

# Angewandte Statistik

Dr. Roger Filiger

2013

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>5</b>
<b>I</b>	<b>Beschreibende Statistik</b>	<b>8</b>
<b>2</b>	<b>Statistische Variablen</b>	<b>9</b>
<b>3</b>	<b>Graphische Darstellungen</b>	<b>11</b>
3.1	Einführung . . . . .	11
3.2	Daten ohne Klasseneinteilung . . . . .	11
3.3	Daten mit Klasseneinteilung . . . . .	12
3.4	Graphische Darstellungen . . . . .	13
3.4.1	Stab-, Balken- und Kreisdiagramme . . . . .	13
3.4.2	Histogramme . . . . .	14
3.5	Kumulierte Häufigkeitsverteilungen . . . . .	15
3.5.1	Daten ohne Klasseneinteilung . . . . .	15
3.5.2	Daten mit Klasseneinteilung . . . . .	16
3.6	Useful Matlab codes . . . . .	17
<b>4</b>	<b>Statistische Masszahlen</b>	<b>18</b>
4.1	Einführung . . . . .	18
4.2	Lagemasszahlen . . . . .	18
4.2.1	Der arithmetische Mittelwert . . . . .	18
4.2.2	Der Median oder Zentralwert . . . . .	19
4.2.3	Der Modus oder Modalwert . . . . .	20
4.3	Quantil und Box-Plot . . . . .	20
4.4	Streumasse . . . . .	21
4.4.1	Die Varianz und die Standardabweichung . . . . .	21
4.4.2	Standardisierte Variablen . . . . .	22
4.5	Übungen . . . . .	23
4.6	Useful Matlab codes . . . . .	24
<b>5</b>	<b>Lineare Regression</b>	<b>25</b>
5.1	Einführung . . . . .	25
5.2	Graphische Methode. . . . .	26
5.3	Methode der kleinsten Quadrate . . . . .	26
5.3.1	Matrizennotation und Verallgemeinerung . . . . .	30
5.4	Useful Matlab codes . . . . .	33
5.5	Aufgaben . . . . .	34
<b>6</b>	<b>Solutions I</b>	<b>35</b>

<b>II</b>	<b>Wahrscheinlichkeitsverteilungen</b>	<b>36</b>
<b>7</b>	<b>Der Begriff des Zufallsexperiments</b>	<b>37</b>
7.1	Einführung . . . . .	37
7.2	Zufallsexperiment . . . . .	37
7.3	Der Begriff der Wahrscheinlichkeit . . . . .	38
7.3.1	Bedingte Wahrscheinlichkeit und Unabhängigkeit . . . . .	39
7.3.1.1	Bedingte Wahrscheinlichkeit . . . . .	40
7.3.1.2	Unabhängigkeit . . . . .	41
7.4	Aufgaben . . . . .	42
<b>8</b>	<b>Zufallsvariablen</b>	<b>45</b>
8.1	Beispiele von Zufallsvariablen . . . . .	46
8.2	Verteilungsfunktion . . . . .	48
8.3	Diskrete Zufallsvariablen . . . . .	50
8.4	Erwartungswert . . . . .	51
8.5	Varianz und Standardabweichung . . . . .	52
8.6	Die wichtigsten diskreten Verteilungen . . . . .	53
8.6.1	Bernoulli-Variable. . . . .	53
8.6.2	Binomialverteilung. . . . .	53
8.6.3	Poisson Zufallsvariable. . . . .	54
8.7	Stetige Zufallsvariablen . . . . .	56
8.8	Wichtige stetige Verteilungen . . . . .	60
8.8.1	Gleichverteilung . . . . .	60
8.8.2	Normalverteilung. . . . .	60
8.8.3	Approximation der Binomialverteilung durch eine Normalverteilung . . . . .	62
8.8.4	Exponentialverteilung. . . . .	64
8.8.5	$\chi^2$ -Verteilung. . . . .	65
8.8.6	$t$ -Verteilung (auch Student-Verteilung). . . . .	66
8.9	Gemeinsame Verteilung und Unabhängigkeit . . . . .	67
8.10	Aufgaben . . . . .	69
<b>9</b>	<b>Grenzwertsätze</b>	<b>74</b>
9.1	Gesetz der grossen Zahlen . . . . .	74
9.2	Zentraler Grenzwertsatz . . . . .	74
9.3	Aufgaben . . . . .	75
<b>10</b>	<b>Solutions II</b>	<b>76</b>
<b>III</b>	<b>Schliessende Statistik</b>	<b>86</b>
<b>11</b>	<b>Stichproben</b>	<b>87</b>
11.1	Aufgaben . . . . .	89
<b>12</b>	<b>Schätzmethoden</b>	<b>90</b>
12.1	Punktschätzung . . . . .	90
12.2	Schätzung durch Konfidenzintervalle . . . . .	92
12.2.1	Konfidenzintervall für den Erwartungswert . . . . .	93
12.2.2	Konfidenzintervall für die Varianz . . . . .	95
12.2.3	Konfidenzintervall für Anteilswert . . . . .	96
12.3	Useful Matlab codes . . . . .	99

12.4 Aufgaben	99
<b>13 Statistischer Test</b>	<b>102</b>
13.1 Parametrischer Test	102
13.1.1 Sprachgebrauch.	102
13.1.2 Test für Mittelwert, Standardabweichung bekannt.	105
13.1.3 Test für Mittelwert, Standardabweichung unbekannt.	108
13.1.4 Test für Varianz einer normalverteilten Zufallsvariablen.	108
13.1.5 Test auf Anteilswert	109
13.1.6 Vergleich der Mittel zweier Populationen	110
13.2 Nichtparametrische Tests	113
13.2.1 Schätzung einer Verteilungsfunktion	113
13.2.2 Annäherung an eine Verteilung via p-p-Plot	114
13.2.3 Annäherung an eine Verteilung via q-q-Plot	115
13.3 Aufgaben	116
<b>14 Solutions III</b>	<b>118</b>

# Kapitel 1

## Einführung

Die Statistik beschäftigt sich mit wissenschaftlichen Methoden des Sammelns, der Gliederung, der Zusammenfassung, der Darstellung und der Analyse von Daten, ebenso wie dem Ziehen von gültigen Schlüssen und dem Treffen von vernünftigen Entscheidungen auf Grund einer solchen Analyse. Wichtige Anwendungen der Statistik befinden sich im Bereich der Qualitätssicherung.

**Beispiel 1** Eine Dreherei fertigt Bolzen aus Stahl mit Nenndurchmesser 22mm. Mit dem Kunden wurde eine Toleranz von  $\pm 0.1\text{mm}$  vereinbart, d.h. die Durchmesser der Bolzen müssen im Bereich zwischen 21.90 und 22.10mm liegen.

Ein Mitarbeiter der Qualitätssicherung hat nun die Aufgabe, die Einhaltung dieser Spezifikation in der Fertigung zu überwachen. Der technisch optimale Weg, alle Teile zu vermessen und nur die in der Toleranz liegenden Bolzen an den Kunden auszuliefern, scheitert am hohen Messaufwand und an den damit verbundenen Kosten. So kommt der Mitarbeiter schnell auf die Idee, nicht alle, sondern nur "bestimmte" Bolzen zu vermessen. Doch muss er insbesondere entscheiden:

- Welche Systematik soll er für die Auswahl dieser sogenannten *Stichproben* anwenden?
- Wie gross soll der Stichprobenumfang sein, d.h. wie viele Teile soll er aussuchen?
- Welche Kennzahl soll er aus der Vermessung der Bolzen zur Beurteilung der Fertigungsqualität bilden?

Alle diese Fragen sind *vor* der Einleitung von Massnahmen zur Qualitätssicherung zu beantworten. Es geht hier um die statistische Erfassung des Problems und die daran anschliessende Planung statistischer Experimente.

Der Qualitätssicherer könnte diese Fragen folgenderweise beantworten: er bestimmt den Durchmesser von *jedem hundertsten* Bolzen einer Schichtproduktion und berechnet daraus den *Mittelwert*. Nehmen wir an, seine Berechnung ergäbe für 73 Bolzendurchmesser den Mittelwert 22.04mm.

Nun stellt sich die Frage des statistischen Schlusses, nämlich:

- Wie und mit welcher Aussagesicherheit kann er aus den Stichprobenergebnissen Rückschlüsse auf die Gesamtproduktion ziehen?

Mit dem erhaltenen Ergebnis könnte man ja nun einfach sagen, dass die Produktion richtig läuft, da ja leicht zu sehen ist, dass die in der Spezifikation angegebenen Toleranzen eingehalten werden. Unser Mitarbeiter aus der Qualitätssicherung neigt natürlich zu derselben Aussage, denn es spricht ja anhand des gefundenen Wertes von 22.04mm nichts dagegen. Er muss aber versuchen, diese Aussage statistisch zu untermauern. Denn nur dann darf er vor dem Kunden behaupten:

„*Mit einer statistischen Sicherheit von 95% wird der Mittelwert von 22.00mm bei einer Toleranz von  $\pm 0.08\text{mm}$  eingehalten.*“

Wie er zu dieser Aussage und Angabe der 95% kommt und wie er den so genannten *Vertrauensbereich* aus den Daten der Stichprobe berechnet, ist Ergebnis einer Methode, die in diesem Kurs behandelt wird. ◇

Veranschaulicht durch dieses Beispiel, müssen in einer typischen Aufgabe der Statistik Schlussfolgerungen über eine ganze Gruppe von Objekten oder Personen gezogen werden. Diese ganze Gruppe heisst die **Grundgesamtheit** – hier handelt es sich um die Gesamtproduktion von Bolzen. Es ist oft unmöglich oder unpraktisch, die gesamte Gruppe zu beobachten, vor allem dann, wenn sie sehr umfangreich ist. Um diese Grundgesamtheit zu untersuchen, muss also ein kleiner Teil davon betrachtet werden. Eine **Stichprobe** wird erhoben – bestehend hier aus jedem hundertsten Bolzen – und analysiert – hier durch Messung der Durchmesser und Berechnung des Mittelwertes. Ist eine Stichprobe für eine Grundgesamtheit repräsentativ, so können aus ihrer Analyse oft wichtige Schlüsse auf die Grundgesamtheit gezogen werden – hier in welchem Intervall der Mittelwert mit einer statistischen Sicherheit von 95% eingehalten wird.

Die Analyse einer Stichprobe ist eine Aufgabe der **beschreibenden Statistik**. Ausgangspunkt ist ein mehr oder weniger umfangreicher Datensatz in der Form

$$x_1, x_2, \dots, x_n$$

Dabei bedeutet  $x_i$  das interessierende Messergebnis der  $i$ -ten Ziehung aus der Grundmenge (z.B. der Durchmesser des  $i$ ten Bolzen). In der beschreibenden Statistik werden diese Daten dann nach Wunsch behandelt, d.h.

- klassiert
- evtl. in Klassen eingeteilt  
→ *absolute und relative Häufigkeiten*
- graphisch dargestellt  
→ *Linien- und Stabdiagramme, Histogramme*
- durch statistische Masszahlen beschrieben  
→ *Mittelwert, Median, Varianz, Standardabweichung*

Der Bereich der Statistik, der Verfahren angibt, Schlussfolgerungen auf die Grundgesamtheit zu ziehen, wird **induktive Statistik** (auch mathematische Statistik, schliessende Statistik oder Inferenzstatistik) genannt. Neben den schon erwähnten Vertrauensintervallen umfasst die induktive Statistik die folgenden Methoden:

- Parameterschätzung  
*Welche Schätzfunktion muss auf der Stichprobe definiert werden, um einen gegebenen Parameter der Grundgesamtheit zu bestimmen?*
- Statistischer Test  
*Mit welcher Sicherheit lässt sich behaupten, dass die Grundgesamtheit einen bestimmten Mittelwert oder eine bestimmte Varianz besitzt? Welche Werte einer Stichprobe sind als „Ausreisser“ zu betrachten?*

Diese Vorgehen können nie absolut sicher sein: absolute Sicherheit ist nur durch eine sogenannte Vollerhebung erreichbar. Rückschlüsse auf die Grundgesamtheit sind nur mit gewissen Unsicherheiten möglich, bedingt durch statistische Schwankungen der Zufallsauswahl der Stichprobe. Bei der Angabe von Schlussfolgerungen wird also die Sprache der **Wahrscheinlichkeit** verwendet. So wird also in der induktiven Statistik über fiktive (da noch nicht gezogene) Stichproben *raisonniert*. Der syntaktische Unterschied zur beschreibenden Statistik besteht darin, dass der Datensatz  $x_1, \dots, x_n$  einer Stichprobe zu einer Ansammlung von sogenannten Zufallsvariablen  $X_1, \dots, X_n$  mutiert:

$$\text{reale Werte : } x_1, \dots, x_n \rightarrow \text{mögliche Werte : } X_1, \dots, X_n$$

Dabei kann  $X_i$  als die Menge der möglichen Ausgänge der  $i$ ten Ziehung aus der Grundmenge angesehen werden. Die Wahrscheinlichkeitstheorie gibt den begrifflichen Rahmen, um zu fragen, wie wahrscheinlich es ist, dass  $X_i$  bei der realen Ziehung den Wert  $x_i$  annimmt. Diese Quantifizierung lässt dann Rückschlüsse auf die Grundgesamtheit zu.

## Teil I

# Beschreibende Statistik



# Kapitel 2

## Statistische Variablen

Untersucht man eine gewisse Eigenschaft (auch Merkmal oder Variable) von Personen oder Objekten, dann erhält man *Daten*. Diese Daten werden *Werte*, oder *Ausprägungen* der entsprechenden *Variable* genannt.

**Beispiel 2** Das Alter der Studenten einer gewissen Klasse der HTI-Biel sei durch die folgende Liste gegeben:

$\{22, 22, 24, 20, 23, 25, 23, 21, 20, 22, 24, 27, 25, 23, 22\}$  .

Diese Zahlen sind die Werte der Variable „Alter der Studenten in dieser Klasse“. Es ist zu vermerken, dass nur sieben verschiedene Ausprägungen vorkommen für insgesamt 15 Werte. Weiter wird der mögliche Wert 26 nicht beobachtet.  $\diamond$

Es gibt verschiedene Merkmalstypen:

- Bei Daten, die ein Ausmass widerspiegeln, wie *Messungen* oder Abzählungen, spricht man von *quantitativen Merkmalen*  
*Beispiel:* Körpergrösse, mit Werten zwischen 140 und 220cm; Alter im Beispiel 2.
- Geben die Ausprägungen hingegen eine *Qualität* und nicht ein Ausmass wieder, dann spricht man von einem *qualitativen Merkmal*.  
*Beispiele:* Geschlecht, mit Ausprägungen männlich/weiblich; Schulnote, mit Ausprägungen sehr gut, gut, befriedigend, ausreichend, mangelhaft

Es gibt noch weitere Kategorien: für die Schulnoten beispielsweise existiert eine Ordnung, während die Ausprägungen für das Geschlecht nicht sinnvoll in eine Reihenfolge gebracht werden können. Wir werden uns vor allem mit quantitativen Merkmalen, oder Variablen befassen. Solche Variable können diskret oder stetig sein:

- Eine *diskrete Variable* kann nur endlich viele oder abzählbar unendlich viele verschiedene Werte annehmen. Sie beschreibt diskrete Daten, die im allgemeinen durch *Abzählungen* gewonnen werden. *Beispiel:* die Anzahl Personen in allen Autos, die während einer Stunde an einer Stelle passieren.
- Eine *stetige Variable* kann theoretisch jeden beliebigen Wert innerhalb eines vorgegebenen Intervalls annehmen. Sie beschreibt stetige Daten, die im allgemeinen durch *Messungen* gewonnen werden.  
*Beispiel:* Die schon erwähnte Körpergrösse.

Der Unterschied zwischen diskreten und stetigen Variablen ist manchmal willkürlich. Tatsächlich könnte eine diskrete Variable mit einer grossen Anzahl möglicher Werte als eine stetige Variable betrachtet werden.

*Beispiel:* die Anzahl Einwohner in Städten einer Region

Im Gegenteil, für eine stetige Variable wird manchmal keine hohe Messgenauigkeit verlangt. Man wird sie dann wie eine diskrete Variable behandeln.

*Beispiel:* die Wohnfläche aller Wohnungen einer Stadt kann durch Abrundung als eine diskrete Variable mit ganzzahligen Werten betrachtet werden.

In solchen Zwischenfällen kann man von *quasi-stetigen* Variablen sprechen.

# Kapitel 3

## Graphische Darstellungen

### 3.1 Einführung

Bei statistischen Erhebungen werden die beobachteten oder gemessenen Werte üblicherweise in der Reihenfolge aufgeschrieben, in der sie anfallen. Wir gehen von einer solchen Erhebung vom Umfang  $n$  aus, und bezeichnen mit  $x_1, \dots, x_n$  die Ausprägungen eines Merkmals  $X$ . Die so entstehende Liste wird **Urliste**<sup>1</sup> genannt. Auch bei kleinem oder mittlerem Umfang  $n$  wird eine reine Auflistung der Rohdaten schnell unübersichtlich. Bei kleinen Stichproben hilft es oft schon, wenn man die Daten der Grösse nach ordnet. In vielen Anwendungen ist  $n$  so gross, dass eine zusammenfassende Darstellung notwendig ist.

Zu diesem Zweck muss noch die Anzahl  $k$  der möglichen Ausprägungen berücksichtigt werden. Falls diese gross ist, dann müssen die Daten in Klassen eingeteilt werden (→ Abschnitt 3.3)

### 3.2 Daten ohne Klasseneinteilung

Um die Daten soweit wie möglich zusammenzufassen, wird die Urliste nach den verschiedenen vorkommenden Ausprägungen durchsucht. Wir bezeichnen diese Werte mit  $a_1, a_2, \dots, a_n$  und nehmen an, dass sie schon nach der Grösse geordnet sind, d.h. dass  $a_1 < a_2 < \dots < a_n$ . Die absoluten bzw. relativen Häufigkeiten werden folgenderweise definiert:

- Die **absolute Häufigkeit**  $h_k$  der Ausprägung  $a_k$  ist die Anzahl von Werten in der Urliste, die mit  $a_k$  übereinstimmen.
- Die **relative Häufigkeit**  $f_k$  der Ausprägung  $a_k$  ist der Anteil von Werten in der Urliste, die mit  $a_k$  übereinstimmen.

Die Häufigkeiten  $f_1, \dots, f_n$  bzw.  $h_1, \dots, h_n$  fasst man in einer **Häufigkeitstabelle** zusammen.

**Beispiel 3** Bei Variablen mit einer geringen Anzahl von Ausprägungen, wie im Beispiel 2 (Alter der Studenten), kann die Häufigkeitstabelle von Hand erstellt werden. Zu diesem Zweck ist es oft nützlich, die beobachteten Ausprägungen unter der Form einer **Strichliste** darzustellen:

---

<sup>1</sup>auch Roh- oder Primärdaten

Alter	
20	
21	
22	
23	
24	
25	
26	
27	

Für die Variable „Alter der Studenten in der Klasse“ erhalten wir also die folgende Häufigkeitstabelle:

Alter $a_k$	absolute Häufigkeit $h_k$	relative Häufigkeit $f_k$
20	2	$2/15=0.133$
21	1	$1/15=0.067$
22	4	$4/15=0.267$
23	3	$3/15=0.200$
24	2	$2/15=0.133$
25	2	$2/15=0.133$
26	0	$0/15=0.000$
27	1	$1/15=0.067$

Tabelle 3.1: Absolute und relative Häufigkeiten

◇

In vielen Erhebungen ist der Umfang  $n$  deutlich grösser, so dass das Abzählen von Häufigkeiten sinnvollerweise mit Hilfe eines Computers erfolgt.

### 3.3 Daten mit Klasseneinteilung

Die Komprimierung der Urliste zu einer deutlich kleineren Menge  $a_1, a_2, \dots, a_n$  von Werten ist oft nicht möglich, insbesondere bei quasi-stetigen Variablen. In diesem Fall kommen viele (denkbare) Werte überhaupt nicht vor, und haben also die Häufigkeit 0. Die Idee besteht darin, die Daten in **Klassen** zu gruppieren, und eine Häufigkeitstabelle für die gruppierten Daten zu erstellen.

**Beispiel 4** Die Länge (in mm) von vierzig Schrauben wurde gemessen.

138 164 150 132 144 125 149 157 146 158  
 140 147 136 148 152 144 168 126 138 176  
 163 119 154 165 146 173 142 147 135 153  
 140 135 161 145 135 142 150 156 145 128

Alle Werte liegen im Intervall zwischen 119 und 176mm, aber von den 58 möglichen Ausprägungen werden nur 30 beobachtet.

Gruppiert man die Urliste in 7 Klassen, dann erhält man die Häufigkeitstabelle 3.2.

