Stochastik

Eine Vorlesung für das Lehramtsstudium

Franz Hofbauer

Oktober 2020

Vorwort

Der Begriff Wahrscheinlichkeit wird üblicherweise mit Häufigkeit assoziiert. Was oft eintritt, hat hohe Wahrscheinlichkeit, was selten eintritt, hat niedrige Wahrscheinlichkeit. Daher wird in diesem Skriptum Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten definiert. Damit wird der Additionssatz für endlich viele Ereignisse bewiesen, auf dem dann die Wahrscheinlichkeitstheorie aufbaut. Man kommt auch ohne σ-Additivität aus.

Das Skriptum beginnt mit der für die Wahrscheinlichkeitstheorie notwendigen Kombinatorik, bevor dann im zweiten Kapitel Wahrscheinlichkeit und bedingte Wahrscheinlichkeit definiert und die wichtigsten Sätze zum Berechnen von Wahrscheinlichkeiten bewiesen werden. Diese sind der Satz über gleichwahrscheinliche Ausfälle (Laplace-Wahrscheinlichkeit), der Multiplikationssatz, der Satz von der totalen Wahrscheinlichkeit und die Formel von Bayes.

Im dritten Kapitel werden diskrete und kontinuierliche Zufallsvariable und deren Verteilungen eingeführt. Ausführlich behandelt werden Binomialverteilung, geometrische und hypergeometrische Verteilung, Poissonverteilung, Exponentialverteilung, Gammaverteilung und Normalverteilung. Die Approximation der Binomialverteilung durch die Normalverteilung wird nur heuristisch hergeleitet.

Im vierten Kapitel wird die gemeinsame Verteilung mehrerer Zufallsvariable und die Unabhängigkeit von Zufallsvariablen eingeführt. Dann wird mit Zufallsvariablen gerechnet. Für die Wurzel aus einer Zufallvariable, für die Summe und für den Quotienten zweier Zufallsvariable werden die Wahrscheinlichkeitsdichten bestimmt. Es wird gezeigt, dass eine Summe unabhängiger normalverteilter Zufallsvariablen wieder normalverteilt ist. Schließlich werden noch Rechenregeln für Erwartungswert und Varianz behandelt. Bewiesen werden sie jedoch nur für diskrete Zufallsvariable.

Das fünfte Kapitel ist der Statistik gewidmet. Zuerst wird die Methode der kleinsten Quadrate behandelt. So gewinnt man Formeln zum Schätzen von Parametern aus vorliegenden Stichproben. Dann folgen Konfidenzintervalle und statistische Tests. Es wird auch der χ^2-Test behandelt, da er früher im Lehrplan war.

Das Skriptum enthält eine große Anzahl durchgerechneter Beispiele. Viele dieser Beispiele wurden in den Jahren 1997 bis 1999 aus den damaligen Schulbüchern übernommen.

Das Skriptum hat zwei Anhänge. Im ersten findet man einige Anwendungen der Wahrscheinlichkeitstheorie, im zweiten werden notwendige Vorkenntnisse aus der Analysis zusammengestellt.

I. Kombinatorik

Stichproben spielen in der Wahrscheinlichkeitstheorie eine wesentliche Rolle. In der Statistik zum Beispiel versucht man, Entscheidungen auf Grund einer zufällig gezogenen Stichprobe zu treffen. Auch die Lottoziehung kann man als Ziehen einer Stichprobe auffassen. In diesem ersten Teil des Skriptums beschäftigen wir uns mit dem Abzählen von Stichproben, die aus einer vorgegebenen Menge gezogen werden. Dazu verwenden wir Methoden der Kombinatorik. Wir beginnen mit geordneten Stichproben, behandeln dann ungeordnete Stichproben und schließlich Zerlegungen von Mengen und Anordnungen von vorgegebenen Objekten.

1. Geordnete Stichproben

Wir haben eine Menge M vor uns, aus der eine Stichprobe gezogen wird (zum Beispiel die Menge $\{1, 2, 3, \dots, 45\}$, aus der bei der Lottoziehung sechs Zahlen gezogen werden, oder eine Menge von Losen, aus der die Preisträger gezogen werden). Mit n bezeichnen wir die Anzahl der Elemente der Menge M. Aus dieser n-elementigen Menge M ziehen wir der Reihe nach k Elemente. Es gibt also einen ersten Zug, einen zweiten Zug, und so weiter bis zum k-ten Zug. Die Ordnung in der Stichprobe ist wesentlich, daher spricht man von geordneten Stichproben. Die Anzahl k der Elemente in der Stichprobe nennt man den Stichprobenumfang.

Wir unterscheiden geordnete Stichproben mit und ohne Zurücklegen. Geordnete Stichproben mit Zurücklegen vom Umfang k erhält man, wenn man aus der Menge M der Reihe nach k Elemente zieht und jedes sofort wieder zurücklegt. Es wird also jedes Mal aus der ursprünglichen Menge M gezogen. Jedes Element kann öfter in der Stichprobe vorkommen. Geordnete Stichproben ohne Zurücklegen vom Umfang k erhält man, wenn man aus der Menge M der Reihe nach k Elemente zieht und nicht zurücklegt. Jedes Element kann nur einmal in der Stichprobe vorkommen.

Geordnete Stichproben schreibt man in der Form $(x_1, x_2, \dots, x_k)$ im Gegensatz zu Mengen, die ungeordnet sind und für die man geschwungene Klammern verwendet. Geordnete Stichproben vom Umfang 2 sind Paare (x_1, x_2) und geordnete Stichproben vom Umfang 3 sind Tripel (x_1, x_2, x_3).

Beispiel 1: Man schreibe alle geordneten Stichproben vom Umfang 2 aus der 4-elementigen Menge $\{a, b, c, d\}$ auf.

Mit Zurücklegen:

$$\begin{array}{cccc} (a,a) & (a,b) & (a,c) & (a,d) \\ (b,a) & (b,b) & (b,c) & (b,d) \\ (c,a) & (c,b) & (c,c) & (c,d) \\ (d,a) & (d,b) & (d,c) & (d,d) \end{array}$$

Ohne Zurücklegen:

$$\begin{array}{ccc} (a,b) & (a,c) & (a,d) \\ (b,a) & (b,c) & (b,d) \\ (c,a) & (c,b) & (c,d) \\ (d,a) & (d,b) & (d,c) \end{array}$$

Man sieht hier schon, wie man diese Stichproben zählen kann. Für den ersten Zug gibt es 4 Möglichkeiten, nämlich alle Elemente der Menge. Wird zurückgelegt, dann gibt es für den zweiten Zug ebenfalls 4 Möglichkeiten. Da man jeden der 4 möglichen zweiten Züge an jeden der 4 möglichen ersten Züge anfügen kann, erhält man $4 \cdot 4 = 16$ geordnete Stichproben mit Zurücklegen.

Wird nicht zurückgelegt, dann gibt es für den zweiten Zug nur 3 Möglichkeiten. Welche Möglichkeiten das sind, hängt davon ab, wie der erste Zug ausgefallen ist. Es sind aber immer 3 Möglichkeiten. Daher gibt es $4 \cdot 3 = 12$ geordnete Stichproben ohne Zurücklegen.

Wir wollen diese Anzahlen allgemein ausrechnen. Zuvor eine Definition.

Definition: Wir definieren $n! = n(n-1)(n-2)\dots 2 \cdot 1$ für $n \in \mathbb{N}$ und $0! = 1$. Man liest $n!$ als "n-Faktorielle".

Satz 1: *Die Anzahl der geordneten Stichproben aus einer n-elementigen Menge M vom Umfang k mit Zurücklegen ist n^k. Die Anzahl der geordneten Stichproben aus einer n-elementigen Menge M vom Umfang k ohne Zurücklegen ist $n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$.*

Beweis: Mit Zurücklegen: Für den ersten Zug kommen alle Elemente der Menge M in Frage, also gibt es n Möglichkeiten. Da wir zurücklegen, wird beim zweiten Mal ebenfalls aus der ursprünglichen Menge M gezogen, also gibt es auch für den zweiten Zug n Möglichkeiten. Das geht so weiter bis zum k-ten Zug. Für jeden gibt es n Möglichkeiten. Wir erhalten also $n \cdot n \cdot \ldots \cdot n = n^k$ Stichproben.

Ohne Zurücklegen: Für den ersten Zug kommen alle Elemente der Menge M in Frage, also gibt es n Möglichkeiten. Da wir nicht zurücklegen, wird beim zweiten Mal aus einer Menge gezogen, die um ein Element weniger hat, also gibt es für den zweiten Zug $n-1$ Möglichkeiten. Beim dritten Mal wird aus einer Menge gezogen, die um zwei Elemente weniger hat, also gibt es für den dritten Zug $n-2$ Möglichkeiten. Das geht so weiter bis zum k-ten Zug, für den es dann nur mehr $n-k+1$ Möglichkeiten gibt. Wir erhalten also $n(n-1)(n-2)\dots(n-k+1)$ Stichproben. $\square$

Beispiel 2: Wie viele mögliche Tipps erlaubt ein Totoschein?

Ein Tipp auf dem Totoschein besteht darin, dass man zu jedem der 12 Spiele eines der Zeichen 1, 2, oder X hinschreibt. Für das erste Spiel wählt man ein Element aus der Menge $\{1, 2, X\}$, für das zweite Spiel wählt man ebenfalls ein Element aus der Menge $\{1, 2, X\}$, und so tut man weiter bis zum 12-ten Spiel. Es gibt also $3 \cdot 3 \cdot \ldots \cdot 3 = 3^{12}$ mögliche Tipps. Die möglichen Tipps sind die geordneten Stichproben vom Umfang 12 (es sind 12 Spiele) aus der 3-elementigen Menge $\{1, 2, X\}$ mit Zurücklegen (es wird immer aus derselben Menge $\{1, 2, X\}$ gewählt).

Beispiel 3: In einem Hotel sind 6 Einbettzimmer frei. Es kommen 4 Gäste. Wie viele Möglichkeiten gibt es, diese auf die 6 Zimmer zu verteilen?

Die Gäste kommen einer nach dem anderen dran. Für den ersten Gast gibt es 6 mögliche Zimmer. Für den zweiten Gast, der als nächster drankommt, gibt es dann nur mehr 5 freie Zimmer. Für den dritten Gast gibt es nur mehr 4 und für den vierten Gastnur mehr 3 freie Zimmer. Also haben wir $6 \cdot 5 \cdot 4 \cdot 3$ mögliche Zimmereinteilungen. Hier handelt es sich also um geordnete Stichproben vom Umfang 4 aus einer 6-elementigen Menge ohne Zurücklegen.

Stimmt der Stichprobenumfang k mit der Anzahl n der Elemente der Menge, aus der gezogen wird, überein, dann sind die geordneten Stichproben ohne Zurücklegen gerade die verschiedenen möglichen Anordnungen der n Elemente der Menge. Wir schreiben das als eigenen Satz auf.

Satz 2: *Die Anzahl aller möglichen Anordnungen von n verschiedenen Objekten ist $n!$.*

Beweis: Genauso wie bei den geordneten Stichproben ohne Zurücklegen. Auf den ersten Platz können wir jedes der n Objekte setzen. Ist der erste Platz besetzt, dann sind noch $n-1$ Objekte übrig, die wir auf den zweiten Platz setzen können. Für den dritten Platz gibt es noch $n-2$ Besetzungsmöglichkeiten und so weiter. Für den n-ten Platz gibt es nur mehr eine Möglichkeit. Also haben wir $n(n-1)\dots 1 = n!$ mögliche Anordnungen. □

Wir verallgemeinern die geordneten Stichproben, indem wir bei jedem Zug aus einer anderen Menge ziehen.

Satz 3: *Seien $M_1, M_2, \dots, M_k$ Mengen, wobei n_j die Anzahl der Elemente der Menge M_j ist. Die Anzahl aller geordneten Stichproben vom Umfang k, wobei beim j-ten Mal aus der Menge M_j gezogen wird, ist $n_1 n_2 \dots n_k$.*

Beweis: Für den ersten Zug kommen alle Elemente der Menge M_1 in Frage, also gibt es n_1 Möglichkeiten. Für den zweiten Zug kommen alle Elemente der Menge M_2 in Frage, also gibt es n_2 Möglichkeiten. So geht es weiter bis zum letzten Zug, für den es n_k Möglichkeiten gibt. Also haben wir $n_1 n_2 \dots n_k$ verschiedene Stichproben. □

Beispiel 4: Eine Autonummer ist eine Folge von Zeichen, die Ziffern oder Buchstaben sein können. Wie viele 4-stellige Autonummern gibt es? Wie viele 4-stellige Autonummern gibt es, die mit einer Ziffer enden? Wie viele 4-stellige Autonummern gibt es, die mit zwei Buchstaben beginnen? Wie viele 4-stellige Autonummern gibt es, die abwechselnd aus Ziffern und Buchstaben bestehen?

Die 4-stelligen Autonummern sind die geordneten Stichproben vom Umfang 4 aus einer 36-elementigen Menge mit Zurücklegen. Es gibt also 36^4 4-stellige Autonummern.

Soll die Autonummer mit einer Ziffer enden, dann werden die ersten drei Zeichen aus einer 36-elementigen Menge gewählt, das vierte Zeichen, das eine Ziffer ist, jedoch aus einer 10-elementigen Menge. Daher gibt es $36 \cdot 36 \cdot 36 \cdot 10$ Autonummern, die mit einer Ziffer enden.

Soll die Autonummer mit zwei Buchstaben beginnen, dann werden die ersten beiden Zeichen, die ja Buchstaben sind, aus einer 26-elementigen Menge gewählt, die anderen beiden Zeichen aus einer 36-elementigen Menge. Daher gibt es $26 \cdot 26 \cdot 36 \cdot 36$ Autonummern, die mit zwei Buchstaben beginnen.

Wechseln Ziffern und Buchstaben ab, dann gibt es zwei Möglichkeiten. Entweder beginnt die Autonummer mit einer Ziffer oder mit einem Buchstaben. Im ersten Fall stehen an der ersten und dritten Stelle Ziffern, an der zweiten und vierten Stelle Buchstaben. Für diesen Fall gibt es $10 \cdot 26 \cdot 10 \cdot 26$ Möglichkeiten. Im zweiten Fall stehen an der ersten und dritten Stelle Buchstaben, an der zweiten und vierten Stelle Ziffern. Für diesen Fall gibt es $26 \cdot 10 \cdot 26 \cdot 10$ Möglichkeiten. Es gibt also insgesamt $2 \cdot 10^2 \cdot 26^2$ Autonummern, die abwechselnd aus Ziffern und Buchstaben bestehen.

2. Ungeordnete Stichproben

Im Gegensatz zu den geordneten Stichproben spielt bei den ungeordneten Stichproben die Reihenfolge innerhalb der Stichprobe keine Rolle. Das Ziehen einer ungeordneten Stichprobe vom Umfang k stellt man sich am besten so vor, dass man mit einem Griff k Elemente aus einer n-elementigen Menge M zieht. Die ungeordneten Stichproben vom Umfang k sind also die k-elementigen Teilmengen dieser Menge.

Beispiel 5: Man schreibe alle ungeordneten Stichproben vom Umfang 3, also alle 3-elementigen Teilmengen, aus der Menge $\{a, b, c, d, e\}$ auf.

$\{a, b, c\}$, $\{a, b, d\}$, $\{a, b, e\}$, $\{a, c, d\}$, $\{a, c, e\}$, $\{a, d, e\}$, $\{b, c, d\}$, $\{b, c, e\}$, $\{b, d, e\}$, $\{c, d, e\}$

Die Anzahl der 3-elementigen Teilmengen aus der 5-elementigen Menge $\{a, b, c, d, e\}$ ist somit 10. Wir suchen eine Formel für die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Zuvor eine Definition.

Definition: Für $n \geq 0$ und $0 \leq k \leq n$ definieren wir $\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$, die sogenannten Binomialkoeffizienten. Man liest n über k. Manchmal setzt man $\binom{n}{k} = 0$, wenn $k < 0$ oder $k > n$.

Satz 4: *Sei $0 \leq k \leq n$. Die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge M ist $\binom{n}{k}$.*

Beweis: Den Fall $k = 0$ behandeln wir zuerst. Es gibt genau eine 0-elementige Teilmenge, nämlich die leere Menge. Wegen $\binom{n}{0} = 1$ stimmt die angegebene Formel für $k = 0$.

Sei jetzt $k \geq 1$ und a_k die Anzahl der k-elementigen Teilmengen der n-elementigen Menge M. Wir leiten eine Gleichung für a_k her, aus der wir dann a_k berechnen. Dazu stellen wir folgende Überlegung an. Nach Satz 2 gibt es für jede k-elementige Teilmenge $k!$ verschiedene Anordnungen. Schreiben wir alle diese möglichen Anordnungen für die a_k Teilmengen auf, dann erhalten wir insgesamt $a_k k!$ geordneten Stichproben. Das sind dann alle geordneten Stichproben vom Umfang k ohne Zurücklegen aus der n-elementigen Menge M. Ihre Anzahl ist $\frac{n!}{(n-k)!}$ nach Satz 1. Daher muss $a_k k! = \frac{n!}{(n-k)!}$ gelten. Es folgt $a_k = \frac{n!}{k!(n-k)!}$, die gesuchte Formel. $\square$

Beispiel 6: Wie viele Diagonalen hat ein regelmäßiges n-Eck.

Die Anzahl aller Geraden, die durch je 2 Punkte dieser n Eckpunkte gehen, ist gleich der Anzahl der 2-elementigen Teilmengen aus den n Eckpunkten, also $\binom{n}{2}$. Daher ist die Anzahl der Seiten und Diagonalen zusammen gleich $\binom{n}{2}$. Da es n Seiten gibt, ist die Anzahl der Diagonalen gleich $\binom{n}{2} - n$.

Beispiel 7: Ein Verein, der 22 Mitglieder hat, will einen Ausschuss von 5 Personen einsetzen. Wie viele Möglichkeiten gibt es? Von den 22 Mitgliedern sind 14 Frauen und 8 Männer. Wie viele Möglichkeiten gibt es, wenn der Ausschuss aus 3 Frauen und 2 Männern bestehen soll?

Ein Ausschuss ist eine Teilmenge. Daher gibt es $\binom{22}{5}$ Möglichkeiten, einen 5-köpfigen Ausschuss aus den 22 Mitgliedern zu wählen.

Soll der Ausschuss 3 Frauen und 2 Männer enthalten, dann wählen wir die Frauen und Männer getrennt aus. Sei $\mathcal{T}_1$ die Menge aller 3-elementigen Teilmengen aus der Menge der 14 Frauen. Sei $\mathcal{T}_2$ die Menge aller 2-elementigen Teilmengen aus der Menge der 8 Männer. Einen 5-köpfigen Ausschuss mit 3 Frauen und 2 Männern erhält man dann, indem man eine der Mengen aus $\mathcal{T}_1$ mit einer der Mengen aus $\mathcal{T}_2$ zusammensetzt. Die Anzahl aller möglichen Ausschüsse ist nach Satz 3 gleich $n_1 n_2$, wobei $n_1 = \binom{14}{3}$ die Anzahl der Elemente von $\mathcal{T}_1$ ist und $n_2 = \binom{8}{2}$ die Anzahl der Elemente von $\mathcal{T}_2$ ist. Es gibt daher $\binom{14}{3}\binom{8}{2}$ mögliche Ausschüsse mit 3 Frauen und 2 Männern.

Beispiel 8: Aus 52 Spielkarten (13 $\heartsuit$-Karten, 13 $\diamondsuit$-Karten, 13 $\clubsuit$-Karten und 13 $\spadesuit$-Karten) wird eine ungeordnete Stichprobe vom Umfang 11 gezogen. Wie viele solche Stichproben gibt es, die 6 ♡-Karten, 3 ♢-Karten und 2 ♣-Karten enthalten?

Wir gehen wie in Beispiel 7 vor. Sei $\mathcal{T}_1$ die Menge aller 6-elementigen Teilmengen aus den ♡-Karten. Ihre Anzahl ist $\binom{13}{6}$. Sei $\mathcal{T}_2$ die Menge aller 3-elementigen Teilmengen aus den ♢-Karten. Ihre Anzahl ist $\binom{13}{3}$. Sei $\mathcal{T}_3$ die Menge aler 2-elementigen Teilmengen aus den ♣-Karten. Ihre Anzahl ist $\binom{13}{2}$. Eine 11-elementige Teilmenge mit 6 ♡-Karten, 3 ♢-Karten und 2 ♣-Karten erhält man dadurch, dass man eine Menge aus $\mathcal{T}_1$, eine Menge aus $\mathcal{T}_2$, und eine Menge aus $\mathcal{T}_3$ zusammensetzt. Gemäß Satz 3 gibt es dafür $\binom{13}{6}\binom{13}{3}\binom{13}{2}$ Möglichkeiten.

Bemerkung: Sei $0 \le M \le N$. Wir verwenden die Methode aus Beispiel 7, um die Gleichung $\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$ für $0 \le n \le N$ zu zeigen, wobei $\binom{u}{v} = 0$ zu setzen ist, wenn $v > u$ gilt. Das geht so: Aus einer Menge von N Kugeln, von denen M weiß und $N-M$ schwarz sind, werden n-elementige Teilmengen gezogen. Die Anzahl der n-elementigen Teilmengen, die k weiße Kugeln und $n-k$ schwarze Kugeln enthalten, ist $\binom{M}{k}\binom{N-M}{n-k}$, da sich jede dieser Teilmengen aus einer k-elementigen Teilmenge der M weißen Kugeln und einer $n-k$-elementigen Teilmenge der $N-M$ schwarzen Kugeln zusammensetzen lässt. Die Anzahl aller n-elementigen Teilmengen ist dann $\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k}$. Die Anzahl aller n-elementigen Teilmengen ist aber auch $\binom{N}{n}$. Daher gilt $\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$.

Wir versuchen jetzt das selbe Problem für geordneten Stichproben zu lösen. Vorhin haben wir die Anzahl der ungeordneten Stichproben vom Umfang 5, die 3 Frauen und 2 Männer enthalten, berechnet. Jetzt berechnen wir die Anzahl der geordneten Stichproben vom Umfang 5, die 3 Buchstaben und 2 Ziffern enthalten.

Beispiel 9: Autonummern bestehen aus Buchstaben und Ziffern. Wie viele verschiedene 5-stellige Autonummern gibt es, die 3 Buchstaben und 2 Ziffern enthalten? Wie viele davon bestehen aus lauter verschiedenen Zeichen? Gefragt ist die Anzahl der geordneten Stichproben vom Umfang 5 (mit und ohne Zurücklegen), die 3 Buchstaben und 2 Ziffern enthalten.

Wir stellen uns 5 Plätze vor, auf die die Buchstaben und Ziffern gestellt werden. Von diesen 5 Plätzen werden 3 mit Buchstaben (B) und 2 mit Ziffern (Z) besetzt. Es gibt folgende Möglichkeiten, die Plätze aufzuteilen: BBBZZ BBZBZ BBZZB BZBBZ BZBZB BZZBB ZBBBZ ZBBZB ZBZBB ZZBBB.

Wie viele geordnete Stichproben gibt es für die Platzaufteilung BBBZZ, also wenn die ertsten 3 Plätze mit Buchstaben und die letzten 2 Plätze mit Ziffern besetzt werden. Für den ersten Platz kommem 26 Buchstaben in Frage, ebenso für den zweiten und dritten, da zurückgelegt wird. Für den vierten Platz kommen 10 Ziffern in Frage und ebenso für den fünften. Es gibt also $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10$ Stichproben, die Platzaufteilung BBBZZ haben.

Ebenso kann man die Anzahl der Stichproben für jede der anderen Platzaufteilungen ausrechnen. Die Anzahl für die Platzaufteilung BBZBZ ist $26 \cdot 26 \cdot 10 \cdot 26 \cdot 10$, die Anzahl für BBZZB ist $26 \cdot 26 \cdot 10 \cdot 10 \cdot 26$, und so weiter. Man sieht, dass die Anzahl der Stichproben für jede Platzaufteilung dieselbe ist, nämlich $26^3 10^2$. Da es insgesamt 10 verschiedene Platzaufteilungen gibt, ist $10 \cdot 26^3 10^2$ die gesuchte Anzahl der geordneten Stichproben vom Umfang 5 mit Zurücklegen, die 3 Buchstaben und 2 Ziffern enthalten.

Wird nicht zurückgelegt, dann ist $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9$ die Anzahl der Stichproben, die Platzaufteilung BBBZZ haben, da für den zweiten Platz nur mehr die 25 übriggebliebenen Buchstaben in Frage kommen und für den dritten Platz nur mehr die 24 übriggebliebenen.

Ebenso kommen für den fünften Platz nur mehr die 9 übriggebliebenen Ziffern in Frage. Die Anzahl der Stichproben mit Platzaufteilung BBZBZ ist $26 \cdot 25 \cdot 10 \cdot 24 \cdot 9$ und die für BZBZB ist $26 \cdot 10 \cdot 25 \cdot 9 \cdot 24$. Man sieht wieder, dass die Anzahl der Stichproben für jede Platzaufteilung dieselbe ist, nämlich $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9$. Daher ist $10 \cdot 26 \cdot 25 \cdot 24 \cdot 10 \cdot 9$ die gesuchte Anzahl der geordneten Stichproben vom Umfang 5 ohne Zurücklegen, die 3 Buchstaben und 2 Ziffern enthalten.

Wir hätten in diesem Beispiel die Platzaufteilungen nicht auflisten müssen. Es genügt ja, die Anzahl der Platzaufteilungen zu kennen. Diese kann man mit Hilfe von Satz 4 ermitteln. Die mit B besetzten Plätze bilden jeweils eine 3-elementige Teilmenge aus der 5-elementigen Menge der Plätze. Die möglichen Platzaufteilungen entsprechen daher den 3-elementigen Teilmengen einer 5-elementigen Menge. Ihre Anzahl ist $\binom{5}{3} = 10$. (Wenn man dieselbe Überlegung mit Z statt mit B anstellt, dann erhält man als Anzahl $\binom{5}{2}$, also ebenfalls 10.)

Beispiel 10: Aus 52 Spielkarten (13 $\heartsuit$-Karten, 13 $\diamondsuit$-Karten, 13 $\clubsuit$-Karten und 13 $\spadesuit$-Karten) wird eine geordnete Stichprobe vom Umfang 17 mit Zurücklegen gezogen. Wie viele solche Stichproben gibt es, die 8 $\heartsuit$-Karten und 9 $\diamondsuit$-Karten enthalten?

Das ist dieselbe Aufgabenstellung wie im letzten Beispiel, nur ist es jetzt nicht mehr möglich, alle Platzaufteilungen aufzulisten. Da jetzt 8 der 17 Plätze mit $\heartsuit$-Karten und die übrigen 9 mit $\diamondsuit$-Karten besetzt werden, gibt es $\binom{17}{8}$ mögliche Platzaufteilungen. Die Anzahl der Stichproben, wo die ersten 8 Plätze mit $\heartsuit$-Karten und die letzten 9 Plätze mit $\diamondsuit$-Karten besetzt sind, ist $13^8 13^9$. Dieselbe Anzahl erhält man auch für alle anderen Platzaufteilungen. Daher gibt es $\binom{17}{8}13^8 13^9$ geordnete Stichproben vom Umfang 17, die 8 $\heartsuit$-Karten und 9 $\diamondsuit$-Karten enthalten.

Zum Abschluss soll noch ein Beispiel behandelt werden, das die verschiedenen Arten von Stichproben einander gegenüberstellt. Statt Buchstaben und Ziffern, statt $\heartsuit$-Karten und $\diamondsuit$-Karten, verwenden wir jetzt rote und grüne Kugeln. Dabei muss man sich jedoch auch gleichfarbige Kugeln als unterscheidbar (nummeriert) vorstellen, so wie die Buchstaben voneinander unterscheidbar sind, und wie auch die $\heartsuit$-Karten voneinander unterscheidbar sind.

Beispiel 11: Aus einer Menge von 14 Kugeln, von denen 9 rot und 5 grün sind, wird eine Stichprobe vom Umfang 7 gezogen. Wie viele verschiedene Stichproben, die 3 rote und 4 grüne Kugeln enthalten, gibt es, wenn
(a) ungeordnet gezogen wird?
(b) geordnet mit Zurücklegen gezogen wird?
(c) geordnet ohne Zurücklegen gezogen wird?

(a) Es gibt $\binom{9}{3}$ Möglichkeiten, eine Teilmenge vom Umfang 3 aus den 9 roten Kugeln zu ziehen. Es gibt $\binom{5}{4}$ Möglichkeiten, eine Teilmenge vom Umfang 4 aus den 5 grünen Kugeln zu ziehen. Indem man jeweils eine dieser Teilmengen aus den roten Kugeln mit einer der Teilmengen aus den grünen Kugeln zusammensetzt, erhält man alle möglichen ungeordneten Stichproben. Ihre Anzahl ist daher $\binom{9}{3}\binom{5}{4}$.

(b) Jede Stichprobe ist eine geordnete Folge von 7 Kugeln, wobei 3 Plätze in dieser Folge von roten Kugeln besetzt werden und 4 Plätze von grünen. Die Anzahl der möglichen Platzaufteilungen ist $\binom{7}{3}$. Die Anzahl der Stichproben mit einer bestimmten Platzaufteilung ist $9^3 5^4$, da zurückgelegt wird. Die Anzahl aller geordneten Stichproben mit Zurücklegen, die 3 rote und 4 grüne Kugeln enthalten, ist daher $\binom{7}{3}9^3 5^4$.

(c) Hier geht man genauso vor wie in (b). Da jetzt aber nicht zurückgelegt wird, gibt es für jede der $\binom{7}{3}$ Platzaufteilungen nur $9 \cdot 8 \cdot 7 \cdot 5 \cdot 4 \cdot 3 \cdot 2$ mögliche Besetzungen mit Kugeln. Die Anzahl aller geordneten Stichproben ohne Zurücklegen ist daher $\binom{7}{3} \cdot 9 \cdot 8 \cdot 7 \cdot 5 \cdot 4 \cdot 3 \cdot 2$. Man kann diese Anzahl auch auf andere Weise erhalten, indem man von (a) ausgeht. Man erhält die geordneten Stichproben ohne Zurücklegen nämlich dadurch, dass man die ungeordneten Stichproben aus (a) auf alle möglichen Arten anordnet. Da man jede dieser ungeordneten Stichproben auf 7! Arten anordnen kann, ist die Anzahl der geordneten Stichproben ohne Zurücklegen gleich $\binom{9}{3}\binom{5}{4}7!$. Das ist dieselbe Anzahl wie vorhin.

Bemerkung: Die hier behandelten ungeordneten Stichproben sind ungeordnete Stichproben ohne Zurücklegen. Es gibt auch ungeordnete Stichproben mit Zurücklegen. (Man zieht geordnet und vergisst die Anordnung. Man erhält "Mengen", in denen Elemente mehrfach vorkommen können.) Man kann zeigen, dass die Anzahl der ungeordneten Stichproben vom Umfang k aus einer n-elementigen Menge mit Zurücklegen gleich $\binom{n+k-1}{k}$ ist. Ein Beispiel für ungeordnete Stichproben mit Zurücklegen vom Umfang 2 aus der Menge $\{0, 1, 2, 3, 4, 5, 6\}$ sind Dominosteine. Ihre Anzahl ist $\binom{7+2-1}{2} = 28$.

3. Zerlegungen einer Menge

In Beispiel 10 kommen nur ♡-Karten und ♢-Karten in der Stichprobe vor. Fragen wir nach der Anzahl der Stichproben vom Umfang 18, die fünf ♡-Karten, drei ♢-Karten, zwei ♣-Karten und acht ♠-Karten enthalten, dann müssen wir zuerst die Anzahl der möglichen Aufteilungen der 18 Plätze in fünf Plätze für die ♡-Karten, in drei Plätze für die ♢-Karten, in zwei Plätze für die ♣-Karten und in acht Plätze für die ♠-Karten bestimmen. Eine Platzaufteilung entspricht einer Zerlegung der Menge der Plätze in Teilmengen, wobei die Anzahl der Elemente dieser Teilmengen vorgegeben ist. Mit solchen Zerlegungen wollen wir uns jetzt beschäftigen.

Vorgegeben ist eine n-elementige Menge M. Zu bestimmen ist die Anzahl aller möglichen Zerlegungen dieser Menge in j Teilmengen, von denen die erste k_1 Elemente, die zweite k_2 Elemente und schließlich die j-te k_j Elemente enthält.

Beispiel 12: Man schreibe alle Zerlegungen der Menge $\{a, b, c, d, e\}$ in 3 Teilmengen auf, von denen die erste 2 Elemente, die zweite 1 Element und die dritte 2 Elemente enthält.

$\{a,b\},\ \{c\},\ \{d,e\}$	$\{a,b\},\ \{d\},\ \{c,e\}$	$\{a,b\},\ \{e\},\ \{c,d\}$
$\{a,c\},\ \{b\},\ \{d,e\}$	$\{a,c\},\ \{d\},\ \{b,e\}$	$\{a,c\},\ \{e\},\ \{b,d\}$
$\{a,d\},\ \{b\},\ \{c,e\}$	$\{a,d\},\ \{c\},\ \{b,e\}$	$\{a,d\},\ \{e\},\ \{b,c\}$
$\{a,e\},\ \{b\},\ \{c,d\}$	$\{a,e\},\ \{c\},\ \{b,d\}$	$\{a,e\},\ \{d\},\ \{b,c\}$
$\{b,c\},\ \{a\},\ \{d,e\}$	$\{b,c\},\ \{d\},\ \{a,e\}$	$\{b,c\},\ \{e\},\ \{a,d\}$
$\{b,d\},\ \{a\},\ \{c,e\}$	$\{b,d\},\ \{c\},\ \{a,e\}$	$\{b,d\},\ \{e\},\ \{a,c\}$
$\{b,e\},\ \{a\},\ \{c,d\}$	$\{b,e\},\ \{c\},\ \{a,d\}$	$\{b,e\},\ \{d\},\ \{a,c\}$
$\{c,d\},\ \{a\},\ \{b,e\}$	$\{c,d\},\ \{b\},\ \{a,e\}$	$\{c,d\},\ \{e\},\ \{a,b\}$
$\{c,e\},\ \{a\},\ \{b,d\}$	$\{c,e\},\ \{b\},\ \{a,d\}$	$\{c,e\},\ \{d\},\ \{a,b\}$
$\{d,e\},\ \{a\},\ \{b,c\}$	$\{d,e\},\ \{b\},\ \{a,c\}$	$\{d,e\},\ \{c\},\ \{a,b\}$

An diesem Beispiel kann man auch schon erkennen, wie man die Anzahl der möglichen Zerlegungen ermitteln kann. Es gibt $\binom{5}{2} = 10$ Möglichkeiten, die erste Teilmenge zu wählen. Zu jeder dieser Möglichkeiten kann man auf $\binom{3}{1} = 3$ Arten die zweite Teilmenge wählen. Für die dritte Teilmenge gibt es dann nur mehr $1 = \binom{2}{2}$ Möglichkeit. Insgesamt haben wir also $\binom{5}{2}\binom{3}{1}\binom{2}{2} = 30$ Möglichkeiten.

Satz 5: *Sei $n = k_1 + k_2 + \cdots + k_j$. Die Anzahl der möglichen Zerlegungen einer n-elementigen Menge M in j Teilmengen, von denen die erste k_1 Elemente, die zweite k_2 Elemente, … und die j-te k_j Elemente hat, ist $\frac{n!}{k_1!k_2!k_3!\ldots k_j!}$.*

Beweis: Es gibt $\binom{n}{k_1}$ Möglichkeiten, die erste Teilmenge aus M zu wählen. Es verbleiben $n - k_1$ Elemente in M. Es gibt dann $\binom{n-k_1}{k_2}$ Möglichkeiten, die zweite Teilmenge zu wählen. Es bleiben $n - k_1 - k_2$ Elemente in M. Somit gibt es $\binom{n-k_1-k_2}{k_3}$ Möglichkeiten, die dritte Teilmenge zu wählen. So tut man weiter. Für die letzte Teilmenge bleiben $n - k_1 - \cdots - k_{j-1} = k_j$ Elemente übrig. Daher gibt es nur mehr $1 = \binom{k_j}{k_j} = \binom{n-k_1-\cdots-k_{j-1}}{k_j}$ Möglichkeit für die letzte Teilmenge. Die Anzahl aller möglichen Zerlegungen ist also $\binom{n}{k_1}\binom{n-k_1}{k_2}\ldots\binom{n-k_1-\cdots-k_{j-1}}{k_j} = \frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\ldots\frac{(n-k_1-\cdots-k_{j-1})!}{k_j!0!} = \frac{n!}{k_1!k_2!\ldots k_j!}$. Damit ist die gewünschte Formel gefunden. ∎

Beispiel 13: Für ein Photo sollen sich 22 Personen, die verschiedene Größe haben, in zwei Reihen aufstellen, wobei die vorne stehende Person jeweils kleiner als die dahinterstehende Person ist. Auf wie viele Arten ist das möglich?

Eine 22-elementige Menge ist in 11 Teilmengen aufzuteilen, von denen jede 2 Elemente enthält. Die beiden Personen in der ersten Teilmenge stellen sich ganz links auf, die beiden Personen in der zweiten Teilmenge daneben und so weiter, wobei jeweils die kleinere Person vorne und die größere hinten steht. Nach Satz 5 ist so eine Aufteilung auf $\frac{22!}{(2!)^{11}}$ verschiedene Arten möglich.

Beispiel 14: Aus 52 Spielkarten, die aus 13 $\heartsuit$-Karten, aus 13 $\diamondsuit$-Karten, aus 13 $\clubsuit$-Karten und aus 13 $\spadesuit$-Karten bestehen, wird eine geordnete Stichprobe vom Umfang 17 mit Zurücklegen gezogen. Wie viele solche Stichproben gibt es, die 5 $\heartsuit$-Karten, 3 $\diamondsuit$-Karten, 7 $\clubsuit$-Karten und 2 $\spadesuit$-Karten enthalten.

Wir gehen vor wie in Beispiel 10. Es werden zuerst die 17 Plätze aufgeteilt und zwar in 5 Plätze für die $\heartsuit$-Karten, in 3 Plätze for die $\diamondsuit$-Karten, in 7 Plätze für die $\clubsuit$-Karten und in 2 Plätze für die $\spadesuit$-Karten. Die Anzahl der möglichen Platzaufteilungen ist gerade die Anzahl der möglichen Zerlegungen einer 17-elementigen Menge in 4 Teilmengen, von denen die erste 5 Elemente, die zweite 3 Elemente, die dritte 7 Elemente und die vierte 2 Elemente hat. Diese Anzahl ist $\frac{17!}{5!3!7!2!}$ nach Satz 5.

Gibt man eine Platzaufteilung vor, dann gibt es 13^5 Möglichkeiten, die 5 Plätze für die $\heartsuit$-Karten zu besetzen. Weiters gibt es 13^3 Möglichkeiten, die 3 Plätze für die $\diamondsuit$-Karten zu besetzen, 13^7 Möglichkeiten, die 7 Plätze für die $\clubsuit$-Karten zu besetzen, und schließlich 13^2 Möglichkeiten, die 2 Plätze für die $\clubsuit$-Karten zu besetzen. Es gibt also $\frac{17!}{5!3!7!2!}13^5 13^3 13^7 13^2$ geordnete Stichproben vom Umfang 17 mit Zurücklegen, die 5 $\heartsuit$-Karten, 3 $\diamondsuit$-Karten, 7 $\clubsuit$-Karten und 2 $\spadesuit$-Karten enthalten.

4. Anordnungen (Permutationen)

Die Anzahl der möglichen Anordnungen von n verschiedenen Objekten ist $n!$ wie in Satz 2 gezeigt wurde. Wie viele Anordnungen gibt es nun, wenn die n Objekte nicht verschieden sind? Wir haben n Buchstaben, unter denen k_1 mal Buchstabe B_1, k_2 mal Buchstabe B_2, ... und k_j mal Buchstabe B_j vorkommt, wobei $n = k_1 + k_2 + \cdots + k_j$ gilt. Diese n Buchstaben werden auf n Plätzen angeordnet. Jede Anordnung entspricht einer Zerlegung der Menge der n Plätze in j Teilmengen, wobei die erste Teilmenge k_1 Elemente hat (auf diesen Plätzen steht B_1), die zweite Teilmenge k_2 Elemente hat (auf diesen Plätzen steht B_2), ... und die j-te Teilmenge k_j Elemente hat (auf diesen Plätzen steht B_j). Nach Satz 5 ist die Anzahl dieser Zerlegungen und damit auch die Anzahl der möglichen Anordnungen dieser Buchstaben gleich $\frac{n!}{k_1!k_2!k_3!\dots k_j!}$.

Beispiel 15: Auf wie viele Arten kann man die acht Buchstaben AAABBBCC anordnen, wenn an erster Stelle kein A stehen soll?

Die Anzahl aller möglichen Anordnungen dieser acht Buchstaben ist $\frac{8!}{3!3!2!} = 560$. Um die Anzahl der Anordnungen, die A an erster Stelle haben, zu bestimmen, schreiben wir ein A an die erste Stelle. Da die übrigen sieben Buchstaben beliebig auf den restlichen sieben Plätzen angeordnet werden können, ist diese Anzahl $\frac{7!}{2!3!2!} = 210$. Es gibt $560 - 210 = 350$ Anordnungen, die A nicht an erster Stelle haben.

Genausogut kann man die Anzahl der Anordnungen berechnen, die B an erster Stelle haben, es sind $\frac{7!}{3!2!2!} = 210$, und die Anzahl der Anordnungen, die C an erster Stelle haben, das sind $\frac{7!}{3!3!1!} = 140$. Die Anzahl der Anordnungen, die A nicht an erster Stelle haben, ist daher $210 + 140 = 350$.

Beispiel 16: Jemand hat in seinem Zimmer 8 Lampen. Er hat 8 LED-Birnen, von denen je zwei in den Farben weiß, rot, grün und blau leuchten. Wie viele verschiedene Beleuchtungen gibt es? Wie viele gibt es, wenn die Tischlampe nicht rot leuchten soll?

Die 8 Lampen fassen wir als 8 Plätze auf, auf denen die LED-Birnen angeordnet werden. Die verschiedenen Beleuchtungsmöglichkeiten sind die verschiedenen Anordnungen der LED-Birnen. Daher gibt es $\frac{8!}{2!2!2!2!} = 2520$ mögliche Beleuchtungen.

Die Anzahl der möglichen Beleuchtungen mit einer roten LED-Birne in der Tischlampe erhält man, indem man die übrigen 7 LED-Birnen auf die übrigen 7 Lampen verteilt. Diese Anzahl ist $\frac{7!}{2!1!2!2!} = 630$. Daher gibt es $2520 - 630 = 1890$ mögliche Beleuchtungen mit nicht rot leuchtender Tischlampe.

II. Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeit ist eine Maßzahl für die Häufigkeit des Auftretens eines Ereignisses bei der Durchführung eines Zufallsexperiments. Entsprechend werden wir die Wahrscheinlichkeit auch definieren. Wir beweisen Sätze und Formeln für die Wahrscheinlichkeit und verwenden diese zum Rechnen von Beispielen. Die wichtigsten Sätze sind der Additionssatz, die Formel für gleichwahrscheinliche Ausfälle, der Multiplikationssatz, die Formel für die totale Wahrscheinlichkeit und die Formel von Bayes.

1. Zufallsexperiment, Ausfall, Ereignis

Ein Zufallsexperiment ist ein Experiment mit verschiedenen möglichen Ausfällen (Ergebnissen). Beispiele für Zufallsexperimente sind das Werfen eines Würfels oder die Lottoziehung. Die möglichen Ausfälle beim Würfeln sind die Augenzahlen 1 bis 6. Die möglichen Ausfälle des Zufallsexperiments Lottoziehung sind alle 6-elementigen Teilmengen der Menge $\{1, 2, 3, \dots, 45\}$. Die möglichen Ausfälle eines Zufallsexperiments fassen wir zu einer Menge zusammen, die üblicherweise mit Ω bezeichnet wird. Wir nennen Ω die Ausfallsmenge. Beim Zufallsexperiment Würfeln haben wir $\Omega = \{1, 2, 3, 4, 5, 6\}$. Beim Zufallsexperiment Lottoziehung ist Ω die Menge aller sechselementigen Teilmengen aus der Menge der ersten 45 natürlichen Zahlen.

Der zentrale Begriff der Wahrscheinlichkeitstheorie ist der des Ereignisses. Ereignisse werden üblicherweise in Worten beschrieben. Man spricht zum Beispiel vom Ereignis, eine gerade Zahl zu würfeln. Oft genügt diese Art der Beschreibung. Um die Wahrscheinlichkeit von Ereignissen berechnen zu können, ist es jedoch manchmal notwendig, in eine Mengensprache überzuwechseln. Ereignisse werden dann als Teilmengen der Ausfallsmenge Ω aufgefasst. Das Ereignis "gerade Zahl würfeln" ist zum Beispiel die Teilmenge $\{2, 4, 6\}$. Das Ereignis "die Lottoziehung ergibt nur Zahlen ≤ 7" ist die Teilmenge $\{\{1, 2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 7\}, \{1, 2, 3, 4, 6, 7\}, \{1, 2, 3, 5, 6, 7\}, \{1, 2, 4, 5, 6, 7\}, \{1, 3, 4, 5, 6, 7\}, \{2, 3, 4, 5, 6, 7\}\}$ aus der oben beschriebenen Ausfallsmenge Ω für das Zufallsexperiment Lottoziehung. Ein als Menge geschriebenes Ereignis A tritt genau dann ein, wenn das Zufallsexperiment einen Ausfall liefert, der in A liegt.

Wirft man zwei Würfel gleichzeitig, so nimmt man unterscheidbare Würfel (einen roten und einen grünen). Die Ausfälle des Zufallsexperiments sind dann Paare von Augenzahlen, wobei die Augenzahl des roten Würfels an die erste Stelle und die Augenzahl des grünen Würfels an die zweite Stelle geschrieben wird. Diese Vorgangsweise wird später das Berechnen von Wahrscheinlichkeiten erleichtern. Die Ausfallsmenge Ω ist dann $\{(i, j) : 1 \le i, j \le 6\}$. Das Ereignis "Augensumme = 4" ist die Teilmenge, die alle Paare (i, j) von Augenzahlen mit $i + j = 4$ enthält, also die Menge $\{(1, 3), (2, 2), (3, 1)\}$. Das Ereignis "Maximum der Augenzahlen ≤ 2" ist die Teilmenge $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$.

Genauso geht man bei mehrmaligem Münzenwerfen vor. Wirft man eine Münze dreimal hintereinander, so sind die Ausfälle Tripel, an deren erster Stelle das Ergebnis des ersten Wurfes, an deren zweiter Stelle das Ergebnis des zweiten Wurfes und an deren dritter Stelle das Ergebnis des dritten Wurfes steht. Nennt man die Seiten der Münze 0 und 1, so erhält man $\Omega = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$. Das Ereignis "keine gleichen Ergebnisse hintereinander" ist die Teilmenge $\{(0, 1, 0), (1, 0, 1)\}$.

Zieht man eine ungeordnete Stichprobe vom Umfang 2 aus der Menge $\{a, b, c, d\}$, dann ist die Ausfallsmenge $\Omega = \{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}\}$. Zieht man eine geordnete Stichprobe vom Umfang 2 ohne Zurücklegen aus $\{a, b, c, d\}$, dann ist die Ausfallsmenge $\Omega = \{(a, b), (a, c), (a, d), (b, a), (b, c), (b, d), (c, a), (c, b), (c, d), (d, a), (d, b), (d, c)\}$. Das Ereignis "aufeinanderfolgende Buchstaben ziehen" ist $\{(a, b), (b, c), (c, d)\}$.

Ereignisse kann man auch miteinander verknüpfen. Verwendet man die umgangssprachliche Beschreibung, so ist es naheliegend logische Operatoren ($\wedge$,$\vee$,$\neg$) zu verwenden. Arbeitet man mit der Mengenschreibweise, so wird man Mengenoperatoren verwenden. Mit $A \wedge B$ bezeichnet man das Ereignis, dass sowohl A als auch B eintritt. Ist A das Ereignis "gerade Zahl würfeln" und B das Ereignis "eine Zahl ≤ 3 würfeln", dann ist $A \wedge B$ das Ereignis "die Zahl 2 würfeln". In der Mengenschreibweise schreibt man dafür $A \cap B$. Man hat dann $A = \{2, 4, 6\}$, $B = \{1, 2, 3\}$ und $A \cap B = \{2\}$. Das Ereignis, sowohl A als auch B tritt ein, wird durch den Durchschnitt richtig dargestellt, da es ja genau die Ausfälle enthalten muss, die sowohl in A als auch in B liegen.

Das Ereignis, A tritt nicht ein, wird mit $\neg A$ bezeichnet. Im obigen Beispiel ist $\neg A$ das Ereignis "ungerade Zahl würfeln". In der Mengenschreibweise entspricht das der Komplementärmenge A', da es ja genau die Ausfälle enthalten muss, die A nicht enthält.

Genauso kann man andere Mengenoperationen interpretieren. Wir tun das in folgender Tabelle.

logische Schreibweise	Mengenschreibweise	ist das Ereignis, dass
$A \wedge B$	$A \cap B$	A und B eintreten
$A \vee B$	$A \cup B$	A oder B oder beide eintreten
$\neg A$	$A' = \Omega \setminus A$	A nicht eintritt
$A \wedge \neg B$	$A \setminus B$	A eintritt, aber B nicht
$\neg A \wedge \neg B$	$A' \cap B'$	weder A noch B eintritt

Die leere Menge ist das Ereignis, das nie eintritt. Man spricht daher auch vom unmöglichen Ereignis. Die Menge Ω ist das Ereignis, das immer eintritt. Man spricht daher vom sicheren Ereignis.

Zwei Ereignisse A und B heißen unvereinbar, wenn sie nicht gleichzeitig eintreten können. In der Mengensprache bedeutet das, dass die beiden Teilmengen A und B von Ω disjunkt sind, also $A \cap B = \emptyset$ gilt. Sie können ja genau dann nicht gleichzeitig eintreten, wenn es keinen Ausfall gibt, der in beiden Mengen liegt.

Mindestens eines der Ereignisse A oder B tritt ein, wenn jeder mögliche Ausfall entweder in A oder in B liegt. In der Mengensprache bedeutet das $A \cup B = \Omega$. Gilt sowohl $A \cap B = \emptyset$ als auch $A \cup B = \Omega$, dann tritt genau eines der beiden Ereignisse A oder B ein. Man sagt dann, die Ereignisse A und B bilden eine Zerlegung von Ω.

2. Wahrscheinlichkeit

Die Wahrscheinlichkeit eines Ereignisses ist eine Maßzahl für die Häufigkeit, mit der das Ereignis eintritt. Um den Begriff der Wahrscheinlichkeit zu definieren, gehen wir folgendermaßen vor. Wir wiederholen das Zufallsexperiment (zum Beispiel das Zufallsexperiment Würfeln) k Mal und zählen, wie oft ein Ereignis A (zum Beispiel das Ereignis "gerade Zahl würfeln") eintritt. Diese Anzahl bezeichnen wir mit $N_k(A)$. Der Quotient

$\frac{N_k(A)}{k}$ ist dann der Anteil (relative Häufigkeit) der Wiederholungen, bei denen A eintritt. Die Wahrscheinlichkeit $P(A)$ des Ereignisses A definieren wir dann durch

$$P(A) = \lim_{k\to\infty} \frac{N_k(A)}{k}$$

Da man ein Zufallsexperiment nicht unendlich oft wiederholen kann, lässt sich nicht entscheiden, ob dieser Grenzwert existiert. Wir nehmen seine Existenz einfach an.

Wir beweisen die grundlegenden Eigenschaften und Rechenregeln für die Wahrscheinlichkeit, die sich aus dieser Definition ergeben.

Satz 6: *Es gilt*
(a) $0 \le P(A) \le 1$ *für alle Ereignisse* A
(b) $P(\emptyset) = 0$ *und* $P(\Omega) = 1$

Beweis: Da $N_k(A)$ die Anzahl der insgesamt k Wiederholungen ist, bei denen A eintritt, folgt $0 \le N_k(A) \le k$ und daraus $0 \le \frac{N_k(A)}{k} \le 1$. Mit k gegen ∞ erhält man (a).
Das Ereignis $\emptyset$ tritt nie ein. Daher gilt $N_k(\emptyset) = 0$, woraus $P(\emptyset) = 0$ folgt. Das Ereignis Ω tritt immer ein. Daher gilt $N_k(\Omega) = k$, woraus $P(\Omega) = 1$ folgt. □

Satz 7 (Additionssatz) *Die Ereignisse $A_1, A_2, \dots, A_n$ seien unvereinbar (paarweise disjunkt), das heißt keine zwei dieser Ereignisse können gleichzeitig eintreten. Dann gilt*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

wobei wir die Mengenschreibweise verwendet haben. Genauso könnte man statt $\cup$ auch $\vee$ schreiben.

Beweis: Wir wiederholen das Zufallsexperiment k Mal. Die Anzahl der Wiederholungen, bei denen A_1 eintritt, ist $N_k(A_1)$. Die Anzahl der Wiederholungen, bei denen A_2 eintritt, ist $N_k(A_2)$ und so weiter. Daher ist $N_k(A_1) + N_k(A_2) + \cdots + N_k(A_n)$ die Anzahl der Wiederholungen, bei denen mindestens eines der Ereignisse $A_1, A_2, \dots, A_n$ eintritt. Keine der k Wiederholungen wird mehrfach gezählt, da die Ereignisse ja unvereinbar sind und daher bei jeder Wiederholung höchstens eines der Ereignisse $A_1, A_2, \dots, A_n$ eintritt.
Andererseits ist $A_1 \cup A_2 \cup \cdots \cup A_n$ gerade das Ereignis, dass mindestens eines der Ereignisse $A_1, A_2, \dots, A_n$ eintritt. Daraus folgt

$$N_k(A_1 \cup A_2 \cup \cdots \cup A_n) = N_k(A_1) + N_k(A_2) + \cdots + N_k(A_n)$$

Dividiert man durch k und lässt k gegen ∞ gehen, so folgt das gewünschte Resultat. □

Aus dem Additionssatz erhält man weitere Rechenregeln für die Wahrscheinlichkeit. Wir verwenden dazu wieder die Mengensprache.

