 1) =$
$P_i(0) + P_i(1) = q_i + e^{-p_i} p_i$.
Also: $P(X_i \neq Y_i) = 1 - q_i - e^{-p_i} \cdot p_i = p_i(1 - e^{-p_i}) \leq p_i^2$
weil $e^{-x} \geq 1 - x$ für $x \geq 0$ (vgl. Funktionsgraphen). □

13.7 Folgerung: Ist $p(n)$ eine Folge mit $0 \leq p(n) \leq 1$ und $\lim_{n\to\infty} n \cdot p(n) = \lambda$, so gilt

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

Beweis: Wir setzen in (13.6) $p_i = p(n)$ für $i = 1, \ldots, n$. Dann ist

$$P(S = k) = b(k; n, p(n))$$

Mit $\lambda(n) = n \cdot p(n)$ folgt

$$0 \leq \left| b(k; n, p(n)) - e^{-\lambda(n)} \frac{\lambda(n)^k}{k!} \right| \leq 2 \cdot n \cdot p(n)^2.$$

Da $\lim\limits_{n\to\infty} \frac{\lambda}{n} = 0$, folgt die Aussage durch Übergang zum Grenzwert. $\square$

Die Folgerung 13.7 kann man auch ziemlich einfach direkt beweisen. Satz 13.6 hat aber den Vorteil, eine Abschätzung des Fehlers zwischen der Binomial- und der Poissonverteilung anzugeben.

In praktischen Anwendungen verwendet man die Poisson-Verteilung als Modell immer dann, wenn untersucht wird, wieviele von vielen möglichen, aber einzeln relativ unwahrscheinlichen unabhängigen Ereignissen eintreten.

13.8 Beispiel: Ein Wasserbehälter von $5m^3$ enthält 10^5 gefährliche Bakterien. Es wird eine Wasserprobe von $50cm^3$ entnommen. Wie groß ist die Wahrscheinlichkeit dafür, daß sie k gefährliche Bakterien enthält? Wie groß ist die Wahrscheinlichkeit, daß sie bakterienfrei ist?

Antwort: Die Wahrscheinlichkeit dafür, eine bestimmte Bakterie zu erwischen ist

$$p = \frac{50cm^3}{5 \cdot 100^3 cm^3} = 10^{-5}.$$

Nach(13.6) gilt mit $n = 10^5$ und $\lambda = 10^5 \cdot p = 1$

$$\sum_{k=0}^{\infty} \left| b(k; n, p) - e^{-\lambda} \frac{\lambda^k}{k!} \right| \leq 2 \cdot 10^5 (10^{-5})^2 = 2 \cdot 10^{-5}.$$

Bis auf einen Fehler $\leq \frac{2}{10^5}$ ist die Antwort also

$$e^{-1} \cdot \frac{1}{k!}.$$

Insbesondere ist die Wahrscheinlichkeit W dafür, daß die Probe bakterienfrei ist, durch die Ungleichung

$$0{,}36785 \leq \frac{1}{e} - 2 \cdot 10^{-5} \leq W \leq \frac{1}{e} + 2 \cdot 10^{-5} \leq 0{,}36790$$

näherungsweise bestimmt.

14 Die Normalverteilung, ein Bericht

In diesem Abschnitt möchte ich auf einen der zentralen Sätze der Wahrscheinlichkeitstheorie eingehen. Den Beweis dieses Satzes will ich aber nicht vorführen, da dafür die Vorlesung "Einführung in die Analysis I" Voraussetzung ist.

14.1 Definition: Eine Zufallsvariable X heißt *normiert*, wenn $E(X) = 0$ und $Var(X) = 1$.

14.2 Aufgabe: Sei X eine Zufallsvariable mit $Var(X) \neq 0$. Dann ist

$$X^* = \frac{1}{\sigma_X} \cdot (X - E(X))$$

normiert. Man nennt X^* die *normierte Form* von X.

Wir betrachten nun folgende Situation: Gegeben ist eine Folge $X_1, X_2, \ldots$ paarweise unabhängiger Zufallsvariabler, die alle **dieselbe** Verteilung besitzen. Wir bezeichnen ihren gemeinsamen Erwartungswert mit E und ihre gemeinsame Standardabweichung mit σ. Setzen wir

$$\begin{aligned} S_n &:= X_1 + \ldots + X_n \\ H_n &:= \tfrac{1}{n} \cdot S_n \end{aligned}$$

so gilt
$$\begin{aligned} E(H_n) &= \tfrac{1}{n}(E(X_1) + \ldots + E(X_n)) = E \\ Var(H_n) &= \tfrac{1}{n^2}(Var(X_1) + \ldots + Var(X_n)) = \tfrac{\sigma^2}{n} \end{aligned}$$

Da $Var(H_n)$ mit wachsendem n gegen 0 geht, wird nach der Tschebyscheff'schen Ungleichung die Verteilung von H_n mit wachsendem n mehr und mehr um E konzentriert. Das ist die Aussage des Gesetzes der großen Zahl.

