

*B. Schmalfuß*

# **Stochastik für Informatik und Lehramt**

Institut für Mathematik  
Friedrich Schiller Universität Jena  
Ernst Abbe Platz 2  
07737 Jena

**Urheberrechtlicher Hinweis:** Der Autor untersagt ausdrücklich jegliche Art der nichtautorisierten Kopie dieses Lehrmaterials oder von Teilen desselben.

## Vorwort

Dieses Vorlesungsskript enthält **Teile** der Vorlesung *Stochastik für Informatiker, Stochastik für Ingenieure* beziehungsweise *Stochastik für Lehramt*. Einige Teile der in der Vorlesung behandelten Themen müssen noch in den nächsten Jahren ergänzt werden. Das bezieht sich speziell auf Themen der Statistik.

Das Manuskript ist ursprünglich für die Lehre an einer Fachhochschule verfasst worden und kann somit hier nur als Lehrmaterial betrachtet werden.

**Die Prüfungsthemen richten sich nicht nach dem Inhalt des Skriptes sondern nach dem in der Vorlesung behandelten Stoffes.**

Da es sich bei diesem Skript um die erste Ausgabe handelt, befinden sich wahrscheinlich noch viele Tippfehler im Text. Für entsprechende Hinweise und andere kritische Bemerkungen wäre ich sehr dankbar.

**Achtung: Die Normalverteilung mit Mittelwert  $\mu$  und Varianz  $\sigma^2$  wird in diesem Skript mit  $N(\mu, \sigma)$  bezeichnet!!!!!!!!!!**

Björn Schmalfuß

27. Januar 2014

email: [bjoern.schmalfuss@uni-jena.de](mailto:bjoern.schmalfuss@uni-jena.de)



# Inhaltsverzeichnis

Vorwort	3
Kapitel 1. Einleitung	7
Kapitel 2. Zufällige Ereignisse und Wahrscheinlichkeiten	13
1. Das Rechnen mit zufälligen Ereignissen	13
2. Heuristische Einführung der Wahrscheinlichkeit	17
3. Die axiomatische Definition einer Wahrscheinlichkeit	20
Kapitel 3. Die bedingte Wahrscheinlichkeit	27
1. Die Definition der bedingten Wahrscheinlichkeit	27
2. Einige Sätze über die bedingte Wahrscheinlichkeit	29
3. Unabhängigkeit von Ereignissen	30
Kapitel 4. Zufallsvariable und deren Verteilung	33
1. Einführung von Zufallsvariablen	33
2. Wahrscheinlichkeitsverteilungen	34
3. Parameter von Zufallsvariablen	50
Kapitel 5. Die Erzeugung von Zufallszahlen	57
1. Die Simulation	57
2. Die Erzeugung gleichverteilter Zufallszahlen	57
3. Erzeugung von Zufallszahlen mittels der Verteilungsfunktion	60
4. Normalverteilte Zufallsvariablen und zentraler Grenzwertsatz	63
Kapitel 6. Mehrdimensionale Wahrscheinlichkeitsverteilungen	67
1. Allgemeine Eigenschaften von mehrdimensionalen Verteilungen	67
2. Randverteilungen und Unabhängigkeit	69
3. Maße für die Abhängigkeit von zwei Zufallsvariablen.	71
Kapitel 7. Markov-Ketten	75
1. Grundlegende Definitionen	75
2. Klassifikation der Zustände einer Markov-Kette	78
3. Stationäres Verhalten der Markov-Kette	81
4. Markov-Ketten mit endlich vielen Zuständen	83
Kapitel 8. Schließende Statistik	87
1. Grundgesamtheit und Stichprobe	87
2. Eigenschaften von Punktschätzungen	91
3. Intervallschätzungen	93
4. Maximum Likelihood-Schätzung	99

Kapitel 9. Statistische Tests	103
1. Die Grundidee eines Tests	103
2. Test des Mittelwertes	105
3. Statistische Tests für die Varianz	111
4. Das Testen einer Verteilung	114
5. Der $\beta$ -Fehler	117
Kapitel 10. Regressionsanalyse	121
1. Einleitung	121
2. Differentiation von Matrizen	123
3. Die Methode der kleinsten Quadrate	124
4. Bewertung der Regressionsrechnung	129
5. Das Gauß-Markov Modell der Regression	132
Anhang A. Kombinatorik	137
Literaturverzeichnis	141

## KAPITEL 1

### Einleitung

eder Mensch hat schon einmal in seinem Leben mit Ereignissen zu tun gehabt, deren Ausgang er nicht 100%ig vorhersagen konnte. Man denke zum Beispiel nur an ein Lottospiel. Im Moment der Abgabe des Lottoscheins weiß man nicht, ob man am nächsten Wochenende reich ist (was ausgesprochen selten der Fall sein wird), oder ob man den Betrag den man entrichten musste, zum Fenster heraus geworfen hat.

Ähnliche Unbestimmtheiten gibt es natürlich auch in den Natur-, Ingenieur- oder Gesellschaftswissenschaften und natürlich auch in der Informatik. Bei der Analyse solcher Ereignisse stellt man fest, dass man diese eigentlich in zwei große Richtungen einteilen kann. Diese verschiedenen Richtungen spiegeln sich auch in der Philosophie wider. Die erste Richtung beschreibt Ereignisse, die nicht vollständig bestimmbar sind aufgrund der *Unfähigkeit* eines Menschen, den Zustand des betrachteten Objektes mit *unendlicher* Genauigkeit beschreiben zu können. Beispielsweise kann man beim Werfen eines Würfels nicht genau sagen, mit welchem Impuls und aus welcher Lage heraus man den Würfel wirft. Deswegen wird man nicht in der Lage sein, zu berechnen welche Ziffer am Ende oben liegt. In der Philosophie wird diese Situation durch den *Laplaceschen Dämon* beschrieben, einem Wesen, welches über die Unzulänglichkeit des Menschen erhaben ist, die Realität nur mit beschränkter Genauigkeit zu kennen. Zufall wird also als die Unfähigkeit des Menschen interpretiert, die Realität mit unendlicher Genauigkeit zu beschreiben, obwohl die Realität an sich *vollständig bestimmt* ist.

Diese Einschätzung änderte sich grundlegend zu Beginn des letzten Jahrhunderts, bedingt durch Forschungsergebnisse aus der Atomphysik. Speziell ist es beim Zerfall von radioaktiven Teilchen nur möglich auszusagen, wieviel Prozent der vorhandenen Teilchen in einem gewissen Zeitraum zerfallen, welches der Teilchen in einem bestimmten Zeitraum zerfällt entzieht sich völlig der Kenntnis des Physikers. Schnell erkannte man auch, dass Gesetzmäßigkeiten der Quantenphysik oder statistischen Quantenmechanik nur dann formuliert werden können, falls man die Unbestimmtheit zugrunde legt. Der Zufall kann also als etwas *real existierendes* angesehen werden. Dies ist die zweite Richtung der Interpretation des Zufalls in der Philosophie

In dieser Vorlesung soll die Definition des Zufalls die beiden gegebenen Interpretationen einschließen:

**Ein Ereignis heißt zufällig, falls man nicht in der Lage ist, mit 100%iger Sicherheit das Eintreten oder Nichteintreten vorherzusagen zu können.<sup>1</sup>**

Um mit dem Begriff des Zufalls in den Wissenschaften arbeiten zu können, ergibt sich die Aufgabe, die Unbestimmtheit des Auftretens von zufälligen Ereignissen quantifizieren oder messen zu können. Diese Messgröße wird als *Wahrscheinlichkeit* bezeichnet und ist gegeben durch eine Zahl zwischen Null und Eins, wobei eine Wahrscheinlichkeit ungefähr Null bedeutet, dass das Ereignis relativ selten auftritt, während eine Wahrscheinlichkeit in der Nähe von Eins bedeutet, dass das Ereignis relativ oft auftritt. Das Rechnen mit Wahrscheinlichkeiten unterliegt bestimmten Gesetzen, die wir in den nächsten Kapiteln kennenlernen werden.

Wir wollen nun einige Beispiele aus der Praxis betrachten, in denen der Zufall eine große Rolle spielt.

**Warteschlange an Rechenanlage:** Jobs kommen zu zufälligen Zeitpunkten an einer Rechenanlage an und müssen eine zufällige Zeit warten, bis sie bearbeitet werden. Die Bearbeitungszeit der Jobs ist zufällig. Es baut sich eine Warteschlange von Jobs auf, die eine zufällige Länge hat. Weiterhin ist die Zeit, die ein Job bis zu seiner Bearbeitung warten muss, zufällig. Um den Zustand der Anlage mindestens teilweise beschreiben zu können, werden Kenngrößen, wie *mittlere Wartezeit* und *mittlere Warteschlangenlänge* eingeführt. Diese Größen beschreiben also nicht die Wartezeit eines speziellen Jobs, sondern den Mittelwert, der sich aus den Wartezeiten der Jobs ergibt über eine gewisse Zeitspanne betrachtet. Aus diesen Kenngrößen können Rückschlüsse auf die Organisation der Anlage gezogen werden.

**Belegung des Hauptspeichers eines Computers:** Bei der Belegung des Hauptspeichers durch Programmblocke gibt es verschiedene Strategien. Bei der *first fit* Strategie wird ein Programmblock mit zufälliger Länge in den ersten verfügbaren Speicherraum mit hinreichender Größe geladen. Dadurch wird einerseits wenig Zeit für das Suchen von Speicherplatz verwendet, andererseits wird unter Umständen viel Speicherplatz verschwendet, da der zusammenhängende Speicherplatz viel größer sein kann als der Programmblock. Für einen folgenden größeren Block ist dann unter Umständen kein Speicherplatz mehr vorhanden. Eine andere Strategie ist die *best fit* Strategie, wobei ein Programmblock in den Speicherblock geladen wird, der unter allen Speicherblöcken, die eine Länge größer als der Programmblock haben die minimale Länge besitzt. Ob die eine oder die andere Strategie die günstigere ist, hängt zum Beispiel von dem zur Verfügung stehenden Speicherplatz, aber auch von der Größe der zu ladenden Blöcke ab, die aber nicht a priori bekannt ist. Somit kann die Abarbeitung eines Programmes als zufällig angesehen werden. Man kann also nur sagen, dass *durchschnittlich* die eine Strategie besser ist als die andere, falls das zu bearbeitende Programm zum Beispiel aus vielen kleinen Programmblocken besteht.

**Zufallsgeneratoren:** Viele Programmiersprachen enthalten Funktionen, die zufällige Zahlen erzeugen, die unabhängig von den zuvor erzeugten Zufallszahlen sind. Die Erzeugung dieser Zahlen basiert auf gewissen Iterationsformeln. Erzeugt man

---

<sup>1</sup>Diese Definition wird im Folgenden noch adaptiert werden müssen

sehr viele dieser Zahlen mittels eines Zufallsgenerators, so erkennt man, dass bei diesen Zahlen eine gewisse Periodizität auftritt. So eine Periodizität drückt aber doch eine Abhängigkeit der Zahlen untereinander aus, was allerdings der Zufälligkeit widerspricht. Man spricht deswegen häufig nur von *Pseudozufallszahlen*. In Abbildung 1 ist ein Polygon mittels eines Zufallsgenerators erzeugt worden.

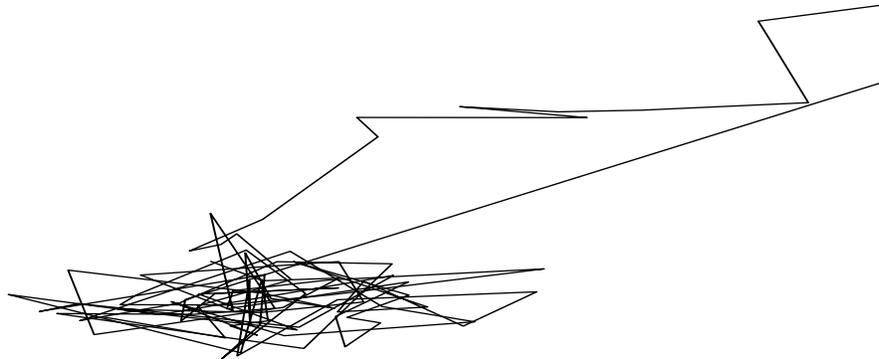


ABBILDUNG 1. Durch einen Zufallsgenerator erzeugter zufälliger Weg

**Filtertheorie:** Messsignale, die von einem Messfühler zum Ableseinstrument übertragen werden, unterliegen bei dieser Übertragung häufig *zufälligen Störungen*. Man sagt auch, dass die Messergebnisse durch ein Rauschen verfälscht werden. Die Aufgabe besteht nun darin, Verfahren anzugeben, die das Rauschen aus den Messwerten wenigstens teilweise herausfiltern. Die robusteste Methode, um diese Filterung vorzunehmen, ist die Überlagerungstechnik. Dabei wird ein Messsignal über mehrere sich nicht beeinflussende Leitungen zum Ableseinstrument übertragen und gemittelt. Durch diesen Mittelungsprozess wird erreicht, dass sich die einzelnen Rauschanteile aufheben. Diese Methode ist allerdings sehr grob und kostenintensiv, da viele parallele Leitungen notwendig sind. In der Praxis kommt der *Bucy-Kalman Filter* zur Anwendung, eine durch die Mathematik fundierte Methode, die das Rauschen aus den Messwerten herausfiltert.

**Erneuerungstheorie:** Die Erneuerungstheorie beschäftigt sich mit dem Ausfallen, Reparieren und Ersetzen von Teilen eines Systems. Dabei geht man davon aus, dass die Lebensdauer eines Teiles im Allgemeinen unbekannt ist. Sie wird also als zufällig angenommen. Die Aufgabe der Erneuerungstheorie ist es, Wartungsstrategien zu entwickeln, so dass das Betreiben einer Anlage hinsichtlich der Erneuerung und Reparatur von Teilen *durchschnittlich* sich kostengünstig gestaltet.

**Statistische Physik:** Teilgebiet der Physik, das die Eigenschaft der Materie auf die Eigenschaft der entsprechenden Moleküle und Atome zurückführt, die in einer sehr großen Anzahl vorliegen. Somit kann man nicht die Bewegung einzelner Moleküle und Atome, die zufällig ist, beschreiben, sondern man muss die Methoden der Wahrscheinlichkeitsrechnung anwenden. Größen wie Druck, Temperatur und Wärmeleitfähigkeit können als Mittelwert bestimmter Bewegungsgrößen sehr vieler

Teilchen beschrieben werden.

**Monte Carlo Simulation:** Methode zur Approximation sehr komplizierter beziehungsweise sehr hochdimensionaler bestimmter Integrale. Dabei werden *gleichmäßig verteilte* zufällige Punkte in einen Rechteckbereich gelegt, in dem sich der Graph der Funktion befindet. Das Verhältnis der zufälligen Punkte *unter* der Funktion zu der Gesamtzahl der Punkte erlaubt Rückschlüsse auf den Wert des bestimmten Integrals, siehe Abbildung 2.

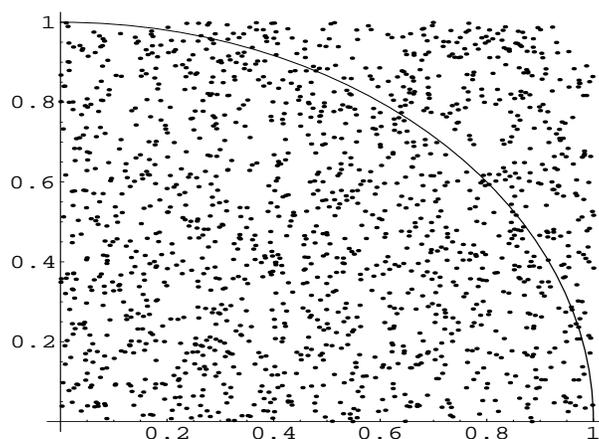


ABBILDUNG 2. Monte Carlo Simulation

Die Wissenschaft, die sich mit der Beschreibung des Zufalls beschäftigt, ist die *Stochastik*. Sie ist eine Teildisziplin der Mathematik, hat aber aufgrund der Bedeutung für andere Wissenschaften in diesen Wissenschaften ihren eigenen Charakter entwickelt.

Die Stochastik unterteilt sich in drei Teildisziplinen, siehe Abbildung 3:

- Die *Wahrscheinlichkeitstheorie* stellt die mathematischen Instrumente bereit, um mit zufälligen Ereignissen rechnen zu können beziehungsweise um Wahrscheinlichkeiten und alle darauf aufbauenden Begriffe definieren zu können.
- Die *Beschreibende Statistik* hat die Aufgabe große Datenmengen durch Einführung von Kenngrößen verdichten zu können. Weiterhin beschäftigt sich die Beschreibende Statistik mit einer übersichtlichen Darstellung dieser Datensätze.
- Die *Schließende Statistik* beschäftigt sich mit der Analyse einer großen Datenmenge (Grundgesamtheit) durch eine relativ kleine Datenmenge (Stichprobe). So werden zum Beispiel aus der Stichprobe gewisse Parameter der Grundgesamtheit, wie Mittelwert oder Streuung der Grundgesamtheit geschätzt.

Beispiele sind Wahlhochrechnungen oder die Ermittlung der Einschaltquote beim Fernsehen. Weiterhin werden in der statistischen Qualitätskontrolle Hypothesen über die Qualität eines Produktes geprüft.

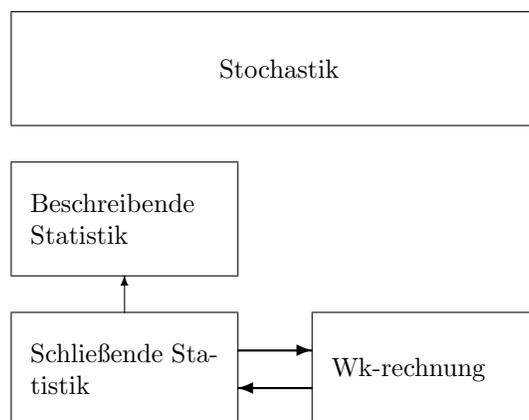


ABBILDUNG 3. Teilgebiete der Stochastik

Die Bedeutung der Statistik liegt vor allem darin, dass sie die empirischen Wissenschaften unterstützt (Empirie-Lernen aus der Erfahrung). Das heißt, dass aus Datensätzen beziehungsweise aus Stichproben Aussagen über vollständig oder teilweise bekannte Gesetzmäßigkeiten gewonnen werden.

Standardmethoden der Statistik sind häufig in Programmpaketen implementiert. Dabei sind stellvertretend zu nennen SAS, SPSS oder SPLUS. Bei diesen Methoden gibt es häufig ein enges Zusammenspiel zwischen Mathematik, Stochastik und Informatik. Das soll am Beispiel der Regressionsrechnung demonstriert werden.

Es besteht die Aufgabe, eine *Punktwolke* von Messdaten  $(x_i, y_i)_{i=1, \dots, n}$  durch eine Gerade  $y = \hat{a}_0 + \hat{a}_1 x$  zu beschreiben. Damit kann das *gemittelte Verhalten* des funktionalen Zusammenhangs der Messwerte beschrieben werden. Die Frage ist nun, welches Kriterium legt man an, um die Punktwolke durch eine Gerade *bestmöglich* zu beschreiben, siehe Abbildung 4. In der Praxis hat sich die Methode der kleinsten Quadrate bewährt:

**Man bestimme eine Gerade so, dass die Summe der Quadratabstände zwischen den  $y$ -Koordinaten  $y_i$  der Messwerte und den Funktionswerten  $y(x_i)$  der Geraden minimal wird.**

Der oben beschriebene Quadratabstand einer beliebigen Geraden

$$y(x) = a_0 + a_1 x$$

zu den Messwerten kann durch eine Funktion beschrieben werden:

$$S(a_0, a_1) = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))^2.$$

Die Koeffizienten  $\hat{a}_0$ ,  $\hat{a}_1$  der besten Gerade im Sinne der Methode der kleinsten Quadrate ergeben sich nun als die Minimalstelle eines Extremwertproblems für eine Funktion mit zwei Veränderlichen. Die Mathematik stellt nun Methoden bereit, diese Minimalstelle zu berechnen. Dazu muss ein lineares Gleichungssystem mit zwei Veränderlichen gelöst werden. Die Berechnung der Koeffizienten  $\hat{a}_0$ ,  $\hat{a}_1$  ist nun in gebräuchlichen Statistik-Programmpaketen implementiert. Weiterhin sind Methoden implementiert, die eine graphische Auswertung der Problemstellung erlauben. In [Abbildung 4](#) wird die Entwicklung der Temperatur in Sheerbroke/Kanada von 1900 bis 1990 beschrieben, siehe [\[21\]](#). Der Trend der Temperaturentwicklung wird mittels Methode der kleinsten Quadrate bestimmt.

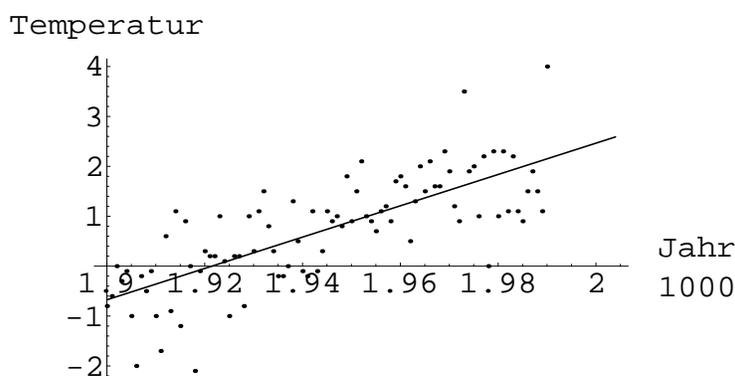


ABBILDUNG 4. Entwicklung der mittleren Temperatur einer Stadt in Kanada und die Beschreibung dieser Punktwolke durch eine Gerade

## Zufällige Ereignisse und Wahrscheinlichkeiten

### 1. Das Rechnen mit zufälligen Ereignissen

Ein *zufälliger Versuch* ist dadurch bestimmt, dass seine Ergebnisse nicht vollständig vorhersagbar sind. Der Terminus zufälliger Versuche kann stehen für das Durchführen einer Messung, Entnahme einer Probe, Durchführung einer Erhebung und so weiter. Wir wollen im Folgenden annehmen, dass so ein zufälliger Versuch unter annähernd gleichen Bedingungen sehr häufig durchgeführt werden kann. Idealisiert man diese Bedingungen, so bedeutet dies, dass man annehmen kann, dass ein zufälliger Versuch unter gleichbleibenden Bedingungen beliebig oft ausgeführt werden kann. Die Ergebnisse, die ein zufälliger Versuch liefert, nennt man zufällige Ereignisse. Es ist im allgemeinen nicht klar, ob bei der einmaligen Durchführung eines zufälligen Versuches ein bestimmtes zufälliges Ereignis eintritt oder nicht.

**Beispiel:** Der einfachste zufällige Versuch ist der Münzwurf. Es bereitet bestimmt keine Probleme sich vorzustellen, dass man eine Münze beliebig oft unter gleichen Bedingungen werfen kann. Man erkennt sofort zwei zufällige Ereignisse, nämlich *Werfen von Zahl* beziehungsweise *Werfen von Kopf*.

Zufällige Ereignisse, wir wollen im Folgenden kurz nur noch Ereignisse sagen, werden mit großen Buchstaben  $A, B, \dots$  bezeichnet und sehr häufig durch Zahlenmengen charakterisiert.

**Beispiel:** Betrachtet man die Lebensdauer einer Festplatte aus einer bestimmten Bauserie, so interessiert das zufällige Ereignis  $A$ , dass diese Festplatte länger als fünf Jahre lebt. Bezeichnet  $T$  die zufällige Zeitspanne, die die Festplatte funktioniert hat, so kann das Ereignis  $A$  beschrieben werden:

$$A = \{T > 5\} = (5, \infty).$$

Es sei  $B$  das Ereignis, dass weniger als vier Speicherchips in Lieferpackungen von je 100 Stück defekt sind:

$$B = \{0, 1, 2, 3\}.$$

Um mit Wahrscheinlichkeiten für das Auftreten von Ereignissen rechnen zu können, müssen zuerst Rechenregeln für den Umgang mit Ereignissen eingeführt werden.

**DEFINITION 2.1.** *Gegeben seien zwei Ereignisse  $A$  und  $B$ . Wir führen ein Ereignis ein, das eintritt, wenn entweder  $A$  oder  $B$  eintritt. Dieses neue zufällige Ereignis  $C$  wird bezeichnet mit*

$$C = A \cup B \text{ oder } A \vee B.$$

Man nennt es die Vereinigung von  $A$  und  $B$  (oder  $A$  vereinigt  $B$ ) oder die Summe von  $A$  und  $B$ . Man kann dieses Ereignis auch mit  $A$  **oder**  $B$  ansprechen.

**Beispiel:** Es sei  $A$  das Ereignis beim einmaligen Werfen eines Würfels eine gerade Zahl zu erhalten,  $B$  sei das Ereignis eine Primzahl  $p > 1$  zu werfen. Dann ist

$$C = A \cup B = \{2, 3, 4, 5, 6\}.$$

Anders als in der Umgangssprache (heute ist Montag oder Dienstag) können zwei durch *oder* verknüpfte Ereignisse auch gleichzeitig eintreten. So ist im obigen Beispiel die Zahl 2 sowohl eine gerade Zahl als auch eine Primzahl.

Interpretiert man Ereignisse als Punktmenge auf der Ebene, so erhält man die Vereinigung dieser Mengen als alle Punkte, die (mindestens)

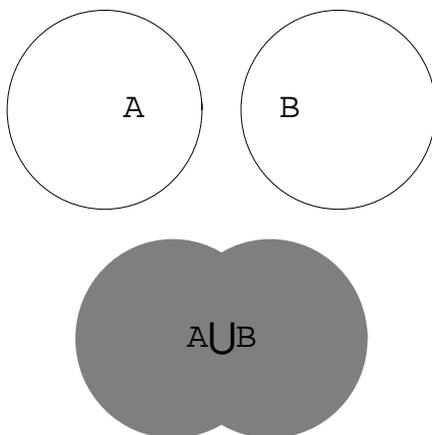


ABBILDUNG 1. Die Vereinigung  $A \cup B$  zweier Ereignisse

in einer dieser Mengen liegen.

Wir führen nun die zweite Rechenoperation für Ereignisse ein:

**DEFINITION 2.2.** Gegeben seien die Ereignisse  $A, B$ . Das neue Ereignis  $C$  tritt genau dann ein, falls  $A$  und gleichzeitig  $B$  eintreten. Die Bezeichnung für dieses Ereignis ist

$$A \cap B, \text{ oder } A \wedge B.$$

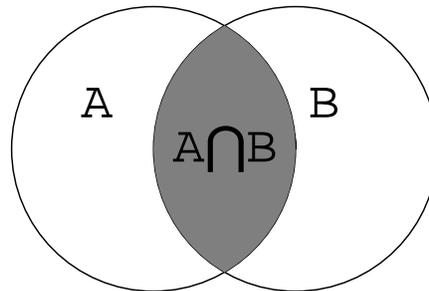
Man sagt  $C$  ist der Durchschnitt oder das Produkt von  $A$  und  $B$  beziehungsweise  $C$  ergibt sich aus  $A$  geschnitten  $B$ ,  $A$  **und**  $B$ .

**Beispiel:** Es sei  $A$  und  $B$  wie im letzten Beispiel definiert, so ist

$$C = A \cap B = \{2\}.$$

Die geometrische Interpretation des Durchschnittes ist in Abbildung 2 dargestellt.

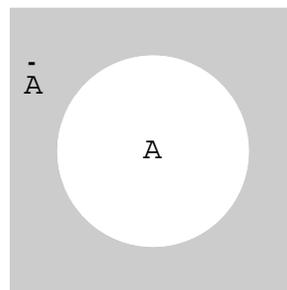
Eine weitere Grundoperation für Ereignisse ist die Negation:

ABBILDUNG 2. Der Durchschnitt  $A \cap B$  zweier Ereignisse

DEFINITION 2.3. Gegeben sei ein Ereignis  $A$ . Dann bezeichnet  $\bar{A}$  oder  $A^c$  das Ereignis, das eintritt, falls  $A$  **nicht** eintritt. Es wird mit Negation von  $A$  oder Komplement von  $A$  bezeichnet.

**Beispiel:** Falls man keine gerade Zahl würfelt (Ereignis gerade Zahl:  $A = \{2, 4, 6\}$ ), so würfelt man eine ungerade Zahl, das heißt:

$$\bar{A} = \{1, 3, 5\}.$$

ABBILDUNG 3. Das Komplement  $\bar{A}$  von  $A$ .

Bildet man die Negation der Negation eines Ereignisses, so erhält man das Ereignis selbst:

$$A = \overline{(\bar{A})}.$$

Es gibt noch weitere Operationen, die sich aber auf diese Grundoperationen zurückführen lassen. Ein Beispiel ist die Ereignisdifferenz  $A \setminus B$  (als  $A$  minus  $B$  ausgesprochen) die genau dann eintritt, falls  $A$  eintritt,  $B$  aber nicht, siehe Abbildung 4.

LEMMA 2.4. Für beliebige Ereignisse  $A, B$  erhält man  $A \setminus B = A \cap \bar{B}$ .

Man mache sich diese einfache Beziehung klar .

Die folgende Relation kann zwischen zwei Ereignissen eingeführt werden:

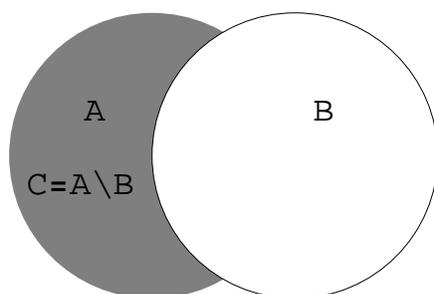


ABBILDUNG 4. Die Ereignisdifferenz  $A \setminus B$ .

DEFINITION 2.5. *Man sagt: Das Ereignis  $A$  zieht das Ereignis  $B$  nach sich, falls aus dem Eintreten von  $A$  das Eintreten von  $B$  folgt. Für diese Relation schreibt man  $A \subset B$ .*

**Beispiel:** Falls  $A = \{2, 3\}$  gewürfelt wird, dann zieht dieses Ereignis nach sich, dass auch das Ereignis  $B$ , eine Primzahl zu würfeln,  $B = \{2, 3, 5\}$  eintritt.

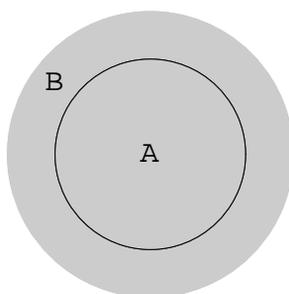


ABBILDUNG 5.  $A \subset B$

Mit Hilfe der Relation  $\subset$  kann man entscheiden, ob zwei Ereignisse  $A, B$  gleich sind:  $A$  und  $B$  sind genau dann gleich, falls  $A \subset B$  und  $B \subset A$  gilt.

Sicher hat der Leser schon bemerkt, dass das Rechnen mit Ereignissen dem Rechnen mit Mengen entspricht. Allgemeiner kann man feststellen, dass die Ereignisoperationen eine *Boolesche Algebra* bilden. Das bedeutet, dass die folgenden Gesetze erfüllt sind:

**Kommutatives Gesetz:**

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

**Assoziatives Gesetz:**

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$

Desweiteren gelten die **distributiven Gesetze:**

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$$

Ersetzt man in der ersten Zeile in den letzten Formeln  $\cup$  durch  $+$  und  $\cap$  durch  $\cdot$ , dann hat dieses Gesetz die gleiche Gestalt, wie das distributive Gesetz für das Rechnen mit reellen Zahlen. Die zweite Zeile würde für das Rechnen mit Zahlen keinen Sinn mehr machen.

Ähnlich wie beim Rechnen mit reellen Zahlen, wo die Punktrechnung vor der Strichrechnung ausgeführt wird, gibt es auch eine Klammerregel für das Rechnen mit Ereignissen. Und zwar gilt  $\cap$  vor  $\cup$ . Speziell könnten im ersten distributiven Gesetz die Klammern auf der rechten Seite und im zweiten auf der linken Seite weggelassen werden. Der Übersichtlichkeit halber empfiehlt es sich aber, lieber ein Klammerpaar mehr als ein Klammerpaar zu wenig zu setzen.

Wir betrachten nun weitere Eigenschaften für das Rechnen mit Ereignissen. Speziell gelten die *De Morganschen Regeln*:

SATZ 2.6. Für beliebige Ereignisse  $A, B$  gilt:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Man veranschauliche sich diese Beziehungen graphisch. Weiterhin zeige man, dass aus einer dieser Gleichungen die andere folgt.

Diese Beziehungen bleiben sinngemäß richtig, falls man anstelle von zwei Ereignissen mehr als zwei (also beliebig viele) betrachtet.

Um mit Ereignissen rechnen zu können, führt man noch zwei Ereignisse ein. Diese Ereignisse sind das *sichere Ereignis*  $\Omega$  und das *unmögliche Ereignis*  $\emptyset$ , wobei das sichere Ereignis immer eintritt, das unmögliche Ereignis niemals. Bei der geometrischen Interpretation von Ereignissen als Menge von Punkten entspricht dem sicheren Ereignis  $\Omega$  die Menge *aller* Punkte, das unmögliche Ereignis ist die *leere Menge*. Speziell gilt für das Rechnen mit diesen Ereignissen, falls  $A$  ein beliebiges Ereignis ist:

$$A \cap \Omega = A \quad A \cup \Omega = \Omega$$

$$A \cap \emptyset = \emptyset \quad A \cup \emptyset = A$$

$$\bar{\Omega} = \emptyset \quad \bar{\emptyset} = \Omega.$$

Leicht lässt sich auch die Richtigkeit der folgenden Beziehungen prüfen:

$$A \cap A = A, \quad A \cup A = A, \quad A \cap \bar{A} = \emptyset, \quad A \cup \bar{A} = \Omega.$$

## 2. Heuristische Einführung der Wahrscheinlichkeit

In diesem und im nächsten Abschnitt werden wir definieren, was die Wahrscheinlichkeit eines Ereignisses ist. Im täglichen Leben hört man häufig solche Sätze wie: *Es ist ziemlich sicher, dass es heute regnen wird* oder *Es ist unwahrscheinlich, dass*

*ich eine Gehaltserhöhung bekomme.* Solche Aussagen charakterisieren qualitativ die Wahrscheinlichkeit, dass ein zufälliges Ereignis eintritt oder nicht. Für viele Anwendungen in den Wissenschaften sind solche qualitativen Aussagen zu ungenau. Man möchte ganz gerne die Wahrscheinlichkeit *messen*, mit der gewisse Ereignisse eintreten oder nicht eintreten. Dieses Maß wird im folgenden mit  $\mathbb{P}$  bezeichnet, denn Wahrscheinlich(keit) heißt im lateinischen *probabilis*. In diesem Abschnitt wollen wir die ersten Maße beschreiben, die es erlauben, die Wahrscheinlichkeiten des Eintretens eines Ereignisses zu messen. Diese Methoden sollen dann im nächsten Abschnitt systematisch ausgebaut werden.

Die Stochastik hat ihren Ursprung in der Analyse von Glücksspielen. Man wollte schon frühzeitig wissen, wie hoch die Gewinnerwartungen bei solchen Spielen sind. Die Wahrscheinlichkeit, die dafür verwendet wurde, wird als *klassische Wahrscheinlichkeit* bezeichnet. Sie geht von speziellen Voraussetzungen für die zu betrachtenden Ereignisse aus. Man bezeichnet ein Ereignis als *atomar*, wenn es nicht mehr in echte Teilereignisse zerlegt werden kann. Betrachtet man als Beispiel den zufälligen Versuch des Werfens mit einem Würfel, so sind die atomaren Ereignisse gegeben durch das Würfeln der Zahlen von Eins bis Sechs. Ein nicht atomares Ereignis wäre eine gerade Zahl zu würfeln. Dieses Ereignis enthält die *kleineren* Ereignisse  $\{2\}$ ,  $\{4\}$ ,  $\{6\}$ . Folgende Voraussetzungen müssen erfüllt sein, damit man die Definition der klassischen Wahrscheinlichkeit anwenden kann:

- Es gibt nur endlich viele atomare Ereignisse.
- Es gibt keinen Grund zur Annahme, dass eines dieser zufälligen Ereignisse bevorzugt auftritt. Die atomaren Ereignisse sind also gleichberechtigt.

Das Werfen eines fairen Würfels ist ein Beispiel, bei dem die Voraussetzungen für die klassische Wahrscheinlichkeit erfüllt sind. In der Tat gibt es keinen Grund zur Annahme, dass eine Sechs durchschnittlich häufiger auftritt als eine Eins.

DEFINITION 2.7. *Angenommen, die obigen Voraussetzungen sind erfüllt. Dann bezeichnen wir als klassische Wahrscheinlichkeit eines Ereignisses  $A$  den Quotienten*

$$\mathbb{P}(A) = \frac{\text{Anzahl der verschiedenen atomaren Ereignisse, die } A \text{ ausfüllen}}{\text{Gesamtanzahl der atomaren Ereignisse}}.$$

Es ist klar, dass  $\mathbb{P}(A)$  eine Zahl zwischen Null und Eins ist. Als einfaches Beispiel betrachten wir das Würfeln mit einem fairen Würfel. Sei  $A$  das Ereignis, eine gerade Zahl zu würfeln, dann setzt es sich aus drei atomaren Ereignissen zusammen. Insgesamt gibt es sechs atomare Ereignisse. Damit ist die klassische Wahrscheinlichkeit von  $A$

$$\mathbb{P}(A) = \frac{3}{6} = 0.5.$$

Dieses Ergebnis spiegelt wider, dass man in 50% der Würfe durchschnittlich eine gerade Zahl erhalten wird.

Wir betrachten zwei weitere Beispiele: Wir fragen nach der Wahrscheinlichkeit, beim Lotto *6 aus 49* einen *Sechser* und einen *Vierer* zu gewinnen. Ein atomares Ereignis besteht darin, dass genau eine Kombination von sechs Zahlen gezogen wird.

Also ist die Anzahl aller atomaren Ereignisse gleich  $K_{n,k} = \binom{n}{k} = 13983816$ , wobei  $n = 49$  und  $k = 6$  ist. Die Wahrscheinlichkeit, dass genau ein Tippschein die zufällig gezogenen sechs richtigen Zahlen enthält ist gleich

$$\mathbb{P}(\text{Sechser}) = \frac{1}{13983816} \approx 7.15 \cdot 10^{-8}.$$

Um die Wahrscheinlichkeit für einen Vierer zu berechnen, müssen wir die Anzahl der atomaren Ereignisse bestimmen, die den Vierer ausfüllen. Dazu wollen wir ohne Einschränkung der Allgemeinheit annehmen, dass die gezogenen sechs Zahlen die Zahlen von Eins bis Sechs sind. So könnte eine Möglichkeit für einen Vierer gegeben sein durch

$$(1, 2, 3, 4, *, *),$$

wobei \* für eine *nichtrichtige Zahl*, also eine Zahl von 7 bis 49, steht. Die Anzahl dieser Vierer beträgt  $\binom{43}{2} = 903$ . Entsprechend würde es auch 903 Vierer mit festgelegten vier Richtigen

$$(1, 2, 3, 5, *, *),$$

und so weiter geben. Da es insgesamt  $\binom{6}{4} = 15$  verschiedene Vierer gibt und die Anzahl der verschiedenen Tippmöglichkeiten 13983816 beträgt, erhält man

$$\mathbb{P}(\text{Vierer}) = \frac{903 \cdot 15}{13983816} = 0.00096.$$

Ein weiteres Beispiel ist das Werfen zweier unterscheidbarer Würfel. Wir nehmen an, dass der eine Würfel rot und der andere Würfel blau ist. So würde  $(1, 4)$  das Ereignis charakterisieren, mit dem roten Würfel eine 1 und mit dem blauen Würfel eine 4 zu würfeln. Die atomaren Ereignisse sind die geordneten Paare der Zahlen von 1 bis 6. Die Anzahl der stochastisch gleichberechtigten atomaren Ereignisse ist 36. Das Ereignis  $A$ , das die Augenzahl des roten Würfels gleich der Augenzahl des blauen Würfels ist, beträgt damit

$$\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6},$$

denn es gilt:

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\},$$

was bedeutet, dass  $A$  aus sechs atomaren Ereignissen besteht.

Der eben eingeführten Methode, Wahrscheinlichkeiten auf theoretische Weise zu bestimmen, soll nun eine Methode gegenübergestellt werden, die als *empirisch* angesehen werden kann. Solche Voraussetzungen, wie für die Einführung der klassischen Wahrscheinlichkeit erforderlich sind, müssen jetzt nicht gemacht werden. Dazu betrachten wir das Ereignis  $A$ , welches das Ergebnis eines zufälligen Versuches sein kann. Der Versuch werde nun  $n$  mal hintereinander ausgeführt. Dabei zählen wir, wie häufig  $A$  eintritt und berechnen daraus die relative Häufigkeit

$$h_n(A) = \frac{\text{Anzahl des Auftretens von } A \text{ bei } n \text{ Versuchen}}{n}.$$

Wir interessieren uns nun für  $h_n(A)$ , falls  $n$  gegen unendlich konvergiert. Das Diagramm (2) zeigt die relative Häufigkeit des Ereignis *Zahl* beim Münzwurfversuch in Abhängigkeit von  $n$ .

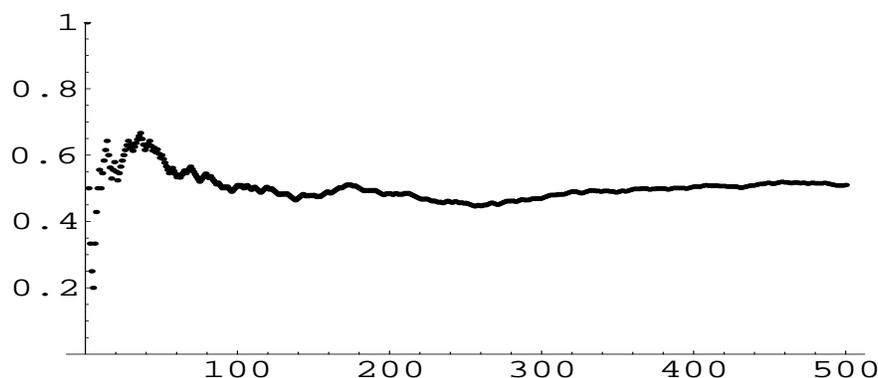


ABBILDUNG 6. Relative Häufigkeit für das Ereignis *Zahl* in Abhängigkeit von  $n$ .

Man sieht, dass sich nach Schwankungen des Wertes für die relative Häufigkeit, ein fester Wert für einen großen Wert von  $n$  einstellt. Dieses Verhalten von  $h_n(A)$  kann als Konvergenz interpretiert werden.

DEFINITION 2.8. *Der Grenzwert von  $h_n(A)$  für  $n \rightarrow \infty$  ist die Wahrscheinlichkeit von  $A$ :*

$$\lim_{n \rightarrow \infty} h_n(A) = \mathbb{P}(A).$$

Der Grenzwert  $\mathbb{P}(A)$  idealisiert das Verhalten von  $h_n(A)$ , in dem Sinn, dass sich für  $h_n(A)$  für einen großen Wert  $n$  ein fester Durchschnittswert einstellt. Für das Ereignis *Zahl* beim Würfeln ist dieser Wert 0.5.

Die so gebildete Wahrscheinlichkeit wird als *statistische Wahrscheinlichkeit* bezeichnet.

Die Definition der statistischen Wahrscheinlichkeit kann an folgendem Beispiel verdeutlicht werden. Betrachtet man die Hochrechnungen im Verlaufe eines Wahlabends, so werden diese mit zunehmender Zahl der ausgewerteten Stimmabgaben immer genauer. Die Schwankungen gegenüber der endgültigen Stimmenverteilung flachen im Laufe der Zeit immer mehr ab, was als Konvergenz interpretiert werden kann.

### 3. Die axiomatische Definition einer Wahrscheinlichkeit

Die im letzten Abschnitt eingeführte *klassische Wahrscheinlichkeit* hat den Nachteil, dass sie an sehr spezielle Voraussetzungen gebunden ist. Diese Voraussetzungen sollen im folgenden fallen gelassen werden. Wir wollen somit eine sehr allgemein anwendbare Definition der Wahrscheinlichkeit geben. Bevor wir aber diese Definition geben können, müssen wir noch Familien von Ereignissen betrachten, für die die Definition der Wahrscheinlichkeit dann Sinn machen wird.

Wir betrachten die elementaren Werte (Elementarereignisse) eines zufälligen Versuches. Darunter verstehen wir die (Zahlen)werte, die das Ergebnis eines zufälligen Versuches sind. Da es sicher ist, dass bei der Versuchsdurchführung irgendeiner dieser Werte auftritt, ist die Menge *aller* möglichen elementaren Versuchsausgänge

gleich dem sicheren Ereignis  $\Omega$ . Würde beispielsweise ein Versuch in der Bestimmung der Lebensdauer einer Glühbirne einer bestimmten Produktionsserie bestehen, so wäre dieser elementare Wert die Zahl, die der Länge der Brenndauer einer zufällig ausgewählten Glühbirne aus dieser Serie entspricht. Diese Zeit existiert zwar, sie kann aber nicht genau aufgrund von Messungenauigkeiten bestimmt werden. Man ist deswegen auch nur in der Lage, gewisse Intervalle anzugeben, in der die Glühbirne ausfällt beziehungsweise nicht ausfällt. Vom Standpunkt der Qualitätskontrolle ist man an Ereignissen wie *Eine Glühbirne lebt länger als 1000 Stunden* interessiert. Diesem Ereignis entspricht das Intervall  $(1000, \infty)$ . Diese interessierenden Ereignisse und weitere Ereignisse, die notwendig für die Definition einer Wahrscheinlichkeit sind, werden zusammengefasst:

DEFINITION 2.9. *Gegeben sei die Menge  $\Omega$  aller elementaren Ereignisse eines zufälligen Versuches. Eine Teilmenge der Potenzmenge<sup>1</sup> von  $\Omega$  heißt Ereignisfeld  $\mathcal{E}$ , falls gilt:*

- *Das sichere Ereignis  $\Omega \in \mathcal{E}$ .*
- *Falls  $A \in \mathcal{E}$ , dann ist auch  $\bar{A} \in \mathcal{E}$ .*
- *Falls  $A_1, A_2, \dots$  abzählbar viele dieser Ereignisse aus  $\mathcal{E}$  sind, dann ist auch die Vereinigung*

$$\bigcup_{i=1}^{\infty} A_i$$

*in  $\mathcal{E}$  enthalten.*

In der obigen Definition ist keine Aussage über den Durchschnitt und über das unmögliche Ereignis  $\emptyset$  gemacht worden. Aus diesen Axiomen erhalten wir die Folgerung:

LEMMA 2.10. *Falls  $A_1, A_2, \dots \in \mathcal{E}$ , dann ist auch*

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{E}.$$

*Außerdem ist das unmögliche Ereignis  $\emptyset \in \mathcal{E}$ .*

BEWEIS. Es gilt nach den De Morganschen Regeln:

$$\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \bar{A}_i} \in \mathcal{E}.$$

Da die Komplemente von Ereignissen aus  $\mathcal{E}$  in  $\mathcal{E}$  liegen, gehören die  $\bar{A}_i$  zu  $\mathcal{E}$ . Weiterhin ist deren Vereinigung in  $\mathcal{E}$  enthalten.

Wegen  $\Omega \in \mathcal{E}$  ist auch  $\bar{\Omega} = \emptyset \in \mathcal{E}$ . □

**Beispiel:** Wir betrachten eine Anzahl von gleichgroßen Kugeln, die in einer nicht-einsehbaren Kiste liegen. Diese Kugeln sind von 1 beginnend durchnummeriert. Weiterhin sind diese Kugeln unterschiedlich gefärbt. Speziell sind in dieser Kiste  $s$  weiße,  $t$  schwarze,  $u$  rote und  $v$  grüne Kugeln, wobei die Kugeln mit den Nummern von Eins bis  $s$  die weißen Kugeln, von  $s + 1$  bis  $s + t$  die schwarzen, und so weiter, sind. Insgesamt sind also  $n = s + t + u + v$  Kugeln in der Kiste.

Der zufällige Versuch besteht in der zufälligen Entnahme einer Kugel, wobei wir

<sup>1</sup>Die Potenzmenge ist die Menge aller Teilmengen von  $\Omega$ , einschließlich  $\Omega$  und der leeren Menge  $\emptyset$ .

uns die *Nummer* der entnommenen Kugel notieren. Die Elementarereignisse sind also die Nummern der gezogenen Kugeln. Das elementare Ereignis des Ziehens der Kugel  $i$  werde mit  $\omega_i$  bezeichnet. Die Menge aller Elementarereignisse ist dann gegeben durch

$$\{\omega_1, \omega_2, \dots, \omega_n\} = \Omega.$$

Das Ereignisfeld besteht in diesem Fall aus der Menge aller Teilmengen von  $\Omega$ , also der Potenzmenge  $2^\Omega$ . Ein Element dieses Ereignisfeldes ist das Ereignis, eine Kugel mit ungerader Numerierung zu ziehen. Ein anderes Ereignis wäre, die Kugel mit der Nummer 1 zu ziehen, oder eine rote Kugel zu ziehen.

