

Mathematik für das Lehramt an der Grund- und
Mittelstufe sowie an Sonderschulen
Teil IV — SoSe 05
Einführung in elementare Stochastik
Wahrscheinlichkeitsrechnung und Statistik

Bodo Werner

<mailto:werner@math.uni-hamburg.de>

12. Juli 2005

Inhaltsverzeichnis

IV	Stochastik	7
1	Vorwort	9
1.1	Internet-Seiten	10
1.2	Lehrbücher	11
2	Beschreibende Statistik	13
2.1	Einführung	13
2.1.1	Die wichtigsten Begriffe	16
2.2	Merkmalraum	16
2.2.1	Stichprobe	16
2.2.2	Diskrete und kontinuierliche Merkmale	17
2.3	Grafische Darstellungen von Erhebungen	19
2.3.1	Beispiel	19
2.3.2	Diagramme in Excel	20
2.3.3	Histogramme	22
2.3.4	Eigenschaften von Häufigkeitsverteilungen	25
2.4	Maßzahlen	26
2.4.1	Mittelwert	26
2.4.2	Charakterisierung des Mittelwertes	27
2.4.3	Mittelwert als Schwerpunkt	27
2.4.4	Skaleneinfluss auf den Mittelwert	28
2.4.5	Varianz, Streuung	29
2.4.6	Skaleneinfluss	30
2.4.7	Berechnungen mit Excel	31
2.4.8	Median, Quantile, Quartile	32
2.4.9	Empirische Verteilungsfunktion	33
2.5	Statistik und Wahrscheinlichkeitstheorie	37
2.6	Korrelation zweier Merkmale	38
2.6.1	Kontingenztafel	39
2.6.2	Korrelationskoeffizient	39
2.6.3	Regression	41

2.6.4	Eine Warnung	43
3	Einführung in die Wahrscheinlichkeitsrechnung	45
3.1	Einführung	45
3.1.1	Elementare Anwendungen der Kombinatorik auf die Wahrscheinlichkeits- Rechnung	45
3.1.2	Beispiele für Fragen aus der Stochastik	47
3.2	Kombinatorik - Wiederholung	50
3.3	Merkmalraum	52
3.3.1	Zufallsvariablen: Ein erster Zugang	54
3.3.2	Zufällige Ereignisse	54
3.4	Häufigkeit und Wahrscheinlichkeit	55
3.4.1	Häufigkeiten	55
3.4.2	Wahrscheinlichkeit bei Zufallsexperimenten	56
3.4.3	Wahrscheinlichkeiten in der Statistik	57
3.4.4	Wahrscheinlichkeit: Axiome von Kolmogoroff	57
3.4.5	Einschluss (Inklusion) - Ausschluss (Exklusion)-Formel	60
3.5	Wahrscheinlichkeits-Modelle, Verteilungen	62
3.5.1	Diskrete Verteilungsfunktion bei quantitativen Merkmalen	63
3.5.2	Bernoulli-, Binomial-, Laplace-Modelle und ihre Verteilungen	65
3.5.3	Poissonverteilung	68
3.5.4	Geometrische Verteilung	70
3.5.5	Hypergeometrische Verteilung	71
3.6	Kontinuierliche Verteilungen	71
3.6.1	Rechtecksverteilung	72
3.6.2	Wahrscheinlichkeits-Dichte	73
3.6.3	Exkurs Integration	74
3.6.4	Normalverteilung - erster Zugang	76
3.6.5	Bemerkungen zu kontinuierlichen Verteilungen in der Statistik	77
3.7	Bedingte Wahrscheinlichkeiten, unabhängige Ereignisse	78
3.7.1	Beispiel für bedingte Wahrscheinlichkeiten aus der Medizin	80
3.8	Reelle Zufallsvariable	81
3.8.1	Verteilung und Verteilungsfunktion	82
3.9	Kenngößen von Zufallsvariablen	84
3.9.1	Erwartungswert	84
3.9.2	Berechnung der Erwartungswerte für bestimmte Verteilungen	85
3.9.3	Median, Quantile	87
3.9.4	Varianz, Streuung	87
3.9.5	Varianz von bestimmten Verteilungen	88
3.9.6	Kenngößen bei kontinuierlichen Verteilungen	88
3.9.7	Rechnen mit Zufallsvariablen	89

3.10	Unabhängigkeit und Kovarianz von Zufallsvariablen	90
3.10.1	Zufallsstichprobe und stochastische Unabhängigkeit	90
3.10.2	Stochastische Unabhängigkeit	90
3.10.3	Kovarianz	90
3.11	Normalverteilung	91
3.11.1	Zentraler Grenzwertsatz	94

Teil IV
Stochastik

Kapitel 1

Vorwort

Die mit den Begriffen Zufall, Wahrscheinlichkeit und Statistik verbundene *Stochastik*¹ spielt auch außerhalb der Mathematik eine herausragende Rolle. Sei es die Wahrscheinlichkeit, im Glücksspiel zu gewinnen, die morgige „Regenwahrscheinlichkeit“ in einem Wetterbericht, die Wahrscheinlichkeit, an einer bestimmten Krankheit zu erkranken, die neuesten Arbeitslosenstatistiken oder eine Umfrage vor der Bundestagswahl.

Die gesamte Versicherungsbranche ruht auf statistischen Grundlagen wie die (steigende) Lebenserwartungen bei Lebensversicherungen oder die Häufigkeit von Sturmschäden bei Gebäudeversicherungen.

Auch die Medizin kommt ohne statistische Fallstudien bei der Erprobung neuer Medikamente oder bei der Frage, ob gewisse Umweltrisiken verantwortlich sind für das Auftreten bestimmter Krankheiten, nicht aus.

Stochastik wird auch auf Warteschlangen an Fahrkartenschaltern der Bahn oder in Wartezimmern angewendet. Allgemein befasst sie sich mit Untersuchungen von Ereignissen, die vom Zufall beeinflusst werden.

Zufällige Ereignisse werden oft durch erhobene Daten („Zufällige Stichproben“) dokumentiert, für deren Analyse die Statistik – ein Teilgebiet der Stochastik – geeignete Methoden bereitstellt. Für viele Studierende der Wirtschaftswissenschaften, Psychologie, Biologie, u.a. sind Mathematik und Statistik identisch.

Wegen ihrer großen Bedeutung zählt die Stochastik heute zur Schulmathematik. Sie gehört zum Kerncurriculum für Lehramtsstudienordnungen, ist also ein Pflichtfach. Dieses Manuskript versucht einen Einstieg in die Stochastik zu geben. Dass der Autor von Hause kein „Stochastiker“ ist, ist vielleicht ein Vorteil, wird aber sicher auch zu Ungereimtheiten führen, zuweilen sogar zu Irrtümern.

Ziel der Lehrveranstaltung ist es nicht, Rezepte und Schemata zu vermitteln, wie man Wahrscheinlichkeiten berechnet oder wie man Daten einer Stichprobe auswertet. Vielmehr ist das

¹Der Begriff Stochastik stammt ursprünglich aus dem Griechischen und bedeutet dort: die Kunst des geschickten Vermutens. Er umfasst sowohl die Wahrscheinlichkeitstheorie, auch Wahrscheinlichkeitsrechnung genannt, als auch die Statistik.

Ziel, dass die grundlegenden *Konzepte* der Stochastik so weit verstanden werden, dass wenigstens exemplarisch Berechnungen und Auswertungen durchgeführt werden können und dass das nötige Knowhow vorliegt, um im gesellschaftlichen Umfeld mitdenken und mitargumentieren zu können. Diese Konzepte der Stochastik verbinden sich in der Regel mit speziellen *sprachlichen Begriffen* („Vokabeln“) wie **diskreter** und **kontinuierlicher, qualitativer** und **quantitativer Merkmalraum, Häufigkeiten, (Häufigkeits- und Wahrscheinlichkeits-) Verteilungen von Stichproben** bzw. **Zufallsvariablen** und (empirische) **Verteilungsfunktionen** und deren Kenngrößen wie **Mittelwert, Erwartungswert, Standardabweichung, Streuung, Varianz, Median, Quantile, Quartile**, ferner **Ereignis, Wahrscheinlichkeit(s-Maß), bedingte Wahrscheinlichkeit**, und spezielle Verteilungen wie **Bernoulli-, Binomial-, Poisson-, Laplace-, Normalverteilung** und deren **Wahrscheinlichkeitsdichte, stochastische Unabhängigkeit** von Ereignissen und Zufallsvariablen, sowie die **Schätzung** der Kenngrößen zugehöriger Zufallsvariable und **Tests** zur Absicherung von **Hypothesen**.

Ich habe mich entschlossen, mit der sog. *Beschreibenden Statistik* zu beginnen, weil diese zum einen wesentlich einfacher ist als die Wahrscheinlichkeits-Rechnung, zum anderen aber auch für den Alltag gegenwärtiger ist. Ich behaupte auch, dass die Statistik eine größere gesellschaftliche Bedeutung als die reine Wahrscheinlichkeitsrechnung hat. Ich denke nur an die Medizin und die Wirtschaftsstatistik. Allerdings sind beide Gebiete sehr eng verwoben, was hoffentlich deutlich werden wird.

Methodisch ist diese Reihenfolge fragwürdig, weil man *Stichproben* — ein Stichprobenvektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ist Ausgangspunkt aller Ausführungen der *Beschreibenden Statistik* — als einzige Möglichkeit auffassen kann, *Zufallsexperimente* und die mit ihnen verbundenen Zufallsvariablen zu realisieren. Daher wird es viel Redundanz geben. Alleine die Namen *Verteilung* und *Verteilungsfunktion* tauchen in verschiedenen Gewändern auf.

1.1 Internet-Seiten

1. In den folgenden Seiten werden sehr häufig Bezüge zu einem hervorragenden Multi-Media-Manuskripts über Biometrie² der Uni Münster (Autoren: ACHIM HEINECKE und WOLFGANG KÖPCKE) hergestellt:

Java unterstützte Münsteraner Biometrie-Oberfläche³

Viele Beispiele und Grafiken verdanke ich diesem Skript.

Die können eine CD-ROM mit allen Unterlagen bei Prof. Köpcke für nur 8 Euro bestellen.

Ich werde mich auf dieses Skript unter dem Namen JUMBO beziehen, da dies die Kurzform des Arbeitstitels des Skripts (Java-Unterstützte Münsteraner Biometrie-Oberfläche) ist, siehe Abb. 1.1.

²Es geht im Wesentlichen um Statistik mit Anwendungen in der Medizin

³<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/bio.html>



Abbildung 1.1: JUMBO

Ich zitiere (H. Grahlki: Die akademische Lehre im Netz. Forschung und Lehre 2 (1998) 69-71): „Es ergibt sich manchmal der Eindruck, dass die Begeisterung der Autoren über ihre Produkte in einem umgekehrten Verhältnis zu der Bereitschaft der Adressaten steht, die Systeme wirklich systematisch für Lernzwecke zu nutzen.“

2. VOLKER SCHMIDT: **Wahrscheinlichkeitsrechnung - Skript im Internet**⁴ (Uni Ulm)
und
Statistik I - Skript im Internet⁵ (Uni Ulm)

1.2 Lehrbücher

1. GERHARD HÜBNER: **Stochastik**. Vieweg.
2. Nach Fertigstellung des Skripts bin ich auf das gerade erschienene Buch
G. FISCHER: **Stochastik einmal anders**. Vieweg 2005
gestoßen, dessen Konzept mit dem Skript ziemlich übereinstimmt, auch im Abstraktheitsgrad.
Sehr empfehlenswert!!!
3. WALTER KRÄMER: **Statistik verstehen**. Piper.
4. JOHANN PFANZAGL: **Elementare Wahrscheinlichkeitsrechnung**. de Gruyter.
5. REGINA STORM: **Wahrscheinlichkeitsrechnung, Mathematische Statistik, Statistische Qualitätskontrolle**. VEB.

⁴http://www.mathematik.uni-ulm.de/stochastik/lehre/ws01_02/wr/vs1/vs1.html

⁵<http://www.mathematik.uni-ulm.de/stochastik/lehre/ss02/statistik1/skript/>

Kapitel 2

Beschreibende Statistik

2.1 Einführung

Alle empirischen Untersuchungen in Pädagogik, Psychologie, Marktforschung, Medizin, Biologie, Sozialwissenschaften etc. bedienen sich der *beschreibenden (deskriptiven) Statistik*, um ihre quantitativen Untersuchungsergebnisse für eine Analyse in Form von Tabellen, Grafiken und statistischen Maßzahlen aufzubereiten. Die *Schließende Statistik* dagegen befasst sich mit den Schlussfolgerungen aus den erhobenen Daten.

Ohne verlässliches Zahlenmaterial kann man nur schwerlich planen. In der Politik ist es Aufgabe des *Statistischen Bundesamtes* und der *statistischen Landesämter*, solches Zahlenmaterial zur Information der Bevölkerung, der Gesetzgebung und Verwaltung, aber auch als Grundlage von Entscheidungen und zu wissenschaftlichen Analysen zu ermitteln. Dabei begnügt man sich mit Schätzungen auf Grund von (Teil-) Erhebungen, da nur selten die wahren Zahlen bekannt sind — hierzu bedarf es einer Totalerhebung. Das Wort *Statistik* hat etwas mit dem Staat zu tun, es bedeutet ursprünglich eine Art *Staatskunde*, einer *Staatsbeschreibung*, in der es vor allem um Zahlen geht! Schon in der Weihnachtsgeschichte („...auf dass alle Welt geschätzt werde...“) geht es um eine „Volkszählung“, die der römische Kaiser Augustus befahl. Noch heute ist die *Bevölkerungsstatistik* eine ganz wesentlicher Anlass zu politischen Diskussionen, von der *Arbeitslosenstatistik* ganz zu schweigen. Umweltschutz ist ohne Statistik gar nicht denkbar, da man nur durch Messungen auf Probleme hinweisen kann!

Die Beziehung zur Wahrscheinlichkeitsrechnung der Stochastik kommt dann zum Tragen, wenn man den *empirischen Verteilungsfunktionen* bekannte Verteilungen der Stochastik (Binomialverteilung, Normalverteilung, etc.) gegenüberstellt. Wie im Vorwort zu diesem Skript ausgeführt, kann man *Stichproben* als eine Möglichkeit auffassen, *Zufallsexperimente* und die mit ihnen verbundenen *Zufallsvariablen* zu realisieren.

Ausgangspunkt jeder Statistik sind *Erhebungen*, meist Teilerhebungen in Form von (*Zufalls-*)

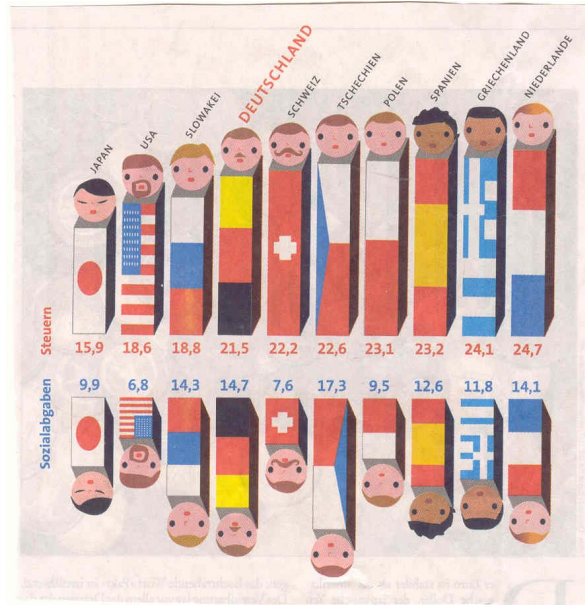


Abbildung 2.1: Sozialabgaben versus Steuern

Stichproben. Mathematisch formuliert wird bei *quantitativen Merkmalen* eine Stichprobe im einfachsten Fall durch einen *Vektor* $\mathbf{x} := (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ wiedergegeben, den ich **Stichprobenvektor** nenne¹. Dabei ist $n \in \mathbb{N}$ der **Umfang** der Stichprobe, die Komponenten x_k sind die am k -ten „Objekt“ (oder auch „Individuum“) erhobenen Daten. Zum Beispiel könnte es sich bei x_k um das Geburtsgewicht des k -ten Neugeborenen in einem Krankenhaus handeln, n ist dann die Anzahl der Neugeborenen, deren Gewicht man in der Erhebung misst. In den meisten Anwendungen, die hier betrachtet werden, ist n relativ groß und die einzelnen Objekte „gehen in der Statistik unter“. Ferner sind die x_k meist Zahlen. In solchen Fällen ist die typische grafische Darstellung ein *Histogramm* mit *klassierten Daten*.

Häufig werden aber nicht nur einzelne Zahlen x_k erhoben, sondern gleich ein ganzer *Datensatz* wie z.B. bei Neugeborenen Name, Alter und Beruf der Eltern, Blutgruppe und Körperlänge des Neugeborenen, Datum und Uhrzeit der Geburt, etc. Dabei müssen die einzelnen Merkmale nicht quantitativ sein, wie z.B. Geschlecht und Blutgruppe. Solche nicht quantitativen Merkmale heißen auch *qualitativ*.

Ein Beispiel dieser Art stellt Abb. 2.1 dar. Hier werden *zwei* verschiedene Merkmale erfasst — die Steuerlast und die Sozialabgaben. Die „befragten“ Individuen sind Länder. Man kann nicht von einer Stichprobe sprechen, sondern von einer Erhebung, dessen Umfang n sehr klein ist. Im Gegensatz zu den Histogrammen in Kap. 2.3 werden hier die einzelnen „befragten Individuen“ auf der horizontalen Achse und die einzelnen Daten auf den vertikalen Achsen dargestellt.

Die gesamte Versicherungsbranche ruht auf statistischen Grundlagen wie die (steigende) Le-

¹Ich habe auch den Namen *Urliste* gefunden.

benserwartungen bei Lebensversicherungen oder die Häufigkeit von Sturmschäden bei Gebäudeversicherungen. Hier liefert jeder Todesfall oder jeder Sturmschaden einen „Datensatz“.

Auch die Medizin kommt ohne statistische Fallstudien bei der Erprobung neuer Medikamente oder bei der Frage, ob gewisse Umweltrisiken verantwortlich sind für das Auftreten bestimmter Krankheiten, nicht aus.

Stichproben haben in der Regel ein Ziel. So will man beispielsweise durch eine Umfrage ermitteln, wie die Altersverteilung von Raucherinnen und Raucher ist. Man stellt sich z.B. die Frage: Wieviel Prozent aller 17-jährigen Jugendlichen rauchen mehr als 5 Zigaretten am Tag? Das „Objekt der Begierde“ ist dann diese (unbekannte) Prozentzahl, die mit Hilfe einer Stichprobe *geschätzt* wird, man will von der Stichprobe auf die *Grundgesamtheit* schließen. Die k -te Komponente des Stichprobenvektors \mathbf{x} gibt dann die Anzahl der täglichen Zigaretten des k -ten befragten 17-jährigen Jugendlichen. Allgemein will man aus einer kleinen Teilgesamtheit Rückschlüsse auf die Grundgesamtheit ziehen.

An dem letzten Beispiel kann man auch den Bezug zur *Wahrscheinlichkeitsrechnung* erkennen: Die gesuchte Prozentzahl kann man auch als Wahrscheinlichkeit interpretieren, dass ein zufällig ausgewählter 17-jähriger Jugendliche mehr als 5 Zigaretten raucht. Diese Wahrscheinlichkeit wird nun mit Hilfe einer *relativen Häufigkeit* aller derjenigen 17-jährigen Jugendlichen der Stichprobe geschätzt, die mehr als 5 Zigaretten rauchen.

Einen noch engeren Bezug der Statistik zur Wahrscheinlichkeitsrechnung wird deutlich, wenn man das immer wieder strapazierte Würfelspiel nimmt. Angenommen, man möchte empirisch die Wahrscheinlichkeit dafür berechnen, dass man mit zwei Würfeln die Augensumme 10 würfelt. Dann kann man eine „Stichprobe“ von 100 Würfeln mit zwei Würfeln durchführen und die Anzahl der Würfe zählen, die 10 als Ergebnis haben. Die relative Häufigkeit für ein solches Ereignis ist dann das Ergebnis der Stichprobe und kann als Schätzung für die gesuchte Wahrscheinlichkeit dienen. Das empirische Gesetz der großen Zahl ist es dann auch, was über diesen Weg zum Wahrscheinlichkeitsbegriff der Kolmogoroff-Axiome führt.

Durch diese Einführung sollte schon deutlich werden, dass es eine enge Beziehung zwischen *relativen Häufigkeit* und *Wahrscheinlichkeit* gibt.

Ziel des vorliegenden Skripts über Statistik ist nicht die Fähigkeit, den praktischen Umgang mit Statistiken einzuüben. Ziel ist es vielmehr, die mathematischen Grundprinzipien der Statistik als ein reiches Anwendungsfeld der Mathematik kennen und verstehen zu lernen, zumal diese im gesellschaftspolitischen Alltag eine große Bedeutung haben. Viele Ihrer Freundinnen und Freunde, die im Studium mit Mathematik zu tun haben, werden sich mit Statistik befassen müssen. Es wäre doch schön, wenn Sie mitreden können.

Ich empfehle Ihnen, auf die grafischen Darstellungen in den Medien zu achten und zu versuchen, diese hier einzuordnen. Dabei werden Sie auf einen Typ stoßen, den ich hier nicht behandeln werde: Grafiken, die die *zeitliche Entwicklung* irgendeiner Wachstumsgröße (z.B. Arbeitslosenzahlen) beschreiben. Bei der Analyse und Darstellung solcher *Zeitreihen* treten andere Fragestellungen auf.

2.1.1 Die wichtigsten Begriffe

Die folgenden Begriffe sollen Sie am Ende der Vorlesung erklären können:

Relative Häufigkeit eines **Merkmals** einer **Stichprobe** vom **Umfang** n , die **Häufigkeitsverteilung** aller Merkmale in einer Stichprobe und ihre grafische Darstellung z.B. durch **Histogramme**, die zugehörige **empirische Verteilungsfunktion**, die Lagemaße **Mittelwert**, **Median**, **Quantile**, **Quartile** einer Stichprobe, die Streumaße **Varianz** und **Standardabweichung**.

Korrelationskoeffizient zwischen zwei Stichprobenvektoren. **Lineare Regression** als Methode, durch eine Punktwolke von Datenpunkten eine Gerade zu legen.

2.2 Merkmalraum

Zentraler Begriff sowohl der Statistik als auch der Wahrscheinlichkeits-Rechnung ist der des **Merkmalraums**, den wir Ω nennen. Dieser enthält alle möglichen² **Ergebnisse** einer Erhebung, auch *Merkmalsausprägungen* des jeweiligen Merkmals genannt. Die Elemente von Ω heißen **Elementarereignisse** und werden — bei endlichem Ω — mit $\omega_1, \omega_2, \dots, \omega_m$ durchnummeriert.

In der Wahrscheinlichkeitsrechnung enthält der Merkmalraum die Ergebnisse von *Zufallsexperimenten*.

2.2.1 Stichprobe

Da eine Totalerhebung nur bei kleinen „Grundgesamtheiten“ möglich ist, ist man meist auf Teilerhebungen in Form von (Zufalls-) Stichproben angewiesen.

Diese Ergebnisse sind gerade die Komponenten des **Stichprobenvektors** $\mathbf{x} = (x_1, \dots, x_n)$, d.h. es gilt $x_k \in \Omega, k = 1, 2, \dots, n$. Man kann die Stichprobe auch als eine Funktion $X : M \rightarrow \Omega$ auffassen mit einer Menge M von n „Individuen“. Mit einer Durchnummerierung von $M := \{1, 2, \dots, n\}$ ist dann $x_j = X(j)$ gerade der Funktionswert der Stichprobenfunktion X .

Auch Stichproben in der Statistik haben einen Zufallsaspekt: Das Ergebnis ist von vornerein nicht bekannt, es erscheint zuweilen „zufällig“. In diesem Sinne ist z.B. die Beobachtung der Blutgruppe eines Neugeborenen auch ein Zufallsexperiment, später werden wir jede Beobachtung einer Stichprobe mit einer Zufallsvariablen verbinden.

Später werden Sie lernen, dass wir formal dann n Zufallsvariable X_1, X_2, \dots, X_n haben, von denen eine „echte“ Zufallsstichprobe annehmen muss, dass alle $X_k, k = 1, 2, \dots, n$ *identisch verteilt* und *stochastisch unabhängig* sind.

Beispiele von Erhebungen aus der Statistik:

²Man sagt, dass die durch die Merkmale gegebene Klasseneinteilung aller Erhebungsgegenstände disjunkt und erschöpfend sein muss.

- Eine Erhebung, die das Merkmal Geschlecht einer Person betrifft. Ω besteht aus den beiden Merkmalsausprägungen „weiblich“ und „männlich“.
- Pisastudie Mathematik: Eine Testperson bearbeitet einen Aufgabensatz und wird mit einer Punktzahl zwischen 0 und 1000 bewertet, d.h. $\Omega = \{n \in \mathbb{N}_0 : 0 \leq n \leq 1000\} \subset \mathbb{N}$.
- Wahlerhebung: Ω besteht aus allen zur Wahl stehenden Parteien. Es werden n WählerInnen befragt.
- Länge (in Metern) eines Menschen: $\Omega = [0, 3] \subset \mathbb{R}$. n ist die Anzahl der Menschen, die ausgemessen werden.
- Es wird das Geschlecht, das Alter und die Anzahl der täglichen Zigaretten von n Erwachsenen erhoben. Dann ist Ω das *kartesische Produkt* $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ mit $\Omega_1 := \{\text{„weiblich“}, \text{„männlich“}\}$, $\Omega_2 := [18, 120] \subset \mathbb{N}$, $\Omega_3 = \mathbb{N}_0$. Man könnte \mathbb{N}_0 durch ein Intervall $[0, 1000]$ ersetzen, wenn man annimmt, dass kein Mensch mehr als tausend Zigaretten täglich rauchen wird.
- Es wird die Anzahl von Blitzen in einer Minute gezählt. Dann ist $\Omega := \mathbb{N}_0$. Auch hierbei wird keine Obergrenze für die Anzahl der Blitze angenommen. n ist die Anzahl der eine Minute dauernden Experimente.

Bemerkung: Die Abb. 2.1 ist zwar sehr attraktiv, passt aber nicht so richtig in unsere „Stichprobenkonzeption“. Allenfalls künstlich: Man einen zufällig in Europa aufgefundenen Sozialabgabepflichtigen Euro, in welchem Land er erhoben wurde.

Man unterscheidet **quantitative (zahlenmäßige)** und **qualitative (begriffliche) Merkmale**, je nachdem, ob man sie durch „Messen“ oder „Zählen“ gewinnen kann oder nicht. So sind die Merkmale „Geschlecht“ und „Blutgruppe“ qualitative Merkmale, „Alter“ und „Gewicht“ sind quantitative.

Achtung: Ω ist ein kartesisches Produkt von Mengen bei der Erhebung mehrerer „verbundener“ Daten, d.h. z.B. die Daten zu jeweils einer Person, die durch einen Fragebogen ermittelt wurden, zusammengestellt zu einem *Datensatz*. n ist dann die Anzahl der erhobenen Datensätze.

Bemerkung zum Begriff *Datensatz*: Dieser ist zentral für eine *Datenbank*, deren Tabellen aus lauter Datensätzen bestehen. Die Anzahl n der Datensätze entspricht dem Umfang der Stichprobe, die *Datenfelder* eines Datensatzes sind von verschiedenem Typ, z.B. vom Typ *Text* (Name), vom Typ *Datum* (Geburtsdatum), vom Typ *Zahl* (Gewicht) — ein quantitatives Merkmal — oder auch vom Typ einer Grafik (Foto).

2.2.2 Diskrete und kontinuierliche Merkmale

Ist Ω endlich oder wenigstens abzählbar, so heißt der Merkmalraum **diskret**, ansonsten **kontinuierlich** (manchmal auch **stetig**). Kontinuierliche Merkmale treten häufig bei statistischen

Erhebungen der Länge³, Gewicht, des Blutdrucks, der Temperatur, etc. auf. Fasst man gewisse Intervalle zu einer Merkmalsausprägung zusammen, so spricht man von **Klassierung** (z.B. „Gewichtsklassen“). So kann man aus kontinuierlichen Merkmalen diskrete machen. Eine solche Klassierung ist immer dann angebracht, wenn der Umfang n der Stichprobe bzw. die Anzahl n der befragten Individuen sehr groß ist.

Qualitative Merkmale sind immer diskret (z.B. Nationalität, Blutgruppe, Beruf, Partei,...).

Ist der Merkmalraum diskret, so kann man jedem *Elementarereignis* (auch *Merkmalsausprägung* genannt) $\omega \in \Omega$ eine *absolute* und eine *relative Häufigkeit* innerhalb einer Stichprobe vom Umfang n zuordnen — der vielleicht wichtigste Begriff der Statistik:

Definition 2.1. Die Anzahl n_ω der Beobachtungen innerhalb der Stichprobe mit Ergebnis ω heißt **absolute Häufigkeit**, während der Quotient n_ω/n die **relative Häufigkeit** von ω ist.

Anders formuliert: Sei $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega^n$ mit $x_k \in \Omega$ das Ergebnis der Stichprobe. Dann gilt

$$n_\omega := |\{k : 1 \leq k \leq n \text{ und } x_k = \omega\}|.$$

Nummeriert man die Elementarereignisse in Ω zu

$$\Omega := \{\omega_1, \omega_2, \dots, \omega_m\},$$

so ist die Bezeichnung $n_j := n_{\omega_j}$ als *absolute Häufigkeit* und $\mathbf{h}_j := \mathbf{n}_j/\mathbf{n}$ als *relative Häufigkeit* des j -ten Merkmals ω_j sinnvoll. Natürlich muss

$$n_1 + n_2 + \dots + n_m = n \text{ bzw. } h_1 + h_2 + \dots + h_m = 1$$

gelten. Die relativen Häufigkeiten werden auch in Prozent angegeben — das kennen Sie von den relativen Wachstumsraten. So gilt beispielsweise $h_j = 0.31 = 31\%$. Die Zusammenstellung aller relativer Häufigkeiten zu einem Vektor kann man auch als *Häufigkeitsverteilung* bezeichnen, deren grafische Darstellung im nächsten Abschnitt Kap. 2.3 behandelt wird.

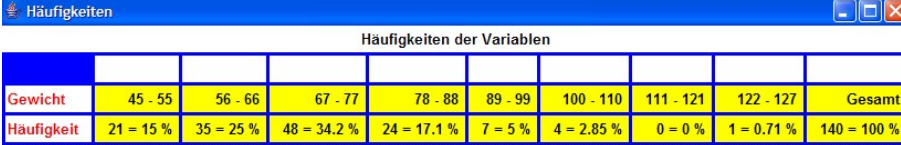
Schon jetzt sollte man darauf hinweisen, dass eine solche Konzeption etwa bei der Ermittlung der Körperlänge oder auch des Alters nur Sinn macht, wenn man die Daten zu *Klassen* zusammenfasst.

Definition 2.2. Sei $h_j, j = 1, 2, \dots, m$ die relative Häufigkeit von $\omega_j \in \Omega$ im Stichprobenvektor \mathbf{x} . Dann heißt der Vektor

$$\mathbf{h} := (h_1, h_2, \dots, h_m)$$

Häufigkeitsverteilung der Stichprobe.