◇

Die Häufigkeiten werden hier nicht einzelnen Ausprägungen, sondern Intervallen zugeordnet. Es gibt immer mehrere mögliche Einteilungen; als Faustregel wird die Anzahl  $k$  der Klassen gemäss der empirischen Formel

$$k \approx \sqrt{n}$$

Klasse Nr	Untere Klassengrenze	Obere Klassengrenze	Klassenmitte	absolute Häufigkeit $h_j$	relative Häufigkeit $f_j$
1	110	120	115	1	$1/40=0.025$
2	120	130	125	3	$3/40=0.075$
3	130	140	135	9	$9/40=0.225$
4	140	150	145	14	$14/40=0.350$
5	150	160	155	6	$6/40=0.150$
6	160	170	165	5	$5/40=0.125$
7	170	180	175	2	$2/40=0.050$

Tabelle 3.2: Häufigkeitstabelle Schraubenlänge

gewählt, wobei  $n$  den Stichprobenumfang bezeichnet.

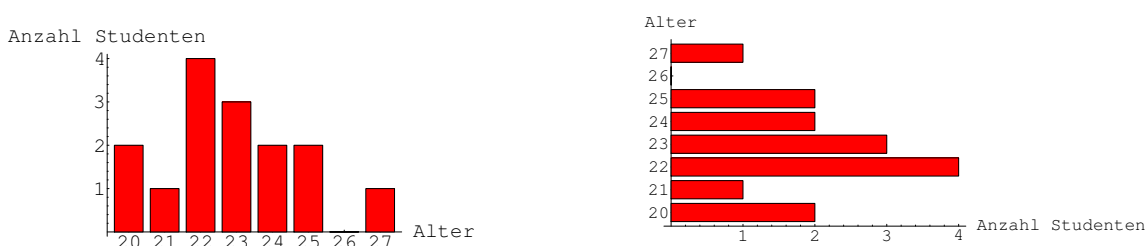
Bei der Klassierung kann es – das ist der Fall in unserem Beispiel – vorkommen, dass gewisse Werte genau auf die Klassengrenze fallen. Dann muss man entscheiden, solche Werte entweder der kleineren oder der grösseren Klasse zuzuordnen. Hier haben wir Werte, die auf Klassengrenzen fallen, jeweils der nächst *niedrigeren* Klasse zugerechnet. Dies bedeutet beispielsweise, dass die zweite Klasse gleich dem *halboffenen* Intervall  $]120, 130]$  ist.

## 3.4 Graphische Darstellungen

### 3.4.1 Stab-, Balken- und Kreisdiagramme

Bei einem *Stabdiagramm* (auch *Säulendiagramm*) werden auf der horizontalen Achse die Ausprägungen des Merkmals abgetragen und auf der vertikalen Achse die absoluten (oder die relativen) Häufigkeiten der jeweiligen Ausprägungen in Form eines Stabes. Das *Balkendiagramm* ergibt sich als eine Variante des Stabdiagramms, indem man die Ausprägungen auf der vertikalen Achse abträgt.

**Beispiel 5** Dem Beispiel 2 entspricht das folgende Stab- bzw. Balkendiagramm :



◇

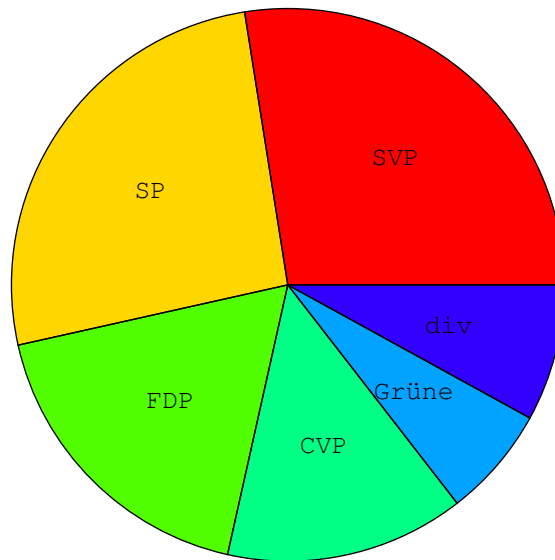
Eine weitere Darstellungsform ist das *Kreisdiagramm*, bei dem der Winkel, der den Kreis-ausschnitt einer Kategorie oder Ausprägung festlegt, proportional zur absoluten (oder relativen) Häufigkeit ist. Damit ist natürlich auch die Fläche des Kreissektors proportional zur Häufigkeit.

**Beispiel 6** Die Parteistärken im Nationalrat (im Jahr 2009) sind in der folgenden Tabelle zusammengefasst:

SVP	SP	FDP	CVP	Grüne	div.
55	52	36	28	13	16

Um diese Stärke zu vergleichen ist ein Kreisdiagramm geeignet:

Parteistärke im Nationalrat



◇

Die Stab-, Balken- und Kreisdiagramme können nur dann verwendet werden, wenn die Anzahl verschiedener Ausprägungen relativ klein ist. In anderen Fällen sind alternative Darstellungsformen besser geeignet. Die wichtigste davon ist das *Histogramm*.

### 3.4.2 Histogramme

Für eine Variable mit vielen verschiedenen Werten kann man annehmen, dass die Daten in Klassen eingeteilt sind (vgl. Abschnitt 3.3). Die Klassen sind benachbarte halboffene Intervalle

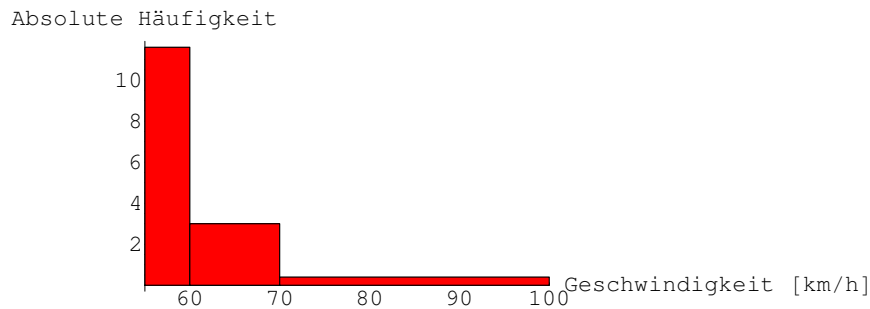
$$]c_0, c_1], ]c_1, c_2], \dots ]c_{k-1}, c_k] .$$

Über jedem von diesen Intervallen wird ein Rechteck so konstruiert, dass seine Fläche proportional zu der absoluten bzw. relativen Häufigkeit der entsprechenden Klasse ist. Die abzutragende Höhe des Rechtecks Nr  $j$  muss also gleich oder proportional zu  $h_j/d_j$  bzw.  $f_j/d_j$  gewählt werden, wobei  $d_j = c_j - c_{j-1}$  die Klassenbreite bezeichnet.

**Beispiel 7** Bei einer Radarkontrolle in einer Stadt wurden diejenigen Fahrzeuge registriert, welche die Geschwindigkeit von 55km/h überschritten haben. Bezüglich der Höhe des Verwarnungsgeldes wurden drei Klassen gebildet:

Klasse	Geschwindigkeit	absolute Häufigkeit
1	55 bis 60 km/h	58
2	60 bis 70 km/h	30
3	70 bis 100 km/h	12

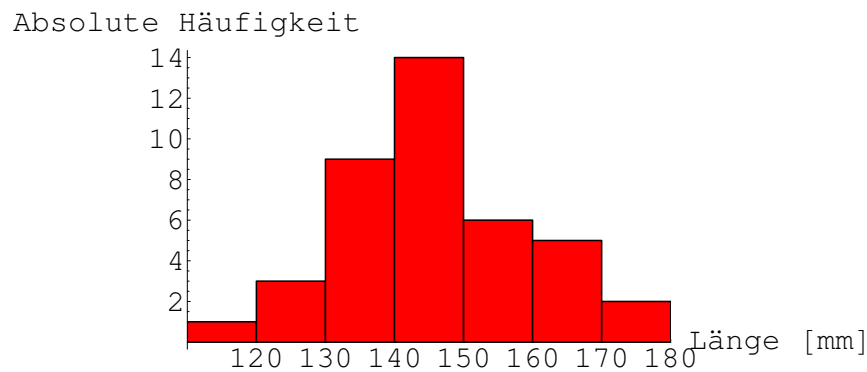
Die Klassenbreiten sind hier gegeben durch  $d_1 = 5$ ,  $d_2 = 10$  und  $d_3 = 30$ . Die Höhe der Rechtecke können also folgenderweise definiert werden:  $58/5 = 11.6$  (Klasse 1),  $30/10 = 3$  (Klasse 2) und  $12/30 = 0.4$  (Klasse 3).



◇

Falls möglich und sinnvoll, sollten die Klassenbreiten  $d_j$  gleich gross sein. Dann können als Höhe der Rechtecke auch die absoluten oder die relativen Häufigkeiten gewählt werden. Weiter sollten offene Randklassen vermieden werden.

**Beispiel 8** Aus der Häufigkeitstabelle 3.2 ergibt sich das folgende Histogramm:



Solche Histogramme werden heutzutage mit Hilfe von Softwares oder Taschenrechnern erhalten. ◇

## 3.5 Kumulierte Häufigkeitsverteilungen

### 3.5.1 Daten ohne Klasseneinteilung

Seien  $a_1 < a_2 < \dots < a_k$  die verschiedenen vorkommenden Ausprägungen eines Merkmals  $X$ . Die **absolute kumulierte Häufigkeitsverteilung**  $H(x)$  gibt die Anzahl der Werte an, die kleiner oder gleich einem gegebenen Wert  $x$  sind. Diese Funktion ist durch

$$H(x) = \sum_{a_i \leq x} h_i$$

gegeben, wobei  $h_i$  die absolute Häufigkeit von  $a_i$  bezeichnet.

Ähnlich wird die relative kumulierte Häufigkeitsverteilung, oder **empirische Verteilungsfunktion**, definiert:

$$F(x) = H(x)/n = \sum_{a_i \leq x} f_i$$

wobei  $f_i$  die relative Häufigkeit von  $a_i$  bezeichnet. Diese Funktion beschreibt den *Anteil* von Daten, die kleiner oder gleich einem gegebenen Wert  $x$  sind. Man nennt sie *empirisch*, um zu verdeutlichen, dass sie – im Gegensatz zum entsprechenden Begriff für Zufallsvariablen in der Wahrscheinlichkeitsrechnung – aus konkreten Daten berechnet wird.

Beide Funktionen  $H(x)$  und  $F(x)$  sind *monoton wachsende Treppenfunktionen*. Es gilt

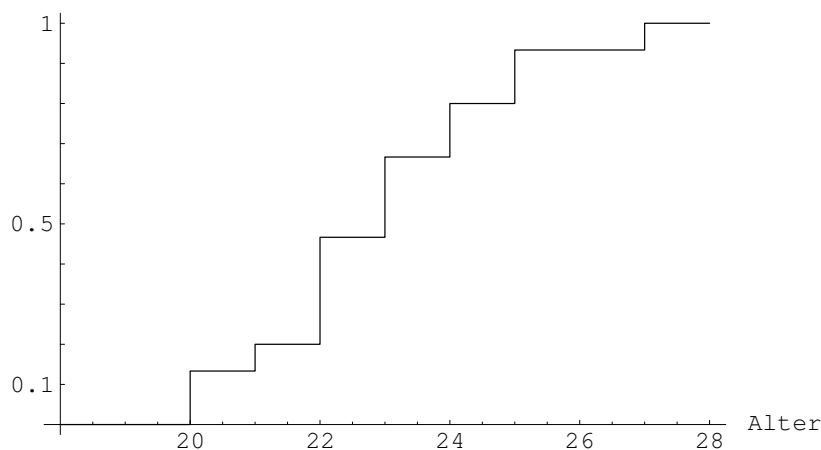
$$\begin{aligned} F(x) &= 0, & \text{falls } x < a_1 \\ F(x) &= f_1, & \text{falls } a_1 \leq x < a_2 \\ F(x) &= f_1 + f_2, & \text{falls } a_2 \leq x < a_3 \\ &\vdots & \vdots \\ F(x) &= f_1 + f_2 + \dots + f_k = 1, & \text{falls } a_k \leq x \end{aligned}$$

*Beispiel:* bei dem Alter der Studenten (Tabelle 3.1) erhalten wir die erweiterte Häufigkeitstabelle:

Alter $a_i$	$h_i$	$f_i$	$H(a_i)$	$F(a_i)$
20	2	2/15=0.133	2	2/15=0.133
21	1	1/15=0.067	3	3/15=0.200
22	4	4/15=0.267	7	7/15=0.467
23	3	3/15=0.200	10	10/15=0.667
24	2	2/15=0.133	12	12/15=0.800
25	2	2/15=0.133	14	14/15=0.933
26	0	0/15=0.000	14	14/15=0.933
27	1	1/15=0.067	15	15/15=1.000

Nun kann die empirische Verteilungsfunktion graphisch dargestellt werden:

empirische Verteilungsfunktion



### 3.5.2 Daten mit Klasseneinteilung

Sind die Daten in Klassen gruppiert, dann ist die Verteilung der Daten innerhalb jeder Klasse verloren. Die empirische Verteilungsfunktion ist also nur an den Klassengrenzen bestimmbar. Dort



ist

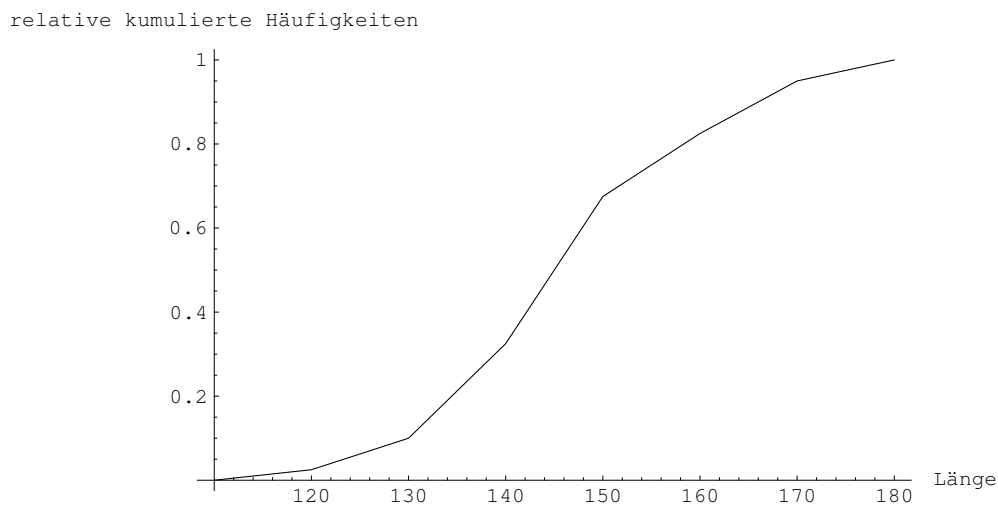
$$F(c_j) = \sum_{i \leq j} f_i, \quad j = 1, \dots, k.$$

*Beispiel:* Bei der Schraubenlänge erhalten wir die folgenden relativen kumulierten Häufigkeiten: Das

Klasse $j$	Untere Klassengrenze	Obere Klassengrenze $c_j$	Klassenmitte	relative kumulierte Häufigkeit $F(c_j)$
1	110	120	115	1/40=0.025
2	120	130	125	4/40=0.100
3	130	140	135	13/40=0.325
4	140	150	145	27/40=0.675
5	150	160	155	33/40=0.825
6	160	170	165	38/40=0.950
7	170	180	175	40/40=1.000

Tabelle 3.3: Schraubenlänge: kumulierte relative Häufigkeiten an Klassengrenzen

*Summenhäufigkeitspolygon* ist die graphische Darstellung der relativen Summenhäufigkeit  $F(c_j)$  gegen die obere Klassengrenze  $c_j$ . Zwischen den Eckpunkten wird die Verteilungsfunktion linear interpoliert.



Damit lassen sich gewisse statistische Masszahlen ermitteln.

### 3.6 Useful Matlab codes

- If  $x$  is a column vector of data points one may generate and plot an empirical cumulative distribution function using the commands:

```
[f,xf] = ecdf(x);
stairs(xf,f)
```

- To produce a disk-like pie-chart with exposed parts you may expand on the following example:

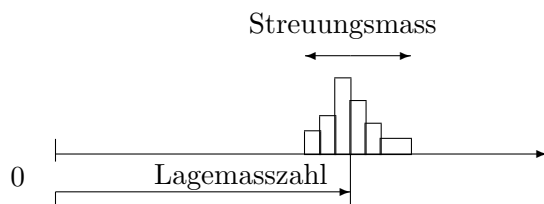
```
x = [1 3 0.5 2.5 2];
expo = [0 1 0 0 0];
pie(x,expo)
```

# Kapitel 4

## Statistische Masszahlen

### 4.1 Einführung

Die graphische Darstellung einer Häufigkeitsverteilung ist zwar nützlich, sie bringt aber keine Antwort zu den folgenden Fragen: wo liegt das Zentrum der Daten? Wie breit sind die Daten um das Zentrum verteilt? Zu diesem Zweck müssen die Daten auf *Masszahlen*, oder *Parameter*, reduziert werden. Man unterscheidet zwischen *Lagemasszahlen*, die angeben, wie gross die Stichprobenwerte etwa sind, und *Streuungsmaasse*, die besagen wie breit die Daten gestreut sind.



### 4.2 Lagemasszahlen

Die bekannteste und wichtigste Lagemasszahl ist der arithmetische Mittelwert:

#### 4.2.1 Der arithmetische Mittelwert

Der *arithmetische Mittelwert* von  $n$  Zahlen  $x_1, x_2, \dots, x_n$  ist

$$\bar{x} := \frac{1}{n} \sum_{i=1}^n x_i$$

Für Häufigkeitsdaten mit Ausprägungen  $a_1, \dots, a_k$  und relativen Häufigkeiten  $f_1, \dots, f_k$  gilt

$$\bar{x} := \sum_{i=1}^k a_i \cdot f_i$$

Man spricht dann von einem *gewogenen* oder *gewichteten* arithmetischen Mittelwert.

Sind die Daten bereits in Klassen gruppiert, so werden die Ausprägungen durch die Klassenmitten  $c_i$  ersetzt. Da ein Teil der Information in der Klasseneinteilung verloren ist, erhält man auf dieser Weise einen *Näherungswert*:

$$\bar{x} \approx \sum_{i=1}^k c_i \cdot f_i.$$

Die Berechnung des Mittelwertes mit den Klassenmitten ist also ein gewogener Mittelwert der Klassenmitten. Die Gewichte sind die relativen Häufigkeiten der Daten der jeweiligen Klasse.

### 4.2.2 Der Median oder Zentralwert

In gewissen Fällen ist der arithmetische Mittelwert nicht die beste Masszahl, um das Zentrum der Daten zu beschreiben. Dies wird anhand des folgenden Beispiels veranschaulicht.

*Beispiel:* Die Löhne einer Stichprobe von Mitarbeitern seien in der folgenden Tabelle gegeben:

Lohn (CHF)	4300	5100	6200	6750	9300	12800	18600
------------	------	------	------	------	------	-------	-------

Als mittlerer Lohn erhalten wir 9007 CHF. Dieser Wert ist offensichtlich wegen der einigen grossen Ausprägungen nach oben gezogen. Wird der grösste Lohn weiter auf CHF 21000 geändert, dann verändert sich der mittlere Lohn zu 9350 CHF. Der arithmetische Mittelwert reagiert also empfindlich auf extreme Werte (Ausreisser). Dies kann unerwünscht sein, beispielsweise wenn solche Ausreisser durch Fehler bei der Datenerhebung oder Messung verursacht wurden. Masszahlen, die den Einfluss solcher Grenzwerte begrenzen, heissen *resistent* oder *robust*.

Der **Median** wird so definiert, dass er die Stichprobe *in zwei gleiche Hälften teilt*. Sind die Daten einer Stichprobe der Grösse nach geordnet, so ist der Median, geschrieben  $\tilde{x}$  oder  $x_{1/2}$ , der Wert in der Mitte der Liste (wenn eine ungerade Anzahl von Daten vorliegt) oder der Mittelwert der beiden Werte in der Mitte (wenn eine gerade Anzahl von Daten vorliegt). Bezeichnet  $x_{(i)}$  der  $i$ -kleinste Wert, also insbesondere

$$x_{(1)} = \min_{1 \leq i \leq n} x_i \quad \text{und} \quad x_{(n)} = \max_{1 \leq i \leq n} x_i,$$

dann erhalten wir für den **Median** die folgende Formel:

$$x_{1/2} := \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ ungerade} \\ \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right), & n \text{ gerade} \end{cases}$$

Im obigen Beispiel ergibt sich als Median der Wert Nr 4, also 6750 CHF, was eine bessere Idee des Zentrums gibt als der Mittelwert. Weiter bleibt der Median unverändert, wenn z.B. der grösste Wert verändert wird: in diesem Sinn handelt es sich um eine robuste Masszahl.

Falls die Daten in Klassen gruppiert sind, muss man zuerst die *Einfallsklasse* bestimmen, d.h. die Klasse in welcher die Folge der relativen kumulierten Häufigkeiten  $F(c_i)$  erstmals den Wert 0.5 überschreitet. Da die Urliste nicht mehr zur Verfügung steht, nimmt man an, dass die Daten in der Einfallsklasse gleichmässig verteilt sind.