Satz 8: *Seien A und B Ereignisse. Dann gilt*
(a) $P(A \setminus B) = P(A) - P(A \cap B)$
(b) $B \subseteq A \;\Rightarrow\; P(A \setminus B) = P(A) - P(B)$
(c) $B \subseteq A \;\Rightarrow\; P(B) \le P(A)$
(d) $P(A') = 1 - P(A)$
(e) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(f) $P(A \cup B) \le P(A) + P(B)$

Beweis: (a) Die Ereignisse $A \setminus B$ und $A \cap B$ sind disjunkt. Ihre Vereinigung ist A. Aus dem Additionssatz folgt daher $P(A) = P(A \setminus B) + P(A \cap B)$. Berechnet man daraus $P(A \setminus B)$, so hat man (a).

(b) Gilt $B \subseteq A$, dann auch $A \cap B = B$. Aus (a) erhalten wir $P(A \setminus B) = P(A) - P(B)$ und (b) ist gezeigt.

(c) Aus (b) erhalten wir $P(A) - P(B) = P(A \setminus B)$. Aus Satz 6 (a) folgt $P(A \setminus B) \geq 0$, womit (c) gezeigt ist.

(d) Die Mengen A und A' sind disjunkt und ihre Vereinigung ist Ω. Der Additionssatz ergibt daher $P(A)+P(A') = P(\Omega)$. Satz 6 (b) besagt, dass $P(\Omega) = 1$ gilt. Daraus erhalten wir $P(A') = 1 - P(A)$ und (d) ist gezeigt.

(e) Die Ereignisse B und $A \setminus B$ sind disjunkt. Ihre Vereinigung ist $A \cup B$. Daher erhalten wir aus dem Additionssatz, dass $P(A \cup B) = P(B) + P(A \setminus B)$ gilt. In (a) wurde $P(A \setminus B) = P(A) - P(A \cap B)$ gezeigt. Setzt man das ein, so hat man bereits (e).

(f) Da $P(A \cap B) \geq 0$ wegen Satz 6 (a) gilt, folgt (f) aus (e). $\square$

3. Gleichwahrscheinliche Ausfälle

Bei vielen Zufallsexperimenten haben alle Ausfälle die gleiche Wahrscheinlichkeit. Beispiele dafür sind das Werfen eines fairen Würfels und die Lottoziehung. Beim Werfen von zwei Würfeln und beim mehrmaligen Münzenwerfen haben wir die Ausfallsmenge Ω so gewählt, dass alle Ausfälle gleichwahrscheinlich sind.

Für eine endliche Menge X sei $|X|$ die Anzahl der Elemente von X. Der folgende Satz führt das Berechnen der Wahrscheinlichkeit auf das Abzählen der Elemente von Mengen zurück. Es ist der einzige Satz, der die Mengendarstellung der Ereignisse verlangt.

Satz 9 (Formel für gleichwahrscheinliche Ausfälle) *Ein Zufallsexperiment mit endlicher Ausfallsmenge Ω habe gleichwahrscheinliche Ausfälle. Sei $A \subseteq \Omega$ ein Ereignis. Dann gilt*

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

Beweis: Sei q die Wahrscheinlichkeit, mit der jeder der Ausfälle eintritt, oder genauer, mit der jede einelementige Teilmenge von Ω eintritt. Sei $k = |A|$ die Anzahl der Elemente der Menge A. Weiters seien $A_1, A_2, \dots, A_k$ die einelementigen Teilmengen von A. Da diese paarweise disjunkt sind und ihre Vereinigung gleich A ist, erhalten wir aus dem Additionssatz, dass $P(A) = P(A_1) + P(A_2) + \cdots + P(A_k)$ gilt. Da die Ereignisse A_j einelementig sind, gilt $P(A_j) = q$ für $1 \leq j \leq k$. Es folgt $P(A) = qk = q|A|$.

Damit ist $P(A)$ bereits berechnet. Es ist nur noch q zu bestimmen. Dazu setzen wir $A = \Omega$ und erhalten $P(\Omega) = q|\Omega|$. Nach Satz 6 gilt $P(\Omega) = 1$, sodass $q = \frac{1}{|\Omega|}$ folgt. Damit erhalten wir $P(A) = q|A| = \frac{|A|}{|\Omega|}$, was zu zeigen war. $\square$

Mit Hilfe dieses Satzes und den Formeln aus der Kombinatorik kann man jetzt viele Beispiele rechnen.

Beispiel 17: Wie groß ist die Wahrscheinlichkeit, eine gerade Zahl zu würfeln?

Die Ausfallsmenge beim Würfeln ist $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis "gerade Augenzahl" ist $A = \{2, 4, 6\}$. Da die Ausfälle gleichwahrscheinlich sind, gilt $P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6}$.

Beispiel 18: Es wird mit 2 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit,

(a) Augensumme 5 zu erhalten?

(b) dass mindestens eine 6 auftritt?

Die Ausfallsmenge ist $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$, die Menge aller Paare von Augenzahlen. Sie hat $6^2 = 36$ Elemente. Alle Ausfälle sind gleichwahrscheinlich. Wir wenden Satz 9 an.

(a) Um das Ereignis “Augensumme 5” zu bestimmen, müssen wir alle Paare von Augenzahlen finden, deren Summe 5 ist. Wir erhalten $A = \{(1,4),(2,3),(3,2),(4,1)\}$. Das ergibt $P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36}$.

(b) Wir bestimmen das Ereignis “mindestens eine 6”, indem wir Ω durchsuchen und erhalten $A = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(6,1),(6,2),(6,3),(6,4),(6,5)\}$. Das ergibt $P(A) = \frac{|A|}{|\Omega|} = \frac{11}{36}$.

Beispiel 19: Es wird mit 3 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit,
(a) Augensumme 5 zu erhalten?
(b) dass keine 6 auftritt?
(c) dass genau zwei Mal 6 auftritt?

Die Ausfallsmenge ist $\Omega = \{(i,j,k) : 1 \le i,j,k \le 6\}$, die Menge aller Tripel von Augenzahlen. Sie hat $6^3 = 216$ Elemente. Alle Ausfälle sind gleichwahrscheinlich. Wir können Satz 9 anwenden.

(a) Das Ereignis “Augensumme 5” besteht aus allen Tripeln von Augenzahlen, deren Summe 5 ist. Wir erhalten $A = \{(1,1,3),(1,3,1),(3,1,1),(1,2,2),(2,1,2),(2,2,1)\}$. Das ergibt $P(A) = \frac{|A|}{|\Omega|} = \frac{6}{216}$.

(b) Das Ereignis “keine 6” ist $A = \{(i,j,k) : 1 \le i,j,k \le 5\}$, die Menge aller Tripel, die keine 6 enthalten. Wegen $|A| = 5^3 = 125$ folgt $P(A) = \frac{|A|}{|\Omega|} = \frac{125}{216}$.

(c) Das Ereignis “zwei Mal 6” ist die Menge A aller Tripel, die genau zwei 6 enthält. Man kann die Menge A aufschreiben. Wir wollen aber versuchen, ihre Elemente zu zählen, ohne sie aufzuschreiben. Es gibt $\binom{3}{2}$ Möglichkeiten, die beiden Plätze auszuwählen, auf denen 6 steht. Ist eine Platzaufteilung festgelegt, dann gibt es für die beiden ausgewählten Plätze nur eine Möglichkeit, nämlich 6, und für den dritten Platz gibt es 5 Möglichkeiten, nämlich alle Augenzahlen außer 6. Für jede Platzaufteilung gibt es also $1 \cdot 1 \cdot 5 = 5$ Möglichkeiten, woraus $|A| = \binom{3}{2} \cdot 5 = 15$ folgt. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{15}{216}$.

Beispiel 20: Wie groß ist die Wahrscheinlichkeit mit sechs Würfeln sechs verschiedene Augenzahlen zu erhalten?

Die Ausfallsmenge Ω ist die Menge aller geordneten Stichproben vom Umfang 6 aus der Menge $\{1,2,3,4,5,6\}$ mit Zurücklegen. Also gilt $|\Omega| = 6^6$. Das Ereignis “sechs verschiedene Augenzahlen” ist die Menge A aller geordneten Stichproben vom Umfang 6 ohne Zurücklegen aus der Menge $\{1,2,3,4,5,6\}$. Daher gilt $|A| = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6!$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{6!}{6^6}$.

Bis jetzt haben wir nur Würfelbeispiele gerechnet. Diese entsprechen geordneten Stichproben aus der Menge $\{1,2,3,4,5,6\}$. Wir wollen jetzt ungeordnete Stichproben behandeln, wobei so gezogen wird, dass jede Teilmenge die gleiche Chance hat dranzukommen. Dann sind alle Ausfälle gleichwahrscheinlich und wir können Satz 9 anwenden.

Beispiel 21: Wie groß ist die Wahrscheinlichkeit, dass beim Lotto 6 aus 45 von den sechs Zahlen, die ich getippt habe, genau fünf gezogen werden?

Jetzt ist die Ausfallsmenge Ω die Menge aller ungeordneten Stichproben vom Umfang 6, also die Menge aller 6-elementigen Teilmengen aus der Menge der ersten 45 natürlichen Zahlen. Somit gilt $|\Omega| = \binom{45}{6}$. Das Ereignis “genau 5 meiner getippten Zahlen werden gezogen” ist die Menge A aller 6-elementigen Teilmengen, die fünf der 6 getippten und eine der 39 nicht getippten Zahlen enthalten. Es gibt $\binom{6}{5}$ fünfelementige Teilmengen aus den getippten Zahlen und $\binom{39}{1}$ einelementige Teilmengen aus den nichtgetippten Zahlen. Daraus folgt $|A| = \binom{6}{5}\binom{39}{1}$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{6}{5}\binom{39}{1}}{\binom{45}{6}}$.

Beispiel 22: Aus einem Kartenspiel mit 20 Karten (5 $\heartsuit$-Karten, 5 $\diamondsuit$-Karten, 5 $\clubsuit$-Karten und 5 $\spadesuit$-Karten) werden 7 Karten ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, 2 $\heartsuit$-Karten, 2 $\diamondsuit$-Karten und 3 $\clubsuit$-Karten zu erhalten?

Die Ausfallsmenge Ω ist die Menge aller 7-elementigen Teilmengen aus den 20 Karten. Somit gilt $|\Omega| = \binom{20}{7}$. Das gefragte Ereignis ist die Menge A aller 7-elementigen Teilmengen, die 2 $\heartsuit$-Karten, 2 $\diamondsuit$-Karten und 3 $\clubsuit$-Karten enthalten. Da es $\binom{5}{2}$ 2-elementige Teilmengen aus den 5 $\heartsuit$-Karten, $\binom{5}{2}$ 2-elementige Teilmengen aus den 5 $\diamondsuit$-Karten und $\binom{5}{3}$ 3-elementige Teilmengen aus den 5 $\clubsuit$-Karten gibt, erhalten wir $|A| = \binom{5}{2}\binom{5}{2}\binom{5}{3}$. Daraus ergibt sich dann $P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{5}{2}\binom{5}{2}\binom{5}{3}}{\binom{20}{7}}$.

Man könnte die letzten beiden Beispiele genausogut mit geordneten Stichproben ohne Zurücklegen rechnen. Im vorletzten Beispiel würde man ein $|\Omega|$ und ein $|A|$ erhalten, das jeweils mit 6! multipliziert ist. Im letzten Beispiel würde man ein $|\Omega|$ und ein $|A|$ erhalten, das jeweils mit 7! multipliziert ist. Die Wahrscheinlichkeit $P(A)$ wäre in beiden Beispielen dieselbe wie die, die wir berechnet haben.

Das ist immer so. Werden Stichproben ohne Zurücklegen gezogen, dann erhält man dieselben Wahrscheinlichkeiten, ob man die Stichproben als ungeordnete oder als geordnete auffasst. Wir werden Stichproben ohne Zurücklegen daher immer als ungeordnete auffassen, da sich in diesem Fall die Anzahlen leichter berechnen lassen.

4. Geometrische Wahrscheinlichkeit

Wieder hat das Zufallsexperiment gleichwahrscheinliche Ausfälle, aber die Ausfallsmenge Ω ist nicht mehr endlich, sondern ein Intervall oder eine Teilmenge des $\mathbb{R}^2$ (könnte auch $\mathbb{R}^n$ für $n \geq 3$ sein). Analog zu Satz 9 verwenden wir wieder die Formel

$$P(A) = \frac{|A|}{|\Omega|} \quad \text{für} \quad A \subseteq \Omega$$

wobei $|X|$ die Länge von X bedeutet, wenn X ein Intervall ist, und $|X|$ die Fläche von X bedeutet, wenn X eine Teilmenge des $\mathbb{R}^2$ ist.

Wir untersuchen einige Beispiele.

Beispiel 23: Jemand lässt seine Uhr auslaufen. Wie groß ist die Wahrscheinlichkeit, dass der große Zeiger zwischen 2 und 4 stehen bleibt?

Die Ausfallsmenge ist $\Omega = [0, 12)$, die Menge aller Punkte, in denen der große Zeiger stehen bleiben kann. Dieses Intervall hat Länge 12. Das gefragte Ereignis ist das Intervall $A = (2, 4)$. Es hat Länge 2. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{12}$.

Beispiel 24: Zwei Personen vereinbaren, zufällig zwischen 5 und 6 Uhr zu kommen und 15 Minuten zu warten. Wie groß ist die Wahrscheinlichkeit, dass sie einander treffen?

Die Ausfallsmenge ist $\Omega = [5, 6] \times [5, 6] = \{(x, y) : 5 \leq x, y \leq 6\}$, wobei x der Zeitpunkt ist, zu dem die erste Person kommt, und y der Zeitpunkt, zu dem die zweite Person kommt. Durch welche Teilmenge A wird das Ereignis "sie treffen einander" dargestellt? Da jede eine Viertelstunde wartet, treffen sie einander genau dann, wenn ihre Ankunftszeitpunkte

x und y höchstens $\frac{1}{4}$ voneinander entfernt sind. Das heißt $A = \{(x, y) \in \Omega : |x - y| \leq \frac{1}{4}\}$. Wegen $|\Omega| = 1$ und $|A| = 1 - \frac{3}{4} \cdot \frac{3}{4} = \frac{7}{16}$ folgt $P(A) = \frac{7}{16}$.

Beispiel 25: Sei $0 < t < 1$. Zwei Zahlen x und y werden zufällig im Intervall $[0, 1]$ gewählt. Wie groß ist die Wahrscheinlichkeit, dass ihr Produkt $\leq t$ ist?

Die Ausfallsmenge ist $\Omega = [0, 1] \times [0, 1] = \{(x, y) : 0 \leq x, y \leq 1\}$. Das gefragte Ereignis ist $A = \{(x, y) \in \Omega : xy \leq t\}$. Es gilt $|\Omega| = 1$. Die Fläche $|A|$ von A erhält man als Summe der Fläche des Rechtecks $[0, t] \times [0, 1]$ und der Fläche über dem Intervall $[t, 1]$ unter der Funktion $x \mapsto \frac{t}{x}$. Diese Fläche ist $t + \int_t^1 \frac{t}{x}\,\mathrm{d}x = t - t \ln t$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = t - t \ln t$.

5. Bedingte Wahrscheinlichkeit

Wir führen ein Zufallsexperiment durch und interessieren uns dafür, ob das Eintreten eines Ereignisses B ein anderes Ereignis A begünstigt. Dazu wiederholen wir das Zufallsexperiment k Mal. Wir berücksichtigen jedoch nur die Wiederholungen, wo das Ereignis B eintritt. Unter diesen zählen wir die Häufigkeit des Ereignisses A. Diese ist die Anzahl der Wiederholungen, wo A und B eintreten, also $N_k(A \cap B)$, wobei wir die Bezeichnung aus Kapitel 2 verwenden. Die bedingte relative Häufigkeit des Eintretens von A unter B ist dann $N_k(A \cap B)/N_k(B)$. Lässt man k gegen ∞ gehen, so erhält man wieder eine Wahrscheinlichkeit, diesmal die bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B (oder gegeben B), die mit $P(A|B)$ bezeichnet wird. Wir definieren also

$$P(A|B) = \lim_{k \to \infty} N_k(A \cap B)/N_k(B)$$

Die bedingte Wahrscheinlichkeit $P(A|B)$ ist eine Maßzahl für die Häufigkeit des Eintretens von A unter den Wiederholungen des Zufallsexperiments, bei denen B eintritt.

Gilt $P(A|B) > P(A)$, dann sagt man, dass das Ereignis A durch das Eintreten von B begünstigt wird. Gilt $P(A|B) < P(A)$, dann sagt man, dass das Ereignis A durch das Eintreten von B benachteiligt wird. Gilt $P(A|B) = P(A)$, dann sagt man, die Ereignisse A und B sind unabhängig.

Aus den Definitionen von Wahrscheinlichkeit und bedingter Wahrscheinlichkeit folgt

$$P(A|B) = \lim_{k \to \infty} \frac{N_k(A \cap B)}{N_k(B)} = \lim_{k \to \infty} \frac{N_k(A \cap B)/k}{N_k(B)/k} = P(A \cap B)/P(B)$$

Wir haben die Ereignisse A und B unabhängig genannt, wenn $P(A|B) = P(A)$ gilt. Wegen $P(A|B) = P(A \cap B)/P(B)$ ist das äquivalent zu $P(A \cap B) = P(A)P(B)$.

Die Bedingung B kann man als schon vorhandene Information auffassen. Wird nach der Wahrscheinlichkeit eines Ereignisses A gefragt, und weiß man, dass B eingetreten ist, so berechnet man nicht $P(A)$ sondern $P(A|B)$. Dazu folgendes Beispiel und zuvor ein Satz, der bei der Berechnung der bedingten Wahrscheinlichkeit hilft.

Satz 10: *Sind alle Ausfälle gleichwahrscheinlich, dann gilt* $P(A|B) = \frac{|A \cap B|}{|B|}$.

Beweis: Da die Ausfälle gleichwahrscheinlich sind, haben wir $P(A \cap B) = \frac{|A \cap B|}{|\Omega|}$ und $P(B) = \frac{|B|}{|\Omega|}$. Es folgt $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|}$. □

Beispiel 26: Es wird zweimal gewürfelt. Man erfährt, dass die Augensumme 7 ist. Wie groß ist die Wahrscheinlichkeit, dass mindestens einmal 6 aufgetreten ist.

Die Ausfallsmenge Ω ist die Menge aller Paare von Augenzahlen. Sie hat 36 Elemente. Gefragt wird nach dem Ereignis A, dass mindestens einmal 6 aufgetreten ist. Das Ereignis A wurde in Beispiel 18 bestimmt. Wir haben die Information, dass die Augensumme 7 ist, also dass das Ereignis $B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ eingetreten ist. Zu berechnen ist daher $P(A|B)$. Es gilt $A \cap B = \{(1,6),(6,1)\}$, also $|A \cap B| = 2$, und $|B| = 6$. Aus Satz 10 folgt $P(A|B) = \frac{2}{6}$.

Einer der wichtigsten Sätze, die mit bedingten Wahrscheinlichkeiten zu tun haben, ist der sogenannte Multiplikationssatz. Er berechnet die Wahrscheinlichkeit des Durchschnitts von Ereignissen.

Satz 11 (Multiplikationssatz) *Für Ereignisse $A_1, A_2, \dots, A_n$ gilt*

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

Es wird dabei vorausgesetzt, dass die Ereignisse, die als Bedingungen auftreten, Wahrscheinlichkeit > 0 haben.

Beweis: Wir haben $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$, ebenso $P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}$ und so fort bis $P(A_n|A_1 \cap \dots \cap A_{n-1}) = \frac{P(A_1 \cap A_2 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})}$. Setzt man das in die rechte Seite der Formel ein, so kürzt sich alles weg, es bleibt nur die linke Seite übrig. $\square$

Typische Anwendungsbeispiele für den Multiplikationssatz sind geordnete Stichproben. Dazu gehört auch wiederholtes Würfeln und Münzenwerfen.

Beispiel 27: Aus einer Menge von 2 roten und 5 gelben Kugeln zieht man eine geordnete Stichprobe vom Umfang 3. Wie groß ist die Wahrscheinlichkeit, zuerst eine rote, dann eine gelbe und zuletzt wieder eine rote Kugel zu ziehen?

Sei A_1 das Ereignis "erste Kugel ist rot", A_2 das Ereignis "zweite Kugel ist gelb" und A_3 das Ereignis "dritte Kugel ist rot". Gesucht ist $P(A_1 \cap A_2 \cap A_3)$. Aus dem Multiplikationssatz folgt $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$. Wir müssen also die drei Wahrscheinlichkeiten, deren Produkt auf der rechten Seite steht, berechnen.

Wir berechnen diese Wahrscheinlichkeiten zuerst für den Fall, dass nicht zurückgelegt wird. Da die Menge 2 rote und 5 gelbe Kugeln enthält, folgt $P(A_1) = \frac{2}{7}$.

Wir berechnen $P(A_2|A_1)$. Die Bedingung A_1 besagt, dass beim ersten Zug eine rote Kugel gezogen wurde, die Menge beim zweiten Zug also nur mehr 1 rote und 5 gelbe Kugeln enthält. Daraus ergibt sich $P(A_2|A_1) = \frac{5}{6}$ als bedingte Wahrscheinlichkeit, beim zweiten Zug eine gelbe Kugel zu ziehen.

Wir berechnen $P(A_3|A_1 \cap A_2)$. Die Bedingung $A_1 \cap A_2$ besagt, dass bei den ersten beiden Zügen eine rote und eine gelbe Kugel gezogen wurden, die Menge beim dritten Zug also nur mehr 1 rote und 4 gelbe Kugeln enthält. Daraus ergibt sich $P(A_3|A_1 \cap A_2) = \frac{1}{5}$ als bedingte Wahrscheinlichkeit, beim dritten Zug eine rote Kugel zu ziehen.

Die gesuchte Wahrscheinlichkeit, bei Ziehen ohne Zurücklegen zuerst eine rote, dann eine gelbe und zuletzt wieder eine rote Kugel zu ziehen, ist also $\frac{2}{7} \cdot \frac{5}{6} \cdot \frac{1}{5}$.

Jetzt behandeln wir den Fall, dass zurückgelegt wird. Die Menge enthält 2 rote und 5 gelbe Kugeln, daher gilt wieder $P(A_1) = \frac{2}{7}$. Da zurückgelegt wird, wird die Menge durch den ersten Zug nicht verändert. Beim zweiten Mal wird wieder aus derselben Menge gezogen, sodass $P(A_2|A_1) = \frac{5}{7}$ gilt. Dasselbe gilt für den dritten Zug. Auch beim dritten Mal wird aus der ursprünglichen Menge gezogen, sodass $P(A_3|A_1 \cap A_2) = \frac{2}{7}$ gilt. Die gesuchte Wahrscheinlichkeit, bei Ziehen mit Zurücklegen zuerst eine rote, dann eine gelbe und zuletzt wieder eine rote Kugel zu ziehen, ist also $\frac{2}{7} \cdot \frac{5}{7} \cdot \frac{2}{7}$.

Bei vielen Beispielen tritt der Multiplikationssatz in Kombination mit dem Additionssatz auf. Dabei wird das Ereignis, nach dessen Wahrscheinlichkeit gefragt wird, in unvereinbare Teilereignisse zerlegt. Die Wahrscheinlichkeiten der Teilereignisse berechnet man mit dem Multiplikationssatz. Der Additionssatz liefert dann die gesuchte Wahrscheinlichkeit als Summe der Wahrscheinlichkeiten der Teilereignisse. Oft wird die Lösung solcher Beispiele in Form eines Baumes aufgeschrieben. Hier wird eine Tabellenschreibweise verwendet.

Beispiel 28: Aus einer Menge von 2 roten und 5 gelben Kugeln zieht man eine geordnete Stichprobe vom Umfang 3. Wie groß ist die Wahrscheinlichkeit, eine ungerade Anzahl von gelben Kugeln zu erhalten?

Wir kürzen "rote Kugel" mit R und "gelbe Kugel" mit G ab. Das Ereignis "ungerade Anzahl von gelben Kugeln" tritt genau dann ein, wenn GGG, GRR, RGR oder RRG gezogen wird. Es lässt sich in diese vier unvereinbaren Teilereignisse zerlegen.

Wir behandeln zuerst den Fall, dass ohne Zurücklegen gezogen wird. In Beispiel 27 wurde $\frac{2}{7} \cdot \frac{5}{6} \cdot \frac{1}{5}$ als Wahrscheinlichkeit für das Ereignis RGR gefunden. Genauso findet man die Wahrscheinlichkeiten der anderen Teilereignisse. Wir tun das in folgender Tabelle:

Teilereignis	Wahrscheinlichkeit nach dem Multiplikationssatz
GGG	$\frac{5}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{60}{210}$
GRR	$\frac{5}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} = \frac{10}{210}$
RGR	$\frac{2}{7} \cdot \frac{5}{6} \cdot \frac{1}{5} = \frac{10}{210}$
RRG	$\frac{2}{7} \cdot \frac{1}{6} \cdot \frac{5}{5} = \frac{10}{210}$
Summe (nach dem Additionssatz)	$= \frac{90}{210}$

Da die Teilereignisse unvereinbar sind, erhält man die Wahrscheinlichkeit des gefragten Ereignisses mit dem Additionssatz durch Aufsummieren der Wahrscheinlichkeiten der Teilereignisse. Das wurde bereits in der Tabelle durchgeführt. Die Wahrscheinlichkeit, eine ungerade Anzahl von gelben Kugeln zu erhalten, ist $\frac{90}{210}$, wenn nicht zurückgelegt wird.

Zieht man die Stichprobe mit Zurücklegen, dann kann man genauso vorgehen. Die entsprechende Tabelle sieht so aus:

Teilereignis	Wahrscheinlichkeit nach dem Multiplikationssatz
GGG	$\frac{5}{7} \cdot \frac{5}{7} \cdot \frac{5}{7} = \frac{125}{343}$
GRR	$\frac{5}{7} \cdot \frac{2}{7} \cdot \frac{2}{7} = \frac{20}{343}$
RGR	$\frac{2}{7} \cdot \frac{5}{7} \cdot \frac{2}{7} = \frac{20}{343}$
RRG	$\frac{2}{7} \cdot \frac{2}{7} \cdot \frac{5}{7} = \frac{20}{343}$
Summe (nach dem Additionssatz)	$= \frac{185}{343}$

Die Wahrscheinlichkeit, eine ungerade Anzahl von gelben Kugeln zu erhalten, ist $\frac{185}{343}$, wenn zurückgelegt wird.

Beispiel 29: Eine unfaire Münze, bei der Wappen mit Wahrscheinlichkeit $\frac{1}{3}$ und Zahl mit Wahrscheinlichkeit $\frac{2}{3}$ auftritt, wird fünf Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass genau dreimal Wappen auftritt?

Wir kürzen "Wappen" mit W und "Zahl" mit Z ab. Genau dreimal Wappen tritt auf, wenn das fünfmalige Werfen eine der Folgen WWWZZ, WWZWZ, WWZZW, WZWWZ, WZWZW, WZZWW, ZWWWZ, ZWWZW, ZWZWW, ZZWWW ergibt. Das sind die zehn unvereinbaren Ereignisse, in die sich das Ereignis "genau dreimal Wappen" zerlegen lässt. Die Berechnung der Wahrscheinlichkeiten führen wir in folgender Tabelle durch:

Teilereignis	Wahrscheinlichkeit nach dem Multiplikationssatz
WWWZZ	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}=\frac{4}{243}$
WWZWZ	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{4}{243}$
WWZZW	$\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}=\frac{4}{243}$
WZWWZ	$\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{4}{243}$
WZWZW	$\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}=\frac{4}{243}$
WZZWW	$\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{4}{243}$
ZWWWZ	$\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}=\frac{4}{243}$
ZWWZW	$\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}=\frac{4}{243}$
ZWZWW	$\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{4}{243}$
ZZWWW	$\frac{2}{3}\cdot\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{4}{243}$
Summe (nach dem Additionssatz)	$=\frac{40}{243}$

Die Wahrscheinlichkeit, dass genau zweimal Wappen auftritt, ist also $\frac{40}{243}$.

Beispiel 30: Aus einer Menge von 3 roten, 2 gelben und 2 schwarzen Kugeln ziehen wir solange ohne Zurücklegen, bis eine schwarze Kugel kommt. Wie groß ist die Wahrscheinlichkeit, eine gelbe Kugel zu ziehen?

Das Ereignis, eine gelbe Kugel zu ziehen bevor eine schwarze kommt, zerfällt in die Teilereignisse, die den Zugfolgen G, RG, RRG und RRRG entsprechen. Die Zugfolgen sind verschieden lang. Die Rechnung führen wir in folgender Tabelle durch

Teilereignis	Wahrscheinlichkeit nach dem Multiplikationssatz
G	$\frac{2}{7}=\frac{60}{210}$
RG	$\frac{3}{7}\cdot\frac{2}{6}=\frac{30}{210}$
RRG	$\frac{3}{7}\cdot\frac{2}{6}\cdot\frac{2}{5}=\frac{12}{210}$
RRRG	$\frac{3}{7}\cdot\frac{2}{6}\cdot\frac{1}{5}\cdot\frac{2}{4}=\frac{3}{210}$
Summe (nach dem Additionssatz)	$=\frac{105}{210}$

Die Wahrscheinlichkeit, eine gelbe Kugel zu ziehen bevor eine schwarze Kugel kommt, ist somit $\frac{105}{210}=\frac{1}{2}$.

Beispiel 31: Anna und Barbara haben beide die gleiche Schachtel mit vier weißen und einer schwarzen Kugel. Jede zieht zufällig aus ihrer Schachtel ohne Zurücklegen und zwar zuerst einmal Anna, dann zweimal Barbara, dann zweimal Anna, dann zweimal Barbara, und immer so weiter. Wer zuerst eine schwarze Kugel zieht gewinnt. Man berechne die Wahrscheinlichkeit, dass Barbara gewinnt.

Die Zugfolgen, die zu einem Sieg von Barbara führen, sind W̃S, W̃WS, W̃WWW̃W̃S und W̃WWW̃W̃WS, wobei ˜ bedeutet, dass Anna zieht. Sonst zieht Barbara. Die Wahrscheinlichkeit, dass Anna W̃W̃W̃ zieht, ist $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3}$. Die Wahrscheinlichkeit, dass Barbara WWS zieht, ist $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3}$. Die Wahrscheinlichkeit für die Zugfolge W̃WWW̃W̃S ist somit $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3}$. Die vollständige Rechnung findet man in folgender Tabelle.

Teilereignis	Wahrscheinlichkeit nach dem Multiplikationssatz
W̃S	$\frac{4}{5} \cdot \frac{1}{5} = \frac{4}{25}$
W̃WS	$\frac{4}{5} \cdot \frac{4}{5} \cdot \frac{1}{4} = \frac{4}{25}$
W̃WWW̃W̃S	$\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{2}{25}$
W̃WWW̃W̃WS	$\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{25}$
Summe (nach dem Additionssatz)	$= \frac{12}{25}$

Barbara gewinnt mit Wahrscheinlichkeit $\frac{12}{25}$.

6. Totale Wahrscheinlichkeit

Manches Mal sind für ein Ereignis A bedingte Wahrscheinlichkeiten vorgegeben und man soll daraus die (totale) Wahrscheinlichkeit des Ereignisses A bestimmen. Wir leiten die dafür notwendigen Formeln her. Dazu benötigen wir folgende Definition.

Definition: Man sagt, die Ereignisse $B_1, B_2, \ldots, B_n$ bilden eine Zerlegung der Ausfallsmenge Ω, wenn sie paarweise disjunkt sind und wenn $\bigcup_{j=1}^n B_j = \Omega$ gilt.

Man kann das auch so ausdrücken: Die Ereignisse $B_1, B_2, \ldots, B_n$ bilden eine Zerlegung der Ausfallsmenge Ω, wenn jeder Ausfall in genau einer dieser Mengen liegt. Das aber bedeutet, dass genau eines der Ereignisse $B_1, B_2, \ldots, B_n$ eintritt.

Satz 12: *Es liege eine Zerlegung $B_1, B_2, \ldots, B_n$ der Ausfallsmenge Ω vor. Sei weiters $A \subseteq \Omega$ ein Ereignis. Dann gilt*
(a) $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$
(b) $P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)}$ *für* $1 \leq j \leq n$
Man nennt (a) *die Formel für die totale Wahrscheinlichkeit und* (b) *die Formel von Bayes.*

Beweis: Wir zeigen (a). Es gilt $\sum_{j=1}^n P(A|B_j)P(B_j) = \sum_{j=1}^n P(A \cap B_j)$ nach Definition der bedingten Wahrscheinlichkeit. Da die Ereignisse B_j paarweise disjunkt sind, sind es auch die Ereignisse $A \cap B_j$. Daher folgt $\sum_{j=1}^n P(A \cap B_j) = P(\bigcup_{j=1}^n (A \cap B_j))$ aus dem Additionssatz. Wegen $\bigcup_{j=1}^n B_j = \Omega$ folgt $\bigcup_{j=1}^n (A \cap B_j) = A \cap \bigcup_{j=1}^n B_j = A \cap \Omega = A$. Setzt man das zusammen, dann ergibt sich (a).

Es gilt $P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}$ und $P(A|B_j) = \frac{P(A \cap B_j)}{P(B_j)}$ nach Definition der bedingten Wahrscheinlichkeit. Daraus folgt (b) sofort. □

Wir rechnen einige typische Beispiele zu diesen Formeln.

Beispiel 32: Drei Maschinen stellen das gleiche Produkt her. Die Tagesproduktion von Maschine I beträgt 8000 Stück mit einem Ausschussanteil von 2%, die von Maschine II 12000 Stück mit einem Ausschussanteil von 1% und die von Maschine III 30000 Stück mit einem Ausschussanteil von 5%. Mit welcher Wahrscheinlichkeit ist ein der Gesamtproduktion zufällig entnommenes Stück Ausschuss?

Sei B_1 das Ereignis, dass das zufällig entnommene Stück von Maschine I stammt, B_2 das Ereignis, dass es von Maschine II stammt, und B_3 das Ereignis, dass es von Maschine III stammt. Es tritt genau eines dieser drei Ereignisse ein, sie bilden eine Zerlegung. Aus der Angabe folgt $P(B_1) = \frac{8000}{50000} = \frac{4}{25}$, $P(B_2) = \frac{12000}{50000} = \frac{6}{25}$ und $P(B_3) = \frac{30000}{50000} = \frac{15}{25}$.

Wenn das zufällig gewählte Stück von Maschine I stammt, dann ist es mit Wahrscheinlichkeit $\frac{2}{100}$ Ausschuss. Wir haben also $P(A|B_1) = \frac{2}{100}$. Genauso erhält man $P(A|B_2) = \frac{1}{100}$ und $P(A|B_3) = \frac{5}{100}$. Mit Hilfe der Formel für die totale Wahrscheinlichkeit erhalten wir $P(A) = P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+P(A|B_3)P(B_3) = \frac{2}{100}\frac{4}{25}+\frac{1}{100}\frac{6}{25}+\frac{5}{100}\frac{15}{25} = \frac{89}{2500}$

Das ist die Wahrscheinlichkeit, dass ein zufällig entnommenes Stück Ausschuss ist. Man kann diese Zahl auch als Ausschussanteil in der Gesamtproduktion interpretieren.

Beispiel 33: Eine Versicherung teilt die Autofahrer in zwei Typen ein, in Risikofahrer und in Sicherheitsfahrer. Die Wahrscheinlichkeit, dass ein Sicherheitsfahrer in einem Jahr einen Unfall hat, ist 0.06. Die Wahrscheinlichkeit, dass ein Risikofahrer in einem Jahr einen Unfall hat, ist 0.6. Die Versicherung weiß aus Erfahrung, dass $\frac{5}{6}$ der Autofahrer Sicherheitsfahrer und $\frac{1}{6}$ der Autofahrer Risikofahrer sind. Ein Autofahrer schließt eine Versicherung ab (man sieht ihm natürlich nicht an, von welchem Typ er ist). Wie groß ist die Wahrscheinlichkeit, dass er nächstes Jahr einen Unfall haben wird?

Sei B_1 das Ereignis, dass der Autofahrer ein Sicherheitsfahrer ist, und B_2 das Ereignis, dass der Autofahrer ein Risikofahrer ist. Es tritt genau eines dieser beiden Ereignisse ein. Die Ereignisse B_1 und B_2 bilden eine Zerlegung.

Gefragt ist nach der Wahrscheinlichkeit des Ereignisses A, dass der Autofahrer innerhalb des nächsten Jahres einen Unfall hat. Wenn der Autofahrer ein Sicherheitsfahrer ist, dann ist diese Wahrscheinlichkeit 0.06, das heißt $P(A|B_1) = 0.06$. Wenn der Autofahrer ein Risikofahrer ist, dann ist diese Wahrscheinlichkeit 0.6, das heißt $P(A|B_2) = 0.6$. Jetzt können wir in die Formel für die totale Wahrscheinlichkeit einsetzen und erhalten

$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0.06 \cdot \frac{5}{6} + 0.6 \cdot \frac{1}{6} = 0.15$.

Die Wahrscheinlichkeit für einen Unfall im nächsten Jahr beträgt also 0.15.

Beispiel 34: Seit sich der Autofahrer aus dem letzten Beispiel versichern ließ, ist ein Jahr vergangen. Es hat sich herausgestellt, dass er einen Unfall hatte. Wie groß ist die Wahrscheinlichkeit, dass der Autofahrer ein Risikofahrer ist?

Wonach wird hier gefragt? Wir haben die Information, dass der Autofahrer einen Unfall hatte. Gefragt ist daher die bedingte Wahrscheinlichkeit dafür, dass er ein Risikofahrer ist. Verwendet man die Bezeichnung aus dem letzten Beispiel, dann ist das die bedingte Wahrscheinlichkeit $P(B_2|A)$.

Die Formel von Bayes besagt $P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)}$. Aus dem letzten Beispiel wissen wir $P(A|B_2)P(B_2) = 0.6 \cdot \frac{1}{6} = 0.1$ und $P(A) = 0.15$. Wir erhalten $P(B_2|A) = \frac{0.1}{0.15} = \frac{2}{3}$. Nach diesem Unfall ist der Autofahrer verdächtig, ein Risikofahrer zu sein.

Beispiel 35: Eine Firma stellt USB-Sticks her, wobei der Ausschussanteil 30 % beträgt. Damit diese nicht in den Verkauf gehen, findet eine Kontrolle statt, bei der ein defekter USB-Stick mit Wahrscheinlichkeit 0.97 als solcher erkannt und ausgeschieden wird, bei der aber auch ein intakter USB-Stick mit Wahrscheinlichkeit 0.05 irrtümlich ausgeschieden wird. Wie groß ist die Wahrscheinlichkeit, dass ein USB-Stick, der in den Verkauf geht, intakt ist?

Sei B_1 das Ereignis, dass ein USB-Stick intakt ist, und B_2 das Ereignis, dass ein USB-Stick defekt ist. Dann gilt $P(B_1) = \frac{7}{10}$ und $P(B_2) = \frac{3}{10}$. Die Ereignisse B_1 und B_2 bilden eine Zerlegung. Es tritt ja genau eines ein. Sei A das Ereignis, dass ein USB-Stick durch die Kontrolle geht, sodass $P(B_1|A)$ die gesuchte bedingte Wahrscheinlichkeit ist. Aus der Angabe erhalten wir $P(A|B_1) = 0.95$. Das ist die Wahrscheinlichkeit, dass ein intakter USB-Stick durch die Kontrolle geht. Weiters erhalten wir $P(A|B_2) = 0.03$. Das ist die Wahrscheinlichkeit, dass ein defekter USB-Stick durch die Kontrolle geht. Wir müssen zuerst $P(A)$ mit Hilfe der Formel für die totale Wahrscheinlichkeit berechnen: $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0.95 \cdot \frac{7}{10} + 0.03 \cdot \frac{3}{10} = 0.674$. Jetzt ergibt sich $P(B_1|A) = \frac{0.95 \cdot 0.7}{0.674} = 0.987$ aus der Formel von Bayes.

Die Wahrscheinlichkeit, dass ein USB-Stick, der in den Verkauf geht, auch intakt ist, beträgt 0.987.

Beispiel 36: Morsezeichen Punkt und Strich werden im Verhältnis 3:4 gesendet. Durch Übertragungsfehler wird Punkt mit Wahrscheinlichkeit $\frac{1}{9}$ zu Strich und Strich mit Wahrscheinlichkeit $\frac{1}{8}$ zu Punkt. Der Empfänger registriert Punkt. Mit welcher Wahrscheinlichkeit wurde Punkt gesendet?

Sei B_1 das Ereignis, dass Punkt gesendet wurde, und B_2 das Ereignis, dass Strich gesendet wurde. Die Ereignisse B_1 und B_2 bilden dann eine Zerlegung, da genau eines eintritt. Aus der Angabe entnehmen wir $P(B_1) = \frac{3}{7}$ und $P(B_2) = \frac{4}{7}$. Sei A das Ereignis, dass der Empfänger Punkt registriert. Dann folgt $P(A|B_1) = \frac{8}{9}$ und $P(A|B_2) = \frac{1}{8}$ aus der Angabe. Gefragt ist nach der bedingten Wahrscheinlichkeit $P(B_1|A)$. Bevor wir die Formel von Bayes anwenden, berechnen wir $P(A)$ nach der Formel für die totale Wahrscheinlichkeit: $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = \frac{8}{9} \cdot \frac{3}{7} + \frac{1}{8} \cdot \frac{4}{7} = \frac{19}{42}$. Die Formel von Bayes ergibt jetzt $P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{8}{9} \cdot \frac{3}{7} \cdot \frac{42}{19} = \frac{16}{19}$.

Die Wahrscheinlichkeit, dass auch Punkt gesendet wurde, wenn der Empfänger Punkt registriert, ist also $\frac{16}{19}$.

III. Zufallsvariable

Oft hat man es mit Zufallsexperimenten zu tun, deren Ausfälle Messwerte oder Anzahlen sind, jedenfalls reelle Zahlen. Das ermöglicht es, den Ausfall des Zufallsexperiments mit einer Variablen zu bezeichnen und mit diesen Variablen zu rechnen.

1. Darstellung von Ereignissen durch Zufallsvariable

Um das Ereignis "höchstens k schwarze Kugeln in der Stichprobe" einfach aufschreiben zu können, bezeichnet man die Anzahl der schwarzen Kugeln in der Stichprobe mit X. Das Ereignis lässt sich dann einfach durch $X \leq k$ ausdrücken. Diese Variablen, die man zur Beschreibung von Ereignissen verwendet, heißen Zufallsvariable. Sie bezeichnen immer den Ausfall eines Zufallsexperiments. Für Zufallsvariable werden üblicherweise Großbuchstaben verwendet.

Mit Variablen kann man auch rechnen. Es gilt zum Beispiel

$$3X + 5 \leq 11 \Leftrightarrow 3X \leq 6 \Leftrightarrow X \leq 2$$

Die durch diese drei Ungleichungen dargestellten Ereignisse sind gleich. Gleiche Ereignisse haben gleiche Wahrscheinlichkeit. Daher gilt

$$P(3X + 5 \leq 11) = P(3X \leq 6) = P(X \leq 2)$$

An diesem Beispiel sieht man, wie man Ereignisse umformen kann.

Die Darstellung der Ereignisse mit Hilfe von Zufallsvariablen entspricht der umgangssprachlichen Beschreibung. Sie wird in eine mathematische Darstellung mit Hilfe von Zufallsvariablen übersetzt. Um die so dargestellten Ereignisse miteinander zu verknüpfen, verwenden wir die logischen Zeichen. Der Additionssatz und die daraus folgenden Rechenregeln aus Satz 8 gelten natürlich weiterhin, da sie ja nicht davon abhängen, wie man Ereignisse darstellt. So gilt zum Beispiel

$$X \leq 3 \Leftrightarrow X \leq 2 \;\vee\; X \in (2, 3]$$

wobei die Ereignisse $X \leq 2$ und $X \in (2, 3]$ unvereinbar sind. Aus dem Additionssatz folgt

$$P(X \leq 3) = P(X \leq 2) + P(X \in (2, 3])$$

Ebenso ist $X \leq 3$ das Gegenereignis zu $X > 3$. Aus Satz 8 (d) folgt daher

$$P(X > 3) = 1 - P(X \leq 3)$$

Im Folgenden werden wir solche Umformungen häufig verwenden.

2. Wahrscheinlichkeitsvektoren und Wahrscheinlichkeitsdichten

Wir haben gesehen, wie man Ereignisse mit Hilfe von Zufallsvariablen darstellen kann und wie man die so dargestellten Ereignisse umformt. Jetzt geht es darum die Wahrscheinlichkeiten dieser Ereignisse auszurechnen. Dazu unterscheiden wir diskrete und kontinuierliche Zufallsvariable. Bei diskreten Zufallsvariablen liegen die möglichen Werte, also die Ausfälle des Zufallsexperiments, in einer endlichen oder abzählbaren Menge, während sie bei einer kontinuierlichen Zufallsvariable ein (oft unbeschränktes) Intervall ausfüllen.

Definition: Sei S eine endliche oder abzählbare Menge (meistens eine Teilmenge von $\mathbb{Z}$). Ein Vektor $(w(i))_{i \in S}$ heißt Wahrscheinlichkeitsvektor, wenn $w(i) \geq 0$ für alle $i \in S$ und $\sum_{i \in S} w(i) = 1$ gilt. Eine Zufallsvariable X heißt diskret, wenn ein Wahrscheinlichkeitsvektor $(w(i))_{i \in S}$ existiert mit $w(i) = P(X = i)$ für alle $i \in S$.

Beispiel 37: Die Zufallsvariable X bezeichne den Ausfall eines Würfels. Man bestimme den Wahrscheinlichkeitsvektor von X.

Sei $S = \{1, 2, 3, 4, 5, 6\}$ und $w(i) = \frac{1}{6}$ für alle $i \in S$. Es gilt $\sum_{i=1}^{6} w(i) = 1$. Also ist $(w(i))_{i \in S}$ ein Wahrscheinlichkeitsvektor. Da jede Augenzahl mit Wahrscheinlichkeit $\frac{1}{6}$ auftritt, gilt auch $w(i) = P(X = i)$ für alle $i \in S$.

Kennt man den Wahrscheinlichkeitsvektor einer Zufallsvariablen, so kann man damit Wahrscheinlichkeiten von Ereignissen berechnen.

Satz 13: *Sei $(w(i))_{i \in S}$ der Wahrscheinlichkeitsvektor einer diskreten Zufallsvariable X. Für jede Teilmenge B von S gilt dann $P(X \in B) = \sum_{i \in B} w(i)$.*

Beweis: Sei zuerst $B = \{i_1, i_2, \dots, i_k\}$, eine endliche Teilmenge von S. Es gilt dann $X \in B \Leftrightarrow X = i_1 \vee X = i_2 \vee \dots \vee X = i_k$. Da die rechtsstehenden Ereignisse unvereinbar sind, folgt $P(X \in B) = P(X = i_1) + \dots + P(X = i_k) = w(i_1) + \dots + w(i_k) = \sum_{i \in B} w(i)$ mit Hilfe des Additionssatzes.

Ist S eine endliche Menge, dann sind wir schon fertig. Sei S also abzählbar. Sei $\varepsilon > 0$. Wir verwenden Satz 53 und Satz 55 aus dem Anhang. Wegen $\sum_{i \in S} w(i) = 1$ existiert nach Satz 53 eine endliche Teilmenge U von S mit $\sum_{i \in U} w(i) > 1 - \varepsilon$. Da U endlich ist, folgt $P(X \in U) = \sum_{i \in U} w(i)$ aus dem ersten Teil des Beweises.

Sei jetzt $B \subseteq S$ beliebig. Wir setzen $C = B \cap U$ und $R = B \setminus U$. Es gilt dann $U \cap R = \emptyset$. Aus dem Additionssatz erhalten wir $P(X \in U) + P(X \in R) = P(X \in U \cup R) \le 1$ und $\sum_{i \in U} w(i) + \sum_{i \in R} w(i) = \sum_{i \in U \cup R} w(i) \le \sum_{i \in S} w(i) = 1$ folgt aus Satz 55 (a) und (c). Wegen $P(X \in U) = \sum_{i \in U} w(i) > 1 - \varepsilon$ ergibt sich $P(X \in R) < \varepsilon$ und $\sum_{i \in R} w(i) < \varepsilon$.

Es gilt $B = C \cup R$ mit $C \cap R = \emptyset$ und C ist endlich. Aus dem Additionssatz ergibt sich $P(X \in B) = P(X \in C) + P(X \in R)$ und $\sum_{i \in B} w(i) = \sum_{i \in C} w(i) + \sum_{i \in R} w(i)$ folgt aus Satz 55 (a). Aus diesen Gleichungen und aus den obigen Abschätzungen erhalten wir $P(X \in C) \le P(X \in B) < P(X \in C) + \varepsilon$ und $\sum_{i \in C} w(i) \le \sum_{i \in B} w(i) \le \sum_{i \in C} w(i) + \varepsilon$. Da C endlich ist, folgt $P(X \in C) = \sum_{i \in C} w(i)$ aus dem ersten Teil des Beweises. Setzt man das alles zusammen, so hat man $|P(X \in B) - \sum_{i \in B} w(i)| < \varepsilon$. Da ε beliebig nahe bei 0 gewählt werden kann, muss $P(X \in B) = \sum_{i \in B} w(i)$ gelten. $\square$

Bemerkung: Ist X eine diskrete Zufallsvariable mit Wahrscheinlichkeitsvektor $(w(i))_{i \in S}$, dann folgt $P(X \in S) = \sum_{i \in S} w(i) = 1$ aus Satz 13. Es gilt dann auch $P(X \notin S) = 1 - P(X \in S) = 0$. Somit nimmt X keine Werte außerhalb von S an. Man nennt daher S den Wertebereich der Zufallsvariable X.

Wir definieren Erwartungswert und Varianz einer diskreten Zufallsvariablen X. Der Erwartungswert ist ein Durchschnittswert aller möglichen Werte von X, wobei jedoch jeder Wert i mit seiner Wahrscheinlichkeit $w(i)$ gewichtet wird. Jeder Wert i geht in den Erwartungswert entsprechend der Häufigkeit seines Auftretens ein. Die Varianz misst, wie stark eine Zufallsvariable X um ihren Erwartungswert schwankt.

Definition: Der Erwartungswert $\mathrm{E}(X)$ einer diskreten Zufallsvariablen X mit Wahrscheinlichkeitsvektor $(w(i))_{i \in S}$ ist definiert durch $\mathrm{E}(X) = \sum_{i \in S} i w(i)$.

Wenn S nicht endlich ist, dann muss man sich um die Konvergenz kümmern. Deshalb verlangen wir zusätzlich, dass $\sum_{i \in S} |i| w(i) < \infty$ gilt. Ansonsten sagen wir, dass der Erwartungswert nicht existiert. Dadurch ist garantiert, dass die Reihe $\sum_{i \in S} i w(i)$ immer denselben endlichen Wert hat, wie immer man sie auch anordnet (siehe Satz 54).

Definition: Die Varianz $\mathrm{V}(X)$ einer diskreten Zufallsvariablen X mit Wahrscheinlichkeitsvektor $(w(i))_{i\in S}$ ist definiert durch $\mathrm{V}(X) = \sum_{i\in S}(i-m)^2 w(i)$ mit $m = \mathrm{E}(X)$. Weiters nennt man $\sqrt{\mathrm{V}(X)}$ die Standardabweichung der Zufallsvariablen X.

Um Erwartungswert und Varianz zu berechnen, bestimmt man den Wahrscheinlichkeitsvektor und setzt in die Formeln ein.

Beispiel 38: Bei einer Verlosung bringen 1% der Lose einen Gewinn von 1000 Euro, 9% der Lose bringen einen Gewinn von 100 Euro und 90% der Lose bringen keinen Gewinn. Welchen Gewinn muss man durchschnittlich pro Los auszahlen? Wie hoch soll man den Lospreis festsetzen?

Sei X der Gewinn eines (zufällig gewählten) Loses. Wir berechnen $\mathrm{E}(X)$. Die möglichen Gewinne, also die möglichen Werte von X sind 1000, 100 und 0 mit Wahrscheinlichkeiten $w(1000) = 0.01$, $w(100) = 0.09$ und $w(0) = 0.9$. Aus der Formel für den Erwartungswert erhalten wir $\mathrm{E}(X) = 1000 \cdot 0.01 + 100 \cdot 0.09 + 0 \cdot 0.9 = 19$. Jedenfalls sollte man den Lospreis höher als 19 Euro festsetzen.

Wir berechnen die Varianz des Gewinns: $\mathrm{V}(X) = 981^2 \cdot 0.01 + 81^2 \cdot 0.09 + (-19)^2 \cdot 0.9 = 10539$. Die Standardabweichung ist $\sqrt{\mathrm{V}(X)} = 102.66$.

Jetzt behandeln wir die sogenannten kontinuierlichen Zufallsvariablen. Man kann sie als kontinuierliches Analogon zu den diskreten Zufallsvariablen auffassen, wobei Summen durch Integrale ersetzt werden.