³Da man nur mit einer gewissen Genauigkeit misst, hat man es im Grunde auch nur mit endlich vielen Merkmalen zu tun. Dies sind aber zu viele, mathematisch ist es einfacher, mit kontinuierlichen Merkmalen zu arbeiten



Häufigkeiten der Variablen									
Gewicht	45 - 55	56 - 66	67 - 77	78 - 88	89 - 99	100 - 110	111 - 121	122 - 127	Gesamt
Häufigkeit	21 = 15 %	35 = 25 %	48 = 34.2 %	24 = 17.1 %	7 = 5 %	4 = 2.85 %	0 = 0 %	1 = 0.71 %	140 = 100 %

Abbildung 2.2: Häufigkeitstabelle

Bemerkung: Es ist nötig, die Ergebnisse x_k der Stichprobe von den potentiellen Merkmalsausprägungen ω_j zu unterscheiden. Ist n der Umfang der Stichprobe, so heißt dies nicht, dass es n verschiedene Ergebnisse x_k gibt; Vielmehr werden i.A. mehrere x_k mit einem ω_j zusammenfallen, nämlich dann, wenn $n_j > 1$. Meist gilt $n \gg m$, so dass zwangsläufig einige $n_j > 1$ sein müssen.

Ich versuche im Folgenden, den Index j für die Merkmalsausprägung ω_j und den Index k für das Stichprobenergebnis x_k zu verwenden.

2.3 Grafische Darstellungen von Erhebungen

Die einfachste Form der grafischen Darstellung ist die durch ein **Blockdiagramm**, s. Abb. 2.3. Die *Merkmalausprägungen* werden an einer Achse in beliebiger bzw. in der natürlichen Reihenfolge (bei sogenannten ordinalen Merkmalen) angetragen. Darüber wird ein Block gezeichnet, dessen Höhe der absoluten bzw. der relativen Häufigkeit des jeweiligen Merkmals entspricht. Die Breite der Blöcke ist beliebig, sie soll aber für alle Blöcke gleich sein.

Bei einem **Kreisdiagramm** (s. Abb. 2.4) entspricht der absoluten bzw. der relativen Häufigkeit der Ausprägung der zentrale Winkel des zugeordneten Kreissegments.

Bei einem **Flächendiagramm** (s. Abb. 2.5) entspricht der absoluten bzw. der relativen Häufigkeit der Ausprägung der Flächeninhalt des zugeordneten Segments.

Neben grafischen Darstellungen kann man natürlich auch Tabellen verwenden, s. Abb. 2.2 mit klassierten Daten.

2.3.1 Beispiel

Bei einer Stichprobe von Patienten, die unter Krampfadern im Unterschenkelbereich litten⁴, wurde eine Salbe zur Linderung der Beschwerden angewandt. Eine halbe Stunde nach Auftragen der Salbe wurden die Patienten befragt, ob eine Besserung eingetreten sei. Es ergab sich folgende Liste:

⁴Dieses Beispiel stammt aus JUMBO.

Besserung	absolute Häufigkeit	relative Häufigkeit
keine	3	12.5%
gering	10	41.7%
deutlich	7	29.2%
keine Angabe	4	16.6%
Gesamt:	24	100%

Tabelle 2.1: Auswertung der Krampfadernbehandlung

Patient	Besserung	Patient	Besserung
1	gering	13	gering
2	deutlich	14	gering
3	gering	15	keine
4	deutlich	16	keine Angabe
5	gering	17	gering
6	keine	18	deutlich
7	deutlich	19	deutlich
8	deutlich	20	gering
9	keine Angabe	21	keine Angabe
10	gering	22	gering
11	keine	23	gering
12	keine Angabe	24	deutlich

Der Merkmalraum Ω besteht aus den vier möglichen Antworten „keine“, „gering“, „deutlich“ und „keine Angabe“.

Hieraus ergibt sich die Tabelle 2.1 der absoluten und relativen Häufigkeiten:

Eine graphische Darstellung durch Block-, Kreis- und Flächendiagramme findet man in den Abb. 2.3-2.5.

2.3.2 Diagramme in Excel

Das Microsoft-Tabellenkalkulationsprogramm *Excel* erlaubt statistische Berechnungen mit grafischer Aufbereitung mittels *Diagramme*.

Beispiel: Man trage die absoluten Häufigkeiten der Tabelle 2.1 in eine Excel-Spalte, kann diese aufsummieren und in einer Nachbarspalte die relativen Häufigkeiten (in Prozent) berechnen. Sodann kann man in der Menüleiste *Einfügen* mit der Auswahl *Diagramm* wählen, den *Diagrammtyp* auswählen (z.B. *Säulen-* oder *Kreisdiagramm* — die Bezeichnungen sind etwas anders als oben).

Das Arbeits-Excelblatt finden Sie in Abb. 2.6.

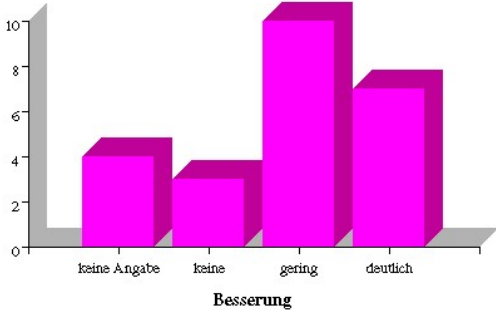


Abbildung 2.3: Blockdiagramm

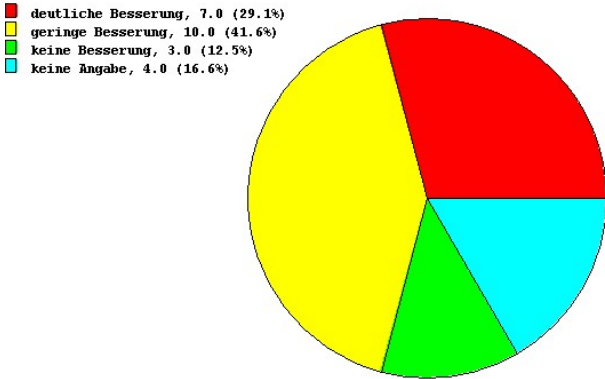


Abbildung 2.4: Kreisdiagramm

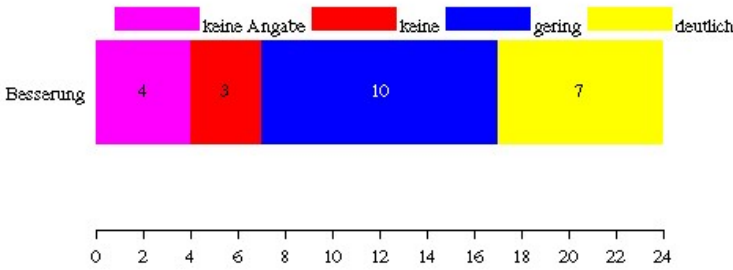


Abbildung 2.5: Flaechendiagramm

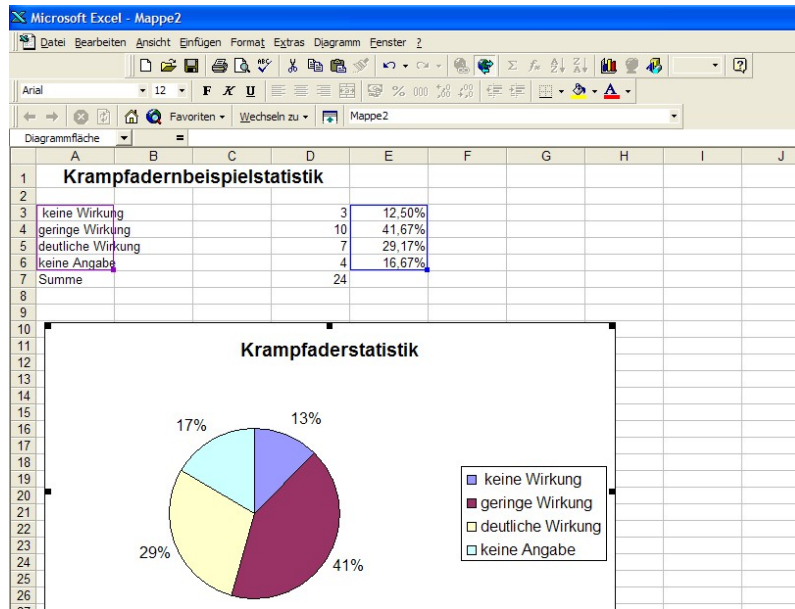


Abbildung 2.6: Kreisdiagramm mit Excel

2.3.3 Histogramme

Bei diskreten Merkmalen unterscheiden sich qualitative und quantitative Merkmale nur insofern, als dass letztere auf natürliche Weise angeordnet sind. Statt von Blockdiagrammen spricht man in solchen Fällen auch von *Stabdiagrammen*. Ordnet man die quantitativen Merkmale zu

$$\omega_1 < \omega_2 < \dots < \omega_m$$

an, so kann man

$$N_j := n_1 + \dots + n_j, \quad H_j := h_1 + \dots + h_j, \quad j = 1, 2, \dots, m$$

als **kumulierte** absolute bzw. relative Häufigkeiten dafür betrachten, dass in der Stichprobe ein Merkmal $\leq \omega_j$ ausfällt, siehe auch die *empirische Verteilungsfunktion* in Kap. 2.4.9.

Kontinuierliche Merkmalräume kann und sollte man durch *Klassierung* zu diskreten Merkmalräumen machen. Zum Beispiel, indem man die Körpergröße in Klassen mit einer gewissen *Klassenbreite* einteilt. Die entsprechenden Diagramme heißen **Histogramme**, siehe Abb. 2.7-2.8. Wenn man dies nicht täte und z.B. die Längen von $n = 1000$ Individuen auf Zentimeter genau misst und grafisch darstellt, bekommt man lauter Stäbe fast gleicher Höhe. FISCHER nennt dies einen Datenfriedhof, indem er die Stäbe mit einem Grabstein vergleicht.

Aber auch bei diskreten Merkmalräumen mit sehr vielen Ausprägungen kann es sinnvoll sein zu klassieren. In den Abb. 2.9-2.10 kann man z.B. die Verteilung der Punktzahlen der ersten Zwischenprüfung mit $n = 71$ TeilnehmerInnen einmal fast unklassiert (bei halben Punktzahlen

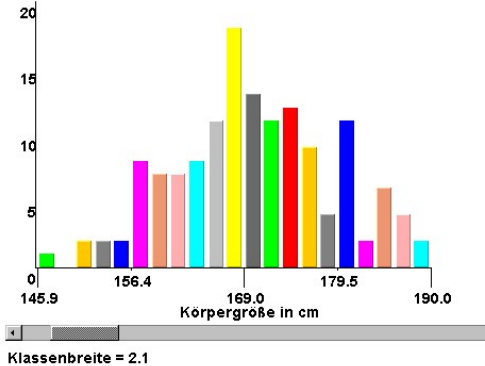


Abbildung 2.7: Histogramm mit Klassenbreite 2.1

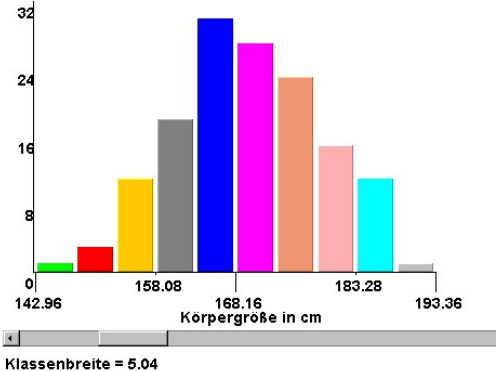
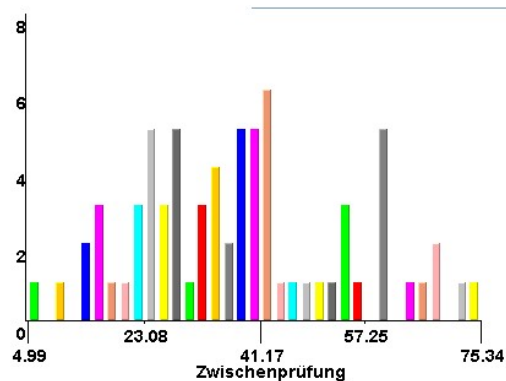


Abbildung 2.8: Histogramm mit Klassenbreite 5.04



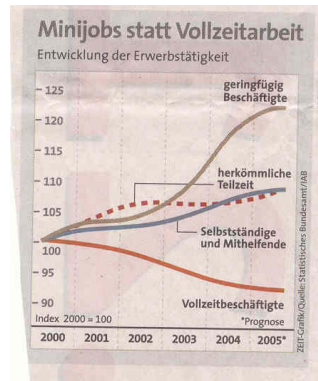


Abbildung 2.11: Minijobs: Grafik in „Die ZEIT“

2.3.4 Eigenschaften von Häufigkeitsverteilungen

Im nächsten Kap. 2.4 werden Sie *Kenngrößen* in Form von *Maßzahlen* von Häufigkeitsverteilungen. Es gibt aber auch Eigenschaften wie *links-* und *rechtsgipflig* oder *symmetrisch*, die Häufigkeitsverteilungen zugeschrieben werden können. Die Namen sprechen für sich.

Ist z.B. die Häufigkeitsverteilung des Alters von Versicherten linksgipflig, so sind Jüngere stärker vertreten. Auch die morgentlichen Körpertemperaturen sind linksgipflig verteilt, während die Gewichtsverteilung älterer Menschen wohl rechtsgipflig ist.

Manipulationen

„Ich glaube nur der Statistik, die ich selbst gefälscht habe“. Dieses angeblich von Winston Churchill stammende Zitat kennen Sie wahrscheinlich. Danach ist Statistik eine Form der Lüge. Ein Buch von W. KRÄMER „So lügt man mit Statistik“ gibt einen Einblick in verschiedene Manipulationsmöglichkeiten.

Eine sehr häufig benutzte Form der grafischen Manipulation ist die, dass die Ordinate nicht bei Null beginnt, sondern bei einem vergleichsweise hohen Wert. Dadurch werden Unterschiede in den Häufigkeiten optisch verstärkt. Steigt der Umsatz in einem Jahr nur um 2%, so kann man dies als immensen Erfolg darstellen, indem man die Ordinate bei etwa 80% des Umsatzes beginnen lässt. Dann wirken 2% Zuwachs optisch wie 10%.

Ein Beispiel dieser Art zeigt eine Grafik aus DIE ZEIT in Abb. 2.11.

Häufig werden aber auch falsche Schlussfolgerungen aus Statistiken gezogen. Untersucht man z.B. das Alter von 12 Todesfällen bei Studenten innerhalb eines Zeitraums, so seien hiervon 3 jünger als 20, 4 zwischen 20 und 24 und 5 zwischen 25 und 30, während keiner älter als 30 war. Kann man daraus schließen, dass Studenten nicht älter als 30 Jahre werden?

Ein weiteres Zitat: „Statistik ist die Lehre, dass es besser sei, an Typhus zu erkranken, als Millionär zu werden. Denn: Alle Millionäre sterben; von den Typhuskranken sterben nur siebzehn Prozent.“

2.4 Maßzahlen

In diesem Kapitel setze ich voraus, dass wir es mit einem diskreten, quantitativen Merkmalraum $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ mit den Ausprägungen

$$\omega_1 < \omega_2 < \dots < \omega_m$$

zu tun haben — möglicherweise nach einer Klassierung der Daten.

Man unterscheidet *Lagemaße* (wie *Mittelwert* und *Median*) und *Streuemaße* (wie *Standardabweichung*, *Varianz*, *Quartilsabstand*). Mit diesen Maßzahlen (auch Kenngrößen genannt) wird versucht, die Daten zu strukturieren bzw. besser noch zu *verdichten*.

2.4.1 Mittelwert

Vorweg: „Steht jemand mit einem Fuß auf der Herdplatte und mit dem anderen im Eiskasten, dann sagt der Statistiker: Im Durchschnitt ist ihm angenehm warm“.

Die bekannteste Maßzahl einer Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)$ zu einem quantitativen Merkmal ist der *Mittelwert*, auch *Durchschnittswert* genannt:

Definition 2.3.

$$\bar{x} = \frac{\sum_{k=1}^n x_k}{n}$$

heißt **Mittelwert** der Stichprobe $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

Es sollte klar sein, dass der Mittelwert diesen Namen zu Recht trägt: \bar{x} liegt in der Tat irgendwo zwischen den x_k . Sehr einfach einsehbar, ist

$$m_0 := \min_{k=1,2,\dots,n} x_k \leq \bar{x} \leq M_0 := \max_{k=1,2,\dots,n} x_k.$$

Wenn nämlich $\bar{x} < m_0$ (indirekter Beweis!), so gilt $\bar{x} < m_0 \leq x_k, k = 1, 2, \dots, n$. Dann gilt aber auch

$$x_1 + x_2 + \dots + x_n > \bar{x} + \bar{x} + \dots + \bar{x} = n \cdot \bar{x}.$$

Dividiert man diese Ungleichung durch n , so erhält man den Widerspruch $\bar{x} > \bar{x}$.

Durch die Schreibweise \bar{x} für den Mittelwert wird der funktionale Zusammenhang zwischen der Stichprobe, genauer dem Stichprobenvektor \mathbf{x} , und dem Mittelwert angedeutet. In den Übungen wird untersucht, wie sich Änderungen eines der Daten x_k auf den Mittelwert auswirkt. Analoge Fragestellungen kann man auch für andere Kenngrößen wie den Median stellen.

Jetzt kann man die relativen Häufigkeiten $h_j, j = 1, 2, \dots, m$, ins Spiel bringen.

Satz 2.4. Es gilt für den Mittelwert $\bar{x} = \frac{\sum_{k=1}^n x_k}{n}$ die Formel

$$\bar{x} = \sum_{j=1}^m h_j \omega_j.$$

Beweis: Zu gegebenem ω_j fasse man alle x_k mit $x_k = \omega_j$ zusammen. Davon gibt es n_j Stück, so dass deren Beitrag zum Mittelwert gerade $n_j\omega_j/n = h_j\omega_j$ ist. ■

Es wurde schon erwähnt, dass die relativen Häufigkeiten h_j mit Wahrscheinlichkeiten einer Zufallsvariable für das Auftreten von ω_j , die ich p_j nenne, zusammenhängen. Dann wird die zweite Formel in Satz 2.4 zu einem *Erwartungswert* dieser Zufallsvariable. \bar{x} wird zu einem *gewichteten Mittelwert* der ω_j mit *Gewichten* h_j . Näheres im Kapitel über Stochastik.

2.4.2 Charakterisierung des Mittelwertes

Die Formel für den Mittelwert scheint irgendwie einleuchtend, sie erscheint „gerecht“, weil sie kein Ergebnis x_k bevorzugt. Die Formel verallgemeinert das *arithmetische Mittel* $\frac{a+b}{2}$ zweier Zahlen a und b . Daher wird \bar{x} auch **arithmetisches Mittel** aller Daten $x_k, k = 1, 2, \dots, n$, genannt.

Eine andere Sichtweise wird durch die Eigenschaft

$$\sum_{k=1}^n (x_k - \bar{x}) = 0$$

wiedergegeben. Die Summe aller vorzeichenbehafteten Abstände der Daten zum Mittelwert verschwindet.

In Kap. 2.4.5 wird die Varianz mit Hilfe der *Abstandskquadrate* $(x_k - \bar{x})^2$ definiert, und es kommt die *Fehlerquadratsumme* $\sum_{k=1}^n (x_k - \bar{x})^2$ ins Spiel, deren funktionale Abhängigkeit von \bar{x} jetzt untersucht wird, d.h. wir betrachten

$$q : \mathbb{R} \rightarrow \mathbb{R}, q(x) = \sum_{k=1}^n (x_k - x)^2.$$

Die entscheidende Eigenschaft lautet

Satz 2.5. *Der Mittelwert \bar{x} ist Tiefpunkt (Minimum) von q , d.h. es gilt*

$$q(\bar{x}) \leq q(x) \text{ für alle } x \in \mathbb{R}.$$

Beweis: Mit Hilfe der Schulmathematik. Es gilt $q'(\bar{x}) = 0$ und $q''(\bar{x}) > 0$. ■

2.4.3 Mittelwert als Schwerpunkt

Wir denken uns die Zahlengerade als gewichtslosen, horizontalen Stab, auf dem an den Positionen x_k Punktmassen gleicher Masse angebracht werden. Ist die absolute Häufigkeit von ω_j durch $n_j \geq 1$ gegeben, so ist in ω_j die n_j -fache Punktmasse loziert. Unterstützt man den Stab am Punkt des Mittelwerts \bar{x} , so hält dieser den Stab im Gleichgewicht — man kann \bar{x} auch als den **Schwerpunkt** der x_k ansehen.

2.4.4 Skaleneinfluss auf den Mittelwert

Wenn man die Daten einer Stichprobe in verschiedenen Skalen messen würde, erwartet man, dass sich die Mittelwerte durch eine Skalentransformation auseinander ergeben.

Beispiel: Die Beziehung zwischen der Celsius-Skala (x) und der Fahrenheitskala (y) wird durch die Skalentransformation

$$y = f(x) := \frac{9}{5}x + 32$$

beschrieben. Wenn der Stichprobenvektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$ in Celsius gemessen wird, ist $\mathbf{y} = (y_1, y_2, \dots, y_n)$ mit $y_k = f(x_k), k = 1, 2, \dots, n$ der entsprechende Stichprobenvektor in Fahrenheit. Für die Mittelwerte erwarten wir die Beziehung $\bar{y} = f(\bar{x})$. Alles andere wäre eine Überraschung. Wie sieht man dies ein?

Satz 2.6. Sei $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ein Stichprobenvektor mit Mittelwert \bar{x} . Sei $f(x) := ax + b$ eine „Skalentransformation“ mit Konstanten a und b . Sei $\mathbf{y} = (y_1, y_2, \dots, y_n)$ mit $y_k = f(x_k), k = 1, 2, \dots, n$ der Stichprobenvektor in der anderen Skala. Dann gilt für den Mittelwert \bar{y} von \mathbf{y} die Beziehung $\bar{y} = f(\bar{x})$.

Beweis: Es gilt nach Definition

$$\bar{y} = \frac{\sum_{k=1}^n y_k}{n} = \frac{1}{n} \sum_{k=1}^n (a \cdot x_k + b).$$

Benutzt man die Rechenregeln für das Summensymbol, so erhält man

$$\bar{y} = \frac{1}{n} (a \sum_{k=1}^n x_k + bn),$$

woraus sich

$$\bar{y} = a\bar{x} + b = f(\bar{x})$$

ergibt. ■

Später, wenn wir Zufallsvariable kennen gelernt haben und \mathbf{x} mit der Zufallsvariable X sowie \mathbf{y} mit der Zufallsvariable Y verbinden, können wir einen funktionalen Zusammenhang $Y = f(X)$ zwischen den beiden Zufallsvariablen formulieren. I.A. gibt es für beliebige Abbildungen f keine einfache Formel $\bar{y} = f(\bar{x})$, die der Beziehung $EY = f(EX)$ zwischen den *Erwartungswerten* von X und Y entsprechen würde.

Mittelwerte von klassierten Daten

Bei kontinuierlichen Merkmalen führt eine Klassierung der Daten formal zu qualitativen Merkmalen, da die Ausprägungen Intervalle sind. Es ist jedoch naheliegend, die Mittelpunkte der Intervalle (*Klassenmitten*) als Merkmalsausprägungen zu nehmen, so dass man wieder ein quantitatives Merkmal erhält. Es stellt sich die Frage, wie sich die Klassierung auf die Mittelwerte auswirken. Die Klassierung kann diese durchaus verändern, wie wir in den Übungen sehen werden.

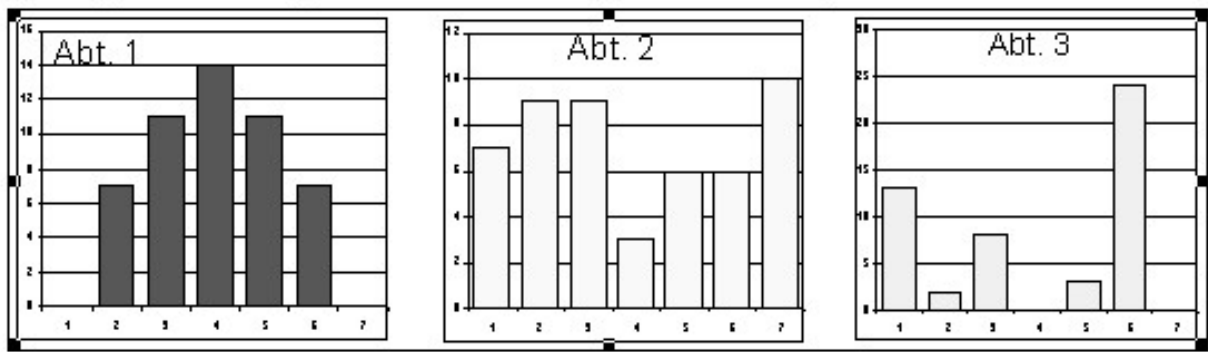


Abbildung 2.12: Drei verschiedene Häufigkeitsverteilungen mit demselben Mittelwert

2.4.5 Varianz, Streuung

Es kann qualitativ sehr unterschiedliche Häufigkeitsverteilungen mit dem gleichen Mittelwert geben, siehe Abb. 2.12.

Streuungsmaße sind Maßzahlen für die Abweichung der Messwerte vom Durchschnittswert. Die bekanntesten Streuungsmaße sind die die **Standardabweichung**

$$s := \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2},$$

die Wurzel aus der **Varianz** s^2 , und die Varianz selbst. Wir schreiben auch $V(\mathbf{x})$ für die Varianz s^2 und $Str(\mathbf{x})$ für die Standardabweichung, womit wieder der funktionale Zusammenhang zwischen Stichprobe und Kennzahl angedeutet wird.

Bemerkungen:

1. Die Berechnungen der Streumaße beruhen auf der **Fehlerquadratsumme**

$$q(x) := \sum_{k=1}^n (x_k - x)^2,$$

die uns schon in Satz 2.5 begegnet ist. Mit ihr drückt sich

$$s = \sqrt{\frac{1}{n-1} q(\bar{x})}$$

und

$$V(\mathbf{x}) = \frac{1}{n-1} q(\bar{x})$$

aus.

2. Wahrscheinlich wundern Sie sich über den Nenner $n - 1$ in der Formel für die Standardabweichung s . Es gibt auch Formeln in der Literatur mit dem Nenner n . Ein Grund für obige Formel ist, dass diese nur für $n \geq 2$ Sinn macht. Die alternative Formel würde für $n = 1$ stets den Wert $s = 0$ ergeben, was wenig sinnvoll erscheint.

Der tiefere Grund für den Nenner $n - 1$ liegt darin, dass so die Standardabweichung eine *erwartungstreue Schätzung* für die entsprechende Kenngröße der einer Zufallsvariablen ist.

Für den Nenner n spricht eine Vereinfachung der Formel in Gestalt von

$$s = \sqrt{\sum_{j=1, \dots, m} h_j (\omega_j - \bar{x})^2}$$

analog zu Satz 2.4.

3. Man kann sich fragen, wieso man die *Quadrate* der Abweichungen vom Mittelwert berechnet und aufaddiert (und schließlich die Wurzel zieht) und nicht einfach die Abstände $|\omega_j - \bar{x}|$. Die Antwort hängt mit Satz 2.5 zusammen, der einen Zusammenhang zwischen Mittelwert \bar{x} und der *Fehlerquadratsumme*

$$q(x) = \sum_{k=1}^n (x_k - x)^2$$

herstellt.

4. Es gibt noch weitere Streumaße wie z.B. der *Quartilsabstand* zwischen dem 1. und dem 3. Quartil, die in Kap. 2.4.8 definiert werden.

2.4.6 Skaleneinfluss

Hat man eine funktionale Beziehung $y_k = f(x_k)$, $k = 1, 2, \dots, n$ zwischen zwei Stichprobenvektoren \mathbf{x} und \mathbf{y} , so gilt i.A. für ihre Mittelwerte

$$\bar{y} \neq f(\bar{x}),$$

es sei denn, f hat eine Gerade als Funktionsbild, d.h., dass $f(x) = ax + b$ mit Konstanten $a, b \in \mathbb{R}$. In letzterem Fall rechnet man

$$V(\mathbf{y}) = |a|^2 V(\mathbf{x}), \quad Str(\mathbf{y}) = |a| Str(\mathbf{x})$$

nach — das b hat keinen Einfluss.

In Abb. 2.13 finden Sie die Preise für 1 Liter Dieselkraftstoff an 10 verschiedenen Tankstellen in Deutschland, Österreich und Tschechien.

<i>Deutschland</i>	<i>Österreich</i>	<i>Tschechien</i>
Preise in Euro	Preise in Euro	Preise in Kronen
0.863	0.689	20.9
0.875	0.695	21.1
0.889	0.710	21.6
0.889	0.710	22.3
0.889	0.715	22.3
0.895	0.719	22.4
0.895	0.721	22.5
0.899	0.725	22.8
0.901	0.743	22.9
0.905	0.779	23.0
$s_X = 0.012$	$s_X = 0.024$	$s_X = 0.700$
$V_X = 0.013$	$V_X = 0.034$	$V_X = 0.031$

Abbildung 2.13: Preise für eine Liter Diesel

Da ein Euro ca. 26 tschechische Kronen Wert ist, ist es nicht überraschend, dass die Streuung $s_X (= Str(\mathbf{x}))$ in Tschechien deutlich höher ausfällt als in Österreich und Deutschland. Da sich die Preise jedoch nicht durch direkte Umrechnung mit Hilfe des Wechselkurses berechnet, gibt $Str(\mathbf{y}) = |a|Str(\mathbf{x})$ mit $a = 26$ und den Euro-Preisen \mathbf{x} sowie den Kronen-Preisen \mathbf{y} nur einen Anhaltspunkt. Will man die Währung „herausrechnen“, so bietet es sich an, die Streuungen durch den Mittelwert (wenn dieser positiv ist) zu dividieren. Man nennt

$$V_X := \frac{s_X}{\bar{x}}$$

auch den **Variationskoeffizienten** von \mathbf{x} . Jetzt sieht man, dass die Benzinpreise in Österreich und Tschechien deutlich stärker streuen als in Deutschland!

Es gibt einen überraschenden Zusammenhang zwischen Varianz und Mittelwerte. Bezeichnet man nämlich mit $\overline{x^2}$ den Mittelwert von $\mathbf{x}^2 := (x_1^2, x_2^2, \dots, x_n^2)$, also von allen Quadraten der Daten, so gilt

$$V(\mathbf{x}) = \overline{x^2} - \bar{x}^2.$$

Hieraus folgt zum einen, dass i.A. $\overline{x^2} > \bar{x}^2$, d.h. dass schon für die simple Quadratfunktion $y = f(x) = x^2$ gilt, dass $\bar{y} \neq f(\bar{x})$.

2.4.7 Berechnungen mit Excel

Ich setze hier voraus, dass Sie im Prinzip den Einsatz einer „Formel“ in Excel kennen.

