Wahrscheinlichkeitstheorie

Eine Vorlesung für das

Lehramtsstudium

Franz Hofbauer

SS 1999

Einleitung

Auch wenn man zufällige Ereignisse nicht voraussagen kann, unterliegen sie doch gewissen Gesetzmäßigkeiten. So kann man nicht vorhersehen, wie ein Würfel fällt, weiß aber, dass keine Augenzahl bevorzugt wird. Die Lottoziehung wird so durchgeführt, dass jedes mögliche Ziehungsergebnis die gleiche Chance hat, gezogen zu werden. Und ein Statistiker muss die Meinungsumfrage so planen, dass die Befragten rein zufällig, also ohne irgendwen zu bevorzugen, ausgewählt werden.

Die Aufgabe der Wahrscheinlichkeitstheorie ist es nun, eine mathematische Beschreibung dieser Gesetzmäßigkeiten zu geben und daraus Resultate und Formeln herzuleiten, die zum Lösen praktischer Fragestellungen dienen können. Viele dieser Fragestellungen kommen aus der Statistik, wo man versucht aus einer zufällig gezogenen Stichprobe Schlussfolgerungen zu erhalten. Aber auch Glücksspiele liefern Stoff für die Wahrscheinlichkeitstheorie. Ein Lottospieler möchte schließlich seine Gewinnchancen kennen.

Im vorliegenden Text wird die mathematische Theorie soweit entwickelt, dass man einfache Fragestellungen behandeln kann. Viele Beispiele, die oft aus Schulbüchern stammen, werden durchgerechnet, um das Anwenden der Theorie zu zeigen.

Es wird mit einer mengentheoretischen Betrachtungsweise der Wahrscheinlichkeitstheorie begonnen. Ereignisse werden durch Mengen dargestellt und die Mengenoperatoren erhalten eine wahrscheinlichkeitstheoretische Interpretation. Beim Berechnen der Wahrscheinlichkeit von Ereignissen spielt das Ermitteln von Anzahlen, insbesondere das Abzählen von Stichproben, eine wesentliche Rolle. Ein weiteres wichtiges Hilfsmittel ist die bedingte Wahrscheinlichkeit.

Hat man es mit Anzahlen oder Messwerten zu tun (zum Beispiel die Anzahl der defekten Glühbirnen in einer gezogenen Stichprobe oder die Füllmenge einer zufällig gewählten Waschmittelpackung), dann erweist sich die Verwendung von Zufallsvariablen als nützlich. Ereignisse werden durch Gleichungen und Ungleichungen, die Zufallsvariable enthalten, dargestellt. Man unterscheidet diskrete Zufallsvariable, die in der Regel Anzahlen darstellen, und kontinuierliche Zufallsvariable, die meistens Messwerte bezeichnen. Die Wahrscheinlichkeit von Ereignissen lässt sich im diskreten Fall mit Hilfe des Wahrscheinlichkeitsvektors und im kontinuierlichen Fall mit Hilfe der Wahrscheinlichkeitsdichte der Zufallsvariablen berechnen.

Die wichtigsten Verteilungen wie zum Beispiel die Binomial– und die Normalverteilung und einige andere werden eingeführt und deren Anwendung erläutert. Ebenso werden Erwartungswert und Varianz einer Zufallsvariablen und ihre wichtigsten Eigenschaften behandelt.

Den Abschluss bilden einige statistische Verfahren, nämlich das Berechnen von Vertrauensintervallen und das Durchführen von Tests. Die mathematische Grundlage dafür bildet wieder das Arbeiten mit Zufallsvariablen. Eine wesentliche Rolle spielen dabei die Binomial– und die Normalverteilung.

I. Kombinatorik

Stichproben spielen in der Wahrscheinlichkeitstheorie eine wesentliche Rolle. In der Statistik zum Beispiel besteht die grundlegende Methode ja gerade darin, aus einer zufällig gezogenen Stichprobe Folgerungen zu ziehen. Ein anderes Beispiel für Stichproben ist das Ziehen der Preise bei einem Glücksspiel. Das Ausrechnen der Wahrscheinlichkeit, bei zufälligem Ziehen gewisse Stichproben zu erhalten, führt auf das Problem, die Anzahl aller möglichen Stichproben oder die Anzahl der Stichproben mit einer gewissen Eigenschaft zu ermitteln. Damit setzen wir uns in diesem ersten Teil auseinander. Das Teilgebiet der Mathematik, das sich mit dem Ermitteln von Anzahlen beschäftigt, heißt 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 vor uns, aus der eine Stichprobe gezogen wird (zum Beispiel eine Menge von Glühbirnen, aus der eine Stichprobe zur Qualitätskontrolle gezogen wird, oder eine Menge von Losen, aus der die Preisträger gezogen werden). Mit n bezeichnen wir die Anzahl der Elemente dieser Menge. Aus dieser n-elementigen Menge 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 heißt Stichprobenumfang.

Wir unterscheiden geordnete Stichproben mit und ohne Zurücklegen. Geordnete Stichproben mit Zurücklegen vom Umfang k erhält man, wenn man aus der vorgegebenen n-elementigen Menge der Reihe nach k Elemente zieht und jedes sofort wieder zurücklegt. Es wird also jedes Mal aus der ursprünglichen Menge gezogen. Jedes Element kann öfter in der Stichprobe vorkommen. Geordnete Stichproben ohne Zurücklegen vom Umfang k erhält man, wenn man aus der vorgegebenen n-elementigen Menge der Reihe nach k Elemente zieht und nicht zurücklegt. Es wird also jedes Mal aus einer kleineren Menge gezogen. 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 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 vom Umfang k mit Zurücklegen ist n^k. Die Anzahl der geordneten Stichproben aus einer n-elementigen Menge 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 in Frage, also gibt es n Möglichkeiten. Da wir zurücklegen, wird beim zweiten Mal ebenfalls aus der ursprünglichen Menge 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 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. □

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 einer Klasse mit 20 Schülern werden ein Klassensprecher, sein Stellvertreter und ein Kassier gewählt. Wie viele Möglichkeiten gibt es, wenn Ämterkumulierungen ausgeschlossen sind?

Wir müssen zuerst entscheiden, in welcher Reihenfolge gewählt wird. Wir beginnen mit dem Klassensprecher. Es sind 20 Schüler, daher auch 20 Möglichkeiten, den Klassensprecher zu wählen. Als nächstes wählen wir den Stellvertreter. Da Ämterkumulierungen ausgeschlossen sind, stehen jetzt nur mehr 19 Schüler zur Verfügung, aus denen gewählt werden kann. Für den Stellvertreter gibt es also 19 Möglichkeiten. Es bleiben 18 Schüler, aus denen schließlich der Kassier gewählt wird. Es gibt daher $20 \cdot 19 \cdot 18$ mögliche Ämterverteilungen. Hier handelt es sich also um geordnete Stichproben vom Umfang 3 aus einer 20-elementigen Menge ohne Zurücklegen.

Beispiel 4: 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. Der Gast 1 wird zuerst eingeteilt. Für ihn gibt es 6 mögliche Zimmer. Für den Gast 2, der als nächster drankommt, gibt es dann nur mehr 5 freie Zimmer. Für den Gast 3 gibt es nur mehr 4 und für den Gast 4 nur 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. $\square$

Wir wollen die geordneten Stichproben verallgemeinern, 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. $\square$

Beispiel 5: 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, wo die ersten beiden Zeichen Buchstaben sind? 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 zieht. Die ungeordneten Stichproben vom Umfang k sind also die k-elementigen Teilmengen dieser Menge.

Beispiel 6: 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 also 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 ist $\binom{n}{k}$.*

Beweis: Den Fall $k = 0$ behandeln wir zuerst. Einerseits gibt es genau eine 0-elementige Teilmenge, nämlich die leere Menge, andererseits gilt $\binom{n}{0} = 1$. Also 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. Wir leiten eine Gleichung für a_k her, aus der wir dann a_k berechnen. Dazu stellen wir folgende Überlegung an. Wir stellen uns vor, dass wir eine Liste aller a_k k-elementigen Teilmengen vor uns haben. Wir nehmen die erste dieser Teilmengen und schreiben alle möglichen Anordnungen dieser Teilmenge in eine Zeile. Da sie k Elemente enthält, gibt es $k!$ verschiedene Anordnungen nach Satz 2. Dann nehmen wir die zweite dieser Teilmengen und schreiben alle möglichen Anordnungen dieser Teilmenge in die zweite Zeile darunter. Da sie k Elemente enthält, gibt es ebenfalls $k!$ verschiedene Anordnungen. So machen wir es mit allen Teilmengen. In der a_k-ten Zeile stehen dann die $k!$ möglichen Anordnungen der letzten Teilmenge. Wir haben jetzt alle möglichen Anordnungen aller k-elementigen Teilmengen vor uns, also alle geordneten Stichproben vom Umfang k ohne Zurücklegen aus einer n-elementigen Menge. Ihre Anzahl ist $\frac{n!}{(n-k)!}$ nach Satz 1. Andererseits sind diese geordneten Stichproben in einem Rechteck aus a_k Zeilen und $k!$ Spalten angeordnet. Also ist ihre Anzahl auch gleich $a_k k!$ und wir erhalten $a_k k! = \frac{n!}{(n-k)!}$ durch Gleichsetzen mit der vorhin ermittelten Anzahl. Daraus folgt $a_k = \frac{n!}{k!(n-k)!}$, die gesuchte Formel. $\square$

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 M_1 die Menge aller 3-elementigen Teilmengen aus der Menge der 14 Frauen. Sei M_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 M_1 mit einer der Mengen aus M_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 M_1 ist und $n_2 = \binom{8}{2}$ die Anzahl der Elemente von M_2 ist. Es gibt daher $\binom{14}{3}\binom{8}{2}$ mögliche Ausschüsse mit 3 Frauen und 2 Männern.

Beispiel 8: 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 9: 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 $\heartsuit$-Karten, 3 $\diamondsuit$-Karten und 2 $\clubsuit$-Karten enthalten?

Wir gehen wie in Beispiel 7 vor. Sei M_1 die Menge aller 6-elementigen Teilmengen aus den $\heartsuit$-Karten. Ihre Anzahl ist $\binom{13}{6}$. Sei M_2 die Menge aller 3-elementigen Teilmengen aus den $\diamondsuit$-Karten. Ihre Anzahl ist $\binom{13}{3}$. Sei M_3 die Menge aler 2-elementigen Teilmengen aus den $\clubsuit$-Karten. Ihre Anzahl ist $\binom{13}{2}$. Alle 11-elementigen Teilmengen mit 6 $\heartsuit$-Karten, 3 $\diamondsuit$-Karten und 2 $\clubsuit$-Karten erhält man dadurch, dass man eine 6-elementige Teilmenge aus den $\heartsuit$-Karten, eine 3-elementige Teilmenge aus den $\diamondsuit$-Karten und eine 2-elementige Teilmenge aus den $\clubsuit$-Karten zusammensetzt. Gemäß Satz 3 gibt es dafür $\binom{13}{6}\binom{13}{3}\binom{13}{2}$ Möglichkeiten.

Bemerkung: Die Methode aus Beispiel 7 und aus Beispiel 9 kann man verwenden, um die Gleichung $\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$ zu zeigen, wobei $\binom{u}{v} = 0$ zu setzen ist, wenn $v > u$ gilt. Das geht so: Aus einer Menge von $n+m$ Kugeln, von denen n weiß und m schwarz sind, werden k-elementige Teilmengen gezogen. Die Anzahl der k-elementigen Teilmengen, die j weiße Kugeln und $k-j$ schwarze Kugeln enthalten, ist $\binom{n}{j}\binom{m}{k-j}$, da sich jede dieser Teilmengen aus einer j-elementigen Teilmenge der n weißen Kugeln und einer $k-j$-elementigen Teilmenge der m schwarzen Kugeln zusammensetzen lässt. Die Anzahl aller k-elementigen Teilmengen ist dann $\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j}$. Die Anzahl aller k-elementigen Teilmengen ist aber auch $\binom{n+m}{k}$. Daher gilt $\sum_{j=0}^{k}\binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$.

Wir kehren zu den geordneten Stichproben zurück. Wir suchen jetzt nicht die Anzahl aller Stichproben, sondern die Anzahl derer, die eine zusätzliche Eigenschaft erfüllen. Dabei kommt Satz 4 ins Spiel. Um das zu sehen, rechnen wir folgendes Beispiel.

Beispiel 10: Im Verein aus Beispiel 7, der 14 Frauen und 8 Männer als Mitglieder hat, werden 5 Preise verlost, ein erster, ein zweiter, ein dritter, ein vierter und ein fünfter. Wie viele Verteilungen der Preise gibt es, sodass unter den Preisträgern 3 Frauen und 2 Männer sind? Anders formuliert: Wie viele verschiedene geordnete Stichproben vom Umfang 5 mit Zurücklegen gibt es, die 3 Frauen und 2 Männer enthalten? Wie viele gibt es, wenn nicht zurückgelegt wird, das heißt wenn Preisträger von der weiteren Verlosung ausgeschlossen sind?

Wir stellen uns 5 Plätze vor, auf die die ausgelosten Frauen und Männer gestellt werden. Von diesen 5 Plätzen werden 3 mit Frauen (W) und 2 mit Männern (M) besetzt. Es gibt dafür folgende Möglichkeiten: WWWMM WWMWM WWMMW WMWWM WMWMW WMMWW MWWWM MWWMW MWMWW MMWWW.

Wie viele geordnete Stichproben gibt es für die Platzaufteilung WWWMM, also wenn die ertsten 3 Plätze mit Frauen und die letzten 2 Plätze mit Männern besetzt werden. Für den ersten Platz kommem 14 Frauen in Frage, ebenso für den zweiten und dritten, da zurückgelegt wird. Für den vierten Platz kommen 8 Männer in Frage und ebenso für den fünften. Es gibt also $14 \cdot 14 \cdot 14 \cdot 8 \cdot 8$ Stichproben, die Platzaufteilung WWWMM haben.

Ebenso kann man die Anzahl der Stichproben für jede der anderen Platzaufteilungen ausrechnen. Die Anzahl für die Platzaufteilung WWMWM ist $14 \cdot 14 \cdot 8 \cdot 14 \cdot 8$, die Anzahl für WWMMW ist $14 \cdot 14 \cdot 8 \cdot 8 \cdot 14$, und so weiter. Man sieht, dass die Anzahl der Stichproben für jede Platzaufteilung dieselbe ist, nämlich $14^3 8^2$. Da es insgesamt 10 verschiedene Platzaufteilungen gibt, ist $10 \cdot 14^3 8^2$ die gesuchte Anzahl der geordneten Stichproben vom Umfang 5 mit Zurücklegen, die 3 Frauen und 2 Männer enthalten.

Wird nicht zurückgelegt, dann ist $14 \cdot 13 \cdot 12 \cdot 8 \cdot 7$ die Anzahl der Stichproben, die Platzaufteilung WWWMM haben, da für den zweiten Platz nur mehr die 13 übriggebliebenen Frauen in Frage kommen und für den dritten Platz nur mehr die 12 übriggebliebenen. Ebenso kommen für den fünften Platz nur mehr die 7 übriggebliebenen Männer in Frage. Die Anzahl der Stichproben mit Platzaufteilung WWMWM ist $14 \cdot 13 \cdot 8 \cdot 12 \cdot 7$ und die für WMWMW ist $14 \cdot 8 \cdot 13 \cdot 7 \cdot 12$. Man sieht wieder, dass die Anzahl der Stichproben für jede Platzaufteilung dieselbe ist, nämlich $14 \cdot 13 \cdot 12 \cdot 8 \cdot 7$. Daher ist $10 \cdot 14 \cdot 13 \cdot 12 \cdot 8 \cdot 7$ die gesuchte Anzahl der geordneten Stichproben vom Umfang 5 ohne Zurücklegen, die 3 Frauen und 2 Männer 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 W 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 M statt mit W anstellt, dann erhält man als Anzahl $\binom{5}{2}$, also ebenfalls 10.)

Beispiel 11: 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 Frauen und Männern, statt $\heartsuit$-Karten und $\diamondsuit$-Karten, verwenden wir jetzt rote und grüne Kugeln. Dabei muss man sich jedoch auch gleichfarbige Kugeln als unterscheidbar vorstellen, so wie die Frauen voneinander unterscheidbar sind, und wie auch die $\heartsuit$-Karten voneinander unterscheidbar sind.

Beispiel 12: 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
(i) ungeordnet gezogen wird?
(ii) geordnet mit Zurücklegen gezogen wird?
(iii) geordnet ohne Zurücklegen gezogen wird?

(i) 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 $\binom{9}{3}\binom{5}{4}$.

(ii) 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 der geordneten Stichproben mit Zurücklegen, die 3 rote und 4 grüne Kugeln enthalten, ist daher $\binom{7}{3}9^3 5^4$.

(iii) Hier kann man genauso vorgehen wie in (ii). Da jetzt aber nicht zurückgelegt wird, gibt es für jede der $\binom{7}{3}$ Platzaufteilungen nur mehr $9 \cdot 8 \cdot 7 \cdot 5 \cdot 4 \cdot 3 \cdot 2$ mögliche Besetzungen mit Kugeln. Die Anzahl der 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 aber auch auf andere Weise erhalten, indem man von (i) ausgeht. Man erhält die geordneten Stichproben ohne Zurücklegen nämlich dadurch, dass man die ungeordneten Stichproben aus (i) 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 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 11 kommen nur $\heartsuit$-Karten und $\diamondsuit$-Karten in der Stichprobe vor. Wie kann man dieses Beispiel verallgemeinern und nach der Anzahl der Stichproben fragen, die eine bestimmte Anzahl von ♡-Karten, von ♢-Karten, von ♣-Karten und von ♠-Karten enthalten. Das kann man genauso behandeln, wie es in Beispiel 11 getan wurde, nur wissen wir noch nicht, wie man die Anzahl der möglichen Platzaufteilungen bestimmt. Eine Platzaufteilung entspricht dann nicht mehr der Auswahl einer Teilmenge von vorgegebenem Umfang, sondern der Zerlegung der Menge der Plätze in Teilmengen, wobei die Anzahl der Elemente dieser Teilmengen vorgegeben ist. Mit solchen Zerlegungen einer Menge wollen wir uns jetzt beschäftigen.

Vorgegeben ist eine n-elementige Menge. 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 13: 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 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 zu wählen. Es verbleiben $n - k_1$ Elemente in der Menge. Es gibt dann $\binom{n-k_1}{k_2}$ Möglichkeiten, die zweite Teilmenge zu wählen. Es bleiben $n - k_1 - k_2$ Elemente in der Menge. 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 $1 = \binom{k_j}{k_j} = \binom{n-k_1-\cdots-k_{j-1}}{k_j}$ Möglichkeit für die letzte Teilmenge. Die Anzahl aller möglichen Zerlegungen ist also

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\cdots-k_{j-1}}{k_j} = \frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdots\frac{(n-k_1-\cdots-k_{j-1})!}{k_j!0!} = \frac{n!}{k_1!k_2!\ldots k_j!}.$$

Damit ist die gewünschte Formel gefunden. □

Beispiel 14: An einem Kinderfest nehmen 22 Mädchen teil. Wie viele Möglichkeiten gibt es, die 22 Mädchen in 11 Tanzpaare aufzuteilen?

