

# Stochastik

Eine Vorlesung für das  
Lehramtsstudium

Verfasst von Franz Hofbauer,  
mit einigen Abänderungen durch Gernot Greschonig

Sommersemester 2017

## Vorwort

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

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

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

Im vierten Kapitel wird die gemeinsame Verteilung mehrerer Zufallsvariable und die Unabhängigkeit von Zufallsvariablen eingeführt. Es geht dann ums Rechnen mit Zufallsvariablen. Für die Wurzel aus einer Zufallsvariable, für die Summe und für den Quotienten zweier Zufallsvariable werden die Wahrscheinlichkeitsdichten bestimmt. Es wird gezeigt, dass eine Summe unabhängiger normalverteilter Zufallsvariablen wieder normalverteilt ist. Schließlich werden noch Rechenregeln für Erwartungswert und Varianz behandelt. Bewiesen werden sie jedoch nur für diskrete Zufallsvariable. Schließlich gibt es noch einen kurzen Ausblick auf den Zentralen Grenzwertsatz und das Gesetz der großen Zahlen.

Das fünfte Kapitel des Skriptums ist der Statistik gewidmet. Zuerst wird die Methode der kleinsten Quadrate behandelt. So gewinnt man Formeln zum Schätzen von Parametern aus vorliegenden Stichproben. Dann folgen Konfidenzintervalle und statistische Tests.

Das Skriptum enthält eine große Anzahl durchgerechneter Beispiele, von denen viele auch im Schulunterricht verwendet werden können.

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

# I. Kombinatorik

Stichproben spielen in der Wahrscheinlichkeitstheorie eine wesentliche Rolle. In der Statistik zum Beispiel besteht die grundlegende Methode ja gerade darin, aus einer zufällig gezogenen Stichprobe Folgerungen zu ziehen. Das Ausrechnen von Wahrscheinlichkeiten kann man oft auf das Abzählen von Stichproben zurückführen. Damit setzen wir uns in diesem ersten Teil auseinander. Das Teilgebiet der Mathematik, das sich mit dem Berechnen 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  $M$  vor uns, aus der eine Stichprobe gezogen wird (zum Beispiel eine Menge von Glühlampen, 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 der Menge  $M$ . Aus dieser  $n$ -elementigen Menge  $M$  ziehen wir der Reihe nach  $k$  Elemente. Es gibt also einen ersten Zug, einen zweiten Zug, und so weiter bis zum  $k$ -ten Zug. Die Ordnung in der Stichprobe ist wesentlich, daher spricht man von geordneten Stichproben. Die Anzahl  $k$  der Elemente in der Stichprobe heißt Stichprobenumfang.

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

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

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

Mit Zurücklegen:				Ohne Zurücklegen:		
$(a, a)$	$(a, b)$	$(a, c)$	$(a, d)$	$(a, b)$	$(a, c)$	$(a, d)$
$(b, a)$	$(b, b)$	$(b, c)$	$(b, d)$	$(b, a)$	$(b, c)$	$(b, d)$
$(c, a)$	$(c, b)$	$(c, c)$	$(c, d)$	$(c, a)$	$(c, b)$	$(c, d)$
$(d, a)$	$(d, b)$	$(d, c)$	$(d, d)$	$(d, a)$	$(d, b)$	$(d, c)$

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

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

Wir wollen diese Anzahlen allgemein ausrechnen. Zuvor eine Definition.

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

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

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

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

**Beispiel 2:** Wie viele mögliche Tipps erlaubt ein Totoschein?

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

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

Die Gäste kommen einer nach dem anderen dran. Für den ersten Gast gibt es 6 mögliche Zimmer. Für den zweiten Gast, der als nächster drankommt, gibt es dann nur mehr 5 freie Zimmer. Für den dritten Gast gibt es nur mehr 4 und für den vierten Gast 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 verallgemeinern die geordneten Stichproben, indem wir bei jedem Zug aus einer anderen Menge ziehen.

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

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

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

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

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

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

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

## 2. Ungeordnete Stichproben

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

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

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

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

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

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

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

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

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

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

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

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

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

**Beispiel 8:** Aus 52 Spielkarten (13  $\heartsuit$ -Karten, 13  $\diamondsuit$ -Karten, 13  $\clubsuit$ -Karten und 13  $\spadesuit$ -Karten) wird eine ungeordnete Stichprobe vom Umfang 11 gezogen. Wie viele solche Stich-

proben gibt es, die 6 ♡-Karten, 3 ♠-Karten und 2 ♣-Karten enthalten?

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

**Bemerkung:** Die Methode aus Beispiel 7 und aus Beispiel 8 kann man verwenden, um die Gleichung  $\sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k} = \binom{N}{n}$  zu zeigen, wobei  $\binom{u}{v} = 0$  zu setzen ist, wenn  $v > u$  gilt. Das geht so: Aus einer Menge von  $N$  Kugeln, von denen  $M$  weiß und  $N - M$  schwarz sind, werden  $n$ -elementige Teilmengen gezogen. Die Anzahl der  $n$ -elementigen Teilmengen, die  $k$  weiße Kugeln und  $n - k$  schwarze Kugeln enthalten, ist  $\binom{M}{k} \binom{N-M}{n-k}$ , da sich jede dieser Teilmengen aus einer  $k$ -elementigen Teilmenge der  $M$  weißen Kugeln und einer  $n - k$ -elementigen Teilmenge der  $N - M$  schwarzen Kugeln zusammensetzen lässt. Die Anzahl aller  $n$ -elementigen Teilmengen ist dann  $\sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k}$ . Die Anzahl aller  $n$ -elementigen Teilmengen ist aber auch  $\binom{N}{n}$ . Daher gilt  $\sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k} = \binom{N}{n}$ .

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

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

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

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

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

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

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

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

**Beispiel 10:** Aus 52 Spielkarten (13 ♡-Karten, 13 ◇-Karten, 13 ♣-Karten und 13 ♠-Karten) wird eine geordnete Stichprobe vom Umfang 17 mit Zurücklegen gezogen. Wie viele solche Stichproben gibt es, die 8 ♡-Karten und 9 ◇-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 ♡-Karten und die übrigen 9 mit ◇-Karten besetzt werden, gibt es  $\binom{17}{8}$  mögliche Platzaufteilungen. Die Anzahl der Stichproben, wo die ersten 8 Plätze mit ♡-Karten und die letzten 9 Plätze mit ◇-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 ♡-Karten und 9 ◇-Karten enthalten.

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

**Beispiel 11:** Aus einer Menge von 14 Kugeln, von denen 9 rot und 5 grün sind, wird eine Stichprobe vom Umfang 7 gezogen. Wie viele verschiedene Stichproben, die 3 rote und 4 grüne Kugeln enthalten, gibt es, wenn

- ungeordnet gezogen wird?
- geordnet mit Zurücklegen gezogen wird?
- geordnet ohne Zurücklegen gezogen wird?

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

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



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

**Bemerkung:** Die hier behandelten ungeordneten Stichproben sind ungeordnete Stichproben ohne Zurücklegen. Es gibt auch ungeordnete Stichproben mit Zurücklegen. Man kann zeigen, dass die Anzahl der ungeordneten Stichproben vom Umfang  $k$  aus einer  $n$ -elementigen Menge mit Zurücklegen gleich  $\binom{n+k-1}{k}$  ist. Ein Beispiel für ungeordnete Stichproben mit Zurücklegen vom Umfang 2 aus der Menge  $\{0, 1, 2, 3, 4, 5, 6\}$  sind Dominosteine. Ihre Anzahl ist  $\binom{7+2-1}{2} = 28$ .

### 3. Zerlegungen einer Menge

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

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

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

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

An diesem Beispiel kann man auch schon erkennen, wie man die Anzahl der möglichen Zerlegungen ermitteln kann. Es gibt  $\binom{5}{2} = 10$  Möglichkeiten, die erste Teilmenge zu

wählen. Zu jeder dieser Möglichkeiten kann man auf  $\binom{3}{1} = 3$  Arten die zweite Teilmenge wählen. Für die dritte Teilmenge gibt es dann nur mehr  $1 = \binom{2}{2}$  Möglichkeit. Insgesamt haben wir also  $\binom{5}{2} \binom{3}{1} \binom{2}{2} = 30$  Möglichkeiten.

**Satz 5:** Sei  $n = k_1 + k_2 + \dots + k_j$ . Die Anzahl der möglichen Zerlegungen einer  $n$ -elementigen Menge  $M$  in  $j$  Teilmengen, von denen die erste  $k_1$  Elemente, die zweite  $k_2$  Elemente, ... und die  $j$ -te  $k_j$  Elemente hat, ist  $\frac{n!}{k_1!k_2!k_3!\dots k_j!}$ . Dieser Ausdruck heißt *Polynomial- oder Multinomialkoeffizient* und wird analog dem Binomialkoeffizienten mit  $\binom{n}{k_1, k_2, \dots, k_j}$  bezeichnet. Für  $j = 2$  (und daher  $k_1 + k_2 = n$ ) gilt  $\binom{n}{k_1, k_2} = \binom{n}{k_1} = \binom{n}{k_2}$

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

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

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

**Beispiel 14:** Aus 52 Spielkarten, die aus 13 ♡-Karten, aus 13 ◇-Karten, aus 13 ♣-Karten und aus 13 ♠-Karten bestehen, wird eine geordnete Stichprobe vom Umfang 17 mit Zurücklegen gezogen. Wie viele solche Stichproben gibt es, die 5 ♡-Karten, 3 ◇-Karten, 7 ♣-Karten und 2 ♠-Karten enthalten.

Wir gehen vor wie in Beispiel 10. Es werden zuerst die 17 Plätze aufgeteilt und zwar in 5 Plätze für die ♡-Karten, in 3 Plätze für die ◇-Karten, in 7 Plätze für die ♣-Karten und in 2 Plätze für die ♠-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 ♡-Karten zu besetzen. Weiters gibt es  $13^3$  Möglichkeiten, die 3 Plätze für die ◇-Karten zu besetzen,  $13^7$  Möglichkeiten, die 7 Plätze für die ♣-Karten zu besetzen, und schließlich  $13^2$  Möglichkeiten, die 2 Plätze für die ♠-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 ♡-Karten, 3 ◇-Karten, 7 ♣-Karten und 2 ♠-Karten enthalten.

#### 4. Anordnungen (Permutationen)

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

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

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

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

**Beispiel 16:** Jemand hat in seinem Zimmer 8 Lampen. Er hat 8 Glühlampen, 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 Glühlampen angeordnet werden. Die verschiedenen Beleuchtungsmöglichkeiten sind die verschiedenen Anordnungen der Glühlampen. 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 Glühlampen auf die übrigen 7 Lampen verteilt. Diese Anzahl ist  $\frac{7!}{2!1!2!2!} = 630$ . Daher gibt es  $2520 - 630 = 1890$  mögliche Beleuchtungen mit nicht rot leuchtender Tischlampe.

## II. Wahrscheinlichkeitstheorie

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

### 1. Zufallsexperiment, Ausfall, Ereignis

Ein Zufallsexperiment ist ein Experiment mit verschiedenen möglichen Ausfällen (Ergebnissen). Beispiele für Zufallsexperimente sind das Werfen eines Würfels oder die Lottoziehung. Die möglichen Ausfälle beim Würfeln sind die Augenzahlen 1 bis 6. Die möglichen Ausfälle des Zufallsexperiments Lottoziehung sind alle 6-elementigen Teilmengen der Menge  $\{1, 2, 3, \dots, 45\}$ . Die möglichen Ausfälle eines Zufallsexperiments fassen wir zu einer Menge zusammen, die üblicherweise mit  $\Omega$  bezeichnet wird. Wir nennen  $\Omega$  die Ausfallsmenge. Beim Zufallsexperiment Würfeln haben wir  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Beim Zufallsexperiment Lottoziehung ist  $\Omega$  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  $\Omega$  aufgefasst. Das Ereignis „gerade Zahl würfeln“ ist zum Beispiel die Teilmenge  $\{2, 4, 6\}$ . Das Ereignis „die Lottoziehung ergibt nur Zahlen  $\leq 7$ “ 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  $\Omega$  für das Zufallsexperiment Lottoziehung. Ein als Menge geschriebenes Ereignis  $A$  tritt genau dann ein, wenn das Zufallsexperiment einen Ausfall liefert, der in  $A$  liegt.

Wirft man zwei Würfel gleichzeitig, so nimmt man unterscheidbare Würfel (einen roten und einen grünen). Die Ausfälle des Zufallsexperiments sind dann Paare von Augenzahlen, wobei die Augenzahl des roten Würfels an die erste Stelle und die Augenzahl des grünen Würfels an die zweite Stelle geschrieben wird. Diese Vorgangsweise wird später das Berechnen von Wahrscheinlichkeiten erleichtern. Die Ausfallsmenge  $\Omega$  ist dann  $\{(i, j) : 1 \leq i, j \leq 6\}$ . 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  $\leq 2$ “ ist die Teilmenge  $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$ .

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

Zieht man eine ungeordnete Stichprobe vom Umfang 2 aus der Menge  $\{a, b, c, d\}$ , dann

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

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

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

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

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

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

Zwei Ereignisse  $A$  und  $B$  heißen unvereinbar, wenn sie nicht gleichzeitig eintreten können. In der Mengensprache bedeutet das, dass die beiden Teilmengen  $A$  und  $B$  von  $\Omega$  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  $\Omega$ .

## 2. Wahrscheinlichkeit

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

„gerade Zahl würfeln“) eintritt. Diese Anzahl bezeichnen wir mit  $N_k(A)$ . Der Quotient  $\frac{N_k(A)}{k}$  ist dann der Anteil (relative Häufigkeit) der Wiederholungen, bei denen  $A$  eintritt. Die Wahrscheinlichkeit  $P(A)$  des Ereignisses  $A$  definieren wir dann durch

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

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

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

**Satz 6:** *Es gilt*

- (a)  $0 \leq P(A) \leq 1$  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  $0 \leq N_k(A) \leq k$  und daraus  $0 \leq \frac{N_k(A)}{k} \leq 1$ . Mit  $k$  gegen  $\infty$  erhält man (a).

Das Ereignis  $\emptyset$  tritt nie ein. Daher gilt  $N_k(\emptyset) = 0$ , woraus  $P(\emptyset) = 0$  folgt. Das Ereignis  $\Omega$  tritt immer ein. Daher gilt  $N_k(\Omega) = k$ , woraus  $P(\Omega) = 1$  folgt.  $\square$

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

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

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  $N_k(A_1) + N_k(A_2) + \dots + N_k(A_n)$  die Anzahl der Wiederholungen, bei denen mindestens eines der Ereignisse  $A_1, A_2, \dots, A_n$  eintritt. Keine der  $k$  Wiederholungen wird mehrfach gezählt, da die Ereignisse ja unvereinbar sind und daher bei jeder Wiederholung höchstens eines der Ereignisse  $A_1, A_2, \dots, A_n$  eintritt.

Andererseits ist  $A_1 \cup A_2 \cup \dots \cup A_n$  gerade das Ereignis, dass mindestens eines der Ereignisse  $A_1, A_2, \dots, A_n$  eintritt. Daraus folgt

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

Dividiert man durch  $k$  und lässt  $k$  gegen  $\infty$  gehen, so folgt das gewünschte Resultat.  $\square$

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

**Satz 8:** *Seien  $A$  und  $B$  Ereignisse. Dann gilt*

- (a)  $P(A \setminus B) = P(A) - P(A \cap B)$
- (b)  $B \subseteq A \Rightarrow P(A \setminus B) = P(A) - P(B)$
- (c)  $B \subseteq A \Rightarrow P(B) \leq P(A)$
- (d)  $P(A') = 1 - P(A)$
- (e)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (f)  $P(A \cup B) \leq P(A) + P(B)$

**Beweis:** (a): Die Ereignisse  $A \setminus B$  und  $A \cap B$  sind disjunkt und ihre Vereinigung ist  $A$ . Aus dem Additionssatz folgt daher  $P(A) = P(A \setminus B) + P(A \cap B)$  und man erhält (a).

- (b): Gilt  $B \subseteq A$ , dann auch  $A \cap B = B$ . Aus (a) erhalten wir  $P(A \setminus B) = P(A) - P(B)$ .
- (c): Aus (b) erhalten wir  $P(A) - P(B) = P(A \setminus B)$ . Aus Satz 6 (a) folgt  $P(A \setminus B) \geq 0$ , womit (c) gezeigt ist.
- (d): Die Mengen  $A$  und  $A'$  sind disjunkt und ihre Vereinigung ist  $\Omega$ . Der Additionssatz ergibt daher  $P(A) + P(A') = P(\Omega)$ . Satz 6 (b) besagt, dass  $P(\Omega) = 1$  gilt. Daraus erhalten wir  $P(A') = 1 - P(A)$  und (d) ist gezeigt.
- (e): Die Ereignisse  $B$  und  $A \setminus B$  sind disjunkt. Ihre Vereinigung ist  $A \cup B$ . Daher erhalten wir aus dem Additionssatz, dass  $P(A \cup B) = P(B) + P(A \setminus B)$  gilt. In (a) wurde  $P(A \setminus B) = P(A) - P(A \cap B)$  gezeigt. Setzt man das ein, so hat man bereits (e).
- (f): Da  $P(A \cap B) \geq 0$  wegen Satz 6 (a) gilt, folgt (f) aus (e).  $\square$

### 3. Gleichwahrscheinliche Ausfälle

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

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

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

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

**Beweis:** Sei  $q$  die Wahrscheinlichkeit, mit der jeder der Ausfälle eintritt, oder genauer, mit der jede einelementige Teilmenge von  $\Omega$  eintritt. Sei  $k = |A|$  die Anzahl der Elemente der Menge  $A$ . Weiters seien  $A_1, A_2, \dots, A_k$  die einelementigen Teilmengen von  $A$ . Da diese disjunkt sind und ihre Vereinigung gleich  $A$  ist, erhalten wir aus dem Additionssatz, dass  $P(A) = P(A_1) + P(A_2) + \dots + P(A_k)$  gilt. Da die Ereignisse  $A_j$  einelementig sind, gilt  $P(A_j) = q$  für  $1 \leq j \leq k$ . Es folgt  $P(A) = qk = q|A|$ . Es ist nur noch  $q$  zu bestimmen. Dazu setzen wir  $A = \Omega$  und erhalten  $P(\Omega) = q|\Omega|$ . Nach Satz 6 gilt  $P(\Omega) = 1$ , sodass  $q = \frac{1}{|\Omega|}$  folgt. Damit erhalten wir  $P(A) = q|A| = \frac{|A|}{|\Omega|}$ , was zu zeigen war.  $\square$

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

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

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

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

- (a) Augensumme 5 zu erhalten?  
 (b) dass mindestens eine 6 auftritt?

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

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

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

**Beispiel 19:** Es wird mit 3 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit,

- (a) Augensumme 5 zu erhalten?
- (b) dass keine 6 auftritt?
- (c) dass genau zwei Mal 6 auftritt?

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

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

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

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

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

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

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

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

Jetzt ist die Ausfallsmenge  $\Omega$  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 22:** Aus einem Kartenspiel mit 20 Karten (5 ♡-Karten, 5 ◇-Karten, 5 ♣-Karten und 5 ♠-Karten) werden 7 Karten ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, 2 ♡-Karten, 2 ◇-Karten und 3 ♣-Karten zu erhalten?

Die Ausfallsmenge  $\Omega$  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 ♡-Karten, 2 ◇-Karten und 3 ♣-Karten enthalten. Da es  $\binom{5}{2}$  2-elementige Teilmengen aus den 5 ♡-Karten,  $\binom{5}{2}$  2-elementige Teilmengen aus den 5 ◇-Karten und  $\binom{5}{3}$  3-elementige Teilmengen aus den 5 ♣-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 genauso gut 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.

Ganz generell erhält man bei Stichproben ohne Zurücklegen dieselben Wahrscheinlichkeiten unabhängig davon, 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 leicht mittels Binomialkoeffizienten berechnen lassen.

#### 4. Geometrische Wahrscheinlichkeit

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

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

wobei  $|X|$  die Länge von  $X$  bedeutet, wenn  $X$  ein Intervall ist, und  $|X|$  die Fläche von  $X$  bedeutet, wenn  $X$  eine Teilmenge des  $\mathbb{R}^2$  ist. Im Falle einer zusammengesetzten Menge (beispielsweise die Vereinigung mehrerer disjunkter Intervalle in  $\mathbb{R}$ ) verwenden wir den Additionssatz. Wesentlich ist hier die Bedingung  $0 < |\Omega| < \infty$ .

Wir untersuchen einige Beispiele.

**Beispiel 23:** Jemand lässt seine mechanische 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 24:** Zwei Personen treffen folgende Vereinbarung. Jede 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 jede eine Viertelstunde wartet, treffen sie einander genau dann, wenn ihre Ankunftszeitpunkte  $x$  und  $y$  höchstens  $\frac{1}{4}$  voneinander entfernt sind. Das heißt  $A = \{(x, y) \in \Omega : |x - y| \leq \frac{1}{4}\}$ . Wegen  $|\Omega| = 1$  und  $|A| = 1 - \frac{3}{4} \cdot \frac{3}{4} = \frac{7}{16}$  folgt  $P(A) = \frac{7}{16}$ .

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

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

## 5. Bedingte Wahrscheinlichkeit

Wir führen ein Zufallsexperiment durch und interessieren uns dafür, ob das Eintreten eines Ereignisses  $B$  ein anderes Ereignis  $A$  begünstigt. Dazu wiederholen wir das Zufallsexperiment  $k$  Mal. Wir berücksichtigen jedoch nur die Wiederholungen, wo das Ereignis  $B$  eintritt. Unter diesen zählen wir die Häufigkeit des Ereignisses  $A$ . Diese ist die Anzahl der Wiederholungen, wo  $A$  und  $B$  eintreten, also  $N_k(A \cap B)$ , wobei wir die Bezeichnung aus Kapitel 2 verwenden. Die bedingte relative Häufigkeit des Eintretens von  $A$  unter  $B$  ist dann  $N_k(A \cap B)/N_k(B)$ . Wir werden hier immer voraussetzen, dass die Wahrscheinlichkeit des Ereignisses  $B$  positiv ist. Dann ist für hinreichend große  $k$  die relative Häufigkeit  $N_k(B)/k$  positiv und damit auch  $N_k(B)$ . Lässt man  $k$  gegen  $\infty$  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

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

das Vertauschen von Limes und Quotientenbildung ist aufgrund der Bedingung  $P(B) = \lim_{k \rightarrow \infty} N_k(B)/k > 0$  zulässig.

Die bedingte Wahrscheinlichkeit  $P(A|B)$  ist eine Maßzahl für die Häufigkeit des Eintretens von  $A$  unter jenen 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. Wegen  $P(A|B) = P(A \cap B)/P(B)$  ist das äquivalent zu  $P(A \cap B) = P(A)P(B)$ .

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

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

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

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

Die Ausfallsmenge  $\Omega$  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 18 bestimmt. Wir haben die Information, dass die Augensumme 7 ist, also dass das Ereignis  $B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$  eingetreten ist. Zu berechnen ist daher  $P(A|B)$ . Es gilt  $A \cap B = \{(1, 6), (6, 1)\}$ , also  $|A \cap B| = 2$ , und  $|B| = 6$ . Aus Satz 10 folgt  $P(A|B) = \frac{2}{6}$ .

Einer der wichtigsten Sätze über bedingte Wahrscheinlichkeiten ist der sogenannte Multiplikationssatz. Er berechnet die Wahrscheinlichkeit des Durchschnitts von Ereignissen.

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

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

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

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

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

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

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

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

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

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

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

Jetzt behandeln wir den Fall, dass zurückgelegt wird. Die Menge enthält 2 rote und 5 gelbe Kugeln, daher gilt wieder  $P(A_1) = \frac{2}{7}$ . Da zurückgelegt wird, wird die Menge durch den ersten Zug nicht verändert. Beim zweiten Mal wird wieder aus derselben Menge gezogen, sodass  $P(A_2|A_1) = \frac{5}{7}$  gilt. Dasselbe gilt für den dritten Zug. Auch beim dritten Mal wird aus der ursprünglichen Menge gezogen, sodass  $P(A_3|A_1 \cap A_2) = \frac{2}{7}$  gilt. Die

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Das Ereignis, zumindest eine gelbe Kugel vor der ersten schwarzen Kugel zu ziehen, zerfällt in die Teilereignisse, die jeweils den Zugfolgen mit den Anfangssequenzen G, RG, RRG und RRRG entsprechen. Diese Teilereignisse sind unvereinbar, da die erste gelbe Kugel in verschiedenen Ziehungen auftritt, und es tritt sicher eine schwarze Kugel auf.

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

Die Wahrscheinlichkeit, zumindest eine gelbe Kugel vor der ersten schwarzen Kugel zu ziehen, ist somit  $\frac{105}{210} = \frac{1}{2}$ .

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

Die Zugfolgen, die zu einem Sieg von Barbara führen, sind  $\tilde{W}S$ ,  $\tilde{W}WS$ ,  $\tilde{W}WW\tilde{W}S$  und  $\tilde{W}WW\tilde{W}WS$ , wobei  $\tilde{\phantom{x}}$  bedeutet, dass Anna zieht. Sonst zieht Barbara. Die Wahrscheinlichkeit, dass Anna  $\tilde{W}\tilde{W}\tilde{W}$  zieht, ist  $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3}$ . Die Wahrscheinlichkeit, dass Barbara  $WWS$  zieht, ist  $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3}$ . Die Wahrscheinlichkeit für die Zugfolge  $\tilde{W}WW\tilde{W}S$  ist somit  $\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3}$ . Die vollständige Rechnung findet man in folgender Tabelle.

Teilergebnis	Wahrscheinlichkeit nach dem Multiplikationssatz
$\tilde{W}S$	$\frac{4}{5} \cdot \frac{1}{5} = \frac{4}{25}$
$\tilde{W}WS$	$\frac{4}{5} \cdot \frac{4}{5} \cdot \frac{1}{4} = \frac{4}{25}$
$\tilde{W}WW\tilde{W}S$	$\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{2}{25}$
$\tilde{W}WW\tilde{W}WS$	$\frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{2}{25}$
Summe (nach dem Additionssatz)	$= \frac{12}{25}$

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

## 6. Totale Wahrscheinlichkeit

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

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

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

**Satz 12:** Es liege eine Zerlegung  $B_1, B_2, \dots, B_n$  der Ausfallsmenge  $\Omega$  vor. Seien weiters  $A, B \subseteq \Omega$  zwei Ereignisse. Dann gilt

$$(a) \quad P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$

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

Man nennt (a) die Formel für die totale Wahrscheinlichkeit und (b) die Formel von Bayes.