Man schreibe den Stichprobenvektor in die erste Spalte A1 : An. Nun wähle man den Menüpunkt *Einfügen* und hier *Funktion...*, sodann im Fenster *Funktion einfügen* die Funktionen MITTELWERT, STABW oder VARIANZ in Gestalt einer Formel MITTELWERT(A1:AN), STABW(A1:AN) oder VARIANZ(A1:AN).

Ein Testlauf ergibt, dass Excel ebenfalls die Standardabweichung mit dem Nenner $n - 1$ berechnet.

2.4.8 Median, Quantile, Quartile

Während der Mittelwert einer Stichprobe durch „Ausreißer“ beeinflusst wird, ist das für seinen „Gegenspieler“, den *Median*, nicht der Fall. Wenn ich als Prüfungsausschussvorsitzender des Studiengangs Mathematik-Diplom in einem Jahresbericht etwas zu den *Studienzeiten* der AbsolventInnen sage, gebe ich immer den Median an, weil dieser wesentlich aussagekräftiger ist als die mittlere Studienzeit. Denn ein Abschluss nach 40 (!) Fachsemestern lässt die mittlere Studienzeit gewaltig steigen, während er auf den Median kaum Einfluss hat. I.A. ist der Median von Studienzeiten kleiner als ihr Mittelwert, da die Verteilung der Studienzeiten linksgipflig ist.

Als „Arbeitsdefinition“ ist der **Median** der Stichprobe der kleinste Wert \tilde{x} ist, für den 50% der Stichprobenwerte x_k kleiner oder gleich \tilde{x} und die anderen 50% größer oder gleich \tilde{x} ausfallen. Hat man z. B. eine Stichprobe vom Umfang $n = 3$ und kennt $x_1 = 2, x_2 = 4$ und weiß man von x_3 nur, dass $x_3 > x_2 = 4$, dann kann man den Mittelwert nicht angeben, aber für den Median gilt $\tilde{x} = 4$, gleichgültig, wie groß x_3 ausfällt. Diese Beobachtung zeigt, dass der empirische Median robust ist gegenüber Ausreißern.

Die Berechnung des Medians ist besonders einfach, wenn man die Daten der Stichprobe der Größe nach zu

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

anordnet⁵. Die „Mitte“ dieser so angeordneten Daten ist dann der Median, genauer: Es gilt

$$\tilde{x} := \begin{cases} x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{falls } n \text{ gerade.} \end{cases}$$

Entsprechend ist das **25%-Quantil** (auch **1.Quartil** genannt) diejenige Zahl q , für die 25% der Stichprobenwerte x_k kleiner oder gleich q und die anderen 75% größer oder gleich q ausfallen. Statt von *Quantil* spricht man zuweilen auch vom **Perzentil**.

Es dürfte jetzt nicht schwerfallen, ein **75%-Quantil** (auch **3.Quartil** genannt) zu definieren und zu erkennen, dass der Median gerade das **50%-Quantil** (oder **2.Quartil**) ist.

Jetzt kann man auch zu $0 < p < 1$ **Quantile** Q_p als diejenigen Zahlen definieren, für die p Anteile der Stichprobenwerte x_k kleiner oder gleich Q_p und die anderen $1 - p$ Anteile größer oder gleich Q_p ausfallen.

Dann ist $Q_{0.5}$ der Median, $Q_{0.25}$ das 1.Quartil, $Q_{0.75}$ das 3.Quartil, etc.

Dass dies eine etwas unpräzise Arbeitsdefinition ist, erkennt man an dem Beispiel einer Notenverteilung mit 10 Einsen und 10 Zweien. Hier ist der Median $\tilde{x} = 1.5$. Als 25%-Quantil erscheint $q = 1$ die einzige sinnvolle Zahl. Es sind jedoch 50% der Stichprobenwerte x_k kleiner oder gleich $q = 1$ und 100% größer oder gleich $q = 1$.

⁵Dies ist eine raffinierte Bezeichnung. Durch $k \mapsto (k), k = 1, 2, \dots, n$ ist eine Permutation von $\{1, 2, \dots, n\}$ gegeben.

Bei den Pisa-Tests sind die 5%- und 95%-Quantile besonders interessant. Wenn ein Land A ein wesentlich höheres 5% (95%)-Quantil als ein anderes Land B hat, sind die schlechtesten (besten) Schüler in A wesentlich besser als die schlechtesten (besten) Schüler in Land B.

Bei Grenzwerten im Umweltschutz bedient man sich zuweilen auch der Quantile, etwa in dem folgenden Sinn: In 98% aller Kontrollmessungen darf der gemessene Wert einen bestimmten *Grenzwert* nicht überschreiten. Haben Sie von der neuen EU-Richtlinie in Bezug auf Feinstaub durch Dieselmotoren gehört? Danach darf ein Grenzwert von 50 Mikrogramm pro Kubikmeter nur an höchstens 35 Tagen im Jahr überschritten werden.

Um eine präzise Definition für Quantile zu geben, brauchen wir den Begriff (*empirische*) *Verteilungsfunktion*:

2.4.9 Empirische Verteilungsfunktion

Definition 2.7. Zu den Daten x_k einer Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ des Umfangs n gehört eine **empirischen Verteilungsfunktion** $F : \mathbb{R} \rightarrow \mathbb{R}$ definieren, wobei $F(x)$ der Anteil der Daten x_j mit $x_k \leq x$ ist:

$$F(x) := \frac{|\{k \in \mathbb{N} : 1 \leq k \leq n \text{ und } x_k \leq x\}|}{n}.$$

Dieser Begriff ist außerordentlich wichtig und wird bei kontinuierlichen Wahrscheinlichkeits-Verteilungen mit den Wahrscheinlichkeitsdichten korrespondieren.

Ist der Merkmalraum $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ mit

$$\omega_1 < \omega_2 < \dots < \omega_m,$$

so ist F eine monoton wachsende *Treppenfunktion*, deren Aussehen durch die relativen Häufigkeiten h_j von ω_j in der Stichprobe \mathbf{x} bestimmt ist. Es gilt $F(x) = 0$ für $x < \omega_1$, $F(\omega_1) = h_1$, sogar $F(x) = h_1$ für $\omega_1 \leq x < \omega_2$. Den nächsten *Sprung* macht F bei $x = \omega_2$: Es gilt $F(x) = h_1 + h_2$ für $\omega_2 \leq x < \omega_3$.

Schließlich gilt

$$F(x) = h_1 + h_2 + \dots + h_j \text{ für } \omega_j \leq x < \omega_{j+1} \quad (2.1)$$

und $F(\omega_m) = 1$, ja sogar $F(x) = 1$ für $x \geq \omega_m$.

In jedem Merkmal ω_j springt⁶ F von $h_1 + \dots + h_{j-1}$ auf $h_1 + \dots + h_{j-1} + h_j$. Vergleiche mit der Setzung $H_j := h_1 + \dots + h_{j-1} + h_j$ in Kap. 2.3.3. Aus diesem Grunde nennt man die empirische Verteilungsfunktion auch *kumulierte relative Häufigkeit*.

In Abb. 2.14 sehen Sie eine solche empirische Verteilungsfunktion einer Stichprobe von $n = 16$ Personen⁷, deren Körpergröße ermittelt wurden. Die Sprunghöhen müssen von der Form $h_j = \frac{r}{16}$ mit $r \in \mathbb{N}$ sein. Die größte Sprunghöhe ($r = 4$) ist bei 165 cm, d.h. vier Personen sind 165 cm lang.

⁶Ein echter Sprung liegt nur für $h_j > 0$ vor.

⁷aus JUMBO-Skript

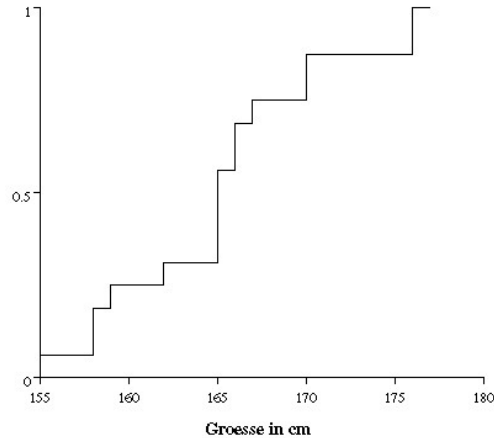


Abbildung 2.14: Eine empirische Verteilungsfunktion zu $n = 16$

Bemerkung: Sind die Daten zu $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ angeordnet, so gilt also $F(x_{(n)}) = 1$ und $F(x_{(k)}) \geq k/n$.

Wenn man die eingangs erwähnte Funktion $X : M \rightarrow \mathbb{R}$ mit der Menge $M := \{1, 2, \dots, n\}$ aller Individuen und $X(k) := x_k$ heranzieht, so kann man sich

$$F(x) := h(X \leq x)$$

vielleicht besser merken⁸. h steht wieder für die relative Häufigkeit, mit der hier die Daten $\leq x$ ausfallen. Beachten Sie, dass

$$h(X > x) = 1 - F(x),$$

da $H(X > x)$ die „komplementäre“ relative Häufigkeit zu $H(X \leq x)$ ist.

Die (empirische) Verteilungsfunktion einer Stichprobe ist auch geeignet, die Anteile der Daten der Stichprobe zu berechnen, die zwischen zwei vorgegebenen Werten liegen, genauer:

Satz 2.8. *Sei F die empirische Verteilungsfunktion einer Stichprobe \mathbf{x} . Es gelte $a < b$. Dann ist $F(b) - F(a)$ der Anteil der Stichprobendaten, die größer als a und kleiner gleich b ausfallen:*

$$F(b) - F(a) = \frac{|\{k \in \mathbb{N} : 1 \leq k \leq n \text{ und } a < x_k \leq b\}|}{n}.$$

Oder knapper:

$$F(b) - F(a) = h(a < X \leq b).$$

⁸In Worten: $F(x)$ ist die relative Häufigkeit, dass $X \leq x$. Diese Schreibweise passt nicht in die bisherigen, mit Funktionen zusammenhängenden Bezeichnungen und dient ausschließlich didaktischen Zwecken.

Jetzt kommt die präzise

Definition 2.9. Sei $0 < p < 1$. Eine Zahl Q_p heißt **p-Quantil** einer Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ genau dann, wenn

$$F(Q_p) \geq p \quad (2.2)$$

und

$$F(q) \leq p \text{ für } q < Q_p. \quad (2.3)$$

Dies ist gewöhnungsbedürftig. Vielleicht wird es klarer, wenn wir die Häufigkeitsfunktion h verwenden. Dann ist Q_p in (2.2) durch

$$h(X \leq Q_p) \geq p$$

und in (2.3) durch

$$h(X \leq q) \leq p \text{ für } q < Q_p$$

definiert. Letzteres kann man kürzer als

$$h(X < Q_p) \leq p$$

schreiben. Beachtet man noch, dass $h(X \leq Q_p) \geq p$ äquivalent zu $h(X > Q_p) \leq 1 - p$ ist, so erhalten wir eine äquivalente Charakterisierung von Q_p , die aber auch ein wenig haarspalterisch erscheint:

Satz 2.10. Sei $0 < p < 1$. Eine Zahl Q_p heißt **p-Quantil** einer Stichprobe $\mathbf{x} = (x_1, \dots, x_n)$ genau dann, wenn

$$h(X < Q_p) \leq p$$

und

$$h(X > Q_p) \leq 1 - p$$

Damit sind auch **Median** als 0.5-Quantil und die anderen Quartile als 0.25- und 0.75-Quantile definiert.

Während der Median anfangs eindeutig definiert wurde, ist dies jetzt nicht mehr der Fall. Es muss nur $h(X < \tilde{x}) \leq 0.5$ und $h(X > \tilde{x}) \leq 0.5$ zu gelten. Oder anders ausgedrückt: Der Median ist dadurch bestimmt, dass höchstens 50% der Daten kleiner und höchstens 50% der Daten größer als der Median sind.

Gehen wir zurück zu unserem Beispiel mit 10 Einsen und 10 Zweien. Jetzt ist *jede* Zahl q mit $1 \leq q \leq 2$ ein Median $Q_{0.5}$, während $Q_{0.25} = 1$ gilt. Häufig wird als *der* Median der Mittelpunkt des Median-Intervalls genommen. Dann wäre der Median 1,5. Entsprechendes gilt für die Quantile.

Die präzise Definition macht wohl Probleme, die man gut an Hand der Verteilungsfunktion in Abb. 2.14 erläutern kann. Das **p-Quantil** $q = Q_p$ ist in der ersten Arbeitsdefinition Lösung

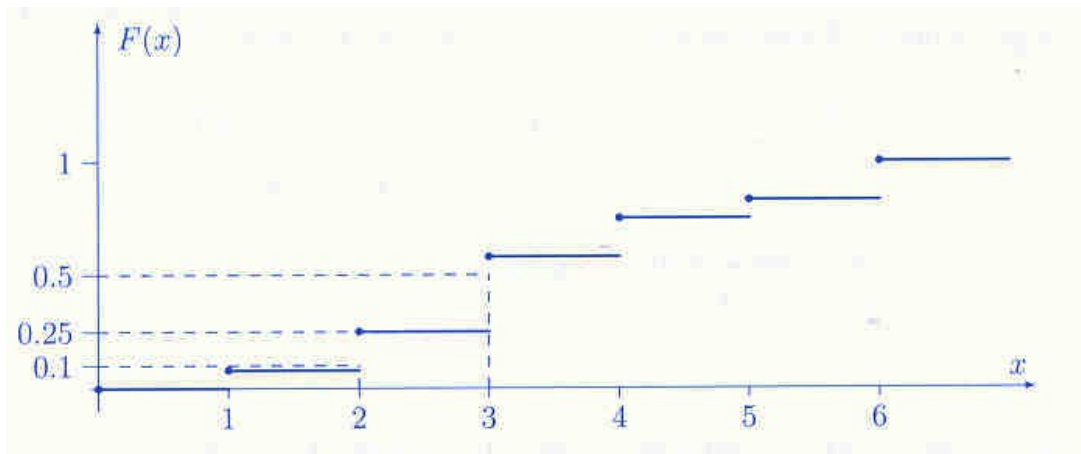


Abbildung 2.15: Empirische Verteilungsfunktion

der Gleichung $F(q) = h(X \leq Q_p) = p$. Da F aber „nur“ eine Treppenfunktion ist, gibt es hierbei Probleme. Trägt man p an der Ordinate ab und zeichnet eine Parallele zur x -Achse, so gibt es zwei Möglichkeiten: Entweder diese Parallele schneidet den Graphen von F zwischen zwei Stufen oder sie stimmt mit einer der Stufen überein. Im ersten (wahrscheinlichen) Fall ist $F(q) = p$ zwar nicht lösbar, aber das Quantil Q_p ist eindeutig der Wert, an dem der Graph von F das Niveau p überspringt. Dieser Wert ist dann für ein ganzes Intervall von p -Werten ein Quantil. Im zweiten Fall gibt es ein Intervall von p -Quantilen, nämlich alle x -Werte, die die Stufe in p -Höhe definieren.

Diese wird sehr schön deutlich an Hand der Abb. 2.15, die die Verteilungsfunktion F von 12 Klausurnoten zeigt (aus FISCHER).

$F(x) = \frac{1}{2}$ hat keine Lösung, der Median ist daher $\tilde{x} = 3$. Das 25%-Quantil ist jedoch nicht eindeutig, da $F(x) = 0.25$ durch alle $x \in [2, 3)$ gelöst wird. Hier ist jedes $Q_{0.25} \in [2, 3]$ (auch 3!!!) ein 25%-Quantil. Allerdings nimmt man in der Praxis die Mitte dieses Intervalls, bezeichnet also $Q_{0.25} = 2.5$ als 25%-Quantil.

Bemerkung: Die Excel-Funktion QUANTIL() liefert andere Werte! Offensichtlich wird die Funktion $p \mapsto Q_p$ an den Sprungstellen etwas verschmiert.

Schauen Sie sich Abb. 2.14 an. Das 1. Quartil liegt bei 159 cm, wenn man letzterem Vorschlag folgt. Wendet man aber Def. 2.9 an, so ist jede Zahl aus $[159, 162]$ ein mögliches 1. Quartil, so dass deren Mittelwert, also 160.5, auch als 1. Quartil bezeichnet werden könnte. Der Median liegt bei 165 cm, das 3. Quartil bei 167 cm.

Bemerkung: Zuweilen erhält die Verteilungsfunktion F einen Index n , und man schreibt F_n an Stelle von F . Dies soll daran erinnern, dass der Umfang n der Stichprobe ein wichtiger Parameter ist, der die Aussagekraft der Empirik bestimmt. Je größer n , desto aussagekräftiger. In der Theorie der Statistik lässt man gedanklich n gegen ∞ konvergieren und untersucht die

„Konvergenz“ der empirischen Verteilungsfunktion gegen die „wahre“ Verteilungsfunktion der Grundgesamtheit. Der *Zentrale Grenzwertsatz* sagt aus, dass unter gewissen Bedingungen F_n gegen die Verteilungsfunktion einer *Normalverteilung* konvergiert.

Die Maßzahlen dieser Grenz-Verteilungsfunktion werden durch die Maßzahlen der Stichprobe geschätzt. Daher sollte man dem Begriff *Median*, *Quartil*, *Quantil*,.. noch den Zusatz *empirisch* geben.

Was halten Sie von folgender Aussage: „Skandal: Über die Hälfte der Haushalte in Deutschland verdienen weniger als im Durchschnitt!“.

Nun, die mathematische Formulierung lautet: Bei der Verteilung der Haushalte in Deutschland nach deren Einkommen ist der Median kleiner als der Mittelwert. Das ist ganz sicher so, da die wenigen Haushalte mit riesigem Einkommen zwar den Mittelwert, aber nicht den Median nach oben treiben. Können Sie sich noch weitere Beispiele für Verteilungen finden, bei denen von vornerein klar ist, dass der Median kleiner oder größer als der Mittelwert ist?

2.5 Statistik und Wahrscheinlichkeitstheorie

Die beschreibende Statistik kommt ohne den Begriff *Wahrscheinlichkeit* aus. An dessen Stelle tritt der hiermit verwandte Begriff *relative Häufigkeit*. Später werden wir sehen, dass an die Stelle einer *Häufigkeitsverteilung* die *Wahrscheinlichkeitsverteilung* und an die Stelle der *empirischen Verteilungsfunktion* die *Verteilungsfunktion* tritt. Statt mit einer *Stichprobe* hat man es dann mit einer *Zufallsvariablen* zu tun, wobei der *Mittelwert* zum *Erwartungswert* und die *Standardabweichung* zur *Streuung* wird. Die Namen *Varianz*, *Median*, *Quantile*, *Quartile* etc. bleiben unverändert. Eigentlich müssten alle statistischen Begriffe den Zusatz *empirisch* tragen.

Die Verbindung zur Wahrscheinlichkeitstheorie wird auch über den Zufallsaspekt einer *Stichprobe* hergestellt. Der Name *Stichprobe* gibt eigentlich nur bei einer *Teilerhebung* einen Sinn. Ziel einer Teilerhebung kann nur sein, von einer Teilgesamtheit auf die Grundgesamtheit zu schließen. Wenn wir einmal unterstellen, dass die Anzahl der Individuen der Grundgesamtheit endlich ($=N$) ist, so ist das eigentliche Ziel einer Teilerhebung vom Umfang $n \ll N$ auf die „wahre“ Häufigkeitsverteilung mit ihren Maßzahlen zu schließen. Die „wahren“ Werte könnte man theoretisch durch eine Totalerhebung erhalten. Für kleine N ist dies auch durchaus möglich, wenn man etwa an die Studierendenstatistik aller an der Uni HH eingeschriebenen Studierenden denkt. Oder an die Statistik der Zwischenprüfung. Wenn es aber um die durchschnittliche Quadratmeterzahl von Wohnungen oder um den genauen Anteil von Personen weiblichen Geschlechts an der Bevölkerung in Deutschland denkt, so ist kann man diese Zahlen nicht genau erheben, zumal sich ja ständig etwas ändert (Wohnungsbau, Geburten, Todesfälle,...). Hier bedarf es Stichproben, mit deren Hilfe die wahren Werte *geschätzt* werden sollen. Jetzt wird der „wahre“ Wert für den Anteil von Personen weiblichen Geschlechts zu einer Wahrscheinlichkeit, bei einer zufällig ausgewählten Person die Merkmalsausprägung „weiblich“ festzustellen. Man kann von einem Zufallsexperiment oder einer Zufallsvariablen sprechen, das in der Feststellung des Geschlechts einer zufällig ausgewählten Person besteht.

Noch deutlicher wird die Verbindung zur Wahrscheinlichkeitsrechnung, wenn man sich eine unendliche Grundgesamtheit vorstellt. Zum Beispiel *alle* möglichen Würfe mit 2 Würfeln verbunden mit der Ermittlung der Augensumme. Hier kann es keine Totalerhebung geben, die etwa die relative Häufigkeit für die Augensumme=10 ergibt. Hier macht man ein Gedankenexperiment mit einer Stichprobe vom Umfang n , der gegen ∞ konvergiert und postuliert — das ist das Gesetz der großen Zahl —, dass die relative Häufigkeit der Stichprobe für $n \rightarrow \infty$ konvergiert. Der Grenzwert heißt dann *Wahrscheinlichkeit* für das *Ereignis* „Augensumme=10“.

Historisch ist die Wahrscheinlichkeitsrechnung eng mit dem Glücksspiel verbunden. Dies kann aber den Blick sehr verengen. So hat man eine unendliche Grundgesamtheit auch bei allen zukünftigen Geburten in Deutschland. Interessiert man sich für das Gewicht von Neugeborenen, so ist deren Verteilung also unbekannt. Aus vorliegenden Zahlenmaterialien kann man jedoch ziemlich genaue Annahmen über diese Verteilung machen, z.B., dass eine Normalverteilung mit einem bestimmten Erwartungswert und einer bestimmten Streuung vorliegt. Unter dieser Annahme ist die Wahrscheinlichkeit, dass ein Neugeborenes zwischen 3000g und 5000g wiegt, bekannt.

Interessant ist es nun, wenn z.B. in einem Krankenhaus „Abweichungen von der Norm“ festgestellt werden, wenn also über einen längeren Zeitraum deutlich andere Mittelwerte als der Erwartungswert gemessen wurden. Dann stellt sich die Frage: Ist dies Zufall? Oder gibt es systematische Gründe für die Abweichung (Umwelt,...)?

Solche Fragen geben Anlass zu *Hypothesen*, die mit Hilfe *statistischer Tests* untersucht werden.

2.6 Korrelation zweier Merkmale

Der Begriff *Korrelation* hat Eingang in die Umgangssprache gefunden. „Es gibt eine hohe Korrelation zwischen Terrorismus und Armut in der Welt, zwischen der sozialen Lage von Jugendlichen und ihrem PISA-Abschneiden, zwischen Übergewicht und Bluthochdruck, zwischen Lungenkrebs und Rauchen,...“. Diese Liste könnte endlos weitergeführt werden.

Ziel dieses Kapitels ist es, den Begriff *Korrelation* mathematisch präzise zu definieren. Hierzu wird ein *Korrelationskoeffizient* r zwischen zwei Stichprobenvektoren zu verschiedenen Merkmalen eingeführt, der sich im wesentlichen als Skalarprodukt zweier Vektoren entpuppt. Zwei Merkmale können dann als *korreliert* angesehen werden, wenn bei großen Stichprobenumfängen n der Betrag $|r|$ des *Korrelationskoeffizienten* nahezu Eins ist. Ist er dagegen nahe Null, so gelten die Merkmale als *unkorreliert*.

Bei der Korrelation geht es um die *Abhängigkeit* zweier Merkmale — eine Vorstufe der *stochastischen Unabhängigkeit* zweier Zufallsvariablen.

Die bisherigen Auswertungsmethoden beschränkten sich auf die Betrachtung *eines* Merkmals. Hier werden erstmalig mehrere (zwei) Merkmale in die Auswertung einbezogen. Ziel ist es, deren Abhängigkeit zu untersuchen.

Noch eine Bemerkung zur grafischen Darstellung von zwei Merkmalen: Hier gibt es auch eine Fülle von Möglichkeiten. Gehen Sie einmal auf die Suche in den Tageszeitungen! Die Abb. 2.1

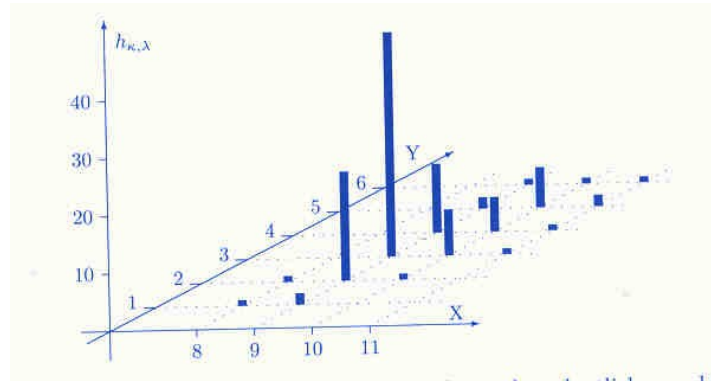


Abbildung 2.16: 3D-Balkendiagramm Studiendauer (X) versus Abschlussnote (Y)

fällt darunter.

2.6.1 Kontingenztafel

Die beiden Stichprobenvektoren \mathbf{x} und \mathbf{y} wurden an denselben Individuen erhoben, haben also gleiche Länge n . Man kann die Daten zum k -ten Individuum zu einem *Datenpunkt* $P_k := (x_k, y_k) \in \mathbb{R}^2, k = 1, 2, \dots, n$ zusammenfassen kann. Jeder solcher Datenpunkt stimmt mit einem Merkmalpaar (α_i, β_j) überein, wenn $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ der (diskrete) Merkmalraum von \mathbf{x} und $\{\beta_1, \beta_2, \dots, \beta_\ell\}$ der von \mathbf{y} ist. Dann ist $\Omega := \{\alpha_1, \alpha_2, \dots, \alpha_m\} \times \{\beta_1, \beta_2, \dots, \beta_\ell\}$ der Merkmalraum der Datenpunkte. Jede Merkmalausprägung (α_i, β_j) hat nun auch eine absolute und eine relative Häufigkeit, die man doppelindizieren sollte (n_{ij} bzw. h_{ij}), und die man zu einer Matrix in Form einer $m \times \ell$ -Tabelle zusammenstellen kann. Sie heißt **Kontingenztafel**. Diese kann man auf mannigfache Weisen grafisch darstellen. In Abb. 2.16 (aus FISCHER) werden $n = 100$ Studierende im Hinblick auf Studiendauer und Abschlussnote grafisch erfasst.

2.6.2 Korrelationskoeffizient

Gegeben seien zwei verschiedene quantitative Merkmale, die durch eine Stichprobe an denselben Objekten ermittelt werden. Mathematisch führt dies zu zwei Stichprobenvektoren $\mathbf{x} \in \mathbb{R}^n$ und $\mathbf{y} \in \mathbb{R}^n$ des gleichen Umfangs n mit Mittelwerten \bar{x} und \bar{y} .

Anschaulich würde man von Korrelation reden, wenn positive (negative) $x_k - \bar{x}$ auch stets positive (negative) $y_k - \bar{y}$ zur Folge hätten. Vielleicht nicht immer, aber häufig. Hier bietet sich an, den Wert

$$S_{xy} := \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) \quad (2.4)$$

zu betrachten. Für $\mathbf{x} = \mathbf{y}$ kennen wir diesen Ausdruck von der empirischen Varianz

$$s_{xx} = \frac{1}{n-1} S_{xx}.$$

Analog nennt man

$$s_{xy} := \frac{S_{xy}}{n-1} \quad (2.5)$$

die (**empirische**) **Kovarianz** der beiden Merkmale.

Erkennen Sie das Skalarprodukt (siehe Kap. II.3.2) in (2.4)? Ich nenne die „zurecht gerückten“ Vektoren⁹

$$\begin{aligned} \mathbf{x}^m &:= (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}), \\ \mathbf{y}^m &:= (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y}). \end{aligned}$$

Dann ist

$$S_{xy} = \mathbf{x}^m \cdot \mathbf{y}^m,$$

gerade das Skalarprodukt der beiden „zurecht gerückten“ Stichprobenvektoren \mathbf{x}^m und \mathbf{y}^m .

Die Kovarianz ändert sich nun offensichtlich nicht, wenn man den Ursprung der Maßeinheiten für die beiden quantitativen Merkmale verschiebt. Um sie aber auch skalierungsunabhängig zu machen, machen wir etwas, was wir aus der Linearen Algebra kennen: Wir *normieren* die beiden Vektoren \mathbf{x}^m und \mathbf{y}^m zur Länge Eins, d.h. wir berechnen

$$\begin{aligned} S_{xx} &:= |\mathbf{x}^m|^2 = \sum_{k=1}^n (x_k - \bar{x})^2, \\ S_{yy} &:= |\mathbf{y}^m|^2 = \sum_{k=1}^n (y_k - \bar{y})^2, \end{aligned}$$

normieren \mathbf{x}^m und \mathbf{y}^m zur Länge Eins durch $\mathbf{x}^m/|\mathbf{x}^m|$ und $\mathbf{y}^m/|\mathbf{y}^m|$ und definieren:

Definition 2.11. *Der Korrelationskoeffizient r_{xy} zwischen den Stichprobenvektoren \mathbf{x} und \mathbf{y} ist durch das Skalarprodukt*

$$r := r_{xy} = (\mathbf{x}^m / \sqrt{S_{xx}}) \cdot (\mathbf{y}^m / \sqrt{S_{yy}})$$

gegeben.

Man erkennt sofort, dass

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}},$$

also (siehe (II.3.1))

$$r_{xy} = \cos \angle(\mathbf{x}^m, \mathbf{y}^m).$$

Aus der Linearen Algebra wissen wir, dass $|r_{xy}| \leq 1$ – dies folgt aus Satz II.3.3. (Cauchy-Schwarzsche Ungleichung). Korreliertheit kann in der Sprache der Linearen Algebra auch so ausgedrückt werden: Der Winkel zwischen \mathbf{x}^m und \mathbf{y}^m ist nahe Null ($r_{xy} = 1$) oder π ($r_{xy} = -1$) (im letzteren Fall spricht man auch von *anti-korreliert*). Ist der Winkel nahe einem rechten, so kann man von Unkorreliertheit sprechen ($r_{xy} = 0$).

⁹Ihre Mittelwerte verschwinden.

2.6.3 Regression

Angenommen, man hätte an dem selben Ort eine Messreihe mit Celsius- und eine mit Fahrenheittemperaturen erstellt und so zwei Stichprobenvektoren \mathbf{x} und \mathbf{y} erhalten. Würde es Sie sehr überraschen, dass die beiden Temperaturmerkmale hoch korreliert sind? Sicher nicht.