Das Bestimmen kann graphisch erzielt werden, indem man den Schnittpunkt zwischen dem Summenhäufigkeitspolygon und der Gerade  $y = 0.5$  bestimmt.

Der Begriff Median lässt sich einfach verallgemeinern: das führt zu dem Begriff vom *Quantil*, vgl. Abschnitt 4.3.

### 4.2.3 Der Modus oder Modalwert

Der *Modus* einer Stichprobe ist derjenige Wert, der am häufigsten auftritt. Eine Stichprobe kann einen (*unimodale* Verteilung), mehrere (*multimodale* Verteilung) oder überhaupt keinen Modus aufweisen. In der Darstellung durch Stabdiagramme ist der Modus die Ausprägung mit dem höchsten Stab. Der Modus einer Stichprobe  $x_1, x_2, \dots, x_n$  wird oft mit  $\hat{x}$  bezeichnet.

Bei gruppierten Daten ist es sinnvoll die Klasse mit der grössten Beobachtungszahl (*Modal-klasse*) zu betrachten. Die Mitte dieser Modalklasse wird als Näherungswert für denjenigen Modus verwendet. Es ist aber zu beachten, dass der wahre Modus, d.h. der der Urliste, noch nicht einmal in der Modalklasse liegen muss!

## 4.3 Quantil und Box-Plot

Sei  $0 < p < 1$  eine Zahl. Das  $p$ -Quantil einer Verteilung wird so definiert, dass es die Daten in zwei Teile trennt und zwar so, dass etwa  $100\%p$  der Daten darunter und  $100\%(1 - p)$  darüber liegen.

Wir bezeichnen, wie bei der Definition des Medians, die der Grösse nach geordnete Liste der Daten mit Klammern

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

Das  $p$ -Quantil  $x_p$  der Stichprobe ist definiert durch

$$x_p := \begin{cases} x_{(\lfloor n \cdot p + 1 \rfloor)}, & n \cdot p \notin \mathbb{Z} \\ \frac{1}{2} (x_{(n \cdot p)} + x_{(n \cdot p + 1)}), & n \cdot p \in \mathbb{Z} \end{cases}$$

wobei  $\lfloor x \rfloor$  die grösste natürliche Zahl bezeichnet, welche kleiner oder gleich  $x$  ist:

$$\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k \leq x\}.$$

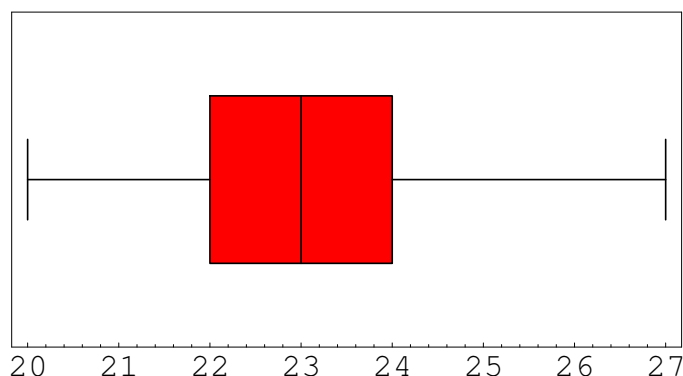
Es folgt unmittelbar aus dieser Definition, dass mindestens  $p \cdot 100\%$  (bzw.  $(1 - p) \cdot 100\%$ ) aller Stichprobenwerte kleiner oder gleich (bzw. grösser oder gleich)  $x_p$  sind. Neben dem Median als 0.5-Quantil besitzen auch weitere häufig verwendete Quantile eigene Namen. So heissen  $x_{0.25}$  und  $x_{0.75}$  das *untere* bzw. *obere Quartil*, und  $x_{j \cdot 0.1}$  das  $j$ -te Dezil ( $j = 1, \dots, 9$ ).

Zusammen mit dem kleinsten und grössten Wert geben die Quartile und der Median Informationen über die Verteilung der Daten. Diese Beschreibung, bestehend aus den Werten

$$x_{(1)}, x_{0.25}, x_{0.5}, x_{0.75}, x_{(n)},$$

heisst *Fünf-Punkte-Zusammenfassung* der Verteilung. Diese Zusammenfassung führt zur komprimierten Visualisierung einer Verteilung durch den sogenannten *Box-Plot*. Der Box-Plot besteht aus einer Schachtel („Box“), mit Anfang bzw. Ende bei dem unteren bzw. oberen Quartil. In der Box wird der Median durch einen vertikalen Strich dargestellt. Weiter gehen zwei Linien („Whiskers“) aus der Box bis zu  $x_{(1)}$  und  $x_{(n)}$ .

**Beispiel 9** Die Daten im Beispiel 2 werden folgenderweise durch den Box-Plot zusammengefasst:



◇

Verschiedene Varianten des Box-Plots stehen zu Verfügung!

## 4.4 Streumasse

Die Lagemasszahlen sagen nur, wie gross die Stichprobenwerte im Durchschnitt sind; sie geben aber keine Information über die Streuung dieser Daten um das Zentrum. Zum Beispiel besitzen die Stichproben  $\{-1, 0, 1\}$  und  $\{-10, 0, 10\}$  den gleichen arithmetischen Mittelwert; die Werte der zweiten Stichprobe sind aber offensichtlich breiter verteilt als die Werte der ersten Stichprobe. Ein *Streuungsmaß* ist ein Mass, das beschreibt, wie breit die Stichprobenwerte um den Mittelwert verteilt sind.

In diesem Abschnitt werden verschiedene Streuungsmasse vorgestellt.

### 4.4.1 Die Varianz und die Standardabweichung

Um die Streuung der Daten  $x_1, x_2, \dots, x_n$  um ihren Mittelwert  $\bar{x}$  zu messen, müssen die Abweichungen  $x_i - \bar{x}$  betrachtet werden. Dabei treten aber sowohl positive wie negative Werte auf, so dass die Summe aller Abweichungen keine geeignete Masszahl für die Streuung ist. Tatsächlich ist diese Summe gleich Null:

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

Die *mittlere Abweichung* von  $n$  Zahlen  $x_1, x_2, \dots, x_n$  ist

$$e := \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Die Absolutbeträge in dieser Formel gewährleisten, dass positive und negative Abweichungen der Daten vom Mittelwert in gleicher Weise berücksichtigt werden. Diese Absolutbeträge bilden aber auch einen grossen Nachteil dieser Masszahl, in dem Sinne, dass sie die Rechnung recht umständlich machen. Die mittlere Abweichung wird aus diesem Grund wenig gebraucht.

Eine andere Methode positive und negative Abweichungen in gleicher Weise zu berücksichtigen besteht darin, den quadratischen Mittelwert der Abweichungen zu betrachten: das führt zu der Definition der *Varianz*:

Die *Varianz* von  $n$  Zahlen  $x_1, x_2, \dots, x_n$  ist der Mittelwert der quadratischen Abweichungen:

$$s^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

und die **Standardabweichung** ist die Quadratwurzel aus der Varianz, nämlich

$$s := \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Infolge des Quadrierens hat  $s^2$  nicht die gleiche Masseinheit wie die Werte selbst. Die Standardabweichung  $s$  hingegen misst die Streuung um den Mittelwert  $\bar{x}$  mit der gleichen Masseinheit.

Für Häufigkeitsdaten mit Ausprägungen  $a_1, \dots, a_k$  und relative Häufigkeiten  $f_1, \dots, f_k$  wird die Varianz  $s^2$  durch den folgenden gewichteten Mittelwert erhalten:

$$s^2 = \sum_{i=1}^n (a_i - \bar{x})^2 f_i.$$

Sind die Daten bereits in Klassen gruppiert, so werden die Ausprägungen durch die Klassenmitten  $c_i$  ersetzt, was uns zu einem *Näherungswert*

$$s^2 \approx \sum_{i=1}^k (c_i - \bar{x})^2 f_i$$

führt. Die Varianz wird (vor allem in der Statistik) auch oft in leicht modifizierter Weise definiert, indem man statt  $n$  durch  $n - 1$  dividiert:

$$s'^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2;$$

man spricht dann von der **Stichprobenvarianz**. Es ist die Definition die wir in diesem Kurs meistens verwenden werden. Entsprechend gilt für die **Standardabweichung der Stichprobe**

$$s' := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Die Mittelung durch  $n - 1$  statt durch  $n$  lässt sich folgenderweise rechtfertigen: da die Summe aller Abweichungen gleich Null ist, ist z.B. die letzte Abweichung  $x_n - \bar{x}$  von den  $n - 1$  ersten abhängig. Somit variieren nur  $n - 1$  Abweichungen frei und man mittelt deshalb durch die Anzahl  $n - 1$  dieser sogenannten *Freiheitsgrade*. Bei Grösserem Umfang  $n$  ist der Unterschied zwischen  $s^2$  und  $s'^2$  in den meisten Fällen vernachlässigbar.

#### 4.4.2 Standardisierte Variablen

Eine Variable heisst **standardisiert**, wenn sie in bezug auf den Mittelwert und in Einheiten der Standardabweichung gemessen wird, also

$$z_i := \frac{x_i - \bar{x}}{s'}$$

Standardisierte Variablen sind reine Verhältniszahlen und damit unabhängig von der verwendeten Einheit. Beim Vergleich von Verteilungen sind sie besonders nützlich.

## 4.5 Übungen

Die Übungen dürfen auch mit dem Statistikprogramm 'R' gelöst werden. Es ist frei auf Internet zugänglich.

**Übung 1** Die für's Jahr 2006 zusätzlichen Strassenfahrzeuge auf Schweizer-strassen von Haltern mit Wohnsitz in der Schweiz ist in der folgenden Tabelle gegeben<sup>1</sup>:

Total Motorfahrzeuge	338'615
Personenwagen	260'682
Personentransportfahrzeuge	2'785
Sachentransportfahrzeuge	23'535
Landwirtschaftsfahrzeuge	3'371
Industriefahrzeuge	3'012
Motorräder	45'230

Stellen Sie diese Daten graphisch dar, mittels

- (a) Stabdiagramm
- (b) Kreisdiagramm

◊

**Übung 2** Der Durchmesser von 50 Bolzen wurde in mm gemessen:

21.8 21.9 21.9 22.2 22.0 22.1 22.0 22.0 21.8 21.8  
 22.0 22.0 21.8 22.1 22.0 22.2 22.0 21.9 22.0 22.0  
 22.2 22.1 22.2 22.0 21.6 22.0 21.9 21.5 22.0 22.3  
 21.6 22.3 21.9 21.9 21.7 22.4 22.1 22.3 22.2 22.1  
 21.8 22.2 22.2 22.1 22.0 22.1 22.5 22.4 22.1 21.8

- (a) Geben Sie die Häufigkeitstabelle an.
- (b) Stellen Sie die Daten graphisch dar.
- (c) Berechnen Sie den arithmetischen Mittelwert sowie die Standardabweichung.
- (d) Zeichnen Sie die empirische Verteilungsfunktion.

◊

**Übung 3** Auf einer Hauptstrasse, wo die Geschwindigkeit auf 80 km/h begrenzt ist, misst ein Radar die Geschwindigkeit aller Autos während einer Stunde.

84 81 76 71 80 81 83 84 80 83  
 74 75 92 76 80 82 94 73 83 83  
 75 81 79 97 78 82 76 78 82 82  
 78 81 91 68 82 73 82 79 75 77  
 83 80 77 81 69 78 81 83 87 87

<sup>1</sup>Bundesamt für Statistik (BFS), Website Statistik Schweiz 2006

- (a) Fassen Sie diese Daten in eine Häufigkeitstabelle zusammen und bilden Sie eine graphische Darstellung.
- (b) Bestimmen Sie daraus den Mittelwert und die Standardabweichung.
- (c) Verwenden Sie die kumulierten Häufigkeiten, um den Median sowie die unteren und oberen Quartile zu berechnen.
- (d) Zeichnen Sie den Box-Plot.

◉

**Übung 4** Die Brenndauer (in Stunden) von 50 Glühlampen wurde gemessen:

15	238	164	222	764	501	2	43	140	104
492	158	85	311	432	130	308	954	489	491
335	60	209	104	286	229	22	347	326	332
20	225	89	125	61	34	3	287	125	318
91	305	192	491	209	168	869	183	541	552

- (a) Fassen Sie diese Daten in eine Häufigkeitstabelle zusammen
- (b) Bestimmen Sie daraus den Mittelwert und die Standardabweichung.
- (c) Verwenden Sie eine geeignete graphische Darstellung, um den Median sowie untere und obere Quartile zu berechnen.
- (d) Zeichnen Sie den Box-Plot.

◉

## 4.6 Useful Matlab codes

- Loading data from a file you may use:

```
data=importdata('filename');
```

- If `data` is a column vector containing numerical values you may calculate the mean, standard deviation, median and drawing a histogramm by using:

```
mean(data), std(data), median(data), hist(data)
```

- The five point summary can be achieved using

```
y = quantile(x, [.025 .25 .50 .75 .975]);
```

- Box plot is easily done using (not available for Octave):

```
boxplot(x);
```



# Kapitel 5

## Lineare Regression

### 5.1 Einführung

Wir haben gelernt einen Datensatz  $x_1, \dots, x_n$  zusammenzufassen. Wir gehen dabei davon aus, dass diese Daten aus einer statistischen Erhebung **eines** Charakters  $X$  aus einer Grundmenge hervorgeht. Oft ist man aber mit dem Problem konfrontiert, Abhängigkeiten zwischen zwei oder mehreren Merkmalen  $X, Y, Z, \dots$  zu erkennen und zu beschreiben. Die Analyse statistischer Datensätze mit mehreren quantitativen Merkmalen besteht darin, die möglichen Relationen zwischen diesen Ausprägungen einer Grundmenge zu studieren. Wir werden uns hier auf das Untersuchen von Beziehungen zwischen zwei Merkmalen  $X$  und  $Y$  beschränken.

#### Beispiele:

- (i) An einer bestimmten Anzahl Stahlstäben gleicher Länge und gleichen Durchmessers wird der Kohlenstoffgehalt  $X$  und die Zugfestigkeit  $Y$  gemessen.
- (ii) Bei jeder Person  $i$  einer Grundgesamtheit  $I$  werde Körpergrösse  $x_i$  und Körpergewicht  $y_i$  gemessen.
- (iii) In einer Gruppe von Ehepaaren bestimmt man das Alter des Ehemannes  $X$  und der Ehefrau  $Y$  bei der Eheschliessung.
- (iv) Bei einer Anzahl von Personenwagen ermittelt man den Hubraum  $X$  und die Leistung  $Y$ .

Obwohl in diesen Fällen kein direkter funktionaler Zusammenhang zwischen  $X$  und  $Y$  besteht, ist klar, dass die Grössen nicht ganz unabhängig voneinander sind. Man sagt, die beiden Grössen  $X$  und  $Y$  seien **korreliert**.

Wir betrachten also eine Stichprobe von  $n$  Wertepaaren  $(x_i, y_i)$  und stellen uns folgende Fragen:

- (i) **Regressionsproblem:** Kann der Zusammenhang zwischen den beiden Werten durch eine möglichst einfache Gleichung angenähert werden?
- (ii) **Korrelationsproblem:** Kann die Güte einer solchen Regression durch ein geeignetes Mass charakterisiert werden?

Wir werden uns hier auf das Aufsuchen einer linearen Relation

$$y = ax + b$$

zwischen zwei Merkmalen  $X$  und  $Y$  eines statistischen Datensatzes  $\{(x_i, y_i)\}_{i \in I}$  beschränken. Solch eine Beziehung ist allgemein als **lineare Regression** bekannt.

Es sei angemerkt, dass es sich im Allgemeinen bei der Lösung eines Regressionsproblems um eine reine mathematische Beziehung handelt. Von ihr ausgehend darf nicht einfach auf eine physikalische Ursache-Wirkung Beziehung zwischen  $Y$  und  $X$  geschlossen werden.

## 5.2 Graphische Methode.

Messung $N^o$	1	...	$i$	...	$n$
$X$	$x_1$	...	$x_i$	...	$x_n$
$Y$	$y_1$	...	$y_i$	...	$y_n$

Als Ausgangslage dient die kartesische Darstellung der Punkte  $M_i(x_i, y_i)$  in der Zahlenebene. Man erhält so eine Punktwolke. Die Form dieser Punktmenge suggeriert eine allfällige affine Relation, falls sich die Menge genügend nahe an ein Geradenstück schmiegt. Diese subjektive Einschätzung und das Angeben einer entsprechenden Geraden kann als eine erste Annäherung zum Verständnis der Daten viel beitragen.

**Beispiel 10** Folgende Messdaten seien gegeben:

Messung $N^o$	1	2	3	4	5	6	7	8	9	10
$X$	3	5	6	6	8	10	10	12	13	14
$Y$	5	5	6	7	7	8	9	8	10	9

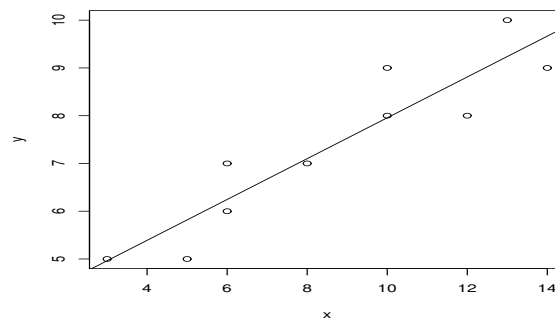


Abbildung 5.1: Regressionsgerade  $y = 0.43x + 3.68$ .

Die Form der Menge suggeriert eine affine Relation zwischen  $Y$  und  $X$  durch die Geradengleichung  $y = 0.43x + 3.68$ . Die Gerade ist so gewählt, dass sie durch den Schwerpunkt  $G(\bar{x}, \bar{y})$  dieser Punktwolke geht.  $\diamond$

## 5.3 Methode der kleinsten Quadrate

Das obige doch eher subjektive Verfahren wird mit der Methode der kleinsten Quadrate formalisiert. Sie liefert die Lösung des Regressionsproblems: Sei

$$D_{Y/X} : \hat{y} = ax + b$$

die Geradengleichung, die die **lineare Regression von  $Y$  bezüglich  $X$**  beschreibt. Gesucht sind die Parameter  $a$  und  $b$ .

Der Punkt  $\hat{y}_i$  sei der Ordinatenwert der Geraden  $D_{Y/X}$  beim Abszissenwert  $x_i$  ( $i$ -ter Wert der Ausprägung  $X$ ). Wir setzen

$$\varepsilon_i := y_i - \hat{y}_i.$$

$\varepsilon_i$  ist also die Differenz zwischen dem Messwert  $y_i$  und dem Ordinatenwert  $\hat{y}_i$ . Die Methode der kleinsten Quadrate besteht darin, die Parameterwerte  $a$  und  $b$  so zu bestimmen, dass die Summe  $S$  der quadrierten Differenzen

$$S = \sum_{i=1}^n \varepsilon_i^2$$

für die gegebene Menge aus  $n$  Punkten minimal wird.

**Beispiel 11** Mit den Werten aus obigem Beispiel ist  $\varepsilon_i$  jeweils als Distanz zwischen  $M_i$  und der Regressionsgeraden gegeben.  $\diamond$

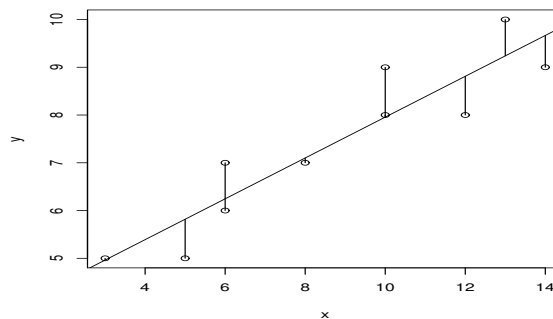


Abbildung 5.2: Darstellung der Abstände  $\varepsilon_i$ .

**Analytische Lösung.** Die Abstände berechnen sich zu

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (ax_i + b)$$

also

$$\varepsilon_i^2 = y_i^2 - 2(ax_i + b)y_i + (ax_i + b)^2.$$

Im Ausdruck  $\sum \varepsilon_i^2$  sind  $a$  und  $b$  die Unbekannten, die Werte  $x_i$  und  $y_i$  sind bekannt. Wir entwickeln  $S$  nach  $a$  und  $b$ :

$$S(a, b) = a^2 \sum x_i^2 + 2ab \sum x_i + nb^2 - 2a \sum x_i y_i - 2b \sum y_i + \sum y_i^2$$

und erhalten eine Gleichung zweiter Ordnung in den Variablen  $a$  und  $b$ . Da die Koeffizienten von  $a^2$  und  $b^2$  positiv sind, hat  $S(a, b)$  ein Minimum das durch die üblichen Bedingungen gefunden wird

$$\frac{\partial S}{\partial a} = 0 \tag{5.1}$$

$$\frac{\partial S}{\partial b} = 0. \tag{5.2}$$

Wir haben

$$\begin{aligned} \frac{\partial S}{\partial a} &= 2a \sum x_i^2 + 2b \sum x_i - 2 \sum x_i y_i, \\ \frac{\partial S}{\partial b} &= 2a \sum x_i + 2nb - 2 \sum y_i = 2n(b + a\bar{x} - \bar{y}). \end{aligned} \tag{5.3}$$

Es folgt aus der letzten Gleichung, dass  $\frac{\partial S}{\partial b} = 0$  die Gleichung  $b = \bar{y} - a\bar{x}$  impliziert. Anders gesagt,  $\bar{y} = a\bar{x} + b$  und wir erkennen: die Regressionsgerade geht durch den Schwerpunkt  $G(\bar{x}, \bar{y})$  der Punktmenge.