**Beweis:** Wir zeigen (a). Es gilt  $\sum_{j=1}^n P(A|B_j)P(B_j) = \sum_{j=1}^n P(A \cap B_j)$  nach Definition der bedingten Wahrscheinlichkeit. Da die Ereignisse  $B_j$  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  $\bigcup_{j=1}^n (A \cap B_j) = A \cap \bigcup_{j=1}^n B_j = A \cap \Omega = A$  und (a) ist bewiesen.

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

Wir rechnen einige typische Beispiele zu diesen Formeln.

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

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

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

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

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

**Beispiel 33:** Eine Versicherung teilt die Autofahrer in zwei Typen ein, in Risikofahrer und in Sicherheitsfahrer. Die Wahrscheinlichkeit, dass ein Sicherheitsfahrer in einem Jahr einen Unfall hat, ist 0,06. Die Wahrscheinlichkeit, dass ein Risikofahrer in einem Jahr einen Unfall hat, ist 0,6. Die Versicherung weiß aus Erfahrung, dass  $\frac{5}{6}$  der Autofahrer Sicherheitsfahrer und  $\frac{1}{6}$  der Autofahrer Risikofahrer sind. Ein Autofahrer schließt eine Versicherung ab (man sieht ihm natürlich nicht an, von welchem Typ er ist). Wie groß ist die Wahrscheinlichkeit, dass er nächstes Jahr einen Unfall haben wird?

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

Gefragt ist nach der Wahrscheinlichkeit des Ereignisses  $A$ , dass der Autofahrer innerhalb des nächsten Jahres einen Unfall hat. Wenn der Autofahrer ein Sicherheitsfahrer ist, dann ist diese Wahrscheinlichkeit 0,06, das heißt  $P(A|B_1) = 0,06$ . Wenn der Autofahrer ein Risikofahrer ist, dann ist diese Wahrscheinlichkeit 0,6, das heißt  $P(A|B_2) = 0,6$ . Jetzt können wir in die Formel für die totale Wahrscheinlichkeit einsetzen und erhalten

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0,06 \cdot \frac{5}{6} + 0,6 \cdot \frac{1}{6} = 0,15.$$

Die Wahrscheinlichkeit für einen Unfall im nächsten Jahr beträgt also 0,15.

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

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

Die Formel von Bayes besagt  $P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)}$ . Aus dem letzten Beispiel wissen wir  $P(A|B_2)P(B_2) = 0,6 \cdot \frac{1}{6} = 0,1$  und  $P(A) = 0,15$ . Wir erhalten  $P(B_2|A) = \frac{0,1}{0,15} = \frac{2}{3}$ . Nach diesem Unfall ist der Autofahrer verdächtig, ein Risikofahrer zu sein.

**Beispiel 35:** Eine Firma stellt 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?

Sei  $B_1$  das Ereignis, dass ein 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. Wir müssen zuerst  $P(A)$  mit Hilfe der Formel für die totale Wahrscheinlichkeit berechnen:  $P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = 0,95 \cdot \frac{7}{10} + 0,03 \cdot \frac{3}{10} = 0,674$ . Jetzt ergibt sich  $P(B_1|A) = \frac{0,95 \cdot 0,7}{0,674} = 0,987$  aus der Formel von Bayes.

Die Wahrscheinlichkeit, dass ein Computerchip, der in den Verkauf geht, auch intakt ist, beträgt 0,987.

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

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

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



### III. Zufallsvariable

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

#### 1. Darstellung von Ereignissen durch Zufallsvariable

Um das Ereignis „höchstens  $k$  defekte Glühlampen in der Stichprobe“ einfach aufschreiben zu können, bezeichnet man die Anzahl der defekten Glühlampen 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 beschreiben immer Ausfälle eines Zufallsexperiments, das heißt für jede reelle Zahl  $t$  definiert  $X \leq t$  ein Ereignis. Für Zufallsvariable werden üblicherweise Großbuchstaben  $X, Y, Z, \dots$  verwendet.

Mit Zufallsvariablen kann man auch rechnen. Es gilt zum Beispiel

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

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

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

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

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

beziehungsweise  $P(X \in (2, 3]) = P(X \leq 3) - P(X \leq 2)$ .

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

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

Im Folgenden werden wir solche Umformungen häufig verwenden.

#### 2. Wahrscheinlichkeitsvektoren und Wahrscheinlichkeitsdichten

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

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

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

Sei  $S = \{1, 2, 3, 4, 5, 6\}$ . Es gilt  $w(i) = P(X = i) = \frac{1}{6}$  für alle  $i \in S$ , da ja jede Augenzahl mit Wahrscheinlichkeit  $\frac{1}{6}$  auftritt. Man prüft  $\sum_{i=1}^6 w(i) = 1$  nach.

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

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

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

Ist  $S$  eine endliche Menge, dann sind wir schon fertig. Sei  $S$  also abzählbar. Sei  $\varepsilon > 0$ . Wegen  $\sum_{i \in S} w(i) = 1$  existiert eine endliche Teilmenge  $U$  von  $S$  mit  $\sum_{i \in U} w(i) > 1 - \varepsilon$ , da ja die Partialsummen gegen 1 konvergieren. Es folgt  $\sum_{i \in U'} w(i) = 1 - \sum_{i \in U} w(i) < \varepsilon$ . Da  $U$  endlich ist, erhalten wir  $P(X \in U) = \sum_{i \in U} w(i) > 1 - \varepsilon$  aus dem ersten Teil des Beweises und damit  $P(X \in U') = 1 - P(X \in U) < \varepsilon$ .

Ist  $B$  jetzt eine beliebige Teilmenge von  $S$ , dann sei  $C = B \cap U$  und  $D = B \setminus U$ . Da  $B$  die disjunkte Vereinigung von  $C$  und  $D$  ist, haben wir  $\sum_{i \in B} w(i) = \sum_{i \in C} w(i) + \sum_{i \in D} w(i)$  und  $P(X \in B) = P(X \in C) + P(X \in D)$  ergibt sich aus dem Additionssatz. Aus dem ersten Teil des Beweises folgt  $P(X \in C) = \sum_{i \in C} w(i)$ , da  $C$  endlich ist. Wegen  $D \subseteq U'$  erhalten wir  $P(X \in D) \leq P(X \in U') < \varepsilon$  und  $\sum_{i \in D} w(i) \leq \sum_{i \in U'} w(i) < \varepsilon$ . Es folgt  $|P(X \in B) - \sum_{i \in B} w(i)| = |P(X \in C) + P(X \in D) - \sum_{i \in C} w(i) - \sum_{i \in D} w(i)| = |P(X \in D) - \sum_{i \in D} w(i)| < 2\varepsilon$ . Da  $\varepsilon$  beliebig war, ist  $P(X \in B) = \sum_{i \in B} w(i)$  gezeigt.  $\square$

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

Wir definieren den Erwartungswert einer diskreten Zufallsvariablen  $X$ . Der Erwartungswert ist ein Durchschnittswert aller möglichen Werte von  $X$ , wobei jedoch jeder Wert  $i$  mit seiner Wahrscheinlichkeit  $w(i)$  gewichtet wird. Jeder Wert  $i$  geht in den Erwartungswert gemäß der Häufigkeit seines Auftretens ein.

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

Es kann vorkommen, dass  $E(X)$  nicht existiert, nämlich dann, wenn  $S$  unendlich ist und die Reihe  $\sum_{i \in S} i w(i)$  nicht absolut konvergent ist (d.h.  $\sum_{i \in S} |i| w(i)$  ist unendlich). Wie stark eine Zufallsvariable  $X$  um ihren Erwartungswert  $E(X)$  schwankt, wird durch die Varianz gemessen.

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

Auch die Varianz existiert nicht notwendigerweise. Wenn für eine diskrete Zufallsvariable  $X$  mit Wahrscheinlichkeitsvektor  $(w(i))_{i \in S}$  die Reihe  $\sum_{i \in S} i^2 w(i)$  konvergiert, dann existiert aufgrund der Ungleichung  $|i| \leq 1 + i^2$  der Erwartungswert  $m = E(X)$ . Wegen  $(i - m)^2 = i^2 - 2im + m^2$  konvergiert nun auch die Reihe  $\sum_{i \in S} (i - m)^2 w(i)$ , daher besitzt die Zufallsvariable  $X$  eine Varianz. Um Erwartungswert und Varianz zu berechnen, bestimmt man den Wahrscheinlichkeitsvektor und setzt in die Formeln ein.

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

Sei  $X$  der Gewinn eines (zufällig gewählten) Loses. Wir berechnen  $E(X)$ . 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  $E(X) = 1000 \cdot 0,01 + 100 \cdot 0,09 + 0 \cdot 0,9 = 19$ . Jedenfalls sollte man den Lospreis höher als 19 Euro festsetzen.

Wir berechnen die Varianz des Gewinns:  $V(X) = 981^2 \cdot 0,01 + 81^2 \cdot 0,09 + (-19)^2 \cdot 0,9 = 10539$ . Die Standardabweichung ist  $\sqrt{V(X)} = 102,66$ .

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

**Definition:** Eine integrierbare Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  mit  $\int_{-\infty}^{\infty} f(x) dx = 1$  heißt Wahrscheinlichkeitsdichte. Eine Zufallsvariable  $X$  heißt kontinuierlich, wenn eine Wahrscheinlichkeitsdichte  $f$  existiert, sodass  $P(X \leq t) = \int_{-\infty}^t f(x) dx$  für alle  $t \in \mathbb{R}$  gilt.

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 14:** Die Verteilungsfunktion  $F$  einer Zufallsvariablen  $X$  ist eine monoton 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 6. Somit ist  $F$  monoton wachsend. Ebenfalls aus Satz 6 folgt  $0 \leq P(X \leq t) \leq 1$ , sodass  $F(t) \in [0, 1]$  gilt.  $\square$

Wahrscheinlichkeiten lassen sich mit Hilfe der Wahrscheinlichkeitsdichte berechnen.

**Satz 15:** Sei  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  die Wahrscheinlichkeitsdichte einer kontinuierlichen Zufallsvariablen  $X$  und  $F : \mathbb{R} \rightarrow [0, 1]$  ihre Verteilungsfunktion. Dann gilt

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

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

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 8 (c) und  $P(X \in (u - \varepsilon, u]) = \int_{u-\varepsilon}^u f(x) dx$  wurde oben bewiesen. Da  $f$  eine integrierbare Funktion ist, gilt  $\lim_{\varepsilon \rightarrow 0} \int_{u-\varepsilon}^u f(x) dx = 0$ . Daraus ergibt sich  $\lim_{\varepsilon \rightarrow 0} P(X \in (u - \varepsilon, u]) = 0$ . Also muss auch  $P(X = u) = 0$  gelten. Die Wahrscheinlichkeit  $P(X \in B)$  ändert sich also nicht, wenn man einen Endpunkt des Intervalls  $B$  dazugibt oder weglässt (Additionssatz). Sind  $r$  und  $s$  die Endpunkte von  $B$ , dann gilt  $P(X \in B) = P(X \in (r, s])$  und (a) ist gezeigt.

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

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

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

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

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

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

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

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

**Bemerkung:** Im Beispiel 39 ist die Verteilungsfunktion  $F(t)$  im Punkt  $t = 0$  nicht differenzierbar, jedoch definiert die angegebene Wahrscheinlichkeitsdichte die geforderte Verteilungsfunktion. Wie man den Funktionswert von  $f$  im Punkt 0 wählt, hat keinen Einfluss auf die mit Hilfe von  $f$  berechneten Wahrscheinlichkeiten, da diese ja Integrale über  $f$  sind.

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  $E(X)$  einer kontinuierlichen Zufallsvariablen  $X$  mit Wahrscheinlichkeitsdichte  $f$  ist definiert durch  $E(X) = \int_{-\infty}^{\infty} xf(x)dx$ . Die Varianz  $V(X)$  ist definiert durch  $V(X) = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx$  mit  $m = E(X)$ . Wieder nennt man  $\sqrt{V(X)}$  die Standardabweichung der Zufallsvariablen  $X$ .

Auch eine kontinuierliche Zufallsvariable besitzt nicht notwendigerweise einen Erwartungswert beziehungsweise eine Varianz. Die Bedingung ist hier die Konvergenz der entsprechenden uneigentlichen Integrale. Analog dem Fall einer diskreten Zufallsvariable folgt auch hier aus der Konvergenz des uneigentlichen Integrals  $\int_{-\infty}^{\infty} x^2 f(x) dx$  die Existenz der Varianz (und damit auch des Erwartungswertes).

**Beispiel 40:** Die Wahrscheinlichkeitsdichte  $f$  der Zufallsvariablen  $X$  sei gegeben durch  $f(x) = \frac{1}{b-a}$  für  $x \in [a, b]$  und  $f(x) = 0$  für  $x \notin [a, b]$ . Gesucht sind  $E(X)$  und  $V(X)$ .

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

$$E(X) = \int_a^b xf(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{a+b}{2}$$

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

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.

### 3. Binomialverteilung und geometrische Verteilung

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

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

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

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

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

**Beweis:** Sei  $S = \{0, 1, 2, \dots, n\}$ . Wir müssen  $P(X = i)$  für  $i \in S$  berechnen. Gefragt ist nach der Wahrscheinlichkeit, dass unter den  $n$  Würfen genau  $i$  Mal „Wappen“ auftritt. Für  $n = 5$  und  $i = 3$  wurde das bereits in Beispiel 29 durchgeführt. Die dort angewendete

Methode lässt sich auch hier anwenden. Wir zerlegen in Teilereignisse. Die Teilereignisse entsprechen gerade den Folgen der Länge  $n$ , die  $i$  Mal W und  $n - i$  Mal Z enthalten. Nach Satz 4 ist die Anzahl dieser Teilereignisse gleich  $\binom{n}{i}$ .

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

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

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

Es liegt ein Bernoulliexperiment vor, wobei wir Augenzahl 6 als Erfolg auffassen und alle anderen Augenzahlen als Misserfolg. Die Wahrscheinlichkeit für Erfolg ist  $\frac{1}{6}$ . Sei  $X$  die Anzahl, mit der Erfolg, das heißt die Augenzahl 6, unter den 10 Würfeln auftritt. Nach Satz 16 hat  $X$  die  $B(n, p)$ -Verteilung mit  $n = 10$  und  $p = \frac{1}{6}$ , das heißt der Wahrscheinlichkeitsvektor der Zufallsvariable  $X$  ist durch  $w(i) = \binom{10}{i}(\frac{1}{6})^i(\frac{5}{6})^{10-i}$  mit  $0 \leq i \leq 10$  gegeben. Gesucht ist die Wahrscheinlichkeit, dass das Ereignis  $X \leq 3$  eintritt. Aus Satz 13 folgt  $P(X \leq 3) = w(0) + w(1) + w(2) + w(3)$ . 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 42:** In einer Pension gibt es 16 Einzelzimmer. Man weiß, dass im Durchschnitt 20% der bestellten Zimmer nicht belegt werden. Der Pensionsinhaber nimmt für die Ferienwoche 18 Buchungen entgegen. Mit welcher Wahrscheinlichkeit wird er Schwierigkeiten haben?

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

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

Die Produktion einer Schraube ist ein Bernoulliexperiment. Die Wahrscheinlichkeit für einen Erfolg (brauchbare Schraube) ist  $p = 0,9$ . Sei  $X$  die Anzahl der brauchbaren Schrauben in der Packung von 100 Stück. Nach Satz 16 hat  $X$  die  $B(n, p)$ -Verteilung mit  $n = 100$  und  $p = 0,9$ . Gesucht ist  $P(X \leq 92)$ . Wir gehen zur Gegenwahrscheinlichkeit

über:  $P(X \geq 93) = \sum_{i=93}^{100} w(i) = \sum_{i=93}^{100} \binom{100}{i} 0,9^i 0,1^{100-i}$ . Die gesuchte Wahrscheinlichkeit ist  $1 - P(X \geq 93)$ . Wir werden später sehen, wie man diese Wahrscheinlichkeit durch Approximation mit der Normalverteilung näherungsweise berechnen kann.

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

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

Für eine  $B(n, p)$ -verteilte Zufallsvariable  $X$  gilt also  $E(X) = np$ . Der Erwartungswert gibt den durchschnittlichen Wert einer Zufallsvariable an. Der Pensionsinhaber aus Beispiel 42 kann also erwarten, dass durchschnittlich  $18 \cdot 0,8 = 14,4$  der 18 Personen, die gebucht haben, kommen. In einer Packung Schrauben aus Beispiel 43 wird man durchschnittlich  $100 \cdot 0,9 = 90$  brauchbare Schrauben finden.

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

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

Aus dem binomischen Lehrsatz folgt, dass  $\sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} = 1$  gilt. Weiters wurde die Summe  $\sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} = np$  oben bei der Berechnung des Erwartungswertes bestimmt. In der Summe  $\sum_{i=0}^n i(i-1) \binom{n}{i} p^i (1-p)^{n-i}$  sind die Summanden für  $i = 0, 1$  gleich null und für  $i \geq 2$  gilt  $i(i-1) \binom{n}{i} = n(n-1) \binom{n-2}{i-2}$ . Es folgt

$$\begin{aligned}
 \sum_{i=0}^n i(i-1) \binom{n}{i} p^i (1-p)^{n-i} &= n(n-1) \sum_{i=2}^n \binom{n-2}{i-2} p^i (1-p)^{n-i} = \\
 &= n(n-1) p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} = n(n-1) p^2
 \end{aligned}$$

Wir erhalten somit für eine  $B(n, p)$ -verteilte Zufallsvariable  $X$ :

$$V(X) = n(n-1)p^2 + (1-2np)np + n^2p^2 = np(1-p)$$

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

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

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

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

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

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

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

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

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

Sei  $X$  die Anzahl der Würfe, bis zum ersten Mal 6 auftritt. Nach Satz 17 hat  $X$  die  $G(p)$ -Verteilung mit  $p = \frac{1}{6}$ . Der Wahrscheinlichkeitsvektor von  $X$  ist  $w(i) = (\frac{5}{6})^{i-1} \frac{1}{6}$  mit  $i \geq 1$ . Barbara gewinnt, wenn  $X$  einen Wert aus der Menge  $M = \{2, 3, 6, 7, 10, 11, \dots\}$  annimmt. Es gilt  $P(X \in M) = w(2) + w(3) + w(6) + w(7) + w(10) + w(11) + \dots = (\frac{5}{6} \frac{1}{6} + (\frac{5}{6})^2 \frac{1}{6})(1 + (\frac{5}{6})^4 + (\frac{5}{6})^8 + \dots) = \frac{55}{6^3} \frac{1}{1 - (\frac{5}{6})^4} = \frac{330}{6^4 - 5^4} = \frac{330}{671}$ . Barbara gewinnt mit Wahrscheinlichkeit  $\frac{330}{671} = 0,4918$ .

Den Erwartungswert einer  $G(p)$ -verteilten Zufallsvariablen  $X$  berechnen wir mit folgendem Trick. Wir gehen von der geometrischen Reihe  $\sum_{i=1}^{\infty} x^i = \frac{1}{1-x} - 1$  aus und differenzieren diese Potenzreihe innerhalb des Konvergenzintervalls  $(-1, 1)$  gliedweise. Es folgt  $\sum_{i=1}^{\infty} i x^{i-1} = \frac{1}{(1-x)^2}$  und daraus ergibt sich

$$E(X) = \sum_{i=1}^{\infty} i w(i) = \sum_{i=1}^{\infty} i (1-p)^{i-1} p = \frac{p}{(1 - (1-p))^2} = \frac{p}{p^2} = \frac{1}{p}$$



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

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

#### 4. Die hypergeometrische Verteilung

Aus einer Menge von  $N$  Kugeln, von denen  $M$  weiß und  $N - M$  schwarz sind, wird eine Stichprobe vom Umfang  $n$  gezogen. Sei  $X$  die Anzahl der weißen Kugeln in der Stichprobe. Zieht man geordnet mit Zurücklegen, dann hat  $X$  die  $B(n, p)$ -Verteilung mit  $p = \frac{M}{N}$ , da sich der Anteil  $p$  der weißen Kugeln nicht ändert. Es liegt ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit  $p$  vor. Zieht man jedoch ungeordnet ohne Zurücklegen, dann gilt  $P(X = i) = \binom{M}{i} \binom{N-M}{n-i} / \binom{N}{n}$ , da  $\binom{N}{n}$  die Anzahl aller  $n$ -elementigen Teilmengen ist und  $\binom{M}{i} \binom{N-M}{n-i}$  die Anzahl der Teilmengen, die  $i$  weiße und  $n - i$  schwarze Kugeln enthalten. Es sei daran erinnert, dass  $\binom{u}{v} = 0$  gilt für  $v < 0$  und  $v > u$ .

Wir führen dazu eine Verteilung ein, diese Verteilung nähert sich bei kleinem Stichprobenumfang  $n$  (im Vergleich zu  $N$ ) der Binomialverteilung  $B(n, p)$  mit  $p = M/N$  an.

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

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

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

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

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

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

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

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

In der Summe  $\sum_{i=0}^n i(i-1) \binom{M}{i} \binom{N-M}{n-i}$  sind wieder die Summanden für  $i = 0, 1$  gleich null und für  $i \geq 2$  gilt  $i(i-1) \binom{M}{i} = M(M-1) \binom{M-2}{i-2}$ . Für  $n \geq 2$  gilt  $\sum_{i=0}^n i(i-1) \binom{M}{i} \binom{N-M}{n-i} =$

$M(M-1) \sum_{i=2}^n \binom{M-2}{i-2} \binom{N-M}{n-i} = M(M-1) \sum_{j=0}^{n-2} \binom{M-2}{j} \binom{N-2-(M-2)}{n-2-j} = M(M-1) \binom{N-2}{n-2}$ .  
 Weiters hatten wir  $\sum_{i=0}^n i \binom{M}{i} \binom{N-M}{n-i} = M \binom{N-1}{n-1}$  und  $\sum_{i=0}^n \binom{M}{i} \binom{N-M}{n-i} = \binom{N}{n}$ . Damit ergibt sich

$$\binom{N}{n} V(X) = M(M-1) \binom{N-2}{n-2} + (1 - 2n \frac{M}{N}) M \binom{N-1}{n-1} + (n \frac{M}{N})^2 \binom{N}{n}$$

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

$$\begin{aligned} V(X) &= M(M-1) \frac{n(n-1)}{N(N-1)} + (1 - 2n \frac{M}{N}) \frac{nM}{N} + (n \frac{M}{N})^2 = (\frac{(M-1)(n-1)}{N-1} + \frac{N-nM}{N}) n \frac{M}{N} = \\ &= \frac{MNn - MN - Nn + N + N^2 - MNn - N + Mn}{N(N-1)} n \frac{M}{N} = n \frac{M}{N} \frac{N-M}{N} \frac{N-n}{N-1} \end{aligned}$$

Das ist die Varianz einer  $H(N, M, n)$ -verteilten Zufallsvariable  $X$ . Diese Formel gilt auch für  $n = 1$ , auch wenn in diesem Fall obige Rechnung nicht funktioniert. Es folgt jedoch  $\binom{N}{1} V(X) = \sum_{i=0}^1 (i - \frac{M}{N})^2 \binom{M}{i} \binom{N-M}{1-i} = (\frac{M}{N})^2 (N-M) + (1 - \frac{M}{N})^2 M = \frac{M^2}{N} + M - 2 \frac{M^2}{N} = \frac{MN - M^2}{N}$ , daher gilt  $V(X) = \frac{M}{N} \frac{N-M}{N}$  für eine  $H(N, M, 1)$ -verteilte Zufallsvariable  $X$ .

## 5. Poissonverteilung

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

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

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

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

Wir entwickeln zur Poissonverteilung folgendes Modell: Bei der Telefonauskunft treffen zu zufälligen Zeitpunkten Telefonanrufe ein. Wir legen einen Zeitpunkt als Nullpunkt fest und bezeichnen die Anzahl der Anrufe im Zeitintervall  $[0, t]$  mit  $X$ .

Um den Wahrscheinlichkeitsvektor von  $X$  zu berechnen, müssen wir einige Annahmen zugrundelegen. Sei  $z > t$  und  $n$  die Anzahl der Personen, die im Zeitintervall  $[0, z]$  anrufen. Wir nehmen an, dass jede Person unabhängig von den anderen zufällig einen Zeitpunkt in  $[0, z]$  auswählt, zu dem sie anruft. Die Wahrscheinlichkeit, dass ein bestimmter Anrufer im Zeitintervall  $[0, t]$  anruft, ist dann  $\frac{t}{z}$  und die Wahrscheinlichkeit, dass er im Zeitintervall  $(t, z]$  anruft, ist  $\frac{z-t}{z} = 1 - \frac{t}{z}$ . Es liegt für jeden der  $n$  Anrufer ein Bernoulliexperiment vor. Ein Erfolg (eine bestimmte Person ruft im Zeitintervall  $[0, t]$  an) tritt mit Wahrscheinlichkeit  $\frac{t}{z}$  auf. Jeder der  $n$  Anrufe entspricht einer unabhängigen Wiederholung des Bernoulliexperiments. Da  $X$  die Anzahl der Anrufe im Zeitintervall  $[0, t]$  ist, also die Anzahl der Erfolge, hat  $X$  die  $B(n, \frac{t}{z})$ -Verteilung. Wir halten  $\mu = \frac{n}{z}$ , die durchschnittliche Anzahl der Anrufe pro Zeiteinheit, fest und lassen  $n$  und damit auch  $z$  gegen  $\infty$  gehen.

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

**Beweis:** Wir berechnen  $P(X = i)$ . Wir setzen  $\frac{1}{z} = \frac{\mu}{n}$  ein und formen um

$$\begin{aligned} P(X = i) &= \binom{n}{i} \left(\frac{t}{z}\right)^i \left(1 - \frac{t}{z}\right)^{n-i} = \binom{n}{i} \left(\frac{\mu t}{n}\right)^i \left(1 - \frac{\mu t}{n}\right)^{n-i} \\ &= \frac{(\mu t)^i}{i!} \frac{n(n-1)\dots(n-i+1)}{n^i} \left(1 - \frac{\mu t}{n}\right)^n \left(1 - \frac{\mu t}{n}\right)^{-i} \end{aligned}$$

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

Wir berechnen noch Erwartungswert und Varianz einer  $P(\lambda)$ -verteilten Zufallsvariable  $X$ . Es gilt  $E(X) = \sum_{i=0}^{\infty} i w(i) = e^{-\lambda} \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} = \lambda$ . Für die Varianz erhalten wir

$$V(X) = \sum_{i=0}^{\infty} (i-\lambda)^2 w(i) = e^{-\lambda} \sum_{i=0}^{\infty} (i(i-1) + (1-2\lambda)i + \lambda^2) \frac{\lambda^i}{i!} = \lambda^2 + \lambda(1-2\lambda) + \lambda^2 = \lambda,$$

denn es gilt  $\sum_{i=0}^{\infty} i(i-1) \frac{\lambda^i}{i!} = \sum_{i=2}^{\infty} i(i-1) \frac{\lambda^i}{i!} = \sum_{j=0}^{\infty} \frac{\lambda^{j+2}}{j!} = \lambda^2 e^\lambda$ .