Dies kann begründet werden: Sei

$$y_k = ax_k + b, k = 1, 2, \dots, n$$

mit Skalierungs-Konstanten $a \neq 0, b$. Wir wissen schon, dass

$$\bar{y} = a\bar{x} + b.$$

Hieraus gewinnen wir

$$y_k - \bar{y} = a(x_k - \bar{x}),$$

also $\mathbf{y}^m = a \cdot \mathbf{x}^m$, d.h. \mathbf{y}^m ist gerade das a -fache von \mathbf{x}^m . Damit gilt für die normierten Vektoren

$$\mathbf{x}^m / |\mathbf{x}^m| = \pm \mathbf{y}^m / |\mathbf{y}^m|,$$

je nachdem, ob $a > 0$ oder $a < 0$. Wie erwartet ist der Korrelationskoeffizient $r_{xy} = 1$ für $a > 0$ und $r_{xy} = -1$ für $a < 0$.

Es gilt sogar die Umkehrung:

Satz 2.12. *Es gilt $|r_{xy}| = 1$ genau dann, wenn es $a \neq 0$ und $b \in \mathbb{R}$ gibt mit $y_k = ax_k + b, k = 1, 2, \dots, n$. Dabei gilt $r_{xy} = 1$ für $a > 0$ und $r_{xy} = -1$ für $a < 0$.*

Grafisch kann man die Korreliertheit von \mathbf{x} und \mathbf{y} auch daran erkennen, dass die *Datenpunkte* $P_k := (x_k, y_k) \in \mathbb{R}^2, k = 1, 2, \dots, n$ (sie bilden eine **Punktvolke**) auf einer Geraden liegen.

Auch wenn die beiden erhobenen Merkmale korreliert sind, werden diese Datenpunkte P_k nur angenähert auf einer Geraden liegen. Die auf GAUSS zurückgehende Methode, die Konstanten a und b , die die Korrelation herstellen, zu bestimmen, heißt *Lineare Regression*¹⁰. Dies ist eine Methode, mit der eine Gerade der Gleichung $y = ax + b$ angenähert durch alle Datenpunkte gelegt wird. Sie führt auf eine Formel

$$a = \frac{s_{xy}}{s_{xx}},$$

die im Zähler die *empirische Kovarianz* (2.5) enthält.

Diese Methode wäre auch bei der Bestimmung eines Potenzwachstums nützlich gewesen, wobei in einem log-log-Diagramm eine Gerade durch Messpunkte gelegt wird, deren Steigung gerade der Exponent des Potenzwachstums ist, siehe Kap. III.2.4.2 und Übungsaufgabe III.37.

In Abb. 2.17 sehen Sie eine Punktvolke zum diastolischen und systolischen Blutdruck von 14 Patienten.

¹⁰Möglicher Proseminarvortrag

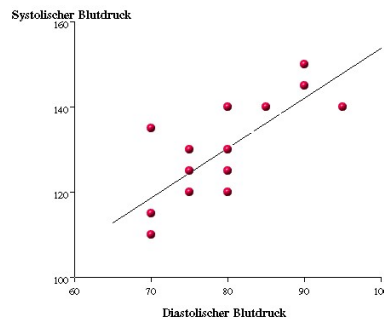


Abbildung 2.17: Regressionsgerade Blutdruck

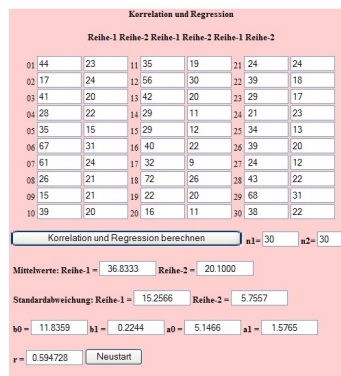


Abbildung 2.18: Korrelationskoeffizient Klausuren

Zwischenprüfung und erster Test

Mit dem **Java-Skript Korrelation und Regression (JUMBO)** kann man den Korrelationskoeffizienten zwischen zwei Stichprobenvektoren ausrechnen. Ich habe dies einmal für eine Teil der Klausurergebnisse getan, wobei das erste Merkmal die erste Zwischenprüfung und das zweite Merkmal der erste Test im WiSe 04/05 ist. In Abb. 2.18 sehen Sie die Daten und die Ergebnisse. Es kann in dem Java-Skript von JUMBO nur ein Stichprobenumfang von maximal $n = 30$ gewählt werden. Das Ergebnis ist ein Korrelationskoeffizient von $r = 0.5947$. Was bedeutet dies? Wäre $r = 1$, hätte der Test nur die Zwischenprüfungsergebnisse bestätigt. Aber einige, die in der ersten ZP-Klausur unter dem Schnitt waren, waren im Test besser als der Durchschnitt und umgekehrt. Aber „im Mittel“ gibt es dieselbe Tendenz, wenn auch nicht sehr stark ausgeprägt. Abb. 2.19 zeigt eine typische Punktwolke mit $r = 0.6$.

Wenn ich mich auf alle diejenigen beschränke, die in beiden Klausuren mindestens ausreichend abschnitten, sieht das Ergebnis schon anders aus: Jetzt ist $r = 0.743$.

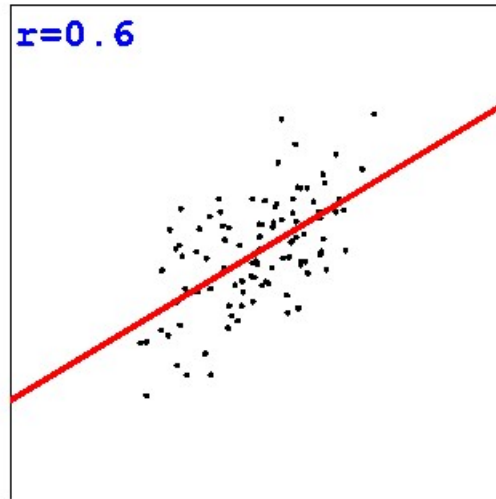


Abbildung 2.19: Punktwolke beim Korrelationskoeffizienten $r = 0.6$

2.6.4 Eine Warnung

Auf keinen Fall darf man aus einer Korrelation zweier Merkmale auf Grund eines Korrelationskoeffizienten $r_{xy} \approx 1$ schließen, dass es einen kausalen Zusammenhang gibt. Berühmtestes Beispiel: Wenn man die Entwicklung der Storchpopulationen mit den Geburtenraten in den letzten Jahren vergleicht, gibt es eine relativ hohe Korrelation. Auch dass, Lungenkrebs und Rauchen deutlich korreliert ist, ergibt nicht zwingend, dass Rauchen für Lungenkrebs verantwortlich ist.

Kapitel 3

Einführung in die Wahrscheinlichkeitsrechnung

3.1 Einführung

Im einführenden Kapitel über *Beschreibende Statistik* klangen schon Bezüge zur Wahrscheinlichkeitsrechnung an, siehe Kap. 2.5. Auch hier gibt es einen **Merkmalsraum** Ω , dessen Elemente Ergebnisse von **Zufallsexperimenten** ist, zu denen auch eine zufällige Stichprobe gezählt werden kann. Man halte sich nur das schon abgestandene Beispiel des Würfels vor Augen, das auf $\Omega = \{1, 2, 3, 4, 5, 6\}$ führt oder den Münzwurf mit den beiden Merkmalsausprägungen „Wappen“ und „Zahl“. Etwas allgemeiner ist der Begriff der **Zufallsvariablen**, die an die Stelle der *Datenfunktion* X in der beschreibenden Statistik tritt.

In der Wahrscheinlichkeits-Theorie führt das *empirische Gesetz der großen Zahl* zu dem Begriff **Wahrscheinlichkeit** eines gewissen **Ereignisses** und zum Begriff **Wahrscheinlichkeitsmodell** mit einer **Verteilung**, die uns schon als Häufigkeitsverteilung in der Statistik mit den *relativen Häufigkeiten* an Stelle der *Wahrscheinlichkeiten* begegnet ist. Die wichtigsten Wahrscheinlichkeits-Modelle führen auf die **Binomialverteilung**, die **Poisson-Verteilung** und die **Normalverteilung**. Letztere erhält ihre überragende Bedeutung durch den *Zentralen Grenzwertsatz*. Zu jeder Verteilung gehören *Kenngrößen* einer zugehörigen **Zufallsvariablen** wie **Erwartungswert**, **Varianz**, **Streuung**, **Median** und **Quantile**.

An die Stelle der *empirischen Verteilungsfunktion* tritt jetzt die **Verteilungsfunktion**, die bei kontinuierlichen Verteilungen auch eine **Wahrscheinlichkeitsdichte** haben kann.

3.1.1 Elementare Anwendungen der Kombinatorik auf die Wahrscheinlichkeits-Rechnung

Wir hatten in Kap. I.7 (*Abschnitt Kombinatorik*) vier Grundmuster kennengelernt. Es ging um eine Urne mit n Kugeln (Menge $A = \{a_1, \dots, a_n\}$), aus der man nacheinander in k Zie-

hungen Kugeln zieht. Dabei unterscheidet man *mit* und *ohne Zurücklegen* und *mit*¹ und *ohne*² Berücksichtigung der *Reihenfolge*. Die Formeln für die Anzahl von möglichen Ziehungsergebnissen lauten n^k (mit Reihenfolge, mit Zurücklegen), $(n)_k := n \cdot (n-1) \cdots (n-k+1)$ (ohne Zurücklegen, mit Reihenfolge) mit dem Spezialfall $n!$ bei $k = n$ (Permutationen), $\binom{n}{k}$ (ohne Zurücklegen, ohne Anordnung) sowie $\binom{n+k-1}{k}$ (mit Zurücklegen, ohne Reihenfolge). In Kap. 3.2 werden einige aus Sicht der Wahrscheinlichkeitsrechnung wesentliche Punkte wiederholt.

Will man diese „Zählkunst“ auf die Berechnung von Wahrscheinlichkeiten anwenden, so muss man annehmen, dass *jedes Ziehungsergebnis gleich wahrscheinlich* ist. Bezeichnen wir mit Ω die von k und n abhängende Menge aller möglichen Ziehungsergebnisse und betrachten eine Teilmenge $E \subset \Omega$, so interessieren wir uns für die Wahrscheinlichkeit, dass ein Ziehungsergebnis zur „Ergebnismenge“ E gehört. Diese **Wahrscheinlichkeit** ist

$$p = \frac{|E|}{|\Omega|}, \quad (3.1)$$

das Verhältnis der *günstigen* zu *allen* Möglichkeiten. Siehe auch das unten aufgeführte „Laplace-Experiment“ zur rigorosen axiomatischen Begründung dieses Wahrscheinlichkeits-Begriffs.

Diese „Gleich-Wahrscheinlichkeit“ der Ziehungsergebnisse ist der Grund, warum der schwierigste Fall der vier Grundmuster (mit Zurücklegen, ohne Reihenfolge) in der Wahrscheinlichkeitsrechnung meines Wissens keine Anwendung hat. Zum Beispiel sind nicht alle „Torschützenlisten“ mit 5 Toren gleich wahrscheinlich, auch wenn man unterstellt, dass jeder Stürmer mit gleicher Wahrscheinlichkeit trifft. Aus demselben Grund, warum es unwahrscheinlicher ist, zwei Einsen als eine Eins und eine Zwei zu würfeln.

Beispiele

Hier sollen Wahrscheinlichkeiten auf Grund der Formel (3.1) berechnet werden.

1. Ein Alphabet bestehe aus 5 Buchstaben. Wie groß ist die Wahrscheinlichkeit, dass ein „zufällig“ gebildetes Wort der Länge drei mit genau zwei Buchstaben auskommt? Es ist $n = 5$ und $k = 3$, es wird zurückgelegt, und es kommt auf die Reihenfolge der Buchstaben an. Es gibt $|\Omega| = n^k = 5^3 = 125$ Worte der Länge drei. Legt man sich auf den doppelt vorkommenden und den einfach vorkommenden Buchstaben fest, so gibt es immer noch drei verschiedene Worte mit diesen beiden Buchstaben. Es gibt $20 = 5 \cdot 4$ mögliche Buchstaben-Paare, von denen der erste doppelt und der zweite einfach vorkommen soll. Also gibt es $|E| = 3 \cdot 20 = 60$ Möglichkeiten von Worten der Länge drei mit zwei Buchstaben. Die Wahrscheinlichkeit ist $|E|/|\Omega| = 60/125 = 12/25$, rund 50%, wenn wir unterstellen, dass jeder Buchstabe an jeder Stelle des Wortes gleichwahrscheinlich vorkommt³.

¹Variationen

²Kombinationen

³Man kann auch „komplementär“ rechnen. Es gibt $5 \cdot 4 \cdot 3 = 60$ Worte mit 3 verschiedenen Buchstaben und 5 Worte mit drei gleichen Buchstaben. Also sind $125 - 65 = 60$ Worte von dem gewünschten Typ.

2. Wir haben $n = 5$ gleich „treffsichere“ Fußballspieler, die $k = 3$ Tore schießen. Wie groß ist die Wahrscheinlichkeit, dass ein Spieler alle drei Tore erzielt? Eine Antwort könnte so aussen: Es gibt $\binom{n+k-1}{k} = \binom{7}{3} = 35$ verschiedene Torschützenlisten (mit Zurücklegen, ohne Reihenfolge). 5 von ihnen enthalten nur einen Namen. Also ist die Wahrscheinlichkeit $p = 1/7$.

Kommt diese Wahrscheinlichkeit einem nicht zu groß vor? Was halten Sie von folgender Alternativrechnung? Wir achten auf die Reihenfolge der Torschützen. Dann gibt es $n^k = 125$ verschiedene Torschützenlisten. Von denen gibt es 5, in denen nur ein Schütze auftritt. Also ist die Wahrscheinlichkeit $5/125=1/25$.

Was ist der Grund für den Widerspruch? Im ersten Fall haben wir übersehen, dass nicht alle Torschützenlisten gleich wahrscheinlich sind. In der zweiten Rechnung wurde richtig argumentiert, sofern alle Spieler mit gleicher Wahrscheinlichkeit treffen.

3. Wie groß ist die Wahrscheinlichkeit, dass 2 Buben im Skat liegen? Ein Kartenspiel hat 32 Karten, der Skat besteht aus 2 Karten. Also gibt es $\binom{32}{2} = 496$ mögliche (und gleichwahrscheinliche) Skate. Es gibt 4 Buben, also $\binom{4}{2} = 6$ verschiedenen Möglichkeiten für einen reinen Bubenskat. Also ist die Wahrscheinlichkeit gleich $6/496 = 3/248$.

Anders sieht es aus der Sicht eines Spielers aus, dessen 10 Karten keinen Buben enthalten. Hier gibt es nur noch $\binom{22}{2} = 231$ Möglichkeiten für einen Skat, die Wahrscheinlichkeit ist $6/231$.

4. Das allseits bekannte (?) **Geburtstagsproblem**: Wie groß ist die Wahrscheinlichkeit, dass in einer Gruppe von k Menschen mindestens zwei am selben Tag Geburtstag haben?

Ich gehe von $n = 365$ Tagen aus. Jede Person „zieht“ einen Tag. Wir stellen uns das Ziehungsergebnis als k -Tupel von Daten vor. Von diesen gibt es $|\Omega| = 365^k$ Stück. Von diesen führen $(365)_k := 365 \cdot 364 \cdot \dots \cdot (365 - k)$ auf lauter verschiedene Tage, also sind $|E| = 365^k - (365)_k$ Ziehungsergebnisse in unserem Fall „günstig“. Die gesuchte Wahrscheinlichkeit ist

$$p(k) := \frac{|E|}{|\Omega|} = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - k)}{365^k},$$

eine Zahl, die für relativ kleine k schon ziemlich groß ist.

k	5	10	20	22	23	24	50	70	100
p(k)	0.027	0.117	0.411	0.476	0.507	0.538	0.970	0.999	1.000

3.1.2 Beispiele für Fragen aus der Stochastik

Wir wollen hier an Hand von Beispielen gewissen stochastischen Fragestellungen nachgehen, um für die nachfolgenden Begriffsbildungen zu motivieren. Dabei werden wir keine vollständigen Antworten auf die Fragen anstreben, jedoch einen ersten Gewöhnungsprozess in Bezug auf die neuen Begriffe einleiten.

Es sollen nicht Glücksspiele im Vordergrund stehen, sondern „echte“ Fragestellungen aus den Anwendungen.

Umfrageergebnisse

Wir beginnen mit einem Beispiel aus der *Schließenden Statistik*.

Eine Umfrage bei n angeblich repräsentativen Personen ergibt für die 5 Parteien SPD, CDU, FDP, Grüne und PDS ein gewisses Ergebnis. Für die FDP laute dieses 7%.

Kann man (vorausgesetzt, es haben alle die Wahrheit gesagt) hieraus schließen, dass die FDP mit einer Wahrscheinlichkeit von 95% mindestens 6% erhalten würde? Anders gefragt: Gibt es ein **Konfidenzintervall**, in dem der wahre Stimmanteil der FDP mit einem angebbaren *Signifikanzniveau* liegt? Die Antwort hängt natürlich von dem Umfang der Umfrage n ab.

Wie groß ist die Wahrscheinlichkeit für ein solches (irreführendes) Umfrageergebnis, wenn in Wirklichkeit nur 5% der Wähler die FDP wählen würden⁴? Dieselbe Frage lautet aus der Sicht der Statistik: Kann die *Hypothese*, dass die FDP auf einen Stimmanteil von 5% oder weniger kommt, mit einer gewissen *Irrtums-Wahrscheinlichkeit* widerlegt werden, wenn in einer Stichprobe von n Personen sich 7% für die FDP aussprechen?

Bei einer späteren Erörterung solcher Fragen wird unterstellt werden, dass zum einen die Befragenden zufällig ausgewählt wurden, d.h. dass jede WählerIn mit gleicher Wahrscheinlichkeit befragt wird und dass die Angaben genau dem tatsächlichen Wahlverhalten entsprechen. Beide Annahmen treffen in der Realität nicht zu. Daher werden die Umfrageergebnisse auch nachträglich „manipuliert“, wobei die Erfahrungen früherer Umfragen und der Vergleich mit den tatsächlichen Wahlergebnissen einfließen.

Schadhafte Produkte

Eine Bierflasche wird beim Abfüllen und dem späteren Abpacken in einen Bierkasten mit einer bekannten (kleinen) Wahrscheinlichkeit von $p\%$ beschädigt. Ein Kasten Bier enthält 24 Flaschen. Wie groß ist die Wahrscheinlichkeit, dass ein Kasten Bier mindestens eine beschädigte Flasche enthält?

Dies ist eine reine Wahrscheinlichkeits-Frage. Will man die Fehlerquote p ermitteln, kommt die Statistik ins Spiel: mit Hilfe einer Stichprobe eines geeigneten Umfangs wird p „geschätzt“.

Krebs

Der Anteil der Krebskranken einer (Vorsorge-) Untersuchung sei $p\%$. Die Methode zur Krebserkennung liefert bei nicht an Krebs Erkrankten ein medizinisch positives Resultat bei $q\%$ der Patienten, bei den Krebskranken selbst sei dieser Test sicher. Wie groß ist die Wahrscheinlichkeit, dass ein positiv getesteter Patient wirklich Krebs hat?

⁴Bei $n = 100$ beträgt diese Wahrscheinlichkeit immerhin 23%, wenn das Umfrageergebnis „7 oder mehr“ lautet.

Roulette

Ein Roulette-Gerät soll darauf getestet werden, ob alle Zahlen mit der gleichen Wahrscheinlichkeit von $1/37$ auftreten, indem man eine gewisse Anzahl von Spielrunden durchführt und die Häufigkeit der einzelnen Zahlen zählt (s. HÜBNER, S.187).

Ausbreitung der Euromünzen

In 12 Ländern gibt es unterschiedliche Euromünzen. Am 1.1.2002 wurden in den jeweiligen Ländern nur die landeseigenen Münzen ausgegeben. Wie mischen sich diese Münzen? Wie groß ist die Wahrscheinlichkeit, dass ich Anfang 2006 als Wechselgeld eine „ausländische“ Münze erhalte⁵?

Leukämieerkrankungen

In verschiedenen Gebieten Deutschlands gibt es unterschiedliche **relative Häufigkeiten** von Leukämiekranken. Wie hoch muss diese sein, damit man auf besondere Risiken (Kernkraftwerk, Baumschulen, Chemiefabriken, Elektrosmog, ...) in einem Gebiet schließen kann?

Unser mathematisches Modell kann so aussehen: Wir nehmen eine gewisse Anzahl (K) von „Kranken“, die wir zufällig auf eine gewisse Anzahl (n) von Gebieten verteilen. Dabei sei die Wahrscheinlichkeit, dass ein Kranker einem bestimmten Gebiet „zugelost“ wird, stets gleich, also $1/n$.

Wie groß ist die Wahrscheinlichkeit, dass einem bestimmten Gebiet genau k Kranke ($0 \leq k \leq K$) „zugelost“ werden?

Wieviele Kranke sind in einem bestimmten Gebiet „zu erwarten“? (klar: K/n)

Wie groß ist die Wahrscheinlichkeit, dass es in einem vorgegebenen Gebiet mehr als doppelt so viele Kranke gibt, als zu erwarten ist?

Wie groß ist die Wahrscheinlichkeit, dass es ein Gebiet mit mehr als doppelt so vielen Kranken wie zu erwarten gibt?

Diese Fragen können auch kombinatorisch angegangen werden. Jeder Kranke „zieht“ ein Gebiet.

Ziegenproblem

Sie sehen drei Zimmer mit verschlossenen Türen. Hinter einer Tür befindet sich der Hauptpreis, ein Auto. Hinter den anderen beiden Türen ein Trostpreis, eine Ziege. Sie haben 2 Versuche. Nach dem ersten wird Ihnen eine Tür geöffnet, hinter der eine Ziege steht. Jetzt haben Sie 2 Möglichkeiten: bei Ihrer Wahl der Tür zu bleiben oder die andere, noch geschlossene Tür zu wählen. Was ist aussichtsreicher?

Im Internet findet man dieses Spiel unter dem Namen **Schachtelspiel**⁶ (Biometrie Münster)

⁵Ich bin immer wieder überrascht, wie wenig sich vor allem die Kupfermünzen mischen. Liegt das darn, dass immer wieder neu geprägt wird?

⁶<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biometrie/bio/eda/monty.html>

3.2 Kombinatorik - Wiederholung

Ein grundlegender Begriff ist der des k -Tupels einer n -elementigen Menge A . Die didaktisch geschickte Bezeichnung ist nicht so offensichtlich: Die n Elemente von A kann man zu $A := \{1, 2, \dots, n\}$ durchnummerieren⁷, und sich als wohl unterscheidbare Kugeln in einer Urne vorstellen. Ein k -Tupel \mathbf{a} ist aus dem kartesischen Produkt $A^k := A \times A \times \dots \times A$ (k -mal) und hat die Gestalt $\mathbf{a} = (a_1, \dots, a_k)$ mit $a_j \in A, j = 1, 2, \dots, k$.

Ein k -Tupel beschreibt ein Ziehungsergebnis mit Anordnung — vergleiche mit den geordneten Paaren (a, b) , für die ja $(a, b) \neq (b, a)$ für $a \neq b$ gilt. Wieviele k -Tupel einer n -elementigen Menge A gibt es? Für die erste Komponente gibt es n Möglichkeiten, für die zweite Komponente ebenfalls,....., also insgesamt n^k , kurz

$$|A^k| = |A|^k.$$

Der Merkmalraum für das Ziehen mit Anordnung und mit Zurücklegen ist daher

$$\Omega = \Omega_1 := A^k = \{(a_1, a_2, \dots, a_k) : a_j \in A, j = 1, 2, \dots, k\}$$

Es gilt

$$|\Omega_1| = n^k.$$

„Mit Anordnung und ohne Zurücklegen“ führt auf k -Tupel mit paarweise verschiedenen Komponenten. Für die erste Komponente gibt es wieder n Möglichkeiten, für die zweite Komponente aber nur noch $n - 1, \dots$, für die k -te Komponente gibt es $n - k + 1$ Möglichkeiten, also insgesamt

$$(n)_k := n(n - 1) \cdots (n - k + 1).$$

Der Merkmalraum aller solcher Ziehungsergebnisse ist

$$\Omega = \Omega_2 := \{(a_1, a_2, \dots, a_k) : a_j \in A, a_i \neq a_j, i, j = 1, 2, \dots, k, \text{ für } i \neq j\}$$

Es gilt

$$|\Omega_2| = (n)_k.$$

Für $k = n$ führt dies auf $(n)_n = n!$ (n -Fakultät). Diese Zahl kann am besten durch die Anzahl aller Permutationen von n Elementen interpretiert werden. Die Rekursionsformel

$$n! = n \cdot (n - 1)!$$

hat dann eine einfache kombinatorische Interpretation: Man erhält alle $n!$ Permutationen, wenn man ein Element von A auszeichnet. Für dieses Element gibt es n mögliche Plätze. Für jeden dieser Plätze sind die restlichen $n - 1$ Elemente noch auf $(n - 1)!$ Weisen anordnungsbar.

⁷Im Gegensatz zu Kap. I.7 benutze ich nicht a_j als Namen für die j -te Kugel in der Urne, sondern einfach j . Dadurch stehen die Bezeichner a_k wieder zur Verfügung. Dies habe ich FISCHER entnommen.

Die für die Wahrscheinlichkeitsrechnung wichtigste Größe ist der *Binomialkoeffizient* $\binom{n}{k}$ (n über k , auch kurz „ k aus n “). Man kann auf $\binom{n}{k}$ Weisen k Elemente aus einer Menge mit n Elementen auswählen, d.h. es gibt $\binom{n}{k}$ Teilmengen mit k Elementen einer n -elementigen Menge. Dies entspricht genau der Anzahl der Ziehungsergebnisse ohne Zurücklegen und ohne Anordnung! Man kann mit

$$\Omega_3 := \{(a_1, \dots, a_k) \in \mathbb{N}^k : 1 \leq a_1 < a_2 < \dots < a_k \leq n\}$$

die Menge aller k -elementigen Teilmengen von A durch spezielle k -Tupel darstellen (dies ist neu gegenüber Kap. I.7, s. FISCHER, S.77). Kann man mit dieser Notation leichter als bisher die grundlegende kombinatorische Formel

$$|\Omega_3| = \binom{n}{k}$$

einsehen?

Nun, die Ziehungsergebnisse ohne Zurücklegen und *mit* Anordnung kann man ja durch

$$\Omega_2 := \{(a_1, a_2, \dots, a_k) \in A^k : a_i \neq a_j \text{ für } i \neq j\}, i, j = 1, 2, \dots, k$$

beschreiben. Offensichtlich führen $k!$ Permutationen eines k -Tupels von Ω_2 zum selben k -Tupel von Ω_3 . Daher gilt

$$k!|\Omega_3| = |\Omega_2|$$

und wegen $|\Omega_2| = (n)_k$ folgt die Behauptung.

Es gilt

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

Berechnen sollte man diese Zahl aber nach Kürzen, z.B.

$$\binom{n}{k} = \frac{(n)_k}{k!}$$

Eine für die *Binomialverteilung* wichtige Interpretation ist die Folgende: Auf n Plätze verteile man k Einsen und $n - k$ Nullen. Das geht gerade auf $\binom{n}{k}$ verschiedene Weisen! Oder: Wieviele n -Tupel von $\{0, 1\}$ haben genau k Einsen und $n - k$ Nullen als Komponenten? Eine k -elementige Teilmenge von $A = \{1, 2, \dots, n\}$ kann ich dadurch codieren, dass ich an den j -ten „Platz j “ eine Eins eintrage, wenn j zu der Teilmenge gehört, und eine Null sonst.

Wegen $|\text{Pot}A| = 2^n$ folgt jetzt sofort

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

Die *binomische Formel*

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

kann man sich sehr rasch kombinatorisch klarmachen!

Das Pascalsche Dreieck beruht auf

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},$$

was leicht kombinatorisch interpretiert werden kann, wenn man ein Element von A auszeichnet und unterscheidet, ob dieses zu einer k -elementigen Teilmenge von A gehört oder nicht.

Auch wenn der vierte kombinatorische Fall (mit Zurücklegen, ohne Anordnung) für die Wahrscheinlichkeitsrechnung unwichtig ist, will ich ihn kurz skizzieren. Die Menge aller Ziehungsergebnisse kann man mit

$$\Omega_4 := \{(a_1, a_2, \dots, a_k) \in A^k : 1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n\}$$

gleichsetzen. Das auch jetzt nicht so einfach einsehbare Ergebnis lautet

$$|\Omega_4| = \binom{n+k-1}{k}.$$

3.3 Merkmalraum

Zentraler Begriff der Wahrscheinlichkeits-Rechnung wie schon der Statistik ist der des **Merkmalraums**, den wir wieder Ω nennen. Dieser enthält die **Ergebnisse** von **Zufallsexperimenten**. Hierunter versteht man im weitesten Sinn die Ergebnisse von zufälligen und wiederholbaren, aber auch von geplanten Beobachtungen (wie durch Erhebungen in der Statistik). Die Elemente von Ω heißen hier **Elementarereignisse**⁸. Alle weiteren Begriffe wie *diskreter*, *kontinuierlicher Merkmalraum* findet man schon im vorangehenden Kap. 2.2 im Rahmen der *Beschreibenden Statistik*.

Der Zufall ist besonders offensichtlich bei Glücksspielen, er kommt aber auch schon immer dann ins Spiel, wenn man das Ergebnis einer Beobachtung nicht kennt. In diesem Sinne ist z.B. die Beobachtung der Blutgruppe eines Neugeborenen auch ein Zufallsexperiment.

Bei den im folgenden aufgeführten Zufallsexperimenten sind die Wahrscheinlichkeiten i.A. von vornerein bekannt - diese ordnen sich der Wahrscheinlichkeitstheorie zu. Zufallsexperimente der Statistik führen i.a. auf unbekannte Wahrscheinlichkeitsverteilungen, deren Kenngrößen „geschätzt“ werden sollen.

Beispiele für Merkmalräume in der Wahrscheinlichkeitstheorie

⁸Teilmengen von Ω werden *Ereignisse* genannt, siehe Def. 3.1.



Abbildung 3.1: Wappen und Zahl beim Münzwurf

- Münzwurf: Ω enthält die beiden Ergebnisse „Wappen“ und „Zahl“, siehe Abb. 3.1. Dies ist das einfachste Zufallsexperiment. Codiert man die Ergebnisse mit 0 bzw. mit 1 und setzt $\Omega = \{0, 1\}$, so spricht man von einem **Bernoulli-Experiment**. Während man beim Münzwurf meist davon ausgeht, dass beide „Elementarereignisse“ mit gleicher Wahrscheinlichkeit auftritt, ist dies beim Bernoulli-Experiment nicht zwingend der Fall. Man denke an den Wurf einer Reißzwecke auf eine poröse Unterlage, die es auch erlaubt, dass die Reißzwecke sich mit der Spitze in die Unterlage bohrt.
- Würfeln mit einem Spielwürfel: $\Omega = \{1, 2, \dots, 6\}$,
- Skat: Im „Skat“ des Kartenspiels „Skat“ liegen zwei von 32 Spielkarten. Ω enthält alle 2-elementigen Teilmengen von K , wobei K die Menge aller 32 Spielkarten ist. Wenn man sich nur für die Anzahl der Buben im Skat interessiert, kommt man mit $\Omega = \{0, 1, 2\}$ aus. Letzteren „reduzierten“ Merkmalraum würde man auch erhalten, wenn man auf dem ursprünglichen Ω eine *reelle Zufallsvariable* (siehe Kap. 3.3.1 und Kap. 3.8) mit Werten aus $\{0, 1, 2\}$ einführen würde.
- 2-maliges Würfeln mit einem Spielwürfel:

1. sehr differenziert

$$\Omega = \{(1, 1), \dots, (1, 6), \dots, (6, 1), \dots, (6, 6)\} = \{1, 2, \dots, 6\}^2. \quad (3.2)$$

Ω enthält $6^2 = 36$ Elemente⁹. Alle Elementarereignisse sind gleichwahrscheinlich.