Definition: Eine integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}^+$ mit $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ heißt Wahrscheinlichkeitsdichte. Eine Zufallsvariable X heißt kontinuierlich, wenn eine Wahrscheinlichkeitsdichte f existiert, sodass $P(X \le t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ für alle $t \in \mathbb{R}$ gilt.

Die Funktion $F : \mathbb{R} \to \mathbb{R}^+$ definiert durch $F(t) = P(X \le t)$ heißt Verteilungsfunktion der Zufallsvariablen X. Die Wahrscheinlichkeitsdichte kann man daher auch als Ableitung der Verteilungsfunktion auffassen.

Satz 14: *Die Verteilungsfunktion F einer Zufallsvariablen X ist eine monton wachsende Funktion von $\mathbb{R}$ nach $[0,1]$. Für $r < s$ gilt $P(r < X \le s) = F(s) - F(r)$.*

Beweis: Es gilt $X \le s \Leftrightarrow X \le r \ \vee \ r < X \le s$, wobei die Ereignisse $X \le r$ und $r < X \le s$ unvereinbar sind. Daher folgt $P(X \le s) = P(X \le r) + P(r < X \le s)$ aus dem Additionssatz. Wir erhalten

$$P(r < X \le s) = P(X \le s) - P(X \le r) = F(s) - F(r)$$

mit Hilfe der Definition der Verteilungsfunktion.

Sind r und s in $\mathbb{R}$ beliebig mit $r < s$, dann gilt $F(s) - F(r) = P(r < X \le s) \ge 0$ nach Satz 6. Somit ist F monoton wachsend. Ebenfalls aus Satz 6 folgt $0 \le P(X \le t) \le 1$, sodass $F(t) \in [0,1]$ für alle $t \in \mathbb{R}$ gilt. $\square$

Wahrscheinlichkeiten lassen sich mit Hilfe der Wahrscheinlichkeitsdichte berechnen.

Satz 15: *Sei $f : \mathbb{R} \to \mathbb{R}^+$ die Wahrscheinlichkeitsdichte einer kontinuierlichen Zufallsvariablen X und $F : \mathbb{R} \to [0,1]$ ihre Verteilungsfunktion. Dann gilt*

(a) $P(X \in B) = F(s) - F(r) = \int_r^s f(x)\,\mathrm{d}x$ *für ein Intervall B mit Endpunkten r und s*

(b) $P(X < s) = P(X \le s) = F(s) = \int_{-\infty}^{s} f(x)\,\mathrm{d}x$ *für $s \in \mathbb{R}$*

(c) $P(X > r) = P(X \ge r) = 1 - F(r) = \int_r^{\infty} f(x)\,\mathrm{d}x$ *für $r \in \mathbb{R}$*

Beweis: In Satz 14 wurde $P(X \in B) = F(s) - F(r)$ für $B = (r, s]$ gezeigt. Aus der Definition der Wahrscheinlichkeitsdichte folgt

$$F(s) - F(r) = \int_{-\infty}^{s} f(x)\,\mathrm{d}x - \int_{-\infty}^{r} f(x)\,\mathrm{d}x = \int_{r}^{s} f(x)\,\mathrm{d}x$$

womit (a) für $B = (r, s]$ bereits bewiesen ist.

Wir zeigen, dass $P(X = u) = 0$ für alle $u \in \mathbb{R}$ gilt. Wenn das Ereignis $X = u$ eintritt, dann tritt für jedes $\varepsilon > 0$ auch das Ereignis $X \in (u - \varepsilon, u]$ ein. Für alle $\varepsilon > 0$ gilt $P(X = u) \leq P(X \in (u - \varepsilon, u])$ nach Satz 8 (c) und $P(X \in (u - \varepsilon, u]) = \int_{u-\varepsilon}^{u} f(x)\,\mathrm{d}x$ wurde oben bewiesen. Da f eine integrierbare Funktion ist, gilt $\lim_{\varepsilon \to 0} \int_{u-\varepsilon}^{u} f(x)\,\mathrm{d}x = 0$. Daraus ergibt sich $\lim_{\varepsilon \to 0} P(X \in (u - \varepsilon, u]) = 0$. Also muss auch $P(X = u) = 0$ gelten. Die Wahrscheinlichkeit $P(X \in B)$ ändert sich also nicht, wenn man einen Endpunkt des Intervalls B dazugibt oder weglässt (Additionssatz). Sind r und s die Endpunkte von B, dann gilt $P(X \in B) = P(X \in (r, s])$ und (a) ist gezeigt.

Aus den Definitionen folgt $P(X \leq s) = F(s) = \int_{-\infty}^{s} f(x)\,\mathrm{d}x$. Da $P(X = s) = 0$ gilt, ergibt sich $P(X < s) = P(X < s) + P(X = s) = P(X \leq s)$ und (b) ist gezeigt.

Um (c) zu beweisen, gehen wir zum Gegenereignis über und verwenden (b). Mit Satz 8 (d) folgt $P(X > r) = 1 - P(X \leq r) = 1 - F(r)$. Mit Hilfe von $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ erhält man

$$P(X > r) = 1 - P(X \leq r) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x - \int_{-\infty}^{r} f(x)\,\mathrm{d}x = \int_{r}^{\infty} f(x)\,\mathrm{d}x$$

Da $P(X = r) = 0$ gilt, folgt $P(X \geq r) = P(X > r)$ und (c) ist gezeigt. $\square$

Bemerkung: Ist F die Verteilungsfunktion einer kontinuierlichen Zufallsvariable, dann folgt $\lim_{s \to \infty} F(s) = 1$ und $\lim_{r \to -\infty} F(r) = 0$ aus Satz 15, da ja $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ gilt.

Oft kommt es vor, dass die Wahrscheinlichkeitsdichte f einer Zufallsvariablen X außerhalb eines (meist unbeschränkten) Intervalls W gleich 0 ist. Aus Satz 15 folgt $P(X \in W) = \int_W f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, sodass X mit Wahrscheinlichkeit 1, also immer, in W liegt. Wir nennen daher W den Wertebereich der Zufallsvariablen X.

Will man die Wahrscheinlichkeitsdichte f einer Zufallsvariablen X berechnen, so berechnet man zuerst die Verteilungsfunktion $F(t) = P(X \leq t)$ für $t \in \mathbb{R}$. Man erhält dann die Wahrscheinlichkeitsdichte f als Ableitung von F.

Beispiel 39: Zwei Punkte a und b werden zufällig im Intervall $[0, 1]$ gewählt. Sei X das Minimum der beiden Punkte. Gesucht ist die Wahrscheinlichkeitsdichte f von X.

Für $t \in \mathbb{R}$ berechnen wir $F(t) = P(X \leq t)$. Wir verwenden dazu die geometrische Wahrscheinlichkeit. Die Ausfallsmenge unseres Zufallsexperiments ist $\Omega = [0, 1] \times [0, 1]$. Das Ereignis $X \leq t$ entspricht der Teilmenge $A = \{(a, b) \in \Omega : \min(a, b) \leq t\}$ von Ω. Für $t < 0$ ist $A = \emptyset$ und somit $|A| = 0$. Für $t \geq 1$ ist $A = \Omega$ und somit $|A| = 1$. Für $0 \leq t < 1$ ist $A = \Omega \setminus (t, 1] \times (t, 1]$ und somit $|A| = 1 - (1 - t)^2 = 2t - t^2$. Wir erhalten $F(t) = 0$ für $t < 0$, $F(t) = 2t - t^2$ für $0 \leq t < 1$ und $F(t) = 1$ für $t \geq 1$. Wegen $f(x) = F'(x)$ folgt $f(x) = 0$ für $x \notin [0, 1]$ und $f(x) = 2 - 2x$ für $x \in [0, 1]$. Der Wertebereich von X ist das Intervall $[0, 1]$.

Bemerkung: Wie man in Beispiel 39 den Funktionswert von f im Punkt 0 wählt, spielt keine Rolle. Das hat keinen Einfluß auf die mit Hilfe von f berechneten Wahrscheinlichkeiten, da diese ja Integrale über f sind. In den Unstetigkeitsstellen einer Wahrscheinlichkeitsdichte hat man die Freiheit, den Funktionswert so zu wählen, wie man will.

Erwartungswert und Varianz für kontinuierliche Zufallsvariable werden analog wie für diskrete Zufallsvariable definiert. Die Summe wird durch das Integral ersetzt und der Wahrscheinlichkeitsvektor durch die Wahrscheinlichkeitsdichte.

Definition: Der Erwartungswert $\mathrm{E}(X)$ einer kontinuierlichen Zufallsvariablen X mit Wahrscheinlichkeitsdichte f ist definiert durch $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$. Analog zum diskreten Fall verlangen wir zusätzlich, dass $\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x < \infty$ gilt. Ansonsten sagen wir, dass der Erwartungswert nicht existiert.
Die Varianz $\mathrm{V}(X)$ ist definiert durch $\mathrm{V}(X) = \int_{-\infty}^{\infty} (x-m)^2 f(x)\,\mathrm{d}x$ mit $m = \mathrm{E}(X)$. Wieder nennt man $\sqrt{\mathrm{V}(X)}$ die Standardabweichung der Zufallsvariablen X.

Beispiel 40: Die Wahrscheinlichkeitsdichte f der Zufallsvariablen X sei gegeben durch $f(x) = \frac{1}{b-a}$ für $x \in [a,b]$ und $f(x) = 0$ für $x \notin [a,b]$. Gesucht sind $\mathrm{E}(X)$ und $\mathrm{V}(X)$.
Der Wertebereich der Zufallsvariablen X ist das Intervall $[a,b]$. Da f außerhalb von $[a,b]$ verschwindet, genügt es, von a bis b zu integrieren. Wir erhalten

$$\mathrm{E}(X) = \int_a^b x f(x)\,\mathrm{d}x = \int_a^b x \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2}$$

$$\mathrm{V}(X) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\,\mathrm{d}x = \frac{(b-a)^3-(a-b)^3}{24}\frac{1}{b-a} = \frac{(b-a)^2}{12}$$

Es gilt auch $\int_a^b |x| f(x)\,\mathrm{d}x < \infty$, da über ein beschränktes Intervall integriert wird.

In den folgenden Kapiteln werden häufig auftretende Wahrscheinlichkeitsvektoren und Wahrscheinlichkeitsdichten behandelt. Man verwendet die Bezeichnung Verteilung als gemeinsamen Überbegriff. Wird nach der Verteilung gefragt, so meint man für diskrete Zufallsvariable den Wahrscheinlichkeitsvektor und für kontinuierliche Zufallsvariable die Wahrscheinlichkeitsdichte oder die Verteilungsfunktion.

3. Binomialverteilung und geometrische Verteilung

Diese beiden Verteilungen treten bei wiederholtem Münzenwerfen auf. Die Binomialverteilung erhält man, wenn man zählt, wie oft "Wappen" fällt.

Definition: Sei $n \geq 1$ und $p \in (0,1)$. Eine diskrete Zufallsvariable X mit Wertebereich $S = \{0,1,2,\dots,n\}$ und Wahrscheinlichkeitsvektor

$$w(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{für} \quad i \in S$$

heißt binomialverteilt mit Parametern n und p oder kurz $B(n,p)$-verteilt. Es gilt $w(i) \geq 0$ für alle $i \in S$ und $\sum_{i \in S} w(i) = 1$ folgt aus dem binomischen Lehrsatz.

Satz 16: *Eine Münze, bei der "Wappen" mit Wahrscheinlichkeit p und "Zahl" mit Wahrscheinlichkeit $1-p$ fällt, wird n Mal geworfen. Sei X die Anzahl mit der "Wappen" unter diesen n Würfen auftritt. Dann hat X die $B(n,p)$-Verteilung.*

Beweis: Sei $S = \{0,1,2,\dots,n\}$. Wir müssen $P(X = i)$ für $i \in S$ berechnen. Gefragt ist nach der Wahrscheinlichkeit, dass unter den n Würfen genau i Mal "Wappen" auftritt. Für $n = 5$ und $i = 3$ wurde das bereits in Beispiel 29 durchgeführt. Die dort angewendete Methode lässt sich auch hier anwenden. Wir zerlegen in Teilereignisse. Die Teilereignisse entsprechen den Folgen der Länge n, die i Mal W und $n-i$ Mal Z enthalten. Nach Satz 4 ist die Anzahl dieser Teilereignisse gleich $\binom{n}{i}$.

Berechnet man die Wahrscheinlichkeiten der Teilereignisse nach dem Multiplikationssatz, so erhält man $p^i(1-p)^{n-i}$ für jedes dieser Teilereignisse. Nach dem Additionssatz erhält man die gesuchte Wahrscheinlichkeit $P(X=i)$, indem man die Wahrscheinlichkeiten der Teilereignisse aufsummiert. Da es $\binom{n}{i}$ Teilereignisse gibt und jedes dieser Teilereignisse dieselbe Wahrscheinlichkeit $p^i(1-p)^{n-i}$ hat, ist die Summe gleich $\binom{n}{i}p^i(1-p)^{n-i}$. Damit haben wir $P(X=i)=\binom{n}{i}p^i(1-p)^{n-i}$ für alle $i \in S$ gezeigt. Die Zufallsvariable X ist daher $B(n,p)$-verteilt. □

Ein Zufallsexperiment heißt Bernoulliexperiment, wenn es nur zwei mögliche Ausfälle hat. Den einen der beiden Ausfälle nennen wir Erfolg, den anderen nennen wir Mißerfolg. Ein Bernoulliexperiment entspricht daher dem Münzenwerfen, wobei wir "Wappen" als Erfolg und "Zahl" als Mißerfolg auffassen. Sei p die Wahrscheinlichkeit für Erfolg, das heißt für das Auftreten von "Wappen". Führt man das Bernoulliexperiment n Mal durch und zählt, wie oft Erfolg ("Wappen") eintritt, dann ist diese Anzahl $B(n,p)$-verteilt.

Beispiel 41: Es wird 10 Mal gewürfelt. Mit welcher Wahrscheinlichkeit tritt höchstens 3 Mal die Augenzahl 6 auf?

Es liegt ein Bernoulliexperiment vor, wobei wir Augenzahl 6 als Erfolg auffassen und alle anderen Augenzahlen als Mißerfolg. Die Wahrscheinlichkeit für Erfolg ist $\frac{1}{6}$. Sei X die Anzahl, mit der Erfolg, das heißt die Augenzahl 6, unter den 10 Würfen auftritt. Nach Satz 16 hat X die $B(n,p)$-Verteilung mit $n=10$ und $p=\frac{1}{6}$, das heißt der Wahrscheinlichkeitsvektor der Zufallsvariable X ist durch $w(i)=\binom{10}{i}(\frac{1}{6})^i(\frac{5}{6})^{10-i}$ für $0 \le i \le 10$ gegeben. Gesucht ist die Wahrscheinlichkeit, dass das Ereignis $X \le 3$ eintritt. Aus Satz 13 folgt $P(X \le 3) = w(0)+w(1)+w(2)+w(3)$. Wir erhalten also $P(X \le 3) = \binom{10}{0}(\frac{1}{6})^0(\frac{5}{6})^{10} + \binom{10}{1}(\frac{1}{6})^1(\frac{5}{6})^9 + \binom{10}{2}(\frac{1}{6})^2(\frac{5}{6})^8 + \binom{10}{3}(\frac{1}{6})^3(\frac{5}{6})^7$.

Beispiel 42: In einer Pension gibt es 16 Einzelzimmer. Man weiß, dass im Durchschnitt 20% der bestellten Zimmer nicht belegt werden. Der Pensionsinhaber nimmt für die Ferienwoche 18 Buchungen entgegen. Mit welcher Wahrscheinlichkeit wird er Schwierigkeiten haben?

Es liegt ein Bernoulliexperiment vor. Unter Erfolg verstehen wir, dass die gebuchte Person kommt. Die Wahrscheinlichkeit für Erfolg ist $p=\frac{4}{5}$. Es werden 18 Buchungen entgegengenommen, das heißt das Bernoulliexperiment wird 18 Mal durchgeführt. Sei X die Anzahl der Erfolge, das ist die Anzahl der Personen, die kommen. Nach Satz 16 hat X die $B(n,p)$-Verteilung mit $n=18$ und $p=\frac{4}{5}$. Es gibt Schwierigkeiten, wenn mehr als 16 kommen, also wenn $X \ge 17$ ist. Nach Satz 13 ist die Wahrscheinlichkeit dafür $P(X \ge 17) = w(17)+w(18) = \binom{18}{17}(\frac{4}{5})^{17}(\frac{1}{5})^1 + \binom{18}{18}(\frac{4}{5})^{18}(\frac{1}{5})^0$.

Beispiel 43: Eine Maschine produziert Schrauben. Der Anteil der unbrauchbaren Schrauben beträgt 10%. Wie groß ist die Wahrscheinlichkeit, dass in einer Packung von 100 Schrauben mehr als 7 unbrauchbare Schrauben sind?

Die Produktion einer Schraube ist ein Bernoulliexperiment. Die Wahrscheinlichkeit für einen Erfolg (brauchbare Schraube) ist $p=0.9$. Sei X die Anzahl der brauchbaren Schrauben in der Packung von 100 Stück. Nach Satz 16 hat X die $B(n,p)$-Verteilung mit $n=100$ und $p=0.9$. Gesucht ist $P(X \le 92)$. Wir gehen zur Gegenwahrscheinlichkeit über: $P(X \ge 93) = \sum_{i=93}^{100} w(i) = \sum_{i=93}^{100}\binom{100}{i}0.9^i 0.1^{100-i}$. Die gesuchte Wahrscheinlichkeit ist $1-P(X \ge 93)$. Wir werden später sehen, wie man diese Wahrscheinlichkeit durch Approximation mit der Normalverteilung näherungsweise berechnen kann.

Wir berechnen den Erwartungswert einer $B(n,p)$-verteilten Zufallsvariable X. Aus der Definition $\mathrm{E}(X)=\sum_{i\in S} iw(i)$ erhalten wir

$$\begin{aligned}
\mathrm{E}(X) &= \textstyle\sum_{i=1}^n i\binom{n}{i}p^i(1-p)^{n-i} && \text{da der Summand für } i=0 \text{ null ist}\\
&= \textstyle\sum_{i=1}^n n\binom{n-1}{i-1}p^i(1-p)^{n-i} && \text{da } i\binom{n}{i}=n\binom{n-1}{i-1} \text{ für } i\ge 1\\
&= np\textstyle\sum_{i=1}^n \binom{n-1}{i-1}p^{i-1}(1-p)^{n-i} &&\\
&= np\textstyle\sum_{j=0}^{n-1} \binom{n-1}{j}p^{j}(1-p)^{n-1-j} && \text{neue Summationsvariable } j=i-1\\
&= np && \text{binomischer Lehrsatz}
\end{aligned}$$

Für eine $B(n,p)$-verteilte Zufallsvariable X gilt also $\mathrm{E}(X)=np$. Der Erwartungswert gibt den durchschnittlichen Wert einer Zufallsvariable an. Der Pensionsinhaber aus Beispiel 42 kann also erwarten, dass durchschnittlich $18\cdot 0.8=14.4$ der 18 Personen, die gebucht haben, kommen. In einer Packung Schrauben aus Beispiel 43 wird man durchschnittlich $100\cdot 0.9=90$ brauchbare Schrauben finden.

Mit den selben Methoden wie den Erwartungswert kann man auch die Varianz $\mathrm{V}(X)$ einer $B(n,p)$-verteilten Zufallsvariablen X berechnen:

$$\mathrm{V}(X)=\textstyle\sum_{i=0}^n(i-np)^2\binom{n}{i}p^i(1-p)^{n-i}=\sum_{i=0}^n\big(i(i-1)+(1-2np)i+n^2p^2\big)\binom{n}{i}p^i(1-p)^{n-i}=$$
$$\textstyle\sum_{i=0}^n i(i-1)\binom{n}{i}p^i(1-p)^{n-i}+(1-2np)\sum_{i=0}^n i\binom{n}{i}p^i(1-p)^{n-i}+n^2p^2\sum_{i=0}^n\binom{n}{i}p^i(1-p)^{n-i}$$

Aus dem binomischen Lehrsatz folgt, dass $\sum_{i=0}^n\binom{n}{i}p^i(1-p)^{n-i}=1$ gilt. Weiters wurde die Summe $\sum_{i=0}^n i\binom{n}{i}p^i(1-p)^{n-i}=np$ oben bei der Berechnung des Erwartungswertes bestimmt. Genauso erhält man $\sum_{i=0}^n i(i-1)\binom{n}{i}p^i(1-p)^{n-i}=n(n-1)p^2$, da die Summanden für $i=0$ und $i=1$ null sind und $i(i-1)\binom{n}{i}=n(n-1)\binom{n-2}{i-2}$ gilt. Wir erhalten somit $\mathrm{V}(X)=n(n-1)p^2+(1-2np)np+n^2p^2=np(1-p)$ für eine $B(n,p)$-verteilte Zufallsvariable X.

Wir führen noch eine weitere Verteilung ein, die mit Bernoulliexperimenten zu tun hat. Jetzt geht es um die Wartezeit auf den ersten Erfolg.

Definition: Sei $p\in(0,1)$. Eine diskrete Zufallsvariable X mit Wertebereich $S=\{1,2,3,\dots\}$ und Wahrscheinlichkeitsvektor

$$w(i)=(1-p)^{i-1}p \quad \text{für} \quad i\in S$$

heißt geometrisch verteilt mit Parameter p oder kurz $G(p)$-verteilt. Es gilt $w(i)\ge 0$ für alle $i\in S$ und $\sum_{i\in S}w(i)=1$ folgt aus der Formel für die geometrische Reihe.

Satz 17: *Ein Bernoulliexperiment, bei dem die Wahrscheinlichkeit für Erfolg gleich p ist, wird so oft durchgeführt, bis der erste Erfolg auftritt. Sei X die Anzahl der dafür notwendigen Wiederholungen. Dann hat X die $G(p)$-Verteilung.*

Beweis: Sei $S=\{1,2,3,\dots\}$ und $i\in S$. Das Ereignis $X=i$ tritt genau dann ein, wenn die ersten $i-1$ Wiederholungen des Bernoulliexperiments Mißerfolg und die i-te Wiederholung Erfolg ergibt. Die Wahrscheinlichkeit dafür ist $(1-p)^{i-1}p$ nach dem Multiplikationssatz, da Mißerfolg ja mit Wahrscheinlichkeit $1-p$ und Erfolg mit Wahrscheinlichkeit p auftritt. Wir haben somit $P(X=i)=(1-p)^{i-1}p$ für alle $i\in S$ gezeigt, das heißt die Zufallsvariable X hat die $G(p)$-Verteilung. □

Beispiel 44: Man würfelt so lange, bis 6 kommt. Wie groß ist die Wahrscheinlichkeit, dass man höchstens 10 Mal würfeln muss?

Es liegt ein Bernoulliexperiment vor, wobei Augenzahl 6 Erfolg bedeutet und alle anderen Augenzahlen Mißerfolg. Die Wahrscheinlichkeit für Erfolg ist $\frac{1}{6}$. Sei X die Anzahl der Würfe, die man benötigt, bis zum ersten Mal 6 auftritt. Gesucht ist $P(X \leq 10)$. Nach Satz 17 hat X die $G(p)$-Verteilung mit $p = \frac{1}{6}$. Für die gesuchte Wahrscheinlichkeit erhalten wir $P(X \leq 10) = \sum_{i=1}^{10} (\frac{5}{6})^{i-1}\frac{1}{6} = 1 - (\frac{5}{6})^{10}$.

Beispiel 45: Anna und Barbara würfeln, und zwar zuerst einmal Anna, dann zweimal Barbara, dann zweimal Anna, dann zweimal Barbara, und immer so weiter. Wer zuerst 6 würfelt, gewinnt. Man berechne die Wahrscheinlichkeit, dass Barbara gewinnt.

Sei X die Anzahl der Würfe, bis zum ersten Mal 6 auftritt. Nach Satz 17 hat X die $G(p)$-Verteilung mit $p = \frac{1}{6}$. Der Wahrscheinlichkeitsvektor von X ist $w(i) = (\frac{5}{6})^{i-1}\frac{1}{6}$ für $i \geq 1$. Barbara gewinnt, wenn X einen Wert aus der Menge $M = \{2, 3, 6, 7, 10, 11, \dots\}$ annimmt. Es gilt $P(X \in M) = w(2) + w(3) + w(6) + w(7) + w(10) + w(11) + \dots = (\frac{5}{6}\frac{1}{6} + (\frac{5}{6})^2\frac{1}{6})(1 + (\frac{5}{6})^4 + (\frac{5}{6})^8 + \dots) = \frac{55}{6^3}\frac{1}{1-(\frac{5}{6})^4} = \frac{330}{6^4-5^4} = \frac{330}{671}$. Barbara gewinnt mit Wahrscheinlichkeit $\frac{330}{671} = 0.4918$.

Den Erwartungswert einer $G(p)$-verteilten Zufallsvariable X berechnen wir mit folgendem Trick. Sei $x \in (0, 1)$. Es gilt dann die Gleichung $\sum_{i=1}^{n-1} x^i = \frac{1-x^n}{1-x} - 1$. Durch Differenzieren erhalten wir $\sum_{i=1}^{n-1} ix^{i-1} = \frac{-nx^{n-1}(1-x)+1-x^n}{(1-x)^2}$. Wir lassen n gegen ∞ gehen. Wegen $0 < x < 1$ ergibt sich die Gleichung $\sum_{i=1}^{\infty} ix^{i-1} = \frac{1}{(1-x)^2}$. Mit Hilfe dieser Gleichung berechnen wir $\mathrm{E}(X) = \sum_{i=1}^{\infty} iw(i) = \sum_{i=1}^{\infty} i(1-p)^{i-1}p = \frac{p}{p^2} = \frac{1}{p}$.

Beispiel 46: Wie oft muss man durchschnittlich würfeln, bis die Augenzahl 6 kommt?

Die Anzahl X der Würfe, die man benötigt, bis 6 kommt, ist $G(p)$-verteilt mit $p = \frac{1}{6}$. Die durchschnittliche Anzahl ist daher $\mathrm{E}(X) = \frac{1}{p} = 6$.

4. Die hypergeometrische Verteilung

Aus einer Menge von N Kugeln, von denen M weiß und $N - M$ schwarz sind, wird eine Stichprobe vom Umfang n gezogen. Sei X die Anzahl der weißen Kugeln in der Stichprobe. Zieht man geordnet mit Zurücklegen, dann hat X die $B(n, p)$-Verteilung mit $p = \frac{M}{N}$, da sich der Anteil p der weißen Kugeln in der Menge, aus der gezogen wird, nicht ändert und daher die Wahrscheinlichkeit, eine weiße Kugel zu ziehen, bei jedem Zug gleich p ist. Es liegt ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit p vor.

Zieht man jedoch ungeordnet ohne Zurücklegen, dann gilt $P(X = i) = \binom{M}{i}\binom{N-M}{n-i}/\binom{N}{n}$, da $\binom{N}{n}$ die Anzahl aller n-elementigen Teilmengen der Menge, aus der gezogen wird, ist und $\binom{M}{i}\binom{N-M}{n-i}$ die Anzahl der Teilmengen, die i weiße und $n - i$ schwarze Kugeln enthalten. Wir führen dazu eine eigene Verteilung ein.

Es sei daran erinnert, dass $\binom{u}{v} = 0$ gesetzt wird, wenn $v < 0$ oder $v > u$ gilt.

Definition: Seien N und M in $\mathbb{N}$ mit $0 < M < N$ und sei $n \geq 1$. Eine diskrete Zufallsvariable X mit Wertebereich $S = \{0, 1, 2, \dots, n\}$ und Wahrscheinlichkeitsvektor

$$w(i) = \binom{M}{i}\binom{N-M}{n-i}\Big/\binom{N}{n} \quad \text{für} \quad i \in S$$

heißt hypergeometrisch verteilt mit Parametern N, M und n oder $H(N, M, n)$-verteilt. Es gilt $w(i) \geq 0$ für $i \in S$ und $\sum_{i \in S} w(i) = 1$ folgt aus der Gleichung $\sum_{i=0}^{n} \binom{M}{i}\binom{N-M}{n-i} = \binom{N}{n}$, die in einer Bemerkung nach Beispiel 8 bewiesen wurde.

Wir berechnen Erwartungswert und Varianz einer $H(N,M,n)$-verteilten Zufallsvariable. Das funktioniert so wie für eine $B(n,p)$-verteilte Zufallsvariable.

$$\begin{aligned}
\tbinom{N}{n}\mathrm{E}(X) &= \textstyle\sum_{i=0}^{n} i\binom{M}{i}\binom{N-M}{n-i} \\
&= \textstyle\sum_{i=1}^{n} i\binom{M}{i}\binom{N-M}{n-i} && \text{da der Summand für } i=0 \text{ null ist} \\
&= \textstyle\sum_{i=1}^{n} M\binom{M-1}{i-1}\binom{N-1-(M-1)}{n-1-(i-1)} && \text{da } i\tbinom{M}{i} = M\tbinom{M-1}{i-1} \text{ für } i\ge 1 \\
&= M\textstyle\sum_{j=0}^{n-1}\binom{M-1}{j}\binom{N-1-(M-1)}{n-1-j} && \text{neue Summationsvariable } \ j=i-1 \\
&= M\tbinom{N-1}{n-1} && \text{wegen } \textstyle\sum_{j=0}^{n-1}\binom{M-1}{j}\binom{N-1-(M-1)}{n-1-j} = \binom{N-1}{n-1}
\end{aligned}$$

Wegen $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1}$ erhalten wir somit, dass $\mathrm{E}(X) = n\frac{M}{N}$ für eine $H(N,M,n)$-verteilte Zufallsvariable X gilt.

Auch die Berechnung der Varianz ist ähnlich wie für die $B(n,p)$-Verteilung:

$\binom{N}{n}\mathrm{V}(X) = \sum_{i=0}^{n}(i-n\frac{M}{N})^2\binom{M}{i}\binom{N-M}{n-i} = \sum_{i=0}^{n}\big(i(i-1)+(1-2n\frac{M}{N})i+(n\frac{M}{N})^2\big)\binom{M}{i}\binom{N-M}{n-i}$

Wir erhalten $\sum_{i=0}^{n} i(i-1)\binom{M}{i}\binom{N-M}{n-i} = M(M-1)\binom{N-2}{n-2}$ wie in obiger Rechnung, da die Summanden für $i=0$ und $i=1$ null sind und $i(i-1)\binom{M}{i} = M(M-1)\binom{M-2}{i-2}$ gilt. Weiters hatten wir $\sum_{i=0}^{n} i\binom{M}{i}\binom{N-M}{n-i} = M\binom{N-1}{n-1}$ und $\sum_{i=0}^{n}\binom{M}{i}\binom{N-M}{n-i} = \binom{N}{n}$. Damit ergibt sich

$$\tbinom{N}{n}\mathrm{V}(X) = M(M-1)\tbinom{N-2}{n-2} + (1-2n\tfrac{M}{N})M\tbinom{N-1}{n-1} + (n\tfrac{M}{N})^2\tbinom{N}{n}$$

Wegen $\binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1} = \frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}$ erhalten wir

$$\mathrm{V}(X) = M(M-1)\tfrac{n(n-1)}{N(N-1)} + (1-2n\tfrac{M}{N})\tfrac{nM}{N} + (n\tfrac{M}{N})^2 = n\tfrac{M}{N}\tfrac{N-M}{N}\tfrac{N-n}{N-1}$$

Das ist die Varianz einer $H(N,M,n)$-verteilten Zufallsvariable X.

Diese Formel gilt auch für $n=1$, auch wenn in diesem Fall obige Rechnung nicht funktioniert. Es gilt $\binom{N}{1}\mathrm{V}(X) = \sum_{i=0}^{1}(i-\frac{M}{N})^2\binom{M}{i}\binom{N-M}{1-i} = (\frac{M}{N})^2(N-M)+(1-\frac{M}{N})^2 M$. Durch Umformen folgt $\mathrm{V}(X) = \frac{M}{N}\frac{N-M}{N}$.

5. Poissonverteilung

Es soll noch eine weitere diskrete Verteilung besprochen werden, die wir durch einen Grenzübergang aus der Binomialverteilung erhalten.

Definition: Sei $\lambda > 0$. Eine diskrete Zufallsvariable X mit Wertebereich $S = \{0,1,2,\dots\}$ und Wahrscheinlichkeitsvektor

$$w(i) = \frac{\lambda^i}{i!}e^{-\lambda} \quad \text{für} \quad i \in S$$

heißt Poissonverteilt mit Parameter λ oder kurz $P(\lambda)$-verteilt. Es gilt $w(i) \ge 0$ für alle $i \in S$ und $\sum_{i=0}^{\infty} w(i) = 1$ folgt aus $\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = e^{\lambda}$.

Bei der Telephonauskunft treffen zu zufälligen Zeitpunkten Telephonanrufe ein. Wir legen einen Zeitpunkt als Nullpunkt fest und bezeichnen die Anzahl der Anrufe im Zeitintervall $[0,t]$ mit X.

Um den Wahrscheinlichkeitsvektor von X zu berechnen, müssen wir einige Annahmen zugrundelegen. Sei $z > t$ und n die Anzahl der Personen, die im Zeitintervall $[0,z]$ anrufen. Wir nehmen an, dass jede Person unabhängig von den anderen zufällig einen Zeitpunkt in $[0,z]$ auswählt, zu dem sie anruft. Die Wahrscheinlichkeit, dass eine Person im Zeitintervall $[0,t]$ anruft, ist dann $\frac{t}{z}$ und die Wahrscheinlichkeit, dass sie im Zeitintervall $(t,z]$

anruft, ist $\frac{z-t}{z} = 1 - \frac{t}{z}$. Es liegt ein Bernoulliexperiment vor. Ein Erfolg (Person ruft im Zeitintervall $[0,t]$ an) tritt mit Wahrscheinlichkeit $\frac{t}{z}$ auf. Jeder der n Anrufe entspricht einer Wiederholung des Bernoulliexperiments. Da X die Anzahl der Anrufe im Zeitintervall $[0,t]$ ist, also die Anzahl der Erfolge, hat X die $B(n, \frac{t}{z})$-Verteilung. Wir halten $\mu = \frac{n}{z}$, die durchschnittliche Anzahl der Anrufe pro Zeiteinheit, fest und lassen n und damit auch z gegen ∞ gehen.

Satz 18: *Die Zufallsvariable X sei $B(n, \frac{t}{z})$-verteilt. Wenn n und z gegen ∞ gehen, sodass $\frac{n}{z} = \mu$ fest bleibt, dann hat X im Grenzwert die $P(\mu t)$-Verteilung.*

Beweis: Sei $i \in \mathbb{N} \cup \{0\}$. Wir suchen $P(X = i)$. Wir setzen $\frac{1}{z} = \frac{\mu}{n}$ ein und formen um

$$\begin{aligned} P(X = i) &= \binom{n}{i}(\tfrac{t}{z})^i(1 - \tfrac{t}{z})^{n-i} = \binom{n}{i}(\tfrac{\mu t}{n})^i(1 - \tfrac{\mu t}{n})^{n-i} \\ &= \tfrac{(\mu t)^i}{i!}\, \tfrac{n(n-1)\ldots(n-i+1)}{n^i}(1 - \tfrac{\mu t}{n})^n(1 - \tfrac{\mu t}{n})^{-i} \end{aligned}$$

Wegen $\lim_{n\to\infty} \frac{n(n-1)\ldots(n-i+1)}{n^i} = 1$ und $\lim_{n\to\infty}(1 - \frac{\mu t}{n})^n = e^{-\mu t}$ erhalten wir aus obiger Gleichung, dass $\lim_{n\to\infty} P(X = i) = \frac{(\mu t)^i}{i!} e^{-\mu t}$ gilt. Das aber besagt, dass X im Grenzwert die $P(\mu t)$-Verteilung hat. □

Beispiel 47: Die Anzahl der Anrufe in einem Zeitintervall von t Stunden habe die $P(\mu t)$-Verteilung mit $\mu = 12$ Anrufen pro Stunde. Wie groß ist die Wahrscheinlichkeit, dass zwischen 16.00 und 16.15 höchstens ein Anruf kommt? Wie groß ist die Wahrscheinlichkeit, dass zwischen 16.00 und 16.15 mindestens drei Anrufe kommen?

Sei X die Anzahl der Anrufe zwischen 16.00 und 16.15, das heißt X ist $P(\lambda)$-verteilt mit $\lambda = 12 \cdot \frac{1}{4} = 3$. Daher gilt

$P(X \leq 1) = w(0) + w(1) = \frac{3^0}{0!}e^{-3} + \frac{3^1}{1!}e^{-3} = 4e^{-3} = 0.199$

$P(X \geq 3) = 1 - P(X \leq 2) = 1 - w(0) - w(1) - w(2) = 1 - (\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!})e^{-3} = 0.5768$

Mit Wahrscheinlichkeit 0.199 kommt höchstens ein Anruf. Mit Wahrscheinlichkeit 0.577 kommen mindestens drei Anrufe.

Die Poissonverteilung wird immer dann verwendet, wenn ein Verhalten vorliegt, wie es oben für die Telephonanrufe beschrieben wurde. Das gilt für die Personen, die ein Geschäft betreten, für die Defekte eines Gerätes, das immer wieder repariert wird, und für die Schadensmeldungen, die bei einer Versicherung eintreffen.

6. Exponentialverteilung und Gammaverteilung

Jetzt kommen wir zu den kontinuierlichen Zufallsvariablen. Wir bleiben bei den Telephonanrufen, fragen jetzt aber nicht nach der Anzahl, sondern nach der Wartezeit auf einen Anruf. Das ist natürlich eine kontinuierliche Größe.

Definition: Sei $\lambda > 0$. Eine kontinuierliche Zufallsvariable X mit Wertebereich $\mathbb{R}^+$ und Wahrscheinlichkeitsdichte

$$f(x) = \lambda e^{-\lambda x} \quad \text{für} \quad x \in \mathbb{R}^+$$

heißt exponentialverteilt mit Parameter λ oder kurz $E(\lambda)$-verteilt. Für $x \in \mathbb{R}^-$ setzt man $f(x) = 0$. Es gilt $f(x) \geq 0$ für $x \in \mathbb{R}$ und $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^{\infty} f(x)\,\mathrm{d}x = 1$.

Satz 19: *Die Anzahl der Anrufe im Zeitintervall $[0,t]$ sei $P(\lambda t)$-verteilt, wobei λ die durchschnittliche Anzahl der Anrufe pro Zeiteinheit ist (letztes Kapitel). Sei Y die Wartezeit vom Zeitpunkt 0 bis zum ersten Anruf. Dann hat Y die $E(\lambda)$-Verteilung.*

Beweis: Sei $t > 0$ beliebig und X die Anzahl der Anrufe im Zeitintervall $[0, t]$. Die Wartezeit auf den ersten Anruf ist genau dann $\leq t$, wenn im Zeitintervall $[0, t]$ mindestens ein Anruf kommt. Das heißt $Y \leq t$ tritt genau dann ein, wenn $X \geq 1$ eintritt. Daher gilt $F(t) = P(Y \leq t) = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda t}$. Die Dichte f von Y erhält man als Ableitung von F. Es folgt $f(x) = (1 - e^{-\lambda x})' = \lambda e^{-\lambda x}$ für $x \in \mathbb{R}^+$.
Die Wartezeit Y kann nicht negativ sein. Für $t \leq 0$ gilt daher $F(t) = P(Y \leq t) = 0$. Es folgt $f(x) = 0$ für $x \leq 0$. Der Wertebereich der Zufallsvariablen Y ist gleich $\mathbb{R}^+$. Die Wartezeit Y auf den ersten Anruf ist $E(\lambda)$-verteilt. $\square$

Eine weitere Anwendung der Exponentialverteilung ist die Lebensdauer eines Gerätes. Sie ist ja die Wartezeit auf den ersten Defekt. Allerdings muss man dabei voraussetzen, dass Defekte rein zufällig auftreten, genauso wie wir es für die Telephonanrufe angenommen haben.

Beispiel 48: Die Lebensdauer eines elektronischen Bauteiles sei $E(\lambda)$-verteilt. Man weiß aus Erfahrung, dass die Lebensdauer mit Wahrscheinlichkeit 0.9 mindestens 500 Stunden beträgt. Wie groß ist λ?

Die Lebensdauer des elektronischen Bauteiles bezeichnen wir mit Y. Die Wahrscheinlichkeit, dass Y größer oder gleich 500 ist, ist 0.9, das heißt $P(Y \geq 500) = 0.9$. Wegen Satz 15 (c) gilt $P(Y \geq 500) = \int_{500}^{\infty} \lambda e^{-\lambda x}\, \mathrm{d}x = e^{-500\lambda}$, da Y ja $E(\lambda)$-verteilt ist. Aus $e^{-500\lambda} = 0.9$ folgt $\lambda = \frac{-\ln 0.9}{500} = 0.0002107$.

Wir berechnen Erwartungswert und Varianz einer $E(\lambda)$-verteilten Zufallsvariablen X. Setzt man die Wahrscheinlichkeitsdichte der Exponentialverteilung in die Formel für den Erwartungswert ein, so erhält man durch partielle Integration
$\mathrm{E}(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\, \mathrm{d}x = -xe^{-\lambda x}\big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, \mathrm{d}x = -\frac{e^{-\lambda x}}{\lambda}\big|_0^{\infty} = \frac{1}{\lambda}$.
Ebenso erhält man durch zweimalige partielle Integration
$\mathrm{V}(X) = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x}\, \mathrm{d}x = -(x - \frac{1}{\lambda})^2 e^{-\lambda x}\big|_0^{\infty} - 2(x - \frac{1}{\lambda})\frac{e^{-\lambda x}}{\lambda}\big|_0^{\infty} + \int_0^{\infty} 2\frac{e^{-\lambda x}}{\lambda}\, \mathrm{d}x = \frac{1}{\lambda^2}$.
Wir haben also $\mathrm{E}(X) = \frac{1}{\lambda}$ und $\mathrm{V}(X) = \frac{1}{\lambda^2}$ erhalten. Die durchschnittliche Lebensdauer des elektronischen Bauteils aus Beispiel 48 beträgt $\frac{500}{-\ln 0.9} = 4745.6$ Stunden.

Jetzt kehren wir zurück zur Telephonauskunft. Um die Wartezeit auf den n-ten Anruf zu berechnen, führen wir eine weitere Verteilung ein.

Definition: Seien r und λ positive reelle Zahlen. Eine kontinuierliche Zufallsvariable X mit Wertebereich $\mathbb{R}^+$ und Wahrscheinlichkeitsdichte

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \quad \text{für} \quad x \in \mathbb{R}^+$$

heißt gammaverteilt (Erlangverteilt) mit Parametern r und λ, oder kurz $E(r, \lambda)$-verteilt, wobei $\Gamma(r)$ definiert ist durch $\Gamma(r) = \int_0^{\infty} y^{r-1} e^{-y}\, \mathrm{d}y$. Es gilt $f(x) > 0$ für $x \in \mathbb{R}^+$ und $\int_0^{\infty} f(x)\, \mathrm{d}x = 1$ folgt aus der Definition von $\Gamma(r)$. Für $x \in \mathbb{R}^-$ setzt man $f(x) = 0$.

Für spezielle Werte von r kann man $\Gamma(r)$ explizit berechnen. Für $n \in \{1, 2, 3, \dots\}$ gilt $\Gamma(n) = (n-1)!$ und für $n \in \{0, 1, 2, \dots\}$ gilt $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{n! 2^{2n}} \sqrt{\pi}$.

Die $E(1, \lambda)$-Verteilung ist die $E(\lambda)$-Verteilung. Die $E(\frac{m}{2}, \frac{1}{2})$-Verteilung spielt in der Statistik eine wichtige Rolle und heißt dort chi-quadrat-Verteilung.

Satz 20: *Die Anzahl der Anrufe im Zeitintervall $[0, t]$ sei $P(\lambda t)$-verteilt. Sei Z die Wartezeit vom Zeitpunkt 0 bis zum n-ten Anruf. Dann hat Z die $E(n, \lambda)$-Verteilung.*

Beweis: Sei $t > 0$ beliebig und X die Anzahl der Anrufe im Zeitintervall $[0,t]$. Die Wartezeit auf den n-ten Anruf ist genau dann $\leq t$, wenn im Zeitintervall $[0,t]$ mindestens n Anrufe kommen, das heißt $Z \leq t$ tritt genau dann ein, wenn $X \geq n$ eintritt. Daher gilt
$F(t) = P(Z \leq t) = P(X \geq n) = 1 - P(X \leq n-1) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \dots - \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}$
Die Wahrscheinlichkeitsdichte f der Zufallsvariable Z erhält man als Ableitung von F. Berechnet man diese mit Hilfe der Produktregel, so kürzen sich alle dabei auftretenden Summanden weg bis auf einen. Man erhält $f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$ für $x \geq 0$. Da Z außerdem als Wartezeit nur positive Werte annehmen kann, ist $\mathbb{R}^+$ der Wertebereich von Z und $f(x) = 0$ für $x < 0$. □

7. Normalverteilung

Misst man eine (physikalische) Größe mehrmals, so erhält man Messwerte, die wegen zufälliger Störungen beim Messvorgang leicht voneinander abweichen. Werden Milchflaschen von einer Abfüllanlage gefüllt, so wird man Schwankungen bei den eingefüllten Mengen feststellen, die auf die Ungenauigkeit der Abfüllanlage zurückzuführen sind. Messwerte und Füllmengen unterliegen zufälligen Einflüssen und werden daher als Zufallsvariable aufgefasst. Für solche Zufallsvariablen verwendet man die Normalverteilung.

Definition: Sei $\mu \in \mathbb{R}$ und sei σ eine positive reelle Zahl. Eine kontinuierliche Zufallsvariable X mit Wertebereich $\mathbb{R}$ und Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für} \quad x \in \mathbb{R}$$

heißt normalverteilt mit Parametern μ und σ oder kurz $N(\mu,\sigma)$-verteilt.

Um Beispiele zur Normalverteilung zu rechnen, benötigen wir einige Vorbereitungen.

Satz 21: *Sei X eine Zufallsvariable mit Wahrscheinlichkeitsdichte f. Sei $a \in \mathbb{R}$, $b > 0$ und $Y = \frac{X-a}{b}$. Dann hat Y die Wahrscheinlichkeitsdichte $g(x) = bf(bx+a)$.*

Beweis: Sei $F(t) = P(X \leq t)$ die Verteilungsfunktion von X, sodass $F'(t) = f(t)$ gilt. Es folgt $P(Y \leq t) = P(\frac{X-a}{b} \leq t) = P(X \leq bt+a) = F(bt+a)$. Die Wahrscheinlichkeitsdichte g von Y ist die Ableitung dieser Funktion, also $g(t) = F'(bt+a)b = bf(bt+a)$. □

Nun können wir eine $N(\mu,\sigma)$-verteilte Zufallsvariable auf eine $N(0,1)$-verteilte Zufallsvariable zurückführen. Die $N(0,1)$-Verteilung nennt man auch Standardnormalverteilung.

Satz 22: *Wenn X die $N(\mu,\sigma)$-Verteilung hat, dann hat $Y = \frac{X-\mu}{\sigma}$ die $N(0,1)$-Verteilung.*

Beweis: Die Wahrscheinlichkeitsdichte von X ist $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$. Nach Satz 21 hat Y die Wahrscheinlichkeitsdichte $g(x) = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-(\sigma x+\mu-\mu)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Das ist die Wahrscheinlichkeitsdichte der $N(0,1)$-Verteilung. □

Um die Wahrscheinlichkeitsdichte der $N(0,1)$-Verteilung nicht immer ausschreiben zu müssen, wurde dafür eine Bezeichnung eingeführt. Man setzt $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Man kann zeigen, dass $\int_{-\infty}^{\infty} e^{-x^2/2} \, \mathrm{d}x = \sqrt{2\pi}$ gilt, das heißt $\int_{-\infty}^{\infty} \varphi(x) \, \mathrm{d}x = 1$. Daraus folgt dann $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \, \mathrm{d}x = 1$. Die Verteilungsfunktion der $N(0,1)$-Verteilung wird mit Φ bezeichnet, das heißt $\Phi(t) = \int_{-\infty}^{t} \varphi(x) \, \mathrm{d}x$ für $t \in \mathbb{R}$.

Satz 23: *Die Funktionen φ und Φ haben folgende Eigenschaften*
(a) *Es gilt $\varphi(-x) = \varphi(x)$ für alle $x \in \mathbb{R}$.*
(b) *Es gilt $\Phi(-t) = 1 - \Phi(t)$ für alle $t \in \mathbb{R}$.*
(c) *Es gilt $u \le v \Leftrightarrow \Phi(u) \le \Phi(v)$ und $u < v \Leftrightarrow \Phi(u) < \Phi(v)$ für alle $u, v \in \mathbb{R}$.*
(d) *Die Abbildung $\Phi : \mathbb{R} \to (0,1)$ ist bijektiv.*

Beweis: Es gilt (a), da in der Formel für $\varphi(x)$ nur x^2 vorkommt, und (b) rechnet man nach $\Phi(-t) = \int_{-\infty}^{-t} \varphi(x)\,\mathrm{d}x = \int_{-\infty}^{-t} \varphi(-x)\,\mathrm{d}x = -\int_{\infty}^{t} \varphi(y)\,\mathrm{d}y = \int_{t}^{\infty} \varphi(y)\,\mathrm{d}y = 1 - \Phi(t)$, wobei zuerst (a) verwendet und dann die Integralsubstitution $y = -x$ durchgeführt wurde.
Es gilt $\Phi'(t) = \varphi(t) > 0$ für alle $t \in \mathbb{R}$. Nach dem Mittelwertsatz der Differentialrechnung existiert ein t zwischen u und v, für das $\Phi(v) - \Phi(u) = \varphi(t)(v - u)$ gilt. Daraus folgt dann $v - u \ge 0 \Rightarrow \Phi(v) - \Phi(u) \ge 0$ und $v - u > 0 \Rightarrow \Phi(v) - \Phi(u) > 0$. Wir können diese Gleichung auch als $v - u = \frac{1}{\varphi(t)}(\Phi(v) - \Phi(u))$ schreiben. Daraus folgt $\Phi(v) - \Phi(u) \ge 0 \Rightarrow v - u \ge 0$ und $\Phi(v) - \Phi(u) > 0 \Rightarrow v - u > 0$. Damit ist (c) gezeigt.
Nach (c) ist die Abbildung $\Phi : \mathbb{R} \to (0,1)$ streng monoton wachsend und daher injektiv. In einer Bemerkung nach Satz 15 wurde $\lim_{s\to\infty} \Phi(s) = 1$ und $\lim_{r\to-\infty} \Phi(r) = 0$ gezeigt. Aus dem Zwischenwertsatz und der Stetigkeit von Φ folgt, dass für jedes $u \in (0,1)$ ein $t \in \mathbb{R}$ existiert mit $\Phi(t) = u$. Die Abbildung $\Phi : \mathbb{R} \to (0,1)$ ist surjektiv. Es folgt (d). □

Tritt in einem Beispiel eine normalverteilte Zufallsvariable auf, so führt man sie mit Satz 22 auf eine $N(0,1)$-verteilte Zufallsvariable Y zurück. Wahrscheinlichkeiten für diese Zufallsvariable kann man mit Hilfe ihrer Verteilungsfunktion Φ ausdrücken. Wegen Satz 15 gilt $P(Y \le t) = \Phi(t)$, $P(Y \ge t) = 1 - \Phi(t)$ und $P(s \le Y \le t) = \Phi(t) - \Phi(s)$, wobei statt $\le$ auch $<$ und statt $\ge$ auch $>$ stehen darf. Die Werte von $\Phi(t)$ für $t \ge 0$ findet man in einer Tabelle. Die Werte von $\Phi(t)$ für $t < 0$ findet man mit Hilfe der Formel $\Phi(-t) = 1 - \Phi(t)$.

Beispiel 49: Eine Maschine stellt Spanplatten her. Ihre Dicke ist $N(\mu, \sigma)$-verteilt mit $\mu = 19$ mm und $\sigma = 0.05$ mm. Die Platten sollen zwischen 18.95 mm und 19.10 mm stark sein. Wieviel Prozent Ausschuss produziert die Maschine?

Sei X die Dicke einer (zufällig gewählten) Platte. Dann hat X die $N(\mu, \sigma)$-Verteilung. Die Platte ist Ausschuss, wenn $X \le 18.95$ oder $X \ge 19.10$ gilt. Wir berechnen die Wahrscheinlichkeit, dass sie kein Ausschuss ist, das heißt, dass $18.95 < X < 19.10$ gilt.

$$\begin{aligned}
P(18.95 < X < 19.10) &= P(\tfrac{18.95-\mu}{\sigma} < Y < \tfrac{19.10-\mu}{\sigma}) && \text{wobei } Y = \tfrac{X-\mu}{\sigma} \\
&= P(-1 < Y < 2) \\
&= \Phi(2) - \Phi(-1) && \text{da } Y \text{ die } N(0,1)\text{-Verteilung hat} \\
&= \Phi(2) - 1 + \Phi(1) && \text{Satz 23 (b)} \\
&= 0.977 - 1 + 0.841 = 0.818 && \text{Tabelle}
\end{aligned}$$

Die Wahrscheinlichkeit, dass eine (zufällig gewählte) Platte Ausschuss ist, beträgt daher $1 - 0.818 = 0.182$. Die Maschine produziert 18.2% Ausschuss.