**Beispiel 47:** Die Anzahl der Anrufe in einem Zeitintervall von  $t$  Stunden habe die  $P(\mu t)$ -Verteilung mit  $\mu = 12$  Anrufen pro Stunde. Wie groß ist die Wahrscheinlichkeit, dass zwischen 16,00 und 16,15 höchstens ein Anruf kommt? Wie groß ist die Wahrscheinlichkeit, dass zwischen 16,00 und 16,15 mindestens drei Anrufe kommen?

Sei  $X$  die Anzahl der Anrufe zwischen 16,00 und 16,15, das heißt  $X$  ist  $P(\lambda)$ -verteilt mit  $\lambda = 12 \cdot \frac{1}{4} = 3$ . Daher gilt  $E(X) = 3$  und

$$P(X \leq 1) = w(0) + w(1) = \frac{3^0}{0!} e^{-3} + \frac{3^1}{1!} e^{-3} = 4e^{-3} = 0,199$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - w(0) - w(1) - w(2) = 1 - (\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}) e^{-3} = 0,5768$$

Mit Wahrscheinlichkeit 0,199 kommt höchstens ein Anruf. Mit Wahrscheinlichkeit 0,577 kommen mindestens drei Anrufe.

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

Eine weitere Anwendung der Poissonverteilung besteht in der zufälligen Verteilung von Punkten in einem euklidischen Raum beliebiger Dimension. Dabei wird eine durchschnittliche Dichte  $\mu$  der Punkte pro Flächeneinheit festgelegt. Nun sei eine beliebige endliche Fläche  $A$  mit dem Flächeninhalt  $|A|$  vorgegeben, dann ist die Zufallsvariable  $X$  der Anzahl von Punkten in dieser Fläche  $P(\mu |A|)$ -verteilt. Das Modell wird ganz analog mit der geometrischen Wahrscheinlichkeit und der Binomialverteilung entwickelt.

## 6. Exponentialverteilung und Gammaverteilung

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

**Definition:** Sei  $\lambda > 0$ . Eine kontinuierliche Zufallsvariable  $X$  mit Wertebereich  $\mathbb{R}_0^+$  und Wahrscheinlichkeitsdichte

$$f(x) = \lambda e^{-\lambda x} \quad \text{für } x \in \mathbb{R}_0^+$$

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

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

**Beweis:** Sei  $t > 0$  beliebig und  $X$  die Anzahl der Anrufe im Zeitintervall  $[0, t]$ . Die Wartezeit auf den ersten Anruf ist genau dann  $\leq t$ , wenn im Zeitintervall  $[0, t]$  mindestens ein Anruf kommt. Das heißt  $Y \leq t$  tritt genau dann ein, wenn  $X \geq 1$  eintritt. Daher gilt  $F(t) = P(Y \leq t) = P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda t}$ . Die Dichte  $f$  von  $Y$  erhält man als Ableitung von  $F$ . Es folgt  $f(x) = (1 - e^{-\lambda x})' = \lambda e^{-\lambda x}$  für  $x \in \mathbb{R}_0^+$ .

Die Wartezeit  $Y$  kann nicht negativ sein. Für  $t \leq 0$  gilt daher  $F(t) = P(Y \leq t) = 0$  und  $f(x) = 0$  für  $x \leq 0$ . Der Wertebereich der Zufallsvariablen  $Y$  ist gleich  $\mathbb{R}_0^+$ . Die Wartezeit  $Y$  auf den ersten Anruf ist also  $E(\lambda)$ -verteilt.  $\square$

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

**Beispiel 48:** Die Lebensdauer eines elektronischen Bauteiles sei  $E(\lambda)$ -verteilt. Man weiß aus Erfahrung, dass die Lebensdauer mit Wahrscheinlichkeit 0,9 mindestens 500 Stunden beträgt. Wie groß ist  $\lambda$ ?

Die Lebensdauer des elektronischen Bauteiles bezeichnen wir mit  $Y$ . Die Wahrscheinlichkeit, dass  $Y$  größer oder gleich 500 ist, ist 0,9, das heißt  $P(Y \geq 500) = 0,9$ . Wegen Satz 15 (c) gilt  $P(Y \geq 500) = \int_{500}^{\infty} \lambda e^{-\lambda x} dx = e^{-500\lambda}$ , da  $Y$  ja  $E(\lambda)$ -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

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = -\frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} = \frac{1}{\lambda}.$$

Ebenso erhält man durch zweimalige partielle Integration

$$V(X) = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = -(x - \frac{1}{\lambda})^2 e^{-\lambda x} \Big|_0^{\infty} - 2(x - \frac{1}{\lambda}) \frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} + \int_0^{\infty} 2 \frac{e^{-\lambda x}}{\lambda} dx = \frac{1}{\lambda^2}.$$

Wir haben also  $E(X) = \frac{1}{\lambda}$  und  $V(X) = \frac{1}{\lambda^2}$  erhalten. Die durchschnittliche Lebensdauer des elektronischen Bauteils aus Beispiel 48 beträgt  $\frac{500}{-\ln 0,9} = 4745,6$  Stunden.

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

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

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

heißt gammaverteilt (Erlangverteilt) mit Parametern  $r$  und  $\lambda$ , oder kurz  $E(r, \lambda)$ -verteilt, wobei  $\Gamma(r)$  definiert ist durch  $\Gamma(r) = \int_0^{\infty} y^{r-1} e^{-y} dy$ . Es gilt  $f(x) > 0$  für  $x \in \mathbb{R}^+$  und  $\int_0^{\infty} f(x) dx = 1$  folgt aus der Definition von  $\Gamma(r)$ . Für  $x \in \mathbb{R}_0^-$  setzt man  $f(x) = 0$ .

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

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

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

**Beweis:** Sei  $t > 0$  beliebig und  $X$  die Anzahl der Anrufe im Zeitintervall  $[0, t]$ . Die Wartezeit auf den  $n$ -ten Anruf ist genau dann  $\leq t$ , wenn im Zeitintervall  $[0, t]$  mindestens  $n$  Anrufe kommen, das heißt  $Z \leq t$  tritt genau dann ein, wenn  $X \geq n$  eintritt. Daher gilt

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

Die Wahrscheinlichkeitsdichte  $f$  der Zufallsvariable  $Z$  erhält man als Ableitung von  $F$ . Berechnet man diese mit Hilfe der Produktregel, so kürzen sich alle dabei auftretenden Summanden weg bis auf einen:

$$F'(t) = \lambda e^{-\lambda t} - \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t} - \dots - (n-1) \lambda \frac{(\lambda t)^{n-2}}{(n-1)!} e^{-\lambda t} + \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}$$

Man erhält  $f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$  für  $x \geq 0$ . Da  $Z$  außerdem als Wartezeit keine negativen Werte annehmen kann, ist  $\mathbb{R}_0^+$  der Wertebereich von  $Z$  und  $f(x) = 0$  für  $x < 0$ .  $\square$

### 7. Normalverteilung

Misst man eine (physikalische) Größe mehrmals, so erhält man Messwerte, die wegen zufälliger Störungen beim Messvorgang leicht voneinander abweichen. Werden Flaschen oder Verpackungen von einer Abfüllanlage gefüllt, so wird man Schwankungen bei den Füllmengen feststellen, die auf die Ungenauigkeit der Abfüllanlage zurückzuführen sind. Ebenso sind die Abmessungen eines Werkstücks nach einem Produktionsprozess Schwankungen unterworfen. Messwerte und Füllmengen unterliegen zufälligen Einflüssen und werden daher als Zufallsvariable aufgefasst. Für solche Zufallsvariablen verwendet man die Normalverteilung.

**Definition:** Seien  $\mu \in \mathbb{R}$  und  $\sigma > 0$ . Eine kontinuierliche Zufallsvariable  $X$  mit Wertebereich  $\mathbb{R}$  und Wahrscheinlichkeitsdichte

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

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

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

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

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

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

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

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

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

**Satz 23:** *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)  $\Phi : \mathbb{R} \rightarrow (0, 1)$  ist streng monoton wachsend und invertierbar.

**Beweis:** Es gilt (a), da in der Formel für  $\varphi(x)$  nur  $x^2$  vorkommt, und (b) rechnet man nach  $\Phi(-t) = \int_{-\infty}^{-t} \varphi(x) dx = \int_{-\infty}^{-t} \varphi(-x) dx = -\int_{\infty}^t \varphi(y) dy = \int_t^{\infty} \varphi(y) dy = 1 - \Phi(t)$ , wobei zuerst (a) verwendet und dann die Integralsubstitution  $y = -x$  durchgeführt wurde. Da  $\Phi'(t) = \varphi(t) > 0$  für alle  $t$  gilt, ist  $\Phi$  eine streng monoton wachsende Funktion und daher auch injektiv. Weiters berechnet man  $\lim_{t \rightarrow -\infty} \Phi(t) = \lim_{t \rightarrow -\infty} \int_{-\infty}^t \varphi(x) dx = 0$  und  $\lim_{t \rightarrow \infty} \Phi(t) = \lim_{t \rightarrow \infty} \int_{-\infty}^t \varphi(x) dx = 1$ . Daher ist  $\Phi : \mathbb{R} \rightarrow (0, 1)$  auch surjektiv. Damit ist gezeigt, dass  $\Phi$  invertierbar ist, und (c) ist bewiesen.  $\square$

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

**Beispiel 49:** Eine Maschine stellt Spanplatten her. Ihre Dicke ist  $N(\mu, \sigma)$ -verteilt mit  $\mu = 19$  mm und  $\sigma = 0,05$  mm. Die Platten sollen zwischen 18,95 mm und 19,10 mm stark sein. Wieviel Prozent Ausschuss produziert die Maschine?

Sei  $X$  die Dicke einer (zufällig gewählten) Platte. Dann hat  $X$  die  $N(\mu, \sigma)$ -Verteilung. Die Platte ist Ausschuss, wenn  $X \leq 18,95$  oder  $X \geq 19,10$  gilt. Wir berechnen die Wahrscheinlichkeit, dass sie kein Ausschuss ist, das heißt, dass  $18,95 < X < 19,10$  gilt.

$$\begin{aligned} P(18,95 < X < 19,10) &= P\left(\frac{18,95-\mu}{\sigma} < Y < \frac{19,10-\mu}{\sigma}\right) \quad \text{wobei } Y = \frac{X-\mu}{\sigma} \\ &= P(-1 < Y < 2) \\ &= \Phi(2) - \Phi(-1) && \text{da } Y \text{ die } N(0, 1)\text{-Verteilung hat} \\ &= \Phi(2) - 1 + \Phi(1) && \text{Satz 23 (b)} \\ &= 0,977 - 1 + 0,841 = 0,818 && \text{Tabelle} \end{aligned}$$

Die Wahrscheinlichkeit, dass eine (zufällig gewählte) Platte Ausschuss ist, beträgt daher  $1 - 0,818 = 0,182$ . Die Maschine produziert 18,2% Ausschuss.

**Beispiel 50:** Auf einer Maschine werden Waschmittelpackungen abgefüllt. Der Inhalt der Packungen ist  $N(\mu, \sigma)$ -verteilt. Die Arbeitsgenauigkeit der Abfüllanlage ist bekannt und durch  $\sigma = 0,05$  kg gegeben. Die durchschnittliche Abfüllmenge  $\mu$  kann man einstellen.

- (a) Die Maschine wird auf  $\mu = 3$  kg eingestellt. Wie groß ist die Wahrscheinlichkeit, dass die am Etikett angegebene Mindestfüllmenge von 2,9 kg unterschritten wird?
- (b) Es wird wieder  $\mu = 3$  kg eingestellt. Welche Mindestfüllmenge  $t$  ist auf das Etikett zu

drucken, wenn höchstens 1% der Packungen diese Menge  $t$  unterschreiten dürfen?

(c) Auf welchen Mittelwert  $\mu$  muss man die Maschine einstellen, wenn die Mindestfüllmenge von 2,9 kg nur von 1% der Packungen unterschritten werden darf?

(a) Sei  $X$  der Inhalt einer (zufällig gewählten) Packung. Gesucht ist  $P(X < 2,9)$ .

$$\begin{aligned} P(X < 2,9) &= P\left(Y < \frac{2,9-3}{0,05}\right) && \text{wobei } Y = \frac{X-\mu}{\sigma} = \frac{X-3}{0,05} \\ &= P(Y < -2) \\ &= \Phi(-2) && \text{da } Y \text{ die } N(0,1)\text{-Verteilung hat} \\ &= 1 - \Phi(2) && \text{Satz 23 (b)} \\ &= 1 - 0,977 = 0,023 && \text{Tabelle} \end{aligned}$$

Die Mindestfüllmenge wird von 2,3% der Packungen unterschritten.

(b) Gesucht ist  $t$ , sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt  $X$  kleiner als  $t$  ist, höchstens 0,01 beträgt. Es muss also  $P(X < t) \leq 0,01$  gelten. Wir erhalten

$$P(X < t) = P\left(\frac{X-\mu}{\sigma} < \frac{t-3}{0,05}\right) = \Phi\left(\frac{t-3}{0,05}\right)$$

Wir suchen  $t$ , sodass  $\Phi\left(\frac{t-3}{0,05}\right) \leq 0,01$  gilt. Da  $\Phi$  streng monoton wachsend ist, ist das äquivalent zu  $\frac{t-3}{0,05} \leq \Phi^{-1}(0,01)$ . In der Tabelle findet man  $\Phi(2,33) = 0,99$ . Wegen Satz 23 (b) folgt  $\Phi(-2,33) = 0,01$ , das heißt  $\Phi^{-1}(0,01) = -2,33$ . Wir suchen also  $t$ , sodass  $\frac{t-3}{0,05} \leq -2,33$  gilt. Es folgt  $t \leq 2,884$ . Drückt man eine Mindestfüllmenge, die  $\leq 2,884$  ist, auf das Etikett, dann ist die verlangte Forderung erfüllt.

(c) Gesucht ist  $\mu$ , sodass die Wahrscheinlichkeit dafür, dass der Packungsinhalt  $X$  kleiner als 2,9 ist, höchstens 0,01 beträgt. Es muss also  $P(X < 2,9) \leq 0,01$  gelten. Wir erhalten

$$P(X < 2,9) = P\left(\frac{X-\mu}{\sigma} < \frac{2,9-\mu}{0,05}\right) = \Phi\left(\frac{2,9-\mu}{0,05}\right)$$

Wir suchen  $\mu$ , sodass  $\Phi\left(\frac{2,9-\mu}{0,05}\right) \leq 0,01$  gilt. Da  $\Phi$  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  $\mu$ , sodass  $\frac{2,9-\mu}{0,05} \leq -2,33$  gilt. Es folgt  $\mu \geq 3,016$ . Stellt man die Maschine auf einen Mittelwert  $\mu$  ein, der  $\geq 3,016$  ist, dann ist die verlangte Forderung erfüllt.

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

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

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

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

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

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

## 8. Approximation der Binomialverteilung

Die Berechnung des Wahrscheinlichkeitsvektors einer  $B(n, p)$ -verteilten Zufallsvariable  $X$  ist für große Werte von  $n$  aufwendig. Noch komplizierter ist es, bei einer gegebenen Wahrscheinlichkeit  $P(X \leq k)$  für eine feste Zahl  $k \geq 0$  und einem gegebenen Parameter  $p$  den Parameter  $n$  zu bestimmen. 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  $g : \mathbb{R} \rightarrow \mathbb{R}_0^+$  definiert durch

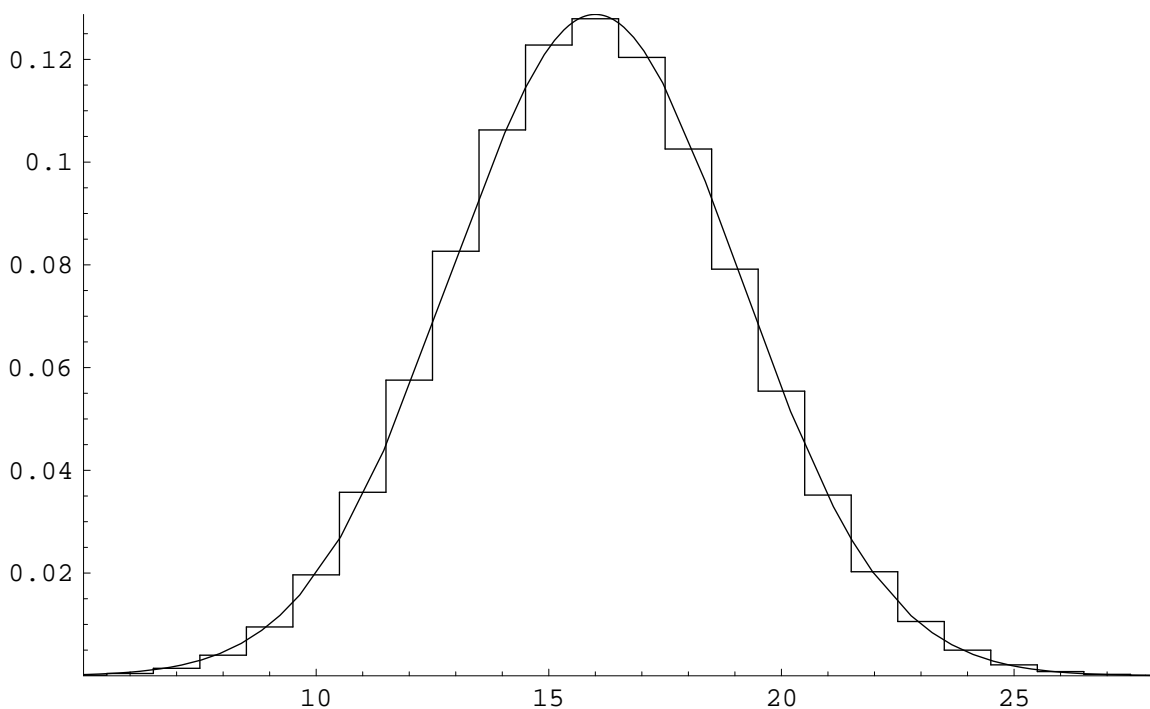
$$g(x) = w(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad \text{für } x \in (i - \frac{1}{2}, i + \frac{1}{2}] \quad \text{und für } 0 \leq i \leq n$$

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

Vergleicht man die Treppenfunktion  $g$  mit der Wahrscheinlichkeitsdichte

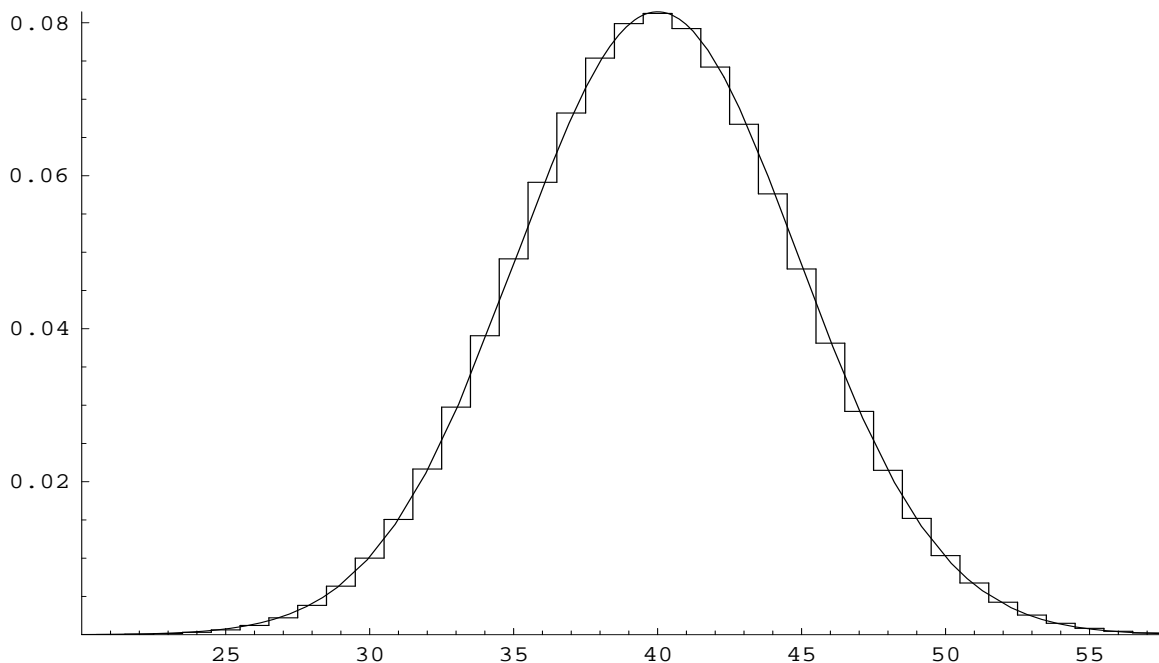
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{für } x \in \mathbb{R}$$

so sieht man, dass die beiden Funktionen, abgesehen von den Sprungstellen, eine ähnliche Form haben. Durch die Wahl von  $\mu$  kann man die Funktion  $f$  nach links oder rechts verschieben, und durch die Wahl von  $\sigma$  kann man  $f$  breiter und niedriger oder schmaler und höher machen. (Die Fläche unter  $f$  ist immer 1.) Wir wählen  $\mu$  und  $\sigma$  so, dass Erwartungswert und Standardabweichung der Binomialverteilung und der Normalverteilung übereinstimmen, also  $\mu = np$  und  $\sigma = \sqrt{np(1-p)}$ . Dadurch wird die Wahrscheinlichkeitsdichte  $f$  recht gut an die Treppenfunktion  $g$  angepasst. Die folgende Abbildung zeigt  $g$  mit  $n = 40$  und  $p = 0,4$  und die dazupassende Wahrscheinlichkeitsdichte der Normalverteilung.





Die Approximation ist umso besser, je größer  $n$  ist. Wenn man  $n$  vergrößert, werden die Sprünge der Treppenfunktion kleiner. Die folgende Abbildung zeigt  $g$  mit  $n = 100$  und  $p = 0,4$  und die dazupassende Wahrscheinlichkeitsdichte der Normalverteilung.



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

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

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

wobei die neue Integrationsvariable  $y = \frac{x-\mu}{\sigma}$  eingeführt wurde. Das ergibt eine gute Approximation der Binomialverteilung durch die Normalverteilung.

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 diese Approximation verwenden kann.

**Beispiel 51:** Auf einer Maschine werden Schrauben erzeugt, von denen 1% schadhaft sind. Die Schrauben werden in Packungen zu 1600 Stück verkauft.

(a) Wie groß ist die Wahrscheinlichkeit, dass man höchstens 9 schadhafte Schrauben in einer Packung findet?

(b) Welche Anzahl von brauchbaren Schrauben in einer Packung kann der Hersteller garantieren, wenn die Garantie von höchstens 2% der Packungen unterschritten werden darf?

(a) Sei  $X$  die Anzahl der schadhafte Schrauben in einer (zufällig gewählten) Packung. Dann hat  $X$  die  $B(n, p)$ -Verteilung mit  $n = 1600$  und  $p = 0,01$ . Gesucht ist die Wahr-

scheinlichkeit, dass  $X \leq 9$  ist.

Um das zu berechnen, verwenden wir die Approximation durch die Normalverteilung. Wir haben  $\mu = np = 1600 \cdot 0,01 = 16$  und  $\sigma = \sqrt{np(1-p)} = \sqrt{1600 \cdot 0,01 \cdot 0,99} = 3,980$ . Wegen  $\sigma \geq 3$  ist die Approximation ausreichend gut. Wir erhalten

$$P(X \leq 9) \approx \Phi\left(\frac{9 + \frac{1}{2} - 16}{3,980}\right) = \Phi(-1,633) = 1 - \Phi(1,633) = 1 - 0,949 = 0,051$$

wobei  $\Phi(1,63)$  in einer Tabelle gefunden wurde.

Die Wahrscheinlichkeit, dass höchstens 9 schadhafte Schrauben in einer Packung sind, beträgt 0,051. Der exakte Wert nach der Binomialverteilung beträgt 0,043.

(b) Sei  $Y$  die Anzahl der brauchbaren Schrauben in einer (zufällig gewählten) Packung und  $k$  die Anzahl der brauchbaren Schrauben, die der Hersteller in einer Packung garantieren kann. Gesucht ist  $k$ , sodass  $P(Y < k) \leq 0,02$  gilt. Die Zufallsvariable  $Y$  hat jetzt die  $B(n, p)$ -Verteilung mit  $n = 1600$  und  $p = 0,99$ .

Wir verwenden die Approximation durch die Normalverteilung. Dazu berechnen wir  $\mu = np = 1600 \cdot 0,99 = 1584$  und  $\sigma = \sqrt{np(1-p)} = \sqrt{1600 \cdot 0,99 \cdot 0,01} = 3,98$ . Weil  $\sigma \geq 3$  gilt, ist die Approximation ausreichend gut.

Wegen  $P(Y < k) = P(Y \leq k-1) \approx \Phi\left(\frac{k-1 + \frac{1}{2} - 1584}{3,98}\right) = \Phi\left(\frac{k-1584,5}{3,98}\right)$  ist  $k$  so zu bestimmen, dass  $\Phi\left(\frac{k-1584,5}{3,98}\right) \leq 0,02$  gilt. Aus  $\Phi(2,05) = 0,98$  folgt  $\Phi(-2,05) = 0,02$  und  $-2,05 = \Phi^{-1}(0,02)$ . Da  $\Phi$  streng monoton wachsend ist, erhalten wir die äquivalente Ungleichung  $\frac{k-1584,5}{3,98} \leq -2,05$ . Das ergibt  $k \leq 1576,3$ .