Gefragt ist, auf wie viele Arten es möglich ist, eine 22-elementige Menge in 11 Teilmengen aufzuteilen, von denen jede 2 Elemente enthält. Satz 5 gibt die Antwort auf diese Frage. Es ist auf $\frac{22!}{(2!)^{11}}$ verschiedene Arten möglich.

Beispiel 15: 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 11. 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? Haben wir n Kugeln, von denen k_1 die Farbe 1, k_2 die Farbe 2, ... und k_j die Farbe j haben, wobei $n = k_1 + k_2 + \cdots + k_j$ gilt, wie viele verschiedene Anordnungen dieser n Kugeln gibt es dann, wenn Anordnungen, die durch Vertauschen gleichfarbiger Kugeln ineinander übergehen, als gleich gelten? Hier geht es um die Anordnung der Farben, nicht um die der Kugeln. Wir schauen uns zuerst ein Beispiel an.

Beispiel 16: Es sollen alle Anordnungen von 5 Kugeln aufgeschrieben werden, von denen zwei rot (R), eine blau (B) und zwei weiß (W) sind?

R R B W W	R R W B W	R R W W B
R B R W W	R W R B W	R W R W B
R B W R W	R W B R W	R W W R B
R B W W R	R W B W R	R W W B R
B R R W W	W R R B W	W R R W B
B R W R W	W R B R W	W R W R B
B R W W R	W R B W R	W R W B R
B W R R W	W B R R W	W W R R B
B W R W R	W B R W R	W W R B R
B W W R R	W B W R R	W W B R R

Wie kommt man zu den Anordnungen der 5 Kugeln in diesem Beispiel? Wie wir schon in Beispiel 15 gesehen haben, entsprechen die Anordnungen der Kugeln den Zerlegungen der Menge der Plätze. Die beiden Plätze für die roten Kugeln kann man auf $\binom{5}{2}$ Arten aus den insgesamt 5 Plätzen wählen. Aus den verbleibenden 3 Plätzen kann man den Platz für die blaue Kugel auf $\binom{3}{1}$ Arten wählen. Für die beiden weißen Kugeln bleiben zwei Plätze übrig, also nur mehr $1 = \binom{2}{2}$ Möglichkeit. Insgesamt gibt es also $\binom{5}{2}\binom{3}{1}\binom{2}{2} = 30$ mögliche Anordnungen der 5 Kugeln.

Jede Anordnung der Kugeln entspricht genau einer Zerlegung einer 5-elementigen Menge in 3 Teilmengen, von denen die erste 2, die zweite 1 und die dritte 2 Elemente enthält. Die Plätze der roten Kugeln bilden die erste Teilmenge, der Platz der blauen Kugel bildet die zweite Teilmenge und die Plätze der weißen Kugeln bilden die dritte Teilmenge. Der Beweis des folgenden Satzes ist deshalb derselbe wie der von Satz 5.

Satz 6: *Sei $n = k_1 + k_2 + \cdots + k_j$. Wir haben n Kugeln, von denen k_1 die Farbe 1, k_2 die Farbe 2, ... und k_j die Farbe j haben. Die Anzahl der möglichen Anordnungen dieser Kugeln ist $\frac{n!}{k_1!k_2!k_3!\ldots k_j!}$.*

Beweis: Es gibt $\binom{n}{k_1}$ Möglichkeiten, die Plätze für die Kugeln der Farbe 1 zu wählen. Es bleiben $n - k_1$ Plätze übrig. Es gibt dann $\binom{n-k_1}{k_2}$ Möglichkeiten, die Plätze für die Kugeln der Farbe 2 zu wählen. So tut man weiter. Für die Kugeln der Farbe j bleiben $n - k_1 - \cdots - k_{j-1} = k_j$ Plätze übrig. Daher gibt es nur $1 = \binom{k_j}{k_j} = \binom{n-k_1-\cdots-k_{j-1}}{k_j}$ Möglichkeit, für diese die Plätze zu wählen. Die Anzahl aller Anordnungen ist also $\binom{n}{k_1}\binom{n-k_1}{k_2}\ldots\binom{n-k_1-\cdots-k_{j-1}}{k_j} = \frac{n!}{k_1!(n-k_1)!}\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdots\frac{(n-k_1-\cdots-k_{j-1})!}{k_j!0!} = \frac{n!}{k_1!k_2!\ldots k_j!}$. Damit ist die gewünschte Formel gefunden. □

Beispiel 17: 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 18: Jemand hat in seinem Zimmer 8 Lampen. Er hat 8 Glühbirnen, 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 Glühbirnen angeordnet werden. Die verschiedenen Beleuchtungsmöglichkeiten sind die verschiedenen Anordnungen der Glühbirnen. Daher gibt es $\frac{8!}{2!2!2!2!} = 2520$ mögliche Beleuchtungen.

Die Anzahl der möglichen Beleuchtungen mit einer roten Glühbirne in der Tischlampe erhält man, indem man die übrigen 7 Glühbirnen auf die übrigen 7 Lampen verteilt. Diese Anzahl ist $\frac{7!}{2!1!2!2!} = 630$. Daher gibt es $2520-630 = 1890$ mögliche Beleuchtungen, wo die Tischlampe keine rote Glühbirne erhält.

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. Zuvor jedoch müssen wir uns damit beschäftigen, was wir unter einem Ereignis verstehen und welche mathematische Sprache wir verwenden wollen. Dafür bietet sich sowohl die logische Sprache als auch die Mengensprache an.

Ist die Definition von Ereignis und Wahrscheinlichkeit gelungen, dann kann man Sätze und Formeln beweisen und diese zum Rechnen von Beispielen verwenden. Der wichtigste Satz ist der Additionssatz, aus dem weitere Eigenschaften der Wahrscheinlichkeit folgen und der in vielen Beweisen Verwendung findet. Die Formel für gleichwahrscheinliche Ausfälle führt das Berechnen von Wahrscheinlichkeiten auf das Abzählen von Mengen zurück, wozu die Formeln aus der Kombinatorik nützlich sind.

Hat man teilweise Information über den Ausfall des Zufallsexperiments, dann verwendet man die bedingte Wahrscheinlichkeit, die die vorhandene Information mitberücksichtigt. Mit Hilfe der bedingten Wahrscheinlichkeit erhält man wieder Sätze und Formeln, um damit Beispiele zu rechnen. Der Multiplikationssatz tritt oft in Kombination mit dem Additionssatz auf. Ebenso häufig findet man Anwendungen, die mit der Formel für die totale Wahrscheinlichkeit und und der Formel von Bayes gerechnet werden.

5. Zufallsexperiment, Ausfall, Ereignis

Ein Zufallsexperiment (Zufallsversuch) 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 aus der Menge der ersten 45 natürlichen Zahlen. Die Menge aller möglichen Ausfälle fassen wir zu einer Menge zusammen, die üblicherweise mit Ω bezeichnet wird. Wir nennen Ω die Ausfallsmenge des Zufallsexperiments. 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 $\{2, 4, 6\}$. Wie kommt man zu der Teilmenge von Ω, die das Ereignis darstellt? Das Zufallsexperiment liefert immer nur einen Ausfall, an dem man erkennt, ob das Ereignis eingetreten ist. Diese Ausfälle, die das Eintreten des Ereignisses bedeuten, fasst man zu einer Menge zusammen, die dann das Ereignis darstellt. 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.

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 auch 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. Welchen Ausfall wir auch erhalten, er liegt nicht in der leeren Menge. Man spricht daher auch vom unmöglichen Ereignis. Die Menge Ω ist das Ereignis, das immer eintritt. Welchen Ausfall wir auch erhalten, er liegt in Ω. 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.

6. Wahrscheinlichkeit

Die Wahrscheinlichkeit eines Ereignisses ist ein Maß 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. Die Anzahl der Wiederholungen des Zufallsexperiments, bei denen A eintritt, 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

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

wobei wir die Existenz des Grenzwertes einfach annehmen. Die Wahrscheinlichkeit $P(A)$ eines Ereignisses A ist also eine Maßzahl für die Häufigkeit des Eintretens von A.

Wir beweisen die grundlegenden Eigenschaften und Rechenregeln für die Wahrscheinlichkeit.

Satz 7: *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 $0 \le \frac{N_k(A)}{k} \le 1$. Lässt man k gegen ∞ gehen, so 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 8 (Additionssatz): *Die Ereignisse $A_1, A_2, \dots, A_n$ seien unvereinbar (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 ja die Ereignisse 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 $A_1, A_2, \ldots, A_n$ 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. $\square$

Manchmal ist dieser Additionssatz für endlich viele Ereignisse nicht ausreichend. Man benötigt eine Version für unendlich viele Ereignisse: Sind $A_1, A_2, A_3, \ldots$ Ereignisse, die unvereinbar sind, dann gilt

$$P\Big(\bigcup\nolimits_{j=1}^{\infty} A_j\Big) = \sum\nolimits_{j=1}^{\infty} P(A_j)$$

Diesen Additionssatz für unendlich viele Ereignisse verwenden wir hier nicht zum Rechnen von Beispielen. Er ist aber notwendig zum Beweis einiger Sätze.

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

Satz 9: *Seien A und B Ereignisse. Dann gilt*
(a) $P(A \setminus B) = P(A) - P(A \cap B)$
(b) $B \subset A \;\Rightarrow\; P(A \setminus B) = P(A) - P(B)$
(c) $B \subset A \;\Rightarrow\; P(B) \le P(A)$
(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) Wenn $B \subset A$, dann gilt $A \cap B = B$. 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 7 (a) folgt $P(A \setminus B) \ge 0$, 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 7 (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 $P(A \cap B) \ge 0$ wegen Satz 7 (a) gilt, folgt (f) aus (e). $\square$

7. 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 10 (Formel für gleichwahrscheinliche Ausfälle): *Bei einem Zufallsexperiment sei die Ausfallsmenge Ω endlich und alle Ausfälle haben gleiche Wahrscheinlichkeit. Sei $A \subset \Omega$ ein Ereignis. Dann gilt*

$$P(A) = |A|/|\Omega|$$

Beweis: Sei q die Wahrscheinlichkeit, mit der jeder der Ausfälle eintritt, oder genauer, mit der jedes Ereignis, das nur aus einem Ausfall besteht, eintritt. Sei $B \subset \Omega$ ein beliebiges Ereignis und $k = |B|$ die Anzahl der Elemente von B. Seien $B_1, B_2, \ldots, B_k$ die einelementigen Teilmengen von B. Da diese disjunkt sind und ihre Vereinigung B ist, folgt aus dem Additionssatz, dass $P(B) = P(B_1) + P(B_2) + \cdots + P(B_k)$ gilt. Da die Ereignisse B_j einelementig sind, gilt $P(B_j) = q$ für $1 \le j \le k$. Es folgt $P(B) = qk = q|B|$.
Für $B = \Omega$ heißt das $P(\Omega) = q|\Omega|$. Nach Satz 7 gilt $P(\Omega) = 1$, sodass $q = \frac{1}{|\Omega|}$ folgt. Für $B = A$ erhalten wir jetzt $P(A) = q|A| = \frac{|A|}{|\Omega|}$, was zu zeigen war. $\square$

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

Beispiel 19: 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 20: Es wird mit 2 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit,
(i) Augensumme 5 zu erhalten?
(ii) dass mindestens eine 6 auftritt?

Die Ausfallsmenge ist $\Omega = \{(i, j) : 1 \le i, j \le 6\}$, die Menge aller Paare von Augenzahlen. Sie hat $6^2 = 36$ Elemente. Alle Ausfälle sind gleichwahrscheinlich. Wir können Satz 10 anwenden.

(i) 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 $P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36}$.

(ii) 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 21: Es wird mit 3 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit,
(i) Augensumme 5 zu erhalten?
(ii) dass keine 6 auftritt?
(iii) 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 10 anwenden.

(i) 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 $P(A) = \frac{|A|}{|\Omega|} = \frac{6}{216}$.

(ii) 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 $P(A) = \frac{|A|}{|\Omega|} = \frac{125}{216}$.

(iii) 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 22: 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 10 anwenden.

Beispiel 23: 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 ist $|\Omega| = \binom{45}{6}$. 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 24: 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 ist $|\Omega| = \binom{20}{7}$. 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.

8. 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 10 verwenden wir wieder die Formel

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

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 25: 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. Das gefragte Ereignis ist $A = (2, 4)$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{12}$.

Beispiel 26: Zwei Personen treffen folgende Vereinbarung. Jeder kommt zufällig zwischen 5 und 6 Uhr und wartet eine Viertelstunde. 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 jeder 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 27: 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 $t + \int_t^1 \frac{t}{x}\, dx = t - t \ln t$. Wir erhalten $P(A) = \frac{|A|}{|\Omega|} = t - t \ln t$.

9. 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 6 verwenden. Der bedingte Anteil oder die bedingte relative Häufigkeit des

Eintretens von A unter B ist dann $N_k(A \cap B)/N_k(B)$, da ja nur die Wiederholungen des Zufallsexperiments berücksichtigt werden, bei denen B eintritt. 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 des Ereignisses 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.

Zur Illustration sei ein Beispiel angeführt. In einer Fabrik werden von drei Maschinen Schrauben erzeugt, die in einem gemeinsamen Vorratsbehälter landen. Man fragt nach der Wahrscheinlichkeit, dass eine Schraube defekt ist. Will man das für die drei Maschinen getrennt tun, dann kommt die bedingte Wahrscheinlichkeit ins Spiel. Das Zufallsexperiment ist jetzt die Produktion einer Schraube mit den möglichen Ausfällen "brauchbar" und "unbrauchbar". Sei A das Ereignis "Schraube ist unbrauchbar" und B das Ereignis "Schraube stammt von Maschine 1". Hat man k Schrauben aus der Gesamtproduktion geprüft, dann ist $N_k(A \cap B)$ die Anzahl der unbrauchbaren Schrauben unter denen, die von Maschine 1 stammen, und die bedingte relative Häufigkeit $N_k(A \cap B)/N_k(B)$ ist der Anteil der unbrauchbaren Schrauben unter denen, die von Maschine 1 stammen. Als Grenzwert für $k \to \infty$ erhalten wir die bedingte Wahrscheinlichkeit $P(A|B)$, das ist die Wahrscheinlichkeit, dass eine von Maschine 1 produzierte Schraube unbrauchbar ist.

Der folgende Satz gibt eine Formel für die bedingte Wahrscheinlichkeit.

Satz 11: *Seien A und B Ereignisse mit $P(B) \neq 0$. Dann gilt*
(a) $P(A|B) = P(A \cap B)/P(B)$
(b) *A und B sind unabhängig* $\Leftrightarrow P(A \cap B) = P(A)P(B)$

Beweis: (a) folgt aus den Definitionen von Wahrscheinlichkeit und bedingter Wahrscheinlichkeit: $P(A|B) = \lim_{k \to \infty} \frac{N_k(A \cap B)}{N_k(B)} = \lim_{k \to \infty} \frac{N_k(A \cap B)/k}{N_k(B)/k} = P(A \cap B)/P(B)$.
Die Definition von unabhängig war $P(A|B) = P(A)$. Wegen (a) ist das äquivalent zu $P(A \cap B) = P(A)P(B)$, womit auch (b) gezeigt ist. □

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 $P(A|B)$. Kaufe ich eine Schraube aus der Fabrik in obigem Beispiel, dann werde ich als Wahrscheinlichkeit, dass sie unbrauchbar ist, die bedingte Wahrscheinlichkeit verwenden, wenn ich zusätzlich weiß, dass sie von Maschine 1 stammt. Dazu folgende Beispiele und zuvor ein Satz, der bei der Berechnung der bedingten Wahrscheinlichkeit hilft.

Satz 12: *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 $P(B) = \frac{|B|}{|\Omega|}$. Aus Satz 11 (a) folgt jetzt die gewünschte Formel. □

Beispiel 28: 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 ist nach dem Ereignis A, dass mindestens einmal 6 aufgetreten ist. Das Ereignis A wurde in Beispiel 20 (ii) bestimmt. Wir haben die Information, dass die Augensumme 7 ist, also 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 12 folgt $P(A|B) = \frac{2}{6}$.

Beispiel 29: Zwei Zahlen werden zufällig in $[0,1]$ gewählt. Man erfährt, dass ihr Abstand $\leq \frac{1}{3}$ ist. Wie groß ist die Wahrscheinlichkeit, dass eine der beiden $\leq \frac{1}{3}$ ist?

Die Ausfallsmenge ist $\Omega = [0,1] \times [0,1]$. Das Ereignis "Abstand der beiden Punkte $\leq \frac{1}{3}$" ist die Menge $\{(x,y) \in \Omega : x - \frac{1}{3} \leq y \leq x + \frac{1}{3}\}$. Die Fläche $|B|$ von B ist $1 - \frac{2}{3} \cdot \frac{2}{3} = \frac{5}{9}$. Das Ereignis "einer der Punkte $\leq \frac{1}{3}$" ist $[0,1] \times [0,\frac{1}{3}] \cup [0,\frac{1}{3}] \times [\frac{1}{3},1]$. Da $A \cap B$ aus einem Quadrat der Seitenlänge $\frac{1}{3}$ und aus zwei Dreiecken, die sich ebenfalls zu einem Quadrat der Seitenlänge $\frac{1}{3}$ zusammensetzen lassen, besteht, ist $|A \cap B| = \frac{2}{9}$. Aus Satz 12 folgt $P(A|B) = \frac{2}{5}$.

Der wohl wichtigste Satz, der mit bedingten Wahrscheinlichkeiten zu tun hat, ist der sogenannte Multiplikationssatz. Er berechnet die Wahrscheinlichkeit des Durchschnitts von Ereignissen.

Satz 13 (Multiplikationssatz): *Für Ereignisse $A_1, A_2, \ldots, A_n$ gilt*

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

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

Beweis: Satz 11 (a) liefert $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$, 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 $P(A_n|A_1 \cap \cdots \cap A_{n-1}) = \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})}$. 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 30: 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?

Hier ist nach drei Ereignissen gefragt und der Wahrscheinlichkeit, dass alle drei eintreten. Wir bezeichnen diese Ereignisse mit A_1, A_2 und A_3, das heißt A_1 ist das Ereignis "erste Kugel ist rot", A_2 ist das Ereignis "zweite Kugel ist gelb" und A_3 ist 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 $P(A_1) = \frac{2}{7}$.

Um $P(A_2|A_1)$ zu berechnen, müssen wir zuerst überlegen, welche Information uns das als Bedingung auftretende Ereignis A_1 liefert. Es 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.

Um $P(A_3|A_1 \cap A_2)$ zu berechnen, müssen wir wieder überlegen, was das als Bedingung auftretende Ereignis $A_1 \cap A_2$ besagt. Es 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. Wir wollen hier eine Tabellenschreibweise verwenden.