Beispiel 50: Auf einer Maschine werden Waschmittelpackungen abgefüllt. Der Inhalt der Packungen ist $N(\mu, \sigma)$-verteilt. Die Arbeitsgenauigkeit der Abfüllanlage ist bekannt und durch $\sigma = 0.05$ kg gegeben. Die durchschnittliche Abfüllmenge μ kann man einstellen.
(a) Die Maschine wird auf $\mu = 3$ kg eingestellt. Wie groß ist die Wahrscheinlichkeit, dass die am Etikett angegebene Mindestfüllmenge von 2.9 kg unterschritten wird?
(b) Es wird wieder $\mu = 3$ kg eingestellt. Welche Mindestfüllmenge t ist auf das Etikett zu drucken, wenn höchstens 1% der Packungen diese Menge t unterschreiten dürfen?

(c) Auf welchen Mittelwert μ muss man die Maschine einstellen, damit mindestens 99% der Packungen mehr als 2.9 kg enthalten?

(a) Sei X der Inhalt einer (zufällig gewählten) Packung. Gesucht ist $P(X < 2.9)$.

$$\begin{aligned} P(X < 2.9) &= P(Y < \tfrac{2.9-3}{0.05}) && \text{wobei } Y = \tfrac{X-\mu}{\sigma} = \tfrac{X-3}{0.05} \\ &= P(Y < -2) \\ &= \Phi(-2) && \text{da } Y \text{ die } N(0,1)\text{-Verteilung hat} \\ &= 1 - \Phi(2) && \text{Satz 23 (b)} \\ &= 1 - 0.977 = 0.023 && \text{Tabelle} \end{aligned}$$

Die Mindestfüllmenge wird von 2.3% der Packungen unterschritten.

(b) Gesucht ist t, sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt X kleiner als t ist, höchstens 0.01 beträgt. Es muss also $P(X < t) \le 0.01$ gelten. Da $\frac{X-\mu}{\sigma}$ die $N(0,1)$-Verteilung hat, erhalten wir $P(X < t) = P(\frac{X-\mu}{\sigma} < \frac{t-3}{0.05}) = \Phi(\frac{t-3}{0.05})$. Wir schreiben 0.01 als Funktionswert von Φ. In der Tabelle finden wir $\Phi(2.33) = 0.99$. Wegen Satz 23 (b) folgt $\Phi(-2.33) = 0.01$. Wir erhalten somit

$$P(X < t) \le 0.01 \quad \Leftrightarrow \quad \Phi(\tfrac{t-3}{0.05}) \le 0.01 \quad \Leftrightarrow \quad \Phi(\tfrac{t-3}{0.05}) \le \Phi(-2.33) \quad \Leftrightarrow \quad \tfrac{t-3}{0.05} \le -2.33$$

Für die letzte Äquivalenz haben wir Satz 23 (c) verwendet. Wir lösen die Ungleichung $\frac{t-3}{0.05} \le -2.33$ nach t auf und erhalten $t \le 2.884$. Druckt man eine Mindestfüllmenge, die ≤ 2.884 ist, auf das Etikett, dann ist die verlangte Forderung erfüllt.

(c) Gesucht ist μ, sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt X größer als 2.9 ist, mindestens 0.99 beträgt. Es muss also $P(X > 2.9) \ge 0.99$ gelten. Da $\frac{X-\mu}{\sigma}$ die $N(0,1)$-Verteilung hat, erhalten wir $P(X > 2.9) = P(\frac{X-\mu}{\sigma} > \frac{2.9-\mu}{0.05}) = 1 - \Phi(\frac{2.9-\mu}{0.05})$. Wir müssen wieder 0.01 als Funktionswert von Φ ausdrücken. Wie oben finden wir $\Phi(2.33) = 0.99$ und daraus dann $\Phi(-2.33) = 0.01$. Wir erhalten somit

$$P(X > 2.9) \ge 0.99 \;\Leftrightarrow\; 1 - \Phi(\tfrac{2.9-\mu}{0.05}) \ge 0.99 \;\Leftrightarrow\; \Phi(\tfrac{2.9-\mu}{0.05}) \le \Phi(-2.33) \;\Leftrightarrow\; \tfrac{2.9-\mu}{0.05} \le -2.33$$

Für die letzte Äquivalenz haben wir Satz 23 (c) verwendet. Wir lösen die Ungleichung $\frac{2.9-\mu}{0.05} \le -2.33$ nach μ auf und erhalten $\mu \ge 3.016$. Stellt man die Maschine auf einen Mittelwert μ ein, der ≥ 3.016 ist, dann ist die verlangte Forderung erfüllt.

Schließlich berechnen wir noch Erwartungswert und Varianz einer $N(\mu, \sigma)$-verteilten Zufallsvariablen X. Durch Substitution der neuen Variablen $y = \frac{x-\mu}{\sigma}$ erhalten wir

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x = \int_{-\infty}^{\infty} (\sigma y + \mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \,\mathrm{d}y = \sigma \int_{-\infty}^{\infty} y\varphi(y) \,\mathrm{d}y + \mu \int_{-\infty}^{\infty} \varphi(y) \,\mathrm{d}y.$$

Wegen $\varphi'(y) = -y\varphi(y)$ ergibt sich $\int_{-\infty}^{\infty} y\varphi(y)\,\mathrm{d}y = -\varphi(y)\big|_{-\infty}^{\infty} = 0 - 0 = 0$. Weiters gilt auch $\int_{-\infty}^{\infty} \varphi(y)\,\mathrm{d}y = 1$. Somit folgt $\mathrm{E}(X) = \mu$. (Man sollte auch die Bedingung $\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x < \infty$ nachprüfen. Sie ergibt sich durch eine ähnliche Rechnung mit $|\sigma y + \mu| \le \sigma|y| + |\mu|$ und $\int_{-\infty}^{\infty} |y|\varphi(y)\,\mathrm{d}y = \int_0^{\infty} y\varphi(y)\,\mathrm{d}y - \int_{-\infty}^0 y\varphi(y)\,\mathrm{d}y = 2\varphi(0) < \infty$.)

Analog berechnen wir die Varianz der $N(\mu, \sigma)$-verteilten Zufallsvariable X. Durch Substitution der neuen Variablen $y = \frac{x-\mu}{\sigma}$ erhalten wir

$$\mathrm{V}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x = \int_{-\infty}^{\infty} (\sigma y)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \,\mathrm{d}y = \sigma^2 \int_{-\infty}^{\infty} y^2 \varphi(y) \,\mathrm{d}y.$$

Wir berechnen $\int_{-\infty}^{\infty} y^2\varphi(y)\,\mathrm{d}y$. Wegen $\varphi'(y) = -y\varphi(y)$ folgt mit partieller Integration $\int_{-\infty}^{\infty} y \cdot y\varphi(y)\,\mathrm{d}y = -y\varphi(y)\big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(y)\,\mathrm{d}y = 0 + 1 = 1$. Somit erhalten wir $\mathrm{V}(X) = \sigma^2$.

8. Approximation der Binomialverteilung

Es ist mühsam, mit dem Wahrscheinlichkeitsvektor der Binomialverteilung zu rechnen. Deshalb wird die Binomialverteilung oft durch die Normalverteilung approximiert.

Die Binomialverteilung ist diskret, während die Normalverteilung kontinuierlich ist. Deshalb führen wir eine Art Wahrscheinlichkeitsdichte für eine $B(n,p)$-verteilte Zufallsvariable X ein. Sei $g : \mathbb{R} \to \mathbb{R}^+$ definiert durch

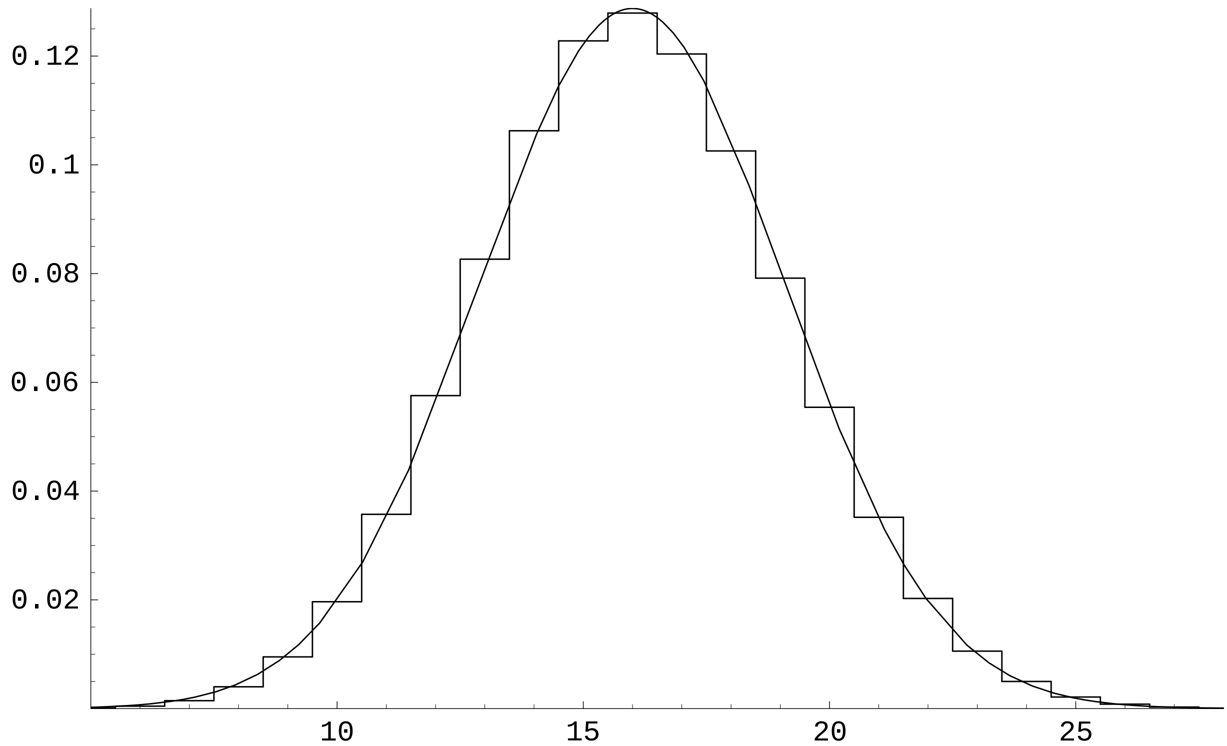
$$g(x) = w(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{für} \quad x \in (i - \tfrac{1}{2}, i + \tfrac{1}{2}] \quad \text{und für} \quad 0 \leq i \leq n$$

und $g(x) = 0$ für $x \notin (-\frac{1}{2}, n + \frac{1}{2}]$. Durch g ist eine Treppenfunktion definiert, sodass die Fläche des Rechtecks, das über dem Intervall $(i - \frac{1}{2}, i + \frac{1}{2}]$ und unter g liegt, gleich $w(i)$ ist. Deshalb ist $\sum_{i=0}^{k} w(i)$ gerade die links von $k + \frac{1}{2}$ liegende Fläche unter der Treppenfunktion g. Für $k \in \{0, 1, \ldots, n\}$ gilt daher $P(X \leq k) = \sum_{i=0}^{k} w(i) = \int_{-\infty}^{k+\frac{1}{2}} g(x)\,\mathrm{d}x$.

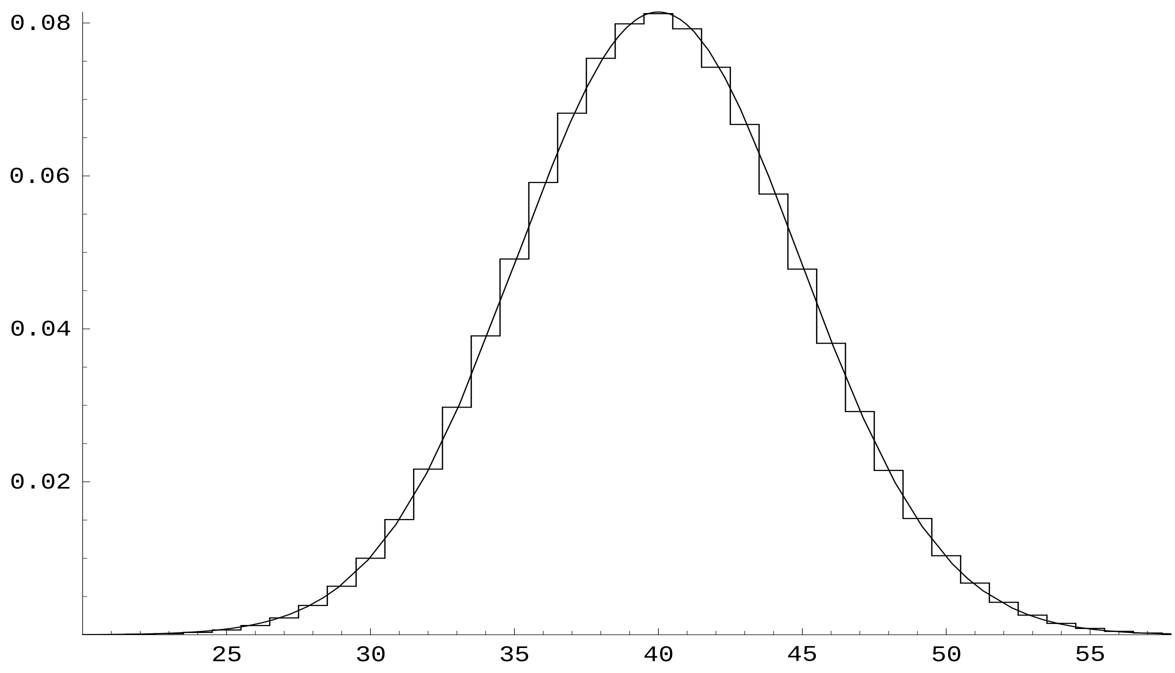
Vergleicht man die Treppenfunktion g mit der Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für} \quad x \in \mathbb{R}$$

so sieht man, dass die beiden Funktionen, abgesehen von den Sprungstellen, eine ähnliche Form haben. Durch die Wahl von μ kann man die Funktion f nach links oder rechts verschieben, und durch die Wahl von σ kann man f breiter und niedriger oder schmäler und höher machen. (Die Fläche unter f ist immer 1.) Wir wählen μ und σ so, dass Erwartungswert und Standardabweichung der Binomialverteilung und der Normalverteilung übereinstimmen, also $\mu = np$ und $\sigma = \sqrt{np(1-p)}$. Dadurch wird die Wahrscheinlichkeitsdichte f recht gut an die Treppenfunktion g angepasst. Die folgende Abbildung zeigt g mit $n = 40$ und $p = 0.4$ und die dazupassende Wahrscheinlichkeitsdichte der Normalverteilung.



Die Approximation ist umso besser, je größer n ist. Wenn man n vergrößert, werden die Sprünge der Treppenfunktion kleiner. Die folgende Abbildung zeigt g mit $n = 100$ und $p = 0.4$ und die dazupassende Wahrscheinlichkeitsdichte der Normalverteilung.



Als Faustregel, wann n groß genug ist, um eine ausreichende Approximation zu gewährleisten, verwendet man $\sigma = \sqrt{np(1-p)} \geq 3$.

Man kann daher eine $B(n,p)$-verteilte Zufallsvariable X näherungsweise so behandeln, als ob sie $N(\mu, \sigma)$-verteilt wäre mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$. Zu Beginn dieses Kapitels haben wir uns für ganzzahliges k überlegt, dass $P(X \leq k) = \int_{-\infty}^{k+\frac{1}{2}} g(x)\,\mathrm{d}x$ gilt. Mit Hilfe dieser Gleichung erhalten wir für $k \in \{0, 1, \ldots, n\}$

$$P(X \leq k) = \int_{-\infty}^{k+\frac{1}{2}} g(x)\,\mathrm{d}x \approx \int_{-\infty}^{k+\frac{1}{2}} f(x)\,\mathrm{d}x = \int_{-\infty}^{\frac{k+\frac{1}{2}-\mu}{\sigma}} \varphi(y)\,\mathrm{d}y = \Phi(\tfrac{k+\frac{1}{2}-\mu}{\sigma})$$

wobei die neue Integrationsvariable $y = \frac{x-\mu}{\sigma}$ eingeführt wurde. Wir erhalten

$$P(X \leq k) \approx \Phi(\tfrac{k+\frac{1}{2}-\mu}{\sigma}) \quad \text{für} \quad k \in \{0, 1, \ldots, n\}$$

als Approximationsformel, die die Wahrscheinlichkeiten einer $B(n,p)$-verteilten Zufallsvariable X mit Hilfe der Verteilungsfunktion Φ der Standardnormalverteilung approximiert.

Bemerkung: Ist Y eine $N(\mu, \sigma)$-verteilte Zufallsvariable und f ihre Wahrscheinlichkeitsdichte, dann gilt $P(Y \leq t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$ für alle $t \in \mathbb{R}$. Vielleicht möchte man diese Formel direkt als Normalverteilungsapproximation für eine $B(n,p)$-verteilte Zufallsvariable X übernehmen. Man erhält $P(X \leq k) \approx \int_{-\infty}^{k} f(x)\,\mathrm{d}x = \Phi(\frac{k-\mu}{\sigma})$ für $k \in \{0, 1, 2, \ldots, n\}$. Das ist jedoch problematisch. Für die $B(n,p)$-verteilte Zufallsvariable X und für ganzzahliges k gilt nämlich $P(X \leq k) = P(X \leq t)$ für alle $t \in [k, k+1)$. Durch $X \leq t$ wird für jede Zahl t im Intervall $[k, k+1)$ dasselbe Ereignis dargestellt. Welche dieser Zahlen soll man dann als obere Integrationsgrenze wählen? Auch daraus ist zu ersehen, dass vermutlich der Mittelpunkt dieses Intervalls, das ist die Zahl $k + \frac{1}{2}$, am besten als obere Integrationsgrenze geeignet ist.

In den Beispielen hat man es mit Wahrscheinlichkeiten $P(X \leq k)$ für ganzzahliges k oder mit $P(k_1 \leq X \leq k_2) = P(X \leq k_2) - P(X \leq k_1 - 1)$ für ganzzahlige k_1 und k_2 zu tun, sodass man diese Approximation verwenden kann.

Beispiel 51: Auf einer Maschine werden Schrauben erzeugt, von denen 1% schadhaft sind. Die Schrauben werden in Packungen zu 1600 Stück verkauft.
(a) Wie groß ist die Wahrscheinlichkeit, dass man höchstens 9 schadhafte Schrauben in einer Packung findet?
(b) Welche Anzahl von brauchbaren Schrauben in einer Packung kann der Hersteller garantieren, wenn die Garantie von höchstens 2% der Packungen unterschritten werden darf?

(a) Sei X die Anzahl der schadhaften Schrauben in einer (zufällig gewählten) Packung. Dann hat X die $B(n,p)$-Verteilung mit $n = 1600$ und $p = 0.01$. Gesucht ist die Wahrscheinlichkeit, dass $X \leq 9$ gilt.

Um das zu berechnen, verwenden wir die Approximation durch die Normalverteilung. Wir haben $\mu = np = 1600 \cdot 0.01 = 16$ und $\sigma = \sqrt{np(1-p)} = \sqrt{1600 \cdot 0.01 \cdot 0.99} = 3.98$. Wegen $\sigma \geq 3$ ist die Approximation ausreichend gut. Wir erhalten

$$P(X \leq 9) \approx \Phi(\tfrac{9+\frac{1}{2}-16}{3.98}) = \Phi(-1.63) = 1 - \Phi(1.63) = 1 - 0.95 = 0.05$$

wobei $\Phi(1.63)$ in einer Tabelle gefunden wurde.

Die Wahrscheinlichkeit, dass höchstens 9 schadhafte Schrauben in einer Packung sind, beträgt 0.05.

(b) Sei Y die Anzahl der brauchbaren Schrauben in einer (zufällig gewählten) Packung und k die Anzahl der brauchbaren Schrauben, die der Hersteller in einer Packung garantieren kann. Gesucht ist k, sodass $P(Y < k) \leq 0.02$ gilt. Die Zufallsvariable Y hat jetzt die $B(n,p)$-Verteilung mit $n = 1600$ und $p = 0.99$.

Wir verwenden die Approximation durch die Normalverteilung. Dazu berechnen wir $\mu = np = 1600 \cdot 0.99 = 1584$ und $\sigma = \sqrt{np(1-p)} = \sqrt{1600 \cdot 0.99 \cdot 0.01} = 3.98$. Wegen $\sigma \geq 3$ ist die Approximation ausreichend gut.

Wegen $P(Y < k) = P(Y \leq k-1) \approx \Phi(\frac{k-1+\frac{1}{2}-1584}{3.98}) = \Phi(\frac{k-1584.5}{3.98})$ ist k so zu bestimmen, dass $\Phi(\frac{k-1584.5}{3.98}) \leq 0.02$ gilt. Wir finden $\Phi(2.05) = 0.98$ in einer Tabelle. Daraus folgt $\Phi(-2.05) = 0.02$. Somit ist k so zu bestimmen, dass $\Phi(\frac{k-1584.5}{3.98}) \leq \Phi(-2.05)$ gilt. Nach Satz 23 (c) ist diese Ungleichung äquivalent zu $\frac{k-1584.5}{3.98} \leq -2.05$. Löst man diese Ungleichung auf, dann erhält man $k \leq 1576.3$.

Der Hersteller kann also 1576 brauchbare Schrauben pro Packung garantieren, natürlich auch jede kleinere Anzahl.

Beispiel 52: Bei einem Fährunternehmen weiß man, dass durchschnittlich 10% der Personen, die einen Platz für ihren PKW gebucht haben, zur Abfahrt nicht erscheinen. Es gibt 120 PKW–Plätze. Wieviele Buchungen dürfen höchstens vorgenommen werden, wenn die Wahrscheinlichkeit, dass alle, die zur Abfahrt erscheinen, einen Platz bekommen, mindestens 0.95 betragen soll?

Sei n die Anzahl der vorgenommenen Buchungen und X die Anzahl der Gebuchten, die auch zur Abfahrt erscheinen. Jeder der n Gebuchten erscheint mit Wahrscheinlichkeit 0.9. Daher hat X die $B(n,p)$-Verteilung mit $p = 0.9$. Wir müssen n so bestimmen, dass $P(X \leq 120) \geq 0.95$ gilt.

Wir verwenden die Approximation durch die Normalverteilung. Dazu berechnen wir $\mu = np = 0.9 \cdot n$ und $\sigma = \sqrt{np(1-p)} = 0.3\sqrt{n}$. Da n jedenfalls ≥ 120 sein muss, ist σ jedenfalls ≥ 3, und wir können die Approximation verwenden.

Wegen $P(X \leq 120) = \Phi(\frac{120.5-0.9\cdot n}{0.3\sqrt{n}})$ und $0.95 = \Phi(1.65)$ ist n so zu bestimmen, dass $\Phi(\frac{120.5-0.9\cdot n}{0.3\sqrt{n}}) \geq \Phi(1.65)$ gilt. Diese Ungleichung ist äquivalent zu $\frac{120.5-0.9\cdot n}{0.3\sqrt{n}} \geq 1.65$ nach Satz 23 (c). Durch Umformen erhalten wir dann

$$n + 0.55\sqrt{n} - 133.9 \leq 0$$

Setzt man $x = \sqrt{n}$ so erhält man die quadratische Ungleichung $x^2 + 0.55x - 133.9 \leq 0$. Die durch die Funktion $x \mapsto x^2 + 0.55x - 133.9$ dargestellte Parabel ist nach oben offen. Die Ungleichung ist also für alle x zwischen x_1 und x_2 erfüllt, wobei x_1 und x_2 die beiden Nullstellen der Parabel sind. Löst man die quadratische Gleichung $x^2+0.55x-133.9 = 0$, so erhält man $x_1 = -11.85$ und $x_2 = 11.3$. Da $x = \sqrt{n}$ aber positiv sein muss, erhalten wir als Lösungsmenge für x das Intervall $[0, 11.3]$ und als Lösungsmenge für $n = x^2$ das Intervall $[0, 127.7]$.

Damit alle, die zur Abfahrt erscheinen, mit einer Wahrscheinlichkeit von mindestens 0.95 einen Platz erhalten, dürfen höchstens 127 Buchungen vorgenommen werden.

Beispiel 53: Wie oft muss man würfeln, damit mit Wahrscheinlichkeit 0.9 mindestens dreimal 6 unter den gewürfelten Ergebnissen auftritt?

Sei n die Anzahl der Würfe und X die Anzahl, mit der 6 auftritt. Dann hat X die $B(n,p)$-Verteilung mit $p = \frac{1}{6}$. Wir müssen n so bestimmen, dass $P(X \geq 3) \geq 0.9$ gilt.

Wir verwenden die Approximation durch die Normalverteilung. Wir finden $\mu = np = \frac{n}{6}$ und $\sigma = \sqrt{np(1-p)} = \frac{\sqrt{5n}}{6}$. Es gilt $P(X \geq 3) = 1 - P(X \leq 2) \approx 1 - \Phi(\frac{2.5-\mu}{\sigma})$. Wegen $P(X \geq 3) \geq 0.9$ erhält man $\Phi(\frac{2.5-\mu}{\sigma}) \leq 0.1$. Nun gilt $\Phi(1.28) = 0.9$ und somit $\Phi(-1.28) = 0.1$. Nach Satz 23 (c) ist diese Ungleichung äquivalent zu $\frac{2.5-\mu}{\sigma} \leq -1.28$ oder $\mu - 2.5 - 1.28\sigma \geq 0$. Setzt man für μ und σ ein, so hat man $n - 1.28\sqrt{5}\sqrt{n} - 15 \geq 0$. Die Lösungen der quadratischen Gleichung $x^2 - 1.28\sqrt{5}\,x - 15 = 0$ sind $x_1 = -2.698$ und $x_2 = 5.560$. Es folgt $\sqrt{n} \leq -2.698$ oder $\sqrt{n} \geq 5.560$. Da das erste nicht möglich ist, ergibt sich $\sqrt{n} \geq 5.560$ oder $n \geq 30.91$. Man muss mindestens 31 Mal würfeln, um mit Wahrscheinlichkeit 0.9 mindestens dreimal 6 zu erhalten.

Da für $n = 31$ aber $\sigma < 3$ gilt, sollte man die Probe mit Hilfe der Binomialverteilung machen: $P(X \geq 3) = 1 - P(X \leq 2) = 1 - (\frac{5}{6})^{31} - \binom{31}{1}\frac{1}{6}(\frac{5}{6})^{30} - \binom{31}{2}(\frac{1}{6})^2(\frac{5}{6})^{29} = 0.9094$. Für $n = 30$ würde man 0.8972 erhalten, also weniger als 0.9.

IV. Mehrere Zufallsvariable

In diesem Kapitel arbeiten wir gleichzeitig mit mehreren Zufallsvariablen. Es geht unter anderem um gemeinsame Verteilungen von Zufallsvariablen, um Summe und Quotient von zwei Zufallsvariablen und um Rechenregeln für den Erwartungswert.

1. Gemeinsame Verteilungen von Zufallsvariablen

Um zwei Zufallsvariable X und Y gleichzeitig untersuchen zu können, geben wir folgende Definitionen und beweisen einige Sätze.

Definition: Seien S und T endliche oder abzählbare Mengen. Ein Vektor $(u(i,j))_{i\in S,j\in T}$ mit nicht negativen Eintragungen und $\sum_{(i,j)\in S\times T} u(i,j) = 1$ heißt gemeinsamer Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y, wenn $u(i,j) = P(X = i \ \wedge \ Y = j)$ für alle $i \in S$ und $j \in T$ gilt. Wir fassen die Zufallsvariablen X und Y zu einem Vektor (X,Y) zusammen. Solche Vektoren nennt man Zufallsvektoren.

Satz 24: *Sei $(u(i,j))_{i\in S,j\in T}$ der gemeinsame Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y. Für jede Teilmenge B von $S \times T$ gilt $P((X,Y) \in B) = \sum_{(i,j)\in B} u(i,j)$.*

Beweis: Der Beweis ist derselbe wie für Satz 13. Im Beweis von Satz 13 ersetzt man S durch $S \times T$, X durch (X,Y), i durch (i,j) und $w(i)$ durch $u(i,j)$. □

Satz 25: *Sei $(u(i,j))_{i\in S,j\in T}$ der gemeinsame Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y. Für $i \in S$ sei $w_1(i) = \sum_{j\in T} u(i,j)$ und für $j \in T$ sei $w_2(j) = \sum_{i\in S} u(i,j)$. Dann ist $(w_1(i))_{i\in S}$ der Wahrscheinlichkeitsvektor der Zufallsvariable X und $(w_2(j))_{j\in T}$ der Wahrscheinlichkeitsvektor der Zufallsvariable Y.*

Beweis: Sei $j \in T$ beliebig und $H_j = \{(i,j) : i \in S\}$. Dann gilt $H_j \subseteq S\times T$. Mit Hilfe von Satz 24 folgt $P(Y = j) = P((X,Y) \in H_j) = \sum_{(m,n)\in H_j} u(m,n) = \sum_{i\in S} u(i,j)$. Damit ist $P(Y = j) = w_2(j)$ für alle $j \in T$ gezeigt.
Wir wenden Satz 56 mit $M = S \times T$ und $G = T$ an. Wegen $\sum_{(m,n)\in S\times T} u(m,n) = 1$ und $w_2(j) = \sum_{(m,n)\in H_j} u(m,n)$ für $j \in T$ ergibt sich $\sum_{j\in T} w_2(j) = \sum_{(m,n)\in S\times T} u(m,n) = 1$. Wir haben bewiesen, dass $(w_2(j))_{j\in T}$ der Wahrscheinlichkeitsvektor der Zufallsvariablen Y ist. Der Beweis für die Zufallsvariable X läuft analog. □

Beispiel 54: Sei X der Ausfall eines Münzenwurfs. Ergibt der Münzenwurf 0, dann ziehen wir aus $1,1,1,2,3,3$. Ergibt der Münzenwurf 1, dann ziehen wir aus $1,2,2,2,2,3$. Sei Y das Ziehungsergebnis. Gesucht ist der gemeinsame Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y.

Sei $S = \{0,1\}$ die Menge aller möglichen Ausfälle des Münzenwurfs und $T = \{1,2,3\}$ die Menge aller möglichen Ziehungsergebnisse. Es gilt dann $P(X = 0 \wedge Y = 1) = P(X = 0)P(Y = 1|X = 0) = \frac{1}{2} \cdot \frac{3}{6} = \frac{1}{4}$. Somit haben wir $u(0,1) = \frac{1}{4}$. Analog folgt $u(0,2) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$ und $u(0,3) = \frac{1}{2} \cdot \frac{2}{6} = \frac{1}{6}$. In den Fällen, wo der Münzenwurf 1 ergibt, erhalten wir $u(1,1) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$, $u(1,2) = \frac{1}{2} \cdot \frac{4}{6} = \frac{1}{3}$ und $u(1,3) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$. Damit ist der gemeinsame Wahrscheinlichkeitsvektor $(u(i,j))_{i\in S,j\in T}$ von X und Y berechnet.

Es folgt $w_1(0) = u(0,1) + u(0,2) + u(0,3) = \frac{1}{2}$ und $w_1(1) = u(1,1) + u(1,2) + u(1,3) = \frac{1}{2}$. Das ist der Wahrscheinlichkeitsvektor der Zufallsvariable X. Ebenso erhalten wir $w_2(1) = u(0,1) + u(1,1) = \frac{1}{3}$, $w_2(2) = u(0,2) + u(1,2) = \frac{5}{12}$ und $w_2(3) = u(0,3) + u(1,3) = \frac{1}{4}$.

Das ist der Wahrscheinlichkeitsvektor der Zufallsvariable Y, der die Wahrscheinlichkeit für das Ziehen jeder der drei Zahlen angibt.

Definition: Eine integrierbare Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$ mit $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x = 1$ heißt gemeinsame Wahrscheinlichkeitsdichte der beiden Zufallsvariablen X und Y, wenn $P(X \leq s \ \wedge\ Y \leq t) = \int_{-\infty}^{s}\int_{-\infty}^{t} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$ für alle $s \in \mathbb{R}$ und alle $t \in \mathbb{R}$ gilt.

Die Integration von Funktionen $g : \mathbb{R}^2 \to \mathbb{R}^+$, insbesondere die Berechnung eines Integrals über einen Normalbereich, wird im Anhang behandelt.

Satz 26: *Sei $g : \mathbb{R}^2 \to [0,\infty)$ gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen X und Y. Dann gilt $\mathrm{P}((X,Y) \in B) = \int_B g(x,y)\,\mathrm{d}(x,y)$ für alle Normalbereiche B des $\mathbb{R}^2$, die auch unbeschränkt sein können.*

Beweis: Wir führen den Beweis nur für eine beschränkte Funktion g durch. Sei B zuerst ein Rechteck $C = (r,s] \times (u,v]$ mit $r < s$ und $u < v$. Wegen des Additionssatzes gilt $\mathrm{P}(X \leq s \wedge Y \leq v) = \mathrm{P}(X \leq s \wedge Y \leq u) + \mathrm{P}(X \leq s \wedge u < Y \leq v)$ und aus obiger Definition folgt $\mathrm{P}(X \leq s \wedge Y \leq v) - \mathrm{P}(X \leq s \wedge Y \leq u) = \int_{-\infty}^{s}\int_{-\infty}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x - \int_{-\infty}^{s}\int_{-\infty}^{u} g(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{s}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$. Setzt man das zusammen, dann ergibt sich

$$\mathrm{P}(X \leq s \wedge u < Y \leq v) = \int_{-\infty}^{s}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

Ersetzt man in dieser Gleichung s durch r, so hat man auch

$$\mathrm{P}(X \leq r \wedge u < Y \leq v) = \int_{-\infty}^{r}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

Der Additionssatz liefert auch $\mathrm{P}(X \leq s \wedge u < Y \leq v) = \mathrm{P}(X \leq r \wedge u < Y \leq v) + \mathrm{P}(r < X \leq s \wedge u < Y \leq v)$. Somit ist $\mathrm{P}(r < X \leq s \wedge u < Y \leq v)$ nach obigen Gleichungen gleich $\int_{-\infty}^{s}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x - \int_{-\infty}^{r}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{r}^{s}\int_{u}^{v} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$. Damit haben wir $\mathrm{P}((X,Y) \in C) = \int_C g(x,y)\,\mathrm{d}(x,y)$ für das Rechteck $C = (r,s] \times (u,v]$ gezeigt.

Der Bereich $R = \{(x,y) \in \mathbb{R}^2 : a < x \leq b,\ \varphi(x) < y \leq \psi(x)\}$ sei ein Rechteckbereich, das heißt es gilt $a < b$ und die Funktionen $\varphi : (a,b] \to \mathbb{R}$ und $\psi : (a,b] \to \mathbb{R}$ sind Treppenfunktionen mit $\varphi \leq \psi$. Es existiert eine Zerlegung $a = c_0 < c_1 < \cdots < c_n = b$ des Intervalls $(a,b]$, sodass die Funktionen φ und ψ auf den Intervallen $(c_{i-1}, c_i]$ für $1 \leq i \leq n$ konstant sind. Die Mengen $C_i = \{(x,y) \in \mathbb{R}^2 : c_{i-1} < x \leq c_i,\ \varphi(x) < y \leq \psi(x)\}$ für $1 \leq i \leq n$ sind paarweise disjunkte Rechtecke, deren Vereinigung R ist, wobei ein Rechteck auch die leere Menge sein kann. Wir haben dann die Gleichung $\int_a^b u(x)\,\mathrm{d}x = \sum_{i=1}^{n}\int_{c_{i-1}}^{c_i} u(x)\,\mathrm{d}x$. Mit $u(x) = \int_{\varphi(x)}^{\psi(x)} g(x,y)\,\mathrm{d}y$ erhalten wir $\int_R g(x,y)\,\mathrm{d}(x,y) = \sum_{i=1}^{n}\int_{C_i} g(x,y)\,\mathrm{d}(x,y)$. Aus dem Additionssatz folgt $\mathrm{P}((X,Y) \in R) = \sum_{i=1}^{n}\mathrm{P}((X,Y) \in C_i)$, da das Ereignis $(X,Y) \in R$ genau dann eintritt, wenn eines der Ereignisse $(X,Y) \in C_i$ für $1 \leq i \leq n$ eintritt, und diese Ereignisse außerdem unvereinbar sind. Da aber $\mathrm{P}((X,Y) \in C) = \int_C g(x,y)\,\mathrm{d}(x,y)$ für alle Rechtecke C bereits oben bewiesen wurde, ist $\mathrm{P}((X,Y) \in R) = \int_R g(x,y)\,\mathrm{d}(x,y)$ für alle Rechteckbereiche R gezeigt.

Sei B ein beschränkter Normalbereich. Sei $\varepsilon > 0$. Da wir g als beschränkt annehmen, existieren Rechteckbereiche R_1 und R_2, mit folgenden Eigenschaften (siehe Anhang)

$$R_1 \subseteq B \subseteq R_2 \quad \text{und} \quad \int_{R_2} g(x,y)\,\mathrm{d}(x,y) - \int_{R_1} g(x,y)\,\mathrm{d}(x,y) < \varepsilon$$

Aus $R_1 \subseteq B \subseteq R_2$ folgt $\mathrm{P}((X,Y) \in R_1) \leq \mathrm{P}((X,Y) \in B) \leq \mathrm{P}((X,Y) \in R_2)$ und auch $\int_{R_1} g(x,y)\,\mathrm{d}(x,y) \leq \int_B g(x,y)\,\mathrm{d}(x,y) \leq \int_{R_2} g(x,y)\,\mathrm{d}(x,y)$, da ja $g \geq 0$ gilt. Wir haben $\mathrm{P}((X,Y) \in R_1) = \int_{R_1} g(x,y)\,\mathrm{d}(x,y)$ und $\mathrm{P}((X,Y) \in R_2) = \int_{R_2} g(x,y)\,\mathrm{d}(x,y)$ oben bewiesen. Daher liegen $\mathrm{P}((X,Y) \in B)$ und $\int_B g(x,y)\,\mathrm{d}(x,y)$ in einem Intervall, das Länge

ε hat. Damit ist $|\mathrm{P}((X,Y)\in B)-\int_B g(x,y)\,\mathrm{d}(x,y)|<\varepsilon$ gezeigt. Da $\varepsilon>0$ beliebig klein gewählt werden kann, muss $\mathrm{P}((X,Y)\in B)=\int_B g(x,y)\,\mathrm{d}(x,y)$ gelten.

Jetzt zu den unbeschränkten Normalbereichen. Sei $\varepsilon>0$. Wegen $\int_{\mathbb{R}^2} g(x,y)\,\mathrm{d}(x,y)=1$ finden wir reelle Zahlen $m>0$ und $n>0$, sodass $\int_U g(x,y)\,\mathrm{d}(x,y)>1-\varepsilon$ für das beschränkte Rechteck $U=[-m,m]\times[-n,n]$ gilt. Wir setzen $V=\mathbb{R}^2\setminus U$. Es gilt dann $\int_V g(x,y)\,\mathrm{d}(x,y)=\int_{\mathbb{R}^2} g(x,y)\,\mathrm{d}(x,y)-\int_U g(x,y)\,\mathrm{d}(x,y)=1-\int_U g(x,y)\,\mathrm{d}(x,y)<\varepsilon$. Aus dem Additionssatz folgt $\mathrm{P}((X,Y)\in U)+\mathrm{P}((X,Y)\in V)=\mathrm{P}((X,Y)\in\mathbb{R}^2)\le 1$. Da U ein beschränkter Normalbereich ist, ist $\mathrm{P}((X,Y)\in U)=\int_U g(x,y)\,\mathrm{d}(x,y)>1-\varepsilon$ bereits gezeigt. Es folgt $\mathrm{P}((X,Y)\in V)<\varepsilon$.

Sei B ein unbeschränkter Normalbereich und $C=B\setminus V=B\cap U$. Dann ist C ein beschränkter Normalbereich. Es gilt $C\subseteq B\subseteq C\cup V$. Außerdem sind C und V disjunkt. Wegen $\mathrm{P}((X,Y)\in C\cup V)=\mathrm{P}((X,Y)\in C)+\mathrm{P}((X,Y)\in V)<\mathrm{P}((X,Y)\in C)+\varepsilon$ folgt

$$\mathrm{P}((X,Y)\in C)\le\mathrm{P}((X,Y)\in B)<\mathrm{P}((X,Y)\in C)+\varepsilon.$$

Da $\int_{C\cup V} g(x,y)\,\mathrm{d}(x,y)=\int_C g(x,y)\,\mathrm{d}(x,y)+\int_V g(x,y)\,\mathrm{d}(x,y)<\int_C g(x,y)\,\mathrm{d}(x,y)+\varepsilon$ ebenso nach obigen Resultaten gilt, erhalten wir auch

$$\int_C g(x,y)\,\mathrm{d}(x,y)\le\int_B g(x,y)\,\mathrm{d}(x,y)<\int_C g(x,y)\,\mathrm{d}(x,y)+\varepsilon.$$

Da C beschränkt ist, wurde $\mathrm{P}((X,Y)\in C)=\int_C g(x,y)\,\mathrm{d}(x,y)$ bereits oben bewiesen. Aus den soeben gezeigten Ungleichungen folgt dann $|\mathrm{P}((X,Y)\in B)-\int_B g(x,y)\,\mathrm{d}(x,y)|<\varepsilon$. Da $\varepsilon>0$ beliebig klein gewählt werden kann, erhalten wir $\mathrm{P}((X,Y)\in B)=\int_B g(x,y)\,\mathrm{d}(x,y)$. (Der Bereich V ist kein Normalbereich, jedoch lässt sich $\int_V g(x,y)\,\mathrm{d}(x,y)$ genauso berechnen wie für einen Normalbereich, wobei statt des inneren Integrals eine Summe von zwei Integralen auftreten kann. Ähnliches gilt auch für den Bereich $C\cup V$.) $\square$

Satz 27: *Sei $g:\mathbb{R}^2\to[0,\infty)$ eine gemeinsame Wahrscheinlichkeitsdichte der beiden Zufallsvariablen X und Y. Für $x\in\mathbb{R}$ sei $f_1(x)=\int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y$ und für $y\in\mathbb{R}$ sei $f_2(y)=\int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}x$. Dann ist f_1 eine Wahrscheinlichkeitsdichte von X und f_2 eine Wahrscheinlichkeitsdichte von Y.*

Beweis: Sei $t\in\mathbb{R}$ beliebig. Sei $B=(-\infty,t]\times\mathbb{R}$. Mit Hilfe von Satz 26 erhalten wir $P(X\le t)=P((X,Y)\in B)=\int_B g(x,y)\,\mathrm{d}(x,y)=\int_{-\infty}^{t}\int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$. Damit ist $P(X\le t)=\int_{-\infty}^{t} f_1(x)\,\mathrm{d}x$ für alle $t\in\mathbb{R}$ gezeigt, das heißt f_1 ist eine Wahrscheinlichkeitsdichte der Zufallsvariable X. (Es gilt ja auch $\int_{-\infty}^{\infty} f_1(x)\,\mathrm{d}x=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x=1$.) Der Beweis für die Zufallsvariable Y läuft analog. $\square$

Besonders wichtig ist der Fall von unabhängigen Zufallsvariablen X und Y. Hier lässt sich der gemeinsame Wahrscheinlichkeitsvektor durch die Wahrscheinlichkeitsvektoren der beiden Zufallsvariablen X und Y ausdrücken.

Definition: Sei X eine diskrete Zufallsvariable mit Wertebereich S und Y eine diskrete Zufallsvariable mit Wertebereich T. Die Zufallsvariablen X und Y heißen unabhängig, wenn $P(X=i\ \wedge\ Y=j)=P(X=i)P(Y=j)$ für alle $i\in S$ und $j\in T$ gilt. Das ist äquivalent zur Unabhängigkeit der Ereignisse $X=i$ und $Y=j$.

Bemerkung: Seien $(w_1(i))_{i\in S}$ und $(w_2(j))_{j\in T}$ die Wahrscheinlichkeitsvektoren der Zufallsvariablen X und Y. Dann sind X und Y genau dann unabhängig, wenn durch $u(i,j)=w_1(i)w_2(j)$ für $i\in S$ und $j\in T$ der gemeinsame Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y gegeben ist (siehe Beweis von Satz 34 (c)).

Eine analoge Definition geben wir für kontinuierliche Zufallsvariable.

Definition: Seien $f_1 : \mathbb{R} \to \mathbb{R}^+$ und $f_2 : \mathbb{R} \to \mathbb{R}^+$ Wahrscheinlichkeitsdichten der Zufallsvariablen X und Y. Sei $g : \mathbb{R}^2 \to \mathbb{R}^+$ definiert durch $g(x, y) = f_1(x)f_2(y)$ für alle $x \in \mathbb{R}$ und $y \in \mathbb{R}$. Die Zufallsvariablen X und Y heißen dann unabhängig, wenn g eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen X und Y ist.

Satz 28: *Die Unabhängigkeit der kontinuierlichen Zufallsvariablen X und Y ist äquivalent dazu, dass $P(X \le s \ \wedge \ Y \le t) = P(X \le s)P(Y \le t)$ für alle $s \in \mathbb{R}$ und alle $t \in \mathbb{R}$ gilt.*

Beweis: Seien $f_1 : \mathbb{R} \to \mathbb{R}^+$ und $f_2 : \mathbb{R} \to \mathbb{R}^+$ Wahrscheinlichkeitsdichten der Zufallsvariablen X und Y. Die Aussage, dass $P(X \le s \ \wedge \ Y \le t) = P(X \le s)P(Y \le t)$ für alle $s, t \in \mathbb{R}$ gilt, ist äquivalent zu $P(X \le s \ \wedge \ Y \le t) = \int_{-\infty}^{s} f_1(x)\,\mathrm{d}x \int_{-\infty}^{t} f_2(y)\,\mathrm{d}y = \int_{-\infty}^{s} f_1(x) \int_{-\infty}^{t} f_2(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{s}\int_{-\infty}^{t} f_1(x)f_2(y)\,\mathrm{d}y\,\mathrm{d}x$ für alle $s, t \in \mathbb{R}$ nach Definition der Wahrscheinlichkeitsdichte. Das aber ist äquivalent dazu, dass $g(x, y) = f_1(x)f_2(y)$ eine gemeinsame Wahrscheinlichkeitsdichte von X und Y ist, also äquivalent zur Unabhängigkeit von X und Y. (Es gilt ja auch $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x)f_2(y)\,\mathrm{d}y\,\mathrm{d}x = 1$.) □

2. Rechnen mit Zufallsvariablen

In diesem Kapitel soll das Rechnen mit Zufallsvariablen systematisch betrieben werden. Ansatzweise wurde das schon in den letzten Kapiteln getan, insbesondere in Satz 21, wo die Wahrscheinlichkeitsdichte einer linearen Funktion $\frac{X-a}{b}$ der Zufallsvariablen X berechnet wurde. Wir tun das jetzt auch für andere Funktionen. Solche Resultate spielen in der Statistik eine wesentliche Rolle.

Wir beginnen mit dem Quadrat einer Zufallsvariablen.

Satz 29: *Sei X eine Zufallsvariable mit Wahrscheinlichkeitsdichte f. Dann hat X^2 Wertebereich $\mathbb{R}^+$ und Wahrscheinlichkeitsdichte $g(x) = \frac{1}{2\sqrt{x}}(f(\sqrt{x}) + f(-\sqrt{x}))$ für $x \in \mathbb{R}^+$.*

Beweis: Sei F die Verteilungsfunktion der Zufallsvariable X, also eine Stammfunktion von f. Sei weiters G die Verteilungsfunktion von X^2, sodass $g = G'$ gilt.
Da X^2 nicht negativ sein kann, gilt $G(t) = P(X^2 \le t) = 0$ und $g(t) = G'(t) = 0$ für $t \le 0$.
Der Wertebereich von X^2 ist $\mathbb{R}^+$. Für $t > 0$ gilt

$$G(t) = P(X^2 \le t) = P(-\sqrt{t} \le X \le \sqrt{t}) = F(\sqrt{t}) - F(-\sqrt{t})$$

und durch Differenzieren nach t erhalten wir

$$g(t) = G'(t) = f(\sqrt{t})\frac{1}{2\sqrt{t}} + f(-\sqrt{t})\frac{1}{2\sqrt{t}} = \frac{1}{2\sqrt{t}}(f(\sqrt{t}) + f(-\sqrt{t}))$$

womit die behauptete Formel bereits gefunden ist. □

Dazu rechnen wir ein Beispiel.

Beispiel 55: Die Zufallsvariable X sei $N(0, 1)$-verteilt. Welche Verteilung hat X^2?

Die Wahrscheinlichkeitsdichte von X ist $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Nach Satz 29 hat X^2 den Wertebereich $\mathbb{R}^+$ und Wahrscheinlichkeitsdichte

$$g(x) = \frac{1}{2\sqrt{x}}(\varphi(\sqrt{x}) + \varphi(-\sqrt{x})) = \frac{1}{2\sqrt{x}}\frac{2}{\sqrt{2\pi}}e^{-\frac{x}{2}} = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\sqrt{\pi}}x^{\frac{1}{2}-1}e^{-\frac{1}{2}x}$$

Das ist die Wahrscheinlichkeitsdichte der $E(\frac{1}{2}, \frac{1}{2})$-Verteilung, die somit die Verteilung der Zufallsvariablen X^2 ist.

Wir beweisen Formeln für Summe und Quotient von zwei Zufallsvariablen.

Satz 30: *Haben die beiden Zufallsvariablen X und Y gemeinsame Wahrscheinlichkeitsdichte $g : \mathbb{R}^2 \to [0, \infty)$, dann hat die Zufallsvariable $X + Y$ Wahrscheinlichkeitsdichte $h(x) = \int_{-\infty}^{\infty} g(x-y, y)\,\mathrm{d}y$ für $x \in \mathbb{R}$.*

Beweis: Sei $t \in \mathbb{R}$ beliebig gewählt und sei $B = \{(x, y) \in \mathbb{R}^2 : x + y \leq t\}$. Mit Hilfe von Satz 26 folgt $P(X+Y \leq t) = P((X, Y) \in B) = \int_B g(x, y)\,\mathrm{d}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{t-y} g(x, y)\,\mathrm{d}x\,\mathrm{d}y$. Führt man im inneren Integral die neue Integrationsvarible $z = x + y$ ein, so erhält man $P(X + Y \leq t) = \int_{-\infty}^{\infty} \int_{-\infty}^{t} g(z-y, y)\,\mathrm{d}z\,\mathrm{d}y$. Wir vertauschen die Integrale und schreiben statt z wieder x. Dann haben wir $P(X+Y \leq t) = \int_{-\infty}^{t} \int_{-\infty}^{\infty} g(x-y, y)\,\mathrm{d}y\,\mathrm{d}x$. Die Funktion $h(x) = \int_{-\infty}^{\infty} g(x-y, y)\,\mathrm{d}y$, die im Integral $\int_{-\infty}^{t} \dots \mathrm{d}x$ steht, ist eine Wahrscheinlichkeitsdichte der Zufallsvariablen $X + Y$. $\square$

Satz 31: *Haben die beiden Zufallsvariablen X und Y gemeinsame Wahrscheinlichkeitsdichte $g : \mathbb{R}^2 \to [0, \infty)$, dann hat die Zufallsvariable $\frac{X}{Y}$ Wahrscheinlichkeitsdichte $h(x) = \int_{-\infty}^{\infty} g(xy, y)|y|\,\mathrm{d}y$ für $x \in \mathbb{R}$. (Nach Satz 27 hat Y eine Wahrscheinlichkeitsdichte. Daher gilt $P(Y = 0) = 0$ und man kann annehmen, dass $Y \neq 0$ gilt.)*

Beweis: Sei $t \in \mathbb{R}$ und sei $B = \{(x, y) \in \mathbb{R}^2 : \frac{x}{y} \leq t, y \neq 0\}$. Mit Satz 26 folgt $P(\frac{X}{Y} \leq t) = P((X, Y) \in B) = \int_B g(x, y)\,\mathrm{d}(x, y) = \int_{-\infty}^{0} \int_{ty}^{\infty} g(x, y)\,\mathrm{d}x\,\mathrm{d}y + \int_{0}^{\infty} \int_{-\infty}^{ty} g(x, y)\,\mathrm{d}x\,\mathrm{d}y$. Führt man in den beiden inneren Integralen die neue Integrationsvarible $z = \frac{x}{y}$ ein, so erhält man $P(\frac{X}{Y} \leq t) = \int_{-\infty}^{0} \int_{t}^{-\infty} g(zy, y)y\,\mathrm{d}z\,\mathrm{d}y + \int_{0}^{\infty} \int_{-\infty}^{t} g(zy, y)y\,\mathrm{d}z\,\mathrm{d}y$. Dabei ist $y < 0$ beim ersten Integral zu beachten. Da $\int_{t}^{-\infty} g(zy, y)y\,\mathrm{d}z = -\int_{-\infty}^{t} g(zy, y)y\,\mathrm{d}z = \int_{-\infty}^{t} g(zy, y)|y|\,\mathrm{d}z$ für $y < 0$ gilt, erhalten wir $P(\frac{X}{Y} \leq t) = \int_{-\infty}^{\infty} \int_{-\infty}^{t} g(zy, y)|y|\,\mathrm{d}z\,\mathrm{d}y$. Wir vertauschen die Integrale und schreiben statt z wieder x. Das ergibt $P(\frac{X}{Y} \leq t) = \int_{-\infty}^{t} \int_{-\infty}^{\infty} g(xy, y)|y|\,\mathrm{d}y\,\mathrm{d}x$. Die Funktion $h(x) = \int_{-\infty}^{\infty} g(xy, y)|y|\,\mathrm{d}y$, die im Integral $\int_{-\infty}^{t} \dots \mathrm{d}x$ steht, ist eine Wahrscheinlichkeitsdichte der Zufallsvariablen $\frac{X}{Y}$. $\square$

Bemerkung: Sind X und Y unabhängige Zufallsvariable mit Wahrscheinlichkeitsdichten f_1 und f_2, dann hat die Zufallsvariable $X + Y$ nach Satz 30 die Wahrscheinlichkeitsdichte $h(x) = \int_{-\infty}^{\infty} f_1(x-y) f_2(y)\,\mathrm{d}y$, da $g(x, y) = f_1(x) f_2(y)$ eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen X und Y ist. Man sagt, die Funktion h ist die Faltung der Funktionen f_1 und f_2.
Genauso erhält man, dass $h(x) = \int_{-\infty}^{\infty} f_1(xy) f_2(y)|y|\,\mathrm{d}y$ eine Wahrscheinlichkeitsdichte der Zufallsvariablen $\frac{X}{Y}$ ist.