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

**Beispiel 52:** Bei einem Fährunternehmen weiß man, dass durchschnittlich 10% der Personen, die einen Platz für ihren PKW gebucht haben, zur Abfahrt nicht erscheinen. Es gibt 120 PKW-Plätze. Wieviele Buchungen dürfen höchstens vorgenommen werden, wenn die Wahrscheinlichkeit, dass alle, die zur Abfahrt erscheinen, einen Platz bekommen, mindestens 0,95 betragen soll?

Sei  $n$  die Anzahl der vorgenommenen Buchungen und  $X$  die Anzahl der Gebuchten, die auch zur Abfahrt erscheinen. Jeder der  $n$  Gebuchten erscheint mit Wahrscheinlichkeit 0,9. Daher hat  $X$  die  $B(n, p)$ -Verteilung mit  $p = 0,9$ . Wir müssen  $n$  so bestimmen, dass  $P(X \leq 120) \geq 0,95$  gilt.

Wir verwenden die Approximation durch die Normalverteilung. Dazu berechnen wir  $\mu = np = 0,9 \cdot n$  und  $\sigma = \sqrt{np(1-p)} = 0,3\sqrt{n}$ . Da  $n$  jedenfalls  $\geq 120$  sein muss, ist  $\sigma$  jedenfalls  $\geq 3$ , und wir können die Approximation verwenden.

Wegen  $P(X \leq 120) = \Phi\left(\frac{120,5 - 0,9 \cdot n}{0,3\sqrt{n}}\right)$  und  $0,95 = \Phi(1,65)$  ist  $n$  so zu bestimmen, dass  $\Phi\left(\frac{120,5 - 0,9 \cdot n}{0,3\sqrt{n}}\right) \geq \Phi(1,65)$  gilt. Wegen der Monotonie von  $\Phi$  ist das wieder äquivalent zu  $\frac{120,5 - 0,9 \cdot n}{0,3\sqrt{n}} \geq 1,65$ . Durch Umformen ergibt sich

$$n + 0,55\sqrt{n} - 133,9 \leq 0$$

Setzt man  $x = \sqrt{n}$  so erhält man die quadratische Ungleichung  $x^2 + 0,55x - 133,9 \leq 0$ . Die durch die Funktion  $x \mapsto x^2 + 0,55x - 133,9$  dargestellte Parabel ist nach oben offen. Die Ungleichung ist also für alle  $x$  zwischen  $x_1$  und  $x_2$  erfüllt, wobei  $x_1$  und  $x_2$  die beiden Nullstellen der Parabel sind. Löst man die quadratische Gleichung  $x^2 + 0,55x - 133,9 = 0$ , so erhält man  $x_1 = -11,85$  und  $x_2 = 11,3$ . Da  $x = \sqrt{n}$  aber positiv sein muss, erhalten wir als Lösungsmenge für  $x$  das Intervall  $[0, 11,3]$  und als Lösungsmenge für  $n = x^2$  das

Intervall  $[0, 127,7]$ .

Damit alle, die zur Abfahrt erscheinen, mit einer Wahrscheinlichkeit von mindestens 0,95 einen Platz erhalten, dürfen höchstens 127 Buchungen vorgenommen werden.

**Beispiel 53:** Wie oft muss man würfeln, damit mit Wahrscheinlichkeit 0,9 mindestens dreimal 6 unter den gewürfelten Ergebnissen auftritt?

Sei  $n$  die Anzahl der Würfe und  $X$  die Anzahl, mit der 6 auftritt. Dann hat  $X$  die  $B(n, p)$ -Verteilung mit  $p = \frac{1}{6}$ . Wir müssen  $n$  so bestimmen, dass  $P(X \geq 3) \geq 0,9$  gilt.

Wir verwenden die Approximation durch die Normalverteilung. Dazu berechnen wir  $\mu = np = \frac{n}{6}$  und  $\sigma = \sqrt{np(1-p)} = \frac{\sqrt{5n}}{6}$ . Es gilt  $P(X \geq 3) = 1 - P(X \leq 2) \approx 1 - \Phi\left(\frac{2,5-\mu}{\sigma}\right)$ . Wegen  $P(X \geq 3) \geq 0,9$  erhält man  $\Phi\left(\frac{2,5-\mu}{\sigma}\right) \leq 0,1$ . Nun gilt  $\Phi(1,28) = 0,9$  und somit  $\Phi(-1,28) = 0,1$ . Da  $\Phi$  streng monoton wachsend ist, ergibt sich  $\frac{2,5-\mu}{\sigma} \leq -1,28$  oder  $\mu - 2,5 - 1,28\sigma \geq 0$ . Setzt man für  $\mu$  und  $\sigma$  ein, so hat man  $n - 1,28\sqrt{5}\sqrt{n} - 15 \geq 0$ . Die Lösungen der quadratischen Gleichung  $x^2 - 1,28\sqrt{5}x - 15 = 0$  sind  $x_1 = -2,698$  und  $x_2 = 5,560$ . Es folgt  $\sqrt{n} \leq -2,698$  oder  $\sqrt{n} \geq 5,560$ . Da das erste nicht möglich ist, ergibt sich  $\sqrt{n} \geq 5,560$  oder  $n \geq 30,91$ . Man muss mindestens 31 Mal würfeln, um mit Wahrscheinlichkeit 0,9 mindestens dreimal 6 zu erhalten.

Da für  $n = 31$  aber  $\sigma < 3$  gilt, sollte man die Probe mit Hilfe der Binomialverteilung machen:  $P(X \geq 3) = 1 - P(X \leq 2) = 1 - \left(\frac{5}{6}\right)^{31} - \binom{31}{1}\frac{1}{6}\left(\frac{5}{6}\right)^{30} - \binom{31}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{29} = 0,9094$ . Für  $n = 30$  würde man 0,8972 erhalten, also weniger als 0,9.

## IV. Mehrere Zufallsvariable

In diesem Kapitel werden Zusammenhänge zwischen Zufallsvariablen untersucht. Es geht um gemeinsame Verteilungen von Zufallsvariablen, um Summe und Quotient von zwei Zufallsvariablen und um Rechenregeln für Erwartungswert und Varianz.

### 1. Gemeinsame Verteilungen von Zufallsvariablen

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

**Definition:** Seien  $S$  und  $T$  endliche oder abzählbare Mengen. Ein Vektor  $(u(i, j))_{i \in S, j \in T}$  mit nicht negativen Eintragungen und  $\sum_{i \in S, j \in T} u(i, j) = 1$  heißt gemeinsamer Wahrscheinlichkeitsvektor der Zufallsvariablen  $X$  und  $Y$ , wenn  $u(i, j) = P(X = i \wedge Y = j)$  für alle  $i \in S$  und  $j \in T$  gilt.

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

**Beweis:** Genauso wie für Satz 13. Im Beweis von Satz 13 ersetzt man  $S$  durch  $S \times T$ . Man stellt sich den Buchstaben  $i$  fettgedruckt vor und interpretiert ihn als ein Element von  $S \times T$ .  $\square$

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

**Beweis:** Sei  $i \in S$  beliebig. Sei  $B = \{(i, j) : j \in T\}$ . Dann gilt  $B \subseteq S \times T$ . Mit Hilfe von Satz 24 folgt  $P(X = i) = P((X, Y) \in B) = \sum_{(k, j) \in B} u(k, j) = \sum_{j \in T} u(i, j)$ . Damit ist  $P(X = i) = w_1(i)$  für alle  $i \in S$  gezeigt, das heißt  $(w_1(i))_{i \in S}$  ist Wahrscheinlichkeitsvektor der Zufallsvariablen  $X$ . Der Beweis für die Zufallsvariable  $Y$  läuft analog.  $\square$

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

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

Es folgt  $w_1(0) = u(0, 1) + u(0, 2) + u(0, 3) = \frac{1}{2}$  und  $w_1(1) = u(1, 1) + u(1, 2) + u(1, 3) = \frac{1}{2}$ . Das ist der Wahrscheinlichkeitsvektor der Zufallsvariable  $X$ . Ebenso erhalten wir  $w_2(1) = u(0, 1) + u(1, 1) = \frac{1}{3}$ ,  $w_2(2) = u(0, 2) + u(1, 2) = \frac{5}{12}$  und  $w_2(3) = u(0, 3) + u(1, 3) = \frac{1}{4}$ . Das ist der Wahrscheinlichkeitsvektor der Zufallsvariable  $Y$ , der die Wahrscheinlichkeit für das Ziehen jeder der drei Zahlen angibt.

**Definition:** Eine integrierbare Funktion  $g : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$  mit  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) dx_2 dx_1 = 1$  heißt gemeinsame Wahrscheinlichkeitsdichte der beiden Zufallsvariablen  $X_1$  und  $X_2$ , wenn  $P(X_1 \leq s \wedge X_2 \leq t) = \int_{-\infty}^s \int_{-\infty}^t g(x_1, x_2) dx_2 dx_1$  für alle  $s \in \mathbb{R}$  und alle  $t \in \mathbb{R}$  gilt.

**Satz 26:** Sei  $g : \mathbb{R}^2 \rightarrow [0, \infty)$  gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X_1$  und  $X_2$ . Dann gilt  $P((X_1, X_2) \in B) = \int_B g(x_1, x_2) d(x_1, x_2)$  für alle Normalbereiche  $B$  des  $\mathbb{R}^2$ , die auch unbeschränkt sein können (Definition im Anhang).

**Beweis:** Sei  $B$  zuerst ein Rechteck  $C = (r_1, s_1] \times (r_2, s_2]$ . Aus obiger Definition erhalten wir  $P(X_1 \leq s_1 \wedge X_2 \leq s_2) - P(X_1 \leq s_1 \wedge X_2 \leq r_2) = \int_{-\infty}^{s_1} \int_{-\infty}^{s_2} g(x_1, x_2) dx_2 dx_1 - \int_{-\infty}^{s_1} \int_{-\infty}^{r_2} g(x_1, x_2) dx_2 dx_1 = \int_{-\infty}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$ . Wegen des Additionssatzes gilt  $P(X_1 \leq s_1 \wedge X_2 \leq s_2) = P(X_1 \leq s_1 \wedge X_2 \leq r_2) + P(X_1 \leq s_1 \wedge r_2 < X_2 \leq s_2)$ . Es folgt

$$P(X_1 \leq s_1 \wedge r_2 < X_2 \leq s_2) = \int_{-\infty}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$$

Ersetzt man in dieser Gleichung  $s_1$  durch  $r_1$ , so hat man auch

$$P(X_1 \leq r_1 \wedge r_2 < X_2 \leq s_2) = \int_{-\infty}^{r_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$$

Der Additionssatz liefert auch  $P(X_1 \leq s_1 \wedge r_2 < X_2 \leq s_2) = P(X_1 \leq r_1 \wedge r_2 < X_2 \leq s_2) + P(r_1 < X_1 \leq s_1 \wedge r_2 < X_2 \leq s_2)$ . Mit Hilfe der beiden obigen Gleichungen ergibt sich  $P(r_1 < X_1 \leq s_1 \wedge r_2 < X_2 \leq s_2) = \int_{-\infty}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1 - \int_{-\infty}^{r_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1 = \int_{r_1}^{s_1} \int_{r_2}^{s_2} g(x_1, x_2) dx_2 dx_1$ . Damit haben wir  $P((X_1, X_2) \in C) = \int_C g(\mathbf{x}) d\mathbf{x}$  für das Rechteck  $C = (r_1, s_1] \times (r_2, s_2]$  gezeigt, wobei  $\mathbf{x} = (x_1, x_2)$ .

Sei jetzt  $B = C_1 \cup C_2 \cup \dots \cup C_j$  eine disjunkte Vereinigung von endlich vielen solchen Rechtecken. Das Ereignis  $(X_1, X_2) \in B$  tritt genau dann ein, wenn eines der Ereignisse  $(X_1, X_2) \in C_1, (X_1, X_2) \in C_2, \dots, (X_1, X_2) \in C_j$  eintritt. Wegen der Disjunktheit der Rechtecke  $C_i$  sind diese Ereignisse unvereinbar. Aus dem Additionssatz erhalten wir dann  $P((X_1, X_2) \in B) = P((X_1, X_2) \in C_1) + P((X_1, X_2) \in C_2) + \dots + P((X_1, X_2) \in C_j)$ . Da  $P((X_1, X_2) \in C_i) = \int_{C_i} g(\mathbf{x}) d\mathbf{x}$  für  $1 \leq i \leq j$  bereits oben bewiesen wurde, erhalten wir  $P((X_1, X_2) \in B) = \int_{C_1} g(\mathbf{x}) d\mathbf{x} + \int_{C_2} g(\mathbf{x}) d\mathbf{x} + \dots + \int_{C_j} g(\mathbf{x}) d\mathbf{x} = \int_B g(\mathbf{x}) d\mathbf{x}$ , wobei die letzte Gleichheit wegen der Disjunktheit der Rechtecke  $C_i$  folgt.

Sei  $B$  jetzt ein beschränkter Normalbereich. Sei weiters  $\varepsilon > 0$ . Dann existieren Mengen  $B_1$  und  $B_2$ , die disjunkte Vereinigungen von endlich vielen Rechtecken wie oben sind, sodass  $B_1 \subseteq B \subseteq B_2$  und  $\int_{B_2} g(\mathbf{x}) d\mathbf{x} - \int_{B_1} g(\mathbf{x}) d\mathbf{x} < \varepsilon$  gilt. Wegen  $B_1 \subseteq B \subseteq B_2$  erhalten wir  $P((X_1, X_2) \in B_1) \leq P((X_1, X_2) \in B) \leq P((X_1, X_2) \in B_2)$ . Ebenfalls wegen  $B_1 \subseteq B \subseteq B_2$  erhalten wir  $\int_{B_1} g(\mathbf{x}) d\mathbf{x} \leq \int_B g(\mathbf{x}) d\mathbf{x} \leq \int_{B_2} g(\mathbf{x}) d\mathbf{x}$ , da ja  $g \geq 0$  gilt. Wir haben  $P((X_1, X_2) \in B_1) = \int_{B_1} g(\mathbf{x}) d\mathbf{x}$  und  $P((X_1, X_2) \in B_2) = \int_{B_2} g(\mathbf{x}) d\mathbf{x}$  bereits oben bewiesen. Daher liegen  $P(X \in B)$  und  $\int_B g(\mathbf{x}) d\mathbf{x}$  in einem Intervall, das Länge  $\varepsilon$  hat. Damit ist  $|P((X_1, X_2) \in B) - \int_B g(\mathbf{x}) d\mathbf{x}| < \varepsilon$  gezeigt. Da  $\varepsilon > 0$  beliebig klein gewählt werden kann, muss auch  $P((X_1, X_2) \in B) = \int_B g(\mathbf{x}) d\mathbf{x}$  gelten.

Um unbeschränkte Normalbereiche zu behandeln, sei  $Q_m = [-m, m] \times [-m, m]$  für  $m \geq 1$ . Da  $Q_m$  beschränkt ist, ist  $P((X_1, X_2) \in Q_m) = \int_{Q_m} g(\mathbf{x}) d\mathbf{x}$  gezeigt. Wegen  $\int_{\mathbb{R}^2} g(\mathbf{x}) d\mathbf{x} = 1$  erhalten wir  $\lim_{m \rightarrow \infty} \int_{Q_m} g(\mathbf{x}) d\mathbf{x} = 1$ . Sei jetzt  $\varepsilon > 0$  beliebig. Dann existiert ein  $m$  mit  $1 - \int_{Q_m} g(\mathbf{x}) d\mathbf{x} < \varepsilon$ . Wir setzen  $U = \mathbb{R}^2 \setminus Q_m$ . Dann gilt  $\int_U g(\mathbf{x}) d\mathbf{x} = 1 - \int_{Q_m} g(\mathbf{x}) d\mathbf{x} < \varepsilon$  und  $P((X_1, X_2) \in U) = 1 - P((X_1, X_2) \in Q_m) = 1 - \int_{Q_m} g(\mathbf{x}) d\mathbf{x} < \varepsilon$ .

Sei  $B$  jetzt ein unbeschränkter Normalbereich im  $\mathbb{R}^2$ . Sei  $C = B \setminus U = B \cap Q_m$  und  $D = B \cap U$ . Wegen  $C \subseteq Q_m$  ist  $C$  ein beschränkter Normalbereich. Wegen  $D \subseteq U$  ergibt sich  $P((X_1, X_2) \in D) \leq P((X_1, X_2) \in U) < \varepsilon$ . Da  $B$  die disjunkte Vereinigung

von  $C$  und  $D$  ist, folgt  $P((X_1, X_2) \in B) = P((X_1, X_2) \in C) + P((X_1, X_2) \in D)$  aus dem Additionssatz. Damit erhalten wir dann

$$P((X_1, X_2) \in C) \leq P((X_1, X_2) \in B) \leq P((X_1, X_2) \in C) + \varepsilon$$

Wegen  $C \subseteq B = C \cup D$  und  $\int_D g(\mathbf{x}) d\mathbf{x} \leq \int_U g(\mathbf{x}) d\mathbf{x} < \varepsilon$  gilt auch

$$\int_C g(\mathbf{x}) d\mathbf{x} \leq \int_B g(\mathbf{x}) d\mathbf{x} \leq \int_C g(\mathbf{x}) d\mathbf{x} + \varepsilon$$

Da  $C$  beschränkt ist, wurde  $P((X_1, X_2) \in C) = \int_C g(\mathbf{x}) d\mathbf{x}$  bereits oben gezeigt. Es folgt  $|P((X_1, X_2) \in B) - \int_B g(\mathbf{x}) d\mathbf{x}| < \varepsilon$ . Da  $\varepsilon > 0$  beliebig klein gewählt werden kann, erhalten wir  $P((X_1, X_2) \in B) = \int_B g(\mathbf{x}) d\mathbf{x}$ .  $\square$

**Satz 27:** Sei  $g : \mathbb{R}^2 \rightarrow [0, \infty)$  eine gemeinsame Wahrscheinlichkeitsdichte der beiden Zufallsvariablen  $X$  und  $Y$ . Für  $x \in \mathbb{R}$  sei  $f_1(x) = \int_{-\infty}^{\infty} g(x, y) dy$  und für  $y \in \mathbb{R}$  sei  $f_2(y) = \int_{-\infty}^{\infty} g(x, y) dx$ . Dann ist  $f_1$  eine Wahrscheinlichkeitsdichte von  $X$  und  $f_2$  eine Wahrscheinlichkeitsdichte von  $Y$ , man bezeichnet diese auch als Randdichten von  $(X, Y)$ .

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

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

**Definition:** Seien  $X$  und  $Y$  Zufallsvariable und  $S$  und  $T$  endliche oder abzählbare Mengen, sodass  $X$  Werte in  $S$  und  $Y$  Werte in  $T$  annimmt. Die Zufallsvariablen  $X$  und  $Y$  heißen unabhängig, wenn  $P(X = i \wedge Y = j) = P(X = i)P(Y = j)$  für alle  $i \in S$  und  $j \in T$  gilt. Das ist äquivalent zur Unabhängigkeit der Ereignisse  $X = i$  und  $Y = j$ .

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

Die entsprechende Definition für kontinuierliche Zufallsvariable benutzt die Ordnung auf den reellen Zahlen.

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

**Satz 28:** Die Unabhängigkeit der kontinuierlichen Zufallsvariablen  $X$  mit Wahrscheinlichkeitsdichte  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  und  $Y$  mit Wahrscheinlichkeitsdichte  $g : \mathbb{R} \rightarrow \mathbb{R}_0^+$  ist äquivalent dazu, dass die Funktion  $h : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$  mit  $h(x, y) = f(x)g(y)$  eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X$  und  $Y$  ist.

**Beweis:** Die Aussage, dass  $P(X \leq s \wedge Y \leq t) = P(X \leq s)P(Y \leq t)$  für alle  $s, t \in \mathbb{R}$  gilt, ist äquivalent zu  $P(X \leq s \wedge Y \leq t) = \int_{-\infty}^s f(x) dx \int_{-\infty}^t g(y) dy = \int_{-\infty}^s f(x) \int_{-\infty}^t g(y) dy dx = \int_{-\infty}^s \int_{-\infty}^t f(x)g(y) dy dx$  für alle  $s, t \in \mathbb{R}$  nach Definition der Wahrscheinlichkeitsdichte. Das aber ist äquivalent dazu, dass  $h(x, y) = f(x)g(y)$  eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X$  und  $Y$  ist.  $\square$

## 2. Rechnen mit Zufallsvariablen

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

Wir beginnen mit dem Quadrat einer Zufallsvariablen.

**Satz 29:** Sei  $X$  eine kontinuierliche Zufallsvariable mit Wahrscheinlichkeitsdichte  $f$ . Dann ist  $X^2$  eine kontinuierliche Zufallsvariable mit Wertebereich  $\mathbb{R}_0^+$  und Wahrscheinlichkeitsdichte  $g(x) = \frac{1}{2\sqrt{x}}(f(\sqrt{x}) + f(-\sqrt{x}))$  für  $x \in \mathbb{R}^+$ .

**Beweis:** Sei  $F$  die Verteilungsfunktion der Zufallsvariablen  $X$ , also eine Stammfunktion von  $f$ . Sei weiters  $G$  die Verteilungsfunktion von  $X^2$  und  $g$  bezeichnet die Wahrscheinlichkeitsdichte von  $X^2$ . Da  $X^2 \geq 0$  gilt sowie  $P(X^2 = 0) = P(X = 0) = 0$ , folgt  $G(t) = P(X^2 \leq t) = 0$  für  $t \leq 0$ . Der Wertebereich von  $X^2$  ist  $\mathbb{R}_0^+$ . Für  $t > 0$  gilt

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

und durch Differenzieren nach  $t$  erhalten wir

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

Dazu rechnen wir ein Beispiel.

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

Die Wahrscheinlichkeitsdichte von  $X$  ist  $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ . Nach Satz 29 hat  $X^2$  den Wertebereich  $\mathbb{R}_0^+$  und für  $x > 0$  die Wahrscheinlichkeitsdichte

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

Das ist die Wahrscheinlichkeitsdichte der  $E(\frac{1}{2}, \frac{1}{2})$ -Verteilung (chi-quadrat-Verteilung mit einem Freiheitsgrad).

Wir beweisen auch Formeln für Summe und Quotient von zwei kontinuierlichen Zufallsvariablen.

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

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

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

**Beweis:** Sei  $t \in \mathbb{R}$  und sei  $B = \{(x, y) \in \mathbb{R}^2 : \frac{x}{y} \leq t, y \neq 0\}$ . Mit Satz 26 folgt  $P(\frac{X}{Y} \leq t) = P((X, Y) \in B) = \int_B g(x, y) d(x, y) = \int_{-\infty}^0 \int_{ty}^{\infty} g(x, y) dx dy + \int_0^{\infty} \int_{-\infty}^{ty} g(x, y) dx dy$ . Führt man in den beiden inneren Integralen die neue Integrationsvariable  $z = \frac{x}{y}$  ein, so erhält man  $P(\frac{X}{Y} \leq t) = \int_{-\infty}^0 \int_t^{-\infty} g(z y, y) y dz dy + \int_0^{\infty} \int_{-\infty}^t g(z y, y) y dz dy$ . Dabei ist  $y < 0$  beim ersten Integral zu beachten. Da  $\int_t^{-\infty} g(z y, y) y dz = - \int_{-\infty}^t g(z y, y) y dz = \int_{-\infty}^t g(z y, y) |y| dz$  für  $y < 0$  gilt erhalten wir  $P(\frac{X}{Y} \leq t) = \int_{-\infty}^{\infty} \int_{-\infty}^t g(z y, y) |y| dz dy$ . Wir vertauschen die Integrale und schreiben statt  $z$  wieder  $x$ . Das ergibt  $P(\frac{X}{Y} \leq t) = \int_{-\infty}^t \int_{-\infty}^{\infty} g(x y, y) |y| dy dx$ , daher ist die Funktion  $h(x) = \int_{-\infty}^{\infty} g(x y, y) |y| dy$  die Wahrscheinlichkeitsdichte der Zufallsvariablen  $\frac{X}{Y}$ .  $\square$

**Bemerkung:** Sind  $X$  und  $Y$  unabhängige Zufallsvariable mit Wahrscheinlichkeitsdichten  $f$  und  $g$ , dann hat die Zufallsvariable  $X + Y$  nach Satz 30 die Wahrscheinlichkeitsdichte  $(f * g)(x) := \int_{-\infty}^{\infty} f(x - y) g(y) dy$ , da  $h(x, y) = f(x) g(y)$  nach Satz 28 eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X$  und  $Y$  ist. Man bezeichnet die Funktion  $f * g$  als die Faltung der Funktionen  $f$  und  $g$ .

Genauso erhält man, dass  $\int_{-\infty}^{\infty} f(x y) g(y) |y| dy$  die Wahrscheinlichkeitsdichte der Zufallsvariablen  $\frac{X}{Y}$  ist.

Dazu rechnen wir zwei Beispiele.

**Beispiel 56:** Die Zufallsvariablen  $X$  und  $Y$  seien voneinander unabhängig. Weiters habe  $X$  die  $N(\mu_1, \sigma_1)$ -Verteilung und  $Y$  die  $N(\mu_2, \sigma_2)$ -Verteilung. Wir zeigen, dass  $X + Y$  dann die  $N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ -Verteilung hat.

Sei  $f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$ . Die Wahrscheinlichkeitsdichte von  $X$  ist dann  $f_{\mu_1, \sigma_1}$  und die von  $Y$  ist  $f_{\mu_2, \sigma_2}$ . Setzt man in die Formel aus obiger Bemerkung ein, so erhält man

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

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

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

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

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

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

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

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

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

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



**Beispiel 57:** Seien  $X$  und  $Y$  unabhängig und beide  $E(\lambda)$ -verteilt. Welche Dichte hat  $\frac{X}{Y}$ ?

Sei  $f(x) = \lambda e^{-\lambda x}$  für  $x \geq 0$  und  $f(x) = 0$  für  $x < 0$ . Das ist die Wahrscheinlichkeitsdichte von  $X$  und  $Y$ . Die Wahrscheinlichkeitsdichte von  $\frac{X}{Y}$  ist dann  $h(x) = \int_{-\infty}^{\infty} f(xy)f(y)|y|dy$ . Da  $f(y) = 0$  für  $y < 0$  gilt, erhält man  $h(x) = \int_0^{\infty} f(xy)f(y)ydy$ . Da  $f(xy) = 0$  für  $x < 0$  und  $y > 0$  gilt, ergibt sich  $h(x) = 0$  für  $x < 0$ .