2. wenn es nur auf die Augensumme ankommt:

$$\Omega = \{2, 3, \dots, 12\}.$$

Auch diesen „reduzierten“ Merkmalraum erhält man durch Betrachtung einer *reellen Zufallsvariablen* (siehe Kap. 3.3.1 und Kap. 3.8) auf dem Merkmalraum (3.2). Jetzt sind die Elementarereignisse nicht mehr gleichwahrscheinlich.

- k -maliges Bernoulli-Experiment

⁹Erinnern Sie sich an die „Produktregel“ $|A \times B| = |A| \times |B|$ aus Kap. I.3.3.5?

1. sehr differenziert: $\Omega = \{0, 1\}^k$. Es gilt $|\Omega| = 2^k$.
 2. Anzahl der „Treffer“ (Einsen): $\Omega = \{0, 1, \dots, k\}$.
- Mehrfaches (k -faches) Ziehen einer Kugel aus einer Urne. Identifiziert man die Urne mit einer Menge A , so kann $\Omega := A^k = A \times A \times \dots \times A$ (k -mal) gesetzt werden, wenn k -mal gezogen wird. Wenn nicht zurückgelegt wird und es nicht auf die Reihenfolge ankommt, besteht Ω aus allen k -elementigen Teilmengen¹⁰ von A . Kommt es nur auf die Anzahl („Treffer“) an, mit der eine ganz bestimmte Kugel gezogen wird, so ist $\Omega := \{0, 1, \dots, k\}$. In diesem Fall reduziert sich eine Ziehung auf ein Bernoulliexperiment, das k -fache Ziehen auf ein k -maliges Bernoulli-Experiment: die ausgezeichnete Kugel wird gezogen oder nicht.

3.3.1 Zufallsvariablen: Ein erster Zugang

Zuweilen errechnet man aus den Ergebnissen eines Zufallsexperiments eine (weitere) Zahl. Wenn man z.B. mit 2 Würfeln würfelt, kann man die Augensumme bilden. Wenn man man die beiden Cholesterinwerte LDL und HDL misst, kann man neben den Werten selbst auch den Quotienten bilden.

Eine Funktion auf einem Merkmalraum mit Werten in \mathbb{R} (also $X : \Omega \rightarrow \mathbb{R}$) nennt man eine **Zufallsvariable**. Wenn man nur an den Werten der Zufallsvariablen interessiert ist, kann man auch als Merkmalraum den Wertebereich der Zufallsvariablen nehmen, dies ist eine Art „Reduzierung“ des Merkmalraums.

Bei jeder Stichprobe, die reelle Daten $x_k, k = 1, 2, \dots, n$, erhebt, kann man die x_k als Werte einer Zufallsvariablen auffassen. Dabei müssen die Erhebungen verschiedener Werte *unabhängig* von einander sein¹¹.

Mit Zufallsvariablen kann man rechnen, z.B. zwei Zufallsvariable (wie LDL und HDL im obigen Beispiel) dividieren, und erhält eine neue Zufallsvariable. Ausführlicheres enthält Kap. 3.8.

Zufallsvariable werden mit einem großen Buchstaben X, Y, \dots bezeichnet. Während man sich zunächst für die Wahrscheinlichkeit $P(A)$ für das Eintreten eines Ereignisses $A \subset \Omega$ interessiert, wird es in Verbindung mit einer Zufallsvariablen z.B. auf $P(a \leq X \leq b)$ als die Wahrscheinlichkeit ankommen, dass der Wert der Zufallsvariablen zwischen a und b liegt. Ähnliches kennen Sie von Stichproben mit einer Stichprobenfunktion X . An Stelle von Wahrscheinlichkeit war von relativer Häufigkeit die Rede.

3.3.2 Zufällige Ereignisse

Definition 3.1. *Ein Ereignis ist eine Teilmenge $A \subset \Omega$ des Merkmalraumes. Die einelementigen Teilmengen des Merkmalraums (oder die Elemente von Ω selbst) heißen **Elementarereignisse**.*

¹⁰Es gilt $|\Omega| = \binom{n}{k}$, siehe Satz I.7.2.

¹¹Daher werden formal x_k als Wert einer Zufallsvariablen X_k aufgefasst, wobei X_1, \dots, X_n *identisch verteilt* und *stochastisch unabhängig* sein sollen, s. HÜBNER, S.177

Allgemein ist ein **Ereignissystem** eine Teilmenge der Potenzmenge von Ω , die gewissen Eigenschaften genügt und für deren Wahrscheinlichkeit man sich interessiert: In der Wahrscheinlichkeitstheorie wird ein **W-Maß** auf dem Ereignissystem eingeführt, d.h. jedem Ereignis $A \subset \Omega$ wird eine Wahrscheinlichkeit $P(A)$ zugeordnet. In der Statistik ist $P(A)$ der *Anteil* von Objekten mit der „Eigenschaft“ A an einer *Grundgesamtheit*, die man durch eine Totalerhebung ermitteln könnten.

Für diskrete Merkmalräume wird *jede* Teilmenge A von Ω als ein Ereignis aufgefasst, d.h. wir betrachten die gesamte Potenzmenge $Pot(\Omega)$ als Ereignissystem. Bei kontinuierlichen Merkmalen muss man echte Teilmengen von $Pot(\Omega)$ als Ereignissystem betrachten¹². Dies führt auf das Konzept einer σ -Algebra, deren Einführung hier aber vermieden werden soll. Das hat zur Folge, dass bei kontinuierlichen Merkmalräumen wenigen „exotischen“¹³ Teilmengen $A \subset \mathbb{R}$ kein W-Maß $P(A)$ zugeordnet werden kann.

Mengentheoretische Konzepte

Die Mengenoperationen \cap (Durchschnitt), \cup (Vereinigung), \setminus (Mengendifferenz) und c (Komplementbildung), siehe Kap. I.3.3.2, erfahren jetzt eine Interpretation und Anwendung: Das Ereignis $A_1 \cup A_2$ ist das Ereignis, dass A_1 **oder** A_2 eintritt, während $A_1 \cap A_2$ das Ereignis ist, dass A_1 **und** A_2 (gemeinsam) eintreten (man schreibt auch $A_1 A_2$). Man nennt zwei Ereignisse A_1 und A_2 **disjunkt**, wenn $A_1 A_2 = \emptyset$. In diesem Fall schreibt man auch $A_1 + A_2$ an Stelle von $A_1 \cup A_2$. Das Ereignis \bar{A} bedeutet, dass A nicht eintritt.

Bei den Wahrscheinlichkeits-Berechnungen von Ereignissen kommen **Rechenregeln für Mengenoperationen** zum Zuge wie die der Assoziativität, Kommutativität und Distributivität von \cup, \cap und die *Regeln von de Morgan* in Satz I.3.5

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Internet-Links

Applet Venn Diagramme und Wahrscheinlichkeiten¹⁴ (Biometrie Münster)
Einführung in Mengen¹⁵ (Mathe Online)

3.4 Häufigkeit und Wahrscheinlichkeit

3.4.1 Häufigkeiten

Gegeben sei ein Zufallsexperiment mit Merkmalraum Ω . Sei $\omega \in \Omega$ ein Elementar-Ereignis. Führen wir (eine endliche Zahl) n Zufallsexperimente (z.B. eine **Stichprobe vom Umfang**

¹²Dies führt auf den Begriff der *messbaren* Mengen

¹³die nicht messbaren Mengen

¹⁴<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/Venn2.html>

¹⁵<http://www.mathe-online.at/mathint/mengen/i.html>

n) durch, so können wir die **absolute Häufigkeit** von ω , d.h. die Anzahl n_ω der Experimente bzw. der Stichproben, deren Ergebnisse gleich ω sind, bestimmen, siehe Kap. 2.2.2. Die **relative Häufigkeit** ist dann

$$h_n(\omega) := \frac{n_\omega}{n},$$

vielleicht *der* zentrale Begriff der *Beschreibenden Statistik*.

Ganz entsprechend kann man von der absoluten (n_A) und relativen Häufigkeit $h_n(A)$ eines Ereignisses A sprechen.

3.4.2 Wahrscheinlichkeit bei Zufallsexperimenten

Jetzt geht es um den Begriff **Wahrscheinlichkeit** bei Zufallsexperimenten, genauer um die Wahrscheinlichkeit $P(A)$ für irgend ein Ereignis, d.h. für irgendeine Teilmenge $A \subset \Omega$. $P(A)$ ist dadurch bestimmt, dass wir

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

setzen (**empirisches Gesetz der großen Zahlen**), indem wir wenigstens gedanklich eine hinreichend große Zahl n von Zufallsexperimenten durchführen. In Kap. III.1.1.7 hatte ich dieses Gedanken im Hinblick auf die *Konvergenz von Folgen* erwähnt. Der Buchstabe P in $P(A)$ kommt vom englischen Wort **Probability** für Wahrscheinlichkeit.

Beispiel: Es wird $n = 1000$ -mal gewürfelt und die Anzahl der Sechsen gezählt. Man erwartet eine relative Häufigkeit des Elementarereignisses $\omega = 6$ in der Nähe von $1/6$. Würfelt man mit 2 Würfeln, so kann man A definieren als das Ereignis, dass die Augensumme 10 ist. Dann ist n_A die Anzahl von Würfeln, die diese Augensumme 10 ergeben hat. Für die relative Häufigkeit erwarten wir ein Ergebnis in der Nähe von $1/12$ (warum?).

JUMBO bietet hierzu einige Applets:

- **Applet - Relative Häufigkeiten und Wahrscheinlichkeiten** ¹⁶

Es werden unabhängige n „Münzwürfe“ (s. Bernoulli-Experimente) durchgeführt, bei denen ein Ereignis A mit Wahrscheinlichkeit p auftritt und die relativen Häufigkeiten für A ermittelt und mit p verglichen. p und n sowie die Ablaufgeschwindigkeit der Simulation sind einstellbar.

- **Applet - Würfelsimulation - Gesetz der großen Zahl und zentraler Grenzwertsatz** ¹⁷ Hier kann man mit $k \leq 9$ Würfeln n -mal würfeln ($n = 1, 10, 20, 100$), wobei nach jedem Wurf die Augensumme der k Würfel gezählt wird, und die empirische Verteilung (s. Kap. 2.4.9) auf die verschiedenen Augensummen ermitteln. Zum Vergleich kann die *Normalverteilung* (s. Kap. 3.6 und Kap. 3.11) eingeblendet werden, die die empirischen Verteilung für nicht zu kleine k gut approximiert¹⁸.

¹⁶<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/prob5.html>

¹⁷<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/grza2.html>

¹⁸Dies liegt am *Zentralen Grenzwertsatz*

Häufig kann man die Wahrscheinlichkeiten aber auch ohne praktische Durchführung sehr vieler Zufallsexperimente ermitteln, etwa durch Symmetrieüberlegungen wie beim Würfel (alle Elementar-Ergebnisse bei einem Würfel haben die gleiche Wahrscheinlichkeit). Dies liegt der Konstruktion der wichtigsten W-Modelle in Kap. 3.5.2 zu Grunde.

Beispiel: Beim 2-maligen Würfeln sei $A_k \subset \Omega$ durch $(a, b) \in A_k$, wenn $a + b = k$, definiert. Es ist $P(A_{10}) = 1/12$ und $P(A_9) = 1/9$.

3.4.3 Wahrscheinlichkeiten in der Statistik

In der Statistik strebt man durch einen hinreichend großen Stichprobenumfang n an, den **Anteil** $P(A)$ an der **Grundgesamtheit** zu bestimmen, dem die durch A gegebene Eigenschaft zukommt. Diesen Anteil kann man als Wahrscheinlichkeit auffassen, dass sich bei *einer* zufälligen Stichprobe ein Ergebnis aus A ergibt.

Beispiel: Die Grundgesamtheit bestehe aus allen 15-jährigen Schülern in Deutschland. In einer Stichprobe werden deren Mathematik-Kenntnisse an Hand von Aufgaben getestet und mit einer Punktzahl zwischen 0 und 1000 bewertet ($\Omega := \{0, 1, \dots, 1000\}$). Ist A das Intervall zwischen 400 und 499, so ist $P(A)$ der Anteil der Schüler, die eine Punktzahl zwischen 400 und 499 erreichen. Eine Stichprobe in einem gewissen Umfang n würde $P(A)$ durch die relative Häufigkeit $h_n(A)$ annähern („schätzen“).

3.4.4 Wahrscheinlichkeit: Axiome von Kolmogoroff

Für jedes n gilt offensichtlich $h_n(\Omega) = 1$, da jedes Ergebnis eines Zufallsperiments ein Elementarereignis ist, und $h_n(\emptyset) = 0$, da jedes Zufallsexperiment ein Ergebnis hat¹⁹. Ansonsten gilt $0 \leq h_n(A) \leq 1$, da $h_n(A)$ als relative Häufigkeit ein Anteil²⁰ ist. Ferner gilt offensichtlich $h_n(A + B) = h_n(A) + h_n(B)$, wenn A und B disjunkte Ereignisse sind.

Im Zusammenhang mit dem *empirischen Gesetz der großen Zahl* ergeben sich folgende „einleuchtende“ Regeln (**Kolmogoroff-Axiome**)(1933) für ein **W-Maß** P (alle folgenden Mengen sind beliebige Teilmengen von Ω):

Definition 3.2. Eine Funktion $P : Pot\Omega \rightarrow \mathbb{R}_+$, die jedem „Ereignis“ $A \subset \Omega$ eine „Wahrscheinlichkeit“ $P(A)$ zuordnet, heißt **Wahrscheinlichkeits-Maß** (kurz **W-Maß**) genau dann, wenn gilt

- $P(\Omega) = 1, P(\emptyset) = 0$
- $1 \geq P(A) \geq 0$
- $P(A + B) = P(A) + P(B)$, falls A und B disjunkt sind.

Hieraus folgen mit Hilfe der Regeln für Mengenoperationen (s.o.) weitere Regeln:

¹⁹Man nennt die leere Menge ein *unmögliches Ereignis*

²⁰Manchmal werden Anteile auch in Prozent (%) gemessen. Dann muss mit 100 multipliziert werden

Satz 3.3. 1. $P(\bar{A}) = 1 - P(A)$

2. $P(A \setminus B) = P(A) - P(AB)$

3. $A \subset B \implies P(A) \leq P(B)$

4. $P(A \cup B) = P(A) + P(B) - P(AB)$

Beweis:

1. Wegen der Partition²¹ $\Omega = A \cup \bar{A}$, $P(\Omega) = 1$ und des letzten Axioms.

2. Folgt ebenfalls aus dem letzten Axiom, da $A = (A \cap B) + A \setminus B$ eine Partition von A ist.

3. Falls $A \subset B$, ist $B = A \cup (B \setminus A)$ eine Partition. Also gilt

$$P(B) = P(A) + P(B \setminus A),$$

woraus wegen $P(B \setminus A) \geq 0$ die Behauptung folgt.

4. Aufgabe: Leite diese Regel aus der letzten her, indem von

$$A \cup B = A + (B \setminus A), \quad B = AB + B \setminus A$$

ausgegangen wird.

■

Ersetzt man bei endlichen Mengen $P(A)$ durch $|A|$ (die *Anzahl* der Elemente von A), so sind alle Axiome – bis auf $P(\Omega) = 1$ und $P(\bar{A}) = 1 - P(A)$ – Aussagen der elementaren Mengenlehre. Es lohnt sich, insbesondere das Axiom $P(A \cup B) = P(A) + P(B) - P(AB)$ unter diesem Gesichtspunkt zu betrachten. Es ist die einfachste Form des Prinzips der Inklusion-Exklusion, das in den Übungen zu Mathe I vorkam, sieh auch Kap. 3.4.5.

Eine physikalische Veranschaulichung liefert eine Massenbelegung eines Bereiches $\Omega \subset \mathbb{R}^p$, $p = 1, 2, 3$ mit Gesamtmasse „Eins“. Dann setze man $P(A)$ als Masse von $A \subset \Omega$ oder – falls die Gesamtmasse nicht Eins ist – als der Massenanteil von A an der Gesamtmasse.

Die offensichtlichste Veranschaulichung ist die, dass man Ω als einen Kreis oder ein Quadrat in der Ebene \mathbb{R}^2 annimmt und als $P(A)$ den Anteil von $A \subset \Omega$ an der Fläche von Ω interpretiert. Das hat zwar theoretische Macken (nicht jeder Menge kann man eine Fläche zuordnen), aber man kann die Wahrscheinlichkeit-Axiome und ihre Folgerungen leicht an Hand von Mengen-Diagrammen deuten:

Die Abb. 3.2 - 3.5 aus FISCHER sprechen für sich.

Die folgenden drei Applets stammen aus dem JUMBO-Skript (Kap. 4.3) und visualisieren den Zusammenhang zwischen Wahrscheinlichkeiten und Mengenoperationen. Der Merkmalraum besteht aus 20 „Kugeln“ unterschiedlicher Farbe und Nummern. Alle Elementarereignisse sollen gleiche Wahrscheinlichkeit haben.

²¹ $X = A \cup B$ heißt *Partition* von X , falls A und B disjunkt sind.

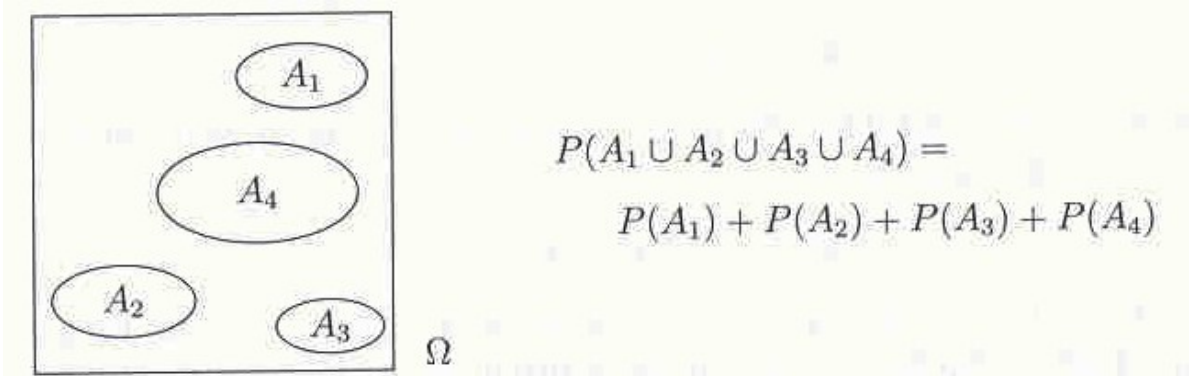


Abbildung 3.2:

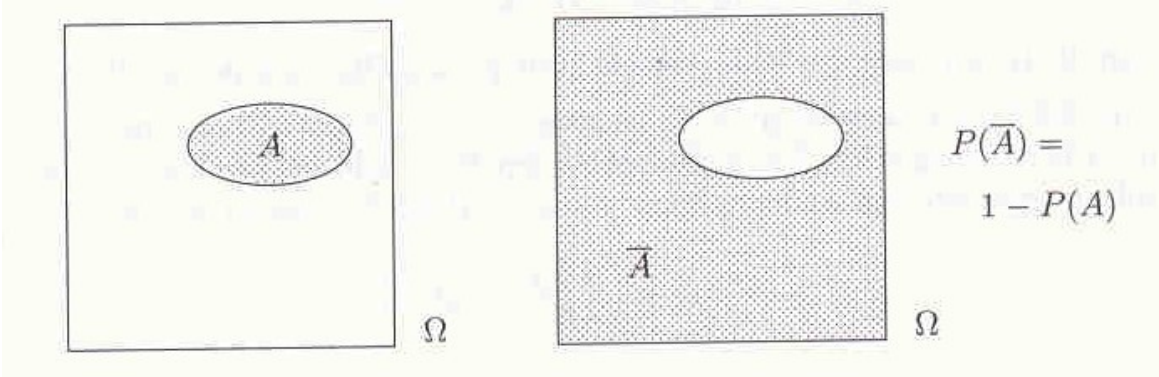


Abbildung 3.3:

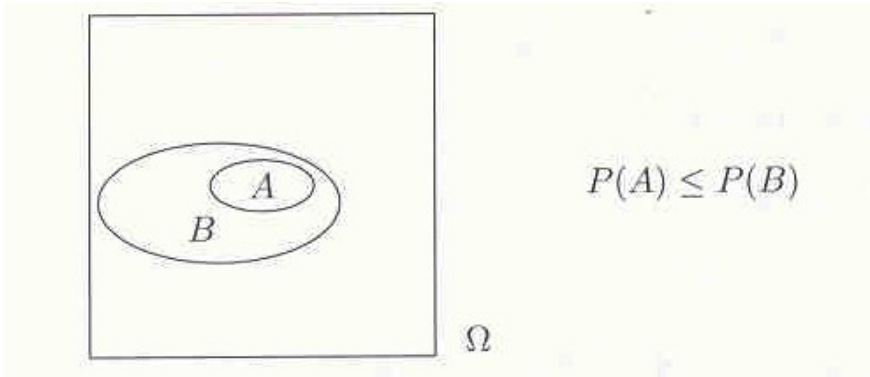


Abbildung 3.4:

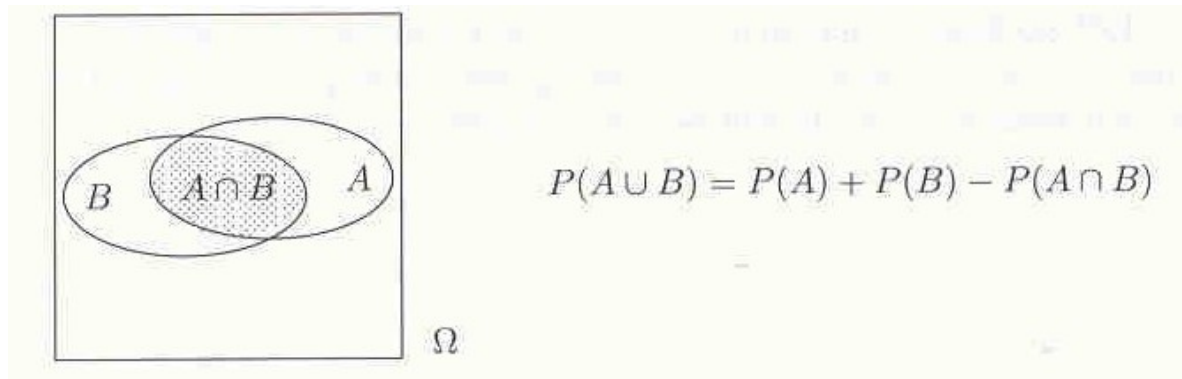


Abbildung 3.5:

- **Applet - Wahrscheinlichkeit von Ereignissen** ²²
- **Durchschnitt und Wahrscheinlichkeiten** ²³
- **Vereinigung und Wahrscheinlichkeiten** ²⁴

3.4.5 Einschluss (Inklusion) - Ausschluss (Exklusion)-Formel

Sie erinnern sich (?):

In Mathe I gab es eine Übungsaufgabe mit folgendem Text:

Die **Summenregel** $|A \cup B| = |A| + |B|$ ist nur für disjunkte A und B richtig. Allgemein gilt

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Noch allgemeiner gilt für 3 Mengen (s.Abb.3.4.5)

$$|A \cup B \cup C| = |A| + |B| + |C| - (|A \cap B| + |B \cap C| + |A \cap C| - |A \cap B \cap C|)$$

Wenn man jetzt die „Inhalte“ $|A|, \dots, |A \cup B \cup C|$ durch Wahrscheinlichkeiten $P(A), \dots, P(A \cup B \cup C)$ ersetzt, bleiben die Formeln, die *Einschluss-Ausschluss-Formeln* oder *Inklusion-Exklusion-Formeln* heißen, richtig:

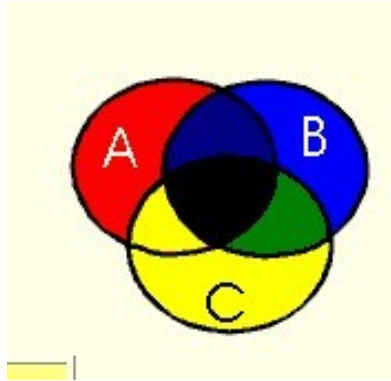
Satz 3.4. *Es gilt*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - (P(A \cap B) + P(B \cap C) + P(A \cap C) - P(A \cap B \cap C))$$

²²<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/prob1.html>

²³<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/prob3.html>

²⁴<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/prob4.html>

Abbildung 3.6: $|A \cup B \cup C| = ?$

Beweis: Diese Formel wird wegen $A \cup B \cup C = (A \cup B) \cup C$ auf die dritte Formel in Satz 3.3 zurückgeführt. Es gilt danach ja

$$P(A \cup B \cup C) = P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C).$$

Nun nutze die Regel $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ aus und wende wieder zwei Mal die dritte Formel in Satz 3.3 an. ■

Man kann diese Formel auf beliebig viele „Summanden“ $A_j, j = 1, 2, \dots, n$ in $P(A_1 \cup A_2 \cup \dots \cup A_n)$ verallgemeinern.

Von dieser allgemeinen Formel soll hier nur ein Spezialfall interessieren, in dem $c(m) := P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_m})$ nur von der Zahl m der „Faktoren“ abhängt:

Satz 3.5.

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{m=1}^n (-1)^{m+1} \binom{n}{m} c(m).$$

Anwendungsbeispiele:

- $k \geq n$ Kugeln sollen auf n Fächer verteilt werden, wobei jedes Fach mit gleicher Wahrscheinlichkeit von einer Kugel (auch mehrfach) besetzt wird. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Fach leer bleibt?

Sei A_j das Ereignis, dass das j -te Fach leer bleibt. Offensichtlich ist $P(A_1 \cup A_2 \cup \dots \cup A_n)$ gesucht. Wir müssen also nur noch $c(m) := P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_m})$ berechnen, das ist die Wahrscheinlichkeit, dass ein ganz bestimmter Satz von m Fächern mit den Nummern j_1, j_2, \dots, j_m leer bleibt. $A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_m}$ ist also genau das Ereignis, dass alle k Kugeln in den restlichen $n - m$ Fächern unterkommen. Hierfür gibt es $(n - m)^k$ mögliche Zuordnungen, so dass $c(m) = (n - m)^k / n^k$. Damit ist die gesuchte Wahrscheinlichkeit

$$\sum_{m=1}^n (-1)^{m+1} \binom{n}{m} \left(1 - \frac{m}{n}\right)^k.$$

- n Paare besuchen eine Party. Für ein Tanzspiel werden Tanzpaare per Los zusammengestellt. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Paar zusammentrifft?

Sei A_j das Ereignis, dass das j -te Paar zu einem Tanzpaar wird. Wieder ist $P(A_1 \cup A_2 \cup \dots \cup A_n)$ gesucht. Wir müssen auch hier $c(m) := P(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_m})$ berechnen, das ist die Wahrscheinlichkeit, dass ein ganz bestimmter Satz von m Paaren zusammentrifft. Da es $n!$ verschiedene Tanzpaarungen gibt (warum?), gibt es unter diesen nur noch $(n-m)!$ Tanzpaarungen mit den ausgezeichneten m Paaren als Tanzpaare. Also gilt $c(m) = \frac{(n-m)!}{n!}$, so dass

$$\sum_{m=1}^n (-1)^{m+1} \binom{n}{m} \frac{(n-m)!}{n!}$$

die gesuchte Wahrscheinlichkeit p_n ist, welche sich zu

$$p_n = \sum_{m=1}^n (-1)^{m+1} \frac{1}{m!}$$

vereinfachen lässt. Aus der Exponentialreihe folgt übrigens

$$\lim_{n \rightarrow \infty} p_n = 1 - e^{-1} \approx 0.63,$$

d.h. für große n kann man wetten, dass mindestens ein Paar wieder zusammenfindet.

3.5 Wahrscheinlichkeits-Modelle, Verteilungen

Definition 3.6. Ein Wahrscheinlichkeits-Modell (kurz **W-Modell**) ist durch durch drei Dinge gekennzeichnet: den Merkmalraum Ω , das Ereignissystem (hier meist $\text{Pot}(\Omega)$) und ein W-Maß P .

Man beachte, dass bei diskreten Modellen ein W-Modell dadurch gegeben ist, dass für jedes $\omega \in \Omega$ die „Elementar-Wahrscheinlichkeit“ $f(\omega) := P(\{\omega\})$ festgesetzt wird. Es muss nur

$$\sum_{\omega \in \Omega} f(\omega) = 1$$

und

$$f(\omega) \geq 0 \text{ für alle } \omega \in \Omega$$

gelten. Aus obigen Axiomen folgt dann

$$P(A) = \sum_{\omega \in A} f(\omega).$$

Definition 3.7. Ist der Merkmalraum $\Omega = \{\omega_1, \dots, \omega_m\}$ eines W -Modells diskret, so heißt $f : \Omega \rightarrow \mathbb{R}_+$, definiert durch $f(\omega) := P(\{\omega\})$ **Zähldichte** oder auch (**diskrete Wahrscheinlichkeits-**) **Verteilung**. Diese ist durch einen **Wahrscheinlichkeits-Vektor**²⁵ $\mathbf{p} = (p_1, p_2, \dots, p_m)$ mit $p_j := f(\omega_j), j = 1, 2, \dots, m$ gegeben.

Man beachte, dass jede Stichprobe vom Umfang n durch Ermittlung der relativen Häufigkeit $f(\omega) := h_n(\omega)$ eines Elementarereignisses $\omega \in \Omega$ auch schon ein – wenn auch nicht das der Grundgesamtheit – W -Maß definiert. Eine solche diskrete Verteilung (die *Häufigkeitsverteilung* genannt wurde und durch die relativen Häufigkeiten h_j an Stelle von p_j bestimmt wird, siehe Kap. 2.2.2) lässt sich leicht visualisieren, wie wir in Kap. 2.3 gesehen haben. Dabei muss man zwischen der „wahren“ Verteilung, die durch das W -Maß gegeben ist, und einer „empirischen“ Verteilung, die man durch n Zufallsexperimente oder Stichproben gewinnt, unterscheiden. Man beachte, dass die hier definierte Wahrscheinlichkeits-Verteilung einen *diskreten*, aber nicht zwingend einen quantitativen Merkmalraum voraussetzt.

3.5.1 Diskrete Verteilungsfunktion bei quantitativen Merkmalen

Ist $\Omega := \{\omega_j, j = 1, 2, \dots, m\}$ ein endlicher Merkmalraum eines W -Modells, so sind nichtnegative reelle Zahlen $p_j := f(\omega_j), j = 1, 2, \dots, m$ definiert, wobei f die Zähldichte ist. Handelt es sich um einen quantitativen Merkmalraum, so liefert ein Zufallsexperiment als Ergebnis stets einer der m Zahlen $\omega_j \in \mathbb{R}$, wir sprechen auch von einer *Zufallsvariablen*²⁶ X (s. Kap. 3.8).

