

EINFÜHRUNG IN DIE WAHRSCHEINLICHKEITSRECHNUNG UND STATISTIK

GUNTHER H. PEICHL

Skriptum zur Vorlesung im SS 1999

INSTITUT FÜR MATHEMATIK
KARL-FRANZENS-UNIVERSITÄT GRAZ

VORWORT

Im Alltag begegnen wir immer wieder Fragestellungen aus der Wahrscheinlichkeitstheorie und Statistik, sei es nun beim wöchentlichen Lottofieber, bei der Beurteilung der Erfolgsaussichten bei einer Prüfung, beim Abschätzen des Marktanteiles eines neuen Produktes, bei der Interpretation von Meinungsumfragen usw. Dieser Omnipräsenz von Zufallseffekten steht allerdings ein weitverbreitetes Unwissen und Unverständnis wahrscheinlichkeitstheoretischer Zusammenhänge gegenüber.

Eine dreistündige Vorlesung "Einführung in die Wahrscheinlichkeitsrechnung und Statistik" stellt notwendigerweise einen mehrfachen Kompromiß dar. Einerseits würde sowohl die Wahrscheinlichkeitsrechnung, als auch die Statistik jeweils eine mehrstündige Einführung rechtfertigen. Ich habe mich daher entschieden, die Wahrscheinlichkeitsrechnung nur soweit aufzubereiten, als dies für das Verständnis der fundamentalen Techniken in der Statistik notwendig ist. Selbst dies erforderte noch Abstriche: aus Zeitgründen konnte ich beispielsweise keine eingehendere Analyse des Begriffs der stochastischen Unabhängigkeit einbauen, es fehlen u.a. auch die statistische Behandlung qualitativer Daten und Methoden der nichtparametrischen Statistik. Andererseits mußte auch ein Kompromiß in der mathematischen Strenge der Darstellung gefunden werden. Um die Einfachheit der grundlegenden Konzepte nicht zu verschleiern, habe ich mich entschlossen, gänzlich auf die maßtheoretische Fundierung der Wahrscheinlichkeitsrechnung zu verzichten. Es war mir auch ein besonderes Anliegen, die Anwendbarkeit und den praktischen Einsatz der vorgestellten Konzepte an konkreten Beispielen zu demonstrieren. Um in 3 Stunden doch einen einigermaßen repräsentativen Einblick in die Welt der Wahrscheinlichkeitsrechnung und Statistik vermitteln zu können, und den Teilnehmern der Lehrveranstaltung die Möglichkeit zu geben, sich auf die wesentlichen Inhalte zu konzentrieren, habe ich mich entschieden, das vorliegende Skriptum auszuarbeiten. Da das Skriptum parallel zur Vorlesung entwickelt wurde, ist es noch mit allen Kinderkrankheiten eines Prototyps behaftet: gelegentliche Inkonsistenzen in der Bezeichnung, suboptimale Reihung des Stoffes, Irrtümer und Tippfehler sind nicht zu vermeiden, beeinträchtigen aber hoffentlich nicht in zu großem Ausmaß dessen Lesbarkeit. Obwohl diese Unterlage weitgehend in sich abgeschlossen ist, kann es nicht das begleitende Studium der Lehrbuchliteratur ersetzen. Besonders empfehlen möchte ich zwei Lehrbücher auf mittlerem Niveau, auf welchen dieses Skriptum aufbaut:

- D. Wackerly, W. Mendenhall und R. Scheaffer, *Mathematical Statistics with Applications*, Duxbury Press, 1996
- J.A. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press, 1995

Contents

Kapitel 1. Beschreibende Statistik	1
1. Darstellung von Daten	1
2. Klassenbildung	6
3. Zeitreihen	8
4. Statistische Kenngrößen	9
Kapitel 2. Elementare Wahrscheinlichkeit	15
1. Grundbegriffe der Wahrscheinlichkeitsrechnung	15
2. Diskrete Wahrscheinlichkeitsmaße	20
3. Stetige Wahrscheinlichkeitsmaße	22
4. Verteilungsfunktionen	26
5. Laplace'sche Zufallsexperimente	28
6. Bedingte Wahrscheinlichkeit, Stochastische Unabhängigkeit	36
7. Diskrete Zufallsvariable	43
8. Spezielle diskrete Verteilungen	47
9. Stetige Zufallsvariable	57
10. Ungleichung von Tschebyscheff	66
11. Mehrdimensionale Zufallsvariable	67
12. Unabhängige Zufallsvariable	71
13. Bedingte Verteilungen	72
14. Erwartungswert einer Funktion von mehrdimensionalen Zufallsvariablen	75
15. Bedingte Erwartungswerte	76
16. Kovarianz und Korrelation	79
17. Funktionen von mehrdimensionalen Zufallsvariablen	81
18. Ordnungsstatistik	85
19. Momenterzeugende Funktion	86
20. Testverteilungen	88
21. Grenzwertsätze	90
Kapitel 3. Schließende Statistik	97
1. Punktschätzverfahren	97
2. Konfidenzintervalle	108
3. Konfidenzintervalle für μ und $\mu_1 - \mu_2$	112
4. Konfidenzintervalle für σ^2	116

5.	Momentenmethode	117
6.	Maximum Likelihood Methode	119
7.	Asymptotische Eigenschaften eines MLS	121
8.	Konsistenz von Schätzfunktionen	123
9.	Relative Effizienz	126
10.	Suffizienz	128
11.	Testen von Hypothesen	134
12.	Der Z -Test für große Stichproben	136
13.	Der p -Wert eines Testes	142
14.	Test von Hypothesen über μ bzw. $\mu_1 - \mu_2$ bei kleinen Stichproben	143
15.	Testen von Hypothesen über Varianzen	145
16.	Dualität zwischen Konfidenzintervall und Hypothesentest	148
17.	Die Macht eines Testes und das Neyman-Pearson Lemma	150
18.	Der verallgemeinerte Likelihood Quotiententest	153

KAPITEL 1

Beschreibende Statistik

1. Darstellung von Daten

Entscheidungen werden in immer stärkeren Ausmaß durch statistische Methoden abgesichert: z.B. Meinungsforschung, Erhebung von Marktanteilen, Reichweiten etc. Am sichersten wäre es natürlich, jedes Mitglied der betreffenden Zielgruppe zu befragen. Für die meisten Anlässe ist diese Vorgangsweise viel zu aufwendig. Es wird daher nur eine repräsentative **Stichprobe** untersucht. In diesem Abschnitt sollen verschiedene Möglichkeiten vorgestellt werden, umfangreiche Datensätze so darzustellen, daß deren wesentlichen Eigenschaften leicht erkennbar sind.

BEISPIEL 1.1. Aus der Ergebnisliste einer Übungsgruppe Analysis II entnehmen wir folgende Daten:

Diese Daten bilden eine Stichprobe vom **Umfang** $n = 12$ aus der **Grundgesamtheit** aller Hörer der Analysis II des betreffenden Jahrganges. Allgemein setzt sich eine Grundgesamtheit aus **Versuchseinheiten**, hier die Studierenden, zusammen. Die Versuchseinheiten sind Träger von **Merkmalen**: das Geschlecht, die Semesterzahl und die Beurteilung. Jedes dieser Merkmale kommt in verschiedenen **Ausprägungen** vor: für das Geschlecht männlich und weiblich, die Semesterzahl

Nr.	Geschlecht	Semesterzahl	Note
1	w	4	NG
2	w	6	GN
3	m	4	GN
4	m	2	SG
5	m	4	GN
6	m	2	BF
7	w	6	GT
8	m	6	BF
9	w	8	NG
10	w	2	NG
11	m	2	NG
12	w	6	NG

TABELLE 1.1. Ergebnisliste

kann theoretisch jede natürliche Zahl sein, die Beurteilung benutzt eine 5–stufige Skala. Das Beispiel zeigt die drei wesentlichsten Merkmalstypen:

- qualitative (nominale) Merkmale
- ordinale Merkmale
- quantitative (metrische) Merkmale

Qualitative Merkmale sind beispielsweise Geschlecht, Staatsangehörigkeit, Blutgruppe, Beruf, Familienstand, etc. Manchmal verwendet man einen Zahlenschlüssel, etwa 0 = weiblich, 1 = männlich. Da die Wahl des Schlüssels vollkommen beliebig ist, ist eine weitere numerische Verarbeitung, etwa die Berechnung eines Mittelwertes, dieser Daten nicht sinnvoll.

Ordinale Merkmale haben als Ausprägungen Ränge, die sich anordnen lassen, beispielsweise Noten, Interesse für Mathematik, Engagement für die Umwelt, Platzierung beim Wettkampf, etc. Für manche ordinale Merkmale gibt es standardisierte Zahlenskalen, etwa 1 – 5 für die Benotung, jedoch kann man diese ohne Informationsverlust stets einer streng monotonen Transformation unterwerfen. Eine Notenskala 10, 20, . . . ,50 leistet dieselben Dienste. Rechnungen mit ordinalen Daten werden gelegentlich durchgeführt (Notendurchschnitt), sie können aber durch sinnvollere Kenngrößen ersetzt werden.

Quantitative Merkmale können (zumindest prinzipiell) gemessen werden: Semesterzahl eines Studierenden, Entfernung Wohnung – Arbeitsplatz, Frequenz der öffentlichen Verkehrsmittel, Knochendichte, etc. Die Ausprägungen werden durch geeignete Teilmengen der reellen Zahlen beschrieben. Quantitative Daten können daher numerisch bearbeitet werden. Der numerische Wert einer Ausprägung ist eindeutig bis auf die Wahl einer Einheit und birgt mehr Information als der eines ordinalen Merkmals: eine 2 Literflasche enthält die doppelte Menge einer 1 Literflasche, ein Schüler mit der Note 2 ist aber nicht unbedingt doppelt so tüchtig wie ein Schüler mit der Note 1.

Die Ergebnisliste 1.1 ist die **Urliste** unserer kleinen Erhebung. Unser eigentliches Interesse gilt aber nicht den Eigenschaften jedes einzelnen Individuums, sondern vielmehr der Verteilung der Merkmale in der Grundgesamtheit. Da Informationen nur über die Stichprobe zur Verfügung stehen, betrachten wir die Verteilung der Merkmale in der Stichprobe und stellen zuerst fest, wie oft die einzelnen Ausprägungen eines jeden Merkmals in der Stichprobe auftreten. Dies ergibt die **absolute Häufigkeit** $h_a(x)$ einer Merkmalsausprägung x . Bezieht man $h_a(x)$ auf den Stichprobenumfang n , erhält man die **relative Häufigkeit** $h_r(x)$ der Ausprägung x :

$$h_r(x) = \frac{h_a(x)}{n}.$$

Unmittelbar einsichtig ist die Ungleichung:

$$0 \leq h_r(x) \leq 1.$$

Sem	absolute relative	
	Häufigkeit	
2		4 33.3 %
4		3 25 %
6		4 33.3%
8		1 8.3%

TABELLE 1.2. Verteilung des Merkmals Semesterzahl

Bei der Bestimmung der absoluten bzw. relativen Häufigkeiten ist es manchmal praktisch, eine Strichliste für die einzelnen Ausprägungen eines Merkmals anzufertigen. Für die graphische Darstellung von Daten verwendet man häufig **Kreisdiagramme** und **Histogramme**. Ein Kreisdiagramm für ein Merkmal mit k Ausprägungen x_i in der Stichprobe entsteht, indem man einen Kreis in k Sektoren unterteilt, deren Flächeninhalt proportional sind zur absoluten bzw. relativen Häufigkeit von x_i , $i = 1, \dots, k$. Ein Kreisdiagramm eignet sich besonders für qualitative Merkmale. Es verdeutlicht besonders anschaulich die Größenverhältnisse der Anteile der einzelnen Datengruppen. Es eignet sich nicht zur Darstellung eines Merkmals mit vielen verschiedenen Ausprägungen, da die Sektoren zu schmal werden. Die Verteilung des Merkmals in der Stichprobe kann man besonders gut in einem Histogramm erkennen. Ein Histogramm erstellt man, indem man in einem kartesischen Koordinatensystem auf der Abszisse die Merkmalsausprägungen x_i aufträgt und symmetrisch über x_i einen Balken errichtet, dessen Flächeninhalt proportional zu $h_a(x_i)$ bzw. $h_r(x_i)$, $i = 1, \dots, k$, ist. Verwendet man in der Darstellung relative Häufigkeiten, sollte der Stichprobenumfang in die Abbildung aufgenommen werden, um die Aussagekraft der Darstellung zu erhalten. Man überzeuge sich von dem unterschiedlichen optischen Eindruck und der jeweiligen suggestiven Wirkung eines Histogramms mit niedrigen und breiten, bzw. mit hohen und schmalen Säulen.

Die Anzahl oder den Anteil der Studierenden, welche sich höchstens im x -ten Semester befinden, kann man bequem aus der kumulativen absoluten, bzw. kumulativen relativen Häufigkeit ablesen. Allgemein kann man ordinale und quantitative Daten kumulieren: man definiert

$$(1.1) \quad F(x) = \sum_{y \leq x} h_r(y) \quad \text{kumulative relative Häufigkeit}$$

$$(1.2) \quad G(x) = \sum_{y \leq x} h_a(y) \quad \text{kumulative absolute Häufigkeit.}$$

(Wir betrachten vorerst nur Merkmale mit endlich vielen Ausprägungen, die vorstehenden Summen sind also wohldefiniert).

Tabelle 1.3 illustriert die kumulierten Häufigkeiten aus Tabelle 1.2: Handelt es sich um ein metrisches Merkmal, kann man die kumulative relative Häufigkeit auffassen als

Sem	x	$h_a(x)$	$h_r(x)$	$G(x)$	$F(x)$
2	4	$\frac{1}{3}$	4	$\frac{1}{3}$	
4	3	$\frac{1}{4}$	7	$\frac{7}{12}$	
6	4	$\frac{1}{3}$	11	$\frac{11}{12}$	
8	1	$\frac{1}{12}$	12	1	

TABELLE 1.3. Verteilung des Merkmals Semesterzahl

Abbildung $F: \mathbb{R} \rightarrow [0, 1]$, vgl. Abbildung 1.3. Offensichtlich ist F rechtsseitig stetig, Sprünge treten bei jenen Ausprägungen auf, welche in der Stichprobe tatsächlich vorkommen.

Quantitative Datensätze, welche nicht zu umfangreich sind, kann man als **Stengel-Blatt Diagramm** darstellen. Wir demonstrieren das Vorgehen an einem Beispiel:

BEISPIEL 1.2. In einer amerikanischen Stadt wurde bei allen Eheschließungen im Laufe einer Woche das Alter der Braut festgehalten:

Der Stamm der Stengel-Blatt Darstellung ist ein senkrechter Strich. Von diesem Stamm zweigen links und rechts Äste ab: die linken Äste sind die Zehnerstellen des Alters, die rechten Äste die Einerstellen, welche entsprechend der Häufigkeit des betreffenden Alters notiert werden. Um die Darstellung übersichtlicher zu gestalten, wird jede Dekade in 2 Abschnitte zerlegt: 2 auf der rechten Seite des Stammes steht beispielsweise für das Alter 20 – 24, 2+ für 25 – 29. In Tabelle 1.5 wurden die Daten

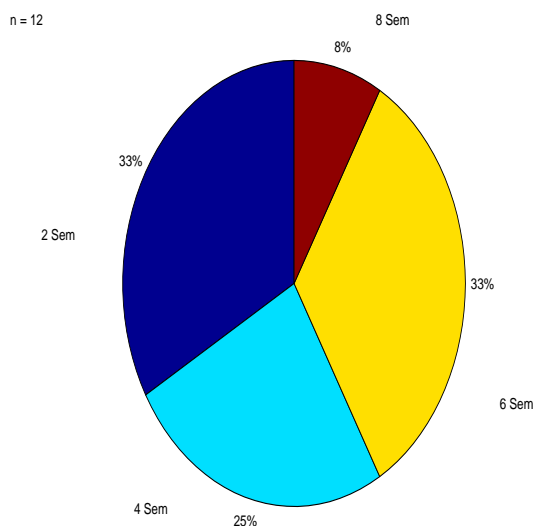


ABB. 1.1. Kreisdiagramm

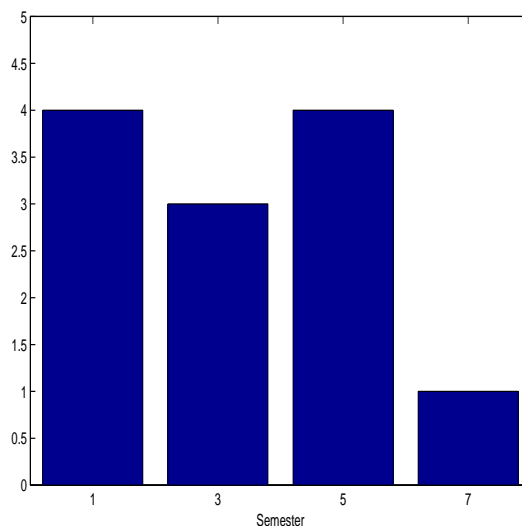


ABB. 1.2. Histogramm

30	27	56	40	30	26	31	24	23	
25	29	33	29	22	33	29	46	25	
34	19	23	23	44	29	30	25	23	
60	25	27	37	24	22	27	31	24	26

TABELLE 1.4. Beispiel 1.2

außerdem noch der Größe nach geordnet. Auch ein Stengel–Blatt Diagramm ver-

6		0
5+		6
5		
4+		6
4		04
3+		7
3		0 0 0 1 1 3 3 4
2+		5 5 5 5 6 6 7 7 7 9 9 9 9
2		2 2 3 3 3 3 4 4 4
1+		9

TABELLE 1.5. Stengel–Blatt Diagramm

anschaulicht bereits recht gut und ohne großen zusätzlichen Aufwand die Verteilung

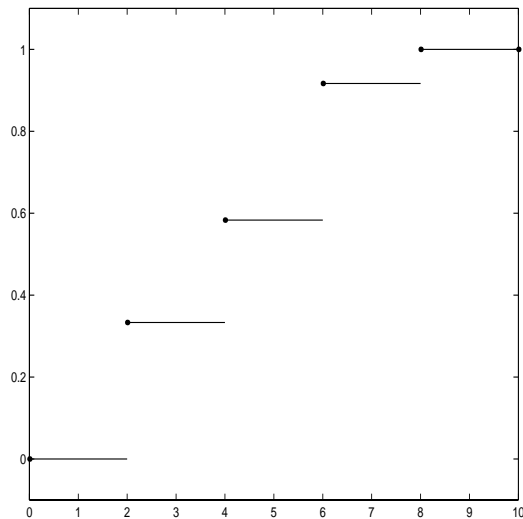


ABB. 1.3. Kumulative relative Häufigkeit

eines Merkmals in der Stichprobe. Außerdem bleiben die Originaldaten erhalten. Es ist daher für umfangreiche Datensätze nicht geeignet.

Das Histogramm der Rohdaten, vgl. Abbildung 1.4, ist unbefriedigend: es enthält offensichtlich zu viele Details, sodaß die zufälligen Schwankungen des Brautalters voll zum Tragen kommen. Dieser Effekt kann durch Gruppierung der Daten gemildert werden.

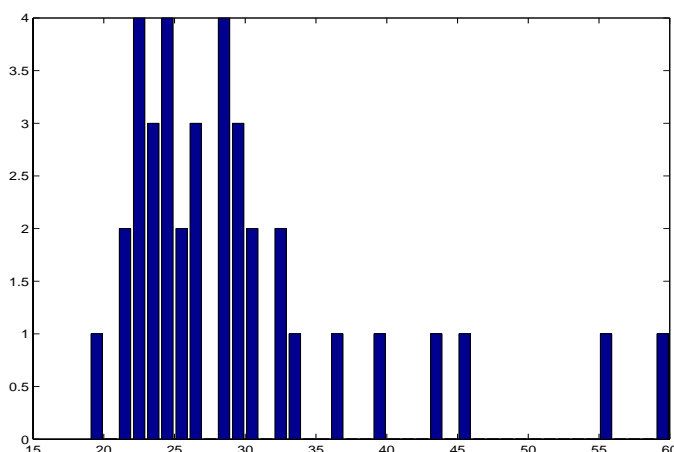


ABB. 1.4. Histogramm Brautalter

2. Klassenbildung

Treten in einer Stichprobe sehr viele verschiedene Ausprägungen eines Merkmals auf, ist es zweckmäßig, die Stichprobe zu vereinfachen, indem man verschiedene Ausprägungen jeweils in einer **Klasse** zusammenfaßt. Die Klassenbildung sollte so erfolgen, daß nur unwesentliche Einzelheiten ausgeschieden werden. Allgemein sollte man nicht mehr als 20 Klassen verwenden. Durch Klassenbildung geht allerdings Information verloren: die einzelnen Stichprobenwerte einer Klasse treten nicht mehr auf. Bei der weiteren Verarbeitung der Stichprobe nimmt man daher bei einem quantitativen Merkmal häufig an, daß alle Werte einer Klasse in der jeweiligen Klassenmitte konzentriert sind. Die absolute Häufigkeit einer Klasse ist die Summe der absoluten Häufigkeiten aller Ausprägungen, welche zu dieser Klasse gehören. Dividiert man die absoluten Klassenhäufigkeiten durch den Stichprobenumfang, ergeben sich die relativen Klassenhäufigkeiten.

BEISPIEL 1.3. Gruppiert man die Daten aus Beispiel 1.2 in die Klassen 15 – 19, 20 – 24, 25 – 29, ..., ergeben sich die Häufigkeitsverteilungen aus Tabelle 1.6 und das vereinfachte Histogramm 1.5.

Man beachte die große Ähnlichkeit des Histogramms 1.5 mit dem Stengel-Blatt Diagramm in Tabelle 1.5. Dies liegt daran, daß die Wahl der Dezimalstellen auf der

	Klassenhäufigkeit			Klassenhäufigkeit	
	absolut	relativ		absolut	relativ
15 – 19	1	0.03	40 – 44	2	0.05
20 – 24	9	0.24	45 – 49	1	0.03
25 – 29	13	0.35	50 – 54	0	0.00
30 – 34	8	0.22	55 – 59	1	0.03
35 – 39	1	0.03	60 – 64	1	0.03

TABELLE 1.6

linken Seite des Stammes eine Klasseneinteilung der Daten induziert. Diese ist im Beispiel mit der Klasseneinteilung in Tabelle 1.6 identisch.

Haben alle Klassen dieselbe Breite, kann man die Höhe der Stäbe im Histogramm proportional zur Klassenhäufigkeit wählen. Bei unterschiedlichen Klassenbreiten ist die Klassenhäufigkeit durch die jeweilige Klassenbreite zu dividieren.

Das folgende Beispiel zeigt, daß Klassenbildung auch für manipulative Zwecke mißbraucht werden kann:

BEISPIEL 1.4. Eine Firma habe folgende Gehaltsstruktur in Tabelle 1.7. Die Abbildung 1.6 zeigt das Histogramm der tatsächlichen Gehaltsstruktur. In Abbildung 1.7 wurden die drei niedersten und die drei höchsten Gehaltsstufen zu jeweils einer Klasse zusammengefaßt. Die Klassenbildung überdeckt den Umstand, daß die

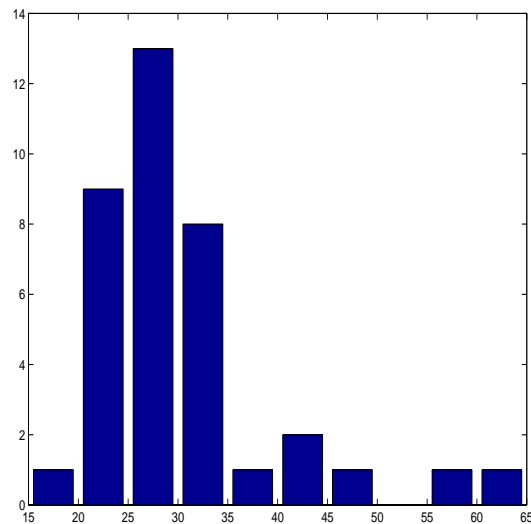


ABB. 1.5. Klassenhistogramm

Gehaltsklasse	Anzahl	Gehaltsklasse	Anzahl
5000 – 10000	3	20001 – 25000	1
10001 – 15000	10	25001 – 30000	2
15001 – 20000	5	30000 – 35000	4

TABELLE 1.7. Lohnniveau

Mehrheit der Arbeitnehmer in den beiden untersten Gehaltsklassen angesiedelt ist und die mittlere Ebene sehr dünn besetzt ist.

3. Zeitreihen

Oft werden Daten über eine gewisse Zeitspanne in regelmäßigen oder unregelmäßigen Abständen erhoben. Einen derartigen Datenbestand nennt man **Zeitreihe**. Natürlich kann man eine Zeitreihe auch durch ein Histogramm darstellen. Die dynamische Entwicklung des Merkmals kommt allerdings deutlicher zum Ausdruck, wenn man das Merkmal als Funktion von der Zeit präsentiert.

BEISPIEL 1.5. Als Beispiel betrachten wir die Preisentwicklung eines bestimmten Produktes über einen Zeitraum von 10 Jahren. Der Preis wurde jeweils zu Beginn des Jahres festgestellt. Die Abbildungen 1.8 und 1.9 zeigen unterschiedliche, korrekte Darstellungen der Preisentwicklung. Welche Darstellung würde der Handel, welche der Verein für Konsumentenschutz bei Verhandlungen über Preiserhöhungen als Argumentationshilfe verwenden?

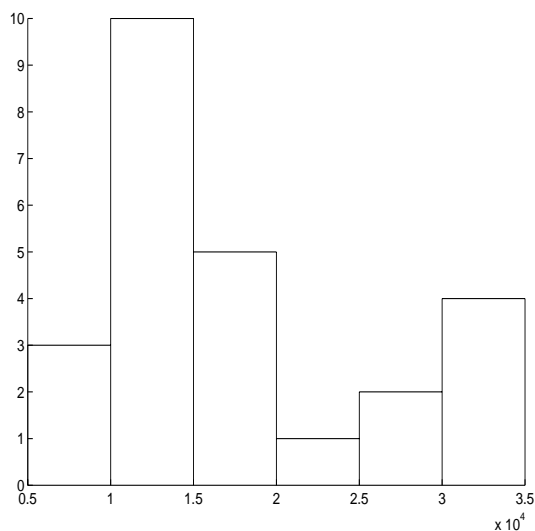


ABB. 1.6. passende Klassenzahl

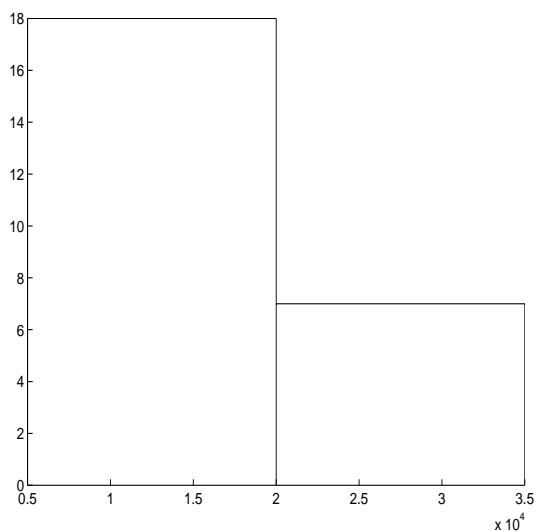


ABB. 1.7. zuwenig Klassen

Jahr	Preis	Jahr	Preis
1985	72.4	1990	76.8
1986	73.8	1991	77.1
1987	75.0	1992	77.3
1988	76.0	1993	77.9
1989	76.2	1994	80.2

TABELLE 1.8. Zeitreihe

4. Statistische Kenngrößen

In diesem Abschnitt besprechen wir mehrere Möglichkeiten, viele Daten durch eine einzige Kenngröße zu beschreiben und die Unterschiedlichkeit der Daten untereinander zu erfassen. Natürlich ist dies mit einem großen Informationsverlust verbunden. Wir bedienen uns dazu der **Lage-** und der **Streuparameter**. Ein Lageparameter gibt an, wo sich die Ausprägungen eines Merkmales häufen. Ein Streuparameter beschreibt, wie stark die Ausprägungen variieren. Zu den Lageparametern zählen

- Modal
- Median
- Perzentile und Quartil
- Mittelwert

Als Beispiele für Streuparameter betrachten wir

- Spannweite, Interquartil-Spannweite

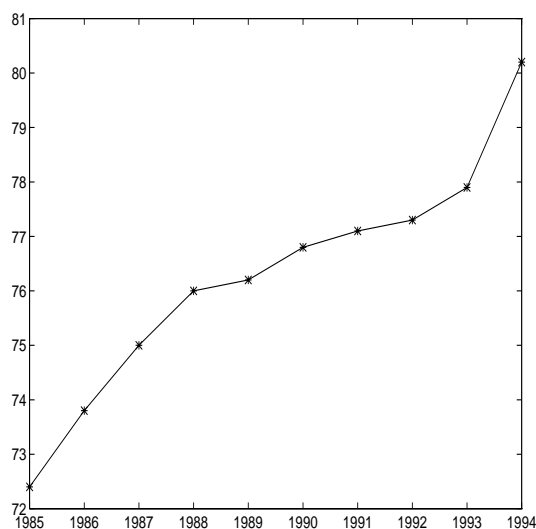


ABB. 1.8. ohne Nullniveau

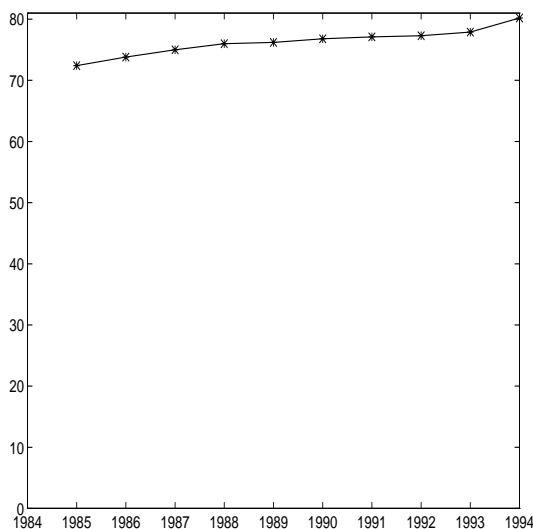


ABB. 1.9. mit Nullniveau

- Varianz, Streuung

4.1. Modal. Der **Modal** eines Merkmals ist die Ausprägung mit der größten absoluten Häufigkeit. Wurde bereits eine Klasseneinteilung vorgenommen, ist der Modal (Modalklasse) die Klasse mit der größten absoluten Klassenhäufigkeit. Bei einem quantitativen Merkmal ersetzt man die Modalklasse häufig durch ihren Mittelpunkt. Ein Merkmal kann durchaus mehrere Modale haben: das Merkmal Semester in Beispiel 1.1 hat die Modale 2 und 6, die Altersverteilung in Beispiel 1.2 hat die Modale 23, 25 und 29, nach der Klasseneinteilung ergibt sich der Modalwert 27.5.

Da in den Modal nur die absolute Häufigkeit einer Ausprägung einfließt, reagiert er sehr empfindlich auf Ausreißer in den Daten. Er kann als einziger der betrachteten Lageparameter auch auf qualitative Merkmale angewendet werden.

4.2. Median, Perzentile und Quartil. Was ist das mittlere Alter der Braut in Beispiel 1.2. Eine Antwort darauf gibt der Mittelwert, den wir im folgenden Abschnitt besprechen. Eine andere Antwort geht von der Vorstellung aus, daß mit mittlerem Alter jenes gemeint ist, für welches die gleiche Anzahl von Frauen jünger bzw. älter ist, also jenes Alter, welches genau in der Mitte der Daten liegt. Zur Bestimmung dieses Alters ordnet man die Daten der Größe nach. Dies ist besonders einfach, wenn die Daten bereits in einem Stengel-Blatt Diagramm vorliegen: es genügt die rechten Blätter zu sortieren. Die Daten in Tabelle 1.5 sind bereits der Größe nach geordnet. Der 19. Eintrag (vom kleinsten aus gezählt), also 27 Jahre, ist daher das gesuchte Alter, der sogenannte Median.

Der **Median**, m , eines ordinalen Merkmals ist jene Ausprägung, welche die Daten in zwei gleich große Hälften teilt. Bei einer ungeraden Anzahl von Daten ist diese Ausprägung eindeutig bestimmt. Wie geht man bei einer geraden Anzahl vor? Betrachten wir folgende Notenverteilung:

1 2 3 3 3 4 4 5 5 5

Die beiden mittleren Noten sind 3 und 4: für jede Note dazwischen, ist die eine Hälfte der Schüler besser, die andere Hälfte schlechter bewertet worden. Bei einer geraden Anzahl von Daten hat sich die Konvention durchgesetzt, als Median den Mittelwert der beiden mittleren Daten zu nehmen, hier also 3.5. Für den Median m gilt also

$$m = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & n \text{ gerade} \end{cases}$$

Welches Alter wird von mindestens 90% der Frauen in der Stichprobe 1.2 nicht überschritten? Der 34. Eintrag ($37 \cdot 0.9 = 33.3$) im (geordneten) Stengel-Blatt Diagramm ist das Alter 44 Jahre. Dieses Alter gibt die 90%-Perzentile an.

Die **p%-Perzentile** eines ordinalen oder metrischen Merkmals ($p \in [0, 100]$) ist jene Ausprägung, für welche mindestens $\frac{pn}{100}$ Einheiten kleinere und $\frac{n(1-p)}{100}$ Einheiten

größere Ausprägungen annehmen. Der Median ist also die 50%-Perzentile, die 25%-Perzentile nennt man **1. Quartil**, Q_I , die 75%-Perzentile heißt **3. Quartil**, Q_{III} . (Der 2. Quartil ist der Median).

Perzentilen können auch bequem aus der kumulativen relativen Häufigkeit abgelesen werden. Nach Abbildung 1.3 liegt die 20%-Perzentile beispielsweise bei 2 Semester (sowie jede weitere $p\%$ -Perzentile mit $0 < p < \frac{100}{3}$). Die $\frac{100}{3}\%$ -Perzentile ist dadurch charakterisiert, daß mindestens 4 Studenten in einem niedrigeren (oder gleichen) und mindestens 8 Studenten in einem höheren (oder gleichen) Semester inskribiert sind. Läßt man reelle Semesterzahlen zu, ist dies für jede Zahl in $[2, 4]$ der Fall. Üblicherweise nimmt man den Mittelwert, also 3 Semester.

Wie weit streut das Alter der Frauen in Beispiel 1.2? Nun, die jüngste Braut war 19, die älteste 60 Jahre alt, ihr Altersunterschied also 41 Jahre. Man nennt die Differenz zwischen der größten und kleinsten Ausprägung eines quantitativen Merkmales **Spannweite**. Diese ist zwar leicht zu bestimmen, wird aber möglicherweise durch Ausreißer in den Daten bestimmt: Werden die beiden ältesten Frauen nicht berücksichtigt, beträgt die neue Spannweite nur mehr 27 Jahre. Die Spannweite gibt ferner keine Information, wie sich die Daten im Intervall $[x_{min}, x_{max}]$ verteilen. Betrachten wir die beiden Quartile $Q_I = 24$ und $Q_{III} = 32$. Aus deren Definition folgt, daß jeweils 25% der Frauen jünger als 24 Jahre und älter als 32 Jahre sind. Mit anderen Worten: 50 % der Frauen sind zwischen 24 und 32 Jahre alt. Der Abstand $Q_{III} - Q_I$ heißt **Interquartil-Spannweite (Quartilabstand)**. Spannweite und Quartilabstand werden im Box-Whisker Diagramm dargestellt, aus dem die Streuung der Daten besonders anschaulich hervorgeht.

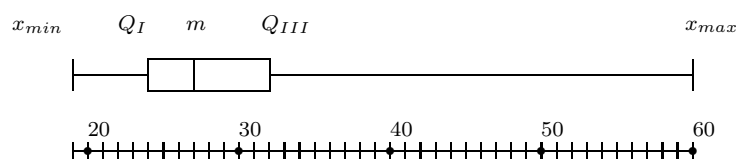


ABB. 1.10. Box-Whisker Diagramm

Ein Box-Whisker Diagramm besteht aus einer horizontalen (vertikalen) Skala, welche die erreichten Werte der Ausprägungen umfaßt, einem Rechteck, welches sich vom 1. zum 3. Quartil erstreckt, einem Teilstrich am Median und T-förmigen Fortsätzen zur minimalen und maximalen Ausprägung. Das Rechteck markiert also den Bereich, in dem 50 % aller Daten liegen.

Aus Abbildung 1.10 erkennt man auch ohne Kenntnis der Originaldaten, daß x_{max} einen Ausreißer darstellt. Ein Box-Whisker Diagramm eignet sich besonders gut zu Vergleich verschiedener Gruppen. Die Originaldaten gehen allerdings verloren.

4.3. Mittelwert und Streuung. Der **Mittelwert** wohl die bekannteste Möglichkeit, einen umfangreichen Datensatz durch eine einzige Zahl zu beschreiben. Er sollte nur für ein quantitatives Merkmal verwendet werden. Wir erinnern an die Definition

$$(1.3) \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Es bezeichnet n den Umfang der Stichprobe, x_i , $i = 1, \dots, n$, die Ausprägungen des Merkmals in der Stichprobe.

Die äquivalente Umformung von (1.3),

$$0 = n\bar{x} - \sum_{i=1}^n x_i = \sum_{i=1}^n (\bar{x} - x_i)$$

beweist den ersten Teil der folgenden Behauptung:

THEOREM 1.1. *Der Mittelwert von n reellen Zahlen wird durch jede der folgenden Behauptungen charakterisiert:*

1. *Der Mittelwert \bar{x} ist jene Zahl, für welche die Summe der Abweichungen von den Daten verschwindet:*

$$\sum_{i=1}^n (\bar{x} - x_i) = 0.$$

2. *Der Mittelwert \bar{x} ist jene Zahl, für welche die Summe der quadratischen Abweichungen von den Daten minimal ist:*

$$\sum_{i=1}^n (\bar{x} - x_i)^2 \leq \sum_{i=1}^n (x - x_i)^2, \quad x \in \mathbb{R}.$$

BEWEIS. Um die 2. Behauptung einzusehen, betrachtet man die Funktion

$$f(x) = \sum_{i=1}^n (x - x_i)^2 = nx^2 - 2x \sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2.$$

Eine einfache Rechnung zeigt nun, daß f das globale Minimum genau im Mittelwert annimmt. \square

Setzt man $x_{min} = \min\{x_1, \dots, x_n\}$ und $x_{max} = \max\{x_1, \dots, x_n\}$, folgt aus der Abschätzung

$$nx_{min} \leq \sum_{i=1}^n x_i \leq nx_{max},$$

daß der Mittelwert stets im Intervall $[x_{min}, x_{max}]$ liegt:

$$x_{min} \leq \bar{x} \leq x_{max}.$$

In den Mittelwert fließt Information von jedem einzelnen Datum ein: die Veränderung auch nur eines einzigen Datums schlägt im Gegensatz zum Modal und Median im Allgemeinen auf den Mittelwert durch. Dies spricht einerseits für den Mittelwert, ist andererseits aber auch Ursache seiner großen Sensitivität gegenüber Ausreißern. Das mediane Alter in Beispiel 1.1 beträgt $m = 27$, der Stichprobenmittelwert $\bar{x} = 30.0$. Entfernt man die beiden Ausreißer 59 und 60 sinkt der Mittelwert auf 28.4 Jahre, während der Median gleich bleibt.

Satz 1.1 zeigt, daß ein natürliches Maß für die Variabilität der Daten durch ihre mittlere quadratische Abweichung, $\frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2$, vom Mittelwert gegeben ist (die Division durch n verhindert das Anwachsen der Summe allein auf Grund einer Vergrößerung des Stichprobenumfanges). Wie wir später sehen werden, ist jedoch folgende Definition des Streumaßes zweckmäßiger:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{x} - x_i)^2 \quad \text{Varianz}$$

$$\sigma = \sqrt{\sigma^2} \quad \text{Standardabweichung}$$

Es mag vielleicht erstaunen, daß man für σ und σ^2 eigene Bezeichnungen verwendet. Dies liegt daran, daß σ wieder die Dimension von x_i trägt, und daher leichter zu interpretieren ist, σ^2 hingegen ist leichter zu manipulieren. Die Definition von Varianz und Streuung ist in der Literatur nicht einheitlich: manche Autoren verwenden den Vorfaktor $\frac{1}{n}$!

Die Varianz kann etwas einfacher über den Verschiebungssatz berechnet werden:

$$(1.4) \quad \sigma^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

Dies ergibt sich mit Satz 1.1 und (1.3) folgendermaßen:

$$(n-1)\sigma^2 = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})x_i = \sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i^2 - n\bar{x}^2.$$

Es seien nun x_j , $j = 1, \dots, k$, die zahlenmäßig verschiedenen Ausprägungen eines quantitativen Merkmales mit relativen Häufigkeiten $h_r(x_j)$. Mittelwert und Varianz können dann folgendermaßen berechnet werden:

$$(1.5) \quad \bar{x} = \sum_{j=1}^k x_j h_r(x_j),$$

$$(1.6) \quad \sigma^2 = \frac{n}{n-1} \left[\sum_{j=1}^k x_j^2 h_r(x_j) - \bar{x}^2 \right].$$

KAPITEL 2

Elementare Wahrscheinlichkeit

1. Grundbegriffe der Wahrscheinlichkeitsrechnung

1.1. Zufallsexperiment, Ereignis. Das klassische Beispiel eines **Zufallsexperimentes** oder einer **Zufallsbeobachtung** ist das Werfen einer Münze: der Vorgang kann beliebig oft wiederholt werden, alle möglichen Ausgänge sind bekannt (Kopf oder Zahl), aber es ist unmöglich, das konkrete Ergebnis einer Durchführung des Experimentes vorherzusagen. Die möglichen Ergebnisse eines Zufallsexperiments heißen **Elementarereignisse**. Die Menge aller Elementarereignisse nennt man **Ereignisraum**, Ω .

- BEISPIEL 2.1. • Werfen einer Münze: $\Omega = \{K, Z\}$
- Werfen von 2 Münzen: $\Omega = \{KK, ZK, KZ, ZZ\}$
Dies beschreibt auch das zweimalige Werfen derselben Münze.
 - Werfen eines Würfels: $\Omega = \{1, 2, 3, 4, 5, 6\}$
 - Werfen von 2 Würfeln: $\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$
 - Messen der Länge eines zufällig gewählten Pfostens: $\Omega = [\bar{\ell} - s, \bar{\ell} + s]$. Es bezeichnet $\bar{\ell}$ den Sollwert und s ein Maß für die Ungenauigkeit der Säge.
 - Lebensdauer einer zufällig gewählten Glühbirne: $\Omega = [0, \infty)$.

Die Teilmengen von Ω nennt man **Ereignisse** (bei unendlichen Ereignisräumen führt aber nicht jede Teilmenge von Ω zu einem sinnvollen Ereignis). Beim zweimaligen Werfen einer Münze steht $A = \{KK, KZ\}$ für das Ereignis, daß die zuerst geworfene Münze Kopf zeigt. Nicht immer sind die Elementarereignisse selbst interessant: es ist beispielsweise unmöglich festzustellen, ob die Länge eines Pfostens *exakt* ℓ Meter beträgt. In der Praxis genügt es zu wissen, daß die Abweichungen vom Sollwert gewisse Toleranzen α nicht überschreiten. Man interessiert sich also für das Ereignis $A = \{l \in \Omega: l \in [\ell - \alpha, \ell + \alpha]\}$. Führt man das Zufallsexperiment durch, tritt das Ereignis A genau dann ein, wenn für das Ergebnis ω des Zufallsexperimentes $\omega \in A$ gilt. Ist das Ereignis A nicht eingetreten, dann ist das **Gegenereignis** $\bar{A} = \Omega \setminus A$ eingetreten. Bei einem Zufallsexperiment mit Ereignisraum Ω bezeichnet Ω das sichere Ereignis und \emptyset das unmögliche Ereignis. Wenn A und B Ereignisse sind, dann sind auch $A \cup B$ und $A \cap B$ wieder Ereignisse. $A \cup B$ tritt genau dann ein, wenn mindestens eines der Ereignisse A oder B eintritt, $A \cap B$ tritt genau dann ein, wenn beide Ereignisse A und B eintreten. Induktion folgt, daß man auf diese Weise endlich viele Ereignisse verknüpfen kann.

Fassen wir die Ereignisse in einer Menge \mathcal{E} zusammen. Unsere Diskussion zeigt, daß \mathcal{E} folgende Eigenschaften aufweisen soll: $\Omega \in \mathcal{E}$, aus $A \in \mathcal{E}$ folgt $\bar{A} \in \mathcal{E}$ und für endlich viele Ereignisse $A_1, \dots, A_n \in \mathcal{E}$ gilt $\cup_{i=1}^n A_i \in \mathcal{E}$. Es hat sich herausgestellt, daß es zweckmäßig ist, die Abgeschlossenheit von \mathcal{E} gegenüber der Vereinigung von abzählbar vielen Ereignissen zu fordern. Dies motiviert folgenden Begriff:

DEFINITION 2.1. *Ein System \mathcal{E} von Teilmengen von Ω heißt σ -Algebra, wenn es folgende Eigenschaften besitzt:*

$$(2.1) \quad \Omega \in \mathcal{E}$$

$$(2.2) \quad A \in \mathcal{E} \Rightarrow \bar{A} \in \mathcal{E}$$

$$(2.3) \quad A_i \in \mathcal{E}, \quad i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$$

Eine unmittelbare Folgerung dieser Eigenschaften ist:

$$\emptyset \in \mathcal{E}$$

$$A_i \in \mathcal{E}, \quad i \in \mathbb{N} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{E}$$

1.2. Der Wahrscheinlichkeitsbegriff. Es gibt viele Ansätze, den Begriff der Wahrscheinlichkeit zu definieren. Manche davon sind zu speziell, andere führen schließlich auf ernste mathematische Schwierigkeiten. Erst die axiomatische Betrachtungsweise brachte ein befriedigendes theoretisches Fundament. Ausgehend von unserer intuitiven Vorstellung von Wahrscheinlichkeit, wollen wir uns jene Eigenschaften plausibel machen, welche die axiomatische Grundlage der Wahrscheinlichkeitsrechnung ausmachen. Betrachten wir den Wurf einer fairen Münze: fair bedeutet, daß das Ergebnis Kopf bzw. Zahl gleich wahrscheinlich ist. Damit meinen wir nicht, daß etwa bei einer Serie von 10 Würfeln 5 mal Kopf und 5 mal Zahl fällt. Vielmehr stellen wir uns vor, daß die relative Häufigkeit des Ereignisses K bzw. Z sich immer weniger vom Wert $\frac{1}{2}$ unterscheidet und drücken dies mit $P(K) = P(Z) = \frac{1}{2}$ aus. Etwas allgemeiner: $P(A) = \alpha$ soll $\lim_{n \rightarrow \infty} h_r(A, n) = \alpha$ zum Ausdruck bringen. Dabei bezeichnet $h_r(A, n)$ die relative Häufigkeit des Eintreffens des Ereignisses A bei einer n -maligen Wiederholung des Zufallsexperimentes. Den genauen Zusammenhang zwischen relativer Häufigkeit und Wahrscheinlichkeit können wir erst später mit dem Gesetz der großen Zahlen klären. Abbildung 2.1 zeigt eine Computersimulation einer Serie von 100.000 Würfeln einer Münze, bei der die relative Häufigkeit des Ereignisses Kopf verfolgt wurde. Wie erwartet, stabilisieren sich die relativen Häufigkeiten sehr rasch beim Wert $\frac{1}{2}$.

Im Folgenden betrachten wir eine feste Anzahl von Wiederholungen des Zufallsexperimentes und schreiben daher $h_r(A)$ anstelle von $h_r(A, n)$. Unmittelbar einsichtig

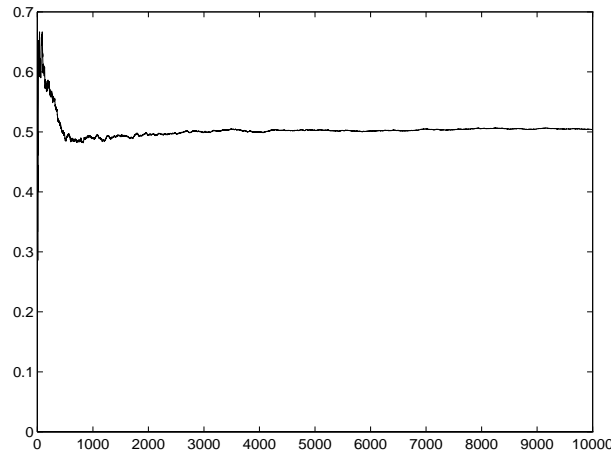


ABB. 2.1. Relative Häufigkeit von Kopf bei 100.000 Münzwürfen

ist

$$(2.4) \quad 0 \leq h_r(A) \leq 1.$$

Da $A = \Omega$ immer eintritt, gilt

$$(2.5) \quad h_r(\Omega) = 1.$$

Es seien nun A und B Ereignisse, welche in der Versuchsreihe mit den relativen Häufigkeiten $h_r(A)$ und $h_r(B)$ aufgetreten sind. Für einen endlichen Ereignisraum Ω gilt dann

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

Dies ist eine unmittelbare Konsequenz der disjunkten Zerlegungen

$$\begin{aligned} A \cup B &= (A \setminus B) \cup (B \setminus A) \cup (A \cap B) \\ A &= (A \setminus B) \cup (A \cap B) \quad A = (B \setminus A) \cup (A \cap B) \end{aligned}$$

indem man auf die Anzahl der Elemente übergeht und durch n dividiert. Insbesondere gilt für Ereignisse, welche sich gegenseitig ausschließen, für welche also $A \cap B = \emptyset$ zutrifft

$$(2.6) \quad h_r(A \cup B) = h_r(A) + h_r(B).$$

Es ist erstaunlich, daß (2.4), (2.5) und (2.6) ausreichen, um eine so reichhaltige Disziplin wie die Wahrscheinlichkeitsrechnung zu begründen. Es ist lediglich notwendig, die Additivität (2.6) auf abzählbar unendliche Systeme paarweise disjunkter Mengensysteme auszudehnen. Eine Familie A_i , $i \in \mathbb{N}$, heißt paarweise disjunkt, wenn für jedes Paar (i, j) , $i \neq j$, $A_i \cap A_j = \emptyset$ zutrifft.

DEFINITION 2.2 (Kolmogorov'sches Axiomensystem, 1933). *Es sei \mathcal{E} eine σ -Algebra von Ereignissen in Ω und p eine Abbildung $p: \mathcal{E} \rightarrow \mathbb{R}$. P heißt **Wahrscheinlichkeitsmaß** (auch **Wahrscheinlichkeitsverteilung**), wenn folgende Axiome erfüllt sind:*

$$(2.7) \quad P(A) \geq 0 \quad \text{für alle } A \in \mathcal{E},$$

$$(2.8) \quad P(\Omega) = 1,$$

(2.9) P ist σ -additiv: Für alle paarweise disjunkten Ereignisse A_i , $i \in \mathbb{N}$ gilt

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

Das Tripel (Ω, \mathcal{E}, P) heißt **Wahrscheinlichkeitsraum**.

Der wichtigste Schritt bei der Modellierung eines Zufallsexperimentes ist die Konstruktion eines passenden Wahrscheinlichkeitsraumes. Es ist möglich, daß man für ein Zufallsexperiment verschiedene Wahrscheinlichkeitsräume konstruieren kann, welche zu unterschiedlichen Ergebnissen führen. Vor der Einführung des Konzeptes eines Wahrscheinlichkeitsraumes wurden derartige Situationen als paradox empfunden.

Wir ziehen nun einige Schlußfolgerungen aus den Axiomen: unmittelbar klar ist die endliche Additivität

$$(2.10) \quad P(A \cup B) = P(A) + P(B)$$

für disjunkte Ereignisse. Setzt man $A = B = \emptyset$ folgt

$$P(\emptyset) = 0.$$

Wählt man $B = \bar{A}$, ergibt sich

$$(2.11) \quad P(\bar{A}) = 1 - P(A).$$

Jedes Wahrscheinlichkeitsmaß ist monoton:

$$(2.12) \quad A \subset B \Rightarrow P(A) \leq P(B), \quad A, B \in \mathcal{E}$$

Dies ergibt sich unmittelbar aus der disjunkten Zerlegung

$$B = A \cup (B \setminus A)$$

(2.10) und Axiom 2.8. Die Monotonie und Axiom 2.9 zeigen insbesondere

$$P(A) \leq 1, \quad A \in \mathcal{E}.$$

Die σ -Additivität faßt zwei grundverschiedene Eigenschaften zusammen: Die Additivität und die Stetigkeit im folgenden Sinne:

DEFINITION 2.3. *Es sei \mathcal{E} eine σ -Algebra über Ω und $P: \mathcal{E} \rightarrow \mathbb{R}$ eine Mengenfunktion.*

- P heißt **aufsteigend stetig**, wenn für jede Folge $(A_i) \subset \mathcal{E}$ mit $A_i \subset A_{i+1}$, $i \in \mathbb{N}$ gilt:

$$(2.13) \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$$

- P heißt **absteigend stetig**, wenn für jede Folge $(A_i) \subset \mathcal{E}$ mit $A_i \supset A_{i+1}$, $i \in \mathbb{N}$ gilt:

$$(2.14) \quad P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i)$$

PROPOSITION 2.1. Für Mengenfunktionen, die additiv sind und das Axiom 2.9 erfüllen, gilt:

- Auf- und absteigende Stetigkeit sind äquivalent;
- Die aufsteigende (absteigende) Stetigkeit ergibt sich bereits aus der Stetigkeit für Folgen, welche zu Ω aufsteigen (zu \emptyset absteigen).