Wir erhalten entsprechend

$$\frac{\partial S}{\partial a} = 2a \sum x_i^2 + 2(\bar{y} - a\bar{x}) \sum x_i - 2 \sum x_i y_i,$$

und die Bedingung  $\frac{\partial S}{\partial a} = 0$  liefert

$$a = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\frac{1}{n-1} \sum (x_i y_i - \bar{x}\bar{y})}{\frac{1}{n-1} \sum (x_i^2 - \bar{x}^2)}.$$

der Nenner ist gerade die empirische Varianz der  $x_i$  (Übung!)

$$s_X^2 = \text{Var}(X) = \frac{1}{n-1} \sum (x_i^2 - \bar{x}^2) = \frac{1}{n-1} \sum (x_i - \bar{x})^2.$$

Was den Zähler angeht, ist seine Struktur der der Varianz ähnlich. Man nennt diesen Ausdruck die **Kovarianz** (genauer, Kovarianz der Stichprobe) von  $X$  und  $Y$ :

$$\boxed{\text{Cov}(X, Y) := \frac{1}{n-1} \sum (x_i y_i - \bar{x}\bar{y}) = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) \stackrel{\text{not.}}{=} s_{XY}.}$$

Die **Koeffizienten der linearen Regression** von  $Y$  bezüglich  $X$  sind entsprechend

$$\boxed{\begin{aligned} \hat{y} &= ax + b & \text{mit} \\ a &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} & \text{und} \\ b &= \bar{y} - a\bar{x}. \end{aligned}}$$

**Bemerkung.** Die Regressionsgerade von  $Y$  bezüglich  $X$  gibt jedem  $x$ -Wert einen ausgeglichenen  $y$ -Wert. Die  $x$  Variable wird dabei als die *unabhängige Variable* betrachtet; die  $y$ -Variable ist entsprechend die *abhängige Variable*. Umgekehrt kann auch die Regressionsgerade von  $X$  bezüglich  $Y$  ausgerechnet werden (einfache Vertauschung der Variablen). In den Anwendungen ist oft im Voraus klar welche der Variable die Unabhängige ist. Entsprechend können die Gleichungen angesetzt werden.

### Korrelationsproblem, linearer Korrelationskoeffizient.

Abgesehen von einigen wenigen Ausnahmen können die Koeffizienten  $a$  und  $b$  immer ausgerechnet werden. Der graphische Vergleich der Punktmenge und der Geraden zeigt jedoch, dass die lineare Regression nicht immer sinnvoll ist.

Im Falle der Methode der kleinsten Quadrate ist es angezeigt ein mathematisches Kriterium zu haben, das uns grünes Licht für ihre Anwendung gibt.

Schauen wir uns dazu die Variation von  $\sum \varepsilon_i^2$  an; sie ist die entscheidende Grösse für unser Bedürfnis. Falls die Variation klein ist, ist die Regressionsgerade eine sinnvolle Annäherung an die Punktwolke. Die Abstände sind:

$$\varepsilon_i = y_i - \hat{y}_i = y_i - (ax_i + b) = y_i - \bar{y} - a(x_i - \bar{x}).$$

Wird die Summe der Abstandsquadrate ausgerechnet (Übung!) folgt:

$$\sum \varepsilon_i^2 = n[\text{Var}(Y) - a^2 \text{Var}(X)]$$

Die Variation von  $\sum \varepsilon_i^2$  hängt also vom absoluten Unterschied von  $[\text{Var}(Y) - a^2 \text{Var}(X)]$  bezüglich 0 oder vom relativen Unterschied von  $\frac{a^2 \text{Var}(X)}{\text{Var}(Y)}$  bezüglich 1 ab.

Betrachten wir nun den Ausdruck

$$r^2 = \frac{a^2 \text{Var}(X)}{\text{Var}(Y)} = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X)\text{Var}(Y)}.$$

Es gilt  $r^2 \geq 0$ , und da  $\sum \varepsilon_i^2 \geq 0$  gilt, ist  $\text{Var}(Y) \geq a^2 \text{Var}(X)$  und also  $r^2 \leq 1$ . Falls alle Punkte  $M_i(x_i, y_i)$  auf einer Geraden liegen gilt  $\varepsilon_i = 0$  für alle  $i$  und deshalb  $r^2 = 1$ . Umgekehrt, falls  $r^2 = 1$  gilt, ist  $\sum \varepsilon_i^2 = 0$  und die Punkte  $M_i$  sind auf einer Geraden. Deshalb gilt,  $r^2 = 1$  genau dann wenn die  $M_i$ 's auf einer Geraden sind. Die Grösse  $r$  ist somit ein geeignetes Mass um die Stärke des linearen Zusammenhangs zwischen den  $x_i$  und  $y_i$  auszudrücken. Wir definieren:

**Der lineare Korrelationskoeffizient** zwischen  $X$  und  $Y$  ist:

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

**Bemerkungen.**

- (i)  $r_{X,Y}$  zeigt an, wie sehr die Grössen  $X$  und  $Y$  (linear) korrelieren.
- (ii) Der Korrelationskoeffizient kann Werte zwischen  $-1$  und  $1$  annehmen. Falls  $r_{X,Y} = 0$  ist, sind die Variablen  $X$  und  $Y$  **unkorreliert**. Ansonsten sind sie korreliert.
- (iii) Generell wird die Korrelation ab einem Korrelationskoeffizient von  $|r| \geq 0.9$  als gut eingestuft. In diesem Fall liefert die lineare Regression eine gute Zusammenfassung des Datensatzes.
- (iv) Falls  $r_{X,Y} < 0$  ist, sind  $X$  und  $Y$  negativ korreliert und die Regressionsgerade fällt. Falls  $r_{X,Y} > 0$  ist, sind  $X$  und  $Y$  positiv korreliert und die Regressionsgerade steigt.

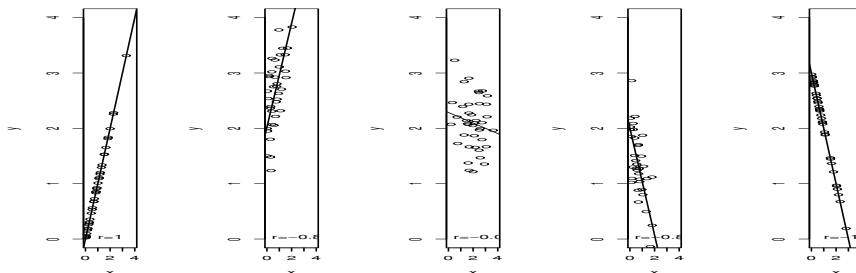


Abbildung 5.3: Datensätze mit stark positiver, schwacher und stark negativer Korrelation.

**Beispiel 12** Schauen wir uns nochmals das Beispiel 10 an.

Mesure $N^o$	1	2	3	4	5	6	7	8	9	10
$X$	3	5	6	6	8	10	10	12	13	14
$Y$	5	5	6	7	7	8	9	8	10	9

Es gilt

$$\bar{x} = 8.7; \quad \bar{y} = 7.4; \quad \text{Cov}(X, Y) = \frac{1}{10} \sum x_i y_i - \bar{x}\bar{y} = 5.22$$

$$\text{Var}(X) = \bar{x^2} - \bar{x}^2 = 12.21; \quad \text{Var}(Y) = \bar{y^2} - \bar{y}^2 = 2.64$$

$$a = \frac{5.22}{12.21} = 0.4275; \quad b = \bar{y} - a\bar{x} = 3.6806.$$

Die Regressionsgerade hat die Gleichung

$$y = 0.4275x + 3.6806$$

und der Korrelationskoeffizient  $r_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} > 0.9$  ist sehr gut.  $\diamond$

### Achtung!

Bei der Interpretation von Korrelationen ist Vorsicht geboten. Bei 6-10 jährigen Schulkindern wurde zum Beispiel eine deutliche Korrelation zwischen Körpergewicht und manueller Geschicklichkeit gefunden, die auf den ersten Blick verblüfft. Sie kann jedoch dadurch erklärt werden, dass die Kinder mit steigendem Alter im Mittel sowohl schwerer als auch geschickter werden.

Bei Zeitreihen muss man besonders aufpassen. So hat zwischen 1900 und 1970 die Zahl der Störche wie auch die Zahl der Neugeborenen zugenommen. Die beiden Zuwachsraten sind also positiv korreliert. Aber dies ist kein Beweis, dass der Klapperstorch die Babies bringt!

### 5.3.1 Matrizennotation und Verallgemeinerung

Die analytische Lösung des Regressionsproblems lässt sich sehr elegant mit Hilfe der Matrizenrechnung herleiten. Dazu nehmen wir die Gleichungen (5.1) in der Form

$$\begin{aligned} \frac{\partial S}{\partial a} &= 2a \sum x_i^2 + 2b \sum x_i - 2 \sum x_i y_i = 0, \\ \frac{\partial S}{\partial b} &= 2a \sum x_i + 2nb - 2 \sum y_i = 0. \end{aligned} \quad (5.4)$$

wieder auf. Nun führen wir folgende Abkürzungen ein:

$$\begin{aligned} S_x &:= \sum_{i=1}^n x_i, & S_y &:= \sum_{i=1}^n y_i \\ S_{xx} &:= \sum_{i=1}^n x_i^2, & S_{xy} &:= \sum_{i=1}^n x_i y_i \end{aligned}$$

Die vier Summen lassen sich leicht berechnen, falls die Punkte  $(x_i, y_i)$  der Reihe nach eingegeben werden. Die meisten Taschenrechner haben diese Befehle fest eingebaut. Dann sind  $a$  und  $b$  bestimmt durch

$$\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad (5.5)$$

Diese  $2 \times 2$ -Matrix kann invertiert werden:

$$\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix}^{-1} = \frac{1}{nS_{xx} - S_x^2} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & n \end{pmatrix}$$

und erhalten für die Koeffizienten der linearen Regressionsgeraden  $y = ax + b$ :

$$\boxed{\begin{aligned} \Delta &= nS_{xx} - S_x^2 && \text{und} \\ a &= \frac{1}{\Delta}(S_{xx}S_y - S_xS_{xy}) && \text{und} \\ b &= \frac{1}{\Delta}(nS_{xy} - S_xS_y) \end{aligned}}$$

Ausgehend von einer Messreihe  $(x_i, y_i)$  sind wir also auf das Gleichungssystem

$$\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \quad (5.6)$$

gestossen. Diese Gleichung werden wir nun elementar umschreiben mit Hilfe einer Matrix  $\mathbf{F}$  und eines Vektors  $\vec{y}$ , welche die Daten enthalten.

$$\mathbf{F} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \text{ and } \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (5.7)$$

Es gilt

$$\mathbf{F}^T \mathbf{F} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \quad (5.8)$$

und

$$\begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{F}^T \cdot \vec{y}.$$

Somit kann das obige System auch geschrieben werden als

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y} \quad (5.9)$$

wobei  $\vec{p} = (b, a)^T$  ist.

Für den Fall einer Regressionsgerade bringt diese Notationsänderung keine Vorteile. Aber mit derselben Notation können auch komplexere Probleme beschrieben werden. Die obigen Ideen können nicht nur für das Einpassen einer Geraden in eine Menge von Datenpunkten verwendet werden. Man kann auch mehrere unabhängige Variablen betrachten und muss auch nicht unbedingt Geraden (oder Ebenen) verwenden. Die grossen zusätzlichen Schwierigkeiten liegen vor allem in der Notation und etwas längeren Rechnungen, die aber ein Computer gerne übernimmt. Wir betrachten also mehrere Werte der unabhängigen Variablen  $x_1, x_2, \dots, x_n$  und Funktionen der Form

$$Y(x) = \sum_{j=1}^k p_j f_j(x)$$

wobei die Funktionen  $f_j(x)$  vorgegeben sind. Die Parameter  $p_1, \dots, p_k$  sind zu berechnen. Für die gegebenen Datenpunkte  $(x_i, y_i)$  soll der Ausdruck

$$S^2(p_1, \dots, p_k) = \sum_i^n (Y(x_i) - y_i)^2 = \sum_i^n \left( \sum_j^k p_j f_j(x_i) - y_i \right)^2$$

minimiert werden. Die Bedingung, dass alle partiellen Ableitungen bezüglich den  $p_j$  Null sein müssen führt wieder auf ein analoges System von  $k$  linearen Gleichungen für die  $k$  Unbekannten  $p_j$ .

**Beispiel 13** Mit den Parametern  $p_1 = a, p_2 = b, p_3 = c$  und den Funktionen  $f_1(x) = x^2, f_2(x) = x$  und  $f_3(x) = 1$  wird

$$Y(x) = ax^2 + bx + c$$

und wir suchen eine optimale Parabel durch die gegebenen Punkte. Weiter unten werden wir sehen, dass in diesem Fall die Matrix  $\mathbf{F}$  gegeben ist durch

$$\mathbf{F} = \begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{pmatrix}$$

und die im System  $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$  zu untersuchende Matrix  $M = \mathbf{F}^T \cdot \mathbf{F}$  ist eine  $3 \times 3$  Matrix.  $\diamond$

Es muss unbedingt herausgestrichen werden, dass die Funktionen  $f_i(x)$  durchaus nichtlinear sein können. Die Linearität bezieht sich auf die Parameter  $p_j$ .

### Herleitung des Gleichungssystems

Für gegebene Datenpunkte  $(x_i, y_i)$  für  $i = 1, \dots, n$  und eine geeignete Wahl der Basisfunktionen  $f_j(x)$  für  $j = 1, \dots, k$  suchen wir die optimalen Werte der Parameter  $p_j$ , sodass die Funktion

$$Y(x) = \sum_{j=1}^k p_j f_j(x)$$

sich möglichst gut an die gegebenen Punkte anpasst. Die Grundidee ist die selbe wie für die einfache lineare Regression.

Die zu minimierende Funktion ist

$$S^2(\vec{p}) := \sum_{i=1}^n (Y(x_i) - y_i)^2 = \sum_{i=1}^n \left( \sum_{j=1}^k p_j f_j(x_i) - y_i \right)^2.$$

Die partiellen Ableitungen sind

$$\frac{\partial Y(x_i)}{\partial p_l} = \frac{\partial}{\partial p_l} \sum_{j=1}^k p_j \cdot f_j(x_i) = f_l(x_i) \quad \text{für } l = 1, 2, \dots, k$$

und somit müssen die Gleichungen

$$\frac{\partial S(\vec{p})}{\partial p_l} = 2 \sum_{i=1}^n \left( \sum_{j=1}^k p_j \cdot f_j(x_i) - y_i \right) \cdot f_l(x_i) = 0 \quad \text{für } l = 1, 2, \dots, k$$

gelöst werden. Das kann umgeschrieben werden zu

$$\sum_{j=1}^k p_j \left( \sum_{i=1}^n f_j(x_i) \cdot f_l(x_i) \right) = \sum_{i=1}^n f_l(x_i) \cdot y_i \quad \text{für } l = 1, \dots, k. \quad (5.10)$$

Diese Rechnungen können mit Hilfe von Matrizen übersichtlicher dargestellt werden. Dazu definieren wir:

$$\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \dots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_k(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_k(x_n) \end{pmatrix}$$

Eine detaillierte Rechnung zeigt nun, dass die Lösung des Systems (5.10) identisch ist mit der Lösung  $\vec{p}$  von:

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}.$$



Solch ein System ist mit Matlab/Octave einfach durch die Befehlszeile:

$$\vec{p} = (\mathbf{F}^T \cdot \mathbf{F}) \setminus \mathbf{F}^T \cdot y$$

zu lösen.

**Beispiel 14** Muss eine Kurve der Form

$$f(x) = a \cdot x^b$$

durch Messpunkte gefittet werden, so verwendet man

$$\ln y = \ln a + b \ln x$$

und somit kann auch das Minimum der Funktion

$$\sum_{i=1}^n (\ln a - b \ln x_i - \ln y_i)^2$$

gesucht werden. Sei konkret

$$\begin{aligned} x &= [0.6; 0.7; 0.8; 0.9; 1.; 1.1; 1.2; 1.3] \\ y &= [2.61445; 2.78728; 2.69749; 2.89113; 3.0522; 3.22576; 3.4746; 3.68019] \end{aligned}$$

Die Lösung von  $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$  wobei die  $i$ -te Komponente von  $\vec{y}$  gleich  $\ln(y_i)$  ist, ist gegeben durch  $\vec{p} = (0.43151, 1.1448)^T$  und führt zum Resultat  $\ln(y) = 1.1448 + 0.43151 \ln x$  und schlussendlich

$$y = e^{1.1448} \cdot e^{0.43151 \ln x} = 3.1419 \cdot x^{0.43151}.$$

◇

## 5.4 Useful Matlab codes

- If  $x$  and  $y$  are two column vectors of same dimension you can calculate the covariance matrix and the correlation coefficient using

$$\text{cov}(x,y), \text{corrcoef}(x,y)$$

- The command `polyfit(x,y,n)` finds the coefficients of a polynomial  $p(x)$  of degree  $n$  that fits the data,  $p(x(i))$  to  $y(i)$ , in a least squares sense. For  $n = 2$  this corresponds to a first order linear regression:

$$p = \text{polyfit}(x,y,2)$$

the linear regression line can be plotted using

$$\text{plot}(x,\text{polyval}(p,x))$$

## 5.5 Aufgaben

**Übung 5** Die Messung eines temperaturabhängigen Widerstandes ergab folgenden Zusammenhang zwischen der Temperatur  $X$  ( $[C]$ ) und dem spezifischen Widerstand  $Y$  ( $[m\Omega \cdot mm^2/m]$ ):

Mesure $N^o$	1	2	3	4	5	6	7	8	9	10
$X$	0	10	20	30	40	50	60	70	80	90
$Y$	16.3	17.0	17.6	18.3	19.0	19.8	20.7	21.5	22.5	23.3

Bestimmen Sie die Mittelwerte  $\bar{x}$  und  $\bar{y}$ , die Varianzen  $s_x^2$  und  $s_y^2$ , die Kovarianz  $Cov(X, Y)$ , die Gleichung der Regressionsgeraden:  $y = ax + b$ , und geben Sie den Korrelationskoeffizienten  $r_{XY}$  an. Kommentar!  $\odot$

**Übung 6** Geben Sie mit Hilfe einer linearen Regression einen Überblick über den Energieverbrauch der Schweiz in Form von Erdölbrennstoff und Elektrizität während den Jahren 1999 bis 2007 und schätzen Sie durch Extrapolation den entsprechenden Verbrauch für 2008 und 2009 ab. Benutzen Sie dazu die Unterlagen des Bundesamtes für Energie unter:

[http://www.bfe.admin.ch/themen/00526/00541/00542/00631/index.html?lang=de&dossier\\\_id=00867](http://www.bfe.admin.ch/themen/00526/00541/00542/00631/index.html?lang=de&dossier\_id=00867).

$\odot$

# Kapitel 6

## Solutions I

**2 c)** Mean:  $\bar{x} = 22.02mm$  standard deviation:  $\sigma = 0.21mm$

**3 b)** Mean:  $\bar{x} = 80.34km/h$ , standard deviation:  $\sigma = 5.71km/h$ , (median:  $m = 81km/h$ ,  $x_{0.25} = 77km/h$ ,  $x_{0.75} = 83km/h$ )

**4 b)** Mean:  $\bar{x} = 263.62h$ , standard deviation:  $\sigma = 218.65h$ , (median:  $m = 215.5h$ ,  $x_{0.25} = 104h$ ,  $x_{0.75} = 334.5h$ )

**5)** Mean:  $\bar{x} = 45$ ,  $\bar{y} = 19.6$ , variance:  $\sigma_x^2 = 916.7$ ,  $\sigma_y^2 = 5.6$ , regression:  $Y = 16.08X + 0.08$ ,  $r_{xy} = 0.997$  (excellent)

**6a)** Regression  $Y$  =electric power consumption [TJ],  $t$  =Year:  $Y = 3513t - 6837317$ ,  $r_{xy} = 0.994$  (excellent). Extrapolation:

$$Y(2006) = 209761 \text{ [TJ]}, \quad Y(2007) = 213274 \text{ [TJ]}.$$

**6b)** Regression  $Y$  =petrol related power consumption [TJ],  $t$  =Year:  $Y = -3327t + 7169289$ ,  $r_{xy} = 0.78$  (not too reliable). Extrapolation:

$$Y(2006) = 495327 \text{ [TJ]}, \quad Y(2007) = 492000 \text{ [TJ]}.$$

## Teil II

# Wahrscheinlichkeitsverteilungen

# Kapitel 7

## Der Begriff des Zufallsexperiments

### 7.1 Einführung

Im täglichen Sprachgebrauch und in der klassischen Physik wird das Eintreffen eines Ereignisses als das Zusammentreffen klar definierter vorangehender Vorgänge, seine Ursachen, verstanden.