Sei nun

14.3 $$Y_n := \text{ normierte Form von } H_n = \frac{S_n - nE}{\sigma \cdot \sqrt{n}} = \frac{\sqrt{n}(H_n - E)}{\sigma}$$

Eines der wichtigsten Resultate der Wahrscheinlichkeitstheorie ist nun, daß die Verteilung von Y_n mit wachsendem n gegen die sogenannte *Normalverteilung* konvergiert. Die genaue Aussage ist

14.4 Zentraler Grenzwertsatz: Ist $X_1, X_2, \ldots$ eine Folge paarweise unabhängiger Zufallsvariabler, die alle dieselbe Verteilung besitzen. Sei Y_n die standardisierte Summe, gegeben in (14.3). Dann existiert für jedes $t \in \mathbb{R}$

$$\Phi(t) := \lim_{n \to \infty} P(Y_n < t)$$

und es gilt

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

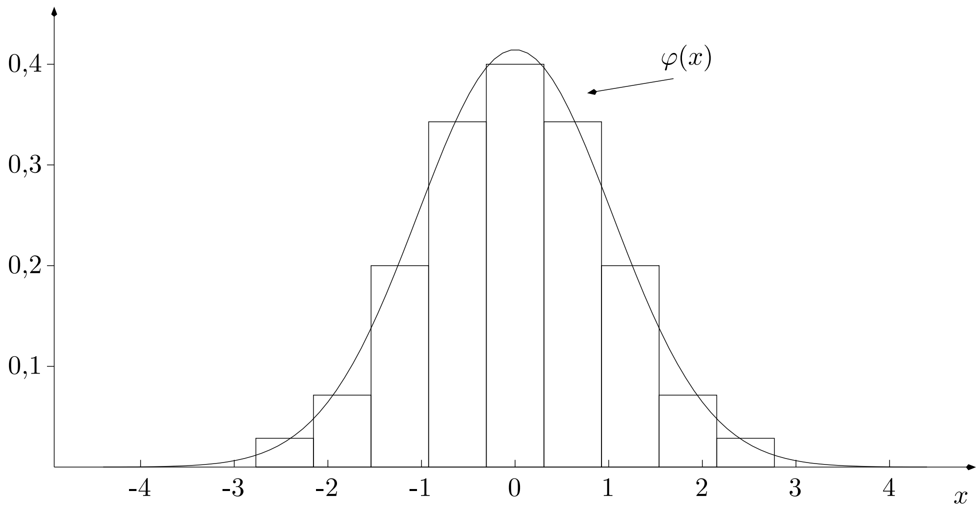
Das Resultat ist äußerst überraschend, weil $\Phi(t)$ nicht von der speziellen Verteilung der X_i abhängig ist.

Die Funktion

$$\varphi = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} : \mathbb{R} \to \mathbb{R}$$

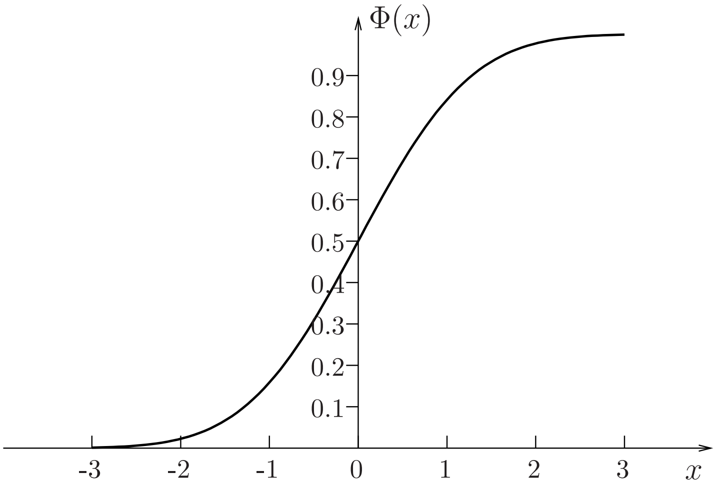
ist die bekannte *Glockenkurve*, die auch jeden 10 DM-Schein ziert.

Die folgenden beiden Skizzen zeigen die Approximation der Bernoulli-Verteilung durch φ und die Verteilungsfunktion Φ von φ.



Approximation der Binomialverteilung mit $p = \frac{2}{5}$ und $n = 10$ durch φ

14.5 Definition: φ nennt man *Normalverteilung*, Φ ist die zugehörige *Verteilungsfunktion* (vergl. (7.7), nur haben wir es jetzt mit Verteilungen auf nicht-diskreten W-Räumen zu tun).



14.6 Anwendungsbeispiel: Wir würfeln 600 mal mit einen idealen Würfel und fragen nach der Wahrscheinlichkeit, daß wir mindestens 90 und höchstens 100 Sechsen erhalten.
$\Omega = \{1, \ldots, 6\}^{600}$. Sei $X_i : \Omega \to \mathbb{R}$, die Zufallsvariable, gegeben durch

$$X_i(w) = \begin{cases} 1 & i-\text{ter Wurf ist } 6 \\ 0 & \text{sonst} \end{cases}$$

Dann gibt $S_{600} = X_1 + \ldots + X_{600}$ die Anzahl der Sechsen an. Gesucht wird

$$P(90 \leq S_n \leq 100) \quad n = 600$$

$E = E(X_i) = 1 \cdot \frac{1}{6}$, $Var(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{1}{6} - \frac{1}{36} = \frac{5}{36}$,
$\sigma = \sigma_{X_i} = \frac{1}{6}\sqrt{5}$. Es folgt

$$Y_n = \frac{S_n - 100}{\sqrt{600} \cdot \frac{1}{6}\sqrt{5}} = \frac{\sqrt{30}}{50}(S_n - 100)$$

$$\begin{aligned} P(90 \leq S_n \leq 100) &= P(-10\tfrac{\sqrt{30}}{50} \leq Y_n \leq 0) \\ &= P(Y_n \leq 0) - P(Y_n < -\tfrac{\sqrt{30}}{50}) \\ &\approx \Phi(0) - \Phi(-1,0954\ldots) = 0,5 - (1 - \Phi(1,0954\ldots)) \\ &\approx 0,36 \end{aligned}$$