Dazu rechnen wir zwei Beispiele.

Beispiel 56: Seien X und Y unabhängig und beide $E(\lambda)$-verteilt. Welche Dichte hat $\frac{X}{Y}$?

Sei $f(x) = \lambda e^{-\lambda x}$ für $x \geq 0$ und $f(x) = 0$ for $x < 0$. Das ist die Wahrscheinlichkeitsdichte von X und Y. Die Wahrscheinlichkeitsdichte von $\frac{X}{Y}$ ist dann $h(x) = \int_{-\infty}^{\infty} f(xy) f(y)|y|\,\mathrm{d}y$. Da $f(y) = 0$ für $y < 0$ gilt, erhält man $h(x) = \int_{0}^{\infty} f(xy) f(y) y\,\mathrm{d}y$. Da $f(xy) = 0$ für $x < 0$ und $y > 0$ gilt, ergibt sich $h(x) = 0$ für $x < 0$.

Es bleibt also $h(x)$ für $x \geq 0$ zu berechnen. In diesem Fall kann man einsetzen und erhält $h(x) = \int_{0}^{\infty} \lambda e^{-\lambda xy} \lambda e^{-\lambda y} y\,\mathrm{d}y = \lambda^2 \int_{0}^{\infty} e^{-\lambda(x+1)y} y\,\mathrm{d}y$. Mit Hilfe partieller Integration berechnet man $h(x) = \frac{1}{(1+x)^2}$.

Beispiel 57: Die Zufallsvariablen X und Y seien voneinander unabhängig. Weiters habe X die $N(\mu_1, \sigma_1)$-Verteilung und Y die $N(\mu_2, \sigma_2)$-Verteilung. Wir zeigen, dass $X+Y$ dann die $N(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2})$-Verteilung hat.

Sei $f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$. Die Wahrscheinlichkeitsdichte von X ist dann f_{μ_1,σ_1} und die von Y ist f_{μ_2,σ_2}. Setzt man in die Formel aus obiger Bemerkung ein, so erhält man

$$h(x) = \int_{-\infty}^{\infty} f_{\mu_1,\sigma_1}(x-y) f_{\mu_2,\sigma_2}(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-y-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}}\,\mathrm{d}y$$

als Wahrscheinlichkeitsdichte von $X+Y$. Wir formen den Exponenten um

$$\frac{(x-y-\mu_1)^2}{2\sigma_1^2} + \frac{(y-\mu_2)^2}{2\sigma_2^2} = \frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)} + \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}(y+C)^2$$

wobei $C = \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2 - x\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ gilt. Es ist zu beachten, dass C zwar von x, jedoch nicht von y abhängt und daher bei der weiter unten folgenden Integration nach y als Konstante zu behandeln ist. Setzt man diesen umgeformten Exponenten ein, so folgt

$$h(x) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}(y+C)^2}\,\mathrm{d}y$$

Führt man in diesem Integral die neue Variable $z = \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}(y+C)$ ein, so ergibt sich

$$\int_{-\infty}^{\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}(y+C)^2}\,\mathrm{d}y = \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,\mathrm{d}z = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\sqrt{2\pi}$$

wobei die letzte Gleichheit wegen $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z = 1$ gilt. Einsetzen ergibt

$$h(x) = \frac{1}{2\pi\sigma_1\sigma_2}\, e^{-\frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \frac{\sigma_1\sigma_2\sqrt{2\pi}}{\sqrt{\sigma_1^2+\sigma_2^2}} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\, e^{-\frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} = f_{\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2}}(x)$$

Das ist die Wahrscheinlichkeitsdichte der $N(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2})$-Verteilung.

Das Resultat aus Beispiel 57 kann man auf eine Summe von n unabhängigen normalverteilten Zufallsvariablen verallgemeinern (wichtig für die Statistik). Dazu müssen wir auch die Definitionen verallgemeinern.

Definition: Eine Funktion $g : \mathbb{R}^n \to \mathbb{R}^+$, die integrierbar ist (über entsprechend definierte Normalbereiche) und für die $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} g(x_1,x_2,\dots,x_n)\,\mathrm{d}x_n \dots \mathrm{d}x_2\,\mathrm{d}x_1 = 1$ gilt, heißt gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen $X_1, X_2, \dots, X_n$, wenn $P(X_1 \le t_1 \wedge X_2 \le t_2 \wedge \dots \wedge X_n \le t_n) = \int_{-\infty}^{t_1}\int_{-\infty}^{t_2}\dots\int_{-\infty}^{t_n} g(x_1,x_2,\dots,x_n)\,\mathrm{d}x_n \dots \mathrm{d}x_2\,\mathrm{d}x_1$ für alle $t_1, t_2, \dots, t_n \in \mathbb{R}$ erfüllt ist.

Definition: Seien $X_1, X_2, \dots, X_n$ Zufallsvariablen und $f_1, f_2, \dots, f_n$ deren Wahrscheinlichkeitsdichten. Man nennt die Zufallsvariablen $X_1, X_2, \dots, X_n$ unabhängig, wenn sie $g(x_1,\dots,x_n) = f_1(x_1)f_2(x_2)\dots f_n(x_n)$ als gemeinsame Wahrscheinlichkeitsdichte haben.

Es gilt dann folgender Satz

Satz 32: *Sind die Zufallsvariablen $X_1, X_2, \dots, X_n$ unabhängig und alle $N(\mu,\sigma)$-verteilt, dann hat $X_1+X_2+\dots+X_n$ die $N(n\mu, \sqrt{n}\sigma)$-Verteilung.*

Beweis: Für $k \geq 1$ sei $Y_k = X_1 + X_2 + \cdots + X_k$. Für $1 \leq k \leq n$ zeigen wir
(a) $Y_k, X_{k+1}, X_{k+2}, \ldots, X_n$ sind unabhängig
(b) Y_k ist $N(k\mu, \sqrt{k}\sigma)$-verteilt
Wegen $Y_1 = X_1$ gelten (a) und (b) für $k = 1$ nach Voraussetzung. Wir machen weiter mit Induktion. Wir bezeichnen die Wahrscheinlichkeitsdichte der $N(\mu, \sigma)$-Verteilung mit f und die Wahrscheinlichkeitsdichte der $N(k\mu, \sqrt{k}\sigma)$-Verteilung mit h_k. Wir nehmen an, dass $k < n$ gilt und (a) und (b) bereits bewiesen sind. Wir müssen dann (a) und (b) für $k + 1$ anstelle von k zeigen.

Nach Induktionsvoraussetzung ist $g(x_0, x_1, \ldots, x_{n-k}) = h_k(x_0)f(x_1)\ldots f(x_{n-k})$ eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen $Y_k, X_{k+1}, X_{k+2}, \ldots, X_n$. Weiters gilt $Y_{k+1} = Y_k + X_{k+1}$. Wir gehen ähnlich vor wie im Beweis von Satz 30 und setzen $B = \{(x_0, x_1, \ldots, x_{n-k}) \in \mathbb{R}^{n-k+1} : x_0 + x_1 \leq t_1, x_2 \leq t_2, \ldots, x_{n-k} \leq t_{n-k}\}$. Wir verwenden eine Verallgemeinerung von Satz 26 und berechnen damit

$$
\begin{aligned}
P(Y_{k+1} \leq t_1 \wedge X_{k+2} \leq t_2 \wedge \cdots \wedge X_n \leq t_{n-k}) &= P((Y_k, X_{k+1}, X_{k+2}, \ldots, X_n) \in B) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t_1 - x_0} \int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} h_k(x_0)f(x_1)f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x_0 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t_1 - x_0} h_k(x_0)f(x_1) \int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x_0 \\
&= \int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2 \int_{-\infty}^{\infty} \int_{-\infty}^{t_1 - x_0} h_k(x_0)f(x_1)\,\mathrm{d}x_1\,\mathrm{d}x_0
\end{aligned}
$$

wobei zuerst $h_k(x_0)f(x_1)$ aus den Integralen über die Variablen $x_{n-k}, \ldots, x_3, x_2$ herausgehoben wurde und anschließend $\int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2$ aus den Integralen über die Variablen x_1 und x_0. Nun gilt

$$
\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{t_1 - x_0} h_k(x_0)f(x_1)\,\mathrm{d}x_1\,\mathrm{d}x_0 &= \int_{-\infty}^{\infty} \int_{-\infty}^{t_1} h_k(x_0)f(z - x_0)\,\mathrm{d}z\,\mathrm{d}x_0 \\
&= \int_{-\infty}^{t_1} \int_{-\infty}^{\infty} h_k(x_0)f(z - x_0)\,\mathrm{d}x_0\,\mathrm{d}z
\end{aligned}
$$

wobei zuerst statt x_1 die neue Integrationsvariable $z = x_0 + x_1$ eingeführt und dann die Integrale vertauscht wurden. Weiters gilt $\int_{-\infty}^{\infty} h_k(x_0)f(z - x_0)\,\mathrm{d}x_0 = h_{k+1}(z)$, wie die Rechnung in Beispiel 57 zeigt. Setzt man ein und schreibt statt z wieder x_1, so ergibt sich

$$
\begin{aligned}
&P(Y_{k+1} \leq t_1 \wedge X_{k+2} \leq t_2 \wedge \cdots \wedge X_n \leq t_{n-k}) \\
&\quad = \int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2 \int_{-\infty}^{t_1} h_{k+1}(x_1)\,\mathrm{d}x_1 \\
&\quad = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \ldots \int_{-\infty}^{t_{n-k}} h_{k+1}(x_1)f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2\,\mathrm{d}x_1
\end{aligned}
$$

wobei wir anschließend das oben durchgeführte Herausheben wieder rückgängig gemacht haben. Wir sehen, dass $\tilde{g}(x_1, \ldots, x_{n-k}) = h_{k+1}(x_1)f(x_2)\ldots f(x_{n-k})$ eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen $Y_{k+1}, X_{k+2}, \ldots, X_n$ ist.

Um (b) zu zeigen, sei $C = \{(x_0, x_1, \ldots, x_{n-k}) \in \mathbb{R}^{n-k+1} : x_0 + x_1 \leq t\}$. Dann haben wir $P(Y_{k+1} \leq t) = P((Y_k, X_{k+1}, \ldots, X_n) \in C)$. Es gilt $\int_{-\infty}^{\infty} cf(x)\,\mathrm{d}x = c$ für alle $c \in \mathbb{R}$. Wir erhalten dann $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_2)\ldots f(x_{n-k})\,\mathrm{d}x_{n-k}\ldots \mathrm{d}x_2 = 1$ durch wiederholtes Anwenden dieses Resultats. Obige Rechnung mit $t_1 = t$ und $t_2 = t_3 = \cdots = t_{n-k} = \infty$ zeigt dann, dass $P(Y_{k+1} \leq t) = \int_{-\infty}^{t} h_{k+1}(x_1)\,\mathrm{d}x_1$ gilt. Somit ist h_{k+1} eine Wahrscheinlichkeitsdichte der Zufallsvariablen Y_{k+1}, das heißt (b) ist für $k + 1$ gezeigt. Da aber $\tilde{g}(x_1, \ldots, x_{n-k}) = h_{k+1}(x_1)f(x_2)\ldots f(x_{n-k})$ eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen $Y_{k+1}, X_{k+2}, \ldots, X_n$ ist, sind diese Zufallsvariablen unabhängig und somit ist auch (a) für $k + 1$ gezeigt.

Die Induktion ist fertig. Wir haben (a) und (b) für $1 \leq k \leq n$ gezeigt. Insbesondere gilt (b) für $k = n$. Damit ist der Satz bewiesen. □

3. Erwartungswert

In diesem Kapitel leiten wir einige Rechenregeln für der Erwartungswert her. Wir geben die Beweise nur für diskrete Zufallsvariable. Die Formeln gelten auch für kontinuierliche Zufallsvariable, aber die Beweise sind im kontinuierlichen Fall schwieriger. Die Grundlage für das Rechnen mit Erwartungswerten bildet der folgende Satz.

Satz 33: *(Erwartungswert einer Funktion von Zufallsvariablen)*
(a) *Es sei X eine diskrete Zufallsvariable mit Wertebereich S und Wahrscheinlichkeitsvektor $(w(i))_{i\in S}$. Weiters sei $\gamma : S \to \mathbb{R}$ eine Abbildung und $Z = \gamma(X)$ (zum Beispiel: wenn $\gamma(x) = x^2$, dann $Z = X^2$). Wenn $\sum_{i\in S} |\gamma(i)|w(i) < \infty$ gilt, dann existiert $\mathrm{E}(Z)$ und es gilt $\mathrm{E}(Z) = \sum_{i\in S} \gamma(i)w(i)$.*
(b) *Seien X und Y diskrete Zufallsvariable mit gemeinsamen Wahrscheinlichkeitsvektor $(u(i,j))_{i\in S, j\in T}$. Weiters sei $\gamma : S\times T \to \mathbb{R}$ eine Abbildung und $Z = \gamma(X,Y)$ (zum Beispiel: wenn $\gamma(x,y) = x+y$, dann $Z = X+Y$). Wenn $\sum_{(i,j)\in S\times T} |\gamma(i,j)|u(i,j) < \infty$ gilt, dann existiert $\mathrm{E}(Z)$ und es gilt $\mathrm{E}(Z) = \sum_{(i,j)\in S\times T} \gamma(i,j)u(i,j)$.*

Beweis: Wir beweisen (a). Wir berechnen zuerst den Wahrscheinlichkeitsvektor der Zufallsvariable Z. Die Menge $\gamma(S)$ bezeichnen wir mit G. Da die Zufallsvariable X Werte in S annimmt, nimmt die Zufallsvariable Z Werte in G an. Wir setzen $H_k = \gamma^{-1}(k)$ und $v(k) = \sum_{i\in H_k} w(i)$ für $k \in G$. Die Mengen H_k für $k \in G$ bilden dann eine Zerlegung der Menge S. Für alle $k \in G$ erhalten wir mit Hilfe von Satz 13

$$P(Z = k) = P(\gamma(X) = k) = P(X \in \gamma^{-1}(k)) = P(X \in H_k) = \textstyle\sum_{i\in H_k} w(i) = v(k)$$

Es gilt $\sum_{i\in S} w(i) = 1$. Aus Satz 56 mit $M = S$ folgt $\sum_{k\in G} v(k) = \sum_{i\in S} w(i) = 1$. Also ist $(v(k))_{k\in G}$ der Wahrscheinlichkeitsvektor der Zufallsvariable Z.
Wir wenden Satz 56 mit $M = S$ und mit $r_i = \gamma(i)w(i)$ für $i \in M = S$ an. Es wird ja $\sum_{i\in S} |r_i| = \sum_{i\in S} |\gamma(i)|w(i) < \infty$ vorausgesetzt. Wir berechen s_k für $k \in G$. Da $\gamma(i) = k$ für alle $i \in H_k$ gilt, erhalten wir $s_k = \sum_{i\in H_k} \gamma(i)w(i) = k\sum_{i\in H_k} w(i) = kv(k)$. Aus Satz 56 folgt dann sowohl $\sum_{k\in G} |k|v(k) < \infty$, das heißt $\mathrm{E}(Z)$ existiert, als auch $\sum_{k\in G} kv(k) = \sum_{i\in S} \gamma(i)w(i)$, das heißt $\mathrm{E}(Z) = \sum_{i\in S} \gamma(i)w(i)$. Damit ist (a) gezeigt.
Der Beweis von (b) ist derselbe. Man ersetzt S durch $S\times T$, X durch (X,Y), i durch (i,j) und $w(i)$ durch $u(i,j)$. □

Jetzt können wir die Rechenregeln für den Erwartungswert herleiten.

Satz 34: *Seien X und Y Zufallsvariable, deren Erwartungswerte existieren. Weiters seien a und b reelle Zahlen. Dann gilt*
(a) $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$
(b) $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
(c) $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$, *wenn X und Y unabhängig sind.*

Beweis: Wir führen den Beweis nur für den Fall diskreter Zufallsvariable durch. Sei S der Wertebereich von X und T der von Y. Für $n \ge 1$ sei S_n eine endliche Teilmenge von S, sodass $S_1 \subseteq S_2 \subseteq S_3 \subseteq \dots$ und $\bigcup_{n=1}^{\infty} S_n = S$ gilt. Ebenso für $n \ge 1$ sei T_n eine endliche Teilmenge von T, sodass $T_1 \subseteq T_2 \subseteq T_3 \subseteq \dots$ und $\bigcup_{n=1}^{\infty} T_n = T$ gilt.
(a) Wir wenden Satz 33 (a) mit $\gamma(x) = ax + b$ an. Sei $(w(i))_{i\in S}$ der Wahrscheinlichkeitsvektor der Zufallsvariablen X. Da die Existenz von $\mathrm{E}(X)$ vorausgesetzt wird, haben wir $\sum_{i\in S} |i|w(i) =: C < \infty$. Wegen $\sum_{i\in S_n} w(i) \le \sum_{i\in S} w(i) = 1$ (Satz 53) gilt für $n \ge 1$

$$\textstyle\sum_{i\in S_n} |ai + b|w(i) \le |a|\sum_{i\in S_n} |i|w(i) + |b|\sum_{i\in S_n} w(i) \le |a|\sum_{i\in S} |i|w(i) + |b| = |a|C + |b|.$$

Lassen wir n gegen ∞ gehen, dann erhalten wir $\sum_{i\in S}|ai+b|w(i)\le |a|C+|b|$ mit Satz 53. Nach Satz 33 (a) existiert $\mathrm{E}(aX+b)$ und ist gleich $\sum_{i\in S}(ai+b)w(i)$. Aus der Gleichung $\sum_{i\in S_n}(ai+b)w(i)=a\sum_{i\in S_n}iw(i)+b\sum_{i\in S_n}w(i)$ folgt jetzt mit $n\to\infty$

$$\mathrm{E}(aX+b)=\sum_{i\in S}(ai+b)w(i)=a\sum_{i\in S}iw(i)+b\sum_{i\in S}w(i)=a\mathrm{E}(X)+b$$

da alle vorkommenden Reihen absolut konvergieren, wofür wir Satz 54 verwenden.

(b) Sei $(u(i,j))_{i\in S,j\in T}$ der gemeinsame Wahrscheinlichkeitsvektor von X und Y. Weiters sei $(w_1(i))_{i\in S}$ der Wahrscheinlichkeitsvektor von X und $(w_2(j))_{j\in T}$ der von Y. Nach Satz 25 gilt $\sum_{j\in T}u(i,j)=w_1(i)$ für $i\in S$ und $\sum_{j\in S}u(i,j)=w_2(j)$ für $j\in T$. Da die Existenz von $\mathrm{E}(X)$ und $\mathrm{E}(Y)$ vorausgesetzt wird, haben wir $\sum_{i\in S}|i|w_1(i)=:C<\infty$ und $\sum_{j\in T}|j|w_2(j)=:D<\infty$. Wegen $\sum_{j\in T_n}u(i,j)\le\sum_{j\in T}u(i,j)=w_1(i)$ gilt für $n\ge 1$

$$\sum_{(i,j)\in S_n\times T_n}|i|u(i,j)=\sum_{i\in S_n}|i|\sum_{j\in T_n}u(i,j)\le\sum_{i\in S_n}|i|w_1(i)\le\sum_{i\in S}|i|w_1(i)=C.$$

Für $n\to\infty$ erhalten wir $\sum_{(i,j)\in S\times T}|i|u(i,j)\le C$ mit Hilfe von Satz 53. Aus Satz 33 (b) mit $\gamma(x,y)=x$ folgt dann $\mathrm{E}(X)=\sum_{(i,j)\in S\times T}iu(i,j)$. Ebenso zeigt man

$$\sum_{(i,j)\in S_n\times T_n}|j|u(i,j)\le\sum_{j\in T}|j|w_2(j)=D$$

für alle $n\ge 1$ und damit $\sum_{(i,j)\in S\times T}|j|u(i,j)\le D$. Aus Satz 33 (b) mit $\gamma(x,y)=y$ folgt dann $\mathrm{E}(Y)=\sum_{(i,j)\in S\times T}ju(i,j)$. Wieder für alle $n\ge 1$ gilt jetzt

$$\sum_{(i,j)\in S_n\times T_n}|i+j|u(i,j)\le\sum_{(i,j)\in S_n\times T_n}|i|u(i,j)+\sum_{(i,j)\in S_n\times T_n}|j|u(i,j)\le C+D.$$

Für $n\to\infty$ erhalten wir $\sum_{(i,j)\in S\times T}|i+j|u(i,j)\le C+D$ mit Hilfe von Satz 53. Aus Satz 33 (b) mit $\gamma(x,y)=x+y$ folgt dann, dass der Erwartungswert $\mathrm{E}(X+Y)$ existiert und dass $\mathrm{E}(X+Y)=\sum_{(i,j)\in S\times T}(i+j)u(i,j)$ gilt. Für alle $n\ge 1$ gilt auch die Gleichung

$$\sum_{(i,j)\in S_n\times T_n}(i+j)u(i,j)=\sum_{(i,j)\in S_n\times T_n}iu(i,j)+\sum_{(i,j)\in S_n\times T_n}ju(i,j).$$

Für $n\to\infty$ konvergiert die linke Seite dieser Gleichung gegen $\sum_{(i,j)\in S\times T}(i+j)u(i,j)$ und die rechte Seite gegen $\sum_{(i,j)\in S\times T}iu(i,j)+\sum_{(i,j)\in S\times T}ju(i,j)$, wozu wir Satz 54 verwenden. Damit ist $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$ gezeigt.

(c) Sei $(w_1(i))_{i\in S}$ der Wahrscheinlichkeitsvektor von X und $(w_2(j))_{j\in T}$ der von Y. Sei $u(i,j)=w_1(i)w_2(j)$ für $i\in S$ und $j\in T$. Da die Zufallsvariablen X und Y unabhängig sind, ist $(u(i,j))_{i\in S,j\in T}$ der gemeinsame Wahrscheinlichkeitsvektor von X und Y (es gilt ja $P(X=i\ \wedge\ Y=j)=P(X=i)P(Y=j)=w_1(i)w_2(j)=u(i,j)$ für alle $i\in S$ und alle $j\in T$ und es folgt auch $\sum_{(i,j)\in S\times T}u(i,j)=\lim_{n\to\infty}\sum_{(i,j)\in S_n\times T_n}w_1(i)w_2(j)=\lim_{n\to\infty}\sum_{i\in S_n}w_1(i)\sum_{j\in T_n}w_2(j)=\sum_{i\in S}w_1(i)\sum_{j\in T}w_2(j)=1$ mit Satz 53).

Weiters haben wir auch $\sum_{i\in S}|i|w_1(i)=:C<\infty$ und $\sum_{j\in T}|j|w_2(j)=:D<\infty$, da die Existenz von $\mathrm{E}(X)$ und $\mathrm{E}(Y)$ vorausgesetzt wird. Für $n\ge 1$ berechnen wir

$$\begin{aligned}\sum_{(i,j)\in S_n\times T_n}|ij|u(i,j)&=\sum_{(i,j)\in S_n\times T_n}|ij|w_1(i)w_2(j)=\sum_{i\in S_n}|i|w_1(i)\sum_{j\in T_n}|j|w_2(j)\\&\le\sum_{i\in S}|i|w_1(i)\sum_{j\in T}|j|w_2(j)=CD.\end{aligned}$$

Für $n\to\infty$ erhalten wir $\sum_{(i,j)\in S\times T}|ij|u(i,j)\le CD$ wegen Satz 53. Aus Satz 33 (b) mit $\gamma(x,y)=xy$ folgt dann, dass $\mathrm{E}(XY)$ existiert und dass $\mathrm{E}(XY)=\sum_{(i,j)\in S\times T}iju(i,j)$ gilt. Eine analoge Rechnung wie oben ergibt für $n\ge 1$

$$\sum_{(i,j)\in S_n\times T_n}iju(i,j)=\sum_{i\in S_n}iw_1(i)\sum_{j\in T_n}jw_2(j).$$

Für $n\to\infty$ konvergiert die linke Seite dieser Gleichung gegen $\sum_{(i,j)\in S\times T}iju(i,j)$ und die rechte Seite gegen $\sum_{i\in S}iw_1(i)\sum_{j\in T}jw_2(j)$, wozu wir Satz 54 verwenden. Damit ist dann $\mathrm{E}(XY)=\mathrm{E}(X)\mathrm{E}(Y)$ gezeigt. $\square$

V. Statistik

In der Statistik geht es darum, aus gesammelten Daten Schlussfolgerungen zu ziehen. Um eine zu messende Größe möglichst genau zu bestimmen, führt man die Messung mehrmals durch. Um die durchschnittliche Füllmenge von Milchflaschen zu ermitteln, zieht man eine zufällige Stichprobe von n Flaschen und misst deren Inhalte. Um den Anteil der Wähler der Partei A bei der bevorstehenden Wahl vorherzusagen, befragt man n zufällig ausgewählte Wahlberechtigte. Um zu testen, ob ein Würfel fair ist, wirft man den Würfel n Mal und versucht aus den geworfenen Augenzahlen Schlüsse zu ziehen.

Wir werden zwei statistische Verfahren kennenlernen, nämlich Konfidenzintervalle und statistische Tests. Ein Konfidenzintervall ist ein Intervall, das aus den erhobenen Daten berechnet wird und die gesuchte Größe (die durchschnittliche Füllmenge, den Anteil der Wähler der Partei A) mit einer vorgegebenen Wahrscheinlichkeit enthält. Bei einem statistischen Test geht es darum, eine im voraus aufgestellte Aussage (Hypothese) durch die erhobenen Daten zu entkräften oder zu bestätigen.

1. Parameterschätzer

Bevor wir auf die eigentlichen statistische Verfahren eingehen, beschäftigen wir uns mit dem Schätzen von Parametern aus Stichproben. Diese Stichproben werden aus einer zuvor festgelegten Grundgesamtheit gezogen. Beispiele für Grundgesamtheiten sind die von einer Firma abgefüllten Waschmittelpackungen oder die erwachsenen Einwohner Österreichs. Aus dieser Grundgesamtheit wird eine zufällige Stichprobe vom Umfang n gezogen und ausgewertet. Wir bezeichnen die gefundenen Werte mit $x_1, x_2, \dots, x_n$. Das können zum Beispiel die Füllmengen von n zufällig gewählten Waschmittelpackungen sein oder die Körpergrößen von n zufällig gewählten Personen.

Aus dieser Stichprobe soll der Mittelwert in der Grundgesamtheit geschätzt werden. In den Beispielen sind das die durchschnittliche Füllmenge der Waschmittelpackungen und die durchschnittliche Körpergröße der erwachsenen Österreicher. Als Schätzer für diesen Mittelwert nehmen wir den Wert m, dessen durchschnittliche quadratische Abweichung $\frac{1}{n}\sum_{j=1}^n (x_j - m)^2$ von den Stichprobenwerten $x_1, x_2, \dots, x_n$ minimal wird. Es gilt

$$\tfrac{1}{n}\textstyle\sum_{j=1}^n (x_j - m)^2 = \tfrac{1}{n}\sum_{j=1}^n x_j^2 - 2m\tfrac{1}{n}\sum_{j=1}^n x_j + m^2$$

Das ist eine quadratische Funktion in m, die ihr Minimum für $m = \frac{1}{n}\sum_{j=1}^n x_j$ annimmt. Üblicherweise nennt man $m = \frac{1}{n}\sum_{j=1}^n x_j$ das Stichprobenmittel. Es ist ein Schätzer für den Mittelwert in der Grundgesamtheit. Die durchschnittliche qudratische Abweichung $\frac{1}{n}\sum_{j=1}^n (x_j - m)^2$ der Stichprobenwerte vom Stichprobenmittel nennt man die Stichprobenvarianz. Wir bezeichnen sie mit v. Sie schätzt die Größe der Schwankungen um den Mittelwert. Genauere Aussagen über den Mittelwert in der Grundgesamtheit gewinnt man mit Hilfe von Konfidenzintervallen, die später behandelt werden.

Wir untersuchen zwei zusammenhängende Größen, zum Beispiel Körpergröße und Körpergewicht. Aus den erwachsenen Einwohnern Österreichs wird eine zufällige Stichprobe vom Umfang n gezogen. Die j-te Person habe Körpergröße x_j und Körpergewicht y_j. Das ergibt zwei miteinander verbundene Stichproben $x_1, x_2, \dots, x_n$ und $y_1, y_2, \dots, y_n$. Seien m_x und m_y die Stichprobenmittel und v_x und v_y die Stichprobenvarianzen. Analog zur Stichprobenvarianz definieren wir die Stichprobencovarianz dieser beiden Stichproben durch $c_{xy} = \frac{1}{n}\sum_{j=1}^n (x_j - m_x)(y_j - m_y)$. Es gilt dann $v_x = c_{xx}$ und $v_y = c_{yy}$. Wir nehmen an, dass die Körpergrößen in der Stichprobe nicht alle gleich groß sind. Dasselbe nehmen wir für die Körpergewichte an. Dann gilt $v_x > 0$ und $v_y > 0$.

Wir stellen die Frage, ob man eine lineare Beziehung $y = ax + b$ zwischen Körpergröße und Körpergewicht finden kann. Wir schätzen die Parameter a und b aus den Stichproben, sodass die Qudratsumme der Abstände der beobachteten Körpergewichte von den durch die Formel $y = ax+b$ vorhergesagten Körpergewichten minimal wird. Wir berechnen zuerst

$$c_{xy} = \frac{1}{n}\sum_{j=1}^n (x_j - m_x)(y_j - m_y) = \frac{1}{n}\sum_{j=1}^n x_j y_j - \frac{1}{n}\sum_{j=1}^n x_j m_y - \frac{1}{n}\sum_{j=1}^n m_x y_j + m_x m_y$$
$$= \frac{1}{n}\sum_{j=1}^n x_j y_j - m_x m_y - m_x m_y + m_x m_y = \frac{1}{n}\sum_{j=1}^n x_j y_j - m_x m_y$$

Es folgt $\frac{1}{n}\sum_{j=1}^n x_j y_j = m_x m_y + c_{xy}$. Als Spezialfälle erhalten wir $\frac{1}{n}\sum_{j=1}^n x_j^2 = m_x^2 + v_x$ und $\frac{1}{n}\sum_{j=1}^n y_j^2 = m_y^2 + v_y$. Mit Hilfe dieser Gleichungen ergibt sich

$$\frac{1}{n}\sum_{j=1}^n (y_j - ax_j - b)^2 = \frac{1}{n}\sum_{j=1}^n (y_j^2 + a^2x_j^2 + b^2 - 2ax_jy_j - 2by_j + 2abx_j)$$
$$= m_y^2 + v_y + a^2m_x^2 + a^2v_x + b^2 - 2am_xm_y - 2ac_{xy} - 2bm_y + 2abm_x$$
$$= (m_y - am_x - b)^2 + (a\sqrt{v_x} - \tfrac{c_{xy}}{\sqrt{v_x}})^2 + v_y - \tfrac{c_{xy}^2}{v_x}$$

Daraus erkennt man, dass $\frac{1}{n}\sum_{j=1}^n (y_j - ax_j - b)^2$ genau dann minimal wird, wenn $a = \frac{c_{xy}}{v_x}$ und $b = m_y - am_x$ gilt. Somit ist $y = ax + b = \frac{c_{xy}}{v_x}(x - m_x) + m_y$ die am besten angepasste Gerade. Sie geht durch den Punkt (m_x, m_y) und hat Anstieg $\frac{c_{xy}}{v_x}$.

Wir definieren den Korrelationskoeffizienten der beiden Stichproben durch $k_{xy} = \frac{c_{xy}}{\sqrt{v_xv_y}}$. Dieser hat folgende Eigenschaften

Satz 35: *Für den Korrelationskoeffizienten k_{xy} gilt $-1 \le k_{xy} \le 1$. Ist $k_{xy} = 1$, dann liegen die Punkte (x_j, y_j) für $1 \le j \le n$ auf einer Geraden mit Anstieg > 0. Ist $k_{xy} = -1$, dann liegen diese Punkte auf einer Geraden mit Anstieg < 0.*

Beweis: Nach obiger Rechnung ist $v_y - \frac{c_{xy}^2}{v_x} = v_y(1 - k_{xy}^2)$ das Minimum der Quadratsumme $\frac{1}{n}\sum_{j=1}^n (y_j - ax_j - b)^2$. Es muss ≥ 0 sein. Es folgt $k_{xy}^2 \le 1$, das heißt $-1 \le k_{xy} \le 1$. Seien jetzt $a = \frac{c_{xy}}{v_x}$ und $b = m_y - am_x$ die Werte, für die dieses Minimum angenommen wird. Gilt $k_{xy}^2 = 1$, dann gilt auch $\frac{1}{n}\sum_{j=1}^n (y_j - ax_j - b)^2 = 0$, das heißt die Punkte (x_j, y_j) für $1 \le j \le n$ liegen alle auf der Geraden $y = ax + b$. Ist $k_{xy} = 1$, dann gilt $c_{xy} > 0$ und somit $a > 0$. Ist $k_{xy} = -1$, dann gilt $c_{xy} < 0$ und somit $a < 0$. □

Der Korrelationskoeffizienten k_{xy} ist eine Maßzahl dafür, wie sehr die Punktemenge $\{(x_j, y_j) : 1 \le j \le n\}$ von der am besten angepassten Gerade abweicht. Je näher k_{xy} bei 1 oder -1 liegt, umso besser lässt sich diese Punktemenge durch eine Gerade approximieren. Je näher k_{xy} bei 0 liegt, umso weniger ist das möglich.

Bemerkung: Wir haben hier bereits gezogene Stichproben behandelt. In diesem Fall liegen n Werte $x_1, x_2, \dots, x_n$ vor. Allerdings sind diese Werte Ergebnisse eines Zufallsexperiments, nämlich des zufälligen Ziehens einer Stichprobe. Daher werden wir sie im Folgenden auch als Zufallsvariable $X_1, X_2, \dots, X_n$ schreiben. Dann ist auch das Stichprobenmittel $M = \frac{1}{n}\sum_{j=1}^n X_j$ eine Zufallsvariable, und ebenso die Stichprobenvarianz $S^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - M)^2$. Die Stichprobenvarianz, die wir jetzt mit S^2 bezeichnen, wird außerdem ein wenig modifiziert. Statt durch n wird durch $n-1$ dividiert. Man tut das deshalb, da man dann zeigen kann, dass $\mathrm{E}(S^2)$ gleich der Varianz in der Grundgesamtheit ist. Man sagt, S^2 ist ein erwartungstreuer Schätzer.

2. Konfidenzintervalle für normalverteilte Messwerte

Misst man eine (physikalische) Größe mehrmals, so wird man Messwerte erhalten, die wegen zufälliger Störungen beim Messvorgang leicht voneinander abweichen. Diese zufälligen Schwankungen nimmt man üblicherweise als normalverteilt an, sodass die Messergebnisse $N(\mu,\sigma)$-verteilte Zufallsvariable sind, wobei μ die zu messende unbekannte Größe ist und σ die Genauigkeit des Messgerätes angibt.

Man führt den Messvorgang n Mal durch. Die dabei gemessenen Werte $X_1, X_2, \dots, X_n$ bilden die Stichprobe. Jede dieser n Zufallsvariablen hat nach den obigen Annahmen die $N(\mu,\sigma)$-Verteilung. Da die Messungen unabhängig voneinander durchgeführt werden, sind auch die Zufallsvariablen $X_1, X_2, \dots, X_n$ unabhängig.

Um die zu messende unbekannte Größe μ ausreichend genau zu bestimmen, gibt man ein $\gamma \in (0,1)$ vor (übliche Werte für γ sind 0.95 oder 0.99) und berechnet aus den gemessenen Werten $X_1, X_2, \dots, X_n$ ein Intervall I, das μ mit Wahrscheinlichkeit γ enthält. So ein Intervall I nennt man γ-Konfidenzintervall für die Größe μ. Für γ gibt es verschiedene Namen, wie zum Beispiel Sicherheitswahrscheinlichkeit oder Konfidenzniveau. Das Konfidenzintervall hängt von den gemessenen Werten $X_1, X_2, \dots, X_n$ ab, also von zufälligen Einflüssen. Führt man so eine Serie von n Messungen mehrere Male durch und berechnet aus jeder das γ-Konfidenzintervall, dann wird man verschiedene Intervalle erhalten. Auch wird μ nicht immer im Konfidenzintervall enthalten sein. Ist zum Beispiel $\gamma = 0.95$, dann wird μ mit einer Wahrscheinlichkeit von 95 % im Konfidenzintervall liegen, das heißt bei wiederholter Bestimmung des Konfidenzintervalls wird sich μ im Durchschnitt jedes zwanzigste Mal außerhalb des Intervalls befinden.

Um Konfidenzintervalle zu berechnen, benötigen wir folgende Definition.

Definition: Für $\beta \in (0,1)$ sei z_β die eindeutige Lösung der Gleichung $\Phi(z_\beta) = 1 - \beta$. Diese Lösung ist eindeutig, da $\Phi : \mathbb{R} \to (0,1)$ bijektiv ist (Satz 23 (d)).

Die Werte von z_β findet man in einer Tabelle. Das im folgenden Satz bestimmte Intervall heißt zweiseitiges γ-Konfidenzintervall. Wir nehmen an, dass wir die Genauigkeit des Messgerätes, also σ, kennen.

Satz 36: *Sei $0 < \gamma < 1$ gegeben. Seien $X_1, X_2, \dots, X_n$ unabhängige $N(\mu,\sigma)$-verteilte Messwerte, wobei σ bekannt ist. Sei $M = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ der Mittelwert der Messwerte (das sogenannte Stichprobenmittel) und $\beta = \frac{1-\gamma}{2}$. Für $I = [M - \frac{\sigma z_\beta}{\sqrt{n}}, M + \frac{\sigma z_\beta}{\sqrt{n}}]$ gilt dann $P(\mu \in I) = \gamma$, das heißt I ist ein γ-Konfidenzintervall für μ.*

Beweis: Aus Satz 32 folgt, dass $X_1 + X_2 + \cdots + X_n$ die $N(n\mu, \sqrt{n}\sigma)$-Verteilung hat. Wegen Satz 22 hat dann $\sqrt{n}\frac{M-\mu}{\sigma} = \frac{X_1+X_2+\cdots+X_n-n\mu}{\sqrt{n}\sigma}$ die $N(0,1)$-Verteilung. Daher gilt

$$P(-z_\beta \le \sqrt{n}\frac{M-\mu}{\sigma} \le z_\beta) = \Phi(z_\beta) - \Phi(-z_\beta) = \Phi(z_\beta) - (1 - \Phi(z_\beta)) = 1 - \beta - \beta = \gamma$$

Wir formen dieses Ereignis um

$$-z_\beta \le \sqrt{n}\frac{M-\mu}{\sigma} \le z_\beta \;\Leftrightarrow\; -\frac{\sigma z_\beta}{\sqrt{n}} \le M - \mu \le \frac{\sigma z_\beta}{\sqrt{n}} \;\Leftrightarrow\; M - \frac{\sigma z_\beta}{\sqrt{n}} \le \mu \le M + \frac{\sigma z_\beta}{\sqrt{n}} \;\Leftrightarrow\; \mu \in I$$

Setzt man dieses umgeformte Ereignis oben ein, so erhält man $P(\mu \in I) = \gamma$, das gewünschte Resultat. □

Jetzt ist es leicht Beispiele zu rechnen. Man braucht ja nur in die Formel, die wir in Satz 36 gefunden haben, einzusetzen.

Beispiel 58: Eine unbekannte Größe μ wird 25 Mal gemessen. Die Genauigkeit des Messvorgangs wird mit $\sigma = 5$ cm angegeben. Als Mittelwert der Messwerte ergibt sich 3.24 m. Gesucht ist ein 95%-Konfidenzintervall.

Wir setzen in die Formel ein. Wegen $\gamma = 0.95$ folgt $\beta = \frac{1-\gamma}{2} = 0.025$. Man findet $z_\beta = 1.96$ in einer Tabelle. Jetzt erhalten wir $\frac{\sigma z_\beta}{\sqrt{n}} = \frac{0.05 \cdot 1.96}{\sqrt{25}} = 0.02$ und daraus ergibt sich $I = [3.24 - 0.02, 3.24 + 0.02] = [3.22, 3.26]$. Mit einer Wahrscheinlichkeit von 0.95 ist μ in diesem Intervall enthalten.

Beispiel 59: Die Füllgewichte von Waschmittelpackungen einer bestimmten Sorte seien $N(\mu, \sigma)$-verteilt. Die Genauigkeit der Abfüllmaschine wird mit 0.02 kg angegeben. Man wiegt 10 zufällig gewählte Packungen und erhält die Inhalte 1.05, 0.99, 1.03, 1.03, 1.01, 1.02, 1.01, 0.97, 1.01, 0.98 in kg. Gesucht ist ein 99%-Konfidenzintervall für den durchschnittlichen Packungsinhalt μ.

Wegen $\gamma = 0.99$ folgt $\beta = \frac{1-\gamma}{2} = 0.005$. Man findet $z_\beta = 2.58$ in einer Tabelle. Weiters berechnen wir $M = \frac{1}{10}(1.05 + 0.99 + \cdots + 0.98) = \frac{10.1}{10} = 1.01$ und $\frac{\sigma z_\beta}{\sqrt{n}} = \frac{0.02 \cdot 2.58}{\sqrt{10}} = 0.019$. Daraus ergibt sich dann $I = [1.01 - 0.019, 1.01 + 0.019] = [0.992, 1.029]$. Mit einer Wahrscheinlichkeit von 0.99 ist μ in diesem Intervall enthalten.

Bevor man ein Konfidenzintervall ermittelt, muss man zwei Entscheidungen treffen. Die eine ist die Wahl der statistischen Sicherheit γ, die angibt mit welcher Wahrscheinlichkeit die unbekannte Größe μ im Konfidenzintervall liegt. Die andere ist die Wahl des Stichprobenumfangs. Hat man sich für eine statistische Sicherheit γ entschieden, dann ist auch z_β festgelegt und man kann die Länge $|I|$ des Konfidenzintervalls I durch den Stichprobenumfang n steuern. Aus der Formel in Satz 36 folgt $|I| = \frac{2\sigma z_\beta}{\sqrt{n}}$. Vergrößert man den Stichprobenumfang n, dann wird das Konfidenzintervall kleiner.

Beispiel 60: Wieviele Packungen muss man in Beispiel 59 überprüfen, damit die Länge des 99%-Konfidenzintervalls höchstens 0.01 kg beträgt?

Für $\gamma = 0.99$ haben wir $z_\beta = 2.58$ gefunden. Es wird verlangt, dass $|I| = \frac{2\sigma z_\beta}{\sqrt{n}} \le 0.01$ gilt. Daraus folgt $\sqrt{n} \ge \frac{2\sigma z_\beta}{0.01} = \frac{2 \cdot 0.02 \cdot 2.58}{0.01} = 10.32$ und $n \ge 106.5$. Man muss mindestens 107 Packungen überprüfen.

Neben den zweiseitigen Konfidenzintervallen gibt es auch einseitige Konfidenzintervalle, die wir im nächsten Satz behandeln.

Satz 37: *Sei $0 < \gamma < 1$ gegeben. Seien $X_1, X_2, \dots, X_n$ unabhängige $N(\mu, \sigma)$-verteilte Messwerte, wobei σ bekannt ist. Sei $M = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ der Mittelwert der Messwerte. Dann sind die beiden Intervalle $I_1 = (-\infty, M + \frac{\sigma z_{1-\gamma}}{\sqrt{n}}]$ und $I_2 = [M - \frac{\sigma z_{1-\gamma}}{\sqrt{n}}, \infty)$ γ-Konfidenzintervalle für μ.*

Beweis: Da $\sqrt{n}\frac{M-\mu}{\sigma}$ die $N(0,1)$-Verteilung hat (siehe Beweis von Satz 36), folgt

$$P(-z_{1-\gamma} \le \sqrt{n}\frac{M-\mu}{\sigma}) = 1 - \Phi(-z_{1-\gamma}) = \Phi(z_{1-\gamma}) = \gamma$$

Daraus folgt $P(\mu \le M + \frac{\sigma z_{1-\gamma}}{\sqrt{n}}) = \gamma$, also $P(\mu \in I_1) = \gamma$. Somit ist das Intervall I_1 ein γ-Konfidenzintervall für μ.

Analog folgt aus $P(\sqrt{n}\frac{M-\mu}{\sigma} \le z_{1-\gamma}) = \Phi(z_{1-\gamma}) = \gamma$, dass $P(\mu \in I_2) = \gamma$ gilt. Somit ist auch I_2 ein γ-Konfidenzintervall für μ. □

Beispiel 61: Aus den in Beispiel 59 erhobenen Packungsinhalten soll das einseitige nach oben offene 99%-Konfidenzintervall I_2 berechnet werden.

Es wurde $\sigma = 0.02$ kg angegeben und $M = 1.01$ kg berechnet. Für $\gamma = 0.99$ findet man $z_{1-\gamma} = 2.33$ in einer Tabelle. Es folgt $\frac{\sigma z_{1-\gamma}}{\sqrt{n}} = \frac{0.02 \cdot 2.33}{\sqrt{10}} = 0.0145$ und daraus $I_2 = [0.9955, \infty)$. Der durchschnittliche Packungsinhalt μ beträgt mit Wahrscheinlichkeit 0.99 mindestens 0.9955 kg.

Es bleibt noch die Frage, was man tut, wenn σ nicht bekannt ist. In diesem Fall ersetzt man σ durch $S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - M)^2}$. Das ist ein aus der Stichprobe $X_1, X_2, \ldots, X_n$ berechneter Schätzwert für σ. Man kann zeigen, dass die Zufallsvariable $\sqrt{n}\frac{M-\mu}{S}$ die sogenannte $T(m)$-Verteilung mit Parameter $m = n-1$ hat. Dann geht man so vor wie oben. Sei $\beta = \frac{1-\gamma}{2}$ und $v_{\beta,m}$ genauso definiert wie z_β, wobei jedoch statt der $N(0,1)$-Verteilung die $T(m)$-Verteilung verwendet wird. Man findet $v_{\beta,m}$ in einer Tabelle. Das Intervall $K = [M - \frac{S v_{\beta,n-1}}{\sqrt{n}}, M + \frac{S v_{\beta,n-1}}{\sqrt{n}}]$ ist dann ein γ-Konfidenzintervall für μ. Entsprechendes gilt für die einseitigen Konfidenzintervalle. Man ersetzt σ durch S und $z_{1-\gamma}$ durch $v_{1-\gamma,n-1}$.

Beispiel 62: In Beispiel 59 sei σ unbekannt. Aus den dort angegebenen Packungsinhalten bestimme man ein 99%-Konfidenzintervall.

Aus $\gamma = 0.99$ folgt $\beta = 0.005$ und man findet $v_{\beta,9} = 3.25$ in einer Tabelle. Der Stichprobenumfang n beträgt 10. In Beispiel 59 wurde $M = 1.01$ ermittelt. Weiters ergibt sich

$$S = \sqrt{\tfrac{1}{9}((1.05 - 1.01)^2 + (0.99 - 1.01)^2 + (1.03 - 1.01)^2 + \cdots + (0.98 - 1.01)^2)} = 0.024$$

Jetzt folgt $\frac{S v_{\beta,n-1}}{\sqrt{n}} = \frac{0.024 \cdot 3.25}{\sqrt{10}} = 0.025$ und $K = [0.985, 1.035]$. Mit Wahrscheinlichkeit 0.99 liegt der durchschnittliche Packungsinhalt μ im Intervall K.

3. Konfidenzintervalle für Prozentsätze

Es sollen unbekannte Prozentsätze (man sagt auch relative Häufigkeiten oder Anteile) geschätzt werden, zum Beispiel der unbekannte Anteil der Raucher in der Gesamtbevölkerung, der Prozentsatz der Wähler der Partei A, oder die relative Häufigkeit von defekten LED-Leuchten in der Tagesproduktion einer Firma.

Wir gehen so vor: Wir legen eine Grundgesamtheit fest (Gesamtbevölkerung, Menge aller Wahlberechtigten, Tagesproduktion der Firma) und interessieren uns für eine Eigenschaft (ist Raucher, wählt Partei A, ist defekt), die in der Grundgesamtheit mit relativer Häufigkeit p auftritt. Wir geben eine Sicherheitswahrscheinlichkeit γ vor und ziehen eine zufällige Stichprobe vom Umfang n mit Zurücklegen aus der Grundgesamtheit. Sei X die Anzahl mit der die Eigenschaft in der Stichprobe auftritt (Anzahl der Raucher, Anzahl der Wähler der Partei A, Anzahl der defekten LED-Leuchten). Aus dieser Anzahl X soll ein Intervall I berechnet werden, das p mit einer Wahrscheinlichkeit $\geq \gamma$ enthält. Man nennt I ein γ-Konfidenzintervall für den Prozentsatz p. Dabei ist zu beachten, dass I vom Zufall abhängt, nicht aber p, das zwar unbekannt, aber fest ist.

Da mit Zurücklegen gezogen wird, hat X die $B(n,p)$-Verteilung. Üblicherweise ist der Stichprobenumfang n groß genug, um die Approximation durch die Normalverteilung zu rechtfertigen. Wir nehmen daher an, dass die Zufallsvariable X die $N(\mu, \sigma)$-Verteilung hat mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$. Auch ist es so, dass in der Praxis natürlich ohne Zurücklegen gezogen wird, also schon in der Annahme der Binomialverteilung für X ein

Fehler enthalten ist. Die Stichprobe ist jedoch gegenüber der Grundgesamtheit, aus der gezogen wird, so klein, dass auch dieser Fehler vernachlässigt werden kann.

Um ein γ-Konfidenzintervall zu berechnen, verwenden wir den folgenden Hilfssatz, der leicht zu beweisen ist.

Hilfssatz A: *Die Funktion $s(x) = \sqrt{x(1-x)}$ für $x \in [0,1]$ stellt einen Halbkreis mit Mittelpunkt $(\frac{1}{2}, 0)$ und Radius $\frac{1}{2}$ dar. Insbesondere ist s monoton wachsend auf $[0, \frac{1}{2}]$, hat das Maximum $\frac{1}{2}$ im Punkt $\frac{1}{2}$ und ist monoton fallend auf $[\frac{1}{2}, 1]$.*

Satz 38: *Sei p die unbekannte relative Häufigkeit einer Eigenschaft in der Grundgesamtheit. Sei n der Stichprobenumfang und X die Anzahl mit der die Eigenschaft in der Stichprobe auftritt. Man wählt $\gamma \in (0,1)$. Sei $\beta = \frac{1-\gamma}{2}$ und $I = [\frac{1}{n}X - \frac{z_\beta}{2\sqrt{n}}, \frac{1}{n}X + \frac{z_\beta}{2\sqrt{n}}]$. Dann gilt $P(p \in I) \geq \gamma$, das heißt I ist ein γ-Konfidenzintervall für p.*

Beweis: Da wir X als $N(\mu, \sigma)$-verteilt mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$ annehmen, hat $Y = \frac{X-np}{\sqrt{np(1-p)}}$ die $N(0,1)$-Verteilung. Daraus folgt

$$P(-z_\beta \leq Y \leq z_\beta) = \Phi(z_\beta) - \Phi(-z_\beta) = \Phi(z_\beta) - (1 - \Phi(z_\beta)) = 1 - \beta - \beta = \gamma$$

Wir formen dieses Ereignis um

$$\begin{aligned} -z_\beta \leq Y \leq z_\beta \;&\Leftrightarrow\; -\frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)} \leq \frac{1}{n}X - p \leq \frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)} \\ &\Leftrightarrow\; \frac{1}{n}X - \frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)} \leq p \leq \frac{1}{n}X + \frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)} \end{aligned}$$

Sei $\tilde{I} = [\frac{1}{n}X - \frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)}, \frac{1}{n}X + \frac{z_\beta}{\sqrt{n}}\sqrt{p(1-p)}]$. Wir haben $P(p \in \tilde{I}) = \gamma$ gezeigt. Nun gilt $\sqrt{p(1-p)} \leq \frac{1}{2}$ für alle $p \in [0,1]$ nach dem Hilfssatz. Ersetzen wir $\sqrt{p(1-p)}$ in $\tilde{I}$ durch $\frac{1}{2}$, so wandert der linke Endpunkt nach links und der rechte Endpunkt nach rechts. Wir erhalten dadurch das Intervall I. Es gilt also $\tilde{I} \subseteq I$. Ist $p \in \tilde{I}$ erfüllt, dann auch $p \in I$. Aus $P(p \in \tilde{I}) = \gamma$ folgt daher $P(p \in I) \geq \gamma$. □

Bemerkung: Im Beweis von Satz 38 sind wir etwas großzügig vorgegangen. Um das im Intervall $\tilde{I}$ vorkommende unbekannte p loszuwerden, haben wir das Intervall vergrößert und so I erhalten. Hat man bei diesem Übergang wenig verschenkt, dann wird I ein gutes Konfidenzintervall sein. Hat man dabei viel verschenkt, so wird I wenig Aussagekraft haben, weil es zu groß ist. Wir vergleichen die Längen $|I|$ und $|\tilde{I}|$ dieser beiden Intervalle. Der Quotient $\frac{|\tilde{I}|}{|I|} = 2\sqrt{p(1-p)}$ gibt an, um wieviel $\tilde{I}$ kleiner als I ist.