In dem Masse wie wir jedoch die Ursachen eines Ereignisses nicht beherrschen, sprechen wir von einem mehr oder weniger zufälligen Ereignis. Ein solches Ereignis soll als **Zufallsereignis** angesprochen werden.

Die Vorhersage eines Zufallsereignisses führt uns auf den Begriff der **Wahrscheinlichkeit** seines Auftretens. Je nach dem wie wahrscheinlich ein Eintreten ist, sprechen wir umgangssprachlich von unmöglichen, wahrscheinlichen oder sicheren Ereignissen.

Der so eingeführte Begriff der Wahrscheinlichkeit ist sehr vage und wir werden ihn in diesem Kapitel präzisieren und quantifizieren.

### 7.2 Zufallsexperiment

Die Wahrscheinlichkeitstheorie bedient sich zu ihrer experimentellen Grundlage sogenannter **Zufallsexperimente**. Das sind Experimente, die sich unter gleichen Bedingungen wiederholen lassen ohne dass das Resultat zwingend gleich ausfällt, seine Ausgänge jedoch gewissen statistischen Gesetzmässigkeiten genügen. Der Begriff zufällig kann nun folgende Bedeutung annehmen

- Der Ausgang des Experiments ist nicht vorhersehbar und ist nur statistisch erfassbar. (*Beispiel*: Beobachtung eines radioaktiven Zerfalls eines Atoms).
- Der Ausgang des Experiments ist im Prinzip vorhersehbar hängt aber von zu vielen Parametern ab, um vorhergesagt werden zu können. (*Beispiel*: Wurf eines Würfels. Bei genauer Kenntnis der Anfangsbedingungen könnte an sich ein Wurf physikalisch modelliert und sein Resultat vorhergesagt werden. In der Praxis ist sowas jedoch undenkbar).

In beiden Fällen werden wir sagen das Experiment sei vom Zufall regiert.

Das Resultat eines Zufallsexperiments ist seine Beobachtung. Die Menge aller möglichen Resultate eines Zufallsexperiments wird als **Fundamentalmenge** bezeichnet; ganz allgemein wird sie mit  $\Omega$  bezeichnet.

**Beispiel 15** Die möglichen Ausgänge des “Kopf oder Zahl” Spiels sind «Kopf» und «Zahl.» Die Fundamentalmenge ist

$$\Omega = \{Kopf, Zahl\}.$$

◇

**Beispiel 16** Die möglichen Ausgänge des Würfelwurfes sind  $1, 2, \dots, 6$ . Also,

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

◇

**Beispiel 17** Wird die Gesamtproduktion der Bolzen aus Beispiel 1  $\Omega$  genannt, so ist die zufällige Wahl eines Bolzens ein Zufallsexperiment. Das Resultat ist ein spezieller Bolzen  $\omega \in \Omega$ .

◇

Beispiel 17 ist symptomatisch für die statistische Anwendung. In der Regel ist eine gewisse Grundmenge vorgegeben (z.B. Menge der an einem Tag  $x$  produzierten Bolzen) und das Experiment besteht darin, ein Element dieser Menge zufällig auszuwählen. Entsprechend ist die Fundamentalmenge  $\Omega$  identisch mit der Grundmenge. **Achtung** bei mehrmaligem zufälligen Ziehen aus der Grundmenge ist Vorsicht geboten beim beschreiben von  $\Omega$  (siehe Übung 7).

Was heisst nun zufälliges Ziehen? In diesem Kurs werden wir diese delikate Frage nur intuitiv beantworten. Das wird für unsere Anwendungen vollauf genügen.

*Ein Element einer endlichen Grundmenge  $G$  mit  $|G| = n$  Elementen wird **zufällig gezogen**, falls seiner Wahl ein Zufallsexperiment zu Grunde liegt, so dass jedes Element von  $G$  mit gleicher Wahrscheinlichkeit, nämlich  $\frac{1}{n}$ , selektiert wird.*

Anders gesagt, die Wahl, das eine oder andere Element aus  $G$  zu ziehen, ist absolut willkürlich. Eine Realisation einer zufälligen Ziehung kann im Prinzip durch Würfeln eines (fiktiven) Würfels mit  $|G| = n$  Seiten durchgeführt werden. In der Praxis stützt man sich heutzutage auf die computerunterstützte Generierung von Zufallszahlen via Zufallsgeneratoren.

Es ist von entscheidender Bedeutung in der Statistik sogenannte Stichproben aus einer Grundmenge zufällig zu ziehen. Die Willkür des Zufalls garantiert uns, dass wir keine zusätzlichen Informationen in die Stichprobe einbringen. In unserem einführenden Beispiel 1 wurde vorgeschlagen jeden 100sten Bolzen zu testen. Das ist natürlich *keine* zufällige Wahl der Stichprobe, und es dürfen nur bedingt Rückschlüsse auf die Gesamtmenge gemacht werden.

### 7.3 Der Begriff der Wahrscheinlichkeit

Die Wahrscheinlichkeitstheorie ist eine moderne und exakte Wissenschaft! Ohne in die Details zu gehen geben wir hier eine kleine intuitive Begriffsklärung. In dieser Theorie ist die Wahrscheinlichkeit eine Funktion, definiert auf sogenannten Ereignismengen mit Werten im Intervall  $[0, 1]$ . Hier einige Definitionen:

- Wir nennen **Zufallsereignis** jede Teilmenge  $A$  der Fundamentalmenge  $\Omega : A \subseteq \Omega$ .
- **Eintreten** : das Ereignis  $A$  tritt ein, falls der Ausgang des Experiments ein Element der Menge  $A$  ist.

**Beispiel 18** Werfen eines Würfels.  $\Omega = \dots$

Eine gerade Zahl werfen  $A = \dots$

Falls der Ausgang die Zahl 4 ist, so ist das Ereignis  $A$  eingetroffen.

◇

- $\emptyset \subseteq \Omega$  :  $\emptyset$  ist das **unmögliche** Ereignis ;
- $\Omega \subseteq \Omega$  :  $\Omega$  ist das **sichere** Ereignis;
- $A \cup B$  ist das Ereignis “**A oder B**” (“oder” nicht ausschliessend) ;
- $A \cap B$  ist das Ereignis “**A und B**” ;
- $\bar{A} = \Omega \setminus A$  ist das Ereignis “**nicht A**” auch das **konträre** Ereignis von  $A$ . (Merke  $A \cup \bar{A} = \Omega$  und  $A \cap \bar{A} = \emptyset$ ) ;
- Falls  $A \cap B = \emptyset$ , sagen wir dass  $A$  und  $B$  **inkompatibel** sind.

**Bemerkung:** Falls  $A \subseteq B$  ist, sagen wir, dass  $A$  das Ereignis  $B$  **impliziert**.  
Es gilt dann  $A \cap B = A$  und  $A \cup B = B$ .

### Wahrscheinlichkeit

Wir stellen uns ein Zufallsexperiment vor, das wir  $n$  mal unter gleichen Bedingungen wiederholen. Wenn sich ein Ereignis  $A$   $k$ -mal wiederholt, nennen wir die **Frequenz** von  $A$  die Zahl  $f(A) = \frac{k}{n}$ . Sehr häufig wird sich nun, falls  $n$  sehr gross ist, die Frequenz  $f(A)$  um einen bestimmten Wert stabilisieren. Dieser Grenzwert ist die **Wahrscheinlichkeit** von  $A$ . Er ist der Limes von  $f(A)$  falls  $n$  gegen unendlich strebt:

$$P(A) = \lim_{n \rightarrow \infty} f(A) = \lim_{n \rightarrow \infty} \frac{k}{n}.$$

**Definition :** Eine Wahrscheinlichkeit ist eine Funktion die jedem Ereignis  $A$  einer Grundmenge  $\Omega$  eine reelle Zahl  $P(A)$  so zuweist, dass folgende drei **Axiome** gelten:

- I.  $P(A) \geq 0$  für alle  $A$ ,  $A \subseteq \Omega$
- II.  $P(\Omega) = 1$
- III.  $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$

Aus dieser Definition folgen untenstehende Sätze

1.  $P(\emptyset) = 0$
2.  $P(\bar{A}) = 1 - P(A)$
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4.  $P(A) \leq 1$
5.  $A \subseteq B \implies P(A) \leq P(B)$
6. Falls die  $n$  Ausgänge eines Experiments alle die gleich Wahrscheinlichkeit  $1/n$  haben sagen wir, sie seien gleichwahrscheinlich. Falls ein Ereignis  $A$  die Vereinigung von  $k$  gleichwahrscheinlichen Ausgängen ist, gilt

$$P(A) = \frac{k}{n} = \frac{\text{Anzahl der "günstigen" Fälle}}{\text{Anzahl der Fälle}}$$

#### 7.3.1 Bedingte Wahrscheinlichkeit und Unabhängigkeit

In diesem Paragraph wird der Begriff der bedingten Wahrscheinlichkeit und der Unabhängigkeit eingeführt. Das Konzept der bedingten Wahrscheinlichkeit ist wichtig falls Wahrscheinlichkeiten von Ereignissen berechnet werden, von denen gewisse Teilinformationen bekannt sind. In solch einer Situation wird durch den Begriff der bedingten Wahrscheinlichkeit dem Wissen um diese Information Rechnung getragen.

Das Konzept der Unabhängigkeit formalisiert den Umstand dass zwei Ereignisse, die nichts miteinander zu tun haben, zusammen (oder einzeln) auftreten können, ohne dass die Präsenz des einen Ereignisses die Wahrscheinlichkeit der Präsenz des anderen beeinflusst.

### 7.3.1.1 Bedingte Wahrscheinlichkeit

**Intuitiver Zugang.** Wir nehmen an, dass das Würfeln zweier unterscheidbarer Würfel alle 36 Elementarereignisse mit gleicher Wahrscheinlichkeit produziert, nämlich  $\frac{1}{36}$ . Nehmen wir nun an, wir wissen dass der erste Würfel eine 3 zeigt. Wie gross ist nun, im Wissen um diese zusätzliche Information, die Wahrscheinlichkeit, dass die Summe der beiden 8 ergibt?

Um diese Wahrscheinlichkeit auszurechnen kann wie folgt vorgegangen werden: der erste Würfel zeigt 3 Augen. Es kommen also genau folgende 6 Elementarereignisse in Frage:  $(3, 1)$ ,  $(3, 2)$ ,  $(3, 3)$ ,  $(3, 4)$ ,  $(3, 5)$  und  $(3, 6)$ . Da jede dieser Kombinationen anfänglich die gleiche Wahrscheinlichkeit hatte gewürfelt zu werden, haben sie auch nach dem Würfeln der 3 die gleiche Chance gewürfelt zu werden. Anders gesagt, gegeben dass der erste Wurf eine 3 bringt, ist die (bedingte) Wahrscheinlichkeit der Ereignisse  $(3, 1)$ ,  $(3, 2)$ ,  $(3, 3)$ ,  $(3, 4)$ ,  $(3, 5)$  und  $(3, 6)$  für alle  $\frac{1}{6}$ , während die (bedingte) Wahrscheinlichkeit aller anderen 30 Ereignisse 0 ist. Die gesuchte Wahrscheinlichkeit ist also  $\frac{1}{6}$ .

Steht nun  $E$  respektive  $F$  für das Ereignis “die Summe der Augen ist 8” und “der erste Wurf gibt 3”, so ist die obige Wahrscheinlichkeit ein Beispiel für die bedingte Wahrscheinlichkeit von  $E$  mit dem Zusatzwissen  $F$ . Sie wird mit dem Ausdruck  $P(E|F)$  bezeichnet.

#### Verallgemeinerung.

Wir lassen uns durch obiges Beispiel leiten, um einen Ausdruck  $P(E|F)$  zu erhalten für beliebige Ereignisse  $E$  und  $F$ : falls  $F$  eingetroffen ist, wird  $E$  genau dann eintreffen, falls beide Ereignisse  $E$  und  $F$  eintreffen. Anders gesagt, wir suchen die Wahrscheinlichkeit von  $E \cap F$ . Da wir aber wissen, dass  $F$  eingetroffen ist, wird dieses Ereignis die neue (reduzierte) Fundamentalmenge sein. Es folgt, dass die bedingte Wahrscheinlichkeit von  $E$  bei Präsenz von  $F$  gegeben ist durch den Vergleich der (unbedingten) Wahrscheinlichkeiten der Ereignisse  $E \cap F$  und  $F$ . Daraus leiten wir folgende Definition ab

**Für  $P(F) > 0$  ist die bedingte Wahrscheinlichkeit von  $E$  gegeben  $F$**

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

**Beispiel 19** Eine Münze wird zweimal geworfen. Nehmen wir an, dass die vier möglichen Ausgänge  $\Omega = \{(k, k), (k, z), (z, k), (z, z)\}$  gleichwahrscheinlich sind, wie gross ist die bedingte Wahrscheinlichkeit beide Male Kopf geworfen zu haben falls wir wissen, dass der erste Wurf Kopf ergab?

*Lösung.* Sei  $E = \{(k, k)\}$  das Ereignis “beide Würfe ergeben Kopf” und  $F = \{(k, k), (k, p)\}$  “der erste Wurf ergibt Kopf”. Die gesuchte Wahrscheinlichkeit ist

$$P(E|F) = \frac{P(\{(k, k)\})}{P(\{(k, k), (k, p)\})} = \frac{1/4}{2/4} = \frac{1}{2}.$$

◇

**Beispiel 20** Eine Urne enthält 8 rote Kugeln und 4 weisse Kugeln. Man zieht ohne Zurücklegen zwei Kugeln und geht davon aus, dass bei jeder Ziehung die möglichen Elementarereignisse gleichwahrscheinlich sind. Wie gross ist die Wahrscheinlichkeit, dass die beiden Kugeln rot sind?



*Lösung.* Wir nennen  $R_1$  respektive  $R_2$  die Ereignisse “die erste Kugel ist rot” und “die zweite ist rot”. Falls die erste Kugel rot ist, bleiben noch 7 rote Kugeln und 4 weisse Kugeln für die zweite Ziehung. Deshalb  $P(R_2|R_1) = \frac{7}{11}$ ; da  $P(R_1) = \frac{8}{12}$  ist, finden wir für die gesuchte Wahrscheinlichkeit

$$P(R_1 \cap R_2) = P(R_1)P(R_2|R_1) = \frac{2}{3} \frac{7}{11} = \frac{14}{33} = \left( \frac{\binom{8}{2}}{\binom{12}{2}} \right)$$

◇

### 7.3.1.2 Unabhängigkeit

Zwei Ereignisse  $E$  und  $F$  sind unabhängig falls gilt

$$P(E \cap F) = P(E)P(F) .$$

Gleichbedeutend:  $E$  und  $F$  sind unabhängig falls gilt

$$P(E | F) = P(E)$$

Das heisst, dass  $E$  und  $F$  unabhängig sind, falls die Information, dass sich  $F$  ereignet hat, die Wahrscheinlichkeit von  $E$  nicht ändert.

**Bemerkungen.** 1) Zwei Ereignisse  $E$  und  $F$ , die nicht unabhängig sind, sind abhängig.  
2) Es ist einfach zu zeigen, dass, falls  $A$  und  $B$  unabhängig sind, das auch auf  $A$  und  $\bar{B}$ ,  $\bar{A}$  und  $B$  und auf die Ereignisse  $\bar{A}$  und  $\bar{B}$  zutrifft. Zum Beispiel ist

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

somit

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

**Beispiel 21** Suppose we toss two fair dice. Let  $E_1$  denote the event that the sum of the dice is six and  $F$  denote the event that the first die equals four. Then

$$P(E_1 \cap F) = P(\{(4, 2)\}) = \frac{1}{36} \neq \frac{5}{36} \frac{1}{6} = P(E_1)P(F).$$

Hence,  $E_1$  and  $F$  are not independent. Intuitively, the reason for this is clear for if we are interested in the possibility of throwing a sum of six (with two dice), then we will be quite happy if the first die lands four because then we still have a possibility to get a total of six. ◇

**Beispiel 22** An Internet packet travels from its source to router 1, from router 1 to router 2, and from router 2 to its destination. If routers drop packets independently with probability  $p$ , what is the probability that a packet is successfully transmitted from its source to its destination?

**Solution.** A packet is successfully transmitted if and only if neither router drops it. Let  $D_i$ ,  $i = 1, 2$  denote the event that the packet is dropped by router  $i$  and let  $S$  be the event that the packet is successfully transmitted. Then  $S$  occurs if and only if the packet is not dropped by router 1 and it is not dropped by router 2. We can write this symbolically as

$$S = \bar{D}_1 \cap \bar{D}_2.$$

Since the problem tells us that  $D_1$  and  $D_2$  are independent, so are  $\bar{D}_1$  and  $\bar{D}_2$ . Hence

$$P(S) = P(\bar{D}_1 \cap \bar{D}_2) = P(\bar{D}_1)P(\bar{D}_2) = (1 - p)(1 - p).$$

◇

## 7.4 Aufgaben

**Übung 7** Bestimmen Sie die Fundamentalmenge  $\Omega$  des Zufallsexperiments, das durch zweimaliges, zufälliges Ziehen aus einer Grundmenge  $G$  besteht falls

- a) die zufällige Ziehung ohne Zurücklegen erfolgt.
- b) die zufällige Ziehung mit Zurücklegen erfolgt (i.e. nach dem ersten Ziehen wird das Element, sagen wir  $g \in G$ , wieder in die Grundmenge  $G$  zurückgelegt).

◊

**Übung 8** Wir betrachten die Ereignisse  $A = \{1, 2, 3\}$ ,  $B = \{4, 5, 6\}$ ,  $C = \{1, 3\}$  in  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Berechnen Sie  $A \cup B$ ,  $A \cap B$ ,  $A \cap C$  und  $\bar{A}$ .

◊

**Übung 9** Man werfe einen Würfel und betrachte die Ereignisse  $A = \{2\}$ ,  $B = \{2; 4; 6\}$ ,  $C = \{1; 2\}$ ,  $D = \{1; 2; 4; 6\}$ ,  $E = \{3; 5\}$  und  $\Omega$ .

Bestimmen Sie die Ereignisse  $B \cap C$ ,  $A \cup B$ ,  $B \cup C$ ,  $C \cap E$ ,  $\bar{A}$ ,  $B \setminus C$ ,  $\bar{A} \cup D$  und  $\bar{A} \cap D$ .

◊

**Übung 10** Seien  $A$  und  $B$  zwei Ereignisse; beschreibe und repräsentiere anschliessend in einem Venn Diagramm die Ereignisse

- (a)  $A$  aber nicht  $B$
- (b) weder  $A$  noch  $B$
- (c)  $A$  oder  $B$ , aber nicht beide gleichzeitig.

◊

**Übung 11** Bestimmen Sie die Grundmenge  $\Omega$  folgender Zufallsexperimente

- a) Man werfe zwei Geldstücke ;      b) Man werfe zwei Würfel ;
- c) Man werfe ein Geldstück so oft, bis Zahl fällt.

◊

**Übung 12** Man werfe ein Geldstück und einen Würfel.

- 1) Geben Sie  $\Omega$  an.
- 2) Beschreiben Sie folgende Ereignisse.
  - ◊  $A =$  “Kopf und eine gerade Zahl treten auf” ;
  - ◊  $B =$  “eine Primzahl wird geworfen” ;
  - ◊  $C =$  “Zahl und eine ungerade Augenzahl werden geworfen”.

3) Beschreiben Sie explizit folgende Ereignisse.

◇  $A$  und  $B$  ereignen sich; ◇  $B$  und  $C$  ereignen sich; ◇ nur  $B$  ereignet sich.

4) Gibt es zwischen  $A$ ,  $B$  und  $C$  inkompatible Ereignisse?

⊙

**Übung 13** a) Man werfe zwei mal hintereinander eine Münze. Wieviele mögliche Ausgänge und wie viele mögliche Ereignisse gibt es?

b) Ein Zufallsexperiment habe drei mögliche Ausgänge. Wie viele mögliche Ereignisse gibt es?

c) Allgemein, ein Zufallsexperiment habe  $n$  Ausgänge, wie viele Ereignisse gibt es?

⊙

**Übung 14** Beweisen Sie diese Aussagen.

⊙

**Übung 15** In einer Schule mit 400 Schülern lernen 300 Englisch, 160 Spanisch und 120 gleichzeitig Spanisch und Englisch. Falls zufällig ein Schüler der Schule herausgegriffen wird, wie gross ist die Wahrscheinlichkeit, dass er

a) Englisch aber nicht Spanisch lernt?

b) Englisch oder Spanisch lernt?

c) weder Englisch noch Spanisch lernt?

⊙

**Übung 16** Wir betrachten Ereignisse  $A$  und  $B$ , so dass  $P(A \cap B) = \frac{1}{4}$ ,  $P(\bar{A}) = \frac{1}{3}$  und  $P(B) = \frac{1}{2}$ . Berechnen Sie  $P(A \cup B)$ ,  $P(\bar{A} \cap \bar{B})$ ,  $P(B \setminus A)$  und  $P(A \cup \bar{B})$ .