Definition 3.8. Zu einem endlichen Merkmalraum in einem W -Modell ist die **Verteilungsfunktion** durch

$$F^X(x) := P(X \leq x) = \sum_{\omega_j \leq x} p_j, \quad x \in \mathbb{R} \quad (3.3)$$

definiert.

Beachte, dass man aus der diskreten Verteilung die Verteilungsfunktion und umgekehrt gewinnen kann. Die Verteilungsfunktion beginnt bei Null, ist monoton wachsend und endet bei Eins. Es handelt sich um eine *Treppenfunktion* mit Stufen bei ω_j . Man mache sich den Unterschied zwischen *Verteilung* und *Verteilungsfunktion* klar. Vielleicht ist es „didaktisch“ klüger, wie in HÜBNER von *Zähldichte* an Stelle von *Verteilung* zu sprechen.

In Abb. 3.7 sehen Sie die Verteilungsfunktion²⁷ zum „perfekten“ Spielwürfel.

Ersetzt man bei Stichproben die Wahrscheinlichkeit p_j durch die relativen Häufigkeiten h_j , so erhält man gerade die *empirische Verteilungsfunktion* F , siehe auch (2.1) in Kap. 2.4.9.

²⁵Man spricht von einem **W-Vektor**, wenn dessen Komponenten alle nichtnegativ sind und sich zu Eins aufsummieren.

²⁶Die nachfolgende Definition wird auf allgemeine Zufallsvariablen X verallgemeinert werden. Daher fügen wir jetzt schon einen oberen Index X an.

²⁷Diese Abbildung stammt aus JUMBO. Sie ist streng genommen nicht korrekt, da die vertikalen Teile nicht mitgezeichnet werden dürfen.

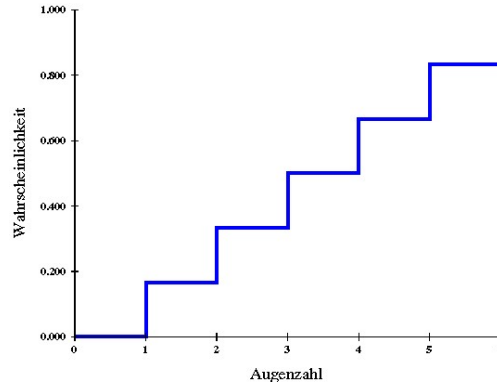


Abbildung 3.7: Verteilungsfunktion Würfeln

Wir werden sehen, dass der Begriff der *Verteilungsfunktion* auch und gerade dort für *kontinuierliche* Merkmalräume bzw. für *kontinuierliche* Zufallsvariable eine wesentliche Rolle spielt. Dabei tritt der Begriff *Wahrscheinlichkeits-Dichte* an die Stelle der *Zähldichte*²⁸.

Man beachte, dass man mit Hilfe der Verteilungsfunktion F^X

$$P(a < X \leq b) = F^X(b) - F^X(a) = \sum_{a < \omega_j \leq b} p_j$$

ausdrücken kann. Das hatten wir für empirische Verteilungsfunktionen schon in Satz 2.8 notiert, wobei die Merkmalsausprägungen zu

$$\omega_1 < \omega_2 < \dots < \omega_m \tag{3.4}$$

angeordnet wurden.

Bemerkung: Man beachte die unterschiedlichen Ungleichheitszeichen in $P(a < X \leq b)$. Wenn $a \neq \omega_j, j = 1, 2, \dots, m$, so gilt ebenfalls $P(a \leq X \leq b) = F^X(b) - F^X(a)$. Wenn jedoch $a = \omega_j$ und die Anordnung (3.4) gilt, so haben wir

$$P(a \leq X \leq b) = F^X(b) - F^X(\omega_{j-1}).$$

Besonders einfach ist es, die Wahrscheinlichkeit $P(X > a)$ mit Hilfe der Verteilungsfunktion F^X auszudrücken: Es gilt

$$P(X > a) = 1 - F^X(a). \tag{3.5}$$

Man beachte, dass „ $X > a$ “ mit „nicht $X \leq a$ “ logisch äquivalent ist.

Zum Abschluss zeigen wir in Abb. 3.8 die Verteilungsfunktion zum Geburtstagsproblem (aus FISCHER, S.84), wobei $p(k) = F^X(k)$ die Wahrscheinlichkeit ist, dass in einer Gruppe von k Menschen mindestens zwei am selben Tag Geburtstag haben²⁹.

²⁸Man beachte, dass es bei kontinuierlichen Merkmalen i.a. keinen Sinn gibt, für ein Elementarereignis $\omega \in \Omega$

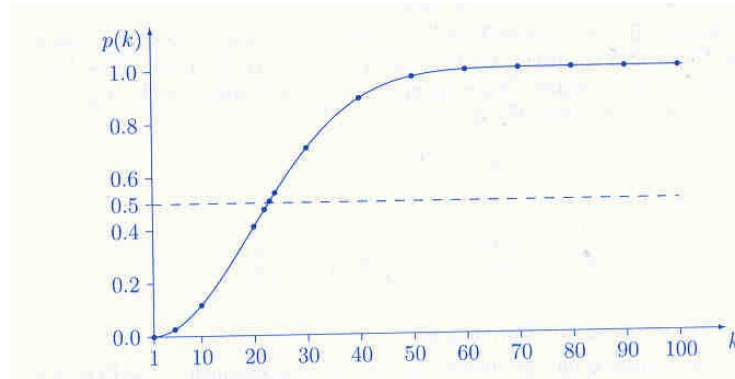


Abbildung 3.8: Wahrscheinlichkeiten beim Geburtstagsproblem

3.5.2 Bernoulli-, Binomial-, Laplace-Modelle und ihre Verteilungen

Für diese Beispiele von Zufallsexperimenten werden Elementar-Wahrscheinlichkeiten in Übereinstimmung mit kombinatorischen Überlegungen definiert. Dadurch werden hier die ersten konkreten Wahrscheinlichkeits-Modelle eingeführt. In diesen Beispielen handelt es sich um diskrete Merkmalräume, die jeweilige W-Verteilung (Zähldichte) wird auf Grund von kombinatorischen Symmetrieüberlegungen bestimmt.

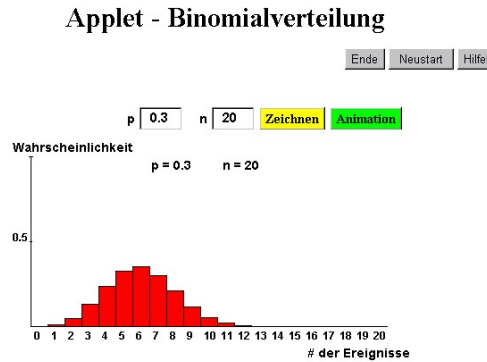
1. Ein **Bernoulli-Modell** $\mathbf{B}(p)$ ist durch $p := P(\{1\})$ definiert. $q := 1 - p = P(\{0\})$ ist die Komplementär-Wahrscheinlichkeit. Man zeichne die zugehörige Verteilungsfunktion!
2. Ein **Binomial-Modell** $\mathbf{B}(n, p)$ beruht auf der n -fachen Hintereinanderausführung eines Bernoulliexperimentes. Der Merkmalraum Ω besteht aus n -Tupeln von Nullen und Einsen, in reduzierter Form auch nur aus $\Omega = \{0, 1, \dots, n\}$, den möglichen Anzahlen von Einsen (den „Treffern“). Die Wahrscheinlichkeit p_j für ein ganz bestimmtes n -Tupel mit j Einsen (und $n - j$ Nullen) ergibt sich bei stochastischer Unabhängigkeit der einzelnen Ausgänge (siehe Kap. 3.7) zu $p^j(1 - p)^{n-j}$. Da es $\binom{n}{j}$ verschiedene n -Tupel mit j Einsen gibt (Kombinatorik!), ist die Wahrscheinlichkeit für genau j Einsen

$$b(n, p; j) := \binom{n}{j} p^j (1 - p)^{n-j}.$$

Wenn man also den Merkmalraum zu $\Omega := \{0, 1, \dots, n\}$ reduziert, weil es nur auf die Gesamtzahl von Einsen ankommt, ist $b(n, p; j)$ die Elementar-Wahrscheinlichkeit des Elementar-Ereignisses $\{j\}$. Hierdurch ist eine Verteilung mit den beiden *Parametern*

eine Wahrscheinlichkeit $f(\omega) = P(\{\omega\})$ zu definieren. Was für einen Sinn beispielsweise soll es haben, wenn man von der Wahrscheinlichkeit spricht, dass eine zufällig ausgewählte Person genau $\pi/2$ Meter lang ist

²⁹Dieses Beispiel ist etwas unglücklich, da kein Zufallsexperiment bzw. eine Zufallsvariable angebar ist, so dass $p(k) = P(X \leq k)$

Abbildung 3.9: Binomialverteilung für $p = 0.3$, $n = 20$

und p definiert, die **Binomialverteilung** heißt und auf JACOB BERNOULLI (1654-1705) zurückgeht. Man sollte auf jeden Fall einmal das Stabdiagramm zur Binomialverteilung für feste n und p studieren. Für große n hat es die Gestalt einer *Glockenkurve*, die in ihrer „reinsten“ Form bei der Normalverteilung (Kap. 3.6) auftritt³⁰. Macht man m Zufallsexperimente mit den Parametern n und j , so erhält man eine *empirische Verteilung*, die die Binomialverteilung approximieren sollte.

JUMBO-Applets:

- **Applet - Binomialverteilung**³¹. Hier kann man interaktiv p und $n \leq 99$ wählen, sogar die Verteilung animieren, indem man p variiert.
- **Javascript und Applet - diskrete Verteilungen**³² Hier kann man den Binomialkoeffizienten $\binom{n}{j}$, die Binomialverteilung und die zugehörige Verteilungsfunktion

$$F_{n,p}(x) := P(X \leq x) = \sum_{j \leq x} b(n, p; j)$$

(desgleichen für die Poisson-Verteilung in Kap. 3.5.3) berechnen.

3. Ein **Laplace-Modell** liegt vor, wenn $\Omega = \{1, 2, \dots, N\}$ endlich ist und alle Elementarereignisse $\{j\}$ gleiche Wahrscheinlichkeit $p_j = p = 1/N$, $j = 1, \dots, N$ haben. Man spricht auch von einer (diskreten) **Gleichverteilung**.

Für $A \subset \Omega$ folgt dann

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

³⁰Der *zentrale Grenzwertsatz* liefert die Begründung hierfür.

³¹<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomasche/bio/bern.html>

³²<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomasche/bio/diskret1.html>

(siehe auch Abschnitt 3.1.1.)

Hier sind wir das erste Mal auf konkrete diskrete **Verteilungen** gestoßen. Immer dann, wenn man verschiedene Ereignisse unter dem Blickwinkel ihrer Wahrscheinlichkeit oder (in der Statistik) ihrer *Häufigkeit* ordnet, spricht man von *Verteilungen*, in der Statistik auch von *empirischen Verteilungen*. Besonders häufig und wichtig ist der Fall, dass der Merkmalraum etwas mit Zahlen zu tun hat, wenn er also quantitativ ist. Es gibt noch viele weitere wichtige Verteilungen, unter denen die *Normalverteilung* die vielleicht bedeutendste ist. Sie ist jedoch eine *kontinuierliche* Verteilung (s. Kap. 3.6), die etwas schwieriger zu verstehen ist als diskrete Verteilungen.

Beispiele für Binomialverteilungen

- Die Wahrscheinlichkeit für eine bestimmte Eigenschaft („weiblich“, „Nichtwähler“, „Blutgruppe 0“, „RaucherIn“, ...) einer zufällig herausgegriffenen Person sei p . Wie groß ist die Wahrscheinlichkeit, dass von $n = 100$ zufällig herausgegriffenen Personen genau k (bzw. höchstens k) diese Eigenschaft haben? Die Antwort lautet $b(n, p; k)$ (bzw. $F_{n,p}(k)$).

Fragt man nach der Wahrscheinlichkeit, dass *mindestens* k Personen diese Eigenschaft haben, so ist die Antwort $1 - F_{n,p}(k - 1)$ ³³.

- *Leukämie* (Kap. 3.1.2): Sei $p := 1/n$ die Wahrscheinlichkeit, dass ein Kranker unter K Kranken einem bestimmten Gebiet „zugelost“ wird. Wie groß ist die Wahrscheinlichkeit, dass einem bestimmten Gebiet genau k Kranke ($0 \leq k \leq K$) „zugelost“ werden? Antwort: $b(K, p; k)$.

Wie groß ist die Wahrscheinlichkeit w , dass es in einem vorgegebenen Gebiet mehr als doppelt so viele Kranke gibt, als zu erwarten ist? Antwort: K/n Kranke sind zu erwarten, also (vergleiche mit (3.5))

$$w = \sum_{k > \frac{2K}{n}} b(K, p; k) \quad \left(= 1 - F_{n,p} \left(\frac{2K}{n} \right) \right).$$

Sei $K = 100$ und $n = 10$. Dann ist $p = 0.1$ und es ergibt sich $w = 0.002$, die gesuchte Wahrscheinlichkeit ist 0,2%. Das selbe Ergebnis erhält man, wenn nur $K = 10n$.

Wie groß ist die Wahrscheinlichkeit, dass es ein Gebiet mit mehr als doppelt so vielen Kranken wie zu erwarten gibt (ein solches Gebiet wollen wir „kritisch“ nennen)?

Hier hilft wieder die Binomialverteilung. Ein gegebenes Gebiet sei kritisch mit der Wahrscheinlichkeit w , z.B. mit $w = 0.002$. Wir stellen uns vor, dass mit einem Bernoulliexperiment $B(w)$ ermittelt wird, ob ein Gebiet kritisch ist oder nicht. Dieses Experiment wird

³³„mindestens k “ bedeutet „ k oder mehr“ oder auch „nicht ($k-1$) oder weniger“.

nach und nach für jedes der n Gebiete wiederholt. Dann ist die gesuchte Wahrscheinlichkeit gerade $1 - b(n, w; 0) = q - (1 - w)^n$. Für $n = 1000$ und $w = 0.002$ kommt hier schon die Wahrscheinlichkeit 86% heraus, dass mindestens ein Gebiet kritisch ist.

- *Bierflaschen (Kap. 3.1.2)*: Eine Bierflasche werde mit der Wahrscheinlichkeit von $p = 0.01$ (1%) beschädigt. Wie groß ist die Wahrscheinlichkeit, dass ein Kasten Bier (24 Flaschen) mindestens eine³⁴ beschädigte Flasche enthält? Antwort: $1 - b(24, p; 0) = 0.214$ (21,4%), da $b(24, p; 0)$ die Wahrscheinlichkeit ist, dass keine Flasche beschädigt ist.
- *Wahlen (Kap. 3.1.2)*: Jeder zwanzigste möge FDP wählen ($p = 0.05$). Wie groß ist die Wahrscheinlichkeit, dass in einer Stichprobe von 1000 Personen 7% oder mehr (also mindestens 70) FDP-Wähler sind? Antwort:

$$\sum_{k \geq 70} b(1000, 0.05; k) = 1 - F_{n,p}(69) = 0.00347436 (= 0,34\%).$$

Bei einer Stichprobe von nur 100 Personen beträgt die Wahrscheinlichkeit immerhin schon 23,4%

Mit statistischen Worten kann man sagen, dass die Null-Hypothese, die FDP würde mit 5% oder weniger abschneiden, bei einer Irrtums-Wahrscheinlichkeit von 0,34% abgelehnt werden müsste, wenn bei einer Stichprobe von 1000 Personen 70 angegeben, die FDP wählen zu wollen.

Die numerischen Ergebnisse habe ich mit Hilfe von **Javascript und Applet - diskrete Verteilungen** (JUMBO) erzielt.

3.5.3 Poissonverteilung

In einem Geigerzähler werden in einem bestimmten Zeitraum (etwa 1 Minute) die radiaktiven Zerfälle gezählt. Da es keine obere Schranke gibt, ist theoretisch $\Omega = \mathbb{N}_0$. Mit einer charakteristischen Zahl $\lambda > 0$, dem Parameter der Poisson-Verteilung, kann man dann als Elementar-Wahrscheinlichkeit

$$f(j) = \pi(\lambda; j) := e^{-\lambda} \frac{\lambda^j}{j!}, j \in \mathbb{N}_0 \quad (3.6)$$

der **Poissonverteilung** (S.D. POISSON, 1781-1840) annehmen. Dass sich alle Wahrscheinlichkeiten zu Eins addieren, liegt an $\sum_{j=0}^{\infty} f(j) = 1$, was aus der Exponentialreihe

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

folgt, siehe Kap. III.1.3.5.

³⁴nicht keine

Sehr interessant und wichtig (auch für die Herleitung von (3.6)) ist die Tatsache, dass die Binomialverteilung in gewissen Fällen sehr gut durch die (einfacher zu berechnende) Poissonverteilung ersetzt werden kann: Wenn für die Binomialverteilung p sehr klein und n sehr groß ist, so kann diese durch die Poissonverteilung ersetzt werden, wenn man $\lambda := np$ setzt. Übrigens ist np bzw. λ der Erwartungswert (s.Kap. 3.9.1) der Binomial- bzw. Poissonverteilung. Umgekehrt ist die Poisson-Verteilung Grenzwert von Binomialverteilungen $B(n, p_n)$ mit $n \rightarrow \infty$ und $np_n \rightarrow \lambda$.

Dies wollen wir genauer mit den Hilfsmitteln der Analysis untersuchen:

Satz 3.9. *Sei (p_n) eine Folge von Zahlen mit $0 < p_n < 1$, für die die Folge (np_n) gegen den Grenzwert λ konvergiert. Dann gilt für jedes feste k*

$$b(n, p_n; k) \rightarrow \pi(\lambda; k), \quad n \rightarrow \infty.$$

Beweis: (HÜBNER, S.64 unten)

Wir sind am Ziel, wenn mit

$$(n)_k := n(n-1) \cdots (n-k+1)$$

gezeigt werden kann, dass³⁵

$$\frac{(n)_k}{k!} p_n^k (1-p_n)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

Der Nenner $k!$ kommt auf beiden Seiten vor – kein Problem. Wie kommt nun $e^{-\lambda}$ ins Spiel? Wir erinnern uns von der kontinuierlichen Verzinsung aus Kap. III.1.2.4 an

$$\left(1 + \frac{x}{n}\right)^n \rightarrow e^x,$$

bzw. an

$$\left(1 - \frac{x_n}{n}\right)^n \rightarrow e^{-x}, \tag{3.7}$$

falls $x_n \rightarrow x$ (Gleich wird $x_n := np_n \rightarrow x := \lambda$ gesetzt werden). Nun betrachten wir den letzten Faktor von $b(n, p_n; k)$,

$$(1-p_n)^{n-k} = \left(1 - \frac{np_n}{n}\right)^{n-k} = \left(1 - \frac{np_n}{n}\right)^n (1-p_n)^{-k}.$$

Da $p_n \rightarrow 0$ und k von n unabhängig ist, geht der letzte Faktor gegen 1, der erste wegen (3.7) gegen $e^{-\lambda}$ – man setze $x_n := np_n$.

Jetzt fehlt nur noch

$$(n)_k p_n^k \rightarrow \lambda^k.$$

Das folgt aber einfach aus

$$(n)_k p_n^k = \frac{(n)_k}{n^k} (np_n)^k,$$

da der erste Faktor gegen 1 konvergiert, der zweite aber gegen λ^k , da $np_n \rightarrow \lambda$ und die Funktion $x \mapsto x^k$ stetig ist. ■

³⁵Die linke Seite ist gerade $b(n, p_n; k)$, die rechte $\pi(\lambda; k)$. Es gilt $\binom{n}{k} = \frac{(n)_k}{k!}$.

Poissonverteilungen treten bei *seltenen* Ereignissen auf, λ ist dabei die zu erwartende Anzahl von Ereignissen in der jeweiligen Zeiteinheit. Mit einer Poissonverteilung kann z.B. die Wahrscheinlichkeit für die Anzahl tödlicher Verkehrsunfälle pro Jahr auf einem bestimmten Autobahnabschnitt beschrieben werden, wenn die zu erwartende Anzahl von Unfällen pro Jahr als gegeben angesehen wird.

JUMBO-Applets:

- **Poissonverteilung**³⁶. Hier wird angenommen, dass die Anzahl von Haifischen, die sich pro Stunde in einem bestimmten Meeresabschnitt aufhalten, poissonverteilt ist. Dann entspricht der Parameter λ der *mittleren Zahl von Haifischen* pro Stunde in dem Meeresabschnitt.
- **Binomial- und Poissonverteilung**

3.5.4 Geometrische Verteilung

Wiederholen wir ein Bernoulli-Experiment so lange, bis eine Eins kommt, so kann man nach der Wahrscheinlichkeit $f(k)$ fragen, dass die „Eins“ nach k -ten „Fehlversuchen“ eintritt. Ist p die Wahrscheinlichkeit für eine „Eins“, so ist

$$f(k) = (1 - p)^k \cdot p. \quad (3.8)$$

Der Merkmalraum ist $\Omega = \mathbb{N}_0$, da im Prinzip beliebig viele Fehlversuche denkbar sind. Dass es sich um eine Wahrscheinlichkeits-Verteilung handelt, liegt an

$$\sum_{k=0}^{\infty} f(k) = p \cdot \sum_{k=0}^{\infty} (1 - p)^k = 1,$$

da die geometrische Reihe $\sum_{k=0}^{\infty} (1 - p)^k$ den Wert $1/p$ hat (siehe Satz III.1.25). Daher nennt man die Verteilung (3.8) **geometrische Verteilung**.

Anwendungen dieser Verteilung gibt es viele:

- Die Wahrscheinlichkeit, mit dem Fahrrad auf dem täglichen Weg zu verunglücken, sei p . Wie groß ist die Wahrscheinlichkeit, dass man mindestens 1000 Tage unfallfrei fährt? Wie groß ist die Wahrscheinlichkeit, am (bösen) 7. Tag zu verunglücken?
- Die Wahrscheinlichkeit, beim Roulettespiel durch „Setzen auf Rot“ zu gewinnen, sei p (≈ 0.5). Der Spieler wählt die Verdoppelungsstrategie, Einsatz 1 Euro. Er hat 2000 Euro zur Verfügung. Wie groß ist die Wahrscheinlichkeit, dass er die Verdopplungsstrategie nicht durchhalten kann?

³⁶<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomaethe/bio/poiss.html>

3.5.5 Hypergeometrische Verteilung

Die von drei Parametern N, K, n abhängige **hypergeometrische Verteilung** ist verwandt mit der Binomialverteilung. Wenn man \mathbf{N} wohl unterscheidbare Exemplare in der Urne hat (z.B. Lose mit jeweils einer Nummer), von denen \mathbf{K} ausgezeichnet sind (z.B. als Gewinnlose) und zieht man \mathbf{n} mal, ohne zurückzulegen (Anzahl der Lose, die man kauft), so kann man nach der Wahrscheinlichkeit fragen, dass genau \mathbf{j} (Gewinnlose) ausgezeichnete Exemplare gezogen werden. Das Ergebnis hängt natürlich von den drei Parametern N, K und n sowie von j ab und errechnet sich zu (Vorlesung??)

$$h(N, K, n; j) := \frac{\binom{K}{j} \binom{N-K}{n-j}}{\binom{N}{n}}, 0 \leq j \leq n.$$

Durch $f(k) := h(N, K, n; j), j = 1, 2, \dots, n$ ist also die Zähldichte der hypergeometrischen Verteilung gegeben.

Mit Zurücklegen wäre $p := K/N$ die Treffer-Wahrscheinlichkeit und wir hätten die Binomialverteilung. Falls $N > K \gg n \geq j$ gilt, erwartet man, dass es kaum einen Einfluss hat, ob zurückgelegt wird oder nicht. Das sieht man ein, wenn man erst zeigt, dass (HÜBNER, S.52)

$$h(N, K, n; j) = \binom{n}{j} \frac{(K)_k (N-K)_{n-j}}{(N)_n}$$

und diesen Ausdruck mit

$$b(n, K/N; j) = \frac{\binom{n}{j} K^k (N-K)^{n-j}}{N^n}$$

vergleicht (HÜBNER, S.63).

Eine weitere Anwendung der hypergeometrischen Verteilung ist die der *Qualitätskontrolle*. Wenn von N Produkten K schadhaft sind und man n Produkte in einer Stichprobe herausgreift, gibt $h(N, K, n; j)$ die Wahrscheinlichkeit an, dass genau j Produkte der Stichprobe schadhaft sind.

Eine andere Anwendung: Eine Lieferung enthält N Produkte, von denen K schadhaft sind. Es wird eine Stichprobe vom Umfang n genommen. Wie groß ist die Wahrscheinlichkeit, dass unter diesen j schadhaft sind?

K ist meist unbekannt. Der Kunde kann die Lieferung akzeptieren, wenn K einen bestimmten Wert K_0 nicht überschreitet (Ausschuss-Anteil K_0/N noch akzeptabel). Mit Hilfe der Stichprobe will er K_0/N schätzen (durch k/n , wenn k die Anzahl der schadhaften Produkte der Stichprobe ist. Wie groß muss er n wählen, um mit einer Irrtums-Wahrscheinlichkeit von 5% die Lieferung zu akzeptieren, obwohl sie kritisch ist? PFLANZAGL, S.41.

3.6 Kontinuierliche Verteilungen

Die einfachsten, aber wichtigsten **kontinuierlichen** Wahrscheinlichkeits-Modelle haben $\Omega = \mathbb{R}$ als Merkmalraum, ein Zufallsexperiment hat als Ergebnis eine Zahl, kann also auch als

reelle Zufallsvariable X angesehen werden. Hier kann es um gemessene Längen, Temperaturen, Zeiten (Lebensdauern), usw, alles also kontinuierliche Größen gehen. Man interessiert sich i.a. für $P(I) = P(a \leq X \leq b)$, wobei $I = [a, b]$ ein Intervall ist. $P(I)$ ist die Wahrscheinlichkeit, dass das Ergebnis des Zufallsexperiments in I liegt. Auch hier macht der Begriff der **Verteilungsfunktion** $F^X : \mathbb{R} \rightarrow \mathbb{R}$ Sinn (vergleiche Def. 3.3 für diskrete Modelle):

Definition 3.10. Zu einem beliebigem W -Modell mit W -Maß P ist die **Verteilungsfunktion** $F^X : \mathbb{R} \rightarrow \mathbb{R}_+$ durch

$$F^X(x) := P(X \leq x)$$

definiert.

F^X ist eine auf ganz \mathbb{R} definierte reelle Funktion mit Werten zwischen 0 und 1,

$$0 \leq F^X(x) \leq 1 \text{ für alle } x \in \mathbb{R},$$

sie ist natürlich monoton wachsend, hat aber im Gegensatz zu diskreten Verteilungsfunktionen – dort stellt sie eine Treppenfunktion mit Stufen bei $\omega \in \Omega$ dar – nicht zwingend mehr Sprungstellen, kann also durchaus **stetig**³⁷ sein. F^X wächst monoton, „beginnend“ mit dem Wert 0 und „endend“ bei dem Wert 1³⁸.

Wie auch im diskreten Fall gilt $P(\alpha < X \leq \beta) = F^X(\beta) - F^X(\alpha)$ für alle $\alpha, \beta \in \mathbb{R}$. Ist F^X stetig, so gilt auch

$$P(\alpha \leq X \leq \beta) = F^X(\beta) - F^X(\alpha),$$

was im diskreten Fall nur für $\alpha \notin \Omega$ richtig ist.

3.6.1 Rechtecksverteilung

Beispiel: Stellen wir uns einen Stab der Länge 3 Meter vor, der an einer zufällig ausgewählten Stelle x durchgeschnitten werden soll. Jede Schnittstelle soll gleich wahrscheinlich sein. Wir legen den Stab in das Intervall $[0, 3]$ der reellen Achse. Dann ist die Wahrscheinlichkeit, dass die Schnittstelle unterhalb der Position x ist, gerade $F(x) = x/3$ für $0 \leq x \leq 3$, also für $x \in [0, 3]$. Es ist $F(x) = 0$ für $x < 0$ und $F(x) = 1$ für $x > 3$. Man spricht von einer *Rechtecksverteilung*³⁹, die bezogen auf $[a, b]$ an Stelle von $[0, 3]$ so definiert ist:

Definition 3.11. Sei zu $a < b$ das Intervall $[a, b]$ gegeben. Dann heißt

$$F^R(x) = \begin{cases} 0 & \text{für } x < a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{sonst} \end{cases}$$

Verteilungsfunktion zur Rechtecksverteilung zu $[a, b]$.

³⁷engl. *continuous*, daher wird *kontinuierlich* zuweilen auch durch *stetig* ersetzt

³⁸Für kleine t ist $F^X(t) \approx 0$ und für große x gilt $F^X(x) \approx 1$, genauer: $\lim_{x \rightarrow -\infty} F^X(x) = 0, \lim_{x \rightarrow +\infty} F^X(x) =$

1

³⁹Der Name kommt von der Gestalt der zugehörigen W -Dichte, s.u.

Sie definiert durch

$$P([\alpha, \beta]) (= P(\alpha \leq X \leq \beta)) = F^R(\beta) - F^R(\alpha)$$

ein W-Maß.

3.6.2 Wahrscheinlichkeits-Dichte

Jetzt kommen wir zum wichtigen und schwierigen Begriff einer (kontinuierlichen) *Wahrscheinlichkeits-Dichte* zu einem kontinuierlichen W-Maß. Hierzu wird auf den *Integralbegriff* (siehe Kap. III.3.7) zurückgegriffen.

Definition 3.12. Gegeben sei ein kontinuierliches W-Modell mit W-Maß P . Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}_+$, die für alle $\alpha < \beta$

$$P(\alpha \leq X \leq \beta) = P([\alpha, \beta]) = \int_{\alpha}^{\beta} f(x) dx \quad (3.9)$$

erfüllt, heißt **Wahrscheinlichkeits-Dichte** (oder auch kurz **W-Dichte**) zum W-Modell. Man sagt, dass die zugehörige Verteilungsfunktion $F^X(x)$ eine **kontinuierliche Dichte besitzt**.

Wenn wir uns

$$P(\alpha \leq X \leq \beta) = F^X(\beta) - F^X(\alpha)$$

vor Augen führen und uns an den Begriff *Stammfunktion* erinnern, so erhalten wir

Satz 3.13. Die Verteilungsfunktion $F^X(x)$ besitzt genau dann eine kontinuierliche W-Dichte f , wenn f F^X als Stammfunktion besitzt⁴⁰. Es gilt

$$F^X(\beta) - F^X(\alpha) = \int_{\alpha}^{\beta} f(x) dx.$$

Bemerkung: Aus den W-Maß-Eigenschaften kann man mit Hilfe *uneigentlicher Integrale*⁴¹

$$F^X(t) = \int_{-\infty}^t f(x) dx$$

und

$$P(\mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx = 1$$

gewinnen. Daher kann man zu irgendeiner nichtnegativen reellen Funktion $f : \mathbb{R} \rightarrow \mathbb{R}_+$, die

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

⁴⁰d.h., falls $F' = f$, genauer falls $F'(x) = f(x)$ für alle x (bis auf endlich viele)

⁴¹Das sind Integrale mit $\pm\infty$ als Integrationsgrenzen.

erfüllt, ein kontinuierliches W-Modell durch (3.9) gewinnen, das dann auf die Verteilungsfunktion

$$F^X(t) := \int_{-\infty}^t f(x)dx$$

führt.