Es bleibt also  $h(x)$  für  $x \geq 0$  zu berechnen. In diesem Fall kann man einsetzen und erhält  $h(x) = \int_0^{\infty} \lambda e^{-\lambda xy} \lambda e^{-\lambda y} y dy = \lambda^2 \int_0^{\infty} e^{-\lambda(x+1)y} y dy$ . Mit Hilfe partieller Integration berechnet man  $h(x) = \frac{1}{(1+x)^2}$ , daher gilt  $F(t) = 0$  für  $x < 0$  und  $F(t) = \frac{t}{1+t}$  für  $t \geq 0$ .

Das Resultat aus Beispiel 56 kann man auf eine Summe von  $n$  unabhängigen normalverteilten Zufallsvariablen verallgemeinern (wichtig für die Statistik). Dazu müssen wir auch die Definitionen verallgemeinern.

**Definition:** Man nennt die Zufallsvariablen  $X_1, X_2, \dots, X_n$  unabhängig, wenn für alle  $t_1, \dots, t_n \in \mathbb{R}$  gilt

$$P(X_1 \leq t_1 \wedge X_2 \leq t_2 \wedge \dots \wedge X_n \leq t_n) = P(X_1 \leq t_1) P(X_2 \leq t_2) \dots P(X_n \leq t_n).$$

**Definition:** Eine integrierbare Funktion  $g : \mathbb{R}^n \rightarrow \mathbb{R}_0^+$  mit  $\int_{\mathbb{R}^n} g(\mathbf{x})d\mathbf{x} = 1$  heißt gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X_1, X_2, \dots, X_n$ , wenn für alle  $t_1, \dots, t_n \in \mathbb{R}$  gilt  $P(X_1 \leq t_1 \wedge \dots \wedge X_n \leq t_n) = \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_n} g(x_1, \dots, x_n) dx_n \dots dx_1$ .

**Satz 32:** Die kontinuierlichen Zufallsvariablen  $X_1, X_2, \dots, X_n$  mit den Wahrscheinlichkeitsdichten  $f_1 : \mathbb{R} \rightarrow \mathbb{R}_0^+, \dots, f_n : \mathbb{R} \rightarrow \mathbb{R}_0^+$  sind unabhängig genau dann, wenn die Funktion  $g : \mathbb{R}^n \rightarrow \mathbb{R}_0^+$  mit  $g(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n)$  eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $X_1, \dots, X_n$  ist.

**Beweis:** Analog Satz 28. □

**Satz 33:** Sind die kontinuierlichen Zufallsvariablen  $X_1, X_2, \dots, X_n$  mit Wahrscheinlichkeitsdichten  $f_1, f_2, \dots, f_n$  unabhängig, dann ist für jedes  $1 \leq k \leq n$  die Zufallsvariable  $Y_k = X_1 + X_2 + \dots + X_k$  kontinuierlich und  $Y_k, X_{k+1}, X_{k+2}, \dots, X_n$  sind unabhängig.

**Beweis:** Der Beweis erfolgt mit Induktion. Wegen  $Y_1 = X_1$  ist der Induktionsanfang erfüllt. Sei nun die Induktionsvoraussetzung für  $1 \leq k < n$  erfüllt. Wir bezeichnen die Wahrscheinlichkeitsdichte von  $Y_k$  mit  $h_k$  und schließen mit Satz 32, dass

$$g(y, x_{k+1}, \dots, x_n) = h_k(y) f_{k+1}(x_{k+1}) \dots f_n(x_n)$$

eine gemeinsame Wahrscheinlichkeitsdichte der Zufallsvariablen  $Y_k, X_{k+1}, X_{k+2}, \dots, X_n$  ist. Weiters gilt  $Y_{k+1} = Y_k + X_{k+1}$ . Wir setzen

$$B = \{(y, x_{k+1}, \dots, x_n) \in \mathbb{R}^{n-k+1} : y + x_{k+1} \leq t_{k+1}, x_{k+2} \leq t_{k+2}, \dots, x_n \leq t_n\}$$

und verwenden eine höherdimensionale Verallgemeinerung von Satz 26, die genauso bewiesen wird. Mittels Vertauschung der Integrationsreihenfolge erhalten wir

$$\begin{aligned} P(Y_{k+1} \leq t_{k+1} \wedge X_{k+2} \leq t_{k+2} \wedge \dots \wedge X_n \leq t_n) &= P((Y_k, X_{k+1}, X_{k+2}, \dots, X_n) \in B) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{t_{k+1}-y} \int_{-\infty}^{t_{k+2}} \dots \int_{-\infty}^{t_n} h_k(y) f_{k+1}(x_{k+1}) f_{k+2}(x_{k+2}) \dots f_n(x_n) dx_n \dots dx_{k+2} dx_{k+1} dy \\ &= \int_{-\infty}^{t_{k+2}} \dots \int_{-\infty}^{t_n} \int_{-\infty}^{\infty} \int_{-\infty}^{t_{k+1}-y} h_k(y) f_{k+1}(x_{k+1}) dx_{k+1} dy f_{k+2}(x_{k+2}) \dots f_n(x_n) dx_n \dots dx_{k+2} \end{aligned}$$

Wir erhalten durch die neue Variable  $z = y + x_{k+1}$  und Vertauschung der Integrale

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{t_{k+1}-y} h_k(y) f_{k+1}(x_{k+1}) dx_{k+1} dy &= \int_{-\infty}^{\infty} \int_{-\infty}^{t_{k+1}} h_k(y) f_{k+1}(z-y) dz dy \\ &= \int_{-\infty}^{t_{k+1}} \int_{-\infty}^{\infty} h_k(y) f_{k+1}(z-y) dy dz \end{aligned}$$

Setzt man  $(f_{k+1} * h_k)(z) = \int_{-\infty}^{\infty} h_k(y) f_{k+1}(z-y) dy = h_{k+1}(z)$ , so ergibt sich

$$\begin{aligned} P(Y_{k+1} \leq t_{k+1} \wedge X_{k+2} \leq t_{k+2} \wedge \dots \wedge X_n \leq t_n) \\ = \int_{-\infty}^{t_{k+1}} \int_{-\infty}^{t_{k+2}} \dots \int_{-\infty}^{t_n} h_{k+1}(z) f(x_{k+2}) \dots f(x_n) dx_n \dots dx_{k+2} dz \end{aligned}$$

Das bedeutet, dass  $\tilde{g}(z, x_{k+2}, \dots, x_n) = h_{k+1}(z) f(x_{k+2}) \dots f(x_n)$  die gemeinsame Wahrscheinlichkeitsdichte der unabhängigen Zufallsvariablen  $Y_{k+1}, X_{k+2}, \dots, X_n$  ist.  $\square$

**Satz 34:** Sind die Zufallsvariablen  $X_1, X_2, \dots, X_n$  unabhängig und alle  $N(\mu, \sigma)$ -verteilt, dann hat  $X_1 + X_2 + \dots + X_n$  die  $N(n\mu, \sqrt{n}\sigma)$ -Verteilung.

Sei nun  $X_1, X_2, X_3, \dots$  eine Folge von  $N(\mu, \sigma)$ -verteilten Zufallsvariablen, sodass für jedes  $n \in \mathbb{N}$  die Zufallsvariablen  $X_1, \dots, X_n$  unabhängig sind. Dann besitzt für jedes  $n \in \mathbb{N}$  die Zufallsvariable

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma}$$

die Standardnormalverteilung (Zentraler Grenzwertsatz für normalverteilte Zufallsvariablen).

**Beweis:** Sei  $f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$ . Wir berechnen die Wahrscheinlichkeitsdichte  $h_k$  von  $X_1 + X_2 + \dots + X_k$  mit Induktion. Es gilt  $h_1(x) = f_{\mu, \sigma}(x)$ , das ist die Wahrscheinlichkeitsdichte von  $X_1$ . Angenommen  $h_k(x) = f_{k\mu, \sqrt{k}\sigma}(x)$  ist bereits bewiesen. Nach Satz 33 sind auch die Zufallsvariablen  $X_1 + \dots + X_k$  und  $X_{k+1}$  unabhängig, daher gilt nun  $h_{k+1}(x) = \int_{-\infty}^{\infty} h_k(x-y) f_{\mu, \sigma}(y) dy$ . Aus der Rechnung in Beispiel 56 folgt  $h_{k+1}(x) = f_{(k+1)\mu, \sqrt{k+1}\sigma}(x)$ . Damit ist  $h_n(x) = f_{n\mu, \sqrt{n}\sigma}(x)$  durch Induktion bewiesen und  $X_1 + X_2 + \dots + X_n$  hat die  $N(n\mu, \sqrt{n}\sigma)$ -Verteilung. Die Aussage über  $Z_n$  folgt nun unmittelbar aus Satz 22.  $\square$

### 3. Erwartungswert und Varianz

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

**Satz 35:** (Erwartungswert einer Funktion von Zufallsvariablen)

(a) Sei  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  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

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

wobei  $(w(i))_{i \in S}$  der Wahrscheinlichkeitsvektor von  $X$  ist.

(b) Sei  $\gamma : \mathbb{R}^2 \rightarrow \mathbb{R}$  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

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

wobei  $(u(i, j))_{i \in S, j \in T}$  der gemeinsame Wahrscheinlichkeitsvektor von  $X$  und  $Y$  ist.

**Beweis:** Wir beweisen (a). Wir berechnen zuerst den Wahrscheinlichkeitsvektor von  $Z$ . Sei  $U = \gamma(S)$ . Da  $X$  Werte in  $S$  annimmt, nimmt  $Z$  Werte in  $U$  an. Für  $k \in U$  gilt

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

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

$$\begin{aligned} E(Z) &= \sum_{k \in U} k v(k) = \sum_{k \in U} \sum_{i \in \gamma^{-1}(k) \cap S} k w(i) \\ &= \sum_{k \in U} \sum_{i \in \gamma^{-1}(k) \cap S} \gamma(i) w(i) && \text{weil } k = \gamma(i) \text{ aus } i \in \gamma^{-1}(k) \text{ folgt} \\ &= \sum_{i \in \gamma^{-1}(U) \cap S} \gamma(i) w(i) \\ &= \sum_{i \in S} \gamma(i) w(i) && \text{weil } \gamma^{-1}(U) \cap S = S \end{aligned}$$

Die Bedingung für diese Umformungen ist die absolute Konvergenz von  $\sum_{i \in S} \gamma(i) w(i)$ , denn dann kann die Reihe beliebig umgeordnet werden. Damit ist (a) bewiesen.

Der Beweis von (b) verläuft analog. Man ersetzt  $S$  durch  $S \times T$  und stellt sich den Buchstaben  $i$  fettgedruckt als ein Element von  $S \times T$  vor.  $\square$

**Bemerkung:** Nun sei  $X$  eine kontinuierliche Zufallsvariable mit einer Wahrscheinlichkeitsdichte  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  und sei  $\gamma : \mathbb{R} \rightarrow \mathbb{R}$  eine stetige Funktion. Die entsprechende Formel für den Erwartungswert von  $Z = \gamma(X)$  lautet

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

Die Bedingung für die Existenz dieses Erwartungswertes ist die Konvergenz des Integrals  $\int_{-\infty}^{\infty} |\gamma(x)| f(x) dx$ .

Jetzt können wir die Eigenschaften des Erwartungswertes herleiten.

**Satz 36:** Seien  $X$  und  $Y$  Zufallsvariable und  $a$  und  $b$  in  $\mathbb{R}$ . Dann gilt

- (a)  $E(aX + b) = aE(X) + b$
- (b)  $E(X + Y) = E(X) + E(Y)$
- (c)  $E(XY) = E(X)E(Y)$ , wenn  $X$  und  $Y$  unabhängig sind.

**Beweis:** Wir führen den Beweis nur für den Fall diskreter Zufallsvariable durch.

(a) Wir wenden Satz 35 (a) mit  $\gamma(x) = ax + b$  an. Ist  $(w(i))_{i \in S}$  der Wahrscheinlichkeitsvektor der Zufallsvariablen  $X$ , so haben wir

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

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

(b) Sei  $(u(i, j))_{i \in S, j \in T}$  der gemeinsame Wahrscheinlichkeitsvektor von  $X$  und  $Y$ . Aus Satz 35 (b) mit  $\gamma(x, y) = x + y$  folgt  $E(X + Y) = \sum_{i \in S} \sum_{j \in T} (i + j) u(i, j)$ . Aus Satz 35 (b) mit  $\gamma(x, y) = x$  folgt  $E(X) = \sum_{i \in S} \sum_{j \in T} i u(i, j)$ . Aus Satz 35 (b) mit  $\gamma(x, y) = y$  folgt  $E(Y) = \sum_{i \in S} \sum_{j \in T} j u(i, j)$ . Das ergibt  $E(X + Y) = E(X) + E(Y)$ .

(c) Sei  $(w_1(i))_{i \in S}$  der Wahrscheinlichkeitsvektor von  $X$  und  $(w_2(j))_{j \in T}$  der von  $Y$ . Da die Zufallsvariablen  $X$  und  $Y$  unabhängig sind, ist durch  $u(i, j) = w_1(i) w_2(j)$  mit  $i \in S$  und  $j \in T$  der gemeinsame Wahrscheinlichkeitsvektor von  $X$  und  $Y$  gegeben. Wir wenden Satz 35 (b) mit  $\gamma(x, y) = xy$  an und erhalten

$$E(XY) = \sum_{i \in S} \sum_{j \in T} ij u(i, j) = \sum_{i \in S} \sum_{j \in T} ij w_1(i) w_2(j) = \sum_{i \in S} i w_1(i) \sum_{j \in T} j w_2(j)$$

Wegen  $E(X) = \sum_{i \in S} i w_1(i)$  und  $E(Y) = \sum_{j \in T} j w_2(j)$  ist damit auch (c) gezeigt.  $\square$

Die Varianz  $V(X)$  einer Zufallsvariablen  $X$  kann man ganz allgemein definieren durch  $V(X) = E((X - m)^2)$  mit  $m = E(X)$ . Die in Kapitel III.2 gegebene Definition der Varianz folgt aus Satz 35 mit  $\gamma(x) = (x - m)^2$ . Aus den in Satz 36 bewiesenen Eigenschaften des Erwartungswertes kann man die Eigenschaften der Varianz herleiten.

**Satz 37:** Seien  $X$  und  $Y$  Zufallsvariable und  $a$  und  $b$  in  $\mathbb{R}$ . Dann gilt

- (a)  $V(X) = E(X^2) - (E(X))^2$
- (b)  $V(aX + b) = a^2V(X)$
- (c)  $V(X + Y) = V(X) + V(Y)$ , wenn  $X$  und  $Y$  unabhängig sind

**Beweis:** Wir setzen  $m = E(X)$  und  $l = E(Y)$ .

(a) Mit Hilfe von Satz 36 (a) und (b) folgt

$$V(X) = E((X - m)^2) = E(X^2 - 2mX + m^2) = E(X^2) - 2mE(X) + m^2 = E(X^2) - m^2$$

Damit ist  $V(X) = E(X^2) - (E(X))^2$  gezeigt.

(b) Aus Satz 36 (a) folgt  $E(aX + b) = am + b$ . Damit erhalten wir

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

wobei noch einmal Satz 36 (a) verwendet wurde, und (b) ist gezeigt.

(c) Aus Satz 36 (b) folgt  $E(X + Y) = m + l$ . Mit Hilfe von (a) erhalten wir

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

Da  $E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2)$  aus Satz 36 (b) und  $E(XY) = ml$  aus Satz 36 (c) folgt, ergibt sich aus obiger Gleichung

$$V(X + Y) = E(X^2) - m^2 + E(Y^2) - l^2 = V(X) + V(Y)$$

wobei nochmals (a) verwendet wurde. □

## V. Statistik

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

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

### 1. Parameterschätzer

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

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

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

Das ist eine quadratische Funktion in  $m$ , die ihr Minimum für  $m = \frac{1}{n} \sum_{j=1}^n x_j$  annimmt. Üblicherweise nennt man

$$m = \frac{1}{n} \sum_{j=1}^n x_j$$

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

Wir untersuchen zwei zusammenhängende Größen, zum Beispiel Körpergröße und Körpergewicht. Aus den erwachsenen Einwohnern Österreichs wird eine zufällige Stichprobe

vom Umfang  $n$  gezogen. Die  $j$ -te Person habe Körpergröße  $x_j$  und Körpergewicht  $y_j$ . Das ergibt zwei miteinander verbundene Stichproben  $x_1, x_2, \dots, x_n$  und  $y_1, y_2, \dots, y_n$ . Seien  $m_x$  und  $m_y$  die Stichprobenmittel und  $v_x$  und  $v_y$  die Stichprobenvarianzen. Analog zur Stichprobenvarianz definieren wir die Stichprobenkovarianz dieser beiden Stichproben durch

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

Es gilt dann  $v_x = c_{xx}$  und  $v_y = c_{yy}$ . Wir nehmen an, dass die Körpergrößen in der Stichprobe nicht alle gleich groß sind. Dasselbe nehmen wir für die Körpergewichte an. Dann gilt  $v_x > 0$  und  $v_y > 0$ .

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

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

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

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

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

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

**Satz 38:** Für den Korrelationskoeffizienten  $k_{xy}$  gilt  $-1 \leq k_{xy} \leq 1$ . Ist  $k_{xy} = 1$ , dann liegen die Punkte  $(x_j, y_j)$  für  $1 \leq j \leq n$  auf einer Geraden mit Anstieg  $> 0$ . Ist  $k_{xy} = -1$ , dann liegen diese Punkte auf einer Geraden mit Anstieg  $< 0$ .

**Beweis:** Nach obiger Rechnung ist  $v_y - \frac{c_{xy}^2}{v_x} = v_y(1 - k_{xy}^2)$  das Minimum der Quadratsumme  $\frac{1}{n} \sum_{j=1}^n (y_j - ax_j - b)^2$ . Es muss  $\geq 0$  sein. Es folgt  $k_{xy}^2 \leq 1$ , das heißt  $-1 \leq k_{xy} \leq 1$ . Gilt  $k_{xy}^2 = 1$ , dann gilt auch  $\frac{1}{n} \sum_{j=1}^n (y_j - ax_j - b)^2 = 0$ , das heißt die Punkte  $(x_j, y_j)$  für  $1 \leq j \leq n$  liegen alle auf der Geraden  $y = ax + b$  mit Anstieg  $a = \frac{c_{xy}}{v_x}$ . Ist  $k_{xy} = 1$ , dann gilt  $c_{xy} > 0$  und somit  $a > 0$ . Ist  $k_{xy} = -1$ , dann gilt  $c_{xy} < 0$  und somit  $a < 0$ .  $\square$

Der Korrelationskoeffizienten  $k_{xy}$  ist eine Maßzahl dafür, wie sehr die Punktemenge  $\{(x_j, y_j) : 1 \leq j \leq n\}$  von der am besten angepassten Gerade abweicht. Je näher  $k_{xy}$  bei 1

oder  $-1$  liegt, umso besser lässt sich diese Punktemenge durch eine Gerade approximieren. Je näher  $k_{xy}$  bei  $0$  liegt, umso weniger ist das möglich.

Wir haben hier bereits gezogene Stichproben behandelt. In diesem Fall liegen  $n$  Werte  $x_1, x_2, \dots, x_n$  vor. Allerdings sind diese Werte Ergebnisse eines Zufallsexperiments, nämlich des zufälligen Ziehens einer Stichprobe. Daher werden wir sie im Folgenden auch als Zufallsvariable  $X_1, X_2, \dots, X_n$  schreiben. Da es sich um die unabhängige Wiederholung ein und desselben Zufallsvorgangs handelt, setzen wir die Zufallsvariablen  $X_1, X_2, \dots, X_n$  als unabhängig voraus und gehen von derselben Verteilung aus. Somit stimmen auch die Erwartungswerte und Varianzen überein und es folgt für die Zufallsvariable des Mittelwerts  $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  mittels Satz 36

$$E(M) = \frac{1}{n}(E(X_1) + \dots + E(X_n)) = E(X_1) = \dots = E(X_n)$$

Aufgrund der Unabhängigkeit folgt für die Varianz von  $M$  mittels Satz 37

$$V(M) = V\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n^2}(V(X_1) + \dots + V(X_n)) = \frac{1}{n}V(X_1) = \dots = \frac{1}{n}V(X_n)$$

Während der Erwartungswert des Mittelwerts bei steigender Anzahl  $n$  konstant ist, nehmen die Varianz mit  $1/n$  und die Standardabweichung mit  $1/\sqrt{n}$  ab. Damit werden große Abweichungen von  $M$  von seinem Erwartungswert  $V(X_1) = \dots = V(X_n)$  mit steigendem Stichprobenumfang  $n$  immer unwahrscheinlicher.

Wir berechnen nun noch den Erwartungswert der Zufallsvariablen  $\sum_{j=1}^n (X_j - M)^2$  mit den Sätzen 36 und 37. Im Verlauf der Rechnung wird die Zufallsvariable  $X_j M$  in die Summe von  $X_j(M - \frac{1}{n}X_j) = X_j \frac{1}{n}(X_1 + X_2 + \dots + X_{j-1} + X_{j+1} + \dots + X_n)$  (d.h. ein Produkt von zwei unabhängigen Zufallsvariablen) und  $\frac{1}{n}X_j^2$  zerlegt:

$$\begin{aligned} E\left(\sum_{j=1}^n (X_j - M)^2\right) &= \sum_{j=1}^n E((X_j - M)^2) = \sum_{j=1}^n (E(X_j^2) - E(2X_j M) + E(M^2)) = \\ &= n(V(X_1) + (E(X_1))^2) - 2 \sum_{j=1}^n E(X_j(M - \frac{1}{n}X_j)) - \frac{2}{n} \sum_{j=1}^n E(X_j^2) + n E(M^2) = \\ &= n(V(X_1) + (E(X_1))^2 + V(M) + (E(M))^2) - 2 \sum_{j=1}^n E(X_j) E(M - \frac{1}{n}X_j) - \frac{2}{n} \sum_{j=1}^n E(X_j^2) \\ &= (n+1)V(X_1) + 2n(E(X_1))^2 - 2n(E(X_1))^2 \frac{n-1}{n} - 2V(X_1) - 2(E(X_1))^2 = \\ &= (n-1)V(X_1) \end{aligned}$$

Somit besitzt die Zufallsvariable

$$\frac{1}{n-1} \sum_{j=1}^n (X_j - M)^2$$

den Erwartungswert  $V(X_1) = \dots = V(X_n)$ . Daher verwendet man für die Stichprobenvarianz oft eine korrigierte Version

$$v_x^* = \frac{1}{n-1} \sum_{j=1}^n (x_j - m_x)^2$$

mit dem Nenner  $n-1$ , denn bei dieser handelt es sich um eine sogenannte erwartungstreue Schätzfunktion.

## 2. Konfidenzintervalle für normalverteilte Messwerte

Misst man eine (physikalische) Größe mehrmals, so werden die Messwerte 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. Dabei ist  $\mu$  die zu messende unbekannte Größe und  $\sigma$  gibt die Genauigkeit des Messgerätes an.

Man führt den Messvorgang  $n$  Mal durch, die dabei gemessenen Werte  $X_1, X_2, \dots, X_n$  bilden eine Stichprobe. 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  $X_1, X_2, \dots, X_n$  unabhängig.

Um die zu messende unbekannte Größe  $\mu$  ausreichend genau zu bestimmen, gibt man ein  $\gamma \in (0, 1)$  vor (übliche Werte für  $\gamma$  sind 0,95 oder 0,99) und berechnet aus den gemessenen Werten  $X_1, X_2, \dots, X_n$  ein Intervall  $I$ , das den unbekannt Parameter  $\mu$  mit Wahrscheinlichkeit  $\gamma$  enthält. So ein Intervall  $I$  nennt man  $\gamma$ -Konfidenzintervall für die Größe  $\mu$ . Für  $\gamma$  gibt es verschiedene Namen, wie zum Beispiel Sicherheitswahrscheinlichkeit oder Konfidenzniveau. Das Konfidenzintervall ist eine Zufallsvariable und hängt von den gemessenen Werten  $X_1, X_2, \dots, X_n$  ab, also von zufälligen Einflüssen. Führt man so eine Serie von  $n$  Messungen mehrere Male durch und berechnet aus jeder das  $\gamma$ -Konfidenzintervall, dann wird man verschiedene Intervalle erhalten. Auch wird  $\mu$  nicht immer im Konfidenzintervall enthalten sein. Ist zum Beispiel  $\gamma = 0,95$ , dann wird das Konfidenzintervall den festen Parameter  $\mu$  mit einer Wahrscheinlichkeit von 95% enthalten, das heißt bei wiederholter Durchführung einer Serie von  $n$  Messungen und Bestimmung des Konfidenzintervalls für jede einzelne Serie wird sich  $\mu$  im Durchschnitt bei jeder zwanzigsten Serie außerhalb des Intervalls befinden.

Es ist methodisch nicht korrekt, eine einzelne Serie von  $n$  Messungen durchzuführen, daraus ein Konfidenzintervall zu ermitteln und zu behaupten:  $\mu$  liegt mit einer Wahrscheinlichkeit von  $\gamma$  in dem Konfidenzintervall. Dieses eine Konfidenzintervall enthält den unbekannt, aber festen Parameter  $\mu$  oder es enthält den Parameter eben nicht.

Um dieses so definierte  $\gamma$ -Konfidenzintervall zu berechnen, benötigen wir folgende Definition.

**Definition:** Für  $\beta \in (0, 1)$  sei  $z_\beta$  die eindeutige Lösung von  $\Phi(z_\beta) = 1 - \beta$ . Diese Lösung ist eindeutig, da  $\Phi : \mathbb{R} \rightarrow (0, 1)$  eine invertierbare Abbildung ist (Satz 23 (c)).

Die Werte von  $z_\beta$  findet man in einer Tabelle. Das im folgenden Satz bestimmte Intervall heißt zweiseitiges  $\gamma$ -Konfidenzintervall. Wir nehmen an, dass wir die Genauigkeit des Messgerätes, also  $\sigma$ , kennen.

**Satz 39:** Sei  $0 < \gamma < 1$  gegeben. Seien  $X_1, X_2, \dots, X_n$  unabhängige  $N(\mu, \sigma)$ -verteilte Messwerte, wobei  $\sigma$  bekannt ist. Sei  $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  der Mittelwert der Messwerte (das sogenannte Stichprobenmittel) und  $\beta = \frac{1-\gamma}{2}$ . Für  $I = [M - \frac{\sigma z_\beta}{\sqrt{n}}, M + \frac{\sigma z_\beta}{\sqrt{n}}]$  gilt dann  $P(\mu \in I) = \gamma$ , das heißt  $I$  ist ein  $\gamma$ -Konfidenzintervall für  $\mu$ .