⊙

**Übung 17** Bestimmen Sie die Wahrscheinlichkeit folgender Ereignisse.

a) Eine gerade Zahl wird mit einem normalen Würfel geworfen.

b) Kopf erscheint mindestens einmal bei dreimaligem Werfen eines Geldstückes.

c) Eine rote Kugel wird zufällig aus einem Sack von vier weissen, drei roten und 5 blauen Kugeln gezogen.

⊙

**Übung 18** Zwei Würfel werden geworfen. Berechnen Sie

a) die Wahrscheinlichkeit eine 3 und eine 5 zu würfeln?

b) die Wahrscheinlichkeit zwei mal die 3 zu würfeln?

c) die Wahrscheinlichkeit eine gerade Summenzahl zu werfen?

d) die Wahrscheinlichkeit zwei gleiche Zahlen zu werfen?

e) die Wahrscheinlichkeit ein Total von 8 zu werfen?

f) die Wahrscheinlichkeit mindestens eine 6 zu werfen?

g) die Wahrscheinlichkeit keine 6 zu werfen?

⊙

**Übung 19** Wir nehmen an, dass ein medizinischer Test für einen spezifischen Krebs eine Sicherheit von 95% aufweist. Das heisst, von hundert Krebserkrankten werden im Mittel 95 entdeckt und von hundert gesunden Getesteten werden im Mittel bei 5 fälschlicherweise Krebs diagnostiziert. Wir nehmen nun an, dass 0,4% der Bevölkerung unter diesem Krebs leiden. Wie gross ist die Wahrscheinlichkeit, dass eine Person Krebs hat, wenn ihr Test positiv ausfällt?  $\odot$

**Übung 20** Mit den gleichen Bezeichnungen wie oben sei  $E_2$  das Ereignis "Summe der Augen der beiden Würfel= 7." Sind  $E_2$  und  $F$  unabhängig? (Antwort: ja).  $\odot$

# Kapitel 8

## Zufallsvariablen

Nach der Realisierung eines Zufallsexperiments wird oft nicht nach dem Ausgang sondern nach einer Funktion, die vom Ausgang abhängt, gefragt.

**Beispiel 23** Bei der Stichprobe von gedrehten Bolzen liegt als Resultat ein zufällig gewählter Bolzen  $\omega$  vor. Nun interessiert uns aber zum Beispiel nur der Durchmesser des Bolzens und nicht der ganze Bolzen als solcher.

Wenn  $X$  die Vorschrift “Durchmesser von” bedeutet, so ist  $X(\omega)$  also der Durchmesser des zufällig gezogenen Bolzens  $\omega$ . Somit ist  $X$  eine wohldefinierte Funktion dessen Argument zufällig gewählt wurde (also in der Fundamentalmenge  $\Omega$  liegt). Dies ist ein Beispiel einer Zufallsvariablen.  $\diamond$

**Definition 1** Eine Zufallsvariable  $X$  eines Zufallsexperiments ist eine Funktion mit Werten in  $\mathbb{R}$ , die auf der Grundmenge  $\Omega$  definiert ist

$$X : \Omega \rightarrow \mathbb{R}$$

Diese Funktionen werden oft mit den Namen  $X, Y, Z$  etc. belegt. Wir unterscheiden dabei zwei Arten von Zufallsvariablen:

- (i) **Diskrete Zufallsvariable.** Kann die Zufallsvariable  $X$  nur eine endliche oder abzählbar unendliche Menge von Werten in  $\mathbb{R}$  annehmen, so nennen wir  $X$  eine diskrete Zufallsvariable.
- (ii) **Stetige Zufallsvariable.** Kann die Zufallsvariable  $X$  alle Werte in einem Teilintervall von  $\mathbb{R}$  annehmen ( $\mathbb{R}$  eingeschlossen), so nennen wir  $X$  eine stetige Zufallsvariable.

**Beispiel 24** Wir haben in Beispiel 17 bemerkt, dass die zufällige Wahl eines Bolzens  $\omega \in \Omega$  ein Zufallsexperiment ist. Wird nun  $X$  als der Durchmesser von  $\omega$  definiert, so ist  $X$  eine stetige Zufallsvariable (der Durchmesser wird je nach  $\omega$  beliebige Werte um den Nominalwert des Durchmesseres annehmen können).

Definieren wir  $\omega \mapsto X(\omega)$  als eine Funktion die gleich 1 ist für einen toleranzhaltigen Bolzen  $\omega$  und gleich 0 für einen nicht toleranzhaltigen Bolzen, so ist  $X$  eine diskrete Zufallsvariable.



Da der Wert von  $X$  durch den Ausgang des Zufallsexperiments (also die Wahl von  $\omega$ ) perfekt bestimmt ist, ist es möglich den Werten von  $X$  die Wahrscheinlichkeit ihres Auftretens zuzuordnen. Diese Angabe nennen wir **die Wahrscheinlichkeitsverteilung** (kurz Verteilung) von  $X$ . Wir unterscheiden dabei zwei Arten von Verteilungen:

- (i) **Diskrete Verteilung.** Eine diskrete Zufallsvariable  $X$  besitzt eine diskrete Verteilung. Wir schreiben

$$P(X = x)$$

für die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  den Wert  $x$  annimmt. Wir kennen das Wahrscheinlichkeitsgesetz von  $X$  falls wir  $P(X = x)$  für alle Werte  $x$  kennen.

- (ii) **Stetige Verteilung.** Eine stetige Zufallsvariable  $X$  besitzt eine Wahrscheinlichkeitsdichte  $f$  so dass gilt

$$P(a < X < b) = \int_a^b f(x) dx$$

für die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  Werte im Intervall  $[a, b]$  annimmt. Wir kennen das Wahrscheinlichkeitsgesetz von  $X$  falls wir  $P(a < X < b)$  für alle Werte  $a, b$  kennen.

**Kennen wir die verschiedenen Werte von  $X$  zusammen mit den Wahrscheinlichkeiten ihres Auftretens, dann ist vom Standpunkt der Wahrscheinlichkeitstheorie alles von der Zufallsvariablen  $X$  bekannt.**

Die konkrete Bestimmung dieser Wahrscheinlichkeiten ist keinesfalls trivial. Für die statistische Anwendung ist zu bemerken, dass die Verteilungen theoretische Annäherungen an die Histogramme (siehe Histogramme und relative Häufigkeiten im Kapitel 3) sind.

Es ist deshalb verständlich, dass die Wahrscheinlichkeiten als Annäherung an relative Häufigkeiten Zahlen grösser gleich 0 sind. Genau so wie die Fläche unter dem Histogramm ist die Fläche unter der stetigen Verteilung bzw. die totale Länge der Picks der diskreten Verteilung in Abbildung 8.1 gleich 1. Dies ist nichts anderes als eine Normalisierung. Es gilt also für eine beliebige Menge  $B \subseteq \mathbb{R}$ :

$$0 \leq P(X \in B) \leq 1 \quad \text{und} \quad P(X \in \mathbb{R}) = 1.$$

## 8.1 Beispiele von Zufallsvariablen

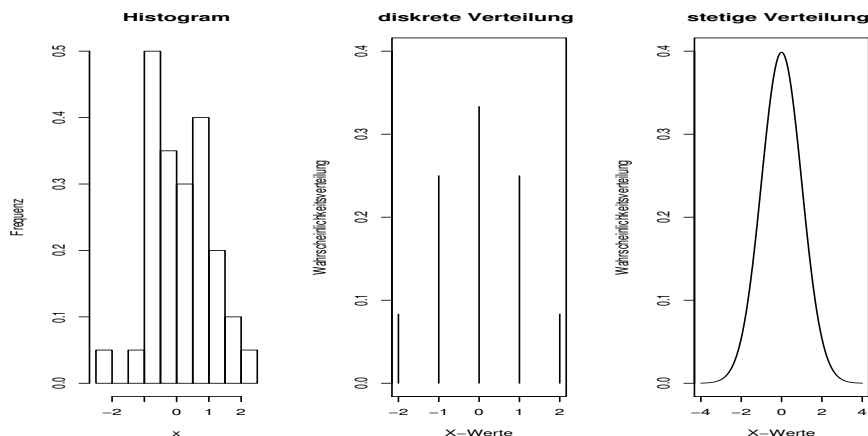


Abbildung 8.1: Links: Histogramm aus Messreihe. Mitte: diskrete Verteilung. Rechts: stetige Verteilung.

**Beispiel 25** Unser Experiment bestehe aus dem Werfen einer Münze. Sei  $X$  die Zufallsvariable, die dem Ausgang Kopf die Zahl 1 und dem Ausgang Zahl die Zahl 0 zuordnet.  $X$  ist eine Zufallsvariable mit möglichen Werten 0, 1 und zugehörigen Wahrscheinlichkeiten

$$P(\{X = 0\}) = P(\{X = 1\}) = \frac{1}{2}$$

Da nach Ausführen des Experiments die Zufallsvariable  $X$  genau einen ihrer möglichen Werte 0 oder 1 annimmt gilt

$$P\left(\bigcup_{i=0}^1 \{X = i\}\right) = P(X = 0 \text{ oder } X = 1) = \frac{1}{2} + \frac{1}{2} = 1.$$

◇

**Beispiel 26** Schauen wir uns nochmal den Kontext von Beispiel 1 an und definieren die Zufallsvariable  $X$  wie folgt:  $X(\omega) = 1$  falls der Bolzendurchmesser des Bolzens  $\omega$  toleranzhaltig ist und  $X(\omega) = 0$  falls nicht.  $X$  ist eine diskrete Zufallsvariable mit Werten 0 und 1 und

$$P(X = 1) = p, \text{ und } P(X = 0) = 1 - p.$$

wobei  $p$  eine unbekannte Zahl zwischen 0 und 1 ist. Auch hier gilt, dass nach Ausführen des Experiments die Zufallsvariable  $X$  genau einen ihrer möglichen Werte 0 oder 1 annimmt

$$P\left(\bigcup_{i=0}^1 \{X = i\}\right) = P(X = 0 \text{ oder } X = 1) = 1.$$

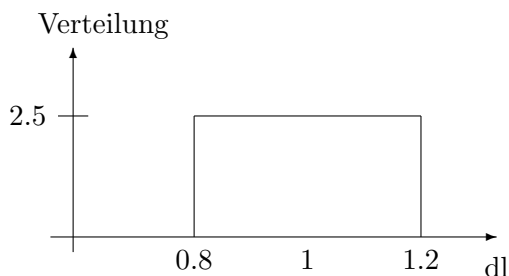
◇

**Beispiel 27** Unser Experiment bestehe nun aus dem Werfen von drei Münzen. Sei  $Y$  die Zufallsvariable, die die Anzahl von Kopfergebnissen registriert.  $Y$  ist eine Zufallsvariable mit möglichen Werten 0, 1, 2 und 3 und zugehörigen Wahrscheinlichkeiten

$$\begin{aligned} P(\{Y = 0\}) &= P(\{(z, z, z)\}) = \frac{1}{8} \\ P(\{Y = 1\}) &= P(\{(z, z, k), (z, k, z), (k, z, z)\}) = \frac{3}{8} \\ P(\{Y = 2\}) &= P(\{(z, k, k), (k, z, k), (k, k, z)\}) = \frac{3}{8} \\ P(\{Y = 3\}) &= P(\{(k, k, k)\}) = \frac{1}{8} \end{aligned}$$

◇

**Beispiel 28** Sei  $X$  die Zufallsvariable, die die Füllmenge einer Kaffeetasse registriert, die von einer bestimmten Kaffeemaschine geliefert wird. Wir modellieren  $X$  als eine stetige Zufallsvariable mit möglichen Werten im Intervall  $[0.8 \text{ dl}; 1.2 \text{ dl}]$ , und wir nehmen an, alle Werte seien gleichwahrscheinlich. Der Graph der Wahrscheinlichkeitsdichte ist entsprechend.



Die Höhe 2.5 ist natürlich nicht als Wahrscheinlichkeit zu interpretieren. Vielmehr gibt die Dichtefunktion eine Intensität wieder. Wollen wir die Wahrscheinlichkeit kennen, die Tasse beinhalte weniger als einen Deziliter, so müssen wir  $P(X \leq 1)$  ausrechnen. ◇

Bevor wir spezifisch auf diskrete und stetige Verteilungen eingehen, werden wir noch einen weiteren Begriff einführen, der für beide Klassen von Zufallsvariablen zusammen definiert wird. Er entspricht den kumulierten Häufigkeitsdiagrammen aus Kapitel 3.

## 8.2 Verteilungsfunktion

**Definition 2** Die Verteilungsfunktion  $F_X$  einer Zufallsvariable  $X$  ist für alle reellen Zahlen wie folgt definiert:

$$\begin{array}{l} F_X : \mathbb{R} \rightarrow [0, 1] \\ x \mapsto F_X(x) = P(\{X \leq x\}) \end{array}$$

Anders gesagt,  $F_X(x)$  ist die Wahrscheinlichkeit, dass die Zufallsvariable  $X$  Werte kleiner gleich  $x$  annimmt. Es müssen also alle Wahrscheinlichkeiten “kumuliert” werden bis und mit der Wahrscheinlichkeit von “ $X = x$ ”.

**Beispiel 29** Sei  $X$  die Zufallsvariable aus obigem Beispiel 27. Der Graph von  $F_X$  finden Sie in Abbildung 8.2. ◇

**Beispiel 30** Sei  $X$  die stetige Zufallsvariable aus obigem Beispiel 28. Der Graph von  $F_X$  finden Sie in Abbildung 8.3. ◇



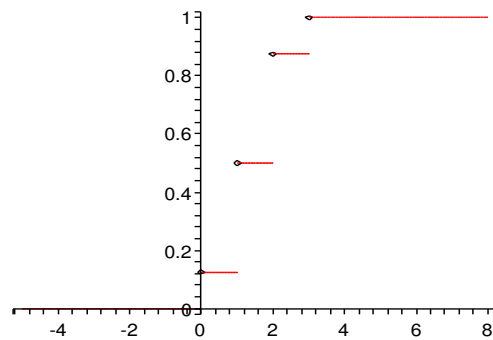


Abbildung 8.2: Verteilungsfunktion der Zufallsvariablen aus Beispiel 27.  
Verteilungsfunktion

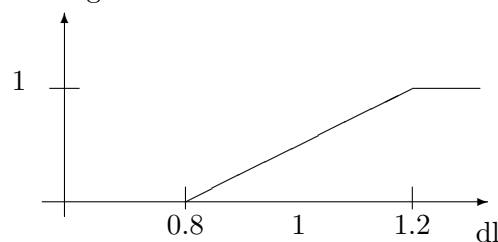


Abbildung 8.3: Verteilungsfunktion der Zufallsvariablen aus Beispiel 28.

Ganz allgemein sieht man, dass stetige Zufallsvariablen eine stetige Verteilungsfunktion besitzen, diskrete jedoch eine unstetige. Die Höhe einer Unstetigkeitsstelle bei  $x$  entspricht dabei gerade der Wahrscheinlichkeit  $P(X = x)$ .

**Eigenschaften von Verteilungsfunktionen.** Sei  $X$  eine Zufallsvariable mit Werten in  $\mathbb{R}$ . Hier einige allgemeine Eigenschaften der Verteilungsfunktion  $F_X$ :

- (i)  $F_X$  ist eine steigende Funktion. Anders gesagt, falls  $a \leq b$  dann ist  $F_X(a) \leq F_X(b)$ .
- (ii)  $\lim_{x \rightarrow \infty} F_X(x) = 1$ ,
- (iii)  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ ,

Es genügt für diese Eigenschaften ein intuitives Verständnis zu haben. Die erste Eigenschaft folgt aus dem Umstand, dass für  $a < b$  gilt: falls für ein  $\omega$  die Ungleichung  $\{X(\omega) \leq a\}$  gilt, ist natürlich automatisch  $\{X(\omega) \leq b\}$  erfüllt. Die Wahrscheinlichkeit des ersten Ereignisses ist also kleiner oder gleich die des zweiten. Was die anderen Eigenschaften betrifft sei angemerkt, dass das Ereignis  $\{X < x\}$  sich dem sicheren Ereignis  $\{X < \infty\} = \Omega$  annähert (resp. gegen das unmögliche Ereignis  $\emptyset$ ) falls  $x$  gegen Unendlich strebt (resp. gegen  $-\infty$ ).

**Verteilungsfunktionen und Wahrscheinlichkeiten.** Alle wahrscheinlichkeitstheoretischen Betrachtungen, die  $X$  betreffen, können mit Hilfe der Verteilungsfunktion behandelt werden. Es gilt,

$$P(a < X \leq b) = F(b) - F(a), \quad \text{für alle } a < b$$

Wir bemerken, dass die Werte  $P(X < b)$  und  $F(b)$  im Allgemeinen voneinander verschieden sind, da letzterer Wert auch die Wahrscheinlichkeit  $P(\{X = b\})$  beinhaltet.

### 8.3 Diskrete Zufallsvariablen

**Wir erinnern:** Eine Zufallsvariable die nur eine abzählbare Anzahl von verschiedenen Werten annehmen kann, nennen wir eine **diskrete Zufallsvariable**.

**Schreibweise.** Für eine diskrete Zufallsvariable  $X$  definieren wir die **Wahrscheinlichkeitsverteilung**  $p$  (oder Wahrscheinlichkeitsgesetz, oder Verteilung) durch

$$p(x) = P(X = x)$$

Diese Verteilung kann nur für eine höchstens abzählbare Grundmenge strikt positiv sein. Anders gesagt, falls  $X$  die Werte  $x_1, x_2, \dots$ , annehmen kann, ist

$$\begin{aligned} p(x_i) &\geq 0 & i = 1, 2, \dots \\ p(x) &= 0 & \text{für alle anderen Werte von } x \end{aligned}$$

Da  $X$  aber einen dieser Werte  $x_i$  mit Sicherheit annehmen muss, gilt

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

*Die Zufallsvariable  $X$  ist durch Angabe der möglichen Werte  $\{x_1, \dots, x_n, \dots\}$  sowie den zugehörigen Wahrscheinlichkeiten  $p(x_1), p(x_2), \dots, p(x_n), \dots$  vollständig definiert.*

Umgekehrt kann durch die Angabe einer Zahlenfolge  $p_1, p_2, \dots, p_n, \dots$  mit den Eigenschaften  $p_i \geq 0$  und  $\sum p_i = 1$  eine Zufallsvariable definiert werden. Man setzt zum Beispiel den Wertebereich von  $X$  gleich  $\mathbb{N}$  und definiert  $P(X = i) := p_i$ .

**Beispiel 31** Wir schauen uns ein Kopf und Zahl Spiel an. Sei  $X$  die Zufallsvariable die den Wert 1 annimmt, falls Kopf geworfen wird, und bei Zahl den Wert 0 hat. Das Wahrscheinlichkeitsgesetz von  $X$  ist:  $p(1) = P(X = 1)$  und  $p(0) = P(X = 0)$ . Bei einer unpräparierten Münze ist natürlich,

$$p(0) = p(1) = \frac{1}{2}.$$

◇

Es ist oft instruktiv die Wahrscheinlichkeitsverteilung aufzuzeichnen. Hierbei werden die Werte  $p(x_i)$  auf der Ordinate und die  $x_i$  auf der  $x$ -Achse abgetragen. Als Beispiel geben wir in Abbildung 8.4 (links) den Graphen der Verteilung

$$p(0) = \frac{1}{4}, \quad p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{4}.$$

Rechts in der Abbildung ist der Graph der Wahrscheinlichkeitsverteilung einer Zufallsvariable, die die Summe zweier Würfel zählt.

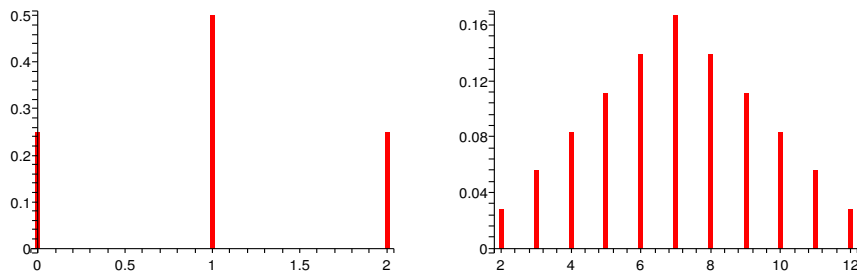


Abbildung 8.4: Wahrscheinlichkeitsverteilungen.

## 8.4 Erwartungswert

**Definition 3** Für eine diskrete Zufallsvariable  $X$  mit Wahrscheinlichkeitsverteilung  $p(\cdot)$ , definieren wir den Erwartungswert  $X$ , Bezeichnung  $E(X)$ , durch den Ausdruck

$$E(X) = \sum_x xp(x).$$

Dabei wird über die möglichen Werte von  $X$  summiert. Konkret: **der Erwartungswert von  $X$  ist das gewichtete Mittel der Werte, die  $X$  annehmen kann.** Die Gewichte entsprechen dabei den Wahrscheinlichkeiten, dass diese Werte angenommen werden.