BEWEIS. Den Beweis der ersten Behauptung überlassen wir dem Leser als einfache Übung. Es sei nun $(A_i) \subset \mathcal{E}$ eine beliebige aufsteigende Folge: $A_i \subset A_{i+1}$, $i \in \mathbb{N}$ und $A = \bigcup_{i=1}^{\infty} A_i$. Ferner setzen wir $B_i = \bar{A} \cup A_i$. Es gilt $B_i \subset B_{i+1}$ und $\bigcup_{i=1}^{\infty} B_i = \bar{A} \cup \bigcup_{i=1}^{\infty} A_i = \Omega$. Aus der Voraussetzung folgt daher $\lim_{i \rightarrow \infty} P(\bar{A} \cup A_i) = \lim_{i \rightarrow \infty} P(B_i) = 1$ und weiter mit (2.10) und (2.11)

$$\lim_{i \rightarrow \infty} (P(\bar{A}) + P(A_i)) = 1 - P(A) + \lim_{i \rightarrow \infty} P(A_i) = 1.$$

□

PROPOSITION 2.2. Eine Mengenfunktion ist genau dann σ -additiv, wenn sie additiv und aufsteigend (oder absteigend) stetig ist.

BEWEIS. Wir zeigen zuerst, daß aus der σ -Additivität die aufsteigende Stetigkeit, wegen Proposition 2.1 somit auch die absteigende Stetigkeit, folgt. Es sei also A_i , $i \in \mathbb{N}$, eine aufsteigende Folge. Die Folge $B_1 = A_1$ und $B_i = A_i \setminus A_{i-1}$, $i \geq 2$, ist paarweise disjunkt, somit folgt

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Umgekehrt sei nun P aufsteigend stetig und endlich additiv und $B_i \in \mathcal{E}$, $i \in \mathbb{N}$, paarweise disjunkt. Die Folge $A_n = \bigcup_{i=1}^n B_i$ ist aufsteigend. Somit folgt

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} B_i\right) &= P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \sum_{i=1}^{\infty} P(B_i) \end{aligned}$$

die σ -Additivität von P . Im obigen Argument gilt die zweite Gleichheit wegen der aufsteigenden Stetigkeit, die endliche Additivität sichert die vorletzte Gleichheit. \square

2. Diskrete Wahrscheinlichkeitsmaße

Ist Ω höchstens abzählbar, kann man als Ereignis- σ -Algebra $\mathcal{P}(\Omega)$ wählen. Betrachten wir zuerst einen endlichen Ereignisraum:

BEISPIEL 2.2 (Die diskrete Gleichverteilung). Es sei $\Omega = \{\omega_1, \dots, \omega_n\}$ eine endliche Menge und $\mathcal{E} = \mathcal{P}(\Omega)$. Ferner nehmen wir an, daß alle Elementarereignisse $\omega_1, \dots, \omega_n$ gleich wahrscheinlich sind, also $P(\omega_i) = p$, $i = 1, \dots, n$. Wegen $\Omega = \bigcup_{i=1}^n \omega_i$ folgt mit den Axiomen 2.8 und 2.9

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \omega_i\right) = \sum_{i=1}^n P(\omega_i) = np,$$

also,

$$p = \frac{1}{n},$$

was wir erwartet haben. Es sei nun $A = \{\omega_{i_1}, \dots, \omega_{i_k}\}$, $k \leq n$ ein zusammengesetztes Ereignis. Analog wie vorhin schließt man

$$P(A) = \sum_{j=1}^k P(\omega_{i_j}) = \frac{k}{n}.$$

Berücksichtigt man noch, daß k die Anzahl genau jener Ergebnisse des Zufallsexperimentes sind, bei welchen das Ereignis A eintritt, welche also "günstig" für A sind, und daß n die Anzahl aller möglichen Ausgänge angibt, erhält man den Wahrscheinlichkeitsbegriff von Laplace

$$(2.15) \quad P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

Es gibt allerdings auch Situationen, für welche ein endlicher Ereignisraum nicht geeignet ist. Man denke z.B. an die Brenndauer von Glühbirnen in Stunden, die Anzahl der Tippfehler pro Seite in diesem Skriptum, Anzahl der Blattläuse auf einem Apfelbaum einer Apfelplantage,...

BEISPIEL 2.3 (Diskrete Wahrscheinlichkeitsmaße). Es sei $\Omega = \{\omega_i : i \in \mathbb{N}\}$ höchstens abzählbar und $\mathcal{E} = \mathcal{P}(\Omega)$. Jedem Elementarereignis ω_i wird eine Wahrscheinlichkeit p_i zugeordnet, sodaß

$$p_i \geq 0, \quad i \in \mathbb{N} \quad \text{und} \quad \sum_{i=1}^{\infty} p_i = 1$$

zutrifft. Für jedes Ereignis $A \in \mathcal{E}$ setzt man

$$(2.16) \quad P(A) = \sum_{i=1}^{\infty} p_i \chi_A(\omega_i).$$

Dabei bezeichnet χ_A die charakteristische Funktion von A , es gilt also

$$\chi_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{sonst.} \end{cases}$$

Wir zeigen, daß (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum ist. Es ist klar, daß die Axiome 2.8 und 2.9 erfüllt sind. Es seien also $A_i, i \in \mathbb{N}$ paarweise disjunkte Ereignisse. Somit gilt

$$(2.17) \quad \chi_{\cup_{i=1}^{\infty} A_i} = \sum_{i=1}^{\infty} \chi_{A_i},$$

denn die linke Seite ist für ein Elementarereignis ω genau dann gleich 1, wenn $\omega \in \cup_{i=1}^{\infty} A_i$ liegt. Da die Ereignisse A_i paarweise disjunkt sind, gilt $\omega \in A_{i_0}$ für genau ein $i_0 \in \mathbb{N}$. Folglich reduziert sich auch die Summe auf der rechten Seite auf $\chi_{A_{i_0}}(\omega) = 1$. Wir schließen weiter

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{k=1}^{\infty} p_k \chi_{\cup_{i=1}^{\infty} A_i}(\omega_k) = \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} p_k \chi_{A_i}(\omega_k) \\ &= \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} p_k \chi_{A_i}(\omega_k) = \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

(die Vertauschung der Summationsreihenfolge ist gerechtfertigt, da alle Summanden nicht negativ sind).

BEISPIEL 2.4. Jeder hat sich sicherlich schon einmal über eine lange Pechsträhne bei "Mensch Ärgere Dich Nicht" geärgert und gefragt, wie groß eigentlich die Wahrscheinlichkeit sei, daß eine Sechs etwa erst beim 5. Wurf fällt oder irgendwann einmal fällt. Dazu betrachten wir die Ereignisse

$A_k \equiv$ eine Sechs fällt zum ersten Mal beim k -ten Wurf, $k \in \mathbb{N}$

$A \equiv$ eine Sechs fällt irgendwann

Offensichtlich wird durch $k \in \mathbb{N}$ das Ereignis A_k eindeutig bestimmt und umgekehrt. Man kann also als Ereignisraum $\Omega = \mathbb{N}$ wählen. Als Wahrscheinlichkeiten für die Elementarereignisse setzen wir fest

$$(2.18) \quad P(A_k) = P(k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}, \quad k \in \mathbb{N}.$$

Dies ergibt

$$P(A) = P\left(\bigcup_{k=1}^{\infty} k\right) = \sum_{k=1}^{\infty} P(k) = \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \frac{1}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k = \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 1.$$

Dies bedeutet aber nicht, daß mit *Sicherheit* irgendwann einmal eine Sechs fallen muß. Wegen $A = \Omega$ ist a posteriori das Ergebnis nur die Bestätigung, daß (2.18) ein Wahrscheinlichkeitsmaß definiert. Wie kommt man zu den Wahrscheinlichkeiten in (2.18)? Dazu fixieren wir ein beliebiges k . Der Ereignisraum Ω_k für eine Serie von k Würfeln ist $\Omega_k = \{1, 2, 3, 4, 5, 6\}^k$ und enthält 6^k Elementarereignisse. Die i -te Koordinate des k -Tupels beschreibt das Ergebnis des i -ten Wurfes. Alle k -Tupel sind gleichwahrscheinlich. Ferner gilt

$$A_k = \{(i_1, \dots, i_{k-1}, 6) : i_j \in \{1, \dots, 5\}, j = 1, \dots, k-1\}.$$

Das Ereignis A_k wird also durch genau 5^{k-1} günstige Elementarereignisse realisiert, aus (2.15) folgt daher (2.18). Wir werden später eine einfachere Möglichkeit kennenlernen, derartige Wahrscheinlichkeiten zu berechnen.

3. Stetige Wahrscheinlichkeitsmaße

Das theoretische Fundament diskreter und stetiger Wahrscheinlichkeitsmaße wird in der Maßtheorie gelegt. Wir haben bereits darauf hingewiesen, daß für $\Omega = \mathbb{R}^m$ in den Anwendungen die interessierenden Ereignisse meist von der Form $\omega \leq b$, $a \leq \omega < b$ oder $a \leq \omega \leq b$, etc. sind (für $m > 1$ sind die Ungleichungen koordinatenweise zu interpretieren). In der Maßtheorie wird gezeigt, daß das System der m -dimensionalen Quader eine eindeutig bestimmte σ -Algebra über \mathbb{R}^m erzeugt. Dies ist die Borel'sche σ -Algebra \mathcal{B}^m . Sie ist außerordentlich umfangreich: beispielsweise sind sämtliche offenen und abgeschlossenen Teilmengen von \mathbb{R}^m Borelmengen. Es ist sogar sehr schwierig, eine Menge zu konstruieren, welche nicht in \mathcal{B}^m liegt. Allerdings genügt es, ein Wahrscheinlichkeitsmaß nur für Quader (oder für ein anderes Erzeugendensystem) zu definieren. Es ist dann automatisch auf der gesamten Borel'schen σ -Algebra festgelegt.

DEFINITION 2.4. Eine Abbildung $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ heißt **Borel meßbar**, wenn das Urbild jeder Borelmenge in \mathbb{R}^n wieder eine Borelmenge des \mathbb{R}^m ist.

Insbesondere sind stetige Funktionen, stückweise stetige Funktionen Borel meßbar.

DEFINITION 2.5. Es sei $p: \mathbb{R}^m \rightarrow \mathbb{R}$ eine Borel meßbare Funktion mit folgenden Eigenschaften:

$$(2.19) \quad \begin{aligned} p(x) &\geq 0, && \text{für alle } x \in \mathbb{R}^m \\ \int_{\mathbb{R}^m} p(x) dx &= 1 \end{aligned}$$

Auf \mathcal{B}^m definieren wir ein **stetiges Wahrscheinlichkeitsmaß** durch

$$P(A) = \int_{\mathbb{R}^m} p(x) \chi_A(x) dx, \quad \text{für alle } A \in \mathcal{B}^m.$$

p heißt **Dichte** von P .

Es sei darauf hingewiesen, daß das Integral im Sinne von Lebesgue zu bilden ist. Im Folgenden werden wir uns meist mit Wahrscheinlichkeitsmaßen mit stetiger Dichte beschäftigen, für welche der Lebesgue'sche und der Riemann'sche Integralbegriff zusammenfallen. Der Zusatz *stetig* beim Wahrscheinlichkeitsmaß hat übrigens nichts mit der Stetigkeit von p zu tun, sondern weist darauf hin, daß das Wahrscheinlichkeitsmaß eine integrierbare Dichte besitzt.

Wir zeigen nun, daß $(\mathbb{R}^m, \mathcal{B}^m, P)$ ein Wahrscheinlichkeitsraum ist. Die beiden Eigenschaften 2.8 und 2.9 sind klar wegen (2.19). Es seien nun A_n , $n \in \mathbb{N}$, paarweise disjunkte Borelmengen. Erinnern wir uns an (2.17), so erhalten wir

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int p(x) \chi_{\bigcup_{i=1}^{\infty} A_i}(x) dx = \int \sum_{i=1}^{\infty} p(x) \chi_{A_i}(x) dx \\ &= \sum_{i=1}^{\infty} \int p(x) \chi_{A_i}(x) dx = \sum_{i=1}^{\infty} P(A_i). \end{aligned}$$

Die Vertauschung von Summation und Integration kann mit dem Satz von der dominierten Konvergenz gerechtfertigt werden. Wir haben zur Vereinfachung der Schreibweise

$$\int f(x) dx \quad \text{anstelle von} \quad \int_{\mathbb{R}^m} f(x) dx$$

geschrieben.

BEISPIEL 2.5 (Stetige Gleichverteilung). Es sei $B \in \mathcal{B}^m$ eine feste Borelmenge mit $0 < \int_B dx < \infty$. Wählt man die Dichte

$$p = \frac{\chi_B}{\int_B dx}$$

erhält man die stetige Gleichverteilung über B

$$P(A) = \frac{\int_{A \cap B} dx}{\int_B dx}, \quad A \in \mathcal{B}^m.$$

Es sei $(\mathbb{R}^m, \mathcal{B}^m, P)$ ein stetiger Wahrscheinlichkeitsraum mit stetiger Dichte p und $x \in \mathbb{R}^m$ ein beliebiges Elementarereignis. Wir können x in eine Folge von abgeschlossenen Quadern Q_i so einschließen, daß $\bigcap_{i=1}^{\infty} Q_i = \{x\}$ und $Q_i \supset Q_{i+1}$, $i \in \mathbb{N}$ gilt. Da p stetig und Q_1 kompakt ist, ist p auf Q_1 beschränkt, $0 \leq p(x) \leq M$, $x \in Q_1$. Somit schließen wir

$$P(Q_i) = \int_{Q_i} p(x) dx \leq M \int_{Q_i} dx \xrightarrow{i \rightarrow \infty} 0.$$

Aus der absteigenden Stetigkeit eines Wahrscheinlichkeitsmaßes folgt nun

$$P(\{x\}) = P(\bigcap Q_i) = \lim_{i \rightarrow \infty} P(Q_i) = 0.$$

Diese Beobachtung untermauert die eingangs getroffenen Bemerkung, daß in einem stetigen Wahrscheinlichkeitsraum die Elementarereignisse nur eine untergeordnete Bedeutung besitzen.

3.1. Paradoxon von Bertrand. Es wurde bereits erwähnt, daß ein und dasselbe Zufallsexperiment durch verschiedenen Wahrscheinlichkeitsräume beschrieben werden kann. Dazu betrachten wir folgendes Beispiel:

BEISPIEL 2.6. In einen Kreis mit dem Radius r wird willkürlich eine Sehne eingezeichnet. Wie groß ist die Wahrscheinlichkeit, daß die Länge l der Sehne größer ist, als die Länge a der Seite des eingeschriebenen gleichseitigen Dreiecks.

1. *Lösungsmöglichkeit:* Kein Punkt auf der Peripherie des Kreises ist vor einem anderen ausgezeichnet: wir können also den Anfangspunkt P_0 der Sehne beliebig wählen. Betrachtet man jenes eingeschriebene gleichseitige Dreieck, dessen Streckensymmetrale normal zur Tangente in P_1 ist, dann tritt das Ereignis " $l > a$ " genau dann ein, wenn der Endpunkt der Sehne auf dem (kürzeren) Kreisbogen $\bar{A}B$ liegt: der Vergleich des günstigen und des möglichen Winkelbereiches für die Sehne ergibt, vgl. Abbildung 2.2

$$P(l > a) = \frac{\frac{\pi}{3}}{\pi} = \frac{1}{3}.$$

2. *Lösungsmöglichkeit:* Keine Richtung ist im Kreis vor einer anderen ausgezeichnet: wir können also die Richtung der Sehne beliebig wählen. Das Ereignis " $l > a$ " tritt nun genau dann ein, wenn die Sehne zwischen l_0 und l_1 liegt und parallel zu l_0 verläuft. Die Lage der Sehne kann durch ihren Abstand von P_0 charakterisiert werden, vgl. Abbildung 2.3. Somit findet man

$$P(l > a) = \frac{\frac{r}{4} + \frac{r}{4}}{r} = \frac{1}{2}.$$

3. *Lösungsmöglichkeit:* Kein Punkt im Kreis ist vor einem anderen ausgezeichnet: jeder kommt mit gleicher Wahrscheinlichkeit als Mittelpunkt M der Sehne in Frage.

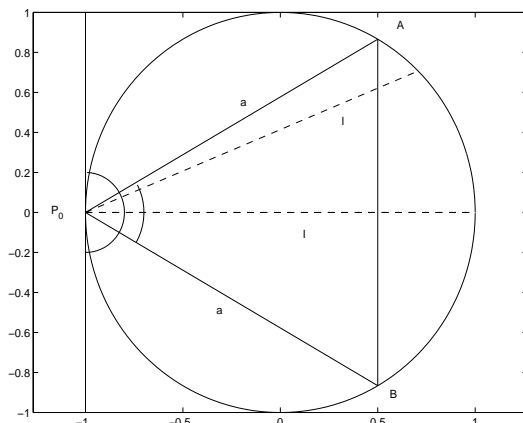


ABB. 2.2

Das Ereignis " $l > a$ " tritt genau dann ein, wenn M im Inneren des konzentrischen Kreises mit Radius $\frac{r}{2}$ liegt, vgl. Abbildung 2.4. Die Laplace'sche Wahrscheinlichkeit für das Ereignis " $l > a$ " ist nun

$$P(l > a) = \frac{\left(\frac{r}{2}\right)^2 \pi}{r^2 \pi} = \frac{1}{4}.$$

Wie ist es möglich, daß wir auf dieselbe Frage unterschiedliche Antworten bekommen? Der Punkt ist, daß die Frage eben nicht immer gleich war: Im ersten Fall ist der Ereignisraum $\Omega_1 = [0, \pi]$, im zweiten Fall $\Omega_2 = [0, r]$, bei der dritten Betrachtungsweise schließlich $\Omega_3 = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$. Da wir bei der Interpretation von "gleichwahrscheinlich" jeweils von verschiedenen Ereignisräumen, also von verschiedenen Annahmen über das Experiment, ausgegangen sind, ist es verständlich und

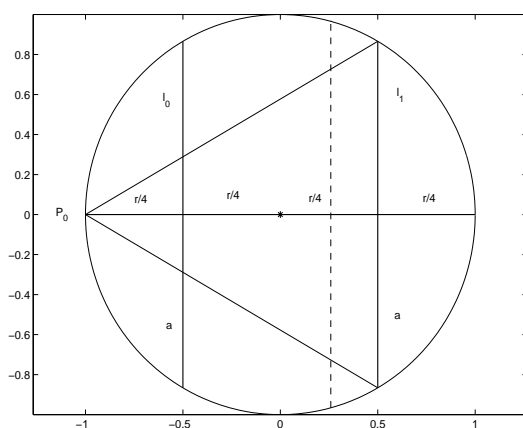


ABB. 2.3

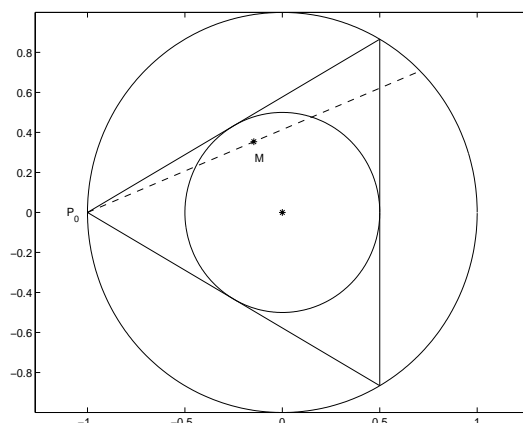


ABB. 2.4

keineswegs mehr paradox, daß wir auch zu unterschiedlichen Wahrscheinlichkeiten gelangen.

4. Verteilungsfunktionen

Wir haben bereits gesehen, daß die kumulative relative Häufigkeitsverteilung bei der Datenanalyse nützlich sein kann. Das wahrscheinlichkeitstheoretische Analogon sind die Verteilungsfunktionen. Wir beschränken uns hier auf Wahrscheinlichkeitsmaße über \mathbb{R} .

DEFINITION 2.6. *Es sei $(\mathcal{B}, \mathbb{R}, P)$ ein Wahrscheinlichkeitsraum. Die Funktion $F: \mathbb{R} \rightarrow [0, 1]$, definiert durch*

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

heißt **Verteilungsfunktion** von P .

THEOREM 2.1. *Jede Verteilungsfunktion F hat folgende Eigenschaften:*

$$(2.20) \quad F \text{ ist monoton steigend.}$$

$$(2.21) \quad \lim_{r \rightarrow -\infty} F(r) = 0, \quad \lim_{r \rightarrow \infty} F(r) = 1.$$

$$(2.22) \quad F \text{ ist rechtsseitig stetig.}$$

BEWEIS. ad 1) Aus $a < b$ folgt

$$F(b) = P((-\infty, b]) = P((-\infty, a]) + P((a, b]) \geq P((-\infty, a]) = F(a).$$

ad 2) Es sei (r_n) eine beliebige Folge mit $\lim_{n \rightarrow \infty} r_n = -\infty$. Somit ist $(-\infty, r_n]$ eine absteigende Folge von Intervallen mit $\bigcap (-\infty, r_n] = \emptyset$. Die Behauptung folgt nun aus

der absteigenden Stetigkeit von P (vgl. Proposition 2.2 und (2.14))

$$\lim_{n \rightarrow \infty} F(r_n) = \lim_{n \rightarrow \infty} P((-\infty, r_n]) = P\left(\bigcap_{n=1}^{\infty} (-\infty, r_n]\right) = P(\emptyset) = 0.$$

Die zweite Behauptung folgt analog.

ad 3) Wir betrachten eine beliebige Folge $s_n \geq r$ mit $\lim_{n \rightarrow \infty} s_n = r$. Wie vorhin schließen wir aus $(-\infty, r] = \bigcap_{n=1}^{\infty} (-\infty, s_n]$ mit Hilfe der absteigenden Stetigkeit auf

$$F(r) = P((-\infty, r]) = P\left(\bigcap_{n=1}^{\infty} (-\infty, s_n]\right) = \lim_{n \rightarrow \infty} P((-\infty, s_n]) = \lim_{n \rightarrow \infty} F(s_n).$$

□

Eine Verteilungsfunktion ist im Allgemeinen nicht auch noch linksseitig stetig: ist etwa $s_n \leq r$, und $\lim_{n \rightarrow \infty} s_n = r$, dann folgt aus $\bigcup_{n=1}^{\infty} (-\infty, s_n] = (-\infty, r)$ und der aufsteigenden Stetigkeit von P nur

$$\lim_{n \rightarrow \infty} F(s_n) = \lim_{n \rightarrow \infty} P((-\infty, s_n]) = P\left(\bigcup_{n=1}^{\infty} (-\infty, s_n]\right) = P((-\infty, r)).$$

Die Verteilungsfunktion ist also genau dann stetig an der Stelle r , wenn

$$P((-\infty, r)) = P((-\infty, r]) \quad (= P((-\infty, r)) + P(\{r\}))$$

also

$$P(\{r\}) = 0$$

gilt. Punkte r mit $P(\{r\}) > 0$ nennt man **Atome** von P . Da eine monotone und beschränkte Funktion höchstens abzählbar viele Sprungstellen besitzen kann, hat ein Wahrscheinlichkeitsmaß auf \mathbb{R} höchstens abzählbar viele Atome.

Der nächste Satz zeigt, daß durch die Verteilungsfunktion ein Wahrscheinlichkeitsmaß bereits eindeutig festgelegt ist.

THEOREM 2.2. *Zu jeder Funktion $F: \mathbb{R} \rightarrow [0, 1]$ mit den Eigenschaften (2.20)–(2.22) gibt es genau ein Wahrscheinlichkeitsmaß, dessen Verteilungsfunktion F ist.*

Wir betrachten nun den Zusammenhang zwischen Verteilungsfunktion und Dichte. Wir zitieren dazu einen tiefliegenden Satz aus der Maßtheorie:

THEOREM 2.3. *Wenn P eine Dichte besitzt, dann ist auch die Ableitung der Verteilungsfunktion eine Dichte.*

Nach diesem Satz kann man Dichten durch Differenzieren der Verteilungsfunktion berechnen. Dabei wird aber vorausgesetzt, daß eine Dichte tatsächlich existiert. Die Abbildung 1.3 veranschaulicht die Problematik: diese Verteilungsfunktion ist mit Ausnahme von endlich vielen Stellen differenzierbar, die Ableitung ist identisch Null, also sicherlich keine Dichte. Eine in diesem Zusammenhang nützliche hinreichende Bedingung ist:

THEOREM 2.4. *Ist F überall differenzierbar, dann ist die Ableitung F' eine Dichte.*

5. Laplace'sche Zufallsexperimente

Wir betrachten Zufallsexperimente mit einem endlichen Ereignisraum Ω , sämtliche Elementarereignisse seien gleichwahrscheinlich. Wir haben uns bereits überlegt, daß in diesem Fall die Wahrscheinlichkeit eines zusammengesetzten Ereignisses gegeben ist durch (2.15)

$$(2.23) \quad P(A) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

Für die Berechnung dieser Wahrscheinlichkeit ist es notwendig, die Anzahl der jeweils günstigen, bzw. möglichen Fälle zu bestimmen. Die Theorie des Abzählens von Mengen wird in der Kombinatorik entwickelt. Wir erinnern hier an einige nützliche Grundbegriffe.

5.1. Grundbegriffe der Kombinatorik. Wir betrachten 2 Klassen von Problemen: Reihenfolgeprobleme (Permutationen) und Auswahlprobleme (Kombinationen, Variationen). In den Anwendungen ist es allerdings meist zweckmäßiger, sich die entsprechenden Anzahlen direkt zu überlegen, als einen fertigen Formalismus zu verwenden. Die Heuristiken, welche den Resultaten zugrunde liegen, sind wichtiger als die Formeln selbst. Im Folgenden wird mit $|A|$ die Anzahl der Elemente einer endlichen Menge bezeichnet. Wir beginnen mit einem fundamentalen Zählprinzip:

PROPOSITION 2.3. *Es seien A_1, \dots, A_k endliche Mengen. Dann gilt*

$$|A_1 \times A_2 \times \dots \times A_k| = |A_1| \times |A_2| \times \dots \times |A_k|.$$

5.1.1. *Permutationen ohne Wiederholung.* Unter einer *Permutation* von n Elementen versteht man eine bestimmte Anordnung dieser Elemente (also eine Bijektion dieser Elemente auf $\{1, \dots, n\}$).

PROPOSITION 2.4. 1. *Die Anzahl der verschiedenen Anordnungen von n unterschiedlichen Elementen beträgt $n!$.*

2. *Die Anzahl der Reihenfolgen, in denen man n unterschiedliche Elemente auf $k \leq n$ Plätze verteilen kann, beträgt $n \cdot (n - 1) \cdot \dots \cdot (n - k + 1) = \frac{n!}{(n-k)!}$.*

Dieses Ergebnis kann man sich leicht plausibel machen (für einen exakten Beweis verweisen wir auf die Lehrbücher der Kombinatorik): Wir möchten n unterschiedliche Elemente auf n Plätze verteilen. Für den 1. Platz stehen alle n Elemente zur Verfügung, für den 2. Platz nur mehr $n - 1$ Elemente, für den k -ten schließlich nur mehr $n - (k - 1)$ Elemente. Die Behauptung ergibt sich nun aus Proposition 2.3.

Wegen des enormen Wachstums der Fakultät ist die Berechnung von $n!$ für große Werte von n mühsam. Folgende Approximation ist nützlich:

PROPOSITION 2.5 (Stirling'sche Formel).

$$n! \approx \sqrt{2\pi n} n^n e^{-n}$$

5.1.2. *Permutationen mit Wiederholung.*

BEISPIEL 2.7. Vor der Kasse eines Supermarktes stehen 20 Personen an. Davon sind 9 Frauen, 8 Männer und 3 Kinder. Wieviele verschiedene Warteschlangen sind möglich, wenn die Wartenden nur danach unterschieden werden, ob sie Frau, Mann oder Kind sind.

Die Lösung überlegt man sich folgendermaßen: Geht man von 20 unterscheidbaren Personen aus, gibt es $20!$ verschiedene Warteschlangen. Betrachten wir nun eine feste Warteschlange. Nach dem Zählprinzip in Proposition 2.3 gibt es $9! 8! 3!$ Permutationen dieser Warteschlange mit derselben Abfolge von Frau–Mann–Kind. Gibt man die Unterscheidbarkeit der Frauen, Männer und Kinder auf, sind diese Permutationen nicht mehr unterscheidbar. Die gesuchte Anzahl z ergibt sich daher aus $20! = 9! 8! 3! z$ zu $z = \frac{20!}{9! 8! 3!}$.

Allgemeiner gilt

PROPOSITION 2.6. *Partitioniert man eine Menge von n Elementen in r Klassen, welche je k_i , $i = 1, \dots, r$ Elemente enthalten, dann ist die Anzahl der möglichen Partitionen (Reihenfolgen) gegeben durch*

$$\frac{n!}{k_1! k_2! \dots k_r!}, \quad k_1 + k_2 + \dots + k_r = n.$$

Wir betrachten nun die Anzahl der Möglichkeiten, k Elemente aus einer Menge von n Elementen auszuwählen. Dabei ist zu beachten, ob die Reihenfolge, in der die Elemente gezogen werden, wesentlich ist und ob ein Element mehrfach gezogen werden kann.

5.1.3. *Kombinationen ohne Wiederholung.* Wir betrachten zuerst den Fall, daß die Reihenfolge der Ziehung berücksichtigt werden muß.

PROPOSITION 2.7. *Aus einer Menge von n Elementen kann man auf*

$$n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

verschiedene Arten geordnete Stichproben vom Umfang $k \leq n$ ohne Wiederholung ziehen.

Dies ist natürlich nur eine andere Formulierung von Proposition 2.4. Soll die Reihenfolge der Elemente in der Stichprobe nicht berücksichtigt werden, sind jeweils sämtliche $k!$ Stichproben, bei welchen dieselben Elemente ausgewählt wurden, zu identifizieren:

PROPOSITION 2.8. *Aus einer Menge von n Elementen kann man auf*

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

verschiedene Arten ungeordnete Stichproben vom Umfang $k \leq n$ ohne Wiederholung ziehen. Dies entspricht der Anzahl der k -elementigen Teilmengen der Ausgangsmenge.

5.1.4. *Kombinationen mit Wiederholung.* Stellt man sich vor, daß eine Auswahl mit Wiederholungen dadurch realisiert wird, daß das jeweils ausgewählte Element wieder in die Grundmenge zurückgelegt wird, stehen bei jedem Zug wieder alle Elemente der Grundmenge zur Verfügung. Eine unmittelbare Folgerung aus Proposition 2.3 ist somit

PROPOSITION 2.9. *Aus einer Menge von n Elementen kann man auf n^k verschiedene Arten geordnete Stichproben mit Wiederholung vom Umfang k ziehen.*

Ziehen wir nun eine ungeordnete Stichprobe vom Umfang k mit Wiederholung aus einer Menge mit n Elementen. Ohne Beschränkung der Allgemeinheit kann man M mit $\{1, \dots, n\}$ identifizieren. Die Stichprobe notieren wir in der Form $[a_1, a_2, \dots, a_k]$ (die Verwendung von Mengenklammern wäre an dieser Stelle unzulässig!). Da die Reihenfolge der gezogenen Elemente nicht beachtet wird, können wir ohne weiteres $a_1 \leq a_2 \leq \dots \leq a_k$ annehmen. Jeder Stichprobe $[a_1, a_2, \dots, a_k]$ kann man nun eine ungeordnete Stichprobe $\{b_1, \dots, b_k\}$ von unterscheidbaren Elementen zuordnen; zum Beispiel durch die Vorschrift:

$$[a_1, a_2, \dots, a_k] \leftrightarrow \{a_1, a_2 + 1, \dots, a_k + k - 1\} = \{b_1, b_2, \dots, b_k\}.$$

Für die Zahlen b_i gilt offenbar $1 \leq b_i \leq n + k - 1$, $i = 1, \dots, k$. Jeder Stichprobe $[a_1, a_2, \dots, a_k]$ wird auf diese Weise genau eine k -elementige Teilmenge von $\{1, \dots, n + k - 1\}$ zugeordnet und umgekehrt. Aus Proposition 2.8 folgt somit

PROPOSITION 2.10. *Aus einer Menge von n Elementen kann man auf*

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

verschiedene Arten ungeordnete Stichproben mit Wiederholung vom Umfang k entnehmen.

5.2. Beispiele.

BEISPIEL 2.8 (Paradoxon von de Méré). Wir betrachten 2 Würfelspiele: Beim ersten Spiel wird ein Würfel 4mal geworfen und der Spieler gewinnt, wenn mindestens eine Sechs fällt. Im anderen Spiel werden 2 Würfel 24mal geworfen, der Spieler gewinnt, wenn mindestens eine doppelte Sechs fällt. Der Marquis meinte, die Gewinnwahrscheinlichkeit sei bei beiden Spielen gleich groß, nämlich $\frac{4}{6}$ bei der ersten Variante und $\frac{24}{36}$ bei der zweiten Variante.

Das erste Spiel kann als Laplace Experiment mit Grundraum

$$\Omega_1 = \{(w_1, w_2, w_3, w_4) : w_i \in \{1, \dots, 6\}, i = 1, \dots, 4\}$$

gedeutet werden. Jedes Spielergebnis ist gleichwahrscheinlich, $p = \frac{1}{6^4}$. Der Spieler gewinnt, wenn folgendes Ereignis eintritt:

$$G_1 = \{(w_1, w_2, w_3, w_4) : \exists i(1 \leq i \leq 4) \text{ mit } x_i = 6\},$$

er verliert, wenn das Gegenereignis \bar{G}_1 eintritt:

$$\bar{G}_1 = \{(w_1, w_2, w_3, w_4) : w_i \in \{1, \dots, 5\}, i = 1, \dots, 4\}.$$

Aus $|\bar{G}_1| = 5^4$ folgt $P(\bar{G}_1) = \frac{5^4}{6^4}$ und somit

$$P(G_1) = 1 - P(\bar{G}_1) = 1 - \frac{5^4}{6^4} \approx 0.518.$$

Der Ereignisraum der zweiten Spielvariante ist

$$\Omega_2 = \{((x_1, y_1), (x_2, y_2), \dots, (x_{24}, y_{24})) : (x_i, y_i) \in \{1, \dots, 6\}^2, i = 1, \dots, 24\}$$

mit $|\Omega_2| = 36^{24}$ möglichen gleichwahrscheinlichen Ergebnissen. Der Spieler gewinnt, falls das Ereignis

$$G_2 = \{((x_1, y_1), (x_2, y_2), \dots, (x_{24}, y_{24})) : \exists i(1 \leq i \leq 24) \text{ mit } (x_i, y_i) = (6, 6)\},$$

und er verliert, wenn das Gegenereignis \bar{G}_2 eintritt:

$$\bar{G}_2 = \{((x_1, y_1), (x_2, y_2), \dots, (x_{24}, y_{24})) : (x_i, y_i) \in \{1, \dots, 6\}^2 \setminus \{(6, 6)\}, i = 1, \dots, 24\}$$

(es gilt nicht $\bar{G}_2 = \Omega_2 \setminus \{(6, 6)\}$!). Für jeden Wurf eines Spieles, welches verloren wird, gibt es also nur 35 mögliche Ergebnisse, somit $|\bar{G}_2| = 35^{24}$. Dies ergibt

$$P(G_2) = 1 - P(\bar{G}_2) = 1 - \frac{35^{24}}{36^{24}} \approx 0,491.$$

BEISPIEL 2.9. Wir betrachten nun ein Laplace Experiment auf dem Ereignisraum

$$\Omega = \underbrace{\Omega_0 \times \dots \times \Omega_0}_{k\text{-mal}},$$

$$(|\Omega_0| = n, \quad k \leq n).$$

Diese Situation lag im vorangehenden Beispiel vor. Gesucht ist die Wahrscheinlichkeit des Ereignisses

$$E = \{(\omega_1, \dots, \omega_k) \in \Omega : \exists i, j \in \{1, \dots, k\} \text{ mit } i \neq j \text{ und } \omega_i = \omega_j\}.$$

Wie vorhin betrachten wir das Gegenereignis

$$\bar{E} = \{(\omega_1, \dots, \omega_k) \in \Omega : \omega_i \text{ sind paarweise verschieden, } 1 \leq i \leq k\}.$$

Aus $|\bar{E}| = n(n-1) \cdots (n-k+1)$ folgt

$$P(\bar{E}) = \frac{n(n-1) \cdots (n-k+1)}{n^k} = \prod_{i=1}^k \left(1 - \frac{i-1}{n}\right)$$

$$= e^{\sum_{i=1}^k \ln\left(1 - \frac{i-1}{n}\right)}.$$

Diese Wahrscheinlichkeit kann man mit dem Mittelwertsatz der Differentialrechnung abschätzen: für $0 \leq x < 1$ gilt

$$\ln(1-x) = \ln(1-x) - \ln 1 = \frac{1}{1-\xi}(-x) \leq -x,$$

für ein $\xi \in (0, x)$. Somit folgt

$$P(\bar{E}) \leq e^{-\sum_{i=1}^k \frac{i-1}{n}} = e^{-\frac{k(k-1)}{2n}},$$

und schließlich

$$P(E) \geq 1 - e^{-\frac{k(k-1)}{2n}}.$$

Eine populäre Anwendung dieses Beispiels ist die Frage nach der Wahrscheinlichkeit, daß in einer Gruppe von k Personen mindestens 2 am selben Tag Geburtstag haben (dabei wird von Schaltjahren abgesehen). Dies kann man auf die eben betrachtete Situation zurückführen, indem man $\Omega_0 = \{1, \dots, 365\}$, also $n = 365$ setzt. Demnach ist

$$P(E) \geq 1 - e^{-\frac{k(k-1)}{2 \cdot 365}},$$

überraschenderweise ist $P(E) > 0.5$ bereits ab $k = 23$.

BEISPIEL 2.10 (Rencontre Problem). Wir betrachten den Ereignisraum

$$\Omega = \{\pi : \pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}, \pi \text{ ist bijektiv}\},$$

Ω ist also die Menge aller Permutationen der Elemente von $\{1, \dots, n\}$. Somit gilt $|\Omega| = n!$. Wir suchen die Wahrscheinlichkeit, daß eine beliebige Permutation keine Fixpunkte besitzt, also die Wahrscheinlichkeit des Ereignisses

$$E_n = \{\pi \in \Omega : \pi(i) \neq i, 1 \leq i \leq n\}.$$

Dazu muß die Anzahl a_n der Elemente von E_n bestimmt werden. Offenbar ist $a_1 = 0$, $a_2 = 1$. Es sei also $n \geq 3$ und $k \in \{2, \dots, n\}$. Wir betrachten folgende Teilmengen von E_n

$$P_k = \{\pi \in \Omega : \pi(i) \neq i, 1 \leq i \leq n, \pi(k) = 1\}.$$

Die Mengen P_k sind paarweise disjunkt, aus Symmetriegründen gilt

$$|P_2| = |P_3| = \dots = |P_n|,$$

was

$$a_n = |E_n| = (n-1)|P_2|$$

zur Folge hat. Wir zerlegen nun P_2 in zwei disjunkte Teilmengen

$$P_2 = P_{21} + P_{22}$$

mit

$$P_{21} = \{\pi \in P_2: \pi(1) = 2\}$$

$$P_{22} = \{\pi \in P_2: \pi(1) \neq 2\}.$$

Da P_{21} bzw. P_{22} die Menge der fixpunktfreien Permutationen von n Elementen ist, bei denen 2 bzw. 1 Bild fixiert sind, ist P_{21} isomorph zu E_{n-2} und P_{22} isomorph zu E_{n-1} . Daraus ergibt sich

$$|P_{21}| = |E_{n-2}| = a_{n-2}, \quad |P_{22}| = |E_{n-1}| = a_{n-1},$$

insgesamt also

$$a_n = (n-1)(a_{n-1} + a_{n-2}), \quad n \geq 3.$$

Die Lösung dieser zweistufigen Rekursion ist gegeben durch

$$a_n = n! \sum_{i=0}^n (-1)^i \frac{1}{i!}.$$

Der Beweis dieser Behauptung wird mit vollständiger Induktion geführt: Die Behauptung stimmt für $n = 3$. Der Induktionsschritt verwendet die Rekursion

$$\begin{aligned} a_{n+1} &= n(a_n + a_{n-1}) = n n! \sum_{i=0}^n (-1)^i \frac{1}{i!} + n! \sum_{i=0}^{n-1} (-1)^i \frac{1}{i!} \\ &= n! \left(n \sum_{i=0}^n (-1)^i \frac{1}{i!} + \sum_{i=0}^{n-1} (-1)^i \frac{1}{i!} - (-1)^n \frac{1}{n!} \right) \\ &= (n+1)! \sum_{i=0}^{n+1} (-1)^i \frac{1}{i!}. \end{aligned}$$

Die Anzahl der fixpunktfreien Permutationen von n Elementen beträgt also

$$(2.24) \quad |E_n| = n! \sum_{i=0}^n (-1)^i \frac{1}{i!}.$$

Die Wahrscheinlichkeit einer zufälliger Auswahl einer fixpunktfreien Permutation aus Ω beträgt daher

$$(2.25) \quad P(E_n) = \sum_{i=0}^n (-1)^i \frac{1}{i!}.$$

Als populäre Einkleidung dieses Beispiels findet man oft: n Teilnehmer einer Party bringen je ein Geschenk mit. Die Geschenke werden durch Los auf die Gäste verteilt. Wie groß ist die Wahrscheinlichkeit, daß mindestens ein Teilnehmer sein eigenes Geschenk zurück erhält.

5.3. Siebformel von Sylvester Poincaré.

PROPOSITION 2.11. *Es seien A_1, \dots, A_n Teilmengen von Ω (Ω kann auch eine unendliche Menge sein). Dann ist die Wahrscheinlichkeit, daß mindestens eines der Ereignisse A_i , $i = 1, \dots, n$ eintritt, gegeben durch*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

Bevor wir den Beweis der Siebformel skizzieren, notieren wir zwei Spezialfälle

$$(2.26) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$(2.27) \quad P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Der Nachweis von 2.26 ist identisch zur Berechnung von $h_r(A \cup B)$.

BEWEIS. Der Beweis wird über vollständige Induktion geführt. Wegen (2.26) stimmt die Behauptung für $n = 2$. Die Behauptung sei richtig für ein $n \geq 3$ und A_1, \dots, A_{n+1} seien Teilmengen von Ω . Aus (2.26) folgt

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P(A_1) + P\left(\bigcup_{i=2}^{n+1} A_i\right) - P\left(A_1 \cap \bigcup_{i=2}^{n+1} A_i\right) \\ &= P(A_1) + P\left(\bigcup_{i=2}^{n+1} A_i\right) - P\left(\bigcup_{i=2}^{n+1} (A_1 \cap A_i)\right). \end{aligned}$$

Mit Hilfe der Induktionsvoraussetzung folgt weiter

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P(A_1) + \sum_{k=1}^n (-1)^{k+1} \sum_{2 \leq i_1 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &\quad - \sum_{k=1}^n (-1)^{k+1} \sum_{2 \leq i_1 < \dots < i_k \leq n+1} P(A_1 \cap A_{i_1} \cap \dots \cap A_{i_k}) \\ &= P(A_1) + \sum_{2 \leq i_1 \leq n+1} P(A_{i_1}) + \sum_{k=2}^n (-1)^{k+1} \sum_{2 \leq i_1 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &\quad - \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{2 \leq i_1 < \dots < i_k \leq n+1} P(A_1 \cap A_{i_1} \cap \dots \cap A_{i_k}) - (-1)^{n+1} P(A_1 \cap \dots \cap A_{n+1}) \\ &= \sum_{k=1}^{n+1} P(A_k) + \sum_{k=2}^n (-1)^{k+1} \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n+1 \\ i_1 > 1}} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned}$$

$$\begin{aligned}
& + \sum_{k=2}^n (-1)^{k+1} \sum_{\substack{1 \leq i_1 < \dots < i_k \leq n+1 \\ i_1=1}} P(A_{i_1} \cap \dots \cap A_{i_k}) + (-1)^{n+2} P(A_1 \cap \dots \cap A_{n+1}) \\
& = \sum_{k=1}^{n+1} (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_k}).
\end{aligned}$$

□

BEISPIEL 2.11. Wie groß ist die Wahrscheinlichkeit, daß beim Werfen mit 3 Würfeln mindestens eine Sechs fällt?

Es bezeichne A_i das Ereignis, eine Sechs fällt beim i -ten Wurf. Dann bedeutet $A_i \cap A_j$, daß beim i -ten und j -ten Wurf eine Sechs gefallen ist. Nach der Siebformel folgt

$$P\left(\bigcup_{i=1}^3 A_i\right) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} - \frac{1}{36} - \frac{1}{36} - \frac{1}{36} + \frac{1}{216} \approx 0,42.$$

Als weiteres nichttriviales Beispiel betrachten wir folgendes Rencontre Problem:

BEISPIEL 2.12. Ein Ball wird von n Ehepaaren besucht. Für die Mitternachts-einlage werden die Tanzpartner ausgelost. Wie groß ist die Wahrscheinlichkeit, daß dabei mindestens ein Ehepaar zusammen tanzt.

Numerieren wir die Ehepaare durch, dann entspricht jeder Auslosung eine Permutation π der Zahlen $\{1, \dots, n\}$. Das i -te Ehepaar trifft bei der Auslosung zusammen, falls $\pi(i) = i$ ist. Das Spiel kann als Laplace-Experiment auf Ω , der Menge aller Permutationen von $\{1, \dots, n\}$, beschrieben werden. Wir interessieren uns für das Ereignis $\bigcup_{i=1}^n A_i$ mit

$$A_i = \{\pi \in \Omega : \pi(i) = i\}, \quad i = 1, \dots, n.$$

Um die Siebformel anwenden zu können benötigen wir noch $P(A_{i_1} \cap \dots \cap A_{i_k})$ für ein beliebiges k -Tupel (i_1, \dots, i_k) . Das Ereignis $A_{i_1} \cap \dots \cap A_{i_k}$ wird genau durch jene Permutationen realisiert, für welche

$$\pi(i_j) = i_j, \quad j = 1, \dots, k$$

gilt. Daher ist $|A_{i_1} \cap \dots \cap A_{i_k}| = (n - k)!$, also

$$(2.28) \quad P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n - k)!}{n!}.$$

Insgesamt gibt es $\binom{n}{k}$ Permutationen mit genau k Fixpunkten, welche alle dieselbe Wahrscheinlichkeit (2.28) besitzen. Die Siebformel ergibt nun

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{(n - k)!}{n!} = \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!}.$$

Da die Exponentialreihe sehr rasch konvergiert, ergibt sich bereits für moderate Werte von n die überraschend große Wahrscheinlichkeit $P(\bigcup_{i=1}^n A_i) \approx 1 - e^{-1} \approx 0,63$. Man vergleiche diesen Lösungsansatz mit jenem in Beispiel 2.10.

6. Bedingte Wahrscheinlichkeit, Stochastische Unabhängigkeit

Die Wahrscheinlichkeit, mit einem vollkommen symmetrischen Würfel eine Sechs zu würfeln (Ereignis A), beträgt $P(A) = \frac{1}{6}$. Werfen wir den Würfel noch einmal, ohne auf die Augenzahl zu achten. Bekommen wir zusätzlich die Information, eine gerade Zahl sei gefallen (Ereignis B), erhöht sich die Wahrscheinlichkeit, daß es eine Sechs ist, auf $\frac{1}{3}$. Man schreibt: $P(A|B) = \frac{1}{3}$ und meint die Wahrscheinlichkeit des Eintreffens von A unter der Voraussetzung, daß das Ereignis B bereits eingetreten ist.

BEISPIEL 2.13. Bei der Volkszählung 1981 ergaben sich die in Tabelle 2.1 angegebenen Anzahlen für die Österreicher unter bzw. über 20 Jahre: Betrachten wir die Ereignisse W , eine zufällig ausgewählte Person ist weiblich, und U , eine zufällig ausgewählte Person ist höchstens 20 Jahre alt. Dann gilt

$$P(W) = \frac{3.975.122}{7.555.338} = 52,6\%, \quad P(U) = \frac{2.165.393}{7.555.338} = 28,7\%.$$

Die Wahrscheinlichkeit, daß eine zufällig ausgewählte Frau höchstens 20 Jahre alt ist, ergibt sich zu

$$P(U|W) = \frac{1.057.605}{3.975.122} = 26,6\%$$

Das Ereignis $U|W$ ist zu unterscheiden vom Ereignis $U \cap W$, welches genau dann eintritt, wenn die zufällig gewählte Person eine Frau und nicht älter als 20 Jahre ist. Allerdings besteht ein enger Zusammenhang zwischen $P(U|W)$ und $P(U \cap W) = \frac{1.057.605}{7.555.338}$.

$$P(U|W) = \frac{\frac{1.057.605}{7.555.338}}{\frac{3.975.122}{7.555.338}} = \frac{P(U \cap W)}{P(W)}.$$

	Unter 20 J.	Über 20 Jahre	Insgesamt
Frauen	1.057.605	2.917.517	3.975.122
Männer	1.107.788	2.472.428	3.580.216
Insgesamt	2.165.393	5.389.945	7.555.338

TABELLE 2.1

DEFINITION 2.7. *Es sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $B \in \mathcal{E}$ ein festes Ereignis mit $P(B) > 0$. Dann heißt*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{E}$$

bedingte Wahrscheinlichkeit von A unter B.

PROPOSITION 2.12. *Es sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $B \in \mathcal{E}$ ein festes Ereignis mit $P(B) > 0$. Dann ist die bedingte Wahrscheinlichkeit $P(\cdot|B): \mathcal{E} \rightarrow \mathbb{R}$ ein Wahrscheinlichkeitsmaß auf Ω .*

BEWEIS. Mit $A, B \in \mathcal{E}$ folgt auch $A \cap B \in \mathcal{E}$. Die Eigenschaften $P(A|B) \geq 0$ und $P(\Omega|B) = 1$ sind unmittelbar einsichtig. Um die σ -Additivität von $P(\cdot|B)$ nachzuweisen, betrachten wir eine Folge paarweiser disjunkter Ereignisse $(A_i) \subset \mathcal{E}$. Dann sind auch die Ereignisse $A_i \cap B, i \in \mathbb{N}$ paarweise disjunkt. Somit folgt

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i|B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B). \end{aligned}$$

□

Als nützliche Folgerung notieren wir die bedingte Wahrscheinlichkeit des Gegenereignisses

$$(2.29) \quad P(\bar{A}|B) = 1 - P(A|B), \quad A \in \mathcal{E}.$$

In folgender Interpretation der bedingten Wahrscheinlichkeit wird die zusätzliche Information, das Ereignis B sei eingetreten, dadurch berücksichtigt, daß der Ereignisraum Ω durch B ersetzt wird: durch die Zusatzinformation wissen wir ja, daß nur mehr die Elementarereignisse, welche B realisieren, möglich sind. Unter dieser Voraussetzung sind nur mehr die Ereignisse in $\mathcal{E}_B = \{A \cap B: A \in \mathcal{E}\}$ sinnvoll. Es ist nicht schwer zu zeigen, daß auch \mathcal{E}_B eine σ -Algebra darstellt. Ferner gilt

$$\mathcal{E}_B = \{A \in \mathcal{E}: A \subset B\},$$

\mathcal{E}_B ist also auch eine σ -Algebra in B . Setzt man nun noch

$$P_B(A) = P(A|B), \quad A \in \mathcal{E}_B$$

zeigt man wie in Proposition 2.12, daß (B, \mathcal{E}_B, P_B) einen Wahrscheinlichkeitsraum auf B definiert. Diese Betrachtungsweise entspricht dem intuitiven Vorgehen in den beiden motivierenden Beispielen.

In manchen Anwendungen ist es relativ leicht, sich die bedingte Wahrscheinlichkeit $P(A|B)$ oder $P(B|A)$ zu überlegen. Dann kann man die Wahrscheinlichkeit

des Durchschnittes berechnen aus

$$(2.30) \quad P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \quad A, B \in \mathcal{E}.$$

BEISPIEL 2.14. In einem Schaltkreis sind zwei Relais parallel geschaltet, sodaß der gesamte Schaltkreis genau dann unterbrochen wird, wenn beide Relais ausfallen. Die Wahrscheinlichkeit, daß Relais 1 ausfällt ist 10%, wenn Relais 1 ausgefallen ist, fällt Relais 2 mit einer Wahrscheinlichkeit von 5% aus. Mit welcher Wahrscheinlichkeit fällt der gesamte Schaltkreis aus?

Der Ausfall des i -ten Relais, $i = 1, 2$, entspricht dem Ereignis R_i . Dann ist $P(R_1) = 0,1$ und $P(R_2|R_1) = 0,05$. Aus (2.30) folgt $P(R_1 \cap R_2) = P(R_2|R_1)P(R_1) = 0,05 \cdot 0,1 = 0,005$.

Tabelle 2.1 ist ein Beispiel einer Vierfeldertafel, einem nützlichen Hilfsmittel zum Bestimmen verschiedener Wahrscheinlichkeiten, welche im Zusammenhang mit 2 Ereignissen auftreten. Eine Vierfeldertafel für die Ereignisse A und B ist folgendermaßen aufgebaut.