**Beweis:** Nach Satz 34 hat  $\sqrt{n} \frac{M - \mu}{\sigma} = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma}$  die  $N(0, 1)$ -Verteilung. Daher gilt

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



Wir formen dieses Ereignis um

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

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

**Bemerkung:** In den formalen Umformungen im Beweis wurde ein Intervall  $I$  für den Parameter  $\mu$  ermittelt, sodass für alle Parameter  $\mu \in I$  das ermittelte Stichprobenmittel  $M$  oberhalb des  $\beta$ -Quantils und unterhalb des  $1 - \beta$ -Quantils der betreffenden  $N(\mu, \sigma/\sqrt{n})$ -Normalverteilung liegt. Das heißt, für einen Parameter im Konfidenzintervall ist die erhaltene Stichprobe nicht zu unwahrscheinlich.

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

**Beispiel 58:** Eine unbekannte Größe  $\mu$  wird 25 Mal gemessen. Die Genauigkeit des Messvorgangs wird mit  $\sigma = 5$  cm angegeben. Als Mittelwert der Messwerte ergibt sich 3,24 m. Gesucht ist ein 95%-Konfidenzintervall.

Wir setzen in die Formel ein. Wegen  $\gamma = 0,95$  folgt  $\beta = \frac{1-\gamma}{2} = 0,025$ . Man findet  $z_\beta = 1,96$  in einer Tabelle. Jetzt erhalten wir  $\frac{\sigma z_\beta}{\sqrt{n}} = \frac{0,05 \cdot 1,96}{\sqrt{25}} = 0,02$  und daraus ergibt sich  $I = [3,24 - 0,02; 3,24 + 0,02] = [3,22; 3,26]$  für das 95%-Konfidenzintervall.

**Beispiel 59:** Die Füllgewichte von Waschmittelpackungen einer bestimmten Sorte seien  $N(\mu, \sigma)$ -verteilt. Die Genauigkeit der Abfüllmaschine wird mit 0,02 kg angegeben. Man wiegt 10 zufällig gewählte Packungen und erhält die Inhalte 1,05; 0,99; 1,03; 1,03; 1,01; 1,02; 1,01; 0,97; 1,01; 0,98 in kg. Gesucht ist ein 99%-Konfidenzintervall für den durchschnittlichen Packungsinhalt  $\mu$ .

Wegen  $\gamma = 0,99$  folgt  $\beta = \frac{1-\gamma}{2} = 0,005$ . Man findet  $z_\beta = 2,58$  in einer Tabelle. Weiters berechnen wir  $M = \frac{1}{10}(1,05 + 0,99 + \dots + 0,98) = \frac{10,1}{10} = 1,01$  und  $\frac{\sigma z_\beta}{\sqrt{n}} = \frac{0,02 \cdot 2,58}{\sqrt{10}} = 0,019$ . Daraus ergibt sich dann  $I = [1,01 - 0,019; 1,01 + 0,019] = [0,992; 1,029]$  für das 99%-Konfidenzintervall.

Bevor man ein Konfidenzintervall ermittelt, muss man zwei Entscheidungen treffen. Die eine ist die Wahl der statistischen Sicherheit  $\gamma$ , die angibt mit welcher Wahrscheinlichkeit das Konfidenzintervall die unbekannte Größe  $\mu$  enthält. Die andere ist die Wahl des Stichprobenumfangs. Hat man sich für eine statistische Sicherheit  $\gamma$  entschieden, dann ist auch  $z_\beta$  festgelegt und man kann die Länge  $|I|$  des Konfidenzintervalls  $I$  durch den Stichprobenumfang  $n$  steuern. Aus der Formel in Satz 39 folgt  $|I| = \frac{2\sigma z_\beta}{\sqrt{n}}$ . Vergrößert man den Stichprobenumfang  $n$ , dann wird das Konfidenzintervall kleiner.

**Beispiel 60:** Wieviele Packungen muss man in Beispiel 59 überprüfen, damit die Länge des 99%-Konfidenzintervalls höchstens 0,01 kg beträgt?

Für  $\gamma = 0,99$  haben wir  $z_\beta = 2,58$  gefunden. Es wird verlangt, dass  $|I| = \frac{2\sigma z_\beta}{\sqrt{n}} \leq 0,01$  gilt. Daraus folgt  $\sqrt{n} \geq \frac{2\sigma z_\beta}{0,01} = \frac{2 \cdot 0,02 \cdot 2,58}{0,01} = 10,32$  und  $n \geq 106,5$ . Man muss mindestens 107 Packungen überprüfen.

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

**Satz 40:** Sei  $0 < \gamma < 1$  gegeben. Seien  $X_1, X_2, \dots, X_n$  unabhängige  $N(\mu, \sigma)$ -verteilte Messwerte, wobei  $\sigma$  bekannt ist. Sei  $M = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$  der Mittelwert der Messwerte. Dann sind die beiden Intervalle  $I_1 = (-\infty, M + \frac{\sigma z_{1-\gamma}}{\sqrt{n}}]$  und  $I_2 = [M - \frac{\sigma z_{1-\gamma}}{\sqrt{n}}, \infty)$   $\gamma$ -Konfidenzintervalle für  $\mu$ .

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

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

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

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

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

Es wurde  $\sigma = 0,02$  kg angegeben und  $M = 1,01$  kg berechnet. Für  $\gamma = 0,99$  findet man  $z_{1-\gamma} = 2,33$  in einer Tabelle. Es folgt  $\frac{\sigma z_{1-\gamma}}{\sqrt{n}} = \frac{0,02 \cdot 2,33}{\sqrt{10}} = 0,0145$  und daraus  $I_2 = [0,9955; \infty)$ .

Häufig treten auch Situationen auf, in denen der Wert von  $\sigma$  nicht bekannt ist. In diesem Fall ersetzt man  $\sigma$  durch den aus der Stichprobe  $X_1, X_2, \dots, X_n$  berechneten Schätzwert

$$S = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (X_j - M)^2}.$$

Man kann zeigen, dass die Zufallsvariable  $\sqrt{n} \frac{M - \mu}{S}$  die sogenannte  $T(m)$ -Verteilung (auch Student-Verteilung) mit Parameter  $m = n - 1$  hat. Davon abgesehen geht man so vor wie oben. Sei  $\beta = \frac{1-\gamma}{2}$  und  $v_{\beta, m}$  genauso definiert wie  $z_{\beta}$ , wobei jedoch statt der  $N(0, 1)$ -Verteilung die  $T(m)$ -Verteilung verwendet wird. Man findet  $v_{\beta, m}$  in einer Tabelle, mit steigendem  $m = n - 1$  nähert sich der Wert  $v_{\beta, m}$  von oben her dem Wert  $z_{\beta}$  an. Das Intervall

$$K = \left[ M - \frac{S v_{\beta, n-1}}{\sqrt{n}}, M + \frac{S v_{\beta, n-1}}{\sqrt{n}} \right]$$

ist dann ein  $\gamma$ -Konfidenzintervall für  $\mu$ . Entsprechendes gilt für die einseitigen Konfidenzintervalle, man ersetzt  $\sigma$  durch  $S$  und  $z_{1-\gamma}$  durch  $v_{1-\gamma, n-1}$ .

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

Aus  $\gamma = 0,99$  folgt  $\beta = 0,005$  und man findet  $v_{\beta, 9} = 3,25$  in einer Tabelle. Der Stichprobenumfang  $n$  beträgt 10. In Beispiel 59 wurde  $M = 1,01$  ermittelt. Weiters ergibt sich

$$S = \sqrt{\frac{1}{9}((1,05 - 1,01)^2 + (0,99 - 1,01)^2 + (1,03 - 1,01)^2 + \dots + (0,98 - 1,01)^2)} = 0,024$$

Jetzt folgt  $\frac{S v_{\beta, n-1}}{\sqrt{n}} = \frac{0,024 \cdot 3,25}{\sqrt{10}} = 0,025$  und  $K = [0,985; 1,035]$ . Das erhaltene Konfidenzintervall ist nun größer als in Beispiel 59.

### 3. Konfidenzintervalle für Prozentsätze

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

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

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

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

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

**Satz 41:** Sei  $p$  die unbekannte relative Häufigkeit einer Eigenschaft in der Grundgesamtheit. Sei  $n$  der Stichprobenumfang und  $X$  die Anzahl mit der die Eigenschaft in der Stichprobe auftritt. Man wählt  $\gamma \in (0, 1)$ . Sei  $\beta = \frac{1-\gamma}{2}$  und  $I = [\frac{1}{n}X - \frac{z_\beta}{2\sqrt{n}}, \frac{1}{n}X + \frac{z_\beta}{2\sqrt{n}}]$ . Dann gilt  $P(p \in I) \geq \gamma$ , das heißt  $I$  ist ein  $\gamma$ -Konfidenzintervall für  $p$ .

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

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

Wir formen dieses Ereignis um

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

Sei  $\tilde{I} = [\frac{1}{n}X - \frac{z_\beta}{\sqrt{n}} \sqrt{p(1-p)}, \frac{1}{n}X + \frac{z_\beta}{\sqrt{n}} \sqrt{p(1-p)}]$ . Wir haben  $P(p \in \tilde{I}) = \gamma$  gezeigt. Nun gilt  $\sqrt{p(1-p)} \leq \frac{1}{2}$  für alle  $p \in [0, 1]$  nach dem Hilfssatz. Ersetzen wir  $\sqrt{p(1-p)}$  in  $\tilde{I}$  durch  $\frac{1}{2}$ , so wandert der linke Endpunkt nach links und der rechte Endpunkt nach rechts.

Wir erhalten dadurch das Intervall  $I$ . Es gilt also  $\tilde{I} \subseteq I$ . Ist  $p \in \tilde{I}$  erfüllt, dann auch  $p \in I$ . Aus  $P(p \in \tilde{I}) = \gamma$  folgt daher  $P(p \in I) \geq \gamma$ .  $\square$

**Bemerkung:** Im Beweis von Satz 41 sind wir etwas großzügig vorgegangen. Um das im Intervall  $\tilde{I}$  vorkommende unbekannte  $p$  loszuwerden, haben wir das Intervall vergrößert und so  $I$  erhalten. Hat man bei diesem Übergang wenig verschenkt, dann wird  $I$  ein gutes Konfidenzintervall sein. Hat man dabei viel verschenkt, so wird  $I$  wenig Aussagekraft haben, weil es zu groß ist. Wir vergleichen die Längen  $|I|$  und  $|\tilde{I}|$  dieser beiden Intervalle. Der Quotient  $\frac{|\tilde{I}|}{|I|} = 2\sqrt{p(1-p)}$  gibt an, um wieviel  $\tilde{I}$  kleiner als  $I$  ist.

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 Konfidenzintervall. Das gilt auch noch, wenn  $p$  zwischen 0,2 und 0,8 liegt. Je näher jedoch  $p$  bei 0 oder 1 liegt, umso unbrauchbarer wird  $I$ , weil es dann viel zu groß ist.

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

Wir berechnen das Intervall  $I$  aus Satz 41. Wegen  $\gamma = 0,95$  folgt  $\beta = \frac{1-\gamma}{2} = 0,025$ . Man findet  $z_\beta = 1,96$  in einer Tabelle. Jetzt erhalten wir  $\frac{z_\beta}{2\sqrt{n}} = \frac{1,96}{2\sqrt{400}} = 0,049$ . Wegen  $\frac{1}{n}X = \frac{164}{400} = 0,41$  folgt  $I = [0,361; 0,459]$ .

**Beispiel 64:** Wie groß muss man den Stichprobenumfang  $n$  wählen, wenn die Länge des 95%-Konfidenzintervalls  $I$  höchstens 0,02 sein soll?

Die Länge von  $I$  ist  $\frac{z_\beta}{\sqrt{n}}$ . Es wird also verlangt, dass  $\frac{z_\beta}{\sqrt{n}} \leq 0,02$  gilt. Für  $\gamma = 0,95$  wurde  $z_\beta = 1,96$  in Beispiel 63 ermittelt. Es folgt  $\sqrt{n} \geq \frac{1,96}{0,02} = 98$ , das heißt  $n \geq 9604$ .

Genauso wie das zweiseitige Konfidenzintervall  $I$  in Satz 41 berechnet man auch die einseitigen  $\gamma$ -Konfidenzintervalle  $I_1 = (-\infty, \frac{1}{n}X + \frac{z_{1-\gamma}}{2\sqrt{n}}]$  und  $I_2 = [\frac{1}{n}X - \frac{z_{1-\gamma}}{2\sqrt{n}}, \infty)$ . Da  $p$  aber ein Prozentsatz ist und daher immer im Intervall  $[0, 1]$  liegt, kann man auch  $I_1 = [0, \frac{1}{n}X + \frac{z_{1-\gamma}}{2\sqrt{n}}]$  und  $I_2 = [\frac{1}{n}X - \frac{z_{1-\gamma}}{2\sqrt{n}}, 1]$  schreiben.

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

Für  $\gamma = 0,95$  finden wir  $z_{1-\gamma} = 1,65$  in einer Tabelle. Es folgt  $\frac{z_{1-\gamma}}{2\sqrt{n}} = \frac{1,65}{2\sqrt{400}} = 0,041$ . Wegen  $\frac{1}{n}X = \frac{164}{400} = 0,41$  erhält man  $I_2 = [0,369; 1]$ .

Was soll man tun, wenn  $p$  in der Nähe von 0 oder von 1 liegt, sodass man das Konfidenzintervall  $I$  nicht mehr verwenden will? Eine – etwas unexakte – Möglichkeit besteht darin, dass man das im Beweis von Satz 41 berechnete Intervall  $\tilde{I}$  nimmt und dort den unbekanntes Prozentsatz  $p$  durch den Prozentsatz  $\frac{X}{n}$ , der in der Stichprobe auftritt, ersetzt. Man erhält dadurch das Intervall

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

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

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

Da  $p$  klein ist, berechnen wir  $J$ . Wegen  $\gamma = 0,99$  folgt  $\beta = \frac{1-\gamma}{2} = 0,005$ . Man findet  $z_\beta = 2,58$  in einer Tabelle. Weiters folgt  $\frac{1}{n}X = \frac{128}{1600} = 0,08$ . Jetzt erhalten wir  $\frac{z_\beta}{\sqrt{n}}\sqrt{\frac{X}{n}(1-\frac{X}{n})} = \frac{2,58}{\sqrt{1600}}\sqrt{0,08 \cdot 0,92} = 0,0175$ . Daraus ergibt sich  $J = [0,0625; 0,0975]$ .

Für das Intervall  $I$  hätte man  $I = [0,048; 0,112]$ , also etwas viel schlechteres, erhalten.

**Bemerkung:** Es gibt auch eine Möglichkeit, auf exakte Weise ein Konfidenzintervall zu bestimmen. Wir gehen von der im Beweis von Satz 41 bewiesenen Gleichung

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

aus und formen das darin vorkommende Ereignis um

$$\begin{aligned} -z_\beta \leq \frac{X - np}{\sqrt{np(1-p)}} \leq z_\beta &\Leftrightarrow \frac{(X - np)^2}{np(1-p)} \leq z_\beta^2 \\ &\Leftrightarrow X^2 - 2Xnp + n^2p^2 \leq z_\beta^2 np - z_\beta^2 np^2 \\ &\Leftrightarrow p^2(n^2 + nz_\beta^2) - 2p(nX + \frac{1}{2}nz_\beta^2) + X^2 \leq 0 \\ &\Leftrightarrow \lambda_1 \leq p \leq \lambda_2 \end{aligned}$$

wobei  $\lambda_1$  und  $\lambda_2$  die beiden Lösungen der quadratischen Gleichung

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

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

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

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

**Beispiel 67:** Wir berechnen für das Beispiel 66 das Konfidenzintervall  $K$ . Es gilt

$$Y = z_\beta \sqrt{\frac{X(n-X)}{n} + \frac{1}{4}z_\beta^2} = 2,58 \sqrt{\frac{128(1600-128)}{1600} + \frac{1}{4} \cdot 2,58^2} = 28,19$$

$$X + \frac{1}{2}z_\beta^2 = 128 + \frac{1}{2} \cdot 2,58^2 = 131,33$$

$$n + z_\beta^2 = 1600 + 2,58^2 = 1606,66$$

Setzt man das in die obige Formel für  $K$  ein, so erhält man

$$K = \left[ \frac{131,33-28,19}{1606,66}, \frac{131,33+28,19}{1606,66} \right] = [0,0642; 0,0993].$$

**Beispiel 68:** Von 900 überprüften Transistoren waren 30 defekt. Gesucht ist ein 95%-Konfidenzintervall für den Anteil  $p$  von defekten Transistoren in der Gesamtproduktion.

Wegen  $\gamma = 0,95$  folgt  $\beta = \frac{1-\gamma}{2} = 0,025$ . Man findet  $z_\beta = 1,96$  in einer Tabelle. Berechnet man  $I$ , so ergibt sich  $I = \left[ \frac{30}{900} - \frac{1,96}{2\sqrt{900}}, \frac{30}{900} + \frac{1,96}{2\sqrt{900}} \right] = [0,0006, 0,066]$ .

Zum Vergleich bestimmen wir das Konfidenzintervall  $K$ . Wir berechnen

$$Y = z_\beta \sqrt{\frac{X(n-X)}{n} + \frac{1}{4}z_\beta^2} = 1,96 \sqrt{\frac{30(900-30)}{900} + \frac{1}{4} \cdot 1,96^2} = 10,73$$

$$X + \frac{1}{2}z_\beta^2 = 30 + \frac{1}{2} \cdot 1,96^2 = 30,96$$

$$n + z_\beta^2 = 900 + 1,96^2 = 903,84$$

Setzt man das in die obige Formel für  $K$  ein, so erhält man

$$K = \left[ \frac{30,96-10,73}{903,84}, \frac{30,96+10,73}{903,84} \right] = [0,0224, 0,0461]. \quad \text{Das Intervall } I \text{ ist ca. dreimal so lang.}$$

#### 4. Statistische Tests für Prozentsätze

Statistische Tests verwendet man, um mittels einer Stichprobe eine Aussage über eine Grundgesamtheit nachzuweisen. Bei der Produktion von Glühlampen zum Beispiel kann man einen noch tolerierbaren Ausschussanteil  $p_0$  festlegen. Der Hersteller will dann nachweisen, dass der unbekannte Anteil  $p$  an defekten Glühlampen  $< p_0$  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 soll ein neues Medikament gegen dieselbe Krankheit auf den Markt gebracht werden. In einer klinischen Studie will man nachweisen, dass der noch unbekannte Prozentsatz  $p$  der behandelten Patienten, bei denen es hilft, größer als  $p_0$  ist.

Die Vorgangsweise beim statistischen Test ist folgende: Es liegt eine Grundgesamtheit (Gesamtproduktion der Glühlampen, Menge aller Patienten) vor, in der eine Eigenschaft (Glühbirne defekt, Medikament hilft) mit unbekannter relativer Häufigkeit  $p$  auftritt. Das Gegenteil der nachzuweisenden Aussage wird als Hypothese  $H_0$  und die nachzuweisende Aussage als Alternative  $H_1$  formuliert. Im Beispiel mit den Glühlampen 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 zufälligen Stichprobe getroffen, die aus der Grundgesamtheit gezogen wird. Daher wird diese Entscheidung auch nur mit einer gewissen Wahrscheinlichkeit, die groß sein soll, richtig ausfallen. Man gibt eine Irrtumswahrscheinlichkeit  $\alpha$  vor (typische Werte für  $\alpha$  sind 0,05 und 0,01) und legt den Stichprobenumfang  $n$  fest. Dann zieht man eine Stichprobe aus der Grundgesamtheit. Sei  $X$  die Anzahl (Häufigkeit), mit der die Eigenschaft in dieser zufälligen Stichprobe auftritt. Auf Grund dieser Anzahl  $X$  trifft man die Entscheidung, die Hypothese  $H_0$  zu verwerfen oder sie beizubehalten. Die Entscheidungsregel soll jedoch so sein, dass die Wahrscheinlichkeit, die Hypothese  $H_0$  zu verwerfen, obwohl sie richtig ist, höchstens  $\alpha$  beträgt.

Wir suchen eine entsprechende Entscheidungsregel. Die möglichen Werte von  $X$  sind die Anzahlen, mit der die Eigenschaft in der Stichprobe vom Umfang  $n$  auftritt. Daher hat  $X$  den Wertebereich  $S = \{0, 1, 2, \dots, n\}$ . Wir wählen eine Teilmenge  $V$  von  $S$  und verwerfen die Hypothese  $H_0$ , wenn sich nach dem Ziehen der Stichprobe herausstellt, dass  $X$  in  $V$  liegt. Man nennt  $V$  den Verwerfungsbereich (Ablehnungsbereich) des Tests. Aus den oben durchgeführten Überlegungen ergeben sich folgende Bedingungen:

- (1)  $V$  enthält die Werte aus  $S$ , die am meisten gegen  $H_0$  (und daher für  $H_1$ ) sprechen
- (2)  $V$  ist maximal, aber so, dass  $P(X \in V) \leq \alpha$  gilt, wenn  $H_0$  richtig ist

Da  $H_0$  verworfen wird, wenn  $X$  in  $V$  fällt, ist (1) klar. In (2) findet man die oben formulierte Bedingung. Die Wahrscheinlichkeit, die Hypothese  $H_0$  zu verwerfen, das ist  $P(X \in V)$ , soll  $\leq \alpha$  sein, wenn  $H_0$  richtig ist. Unter dieser Bedingung soll jedoch  $V$  möglichst groß sein. Das Risiko,  $H_0$  bei Richtigkeit zu verwerfen, kontrolliert man durch die Irrtumswahrscheinlichkeit  $\alpha$ . Das Risiko,  $H_0$  bei Unrichtigkeit nicht zu verwerfen, macht man dadurch möglichst klein, dass man  $V$  maximal wählt.

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

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

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

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

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

Der Satz ist bewiesen, wenn wir  $h'(p) \leq 0$  zeigen. Mit Hilfe der Produktregel kann man die Summanden in obiger Summe der Reihe nach differenzieren. Man erhält

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

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

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

Da  $h'(p) \leq 0$  gilt ist die Funktion  $h$  monoton fallend. Somit ist der Hilfssatz bewiesen.  $\square$

**Satz 42:** Sei  $p$  die unbekannte relative Häufigkeit, mit der eine Eigenschaft in der Grundgesamtheit auftritt. Der Stichprobenumfang sei  $n$ , die Irrtumswahrscheinlichkeit sei  $\alpha$  und der durchzuführende Test sei  $H_0 : p \leq p_0$  gegen  $H_1 : p > p_0$ . Bestimmt man das minimale  $k_0$ , sodass  $\Psi_{n,p_0}(k_0 - 1) \geq 1 - \alpha$  gilt, und wählt  $V = \{k_0, k_0 + 1, \dots, n\}$ , dann erfüllt  $V$  die Bedingungen (1) und (2) für diesen Test, ist also der Verwerfungsbereich.

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

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

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

**Beispiel 69:** Für den Test  $H_0 : p \leq 0,3$  gegen  $H_1 : p > 0,3$  ist der Verwerfungsbereich bei einem Stichprobenumfang  $n = 10$  und verschiedenen Irrtumswahrscheinlichkeiten gesucht.

Der Verwerfungsbereich für diesen Test ist  $V = \{k_0, k_0 + 1, \dots, 10\}$ . Um  $k_0$  zu bestimmen, erstellen wir eine Wertetabelle für die Abbildung  $k \mapsto \Psi_{10;0,3}(k)$ .

$k$	0	1	2	3	4	5	6	7	8	9	10
$\Psi_{10;0,3}(k)$	0,028	0,149	0,383	0,650	0,850	0,953	0,989	0,998	1,000	1,000	1,000

Nach Satz 42 ist  $k_0$  minimal zu wählen, sodass  $\Psi_{10;0,3}(k_0 - 1) \geq 1 - \alpha$  gilt. Man kann  $k_0$  jetzt leicht aus der Tabelle ablesen. Für  $\alpha = 0,05$  ist  $1 - \alpha = 0,95$  und  $k_0 - 1 = 5$ , also  $V = \{6, 7, 8, 9, 10\}$ . Für  $\alpha = 0,02$  ist  $1 - \alpha = 0,98$  und  $k_0 - 1 = 6$ , also  $V = \{7, 8, 9, 10\}$ . Für  $\alpha = 0,01$  ist  $1 - \alpha = 0,99$  und  $k_0 - 1 = 7$ , also  $V = \{8, 9, 10\}$ .

**Satz 43:** Sei  $p$  die unbekannte relative Häufigkeit, mit der eine Eigenschaft in der Grundgesamtheit auftritt. Der Stichprobenumfang sei  $n$ , die Irrtumswahrscheinlichkeit sei  $\alpha$  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  $V = \{0, 1, \dots, k_0\}$ , dann erfüllt  $V$  die Bedingungen (1) und (2) für diesen Test, ist also der Verwerfungsbereich.

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

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

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

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

**Beispiel 70:** 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  $\Psi_{400;0,5}(k_0) \approx \Phi\left(\frac{k_0 + \frac{1}{2} - 400 \cdot 0,5}{\sqrt{400 \cdot 0,5 \cdot 0,5}}\right) = \Phi\left(\frac{k_0 + \frac{1}{2} - 200}{10}\right)$  ist diese Ungleichung äquivalent zu  $\Phi\left(\frac{k_0 + \frac{1}{2} - 200}{10}\right) \leq 0,01 = \Phi(-2,33)$ . Da  $\Phi$  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  $V = \{0, 1, 2, \dots, 176\}$  der Verwerfungsbereich.

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

**Satz 44:** Die Irrtumswahrscheinlichkeit sei  $\alpha$ . Nach dem Ziehen einer Stichprobe vom Umfang  $n$  stellt man fest, dass die Eigenschaft mit Häufigkeit  $s$  in dieser vorkommt.

(a) Sei  $V$  der Verwerfungsbereich für den Test  $H_0 : p \leq p_0$  gegen  $H_1 : p > p_0$ . Dann gilt  $s \in V \Leftrightarrow \Psi_{n,p_0}(s-1) \geq 1 - \alpha$ .

(b) Sei  $V$  der Verwerfungsbereich für den Test  $H_0 : p \geq p_0$  gegen  $H_1 : p < p_0$ . Dann gilt  $s \in V \Leftrightarrow \Psi_{n,p_0}(s) \leq \alpha$ .