Hat man $0.3 \leq p \leq 0.7$, dann gilt $1 \geq 2\sqrt{p(1-p)} \geq 0.9$, wie aus dem Hilfssatz leicht folgt. Daraus erhält man $|I| \geq |\tilde{I}| \geq 0.9|I|$. In diesem Fall hat man höchstens 10% verschenkt.

Hat man $0.2 \leq p \leq 0.8$, dann gilt $1 \geq 2\sqrt{p(1-p)} \geq 0.8$, wie wieder aus dem Hilfssatz folgt. Daraus erhält man $|I| \geq |\tilde{I}| \geq 0.8|I|$. In diesem Fall hat man höchstens 20% verschenkt.

Weiß man also, dass das unbekannte p zwischen 0.3 und 0.7 liegt, so ist I ein gutes Konfidenzintervall. Das gilt auch noch, wenn p zwischen 0.2 und 0.8 liegt. Je näher jedoch p bei 0 oder 1 liegt, umso unbrauchbarer wird I, weil es dann viel zu groß ist.

Beispiel 63: In einer Stichprobe von $n = 400$ Wahlberechtigten waren 164 Wähler der Partei A. Gesucht ist ein 95%-Konfidenzintervall für den Anteil p der A-Wähler unter allen Wahlberechtigten.

Wir berechnen das Intervall I aus Satz 38. Wegen $\gamma = 0.95$ folgt $\beta = \frac{1-\gamma}{2} = 0.025$. Man findet $z_\beta = 1.96$ in einer Tabelle. Jetzt erhalten wir $\frac{z_\beta}{2\sqrt{n}} = \frac{1.96}{2\sqrt{400}} = 0.049$. Wegen $\frac{1}{n}X = \frac{164}{400} = 0.41$ folgt $I = [0.361, 0.459]$.

Beispiel 64: Wie groß muss man den Stichprobenumfang n wählen, wenn die Länge des 95%-Konfidenzintervalls I höchstens 0.02 sein soll?

Die Länge von I ist $\frac{z_\beta}{\sqrt{n}}$. Es wird also verlangt, dass $\frac{z_\beta}{\sqrt{n}} \leq 0.02$ gilt. Für $\gamma = 0.95$ wurde $z_\beta = 1.96$ in Beispiel 63 ermittelt. Es folgt $\sqrt{n} \geq \frac{1.96}{0.02} = 98$, das heißt $n \geq 9604$.

Genauso wie das zweiseitige Konfidenzintervall I in Satz 38 berechnet man auch die einseitigen γ-Konfidenzintervalle $I_1 = (-\infty, \frac{1}{n}X + \frac{z_{1-\gamma}}{2\sqrt{n}}]$ und $I_2 = [\frac{1}{n}X - \frac{z_{1-\gamma}}{2\sqrt{n}}, \infty)$. Da p aber ein Prozentsatz ist und daher immer im Intervall $[0,1]$ liegt, kann man auch $I_1 = [0, \frac{1}{n}X + \frac{z_{1-\gamma}}{2\sqrt{n}}]$ und $I_2 = [\frac{1}{n}X - \frac{z_{1-\gamma}}{2\sqrt{n}}, 1]$ schreiben.

Beispiel 65: Für das Umfrageergebnis aus Beispiel 63 soll das nach oben offene einseitige 95%-Konfidenzintervall I_2 berechnet werden.

Für $\gamma = 0.95$ finden wir $z_{1-\gamma} = 1.65$ in einer Tabelle. Es folgt $\frac{z_{1-\gamma}}{2\sqrt{n}} = \frac{1.65}{2\sqrt{400}} = 0.041$. Wegen $\frac{1}{n}X = \frac{164}{400} = 0.41$ erhält man $I_2 = [0.369, 1]$.

Was soll man tun, wenn man nicht so viel verschenken will? Eine Möglichkeit, die etwas unexakt ist, besteht darin, das im Beweis von Satz 38 berechnete Intervall $\tilde{I}$ zu verwenden und dort den unbekannten Prozentsatz p durch den Prozentsatz $\frac{X}{n}$, der in der Stichprobe auftritt, zu ersetzen. Man erhält dadurch das Intervall

$$J = \Big[\ \frac{X}{n} - \frac{z_\beta}{\sqrt{n}}\sqrt{\frac{X}{n}(1-\frac{X}{n})}, \frac{X}{n} + \frac{z_\beta}{\sqrt{n}}\sqrt{\frac{X}{n}(1-\frac{X}{n})}\ \Big]$$

als γ-Konfidenzintervall für p, wobei wieder $\beta = \frac{1-\gamma}{2}$ zu setzen ist.

Beispiel 66: Der Anteil p der defekten LED-Leuchten in der Tagesproduktion einer Fabrik ist sicher kleiner als 10%. Man zieht eine Stichprobe von 1600 LED-Leuchten und findet 128 defekte darunter. Gesucht ist ein 99%-Konfidenzintervall für p.

Da p klein ist, berechnen wir J. Wegen $\gamma = 0.99$ folgt $\beta = \frac{1-\gamma}{2} = 0.005$. Man findet $z_\beta = 2.58$ in einer Tabelle. Weiters folgt $\frac{1}{n}X = \frac{128}{1600} = 0.08$. Jetzt erhalten wir $\frac{z_\beta}{\sqrt{n}}\sqrt{\frac{X}{n}(1-\frac{X}{n})} = \frac{2.58}{\sqrt{1600}}\sqrt{0.08 \cdot 0.92} = 0.0175$. Daraus ergibt sich $J = [0.0625, 0.0975]$.
Für das Intervall I hätte man $I = [0.048, 0.112]$, also etwas viel ungenaueres, erhalten.

Bemerkung: Die Ungleichung $-z_\beta \leq \frac{X-np}{\sqrt{np(1-p)}} \leq z_\beta$ aus dem Beweis von Satz 38 ist äquivalent zu $\frac{(X-np)^2}{np(1-p)} \leq z_\beta^2$ und diese wieder zu $p^2(n^2 + nz_\beta^2) - 2p(nX + \frac{1}{2}nz_\beta^2) + X^2 \leq 0$. Sie lässt sich exakt nach p auflösen. (Sind λ_1 und λ_2 die beiden reellen Lösungen einer quadratischen Gleichung $x^2 + ux + v = 0$ und $\lambda_1 < \lambda_2$, dann ist das Intervall $[\lambda_1, \lambda_2]$ die Lösungsmenge der Ungleichung $x^2 + ux + v \leq 0$.) Dadurch erhält man, dass

$$K = \Big[\ \frac{X+\frac{1}{2}z_\beta^2-Y}{n+z_\beta^2}, \frac{X+\frac{1}{2}z_\beta^2+Y}{n+z_\beta^2}\ \Big] \quad \text{mit} \quad Y = z_\beta\sqrt{\frac{X(n-X)}{n} + \frac{1}{4}z_\beta^2} \quad \text{und} \quad \beta = \frac{1-\gamma}{2}$$

ein γ-Konfidenzintervall für p ist. Dieses Konfidenzintervall ist zwar komplizierter als die anderen, aber es ist ein exaktes Konfidenzintervall. Es sollte jedenfalls verwendet werden, wenn p sehr nahe bei 0 oder sehr nahe bei 1 liegt.

4. Statistische Tests für Prozentsätze

Statistische Tests verwendet man, um etwas nachzuweisen. Bei der Produktion von LED-Leuchten zum Beispiel kann man einen tolerierbaren Ausschussanteil p_0 festlegen. Die Firma, die die LED-Leuchten produziert, will dann nachweisen, dass der unbekannte Anteil p an defekten LED-Leuchten kleiner als p_0 ist. Ein anderes Beispiel ist das Testen eines Medikaments. Das derzeit am Markt befindliche Medikament gegen eine bestimmte Krankheit hilft bei einem Prozentsatz p_0 der behandelten Patienten. Ein neues Medikament gegen dieselbe Krankheit kommt auf den Markt. Man will nachweisen, dass der noch unbekannte Prozentsatz p der behandelten Patienten, bei denen das neue Medikament hilft, größer als p_0 ist.

Die Vorgangsweise beim statistischen Test ist folgende: Es liegt eine Grundgesamtheit (Gesamtproduktion der LED-Leuchten, Menge aller Patienten) vor, in der eine Eigenschaft (LED-Leuchte defekt, Medikament hilft) mit unbekannter relativer Häufigkeit p auftritt. Das Gegenteil der nachzuweisenden Aussage wird als Hypothese H_0 und die nachzuweisende Aussage als Alternative H_1 formuliert. Im Beispiel mit den LED-Leuchten führt das zu folgendem Test

$$H_0 : p \geq p_0 \quad \text{gegen} \quad H_1 : p < p_0$$

Im zweiten Beispiel, in dem das neue Medikament geprüft wird, wird

$$H_0 : p \leq p_0 \quad \text{gegen} \quad H_1 : p > p_0$$

getestet. Die Entscheidung, ob H_0 verworfen werden soll, wird auf Grund einer zufälligen Stichprobe getroffen, die aus der Grundgesamtheit gezogenen wird. Daher wird diese Entscheidung auch nur mit einer gewissen Wahrscheinlichkeit, die groß sein soll, richtig ausfallen. Man gibt eine Irrtumswahrscheinlichkeit α vor (typische Werte für α sind 0.05 und 0.01) und legt den Stichprobenumfang n fest. Dann zieht man eine Stichprobe aus der Grundgesamtheit (mit Zurücklegen). Sei X die Anzahl (Häufigkeit), mit der die Eigenschaft in dieser zufälligen Stichprobe auftritt. Auf Grund dieser Anzahl X trifft man dann die Entscheidung, die Hypothese H_0 zu verwerfen oder nicht zu verwerfen. Die Entscheidungsregel soll so gewählt werden, dass die Wahrscheinlichkeit, die Hypothese H_0 zu verwerfen, obwohl sie richtig ist, höchstens α beträgt.

Wir suchen eine entsprechende Entscheidungsregel. Die möglichen Werte von X sind die Anzahlen, mit der die Eigenschaft in der Stichprobe vom Umfang n auftreten kann. Daher hat X den Wertebereich $S = \{0, 1, 2, \ldots, n\}$. Wir wählen eine Teilmenge V von S und verwerfen die Hypothese H_0, wenn sich nach dem Ziehen der Stichprobe herausstellt, dass X in V liegt. Man nennt V den Verwerfungsbereich (Ablehnungsbereich) des Tests. Aus den oben durchgeführten Überlegungen ergeben sich folgende Bedingungen:
(1) V enthält die Werte aus S, die am meisten gegen H_0 (und daher für H_1) sprechen
(2) V ist maximal, aber so, dass $P(X \in V) \leq \alpha$ gilt, wenn H_0 richtig ist
Da H_0 verworfen wird, wenn X in V fällt, ist (1) klar. In (2) findet man die oben formulierte Bedingung. Die Wahrscheinlichkeit, die Hypothese H_0 zu verwerfen, das ist $P(X \in V)$, soll $\leq \alpha$ sein, wenn H_0 richtig ist. Unter dieser Bedingung soll jedoch V möglichst groß sein. Das Risiko, H_0 bei Richtigkeit zu verwerfen, kontrolliert man durch die Irrtumswahrscheinlichkeit α. Das Risiko, H_0 bei Unrichtigkeit nicht zu verwerfen, macht man dadurch möglichst klein, dass man V maximal wählt.

Wir verwenden die in (1) und (2) formulierten Bedingungen, um den Verwerfungsbereich auszurechnen. Dazu führen wir eine Bezeichnung für die Verteilungsfunktion der Binomialverteilung ein und beweisen einen Hilfssatz.

Definition: Für $0 \leq k \leq n$ definieren wir $\Psi_{n,p}(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$, das heißt $\Psi_{n,p}(k) = P(X \leq k)$ für eine $B(n,p)$-verteilte Zufallsvariable X.

Hilfssatz B: *Für festes n und k gilt $\Psi_{n,p}(k) \geq \Psi_{n,q}(k)$, wenn $p \leq q$ ist.*

Beweis: Wir definieren die Funktion $h(p)$ für $p \in (0,1)$ durch
$h(p) = \Psi_{n,p}(k) = \binom{n}{0}(1-p)^n + \binom{n}{1}p(1-p)^{n-1} + \binom{n}{2}p^2(1-p)^{n-2} + \cdots + \binom{n}{k}p^k(1-p)^{n-k}$
Der Satz ist bewiesen, wenn wir $h'(p) \leq 0$ zeigen. Mit Hilfe der Produktregel kann man die Summanden in obiger Summe der Reihe nach differenzieren. Man erhält

$$\begin{aligned} h'(p) = &- \tbinom{n}{0}n(1-p)^{n-1} + \tbinom{n}{1}(1-p)^{n-1} - \tbinom{n}{1}(n-1)p(1-p)^{n-2} \\ &+ \tbinom{n}{2}2p(1-p)^{n-2} - \tbinom{n}{2}(n-2)p^2(1-p)^{n-1} + \dots \\ &\cdots + \tbinom{n}{k}kp^{k-1}(1-p)^{n-k} - \tbinom{n}{k}(n-k)p^k(1-p)^{n-k-1} \end{aligned}$$

Da $\binom{n}{i-1}(n-i+1) = \frac{n!}{(i-1)!(n-i+1)!}(n-i+1) = \frac{n!}{(i-1)!(n-i)!} = \frac{n!}{i!(n-i)!}i = \binom{n}{i}i$ für alle i mit $1 \leq i \leq n$ gilt, kürzen sich in dieser Summe immer zwei aufeinanderfolgende Summanden weg. Es bleibt nur der letzte Summand übrig, also

$$h'(p) = -\tbinom{n}{k}(n-k)p^k(1-p)^{n-k-1}$$

Wir haben $h'(p) \leq 0$ gezeigt. Der Hilfssatz ist bewiesen. □

Satz 39: *Sei p die unbekannte relative Häufigkeit, mit der eine Eigenschaft in der Grundgesamtheit auftritt. Der Stichprobenumfang sei n, die Irrtumswahrscheinlichkeit sei α und der durchzuführende Test sei $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$. Bestimmt man das minimale k_0, sodass $\Psi_{n,p_0}(k_0 - 1) \geq 1 - \alpha$ gilt, und wählt $V = \{k_0, k_0 + 1, \dots, n\}$, dann erfüllt V die Bedingungen (1) und (2) für diesen Test, ist also der Verwerfungsbereich.*

Beweis: Die Zufallsvariable X, die die Häufigkeit angibt, mit der die Eigenschaft in der Stichprobe vom Umfang n auftritt, hat nach Satz 16 die $B(n,p)$-Verteilung.

Große Werte von X sprechen am meisten gegen die Hypothese H_0, die ja behauptet, dass die Eigenschaft in der Grundgesamtheit mit relativer Häufigkeit $\leq p_0$ vorkommt. Um (1) zu erfüllen, wählen wir daher $V = \{k_0, k_0 + 1, \dots, n\}$ für ein k_0 als Verwerfungsbereich.

Um (2) zu erfüllen, muss V möglichst groß sein, aber $P(X \in V) \leq \alpha$, wenn H_0 gilt. Es gilt $P(X \in V) = P(X \geq k_0) = 1 - P(X \leq k_0 - 1) = 1 - \Psi_{n,p}(k_0 - 1)$, da ja X die $B(n,p)$-Verteilung hat. Daher ist k_0 minimal zu wählen, sodass aber $1 - \Psi_{n,p}(k_0 - 1) \leq \alpha$ für alle $p \leq p_0$ gilt. Aus dem Hilfssatz folgt $1 - \Psi_{n,p}(k_0 - 1) \leq 1 - \Psi_{n,p_0}(k_0 - 1)$ für alle $p \leq p_0$. Also ist (2) äquivalent dazu, k_0 minimal zu wählen, sodass $1 - \Psi_{n,p_0}(k_0 - 1) \leq \alpha$ gilt, das heißt $\Psi_{n,p_0}(k_0 - 1) \geq 1 - \alpha$. Damit ist der Satz bewiesen. □

Beispiel 67: Für den Test $H_0 : p \leq 0.3$ gegen $H_1 : p > 0.3$ ist der Verwerfungsbereich bei einem Stichprobenumfang $n = 10$ und verschiedenen Irrtumswahrscheinlichkeiten gesucht.

Der Verwerfungsbereich für diesen Test ist $V = \{k_0, k_0+1, \dots, 10\}$. Um k_0 zu bestimmen, erstellen wir eine Wertetabelle für die Abbildung $k \mapsto \Psi_{10,0.3}(k)$.

k	0	1	2	3	4	5	6	7	8	9	10
$\Psi_{10,0.3}(k)$	0.028	0.149	0.383	0.650	0.850	0.953	0.989	0.998	1.000	1.000	1.000

Nach Satz 39 ist k_0 minimal zu wählen, sodass $\Psi_{10,0.3}(k_0 - 1) \geq 1 - \alpha$ gilt. Man kann k_0 jetzt leicht aus der Tabelle ablesen. Für $\alpha = 0.05$ ist $1 - \alpha = 0.95$ und $k_0 - 1 = 5$, also $V = \{6, 7, 8, 9, 10\}$. Für $\alpha = 0.02$ ist $1 - \alpha = 0.98$ und $k_0 - 1 = 6$, also $V = \{7, 8, 9, 10\}$. Für $\alpha = 0.01$ ist $1 - \alpha = 0.99$ und $k_0 - 1 = 7$, also $V = \{8, 9, 10\}$.

Satz 40: *Sei p die unbekannte relative Häufigkeit, mit der eine Eigenschaft in der Grundgesamtheit auftritt. Der Stichprobenumfang sei n, die Irrtumswahrscheinlichkeit sei α und der durchzuführende Test sei $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$. Bestimmt man das maximale k_0, sodass $\Psi_{n,p_0}(k_0) \leq \alpha$ gilt, und wählt $V = \{0, 1, \ldots, k_0\}$, dann erfüllt V die Bedingungen (1) und (2) für diesen Test, ist also der Verwerfungsbereich.*

Beweis: Die Zufallsvariable X, die die Häufigkeit angibt, mit der die Eigenschaft in der Stichprobe vom Umfang n auftritt, hat nach Satz 16 die $B(n, p)$-Verteilung.

Kleine Werte von X sprechen am meisten gegen die Hypothese H_0, die ja einen Anteil $\geq p_0$ für die Eigenschaft in der Grundgesamtheit behauptet. Um (1) zu erfüllen, wählen wir $V = \{0, 1, \ldots, k_0\}$ für ein k_0 als Verwerfungsbereich.

Um (2) zu erfüllen, muss V möglichst groß sein, aber $P(X \in V) \leq \alpha$, wenn H_0 gilt. Da X die $B(n, p)$-Verteilung hat, und daher $P(X \in V) = P(X \leq k_0) = \Psi_{n,p}(k_0)$ gilt, bedeutet das, k_0 maximal zu wählen, sodass $\Psi_{n,p}(k_0) \leq \alpha$ für alle $p \geq p_0$ gilt. Aus dem Hilfssatz folgt $\Psi_{n,p}(k_0) \leq \Psi_{n,p_0}(k_0)$ für alle $p \geq p_0$. Daher ist (2) äquivalent dazu, k_0 maximal zu wählen, sodass $\Psi_{n,p_0}(k_0) \leq \alpha$ gilt. Damit ist der Satz bewiesen. □

Meistens ist der Stichprobenumfang n so groß, dass es schwer ist, mit der Binomialverteilung zu rechnen. Man verwendet dann wieder die Approximation durch die Normalverteilung. Für ganzzahliges k gilt ja $\Psi_{n,p}(k) \approx \Phi(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}})$ nach Kapitel 8.

Beispiel 68: Gesucht ist der Verwerfungsbereich für $H_0 : p \geq 0.5$ gegen $H_1 : p < 0.5$ bei einem Stichprobenumfang $n = 400$ und Irrtumswahrscheinlichkeit $\alpha = 0.01$.

Um den Verwerfungsbereich zu bestimmen, müssen wir das maximale k_0 finden, sodass $\Psi_{400,0.5}(k_0) \leq 0.01$ erfüllt ist. Wegen $\Psi_{400,0.5}(k_0) \approx \Phi(\frac{k_0+\frac{1}{2}-400\cdot 0.5}{\sqrt{400\cdot 0.5\cdot 0.5}}) = \Phi(\frac{k_0+\frac{1}{2}-200}{10})$ ist diese Ungleichung äquivalent zu $\Phi(\frac{k_0+\frac{1}{2}-200}{10}) \leq 0.01 = \Phi(-2.33)$. Wegen Satz 23 (c) erhalten wir dann $\frac{k_0+\frac{1}{2}-200}{10} \leq -2.33$. Es folgt $k_0 \leq 176.2$. Das maximale k_0, das diese Ungleichung erfüllt, ist 176. Daher ist $V = \{0, 1, 2, \ldots, 176\}$ der Verwerfungsbereich.

Um zu entscheiden, ob die Hypothese verworfen wird, ist es nicht notwendig, den Verwerfungsbereich zu bestimmen.

Satz 41: *Die Irrtumswahrscheinlichkeit sei α. Nach dem Ziehen einer Stichprobe vom Umfang n stellt man fest, dass die Eigenschaft mit Häufigkeit s in dieser vorkommt.*
(a) *Sei V der Verwerfungsbereich für den Test $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$. Dann gilt $s \in V \;\Leftrightarrow\; \Psi_{n,p_0}(s-1) \geq 1-\alpha$.*
(b) *Sei V der Verwerfungsbereich für den Test $H_0 : p \geq p_0$ gegen $H_1 : p < p_0$. Dann gilt $s \in V \;\Leftrightarrow\; \Psi_{n,p_0}(s) \leq \alpha$.*

Beweis: Es gilt $\Psi_{n,p_0}(k) = \sum_{i=0}^{k} \binom{n}{i} p_0^i (1-p_0)^{n-i}$. Wenn man k vergrößert, dann wird auch $\Psi_{n,p_0}(k)$ größer. In (a) ist $V = \{k_0, k_0+1, \ldots, n\}$ der Verwerfungsbereich, wobei k_0 minimal gewählt wird mit $\Psi_{n,p_0}(k_0-1) \geq 1-\alpha$. Deshalb gilt $s \in V$, also $s \geq k_0$ genau dann, wenn $\Psi_{n,p_0}(s-1) \geq 1-\alpha$ gilt.

Analog erhalten wir (b). Jetzt ist $V = \{0, 1, 2, \ldots, k_0\}$ der Verwerfungsbereich, wobei k_0 maximal gewählt wird mit $\Psi_{n,p_0}(k_0) \leq \alpha$. Da $\Psi_{n,p_0}(k)$ monoton wachsend in k ist, gilt $s \in V$, also $s \leq k_0$ genau dann, wenn $\Psi_{n,p_0}(s) \leq \alpha$ gilt. □

Wendet man Satz 41 in Beispielen an, so kann man entweder mit der Binomialverteilung rechnen oder mit der Approximation durch die Normalverteilung. Dazu je ein Beispiel.

Beispiel 69: Bei LED-Leuchten wird ein Ausschussanteil von 0.015 toleriert. Jemand möchte nachweisen, dass er größer ist. Aus einer Lieferung LED-Leuchten zieht er eine Stichprobe vom Umfang $n = 20$ und findet $s = 2$ defekte LED-Leuchten. Wird die Hypothese $H_0 : p \leq 0.015$ bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ abgelehnt?

Wir verwenden Satz 41 (a). Es wird genau dann verworfen, wenn $\Psi_{n,p_0}(s-1) \geq 1-\alpha$ ist, also $\Psi_{20,0.015}(1) \geq 0.95$. Es gilt $\Psi_{20,0.015}(1) = \binom{20}{0}0.985^{20} + \binom{20}{1}0.015 \cdot 0.985^{19} = 0.964$. Wegen $0.964 \geq 0.95$ wird die Hypothese H_0 verworfen. Wir können es daher als erwiesen ansehen, dass der Anteil an defekten LED-Leuchten über der Toleranzgrenze liegt. Die Wahrscheinlichkeit für einen Irrtum ist dabei höchstens 0.05.

Beispiel 70: Nach der letzten Statistik rauchen 40% der Männer. Nach einer Antiraucherkampagne findet man unter 1000 zufällig gewählten Männern 366 Raucher. Hat sich der Raucheranteil verringert? Man teste mit Irrtumswahrscheinlichkeit $\alpha = 0.05$.

Es geht darum, einen verringerten Raucheranteil nachzuweisen. Daher nehmen wir das Gegenteil als Hypothese und testen $H_0 : p \geq 0.4$ gegen $H_1 : p < 0.4$.

Wir verwenden Satz 41 (b). Es wird genau dann verworfen, wenn $\Psi_{n,p_0}(s) \leq \alpha$ ist, also $\Psi_{1000,0.4}(366) \leq 0.05$. Wir verwenden die Approximation durch die Normalverteilung. Es gilt $\Psi_{1000,0.4}(366) \approx \Phi(\frac{366+\frac{1}{2}-1000\cdot 0.4}{\sqrt{1000\cdot 0.4\cdot 0.6}}) = \Phi(-2.16) = 1 - \Phi(2.16) = 0.016$. Wegen $0.016 \leq 0.05$ wird H_0 verworfen. Bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ kann es als erwiesen gelten, dass der Raucheranteil durch die Kampagne kleiner geworden ist.

Neben den bisher behandelten einseitigen Tests gibt es auch einen zweiseitigen Test. Sei p wieder die unbekannte relative Häufigkeit, mit der eine Eigenschaft in der Grundgesamtheit auftritt. Beim zweiseitigen Test wird

$$H_0 : p = p_0 \qquad \text{gegen} \qquad H_1 : p \neq p_0$$

getestet. Sei n der Stichprobenumfang und α die vorgegebene Irrtumswahrscheinlichkeit. Sei X die Häufigkeit, mit der die Eigenschaft in der Stichprobe auftritt. Die möglichen Werte von X sind $0, 1, 2, \dots, n$. Wir suchen einen Verwerfungsbereich V, der (1) und (2) erfüllt. Beim zweiseitigen Test sprechen sowohl kleine als auch große Werte von X gegen die Hypothese. Daher wählen wir $V = \{0, 1, \dots, k_0, k_1, k_1+1, \dots, n\}$. Wir wählen k_0 maximal mit $\Psi_{n,p_0}(k_0) \leq \frac{\alpha}{2}$ und k_1 minimal mit $1-\Psi_{n,p_0}(k_1-1) \leq \frac{\alpha}{2}$. Wenn jetzt $H_0 : p = p_0$ richtig ist, dann gilt $P(X \in V) = P(X \leq k_0) + P(X \geq k_1) = \Psi_{n,p_0}(k_0) + 1 - \Psi_{n,p_0}(k_1 - 1) \leq \alpha$. Wir haben also wieder V maximal gewählt, sodass die Wahrscheinlichkeit, die Hypothese zu verwerfen, obwohl sie richtig ist, höchstens gleich α ist.

Beispiel 71: Wir prüfen einen Würfel. Sei p die unbekannte Wahrscheinlichkeit (relative Häufigkeit), mit der 6 auftritt. Wir testen $H_0 : p = \frac{1}{6}$ gegen $H_1 : p \neq \frac{1}{6}$. Wir würfeln 80 Mal. Gesucht ist der Verwerfungsbereich bei Irrtumswahrscheinlichkeit $\alpha = 0.05$.

Wir bestimmen k_0 maximal, sodass $\Psi_{n,p_0}(k_0) \leq \frac{\alpha}{2}$ erfüllt ist, und k_1 minimal, sodass $\Psi_{n,p_0}(k_1 - 1) \geq 1 - \frac{\alpha}{2}$ gilt. Mit Hilfe der Approximation durch die Normalverteilung erhalten wir $\Psi_{80,\frac{1}{6}}(k_0) \approx \Phi(\frac{k_0+\frac{1}{2}-80\cdot\frac{1}{6}}{\sqrt{80\cdot\frac{1}{6}\cdot\frac{5}{6}}}) = \Phi(\frac{k_0-12.83}{3.33})$. Wegen $\frac{\alpha}{2} = 0.025 = \Phi(-1.96)$ wird $\Psi_{n,p_0}(k_0) \leq \frac{\alpha}{2}$ zu $\frac{k_0-12.83}{3.33} \leq -1.96$, woraus $k_0 \leq 6.30$ folgt. Da k_0 maximal sein soll, ergibt sich $k_0 = 6$.

Genauso wird k_1 bestimmt. Es gilt $\Psi_{80,\frac{1}{6}}(k_1 - 1) = \Phi(\frac{k_1-\frac{1}{2}-80\cdot\frac{1}{6}}{\sqrt{80\cdot\frac{1}{6}\frac{5}{6}}}) = \Phi(\frac{k_1-13.83}{3.33})$. Wegen $1 - \frac{\alpha}{2} = 0.975 = \Phi(1.96)$ ist das minimale k_1 zu bestimmen, sodass $\frac{k_1-13.83}{3.33} \geq 1.96$ gilt, woraus $k_1 \geq 20.36$ folgt. Da k_1 minimal sein soll, ergibt sich $k_1 = 21$. Wir erhalten den Verwerfungsbereich $V = \{0, 1, \ldots, 6, 21, 22, \ldots, 80\}$. Liegt die Anzahl, mit der 6 unter den 80 Würfen auftritt, in V, dann kann man bei Irrtumswahrscheinlichkeit 0.05 schließen, dass der Würfel die Augenzahl 6 nicht mit Wahrscheinlichkeit $\frac{1}{6}$ liefert.

Auch beim zweiseitigen Test kann man entscheiden, ob eine Zahl s in den Verwerfungsbereich V fällt, ohne diesen auszurechnen. Da sich V aus den Verwerfungsbereichen von zwei einseitigen Tests zusammensetzt, erhalten wir mit Hilfe von Satz 41

$$s \in V \quad \Leftrightarrow \quad s \le k_0 \text{ oder } s \ge k_1 \quad \Leftrightarrow \quad \Psi_{n,p_0}(s) \le \tfrac{\alpha}{2} \text{ oder } \Psi_{n,p_0}(s-1) \ge 1 - \tfrac{\alpha}{2}$$

Das verwenden wir im folgenden Beispiel.

Beispiel 72: Bei 900 Münzenwürfen tritt 473 Mal "Kopf" auf. Ist die Münze fair?

Wir testen $H_0 : p = \frac{1}{2}$ gegen $H_1 : p \neq \frac{1}{2}$ bei Irrtumswahrscheinlichkeit $\alpha = 0.05$. Um zu überprüfen, ob $s = 473$ im Verwerfungsbereich liegt, verwenden wir obige Bedingungen und Approximation durch die Normalverteilung.

$$\Psi_{n,p_0}(s) \approx \Phi\Big(\tfrac{s+\frac{1}{2}-np_0}{\sqrt{np_0(1-p_0)}}\Big) = \Phi\Big(\tfrac{473+\frac{1}{2}-900\cdot\frac{1}{2}}{\sqrt{900\cdot\frac{1}{2}\cdot\frac{1}{2}}}\Big) = \Phi(1.57) = 0.942$$

$$\Psi_{n,p_0}(s-1) \approx \Phi\Big(\tfrac{s-\frac{1}{2}-np_0}{\sqrt{np_0(1-p_0)}}\Big) = \Phi\Big(\tfrac{473-\frac{1}{2}-900\cdot\frac{1}{2}}{\sqrt{900\cdot\frac{1}{2}\cdot\frac{1}{2}}}\Big) = \Phi(1.50) = 0.933$$

Es gilt weder $\Psi_{n,p_0}(s) = 0.942 \le \frac{\alpha}{2} = 0.025$ noch $\Psi_{n,p_0}(s-1) = 0.933 \ge 1 - \frac{\alpha}{2} = 0.975$. Daher liegt $s = 473$ nicht im Verwerfungsbereich. Der Nachweis, dass die Münze unfair ist, ist nicht gelungen. Das Ergebnis ist jedoch sehr knapp.

Zum Abschluß dieses Kapitels beschäftigen wir uns noch mit der Bestimmung des Stichprobenumfangs bei den einseitigen Tests. Durch die Vorgangsweise bei einem Test wird garantiert, dass die Hypothese H_0 bei Richtigkeit nur mit kleiner Wahrscheinlichkeit verworfen wird. Um auch zu garantieren, dass die Hypothese H_0 bei Unrichtigkeit mit großer Wahrscheinlichkeit verworfen wird, muss man den Stichprobenumfang groß genug machen.

Wir tun das für den Test $H_0 : p \le p_0$ gegen $H_1 : p > p_0$, der den Verwerfungsbereich $V = \{k_0, k_0 + 1, \ldots, n\}$ hat. Wir geben noch ein $p_1 > p_0$ und ein $\tilde{\alpha}$ vor und verlangen

(i) $P(X \in V) \le \alpha$ wenn $p \le p_0$

(ii) $P(X \in V) \ge 1 - \tilde{\alpha}$ wenn $p \ge p_1$

Die Bedingung (i) wird bei jedem Test verlangt. Sie besagt, dass H_0 bei Richtigkeit nur mit Wahrscheinlichkeit $\le \alpha$ verworfen wird. Es kommt (ii) dazu: Wenn $p \ge p_1$ gilt (das ist fast H_1), dann wird H_0 mit Wahrscheinlichkeit $\ge 1 - \tilde{\alpha}$ verworfen. Wir bestimmen einen (möglichst kleinen) Stichprobenumfang n und k_0 so, dass (i) und (ii) gelten.

Nach dem Hilfssatz ist $P(X \in V) = 1 - \Psi_{n,p}(k_0 - 1)$ monoton wachsend in p. Daher ist (i) äquivalent zu $1 - \Psi_{n,p_0}(k_0 - 1) \le \alpha$, das heißt zu $\Psi_{n,p_0}(k_0 - 1) \ge 1 - \alpha$, und (ii) ist äquivalent zu $1 - \Psi_{n,p_1}(k_0 - 1) \ge 1 - \tilde{\alpha}$, das heißt zu $\Psi_{n,p_1}(k_0 - 1) \le \tilde{\alpha}$.

Beispiel 73: Sei $p_0 = 0.1$ der Anteil an defekten LED-Leuchten, der toleriert wird. Es soll $H_0 : p \le p_0$ gegen $H_1 : p > p_0$ bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ getestet werden. Stichprobenumfang n und Verwerfungsbereich $V = \{k_0, k_0 + 1, \ldots, n\}$ sind so zu bestimmen, dass H_0 mit Wahrscheinlichkeit ≥ 0.95 abgelehnt wird, wenn $p \ge 0.15$ gilt.

Es wird verlangt, dass (i) und (ii) erfüllt sind mit $\alpha = \tilde{\alpha} = 0.05$, mit $p_0 = 0.1$ und mit $p_1 = 0.15$. Wegen $\Psi_{n,p}(k_0-1) \approx \Phi\big(\frac{k_0-\frac{1}{2}-np}{\sqrt{np(1-p)}}\big)$ erhalten wir die zu (i) und (ii) äquivalenten

Ungleichungen

$$\Phi\left(\frac{k_0-\frac{1}{2}-0.1n}{\sqrt{0.1\cdot 0.9\cdot n}}\right) \ge 0.95 \qquad \text{und} \qquad \Phi\left(\frac{k_0-\frac{1}{2}-0.15n}{\sqrt{0.15\cdot 0.85\cdot n}}\right) \le 0.05$$

$$\frac{k_0-\frac{1}{2}-0.1n}{\sqrt{0.1\cdot 0.9\cdot n}} \ge 1.65 \qquad \text{und} \qquad \frac{k_0-\frac{1}{2}-0.15n}{\sqrt{0.15\cdot 0.85\cdot n}} \le -1.65$$

$$k_0 - \tfrac{1}{2} - 0.1n \ge 0.495\sqrt{n} \qquad \text{und} \qquad k_0 - \tfrac{1}{2} - 0.15n \le -0.59\sqrt{n}$$

$$\tfrac{1}{2} + 0.1n + 0.495\sqrt{n} \le k_0 \le \tfrac{1}{2} + 0.15n - 0.59\sqrt{n}$$

Wir müssen ein möglichst kleines n und ein k_0 finden, sodass diese Ungleichungen erfüllt sind. Wegen $\frac{1}{2} + 0.1n + 0.495\sqrt{n} \le \frac{1}{2} + 0.15n - 0.59\sqrt{n}$ folgt $1.085\sqrt{n} \le 0.05n$ und $n \ge 470.89$. Wir rechnen für einige n nach:

n:	470	480	490	500	510	520
$\frac{1}{2} + 0.1n + 0.495\sqrt{n}$:	58.23	59.34	60.46	61.57	62.68	63.79
$\frac{1}{2} + 0.15n - 0.59\sqrt{n}$:	58.21	59.57	60.94	62.31	63.67	65.05

Da k_0 zwischen diesen Werten liegen muss, können wir $n = 500$ und $k_0 = 62$, also $V = \{62, 63, \dots, 500\}$ wählen (eine andere Lösung ist $n = 476$ und $k_0 = 59$).

Für diesen Test gilt: Endet die Durchführung des Tests mit dem Verwerfen der Hypothese H_0, dann schließen wir, dass $p > p_0$ gilt. Die Wahrscheinlichkeit für einen Irrtum ist dann $\le \alpha$ wegen (i).

Endet die Durchführung des Tests hingegen mit dem Nichtverwerfen der Hypothese H_0, dann schließen wir, dass $p < p_1$ gilt. Die Wahrscheinlichkeit für einen Irrtum ist dann $\le \tilde{\alpha}$, da wegen (ii) ja $P(X \notin V) \le \tilde{\alpha}$ für $p \ge p_1$ gilt.

Wir haben also in jedem Fall eine Schlussfolgerung.

5. Tests für den Erwartungswert einer Normalverteilung

Neben den Tests für Prozentsätze gibt es viele weitere statistische Tests. Wir behandeln hier kurz einen Test für den Erwartungswert μ einer Normalverteilung, der den Tests für Prozentsätze sehr ähnlich ist. Wir können von Beispiel 59 ausgehen. Die Füllgewichte von Waschmittelpackungen einer bestimmten Sorte sind $N(\mu, \sigma)$-verteilt. Die Genauigkeit σ der Abfüllmaschine ist bekannt. Wir wollen überprüfen, ob das durchschnittliche Füllgewicht μ mindestens $\mu_0 = 1$ kg beträgt. Das führt zum Test

$$H_0 : \mu \le \mu_0 \qquad \text{gegen} \qquad H_1 : \mu > \mu_0$$

Wir ziehen eine Stichprobe $X_1, X_2, \dots, X_n$, das sind die Füllgewichte von n zufällig gewählten Waschmittelpackungen. Diese Zufallsvariablen sind dann unabhängig und alle $N(\mu, \sigma)$-verteilt. Wir berechnen das Stichprobenmittel $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Die Zufallsvariable M nimmt Werte in $\mathbb{R}$ an. Wir legen eine Teilmenge V von $\mathbb{R}$ fest und verwerfen H_0 immer dann, wenn M in die Menge V fällt. Dieser Verwerfungsbereich V ist so zu bestimmen, dass (1) und (2) gelten.

Satz 42: *Der Stichprobenumfang sei n, die Irrtumswahrscheinlichkeit sei α und der durchzuführende Test sei $H_0 : \mu \le \mu_0$ gegen $H_1 : \mu > \mu_0$. Der Verwerfungsbereich ist dann $V = [\mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)\, , \infty)$.*

Beweis: Es ist zu erwarten, dass M nahe μ liegt, da M das Stichprobenmittel ist und μ der Erwartungswert des Füllgewichts. Daher sprechen große Werte von M gegen H_0. Wegen (1), wobei jetzt $S = \mathbb{R}$ gilt, ist $V = [c, \infty)$ für ein $c \in \mathbb{R}$ zu wählen.

Wegen (2) ist der Verwerfungsbereich V möglichst groß und somit c möglichst klein zu wählen, sodass $P(M \in V) = P(M \geq c) \leq \alpha$ gilt, wenn H_0 erfüllt ist, das heißt für alle $\mu \leq \mu_0$. Im Beweis von Satz 36 wurde gezeigt, dass $\sqrt{n}\frac{M-\mu}{\sigma}$ die $N(0,1)$-Verteilung hat. Es folgt $P(M \geq c) = P(\sqrt{n}\frac{M-\mu}{\sigma} \geq \sqrt{n}\frac{c-\mu}{\sigma}) = 1 - \Phi(\sqrt{n}\frac{c-\mu}{\sigma})$. Somit gilt $P(M \geq c) \leq \alpha$ für alle $\mu \leq \mu_0$ genau dann, wenn $\Phi(\sqrt{n}\frac{c-\mu}{\sigma}) \geq 1 - \alpha$ für alle $\mu \leq \mu_0$ gilt. Das ist wieder äquivalent zu $\Phi(\sqrt{n}\frac{c-\mu_0}{\sigma}) \geq 1 - \alpha$, da Φ monoton wachsend ist. Unter dieser Bedingung ist c minimal zu wählen. Daher ist c die Lösung der Gleichung $\Phi(\sqrt{n}\frac{c-\mu_0}{\sigma}) = 1 - \alpha$. Löst man diese Gleichung, so ergibt sich $c = \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)$. Damit ist V berechnet. □

Beispiel 74: Die Füllgewichte von Waschmittelpackungen einer bestimmten Sorte seien $N(\mu, \sigma)$-verteilt. Die Genauigkeit der Abfüllmaschine wird mit 0.02 kg angegeben. Es soll $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ mit $\mu_0 = 1$ kg getestet werden. Der Stichprobenumfang ist 10. Gesucht ist der Verwerfungsbereich V für die Irrtumswahrscheinlichkeit $\alpha = 0.05$.

Wir finden $\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) = 1.65$ in einer Tabelle. Es folgt $\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha) = \frac{0.02 \cdot 1.65}{\sqrt{10}} = 0.0104$. Somit ist $V = [1.0104, \infty)$ der Verwerfungsbereich. Für die Stichprobe aus Beispiel 59 haben wir $M = 1.01$ berechnet. Wegen $M \notin V$ wird die Hypothese nicht verworfen. Der Nachweis, dass die durchschnittliche Füllmenge mindestens 1 kg ist, ist nicht gelungen. (Man sollte einen größeren Stichprobenumfang wählen.)

Ganz analog bestimmt man den Verwerfungsbereich für den Test $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$. Man erhält $V = (-\infty\,, \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)] = (-\infty\,, \mu_0 - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)]$. Den Verwerfungsbereich für den zweiseitigen Test $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ kann man aus den Verwerfungsbereichen von zwei einseitigen Tests mit Irrtumswahrscheinlichkeit $\frac{\alpha}{2}$ zusammensetzen. Man erhält $V = (-\infty\,, \mu_0 - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2})] \cup [\mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{\alpha}{2})\,, \infty)$.

6. Der χ^2-Test (chi-quadrat-Test)

Ein Würfel wird 300 Mal geworfen. Die Augenzahlen 1 bis 6 treten dabei mit Häufigkeiten 62, 49, 45, 39, 57, 48 auf. Ist der Würfel fair?

Das ist ein Beispiel für folgende Situation. Ein Zufallsexperiment hat die Ausfälle $1, 2, 3, \dots, k$. Weiters sind Wahrscheinlichkeiten $q_1, q_2, \dots, q_k$ vorgegeben, die natürlich ≥ 0 sind und $\sum_{j=1}^{k} q_j = 1$ erfüllen. Es soll die Hypothese getestet werden, ob für $1 \leq j \leq k$ der Ausfall j mit Wahrscheinlichkeit q_j auftritt.

Um diese Hypothese zu testen, wird das Zufallsexperiment n Mal beobachtet. Die Häufigkeiten, mit denen die Ausfälle $1, 2, \dots, k$ dabei auftreten, seien $N_1, N_2, \dots, N_k$, wobei $N_1 + N_2 + \dots + N_k = n$ gilt. Wir bezeichnen diese Häufigkeiten mit Großbuchstaben, da sie vom Zufall abhängen, also Zufallsvariable sind. Wenn die Hypothese, dass der Ausfall j mit Wahrscheinlichkeit q_j auftritt, richtig ist, dann ist nq_j der Erwartungswert für die Häufigkeit des Ausfalls j unter diesen n Wiederholungen des Zufallsexperiments. Je näher der Vektor $(N_1, N_2, \dots, N_k)$ der beobachteten Häufigkeiten dem Vektor $(nq_1, nq_2, \dots, nq_k)$ der bei Gültigkeit der Hypothese zu erwartenden Häufigkeiten liegt, umso mehr wird die Hypothese bestätigt. Wir messen den Abstand dieser beiden Vektoren durch

$$D = \sum_{j=1}^{k} \frac{(N_j - nq_j)^2}{nq_j}$$

Große Werte von D sprechen gegen die Hypothese, während kleine Werte von D für die Hypothese sprechen.

Um den Verwerfungsbereich zu bestimmen, müssen wir die Verteilung der Zufallsvariablen D kennen. Wenn die Hypothese, dass der Ausfall j mit Wahrscheinlichkeit q_j auftritt, richtig ist, dann hat D (näherungsweise) die $E(\frac{k-1}{2}, \frac{1}{2})$-Verteilung. Der Beweis dafür ist schwer. Diese Verteilung heißt auch χ^2-Verteilung. Wieder definiert man $u_{\alpha,m}$ durch $\int_{u_{\alpha,m}}^{\infty} g_m(x)\,\mathrm{d}x = \alpha$, wobei g_m die Wahrscheinlichkeitsdichte der $E(\frac{m}{2}, \frac{1}{2})$-Verteilung ist. Wie immer ist auch $u_{\alpha,m}$ in Tabellen zu finden. Als Verwerfungsbereich zur Irrtumswahrscheinlichkeit α erhalten wir

$$V = [u_{\alpha,k-1}, \infty)$$

Einerseits liegen die großen Werte von D, das sind die, die gegen die Hypothese sprechen, in V, andererseits gilt $P(D \in V) = P(D \geq u_{\alpha,k-1}) = \int_{u_{\alpha,k-1}}^{\infty} g_{k-1}(x)\,\mathrm{d}x = \alpha$, wenn die Hypothese richtig ist. Würde man V vergrößern, dann würde $P(D \in V) > \alpha$ gelten.

Somit sind die beiden Bedingungen (1) und (2) für den Verwerfungsbereich V erfüllt: Er enthält die Werte von D, die am meisten gegen die Hypothese sprechen, und er ist maximal, sodass aber bei Gültigkeit der Hypothese die Wahrscheinlichkeit, die Hypothese zu verwerfen, $\leq \alpha$ ist.

Jetzt haben wir alles, um den χ^2-Test durchzuführen. Wir kehren zum eingangs angeführten Beispiel zurück.

Beispiel 75: Ein Würfel wird 300 Mal geworfen. Die Augenzahlen 1 bis 6 treten dabei mit Häufigkeiten 62, 49, 45, 39, 57, 48 auf. Mit Hilfe des χ^2-Tests teste man mit Irrtumswahrscheinlichkeit $\alpha = 0.05$, ob jede Augenzahl mit Wahrscheinlichkeit $\frac{1}{6}$ auftritt.

In diesem Beispiel gilt $k = 6$ und $q_j = \frac{1}{6}$ für alle j. Das Zufallsexperiment Würfeln wird $n = 300$ Mal beobachtet. Wir tragen die beobachteten Häufigkeiten und die zu erwartenden Häufigkeiten für die einzelnen Augenzahlen in eine Tabelle ein. Die zu erwartenden Häufigkeiten sind $nq_j = 300 \cdot \frac{1}{6} = 50$ für alle j.

Augenzahl	1	2	3	4	5	6
beobachtete Häufigkeit	62	49	45	39	57	48
erwartete Häufigkeit	50	50	50	50	50	50

Daraus berechnet man leicht

$$D = \frac{(62-50)^2}{50} + \frac{(49-50)^2}{50} + \frac{(45-50)^2}{50} + \frac{(39-50)^2}{50} + \frac{(57-50)^2}{50} + \frac{(48-50)^2}{50} = 6.88$$

In einer Tabelle finden wir $u_{\alpha,k-1} = u_{0.05,5} = 11.07$. Der Wert 6.88 von D fällt nicht in den Verwerfungsbereich $V = [11.07, \infty)$. Es konnte nicht nachgewiesen werden, dass der Würfel unfair ist.

Den chi-quadrat-Test kann man verwenden, um zu testen, ob eine bestimmte Verteilung vorliegt. Im nächsten Beispiel ist das die Poissonverteilung.

Beispiel 76: In einer Kleinstadt wird an 100 Tagen die Anzahl der Autounfälle registriert. An 42 Tagen gab es keinen Unfall, an 36 Tagen gab es einen Unfall, an 14 Tagen gab es zwei Unfälle, an 6 Tagen gab es drei Unfälle und an 2 Tagen gab es mehr als drei Unfälle. Mit Hilfe eines χ^2-Tests zur Irrtumswahrscheinlichkeit $\alpha = 0.05$ ist zu prüfen, ob die Anzahl der Unfälle an einem zufällig gewählten Tag $P(\lambda)$-verteilt ist mit $\lambda = 0.9$.

In diesem Beispiel ist $k = 5$ und $n = 100$. Die Wahrscheinlichkeiten q_j und die zu erwartenden Häufigkeiten nq_j müssen wir mit Hilfe der $P(\lambda)$-Verteilung ermitteln.

$q_0 = \frac{\lambda^0}{0!}e^{-\lambda} = 0.4066$, woraus $nq_0 = 40.66$ folgt.
$q_1 = \frac{\lambda^1}{1!}e^{-\lambda} = 0.3659$, woraus $nq_1 = 36.59$ folgt.
$q_2 = \frac{\lambda^2}{2!}e^{-\lambda} = 0.1647$, woraus $nq_2 = 16.47$ folgt.
$q_3 = \frac{\lambda^3}{3!}e^{-\lambda} = 0.0494$, woraus $nq_3 = 4.94$ folgt.
$q_4 = \sum_{j=4}^{\infty} \frac{\lambda^j}{j!}e^{-\lambda} = 1 - q_0 - q_1 - q_2 - q_3 = 0.0135$, woraus $nq_4 = 1.35$ folgt.

Wir tragen die Häufigkeiten wieder in eine Tabelle ein.

Anzahl der Unfälle	0	1	2	3	≥ 4
beobachtete Häufigkeit	42	36	14	6	2
erwartete Häufigkeit	40.66	36.59	16.47	4.94	1.35

Daraus berechnet man leicht

$$D = \frac{(42-40.66)^2}{40.66} + \frac{(36-36.59)^2}{36.59} + \frac{(14-16.47)^2}{16.47} + \frac{(6-4.94)^2}{4.94} + \frac{(2-1.35)^2}{1.35} = 0.969$$

In einer Tabelle finden wir $u_{\alpha,k-1} = u_{0.05,4} = 9.49$. Der Wert 0.969 von D fällt nicht in den Verwerfungsbereich $V = [9.49, \infty)$. Dieses Ergebnis spricht also nicht gegen die Annahme, dass die Anzahl der Unfälle pro Tag poissonverteilt ist mit $\lambda = 0.9$.

Eine andere Version des χ^2-Tests kann auch zum Testen der Unabhängigkeit zweier zufälliger Merkmale, zum Beispiel Augenfarbe und Haarfarbe, verwendet werden. In einer zufällig gewählten Stichprobe von 100 Personen wurden Augenfarbe und Haarfarbe erhoben. Man erhielt die in der Tabelle angegebenen Anzahlen

	blondes Haar	braunes Haar	schwarzes Haar
blaue Augen	23	17	7
braune Augen	8	21	24

In diesem Fall liegt folgende Situation vor. Wir fassen das erste Merkmal als Zufallsvariable X mit den möglichen Werten $1, 2, \ldots, k$ auf und das zweite Merkmal als Zufallsvariable Y mit den möglichen Werten $1, 2, \ldots, l$. Im Beispiel ist X die Augenfarbe mit den möglichen Werten 1=blau und 2=braun und Y die Haarfarbe mit den möglichen Werten 1=blond, 2=braun und 3=schwarz. Es wird eine zufällige Stichprobe vom Umfang n gezogen. Mit N_{ij} wird die Anzahl der Elemente in der Stichprobe bezeichnet, für die die Werte i und j auftreten. Es gilt $n = \sum_{i=1}^{k} \sum_{j=1}^{l} N_{ij}$. Im Beispiel haben wir $N_{11} = 23$, $N_{12} = 17$, $N_{13} = 7$, $N_{21} = 8$, $N_{22} = 21$, $N_{23} = 24$.

Es soll getestet werden, ob die Zufallsvariablen X und Y unabhängig sind. Sie sind unabhängig genau dann, wenn die Wahrscheinlichkeiten $w(i,j) = P(X = i \wedge Y = j)$ in ein Produkt $w_1(i)w_2(j)$ zerfallen, wobei $w_1(i) = P(X = i)$ und $w_2(j) = P(Y = j)$ gilt. Wir ersetzen die Wahrscheinlichkeiten durch die relativen Häufigkeiten aus der Stichprobe:

$$w(i,j) \text{ durch } \frac{N_{ij}}{n} \qquad w_1(i) \text{ durch } \frac{\sum_{j=1}^{l} N_{ij}}{n} =: r_i \qquad w_2(j) \text{ durch } \frac{\sum_{i=1}^{k} N_{ij}}{n} =: s_j$$

Für die Unabhängigkeit der beiden Zufallsvariablen spricht, wenn $\frac{N_{ij}}{n} \approx r_i s_j$ erfüllt ist, das heißt, wenn $N_{ij} \approx n r_i s_j$ für $1 \leq i \leq k$ und $1 \leq j \leq l$ gilt. Wir messen den Abstand zwischen diesen Zahlen, die bei Unabhängigkeit nahe sein sollten, durch

$$U = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(N_{ij} - nr_i s_j)^2}{nr_i s_j}$$

Große Werte von U sprechen gegen die Hypothese, dass die beiden Zufallsvariablen unabhängig sind. Wenn die Hypothese, dass Unabhängigkeit vorliegt, richtig ist, dann hat U (näherungsweise) die $E(\frac{m}{2}, \frac{1}{2})$-Verteilung mit $m = (k-1)(l-1)$. Wir erhalten genauso wie oben $V = [u_{\alpha,m}, \infty)$ als Verwerfungsbereich zur Irrtumswahrscheinlichkeit α.