Wir führen den Versuch noch einmal durch, wobei jetzt nicht die Nummer der gezogenen Kugel, sondern deren Farbe interessiert. Die Ereignisse des Ereignisfeldes sind jetzt gegeben durch:

$$\begin{aligned} A &= \{\omega_1, \dots, \omega_s\}, \\ B &= \{\omega_{s+1}, \dots, \omega_{s+t}\}, \\ C &= \{\omega_{s+t+1}, \dots, \omega_{s+t+u}\}, \\ D &= \{\omega_{s+t+u+1}, \dots, \omega_{s+t+u+v}\}. \end{aligned}$$

In diesem Fall besteht das Ereignisfeld aus

$$\{\emptyset, A, B, C, D, A \cup B, A \cup C, A \cup D, B \cup C, B \cup D, C \cup D, A \cup B \cup C, A \cup B \cup D, A \cup C \cup D, B \cup C \cup D, \Omega\}.$$

(Man überprüfe diese Aussage.) Speziell zeigt das letzte Beispiel, dass die elementaren Ereignisse nicht unbedingt zum Ereignisfeld gehören müssen.

Wir sind nun in der Lage, die mathematischen Eigenschaften eines Maßes anzugeben, das zur Charakterisierung der durchschnittlichen Häufigkeit des Eintretens von zufälligen Ereignissen dienen soll.

DEFINITION 2.11. *Eine Funktion  $\mathbb{P}$ , definiert auf einem Ereignisfeld  $\mathcal{E}$  und einem Wertevorrat enthalten in  $[0, 1]$ , heißt Wahrscheinlichkeit, falls*

$$(1) \quad \mathbb{P}(\Omega) = 1.$$

*Es seien  $A_1, A_2, \dots$  abzählbar unendlich viele paarweise unvereinbare Ereignisse aus  $\mathcal{E}$ . Dann ist*

$$(2) \quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

*Dabei bedeutet paarweise unvereinbar, dass für  $A_i, A_j$  mit Indices  $i \neq j$  die Eigenschaft  $A_i \cap A_j = \emptyset$  gilt.*

Wir betrachten einige elementare Folgerungen aus der Definition der Wahrscheinlichkeit:

FOLGERUNG 2.12. *Es gilt  $\mathbb{P}(\emptyset) = 0$ .*

Man setze  $A_1 = \Omega, A_2 = A_3 = \dots = \emptyset$ . Diese Ereignisse sind paarweise unvereinbar. Somit gilt nach (1) und (2)

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup A_2 \cup A_3 \cup \dots) = \mathbb{P}(\Omega) + \mathbb{P}(A_2) + \mathbb{P}(A_3) + \dots,$$

woraus folgt, dass  $\mathbb{P}(A_2) = 0$  ist.

FOLGERUNG 2.13. Die Gleichheit im Axiom (2) ist auch erfüllt für endlich viele Ereignisse  $A_i$  falls diese unvereinbar sind:

$$\mathbb{P}\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k \mathbb{P}(A_i).$$

Man setze  $A_{k+1}, A_{k+2}, \dots = \emptyset$  und wende  $\mathbb{P}(A_{k+j}) = 0, j = 1, 2, \dots$  an. Speziell gilt für zwei unvereinbare Ereignisse  $A, B$ :

$$(3) \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

FOLGERUNG 2.14. Für jedes Ereignis aus  $\mathcal{E}$  gilt:

$$(4) \quad \mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$$

Es gilt  $A \cup \bar{A} = \Omega, A \cap \bar{A} = \emptyset$ . Wegen der letzten Beziehung kann (3) für  $B = \bar{A}$  angewandt werden.

FOLGERUNG 2.15. Falls für zwei Ereignisse  $A, B$  aus  $\mathcal{E}$  gilt  $A \subset B$  dann ist  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

Man benutze  $B = A \cup (B \setminus A)$ , wobei  $A \cap (B \setminus A) = \emptyset$  ist. Dann wende man (3) an.

Die Axiome und die elementaren Folgerungen reflektieren die elementaren Eigenschaften des Messens. Misst man beispielsweise die Relativgewichte zweier unterschiedlicher Massen eines Systems, so ist die Gesamtmasse gleich der Summe der Massen der beiden Teilmassen, das heißt es gilt (3).

Wir wollen untersuchen, ob die von uns eingeführten Wahrscheinlichkeiten den obigen Axiomen entsprechen. Zuerst betrachten wir die *klassische Wahrscheinlichkeit*. Wir nehmen an, dass ein zufälliger Versuch endlich viele gleichberechtigte elementare Ereignisse besitzt. Dann können wir annehmen, dass die atomaren Ereignisse durch die elementaren Ereignisse gegeben sind. Als Ereignisfeld wählen wir die Potenzmenge der atomaren Ereignisse. Da

$$\mathbb{P}(\Omega) = \frac{\text{Anzahl aller atomaren Ereignisse}}{\text{Anzahl aller atomaren Ereignisse}} = 1,$$

ist das erste Axiom einer Wahrscheinlichkeit erfüllt. Das Additionsaxiom betrachten wir nur für zwei Ereignisse. Für endlich viele Ereignisse oder für abzählbar unendlich viele Ereignisse gilt es natürlich auch. Im letzten Fall müssen dann aber außer endlich vielen Ereignissen, unendlich viele Ereignisse als *unmögliches Ereignis* angesetzt werden. Für die Vereinigung von zwei Ereignissen, die unvereinbar sind, ergibt sich die Anzahl der enthaltenen atomaren Ereignisse als Summe der atomaren Ereignisse der einzelnen Ereignisse. Das heißt, falls  $\#A$  die Anzahl der atomaren Ereignisse in  $A$  bezeichnet:

$$\#(A_1 \cup A_2) = \#A_1 + \#A_2.$$

Daraus ergibt sich trivialerweise die Additionsbeziehung für Wahrscheinlichkeiten.

Ähnlich für die *statistische Wahrscheinlichkeit*: Da für jedes  $n$  gilt:

$$h_n(\Omega) = 1$$

gilt auch für den Grenzwert

$$\lim_{n \rightarrow \infty} h_n(\Omega) = 1 = \mathbb{P}(\Omega).$$

Weiterhin gilt für zwei unvereinbare Ereignisse  $A, B$ :

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

Beim Grenzübergang für  $n \rightarrow \infty$  bleibt diese Beziehung erhalten. Damit gilt das Additionsaxiom.

Wir definieren ein Wahrscheinlichkeitsmaß auf der Menge der reellen Zahlen  $\mathbb{R}$ . Es sei  $f(x)$  eine Funktion mit nicht negativen Werten und mit Definitionsbereich  $\mathbb{R}$ . Der Flächeninhalt unter der Kurve sei Eins:

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Für ein Intervall aus  $\mathbb{R}$  bilden wir eine Maßzahl wie folgt:

$$\mathbb{P}(I) = \text{Fläche über dem Intervall unter der Funktion.}$$

Dann ist  $\mathbb{P}(\mathbb{R}) = 1$  und der Flächeninhalt über zwei disjunkten Teilintervallen ist gleich der Summe der Flächeninhalte über beiden Intervallen.

Wir haben die Frage offen gelassen, ob die Intervalle ein Ereignisfeld bilden. Das ist offenbar nicht so, denn die Vereinigung zweier Intervalle ist im Allgemeinen kein Intervall. Als Ereignisfeld wählt man die sogenannte Borel- $\sigma$ -Algebra. Das ist das kleinste Ereignisfeld, das alle Intervalle enthält, siehe [2] Seite 165.

Der folgende Satz erlaubt eine Verallgemeinerung von (3), falls die Voraussetzung, dass  $A, B$  unvereinbar sind, nicht erfüllt ist.

**SATZ 2.16. (Additionssatz)** *Gegeben seien zwei beliebige Ereignisse  $A, B$  aus  $\mathcal{E}$ . Dann gilt:*

$$(5) \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

**BEWEIS.** Es sei  $C := A \setminus (A \cap B)$ . Wegen der Definition von  $C$  ist der Durchschnitt von  $C$  und der Menge  $B$  leer (siehe Abbildung 7).

Es gilt  $C \cup (A \cap B) = A$ ,  $C \cap B = \emptyset$ . Da  $A \cap B \subset B$  ist auch

$$C \cap (A \cap B) = \emptyset$$

und weiterhin  $C \cup B = A \cup B$ . Damit gilt nach (3)

$$\mathbb{P}(A) = \mathbb{P}(C \cup (A \cap B)) = \mathbb{P}(C) + \mathbb{P}(A \cap B)$$

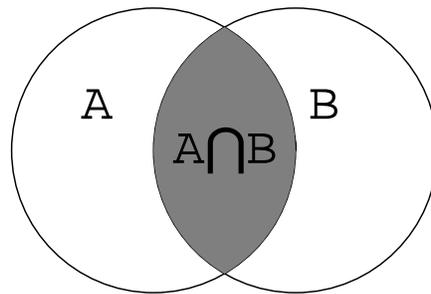


ABBILDUNG 7.  $C = A \setminus B = A \setminus (A \cap B)$

und somit

$$\mathbb{P}(C) = \mathbb{P}(A) - \mathbb{P}(A \cap B).$$

Ähnlich erhält man

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

□

Man überlege, wie der Additionssatz für drei oder mehr Ereignisse

$$\mathbb{P}(A \cup B \cup C)$$

hergeleitet werden kann.

**Beispiel:** Es sei  $A_1$  das Ereignis, dass eine Person regelmäßig ins Kino geht und  $A_2$ , dass eine Person regelmäßig TV schaut. Bei einer statistischen Erhebung wurde festgestellt, dass eine Person mit Wahrscheinlichkeit von 0.2 regelmäßig ins Kino geht und regelmäßig TV sieht (Ereignis  $A_1 \cap A_2$ ). Mit einer Wahrscheinlichkeit von 0.6 sieht eine Person nicht regelmäßig Fernsehen und mit einer Wahrscheinlichkeit von 0.3 geht sie regelmäßig ins Kino. Man berechne die Wahrscheinlichkeit von  $A_2$ ,  $A_1 \cap \bar{A}_2$  und  $A_1 \cup A_2$ .

Es gilt:

$$\mathbb{P}(A_2) = 1 - \mathbb{P}(\bar{A}_2) = 1 - 0.6 = 0.4.$$

Desweiteren gilt:

$$(A_1 \cap \bar{A}_2) \cup (A_1 \cap A_2) = A_1, \quad (A_1 \cap \bar{A}_2) \cap (A_1 \cap A_2) = \emptyset.$$

Damit gilt nach dem Additionsaxiom

$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap \bar{A}_2), \\ \mathbb{P}(A_1 \cap \bar{A}_2) &= \mathbb{P}(A_1) - \mathbb{P}(A_1 \cap A_2) = 0.3 - 0.2 = 0.1. \end{aligned}$$

Nach dem Additionssatz erhält man

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) = 0.3 + 0.4 - 0.2 = 0.5.$$

Es kann auch noch  $\mathbb{P}(\bar{A}_1 \cup \bar{A}_2)$  ausgerechnet werden. Es gilt nach den De Morganschen Regeln:

$$\mathbb{P}(\bar{A}_1 \cup \bar{A}_2) = \mathbb{P}(\overline{A_1 \cap A_2}) = 1 - \mathbb{P}(A_1 \cap A_2) = 1 - 0.2 = 0.8.$$

Später benötigen wir auch noch die folgende Definition:

DEFINITION 2.17. *Wir nennen eine Folge von Ereignissen  $(A_i)_{i \in \mathbb{N}}$  monoton fallend, falls gilt:*

$$A_1 \supset A_2 \supset A_3 \supset \dots.$$

*Eine Folge von Ereignissen  $(A_i)_{i \in \mathbb{N}}$  wird monoton wachsend genannt, falls gilt:*

$$A_1 \subset A_2 \subset A_3 \subset \dots.$$

Die im folgenden Satz formulierte Eigenschaft der Wahrscheinlichkeit heißt *Stetigkeit der Wahrscheinlichkeit*.

SATZ 2.18. *Gegeben sei eine monoton fallende Folge von Ereignissen  $(A_i)_{i \in \mathbb{N}}$ . Dann gilt*

$$\mathbb{P}\left(\bigcap_{i \in \mathbb{N}} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i).$$

*Sei nun die Folge  $(A_i)_{i \in \mathbb{N}}$  monoton wachsend, dann ist*

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i).$$

## Die bedingte Wahrscheinlichkeit

### 1. Die Definition der bedingten Wahrscheinlichkeit

Häufig werden Wahrscheinlichkeiten benötigt, für die gewisse Bedingungen vorausgesetzt werden. Spezielle Anwendungen finden sich bei den Lebensversicherungen. Um die Versicherungsraten errechnen zu können, die ein Versicherter zahlen muss, müssen aus Statistiken zum Beispiel folgende Wahrscheinlichkeiten ermittelt werden:

*Wahrscheinlichkeit, dass ein Mann 70 wird, unter der Bedingung, dass er 30 geworden ist.*

Es wird also nicht die Wahrscheinlichkeit ermittelt, dass irgendein Mann 70 wird. Es interessiert die Wahrscheinlichkeit, dass irgendein Mann, der die Bedingung erfüllt 30 zu sein, 70 wird.

Entsprechendes statistisches Zahlenmaterial ist in den sogenannten *Sterbetafeln* zusammengefasst, die zum Handwerkszeug jedes Versicherungsmathematikers gehören.

**DEFINITION 3.1.** *Es sei  $A \in \mathcal{E}$  ein Ereignis,  $\mathbb{P}(A) > 0$ . Die bedingte Wahrscheinlichkeit von  $B \in \mathcal{E}$  ist definiert durch*

$$(6) \quad \mathbb{P}(B|A) = \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

$\mathbb{P}(B|A)$  wird als *Wahrscheinlichkeit von  $B$  unter der Bedingung von  $A$*  bezeichnet. Die bedingte Wahrscheinlichkeit bringt eine gewisse Anteiligkeit zum Ausdruck. Diese Anteiligkeit (Abbildung 1) erlaubt es, eine neue Wahrscheinlichkeit zu definieren. Betrachtet man  $\mathbb{P}(\cdot|A)$ , wobei für  $\cdot$  ein beliebiges Ereignis aus dem Ereignisfeld  $\mathcal{E}$  eingesetzt werden kann, so definiert der Ausdruck  $\mathbb{P}(\cdot|A)$  eine Funktion auf den Ereignissen aus  $\mathcal{E}$ , die alle Voraussetzungen der Definition 2.11 erfüllt. Es gilt also das Additionsaxiom (2) und  $\mathbb{P}(\Omega|A) = 1$ . Da aber diese Axiome für die bedingte Wahrscheinlichkeit erfüllt sind, können auch alle anderen allgemeinen Eigenschaften einer Wahrscheinlichkeit gefolgert werden, wie zum Beispiel der Additionssatz

$$\mathbb{P}(C \cup B|A) = \mathbb{P}(C|A) + \mathbb{P}(B|A) - \mathbb{P}(C \cap B|A).$$

Wir haben bisher nur die Wahrscheinlichkeiten von Vereinigungen untersuchen können. Dazu haben wir den Additionssatz oder das Additionsaxiom benutzt. Aufgrund der Definition der bedingten Wahrscheinlichkeit sind wir in der Lage, die Wahrscheinlichkeit für den Durchschnitt berechnen zu können.

**SATZ 3.2.** *Falls  $\mathbb{P}(A) > 0$ , so gilt:*

$$(7) \quad \mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A).$$

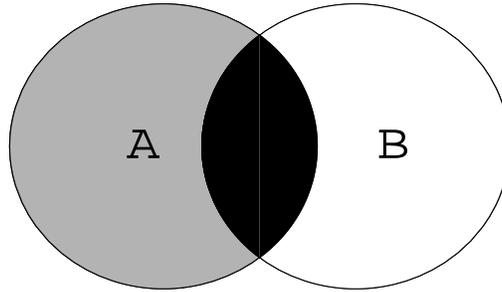


ABBILDUNG 1. Anteil der schwarzen Fläche  $A \cap B$  an  $A$

Falls auch noch  $\mathbb{P}(B) > 0$  gilt, so ist

$$(8) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B).$$

Diese Formeln erhält man ganz einfach durch Umstellen der obigen Definitionsformel (6) der bedingten Wahrscheinlichkeit. Satz 3.2 wird Produktsatz genannt.

Wir wollen nun als Verallgemeinerung von Satz 3.2 eine Formel für

$$\mathbb{P}(A \cap B \cap C)$$

herleiten. Aufgrund von Satz 3.2 gilt:

$$(9) \quad \mathbb{P}(A \cap B \cap C) = \mathbb{P}((B \cap C) \cap A) = \mathbb{P}(B \cap C|A) \cdot \mathbb{P}(A),$$

falls  $\mathbb{P}(A) > 0$ . Wir benutzen für die bedingte Wahrscheinlichkeit  $\mathbb{P}(\cdot|A)$  das Symbol  $\mathbb{P}_A(\cdot)$ . Da  $\mathbb{P}_A$  alle Eigenschaften einer Wahrscheinlichkeit besitzt, gilt natürlich auch der Produktsatz für die bedingten Wahrscheinlichkeiten:

$$\mathbb{P}_A(B \cap C) = \mathbb{P}_A(B) \cdot \mathbb{P}_A(C|B).$$

Für  $\mathbb{P}_A(C|B)$  können wir somit schreiben:

$$\begin{aligned} \mathbb{P}_A(C|B) &= \frac{\mathbb{P}_A(B \cap C)}{\mathbb{P}_A(B)} \\ &= \frac{\frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A)}}{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A \cap B)} = \mathbb{P}(C|A \cap B) \end{aligned}$$

falls  $\mathbb{P}(A \cap B) =$ . Fassen wir alle Formeln zusammen, so erhalten wir wegen  $\mathbb{P}_A(B) = \mathbb{P}(B|A)$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B).$$

**Beispiel:** Eine Firma baut Computer. Es bezeichne  $A_1$  das Ereignis, dass ein produzierter Computer nicht funktionstüchtig ist. Die Wahrscheinlichkeit dafür sei 0.2. Diese nicht funktionstüchtigen Computer müssen nachgebessert werden. Es sei  $A_2$  das Ereignis, dass ein Computer nach der ersten Nachbesserung nicht betriebsbereit ist und entsprechend  $A_3$  sei das Ereignis, dass ein Computer nach der ersten und zweiten Nachbesserung nicht betriebsbereit ist. 12% der einmal nachgebesserten Computer sind nicht betriebsbereit und 5% der zweimal nachgebesserten Computer sind nicht betriebsbereit. Wie groß ist die Wahrscheinlichkeit, dass mehr als

zwei Nachbesserungen notwendig sind?

Das dazu gehörige Ereignis kann wie folgt ausgedrückt werden:

$$A_1 \cap A_2 \cap A_3.$$

Die Prozentzahlen in der Aufgabe beziehen sich auf den Anteil, der im vorangehenden Produktionsschritt noch defekten Computer. Sie können als bedingte Wahrscheinlichkeiten  $\mathbb{P}(A_2|A_1)$  und  $\mathbb{P}(A_3|A_2 \cap A_1)$  interpretiert werden. Nach (9) gilt

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_2 \cap A_1) = 0,0012 = 0,2 \cdot 0,12 \cdot 0,05.$$

Die folgende Anwendungen können mittels des Produktsatzes hergeleitet werden.

## 2. Einige Sätze über die bedingte Wahrscheinlichkeit

Die im letzten Abschnitt hergeleiteten Eigenschaften für die Wahrscheinlichkeit des Produktes zweier Ereignisse erlauben uns den *Satz von der bedingten Wahrscheinlichkeit* zu formulieren. Um den Satz von der totalen Wahrscheinlichkeit formulieren zu können, benötigen wir die Definition eines vollständigen Systems von Ereignissen:

DEFINITION 3.3. Eine Menge von Ereignissen  $(A_i)_{i=1,\dots,n}$  wird vollständiges System von Ereignissen genannt, falls gilt:

$$\bigcup_{i=1}^n A_i = \Omega, \quad A_i \cap A_j = \emptyset, \text{ falls } i \neq j, \quad \mathbb{P}(A_i) \neq 0.$$

Darauf aufbauend gilt der Satz von der totalen Wahrscheinlichkeit:

SATZ 3.4. Es sei  $(A_i)_{i=1,\dots,n}$  ein vollständiges System von Ereignissen mit einer Wahrscheinlichkeit  $\mathbb{P}(A_i) > 0$ . Dann gilt für ein beliebiges Ereignis  $B$ :

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

BEWEIS. Da  $(A_i)_{i=1,\dots,n}$  ein vollständiges System von Ereignissen ist, kann  $B$  wie folgt geschrieben werden

$$B = \bigcup_{i=1}^n (B \cap A_i), \quad (B \cap A_i) \cap (B \cap A_j) = \emptyset \text{ für } i \neq j.$$

Damit gilt nach dem Additionsaxiom (2) und der Definition der bedingten Wahrscheinlichkeit

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

□

**Beispiel:** Bei der Gütekontrolle wird ein Ausschussteil mit Wahrscheinlichkeit von 0.9 erkannt. Ein einwandfreies Teil wird mit Wahrscheinlichkeit von 0.85 als solches erkannt. Der Ausschussprozentsatz beträgt 5% (Ereignis  $A_2$ ). Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Teil als fehlerhaft eingestuft wird? Das Ereignis, dass ein Teil als fehlerhaft eingestuft wird, sei mit  $B$  bezeichnet.

Falls  $A_1$  das Ereignis *kein Ausschuss* bezeichnet, so ist  $(A_i)_{i=1,2}$  ein vollständiges System von Ereignissen. Weiterhin gilt:

$$\begin{aligned}\mathbb{P}(B|A_1) &= 1 - \mathbb{P}(\bar{B}|A_1) = 1 - 0.85 = 0.15 \\ \mathbb{P}(B|A_2) &= 0.9.\end{aligned}$$

Nach dem Satz von der totalen Wahrscheinlichkeit folgt somit:

$$\mathbb{P}(B) = \mathbb{P}(B|A_1)\mathbb{P}(A_1) + \mathbb{P}(B|A_2)\mathbb{P}(A_2) = 0.1875.$$

Der folgende Satz erlaubt die Berechnung von bedingten Wahrscheinlichkeiten aus anderen bedingten Wahrscheinlichkeiten, wobei sich Ereignis und Bedingung austauschen.

**SATZ 3.5.** (Bayes Formel) *Es sei  $(A_i)_{i=1,\dots,n}$  ein vollständiges System von Ereignissen mit  $\mathbb{P}(A_i) > 0$  und  $B$  ein Ereignis mit  $\mathbb{P}(B) > 0$ . Dann gilt:*

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

Dieser Satz erlaubt also  $\mathbb{P}(A_i|B)$  zu berechnen, falls  $\mathbb{P}(B|A_j)$ ,  $j = 1, \dots, n$  bekannt ist. Der Term im Nenner im letzten Satz ist nach dem Satz der totalen Wahrscheinlichkeit gleich  $\mathbb{P}(B)$ .

**Beispiel:** Bezugnehmend auf das letzte Beispiel interessiert die Wahrscheinlichkeit, dass sich ein einwandfreies Teil unter den als mangelhaft eingeschätzten Teilen befindet. Nach der Bayesschen Formel gilt:

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(B|A_1)\mathbb{P}(A_1)}{\mathbb{P}(B)} = 0.76.$$

### 3. Unabhängigkeit von Ereignissen

Es seien  $A, B$  zwei Ereignisse. Diese Ereignisse werden als stochastisch *unabhängig* bezeichnet, falls das Eintreten von  $A$  nicht die Wahrscheinlichkeit des Eintretens von  $B$  beeinflusst. Diese Beziehung kann mittels der Definition der bedingten Wahrscheinlichkeit auch wie folgt formuliert werden.

**DEFINITION 3.6.** *Gilt für ein Ereignis  $A$  die Beziehung  $\mathbb{P}(A) \neq 0$ , dann sind  $A$  und  $B$  unabhängig, falls*

$$(10) \quad \mathbb{P}(B) = \mathbb{P}(B|A).$$

Durch Anwendung von (7) erhält man aus (10)

$$(11) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Diese Beziehung ist auch richtig, falls  $\mathbb{P}(A), \mathbb{P}(B) = 0$  sind. Man nennt (10) den *Produktsatz für unabhängige Ereignisse*. Da man den Durchschnitt von Ereignissen auch als Produkt von Ereignissen bezeichnen kann, kann diese Version des (allgemeinen) Produktsatzes auch wie folgt formuliert werden:

*Die Wahrscheinlichkeit des Produktes zweier Ereignisse ist gleich dem Produkt der Wahrscheinlichkeiten.*

In der obigen Formulierung der Beziehung *Unabhängigkeit* ist kein Unterschied zwischen  $A$  und  $B$  zu erkennen, obwohl in der Formel (10)  $A$  und  $B$  nicht symmetrisch

auftreten. In der Tat erhält man, falls  $\mathbb{P}(B) \neq 0$  gilt, durch Umstellen der in (10) vorkommenden bedingten Wahrscheinlichkeit nach  $\mathbb{P}(A)$ :

$$\mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B).$$

Somit ist falls  $B$  unabhängig von  $A$  auch  $A$  unabhängig von  $B$ . Damit ist also die Sprechweise *A und B sind unabhängig*, die die Gleichberechtigung von  $A$  und  $B$  bezüglich stochastischen Unabhängigkeit ausdrückt, gerechtfertigt.

Falls zwei Ereignisse nicht unabhängig sind, werden sie als abhängig bezeichnet. Ähnlich wie bei der Unabhängigkeit ist  $A$  abhängig von  $B$  genau dann, wenn  $B$  abhängig von  $A$  ist. Gewisse Kausalitäten zwischen  $A$  und  $B$  können nicht durch diese stochastische Definition der Unabhängigkeit beschrieben werden.

Die Unabhängigkeit von mehr als zwei Ereignissen wird wie folgt definiert:

**DEFINITION 3.7.** *Die Ereignisse  $A_1, \dots, A_n$  heißen unabhängig voneinander, falls für jede aus mindestens zwei Ereignissen bestehende Teilmenge  $A_{n_1}, \dots, A_{n_j}$ ,  $1 \leq j \leq n$ ,  $n_i \in \{1, \dots, n\}$  von  $A_1, \dots, A_n$  gilt:*

$$\mathbb{P}(A_{n_1} \cap \dots \cap A_{n_j}) = \mathbb{P}(A_{n_1}) \cdot \dots \cdot \mathbb{P}(A_{n_j}).$$

Unabhängigkeit spielt beim Verständnis vieler Modelle aus Natur- und Ingenieurwissenschaften eine wichtige Rolle. Wir betrachten den Weg eines Moleküls. Da so ein Molekül ständig mit anderen Molekülen zusammenstößt, geht sehr schnell jede Information über die Wegrichtung verloren. Kleine Wegstückchen sind also unabhängig voneinander.

**BEMERKUNG 3.8.** Sind  $A, B$  unabhängig, so sind auch  $A, \bar{B}$  oder  $\bar{A}, B$  oder  $\bar{A}, \bar{B}$  unabhängig.

**Beispiel:** Das Ereignis  $A$  beim Wurf mit zwei Würfeln eine *Sechs* mit dem ersten Würfel zu erzielen, ist unabhängig von dem Ereignis  $B$ , mit dem zweiten Würfel eine *Sechs* zu erreichen. Das ist intuitiv klar, kann aber auch mittels der obigen Definition nachgerechnet werden. Es gilt

$$\begin{aligned} A &= \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\} \\ B &= \{(1, 6), (2, 6), (3, 6), (4, 6), (5, 6), (6, 6)\} \\ A \cap B &= \{(6, 6)\}. \end{aligned}$$

Damit ist

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6} = \mathbb{P}(B).$$

Eine weitere Anwendung bezieht sich auf die Zuverlässigkeit von komplexen Systemen. So ein System setzt sich aus einfachen Teilsystemen zusammen. Die Systemfunktion eines solchen komplexen Systems beschreibt die Funktionstüchtigkeit des komplexen Systems in Abhängigkeit der Funktionstüchtigkeit der Teilsysteme. Die einfachsten Systeme, aus denen man kompliziertere Systeme zusammensetzen kann, sind das serielle und das parallele System (siehe Abbildung 2, 3).

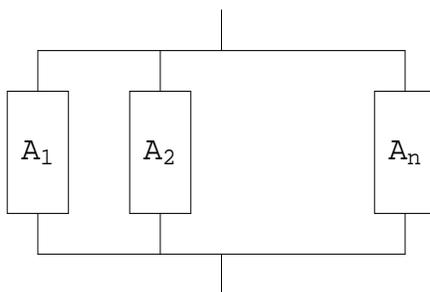


ABBILDUNG 2. Parallelschaltung

Es bezeichne  $A_i$  das Ereignis, dass das Teilsystem  $i$  funktioniert. Desweiteren nehmen wir an, dass die Funktionsweise der einzelnen Systeme untereinander unabhängig ist, siehe Definition 3.7. Es sei  $S$  das Ereignis, dass das Gesamtsystem funktioniert. Dann gilt für das serielle System:

$$S = A_1 \cap \dots \cap A_n, \quad \bar{S} = \bar{A}_1 \cup \dots \cup \bar{A}_n.$$

Für das parallele System erhält man

$$S = A_1 \cup \dots \cup A_n, \quad \bar{S} = \bar{A}_1 \cap \dots \cap \bar{A}_n.$$

Somit gilt für das serielle System 3)

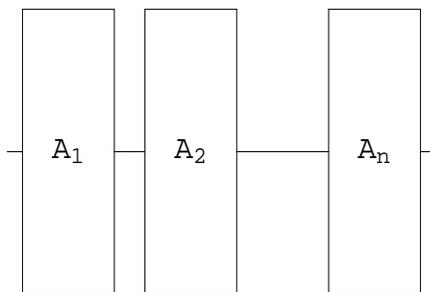


ABBILDUNG 3. serielle Schaltung

$$\mathbb{P}(S) = \prod_{i=1}^n \mathbb{P}(A_i).$$

Für das parallele System könnte  $\mathbb{P}(S)$  mittels des Additionssatzes ausgerechnet werden, wobei die dabei auftretenden Wahrscheinlichkeiten von Durchschnitten mittels Produktsatz gebildet werden können. Schneller geht es wie folgt:

$$\begin{aligned} \mathbb{P}(S) &= 1 - \mathbb{P}(\bar{S}) = 1 - \mathbb{P}(\bar{A}_1) \cdot \dots \cdot \mathbb{P}(\bar{A}_n) \\ &= 1 - (1 - \mathbb{P}(A_1)) \cdot \dots \cdot (1 - \mathbb{P}(A_n)). \end{aligned}$$

## Zufallsvariable und deren Verteilung

### 1. Einführung von Zufallsvariablen

Ergebnisse eines zufälligen Versuches sind Ereignisse, die mit einer gewissen Wahrscheinlichkeit auftreten. Im Folgenden wollen wir die *Wahrscheinlichkeitsverteilung* eines zufälligen Versuches beschreiben. Das bedeutet, wir beschreiben, wie sich die Wahrscheinlichkeitsmasse (die ja Eins ist) auf die einzelnen Ereignisse verteilt.

Es ist im Allgemeinen möglich, die Ergebnisse eines zufälligen Versuches als Zahlenwerte zu interpretieren. Liefert der zufällige Versuch *Münzwurf* die beiden Versuchsausgänge *Kopf* oder *Zahl*, so kann *Kopf* als 0 und *Zahl* als 1 kodiert werden. Diese zahlenmäßige Interpretation eines zufälligen Versuches bezeichnet man als *Zufallsvariable* oder *Zufallsgröße*. Wir werden im weiteren beschreiben, wie sich die Wahrscheinlichkeiten auf den Werten der Zufallsvariablen verteilen. Man spricht deswegen auch von der Verteilung einer Zufallsvariablen.

In der Praxis stellt man fest, dass es Verteilungen für Zufallsvariablen gibt, die in den verschiedensten Situationen immer wieder auftreten. Die bekanntesten dieser Verteilungen sind unter anderem die Normalverteilung (weißes Rauschen, Fehlerrechnung), Exponentialverteilung (Zuverlässigkeitstheorie), Weibull-Verteilung (zufällige Korngrößen von Siebgut, zufällige Ausfallzeiten von Geräten unter Verschleiß), Poisson-Verteilung (zufällige Länge von Warteschlangen) und die hypergeometrische Verteilung (Qualitätskontrolle).

Als eine Anwendung, bei der die Zufallsvariablen eine Rolle spielen, sei auch noch die Simulation erwähnt. Um Simulationen sachgerecht durchführen zu können, muss

Wert $z$ von $Z$	atomare Ereignisse	$\mathbb{P}(Z = z)$
2	(1,1)	$\frac{1}{36}$
3	(1,2), (2,1)	$\frac{2}{36}$
4	(1,3), (2,2), (3,1)	$\frac{3}{36}$
5	(1,4), (2,3), (3,2), (4,1)	$\frac{4}{36}$
6	(1,5), (2,4), (3,3), (4,2), (5,1)	$\frac{5}{36}$
7	(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)	$\frac{6}{36}$
8	(2,6), (3,5), (4,4), (5,3), (6,2)	$\frac{5}{36}$
9	(3,6), (4,5), (5,4), (6,3)	$\frac{4}{36}$
10	(4,6), (5,5), (6,4)	$\frac{3}{36}$
11	(5,6), (6,5)	$\frac{2}{36}$
12	(6,6)	$\frac{1}{36}$

TABELLE 1. Der Zwei-Würfel Versuch

über die auftretenden Zufallsvariablen mindestens näherungsweise deren Wahrscheinlichkeitsverteilung bekannt sein.

Wir betrachten nun einige einfache Beispiele:

**Beispiel:** Wir würfeln mit zwei Würfeln. Die Zufallsvariable  $Z$  ist bestimmt durch die Summe der beiden Augenzahlen. Es sei  $\omega_i, i = 1, 2$  das Ergebnis des  $i$ -ten Wurfes, dann gilt  $Z = Z(\omega) = \omega_1 + \omega_2, \omega = (\omega_1, \omega_2)$ . Der Wertevorrat von  $Z$  ist gegeben durch

$$\{2, 3, 4, \dots, 12\}.$$

Die Wahrscheinlichkeitsverteilung wird nun dadurch bestimmt, wie wahrscheinlich es ist, dass die verschiedenen Werte von  $Z$  angenommen werden:

$$\mathbb{P}(Z = z) \quad \text{für } z = 2, \dots, 12.$$

Die Verteilung von  $Z$  wird durch die Tabelle 1 charakterisiert.

Eine Möglichkeit der graphischen Beschreibung ist durch das Diagramm in Abbildung 1 gegeben.

**Beispiel:** Einem Warenposten von 100 Teilen werden zur Qualitätskontrolle 10 Teile entnommen, wobei die entnommenen Teile nicht wieder zurückgelegt werden. Falls das  $i$ -te entnommene Teil qualitätsgerecht ist, notieren wir  $\omega_i = 0, i = 1, \dots, 10$ . Anderenfalls setzen wir  $\omega_i = 1$ . Die Zufallsvariable  $Z$  gibt die Anzahl der defekten Teile an:

$$Z = Z(\omega) = \sum_{i=1}^{10} \omega_i, \quad \omega = (\omega_1, \dots, \omega_{10}).$$

Die zugehörige Verteilung ist die hypergeometrische Verteilung, die im nächsten Abschnitt beschrieben wird.

**Beispiel:** Die Zufallsvariable  $T$  sei gegeben durch die zufällige Lebensdauer einer Glühbirne aus einer bestimmten Produktionsserie. Im Unterschied zu den obigen Beispielen liegen die Werte dieser Zufallsvariablen in einem Kontinuum und sind nicht durch eine endliche oder abzählbar unendliche Menge gegeben.

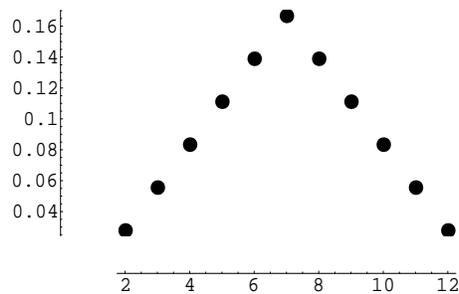


ABBILDUNG 1. Einzelwahrscheinlichkeiten der Summe zweier Würfel

## 2. Wahrscheinlichkeitsverteilungen

Wir wollen im Folgenden verschiedene Möglichkeiten kennenlernen, wie Wahrscheinlichkeitsverteilungen beschrieben werden können.

DEFINITION 4.1. Gegeben sei eine Zufallsvariable  $Z$ . Die durch

$$F(z) = \mathbb{P}(Z \leq z) = \mathbb{P}((-\infty, z]), \quad z \in \mathbb{R}$$

definierte Funktion  $F(z)$  heißt die Verteilungsfunktion der Zufallsvariablen  $Z$ .

Der Wert eines Argumentes  $z$  ist also die *Wahrscheinlichkeit*, dass die Zufallsvariable  $Z$  Werte kleiner oder gleich  $z$  annimmt. Der Definitionsbereich von  $F(z)$  sind die reellen Zahlen, auch wenn die Zufallsvariable nur Werte in einer kleineren Menge als  $\mathbb{R}$  annimmt. Da eine Wahrscheinlichkeit immer Werte im Intervall  $[0, 1]$  annimmt, überträgt sich das auch auf die Verteilungsfunktion:

$$\text{Db}(F) = \mathbb{R}, \quad \text{Wv}(F) \subset [0, 1].$$

**Beispiel:** Die Verteilungsfunktion für das Würfelexperiment: Die Werte, die die Zufallsvariable  $Z$ , die durch die gewürfelten Augen gegeben ist, annimmt sind

$$\{1, 2, 3, 4, 5, 6\}$$

und  $\mathbb{P}(Z = i) = \frac{1}{6}$  für  $i = 1, \dots, 6$ . Wir erhalten die Abbildung 2

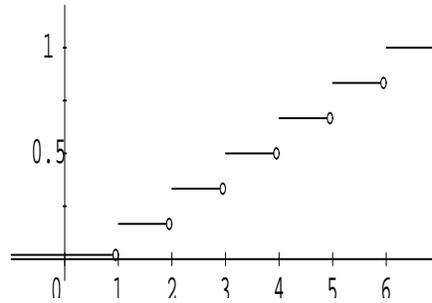


ABBILDUNG 2. Verteilungsfunktion des Würfelversuches

für die Verteilungsfunktion  $F(z)$ .

Aus der Definition ergibt sich für die Wahrscheinlichkeit, dass eine Zufallsvariable  $Z$  in einem Intervall  $(a, b]$  liegt

$$(12) \quad \mathbb{P}(Z \in (a, b]) = F(b) - F(a).$$

Jede Verteilungsfunktion besitzt die folgenden Eigenschaften:

SATZ 4.2. Es sei  $F(z)$  die Verteilungsfunktion einer Zufallsvariablen  $Z$ .

(i)  $F(z)$  ist monoton wachsend: aus  $z_1 < z_2$  folgt  $F(z_1) \leq F(z_2)$ .

(ii) Es gilt  $\lim_{z \rightarrow -\infty} F(z) = 0$  und  $\lim_{z \rightarrow \infty} F(z) = 1$ .

(iii)  $F(z)$  ist rechtsseitig stetig.  $\lim_{z \downarrow z_0} F(z) = F(z_0)$  für  $z_0 \in \mathbb{R}$ .

BEWEIS. (i) Aus  $z_1 < z_2$  folgt  $\mathbb{P}(Z \in (-\infty, z_1]) \leq \mathbb{P}(Z \in (-\infty, z_2])$ , was nach der Definition der Verteilungsfunktion äquivalent ist zu  $F(z_1) \leq F(z_2)$ .

(ii) Wir beschränken uns darauf zu zeigen, dass

$$\lim_{z_n \rightarrow -\infty} F(z_n) = 0.$$

Wir nehmen ohne Einschränkung der Allgemeinheit an, dass  $(z_n)$  eine monoton fallende Folge ist. Für  $A_n = (-\infty, z_n]$  gilt:

$$\bigcap_{n=1}^{\infty} A_n = \emptyset.$$

Speziell ist die Folge von Ereignissen  $A_n$  monoton fallend, siehe Definition 2.17. Nun wenden wir den Satz 2.18 an. Wir erhalten

$$0 = \mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \lim_{n \rightarrow \infty} F(z_n).$$

(iii) Rechtsseitig stetig heißt, dass eine Verteilungsfunktion Sprünge (hebbare Unstetigkeiten) haben kann, dass aber auf der rechten Seite der Unstetigkeitsstelle der Funktionswert definiert ist.

Für eine monoton fallende Folge  $z_n$  mit Grenzwert  $z_0$  gilt:

$$\bigcap_{n=1}^{\infty} (-\infty, z_n] = (-\infty, z_0].$$

Wendet man nochmals Satz 2.18 an, so ergibt sich

$$\begin{aligned} F(z_0) = \mathbb{P}(Z \in (-\infty, z_0]) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} (-\infty, z_n]\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}((-\infty, z_n]) = \lim_{n \rightarrow \infty} F(z_n). \end{aligned}$$

□

Im Folgenden werden wir zwei große Klassen von sich qualitativ unterscheidenden Zufallsvariablen betrachten. Dabei handelt es sich um die *diskreten* und die *stetigen* Zufallsvariablen.

**DEFINITION 4.3.** *Eine Zufallsvariable heißt diskret, wenn sie endlich viele oder unendlich abzählbar viele Werte annimmt. Eine Zufallsvariable heißt stetig, falls ihre Werte ein Intervall, bestehend aus mehr als einem Punkt, ausfüllen.*

So definiert der Zufallsversuch des Würfels, als auch der Qualitätskontrolle von 10 Teilen aus einen Warenposten von 100 Teilen diskrete Zufallsvariable. Im Unterschied dazu ist die Lebensdauer einer Glühbirne eine stetige Zufallsvariable. Denn als Ausfallszeitpunkt kommt jede positive reelle Zahl in Frage.

Wir wollen nun einige diskrete Zufallsvariable detailliert betrachten. Wir nehmen an, dass die Zufallsvariable  $Z$  die Werte

$$z_1, z_2, z_3, z_4 \dots$$

besitzt. Die Wahrscheinlichkeiten, dass diese Werte durch die Zufallsvariable angenommen werden, bezeichnen wir mit  $p_i$ :

$$\mathbb{P}(Z = z_i) = p_i.$$

Will man die Wahrscheinlichkeit berechnen, dass  $Z$  Werte in einer Menge  $A$  annimmt, so gilt:

$$\mathbb{P}(Z \in A) = \sum_{z_i \in A} p_i.$$

Wir betrachten nun einige spezielle Verteilungen, die in den verschiedensten Natur- und Ingenieurwissenschaften ihre Anwendung finden.

Wir beginnen mit der *Binomialverteilung*.

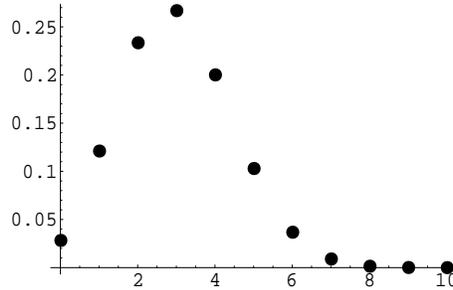


ABBILDUNG 3. Wahrscheinlichkeiten der (10, 0.3)-Binomialverteilung

Angenommen, ein zufälliger Versuch wird unter gleichbleibenden Bedingungen  $n$ -mal ausgeführt. Die Ausgänge der einzelnen Versuche beeinflussen sich nicht gegenseitig. Diese Bedingungen werden als *Bernoulli Versuch* bezeichnet. Bei jedem dieser  $n$  Versuche interessieren nur zwei Ereignisse. Entweder das eine Ereignis  $A$  tritt ein. Falls dieses Ereignis  $A$  nicht eintritt, so tritt  $\bar{A}$ , das komplementäre Ereignis, ein. Falls im  $i$ -ten Versuch  $A$  beziehungsweise  $\bar{A}$  eintritt, so schreiben wir dafür  $A_i$ ,  $\bar{A}_i$ ,  $i = 1, \dots, n$ . Man kann sich vorstellen, dass aus einem Warenposten  $n$  Teile nacheinander entnommen werden und je nach dem, ob das  $i$ -te Teil defekt ist, oder nicht,  $A_i$  oder  $\bar{A}_i$  eintritt. Nach der Entnahme und Test des  $i$ -ten Teiles wird dieses wieder in den Warenposten zurückgelegt. Damit wird sichergestellt, dass die einzelnen Versuche unabhängig voneinander stattfinden und dass die Versuchsbedingungen unverändert bleiben. Somit tritt  $A$  beziehungsweise  $\bar{A}$  in jedem dieser  $n$  Versuche mit gleicher Wahrscheinlichkeit ein. Wir nehmen also an, dass gilt:

$$\mathbb{P}(A_i) = p, \quad \mathbb{P}(\bar{A}_i) = 1 - p.$$

Die Binomialverteilung beschreibt die Wahrscheinlichkeiten, dass bei  $n$  Versuchen  $k$ -mal das Ereignis  $A$  und  $n - k$  mal das Ereignis  $\bar{A}$  auftritt. Diese Ereignisse werden mit  $B_k$  bezeichnet. Wir berechnen im Folgenden die Wahrscheinlichkeiten dieser Ereignisse.

Für  $B_0$  erhalten wir:

$$B_0 = \bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n.$$

Aufgrund der Unabhängigkeit dieser Ereignisse kann die Formel in Definition 3.7 für unabhängige Ereignisse angewandt werden. Somit erhalten wir die Wahrscheinlichkeit

$$\mathbb{P}(B_0) = \mathbb{P}(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) = \mathbb{P}(\bar{A}_1) \cdot \dots \cdot \mathbb{P}(\bar{A}_n) = (1 - p)^n.$$

Für das Ereignis  $B_1$  erhalten wir:

$$B_1 = (A_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) \cup (\bar{A}_1 \cap A_2 \cap \bar{A}_3 \cap \dots \cap \bar{A}_n) \cup \dots \cup (\bar{A}_1 \cap \dots \cap \bar{A}_{n-1} \cap A_n).$$

Aufgrund des Additionaxioms 2 (die eingeklammerten Ereignisse sind paarweise unvereinbar) und des Produktsatzes für unabhängige Ereignisse gilt

$$\mathbb{P}(B_1) = np(1 - p)^{n-1}.$$

Für  $B_2$  gilt:

$$B_2 = (A_1 \cap A_2 \cap \bar{A}_3 \cap \cdots \cap \bar{A}_n) \cup (A_1 \cap \bar{A}_2 \cap A_3 \cap \bar{A}_4 \cap \cdots \cap \bar{A}_n) \cup \cdots .$$

Die Anzahl der in Klammern zusammengefaßten Ereignisse, die vereinigt werden, kann mittels der Kombinatorik berechnet werden. Die Indizes der nicht überstrichenen Ereignisse in den einzelnen Klammern können zu Zweierklassen zusammengestellt werden. Dabei können in diesen Klassen  $n$  verschiedene Elemente stehen. Da Wiederholungen nicht möglich sind und jedes Element höchstens einmal in einer Zusammenstellung auftritt, wird die Gesamtheit dieser Klassen durch Kombination ohne Wiederholung gebildet. Es gibt also  $\binom{n}{2}$  solcher Zweierzusammenstellungen. Damit ist die Wahrscheinlichkeit für das Eintreten von  $B_2$  unter Berücksichtigung des Produktsatzes für unabhängige Ereignisse

$$\mathbb{P}(B_2) = \binom{n}{2} p^2 (1-p)^{n-2}.$$

Ähnlich können die Wahrscheinlichkeiten von  $\mathbb{P}(B_k)$ ,  $k = 3, \dots, n$  berechnet werden

DEFINITION 4.4. Eine Zufallsvariable  $Z$  mit Werten  $0, 1, \dots, n$  heißt  $(n, p)$ -binomialverteilt, falls

$$\mathbb{P}(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

für  $k = 0, 1, \dots, n$  gilt.

**Beispiel:** Es soll berechnet werden, wie groß die Wahrscheinlichkeit ist, beim fünfmaligen Werfen einer Münze dreimal *Wappen* zu erhalten. Es sei  $Z$  die Zufallsvariable, die zählt, wie häufig bei fünf Versuchen *Wappen* auftritt. Wir können beim Münzwurf davon ausgehen, dass die entsprechenden Versuche unabhängig unter gleichen Bedingungen durchgeführt werden, so dass die Voraussetzungen für die Binomialverteilung der Zufallsvariablen  $Z$  erfüllt ist. Die Wahrscheinlichkeiten, dass *Zahl* beziehungsweise *Wappen* auftritt, sind jeweils 0.5. Es ergibt sich:

$$\mathbb{P}(Z = 3) = \binom{5}{3} 0.5^3 0.5^2 = 0.3125.$$

Um die *geometrische Verteilung* zu beschreiben, nehmen wir an, dass ein zufälliger Versuch unter gleichbleibenden Bedingungen beliebig oft wiederholt werden kann. Dabei können nur die Ereignisse  $A, \bar{A}$  eintreten. Wir können uns also einen Bernoulli-Versuch vorstellen mit abzählbar unendlich vielen Teilversuchen. Die Wahrscheinlichkeit für das Eintreten von  $A, \bar{A}$  seien  $p, 1-p$ . Die Zufallsvariable  $Z$  gibt an, bei welchem Versuch das erste Mal  $\bar{A}$  eintritt. Es bezeichne  $B_i$  das Ereignis, dass  $\bar{A}$  das erste mal beim  $i$ -ten Versuch eintritt und  $A_i$  das Ereignis, dass beim  $i$ -ten Versuch  $A$  eintritt. Es gilt:

$$B_i = A_1 \cap \cdots \cap A_{i-1} \cap \bar{A}_i.$$

Die Wahrscheinlichkeit für  $B_i$  ergibt sich nach dem Produktsatz für unabhängige Ereignisse:

$$\mathbb{P}(B_i) = \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_{i-1}) \mathbb{P}(\bar{A}_i) = p^{i-1} (1-p).$$

Speziell für  $i = 1$  erhalten wir die Wahrscheinlichkeit  $\mathbb{P}(B_1) = 1-p$ .