	B	\bar{B}	
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
	$P(B)$	$P(\bar{B})$	1

TABELLE 2.2. Vierfeldertafel

Als Übung überlege man sich die Zusammenhänge zwischen den Eintragungen in die Vierfeldertafel. Man beachte, daß die Verhältnisse der Innenfelder zu den Feldern der Randzeile (Randspalte) bedingte Wahrscheinlichkeiten darstellen.

In einer Reihe von Anwendungen gibt es mehrere Alternativen A_i , unter denen ein bestimmtes Ereignis E eintreten kann und man kennt die bedingten Wahrscheinlichkeiten $P(E|A_i)$ und die Wahrscheinlichkeiten für das Eintreten der Alternativen $P(A_i)$. Dann kann man die unbedingte (totale) Wahrscheinlichkeit von A folgendermaßen ermitteln:

THEOREM 2.5 (Satz von der totalen Wahrscheinlichkeit). *Es seien (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $(A_i)_{i \in \mathbb{N}} \subset \mathcal{E}$ eine Folge von paarweise disjunkten Ereignissen mit $P(A_i) > 0$, $i \in \mathbb{N}$. Für jedes Ereignis $E \in \mathcal{E}$ mit $E \subset \bigcup_{i=1}^{\infty} A_i$ gilt dann*

$$(2.31) \quad P(E) = \sum_{i=1}^{\infty} P(E|A_i)P(A_i).$$

Bei Anwendungen dieses Satzes ist häufig $\bigcup_{i=1}^{\infty} A_i = \Omega$.

BEWEIS. Das Ereignis E erfülle die Voraussetzungen des Satzes. Wegen der paarweisen Disjunktheit von A_i sind auch die Durchschnitte $E \cap A_i$ paarweise disjunkt. Aus

$$E = \bigcup_{i=1}^{\infty} (E \cap A_i)$$

folgt mit der σ -Additivität von P und (2.30)

$$P(E) = \sum_{i=1}^{\infty} P(E \cap A_i) = \sum_{i=1}^{\infty} P(E|A_i)P(A_i).$$

□

BEISPIEL 2.15. Ein "zuverlässiger" Test für die Diagnose von TBC führt in 94% aller Fälle, in denen die Testperson nicht erkrankt ist, zu einem negativen Ergebnis; er ist positiv in 96% aller Fälle, in denen der Proband tatsächlich an TBC leidet. In einer bestimmten Zielgruppe beträgt die Wahrscheinlichkeit, an TBC erkrankt zu sein, 1:145. Wie groß ist die Wahrscheinlichkeit, daß eine Person aus dieser Zielgruppe tatsächlich das Virus in sich trägt, wenn der Test positiv ausgefallen ist.

Für die Ereignisse

- N Test fällt negativ aus
- T Testperson ist an TBC erkrankt

gilt dann

$$P(N|\bar{T}) = 0,94, \quad P(\bar{N}|T) = 0,96, \quad P(T) = \frac{1}{145}, \quad P(\bar{T}) = \frac{144}{145}.$$

Zu bestimmen ist

$$P(T|\bar{N}) = \frac{P(T \cap \bar{N})}{P(\bar{N})}.$$

Mit Hilfe von (2.30) findet man

$$P(T \cap \bar{N}) = P(\bar{N}|T)P(T) = 0,96 \cdot \frac{1}{145} = 0,0066.$$

Die Wahrscheinlichkeit von \bar{N} berechnen wir mit (2.31) und (2.29):

$$\begin{aligned} P(\bar{N}) &= P(\bar{N}|T)P(T) + P(\bar{N}|\bar{T})P(\bar{T}) = P(\bar{N}|T)P(T) + (1 - P(N|\bar{T}))P(\bar{T}) \\ &= 0,96 \cdot \frac{1}{145} + 0,06 \cdot \frac{144}{145} = 0,0652, \end{aligned}$$

also $P(T|\bar{N}) = 0,10$.

Der Vollständigkeit halber sei erwähnt, daß $\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$ einen möglichen Ereignisraum für dieses Zufallsexperiment darstellt. Dabei bedeutet eine 1 bzw. 0 an der 1. Stelle des geordneten Paares, daß die Testperson TBC hat bzw. nicht

hat, eine 1 bzw. 0 als 2. Koordinate, daß der Test positiv bzw. negativ ausgefallen ist. Es liegt allerdings kein Laplace Experiment vor.

THEOREM 2.6 (Formel von Bayes). *Es seien (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $(A_i)_{i \in \mathbb{N}} \subset \mathcal{E}$ eine Folge von paarweise disjunkten Ereignissen mit $P(A_i) > 0$, $i \in \mathbb{N}$. Für jedes Ereignis $E \in \mathcal{E}$ mit $E \subset \bigcup_{i=1}^{\infty} A_i$ und $P(E) > 0$ gilt dann*

$$(2.32) \quad P(A_k|E) = \frac{P(E|A_k)P(A_k)}{\sum_{i=1}^{\infty} P(E|A_i)P(A_i)}, \quad k \in \mathbb{N}.$$

BEWEIS. Nach (2.30) gilt

$$P(A_k|E) = \frac{P(A_k \cap E)}{P(E)} = \frac{P(E|A_k)P(A_k)}{P(E)}.$$

Stellt man $P(E)$ mit Hilfe des Satzes von der totalen Wahrscheinlichkeit 2.5 dar, folgt die Behauptung. \square

BEISPIEL 2.16 (Gestörter Nachrichtenkanal). Bei der Übertragung der Zeichen "Punkt" und "Strich" in einem Fernmeldesystem werden durch Störungen im Mittel 5% der gesendeten Punkte als Striche und 3% der gesendeten Striche als Punkte empfangen. Das Verhältnis von gesendeten Punkten zu Strichen ist $p = \frac{3}{5}$. Wie groß ist die Wahrscheinlichkeit, daß das richtige Zeichen empfangen wurde, falls "Punkt" empfangen wurde.

Wir identifizieren das Symbol "Strich" mit "1" und "Punkt" mit "0" und wählen als Ereignisraum dieses Zufallsexperimentes $\Omega = \{0, 1\}^2$. Die 1. Koordinate stehe für das gesendete, die zweite für das empfangene Signal. Wir betrachten die Ereignisse

S_i es wird i gesendet
 E_i es wird i empfangen,

$i = 0, 1$. Die Ereignisse $S_0 = \{(0, 0), (0, 1)\}$ und $S_1 = \{(1, 0), (1, 1)\}$ bilden eine disjunkte Partition von Ω . Weiters wissen wir

$$P(E_1|S_0) = 0,05, \quad P(E_0|S_1) = 0,03.$$

Setzt man schließlich noch

$$p = \frac{P(S_0)}{P(S_1)} = \frac{3}{5}$$

folgt aus der Bayes'schen Regel für die gesuchte Wahrscheinlichkeit

$$\begin{aligned} P(S_0|E_0) &= \frac{P(E_0|S_0)P(S_0)}{P(E_0|S_0)P(S_0) + P(E_0|S_1)P(S_1)} \\ &= \left(1 + \frac{P(E_0|S_1)P(S_1)}{P(E_0|S_0)P(S_0)}\right)^{-1} = \left(1 + \frac{1}{p} \frac{0,03}{1 - 0,05}\right)^{-1} = 0,95. \end{aligned}$$

Im letzten Schritt wurde $P(E_0|S_0) = 1 - P(E_1|S_0)$ verwendet.

THEOREM 2.7 (Multiplikationssatz). *Es sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und A_0, \dots, A_n seien Ereignisse mit $P(A_0 \cap \dots \cap A_n) > 0$. Dann gilt*

$$(2.33) \quad P\left(\bigcap_{i=0}^n A_i\right) = P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_n|A_0 \cap \dots \cap A_{n-1}).$$

BEWEIS. Wegen der Abgeschlossenheit einer σ -Algebra gegenüber der Bildung von abzählbar vielen Durchschnitten, und wegen $0 < P(A_0 \cap \dots \cap A_n) \leq P(A_0 \cap \dots \cap A_{n-1}) \leq \dots \leq P(A_0)$ sind die Ereignisse in (2.33) und damit die auftretenden bedingten Wahrscheinlichkeiten wohldefiniert. Der Beweis wird durch Induktion nach n geführt: Für $n = 1$ folgt die Gültigkeit der Behauptung aus (2.30). Ist die Behauptung richtig für n , dann folgt

$$\begin{aligned} P\left(\bigcap_{i=0}^{n+1} A_i\right) &= P\left(\bigcap_{i=0}^n A_i\right)P(A_{n+1}|A_0 \cap \dots \cap A_n) \\ &= P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_{n+1}|A_0 \cap \dots \cap A_n) \end{aligned}$$

□

Der Multiplikationssatz ist die theoretische Grundlage für die häufig verwendete Technik, Wahrscheinlichkeiten an Hand eines Wahrscheinlichkeitsbaumes zu berechnen

BEISPIEL 2.17. In einer Lade seien 7 braune und 5 graue Socken. Wie groß ist die Wahrscheinlichkeit, bei willkürlicher Auswahl ein farbgleiches Paar zu ziehen. Wir betrachten die Ereignisse

- B_i eine braune Socke wird beim i -ten Zug gezogen
- G_i eine graue Socke wird beim i -ten Zug gezogen

$i = 1, 2$. Von Interesse ist das Ereignis $E = (B_1 \cap B_2) \cup (G_1 \cap G_2)$. Aus dem Multiplikationssatz folgt

$$P(B_1 \cap B_2) = P(B_2|B_1)P(B_1), \quad P(G_1 \cap G_2) = P(G_2|G_1)P(G_1).$$

Die erforderlichen Wahrscheinlichkeiten überlegt man sich am leichtesten mit dem Wahrscheinlichkeitsbaum aus Abbildung 2.5. Der allgemeine Zusammenhang zwischen einem Wahrscheinlichkeitsbaum und Satz 2.7 wird in Abbildung 2.6 angedeutet.

Wenn wir eine faire Münze 2-mal werfen und beim 1. Wurf Zahl fällt, dann hat dieses Ergebnis keinen Einfluß auf das Ausfallen des 2. Wurfes. Dies ist ja gerade die Voraussetzung für ein Laplace Experiment. Bezeichnet $Z_i, i = 1, 2$ das Ereignis “Zahl fällt beim i -ten Wurf”, dann gilt $P(Z_2|Z_1) = \frac{1}{2} = P(Z_2)$. Man sagt, die Ereignisse Z_1 und Z_2 sind stochastisch unabhängig. Allgemeiner definiert man:

DEFINITION 2.8. *Eine höchstens abzählbar unendliche Familie von Ereignissen $\{A_i: i \in I\}$, $I \subset \mathbb{N}$, in einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) heißt **stochastisch***

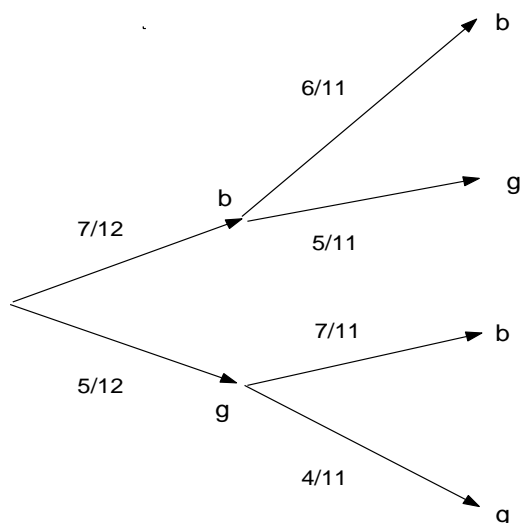


ABB. 2.5

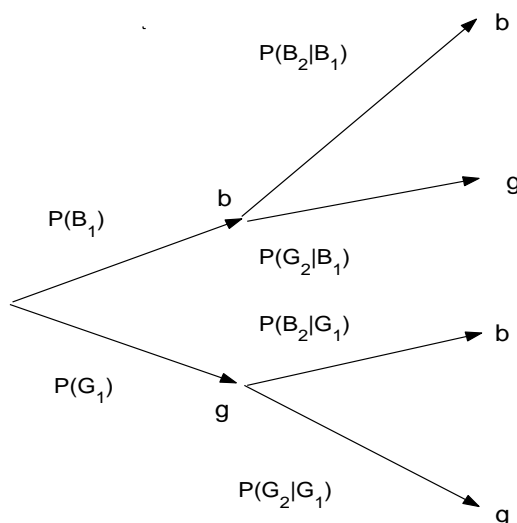


ABB. 2.6

unabhängig, wenn für je endlich viele paarweise verschieden Indizes $i_1, \dots, i_n \in I$ gilt

$$(2.34) \quad P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_n}).$$

PROPOSITION 2.13. *Es seien A und B stochastisch unabhängige Ereignisse in (Ω, \mathcal{E}, P) . Dann sind folgende Aussagen äquivalent:*

- $P(A \cap B) = P(A)P(B)$,
- $P(A|B) = P(A)$, $P(B|A) = P(B)$,
- A und \bar{B} sind stochastisch unabhängig.

Ohne Beweis zitieren wir eine etwas einfachere Charakterisierung der stochastischen Unabhängigkeit:

PROPOSITION 2.14. *Die Ereignisse A_1, \dots, A_n in einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) sind stochastisch unabhängig genau dann, wenn*

$$(2.35) \quad P(B_1 \cap \dots \cap B_n) = \prod_{i=1}^n P(B_i)$$

für jede Wahl von Mengen $B_i \in \{A_i, \bar{A}_i\}$, $1 \leq i \leq n$, gilt.

Für unabhängige Ereignisse vereinfacht sich die Sylvestersche Siebformel:

PROPOSITION 2.15. *Es seien A_1, \dots, A_n stochastisch unabhängige Ereignisse in einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) . Dann gilt*

$$(2.36) \quad P\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n (1 - P(A_i)).$$

BEWEIS. Es ist einfacher, sich (2.36) direkt zu überlegen als aus Proposition 2.11 abzuleiten:

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\overline{\bigcup_{i=1}^n A_i}\right) = 1 - P\left(\bigcap_{i=1}^n \bar{A}_i\right) = 1 - \prod_{i=1}^n (1 - P(A_i)),$$

im letzten Schritt wurde von die stochastische Unabhängigkeit der Ereignisse $\bar{A}_1, \dots, \bar{A}_n$ verwendet. \square

7. Diskrete Zufallsvariable

Bisher haben wir uns mit Zufallsexperimenten beschäftigt und verschiedene Möglichkeiten gesehen, die Wahrscheinlichkeit eines Ereignisses zu bestimmen. Oft ist man aber nicht unmittelbar an der Wahrscheinlichkeit eines Ereignisses aus jenem Ereignisraum interessiert, in welchem das Zufallsexperiment durchgeführt wird, sondern an der Wahrscheinlichkeit einer vom Ausgang des Experimentes abhängigen Größe:

BEISPIEL 2.18. Das ursprüngliche Zufallsexperiment sei das Werfen von 2 homogenen Würfeln. Dies ist ein Laplace Experiment in $\Omega = \{\omega = (i, j) : i, j \in \{1, \dots, 6\}\}$. Für jeden Wurf $\omega = (i, j)$ wird die Augensumme $X(\omega) = i + j$ berechnet. Wir interessieren uns für die Wahrscheinlichkeit des Auftretens einer bestimmten Augensumme, also für $P(\{\omega : X(\omega) = k\})$, $2 \leq k \leq 12$. Vereinfacht schreiben wir diese Wahrscheinlichkeit $P(X = k)$. Aus Symmetriegründen gilt $P(X = k) = P(X = 12 - k + 2)$, $k = 2, \dots, 7$. Eine einfache Überlegung zeigt nun

$$P(X = k) = \frac{k-1}{36}, \quad k = 2, \dots, 7.$$

Wegen

$$\sum_{k=2}^{12} P(X = k) = 1$$

definiert die Funktion $P_X : \{2, \dots, 12\} \rightarrow [0, 1]$, $P_X(k) = P(X = k)$ ein Wahrscheinlichkeitsmaß auf $\{2, \dots, 12\}$.

Dieses Beispiel ist ein Spezialfall folgender allgemeinerer Situation: Gegeben sei ein Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) , eine Menge $\tilde{\Omega}$, eine σ -Algebra $\tilde{\mathcal{E}}$ in $\tilde{\Omega}$ und eine meßbare Abbildung $X : (\Omega, \mathcal{E}) \rightarrow (\tilde{\Omega}, \tilde{\mathcal{E}})$. Definiert man die Mengenfunktion

$$(2.37) \quad P_X : \begin{cases} \tilde{\mathcal{E}} \rightarrow [0, 1] \\ P_X(B) = P(X^{-1}(B)) \end{cases}$$

dann ist $(\tilde{\Omega}, \tilde{\mathcal{E}}, P_X)$ ein Wahrscheinlichkeitsraum. Offensichtlich gilt $P_X(B) \geq 0$ und $P_X(\tilde{\Omega}) = 1$ ($X^{-1}(\tilde{\Omega}) = \Omega$). Es seien also B_i , $i \in \mathbb{N}$ paarweise disjunkte Mengen aus $\tilde{\mathcal{E}}$. Wegen der Meßbarkeit von X sind die Urbilder $X^{-1}(B_i)$ Mengen in \mathcal{E} . Da

$\omega \in X^{-1}(B_i) \cap X^{-1}(B_j)$ äquivalent ist zu $X(\omega) \in B_i \cap B_j$, sind die Ereignisse $X^{-1}(B_i)$ ebenfalls paarweise disjunkt. Somit überträgt sich die σ -Additivität von P auf P_X :

$$\begin{aligned} P_X\left(\bigcup_{i=1}^{\infty} B_i\right) &= P\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right) \\ &= \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_X(B_i). \end{aligned}$$

DEFINITION 2.9. *Das durch (2.37) definierte Wahrscheinlichkeitsmaß heißt induziertes Wahrscheinlichkeitsmaß oder Wahrscheinlichkeitsverteilung von X . Im Falle $\tilde{\Omega} = \mathbb{R}^n$ und $\tilde{\mathcal{E}} = \mathcal{B}^n$, nennt man die induzierende Abbildung $X: (\Omega, \mathcal{E}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ Zufallsvariable. Ist speziell $\tilde{\Omega} = \mathbb{N}$ liegt eine diskrete Zufallsvariable vor.*

Wir haben bereits gesehen, daß eine Wahrscheinlichkeitsverteilung auch durch ihre Verteilungsfunktion beschrieben werden kann. Es sei $F_X: \mathbb{R} \rightarrow \mathbb{R}$ die Verteilungsfunktion von P_X , also

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}((-\infty, x])).$$

Etwas schlampiger nennt man F_X auch Verteilungsfunktion der Zufallsvariablen X . Wenn keine Verwechslungen möglich sind, werden wir F anstelle von F_X schreiben. In diesem Abschnitt werden wir ausschließlich diskrete Zufallsvariable betrachten. In diesem Falle ergibt sich für die Verteilungsfunktion

$$(2.38) \quad F_X(x) = \sum_{\substack{k \leq x \\ k \in \mathbb{N}}} P(X = k).$$

Der Einfachheit halber schreiben wir $P(X = k)$, $P(X < k)$ etc. für $P(X^{-1}(\{k\}))$, $P(X^{-1}((-\infty, k)))$ etc.

BEISPIEL 2.19. Man bietet Ihnen folgende Wette an: Es soll n -mal gewürfelt werden. Fällt eine 1 oder 2, gewinnen Sie S 300, andernfalls verlieren Sie S 100. Sollten Sie auf diese Wette einsteigen?

Bevor Sie sich entscheiden, sollten Sie folgende Überlegung anstellen: bei einer Serie von n Spielen sollte etwa bei einem Drittel der Würfe 1 oder 2 gefallen sein, bei welchen Sie also S 300 gewinnen. Ihre Gewinnerwartung bei einer derartigen Serie liegt also bei

$$300 \frac{n}{3} - 100 \frac{2n}{3} = 100 \frac{n}{3},$$

pro Spiel können Sie also im Durchschnitt mit einem Gewinn von

$$300 \frac{1}{3} - 100 \frac{2}{3} = 100 \frac{1}{3}$$

rechnen.

In diesem Beispiel ist der Ausgangspunkt das Laplace Experiment in $\Omega = \{1, \dots, 6\}$. Abhängig vom zufälligen Ergebnis ω dieses Experimentes wird ein Gewinn

$$X(\omega) = \begin{cases} 300 & \omega \in \{1, 2\} \\ -100 & \omega \in \{3, 4, 5, 6\} \end{cases}$$

ausbezahlt (ein negativer Gewinn bedeutet natürlich eine Zahlung, welche Sie leisten müssen). Der Gewinn ist also eine Zufallsvariable $X: \Omega \rightarrow \{300, -100\}$. Die Gewinnerwartung $E(X)$ pro Spiel setzen wir intuitiv an als

$$E(X) = 300P(X = 300) - 100P(X = -100).$$

Allgemeiner definiert man:

DEFINITION 2.10. *Es sei X eine diskrete Zufallsvariable auf (Ω, \mathcal{E}, P) . Ist die Reihe*

$$(2.39) \quad E(X) = \sum_{x \in X(\Omega)} xP(X = x) = \sum_{x \in X(\Omega)} xP_X(\{x\})$$

absolut konvergent, so heißt ihr Wert Erwartungswert oder die Erwartung von X .

Bei der Berechnung des Erwartungswertes von $X: \Omega \rightarrow \mathbb{N}$ nach (2.39) hat man also die möglichen Werte von X mit der Wahrscheinlichkeit ihres Auftretens zu multiplizieren. Manchmal ist es zweckmäßig, $E(X)$ aus den Wahrscheinlichkeiten der Elementarereignisse in Ω zu bestimmen.

PROPOSITION 2.16. *Es sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{N}$ eine diskrete Zufallsvariable. Der Erwartungswert von X existiert genau dann, wenn die Reihe*

$$(2.40) \quad E(X) = \sum_{\omega \in \Omega} X(\omega)p_\omega,$$

$p_\omega = P(\{\omega\})$, absolut konvergent ist.

BEWEIS. Existiert der Erwartungswert von X , folgt aus der absoluten Konvergenz von (2.39)

$$\begin{aligned} E(X) &= \sum_{x \in X(\Omega)} xP(X = x) = \sum_{x \in X(\Omega)} x \sum_{\omega: X(\omega)=x} p_\omega = \sum_{x \in X(\Omega)} \sum_{\omega: X(\omega)=x} xp_\omega \\ &= \sum_{x \in X(\Omega)} \sum_{\omega: X(\omega)=x} X(\omega)p_\omega = \sum_{x \in X(\Omega)} \sum_{\omega: X^{-1}(\{x\})} X(\omega)p_\omega = \sum_{\omega \in \Omega} X(\omega)p_\omega \end{aligned}$$

Die letzte Gleichheit folgt aus dem Umstand, daß

$$\Omega = \bigcup_{x \in X(\Omega)} \{\omega: X(\omega) = x\} = \bigcup_{x \in X(\Omega)} X^{-1}(\{x\})$$

eine disjunkte Partition von Ω bildet. Mit dem Umordnungssatz für absolut konvergent Reihen erhält man nun die Behauptung. Die Umkehrung beweist man vollkommen analog. \square

Mit Hilfe dieser Proposition kann man also den erwarteten Gewinn aus Beispiel 2.19 auch folgendermaßen berechnen:

$$\begin{aligned} E(X) &= X(1)\frac{1}{6} + X(2)\frac{1}{6} + X(3)\frac{1}{6} + X(4)\frac{1}{6} + X(5)\frac{1}{6} + X(6)\frac{1}{6} \\ &= (300 + 300 - 100 - 100 - 100 - 100)\frac{1}{6} = \frac{100}{3}. \end{aligned}$$

Eine unmittelbare Folgerung aus der Definition bzw. einfacher aus Proposition 2.16 sind folgende Eigenschaften des Erwartungswertes einer (diskreten) Zufallsvariablen:

PROPOSITION 2.17. *Es seien X und Y diskrete Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) mit Erwartungswert $E(X)$ und $E(Y)$. Dann gilt*

1. *Es existieren die Erwartungswerte von $X + Y$ und αX , $\alpha \in \mathbb{R}$, und es gilt*

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ E(\alpha X) &= \alpha E(X) \end{aligned}$$

Der Erwartungswert ist also ein lineares Funktional.

2. *Der Erwartungswert ist monoton:*

$$X \leq Y \Rightarrow E(X) \leq E(Y)$$

3. *Gilt $E(X) = E(Y)$ und $X \leq Y$, dann folgt $P(X = Y) = 1$.*
4. *Für konstante Zufallsvariable $X \equiv \alpha$ gilt*

$$E(\alpha) = \alpha.$$

Wir werden später sehen, daß der Erwartungswert einer Zufallsvariablen X , dem Grenzwert des Mittelwertes von X bei oftmaligem Wiederholen des Zufallsexperimentes entspricht. Jede Wiederholung ergibt einen Wert $X(\omega)$, welcher in unregelmäßiger Weise um $E(X)$ schwankt. In Anlehnung an die Diskussion der Varianz von metrischen Daten legen wir folgendes Maß für die Fluktuationen von $X(\omega)$ fest:

DEFINITION 2.11. *Es sei X eine diskrete Zufallsvariable auf (Ω, \mathcal{E}, P) . Konvergiert die Reihe*

$$(2.41) \quad V(X) = E((X - E(X))^2) = \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 p_\omega$$

*dann heißt $V(X)$ **Varianz von X** . Die positive Quadratwurzel aus der Varianz von X heißt **Standardabweichung oder Streuung von X** , $\sigma(X)$.*

Verwendet man die Eigenschaften des Erwartungswertes aus Proposition 2.17 ergibt sich für die Varianz einer Zufallsvariablen

$$V(X) = E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - 2E(X)E(X) + E(X)^2,$$

also

$$(2.42) \quad V(X) = E(X^2) - E(X)^2.$$

Die Berechnung von $E(X^2)$ wird durch folgende Formel erleichtert:

PROPOSITION 2.18. *Es sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum, $X: \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable und $f: \mathbb{R} \rightarrow \mathbb{R}$ eine Abbildung, für welche $E(f \circ X)$ existiert. Dann gilt*

$$(2.43) \quad E(f \circ X) = \sum_{x \in X(\Omega)} f(x)P_X(\{x\}) = \sum_{\omega \in \Omega} f(X(\omega))p_\omega.$$

BEWEIS. Wir gehen ähnlich wie im Beweis von Proposition 2.16 vor:

$$\begin{aligned} E(f \circ X) &= \sum_{t \in f \circ X(\Omega)} tP_{f \circ X}(\{t\}) = \sum_{t \in f \circ X(\Omega)} tP(X^{-1}(f^{-1}(\{t\}))) = \sum_{t \in f \circ X(\Omega)} tP_X(f^{-1}(\{t\})) \\ &= \sum_{t \in f \circ X(\Omega)} \sum_{x \in f^{-1}(\{t\})} tP_X(\{x\}) = \sum_{t \in f \circ X(\Omega)} \sum_{x: f(x)=t} tP_X(\{x\}) \\ &= \sum_{t \in f \circ X(\Omega)} \sum_{x: f(x)=t} f(x)P_X(\{x\}) = \sum_{x \in X(\Omega)} f(x)P_X(\{x\}) \end{aligned}$$

Die letzte Gleichheit folgt aus der absoluten Konvergenz der Reihe für $E(f \circ X)$. Der Rest der Behauptung folgt wie in Proposition 2.16. \square

8. Spezielle diskrete Verteilungen

8.1. Bernoulli- oder Binomialverteilung.

BEISPIEL 2.20. In einer Charge von 500 Sicherungen sind 5% defekt. Wie groß ist die Wahrscheinlichkeit, daß in einer Stichprobe von 5 (zufällig gewählten) Sicherungen mindestens 1 Stück defekt ist.

Bei jeder Entnahme wird die Sicherung getestet und festgehalten, ob sie defekt ist (Ereignis A tritt ein) oder nicht (Gegenereignis \bar{A} tritt ein). Da der Umfang der Stichprobe im Verhältnis zum Umfang der Charge klein ist, kann man annehmen, daß die Wahrscheinlichkeit des Eintreffens von A durch die Entnahme von einigen wenigen Sicherungen praktisch nicht verändert wird. Von Interesse ist die Häufigkeit des Eintreffens von A in der Stichprobe. Ein derartiges Zufallsexperiment heißt **Bernoulli Experiment**. Die charakteristischen Eigenschaften eines Bernoulli Experimentes sind allgemein

- Es besteht aus n Wiederholungen desselben Experimentes
- Das Experiment hat nur 2 mögliche Ergebnisse: Erfolg E und Mißerfolg \bar{E}

- $P(E) = p$ bzw. $P(\bar{E}) = q = 1 - p$ bei jeder Wiederholung des Experimentes
- Die einzelnen Wiederholungen sind unabhängig
- Von Interesse ist die Zufallsvariable X , die Anzahl der erfolgreichen Experimente unter den n Wiederholungen

Als Ereignisraum für ein Bernoulli Experiment wählen wir $\Omega = \Omega_0^n$, $\Omega_0 = \{1, 0\}$, wobei eine 1 einen Erfolg, eine 0 einen Mißerfolg bedeutet. Definieren wir ferner die Ereignisse

$$(2.44) \quad E_i = \underbrace{\Omega_0 \times \cdots \times \Omega_0}_{(i-1)\text{-mal}} \times \{1\} \times \underbrace{\Omega_0 \times \cdots \times \Omega_0}_{(n-i)\text{-mal}}$$

also "Erfolg bei der i -ten Wiederholung", dann gilt nach Voraussetzung $P(E_i) = p$, und $P(\bar{E}_i) = q$, $i = 1, \dots, n$. Ein typischer Ausgang des Bernoulli Experimentes, $\omega = (\omega_1, \dots, \omega_n)$, bei dem genau k Erfolge und zwar bei den Wiederholungen i_1, \dots, i_k , folglich Mißerfolge bei den Wiederholungen j_1, \dots, j_{n-k} , $j_i \in \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$, aufgetreten sind, kann also aufgefaßt werden als

$$\omega = E_{i_1} \cap \cdots \cap E_{i_k} \cap \bar{E}_{j_1} \cap \cdots \cap \bar{E}_{j_{n-k}}$$

Interpretiert man die experimentelle Unabhängigkeit im mathematischen Modell als stochastische Unabhängigkeit, dann sind die Ereignisse E_i , $i = 1, \dots, n$, stochastisch unabhängig. Aus Proposition 2.14 folgt dann

$$p_\omega = p^k q^{n-k}.$$

Da es $\binom{n}{k}$ verschiedene Bernoulliexperimente mit genau k Erfolgen gibt, erhält man schließlich

$$(2.45) \quad P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, \dots, n.$$

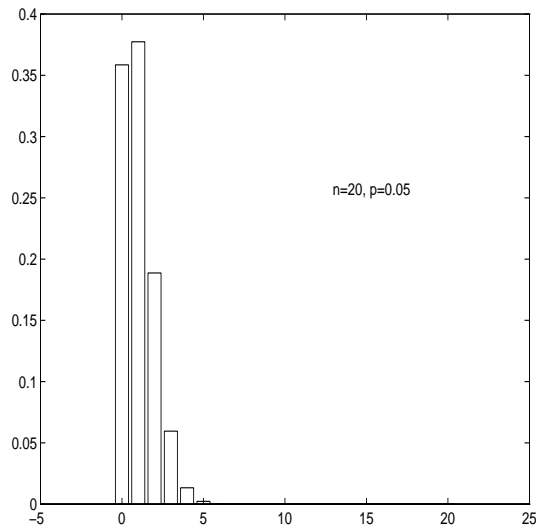
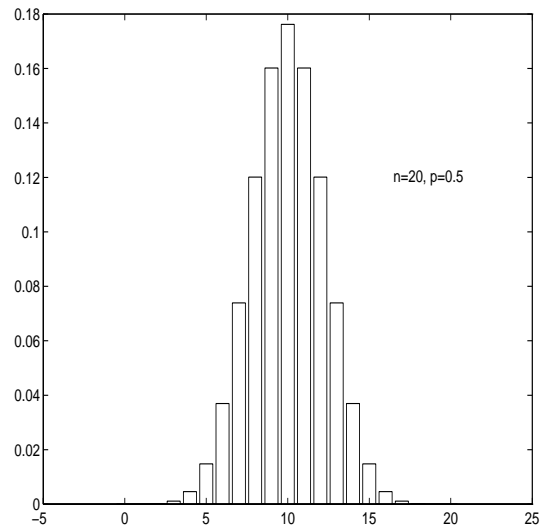
Der binomische Lehrsatz zeigt,

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1,$$

daß P_X ein Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ definiert, vgl Beispiel 2.3. Die Verteilung von P_X nennt man **Binomial-** oder **Bernoulliverteilung**. Man schreibt auch: die Zufallsvariable X ist $B(n; p)$ verteilt.

Die Abbildungen 2.7 und 2.8 zeigen Histogramme verschiedener Binomialverteilungen. Man beachte die großen qualitativen Unterschiede der Histogramme für kleine bzw. große Werte von p .

BEISPIEL 2.21 (Fortsetzung von Beispiel 2.20). Es sei X die Anzahl der defekten Sicherungen in der Stichprobe. Nach den vorausgehenden Ausführungen genügt X einer $B(5; 0,05)$ Verteilung. Für das Ereignis A , die Stichprobe enthält mindestens

ABB. 2.7. $B(20; 0,05)$ -VerteilungABB. 2.8. $B(20; 0,5)$ -Verteilung

eine defekte Sicherung, findet man daher

$$P(A) = 1 - P(\bar{A}) = 1 - P(X = 0) = 1 - \binom{5}{0} 0,05^0 \cdot 0,95^5 = 0,226$$

eine in Anbetracht des geringen Umfangs der Stichprobe überraschend große Wahrscheinlichkeit.

PROPOSITION 2.19. Für eine binomial ($B(n; p)$) verteilte Zufallsvariable X gilt

$$(2.46) \quad E(X) = np$$

$$(2.47) \quad V(X) = npq$$

BEWEIS. Die Zufallsvariable X sei $B(n; p)$ verteilt. Dann gilt

$$\begin{aligned} E(X) &= \sum_{k=0}^n k P(X = k) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{n-1} q^{n-k} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} = np \end{aligned}$$

Für die Berechnung der Varianz von X stützen wir uns auf (2.42). Wir benötigen

$$\begin{aligned} E(X^2) &= \sum_{k=1}^n k^2 \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \frac{kn!}{(k-1)!(n-k)!} p^k q^{n-k} \\ &= np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^j q^{n-1-j} = np \left[\sum_{j=0}^{n-1} j \binom{n-1}{j} p^j q^{n-1-j} + 1 \right] \\ &= np [E(X_{n-1,p}) + 1] = np((n-1)p + 1) \end{aligned}$$

Setzt man $E(X^2)$ in (2.42) ein, erhält man

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

□

8.2. Geometrische Verteilung. Wir betrachten nun ein dem Bernoulli Experiment sehr ähnliches Zufallsexperiment: der einzige Unterschied besteht darin, daß wir diesmal daran interessiert sind, wann zum ersten Male das Experiment erfolgreich ausfällt. Die Zufallsvariable X beschreibt also die "Wartezeit" bis zum ersten "Treffer". Der Ereignisraum Ω für dieses Experiment besteht aus abzählbar unendlich vielen Elementarereignissen

$$\Omega = \{E, \bar{E}E, \bar{E}\bar{E}E, \bar{E}\bar{E}\bar{E}E, \dots\}.$$

Es seien E_i die Ereignisse (2.44) (mit $n = k$), dann gilt

$$\underbrace{\bar{E} \dots \bar{E}}_{(k-1)\text{-mal}} E = \bar{E}_1 \cap \dots \cap \bar{E}_{k-1} \cap E_k.$$

Die stochastische Unabhängigkeit der Ereignisse E_i ergibt

$$(2.48) \quad P(X = k) = q^{k-1}p, \quad k = 1, 2, \dots$$

Wegen

$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=1}^{\infty} q^{k-1} = \frac{p}{1-q} = 1$$

definiert P_X eine Wahrscheinlichkeitsverteilung auf \mathbb{N} , die **geometrische** oder **Pascal Verteilung**.

PROPOSITION 2.20. *Für eine geometrisch verteilte Zufallsvariable mit Parameter $p \in (0, 1)$ gilt*

$$(2.49) \quad E(X) = \frac{1}{p}$$

$$(2.50) \quad V(X) = \frac{q}{p^2}$$

BEWEIS. Der Erwartungswert von X ergibt sich aus folgender Überlegung

$$\begin{aligned} E(X) &= p \sum_{k=1}^{\infty} kq^{k-1} = p \sum_{k=0}^{\infty} \frac{d}{dq} q^k \\ &= p \frac{d}{dq} \frac{1}{1-q} = \frac{p}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

Die Berechnung der Varianz folgt demselben Schema. Wir berechnen zuerst

$$\begin{aligned} E(X^2) &= p \sum_{k=1}^{\infty} k^2 q^{k-1} = p \sum_{k=1}^{\infty} k(k-1)q^{k-1} + p \sum_{k=1}^{\infty} kq^{k-1} \\ &= pq \sum_{k=2}^{\infty} \frac{d^2}{dq^2} q^k + E(X) = pq \frac{2}{p^3} + \frac{1}{p} = \frac{2q}{p^2} + \frac{1}{p}. \end{aligned}$$

Die Varianz ergibt sich aus (2.42)

$$V(X) = E(X^2) - E(X)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

□

BEISPIEL 2.22. Bei der Erdölprospektion werden in einem bestimmten Gebiet Bohrungen niedergebracht. Die Wahrscheinlichkeit, bei einer Bohrung fündig zu werden, betrage 20%. Mit welcher Wahrscheinlichkeit trifft man bei der 4. Bohrung zum ersten Male auf Erdöl?

Die Zufallsvariable X , ($X(\omega) = k$ bedeute, bei der k -ten Bohrung wird man zum ersten Male fündig), ist geometrisch mit Parameter $p = 0,2$ verteilt. Somit findet man

$$P(X = 3) = pq^3 = 0,2 \cdot 0,8^3 \approx 0,10.$$

8.3. Poisson Verteilung.

BEISPIEL 2.23. Eine Telephonzentrale erhält im Laufe einer Stunde durchschnittlich 60 Anrufe. Wie groß ist die Wahrscheinlichkeit dafür, daß innerhalb von 30 Sekunden, in denen sich die Telephonistin entfernt a) kein Anruf, b) genau ein Anruf eintrifft.

Für die Lösung dieser Aufgabe benötigt man die Verteilung der Zufallsvariablen X , welche die innerhalb von jeweils 30 Sekunden in der Telephonzentrale eintreffenden Anrufe zählt. Offenbar handelt es sich um eine diskrete Zufallsvariable, welche nur die Werte $0, 1, \dots$, annehmen kann. Mit folgendem Kunstgriff führen wir die unbekannte Verteilung von X auf die Binomialverteilung zurück: Wir denken uns das relevante Zeitintervall (hier 30 Sekunden) in n so kleine Abschnitte unterteilt, daß in jedem Abschnitt höchstens ein Ereignis (hier Anruf) mit Wahrscheinlichkeit $p_n > 0$ eintreten

kann. Natürlich treffen die Anrufe unabhängig voneinander in der Telephonzentrale ein. Führt man also eine Zufallsvariable X_n mit der Bedeutung

$X_n = k$ in genau k Abschnitten wird ein Ereignis registriert, $k = 0, 1, \dots, n$,

ein, kann man für X_n eine $B(n, p_n)$ Verteilung annehmen mit einer noch unbekanntem Erfolgswahrscheinlichkeit p_n . Somit gilt

$$P_n(X_n = k) = \binom{n}{k} p_n^k q_n^{n-k}$$

und

$$E(X_n) = np_n.$$

Der Erwartungswert von X_n ist die im Mittel zu erwartende Anzahl von Ereignissen im gesamten Zeitintervall (hier: Anrufe innerhalb von 30 Sekunden). Somit ist $E(X_n)$ unabhängig von n . Wir setzen $\lambda = np_n$. Da wir weder p_n noch n kennen, liegt es nahe, die Unterteilung immer feiner zu machen und schließlich den Grenzfall $\lim_{n \rightarrow \infty} P_n(X_n = k)$ mit der Nebenbedingung $np_n = \lambda$ zu untersuchen:

$$\begin{aligned} P_n(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \underbrace{\prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right)}_{\xrightarrow{n \rightarrow \infty} 1} \underbrace{\frac{\lambda^k}{k!}}_{\xrightarrow{n \rightarrow \infty} e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n-k}}_{\xrightarrow{n \rightarrow \infty} 1} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

Wir erhalten also für die gesuchte Verteilung von X

$$(2.51) \quad P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Eine einfache Rechnung zeigt

$$\sum_{k=0}^{\infty} P(X = k) = 1,$$

somit definiert (2.51) eine diskrete Verteilung auf \mathbb{N}_0 , die **Poisson Verteilung**. Wegen der Konstruktion der Poissonverteilung ist es nicht überraschend, daß der positive Parameter λ der Erwartungswert von X ist:

PROPOSITION 2.21. *Für eine Poisson verteilte Zufallsvariable X mit Parameter $\lambda > 0$ gilt*

$$(2.52) \quad E(X) = V(X) = \lambda$$

BEWEIS. Der Erwartungswert von X folgt unmittelbar aus der Exponentialreihe. Zur Berechnung der Varianz betrachten wir wieder zuerst

$$\begin{aligned} E(X^2) &= e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \left[\sum_{k=1}^{\infty} k(k-1) \frac{\lambda^k}{k!} + \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \right] \\ &= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + \lambda = \lambda^2 + \lambda \end{aligned}$$

und somit

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

□

BEISPIEL 2.24 (Fortsetzung von Beispiel 2.23). Da 1 Stunde aus 120 Abschnitten von je 30 Sekunden besteht und in 1 Stunde im Mittel 60 Anrufe eintreffen, ergibt sich aus den Überlegungen, welche zur Poisson Verteilung führten

$$60 = 120 \cdot np_n = 120\lambda = 120E(X),$$

also $E(X) = \frac{1}{2}$. Die gesuchten Wahrscheinlichkeiten sind demnach

- a) $P(X = 0) = \frac{0,5^0}{0!} e^{-0,5} \approx 0,61$
- b) $P(X = 1) = \frac{0,5^1}{1!} e^{-0,5} \approx 0,30$

Die Poisson Verteilung ist oft ein gutes Modell für die Wahrscheinlichkeitsverteilung der Anzahl von Ereignissen, welche während einer gewissen Zeitspanne, in einem bestimmten Volums- oder Flächenelement eintreten und dort den Erwartungswert λ besitzen. Weitere Beispiele sind die Anzahl der von einer radioaktiven Substanz während einer Zeiteinheit emittierten Partikel, die Anzahl der monatlichen Unfälle in einer Fabrik, usw. In Hinsicht auf ihre Herleitung ist klar, daß die Poisson Verteilung auch zur Approximation der Binomialverteilung verwendet werden kann. Tabelle 2.3 demonstriert die Güte der Approximation:

8.4. Hypergeometrische Verteilung. Die Binomialverteilung wurde unter der Voraussetzung abgeleitet, daß die Wahrscheinlichkeit p für das Eintreten des beobachteten Ereignisses durch die Entnahme der Stichprobe kaum beeinflusst wird. Diese Annahme ist gerechtfertigt, wenn die Gesamtpopulation groß ist im Vergleich zum Umfang der Stichprobe. Wir betrachten nun Situationen, in denen diese Annahme nicht zutrifft.

BEISPIEL 2.25. Ein bestimmtes Produkt wird in Packungen zu je 20 Stück an die Kunden versandt. Da die Qualitätskontrolle für die gesamte Produktion einerseits zu teuer wäre, andererseits der Betrieb seine Abnehmer nicht mit der Belieferung defekter Produkte verärgern möchte, wird folgende Stichprobenkontrolle durchgeführt: Aus jeder Packung wird eine Stichprobe von 5 Stücken entnommen und die Packung zurückgewiesen, wenn mehr als ein defektes Stück gefunden wird

k	Binomialverteilung			Poisson Verteilung
	$n = 4, p = \frac{1}{4}$	$n = 8, p = \frac{1}{8}$	$n = 100, p = \frac{1}{100}$	$\lambda = 1$
0	0,316	0,344	0,366	0,368
1	0,422	0,393	0,370	0,368
2	0,211	0,196	0,185	0,184
3	0,047	0,056	0,061	0,061
4	0,004	0,010	0,015	0,015
5	--	0,001	0,003	0,003

TABELLE 2.3. Approximation der Binomialverteilung durch die Poisson Verteilung

(die zurückgewiesenen Packungen werden später einzeln untersucht). Wie groß ist die Wahrscheinlichkeit, daß eine Packung, welche 4 schadhafte Produkte enthält, zurückgewiesen wird.

Diesem Beispiel liegt folgende Struktur zugrunde: Eine Population besteht aus N Individuen, $M \leq N$ dieser Individuen tragen ein bestimmtes Merkmal. Aus der Population wird eine zufällige Stichprobe vom Umfang n entnommen (ohne Zurücklegen und ohne Beachtung der Reihenfolge) und man interessiert sich für die Anzahl der Individuen in der Stichprobe, welche das untersuchte Merkmal tragen. Zur Erleichterung der Sprechweise betrachten wir N Kugeln, von denen M rot und $N - M$ schwarz sind. Beobachtet wird die Anzahl der roten Kugeln in einer Stichprobe vom Umfang n . Zählt die Zufallsvariable X die roten Kugeln in der Stichprobe, dann kann X offensichtlich nur Werte k annehmen, für welche

$$0 \leq k \leq M, \quad 0 \leq n - k \leq N - M$$

gilt. Insgesamt gibt es $\binom{N}{n}$ Möglichkeiten ungeordnete Stichproben zu ziehen. Wegen der Zufälligkeit der Entnahme, sind diese Stichproben gleichwahrscheinlich (man denke sich die Kugeln von $1 \dots, N$ durchnummeriert, die Farbe wird nicht beachtet). Enthält die Stichprobe genau k rote Kugeln, sind notwendigerweise $n - k$ Kugeln schwarz. Die roten Kugeln in der Stichprobe stammen aus der Teilmenge der roten Kugeln in der Gesamtpopulation, dies ergibt insgesamt $\binom{M}{k}$ Möglichkeiten k rote Kugeln auszuwählen, die schwarzen Kugeln in der Stichprobe stammen aus der Teilmenge der schwarzen Kugeln in der Gesamtpopulation, dies ergibt insgesamt $\binom{N-M}{n-k}$ Möglichkeiten $n - k$ schwarze Kugeln auszuwählen. Da jede Auswahl von k roten mit jeder Auswahl von $n - k$ schwarzen Kugeln kombiniert werden kann, gibt es insgesamt $\binom{M}{k} \binom{N-M}{n-k}$ Stichproben mit genau k roten Kugeln. Die Zufallsvariable X ist daher

folgendermaßen verteilt

$$(2.53) \quad P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, n.$$

(Wir erinnern an die Konvention $\binom{a}{b} = 0$ für $b > a$. Somit ist (2.53) sinnvoll auch für $k > M$ und $k < n - (N - M)$.) Wir zeigen zuerst, daß (2.53) tatsächlich ein diskretes Wahrscheinlichkeitsmaß definiert: Dazu denken wir uns die Kugeln von $1, \dots, n$ so nummeriert, daß die ersten M Kugeln rot sind. Das Ziehen einer Stichprobe vom Umfang n kann als Laplace Experiment in

$$\Omega = \{\omega : \omega \subset \{1, \dots, N\}, |\omega| = n\}$$

und P als Gleichverteilung auf Ω aufgefaßt werden. Die Zufallsvariable $X: \Omega \rightarrow \mathbb{N}_0$ ist formal definiert durch

$$X(\omega) = |\omega \cap \{1, \dots, M\}|$$

Die Abbildung (2.53) kann dann als induziertes Maß P_X (2.37) aufgefaßt werden. Man nennt (2.53) **hypergeometrische Verteilung** $\mathcal{H}(N, M, n)$ mit den Parametern n , $M \leq N$. Für die Berechnung des Erwartungswertes und der Varianz einer hypergeometrisch verteilten Zufallsvariablen benötigen wir folgende Identität, welche auch für den Nachweis benützt werden kann, daß $\mathcal{H}(N, M, n)$ ein Wahrscheinlichkeitsmaß definiert:

LEMMA 2.1. *Es seien $m \leq n$ natürliche Zahlen. Dann gilt*

$$(2.54) \quad \sum_{k=0}^l \binom{n}{k} \binom{m}{l-k} = \binom{n+m}{l}, \quad l = 0, \dots, n+m.$$

BEWEIS. Wir gehen aus einerseits von

$$(1+x)^n (1+x)^m = (1+x)^{n+m} = \sum_{l=0}^{n+m} \binom{n+m}{l} x^l,$$

andererseits gilt aber auch

$$\begin{aligned} (1+x)^n (1+x)^m &= \sum_{k=0}^n \binom{n}{k} x^k \sum_{i=0}^m \binom{m}{i} x^i = \sum_{k=0}^n \sum_{i=0}^m \binom{n}{k} \binom{m}{i} x^{k+i} \\ &= \sum_{k=0}^n \sum_{l=k}^{m+k} \binom{n}{k} \binom{m}{l-k} x^l. \end{aligned}$$

Vertauscht man die Summationsreihenfolge, zerfällt die Doppelsumme in drei Teilsommen, welche wieder zusammengeführt werden können, wenn man $\binom{a}{b} = 0$ setzt,

falls $b < 0$ oder $a \geq 0$ und $b > a$ gilt:

$$\begin{aligned} (1+x)^n(1+x)^m &= \left[\sum_{l=0}^m \sum_{k=0}^l + \sum_{l=m+1}^n \sum_{k=l-m}^l + \sum_{l=n+1}^{n+m} \sum_{k=l-m}^n \right] \binom{n}{k} \binom{m}{l-k} x^l \\ &= \sum_{l=0}^{n+m} \left[\sum_{k=0}^l \binom{n}{k} \binom{m}{l-k} \right] x^l \end{aligned}$$

Die Behauptung folgt nun aus einem Vergleich der Koeffizienten in den beiden Darstellungen für $(1+x)^n(1+x)^m$. \square

PROPOSITION 2.22. *Es sei X eine hypergeometrisch $\mathcal{H}(N, M, n)$ verteilte Zufallsvariable. Dann gilt*

$$(2.55) \quad E(X) = n \frac{M}{N}$$

$$(2.56) \quad V(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}, \quad n \geq 2,$$

$$(2.57) \quad V(X) = \frac{M}{N} \left(1 - \frac{M}{N}\right), \quad n = 1.$$

BEWEIS. Wegen $k \binom{M}{k} = M \binom{M-1}{k-1}$ erhält man für den Erwartungswert vorerst

$$\begin{aligned} E(X) &= \sum_{\substack{\min\{M,n\} \\ \max\{0, n-(N-M)\}}} k \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{k=1}^n k \binom{M}{k} \binom{N-M}{n-k} \\ &= \frac{M}{\binom{N}{n}} \sum_{k=1}^n \binom{M-1}{k-1} \binom{N-M}{n-k} = \frac{M}{\binom{N}{n}} \sum_{k=0}^{n-1} \binom{M-1}{k} \binom{N-M}{n-1-k} \\ &= \frac{M}{\binom{N}{n}} \binom{N-1}{n-1} = n \frac{M}{N}. \end{aligned}$$

In der vorletzten Gleichung wurde Lemma 2.1 mit den Werten $(l, n, m) \leftrightarrow (n-1, M-1, N-M)$ verwendet. Zur Berechnung der Varianz für $n \geq 2$ betrachten wir wieder

$$\begin{aligned} E(X^2) &= \frac{1}{\binom{N}{n}} \sum_{k=0}^n k^2 \binom{M}{k} \binom{N-M}{n-k} = \frac{1}{\binom{N}{n}} \sum_{k=2}^n k(k-1) \binom{M}{k} \binom{N-M}{n-k} + E(X) \\ &= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=2}^n \binom{M-2}{k-2} \binom{N-M}{n-k} + E(X) \\ &= \frac{M(M-1)}{\binom{N}{n}} \sum_{k=0}^{n-2} \binom{M-2}{k} \binom{N-M}{n-2-k} + E(X) \\ &= \frac{M(M-1)}{\binom{N}{n}} \binom{N-2}{n-2} + n \frac{M}{N}, \end{aligned}$$

im letzten Schritt wurde wieder Lemma 2.1 eingesetzt. Eine einfache Rechnung führt nun zu dem gesuchten Ergebnis

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = \frac{M(M-1)n(n-1) + nM(N-1)}{N(N-1)} - \frac{n^2M^2}{N^2} \\ &= \frac{nM}{N^2(N-1)} (-nN - NM + N^2 + nM) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}. \end{aligned}$$

Im Fall $n = 1$ erhält man

$$E(X^2) = \frac{\binom{M}{1} \binom{N-M}{0}}{\binom{N}{1}} = \frac{M}{N} = E(X)$$

also

$$V(X) = E(X)(1 - E(X)) = \frac{M}{N} \left(1 - \frac{M}{N}\right).$$

□

BEISPIEL 2.26 (Fortsetzung von Beispiel 2.25). Bezeichnet man mit D die Anzahl der defekten Stücke in der Stichprobe, dann wird die Packung zurückgewiesen, wenn D die Werte 2, 3 und 4 annimmt. Die Zufallsvariable D ist hypergeometrisch $\mathcal{H}(20, 4, 5)$ verteilt. Die Wahrscheinlichkeit der Zurückweisung beträgt daher

$$\begin{aligned} P(D \geq 2) &= P(D = 2) + P(D = 3) + P(D = 4) = 1 - P(D = 0) - P(D = 1) \\ &= 1 - \frac{\binom{4}{0} \binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1} \binom{16}{4}}{\binom{20}{5}} \approx 1 - 0,28 - 0,47 = 0,25 \end{aligned}$$

9. Stetige Zufallsvariable

Bisher haben wir nur diskrete Zufallsvariable betrachtet, deren Bildbereich höchstens abzählbar ist und aus isolierten Punkten besteht. Es ist allerdings nicht schwer, Beispiele anzugeben, welche nicht in das Konzept einer diskreten Zufallsvariablen passen. Man denke beispielsweise an die tägliche Niederschlagsmenge, an die Lebensdauer eines elektronischen Bauteils usw., welche jeden Wert in einem sinnvollen Intervall annehmen können. Wir nennen derartige Zufallsvariable **kontinuierlich**. Im folgenden werden wir stets voraussetzen, daß das zugrundeliegende Zufallsexperiment durch ein stetiges Wahrscheinlichkeitsmaß gesteuert wird. Die Verteilungsfunktion der Zufallsvariablen kann also in der Form

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

geschrieben werden. Dies hat $P(X = a) = 0$ zur Folge. Ferner gilt

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

Wir nennen Zufallsvariable, deren Verteilungsfunktion eine Dichte besitzt, **stetig**.