Beispiel 77: In einer zufällig gewählten Stichprobe von $n = 100$ Personen wurden Augenfarbe und Haarfarbe erhoben. Man erhielt die in obiger Tabelle angegebnen Anzahlen. Mit Hilfe eines χ^2-Tests zur Irrtumswahrscheinlichkeit $\alpha = 0.05$ soll getestet werden, ob Augenfarbe und Haarfarbe unabhängig sind.

Wir ergänzen die Tabelle mit den Anzahlen durch Zeilen– und Spaltensummen.

	1=blond	2=braun	3=schwarz	Zeilensumme
1=blau	23	17	7	47
2=braun	8	21	24	53
Spaltensumme	31	38	31	100

Man erhält r_1 und r_2, indem man die Zeilensummen durch $n = 100$ dividiert:
$r_1 = \frac{47}{100} = 0.47$, $r_2 = \frac{53}{100} = 0.53$
Man erhält s_1, s_2 und s_3, indem man die Spaltensummen durch $n = 100$ dividiert:
$s_1 = \frac{31}{100} = 0.31$, $s_2 = \frac{38}{100} = 0.38$, $s_3 = \frac{31}{100} = 0.31$
Daraus berechnet man nr_is_j für $1 \le i \le 2$ und $1 \le j \le 3$. Wir schreiben die Resultate in eine Tabelle

	1=blond	2=braun	3=schwarz
1=blau	14.57	17.86	14.57
2=braun	16.43	20.14	16.43

Daraus können wir U berechnen
$U = \frac{(23-14.57)^2}{14.57} + \frac{(17-17.86)^2}{17.86} + \frac{(7-14.57)^2}{14.57} + \frac{(8-16.43)^2}{16.43} + \frac{(21-20.14)^2}{20.14} + \frac{(24-16.43)^2}{16.43} = 16.7$
Für $\alpha = 0.05$ und $m = (k-1)(l-1) = 1 \cdot 2 = 2$ finden wir $u_{\alpha,m} = 5.99$ in einer Tabelle. Der Wert 16.7 von U fällt in den Verwerfungsbereich $V = [5.99, \infty)$. Bei Irrtumswahrscheinlichkeit 0.05 können wir es als erwiesen ansehen, dass Augenfarbe und Haarfarbe nicht unabhängig sind.

VI. Anhang: Einige Anwendungen

In diesem Anhang werden einige Anwendungen der Wahrscheinlichkeitstheorie behandelt. Wir beginnen mit der Einschluss-Ausschluss-Formel, dann beschäftigen wir uns mit Fragen, die beim Ziehen von Stichproben auftreten, nämlich mit dem Problem der ersten Wiederkehr und mit Sammlerproblemen. Zum Abschluss beschäftigen wir uns mit dem Hardy-Weinberg-Gesetz aus der Populationsgenetik und betreiben ein wenig Informationstheorie.

1. Einschluss–Ausschluss–Formel

Für ein Ereignis A definieren wir die Zufallsvariable L_A durch $L_A = 1$, wenn A eintritt, und $L_A = 0$, wenn A nicht eintritt.

Satz 43: *Für Ereignisse A und B gilt $L_{A\cap B} = L_A L_B$ und $L_{A'} = 1 - L_A$.*

Beweis: Wenn $A \cap B$ eintritt, dann treten auch A und B ein. Es gilt also $L_{A\cap B} = 1$ und $L_A = L_B = 1$, sodass $L_{A\cap B} = L_A L_B$ erfüllt ist. Wenn $A \cap B$ nicht eintritt, dann tritt A oder B nicht ein. Es gilt also $L_{A\cap B} = 0$ und mindestens eins von L_A und L_B ist 0, sodass $L_{A\cap B} = L_A L_B$ erfüllt ist. Wir haben gezeigt, dass $L_{A\cap B} = L_A L_B$ immer gilt.
Wenn A eintritt, dann tritt A' nicht ein, sodass $L_A = 1$ und $L_{A'} = 0$ gilt. Wenn A nicht eintritt, dann tritt A' ein, sodass $L_A = 0$ und $L_{A'} = 1$ gilt. Wir haben gezeigt, dass $L_{A'} = 1 - L_A$ immer gilt. □

Satz 44: *Für ein Ereignis A gilt* $\mathrm{E}(L_A) = \mathrm{P}(A)$.

Beweis: Die Zufallsvariable L_A nimmt nur die Werte 0 und 1 an. Daher erhalten wir $\mathrm{E}(L_A) = 0 \cdot P(L_A = 0) + 1 \cdot P(L_A = 1) = P(L_A = 1)$. Da $L_A = 1$ genau dann gilt, wenn A eintritt, haben wir auch $P(L_A = 1) = P(A)$. □

Für Ereignisse A und B erhalten wir $L_{A\cup B} = 1 - L_{(A\cup B)'} = 1 - L_{A'\cap B'} = 1 - L_{A'}L_{B'} = 1 - (1 - L_A)(1 - L_B) = L_A + L_B - L_A L_B = L_A + L_B - L_{A\cap B}$ aus Satz 43. Es folgt $\mathrm{E}(L_{A\cup B}) = \mathrm{E}(L_A) + \mathrm{E}(L_B) - \mathrm{E}(L_{A\cap B})$, da der Erwartungswert linear ist. Wegen Satz 44 ergibt sich $P(A\cup B) = P(A) + P(B) - P(A\cap B)$. Das ist eine etwas umständliche Methode, diese Formel zu beweisen. Aber sie lässt sich leicht verallgemeinern.

Satz 45: *Sei $n \ge 2$. Seien $A_1, A_2, \dots, A_n$ Ereignisse und $C = A_1 \cup A_2 \cup \dots \cup A_n$. Dann gilt $\mathrm{P}(C) = \sum_{j=1}^{n} (-1)^{j-1} S_j$, wobei $S_j = \sum_{I\in\mathcal{M}_j} \mathrm{P}(\bigcap_{i\in I} A_i)$ und $\mathcal{M}_j$ die Menge aller j-elementigen Teilmengen von $\{1, 2, \dots, n\}$ ist.*

Beweis: Es gilt $C' = A_1' \cap A_2' \cap \dots \cap A_n'$. Wir erhalten $L_{C'} = L_{A_1'} L_{A_2'} \dots L_{A_n'}$ und daraus $L_{C'} = (1 - L_{A_1})(1 - L_{A_2}) \dots (1 - L_{A_n})$ mit Hilfe von Satz 43. Wir überlegen uns, was sich durch Ausmultiplizieren der rechten Seite ergibt. Sei $R_j = \sum_{I\in\mathcal{M}_j} \prod_{i\in I} L_{A_i}$ für $1 \le j \le n$. Insbesondere gilt $R_n = L_{A_1} L_{A_2} \dots L_{A_n}$. Man erhält dann

$$L_{C'} = 1 + \sum_{j=1}^{n} (-1)^j R_j$$

da $(-1)^j R_j$ alle Summanden enthält, bei denen aus $n - j$ Klammern der Einser genommen wird und aus den anderen j Klammern L_{A_i}. Aus Satz 43 folgt $R_j = \sum_{I\in\mathcal{M}_j} L_{\cap_{i\in I} A_i}$ und daraus $\mathrm{E}(R_j) = \sum_{I\in\mathcal{M}_j} \mathrm{E}(L_{\cap_{i\in I} A_i})$, da der Erwartungswert linear ist. Wegen Satz 44 erhalten wir $\mathrm{E}(R_j) = \sum_{I\in\mathcal{M}_j} \mathrm{P}(\bigcap_{i\in I} A_i)$. Da auch $L_C = 1 - L_{C'} = \sum_{j=1}^{n} (-1)^{j-1} R_j$

gilt und der Erwartungswert linear ist, erhalten wir $\mathrm{E}(L_C) = \sum_{j=1}^{n}(-1)^{j-1}\mathrm{E}(R_j)$. Da $\mathrm{E}(L_C) = P(C)$ nach Satz 44 gilt, ergibt sich $P(C) = \sum_{j=1}^{n}(-1)^{j-1}\sum_{I\in\mathcal{M}_j}\mathrm{P}(\bigcap_{i\in I}A_i)$. Das ist die gesuchte Formel. $\square$

Beispiel 78: Es werden m unterscheidbare Kugeln auf n Zellen verteilt. Jede Zelle darf mehrfach besetzt werden. Es wird angenommen, dass jede Verteilung der Kugeln gleich wahrscheinlich ist. Wie groß ist die Wahrscheinlichkeit, dass alle Zellen besetzt sind?

Sei A_i das Ereignis, dass die i-te Zelle leer ist. Sei $C = A_1 \cup A_2 \cup \cdots \cup A_n$. Das ist das Ereignis, dass mindestens eine Zelle leer ist. Gesucht ist also $P(C') = 1 - P(C)$. Für $1 \le j \le n$ und $I \in \mathcal{M}_j$ gilt $\mathrm{P}(\bigcap_{i\in I}A_i) = (\frac{n-j}{n})^m$, das ist die Wahrscheinlichkeit, dass die j Zellen, deren Nummer in I vorkommt, leer bleiben. Die Wahrscheinlichkeit, dass alle n Zellen leer bleiben, ist natürlich null. Da die Anzahl der j-elementigen Teilmengen von $\{1, 2, \dots, n\}$ gleich $\binom{n}{j}$ ist, erhalten wir $|\mathcal{M}_j| = \binom{n}{j}$ und $S_j = \binom{n}{j}(\frac{n-j}{n})^m$. Es folgt $P(C) = \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}(\frac{n-j}{n})^m$ und $P(C') = \sum_{j=0}^{n}(-1)^{j}\binom{n}{j}(\frac{n-j}{n})^m$.

Satz 46 (Einschluss-Ausschluss-Formel) *Sei $n \ge 2$. Seien $A_1, A_2, \dots, A_n$ endliche Mengen und $C = A_1 \cup A_2 \cup \cdots \cup A_n$. Dann gilt $|C| = \sum_{j=1}^{n}(-1)^{j-1}S_j$, wobei $S_j = \sum_{I\in\mathcal{M}_j}|\bigcap_{i\in I}A_i|$ und $\mathcal{M}_j$ die Menge aller j-elementigen Teilmengen von $\{1, 2, \dots, n\}$ ist.*

Beweis: Sei Ω eine endliche Menge, die C enthält. Alle Elemente von Ω, also alle Ausfälle, seien gleichwahrscheinlich. Für alle $B \subseteq \Omega$ gilt dann $P(B) = \frac{|B|}{|\Omega|}$. Aus Satz 45 ergibt sich $P(C) = \sum_{j=1}^{n}(-1)^{j-1}\sum_{I\in\mathcal{M}_j}\mathrm{P}(\bigcap_{i\in I}A_i)$, das heißt $\frac{|C|}{|\Omega|} = \sum_{j=1}^{n}(-1)^{j-1}\sum_{I\in\mathcal{M}_j}\frac{|\bigcap_{i\in I}A_i|}{|\Omega|}$. Multipliziert man mit $|\Omega|$, so ergibt sich $|C| = \sum_{j=1}^{n}(-1)^{j-1}\sum_{I\in\mathcal{M}_j}|\bigcap_{i\in I}A_i|$, das ist die gesuchte Formel. $\square$

Beispiel 79: Sei U eine m-elementige Menge und $V = \{1, 2, \dots, n\}$. Wie viele surjektive Abbildungen von U nach V gibt es?

Für $1 \le i \le n$ sei A_i die Menge der Abbildungen $f : U \to V$ mit $f(U) \subseteq V \setminus \{i\}$. Sei $C = A_1 \cup A_2 \cup \cdots \cup A_n$. Das ist die Menge aller nicht surjektiven Abbildungen. Gesucht ist also $n^m - |C|$, da n^m die Anzahl aller Abbildungen von U nach V ist. Für $1 \le j \le n$ und $I \in \mathcal{M}_j$ enthält $\bigcap_{i\in I}A_i$ alle Abbildungen $f : U \to V$ mit $f(U) \subseteq V \setminus I$, sodass $|\bigcap_{i\in I}A_i| = (n-j)^m$ gilt. Da die Anzahl der j-elementigen Teilmengen von $\{1, 2, \dots, n\}$ gleich $\binom{n}{j}$ ist, erhalten wir $|\mathcal{M}_j| = \binom{n}{j}$ und $S_j = \binom{n}{j}(n-j)^m$. Es folgt $|C| = \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}(n-j)^m$, sodass $\sum_{j=0}^{n}(-1)^{j}\binom{n}{j}(n-j)^m$ die Anzahl aller surjektiven Abbildungen von U nach V ist.

Beispiel 80: Wie viele fixpunktfreie Permutationen der Zahlen $1, 2, \dots, n$ gibt es?

Für $1 \le i \le n$ sei A_i die Menge der Permutationen, bei denen i ein Fixpunkt ist, das heißt i auf dem Platz i steht. Sei $C = A_1 \cup A_2 \cup \cdots \cup A_n$. Das ist die Menge aller Permutationen mit mindestens einem Fixpunkt. Gesucht ist also $n! - |C|$, da $n!$ die Anzahl aller Permutationen ist. Für $1 \le j \le n$ und $I \in \mathcal{M}_j$ enthält $\bigcap_{i\in I}A_i$ alle Permutationen, für die alle $i \in I$ Fixpunkte sind, sodass $|\bigcap_{i\in I}A_i| = (n-j)!$ gilt. Da die Anzahl der j-elementigen Teilmengen von $\{1, 2, \dots, n\}$ gleich $\binom{n}{j}$ ist, erhalten wir $S_j = \binom{n}{j}(n-j)!$ und $|C| = \sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}(n-j)!$. Somit ist $\sum_{j=0}^{n}(-1)^{j}\binom{n}{j}(n-j)! = n!\sum_{j=0}^{n}(-1)^{j}\frac{1}{j!}$ die Anzahl der fixpunktfreien Permutationen.

2. Erste Wiederkehr

Dieser und der nächste Abschnitt stammen im Wesentlichen aus dem Buch "Henze: Stochastik für Einsteiger".

Sei $n \geq 2$ fest gewählt. Wir ziehen aus der Menge $\{1, 2, \ldots, n\}$ mit Zurücklegen so lange bis eine Zahl auftritt, die vorher schon einmal gezogen wurde. Sei X_n die Anzahl der Züge bis zu dieser ersten Wiederkehr.

Satz 47: *Es gilt* $P(X_n \leq k) = 1 - \prod_{j=1}^{k-1}(1 - \frac{j}{n})$ *und* $P(X_n = k) = \frac{k-1}{n} \prod_{j=1}^{k-2}(1 - \frac{j}{n})$.

Beweis: Die erste Wiederkehr tritt frühestens nach 2 Zügen und spätestens nach $n+1$ Zügen ein. Daher nimmt die Zufallsvariable X_n die Werte $2, 3, \ldots, n+1$ an. Das Ereignis $X_n \geq k+1$ tritt genau dann ein, wenn bei den ersten k Zügen verschiedene Zahlen gezogen wurden. Es gilt daher $P(X_n \geq k+1) = \frac{n(n-1)\ldots(n-k+1)}{n^k} = \prod_{j=1}^{k-1} \frac{n-j}{n} = \prod_{j=1}^{k-1}(1 - \frac{j}{n})$. Die beiden gesuchten Formeln ergeben sich dann aus $P(X_n \leq k) = 1 - P(X_n \geq k+1)$ und $P(X_n = k) = P(X_n \geq k) - P(X_n \geq k+1)$. □

Sei $t > 0$ fest gewählt. Wir berechnen $\lim_{n\to\infty} P(X_n \leq \sqrt{n}\, t)$. Dazu zwei Hilfssätze

Hilfssatz C: *Für* $0 < x < 1$ *gilt* $e^{-\frac{x}{1-x}} \leq 1 - x \leq e^{-x}$.

Beweis: Es genügt $-\frac{x}{1-x} \leq \log(1-x) \leq -x$ zu zeigen. Wir zeigen, dass die Funktionen $f(x) = \log(1-x) + \frac{x}{1-x}$ und $g(x) = -x - \log(1-x)$ auf $[0, 1)$ nicht negativ sind.
Es gilt $f(0) = 0$ und $f'(x) = -\frac{1}{1-x} + \frac{1}{(1-x)^2} = \frac{x}{(1-x)^2} > 0$ für $x \in (0, 1)$. Also ist f monoton wachsend und somit ≥ 0 auf dem Intervall $[0, 1)$.
Es gilt $g(0) = 0$ und $g'(x) = -1 + \frac{1}{1-x} = \frac{x}{1-x} > 0$ für $x \in (0, 1)$. Also ist auch g monoton wachsend und somit ≥ 0 auf dem Intervall $[0, 1)$. □

Sei k_n die größte ganze Zahl kleiner oder gleich $\sqrt{n}\, t$.

Hilfssatz D: *Es gilt* $\lim_{n\to\infty} \frac{k_n(k_n-1)}{n} = t^2$ *und* $\lim_{n\to\infty} \frac{k_n(k_n+1)}{n-k_n} = t^2$.

Beweis: Seien c und d beliebige Konstante aus $\mathbb{R}$. Dann gilt

$$\lim_{n\to\infty} \frac{(\sqrt{n}\, t+c)(\sqrt{n}\, t+d)}{n} = \lim_{n\to\infty} (t + \tfrac{c}{\sqrt{n}})(t + \tfrac{d}{\sqrt{n}}) = t^2$$

Nach Definition von k_n gilt $k_n \leq \sqrt{n}\, t < k_n + 1$ und daher auch $\sqrt{n}\, t - 1 < k_n \leq \sqrt{n}\, t$. Für große n folgt $\frac{(\sqrt{n}\, t-1)(\sqrt{n}\, t-2)}{n} \leq \frac{k_n(k_n-1)}{n} \leq \frac{k_n(k_n+1)}{n-k_n} \leq \frac{\sqrt{n}\, t(\sqrt{n}\, t+1)}{n} \frac{n}{n-\sqrt{n}\, t}$. Nach obiger Rechnung gilt $\lim_{n\to\infty} \frac{(\sqrt{n}\, t-1)(\sqrt{n}\, t-2)}{n} = t^2$ und $\lim_{n\to\infty} \frac{\sqrt{n}\, t(\sqrt{n}\, t+1)}{n} = t^2$. Weiters gilt $\lim_{n\to\infty} \frac{n}{n-\sqrt{n}\, t} = \lim_{n\to\infty} \frac{1}{1-t/\sqrt{n}} = 1$ und somit $\lim_{n\to\infty} \frac{\sqrt{n}\, t(\sqrt{n}\, t+1)}{n} \frac{n}{n-\sqrt{n}\, t} = t^2$. Daher muss sowohl $\lim_{n\to\infty} \frac{k_n(k_n-1)}{n} = t^2$ als auch $\lim_{n\to\infty} \frac{k_n(k_n+1)}{n-k_n} = t^2$ gelten. □

Satz 48: *Für alle* $t > 0$ *gilt* $\lim_{n\to\infty} P(X_n \leq \sqrt{n}\, t) = 1 - e^{-\frac{1}{2}t^2}$.

Beweis: Nach Definition von k_n gilt $k_n \leq \sqrt{n}\, t < k_n + 1$. Somit gilt auch

$$P(X_n \leq k_n) \leq P(X_n \leq \sqrt{n}\, t) \leq P(X_n \leq k_n + 1)$$

Da wir n gegen ∞ gehen lassen, können wir annehmen, dass $n > \max(t^2, \frac{4}{t^2})$ gilt.
Wegen $n > \frac{4}{t^2}$ gilt $\sqrt{n}\, t > 2$ und daher haben wir auch $k_n \geq 2$. Wegen $n > t^2$ ergibt sich $k_n \leq \sqrt{n}\, t < n$ und daher auch $0 < \frac{j}{n} < 1$ für $1 \leq j \leq k_n$. Aus Hilfssatz C erhalten

wir dann $e^{-\frac{j}{n-j}} \leq 1-\frac{j}{n} \leq e^{-\frac{j}{n}}$, also auch $e^{-\frac{j}{n-k_n}} \leq 1-\frac{j}{n} \leq e^{-\frac{j}{n}}$ für $1 \leq j \leq k_n$. Produktbildung ergibt $e^{-\frac{1+2+\cdots+k_n}{n-k_n}} \leq \prod_{j=1}^{k_n}(1-\frac{j}{n})$ und $\prod_{j=1}^{k_n-1}(1-\frac{j}{n}) \leq e^{-\frac{1+2+\cdots+(k_n-1)}{n}}$. Es folgt $1-e^{-\frac{1}{2}\frac{k_n(k_n+1)}{n-k_n}} \geq P(X_n \leq k_n+1)$ und $P(X_n \leq k_n) \geq 1-e^{-\frac{1}{2}\frac{k_n(k_n-1)}{n}}$ mit Hilfe von Satz 47. Aus obiger Ungleichung erhalten wir

$$1-e^{-\frac{1}{2}\frac{k_n(k_n-1)}{n}} \leq P(X_n \leq \sqrt{n}\,t) \leq 1-e^{-\frac{1}{2}\frac{k_n(k_n+1)}{n-k_n}}$$

Aus Hilfssatz D folgt $\lim_{n\to\infty} e^{-\frac{1}{2}\frac{k_n(k_n-1)}{n}} = e^{-\frac{1}{2}t^2}$ und $\lim_{n\to\infty} e^{-\frac{1}{2}\frac{k_n(k_n+1)}{n-k_n}} = e^{-\frac{1}{2}t^2}$. Daraus ergibt sich dann $\lim_{n\to\infty} P(X_n \leq \sqrt{n}\,t) = 1-e^{-\frac{1}{2}t^2}$. □

Beispiel 81: Nach wie vielen Ziehungen beim Lotto 6 aus 45 kann man erwarten, das ein Ziehungsergebnis wieder auftritt?

Beim Lotto 6 aus 45 gibt es $\binom{45}{6} = 8145060$ mögliche Ziehungsergebnisse. Sei X die Anzahl der Ziehungen, bis eine zuvor bereits gezogene Zahlenreihe wieder auftritt. Das ist die Frage nach der ersten Wiederkehr bei $n = 8145060$ möglichen Ziehungsergebnissen. Wegen Satz 48 gilt $P(X_n \leq \sqrt{n}\,t) \approx 1-e^{-\frac{1}{2}t^2}$, wenn n nur groß genug ist. Wir haben $n = 8145060$, das sollte groß genug sein.

Wir setzen $t = \sqrt{2\log 2}$. Dann gilt $1-e^{-\frac{1}{2}t^2} = \frac{1}{2}$. Es folgt $P(X_n \leq \sqrt{2n\log 2}) \approx \frac{1}{2}$. Für $n = 8145060$ erhalten wir $\sqrt{2n\log 2} = 3360{,}28$. Somit gilt $P(X \leq 3360) \approx \frac{1}{2}$. Mit Wahrscheinlichkeit $\frac{1}{2}$ tritt unter den ersten 3360 Ziehungen eine Wiederkehr auf.

Sei $t = \sqrt{-2\log 0.05}$. Dann gilt $1-e^{-\frac{1}{2}t^2} = 0.95$. Es folgt $P(X_n \leq \sqrt{-2n\log 0.05}) \approx 0.95$. Für $n = 8145060$ erhalten wir $\sqrt{-2n\log 0.05} = 6985{,}76$. Somit gilt $P(X \leq 6985) \approx 0.95$. Mit Wahrscheinlichkeit 0.95 tritt unter den ersten 6985 Ziehungen eine Wiederkehr auf.

Will man eine exakte Abschätzung haben, dann kann man die im Beweis von Satz 48 gefundenen Ungleichungen $1-e^{-\frac{1}{2}\frac{k_n(k_n-1)}{n}} \leq P(X_n \leq \sqrt{n}\,t) \leq 1-e^{-\frac{1}{2}\frac{k_n(k_n+1)}{n-k_n}}$ verwenden, wobei k_n die größte ganze Zahl kleiner oder gleich $\sqrt{n}\,t$ ist. Für $n = 8145060$ und $t = \sqrt{2\log 2}$ hatten wir $\sqrt{n}\,t = 3360{,}28$. Somit haben wir $k_n = 3360$. Damit ergibt sich $1-e^{-\frac{1}{2}\frac{k_n(k_n-1)}{n}} = 0.49984$ und $1-e^{-\frac{1}{2}\frac{k_n(k_n+1)}{n-k_n}} = 0.50019$. Wir erhalten daraus die Abschätzung $0.49984 \leq P(X \leq 3360) \leq 0.50019$. Man sieht, dass obige Approximation sehr gut ist.

3. Ein Sammlerproblem

Aus der Menge $M = \{1, 2, \ldots, n\}$ wird so lange mit Zurücklegen gezogen, bis jede Zahl mindestens einmal gezogen wurde. Sei X die Anzahl der dafür benötigten Züge. So lange dauert es, bis man alle Elemente der Menge M gesammelt hat.

Satz 49: *Es gilt* $\mathrm{E}(X) = n\,(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n})$ *und* $\mathrm{V}(X) = \sum_{j=1}^{n-1}\frac{n(n-j)}{j^2}$.

Beweis: Sei $Z_1 = 1$ die Anzahl der Züge bis eine Zahl m_1 aus M auftritt. Das geschieht natürlich gleich beim ersten Zug. Es folgt $\mathrm{E}(Z_1) = 1$ und $\mathrm{V}(Z_1) = 0$.

Nachdem m_1 aufgetreten ist, ziehen wir so lange mit Zurücklegen bis eine Zahl m_2 aus $M \setminus \{m_1\}$ auftritt. Sei Z_2 die Anzahl der Züge. Dann ist Z_2 von Z_1 unabhängig und hat die $G(p)$-Verteilung mit $p = \frac{n-1}{n}$. Es folgt $\mathrm{E}(Z_2) = \frac{1}{p} = \frac{n}{n-1}$ und $\mathrm{V}(Z_2) = \frac{1-p}{p^2} = \frac{n}{(n-1)^2}$.

So geht das weiter. Sei Z_3 die Anzahl der Züge bis eine Zahl m_3 aus $M \setminus \{m_1, m_2\}$ auftritt. Dann ist Z_3 von Z_2 und Z_1 unabhängig und hat die $G(p)$-Verteilung mit $p = \frac{n-2}{n}$. Es folgt $\mathrm{E}(Z_3) = \frac{1}{p} = \frac{n}{n-2}$ und $\mathrm{V}(Z_3) = \frac{1-p}{p^2} = \frac{2n}{(n-2)^2}$.

Es endet mit Z_n, der Anzahl der Züge bis die einzige Zahl m_n aus $B \setminus \{m_1, m_2, \dots, m_{n-1}\}$ auftritt. Wieder ist Z_n von $Z_1, \dots, Z_{n-1}$ unabhängig und $G(p)$-verteilt mit $p = \frac{1}{n}$. Es folgt $\mathrm{E}(Z_n) = \frac{1}{p} = \frac{n}{1}$ und $\mathrm{V}(Z_b) = \frac{1-p}{p^2} = \frac{n(n-1)}{1}$.
Für die Wartezeit X, bis alle Zahlen aus M aufgetreten sind, gilt $X = Z_1 + Z_2 + \dots + Z_n$. Es folgt $\mathrm{E}(X) = \mathrm{E}(Z_1) + \mathrm{E}(Z_2) + \dots + \mathrm{E}(Z_n) = n(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{2} + 1)$, da der Erwartungswert linear ist. Die Zufallsvariablen sind unabhängig, also ist auch die Varianz linear. Es gilt $\mathrm{V}(X) = \mathrm{V}(Z_1) + \mathrm{V}(Z_2) + \dots + \mathrm{V}(Z_n) = n(\frac{1}{(n-1)^2} + \dots + \frac{n-2}{2^2} + \frac{n-1}{1^2}) = n \sum_{j=1}^{n} \frac{n-j}{j^2}$. □

Beispiel 82: Man würfelt so lange, bis jede Augenzahl einmal aufgetreten ist. Wie oft muss man durchschnittlich würfeln?

Wir ziehen aus der Menge $M = \{1, 2, 3, 4, 5, 6\}$ so lange mit Zurücklegen, bis jedes Element einmal aufgetreten ist. Aus Satz 49 mit $n = 6$ folgt $\mathrm{E}(X) = 6+3+2+\frac{3}{2}+\frac{6}{5}+1 = 14.7$ und $\mathrm{V}(X) = \sum_{i=1}^{5} \frac{6(6-i)}{i^2} = 6(5 + 1 + \frac{1}{3} + \frac{1}{8} + \frac{1}{25}) = 38.99$, sodass 6.24 die Standardabweichung von X ist.

Wir verallgemeinern das Sammlerproblem. Sei $2 \le s < n$. Bei jedem Zug wird nicht eine, sondern s Zahlen aus der Menge $M = \{1, 2, \dots, n\}$ gezogen und diese dann wieder zurückgelegt. Es wird so lange gezogen, bis jede Zahl mindestens einmal aufgetreten ist. Sei X die Anzahl der dafür benötigten Züge.

Satz 50: *Sei* $q_j = \binom{n-j}{s} / \binom{n}{s}$. *Für* $k \ge 2$ *gilt dann* $P(X \ge k) = \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} q_j^{k-1}$ *und* $P(X = k) = \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} (1 - q_j) q_j^{k-1}$. *Weiters gilt für den Erwartungswert* $\mathrm{E}(X) = \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} \frac{2q_j - q_j^2}{1-q_j}$.

Beweis: Sei W_i die Anzahl der Züge, bis die Zahl i zum ersten Mal auftritt. Dann gilt $X = \max(W_1, W_2, \dots, W_n)$. Sei $k \ge 2$ fest gewählt. Sei A_i das Ereignis $W_i \ge k$. Das Ereignis $X \ge k$ tritt genau dann ein, wenn mindestens eines der Ereignisse $A_1, A_2, \dots, A_n$ eintritt, das heißt wenn $\bigcup_{i=1}^{n} A_i$ eintritt. Somit gilt $P(X \ge k) = P(\bigcup_{i=1}^{n} A_i)$. Es liegt nahe, Satz 45 zu verwenden.
Sei $\mathcal{M}_j$ die Menge aller j-elementigen Teilmengen von $\{1, 2, \dots, n\}$. Für $I \in \mathcal{M}_j$ müssen wir $\mathrm{P}(\bigcap_{i \in I} A_i)$ berechnen. Das Ereignis $\bigcap_{i \in I} A_i$ tritt genau dann ein, wenn bei den ersten $k - 1$ Zügen keine Zahl aus I gezogen wurde. Da mit Zurücklegen gezogen wird, haben wir $\mathrm{P}(\bigcap_{i \in I} A_i) = q_j^{k-1}$, wobei q_j die Wahrscheinlichkeit ist bei einem Zug keine Zahl aus I zu ziehen.
Da wir bei einem Zug s Elemente aus der Menge $\{1, 2, \dots, n\}$ ziehen, gibt es $\binom{n}{s}$ mögliche Ziehungsergebnisse. Da I eine j-elementige Teilmenge von $\{1, 2, \dots, n\}$ ist, gibt es $\binom{n-j}{s}$ mögliche Ziehungsergebnisse, die keine Zahl aus I enthalten. Da alle Ziehungsergebnisse gleich wahrscheinlich sind, erhalten wir $q_j = \binom{n-j}{s} / \binom{n}{s}$, wenn $n - j \ge s$ ist. Ist $n - j < s$, dann gilt $q_j = 0$, da weniger als s Zahlen ausserhalb von I liegen.
Nun ergibt sich $S_j = \sum_{I \in \mathcal{M}_j} \mathrm{P}(\bigcap_{i \in I} A_i) = \sum_{I \in \mathcal{M}_j} q_j^{k-1} = |\mathcal{M}_j| q_j^{k-1} = \binom{n}{j} q_j^{k-1}$. Nach Satz 45 gilt dann $\mathrm{P}(\bigcup_{i=1}^{n} A_i) = \sum_{j=1}^{n} (-1)^{j-1} S_j$. Setzen wir ein, so erhalten wir

$$P(X \ge k) = \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} q_j^{k-1}$$

Als obere Summationsgrenze schreiben wir $n - s$, da für $j > n - s$ ja $q_j = 0$ gilt. Es folgt

$$\begin{aligned} P(X = k) = P(X \ge k) - P(X \ge k+1) &= \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} (q_j^{k-1} - q_j^k) \\ &= \sum_{j=1}^{n-s} (-1)^{j-1} \binom{n}{j} (1 - q_j) q_j^{k-1} \end{aligned}$$

Damit sind die ersten beiden Gleichungen bereits gezeigt.

Nach Definition des Erwartungswerts gilt $\mathrm{E}(X) = \sum_{k=1}^{\infty} k\,P(X = k)$. Wegen $s < n$ kann beim ersten Zug nicht jede Zahl aus der Menge $M = \{1, 2, \ldots, n\}$ auftreten. Es gilt somit $P(X = 1) = 0$ und $\mathrm{E}(X) = \sum_{k=2}^{\infty} k\,P(X = k)$. Setzt man für $P(X = k)$ ein und vertauscht die Summen, so erhält man $\mathrm{E}(X) = \sum_{j=1}^{n-s}(-1)^{j-1}\binom{n}{j}(1 - q_j)\sum_{k=2}^{\infty} kq_j^{k-1}$. Differenziert man die Gleichung $\sum_{k=2}^{\infty} x^k = \frac{x^2}{1-x}$, dann ergibt sich $\sum_{k=2}^{\infty} kx^{k-1} = \frac{2x-x^2}{(1-x)^2}$. Somit ist $\sum_{k=2}^{\infty} kq_j^{k-1} = \frac{2q_j-q_j^2}{(1-q_j)^2}$ gezeigt. Setzt man das oben in die Formel für $\mathrm{E}(X)$ ein, dann erhält man $\mathrm{E}(X) = \sum_{j=1}^{n-s}(-1)^{j-1}\binom{n}{j}\frac{2q_j-q_j^2}{1-q_j}$, das gesuchte Resultat. $\square$

Bemerkung: Aus Satz 50 kann man eine kombinatorische Identität gewinnen. Sei a die kleinste ganze Zahl, die größer oder gleich $\frac{n}{s}$ ist. Es sind mindestens a Züge notwendig, um jede Zahl aus der Menge $M = \{1, 2, \ldots, n\}$ mindestens einmal zu ziehen. Daher muss $P(X \geq k) = 1$ gelten, wenn $k \leq a$ ist. Setzt man für $P(X \geq k)$ gemäß Satz 50 ein, so erhält man, dass $\sum_{j=1}^{n-s}(-1)^{j-1}\binom{n}{j}\binom{n-j}{s}^{k-1} = \binom{n}{s}^{k-1}$ für $2 \leq k \leq a$ gilt.

Beispiel 83: Nach wie vielen Ziehungen beim Lotto 6 aus 45 kann man erwarten, dass jede Zahl mindestens einmal gezogen wurde?

Bei einer Ziehung werden 6 Zahlen aus der Menge der ersten 45 Zahlen gezogen. Es gilt also $n = 45$ und $s = 6$. Sei X die Anzahl der Ziehungen, bis jede der 45 Zahlen einmal aufgetreten ist. Nach Satz 50 gilt $\mathrm{E}(X) = \sum_{j=1}^{39}(-1)^{j-1}\binom{45}{j}\frac{2q_j-q_j^2}{1-q_j}$, wobei jetzt $q_j = \binom{45-j}{6}/\binom{45}{6}$ zu wählen ist. Mit Hilfe eines Computers erhält man $\mathrm{E}(X) = 31.5$.

Nach Satz 50 gilt auch $P(X \geq k) = \sum_{j=1}^{39}(-1)^{j-1}\binom{45}{j}q_j^{k-1}$. Das kann man für verschiedene Werte von k ausrechnen. Man erhält $P(X \geq 18) = 0.99$ und $P(X \geq 60) = 0.01$. Daher liegt die Anzahl der Ziehungen, die man abwarten muss, bis alle Zahlen zumindest einmal gezogen wurden, mit sehr großer Wahrscheinlichkeit zwischen 18 und 60. Es gilt auch $P(X \geq 30) = 0.52$ und $P(X \geq 31) = 0.47$. Die Wahrscheinlichkeit, dass man weniger als 30 Ziehungen warten muss, ist ungefähr $\frac{1}{2}$.

Beispiel 84: Es gibt 200 Bilder zum Sammeln. Sie sind in Packungen zu 5 Stück erhältlich. Nach wie vielen Packungen kann man erwarten, dass man alle Bilder hat?

Es gilt also $n = 200$ und $s = 5$. Sei X die Anzahl der Packungen, die man erwerben muss, bis man alle Bilder gesammelt hat. Nach Satz 50 gilt $\mathrm{E}(X) = \sum_{j=1}^{195}(-1)^{j-1}\binom{200}{j}\frac{2q_j-q_j^2}{1-q_j}$, wobei $q_j = \binom{200-j}{5}/\binom{200}{5}$ zu wählen ist. Lässt man das von einem Computer ausrechnen, so erhält man $\mathrm{E}(X) = 233.2$.

Nach Satz 50 gilt $P(X \geq k) = \sum_{j=1}^{195}(-1)^{j-1}\binom{200}{j}q_j^{k-1}$. Das kann man für verschiedene Werte von k ausrechnen. Man erhält $P(X \geq 152) = 0.99$ und $P(X \geq 392) = 0.01$. Daher liegt die Anzahl der Packungen, die man erwerben muss, bis man alle Bilder hat, mit sehr großer Wahrscheinlichkeit zwischen 152 und 392. Es gilt auch $P(X \geq 225) = 0.502$. Die Wahrscheinlichkeit, dass man weniger als 225 Packungen erwerben muss, ist ungefähr $\frac{1}{2}$.

4. Das Hardy–Weinberg–Gesetz (Populationsgenetik)

Gene sind auf den Chromosomenfäden angeordnet. Die verschiedenen Zustände eines Gens heißen Allele. Chromosomenfäden treten paarweise auf. Einen erbt man von der Mutter und einen vom Vater.

von der Mutter

vom Vater

Wir betrachten einen Genort am oberen Chromosom. Der Zustand (das Allel) des dortigen Gens wurde mit A bezeichnet. Am entsprechenden Ort des unteren Chromosoms befindet sich dasselbe Gen, eventuell in einem anderen Zustand a.

Wenn ein Gen zwei Allele (Zustände) A und a hat, so gibt es bezüglich dieses Gens drei Genotypen, nämlich AA, Aa und aa. Die Genotypen AA und aa heißen homozygot. Der Genotyp Aa heißt heterozygot. Wenn AA–Individuen und Aa–Individuen gleich aussehen, dann wird das merkmalbestimmende Allel A dominant und das unterlegene Allel a rezessiv genannt. Man muss also den Genotyp, das heißt die Erbanlagen, vom Phänotyp, den sichtbaren Merkmalen, unterscheiden.

Was geschieht bei der Vererbung? Wir betrachten ein Gen mit zwei Allelen A und a. Es gibt die Genotypen AA, Aa und aa. Daraus ergeben sich folgende Kreuzungsmöglichkeiten:

$AA \times AA \quad AA \times Aa \quad Aa \times AA \quad AA \times aa \quad aa \times AA \quad Aa \times Aa \quad Aa \times aa \quad aa \times Aa \quad aa \times aa$

Seien u, v und w die Häufigkeiten der drei Genotypen AA, Aa und aa in einer Bevölkerung, sodass $u + v + w = 1$ gilt. Wir berechnen diese Häufigkeiten für die nächste Generation, wenn die Eltern eines Nachkommen zufällig gemäß diesen Häufigkeiten gewählt werden. Der Nachkomme erhält zufällig ein Allel von jedem Elternteil. Das gibt folgende Tabelle:

Genotypen der Eltern	Häufigkeit dieser Eltern	Häufigkeit der Nachkommen AA	Aa	aa
$AA \times AA$	u^2	u^2	0	0
$AA \times Aa$	uv	$\frac{1}{2}uv$	$\frac{1}{2}uv$	0
$Aa \times AA$	vu	$\frac{1}{2}vu$	$\frac{1}{2}vu$	0
$Aa \times Aa$	v^2	$\frac{1}{4}v^2$	$\frac{1}{2}v^2$	$\frac{1}{4}v^2$
$AA \times aa$	uw	0	uw	0
$aa \times AA$	wu	0	wu	0
$Aa \times aa$	vw	0	$\frac{1}{2}vw$	$\frac{1}{2}vw$
$aa \times Aa$	wv	0	$\frac{1}{2}wv$	$\frac{1}{2}wv$
$aa \times aa$	w^2	0	0	w^2

Es werden zufällig zwei Eltern ausgewählt. Die Wahrscheinlichkeit, dass beide vom Typ AA sind, ist u^2. Der Nachkomme ist dann immer vom Typ AA. Das ergibt die erste Zeile der Tabelle. Die Wahrscheinlichkeit, dass die Mutter vom Typ AA und der Vater vom Typ Aa ist, ist uv. Der Nachkomme erbt von der Mutter das Allel A und vom Vater das Allel A oder das Allel a jeweils mit Wahrscheinlichkeit $\frac{1}{2}$. Daher ist der Nachkomme vom Typ AA oder Aa jeweils mit Wahrscheinlichkeit $\frac{1}{2}uv$. Das ist die zweite Zeile der Tabelle.

Ebenso erhält man die anderen Zeilen der Tabelle.

Die Häufigkeiten der Genotypen AA, Aa und aa in der ersten Generation sind daher

$$\begin{aligned}
u_1 &= u^2 + \tfrac{1}{2}uv + \tfrac{1}{2}vu + \tfrac{1}{4}v^2 = (u + \tfrac{v}{2})^2 \\
v_1 &= \tfrac{1}{2}uv + \tfrac{1}{2}vu + uw + wu + \tfrac{1}{2}v^2 + \tfrac{1}{2}vw + \tfrac{1}{2}wv = 2(u + \tfrac{v}{2})(\tfrac{v}{2} + w) \\
w_1 &= \tfrac{1}{4}v^2 + \tfrac{1}{2}vw + \tfrac{1}{2}wv + w^2 = (\tfrac{v}{2} + w)^2
\end{aligned}$$

Sei $p = u + \frac{v}{2}$ und $q = \frac{v}{2} + w$, sodass $p + q = 1$ gilt. Man kann p als die Häufigkeit des Allels A und q als die Häufigkeit des Allels a interpretieren. Es folgt dann $u_1 = p^2$, $v_1 = 2pq$ und $w_1 = q^2$. Es gilt $v_1^2 = 4u_1w_1$.

Nachdem damit die Häufigkeiten der einzelnen Genotypen in der ersten Generation berechnet sind, kann man diese als Elterngeneration auffassen und auf dieselbe Art die Häufigkeiten der Genotypen in der zweiten Generation berechnen:

$$\begin{aligned}
u_2 &= (u_1 + \tfrac{v_1}{2})^2 = (p^2 + pq)^2 = p^2 = u_1 \\
v_2 &= 2(u_1 + \tfrac{v_1}{2})(\tfrac{v_1}{2} + w_1) = 2(p^2 + pq)(pq + q^2) = 2pq = v_1 \\
w_2 &= (\tfrac{v_1}{2} + w_1)^2 = (pq + q^2)^2 = q^2 = w_1
\end{aligned}$$

Man sieht, dass sich die Häufigkeiten der Genotypen ab der ersten Generation nicht mehr ändern. Das nennt man das Hardy–Weinberg–Gesetz. Man sagt, eine Bevölkerung ist im Hardy–Weinberg–Gleichgewicht, wenn für die Häufigkeiten u, v und w der Genotypen AA, Aa und aa die Gleichung $v^2 = 4uw$ erfüllt ist.

Das Gen, das die Blutgruppe bestimmt, hat drei Allele. Man bezeichnet sie mit A, B und 0. Das ergibt sechs mögliche Genotypen. Der Genotyp bestimmt die Blutgruppe.

Genotyp	AA	$A0$	AB	BB	$B0$	00
Blutgruppe	A	A	AB	B	B	0

Genau auf dieselbe Art wie für zwei Allele kann man das Hardy–Weinberggesetz auch für drei Allele herleiten. Nach einer Generation ändern sich die Häufigkeiten der Genotypen nicht mehr. Sind p, q und r die Häufigkeiten, mit denen die Allele A, B und 0 in der Bevölkerung vorhanden sind, dann sind die Häufigkeiten der Genotypen gegeben durch

Genotyp	AA	BB	00	$A0$	$B0$	AB
Häufigkeit	p^2	q^2	r^2	$2pr$	$2qr$	$2pq$

Daraus ergeben sich dann die Häufigkeiten der einzelnen Blutgruppen, die sich ebenfalls nicht mehr im Laufe der Generationen verändern.

Blutgruppe	A	B	AB	0
Häufigkeit	$p^2 + 2pr$	$q^2 + 2qr$	$2pq$	r^2

Kennt man die Häufigkeiten der Blutgruppen einer Bevölkerung, so kann man mit Hilfe dieser Formeln die Häufigkeiten der Allele berechnen.

5. Information und Entropie

Es soll der Informationsgehalt einer Sprache, zum Beispiel der deutschen, gemessen werden. Wieviel Information pro Buchstabe gewinnt man beim Lesen eines deutschen Textes? Dazu wird der Text als Nachricht in den beiden Binärzeichen 0 und 1 codiert. Wir machen das so, dass die codierte Nachricht möglichst kurz wird. Die enthaltene Information wird also möglichst dicht verpackt. Die durchschnittliche Anzahl von Binärzeichen, die man dabei zum Codieren eines Buchstaben braucht, nennt man den Informationsgehalt oder die Entropie der Sprache. Sie wird in Binärzeichen (bit) pro Buchstaben gemessen.

Natürlich darf beim Codieren keine Information verlorengehen. Es muss möglich sein, aus dem codierten Text den ursprünglichen Text wiederzugewinnen. Eine Codierung ordnet jedem Buchstaben eine 0-1-Folge zu. Damit man wieder zurückübersetzen kann, darf keine dieser Folgen ein Anfangsstück einer anderen sein. Besteht das Alphabet aus den sieben Buchstaben a, b, c, d, e, f und g, dann könnte man zum Beispiel folgende Codierung wählen:

a	1 1 0	0 0	e
b	1 1 1 0	0 1	g
c	1 0 1	1 0 0	f
d	1 1 1 1	1 0 1	c
e	0 0	1 1 0	a
f	1 0 0	1 1 1 0	b
g	0 1	1 1 1 1	d

Die linke Tabelle verwendet man, um einen Text zu codieren. Rechts steht dieselbe Tabelle, jedoch zum Zurückübersetzen. Das funktioniert, weil keine der gewählten Codefolgen ein Anfangsstück einer anderen ist. Ist 01100101010111010111101 die Übersetzung einer Buchstabenfolge, so gewinnt man diese mit Hilfe der rechten Tabelle zurück: $gfcggadg$.

Die einzelnen Buchstaben einer Sprache treten nicht mit gleichen Wahrscheinlichkeiten (relativen Häufigkeiten) auf. Die codierten Nachrichten werden dann möglichst kurz sein, wenn man den Buchstaben, die häufiger auftreten, kürzere Codefolgen, und denen, die seltener auftreten, längere Codefolgen zuordnet. Sei A das Alphabet, die Menge aller Buchstaben einer Sprache. Ist eine Codierung gegeben, dann sei $L(i)$ die Länge des Codes für den Buchstaben i. In obigem Beispiel ist $L(a) = 3$, $L(b) = 4$ und so weiter.

Satz 51: *Wenn keine Codefolge ein Anfangsstück einer anderen Codefolge ist, dann gilt* $\sum_{i\in A} 2^{-L(i)} \le 1$. *Sind umgekehrt Zahlen* $L(i) \in \mathbb{N}$ *für alle* $i \in A$ *vorgegeben, die diese Ungleichung erfüllen, dann gibt es eine Codierung mit diesen Codelängen, sodass keine Codefolge ein Anfangsstück einer anderen ist.*

Beweis: Wir zeigen zuerst den ersten Teil des Satzes. Sei $N = \max\{L(i) : i \in A\}$. Die Codefolge des Buchstaben i der vorliegenden Codierung lässt sich auf $2^{N-L(i)}$ Arten zu einer 0-1-Folge der Länge N erweitern. Tut man das für alle $i \in A$, so erhält man lauter verschiedene 0-1-Folgen der Länge N. Da es insgesamt höchstens 2^N solche Folgen gibt, erhalten wir $\sum_{i\in A} 2^{N-L(i)} \le 2^N$ oder $\sum_{i\in A} 2^{-L(i)} \le 1$.

Sind $L(i)$ vorgegeben, die $\sum_{i\in A} 2^{-L(i)} \le 1$ erfüllen, so sei wieder $N = \max\{L(i) : i \in A\}$. Es gilt $\sum_{i\in A} 2^{N-L(i)} \le 2^N$. Seien $i_1, i_2, \ldots \in A$ so nummeriert, dass $L(i_1) \le L(i_2) \le \ldots$ gilt. Wir schreiben alle 0-1-Folgen der Länge N untereinander, und zwar lexicographisch geordnet. Das gemeinsame Anfangsstück der ersten $2^{N-L(i_1)}$ Folgen (es hat Länge $L(i_1)$) erklären wir zur Codefolge des Buchstaben i_1. Wegen $L(i_2) \geq L(i_1)$ haben die nächsten $2^{N-L(i_2)}$ Folgen ein gemeinsames Anfangsstück der Länge $L(i_2)$. Dieses erklären wir zur Codefolge des Buchstaben i_2. So kann man fortsetzen. Wegen $\sum_{i\in A} 2^{N-L(i)} \leq 2^N$ hat man genug 0-1-Folgen für alle Buchstaben in A. Gilt echte Ungleichung, dann bleiben sogar 0-1-Folgen übrig. □

In obigem Beispiel kann man leicht nachprüfen, dass die Ungleichung aus dem letzten Satz erfüllt ist.

Die durchschnittliche Codelänge ist $\sum_{i\in A} p_i L(i)$. Es wird mit den relativen Häufigkeiten p_i der einzelnen Buchstaben gewichtet. Jetzt geht es darum eine Codierung mit minimaler durchschnittlicher Codelänge zu finden.

Satz 52: *Sei p_i die Wahrscheinlichkeit mit der der Buchstabe i auftritt, sodass $p_i > 0$ und $\sum_{i\in A} p_i = 1$ gilt. Unter der Bedingung $\sum_{i\in A} 2^{-L(i)} \leq 1$ nimmt $\sum_{i\in A} p_i L(i)$ das Minimum für $L(i) = -\log p_i$ an, wobei* log *der Logarithmus zur Basis* 2 *ist.*

Beweis: Es gilt $\log x \leq \frac{1}{\ln 2}(x-1)$ für alle $x > 0$, wobei Gleichheit genau dann gilt, wenn $x = 1$ ist. Nun folgt

$$\begin{aligned}\sum_{i\in A} p_i(-\log p_i) - \sum_{i\in A} p_i L(i) &= \sum_{i\in A} p_i \log \frac{2^{-L(i)}}{p_i} \leq \sum_{i\in A} p_i \frac{1}{\ln 2}\Big(\frac{2^{-L(i)}}{p_i} - 1\Big) \\ &= \frac{1}{\ln 2}\Big(\sum_{i\in A} 2^{-L(i)} - \sum_{i\in A} p_i\Big) \leq 0\end{aligned}$$

da $\sum_{i\in A} 2^{-L(i)} \leq 1$ und $\sum_{i\in A} p_i = 1$ ist. Damit ist $\sum_{i\in A} p_i(-\log p_i) \leq \sum_{i\in A} p_i L(i)$ gezeigt. Gleichheit kann nur dann gelten, wenn $\frac{2^{-L(i)}}{p_i} = 1$, das heißt $L(i) = -\log p_i$ für alle $i \in A$ gilt. In diesem Fall ist auch die Bedingung $\sum_{i\in A} 2^{-L(i)} \leq 1$ erfüllt. □

Damit haben wir das zu Beginn gestellte Problem gelöst. Wenn p_i die Buchstabenhäufigkeiten sind, dann ist $-\sum_{i\in A} p_i \log p_i$ die minmale mögliche durchschnittliche Codelänge. Sie wird als Informationsgehalt oder Entropie der Sprache bezeichnet. Allerdings ist diese durchschnittliche Codelänge nur dann realisierbar, wenn $p_i = 2^{-n_i}$ mit $n_i \in \mathbb{N}$ für alle Buchstaben $i \in A$ gilt. Dann wählt man $L(i) = n_i$. Das obige Beispiel ist eine in diesem Sinn optimale Codierung, wenn $p_e = p_g = \frac{1}{4}$, $p_a = p_c = p_f = \frac{1}{8}$ und $p_b = p_d = \frac{1}{16}$ gilt.

Ist $-\log p_i$ nicht ganzzahlig, so kann man $L(i)$ als die eindeutig bestimmte ganze Zahl im Intervall $[-\log p_i, -\log p_i + 1)$ wählen. Dann gilt $\sum_{i\in A} 2^{-L(i)} \leq 1$ und $\sum_{i\in A} p_i L(i) \leq -\sum_{i\in A} p_i \log p_i + 1$. Die minimale mögliche durchschnittliche Codelänge wird höchstens um 1 überschritten.

VII. Anhang: Vorkenntnisse aus der Analysis

Es werden einige Resultate aus der Analysis zusammengestellt, die in der Stochastik verwendet werden. Dabei wird nur die Analysis in einer Variablen vorausgesetzt.