DEFINITION 4.5. Eine Zufallsvariable  $Z$  mit Wertevorrat  $\mathbb{N}$  heißt *geometrisch verteilt* falls gilt:

$$\mathbb{P}(Z = i) = p^{i-1}(1-p) \quad \text{für } i \in \mathbb{N}.$$

Im Unterschied zu der Binomialverteilung besitzt die geometrische Verteilung einen abzählbar unendlichen Wertevorrat.

Notwendig für die Korrektheit der obigen Definition ist, dass die Summe der Einzelwahrscheinlichkeiten Eins ergibt:

$$\sum_{i=1}^{\infty} \mathbb{P}(Z = i) = 1.$$

Die rechte Seite ist gleich

$$(1-p) \sum_{i=0}^{\infty} p^i = (1-p) \frac{1}{1-p} = 1$$

nach der Formel für die geometrische Reihe.

**Beispiel:** Die Wahrscheinlichkeit, dass nach dem vierten Münzwurf das erste mal *Wappen* auftritt ist

$$\begin{aligned} \mathbb{P}(B_5 \cup B_6 \cup \dots) &= \sum_{i=5}^{\infty} (1-p)p^{i-1} = (1-p)p^4 \sum_{i=0}^{\infty} p^i \\ &= (1-p)p^4 \frac{1}{1-p} = p^4 = 0.0625, \end{aligned}$$

falls wir unter  $A_i$  die Ereignisse verstehen, beim  $i$ -ten Wurf *Zahl* zu werfen mit der Wahrscheinlichkeit von 0.5.

Eine Grundverteilung der Qualitätskontrolle ist die *hypergeometrische Verteilung*. Während wir, um die Voraussetzungen des Bernoulli Versuches zu sichern, bei der Binomialverteilung entnommene Werkstücke wieder zurücklegen, werden nun die getesteten Werkstücke nicht wieder zurückgelegt.

Wir leiten die hypergeometrische Verteilung an einem Beispiel ab:

Gegeben sei ein Warenposten mit  $N$  Werkstücken. Davon seien  $M$  defekt. Es werden  $n$  Werkstücke zufällig entnommen. Wir wollen die Wahrscheinlichkeit berechnen, dass  $k$  der  $n$  entnommenen Werkstücke defekt sind. Die Zufallsvariable  $Z$  zählt die defekten Werkstücke unter den  $n$  entnommenen Werkstücken. Da die Entnahme von beliebigen Kombinationen von  $n$  Teilen gleichberechtigt ist, kann die Definition der *klassischen Wahrscheinlichkeit* zur Anwendung kommen. Damit gilt

$$\mathbb{P}(Z = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

für  $k = 0, \dots, n$ . Die Herleitung dieser Formel ergibt sich auf die gleiche Weise, wie die Wahrscheinlichkeit berechnet wird, bei 6 aus 49 zum Beispiel einen Vierer zu erhalten. Entsprechend versuche man das folgende Schema zu verstehen, falls

$*$   $\in \{M + 1, \dots, N\}$  Nummern von nichtdefekten Teilen sind.

$$\binom{M}{k} \left\{ \begin{array}{l} \underbrace{\{1, 2, \dots, k, *, *, \dots, *\}}_{k} \quad \underbrace{\{*, *, \dots, *\}}_{n-k} \quad \binom{N-M}{n-k} \\ \dots \\ \dots \\ \dots \\ \underbrace{\{M-k+1, \dots, M, *, *, \dots, *\}}_{k} \quad \underbrace{\{*, *, \dots, *\}}_{n-k} \quad \binom{N-M}{n-k} \end{array} \right.$$

Falls in der obigen Formel ein Binomialkoeffizient  $\binom{r}{s}$  mit  $r < s$  auftritt, so wird dieser Null gesetzt.

DEFINITION 4.6. Eine Zufallsvariable  $Z$  mit Wertevorrat  $\{0, 1, \dots, n\}$  heißt hypergeometrisch verteilt, falls für Zahlen  $M, N, n, k \in \mathbb{N}$  gilt

$$\mathbb{P}(Z = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

für  $k = 0, \dots, n$  und  $M \geq k, N - M \geq n - k$  und ansonsten Null.

**Beispiel:** In einem Warenposten von  $N = 100$  Ventilen befinden sich  $M = 5$  defekte Ventile. Es werden 5 Ventile entnommen. Wie groß ist die Wahrscheinlichkeit, dass man Null defekte Teile entnimmt.

Es sei  $Z$  die Zufallsvariable, die die defekten Teile zählt. Wir haben zu berechnen

$$\mathbb{P}(Z = 0) = \frac{\binom{5}{0} \binom{95}{5}}{\binom{100}{5}} = 0.77.$$

Falls  $\frac{n}{M}$  klein ist, so kann die hypergeometrische Verteilung durch die Binomialverteilung ersetzt werden, wobei für die Wahrscheinlichkeit  $p = \frac{M}{N}$  gesetzt wird. In der Tat, wird der Versuchsaufbau wenig geändert, wenn bei einer großen Anzahl von Teilen wenige entfernt werden. Die Abweichung zwischen Binomialverteilung und hypergeometrischer Verteilung ist in der Regel hinreichend klein, falls  $\frac{n}{M} \leq 0.05$ .

Die *Poisson-Verteilung* ist wieder eine Verteilung mit Wertevorrat  $\mathbb{Z}^+$ . Sie wird zum Beispiel dazu genutzt, um zufällige Anzahlen zu charakterisieren. So kann zum Beispiel die zufällige Länge von Warteschlangen von Fahrzeugen an Ampeln oder Jobs an Servern damit modelliert werden.

DEFINITION 4.7. Es sei  $\lambda$  eine positive Zahl. Eine Zufallsvariable  $Z$  mit Wertevorrat  $\mathbb{Z}^+$  heißt *Poisson-verteilt*, falls gilt

$$\mathbb{P}(Z = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{für } k = 0, 1, \dots.$$

Wir wollen noch die für eine Wahrscheinlichkeitsverteilung notwendige Bedingung

$$\sum_{i=0}^{\infty} \mathbb{P}(Z = i) = 1$$

überprüfen. Die Taylorentwicklung der Exponentialfunktion  $e^x$  mit der Entwicklungsstelle Null für  $x = \lambda$  lautet

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}.$$

Damit ist die obige Summe gleich Eins.

Die Binomialverteilung kann unter gewissen Bedingungen durch die Poisson-Verteilung ersetzt werden. Die Abweichungen zwischen diesen Verteilungen sind hinreichend klein, falls die Bedingung  $n \geq 50$ ,  $p \leq 0.1$  erfüllt ist. Als Parameter  $\lambda$  wählt man dann den Wert  $np$ .  $\lambda$  entspricht dem *durchschnittlichen* Wert der Zufallsvariablen  $Z$ .

Trägt man die Einzelwahrscheinlichkeiten  $\mathbb{P}(Z = i) = p(i)$  in einem Diagramm ab, so erhält man Abbildung 4.

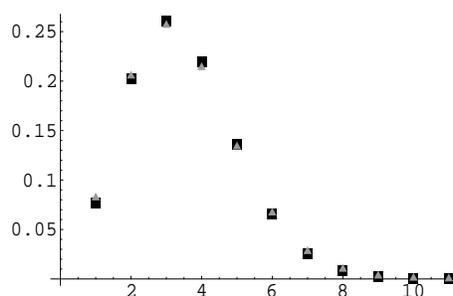


ABBILDUNG 4. Vergleich der Einzelwahrscheinlichkeiten der  $(50, 0.05)$ -Binomialverteilung (graues Dreieck) und der 2.5-Poisson-Verteilung (schwarzes Viereck). Die Dreiecke liegen häufig über den Vierecken.

Wir wenden uns nun den *stetigen* Zufallsvariablen zu. Zur Erinnerung, eine Zufallsvariable  $Z$  heißt stetig, wenn alle Werte in einem Kontinuum angenommen werden können. Die Beschreibung der Wahrscheinlichkeitsverteilung durch die Wahrscheinlichkeiten, dass spezielle Werte von der Zufallsvariablen angenommen werden, ist nicht sinnvoll. Es gilt nämlich

$$\mathbb{P}(Z = z) = 0$$

für jedes  $z$  aus dem Wertevorrat von  $Z$ .

Ein Werkzeug, um solche Wahrscheinlichkeitsverteilungen zu beschreiben, sind *Dichtefunktionen*. Um diese Funktionen einzuführen, unterteilen wir die Zahlengerade in durchschnittsfremde Intervalle  $\pi_n^i$ , mit Länge  $\frac{1}{n}$ . Eine solche Unterteilung bezeichnen wir mit  $\Pi_n$ . Desweiteren betrachten wir die Wahrscheinlichkeiten, dass  $Z$  Werte in den Intervallen  $\pi_n^i$  annimmt:

$$\mathbb{P}(Z \in \pi_n^i), \quad \pi_n^i \in \Pi_n.$$

Wir zeichnen diese Wahrscheinlichkeiten in ein Histogramm für fest gewähltes  $n$  ein.

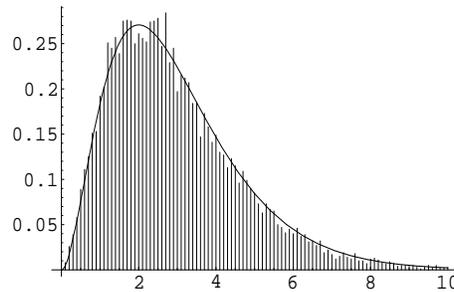


ABBILDUNG 5. Histogramm (für sehr kleine Basislänge) und Konturkurve

Dann verkleinern wir die Intervallängen der Intervalle  $\pi_n^i$ , in dem wir  $n$  gegen unendlich laufen lassen. Dabei bildet sich über den Rechteckflächen eine Konturkurve heraus, siehe Abbildung 5. Diese Konturkurve beschreibt die Dichtefunktion der Zufallsvariablen  $Z$ . Da die Flächeninhalte der Rechtecke, die als Grundseite die Zerlegungsintervalle besitzen, den Wahrscheinlichkeiten entsprechen, die Zufallsvariable  $Z$  Werte in so einem Intervall annimmt, ergibt sich die folgende Definition:

DEFINITION 4.8. Die Funktion  $f(z)$  heißt Dichtefunktion der Zufallsvariablen  $Z$ , falls für alle reellen Zahlen  $a < b$  gilt:

$$\mathbb{P}(Z \in (a, b]) = \int_a^b f(z) dz.$$

Es können also Wahrscheinlichkeiten ausgerechnet werden, indem Integrale bezüglich der Dichtefunktion bestimmt werden. Man erinnere sich daran, dass ein bestimmtes Integral als Grenzwert von Summen von Rechteckflächeninhalten gegeben ist, wobei die Basislänge der Rechtecke gegen Null geht.

Aus der Definition der Dichtefunktion ergeben sich die folgenden Eigenschaften:

SATZ 4.9. Es sei  $f(z)$  eine Dichtefunktion. Dann gilt:

- $f(z) \geq 0$  für  $z \in \mathbb{R}$ .
- $\int_{-\infty}^{\infty} f(z) dz = 1$ .
- Die Verteilungsfunktion kann bestimmt werden als:

$$F(z) = \mathbb{P}(Z \leq z) = \int_{-\infty}^z f(x) dx.$$

Falls die Funktion  $F(z)$  differenzierbar in  $z$  ist, dann gilt:

$$F'(z) = f(z).$$

Umgekehrt gilt, eine Funktion, die die ersten beiden Eigenschaften des obigen Satzes erfüllt, ist eine Dichtefunktion.

**Beispiel:** Gegeben sei eine Funktion

$$f(z) = \begin{cases} \frac{1}{2} + \frac{z}{4} & \text{für } -2 \leq z \leq 0 \\ \frac{1}{2} - \frac{z}{4} & \text{für } 0 < z \leq 2 \\ 0 & \text{sonst} \end{cases}$$

siehe Abbildung 6.

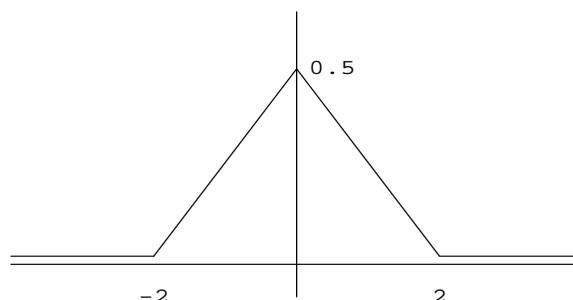


ABBILDUNG 6.

Es soll nachgewiesen werden, dass  $f(z)$  eine Dichtefunktion ist. Desweiteren bestimme man  $\mathbb{P}(Z \geq \frac{1}{2})$ .

Es gilt:

$$(13) \quad \int_{-\infty}^{\infty} f(z) dz = \int_{-\infty}^{-2} f(z) dz + \int_{-2}^0 f(z) dz + \int_0^2 f(z) dz + \int_2^{\infty} f(z) dz,$$

wobei wir für das zweite Integral auf der rechten Seite erhalten:

$$\int_{-2}^0 f(z) dz = \int_{-2}^0 \left( \frac{1}{2} + \frac{z}{4} \right) dz = \left( \frac{1}{2}z + \frac{z^2}{8} \right) \Big|_{-2}^0 = 1 - \frac{1}{2}.$$

und für das dritte Integral

$$\int_0^2 f(z) dz = \frac{1}{2}.$$

Die anderen beiden Integrale in Formel (13) sind Null, da die Funktion  $f(z)$  über dem entsprechenden Integrationsbereich  $(-\infty, -2)$  und  $(2, \infty)$  Null ist. Damit gilt

$$\int_{-\infty}^{\infty} f(z) dz = 1.$$

Desweiteren ist  $f(z) \geq 0$  für alle  $z \in \mathbb{R}$ . Somit ist nachgewiesen, dass  $f(z)$  eine Dichtefunktion ist. Für den zweiten Teil der Aufgabe berechnen wir die Wahrscheinlichkeit:

$$\mathbb{P}(Z \geq 0.5) = \int_{0.5}^{\infty} f(z) dz = \int_{0.5}^2 f(z) dz + \int_2^{\infty} f(z) dz,$$

wobei das zweite Integral auf der rechten Seite Null ist. Für das erste Integral gilt:

$$\int_{0.5}^2 f(z) dz = \int_{0.5}^2 \left( \frac{1}{2} - \frac{z}{4} \right) dz = \left( \frac{1}{2}z - \frac{z^2}{8} \right) \Big|_{0.5}^2 = 1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{32} = \frac{9}{32}.$$

Im Folgenden wollen wir uns mit speziellen stetigen Verteilungen beschäftigen, die zur Modellierung von vielen Prozessen in Natur-, Ingenieur- und Wirtschaftswissenschaften dienen.

Wir beginnen mit der *gleichmäßigen Verteilung*. Dabei wird davon ausgegangen, dass ein zufälliger Versuch Werte in einem endlichen Intervall  $[a, b]$  annimmt. Wir fordern weiterhin, dass die Versuchsausgänge in gleichlangen Teilintervallen von  $[a, b]$  mit gleicher Wahrscheinlichkeit liegen.

ABBILDUNG 7. Dichte der  $(0, 1)$  gleichmäßigen Verteilung

DEFINITION 4.10. Eine Zufallsvariable  $Z$  mit Werten in  $[a, b]$ ,  $a < b$  heißt *gleichmäßig verteilt*, falls sie die Dichte

$$(14) \quad f(z) = \begin{cases} \frac{1}{b-a} & \text{für } z \in [a, b] \\ 0 & \text{sonst.} \end{cases}$$

besitzt.

Es ist leicht zu prüfen, dass  $f(z)$  eine Dichtefunktion ist:

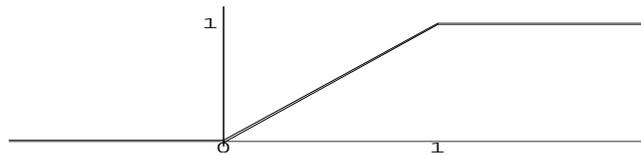
$$\int_{-\infty}^{\infty} f(z) dz = \int_a^b \frac{1}{b-a} dx = 1.$$

Für die Verteilungsfunktion ergibt sich:

$$F(z) = \int_{-\infty}^z f(\xi) d\xi = \begin{cases} 0 & \text{für } z < a \\ \frac{z-a}{b-a} & \text{für } z \in [a, b] \\ 1 & \text{für } z > b. \end{cases}$$

**Beispiel:** Angenommen die Randpunkte des Intervalls  $[a, b]$  seien gegeben durch  $a = 0$ ,  $b = 1$ . Die Wahrscheinlichkeit, dass  $Z$  Werte in  $(\frac{1}{3}, \frac{2}{3}]$  annimmt, ergibt sich durch

$$\mathbb{P}\left(Z \in \left(\frac{1}{3}, \frac{2}{3}\right]\right) = F\left(\frac{2}{3}\right) - F\left(\frac{1}{3}\right) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

ABBILDUNG 8. Verteilungsfunktion der  $(0, 1)$  gleichmäßigen Verteilung

Die gleichmäßige Verteilung beschreibt näherungsweise die von einem Computer bereitgestellten Zufallszahlen.

Die *Exponentialverteilung* ist die Grundverteilung der Zuverlässigkeitstheorie. Sie beschreibt die Lebensdauer von (einfachen) Bauteilen, bei denen die Alterung vernachlässigt werden kann. Desweiteren kann der radioaktive Zerfall damit modelliert werden. Weitere Anwendungen bestehen in der Zeitdauer von Instandsetzungsmaßnahmen in der Länge von Warteschlangen.

DEFINITION 4.11. Es sei  $\lambda > 0$ . Dann heißt eine Zufallsvariable  $Z$  exponentialverteilt, falls sie die Dichte

$$(15) \quad f(z) = \begin{cases} 0 & \text{für } z \leq 0 \\ \lambda e^{-\lambda z} & \text{für } z > 0 \end{cases}$$

besitzt.

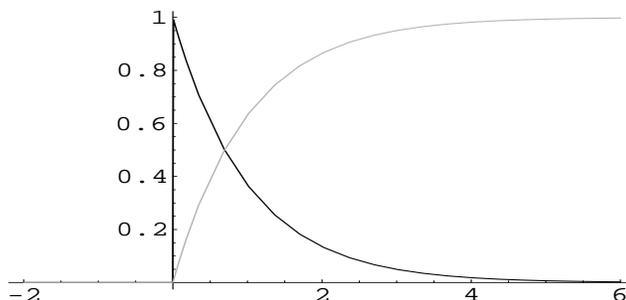


ABBILDUNG 9. Dichte- und Verteilungsfunktion der Exponentialverteilung für  $\lambda = 1$ .

Die Bedeutung des Parameters  $\lambda$  liegt darin, dass  $\lambda^{-1}$  ein mittlerer Wert der Zufallsvariablen  $Z$  ist. Für die Verteilungsfunktion erhalten wir:

$$F(z) = \begin{cases} 0 & \text{für } z < 0 \\ 1 - e^{-\lambda z} & \text{für } z \geq 0. \end{cases}$$

**Beispiel:** Die mittlere Reparaturzeit eines Gerätes beträgt zwei Stunden. Wie groß ist die Wahrscheinlichkeit, dass die Reparaturzeit mindestens drei Stunden beträgt. Für den Parameter gilt:  $\lambda = \frac{1}{2(h)}$ . Mittels der Verteilungsfunktion erhält man:

$$\mathbb{P}(Z > 3) = 1 - \mathbb{P}(Z \leq 3) = 1 - F(3) = 1 - (1 - e^{-0.5 \cdot 3}) = 0.2231.$$

Eine Eigenschaft, die die Exponentialverteilung wichtig für Anwendungen macht, wird als *Gedächtnislosigkeit* dieser Verteilung bezeichnet.

SATZ 4.12. Es gilt:  $Z$  ist exponentialverteilt genau dann, wenn

$$(16) \quad \mathbb{P}(Z > t_1 + t_2 | Z > t_2) = \mathbb{P}(Z > t_1).$$

Zur Interpretation dieser Gleichung nehmen wir an, dass  $Z$  eine zufällige Lebensdauer beschreibt. Die letzte Gleichung besagt, dass die Wahrscheinlichkeit, dass ein Bauteil älter als  $t_1 + t_2$  wird, falls es schon älter als  $t_2$  geworden ist, gleich der Wahrscheinlichkeit ist, älter als  $t_1$  zu werden. Das heißt, die Wahrscheinlichkeit, ein gewisses Alter zu überschreiten, hängt nur von der betrachteten Zeitspanne ab, nicht aber vom Alter des Bauteiles zu Beginn der Beobachtung. Der Anteil der Elemente, die länger als eine Zeitspanne  $t_1$  leben, ist unabhängig von  $t_2$ . Speziell wird die Überlebenswahrscheinlichkeit nicht beeinflusst durch Ereignisse vor dem Zeitpunkt  $t_2$ .

BEWEIS. 1) Nach der Definition der bedingten Wahrscheinlichkeit gilt für eine Exponentialverteilung:

$$\begin{aligned}\mathbb{P}(Z > t_1 + t_2 | Z > t_2) &= \frac{\mathbb{P}((Z > t_1 + t_2) \cap (Z > t_2))}{\mathbb{P}(Z > t_2)} \\ &= \frac{\mathbb{P}(Z > t_1 + t_2)}{\mathbb{P}(Z > t_2)} = \frac{1 - F(t_1 + t_2)}{1 - F(t_2)} = \frac{e^{-\lambda(t_1 + t_2)}}{e^{-\lambda t_2}} \\ &= e^{-\lambda t_1} = 1 - F(t_1) = \mathbb{P}(Z > t_1).\end{aligned}$$

woraus der erste Teil des Beweises folgt.

2) Wir nehmen an, dass (16) erfüllt ist. Dann ergibt sich aus der Definition der Verteilungsfunktion und der bedingten Wahrscheinlichkeit

$$1 - F(t_1 + t_2) = (1 - F(t_1))(1 - F(t_2)) \quad \text{für } t_1, t_2 \geq 0,$$

also

$$h(t_1 + t_2) = h(t_1)h(t_2) \quad \text{für } t_1, t_2 \geq 0.$$

Die einzigen Lösungen sind Funktionen  $1 - F(t) = h(t) = e^{-\lambda t}$ .  $\square$

Eine der wichtigsten Verteilungen ist die *Normalverteilung*, die auch als *Gaußverteilung* bezeichnet wird. Sie findet ihre Anwendung in der Statistik, der Zuverlässigkeitstheorie und in der Fehlerrechnung. Desweiteren bestehen Beziehungen zur Brownschen Molekularbewegung und zu dem Verhalten von Aktienmärkten. Mit Hilfe der Normalverteilung können zufällige Abweichungen vom Nennmaß beschrieben werden. Der Grund dafür ist, dass sich solche Abweichungen aus der Überlagerung von vielen unabhängigen Einflussgrößen ergeben. Diese Eigenschaft steht im Zusammenhang mit der Theorie der Grenzwertsätze, wo die Superposition von unabhängigen Zufallsvariablen untersucht wird.

DEFINITION 4.13. *Es seien  $\mu \in \mathbb{R}$  und  $\sigma > 0$  zwei Parameter. Eine Zufallsvariable  $Z$  heißt normalverteilt, falls sie die Dichte*

$$(17) \quad f(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z - \mu)^2}{2\sigma^2}\right)$$

besitzt.

Diese Funktion ist symmetrisch, wobei die Symmetrieachse senkrecht den Wert  $z = \mu$  auf der reellen Achse schneidet.  $\mu$  gibt auch den mittleren Wert der Zufallsvariablen  $Z$  an, siehe Abbildung 10.

Der Parameter  $\sigma$  bestimmt die *Breite* der Dichtefunktion. Ist  $\sigma$  groß, so ist der Graph der Funktion auseinandergezogen, ist  $\sigma$  klein, so ist der Graph der Funktion sehr spitz.  $\sigma$  beschreibt, wie die Werte der Zufallsvariablen  $Z$  streuen, siehe Abbildung 11.

Ist  $Z$  normalverteilt mit den Parametern  $\mu$  und  $\sigma$ , so schreibt man abkürzend dafür  $N(\mu, \sigma)$ . Ist  $\mu = 0$  und  $\sigma = 1$ , so bezeichnet man die entsprechende Verteilung als *Standardnormalverteilung*. Die zugehörige Verteilungsfunktion der Standardnormalverteilung werde mit  $\Phi(z)$  bezeichnet. Die Verteilungsfunktion einer Normalverteilung ist im Allgemeinen nicht explizit aufschreibbar. Die Werte liegen nur tabelliert für die Standardnormalverteilung vor.

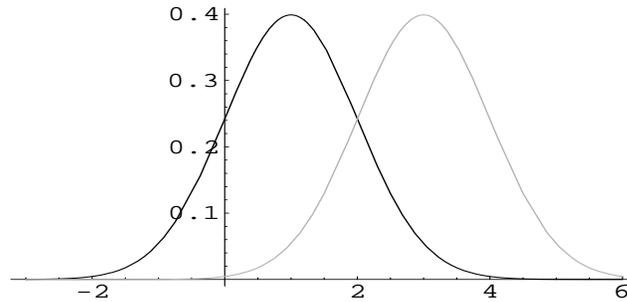


ABBILDUNG 10. Dichte der Normalverteilung für  $\mu = 1$  und  $\mu = 3$ .

Werden die Werte einer normalverteilten Zufallsvariablen linear transformiert, so ist die transformierte Zufallsvariable ebenfalls normalverteilt. Es gilt:

LEMMA 4.14. *Es seien  $a, b$  reelle Zahlen und  $Z$  eine  $N(\mu, \sigma)$  normalverteilte Zufallsvariable. Die Zufallsvariable*

$$\tilde{Z} = aZ + b$$

ist  $N(\tilde{\mu}, \tilde{\sigma})$ -verteilt mit

$$\tilde{\mu} = a\mu + b, \quad \tilde{\sigma} = |a|\sigma.$$

Setzt man speziell  $a = \sigma^{-1}$  und  $b = -\mu\sigma^{-1}$ , dann ist  $\tilde{\mu} = 0$ ,  $\tilde{\sigma} = 1$ .  $\tilde{Z}$  ist somit standardnormalverteilt.  $\tilde{Z}$  ergibt sich in diesem Fall aus  $Z$  durch die Transformation:

$$(18) \quad \tilde{Z} = \frac{Z - \mu}{\sigma}.$$

Um Wahrscheinlichkeiten zu berechnen, dass eine  $N(\mu, \sigma)$ -verteilte Zufallsvariable Werte in einem gewissen Intervall annimmt, kann theoretisch nach (12) berechnet werden. Die Verteilungsfunktion einer normalverteilten Zufallsvariablen ist nicht *explizit* als Formel aufschreibbar, obwohl sie existiert. Mit anderen Worten bedeutet das, dass die Stammfunktion zu (17) nicht angegeben werden kann. Es gibt aber die Möglichkeit, die Integrale über diese Dichtefunktion näherungsweise zu berechnen und die berechneten Werte zu tabellieren. Allerdings scheint der dazu notwendige Aufwand sehr groß zu sein, denn es müsste ja für jedes  $\mu$  und  $\sigma > 0$  eine Tabelle bereitgestellt werden. Durch eine Transformation wie (18) kann jede  $N(\mu, \sigma)$ -verteilte Zufallsvariable  $Z$  in eine  $N(0, 1)$  verteilte Zufallsvariable  $\tilde{Z}$  mit Verteilungsfunktion  $\Phi(z)$  transformiert werden. Deswegen enthalten Formelsammlungen nur die Werte der Verteilungsfunktion der Standardnormalverteilung.

Wir berechnen die Wahrscheinlichkeit, dass eine  $N(\mu, \sigma)$ -verteilte Zufallsvariable  $Z$  Werte in einem Intervall  $(a, b]$  annimmt:

$$\begin{aligned} \mathbb{P}(Z \in (a, b]) &= \mathbb{P}(a < Z \leq b) = \mathbb{P}\left(\frac{a - \mu}{\sigma} < \frac{Z - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\ &= \mathbb{P}\left(\frac{a - \mu}{\sigma} < \tilde{Z} \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \end{aligned}$$

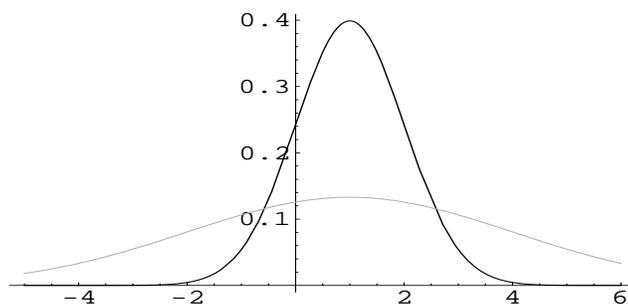


ABBILDUNG 11. Dichte der Normalverteilung für  $\sigma = 1$  und  $\sigma = 3$ , wobei  $\mu = 1$  gewählt ist.

Die Werte der Verteilungsfunktion  $\Phi$  können aus einer Tabelle abgelesen werden, sofern die Argumente von  $\Phi$  nicht negativ sind. Die Funktionswerte für negative Argumente können durch die einfache Umrechnungsformel

$$\Phi(z) = 1 - \Phi(-z), \quad xz < 0$$

berechnet werden. Diese Formel basiert auf der Symmetrie der Dichtefunktion der Standardnormalverteilung. Speziell gilt nach [Abbildung 12](#)

$$\Phi(z) = \mathbb{P}(Z \leq z) = \Phi(Z \geq -z) = 1 - \Phi(-z).$$

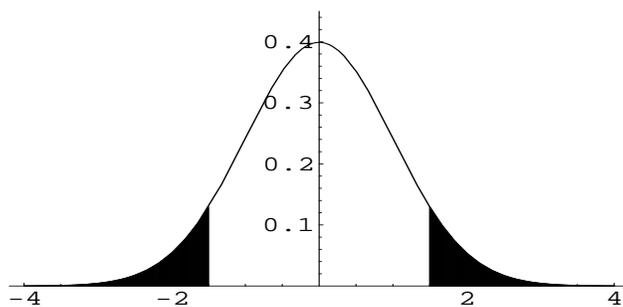


ABBILDUNG 12. Die linke schwarze Fläche ist  $\Phi(z)$ ,  $z < 0$  und die rechte  $1 - \Phi(-z)$ . Beide Flächen sind gleichgroß.

**Beispiel:** Es sei  $Z$  eine  $N(6, 2)$  normalverteilte Zufallsvariable. Man berechne:

$$\begin{aligned}\mathbb{P}(Z \leq 7) &= \mathbb{P}\left(\tilde{Z} \leq \frac{7-6}{2}\right) = \Phi(0.5) = 0.6914 \\ \mathbb{P}(Z \leq 3) &= \mathbb{P}\left(\tilde{Z} \leq \frac{3-6}{2}\right) = \Phi(-1.5) = 1 - 0.93319 = 0.06681 \\ \mathbb{P}(6.2 < Z \leq 8) &= \mathbb{P}\left(\frac{6.2-6}{2} < \tilde{Z} \leq \frac{8-6}{2}\right) \\ &= \Phi(1) - \Phi(0.1) = 0.8413 - 0.5398 = 0.3015 \\ \mathbb{P}(4.6 \leq Z \leq 7) &= \mathbb{P}\left(\frac{4.6-6}{2} \leq \tilde{Z} \leq \frac{7-6}{2}\right) \\ &= \mathbb{P}(-0.7 \leq \tilde{Z} \leq 0.5) \\ &= \Phi(0.5) - (1 - \Phi(0.7)) = 0.6914 + 0.7580 - 1 = 0.4494.\end{aligned}$$

Der folgende Satz beinhaltet eine weitere Eigenschaft der Normalverteilung, die im Folgenden näher erläutert wird. Dabei geht es um die Superposition von normalverteilten Zufallsvariablen

**SATZ 4.15.** *Gegeben seien  $n$  unabhängige normalverteilte Zufallsvariablen  $Z_i$  mit Verteilung  $N(\mu_i, \sigma_i)$ . Dann ist die Überlagerung dieser Zufallsvariablen*

$$Z := Z_1 + Z_2 + \cdots + Z_n$$

wieder normalverteilt mit dem Parameter

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$

und

$$\sigma := \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}.$$

Bei der Analyse von Messfehlern findet häufig das Fehlerfortpflanzungsgesetz seine Anwendung.

**SATZ 4.16.** *Es sei  $g(z)$  eine stetig differenzierbare Funktion und  $Z$  eine  $N(\mu, \sigma)$ -verteilte Zufallsvariable. Unter der Annahme, dass der Parameter  $\sigma$  nicht zu groß ist, ist die Verteilung der Zufallsvariablen  $g(Z)$  näherungsweise normalverteilt mit Parametern  $\tilde{\mu} = g(\mu)$  und  $\tilde{\sigma} = |g'(\mu)|\sigma$ .*

**BEWEIS.** In einer Umgebung um  $\mu$  wird  $g(z)$  näherungsweise durch die Linearisierung  $l(z)$  beschrieben, wobei diese lineare Funktion gegeben ist durch

$$g(z) \approx l(z) = g'(\mu)(z - \mu) + g(\mu).$$

Ist  $\sigma$  nicht zu groß, so liegen die *meisten* zufälligen Werte der Zufallsvariablen  $Z$  mit großer Wahrscheinlichkeit in der Nähe von  $\mu$ . Damit ist  $g(Z)$  näherungsweise gegeben durch

$$l(Z) = g'(\mu)(Z - \mu) + g(\mu).$$

$l(Z)$  ergibt sich aus einer linearen Transformation von  $Z$ , ist also normalverteilt. Die Parameter  $\tilde{\mu}$ ,  $\tilde{\sigma}$  können nach Lemma 4.14 berechnet werden.  $\square$

Als Verallgemeinerung des letzten Satzes erhalten wir

SATZ 4.17. Es sei  $g(z_1, \dots, z_n)$  eine stetig differenzierbare Funktion mit  $n$  Veränderlichen. Desweiteren seien  $Z_1, \dots, Z_n$  unabhängige normalverteilte Zufallsvariablen mit Verteilung  $N(\mu_i, \sigma_i)$ ,  $i = 1, \dots, n$ , wobei die  $\sigma_i$  als nicht zu groß angenommen werden. Dann ist

$$g(Z_1, \dots, Z_n)$$

näherungsweise normalverteilt mit den Parametern  $\tilde{\mu} = g(\mu_1, \dots, \mu_n)$  und

$$\tilde{\sigma} = \sqrt{\left| \frac{\partial g(\mu_1, \dots, \mu_n)}{\partial z_1} \right|^2 \sigma_1^2 + \dots + \left| \frac{\partial g(\mu_1, \dots, \mu_n)}{\partial z_n} \right|^2 \sigma_n^2}.$$

Zum Abschluss sollen noch weitere wichtige Verteilungsfunktionen genannt werden und ein paar Anwendungen aufgezeigt werden:

- Weibull-Verteilung-Korngrößen von Siebgut
- Erlang-Verteilung-Zuverlässigkeitstheorie
- logarithmische Normalverteilung-Zuverlässigkeitstheorie
- Student-Verteilung-Statistik
- Fischer Verteilung-Statistik
- $\chi^2$  Verteilung-Statistik

### 3. Parameter von Zufallsvariablen

Die Kenntnis der Verteilung einer Zufallsvariablen bedeutet, dass die Gesamtheit der stochastischen Information bezüglich dieser Zufallsvariablen bekannt ist. Außerdem ist man nun in der Lage, eine Zufallsvariable durch spezielle Parameter teilweise zu beschreiben. Manchmal reicht auch schon die Kenntnis von wenigen speziellen Parametern aus, die Verteilung der Zufallsvariablen vollständig zu charakterisieren. Das ist zum Beispiel bei der Normalverteilung der Fall, wobei die Kenntnis von  $\mu$  und  $\sigma$  hinreichend für die Kenntnis der gesamten Verteilung ist.

Wir beginnen mit der Beschreibung des Parameters *Erwartungswert*, der auch manchmal als *Mittelwert* bezeichnet wird. Ihm entspricht der durchschnittliche Wert aller Werte, die durch den zufälligen Versuch, den die Zufallsvariable  $Z$  beschreibt, angenommen werden kann. Eine andere Interpretation erlaubt diesen Wert als Schwerpunkt der Wahrscheinlichkeitsverteilung zu charakterisieren, siehe Abbildung 13.

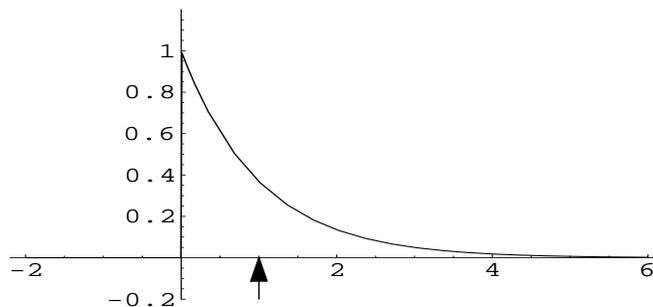


ABBILDUNG 13. Schwerpunkt der Exponentialverteilung mit Parameter  $\lambda = 1$

DEFINITION 4.18. Der Erwartungswert oder Mittelwert einer Zufallsvariablen  $Z$  ist definiert:

- $Z$  sei eine diskrete Zufallsvariable:

$$\mathbb{E} Z = \sum_i z_i p_i.$$

Dabei sind  $z_i$  die Werte einer Zufallsvariablen  $Z$  und  $p_i = \mathbb{P}(Z = z_i)$ .

- $Z$  sei eine stetige Zufallsvariable

$$\mathbb{E} Z = \int_{-\infty}^{\infty} z f(z) dx,$$

wobei  $f(z)$  die Dichtefunktion von  $Z$  ist.

Es sei erwähnt, dass es Verteilungen gibt, für die der Erwartungswert nicht definiert ist.

Der Erwartungswert besitzt die folgenden Eigenschaften:

SATZ 4.19. Es sei  $Z_i, i = 1, \dots, n$  eine Menge von Zufallsvariablen. Dann gilt:

$$\mathbb{E}(Z_1 + Z_2 + \dots + Z_n) = \mathbb{E}Z_1 + \mathbb{E}Z_2 + \dots + \mathbb{E}Z_n.$$

Weiterhin gilt für eine reelle Zahl  $\lambda$ :

$$\mathbb{E}(\lambda Z) = \lambda \mathbb{E} Z.$$

**Beispiel:** Es sei  $Z$  eine  $(n, p)$ -binomialverteilte Zufallsvariable. Dann gilt nach Definition 4.4

$$\mathbb{E} Z = \sum_{i=0}^n i \binom{n}{p} p^i (1-p)^{n-i}.$$

Eine Berechnung dieses Ausdruckes ist möglich, mittels der Eigenschaften des Binomialkoeffizientens. Wir wollen aber einen anderen Weg wählen, der auf Satz 4.19 basiert. Die Zufallsvariable  $Z$  zählt, wie häufig ein Ereignis  $A$  bei der Durchführung von  $n$  unabhängigen Versuchen eintritt. Das Ereignis  $A$  tritt bei jedem der Versuche mit Wahrscheinlichkeit  $p$  ein. Wir können  $Z$  wie folgt definieren:

$$Z = Z_1 + Z_2 + \dots + Z_n,$$

wobei die Zufallsvariablen  $Z_i$  nur die Werte 0, 1 annehmen können.  $Z_i$  besitzt den Wert 1 genau dann, wenn beim  $i$ -ten Versuch  $A$  eintritt. Das geschieht mit einer Wahrscheinlichkeit von  $p$ .  $Z_i$  besitzt den Wert 0 genau dann, wenn beim  $i$ -ten Versuch das Ereignis  $\bar{A}$  eintritt. Das geschieht mit Wahrscheinlichkeit  $1-p$ . Der Erwartungswert der Zufallsvariablen  $Z_i$  kann nach Definition 4.18 wie folgt berechnet werden:

$$\mathbb{E} Z_i = 1p + 0(1-p) = p.$$

Nach Satz 4.19 ist dann der Erwartungswert von  $Z$ :

$$\mathbb{E} Z = \mathbb{E} Z_1 + \mathbb{E} Z_2 + \dots + \mathbb{E} Z_n = np.$$

**Beispiel:** Es sei  $Z$  eine Poisson-verteilte Zufallsvariable mit einem Parameter  $\lambda > 0$ . Die Einzelwahrscheinlichkeiten sind definiert durch:

$$\mathbb{P}(Z = i) = \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{für } i = 0, 1, \dots,$$

siehe Definition 4.7. Die folgende Berechnung beruht wieder auf der Taylor-Entwicklung der  $e$ -Funktion:

$$\begin{aligned}\mathbb{E}Z &= \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} + 0 \\ &= \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda.\end{aligned}$$

**Beispiel:** Wir berechnen den Erwartungswert der gleichmäßigen Verteilung, welche die Dichtefunktion (14) besitzt. Die Integration nach Definition 4.18 ergibt:

$$\begin{aligned}(19) \quad \int_{-\infty}^{\infty} z f(z) dz &= \int_a^b \frac{z}{b-a} dz = \frac{1}{b-a} \frac{z^2}{2} \Big|_a^b \\ &= \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{1}{2}(b+a).\end{aligned}$$

Der Erwartungswert ist also gerade der Mittelpunkt des Intervalls, auf dem die gleichverteilte Zufallsvariable  $Z$  konzentriert ist.

**Beispiel:** Der Erwartungswert einer exponentiell verteilten Zufallsvariablen  $Z$  kann durch partielle Integration berechnet werden.

$$\begin{aligned}\mathbb{E}Z &= \int_0^{\infty} z \lambda e^{-\lambda z} dz = z \frac{-1}{\lambda} \lambda e^{-\lambda z} \Big|_0^{\infty} \\ &\quad - \int_0^{\infty} \lambda \frac{-1}{\lambda} e^{-\lambda z} dz = 0 + \int_0^{\infty} e^{-\lambda z} dz = \frac{1}{\lambda}.\end{aligned}$$

SATZ 4.20. *Es sei  $Z$  eine  $N(\mu, \sigma)$ -verteilte Zufallsvariable, dann gilt  $\mathbb{E}Z = \mu$ .*

Der Beweis wird später mittels der momenterzeugenden Funktionen erbracht. Abschließend soll noch der Erwartungswert einer durch eine Funktion  $g(x)$  transformierten Zufallsvariablen beschrieben werden. Es sei  $Z$  eine Zufallsvariable. Dann ist auch  $g(Z)$  eine Zufallsvariable.

SATZ 4.21. *Der Erwartungswert von  $g(Z)$  ist gegeben durch*

$$(20) \quad \mathbb{E}g(Z) = \begin{cases} \sum_i g(z_i) p_i & Z \text{ diskret} \\ \int_{-\infty}^{\infty} g(z) f(z) dz & Z \text{ stetig mit Dichte } f(z). \end{cases}$$

Die *Varianz* ist ein Parameter, der beschreibt, ob die Ausgänge des zufälligen Versuches im Mittel einen großen oder kleinen Abstand vom Erwartungswert besitzen. Speziell betrachtet man den *quadratischen* Abstand vom Erwartungswert  $\mathbb{E}Z$ . Der quadratische Abstand eines Punktes  $x$  vom Erwartungswert ist gegeben durch

$$g(x) = (x - \mathbb{E}Z)^2.$$

Berechnet man den mittleren quadratischen Abstand der Zufallsvariablen  $Z$  von  $\mathbb{E}Z$ , so ergibt sich aus Satz 4.21

$$\mathbb{E}(Z - \mathbb{E}Z)^2.$$

DEFINITION 4.22. Die Varianz, manchmal auch als Streuung bezeichnet, einer Zufallsvariablen ist gegeben durch

$$D^2 Z = \mathbb{E}(Z - \mathbb{E}Z)^2.$$

Die Quadratwurzel  $\sqrt{D^2 Z}$  der Varianz bezeichnet man als Standardabweichung.

Manchmal kann man den Rechenaufwand vermindern, wenn man die Varianz nach der folgenden Formel berechnet:

SATZ 4.23. Es gilt:

$$D^2 Z = \mathbb{E}(Z^2) - (\mathbb{E}Z)^2.$$

BEWEIS. Nach der binomischen Formel erhalten wir

$$\begin{aligned} D^2 Z &= \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathbb{E}(Z^2 - 2Z\mathbb{E}Z + (\mathbb{E}Z)^2) \\ &= \mathbb{E}Z^2 - 2\mathbb{E}Z\mathbb{E}Z + \mathbb{E}(\mathbb{E}Z)^2 \\ &= \mathbb{E}Z^2 - 2(\mathbb{E}Z)^2 + (\mathbb{E}Z)^2 = \mathbb{E}Z^2 - (\mathbb{E}Z)^2. \end{aligned}$$

Bei der Berechnung haben wir genutzt, dass der Erwartungswert einer Konstanten die Konstante selbst ist.  $\square$

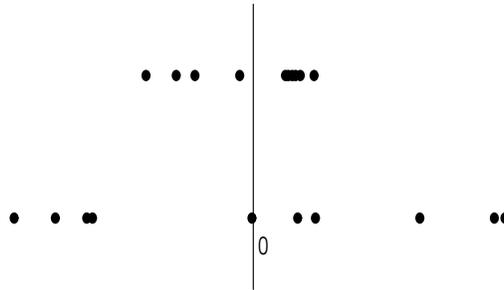


ABBILDUNG 14. Oben: Punkte mit kleiner Streuung um den Mittelwert 0 herum, unten große Streuung

Aufgrund von Satz 4.21 ist es nun möglich, Formeln für die Berechnung der Varianz anzugeben:

SATZ 4.24. Für eine diskrete Zufallsvariable  $Z$  ergibt sich die Varianz:

$$D^2 Z = \sum_i (z_i - \mathbb{E}Z)^2 p_i = \sum_i z_i^2 p_i - (\mathbb{E}Z)^2.$$

Für eine stetige Zufallsvariable  $Z$  mit Dichte  $f(z)$  erhalten wir für die Varianz

$$D^2 Z = \int_{-\infty}^{\infty} (z - \mathbb{E}Z)^2 f(z) dz = \int_{-\infty}^{\infty} z^2 f(z) dz - (\mathbb{E}Z)^2.$$

Es sei  $Z$  eine Zufallsvariable, die durch Superposition von Zufallsvariablen  $Z_i$  entsteht:

$$Z = Z_1 + \dots + Z_n.$$

Leider ergibt sich im Allgemeinen die Varianz von  $Z$  nicht als Summe der Varianzen von  $Z_i$ . Diese Eigenschaft ist nur unter speziellen Voraussetzungen richtig:

SATZ 4.25. Die Zufallsvariablen  $Z_i$ ,  $i = 1, \dots, n$  seien untereinander vollständig unabhängig<sup>1</sup>. Dann gilt:

$$D^2 Z = D^2 Z_1 + \dots + D^2 Z_n.$$

Desweiteren gilt:

$$D^2(aZ + b) = |a|^2 D^2 Z.$$

Die letzte Eigenschaft bedeutet, dass sich die Varianz nicht bei der *Verschiebung* von Zufallsvariablen ändert.

**Beispiel:** Wir berechnen die Varianz für die Gleichverteilung, wobei die Dichte in (14) definiert ist. Der Erwartungswert ist nach (19) gegeben durch

$$\mathbb{E}Z = \frac{a+b}{2}.$$

Somit ist

$$\begin{aligned} D^2 Z &= \int_a^b \frac{z^2}{b-a} dx - (\mathbb{E}Z)^2 = \frac{z^3}{3(b-a)} \Big|_a^b - \frac{(b+a)^2}{4} \\ &= \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4} = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} \\ &= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}, \end{aligned}$$

wobei wir benutzt haben, dass

$$\frac{b^3 - a^3}{b-a} = b^2 + ab + a^2$$

ist.

Aus dieser Formel geht hervor, dass wenn die Randpunkte  $a, b$  des Intervalls, auf dem sich die Werte der gleichmäßigen Verteilung konzentrieren, nahe beieinander liegen, dass dann auch die Varianz klein ist und umgekehrt.

Wir wollen nun die Methode der *momentenerzeugenden Funktion* anwenden, um gemeinsam den Erwartungswert und die Varianz einer normalverteilten Zufallsvariablen  $Z \sim N(\mu, \sigma)$  zu berechnen. Wir definieren eine Funktion

$$m(t) = \mathbb{E}e^{tZ} = e^{\mu t} \mathbb{E}e^{t(Z-\mu)}.$$

Aufgrund von Satz 4.21 mit  $g(x) = e^{t(x-\mu)}$  kann dafür geschrieben werden

$$e^{\mu t} \mathbb{E}e^{t(Z-\mu)} = \frac{e^{t\mu}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}((z-\mu)^2 - 2\sigma^2 t(z-\mu))\right) dz.$$

Der Ausdruck im Exponenten innerhalb des Integrals kann wie folgt durch eine quadratische Ergänzung umgeformt werden:

$$(z-\mu)^2 - 2\sigma^2 t(z-\mu) = ((z-\mu) - \sigma^2 t)^2 - \sigma^4 t^2.$$

Somit gilt:

$$m(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{(z - (\mu + \sigma^2 t))^2}{2\sigma^2}\right) dz.$$

<sup>1</sup>Die *Unabhängigkeit von Zufallsvariablen* ist hier noch nicht in aller Strenge definiert. Das geschieht erst in Abschnitt 2. Bis dahin stelle man sich unter der Unabhängigkeit von Zufallsvariablen vor, dass sich die den Zufallsvariablen zugrunde liegenden zufälligen Versuche nicht gegenseitig beeinflussen.