DEFINITION 2.12. *Es sei X eine stetige Zufallsvariable auf \mathbb{R} . Existiert das uneigentliche Integral $\int_{-\infty}^{\infty} |x|f(x) dx$, dann nennt man*

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

den **Erwartungswert** von X . Die **Varianz** von X wird wie im diskreten Fall durch

$$V(X) = E((X - E(X))^2),$$

definiert.

Man überzeugt sich leicht davon, daß obige Definition des Erwartungswertes einer stetigen Zufallsvariablen den Erwartungswert einer diskreten Zufallsvariablen umfaßt. Wegen der Linearität des Integrals ist der Erwartungswert ein lineares Funktional. Die Varianz kann daher wieder nach (2.42) berechnet werden. Manchmal ist man am Erwartungswert einer Funktion einer Zufallsvariablen interessiert. Ohne Beweis notieren wir folgendes Resultat:

PROPOSITION 2.23. *Es sei X eine stetige Zufallsvariable auf \mathbb{R} und $g: X(\mathbb{R}) \rightarrow \mathbb{R}$ eine meßbare Funktion. Dann gilt*

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx,$$

sofernne $\int_{-\infty}^{\infty} |g(x)|f(x) dx$ existiert.

BEISPIEL 2.27. Gegeben sei

$$f(x) = \begin{cases} cx^2 & x \in [0, 2] \\ 0 & \text{sonst} \end{cases}.$$

Man bestimme die Konstante $c > 0$ so, daß f eine zuläßige Wahrscheinlichkeitsdichte darstellt und bestimme Erwartungswert und Varianz jener Zufallsvariablen, deren Verteilungsfunktion die Dichte f hat.

Die Konstante c ergibt sich aus der Forderung $\int_{-\infty}^{\infty} f(x) dx = 1$ zu $c = \frac{3}{8}$. Für den gesuchten Erwartungswert erhält man

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx = \int_0^2 x \frac{3}{8} x^2 dx = \frac{3}{2},$$

für die Varianz berechnet man wieder zuerst

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{3}{8} \int_0^2 x^4 dx = 2,4.$$

Dies ergibt

$$V(X) = E(X^2) - E(X)^2 = 2,4 - 1,5^2 = 0,15.$$

9.1. Gleichförmige Verteilung. Eine Zufallsvariable X besitzt über dem Intervall $[a, b]$ eine **gleichförmige Verteilung** (ist über dem Intervall $[a, b]$ gleich verteilt), wenn die Verteilungsfunktion die Dichte

$$(2.58) \quad f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{sonst} \end{cases}$$

besitzt. Eine einfache Integration zeigt:

PROPOSITION 2.24. *Es sei X eine auf dem Intervall $[a, b]$ gleichverteilte Zufallsvariable. Dann gilt*

$$(2.59) \quad E(X) = \frac{1}{2}(a + b),$$

$$(2.60) \quad V(X) = \frac{1}{12}(b - a)^2.$$

Eine wichtige Anwendung der Gleichverteilung ist die Simulation statistischer Daten, welche nicht gleichverteilt sind. In vielen Fällen ist es nämlich möglich, eine Transformation von der Gleichverteilung auf die gewünschte Verteilung anzugeben. In diesen Situationen können Simulationen mit dem in jeden Computer eingebauten Zufallszahlengenerator durchgeführt werden, welcher im Intervall $(0, 1)$ gleichverteilte Zufallszahlen erzeugt.

Manche stetige Zufallsvariable in Biologie, Wirtschaft und Naturwissenschaft können als gleichverteilt angenommen werden. Zählt man beispielsweise Ereignisse, welche einer Poissonverteilung genügen, und weiß man, daß in einem bestimmten Intervall $[0, t]$ genau ein Ereignis registriert wurde, dann ist der tatsächliche Zeitpunkt des Eintretens dieses Ereignisses im Intervall $[0, t]$ gleichverteilt.

BEISPIEL 2.28. Die Anzahl der Kunden, welche sich während einer bestimmten Zeitspanne an einer Kasse anstellen, genügt einer Poisson Verteilung. Angenommen, in einer 30-Minuten Periode ist genau ein Kunde zur Kasse gekommen. Wie groß ist die Wahrscheinlichkeit, daß er während der letzten 5 Minuten erschienen ist? Die Zufallsvariable T bezeichne den Zeitpunkt des Erscheinens des Kunden an der Kasse. Wie vorhin erwähnt, ist T auf $[0, 30]$ gleichverteilt. Somit folgt

$$P(25 \leq T \leq 30) = \int_{25}^{30} \frac{1}{30} dt = \frac{1}{6}.$$

Folgende alternative Lösung ist sehr instruktiv: Die Zufallsvariable $X(t_0, t_1)$ bezeichne nun die Anzahl der an der Kasse in der Zeitspanne $[t_0, t_1]$ eintreffenden Kunden. Nach Voraussetzung ist $X(t_0, t_1)$ Poisson verteilt, also

$$P(X(t_0, t_1) = k) = \frac{1}{k!} e^{-\lambda(t_1 - t_0)} (\lambda(t_1 - t_0))^k.$$

(Die Zeit wird in Minuten gemessen). Wenn der Kunde in den letzten 5 Minuten erscheint, ist in den ersten 25 Minuten niemand zur Kasse gekommen. Gesucht ist

also offenbar die bedingte Wahrscheinlichkeit

$$P(X(0, 25) = 0 | X(0, 30) = 1) = \frac{P(X(0, 25) = 0 \text{ und } X(0, 30) = 1)}{P(X(0, 30) = 1)}.$$

Die Wahrscheinlichkeit des Durchschnitts ergibt sich aus

$$\begin{aligned} P(X(0, 25) = 0 \text{ und } X(0, 30) = 1) &= P(X(0, 30) = 1 | X(0, 25) = 0) \cdot P(X(0, 25) = 0) \\ &= P(X(25, 30) = 1) \cdot P(X(0, 25) = 0). \end{aligned}$$

Setzt man in die Poisson Verteilung ein, ergeben sich die Wahrscheinlichkeiten

$$\begin{aligned} P(X(0, 30) = 1) &= 30\lambda e^{-30\lambda}, \\ P(X(0, 25) = 0) &= e^{-25\lambda}, \\ P(X(25, 30) = 1) &= 5\lambda e^{-5\lambda}, \end{aligned}$$

und somit insgesamt

$$P(X(0, 25) = 0 | X(0, 30) = 1) = \frac{5\lambda e^{-5\lambda} e^{-25\lambda}}{30\lambda e^{-30\lambda}} = \frac{5}{30} = \frac{1}{6}.$$

9.2. Normalverteilung. Die Normalverteilung oder die Gauß Verteilung wurde von C.F. Gauß bei der Untersuchung von Meßfehlern eingeführt. Sie ist die bei weitem wichtigste stetige Verteilung. Hiefür gibt es verschiedene Gründe:

- Viele Zufallsvariable, die in der Praxis auftreten, sind normalverteilt.
- Für viele Zufallsvariable, deren empirische kumulative relative Häufigkeit unimodal ist, führt die Annahme einer Normalverteilung zu brauchbaren Ergebnissen.
- Manche nichtnormalverteilte Zufallsvariable lassen sich verhältnismäßig einfach auf normalverteilte Zufallsvariable transformieren.
- Zahlreiche komplizierte Verteilungen können durch die Normalverteilung approximiert werden.
- Zahlreiche Zufallsvariable lassen sich als Summe einer großen Anzahl von unabhängigen Zufallsvariablen auffassen. Der zentrale Grenzwertsatz besagt, daß eine derartige Zufallsvariable “ungefähr” normalverteilt ist.

DEFINITION 2.13. *Eine stetige Zufallsvariable ist **normalverteilt**, kurz $N(\mu, \sigma)$ -verteilt, wenn die Dichte der Verteilungsfunktion gegeben ist durch*

$$(2.61) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}.$$

Abb. 2.9 zeigt die typische Glockenform der Dichten einer Normalverteilung für verschiedene Werte von σ mit $\mu = 0$. Die Kurven sind symmetrisch bezüglich μ , das globale Maximum wird in μ angenommen. Für große Werte von σ ist der Kurvenverlauf flach, für kleine Werte von σ wird der Maximalwert größer und der Anstieg steiler. Durch (2.61) wird tatsächlich eine Wahrscheinlichkeitsdichte definiert. Dies ist eine unmittelbare Konsequenz aus dem Standardintegral

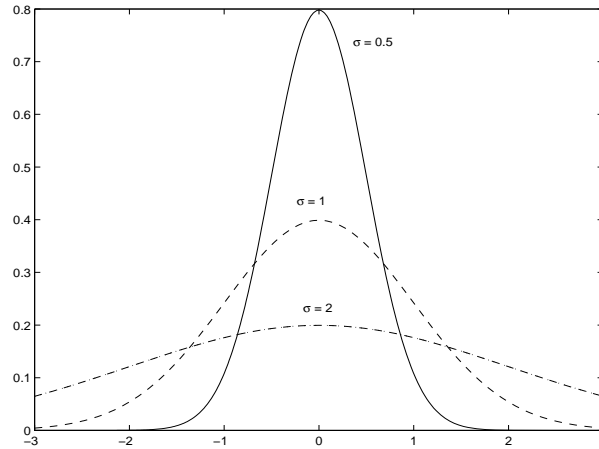


ABB. 2.9. Dichte der Normalverteilung

$$(2.62) \quad \int_{-\infty}^{\infty} e^{-u^2} du = \frac{1}{\sqrt{\pi}}.$$

Ein schöner Beweis verwendet folgenden Trick: man quadriert zuerst das Integral, schreibt das Produkt als Doppelintegral und geht über auf Polarkoordinaten:

$$\begin{aligned} \left[\int_{-\infty}^{\infty} e^{-u^2} du \right]^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-u^2-v^2} du dv = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} r dr d\varphi \\ &= 2\pi \int_0^{\infty} e^{-r^2} r dr = -\pi \int_0^{\infty} \frac{d}{dr} e^{-r^2} dr = \pi. \end{aligned}$$

Wir überlassen es dem Leser als einfache Integrationsübung, Erwartungswert und Varianz einer normalverteilten Zufallsvariablen zu berechnen:

PROPOSITION 2.25. *Es sei X eine $N(\mu, \sigma)$ -verteilte Zufallsvariable. Dann gilt*

$$(2.63) \quad E(X) = \mu,$$

$$(2.64) \quad V(X) = \sigma^2.$$

Einer der großen Vorteile der Normalverteilung besteht darin, daß sich jede $N(\mu, \sigma)$ -verteilte Zufallsvariable X auf eine $N(0, 1)$ verteilte Zufallsvariable Z zurückführen läßt. Dazu setzt man

$$(2.65) \quad Z = \frac{1}{\sigma}(X - \mu), \quad \text{also } X = \sigma Z + \mu.$$

Die Verteilung von Z ergibt sich aus folgender Überlegung:

$$F_Z(z) = P(Z \leq z) = P\left(\frac{1}{\sigma}(X - \mu) \leq z\right) = P(X \leq \sigma z + \mu) = F_X(\sigma z + \mu)$$

woraus sich nach Satz 2.4 die Dichte von Z durch Ableitung ergibt

$$\varphi(z) = f_Z(z) = \frac{d}{dz} F_X(\sigma z + \mu) = \sigma f_X(\sigma z + \mu)$$

Setzt man in (2.61) ein, erhält man die Dichte

$$(2.66) \quad \varphi(z) = f_Z(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, & z \geq 0 \\ 0, & z < 0. \end{cases}$$

Durch Vergleich mit (2.61) erkennt man, daß (2.66) die Dichte einer $N(0, 1)$ -verteilten Zufallsvariablen darstellt. Eine analoge Überlegung zeigt, daß $Z = aX + b$, $a \neq 0$, $b \in \mathbb{R}$, einer $N(a\mu + b, a\sigma)$ Verteilung folgt, falls X eine $N(\mu, \sigma)$ verteilte Zufallsvariable darstellt.

Die $N(0, 1)$ Verteilung heißt **Standardnormalverteilung**. Anstelle von F_Z ist die Bezeichnung $\Phi(z)$ üblich, statt F_X schreiben wir wieder einfacher F . Es ist nicht möglich die Verteilungsfunktion F geschlossen auszuwerten, man muß auf numerische Integrationsmethoden zurückgreifen. Wegen des Zusammenhanges zwischen der $N(\mu, \sigma)$ - und der $N(0, 1)$ -Verteilung genügt es allerdings, Tabellen für die Standardnormalverteilung zu berechnen:

$$(2.67) \quad P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Aus Symmetriegründen folgt die Beziehung

$$\Phi(z) - \Phi(-z) = 1 - 2\Phi(-z), \quad z > 0,$$

also

$$\Phi(-z) = 1 - \Phi(z) \quad z > 0.$$

Aus diesem Grunde ist es nicht notwendig, Φ für negative Werte von z zu berechnen.

Aus diesem Grunde ist es nicht notwendig, Φ für negative Werte von z zu berechnen. Abschließend notieren wir folgende nützliche Regel:

$$(2.68) \quad P(|X - \mu| \leq \sigma) = 0,683 \quad P(|X - \mu| \geq \sigma) = 0,317$$

$$(2.69) \quad P(|X - \mu| \leq 2\sigma) = 0,954 \quad P(|X - \mu| \geq 2\sigma) = 0,046$$

$$(2.70) \quad P(|X - \mu| \leq 3\sigma) = 0,997 \quad P(|X - \mu| \geq 3\sigma) = 0,003$$

$$(2.71)$$

BEISPIEL 2.29. Eine Maschine stellt Platten her, deren Stärke naturgemäß schwankt. Die Plattenstärke X ist daher eine Zufallsvariable und –wie man weiß– normalverteilt mit dem Mittelwert 10 mm (hängt von der Maschineneinstellung ab) und der Standardabweichung 0,02 mm (hängt von der Qualität der Maschine ab). Wieviel Prozent Ausschuß ist zu erwarten, wenn a) die Platten mindestens 9,95 mm stark sein sollen,

b) die Platten höchstens 10,02 mm stark sein dürfen, c) die Plattenstärke zwischen 9,97 mm und 10,03 mm liegen muß.

Lösung: Die Zufallsvariable X ist $N(10; 0, 02)$ verteilt.

$$\text{a) } P(X \leq 9,95) = \Phi\left(\frac{9,95-10}{0,02}\right) = 0,0062.$$

$$\text{b) } P(X > 10,2) = 1 - P(X \leq 10,02) = 1 - \Phi\left(\frac{10,02-10}{0,02}\right) = 1 - \Phi(1) = 1 - 0,8413 = 0,1587.$$

$$\text{c) } 1 - P(97 < X \leq 10,03) = 1 - (\Phi\left(\frac{10,03-10}{0,02}\right) - \Phi\left(\frac{9,97-10}{0,02}\right)) = 1 - (\Phi(1,5) - \Phi(-1,5)) = 2 - 2\Phi(1,5) = 2 - 1,8664 = 0,1336.$$

In der Statistik ist folgende von der Normalverteilung abgeleitete Verteilung von großer Bedeutung.

DEFINITION 2.14. *Es sei X eine standardnormal verteilte Zufallsvariable. Man nennt die Verteilung von $Z = X^2$ χ_1^2 -Verteilung mit einem Freiheitsgrad.*

Um die Verteilungsfunktion von einer χ_1^2 -verteilten Zufallsvariablen $Z = X^2$ aus der Standardnormalverteilung von X abzuleiten, betrachten wir

$$F_Z(z) = P(Z \leq z) = P(-\sqrt{z} \leq X \leq \sqrt{z}) = \Phi(\sqrt{z}) - \Phi(-\sqrt{z}),$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{1}{2} z^{-1/2} [\varphi(\sqrt{z}) + \varphi(-\sqrt{z})] = z^{-1/2} \varphi(\sqrt{z}),$$

(man beachte $\varphi(z) = \Phi(z)'$). Setzt man die Dichte der Standardnormalverteilung ein, ergibt sich die Dichte der χ_1^2 -Verteilung:

$$(2.72) \quad f_Z(z) = \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2}, \quad z \geq 0.$$

Wegen $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ kann man (2.72) auch auffassen als Dichte einer Gamma verteilten Zufallsvariablen mit den Parametern $\alpha = \lambda = \frac{1}{2}$. Eine einfache Überlegung zeigt, wenn X einer $N(\mu, \sigma)$ -Verteilung folgt, dann genügt $(\frac{X-\mu}{\sigma})^2$ einer χ_1^2 -Verteilung.

9.3. Gamma Verteilung. Zahlreiche Zufallsvariable können nur nichtnegative Werte annehmen. Ihre empirische Wahrscheinlichkeitsdichte ist oft unimodal und asymmetrisch, vgl Abb. 2.10. Als Beispiel sei die Zeitspanne zwischen Funktionsstörungen bei Flugzeugmotoren, die Dauer von Routine Untersuchungen bei Flugzeugen oder Autos erwähnt. Derartige Situationen lassen sich nur schlecht mit einer Normalverteilung modellieren, da deren Dichte einerseits symmetrisch um den Erwartungswert ist, andererseits auch negativen Werten positive Wahrscheinlichkeitsdichten zugewiesen werden. In solchen Fällen ist oft der Einsatz der Gamma Verteilung sinnvoll.

DEFINITION 2.15. *Eine stetige Zufallsvariable X besitzt eine **Gamma Verteilung** mit den Parametern $\alpha > 0$ und $\lambda > 0$, wenn ihre Dichte gegeben ist durch*

$$(2.73) \quad f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0 & x < 0. \end{cases}$$

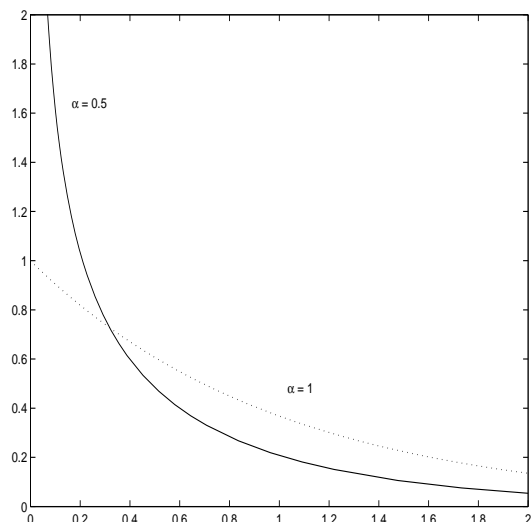


ABB. 2.10

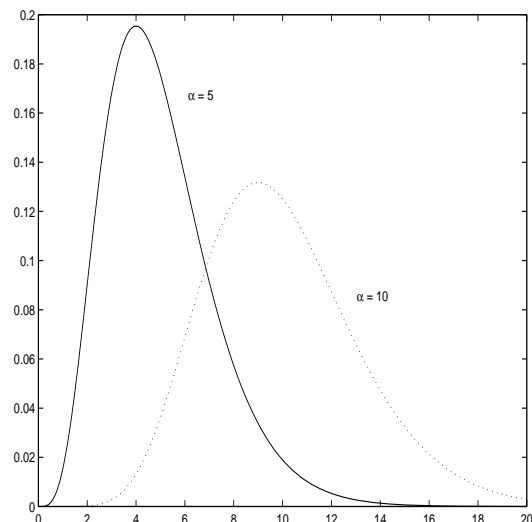


ABB. 2.11

Die Gamma Funktion $\Gamma(\alpha)$ ist definiert durch

$$(2.74) \quad \Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du, \alpha > 0$$

Die Abbildungen 2.10 und 2.11 zeigen die Dichten der Gamma Verteilung für verschiedene Werte des Parameters α . Die Form der Verteilungsfunktion hängt stark von α ab. Man nennt daher α den Formparameter der Gamma Verteilung. Der Parameter λ heißt Skalierungsparameter. Multipliziert man nämlich eine Gamma verteilte Zufallsvariable X mit einer Konstanten β erhält man wieder eine Gamma verteilte Zufallsvariable mit gleichem α , der Parameter λ wird durch $\frac{\lambda}{\beta}$ ersetzt. Einer Änderung von λ entspricht also die Änderung der Maßeinheit beim zugrundeliegenden Zufallsexperiment.

Die Gamma Funktion ist eine Verallgemeinerung der Fakultät: direkte Integration zeigt

$$\Gamma(1) = 1,$$

mit partieller Integration verifiziert man

$$(2.75) \quad \Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \alpha \geq 0,$$

woraus

$$\Gamma(n) = (n - 1)!, \quad n \in \mathbb{N}.$$

folgt. Für $\alpha \notin \mathbb{N}$ ist es nicht möglich, einen geschlossenen Ausdruck für die Wahrscheinlichkeit $P(a < X \leq b)$ anzugeben.

PROPOSITION 2.26. Wenn X eine Gamma verteilte Zufallsvariable mit den Parametern α und λ ist, dann gilt

$$E(X) = \frac{\alpha}{\lambda}$$

$$V(X) = \frac{\alpha}{\lambda^2}.$$

BEWEIS. Der Erwartungswert ergibt sich mit (2.75) aus folgender Rechnung:

$$E(X) = \int_0^{\infty} x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^{\infty} \left(\frac{u}{\lambda}\right)^\alpha e^{-u} du$$

$$= \frac{1}{\lambda} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

Eine ähnliche Rechnung zeigt

$$E(X^2) = \frac{1}{\lambda^2}(\alpha+1)\alpha,$$

somit folgt

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}(\alpha+1)\alpha - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

□

Ein Spezialfall der Gammaverteilung mit $\alpha = 1$ ist die folgende Exponentialverteilung.

9.4. Exponentialverteilung. Für manche Bauteile ist die Annahme sinnvoll, daß zu jedem Zeitpunkt die Wahrscheinlichkeit, daß der Bauteil noch weitere b Zeiteinheiten funktioniert, unabhängig ist von der Dauer, wie lange er bereits eingesetzt war. Genauer: die Wahrscheinlichkeit, daß der Modul mindestens $a + b$ Zeiteinheiten übersteht, sofern er bereits a Zeiteinheiten reibungslos funktioniert hat, ist gleich der Wahrscheinlichkeit, daß die Komponente mindestens b Zeiteinheiten funktioniert, wenn sie zum Zeitpunkt $t = 0$ neu eingesetzt wird. Man nennt diese Eigenschaft Gedächtnislosigkeit des Bauteils. Diese Eigenschaft ist charakteristisch für die Exponentialverteilung:

DEFINITION 2.16. Eine Zufallsvariable X heißt **exponentialverteilt** mit Parameter $\lambda > 0$, wenn ihre Verteilungsfunktion die Dichte

$$(2.76) \quad f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

besitzt.

Man überzeuge sich davon, daß durch (2.76) tatsächlich eine Wahrscheinlichkeitsdichte definiert wird. Eine einfache Rechnung ergibt Erwartungswert und Varianz einer exponentialverteilten Zufallsvariablen:

PROPOSITION 2.27. *Es sei X eine exponentialverteilte Zufallsvariable mit Parameter λ . Dann gilt*

$$(2.77) \quad E(X) = \frac{1}{\lambda},$$

$$(2.78) \quad V(X) = \frac{1}{\lambda^2}.$$

Für die kumulative Verteilungsfunktion von X ergibt sich

$$(2.79) \quad F(x) = \int_{-\infty}^x f(x) dx = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Somit folgt

$$P(X > x) = 1 - F(x) = e^{-\lambda x}.$$

BEISPIEL 2.30. In diesem Beispiel demonstrieren wir, daß die Exponentialverteilung tatsächlich die gewünschte Gedächtnislosigkeit besitzt. Ein elektronischer Bauteil sei also bereits a Stunden in Betrieb. Wir berechnen unter der Annahme einer Exponentialverteilung für die Lebensdauer T dieser Komponente die Wahrscheinlichkeit, daß sie noch weitere b Stunden einsatzfähig bleibt. Wir suchen also $P(T > a + b | T > a)$. Man erhält

$$\begin{aligned} P(T > a + b | T > a) &= \frac{P(T > a + b \text{ und } T > a)}{P(T > a)} = \frac{P(T > a + b)}{P(T > a)} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b}. \end{aligned}$$

10. Ungleichung von Tschebyscheff

Oft interessiert man sich für Wahrscheinlichkeiten der Art $P(|X - E(X)| < k\sigma)$, σ bezeichnet die Standardabweichung von X . Kennt man die Verteilung von X ist die Berechnung dieser Wahrscheinlichkeit vielleicht mühsam, aber prinzipiell durchführbar. In manchen Fällen ist die exakte Wahrscheinlichkeitsverteilung jedoch nicht bekannt, aber man kennt zumindest den Erwartungswert und die Varianz von X . In diesen Fällen ergibt die Ungleichung von Tschebyscheff zumindest eine untere Schranke für die interessierende Wahrscheinlichkeit

THEOREM 2.8. *Es sei X eine diskrete oder kontinuierliche Zufallsvariable mit Erwartungswert $E(X) = \mu$ und Varianz $V(X) = \sigma^2$. Dann gilt für jedes $\lambda > 0$*

$$(2.80) \quad P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2} \quad \text{bzw.} \quad P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}.$$

BEWEIS. Wir führen den Beweis nur für kontinuierliche Zufallsvariable. Es sei f die Wahrscheinlichkeitsdichte von X . Es folgt

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\mu - \lambda\sigma} (x - \mu)^2 f(x) dx \\ &\quad + \int_{\mu - \lambda\sigma}^{\mu + \lambda\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + \lambda\sigma}^{\infty} (x - \mu)^2 f(x) dx \\ &\geq \int_{-\infty}^{\mu - \lambda\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + \lambda\sigma}^{\infty} (x - \mu)^2 f(x) dx. \end{aligned}$$

In beiden Integralen gilt die Abschätzung $|x - \mu| \geq \lambda\sigma$, somit

$$\begin{aligned} V(X) &= \sigma^2 \geq \lambda^2 \sigma^2 \left[\int_{-\infty}^{\mu - \lambda\sigma} f(x) dx + \int_{\mu + \lambda\sigma}^{\infty} f(x) dx \right] \\ &= \lambda^2 \sigma^2 P(|X - \mu| \geq \lambda\sigma). \end{aligned}$$

□

BEISPIEL 2.31. Nehmen wir an, daß erfahrungsgemäß die Dauer X (in Minuten) eines Routineservice für einen elektronischen Bauteil einer Gammaverteilung mit $\alpha = 3.1$ und $\lambda = \frac{1}{2}$ folgt. Ein neuer Servicetechniker benötigt 21.5 Minuten. Hat man eine gute (= flinke) Kraft angeheuert?

Der Erwartungswert und die Varianz für die Dauer des Service sind (vgl. Proposition 2.26)

$$E(X) = \frac{\alpha}{\lambda} = 6,2, \quad V(X) = \frac{\alpha}{\lambda^2} = 12,4 \quad \text{also } \sigma = 3,52.$$

Die Servicedauer von 21,5 Minuten übertrifft den Erwartungswert um 15,3 Minuten, das sind $k = \frac{15,3}{3,52}$ Standardabweichungen. Die Ungleichung von Tschebyscheff ergibt

$$P(|X - 6,2| \geq 15,3) \leq \frac{1}{k^2} = 0,0529.$$

Dieses Resultat läßt zwei Schlüsse zu: entweder ist dieser Arbeiter langsamer als im Mittel die übrigen, oder er hatte zufällig mit einem besonders heiklen Fall zu tun. Die Wahrscheinlichkeit für die zweite Möglichkeit ist allerdings kleiner als 5%.

11. Mehrdimensionale Zufallsvariable

Oft betrachtet man bei einem Zufallsexperiment gleichzeitig mehrere Größen: z. B. bei zufällig ausgewählten Personen Alter und Einkommensklasse, bei Werkstoffen Härte, Zug- und Druckfestigkeit, Gehalt an verschiedenen Komponenten usw. Man hat also mit mehreren Zufallsvariablen zu tun, an deren gemeinsamer Verteilung man interessiert ist. Die gemeinsame Verteilung mehrerer Zufallsvariablen ist aber auch von großer Bedeutung bei der theoretischen Begründung statistischer

Prüfverfahren. Eine rigorose Entwicklung der Theorie mehrdimensionaler Zufallsvariabler sprengt den Rahmen dieser Einführung. Wir beschränken uns daher auf die Diskussion der grundlegenden Ideen. Wir beginnen mit der einfachsten Situation.

Bei einem Zufallsexperiment werden gleichzeitig zwei Größen beobachtet, welche durch die Zufallsvariablen X und Y beschrieben werden. Wir bezeichnen mit $p(x, y) = P(X = x, Y = y)$ beispielsweise die Wahrscheinlichkeit, daß X den Wert x und gleichzeitig Y den Wert y annimmt. Die gemeinsame Verteilung der beiden Zufallsvariablen X und Y ist bestimmt durch die gemeinsame kumulative Verteilungsfunktion

$$(2.81) \quad F(x, y) = P(X \leq x, Y \leq y).$$

Sie gibt die Wahrscheinlichkeit an, mit der X einen Wert annimmt, der höchstens x ist, und mit der Y gleichzeitig einen Wert annimmt, der y nicht übersteigt. Die Verteilungsfunktion F bestimmt die zweidimensionale Wahrscheinlichkeitsverteilung eindeutig durch

$$(2.82) \quad P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Um diese Formel einzusehen, definieren wir die Mengen

$$\begin{aligned} A &= (-\infty, a_1] \times (a_2, b_2] & B &= (a_1, b_1] \times (a_2, b_2] \\ C &= (-\infty, a_1] \times (-\infty, a_2] & D &= (a_1, b_1] \times (-\infty, a_2] \end{aligned}$$

und schreiben $A \cup B \cup C \cup D$ als disjunkte Vereinigung

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

Somit gilt

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

Berücksichtigt man noch

$$\begin{aligned} F(b_1, b_2) &= P(A \cup B \cup C \cup D), & F(a_1, b_2) &= P(A \cup C), \\ F(b_1, a_2) &= P(C \cup D), & F(a_1, a_2) &= P(C) \end{aligned}$$

erhält man (2.82). Die gemeinsame Verteilungsfunktion von n Zufallsvariablen X_1, \dots, X_n ist gegeben durch

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Die Zufallsvariablen X_1, \dots, X_n kann man auch zu einer einzigen n -dimensionalen Zufallsvariablen X zusammenfassen.

11.1. Diskrete mehrdimensionale Zufallsvariable. Sind X_1, \dots, X_n diskrete Zufallsvariable, kann man auch mit der gemeinsamen Wahrscheinlichkeitsverteilung rechnen:

$$(2.83) \quad p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

BEISPIEL 2.32. Eine faire Münze wird dreimal geworfen. Die Zufallsvariable X zähle die Anzahl der Köpfe beim ersten Wurf und Y die Gesamtanzahl der Köpfe. Der Ereignisraum dieses Zufallsexperimentes ist

$$\Omega = \{KKK, KKZ, KZK, KZZ, ZKK, ZKZ, ZZK, ZZZ\},$$

woraus man die gemeinsame Verteilung von X und Y ablesen kann

x	y				P_1
	0	1	2	3	
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
P_2	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

TABELLE 2.4

Beispielsweise gilt $p(1, 2) = P(X = 1, Y = 2) = P(\{KKZ, KZK\}) = \frac{2}{8}$. Wie kann man aus der gemeinsamen Verteilung z.B. die Verteilung von Y gewinnen? Offensichtlich gilt

$$p_2(0) = P(X \in \{0, 1\}, Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = \frac{1}{8} + 0 = \frac{1}{8}.$$

Analog findet man $p_2(1) = p_2(2) = \frac{3}{8}$, $p_2(3) = \frac{1}{8}$ und $p_1(0) = P(X = 0, Y \in \{0, 1, 2, 3\}) = \frac{1}{2}$, $p_1(1) = \frac{1}{2}$ (das war natürlich zu erwarten). Die Wahrscheinlichkeitsverteilungen p_1 und p_2 sind sogenannte Randverteilungen.

DEFINITION 2.17. *Es seien X_1, \dots, X_n diskrete Zufallsvariable und P ihre gemeinsame Wahrscheinlichkeitsverteilung. Die **i -te Randverteilung (Marginalverteilung)** ist die Wahrscheinlichkeitsverteilung P_i , $i = 1, \dots, n$*

$$(2.84) \quad P_i(X_i = x_i) \equiv p_i(x_i) = \sum_{\substack{x_1, \dots, x_j, \dots, x_n \\ j \neq i}} p(x_1, \dots, x_j, \dots, x_n), i = 1, \dots, n.$$

Für den wichtigsten Spezialfall einer zweidimensionalen Zufallsvariablen (X, Y) bedeutet (2.84)

$$p_1(x_k) = \sum_{j=1}^{\infty} p(x_k, y_j), \quad p_2(y_j) = \sum_{k=1}^{\infty} p(x_k, y_j).$$

Ordnet man also die Wahrscheinlichkeitsverteilung einer zweidimensionalen diskreten Zufallsvariablen wie in Tabelle 2.4 an, dann bedeutet die i -te Zeilensumme die Randverteilung P_1 , die j -te Spaltensumme die Randverteilung P_2 dar. Es sei darauf hingewiesen, daß die gemeinsame Verteilung durch die Randverteilungen allein noch nicht eindeutig bestimmt wird. Die kumulative Verteilungsfunktion im Beispiel 2.32 erhält man aus

$$(2.85) \quad F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j).$$

Eine einfache Rechnung ergibt

	$y < 0$	$0 \leq y < 1$	$1 \leq y < 2$	$2 \leq y < 3$	$y \geq 3$
$x < 0$	0	0	0	0	0
$0 \leq x < 1$	0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{4}{8}$
$x \geq 1$	0	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{7}{8}$	1

TABELLE 2.5. gemeinsame kumulative Verteilungsfunktion

11.2. Zweidimensionale stetige Zufallsvariable.

DEFINITION 2.18. *Es seien X und Y stetige Zufallsvariable und $F: \mathbb{R}^2 \rightarrow [0, 1]$ ihre gemeinsame Verteilungsfunktion. Man nennt (X, Y) eine **zweidimensionale stetige Zufallsvariable**, wenn es eine nichtnegative, integrierbare Funktion $f: \mathbb{R}^2 \rightarrow [0, \infty)$ gibt, mit*

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(r, s) dr ds, \quad (x, y) \in \mathbb{R}^2.$$

f heißt *gemeinsame (2-dimensionale) Wahrscheinlichkeitsdichte*

Implizit wurde in Definition 2.18 angenommen, daß die gemeinsame Dichte tatsächlich ein Wahrscheinlichkeitsmaß auf \mathcal{B}^2 induziert, also

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r, s) dr ds = 1$$

erfüllt ist. Ist f stetig auf \mathbb{R}^2 , dann besteht zwischen der Verteilungsfunktion und der Wahrscheinlichkeitsdichte folgender Zusammenhang

$$(2.86) \quad f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

BEISPIEL 2.33. Der Ort (X, Y) eines radioaktiven Teilchens sei im Einheitsquadrat gleichverteilt, die Aufenthaltswahrscheinlichkeit des Teilchens in Gebieten mit gleichem Flächeninhalt ist also gleich groß. Man bestimme $F(0.2, 0.4)$ und $P(0, 1 \leq X \leq 0, 3, 0 \leq Y \leq 0.5)$.

Dieses Problem kann durch das zweidimensionale Analogon der stetigen Gleichverteilung modelliert werden:

$$f(x, y) = \begin{cases} 1, & (x, y) \in [0, 1]^2 \\ 0, & \text{sonst.} \end{cases}$$

Man erhält

$$F(0.2, 0.4) = \int_{-\infty}^{0.2} \int_{-\infty}^{0.4} f(r, s) dr ds = \int_0^{0.2} \int_0^{0.4} 1 dr ds = 0,08$$

bzw.

$$P(0, 1 \leq X \leq 0,3, 0 \leq Y \leq 0.5) = \int_{0,1}^{0,3} \int_0^{0,5} 1 dy dx = 0,1.$$

Ähnlich zu diskreten Zufallsvariablen definieren wir die **kumulative Randverteilung** von X durch

$$(2.87) \quad F_1(x) = P(X \leq x, Y \in \mathbb{R}) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, y) dy du.$$

Ist die Dichte f stetig, ergibt sich man die **Randdichte** von X aus der gemeinsamen Dichte

$$(2.88) \quad f_1(x) = F_1'(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

12. Unabhängige Zufallsvariable

DEFINITION 2.19. *Diskrete oder stetige Zufallsvariable X_1, \dots, X_n heißen **unabhängig**, wenn ihre gemeinsame kumulative Verteilungsfunktion das Produkt der einzelnen kumulativen Randverteilungen ist:*

$$(2.89) \quad F(x_1, \dots, x_n) = F_1(x_1)F_2(x_2) \dots F_n(x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Für stetige Zufallsvariable ist die Unabhängigkeit äquivalent zu der Bedingung, daß die gemeinsame Wahrscheinlichkeitsdichte das Produkt der einzelnen Randdichten ist:

$$(2.90) \quad f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n).$$

Wir skizzieren den Beweis nur für zwei stetige Zufallsvariable X und Y . Sind X und Y unabhängig, findet man

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} F_1(x)F_2(y) = F_1'(x)F_2'(y) = f_1(x)f_2(y).$$

Gilt umgekehrt $f(x, y) = f_1(x)f_2(y)$, folgt für die gemeinsame kumulative Verteilungsfunktion

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du = \int_{-\infty}^x \int_{-\infty}^y f_1(u) f_2(v) dv du \\ &= \int_{-\infty}^x f_1(u) du \int_{-\infty}^y f_2(v) dv = F_1(x) F_2(y) \end{aligned}$$

Für diskrete Zufallsvariable X_1, \dots, X_N ist die Unabhängigkeit gleichwertig mit der Faktorisierung

$$(2.91) \quad p(x_1, \dots, x_n) = p_1(x_1) \dots p_n(x_n).$$

Sind X und Y unabhängige Zufallsvariable, dann sind auch $Z = g(X)$ und $W = h(Y)$ unabhängig. Ferner gilt für beliebige Borel Mengen A und B

$$(2.92) \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

BEISPIEL 2.34. Wir betrachten das Werfen eines roten und eines blauen Würfels. Die Zufallsvariable X sei die Augenzahl des roten Würfels, Y jene des blauen Würfels. Die gemeinsame Verteilung ist gegeben durch $p(i, j) = \frac{1}{36}$, $1 \leq i, j \leq 6$, die Randverteilungen $p_1(i) = p(i) = \frac{1}{6}$, $i = 1, \dots, 6$. Es folgt $p(i, j) = p_1(i)p_2(j)$, also sind X und Y unabhängig, was ja zu erwarten war.

13. Bedingte Verteilungen

13.1. Diskrete Zufallsvariable.

DEFINITION 2.20. Es seien X und Y diskrete Zufallsvariable mit gemeinsamer Wahrscheinlichkeitsverteilung $p(x, y)$ und marginalen Wahrscheinlichkeitsverteilungen p_X bzw. p_Y . Die **bedingte Wahrscheinlichkeitsverteilung von X unter $Y = y$** ist gegeben durch

$$(2.93) \quad p_{X|Y}(x|y) = \begin{cases} \frac{p(x, y)}{p_Y(y)} & \text{falls } p_Y(y) > 0 \\ 0 & \text{sonst,} \end{cases}$$

die bedingte Wahrscheinlichkeitsdichte von Y unter $X = x_i$

$$(2.94) \quad p_{Y|X}(y|x) = \begin{cases} \frac{p(x, y)}{p_X(x)} & \text{falls } p_X(x) > 0 \\ 0 & \text{sonst,} \end{cases}$$

BEISPIEL 2.35 (Fortsetzung von Beispiel 2.32). Die bedingte Wahrscheinlichkeitsverteilung von X unter $Y = 1$ ist gegeben durch

$$p_{X|Y}(0|1) = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}, \quad p_{X|Y}(1|1) = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}.$$

Löst man die Definition der bedingten Wahrscheinlichkeitsverteilung nach der gemeinsamen Verteilung auf, erhält man

$$(2.95) \quad p(x, y) = p_{X|Y}(x|y)p_Y(y).$$

Summiert man beide Seiten über alle zulässigen Werte von y , ergibt sich eine nützliche Variante des Satzes von der totalen Wahrscheinlichkeit

$$(2.96) \quad p_X(x) = \sum_y p(x, y) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

BEISPIEL 2.36. Auf Grund eines Defektes registriert ein Teilchendetektor einfallenden Partikel (unabhängig) mit einer Wahrscheinlichkeit p . Wenn die pro Zeiteinheit eintreffenden Partikel Poisson verteilt sind mit Erwartungswert λ , wie sind die vom Detektor registrierten Teilchen verteilt?

Lösung: Wir bezeichnen die Anzahl der tatsächlich einfallenden Teilchen mit N , die Anzahl der registrierten Partikel mit X . N ist Poisson verteilt mit Parameter λ . Wenn man weiß daß $N = n$ Partikel einfallen, dann ist X binomial, genauer $B(n, p)$, verteilt (n unabhängige Versuche mit Erfolgswahrscheinlichkeit p). Es gilt also $P(X = k|N = n) = \binom{n}{k} p^k (1-p)^{n-k}$. Nach dem Satz von der totalen Wahrscheinlichkeit folgt

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} P(X = k|N = n)P(N = n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{n!} \lambda^n e^{-\lambda} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \lambda^{n-k} \frac{(1-p)^{n-k}}{(n-k)!} = \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned}$$

Die Anzahl der registrierten Teilchen ist demnach ebenfalls Poisson verteilt mit Parameter λp . Es sind zahlreiche andere Formulierungen dieses Beispiels möglich: N kann auch die Anzahl der Verkehrsunfälle in einer bestimmten Zeitspanne sein, welche mit einer Wahrscheinlichkeit p einen letalen Ausgang haben.

13.2. Stetige Zufallsvariable.

DEFINITION 2.21. Es seien X und Y stetige Zufallsvariable und $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ die Dichte der gemeinsamen kumulativen Verteilungsfunktion. Die **bedingte Verteilung von X unter $Y = y$** ist gegeben durch

$$(2.97) \quad F_{X|Y}(x|y) = P(X \leq x|Y = y)$$

Das kontinuierliche Analogum zum Satz von der totalen Wahrscheinlichkeit 2.96 ist (ohne Beweis)

$$(2.98) \quad F_X(x) = \int_{-\infty}^{\infty} F_{X|Y}(x|y) f_Y(y) dy.$$

Andererseits gilt (formal, wegen der Vertauschung der Integrationsreihenfolge)

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(s) ds = \int_{-\infty}^x \int_{-\infty}^{\infty} f(s, t) dt ds \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^x f(s, t) ds dt \end{aligned}$$

Vergleicht man die beiden Darstellungen für F_X , findet man

$$F_{X|Y}(x|y)f_Y(y) = \int_{-\infty}^x f(s, y) ds$$

und damit eine explizite Darstellung der bedingten Verteilung von X unter $Y = y$

$$(2.99) \quad F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f(s, y)}{f_Y(y)} ds, \quad \text{falls } f_Y(y) > 0.$$

Diese Darstellung der bedingten Verteilung von X unter $Y = y$ motiviert folgenden Begriff

DEFINITION 2.22. *Es seien X und Y stetige Zufallsvariable, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ die Dichte der gemeinsamen kumulativen Verteilungsfunktion und f_X bzw. f_Y die Dichten der jeweiligen Randverteilungen. Die **bedingte Wahrscheinlichkeitsdichte von X unter $Y = y$** ist gegeben durch*

$$(2.100) \quad f_{X|Y}(x, y) = \begin{cases} \frac{f(x, y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ 0 & \text{sonst,} \end{cases}$$

die bedingte Wahrscheinlichkeitsdichte von Y unter $X = x$

$$(2.101) \quad f_{Y|X}(x, y) = \begin{cases} \frac{f(x, y)}{f_X(x)} & \text{falls } f_X(x) > 0 \\ 0 & \text{sonst,} \end{cases}$$

BEISPIEL 2.37. Zu Beginn eines jeden Tages enthält ein Getränkeautomat eine zufällige Menge von Y Liter eines bestimmten Getränkes. Während des Tages werden X Liter abgegeben. Da der Automat tagsüber nicht aufgefüllt wird, ist $X \leq Y$. Die gemeinsame Verteilung von X und Y ist erfahrungsgemäß

$$f(x, y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq y, 0 \leq y \leq 20 \\ 0 & \text{sonst} \end{cases}$$

Man bestimme die bedingte Dichte $f_{X|Y}$ unter $Y = y$ und die Wahrscheinlichkeit, daß tagsüber nicht mehrmals 5 Liter verkauft werden, sofern der Automat am Morgen noch 15 Liter enthielt.

Lösung: Die marginale Dichte f_Y ist gegeben durch

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx,$$

also

$$f_Y(y) = \begin{cases} \int_0^y \frac{1}{2} dx = \frac{1}{2}y, & 0 \leq y \leq 20 \\ 0 & \text{sonst} \end{cases}$$

und somit nach Definition 2.22

$$f_{X|Y}(x, y) = \begin{cases} \frac{f(x, y)}{f_Y(y)} & \text{falls } f_Y(y) > 0 \\ 0 & \text{sonst} \end{cases}$$

also

$$f_{X|Y}(x, y) = \begin{cases} \frac{1}{2} & 0 < x \leq y \leq 20 \\ 0 & \text{sonst.} \end{cases}$$

Die gesuchte Wahrscheinlichkeit ergibt sich demnach zu

$$P(X \leq 5 | Y = 15) = \int_{-\infty}^5 f(x | y = 15) dx = \int_0^5 \frac{1}{15} dx = \frac{1}{3}.$$

14. Erwartungswert einer Funktion von mehrdimensionalen Zufallsvariablen

DEFINITION 2.23. *Es sei $g: \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige Funktion.*

a) Sind X_1, \dots, X_n diskrete Zufallsvariable mit der gemeinsamen Wahrscheinlichkeitsverteilung p , dann ist der **Erwartungswert** von $g(X_1, \dots, X_n)$ definiert durch

$$(2.102) \quad E(g(X_1, \dots, X_n)) = \sum_{x_n} \cdots \sum_{x_1} g(x_1, \dots, x_n) p(x_1, \dots, x_n),$$

falls die Reihe absolut konvergent ist.

b) Sind X_1, \dots, X_n stetige Zufallsvariable mit der gemeinsamen Wahrscheinlichkeitsdichte f , dann ist der **Erwartungswert** von $g(X_1, \dots, X_n)$ definiert durch

$$(2.103) \quad E(g(X_1, \dots, X_n)) = \int \cdots \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

sofern

$$\int \cdots \int_{\mathbb{R}^n} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \cdots dx_n < \infty.$$

Die Varianz eines Zufallsvektors $Z = g(X_1, \dots, X_n)$ ist gegeben durch

$$V(Z) = E((Z - E(Z))^2) = E(Z^2) - E(Z)^2.$$

Es folgt unmittelbar aus der Definition, daß der Erwartungswert linear vom Zufallsvektor abhängt. Dies ist nicht der Fall bei der Varianz:

PROPOSITION 2.28. *Es sei X ein Zufallsvektor mit endlicher Varianz. Dann gilt*

$$(2.104) \quad V(\alpha X + \beta) = \alpha^2 V(X)$$

Der einfache Beweis sei dem Leser als Übung überlassen.

PROPOSITION 2.29. *Es seien X_1, \dots, X_n unabhängige Zufallsvariable mit gemeinsamer Wahrscheinlichkeitsdichte f und $g_i: \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, n$ meßbare Funktionen. Dann sind auch die Zufallsvariablen $g_1(X_1), \dots, g_n(X_n)$ unabhängig und es gilt*

$$(2.105) \quad E(g_1(X_1) \cdots g_n(X_n)) = E(g_1(X_1)) \cdots E(g_n(X_n)).$$

BEWEIS. Wir skizzieren den Beweis für zwei stetige unabhängige Zufallsvariable X und Y mit gemeinsamer Wahrscheinlichkeitsdichte f und den Randdichten f_X und f_Y . Die Unabhängigkeit von $g(X)$ und $h(Y)$ haben wir schon einmal festgestellt. Nach Definition 2.23 gilt

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x,y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x) dx \int_{-\infty}^{\infty} h(y)f_Y(y) dy = E(g(X))E(h(Y)). \end{aligned}$$

□

BEISPIEL 2.38. Ein Gastwirt lagert zu Beginn einer Woche X Liter Bier ein und schenkt im Laufe der Woche Y Liter aus. Die gemeinsame Dichte von X und Y sei

$$f(x,y) = \begin{cases} \frac{1}{x}, & 0 < y \leq x \leq 1, \\ 0 & \text{sonst} \end{cases}$$

Man bestimme den Erwartungswert der nicht verkauften Biermenge.

Lösung: Gesucht ist $E(X - Y)$. Setzt man in Definition 2.23 ein, findet man

$$E(X - Y) = \int_0^1 \left[\int_0^x \frac{x-y}{x} dy \right] dx = \int_0^1 \frac{x}{2} dx = \frac{1}{4}.$$

Nebenbemerkung: Die Zufallsvariablen X und Y sind nicht unabhängig!

15. Bedingte Erwartungswerte

DEFINITION 2.24. *Es seien X und Y zufällige Variable und $g: \mathbb{R} \rightarrow \mathbb{R}$ stetig. Der bedingte Erwartungswert von \mathbf{X} unter $\mathbf{Y} = \mathbf{y}$ ist definiert durch*

$$(2.106) \quad E(g(X)|Y = y) = \begin{cases} \sum_x g(x)p_{X|Y}(x|y), & X, Y \text{ diskrete Zufallsvariable,} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y) dx, & X, Y \text{ stetige Zufallsvariable,} \end{cases}$$

falls die Reihe bzw. das Integral absolut konvergent sind.

BEISPIEL 2.39 (Fortsetzung von Beispiel 2.37). Man bestimme den Erwartungswert der verkauften Getränkemenge, falls der Automat am Morgen noch 15 Liter enthält.

Lösung: Wir haben bereits die bedingte Wahrscheinlichkeitsdichte berechnet:

$$f_{X|Y}(x, y) = \begin{cases} \frac{1}{y} & 0 < x \leq y \leq 20 \\ 0 & \text{sonst.} \end{cases}$$

Definition 2.24 ergibt daher für jedes $y \in [0, 20]$

$$E(X|Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx = \int_0^y x \frac{1}{y} dx = \frac{y}{2},$$

und somit $E(X|Y = 15) = 7,5$.

Dieses Beispiel zeigt, daß der bedingte Erwartungswert $E(X|Y = y)$ eine Funktion von y definiert (wir setzen dabei implizit voraus, daß $E(X|Y = y)$ für alle Werte von Y existiert). Die Abbildung $y \rightarrow E(X|Y = y)$ hängt natürlich von der Zufallsvariablen Y ab und ist daher selbst eine Zufallsvariable, welche $E(X|Y)$ geschrieben wird. In Beispiel 2.39 haben wir $E(X|Y) = \frac{1}{2}Y$ berechnet. Wie für jede Zufallsvariable kann man auch für $E(X|Y)$ Erwartungswert und Varianz untersuchen

PROPOSITION 2.30 (Law of total expectation). *Es seien X und Y Zufallsvariable und es existiere $E(Y)$. Dann gilt*

$$(2.107) \quad E(X) = E_y[E_x(X|Y)],$$

(wir haben mit Indizes angedeutet, über welche Variablen bei der Bildung des Erwartungswertes summiert, bzw. integriert werden muß).

BEWEIS. Wir skizzieren den Beweis für stetige Zufallsvariable X, Y mit gemeinsamer Dichte f und den Randdichten f_X und f_Y . Es ist vielleicht natürlicher mit der linken Seite in (2.107) zu beginnen:

$$\begin{aligned} E_y(E_x(X|Y)) &= \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y) dy = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \stackrel{*}{=} \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx = E(X) \end{aligned}$$

Die mit $*$ markierte Vertauschung der Integrationsreihenfolge ist nach dem Satz von Tonelli gerechtfertigt, da das rechtsstehende Integral wegen der Existenz von $E(X)$ absolut konvergiert. Der Beweis für diskrete Zufallsvariable verläuft analog. \square

Wir untersuchen nun die Varianz von $E(X|Y)$. Dazu benötigen wir den Begriff der bedingten Varianz von X unter $Y = y$:

$$(2.108) \quad V(X|Y = y) \equiv E((X - E(X))^2|Y = y) = E(X^2|Y = y) - E(X|Y = y)^2,$$

welche wie vorhin als von Y abhängige Zufallsvariable $V(X|Y)$ aufgefaßt werden kann.