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

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

Wendet man Satz 44 in Beispielen an, so kann man entweder mit der Binomialverteilung rechnen oder mit der Approximation durch die Normalverteilung. Dazu je ein Beispiel.



**Beispiel 71:** Bei Glühlampen wird ein Ausschussanteil von 0,015 toleriert. Ein Kunde möchte nachweisen, dass er größer ist. Aus einer Lieferung Glühlampen zieht er eine Stichprobe vom Umfang  $n = 20$  und findet  $s = 2$  defekte Glühlampen. Wird die Hypothese  $H_0 : p \leq 0,015$  bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$  abgelehnt?

Wir verwenden Satz 44 (a). Es wird genau dann verworfen, wenn  $\Psi_{n,p_0}(s-1) \geq 1 - \alpha$  ist, also  $\Psi_{20;0,015}(1) \geq 0,95$ . Es gilt  $\Psi_{20;0,015}(1) = \binom{20}{0}0,985^{20} + \binom{20}{1}0,015 \cdot 0,985^{19} = 0,964$ . Wegen  $0,964 \geq 0,95$  wird die Hypothese  $H_0$  verworfen. Wir können es daher als erwiesen ansehen, dass der Anteil an defekten Glühlampen über der Toleranzgrenze liegt. Die Wahrscheinlichkeit für einen Irrtum ist dabei  $\leq 0,05$ .

**Beispiel 72:** Nach der letzten Statistik rauchen 40% der Männer. Nach einer Anti-raucherkampagne 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 44 (b). Es wird genau dann verworfen, wenn  $\Psi_{n,p_0}(s) \leq \alpha$  ist, also  $\Psi_{1000;0,4}(366) \leq 0,05$ . Wir verwenden die Approximation durch die Normalverteilung. Es gilt  $\Psi_{1000;0,4}(366) \approx \Phi\left(\frac{366 + \frac{1}{2} - 1000 \cdot 0,4}{\sqrt{1000 \cdot 0,4 \cdot 0,6}}\right) = \Phi(-2,16) = 1 - \Phi(2,16) = 0,016$ . Wegen  $0,016 \leq 0,05$  wird  $H_0$  verworfen. Bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$  kann es als erwiesen gelten, dass der Raucheranteil durch die Kampagne kleiner geworden ist.

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

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

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

**Beispiel 73:** Wir prüfen einen Würfel. Sei  $p$  die unbekannte Wahrscheinlichkeit (relative Häufigkeit), mit der 6 auftritt. Wir testen  $H_0 : p = \frac{1}{6}$  gegen  $H_1 : p \neq \frac{1}{6}$ . Wir würfeln 80 Mal. Gesucht ist der Verwerfungsbereich bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$ .

Wir bestimmen  $k_0$  maximal, sodass  $\Psi_{n,p_0}(k_0) \leq \frac{\alpha}{2}$  erfüllt ist, und  $k_1$  minimal, sodass  $\Psi_{n,p_0}(k_1 - 1) \geq 1 - \frac{\alpha}{2}$  gilt. Mit Hilfe der Approximation durch die Normalverteilung erhalten wir  $\Psi_{80, \frac{1}{6}}(k_0) \approx \Phi\left(\frac{k_0 + \frac{1}{2} - 80 \cdot \frac{1}{6}}{\sqrt{80 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = \Phi\left(\frac{k_0 - 12,83}{3,33}\right)$ . 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  $\Psi_{80, \frac{1}{6}}(k_1 - 1) \approx \Phi\left(\frac{k_1 - \frac{1}{2} - 80 \cdot \frac{1}{6}}{\sqrt{80 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = \Phi\left(\frac{k_1 - 13,83}{3,33}\right)$ . 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 Würfeln auftritt, in  $V$ , dann kann man bei Irrtumswahrscheinlichkeit 0,05 schließen, dass der Würfel die Augenzahl 6 nicht mit Wahrscheinlichkeit  $\frac{1}{6}$  liefert.

Auch beim zweiseitigen Test kann man entscheiden, ob eine Zahl  $s$  in den Verwerfungsbereich  $V$  fällt, ohne diesen auszurechnen. Da sich  $V$  aus den Verwerfungsbereichen von zwei einseitigen Tests zusammensetzt, erhalten wir mit Hilfe von Satz 44

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

Das verwenden wir im folgenden Beispiel.

**Beispiel 74:** Bei 900 Münzenwürfen tritt 473 Mal „Kopf“ auf. Ist die Münze fair?

Wir testen  $H_0 : p = \frac{1}{2}$  gegen  $H_1 : p \neq \frac{1}{2}$  bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$ . Um zu überprüfen, ob  $s = 473$  im Verwerfungsbereich liegt, verwenden wir obige Bedingungen und Approximation durch die Normalverteilung.

$$\Psi_{n,p_0}(s) \approx \Phi\left(\frac{s + \frac{1}{2} - np_0}{\sqrt{np_0(1-p_0)}}\right) = \Phi\left(\frac{473 + \frac{1}{2} - 900 \cdot \frac{1}{2}}{\sqrt{900 \cdot \frac{1}{2} \cdot \frac{1}{2}}}\right) = \Phi(1,57) = 0,942$$

$$\Psi_{n,p_0}(s-1) \approx \Phi\left(\frac{s - \frac{1}{2} - np_0}{\sqrt{np_0(1-p_0)}}\right) = \Phi\left(\frac{473 - \frac{1}{2} - 900 \cdot \frac{1}{2}}{\sqrt{900 \cdot \frac{1}{2} \cdot \frac{1}{2}}}\right) = \Phi(1,50) = 0,933$$

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. Wir können keine Aussage machen.

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

Wir betrachten zuerst den Test  $H_0 : p \leq p_0$  gegen  $H_1 : p > p_0$ , der den Verwerfungsbereich  $V = \{k_0, k_0 + 1, \dots, n\}$  hat. Wir geben noch ein  $p_1 > p_0$  und ein  $\tilde{\alpha}$  vor und verlangen

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

(ii)  $P(X \in V) \geq 1 - \tilde{\alpha}$  wenn  $p \geq p_1 \Leftrightarrow P(X \in \{0, \dots, k_0 - 1\}) \leq \tilde{\alpha}$  wenn  $p \geq p_1$

Die Bedingung (i) wird bei jedem Test verlangt. Sie besagt, dass  $H_0$  bei Richtigkeit nur mit Wahrscheinlichkeit  $\leq \alpha$  verworfen wird. Es kommt (ii) dazu: Wenn  $p \geq p_1$  gilt (diese Hypothese ist nur etwas stärker als  $H_1$ ), dann folgt mit Wahrscheinlichkeit  $\geq 1 - \tilde{\alpha}$   $X \in V$  und die Hypothese  $H_0$  wird verworfen. Dies entspricht einem zweiten Test mit  $\tilde{H}_0 : p \geq p_1$  gegen  $\tilde{H}_1 : p < p_1$  mit Irrtumswahrscheinlichkeit  $\tilde{\alpha}$  und Verwerfungsbereich  $\tilde{V} = \{0, \dots, k_0 - 1\} = \{0, \dots, n\} \setminus V$ . Wir bestimmen einen (möglichst kleinen) Stichprobenumfang  $n$  und  $k_0$  so, dass (i) und (ii) gelten.

Nach dem Hilfssatz ist  $P(X \in V) = 1 - \Psi_{n,p}(k_0 - 1)$  monoton wachsend in  $p$ . Daher 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  $\Psi_{n,p_1}(k_0 - 1) \leq \tilde{\alpha}$ .

Die Verwerfungsbereiche  $V$  und  $\tilde{V}$  bilden eine Zerlegung von  $\{0, \dots, n\}$ , daher gilt für diesen Test: Endet die Durchführung des Tests mit dem Verwerfen der Hypothese  $H_0$ , dann schließen wir, dass  $p > p_0$  gilt. Die Wahrscheinlichkeit für einen Irrtum ist dann  $\leq \alpha$  wegen (i).

Endet die Durchführung des Tests hingegen mit dem Nichtverwerfen der Hypothese  $H_0$ , dann wird die Hypothese  $\tilde{H}_0$  verworfen und wir schließen  $p < p_1$ . Die Wahrscheinlichkeit für einen Irrtum ist dann  $\leq \tilde{\alpha}$ , da wegen (ii) ja  $P(X \notin V) \leq \tilde{\alpha}$  für  $p \geq p_1$  gilt. Wir haben also in jedem Fall eine Schlussfolgerung.

Wir betrachten auch noch den Test  $H_0 : p \geq p_0$  gegen  $H_1 : p < p_0$  mit dem Verwerfungsbereich  $V = \{0, 1, \dots, k_0\}$ . Hier geben wir ein  $p_1 < p_0$  und ein  $\tilde{\alpha}$  vor und verlangen

- (i)  $P(X \in V) \leq \alpha$  wenn  $p \geq p_0$
- (ii)  $P(X \in V) \geq 1 - \tilde{\alpha}$  wenn  $p \leq p_1$

Die Bedingungen (i) und (ii) sind äquivalent zu  $\Psi_{n,p_0}(k_0) \leq \alpha$  und  $\Psi_{n,p_1}(k_0) \geq 1 - \tilde{\alpha}$ . Wenn nun  $X \in V$  eintritt, wird die Hypothese  $H_0$  verworfen und wir schließen  $p < p_0$ . Wenn  $X \notin V$  eintritt und die Hypothese  $H_0$  nicht verworfen wird, dann schließen wir  $p > p_1$ .

**Beispiel 75:** Sei  $p_0 = 0,1$  der Anteil an defekten Glühlampen, der toleriert wird. Es soll  $H_0 : p \leq p_0$  gegen  $H_1 : p > p_0$  bei Irrtumswahrscheinlichkeit  $\alpha = 0,05$  getestet werden. Stichprobenumfang  $n$  und Verwerfungsbereich  $V = \{k_0, k_0 + 1, \dots, n\}$  sind so zu bestimmen, dass  $H_0$  mit Wahrscheinlichkeit  $\geq 0,95$  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  $\Psi_{n,p}(k_0 - 1) \approx \Phi\left(\frac{k_0 - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$  erhalten wir die zu (i) und (ii) äquivalenten Ungleichungen

$$\begin{aligned} \Phi\left(\frac{k_0 - \frac{1}{2} - 0,1n}{\sqrt{0,1 \cdot 0,9 \cdot n}}\right) &\geq 0,95 && \text{und} && \Phi\left(\frac{k_0 - \frac{1}{2} - 0,15n}{\sqrt{0,15 \cdot 0,85 \cdot n}}\right) &\leq 0,05 \\ \frac{k_0 - \frac{1}{2} - 0,1n}{\sqrt{0,1 \cdot 0,9 \cdot n}} &\geq 1,65 && \text{und} && \frac{k_0 - \frac{1}{2} - 0,15n}{\sqrt{0,15 \cdot 0,85 \cdot n}} &\leq -1,65 \\ k_0 - \frac{1}{2} - 0,1n &\geq 0,495\sqrt{n} && \text{und} && k_0 - \frac{1}{2} - 0,15n &\leq -0,59\sqrt{n} \\ \frac{1}{2} + 0,1n + 0,495\sqrt{n} &\leq k_0 &\leq &\frac{1}{2} + 0,15n - 0,59\sqrt{n} \end{aligned}$$

Wir müssen ein möglichst kleines  $n$  und ein  $k_0$  finden, sodass diese Ungleichungen erfüllt sind. Wegen  $\frac{1}{2} + 0,1n + 0,495\sqrt{n} \leq \frac{1}{2} + 0,15n - 0,59\sqrt{n}$  folgt  $1,085\sqrt{n} \leq 0,05n$  und  $n \geq 470,89$ . Wir rechnen für einige  $n$  nach:

$n$ :	470	480	490	500	510	520
$\frac{1}{2} + 0,1n + 0,495\sqrt{n}$ :	58,23	59,34	60,46	61,57	62,68	63,79
$\frac{1}{2} + 0,15n - 0,59\sqrt{n}$ :	58,21	59,57	60,94	62,31	63,67	65,05

Da  $k_0$  zwischen diesen Werten liegen muss, können wir  $n = 500$  und  $k_0 = 62$ , also  $V = \{62, 63, \dots, 500\}$  wählen (eine andere Lösung ist  $n = 476$  und  $k_0 = 59$ ). Zum Vergleich: Wenn wir für  $p_1$  statt 0,15 den Wert 0,12 wählen, dann ergibt sich die Lösung  $n = 2669$  und  $k_0 = 293$ .

### 5. Tests für den Erwartungswert einer Normalverteilung

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

$$H_0 : \mu \leq \mu_0 \quad \text{gegen} \quad H_1 : \mu > \mu_0$$

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

- (1)  $V$  enthält die reellen Zahlen, die am meisten gegen  $H_0$  (und daher für  $H_1$ ) sprechen
- (2)  $V$  ist maximal, aber so, dass  $P(M \in V) \leq \alpha$  gilt, wenn  $H_0$  richtig ist

**Satz 45:** Der Stichprobenumfang sei  $n$  und die Irrtumswahrscheinlichkeit sei  $\alpha$ . Für den Test  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  ist der Verwerfungsbereich dann

$$V = [\mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha), \infty)$$

Der Test  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$  hat den Verwerfungsbereich

$$V = (-\infty, \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)] = (-\infty, \mu_0 - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha)]$$

Der zweiseitige Test  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  hat den Verwerfungsbereich

$$V = (-\infty, \mu_0 - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2})] \cup [\mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \frac{\alpha}{2}), \infty)$$

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

Wegen (2) ist der Verwerfungsbereich  $V$  möglichst groß und somit  $c$  möglichst klein zu wählen, sodass  $P(M \in V) = P(M \geq c) \leq \alpha$  gilt, wenn  $H_0$  erfüllt ist, das heißt für alle  $\mu \leq \mu_0$ . Im Beweis von Satz 39 wurde gezeigt, dass  $\sqrt{n}\frac{M-\mu}{\sigma}$  die  $N(0, 1)$ -Verteilung hat. Es folgt  $P(M \geq c) = P(\sqrt{n}\frac{M-\mu}{\sigma} \geq \sqrt{n}\frac{c-\mu}{\sigma}) = 1 - \Phi(\sqrt{n}\frac{c-\mu}{\sigma})$ . Somit gilt  $P(M \geq c) \leq \alpha$  für alle  $\mu \leq \mu_0$  genau dann, wenn  $\Phi(\sqrt{n}\frac{c-\mu}{\sigma}) \geq 1 - \alpha$  für alle  $\mu \leq \mu_0$  gilt. Das ist wieder äquivalent zu  $\Phi(\sqrt{n}\frac{c-\mu_0}{\sigma}) \geq 1 - \alpha$ , da  $\Phi$  monoton wachsend ist. Unter dieser Bedingung ist  $c$  minimal zu wählen. Daher ist  $c$  die Lösung der Gleichung  $\Phi(\sqrt{n}\frac{c-\mu_0}{\sigma}) = 1 - \alpha$ . Löst man diese Gleichung, so ergibt sich  $c = \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha)$ . Damit ist  $V$  berechnet.

Der Beweis für den Test  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$  erfolgt analog.

Der Verwerfungsbereich des zweiseitigen Tests  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  ist durch große Abstände von  $M$  zu  $\mu_0$  gekennzeichnet. Wegen  $\mu = \mu_0 \Leftrightarrow \mu \leq \mu_0 \wedge \mu \geq \mu_0$  setzt sich der Verwerfungsbereich von  $H_0$  aus den Verwerfungsbereichen von zwei einseitigen Tests mit Irrtumswahrscheinlichkeit  $\frac{\alpha}{2}$  zusammen.  $\square$

**Beispiel 76:** Die Füllgewichte von Waschmittelpackungen einer bestimmten Sorte seien  $N(\mu, \sigma)$ -verteilt. Die Genauigkeit der Abfüllmaschine wird mit 0,02 kg angegeben. Es soll  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  mit  $\mu_0 = 1$  kg getestet werden. Der Stichprobenumfang ist 10. Gesucht ist der Verwerfungsbereich  $V$  für die Irrtumswahrscheinlichkeit  $\alpha = 0,05$ .

Wir finden  $\Phi^{-1}(1 - \alpha) = \Phi^{-1}(0,95) = 1,65$  in einer Tabelle. Es folgt  $\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha) = \frac{0,02 \cdot 1,65}{\sqrt{10}} = 0,0104$ . Somit ist  $V = [1,0104, \infty)$  der Verwerfungsbereich. Für die Stichprobe aus Beispiel 59 haben wir  $M = 1,01$  berechnet. Wegen  $M \notin V$  wird die Hypothese  $H_0$  nicht verworfen. Der Nachweis, dass die durchschnittliche Füllmenge mindestens 1 kg ist, ist nicht gelungen. (Man sollte einen größeren Stichprobenumfang wählen.)

## VI. Anhang: Vorkenntnisse aus der Analysis

Es werden einige Resultate aus der Analysis zusammengestellt, die in der Stochastik verwendet werden. Dabei wird nur die Analysis in einer Variablen vorausgesetzt.

### 1. Reihen

In der Analysis wird zu einer Folge  $u_0, u_1, u_2, u_3, \dots$  von reellen Zahlen die Reihe  $\sum_{j=0}^{\infty} u_j$  als Grenzwert der Partialsummen  $\sum_{j=0}^n u_j$  für  $n \rightarrow \infty$  definiert. In der Wahrscheinlichkeitstheorie hat man Summen  $\sum_{i \in S} w(i)$ , wobei  $S$  eine endliche oder abzählbare Menge ist und  $w(i) \geq 0$  für alle  $i \in S$  gilt. Ist  $S$  endlich, dann ist diese Summe eindeutig, da man ja in beliebiger Reihenfolge addieren kann. Für abzählbares  $S$  muss man eine Anordnung der Reihenglieder wählen, um die Summe  $\sum_{i \in S} w(i)$  als Grenzwert der Partialsummen berechnen zu können. Es stellt sich die Frage, ob man da nicht je nach Anordnung verschiedene Resultate erhalten kann.

**Satz 46:** Sei  $S$  abzählbar und  $w(i) \geq 0$  für alle  $i \in S$ . Dann hat  $\sum_{i \in S} w(i)$  immer denselben Wert  $W$ , wie immer man die Reihe anordnet. Es ist jedoch auch  $W = \infty$  möglich. Ist  $W < \infty$ , dann existiert für jedes  $\varepsilon > 0$  eine endliche Teilmenge  $S_\varepsilon$  von  $S$  mit  $W - \varepsilon < \sum_{i \in S_\varepsilon} w(i) \leq W$ .

**Beweis:** Sei  $a_0, a_1, a_2, a_3, \dots$  irgendeine Anordnung der Zahlen  $w(i)$  mit  $i \in S$ . Zahlen, die öfter vorkommen, müssen natürlich auch in der Anordnung mit derselben Häufigkeit auftreten. Sei  $A_n = \sum_{j=0}^n a_j$  die  $n$ -te Partialsumme. Da  $a_j \geq 0$  für alle  $j \geq 0$  gilt, ist die Folge  $(A_n)_{n \geq 0}$  monoton wachsend. Ist sie beschränkt, dann existiert der Grenzwert  $\lim_{n \rightarrow \infty} A_n$ , den wir mit  $W$  bezeichnen. Ist sie unbeschränkt, dann schreiben wir  $\lim_{n \rightarrow \infty} A_n = \infty$  und setzen  $W = \infty$ . Es gilt auch  $A_n \leq W$  für alle  $n \geq 0$ .

Sei  $b_0, b_1, b_2, b_3, \dots$  eine andere Anordnung der Zahlen  $w(i)$  mit  $i \in S$  und  $B_n = \sum_{j=0}^n b_j$  die  $n$ -te Partialsumme. Wie oben existiert  $B = \lim_{n \rightarrow \infty} B_n$ , kann aber auch  $\infty$  sein. Wieder gilt  $B_n \leq B$  für alle  $n \geq 0$ .

Sei  $n \geq 0$  beliebig. Da die Zahlen  $b_0, b_1, \dots, b_n$  irgendwo in der Folge  $a_0, a_1, a_2, a_3, \dots$  vorkommen, existiert eine Zahl  $m$  mit  $\sum_{j=0}^n b_j \leq \sum_{j=0}^m a_j$ , das heißt  $B_n \leq A_m$ , woraus  $B_n \leq W$  folgt. Da das für alle  $n \geq 0$  gilt, erhalten wir auch  $\lim_{n \rightarrow \infty} B_n \leq W$ , das heißt  $B \leq W$ . Ganz analog beweist man  $W \leq B$ . Damit ist  $B = W$  gezeigt. Wie man die Reihenglieder auch anordnet, man erhält immer  $W$  als Grenzwert.

Sei  $W < \infty$  und  $\varepsilon > 0$ . Wegen  $\lim_{n \rightarrow \infty} A_n = W$  existiert ein  $n_0$  mit  $A_{n_0} > W - \varepsilon$ . Sei  $S_\varepsilon$  die Teilmenge von  $S$ , sodass  $a_0, a_1, \dots, a_{n_0}$  gerade die Zahlen  $w(i)$  mit  $i \in S_\varepsilon$  sind. Dann gilt  $\sum_{i \in S_\varepsilon} w(i) = \sum_{j=0}^{n_0} a_j > W - \varepsilon$ . Wir haben bereits gesehen, dass  $A_{n_0} \leq W$  gilt, das heißt  $\sum_{i \in S_\varepsilon} w(i) \leq W$ .  $\square$

### 2. Uneigentliches Integral

Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine Funktion, die auf jedem beschränkten Intervall integrierbar ist. Wir definieren die uneigentlichen Integrale mit einem unbeschränkten Integrationsintervall durch folgende Limiten, sofern diese existieren:

$$\begin{aligned} \int_a^\infty f(x) dx &= \lim_{n \rightarrow \infty} \int_a^n f(x) dx \\ \int_{-\infty}^b f(x) dx &= \lim_{n \rightarrow \infty} \int_{-n}^b f(x) dx \end{aligned}$$

Das uneigentliche Integral über die gesamten reellen Zahlen definieren wir durch

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^c f(x) dx + \int_c^{\infty} f(x) dx$$

mit einer beliebigen Zahl  $c \in \mathbb{R}$ . Diese Definition hängt nicht von der Zahl  $c \in \mathbb{R}$  ab, denn bei einer anderen Wahl  $\tilde{c} \in \mathbb{R}$  mit  $\tilde{c} > c$  wird das Integral  $\int_c^{\tilde{c}} f(x) dx$  bei  $\int_{-\infty}^c f(x) dx$  addiert und bei  $\int_c^{\infty} f(x) dx$  subtrahiert um  $\int_{-\infty}^{\tilde{c}} f(x) dx + \int_{\tilde{c}}^{\infty} f(x) dx$  zu erhalten. Wenn zusätzlich  $f \geq 0$  gilt, dann bilden die Integrale monoton wachsenden Folgen in  $n$  und die Limiten der Integrale existieren immer, können aber  $\infty$  sein.

Die üblichen Rechenregeln für Integrale bleiben dabei erhalten. Ist zum Beispiel  $a < u$ , dann gilt  $\int_a^n f(x) dx = \int_a^u f(x) dx + \int_u^n f(x) dx$ . Lässt man  $n$  gegen  $\infty$  gehen, dann erhält man  $\int_a^{\infty} f(x) dx = \int_a^u f(x) dx + \int_u^{\infty} f(x) dx$ . Sind  $f$  und  $g$  Funktionen mit Werten in  $\mathbb{R}_0^+$ , dann gilt  $\int_a^n (f(x) + g(x)) dx = \int_a^n f(x) dx + \int_a^n g(x) dx$ . Lässt man  $n$  gegen  $\infty$  gehen, dann erhält man  $\int_a^{\infty} (f(x) + g(x)) dx = \int_a^{\infty} f(x) dx + \int_a^{\infty} g(x) dx$ .

Eine andere Art von uneigentlichen Integralen erhält man, wenn  $f : \mathbb{R} \rightarrow \mathbb{R}$  eine Polstelle hat. Hat  $f$  eine Polstelle bei  $b$  dann definiert man  $\int_a^b f(x) dx = \lim_{\varepsilon \downarrow 0} \int_a^{b-\varepsilon} f(x) dx$ . Hat  $f$  eine Polstelle bei  $a$  dann definiert man  $\int_a^b f(x) dx = \lim_{\varepsilon \downarrow 0} \int_{a+\varepsilon}^b f(x) dx$ . Wenn  $f \geq 0$  gilt, dann sind  $\varepsilon \mapsto \int_a^{b-\varepsilon} f(x) dx$  und  $\varepsilon \mapsto \int_{a+\varepsilon}^b f(x) dx$  monoton fallende Abbildungen. Daher existieren die Grenzwerte, können aber  $\infty$  sein. Gilt  $a < c < b$  und hat  $f$  eine Polstelle bei  $c$ , dann schreibt man  $\int_a^b f(x) dx$  als  $\int_a^c f(x) dx + \int_c^b f(x) dx$  und hat dann die Polstelle wieder in den Endpunkten der Integrationsintervalle.

Die üblichen Rechenregeln für Integrale bleiben dabei erhalten. Hat man zum Beispiel  $a < u < b$  und eine Polstelle bei  $b$ , dann gilt  $\int_a^{b-\varepsilon} f(x) dx = \int_a^u f(x) dx + \int_u^{b-\varepsilon} f(x) dx$ . Lässt man  $\varepsilon$  gegen 0 gehen, dann erhält man  $\int_a^b f(x) dx = \int_a^u f(x) dx + \int_u^b f(x) dx$ .

**Satz 47:** Sei  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  eine integrierbare Funktion. Dann gilt  $\lim_{\varepsilon \rightarrow 0} \int_{b-\varepsilon}^b f(x) dx = 0$ .

**Beweis:** Hat  $f$  eine Polstelle bei  $b$ , dann ist das Integral uneigentlich. Wir wählen  $a < b$ . Nach Definition des uneigentlichen Integrals gilt dann  $\lim_{\varepsilon \rightarrow 0} \int_a^{b-\varepsilon} f(x) dx = \int_a^b f(x) dx$ . Wegen  $\int_{b-\varepsilon}^b f(x) dx = \int_a^b f(x) dx - \int_a^{b-\varepsilon} f(x) dx$  erhalten wir  $\lim_{\varepsilon \rightarrow 0} \int_{b-\varepsilon}^b f(x) dx = 0$ . Hat  $f$  keine Polstelle bei  $b$ , dann existiert ein  $\varepsilon_0 > 0$  und eine Konstante  $c$  mit  $f(x) \leq c$  für alle  $x \in [b - \varepsilon_0, b]$ . Für  $\varepsilon < \varepsilon_0$  erhalten wir dann  $0 \leq \int_{b-\varepsilon}^b f(x) dx \leq \varepsilon c$ . Daraus folgt  $\lim_{\varepsilon \rightarrow 0} \int_{b-\varepsilon}^b f(x) dx = 0$ .  $\square$

### 3. Volumen

Für Volumsberechnungen führt man üblicherweise ein mehrdimensionales Integral ein. Man muss aber nicht unbedingt diesen Aufwand betreiben, sondern kann auch mit dem eindimensionalen Integral auskommen.