Der vorletzte und letzte Faktor zusammen ergeben das Integral über eine Dichte einer  $N(\mu + \sigma^2 t, \sigma)$ -verteilten Zufallsvariablen, so dass das Produkt dieser Ausdrücke Eins ist. Damit erhalten wir

$$m(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

Die Taylor-Entwicklung dieser Funktion mit der Entwicklungsstelle Null ergibt:

$$\begin{aligned} m(t) &= e^{\mu t + \frac{\sigma^2 t^2}{2}} = \sum_{i=0}^{\infty} \frac{(\mu t + \frac{\sigma^2 t^2}{2})^i}{i!} \\ (21) \quad &= 1 + \left( \mu t + \frac{\sigma^2 t^2}{2} \right) + \frac{\mu^2 t^2}{2} + O(t^3) \\ &= 1 + \mu t + \frac{t^2}{2}(\mu^2 + \sigma^2) + O(t^3). \end{aligned}$$

Andererseits erhalten wir

$$\begin{aligned} m(t) &= \mathbb{E}e^{tZ} = \mathbb{E} \sum_{i=0}^{\infty} \frac{t^i Z^i}{i!} = \sum_{i=0}^{\infty} \frac{t^i}{i!} \mathbb{E}Z^i \\ (22) \quad &= 1 + t\mathbb{E}Z + \frac{t^2}{2}\mathbb{E}Z^2 + O(t^3). \end{aligned}$$

Aus einem Koeffizientenvergleich zwischen (21) und (22) folgt:

$$\begin{aligned} \mu t &= t\mathbb{E}Z & \text{also } \mu &= \mathbb{E}Z \\ \mu^2 + \sigma^2 &= \mathbb{E}Z^2 \\ D^2 Z &= \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2. \end{aligned}$$

Wir haben also bewiesen, dass eine  $N(\mu, \sigma)$  normalverteilte Zufallsvariable den Erwartungswert  $\mu$  und die Varianz  $\sigma^2$  besitzt. Das heißt, der Parameter  $\sigma$  entspricht der Standardabweichung.

Abschließend sei noch erwähnt, dass die Varianz einer exponentialverteilten Zufallsvariable mit Parameter  $\lambda$  gleich  $\lambda^{-2}$  und die einer Poisson-verteilten Zufallsvariablen mit Parameter  $\lambda$  gleich  $\lambda$  ist.

Mit der Varianz verwandt sind die sogenannten Momente.

DEFINITION 4.26. *Es sei  $k \in \mathbb{N}$ . Die Größe*

$$M_k(Z) = \mathbb{E}Z^k$$

*wird als Moment  $k$ -ter Ordnung bezeichnet. Die Größe*

$$M_k^a(Z) = \mathbb{E}|Z|^k$$

*wird als absolutes Moment  $k$ -ter Ordnung bezeichnet. Die Größe*

$$M_k^c(Z) = \mathbb{E}(Z - \mathbb{E}Z)^k$$

*wird als zentrales Moment  $k$ -ter Ordnung bezeichnet.*

Speziell ist das zentrale Moment der zweiten Ordnung die Varianz. Die  $k$ -ten Momente können durch die momenterzeugende Funktion berechnet werden, falls diese existieren. Speziell gilt:

$$\mathbb{E}Z^k = \frac{d^k}{dt^k} m(t)|_{t=0}, \quad k \in \mathbb{N}.$$

Wir betrachten nun die sogenannten *Quantile*. Mittels der Quantile ist es möglich Bereiche festzulegen, in denen die Zufallsvariable mit einer gewissen vorgegebenen Wahrscheinlichkeit  $\alpha$  liegt.

DEFINITION 4.27. *Es sei  $\alpha$  eine Zahl aus dem Intervall  $(0, 1)$ . Dann ist für eine Zufallsvariable  $Z$  mit der Verteilungsfunktion  $F(z)$  das  $\alpha$ -Quantil  $z_\alpha$  gegeben durch:*

$$F(z_\alpha) \geq \alpha \quad \text{und} \quad \alpha > F(z) \quad \text{für} \quad z < z_\alpha.$$

Das 0.5-Quantil wird als *Median* bezeichnet.

Falls die Verteilungsfunktion *streng* monoton wachsend in einer Umgebung ist, in der das  $\alpha$ -Quantil liegt, dann kann das  $\alpha$ -Quantil durch die Lösung der folgenden Gleichung bestimmt werden:

$$\alpha = F(z_\alpha),$$

wobei die rechte Seite nach der Definition 4.1 gleich  $\mathbb{P}(Z \leq z_\alpha)$  ist. Das bedeutet, dass die Werte der Zufallsvariablen  $Z$  mit Wahrscheinlichkeit  $\alpha$  links von  $z_\alpha$  liegen. Der Median teilt also die reelle Zahlengerade in zwei Bereiche, die durch die stetige Zufallsvariable  $Z$  mit Wahrscheinlichkeit 0.5 belegt werden.

Die Bestimmung des Quantils kann als inverse Fragestellung zur Bestimmung einer Wahrscheinlichkeit interpretiert werden. In der Tat haben wir zu einem vorgegebenen Zahlenwert die Wahrscheinlichkeit ermittelt, dass die Zufallsvariable  $Z$  kleiner oder gleich diesem Wert ist. Nun ermitteln wir zu einer gegebenen Wahrscheinlichkeit  $\alpha$  einen Zahlenwert  $z_\alpha$ , so dass die Zufallsvariable  $Z$  mit Wahrscheinlichkeit  $\alpha$  kleiner oder gleich diesem Zahlenwert ist.

Für spezielle Verteilungen, wie zum Beispiel für die Student-,  $\chi^2$ - oder  $F$ -Verteilung, sind die Quantile in Tabellen abgelegt. Um ein Quantil für die Normalverteilung zu bestimmen, können wir die Tabelle der Verteilungsfunktion benutzen. Dazu nehmen wir an, dass  $\alpha \in [0.5, 1)$  liegt. Man bestimme einen Wert in der Nähe von  $\alpha$  im Inneren der Tabelle und suche dazu den entsprechenden *Randwert* heraus. Dieser ist näherungsweise das  $\alpha$ -Quantil.

Zum Abschluss wollen wir noch eine Ungleichung kennenlernen, die es erlaubt abzuschätzen, dass die Wahrscheinlichkeit in gewissen Bereichen liegt.

SATZ 4.28. *Es gilt für  $k > 0$ :*

$$\mathbb{P}(|Z - a| > b) \leq \frac{M_k^a(Z - a)}{b^k}$$

Für  $a \in \mathbb{R}$ ,  $b > 0$ ,  $k \in \mathbb{N}$ . Setzt man speziell für  $a = \mathbb{E}Z$  und  $k = 2$ , so erhält man

$$\mathbb{P}(|Z - \mathbb{E}Z| > b) \leq \frac{D^2 Z}{b^2}.$$

Diese Formeln werden als *Tschebyschev-Ungleichungen* bezeichnet.

## Die Erzeugung von Zufallszahlen

### 1. Die Simulation

Bei der Analyse von komplexen Systemen tritt häufig die Schwierigkeit auf, dass die Anfangs- und Randbedingungen nicht genau bekannt sind. Soll man zum Beispiel gewisse Aussagen über die Länge von Fahrzeugschlangen machen, die sich an Ampeln bilden, so weiß man im allgemeinen nicht, wann welches Fahrzeug die Ampel erreicht. Ein anderes Beispiel ist der Vergleich der Arbeitsweise des *englischen* und *deutschen Schaltersystems*, siehe [14]. Während es beim englischen System nur eine Warteschlange gibt, von der aus sich die Kunden auf die verschiedenen Schalter aufteilen, hat beim deutschen System jeder Schalter seine eigene Warteschlange. Entsprechende Unterschiede bei der Abfertigung sind zum Beispiel auf Flughäfen zu finden.

Um nun Aussagen über die Effektivität beider Abfertigungssysteme zu machen, werden die Ankunftszeiten und Bedienzeiten der einzelnen Kunden durch Zufallszahlen simuliert. Mit diesen zufälligen Ankunfts- und Bedienzeiten können durchschnittliche Werte für die Anzahl der abgefertigten Kunden je Stunde gemacht werden. Methoden, um gewisse Aussagen über die Verteilung der zufälligen Ankunfts- und Bedienzeiten zu gewinnen, stellt die Statistik bereit. Es müssen also Zufallszahlen erzeugt werden, die einer vorgeschriebenen Verteilung genügen. Wie solche Zufallszahlen erzeugt werden, lernen wir im folgende kennen.

### 2. Die Erzeugung gleichverteilter Zufallszahlen

Die grundlegenden Zufallszahlen sind gleichverteilte Zufallszahlen, das heißt Zufallszahlen, die zumindestens näherungsweise einer gleichverteilten Zufallsvariablen entsprechen. Solche Zufallszahlen werden von modernen Programmiersprachen wie C++ bereitgestellt. Andererseits können aus gleichverteilten Zufallszahlen mit anderer Verteilung gebildet werden. Wir beginnen deshalb mit der Beschreibung dieser Zufallszahlen.

Bei einer gleichverteilten Zufallsvariablen über dem Intervall  $(0, 1)$  hing die Wahrscheinlichkeit, dass diese Zufallsvariable Werte in einem Teilintervall  $I \subset [0, 1]$  hat, nur von der Länge, nicht aber von der Lage des Intervalls ab. Würde sich nun in einer Box zehn Kugeln mit den Ziffern  $0, \dots, 9$  befinden und würde man nun diese Kugeln zufällig und unendlich oft aus der Box entnehmen, notieren und wieder zurücklegen, so würde man eine Folge von Zahlen

$$i_1, i_2, \dots, i_n, \dots$$

erhalten. Diese Folge bestimmt die zufällige Zahl  $0.i_2i_2 \dots i_n \dots$  aus dem Intervall  $[0, 1]$ . Diese theoretisch gebildeten Zahlen beschreiben gleichverteilte Zufallszahlen

aus dem Intervall  $[0, 1]$ . Bei der praktischen Bestimmung dieser Zahlen würden allerdings unüberwindliche Schwierigkeiten auftreten. So wäre man nicht in der Lage, eine unendliche Folge dieser Ziffern zu bestimmen. Man könnte also nur durch die  $n$ -malige Entnahme von Kugeln Zufallszahlen, die eine *diskrete Gleichverteilung* besitzen, bestimmen. Bei dieser diskreten Gleichverteilung würde jede Dezimalzahl zwischen Null und Eins mit  $n$  Stellen nach dem Komma mit der gleichen Wahrscheinlichkeit angenommen werden. So eine diskrete Verteilung ist eine Approximation der stetigen Gleichverteilung, wenn  $n$  hinreichend groß gewählt wird.

Eine weitere Schwierigkeit, die bei der Realisierung dieser Zufallszahlen auf einem Computer auftreten würden ist, dass ein *Initialzufall* benötigt würde, um die *zufällige* Entnahme der zehn Kugeln aus der Box zu realisieren.

All diese Probleme führen dazu, dass man andere Methoden anwendet, um durch einen Algorithmus auf einem Computer gleichverteilte zufällige Zahlen zu erzeugen. Die erste Methode ist die *lineare Kongruenzmethode*. Dabei werden durch eine Iterationsformel zufällige Zahlen erzeugt.

Es seien  $m, a, b$  natürliche Zahlen und  $z_0 \in \mathbb{Z}^+, z_0 \leq m$ . Man erhält eine Folge von Zahlen wie folgt:

$$z_{j+1} = a z_j + b \pmod{m}. \quad j = 0, 1, \dots$$

Dann werden die Glieder der Folge  $x_j = \frac{z_j}{m}$  als gleichverteilte Zufallszahlen interpretiert.  $\text{mod } m$  ist dabei der Rest, der bei der Division durch  $m$  entsteht. So ergibt sich zum Beispiel:

$$46 \pmod{7} = 4, \quad \text{denn } 46 = 6 \cdot 7 + 4.$$

Setzt man zum Beispiel  $m = 100, a = 31, b = 47$  und  $z_0 = 73$ , so erhält man:

$$\tilde{z}_1 = 31 \cdot 73 + 47 = 2310, \quad z_1 = 2310 \pmod{100} = 10$$

$$\tilde{z}_2 = 31 \cdot 10 + 47 = 357, \quad z_2 = 357 \pmod{100} = 57$$

$$\tilde{z}_3 = 31 \cdot 57 + 47 = 1814, \quad z_3 = 1814 \pmod{100} = 14$$

...

Damit interpretiert man die Folge

$$x_1 = 0.1, x_2 = 0.57, x_3 = 0.14, \dots$$

als Folge von gleichverteilten Zufallsvariablen.

Bei einer günstigen Wahl der Parameter  $a, b, m$  erhält man unregelmäßig verteilte Zahlen im Intervall  $[0, 1]$ , die den Anschein erwecken, zufällig zu sein. Schaut man allerdings etwas genauer auf die Eigenschaften dieser Folge, so erkennt man Eigenschaften, die einer Zufälligkeit widersprechen. Die Folge  $x_1, x_2, \dots$  ist durch den Startwert  $z_0$  *vollständig bestimmt*. Das bedeutet, startet man mit dem gleichen Startwert, so erhält man immer wieder die gleiche Folge. Das widerspricht natürlich der Unregelmäßigkeit, der dem wahren Zufall innewohnen sollte. Eine weitere Eigenschaft, die dem Zufall widerspricht, ist die Periodizität dieser Folge, wobei die maximale Periode  $m$  ist. Aufgrund dieser vom *reinen*<sup>1</sup> Zufall abweichenden Eigenschaften nennt man diese Zahlen *Pseudozufallszahlen*.

Weitere Probleme können auftreten, bei einer gewissen Wahl der Parameter  $a, b$  und  $m$ . Wie man in der ersten Reihe der Abbildung 1 erkennt, treten bei den Parametern

$$m = 500, \quad a = 41, \quad b = 343, \quad z_0 = 251$$

<sup>1</sup>was immer das ist

Muster auf, siehe [7] Seite 160. Dabei ist im ersten Bild die durch die lineare Kongruenzmethode erzeugte Folge  $x_1, \dots, x_{1000}$  von Zufallszahlen dargestellt, während das zweite Bild Paare  $(x_1, x_2), (x_3, x_4), \dots$  enthält. Bei der Wahl von

$$m = 1000000, \quad a = 411111, \quad b = 823543, \quad z_0 = 499777$$

treten solche Muster nicht auf, siehe zweite Reihe der Abbildung 1.

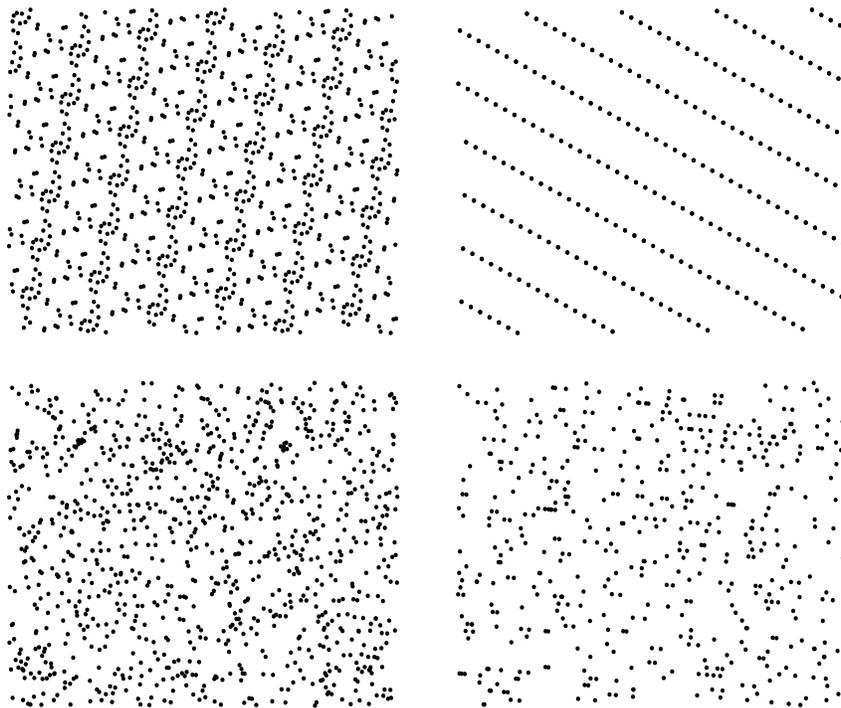


ABBILDUNG 1. lineare Kongruenzmethode zur Erzeugung von Zufallszahlen

Um eine sehr große Periodenlänge durch die Wahl von  $a, b, m$  zu erhalten, sollte man folgendes Kriterium berücksichtigen:

- $b$  und  $m$  sind teilerfremd.
- Jede Primzahl die  $m$  teilt, teilt auch  $a - 1$ .
- Ist  $m$  durch Vier teilbar, so ist auch  $a - 1$  durch 4 teilbar,

siehe [7] Seite 159 oder [10] Seite 16.

Eine andere Methode um gleichverteilte Zufallszahlen zu erzeugen, ist die Quadratmittelmethode. Dabei wird eine natürliche Zahl quadriert und dann der mittlere Teil der Ziffernfolge dieser Zahl zur Bildung einer Zufallszahl herangezogen: Wählt

man als Startwert die  $2n$  stellige Zahl  $z_0 = 5371$  für  $n = 2$ , so ist

$$\begin{aligned} z_0^2 &= 28847641, & z_1 &= 8476 \\ z_1^2 &= 71690089, & z_2 &= 6900 \\ z_2^2 &= 47610000, & z_3 &= 6100. \end{aligned}$$

Falls die entstandene Zahl weniger als  $2n$  Stellen besitzt, werden vorne Nullen ergänzt. Die Folge der gleichverteilten Zufallszahlen wäre dann gegeben durch  $x_1 = 0.8476$ ,  $x_2 = 0.69$ ,  $x_3 = 0.61, \dots$ .

Um die Qualität von Zufallszahlen zu überprüfen, werden verschiedene statistische Tests angewandt.

### 3. Erzeugung von Zufallszahlen mittels der Verteilungsfunktion

Es kann vorkommen, dass zur Durchführung von bestimmten Simulationen andere Verteilungen benötigt werden als die Gleichverteilung. Wir wollen Zufallszahlen mit einer Verteilung  $F_X(x)$  durch Transformation einer  $[0, 1]$ -gleichverteilten Zufallsvariablen  $Z$  erzeugen.  $Z$  entspricht also den Zufallszahlen, die zum Beispiel mit der linearen Kongruenzmethode im letzten Abschnitt erzeugt wurden. Wir erinnern auch daran, dass die Verteilungsfunktion von  $Z$  gegeben ist durch

$$F_Z(z) = \begin{cases} 0 & : & z \leq 0 \\ z & : & z \in (0, 1] \\ 1 & : & z > 1 \end{cases}.$$

Wir wollen zuerst auch annehmen, dass  $x \rightarrow F_X(x)$  invertierbar ist, das heißt, dass  $F_X(x)$  *streng* monoton wachsend ist, so dass die Umkehrfunktion  $F_X^{-1}$  existiert.

SATZ 5.1. *Wir definieren die Zufallsvariable  $X$  durch*

$$(23) \quad X := F_X^{-1}(Z),$$

wobei  $Z$  eine  $[0, 1]$ -gleichverteilte Zufallsvariable ist. Dann besitzt  $X$  die Verteilungsfunktion  $F_X(x)$ .

BEWEIS. Wir berechnen die Verteilungsfunktion von  $X$ :

$$\mathbb{P}(X \leq x) = \mathbb{P}(F_X^{-1}(Z) \leq x).$$

Aufgrund der Existenz der Umkehrfunktion, ist die rechte Seite gleich  $\mathbb{P}(Z \leq F_X(x))$ . Nun erinnern wir uns an die Definition der Verteilungsfunktion von  $Z$  und erhalten:

$$\mathbb{P}(F_X^{-1}(Z) \leq x) = \mathbb{P}(Z \leq F_X(x)) = F_Z(F_X(x)) = F_X(x).$$

□

Hat man nun eine Folge von  $[0, 1]$ -gleichverteilten (Pseudo)zufallszahlen  $z_1, z_2, \dots$ , so besteht die Folge

$$x_1 = F_X^{-1}(z_1), x_2 = F_X^{-1}(z_2), \dots$$

aus (Pseudo)zufallszahlen mit der Verteilung von  $X$ .

Viele Verteilungsfunktionen besitzen nicht die Eigenschaft, *streng* monoton zu sein. Speziell sind die Verteilungsfunktionen aller diskreten Verteilungen nicht *streng* monoton. Um trotzdem mit der Transformationsformel arbeiten zu können (23), benötigen wir die *verallgemeinerte Inverse*. Man erinnere sich noch einmal an den Begriff des *Infimums* einer Menge von reellen Zahlen (Bezeichnung: *inf*), das durch die größte untere Schranke dieser Menge gegeben ist.

DEFINITION 5.2. Für eine Verteilungsfunktion  $F(x)$ , definiert auf  $\mathbb{R}$  und mit einem Wertebereich enthalten in  $\mathbb{R}$  bezeichnen wir

$$F^{-1}(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}$$

als verallgemeinerte Inverse von  $F(x)$ , wobei folgende Konventionen gelten:

$$\inf \mathbb{R} = -\infty, \quad \inf \emptyset = \infty.$$

Die verallgemeinerte Inverse wird mit der gleichen Symbolik bezeichnet wie die Umkehrfunktion. Falls  $F(x)$  eine Umkehrfunktion besitzt, so ist diese gleich der verallgemeinerten Inversen.  $\inf$  ist das Infimum einer Zahlenmenge, das heißt die größte untere Schranke dieser Menge.

**Beispiel:** Es sei  $F(x)$  die Funktion:

$$F(z) = \begin{cases} 0 & : & z < 0 \\ \frac{1}{2} & : & z \in [0, 1) \\ 1 & : & z \geq 1 \end{cases}.$$

Dann gilt für die verallgemeinerte Inverse

$$\begin{aligned} F^{-1}(0) &= \inf\{x \in \mathbb{R}, F(x) \geq 0\} = \inf \mathbb{R} = -\infty \\ F^{-1}(0.25) &= \inf\{x \in \mathbb{R}, F(x) \geq 0.25\} = 0 \\ F^{-1}(0.5) &= \inf\{x \in \mathbb{R}, F(x) \geq 0.5\} = 0 \\ F^{-1}(0.75) &= \inf\{x \in \mathbb{R}, F(x) \geq 0.75\} = 1 \\ F^{-1}(1) &= \inf\{x \in \mathbb{R}, F(x) \geq 1\} = 1. \end{aligned}$$

SATZ 5.3. Angenommen, es gelten die Voraussetzungen von Satz 5.1, wobei  $F_X^{-1}$  die verallgemeinerte Inverse der Verteilungsfunktion von  $F_X$  ist. Dann besitzt die durch (23) generierte Zufallsvariable die Verteilungsfunktion  $F_X(x)$ .

BEWEIS. Die verallgemeinerte Inverse besitzt die Eigenschaft

$$(24) \quad F_X^{-1}(y) \leq x \quad \text{dann und nur dann, wenn} \quad y \leq F_X(x).$$

Damit bleibt die Gleichungskette aus dem Beweis des Satzes 5.1 richtig. Wir beweisen (24). Dazu benutzen wir die Monotonie und rechtseitige Stetigkeit von  $F_X$ .

( $\Rightarrow$ )

$$F_X^{-1}(y) = \inf\{u : F_X(u) \geq y\} =: u_{inf} \leq x$$

nach Definition der verallgemeinerten Inverse von  $F_X$ . Da  $F_X$  als Verteilungsfunktion rechtsseitig stetig ist gilt:  $F_X(u_{inf}) \geq y$  und wegen der Monotonie  $F_X(x) \geq y$ .

( $\Leftarrow$ ) Falls  $F_X(x) \geq y$  für ein  $x$ , dann gilt wegen der Infimumbildung und der Monotonie

$$F_X^{-1}(y) = \inf\{u : F_X(u) \geq y\} \leq x.$$

□

**Beispiel:** Es sei

$$F_X(x) = \begin{cases} 0 & : & x < 0 \\ 1 - e^{-\lambda x} & : & x \geq 0 \end{cases}$$

die Verteilungsfunktion für eine exponentialverteilte Zufallsvariable  $X$  mit Parameter  $\lambda > 0$ . Die verallgemeinerte Inverse ist gegeben durch

$$F_X^{-1}(y) = -\frac{1}{\lambda} \ln(1 - y) \quad \text{für } y \in (0, 1).$$

Ist also  $z_1, z_2, \dots$  eine Folge von gleichverteilten Zufallszahlen, dann ist

$$x_1 = -\frac{1}{\lambda} \ln(1 - z_1), x_2 = -\frac{1}{\lambda} \ln(1 - z_2), \dots$$

eine Folge exponentialverteilter Zufallszahlen. Es sei noch bemerkt, dass eine  $[0, 1]$ -gleichverteilte Zufallsvariable  $Z$  die beiden Randwerte 0, 1 nur mit Wahrscheinlichkeit Null, also praktisch nie annimmt.

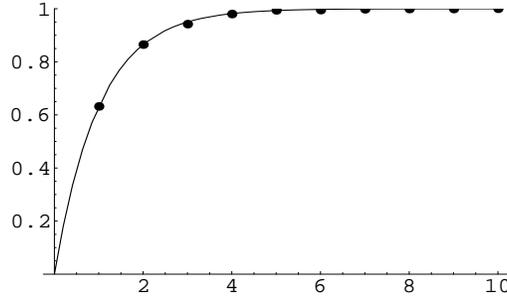


ABBILDUNG 2. Vergleich der empirischen Verteilungsfunktion (Punkte) von durch die Transformationsmethode erzeugte Zufallszahlen mit der tatsächlichen Verteilungsfunktion

**Beispiel:** Wir betrachten Zufallszahlen nach der geometrischen Verteilung mit Parameter  $p \in (0, 1)$ . Eine geometrisch verteilte Zufallsvariable besitzt eine diskrete Verteilung mit der in Abbildung dargestellten Verteilungsfunktion. Um die Verteilungsfunktion durch eine Formel beschreiben zu können, führen wir den Abrundungsoperator  $\lfloor \cdot \rfloor$  und Aufrundungsoperator  $\lceil \cdot \rceil$  ein. Speziell gilt:

$$\lfloor 2.5 \rfloor = 2, \quad \lceil 2.5 \rceil = 3.$$

Mit Hilfe des Abrundungsoperators kann nun die Verteilungsfunktion der geometrischen Verteilung dargestellt werden. Es gilt:

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k=1}^{\lfloor x \rfloor} (1-p)p^{k-1} = (1-p) \sum_{k=1}^{\lfloor x \rfloor} p^{k-1} = 1 - p^{\lfloor x \rfloor}.$$

Durch Umstellen erhält man

$$\lfloor x \rfloor = \frac{\ln(1-z)}{\ln p}.$$

Am Graphen der Funktion  $f(x) = \lfloor x \rfloor$  kann man sich leicht klar machen, dass die verallgemeinerte Inverse von  $\lfloor \cdot \rfloor$  gerade  $\lceil \cdot \rceil$  ist. Also gilt

$$\begin{aligned} F_X^{-1}(z) &= \inf\{x \in \mathbb{R} : F_X(x) \geq z\} = \inf\{x \in \mathbb{R} : 1 - p^{\lfloor x \rfloor} \geq z\} \\ &= \inf\{x \in \mathbb{R} : 1 - z \geq p^{\lfloor x \rfloor}\} = \inf\left\{x \in \mathbb{R} : \frac{\log(1-z)}{\log p} \leq \lfloor x \rfloor\right\} \\ &= \inf\left\{x \in \mathbb{R} : \left\lceil \frac{\log(1-z)}{\log p} \right\rceil \leq x\right\} = \left\lceil \frac{\ln(1-z)}{\ln p} \right\rceil \end{aligned}$$

gilt. Somit ist für  $z_1, z_2, \dots$  eine Folge von gleichverteilten Zufallszahlen

$$x_1 = \left\lceil \frac{\ln(1 - z_1)}{\ln p} \right\rceil, x_2 = \left\lceil \frac{\ln(1 - z_2)}{\ln p} \right\rceil, \dots$$

eine Folge geometrisch verteilter Zufallszahlen.

#### 4. Normalverteilte Zufallsvariablen und zentraler Grenzwertsatz

Die Erzeugung von Zufallszahlen nach der im letzten Abschnitt beschriebenen Methode stößt auf Schwierigkeiten, denn ein expliziter Ausdruck für die Verteilungsfunktion der Standardnormalverteilung  $\Phi(x)$  ist unbekannt. Damit kann aus  $\Phi(x)$  nicht deren inverse Funktion angegeben werden.

Andererseits treten normalverteilte Zufallsvariablen bei vielen Anwendungen in Natur-, Ingenieur- und Wirtschaftswissenschaften auf. So zum Beispiel kann die Bahn eines kleinen Partikels in einer Flüssigkeit durch eine normalverteilte Zufallsvariable beschrieben werden. Speziell ist zu beobachten, dass so ein Partikel nicht im Ruhezustand verharrt, sondern eine Zickzackbewegung vollführt, siehe [11] Seite 45.

Die Erzeugung von normalverteilten Zufallsvariablen wird auf der Überlagerung von anderen Zufallsvariablen mit speziellen mathematischen Eigenschaften beruhen. Dies erklärt, warum Messfehler häufig (näherungsweise) normalverteilt sind, ergeben sie sich doch als Überlagerung vieler Störeinflüsse.

Die mathematische Formulierung dieses Überlagerungsprinzips ist der zentrale Grenzwertsatz.

**SATZ 5.4.** *Es sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge von unabhängigen Zufallsvariablen mit gleicher Verteilung. Weiterhin gilt für  $n \in \mathbb{N}$*

$$D^2 X_n =: \sigma^2 < \infty, \quad \mathbb{E} X_n =: \mu.$$

*Wir definieren eine Folge von Zufallsvariablen  $(S_n)_{n \in \mathbb{N}}$  durch*

$$S_n = X_1 + \dots + X_n.$$

*Dann gilt für  $a < b$*

$$(25) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left( a < \frac{S_n - \mu n}{\sigma \sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a)$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{S_n - \mu n}{\sigma \sqrt{n}} \leq b \right) = \Phi(b).$$

Dieser Satz sagt aus, dass die Zufallsvariablen  $S_n$  näherungsweise normalverteilt sind, falls  $n$  hinreichend groß ist. Die transformierten Zufallsvariablen

$$(26) \quad Y_n = \frac{S_n - \mu n}{\sigma \sqrt{n}}$$

sind näherungsweise standardnormalverteilt.

**Beispiel:** Früher wurde die *Zwölferregel* benutzt, um näherungsweise normalverteilte Zufallsvariablen zu erzeugen. Dabei ergibt sich  $S_{12}$  als eine Summe von zwölf  $[0, 1]$ -gleichverteilten und unabhängigen Zufallsvariablen  $X_1, \dots, X_{12}$ , siehe Abbildung 3.

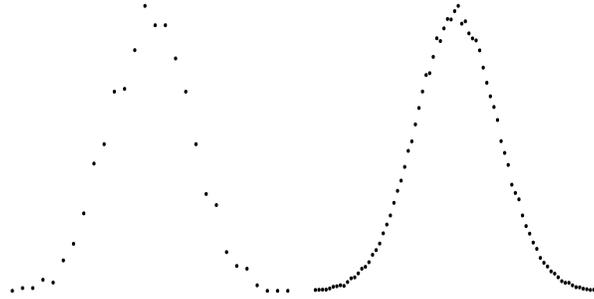


ABBILDUNG 3. Approximation der Normalverteilung durch  $S_{12}$  und  $S_{1000}$

Aufgrund

$$\mathbb{E}X_i = 0.5, \quad D^2X = \frac{1}{12}$$

ist nach Satz 5.4  $S_{12} - 6$  näherungsweise standardnormalverteilt. Wegen  $n = 12$  ist der Nenner in (26) immer 1.

Ein Spezialfall des zentralen Grenzwertsatzes ist der Satz von de Moivre-Laplace. Dabei wird vorausgesetzt, dass die unabhängigen Zufallsvariablen  $X_n$  nur die Werte Null und Eins annehmen.

$$(27) \quad X_i = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p \\ 0 & \text{mit Wahrscheinlichkeit } 1 - p \end{cases}.$$

Setzt man noch voraus, dass  $X_i = 1$  genau dann, wenn ein Ereignis  $A$  eintritt und  $X_i = 0$  genau dann, wenn ein Ereignis  $\bar{A}$  eintritt für  $i = 1, \dots, n$ , so beschreiben die Zufallsvariablen  $X_1, \dots, X_n$  ein Bernoulli-Experiment. Speziell gibt die Zufallsvariable

$$S_n = X_1 + \dots + X_n$$

an, wie oft bei  $n$  Versuchen das Ereignis  $A$  eintritt.  $S_n$  ist also  $(n, p)$ -binomialverteilt. Die Berechnung der Wahrscheinlichkeiten dieser Verteilung kann sehr aufwendig für große  $n$  sein, da in den auftretenden Binomialkoeffizienten Fakultäten von sehr großen Zahlen enthalten sind. Näherungsweise können diese Wahrscheinlichkeiten dann mit dem Satz von de Moivre-Laplace berechnet werden.

SATZ 5.5. *Es sei  $S_n$  eine Folge von  $(n, p)$ -binomialverteilter Zufallsvariablen, wobei  $p \in (0, 1)$  fest gewählt ist. Dann gilt:*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( a < \frac{S_n - pn}{\sqrt{np(1-p)}} \leq b \right) = \Phi(b) - \Phi(a)$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{S_n - pn}{\sqrt{np(1-p)}} \leq b \right) = \Phi(b).$$

Die Zufallsvariablen  $S_n$  sind also wieder näherungsweise normalverteilt.

BEWEIS. Für die in (27) eingeführten Zufallsvariablen gilt:

$$\mathbb{E}X_i = 1 \cdot p + 0 \cdot (1 - p) = p = \mu$$

$$D^2X_i = (1 - p)^2 p + (0 - p)^2 (1 - p) = p(1 - p) = \sigma^2.$$

Es ergibt sich für den Erwartungswert von  $S_n$  als Summe von  $n$  Zufallsvariablen nach Satz 4.19  $\mathbb{E}S_n = np$  und für die Varianz dieser Summe von  $n$  Zufallsvariablen nach Satz 4.25  $D^2S_n = n(1-p)p$ . Die Aussage folgt nun direkt durch die Anwendung des zentralen Grenzwertsatzes 5.4.  $\square$

**Beispiel:** Mit Wahrscheinlichkeit 0.8 gehört ein produzierter Computerchip zur Qualitätsklasse Eins. Die Tagesproduktion beträgt 2000 Stück. Wie groß ist die Wahrscheinlichkeit, dass mehr als 1600 an einem Tage produzierte Chips zur Qualitätsklasse Eins gehören, wenn die Voraussetzungen des Bernoulli-Experimentes erfüllt sind?

Die Zufallsvariable, die die Anzahl der qualitätsgerechten Chips angibt sei  $S_{2000}$ . Die gesuchte Wahrscheinlichkeit ist dann gegeben durch

$$(28) \quad \mathbb{P}(S_{2000} > 1600) = \sum_{j=1601}^{2000} \binom{2000}{j} 0.8^j 0.2^{2000-j},$$

eine zugegeben verausgabende Tätigkeit, all die auftretenden Binomialkoeffizienten zu berechnen. Mit dem Satz von de Moivre-Laplace erhalten wir

$$\begin{aligned} \mathbb{P}(S_{2000} > 1600) &= 1 - \mathbb{P}(S_{2000} \leq 1600) \\ &= 1 - \mathbb{P}\left(\frac{S_{2000} - 2000 \cdot 0.8}{\sqrt{2000 \cdot 0.8 \cdot 0.2}} \leq \frac{\overbrace{1600 - 2000 \cdot 0.8}^{=0}}{\sqrt{2000 \cdot 0.8 \cdot 0.2}}\right) \approx 1 - \Phi(0) = 0.5. \end{aligned}$$

Zum Abschluß seien noch die Stirlingsche Formel erwähnt, die es erlaubt, Fakultäten  $n!$  durch einfacher zu berechnende Ausdrücke zu ersetzen, falls  $n$  sehr groß ist.

SATZ 5.6. *Es gilt*

$$\lim_{n \rightarrow \infty} \frac{n^n e^{-n} \sqrt{2\pi n}}{n!} = 1.$$

*Das bedeutet für große  $n$  kann  $n!$  durch  $n^n e^{-n} \sqrt{2\pi n}$  ersetzt werden.*

Mit dieser Formel erhält man

$$\mathbb{P}(S_n = 1600) = \binom{2000}{1600} 0.8^{1600} 0.2^{400} \approx \frac{1}{\sqrt{2\pi 320}} = 0.0223,$$

wobei die Potenzen durch Kürzen wegfallen.



## Mehrdimensionale Wahrscheinlichkeitsverteilungen

### 1. Allgemeine Eigenschaften von mehrdimensionalen Verteilungen

Im Folgenden wollen wir annehmen, dass bei einem zufälligen Versuch zwei (oder mehrere) zufällige Größen auftreten. Diese beiden zufälligen Parameter werden durch Zufallsvariable charakterisiert. Als Beispiel stelle man sich vor, dass in verschiedenen Proben einer chemischen Substanz einerseits der Anteil  $X$  eines Zugschlagsstoffes, zum anderen die Wärmeleitfähigkeit  $Y$  gemessen werde.

Es interessiert, ob es Zusammenhänge zwischen den Werten der beiden Zufallsvariablen  $(X, Y)$  gibt. Speziell soll der Grad der Abhängigkeit/Unabhängigkeit dieser beiden Zufallsvariablen gemessen werden. Bevor wir aber Maßzahlen einführen, die erlauben Aussagen über den Grad der Abhängigkeit/Unabhängigkeit zu machen, wollen wir zuerst die Wahrscheinlichkeitsverteilung des Paares  $(X, Y)$  beschreiben. Wir betrachten einen zufälligen Versuch, durch den zwei Zufallsvariable  $X, Y$  beobachtet werden. Es könnte zum Beispiel die Wahrscheinlichkeit interessieren, ob das Paar  $(X, Y)$  Werte in einem Rechteck

$$R = \{(x, y) : a < x \leq b, c < y \leq d\}, \quad a \leq b, \quad c \leq d$$

annimmt:  $\mathbb{P}((X, Y) \in R)$ . Um solche Wahrscheinlichkeiten zu berechnen, wird der Begriff der Verteilungsfunktion des Paares von Zufallsvariablen  $(X, Y)$  eingeführt.

DEFINITION 6.1. *Eine Funktion mit dem Definitionsbereich  $\mathbb{R}^2$ , definiert durch*

$$F(x, y) = \mathbb{P}(X \leq x \wedge Y \leq y)$$

*heißt Verteilungsfunktion des zufälligen Paares (Vektors)  $X, Y$ .*

Aus der Definition folgt sofort, dass  $F(x, y)$  Werte zwischen Null und Eins annimmt. Speziell ist die Wahrscheinlichkeit für das Rechteck  $R$  gegeben durch

$$\mathbb{P}((X, Y) \in R) = F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Es ist nun nicht schwer eine Verallgemeinerung der Definition 6.1 für  $r$ -Tupel von Zufallsvariablen  $(X_1, \dots, X_r)$  vorzunehmen.

**Beispiel:** Eine diskrete mehrdimensionale Verteilung ist die Multinomialverteilung. Dazu betrachten wir einen Versuch mit  $r \in \mathbb{N}$  verschiedenen Ausgängen  $A_1, A_2, \dots, A_r$ . Der Versuch werde nun  $n$ -mal unabhängig voneinander und unter gleichen Bedingungen ausgeführt.

$$\begin{aligned} A_1 \text{ tritt } x_1 \text{ mal auf mit } \mathbb{P}(A_1) &= p_1 \\ \dots & \\ A_r \text{ tritt } x_r \text{ mal auf mit } \mathbb{P}(A_r) &= p_r \end{aligned}$$

wobei  $x_1 + \dots + x_r = n$  gilt. Es seien nun  $X_i$  die Zufallsvariablen, die zählen, wie häufig  $A_i$  für  $i = 1, \dots, r$  auftreten. Dann ist:

$$\mathbb{P}(X_1 = x_1, \dots, X_r = x_r) = \frac{n!}{x_1! \cdot \dots \cdot x_r!} p_1^{x_1} \cdot \dots \cdot p_r^{x_r}$$

$$x_i = 1, \dots, n, \quad \sum_{i=1}^r x_i = n.$$

Falls  $r = 2$  gesetzt wird, erhalten wir die wohlbekanntere Binomialverteilung. Nehmen wir nun an, das Paar  $(X, Y)$  bestehe aus zwei stetigen Zufallsvariablen. Dann kann die Wahrscheinlichkeitsverteilung dieses Paares wieder durch eine Dichtefunktion  $f(x, y)$  beschrieben werden. Diese Dichtefunktion hat nun als Definitionsbereich den  $\mathbb{R}^2$ , in dem die Werte der Paare  $(X, Y)$  liegen. Speziell gilt für die Wahrscheinlichkeitsverteilung

$$\mathbb{P}(X \leq x \wedge Y \leq y) = F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv.$$

Als Beispiel sei die zweidimensionale Normalverteilung angeführt. Diese besitzt die Dichte

$$(29) \quad f(\vec{z}) = \frac{1}{2\pi\sqrt{\Delta}} \exp\left(-\frac{1}{2}(\vec{z} - \vec{\mu})^T \Sigma^{-1}(\vec{z} - \vec{\mu})\right),$$

$$\vec{z} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \vec{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$$

$$\Delta = \det(\Sigma), \quad \Sigma^{-1} = \frac{1}{\Delta} \begin{pmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{pmatrix}.$$

Die Größen  $\mu_X, \mu_Y$  entsprechen den Erwartungswerten von  $X, Y$ .  $\sigma_X^2, \sigma_Y^2$  sind die Varianzen von  $X, Y$ . Die Größe  $\sigma_{XY}$  ist die *Kovarianz* zwischen  $X$  und  $Y$ , die die Richtung der Abhängigkeit von  $X$  und  $Y$  beschreibt. Sie wird im übernächsten Abschnitt eingeführt. Ist speziell  $\mu_X = \mu_Y = \sigma_{XY} = 0$  und  $\sigma_X^2 + \sigma_Y^2 = 1$ , so spricht man von einer zweidimensionalen Standardnormalverteilung. In Abbildung 1 sind die Niveaulinien für

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad \Sigma_2 = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$$

dargestellt.

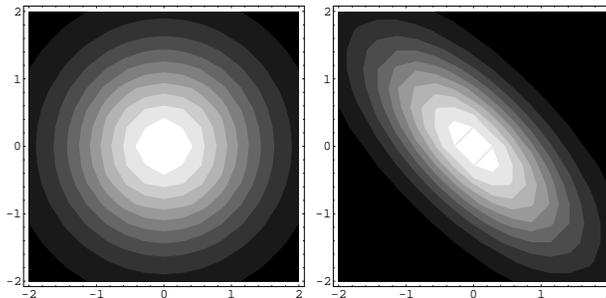


ABBILDUNG 1. Mehrdimensionale Normalverteilung für zwei Matrizen  $\Sigma_1, \Sigma_2$

Um die Wahrscheinlichkeit zu berechnen, dass das Paar  $(X, Y)$  Werte in einem zweidimensionalen Bereich annimmt, wird wie folgt vorgegangen:  
Falls  $X, Y$  diskrete Zufallsvariablen mit den Werten  $x_i, y_j$  sind und die Wahrscheinlichkeiten  $\mathbb{P}(X_i = x_i \wedge Y_j = y_j)$  mit  $p_{ij}$  bezeichnet werden, dann ist

$$\mathbb{P}((X, Y) \in G) = \sum_{(x_i, y_j) \in G} p_{ij}.$$

Falls  $X, Y$  stetige Zufallsvariablen sind, so muss das Bereichsintegral bezüglich der Dichte von  $(X, Y)$  berechnet werden:

$$\mathbb{P}((X, Y) \in G) = \int \int_G f(x, y) dy dx.$$

## 2. Randverteilungen und Unabhängigkeit

Gegeben sei die Verteilungsfunktion des zufälligen Vektors  $(X, Y)$ . Wir fragen, ob es eine Möglichkeit gibt, aus der Verteilungsfunktion  $F(x, y)$  die Verteilungen von  $X$  und  $Y$  zu generieren. Diese Verteilungen werden auch als Randverteilungen von  $(X, Y)$  bezeichnet. Im Falle von stetigen Zufallsvariablen  $X, Y$  gilt für die Verteilungsfunktion von  $X$

$$(30) \quad F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x \wedge Y \in \mathbb{R}) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, v) dv du.$$

Damit hat  $X$  die Dichte

$$f_X(x) = \int_{-\infty}^{\infty} f(x, v) dv,$$

die auch als Randdichte von  $(X, Y)$  bezeichnet wird. Analog kann auch  $F_Y(y)$ ,  $f_Y(y)$  berechnet werden.

Um  $F_X(x)$ , also  $\mathbb{P}(X \leq x)$  auszurechnen, werden an die Zufallsvariable  $Y$  keine Einschränkungen bezüglich des Bereiches gemacht, in dem die Werte von  $Y$  liegen, das heißt  $Y \in (-\infty, \infty)$ . Damit gilt

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x \wedge Y \in (-\infty, \infty)).$$

Daraus resultiert das innere Integral in Formel (30) von  $-\infty$  bis  $\infty$ . Warum man diese Verteilungen gerade Randverteilungen nennt, erklärt das folgende Beispiel eines Paares von diskreten Zufallsvariablen  $(X, Y)$ . Die Wahrscheinlichkeitsverteilung dieses Paares ist durch die Tabelle 1 gegeben.

$X Y$	$y_1$	$y_2$	$y_3$	$y_4$	
$x_1$	0.05	0.1	0.05	0.05	0.25
$x_2$	0.15	0.2	0.1	0.05	0.5
$x_3$	0.1	0.1	0	0.05	0.25
	0.3	0.4	0.15	0.15	1

TABELLE 1. Verteilung einer diskreten zweidimensionalen Zufallsvariablen

So ist zum Beispiel die Wahrscheinlichkeit

$$\mathbb{P}(X = x_1 \wedge Y = y_3) = 0.05.$$

Die Wahrscheinlichkeit, dass  $X = x_2$  ist, kann berechnet werden als

$$\begin{aligned} \mathbb{P}(X = x_2) = & \mathbb{P}(X = x_2 \wedge Y = y_1) + \mathbb{P}(X = x_2 \wedge Y = y_2) + \mathbb{P}(X = x_2 \wedge Y = y_3) \\ & + \mathbb{P}(X = x_2 \wedge Y = y_4). \end{aligned}$$

Diese Wahrscheinlichkeit ergibt sich aus der Summation der Wahrscheinlichkeiten in der entsprechenden Zeile und kann am *Rand* abgelesen werden.

Wir wollen jetzt die Unabhängigkeit von zwei Zufallsvariablen  $X$  und  $Y$  untersuchen. Da den Zufallsvariablen zufällige Versuche entsprechen, kann man heuristisch sagen, dass wenn sich die Versuche  $X$  und  $Y$  nicht beeinflussen Unabhängigkeit zwischen  $X$  und  $Y$  gegeben ist. Wir hatten schon einmal die Unabhängigkeit zwischen Ereignissen  $A$  und  $B$  untersucht, siehe Kapitel 3. Zwei Ereignisse  $A$  und  $B$  heißen unabhängig falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Wir wollen nun die Unabhängigkeit von zwei Zufallsvariablen auf die Unabhängigkeit von Ereignissen zurückführen.

**DEFINITION 6.2.** *Zwei Zufallsvariablen  $X$  und  $Y$  heißen unabhängig, falls alle durch  $X$  generierten Ereignisse von den durch  $Y$  generierten Ereignissen unabhängig sind.*

Dabei versteht man unter den Ereignissen, die durch  $X$  generiert werden, die elementaren Ausgänge des zufälligen Versuches  $X$ , die in einem gewissen Intervall liegen

$$A_{X,a,b} = (a < X \leq b) \quad \text{für } a < b$$

und ähnlich für  $Y$ . Unabhängigkeit heißt damit

$$\mathbb{P}(A_{X,a,b} \cap A_{Y,c,d}) = \mathbb{P}(A_{X,a,b}) \cdot \mathbb{P}(A_{Y,c,d}).$$

Speziell gilt dies auch für die Ereignisse  $\{X \leq x\}$ ,  $\{Y \leq y\}$ , also

$$\mathbb{P}(X \leq x \wedge Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y).$$

Aufgrund der Definition 6.1 der Verteilungsfunktion entspricht dieser Gleichung:

$$F(x, y) = F_X(x) \cdot F_Y(y).$$

Dabei ist  $F(x, y)$  die Verteilungsfunktion des Paares  $(X, Y)$ . Falls  $X$  und  $Y$  stetige Zufallsvariablen sind, gilt nach Differentiation der Verteilungsfunktionen

$$(31) \quad f(x, y) = f_X(x) \cdot f_Y(y).$$

Die Umkehrung gilt auch:

**SATZ 6.3.** *Falls für das Paar  $(X, Y)$  gilt:*

$$F(x, y) = F_X(x) \cdot F_Y(y),$$

*dann sind  $X$  und  $Y$  unabhängig.*

Für unabhängige Zufallsvariable lässt sich besonders einfach der Erwartungswert des Produktes berechnen. Dieser sei definiert durch

$$\mathbb{E}(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dydx.$$

SATZ 6.4. Falls  $X$  und  $Y$  unabhängig sind, gilt:

$$\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y.$$

BEWEIS. Der Beweis wird nur für stetige Zufallsvariable betrachtet. Nach (31) gilt:

$$\begin{aligned} \mathbb{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dydx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dydx \\ &= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy = \mathbb{E}X \cdot \mathbb{E}Y. \end{aligned}$$

□

Im Satz 4.25 wurde festgestellt, dass die Summe der Varianzen von unabhängigen Zufallsvariablen gleich der Varianz der Summe ist. Der Beweis kann nun durch Anwendung des Satzes 6.4 erbracht werden.

$$\begin{aligned} D^2(X + Y) &= \mathbb{E}(X + Y)^2 - (\mathbb{E}(X + Y))^2 \\ &= \mathbb{E}X^2 + 2\mathbb{E}X\mathbb{E}Y + \mathbb{E}Y^2 - (\mathbb{E}X)^2 - 2\mathbb{E}X \cdot \mathbb{E}Y - (\mathbb{E}Y)^2 \\ &= \mathbb{E}X^2 - (\mathbb{E}X)^2 + \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = D^2X + D^2Y. \end{aligned}$$

### 3. Maße für die Abhängigkeit von zwei Zufallsvariablen.

Es ist nicht schwer sich vorzustellen, dass zwei Zufallsvariable völlig unabhängig sein können. Das ist der Fall, falls  $X$  und  $Y$  die gewürfelte Augenzahl zweier unterschiedlicher Würfel bezeichnen. Andererseits können zwei Zufallsvariable auch völlig abhängig sei. Man berechne zum Beispiel  $Y$  als eine lineare Transformation von  $X$ :

$$Y = aX + b.$$

Es gibt weiterhin Situationen, in denen eine gewisse Abhängigkeit vorhanden ist. Diese ist dann allerdings nicht so stark ausgeprägt, wie die soeben geschilderte funktionale Abhängigkeit. Als Beispiel wäre der Zusammenhang zwischen Alter und Blutdruck zu nennen. Wir wollen zwei Maße kennenlernen, die die Tendenz beziehungsweise Stärke der Abhängigkeit beschreiben.