Beispiel 31: 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 30 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 Hilfe des Additionssatzes 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 32: 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 oder 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 $\frac{40}{243}$.

Man kann Beispiel 32 mit Beispiel 10 vergleichen. Diese beiden Beispiele sind sehr ähnlich. In Beispiel 32 wird ein Ereignis in zehn Teilereignisse zerlegt. Es werden die Wahrscheinlichkeiten der Teilereignisse berechnet und aufsummiert, um die gesuchte Wahrscheinlichkeit zu erhalten. In Beispiel 10 wird ein Menge von geordneten Stichproben in zehn Teilmengen zerlegt. Es werden die Anzahlen der Teilmengen berechnet und aufsummiert, um die gesuchte Anzahl zu erhalten. Die Zerlegung in Teilereignisse beziehungsweise Teilmengen ist dieselbe. Auch das Berechnen der Wahrscheinlichkeiten beziehungsweise Anzahlen ist ähnlich.

10. Totale Wahrscheinlichkeit

Manches Mal sind für ein Ereignis A bedingte Wahrscheinlichkeiten vorgegeben und man soll daraus die (totale) Wahrscheinlichkeit des Ereignisses A bestimmen. Es kann dann auch noch nach einer Wahrscheinlichkeit gefragt sein, bei der A als Bedingung auftritt. Wir leiten die dafür notwendigen Formeln her. Dazu benötigen wir folgende Definition.

Definition: Man sagt, die Ereignisse $B_1, B_2, \dots, B_n$ bilden eine Zerlegung der Ausfallsmenge Ω, wenn sie disjunkt sind und wenn $\bigcup_{j=1}^n B_j = \Omega$ gilt. Dass die Ereignisse $B_1, B_2, \dots, B_n$ eine Zerlegung bilden, ist gleichbedeutend damit, dass genau eines der Ereignisse eintritt.

Satz 14: *Seien $B_1, B_2, \dots, B_n$ eine Zerlegung von Ω und $A \subset \Omega$ 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* $1 \le j \le n$
Man nennt (a) die Formel für die totale Wahrscheinlichkeit und (b) die Formel von Bayes.

Beweis: Um (a) zu zeigen, verwenden wir zuerst Satz 11 (a). Aus diesem Satz erhalten wir $\sum_{j=1}^n P(A|B_j)P(B_j) = \sum_{j=1}^n P(A \cap B_j)$. Da die Ereignisse B_j disjunkt sind, sind es auch die Ereignisse $A \cap B_j$ und es folgt $\sum_{j=1}^n P(A \cap B_j) = P(\bigcup_{j=1}^n (A \cap B_j))$ aus dem Additionssatz. Wegen $\bigcup_{j=1}^n B_j = \Omega$ erhalten wir $\bigcup_{j=1}^n (A \cap B_j) = A \cap \bigcup_{j=1}^n B_j = A$ und (a) ist bewiesen.
Aus Satz 11 (a) erhalten wir, dass $P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}$ und $P(A|B_j) = \frac{P(A \cap B_j)}{P(B_j)}$ gilt. Daraus folgt (b) sofort. □

Wir rechnen einige typische Beispiele zu diesen Formeln.

Beispiel 33: 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?

Gefragt ist nach der Wahrscheinlichkeit des Ereignisses A, dass das Stück Ausschuss ist. Wir wissen, wie viel jede Maschine zur Produktion beiträgt und wie viel Ausschuss jede der Maschinen liefert.

Sei B_1 das Ereignis, dass das zufällig gewählte 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 $P(A|B_3) = \frac{5}{100}$. Jetzt können wir in die Formel für die totale Wahrscheinlichkeit einsetzen und erhalten

$P(A) = P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+P(A|B_3)P(B_3) = \frac{2}{100}\frac{4}{25}+\frac{1}{100}\frac{6}{25}+\frac{5}{100}\frac{15}{25} = \frac{89}{2500}$

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

Beispiel 34: 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?

Aus der Angabe des Beispiels erkennt man, dass eine Zerlegung vorliegt. 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.

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 35: 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 36: Eine Firma stellt Computerchips her, wobei der Ausschussanteil 30 % beträgt. Damit diese nicht in den Verkauf gehen, findet eine Kontrolle statt, bei der ein defekter Computerchip mit Wahrscheinlichkeit 0.97 als solcher erkannt und ausgeschieden wird, bei der aber auch ein intakter Computerchip mit Wahrscheinlichkeit 0.05 irrtümlich ausgeschieden wird. Wie groß ist die Wahrscheinlichkeit, dass ein Computerchip, der in den Verkauf geht, intakt ist?

Gesucht ist die Wahrscheinlichkeit, dass ein (zufällig gewählter) Computerchip, der in den Verkauf, also durch die Kontrolle geht, intakt ist. Es ist also eine bedingte Wahrscheinlichkeit zu berechnen.

Was ist gegeben? Wir haben defekte und intakte Computerchips. Sei also B_1 das Ereignis, dass ein (zufällig gewählter) Computerchip intakt ist, und B_2 das Ereignis, dass ein Computerchip 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 Computerchip 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 Computerchip durch die Kontrolle geht. Weiters erhalten wir $P(A|B_2) = 0.03$. Das ist die Wahrscheinlichkeit, dass ein defekter Computerchip durch die Kontrolle geht. Die Formel von Bayes besagt, dass $P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)}$. Zuvor müssen wir $P(A)$ 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 folgt $P(B_1|A) = \frac{0.95 \cdot 0.7}{0.674} = 0.987$.

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

Beispiel 37: 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 von Ω. Aus der Angabe ersehen wir, dass $P(B_1) = \frac{3}{7}$ und $P(B_2) = \frac{4}{7}$. Sei A das Ereignis, dass der Empfänger Punkt registriert. Dann besagt die Angabe, dass $P(A|B_1) = \frac{8}{9}$ und $P(A|B_2) = \frac{1}{8}$. 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 der Empfänger auch Punkt registriert, wenn Punkt gesendet wurde, 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. Das Bestimmen der Wahrscheinlichkeiten von Ereignissen wird dann auf das Berechnen von Summen oder Integralen zurückgeführt. Weiters lässt sich in diesem Fall der Mittelwert (Erwartungswert) der durch die Zufallsvariable bezeichneten Ausfälle definieren und ebenso die Varianz, die angibt, wie stark die Ausfälle um den Mittelwert schwanken.

11. Darstellung von Ereignissen durch Zufallsvariable

Was ist eine Variable? Wozu sind Variable gut? Man kann Mathematik auch ohne Variable betreiben. Das würde so aussehen: Der Flächeninhalt eines Dreiecks ist gleich der Länge einer Seite multipliziert mit der Länge der Höhe auf diese Seite dividiert durch zwei. Das würde kompliziert werden. So führt man Namen ein. Wir bezeichnen die Länge einer Seite mit c, die Länge der zugehörigen Höhe mit h und den Flächeninhalt mit F. Verwendet man auch noch das Gleichheitszeichen, den Multiplikationspunkt und den Bruchstrich, dann lässt sich der oben beschriebene Zusammenhang sehr einfach ausdrücken, nämlich durch $F = \frac{c \cdot h}{2}$. Das hat den zusätzlichen Vorteil, dass man leicht damit rechnen kann. Durch Umformen erhält man zum Beispiel $h = \frac{2F}{c}$.

In der Wahrscheinlichkeitstheorie macht man es genauso. Um das Ereignis "höchstens k defekte Glühbirnen in der Stichprobe" einfach aufschreiben zu können, bezeichnet man die Anzahl der defekten Glühbirnen 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. Im soeben angeführten Beispiel war das Zufallsexperiment das Ziehen einer Stichprobe aus einer Menge von Glühbirnen, wobei die Anzahl der defekten Glühbirnen in der Stichprobe als Ausfall des Zufallsexperiments angesehen wird. Zufallsvariable werden üblicherweise mit Großbuchstaben bezeichnet.

Als weiteres Beispiel wählen wir das Zufallsexperiment Würfeln. Die Zufallsvariable Y bezeichne den Ausfall des Würfels. Das Ereignis, eine Zahl kleiner oder gleich 3 zu würfeln, wird dann durch $Y \leq 3$ dargestellt. Das Ereignis, eine gerade Zahl zu würfeln, wird durch $Y \in \{2, 4, 6\}$ dargestellt.

Mit Variablen kann man auch rechnen. Daher kann man auch mit Ereignissen, die durch Zufallsvariable dargestellt werden, rechnen. Es gilt zum Beispiel

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

Durch diese drei Ungleichungen werden die gleichen Ereignisse dargestellt. 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. Genauso, wie wir eingangs die Beischreibung der Dreiecksflächenberechnung in eine mathematische Darstellung mit Hilfe von Variablen übertragen haben, wird auch die umgangssprachliche Beischreibung von Ereignissen in eine mathematische Darstellung mit Hilfe von Zufallsvariablen übersetzt. Um die so dargestellten Ereignisse miteinander zu verknüpfen, verwenden wir die logische Sprache. Der Additionssatz und die daraus folgenden Rechenregeln aus Satz 9 gelten natürlich weiterhin, da sie ja nicht davon abhängen, wie man Ereignisse darstellt. So gilt zum Beispiel

$$X \le 3 \Leftrightarrow X \le 2 \ \vee \ X \in (2,3]$$

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

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

Ebenso ist $X \le 3$ das Gegenereignis zu $X > 3$. Aus Satz 9 (d) folgt daher

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

Im Folgenden werden wir solche Umformungen häufig verwenden.

12. 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 diskreten, das heißt endlichen oder abzählbaren, Menge, während sie bei einer kontinuierlichen Zufallsvariable ein (oft unbeschränktes) Intervall ausfüllen.

Definition: Eine Zufallsvariable X heißt diskret, wenn sie Werte in einer endlichen oder abzählbaren Menge R annimmt. Wir nennen R den Wertebereich der Zufallsvariablen. Für $i \in R$ definieren wir $w(i) = P(X = i)$. Wir fassen diese Werte zu einem Vektor $(w(i))_{i \in R}$ zusammen, den wir den Wahrscheinlichkeitsvektor der Zufallsvariablen X nennen.

Beispiel 38: Die Zufallsvariable X bezeichne den Ausfall eines Würfels. Man bestimme Wertebereich und Wahrscheinlichkeitsvektor von X.

Die möglichen Werte von X sind die Augenzahlen. Also ist $R = \{1,2,3,4,5,6\}$ der Wertebereich. Es gilt $w(i) = P(X = i) = \frac{1}{6}$ für alle $i \in R$, da ja jede Augenzahl mit Wahrscheinlichkeit $\frac{1}{6}$ auftritt.

Kennt man den Wahrscheinlichkeitsvektor einer Zufallsvariablen oder hat ihn mit Hilfe der Methoden aus den letzten Kapiteln bestimmt, so kann man damit Wahrscheinlichkeiten von Ereignissen berechnen.

Satz 15: *Sei $(w(i))_{i \in R}$ der Wahrscheinlichkeitsvektor einer diskreten Zufallsvariablen X. Dann gilt*
(a) $w(i) \ge 0$ *für* $i \in R$ *und* $\sum_{i \in R} w(i) = 1$
(b) *für* $B \subset R$ *gilt* $P(X \in B) = \sum_{i \in B} w(i)$

Beweis: Wir zeigen zuerst (b). Sei $B = \{i_1, i_2, \dots\}$. Das ist eine endliche oder abzählbare Menge. Es gilt

$$X \in B \ \Leftrightarrow \ X = i_1 \ \vee \ X = i_2 \ \vee \ \dots$$

Da die rechtsstehenden Ereignisse unvereinbar sind, folgt aus dem Additionssatz

$$P(X \in B) = P(X = i_1) + P(X = i_2) + \cdots = w(i_1) + w(i_2) + \cdots = \sum\nolimits_{i \in B} w(i)$$

Damit ist (b) gezeigt.

Die erste Aussage von (a) ist klar, da $w(i) = P(X = i)$ eine Wahrscheinlichkeit und somit ≥ 0 ist. Setzt man $B = R$ in (b), so folgt $P(X \in R) = \sum_{i \in R} w(i)$. Da R alle möglichen Werte von X enthält, gilt $P(X \in R) = 1$. Damit ist auch (a) gezeigt. □

Beispiel 39: Es wird mit zwei Würfeln gewürfelt. Sei X die Augensumme. Man bestimme Wertebereich und Wahrscheinlichkeitsvektor von X.

Der Wertebereich von X ist $R = \{2, 3, 4, \ldots, 12\}$. Durch Abzählen der Augenpaare mit Summe i kann man $w(i) = P(X = i)$ für $i \in R$ ermitteln. Man erhält $w(2){=}w(12){=}\frac{1}{36}$, $w(3){=}w(11){=}\frac{2}{36}$, $w(4){=}w(10){=}\frac{3}{36}$, $w(5){=}w(9){=}\frac{4}{36}$, $w(6){=}w(8){=}\frac{5}{36}$, $w(7){=}\frac{6}{36}$.
Damit kann man die Wahrscheinlichkeit von Ereignissen ausrechnen. Interessiert man sich für das Ereignis "Augensumme kleiner als 5", dann erhält man mit Hilfe von Satz 15 (b) dass $P(X < 5) = P(X \in \{2, 3, 4\}) = w(2) + w(3) + w(4) = \frac{6}{36}$.

Wir wollen auch gleich den Erwartungswert oder Mittelwert einer diskreten Zufallsvariablen X definieren. Der Erwartungswert ist ein Durchschnittswert aller möglichen Werte von X, wobei jedoch nicht alle Werte gleich gewichtet werden, wie es beim arithmetischen Mittel der Fall ist, sondern jeder Wert i mit seiner Wahrscheinlichkeit $w(i)$ gewichtet wird. Jeder Wert i geht in den Erwartungswert gemäß der Häufigkeit seines Auftretens ein.

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

Will man wissen, wie stark die Zufallsvariable X um ihren Erwartungswert $\mathrm{E}(X)$ schwankt, so verwendet man dafür den Mittelwert der Quadrate der Abweichungen der einzelnen Werte i von ihrem Mittelwert $\mathrm{E}(X)$.

Definition: Die Varianz $\mathrm{V}(X)$ einer diskreten Zufallsvariablen X mit Wahrscheinlichkeitsvektor $(w(i))_{i \in R}$ ist definiert durch $\mathrm{V}(X) = \sum_{i \in R}(i - m)^2 w(i)$, wobei $m = \mathrm{E}(X)$. 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 40: Bei einer Verlosung sollen 1 % der Lose einen Gewinn von 1000 bringen, 9 % der Lose sollen einen Gewinn von 100 bringen und 90 % der Lose bringen keinen Gewinn. Welchen Gewinn muss man durchschnittlich pro Los auszahlen, das heißt 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 größer als 19 festsetzen.

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

Definition: Eine Zufallsvariable X heißt kontinuierlich, wenn eine integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}^+$ existiert, sodass $P(X \leq t) = \int_{-\infty}^{t} f(x)\,dx$ für alle $t \in \mathbb{R}$ gilt. Man nennt f die Wahrscheinlichkeitsdichte der Zufallsvariablen X. Die Funktion F definiert durch $F(t) = P(X \leq t)$ heißt Verteilungsfunktion der Zufallsvariablen X. Die Wahrscheinlichkeitsdichte kann man daher auch als Ableitung der Verteilungsfunktion auffassen.

Satz 16: *Die Verteilungsfunktion F einer Zufallsvariablen X ist eine monton wachsende Funktion von $\mathbb{R}$ nach $[0,1]$. Für $r < s$ gilt $P(r < X \leq s) = F(s) - F(r)$.*

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

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

mit Hilfe der Definition der Verteilungsfunktion.

Sind r und s in $\mathbb{R}$ beliebig mit $r < s$, dann gilt $F(s) - F(r) = P(r < X \leq s) \geq 0$ nach Satz 7 und die Monotonie von F ist bereits gezeigt. Ebenfalls aus Satz 7 folgt $0 \leq P(X \leq t) \leq 1$, sodass $F(t) \in [0,1]$ gilt. □

Kennt man die Wahrscheinlichkeitsdichte einer Zufallsvariablen oder hat sie mit Hilfe der Methoden aus den letzten Kapiteln bestimmt, so kann man damit Wahrscheinlichkeiten von Ereignissen berechnen.

Satz 17: *Sei $f : \mathbb{R} \to \mathbb{R}^+$ die Wahrscheinlichkeitsdichte einer kontinuierlichen Zufallsvariablen X und $F : \mathbb{R} \to [0,1]$ ihre Verteilungsfunktion. Dann gilt $\int_{-\infty}^{\infty} f(x)\,dx = 1$ und*
(a) $P(X \in B) = F(s) - F(r) = \int_{r}^{s} f(x)\,dx$ für ein Intervall B mit Endpunkten r und s
(b) $P(X < s) = P(X \leq s) = F(s) = \int_{-\infty}^{s} f(x)\,dx$ für $s \in \mathbb{R}$
(c) $P(X > r) = P(X \geq r) = 1 - F(r) = \int_{r}^{\infty} f(x)\,dx$ für $r \in \mathbb{R}$

Beweis: In Satz 16 wurde $P(X \in B) = F(s) - F(r)$ für $B = (r,s]$ gezeigt. Aus der Definition der Wahrscheinlichkeitsdichte folgt

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

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 9 (c). Aus der zu Beginn dieses Beweises gezeigten Formel folgt $\lim_{\varepsilon \to 0} P(X \in (u - \varepsilon, u]) = \lim_{\varepsilon \to 0} \int_{u-\varepsilon}^{u} f(x)\,dx = 0$, woraus sich $P(X = u) = 0$ ergibt. 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 also $P(X \in B) = P(X \in (r,s])$ und (a) ist gezeigt.

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

Bevor wir (c) zeigen, beweisen wir $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Es gilt

$$-\infty < X < \infty \quad \Leftrightarrow \quad X \le 0 \ \vee \ 0 < X \le 1 \ \vee \ 1 < X \le 2 \ \vee \ \ldots$$

wobei die rechts stehenden Ereignisse unvereinbar sind. Der Additionssatz ergibt

$$\begin{aligned} P(-\infty < X < \infty) &= P(X \le 0) + P(0 < X \le 1) + P(1 < X \le 2) + \ldots \\ &= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \cdots = \int_{-\infty}^{\infty} f(x)\,dx \end{aligned}$$

wobei wir (a) und (b) verwendet haben. Da das Ereignis $-\infty < X < \infty$ immer eintritt, gilt $P(-\infty < X < \infty) = 1$. Somit ist $\int_{-\infty}^{\infty} f(x)\,dx = 1$ bewiesen.