Ist zum Beispiel die Wahrscheinlichkeitsverteilung von  $X$  gegeben durch

$$p(0) = \frac{1}{2} = p(1)$$

dann ist

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

nichts anderes als das arithmetische Mittel der Werte 0 und 1, die  $X$  annehmen kann. Falls jedoch

$$p(0) = \frac{1}{3} \quad p(1) = \frac{2}{3}$$

dann ist

$$E(X) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

das gewichtete Mittel der Werte 0 und 1. Der Wert 1 hat dabei ein doppelt so hohes Gewicht wie 0, da  $p(1) = 2p(0)$  gilt.

**Erwartungswert einer Funktion einer Zufallsvariablen.** Sei  $X$  eine diskrete Zufallsvariable mit gegebener Wahrscheinlichkeitsverteilung. Es kommt vor, dass wir den Erwartungswert einer Funktion von  $X$ , sagen wir  $g(X)$  ausrechnen müssen. Wie ist das Vorgehen? Es ist zu bemerken, dass  $g(X)$  eine diskrete Zufallsvariable ist, deren Verteilung aus der von  $X$  abgeleitet werden kann. Ist die Verteilung von  $g(X)$  bekannt, genügt es, die Definition des Erwartungswerts anzuwenden, um  $E(g(X))$  zu erhalten.

**Beispiel 32** Sei  $X$  eine Zufallsvariable die die Werte  $-1, 0, 1$  mit folgenden Wahrscheinlichkeiten annehmen kann:

$$P(\{X = -1\}) = 0.2, \quad P(\{X = 0\}) = 0.5, \quad P(\{X = 1\}) = 0.3.$$

Berechnen Sie  $E(X^2)$ .

*Lösung.* Sei  $Y = g(X) = X^2$ .  $Y$  nimmt die Werte 0 und 1 an. Die Verteilung von  $Y$  ist

$$\begin{aligned} P(Y = 1) &= P(X = -1) + P(X = 1) = 0.5 \\ P(Y = 0) &= P(X = 0) = 0.5 \end{aligned}$$

Also

$$E(X^2) = E(Y) = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5.$$

Es sei angemerkt, dass  $E(X)^2 \neq E(X^2)$  gilt. ◇

**Rechenvorschrift.** Obwohl mit obiger Prozedur im Prinzip der Erwartungswert einer Funktion von  $X$  immer ausgerechnet werden kann, gibt es eine andere Art  $E(g(X))$  zu berechnen. In der Tat, beachten wir, dass  $g(X)$  gleich  $g(x)$  ist, falls  $X$  gleich  $x$  ist, so ist es naheliegend anzunehmen, dass  $E(g(X))$  das gewichtete Mittel der Werte  $g(x)$  ist.  $g(x)$  wird dabei mit der Wahrscheinlichkeit gewichtet, dass  $X$  gleich  $x$  ist. Intuitiv ist somit folgendes Resultat einzusehen

**Theorem.** Ist  $X$  eine diskrete Zufallsvariable mit Werten  $x_i$ ,  $i \geq 1$ , und respektiver Verteilung  $p(x_i)$ , dann ist für jede reelle Funktion  $g$

$$E(g(X)) = \sum_i g(x_i)p(x_i).$$

Die Demonstration wird an der Tafel geführt. Es sei jedoch obiges Beispiel mit dieser Methode nochmals ausgerechnet:

$$E(X^2) = (-1)^2 0.2 + 0^2 0.5 + 1^2 0.3 = 0.5$$

was natürlich mit obigem Resultat übereinstimmt.

## 8.5 Varianz und Standardabweichung

**Definition 4** Die **Varianz** einer Zufallsvariable  $X$ , kurz  $V(X)$  (auch  $\text{Var}(X)$ ), ist die Grösse

$$V(X) = E[(X - \mu)^2]$$

wobei  $\mu$  der Erwartungswert von  $X$  bedeutet:  $\mu = E(X)$ .

**Definition 5** Die **Standardabweichung** von  $X$ , ist die Wurzel aus  $V(X)$  bezeichnet mit  $\sigma_X$ .

$$\sigma_X = \sqrt{V(X)}$$

## 8.6 Die wichtigsten diskreten Verteilungen

Die diskreten Zufallsvariablen (bzw. ihre Verteilungen) sind oft nach Kategorien geordnet und entsprechend benannt. In diesem Abschnitt werden wir die wichtigsten Typen kennenlernen.

### 8.6.1 Bernoulli-Variable.

Es liege ein Zufallsexperiment vor, dessen Resultat als Erfolg oder Misserfolg interpretiert werden kann. Man definiert dann eine Zufallsvariable  $X$  wie folgt: Falls das Resultat einen Erfolg liefert, ist der Wert von  $X$  gleich 1 ansonsten gleich 0. Die Verteilung von  $X$  ist

$$\begin{aligned} P(\{X = 0\}) &= 1 - p \\ P(\{X = 1\}) &= p \end{aligned}$$

wobei  $p$  die Erfolgswahrscheinlichkeit ist mit  $0 \leq p \leq 1$ . Solche Verteilungen wurden speziell vom Schweizer Mathematiker Jacques Bernoulli untersucht und tragen daher seinen Namen.

Eine Zufallsvariable  $X$  heisst **Bernoulli-Variable**, falls es eine Zahl  $p \in [0, 1]$  so gibt, dass die Verteilung von  $X$  gegeben ist durch

$$P(X = 1) = p = 1 - P(X = 0).$$

Die Verteilung selbst nennt man Bernoulli-Verteilung.

**Beispiel 33** Bedienen wir uns des Beispiels 1 und definieren wir die Zufallsvariable  $X$  mit Werten  $X(\omega) = 1$  falls der Bolzendurchmesser von  $\omega$  toleranzhaltig ist und  $X(\omega) = 0$  falls nicht.  $X$  ist eine Bernoulli-verteilte Zufallsvariable mit unbekanntem  $p$ . Das praktische Problem besteht darin, den wahren Wert von  $p$  auf Grund einer statistischen Erhebung abzuschätzen.  $\diamond$

### 8.6.2 Binomialverteilung.

Wir führen  $n$  unabhängige (d.h. sich nicht beeinflussende) Bernoulli Zufallsexperimente durch. Die zugehörigen Bernoulli Zufallsvariablen  $X_1, X_2, \dots, X_n$  haben alle die gleiche Wahrscheinlichkeit  $p$  für Erfolg (und Wahrscheinlichkeit  $1 - p$  für Misserfolg). Jede dieser Zufallsvariablen kann die Werte 0 oder 1 annehmen je nach dem, ob das entsprechende Experiment erfolgreich war oder nicht. Wir definieren nun die Summe der Zufallsvariablen  $X_i$ :

$$X = \sum_{i=1}^n X_i.$$

Per Definition zählt die Zufallsvariable  $X$  die Anzahl der erfolgreichen Stichproben aus insgesamt  $n$  Messungen (d.h.,  $n$  Realisierungen von Bernoulli-Experimenten).  $X$  kann also die Werte  $\{0, 1, 2, \dots, n\}$  annehmen.

Um die Wahrscheinlichkeitsverteilung von  $X$  abzuleiten bemerken wir, dass jede gegebene Sequenz der Länge  $n$  mit  $i$  Erfolgen und  $n - i$  Misserfolgen mit Wahrscheinlichkeit  $p^i(1 - p)^{n-i}$  eintritt (Dank der Unabhängigkeit der Stichproben). Aus dem Umstand, dass es genau  $\binom{n}{i}$  solcher Sequenzen gibt die  $i$  Erfolge und  $n - i$  Misserfolge beinhalten, folgt nun das Wahrscheinlichkeitsgesetz:

$$P(X = i) = p(i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n.$$

Eine Zufallsvariable  $X$  mit Werten in  $\{0, 1, 2, \dots, n\}$  ist binomialverteilt mit Parameter  $(n, p)$ , falls die Verteilung von  $X$  gegeben ist durch

$$P(X = i) = p(i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n.$$

Eine Bernoulli-Variable ist also nichts anderes als eine binomialverteilte Zufallsvariablen mit Parameter  $(1, p)$ .

**Beispiel 34** Es sei bekannt, dass die Schrauben eines gewissen Herstellers mit Wahrscheinlichkeit 0.01 fehlerhaft sind. Der Zustand einer Schraube sei unabhängig von anderen Schraubenzuständen. Nun ist der Hersteller bereit 10er Pakete genau dann zurückzunehmen, wenn mehr als eine Schraube fehlerhaft ist. Welcher Anteil an verkauften Paketen muss der Hersteller im Mittel zurückkaufen?

*Lösung.* Sei  $X$  die Anzahl der fehlerhaften Schrauben für ein gegebenes Paket.  $X$  ist eine binomialverteilte Zufallsvariable mit Parameter  $(10; 0.01)$ . Die Wahrscheinlichkeit ein Paket ersetzen zu müssen ist

$$1 - P(\{X = 0\}) - P(\{X = 1\}) = 1 - \binom{10}{0} 0.01^0 0.99^{10} - \binom{10}{1} 0.01^1 0.99^9 \approx 0.004$$

Es ist zu erwarten, dass 0.4% der Pakete zu ersetzen sind.  $\diamond$

**Beispiel 35** Ein Kommunikationssystem bestehe aus  $n$  Komponenten; jede von ihnen funktioniert unabhängig von allen anderen Komponenten mit Wahrscheinlichkeit  $p$ . Das Gesamtsystem funktioniert, wenn mindestens die Hälfte der Komponenten funktionieren.

Für welche Werte von  $p$  wird ein System bestehend aus 5 Komponenten öfter operationell sein als ein System bestehend aus 3 Komponenten?

*Lösung.* Die Anzahl der funktionierenden Komponenten ist eine binomialverteilte Zufallsvariable mit Parameter  $(n, p)$  die Wahrscheinlichkeit, dass ein System mit 5 Komponenten operationell ist beträgt

$$\binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + p^5.$$

Für ein System bestehend aus 3 Komponenten ist die Wahrscheinlichkeit

$$\binom{3}{2} p^2 (1-p) + p^3.$$

Es folgt, das System bestehend aus 5 Komponenten ist demjenigen aus 3 Komponenten vorzuziehen, falls

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3,$$

was nach Umformungen gleichbedeutend mit  $p > \frac{1}{2}$  ist.  $\diamond$

### 8.6.3 Poisson Zufallsvariable.

Eine Zufallsvariable  $X$  mit möglichen Werten  $0, 1, 2, \dots$  ist poissonverteilt mit Parameter  $\lambda$  falls es ein  $\lambda > 0$  gibt, so dass

$$p(i) = P(\{X = i\}) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \dots$$

Diese Gleichung definiert in der Tat eine Wahrscheinlichkeitsverteilung, da gilt:

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} e^{\lambda} = 1.$$

Die Poissonverteilung wurde durch Siméon Denis Poisson in einem Werk über Wahrscheinlichkeit und ihre Anwendung in Rechtsfragen eingeführt. Sein Buch wurde 1837 unter dem Titel *Recherche sur la probabilité des jugements en matière criminelle et en matière civile* herausgegeben.

**Anwendung: Approximation einer Binomialverteilung durch eine Poissonverteilung.**

Die poissonverteilten Zufallsvariablen haben einen grossen Anwendungsbereich. Speziell können mit ihrer Hilfe binomialverteilte Zufallsvariablen mit Parameter  $(n, p)$  angenähert werden für  $n$  gross und  $p$  klein, so dass  $np$  von mittlerer Grössenordnung ist. Um sich davon zu überzeugen, sei  $X$  eine binomialverteilte Zufallsvariable mit Parameter  $(n, p)$  und wir setzen

$$\lambda = np.$$

Es gilt

$$\begin{aligned} P(\{X = i\}) &= \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\dots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \end{aligned}$$

Sei nun  $n$  gross und  $\lambda$  von mittlerer Grösse

$$(1 - \lambda/n)^n \approx e^{-\lambda}, \quad \frac{n(n-1)\dots(n-i+1)}{n^i} \approx 1, \quad (1 - \lambda/n)^i \approx 1$$

Es folgt

$$P(\{X = i\}) \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Nachfolgend einige Anwendungsbereiche wo Poissonverteilungen auftauchen.

- (i) Anzahl von Druckfehlern pro Druckseite.
- (ii) Anzahl der Personen älter als 100 Jahre in einer vorgegebenen Gessellschaft.
- (iii) Anzahl der falsch gewählten Telefonnummern pro Tag.
- (iv) Anzahl der Hundebiscuitpakete, die pro Tag in einem gegebenen Verkaufszenter verkauft werden.
- (v) Anzahl der Kunden pro Tag, die ein gegebenes Postbüro besuchen.
- (vi) Anzahl der emittierten  $\alpha$  Partikel, die von einem radioaktiven Material in einer gewissen Zeit abgestrahlt werden.

In all diesen Beispielen – und in vielen anderen – ist die Zufallsvariable immer nur annähernd poissonverteilt. Grund: es wird dadurch eine Binomialverteilung angenähert.

**Beispiel 36** Nehmen wir an, die Anzahl Tippfehler pro Seite in diesem Skript folge einer Poissonverteilung mit Parameter  $\lambda = \frac{1}{2}$ . Berechnen Sie die Wahrscheinlichkeit, dass mindestens ein Fehler auf dieser Seite vorhanden ist.

*Lösung.* Sei  $X$  die Anzahl Fehler auf dieser Seite. Es ist

$$P(\{X \geq 1\}) = 1 - P(\{X = 0\}) = 1 - e^{-1/2} \approx 0.393.$$

◇

**Beispiel 37** Nehmen wir an, die Fehlerwahrscheinlichkeit eines Fabrikationsgegenstandes sei 0.1. Finden Sie die Wahrscheinlichkeit, dass ein Los von 10 dieser Objekte höchstens ein fehlerhaftes Element enthält.

*Lösung.* Die Anzahl der fehlerhaften Objekte im Los folgt einer Binomialverteilung mit Parameter  $(10, 0.1)$ . Die exakte Wahrscheinlichkeit ist demnach

$$\binom{10}{0} 0.1^0 0.9^{10} + \binom{10}{1} 0.1^1 0.9^9 = 0.7361$$

wogegen die Poissonannäherung die Wahrscheinlichkeit  $e^{-1} + e^{-1} \approx 0.7358$  liefert.

◇

## 8.7 Stetige Zufallsvariablen

**Wir erinnern:** Eine Zufallsvariable die alle Werte in einem Teilintervall von  $\mathbb{R}$  (oder in ganz  $\mathbb{R}$ ) annehmen kann, nennen wir eine **stetige Zufallsvariable**. Genauer:

$X$  ist eine stetige Zufallsvariable falls es eine auf  $\mathbb{R}$  definierte, nicht negative, normierte Funktion  $f$  gibt so, dass für alle reellen Zahlenmengen  $B \subset \mathbb{R}$  gilt:

$$P(\{X \in B\}) = \int_B f(x) dx.$$

Die Funktion  $f$  heisst **Wahrscheinlichkeitsdichte** der Zufallsvariable  $X$ . Die obige Gleichung bedeutet: die Wahrscheinlichkeit, dass  $X$  einen Wert in der Menge  $B$  annimmt, ist durch die Integration der Wahrscheinlichkeitsdichte  $f$  über die Menge  $B$  gegeben. Da  $X$  mit Sicherheit einen Wert in  $\mathbb{R}$  annimmt, folgt die **Normierungsbedingung** an  $f$ :

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

Alle wahrscheinlichkeitstheoretischen Fragen können mit der Kenntnis von  $f$  im Prinzip beantwortet werden. Zum Beispiel für  $B = [a, b]$  folgt

$$P(\{a \leq X \leq b\}) = \int_a^b f(x) dx.$$

Für  $a = b$  folgt nun sofort

$$P(\{X = a\}) = \int_a^a f(x) dx = 0.$$



Das bedeutet, die Wahrscheinlichkeit einer stetigen Zufallsvariablen einen spezifischen Wert anzunehmen ist Null! Das mag überraschen, ist aber intuitiv dadurch zu verstehen, dass es viel zu viele Alternativen zum spezifischen Wert  $a$  gibt. Für stetige Variablen kann die Verteilungsfunktion  $F$  geschrieben werden als

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

**Definition 6** Für eine stetige Zufallsvariable  $X$  ist die **Wahrscheinlichkeitsverteilung** durch die Werte  $P(\{X \in B\})$  für beliebige  $B \subset \mathbb{R}$  gegeben.

In diesem Kurs beschränken wir uns auf Zufallsvariablen deren Verteilung  $P(\{X \in B\})$  mit Hilfe einer Wahrscheinlichkeitsdichte  $f$  berechnet werden kann. Es gilt also:

Die stetige Zufallsvariable  $X$  mit Werten in  $\mathbb{R}$  ist vollständig definiert durch Angabe der zugehörigen **Wahrscheinlichkeitsdichte**  $f(x)$ .

**Beispiel 38** Sei  $X$  eine stetige Zufallsvariable mit Verteilungsdichte

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{sonst} \end{cases}$$

Welchen Wert hat  $C$  und wie gross ist  $P(\{X > 1\})$ ?

*Lösung.* Da  $f$  eine Wahrscheinlichkeitsdichte ist, muss sie der Bedingung  $\int_{-\infty}^{\infty} f(x)dx = 1$  genügen. Das gibt

$$C \int_0^2 (4x - 2x^2)dx = C \left[ 2x^2 - \frac{2x^3}{3} \right]_{x=0}^{x=2} = 1$$

und ausgerechnet  $C = \frac{3}{8}$ . Also

$$P(\{x > 1\}) = \int_1^{\infty} f(x)dx = \frac{3}{8} \int_1^2 (4x - 2x^2)dx = \frac{1}{2}.$$

◇

**Beispiel 39** Die Lebensdauer eines Computers vor seiner ersten seriösen Panne ist eine stetige Zufallsvariable  $X$  mit Dichtefunktion

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

(a) Wie gross ist die Wahrscheinlichkeit, dass der Computer ohne seriöse Störung zwischen 50 und 150 Stunden arbeitet?

*Lösung.* Für  $\lambda = \frac{1}{100}$ , ist  $f$  eine Wahrscheinlichkeitsdichte, da  $f(x) \geq 0$  und

$$\int_{-\infty}^{\infty} f(x)dx = -100\lambda e^{-x/100} \Big|_0^{\infty} = 100\lambda = 1.$$

Die gesuchte Wahrscheinlichkeit ist demnach

$$P(\{50 < X < 150\}) = \int_{50}^{150} f(x)dx = -e^{-x/100} \Big|_{50}^{150} \approx 0.383.$$

◇

**Verteilungsfunktion einer stetigen Zufallsvariable.** Die Beziehung zwischen der Verteilungsfunktion  $F_X$  und der Dichte  $f$  einer stetigen Zufallsvariable  $X$  ist gegeben durch

$$F(a) = P(\{X \in ]-\infty, a]\}) = \int_{-\infty}^a f(x)dx$$

Die Ableitung bezüglich  $a$  ergibt dann

$$\frac{d}{da}F(a) = f(a).$$

Die Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen ist also die Ableitung der Verteilungsfunktion.

**Erwartungswert stetiger Zufallsvariablen.** Im Kapitel 7.5 definierten wir den Erwartungswert einer diskreten Zufallsvariable  $X$  durch

$$E(X) = \sum_x xP(\{X = x\}).$$

Ist  $X$  eine stetige Zufallsvariable mit Dichtefunktion  $f(x)$ , wird das Gewicht  $P(X = x)$  durch den Ausdruck

$$f(x)dx = P(\{x \leq X \leq x + dx\}), \text{ für } dx \text{ klein}$$

ersetzt. Es ist so einzusehen, dass die analoge **Definition des Erwartungswerts von  $X$**  wie folgt ist

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

**Beispiel 40** Berechnen Sie  $E(X)$  für die Dichtefunktion

$$f(x) = \begin{cases} 2x & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

*Lösung.*

$$E(X) = \int xf(x)dx = \int_0^1 2x^2dx = \frac{2}{3}.$$

◇

**Beispiel 41** Berechnen Sie  $E(e^X)$  für eine Dichtefunktion von  $X$

$$f(x) = \begin{cases} 1 & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

*Lösung.* Sei  $Y = e^X$ . Wir bestimmen zuerst die Verteilungsfunktion  $F_Y$  von  $Y$ . Für  $1 \leq x \leq e$ ,

$$\begin{aligned} F_Y(x) &= P(\{Y \leq x\}) = P(\{e^X \leq x\}) = P(\{X \leq \log(x)\}) \\ F_Y(x) &= \int_0^{\log(x)} f(y)dy = \log(x). \end{aligned}$$

Ableiten von  $F_Y(x)$  ergibt die Dichte von  $Y$

$$f_Y(x) = \frac{1}{x}, \quad 1 \leq x \leq e.$$

Also

$$E(e^X) = E(Y) = \int_{-\infty}^{\infty} x f_Y(x) dx = \int_1^e dx = e - 1$$

◇

**Rechenregel.** Wie im diskreten Fall existiert auch im stetigen Fall eine weitere Art den Erwartungswert einer Funktion von  $X$  auszurechnen. Das folgende Theorem ist eine Analogie des diskreten Falls.