PROPOSITION 2.31. *Es seien X und Y Zufallsvariable und es existiere $V(Y)$. Dann gilt*

$$(2.109) \quad V(X) = E(V(X|Y)) + V(E(X|Y)).$$

BEWEIS. Die bedingte Varianz $V(X|Y)$ ist gegeben durch

$$V(X|Y) = E(X^2|Y) - E(X|Y)^2,$$

und hat daher den Erwartungswert

$$E[V(X|Y)] = E[E(X^2|Y)] - E[E(X|Y)^2].$$

Ferner gilt stets (wegen Proposition 2.30)

$$\begin{aligned} V[E(X|Y)] &= E[E(X|Y)^2] - E[E(X|Y)]^2, \\ V(X) &= E(X^2) - E(X)^2 = E[E(X^2|Y)] - E[E(X|Y)]^2 \\ &= E[V(X|Y)] + E[E(X|Y)^2] + V[E(X|Y)] - E[E(X|Y)]^2 \\ &= E[V(X|Y)] + V[E(X|Y)] \end{aligned}$$

□

BEISPIEL 2.40. An einer Produktionslinie wird ein bestimmtes Werkstück in grosser Tagesstückzahl hergestellt. Zur Qualitätssicherung wird täglich eine Stichprobe vom Umfang $n = 10$ untersucht. Die Wahrscheinlichkeit ein schadhaftes Stück zu finden sei p . Die Anzahl der schadhaften Werkstücke in der Stichprobe X ist daher $B(n, p)$ verteilt. Allerdings variiere p von Tag zu Tag, ist also selbst Wert einer Zufallsvariablen P . Unter der Annahme, daß P im Intervall $[0, \frac{1}{4}]$ gleichverteilt ist, bestimme man Erwartungswert und Varianz von X .

Lösung: Man weiß, $E(X|p) = np$ und $V(X|p) = npq$. Wegen des Gesetzes von der totalen Erwartung gilt

$$E(X) = E[E(X|P)] = E(nP) = nE(P) = n \frac{1}{2} \left(\frac{1}{4} - 0 \right) = \frac{n}{8},$$

für $n = 10$ ergibt sich somit $E(X) = \frac{5}{4}$. Die Varianz ergibt sich aus Proposition 2.31

$$V(X) = E[V(X|P)] + V(E(X|P)) = E(nPQ) + V(nP) = n(E(P) - E(P^2)) + n^2V(P).$$

Wir benötigen $E(P^2)$ und berechnen dies aus

$$E(P^2) = V(P) + E(P)^2 = \frac{1}{12} \frac{1}{4^2} + \frac{1}{64} = \frac{1}{48}.$$

Dies ergibt schließlich

$$V(X) = n \left(\frac{1}{8} - \frac{1}{48} \right) + n^2 \frac{1}{192} = \frac{5n}{48} + \frac{n^2}{192}.$$

Für $n = 10$ ergibt sich die Varianz $V(X) = 1,5625$.

BEISPIEL 2.41. Bei einer Versicherungsgesellschaft treffen täglich eine zufällige Anzahl von N Schadensmeldungen in der Höhe von jeweils X_i Schillingen ein, $i = 1, \dots, N$. Die Schadenshöhen haben jeweils den Erwartungswert $E(X_i) = \mu$. Das Management interessiert sich für die täglich zu erwartende Gesamtschadenssumme $T = \sum_{i=1}^N X_i$.

Lösung: Offenbar kann man leicht den bedingten Erwartungswert $E(T|N = n)$ berechnen:

$$E(T|N = n) = E\left(\sum_{i=1}^n X_i\right) = n\mu.$$

Also gilt $E(T|N) = \mu N$ und somit schließlich

$$E(T) = E[E(T|N)] = E(\mu N) = \mu E(N)$$

16. Kovarianz und Korrelation

DEFINITION 2.25. Es seien X und Y gemeinsam verteilte Zufallsvariable mit Erwartungswert $E(X)$ und $E(Y)$. Die **Kovarianz von X und Y** ist definiert durch

$$(2.110) \quad \text{Cov}(X, Y) \equiv \sigma_{XY} = E[(X - E(X))(Y - E(Y))],$$

sofern dieser Erwartungswert existiert.

Angenommen zwischen X und Y besteht ein linearer Zusammenhang $Y = aX + b$, $a \neq 0$. In diesem Falle ist das Produkt der Abweichungen $(x - E(X))(y - E(Y))$ stets positiv falls $a > 0$ und negativ falls $a < 0$ ist. Die Kovarianz bildet eine gewichtete Summe dieser Abweichungen. Da es wegen des konstanten Vorzeichens der Summanden zu keinen Auslöschungen kommen kann, ist in diesem Fall $|\text{Cov}(X, Y)|$ groß. Bilden die Daten (X, Y) hingegen eine mehr oder weniger regellose Punktwolke in der Ebene, dann wird das Produkt der Abweichungen für einige Punkte positiv, für einige negativ sein, ihr Mittel wird wegen der unvermeidlichen Auslöschungen dem Betrag nach klein sein.

Für die konkrete Rechnung ist oft folgende Darstellung der Kovarianz vorteilhaft:

PROPOSITION 2.32. Die Kovarianz gemeinsam verteilter Zufallsvariable X und Y ist gegeben durch

$$(2.111) \quad \text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Sind die Zufallsvariablen X und Y unabhängig, dann ist

$$\text{Cov}(X, Y) = 0.$$

BEWEIS. Diese Darstellung folgt aus der Linearität des Erwartungswertes:

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Für unabhängige Zufallsvariable gilt $E(XY) = E(X)E(Y)$ und somit $\text{Cov}(X, Y) = 0$. Die Umkehrung trifft allerdings nicht zu. \square

PROPOSITION 2.33 (Eigenschaften der Kovarianz). *Für die Kovarianz gemeinsam verteilter Zufallsvariable X, Y und Z gilt ($a, b \in \mathbb{R}$)*

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
2. $\text{Cov}(X + a, Y) = \text{Cov}(X, Y)$,
3. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$,
4. $\text{Cov}(X, Y + Z) = \text{Cov}(Y, X) + \text{Cov}(X, Z)$

Der einfache Beweis sei dem Leser als Übung überlassen. Die Abbildung $(X, Y) \rightarrow \text{Cov}(Y, X)$ ist also bilinear. Allgemeiner gilt

PROPOSITION 2.34. *Die Kovarianz der Zufallsvariablen $U = a + \sum_{i=1}^n b_i X_i$ und $V = c + \sum_{j=1}^m d_j Y_j$ beträgt*

$$(2.112) \quad \text{Cov}(U, V) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j).$$

Wegen $V(X) = \text{Cov}(X, X)$ folgt aus dieser Proposition sofort

$$(2.113) \quad V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y).$$

Anders als der Erwartungswert ist also die Varianz kein lineares Funktional. Allgemeiner gilt

COROLLARY 2.1.

$$(2.114) \quad V\left(a + \sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j \text{Cov}(X_i, X_j).$$

Insbesondere gilt demnach

$$(2.115) \quad V(a + bX) = b^2 V(X).$$

COROLLARY 2.2. *Für unabhängige Zufallsvariable X_1, \dots, X_n gilt*

$$(2.116) \quad V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i).$$

Die Kovarianz hat den Nachteil, daß sie nicht skalierungsinvariant ist: Ihr Wert kann durch eine geeignete Wahl der Einheit beliebige Werte annehmen, vgl. Proposition 2.33. Aus diesem Grunde mißt man die Abhängigkeit zwischen zwei Zufallsvariablen besser mit dem dimensionslosen **Korrelationskoeffizienten**

$$(2.117) \quad \rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

PROPOSITION 2.35. *Für den Korrelationskoeffizient gilt $|\rho| \leq 1$. Es ist $|\rho| = 1$ genau dann, wenn es Konstante a, b gibt, sodaß $P(Y = a + bX) = 1$ gilt.*

BEWEIS. Wir setzen $\sigma_X^2 = V(X)$ und $\sigma_Y^2 = V(Y)$. Da die Varianz einer Zufallsvariablen nicht negativ ist, folgt aus (2.113) und (2.115)

$$\begin{aligned} 0 \leq V\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) &= V\left(\frac{X}{\sigma_X}\right) + V\left(\frac{Y}{\sigma_Y}\right) \pm 2\frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} \\ &= \frac{V(X)}{\sigma_X^2} + \frac{V(Y)}{\sigma_Y^2} \pm 2\frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} = 2(1 \pm \rho) \end{aligned}$$

Die Ungleichung $1 \pm \rho \geq 0$ ist gleichwertig mit $|\rho| \leq 1$. Aus der eben hergeleiteten Abschätzung folgt aber auch

$$V\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = 0, \quad \text{falls } \rho = \mp 1.$$

Es ist nun intuitiv klar, daß die Werte einer Zufallsvariablen mit Wahrscheinlichkeit 1 in ihrem Erwartungswert konzentriert sind, falls ihre Varianz verschwindet, mit anderen Worten

$$P\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y} = c\right) = 1$$

mit $c = E\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right)$. □

Wir präzisieren nun das letzte Argument des vorausgehenden Beweises:

PROPOSITION 2.36. *Es sei X eine Zufallsvariable mit Erwartungswert $E(X) = \mu$. Ist $V(X) = 0$, dann gilt $P(X = \mu) = 1$.*

BEWEIS. Die Ungleichung von Tschebyscheff 2.8 kann auch in der Form

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

geschrieben werden (gleicher Beweis). Da $\sigma = 0$ ist und daher $\varepsilon > 0$ beliebig klein gewählt werden kann, ist die Behauptung bewiesen. □

17. Funktionen von mehrdimensionalen Zufallsvariablen

17.1. Summe und Quotient. Wir beschränken uns auf zwei gemeinsam verteilte Zufallsvariable X und Y und betrachten zuerst den diskreten Fall. Die gemeinsame Wahrscheinlichkeitsverteilung sei p und $Z = X + Y$. Die Wahrscheinlichkeitsverteilung von Z ergibt sich aus der Beobachtung, daß $Z = z$ gleichwertig ist mit $X = x$ und $Y = z - x$. Summiert man über alle möglichen Werte von X auf ergibt sich

$$p_Z(z) = P(Z = z) = \sum_x p(x, z - x).$$

Sind die Zufallsvariablen X und Y unabhängig, gilt also $p(x, y) = p_X(x)p_Y(y)$, erhält man

$$(2.118) \quad p_Z(z) = \sum_x p_X(x)p_Y(z - x).$$

Durch (2.118) wird wieder ein diskretes Wahrscheinlichkeitsmaß definiert, welches man die **Faltung** von P_X und P_Y nennt.

Von besonderer Bedeutung ist dieser Begriff bei stetigen Zufallsvariablen: die kumulative Verteilungsfunktion von $Z = X + Y$ ergibt sich in diesem Falle aus

$$F_Z(z) = P(X + Y \leq z) = \iint_{R_z} f(x, y) \, dx dy,$$

wobei der Integrationsbereich gegeben ist durch den Halbraum

$$R_z = \{(x, y) : x + y \leq z\}$$

gegeben ist. Setzt man dies in F_Z ein und substituiert man im inneren Integral $\nu = y + x$ erhält man

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) \, dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^z f(x, \nu - x) \, d\nu dx \\ &\stackrel{*}{=} \int_{-\infty}^z \int_{-\infty}^{\infty} f(x, \nu - x) \, dx d\nu. \end{aligned}$$

Die Vertauschung der Integrationsreihenfolge in $\stackrel{*}{=}$ kann mit dem Satz von Fubini gerechtfertigt werden. Ist z.B. die Dichte f stetig, ergibt sich weiters

$$f_Z(z) = F_Z(z)' = \int_{-\infty}^{\infty} f(x, z - x) \, dx,$$

und somit schließlich für unabhängige Zufallsvariable X und Y

$$(2.119) \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) \, dx.$$

Man nennt f_Z die **Faltung** von f_X und f_Y .

BEISPIEL 2.42. Die Lebensdauer eines Bauteils sei exponentiell mit Parameter $\lambda > 0$ verteilt. Aus Sicherheitsgründen wurde in das System eine identische, unabhängige Kopie eingebaut, sodaß das System so lange funktioniert, so lange einer der beiden Bauteile funktionstüchtig ist. Die gesamte Lebensdauer des Systems ist somit die Summe von zwei unabhängigen, exponentiell verteilten Zufallsvariablen, $S = T_1 + T_2$. Nach (2.119) ist die Wahrscheinlichkeitsdichte von f_S gegeben durch $f_S(s) = \int_{-\infty}^{\infty} f_{T_1}(x) f_{T_2}(s - x) \, dx$. Der Integrand ist genau dann ungleich Null, wenn $x > 0$ und $s - x > 0$ ist. Dies führt zu

$$f_S(s) = \int_0^s \lambda e^{-\lambda x} e^{-\lambda(s-x)} \, dx = \lambda^2 \int_0^s e^{-\lambda s} \, dx = \lambda^2 s e^{-\lambda s}.$$

Die Lebensdauer des Gesamtsystems ist also unter diesen Voraussetzungen Gamma verteilt mit den Parametern $\alpha = 2$ und λ .

PROPOSITION 2.37. *Es seien X_i , $i = 1, \dots, n$ unabhängige $N(\mu_i, \sigma_i)$ verteilte Zufallsvariable. Dann ist $U = \sum_{i=1}^n \alpha_i X_i$, $\sum_{i=1}^n \alpha_i^2 > 0$, ebenfalls normalverteilt, und zwar gilt*

$$(2.120) \quad \begin{aligned} E(U) &= \sum_{i=1}^n \alpha_i \mu_i, \\ V(U) &= \sum_{i=1}^n \alpha_i^2 \sigma_i^2. \end{aligned}$$

BEWEIS. Es genügt, folgende Eigenschaften für normal verteilte Zufallsvariable nachzuweisen: 1) wenn X einer $N(\mu, \sigma)$ -Verteilung folgt, dann ist αX ebenfalls normalverteilt mit $E(\alpha X) = \alpha \mu$ und $V(\alpha X) = \alpha^2 \sigma^2$, 2) wenn X und Y unabhängige $N(\mu_X, \sigma_X)$ - bzw. $N(\mu_Y, \sigma_Y)$ - verteilte Zufallsvariable sind, dann ist $Z = X + Y$ normalverteilt mit $E(X + Y) = \mu_X + \mu_Y$ und $V(X + Y) = \sigma_X^2 + \sigma_Y^2$. Den Beweis der ersten Behauptung überlassen wir dem Leser als Übung. Die Essenz des Beweises der zweiten Behauptung ist der Nachweis, daß die Summe unabhängiger normalverteilter Zufallsvariablen selbst wieder normalverteilt ist. Die Dichte von Z ist nach (2.119) gegeben durch das Faltungsintegral

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} e^{-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}} dx \\ &= \frac{1}{2\pi\sigma_Y} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} e^{-\frac{(z-\sigma_X u - \mu_X - \mu_Y)^2}{2\sigma_Y^2}} du, \end{aligned}$$

wobei wir im letzten Integral $x = \sigma_X u + \mu_X$ substituiert haben. Nach einigen algebraischen Umformungen, kann man den Exponenten in folger Form schreiben

$$\begin{aligned} &\frac{u^2}{2} + \frac{1}{2\sigma_Y^2} (z - \sigma_X u - \mu_X - \mu_Y)^2 \\ &= \frac{1}{2\sigma_Y^2} \left(\sqrt{\sigma_X^2 + \sigma_Y^2} u - \frac{\sigma_X (z - (\mu_X + \mu_Y))}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \right)^2 + \frac{(z - (\mu_X + \mu_Y))^2}{\sigma_X^2 + \sigma_Y^2}. \end{aligned}$$

Setzt man diesen Ausdruck in f_Z ein, bekommt man nach kurzer Rechnung

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_X^2 + \sigma_Y^2}} e^{-\frac{(z - (\mu_X + \mu_Y))^2}{2(\sigma_X^2 + \sigma_Y^2)}},$$

dies ist die Dichte einer $N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$ verteilten Zufallsvariablen. \square

Wir betrachten nun den Quotienten zweier stetiger Zufallsvariabler $Z = \frac{Y}{X}$. Nun gilt

$$\begin{aligned} F_Z(z) &= P\left(\frac{Y}{X} \leq z\right) = P(\{(x, y) : \frac{y}{x} \leq z\}) \\ &= \begin{cases} P(\{(x, y) : y \leq zx\}) & \text{falls } x > 0 \\ P(\{(x, y) : y \geq zx\}) & \text{falls } x < 0 \end{cases} \end{aligned}$$

also

$$F_Z(z) = \int_{-\infty}^0 \int_{xz}^{\infty} f(x, y) \, dy dx + \int_0^{\infty} \int_{-\infty}^{xz} f(x, y) \, dy dx.$$

Um die Abhängigkeit der Grenzen des inneren Integrals von x zu eliminieren, substituieren wir $y = x\nu$ und erhalten

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^0 \int_z^{-\infty} x f(x, \nu x) \, d\nu dx + \int_0^{\infty} \int_{-\infty}^z x f(x, \nu x) \, d\nu dx \\ &= \int_{-\infty}^0 \int_{-\infty}^z (-x) f(x, \nu x) \, d\nu dx + \int_0^{\infty} \int_{-\infty}^z x f(x, \nu x) \, d\nu dx \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} |x| f(x, x\nu) \, dx d\nu. \end{aligned}$$

und daraus durch Ableitung

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz) \, dx.$$

Wenn insbesondere die Zufallsvariablen X und Y unabhängig sind, folgt

$$(2.121) \quad f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) \, dx.$$

BEISPIEL 2.43. Wir wenden nun (2.121) auf unabhängige, standard normalverteilte Zufallsvariable an. Wegen (2.66) ergibt (2.121)

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| e^{-x^2/2} e^{-x^2 z^2/2} \, dx = \frac{1}{\pi} \int_0^{\infty} x e^{-x^2(z^2+1)/2} \, dx \\ &= \frac{1}{2\pi} \int_0^{\infty} e^{-u(z^2+1)/2} \, du. \end{aligned}$$

Eine einfache Integration führt auf

$$(2.122) \quad f_Z(z) = \frac{1}{\pi(z^2 + 1)}, \quad z \in \mathbb{R},$$

die Dichte der Cauchy Verteilung. Zusammenfassend halten wir fest: der Quotient $Z = \frac{Y}{X}$ unabhängiger, standard normalverteilter Zufallsvariablen X und Y ist Cauchy verteilt. Da das uneigentliche Integral

$$\int_{-\infty}^{\infty} \frac{z}{\pi(z^2 + 1)} dz$$

divergiert, besitzt Z keinen Erwartungswert!

18. Ordnungsstatistik

In manchen Fällen interessiert man sich für die relativen Größen von Zufallsvariablen: die schnellste Rundenzeit in einem Autorennen, der größte Weite beim Skispringen, etc. Wir betrachten also unabhängige, identisch verteilte, stetige Zufallsvariable X_1, \dots, X_n . Die Verteilungsfunktion jeder Zufallsvariablen sei F , die Dichte f . Die der Größe nach geordneten Zufallsvariablen bezeichnen wir $X_{(1)}, \dots, X_{(n)}$, also

$$X_{(1)} \leq X_{(2)} \leq \dots, \leq X_{(n)}.$$

Insbesondere ist also $X_{(1)} = \min\{X_1, \dots, X_n\}$ und $X_{(n)} = \max\{X_1, \dots, X_n\}$. Die geordneten Zufallsvariablen nennt man auch **Ordnungsstatistik**. Die Verteilungsfunktion $F_{(n)}$ von $X_{(n)}$ ergibt sich aus folgender einfacher Überlegung: $X_{(n)} \leq x$ ist gleichwertig mit dem simultanen Eintreten der Ereignisse $X_i \leq x$, $i = 1 \dots, n$. Wegen der Unabhängigkeit dieser Ereignisse als Folge der Unabhängigkeit der Zufallsvariablen findet man

$$\begin{aligned} F_{(n)}(x) &= P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) = F(x)^n \end{aligned}$$

Bezeichnet man die Dichte von $X_{(n)}$ mit $g_{(n)}$, erhält man

$$(2.123) \quad g_{(n)} = nF(x)^{n-1}f(x).$$

Die Dichte $g_{(1)}$ der Verteilungsfunktion $F_{(1)}$ von $X_{(1)}$ bestimmt man auf ähnliche Weise.

$$F_{(1)}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x).$$

Da $X_{(1)}$ das Minimum von X_1, \dots, X_n ist, tritt das Ereignis $P(X_{(1)} > x)$ genau dann ein, wenn $X_i > x$, $i = 1 \dots, n$. Wegen $P(X_i > x) = 1 - F(x)$ zeigt ein analoges Argument wie vorhin

$$(2.124) \quad \begin{aligned} F_{(1)}(x) &= 1 - P(X_1 > x, \dots, X_n > x) = 1 - [P(X_1 > x) \cdots P(X_n > x)] \\ &= 1 - [1 - F(x)]^n \end{aligned}$$

also

$$(2.125) \quad g_{(1)}(x) = n[1 - F(x)]^{n-1}f(x).$$

BEISPIEL 2.44. Die Lebensdauer (in Stunden) gewisser elektronischer Komponenten ist exponentiell verteilt mit Dichte

$$f(x) = \begin{cases} \frac{1}{100}e^{-x/100}, & x > 0 \\ 0, & \text{sonst.} \end{cases}$$

a) Angenommen zwei derartige Komponenten operieren unabhängig voneinander und sind in Serie geschaltet. Man bestimme die Dichte l der Verteilung der Lebensdauer L des gesamten Systems.

Lösung: Das System fällt aus, wenn eine der beiden Komponenten ausfällt, also $L = \min\{X_1, X_2\}$, wobei X_i unabhängige Zufallsvariable mit der angegebenen Wahrscheinlichkeitsdichte sind. Die kumulative Verteilungsfunktion ist $F(x) = 1 - e^{-x/100}$, $x \geq 0$, also nach (2.124)

$$l(x) = g_{(1)}(x) = \begin{cases} 2e^{-x/100} \frac{1}{100}e^{-x/100}, & x > 0 \\ 0, & \text{sonst,} \end{cases}$$

somit

$$l(x) = \begin{cases} \frac{1}{50}e^{-x/50} & x > 0 \\ 0, & \text{sonst.} \end{cases}$$

Das Minimum zweier exponentiell verteilter Zufallsvariablen ist demnach wieder exponentiell verteilt. Man beachte, daß die mittlere Lebensdauer der Einzelkomponenten 100 Stunden beträgt, jene des in Serie geschalteten Systems jedoch nur 50 Stunden.

b) Wie oben, jedoch werden die Komponenten nun parallel geschaltet.

Lösung: In diesem Fall fällt das System genau dann aus, wenn beide Komponenten defekt sind. Nun ist also $L = \max\{X_1, X_2\}$. Eine einfache Anwendung von (2.123) ergibt

$$l(x) = g_{(2)} = 2F(x)f(x) = \begin{cases} \frac{1}{50}(e^{-x/100} - e^{-x/50}) & x > 0 \\ 0, & \text{sonst.} \end{cases}$$

Das Maximum von exponentiell verteilten Zufallsvariablen ist also selbst nicht wieder exponentiell verteilt. Die mittlere Lebensdauer des gesamten Systems beträgt nun 150 Stunden.

19. Momenterzeugende Funktion

DEFINITION 2.26. Die **Momenterzeugende Funktion** einer Zufallsvariablen X ist definiert durch $M(t) = E(e^{Xt})$, sofern dieser Erwartungswert existiert. Es folgt

$$(2.126) \quad M(t) = \begin{cases} \sum_x e^{tx} p(x), & \text{für } X \text{ diskret} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{für } X \text{ stetig.} \end{cases}$$

PROPOSITION 2.38. *Wenn die Momenterzeugende Funktion M einer Zufallsvariablen X auf einem offenen Intervall, welches Null enthält, existiert, dann ist die Wahrscheinlichkeitsverteilung von X durch M eindeutig bestimmt.*

Wenn also zwei Zufallsvariable auf einem offenen Intervall, welches Null enthält, dieselbe Momenterzeugende Funktion haben, dann besitzen sie auch dieselbe Verteilung. Der Beweis dieses Satzes erfordert Kenntnisse aus der Theorie der Laplacetransformation und kann daher hier nicht besprochen werden.

Die Bezeichnung Momenterzeugende Funktion deutet auf den engen Zusammenhang mit den Momenten einer Zufallsvariablen hin. Man nennt $\mathbf{E}(\mathbf{X}^r)$ **r -tes Moment** von X , bzw $\mathbf{E}((\mathbf{X} - \mathbf{E}(\mathbf{X}))^r)$ das **r -te zentrale Moment**. Der Erwartungswert einer Zufallsvariablen ist demnach das 1. Moment, die Varianz das zentrale 2. Moment. Bildet man die k -te Ableitung der Momenterzeugende Funktion

$$\frac{d^k}{dt^k}M(t) = \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} x^k e^{tx} f(x) dx$$

(die Vertauschung von Integration und Differentiation kann mit der Existenz von M gerechtfertigt werden) und wertet die Ableitung für $t = 0$ aus, erhält man

$$M^{(k)}(0) = \int_{-\infty}^{\infty} x^k f(x) dx = E(X^k),$$

also das k -te Moment von X . Wir fassen zusammen:

PROPOSITION 2.39. *Wenn die Momenterzeugende Funktion M einer Zufallsvariablen X auf einem offenen Intervall, welches Null enthält, existiert, dann ist M beliebig oft differenzierbar und es gilt $M^{(k)}(0) = E(X^k)$, $k \in \mathbb{N}$.*

Für viele Verteilungen kann man die Momenterzeugende Funktion explizit angeben. Die Berechnung der Momente von X reduziert sich wegen Proposition 2.39 auf eine simple Ableitung.

Verteilung	$M(t)$
$B(n, p)$	$(pe^1 + q)^n$
geometrisch	$\frac{pe^t}{1-qe^t}$
Poisson	$e^{\lambda(e^t-1)}$
Gamma(α, λ)	$(\frac{\lambda}{\lambda-t})^\alpha$
$N(0, 1)$	$e^{t^2/2}$

TABELLE 2.6. Momenterzeugende Funktionen

Weitere nützliche Eigenschaften der Momenterzeugende Funktion sind:

PROPOSITION 2.40. 1.) Die Zufallsvariable X habe die Momenterzeugende Funktion M_X und es sei $Y = \alpha + \beta X$. Dann hat Y die Momenterzeugende Funktion

$$M_Y(t) = e^{\alpha t} M_X(\beta t).$$

2.) Es seien X und Y unabhängige Zufallsvariable mit Momenterzeugenden Funktionen M_X und M_Y . Dann besitzt auch $Z = X + Y$ eine Momenterzeugende Funktion M_Z , welche auf dem Durchschnitt der Existenzintervalle von M_X und M_Y existiert:

$$M_Z(t) = M_X(t)M_Y(t).$$

BEWEIS. Die erste Behauptung folgt aus der Linearität des Erwartungswertes

$$M_Y(t) = E(e^{tY}) = E(e^{(\alpha + \beta X)t}) = E(e^{\alpha t} e^{\beta X t}) = e^{\alpha t} E(e^{\beta X t}) = e^{\alpha t} M_X(\beta t).$$

Die zweite Behauptung ergibt sich auf analoge Weise, wenn man berücksichtigt, daß $E(e^{tX} e^{tY}) = E(e^{tX})E(e^{tY})$ als Folge der stochastischen Unabhängigkeit von X und Y gilt. \square

BEISPIEL 2.45. Es seien X und Y unabhängige, Gamma verteilte Zufallsvariable, X folge einer Gammaverteilung mit den Parametern α und λ , die Parameter der Verteilung von Y seien β und λ . Nach Proposition 2.40 ist die Momenterzeugende Funktion von $Z = X + Y$ gegeben durch

$$M_Z(t) = M_X(t)M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha \left(\frac{\lambda}{\lambda - t}\right)^\beta = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha + \beta}$$

Wir schließen mit Proposition 2.38, daß die Summe zweier Gamma verteilter Zufallsvariablen mit Parametern (α, λ) bzw. (β, λ) wieder Gamma verteilt ist, und zwar mit den Parametern $(\alpha + \beta, \lambda)$.

20. Testverteilungen

Wir haben bisher Verteilungsfunktionen betrachtet, welche für die mathematische Modellierung bestimmter Zufallsexperimente entworfen wurden. Wir wenden uns nun den Testverteilungen (Prüfverteilungen) zu, welche die Grundlage statistischer Tests bilden.

20.1. χ^2 -Verteilung. Wir haben in Definition 2.14 festgelegt, daß das Quadrat einer standardnormalverteilten Zufallsvariablen χ_1^2 -verteilt (1 Freiheitsgrad) ist. Wir haben gesehen, daß die Dichte der χ_1^2 -Verteilung durch die Dichte der Gammaverteilung mit $\alpha = \lambda = \frac{1}{2}$ gegeben ist. Wir betrachten nun die etwas allgemeinere Situation

DEFINITION 2.27. Es seien U_1, \dots, U_n unabhängige χ^2 verteilte Zufallsvariable mit je einem Freiheitsgrad. Die Verteilung von $V = U_1 + \dots + U_n$ heißt χ^2 -Verteilung mit n Freiheitsgraden und wird χ_n^2 bezeichnet.

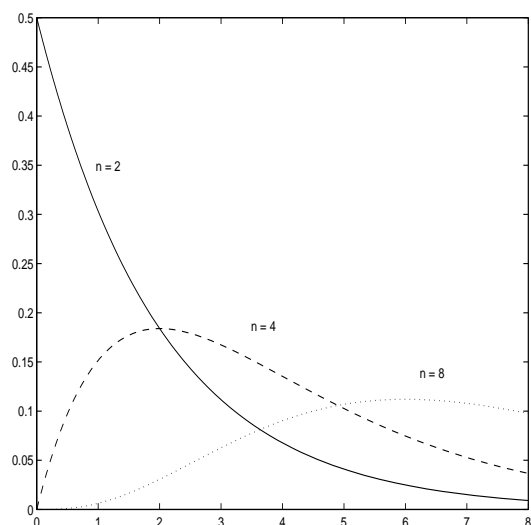
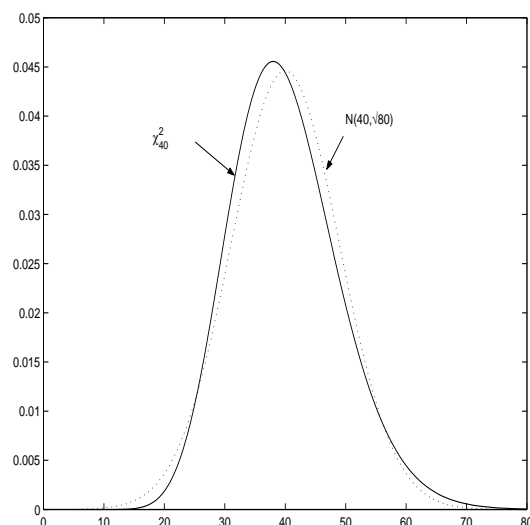
ABB. 2.12. χ^2 Dichte

ABB. 2.13. Normalapproximation

Da die Zufallsvariablen U_i Gamma verteilt mit den Parametern $(\alpha, \lambda) = (\frac{1}{2}, \frac{1}{2})$ sind, folgt aus Beispiel 2.45, daß V ebenfalls Gamma verteilt und zwar mit den Parametern $(\frac{n}{2}, \frac{1}{2})$ ist. Somit erhalten wir für die Wahrscheinlichkeitsdichte und die Momenterzeugende Funktion der χ_n^2 -Verteilung

$$(2.127) \quad f(x) = \frac{1}{2^n \Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad x \geq 0,$$

$$M_V(t) = (1 - 2t)^{-n/2}.$$

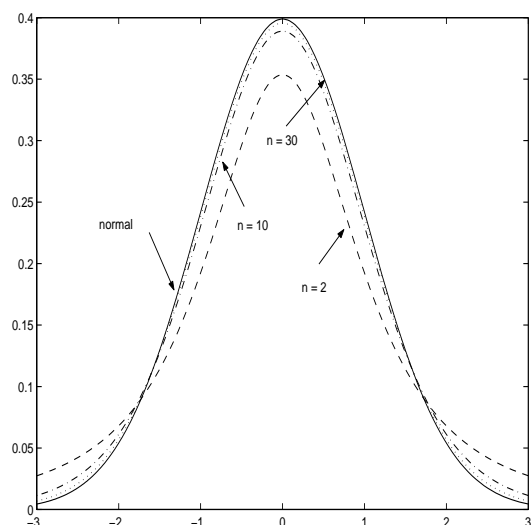
Mit Hilfe von Proposition 2.39 findet man

$$(2.128) \quad E(V) = n, \quad \sigma_V^2 = 2n.$$

Die χ_n^2 -Verteilung besitzt folgende bemerkenswerte Eigenschaft: wenn U einer χ_n^2 -Verteilung, V einer χ_m^2 -Verteilung folgt, dann besitzt $U + V$ eine χ_{n+m}^2 -Verteilung. Abbildung 2.13 zeigt, daß sich die χ_n^2 -Verteilung für große n brauchbar durch die Normalverteilung $N(n, \sqrt{2n})$ approximieren läßt

20.2. t -Verteilung. Die t -Verteilung wurde von W.S.Gosset eingeführt, der unter dem Pseudonym "Student" veröffentlichte. Die t -Verteilung heißt daher auch **Student Verteilung**.

DEFINITION 2.28. Es sei X standard normalverteilt und U folge einer χ_n^2 Verteilung, $n \in \mathbb{N}$. Die Verteilung von $T = \frac{X}{\sqrt{U/n}}$ heißt **t -Verteilung mit n Freiheitsgraden**, kurz t_n -Verteilung.

ABB. 2.14. t -Dichte

Es sei dem Leser als Übung überlassen, als Anwendung von (2.121) die Dichte der t_n -Verteilung zu berechnen. Man erhält nach längerer Rechnung

$$(2.129) \quad f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad t \in \mathbb{R}.$$

Die Dichte ist eine gerade Funktion und besitzt eine große Ähnlichkeit mit der Dichte der Standardnormalverteilung.

20.3. F -Verteilung.

DEFINITION 2.29. *Es seien U und V unabhängige χ_m^2 bzw. χ_n^2 -verteilte Zufallsvariable. Die Dichte von*

$$W = \frac{U/m}{V/n}$$

heißt **F -Verteilung mit m und n Freiheitsgraden**, kurz: $F_{m,n}$.

Man überzeuge sich davon, daß die Dichte der $F_{m,n}$ -Verteilung gegeben ist durch

$$(2.130) \quad f_W(w) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2}.$$

21. Grenzwertsätze

21.1. Gesetz der großen Zahlen. In diesem Abschnitt werden wir unsere naive Vorstellung vom Begriff der Wahrscheinlichkeit theoretisch untermauern: Wir sind von der Vorstellung ausgegangen, daß beim wiederholten Werfen einer Münze die relative Anzahl von Köpfen gegen $p = \frac{1}{2}$ "strebt", wenn die Anzahl der Wiederholungen

über alle Grenzen wächst. Angenommen wir werfen die Münze n mal. Die einzelnen Würfe können als unabhängige Zufallsvariable X_i , $i = 1, \dots, n$, modelliert werden. Es sei $X_i = 1$, wenn der i -te Wurf Kopf zeigt, andernfalls sei $X_i = 0$. Die relative Häufigkeit von "Kopf" bei n Würfungen ist daher

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Das Gesetz der großen Zahlen stellt fest, daß \bar{X}_n tatsächlich in einem bestimmten Sinne gegen $\frac{1}{2}$ konvergiert.

THEOREM 2.9 (Gesetz der großen Zahlen). *Es seien X_i , $i = 1, \dots, n$, unabhängige Zufallsvariable mit $E(X_i) = \mu$ und $V(X_i) = \sigma^2$, $i = 1, \dots, n$. Ferner sei $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Dann gilt für jedes $\varepsilon > 0$*

$$(2.131) \quad \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

BEWEIS. Wegen der Linearität des Erwartungswertes ist $E(\bar{X}_n) = \mu$, wegen der Unabhängigkeit der Zufallsvariablen X_i folgt aus (2.115) und Proposition 2.33

$$V(\bar{X}_n) = \sum_{i=1}^n V\left(\frac{X_i}{n}\right) = \sum_{i=1}^n \frac{1}{n^2} V(X_i) = \frac{\sigma^2}{n}.$$

Die Behauptung folgt nun aus der Ungleichung von Tschebyscheff 2.8:

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

□

Eine Folge von Zufallsvariablen (X_i) , welche derart gegen eine andere Zufallsvariable X konvergiert, daß für jedes $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

gilt, heißt **stochastisch gegen X konvergent**. Neben der stochastischen Konvergenz gibt es einen stärkeren Konvergenzbegriff, die **Konvergenz fast überall** oder **fast sichere Konvergenz**. Fast sichere Konvergenz von (X_n) gegen X liegt vor, wenn

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1,$$

mit anderen Worten, $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ kann höchstens auf einer Menge mit Wahrscheinlichkeit Null verletzt sein. Manchmal nennt man das Gesetz der großen Zahlen, welches nur die stochastische Konvergenz von \bar{X}_n sicher stellt, schwaches Gesetz der großen Zahlen. Es gibt auch ein starkes Gesetz der großen Zahlen, welches unter denselben Voraussetzungen die fast sichere Konvergenz von \bar{X}_n zum Inhalt hat.

21.2. Der zentrale Grenzwertsatz. Wir betrachten nun eine Folge von Zufallsvariablen X_n mit Verteilungsfunktionen F_n . Die Verteilungsfunktionen sind eindeutig bestimmt durch ihre Momenterzeugenden Funktionen M_n . Angenommen, die Momenterzeugenden Funktionen konvergieren gegen eine Funktion M . Was kann man über F_n aussagen? Ohne Beweis zitieren wir folgendes Resultat:

THEOREM 2.10 (Stetigkeitssatz). *Es sei (F_n) eine Folge von Verteilungsfunktionen und (M_n) die Folge ihrer Momenterzeugenden Funktionen. Ferner sei F eine Verteilungsfunktion mit Momenterzeugender Funktion M . Wenn $\lim_{n \rightarrow \infty} M_n(t) = M(t)$ auf einem offenen Intervall welches Null enthält, gilt, dann folgt $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ in allen Stetigkeitsstellen von F .*

BEISPIEL 2.46 (Normalapproximation der Poissonverteilung). Es sei (X_n) eine Folge Poissonverteilter Zufallsvariable mit Mittelwert λ_n und λ_n divergiere gegen ∞ . Es gilt also $E(X_n) = V(X_n) = \lambda_n$. Wegen der Divergenz der Erwartungswerte und Varianz wäre es aussichtslos, X_n durch eine Zufallsvariable mit endlichem Erwartungswert und endlicher Varianz zu approximieren. Diese Schwierigkeit kann man durch Übergang auf die standardisierten Variablen umschiffen. Wir betrachten daher anstelle von X_n die Zufallsvariablen

$$(2.132) \quad Z_n = \frac{X_n - E(X_n)}{\sqrt{V(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}.$$

Eine einfache Rechnung zeigt $E(Z_n) = 0$, $V(Z_n) = 1$. Es sei M_{X_n} die Momenterzeugende Funktion von X_n . Mit Hilfe von Proposition 2.40 findet man

$$(2.133) \quad M_{Z_n}(t) = e^{-E(X_n)t/\sqrt{V(X_n)}} M_{X_n}\left(\frac{t}{\sqrt{V(X_n)}}\right),$$

in diesem Beispiel also ($M_{X_n}(t) = e^{\lambda_n(e^t-1)}$ aus Tabelle 19)

$$M_{Z_n}(t) = e^{-\sqrt{\lambda_n}t} e^{\lambda_n(e^{t/\sqrt{\lambda_n}}-1)}.$$

Man überzeuge sich von

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2},$$

das ist die Momenterzeugende Funktion der Standardnormalverteilung. Satz 2.46 garantiert nun die punktweise Konvergenz von F_n gegen Φ .

Wir wenden uns nun dem zentralen Grenzwertsatz in seiner einfachsten Form zu. Es seien nun X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariable mit Mittelwert μ und Varianz σ^2 . Setzt man

$$S_n = \sum_{i=1}^n X_i,$$

dann konvergiert nach dem Gesetz der großen Zahlen $\frac{S_n}{n}$ stochastisch gegen μ . Der zentrale Grenzwertsatz gibt Auskunft, wie $\frac{S_n}{n}$ um μ schwankt. Um diese Schwankungen zu analysieren geht man wieder auf standardisierte Zufallsvariable über:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Der zentrale Grenzwertsatz stellt fest, daß die Verteilung von Z_n gegen die Standardnormalverteilung konvergiert:

THEOREM 2.11 (Zentraler Grenzwertsatz). *Es sei (X_n) eine Folge unabhängiger, identisch verteilter Zufallsvariable mit Mittelwert 0 und Varianz σ^2 . Die Momenterzeugende Funktion der Verteilungsfunktion F sei in einer Umgebung von Null definiert. Dann gilt*

$$(2.134) \quad \lim_{n \rightarrow \infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad x \in \mathbb{R}$$

mit

$$S_n = \sum_{i=1}^n X_i.$$

BEWEISSKIZZE. Wir setzen $Z_n = \frac{S_n}{\sigma\sqrt{n}}$ und zeigen, daß die Momenterzeugende Funktion von Z_n gegen die Momenterzeugende Funktion der Standardnormalverteilung konvergiert. Da S_n eine Summe unabhängiger, identisch verteilter Zufallsvariable ist, folgt mit Proposition 2.40

$$M_{S_n}(t) = M(t)^n,$$

und daher

$$M_{Z_n}(t) = M\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

Man kann zeigen, daß eine Momenterzeugende Funktion um Null in eine Potenzreihe entwickelt werden kann. Insbesondere gilt

$$M(s) = M(0) + sM'(0) + \frac{s^2}{2}M''(0) + o(s^2).$$

Nach Voraussetzung ist $E(X_n) = M'(0) = 0$ und $V(X_n) = M''(0) = \sigma^2$, also

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)$$

und daher

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right)^n.$$

Berücksichtigt man nun noch den elementaren Grenzwert

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^{\lim_{n \rightarrow \infty} a_n},$$

soferne $\lim_{n \rightarrow \infty} a_n$ existiert und

$$\lim_{n \rightarrow \infty} \frac{o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)}{\left(\frac{t}{\sigma\sqrt{n}}\right)^2} = 0,$$

erhält man

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2}, \quad t \in \mathbb{R},$$

das ist die Momenterzeugende Funktion der Standardnormalverteilung. \square

BEISPIEL 2.47. Wir betrachten in $[0, 1]$ gleichverteilte Zufallsvariable X_1, \dots, X_n . Wegen $E(X_i) = \frac{1}{2}$ und $V(X_i) = \frac{1}{12}$ sind die standardisierten Summen gegeben durch $Z_n = \frac{S_n - n/2}{\sqrt{n/12}}$. Nach dem zentralen Grenzwertsatz konvergieren die Verteilungen von Z_n gegen Φ . Abbildung 2.15 zeigt ein Histogramm von 1000 derartiger Summen Z_n für $n = 12$. Trotz des geringen Stichprobenumfanges (12!) ist die Übereinstimmung bereits überraschend gut. Der Zusammenhang der Abbildung mit dem zentralen Grenzwertsatz wird klar, wenn man berücksichtigt, daß dieser in anderer Formulierung folgendes feststellt:

$$P(a < Z_n \leq b) = F_{Z_n}(b) - F_{Z_n}(a) \xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a) \approx (b - a)\varphi\left(\frac{a + b}{2}\right),$$

(im letzten Schritt wurde der Mittelwertsatz benützt und die Zwischenstelle durch die Intervallmitte ersetzt).

BEISPIEL 2.48 (Meßfehler). Angenommen die Zufallsvariablen X_1, \dots, X_n repräsentieren n unabhängige Messungen einer Größe μ . Die Messungen seien mit keinem systematischen Fehler behaftet und es sei $V(X_i) = \sigma^2$. Es liegt nahe, das Mittel der Messungen $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ als Schätzung für μ zu nehmen. Das Gesetz der großen Zahlen sichert die stochastische Konvergenz von \bar{X}_n gegen μ . Mit Hilfe der Tschebyscheff'schen Ungleichung könnte man die Wahrscheinlichkeit eines Fehlers gegebener Größe abschätzen, der zentrale Grenzwertsatz führt auf ein viel schärferes Resultat. Angenommen, wir möchten $P(|\bar{X}_n - \mu| < c)$ abschätzen. Um den zentralen Grenzwertsatz anwenden zu können, geht man zuerst wieder auf standardisierte Variable über:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}},$$

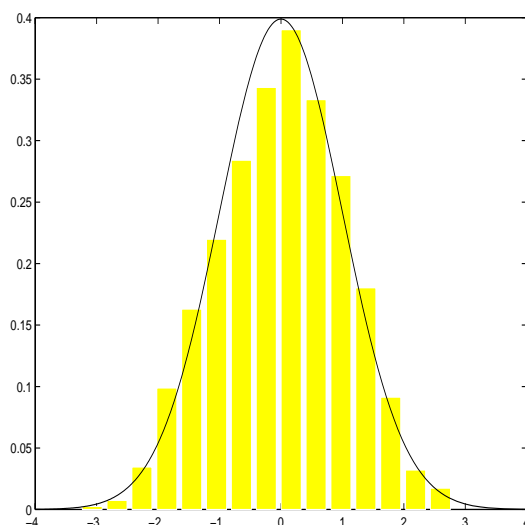


ABB. 2.15

beachte $E(\bar{X}_n) = \mu$ und $V(\bar{X}_n) = \frac{\sigma^2}{n}$. Die gesuchte Wahrscheinlichkeit ergibt sich aus

$$P(|\bar{X}_n - \mu| < c) = P(-c < \bar{X}_n - \mu < c) = P\left(\frac{-c}{\sigma/\sqrt{n}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{c}{\sigma/\sqrt{n}}\right) \\ \approx \Phi\left(\frac{c}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-c}{\sigma/\sqrt{n}}\right).$$

Werden beispielsweise 16 Messungen mit $\sigma = 1$ vorgenommen, ergibt sich für die Wahrscheinlichkeit der Abweichung des Mittels von μ um höchstens $\frac{1}{2}$

$$P(|\bar{X}_n - \mu| < \frac{1}{2}) = \Phi(2) - \Phi(-2) \doteq 0.954.$$

BEISPIEL 2.49 (Normalapproximation der Binomialverteilung). Es gibt Tabellen für die Binomialverteilung für verschiedene Stichprobengrößen. Die direkte Auswertung der Binomialverteilung für nichttabellierte Werte von n ist allerdings für große Stichproben sehr mühsam. Eine Alternative bietet der zentrale Grenzwertsatz: eine $B(n, p)$ verteilte Zufallsvariable X kann nämlich als Summe von sogenannten **Bernoulli Zufallsvariablen** X_i gedeutet werden, also

$$X = \sum_{i=1}^n X_i,$$

wobei

$$X_i = \begin{cases} 1 & \text{wenn die } i\text{-te Wiederholung erfolgreich ist} \\ 0 & \text{sonst} \end{cases}$$

Da ein Bernoulli Experiment zugrunde liegt, sind die Zufallsvariablen X_i unabhängig. Ferner gilt $E(X_i) = p$ und $V(X_i) = pq$, $i = 1, \dots, n$. Die standardisierten Zufallsvariablen sind demnach gegeben durch

$$Z_n = \frac{X - np}{\sqrt{npq}} = \frac{\bar{X} - p}{\sqrt{pq/n}}.$$

Nach dem zentralen Grenzwertsatz kann die Verteilung von Z_n durch Φ approximiert werden. als Faustregel gilt, daß die Approximation brauchbar ist, wenn

$$np > 5, \quad \text{und} \quad nq > 5.$$

BEISPIEL 2.50. Eine Münze zeigt bei 100 Würfeln 60 Mal Kopf. Sind Zweifel an der Fairness der Münze angebracht?

Wenn die Münze fair ist, ist die Anzahl von Köpfen eine binomialverteilte Zufallsvariable mit $n = 100$, $p = .5$, $E(X) = 50$ und $V(X) = 25$. Man überlege sich, daß die Berechnung von $P(X = 60)$ oder $P(X = 50)$ wenig aufschlußreich ist. Wir berechnen daher $P(X \geq 60)$ und erhalten

$$P(X \geq 60) = 1 - P(X < 60) \approx 1 - \Phi\left(\frac{60 - 50}{\sqrt{0.5 \cdot 0.5 \cdot 100}}\right) = 1 - \Phi(2) \doteq 0.0228,$$

es sind also Zweifel an der Korrektheit der Münze angebracht.

KAPITEL 3

Schließende Statistik

Bisher sind wir davon ausgegangen, daß die Parameter der Wahrscheinlichkeitsverteilungen, beispielsweise μ und σ in der Normalverteilung, bekannt sind. Bei praktischen Anwendungen ist dies jedoch selten der Fall. Die schließende Statistik stellt Methoden bereit, mit deren Hilfe man aus Stichproben Information über die interessierenden Parameter gewinnen kann. Da naturgemäß eine Stichprobe nur einen kleinen Teil der Grundpopulation umfaßt, birgt diese Information stets ein bestimmtes Maß an Unsicherheit in sich. Absolut zuverlässige Information wäre nur bei Erfassen der gesamten Grundpopulation zu erzielen. Es ist auch Aufgabe der schließenden Statistik das Ausmaß an Unsicherheit zu quantifizieren. Eine wesentliche Voraussetzung für die Anwendung statistischer Methoden ist die Zufälligkeit der Auswahl der Elemente in der Stichprobe: für jedes Individuum der Grundpopulation muß die Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden, gleich sein. Nur dadurch ist gewährleistet, daß die Stichprobe das Verhalten der Grundpopulation ausreichend widerspiegelt. Auf die Methoden der statistisch korrekten Entnahme einer Stichprobe, ein durchaus schwieriges Problem, kann hier nicht eingegangen werden.

In diesem Kapitel sollen 2 Klassen von Schätzmethoden vorgestellt werden: bei der einen Klasse wird mit Hilfe der Stichprobe ein numerischer Wert für den interessierenden Parameter berechnet. Diese Verfahren werden **Punktschätzungen** genannt. Eine andere Idee besteht darin, die Stichprobe zu benutzen, um ein Intervall zu berechnen, welches den Zielparameter mit einer vorgegebenen Wahrscheinlichkeit enthält. Diese Gruppe von Verfahren nennt man **Intervallschätzungen**.

1. Punktschätzverfahren

Wir betrachten ein bestimmtes Merkmal einer Population mit N Individuen. Jedem Individuum sei ein für das betrachtete Merkmal charakteristischer numerischer Wert x_i , $i = 1, \dots, n$ zugeordnet. Bei einem metrischen Merkmal bezeichnet x_i die jeweilige Ausprägung des Merkmals beim i -ten Individuum, bei einem qualitativen Merkmal kann x_i durch die Werte 0 oder 1 beispielsweise das Fehlen oder Vorhandensein des Merkmals angezeigt werden. Im Folgenden bezeichnen wir den **Populationsmittelwert** stets mit

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

und die **Populationsvarianz** stets mit

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2.$$

Die Definition der Populationsvarianz wurde so gewählt, daß sie im Einklang steht mit der Definition der Varianz einer Zufallsvariablen. Meist sind diese beiden Populationsparameter nicht bekannt. Um Aufschluß über diese Parameter zu erhalten, zieht man eine zufällige Stichprobe. Dies ist prinzipiell auf zwei Arten möglich: Ziehen mit bzw. ohne Zurücklegen. Zieht man die Stichprobe mit Zurücklegen, ist nach jedem Zug der ursprüngliche Umfang der Population wieder hergestellt, sodaß die einzelnen Züge voneinander unabhängig sind. Wird ein bereits ausgewähltes Individuum abermals gezogen, wird dieser Zug nicht berücksichtigt. Die unter dieser Annahme entwickelte Theorie der Stichprobenentnahme heißt Stichprobenentnahme aus einer unendlichen Grundgesamtheit. Beim Ziehen ohne Zurücklegen ist die Unabhängigkeit der einzelnen Züge nur mehr dann gewährleistet, wenn die Grundpopulation beliebig groß ist. Praktisch sieht man die Grundpopulation als unendlich an, wenn der Stichprobenumfang n klein ist im Vergleich zur Größe der Grundpopulation. Hat man auf irgendeine Weise eine Stichprobe vom Umfang n gezogen, liegen Stichprobenwerte x_{i_1}, \dots, x_{i_n} vor, welche zur Berechnung einer Schätzung für den Zielparameter herangezogen werden kann.