Um (c) zu beweisen, gehen wir zum Gegenereignis über und verwenden (b). Mit Hilfe von Satz 9 (d) folgt $P(X > r) = 1 - P(X \le r) = 1 - F(r)$. Verwendet man auch $\int_{-\infty}^{\infty} f(x)\,dx = 1$ so erhält man

$$P(X > r) = 1 - P(X \le r) = \int_{-\infty}^{\infty} f(x)\,dx - \int_{-\infty}^{r} f(x)\,dx = \int_{r}^{\infty} f(x)\,dx$$

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

Oft kommt es vor, dass die Wahrscheinlichkeitsdichte f einer Zufallsvariablen X außerhalb eines (meist unbeschränkten) Intervalls W gleich 0 ist. Wir nennen dann W den Wertebereich der Zufallsvariablen X. Wegen Satz 17 gilt ja $P(X \in W) = \int_W f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 1$, sodass X mit Wahrscheinlichkeit 1, also immer, in W liegt.

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

Beispiel 41: 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.

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

Bemerkung: Wie man im letzten Beispiel 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 immer 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 $\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx$. Die Varianz $\mathrm{V}(X)$

ist definiert durch $\mathrm{V}(X) = \int_{-\infty}^{\infty}(x-m)^2 f(x)\,dx$ mit $m = \mathrm{E}(X)$. Weiters nennt man $\sqrt{\mathrm{V}(X)}$ die Standardabweichung der Zufallsvariablen X.

Beispiel 42: Die Wahrscheinlichkeitsdichte f der Zufallsvariablen X sei definiert 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

$$\mathrm{E}(X) = \int_a^b x f(x)\,dx = \int_a^b x\frac{1}{b-a}\,dx = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{a+b}{2}$$
$$\mathrm{V}(X) = \int_a^b (x-\frac{a+b}{2})^2\frac{1}{b-a}\,dx = \frac{(b-a)^3-(a-b)^3}{24}\frac{1}{b-a} = \frac{a^2+ab+b^2}{12}$$

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.

13. 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 Zufallsvariable X mit Wertebereich $R = \{0,1,2,\dots,n\}$ und Wahrscheinlichkeitsvektor

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

heißt binomialverteilt oder kurz $B(n,p)$-verteilt.

Satz 18: *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: Die möglichen Werte für die Anzahl, mit der "Wappen" unter diesen n Würfen auftritt, sind $0,1,2,\dots,n$. Der Wertebereich von X ist also $R = \{0,1,2,\dots,n\}$.

Wir müssen $w(k) = P(X=k)$ für $k \in R$ berechnen. Gefragt ist nach der Wahrscheinlichkeit, dass unter den n Würfen genau k Mal "Wappen" auftritt. Für $n=5$ und $k=3$ wurde das bereits in Beispiel 32 durchgeführt. Die dort angewendete Methode lässt sich auch hier anwenden. Wir zerlegen in Teilereignisse. Die Teilereignisse sind gerade die Folgen der Länge n die k Mal W und $n-k$ Mal Z enthalten. Nach Satz 4 ist die Anzahl dieser Teilereignisse gleich $\binom{n}{k}$.

Würde man die Berechnung in Form einer Tabelle durchführen, dann hätte diese Tabelle $\binom{n}{k}$ Zeilen. Berechnet man die Wahrscheinlichkeiten der Teilereignisse nach dem Multiplikationssatz, so erhält man für jedes dieser Teilereignisse $p^k(1-p)^{n-k}$. Dieser Wert steht also in jeder Zeile der Tabelle. Nach dem Additionssatz erhält man die gesuchte Wahrscheinlichkeit, indem man die Zeilen aufsummiert. Die Summe ist $\binom{n}{k}p^k(1-p)^{n-k}$. Damit ist $w(k) = P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$ für $k \in R$ gezeigt. Die Zufallsvariable X ist $B(n,p)$-verteilt. $\square$

Hat man ein Beispiel vor sich, so muss man erkennen, ob die in Satz 18 beschriebene Situation, die dem n-maligen Werfen einer Münze entspricht, vorliegt. Dann rechnet man mit der Binomialverteilung.

Beispiel 43: Es wird 10 Mal gewürfelt. Mit welcher Wahrscheinlichkeit tritt höchstens 3 Mal die Augenzahl 6 auf?

Das zehnmalige Würfeln entspricht einem zehnmaligen Münzenwerfen, wobei die eine Seite der Münze - entspricht 6 würfeln - mit Wahrscheinlichkeit $\frac{1}{6}$ fällt, und die andere Seite der Münze - entspricht nicht 6 würfeln - mit Wahrscheinlichkeit $\frac{5}{6}$. Sei X die Anzahl, mit der die Augenzahl 6 unter den 10 Würfen auftritt. Dann hat X die $B(n,p)$-Verteilung mit $n = 10$ und $p = \frac{1}{6}$. Gesucht ist die Wahrscheinlichkeit, dass $X \leq 3$ ist. Aus Satz 15 (b) folgt $P(X \leq 3) = w(0) + w(1) + w(2) + w(3)$, wobei $w(k) = \binom{10}{k}(\frac{1}{6})^k(\frac{5}{6})^{10-k}$ ist. Das ist der Wahrscheinlichkeitsvektor der $B(10, \frac{1}{6})$-Verteilung. Wir erhalten also $P(X \leq 3) = \binom{10}{0}(\frac{1}{6})^0(\frac{5}{6})^{10} + \binom{10}{1}(\frac{1}{6})^1(\frac{5}{6})^9 + \binom{10}{2}(\frac{1}{6})^2(\frac{5}{6})^8 + \binom{10}{3}(\frac{1}{6})^3(\frac{5}{6})^7$.

Beispiel 44: 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?

Jeder der 18 Gebuchten kommt oder kommt nicht. Für jeden ist die Wahrscheinlichkeit, zu kommen, gleich $\frac{4}{5}$. Wir können das so verstehen, dass jeder der 18 Gebuchten eine Münze wirft, bei der "Wappen" mit Wahrscheinlichkeit $\frac{4}{5}$ und "Zahl" mit Wahrscheinlichkeit $\frac{1}{5}$ fällt. Diejenigen, die "Wappen" werfen, kommen.

Sei X die Anzahl derjenigen unter den Gebuchten, die kommen, das heißt die "Wappen" werfen. Nach Satz 18 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 $X \geq 17$ ist. Die Wahrscheinlichkeit dafür ist $P(X \geq 17) = w(17) + w(18) = \binom{18}{17}(\frac{4}{5})^{17}(\frac{1}{5})^1 + \binom{18}{18}(\frac{4}{5})^{18}(\frac{1}{5})^0$.

Beispiel 45: 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 kann man als Wurf einer Münze auffassen, die mit Wahrscheinlichkeit 0.1 auf "unbrauchbar" und mit Wahrscheinlichkeit 0.9 auf "brauchbar" fällt. Sei X die Anzahl der unbrauchbaren Schrauben in der Packung von 100 Stück. Die 100 Stück Schrauben kann man sich als 100 Münzenwürfe vorstellen, wobei X die Anzahl der Würfe ist, die auf "unbrauchbar" fallen. Daher hat X die $B(n,p)$-Verteilung mit $n = 100$ und $p = 0.1$. Gesucht ist
$P(X \geq 8) = 1 - P(X \leq 7) = 1 - \sum_{k=0}^{7} w(k) = 1 - \sum_{k=0}^{7} \binom{100}{k} 0.1^k 0.9^{100-k}$.
Würde man $P(X \geq 8)$ direkt ausrechnen, dann müßte man eine viel größere Summe auswerten.

In den beiden letzten Beispielen kann man auch nach dem Erwartungswert fragen. Der Pensionsinhaber in Beispiel 44 ist daran interessiert, wie viele der 18 Personen, deren Buchung er jedes Jahr für die Ferienwoche entgegennimmt, im Durchschnitt kommen? Der Käufer einer Packung Schrauben aus Beispiel 45 möchte wissen, wie viele unbrauchbare Schrauben durchschnittlich in einer Packung sind?

Um das zu beantworten, wollen wir den Erwartungswert einer $B(n,p)$-verteilten Zufallsvariablen X berechnen. Aus der Definition $\mathrm{E}(X) = \sum_{k\in R} kw(k)$ erhalten wir

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

Für eine $B(n,p)$-verteilte Zufallsvariable X gilt also $\mathrm{E}(X) = np$. Der Pensionsinhaber aus Beispiel 44 kann also erwarten, dass durchschnittlich $18 \cdot 0.8 = 14.4$ der 18 Personen, die gebucht haben, kommen. In einer Packung Schrauben aus Beispiel 45 wird man durchschnittlich $100 \cdot 0.1 = 10$ unbrauchbaree 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.

$$\begin{aligned}
\mathrm{V}(X) &= \sum_{k=0}^{n} (k-np)^2 \binom{n}{k} p^k (1-p)^{n-k}\\
&= \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + (1-2np) \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} + n^2p^2 \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

Aus dem binomischen Lehrsatz folgt, dass $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$ gilt. Weiters wurde die Summe $\sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = np$ oben bei der Berechnung des Erwartungswertes bestimmt. Genauso erhält man auch $\sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} = n(n-1)p^2$. Es gilt also $\mathrm{V}(X) = n(n-1)p^2 + (1-2np)np + n^2p^2 = np(1-p)$ für eine $B(n,p)$-verteilte Zufallsvariable X.

Wir führen noch eine weitere Verteilung ein, die mit Münzenwerfen zu tun hat. Jetzt werfen wir die Münze solange, bis zum ersten Mal "Zahl" fällt.

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

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

heißt geometrisch verteilt.

Satz 19: *Eine Münze, bei der "Wappen" mit Wahrscheinlichkeit p und "Zahl" mit Wahrscheinlichkeit $1-p$ auftritt, wird solange geworfen, bis "Zahl" fällt. Sei X die Anzahl der dafür notwendigen Würfe. Dann ist X geometrisch verteilt.*

Beweis: Das erste Mal "Zahl" kann beim ersten Wurf, beim zweiten Wurf oder bei einem späteren Wurf eintreten. Daher sind $1,2,3,\dots$ die möglichen Werte für X. Der Wertebereich ist also $R = \{1,2,3,\dots\}$. Das Ereignis $X = k$ tritt genau dann ein, wenn die ersten

$k-1$ Würfe "Wappen" und der k-te Wurf "Zahl" ergibt. Nach dem Multiplikationssatz ist die Wahrscheinlichkeit dafür gleich $p^{k-1}(1-p)$. Wir erhalten also $w(k) = p^{k-1}(1-p)$ für $k \in R$ und X ist geometrisch verteilt. □

Beispiel 46: Man würfelt so lange, bis 6 kommt. Wie groß ist die Wahrscheinlichkeit, dass man höchstens 10 Mal würfeln muss?

Sei X die Anzahl der Würfe, die man benötigt, bis zum ersten Mal 6 auftritt. Gesucht ist $P(X \leq 10)$. Dieses Würfelexperiment entspricht dem Werfen einer Münze, bei der "Wappen" mit Wahrscheinlichkeit $\frac{5}{6}$ und "Zahl" mit Wahrscheinlichkeit $\frac{1}{6}$ auftritt. Daher hat X die geometrische Verteilung mit $p = \frac{5}{6}$. Für die gesuchte Wahrscheinlichkeit erhalten wir $P(X \leq 10) = \sum_{k=1}^{10} (\frac{5}{6})^{k-1}\frac{1}{6} = 1 - (\frac{5}{6})^{10}$.

Den Erwartungswert einer geometrisch verteilten Zufallsvariablen X berechnen wir mit folgendem Trick. Wir gehen aus von $\sum_{k=1}^{\infty} x^k = \frac{1}{1-x} - 1$, das ist eine geometrische Reihe. Durch Differenzieren dieser Reihe erhalten wir $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$. Daraus ergibt sich $\mathrm{E}(X) = \sum_{k=1}^{\infty} kw(k) = \sum_{k=1}^{\infty} kp^{k-1}(1-p) = \frac{1-p}{(1-p)^2} = \frac{1}{1-p}$.

Beispiel 47: Wie oft muss man durchschnittlich würfeln, bis die Augenzahl 6 kommt?

Die Anzahl X der Würfe, die man benötigt, bis 6 kommt, ist geometrisch verteilt mit $p = \frac{5}{6}$. Die durchschnittliche Anzahl ist daher $\mathrm{E}(X) = \frac{1}{1-p} = 6$. Im Durchschnitt benötigt man sechs Würfe, bis die Augenzahl 6 zum ersten Mal auftritt.

14. Poissonverteilung

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

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

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

heißt Poissonverteilt oder kurz $P(\lambda)$-verteilt.

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. Da X eine Anzahl ist, sind $0, 1, 2, \dots$ ihre möglichen Werte.

Um den Wahrscheinlichkeitsvektor von X zu berechnen, müssen wir einige Annahmen zugrundelegen. Sei $z > t$ und n die Anzahl der Anrufer im Zeitintervall $[0, z]$. Wir nehmen an, dass jeder Anrufer unabhängig von den anderen zufällig einen Zeitpunkt in $[0, z]$ auswählt, zu dem er anruft. Die Wahrscheinlichkeit, dass ein Anrufer im Zeitintervall $[0, t]$ anruft, ist dann $\frac{t}{z}$ und die Wahrscheinlichkeit, dass er im Zeitintervall $(t, z]$ anruft, ist $\frac{z-t}{z} = 1 - \frac{t}{z}$. Wir können wieder ans Münzenwerfen denken. Jeder Anrufer wirft eine Münze, die mit Wahrscheinlichkeit $\frac{t}{z}$ für $[0, t]$ und mit Wahrscheinlichkeit $1 - \frac{t}{z}$ für $(t, z]$ entscheidet. Da X die Anzahl der Anrufer im Zeitintervall $[0, t]$, also die Anzahl der Münzenwürfe, die für $[0, t]$ entscheiden, ist, hat X die $B(n, \frac{t}{z})$-Verteilung. Wir halten $\mu = \frac{n}{z}$, die Anzahl der Anrufer pro Zeiteinheit, fest und lassen n und damit auch z gegen ∞ gehen.

Satz 20: *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 eine $P(\mu t)$-Verteilung.*

Beweis: Wir berechnen $w(k) = P(X = k)$.

$$P(X=k) = \binom{n}{k}(\tfrac{t}{z})^k(1-\tfrac{t}{z})^{n-k} = \binom{n}{k}(\tfrac{\mu t}{n})^k(1-\tfrac{\mu t}{n})^{n-k}$$
$$= \frac{(\mu t)^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}(1-\tfrac{\mu t}{n})^n(1-\tfrac{\mu t}{n})^{-k}$$

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

Beispiel 48: 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

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

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

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 Kunden, die ein Geschäft betreten, für die Defekte eines Gerätes, oder für die Schadensmeldungen, die bei einer Versicherung eintreffen.

15. Exponentialverteilung und Gammaverteilung

Jetzt kommen wir zu den kontinuierlichen Zufallsvariablen. In diesem Kapitel geht es um Wartezeiten.

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

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

heißt exponentialverteilt oder kurz $E(\lambda)$-verteilt.

Satz 21: *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 $\le t$, wenn im Zeitintervall $[0,t]$ mindestens ein Anruf kommt. Das heißt $Y \le t$ tritt genau dann ein, wenn $X \ge 1$ eintritt. Daher gilt $F(t) = P(Y \le t) = P(X \ge 1) = 1 - P(X = 0) = 1 - e^{-\lambda t}$. Die Dichte f von Y erhält man als Ableitung von F, nämlich $f(x) = (1 - e^{-\lambda x})' = \lambda e^{-\lambda x}$.

Die Wartezeit Y kann nicht negativ sein. Für $t \le 0$ gilt daher $F(t) = P(Y \le t) = 0$ und $f(x) = 0$ für $x \le 0$. Der Wertebereich der Zufallsvariablen Y ist gleich $\mathbb{R}^+$. Die Wartezeit Y auf den ersten Anruf ist also $E(\lambda)$-verteilt. $\square$

Eine beliebte Anwendung der Exponentialverteilung ist die Lebensdauer eines Gerätes. Sie ist ja die Wartezeit auf den ersten Defekt.

Beispiel 49: 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 $P(Y \geq 500) = 0.9$. Es gilt $P(Y \geq 500) = \int_{500}^{\infty} \lambda e^{-\lambda x}\, dx = e^{-500\lambda}$, da Y $E(\lambda)$-verteilt ist. Aus $e^{-500\lambda} = 0.9$ folgt $\lambda = \frac{-\ln 0.9}{500} = 0.0002$.

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
$\mathrm{E}(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx = -xe^{-\lambda x}\big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = -\frac{e^{-\lambda x}}{\lambda}\big|_0^\infty = \frac{1}{\lambda}$.
Ebenso erhält man durch zweimalige partielle Integration
$\mathrm{V}(X) = \int_0^\infty (x-\frac{1}{\lambda})^2 \lambda e^{-\lambda x}\, dx = -(x-\frac{1}{\lambda})^2 e^{-\lambda x}\big|_0^\infty - 2(x-\frac{1}{\lambda})\frac{e^{-\lambda x}}{\lambda}\big|_0^\infty + \int_0^\infty 2\frac{e^{-\lambda x}}{\lambda}\, dx = \frac{1}{\lambda^2}$.
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 49 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 $\lambda > 0$ und $r > 0$. Eine 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 oder kurz $G(r, \lambda)$-verteilt, wobei $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y}\, dy$ ist.

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

Satz 22: *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 $G(n, \lambda)$-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

$$F(t) = P(Z \leq t) = P(X \geq n) = 1 - P(X \leq n-1) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} - \dots - \tfrac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}$$

Die Dichte f von Z erhält man als Ableitung von F. Berechnet man diese, so kürzen sich alle bis auf einen Summanden weg. Man erhält $f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$ für $x \geq 0$. Da Z als Wartezeit nur positive Werte annehmen kann, ist $\mathbb{R}^+$ der Wertebereich von Z und $f(x) = 0$ für $x < 0$. $\square$

16. Normalverteilung

Misst man eine (physikalische) Größe mehrmals, so wird man Messwerte erhalten, die wegen zufälliger Störungen beim Messvorgang leicht voneinander abweichen. Werden Milchflaschen mit einem Liter Inhalt von einer Abfüllanlage gefüllt, so wird man ebenfalls 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 $\sigma > 0$. Eine 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 oder kurz $N(\mu, \sigma)$-verteilt.

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

Satz 23: *Sei X eine Zufallsvariable mit Wahrscheinlichkeitsdichte f. Seien $a \in \mathbb{R}$, $b > 0$ und $Y = \frac{X-a}{b}$. Dann hat Y die Wahrscheinlichkeitsdichte $g(x) = bf(bx + a)$.*

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

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

Satz 24: *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 23 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 $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Man kann zeigen, dass $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi}$ ist, sodass $\int_{-\infty}^{\infty} \varphi(x)\, dx = 1$ gilt, woraus wieder $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = 1$ folgt. Die Verteilungsfunktion der $N(0, 1)$-Verteilung wird mit Φ bezeichnet, das heißt $\Phi(t) = \int_{-\infty}^{t} \varphi(x)\, dx$ für $t \in \mathbb{R}$.