Sei  $E_x$  die Ebene im  $\mathbb{R}^3$  durch den Punkt  $(x, 0, 0)$ , die senkrecht auf die  $x$ -Achse steht. Der Körper, dessen Volumen  $V$  wir berechnen wollen, liege im  $\mathbb{R}^3$  zwischen den Ebenen  $E_a$  und  $E_b$ . Für  $x \in [a, b]$  sei  $u(x)$  die Fläche des Schnittes der Ebene  $E_x$  mit dem Körper. Wir nehmen an, dass  $u : [a, b] \rightarrow \mathbb{R}$  eine integrierbare Funktion ist.

Sei  $a = x_0 < x_1 < \dots < x_n = b$  eine beliebige Zerlegung des Intervalls  $[a, b]$ . Sei  $V_j$  das Volumen des Teils des Körpers, der zwischen den Ebenen  $E_{x_{j-1}}$  und  $E_{x_j}$  liegt. Es gilt

$$(x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} u(x) \leq V_j \leq (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} u(x)$$

Wegen  $\sum_{j=1}^n V_j = V$  erhalten wir dann

$$\sum_{j=1}^n (x_j - x_{j-1}) \inf_{x \in [x_{j-1}, x_j]} u(x) \leq V \leq \sum_{j=1}^n (x_j - x_{j-1}) \sup_{x \in [x_{j-1}, x_j]} u(x)$$

Links steht die Untersumme und rechts die Obersumme des Integrals  $\int_a^b u(x) dx$  zur Zerlegung  $a = x_0 < x_1 < \dots < x_n = b$ . Lässt man die Gitterweite der Zerlegung gegen 0 gehen, dann konvergieren beide Summen gegen dieses Integral. Wir erhalten daher

$$V = \int_a^b u(x) dx$$

und haben damit eine Formel für das Volumen des Körpers gefunden.

Kennt man die Schnittflächen  $u(x)$  des Körpers, so kann man das Volumen berechnen. Bei einem Drehkörper, der durch die Rotation des Flächenstücks unter der stetigen Funktion  $h : [a, b] \rightarrow \mathbb{R}_0^+$  um die  $x$ -Achse entsteht, ist das einfach. Für  $x \in [a, b]$  ist die Fläche des Schnittes der Ebene  $E_x$  mit dem Drehkörper gleich  $u(x) = \pi h(x)^2$ . Deshalb ist

$$V = \pi \int_a^b h(x)^2 dx$$

die Formel für das Volumen des Drehkörpers.

Hier geht es jedoch darum, das Volumen über einem Gebiet  $B \subseteq \mathbb{R}^2$  unter einer Funktion  $g : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$  zu berechnen. Wir schreiben  $\int_B g(x, y) d(x, y)$  für dieses Volumen analog zum eindimensionalen Integral  $\int_{[a,b]} f(x) dx = \int_a^b f(x) dx$ , das die Fläche über dem Intervall  $[a, b]$  unter der Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  angibt. Es gelten entsprechende Rechenregeln, wie zum Beispiel  $\int_{B \cup C} g(x, y) d(x, y) = \int_B g(x, y) d(x, y) + \int_C g(x, y) d(x, y)$  für disjunkte Teilmengen  $B$  und  $C$  des  $\mathbb{R}^2$ , da sich Volumina über disjunkten Bereichen addieren.

Um mit obiger Methode das Volumen  $\int_B g(x, y) d(x, y)$  über einem Gebiet  $B \subseteq \mathbb{R}^2$  unter einer Funktion  $g : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$  zu berechnen, müssen wir annehmen, dass  $B$  ein sogenannter Normalbereich ist. Man nennt  $B$  einen Normalbereich, wenn ein Intervall  $[a, b]$  und stetige Funktionen  $\varphi : [a, b] \rightarrow \mathbb{R}$  und  $\psi : [a, b] \rightarrow \mathbb{R}$  mit  $\varphi \leq \psi$  existieren, sodass

$$B = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \varphi(x) \leq y \leq \psi(x)\}$$

gilt. Die Fläche  $u(x)$  des Schnittes der Ebene  $E_x$  mit dem Körper über  $B$  und unter  $g$  lässt sich dann durch das Integral  $\int_{\varphi(x)}^{\psi(x)} g(x, y) dy$  berechnen. Wir erhalten also

$$\int_B g(x, y) d(x, y) = \int_a^b u(x) dx = \int_a^b \int_{\varphi(x)}^{\psi(x)} g(x, y) dy dx$$

als Volumen über dem Gebiet  $B$  unter der Funktion  $g : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$ .

Wir können bei allen diesen Überlegungen die Rolle der Koordinaten vertauschen. Lässt sich der Bereich  $B$  nicht nur als  $B = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \varphi(x) \leq y \leq \psi(x)\}$  schreiben, sondern auch als  $B = \{(x, y) \in \mathbb{R}^2 : c \leq y \leq d, \varrho(y) \leq x \leq \chi(y)\}$  für ein Intervall  $[c, d]$  und stetige Funktionen  $\varrho : [c, d] \rightarrow \mathbb{R}$  und  $\chi : [c, d] \rightarrow \mathbb{R}$  mit  $\varrho \leq \chi$ , dann ergibt sich wie oben  $\int_B g(x, y) d(x, y) = \int_c^d \int_{\varrho(y)}^{\chi(y)} g(x, y) dx dy$ . Insbesondere haben wir

$$\int_a^b \int_{\varphi(x)}^{\psi(x)} g(x, y) dy dx = \int_c^d \int_{\varrho(y)}^{\chi(y)} g(x, y) dx dy$$

erhalten. Ist  $B$  das Rechteck  $[a, b] \times [c, d]$ , dann wird das zu

$$(*) \quad \int_a^b \int_c^d g(x, y) \, dy \, dx = \int_c^d \int_a^b g(x, y) \, dx \, dy$$

Diese Formel ist manchmal nützlich.

**Approximation durch Rechteckmengen:** Das Integral  $\int_B g(x, y) \, d(x, y)$  soll durch Integrale über Mengen approximiert werden, die Vereinigung von endlich vielen Rechtecken sind, wobei  $B$  ein Normalbereich  $\{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \varphi(x) \leq y \leq \psi(x)\}$  ist.

Sei  $c$  eine obere Schranke für die Funktion  $g$  und  $\varepsilon > 0$  beliebig. Wir finden eine Zerlegung  $a = x_0 < x_1 < \dots < x_n = b$  des Intervalls  $[a, b]$  mit folgenden Eigenschaften: Für  $1 \leq j \leq n$  existieren reelle Zahlen  $p_j$  und  $q_j$ , sodass  $p_j < \varphi(x) < q_j$  für alle  $x \in (x_{j-1}, x_j]$  und  $\sum_{j=1}^n (q_j - p_j)(x_j - x_{j-1}) < \frac{\varepsilon}{2c}$  gilt. Ebenso existieren reelle Zahlen  $u_j$  und  $v_j$ , sodass  $u_j < \psi(x) < v_j$  für alle  $x \in (x_{j-1}, x_j]$  und  $\sum_{j=1}^n (v_j - u_j)(x_j - x_{j-1}) < \frac{\varepsilon}{2c}$  gilt.

Wir definieren Rechtecke. Für  $1 \leq j \leq n$  sei  $C_j = (x_{j-1}, x_j] \times (q_j, u_j]$ , wenn  $q_j < u_j$  ist, und  $C_j = \emptyset$ , wenn  $q_j \geq u_j$  ist. Für  $2 \leq j \leq n$  sei  $D_j = (x_{j-1}, x_j] \times (p_j, v_j]$  und es sei  $D_1 = (a - \delta, x_1] \times (p_1, v_1]$ , wobei  $\delta > 0$  noch geeignet gewählt wird. Schließlich setzen wir  $C = \bigcup_{j=1}^n C_j$  und  $D = \bigcup_{j=1}^n D_j$ , wobei jede dieser Vereinigungen eine Vereinigung von disjunkten Rechtecken ist.

Es gilt dann  $C \subseteq B \subseteq D$ . Weiters gilt  $\int_C g(x, y) \, d(x, y) = \sum_{j=1}^n \int_{C_j} g(x, y) \, d(x, y)$  und  $\int_D g(x, y) \, d(x, y) = \sum_{j=1}^n \int_{D_j} g(x, y) \, d(x, y)$ . Damit erhalten wir

$$\begin{aligned} \int_D g(x, y) \, d(x, y) - \int_C g(x, y) \, d(x, y) &= \sum_{j=1}^n \int_{D_j \setminus C_j} g(x, y) \, d(x, y) \\ &\leq c \sum_{j=1}^n (q_j - p_j)(x_j - x_{j-1}) + c \sum_{j=1}^n (v_j - u_j)(x_j - x_{j-1}) + c\delta(v_1 - p_1) \end{aligned}$$

Da  $c \sum_{j=1}^n (q_j - p_j)(x_j - x_{j-1}) + c \sum_{j=1}^n (v_j - u_j)(x_j - x_{j-1}) < \varepsilon$  gilt, können wir  $\delta > 0$  so wählen, dass auch  $\int_D g(x, y) \, d(x, y) - \int_C g(x, y) \, d(x, y) < \varepsilon$  gilt.

**Uneigentliches Integral:** Wir können auch über ein unbeschränktes Gebiet  $B$  integrieren. Für  $n \geq 1$  sei  $Q_n$  das Quadrat  $[-n, n] \times [-n, n]$ . Wir definieren dann

$$\int_B g(x, y) \, d(x, y) = \lim_{n \rightarrow \infty} \int_{B \cap Q_n} g(x, y) \, d(x, y)$$

Wegen  $g \geq 0$  ist die rechte Seite monoton wachsend in  $n$ . Daher existiert der Grenzwert. Ist  $B$  ein Normalbereich  $\{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \varphi(x) \leq y \leq \psi(x)\}$ , wobei die Grenzen jetzt auch unendlich sein können, dann ergibt sich für  $B \cap Q_n$  der beschränkte Normalbereich  $\{(x, y) \in \mathbb{R}^2 : \max(a, -n) \leq x \leq \min(b, n), \max(\varphi(x), -n) \leq y \leq \min(\psi(x), n)\}$ . Mit dieser Definition des uneigentlichen Integrals gilt die Formel (\*) auch für unbeschränkte Integrationsintervalle.

#### 4. Berechnung eines Integrals

Wir verwenden obige Resultate, um das Integral  $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} \, dx$  zu berechnen. Man erhält dieses Integral als  $\lim_{n \rightarrow \infty} I_n$ , wenn man  $I_n$  für  $\int_{-n}^n e^{-\frac{1}{2}x^2} \, dx$  schreibt.

Die Rechnung führt über ein zweifaches Integral. Wir schreiben  $Q_n$  für das Quadrat  $[-n, n] \times [-n, n]$ . Durch entsprechende Umformungen erhalten wir dann

$$\begin{aligned} I_n^2 &= I_n \int_{-n}^n e^{-\frac{1}{2}x^2} \, dx = \int_{-n}^n e^{-\frac{1}{2}x^2} I_n \, dx = \int_{-n}^n e^{-\frac{1}{2}x^2} \int_{-n}^n e^{-\frac{1}{2}y^2} \, dy \, dx \\ &= \int_{-n}^n \int_{-n}^n e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}y^2} \, dy \, dx = \int_{-n}^n \int_{-n}^n e^{-\frac{1}{2}(x^2+y^2)} \, dy \, dx = \int_{Q_n} e^{-\frac{1}{2}(x^2+y^2)} \, d(x, y) \end{aligned}$$



Wir haben nur die Rechenregeln für einfache Integrale verwendet und das zweifache Integral, das am Ende herauskommt, als Volumen über dem Quadrat  $Q_n$  interpretiert, das unter der Funktion  $e^{-\frac{1}{2}(x^2+y^2)}$  liegt.

Sei  $K_R$  der Kreis mit Mittelpunkt  $(0, 0)$  und Radius  $R$ . Es gilt  $K_n \subseteq Q_n \subseteq K_{2n}$ , woraus  $\int_{K_n} e^{-\frac{1}{2}(x^2+y^2)} d(x, y) \leq \int_{Q_n} e^{-\frac{1}{2}(x^2+y^2)} d(x, y) \leq \int_{K_{2n}} e^{-\frac{1}{2}(x^2+y^2)} d(x, y)$  folgt. Damit haben wir eine Abschätzung für  $I_n^2$  gefunden.

Wir berechnen  $V = \int_{K_R} e^{-\frac{1}{2}(x^2+y^2)} d(x, y)$ . Dazu zerlegen wir den Kreis  $K_R$  in Kreisringe. Sei  $k \geq 1$  und  $r_j = j \frac{R}{k}$  für  $0 \leq j \leq k$ . Sei weiters  $V_j$  das Volumen unter der Funktion  $e^{-\frac{1}{2}(x^2+y^2)}$  über dem Kreisring  $K_{r_j} \setminus K_{r_{j-1}}$ . Es gilt dann  $V = \sum_{j=1}^k V_j$ . Die Funktionswerte auf diesem Kreisring liegen zwischen  $e^{-\frac{1}{2}r_{j-1}^2}$  und  $e^{-\frac{1}{2}r_j^2}$ . Die Fläche des Kreisrings ist  $\pi(r_j^2 - r_{j-1}^2)$ . Daher gilt

$$e^{-\frac{1}{2}r_j^2} \pi(r_j^2 - r_{j-1}^2) \leq V_j \leq e^{-\frac{1}{2}r_{j-1}^2} \pi(r_j^2 - r_{j-1}^2)$$

Wegen  $r_j^2 - r_{j-1}^2 = (r_j + r_{j-1})(r_j - r_{j-1}) = (2r_j - \frac{R}{k})(r_j - r_{j-1}) = (2r_{j-1} + \frac{R}{k})(r_j - r_{j-1})$  und da  $e^{-\frac{1}{2}r^2} \leq 1$  für alle  $r \geq 0$  gilt, erhalten wir

$$2\pi e^{-\frac{1}{2}r_j^2} r_j(r_j - r_{j-1}) - \frac{\pi R}{k}(r_j - r_{j-1}) \leq V_j \leq 2\pi e^{-\frac{1}{2}r_{j-1}^2} r_{j-1}(r_j - r_{j-1}) + \frac{\pi R}{k}(r_j - r_{j-1})$$

Summation über  $j$  von 1 bis  $k$  ergibt schließlich

$$2\pi \sum_{j=1}^k e^{-\frac{1}{2}r_j^2} r_j(r_j - r_{j-1}) - \frac{\pi R^2}{k} \leq V \leq 2\pi \sum_{j=1}^k e^{-\frac{1}{2}r_{j-1}^2} r_{j-1}(r_j - r_{j-1}) + \frac{\pi R^2}{k}$$

Die beiden Summen sind Riemannsummen, die gegen das Integral  $\int_0^R e^{-\frac{1}{2}r^2} r dr$  konvergieren, wenn  $k$  nach  $\infty$  geht. Dadurch und wegen  $\lim_{k \rightarrow \infty} \frac{\pi R^2}{k} = 0$  erhalten wir aus obigen Ungleichungen die Gleichung

$$V = 2\pi \int_0^R e^{-\frac{1}{2}r^2} r dr$$

Dieses Integral lässt sich durch die Substitution  $t = r^2$  leicht ausrechnen. Man erhält

$$V = \pi \int_0^{R^2} e^{-\frac{1}{2}t} dt = -2\pi e^{-\frac{1}{2}t} \Big|_0^{R^2} = 2\pi(1 - e^{-\frac{1}{2}R^2})$$

Damit ist  $V = \int_{K_R} e^{-\frac{1}{2}(x^2+y^2)} d(x, y)$  berechnet. Setzt man das in die oben gefundene Abschätzung für  $I_n^2$  ein, so erhält man

$$2\pi(1 - e^{-\frac{1}{2}n^2}) \leq I_n^2 \leq 2\pi(1 - e^{-2n^2}) \quad \text{oder} \quad \sqrt{2\pi} \sqrt{1 - e^{-n^2/2}} \leq I_n \leq \sqrt{2\pi} \sqrt{1 - e^{-2n^2}}$$

Es folgt  $\lim_{n \rightarrow \infty} I_n = \sqrt{2\pi}$ . Damit haben wir das gewünschte Resultat gefunden. Es gilt

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

Damit ist gezeigt, dass die Wahrscheinlichkeitsdichte der Standardnormalverteilung Integral 1 hat.

Wir berechnen das Volumen  $V$  unter der Funktion  $e^{-\frac{1}{2}(x^2+y^2)}$  über dem Kreis  $K_R$  noch auf eine andere Art, nämlich als Volumen eines Drehkörpers. Wir schneiden die Funktion  $\varphi(x) = e^{-\frac{1}{2}x^2}$  bei  $-R$  und  $R$  ab und lassen sie um die  $y$ -Achse rotieren. Das Volumen des Drehkörpers ist dann  $V$ .

Die Umkehrfunktion von  $\varphi$  erhalten wir durch Auflösen der Gleichung  $e^{-\frac{1}{2}x^2} = y$  nach der Variable  $x$ . Das ergibt  $\varphi^{-1}(y) = \sqrt{-2 \log y}$ . Wir schneiden bei  $R$  ab und erhalten so die Funktion  $h(y) = \min(R, \sqrt{-2 \log y})$ , die auf dem Intervall  $[0, 1]$  definiert ist. Durch

Rotation um die  $y$ -Achse ergibt sich das Volumen  $V$ , das heißt  $V = \pi \int_0^1 h(y)^2 dy$ . Wir teilen das Integrationsintervall bei  $y_0 = e^{-\frac{1}{2}R^2}$ . Es gilt  $h(y) = R$  auf dem Intervall  $[0, y_0]$  und  $h(y) = \sqrt{-2 \log y}$  auf dem Intervall  $[y_0, 1]$ . Damit erhalten wir dann

$$\begin{aligned} V &= \pi \int_0^{y_0} R^2 dy + \pi \int_{y_0}^1 -2 \log y dy = \pi R^2 y_0 - 2\pi(y \log y - y) \Big|_{y_0}^1 \\ &= \pi R^2 y_0 + 2\pi + 2\pi y_0 \log y_0 - 2\pi y_0 = \pi R^2 y_0 + 2\pi - \pi y_0 R^2 - 2\pi e^{-\frac{1}{2}R^2} \\ &= 2\pi(1 - e^{-\frac{1}{2}R^2}) \end{aligned}$$

Damit haben wir für  $V$  denselben Wert wie oben erhalten. Alles andere lässt sich genauso wie oben berechnen.

## 5. Die Gammafunktion

Die Gammafunktion wird durch  $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$  für alle  $r > 0$  definiert.

**Satz 48:** Für  $r > 0$  gilt  $\Gamma(r+1) = r\Gamma(r)$ .

**Beweis:** Das zeigt man mit partieller Integration:

$$\Gamma(r+1) = \int_0^\infty x^r e^{-x} dx = -x^r e^{-x} \Big|_0^\infty + \int_0^\infty r x^{r-1} e^{-x} dx = 0 + r\Gamma(r). \quad \square$$

Für spezielle Werte von  $r$ , zum Beispiel für  $r \in \mathbb{N}$ , kann man  $\Gamma(r)$  explizit berechnen. Es gilt  $\Gamma(1) = \int_0^\infty e^{-x} dx = -e^{-x} \Big|_0^\infty = 1$  und für  $n \in \{1, 2, 3, \dots\}$  gilt  $\Gamma(n) = (n-1)!$ . Das folgt mit Induktion. Für  $n=1$  haben wir es gerade berechnet. Ist  $\Gamma(n) = (n-1)!$  schon gezeigt, dann folgt  $\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n!$  mit Hilfe von Satz 48.

**Satz 49:** Es gilt  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

**Beweis:** Wir haben  $\int_{-\infty}^\infty e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$  gezeigt. Da  $e^{-\frac{1}{2}x^2}$  eine gerade Funktion ist, ist das Integral über  $[0, \infty)$  halb so groß. Wir haben daher  $\int_0^\infty e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{\pi}{2}}$ . Wir führen die neue Integrationsvariable  $y = \frac{1}{2}x^2$  ein. Es gilt  $x = \sqrt{2y}$  und  $dx = \frac{1}{\sqrt{2y}} dy$ . Wir erhalten  $\int_0^\infty e^{-y} \frac{1}{\sqrt{2y}} dy = \sqrt{\frac{\pi}{2}}$ . Multipliziert man diese Gleichung mit  $\sqrt{2}$ , dann steht bereits  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$  da.  $\square$

**Satz 50:** Für  $n \in \{0, 1, 2, \dots\}$  gilt  $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{n! 2^{2n}} \sqrt{\pi}$ .

**Beweis:** Für  $n=0$  ist das Satz 49. Sei  $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{n! 2^{2n}} \sqrt{\pi}$  bereits gezeigt. Aus Satz 48 folgt  $\Gamma(n + \frac{3}{2}) = (n + \frac{1}{2})\Gamma(n + \frac{1}{2}) = \frac{(2n+2)(2n+1)}{4(n+1)} \Gamma(n + \frac{1}{2})$ . Setzt man für  $\Gamma(n + \frac{1}{2})$  ein, so hat man  $\Gamma(n + \frac{3}{2}) = \frac{(2n+2)!}{(n+1)! 2^{2n+2}} \sqrt{\pi}$ , die Formel für  $n+1$ .  $\square$

# Inhaltsverzeichnis

<b>I. Kombinatorik</b>	1
1. Geordnete Stichproben	1
2. Ungeordnete Stichproben	3
3. Zerlegungen einer Menge	7
4. Anordnungen (Permutationen)	9
<b>II. Wahrscheinlichkeitstheorie</b>	10
1. Zufallsexperiment, Ausfall, Ereignis	10
2. Wahrscheinlichkeit	11
3. Gleichwahrscheinliche Ausfälle	13
4. Geometrische Wahrscheinlichkeit	15
5. Bedingte Wahrscheinlichkeit	16
6. Totale Wahrscheinlichkeit	20
<b>III. Zufallsvariable</b>	23
1. Darstellung von Ereignissen durch Zufallsvariable	23
2. Wahrscheinlichkeitsvektoren und Wahrscheinlichkeitsdichten	23
3. Binomialverteilung und geometrische Verteilung	27
4. Die hypergeometrische Verteilung	31
5. Poissonverteilung	32
6. Exponentialverteilung und Gammaverteilung	33
7. Normalverteilung	35
8. Approximation der Binomialverteilung	38
<b>IV. Mehrere Zufallsvariable</b>	42
1. Gemeinsame Verteilungen von Zufallsvariablen	42
2. Rechnen mit Zufallsvariablen	45
3. Erwartungswert und Varianz	48
<b>V. Statistik</b>	51
1. Parameterschätzer	51
2. Konfidenzintervalle für normalverteilte Messwerte	54
3. Konfidenzintervalle für Prozentsätze	57
4. Statistische Tests für Prozentsätze	60
5. Tests für den Erwartungswert einer Normalverteilung	65
<b>VI. Anhang: Vorkenntnisse aus der Analysis</b>	67

### 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 $z_\beta$ und $v_{\beta,n}$ (Normalverteilung und T-Verteilung)

$\beta =$	$z_\beta$	$v_{\beta,7}$	$v_{\beta,8}$	$v_{\beta,9}$	$v_{\beta,10}$	$v_{\beta,12}$	$v_{\beta,14}$	$v_{\beta,16}$	$v_{\beta,18}$	$v_{\beta,25}$	$v_{\beta,40}$
0.05	1.65	1.895	1.860	1.833	1.812	1.782	1.761	1.746	1.734	1.708	1.684
0.025	1.96	2.365	2.306	2.262	2.228	2.179	2.145	2.120	2.101	2.060	2.021
0.01	2.33	2.998	2.896	2.821	2.764	2.681	2.624	2.584	2.552	2.485	2.423
0.005	2.58	3.499	3.355	3.250	3.169	3.055	2.977	2.921	2.878	2.787	2.704

### Tabellen für $u_{\beta,n}$ ( $\chi^2$ -Verteilung)

$\beta =$	$u_{\beta,1}$	$u_{\beta,2}$	$u_{\beta,3}$	$u_{\beta,4}$	$u_{\beta,5}$	$u_{\beta,6}$	$u_{\beta,8}$	$u_{\beta,9}$	$u_{\beta,10}$	$u_{\beta,16}$	$u_{\beta,20}$
0.05	3.84	5.99	7.81	9.49	11.07	12.6	15.51	16.92	18.31	26.30	31.41
0.025	5.02	7.38	9.35	11.1	12.83	14.4	17.53	19.02	20.48	28.85	34.17
0.01	6.63	9.21	11.3	13.3	15.1	16.8	20.1	21.7	23.2	32.0	37.6
0.005	7.88	10.6	12.8	14.9	16.7	18.5	22.0	23.6	25.2	34.3	40.0

## Zusammenstellung der wichtigen Verteilungen

Name	Parameter	Abkürzung	Wertebereich	W-Vektor/Dichte	Ew.	Varianz
Binomial- verteilung	$n \in \mathbb{N}, p \in (0, 1)$	$B(n, p)$	$\{0, 1, \dots, n\}$	$w(j) = \binom{n}{j} p^j (1-p)^{n-j}$	$np$	$np(1-p)$
geometrische Verteilung	$p \in (0, 1)$	$G(p)$	$\{1, 2, \dots\}$	$w(j) = (1-p)^{j-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson- verteilung	$\lambda > 0$	$P(\lambda)$	$\{0, 1, 2, \dots\}$	$w(j) = e^{-\lambda} \frac{\lambda^j}{j!}$	$\lambda$	$\lambda$
Normal- verteilung	$\mu \in \mathbb{R}, \sigma > 0$	$N(\mu, \sigma)$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\mu$	$\sigma^2$
Exponential- verteilung	$\lambda > 0$	$E(\lambda)$	$\mathbb{R}_0^+$	$f(x) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma- verteilung	$\lambda > 0, r > 0$	$E(\lambda, r)$	$\mathbb{R}_0^+$	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$
$T$ -Verteilung	$n \in \mathbb{N}$	$T(n)$	$\mathbb{R}$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$	$0$	