Hierbei haben wir benutzt, daß wegen der Symmetrie $\varphi(x) = \varphi(-x)$ gilt

14.7 $\Phi(-x) = 1 - \Phi(x)$
Die gesuchte Wahrscheinlichkeit ist **näherungsweise** 36%.

Wie wir wissen, ist "näherungsweise" **kein** mathematischer Ausdruck, solange wir nicht wissen, wie groß der Fehler ist. Es gibt viele Untersuchungen über die Qualität der Approximation des zentralen Grenzwertsatzes. Das bedeutenste Resultat ist folgender Satz, der 1941 von Berry und unabhängig davon 1942 von Esseen mit ganz anderen Mitteln bewiesen wurde.

14.8 Satz von Berry und Esseen: Sei $X_1, X_2, \dots$ eine Folge paarweise unabhängiger Zufallsvariablen mit derselben Verteilung, so daß $\forall i \in \mathbb{N}\backslash\{0\}$ gilt

$$E(X_i) = 0, \quad E(X_i^2) = \sigma^2 > 0, \quad E(|X_i|^3) = \rho < \infty$$

Dann gilt für die Folge Y_n der normierten Summen (14.3)

$$\left|P(Y_n < t) - \Phi(t)\right| \leq \frac{C \cdot \rho}{\sigma^3 \cdot \sqrt{n}}$$

mit einer von X_i unabhängigen Konstanten C.
Über den Wert von C wurde viel gearbeitet. Berry berechnete sie mit $C \leq 1{,}88$, aber seine Rechnung war fehlerhaft. Esseen schätzte sie mit 7,59 ab. Das mir bekannte beste Resultat stammt von van Beek aus dem Jahr 1972 mit einem Wert $C \leq 0{,}7975$.

14.9 Beispiel: Wir wenden die Abschätzung auf Beispiel (14.6) an. Dafür müssen wir zunächst X_i durch die Zufallsvariable

$$Z_i = X_i - \frac{1}{6}, \quad \text{d.h. } Z_i(w) = \begin{cases} \frac{5}{6}, & \text{falls } i\text{-ter Wurf } = 6 \\ -\frac{1}{6}, & \text{falls } i\text{-ter Wurf } \neq 6 \end{cases}$$

ersetzen, damit $E(Z_i) = 0$. Dann gilt
$Var(Z_i) = Var(X_i)$, also $\sigma_{Z_i} = \sigma_{X_i} = \frac{1}{6}\sqrt{5}$. Insbesondere sind die normierten Summen Y_n dieselben für die Z_i und X_i. Wir berechnen ρ:

$$\rho = E(|Z_i|^3) = \left(\frac{5}{6}\right)^3 \frac{1}{6} + \left(\frac{1}{6}\right)^3 \frac{5}{6} = \frac{5^3 + 5}{6^4} = \frac{130}{6^4}$$

Es folgt: $\left|P(Y_n < t) - \Phi(t)\right| \leq \dfrac{0{,}7975 \cdot 130}{6^4 \cdot \frac{1}{6^3}\sqrt{5}^3 \cdot \sqrt{600}} = \dfrac{0{,}7975 \cdot 13}{6 \cdot 5 \cdot \sqrt{30}} = 0{,}06309\dots$

In der Herleitung unseres Ergebnisses haben wir die Abschätzung zweimal benutzt, so daß der mögliche Fehler $\leq \pm 12{,}6\%$ ist, eine relativ beachtliche Abweichung von den berechneten 36%.

Die Abschätzung von Berry und Esseen kann schon deshalb nicht optimal für einzelne Probleme sein, weil sie für **alle** Verteilungen gültig ist. Für spezielle Fälle gibt es erheblich bessere Resultate.

14.10 Der Fall der Bernoulli-Verteilung

Wir untersuchen ein Bernoulli-Experiment mit der Wahrscheinlichkeit p für Treffer und $q = 1 - p$ für Niete. Sei X_i die Zufallsvariable mit Wert 1, falls im i-ten Versuch T eintritt und sonst 0.

14.11 Satz: In diesem Fall gilt

$$P(a \leq S_n \leq b) = \left(1 + \frac{\Theta_1 \cdot c}{\sqrt{npq}}\right) \cdot \left(1 + \Theta_2 \frac{1 + 2c(c^2+3)\sqrt{npq}}{3npq}\right) \cdot \int\limits_A^B \varphi(t)dt$$

wobei

(1) $|\Theta_i| < 1$, für $i = 1, 2$

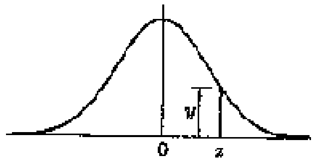
(2) $A = \dfrac{a - np}{\sqrt{npq}}, \quad B = \dfrac{b - np}{\sqrt{npq}}, \qquad c = \max\{|A|, |B|\}.$

In unserem Beispiel liefert diese Abschätzung keine besseren Werte.