Satz 25: *Es gilt*
(a) $\varphi(-x) = \varphi(x)$ *für alle* $x \in \mathbb{R}$
(b) $\Phi(-t) = 1 - \Phi(t)$ *für alle* $t \in \mathbb{R}$
(c) $\Phi : \mathbb{R} \to (0, 1)$ *ist streng monoton wachsend und invertierbar.*

Beweis: Man erkennt (a) sofort aus der Definition von φ. Um (b) zu sehen, rechnet man $\Phi(-t) = \int_{-\infty}^{-t} \varphi(x)\, dx = \int_{-\infty}^{-t} \varphi(-x)\, dx = -\int_{\infty}^{t} \varphi(y)\, dy = \int_{t}^{\infty} \varphi(y)\, dy = 1 - \Phi(t)$.

Da $\Phi'(t) = \varphi(t) > 0$ für alle t gilt, ist Φ eine streng monoton wachsende Funktion und daher auch injektiv. Weiters erhalten wir $\lim_{t\to-\infty}\Phi(t) = \lim_{t\to-\infty}\int_{-\infty}^{t}\varphi(x)\,dx = 0$ und $\lim_{t\to\infty}\Phi(t) = \lim_{t\to\infty}\int_{-\infty}^{t}\varphi(x)\,dx = 1$. Daher ist $\Phi : \mathbb{R} \to (0,1)$ auch surjektiv. Damit ist gezeigt, dass Φ invertierbar ist, und (c) ist bewiesen. $\square$

Tritt in einem Beispiel eine normalverteilte Zufallsvariable auf, so verwendet man zuerst Satz 24, um sie auf eine $N(0,1)$-verteilte Zufallsvariable Y zurückzuführen. Wahrscheinlichkeiten für diese Zufallsvariable kann man mit Hilfe ihrer Verteilungsfunktion Φ ausdrücken. Wegen Satz 17 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 50: 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.

$$
\begin{aligned}
P(18.95 < X < 19.10) &= P(\tfrac{18.95-\mu}{\sigma} < Y < \tfrac{19.10-\mu}{\sigma}) && \text{wobei } Y = \tfrac{X-\mu}{\sigma} \\
&= P(-1 < Y < 2) \\
&= \Phi(2) - \Phi(-1) && \text{da } Y \text{ die } N(0,1)\text{-Verteilung hat} \\
&= \Phi(2) - 1 + \Phi(1) && \text{Satz 25 (b)} \\
&= 0.977 - 1 + 0.841 = 0.818 && \text{Tabelle}
\end{aligned}
$$

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

Beispiel 51: 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.
(i) 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?
(ii) 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?
(iii) Auf welchen Mittelwert μ muss man die Maschine einstellen, wenn die Mindestfüllmenge von 2.9 kg nur von 1 % der Packungen unterschritten werden darf?

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

$$
\begin{aligned}
P(X \le 2.9) &= P(Y \le \tfrac{2.9-3}{0.05}) && \text{wobei } Y = \tfrac{X-\mu}{\sigma} = \tfrac{X-3}{0.05} \\
&= P(Y \le -2) \\
&= \Phi(-2) && \text{da } Y \text{ die } N(0,1)\text{-Verteilung hat} \\
&= 1 - \Phi(2) && \text{Satz 25 (b)} \\
&= 1 - 0.977 = 0.023 && \text{Tabelle}
\end{aligned}
$$

Die Mindestfüllmenge wird von 2.3 % der Packungen unterschritten.

(ii) Gesucht ist t, sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt X kleiner als t ist, höchstens 0.01 beträgt. Das heißt also, dass $P(X < t) \leq 0.01$ gelten muss. Wir erhalten

$$P(X < t) = P(\tfrac{X-\mu}{\sigma} < \tfrac{t-3}{0.05}) = \Phi(\tfrac{t-3}{0.05})$$

Wir suchen t, sodass $\Phi(\frac{t-3}{0.05}) \leq 0.01$ gilt. Da Φ streng monoton wachsend ist, ist das äquivalent zu $\frac{t-3}{0.05} \leq \Phi^{-1}(0.01)$. In der Tabelle findet man $\Phi^{-1}(0.99) = 2.33$. Wegen Satz 25 (b) folgt $\Phi^{-1}(0.01) = -2.33$. Wir suchen also t, sodass $\frac{t-3}{0.05} \leq -2.33$ gilt. Es folgt $t \leq 2.884$. Druckt man eine Mindestfüllmenge, die ≤ 2.884 ist, auf das Etikett, dann ist die verlangte Forderung erfüllt.

(iii) Gesucht ist μ, sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt X kleiner als 2.9 ist, höchstens 0.01 beträgt. Das heißt also, dass $P(X < 2.9) \leq 0.01$ gelten muss. Wir erhalten

$$P(X < 2.9) = P(\tfrac{X-\mu}{\sigma} < \tfrac{2.9-\mu}{0.05}) = \Phi(\tfrac{2.9-\mu}{0.05})$$

Wir suchen μ, sodass $\Phi(\frac{2.9-\mu}{0.05}) \leq 0.01$ gilt. Da Φ streng monoton wachsend ist, ist das äquivalent zu $\frac{2.9-\mu}{0.05} \leq \Phi^{-1}(0.01)$. Wie oben findet man $\Phi^{-1}(0.01) = -2.33$. Wir suchen also μ, sodass $\frac{2.9-\mu}{0.05} \leq -2.33$ gilt. Es folgt $\mu \geq 3.016$. Stellt man die Maschine auf einen Mittelwert μ ein, der ≥ 3.016 ist, dann ist die verlangte Forderung erfüllt.

Zum Abschluss dieses Kapitels wollen wir den Erwartungswert und die Varianz einer $N(\mu, \sigma)$-verteilten Zufallsvariablen X berechnen. Durch Substitution der neuen Variablen $y = \frac{x-\mu}{\sigma}$ erhalten wir

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

Es gilt $\int_{-\infty}^{\infty} y\varphi(y)\,dy = 0$, da $\varphi(y) = \varphi(-y)$ ist und daher $\int_0^{\infty} y\varphi(y)\,dy = -\int_{-\infty}^0 y\varphi(y)\,dy$ gilt. Weiters gilt $\int_{-\infty}^{\infty} \varphi(y)\,dy = 1$. Somit folgt $\mathrm{E}(X) = \mu$.

Analog berechnen wir die Varianz von X. Durch Substitution der neuen Variablen $y = \frac{x-\mu}{\sigma}$ erhalten wir

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

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

17. Approximation der Binomialverteilung

Es ist etwas 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

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

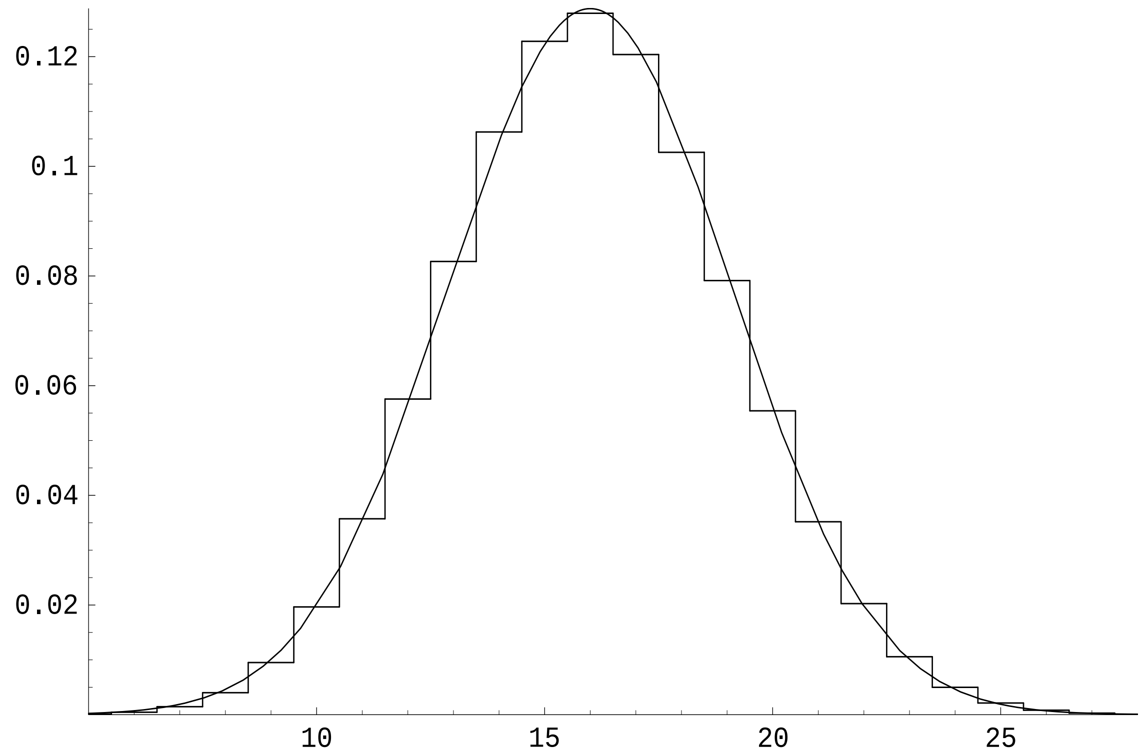
und $g(x) = 0$ für $x \notin (-\frac{1}{2}, n + \frac{1}{2}]$. Durch g ist eine Treppenfunktion definiert, sodass die Fläche des Rechtecks, das über dem Intervall $(j - \frac{1}{2}, j + \frac{1}{2}]$ und unter g liegt, gleich $w(j)$

ist. Deshalb ist $\sum_{j=0}^{k} w(j)$ gerade die links von $k + \frac{1}{2}$ liegende Fläche unter der Treppenfunktion g. Für $k \in \{0, 1, \ldots, n\}$ gilt daher $P(X \leq k) = \sum_{j=0}^{k} w(j) = \int_{-\infty}^{k+\frac{1}{2}} g(x)\, dx$.

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.

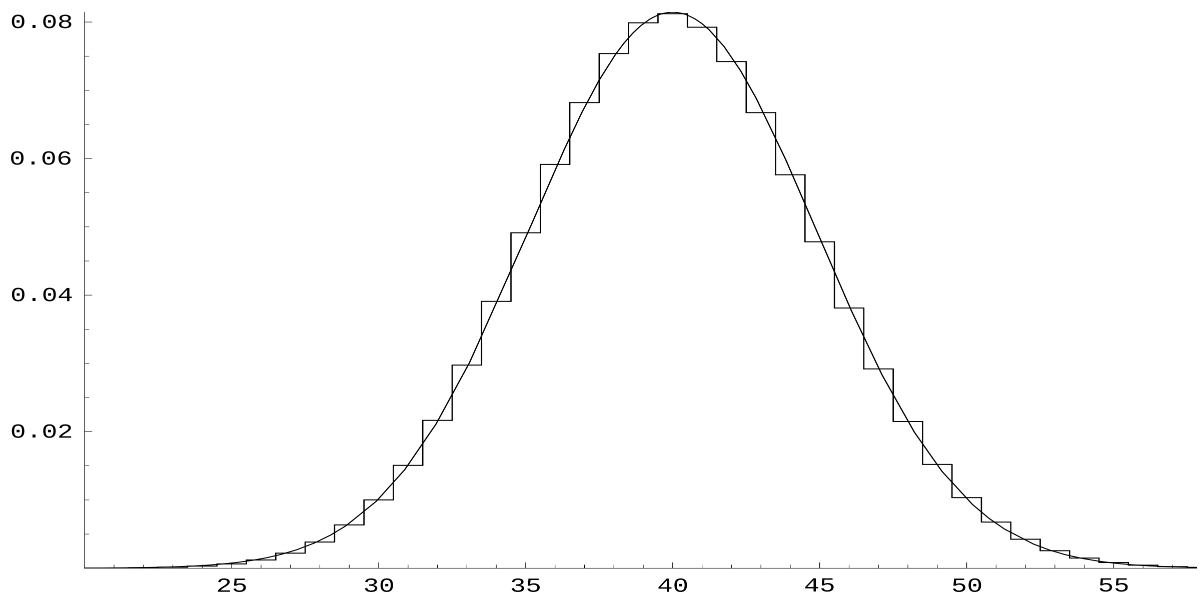


Man kann daher eine $B(n,p)$-verteilte Zufallsvariable X näherungsweise so behandeln, als ob sie $N(\mu, \sigma)$-verteilt wäre mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$. Für $t \in \mathbb{R}$ gilt

$$P(X \leq t) \approx \int_{-\infty}^{t} g(x)\, dx \approx \int_{-\infty}^{t} f(x)\, dx = \int_{-\infty}^{\frac{t-\mu}{\sigma}} \varphi(y)\, dy = \Phi(\tfrac{t-\mu}{\sigma})$$

wobei man die vorletzte Gleichheit dadurch erhält, dass man die neue Variable $y = \frac{x-\mu}{\sigma}$ einführt.

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 $\sigma = \sqrt{np(1-p)} \geq 3$.

In dieser Approximation sind zwei Fehler enthalten. Der eine entsteht durch die Approximation von g durch f. Der andere entsteht, weil auch $P(X \leq t) \approx \int_{-\infty}^{t} g(x)\,dx$ nur näherungsweise stimmt. Zu Beginn dieses Kapitels haben wir uns für ganzzahliges k überlegt, dass $P(X \leq k) = \int_{-\infty}^{k+\frac{1}{2}} g(x)\,dx$ gilt. Mit Hilfe dieser Gleichung erhalten wir für $k \in \{0, 1, \dots, n\}$ wie oben

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

Das ergibt eine bessere Approximation.

In den Beispielen hat man es meistens 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 die verbesserte Approximation verwenden kann.

Beispiel 52: Auf einer Maschine werden Schrauben erzeugt, von denen 1 % schadhaft sind. Die Schrauben werden in Packungen zu 1600 Stück verkauft.
(i) Wie groß ist die Wahrscheinlichkeit, dass höchstens 9 schadhafte Schrauben in einer Packung sind?
(ii) Welche Anzahl von brauchbaren Schrauben in einer Packung kann der Hersteller garantieren, wenn seine Garantie von höchstens 2 % der Packungen unterschritten werden darf?

(i) 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 $X \le 9$ ist.

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 $\sigma \ge 3$ ist die Approximation ausreichend gut. Wir erhalten

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

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

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

(ii) 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 $P(Y < k) \le 0.02$ gilt. Jetzt hat Y 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 $\sigma = \sqrt{np(1-p)} = \sqrt{1600 \cdot 0.99 \cdot 0.01} = 3.98$. Weil $\sigma \ge 3$ gilt, ist die Approximation ausreichend gut.

Wegen $P(Y < k) = P(Y \le k-1) \approx \Phi(\frac{k-1+\frac{1}{2}-1584}{3.98}) = \Phi(\frac{k-1584.5}{3.98})$ ist k so zu bestimmen, dass $\Phi(\frac{k-1584.5}{3.98}) \le 0.02$ gilt. Aus $\Phi(2.05) = 0.98$ folgt $\Phi(-2.05) = 0.02$ und $-2.05 = \Phi^{-1}(0.02)$. Da Φ streng monoton wachsend ist, erhalten wir die äquivalente Ungleichung $\frac{k-1584.5}{3.98} \le -2.05$. Das ergibt $k \le 1576.3$.

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

Beispiel 53: 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 $P(X \le 120) \ge 0.95$ 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 $P(X \le 120) = \Phi(\frac{120.5-0.9\cdot n}{0.3\sqrt{n}})$ und $0.95 = \Phi(1.65)$ ist n so zu bestimmen, dass $\Phi(\frac{120.5-0.9\cdot n}{0.3\sqrt{n}}) \ge \Phi(1.65)$. Wegen der Monotonie von Φ ist das wieder äquivalent zu $\frac{120.5-0.9\cdot n}{0.3\sqrt{n}} \ge 1.65$. Durch Umformen ergibt sich

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

Setzt man $x = \sqrt{n}$ so erhält man die quadratische Ungleichung $x^2 + 0.55x - 133.9 \le 0$. 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.

18. 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 23, wo die Wahrscheinlichkeitsdichte einer linearen Funktion $\frac{X-a}{b}$ der Zufallsvariablen X berechnet wurde. Wir wollen dasselbe auch für andere Funktionen von Zufallsvariablen tun. Solche Resultate bilden auch die Grundlage für Beweise in der Statistik. Dieses Kapitel ist nicht für das Verständnis der folgenden Kapitel notwendig und kann daher weggelassen werden.

Wir beginnen mit dem Quadrat einer Zufallsvariablen.

Satz 26: *Sei X eine Zufallsvariable mit Wahrscheinlichkeitsdichte f. Dann hat X^2 den Wertebereich $\mathbb{R}^+$ und die 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 Zufallsvariablen X, also eine Stammfunktion von f. Sei weiters G die Verteilungsfunktion von X^2, sodass $g=G'$ gilt.

Da X^2 immer positiv ist, 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. □

Beispiel 54: 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 26 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 $G(\frac{1}{2},\frac{1}{2})$-Verteilung, die somit die Verteilung von X^2 ist.

Jetzt wollen wir zwei Zufallsvariablen X und Y gleichzeitig untersuchen und die Wahrscheinlichkeitsdichte von $X+Y$ bestimmen. Zuvor wollen wir die gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen X und Y definieren.

Definition: Eine integrierbare Funktion $h : \mathbb{R}^2 \to \mathbb{R}^+$ heißt gemeinsame Wahrscheinlichkeitsdichte von X und Y, wenn $P(X \le s \ \wedge \ Y \le t) = \int_{-\infty}^{s} \int_{-\infty}^{t} h(x,y)\,dy\,dx$ für alle $s \in \mathbb{R}$ und alle $t \in \mathbb{R}$ gilt.

Satz 27: *Sei $h : \mathbb{R}^2 \to \mathbb{R}^+$ die gemeinsame Wahrscheinlichkeitsdichte von X und Y. Für integrierbare Teilmengen B von $\mathbb{R}^2$ gilt $P((X,Y) \in B) = \int_B h(x,y)\,d(x,y)$.*

Beweis: Der Beweis erfordert die Approximation der Menge B durch Vereinigungen von Rechtecken und soll hier nicht durchgeführt werden. □

Analog kann man einen gemeinsamen Wahrscheinlichkeitsvektor zweier Zufallsvariablen X und Y definieren. Wir werden im Folgenden nur Zufallsvariable mit Wahrscheinlichkeitsdichten untersuchen, wollen aber die Definition des gemeinsamen Wahrscheinlichkeitsvektors angeben.

Definition: Ist X diskret mit Wertebereich R und Y diskret mit Wertebereich S, dann wird durch $u(i,j) = P(X = i \ \wedge \ Y = j)$ für $i \in R$ und $j \in S$ der gemeinsame Wahrscheinlichkeitsvektor der Zufallsvariablen X und Y definiert.

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

Definition: Die Zufallsvariablen X und Y heißen unabhängig, wenn für jedes $s \in \mathbb{R}$ und jedes $t \in \mathbb{R}$ die Ereignisse $X \le s$ und $Y \le t$ unabhängig sind, das heißt wenn $P(X \le s \ \wedge \ Y \le t) = P(X \le s)P(Y \le t)$ gilt.