1. Reihen

In der Analysis wird zu einer Folge $u_0, u_1, u_2, u_3, \dots$ von reellen Zahlen die Reihe $\sum_{j=0}^{\infty} u_j$ als Grenzwert der Partialsummen $\sum_{j=0}^{n} u_j$ für $n \to \infty$ definiert. In der Wahrscheinlichkeitstheorie hat man Summen $\sum_{i \in S} r_i$, wobei S eine endliche oder abzählbare Menge ist. Ist S endlich, dann ist diese Summe eindeutig, da man ja in beliebiger Reihenfolge addieren kann. Für abzählbares S muss man eine Anordnung der Reihenglieder wählen, um die Summe $\sum_{i \in S} r_i$ als Grenzwert der Partialsummen berechnen zu können. Satz 53 und Satz 54 zeigen, unter welchen Voraussetzungen jede Anordnung der Reihenglieder dieselbe Summe ergibt. Satz 55 bringt dann einige einfache Rechenregeln.

Satz 53: *Sei S eine abzählbare Menge und $r_i \geq 0$ für alle $i \in S$. Dann hat $\sum_{i \in S} r_i$ immer denselben Wert W, wie immer man die Reihe anordnet. Es ist jedoch auch $W = \infty$ möglich. Weiters gilt $\sum_{i \in E} r_i \leq W$ für jede endliche Teilmenge E von S. Ist $W < \infty$, dann existiert für jedes $\varepsilon > 0$ eine endliche Teilmenge U von S mit $\sum_{i \in U} r_i > W - \varepsilon$.*

Beweis: Sei $a_0, a_1, a_2, a_3, \dots$ irgendeine Anordnung der Zahlen r_i für $i \in S$. Zahlen, die öfter vorkommen, müssen natürlich auch in der Anordnung mit derselben Häufigkeit auftreten. Sei $A_n = \sum_{j=0}^{n} a_j$ die n-te Partialsumme. Da $a_j \geq 0$ für alle $j \geq 0$ gilt, ist die Folge $(A_n)_{n \geq 0}$ monoton wachsend. Ist sie beschränkt, dann existiert der Grenzwert $\lim_{n \to \infty} A_n$, den wir mit W bezeichnen. Ist sie unbeschränkt, dann schreiben wir $\lim_{n \to \infty} A_n = \infty$ und setzen $W = \infty$.

Da die Folge $(A_n)_{n \geq 0}$ monoton wachsend ist, gilt auch $A_n \leq W$ für alle $n \geq 0$. Ist E eine endliche Teilmenge von S, dann existiert ein $n \geq 0$, sodass die Zahlen r_i für $i \in E$ unter den Zahlen $a_0, a_1, a_2, \dots, a_n$ vorkommen. Daher gilt auch $\sum_{i \in E} r_i \leq W$.

Sei $b_0, b_1, b_2, b_3, \dots$ eine andere Anordnung der Zahlen r_i für $i \in S$ und sei $B_n = \sum_{j=0}^{n} b_j$ die n-te Partialsumme. Wie oben existiert $B = \lim_{n \to \infty} B_n$, kann aber auch ∞ sein. Wieder gilt $B_n \leq B$ für alle $n \geq 0$.

Sei $n \geq 0$ beliebig. Da die Zahlen $b_0, b_1, \dots, b_n$ irgendwo in der Folge $a_0, a_1, a_2, a_3, \dots$ vorkommen, existiert eine Zahl m mit $\sum_{j=0}^{n} b_j \leq \sum_{j=0}^{m} a_j$, das heißt $B_n \leq A_m$, woraus $B_n \leq W$ folgt. Da das für alle $n \geq 0$ gilt, erhalten wir auch $\lim_{n \to \infty} B_n \leq W$, das heißt $B \leq W$. Ganz analog beweist man $W \leq B$. Damit ist $B = W$ gezeigt. Wie man die Reihenglieder auch anordnet, man erhält immer W als Grenzwert.

Sei $W < \infty$ und $\varepsilon > 0$. Wegen $\lim_{n \to \infty} A_n = W$ existiert ein n_0 mit $A_{n_0} > W - \varepsilon$. Sei U die Teilmenge von S, sodass $a_0, a_1, \dots, a_{n_0}$ gerade die Zahlen r_i für $i \in U$ sind. Dann gilt $\sum_{i \in U} r_i = \sum_{j=0}^{n_0} a_j > W - \varepsilon$. □

Satz 54: *Sei S eine abzählbare Menge und seien r_i für $i \in S$ reelle Zahlen, sodass $\sum_{i \in S} |r_i| < \infty$ gilt. Dann hat $\sum_{i \in S} r_i$ immer denselben endlichen Wert, wie immer man die Reihe auch anordnet.*

Beweis: Seien $a_0, a_1, a_2, a_3, \dots$ und $b_0, b_1, b_2, b_3, \dots$ beliebige Anordnungen der Zahlen r_i für $i \in S$. Nach Voraussetzung gilt $\sum_{i \in S} |r_i| =: R < \infty$. Dann gilt auch $\sum_{j=0}^{\infty} |a_j| = R$

und $\sum_{j=0}^{\infty} |b_j| = R$ wegen Satz 53. In der Analysis wird gezeigt, dass absolut konvergente Reihen konvergieren. Daher existieren reelle Zahlen A und B, sodass $\sum_{j=0}^{\infty} a_j = A$ und $\sum_{j=0}^{\infty} b_j = B$ gilt. Wir müssen $A = B$ zeigen.

Sei $\varepsilon > 0$ vorgegeben. Nach Satz 53 existiert eine endliche Teilmenge U von S, sodass $\sum_{i\in U} |r_i| > R - \varepsilon$ gilt. Wir wählen n so groß, dass einerseits $|\sum_{j=0}^{n} a_j - A| < \varepsilon$ gilt und andererseits die Zahlen r_i für $i \in U$ unter den Zahlen $a_0, a_1, a_2, \ldots, a_n$ vorkommen. Ebenso wählen wir m so groß, dass einerseits $|\sum_{j=0}^{m} b_j - B| < \varepsilon$ gilt und andererseits die Zahlen r_i für $i \in U$ unter den Zahlen $b_0, b_1, b_2, \ldots, b_m$ vorkommen.

Sei T diejenige endliche Teilmenge von S, sodass die Zahlen $a_0, a_1, a_2, \ldots, a_n$ gerade die Zahlen r_i für $i \in T$ sind. Es gilt dann $U \subseteq T$. Damit erhalten wir mit Hilfe von Satz 53

$$\begin{aligned} |\textstyle\sum_{j=0}^{n} a_j - \sum_{i\in U} r_i| &= |\textstyle\sum_{i\in T} r_i - \sum_{i\in U} r_i| = |\sum_{i\in T\setminus U} r_i| \le \sum_{i\in T\setminus U} |r_i| \\ &= \textstyle\sum_{i\in T} |r_i| - \sum_{i\in U} |r_i| < R - (R - \varepsilon) = \varepsilon. \end{aligned}$$

Ebenso zeigt man $|\sum_{j=0}^{m} b_j - \sum_{i\in U} r_i| < \varepsilon$. Daraus folgt dann $|\sum_{j=0}^{n} a_j - \sum_{j=0}^{m} b_j| < 2\varepsilon$ mit Hilfe der Dreiecksungleichung. Schließlich erhalten wir

$$|A - B| \le |A - \textstyle\sum_{j=0}^{n} a_j| + |\sum_{j=0}^{n} a_j - \sum_{j=0}^{m} b_j| + |\sum_{j=0}^{m} b_j - B| < 4\varepsilon.$$

Da ε beliebig klein gewählt werden kann, ist damit $A = B$ gezeigt. □

Satz 55: *Sei S abzählbar und seien r_i für $i \in S$ reelle Zahlen mit $\sum_{i\in S} |r_i| < \infty$.*
(a) *Ist $U \subseteq S$ endlich, $V \subseteq S \setminus U$ und $T = U \cup V$, dann gilt $\sum_{i\in T} r_i = \sum_{i\in U} r_i + \sum_{i\in V} r_i$.*
(b) *Für jede Teilmenge R von S gilt $|\sum_{i\in R} r_i| \le \sum_{i\in R} |r_i| < \infty$.*
(c) *Gilt $r_i \ge 0$ für alle $i \in S$ und $P \subseteq R \subseteq S$, dann gilt auch $\sum_{i\in P} r_i \le \sum_{i\in R} r_i$.*

Beweis: Sei R eine abzählbare Teilmenge von S. Sei $a_0, a_1, a_2, a_3, \ldots$ eine Anordnung der Zahlen r_i für $i \in R$. Für alle $n \ge 0$ gilt $\sum_{j=0}^{n} |a_j| \le \sum_{i\in S} |r_i| < \infty$ nach Satz 53. Mit Satz 53 folgt dann $\sum_{i\in R} |r_i| = \lim_{n\to\infty} \sum_{j=0}^{n} |a_j| \le \sum_{i\in S} |r_i| < \infty$. Aus Satz 54 erhalten wir jetzt $\lim_{n\to\infty} \sum_{j=0}^{n} a_j = \sum_{i\in R} r_i$ und daraus $\lim_{n\to\infty} |\sum_{j=0}^{n} a_j| = |\sum_{i\in R} r_i|$. Für $n \ge 0$ folgt $|\sum_{j=0}^{n} a_j| \le \sum_{j=0}^{n} |a_j|$ aus der Dreiecksungleichung. Lässt man n in dieser Ungleichung nach ∞ gehen, dann hat man $|\sum_{i\in R} r_i| \le \sum_{i\in R} |r_i|$. Das ist (b). Ist R eine endliche Teilmenge von S, dann erhält man (b) direkt aus der Dreiecksungleichung.

Wir haben $\sum_{i\in R} |r_i| \le \sum_{i\in S} |r_i|$ für alle Teilmengen R von S gezeigt (für endliches R in Satz 53). Wegen $\sum_{i\in R} |r_i| < \infty$ können wir dieses Resultat auch auf R anstelle von S anwenden. Für jede Teilmenge P von R gilt somit $\sum_{i\in P} |r_i| \le \sum_{i\in R} |r_i|$. Daraus folgt (c).

Seien U, V und T wie in (a). Seien $a_0, a_1, a_2, a_3, \ldots, a_{m-1}$ die Zahlen r_i für $i \in U$ und $a_m, a_{m+1}, a_{m+2}, \ldots$ die Zahlen r_i für $i \in V$. Nach Voraussetzung gilt $\sum_{i\in S} |r_i| < \infty$. Aus (c) folgt $\sum_{i\in V} |r_i| < \infty$ und $\sum_{i\in T} |r_i| < \infty$. Wir erhalten $\lim_{n\to\infty} \sum_{j=0}^{n} a_j = \sum_{i\in T} r_i$ und $\lim_{n\to\infty} \sum_{j=m}^{n} a_j = \sum_{i\in V} r_i$ aus Satz 54. Es gilt $\sum_{j=0}^{n} a_j = \sum_{j=0}^{m-1} a_j + \sum_{j=m}^{n} a_j$ für alle $n \ge m$. Mit $n \to \infty$ folgt dann $\sum_{i\in T} r_i = \sum_{i\in U} r_i + \sum_{i\in V} r_i$. Das ist (a). □

Wir beweisen noch einen Umordnungssatz für Reihen. Im Beweis dieses Satzes und auch im Beweis von Satz 34 werden Satz 53 und Satz 54 auf folgende Weise angewendet. Es seien die Voraussetzungen von Satz 53 oder von Satz 54 erfüllt. Für $n \ge 1$ sei S_n eine endliche Teilmenge von S, sodass $S_1 \subseteq S_2 \subseteq S_3 \subseteq \ldots$ und $\bigcup_{n=1}^{\infty} S_n = S$ gilt. Wir können uns dann die Reihe so angeordnet denken, dass zuerst die Zahlen r_i für $i \in S_1$ kommen, dann die Zahlen r_i für $i \in S_2 \setminus S_1$, dann die Zahlen r_i für $i \in S_3 \setminus S_2$, und immer so weiter. Die Summen $\sum_{i\in S_n} r_i$ sind dann Partialsummen für diese Anordnung. Aus Satz 53 oder Satz 54 folgt dann, dass $\lim_{n\to\infty} \sum_{i\in S_n} r_i = \sum_{i\in S} r_i$ gilt.

Satz 56: *Sei M eine abzählbare Menge. Weiters seien r_i für $i \in M$ reelle Zahlen mit $\sum_{i\in M} |r_i| < \infty$. Die Menge G sei endlich oder abzählbar. Die Mengen H_k für $k \in G$ seien endlich oder abzählbar und bilden eine Zerlegung der Menge M (das heißt sie sind paarweise disjunkt und ihre Vereinigung ist M). Für alle $k \in G$ gilt dann $\sum_{i\in H_k} |r_i| < \infty$ und wir setzen $s_k = \sum_{i\in H_k} r_i$. Weiters gilt auch $\sum_{k\in G} |s_k| < \infty$ und $\sum_{k\in G} s_k = \sum_{i\in M} r_i$.*

Beweis: Für $n \ge 1$ sei G_n eine endliche Teilmenge von G, sodass $G_1 \subseteq G_2 \subseteq G_3 \subseteq \dots$ und $\bigcup_{n=1}^{\infty} G_n = G$ gilt. Für $m \ge 1$ sei H_{km} eine endliche Teilmenge von H_k, sodass $H_{k1} \subseteq H_{k2} \subseteq H_{k3} \subseteq \dots$ und $\bigcup_{m=1}^{\infty} H_{km} = H_k$ gilt, wobei k in G liegt. Für $n \ge 1$ und für $m \ge 1$ setzen wir dann $M_{nm} = \bigcup_{k\in G_n} H_{km}$. Das sind endliche Teilmengen von M.

Nach Voraussetzung gilt $\sum_{i\in M} |r_i| =: C < \infty$. Aus Satz 55 (c) erhalten wir dann, dass $\sum_{i\in H_k} |r_i| \le \sum_{i\in M} |r_i| = C < \infty$ für alle $k \in G$ gilt.

Mit Hilfe dieser oben definierten Mengen und Satz 53 ergibt sich für $n \ge 1$ und $m \ge 1$

$$\sum_{k\in G_n} |\sum_{i\in H_{km}} r_i| \le \sum_{k\in G_n} \sum_{i\in H_{km}} |r_i| = \sum_{i\in M_{nm}} |r_i| \le \sum_{i\in M} |r_i| = C$$

Wir lassen m nach ∞ gehen. Wegen $s_k = \sum_{i\in H_k} r_i$ erhalten wir $\sum_{k\in G_n} |s_k| \le C$ für alle $n \ge 1$ mit Hilfe von Satz 54. Jetzt lassen wir auch n nach ∞ gehen und erhalten $\sum_{k\in G} |s_k| \le C$ mit Hilfe von Satz 53. Wir setzen $\sum_{k\in G} |s_k| =: D < \infty$.

Um $\sum_{k\in G} s_k = \sum_{i\in M} r_i$ zu zeigen, gehen wir von folgender Gleichung aus

$$\sum_{k\in G_n} \sum_{i\in H_{kn}} r_i = \sum_{i\in M_{nn}} r_i$$

Wegen $\sum_{i\in M} |r_i| < \infty$ und $\bigcup_{n=1}^{\infty} M_{nn} = M$ konvergiert die rechte Seite dieser Gleichung für $n \to \infty$ gegen $\sum_{i\in M} r_i$, wofür wir Satz 54 verwenden. Unten zeigen wir, dass die linke Seite dieser Gleichung für $n \to \infty$ gegen $\sum_{k\in G} s_k$ konvergiert. Damit ist dann die Gleichung $\sum_{k\in G} s_k = \sum_{i\in M} r_i$ bewiesen.

Um die gewünschte Konvergenz zu zeigen, sei $\varepsilon > 0$ vorgegeben. Nach Satz 53 existiert eine endliche Teilmenge U von M mit $\sum_{i\in U} |r_i| > C - \varepsilon$. Wir wählen u so groß, dass $U \subseteq M_{uu}$ und $\sum_{k\in G_u} |s_k| > D - \varepsilon$ gilt. Sei a die Anzahl der Elemente der Menge G_u. Dann wählen wir v so groß, dass $|s_k - \sum_{i\in H_{kn}} r_i| < \frac{\varepsilon}{a}$ für alle $k \in G_u$ und alle $n \ge v$ gilt. Sei $n_0 = \max(u, v)$. Für $n \ge n_0$ erhalten wir dann mit Hilfe von Satz 55 (a)

$$\begin{aligned} &|\sum_{k\in G_n} \sum_{i\in H_{kn}} r_i - \sum_{k\in G} s_k| \\ &\quad \le |\sum_{k\in G_u} \big(\sum_{i\in H_{kn}} r_i - s_k\big)| + |\sum_{k\in G_n\setminus G_u} \sum_{i\in H_{kn}} r_i| + |\sum_{k\in G\setminus G_u} s_k|. \end{aligned}$$

Wir schätzen diese drei Summanden ab. Da $|s_k - \sum_{i\in H_{kn}} r_i| < \frac{\varepsilon}{a}$ für alle $k \in G_u$ und alle $n \ge v$ nach Wahl von v gilt, erhalten wir

$$|\sum_{k\in G_u} \big(\sum_{i\in H_{kn}} r_i - s_k\big)| \le \sum_{k\in G_u} |\sum_{i\in H_{kn}} r_i - s_k| < \sum_{k\in G_u} \tfrac{\varepsilon}{a} = \varepsilon.$$

Da $\sum_{i\in M_{un}} |r_i| \ge \sum_{i\in M_{uu}} |r_i| \ge \sum_{i\in U} |r_i| > C - \varepsilon$ nach Wahl von u gilt, folgt mit Satz 53

$$\begin{aligned} |\sum_{k\in G_n\setminus G_u} \sum_{i\in H_{kn}} r_i| &\le \sum_{k\in G_n\setminus G_u} \sum_{i\in H_{kn}} |r_i| = \sum_{k\in G_n} \sum_{i\in H_{kn}} |r_i| - \sum_{k\in G_u} \sum_{i\in H_{kn}} |r_i| \\ &= \sum_{i\in M_{nn}} |r_i| - \sum_{i\in M_{un}} |r_i| < C - (C - \varepsilon) = \varepsilon. \end{aligned}$$

Da $\sum_{k\in G_u} |s_k| > D - \varepsilon$ nach Wahl von u gilt, erhalten wir mit Satz 55 (b) und Satz 55 (a)

$$|\sum_{k\in G\setminus G_u} s_k| \le \sum_{k\in G\setminus G_u} |s_k| = \sum_{k\in G} |s_k| - \sum_{k\in G_u} |s_k| < D - (D - \varepsilon) = \varepsilon.$$

Wir setzen das alles zusammen und erhalten

$$|\sum_{k\in G_n} \sum_{i\in H_{kn}} r_i - \sum_{k\in G} s_k| < 3\varepsilon.$$

Da diese Ungleichung für alle $n \ge n_0$ gilt, ist $\lim_{n\to\infty} \sum_{k\in G_n} \sum_{i\in H_{kn}} r_i = \sum_{k\in G} s_k$ gezeigt und der Beweis ist vollständig. $\square$

2. Uneigentliches Integral

Sei $f : \mathbb{R} \to \mathbb{R}^+$ eine Funktion, die auf jedem beschränkten Intervall integrierbar ist. Wir definieren die uneigentlichen Integrale mit einem unbeschränkten Integrationsintervall: $\int_a^\infty f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_a^n f(x)\,\mathrm{d}x$ und $\int_{-\infty}^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{-n}^b f(x)\,\mathrm{d}x$. Wegen $f \geq 0$ werden die Integrale, deren Grenzwert gebildet wird, größer, wenn man n vergrößert. Wir bilden also Grenzwerte von monoton wachsenden Folgen. Daher existieren die Grenzwerte, können aber ∞ sein.

Die üblichen Rechenregeln für Integrale bleiben dabei erhalten. Sind zum Beispiel a und u in $\mathbb{R}$, dann gilt $\int_a^n f(x)\,\mathrm{d}x = \int_a^u f(x)\,\mathrm{d}x + \int_u^n f(x)\,\mathrm{d}x$. Lässt man n gegen ∞ gehen, dann erhält man $\int_a^\infty f(x)\,\mathrm{d}x = \int_a^u f(x)\,\mathrm{d}x + \int_u^\infty f(x)\,\mathrm{d}x$. Sind f und g integrierbare Funktionen mit Werten in $\mathbb{R}^+$, dann gilt $\int_a^n f(x) + g(x)\,\mathrm{d}x = \int_a^n f(x)\,\mathrm{d}x + \int_a^n g(x)\,\mathrm{d}x$. Lässt man n gegen ∞ gehen, dann erhält man $\int_a^\infty f(x) + g(x)\,\mathrm{d}x = \int_a^\infty f(x)\,\mathrm{d}x + \int_a^\infty g(x)\,\mathrm{d}x$.

Weiters definieren wir $\int_{-\infty}^\infty f(x)\,\mathrm{d}x$ durch $\int_{-\infty}^0 f(x)\,\mathrm{d}x + \int_0^\infty f(x)\,\mathrm{d}x$. Für alle $a \in \mathbb{R}$ gilt dann auch $\int_{-\infty}^\infty f(x)\,\mathrm{d}x = \int_{-\infty}^a f(x)\,\mathrm{d}x + \int_a^\infty f(x)\,\mathrm{d}x$. Aus dieser Definition folgt auch $\int_{-\infty}^\infty f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{-n}^0 f(x)\,\mathrm{d}x + \int_0^\infty f(x)\,\mathrm{d}x = \lim_{n\to\infty} \big(\int_{-n}^0 f(x)\,\mathrm{d}x + \int_0^\infty f(x)\,\mathrm{d}x \big) = \lim_{n\to\infty} \int_{-n}^\infty f(x)\,\mathrm{d}x$. Genauso erhalten wir $\int_{-\infty}^\infty f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{-\infty}^n f(x)\,\mathrm{d}x$ und $\int_{-\infty}^\infty f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_{-n}^n f(x)\,\mathrm{d}x$.

Ist die Funktion $g : \mathbb{R} \to \mathbb{R}$ auf jedem beschränkten Intervall integrierbar, aber nicht ≥ 0, dann muss man vorsichtiger sein. Man verlangt dann, dass $\int_{-\infty}^\infty |g(x)|\,\mathrm{d}x < \infty$ gilt. In diesem Fall kann man die uneigentlichen Integrale wie oben definieren und es gelten wieder die üblichen Rechenregeln. Dieses Problem tritt beim Erwartungswert einer kontinuierlichen Zufallsvariable auf.

Eine andere Art von uneigentlichen Integralen erhält man, wenn $f : [a, b) \to \mathbb{R}^+$ unbeschränkt, jedoch für jedes $\varepsilon > 0$ auf dem Intervall $[a, b - \varepsilon]$ beschränkt und integrierbar ist. Wir sagen dann einfach, dass f eine Polstelle bei b hat. Man definiert dann $\int_a^b f(x)\,\mathrm{d}x = \lim_{\varepsilon\downarrow 0} \int_a^{b-\varepsilon} f(x)\,\mathrm{d}x$. Ist $f : (a, b] \to \mathbb{R}^+$ unbeschränkt, jedoch für jedes $\varepsilon > 0$ auf dem Intervall $[a + \varepsilon, b]$ beschränkt und integrierbar (f hat eine Polstelle bei a), dann definiert man $\int_a^b f(x)\,\mathrm{d}x = \lim_{\varepsilon\downarrow 0} \int_{a+\varepsilon}^b f(x)\,\mathrm{d}x$. Wegen $f \geq 0$ sind $\varepsilon \mapsto \int_a^{b-\varepsilon} f(x)\,\mathrm{d}x$ und $\varepsilon \mapsto \int_{a+\varepsilon}^b f(x)\,\mathrm{d}x$ monoton fallende Abbildungen. Daher existieren die Grenzwerte, können aber ∞ sein. Gilt $a < c < b$ und hat f eine Polstelle bei c, dann schreibt man $\int_a^b f(x)\,\mathrm{d}x$ als $\int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x$ und wendet obige Definitionen an.

Die üblichen Rechenregeln für Integrale bleiben dabei erhalten. Hat man zum Beispiel $a < u < b$ und eine Polstelle bei b, dann gilt $\int_a^{b-\varepsilon} f(x)\,\mathrm{d}x = \int_a^u f(x)\,\mathrm{d}x + \int_u^{b-\varepsilon} f(x)\,\mathrm{d}x$. Lässt man ε gegen 0 gehen, dann erhält man $\int_a^b f(x)\,\mathrm{d}x = \int_a^u f(x)\,\mathrm{d}x + \int_u^b f(x)\,\mathrm{d}x$.

Satz 57: *Sei $f : \mathbb{R} \to \mathbb{R}^+$ eine integrierbare Funktion. Dann gilt $\lim_{\varepsilon\to 0} \int_{b-\varepsilon}^b f(x)\,\mathrm{d}x = 0$.*

Beweis: Hat f eine Polstelle bei b, dann ist das Integral uneigentlich. Wir wählen $a < b$. Nach Definition des uneigentlichen Integrals gilt dann $\lim_{\varepsilon\to 0} \int_a^{b-\varepsilon} f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x$. Wegen $\int_{b-\varepsilon}^b f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x - \int_a^{b-\varepsilon} f(x)\,\mathrm{d}x$ erhalten wir $\lim_{\varepsilon\to 0} \int_{b-\varepsilon}^b f(x)\,\mathrm{d}x = 0$.

Hat f keine Polstelle bei b (wir haben ein eigentliches Integral), dann existiert ein $\varepsilon_0 > 0$ und eine Konstante c mit $f(x) \leq c$ für alle $x \in [b - \varepsilon_0, b]$. Für $\varepsilon < \varepsilon_0$ erhalten wir dann $0 \leq \int_{b-\varepsilon}^b f(x)\,\mathrm{d}x \leq \varepsilon c$. Daraus folgt $\lim_{\varepsilon\to 0} \int_{b-\varepsilon}^b f(x)\,\mathrm{d}x = 0$. $\square$

3. Volumen

Für Volumsberechnungen führt man üblicherweise ein mehrdimensionales Integral ein. Man kann aber auch mit dem eindimensionalen Integral auskommen.

Sei E_x die Ebene im $\mathbb{R}^3$ durch den Punkt $(x,0,0)$, die senkrecht auf die x-Achse steht. Der Körper, dessen Volumen V wir berechnen wollen, liege im $\mathbb{R}^3$ zwischen den Ebenen E_a und E_b. Für $x \in [a,b]$ sei $u(x)$ die Fläche des Schnittes der Ebene E_x mit dem Körper. Wir nehmen an, dass $u : [a,b] \to \mathbb{R}$ eine integrierbare Funktion ist.

Sei $a = x_0 < x_1 < \cdots < x_n = b$ eine beliebige Zerlegung des Intervalls $[a,b]$. Sei V_j das Volumen des Teils des Körpers, der zwischen den Ebenen $E_{x_{j-1}}$ und E_{x_j} liegt. Es gilt

$$(x_j - x_{j-1}) \inf_{x\in[x_{j-1},x_j]} u(x) \le V_j \le (x_j - x_{j-1}) \sup_{x\in[x_{j-1},x_j]} u(x)$$

Wegen $\sum_{j=1}^n V_j = V$ erhalten wir dann

$$\sum_{j=1}^n (x_j - x_{j-1}) \inf_{x\in[x_{j-1},x_j]} u(x) \le V \le \sum_{j=1}^n (x_j - x_{j-1}) \sup_{x\in[x_{j-1},x_j]} u(x)$$

Links steht die Untersumme und rechts die Obersumme des Integrals $\int_a^b u(x)\,\mathrm{d}x$ zur Zerlegung $a = x_0 < x_1 < \cdots < x_n = b$. Lässt man die Gitterweite der Zerlegung gegen 0 gehen, dann konvergieren beide Summen gegen dieses Integral. Wir erhalten daher

$$V = \int_a^b u(x)\,\mathrm{d}x$$

und haben damit eine Formel für das Volumen des Körpers gefunden.

Kennt man die Schnittflächen $u(x)$ des Körpers, so kann man das Volumen berechnen. Bei einem Drehkörper, der durch die Rotation des Flächenstücks unter der stetigen Funktion $h : [a,b] \to \mathbb{R}^+$ um die x-Achse entsteht, ist das einfach. Für $x \in [a,b]$ ist die Fläche des Schnittes der Ebene E_x mit dem Drehkörper gleich $u(x) = \pi h(x)^2$. Deshalb ist

$$V = \pi \int_a^b h(x)^2\,\mathrm{d}x$$

die Formel für das Volumen des Drehkörpers.

Hier geht es jedoch darum, das Volumen über einem Bereich $B \subseteq \mathbb{R}^2$ unter einer Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$ zu berechnen. Wir schreiben $\int_B g(x,y)\,\mathrm{d}(x,y)$ für dieses Volumen analog zum eindimensionalen Integral $\int_{[a,b]} f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x$, das die Fläche über dem Intervall $[a,b]$ unter der Funktion $f : \mathbb{R} \to \mathbb{R}^+$ angibt.

Um mit obiger Methode das Volumen $\int_B g(x,y)\,\mathrm{d}(x,y)$ über einem Bereich $B \subseteq \mathbb{R}^2$ unter einer Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$ zu berechnen, müssen wir annehmen, dass B ein sogenannter Normalbereich ist. Man nennt B einen Normalbereich, wenn ein Intervall $[a,b]$ und stetige Funktionen $\varphi : [a,b] \to \mathbb{R}$ und $\psi : [a,b] \to \mathbb{R}$ mit $\varphi \le \psi$ existieren, sodass

$$B = \{(x,y) \in \mathbb{R}^2 : a \le x \le b,\ \varphi(x) \le y \le \psi(x)\}$$

gilt. Die Fläche $u(x)$ des Schnittes der Ebene E_x mit dem Körper über B und unter g lässt sich dann durch das Integral $u(x) = \int_{\varphi(x)}^{\psi(x)} g(x,y)\,\mathrm{d}y$ berechnen. Wir erhalten also

$$\int_B g(x,y)\,\mathrm{d}(x,y) = \int_a^b u(x)\,\mathrm{d}x = \int_a^b \int_{\varphi(x)}^{\psi(x)} g(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

als Volumen über dem Normalbereich B unter der Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$.

Approximation durch Integrale über Rechteckbereiche: Wir nennen einen Bereich $R = \{(x,y) \in \mathbb{R}^2 : c < x \leq d,\ \varrho(x) < y \leq \chi(x)\}$ Rechteckbereich, wenn $c < d$ gilt und die Funktionen $\varrho : (c,d] \to \mathbb{R}$ und $\chi : (c,d] \to \mathbb{R}$ linksseitig stetige Treppenfunktionen sind, für die $\varrho \leq \chi$ gilt. Für den Normalbereich $B = \{(x,y) \in \mathbb{R}^2 : a \leq x \leq b,\ \varphi(x) \leq y \leq \psi(x)\}$ soll das Integral $\int_B g(x,y)\,\mathrm{d}(x,y)$ durch Integrale über Rechteckbereiche approximiert werden.

Wir nehmen an, dass die Funktion g beschränkt ist. Sei d eine obere Schranke für g und $\varepsilon > 0$ beliebig. Wir finden eine Zerlegung $a = c_0 < c_1 < \dots < c_n = b$ des Intervalls $[a,b]$ mit folgenden Eigenschaften: Für $1 \leq j \leq n$ existieren reelle Zahlen p_j und q_j, sodass $p_j < \varphi(x) < q_j$ für alle $x \in (c_{j-1}, c_j]$ und $\sum_{j=1}^n (q_j - p_j)(c_j - c_{j-1}) < \frac{\varepsilon}{2d}$ gilt. Ebenso existieren reelle Zahlen u_j und v_j, sodass $u_j < \psi(x) < v_j$ für alle $x \in (c_{j-1}, c_j]$ und $\sum_{j=1}^n (v_j - u_j)(c_j - c_{j-1}) < \frac{\varepsilon}{2d}$ gilt. Das ist möglich, da φ und ψ integrierbar sind. Wegen $\varphi \leq \psi$ können wir auch annehmen, dass $p_j \leq u_j$ und $q_j \leq v_j$ für $1 \leq j \leq n$ gilt.

Wir definieren Treppenfunktionen β und γ auf dem Intervall $(a,b]$ und Treppenfunktionen α und δ auf dem Intervall $(a-\eta, b]$, wobei $\eta > 0$ noch geeignet gewählt wird. Für $1 \leq j \leq n$ und $x \in (c_{j-1}, c_j]$ sei $\beta(x) = q_j$ und $\gamma(x) = \max(q_j, u_j)$. Für $2 \leq j \leq n$ und $x \in (c_{j-1}, c_j]$ sei $\alpha(x) = p_j$ und $\delta(x) = v_j$. Für $x \in (a-\eta, c_1]$ sei $\alpha(x) = p_1$ und $\delta(x) = v_1$. Es gilt dann $\alpha \leq \beta \leq \gamma \leq \delta$ auf dem Intervall $(a,b]$. Wir erhalten damit die Rechteckbereiche $C = \{(x,y) \in \mathbb{R}^2 : a < x \leq b,\ \beta(x) < y \leq \gamma(x)\}$ und $D = \{(x,y) \in \mathbb{R}^2 : a - \eta < x \leq b,\ \alpha(x) < y \leq \delta(x)\}$.

Aus diesen Definitionen folgt $C \subseteq B \subseteq D$. Weiters erhalten wir

$$\begin{aligned}
&\textstyle\int_D g(x,y)\,\mathrm{d}(x,y) - \int_C g(x,y)\,\mathrm{d}(x,y) \\
&= \textstyle\int_{a-\eta}^{a} \int_{\alpha(x)}^{\delta(x)} g(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_a^b \int_{\alpha(x)}^{\beta(x)} g(x,y)\,\mathrm{d}y\,\mathrm{d}x + \int_a^b \int_{\gamma(x)}^{\delta(x)} g(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&\leq d\eta(v_1 - p_1) + d\textstyle\sum_{j=1}^n (q_j - p_j)(c_j - c_{j-1}) + d\sum_{j=1}^n (v_j - u_j)(c_j - c_{j-1})
\end{aligned}$$

Da $d\sum_{j=1}^n (q_j - p_j)(c_j - c_{j-1}) + d\sum_{j=1}^n (v_j - u_j)(c_j - c_{j-1}) < \varepsilon$ aus obigen Definitionen folgt, können wir $\eta > 0$ so klein wählen, dass $\int_D g(x,y)\,\mathrm{d}(x,y) - \int_C g(x,y)\,\mathrm{d}(x,y) < \varepsilon$ gilt.

Unbeschränkte Normalbereiche: In der Definition eines Normalbereiches B lassen wir $a = -\infty$ und $b = \infty$ zu und die Funktionen φ und ψ dürfen auch die Werte $-\infty$ und ∞ annehmen. Da wir uneigentliche Integrale bereits in Abschnitt 2 definiert haben, kann man das Volumen $\int_B g(x,y)\,\mathrm{d}(x,y)$ für eine Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$ wie oben berechnen.

Ein Rechteck $R = [a,b] \times [c,d]$ ist ein spezieller Normalbereich, wobei wir auch unbeschränkte Intervalle zulassen. Es gilt dann $\int_R g(x,y)\,\mathrm{d}(x,y) = \int_a^b \int_c^d g(x,y)\,\mathrm{d}y\,\mathrm{d}x$. Wir können die Rollen der beiden Koordinaten vertauschen und y als erste und x als zweite Koordinate auffassen. Wir erhalten dann $\int_R g(x,y)\,\mathrm{d}(x,y) = \int_c^d \int_a^b g(x,y)\,\mathrm{d}x\,\mathrm{d}y$. Es folgt $\int_a^b \int_c^d g(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_c^d \int_a^b g(x,y)\,\mathrm{d}x\,\mathrm{d}y$, da wir in beiden Fällen das Volumen über dem Rechteck $R = [a,b] \times [c,d]$ unter der Funktion $g : \mathbb{R}^2 \to \mathbb{R}^+$ ausrechnen. Wir verwenden diese Formel auch für unbeschränkte Rechtecke. Es können dann auch $-\infty$ und ∞ als Integrationsgrenzen auftreten.

Es gelte $\int_{\mathbb{R}^2} g(x,y)\,\mathrm{d}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x = C < \infty$ und $\varepsilon > 0$ sei beliebig. Wir verwenden die Resultate über uneigentliche Integrale, die wir in Abschnitt 2 gefunden haben. Es existiert ein $m > 0$ mit $\int_{-m}^{m} \int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x > C - \frac{\varepsilon}{2}$. Wir verwenden die Formel $\int_{-m}^{m} \int_{-\infty}^{\infty} g(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty} \int_{-m}^{m} g(x,y)\,\mathrm{d}x\,\mathrm{d}y$. Es existiert ein $n > 0$ mit $\int_{-n}^{n} \int_{-m}^{m} g(x,y)\,\mathrm{d}x\,\mathrm{d}y > C - \varepsilon$. Für jedes $\varepsilon > 0$ existiert somit ein beschränktes Rechteck $U = [-m,m] \times [-n,n]$, für das $\int_U g(x,y)\,\mathrm{d}(x,y) = \int_{-m}^{m} \int_{-n}^{n} g(x,y)\,\mathrm{d}y\,\mathrm{d}x > C - \varepsilon$ gilt.

Sei $B = \{(x,y) \in \mathbb{R}^2 : a \le x \le b,\ \varphi(x) \le y \le \psi(x)\}$ ein Normalbereich. Für ein beschränktes Rechteck $U = [-m,m] \times [-n,n]$ ist dann $B \cap U$ der beschränkte Normalbereich $\{(x,y) \in \mathbb{R}^2 : \max(a,-m) \le x \le \min(b,m),\ \max(\varphi(x),-n) \le y \le \min(\psi(x),n)\}$.

4. Berechnung eines Integrals

Wir verwenden obige Resultate, um das Integral $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,\mathrm{d}x$ zu berechnen. Setzt man $I_n = \int_{-n}^{n} e^{-\frac{1}{2}x^2}\,\mathrm{d}x$, dann erhält man dieses Integral als $\lim_{n\to\infty} I_n$.

Die Rechnung führt über ein zweifaches Integral. Wir schreiben Q_n für das Quadrat $[-n,n] \times [-n,n]$. Durch entsprechende Umformungen erhalten wir dann

$$\begin{aligned} I_n^2 &= I_n \int_{-n}^{n} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{-n}^{n} e^{-\frac{1}{2}x^2} I_n\,\mathrm{d}x = \int_{-n}^{n} e^{-\frac{1}{2}x^2} \int_{-n}^{n} e^{-\frac{1}{2}y^2}\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-n}^{n}\int_{-n}^{n} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2}\,\mathrm{d}y\,\mathrm{d}x = \int_{-n}^{n}\int_{-n}^{n} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}y\,\mathrm{d}x = \int_{Q_n} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y) \end{aligned}$$

Wir haben nur die Rechenregeln für einfache Integrale verwendet und das zweifache Integral, das am Ende herauskommt, als Volumen über dem Quadrat Q_n interpretiert, das unter der Funktion $e^{-\frac{1}{2}(x^2+y^2)}$ liegt.

Sei K_R der Kreis mit Mittelpunkt $(0,0)$ und Radius R. Es gilt $K_n \subseteq Q_n \subseteq K_{2n}$, woraus $\int_{K_n} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y) \le \int_{Q_n} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y) \le \int_{K_{2n}} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y)$ folgt. Damit haben wir eine Abschätzung für I_n^2 gefunden.

Wir berechnen $V = \int_{K_R} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y)$. Dazu zerlegen wir den Kreis K_R in Kreisringe. Sei $k \ge 1$ und $r_j = j\frac{R}{k}$ für $0 \le j \le k$. Sei weiters V_j das Volumen unter der Funktion $e^{-\frac{1}{2}(x^2+y^2)}$ über dem Kreisring $K_{r_j} \setminus K_{r_{j-1}}$. Es gilt dann $V = \sum_{j=1}^{k} V_j$. Die Funktionswerte auf diesem Kreisring liegen zwischen $e^{-\frac{1}{2}r_{j-1}^2}$ und $e^{-\frac{1}{2}r_j^2}$. Die Fläche des Kreisrings ist $\pi(r_j^2 - r_{j-1}^2)$. Daher gilt

$$e^{-\frac{1}{2}r_j^2}\,\pi(r_j^2 - r_{j-1}^2) \le V_j \le e^{-\frac{1}{2}r_{j-1}^2}\,\pi(r_j^2 - r_{j-1}^2)$$

Wegen $r_j^2 - r_{j-1}^2 = (r_j + r_{j-1})(r_j - r_{j-1}) = (2r_j - \frac{R}{k})(r_j - r_{j-1}) = (2r_{j-1} + \frac{R}{k})(r_j - r_{j-1})$ und da $e^{-\frac{1}{2}r^2} \le 1$ für alle $r \ge 0$ gilt, erhalten wir

$$2\pi e^{-\frac{1}{2}r_j^2}\,r_j(r_j - r_{j-1}) - \tfrac{\pi R}{k}(r_j - r_{j-1}) \le V_j \le 2\pi e^{-\frac{1}{2}r_{j-1}^2}\,r_{j-1}(r_j - r_{j-1}) + \tfrac{\pi R}{k}(r_j - r_{j-1})$$

Summation über j von 1 bis k ergibt schließlich

$$2\pi \sum_{j=1}^{k} e^{-\frac{1}{2}r_j^2}\,r_j(r_j - r_{j-1}) - \tfrac{\pi R^2}{k} \le V \le 2\pi \sum_{j=1}^{k} e^{-\frac{1}{2}r_{j-1}^2}\,r_{j-1}(r_j - r_{j-1}) + \tfrac{\pi R^2}{k}$$

Die beiden Summen sind Riemannsummen, die gegen das Integral $\int_0^R e^{-\frac{1}{2}r^2}\,r\,\mathrm{d}r$ konvergieren, wenn k nach ∞ geht. Dadurch und wegen $\lim_{k\to\infty} \frac{\pi R^2}{k} = 0$ erhalten wir

$$V = 2\pi \int_0^R e^{-\frac{1}{2}r^2}\,r\,\mathrm{d}r$$

Dieses Integral lässt sich durch die Substitution $t = r^2$ leicht ausrechnen. Man erhält

$$V = \pi \int_0^{R^2} e^{-\frac{1}{2}t}\,\mathrm{d}t = -2\pi e^{-\frac{1}{2}t}\Big|_0^{R^2} = 2\pi(1 - e^{-\frac{1}{2}R^2})$$

Damit ist $V = \int_{K_R} e^{-\frac{1}{2}(x^2+y^2)}\,\mathrm{d}(x,y)$ berechnet. Setzt man das in die oben gefundene Abschätzung für I_n^2 ein, so erhält man

$$2\pi(1 - e^{-\frac{1}{2}n^2}) \le I_n^2 \le 2\pi(1 - e^{-2n^2}) \quad \text{oder} \quad \sqrt{2\pi}\sqrt{1 - e^{-n^2/2}} \le I_n \le \sqrt{2\pi}\sqrt{1 - e^{-2n^2}}$$

Es folgt $\lim_{n\to\infty} I_n = \sqrt{2\pi}$. Damit haben wir das gewünschte Resultat gefunden. Es gilt

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \sqrt{2\pi}$$

Die Wahrscheinlichkeitsdichte der Standardnormalverteilung hat somit Integral 1.

Wir berechnen das Volumen V unter der Funktion $e^{-\frac{1}{2}(x^2+y^2)}$ über dem Kreis K_R noch auf eine andere Art, nämlich als Volumen eines Drehkörpers. Wir schneiden die Funktion $\varphi(x) = e^{-\frac{1}{2}x^2}$ bei $-R$ und R ab und lassen sie um die y-Achse rotieren. Das Volumen des Drehkörpers ist dann V.

Die Umkehrfunktion von φ erhalten wir durch Auflösen der Gleichung $e^{-\frac{1}{2}x^2} = y$ nach der Variable x. Das ergibt $\varphi^{-1}(y) = \sqrt{-2\log y}$. Wir schneiden bei R ab und erhalten so die Funktion $h(y) = \min(R, \sqrt{-2\log y})$, die auf dem Intervall $[0,1]$ definiert ist. Durch Rotation um die y-Achse ergibt sich das Volumen V, das heißt $V = \pi \int_0^1 h(y)^2 \,\mathrm{d}y$. Wir teilen das Integrationsinterval bei $y_0 = e^{-\frac{1}{2}R^2}$. Es gilt $h(y) = R$ auf dem Intervall $[0, y_0]$ und $h(y) = \sqrt{-2\log y}$ auf dem Intervall $[y_0, 1]$. Damit erhalten wir dann

$$\begin{aligned} V &= \pi \int_0^{y_0} R^2 \,\mathrm{d}y + \pi \int_{y_0}^1 -2\log y \,\mathrm{d}y = \pi R^2 y_0 - 2\pi(y\log y - y)\Big|_{y_0}^1 \\ &= \pi R^2 y_0 + 2\pi + 2\pi y_0 \log y_0 - 2\pi y_0 = \pi R^2 y_0 + 2\pi - \pi y_0 R^2 - 2\pi e^{-\frac{1}{2}R^2} \\ &= 2\pi(1 - e^{-\frac{1}{2}R^2}) \end{aligned}$$

Damit haben wir für V denselben Wert wie oben erhalten. Alles andere lässt sich genauso wie oben berechnen.

5. Die Gammafunktion

Die Gammafunktion wird durch $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}\,\mathrm{d}x$ für alle $r > 0$ definiert.

Satz 58: *Für $r > 0$ gilt $\Gamma(r+1) = r\Gamma(r)$.*

Beweis: Das zeigt man mit partieller Integration:
$\Gamma(r+1) = \int_0^\infty x^r e^{-x}\,\mathrm{d}x = -x^r e^{-x}\Big|_0^\infty + \int_0^\infty r x^{r-1} e^{-x}\,\mathrm{d}x = 0 + r\Gamma(r)$. □

Für spezielle Werte von r, zum Beispiel für $r \in \mathbb{N}$, kann man $\Gamma(r)$ explizit berechnen. Es gilt $\Gamma(1) = \int_0^\infty e^{-x}\,\mathrm{d}x = -e^{-x}\Big|_0^\infty = 1$ und für $n \in \{1,2,3,\dots\}$ gilt $\Gamma(n) = (n-1)!$. Das folgt mit Induktion. Für $n = 1$ haben wir es gerade berechnet. Ist $\Gamma(n) = (n-1)!$ schon gezeigt, dann folgt $\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n!$ mit Hilfe von Satz 58.

Satz 59: *Es gilt $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.*

Beweis: Wir haben $\int_{-\infty}^\infty e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \sqrt{2\pi}$ gezeigt. Da $e^{-\frac{1}{2}x^2}$ eine gerade Funktion ist, ist das Integral über $[0,\infty)$ halb so groß. Wir haben daher $\int_0^\infty e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}$. Wir führen die neue Integrationsvariable $y = \frac{1}{2}x^2$ ein. Es gilt $x = \sqrt{2y}$ und $\mathrm{d}x = \frac{1}{\sqrt{2y}}\,\mathrm{d}y$. Wir erhalten $\int_0^\infty e^{-y}\frac{1}{\sqrt{2y}}\,\mathrm{d}y = \sqrt{\frac{\pi}{2}}$. Multipliziert man diese Gleichung mit $\sqrt{2}$, dann steht bereits $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ da. □

Satz 60: *Für $n \in \{0,1,2,\dots\}$ gilt $\Gamma(n+\frac{1}{2}) = \frac{(2n)!}{n!2^{2n}}\sqrt{\pi}$.*

Beweis: Für $n = 0$ ist das Satz 59. Sei $\Gamma(n+\frac{1}{2}) = \frac{(2n)!}{n!2^{2n}}\sqrt{\pi}$ bereits gezeigt. Aus Satz 58 folgt $\Gamma(n+\frac{3}{2}) = (n+\frac{1}{2})\Gamma(n+\frac{1}{2}) = \frac{(2n+2)(2n+1)}{4(n+1)}\Gamma(n+\frac{1}{2})$. Setzt man für $\Gamma(n+\frac{1}{2})$ ein, so hat man $\Gamma(n+\frac{3}{2}) = \frac{(2n+2)!}{(n+1)!2^{2n+2}}\sqrt{\pi}$, die Formel für $n+1$. □

Inhaltsverzeichnis

Tabelle für die $N(0,1)$-Verteilung

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.500	0.30	0.618	0.60	0.726	0.90	0.816	1.20	0.885	1.50	0.933	2.25	0.988
0.02	0.508	0.32	0.626	0.62	0.732	0.92	0.821	1.22	0.889	1.55	0.939	2.30	0.989
0.04	0.516	0.34	0.633	0.64	0.739	0.94	0.826	1.24	0.893	1.60	0.945	2.35	0.991
0.06	0.524	0.36	0.641	0.66	0.745	0.96	0.832	1.26	0.896	1.65	0.951	2.40	0.992
0.08	0.532	0.38	0.648	0.68	0.752	0.98	0.837	1.28	0.900	1.70	0.955	2.45	0.993
0.10	0.540	0.40	0.655	0.70	0.758	1.00	0.841	1.30	0.903	1.75	0.960	2.50	0.994
0.12	0.548	0.42	0.663	0.72	0.764	1.02	0.846	1.32	0.907	1.80	0.964	2.55	0.995
0.14	0.556	0.44	0.670	0.74	0.770	1.04	0.851	1.34	0.910	1.85	0.968	2.60	0.995
0.16	0.564	0.46	0.677	0.76	0.776	1.06	0.855	1.36	0.913	1.90	0.971	2.65	0.996
0.18	0.571	0.48	0.684	0.78	0.782	1.08	0.860	1.38	0.916	1.95	0.974	2.70	0.997
0.20	0.579	0.50	0.692	0.80	0.788	1.10	0.864	1.40	0.919	2.00	0.977	2.75	0.997
0.22	0.587	0.52	0.699	0.82	0.794	1.12	0.869	1.42	0.922	2.05	0.980	2.80	0.997
0.24	0.595	0.54	0.705	0.84	0.800	1.14	0.873	1.44	0.925	2.10	0.982	2.85	0.998
0.26	0.603	0.56	0.712	0.86	0.805	1.16	0.877	1.46	0.928	2.15	0.984	2.90	0.998
0.28	0.610	0.58	0.719	0.88	0.811	1.18	0.881	1.48	0.931	2.20	0.986	2.97	0.999

Tabellen für $\varrho\beta$ und $v_{\beta,n}$ (Normalverteilung und T–Verteilung)

$\beta =$	$\varrho\beta$	$v_{\beta,7}$	$v_{\beta,8}$	$v_{\beta,9}$	$v_{\beta,10}$	$v_{\beta,12}$	$v_{\beta,14}$	$v_{\beta,16}$	$v_{\beta,18}$	$v_{\beta,25}$	$v_{\beta,40}$
0.05	1.65	1.895	1.860	1.833	1.812	1.782	1.761	1.746	1.734	1.708	1.684
0.025	1.96	2.365	2.306	2.262	2.228	2.179	2.145	2.120	2.101	2.060	2.021
0.01	2.33	2.998	2.896	2.821	2.764	2.681	2.624	2.584	2.552	2.485	2.423
0.005	2.58	3.499	3.355	3.250	3.169	3.055	2.977	2.921	2.878	2.787	2.704

Tabellen für $u_{\beta,n}$ (χ^2-Verteilung)

$\beta =$	$u_{\beta,1}$	$u_{\beta,2}$	$u_{\beta,3}$	$u_{\beta,4}$	$u_{\beta,5}$	$u_{\beta,6}$	$u_{\beta,8}$	$u_{\beta,9}$	$u_{\beta,10}$	$u_{\beta,16}$	$u_{\beta,20}$
0.05	3.84	5.99	7.81	9.49	11.07	12.6	15.51	16.92	18.31	26.30	31.41
0.025	5.02	7.38	9.35	11.1	12.83	14.4	17.53	19.02	20.48	28.85	34.17
0.01	6.63	9.21	11.3	13.3	15.1	16.8	20.1	21.7	23.2	32.0	37.6
0.005	7.88	10.6	12.8	14.9	16.7	18.5	22.0	23.6	25.2	34.3	40.0

Zusammenstellung der wichtigen Verteilungen

Name	Parameter	Abkürzung	Wertebereich	W–Vektor/Dichte	Ew.	Varianz
Binomial-verteilung	$n \in \mathbb{N}$, $p \in (0,1)$	$B(n,p)$	$\{0,1,\ldots,n\}$	$w(j) = \binom{n}{j} p^j (1-p)^{n-j}$	np	$np(1-p)$
geometrische Verteilung	$p \in (0,1)$	$G(p)$	$\{1,2,\ldots\}$	$w(j) = (1-p)^{j-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
hypergeometrische Verteilung	$0 < M < N$, $n \in \mathbb{N}$	$H(N,M,n)$	$\{0,1,\ldots,n\}$	$w(j) = \binom{M}{j}\binom{N-M}{n-j} \Big/ \binom{N}{n}$	$n\frac{M}{N}$	$n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$
Poisson-verteilung	$\lambda > 0$	$P(\lambda)$	$\{0,1,2,\ldots\}$	$w(j) = e^{-\lambda}\frac{\lambda^j}{j!}$	λ	λ
Normal-verteilung	$\mu \in \mathbb{R}$, $\sigma > 0$	$N(\mu,\sigma)$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential-verteilung	$\lambda > 0$	$E(\lambda)$	$\mathbb{R}^+$	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma-verteilung	$\lambda > 0$, $r > 0$	$E(\lambda,r)$	$\mathbb{R}^+$	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$
T-Verteilung	$n \in \mathbb{N}$	$T(n)$	$\mathbb{R}$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}$	0	