DEFINITION 6.5. Es seien  $X$  und  $Y$  zwei Zufallsvariable mit endlichem Erwartungswert. Dann ist die Kovarianz definiert durch

$$\text{cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y).$$

Speziell gilt, falls  $X = Y$  vorausgesetzt wird:

$$\text{cov}(X, X) = D^2X.$$

Die Kovarianz beschreibt das tendenzielle Verhalten der Paare von Versuchsausgängen. Die Kovarianz  $\text{cov}(X, Y)$  ist größer Null, falls bei Vergrößerung der Versuchsausgänge der Zufallsvariablen  $X$ , auch die Versuchsausgänge des zu  $Y$  gehörigen Versuches *tendenziell* größer werden. In diesem Fall sind nämlich die meisten Produkte  $(x - \mathbb{E}X)(y - \mathbb{E}Y)$  positiv siehe Abbildung 2.

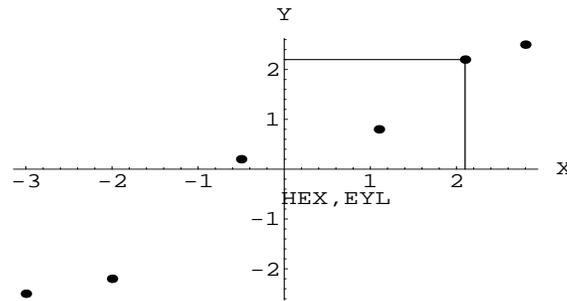


ABBILDUNG 2. positive Kovarianz: die meisten Produkte  $(x - EX)(y - EY)$  sind positiv.

Umgekehrt, die Kovarianz  $\text{cov}(X, Y)$  ist kleiner Null, falls bei Vergrößerung der Versuchsausgänge der Zufallsvariable  $X$  dann auch die Versuchsausgänge des zu  $Y$  gehörigen Versuches *tendenziell* kleiner werden.

Falls  $X$  und  $Y$  unabhängig sind, ist die Kovarianz Null. Die Umkehrung gilt im allgemeinen nicht. Allerdings gibt es Fälle, wo die Umkehrung richtig ist. Besitzt zum Beispiel das Paar  $(X, Y)$  eine zweidimensionale Normalverteilung, so folgt die Unabhängigkeit von  $X$  und  $Y$  aus dem Verschwinden der Kovarianz.

In der Definition der Dichte der zweidimensionalen Normalverteilung (29) ist das Element  $\sigma_{XY}$  in der Matrix  $\Sigma$  die Kovarianz von  $X$  und  $Y$ . Die Niveaulinien der Dichtefunktion sind Ellipsen, siehe Abbildung 1. Falls nun  $\sigma_{XY} \neq 0$ , dann sind die Hauptachsen dieser Ellipsen nicht parallel zu den Koordinatenachsen. Darin spiegelt sich der Fakt wider, dass eine tendenzielle Abhängigkeit zwischen  $X$  und  $Y$  vorliegt. Ist  $\sigma_{XY} = 0$ , dann liegen die Halbachsen parallel zu den Koordinatenachsen. Speziell sind die Zufallsvariablen  $X$  und  $Y$  unabhängig.

Der Nachteil der Kovarianz ist, dass nur etwas über die Richtung der Abhängigkeit, nichts aber über die Stärke der Abhängigkeit ausgesagt werden kann. Speziell führen auch (lineare) Maßstabsänderungen zu einer Änderung der Kovarianz.

Dieser Nachteil wird beseitigt durch die Einführung des Korrelationskoeffizienten:

DEFINITION 6.6. *Es seien  $X$  und  $Y$  zwei Zufallsvariable mit endlichem Erwartungswert. Dann ist der Korrelationskoeffizient definiert durch:*

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D^2 X \cdot D^2 Y}}.$$

Der Korrelationskoeffizient ist immer im Intervall  $[-1, 1]$  enthalten. Ist speziell  $\rho(X, Y) = -1$ , dann sind  $X$  und  $Y$  durch eine lineare funktionale Beziehung gekoppelt:

$$Y = aX + b,$$

wobei der Anstieg  $a$  negativ ist.  $X$  und  $Y$  sind vollständig abhängig, falls  $\rho(X, Y) = +1$  wobei  $a > 0$  gilt. Für  $\rho(X, Y) \in (-1, 1)$  liegt eine schwächere Abhängigkeit vor, wobei das Vorzeichen die Tendenz der Abhängigkeit beschreibt, siehe Abbildung 3.

Falls  $X$  und  $Y$  unabhängig sind, so ist der Korrelationskoeffizient Null. Die Umkehrung gilt allerdings nicht. Speziell kann es passieren, falls  $X$  und  $Y$  durch einen

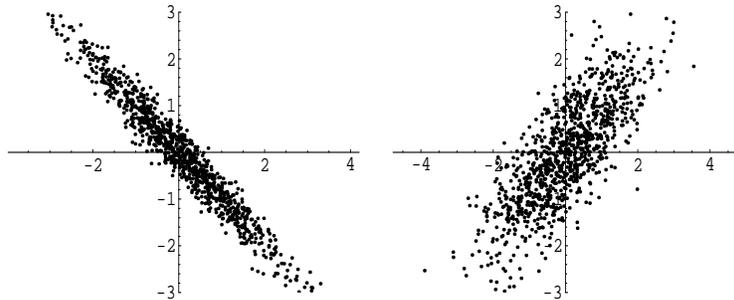


ABBILDUNG 3. Punktwolken mit Korrelationskoeffizient  $\rho(X, Y) = -0.95$  und  $\rho(X, Y) = 0.8$

*nichtlinearen* Zusammenhang gekoppelt sind, dass der Korrelationskoeffizient Null ist.

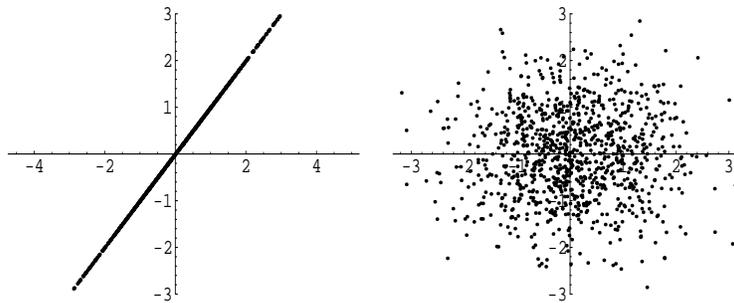


ABBILDUNG 4. Vollständige lineare Abhängigkeit ( $\rho(X, Y) = 1$ ) und Unabhängigkeit ( $\rho(X, Y) = 0$ ) der normalverteilten Zufallsvariablen  $(X, Y)$



## Markov-Ketten

### 1. Grundlegende Definitionen

In diesem Kapitel sollen Systeme untersucht werden, die durch den zufälligen Übergang zwischen verschiedenen Zuständen charakterisiert sind. Mathematische Modelle, die solche Zustände beschreiben, bezeichnet man als *zufällige Prozesse*. Das Folgende basiert auf einer Darstellung von Osaki [17].

DEFINITION 7.1. *Gegeben sei eine Menge  $\mathbb{T}$ , zum Beispiel  $\mathbb{T} = \mathbb{Z}^+$  oder  $\mathbb{T} = \mathbb{R}^+$ . (Diese Menge werde als Zeitmenge interpretiert.) Eine Familie von Zufallsvariablen  $X = (X(t))_{t \in \mathbb{T}}$  heißt zufälliger (stochastischer) Prozess.*

Zufällige Prozesse charakterisieren also die zeitliche Entwicklung von zufälligen Erscheinungen.

Ein Beispiel ist der sogenannte *Wiener Prozess*, der die Brownsche Bewegung, also die zufällige Bewegung eines Moleküls in einer Flüssigkeit, beschreibt. Die zickzackartige Bahnkurve ergibt sich durch die zufälligen Kollisionen mit anderen Molekülen in der Flüssigkeit. Siehe Abbildung 1.

Uns werden hier nur Prozesse interessieren, die eine gewisse Bedeutung für die Informatik besitzen, die Markov-Ketten.

DEFINITION 7.2. *Ein zufälliger Prozess mit Zeitmenge  $\mathbb{T} = \mathbb{Z}^+$   $X = (X(t))_{t \in \mathbb{Z}^+}$ , wobei der Wertevorrat eine endliche oder abzählbar unendliche Menge ist, (wir bezeichnen diesen Wertevorrat mit  $\{0, 1, \dots, N\}$  oder  $\mathbb{Z}^+$ ) heißt Markov-Kette, falls*

$$(32) \quad \begin{aligned} \mathbb{P}(X(n+1) = j | X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n) \\ = \mathbb{P}(X(n+1) = j | X(n) = i_n) =: p_{ij}, \quad i_k \in Wv(X), \end{aligned}$$

*falls  $\mathbb{P}(X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n) \neq 0$  ist.  $p_{ij}$  wird als Übergangswahrscheinlichkeit vom Zustand  $i$  zum Zustand  $j$  bezeichnet.*

BEMERKUNG 7.3. *Markov-Ketten oder allgemeiner Markov-Prozesse werden auch als Prozesse ohne Gedächtnis bezeichnet. Speziell hängt die Wahrscheinlichkeit des Eintretens des Zustandes  $j$  zur Zeit  $n+1$  nur vom Zustand  $i_n$  zur Zeit  $n$  ab. Der Einfluss der weiter zurückliegenden Zustände  $X(n-1) = i_{n-1}, \dots$  spielt für die Wahrscheinlichkeit des Eintretens von  $j$  zur Zeit  $n+1$  keine Rolle. Diese Zustände werden sozusagen vergessen.*

*Es gilt:*

$$p_{ij} \geq 0, \quad \sum_{j=0}^{N/\infty} p_{ij} = 1.$$

Es ist also sicher, von  $i$  zu irgendeinem Zustand zu gelangen. Die Übergangswahrscheinlichkeiten können zu einer (unter Umständen unendlichen) Matrix

$$P = (p_{ij})$$

zusammengefasst werden. Die obige Gleichung zeigt, dass die Summe der Elemente in jeder Zeile Eins ist. Eine Matrix mit dieser Eigenschaft und mit  $p_{ij} \geq 0$  wird als stochastische Matrix bezeichnet.

**Beispiel:** Es seien  $a, b \in [0, 1]$ . Dann beschreibt die Übergangsmatrix

$$\begin{matrix} 0 & \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}. \end{matrix}$$

die durch das folgende Diagramm beschriebenen zufälligen Übergänge.

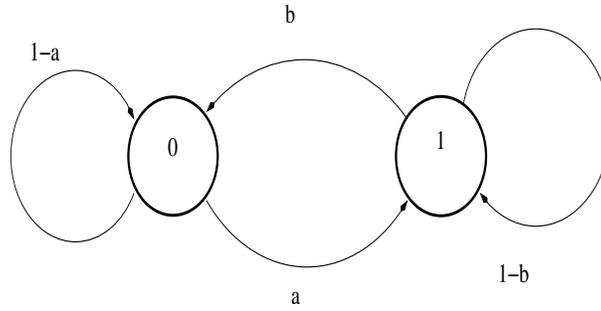


ABBILDUNG 1. Übergangsdiagramm

Wir berechnen die Wahrscheinlichkeit für einen Pfad durch die Zustände des Systems. Speziell gilt nach der Definition der bedingten Wahrscheinlichkeit:

$$\begin{aligned} & \mathbb{P}(X(0) = i_0, X(1) = i_1, \dots, X(n) = i_n) \\ & \mathbb{P}(X(n) = i_n | X(0) = i_0, \dots, X(n-1) = i_{n-1}) \times \\ & \quad \times P(X(0) = i_0, X(1) = i_1, \dots, X(n-1) = i_{n-1}) \\ & = p_{i_{n-1}i_n} \mathbb{P}(X(0) = i_0, \dots, X(n-1) = i_{n-1}) \\ & = \dots \\ & = p_{i_{n-1}i_n} p_{i_{n-2}i_{n-1}} \dots p_{i_0i_1} \mathbb{P}(X(0) = i_0). \end{aligned}$$

Mit  $\pi_{i_0}(0)$  bezeichnen wir die Wahrscheinlichkeit, dass zur Zeit  $t = 0$  der Zustand  $i_0$  angenommen wird:

$$\pi_{i_0}(0) := \mathbb{P}(X(0) = i_0).$$

Falls wir davon ausgehen, dass der Anfangszustand  $i_0$  sicher angenommen wird, so setzen wir die Wahrscheinlichkeit  $\pi_i(0) = 1$ . Speziell ist die Verteilung des zufälligen Anfangszustandes gegeben durch

$$\pi(0) = (\pi_0(0), \pi_1(0), \dots), \quad \sum_{i=0}^{N/\infty} \pi_i(0) = 1.$$

Wir führen die  $n$ -Schritt Übergangswahrscheinlichkeiten ein:

$$p_{ij}^n := \mathbb{P}(X(n+m) = j | X(m) = i)$$

für  $n \in \mathbb{N}$ . Speziell definieren wir für  $n = 0$ :

$$\begin{aligned} 0 & : i \neq j \\ 1 & : i = j \end{aligned}$$

Diese  $n$ -Schritt Übergangswahrscheinlichkeiten können zu der Matrix

$$P^{(n)} = (p_{ij}^n)$$

zusammengefasst werden.

Der folgende Satz ist für das Verständnis der Markov-Ketten von großer Bedeutung:

SATZ 7.4. *Es seien  $r, n$  ganze Zahlen mit  $0 \leq r \leq n$ . Dann gilt:*

$$p_{ij}^n = \sum_{k=0}^{N/\infty} p_{ik}^r p_{kj}^{n-r}.$$

Den Beweis des Satzes kann man sich mittels der Abbildung 2 klarmachen. Die Wahrscheinlichkeit in  $n$

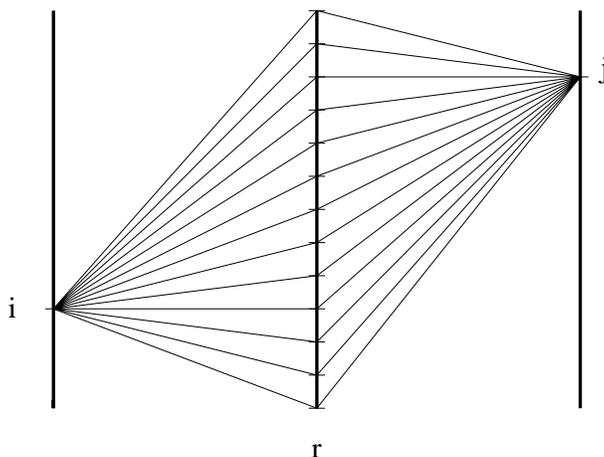


ABBILDUNG 2. Chapman-Kolmogorov-Gleichung

Zeitschritten von  $i$  nach  $j$  überzugehen, ist gleich der Summe der Wahrscheinlichkeiten von  $i$  nach  $j$  in  $n$  Schritten überzugehen und dabei zur Zeit  $r$  irgendeinen der Zustände zu durchlaufen.

Diese Gleichung heißt *Chapman-Kolmogorov-Gleichung*. Sie kann in Matrixform wie folgt geschrieben werden:

$$P^{(n)} = P^{(r)} P^{(n-r)}.$$

Speziell gilt:

$$P^{(n)} = PP^{(n-1)} = PPP^{(n-2)} = \dots = P^n,$$

wobei die rechte Seite die  $n$ -te Potenz der Matrix  $P$  ist.

Es bezeichne

$$\pi(n) = (\pi_0(n), \pi_1(n), \dots), \quad \pi_j(n) = \mathbb{P}(X(n) = j)$$

die Verteilung des Systems zur Zeit  $n$ . Dann gilt:

$$\begin{aligned} \pi_j(n) = \mathbb{P}(X(n) = j) &= \sum_{i=0}^{N/\infty} \mathbb{P}(X(n) = j | X(0) = i) \mathbb{P}(X(0) = i) \\ &= \sum_{i=0}^{N/\infty} \pi_i(0) p_{ij}^n. \end{aligned}$$

Das bedeutet, man kann die Wahrscheinlichkeitsverteilung zur Zeit  $n$  durch eine Matrixmultiplikation der  $n$ -Schritt Übergangsmatrix und der Anfangsverteilung  $\pi(0)$  berechnen:

$$\pi(n) = \pi(0)P^{(n)}.$$

**Beispiel:** Wir betrachten die Übergangsmatrix

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}.$$

Dann erhält man durch das Multiplizieren der Matrix  $P$  mit sich selbst

$$\begin{aligned} P^{(2)} = P^2 &= \begin{pmatrix} (1-a)^2 + ab & (1-a)a + (1-b)a \\ (1-a)b + (1-b)b & ab + (1-b)^2 \end{pmatrix} \\ P^{(n)} = P^n &= \frac{1}{a+b} \begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^n}{a+b} \begin{pmatrix} a & -a \\ -b & b \end{pmatrix}. \end{aligned}$$

Speziell gilt:

$$\lim_{n \rightarrow \infty} P^{(n)} = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix}.$$

## 2. Klassifikation der Zustände einer Markov-Kette

Im Folgenden wollen wir sehen, dass die Werte, die eine Markov-Kette annehmen kann, sich in zwei verschiedene Klassen einteilen lassen. Die eine Klasse besteht aus Zuständen, die unendlich oft mit Wahrscheinlichkeit 1 erreicht werden, während die Zustände der anderen Klasse nur endlich oft durch die Markov-Kette  $X$  erreicht werden.

Wir sagen, der Zustand  $j$  ist vom Zustand  $i$  aus erreichbar (schreibweise  $i \rightarrow j$ ), falls es ein  $n \in \mathbb{Z}^+$  gibt mit  $p_{ij}^n > 0$ . Dies bedeutet, dass wir von  $i$  nach  $j$  nach  $n$  Zeitschritten mit positiver Wahrscheinlichkeit gelangen.

Wir sagen, die Zustände  $i, j$  kommunizieren miteinander ( $i \leftrightarrow j$ ), falls es natürliche Zahlen  $m, n$  gibt mit  $p_{ij}^m > 0, p_{ji}^n > 0$ . Wir gelangen also von  $i$  nach  $j$ , aber auch von  $j$  nach  $i$  in  $m$  beziehungsweise  $n$  Zeitschritten mit positiver Wahrscheinlichkeit.

SATZ 7.5.  $\leftrightarrow$  bildet eine Äquivalenzrelation, das heißt es gilt:

- (i)  $i \leftrightarrow i$
- (ii) falls  $i \leftrightarrow j$  dann  $j \leftrightarrow i$
- (iii) falls  $i \leftrightarrow j$  und  $j \leftrightarrow k$  dann  $i \leftrightarrow k$ .

BEWEIS. (i) und (ii) folgt direkt aus der Definition von  $\leftrightarrow$ . Um (iii) einzusehen, können wir nach der direkten Definition von  $\leftrightarrow$  annehmen, dass es  $m, n$  gibt mit

$$p_{ij}^m > 0, \quad p_{jk}^n > 0.$$

Dann gilt nach der Chapman-Kolmogorov-Gleichung

$$p_{ik}^{m+n} = \sum_{l=0}^{N/\infty} p_{il}^m p_{lk}^n \geq p_{ij}^m p_{jk}^n > 0,$$

also  $i \rightarrow k$ . Die umgekehrte Beziehung erfolgt analog.  $\square$

Beispiel: Wir betrachten eine Markov-Kette mit Übergangswahrscheinlichkeit

$$\begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

zwischen den Zuständen 0, 1, 2. Dann gilt:

$$1 \leftrightarrow 2, \quad 2 \rightarrow 0, \quad 0 \not\leftrightarrow 2.$$

Es gibt also genau zwei Äquivalenzklassen  $\{0\}$  und  $\{1, 2\}$ .

DEFINITION 7.6. Eine Markov-Kette  $X$  heißt irreduzibel, falls  $i \leftrightarrow j$  für alle Zustände  $i, j$  der Markov-Kette gilt.

Es sei  $d(i)$  der größte gemeinsame Teiler aller  $n \geq 1$  mit  $p_{ii}^n > 0$ . Der Zustand  $i$  heißt aperiodisch, falls  $d(i) = 1$  gilt.  $i$  heißt periodisch, falls  $d(i) > 1 \in \mathbb{N}$ .

Die Markov-Kette aus dem letzten Beispiel ist *nicht* irreduzibel, denn es ist nur mit Wahrscheinlichkeit Null möglich, von 0 nach 2 zu gelangen.

Die Periodizität einer Markov-Kette bringt eine gewisse Regelmäßigkeit bezüglich  $n$  zum Ausdruck in der Abfolge der Übergangswahrscheinlichkeiten mit  $p_{ii}^n > 0$ .

**Beispiel:** Wir betrachten die Übergangsmatrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

dann gilt:

$$0 \rightarrow 1 \rightarrow 2 \rightarrow 0,$$

also ist diese Markov-Kette irreduzibel. Weiterhin erhalten wir

$$p_{00}^3 = p_{00}^6 = 1$$

und für  $n$ , die nicht durch drei teilbar sind  $p_{00}^n = 0$ . Damit ist der Zustand 0 periodisch mit  $d(0) = 3$ .

Es sei  $f_{ij}^n$  die Wahrscheinlichkeit für den Übergang von  $i$  nach  $j$  in genau  $n$  Schritten:

$$f_{ij}^n = \mathbb{P}(X(n) = j, X(r) \neq j, r = 1, 2, \dots, n-1 | X(0) = i).$$

Dabei gilt für  $i \neq j$  die Gleichung  $f_{ij}^0 = 0$  und  $f_{ij}^1 = p_{ij}$ . Die Wahrscheinlichkeit, dass  $i$  irgendwann nach  $j$  übergeht ist gerade

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n.$$

**DEFINITION 7.7.** *Ein Zustand  $i$  heißt rekurrent, falls  $f_{ii} = 1$ ;  $i$  heißt transient, falls  $f_{ii} < 1$ .*

Rekurrente Zustände zeichnen sich also dadurch aus, dass die Markov-Kette zu diesem Zustand mit Wahrscheinlichkeit 1 zurückkehrt. Bei transienten Zuständen ist dies nicht der Fall.

Ohne Beweis führen wir den folgenden Satz an, der ein tieferes Verständnis von Rekurrenz und Transienz erlaubt.

**SATZ 7.8.** *Ein Zustand  $i$  ist genau dann rekurrent, falls*

$$\sum_{n=1}^{\infty} p_{ii}^n = \infty.$$

*Ein Zustand  $i$  ist genau dann transient, wenn*

$$\sum_{n=1}^{\infty} p_{ii}^n < \infty.$$

Aufgrund der Konvergenz der zweiten Reihe folgt, dass für  $k \rightarrow \infty$  die Partialsummen  $\sum_{n=k}^{\infty} p_{ii}^n < \infty$  der  $n$ -Schrittübergangswahrscheinlichkeiten gegen Null konvergieren. Transiente Zustände werden also durch die Markov-Kette nur endlich oft mit Wahrscheinlichkeit 1 aufgesucht. Genauso folgt, dass eine Markov-Kette einen rekurrenten Zustand unendlich oft aufsucht mit Wahrscheinlichkeit 1 eintritt. Das kann wie folgt motiviert werden. Es sei  $T$  die erste Zeit, dass der Zustand  $i$  wieder erreicht wird. Dann ist aufgrund der Rekurrenz diese Zeit endlich mit Wahrscheinlichkeit 1. *Startet* man die Markov-Kette wieder in dieser zufälligen Zeit, dann wird der Zustand  $i$  wieder erreicht, und zwar mit Wahrscheinlichkeit 1, und so weiter.

Eine Folgerung aus dem letzten Satz ist, dass ein Zustand, der nicht rekurrent ist, transient ist und umgekehrt.

**Beispiel:** Wir betrachten die Markov-Kette  $X$  mit der Übergangswahrscheinlichkeit

$$\begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Durch Multiplikation dieser Matrix  $n - 1$  mal mit sich selbst erhält man

$$P^{(n)} = \begin{pmatrix} 1 & 0 \\ \frac{2^n - 1}{2^n} & \frac{1}{2^n} \end{pmatrix}.$$

Der Zustand 0 ist rekurrent, da  $\sum_{n=1}^{\infty} p_{00}^n = \infty$  und 1 ist transient wegen

$$\sum_{n=1}^{\infty} p_{11}^n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1 < \infty.$$

**SATZ 7.9.** *Angenommen der Zustand  $i$  ist rekurrent. Weiterhin kommuniziert  $i$  mit  $j$ . Dann ist  $j$  auch rekurrent.*

BEWEIS. Da  $i, j$  kommunizieren, gibt es  $m, n$  mit  $p_{ij}^m > 0, p_{ji}^n > 0$ . Weiterhin gilt die Ungleichung

$$p_{jj}^{m+s+n} > p_{ji}^m p_{ii}^s p_{ij}^n.$$

Die linke Seite beschreibt die Wahrscheinlichkeit von  $j$  nach  $j$  in  $m+s+n$  Zeitschritten überzugehen, während die rechte Seite die Wahrscheinlichkeit des speziellen Ereignisses beschreibt, bei so einem Übergang zuerst von  $j$  nach  $i$  in vorgegebenen  $m$  Schritten, dann in  $s$  Schritten von  $i$  nach  $i$  zu gehen und dann erst wieder  $j$  in genau  $n$  Schritten aufzusuchen. Damit ergibt sich

$$\sum_{s=1}^{\infty} p_{jj}^{m+s+n} \geq p_{ji}^m p_{ij}^n \sum_{s=1}^{\infty} p_{ii}^s = \infty,$$

□

da  $i$  als rekurrent vorausgesetzt ist.

BEMERKUNG 7.10. *Eine irreduzible Markov-Kette besteht nur aus transienten oder rekurrenten Zuständen.*

**Beispiel:**(Random walk) Es sei  $X$  eine Markov-Kette mit Zustandsmenge  $\mathbb{Z}$ . Die Übergänge erfolgen von  $i$  nach  $i+1$  mit Wahrscheinlichkeit  $p \in (0, 1)$  und von  $i$  nach  $i-1$  mit Wahrscheinlichkeit  $1-p=:q$ . Wir berechnen die Wahrscheinlichkeit in  $n$  Schritten von einem Zustand, sagen wir 0 zu diesem Zustand zurückzukommen. Da dazu eine gerade Anzahl von Übergängen notwendig ist, gilt  $p_{00}^{2n+1} = 0$ . Durch Anwendung der Binomialverteilung erhalten wir,

$$\begin{aligned} p_{00}^{2n} &= \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} (pq)^n \\ &\approx \frac{\sqrt{2\pi 2n} (2n)^{2n} e^{-2n}}{2\pi n (n^n)^2 e^{-n} e^{-n}} p^n q^n = \frac{(4pq)^n}{\sqrt{\pi n}}, \end{aligned}$$

wobei für die letzte Zeile die Stirlingsche Formel benutzt wurde. Damit gilt

$$\sum_{n=1}^{\infty} p_{00}^n \approx \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{n\pi}} = \begin{cases} < \infty & : p \neq q \\ \infty & : p = q \end{cases},$$

da für  $p \neq q$  die Beziehung  $4pq < 1$  und damit die geometrische Reihe eine konvergente Majorante für die obige Reihe darstellt. Für  $p = q$  gilt:  $4pq = 1$ . Die dabei vorgenommene Approximation der Koeffizienten besitzt keinen Einfluss auf Konvergenz/Divergenz der Reihe. Also ist die harmonische Reihe für diese Parameter eine divergente Minorante.

Damit ist die Markov-Kette für  $p = q$  rekurrent, und ansonsten transient.

### 3. Stationäres Verhalten der Markov-Kette

In diesem Abschnitt sind wir an einem gewissen typischen Verhalten interessiert, das sich nach einer gewissen Zeit einstellt. Speziell bedeutet typisches Verhalten, dass sich die Wahrscheinlichkeiten der einzelnen Zustände beim Übergang von der Zeit  $n$  zur Zeit  $n+1$  für  $n \in \mathbb{Z}^+$  nicht mehr ändern. Man kann also von einem zufälligen Gleichgewicht sprechen. Da man mittels der Übergangsmatrix  $P$  den Übergang der Wahrscheinlichkeitsverteilung der Zustände von der Zeit  $n$  zu einer späteren Zeit beschreiben kann, gilt:

$$(33) \quad \pi^* = \pi^* P = \pi^* P^{(2)} = \dots = \pi^* P^{(n)},$$

wobei wir mit  $\pi^*$  die Verteilung zu diesem zufälligen Gleichgewicht bezeichnen. Dieses wird auch *stationäre* Verteilung der Markov-Kette  $X$  genannt.

Es sei noch einmal darauf hingewiesen, dass sich bei einem zufälligen Gleichgewicht die Zustände des Systems ändern (was bei einem üblichen Gleichgewicht nicht der Fall wäre), während die Wahrscheinlichkeitsverteilung dieser Zustände konstant bleibt. Wir wollen nun die Frage beantworten, wie diese stationäre Verteilung strukturiert ist.

Dazu sei  $T_i$  der erste zufällige Zeitpunkt, zu dem, falls in  $i$  gestartet wird, der Zustand  $i$  wieder angenommen wird. Dabei gilt  $T_i < \infty$ , falls  $i$  rekurrent ist. Der Erwartungswert von  $T_i$  errechnet sich für rekurrentes  $i$  durch

$$\mathbb{E}T_i = \sum_{n=1}^{\infty} f_{ii}^n n =: \mu_i.$$

Dabei ist es möglich, dass  $\mu_i = \infty$  gilt.

Wir unterteilen also die rekurrenten Zustände  $i$  noch einmal in *null-rekurrente Zustände*, falls  $\mu_i = \infty$  und in *positiv-rekurrente Zustände*, falls  $\mu_i < \infty$ .

Mittels der positiv-rekurrenten Zustände kann nun die stationäre Verteilung der Markov-Kette  $X$  beschrieben werden.

**SATZ 7.11.** *Gegeben sei eine irreduzible Markov-Kette. Dann sind folgende Aussagen äquivalent.*

- (i) *Es existiert eine stationäre Verteilung  $\pi^*$ .*
- (ii) *Alle Zustände sind positiv rekurrent.*

Weiterhin gilt, dass  $\pi^*$  die einzige stationäre Verteilung ist, die durch

$$\pi_i^* = \frac{1}{\mu_i}$$

gegeben ist. Falls alle diese Zustände aperiodisch sind, dann gilt:

$$\lim_{n \rightarrow \infty} p_{ij}^n = \pi_j^*, \quad \lim_{n \rightarrow \infty} |\mathbb{P}(X(n) = j) - \pi_j^*| = 0.$$

**Beispiel:** Gegeben sei eine Markov-Kette mit Übergangsmatrix

$$\begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

Zur Bestimmung von  $\pi^*$  kann (33) herangezogen werden:  $\pi^* = \pi^* P$ . Damit können die Komponenten des Vektors  $\pi^*$  durch die Lösung des linearen Gleichungssystems bestimmt werden:

$$\begin{aligned} \pi_0^* &= 0.8\pi_0^* + 0.3\pi_1^* & \Leftrightarrow & -0.2\pi_0^* + 0.3\pi_1^* = 0 \\ \pi_1^* &= 0.2\pi_0^* + 0.7\pi_1^* & \Leftrightarrow & +0.2\pi_0^* - 0.3\pi_1^* = 0 \end{aligned}.$$

Die Koeffizientenmatrix dieses linearen homogenen Gleichungssystems hat den Rang 1 und somit unendlich viele Lösungen. Da die Lösung eine Wahrscheinlichkeitsverteilung sein soll, muss zusätzlich  $\pi_0^* + \pi_1^* = 1$  gelten. Damit ergibt sich als stationäre Verteilung von  $X$ :

$$\pi_0^* = 0.6, \quad \pi_1^* = 0.4.$$

Die stationäre Verteilung kann mit den Werkzeugen der linearen Algebra (Eigenvektoren) ausgerechnet werden. Damit können dann die mittleren Rekurrenzzeiten  $\mu_i$  bestimmt werden.

#### 4. Markov-Ketten mit endlich vielen Zuständen

Wir nehmen an, dass die Menge der Werte, die eine Markov-Kette annehmen kann, gegeben ist durch  $\{0, 1, \dots, N\}$ . Für Markov-Ketten mit endlich vielen Zuständen gilt der folgende Satz, den wir ohne Beweis formulieren.

**SATZ 7.12.** *Gegeben sei eine Markov-Kette  $X$  mit endlich vielen Zuständen. Dann gilt:*

- (i) *Es gibt keine Null-rekurrenten Zustände.*
- (ii) *Nicht alle Zustände sind transient.*

Gegeben sei eine Markov-Kette  $X$  mit endlich vielen Zuständen. Dann können diese in transiente und positiv-rekurrente Zustände unterteilt werden. Diese wiederum können in untereinander kommunizierende Zustände eingeteilt werden. Wir bezeichnen die Äquivalenzklassen (siehe Satz 7.5) mit  $C_1, \dots, C_l$ . Damit besitzt die Übergangsmatrix die folgende Struktur:

$$P = \begin{pmatrix} P_1 & 0 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & P_l & 0 \\ R_1 & R_2 & \dots & R_l & Q \end{pmatrix} \begin{matrix} C_1 \\ C_2 \\ \\ C_l \\ T \end{matrix}.$$

$T$  bezeichne die transienten Zustände. Die Matrix  $P_r$  ist die Übergangsmatrix für eine Äquivalenzklasse, also von Zuständen, die untereinander kommunizieren. In der letzten Zeile bezeichnen die Matrizen  $R_r$  die Wahrscheinlichkeit des Überganges der transienten Zustände in die entsprechende Klasse von rekurrenten Zuständen. In diese Klassen kommt man zwar hinein, aber nicht wieder heraus. Die Matrix  $Q$  beschreibt die Wahrscheinlichkeit des Verbleibens der transienten Zustände in diesen bei einem Zeitschritt. Wir betrachten die folgende Übergangsmatrix

$$\begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \end{pmatrix}.$$

Nach Umordnung der Zeilen entsprechend der Äquivalenzklasse gilt:

$$\begin{matrix} 3 \\ 5 \\ 2 \\ 4 \\ 0 \\ 6 \\ 1 \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix}.$$

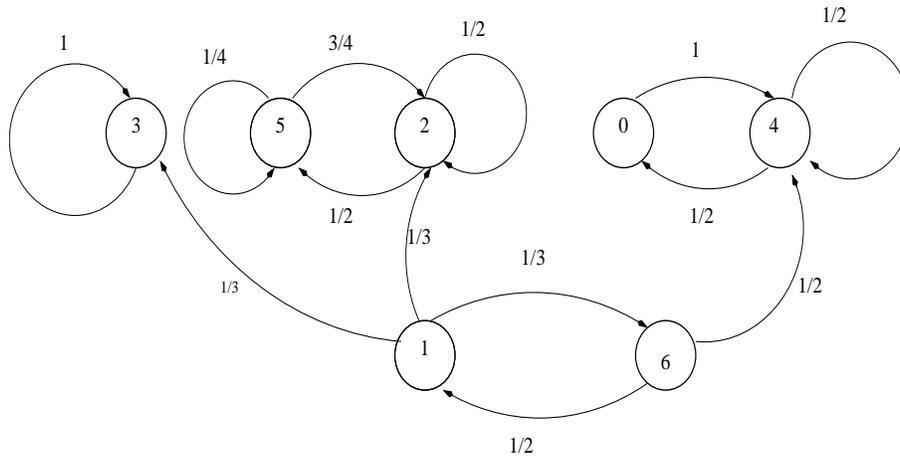


ABBILDUNG 3. Übergangsdiagramm

Auf der Hauptdiagonale findet man die Übergangsmatrizen  $P_1$  (nur ein Element),  $P_2$ ,  $P_3$  zu den Äquivalenzklassen  $C_1$ ,  $C_2$ ,  $C_3$ . Die letzten beiden Zeilen beschreiben die transienten Zustände und deren Übergänge.

Mittels dieser Darstellung können nun die Wahrscheinlichkeiten des Überganges eines transienten Zustandes von  $i$  nach  $j \in C_k$  ausgerechnet werden. Diese Wahrscheinlichkeiten wurden mit  $f_{ij}$  bezeichnet.

**SATZ 7.13.** Für die Wahrscheinlichkeit des Überganges von  $i \in T$  nach  $C_k$  gilt:

$$f_{ij} = \sum_{l \in C_k} p_{il} + \sum_{l \in T} p_{il} f_{lj}$$

Oder in Matrixschreibweise:

$$(f_{ij}) = R_k \mathbf{1} + Q(f_{ij}), \quad (f_{ij}) = (E - Q)^{-1} R_k \cdot \mathbf{1}.$$

$\mathbf{1}$  ist dabei eine Spaltenmatrix, die nur aus Einsen besteht. Dabei beschreibt  $R_k \mathbf{1}$  den Term  $\sum_{l \in C_k} p_{il}$ .

Der Beweis ist nicht schwer. Der erste Term auf der rechten Seite der obigen Formel beschreibt die Wahrscheinlichkeit des direkten Überganges von  $i$  nach  $j \in C_k$  oder, was gleich dazu ist, von  $i$  nach  $l \in C_k$ . Der zweite Term beschreibt die Wahrscheinlichkeit des Überganges von  $T$  nach  $C_k$  mit Zwischenstation in  $T$ .

**Beispiel:** Wir betrachten die oben beschriebene Markov-Kette. Dann gilt:

$$E - Q = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{3} & 1 \end{pmatrix}, \quad (E - Q)^{-1} = \begin{pmatrix} \frac{6}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{3}{5} \end{pmatrix}.$$

Damit ergibt sich für den Übergang nach  $C_2$

$$(E - Q)^{-1} R_2 \cdot \mathbf{1} = \begin{pmatrix} \frac{6}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{3}{5} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} \\ \frac{3}{5} \end{pmatrix}.$$

Wir erhalten

$$f_{6j} = \frac{1}{5}, \quad f_{1j} = \frac{2}{5}, \quad j = 5, 2.$$

In der Tat ist diese Wahrscheinlichkeit unabhängig von den Übergängen in  $C_k$ , denn innerhalb einer Äquivalenzklasse erreicht ein Zustand jeden anderen mit Wahrscheinlichkeit 1 aufgrund der Rekurrenz.



## Schließende Statistik

### 1. Grundgesamtheit und Stichprobe

In diesem Abschnitt wollen wir das Grundmodell der *Schließenden Statistik* betrachten. Dabei geht man von einer großen Menge von Objekten aus, die ein gewisses Merkmal  $X$  tragen. Bezüglich der verschiedenen Objekte variieren die Werte des Merkmals. Zum Beispiel stelle man sich einen Automaten vor, der Mehl in Tüten verpackt. Das Merkmal dieser Objekte ist das Gewicht des abgepackten Mehls. Würde man nun von jeder einzelnen Tüte das Gewicht messen, so würde man feststellen, dass die Gewichte der Tüten variieren, das heißt unterschiedliche Tüten haben ein unterschiedliches Gewicht. So könnten zum Beispiel bei den ersten zehn gewogenen Tüten die folgenden Gewichte ermittelt werden:

501, 503, 497, 500, 501, 499, 502, 500, 498, 501[g].

Würde man zufällig eine Tüte aus der Gesamtproduktion entnehmen, so würde man einen zufälligen Wert für das Gewicht erhalten. Die verschiedenen Werte des Merkmals  $X$  interpretieren eine Zufallsvariable, die wir auch mit  $X$  bezeichnen. Diese Zufallsvariable wird auch als *Grundgesamtheit* bezeichnet. Eine Zufallsvariable wird charakterisiert durch ihre Verteilung beziehungsweise ihre Parameter, wie Erwartungswert oder Varianz. Der Erwartungswert beziehungsweise die Varianz der Gewichte sind wichtige Größen, die die Produktion der Mehltüten beschreiben. Es wäre aber zu kostspielig, jede Tüte zu wiegen und dann als arithmetisches Mittel den Mittelwert der Grundgesamtheit zu bestimmen.

Das Anliegen der Schließenden Statistik ist es nun, mittels einer kleinen Menge von Daten (*Stichprobe*), Aussagen über die Grundgesamtheit zu treffen. So sollen gewisse Parameter der Grundgesamtheit durch die Stichprobe *näherungsweise* ermittelt werden, oder wie man in der Statistik sagt, diese Parameter sollen mittels einer Stichprobe *geschätzt* werden. Desweiteren geht es darum, gewisse Annahmen über die Parameter oder die Verteilung der Grundgesamtheit zu prüfen, was man auch als *Testen von Hypothesen* über die Grundgesamtheit bezeichnet.

Ein anderes Beispiel ist die Wahlvorhersage, bei der aus Befragungen von relativ wenigen Wählern, vorausschauend Aussagen über den Ausgang einer Wahl gemacht werden.

Es ist klar, dass bei der Gewinnung von Informationen über die Grundgesamtheit mittels einer Stichprobe Fehler auftreten können, denn es steht ja nur eine relativ kleine Datenmenge zur Analyse der Grundgesamtheit zur Verfügung. Aufgabe der Schließenden Statistik ist es somit auch, gewisse Aussagen über die Wahrscheinlichkeit von auftretenden Fehlernaussagen oder fehlerhaften Abweichungen zu machen.

Grundlage der Schließenden Statistik ist die Wahrscheinlichkeitstheorie, die im ersten Teil behandelt wurde. Auf der Wahrscheinlichkeitsrechnung basiert auch das Grundmodell der Schließenden Statistik.

DEFINITION 8.1. Gegeben sei eine Zufallsvariable (Grundgesamtheit)  $X$ . Als mathematische Stichprobe vom Umfang  $n$  dieser Grundgesamtheit bezeichnet man eine Menge von  $n$  Zufallsvariablen

$$X_1, X_2, \dots, X_n$$

mit folgenden Eigenschaften:

- Die Zufallsvariablen  $X_i$  besitzen die gleiche Wahrscheinlichkeitsverteilung wie  $X$ .
- Die Zufallsvariablen  $X_i$  sind vollständig untereinander unabhängig.

Ermittelt man  $n$  konkrete Zahlenwerte  $x_1, x_2, \dots, x_n$  aus einer großen Menge der merkmalttragenden Objekte, so bezeichnet man diese  $n$  Zahlen als *konkrete Stichprobe*.

Wir werden nun nachprüfen, inwiefern es sinnvoll ist, Definition 8.1 als Modell für die Stichprobenentnahme zu wählen. Es ist klar, dass ein Element einer Stichprobe immer unabhängig von den anderen Elementen einer Stichprobe entnommen werden muss, um eine Unregelmäßigkeit beziehungsweise Gleichberechtigung der einzelnen Elemente zu garantieren. Diese Unabhängigkeit spiegelt sich in der in der Definition gemachten Voraussetzung der *mathematischen* Unabhängigkeit der Zufallsvariablen  $X_1, \dots, X_n$  wider.

Wir stellen uns vor, dass wir eine große Anzahl ( $k$  Stück) von konkreten Stichproben vom Umfang  $n$  generieren:

$$\begin{array}{c} x_1^1, x_2^1, x_3^1, \dots, x_n^1 \\ x_1^2, x_2^2, x_3^2, \dots, x_n^2 \\ \dots \\ \dots \\ x_1^k, x_2^k, x_3^k, \dots, x_n^k. \end{array}$$

Betrachten wir aus jeder dieser Stichproben das  $i$ -te Element,  $i \in \{1, \dots, n\}$

$$x_i^1, x_i^2, \dots, x_i^k$$

so erhält man  $k$  Elemente, die aus der Gesamtheit der merkmalttragenden Objekte gewonnen wurden. Sie reflektieren damit die Verteilung der Werte des Merkmals  $X$ , weswegen man annehmen kann, dass die Verteilung des mathematischen Stichprobenelementes  $X_i$  gleich der Verteilung der Grundgesamtheit  $X$  ist.

Die Bestimmung von Parametern der Grundgesamtheit erfolgt mittels *mathematischer Stichprobenfunktionen*, die auch *Schätzung* genannt werden.

DEFINITION 8.2. Gegeben sei eine mathematische Stichprobe  $X_1, \dots, X_n$  und eine Funktion<sup>1</sup>

$$T(x_1, \dots, x_n)$$

mit Wertevorrat in  $\mathbb{R}$ . Dann nennt man die Zufallsvariable  $T(X_1, \dots, X_n)$  eine *mathematische Stichprobenfunktion* oder *Schätzung*.

Setzt man als Argumente in die Funktion  $T$  die Zahlenwerte einer konkreten Stichprobe ein, so erhält man konkrete Schätzwerte für den Parameter einer Grundgesamtheit. Es ist natürlich nicht sinnvoll, jede beliebige Funktion  $T$  zur Definition

<sup>1</sup>mit speziellen technischen Eigenschaften, die hier nicht genannt werden sollen

einer mathematischen Stichprobenfunktion heranzuziehen. Weitere Kriterien, die so eine Funktion  $T$  noch erfüllen sollte, werden im Folgenden eingeführt.

**Beispiel:** Die bekannteste mathematische Stichprobenfunktion ist das arithmetische Mittel

$$(34) \quad \bar{X} = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i,$$

welches eine Schätzung des Erwartungswertes der Grundgesamtheit  $X$  repräsentiert. Desweiteren sind als Schätzungen für die Varianz zu erwähnen

$$\begin{aligned} \tilde{S}^2 &= T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\ S^2 &= T(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \end{aligned}$$

Eine Schätzung für die Wahrscheinlichkeit  $\mathbb{P}(A)$ , dass die Grundgesamtheit Werte in einer Menge  $A$  annimmt, ist die relative Häufigkeit:

$$H_n(A) = \frac{1}{n} \sum_{i=1}^n \chi_A(X_i),$$

wobei

$$\chi_A(x) = \begin{cases} 1 & : x \in A \\ 0 & : x \notin A \end{cases}.$$

Eine weitere Schätzung für den Mittelwert der Grundgesamtheit ist der *Stichprobenmedian*. Wir nehmen dazu an, dass  $X_1^*, \dots, X_n^*$  aus  $X_1, \dots, X_n$  gebildet wird, indem wir diese Werte der Größe nach ordnen. Der Stichprobenmedian ist definiert durch

$$X_{0.5} = \begin{cases} X_{\frac{n+1}{2}}^* & : \text{ falls } n \text{ ungerade} \\ \frac{1}{2}(X_{\frac{n}{2}}^* + X_{\frac{n}{2}+1}^*) & : \text{ falls } n \text{ gerade} \end{cases}.$$

(Man beachte den Unterschied zum Median einer Zufallsvariablen, siehe Definition 4.27.)

Wir hatten schon erwähnt, dass bei der Bestimmung von Parametern einer Grundgesamtheit mittels einer Stichprobe Abweichungen des Schätzwertes vom Parameter mit gewissen Wahrscheinlichkeiten auftreten. Diese Wahrscheinlichkeiten können mittels der Verteilungen der mathematischen Stichprobenfunktionen ermittelt beziehungsweise abgeschätzt werden.

Wir wollen uns im Folgenden mit der Interpretation von Schätzungen und deren Genauigkeit beschäftigen.