Satz 28: *Seien $f_1 : \mathbb{R} \to \mathbb{R}^+$ und $f_2 : \mathbb{R} \to \mathbb{R}^+$ die Wahrscheinlichkeitsdichten der Zufallsvariablen X und Y und sei $h : \mathbb{R}^2 \to \mathbb{R}^+$ die gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen X und Y. Dann sind X und Y unabhängig genau dann, wenn $h(x,y) = f_1(x)f_2(y)$ für alle $x \in \mathbb{R}$ und alle $y \in \mathbb{R}$ gilt.*

Beweis: Sind X und Y unabhängig, dann gilt $P(X \le s \ \wedge \ Y \le t) = P(X \le s)P(Y \le t)$ für alle $s, t \in \mathbb{R}$. Mit Hilfe der Definition der Wahrscheinlichkeitsdichte folgt daraus $\int_{-\infty}^{s} \int_{-\infty}^{t} h(x,y)\,dy\,dx = \int_{-\infty}^{s} f_1(x)\,dx \int_{-\infty}^{t} f_2(y)\,dy$ für alle $s, t \in \mathbb{R}$. Durch Differenzieren nach s und nach t erhält man $h(s,t) = f_1(s)f_2(t)$ für alle $s \in \mathbb{R}$ und alle $t \in \mathbb{R}$.
Die Umkehrung funktioniert genauso. Gilt $h(x,y) = f_1(x)f_2(y)$ für alle $x, y \in \mathbb{R}$, so folgt durch Integration, dass $\int_{-\infty}^{s} \int_{-\infty}^{t} h(x,y)\,dy\,dx = \int_{-\infty}^{s} f_1(x)\,dx \int_{-\infty}^{t} f_2(y)\,dy$ für alle $s, t \in \mathbb{R}$ gilt. Das aber hat wieder die Unabhängigkeit von X und Y zur Folge. □

Jetzt können wir die Wahrscheinlichkeitsdichte der Summe zweier unabhängiger Zufallsvariablen berechnen.

Satz 29: *Seien X und Y unabhängige Zufallsvariablen mit Wahrscheinlichkeitsdichten f_1 und f_2. Die Wahrscheinlichkeitsdichte von $X + Y$ ist dann*

$$g(x) = \int_{-\infty}^{\infty} f_1(x-y)f_2(y)\,dy$$

Sind die Wertebereiche von X und Y in $\mathbb{R}^+$ enthalten, also f_1 und f_2 gleich 0 auf $\mathbb{R}^-$, dann gilt $g(x) = \int_0^x f_1(x-y)f_2(y)\,dy$.

Beweis: Da X und Y als unabhängig vorausgesetzt sind, ist $h(x,y) = f_1(x)f_2(y)$ nach Satz 28. Um die Wahrscheinlichkeitsdichte g von $X+Y$ zu bestimmen, müssen wir zuerst $P(X+Y \le t)$ berechnen. Dazu sei $B = \{(x,y) \in \mathbb{R}^2 : x+y \le t\}$. Aus Satz 27 folgt

$$P(X+Y \le t) = P((X,Y) \in B) = \int_B h(x,y)\,d(x,y) = \int_B f_1(x)f_2(y)\,d(x,y)$$

Nun ist $B = \{(x,y) \in \mathbb{R}^2 : -\infty < y < \infty, -\infty < x \le t-y\}$. Setzt man diese Grenzen ein, so ergibt sich

$$P(X+Y \le t) = \int_{-\infty}^{\infty}\int_{-\infty}^{t-y} f_1(x)f_2(y)\,dx\,dy$$

Führt man im inneren Integral die neue Variable $z = x+y$ ein, dann erhält man

$$P(X+Y \le t) = \int_{-\infty}^{\infty}\int_{-\infty}^{t} f_1(z-y)f_2(y)\,dz\,dy = \int_{-\infty}^{t}\int_{-\infty}^{\infty} f_1(z-y)f_2(y)\,dy\,dz$$

Da die Wahrscheinlichkeitsdichte g von $X+Y$ durch $P(X+Y \le t) = \int_{-\infty}^{t} g(z)\,dz$ für alle t definiert ist, haben wir $g(z) = \int_{-\infty}^{\infty} f_1(z-y)f_2(y)\,dy$ gezeigt.

Wenn f_1 und f_2 gleich 0 auf $\mathbb{R}^-$ sind, dann ist $f_2(y) = 0$ für $y < 0$ und $f_1(x-y) = 0$ für $y > x$. Daher genügt es, von 0 bis x zu integrieren. $\square$

Beispiel 55: 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. Unter diesen Voraussetzungen hat $X+Y$ die $N(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2})$-Verteilung.

Das berechnen wir mit Hilfe von Satz 29. Die Wahrscheinlichkeitsdichte von X ist $f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{-(x-\mu_1)^2/2\sigma_1^2}$ und die von Y ist $f_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2}e^{-(x-\mu_2)^2/2\sigma_2^2}$. Setzt man das in die Formel aus Satz 29 ein, so erhält man

$$g(x) = \int_{-\infty}^{\infty} f_1(x-y)f_2(y)\,dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-y-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}}\,dy$$

Durch entsprechendes Umformen des Exponenten ergibt sich

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

wobei $C = \frac{\mu_1\sigma_2^2 - \mu_2\sigma_1^2 - x\sigma_2^2}{\sigma_1^2+\sigma_2^2}$. 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

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

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

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

wobei die letzte Gleichheit wegen $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz = 1$ gilt. Setzt man das oben ein, so folgt schließlich

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

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

Das Resultat aus Beispiel 55 kann man auf eine Summe von n unabhängigen normalverteilten Zufallsvariablen verallgemeinern. Dadurch ergibt sich der folgende Satz, der für die Statistik normalverteilter Messwerte wichtig ist.

Satz 30: *Sind die Zufallsvariablen $X_1, X_2, \ldots, X_n$ unabhängig und alle $N(\mu, \sigma)$-verteilt, dann hat $X_1 + X_2 + \cdots + X_n$ die $N(n\mu, \sqrt{n}\sigma)$-Verteilung.*

Beweis: Für $n = 2$ folgt das sofort aus Beispiel 55. Die Verteilung von $X_1 + X_2$ ist die $N(2\mu, \sqrt{2}\sigma)$-Verteilung.
Um das Resultat für $n > 2$ zu zeigen, machen wir weiter mit Induktion. Wir nehmen an, es ist schon gezeigt, dass $X_1 + X_2 + \cdots + X_k$ die $N(k\mu, \sqrt{k}\sigma)$-Verteilung hat. Die vorausgesetzte Unabhängigkeit (für n Zufallsvariablen haben wir sie gar nicht definiert) verstehen wir so, dass $X_1 + X_2 + \cdots + X_k$ und X_{k+1} unabhängig sind. (Sind die X_j die Messergebnisse von unabhängig durchgeführten Messungen, dann ist X_{k+1} sicherlich auch von der Summe $X_1 + \cdots + X_k$ der vorher erhaltenen Messergebnisse unabhängig.) Jetzt folgt aus Beispiel 55, dass $X_1 + X_2 + \cdots + X_{k+1}$ die $N(k\mu + \mu, \sqrt{k\sigma^2 + \sigma^2})$-Verteilung, also die $N((k+1)\mu, \sqrt{k+1}\,\sigma)$-Verteilung hat. $\square$

19. Erwartungswert und Varianz

In diesem Kapitel werden einige Rechenregeln für Erwartungswert und Varianz hergeleitet. Diese Resultate werden sonst nirgends verwendet, sodass dieses Kapitel weggelassen werden kann.

Die Grundlage für das Rechnen mit Erwartungswerten bildet der folgende Satz.

Satz 31: *Erwartungswert einer Funktion von Zufallsvariablen*
(a) *Sei $\gamma : \mathbb{R} \to \mathbb{R}$ eine Abbildung. Es sei weiters X eine Zufallsvariable und $Z = \gamma(X)$ (zum Beispiel: wenn $\gamma(x) = x^2$, dann $Z = X^2$). Dann gilt*

$$\mathrm{E}(Z) = \sum_{i \in R} \gamma(i) w(i)$$

falls $w(i)$ mit $i \in R$ der Wahrscheinlichkeitsvektor von X ist, und

$$\mathrm{E}(Z) = \int_{-\infty}^{\infty} \gamma(x) f(x)\, dx$$

falls f die Wahrscheinlichkeitsdichte von X und γ integrierbar ist.
(b) *Sei $\gamma : \mathbb{R}^2 \to \mathbb{R}$ eine Abbildung. Seien X und Y Zufallsvariable und $Z = \gamma(X, Y)$ (zum Beispiel: wenn $\gamma(x, y) = x + y$, dann $Z = X + Y$). Dann gilt*

$$\mathrm{E}(Z) = \sum_{i \in R} \sum_{j \in S} \gamma(i, j) u(i, j)$$

falls $u(i, j)$ mit $i \in R$ und $j \in S$ der gemeinsame Wahrscheinlichkeitsvektor von X und Y

ist, und

$$\mathrm{E}(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \gamma(x,y)h(x,y)\,dx\,dy$$

falls h die gemeinsame Wahrscheinlichkeitsdichte von X und Y und γ integrierbar ist.

Beweis: Wir beweisen (a) für eine Zufallsvariable X, die einen Wahrscheinlichkeitsvektor $w(i)$ mit $i \in R$ hat. Wir berechnen zuerst den Wahrscheinlichkeitsvektor von Z. Der Wertebereich von Z ist $T = \gamma(R)$. Für $k \in T$ gilt

$$P(Z = k) = P(\gamma(X) = k) = P(X \in \gamma^{-1}(k)) = \sum_{i\in\gamma^{-1}(k)\cap R} w(i)$$

Also ist $v(k) = \sum_{i\in\gamma^{-1}(k)\cap R} w(i)$ mit $k \in T$ der Wahrscheinlichkeitsvektor von Z. Aus der Definition des Erwartungswertes folgt jetzt

$$\begin{aligned}
\mathrm{E}(Z) &= \sum_{k\in T} kv(k) = \sum_{k\in T}\sum_{i\in\gamma^{-1}(k)\cap R} kw(i) \\
&= \sum_{k\in T}\sum_{i\in\gamma^{-1}(k)\cap R} \gamma(i)w(i) && \text{weil } k = \gamma(i) \text{ aus } i \in \gamma^{-1}(k) \text{ folgt} \\
&= \sum_{i\in\gamma^{-1}(T)\cap R} \gamma(i)w(i) \\
&= \sum_{i\in R} \gamma(i)w(i) && \text{weil } \gamma^{-1}(T)\cap R = R
\end{aligned}$$

Damit ist (a) für diskrete Zufallsvariable bewiesen.

Der Beweis von (b) für Zufallsvariable mit Wahrscheinlichkeitsvektoren geht analog. Der Beweis für Zufallsvariable mit Wahrscheinlichkeitsdichten ist kompliziert und soll nicht ausgeführt werden. □

Jetzt können wir die Eigenschaften des Erwartungswertes herleiten.

Satz 32: *Seien X und Y Zufallsvariable und a und b in $\mathbb{R}$. 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.

(a) Wir wenden Satz 31 (a) mit $\gamma(x) = ax + b$ an. Ist $w(i)$ mit $i \in R$ der Wahrscheinlichkeitsvektor von X, so haben wir

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

da $\sum_{i\in R} iw(i) = \mathrm{E}(X)$ und $\sum_{i\in R} w(i) = 1$ gilt.

(b) Sei $u(i,j)$ mit $i \in R$ und $j \in S$ der gemeinsame Wahrscheinlichkeitsvektor von X und Y. Aus Satz 31 (b) mit $\gamma(x,y) = x + y$ folgt $\mathrm{E}(X + Y) = \sum_{i\in R}\sum_{j\in S}(i + j)u(i,j)$. Aus Satz 31 (b) mit $\gamma(x,y) = x$ folgt $\mathrm{E}(X) = \sum_{i\in R}\sum_{j\in S} iu(i,j)$. Aus Satz 31 (b) mit $\gamma(x,y) = y$ folgt $\mathrm{E}(Y) = \sum_{i\in R}\sum_{j\in S} ju(i,j)$. Das ergibt $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

(c) Seien $w_1(i)$ mit $i \in R$ und $w_2(j)$ mit $j \in S$ die Wahrscheinlichkeitsvektoren von X und Y. Da X und Y unabhängig sind, ist $u(i,j) = w_1(i)w_2(j)$ mit $i \in R$ und $j \in S$

der gemeinsame Wahrscheinlichkeitsvektor von X und Y. In Satz 28 wurde dieses Resultat für Zufallsvariable mit Wahrscheinlichkeitsdichten gezeigt. Es gilt auch für diskrete Zufallsvariable. Wir wenden Satz 31 (b) mit $\gamma(x,y) = xy$ an und erhalten

$$\mathrm{E}(XY) = \sum_{i\in R}\sum_{j\in S} iju(i,j) = \sum_{i\in R}\sum_{j\in S} ijw_1(i)w_2(j) = \sum_{i\in R} iw_1(i)\sum_{j\in S} jw_2(j) = \mathrm{E}(X)\mathrm{E}(Y)$$

Damit ist auch (c) gezeigt.

Für kontinuierliche Zufallsvariable geht alles analog. Statt der Summen hat man Integrale und statt der Wahrscheinlichkeitsvektoren hat man Wahrscheinlichkeitsdichten. $\square$

Die Varianz $V(X)$ einer Zufallsvariablen X kann man ganz allgemein definieren durch $\mathrm{V}(X) = \mathrm{E}((X-m)^2)$ mit $m = \mathrm{E}(X)$. Die in Kapitel 12 gegebenen Definitionen der Varianz für diskrete und kontinuierliche Zufallsvariable sind Spezialfälle davon, wie Satz 31 mit $\gamma(x) = (x-m)^2$ zeigt. Aus den in Satz 32 bewiesenen Eigenschaften des Erwartungswertes kann man die Eigenschaften der Varianz herleiten.

Satz 33: *Seien X und Y Zufallsvariable und a und b in $\mathbb{R}$. Dann gilt*
(a) $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$
(b) $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$
(c) $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$, *wenn X und Y unabhängig sind*

Beweis: Wir setzen $m = \mathrm{E}(X)$ und $l = \mathrm{E}(Y)$.
(a) Mit Hilfe von Satz 32 (a) und (b) folgt

$$\mathrm{V}(X) = \mathrm{E}((X-m)^2) = \mathrm{E}(X^2 - 2mX + m^2) = \mathrm{E}(X^2) - 2m\mathrm{E}(X) + m^2 = \mathrm{E}(X^2) - m^2$$

Damit ist $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$ gezeigt.
(b) Aus Satz 32 (a) folgt $\mathrm{E}(aX+b) = am+b$. Damit erhalten wir

$$\mathrm{V}(aX+b) = \mathrm{E}((aX+b-am-b)^2) = \mathrm{E}(a^2(X-m)^2) = a^2\mathrm{E}((X-m)^2) = a^2\mathrm{V}(X)$$

wobei noch einmal Satz 32 (a) verwendet wurde, und (b) ist gezeigt.
(c) Aus Satz 32 (b) folgt $\mathrm{E}(X+Y) = m+l$. Mit Hilfe von (a) erhalten wir

$$\mathrm{V}(X+Y) = \mathrm{E}((X+Y)^2) - (m+l)^2 = \mathrm{E}(X^2+2XY+Y^2) - m^2 - 2ml - l^2$$

Da $\mathrm{E}(X^2+2XY+Y^2) = \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}(Y^2)$ aus Satz 32 (b) und $\mathrm{E}(XY) = ml$ aus Satz 32 (c) folgt, ergibt sich aus obiger Gleichung

$$\mathrm{V}(X+Y) = \mathrm{E}(X^2) - m^2 + \mathrm{E}(Y^2) - l^2 = \mathrm{V}(X) + \mathrm{V}(Y)$$

wobei nochmals (a) verwendet wurde. $\square$

IV. Statistik

In der Statistik geht es darum, aus gesammelten Daten Schlussfolgerungen zu ziehen. Wir führen einige typische Beispiele an. 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 Vertrauensintervalle und statistische Tests. Ein Vertrauensintervall ist ein Intervall, das aus den erhobenen Daten berechnet wird und das den zu messenden Wert (die durchschnittliche Füllmenge, den Anteil der A-Wähler) mit einer vorgegebenen Wahrscheinlichkeit enthält. Beim statistischen Test geht es darum, eine im voraus aufgestellte Aussage durch die erhobenen Daten zu entkräften oder zu bestätigen.

20. Vertrauensintervalle 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. Durch mehrmaliges Messen versucht man, μ mit größerer Sicherheit zu bestimmen.

Man führt den Messvorgang n Mal durch und bezeichnet die dabei erhaltenen Messwerte mit $X_1, X_2, \ldots, X_n$. Jede dieser n Zufallsvariablen hat nach den oben gemachten Annahmen die $N(\mu, \sigma)$-Verteilung. Da die Messungen unabhängig voneinander durchgeführt werden, sind auch die Zufallsvariablen $X_1, X_2, \ldots, X_n$ voneinander 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 $X_1, X_2, \ldots, X_n$ ein Intervall I, das μ mit Wahrscheinlichkeit γ enthält. So ein Intervall I nennt man γ-Vertrauensintervall (oder γ-Konfidenzintervall) für die Größe μ. Für γ gibt es verschiedene Namen, wie zum Beispiel Sicherheitswahrscheinlichkeit, Konfidenzniveau oder Überdeckungswahrscheinlichkeit. Das Vertrauensintervall hängt von den gemessenen Werten $X_1, X_2, \ldots, X_n$ ab, also von zufälligen Einflüssen. Führt man so eine Serie von n Messungen mehrere Male durch und berechnet aus jeder das γ-Vertrauensintervall, dann wird man verschiedene Intervalle erhalten. Auch wird μ nicht immer im Vertrauensintervall enthalten sein. Ist zum Beispiel $\gamma = 0.95$, dann wird μ mit einer Wahrscheinlichkeit von 95 % im Vertrauensintervall liegen, das heißt bei wiederholter Bestimmung des Vertrauensintervalls wird sich μ im Durchschnitt jedes zwanzigste Mal außerhalb des Intervalls befinden.

Normalverteilte Messwerte kann man auch bei anderen Fragestellungen annehmen. Die Abfüllmengen von maschinell gefüllten Flaschen kann man ebenfalls als Messwerte, die zufälligen Schwankungen unterliegen, auffassen. Hier misst die Abfüllmaschine. Jedem Menschen wird die Körpergröße zufällig (?) von der Natur zugemessen, sodass man auch die Körpergröße oft als normalverteilt annimmt. Die Natur erlaubt sich etwas größere Abweichungen.

Um dieses so definierte γ-Vertrauensintervall zu berechnen, benötigen wir folgende Definition.

Definition: Für $\beta \in (0,1)$ sei t_β die eindeutige Lösung von $\Phi(t_\beta) = 1 - \beta$. Diese Lösung ist eindeutig, da $\Phi : \mathbb{R} \to (0,1)$ eine invertierbare Abbildung ist (Satz 25 (c)).