In der Realität kann die zugehörige Zufallsvariable nicht beliebig kleine (negative) und / oder beliebig große Werte annehmen. Dann gibt es Zahlen $m < M$ mit $F^X(x) = 0$ für $x \leq m$ und $F^X(x) = 1$ für $x \geq M$ und wir haben es mit *eigentlichen* Integralen zu tun:

$$F^X(t) = \int_m^t f(x)dx, \text{ bzw. } P(m \leq X \leq M) = P([m, M]) = \int_m^M f(x)dx = 1. \quad (3.10)$$

Die Dichte $f^R(x)$ zu einer **Rechtecksverteilung** zum Intervall $[a, b]$ hat die *Rechtecks*-Gestalt

$$f^R(x) = \begin{cases} \frac{1}{b-a} & \text{falls } x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

Man beachte, dass die Fläche des Rechtecks 1 ist, so dass es sich wirklich um eine Wahrscheinlichkeits-Dichte handelt. Jetzt erkennt man auch, warum die Rechtecksverteilung auch als **Gleichverteilung** in Bezug auf $[a, b]$ angesehen werden kann. Wie bei obigem Zuschnittproblem ist jedes $x \in [a, b]$ „gleichwahrscheinlich“. Die Zufallsvariable kann allerdings keine Werte außerhalb des Intervalls annehmen.

Die nach der Rechtecksverteilung nächst einfache Verteilung ist eine **Dreiecksverteilung**, deren Wahrscheinlichkeitsdichte durch ein Dreieck über $[a, b]$ der Höhe $(b-a)/2$ gegeben ist (dann ist die Dreiecksfläche wieder 1). Im Gegensatz zur Rechtecksverteilung wird hier die Mitte von $[a, b]$ im Vergleich zum Rand bevorzugt. Die Verteilungsfunktion setzt sich aus zwei Parabelbögen zusammen.

Wenn man den Ansatz in (3.9) verstanden hat, versteht man, warum Integration (s. Kap. III.3.7) in den „Anwendungen“ so wichtig ist: Mit Integralen wird etwas „gemessen“, hier eine Wahrscheinlichkeit, in der Schule war es eine Fläche, in der Physik sind es Ladung, Masse, u.a. Man beachte das diskrete Analogon, für das $P([a, b]) = \sum_{\omega \in I} f(\omega)$ gilt. Integration ist eine „kontinuierliche Summe“.

3.6.3 Exkurs Integration

Ich wiederhole ganz kurz einige Fakten aus Kap. III.3.7.

Mit $\int_a^b f(x)dx$ misst man die Fläche zwischen Graphen von f und x -Achse zwischen $x = a$ und $x = b$, wobei die Flächenteile unterhalb der x -Achse negatives Vorzeichen bekommen.

f besitzt eine **Stammfunktion** F , wenn für die Ableitung $F'(x) = f(x)$ für alle x gilt. Dieser Begriff findet seine Berechtigung in dem *Hauptsatz der Integral- und Differentialrechnung*, der auf

$$\int_a^b f(x)dx = F(b) - F(a)$$

führt.

Eine kontinuierliche Verteilung F besitzt also nur dann eine W-Dichte, falls F differenzierbar ist, die Dichte gewinnt man dann durch *Differentiation*: $f(x) = F'(x)$. Ist F eine Treppenfunktion, so ist F in den Sprungstellen nicht differenzierbar (noch nicht einmal stetig!), es gibt keine Dichte im kontinuierlichen Sinne (nur eine „Zähldichte“).

Im Falle der Rechtecksverteilung ist F nur an zwei Stellen nicht differenzierbar. Dies reicht aus, um die W-Dichte zu erhalten.

Uneigentliche Integrale mit $\pm\infty$ in den Integrationsgrenzen kann man als Grenzwerte von eigentlichen Integralen verstehen, z.B. ist $s := \int_{-\infty}^b f(x)dx$ der Wert des Integrals, wenn für jede gegen $-\infty$ divergierende Folge (a_n) die Folge der bestimmten Integrale $\int_{a_n}^b f(x)dx$ gegen s konvergiert.

Beispiel:

$$\int_{-\infty}^b e^x dx = e^b,$$

da

$$\int_{a_n}^b e^x dx = e^b - e^{a_n}$$

und $e^{a_n} \rightarrow 0$, wenn $a_n \rightarrow -\infty$.

Uneigentliche Integrale müssen nicht *konvergieren*, genauso, wie unendliche Reihen, mit denen sie vieles gemein haben, nicht konvergieren müssen. Zum Beispiel divergiert $\int_1^\infty \frac{1}{x} dx$, da $\ln x$ eine Stammfunktion des Integranden ist und

$$\int_1^{b_n} \frac{1}{x} = \ln b_n.$$

Diese Folge divergiert gegen $+\infty$, wenn $b_n \rightarrow \infty$.

Bei der Normalverteilung haben wir es mit dem Integranden e^{-x^2} zu tun, dessen Stammfunktion man nicht mit den uns bekannten Funktionen ausdrücken kann. Da aber e^{-x^2} sehr rasch abklingt, wenn $x \rightarrow \pm\infty$, ist die Fläche unter seinem glockenförmigen Graphen endlich, auch wenn wir die Berandungen ganz nach außen schieben⁴². Genauer: Es gilt

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi},$$

eine überraschende Verbindung zwischen der Eulerschen Zahl e und der Kreiszahl π .

⁴²Die blaue Fläche in Abb. 3.10 ist endlich.

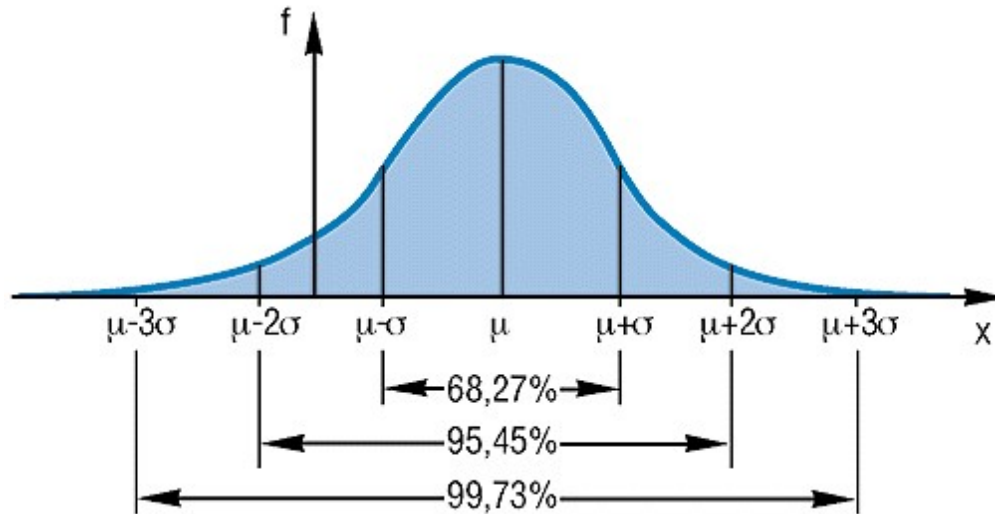


Abbildung 3.10: Dichte der Normalverteilung

3.6.4 Normalverteilung - erster Zugang

Das wichtigste kontinuierliche Beispiel ist die W-Dichte der **Normalverteilung** $\mathfrak{N}(\mu, \sigma^2)$, die durch

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

definiert ist, siehe Abb. 3.10. Die Dichte heißt auch „Glockenkurve“.

In Übereinstimmung mit dem diskreten Fall ist die *Verteilungsfunktion* der Normalverteilung durch

$$F(t) = P(X \leq t) = \int_{-\infty}^t f(x) dx$$

definiert, eine Funktion mit einem sigmoiden Graphen, siehe Abb. 3.11.

Ähnlich wie bei der Poissonverteilung kann die Zufallsvariable einer Normalverteilung theoretisch beliebig kleine und große Werte annehmen. Die Wahrscheinlichkeit hierfür ist zwar immer positiv, aber möglicherweise „verschwindend klein“: die Dichte in Abb. 3.10 klingt für $x \rightarrow \pm\infty$ sehr, sehr rasch ab.

Mehr werden Sie in Kap. 3.11 erfahren, nachdem *Maßzahlen (Kenngrößen)* von Verteilungen wie *Erwartungswert* (hier μ) und *Varianz* (hier σ^2) eingeführt wurden.

Dass die Normalverteilung eine herausragende Bedeutung hat, liegt am *Zentralen Grenzwertsatz*⁴³. In der Statistik beruhen viele Tests auf der durch den Zentralen Grenzwertsatz gestützten

⁴³Dieser besagt vereinfacht, dass eine Summe von n unabhängigen Verteilungen für $n \rightarrow \infty$ gegen eine Normalverteilung konvergiert. Die Binomialverteilung $b(n, p; k)$ ist eine Summe von n Bernoulli-Verteilungen. Daher hat die Binomialverteilung für große n die Form einer Glockenkurve

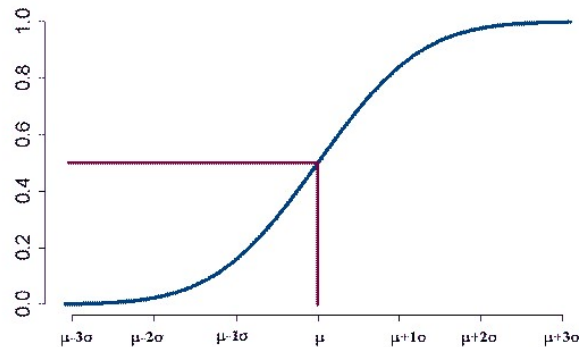


Abbildung 3.11: Verteilungsfunktion der Normalverteilung

Annahme, dass gewisse Merkmale normalverteilt sind, allerdings mit i.a. unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 , welche auf Grund einer Stichprobe *geschätzt* werden können.

Applets:

- In dem **Applet zur Dichte und Verteilungsfunktion der Normalverteilung** (JUMBO) werden wahlweise die Dichte und die Verteilungsfunktion der Normalverteilung für einstellbare Werte von μ und σ gezeigt.
- Eine Animation zu **Galton's Nagelbrett**⁴⁴ will demonstrieren, dass eine Normalverteilung im Spiel ist.

3.6.5 Bemerkungen zu kontinuierlichen Verteilungen in der Statistik

In der Statistik spielen noch weitere kontinuierliche Verteilungen eine Rolle. So z.B. gibt die *Student(n)*-Verteilung an, wie der Mittelwert einer normalverteilten Stichprobe vom Umfang n verteilt ist (s. HÜBNER, S.182).

Die *Chi-Quadrat-Verteilung* (s. HÜBNER, S.71) ist die Verteilung der Zufallsvariablen X^2 , wenn X eine normalverteilte Zufallsvariable ist. Sie kommt ins Spiel, wenn man eine gewisse Verteilung (Zähldichte) einer diskreten Zufallsvariablen absichern will, z.B. die Gleichverteilung für die Zahlen 0, 1, ..., 36 beim Roulette (s. HÜBNER, S.187).

⁴⁴<http://statistik.wu-wien.ac.at/mathstat/hatz/vo/applets/Galton/galton.html>

Generell ist jede „bei Null beginnende, bei Eins endende“ monotone Funktion F mit nichtnegativen Werten eine mögliche Verteilungsfunktion zu einer Zufallsvariablen X . Handelt es sich um eine Treppenfunktion mit endlich vielen Stufen der Höhe p_j an den Stellen $\omega_j, j = 1, 2, \dots, n$, so handelt es sich um eine diskrete Zufallsvariable mit den Werten ω_j , die mit Wahrscheinlichkeit p_j angenommen werden.

3.7 Bedingte Wahrscheinlichkeiten, unabhängige Ereignisse

Wir betrachten zwei Ereignisse A und B eines Wahrscheinlichkeits-Modells. Wenn man nacheinander n Zufallsexperimente dieses W-Modells ausführt, kann man nach der relativen Häufigkeit fragen, dass A eintritt, sofern gleichzeitig B eingetreten ist. Es ist klar, dass die gesuchte Zahl der Quotient aus der absoluten Häufigkeit n_{AB} für AB (A und B) und der von $B - n_B -$ und damit auch der Quotient ihrer relativen Häufigkeiten ist. Daher ist es auf Grund des empirischen Gesetzes der großen Zahlen folgerichtig zu definieren:

Definition 3.14. Die **bedingte Wahrscheinlichkeit** $P(A|B)$ von A unter der Bedingung B wird durch

$$P(A|B) := \frac{P(AB)}{P(B)}. \quad (3.11)$$

definiert.

Hieraus ergibt sich

$$P(AB) = P(B)P(A|B) \quad (3.12)$$

oder auch

$$P(AB) = P(A)P(B|A).$$

Definition 3.15. Man nennt zwei Ereignisse A und B eines Wahrscheinlichkeits-Modells **stochastisch unabhängig**, falls $P(A) = P(A|B)$.

Die stochastische Unabhängigkeit von A und B ist also dann gegeben, wenn die Wahrscheinlichkeit für A unabhängig davon ist, ob gleichzeitig B eintritt oder nicht (oder so ausgedrückt: Das Eintreten des Ereignisses A wird durch B nicht beeinflusst).

Hieraus ergibt sich

Satz 3.16. Zwei Ereignisse A und B eines W-Modells sind stochastisch unabhängig genau dann, wenn

$$P(AB) = P(A) \cdot P(B).$$

Wenn zweimal hintereinander gewürfelt wird, ist man versucht zu akzeptieren, dass das Ergebnis des zweiten Wurfes nicht von dem des ersten beeinflusst wird. Der Würfel hat ja schließlich kein Gedächtnis.

Das können wir mit Hilfe der Annahme, dass alle Ergebnispaare (m, n) mit $1 \leq m, n \leq 6$ gleichwahrscheinlich sind (Wahrscheinlichkeit = $1/36$) und mit Hilfe von Satz 3.16 nachvollziehen, da $1/36 = 1/6 \cdot 1/6$ (Man setze $A :=$ „Im ersten Wurf wird ein m geworfen“ und $B :=$ „Im zweiten Wurf wird ein n geworfen“).

Bei dem Wahrscheinlichkeits-Modell der Binomialverteilung (n -maligen Bernoulliexperiment) wurde der Term $p^j(1-p)^{n-j}$ mit der stochastischen Unabhängigkeit begründet. Dies können wir jetzt nachvollziehen. Wir tun dies für $n = 2, j = 1$: Die Wahrscheinlichkeit, beim ersten Bernoulli-Experiment eine Eins zu erzielen, ist p , die, beim zweiten B-Experiment eine Null zu erzielen, ist $1 - p$. Wenn beide Experimente unabhängig sind, ist die Wahrscheinlichkeit, erst eine Eins und danach eine Null zu „werfen“, nach obiger Definition $p \cdot (1 - p)$.

Beim Münzwurf bedeutet die Unabhängigkeit zweier Würfe, dass das Ergebnis des zweiten Wurfs einer Münze durch das des ersten Wurfs nicht beeinflusst wird. Dennoch führt dies immer wieder zu Irrtümern: Wenn 10 Mal hintereinander „Zahl“ kam, meint man, dass im 11. Wurf die Wahrscheinlichkeit, wieder „Zahl“ zu werfen, kleiner als 0,5 sein sollte. Selbst D’ALEMBERT ließ sich hiervon selbst von EULER nicht abbringen!

Beim Lottospiel *7 aus 49* sind die Ereignisse A „als erste Zahl wird die 13 gezogen“ und B „als zweite Zahl wird die 31 gezogen“ (leicht) stochastisch abhängig, weil $P(AB) = \frac{1}{49 \cdot 48} \neq \frac{1}{49^2} = P(A)P(B)$. Das liegt daran, dass $P(B|A) = 1/48 > P(B) = 1/49$, da für B unter der Bedingung A nur noch 48 statt 49 Zahlen in Frage kommen.

Bei statistischen Stichproben sollen die Daten stochastisch unabhängig von einander erhoben werden. D.h., dass z.B. die verschiedenen Versuchspersonen zufällig und unabhängig von einander ausgewählt werden. Theoretisch muss dies auch dazu führen können, dass innerhalb einer Stichprobe ein Individuum mehrmals untersucht wird. Bei großen Grundgesamtheiten und verhältnismäßig kleinem Umfang der Stichprobe wäre dies sehr unwahrscheinlich, so dass man durchaus von vornerein verschiedene Individuen untersuchen kann.

Es gibt eine schöne geometrische Veranschaulichung in der Ebene. Ω sei das Einheitsquadrat und $A \subset \Omega$ sei eine Teilmenge mit einem Flächeninhalt $P(A)$. Man kann $P(A)$ als die Wahrscheinlichkeit interpretieren, dass man mit einem Wurfpeil A trifft - vorausgesetzt, jeder Punkt von Ω wird mit derselben Wahrscheinlichkeit getroffen.

1. Wenn $B \subset A$ gilt also $P(AB) = P(B)$ und $1 = P(A|B)$. Klar: Wenn B getroffen wurde, dann erst recht A .
2. Wenn $P(AB) > P(A) \cdot P(B)$, also $\frac{P(AB)}{P(B)} > P(A)$, gilt also $P(A|B) > P(A)$, d.h. unter der Annahme B wird A wahrscheinlicher als ohne diese Annahme.
3. Wenn $P(AB) = P(A) \cdot P(B)$, also $P(A|B) = P(A)$, verändert sich die Wahrscheinlichkeit für A nicht, wenn B angenommen wird.
4. Wenn $P(AB) < P(A) \cdot P(B)$, also $\frac{P(AB)}{P(B)} < P(A)$, gilt also $P(A|B) < P(A)$, d.h. unter der Annahme B wird A unwahrscheinlicher als ohne diese Annahme.

5. Wenn $AB = \emptyset$, gilt $0 = P(AB) < P(A) \cdot P(B)$ und insbesondere $0 = P(A|B) < P(A)$, d.h. unter der Annahme B wird A total unwahrscheinlich.

Im **Applet - Bedingte Wahrscheinlichkeiten**⁴⁵ kann man z.B. erkennen, dass die Ereignisse „rot“ und „ungerade“ abhängig, „blau“ und „ungerade“ aber unabhängig sind.

Man kann sich die Definition (3.14) – wie auch die Kolmogoroff-Axiome – sehr gut plausibel machen, indem man Wahrscheinlichkeiten als *Anteil* an einer Fläche vorstellt. $P(A)$ ist dann der Anteil von A an Ω und $P(A|B)$ ist der Anteil von $A \cap B$ an B .

3.7.1 Beispiel für bedingte Wahrscheinlichkeiten aus der Medizin

Sei Ω eine Bevölkerung und A die Menge aller derjenigen Personen dieser Bevölkerung, die an einer bestimmten Krankheit leiden. Für diese Krankheit gebe es einen Test, der bei gewissen Personen (hierdurch wird eine Menge $B \subset \Omega$ definiert) positiv ausfällt. Der Test sei bei Kranken sicher, d.h. es gelte $A \subset B$. Bei Gesunden sei der Test mit einer Wahrscheinlichkeit p falsch, d.h. er falle medizinisch positiv aus. Der Anteil der Kranken an der Bevölkerung sei $a = |A|/|\Omega|$ – man kann auch $P(A) = a$ sagen⁴⁶. Wie groß ist die Wahrscheinlichkeit, dass eine Person wirklich krank ist, wenn der Test positiv war?

Es geht also um die bedingte Wahrscheinlichkeit $P(A|B)$, welche in der Medizin *Spezifität* des Tests genannt wird. Gegeben ist $P(A) = a$, $P(B|A) = 1$ und $P(B|\bar{A}) = p$. Der Allgemeinheit wegen setze ich $P(B|A) = b$, so dass auch der Fall erfasst wird, dass der Test bei Kranken negativ ausfällt⁴⁷. Wegen (3.12) gilt

$$P(A|B) = \frac{P(AB)}{P(B)},$$

d.h. wir müssen $P(AB)$ und $P(B)$ kennen. Nun ist $P(AB) = P(B|A)P(A) = ab$. Fehlt noch $P(B)$. Wegen $B = (A + \bar{A})B = AB + \bar{A}B$ gilt

$$P(B) = P(AB) + P(\bar{A}B).$$

Und wegen

$$p = P(B|\bar{A}) = \frac{P(\bar{A}B)}{P(\bar{A})}$$

folgt $P(\bar{A}B) = p(1 - a)$ und damit $P(B) = ab + p(1 - a)$. Damit ergibt sich

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{ab}{ab + p(1 - a)}.$$

⁴⁵<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomathe/bio/prob2.html>

⁴⁶Man nennt a die Prävalenz.

⁴⁷ b heißt *Sensitivität*.

So ergibt sich z.B. bei $a = 0.01 = p, b = 1$, bei also 1% Kranken und einer Test-Unsicherheit von 1% die Wahrscheinlichkeit von fast 50%, dass eine positiv getestete Person wirklich krank ist.

Solche Beispiele müssen Medizin-StudentInnen rechnen können, was manche sehr quält. Ich habe mal in einem Artikel gelesen, dass der Begriff Wahrscheinlichkeit von Laien nur schwer verständlich ist, erst recht das Rechnen mit Wahrscheinlichkeiten. Leichter sei es, in Anteilen einer (fiktiven) Grundgesamtheit zu argumentieren. Das sähe in unserem Zahlenbeispiel so aus: Von 10000 Personen sind 100 krank ($a=1\%$), von den 9900 Gesunden werden 1%, also 99 positiv getestet. Von insgesamt 199 positiv getesteten Personen sind also nur 100 (ca. 50%) wirklich krank.

Formeln, die von (bekannten) bedingten Wahrscheinlichkeit $P(A|B_j), j = 1, 2, \dots, m$, auf $P(B_k|A), k = 1, 2, \dots, m$ schließen lassen, heißen in der Literatur *Bayes-Umkehrformeln*, in der einfachsten Form

Satz 3.17.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}$$

Überzeugen Sie sich, dass dies genau die obige Formel ist!

Es gibt auch andere medizinische Anwendungen. Z.B., wenn B ein Symptom einer Krankheit A ist und man sich für die Wahrscheinlichkeit interessiert, dass die Krankheit bei dem Symptom vorliegt. Man muss nur $P(B|A)$ (Wahrscheinlichkeit für das Symptom bei Kranken), $P(A)$ (Anteil der Kranken) und $P(B|\bar{A})$ (Wahrscheinlichkeit für das Symptom bei Gesunden) kennen.

3.8 Reelle Zufallsvariable

Ein Wahrscheinlichkeits-Modell ist durch einen Merkmalraum Ω und durch dessen Potenzmenge⁴⁸ als Ereignissystem zusammen mit einem W-Maß P gegeben. Bei *quantitativem* Merkmalraum wurde jedem Elementarereignis eine (natürliche oder reelle) Zahl zugeordnet, oder anders gesagt: jedes Zufallsexperiment hat eine reelle Zahl als Ergebnis, wir hatten in diesem Zusammenhang schon von einer *Zufallsvariablen* gesprochen und diese X genannt. Die genaue Definition ist wie folgt:

Definition 3.18. *Jede reelle Abbildung $X : \Omega \rightarrow \mathbb{R}$, die einem Elementarereignis $\omega \in \Omega$ eine Zahl $X(\omega)$ zuordnet, definiert eine (reelle) **Zufallsvariable**.*

Ist $\Omega \subset \mathbb{R}$, der Merkmalraum also quantitativ, so gibt es natürlich die „triviale Zufallsvariable“, die jedem $\omega \in \Omega$ just dieses ω zuordnet, in Form der *Identität*, die wir in Kap. I.8.4 im Rahmen des Begriffs *Neutrales Element* einer Verknüpfung (hier *Verkettung von Abbildungen*) kennengelernt haben. Diese triviale Zufallsvariable ist es, die wir bisher gemeint haben, wenn wir von

⁴⁸Dies ist eigentlich nur für endliche Merkmalräume richtig, sonst muss man σ -Algebren betrachten

Zufallsvariablen im Zusammenhang mit Binomial-, Poisson-Verteilungen etc. gesprochen hatten. Aber auch bei qualitativem Merkmalraum kann man sich diesen durch eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ „quantifiziert“ vorstellen, indem man z.B. die beiden Geschlechter oder die beiden Seiten einer Münze oder Nationalitäten durch Zahlen „codiert“.

Als einfaches Beispiel betrachten wir nochmals das zweimalige Würfeln mit einem Würfel. Hier ist der differenzierteste Merkmalraum $\Omega = \{1, 2, \dots, 6\}^2$. Die Zufallsvariable „Augensumme“ ordnet jedem Zahlenpaar (a, b) deren Summe $a + b$ aus $\Omega' := \{2, 3, \dots, 12\}$ zu. In diesem Zusammenhang hatten wir von einem „reduzierten“ Merkmalraum gesprochen.

Oder man betrachte ein noch einfacheres Wahrscheinlichkeits-Modell: den Wurf mit nur einem Würfel, der allerdings keine sechs Ziffern, sondern sechs verschiedene Farben auf seinen Seiten aufweist. Wenn wir jetzt jeder Farbe eine Zahl zuordnen, haben wir eine (reelle) Zufallsvariable.

Betrachten wir wieder eine allgemeine Situation, so interessieren wir uns nun für ganz spezielle, mit einer Zufallsvariablen X zusammenhängende Ereignisse, indem wir einem $A' \subset \mathbb{R}$ ein Ereignis $A := \{\omega \in \Omega : X(\omega) \in A'\}$ zuordnen (Urbild von A' unter X), das auch kurz mit $A := \{X \in A'\}$ beschrieben wird.⁴⁹ Ist Ω endlich, so auch $X(\Omega)$, d.h. X nimmt nur endlich viele Werte an, von denen es nicht mehr als $|\Omega|$ geben kann. Blickt man auf das W-Modell durch die X -Brille, hat man i.a. den Ereignisraum reduziert, da nicht alle Teilmengen von Ω Urbilder unter X sind. Man kann auch Ω vergessen und nur $\Omega' = f(\Omega)$ mit Elementarereignissen $\omega' \in \Omega'$ und Elementar-Wahrscheinlichkeiten $P(\{\omega'\}) := P(X = \omega')$ betrachten, d.h. man hat ein reduziertes W-Modell mit einem quantitativen Merkmalraum.

3.8.1 Verteilung und Verteilungsfunktion

Also gibt es Sinn, von einer (Wahrscheinlichkeits-) **Verteilung einer Zufallsvariablen** zu sprechen, wie wir ja schon z.B. bei der Binomialverteilung getan hatten, bei der ja auch eine Zufallsvariable im Spiel war, die die Anzahl der „Einsen“ bei n Bernoulli-Experimenten zählt. Diese Wahrscheinlichkeits-Verteilung entspricht genau der Häufigkeitsverteilung bei Stichproben.

Nimmt die Zufallsvariable nur diskrete Werte an (etwa aus \mathbb{N}_0) – was bei diskreten W-Modellen immer der Fall ist –, so ist die Wahrscheinlichkeits-Verteilung der Zufallsvariablen X durch die Zähldichte

$$f(j) = P(X = j), j \in \mathbb{N}_0,$$

gegeben.

In jedem Fall – diskret oder kontinuierlich – ist die **Verteilungsfunktion**

$$F^X(x) = P(X \leq x), x \in \mathbb{R}$$

der Zufallsvariablen X definiert, so wie schon in Def. 3.10 ausgeführt.

⁴⁹Für nichtdiskrete W-Modelle, bei denen nicht jede Teilmenge von Ω „messbar“ ist, muss man verlangen, dass die Zufallsvariable X messbar ist.

Sie besitzt jedoch im diskreten Fall — wie bei empirischen Verteilungsfunktionen mit diskretem oder klassierten Merkmalraum (s. Kap. 2.4.9)— *keine* kontinuierliche Dichte⁵⁰. Sie gibt sozusagen an, *wie die Zufallsvariable verteilt ist*.

Beispiele:

1. n -maliges Bernoulliexperiment, X = Anzahl der Einsen bei n -maligem Ziehen. Die Verteilung dieses reduzierten W-Modells ist gerade die Binomialverteilung.
2. Eine Lotterie vertreibt 1000 Lose mit den Nummern 000-999, von denen sie 750 verkauft. Es werden 3 Losnummern mit den Gewinnen 10 Euro, 100 Euro und 1000 Euro gezogen. Man kann den Merkmalraum sehr differenziert mit $\Omega := \{0, 1, \dots, 999\}^3$, also mit den möglichen Gewinnlosnummern, ansetzen⁵¹. Definiert man dann $X(\omega)$ als den von der Lotterie auszuzahlenden Gewinn⁵², so hat man eine Zufallsvariable, die allerdings nur endlich viele Werte 0, 10, 100, 110, 1000, 1100, 1010, 1110 annehmen kann. Dennoch, insbesondere bei mehreren Gewinnlosen mit unterschiedlichen Beträgen, ist es sinnvoll, von beliebigen (nichtnegativen) Werten für X auszugehen.
3. Bei einer Blutuntersuchung werden das „gute“ HDL-Cholesterin und das „schlechte“ LDL-Cholesterin als Zahlen zwischen 0 und 500 bestimmt. Man kann also $\Omega := [0, 500]^2$ setzen. Von Interesse ist der Quotient dieser beiden Werte (LDL/HDL) als „Risikofaktor“. Wie ist dieser Risikofaktor in der (deutschen) Bevölkerung verteilt? Welche Werte gelten als „normal“? Gibt es Unterschiede zu einer entsprechenden Verteilung in einem anderen Land?
4. Bei der Geburt eines Kindes werden i.a. mehrere Daten erhoben, insbesondere das Gewicht. Fasst man eine solche Geburt als „Stichprobe“ auf, so ist das ermittelte Gewicht der Wert einer Zufallsvariablen. Bei entsprechend vielen Geburten erhält man eine *empirische Verteilungsfunktion*. Man erwartet eine Normalverteilung (mit Mittelwert $\mu = 3200g$ und Standardabweichung $\sigma = 500g$).
5. In einer Produktion von Lebensmittel werden jeden Tag Stichproben vom Umfang n durchgeführt, die das Gewicht der (abgepackten) Produkte präzise bestimmt. Diese werden in der Regel von dem angestrebten Gewicht etwas abweichen. Jedes Zufallsexperiment besteht hier also aus einer solchen Stichprobe mit der Angabe von n Zahlen G_1, G_2, \dots, G_n . Nun kann man sich für den Mittelwert

$$X(\omega) = \bar{G} = \frac{G_1 + \dots + G_n}{n}$$

⁵⁰Generell besitzen diejenigen Verteilungsfunktionen keine kontinuierliche (stetige) Dichte, die eine Sprungstelle haben

⁵¹Jedes $\omega \in \Omega$ steht für ein Tripel von drei potentiellen Gewinn-Losnummern. Dabei wird hier noch nicht ausgeschlossen, dass zwei oder drei Gewinne auf dieselbe Losnummer fallen können.