1.1. Schätzung des Stichprobenmittels. Es liegt nahe, das Stichprobenmittel

$$\bar{x}_n = \frac{1}{n} \sum_{k=1}^n x_{i_k}$$

als Schätzwert für das Populationsmittel μ und die Stichprobenvarianz

$$s_n^2 = \frac{1}{n-1} \sum_{k=1}^n (x_{i_k} - \bar{x}_n)^2$$

als Schätzwert für die Populationsvarianz zu nehmen (wir werden etwas später klären, warum es günstiger ist die Stichprobenvarianz mit dem Nenner $n-1$ zu definieren). Die konkreten Stichprobenwerte und damit auch das Stichprobenmittel, bzw. die Stichprobenvarianz sind natürlich erst nach der Ziehung der Stichprobe festgelegt. Man kann nun die Stichprobenwerte einerseits auffassen als n beobachtete Werte einer *einzelnen* Zufallsvariablen X , welche die Verteilung des interessierenden Merkmals in der Population beschreibt, es hat sich jedoch als fruchtbarer herausgestellt, die Stichprobenwerte als einzelne Beobachtungen von n Zufallsvariablen X_1, \dots, X_n aufzufassen, welche alle die Verteilung von X besitzen. Dieser Aspekt soll in der folgenden Proposition etwas vertieft werden:

PROPOSITION 3.1. Wir bezeichnen mit ξ_1, \dots, ξ_m die verschiedenen Ausprägungen des Merkmals in der Population und mit n_j die Anzahl der Individuen, bei welchen das Merkmal die Ausprägung ξ_j , $j = 1, \dots, m$, annimmt. Dann ist jedes X_i , $i = 1, \dots, n$ eine Zufallsvariable, deren Wahrscheinlichkeitsverteilung gegeben ist durch

$$P(X_i = \xi_j) = \frac{n_j}{N}.$$

Ferner gilt

$$\begin{aligned} E(X_i) &= \mu, \\ V(X_i) &= \sigma^2, \quad , i = 1, \dots, n. \end{aligned}$$

BEWEIS. Die einzigen Werte, welche X_i annehmen kann, sind ξ_1, \dots, ξ_m . Da jedes Individuum der Population mit gleicher Wahrscheinlichkeit als i -tes Stichprobenelement in Frage kommt, ist die Wahrscheinlichkeit, mit der X_i den Wert ξ_j annimmt, gegeben durch $\frac{n_j}{N}$. Der Erwartungswert von X_i ergibt sich aus

$$E(X_i) = \sum_{j=1}^m \xi_j P(X_i = \xi_j) = \frac{1}{N} \sum_{j=1}^m \xi_j n_j = \frac{1}{N} \sum_{i=1}^N x_i = \mu$$

Analog berechnet man die Varianz von X_i

$$V(X_i) = E(X_i^2) - E(X_i)^2 = \frac{1}{N} \sum_{j=1}^m \xi_j^2 n_j - \mu^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2 = \sigma^2.$$

□

Die Zufallsvariablen X_i , $i = 1, \dots, n$, sind also identisch verteilt. Es sind aber auch sämtliche aus der Stichprobe abgeleitete Größen selbst wieder Zufallsvariable, welche von X_1, \dots, X_n abhängen. Insbesondere sind daher Stichprobenmittel und Stichprobenvarianz Zufallsvariable. Um diesen Aspekt deutlich hervorzuheben, schreiben wir fortan

$$\begin{aligned} (3.1) \quad \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, \\ S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

für das Stichprobenmittel bzw. die Stichprobenvarianz. Man nennt deren Verteilungen **Stichprobenverteilung** des Stichprobenmittels bzw. der Stichprobenvarianz.

PROPOSITION 3.2. 1. Der Erwartungswert (der Stichprobenverteilung) des Stichprobenmittels ist stets das Populationsmittel:

$$(3.2) \quad E(\bar{X}) = \mu.$$

2. Sind die Zufallsvariablen X_1, \dots, X_n unabhängig, dann ist die Varianz (der Stichprobenverteilung) des Stichprobenmittels gegeben durch

$$(3.3) \quad V(\bar{X}) = \frac{\sigma^2}{n}.$$

BEWEIS. Die erste Aussage folgt aus der Linearität des Erwartungswertes, die zweite Aussage aus der stochastischen Unabhängigkeit der Zufallsvariablen X_i , $i = 1, \dots, n$, (2.115) und Korollar 2.2 \square

Es wurde bereits ausgeführt, daß die Unabhängigkeit der Zufallsvariablen X_1, \dots, X_n nur bei der Ziehung der Stichprobe mit Zurücklegen gewährleistet ist. Die Entnahme der Stichprobe ohne Zurücklegen induziert eine Abhängigkeit zwischen den Zufallsvariablen X_1, \dots, X_n , welche die Berechnung der Varianz des Stichprobenmittels etwas kompliziert. Wir berechnen zuerst $\text{Cov}(X_i, X_j)$:

PROPOSITION 3.3. Wird eine Stichprobe vom Umfang n aus einer Population mit N Individuen ohne Zurücklegen entnommen, dann gilt

$$(3.4) \quad \text{Cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}, \quad i \neq j$$

BEWEIS. Nach Proposition 2.32 gilt

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

Der Erwartungswert von $X_i X_j$ ergibt sich aus der gemeinsamen Verteilung von X_i und X_j (ξ_1, \dots, ξ_m bezeichnen die verschiedenen Werte von X_i , $i = 1, \dots, n$):

$$\begin{aligned} E(X_i X_j) &= \sum_{k=1}^m \sum_{l=1}^m \xi_k \xi_l P(X_i = \xi_k \text{ und } X_j = \xi_l) \\ &= \sum_{k=1}^m \xi_k P(X_i = \xi_k) \sum_{l=1}^m \xi_l P(X_j = \xi_l | X_i = \xi_k). \end{aligned}$$

Eine einfache Überlegung ergibt die bedingte Wahrscheinlichkeit

$$P(X_j = \xi_l | X_i = \xi_k) = \begin{cases} \frac{n_l}{N-1} & \text{für } k \neq l \\ \frac{n_l - 1}{N-1} & \text{für } k = l \end{cases}$$

Berücksichtigt man Proposition 3.1 und $\sigma^2 + \mu^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 = \sum_{k=1}^m \xi_k^2 \frac{n_k}{N}$ ergibt sich

$$\begin{aligned} E(X_i X_j) &= \sum_{k=1}^m \xi_k \frac{n_k}{N} \left[\sum_{\substack{l=1 \\ l \neq k}}^m \frac{n_l}{N-1} \xi_l + \frac{n_k-1}{N-1} \xi_k \right] \\ &= \sum_{k=1}^m \xi_k \frac{n_k}{N} \left[\sum_{l=1}^m \frac{n_l}{N-1} \xi_l - \frac{\xi_k}{N-1} \right] \\ &= \frac{N}{N-1} \sum_{k=1}^m \sum_{l=1}^m \xi_k \xi_l \frac{n_k}{N} \frac{n_l}{N} - \sum_{k=1}^m \xi_k^2 \frac{n_k}{N(N-1)} \\ &= \frac{N}{N-1} \mu^2 - \frac{1}{N-1} (\sigma^2 + \mu^2) = \mu^2 - \frac{\sigma^2}{N-1}. \end{aligned}$$

Somit folgt die gesuchte Kovarianz

$$\text{Cov}(X_i, X_j) = \mu^2 - \frac{\sigma^2}{N-1} - \mu^2 = -\frac{\sigma^2}{N-1}.$$

□

Aus diesem Lemma geht hervor, daß die Zufallsvariablen X_1, \dots, X_n beim Ziehen ohne Zurücklegen nicht unabhängig sind, ihre Korrelation jedoch mit wachsender Populationsgröße immer schwächer wird.

PROPOSITION 3.4. *Wird eine Stichprobe vom Umfang n aus einer Population mit N Individuen ohne Zurücklegen entnommen, dann ist die Varianz der Stichprobenverteilung des Stichprobenmittels gegeben durch*

$$(3.5) \quad V(\bar{X}) = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right).$$

BEWEIS. Nach Korollar 2.114 gilt

$$\begin{aligned} V(\bar{X}) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i}^n \text{Cov}(X_i, X_j) \\ &= \frac{\sigma^2}{n} - \frac{1}{n^2} n(n-1) \frac{\sigma^2}{N-1} = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right). \end{aligned}$$

□

Man beachte, daß sich die Varianzen der Verteilungen des Stichprobenmittels, wenn die Stichproben mit bzw. ohne Zurücklegen entnommen werden, nur um den Faktor

$$1 - \frac{n-1}{N-1},$$

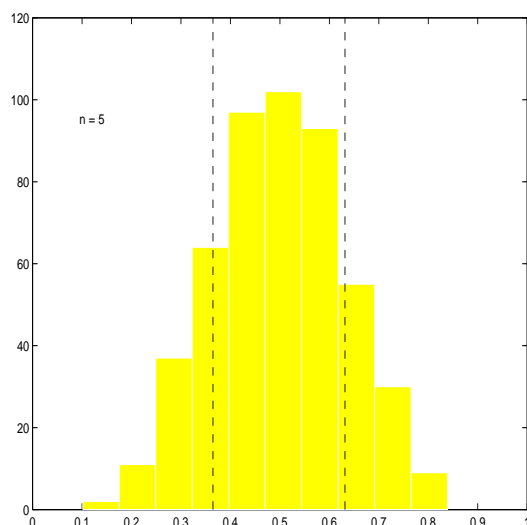


ABB. 3.1

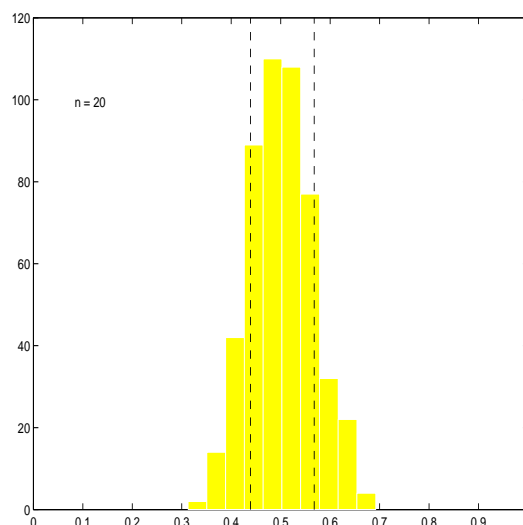


ABB. 3.2

die sogenannte endliche Populationskorrektur, unterscheiden. Ist der Stichprobenanteil $\frac{n}{N}$ klein, dann ist der Unterschied zwischen den beiden Varianzen vernachlässigbar. Aus diesem Grunde werden oft nur Stichproben mit Zurücklegen behandelt.

Die Zufallsvariable \bar{X} ist ein Beispiel eines Punktschätzers für das i.A. unbekanntes Populationsmittel μ . Proposition 3.2 zeigt, daß dieser Schätzer die angenehme Eigenschaft hat, daß sein Mittelwert mit dem Wert des gesuchten Parameters übereinstimmt. Das Gesetz der großen Zahlen garantiert zumindest mit wachsendem Stichprobenumfang die stochastische Konvergenz von \bar{X} gegen μ . Die Standardabweichung, $\sigma_{\bar{X}}$, der Stichprobenverteilung von \bar{X} , in diesem Zusammenhang oft auch **Standardfehler** des Schätzers genannt, ist ein Maß für die Güte der Schätzung. Sieht man von der endlichen Populationskorrektur ab, gilt

$$\sigma_{\bar{X}} \approx \frac{\sigma}{\sqrt{n}}.$$

Die Güte der Schätzung ist demnach unabhängig von der Größe der Population. Sie wird aber wesentlich vom Umfang der Stichprobe beeinflusst: eine Verdoppelung der Genauigkeit der Schätzung muß mit dem vierfachen Stichprobenumfang erkauft werden. Dieser Effekt wird in den Abbildungen 3.1 und 3.2 veranschaulicht. Sie zeigen Histogramme von 500 Stichproben der Mittelwerte von 5 bzw. 20 in $(0,1)$ gleichverteilten Zufallszahlen. Für die Standardfehler ergab sich $\sigma_{\bar{X}_5} = 0.13$ und $\sigma_{\bar{X}_{20}} = 0.065$.

Die Genauigkeit ist aber auch direkt zur Populationsvarianz σ proportional. Dies ist unmittelbar einsichtig: je kleiner σ ist, desto weniger streuen die Werte des beobachteten Merkmales in der Ausgangspopulation um den Mittelwert, desto weniger streuen aber auch die Werte in einer zufällig gezogenen Stichprobe, sodaß ein

umso kleinerer Stichprobenumfang genügt, dieselbe Genauigkeit bei der Schätzung zu erzielen.

Kehren wir nun noch einmal zum allgemeinen Punktschätzproblem zurück. Angenommen wir möchten einen Populationsparameter θ schätzen. Wir bezeichnen die Schätzfunktion mit $\hat{\theta}$. Diese hängt von den Zufallsvariablen X_1, \dots, X_n ab,

$$\hat{\theta} = g(X_1, \dots, X_n),$$

und ist daher selbst eine Zufallsvariable, deren Verteilung die Stichprobenverteilung von $\hat{\theta}$ heißt. Vorhin wurde $\hat{\mu} = \bar{X}$ als Schätzer für $\theta = \mu$ diskutiert. Meist wird es mehrere, gleich plausible Möglichkeiten der Schätzung geben. Wie kann man aus diesen Möglichkeiten eine rationale Auswahl treffen? Eine wünschenswerte Eigenschaft einer Schätzfunktion ist, daß seine Verteilungsfunktion möglich ist um den wahren Parameterwert θ zentriert ist, also $E(\hat{\theta}) = \theta$ gilt.

DEFINITION 3.1. *Ein Punktschätzer $\hat{\theta}$ eines Parameters θ heißt **erwartungstreu (unbiased)**, wenn*

$$(3.6) \quad E(\hat{\theta}) = \theta$$

*zutrifft. Man nennt $B = E(\hat{\theta}) - \theta$ **Bias (systematischer Fehler)** des Punktschätzers.*

Proposition 3.2 stellt also fest, daß \bar{X} ein erwartungstreuer Punktschätzer des Populationsmittels ist. Die Abbildungen 3.3 und 3.4 zeigen zwei mögliche Stichprobenverteilungen ein es erwartungstreuen Punktschätzers für einen Zielparameter θ . Offensichtlich wird man bei sonst gleichen Charakteristika den Schätzer aus Abbildung 3.4 bevorzugen, da dessen kleinere Varianz erwarten läßt, daß bei mehrfachen Stichproben ein höherer Anteil der Schätzungen in der Nähe von θ liegt. Wir werden also trachten, erwartungstreue Punktschätzer minimaler Varianz zu finden. Anstelle der Varianz verwendet man gelegentlich auch die **mittlere quadratische Abweichung** $\hat{\theta} - \theta$ zur Bewertung eines Punktschätzers.

1.2. Schätzen der Populationsvarianz. Wir haben bereits mehrfach festgestellt, daß

$$(3.7) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

einen möglichen Schätzer für die Populationsvarianz darstellt. Wir zeigen nun, daß dieser Schätzer nicht erwartungstreu ist.

PROPOSITION 3.5. *Wird eine Stichprobe vom Umfang n aus einer Population mit N Individuen ohne Zurücklegen gezogen und nimmt man (3.7) als Punktschätzer für die Populationsvarianz σ^2 , dann gilt*

$$(3.8) \quad E(\hat{\sigma}^2) = \sigma^2 \frac{n-1}{n} \frac{N}{N-1}.$$

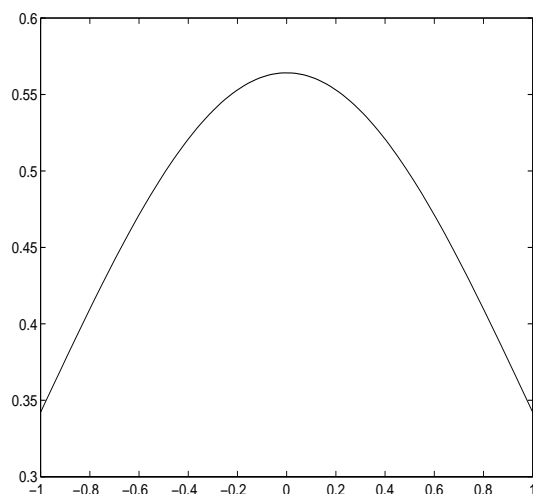


ABB. 3.3

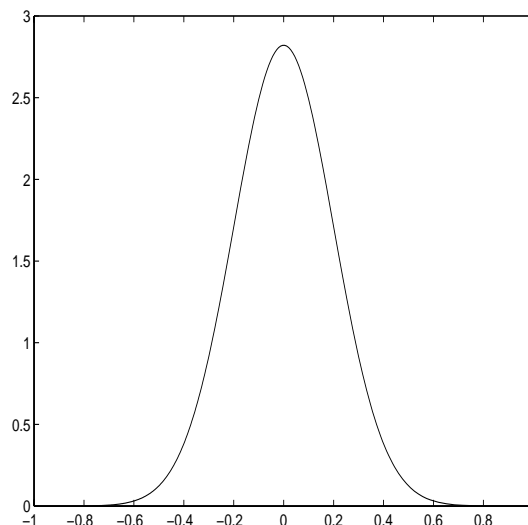


ABB. 3.4

BEWEIS. Es wurde bereits gezeigt, daß (3.7) dargestellt werden kann durch

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Also folgt

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}^2).$$

Aus der stets gültigen Beziehung $V(X) = E(X^2) - E(X)^2$ folgert man mit Proposition 3.1 und Proposition 3.4

$$\begin{aligned} E(X_i^2) &= V(X_i) + E(X_i)^2 = \sigma^2 + \mu^2, \\ E(\bar{X}^2) &= V(\bar{X}) + E(\bar{X})^2 = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right) + \mu^2 \end{aligned}$$

und somit

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n} n(\sigma^2 + \mu^2) - \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right) - \mu^2 \\ &= \sigma^2 \frac{n-1}{n} \frac{N}{N-1}. \end{aligned}$$

□

Wegen $n < N$ ist $\frac{n-1}{n} \frac{N}{N-1} < 1$ und somit auch

$$B = E(\hat{\sigma}^2) - \sigma^2 < 0.$$

Der Punktschätzer (3.7) besitzt also eine negative Bias, er tendiert also dazu, die Populationsvarianz systematisch zu unterschätzen. Wegen der Linearität des Erwartungswertes erkennt man aber, daß

$$(3.9) \quad \hat{\sigma}^2 = \frac{N-1}{N} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ein erwartungstreuer Schätzer für die Populationsvarianz ist, wenn die Stichprobe ohne Zurücklegen gezogen wird. Stichproben, welche mit Zurücklegen gezogen werden, werden erfaßt, indem man in (3.9) den Grenzwert $N \rightarrow \infty$ ausführt. Dies —oder eine direkte Rechnung— führt auf

$$(3.10) \quad \hat{\sigma}^2 \equiv S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Da der relative Unterschied zwischen den beiden Varianzen von der Größenordnung $\frac{1}{N}$ ist, nimmt man meist (3.10) als Punktschätzer für die Populationsvarianz.

Es wurde in Proposition 3.4 gezeigt, daß die Varianz des Stichprobenmittels

$$V(\bar{X}) = \sigma_{\bar{X}}^2 \approx \frac{\sigma^2}{n}$$

bestimmt ist durch die meist unbekannteste Populationsvarianz. Ersetzt man σ^2 durch einen erwartungstreuen Schätzer $\hat{\sigma}^2$ erhält man einen erwartungstreuen Schätzer für die Varianz des Stichprobenmittels:

PROPOSITION 3.6. *Ein erwartungstreuer Schätzer für die Varianz der Verteilung des Stichprobenmittels ist gegeben durch*

$$(3.11) \quad \hat{\sigma}_{\bar{X}}^2 = \begin{cases} \frac{S^2}{n} (1 - \frac{n}{N}) & \text{Stichprobe ohne Zurücklegen} \\ \frac{S^2}{n} & \text{Stichprobe mit Zurücklegen} \end{cases}$$

BEWEIS. Wir betrachten nur den Fall einer Stichprobe, welche ohne Zurücklegen entnommen wird. Der andere Fall ergibt sich wie vorhin, indem man den Grenzübergang $N \rightarrow \infty$ durchführt. Nach Proposition 3.4 gilt

$$V(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}.$$

Es sei $\hat{\sigma}^2$ ein erwartungstreuer Schätzer für σ^2 . Setzt man

$$\hat{\sigma}_{\bar{X}}^2 = \frac{\hat{\sigma}^2}{n} \frac{N-n}{N-1}$$

folgt

$$E(\hat{\sigma}_{\bar{X}}^2) = \frac{N-n}{N-1} \frac{1}{n} E(\hat{\sigma}^2) = \frac{\sigma^2}{n} \frac{N-n}{N-1} = V(\bar{X}),$$

d.h. $\hat{\sigma}_{\bar{X}}^2$ ist ein erwartungstreuer Schätzer von $V(\bar{X})$. Ersetzt man $\hat{\sigma}^2$ durch (3.9) ergibt sich die Behauptung. \square

1.3. Schätzen eines Populationsanteils. Oft ist man an dem Anteil p der Population interessiert, welche eine bestimmte Eigenschaft besitzen. Dieser kann mit den vorhin geschilderten Methoden geschätzt werden, wenn man $X_i = 0$ oder $= 1$ setzt, je nachdem, ob das i -te Stichprobenelement die gewünschte Eigenschaft besitzt oder nicht. Bildet man das Stichprobenmittel \bar{X} , dann bedeutet dies gerade die relative Häufigkeit dieser Eigenschaft in der betreffenden Stichprobe. Es liegt daher nahe, das Stichprobenmittel als Schätzer für den Anteil p zu nehmen, d.h.

$$\hat{p} = \bar{X}.$$

Nach Proposition 3.1 ist \hat{p} ein erwartungstreuer Schätzer für den Populationsanteil p . Die Populationsvarianz ist in diesem Fall durch

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2 = \frac{1}{N} \sum_{i=1}^N x_i - \mu^2 = p(1-p)$$

gegeben (beachte $\mu = p$). Dieselbe Rechnung zeigt

$$S^2 = \frac{n}{n-1} \hat{p}(1-\hat{p}).$$

Als Spezialfall von Proposition 3.6 erhält man daher

COROLLARY 3.1. *Ein erwartungstreuer Punktschätzer der Varianz eines Populationsanteils ist durch*

$$(3.12) \quad \hat{\sigma}_{\hat{p}}^2 = \begin{cases} \frac{\hat{p}(1-\hat{p})}{n-1} (1 - \frac{n}{N}) & \text{für Stichproben ohne Zurücklegen} \\ \frac{\hat{p}(1-\hat{p})}{n-1} & \text{für Stichproben mit Zurücklegen} \end{cases}$$

Bei kleinem Stichprobenanteil $\frac{n}{N}$ kann die endliche Populationskorrektur vernachlässigt werden, ebenso verwendet man manchmal n anstelle von $n-1$ im Nenner.

BEISPIEL 3.1. In einer zufällig gezogenen Stichprobe von $n = 1000$ Wählern, äußerten $x = 560$ der Befragten Sympathie für den Kandidaten A. Was kann man daraus für die Wahlchancen von A ableiten?

Wir verwenden $\hat{p} = \bar{X}$ als Schätzer für den Stimmenanteil von A. Dies ergibt $\hat{p} = \frac{560}{1000} = 0,56$, also einen Stimmenanteil von 56%. Mit welcher Unsicherheit ist dieser Stimmenanteil behaftet? Dazu berechnen wir ein Intervall I um \hat{p} , welches mit einer Wahrscheinlichkeit von mindestens 75% den wahren Stimmenanteil von A enthält. Nach der Ungleichung von Tscheyeff gilt

$$P(|\hat{p} - p| < k\sigma_{\hat{p}}) \geq 1 - \frac{1}{k^2}.$$

Ein Intervall der Länge $2\sigma_{\hat{p}}$ um \hat{p} hat demnach die gewünschten Eigenschaften. Allerdings ist

$$\sigma_{\hat{p}}^2 = \frac{\sigma^2}{n} \left(1 - \frac{n-1}{N-1}\right)$$

nicht zugänglich. Eine brauchbare Abschätzung der Genauigkeit von \hat{p} erhält man, indem man das Intervall $(\hat{p} - 2\sigma_{\hat{p}}, \hat{p} + 2\sigma_{\hat{p}})$ durch das Intervall $(\hat{p} - 2\hat{\sigma}_{\hat{p}}, \hat{p} + 2\hat{\sigma}_{\hat{p}})$ ersetzt. Dies ergibt

$$2\hat{\sigma}_{\hat{p}} = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} = 2\sqrt{\frac{0,56 \cdot 0,44}{999}} = 0.03.$$

Mit einer Wahrscheinlichkeit von mindestens 75% liegt der wahre Stimmenanteil von A zwischen 53% und 59%. Da wegen des großen Stichprobenumfanges die Verteilung von \hat{p} bereits sehr gut durch die Normalverteilung approximiert wird, liegt diese Wahrscheinlichkeit sogar bei 95%.

1.4. Der Vergleich zweier Populationen. Wir betrachten nun zwei verschiedene, unabhängige Populationen mit Populationsmittel μ_1, μ_2 und binomialen Parametern p_1 bzw. p_2 . Wir interessieren uns für den Unterschied in diesen Parametern. Dazu ziehen wir Stichproben vom Umfang n_1 bzw. n_2 aus den beiden Populationen und versuchen mit Hilfe der Stichproben auf $\mu_1 - \mu_2$ bzw. auf $p_1 - p_2$ zu schließen. Es liegt nahe als Schätzer für die Differenz dieser Parameter $\bar{X}_1 - \bar{X}_2$, bzw. $\hat{p}_1 - \hat{p}_2$ zu verwenden. Diese Schätzer sind erwartungstreu. Ihre Varianz ist gegeben durch

$$(3.13) \quad V(\bar{X}_1) - V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

und (der Einfachheit halber betrachten wir nur Stichproben mit Zurücklegen)

$$(3.14) \quad V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

In konkreten Beispielen sind die unbekanntenen Populationsvarianzen durch entsprechende Schätzungen zu ersetzen.

BEISPIEL 3.2. Zwei verschiedene Autoreifenmarken werden einem Verschleißtest unterzogen. Je 100 Reifen wurden so lange gefahren, bis ein bestimmter Abnutzungsgrad erreicht war. Als Testergebnis ergab sich

$$\begin{aligned} \bar{x}_1 &= 26.400 \text{ km} & s_1^2 &= 1.440.000 \\ \bar{x}_2 &= 25.100 \text{ km} & s_2^2 &= 1.960.000. \end{aligned}$$

Sind die Reifen der ersten Firma tatsächlich ausdauernder? Dazu schätzen wir den Unterschied in der mittleren Kilometerleistung $\mu_1 - \mu_2$. Der Punktschätzer $\bar{X}_1 - \bar{X}_2$ ergibt

$$\hat{\mu}_1 - \hat{\mu}_2 = 26.400 - 25.100 = 1300 \text{ km}.$$

Die Güte dieser Schätzung beurteilen wir durch den Standardfehler $\sigma_{\bar{X}_1 - \bar{X}_2}$, in welchem wir die Populationsvarianzen durch die Stichprobenvarianzen ersetzen:

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{1.440.000}{100} + \frac{1.960.000}{100}} = 184.4 \text{ km.}$$

Mit einer Wahrscheinlichkeit von (wegen des großen Stichprobenumfanges) von 95% liegt also der Unterschied in der mittleren Kilometerleistung $\mu_1 - \mu_2$ zwischen 931.2 km und 1668.8 km. Dies ist ein gewichtiges Argument für die Reifen der Firma 1.

Abschließend stellen wir die gängigen Punktschätzer noch einmal zusammen (der Einfachheit halber berücksichtigen wir nur Stichproben, welche mit Zurücklegen entnommen wurden).

θ	n	$\hat{\theta}$	$E(\hat{\theta})$	$\sigma_{\hat{\theta}}$
μ	n	\bar{X}	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	\bar{X}	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 und n_2	$\bar{X}_1 - \bar{X}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1 und n_2	$\bar{X}_1 - \bar{X}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

TABELLE 3.1. Punktschätzer für Populationsparameter

2. Konfidenzintervalle

Bisher haben wir nur Punktschätzer betrachtet: die Stichprobe wurde benutzt, um einen Schätzwert z.B. für einen Populationsparameter zu berechnen. Ein **Intervallschätzer** dagegen berechnet aus der Stichprobe ein Intervall $[\hat{\theta}_u, \hat{\theta}_o]$, welches den Zielparameter θ mit einer vorgegebenen Wahrscheinlichkeit enthält:

$$(3.15) \quad P(\hat{\theta}_u \leq \hat{\theta} \leq \hat{\theta}_o) = 1 - \alpha.$$

Die Intervallgrenzen $\hat{\theta}_u$, $\hat{\theta}_o$ hängen von der jeweiligen Stichprobe ab und variieren daher in zufälliger Weise von Stichprobe zu Stichprobe. Die Schätzintervalle nennt man allgemein **Konfidenzintervalle**, die zugehörige Wahrscheinlichkeit $1 - \alpha$ **Konfidenzniveau** oder Vertrauensniveau. Aus der weiteren Diskussion wird klar, warum es günstig ist, das Vertrauensniveau in der Form $1 - \alpha$ zu schreiben. Es sei ausdrücklich darauf hingewiesen, daß ein Konfidenzintervall den Zielparameter nicht enthalten muß dies tritt allerdings nur mit der geringen Wahrscheinlichkeit α ein.

Konfidenzintervalle vom Typ (3.15) nennt man zweiseitige Konfidenzintervalle. Es sind auch einseitige Konfidenzintervalle

$$P(\theta \leq \hat{\theta}_o) = 1 - \alpha \quad \text{bzw.} \quad P(\hat{\theta}_u \leq \theta) = 1 - \alpha$$

in Gebrauch. Eine häufig verwendete Strategie zur Berechnung von Konfidenzintervallen geht von einer sogenannten Pivotgröße Z aus, welche folgende Eigenschaften besitzen soll:

- Z hängt nur von den Stichprobenwerten und dem Zielparameter θ ab. Dieser ist die einzige Unbekannte in Z .
- Die Wahrscheinlichkeitsverteilung von Z hängt nicht von θ ab.

Man berechnet zuerst ein Konfidenzintervall für Z und rechnet dieses anschließend auf ein Konfidenzintervall für θ um. Folgendes Beispiel illustriert die Anwendung der Pivotstrategie:

BEISPIEL 3.3. Die Wahrscheinlichkeitsdichte einer Zufallsvariablen Y sei gegeben durch die 1-parametrische Familie ($\theta > 0$)

$$f_Y(y) = \begin{cases} \frac{1}{\theta} e^{-y/\theta} & y \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Um ein zweiseitiges Konfidenzintervall mit Konfidenzniveau 0.90 für θ zu berechnen ($\alpha = 0.1$), bilden wir zuerst die Pivotgröße $Z = \frac{Y}{\theta}$. Die Verteilungsdichte von Z ergibt sich aus

$$F_Z(z) = P(Z \leq z) = P\left(\frac{Y}{\theta} \leq z\right) = P(Y \leq \theta z) = F_Y(\theta z)$$

durch Differenzieren

$$f_Z(z) = \begin{cases} e^{-z} & z \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Ein Konfidenzintervall zum Konfidenzniveau 0.90 für Z ist gegeben durch

$$P(a \leq Z \leq b) = \int_a^b f_Z(z) dz = 0.90.$$

Eine Möglichkeit, dies zu erreichen ist die Forderung

$$P(Z \leq a) = \int_0^a e^{-z} dz = 0.05 (= \frac{\alpha}{2})$$

$$P(Z \geq b) = \int_b^\infty e^{-z} dz = 0.05 (= \frac{\alpha}{2}).$$

Dies ergibt

$$a = 0.51, \quad b = 2.996.$$

Die Konstruktion garantiert

$$P(a \leq Z \leq b) = P\left(a \leq \frac{Y}{\theta} \leq b\right) = 0.90.$$

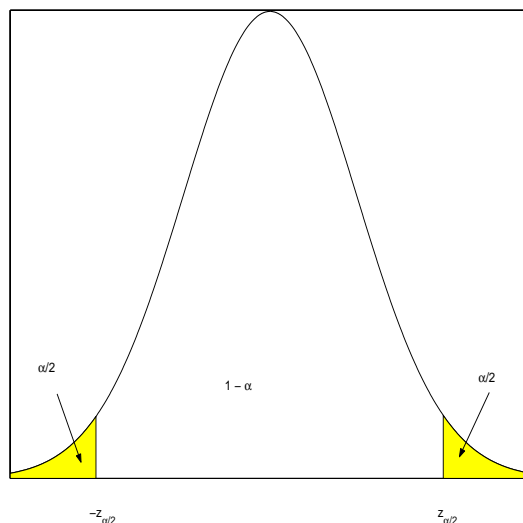


ABB. 3.5

Eine einfache Umformung ergibt nun das Konfidenzintervall für θ :

$$P\left(\frac{Y}{2.996} \leq \theta \leq \frac{Y}{0.051}\right) = 0.90.$$

Der Intervallschätzer ist daher durch die Intervallgrenzen

$$\hat{\theta}_u = \frac{Y}{2.996}, \quad \hat{\theta}_o = \frac{Y}{0.051}$$

gegeben. Ein einziger Beobachtungswert von Y ergibt dann ein Intervall, welches den wahren Parameter mit einer Wahrscheinlichkeit von 90% einschließt.

2.1. Konfidenzintervalle für große Stichproben. Im vorigen Abschnitt haben wir Punktschätzer für verschiedene Populationsparameter besprochen, vgl. Tabelle 3.1. Nach dem zentralen Grenzwertsatz sind die standardisierten Schätzer

$$(3.16) \quad Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

für große Stichprobenumfänge ungefähr standardnormalverteilt. Die Größe Z kann daher als Pivotgröße genommen werden. Ein $1 - \alpha$ Konfidenzintervall für θ ergibt sich daher auf folgende Weise: Wir berechnen zuerst ein $1 - \alpha$ Konfidenzintervall für Z

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$$

$$z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

Der Wert von $z_{\alpha/2}$ kann aus Tabellen entnommen werden. Die Abbildung 3.5 veranschaulicht die zu Grunde liegende Idee.

Substituiert man für Z , ergibt sich

$$P(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}) = 1 - \alpha$$

bzw.

$$P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha.$$

Die Endpunkte eines zweiseitigen $1 - \alpha$ Konfidenzintervalls sind also gegeben durch

$$(3.17) \quad \hat{\theta}_u = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \quad \hat{\theta}_o = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}},$$

die Endpunkte eines einseitigen $1 - \alpha$ Konfidenzintervalles durch

$$(3.18) \quad \hat{\theta}_u = \hat{\theta} - z_{\alpha}\sigma_{\hat{\theta}}, \quad \hat{\theta}_o = \hat{\theta} + z_{\alpha}\sigma_{\hat{\theta}}.$$

BEISPIEL 3.4. Zwei Kühlschranksorten A und B besitzen beide eine Garantiezeit von 1 Jahr. Bei einem Vergleichstest wurden 12 von 50 Geräten der Marke A und 12 von 60 Geräten der Marke B innerhalb der Garantiezeit defekt. Man bestimme ein 98%–Konfidenzintervall für den wahren Unterschied $p_1 - p_2$ der Ausfallsquoten im ersten Jahr.

Das Konfidenzintervall

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

hat nun die Form

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}.$$

Da die wahren Anteile p_1 und p_2 nicht bekannt sind, werden sie durch die Punktschätzungen \hat{p}_1 und \hat{p}_2 ersetzt (konsistenter wäre $\sigma_{\hat{\theta}}$ durch den Punktschätzer $\hat{\sigma}_{\hat{\theta}}$ zu ersetzen, der Unterschied ist jedoch geringfügig). Auf diese Weise erhält man mit $\hat{p}_1 = \frac{12}{50} = 0.24$, $\hat{p}_2 = \frac{12}{60} = 0.2$, $n_1 = 50$, $n_2 = 60$, $z_{0.01} = \Phi^{-1}(0.01) = 2.33$ (aus Tabelle)

$$(0.24 - 0.20) \pm 2.33 \sqrt{\frac{0.24 \cdot 0.76}{50} + \frac{0.20 \cdot 0.80}{60}} = 0.04 \pm 0.1851,$$

also das Intervall

$$[-0.15, 0.23].$$

Da dieses Intervall auch die Null enthält, ist somit sowohl die Hypothese daß sich die beiden Kühlschranksorten eigentlich nicht unterscheiden, aber auch die Hypothese, daß die Ausfallsrate bei Geräten der Marke B um 20% höher ist als jene der Marke A, mit einem Konfidenzniveau von 98% glaubhaft. Der Vergleichstest ist also wenig aussagekräftig.

Man kann diese Technik auch dazu benutzen, um den für ein gewünschtes Konfidenzniveau erforderlichen Stichprobenumfang abzuschätzen. Ein Beispiel veranschaulicht die Vorgangsweise:

BEISPIEL 3.5. Auf einen Reiz in einem psychologischen Experiment seien 2 Reaktionen, A oder B, möglich. Man möchte die Wahrscheinlichkeit abschätzen, daß ein Individuum nach dem Schema A reagiert. Der Schätzfehler soll höchstens 0.04 mit einem Konfidenzniveau von 0.90 betragen. Wieviele Probanden müssen in die Untersuchung aufgenommen werden? Eine rohe Schätzung der gesuchten Wahrscheinlichkeit liegt bei 60%.

Das Konfidenzniveau beträgt $1 - \alpha = 0.90$, also $\frac{\alpha}{2} = 0.05$. Einer Tabelle der Verteilungsfunktion der Standardnormalverteilung entnimmt man $z_{0.05} = 1.645$. Das Konfidenzintervall ist durch $[\hat{p} - z_{0.05}\sigma_{\hat{p}}, \hat{p} + z_{0.05}\sigma_{\hat{p}}]$ gegeben. Es wird also gefordert

$$z_{0.05}\sigma_{\hat{p}} = 1.645\sqrt{\frac{pq}{n}} = 0.04.$$

Der Standardfehler $\sigma_{\hat{p}}$ hängt allerdings von der unbekanntem Wahrscheinlichkeit p ab. Wir benützen daher $p \approx 0.6$ und erhalten

$$1.645\sqrt{\frac{0.6 \cdot 0.4}{n}} = 0.04$$

und daraus $n = 406$. (Ohne die Information $p \approx 0.6$ hätte man das Maximum von n als Funktion von p bestimmen können).

3. Konfidenzintervalle für μ und $\mu_1 - \mu_2$

In diesem Abschnitt betrachten wir Stichproben X_1, \dots, X_n , welche aus einer normalverteilten Grundgesamtheit mit Zurücklegen entnommen wurden. Für die Praxis bedeutet diese Annahme, daß die Verteilung der Population, welche ja meist erst durch die Stichprobe greifbar wird, zumindest glockenförmig ist. Wir berechnen zuerst ein Konfidenzintervall für das Populationsmittel, wenn $V(X_i) = \sigma^2$ nicht bekannt ist und der Stichprobenumfang so klein ist, daß die Methoden des vorigen Abschnittes nicht angewendet werden können. Als Pivotgröße wählen wir in diesem Fall

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Wir zeigen nun, daß T einer t -Verteilung mit $n - 1$ Freiheitsgraden genügt. Dazu benötigen wir folgende Hilfsmittel:

LEMMA 3.1. *Die Zufallsvariablen X_1, \dots, X_n seien $N(\mu, \sigma)$ verteilt und unabhängig. Dann sind das Stichprobenmittel \bar{X} und der Zufallsvektor $(X_1 - \bar{X}), \dots, (X_n - \bar{X})$ stochastisch unabhängig. Darüber hinaus sind auch \bar{X} und S^2 stochastisch unabhängig.*

BEWEIS. Der Beweis dieser Aussage geht über den Umfang dieser Vorlesung hinaus. \square

PROPOSITION 3.7. *Es sei X_1, \dots, X_n eine unabhängige Stichprobe aus einer normalverteilten Grundgesamtheit mit Mittelwert μ und Varianz σ^2 . Dann genügt*

$$U = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

einer χ^2 -Verteilung mit $n-1$ Freiheitsgraden.

BEWEISSKIZZE. Wir bemerken zuerst, daß

$$W = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$$

χ^2 -verteilt ist und n Freiheitsgraden besitzt. Berücksichtigt man $\sum_{i=1}^n (X_i - \bar{X}) = 0$, erhält man

$$\begin{aligned} W &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \equiv U + V. \end{aligned}$$

V ist χ_1^2 verteilt, ferner sind U und V stochastisch unabhängig. Nach Proposition 2.40 gilt daher für die momenterzeugende Funktion von W die Beziehung $M_W(t) = M_U(t)M_V(t)$, also

$$M_U(t) = \frac{M_W(t)}{M_V(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2}.$$

M_U ist demnach die momenterzeugende Funktion einer χ_{n-1}^2 -verteilten Zufallsvariablen. Nach dem Eindeutigkeitsatz Proposition 2.38 ist die Verteilung von U eine χ^2 -Verteilung mit $n-1$ Freiheitsgraden. \square

COROLLARY 3.2. *Die Prüfgröße*

$$(3.19) \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

ist t -verteilt mit $n-1$ Freiheitsgraden.

BEWEIS. Setzt man

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad U = \frac{(n-1)S^2}{\sigma^2}, \quad \nu = n-1,$$

dann ist Z standard normalverteilt, U ist Ξ_{n-1}^2 -verteilt und somit besitzt nach Definition 2.28 $\frac{Z}{\sqrt{U/\nu}}$ eine t -Verteilung mit $n-1$ Freiheitsgraden. Dieser Quotient ist aber

gerade die Prüfgröße T :

$$\frac{Z}{\sqrt{U/\nu}} = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{[(n-1)S^2]/\sigma^2}/(n-1)} = T.$$

□

Wegen der Symmetrie der t -Verteilung um $t = 0$ erhält man ein $1 - \alpha$ Konfidenzintervall für μ aus der Forderung

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$$

nach einfachen Manipulationen zu

$$(3.20) \quad \left[\bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right],$$

wobei $t_{\alpha/2} = F_t^{-1}(1 - \frac{\alpha}{2})$ einer Tabelle der t -Verteilung zu entnehmen ist. Auf analoge Weise findet man, daß

$$\hat{\mu}_u = \bar{X} - t_{\alpha} \frac{S}{\sqrt{n}}$$

eine $1 - \alpha$ untere Konfidenzschranke, bzw.

$$\hat{\mu}_o = \bar{X} + t_{\alpha} \frac{S}{\sqrt{n}}$$

eine $1 - \alpha$ obere Konfidenzschranke für μ darstellt.

BEISPIEL 3.6. Ein Erzeuger von Schießpulver möchte eine neue Mischung testen, und mißt zu diesem Zweck die Mündungsgeschwindigkeit von 8 Projektilen (wegen der Aufwendigkeit der Messungen sind nur kleine Stichproben möglich) (Angaben in

3005	2925	2935	2965
2995	3005	2937	2905

ft/sec). Gesucht ist ein 95%-Konfidenzintervall für die mittlere Mündungsgeschwindigkeit.

Nehmen wir an, daß die Messungen normalverteilt sind. Dann ist ein 95%-Konfidenzintervall durch die Intervallschätzer $\bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$ gegeben. Für die konkrete Meßreihe findet man $\bar{x} = 2959$ und $s = 39.1$. Einer Tabelle der t -Verteilung entnimmt man in der Zeile für $n - 1 = 7$ Freiheitsgrade $t_{\alpha/2} = t_{0.025} = 2.365$. Dies ergibt das Konfidenzintervall

$$2959 \pm 2.365 \frac{39.1}{\sqrt{8}} = 2959 \pm 32.7.$$

Wir wenden uns noch kurz dem Vergleich von 2 *normalverteilten* Populationen zu: die Populationsmittel seien μ_1 , bzw. μ_2 , die Varianzen in beiden Populationen seien

gleich: $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Wir interessieren uns für den Unterschied in den Populationsmitteln $\mu_1 - \mu_2$ und entnehmen zu diesem Zweck aus beiden Populationen voneinander unabhängige Stichproben von Umfang n_1 bzw. n_2 . Ein Konfidenzintervall kann aus der Prüfgröße

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

abgeleitet werden. Da die gemeinsame Varianz σ^2 nicht bekannt ist, konstruieren wir vorerst einen erwartungstreuen Schätzer. Bezeichnen wir die Stichprobenvariablen mit X_{1i}, \dots, X_{n_1i} , $i = 1, 2$. Ein Schätzer für die gemeinsame Varianz ergibt sich, indem man die Varianzen der einzelnen Stichproben folgendermaßen kombiniert:

$$(3.21) \quad S_p^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i2} - \bar{X}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

wobei S_i^2 den erwartungstreue Schätzer (3.10) für die Varianz der Population $i = 1, 2$ bezeichnet. Wegen

$$E(S_p^2) = \frac{n_1 - 1}{n_1 + n_2 - 2} E(S_1^2) + \frac{n_2 - 1}{n_1 + n_2 - 2} E(S_2^2) = \sigma^2$$

ist der kombinierte Schätzer S_p in der Tat erwartungstreu. Die Zufallsvariable

$$W = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} = \sum_{i=1}^{n_1} \left(\frac{X_{i1} - \bar{X}_1}{\sigma}\right)^2 + \sum_{i=1}^{n_2} \left(\frac{X_{i2} - \bar{X}_2}{\sigma}\right)^2$$

ist die Summe zweier χ^2 -verteilter Zufallsvariablen mit je $n_1 - 1$ bzw. $n_2 - 1$ Freiheitsgraden, vgl Proposition 3.7. Daher ist W selbst χ^2 -verteilt mit $\nu = n_1 + n_2 - 2$ Freiheitsgraden. Die Zufallsvariablen Z und W sind stochastisch unabhängig. Somit besitzt die Pivotgröße

$$\begin{aligned} T &= \frac{Z}{\sqrt{W/\nu}} = \frac{\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}{\sqrt{\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} (n_1 + n_2 - 2)}} \\ &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \end{aligned}$$

eine t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden. Für die Grenzen eines $1 - \alpha$ Konfidenzintervalles für $\mu_1 - \mu_2$ erhält man auf dieselbe Weise wie vorhin die Schätzfunktionen

$$(3.22) \quad (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

wobei der Wert von $t_{\alpha/2}$ einer Tabelle der t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden entnommen wird.

BEISPIEL 3.7. Zur Erhöhung der Produktivität werden die Arbeiter an einem bestimmten Fließband einer 3-wöchigen Schulung unterzogen. Es soll nun ein neues Schulungsprogramm getestet werden. Dazu werden je 9 Arbeiter nach der alten bzw. neuen Methode eingeschult. Anschließend wird bei jedem Arbeiter die für die Assemblierung eines Bauteils benötigte Zeit (in Minuten) gemessen

alte Methode	32	37	35	28	41	44	35	31	34
neue Methode	35	31	29	25	34	40	27	32	31

Man bestimme ein 95% Konfidenzintervall für den wahren Unterschied in den mittleren Assemblierungszeiten.

Wir nehmen an, daß diese Zeiten normalverteilt sind, und in beiden Gruppen die gleiche Varianz haben. Die Auswertung der Stichproben ergibt

$$(3.23) \quad \bar{x}_1 = 35.22 \qquad \qquad \qquad \bar{x}_2 = 31.56$$

$$(3.24) \quad \sum_{i=1}^9 (x_{i1} - \bar{x}_1)^2 = 195.56 \qquad \qquad \sum_{i=1}^9 (x_{i2} - \bar{x}_2)^2 = 160.22$$

Der kombinierte Schätzer für die gemeinsame Varianz ergibt den Schätzwert

$$s_p^2 = \frac{195.56 + 160.22}{9 + 9 - 2} = 22.4 \qquad \text{also } s_p = 4.71.$$

Aus der Tabelle für die t -Verteilung mit $9 + 9 - 2 = 16$ Freiheitsgraden entnimmt man $t_{0,025} = 2.120$. Setzt man diese Werte in den Schätzer (3.22) ein, erhält man folgende Grenzen des Konfidenzintervalles

$$(35.22 - 31.56) \pm 2.120 \cdot 4.71 \sqrt{\frac{1}{9} + \frac{1}{9}} = 3.66 \pm 4.71.$$

Das Konfidenzintervall $[-1.05, 8.37]$ ist also sehr weit und enthält sowohl positive wie negative Werte. Eine Verkürzung der mittleren Assemblierungszeiten wird also durch diesen Test nicht belegt (zumindest nicht mit dem hohen Konfidenzniveau von 95 %).

4. Konfidenzintervalle für σ^2

In manchen Anwendungen wird nicht nur ein Schätzwert, sondern auch ein Konfidenzintervall für die Populationsvarianz benötigt. Beispielsweise ist man in der Pharmazie nicht nur am mittleren Wirkstoffgehalt bei Tabletten interessiert, sondern es sollte auch die Streuung der enthaltenen Wirkstoffmenge von Tablette zu Tablette in vorgegebenen Grenzen bleiben.

Wir gehen wieder von einer Stichprobe X_1, \dots, X_n aus einer *normalverteilten* Grundgesamtheit aus. Dann ist die Testgröße

$$U = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma^2} \right)^2 = \frac{(n-1)S^2}{\sigma^2}$$

nach Proposition 3.7 χ^2 -verteilt mit $n - 1$ Freiheitsgraden und entsprechend ein $1 - \alpha$ Konfidenzintervall für U gegeben durch

$$P(\chi_u^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_o^2) = 1 - \alpha.$$

Da die Dichte der χ^2 -Verteilung nicht symmetrisch ist, wählen wir χ_u^2 und χ_o^2 so, daß $F_{\chi_u^2} = \frac{\alpha}{2}$ und $F_{\chi_o^2} = 1 - \frac{\alpha}{2}$ und bezeichnen diese Grenzen mit $\chi_u^2 = \chi_{\alpha/2}^2$ bzw. $\chi_o^2 = \chi_{1-\alpha/2}^2$. Diese Wahl liefert allgemein nicht das kürzeste Konfidenzintervall, ist dafür aber leicht zu berechnen. Die Grenzen des entsprechenden $1 - \alpha$ Konfidenzintervalles für σ^2 sind demnach bestimmt durch die Schätzfunktionen

$$(3.25) \quad \hat{\chi}_u^2 = \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}, \quad \hat{\chi}_o^2 = \frac{(n-1)S^2}{\chi_{\alpha/2}^2}$$

BEISPIEL 3.8. Die Varianz von Messungen ist ein Maß für die Güte der verwendeten Instrumente. Akustische Messungen ergaben die Werte 4.1dB, 5.2dB und 10.2dB. Gesucht ist ein 95% Konfidenzintervall für σ^2 .

Unter der Voraussetzung, daß die verwendete Apparatur normalverteilte Messungen liefert, ist ein 95% Konfidenzintervall für σ^2 durch (3.25) bestimmt: $s^2 = 10.57$, $n = 3$ und $\alpha/2 = 0.05$. Einer Tabelle der χ^2 -Verteilung entnimmt man die Werte $\chi_{0.95}^2 = 5.991$ und $\chi_{0.05}^2 = 0.103$. Dies ergibt die Grenzen

$$\hat{\chi}_u^2 = \frac{2 \cdot 10.57}{5.991} = 3.53, \quad \text{und} \quad \hat{\chi}_o^2 = \frac{2 \cdot 10.57}{0.103} = 205.24.$$

Das 95% Konfidenzintervall ist wegen des geringen Stichprobenumfanges sehr weit.

Wir haben uns bisher fast ausschließlich mit dem Schätzen von Populationsparametern beschäftigt. Es war deshalb meist klar, wie ein Schätzwert aus der Stichprobe gewonnen werden könnte. In den beiden folgenden Abschnitten stellen wir systematische Verfahren zur Konstruktion von Schätzfunktionen vor. Das einfachste Verfahren ist die Momentenmethode.

5. Momentenmethode

Erwartungswert und Varianz sind zwar die wichtigsten Kenngrößen einer Verteilung, durch sie allein ist eine Verteilung jedoch noch nicht festgelegt. Erinnern wir uns: Erwartungswert und Varianz können auch aufgefaßt werden als erstes und zweites Moment einer Verteilung X . Allgemein wurden die höheren Momente definiert durch

$$\mu_k = E(X^k), k = 1, 2, \dots$$

Die höheren Momente bestimmen eindeutig die momenterzeugende Funktion und damit auch die Verteilung. Die **Momentenmethode** beruht auf der Annahme, daß

die k -ten Stichprobenmomente

$$(3.26) \quad m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

eine vernünftige Näherung der entsprechenden Populationsmomente bieten. Angenommen, die Verteilung einer Zufallsvariablen X hängt von k unbekanntem Parametern $\theta_1, \dots, \theta_k$ ab. Diese Abhängigkeit überträgt sich auf die höheren Momente $\mu_i = \mu_i(\theta_1, \dots, \theta_k)$. Bei der Momentenmethode versucht man, Schätzwerte für die Parameter $\theta_1, \dots, \theta_k$ durch Lösen des nichtlinearen Gleichungssystems

$$(3.27) \quad \mu_i(\theta_1, \dots, \theta_k) = m_i, \quad i = 1, \dots, k$$

zu gewinnen. Wir veranschaulichen die Methode an zwei Beispielen:

BEISPIEL 3.9. Eine Stichprobe X_1, \dots, X_n soll einer über $[0, \theta]$ gleichverteilten Grundgesamtheit entnommen und ein Schätzwert für θ bestimmt werden.

Das erste Moment einer gleichverteilten Zufallsvariablen ist gegeben durch

$$\mu_1 = \mu = \frac{\theta}{2}.$$

Das entsprechende Stichprobenmoment ist das Stichprobenmittel

$$m_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Der Schätzwert für θ ergibt sich daher aus

$$\mu_1 = \frac{\theta}{2} = \bar{X}$$

zu

$$\hat{\theta} = 2\bar{X}.$$

Wir bemerken, daß wegen

$$E(\hat{\theta}) = 2E(\bar{X}) = 2\mu = \theta$$

der Schätzer erwartungstreu ist.

BEISPIEL 3.10. Eine Stichprobe X_1, \dots, X_n wird einer Grundgesamtheit entnommen, in welcher das interessierende Merkmal Gamma verteilt ist mit den Parametern (α, λ) .