TABLE 2
COMPARISON OF THE BINOMIAL DISTRIBUTION FOR $n = 100$, $p = 0.3$ AND THE NORMAL APPROXIMATION

Number of successes	Probability	Normal approximation	Percentage error
$9 \leq S_n \leq 11$	0.000 006	0.000 03	+400
$12 \leq S_n \leq 14$	0.000 15	0.000 33	+100
$15 \leq S_n \leq 17$	0.002 01	0.002 83	+40
$18 \leq S_n \leq 20$	0.014 30	0.015 99	+12
$21 \leq S_n \leq 23$	0.059 07	0.058 95	0
$24 \leq S_n \leq 26$	0.148 87	0.144 47	−3
$27 \leq S_n \leq 29$	0.237 94	0.234 05	−2
$31 \leq S_n \leq 33$	0.230 13	0.234 05	+2
$34 \leq S_n \leq 36$	0.140 86	0.144 47	+3
$37 \leq S_n \leq 39$	0.058 89	0.058 95	0
$40 \leq S_n \leq 42$	0.017 02	0.015 99	−6
$43 \leq S_n \leq 45$	0.003 43	0.002 83	−18
$46 \leq S_n \leq 48$	0.000 49	0.000 33	−33
$49 \leq S_n \leq 51$	0.000 05	0.000 03	−40

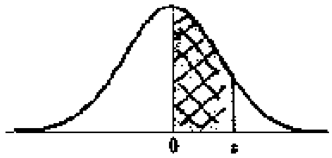
Ordinates (y) of the Standard Normal Curve at z



z	0	1	2	3	4	5	6	7	8	9
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

Tabelle 1: Tafel für $\varphi = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

Areas under the Standard Normal Curve from 0 to z



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

Tabelle 2: Tafel für Φ genauer für $H(z) := \int_0^z \varphi(t)dt$

Teil IV

Elemente der Schätztheorie

15 Schätzprobleme

Schätzprobleme treten in vielfältiger Gestalt auf. Wir erinnern an die Übungsaufgabe, einen Fischbestand mit Hilfe der hypergeometrische Verteilung und Entnahme einer Probe zu schätzen. Ein häufig auftretendes Problem ist das Schätzen der Trefferwahrscheinlichkeit bei einem Bernoulli-Experiment.

Wir beschreiben zunächst den mathematischen Rahmen.

15.1 Definition: Ein *Schätzproblem* mit diskretem Stichprobenraum besteht aus

- einer nicht-leeren höchstens abzählbar großen Menge $\mathcal{X}$ von *Stichproben*
- einer beliebigen Menge $\{P_\vartheta; \vartheta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$
- einer zu schätzenden Funktion $g : \Theta \to \mathbb{R}$.

15.2 Beispiel: Fischbestand

Sei N die Zahl der Fische im Teich, die zu schätzen ist. Man fängt r Fische, kennzeichnet sie und setzt sie wieder aus. Danach fängt man n Fische. Aus der Anzahl x der markierten darunter will man auf N schließen.

In diesem Fall ist $\mathcal{X} = \{0, 1, \ldots, n\}$. Dabei steht $x \in \mathcal{X}$ für die Anzahl der markierten Fische in der Stichprobe $\Theta = \{n, n+1, n+2, \ldots\}$. Für $n \in \Theta$ ist $P_N = h(-; n, r, N-r)$ die hypergeometrische Verteilung, und g ist die Abbildung

$$g : \Theta \to \mathbb{N}, \quad N \to N$$

15.3 Bezeichnung: Jede Abbildung $\hat{g} : \mathcal{X} \to \mathbb{R}$ nennt man einen *Schätzer* von g und $\hat{g}(x)$ den Schätzwert von g aufgrund der Stichprobe $x \in \mathcal{X}$.

15.4 Beispiel: In einer Stadt gibt es N Taxis, die mit Nummern $1, 2, \ldots, N$ sichtbar beschriftet sind. Ein Passant steht an einer belebten Straße und notiert die Nummern von n verschiedenen Taxen $x_1 < x_2 < \ldots < x_n$. Aus diesen Zahlen versucht er, auf N zu schließen.

$$\mathcal{X} = \text{ Menge aller } n\text{-elementigen Teilmengen von } \{1, 2, 3, \ldots, N\}.$$

Die Taxen haben alle denselben festen Standort, so daß sie mit gleicher Wahrscheinlichkeit durch die Straßen kommen. Damit ist

$$P_N(x) = \frac{1}{\binom{N}{n}} \qquad P_\vartheta = P_N,\ N \in \Theta$$

Jetzt ist $\Theta = \{x_n, x_n + 1, \ldots\}$, da $N \geq x_n$ und $g : \Theta \to \mathbb{N}$, $N \mapsto N$.

Ansätze für Schätzer:

(1) $\hat{g}(x) = x_n \quad x = \{x_1 < x_2 < \ldots < x_n\}$

(2) Aus Symmetriegründen dürfte es "im Durchschnitt" genausoviele Taxen mit Nummern $< x_1$ wie Taxen mit Nummern $> x_n$ geben. Daher kann man den Schätzer

$$\hat{g}_1(x) = x_n + x_1 - 1$$

diskutieren.