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

Satz 34: *Sei $0 < \gamma < 1$ gegeben. Seien $X_1, X_2, \ldots, X_n$ 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 und $\beta = \frac{1-\gamma}{2}$. Für $I = [M - \frac{\sigma t_\beta}{\sqrt{n}}, M + \frac{\sigma t_\beta}{\sqrt{n}}]$ gilt dann $P(\mu \in I) = \gamma$, das heißt I ist ein γ-Vertrauensintervall für μ.*

Beweis: Aus Satz 30 folgt, dass $X_1 + X_2 + \cdots + X_n$ die $N(n\mu, \sqrt{n}\sigma)$-Verteilung hat. Nach Satz 24 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

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

Wir formen dieses Ereignis um

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

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

Jetzt ist es leicht Beispiele zu rechnen. Man braucht ja nur in die Formel einsetzen.

Beispiel 56: 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 %-Vertrauensintervall.

Wir setzen in die Formel ein. Wegen $\gamma = 0.95$ folgt $\beta = \frac{1-\gamma}{2} = 0.025$. Man findet $t_\beta = 1.96$ in einer Tabelle. Jetzt erhalten wir $\frac{\sigma t_\beta}{\sqrt{n}} = \frac{0.05 \cdot 1.96}{\sqrt{25}} = 0.02$ 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 57: Die Füllmengen 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 %-Vertrauensintervall für den durchschnittlichen Packungsinhalt μ.

Wegen $\gamma = 0.99$ folgt $\beta = \frac{1-\gamma}{2} = 0.005$. Man findet $t_\beta = 2.58$ in einer Tabelle. Weiters berechnen wir $M = \frac{1}{10}(1.05 + 0.99 + \cdots + 0.98) = \frac{10.1}{10} = 1.01$ und $\frac{\sigma t_\beta}{\sqrt{n}} = \frac{0.02 \cdot 2.58}{\sqrt{10}} = 0.019$. 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 Vertrauensintervall 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 Vertrauensintervall liegt. Die andere ist die Wahl des Stichprobenumfangs. Hat man sich für eine statistische Sicherheit γ entschieden, dann ist auch t_β festgelegt und man kann die Länge $|I|$ des Vertrauensintervalls I durch den Stichprobenumfang n steuern. Aus der Formel in Satz 34 folgt $|I| = \frac{2\sigma t_\beta}{\sqrt{n}}$. Vergrößert man den Stichprobenumfang n, dann wird das Vertrauensintervall kleiner.

Beispiel 58: Wieviele Messungen muss man in Beispiel 56 durchführen, damit die Länge des 95 %-Vertrauensintervalls höchstens 6 cm beträgt?

Für $\gamma = 0.95$ haben wir $t_\beta = 1.96$ gefunden. Es wird verlangt, dass $|I| \le 0.06$ gilt, also $\frac{2\sigma t_\beta}{\sqrt{n}} \le 0.06$. Daraus folgt $\sqrt{n} \ge \frac{2\sigma t_\beta}{0.06} = \frac{2\cdot 0.05\cdot 1.96}{0.06} = 3.27$ und $n \ge 10.48$. Man muss also mindestens 11 Messungen durchführen.

Beispiel 59: Wieviele Packungen muss man in Beispiel 57 überprüfen, damit die Länge des 99 %-Vertrauensintervalls höchstens 0.01 kg beträgt?

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

Neben den zweiseitigen Vertrauensintervallen gibt es auch einseitige Vertrauensintervalle, die wir im nächsten Satz behandeln.

Satz 35: *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 $I_1 = (-\infty, M + \frac{\sigma t_{1-\gamma}}{\sqrt{n}}]$ und $I_2 = [M - \frac{\sigma t_{1-\gamma}}{\sqrt{n}}, \infty)$ γ-Vertrauensintervalle für μ.*

Beweis: Weil $\sqrt{n}\frac{M-\mu}{\sigma}$ die $N(0,1)$-Verteilung hat, folgt

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

Daraus folgt $P(\mu \le M + \frac{\sigma t_{1-\gamma}}{\sqrt{n}}) = \gamma$, also $P(\mu \in I_1) = \gamma$. Somit ist das Intervall I_1 ein γ-Vertrauensintervall für μ.
Analog folgt aus $P(\sqrt{n}\frac{M-\mu}{\sigma} \le t_{1-\gamma}) = \Phi(t_{1-\gamma}) = \gamma$, dass $P(\mu \in I_2) = \gamma$ gilt. Somit ist auch I_2 ein γ-Vertrauensintervall für μ. □

Beispiel 60: Aus den in Beispiel 57 erhobenen Packungsinhalten soll das einseitige nach oben offene 99 %-Vertrauensintervall I_2 berechnet werden.

Es wurde $\sigma = 0.02$ kg angegeben und $M = 1.01$ kg berechnet. Für $\gamma = 0.99$ findet man $t_{1-\gamma} = 2.33$ in einer Tabelle. Es folgt $\frac{\sigma t_{1-\gamma}}{\sqrt{n}} = \frac{0.02\cdot 2.33}{\sqrt{10}} = 0.0145$ und daraus $I_2 = [0.9955, \infty)$. Der Mittelwert der Packungsinhalte 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 $S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - M)^2}$. Man kann zeigen, dass dann das Intervall $K = [M - \frac{S v_{\beta,n-1}}{\sqrt{n}}, M + \frac{S v_{\beta,n-1}}{\sqrt{n}}]$ ein γ-Vertrauensintervall für μ ist, wobei $\beta = \frac{1-\gamma}{2}$ und

$v_{\beta,m}$ genauso definiert wird wie t_β, jedoch statt der $N(0,1)$-Verteilung die sogenannte $T(m)$-Verteilung verwendet wird. Man findet $v_{\beta,m}$ in einer Tabelle.

Beispiel 61: In Beispiel 57 sei σ unbekannt. Aus den dort angegebenen Packungsinhalten bestimme man ein 99 %-Vertrauensintervall.

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 57 wurde $M = 1.01$ ermittelt. Als nächstes berechnen wir

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

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

21. Vertrauensintervalle für Prozentsätze

Will die Partei A den Prozentsatz der Wahlberechtigten, von denen sie bei der bevorstehenden Wahl gewählt wird, wissen, dann muss sie sich damit begnügen, eine nicht zu große Anzahl von zufällig bestimmten Wahlberechtigten zu befragen, und daraus ihre Schlüsse ziehen.

Sei also p der unbekannte Prozentsatz der A-Wähler unter den Wahlberechtigten. Man gibt eine Sicherheitswahrscheinlichkeit γ vor und zieht eine zufällige Stichprobe vom Umfang n mit Zurücklegen aus der Menge aller Wahlberechtigten. Sei X die Anzahl der A-Wähler in der Stichprobe. Aus dieser Anzahl X soll ein Intervall I berechnet werden, das p mit einer Wahrscheinlichkeit $\geq \gamma$ enthält. Man nennt I ein γ-Vertrauensintervall für den Prozentsatz p der A-Wähler.

Natürlich lässt sich diese Vorgangsweise auch in vielen anderen Fällen anwenden, nämlich immer dann, wenn man Prozentsätze schätzen will. Das kann der Prozentsatz der Raucher in der Gesamtbevölkerung sein, oder der Prozentsatz der defekten Glühbirnen in der Tagesproduktion einer Firma, oder vieles mehr. Wir wollen von der zu untersuchenden Eigenschaft sprechen. Die zu untersuchende Eigenschaft ist dann, ein A-Wähler zu sein, ein Raucher zu sein oder, bei den Glühbirnen, defekt zu sein.

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 werden daher X einfach als $N(\mu,\sigma)$-verteilt mit $\mu = np$ und $\sigma = \sqrt{np(1-p)}$ annehmen. 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. Der Stichprobenumfang ist jedoch gegenüber der Gesamtmenge, aus der gezogen wird, so klein, dass auch dieser Fehler vernachlässigt werden kann.

Um ein γ-Vertrauensintervall zu berechnen, verwenden wir den folgenden Hilfssatz, dessen Beweis dem Leser überlassen bleibt.

Hilfssatz: *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 36: *Sei p der unbekannte Prozentsatz mit der die zu untersuchende Eigenschaft in der Gesamtmenge, aus der gezogen wird, vorkommt. Sei n der Stichprobenumfang und*

X die Anzahl, mit der diese Eigenschaft in der Stichprobe auftritt. Man wählt $\gamma \in (0,1)$. Sei $\beta = \frac{1-\gamma}{2}$ und $I = [\frac{1}{n}X - \frac{t_\beta}{2\sqrt{n}}, \frac{1}{n}X + \frac{t_\beta}{2\sqrt{n}}]$. Dann gilt $P(p \in I) \geq \gamma$, das heißt I ist ein γ-Vertrauensintervall 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

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

Wir formen dieses Ereignis um

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

Sei $\tilde{I} = [\frac{1}{n}X - \frac{t_\beta}{\sqrt{n}}\sqrt{p(1-p)}, \frac{1}{n}X + \frac{t_\beta}{\sqrt{n}}\sqrt{p(1-p)}]$. 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 $\tilde{I} \subset I$. Aus $P(p \in \tilde{I}) = \gamma$ folgt dann $P(p \in I) \geq \gamma$. □

Bemerkung: Im Beweis von Satz 36 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 Vertrauensintervall 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.

Ist $0.3 \leq p \leq 0.7$, 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. Ist $0.2 \leq p \leq 0.8$, 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 Vertrauensintervall. 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 und man auf andere Weise bessere Vertrauensintervalle erhält.

Beispiel 62: In einer Stichprobe von $n = 400$ Wahlberechtigten waren 164 A-Wähler. Gesucht ist ein 95 %-Vertrauensintervall für den Anteil p der A-Wähler unter allen Wahlberechtigten.

Wir berechnen das Intervall I aus Satz 36. Wegen $\gamma = 0.95$ folgt $\beta = \frac{1-\gamma}{2} = 0.025$. Man findet $t_\beta = 1.96$ in einer Tabelle. Jetzt erhalten wir $\frac{t_\beta}{2\sqrt{n}} = \frac{1.96}{2\sqrt{400}} = 0.049$. Wegen $\frac{1}{n}X = \frac{164}{400} = 0.41$ folgt $I = [0.361, 0.459]$.

Beispiel 63: Wie groß muss man den Stichprobenumfang n wählen, wenn die Länge des 95 %-Vertrauensintervalls I höchstens 0.02 sein soll?

Die Länge von I ist $\frac{t_\beta}{\sqrt{n}}$. Es wird also verlangt, dass $\frac{t_\beta}{\sqrt{n}} \leq 0.02$ gilt. Für $\gamma = 0.95$ wurde in Beispiel 62 $t_\beta = 1.96$ ermittelt. Es folgt $\sqrt{n} \geq \frac{1.96}{0.02} = 98$, das heißt $n \geq 9604$.

Bemerkung: Bei den hier auftretenden Stichprobenumfängen von einigen Hundert oder sogar Tausend ist die Approximation durch die Normalverteilung sicher gerechtfertigt.

Genauso wie das zweiseitige Vertrauensintervall I in Satz 36 berechnet man auch die einseitigen γ-Vertrauensintervalle $I_1 = (-\infty, \frac{1}{n}X + \frac{t_{1-\gamma}}{2\sqrt{n}}]$ und $I_2 = [\frac{1}{n}X - \frac{t_{1-\gamma}}{2\sqrt{n}}, \infty)$. Da p aber ein Prozentsatz ist und daher immer im Intervall $[0,1]$ liegt, kann man auch $I_1 = [0, \frac{1}{n}X + \frac{t_{1-\gamma}}{2\sqrt{n}}]$ und $I_2 = [\frac{1}{n}X - \frac{t_{1-\gamma}}{2\sqrt{n}}, 1]$ schreiben.

Beispiel 64: Für das Umfrageergebnis aus Beispiel 62 soll das nach oben offene einseitige 95 %-Vertrauensintervall I_2 berechnet werden.
Für $\gamma = 0.95$ finden wir $t_{1-\gamma} = 1.65$ in einer Tabelle. Es folgt $\frac{t_{1-\gamma}}{2\sqrt{n}} = \frac{1.65}{2\sqrt{400}} = 0.041$. Wegen $\frac{1}{n}X = \frac{164}{400} = 0.41$ erhält man $I_2 = [0.369, 1]$.

Was soll man tun, wenn p in der Nähe von 0 oder von 1 liegt, sodass man das Vertrauensintervall I nicht mehr verwenden will? Eine – etwas unexakte – Möglichkeit besteht darin, dass man das im Beweis von Satz 36 berechnete Intervall $\tilde{I}$ nimmt und dort den unbekannten Prozentsatz p durch den Prozentsatz $\frac{X}{n}$, der in der Stichprobe vorkommt, ersetzt. Man erhält dadurch das Intervall

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

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

Beispiel 65: Der Anteil p der defekten Glühbirnen in der Tagesproduktion einer Fabrik ist sicher kleiner als 10 %. Man zieht eine Stichprobe von 1600 Glühbirnen und findet 128 defekte darunter. Gesucht ist ein 99 %-Vertrauensintervall für p.
Da p klein ist, berechnen wir J. Wegen $\gamma = 0.99$ folgt $\beta = \frac{1-\gamma}{2} = 0.005$. Man findet $t_\beta = 2.58$ in einer Tabelle. Weiters folgt $\frac{1}{n}X = \frac{128}{1600} = 0.08$. Jetzt erhalten wir $\frac{t_\beta}{\sqrt{n}}\sqrt{\frac{X}{n}(1-\frac{X}{n})} = \frac{2.58}{\sqrt{1600}}\sqrt{0.08\cdot 0.92} = 0.0175$. Daraus ergibt sich $J = [0.0625, 0.0975]$.
Für das Intervall I hätte man $I = [0.048, 0.112]$, also etwas viel schlechteres, erhalten.

Es gibt auch eine Möglichkeit, auf exakte Weise ein Vertrauensintervall zu bestimmen, das besser ist als I. Wir gehen von der im Beweis von Satz 36 bewiesenen Gleichung

$$P(-t_\beta \le \frac{X-np}{\sqrt{np(1-p)}} \le t_\beta) = \gamma$$

aus und formen das darin vorkommende Ereignis um

$$\begin{aligned}
-t_\beta \le \frac{X-np}{\sqrt{np(1-p)}} \le t_\beta \;&\Leftrightarrow\; \frac{(X-np)^2}{np(1-p)} \le t_\beta^2 \\
&\Leftrightarrow\; X^2 - 2Xnp + n^2p^2 \le t_\beta^2 np - t_\beta^2 np^2 \\
&\Leftrightarrow\; p^2(n^2 + nt_\beta^2) - 2p(nX + \tfrac{1}{2}nt_\beta^2) + X^2 \le 0 \\
&\Leftrightarrow\; \lambda_1 \le p \le \lambda_2
\end{aligned}$$

wobei λ_1 und λ_2 die beiden Lösungen der quadratischen Gleichung

$$\lambda^2(n^2 + nt_\beta^2) - 2\lambda(nX + \tfrac{1}{2}nt_\beta^2) + X^2 = 0$$

sind. Setzt man $K = [\lambda_1, \lambda_2]$, dann haben wir $P(p \in K) = \gamma$ bewiesen, das heißt K ist ein γ-Vertrauensintervall für p. Löst man diese quadratische Gleichung, so erhält man

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

Dieses Vertrauensintervall K ist zwar komplizierter als I, aber wesentlich genauer, wenn p in der Nähe von 0 oder 1 liegt.

Beispiel 66: Von 900 überprüften Transistoren waren 30 defekt. Gesucht ist ein 95 %-Vertrauensintervall für den unbekannten Anteil p von defekten Transistoren in der Gesamtproduktion.

Wegen $\gamma = 0.95$ folgt $\beta = \frac{1-\gamma}{2} = 0.025$. Man findet $t_\beta = 1.96$ in einer Tabelle. Berechnet man I, so ergibt sich $I = [\frac{30}{900} - \frac{1.96}{2\sqrt{900}}, \frac{30}{900} + \frac{1.96}{2\sqrt{900}}] = [0.0006, 0.066]$.
Zum Vergleich wollen wir das Vertrauensintervall K bestimmen. Wir berechnen
$Y = t_\beta\sqrt{\frac{X(n-X)}{n} + \frac{1}{4}t_\beta^2} = 1.96\sqrt{\frac{30(900-30)}{900} + \frac{1}{4} \cdot 1.96^2} = 10.73$
$X + \frac{1}{2}t_\beta^2 = 30 + \frac{1}{2} \cdot 1.96^2 = 30.96$
$n + t_\beta^2 = 900 + 1.96^2 = 903.84$
Setzt man das in die obige Formel für K ein, so erhält man
$K = [\frac{30.96-10.73}{903.84}, \frac{30.96+10.73}{903.84}] = [0.0224, 0.0461]$.
Beide Intervalle, I und K, sind γ-Vertrauensintervalle, aber I ist fast dreimal so lang wie K. Somit hat es sich ausgezahlt, das kompliziertere Intervall K zu berechnen.

22. Statistische Tests für Prozentsätze

Statistische Tests verwendet man, wenn man etwas nachweisen will. Bei der Produktion von Glühbirnen zum Beispiel kann man einen noch tolerierbaren Ausschussanteil p_0 festlegen. Die Firma, die die Glühbirnen produziert, will dann nachweisen, dass der unbekannte Anteil p an defekten Glühbirnen $< p_0$ 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. Jetzt kommt ein neues Medikament gegen dieselbe Krankheit auf den Markt. Man will nachweisen, dass der noch unbekannte Prozentsatz p der behandelten Patienten, bei denen es hilft, größer als p_0 ist.

Die Vorgangsweise bei einem statistischen Test ist die folgende. Das Gegenteil der nachzuweisenden Aussage wird als Hypothese H_0 und die nachzuweisende Aussage als Alternative H_1 formuliert. Im Beispiel mit den Glühbirnen führt das zum Test

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

Im zweiten Beispiel, wo das neue Medikament geprüft wird, wird

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

getestet. Die Entscheidung für H_0 oder für H_1 wird auf Grund einer Stichprobe getroffen. Daher wird diese Entscheidung auch nur mit einer gewissen Wahrscheinlichkeit, die möglichst 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. Sei X die Anzahl, mit der die zu untersuchende Eigenschaft in der Stichprobe vorkommt, im Beispiel mit den Glühbirnen also die Anzahl der defekten Glühbirnen unter den n gezogenen, und bei der Prüfung des neuen Medikaments die Anzahl der Patienten unter den n behandelten, bei denen es wirkt. Auf Grund dieser Anzahl X trifft man die Entscheidung, die Hypothese H_0 zu verwerfen oder nicht zu verwerfen. Die Entscheidungsregel soll jedoch so sein, dass die Wahrscheinlichkeit, die Hypothese H_0 zu verwerfen, obwohl sie richtig ist, höchstens α beträgt.