Die ersten beiden Momente der Gamma Verteilung sind gegeben durch

$$\mu_1 = \frac{\alpha}{\lambda}$$

$$\mu_2 = \mu^2 + \sigma^2 = \frac{\alpha(\alpha + 1)}{\lambda^2}.$$

Die Momentenschätzer $(\hat{\alpha}, \hat{\lambda})$ ergeben sich aus dem Gleichungssystem

$$\begin{aligned}\frac{\alpha}{\lambda} &= m_1, \\ \frac{\alpha(\alpha + 1)}{\lambda^2} &= m_2\end{aligned}$$

zu

$$\begin{aligned}\hat{\lambda} &= \frac{\bar{X}}{\hat{\sigma}^2}, \\ \hat{\alpha} &= \frac{\bar{X}^2}{\hat{\sigma}^2}\end{aligned}$$

zu

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

6. Maximum Likelihood Methode

Angenommen, die Verteilung einer Eigenschaft hängt von Parametern $\theta_1, \dots, \theta_k$ ab, welche mittels einer Stichprobe X_1, \dots, X_n geschätzt werden sollen. Die gemeinsame Wahrscheinlichkeitsdichte (oder -verteilung bei diskreten Variable n) sei

$$f(x_1, \dots, x_n | \theta_1, \dots, \theta_k).$$

Man definiert die **Likelihoodfunktion** durch

$$(3.28) \quad \text{lik}(\theta_1, \dots, \theta_k) = f(X_1, \dots, X_n | \theta_1, \dots, \theta_k).$$

Beobachten wir durch die Stichprobe also die Werte $X_i = x_i$, dann stellt bei einer diskreten, gemeinsamen Verteilung die Likelihoodfunktion gerade die Wahrscheinlichkeit in Abhängigkeit von den Parametern $\theta = (\theta_1, \dots, \theta_k)$ dar, die konkrete Stichprobe zu ziehen. Der **Maximum Likelihood Schätzer (MLS)** für θ ist jener Parametersatz $\hat{\theta}$, welcher diese Wahrscheinlichkeit maximiert.

Im Falle unabhängiger Stichprobenvariabler vereinfacht sich die Berechnung des MLS, indem man zur **logarithmischen Likelihoodfunktion** $\ell(\theta)$ übergeht: in diesem Falle ist nämlich

$$\text{lik}(\theta) = \prod_{i=1}^n f(X_i | \theta_1, \dots, \theta_k)$$

und somit

$$(3.29) \quad \ell(\theta) = \sum_{i=1}^n \ln [f(X_i | \theta_1, \dots, \theta_k)]$$

BEISPIEL 3.11. Die n -malige Wiederholung eines Bernoulliexperimentes ergab die Werte x_1, \dots, x_n . Die Zahlen x_i können nur die Werte 1 oder 0 annehmen, je nachdem die i -te Wiederholung des Experimentes erfolgreich war oder nicht. Der MLS für die Erfolgswahrscheinlichkeit p ergibt sich aus der Likelihoodfunktion für diese Stichprobe

$$\text{lik}(p) = P(X_1 = x_1, \dots, X_n = x_n | p) = p^x (1 - p)^{n-x}$$

mit

$$x = \sum_{i=1}^n x_i.$$

aus der notwendigen Optimalitätsbedingung $\text{lik}(p)' = 0$ zu

$$\hat{p} = \frac{x}{n},$$

also jenen Schätzer, welchen wir bereits auf einer intuitiven Weise eingeführt haben. Es ist klar, daß die Likelihoodfunktion in \hat{p} das globale Maximum annimmt: es ist nämlich $\text{lik}(p) \geq 0$ und \hat{p} der einzige stationäre Punkt in $(0, 1)$.

BEISPIEL 3.12. Wir berechnen nun die MLS für Mittelwert μ und Varianz σ^2 einer Normalverteilung aus einer Stichprobe X_1, \dots, X_n . Die Stichprobenvariablen seien unabhängig.

Die gemeinsame Wahrscheinlichkeitsdichte der Stichprobe $X_i = x_i$, $i = 1, \dots, n$ ist wegen der Unabhängigkeit von X_1, \dots, X_n gegeben durch

$$\begin{aligned} \text{lik}(\mu, \sigma^2) &= f(x_1, \dots, x_n | \mu, \sigma^2) = \prod_{i=1}^n f(x_i | \mu, \sigma^2) \\ &= \prod_{i=1}^n \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned}$$

Für die logarithmische Likelihoodfunktion erhält man also

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Die stationären Punkte der Likelihoodfunktion ergeben sich aus den Nullstellen des Gradienten

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

zu

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Man beachte, daß zwar der Schätzer für das Populationsmittel erwartungstreu ist, nicht aber jener für die Varianz.

BEISPIEL 3.13. Es sei X_1, \dots, X_n eine Stichprobe einer auf $[0, \theta]$ gleichverteilten Zufallsvariablen. Man bestimme den MLS für θ unter der Annahme der Unabhängigkeit von X_1, \dots, X_n .

In diesem Falle ist die Likelihoodfunktion gegeben durch

$$\begin{aligned} \text{lik}(\theta) &= f(x_1, \dots, x_n | \theta) = f(x_1 | \theta) \dots f(x_n | \theta) \\ &= \begin{cases} \frac{1}{\theta^n} & \text{falls } x_i \in [0, \theta], i = 1, \dots, n \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Die Likelihoodfunktion ist streng monoton fallend. Das Maximum wird daher für den kleinsten Wert von θ angenommen, welcher mit den Nebenbedingungen $x_i \leq \theta$, $i = 1, \dots, n$ verträglich ist. Dies ist offenbar für $\hat{\theta} = \max(x_1, \dots, x_n)$ der Fall. Somit ist der MLS für θ

$$\hat{\theta} = \max(X_1, \dots, X_n) = X_{(n)}.$$

Dieser Schätzer ist nicht erwartungstreu. Wegen (2.123) ist die Verteilung von $X_{(n)}$ gegeben durch

$$g_{(n)}(x) = \begin{cases} \frac{n}{\theta^*} \left(\frac{x}{\theta^*}\right)^{n-1} & 0 \leq x \leq \theta^* \\ 0 & \text{sonst,} \end{cases}$$

wobei θ^* den wahren Wert von θ bezeichnet. Der Erwartungswert von $\hat{\theta}$ ist demnach gegeben durch

$$E(\hat{\theta}) = \frac{n}{\theta^*} \int_0^{\theta^*} x^n dx = \frac{n}{n+1} \theta^*.$$

Diese Rechnung zeigt aber auch, daß die einfache Modifikation

$$\frac{n+1}{n} \hat{\theta} = \frac{n+1}{n} X_{(n)}$$

ein erwartungstreuer Schätzer ist.

7. Asymptotische Eigenschaften eines MLS

In diesem Abschnitt skizzieren wir die asymptotische Verteilung eines MLS. Da diese Untersuchungen sehr aufwendig sind, beschränken wir uns auf den Fall eines skalaren Parameters und verzichten auch auf Beweise. Wir bezeichnen den wahren

Wert des Parameters mit θ^* , ferner setzen wir unabhängige, identisch verteilte Stichprobenvariablen X_1, \dots, x_n voraus, sodaß die logarithmische Likelihoodfunktion gegeben ist durch

$$\ell(\theta) = \sum_{i=1}^n \ln f(X_i|\theta)$$

Eine wesentliche Rolle bei der Untersuchung des asymptotischen Verhaltens eines MLS spielt die Größe

$$(3.30) \quad I(\theta) = E\left[\frac{\partial}{\partial\theta} \ln f(X|\theta)\right]^2,$$

wobei X eine der Stichprobenvariablen $X_i, i = 1, \dots, n$ repräsentiert. Die Auswertung von $I(\theta)$ kann manchmal durch die Identität

$$(3.31) \quad I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f(X|\theta)\right]$$

vereinfacht werden.

PROPOSITION 3.8. *Es sei $\hat{\theta}_n$ ein MLS für den unbekannt Parameter θ^* , welcher aus einer unabhängigen, identisch verteilten Stichprobe vom Umfang n berechnet wird. Unter geeigneten Glattheitsvoraussetzungen an die Wahrscheinlichkeitsdichte f gilt dann*

1. *Die Folge der MLS ($\hat{\theta}_n$) konvergiert stochastisch gegen θ^* .*
2. *Die Folge der Verteilungsfunktionen der Zufallsvariablen $(\sqrt{nI(\theta^*)}(\hat{\theta}_n - \theta^*))$ konvergiert punktweise gegen die Verteilungsfunktion der Standardnormalverteilung.*

Der Mittelwert eines MLS ist also asymptotisch gleich θ^* , seine Varianz ist asymptotisch gleich $\frac{1}{nI(\theta^*)}$. Ein MLS ist also zumindest asymptotisch erwartungstreu.

Man kann diese Asymptotik verwenden, um ein näherungsweise Konfidenzintervall für den MLS in jenen Fällen zu berechnen, in welchen keine geeignete Pivotgröße zur Verfügung steht. Man kann dann $\sqrt{nI(\theta^*)}(\hat{\theta}_n - \theta^*)$ als Prüfgröße ansetzen. Da θ^* nicht bekannt ist, ersetzt man $I(\theta^*)$ durch $I(\hat{\theta})$. Dann ist aber auch $\sqrt{nI(\hat{\theta})}(\hat{\theta}_n - \theta^*)$ asymptotisch standard normalverteilt. Somit ergibt sich ein näherungsweise $(1 - \alpha)$ Konfidenzintervall aus der Bedingung

$$P(-z_{\alpha/2} \leq \sqrt{nI(\theta^*)}(\hat{\theta}_n - \theta^*) \leq z_{\alpha/2}) = 1 - \alpha.$$

Die Grenzen eines näherungsweise $(1 - \alpha)$ Konfidenzintervalles sind daher durch die Schätzfunktionen

$$(3.32) \quad \hat{\theta}_u = \hat{\theta}_n - z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta})}}, \quad \hat{\theta}_o = \hat{\theta}_n + z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta})}}$$

gegeben.

BEISPIEL 3.14. Der MLS für den Parameter der Poissonverteilung ist

$$\hat{\lambda} = \bar{X}.$$

Ein exaktes Konfidenzintervall kann aus dem Umstand, daß $n\hat{\lambda} = \sum_{i=1}^n X_i$ selbst wieder einer Poissonverteilung genügt, abgeleitet werden. Für große Stichprobenumfänge kann man einfacher auf (3.32) zurückgreifen. Wir berechnen zuerst

$$I(\lambda) = E\left[\frac{\partial}{\partial \lambda} \ln f(X|\lambda)\right]^2.$$

In diesem Falle ist

$$f(x|\lambda) = e^{-\lambda} \frac{\lambda^x}{x!},$$

also

$$\ell(\lambda) = \ln f(x|\lambda) = -\lambda + x \ln \lambda - \ln x!$$

und daher

$$I(\lambda) = E\left(\frac{X}{\lambda} - 1\right)^2 = \frac{1}{\lambda}.$$

Darin ersetzt man λ durch $\hat{\lambda} = \bar{X}$. Dies ergibt die Grenzen eines approximativen $1 - \alpha$ Konfidenzintervalles

$$\bar{X} \pm z_{\alpha/2} \sqrt{\frac{\bar{X}}{n}}.$$

8. Konsistenz von Schätzfunktionen

Wir haben bereits ausgeführt, daß Erwartungstreue und geringe mittlere quadratische Abweichung wünschenswerte Eigenschaften von Schätzfunktionen sind. Die mittlere quadratische Abweichung ist gegeben durch

$$E((\hat{\theta} - \theta)^2) = V(\theta) + B^2,$$

wobei $B = \hat{\theta} - \theta$ den Bias des Schätzers bezeichnet. Für erwartungstreue Schätzer fällt also die mittlere quadratische Abweichung gerade mit seiner Varianz zusammen. Nach der Ungleichung von Tschebyscheff gilt

$$(3.33) \quad P(|\hat{\theta} - E(\hat{\theta})| > \frac{\varepsilon}{\sigma_{\hat{\theta}}}) \leq \frac{1}{\left(\frac{\varepsilon}{\sigma_{\hat{\theta}}}\right)^2} = \frac{V(\hat{\theta})}{\varepsilon^2}.$$

Je kleiner die Varianz des Schätzers, desto geringer ist die Wahrscheinlichkeit großer Abweichungen des Schätzwertes von seinem Erwartungswert, also dem wahren Parameter, falls der Schätzer erwartungstreu ist. Diese Überlegung motiviert folgenden Begriff

DEFINITION 3.2. Es sei $\hat{\theta}_n$ eine Schätzfunktion, welche eine Stichprobe vom Umfang n verwendet. Der Schätzer $\hat{\theta}_n$ heißt **konsistent**, wenn die Folge $(\hat{\theta}_n)$ stochastisch gegen θ konvergiert, d.h. wenn für jedes $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

gilt.

Nach Proposition 3.8 sind Maximum Likelihood Schätzer für hinreichend glatte Likelihoodfunktionen konsistent. Eine hinreichende Bedingung für Konsistenz folgt aus (3.33):

PROPOSITION 3.9. Es sei $(\hat{\theta}_n)$ eine Folge erwartungstreuer Schätzer für den Parameter θ . Die Schätzer $(\hat{\theta}_n)$ sind konsistent, wenn

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$$

gilt.

BEWEIS. Wegen der Erwartungstreue ist (3.33) gleichwertig mit

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{V(\hat{\theta}_n)}{\varepsilon}.$$

Daraus folgt die Behauptung. □

Es sei beispielsweise X_1, \dots, X_n eine Stichprobe einer Grundgesamtheit mit Mittel μ und Varianz σ^2 . Das Gesetz der großen Zahlen bringt zum Ausdruck, daß das Stichprobenmittel \bar{X}_n ein konsistenter Schätzer für μ ist.

Es ist nicht schwer, sich davon zu überzeugen, daß für stochastisch konvergente Folgen von Zufallsvariablen dieselben Regeln gelten, wie für konvergente Folgen reeller Zahlen. Konvergieren also die Folgen $(\hat{\theta}_n)$, bzw. $(\tilde{\theta}_n)$ stochastisch gegen θ bzw. gegen $\tilde{\theta}$, dann gilt auch

$$\lim_{n \rightarrow \infty} \hat{\theta}_n \pm \tilde{\theta}_n = \theta \pm \tilde{\theta},$$

$$\lim_{n \rightarrow \infty} \hat{\theta}_n \tilde{\theta}_n = \theta \tilde{\theta},$$

$$\lim_{n \rightarrow \infty} \frac{\hat{\theta}_n}{\tilde{\theta}_n} = \frac{\theta}{\tilde{\theta}} \text{ soferne } P(\tilde{\theta} \neq 0) = 1$$

$$\lim_{n \rightarrow \infty} g(\hat{\theta}_n) = g(\theta), \text{ für stetige Funktionen } g$$

Als Anwendung zeigen wir, daß der erwartungstreue Schätzer für die Populationsvarianz

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right)$$

i.A. konsistent ist. Es seien also X_1, \dots, X_n unabhängige Stichprobenvariable, deren Momente $E(X_i) = \mu$, $E(X_i^2) = \mu_2$ und $E(X_i^4) = \mu_4$, $i = 1, \dots, n$ endlich sind. Das zweite Stichprobenmoment $\frac{1}{n} \sum_{i=1}^n X_i^2$ ist das Mittel von unabhängigen identisch verteilten Zufallsvariablen mit $E(X_i^2) = \mu_2$ und $V(X_i^2) = E(X_i^4) - E(X_i^2)^2 = \mu_4 - \mu_2^2 < \infty$. Nach dem Gesetz der großen Zahlen konvergiert $\frac{1}{n} \sum_{i=1}^n X_i^2$ stochastisch gegen μ_2 und \bar{X}_n^2 konvergiert stochastisch gegen μ^2 . In der Folge konvergiert

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2$$

stochastisch gegen $\mu_2 - \mu^2 = \sigma^2$ und damit gilt auch $\lim_{n \rightarrow \infty} S_n^2 = \sigma^2$ im Sinne der stochastischen Konvergenz.

Ohne Beweis zitieren wir einen Spezialfall des Satzes von Slutsky:

PROPOSITION 3.10. *Es sei (U_n) eine Folge von Zufallsvariablen, deren Verteilungsfunktionen punktweise gegen die Verteilungsfunktion der Standardnormalverteilung konvergiert. Die Folge (W_n) konvergiere stochastisch gegen 1. Dann konvergiert die Verteilungsfunktion von $\frac{U_n}{W_n}$ punktweise gegen die Verteilungsfunktion der Standardnormalverteilung.*

Als Anwendung erinnern wir daran, daß ein für große Stichproben gültiges $1 - \alpha$ Konfidenzintervall für das Populationsmittel die Grenzen

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

besitzt. Um die Grenzen auswerten zu können, wurde die unbekannte Populationsvarianz durch den Schätzer S_n^2 ersetzt. Proposition 3.10 bietet die theoretische Rechtfertigung für diese Vorgangsweise: dem approximativen Konfidenzintervall

$$\left[\bar{X} - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right]$$

liegt die Pivotgröße

$$T_n = \sqrt{n} \frac{\bar{X} - \mu}{S_n} = \frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{S_n / \sigma} = \frac{U_n}{W_n}$$

zugrunde. Die Zufallsvariablen U_n und W_n erfüllen die Voraussetzungen von Proposition 3.10, sodaß die Verteilung von T_n punktweise gegen jene der Standardnormalverteilung konvergiert. Die Testgrößen T_n sind also für große Stichprobenumfänge ungefähr standard normalverteilt unabhängig von der Ausgangsverteilung der Stichprobenvariablen X_1, \dots, X_n . Wir erinnern daran, daß T_n einer t -Verteilung mit $n - 1$ Freiheitsgraden genügt, falls die Stichprobe einer normalverteilten Grundgesamtheit entnommen wurde. Wir schließen somit, daß die Verteilung der t -Verteilung punktweise gegen jene der Standardnormalverteilung konvergiert, wenn die Anzahl der Freiheitsgrade beliebig groß wird.

9. Relative Effizienz

Meist gibt es mehrere Möglichkeiten, erwartungstreue Schätzer $\hat{\theta}$ für den Parameter θ zu konstruieren. Wegen der geringeren Streuung der Schätzwerte bei wiederholten Stichprobenentnahmen bevorzugt man meist den Schätzer mit der geringeren Varianz. Der quantitative Vergleich zweier erwartungstreuer Schätzer wird durch folgenden Begriff erleichtert.

DEFINITION 3.3. *Es seien $\hat{\theta}_i$, $i = 1, 2$ erwartungstreue Schätzer des selben Parameters θ . Die **Effizienz von $\hat{\theta}_1$ relative zu $\hat{\theta}_2$** , $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$ ist definiert durch das Verhältnis*

$$(3.34) \quad \text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

BEISPIEL 3.15. Es seien X_1, \dots, X_n unabhängige, identisch verteilte Stichprobenvariable auf einer auf $[0, \theta]$ gleichverteilten Grundgesamtheit. Es wurde bereits gezeigt, daß

$$\hat{\theta}_1 = 2\bar{X}, \quad \text{und} \quad \hat{\theta}_2 = \frac{n+1}{n}X_{(n)}$$

erwartungstreue Schätzer für θ sind. Man verifiziere

$$V(\hat{\theta}_1) = V(2\bar{X}) = 4V(\bar{X}) = 4 \frac{V(X_i)}{n} = \frac{1}{3n}\theta^2,$$

$$V(\hat{\theta}_2) = V\left(\frac{n+1}{n}X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 V(X_{(n)}) = \frac{1}{n(n+2)}\theta^2$$

und somit

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{3}{n+2} < 1, \quad n > 1.$$

Somit hat $\hat{\theta}_2$ eine kleinere Varianz als $\hat{\theta}_1$ und ist daher als Schätzer vorzuziehen.

Im folgenden zeigen wir, daß die Varianz erwartungstreuer Schätzer eines Parameters nicht beliebig klein werden kann:

PROPOSITION 3.11 (Ungleichung von Cramer–Rao). *Es seien X_1, \dots, X_n unabhängige, identisch verteilte Zufallsvariable mit Wahrscheinlichkeitsdichte $f(x|\theta^*)$ und $\hat{\theta} = t(X_1, \dots, X_n)$ ein beliebiger erwartungstreuer Schätzer für θ^* . Wenn die Wahrscheinlichkeitsdichte hinreichend glatt ist, dann gilt*

$$(3.35) \quad V(\hat{\theta}) \geq \frac{1}{nI(\theta^*)}$$

wobei $I(\theta^*)$ gegeben ist durch

$$I(\theta^*) = E\left[\frac{\partial}{\partial \theta} \ln f(X_i|\theta^*)\right]^2$$

Das Bemerkenswerte an dieser Ungleichung ist, daß die untere Schranke scharf ist: erwartungstreue Schätzer, deren Varianz diese untere Schranke annehmen, heißen **effizient**. Wir haben bereits erwähnt, daß die Varianz von Maximum Likelihood Schätzern gegen die Cramer–Rao Schranke konvergieren. ML Schätzer sind daher **asymptotisch effizient**.

BEWEISSKIZZE DER UNGLEICHUNG VON CRAMER–RAO. Wir setzen

$$Z = \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln f(X_i|\theta^*) = \sum_{i=1}^n \frac{\frac{\partial}{\partial \theta} f(X_i|\theta^*)}{f(X_i|\theta^*)},$$

das ist die Ableitung der logarithmischen Likelihoodfunktion. Wir zeigen zuerst

$$E(Z) = 0.$$

Dies folgt aus dem Umstand, daß der erwartungswert jedes summanden von Z Null ist. In der Tat, es gilt doch (man beachte, daß die Stichprobenvariablen identisch verteilt sind)

$$\begin{aligned} E\left(\frac{\frac{\partial}{\partial \theta} f(X_i|\theta^*)}{f(X_i|\theta^*)}\right) &= \int \frac{\partial}{\partial \theta} f(x|\theta^*) f(x|\theta^*) dx \\ &= \int \frac{\frac{\partial}{\partial \theta} f(x|\theta^*)}{f(x|\theta^*)} f(x|\theta^*) dx = \int \frac{\partial}{\partial \theta} f(x|\theta^*) dx \\ &= \frac{\partial}{\partial \theta} \underbrace{\int f(x|\theta^*) dx}_{=1} = 0. \end{aligned}$$

Zusammen mit dem Umstand, daß Z eine Summe von stochastisch unabhängigen Zufallsvariablen ist (dies ohne Beweis), erhält man nun

$$\begin{aligned} V(Z) &= nV\left(\frac{\partial}{\partial \theta} \ln f(X_i|\theta^*)\right) \\ &= nE\left(\frac{\partial}{\partial \theta} \ln f(X_i|\theta^*)\right)^2 = nI(\theta^*). \end{aligned}$$

Da für den Korrelationskoeffizienten ρ zwischen zwei Zufallsvariablen stets die Ungleichung $\rho^2 \leq 1$ gilt, folgt insbesondere

$$\text{Cov}^2(Z, \hat{\theta}) \leq V(Z)V(\hat{\theta}),$$

also

$$V(\hat{\theta}) \geq \frac{\text{Cov}^2(Z, \hat{\theta})}{V(Z)}.$$

Die Behauptung folgt nun aus $\text{Cov}(Z, \hat{\theta}) = 1$. Um dies einzusehen, beachten wir vorerst $\text{Cov}(Z, \hat{\theta}) = E(Z\hat{\theta})$ (wegen $E(Z) = 0$). Somit ergibt sich

$$\begin{aligned} \text{Cov}(Z, \hat{\theta}) &= E(Z\hat{\theta}) = \int \cdots \int t(x_1, \dots, x_n) \underbrace{\sum_{i=1}^n \frac{\frac{\partial}{\partial \theta} f(X_i|\theta^*)}{f(X_i|\theta^*)} \prod_{j=1}^n f(x_j|\theta^*)}_{\frac{\partial}{\partial \theta} \prod_{j=1}^n f(x_j|\theta^*)} dx_1 \cdots dx_n \\ &= \int \cdots \int t(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{j=1}^n f(x_j|\theta^*) dx_1 \cdots dx_n \\ &\stackrel{*}{=} \frac{\partial}{\partial \theta} \int \cdots \int t(x_1, \dots, x_n) \prod_{j=1}^n f(x_j|\theta) dx_1 \cdots dx_n \Big|_{\theta=\theta^*} \\ &= \frac{\partial}{\partial \theta} E(\hat{\theta}) \Big|_{\theta=\theta^*} = \frac{\partial}{\partial \theta} \theta = 1 \end{aligned}$$

In den mit * markierten Gleichungen wurde formal die Reihenfolge von Integration und Differentiation vertauscht. Dies kann bei hinreichender Glattheit des Integranden gerechtfertigt werden. \square

BEISPIEL 3.16. In Beispiel 3.14 haben wir gesehen, daß $\hat{\lambda} = \bar{X}$ ein MLS für den Parameter λ der Poisson Verteilung darstellt. Bei dieser Gelegenheit haben wir auch $I(\lambda)$ berechnet:

$$I(\lambda) = \frac{1}{\lambda}.$$

Somit gilt für jeden erwartungstreuen Schätzer $\hat{\theta}$, der eine unabhängige Stichprobe verwendet, die Cramer–Rao Schranke für dessen Varianz

$$V(\hat{\theta}) \geq \frac{\lambda}{n}.$$

Für die Varianz des MLS findet man

$$V(\hat{\lambda}) = V(\bar{X}) = \frac{\sigma^2}{n} = \frac{\lambda}{n},$$

der MLS ist daher effizient. Kein anderer erwartungstreuer Schätzer kann eine kleinere Varianz haben.

10. Suffizienz

Wir haben nun einige systematische Verfahren zur Konstruktion von Schätzfunktionen für einen Parameter θ kennengelernt. Die Schätzfunktion benützt die Stichprobe, um einen Schätzwert für θ zu berechnen. Wir wenden uns nun der naheliegenden Frage zu, ob eine gegebene Schätzfunktion *alle* Information verwendet, welche in der Stichprobe über den Parameter θ enthalten ist. Als Beispiel betrachten wir eine Folge von Bernoulli Experimenten mit einer unbekanntem Erfolgswahrscheinlichkeit p .

Intuitiv sind wir überzeugt, daß in der Gesamtanzahl von Erfolgen bei n Wiederholungen des Experimentes die gesamte Information über p enthalten ist, also beispielsweise die Reihenfolge, in welcher die Erfolge eintreten, irrelevant ist. Der folgende Begriff formalisiert diese Vorstellung.

DEFINITION 3.4. *Es sei X_1, \dots, X_n Stichprobenvariable aus einer Grundgesamtheit, deren Verteilung von einem Parameter θ abhängt. Eine **Statistik** bezeichnet im Folgenden auch eine Funktion der Stichprobenvariablen. Eine Statistik $T = T(X_1, \dots, X_n)$ heißt **suffizient (erschöpfend)**, wenn die bedingte Verteilung*

$$(3.36) \quad P(X = x_1, \dots, X_n = x_n | T = t, \Theta = \theta)$$

für jeden Wert von t unabhängig ist von θ .

BEISPIEL 3.17. Zur Illustration greifen wir noch einmal das Beispiel von n unabhängigen Wiederholungen eines Bernoulli Experimentes auf. Die i -te Wiederholung wird durch die Zufallsvariable X_i beschrieben, deren Verteilung durch $P(X_i = 1 | \Theta = p) = p$, $P(X_i = 0 | \Theta = p) = 1 - p$, $i = 1, \dots, n$, beschrieben. Wir betrachten die Statistik $T = \sum_{i=1}^n X_i$. Angenommen, bei n Wiederholungen des Experimentes traten t Erfolge auf. Dann ist

$$\begin{aligned} P(X = x_1, \dots, X_n = x_n | T = t, \Theta = p) &= \frac{P(X = x_1, \dots, X_n = x_n, T = t | \Theta = p)}{P(T = t | \Theta = p)} \\ &= \begin{cases} \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{1}{\binom{n}{t}}, & \text{falls } t = \sum_{i=1}^n x_i, \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Da die bedingte Wahrscheinlichkeit (3.36) nicht von p abhängt, ist die Statistik T suffizient.

Wir geben nun eine in vielen Fällen leichter verifizierbare Charakterisierung der Suffizienz einer Statistik.

PROPOSITION 3.12 (Faktorkriterium). *Eine notwendige und hinreichende Bedingung dafür, daß eine Statistik $T(X_1, \dots, X_n)$ suffizient ist für einen Parameter θ , besteht darin, daß die Likelihoodfunktion der Stichprobe faktorisiert in der Form*

$$(3.37) \quad \text{lik}(\theta) = f(x_1, \dots, x_n | \theta) = g(T(x_1, \dots, x_n), \theta) h(x_1, \dots, x_n).$$

Der Faktor g hängt also nur vom Wert der Statistik T für die konkrete Stichprobe und vom Parameter θ ab, der Faktor h ist unabhängig von θ .

BEWEIS. Wir führen den Beweis nur für den diskreten Fall und zeigen zuerst, daß eine Faktorisierung in der angegebenen Form hinreichend ist. Es ist

$$\begin{aligned} P(T = t | \Theta = \theta) &= \sum_{T(x_1, \dots, x_n) = t} P(X_1 = x_1, \dots, X_n = x_n | \Theta = \theta) \\ &= \sum_{T(x_1, \dots, x_n) = t} f(x_1, \dots, x_n | \theta) = g(t, \theta) \sum_{T(x_1, \dots, x_n) = t} h(x_1, \dots, x_n). \end{aligned}$$

Somit erhalten wir für die bedingte Wahrscheinlichkeit (3.36) (wir betrachten nur $T = T(x_1, \dots, x_n)$, andernfalls ist die Wahrscheinlichkeit Null)

$$\begin{aligned} P(X = x_1, \dots, X_n = x_n | T = t, \Theta = \theta) &= \frac{P(X = x_1, \dots, X_n = x_n, T = t | \Theta = \theta)}{P(T = t | \Theta = \theta)} \\ &= \frac{g(t, \theta) h(x_1, \dots, x_n)}{g(t, \theta) \sum_{T(x_1, \dots, x_n) = t} h(x_1, \dots, x_n)} \\ &= \frac{h(x_1, \dots, x_n)}{\sum_{T(x_1, \dots, x_n) = t} h(x_1, \dots, x_n)} \end{aligned}$$

also einen Ausdruck, der von θ nicht abhängt. Die Statistik $T(X_1, \dots, X_n)$ ist daher suffizient.

Umgekehrt sei nun die Statistik $T(X_1, \dots, X_n)$ suffizient, die bedingte Wahrscheinlichkeit (3.36) also von θ unabhängig. Setzt man

$$g(t, \theta) = P(T = t(x_1, \dots, x_n) | \Theta = \theta),$$

$$h(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n | T = t(x_1, \dots, x_n), \Theta = \theta)$$

findet man für die Wahrscheinlichkeit der Stichprobe für einen bestimmten Parameter

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | \Theta = \theta) &= \sum_t P(X_1 = x_1, \dots, X_n = x_n | T = t, \Theta = \theta) P(T = t | \Theta = \theta) \\ &= P(T = t(x_1, \dots, x_n) | \Theta = \theta) P(X_1 = x_1, \dots, X_n = x_n | T = t(x_1, \dots, x_n), \Theta = \theta) \\ &= g(t, \theta) h(x_1, \dots, x_n) \end{aligned}$$

die Faktorisierung (3.37) und der Suffizienz von T . □

COROLLARY 3.3. *Wenn eine Statistik T suffizient ist für einen Parameter θ , dann ist der MLS für θ eine Funktion von T .*

BEWEIS. Nach Proposition 3.12 hat die Likelihoodfunktion die Form

$$\text{lik}(\theta) = g(t, \theta) h(x_1, \dots, x_n).$$

und nimmt daher ihr Maximum an derselben Stelle wie g an. □

Dieser Satz stellt eine systematische Methode zur Verfügung, suffiziente Statistiken zu gewinnen:

BEISPIEL 3.18. Wir betrachten wieder eine Folge von unabhängigen Bernoulli-variablen X_1, \dots, X_n , also $P(X_i = x | \Theta = p) = p^x (1-p)^{1-x}$, $x = 0, 1$. Die Likelihoodfunktion einer Stichprobe ist daher gegeben durch

$$\begin{aligned} \text{lik}(p) &= \prod_{i=1}^n p^{x_i} (1-p)^{n-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \\ &= \left(\frac{p}{1-p} \right)^{\sum_{i=1}^n x_i} (1-p)^n. \end{aligned}$$

Sie besitzt also die Faktorisierung $g(\sum_{i=1}^n x_i, p)h(x_1, \dots, x_n)$ mit

$$g\left(\sum_{i=1}^n x_i, p\right) = \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i} (1-p)^n, \quad h(x_1, \dots, x_n) = 1,$$

dies bestätigt, daß $T = \sum_{i=1}^n X_i$ eine suffiziente Statistik für p ist.

Proposition 3.12 wurde zwar nur für skalare Parameter formuliert, die Charakterisierung gilt aber auch im mehrdimensionalen Fall:

BEISPIEL 3.19. Wir betrachten eine unabhängige Stichprobe aus einer normalverteilten Grundgesamtheit mit Erwartungswert μ und Varianz σ^2 . Die Likelihoodfunktion ist in diesem Fall gegeben durch

$$\begin{aligned} \text{lik}(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right] \\ &= \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \\ &= \frac{1}{\sigma^n(2\pi)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right)\right] \\ &= g\left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2, \mu, \sigma^2\right). \end{aligned}$$

Da dieser Ausdruck außer von den Parametern nur mehr von $t_1 = \sum_{i=1}^n x_i$ und $t_2 = \sum_{i=1}^n x_i^2$ abhängt, sind $T_1 = \sum_{i=1}^n X_i$ und $T_2 = \sum_{i=1}^n X_i^2$ suffiziente Statistiken für die Parameter (μ, σ^2) .

PROPOSITION 3.13 (Satz von Rao–Blackwell). *Es sei $\hat{\theta}$ ein erwartungstreuer Schätzer für einen Parameter θ mit $V(\hat{\theta})$ und T eine suffiziente Statistik für θ . Setzt man*

$$(3.38) \quad \hat{\theta}^* = E(\hat{\theta}|T),$$

dann ist $\hat{\theta}^$ ebenfalls ein erwartungstreuer Schätzer für θ mit*

$$V(\hat{\theta}^*) \leq V(\hat{\theta}).$$

BEWEISSKIZZE. Da T suffizient für θ ist, ist die bedingte Verteilung jeder Statistik (einschließlich $\hat{\theta}$) unter T unabhängig von θ . Somit hängt $\hat{\theta}^* = E(\hat{\theta}|T)$ nur von der Stichprobe, nicht aber von θ ab, ist also selbst eine Statistik. Die Erwartungstreue dieser Statistik ergibt sich aus Proposition 2.30:

$$E(\hat{\theta}^*) = E(E(\hat{\theta}|T)) = E(\hat{\theta}) = \theta.$$

Analog berechnen wir die Varianz von $\hat{\theta}^*$ mit Hilfe von Proposition 2.31

$$\begin{aligned} V(\hat{\theta}) &= V(E(\hat{\theta}|T)) + E(V(\hat{\theta}|T)) \\ &= V(\hat{\theta}^*) + E(V(\hat{\theta}|T)). \end{aligned}$$

Wegen $V(\hat{\theta}|T = t) \geq 0$ für alle t , folgt $E(V(\hat{\theta}|T)) \geq 0$ und somit auch $V(\hat{\theta}^*) \leq V(\hat{\theta})$. \square

Dieser Satz zeigt, daß ein erwartungstreuer Schätzer verbessert werden kann, wenn man eine suffiziente Statistik für T kennt. Kann man den Schätzer von Rao–Blackwell auf diese Weise verbessern? Nach der Definition des iterierten Erwartungswertes ist $\hat{\theta}^*$ eine Funktion von T , z.B. $\hat{\theta}^* = h(T)$, und man kann zeigen, daß daher eine erneute Anwendung des Satzes keine Verbesserung mehr bringt. Dies liegt im wesentlichen an $E(h(T)|T) = h(T)$.

Da es für einen Parameter viele suffiziente Statistiken gibt, stellt sich die Frage, welche dieser Statistiken man im Satz von Rao–Blackwell verwenden soll. Wir müssen uns in dieser Hinsicht auf einige Hinweise beschränken: Typischerweise ergibt der Faktorisierungssatz suffiziente Statistiken, welche in gewisser Weise optimal die Information über den Parameter zusammenfassen. Man nennt derartige Statistiken **minimale hinreichende Statistiken**. Man kann nun zeigen, daß der Rao–Blackwell Schätzer minimale Varianz besitzt, also effizient ist, wenn man eine minimale suffiziente Statistik verwendet. Die direkte Berechnung des bedingten Erwartungswertes kann allerdings sehr aufwendig sein. In der Praxis geht man daher oft folgendermaßen vor: Wenn T eine minimale suffiziente Statistik ist und eine Funktion von T , z.B. $h(T)$, mit $E(h(T)) = \theta$ gefunden werden kann, dann ist $h(T)$ der erwartungstreue Schätzer minimaler Varianz für θ . Wir illustrieren diese Technik an einigen Beispielen.

BEISPIEL 3.20. In Beispiel 3.18 haben wir mit Hilfe des Faktorisierungssatzes gezeigt, daß $T = \sum_{i=1}^n X_i$ eine hinreichende Statistik für die Erfolgswahrscheinlichkeit p bei n Wiederholungen eines Bernoulli Experimentes ist. Aus

$$E(T) = np$$

schließt man, daß

$$\hat{p}^* = \frac{T}{n} = \bar{X}$$

ein erwartungstreuer Schätzer für p ist. Da er mit Hilfe einer suffizienten Statistik gebildet wurde, hat er minimale Varianz. Um diese Bemerkung zu verifizieren, beachten wir einerseits $V(\hat{p}^*) = \frac{p(1-p)}{n}$ und berechnen andererseits die untere Schranke für die Varianz einer erwartungstreuen Schätzers nach Cramer–Rao: dazu benötigt man

$$I(p) = E\left[\frac{\partial}{\partial p} \ln f(X|p)\right]^2 = -E\left[\frac{\partial^2}{\partial p^2} \ln f(X|p)\right]$$

mit

$$\ln f(x|p) = p^x(1-p)^{1-x}.$$

Eine kurze Rechnung ergibt

$$I(p) = \frac{1}{p(1-p)}$$

Nach Cramer–Rao ist daher die Varianz eines beliebigen erwartungstreuen Schätzers nach unten beschränkt durch

$$V(\hat{\theta}) \geq \frac{1}{nI(p)} = \frac{p(1-p)}{n}.$$

BEISPIEL 3.21. Die **Weibullverteilung** ist ein gutes Modell für Lebensdauern. Ihre Dichte hängt von zwei positiven Parametern α und m ab:

$$f(x|\alpha, m) = \frac{mx^{m-1}}{\alpha} e^{-x^m/\alpha}, \quad x > 0.$$

Es sei X_1, \dots, X_n eine unabhängige Stichprobe aus einer Weibull verteilten Population mit $m = 2$. Gesucht ist ein effizienter Schätzer für α .

Als ersten Schritt bestimmen wir eine minimale suffiziente Statistik mit Hilfe des Faktorisierungssatzes:

$$\text{lik}(\alpha) = \prod_{i=1}^n f(x_i|\alpha, 2) = \underbrace{\left(\frac{2}{\alpha}\right)^n \exp\left(-\frac{2}{\alpha} \sum_{i=1}^n x_i^2\right)}_{g(\sum_{i=1}^n x_i^2, \alpha)} \underbrace{\prod_{i=1}^n x_i}_{h(x_1, \dots, x_n)}.$$

Dies ergibt die Statistik $T = \sum_{i=1}^n x_i^2$. Als nächstes bestimmen wir die Verteilung von $W = X_i^2$:

$$F_W(w) = P(W \leq w) = P(X_i \leq \sqrt{w}) = F_{X_i}(\sqrt{w})$$

$$f_W(w) = f_{X_i}(\sqrt{w}) \frac{1}{2\sqrt{w}} = \frac{1}{\alpha} e^{-w/\alpha}.$$

Dies ist eine Exponentialverteilung mit Parameter $\lambda = \frac{1}{\alpha}$. Aus Proposition 2.27 ergibt sich daher

$$E(X_i^2) = E(W) = \alpha \quad \text{also} \quad E\left(\sum_{i=1}^n X_i^2\right) = n\alpha.$$

Die Statistik T ist daher zwar nicht erwartungstreu, aber

$$\hat{\alpha}^* = \frac{1}{n} \sum_{i=1}^n X_i^2$$

ist ein erwartungstreuer Schätzer für den Weibullparameter α . Da er aus einer minimalen suffizienten Statistik abgeleitet wurde, besitzt er minimale Varianz.

11. Testen von Hypothesen

Das Ziel statistischer Verfahren ist stets, Information über unbekannte Populationsparameter aus Stichproben zu gewinnen. Zu diesem Zweck haben wir bereits einige Schätzverfahren besprochen. Auf Populationsparameter kann man aber auch durch Testen von statistischen Hypothesen schließen. In der Statistik versteht man unter einer **Hypothese** eine Annahme über einen Populationsparameter. Ein **Test** ist ein Prüfverfahren, mit dessen Hilfe man entscheidet, ob man die Hypothese annehmen oder verwerfen soll.

11.1. Das Neyman–Pearson Paradigma. Ein weit verbreitetes Testverfahren wurde von Neyman und Pearson entwickelt. Schematisch geht man folgendermaßen vor: Man gruppiert Verteilungen in zwei Teilmengen. Die eine nennt man **Nullhypothese**, die andere nennt man **Alternativhypothese**. Ähnlich dem Vorgehen bei einem Widerspruchsbeweis versucht man die Alternativhypothese zu untermauern, indem man zeigt, daß das Auftreten der gezogenen Stichprobe sehr unwahrscheinlich ist, wenn die Nullhypothese wahr wäre. Wir beschreiben die Vorgangsweise mit einem Beispiel:

BEISPIEL 3.22. Ein Kandidat K bei einer Wahl behauptet, daß sein Stimmenanteil p über 50% liegen würde. Wir zweifeln jedoch an dieser Aussage, und möchten die Annahme stützen, daß der Stimmenanteil von K tatsächlich unter 50% liegt.

Wir formulieren daher die Nullhypothese

$$H_0: p = 50\%$$

und die Alternativhypothese

$$H_a: p < 50\%.$$

Um zwischen den beiden Hypothesen entscheiden zu können, befragen wir $n = 15$ zufällig ausgewählte Wahlberechtigte. Es sei T die Anzahl der Befragten, welche K unterstützen. Wir vereinbaren, die Nullhypothese zu verwerfen, wenn $T \leq 2$ ausfällt. Jeder statistische Test besteht aus diesen Komponenten

- Nullhypothese H_0
- Alternativhypothese H_a
- Teststatistik T
- Ablehnbereich \mathcal{A}

Im Beispiel ist der Ablehnbereich durch $\mathcal{A} = \{0, 1, 2\}$ gegeben. Finden wir beispielsweise $T = 1$, werden wir gemäß unserer Vereinbarung die Nullhypothese verwerfen, und die Alternativhypothese annehmen. Natürlich nehmen wir dabei das Risiko einer Fehlentscheidung in Kauf: Es ist ja nicht ausgeschlossen, daß in einer Stichprobe von 15 Wählern nur 2 den Kandidaten K unterstützen. Allerdings ist die Wahrscheinlichkeit für das Auftreten einer derartigen Stichprobe sehr klein. Allgemein sind folgende Fehlentscheidungen möglich:

DEFINITION 3.5. Ein **Fehler 1.Art** liegt vor, wenn man die Nullhypothese ablehnt, obwohl sie zutrifft. Die Wahrscheinlichkeit, einen solchen Fehler zu begehen, wird mit α bezeichnet und heißt das **Signifikanzniveau** des Tests.

Akzeptiert man hingegen die Nullhypothese, obwohl sie falsch ist, wenn also die Alternativhypothese zutrifft, dann spricht man von einem **Fehler 2.Art**. Die Wahrscheinlichkeit eines fehlers 2. Art wird mit β bezeichnet.

BEISPIEL 3.23 (Fortsetzung von Beispiel 3.22). Im Beispiel der Wahlchancen des Kandidaten K bedeutet ein Fehler 1. Art, daß wir fälschlicherweise schließen, daß K die Wahl verliert, obwohl er sie gewinnen wird. Ein Fehler 2. Art liegt vor, wenn wir einen Wahlsieg von K prognostizieren, obwohl K tatsächlich verlieren wird. Wie groß sind diese beiden Risiken? Da die Teststatistik T unter der Annahme der Gültigkeit von H_0 einer Binomialverteilung mit $n = 15$ und $p = 0.5$ folgt, berechnet man den Fehler 1.Art aus

$$\begin{aligned}\alpha &= P(\text{Verwerfen von } H_0, \text{ wenn } H_0 \text{ zutrifft}) = P(T \in \mathcal{A} | H_0) \\ &= P(T \leq 2 | p = 0.5) = \sum_{t=0}^2 \binom{15}{t} 0.5^{15} = 0.004.\end{aligned}$$

(der Wert von α wurde einer Tabelle der Binomialverteilung entnommen). Die Berechnung der Wahrscheinlichkeit eines Fehlers 2.Art ist ungleich schwieriger. Man kann eine Vorstellung von diesem Risiko bekommen, indem man aus der Menge der Alternativhypothesen eine besonders relevante herausgreift und für diese den Fehler 2. Art berechnet. Sind wir beispielsweise der Meinung, daß der Stimmenanteil von K eher bei $p = 0.3$ liegt, findet man

$$\begin{aligned}\beta &= P(\text{Annahme von } H_0, \text{ wenn } H_a: p = 0.3 \text{ zutrifft}) = P(T \notin \mathcal{A} | p = 0.3) \\ &= P(T > 2 | p = 0.3) = 1 - P(T \leq 2 | p = 0.3) = 0.873.\end{aligned}$$

Der Test wird uns also nahezu immer (mit 87% Wahrscheinlichkeit) dazu führen, einen Wahlsieg von K vorherzusagen, selbst dann, wenn dessen wahrer Stimmenanteil nur bei 30% liegt. Selbst wenn der Stimmenanteil von K nur bei 10% liegt, liegt die Wahrscheinlichkeit eines Fehlers 2. Art noch immer bei 18%!

Eine Möglichkeit, das Risiko eines Fehlers 2.Art zu reduzieren, besteht in der Vergrößerung des Ablehnbereiches für H_0 . Dies erhöht allerdings das Risiko für einen Fehler 1.Art.

BEISPIEL 3.24 (Fortsetzung von Beispiel 3.22). Nimmt man als Ablehnbereich $\mathcal{A} = \{T \leq 4\}$ ist das neue Signifikanzniveau des Testes $\alpha = 0.15$, die Wahrscheinlichkeit des Fehlers 2.Art, falls $p = 0.3$, ergibt sich nun zu $\beta = 0.28$. Die Wahrscheinlichkeiten der beiden Fehler sind nun zwar besser ausgewogen, beide sind aber noch immer inakzeptabel hoch. Die einzige Möglichkeit, *beide* Risiken zu vermindern, besteht in der Erhöhung des Umfanges der Stichprobe.

Solange keine zuverlässigen Schätzungen des Risikos für einen Fehler 2. Art vorliegen, sollte man im Falle $T \notin \mathcal{A}$ die Nullhypothese nicht annehmen, sondern nur schließen, daß die Nullhypothese auf Grund der Stichprobe *nicht verworfen werden kann*.

Wir beenden diesen einführenden Abschnitt mit einigen weiteren Bemerkungen: Nullhypothese und Alternativhypothese gehen asymmetrisch in das Neyman–Pearson Paradigma ein. Die Festlegung der Null- bzw. der Alternativhypothese hängt von der konkreten Anwendung und dem Zweck des Testes ab. Richtlinien für das Festlegen der Nullhypothese sind u.a.:

- In vielen Fällen sind die Konsequenzen einer fälschlichen Verwerfung einer der beiden Hypothesen unterschiedlich schwerwiegend. Als Nullhypothese nimmt man jene, deren irrtümliche Verwerfung die gravierenderen Konsequenzen besitzt, da der Fehler 1. Art durch das Signifikanzniveau des Testes kontrolliert werden kann.
- Als Nullhypothese nimmt man oft jene Hypothese, deren Glaubwürdigkeit durch die Stichprobe erschüttert werden soll.
- In manchen Fällen ist es zweckmäßig, die mathematisch einfachere Hypothese als Nullhypothese anzusetzen.

Der Leser hat sich vielleicht bereits gewundert, warum in Beispiel 3.22 als Nullhypothese nicht $H_0^*: p \geq 0.5$, das logische Gegenteil von H_a , formuliert wurde. Dafür sollen 2 Gründe angeführt werden. Zum einen steht nicht die Nullhypothese im Brennpunkt, sondern wir möchten ja die Alternativhypothese statistisch untermauern. Andererseits stellt sich heraus, daß die Nullhypothese H_0^* zu den gleichen Schlüssen führt wie H_0 .

12. Der Z-Test für große Stichproben

Angenommen wir testen eine Hypothese über einen Populationsparameter θ mittels einer unabhängigen Stichprobe X_1, \dots, x_n . Wir entwickeln einen Test, dem ein erwartungstreuer Punktschätzer $\hat{\theta}$ für θ zugrundegelegt wird, der (zumindest für große Stichproben) annähernd normalverteilt ist mit Mittel θ und Standardfehler $\sigma_{\hat{\theta}}$. Ferner sei $\theta = \theta_0$ ein Parameterwert, der gegen $\theta > \theta_0$ getestet werden soll. Die einfachere der beiden Hypothesen ist $\theta = \theta_0$, welche wir daher als Nullhypothese wählen. Wenn der Schätzwert $\hat{\theta}(x_1, \dots, x_n)$ in der Nähe von θ_0 liegt, haben wir keinen Grund, die Nullhypothese zu verwerfen. Gilt $\theta^* > \theta_0$, für den wahren Wert θ^* von θ , dann werden auch die Schätzwerte mit hoher Wahrscheinlichkeit größer als θ_0 ausfallen. Abbildung 3.6 zeigt die Stichprobenverteilung von $\hat{\theta}$ für $\theta^* \approx \theta_0$ und $\theta^* > \theta_0$.

Dies legt $\mathcal{A} = \{\theta > k\}$ als Ablehnbereich nahe. Die Charakteristika des Testes sind demnach

- $H_0: \theta = \theta_0$
- $H_a: \theta > \theta_0$
- Teststatistik: $\hat{\theta}$

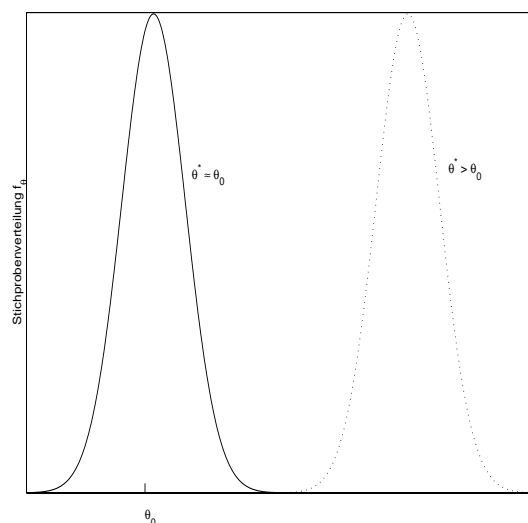


ABB. 3.6. Stichprobenverteilung $F_{\hat{\theta}}$

- Ablehnbereich: $\mathcal{A} = \{\hat{\theta} > k\}$

Der tatsächliche Wert von k wird durch das Risiko eines Fehlers 1. Art, welches wir einzugehen bereit sind, gesteuert. Es sei also α gegeben. Unter der Annahme der Gültigkeit der Nullhypothese ist $\hat{\theta}$ normalverteilt mit Mittel θ_0 und Standardabweichung $\sigma_{\hat{\theta}}$. Bei einem Test mit dem Signifikanzniveau α lehnen wir somit die Nullhypothese ab, wenn

$$P(\hat{\theta} \geq \theta_0 + \delta) = P\left(\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \geq \frac{\delta}{\sigma_{\hat{\theta}}}\right) = \alpha$$

gilt. Da $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$ standard normalverteilt ist, folgt

$$\begin{aligned} \delta &= z_{\alpha} \sigma_{\hat{\theta}} \\ z_{\alpha} &= \Phi^{-1}(1 - \alpha) \end{aligned}$$

Abbildung 3.7 veranschaulicht die Vorgangsweise.

Die vorige Überlegung zeigt, daß der Test äquivalent ist zu

- $H_0: \theta = \theta_0$
- $H_a: \theta > \theta_0$
- Teststatistik: $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$
- Ablehnbereich: $\mathcal{A} = \{z > z_{\alpha}\}$

Natürlich kann man die Nullhypothese $H_0: \theta = \theta_0$ auch gegen $H_a: \theta < \theta_0$ testen. In beiden Fällen ergibt sich ein **einseitiger Test**. Will man hingegen nur irgendeine

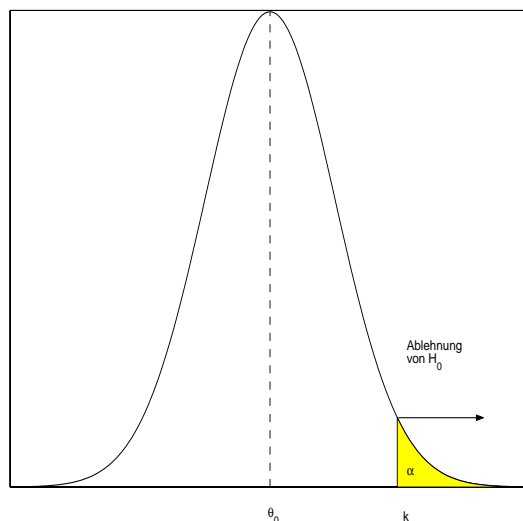


ABB. 3.7. Ablehnbereich

Abweichungen des Parameters von θ_0 aufdecken, dann verwendet man einen **zweiseitigen Test**, bei welchem $H_0: \theta = \theta_0$ gegen $H_a: \theta \neq \theta_0$ getestet wird. Wir stellen in der folgenden Tabelle die verschiedenen Versionen des Z-Testes zusammen:

TABELLE 3.2. Z-Test für große Stichproben

$$\begin{aligned}
 &H_0: \theta = \theta_0 \\
 &H_a: \begin{cases} \theta > \theta_0 \\ \theta < \theta_0 \\ \theta \neq \theta_0 \end{cases} \\
 \text{Teststatistik: } &Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \\
 \text{Ablehnbereich: } &\mathcal{A} = \begin{cases} \{z > z_\alpha\} \\ \{z < -z_\alpha\} \\ \{|z| > z_{\alpha/2}\} \end{cases}
 \end{aligned}$$

Wir weisen auch darauf hin, daß durch die Nullhypothese im Z-Test die jeweilige Verteilung eindeutig festgelegt ist. Hypothesen mit dieser Eigenschaft nennt man **einfach**. Hypothesen, welche nicht einfach sind, heißen **zusammengesetzt**. Die Alternativhypothesen im Z-Test sind Beispiele für zusammengesetzte Hypothesen.