(3) Man könnte auch die Länge der Beobachtungslücke $\{x_n + 1, \ldots, N\}$ durch die mittlere Länge aller Lücken abschätzen:

$$\frac{(x_1 - 1) + (x_2 - x_1 - 1) + (x_3 - x_2 - 1) + \ldots + (x_n - x_{n-1} - 1)}{n} = \frac{x_n - n}{n}$$

Dies führt zu $\hat{g}_2(x) = x_n + \frac{x_n - n}{n} = \frac{(n+1)x_n - n}{n}$
(bzw. der nächsten ganzen Zahl).

Es gibt also eine Reihe heuristischer Argumente für Schätzer.

Bei der Lösung des Beispiels 15.2 wurde als Schätzverfahren die *Maximum-Likelihood-Schätzung* empfohlen, auf die wir jetzt eingehen werden.

15.5 Maximum-Likelihood-Schätzer: Für $x \in \mathcal{X}$ nennt man

$$L_x : \Theta \to \mathbb{R} \quad \vartheta \mapsto P_\vartheta(x)$$

die *Likelihood-Funktion*. Sei $\hat{\vartheta}(x) \in \Theta$ ein Element, für das

$$L_x(\hat{\vartheta}(x)) = \sup\{L_x(\vartheta); \vartheta \in \Theta\},$$

d.h. L_x hat in $\hat{\vartheta}(x)$ seinen maximalen Wert, dann heißt $\hat{\vartheta}(x)$ *Maximum-Likelihood-Schätzung* von ϑ.

15.6 Bemerkung: $\hat{\vartheta}(x)$ braucht nicht zu existieren. Es kann aber auch mehrere Maximum-Likelihood-Schätzer geben.

15.7 Beispiel: Wir ermitteln die Maximum-Likelihood-Schätzung im Beispiel (15.2). Sei x = Zahl der markierten Fische. Dann ist die Likelihood-Funktion

$$L_x : \Theta \to \mathbb{R}, \quad N \mapsto P_N(x) = h(x; n, r, N-r).$$

Wir suchen das Maximum:

$$\begin{aligned}
\frac{P_N(x)}{P_{N-1}(x)} &= \frac{\binom{r}{x} \cdot \binom{N-r}{n-x} \cdot \binom{N-1}{n}}{\binom{N}{n} \cdot \binom{r}{x}\binom{N-1-r}{n-x}} \\
&= \frac{(N-r)! \cdot (N-1)!n!(N-n)!(n-x)!(N-1-r-n+x)!}{(n-x)!(N-r-n+x)!n!(N-1-n)! \cdot N!(N-1-r)!} \\
&= \frac{(N-n)\cdot(N-r)}{N\cdot(N-r-n+x)}
\end{aligned}$$

$$\begin{aligned}
\text{Also: } P_N(x) > P_{N-1}(x) &\iff (N-n)\cdot(N-r) > N(N-r-n+x) \\
&\iff N^2 - Nr - nN + nr > N^2 - Nr - Nn + Nx \\
&\iff nr > Nx
\end{aligned}$$

Die gleich Äquivalenz erhält man, wenn man $>$ durch $\geq, =, <, \leq$ ersetzt. Es folgt: $P_N(x)$ ist für $\hat{N}(x) = \left[\frac{nr}{x}\right] =$ größte Zahl $\leq \frac{nr}{x}$ maximal. Ist $\frac{nr}{x}$ keine ganze Zahl, ist $\hat{N}(x)$ das eindeutige Maximum. Ist $\frac{nr}{x}$ ganze Zahl, erhalten wir $P_N(x) = P_{N-1}(x)$ für $nr = Nx$. Damit tritt das Maximum bei $N = \frac{nr}{x}$ und $N - 1 = \frac{nr}{x} - 1$ auf.

Wie vergleichen das mit der Lösung der Übungsaufgabe. Dort war $r = 50$, $n = 45$, $x = 10$. Die Maximum-Likelihood-Schätzung gibt den Schätzwert

$$\hat{N}(x) = \frac{45 \cdot 50}{10} = 225 \text{ oder } \frac{45 \cdot 50}{10} - 1 = 224$$

Das ist dasselbe Ergebnis wie in unserer Übungsaufgabe.

15.8 Beispiel: Maximum-Likelihood-Schätzung im Taxiproblem (15.4). Likelilhood-Funktion: Sei $x = \{x_1 < \ldots < x_n\}$

$$L_x : \Theta \to \mathbb{R}, \quad N \mapsto P_N(x) = \binom{N}{n}^{-1}$$

$P_N(x)$ ist also umso größer, je kleiner N ist. Das kleinste n ist aber x_n. Also ist $\hat{N}(x) = x_n$ der Maximum-Likelihood-Schätzer.