Wir werden den Erwartungswert einer Grundgesamtheit schätzen und entnehmen dazu eine Stichprobe vom Umfang  $n$ . Dann berechnen wir als Schätzwert das arithmetische Mittel:

$$x_1^1, x_2^1, x_3^1, \dots, x_n^1 \mapsto \bar{x}^1 = \frac{1}{n} \sum_{i=1}^n x_i^1.$$

Würden wir nun aus der Grundgesamtheit eine zweite Stichprobe entnehmen und daraus das arithmetische Mittel bilden, so würden wir erhalten:

$$x_1^2, x_2^2, x_3^2, \dots, x_n^2 \mapsto \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Dabei würden wir (meistens) feststellen, dass  $\bar{x}^1 \neq \bar{x}^2$  ist. Mindestens einer der beiden Werte kann nicht dem Erwartungswert von  $X$  entsprechen. Wir haben also einen Fehler gemacht. Inwiefern ist es trotzdem sinnvoll, das arithmetische Mittel als Schätzung für den Mittelwert von  $X$  zu bestimmen? Dazu betrachten wir wieder  $k$  verschiedene Stichprobenentnahmen. Wir berechnen für jede dieser Stichproben das arithmetische Mittel und bilden aus allen diesen arithmetischen Mitteln noch einmal das arithmetische Mittel. Falls  $k$  eine sehr große Zahl ist, beziehungsweise gegen  $\infty$  konvergiert, dann sollte diese Zahl gleich dem Erwartungswert der Grundgesamtheit sein:

$$\left. \begin{array}{l} x_1^1, x_2^1, x_3^1, \dots, x_n^1 \mapsto \bar{x}^1 \\ x_1^2, x_2^2, x_3^2, \dots, x_n^2 \mapsto \bar{x}^2 \\ \dots \\ \dots \\ x_1^k, x_2^k, x_3^k, \dots, x_n^k \mapsto \bar{x}^k \end{array} \right\} \mathbb{E}X \approx \frac{1}{k} \sum_{j=1}^k \bar{x}^j. \text{ } ^2$$

Das bedeutet, dass man nur weiß, dass im Mittel die aus den Stichproben berechneten arithmetischen Mittel, den Erwartungswert der Grundgesamtheit wiedergeben. Denn  $\frac{1}{k} \sum_{j=1}^k \bar{x}^j$  ist ein arithmetisches Mittel, also ein Mittelwert. Bildet man den Erwartungswert des arithmetischen Mittels, so kann das auch nachgerechnet werden. Dazu wenden wir Satz 4.19 an:

$$(35) \quad \mathbb{E}\bar{X} = \mathbb{E}\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X = \mathbb{E}X.$$

Dabei folgt aus der Definition 8.1, da  $X_i$  und  $X$  die gleiche Verteilung haben, dass auch die Erwartungswerte von  $X_i$  und  $X$  gleich sind. Andererseits können die arithmetischen Mittel von einzelnen Stichproben sehr weit vom Mittelwert der Grundgesamtheit entfernt sein.

Wir wollen nun überlegen, inwiefern eine Schätzung des Erwartungswertes der Grundgesamtheit *besser* wird, falls wir den Stichprobenumfang, also  $n$ , vergrößern. Dazu berechnen wir die Varianz der Stichprobenfunktion  $\bar{X}$ . Um Satz 4.25 anwenden zu können, nutzen wir, dass die Zufallsvariablen  $X_i$  unabhängig sind, siehe Definition 8.1:

$$(36) \quad D^2\bar{X} = D^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2X_i = \frac{1}{n} D^2X.$$

Dabei haben wir wieder angewendet, dass  $X$  und  $X_i$  die gleichen Verteilungen besitzen. Wir haben ausgerechnet, dass die Varianz von  $\bar{X}$  kleiner wird, falls der Umfang der Stichprobe  $n$  größer wird. Das heißt, die aus den verschiedenen Stichproben berechneten arithmetischen Mittel streuen weniger um den Erwartungswert  $\mathbb{E}\bar{X}$  herum, der nach (35) gleich dem Erwartungswert der Grundgesamtheit ist.

<sup>2</sup>Die Konvergenz dieses Ausdrucks für  $k$  gegen unendlich ist durch das Gesetz der großen Zahlen richtig, siehe Bauer [1].

Dieses Resultat kann als Vergrößerung der Genauigkeit der Schätzung interpretiert werden.

Quantitativ lässt sich die Wahrscheinlichkeit für große Abweichungen auch mittels der Tschebyshev-Ungleichung (siehe Satz 4.28) beschreiben. Dazu bezeichnen wir mit  $\bar{X}_n$  das arithmetische Mittel, das aus  $n$  Stichprobenelementen gebildet wird:

$$(37) \quad \mathbb{P}(|\bar{X}_n - \mathbb{E}X| > \varepsilon) = \mathbb{P}(|\bar{X}_n - \mathbb{E}\bar{X}_n| > \varepsilon) \leq \frac{D^2 \bar{X}_n}{\varepsilon^2} = \frac{D^2 X}{n \varepsilon^2}.$$

Der Ausdruck unter der Wahrscheinlichkeit beschreibt, dass die Abweichung zwischen  $\bar{X}_n$  und  $\mathbb{E}X$  größer als  $\varepsilon$  ist. Die rechte Seite konvergiert gegen Null für  $n \rightarrow \infty$ , falls wir  $D^2 X < \infty$  annehmen.

## 2. Eigenschaften von Punktschätzungen

Im letzten Abschnitt haben wir Schätzungen kennengelernt, die dazu dienen, Aussagen über die unbekannt Parameter einer Grundgesamtheit zu machen. Solche Schätzungen werden auch *Punktschätzungen* genannt, da das Ergebnis so einer Schätzung eine Zahl, also ein Punkt auf der Zahlengeraden ist. Wir haben auch gesehen, dass eine Funktion mit Definitionsbereich  $\mathbb{R}^n$  als Schätzung dienen kann. Wir wollen nun Kriterien kennenlernen, die sichern, dass so eine Schätzung sinnvoll für die statistische Analyse ist.

**DEFINITION 8.3.** *Es sei  $X_1, \dots, X_n$  eine mathematische Stichprobe für die Grundgesamtheit  $X$ . Eine Schätzung  $T(X_1, \dots, X_n)$  heißt erwartungstreu für den Parameter  $p$  der Grundgesamtheit  $X$ , falls gilt:*

$$\mathbb{E}T(X_1, \dots, X_n) = p.$$

$T(X_1, \dots, X_n)$  wird asymptotisch erwartungstreu genannt, falls

$$\lim_{n \rightarrow \infty} \mathbb{E}T(X_1, \dots, X_n) = p.$$

Die Erwartungstreue einer Schätzung bedeutet, dass falls viele konkrete Stichproben parallel gezogen werden, im Mittel der durch diese Stichproben berechnete Schätzwert dem unbekannt Parameter  $p$  entspricht. Bei asymptotisch erwartungstreuen Schätzungen stellt sich der unbekannt Parameter  $p$  im Mittel dann ein, falls der Umfang der Stichprobe gegen unendlich geht. Falls eine Schätzung erwartungstreu ist, dann ist sie natürlich auch asymptotisch erwartungstreu.

Im letzten Abschnitt hatten wir das arithmetische Mittel als Schätzung für den Parameter Erwartungswert kennengelernt. Formel (35) zeigt, dass diese Schätzung erwartungstreu ist.

Desweiteren haben wir im letzten Abschnitt gesehen, dass es zwei Schätzungen für den Parameter Varianz ( $D^2 X$ ) der Grundgesamtheit gibt. Wir wollen zuerst prüfen, ob die Schätzung  $\tilde{S}^2$  erwartungstreu ist, das heißt, ob gilt:

$$\mathbb{E}\tilde{S}^2 = D^2 X.$$

Wir erhalten wegen

$$X_i - \bar{X} = (X_i - \mathbb{E}X) - (\bar{X} - \mathbb{E}X) = (X_i - \mathbb{E}X) - \left(\frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}X)\right)$$

nach der binomischen Formel

$$(38) \quad \mathbb{E}(X_i - \bar{X})^2 = \mathbb{E}(X_i - \mathbb{E}X)^2 + \mathbb{E}(\bar{X} - \mathbb{E}X)^2 - \frac{2}{n} \mathbb{E} \sum_{j=1}^n (X_j - \mathbb{E}X)(X_i - \mathbb{E}X).$$

Wir betrachten die Erwartungswerte der Summanden der obigen Summe:

$$(39) \quad \mathbb{E}(X_i - \mathbb{E}X)(X_j - \mathbb{E}X) = \mathbb{E}X_i X_j - \mathbb{E}X_i \mathbb{E}X - \mathbb{E}X \mathbb{E}X_j + \mathbb{E}X \mathbb{E}X.$$

Aufgrund von Definition 8.1 sind die Zufallsvariablen  $X_i, X_j$  für  $j \neq i$  unabhängig. Nach Satz 6.4 folgt daraus, dass

$$\mathbb{E}(X_i X_j) = \mathbb{E}X_i \mathbb{E}X_j = \mathbb{E}X \mathbb{E}X$$

wegen  $\mathbb{E}X_j = \mathbb{E}X$ , da  $X_j$  und  $X$  die gleiche Verteilung besitzen. Damit ist für  $j \neq i$  der Ausdruck in (39) gleich Null. In der Summe von (38) ist nur ein Summand von Null verschieden, und zwar für  $i = j$ . Weiterhin gilt

$$\begin{aligned} \mathbb{E}(X_i - \mathbb{E}X)^2 &= \mathbb{E}(X - \mathbb{E}X)^2 = D^2 X, \\ \mathbb{E}(\bar{X} - \mathbb{E}X)^2 &= \mathbb{E}(\bar{X} - \mathbb{E}\bar{X})^2 = \frac{1}{n} D^2 X. \end{aligned}$$

Dabei haben wir benutzt, dass aufgrund der Definition (8.1)  $\mathbb{E}X_i = \mathbb{E}X = \mathbb{E}\bar{X}$  gilt. Für die letzte Formel haben wir (36) angewendet. Fassen wir diese Rechnungen zusammen, dann erhalten wir

$$\mathbb{E}(X_i - \mathbb{E}\bar{X})^2 = D^2 X + \frac{1}{n} D^2 X - \frac{2}{n} D^2 X = \frac{n-1}{n} D^2 X.$$

Nach Summation über  $i$  und Division durch  $n$  und Erwartungswertbildung ergibt sich daraus:

$$(40) \quad \mathbb{E}\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i - \bar{X})^2 = \frac{n-1}{n} D^2 X \neq D^2 X.$$

Wir haben nachgeprüft, dass die Schätzung  $\tilde{S}^2$  für die Varianz der Grundgesamtheit **nicht** erwartungstreu ist. Multipliziert man  $\tilde{S}^2$  mit  $\frac{n}{n-1}$  so erhält man  $S^2$ . Speziell erlaubt (40) die Formulierung des folgenden Satzes.

SATZ 8.4. *Die Stichprobenfunktion*

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

*ist eine erwartungstreue Schätzung für die Varianz der Grundgesamtheit.*

Aus den obigen Betrachtungen folgt unmittelbar, dass  $\tilde{S}^2$  nur *asymptotisch* erwartungstreu ist, da  $\frac{n}{n-1}$  für  $n \rightarrow \infty$  gegen Eins konvergiert.

$\tilde{S}^2$  repräsentiert das arithmetische Mittel der quadratischen Abweichung der Stichprobenelemente vom Mittelwert. Diese Größe findet ihre Anwendung in der Beschreibenden Statistik. Für die folgenden Anwendungen werden wir in diesem Kapitel immer  $S^2$  benutzen, da vom Standpunkt der Schließenden Statistik die Eigenschaften von  $S^2$  vorteilhafter für die Beschreibung der Varianz der Grundgesamtheit  $X$  sind als die von  $\tilde{S}^2$ . Es sei auch noch darauf hingewiesen, dass auf einem Taschenrechner im Allgemeinen zwei verschiedene Varianzen/Standardabweichungen implementiert sind. Eine dieser Varianzen bezieht sich auf  $\tilde{S}^2$ , die andere auf  $S^2$ . Im Allgemeinen ist der Stichprobenmedian nur eine *asymptotisch* erwartungstreue Schätzung für den Erwartungswert der Grundgesamtheit.

DEFINITION 8.5. *Es sei  $T(X_1, \dots, X_n)$  eine Schätzung für einen Parameter  $p$  der Grundgesamtheit  $X$ . Diese Schätzung wird konsistent genannt, falls für jedes  $\varepsilon > 0$*

$$(41) \quad \lim_{n \rightarrow \infty} \mathbb{P}(|T(X_1, \dots, X_n) - p| > \varepsilon) = 0$$

*gilt.*

Die Interpretation der Ungleichung (41) ist die Folgende: Angenommen wir ziehen verschiedene Stichproben und berechnen daraus verschiedene Werte für die Schätzung  $T$ . Dann ist die Wahrscheinlichkeit, dass diese Werte weiter als  $\varepsilon$  vom (unbekannten) Parameter der Grundgesamtheit entfernt sind klein, falls der Umfang der Stichprobe  $n$  groß gewählt wird. Die Konsistenz des arithmetischen Mittels  $\bar{X}$  als Schätzung für den unbekannt Parameter  $p = \mathbb{E}X$  folgt aus (37).

Die folgende Definition erlaubt zwei Schätzungen untereinander zu vergleichen.

DEFINITION 8.6. *Gegeben seien zwei asymptotisch erwartungstreue Schätzungen  $T_1(X_1, \dots, X_n)$  und  $T_2(X_1, \dots, X_n)$  für den Parameter  $p$  der Grundgesamtheit  $X$ .  $T_1$  wird als effizienter als  $T_2$  bezeichnet, falls*

$$D^2T_1 < D^2T_2$$

*gilt.*

Da die Varianz die mittlere quadratische Abweichung vom Erwartungswert ist, kann sie als Maß für die Abweichung interpretiert werden. Falls  $T_1$  effizienter ist als  $T_2$ , so bedeutet dies, dass im Mittel die aus konkreten Stichproben errechneten Werte vom Erwartungswert weniger abweichen, als die Werte, die mit  $T_2$  errechnet wurden.

Das arithmetische Mittel und der Stichprobenmedian sind asymptotisch erwartungstreue Schätzungen für den Erwartungswert der Grundgesamtheit  $\mathbb{E}X = p$ . Es lässt sich zeigen, dass das arithmetische Mittel unter gewissen Annahmen effizienter ist, als der Stichprobenmedian.

### 3. Intervallschätzungen

Wir hatten im letzten Abschnitt erwähnt, dass die Differenz zwischen einem Schätzwert und dem entsprechenden unbekannt Parameter der Grundgesamtheit groß sein kann. Dabei ist unbekannt, wie groß diese Differenz ist. Falls die Schätzung konsistent ist, so weiß man lediglich, dass bei einem hinreichend großen Stichprobenumfang die Wahrscheinlichkeit für große Abweichungen vom Parameter klein ist.

Solche Fehler sind prinzipiell nicht zu vermeiden, beruhen sie doch auf der relativ kleinen Information über die Grundgesamtheit, die in der Stichprobe *gespeichert* ist.

Eine Methode, die zwar nicht diese Fehler vermeidet, sie aber trotzdem quantitativ mit in die Rechnung einbezieht, sind sogenannte *Intervallschätzungen*. Ziel dieser Methode ist es, ein Intervall anzugeben, das den interessierenden Parameter der Grundgesamtheit enthält. Die Grenzen des Intervalls werden mittels einer Stichprobe berechnet. Da diese Stichprobe wieder nur wenig Information über die Grundgesamtheit enthält, kann es zu Fehlern kommen. Einen Fehler zu machen bedeutet nun, dass man ein Intervall aus der Rechnung erhält, welches den interessierenden Parameter *nicht* enthält. Solche Fehler sind objektiv gegeben und auch durch sehr genaues Arbeiten nicht vermeidbar. Die Wahrscheinlichkeit für so einen Fehler, die man sich vorgibt, wird mit in die Rechnung aufgenommen.

Eine Intervallschätzung liefert ein Intervall, welches den Parameter der Grundgesamtheit mit vorgegebener Wahrscheinlichkeit enthält beziehungsweise nicht enthält. Die vorgegebene Wahrscheinlichkeit  $\alpha \in (0, 1)$ , dass der Parameter im Intervall nicht enthalten ist, wird als *Irrtumswahrscheinlichkeit* bezeichnet. Man irrt sich also mit Wahrscheinlichkeit  $\alpha$ , wenn man annimmt, dass das berechnete Intervall den unbekanntem Parameter enthält. Komplementär dazu wird mit  $1 - \alpha$  die *statistische Sicherheit* bezeichnet, dass das Intervall den Parameter enthält. Die so berechneten Intervalle werden auch als *Konfidenzintervalle* bezeichnet.

DEFINITION 8.7. Es sei  $X_1, \dots, X_n$  eine mathematische Stichprobe aus der Grundgesamtheit  $X$ . Dann ist das (mathematische) zweiseitige Konfidenzintervall für den Parameter  $p$  der Grundgesamtheit  $X$  ( $G_u^\alpha(X_1, \dots, X_n), G_o^\alpha(X_1, \dots, X_n)$ ) in Abhängigkeit von der Irrtumswahrscheinlichkeit  $\alpha$  definiert durch

$$\mathbb{P}(p \in (G_u^\alpha(X_1, \dots, X_n), G_o^\alpha(X_1, \dots, X_n))) = 1 - \alpha.$$

Ein einseitiges Konfidenzintervall ( $G_u^\alpha(X_1, \dots, X_n), \infty$ ) oder  $(-\infty, G_o^\alpha(X_1, \dots, X_n))$  ist gegeben durch

$$\mathbb{P}(p \in (G_u^\alpha(X_1, \dots, X_n), \infty)) = 1 - \alpha$$

beziehungsweise

$$\mathbb{P}(p \in (-\infty, G_o^\alpha(X_1, \dots, X_n))) = 1 - \alpha.$$

Ein mathematisches Konfidenzintervall ist ein zufälliges Intervall, also eine intervallwertige Zufallsvariable. Realisierungen dieses zufälligen Intervalls, die durch eine konkrete Stichprobe berechnet werden, werden mit  $g_u^\alpha = g_u^\alpha(x_1, \dots, x_n)$  beziehungsweise  $g_o^\alpha = g_o^\alpha(x_1, \dots, x_n)$  bezeichnet. Wenn man durch verschiedene konkrete Stichproben verschiedene konkrete Konfidenzintervalle berechnet, so enthalten manche dieser Intervalle den Parameter, manche nicht. Die Wahrscheinlichkeit, dass der Parameter enthalten ist, ist  $1 - \alpha$ ; die Wahrscheinlichkeit, dass der Parameter nicht enthalten ist, ist  $\alpha$ , siehe Abbildung 1.

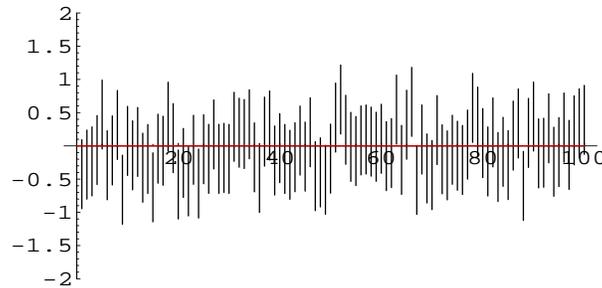


ABBILDUNG 1. Von 100 berechneten Konfidenzintervallen, beinhalten die meisten den Parameter 0 (waagerechte Linie), einige aber nicht.

Bei praktischen Untersuchungen werden üblicherweise für die Irrtumswahrscheinlichkeit  $\alpha$  folgende Werte betrachtet:

$$\alpha = 0.001, 0.005, 0.01, 0.05.$$

Bei der Wahl von  $\alpha$  tritt folgendes Dilemma auf:

$\alpha$ klein ( $\alpha = 0.001, 0.005$ )	große Sicherheit, dass das Intervall den Parameter enthält	Das Intervall ist groß, das heißt ungenaue Eingrenzung des Parameters
$\alpha$ groß ( $\alpha = 0.01, 0.05$ )	kleine Sicherheit, dass das Intervall den Parameter enthält	Das Intervall ist klein, das heißt, genaue Eingrenzung des Parameters

Im Folgenden werden wir noch sehen, dass die Genauigkeit der Eingrenzung durch den Umfang  $n$  der Stichprobe gesteuert werden kann.

Wir wollen im Folgenden Intervallschätzungen für verschiedene Parameter der Grundgesamtheit kennenlernen:

### Intervall für den Mittelwert einer normalverteilten Grundgesamtheit bei bekannter Varianz

Wir nehmen an, dass die Grundgesamtheit  $X$  eine Normalverteilung  $N(\mu, \sigma)$  besitzt, wobei der unbekannte Parameter  $\mathbb{E}X = \mu$  durch ein Intervall eingeschlossen werden soll und die Varianz  $\sigma^2$  als bekannt vorausgesetzt wird.

Mittelpunkt dieses Intervalls soll das arithmetische Mittel  $\bar{X}$  sein. Diese Wahl ist sinnvoll, denn das arithmetische Mittel ist eine erwartungstreue Schätzung für den Erwartungswert. Nach Definition 8.1 besteht also die mathematische Stichprobe  $X_1, \dots, X_n$  aus  $n$  unabhängigen normalverteilten Zufallsvariablen. Wir bestimmen als erstes die Verteilung von  $\bar{X}$ . Nach Satz 4.15 wissen wir, dass

$$X_1 + \dots + X_n$$

eine  $N(n\mu, \sqrt{n}\sigma)$  normalverteilte Zufallsvariable ist. Falls diese Summe noch durch  $n$  dividiert wird, erhalten wir  $\bar{X}$ . Nach Satz 4.19 und Satz 4.25 ergibt sich

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Wir bestimmen jetzt den Wert für  $g$ , so dass das Intervall  $(\bar{X} - g, \bar{X} + g)$  den unbekanntem Erwartungswert  $\mu$  mit Wahrscheinlichkeit  $1 - \alpha$  enthält:

$$\mathbb{P}(\mu \in (\bar{X} - g, \bar{X} + g)) = 1 - \alpha$$

oder was äquivalent dazu ist

$$(42) \quad \mathbb{P}(|\bar{X} - \mu| < g) = \mathbb{P}(-g < \bar{X} - \mu < g) = 1 - \alpha.$$

Um die Grenze  $g$  zu berechnen, transformieren wir  $\bar{X}$  in eine *standard*normalverteilte Zufallsvariable

$$(43) \quad Y = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

Damit gilt, falls die Ungleichungen in (42) noch durch  $\sigma/\sqrt{n}$  dividiert werden

$$\begin{aligned} \mathbb{P}(-g < \bar{X} - \mu < g) &= \mathbb{P}\left(\frac{-g}{\sigma} \sqrt{n} < Y < \frac{g}{\sigma} \sqrt{n}\right) \\ &= \Phi\left(\frac{g}{\sigma} \sqrt{n}\right) - \Phi\left(-\frac{g}{\sigma} \sqrt{n}\right) = 2\Phi\left(\frac{g}{\sigma} \sqrt{n}\right) - 1 = 1 - \alpha, \\ \Phi\left(\frac{g}{\sigma} \sqrt{n}\right) &= 1 - \frac{\alpha}{2}. \end{aligned}$$

Damit ist

$$\frac{g}{\sigma}\sqrt{n} = y_{1-\frac{\alpha}{2}}$$

das  $1 - \frac{\alpha}{2}$ -Quantil  $y_{1-\frac{\alpha}{2}}$  der Standardnormalverteilung, siehe Definition 4.27. Für die Grenzen des konkreten Konfidenzintervalls erhalten wir:

$$g = y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\mu \in \left( \bar{x} - y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right), \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Der Erwartungswert der Grundgesamtheit liegt also in diesem Intervall mit der Wahrscheinlichkeit  $1 - \alpha$ .

Es sei noch bemerkt, dass  $\sqrt{n}$  im Nenner von  $g$  auftritt. Damit wird das Intervall kleiner, falls  $n$  größer wird. Je größer also  $n$  desto kleiner ist die Eingrenzung von  $\mu$  durch das Konfidenzintervall.

**Beispiel:** Es soll ein Intervall berechnet werden, dass die mittlere Reißfestigkeit von Bremsseilen mit einer Wahrscheinlichkeit von 0.95 enthält. Wir nehmen an, dass die Reißfestigkeit durch eine normalverteilte Zufallsvariable mit Standardabweichung von  $10 \text{ N/mm}^2$  beschrieben werden kann. Es werden mit 20 Seilen Reißversuche durchgeführt. Daraus ergeben sich folgende Messdaten:

$$(44) \quad \begin{array}{l} 95.5, 79.1, 79.4, 84.1, 57.2, 84.7, 82.9, 73.0, 83.3, 85.6, 76.9, 66.1, 74.9, \\ 91.6, 82.0, 86.2, 78.4, 83.1, 90.0, 77.8[N/mm^2]. \end{array}$$

Als arithmetisches Mittel ergibt sich

$$\bar{x} = 80.59[N/mm^2].$$

Desweiteren erhalten wir für das Quantil  $y_{1-\frac{\alpha}{2}} = y_{0.975} = 1.96$  (siehe Tabelle in [19]) und  $\sigma = 10[N/mm^2]$  nach Voraussetzung. Damit ist der Erwartungswert  $\mu$  mit Wahrscheinlichkeit 0.95 in dem Intervall

$$(45) \quad \left( 80.59 - 1.96 \frac{10}{\sqrt{20}}, 80.59 + 1.96 \frac{10}{\sqrt{20}} \right) = (76.2, 84.97)$$

enthalten.

Vom Standpunkt eines Qualitätskontrolleurs stelle man sich vor, dass er diese Bremsseile akzeptieren würde, falls 95 % der Seile eine Reißfestigkeit größer als  $75[N/mm^2]$  besitzen. Allerdings scheint eine Begrenzung der Reißfestigkeit nach oben, so wie durch das Intervall (45) gegeben, für die Qualität der Bremsseile unerheblich. Deswegen macht es mehr Sinn, ein einseitiges Konfidenzintervall  $(g_u, \infty)$  für den Erwartungswert zu berechnen. Unter der Voraussetzung der Normalverteiltetheit von  $X$ , wobei die Varianz bekannt ist, erhalten wir:

$$\mathbb{P}(\mu \in (\bar{X} - g, \infty)) = \mathbb{P}(g \leq \bar{X} - \mu) = 1 - \alpha.$$

Eine Transformation auf die Standardnormalverteilung ergibt, dass

$$\mu \in \left( \bar{x} - y_{1-\alpha} \frac{\sigma}{\sqrt{n}}, \infty \right)$$

mit Wahrscheinlichkeit  $1 - \alpha$ .

**Beispiel:** Wir berechnen das einseitige Konfidenzintervall für die Daten aus (44) mit  $\alpha = 0.01$ . Das 0.99-Quantil  $y_{0.99} = 2.326$ , woraus wir das Intervall

$$(75.39, \infty)$$

erhalten. (Man beachte den Unterschied zwischen den Quantilen der zweiseitigen und einseitigen Konfidenzintervalle.)

Ein einseitiges Konfidenzintervall, das in die andere Richtung geöffnet ist, ist gegeben durch

$$\mu \in \left(-\infty, \bar{x} + y_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right).$$

mit Wahrscheinlichkeit  $1 - \alpha$ .

### Intervall für den Mittelwert einer normalverteilten Grundgesamtheit bei unbekannter Varianz

Die für die Berechnung der Konfidenzintervalle gerade gemachte Voraussetzung, dass die Varianz der normalverteilten Grundgesamtheit bekannt ist, ist in der Praxis selten anzutreffen. Wenn keine Informationen über die Varianz vorliegen, muss sie aus der gezogenen Stichprobe geschätzt werden. Dazu benutzen wir die *erwartungstreue* Schätzung für die Varianz

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Die in (43) durchgeführte Transformation in eine Standardnormalverteilung ist nun zu ersetzen durch

$$(46) \quad \frac{\bar{X} - \mu}{S} \sqrt{n} =: T_{n-1},$$

wobei im Nenner jetzt die *Zufallsvariable*  $S = \sqrt{S^2}$  erscheint. Dadurch ist dieser Ausdruck nicht standardnormalverteilt, sondern besitzt eine *Student-Verteilung*, auch als *T-Verteilung* bezeichnet. Diese Verteilung hängt vom Umfang der Stichprobe  $X_1, \dots, X_n$  ab. Diese Abhängigkeit wird durch den Index  $n - 1$  beschrieben. Man sagt, dass  $T_{n-1}$  eine Studentverteilte Zufallsvariable mit  $n - 1$  *Freiheitsgraden* ist.

Die Student-Verteilung beschreibt die Wahrscheinlichkeit einer stetigen Zufallsvariablen. Ihr Graph sieht so ähnlich aus, wie der der Dichte einer standardnormalverteilten Zufallsvariablen, siehe Abbildung 2. Die Unterschiede zwischen der Verteilung einer standardnormalverteilten Zufallsvariablen und einer Studentverteilten Zufallsvariablen sind sehr gering, falls die Anzahl der Freiheitsgrade  $n$  größer als 60 ist. In diesem Fall können die Quantile der Student-Verteilung durch die Quantile der Normalverteilung ersetzt werden.

Die Formel für die Konfidenzintervalle ergeben sich nun analog aus den obigen Berechnungen. Dabei ist  $\sigma$  durch den aus der konkreten Stichprobe berechneten Wert

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

zu ersetzen. Wir erhalten, dass

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}\right)$$

mit Wahrscheinlichkeit  $1 - \alpha$  gilt.  $t_{1-\frac{\alpha}{2}, n-1}$  sind die Quantile der Studentverteilung mit  $n - 1$  Freiheitsgrade. Diese können aus der Tabelle in [19] entnommen werden.

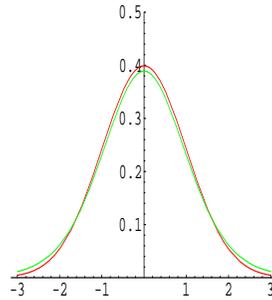


ABBILDUNG 2. Die Dichtefunktionen der Standardnormalverteilung und der Studentverteilung mit 10 Freiheitsgraden übereinandergelegt.

Ähnlich können einseitige Intervalle berechnet werden:

$$\mu \in \left( \bar{x} - t_{1-\alpha, n-1} \frac{s}{\sqrt{n}}, \infty \right)$$

beziehungsweise

$$\mu \in \left( -\infty, \bar{x} + t_{1-\alpha, n-1} \frac{s}{\sqrt{n}} \right)$$

mit Wahrscheinlichkeit  $1 - \alpha$ .

**Beispiel:** Um die Qualität einer Abfüllanlage zu beschreiben, soll ein Intervall angegeben werden, das mit Wahrscheinlichkeit  $0.95 (= 1 - \alpha)$  den Erwartungswert der Grundgesamtheit enthält. Es wird davon ausgegangen, dass die Füllmenge normalverteilt ist. Folgende Stichprobe wurde ermittelt:

549, 547, 549, 549, 545, 550, 550, 545, 550, 544, 543, 549, 548 [g]

Wir berechnen aus der Stichprobe

$$\bar{x} = 547.54, \quad s = 2.47$$

Da  $\sigma^2$  durch  $s^2$  geschätzt wurde, müssen wir die Quantile der Studentverteilung mit  $n - 1 = 12$  berechnen:

$$t_{1-\frac{\alpha}{2}, n-1} = t_{0.975, 12} = 2.179.$$

Damit erhalten wir für das Intervall, das den Erwartungswert  $\mu$  mit Wahrscheinlichkeit von 0.95 enthält:

$$(546.05, 549.03).$$

Zum Abschluß dieses Kapitels wollen wir noch einige Intervallschätzungen für Parameter der Grundgesamtheit erwähnen, ohne auf die Herleitung der entsprechenden Formel einzugehen.

### Intervallschätzung für die Varianz einer normalverteilten Grundgesamtheit

Zweiseitiges Intervall:

$$\left( \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \right).$$

$\chi_{1-\frac{\alpha}{2}, n-1}^2, \chi_{\frac{\alpha}{2}, n-1}^2$  bezeichnet die Quantile der  $\chi^2$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Einseitige Intervalle:

$$\left( \frac{(n-1)s^2}{\chi_{1-\alpha, n-1}^2}, \infty \right), \quad \left( -\infty, \frac{(n-1)s^2}{\chi_{\alpha, n-1}^2} \right).$$

### Intervallschätzung für die Wahrscheinlichkeit $\mathbb{P}(A)$ der Grundgesamtheit

Es sei  $h_n(A)$  die (konkrete) relative Häufigkeit des Auftretens von  $A$  der konkreten Stichprobe  $x_1, \dots, x_n$ . Die mathematische relative Häufigkeit  $H_n(A)$  ist eine erwartungstreue Schätzung für  $\mathbb{P}(A)$ . Wir setzen voraus, dass der Stichprobenumfang nicht zu klein ist. *Näherungsweise* erhalten wir die folgende Formel:

Zweiseitiges Intervall:

$$\left( h_n(A) - y_{\frac{\alpha}{2}} \sqrt{\frac{h_n(A)(1-h_n(A))}{n}}, h_n(A) + y_{\frac{\alpha}{2}} \sqrt{\frac{h_n(A)(1-h_n(A))}{n}} \right).$$

Eine Übersicht über weitere Intervallschätzungen findet man in Beichelt [3].

### Bestimmung von statistischen Anteilsbereichen

Die Bestimmung von statistischen Anteilsbereichen ist verwandt mit der Bestimmung von Konfidenzintervallen. Dabei geht es darum, ein Intervall  $(G_u, G_o)$  zu bestimmen, das mit einer vorgegebenen Wahrscheinlichkeit  $\beta$  mindestens den Anteil  $\gamma$  der Grundgesamtheit  $X$  enthält. Speziell heißt das, dass die Wahrscheinlichkeit

$$\mathbb{P}(G_u < X < G_o) > \gamma$$

gleich  $\beta$  ist. Unter der Voraussetzung, dass  $X$  normalverteilt ist, können aus einer Stichprobe  $x_1, \dots, x_n$  konkrete Werte für die Grenzen  $G_o, G_u$  geschätzt werden. Entsprechende Formeln findet man in Storm [20].

## 4. Maximum Likelihood-Schätzung

In diesem Abschnitt soll es darum gehen, gewisse Parameter einer Grundgesamtheit zu schätzen. Wir können zum Beispiel annehmen, dass zwar der Typ der Verteilung bekannt ist, gewisse Parameter dieser Verteilung aber unbekannt sind und somit geschätzt werden müssen. Wir kennen also die Modellklasse, aus der nun durch die Maximum Likelihood-Schätzung das passende Modell bestimmt werden muss. Speziell soll es darum gehen, das passende Modell so zu bestimmen, dass es am Besten zur *konkreten Stichprobe* passt. Oder mit anderen Worten: Wir bestimmen das passende Modell so, dass die konkrete Stichprobe mit größt möglicher Wahrscheinlichkeit auftritt. Da im Englischen Likelihood auch *Wahrscheinlichkeit* bedeutet, wird somit der Name der Methode klar.

Wir betrachten ein Beispiel. Zur Qualitätskontrolle werden aus einer Packung von Ventilen  $n = 10$  Ventile entnommen und wieder zurückgelegt. Unter der Annahme, dass  $S_n = k \in \{0, 1, \dots, n\}$  defekte Teile auftreten, können wir diesen Qualitätstest durch ein Bernoulli-Experiment beschreiben. Die entsprechende Modellklasse besteht aus den  $(n, p)$ -Binomialverteilungen, wobei der Parameter  $n = 10$  schon bekannt ist. Durch die Maximum Likelihood-Schätzung soll also der Parameter  $p$  der Verteilung so bestimmt werden, dass er am Besten mit der Information der konkreten Stichprobe,  $k$  defekte Ventile unter  $n$  Ventilen, übereinstimmt.

Zu diesem  $k$  wird nun die zugehörige *Maximum Likelihood-Funktion* formuliert:

$$L_k(p) = \text{Bin}(10, p)(S_{10} = k) = \binom{10}{k} p^k (1-p)^{10-k}$$

$L_k(p)$  is also eine Funktion mit Argument  $p \in [0, 1]$ , wobei  $k$  durch das Ergebnis der konkreten Stichprobe, also Anzahl der defekten Teile unter  $n = 10$ , gegeben ist. Der Schätzwert  $p^*$  ist dann gegeben durch

$$L_k(p^*) = \max_{0 \leq p \leq 1} L_k(p).$$

Da  $L_k(p)$  die Wahrscheinlichkeit von dem Ereignis  $\{S_n = k\}$  unter der Annahme, dass bei einer Entnahme mit Wahrscheinlichkeit  $p$  ein defektes Teil auftritt, bestimmen wir mit der obigen Extremwertaufgabe  $p$  so, dass es am Besten zu dem Beobachtungsergebnis der konkreten Stichprobe  $\{S_n = k\}$  passt.

Wir lösen die obige Extremwertaufgabe. Falls  $k = 0$  beobachtet wurde, so lautet die Maximum Likelihood-Funktion

$$L_k(p) = L_0(p) = (1-p)^n.$$

Damit erhält man  $p^* = 0$  als Maximalstelle. Ähnlich für  $k = n$  mit der Maximalstelle  $p^* = 1$ . Für  $0 < p < 1$  überprüfen wir die notwendige Bedingung für einen Extremwert:

$$\frac{d}{dp} L_k(p) = \binom{n}{k} p^{k-1} (1-p)^{n-k-1} (k(1-p) - (n-k)p) = 0.$$

Da ein Produkt Null ist genau dann, falls einer der Faktoren Null ist, und  $0 < p < 1$  gilt, kann die extremwertverdächtige Stelle aus

$$(k(1-p) - (n-k)p) = 0$$

bestimmt werden. Die Lösung dieser Gleichung ist  $p^* = k/n$ . Es ist nicht schwer zu sehen, dass  $\frac{d}{dp} L_k(p)$  für  $p < p^*$  positiv und für  $p > p^*$  negativ ist. Somit ist das berechnete  $p^*$  die Maximum Likelihood-Schätzung für  $p$ .

Zusammenfassend sei also noch einmal bemerkt, dass man um eine Maximum Likelihood-Schätzung durchzuführen, zwei Schrittenotwendig sind:

- Aufstellung einer zur Modellklasse gehörende Maximum Likelihood-Funktion.
- Maximierung dieser Funktion.

Abschließend betrachten wir ein weiteres Beispiel.

Wir nehmen an, dass die Lebensdauer von  $n$  Speicherchips durch eine konkrete Stichprobe  $x_1, \dots, x_n$  bestimmt wurde. Unter der Annahme, dass die Grundgesamtheit exponentialverteilt ist (Modellklasse  $X \sim \text{Exp}(\lambda)$ ,  $\lambda > 0$ ), soll der Parameter *mittlere Lebensdauer*  $\mathbb{E}X = 1/\lambda$  durch die Maximum Likelihood-Schätzung bestimmt werden. Die Dichte der Exponentialverteilung ist gegeben durch (15). Wir bestimmen den Parameter der Dichte nun so, dass er am Besten zu den Werten der konkreten Stichprobe passt. Somit erhalten wir als Maximum Likelihood-Funktion zu einer konkreten Stichprobe  $x_1, \dots, x_n$

$$L_{(x_1, \dots, x_n)}(\lambda) = f_n(x_1, \dots, x_n) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

Speziell beschreibt Dichtefunktion  $f_n(x_1, \dots, x_n)$  auch die Verteilung der mathematischen Stichprobe  $(X_1, \dots, X_n)$ . In manchen Fällen vereinfacht sich die Rechnung, wenn man die Maximum Likelihood-Funktion logarithmiert. Da die Logarithmusfunktion streng wachsend ist, besitzen beide Funktionen (die originale und die logarithmierte  $\hat{L}_{(x_1, \dots, x_n)}$ ) das gleiche Maximum. Es gilt

$$\begin{aligned} \hat{L}_{(x_1, \dots, x_n)}(\lambda) &:= \log(L_{(x_1, \dots, x_n)}(\lambda)) = n \log \lambda - \lambda \sum_{i=1}^n x_i \\ \frac{d}{d\lambda} \hat{L}_{(x_1, \dots, x_n)}(\lambda) &= \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0, \end{aligned}$$

woraus für das Maximum  $\lambda^* = n / \sum_{i=1}^n x_i$  folgt. Also ergibt die Schätzung für die mittlere Lebensdauer gerade das Reziproke des arithmetischen Mittels  $\bar{x}$  der konkreten Stichprobe. Die Maximum Likelihood-Schätzung liefert also in diesem Fall die erwartungstreue Schätzung für den Erwartungswert einer Grundgesamtheit.



## Statistische Tests

### 1. Die Grundidee eines Tests

Statistische Tests (statistische Prüfverfahren) dienen zur Überprüfung von Hypothesen. Diese Hypothesen beziehen sich auf die Grundgesamtheit. So kann zum Beispiel die Hypothese aufgestellt werden, dass die Grundgesamtheit einen bestimmten Erwartungswert besitzt. Die Entscheidung, ob die Hypothese verworfen oder nicht verworfen wird, wird anhand einer Stichprobe gefällt. Da diese Stichprobe wieder nur eine kleine Information bezüglich der Grundgesamtheit enthält, können durchaus Fehlentscheidungen über die Ablehnung beziehungsweise Nichtablehnung der Hypothese auftreten.

Die bei einem Test zu überprüfende Hypothese wird als *Nullhypothese*  $H_0$  bezeichnet. Parallel zur Nullhypothese wird eine *Alternativhypothese*  $H_a$  formuliert, die automatisch angenommen wird, falls das Testverfahren ergibt, dass  $H_0$  abgelehnt werden muss. Beispiele für Hypothesen sind:

Der Mittelwert der vom Abfüllautomaten abgepackten Mehltüten beträgt 500g:

$$H_0 : \mathbb{E}X = 500g$$

Für die zugehörige Alternativhypothese gibt es verschiedene Möglichkeiten:

$$H_a : \mathbb{E}X \neq 500g \quad \text{oder} \quad H_a : \mathbb{E}X < 500g \quad \text{oder} \quad H_a : \mathbb{E}X > 500g.$$

Welche dieser Alternativhypothesen gewählt wird, hängt zum Beispiel davon ab, ob man sich dafür interessiert, ob der Erwartungswert von 500g abweicht, oder ob nur Abweichungen nach unten beziehungsweise nach oben interessieren. Das zu wählende Testverfahren kann von der Alternativhypothese abhängen.

Zwei Produktionsstädten produzieren das gleiche Produkt (zum Beispiel Glühbirnen). Die Hypothese besteht darin, dass man annimmt, dass die mittlere Lebensdauer des in Betrieb 1 produzierten Produktes gleich der mittleren Lebensdauer des in Betrieb 2 produzierten Produktes ist:

$$H_0 : \mathbb{E}T_1 = \mathbb{E}T_2$$

und als Alternativhypothese

$$H_a : \mathbb{E}T_1 \neq \mathbb{E}T_2.$$

In Abschnitt 3 des Kapitels 8 haben wir Intervallschätzungen betrachtet. Dabei haben wir vorausgesetzt, dass die Grundgesamtheit normalverteilt ist. Diese Annahme sollte natürlich durch einen statistischen Test unterlegt werden. Man würde als Nullhypothese formulieren:

$$H_0 : X \sim N(\mu, \sigma), \quad H_a : X \not\sim N(\mu, \sigma).$$

In den folgenden Kapiteln werden wir noch weitere Tests kennenlernen, die sich auf spezielle Problemstellungen der Statistik beziehen.

Wir beschreiben nun das Vorgehen, das über die Ablehnung, Nichtablehnung einer Hypothese entscheidet. Nehmen wir an, dass die Hypothese über den Erwartungswert der Grundgesamtheit  $X$

$$\mathbb{E}X = \mu_0$$

getestet werden soll. Die Information über die Grundgesamtheit liegt neben gewissen Voraussetzungen über die Verteilung (auf die erst später eingegangen wird) als konkrete Stichprobe

$$x_1, x_2, \dots, x_n$$

vor. Da das arithmetische Mittel

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

eine *erwartungstreue* Schätzung für den Erwartungswert der Grundgesamtheit ist (siehe Kapitel 8), sollte das *konkrete* arithmetische Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

näherungsweise  $\mathbb{E}X$  wiedergeben. Die Entscheidung über die Hypothese  $H_0$  wird nun wie folgt getroffen: Falls die Abweichung zwischen  $\bar{x}$  und dem hypothetischen Wert  $\mu_0$  *groß* ist, so wird die Hypothese  $H_0$  abgelehnt. Den Bereich, in dem  $\bar{x}$  liegen muss, damit  $H_0$  abgelehnt wird, bezeichnet man als *kritischen Bereich*. Wenn also das arithmetische Mittel  $\bar{x}$  einen großen Abstand von  $\mu_0$  hat, so geht man davon aus, dass die erwartungstreue Schätzung nicht den Erwartungswert  $\mu_0$  besitzt. *Es ist unwahrscheinlich, dass  $\bar{x}$  einen großen Abstand von dem in der Nullhypothese formulierten Wert hat, obwohl  $H_0$  gilt.*  $\bar{X}$  ist eine erwartungstreue Schätzung, und somit

$$\bar{x} \approx \mathbb{E}\bar{X} = \mathbb{E}X \neq \mu_0,$$

das heißt,  $H_0$  wird verworfen und  $H_a$  wird akzeptiert. Mit anderen Worten: Die Abweichung zwischen  $\bar{x}$  und  $\mu_0$  ist *signifikant*, um  $H_0$  abzulehnen.

Da die Elemente der Stichprobe  $x_1, \dots, x_n$  zufällige Werte sind, ist auch im Falle, dass  $\mathbb{E}X = \mu_0$  ist, das arithmetische Mittel  $\bar{x}$  im Allgemeinen ungleich dem Erwartungswert der Grundgesamtheit. In diesem Fall geht man davon aus, dass die Abweichung zwischen  $\bar{x}$  und  $\mu_0$  nur auf der Zufälligkeit der Stichprobe beruht, die nicht zur Ablehnung von  $H_0$  führt.

Da wieder wenig Information über die Grundgesamtheit vorliegt, können fehlerhafte Schlussfolgerungen über die Grundgesamtheit gezogen werden. Dabei sind zwei qualitativ unterschiedliche Arten von Fehlern möglich. Einerseits kann der sogenannte  *$\alpha$ -Fehler* oder auch als *Fehler 1. Art* bezeichnet auftreten, was bedeutet, dass  $H_0$  abgelehnt wird, obwohl  $H_0$  zutrifft. Das tritt dann auf, wenn die Zufälligkeit der Stichprobenelemente einen Schätzwert ergeben, der sehr weit von dem in  $H_0$  formulierten Wert entfernt ist. *Es ist unwahrscheinlich, dass  $\bar{x}$  einen großen Abstand von dem in der Nullhypothese formulierten Wert hat, obwohl  $H_0$  gilt; es ist aber nicht unmöglich.*

Andererseits kann wieder die Zufälligkeit der Stichprobenelemente einen Schätzwert liefern, der in der Nähe des in  $H_0$  formulierten Wertes ist, obwohl die Grundgesamtheit einen ganz anderen Parameter besitzt. Dieser Fehler heißt  $\beta$ -Fehler oder *Fehler 2. Art*.

Die Situation ist mit einem Gerichtsverfahren vergleichbar. Ist  $H_0$  die vor der Verhandlung bestehende Hypothese, dass der Angeklagte unschuldig ist, dann sind folgende Entscheidungen möglich:

		Wahrheit	
		unschuldig	schuldig
Urteil	schuldig	$\alpha$ -Fehler	
	unschuldig		$\beta$ -Fehler

TABELLE 1.  $\alpha$ - und  $\beta$ -Fehler bei Gerichtsverhandlung

Bevor wir beginnen, ein Testverfahren zu beschreiben, sei auf folgende wichtige Interpretation hingewiesen. Speziell ist dem Leser vielleicht aufgefallen, dass wir die Formulierung *Die Hypothese wird angenommen* vermieden haben. **Wird eine Hypothese angenommen, also nicht abgelehnt, so bedeutet das nur, dass wir nicht hinreichend viel Information aus der Stichprobe erhalten haben, um die Hypothese abzulehnen.**

## 2. Test des Mittelwertes

Wir werden nun ein Testverfahren für die Null- beziehungsweise Alternativhypothese

$$H_0: \mathbb{E}X = \mu_0, \quad H_a: \mathbb{E}X \neq \mu_0$$

herleiten. Wir nehmen für die Grundgesamtheit an, dass sie normalverteilt ist mit bekannter Varianz  $D^2X = \sigma^2$ :

$$X \sim N(\mu, \sigma).$$

Falls die Nullhypothese gilt, ist  $\mathbb{E}X = \mu = \mu_0$ . Eine erwartungstreue Schätzung für den Erwartungswert  $\mu$  ist das arithmetische Mittel.