Kommt es zur Verwerfung von H_0, dann ist die Wahrscheinlichkeit, dass H_0 richtig ist, sehr klein, und die nachzuweisende Aussage H_1 hat eine sehr große Wahrscheinlichkeit. Der Nachweis von H_1 ist dann gelungen. Wird nicht verworfen, dann spricht der Test nicht deutlich genug gegen H_0. Der Nachweis von H_1 ist dann nicht gelungen.

Jetzt müssen wir noch eine entsprechende Entscheidungsregel aufstellen. Die möglichen Werte von X sind die Anzahlen, mit der die zu untersuchende Eigenschaft in der Stichprobe vom Umfang n auftritt. Daher hat X den Wertebereich $R = \{0, 1, 2, \dots, n\}$. Wir wählen eine Teilmenge V von R und verwerfen die Hypothese H_0 immer dann, wenn sich nach dem Ziehen der Stichprobe herausstellt, dass X in V liegt. Man nennt V den Verwerfungsbereich des Tests. Aus den oben durchgeführten Überlegungen ergeben sich folgende Bedingungen.
(1) V enthält die Werte aus R, die am meisten gegen H_0 und für H_1 sprechen
(2) V ist maximal, aber so, dass $P(X \in V) \le \alpha$ 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 $\le \alpha$ sein, wenn H_0 richtig ist. Unter dieser Bedingung soll jedoch V möglichst groß sein. Wenn man V zu klein macht, dann wird zwar das Risiko kleiner, H_0 bei Richtigkeit zu verwerfen, aber die Wahrscheinlichkeit, H_0 bei Unrichtigheit nicht zu verwerfen, wird größer. Das Risiko, H_0 bei Richtigkeit zu verwerfen, kontrolliert man durch die Irrtumswahrscheinlichkeit α. Das Risiko, H_0 bei Unrichtigheit nicht zu verwerfen, macht man dadurch möglichst klein, dass man V maximal wählt.

Jetzt wollen wir die in (1) und (2) formulierten Bedingungen verwenden, um den Verwerfungsbereich auch auszurechnen. Dazu führen wir eine Definition ein und beweisen einen Hilfssatz.

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

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

Beweis: Wir definieren die Funktion $h(p)$ für $p \in (0,1)$ durch

$$h(p) = \Psi_{n,p}(k) = \binom{n}{0}(1-p)^n + \binom{n}{1} p(1-p)^{n-1} + \binom{n}{2} p^2(1-p)^{n-2} + \dots + \binom{n}{k} p^k(1-p)^{n-k}$$

Der Satz ist bewiesen, wenn wir $h'(p) \le 0$ zeigen.

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

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

$$h'(p) = -\binom{n}{k}(n-k)p^k(1-p)^{n-k-1} \leq 0$$

und der Hilfssatz ist bewiesen. □

Satz 37: *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 $1 - \Psi_{n,p_0}(k_0 - 1) \leq \alpha$ gilt, und wählt $V = \{k_0, k_0 + 1, \ldots, n\}$, dann erfüllt V die Bedingungen (1) und (2) für diesen Test, ist also der Verwerfungsbereich.*

Beweis: Sei X die Anzahl, mit der die zu untersuchende Eigenschaft in der Stichprobe vom Umfang n vorkommt. Nach Satz 18 hat X die $B(n,p)$-Verteilung.

Große Werte von X sprechen am meisten gegen die Hypothese H_0, die ja einen Anteil $\leq p_0$ für die zu untersuchende Eigenschaft behauptet. Um (1) zu erfüllen, wählen wir $V = \{k_0, k_0 + 1, \ldots, n\}$. Dabei ist k_0 so zu bestimmen, dass auch (2) gilt.

Um (2) zu erfüllen, müssen wir V möglichst groß machen, sodass jedoch $P(X \in V) \leq \alpha$ immer dann gilt, wenn $p \leq p_0$ erfüllt ist. Da X die $B(n,p)$-Verteilung hat, und daher $P(X \in V) = P(X \geq k_0) = 1 - P(X \leq k_0 - 1) = 1 - \Psi_{n,p}(k_0 - 1)$ gilt, bedeutet das, k_0 minimal zu wählen, sodass $1 - \Psi_{n,p}(k_0 - 1) \leq \alpha$ für alle $p \leq p_0$ gilt. Aus dem Hilfssatz folgt aber, dass $1 - \Psi_{n,p}(k_0 - 1) \leq 1 - \Psi_{n,p_0}(k_0 - 1)$, wenn $p \leq p_0$ ist. Daher ist (2) äquivalent dazu, k_0 minimal zu wählen, sodass $1 - \Psi_{n,p_0}(k_0 - 1) \leq \alpha$ gilt. Der Satz ist 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.

Gesucht ist also das minimale k_0, sodass $1 - \Psi_{10,0.3}(k_0 - 1) \leq \alpha$ gilt. Der Verwerfungsbereich ist dann $V = \{k_0, k_0 + 1, \ldots, 10\}$. Dazu 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

Gesucht ist das minimale k_0, sodass $\Psi_{10,0.3}(k_0 - 1) \geq 1 - \alpha$ gilt. Das kann man 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 38: *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: Sei X die Anzahl, mit der die zu untersuchende Eigenschaft in der Stichprobe vom Umfang n vorkommt. Nach Satz 18 hat X 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 zu untersuchende Eigenschaft behauptet. Um (1) zu erfüllen, wählen wir $V = \{0, 1, \ldots, k_0\}$.

Um (2) zu erfüllen, müssen wir V möglichst groß machen, sodass $P(X \in V) \leq \alpha$ immer dann gilt, wenn $p \geq p_0$ erfüllt ist. Da aber 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. Der Satz ist gezeigt. □

Meistens ist der Stichprobenumfang n so groß, dass es schwer ist, mit der Binomialverteilung zu rechnen. Man verwendet 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 17.

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 $\Phi(\frac{k_0+\frac{1}{2}-200}{10}) \leq 0.01 = \Phi(-2.33)$. Da Φ streng monoton wachsend ist, folgt $\frac{k_0+\frac{1}{2}-200}{10} \leq -2.33$ und daraus $k_0 \leq 176.2$. 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. Nach dem Ziehen der Stichprobe stellt man fest, dass die zu untersuchende Eigenschaft mit Anzahl s in der Stichprobe vorkommt. Der nächste Satz gibt an, wann s im Verwerfungsbereich liegt, ohne diesen berechnen zu müssen.

Satz 39: *Der Stichprobenumfang sei n und die Irrtumswahrscheinlichkeit sei α.*
(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: Wenn man k vergrößert, dann wird auch $\Psi_{n,p_0}(k)$ größer, wie man aus der Definition ersieht. In (a) ist $V = \{k_0, k_0+1, \ldots, n\}$ der Verwerfungsbereich, wobei k_0 minimal ist mit $\Psi_{n,p_0}(k_0-1) \geq 1-\alpha$. Deshalb gilt $s \in V$, also $s \geq k_0$ genau dann, wenn $\Psi_{n,p_0}(s-1) \geq 1-\alpha$ ist.
Analog erhalten wir (b). Jetzt ist $V = \{0, 1, 2, \ldots, k_0\}$ der Verwerfungsbereich, wobei k_0 maximal ist 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 $\Psi_{n,p_0}(s) \leq \alpha$ ist. □

Wendet man Satz 39 in Beispielen an, so kann man entweder mit der Binomialverteilung rechnen oder die Approximation durch die Normalverteilung verwenden. Dazu rechnen wir jeweils ein Beispiel.

Beispiel 69: Bei Glühbirnen wird ein Ausschussanteil von 0.015 toleriert. Aus einer Lieferung Glühbirnen wird eine Stichprobe vom Umfang $n = 20$ gezogen. Man findet 2 defekte Glühbirnen. Wird die Hypothese $H_0 : p \leq 0.015$ bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ abgelehnt?

Wir verwenden Satz 39 (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 verworfen. Bei einer Irrtumswahrscheinlichkeit $\alpha = 0.05$ kann man es daher als nachgewiesen ansehen, dass der Anteil an defekten Glühbirnen nicht unter der Toleranzgrenze liegt.

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 39 (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 folgt $\Psi_{1000,0.4}(366) \approx \Phi(\frac{366+\frac{1}{2}-1000\cdot 0.4}{\sqrt{1000\cdot 0.4\cdot 0.6}}) = \Phi(-2.16) = 1 - \Phi(2.16) = 0.016$. Wegen $0.016 \leq 0.05$ wird H_0 verworfen. Wir können es als erwiesen ansehen, dass der Raucheranteil durch die Antiraucherkampagne kleiner geworden ist.

Neben den bisher behandelten einseitigen Tests gibt es auch einen zweiseitigen Test. In diesem Fall 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 Anzahl, mit der die zu untersuchende Eigenschaft in der Stichprobe auftritt. Wir suchen einen Verwerfungsbereich, der (1) und (2) erfüllt. Beim zweiseitigen Test sprechen sowohl kleine als auch große Werte gegen die Hypothese. Daher wählen wir $V = \{0, 1, \dots, k_0, k_1, k_1 + 1, \dots, n\}$. Wir bestimmen k_0 maximal, sodass $\Psi_{n,p_0}(k_0) \leq \frac{\alpha}{2}$ erfüllt ist, und k_1 minimal, sodass $1-\Psi_{n,p_0}(k_1-1) \leq \frac{\alpha}{2}$ erfüllt ist. 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, mit der 6 auftritt. Wir testen $H_0 : p = \frac{1}{6}$ gegen $H_1 : p \neq \frac{1}{6}$. Der Würfel wird 80 Mal geworfen. 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 $V = \{0, 1, \dots, 6, 21, 22, \dots, 80\}$. Liegt die Anzahl, mit der 6 unter den 80 Würfen auftritt, in V, dann gilt es bei Irrtumswahrscheinlichkeit 0.05 als erwiesen, 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 39

$$s \in V \quad \Leftrightarrow \quad s \leq k_0 \text{ oder } s \geq k_1 \quad \Leftrightarrow \quad \Psi_{n,p_0}(s) \leq \frac{\alpha}{2} \text{ oder } \Psi_{n,p_0}(s-1) \geq 1 - \frac{\alpha}{2}$$

Das verwenden wir im folgenden Beispiel.

Beispiel 72: Eine Münze wird 900 Mal geworfen. Es 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.

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

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

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

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. Wenn H_1 gilt, dann sollte H_0 mit großer Wahrscheinlichkeit verworfen werden. Um das zu garantieren, muss man den Stichprobenumfang groß genug machen.

Wir tun das für den Test $H_0 : p \leq p_0$ 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 (zum Beispiel $\tilde{\alpha} = \alpha$) und verlangen

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

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

Die Bedingung (i) wird bei jedem Test verlangt und besagt, dass H_0 bei Richtigkeit nur mit kleiner Wahrscheinlichkeit verworfen wird. Es kommt (ii) dazu: Wenn $p \geq p_1$ gilt (das ist fast H_1), dann wird H_0 mit großer Wahrscheinlichkeit verworfen. Wir wollen einen (möglichst kleinen) Stichprobenumfang n und k_0 so bestimmen, 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 sind (i) und (ii) äquivalent zu $1 - \Psi_{n,p_0}(k_0 - 1) \leq \alpha$ und $1 - \Psi_{n,p_1}(k_0 - 1) \geq 1 - \tilde{\alpha}$, das heißt zu $\Psi_{n,p_0}(k_0 - 1) \geq 1 - \alpha$ und $\Psi_{n,p_1}(k_0 - 1) \leq \tilde{\alpha}$.

Beispiel 73: Sei $p_0 = 0.1$ der tolerierbare Anteil an defekten Glühbirnen. Es soll $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$ bei Irrtumswahrscheinlichkeit $\alpha = 0.05$ getestet werden. Der Stichprobenumfang n ist so zu bestimmen, dass H_0 mit Wahrscheinlichkeit ≥ 0.95 abgelehnt wird, wenn $p \geq 0.15$ 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 $\Psi_{n,p}(k_0 - 1) \approx \Phi(\frac{k_0 - \frac{1}{2} - np}{\sqrt{np(1-p)}})$ erhalten wir die zu (i) und (ii) äquivalenten Ungleichungen

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

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

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

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

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 $n \ge 470$. 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 $V = \{62, 63, \ldots, 500\}$ wählen.

Für diesen Test gilt: Wenn $p \le p_0$, dann ist es sehr unwahrscheinlich, dass verworfen wird (siehe (i)). Endet die Durchführung des Tests also mit dem Verwerfen der Hypothese, dann können wir mit ziemlicher Sicherheit behaupten, dass $p > p_0$ gilt. Wenn $p \ge p_1$ gilt, dann ist es sehr wahrscheinlich, dass H_0 verworfen wird (siehe (ii)). Endet die Durchführung des Tests also mit dem Nichtverwerfen von H_0, dann können wir mit ziemlicher Sicherheit behaupten, dass $p < p_1$ gilt. Wir haben also in jedem Fall eine Schlussfolgerung.

23. 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? Tritt jede Augenzahl mit Wahrscheinlichkeit $\frac{1}{6}$ auf?

Das ist ein Beispiel für folgende Situation. Ein Zufallsexperiment hat die Ausfälle $1, 2, 3, \ldots, k$. Weiters sind Wahrscheinlichkeiten $q_1, q_2, \ldots, q_k$ vorgegeben, die natürlich ≥ 0 sind und $\sum_{j=1}^{k} q_j = 1$ erfüllen. Es soll die Hypothese getestet werden, ob für $1 \le j \le k$ 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 $1, 2, \ldots, k$ dabei auftreten, seien $N_1, N_2, \ldots, N_k$, wobei $N_1 + N_2 + \cdots + N_k = n$ 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 $(N_1, N_2, \ldots, N_k)$ der beobachteten Häufigkeiten dem Vektor $(nq_1, nq_2, \ldots, nq_k)$ 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 $G(\frac{k-1}{2}, \frac{1}{2})$-Verteilung. Der Beweis dafür ist schwer. Diese Verteilung heißt auch χ^2-Verteilung. Wieder definiert man $u_{\alpha,m}$ durch $\int_{u_{\alpha,m}}^{\infty} g_m(x)\, dx = \alpha$, wobei g_m die Wahrscheinlichkeitsdichte der $G(\frac{m}{2}, \frac{1}{2})$-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 $P(D \in V) = P(D \geq u_{\alpha,k-1}) = \int_{u_{\alpha,k-1}}^{\infty} g_{k-1}(x)\,dx = \alpha$, wenn die Hypothese richtig ist. Würde man V vergrößern, dann wäre $P(D \in V) > \alpha$.

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 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 74: 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 ist $k = 6$ 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

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

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 $V = [11.07, \infty)$. Daher gibt der Test keinen Hinweis darauf, dass der Würfel gefälscht ist.

Beispiel 75: 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

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

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 $V = [9.49, \infty)$. Daher gibt der Test keinen Hinweis darauf, dass die Anzahl der Unfälle pro Tag nicht die Poissonverteilung mit $\lambda = 0.9$ hat.

Der χ^2-Test 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 $1, 2, \dots, k$ auf und das zweite Merkmal als Zufallsvariable Y mit den möglichen Werten $1, 2, \dots, l$. 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 der gemeinsame Wahrscheinlichkeitsvektor $w(i, j)$ der beiden Zufallsvariablen X und Y in ein Produkt $w_1(i)w_2(j)$ zerfällt, wobei $w_1(i)$ der Wahrscheinlichkeitsvektor von X und $w_2(j)$ der Wahrscheinlichkeitsvektor von Y ist. Wir ersetzen die Wahrscheinlichkeitsvektoren durch die relativen Häufigkeiten aus der Stichprobe, nämlich

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

Für die Unabhängigkeit der beiden Zufallsvariablen spricht, wenn $\frac{N_{ij}}{n} \approx r_i s_j$ gilt, also $N_{ij} \approx n r_i s_j$ für $1 \le i \le k$ und $1 \le j \le l$.

Wir messen den Abstand zwischen diesen Zahlen, die bei Unabhängigkeit nahe sein sollten, durch

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

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 $G(\frac{m}{2}, \frac{1}{2})$-Verteilung mit $m = (k-1)(l-1)$. Wir erhalten genauso wie oben $V = [u_{\alpha,m}, \infty)$ als Verwerfungsbereich zur Irrtumswahrscheinlichkeit α.

Beispiel 76: 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 von D fällt in den Verwerfungsbereich $V = [5.99, \infty)$. Wir können es als erwiesen ansehen, dass Augenfarbe und Haarfarbe nicht unabhängig sind.

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 t_β und $v_{\beta,n}$ (Normalverteilung und T–Verteilung)

$\beta =$	t_β	$v_{\beta,7}$	$v_{\beta,8}$	$v_{\beta,9}$	$v_{\beta,10}$	$v_{\beta,12}$	$v_{\beta,14}$	$v_{\beta,16}$	$v_{\beta,18}$	$v_{\beta,25}$	$v_{\beta,40}$
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	6.36	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

Inhaltsverzeichnis

I. Kombinatorik 1

1. Geordnete Stichproben 1
2. Ungeordnete Stichproben 4
3. Zerlegungen einer Menge 7
4. Anordnungen (Permutationen) 9

II. Wahrscheinlichkeitstheorie 11

5. Zufallsexperiment, Ausfall, Ereignis 11
6. Wahrscheinlichkeit 13
7. Gleichwahrscheinliche Ausfälle 14
8. Geometrische Wahrscheinlichkeit 17
9. Bedingte Wahrscheinlichkeit 17
10. Totale Wahrscheinlichkeit 22

III. Zufallsvariable 25

11. Darstellung von Ereignissen durch Zufallsvariable 25
12. Wahrscheinlichkeitsvektoren und Wahrscheinlichkeitsdichten 26
13. Binomialverteilung und geometrische Verteilung 30
14. Poissonverteilung 33
15. Exponentialverteilung und Gammaverteilung 34
16. Normalverteilung 35
17. Approximation der Binomialverteilung 38
18. Rechnen mit Zufallsvariablen 42
19. Erwartungswert und Varianz 45

IV. Statistik 48

20. Vertrauensintervalle für normalverteilte Messwerte 48
21. Vertrauensintervalle für Prozentsätze 51
22. Statistische Tests für Prozentsätze 54
23. Der χ^2-Test (chi-quadrat-Test) 60

Tabellen 64

Zusammenstellung der wichtigen Verteilungen

Name	Parameter	Abkürzung	Wertebereich	W–Vektor/Dichte	Ew.	Varianz
Binomial-verteilung	$n \in \mathbb{N}$, $p \in (0,1)$	$B(n,p)$	$\{0,1,\dots,n\}$	$w(j) = \binom{n}{j} p^j (1-p)^{n-j}$	np	$np(1-p)$
geometrische Verteilung	$p \in (0,1)$		$\{1,2,\dots\}$	$w(j) = (1-p)p^{j-1}$	$\frac{1}{1-p}$	$\frac{p}{(1-p)^2}$
Poisson-verteilung	$\lambda > 0$	$P(\lambda)$	$\{0,1,2,\dots\}$	$w(j) = e^{-\lambda} \frac{\lambda^j}{j!}$	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	$\lambda > 0$, $r > 0$	$G(\lambda,r)$	$\mathbb{R}^+$	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$