15.9 Beispiel: Bei einem Bernoulli-Experiment soll die Wahrscheinlichkeit p für "Treffer" geschätzt werden. Dazu führen wir das Experiment n-mal durch. In diesem Fall ist $\mathcal{X} = \{0, 1, \ldots, n\}$ der Stichprobenraum, wobei k die Anzahl der Treffer der Stichprobe ist, Θ ist das Einheitsintervall $[0, 1]$,

$$P_\vartheta = b(-; n, \vartheta) \qquad \vartheta \in [0, 1] = \Theta$$

ist die Bernoulli-Verteilung und

$$g : \Theta \to \mathbb{R}, \quad \vartheta \mapsto \vartheta$$

die Schätzfunktion. Sei k die Trefferzahl der Stichprobe. Wir haben die Likelihood-Funktion

$$L_k : \Theta \to \mathbb{R}, \quad p \mapsto P_p(k) = b(k; n, p) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}.$$

Gesucht wird das Maximum. Dazu differenzieren wir nach p

$$\begin{aligned} L_k(p)' &= \tbinom{n}{k} \left(k \cdot p^{k-1} \cdot (1-p)^{n-k} - (n-k)(1-p)^{n-k-1} \cdot p^k\right) \\ L_k(p)' &= 0 \iff k \cdot p^{k-1} \cdot (1-p)^{n-k} = (n-k) \cdot (1-p)^{n-k-1} \cdot p^k \end{aligned}$$

Wir wollen zunächst annehmen, daß $p \neq 0$ und $p \neq 1$. Dann folgt

$$L_k(p)' = 0 \iff k \cdot (1-p) = (n-k) \cdot p \iff k = n \cdot p.$$

Wir erhalten als Maximum-Likelihood-Schätzung: $\underline{\underline{\hat{p}(k) = \frac{k}{n}}}$

(man kontrolliert leicht nach, daß bei $\frac{k}{n}$ tatsächlich ein Maximum vorliegt). Ist nun $p = 0$, ist $k = 0$ und $\hat{p}(k) = 0$, ist $p = 1$, ist $k = n$, also auch $\hat{p}(k) = 1$. Damit passen auch diese Fälle zu unserem Schätzwert.

Man kann sich die Herleitung des letzten Ergebnisses einfacher machen. Da die Logarithmusfunktion streng monoton wachsend ist, kann man genausogut den Maximalwert der log-*Likelihood-Funktion* berechnen. Wir setzen

15.10 Bezeichnung: Ist $L_x : \Theta \to \mathbb{R}$ die Likelihood-Funktion, so heißt

$$\mathcal{L}_x := \ln L_x : \Theta \to \mathbb{R}$$

die log-*Likelihood-Funktion*.

15.11 Im Beispiel 15.9 gilt

(i) $\mathcal{L}_k(p) = \ln L_k(p) = \ln \binom{n}{k} + k \cdot \ln p + (n-k) \cdot \ln(1-p)$

(ii) $\mathcal{L}'_k(p) = \frac{1}{L_k(p)} \cdot L'_k(p) = \frac{k}{p} - \frac{n-k}{1-p} \overset{(*)}{=} \frac{k}{pq} - \frac{n}{q}$

(iii) $L'_k(p) = \left(\frac{k}{p} - \frac{n-k}{1-p}\right) \cdot L_k(p)$

zu $(*)$: $= \frac{k}{p} - \frac{n-k}{q} = \frac{k}{p} + \frac{k}{q} - \frac{n}{q} = \frac{qk+pk}{pq} - \frac{n}{q} = \frac{k}{pq} - \frac{n}{q}$

16 Erwartungstreue

Zur Erinnerung: Der Stichprobenraum ersetzt bei Schätzproblemen den W-Raum. Ist

$$T : \mathcal{X} \to \mathbb{R}$$

eine Abbildung, so können wir ihren Erwartungswert bzgl. der W-Maße $P_\vartheta, \vartheta \in \Theta$ ermitteln:

$$E_\vartheta(T) = \sum_{x \in \mathcal{X}} T(x) \cdot P_\vartheta(x) = \sum_{t \in \text{ Bild } T} t \cdot P_\vartheta(T = t).$$

16.1 Definition: Ein Schätzer $\hat{g} : \mathcal{X} \to \mathbb{R}$ von $g : \Theta \to \mathbb{R}$ heißt *erwartungstreu*, wenn für alle $\vartheta \in \Theta$ gilt

$$E_\vartheta(\hat{g}) = g(\vartheta).$$

Ist $\Theta \subseteq \mathbb{R}$, so heißt $\hat{\vartheta} : \mathcal{X} \to \mathbb{R}$ *erwartungstreuer Schätzer von* ϑ, wenn

$$\vartheta = E_\vartheta(\hat{\vartheta}) \quad \forall \vartheta \in \Theta$$

$b(\vartheta, \hat{g}) = E_\vartheta(\hat{g}) - g(\vartheta)$ heißt *Bias* der Schätzung von $\hat{g}$.

16.2 Beispiel: Im Beispiel (15.9) haben wir ein Bernoulli-Experiment der Länge n mit Trefferwahrscheinlichkeit p, $\mathcal{X} = \{0, 1, \ldots, n\}$. Wir zeigten

$$\hat{p} = \mathcal{X} \to \mathbb{R}, \quad k \mapsto \frac{k}{n}$$

ist der Maximum-Likelihood-Schätzer für

$$g : \Theta \to \mathbb{R}, \quad p \mapsto p.$$

Es gilt

$$E_p(\hat{p}) = \sum_{k=0}^{n} \frac{k}{n} \cdot b(k; n, p) = \frac{1}{n} \sum_{k=0}^{n} k \cdot b(k; n, p) = \frac{1}{n} \cdot n \cdot p = p.$$

Also ist $\hat{p}$ erwartungstreuer Schätzer von $p \in \Theta$.