Wir werden nun festlegen, wann der Abstand zwischen  $\mu$  und  $\bar{x}$  als hinreichend groß eingeschätzt wird, damit die Hypothese  $H_0$  abgelehnt wird, oder in anderen Worten, was der kritische Bereich für diesen Test ist. Ausgangspunkt für diesen Test ist der vorgegebene Fehler 1. Art, der mit  $\alpha$  bezeichnet wurde. Dabei ist  $\alpha$  der Wert für eine Wahrscheinlichkeit. Diese Wahrscheinlichkeit beschreibt das Ereignis, dass  $H_0$  aufgrund der Zufälligkeit von  $\bar{X}$  abgelehnt wird, obwohl  $H_0$  gilt. Der  $\alpha$ -Fehler tritt auf, falls man unter der Annahme, dass die Grundgesamtheit  $N(\mu_0, \sigma)$ -verteilt ist, Stichprobenelemente  $x_1, \dots, x_n$  zufällig ermittelt, die Werte besitzen, so dass die Abweichung des arithmetischen Mittels  $\bar{x}$  dieser Stichprobe von  $\mu_0$  groß ist. Der kritische Bereich wird so konstruiert, dass große Abstände des arithmetischen Mittels von  $\mu_0$  mit Wahrscheinlichkeit  $\alpha$  auftreten. Gibt man sich für einen Test die Wahrscheinlichkeit  $\alpha$  vor, so wählt man ein Risiko einen Fehler 1. Art zu begehen. Es sei der kritische Bereich das Komplement des Intervalles  $(\mu_0 - g, \mu_0 + g)$ .  $g$  wird in Abhängigkeit von  $\alpha$ ,  $\sigma$  und dem Stichprobenumfang  $n$  so bestimmt, dass gilt:

$$\mathbb{P}(\bar{X} \in (\mu_0 - g, \mu_0 + g)) = 1 - \alpha,$$

oder was zu dieser Gleichung äquivalent ist:

$$\mathbb{P}(-g < \bar{X} - \mu_0 < g) = 1 - \alpha$$

Bei der Annahme des Vorliegens eines  $\alpha$ -Fehlers, gehen wir davon aus, dass die Grundgesamtheit  $X$  die Hypothese  $H_0$  erfüllt, also  $N(\mu_0, \sigma)$ -verteilt ist. Aufgrund der Erwartungstreue von  $\bar{X}$  und aufgrund von (36) und Satz 4.15 ist die Verteilung von  $\bar{X}$  gegeben durch

$$N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right).$$

Damit ist

$$Y = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$$

standardnormalverteilt:  $Y \sim N(0, 1)$ ,

$$\mathbb{P}\left(-g \frac{\sqrt{n}}{\sigma} < Y < g \frac{\sqrt{n}}{\sigma}\right) = 1 - \alpha.$$

Andererseits gilt nach Definition der  $1 - \frac{\alpha}{2}$ -Quantile der Standardnormalverteilung:

$$\mathbb{P}\left(-y_{1-\frac{\alpha}{2}} < Y < y_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Daraus ergibt sich für  $g$ :

$$g \frac{\sqrt{n}}{\sigma} = y_{1-\frac{\alpha}{2}}, \quad g = y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Bei einem vorgegebenen Fehler 1. Art  $\alpha$ , wird  $H_0$  abgelehnt, falls  $\bar{x}$  im kritischen Bereich liegt, das heißt:

$$(47) \quad \bar{x} \notin \left( \mu_0 - y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu_0 + y_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right).$$

Oder etwas anders formuliert: Unter der Annahme von  $H_0$  ist

$$y = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$$

die Realisierung einer standardnormalverteilten Zufallsvariablen. Der kritische Bereich ist gegeben durch

$$|y| \geq y_{1-\frac{\alpha}{2}}.$$

**Beispiel:** Es soll die Hypothese überprüft werden, dass ein Abfüllautomat normgerecht arbeitet, das heißt, dass das mittlere Gewicht der abgepackten Tüten 500g beträgt. Die Gewichte werden als normalverteilt angenommen. Es wurde aus den in einer Stunde produzierten Mehltüten die folgende Stichprobe ermittelt:

499, 500, 498, 500, 498, 498, 498, 498, 495, 499, 500, 496, 499, 497, 500 in  $g$ .

Die Hypothese lautet:

$$H_0 : \quad \mathbb{E}X = 500g, \quad H_a : \quad \mathbb{E}X \neq 500g.$$

Das Signifikanzniveau sei  $\alpha = 0.01$ . Für die Standardabweichung werde  $\sigma = 1.5$  angenommen. Aus der gegebenen Stichprobe errechnet man für  $\bar{x} = 498.3$ . Für das Quantil der Standardnormalverteilung erhält man

$$\alpha = 0.01, \quad \frac{\alpha}{2} = 0.995, \quad y_{1-\frac{\alpha}{2}} = y_{0.995} = 2.58.$$

Damit ergibt sich für den nichtkritischen Bereich das Intervall

$$\left( 500 - 2.58 \frac{1.5}{3.87}, 500 + 2.58 \frac{1.5}{3.87} \right) = (499, 501).$$

Der Wert  $\bar{x}$  liegt im kritischen Bereich, das heißt die Nullhypothese wird abgelehnt und es wird davon ausgegangen, dass die alternative Hypothese  $H_a : \mathbb{E}X \neq 500g$  zutrifft.

Der betrachtete Test basierte auf der Ablehnung der Nullhypothese, falls  $\bar{x}$  weit genug von  $\mu_0$  in *beiden Richtungen* entfernt ist. Manchmal bedingt die Alternativhypothese, dass nur Abweichungen in einer Richtung zur Ablehnung von  $H_0$  führen. Wir nehmen wieder an, dass eine Grundgesamtheit vorliegt, die normalverteilt ist und von der die Varianz bekannt ist. Zur Nullhypothese  $\mathbb{E}X = \mu_0$  formulieren wir die Alternativhypothese

$$(48) \quad H_a : \mathbb{E}X = \mu_a < \mu_0, \quad \text{oder} \quad \mathbb{E}X = \mu_a > \mu_0.$$

Nehmen wir das Signifikanzniveau  $\alpha$  als gegeben an, dann ergibt sich ein *einseitiger* kritischer Bereich als

$$\bar{x} \notin \left( \mu_0 - y_{1-\alpha} \frac{\sigma}{\sqrt{n}}, \infty \right)$$

oder

$$y \leq -y_{1-\alpha}, \quad y = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}$$

falls die Alternativhypothese einen kleineren Erwartungswert als  $\mu_0$  beschreibt. Umgekehrt, falls der in der Alternativhypothese beschriebene Erwartungswert größer als  $\mu_0$  angenommen wird:

$$\bar{x} \notin \left( -\infty, \mu_0 + y_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right)$$

oder

$$y \geq y_{1-\alpha}, \quad y = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}.$$

Man vergleiche die in den letzten Formeln stehenden Quantile mit dem Quantil in (47).

Wir betrachten den Fall, dass  $X$  normalverteilt, aber  $\sigma$  nicht bekannt ist. Dann muss  $\sigma^2$  durch die erwartungstreue Schätzung für die Varianz  $s^2$  geschätzt werden. Beachtet man wieder die Transformation (46), so gelangt man zu der Zufallsvariable

$$\frac{\bar{X} - \mu_0}{s} \sqrt{n} = T_{n-1}.$$

Diese Zufallsvariable ist *Student-verteilt* mit dem Parameter  $n-1$ . Dieser Parameter wird als *Freiheitsgrad* bezeichnet. Wir können somit den kritischen Bereich für den zweiseitigen Test beschreiben

$$\bar{x} \notin \left( \mu_0 - t_{1-\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}, \mu_0 + t_{1-\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \right)$$

oder

$$|t| \geq t_{1-\frac{\alpha}{2}, n-1}, \quad t = \frac{\bar{x} - \mu_0}{s} \sqrt{n},$$

falls die Hypothese  $H_0 : \mathbb{E}X = \mu_0$ ,  $H_a : \mathbb{E}X \neq \mu_0$  gegeben sind. Dabei ist  $t_{1-\frac{\alpha}{2}, n-1}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Studentverteilung mit  $n-1$  Freiheitsgraden. Im Falle der Hypothesen (48) erhalten wir für den kritischen Bereich

$$\bar{x} \notin \left( \mu_0 - t_{1-\alpha, n-1} \frac{s}{\sqrt{n}}, \infty \right).$$

oder

$$t \leq t_{1-\alpha, n-1}, \quad t = \frac{\bar{x} - \mu_0}{s} \sqrt{n},$$

falls die Alternativhypothese einen kleineren Erwartungswert als  $\mu_0$  beschreibt. Falls der in der Alternativhypothese beschriebene Erwartungswert größer als  $\mu_0$  angenommen wird:

$$\bar{x} \notin \left(-\infty, \mu_0 + t_{1-\alpha, n-1} \frac{s}{\sqrt{n}}\right).$$

oder

$$t \geq t_{1-\alpha, n-1}, \quad t = \frac{\bar{x} - \mu_0}{s} \sqrt{n},$$

**Beispiel:** Ein Chemiker behauptet durch einen neuen Zuschlagstoff die Festigkeit eines Produktes von  $68 \text{ N/mm}^2$  auf  $70 \text{ N/mm}^2$  erhöht zu haben. Seine Kollegen bezweifeln das. Zur Überprüfung wird ein Testverfahren herangezogen. Es wird die Festigkeit einer Stichprobe des neuen Materials untersucht. Falls die Hypothese

$$H_0 : \mathbb{E}X = 70$$

abgelehnt wird, so geht man davon aus, dass die alternative Hypothese  $H_a : \mathbb{E}X = 68$  gilt, das heißt, dass keine Verbesserung der Festigkeit durch den neuen Zuschlagstoff erreicht wird. Die Festigkeit  $X$  werde als normalverteilt vorausgesetzt.

Folgende Werte werden ermittelt:

$$70, 71, 70, 69, 67, 70, 70, 69, 69, 69, 72, 70, 70, 69.$$

Als Signifikanzniveau wird  $\alpha = 0.025$  vorgegeben. Aus den Daten wird berechnet:

$$n = 15, \quad \bar{x} = 69.67, \quad s^2 = 1.24, \quad s = 1.11, \quad t_{0.975, 14} = 2.145.$$

Somit erhalten wir für den kritischen Bereich

$$\bar{x} \notin \left(70 - 2.145 \frac{1.11}{\sqrt{15}}, \infty\right) = (69.39, \infty).$$

Da  $\bar{x} = 69.67$  in dem Intervall liegt, wird  $H_0$  nicht abgelehnt. Man geht davon aus, dass sich die Festigkeit durch den Zuschlagstoff auf  $70 \text{ N/mm}^2$  verbessert hat.

Zum Abschluß dieser einleitenden Beispiele soll noch einmal das allgemeine praktische Vorgehen bei einem Test beschrieben werden, das auch für die folgenden Tests anwendbar ist:

- (1) In Abhängigkeit des interessierenden statistischen Sachverhaltes (Parameter) Aufstellen einer Nullhypothese  $H_0$  und der zugehörigen alternativen Hypothese  $H_a$ . Festlegen des Fehlers 1. Art ( $\alpha$ -Fehler). Festlegen des Schätzwertes für den interessierenden Parameter der Grundgesamtheit, über den eine Hypothese geprüft werden soll. Festlegung des anzuwendenden Testverfahrens.
- (2) Konstruktion des kritischen Bereiches.
- (3) Ermittlung einer konkreten Stichprobe, die nicht zu klein sein sollte. Berechnung des Schätzwertes.
- (4) Entscheidung über die Hypothese
  - Falls Schätzwert im kritischen Bereich liegt, dann Ablehnung von  $H_0$ . Akzeptanz von  $H_a$ .
  - Falls Schätzwert nicht im kritischen Bereich liegt, keine Ablehnung von  $H_0$ .

Die Student-Verteilung kann auch verwendet werden, um die Gleichheit zweier Mittelwerte aus normalverteilten Grundgesamtheiten zu überprüfen.

Man stelle sich die folgende praktische Aufgabenstellung vor: Es soll überprüft werden, ob die Erhöhung des Anteils eines Stoffes in Dünger zur Gewichtsänderung bei Kartoffeln führt. Die Nullhypothese würde darin bestehen, dass die mittleren Gewichte  $\mu_1$ ,  $\mu_2$  der Grundgesamtheiten  $X_1$ ,  $X_2$ , die einmal die mit dem neuen Dünger behandelten Kartoffeln und zum Anderen die mit dem alten Dünger behandelten Kartoffeln repräsentieren, gleich sind. Beispielsweise könnten folgende Stichproben aus den Grundgesamtheiten  $X_1$ ,  $X_2$  entnommen worden sein.

$$(49) \quad \begin{array}{l} X_1 : \quad 500, 494, 511, 490, 495, 496, 499, 496, 517, 503, 481, 515 \\ X_2 : \quad 492, 509, 504, 519, 483, 511, 528, 522, 509, 506, 508, 496. \end{array}$$

Die alternative Hypothese würde den Sachverhalt beschreiben, dass unterschiedliche mittlere Gewichte vorliegen:

$$H_0 : \quad \mu_1 = \mu_2, \quad H_a : \quad \mu_1 \neq \mu_2.$$

Der mit diesen Hypothesen beschriebene Test bezieht sich auf eine zweiseitige Fragestellung, das heißt, *betragsmäßig* große Abweichungen werden mit Ablehnung von  $H_0$  bestraft. Ähnlich wäre es möglich, Hypothesen zu überprüfen, die zum Gegenstand eine Gewichtszunahme (oder Gewichtsabnahme) durch den neuen Dünger haben:

$$H_0 : \quad \mu_1 = \mu_2, \quad H_a : \quad \mu_1 > \mu_2 \quad (H_a : \quad \mu_1 < \mu_2).$$

Die folgenden Tests beziehen sich auf *verbundene* beziehungsweise *unabhängige* Stichproben.

Eine verbundene Stichprobe ist dadurch charakterisiert, dass Paare von Stichprobenelementen gebildet werden, die gleichen oder zumindest ähnlichen Versuchsvoraussetzungen unterliegen. Man könnte sich also vorstellen, dass mehrere Paare von jeweils gleichgroßen Kartoffelpflanzen in einem Gewächshaus (also unter gleichen Bedingungen) zur Untersuchung des Gewichtes herangezogen wurden, wobei die beiden Pflanzen in jedem dieser Paare mit unterschiedlichen Düngern behandelt wurden.

Unabhängige Stichproben wären gegeben, falls von zwei Feldern, die jeweils mit den beiden Düngersorten gedüngt worden sind, Stichproben entnommen werden.

Bei verbundenen Stichproben haben beide Testreihen den gleichen Umfang, bei unabhängigen muss das nicht der Fall sein.

### **Vergleich zweier Mittelwerte, zweiseitige Fragestellung, verbundene Stichproben:**

Aus den normalverteilten Grundgesamtheiten  $X_1$ ,  $X_2$  liegen die verbundenen Stichproben

$$(x_1^1, x_2^1), (x_2^1, x_2^2), \dots, (x_n^1, x_n^2)$$

vor.

*Hypothesen:*

$$H_0 : \quad \mu_1 - \mu_2 = 0, \quad H_a : \quad \mu_1 - \mu_2 \neq 0.$$

*Testgrößen:*

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n (x_i^1 - x_i^2), \quad s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i^1 - x_i^2 - \bar{d})^2.$$

Unter Annahme von  $H_0$  ist

$$t = \frac{\bar{d}}{s_d} \sqrt{n}$$

die Realisierung einer  $T$ -verteilten Zufallsvariablen mit  $n - 1$  Freiheitsgraden.

*Kritischer Bereich:* Für gegebenes Signifikanzniveau  $\alpha > 0$  wird  $H_0$  abgelehnt, falls

$$|t| \geq t_{1-\frac{\alpha}{2}, n-1}.$$

**Beispiel:** Wir nehmen an, dass die untereinander stehenden Daten in (49) die beschriebenen Paare einer verbundenen Stichprobe bilden. Aus diesen Daten berechnet man  $|t| = 1.45$ . Da man für das Quantil der Student-Verteilung  $t_{0.975, 11}$  den Wert 2.2 erhält, wird die Nullhypothese nicht abgelehnt. Da der Wert  $\bar{d}$  die mittlere Abweichung der Differenz der Stichprobenpaare beschreibt, bedeutet das, dass diese Abweichung der arithmetischen Mittel  $\bar{x}^1, \bar{x}^2$  nicht signifikant groß ist, um  $H_0$  abzulehnen.

### Vergleich zweier Mittelwerte, zweiseitige Fragestellung, unabhängige Stichproben:

Aus den normalverteilten Grundgesamtheiten  $X_1, X_2$  liegen die unabhängigen Stichproben

$$x_1^1, \dots, x_{n_1}^1, \quad x_1^2, \dots, x_{n_2}^2$$

vor. Für  $X_1, X_2$  wird noch vorausgesetzt, dass sie die gleiche Varianz haben:  $\sigma_1 = \sigma_2$ . Diese Annahme über die Grundgesamtheiten muss natürlich mittels eines dafür geeigneten Tests geprüft werden. Ein Test, der für diese Hypothese geeignet ist, wird im folgenden Abschnitt beschrieben.

*Hypothesen:*

$$H_0 : \quad \mu_1 - \mu_2 = 0, \quad H_a : \quad \mu_1 - \mu_2 \neq 0.$$

*Testgrößen:* Unter der Hypothese  $H_0$  ist

$$(50) \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_{x_1}^2 + (n_2-1)s_{x_2}^2}{n_1+n_2-2}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

Realisierung einer Student-verteilten Zufallsvariable mit  $m = n_1 + n_2 - 2$  Freiheitsgraden. In der obigen Formel beschreiben  $\bar{x}_1, \bar{x}_2$  die arithmetischen Mittel der beiden Stichproben. Entsprechend sind  $s_{x_1}^2, s_{x_2}^2$  die üblichen erwartungstreuen Schätzungen für die Varianz, siehe 8.4.

*Kritischer Bereich:* Für gegebenes Signifikanzniveau  $\alpha > 0$  wird  $H_0$  abgelehnt, falls

$$|t| \geq t_{1-\frac{\alpha}{2}, n_1+n_2-2}.$$

Bei Ungleichheit der Varianz der beiden normalverteilten Zufallsvariablen  $X_1, X_2$  wird im Unterschied zu (50) die folgende Testgröße berechnet:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2}}}.$$

Diese ist näherungsweise Student-verteilt. Der kritische Bereich wird festgelegt durch

$$|t| \geq t_{1-\frac{\alpha}{2}, m},$$

wobei  $m$  die größte ganze Zahl ist, die kleiner oder gleich

$$M = \frac{1}{\frac{c^2}{n_1-1} + \frac{(1-c)^2}{n_2-1}} \quad \text{mit } c = \frac{\frac{s_{x_1}^2}{n_1}}{\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2}}.$$

### 3. Statistische Tests für die Varianz

Genauso, wie man an Verfahren interessiert ist, die die Überprüfung von Hypothesen über den Mittelwert zum Inhalt haben, besteht häufig auch die Aufgabe, Hypothesen über die Varianz zu testen. Speziell waren Tests über den Mittelwert an gewisse Voraussetzungen über die Varianz gebunden. So mussten beim Test bezüglich des Vergleiches des Mittelwertes bei unabhängigen Stichproben die Voraussetzung erfüllt sein, dass die Varianzen der Zuallsvariablen  $X_1, X_2$  gleich sind:  $\sigma_1 = \sigma_2$ . Häufig sind solche Beziehungen nicht *a priori* gegeben. Sie müssen durch einen Test überprüft werden.

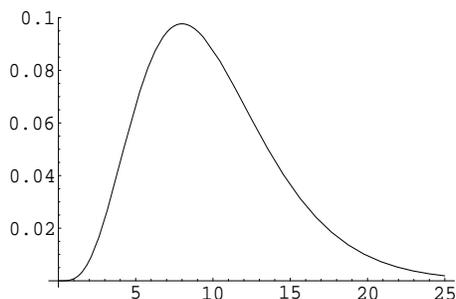


ABBILDUNG 1. Dichtefunktion der  $\chi^2$ -Verteilung mit 10 Freiheitsgraden.

Ein wesentlicher Unterschied zu den oben behandelten Tests, die auf der Student-Verteilung basieren, beruhen die Tests bezüglich der Streuung häufig auf der  $\chi^2$ -Verteilung beziehungsweise die Fisher-Verteilung.

#### Test einer Hypothese über die Varianz einer Grundgesamtheit

Es sei  $X$  eine normalverteilte Grundgesamtheit. Die Varianz der Grundgesamtheit werde mit  $D^2X = \sigma^2$  bezeichnet. Die in der Hypothese formulierte *Sollvarianz* sei  $\sigma_0$ . Es werde eine Stichprobe

$$x_1, x_2, \dots, x_n$$

ermittelt. Weiterhin sei das Signifikanzniveau  $\alpha \in (0, 1)$  gegeben. Wir werden die folgenden beiden Fälle untersuchen.

- (1) Der Erwartungswert der Grundgesamtheit  $X$  werde als bekannt angenommen:  $\mathbb{E}X = \mu_0$ .
- (2) Der Erwartungswert der Grundgesamtheit  $X$  sei unbekannt.

*Hypothesen:* Es werden zweiseitige als auch einseitige Tests formuliert.

- (i)  $H_0 : \sigma = \sigma_0, \quad H_a : \sigma \neq \sigma_0$

$$(ii) H_0 : \sigma = \sigma_0, \quad H_a : \sigma > \sigma_0$$

$$(iii) H_0 : \sigma = \sigma_0, \quad H_a : \sigma < \sigma_0.$$

*Testgrößen:*

Wir betrachten die folgenden Größen:

$$(1) \quad \chi^2 = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma_0^2} \quad \text{ist } \chi_n^2 \text{ - verteilt,}$$

$$(2) \quad \chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad \text{ist } \chi_{n-1}^2 \text{ - verteilt.}$$

*Kritischer Bereich:*

(i)

$$\begin{aligned} \chi^2 &\geq \chi_{1-\frac{\alpha}{2},n}^2 && \text{oder} && \chi^2 \leq \chi_{\frac{\alpha}{2},n}^2 && \text{für Fall (1)} \\ \chi^2 &\geq \chi_{1-\frac{\alpha}{2},n-1}^2 && \text{oder} && \chi^2 \leq \chi_{\frac{\alpha}{2},n-1}^2 && \text{für Fall (2)} \end{aligned}$$

(ii)

$$\chi^2 \geq \begin{cases} \chi_{1-\alpha,n}^2 & \text{für Fall (1)} \\ \chi_{1-\alpha,n-1}^2 & \text{für Fall (2)} \end{cases}$$

(iii)

$$\chi^2 \leq \begin{cases} \chi_{\alpha,n}^2 & \text{für Fall (1)} \\ \chi_{\alpha,n-1}^2 & \text{für Fall (2)}. \end{cases}$$

Dabei bezeichnet  $\chi_{\alpha,n}$  das Quantil der  $\chi^2$ -Verteilung mit  $n$  Freiheitsgraden.

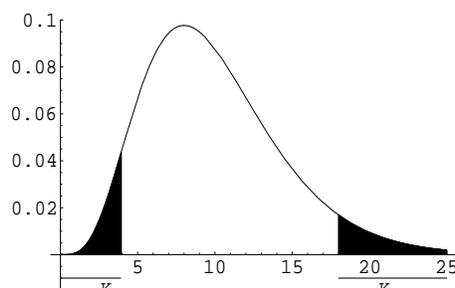


ABBILDUNG 2. Kritischer Bereich eines zweiseitigen Tests der Varianz für  $\alpha = 0.1$

**Beispiel:** Die Breite von Briefumschlägen, die von einem Automaten mit Werbebriefen gefüllt werden, ist normalverteilt. Um Ausfälle des Automaten zu minimieren, dürfen die Briefumschläge nicht zu groß sein. Die Firma, die die Umschläge produziert gibt an, dass die Varianz der Umschläge  $2\text{mm}^2$  beträgt. Es liegt folgende Stichprobe vor:

10.25, 11.32, 9.14, 13.59, 8.01, 10.65, 9.06, 8.43, 13.62, 10.07,  
6.71, 9.51, 9.77, 9.51, 9.87, 10.36, 8.87, 9.56, 13.43, 10.28, 14.02,  
11.25, 9.45, 9.12, 10.42.

Das Signifikanzniveau betrage  $\alpha = 0.05$ . Da wir nur Abweichungen nach oben als signifikant erachten, wählen wir die alternative Hypothese  $\sigma > \sigma_0$ . Der Erwartungswert  $\mu_0$  kann nicht als bekannt angenommen werden.

Die erwartungstreue Schätzung für die Varianz ergibt:

$$s^2 = \frac{1}{25-1} \sum_{i=1}^{25} (x_i - \bar{x})^2 = 3.27.$$

Daraus errechnen wir die Testgröße

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = 39.24$$

Da wir für das Quantil  $\chi_{1-\alpha, n-1}^2 = \chi_{0.95, 24}^2$  den Wert 36.4 erhalten, müssen wir  $H_0$  ablehnen.

### Test auf Gleichheit zweier Varianzen

Um einen Test auf Gleichheit zweier Varianzen durchzuführen, wird eine neue Verteilung, die *Fisher* oder *F*-Verteilung benötigt:

Es seien  $X_1 \sim N(\mu_1, \sigma_1)$ ,  $X_2 \sim N(\mu_2, \sigma_2)$  zwei normalverteilte Zufallsvariablen bezüglich derer die folgenden Stichproben vorliegen

$$x_1^1, \dots, x_{n_1}^1, \quad x_1^2, \dots, x_{n_2}^2.$$

Das Signifikanzniveau betrage  $\alpha$ .

*Hypothesen:*

Wir betrachten die beiden Paare von Hypothesen

$$H_0: \sigma_1 = \sigma_2, \quad H_a: \sigma_1 \neq \sigma_2.$$

beziehungsweise

$$H_0: \sigma_1 = \sigma_2, \quad H_a: \sigma_1 > \sigma_2.$$

*Testgröße:*

Unter Annahme von  $H_0$  gilt, ist der Wert

$$f = \frac{s_1^2}{s_2^2}$$

eine Realisierung einer *F*-verteilten Zufallsvariablen mit  $m_1 = n_1 - 1$ ,  $m_2 = n_2 - 1$  Freiheitsgraden.

*Kritischer Bereich:*

$$f \leq f_{\frac{\alpha}{2}, m_1, m_2} \quad \text{oder} \quad f \geq f_{1-\frac{\alpha}{2}, m_1, m_2}$$

beziehungsweise

$$f \geq f_{1-\alpha, m_1, m_2}.$$

Abschließend sei noch bemerkt, dass die *F*-Verteilung die folgende Symmetrieeigenschaft für die Quantile aufweist:

$$f_{\alpha, m_1, m_2} = f_{1-\alpha, m_2, m_1}^{-1}.$$

#### 4. Das Testen einer Verteilung

Im ersten Teil dieses Abschnittes wollen wir Verfahren angeben, die erlauben eine Hypothese zu überprüfen, dass eine Grundgesamtheit  $X$  eine spezielle Verteilung besitzt. Viele oben beschriebene Tests wurden unter der Voraussetzung formuliert, dass die zugrunde liegende Grundgesamtheit normalverteilt ist. Das Verfahren, das wir kennenlernen, erlaubt es, diese Hypothese zu überprüfen.

Im zweiten Teil werden wir ein Verfahren kennenlernen, das den Vergleich der Verteilung von zwei Grundgesamtheiten erlaubt.

##### Der $\chi^2$ -Anpassungstest

Wir testen die Verteilung einer Grundgesamtheit  $X$  mit Verteilungsfunktion  $F(x)$ . Wir betrachten dabei die *Sollverteilung* einer Zufallsvariablen  $X_0$ , die durch ihre Verteilungsfunktion  $F_0(x)$  beschrieben wird. Desweiteren liegt eine Stichprobe

$$x_1, x_2, \dots, x_n$$

vor. Wir unterteilen weiterhin den Definitionsbereich von  $F_0(x)$  in  $m + 1$  Klassen

$$K_1 = (-\infty, k_1], \quad K_2 = (k_1, k_2], \quad \dots \quad K_{m+1} = (k_m, \infty).$$

Für jede Klasse  $K_j$  wird die Anzahl  $H_j$  der Stichprobenelemente  $x_i$  bestimmt, die in dieser Klasse liegen. Weiterhin wird die theoretische Häufigkeit  $P_j$  gebildet. Diese Größen sind wie folgt definiert:

$$P_j = (F_0(k_j) - F_0(k_{j-1}))n = \mathbb{P}(X_0 \in (k_{j-1}, k_j])n \quad \text{für } j = 1, \dots, m + 1,$$

wobei  $k_0 = -\infty$  und  $k_{m+1} = \infty$  gesetzt wird. Der Wert  $P_j$  ist gleich der Wahrscheinlichkeit, dass die Zufallsvariablen  $X_0$  in einem Intervall  $(k_{j-1}, k_j]$  liegen, multipliziert mit dem Umfang der Stichprobe. Sind nun die Differenzen zwischen den Häufigkeiten  $H_j$  und den theoretischen Häufigkeiten  $P_j$ , mit einem speziellen Maß gemessen, groß, so wird davon ausgegangen, dass die Verteilung der Zufallsvariablen  $X$  nicht  $F_0$  sein kann.

Aus Gründen, die hier nicht angeführt werden sollen, dürfen die Klassen nicht zu klein gewählt werden. Als Faustregel gilt, dass die Werte  $P_j \geq 5$  gelten sollten. Ist das bei einer Klassenunterteilung nicht der Fall, so sollten benachbarte Klassen zusammengelegt werden, bis diese Forderung erfüllt ist.

*Hypothese:*

$$H_0 : F = F_0. \quad H_a : F \neq F_0.$$

*Testgröße:*

Unter der Annahme von  $H_0$  ist der Wert

$$\chi^2 := \sum_{j=1}^{m+1} \frac{(H_j - P_j)^2}{P_j}$$

die Realisierung einer näherungsweise  $\chi^2$ -verteilten Zufallsvariablen mit  $r$  Freiheitsgraden. Die Anzahl der Freiheitsgrade ergibt sich aus der Anzahl der Klassen, also  $m + 1$  um eins reduziert. Müssen zur Beschreibung von  $F_0$  noch weitere Parameter geschätzt werden, dann verringert sich die Anzahl der Freiheitsgrade noch um die Anzahl der zu schätzenden Parameter. Ist  $X_0$  normalverteilt mit den Parametern  $\mu$  und  $\sigma$  und sind diese unbekannt und müssen noch aus der Stichprobe geschätzt werden, dann ist  $r = (m + 1) - 1 - 2$ .

*Kritischer Bereich:*

$H_0$  wird abgelehnt, falls

$$\chi^2 \geq \chi_{1-\alpha, r}^2$$

gilt, wobei  $\chi_{1-\alpha, r}^2$  das  $\alpha$ -Quantil einer  $\chi^2$ -verteilten Zufallsvariablen ist mit  $r$  Freiheitsgraden.

**Beispiel:** Es soll getestet werden, ob die Lebensdauer eines einfachen Aggregates exponentialverteilt ist. Die Verteilungsfunktion ist gegeben durch

$$F_0(x) = \begin{cases} 0 & : x \leq 0 \\ 1 - e^{-\lambda_0 x} & : x > 0. \end{cases}$$

$\lambda_0$  wird durch eine Stichprobe vom Umfang  $n = 120$  geschätzt durch

$$\lambda_0 = (\bar{x})^{-1},$$

siehe Abschnitt 3, wobei sich der Wert  $\lambda_0 = 0.009$  ergibt. Die folgende Klassenunterteilung wurde festgelegt.

$$K_1 = [0, 22] \quad K_2 = (22, 60] \quad K_3 = (60, 100] \quad K_4 = (100, 180] \quad K_5 = (180, \infty).$$

Da diese Wahrscheinlichkeitsverteilung auf  $\mathbb{R}^+$  beschränkt ist, kann  $k_0 = 0$  gesetzt werden. Aus der Stichprobe werden folgende Häufigkeiten  $H_j$  berechnet

$$H_1 = 20 \quad H_2 = 31 \quad H_3 = 23 \quad H_4 = 24 \quad H_5 = 22$$

und als theoretische Häufigkeiten kann man mit Hilfe der Verteilungsfunktion von  $F_0$  berechnen

$$P_1 = 21.56 \quad P_2 = 28.51 \quad P_3 = 21.15 \quad P_4 = 25.05 \quad P_5 = 23.73.$$

Damit ergibt sich

$$\chi^2 = 0.7.$$

Da ein Parameter ( $\lambda_0$ ) geschätzt wurde, ist  $r = 5 - 1 - 1$  und somit

$$\chi_{0.9, 3} = 6.25.$$

Der Wert  $\chi^2$  ist nicht im kritischen Bereich. Die Hypothese über die Verteilung muss also nicht verworfen werden.

Im Folgenden wollen wir *Rangsummentests* kennenlernen, mit denen Hypothesen über die Gleichheit zweier Verteilungen geprüft werden können. Wir wollen im Folgenden die Idee erläutern. Es soll zum Beispiel das Wachstum von Pflanzen unter dem Einfluss unterschiedlicher Dünger untersucht werden. Angenommen zwei der Pflanzen sind mit der Düngersorte A und drei der Pflanzen mit der Düngersorte B behandelt worden. Wir schreiben die folgende Liste auf:

$$A, B, A, B, B.$$

Diese Liste beschreibt die Reihenfolge der Größen der einzelnen Pflanzen mit dem kleinsten Wert beginnend. Die kleinste Größe besitzt eine Pflanze, die mit dem Dünger A behandelt wurde. Es ist also zu erkennen, dass tendenziell die mit Dünger A behandelte Pflanzen im unteren Bereich der Liste anzutreffen sind, während sich die mit B behandelten Pflanzen im oberen der Liste befinden. Es gibt aber eine Pflanze, die mit A behandelt wurde und größer ist als eine mit B behandelte Pflanze. Die Frage, die beantwortet werden muss ist nun, ob die Symbole A und B in der Liste gut gemischt sind, was für die Gleichheit der Größenverteilung der beiden Pflanzen sprechen würde oder ob sich ein Symbol am Anfang oder am Ende der

Liste konzentriert. Im Falle einer guten *Durchmischung* der Symbole  $A$ ,  $B$  kann die Hypothese

$H_0$  : die Verteilungen der Zufallsgrößen beider Gruppen sind gleich

nicht abgelehnt werden.

Die folgende Tabelle beschreibt die möglichen Rangplätze beider Gruppen. Die letzte Spalte beschreibt die Rangsummen  $R_1$  der Gruppe A.

Rangplatz	1	2	3	4	5	Rangsumme A
1	A	A	B	B	B	3
2	A	B	A	B	B	4
3	B	A	A	B	B	5
4	A	B	B	A	B	5
5	A	B	B	B	A	6
6	B	A	B	A	B	6
7	B	B	A	A	B	7
8	B	A	B	B	A	7
9	B	B	A	B	A	8
10	B	B	B	A	A	9

TABELLE 2. Aufistung der Rangsummen

Die Anordnungen von der dritten bis zur achten Spalte können als gut gemischt angesehen werden, so dass bei einer Rangsumme zwischen 5 und 7 keine Gründe für die Ablehnung von  $H_0$  vorliegen.

Unter der Annahme, dass die Zusammenstellungen *gleichberechtigt* sind, läßt sich mittels der klassischen Wahrscheinlichkeit die Verteilung der Rangsummen  $R_1$  angeben. Es ist nun üblich eine Zufallsvariable anzugeben, die die Rangsummen in etwas transformierter Form beschreibt:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1.$$

Die Verteilung dieser Zufallsvariable wird als  $U$ -Verteilung bezeichnet. Diese Verteilung ist diskret. Die natürlichen Zahlen  $n_1$ ,  $n_2$  beschreiben die Anzahl der Elemente in den Gruppen A, B. Der Wertevorrat von  $U_1$  ist die Mengen der natürlichen Zahlen zwischen 0 und  $n_1 n_2$ .

### Der $U$ -Test

*Hypothese:*

$H_0$  : Die Verteilung beider Klassen ist gleich,

$H_a$  : Die Verteilung beider Klassen ist nicht gleich.

*Testgröße:*

$$u_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - r_1,$$

wobei  $r_1$  eine Realisierung der Rangsummen der ersten Gruppe ist.

*Kritischer Bereich:* Aus der Verteilung von  $U_1$  können die kritischen Werte  $u_{\frac{\alpha}{2}}$  und  $u_{1-\frac{\alpha}{2}}$  zu einem Signifikanzniveau  $\alpha$  entnommen werden.  $H_0$  wird abgelehnt, falls

$$u_1 \leq u_{\frac{\alpha}{2}} \quad \text{oder} \quad u_1 \geq u_{1-\frac{\alpha}{2}}.$$

Die Berechnung der Quantile ist eingebettet in der rechen-technischen Implementierung des  $U$ -Testes in statistischer Standardsoftware. Desweiteren findet man die kritischen Werte in Tabellen, siehe zum Beispiel Kähler [9], Sachs [18].

**Beispiel:** Jeweils 10 Schüler aus Deutschland (A) und Belgien (B) schreiben eine Testklausur über den Abiturstoff in Mathematik. Es wird die folgende Rangfolge ermittelt:

B, A, A, A, A, B, B, A, B, A, A, B, A, B, B, A, B, A, B, B.

Es soll die Hypothese überprüft werden, ob die Leistungsverteilung in beiden Klassen gleich ist. Wir erhalten als Rangsumme  $r_1 = 89$  und den zugehörigen  $U_1$ -Wert 66. Da für das Signifikanzniveau  $\alpha = 0.05$  die kritischen Werte 23 und 77 betragen, muss  $H_0$  nicht abgelehnt werden.

Der  $U$ -Test ist für unabhängige Stichproben anwendbar. Für verbundene Stichproben empfiehlt sich der Wilcoxon Test.

Abschließend soll noch bemerkt werden, dass bei hinreichend großen Werten für  $n_1, n_2$  die Zufallsvariable  $U_1$  näherungsweise normalverteilt ist:

$$U_1 \sim N\left(\frac{n_1 n_2}{2}, \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}\right).$$

## 5. Der $\beta$ -Fehler

Wir hatten schon im ersten Abschnitt dieses Kapitels den  $\alpha$ - und  $\beta$ -Fehler kennengelernt. Während der  $\alpha$ -Fehler genutzt wurde, um den kritischen Bereich zu konstruieren, wollen wir jetzt die Bedeutung des  $\beta$ -Fehlers studieren. Dieser Fehler beschreibt die Wahrscheinlichkeit, dass  $H_0$  angenommen wird, obwohl  $H_a$  zutrifft. Wir werden uns im folgenden nur auf den Test des Mittelwertes einer normalverteilten Grundgesamtheit  $X \sim N(\mu, \sigma_0)$  mit bekannter Varianz  $\sigma_0$  beziehen. Wir haben also die Hypothesen:

$$H_0 \quad \text{trifft zu: } \mathbb{E}X = \mu_0, \quad H_a \quad \text{trifft zu: } \mathbb{E}X = \mu_a,$$

wobei  $\mu_0 \neq \mu_a$  ist. Der  $\beta$ -Fehler trifft zu, wenn aufgrund der Variabilität der Daten sich aus der Stichprobe ein arithmetisches Mittel  $\bar{x}$  ergibt, das nicht im kritischen Bereich liegt (also keine Ablehnung von  $H_0$ ), obwohl  $\mu_a$  der tatsächliche Mittelwert der Grundgesamtheit ist. Der  $\beta$ -Fehler hängt unter anderem von dem  $\alpha$ -Fehler,  $\mu_a, \mu_0$  und dem Stichprobenumfang  $n$  ab.

Man stelle sich vor, man teste die Hypothese  $H_0 : \mathbb{E}X = \mu_0$  gegen die Hypothese  $H_a : \mathbb{E}X = \mu_a$ , wobei  $\mu_a > \mu_0$  ist. Unter der Annahme, dass die Grundgesamtheit normalverteilt ist, erhält man bei bekannter Varianz für die kritischen Bereiche

$$K = \left[\mu_0 + y_{1-\alpha} \frac{\sigma_0}{\sqrt{n}}, \infty\right)$$

bei einem vorgegebenen Fehler 1. Art  $\alpha$ . Man kann nun die Wahrscheinlichkeit, einen Fehler 1. Art zu begehen beliebig klein machen, indem man  $\alpha$  gegen Null konvergieren läßt. Dadurch vergrößert sich der Akzeptanzbereich  $\bar{K}$  für  $H_0$ , der

kritische Bereich wird kleiner. Durch die Vergrößerung des Akzeptanzbereiches vergrößert sich auch der  $\beta$ -Fehler. Man bezahlt also für die Verkleinerung des  $\alpha$ -Fehlers mit der Vergrößerung des  $\beta$ -Fehlers.

Die im Folgenden definierte Größe *Trennschärfe* beschreibt den Zusammenhang zwischen  $\alpha$ -Fehler und  $\beta$ -Fehler.

DEFINITION 9.1. Die *Trennschärfe eines Tests* ist definiert als  $1 - \beta$ :

$$T = T(\alpha, \mu_0, \sigma_0, \mu_a, n) := 1 - \beta.$$

Wie wir noch im folgenden Beispiel sehen werden, hängt die Trennschärfe von dem Bruch  $(\mu_0 - \mu_a)/\sigma_0$  ab. Die Funktion, die als Argument diesen Bruch hat und die Werte  $1 - \beta$  besitzt, wird als Gütefunktion bezeichnet. Die Funktion

1 – Gütefunktion

ist die Operationscharakteristik eines Tests, siehe Abbildung 3.

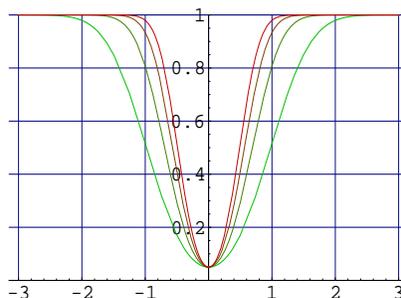


ABBILDUNG 3. Trennschärfekurven für verschiedene  $\alpha$  und  $n$  Werte, wobei auf der Abszisse  $\frac{\mu_0 - \mu_a}{\sigma_0}$  abgetragen wird.

Bei vorgegebenem  $\alpha$  und  $(\mu_0 - \mu_a)/\sigma_0$  sollte der Umfang der Stichprobe so gewählt werden, dass die Trennschärfe über 0.75 liegt.

Die Trennschärfe vergrößert sich ( $\beta$  verkleinert sich), falls der Stichprobenumfang  $n$  vergrößert wird, siehe 4, Abbildung 5.

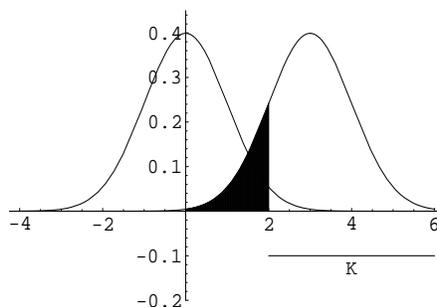


ABBILDUNG 4. Die linke Kurve ist die Dichte bezüglich  $\mu_0$ , die rechte Kurve ist die Dichte bezüglich  $\mu_a$ . Der schwarzen Fläche entspricht der  $\beta$ -Fehler.

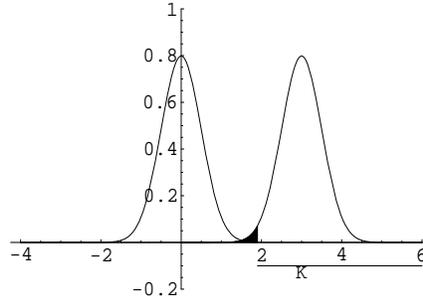


ABBILDUNG 5. Bei Vergrößerung von  $n$  wird die schwarze Fläche kleiner.

Es soll nun die Trennschärfe für einen einseitigen Test des Mittelwertes bei bekannter Varianz  $\sigma_0$  berechnet werden, wobei die Grundgesamtheit als normalverteilt angenommen wird. Wir nehmen an, dass die folgenden Hypothesen formuliert sind:

$$H_0 : \mu_0 = 20 \quad H_a : \mu_a = 24.$$

Der Stichprobenumfang beträgt  $n = 25$ . Wir gehen davon aus, dass  $\sigma_0 = 10$  ist und nehmen den  $\alpha$ -Fehler  $\alpha = 0.05$  an. Wir erhalten das zugehörige Quantil der Standardnormalverteilung  $y_{1-\alpha} = 1.645$ . Da die Testgröße durch das arithmetische Mittel gegeben ist, benötigen wir die Verteilung der Stichprobenfunktion  $\bar{X}$ . Diese Zufallsvariable ist unter der Annahme von  $H_a$ , das heißt, der Annahme, dass die Grundgesamtheit den Mittelwert  $\mu_a$  besitzt,  $N(\mu_a, \sigma_0/\sqrt{n})$  standardnormalverteilt, siehe (36). Es sei der für (48) festgelegte kritische Bereich.

$$K = [\mu_0 + y_{1-\alpha} \frac{\sigma_0}{\sqrt{n}}, \infty).$$

Der  $\beta$ -Fehler tritt auf, wenn das aus der Stichprobe berechnete arithmetische Mittel  $\bar{x}$  im Komplement  $\bar{K}$  des kritischen Bereiches, also im Akzeptanzbereich für  $H_0$ , liegt:

$$\bar{x} \in \bar{K}.$$

Durch Transformation auf Standardnormalverteilung, siehe Lemma 4.14, (18) erhalten wir

$$\begin{aligned} \beta &= \mathbb{P}(\bar{X} \in \bar{K}) = \mathbb{P}(\bar{X} < \mu_0 + y_{1-\alpha} \frac{\sigma_0}{\sqrt{n}}) \\ &= \Phi\left(\frac{\mu_0 + y_{1-\alpha} \frac{\sigma_0}{\sqrt{n}} - \mu_a}{\frac{\sigma_0}{\sqrt{n}}}\right) \\ &= \Phi\left(\frac{\mu_0 - \mu_a}{\sigma_0} \sqrt{n} + y_{1-\alpha}\right) = \Phi(-0.355). \end{aligned}$$

Somit erhalten wir für die Trennschärfe

$$T = 1 - \beta = 1 - \Phi(-0.355) = \Phi(0.355) = 0.64.$$

Der Stichprobenumfang ist also nicht ausreichend, um eine Trennschärfe von 0.75 zu gewährleisten.



## Regressionsanalyse

### 1. Einleitung

Wir nehmen an, dass eine Menge von Daten gegeben ist, die durch Messungen ermittelt wurden:

$$\begin{aligned}(x_{11}, x_{12}, \dots, x_{1k}, y_1) &= (\vec{x}_1, y_1) \\ (x_{21}, x_{22}, \dots, x_{2k}, y_2) &= (\vec{x}_2, y_2) \\ \dots & \\ \dots & \\ (x_{n1}, x_{n2}, \dots, x_{nk}, y_n) &= (\vec{x}_n, y_n).\end{aligned}$$

Gesucht ist eine Funktion mit  $k$  Veränderlichen, die diese Messwerte näherungsweise beschreibt:

$$y = y(x_1, \dots, x_k).$$

Die gemessenen  $y$ -Werte sind mit Messfehlern behaftet. Das gleiche kann natürlich bei der Einstellung der Messstellen  $x_i$  auftreten. Dieser Fall soll jetzt aber außer acht gelassen werden und erst später diskutiert werden. Da die  $y$ -Werte als fehlerhaft angenommen werden, wird es nicht sinnvoll sein, die Funktion  $y(\vec{x})$  so zu konstruieren, dass der Graph der Funktion alle Messwerte enthält. So eine Funktion hätte einen sehr stark schwankenden Verlauf und würde nicht das Wesentliche bezüglich der Abhängigkeit der  $x$ - und  $y$ -Werte wiedergeben. Es scheint also sinnvoller zu sein, die Funktion *ausgleichend* zwischen die Messwerte zu legen.

In [?] sind folgende Daten gegeben, die aus Messungen resultieren, bei denen die Verkehrsdichte und die Geschwindigkeit ermittelt wurden.

Werden diese Messdaten in ein Koordinatensystem eingetragen, so ergibt sich die *Punkt看ke* in Abbildung 1:

Mittels der *Zweipunktegleichung* kann man eine Gerade bestimmen, die mehr oder weniger gut die Punkt看ke beschreibt. Wählt man zum Beispiel die Punkte

$$(100, 3.6) \quad (30, 5.5),$$

die in der Punkt看ke liegen, so erhält man den in Abbildung (1) dargestellten Graphen der linearen Funktion

$$y(x) = 6.3 - 0.027x.$$

Die so ermittelte Funktion hat allerdings den Nachteil, dass nicht alle Punkte mit in die Berechnung eingegangen sind.

Weiterhin sieht man, dass die Punkt看ke eine leicht nach oben gebogene Form hat. Das bedeutet, eine lineare Funktion wird keine optimale Beschreibung der

Dichte $x_i$	$\sqrt{v} y_i$	Dichte $x_i$	$\sqrt{v} y_i$
20.4	6.229	29.5	5.639
27.4	5.612	30.8	5.621
106.2	3.256	26.5	5.831
80.4	4.012	35.7	5.376
141.3	2.775	30.0	5.367
130.9	2.881	106.2	3.240
121.7	2.915	97.0	3.507
106.5	3.332	90.1	3.633
130.5	2.933	106.7	3.376
101.1	3.332	99.3	3.347
144.2	2.793	109.1	3.376
123.9	3.130	107.2	3.209

TABELLE 1. Verkehrsdichte und Wurzel aus der Geschwindigkeit

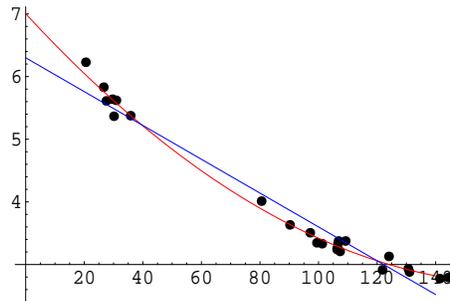


ABBILDUNG 1. Punktwolke und deren Approximation durch eine lineare und quadratische Funktion

Abhängigkeit der  $y$ -Werte von den  $x$ -Werten liefern. Berechnet man die *Residuen*  $e_i$ , das heißt, die Differenzen zwischen den Messwerten  $y_i$  und den Werten der berechneten Funktion an der Stelle  $y(x_i)$

$$e_i = y_i - y(x_i),$$

so erhält man die Abbildung (2):

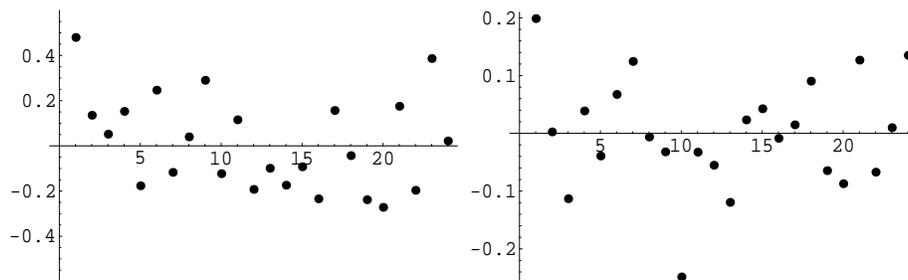


ABBILDUNG 2. Residuen bezüglich der linearen und quadratischen Approximation

Man erkennt, dass für kleine und große  $x$ -Werte die Residuen positiv sind während die Residuen zu  $x$ -Werten mittleren Bereich negativ sind. Dieses Verhalten der Residuen unterstreicht noch einmal, dass das von uns berechnete lineare Modell nicht besonders gut geeignet ist, um die Punktwolke zu beschreiben.