⁵²Da nur 750 Lose verkauft wurden, ist es keineswegs sicher, dass die Lotterie den Maximalbetrag von 1110 Euro auszahlen muss.

oder für die empirische Varianz

$$X(\omega) = V := \frac{1}{n-1} \sum_{j=1}^n (G_j - \bar{G})^2$$

interessieren. Beide definieren wieder Zufallsvariable.

Man kann mit Zufallsvariablen rechnen: man kann sie addieren, multiplizieren, etc. Das haben wir auch schon getan: so ist die Zufallsvariable der Binomialverteilung $B(n, p)$ die Summe von n (stochastisch unabhängigen und identisch verteilten) Zufallsvariablen der Bernoulliverteilung $B(p)$. Siehe auch Kap. 3.9.7.

Wenn man sich für die Augensumme von 2 Spielwürfeln interessiert, addiert man praktisch zwei (identische, stochastisch unabhängige) Zufallsvariable zum Wurf eines Würfels.

Für die Summe zweier *stochastisch unabhängiger* Zufallsvariable $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ (siehe Kap. 3.10.2) gibt es schöne Formeln für die (Zähl-) Dichte.

3.9 Kenngrößen von Zufallsvariablen

Die wichtigsten Kenngrößen sind *Erwartungswert* und *Streuung*, aber auch *Median* und sonstige *Quantile*. Sie kennen diese schon als Maßzahlen bei Stichproben (Kap. 2.4), wobei *Mittelwert* mit *Erwartungswert* korrespondiert.

Wir gehen zunächst wieder von *diskreten* W-Modellen mit endlichem Merkmalraum Ω aus.

Eine Zufallsvariable X ist dann durch endlich viele (verschiedene) Werte⁵³ $x_j, j = 1, 2, \dots, m$,⁵⁴ die sie annehmen kann, definiert. Natürlich spielen die Wahrscheinlichkeiten $p_j, j = 1, 2, \dots, m$, eine Rolle, mit der die Werte x_j angenommen werden, also

$$p_j := P(X = x_j), j = 1, 2, \dots, m.$$

Es muss $\sum_{j=1}^m p_j = 1$ sein.

3.9.1 Erwartungswert

Definition 3.19.

$$EX := \sum_{j=1}^m p_j x_j \tag{3.13}$$

heißt **Erwartungswert** der Zufallsvariablen X .

⁵³Die Bezeichnung x_j steht in einem gewissen Widerspruch zu der in Verbindung mit einem Stichprobenvektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$. Sie entspricht nicht der j -ten Komponente einer Stichprobe, sondern dem Element ω_j des Merkmalraums der Stichprobe. Man muss also aufpassen, wenn man mit Hilfe der Zufallsvariablen X eine Stichprobe vom Umfang n vornimmt. Wenn man die Daten dieser Stichprobe auflistet, sind die Bezeichnungen x_k schon vergeben!

⁵⁴Der Merkmalraum Ω kann mehr als m Elemente besitzen, wenn X für verschiedene Elementarereignisse gleiche Werte annimmt

Statt EX schreibt man auch $E(X)$ ⁵⁵.

Wie kommt man auf diese Formel? Sie ergibt sich aus dem Mittelwert einer Stichprobe über das empirische Gesetz der großen Zahl genauso wie sich der Begriff *Wahrscheinlichkeit* aus dem Begriff *relative Häufigkeit* ergibt:

Wenn $p_j \equiv 1/m$, so haben wir es mit einem simplen (arithmetischen) **Mittelwert** zu tun – alle Werte x_j sind gleich wahrscheinlich. Nun nehmen wir an, wir machen n Zufallsexperimente (eine Stichprobe!!) und zählen die Ausgänge: x_j komme n_j mal vor, d.h. es ist $n = n_1 + \dots + n_m$. Bilden wir dann den Mittelwert gewichtet mit der absoluten Häufigkeit n_j , so erhalten wir $\frac{1}{n} \sum_{j=1}^m n_j x_j$. Diese Formel geht in (3.13) über, wenn man gemäß des Gesetzes der großen Zahl die *relativen Häufigkeiten* n_j/n durch die Wahrscheinlichkeiten p_j ersetzt⁵⁶. In Satz 2.4 steht für den Mittelwert von Stichproben die Formel

$$\bar{x} = \sum_{j=1}^m h_j \omega_j.$$

Wenn Sie hier $h_j = p_j$ und $x_j = \omega_j$ setzen, haben Sie (3.13).

Bei abzählbar unendlich vielen Werten geht die Summe in (3.13) in eine Reihe über.

Bei kontinuierlichen Zufallsvariablen wird die Summe durch ein Integral ersetzt, siehe Kap. 3.9.6.

3.9.2 Berechnung der Erwartungswerte für bestimmte Verteilungen

1. Bernoulli $B(p)$: Die Zufallsvariable nimmt den Wert $x_1 = 0$ oder $x_2 = 1$ an. Es ist $p_2 = p$ (und $p_1 = 1 - p$). Dann gilt $EX = p$.
2. Binomialverteilung $B(n, p)$: Es ist $x_j = j, j = 0, 1, 2, \dots, n$, und

$$p_j = b(n, p; j) = \binom{n}{j} p^j (1 - p)^{n-j}.$$

Dann gilt

Satz 3.20. Für den Erwartungswert der Binomialverteilung $B(n, p)$ gilt

$$EX = \sum_{j=0}^n j \cdot b(n, p; j) = np.$$

⁵⁵Diese Schreibweise verdeutlicht, dass es sich bei E um eine *Abbildung* handelt, die einer Zufallsvariablen X deren Erwartungswert $E(X)$ zuordnet

⁵⁶In der Statistik sind die Wahrscheinlichkeiten p_j i.a. nicht bekannt, sie werden durch die relativen Häufigkeiten geschätzt. Entsprechend wird der Erwartungswert durch den Mittelwert geschätzt.

Beweis:

HÜBNER, S.93. Am einsichtigsten ist dieses Ergebnis, wenn man mit X_j die Bernoulli-Zufallsvariable $B(p)$ bei der j -ten Beobachtung bezeichnet und $X = X_1 + \dots + X_n$ sowie $E(X_1 + \dots + X_n) = EX_1 + \dots + EX_n$ ausnutzt. Formal hat man hier Zufallsvariable addiert: die Zufallsvariable der Binomialverteilung $B(n, p)$ ist die Summe von n identischen Bernoulli-Zufallsvariablen $X_j, j = 1, 2, \dots$,⁵⁷.)

Man kann den Beweis aber auch direkt führen. Hierzu mache man sich zunächst klar, dass für $j \geq 1$

$$j \cdot b(n, p; j) = np \cdot b(n-1, p; j-1).$$

Nun ist

$$\sum_{j=0}^n j \cdot b(n, p; j) = np \sum_{j=1}^n b(n-1, p; j-1) = np \sum_{j=0}^{n-1} b(n-1, p; j) = np.$$

■

3. Poissonverteilung $\pi(\lambda)$: Hier ist $x_j = j, j \in \mathbb{N}_0$ und $p_j = e^{-\lambda} \frac{\lambda^j}{j!}$ für abzählbar unendlich viele $j = 0, 1, 2, \dots$. Hier gilt

Satz 3.21. Für den Erwartungswert der Poissonverteilung $\pi(\lambda)$ gilt

$$EX = \lambda.$$

Beweis: Es ist

$$\begin{aligned} EX &= \sum_{j=0}^{\infty} j e^{-\lambda} \frac{\lambda^j}{j!} \\ &= e^{-\lambda} \sum_{j=0}^{\infty} \frac{j \lambda^j}{j!} \\ &= e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\ &= e^{-\lambda} \lambda e^{\lambda} \\ &= \lambda \end{aligned}$$

⁵⁷In diesem Skript wird die Verteilung der Summe von Zufallsvariablen nur kurz behandelt. Dies gehört allerdings zu dem Standardrepertoire von Stochastikbüchern

3.9.3 Median, Quantile

Diese mit dem Zusatz *empirisch* schon für Stichproben in Kap. 2.4.8 definierte Begriffe lassen sich sofort auf Zufallsvariable übertragen, wenn man deren Verteilungsfunktion F^X für die *empirische Verteilungsfunktion* der Stichprobe einsetzt:

Eine Zahl $m \in \mathbb{R}$ heißt **Median** der Zufallsvariable X , wenn

$$F^X(m) = P(X \leq m) \geq 1/2,$$

aber

$$F^X(m') \leq 1/2 \text{ wenn } m' < m.$$

Gibt es ein m mit $F^X(m) = 1/2$, so ist m auch Median (es kann mehrere Mediane geben). Gibt es kein solches m , so ist die Stelle m , an der der Wert $1/2$ „übersprungen“ wird, der Median.

Eine Zahl $u \in \mathbb{R}$ heißt **25%-Quantil** oder erstes **Quartil**, wenn $F^X(u) \geq 0.25$, aber $F^X(u') \leq 0.25$, wenn $u' < u$. Andere Quantile werden analog definiert. Ein Median ist also ein 50%-Quantil, das dritte Quartil ist ein 75%-Quantil. Man schreibt auch $u_{25\%}$ für ein 25%-Quantil. Etwas unscharf kann man sagen: Ist $0 < q < 1$, so hat ein $100q\%$ -Quantil u die Eigenschaft, dass mit der Wahrscheinlichkeit q die Zufallsvariable einen Wert $\leq u$ und mit der Wahrscheinlichkeit $1 - q$ einen Wert $\geq u$ hat.

Liegt eine kontinuierliche Verteilung vor, so dass die Verteilungsfunktion jeden Wert genau einmal annimmt, so sind alle Quantile eindeutig.

3.9.4 Varianz, Streuung

Kennt man $\mu = EX$, so kann man eine neue Zufallsvariable $Y = (X - \mu)^2$ betrachten, die die quadratische Abweichung vom Erwartungswert misst.

Definition 3.22. Der Erwartungswert EY von $Y = (X - \mu)^2$ mit $\mu = EX$ heißt **Varianz** $V(X) := \sum_{j=1}^m p_j (x_j - \mu)^2$ der Zufallsvariablen X .

Die Berechnung kann auch mit Hilfe des folgenden Satzes vorgenommen werden:

Satz 3.23. Es gilt ⁵⁸

$$V(X) := E((X - EX)^2) = E(X^2) - (EX)^2.$$

Beweis: Es gelten die Rechenregeln $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$ und (Linearität des Erwartungswertes)

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

Setzt man $E(X) = \mu$, so ist man fertig. ■

⁵⁸Diese Formel ist auch für kontinuierliche Zufallsvariable anwendbar.

Definition 3.24. Die positive Wurzel aus der Varianz,

$$\text{Str}(X) := \sigma := \sqrt{V(X)}$$

heißt **Standardabweichung** oder **Streuung** von X .

Es liegt die Frage nahe, welche Beziehung die Begriffe *Varianz* und *Standardabweichung* einer Zufallsvariablen X zu der (*empirischen*) *Varianz* und (*empirischen*) *Standardabweichung* einer Stichprobe haben, siehe Kap. 2.4.5. Diese Frage ist nicht so einfach zu beantworten, weil bei der Stichprobe der Nenner $n - 1$ auftaucht. Ohne diesen (und Nenner n) wäre auch die empirische Varianz ein Mittelwert einer mit Hilfe der Stichprobe \mathbf{x} etwas künstlich erzeugten neuen Stichprobe $(x_1 - \mu)^2, (x_2 - \mu)^2, \dots, (x_n - \mu)^2$. Dann wäre das Vorgehen in Kap. 2.4.5 analog zum hiesigen gewesen!

Die wahre Antwort erhalten Sie im Rahmen der „Erwartungstreue“ von Schätzern.

3.9.5 Varianz von bestimmten Verteilungen

1. Bernoulli $B(p)$: Für die Zufallsvariable X gilt $X^2 = X$, also $V(X) = E(X^2) - (EX)^2 = p - p^2 = p(1 - p)$. Die Varianz ist am größten für $p = 0.5$.
2. Binomialverteilung $B(n, p)$: Wenn man hier im Vorgriff auf Kap. 3.9.7 $V(X + Y) = V(X) + V(Y)$ bei stochastisch unabhängigen Zufallsvariablen benutzt und die Zufallsvariable der Binomialverteilung als Summe von n identisch verteilten unabhängigen Bernoulli-Zufallsvariablen sieht sowie obiges ausnutzt, so ergibt sich

Satz 3.25. Die Varianz der Binomialverteilung $B(n, p)$ ist $V(X) = np(1 - p)$.

3. Poissonverteilung $\pi(\lambda)$: Hier ist die Varianz wie schon der Erwartungswert gleich λ .

3.9.6 Kenngrößen bei kontinuierlichen Verteilungen

Hier möchte ich nicht in die Details gehen. Wichtig ist nur, dass es auch bei kontinuierlichen Zufallsvariablen nicht nur eine mit Hilfe einer *W-Dichte* definierte *Verteilungsfunktion*

$$F^X(t) = \int_{-\infty}^t f(x) dx$$

gibt, sondern dass die eben definierten Kenngrößen bei diskreten Zufallsvariablen auch ihr Gegenstück bei kontinuierlichen Zufallsvariablen haben. Hier gebe ich nur an:

$$EX = \int_{-\infty}^{\infty} x \cdot f(x) dx,$$

$$V(X) = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx.$$

Plausibel kann man diese Formel machen, wenn wir die kontinuierliche Zufallsvariable durch eine Art „Klassierung“ diskretisieren: Wir lassen nur die endlich vielen Werte $x_1 < x_2 < \dots < x_m$ zu und teilen die Fläche unter der Dichte $f(x)$ in n Teile, die wir um die x_j gruppieren und setzen $p_j :=$ Fläche des j -ten Teilstücks. Ganz analog, wie in Kap. III.3.7 das Integral durch eine Summe über endlich viele Rechteckflächen angenähert wurde.

Für die Normalverteilungsdichte

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

erhält man $EX = \mu$ und $V(X) = \sigma^2$. Um dies einzusehen, muss man etwas anspruchsvollere Integralrechnung betreiben.

3.9.7 Rechnen mit Zufallsvariablen

Man kann Zufallsvariable addieren, multiplizieren, man kann sie mit einem Skalar multiplizieren, kurz: man kann mit ihnen „rechnen“. Die Frage stellt sich, wie sich dies auf deren Verteilungen und Maßzahlen auswirkt. Dies wollen wir in diesem Skript nicht behandeln, nur kurz sagen, dass⁵⁹

$$E(X + Y) = EX + EY$$

$$E(\alpha X) = \alpha EX, \quad V(\alpha X) = \alpha^2 V(X)$$

gilt. Falls X und Y stochastisch unabhängig sind (s. nächster Abschnitt), so gilt

$$V(X + Y) = V(X) + V(Y).$$

Diese Formeln können z.B. angewendet werden, wenn man die Binomialverteilung $B(n, p)$ als Summe von n identischen Bernoulliverteilungen $B(p)$ interpretiert. Aber auch, wenn man den Mittelwert einer durch eine Stichprobe erhobenen Zufallsvariablen schätzt.

Es geht aber noch allgemeiner:

Definition 3.26. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ irgendeine reelle Funktion. Dann kann man aus einer Zufallsvariablen X mittels Verkettung $Y := f \circ X$ eine weitere Zufallsvariable Y gewinnen.

Ein Beispiel haben Sie schon in Kap. 2.4.4 kennengelernt, wobei f die Umrechnung von Celsius nach Fahrenheit beschreibt. Hier war $f(x) := ax + b$ eine lineare Skalentransformation. Das dortige Ergebnis

$$\bar{y} = a\bar{x} + b = f(\bar{x})$$

⁵⁹Bei E lässt man zuweilen die Funktionsklammern weg, nicht aber bei V . Warum? Keine Ahnung.

können wir jetzt als

$$EY = aEX + b = f(EX)$$

formulieren, warnen aber vor einer Übertragung auf nichtlineare Funktionen. Ist z.B. $f(x) := x^2$ und $EX = 0$, so wird i.A. $E(X^2) > 0$ sein!

3.10 Unabhängigkeit und Kovarianz von Zufallsvariablen

Wir betrachten zwei Zufallsvariablee X und Y , z.B. die Länge und das Gewicht einer Person. Wann können wir sagen, dass diese beiden Zufallsvariablen stochastisch (un)abhängig sind? Da man grob sagen kann „Je länger, desto schwerer“, sind „Länge“ und „Gewicht“ wohl nicht unabhängig.

Wenn wir jedoch zwei Personen zufällig und „unabhängig von einander“ herausgreifen und nur deren Gewicht messen, sollten dies Messergebnisse unabhängig von einander sein.

3.10.1 Zufallsstichprobe und stochastische Unabhängigkeit

Dass eine Stichprobe vom Umfang n mit einer Zufallsvariablen korrespondiert, ist schon überstrapaziert worden. Formal haben wir sogar n Zufallsvariable X_1, X_2, \dots, X_n bei einer Stichprobe vom Umfang n — für jede einzelne Beobachtung eine.

Wenn man von einer „echten“ Zufallsstichprobe spricht, muss man annehmen, dass alle $X_k, k = 1, 2, \dots, n$ *identisch verteilt*, d.h. gleiche Verteilungsfunktionen haben, und *stochastisch unabhängig* sind. — In HÜBNER wird diese Eigenschaft mit **u.i.v** (unabhängig identisch verteilt) abgekürzt .

3.10.2 Stochastische Unabhängigkeit

Wir hatten schon definiert, wann zwei Ereignisse A und B (Teilmengen von $Pot\Omega$) stochastisch unabhängig sind. Jetzt seien A' und B' beliebige reelle Mengen und $A := \{X \in A'\}, B := \{Y \in B'\}$. Wenn dann A und B unabhängig sind für beliebige A', B' , so heißen die Zufallsvariablen X und Y **stochastisch unabhängig**.

3.10.3 Kovarianz

Diesen Begriff versteht man am leichtesten, wenn man den Statistikzugang der *empirischen Kovarianz*, der wiederum mit der *Korrelation* zusammenhängt, versteht (s. Kap. 2.6.3).

Hier nur kurz: sind x_k Daten einer X -Erhebung und y_k entsprechend die zu Y , so kann man deren Mittelwerte \bar{x} und \bar{y} berechnen (diese sind die empirischen Erwartungswerte von X bzw. Y) und die

Summe

$$S_{xy} := \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$$

betrachten. Ein Summand ist positiv, wenn gleichzeitig x_k und y_k auf derselben Seite des Mittelwerts liegt. Je größer S_{xy} , so mehr sind die Daten x_k und y_k „gleichgeschaltet“.

Das nicht-empirische Gegenstück zu (der **empirischen Kovarianz**) S_{xy} ist

$$\text{Kov}(X, Y) := E((X - EX)(Y - EY)) = E(X \cdot Y) - EX \cdot EY$$

und heißt **Kovarianz** von X und Y .

Man kann zeigen, dass $\text{Kov}(X, Y) = 0$ bzw. $E(X \cdot Y) = E(X)E(Y)$, falls X und Y stochastisch unabhängig sind. In diesem Fall ist der **Korrelationskoeffizient**

$$\text{korr}(X, Y) = \frac{\text{Kov}(X, Y)}{\text{Str}(X) \cdot \text{Str}(Y)}$$

null und wegen

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Kov}(X, Y)$$

folgt $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Stochastisch unabhängige Zufallsvariable sind also *unkorreliert*.

Beispiel: X sei die Anzahl der beschädigten Bierflaschen in einem Bierkasten und Y die Anzahl der Bierflaschen mit schief aufgedrucktem Etikett. Sei $p_j := P(X = j)$ und $q_k := P(Y = k)$. Die beiden Zufallsvariable sind unabhängig, falls $P(X = j, Y = k) = p_j q_k$ gilt. In der Tat folgt hieraus

$$E(XY) = \sum_m m \cdot P(XY = m)$$

und wegen

$$m \cdot P(XY = m) = \sum_{j,k:j+k=m} m \cdot P(X = j, Y = k) = \sum_{j,k:j+k=m} m \cdot p_j q_k$$

folgt

$$E(XY) = \sum_{j,k} jk p_j q_k = \left(\sum_j j \cdot p_j \right) \cdot \left(\sum_k k \cdot q_k \right) = EX \cdot EY$$

3.11 Normalverteilung

Definition 3.27. Die Standardnormalverteilung ist durch die Wahrscheinlichkeitsdichte

$$f_{0,1}(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

ihre Verteilungsfunktion wird mit

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

bezeichnet.



Abbildung 3.12: Gauß'sche Glockenkurve auf dem 10-DM-Schein

Dass $f_{0,1}$ wirklich eine Wahrscheinlichkeitsdichte ist, liegt an

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}.$$

Dessen Graph ist die berühmte Gauß'sche Glockenkurve auf dem ehemaligen 10 DM - Schein, siehe Abb. 3.12.

Dass der Erwartungswert der Standardnormalverteilung $\mu = 0$ ist, erhält man aus der relativ leicht zu zeigenden Formel

$$\mu = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2} dx = 0.$$

Dass die Varianz der Standardnormalverteilung $\sigma = 1$ ist, kann man nicht so einfach ausrechnen.

Ersetzt man x durch $\frac{x-\mu}{\sigma}$ mit $\mu \in \mathbb{R}$ und $\sigma > 0$, so kann man aus der Standardnormalverteilung eine weitere Verteilung mit Erwartungswert μ und Streuung σ gewinnen, wenn man die transformierte Wahrscheinlichkeitsdichte

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f_{0,1}\left(\frac{x-\mu}{\sigma}\right),$$

also

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

betrachtet. Sie ist symmetrisch zu $x = \mu$ und hat bei $x \pm \sigma$ zwei Wendepunkte.

Damit ist die Verteilungsfunktion der Gauß'schen Normalverteilung mit Mittelwert μ und Standardabweichung σ definiert durch

$$F_{\mu,\sigma}(t) := \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Eine kontinuierliche Zufallsvariable, die so verteilt ist, wird mit $\mathfrak{N}(\mu, \sigma^2)$ abgekürzt.

Die Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung, findet man häufig tabelliert. Mittels

$$F_{\mu,\sigma}(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

kann sie in die Verteilungsfunktion der Gauß'schen Normalverteilung mit Mittelwert μ und Standardabweichung σ umgerechnet werden.

Beispiel: Das Geburtsgewicht von Neugeborenen nach unauffälliger Schwangerschaft kann als normalverteilt mit Erwartungswert $\mu = 3500g$ und Standardabweichung $\sigma = 500g$ angenommen werden.

Die Wahrscheinlichkeit, dass ein Neugeborenes aus dieser Grundgesamtheit nicht mehr als 4700 g wiegt, ist dann

$$F(4700) = \Phi\left(\frac{4700 - 3500}{500}\right) = \Phi(2.4) = 0.9918.$$

D.h. in der genannten Grundgesamtheit wiegen damit 99,18 % aller Neugeborenen nicht mehr als 4700 g.

Die Funktion $\Phi(x)$ sollte in allen Softwarepaketen mit mathematischen Anwendungen installiert sein.

Besonders wichtig sind die Quantile der Normalverteilung, wobei es mit der Formel

$$z = \frac{t - \mu}{\sigma}$$

möglich ist, von den Quantilen der Standardnormalverteilung auf die von $\mathfrak{N}(\mu, \sigma)$ zu schließen. Beispiel (s.o.): Es soll eine Grenze für das Geburtsgewicht angegeben werden, die nur vom 2,5 % aller Neugeborenen übertroffen wird.

Da das 97,5%-Quantil der Standardnormalverteilung bei $z_{0.975} = 1.96$ liegt, ergibt sich für $\mu = 3500$ und $\sigma = 500$ ein 97,5%-Quantil von $x_{0.975} = \mu + \sigma z_{0.975} = 3500 + 50 \cdot 1.96 = 4480$. In der genannten Grundgesamtheit wiegen also 97,5 % aller Neugeborenen nicht mehr als 4480 g, d.h. 2,5 % wiegen mindestens 4480 g.

Bei der Erhebung eines Merkmals wie „Größe eines Menschen“ wird häufig eine Normalverteilung unterstellt. Dies kann nur sinnvoll sein, wenn symmetrisch verteilte, eingipflige Merkmale vorliegen. Zum Standardisieren einer Normalverteilung benötigt man deren Erwartungswert und Varianz bzw. Standardabweichung. In der Praxis sind diese häufig nicht bekannt, und man muss sie aus einer Stichprobe durch den arithmetischen Mittelwert und die empirische Varianz bzw. Standardabweichung schätzen. Ein erster Hinweis auf Symmetrie liegt dann vor, wenn der Median und der Mittelwert annähernd gleich sind. Eine optische Überprüfung ist durch ein Histogramm mit einer angepassten Normalverteilungsdichte möglich. Die sigmoide Form der empirischen Verteilungsfunktion ist ebenfalls ein Hinweis auf annähernd normalverteilte Merkmale.

Noch geeigneter ist der sogenannte Normalverteilungsplot, wo mit Hilfe der Normalverteilung die empirische Verteilungsfunktion so transformiert wird, dass bei normalverteilten Merkmalen eine Gerade entsteht.

Häufig (warum?) sind die Logarithmen einer Zufallsvariablen normalverteilt! (Beispiel: LDH-Wert).

3.11.1 Zentraler Grenzwertsatz

Dieser besagt, dass die Summe von n unabhängigen Zufallsvariablen mit gleicher Verteilung für $n \rightarrow \infty$ normalverteilt ist. Bestes Beispiel ist eine binomialverteilte Zufallsvariable (Verteilung $B(n; p)$), die ja die Summe von n Bernoulli-verteilten Zufallsvariablen ist. Daher ist $B(n; p)$ etwa gleich $N(\mu, \sigma)$ mit $\mu := np$ und $\sigma^2 := np(1 - p)$, wenn n groß.

Will man $P(k_1 \leq X \leq k_2)$ mit binomialverteilter Zufallsvariable X näherungsweise durch eine Normalverteilung F berechnen, sollte man dies durch $F(k_2 + 0.5) - F(k_1 - 0.5)$ tun (Stetigkeitskorrektur). Der Zentrale Grenzwertsatz liefert die Aussage, dass die Mittelwerte von Stichproben-Daten für hinreichend großen Umfang n der Stichprobe normalverteilt sind. Fasst man die Daten x_k einer Stichprobe als Werte von identischen Zufallsvariablen X auf, so ist die Schätzung des Mittelwertes von X durch den Mittelwert der Daten wieder eine Zufallsvariable. Deren Mittelwert ist gleich dem von X (erwartungstreuer Schätzer), deren Varianz σ^2/n , wenn σ^2 die Varianz von X ist.

Die Statistik liefert Methoden, die Wahrscheinlichkeit zu schätzen, dass der Mittelwert einer Stichprobe in einem gewissen *Konfidenzintervall* liegt.

Beispiel: Roulette, 18 rote, 18 schwarze, 2 grüne Felder. Der Spieler setzt stets 1 Euro auf rot. Sein Gewinn ist 1 Euro, wenn rot kommt, und ein Euro Verlust sonst. Die Auszahlung im n -ten Spiel ist eine Zufallsvariable X_n mit den Werten $X_n = 1$ mit $P(X_n = 1) = 9/19$ und $X_n = -1$ mit $P(X_n = -1) = 10/19$. Alle X_n sind identisch (Bernoulli-) verteilt, Erwartungswert μ und Varianz σ^2 sind einfach zu berechnen ($EX_n = \mu = -1/19 = -0.05263$, $\sigma^2 = 9/19 \cdot 10/19 = 0.2493$). Nun sei Y_n die Zufallsvariable, die die Auszahlung nach n Spielen liefert, also $Y_n = X_1 + \dots + X_n$. Y_n ist für große n normalverteilt mit der Varianz $n\sigma^2 = n \cdot 0.2493$ und Erwartungswert $-n \cdot 0.05263$, so der Zentrale Grenzwertsatz.

Nach $n = 100$ Spielen erwarten wir einen Verlust von 5,26 Euro ($\mu = -5,26$) mit einer Varianz von 24,93 bzw. einer Streuung von $\sigma = 4,993$ Euro. Will man wissen, mit welcher Wahrscheinlichkeit man mindestens 5 Euro gewinnt, kann man die Normalverteilung heranziehen. Die standardisierte Variable ist $t = (5 - \mu)/\sigma$, also $t = 2.05$. Wegen $\Phi(2.05) = 0.9798$ beträgt diese Gewinn-Wahrscheinlichkeit nur 2,02%. Mit der Binomialverteilung hätte man länger gerechnet.

Ein schönes Beispiel für die Anwendung des Zentralen Grenzwertsatzes (ZGS) findet man in [Normalverteilungs-Applet \(Mathe Online\)](http://www.mathe-online.at/galerie/wstat1/wstat1.html)⁶⁰. Man muss hier zwei Zahlen festlegen: n und N . N ist hier die Anzahl der Zufallsexperimente, n die Anzahl der Summanden bei dem ZGS. Ein einzelnen Summand ist eine Rechteckverteilung auf $[0, 1]$, deren Zufallsvariable jede Zahl aus $[0, 1]$ mit gleicher Wahrscheinlichkeit ergibt. Nennen wir diese Zufallsvariable X , so wird $Y = X_1 + X_2 + \dots + X_n$, wobei die X_j alle wie X verteilt sind (und stochastisch unabhängig sind). Ein Zufallsexperiment liefert einen Wert für Y , d.h. enthält n „interne“ Zufallsvariablen. Nun wird die empirische Verteilung von Y gezeigt, wenn N solcher Zufallsexperimente durchgeführt werden. Je größer n , desto eher entspricht die Dichte von Y der einer Normalverteilung. Da nicht X , sondern eine diskrete Version durch Unterteilung von $[0, 1]$ in 100 Teilintervalle verwendet

⁶⁰<http://www.mathe-online.at/galerie/wstat1/wstat1.html>

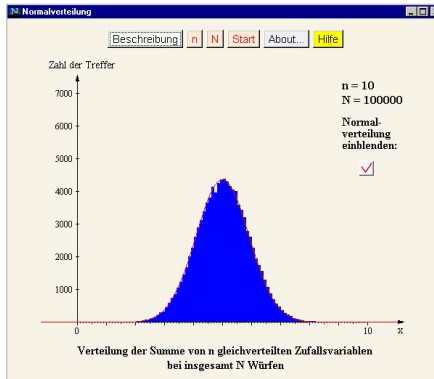


Abbildung 3.13: Applet zum Zentralen Grenzwertsatz

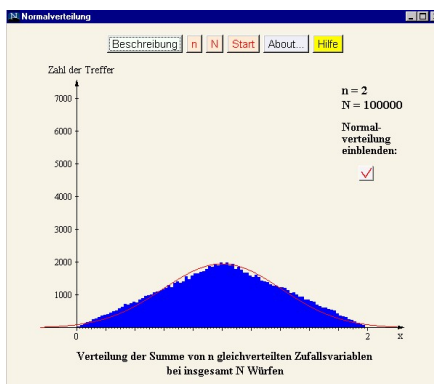


Abbildung 3.14: Applet zum Zentralen Grenzwertsatz

wird, sieht man auch nur eine (empirische) Zähldichte, die aber kaum von einer kontinuierlichen Dichte zu unterscheiden ist. In Abb. 3.13 ist $n = 10$, die Normalverteilung ist gut zu erkennen. In Abb. 3.14 ist $n = 2$, hier handelt es sich um eine Dichte in Dreiecksform, also nicht um eine Normalverteilung.