16.3 Beispiel: Taxiproblem

(1) Der Maximum-Likelihood-Schätzer $\hat{g} : \mathcal{X} \to \mathbb{R}$, $(x_1 < \ldots < x_n) \mapsto x_n$ ist nicht erwartungstreu:

Sei $N \in \Theta = \{x_n, x_{n+1}, \ldots\}$. P_N ist die Gleichverteilung, aber

$$E_n(\hat{g}) = \binom{N}{n}^{-1} \cdot \sum_{x \in \mathcal{X}} \hat{g}(x) \neq N \text{ für einige } N \in \Theta.$$

Man prüft leicht nach, daß $E_N(\hat{g})$ nicht immer gleich N ist, z.B. falls $N = x_n = n + 1$.

(2) Dagegen sind die Schätzer $\hat{g}_1$ und $\hat{g}_2$ erwartungstreu. Auf den Beweis dieser Tatsache wollen wir verzichten.

16.4 Messungen einer Größe Eine Größe μ soll in n unabhängigen Messungen "ermittelt" werden. Seien $X_1, \ldots, X_n$ die unabhängigen Zufallsvariablen, die die Messungen beschreiben. Ihre Werte "streuen" um μ, aber ihre von μ abhängige Verteilung ist unbekannt. Wir nehmen nur an, daß

$$E_\mu(X_i) = \mu.$$

Sei

$$\overline{X} = \frac{1}{n}(X_1, + \ldots + X_n)$$

der **Mittelwert** unserer Messungen. Dann gilt

$$E_\mu(\overline{X}) = \frac{1}{n}(E_\mu(X_1) + \ldots + E_\mu(X_n)) = \mu.$$

Damit ist der Mittelwert eine erwartungstreu Schätzung von μ. Wir nehmen weiter an, daß die X_i die unbekannte Varianz σ^2 haben. Sei P_ϑ die unbekannte Verteilung der X_i. Dann gilt

$$E_\mu(X_i) = \mu =: g_1(\vartheta), \quad \sigma^2 = \mathrm{Var}_\vartheta(X_i) =: g_2(\vartheta).$$

16.5 Behauptung: $s^2 := \frac{1}{n-1} \sum\limits_{i=1}^{n} (X_i - \overline{X})^2$ ist erwartungstreue Schätzung von σ^2.

Beweis: Da die X_i voneinander unabhängig sind, gilt

$$\begin{aligned}
&E_\vartheta((X_i - \overline{X})^2) \\
&= E_\vartheta(((X_i - \mu) - (\overline{X} - \mu))^2) \\
&= \underbrace{E_\vartheta((X_i - \mu)^2)}_{\mathrm{Var}_\vartheta(X_i)=\sigma^2} - 2E_\vartheta((X_i - \mu)\cdot(\overline{X} - \mu)) + \underbrace{E_\vartheta((\overline{X} - \mu)^2)}_{\mathrm{Var}(\overline{X})=\frac{1}{n^2}\cdot n\cdot \mathrm{Var}(X_n)=\frac{\sigma^2}{n}}
\end{aligned}$$

$$\begin{aligned}
&E_\vartheta((X_i - \mu) - (\overline{X} - \mu)) \\
&\qquad = \mathrm{Cov}(X_i, \overline{X}) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = \frac{1}{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}
\end{aligned}$$

Also: $E_\vartheta((X_i - \overline{X})^2) = \sigma^2 - \frac{2}{n}\sigma^2 + \frac{\sigma^2}{n} = \sigma^2(1 - \frac{1}{n})$
und $E_\vartheta(s^2) = \frac{1}{n-1}\cdot n \cdot \sigma^2 \cdot \frac{n-1}{n} = \sigma^2$ □

16.6 Bemerkung: Die Erwartungstreue erzwingt den Faktor $\frac{1}{n-1}$ statt des naheliegenden Faktors $\frac{1}{n}$ für die Schätzung von s^2 von σ^2.

16.7 Folgerung: Da E_ϑ eine linear Funktion ist, folgt:

$$\frac{s^2}{n} \text{ ist erwartungstreue Schätzung der Varianz } \mathrm{Var}_\vartheta(\overline{X}) = \frac{\sigma^2}{n}$$

$$E_\vartheta(\frac{s^2}{n}) = \frac{1}{n}E_\vartheta(s^2) = \frac{1}{n}\sigma^2 = \mathrm{Var}(\overline{X}).$$

17 Der mittlere quadratische Fehler

Die wichtigste Forderung an einen Schätzer $T : \mathcal{X} \to \mathbb{R}$ ist sicherlich, daß die Schätzwerte $T(x)$ nahe bei der zu schätzenden Größe $g(\vartheta)$ liegen.

17.1 Definition: Sei $T : \mathcal{X} \to \mathbb{R}$ ein Schätzer für $g : \Theta \to \mathbb{R}$. Dann heißt

$$R(\vartheta, T) = E_\vartheta((T - g(\vartheta))^2)$$

der *mittlere quadratische Fehler* der Schätzung.

Diese Definition ist natürlich in Anlehnung an die Definition der Varianz gewählt. Man wird Schätzer suchen, für die $R(\vartheta, T)$ möglichst klein ist. Sind T_1 und T_2 zwei Schätzer, so kann es durchaus vorkommen, daß es ϑ_1 und $\vartheta_2 \in \Theta$ gibt, so daß

$$\begin{array}{lrcl} & R(\vartheta_1, T_1) & < & R(\vartheta_1, T_2) \\ \text{aber} & R(\vartheta_2, T_1) & > & R(\vartheta_2, T_2) \end{array}$$

Damit erlaubt dieses Kriterium oft nicht die Wahl eines eindeutigen Schätzers. Die Situation wird besser, wenn man sich auf erwartungstreue Schätzer einschränkt.