Wir werden im übernächsten Abschnitt eine Methode kennenlernen, die ermöglicht ein Polynom zu konstruieren, das gut in die Punktwolke hineinpaßt: Dieses Polynom ist gegeben durch

$$y = 7 - 0.005x + 0.00015x^2,$$

siehe Abbildung 1.

## 2. Differentiation von Matrizen

Als Vorbereitung auf den nächsten Abschnitt wollen wir hier ein paar Hilfsbetrachtungen einfügen, die sich auf die Differentiation von Matrizen beziehen.

Wir betrachten den Vektor

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}.$$

Gegeben sei weiterhin die reelle Funktion mit  $k + 1$  Veränderlichen

$$f(\vec{\beta}) = f(\beta_0, \beta_1, \dots, \beta_k).$$

Unter der Voraussetzung, dass  $f$  differenzierbar ist, kann der Gradient berechnet werden:

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} := \begin{pmatrix} \frac{\partial f(\vec{\beta})}{\partial \beta_0} \\ \vdots \\ \frac{\partial f(\vec{\beta})}{\partial \beta_k} \end{pmatrix} = \nabla f(\vec{\beta}).$$

Für einen  $k + 1$ -dimensionalen Vektor

$$\vec{c} = \begin{pmatrix} c_0 \\ \vdots \\ c_k \end{pmatrix}$$

definieren wir die Funktion

$$f(\vec{\beta}) = \sum_{i=0}^k c_i \beta_i.$$

Dieser Ausdruck entspricht dem Skalarprodukt zwischen den Vektoren  $\vec{c}$  und  $\vec{\beta}$ . Interpretiert man den transponierten Vektor  $\vec{c}^T$  als Zeilenmatrix und  $\vec{\beta}$  als Spaltenmatrix, so kann das berechnete Skalarprodukt durch Matrixmultiplikation gebildet werden:

$$f(\vec{\beta}) = \vec{c}^T \vec{\beta} = \vec{\beta}^T \vec{c}.$$

Für den Gradienten erhalten wir

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = \vec{c} = \begin{pmatrix} c_0 \\ \vdots \\ c_k \end{pmatrix}.$$

Es sei nun  $A$  eine *symmetrische* Matrix mit  $k + 1$  Spalten und Zeilen:  $A = (a_{i,j})_{i,j=0,\dots,k}$ . Bezeichne  $a_0, a_1, \dots, a_k$  die Spalten(vektoren) der Matrix, dann gilt wegen der Symmetrie

$$A = (a_0, \dots, a_k) = \begin{pmatrix} a_0^T \\ \vdots \\ a_k^T \end{pmatrix}.$$

Wir betrachten die Funktion

$$(51) \quad f(\vec{\beta}) = \vec{\beta}^T A \vec{\beta}.$$

LEMMA 10.1. *Es gilt für die in (51) definierte Funktion*

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = 2A\vec{\beta}.$$

BEWEIS. Wird die in dieser Formel auftretende Matrixmultiplikation ausgeführt, so ergibt sich

$$\begin{aligned} f(\vec{\beta}) &= \sum_{i=0}^k \beta_i \left( \sum_{j=0}^k a_{i,j} \beta_j \right) = \sum_{i,j=0}^k a_{i,j} \beta_i \beta_j \\ &= \sum_{i=j} a_{i,i} \beta_i^2 + \sum_{i \neq j} a_{i,j} \beta_i \beta_j = \sum_{i=j} a_{i,i} \beta_i^2 + 2 \sum_{i < j} a_{i,j} \beta_i \beta_j, \end{aligned}$$

wobei die letzte Gleichung aus der Symmetrie von  $A$  folgt.

Damit erhalten wir für die Ableitung:

$$\frac{\partial f(\vec{\beta})}{\partial \beta_0} = 2a_{0,0}\beta_0 + 2 \sum_{j=1}^k a_{0,j}\beta_j = 2a_0^T \vec{\beta}.$$

Ähnlich erhalten wir die Ableitungen nach  $\beta_1, \dots, \beta_k$ . Fassen wir diese Ableitungen zusammen, so gilt:

$$\frac{\partial f(\vec{\beta})}{\partial \vec{\beta}} = 2A\vec{\beta}.$$

□

Zum Abschluß sei noch bemerkt, dass diese Formel nur für symmetrische Matrizen richtig ist.

### 3. Die Methode der kleinsten Quadrate

Wir wollen nun ein Kriterium festlegen, das erfüllt sein muss, damit eine Funktion *optimal* in eine durch Messpunkte gegebene hineinpaßt. Wie dieses Kriterium wirkt, machen wir uns an einer linearen Funktion klar. Später werden wir unser Vorgehen auf andere Klassen von Funktionen verallgemeinern.

Gegeben seien die Messwertpaare

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Aufgrund der Lage dieser Messpunkte entscheiden wir uns dafür, die Punktwolke durch eine Gerade zu approximieren. Wir wählen somit das *lineare Modell*

$$y = \beta_0 + \beta_1 x.$$

Die Koeffizienten  $\beta_0, \beta_1$  sind unbekannt und müssen berechnet werden. Zwischen den Vektoren

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

und allen linearen Funktionen gibt es eine eindeutige Zuordnung. Ist zum Beispiel so ein Vektor  $\vec{\beta}$  gegeben, so sind Anstieg und Absolutglied der Geraden bekannt. Umgekehrt, jeder Geraden, die nicht senkrecht auf der  $x$ -Achse steht, kann ein Wert für den Anstieg  $\beta_1$  und das Absolutglied  $\beta_0$  zugeordnet werden.

Die Gerade, die die Punktwolke am besten beschreibt, wird nach folgendem Kriterium bestimmt:

**Gesucht ist die Gerade  $y(x) = \beta_0 + \beta_1 x$ , so dass die Summe der Quadrate der Abstände zwischen Messwerten  $y_i$  und den Funktionswerten  $y(x_i)$  minimal ist.**

Später werden wir das hier noch sehr spezielle Modell der Geraden durch allgemeinere Modelle ersetzen. Das Prinzip der kleinsten Quadrate wird auch dafür gültig bleiben.

Das Prinzip der kleinsten Quadrate besagt nun, dass die Minimalstelle der Funktion

$$(52) \quad S(\vec{\beta}) = S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

die Koeffizienten  $\hat{\beta}_0, \hat{\beta}_1$  liefert, so dass  $y = \hat{\beta}_0 + \hat{\beta}_1 x$  die Punktwolke am besten beschreibt.

Wir werden nun die Minimalstellen für allgemeinere Modelle berechnen. Wir nehmen an, dass wir  $n$  verschiedene  $k + 1$  Tupel von Messwerten haben.

$$\begin{aligned} (x_{11}, x_{12}, \dots, x_{1k}, y_1) &= (\vec{x}_1, y_1) \\ (x_{21}, x_{22}, \dots, x_{2k}, y_2) &= (\vec{x}_2, y_2) \\ &\dots \\ &\dots \\ (x_{n1}, x_{n2}, \dots, x_{nk}, y_n) &= (\vec{x}_n, y_n). \end{aligned}$$

Wir führen das allgemeine lineare Modell

$$(53) \quad y(\vec{x}) = \beta_0 + \beta_1 x + \dots + \beta_k x_k$$

ein. Diese Modellklasse besteht aus linearen Funktionen mit  $k$  Veränderlichen. So eine Modellklasse lässt viele Spezifikationen zu. So kann eine durch Messwertpaare gegebene Punktwolke durch allgemeinere Funktionen als Geraden approximiert werden. Wollen wir ein Polynom  $k$ -ten Grades zur Beschreibung der Punktwolke heranziehen:

$$y(x) = \beta_0 + \beta_1 x + \dots + \beta_k x^k,$$

so können wir die Substitution

$$x_j = x^j, \quad j = 1, \dots, k$$

in (53) einsetzen.

Das Prinzip der kleinsten Quadrate besagt, dass  $\hat{\vec{\beta}}$  als Minimalstelle der Funktion

$$(54) \quad S(\vec{\beta}) = \sum_{i=1}^n (y_i - \underbrace{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}_{(X\vec{\beta})_i\text{-tes Element}})^2$$

bestimmt werden muss, wobei die Datenmatrix  $X$  definiert ist als

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}.$$

Wir führen weiterhin den Datenvektor

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

ein.

SATZ 10.2. Gegeben seien die Datenmatrix  $X$  und der Datenvektor  $\vec{y}$ . Dann gilt für die Minimalstelle  $\hat{\vec{\beta}}$  von (54)

$$(55) \quad \hat{\vec{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (X^T X)^{-1} X^T \vec{y}.$$

BEWEIS. Es gilt:

$$(56) \quad \begin{aligned} S(\vec{\beta}) &= (\vec{y} - X\vec{\beta})^T (\vec{y} - X\vec{\beta}) = (\vec{y}^T - (X\vec{\beta})^T) (\vec{y} - X\vec{\beta}) \\ &= \vec{y}^T \vec{y} - (X\vec{\beta})^T \vec{y} - \vec{y}^T X\vec{\beta} + (X\vec{\beta})^T (X\vec{\beta}) \\ &= \vec{y}^T \vec{y} - (X\vec{\beta})^T \vec{y} - (X\vec{\beta})^T \vec{y} + \beta^T X^T X \beta \\ &= \vec{y}^T \vec{y} - 2\vec{\beta}^T (X^T \vec{y}) + \beta^T X^T X \beta. \end{aligned}$$

Dabei haben wir das Kommutativgesetz für das Skalarprodukt und die Regel der Matrixmultiplikation  $(AB)^T = B^T A^T$  angewendet. Wir formulieren die notwendige Bedingung für ein Minimum:

$$\frac{\partial S(\vec{\beta})}{\partial \vec{\beta}} = \vec{0}.$$

Differenzieren wir die einzelnen Glieder in (56), so erhalten wir nach Abschnitt 2 und Lemma 10.1

$$\frac{\partial \vec{y}^T \vec{y}}{\partial \vec{\beta}} = 0, \quad -2 \frac{\partial \vec{\beta}^T (X^T \vec{y})}{\partial \vec{\beta}} = -2X^T \vec{y}, \quad \frac{\partial \vec{\beta}^T (X^T X) \vec{\beta}}{\partial \vec{\beta}} = 2X^T X \vec{\beta}.$$

Speziell ist  $X^T X$  eine symmetrische Matrix. Zusammenfassend gilt:

$$\frac{\partial S(\vec{\beta})}{\partial \vec{\beta}} = -2X^T \vec{y} + 2X^T X \vec{\beta} = 0.$$

Ein kritischer Punkt  $\hat{\vec{\beta}}$  ist also Lösung des linearen Gleichungssystems

$$(57) \quad X^T X \hat{\vec{\beta}} = X^T \vec{y}.$$

Da die Matrix  $X^T X$  invertierbar ist, gilt (55).

Um zu sehen, dass  $\hat{\vec{\beta}}$  Minimalstelle ist, muss die Matrix der zweiten Ableitungen gebildet werden:

$$\frac{\partial^2 S(\hat{\vec{\beta}})}{\partial \hat{\vec{\beta}} \partial \hat{\vec{\beta}}} = 2X^T X.$$

Die Matrix  $X$  besitzt vollen Rang. Damit ist  $X^T X$  positiv definit. Daraus folgt, dass  $\hat{\vec{\beta}}$  die einzige Minimalstelle ist.  $\square$

Dieses in Matrixform gegebene lineare Gleichungssystem heißt *System der Normalgleichungen*. Dieses Gleichungssystem kann mittels des Austauschverfahrens, Gauß oder Cramer Algorithmus gelöst werden.

**Beispiel:** Es soll die Säureresistenz einer neuen Plastiksorte in Abhängigkeit von der Menge eines Zuschlagstoffes durch eine Funktion beschrieben werden. Es liegen folgende Messdaten vor:

$x_i$	$y_i$
1	2.044
2	3.800
3	4.528
4	4.006
5	1.391

TABELLE 2. Auflistung der Daten

Eine graphische Analyse der Daten ergibt, dass die Punktwolke durch ein Polynom zweiten Grades

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

gut approximiert wird. Wir setzen im Allgemeinen Modell  $x_{i1} = x_i$ ,  $x_{i2} = x_i^2$ . Daraus ergeben sich die Datenmatrizen

$$\vec{y} = \begin{pmatrix} 2.044 \\ 3.800 \\ 4.528 \\ 4.006 \\ 1.391 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{pmatrix},$$

wobei die letzte Spalte der Matrix  $X$  aus den Quadraten der zweiten Spalte besteht. Wir berechnen weiter

$$X^T X = \begin{pmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{pmatrix} \quad X^T \vec{y} = \begin{pmatrix} 15.771 \\ 46.212 \\ 156.884 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 4.6 & -3.3 & 0.5 \\ -3.3 & 2.671 & -0.428 \\ 0.5 & -0.428 & 0.071 \end{pmatrix}.$$

Daraus errechnet man

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \vec{y} = \begin{pmatrix} -1.51 \\ 4.175 \\ -0.712 \end{pmatrix}.$$

Damit beschreibt die Funktion

$$y(x) = -1.51 + 4.174x - 0.712x^2$$

nach dem Prinzip der kleinsten Quadrate am besten die Punktwolke. Betrachten wir abschließend noch die Residuen  $e_i = y_i - y(x_i)$

$x_i$	$e_i$
1	0.097
2	-0.17
3	-0.053
4	0.248
5	-0.11

TABELLE 3. Liste der Residuen

so sind die Werte  $e_i$  unregelmäßig um Null verteilt. Also ist ein Modell einer quadratischen Funktion passend, um die gegebenen Messdaten zu beschreiben. Mittels des Systems der Normalgleichungen (57) wollen wir noch weitere Eigenschaften des Modells

$$y = \beta_0 + \beta_1 x_n + \cdots + \beta_k x_k$$

herleiten. Wegen

$$\begin{aligned} X\hat{\beta} &= \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \cdots + \hat{\beta}_k x_{1k} \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \cdots + \hat{\beta}_k x_{nk} \end{pmatrix} \\ &= \begin{pmatrix} y(\text{1. Messwert}) \\ \vdots \\ y(\text{n. Messwert}) \end{pmatrix} \end{aligned}$$

ergibt sich aus (57)

$$X^T(\vec{y} - X\hat{\beta}) = X^T\vec{e} = \vec{0},$$

wobei  $\vec{e}$  den Vektor der Residuen bezeichnet. Wir erhalten  $\vec{e} = \vec{y} - X\hat{\beta}$ .

LEMMA 10.3. *Es gilt*

$$\sum_{i=1}^n e_i = 0, \quad \sum_{i=1}^n e_i x_{ij} = 0, \quad \text{für } j = 1, \dots, n$$

und

$$\frac{1}{n} \sum_{i=1}^n y(x_i) = \bar{y} := \frac{1}{n} \sum_{i=1}^n y_i.$$

Zum Beweis der ersten Zeile führe man die folgende Matrixmultiplikation aus:

$$\vec{0} = X^T\vec{e} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{n1} \\ \vdots & & \vdots \\ x_{k1} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}.$$

Unter anderem sagt der Hilfssatz aus, dass die Summe der Fehler (Residuen) gleich Null ist. Die Regressionsfunktion wird so konstruiert, dass sich die Fehler ausgleichen.

#### 4. Bewertung der Regressionsrechnung

Nachdem wir in der Lage sind, eine Funktion zu berechnen, die näherungsweise den in den Daten enthaltenen funktionalen Zusammenhang beschreibt, beschäftigen wir uns mit der Frage, wie gut diese Funktion diesen Zusammenhang beschreibt. Dabei stehen folgende Fragen im Mittelpunkt. Einerseits interessiert uns, ob das von uns gewählte Modell geeignet ist, den funktionalen Zusammenhang auszudrücken. Zum anderen ergibt sich die Frage, ob und mit welcher Intensität die Daten überhaupt so einen Zusammenhang zum Ausdruck bringen können.

Um sich eine Meinung über das gewählte Modell zu verschaffen, empfiehlt sich eine graphische Analyse der Residuen. Falls die Residuen *unregelmäßig* um Null variieren, kann davon ausgegangen werden, ein passendes Modell gewählt zu haben. Zeigen andererseits die Residuen die Tendenz, in gewissen Bereichen konzentriert kleiner oder größer als Null zu sein, so sollte die Wahl des Modells überprüft werden. Ein Wert, der zur Einschätzung der Qualität der Analyse herangezogen wird, ist das *Maß der Abweichung*

$$S = S(\hat{\beta}) = \sum_{i=1}^n e_i^2.$$

Dieser Wert entspricht dem Minimalwert der Funktion  $S(\beta)$ , siehe (52). Anstatt dieses Wertes wird manchmal auch der gemittelte Wert

$$\frac{1}{n} \sum_{i=1}^n e_i^2$$

zu Beschreibung der Abweichung herangezogen. Falls diese Werte *klein* sind, so paßt die berechnete Funktion gut in die Punktwolke, Falls diese Werte *groß* sind, so kann einerseits das Modell falsch gewählt sein, zum anderen ist es möglich, dass die Daten keine oder eine nur sehr schwache funktionale Abhängigkeit wiedergeben. Die Konzentration der Daten, um eine Funktion herum ist also schwach.

Problematisch ist wieder die Einschätzung, wann diese Werte als klein beziehungsweise groß einzuschätzen sind. Ein weiteres Problem ist, dass Maßstabsänderungen die obigen Werte beeinflussen.

Ein Maß, das diese Schwäche nicht besitzt, ist das *Bestimmtheitsmaß*. Wir wollen die Einführung dieses Maßes motivieren. Dazu führen wir die folgenden Größen ein:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \hat{y}_i = y(\vec{x}_i),$$

wobei der letzte Wert der Funktionswert der berechneten Regressionsfunktion an der Messstelle  $\vec{x}_i$  ist. Desweiteren schreiben wir

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i.$$

Wir notieren die Identität, die sich nach einer längeren Rechnung ergibt:

$$\sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 + \sum_{i=1}^n e_i^2.$$

Die in dieser Gleichung auftretenden Terme haben folgende Interpretation:  $y_i - \bar{y}$  beschreibt die Gesamtabweichung der Daten  $y_i$  vom Mittelwert  $\bar{y}$ . Diese Abweichung teilt sich auf in die durch die Regressionsfunktion erklärte Abweichung  $\hat{y}_i - \bar{y}$  und die durch die Regressionsfunktion nicht erklärte Abweichung  $y_i - \hat{y}_i$ , die dem Residuum  $e_i$  entspricht. Die Regressionsfunktion sollte das Problem gut beschreiben, falls das Verhältnis von erklärter und nicht erklärter Abweichung groß ist beziehungsweise die Residuen klein sind. Dieses Verhältnis definiert das Bestimmtheitsmaß

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Es gilt:  $0 \leq R^2 \leq 1$ . Falls  $R^2 = 1$ , so wird die Abweichung vollständig durch das Regressionsmodell erklärt. Aus der obigen Formel folgt speziell, dass die Residuen Null sind. Falls  $R^2 = 0$ , so gibt es keine Erklärung durch das Regressionsmodell. Das ist speziell dann der Fall, wenn die  $\bar{x}$  und  $y$  unabhängig voneinander sind. In diesem Fall kann es keine Funktion geben, die eine Beziehung zwischen den  $x$ - und  $y$ -Werten beschreibt.  $R^2$  kann aber auch gleich Null sein, wenn durch ein total unpassendes Modell versucht wird, die Daten zu beschreiben. Liegt  $R^2$  irgendwo zwischen 0 und 1, so hat man eine mehr oder weniger gute Erklärung der Abweichung  $y_i - \bar{y}$  durch das Regressionsmodell.

**Beispiel:** Legen wir die Daten des letzten Beispiels zu Grunde, so ergibt sich

$$\sum_{i=1}^n e_i^2 = 0.1186, \quad \bar{y} = 3.15423 \quad \text{und} \quad \sum_{i=1}^n (y_i - \bar{y})^2 = 7.3693.$$

Daraus errechnen wir  $R^2 = 0.983$ . Das bedeutet, die Gesamtabweichung der Daten  $y_i$  von  $\bar{y}$  ist sehr klein. Der durch die Daten bestimmte Zusammenhang zwischen  $x$  und  $y$  wird sehr gut durch die Regressionsfunktion beschrieben.

Abschließend wollen wir den Spezialfall der Regressionsgeraden  $y = \beta_0 + \beta_1 x$  betrachten. Dazu führen wir die Datenmatrix

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

ein. Wir berechnen weiterhin

$$X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad X^T \vec{y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{pmatrix}.$$

Löst man die Normalgleichungen (57) mittels der Cramerschen Regel, so ergibt sich

$$(58) \quad \hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}.$$

Wir erinnern uns, dass die empirische Varianz und Kovarianz gegeben sind durch

$$\begin{aligned} s_X^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ s_{XY} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \bar{y} - \frac{1}{n} \sum_{i=1}^n \bar{x} y_i + \frac{1}{n} \sum_{i=1}^n \bar{x} \bar{y} \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}. \end{aligned}$$

Damit erhalten wir die Identität

$$\hat{\beta}_1 = \frac{n^2 s_{XY}}{n^2 s_X^2} = \frac{s_{XY}}{s_X^2}.$$

Da  $s_{XY}$  die Wachstumstendenz einer Punktwolke beschreibt, korrespondiert diese mit dem Anstieg der Regressionsgeraden. Im Falle der Unabhängigkeit von  $x$  und  $y$  ist  $s_{XY} = 0$ . Obwohl es in diesem Fall nicht sinnvoll ist, eine Regressionsgerade zu ermitteln, würde man den Anstieg Null errechnen.

Der Wert für  $\hat{\beta}_0$  ergibt sich aus

$$(59) \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

LEMMA 10.4. *Es sei*

$$\rho = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}}$$

der Korrelationskoeffizient der Daten  $x, y$ . Dann gilt für das Bestimmtheitsmaß

$$R^2 = \rho^2.$$

Der Korrelationskoeffizient ist genau dann  $\pm 1$ , wenn die Daten auf einer Geraden liegen. In diesem Fall ist  $R^2 = \rho^2$ , das heißt die Regressionsgerade ist der Graph der linearen Funktion, auf dem die Datenpaare  $(x_i, y_i)$  liegen.

**Beispiel:** Gegeben seien die folgenden Messdaten

$$\begin{aligned} &(1, 2.78427), (2, 7.35267), (3, 10.3505), (4, 10.2231), (5, 7.89427), \\ &(6, 16.9344), (7, 13.5301), (8, 22.7624), (9, 13.1634), (10, 30.1559), \\ &(11, 23.2748), (12, 26.8797), (13, 31.9496), (14, 34.7657), (15, 30.8041), \\ &(16, 32.8026), (17, 36.9941), (18, 35.9862), (19, 39.9391), (20, 37.2845), \\ &(21, 45.7449), (22, 40.9074), (23, 52.1328), (24, 50.644), (25, 60.2601). \end{aligned}$$

Daraus errechnet man  $\hat{\beta}_1 = 2.04303$  und  $\hat{\beta}_0 = 2.061$ . Die graphische Darstellung der Daten, der Regressionsgeraden und der Residuen ergibt die folgende Graphik.

Da sich die Residuen sehr unregelmäßig um die Abszisse verteilen, beschreibt das Modell sehr gut die Daten. Das wird auch durch die Berechnung des Bestimmtheitsmaßes bestätigt:

$$R^2 = 0.9396.$$

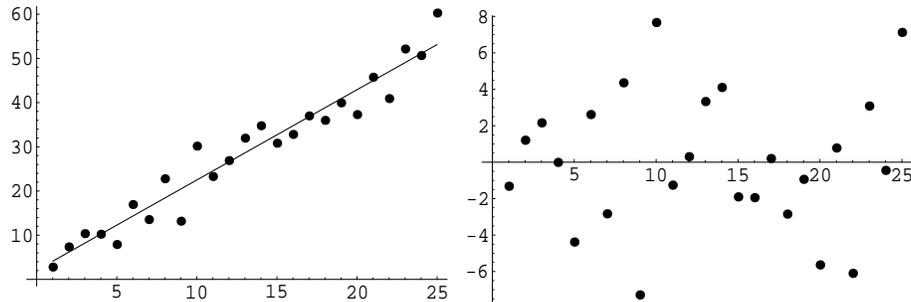


ABBILDUNG 3. Punktwolke und deren Approximation durch eine lineare Funktion und die zugehörigen Residuen

### 5. Das Gauß-Markov Modell der Regression

Wir wollen im Folgenden annehmen, dass die Messpunkte als Ergebnis von auftretenden Fehlereinflüssen bei der Messung entstehen. Wir nehmen an, dass die Abhängigkeit eines Merkmales  $Y$  von einem Merkmal  $X$  durch eine Funktion dargestellt werden kann, wobei davon ausgegangen wird, dass diese Funktion *linear* ist. Kompliziertere Abhängigkeiten können ähnlich behandelt werden. Wir betrachten die Funktion

$$y = b_0 + b_1x,$$

wobei  $b_0, b_1$  unbekannte Koeffizienten sind. Die Aufgabe besteht darin,  $b_0, b_1$  durch eine Schätzung näherungsweise zu bestimmen. Dazu müssen Messwerte ermittelt werden. Aufgrund der auftretenden Messfehler erhalten wir die Paare  $(x_i, y_i)_{i=1, \dots, n}$ , bestimmt durch die folgenden Gleichungen

$$y_i = b_0 + b_1x_i + e_i, \dots i = 1, \dots, n.$$

Die zufälligen und unbekanntenen Messfehler  $e_i$  werden als Zufallsvariablen angenommen. Damit sind auch die  $y_i$  Zufallsvariablen, während die  $x_i$  als *fest* gegebene Werte angenommen werden.

Die folgenden Annahmen über die Messfehler  $e_i$  werden als *Gauß-Markov-Bedingungen* bezeichnet.

- Der mittlere Messfehler ist Null:  $\mathbb{E}e_i = 0$  für  $i = 1, \dots, n$ .
- Die Varianz des Messfehlers ist unabhängig von  $i$ :  $D^2e_i = \mathbb{E}(e_i - \mathbb{E}e_i)^2 = 0$  für  $i = 1, \dots, n$ .
- Die Messfehler sind unkorreliert:  $\mathbb{E}(e_ie_j) = 0$  für  $i \neq j = 1, \dots, n$ .

Speziell folgt die Unkorreliertheit der Messfehler falls man annehmen kann, dass die Messfehler für verschiedene  $i$  unabhängig sind.

Wir haben zu Ende des letzten Abschnittes Berechnungsformeln kennengelernt, für Anstieg und Absolutglied einer Regressionsgeraden. Diese ergeben sich in Abhängigkeit von den zufälligen Werten  $y_i$ , sind somit selbst wieder Zufallsvariablen. In Abschnitt 2 hatten wir Qualitätskriterien für Schätzungen angegeben. Eines dieser Qualitätskriterien war die Erwartungstreue einer Schätzung, siehe Definition 8.3. Die Formeln (58) und (59) liefern solche erwartungstreuen Schätzungen. Es gilt also:

$$\mathbb{E}\hat{\beta}_1 = b_1, \quad \mathbb{E}\hat{\beta}_0 = b_0.$$

Unter der Annahme, dass die Gauß-Markov Bedingungen erfüllt sind, kann auch die Varianz der Koeffizienten berechnet werden. Für die Varianzen der Zufallsvariablen  $\hat{\sigma}_0$ ,  $\hat{\sigma}_1$  gilt:

$$(60) \quad D^2 \hat{\beta}_0 = \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2, \quad D^2 \hat{\beta}_1 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Diese Formeln erlauben die folgende Interpretation. In den Nennern dieser Ausdrücke treten Summen auf, die zu den empirischen Varianzen der Messstellen proportional sind. Das heißt, je weiter die Messstellen auseinander gezogen sind, desto größer ist der Nenner und desto kleiner ist die Varianz von  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ .

Im allgemeinen ist der in den Gauß-Markov-Bedingungen formulierte Wert  $\sigma$ , der die Varianzen von  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  bestimmt, unbekannt. Er muss aus den Daten geschätzt werden. Eine erwartungstreue Schätzung für  $\sigma^2$ ,  $D^2 \hat{\beta}_0$ ,  $D^2 \hat{\beta}_1$  ist

$$(61) \quad s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$s_{\hat{\beta}_0}^2 = \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} s^2, \quad s_{\hat{\beta}_1}^2 = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Wir werden im Folgenden die Annahme machen, dass die Fehler  $e_i$  normalverteilt sind. Zusammen mit den schon gemachten Voraussetzungen über die  $e_i$  gilt

$$e_i \sim N(0, \sigma).$$

**SATZ 10.5.** *Angenommen die Gauß-Markov-Bedingungen sind erfüllt und es gilt (61). Dann sind die Zufallsvariablen  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  normalverteilt mit Mittelwert  $\mathbb{E} \hat{\beta}_i = b_i$  und der in (60) gegebenen Varianz.*

Die folgenden Betrachtungen basieren auf diesem Satz.

Ähnlich wie in Abschnitt 2 tritt das Problem auf, dass die Punktschätzungen von  $\beta_0$ ,  $\beta_1$  sehr weit von den unbekanntenen Werten  $b_0$ ,  $b_1$  entfernt liegen können. Um eine Vorstellung zu bekommen, in welchem Bereich  $b_0$ ,  $b_1$  mit einer vorgegebenen Wahrscheinlichkeit liegt, wird ein Konfidenzintervall berechnet.

Wir legen mit  $\alpha (= 0.1, 0.05, 0.01)$  eine Irrtumswahrscheinlichkeit fest. Dann ist  $b_0$  mit Wahrscheinlichkeit  $1 - \alpha$  im Intervall

$$b_0 \in [\hat{\beta}_0 - s_{\hat{\beta}_0} t_{1-\frac{\alpha}{2}, n-2}, \hat{\beta}_0 + s_{\hat{\beta}_0} t_{1-\frac{\alpha}{2}, n-2}]$$

enthalten. Entsprechend gilt für den Koeffizienten  $b_1$ :

$$b_1 \in [\hat{\beta}_1 - s_{\hat{\beta}_1} t_{1-\frac{\alpha}{2}, n-2}, \hat{\beta}_1 + s_{\hat{\beta}_1} t_{1-\frac{\alpha}{2}, n-2}]$$

mit Wahrscheinlichkeit  $1 - \alpha$ .

Mit  $t_{1-\frac{\alpha}{2}, n-2}$  haben wir das  $1 - \frac{\alpha}{2}$ -Quantil der Student-Verteilung bezeichnet. Jedes der oben berechneten Intervalle enthält mit Wahrscheinlichkeit  $1 - \alpha$  die Koeffizienten  $b_0$ ,  $b_1$ . Wir können aber nicht daraus schließen, dass  $\beta_0$  und  $\beta_1$  gleichzeitig in den Intervallen mit Wahrscheinlichkeit  $1 - \alpha$  enthalten sind.

Um zu sichern, dass  $b_0$  und  $b_1$  gleichzeitig in diesen Intervallen enthalten sind, müssen *Simultane Konfidenzintervalle* berechnet werden. Es gilt mit Wahrscheinlichkeit  $1 - \alpha$

$$b_0 \in [\beta_0 - \sqrt{2s_{\hat{\beta}_0}^2 f_{1-\alpha, 2, n-2}}, \beta_0 + \sqrt{2s_{\hat{\beta}_0}^2 \hat{\beta}_0 f_{1-\alpha, 2, n-2}}] \quad \text{und}$$

$$b_1 \in [\beta_1 - \sqrt{2s_{\hat{\beta}_1}^2 f_{1-\alpha, 2, n-2}}, \beta_1 + \sqrt{2s_{\hat{\beta}_1}^2 f_{1-\alpha, 2, n-2}}].$$

Diese beiden Intervalle überdecken  $b_0, b_1$  gleichzeitig mit der Wahrscheinlichkeit von  $1 - \alpha$ .

In ähnlicher Weise, wie gerade Konfidenzintervalle für die Regressionskoeffizienten berechnet wurden, können auch Konfidenzintervalle für die Funktionswerte  $y = b_0 + b_1x$  berechnet werden. Dazu betrachten wir die Größe  $\mathbb{E}(y|x_0)$ , die gegeben ist als der Erwartungswert der  $y$ -Werte der Regressionsfunktion bei gegebenen  $x_0$ . Geben wir uns eine Irrtumswahrscheinlichkeit von  $\alpha$  vor, so gilt

$$\mathbb{E}(y|x_0) \in [\hat{\beta}_0 + \beta_1 x_0 - g(x_0), \hat{\beta}_0 + \beta_1 x_0 + g(x_0)],$$

wobei

$$g(x_0) = s t_{1-\frac{\alpha}{2}, n-2} \sqrt{\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

ist. Stellt man das Konfidenzintervall in Abhängigkeit von  $x_0$  dar, so erkennt man, dass es an der Stelle  $x = x_0$  am schmalsten ist. Im äußeren Bereich der Punktwolke werden die Konfidenzintervalle breiter.

Ein nach der obigen Formel berechnetes Intervall enthält nur für ein  $x_0$  den Erwartungswert  $\mathbb{E}(y|x_0)$  mit Wahrscheinlichkeit  $1 - \alpha$ . Ein simultanes Konfidenzintervall bezüglich aller  $x_0$ -Werte ist durch die folgende Formel gegeben

$$\mathbb{E}(y|x_0) \in [\hat{\beta}_0 + \beta_1 x_0 - g(x_0), \hat{\beta}_0 + \beta_1 x_0 + g(x_0)],$$

wobei

$$g(x_0) = \sqrt{2s^2 f_{1-\alpha, 2, n-2} \left( \frac{\frac{1}{n} + (\bar{x} - x_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

ist. Ein Prognoseintervall, dass eine Realisierung des geschätzten Wertes der Regressionsgeraden an der Stelle  $x_0$  mit Wahrscheinlichkeit  $1 - \alpha$  enthält, ist gegeben durch

$$y(x_0) \in [\hat{\beta}_0 + \beta_1 x_0 - g(x_0), \hat{\beta}_0 + \beta_1 x_0 + g(x_0)],$$

wobei

$$g(x_0) = s t_{1-\frac{\alpha}{2}, n-2} \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

gilt.

Abschließend sollen noch Hypothesen über die Koeffizienten der Regressionsgerade getestet werden. Wir gehen im Folgenden von hypothetischen Werten für  $b_0, b_1$  und  $\sigma$  aus, die wir mit  $b_0^0, b_1^0$  und  $\sigma_0$  bezeichnen.

### Testen des Absolutgliedes einer Regressionsgeraden

*Hypothese:*

$$H_0 : b_0 = b_0^0$$

$$H_a : b_0 \neq b_0^0$$

*Testgröße:*

$$t = \frac{\hat{\beta}_0 - b_0^0}{s_a}$$

*Kritischer Bereich:*

$$|t| > t_{1-\frac{\alpha}{2}, n-2}.$$

**Testen des Anstieges einer Regressionsgeraden***Hypothese:*

$$H_0 : b_1 = b_1^0$$

$$H_a : b_1 \neq b_1^0$$

*Testgröße:*

$$t = \frac{\hat{\beta}_1 - b_1^0}{s_{\hat{\beta}_1}}.$$

*Kritischer Bereich:*

$$|t| > t_{1-\frac{\alpha}{2}, n-2}.$$

Die Größen  $t_{1-\frac{\alpha}{2}, n-2}$  repräsentieren die  $1 - \frac{\alpha}{2}$ -Quantile der Studentverteilung mit  $n - 2$  Freiheitsgraden.

**Testen der Messfehlervarianz  $\sigma^2$** *Hypothese:*

$$H_0 : \sigma = \sigma^0$$

$$H_a : \sigma \neq \sigma^0$$

*Testgröße:*

$$\chi^2 = \frac{(n-2)s^2}{\sigma_0^2}.$$

*Kritischer Bereich:*

$$\chi^2 \leq \chi_{\frac{\alpha}{2}, n-2}^2 \quad \text{oder} \quad \chi^2 \geq \chi_{1-\frac{\alpha}{2}, n-2}^2.$$

$\chi_{\frac{\alpha}{2}, n-2}^2, \chi_{1-\frac{\alpha}{2}, n-2}^2$  sind die  $\frac{\alpha}{2}$  beziehungsweise  $1 - \frac{\alpha}{2}$  der  $\chi^2$ -Verteilung mit  $n - 2$  Freiheitsgraden.

**Beispiel** (Fortsetzung): Berechnet man die Schätzungen für die Varianz des Fehlers  $\sigma$  und die Varianzen von  $\hat{\beta}_0, \hat{\beta}_1$  so ergibt sich

$$s^2 = 15.1655, \quad s_{\hat{\beta}_1}^2 = 0.0116, \quad s_{\hat{\beta}_0}^2 = 2.57813.$$

Die Konfidenzintervalle für  $\mathbb{E}(y|x)$  sind in der Grafik 4 dargestellt.

Abschließend sollen noch die Hypothesen getestet werden, dass  $b_0 = 3, b_1 = 2$  und die Varianz des Fehlers gleich 10 ist. Wir erhalten folgende Testgrößen für die drei Tests

$$t = -8.68921, \quad t = 0.0267 \quad \text{und} \quad \chi^2 = 34.8806.$$

Für die entsprechenden Quantile gilt:

$$t_{0.975, 23} = 2.06866 \quad \text{und} \quad \chi_{0.975, 23}^2 = 38.0756 \quad \chi_{0.025, 23}^2 = 11.6886.$$

Damit wird die Hypothese  $H_0 : b_0 = 3$  abgelehnt. Die Hypothese  $H_0 : b_1 = 2$  und  $H_0 : \sigma^2 = 10$  werden nicht abgelehnt.

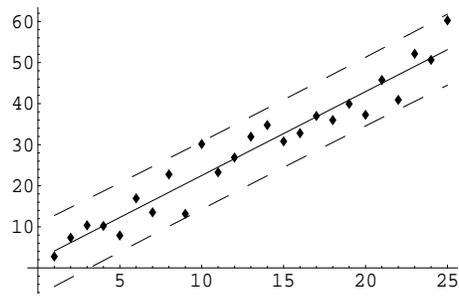


ABBILDUNG 4. Die gestrichelten Linien sind die Grenzen der Konfidenzintervalle die  $E(y|x_0)$  mit der Wahrscheinlichkeit von 0.95 enthalten

## ANHANG A

# Kombinatorik

Die Kombinatorik beschäftigt sich mit der Anzahl von Möglichkeiten, aus einer gegebenen Anzahl von Elementen Gruppen (Klassen) zusammenzustellen. Wir wollen annehmen, daß unterscheidbare Elemente mit  $1, 2, \dots, n$  bezeichnet werden.

Zuerst betrachten wir Zusammenstellungen, bei denen *alle* Elemente in einer Zusammenstellung vorkommen. Sie werden unterschieden, falls sie sich in der Reihenfolge der Anordnung unterscheiden. Diese Zusammenstellungen nennen wir *Permutationen*.

DEFINITION A.1. *Die Zusammenstellungen von  $n$  unterschiedlichen Elementen heißt Permutation ohne Wiederholung.*

Die Anzahl der möglichen Zusammenstellungen berechnet sich aus

$$P_n = n! \quad .$$

$n!$  ist die *Fakultät*:

$$n! = \begin{cases} 1 \cdot 2 \cdot \dots \cdot n & \text{falls } n = 1, 2, \dots \\ 1 & \text{falls } n = 0 \end{cases} .$$

**Beispiel:** Es können  $P_5 = 5! = 120$  verschiedene sechsstellige Zahlen aus den Ziffern von 1 bis 6 gebildet werden, wenn jede Ziffer höchstens einmal vorkommt und an der ersten Position immer die 1 steht.

Nun nehmen wir an, daß es  $n$  Elemente gibt, wobei unter diesen  $n$  Elementen genau  $n_1, n_2, \dots, n_j$  Elemente,

$$n = \sum_{i=1}^j n_i,$$

*nicht* unterscheidbar sind.

DEFINITION A.2. *Die Zusammenstellungen von  $n$  Elementen, wobei genau*

$$n_1, n_2, \dots, n_j$$

*Elemente nicht unterscheidbar sind, heißen eine Permutation mit Wiederholung.*

Die Anzahl der Zusammenstellungen ergibt sich aus

$$P_{n_1, \dots, n_j}^W = \frac{n!}{n_1! n_2! \dots n_j!} .$$

**Beispiel:** In einer Stadt verlaufen alle Straßen parallel in Nord-Süd Richtung und in ost-west Richtung, siehe Abbildung 1.

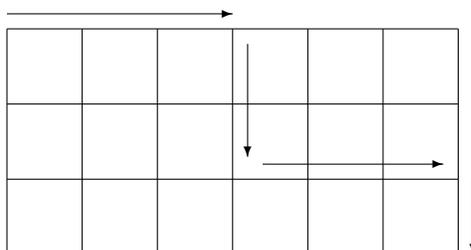


ABBILDUNG 1. Straßenverlauf

Es soll die Frage beantwortet werden, wieviele Wege es aus dem Nordwesten der Stadt (linke obere Ecke) in den Südosten (rechte untere Ecke) gibt. Dabei darf nur nach Osten beziehungsweise nach Süden gegangen werden. Der im Bild eingezeichnete Weg kann durch den Vektor  $(1, 1, 1, 0, 0, 1, 1, 1, 0)$  charakterisiert werden. Dabei wird eine 1 geschrieben, falls man einen Straßenabschnitt nach Osten und eine Null, falls man einen Straßenabschnitt nach Süden geht. Damit gibt es eine eindeutige Zuordnung von Wegen und Vektoren die genau sechs Einsen und drei Nullen enthalten. Die Anzahl dieser Vektoren entspricht also einer Permutation mit Wiederholung, wobei  $n = 9$ ,  $n_1 = 6$ ,  $n_2 = 3$  gilt. Wir erhalten  $P_{6,3}^W = 84$  unterschiedliche Wege.

Bei den folgenden Zusammenstellungen werden  $n$  Elemente zu Klassen der Größe  $k$  zusammengestellt. Angenommen, es seien die Elemente 1, 2, 3, 4 gegeben und die Klassengröße sei  $k = 2$ . Dann wäre eine Möglichkeit der Zusammenstellung zu Klassen:

$$(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)$$

gegeben. Es fällt auf, daß die Reihenfolge, wie die Elemente in den Klassen stehen, nicht berücksichtigt wird. Das heißt, daß zum Beispiel die Zusammenstellungen  $(1, 2)$  und  $(2, 1)$  nur einmal gezählt werden. Zusammenstellungen, in denen die Reihenfolge in den Klassen keine Rolle spielt, heißen eine *Kombination*. Weiterhin erkennt man im obigen Beispiel, daß niemals die gleichen Elemente wie  $(1, 1), \dots$ , in den Klassen auftreten. Wir haben also Zusammenstellungen *ohne Wiederholungen*.

**DEFINITION A.3.** Die Zusammenstellungen von  $n$  unterschiedlichen Elementen zu Klassen der Größe  $k$ , wobei die Reihenfolge der Elemente in den Klassen nicht berücksichtigt wird, und weiterhin jedes Element höchstens einmal in einer Klasse auftritt, heißen Kombination ohne Wiederholung.

Die Anzahl der möglichen Klassen berechnet sich aus

$$K_{n,k} = \binom{n}{k}.$$

$\binom{n}{k}$  ist der bekannte *Binomialkoeffizient*, der sich aus der Formel

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

berechnet.

**Beispiel:** Beim Lotto 6 aus 49 gibt es genau

$$K_{49,6} = 13983816$$

verschiedene Möglichkeiten, einen Tippschein auszufüllen. Denn jeder Tippschein entspricht einer Klasse mit sechs Elementen. Dabei wird berücksichtigt, daß jede Zahl bei der Ziehung höchstens einmal gezogen wird. Die Reihenfolge, in der die Zahlen gezogen werden, hat darauf, ob man gewinnt oder nicht, keinen Einfluß. Spielt wiederum die Reihenfolge einer Zusammenstellung in einer Klasse keine Rolle, können aber Elemente in einer Klasse mehr als einmal vorkommen, so erhält man eine Kombination mit Wiederholung.

DEFINITION A.4. Falls  $n$  Elemente zu Klassen der Größe  $k$  zusammengestellt werden, wobei die Reihenfolge der Anordnung keine Rolle spielt, Elemente in den Klassen öfters als einmal vorkommen können, so spricht man von einer Kombination mit Wiederholung.

Die Anzahl der möglichen Klassen erhält man als

$$K_{n,k}^W = \binom{n+k-1}{k}.$$

**Beispiel:** Ein Dominospiel mit Steinen von  $(0,0)$  bis  $(6,6)$  besitzt genau

$$K_{7,2}^W = 28$$

Steine.

Zusammenstellungen, bei denen die Reihenfolge des Auftretens in den Klassen berücksichtigt wird, das heißt zum Beispiel  $(1,2)$  und  $(2,1)$  als zwei Elemente gezählt werden, bezeichnet man als *Variation*. Parallel zu den Kombinationen definiert man

DEFINITION A.5. Die Zusammenstellungen von  $n$  unterschiedlichen Elementen zu Klassen der Größe  $k$ , wobei die Reihenfolge der Anordnung der Elemente berücksichtigt wird, jedes Element höchstens einmal in einer Klasse auftritt, heißen *Variation ohne Wiederholung*.

Man kann die Anzahl der möglichen Zusammenstellungen berechnen als

$$V_{n,k} = \frac{n!}{(n-k)!}.$$

Diese Formel läßt sich sehr einfach aus der Formel für die Kombination ohne Wiederholung und Permutation ohne Wiederholung herleiten. Man überlege wie?

**Beispiel:** Die Anzahl der Fahrkarten, die für eine Strecke mit 10 Stationen (zwei Endstationen) gedruckt werden müssen beträgt

$$V_{10,2} = 90.$$

BEMERKUNG A.6. Falls  $n = k$  ist, so erhält man

$$P_n = V_{n,n}.$$

DEFINITION A.7. Zusammenstellungen von  $n$  unterschiedlichen Elementen zu Klassen der Größe  $k$ , wobei die Reihenfolge der Anordnung der Elemente berücksichtigt wird, Elemente auch öfters als einmal in einer Klasse auftreten können, heißen *Variation mit Wiederholung*.

Die Formel zur Berechnung der Anzahl der Klassen lautet:

$$V_{n,k}^W = n^k.$$

**Beispiel:** Blindenschrift besteht aus bis zu sechs Positionen, die aus dem Papier gestanzt werden. Eine ausgestanzte Position werde mit 1, eine nichtausgestanzte mit Null kodiert. Da die Position der Ausstanzungen eine Rolle spielt und die beiden Zeichen öfters als einmal vorkommen, hat man

$$V_{2,6}^W = 2^6 = 64$$

Zeichen zur Verfügung.

## Literaturverzeichnis

- [1] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin, 1991.
- [2] K. Behnen and G. Neuhaus. *Grundkurs Stochastik*. B. G. Teubner, 1995.
- [3] F. Beichelt. *Stochastik für Ingenieure*. B. G. Teubner, Stuttgart, 1995.
- [4] O. Beyer, H. Hackel, H. Pieper, and V. Tiedge. *Wahrscheinlichkeitsrechnung und mathematische Statistik*. B. G. Teubner, Leipzig, 2. edition, 1988.
- [5] H. Gillert and V. Nollau. *Übungsaufgaben zur Wahrscheinlichkeitsrechnung und mathematischen Statistik*. Teubner Leipzig, Leipzig, 4. edition, 1990.
- [6] M. Greiner and G. Tinhofer. *Stochastik für Studienanfänger der Informatik*. Hanser, München-Wien, 1996.
- [7] N. Henze. *Stochastik für Einsteiger*. Vieweg, Braunschweig, Wiesbaden, 3. edition, 2000.
- [8] G. Hübner. *Stochastik*. Vieweg, Braunschweig-Wiesbaden, 2. edition, 2000.
- [9] W.-M. Kähler. *Einführung in die statistische Datenanalyse*. Vieweg, Braunschweig/Wiesbaden, 1995.
- [10] D. N. Knuth. *The art of computer programming*. Addison-Wesley Publ. Comp., Reading, Massachusetts, 1997.
- [11] V. Krebs. *Nichtlineare Filterung*. R. Oldenbourg Verlag, München, Wien, 1980.
- [12] J. Lehn, H. Wegmann, and S. Rettig. *Aufgabensammlung zur Einführung in die Statistik*. Teubner, Stuttgart, 1988.
- [13] R. Mather and D. Pfeifer. *Stochastik für Informatiker*. B. G. Teubner, Stuttgart, 1990.
- [14] O. Moeschlin. *Statistik und experimentelle Stochastik*. Birkhäuser, Basel, 1995.
- [15] P. H. Müller. *Wahrscheinlichkeitsrechnung und Mathematische Statistik*. Akademie-Verlag, Berlin, 1970.
- [16] V. Nollau, L. Partzsch, R. Storm, and C. Lange. *Wahrscheinlichkeitsrechnung in Beispielen und Aufgaben*. B. G. Teubner Verlagsgesellschaft, Stuttgart-Leipzig, 1997.
- [17] S. Osaki. *Applied stochastic system modeling*. Springer, 1992.
- [18] L. Sachs. *Angewandte Statistik*. Springer-Verlag, Berlin, Heidelberg, New York, 1999.
- [19] B. Schmalfuß. *Statistik-Formeln*. 1998.
- [20] R. Storm. *Wahrscheinlichkeitsrechnung, Mathematische Statistik und Qualitätskontrolle*. Fachbuchverlag, Leipzig-Köln, 10. edition, 1995.
- [21] H. von Storch, S. Güss, and M. Heinemann. *Das Klimasystem und seine Modellierung*. Springer, Berlin, 1999.
- [22] H. W. Zwanziger. *Statistik-Vadamekum, Lehrmaterial zur Analytikausbildung an der FH-Merseburg*. Merseburg, Merseburg, 2000.