BEISPIEL 3.25. Ein Personalchef einer großen Gesellschaft ist mit den Mitarbeitern des Außendienstes unzufrieden, da sie pro Woche nicht mehr als 15 Abschlüsse im Mittel tätigen. Um diese Behauptung zu testen (und nach Möglichkeit zu widerlegen),

wurden (vom Betriebsrat) 36 Vertreter ausgewählt und die Anzahl der Abschlüsse in einer zufällig ausgewählten Woche registriert. Dies ergab einen Mittelwert von 17 Abschlüssen mit einer Varianz von 9. Kann die Behauptung des Personalchefs widerlegt werden? Man verwende einen Test mit Signifikanzniveau $\alpha = 0.05$.

Da der Betriebsrat die Glaubwürdigkeit der Behauptung des Personalchefs erschüttern will und statistische Evidenz für die Behauptung sucht, für die mittlere Anzahl μ der Abschlüsse gelte doch $\mu > 15$, formulieren wir als Null-, und Alternativhypothese

$$H_0: \mu = 15, \quad H_a: \mu > 15.$$

Als Punktschätzer für μ kann man das Stichprobenmittel verwenden. Dies führt auf die standard normalverteilte Teststatistik

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - 15}{\sigma/\sqrt{n}}$$

Der Ablehnbereich ist gegeben durch $\mathcal{A} = \{z > z_{0.05} = 1.645\}$. Die unbekannte Populationsvarianz σ^2 kann wegen des großen Stichprobenumfanges durch die Stichprobenvarianz $s^2 = 9$ angenähert werden. Der beobachtete Wert der Teststatistik ist daher angenähert gleich

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{17 - 15}{3/\sqrt{36}} = 4.$$

Wegen $z \in \mathcal{A}$ wird die Nullhypothese $H_0: \mu = 15$ verworfen. Das Risiko dabei eine Fehlentscheidung zu treffen beträgt nur 5%.

BEISPIEL 3.26. In einer Studie sollen die Reaktionszeiten von Frauen und Männern auf einen bestimmten Reiz verglichen werden. Ein Experiment mit jeweils zufällig ausgewählten 50 Frauen und Männern ergab

$$\begin{aligned} n_1 &= 50 & n_2 &= 50 \\ \bar{x}_1 &= 3.6 \text{ sec} & \bar{x}_2 &= 3.8 \text{ sec} \\ s_1^2 &= 0.18 & s_2^2 &= 0.14 \end{aligned}$$

Kann man auf Grund der Daten auf einen Unterschied der wahren mittleren Reaktionszeiten von Frauen und Männern schließen (Signifikanzniveau $\alpha = 0.05$.)

Wenn wir die Hypothese, daß die wahren mittleren Reaktionszeiten μ_i , $i = 1, 2$, verschieden sind, festigen wollen, testen wir die Nullhypothese

$$H_0: \mu_1 - \mu_2 = 0$$

gegen die Alternativhypothese

$$H_a: \mu_1 - \mu_2 \neq 0.$$

Der Punktschätzer für $\mu_1 - \mu_2$ ist $\bar{X}_1 - \bar{X}_2$ und erfüllt die Voraussetzungen, um als Grundlage eines Z-Testes dienen zu können. Will man also die Nullhypothese $H_0: \mu_1 - \mu_2 = D_0$ gegen eine Alternativhypothese testen, ist

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

eine geeignete Teststatistik. Im Beispiel ist $D_0 = 0$. Wegen des großen Stichprobenumfanges kann man die Populationsvarianzen σ_i^2 jeweils durch die Stichprobenvarianzen s_i^2 , $i = 1, 2$, approximieren. Der Ablehnbereich bei einer zweiseitigen Alternative ist gegeben durch $\{|z| > z_{0.025} = 1.96\}$. Die Auswertung der Teststatistik ergibt den Wert

$$z = \frac{3.6 - 3.8}{\sqrt{\frac{0.18}{50} + \frac{0.14}{50}}} = -2.5$$

Wegen $z \in \mathcal{A}$ kann man auf dem Signifikanzniveau von $\alpha = 0.05$ die Nullhypothese verwerfen und auf einen Unterschied in den Reaktionszeiten bei Männern und bei Frauen schließen.

12.1. Das Risiko eines Fehlers 2.Art beim Z-Test. Wir skizzieren das Vorgehen für den einseitigen Test

$$H_0: \theta = \theta_0, \quad H_a: \theta > \theta_0.$$

mit dem Ablehnbereich

$$\mathcal{A} = \{\hat{\theta} > k = \theta_0 + z_\alpha \sigma_{\hat{\theta}}\},$$

Da die Alternativhypothese zusammengesetzt ist, können wir die Wahrscheinlichkeit eines Fehlers 2. Art nur für spezielle Parameter aus H_a bestimmen. Angenommen, wir denken an eine spezielle Alternative, $\theta = \theta_a > \theta_0$, dann folgt

$$\begin{aligned} \beta(\theta_a) &= P(T \notin \mathcal{A} | \theta_a) = P(\hat{\theta} \leq k | \theta_a) \\ (3.39) \quad &= P\left(\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} \leq \frac{k - \theta_a}{\sigma_{\hat{\theta}}}\right) = \Phi\left(\frac{k - \theta_a}{\sigma_{\hat{\theta}}}\right), \end{aligned}$$

da ja $\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}}$ standard normalverteilt ist, falls $\theta = \theta_a$ der wahre Wert des Parameters ist.

BEISPIEL 3.27 (Fortsetzung von Beispiel 3.25). Angenommen der Personalchef möchte nun die bereits erhobenen Daten ($n = 36$, $\bar{x} = 17$ und $s^2 = 9$) verwenden um die Nullhypothese $H_0: \mu = 15$ gegen die spezielle Alternativhypothese $H_a: \mu = 16$ zu testen. Man bestimme $\beta(16)$ für diesen Test.

Der Ablehnbereich in Beispiel 3.25 war gegeben durch

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > 1.645 \quad \text{also} \quad \bar{x} > \mu_0 + 1.645 \frac{\sigma}{\sqrt{n}}$$

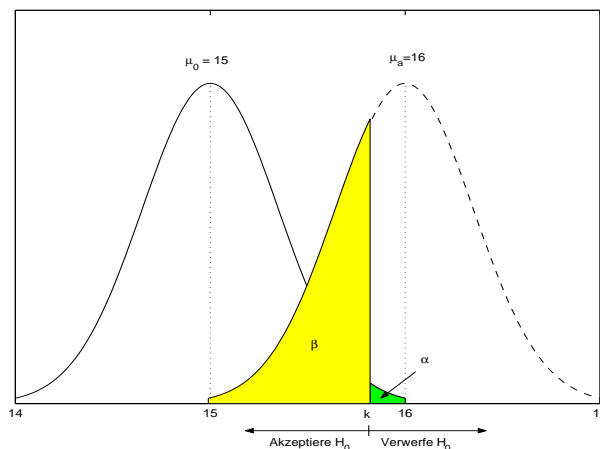


ABB. 3.8

also

$$\bar{x} > 15.823$$

Die Wahrscheinlichkeit eines Fehlers 2. Art folgt aus (3.39)

$$\beta(\mu) = \Phi\left(\frac{15.823 - \mu}{3/\sqrt{36}}\right),$$

und somit

$$\beta(16) = 0.36.$$

Der große Wert von β zeigt, daß mit der Stichprobengröße $n = 36$ der Test nicht die Trennschärfe besitzt, um einen Unterschied von 1 Abschluß in den Mitteln aufzulösen, vgl. Abbildung 3.8

12.2. Die Bestimmung des Stichprobenumfanges. Wir betrachten wieder einen Test von $H_0: \theta = \theta_0$ gegen $H_a: \theta > \theta_0$. Wenn wir die Werte von α und β vorgeben, wobei β für eine spezielle Alternative $\theta_a > \theta_0$ ausgewertet wird, also

$$\alpha = P\left(\frac{\hat{\theta} - \theta_0}{\sigma/\sqrt{n}} > \frac{k - \theta_0}{\sigma/\sqrt{n}} \mid \theta_0\right) = P(Z_0 > z_\alpha)$$

$$\beta = P\left(\frac{\hat{\theta} - \theta_a}{\sigma/\sqrt{n}} \leq \frac{k - \theta_a}{\sigma/\sqrt{n}} \mid \theta_a\right) = P(Z_a \leq -z_\beta)$$

fordern, erhält man die Gleichungen

$$\frac{k - \theta_0}{\sigma/\sqrt{n}} = z_\alpha, \quad \frac{k - \theta_a}{\sigma/\sqrt{n}} = -z_\beta.$$

Daraus folgt der erforderliche Stichprobenumfang

$$(3.40) \quad n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\theta_a - \theta_0)^2}.$$

BEISPIEL 3.28 (Fortsetzung von Beispiel 3.27). Man bestimme den erforderlichen Stichprobenumfang, um die in Beispiel 3.27 benötigte Genauigkeit des Testes mit $\alpha = \beta = 0.05$ zu gewährleisten. Setzt man die Daten in (3.40) ein, erhält man

$$n = \frac{(1.645 + 1.645)^2 \cdot 9}{(16 - 15)^2} = 97.4.$$

13. Der p -Wert eines Testes

Das Signifikanzniveau eines Testes gibt die Wahrscheinlichkeit eines Fehlers 1. Art an. Die Wahl des Signifikanzniveaus unterliegt allerdings keinen objektiven Kriterien. Dies kann zur Folge haben, daß zwei Personen aus demselben Datensatz konträre Schlüsse ziehen. Die eine verwendet einen Test mit dem Signifikanzniveau von $\alpha = 0.05$ und verwirft die Nullhypothese, die andere verwendet denselben Test mit $\alpha = 0.01$ und kommt zum Schluß daß die Daten nicht stark genug, sind die Nullhypothese zu verwerfen. Diese Willkürlichkeit wird durch den Begriff des p -Wertes eliminiert.

DEFINITION 3.6. Der **p -Wert** eines Testes, p^* , ist das kleinste Signifikanzniveau α , für welches die beobachteten Daten das Verwerfen der Nullhypothese rechtfertigen.

Der p -Wert ist die Wahrscheinlichkeit, die konkrete Teststatistik zu beobachten, falls die Gültigkeit der Nullhypothese angenommen wird. Der p -Wert hängt also nur von den Daten ab. Je kleiner der p -Wert ausfällt, desto deutlicher weisen die Daten darauf hin, die Nullhypothese zu verwerfen. Oft teilt man nur den p -Wert mit, und überläßt es dem Benutzer (Auftraggeber), den Test zu interpretieren und ein geeignetes Signifikanzniveau festzulegen.

BEISPIEL 3.29 (Fortsetzung von Beispiel 3.25). In Beispiel 3.25 wurde $H_0: p = 0.5$ gegen $H_a: p < 0.5$ getestet. Als Teststatistik wurde X , die Anzahl der Wähler von K verwendet. Man bestimme den p -Wert, wenn $X = 3$ Personen in einer Stichprobe von 15 Befragten K gewählt haben.

Der größte Ablehnbereich, der mit der Stichprobe verträglich ist, ist gegeben durch $\{X \leq 3\}$. Somit folgt der p -Wert

$$p^* = P(\{X \leq 3 | p = 0.5\}) = 0.018.$$

Jedes Signifikanzniveau $\alpha \geq p^*$ führt zur Ablehnung der Nullhypothese, für $\alpha < p^*$ sind die Daten nicht aussagekräftig genug, um die Nullhypothese abzulehnen.

BEISPIEL 3.30 (Fortsetzung von Beispiel 3.26). Wir bestimmen nun den p -Wert für den zweiseitigen Test $H_0: \mu_1 - \mu_2 = 0$ gegen $H_a: \mu_1 - \mu_2 \neq 0$. Die Teststatistik Z

war standard normalverteilt, beobachtet wurde der Wert $z = -2.5$. Für den p -Wert findet man somit

$$p^* = P(|Z| \geq 2.5) = 0.0124.$$

14. Test von Hypothesen über μ bzw. $\mu_1 - \mu_2$ bei kleinen Stichproben

Bei der Konstruktion des allgemeinen Z -Tests sind wir davon ausgegangen, daß die Stichprobe so groß ist, daß die Testgröße

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

annähernd standard normalverteilt ist. In diesem Abschnitt werden wir von dieser Voraussetzung abgehen und Tests entwickeln, welche für kleine Stichproben geeignet sind. Allerdings ist es nun notwendig, anzunehmen, daß die Stichprobe einer normalverteilten Population entnommen wurde.

14.1. Einfacher T-Test. Es sei X_1, \dots, X_n eine unabhängige Stichprobe aus einer $N(\mu, \sigma)$ verteilten Population. Der Erwartungswert μ und die Varianz σ^2 seien nicht bekannt, sie können jedoch durch das Stichprobenmittel \bar{X} und die Stichprobenvarianz S^2 geschätzt werden. Es wurde in Abschnitt 3 gezeigt, daß die Testgröße

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

einer t -Verteilung mit $n - 1$ Freiheitsgraden folgt, wenn die Nullhypothese $H_0: \mu = \mu_0$ zutrifft. (Es sei darauf hingewiesen, daß die Nullhypothese nicht einfach ist, da sie ja die Varianz der Population offen läßt). Der Ablehnbereich hängt von der Alternativhypothese ab, und wird wie beim Z -Test festgelegt. Dies ergibt folgende Varianten des einfachen T-Tests:

Einfacher T-Test für μ (kleine Stichproben)

$$H_0: \theta = \mu_0$$

$$H_a: \begin{cases} \mu > \mu_0 \\ \mu < \mu_0 \\ \mu \neq \mu_0 \end{cases}$$

$$\text{Teststatistik: } T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}, \quad t\text{-verteilt, } n - 1 \text{ Freiheitsgrade}$$

$$\text{Ablehnbereich: } \mathcal{A} = \begin{cases} \{T > t_\alpha\} \\ \{T < -t_\alpha\} \\ \{|T| > t_{\alpha/2}\} \end{cases}$$

t_α ist bestimmt durch die Forderung $P(T > t_\alpha) = \alpha$.

BEISPIEL 3.31 (Fortsetzung von Beispiel 3.6). In Beispiel 3.6 wurde ein Konfidenzintervall für die Mündungsgeschwindigkeit von Projektilen aus Messungen für 8 Geschosse berechnet. Es ergab sich eine mittlere Mündungsgeschwindigkeit $\bar{x} = 2959$ *ft/sec* und eine Standardabweichung $s = 39.1$ *ft/sec*. Der Erzeuger des Schießpulvers behauptet, daß für das neue Pulver die mittlere Mündungsgeschwindigkeit nicht kleiner sei als 3000 *ft/sec*. Kann man mit Hilfe dieser Stichprobe die Behauptung des Erzeugers mit einem Signifikanzniveau von 0,025 widerlegen?

Unter der Voraussetzung, daß die gemessenen Geschwindigkeiten normal verteilt sind, setzen wir einen T-Test von $H_0: \mu = 3000$ gegen $H_a: \mu < 3000$ an. Die Teststatistik besitzt $n = 7$ Freiheitsgrade. Dies führt auf den Ablehnbereich $\mathcal{A} = \{t < -t_{0,025} = -2.365\}$. Die Auswertung der Teststatistik ergibt

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2959 - 3000}{39.1/\sqrt{8}} = -2.966.$$

Wegen $t \in \mathcal{A}$ kann auf dem Signifikanzniveau 0,025 die Nullhypothese verwerfen, und schließen, daß die mittlere Mündungsgeschwindigkeit tatsächlich kleiner als 3000 *ft/sec* ist.

14.2. Doppelter T-Test für $\mu_1 - \mu_2$. Eine weitere Anwendung der t -Verteilung ergibt sich beim Vergleich von der Populationsmittel von zwei *normalverteilten* Populationen, deren Varianzen gleich, aber nicht notwendig bekannt sein müssen. Es seien also X_{i1}, \dots, X_{in_i} , $i = 1, 2$, unabhängige Stichproben aus den beiden Populationen mit dem Populationsmittel μ_i und der Populationsvarianz σ^2 . Ferner seien \bar{X}_i und S_i^2 die jeweiligen Stichprobenmittel, bzw. die Stichprobenvarianzen. In Abschnitt 3 wurde gezeigt, daß die Prüfgröße

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

t -verteilt ist mit $n_1 + n_2 - 2$ Freiheitsgraden, wobei S_p^2 ein gewichtetes Mittel der Stichprobenvarianzen darstellt,

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Wenn man die Nullhypothese $H_0: \mu_1 - \mu_2 = D_0$ für einen festen Wert von D_0 gegen eine Alternative testet, dann besitzt

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

eine t -Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden, falls die Nullhypothese zutrifft. Wie beim Z -Test erhält man daher folgende Varianten des doppelten T-Testes:

Doppelter T-Test für $\mu_1 - \mu_2$ (kleine Stichproben)

$$H_0: \mu_1 - \mu_2 = D_0$$

$$H_a: \begin{cases} \mu_1 - \mu_2 > D_0 \\ \mu_1 - \mu_2 < D_0 \\ \mu_1 - \mu_2 \neq D_0 \end{cases}$$

$$\text{Teststatistik: } T = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad t\text{-verteilt, } n_1 + n_2 - 2 \text{ Freiheitsgrade}$$

$$\text{Ablehnbereich: } \mathcal{A} = \begin{cases} \{T > t_\alpha\} \\ \{T < -t_\alpha\} \\ \{|T| > t_{\alpha/2}\} \end{cases}$$

t_α ist bestimmt durch die Forderung $P(T > t_\alpha) = \alpha$.

BEISPIEL 3.32. Arbeiter, welche an einem Fließband Bauteile zusammensetzen, werden nach 2 verschiedenen Methoden eingeschult. Um die Effizienz der beiden Methoden zu vergleichen, wurde bei je 9 Arbeitern, die für die Assemblierung benötigte Zeit gemessen. Zusammengefaßt ergaben sich folgende Werte

$$\begin{aligned} n_1 &= 9 & n_2 &= 9 \\ \bar{x}_1 &= 35.22 \text{ sec} & \bar{x}_2 &= 31.56 \text{ sec} \\ \sum_{i=1}^9 (x_{1i} - \bar{x}_1)^2 &= 195.56 & \sum_{i=1}^9 (x_{2i} - \bar{x}_2)^2 &= 160.22 \end{aligned}$$

Kann man auf dem Signifikanzniveau 0.05 behaupten, daß die beiden Ausbildungsmethoden zu unterschiedlichen mittleren Assemblierungszeiten führen?

Da wir $H_0: \mu_1 - \mu_2 = 0$ gegen $H_a: \mu_1 - \mu_2 \neq 0$ testen, ist ein zweiseitiger Test zu verwenden. Der Ablehnbereich ist also $\{|T| > t_{\alpha/2}\}$, mit $t_{\alpha/2} = t_{0.025} = 2.120$ (für 16 Freiheitsgrade) festgelegt. Als Schätzung der gemeinsamen Populationsvarianz erhält man $s_p^2 = 22.24$, also $s_p = 4.72$. Die Auswertung der Teststatistik ergibt

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_1}}} = \frac{35.22 - 31.56}{4.72 \sqrt{\frac{1}{9} + \frac{1}{9}}} = 1.65$$

einen Wert, der nicht in den Ablehnbereich fällt. Die Daten sind daher nicht stark genug, um auf dem Signifikanzniveau von 0.05 die Nullhypothese zu verwerfen. Wie wir bereits ausgeführt haben, führt dieses Testergebnis nicht automatisch zur Annahme von H_0 .

15. Testen von Hypothesen über Varianzen

15.1. χ^2 -Streuungstest. Es sei X_1, \dots, X_n eine unabhängige Stichprobe aus einer $N(\mu, \sigma)$ verteilten Grundgesamtheit. Es soll die Nullhypothese $H_0: \sigma^2 = \sigma_0^2$

gegen verschiedene Alternativen getestet werden. Wenn die Nullhypothese gilt, dann besitzt

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

eine χ^2 -Verteilung mit $n-1$ Freiheitsgraden. Wie soll beispielsweise der Ablehnbereich definiert werden, wenn als Alternativhypothese $H_a: \sigma^2 > \sigma_0^2$ gewählt wird? Wenn die Alternative zutrifft, dann erwarten wir eher einen Wert für die Stichprobenvarianz S^2 mit $S^2 > \sigma_0^2$, d.h. die Teststatistik χ^2 nimmt große Werte an. Das Signifikanzniveau α kann daher sichergestellt werden, wenn man

$$\mathcal{A} = \{\chi^2 > \chi_\alpha^2\}$$

festlegt, wobei χ_α^2 durch

$$P(\chi^2 > \chi_\alpha^2) = \alpha$$

festgelegt wird. Analog geht man bei den anderen Alternativhypothesen vor. Zusammenfassend erhält man folgende Möglichkeiten:

χ^2 - Streuungstest

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_a: \begin{cases} \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \\ \sigma^2 \neq \sigma_0^2 \end{cases}$$

$$\text{Teststatistik: } T = \frac{(n-1)S^2}{\sigma_0^2}, \quad \chi^2\text{-verteilt, } n-1 \text{ Freiheitsgrade}$$

$$\text{Ablehnbereich: } \mathcal{A} = \begin{cases} \{T > \chi_\alpha^2\} \\ \{T < \chi_{1-\alpha}^2\} \\ \{T > \chi_{\alpha/2}^2\} \cup \{T < \chi_{1-\alpha/2}^2\} \end{cases}$$

χ_α^2 ist bestimmt durch die Forderung $P(\chi^2 > \chi_\alpha^2) = \alpha$.

BEISPIEL 3.33. Eine Firma erzeugt Präzisionsteile, deren Durchmesser eine Varianz von höchstens 0.0002 mm aufweisen dürfen. Eine Stichprobe von 10 Teilen ergab eine Stichprobenvarianz von $s^2 = 0.0003$. Man teste auf dem Signifikanzniveau 0.05 die Nullhypothese $H_0: \sigma^2 = 0.0002$ gegen die Alternativhypothese $H_a: \sigma^2 > 0.0002$.

Wenn die Durchmesser der Teile normalverteilt sind, besitzt die Testgröße eine χ^2 -Verteilung mit 9 Freiheitsgraden. Wir lehnen H_0 daher ab, wenn $T > \chi_{0.05}^2 = 16,919$ ausfällt. Der beobachtete Wert der Teststatistik beträgt

$$t = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \cdot 0.0003}{0.0002} = 13.5.$$

Es gibt daher nicht genug Evidenz, um die Nullhypothese auf dem Niveau 0.05 abzulehnen.

15.2. Der F-Test für $\sigma_1^2 - \sigma_2^2$. Wir wenden uns dem Vergleich der Varianzen von zwei Populationen zu. Wir nehmen an, daß beide Populationen *normalverteilt* mit Mittelwert μ_i und Varianz σ_i^2 , $i = 1, 2$, sind. Wir betrachten den Test $H_0: \sigma_1^2 = \sigma_2^2$ gegen $H_a: \sigma_1^2 > \sigma_2^2$ und entnehmen aus den Populationen unabhängige Stichproben X_{i1}, \dots, X_{in_i} . Da die Stichprobenvarianzen S_i^2 Schätzwerte für die Populationsvarianzen sind, liegt es nahe, H_0 zu verwerfen, wenn $S_1^2 \gg S_2^2$ gilt. Dies legt den Ablehnbereich

$$\mathcal{A} = \left\{ \frac{S_1^2}{S_2^2} > k \right\}$$

nahe, wobei k so bestimmt wird, daß die Wahrscheinlichkeit eines Fehlers 1. Art durch α begrenzt ist. Beachtet man, daß $\frac{(n_i-1)S_i^2}{\sigma_i^2}$ unabhängige, χ^2 -verteilte Zufallsvariablen mit $n_i - 1$ Freiheitsgraden sind, folgt aus der Definition 2.29, daß die Testgröße

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

eine F -Verteilung mit $n_1 - 1$ Zähler- und $n_2 - 1$ Nennerfreiheitsgraden besitzt. Unter der Voraussetzung der Gültigkeit der Nullhypothese gilt aber

$$F = \frac{S_1^2}{S_2^2},$$

sodaß der Ablehnbereich geschrieben werden kann in der Form

$$\mathcal{A} = \{F > f_\alpha\}$$

geschrieben werden kann. Da die Indizierung der beiden Populationen vollkommen beliebig ist, kann man vereinbaren, jene Population, deren Varianz in H_a größer angenommen wird, mit dem Index 1 zu versehen. Wenn nur ein Unterschied in den Varianzen aufgedeckt werden soll, wenn also $H_0: \sigma_1^2 = \sigma_2^2$ gegen $H_a: \sigma_1^2 \neq \sigma_2^2$ getestet werden soll, kann man analog vorgehen: die größere Stichprobenvarianz wird in den Zähler der Teststatistik F geschrieben, als Ablehnbereich ist $\mathcal{A} = \{F > f_{\alpha/2}\}$ zu nehmen.

F – Test

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \begin{cases} \sigma_1^2 > \sigma_2^2 \\ \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Teststatistik: $F = \frac{S_1^2}{S_2^2}$, F -verteilt, $n_1 - 1$ Zähler, $n_2 - 1$ Nennerfreiheitsgrade

$$\text{Ablehnbereich: } \mathcal{A} = \begin{cases} \{F > f_\alpha\} \\ \{F > f_{\alpha/2}\} \end{cases}$$

f_α ist bestimmt durch die Forderung $P(F > f_\alpha) = \alpha$.

BEISPIEL 3.34. Ein Experiment zum Vergleich der Schmerzgrenzen gegenüber Elektroschocks bei Männern und Frauen ergab folgendes Resultat:

	Männer	Frauen
n	14	10
\bar{x}	16.2	14.9
s^2	12.	26.4

Kann man auf Grund der Daten auf einem Signifikanzniveau von 0.10 auf einen signifikanten Unterschied in der Variabilität der Schmerzgrenzen bei Männern und Frauen schließen?

Wir gehen wieder von der Annahme aus, daß die Schmerzgrenzen bei beiden Geschlechtern ungefähr normal verteilt sind und testen $H_0: \sigma_M^2 = \sigma_F^2$ gegen $H_a: \sigma_M^2 \neq \sigma_F^2$. Die größere Stichprobenvarianz ist jene der Frauen. Wir setzen die Teststatistik daher an als $F = \frac{S_F^2}{S_M^2}$, welche F -verteilt ist und 9 Zähler- und 13 Nennerfreiheitsgrade besitzt. Der Ablehnbereich ist durch $\mathcal{A} = \{F > f_{\alpha/2} = 2.71\}$ gegeben. Die Auswertung der Teststatistik ergibt $f = \frac{26.4}{12.7} = 2.079$. Wegen $f \notin \mathcal{A}$ kann daher die Nullhypothese nicht verworfen werden.

16. Dualität zwischen Konfidenzintervall und Hypothesentest

In Abschnitt 2 wurde ein $(1 - \alpha)$ – Konfidenzintervall für einen Populationsparameter θ konstruiert, welches für große Stichproben anwendbar ist:

$$\text{conf}(\theta) = [\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}].$$

Dem zweiseitigen Konfidenzintervall entspricht ein zweiseitiger Z -Test $H_0: \theta = \theta_0$ versus $H_a: \theta \neq \theta_0$ zum Signifikanzniveau α , der die Nullhypothese verwirft, wenn die Testgröße

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

in den Ablehnbereich

$$\mathcal{A} = \{|z| > z_{\alpha/2}\}$$

fällt. In diesem Zusammenhang nennt man das Komplement des Ablehnbereiches $\bar{\mathcal{A}}$ **Akzeptanzbereich**. Drückt man den Akzeptanzbereich wieder durch θ aus, erhält man

$$\bar{\mathcal{A}} = \{z \leq z_{\alpha/2}\} = [\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}] = \text{conf}(\theta).$$

Die Nullhypothese $H_0: \theta = \theta_0$ wird also genau dann nicht verworfen, wenn $\theta_0 \in \text{conf}(\theta)$ liegt. Das $(1 - \alpha)$ -Konfidenzintervall für θ enthält also genau jene Parameter, für welche auf dem Signifikanzniveau α die Nullhypothese nicht verworfen wird. Da jeder Wert von θ_0 im $(1 - \alpha)$ -Konfidenzintervall auf dem Signifikanzniveau α akzeptabel ist, ist es nicht sinnvoll, ohne weiter Zusatzinformation, etwa über das Risiko eines Fehlers 2.Art, die Nullhypothese, daß θ gerade den Wert θ_0 annimmt, zu akzeptieren, wenn die Testgröße in den Akzeptanzbereich fällt.

Wir zeigen nun, daß diese Dualität allgemeiner gilt. Dazu betrachten wir eine Familie von Wahrscheinlichkeitsverteilungen, welche nur von einem Parameter $\theta \in \Pi$ abhängt. Ferner fassen wir die Stichprobe in einem Zufallsvektor \vec{X} zusammen. Wir testen die Nullhypothese $H_0: \theta = \theta_0$ gegen eine Alternative und verwenden dafür die Testgröße $T = T(\vec{X})$ mit dem Akzeptanzbereich $\bar{\mathcal{A}}(\theta_0) (= [\theta_0 - z_{\alpha/2}\sigma_{\hat{\theta}}, \theta_0 + z_{\alpha/2}\sigma_{\hat{\theta}}])$ im vorigen Beispiel). Dann gilt

PROPOSITION 3.14. *Für jeden Wert von $\theta \in \Pi$ gebe es einen Test für $H_0: \theta = \theta_0$ gegen eine Alternative zum Signifikanzniveau α . Dann gilt mit den obigen Bezeichnungen*

1. Die Menge

$$C(\vec{X}) = \{\theta : T(\vec{X}) \in \bar{\mathcal{A}}(\theta)\}$$

ist eine $(1 - \alpha)$ -Konfidenzmenge für θ .

2. Ist umgekehrt $C(\vec{X})$ eine $(1 - \alpha)$ -Konfidenzmenge für θ , dann ist

$$A(\theta_0) = \{T(\vec{X}) : \theta_0 \in C(\vec{X})\}$$

ein Akzeptanzbereich für einen Test der Nullhypothese $H_0: \theta = \theta_0$ zum Niveau α .

Die erste Behauptung stellt fest, daß ein $(1 - \alpha)$ -Konfidenzbereich für θ aus jenen Parametern θ_0 besteht, für welche die Nullhypothese $\theta = \theta_0$ zum Niveau α nicht verworfen werden kann. Die zweite Behauptung sagt aus, daß die Nullhypothese nicht verworfen wird, wenn θ_0 im Konfidenzbereich liegt.

BEWEIS. Da $\bar{\mathcal{A}}(\theta)$ der Akzeptanzbereich eines Tests zum Niveau α ist, gilt offensichtlich

$$P(T(\vec{X}) \in \bar{\mathcal{A}}(\theta_0) | \theta = \theta_0) = 1 - \alpha$$

und daher auch wegen der Definition von $C(\vec{X})$

$$P(\theta_0 \in C(\vec{X})|\theta = \theta_0) = P(T(\vec{X}) \in \bar{\mathcal{A}}(\theta)|\theta = \theta_0) = 1 - \alpha.$$

Umgekehrt sei nun $C(\vec{X})$ ein $(1 - \alpha)$ -Konfidenzbereich, d.h. es gilt

$$P(\theta_0 \in C(\vec{X})|\theta = \theta_0) = 1 - \alpha.$$

Für einen Test zum Niveau α folgt dann

$$P(T(\vec{X}) \in \mathcal{A}(\theta_0)|\theta = \theta_0) = P(\theta_0 \in C(\vec{X})|\theta = \theta_0) = 1 - \alpha.$$

□

Diese Dualität kann nützlich sein, wenn zum Beispiel der Akzeptanzbereich für einen Test auf direkte Weise schwer zugänglich ist, aber ein Konfidenzintervall berechnet werden kann.

17. Die Macht eines Testes und das Neyman–Pearson Lemma

Die Güte eines Testes kann durch das Signifikanzniveau α und die Wahrscheinlichkeit eines Fehlers 2. Art β beurteilt werden. Ein verwandtes Konzept ist die Macht eines Testes

DEFINITION 3.7. *Es sei \mathcal{A} der Ablehnbereich eines Testes einer Hypothese über einen Parameter θ und T die dabei verwendete Teststatistik. Die **Macht** des Testes, $G(\tilde{\theta})$, ist die Wahrscheinlichkeit, mit welcher der Test zur Ablehnung der Nullhypothese führt, wenn der wahre Parameter den Wert $\tilde{\theta}$ hat, also*

$$G(\tilde{\theta}) = P(T \in \mathcal{A}|\theta = \tilde{\theta})$$

Die Macht eines Testes ist eng mit der Wahrscheinlichkeit eines Fehlers 2. Art verknüpft. Wenn wir die Nullhypothese $H_0: \theta = \theta_0$ testen, dann gilt offensichtlich $G(\theta_0) = \alpha$. Wählt man einen Parameter θ_a aus der Alternativhypothese, folgt

$$\begin{aligned} G(\theta_a) &= P(\text{verwerfe } H_0|\theta_a) \\ &= 1 - P(\text{akzeptiere } H_0|\theta_a) = 1 - \beta(\theta_a). \end{aligned}$$

Ein idealer Test würde mit Sicherheit jede Abweichung von der Nullhypothese $H_0: \theta = \theta_0$ aufdecken, d.h. es sollte $G(\theta_a) = 1$ für $\theta_a \neq \theta_0$ gelten. Dies ist in der Praxis nicht möglich, da für einen festen Stichprobenumfang α und β nicht gleichzeitig beliebig klein gemacht werden können. In der Praxis geht man daher folgendermaßen vor: man wählt einen festen Wert für α und versucht den Ablehnbereich so zu bestimmen, daß die Wahrscheinlichkeit eines Fehlers 2. Art für jede mögliche Alternative minimal wird, d.h. die Macht des Testes maximal wird. Das Lemma von Neyman–Pearson, für dessen Beweis wir auf die einschlägige Literatur verweisen, gibt eine Lösung in einer einfachen Situation:

PROPOSITION 3.15 (Lemma von Neyman–Pearson). *Wir betrachten einen Test der einfachen Nullhypothese $H_0: \theta = \theta_0$ gegen die einfache Alternativhypothese $H_a: \theta = \theta_a$. Ferner sei $\text{lik}(\theta)$ die Likelihoodfunktion einer Stichprobe X_1, \dots, x_n . Dann hat der **Likelihood Quotienten Test**, welcher die Teststatistik*

$$T = \frac{\text{lik}(\theta_0)}{\text{lik}(\theta_a)}$$

und den Ablehnbereich $T < k$ verwendet, für ein gegebenes Signifikanzniveau α maximale Macht in θ_a . Die Konstante k ist so einzurichten, daß die Wahrscheinlichkeit eines Fehlers 1. Art gerade durch α begrenzt wird.

Diesem Test liegt die Heuristik zugrunde, daß kleine Werte von $\frac{\text{lik}(\theta_0)}{\text{lik}(\theta_a)}$ andeuten, daß die konkrete Stichprobe unter H_0 sehr unwahrscheinlich ist gegenüber dem Auftreten der Stichprobe unter H_1 .

BEISPIEL 3.35. Es sei X eine Einzelbeobachtung einer Zufallsvariablen mit der Verteilungsdichte

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0 & \text{sonst.} \end{cases}$$

Man bestimme den mächtigsten Test zum Signifikanzniveau $\alpha = 0.05$ von $H_0: \theta = 2$ versus $H_a: \theta = 1$.

Da beide Hypothesen einfach sind, kann der mächtigste Test mit Hilfe des Neyman–Pearson Lemmas gefunden werden. Dazu berechnet man zuerst

$$\frac{\text{lik}(\theta_0)}{\text{lik}(\theta_a)} = \frac{f(x|\theta_0)}{f(x|\theta_a)} = 2x, \quad 0 < x < 1.$$

der mächtigste Test verwendet daher einen Ablehnbereich der Form

$$\mathcal{A} = \{2x < k\} \quad \text{bzw.} \quad \mathcal{A} = \{x < k'\},$$

$0 \leq k' \leq 1$. Der aktuelle Wert von k' wird durch das geforderte Signifikanzniveau festgelegt:

$$0.05 = P(x \in \mathcal{A} | \theta = 2) = \int_0^{k'} 2x \, dx = (k')^2.$$

Somit gilt $k' = \sqrt{0.05}$, der optimale Ablehnbereich ist somit durch

$$\mathcal{A} = \{x < \sqrt{0.05}\}$$

festgelegt. Unter allen Tests mit dem Signifikanzniveau 0.05 von H_0 gegen H_a hat dieser Test das kleinste Risiko eines Fehlers 2. Art.

Eine wesentliche Voraussetzung für die Anwendbarkeit des Neyman–Pearson Lemmas ist die Einfachheit von Null- und Alternativhypothese. In manchen Fällen kann

der mächtigste Test auch für $H_0: \theta = \theta_0$ gegen $H_a: \theta > \theta_0$ mit dieser Technik bestimmt werden. Dazu bestimmt man zuerst den mächtigsten Test für $H_0: \theta = \theta_0$ gegen $H_a^*: \theta = \theta_a$ für eine beliebige Wahl von θ_a aus der Alternative. Wenn die Form des Ablehnbereiches unabhängig von θ_a ist, hat man gleichzeitig den sogenannten **gleichmäßig mächtigsten Test** für $H_0: \theta = \theta_0$ gegen $H_a: \theta > \theta_0$ gefunden.

BEISPIEL 3.36. Es sei X_1, \dots, X_n eine unabhängige Stichprobe aus einer normalverteilten Population mit unbekanntem Erwartungswert μ und bekannter Varianz σ^2 . Man bestimme den gleichmäßig mächtigsten Test von $H_0: \mu = \mu_0$ gegen $H_a: \mu > \mu_0$.

Im ersten Schritt bestimmen wir den mächtigsten Test für $H_0: \mu = \mu_0$ gegen $H_a^*: \mu = \mu_a$ für ein beliebiges $\mu_a > \mu_0$. Die Likelihood Funktion der Stichprobe ist gegeben durch

$$\text{lik}(\mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right).$$

Der Ablehnbereich des mächtigsten Testes von H_0 versus H_a^* ist daher bestimmt durch

$$\frac{\text{lik}(\mu_0)}{\text{lik}(\mu_a)} = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \mu_0)^2 - (x_i - \mu_a)^2]\right) < e^{-\frac{1}{2\sigma^2}k}$$

(der Einfachheit halber wurde die Konstante k in der speziellen Form angesetzt). Eine einfache Umformung ergibt die Ungleichung

$$2n\bar{x}(\mu_a - \mu_0) + n(\mu_0^2 - \mu_a^2) > k$$

bzw. ($\mu_a > \mu_0$)

$$\bar{x} > \frac{k + n(\mu_a^2 - \mu_0^2)}{2n(\mu_a - \mu_0)}.$$

Der mächtigste Test von H_0 gegen H_a^* verwendet also die Teststatistik $T = \bar{X}$ und den Ablehnbereich

$$\mathcal{A} = \{\bar{X} > k'\},$$

die Konstante k' wird bestimmt durch das Signifikanzniveau des Testes

$$\alpha = P(\bar{X} \in \mathcal{A} | \mu = \mu_0) = P(\bar{X} > k' | \mu = \mu_0).$$

Der Ablehnbereich ist somit unabhängig von μ_a : jeder Wert von $\mu_a > \mu_0$ führt auf denselben Ablehnbereich. Dieser Ablehnbereich und die Teststatistik T definieren also den gleichmäßig mächtigsten Test von $H_0: \mu = \mu_0$ gegen $H_a: \mu > \mu_0$. Unter der Voraussetzung von H_0 ist die Teststatistik $\bar{X} \sim N(\mu_0, \sigma/\sqrt{n})$ verteilt. Die Bedingung für k' ist somit gleichwertig mit

$$P(\bar{X} > k' | \mu = \mu_0) = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{k' - \mu_0}{\sigma/\sqrt{n}}\right) = \alpha.$$

Setzt man

$$\frac{k' - \mu_0}{\sigma/\sqrt{n}} = z(\alpha)$$

kann der Ablehnbereich geschrieben werden in der Form

$$\mathcal{A} = \{\hat{\mu} > \mu_0 + \frac{\sigma}{\sqrt{n}}z(\alpha)\}.$$

Da $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ standard normalverteilt ist, gilt $z(\alpha) = \Phi^{-1}(1 - \alpha)$. Diese Überlegung zeigt, daß der gleichmäßig mächtigste Test von $H_0: \mu = \mu_0$ gegen $H_a: \mu > \mu_0$ identisch ist mit dem entsprechenden Z-Test aus Abschnitt 12.

18. Der verallgemeinerte Likelihood Quotiententest

Der Likelihood Quotiententest einer einfachen Nullhypothese gegen eine Alternative ist nach dem Neyman-Pearson Lemma optimal. Der verallgemeinerte Likelihood Quotiententest ist in der Lage, zusammengesetzte Null-, und/oder Alternativhypothesen zu verarbeiten. Es kommt beispielsweise häufig vor, daß die Hypothesen die zugrundeliegende Verteilung nicht eindeutig festlegen, da noch weitere Parameter frei bleiben, z.B. $H_0: \mu = \mu_0$ bei einer Normalverteilung, wenn μ und σ^2 unbekannt sind.

Wir fassen im Folgenden alle unbekannt Parameter zu einem Vektor $\theta \in \mathbb{R}^k$ zusammen. Die Nullhypothese spezifiziere $\theta \in \Theta_0$, die Alternativhypothese behaupte $\theta \in \Theta_a$, wobei $\Theta_0 \cap \Theta_a = \emptyset$. Ferner sei $\Theta = \Theta_0 \cup \Theta_a$. Ein plausibles Maß für die relative Gültigkeit der Hypothesen, ist das Verhältnis ihrer Likelihoodfunktionen, in denen man jeden unbekannt Parameter durch den entsprechenden Maximum Likelihoodschätzer ersetzt:

Verallgemeinerter Maximumlikelihood Quotiententest

Es sei $L(\hat{\Theta}_0)$ die Likelihoodfunktion der Stichprobe, in der man alle unbekannt Parameter durch ihre Maximum Likelihoodschätzer unter der Nebenbedingung $\theta \in \Theta_0$ ersetzt. Analog wird $L(\hat{\Theta}_a)$ konstruiert, die MLS werden allerdings der Nebenbedingung $\theta \in \Theta_a$ unterworfen.

Der verallgemeinerte Maximumlikelihood Quotiententest verwendet die Teststatistik

$$T = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\max_{\theta \in \Theta_0} \text{lik}(\theta)}{\max_{\theta \in \Theta} \text{lik}(\theta)}$$

und verwirft die Nullhypothese $H_0: \theta \in \Theta_0$ zugunsten der Alternativhypothese $H_a: \theta \in \Theta_a$ falls $T \leq k$.

Offensichtlich gilt stets $0 \leq T \leq 1$. Kleine Werte von T bedeuten, daß die Wahrscheinlichkeit der Stichprobe unter H_0 klein im Vergleich zur Wahrscheinlichkeit ihres Auftretens, falls H_a gilt. Der tatsächliche Wert von k wird durch das Signifikanzniveau des Testes festgelegt. Der verallgemeinerte Maximumlikelihood Quotiententest ist im allgemeinen nicht optimal.

BEISPIEL 3.37. Es sei X_1, \dots, x_n eine unabhängige Stichprobe aus einer normal verteilten Population mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 . Man konstruiere einen Likelihood Quotiententest von $H_0: \mu = \mu_0$ gegen $H_a: \mu > \mu_0$.

In diesem Beispiel ist $\Theta_0 = \{(\mu_0, \sigma^2): \sigma^2 > 0\}$, $\Theta_a = \{(\mu, \sigma^2): \mu > \mu_0, \sigma^2 > 0\}$ und $\Theta = \{(\mu, \sigma^2): \mu \geq \mu_0, \sigma^2 > 0\}$. Die Likelihoodfunktion der Stichprobe ist gegeben durch

$$\text{lik}(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right).$$

Um $L(\hat{\Theta}_0)$ zu bestimmen, benötigen wir den MLS für σ^2 unter der Nebenbedingung $\mu = \mu_0$. Wie in Beispiel 3.12 findet man

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2.$$

Somit ist $L(\hat{\Theta}_0)$ gegeben durch $\text{lik}(\mu_0, \hat{\sigma}_0^2)$, also

$$L(\hat{\Theta}_0) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} e^{-n/2}.$$

Für die Berechnung von $L(\hat{\Theta})$ ist das Maximum von $\text{lik}(\mu, \sigma^2)$ auf $\Theta = [\mu_0, \infty) \times (0, \infty)$ zu bestimmen. Nach Beispiel 3.12 sind die unrestringierten Maximum Likelihood Schätzer gegeben durch

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}_{\bar{X}}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2.$$

Falls $(\hat{\mu}, \hat{\sigma}^2) \notin \Theta$ wird das Maximum auf dem Rand von Θ angenommen. Wegen $\lim_{\sigma^2 \downarrow 0} \text{lik}(\mu, \sigma^2) = 0$, $\mu \in \mathbb{R}$, wird das Maximum für $\mu = \mu_0$ angenommen. Die MLS für μ und σ^2 sind demnach

$$(\hat{\mu}, \hat{\sigma}^2) = \begin{cases} (\bar{X}, \sigma_{\bar{X}}^2) & \text{falls } \bar{X} > \mu_0 \\ (\mu_0, \sigma_0^2) & \text{falls } \bar{X} \leq \mu_0, \end{cases}$$

und somit

$$L(\hat{\Theta}) = \begin{cases} \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_{\bar{X}}^2}\right)^{n/2} e^{-n/2} & \text{falls } \bar{X} > \mu_0 \\ \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat{\sigma}_0^2}\right)^{n/2} e^{-n/2} & \text{falls } \bar{X} \leq \mu_0 \end{cases}$$

Die Teststatistik ergibt sich daher zu

$$T = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \left(\frac{\hat{\sigma}_{\bar{X}}^2}{\hat{\sigma}_0^2}\right)^{n/2} \\ = \begin{cases} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \mu_0)^2}\right)^{n/2} & \text{falls } \bar{X} > \mu_0 \\ 1 & \bar{X} \leq \mu_0. \end{cases}$$

Wie bereits erwähnt, gilt $0 \leq T \leq 1$. $T = 1$ bedeutet, daß die Stichproben unter beiden Hypothesen gleich wahrscheinlich sind. Es besteht daher kein Grund H_0 zu verwerfen. Für die Schranke k im Ablehnbereich $\mathcal{A} = \{T \leq k\}$ folgt demnach $0 < k < 1$. Wegen

$$\sum_{i=1}^n (X_i - \mu_0)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$$

läßt sich die Bedingung für die Ablehnung der Nullhypothese $T \leq k$ umschreiben in

$$\frac{1}{1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} < k^{2/n},$$

dies ist äquivalent zu

$$\frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} > k^{-2/n} - 1 \equiv k'$$

bzw. zu

$$\frac{n(\bar{X} - \mu_0)^2}{S^2} > (n-1)k'$$

mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Da $T \leq k < 1$ die Ungleichung $\bar{X} > \mu_0$ nach sich zieht, erhält man schließlich

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > \sqrt{(n-1)k'},$$

d.h. der verallgemeinerte Likelihood Quotienten Test ist äquivalent zum einfachen T-Test.

Allerdings führt der Maximum Likelihood Quotiententest nicht immer auf eine Teststatistik, deren Wahrscheinlichkeitsverteilung bekannt ist. Unter bestimmten Regularitätsbedingungen an die Verteilung der Population kann man die asymptotische Verteilung der Teststatistik angeben:

PROPOSITION 3.16. *Wir betrachten den verallgemeinerten Maximum Likelihood Quotiententest für $H_0: \theta \in \Theta_0$ gegen $H_a: \theta \in \Theta_a$. Es sei r_0 die Anzahl der Parameter (= Koordinaten von θ), welche durch $H_0: \theta \in \Theta_0$ festgelegt werden, r sei die Anzahl der durch die Bedingung $\theta \in \Theta = \Theta_0 \cup \Theta_a$ fixierten Parameter, $r_0 > r$. Dann ist $-2 \ln T$ für große Stichproben annähernd χ^2 -verteilt mit $r_0 - r$ Freiheitsgraden.*

BEISPIEL 3.38. Ein Geschäftsmann möchte die Anzahl der wöchentlich aus 2 verschiedenen Filialen seines Unternehmens eintreffenden Beschwerden vergleichen. 100 unabhängige Aufzeichnungen ergaben ein Wochenmittel von $\bar{x}_1 = 20$ für Filiale A, bzw. $\bar{x}_2 = 22$ für Filiale B. Ausgehend von der Annahme, daß die Anzahl der Beschwerden pro Woche in beiden Filialen Poisson verteilt ist mit dem Erwartungswert θ_i , $i = 1, 2$ teste man $H_0: \theta_1 = \theta_2$ gegen $H_a: \theta_1 \neq \theta_2$ auf dem Niveau 0.01.

Die Likelihoodfunktion der beiden Stichproben $X_{i,1}, \dots, X_{i,100}$, $i = 1, 2$, ist durch deren gemeinsame Wahrscheinlichkeitsverteilung gegeben

$$\text{lik}(\theta_1, \theta_2) = \frac{1}{k!} \theta_1^{\sum x_{i1}} e^{-n\theta_1} \theta_2^{\sum x_{i2}} e^{-n\theta_2},$$

$k! = x_{11}! \cdots x_{1n}! \cdot x_{21}! \cdots x_{2n}!$ und $n = 100$. Ferner setzen wir $\Theta_0 = \{(\theta_1, \theta_2): \theta_1 = \theta_2, \theta_i > 0, i = 1, 2\}$, $\Theta_a = \{(\theta_1, \theta_2): \theta_1 \neq \theta_2, \theta_i > 0, i = 1, 2\}$ und $\Theta = \{(\theta_1, \theta_2): \theta_i > 0, i = 1, 2\}$. Wenn die Nullhypothese gilt, hängt die Likelihoodfunktion nur mehr von einem Parameter, dem gemeinsamen Wert θ von θ_i , ab und vereinfacht sich zu

$$\text{lik}(\theta, \theta) = \frac{1}{k!} \theta^{\sum (x_{i1} + x_{i2})} e^{-2n\theta}.$$

Sie nimmt demnach ihr Maximum im MLS für θ an. Eine einfache Rechnung ergibt

$$\hat{\theta} = \frac{1}{2n} \sum (x_{i1} + x_{i2}) = \frac{1}{2} (\bar{X}_1 + \bar{X}_2).$$

Somit folgt

$$L(\Theta_0) = \frac{1}{k!} \hat{\theta}^{n\bar{X}_1 + n\bar{X}_2} e^{-2n\hat{\theta}}.$$

Die Berechnung von $L(\Theta)$ erfordert die MLS für θ_i , welche durch

$$\hat{\theta}_i = \bar{X}_i, \quad i = 1, 2$$

gegeben sind. Dies ergibt

$$L(\Theta) = \frac{1}{k!} \hat{\theta}_1^{n\bar{X}_1} \hat{\theta}_2^{n\bar{X}_2} e^{-n\hat{\theta}_1 - n\hat{\theta}_2}.$$

Somit folgt die Teststatistik

$$T = \frac{L(\Theta_0)}{L(\Theta)} = \frac{\hat{\theta}^{n\bar{X}_1 + n\bar{X}_2}}{\hat{\theta}_1^{n\bar{X}_1} \hat{\theta}_2^{n\bar{X}_2}} = \frac{(\frac{1}{2}(\bar{X}_1 + \bar{X}_2))^{n\bar{X}_1 + n\bar{X}_2}}{(\bar{X}_1)^{n\bar{X}_1} (\bar{X}_2)^{n\bar{X}_2}}.$$

Setzt man die Beobachtungen ein, erhält man

$$T = \left(\frac{1}{2}(20 + 22)\right)^{100(20+22)} 20^{100 \cdot 20} 22^{100 \cdot 22}$$

also

$$-2 \ln T = -2[4200 \ln 21 - 200 \ln 20 - 2200 \ln 22] = 9.53.$$

Die Nullhypothese fixiert einen Freiheitsgrad, $r_0 = 1$, in Θ sind beide Parameter voneinander unabhängig, also $r = 0$. Wegen des großen Umfangs der Stichprobe ist $-2 \ln T$ daher annähernd χ^2 verteilt mit $r_0 - r = 1$ Freiheitsgraden. Kleinen Werten von T entsprechen große Werte von $-2 \ln T$. Da der beobachtete Wert von $-2 \ln T$ größer ist als $\chi_{0.01}^2 = 6.635$ verwerfen wir die Nullhypothese auf dem Niveau 0.01 und schließen, daß die mittlere Anzahl der Beschwerden pro Woche in beiden Filialen tatsächlich verschieden sind.