17.2 Beispiel: (ohne Beweis) Beim Taxiproblem gilt für die erwartungstreuen Schätzer $\hat{g}_1$ und $\hat{g}_2$:

$$R(\vartheta, \hat{g}_2) \leq R(\vartheta, \hat{g}_1),$$

d.h. $\hat{g}_2$ ist der bessere Schätzer.

17.3 Ist T erwartungstreu, so ist $E_\vartheta(t) = g(\vartheta)$, also

$$R(\vartheta, T) = \mathrm{Var}_\vartheta(T)$$

17.4 Satz: Im Beispiel (15.9) ist $\hat{p} : \{0, \ldots, n\} \to \mathbb{R}$, $k \mapsto \frac{k}{n}$ der erwartungstreue Schätzer für $g : \Theta = [0, 1] \to \mathbb{R}$, $p \mapsto p$ mit dem kleinsten quadratischen Fehler.

Beweis: Wir erinnern: Für $k \in \{0, \ldots, n\} = \mathcal{X}$ ist

$$L_k : \Theta = [0, 1] \to \mathbb{R}, \quad p \mapsto \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

die Likelihood-Funktion. Sei $T : \mathcal{X} \to \mathbb{R}$ erwartungstreuer Schätzer für g. Dann gilt

(1) $p = g(p) = E_p(T) = \sum\limits_{k=0}^{n} T(k) \cdot b(k; n, p) = \sum\limits_{k=0}^{n} T(k) \cdot L_k(p)$

(2) $1 = \sum\limits_{k=0}^{n} L_k(p)$

Wir leiten (1) und (2) nach p ab. Unter Benutzung von (15.11) erhalten wir

(3) $1 = \sum\limits_{k=0}^{n} T(k) \cdot L'_k(p) = \sum\limits_{k=0}^{n} T(k) \cdot \mathcal{L}'_k(p) \cdot L_k(p) = E_p(T \cdot \mathcal{L}'_x(p))$

(4) $0 = \sum\limits_{k=0}^{n} L_k'(p) = \sum\limits_{k=0}^{n} \mathcal{L}_k'(p) \cdot L_k(p) = E_p(\mathcal{L}_x'(p))$

$$\begin{aligned}\text{Es folgt: } 1 &= E_p(T \cdot \mathcal{L}_k'(p)) - E_p(T) \cdot E_p(\mathcal{L}_k'(p)) \\ &= E_p(T \cdot \mathcal{L}_k'(p) - E_p(T) \cdot \mathcal{L}_k'(p)) = E_p((T - E_p(T)) \cdot \mathcal{L}_k'(p)) \\ &\overset{10.16}{\leq} \sqrt{E_p((T - E_p(T)^2) \cdot E_p((\mathcal{L}_k'(p))^2)}\end{aligned}$$

Quadrieren wir, so erhalten wir

$$\operatorname{Var}_p(T) \geq \frac{1}{E_p(\mathcal{L}_x'(p)^2)} \overset{s.u.}{=} \frac{pq}{n}$$

Nun gilt:

(a) $\mathcal{L}_k'(p)^2 = \left(\frac{k}{pq} - \frac{n}{q}\right)^2 = \frac{k^2}{p^2q^2} - \frac{2n}{pq^2} \cdot k + \frac{n^2}{q^2}$

(b) Setzen wir $X : \{0, 1, \dots, n\} \to \mathbb{R}$, $k \to k$, so gilt $E_p(X) =$ Erwartungswert der Binomialverteilung, $\operatorname{Var}_p(X) =$ Varianz der Binomialverteilung. Also $E_p(X) = np$, $E_p(X^2) = \operatorname{Var}_p(X) + E_p(X)^2 = npq + n^2p^2$.

Es folgt

$$\begin{aligned}E_p(\mathcal{L}_x'(p)^2) &= \frac{1}{p^2q^2} E_p(X^2) - \frac{2n}{pq^2} E_p(X) + \frac{n^2}{q^2} = \frac{npq}{p^2q^2} + \frac{n^2p^2}{p^2q^2} - \frac{2n^2p}{pq^2} + \frac{n^2}{q^2} \\ &= \frac{n}{pq}\end{aligned}$$

Für den erwartungstreuen Schätzer $\hat{p}$ gilt

$$\begin{aligned}\operatorname{Var}_p(\hat{p}) &= \sum_{k=0}^{n} \hat{p}(k)^2 \cdot b(k; n, p) - E_p(\hat{p})^2 \\ &= \frac{1}{n^2} \sum_{k=0}^{n} (k)^2 \cdot b(k; n, p) - p^2 \\ &= \frac{1}{n^2} \underbrace{\left(\sum_{k=0}^{n} (k)^2 \cdot b(k; n, p) - (np)^2\right)}_{\text{Varianz der Binomialverteilung}} = \frac{1}{n^2} \cdot npq = \frac{pq}{n}\end{aligned}$$

Es folgt: $\operatorname{Var}_p(T) \geq \frac{pq}{n} = \operatorname{Var}_p(\hat{p})$. $\square$