**Theorem** Sei  $X$  eine stetige Zufallsvariable mit Dichte  $f(x)$ , dann gilt für jede reellwertige Funktion  $g$

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

Der Beweis dieses Resultats ist komplizierter als sein diskretes Analogon. Wir werden ihn hier nicht reproduzieren. Verifizieren wir seine Anwendbarkeit mit Hilfe obigen Beispiels:

$$E(e^X) = \int_{-\infty}^{\infty} e^x f(x) dx = \int_0^1 e^x \cdot 1 dx = e - 1.$$

**Varianz stetiger Zufallsvariablen.** Die Varianz stetiger Zufallsvariablen ist genau gleich definiert wie im diskreten Fall. Das heißt, **ist  $X$  eine stetige Zufallsvariable mit Erwartungswert  $\mu$ , dann ist die Varianz von  $X$  definiert als**

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

Die Formel

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

wird genau gleich bewiesen wie im diskreten Fall.

**Beispiel 42** Berechnen Sie  $V(X)$ , falls die Wahrscheinlichkeitsdichte von  $X$  wie folgt definiert ist

$$f(x) = \begin{cases} 2x & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

*Lösung.*

$$E(X^2) = \int x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}.$$

Also, da  $E(X) = \frac{2}{3}$  (siehe Beispiel 40) erhält man  $V(X) = \frac{1}{18}$ .

◇

Man kann zeigen, der Beweis ist dabei analog wie im diskreten zu führen, dass für Konstanten  $a$  und  $b$  gilt

$$V(aX + b) = a^2V(X).$$

## 8.8 Wichtige stetige Wahrscheinlichkeitsverteilungen

Die stetigen Zufallsvariablen sind in Kategorien geordnet und entsprechend benannt. Diese Sektion gibt die wichtigsten Verteilungen an.

### 8.8.1 Gleichverteilung

Gleichverteilung auf  $[0, 1]$

*Eine Zufallsvariable ist gleichverteilt auf dem Intervall  $[0, 1]$ , falls die zugehörige Verteilungsdichte folgende Form hat*

$$f(x) = \begin{cases} 1 & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{falls nicht} \end{cases}$$

**Gleichverteilung auf beliebigem endlichem Intervall.** Eine Zufallsvariable ist gleichverteilt auf  $[\alpha, \beta]$  falls seine Wahrscheinlichkeitsdichte durch folgenden Ausdruck gegeben ist

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{falls } \alpha \leq x \leq \beta \\ 0 & \text{falls nicht} \end{cases}$$

### 8.8.2 Normalverteilung.

Eine Zufallsvariable  $X$  ist **normalverteilt mit Parameter  $\mu$  und  $\sigma^2$** , falls die Dichte von  $X$  gegeben ist durch

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \quad -\infty < x < \infty.$$

Der Graph ist eine Glockenkurve mit vertikaler Symmetrieachse bei  $\mu$  (siehe Abbildung 8.5)

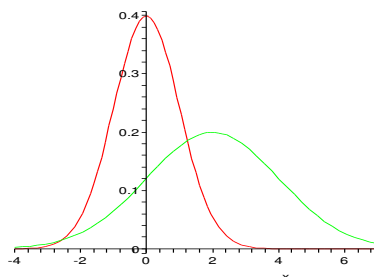


Abbildung 8.5: Dichte zweier Normalverteilungen mit  $\mu = 0, \sigma^2 = 1$  und  $\mu = 2, \sigma^2 = 4$ .

Die Normalverteilung wurde 1733 durch den französischen Mathematiker De Moivre eingeführt; er brauchte sie zur Annäherung von Wahrscheinlichkeiten binomial verteilter Zufallsvariablen mit grossem Parameter  $n$ . Dieses Approximationsverfahren wurde im Verlaufe der Theorie immer mehr entwickelt und ist heute unter dem Namen **Zentraler Grenzwertsatz** bekannt. Dieses Theorem, eines der wichtigsten überhaupt in der Wahrscheinlichkeitsrechnung, ist die Basis zum theoretischen Verständnis der empirischen Tatsache, dass viele Zufallsereignisse approximativ einer Normalverteilung genügen. Als Beispiel sei die Grösse einer zufällig ausgewählten Person genannt, oder die Geschwindigkeitskomponenten eines Gasmoleküls, oder der Messfehler einer physikalischen Grösse bei seiner experimentellen Bestimmung.

**Validierung der Definition.** Es muss zuerst gezeigt werden, dass  $f_X$  eine Wahrscheinlichkeitsdichte ist, d.h. zeige

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

Durch die Substitution  $y = \frac{x-\mu}{\sigma}$ , erhält man

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

und es bleibt zu zeigen, dass

$$\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}.$$

Dazu setzen wir

$$I := \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy.$$

Wir haben

$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy$$

Wenn wir zu Polarkoordinaten wechseln, können wir dieses Integral ausrechnen. Wir setzen  $x = r \cos(\theta)$ ,  $y = r \sin(\theta)$  und  $dy dx = r d\theta dr$ . Nun ist

$$I^2 = \int_0^{\infty} \int_0^{2\pi} e^{-r^2/2} r d\theta dr = 2\pi \int_0^{\infty} e^{-r^2/2} r dr = 2\pi.$$

Dies zeigt, dass  $I$  gleich  $\sqrt{2\pi}$  ist und beendet die Validierung der Definition.

Die Parameter  $\mu$  und  $\sigma$  haben eine natürliche Interpretation als Mittelwert und Standardabweichung (siehe Übung 27). Darüber hinaus ist es für das Verständnis von  $\sigma$  hilfreich, sich folgende Relation vor Augen zu führen (siehe Abbildung 8.6):

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\approx 0.6826 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\approx 0.9546 \end{aligned}$$

**Notation.** Wir schreiben

$$X \sim N(\mu, \sigma^2)$$

um zu sagen, dass  $X$  eine normalverteilte Zufallsvariable mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$  ist.

Die wichtige Konsequenz der Übung 37 ist, dass, wenn  $X$  eine normale Zufallsvariable mit Parameter  $\mu$  und  $\sigma^2$  ist, dann ist die Zufallsvariable

$$Z = (X - \mu)/\sigma$$

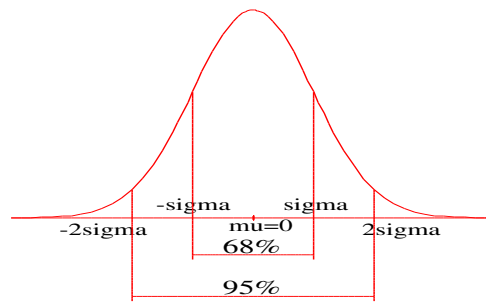


Abbildung 8.6: Die Fläche unter der Dichtefunktion zwischen  $\mu - \sigma$  und  $\mu + \sigma$  entspricht 68% der Gesamtfläche. Die Fläche unter der Dichtefunktion zwischen  $\mu - 2\sigma$  und  $\mu + 2\sigma$  ist bereits 95% der Gesamtfläche.

normalverteilt mit Parameter 0 und 1 ist. Eine solche Zufallsvariable nennen wir **standardisiert**. Es ist gebräuchlich die Verteilungsfunktion einer standardisierten, normalverteilten Zufallsvariablen mit  $\Phi$  zu bezeichnen. Das heißt,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Die Werte von  $\Phi(x)$  sind mit dem Computer auszurechnen oder einer Tabelle (siehe Anhang) zu entnehmen, da diese Funktion keine elementare Stammfunktion besitzt.

**Beispiel 43** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu = 3$  und  $\sigma^2 = 9$ . Berechnen Sie

- (a)  $P(2 < X < 5)$ , (b)  $P(X > 0)$  und (c)  $P(|X - 3| > 6)$

*Lösung.*

$$\begin{aligned} P(2 < X < 5) &= P\left(\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right) = P\left(-\frac{1}{3} < Z < \frac{2}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) = \Phi\left(\frac{2}{3}\right) - 1 + \Phi\left(\frac{1}{3}\right) \approx 0.38 \end{aligned}$$

$$\begin{aligned} P(X > 0) &= P\left(\frac{X-3}{3} > \frac{0-3}{3}\right) = P(Z > -1) \\ &= 1 - \Phi(-1) = \Phi(1) \approx 0.84 \end{aligned}$$

$$\begin{aligned} P(|X - 3| > 6) &= P(X > 9) + P(X < -3) \\ &= P\left(\frac{X-3}{3} > \frac{9-3}{3}\right) + P\left(\frac{X-3}{3} < \frac{-3-3}{3}\right) \\ &= P(Z > 2) + P(Z < -2) = 1 - \Phi(2) + \Phi(-2) \approx 0.046 \end{aligned}$$

◇

### 8.8.3 Approximation der Binomialverteilung durch eine Normalverteilung

Das untenstehende Theorem ist ein Spezialfall des bereits angetönten zentralen Grenzwertsatzes. Diese Version besagt, dass, wenn eine binomialverteilte Zufallsvariable mit Parameter  $n$  und  $p$  durch

Abziehen des Mittelwertes  $np$  und Division durch die Standardabweichung  $\sqrt{np(1-p)}$  "standardisiert" wird, die standardisierte Zufallsvariable (mit Mittelwert 0 und Varianz 1) approximativ und für grosse  $n$ , einer standardisierten Normalverteilung folgt.

**Theorem.** Sei  $S_n$  die Anzahl der Erfolge im Verlaufe von  $n$  Realisierungen von unabhängigen Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit  $p$ . Für alle  $a < b$  gilt

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \Phi(b) - \Phi(a)$$

falls  $n \rightarrow \infty$ .

Eine Demonstration wird im Unterricht gegeben. Wichtiger als die Demonstration ist die Bemerkung, dass zwei verschiedene Approximationen der Binomialverteilung vorgeschlagen wurden.

- (i) Die Poissonapproximation (gut für grosse  $n$  und falls  $np$  nicht extrem ist) und
- (ii) Die Approximation durch eine Normalverteilung. Sie ist zufriedenstellend falls  $np(1-p)$  gross genug ist (Faustregel  $np(1-p) > 10$ ) (siehe Abbildung 8.7).

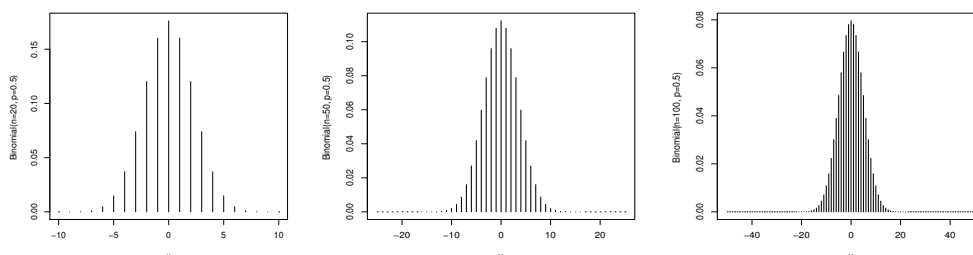


Abbildung 8.7: Die Binomialverteilung nimmt für grosse  $n$  die Form einer Glockenkurve an.

**Beispiel 44** Sei  $X$  die Zufallsvariable, die die Anzahl der Kopfwürfe in 40 Münzwürfen zählt. Es soll  $P(X = 20)$  berechnet werden, einmal exakt und einmal mit Hilfe der Normalapproximation.

*Lösung.* Da  $X$  eine diskrete und der Normaverteilung eine stetige Zufallsvariable zugrundeliegt, ist eine mögliche Annäherung an die gesuchte Wahrscheinlichkeit wie folgt

$$\begin{aligned} P(X = 20) &= P(19.5 < X < 20.5) \\ &= P\left(\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right) \\ &= P\left(-0.16 < \frac{X - 20}{\sqrt{10}} < 0.16\right) = \Phi(0.16) - \Phi(-0.16) \approx 0.1272 \end{aligned}$$

Das exakte Ergebnis ist  $\binom{40}{20} \left(\frac{1}{2}\right)^{40} \approx 0.1254$ .

◇

### 8.8.4 Exponentialverteilung.

Eine Zufallsvariable  $X$  mit folgender Dichtefunktion und positivem Parameter  $\lambda$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$$

ist eine exponentielle Zufallsvariable (oder exponentielle Verteilung) mit Parameter  $\lambda$ .

Die Verteilungsfunktion  $F_X$  einer exponentiellen Zufallsvariable ist

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}$$

und es sei bemerkt, dass  $F(\infty) = 1$  wie es sein muss.

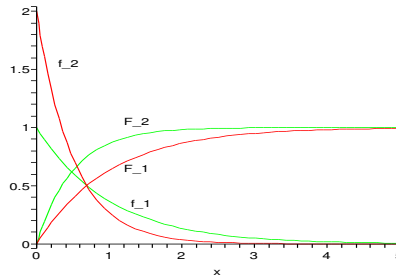


Abbildung 8.8: Dichten von Exponentialverteilungen  $f_1, f_2$  mit Verteilungsfunktion  $F_1$  und  $F_2$  für die Parameter  $\lambda = 1$  und  $\lambda = 2$ .

In der Praxis wird die Exponentialverteilung oft bei Problemen mit Wartezeiten angetroffen. So ist zum Beispiel die Zeit zwischen zwei Erdbeben exponentialverteilt, ebenso die Zeit zwischen zwei falschgewählten Telefonverbindungen...

**Beispiel 45** Es wird davon ausgegangen, dass die Länge eines Telefongesprächs (in Minuten) einer Exponentialverteilung folgt mit Parameter  $\lambda = \frac{1}{10}$ . Sie kommen an eine Kabine die aber gerade besetzt wird. Wie gross ist die Wahrscheinlichkeit das Sie (a) mehr als 10 Minuten warten müssen, (b) zwischen 10 und 20 Minuten warten müssen?

*Lösung.*  $X$  sei die Gesprächslänge der Person die vor Ihnen die Kabine betrat. Die gesuchten Wahrscheinlichkeiten sind

$$a) P(X > 10) = \int_{10}^{\infty} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{\infty} = e^{-1} \approx 0.368$$

$$b) P(10 < X < 20) = \int_{10}^{20} \frac{1}{10} e^{-x/10} dx = e^{-1} - e^{-2} \approx 0.233$$

◇



8.8.5  $\chi^2$ -Verteilung.

Eine Zufallsvariable  $X$  ist  $\chi^2$  (**Chi-Quadrat**) verteilt mit  $n$  **Freiheitsgraden**, falls  $X$  gegeben ist durch die Summe der Quadrate von  $n$  unabhängigen, standardisierten und normalverteilten Zufallsvariablen  $X_1, \dots, X_n$ :

$$X = X_1^2 + X_2^2 + \dots + X_n^2$$

Der Wertebereich von  $X$  ist also  $\mathbb{R}^+$  und die Dichtefunktion ist gegeben durch<sup>1</sup>

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \quad 0 < x < \infty.$$

Der Graph ist eine deformierte Glockenkurve mit Mittelwert  $\mu = n$  und Varianz  $\sigma^2 = 2n$  (siehe Abbildung 8.9). Die  $\chi^2$ -Verteilung hat eine besondere Stellung in der Statistik, da sie natürlich  $\chi^2$  Dichtefunktionen

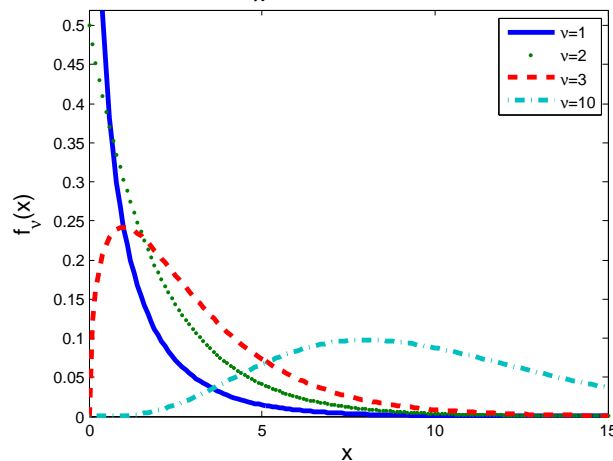


Abbildung 8.9: Dichte von vier  $\chi^2$ -Verteilungen mit  $\nu = 1$ ,  $\nu = 2$ ,  $\nu = 3$  und  $\nu = 10$ .

bei der Schätzung der Varianz auftritt. Zur Illustration stellen wir uns vor, dass wir  $n$  Messungen vornehmen wollen. Jede Messung  $X_i$  sei dabei normalverteilt mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ . Das Messresultat wird dann wie folgt angegeben:

$$\text{Mittel} \pm \text{Standardabweichung}$$

Für den Mittelwert nimmt man das arithmetische Mittel  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  und für die Varianz (bzw. Standardabweichung) nimmt man die empirische Varianz  $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  (bzw. die Wurzel davon). Im letzten Term kommt gerade eine Summe von Quadraten normaler Zufallsvariablen vor. Es folgt: die Zufallsvariable

$$Y = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S'^2}{\sigma^2}$$

<sup>1</sup> Die Funktion  $\Gamma$  (*Gamma*) ist eine Verallgemeinerung der Fakultätsfunktion “!”. Es gilt  $\Gamma(n) = (n-1)!$ . Sie erlaubt die Definition eines Ausdrucks wie  $(-\frac{1}{2})! = \sqrt{\pi}$ ,  $\frac{1}{2}! = \frac{\sqrt{\pi}}{2}$ ,  $\frac{3}{2}! = \frac{3\sqrt{\pi}}{2}$  oder  $\frac{5}{2}! = \frac{15\sqrt{\pi}}{8}$ . Es gilt:  $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$ . Einige Werte sind im Anhang aufgelistet.

ist  $\chi^2$ -verteilt mit  $n - 1$  Freiheitsgraden ( $n - 1$  deshalb weil der Mittelwert  $\bar{X}$  einen Freiheitsgrad einschränkt). Die Verteilung der empirischen Varianz ist also bekannt, falls von normalverteilten Zufallsvariablen ausgegangen wird und falls  $\sigma^2$  bekannt ist! Letzterer Fall ist eher unwahrscheinlich in den Anwendungen. Das ist der Grund für folgende Verteilung

### 8.8.6 $t$ -Verteilung (auch Student-Verteilung).

Eine Zufallsvariable  $T$  ist  **$t$ -verteilt mit  $\nu$  Freiheitsgraden**, falls  $T$  gegeben ist durch den Quotienten einer standardisierten, normalverteilten Zufallsvariable  $X$  und der Wurzel einer durch  $\nu$  geteilten,  $\chi^2$ -verteilten Zufallsvariable  $Y$  mit Freiheitsgrad  $\nu$ :

$$T = \frac{X}{\sqrt{Y/\nu}}$$

Der Wertebereich von  $T$  ist  $\mathbb{R}$  und die Dichtefunktion ist gegeben durch

$$f_{T_\nu}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu\pi}} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}, \nu \in \mathbb{N}.$$

Der Graph ist eine Glockenkurve ähnlich der der normalverteilten mit Mittelwert  $\mu = 0$  und Varianz  $\sigma^2 = \nu/(\nu - 2)$  (siehe Abbildung 8.10). Auch die  $t$ -Verteilung hat eine besondere Stellung

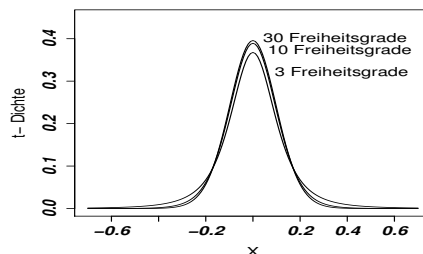


Abbildung 8.10: Dichte dreier  $t$ -Verteilungen mit Freiheitsgrad  $\nu = 3$ ,  $\nu = 10$  und  $\nu = 30$ .

in der Statistik da sie, wie oben angedeutet, auf die Kenntnis der Varianz nicht angewiesen ist. Zur Illustration stellen wir uns nochmals vor,  $n$  Messungen vornehmen zu wollen. Jede Messung sei dabei normalverteilt mit Mittelwert  $\mu$  und Varianz  $\sigma^2$ . Das Messresultat wird dann wie folgt angegeben:

$$\bar{x} \pm s'$$

mit  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  und Standardabweichung  $s' = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ . Es folgt: die Zufallsvariable

$$Y = \frac{\bar{X} - \mu}{S'} = \frac{\bar{X} - \mu}{S'/\sigma} \sim \text{Student}$$

ist  $t$ -verteilt mit  $n - 1$  Freiheitsgraden ( $n - 1$  deshalb weil der Mittelwert  $\bar{X}$  einen Freiheitsgrad einschränkt). Die Verteilung von  $Y$  ist also bekannt, falls von normalverteilten Zufallsvariablen ausgegangen wird. Dabei muss  $\sigma^2$  nicht bekannt sein!

## 8.9 Gemeinsame Verteilung und Unabhängigkeit

Ist die Verteilung einer Zufallsvariable  $X$  bekannt, kann im Prinzip jede probabilistische Frage beantwortet werden. Ist nun gleichzeitig eine weitere Zufallsvariable  $Y$ , deren Verteilung auch bekannt ist, gegeben, können jedoch zusätzliche Fragestellungen erhoben werden, die die Beziehungen zwischen  $X$  und  $Y$  betreffen und die ohne weitere Informationen nicht beantwortet werden können.

Wir betrachten als Beispiel eine statistische Befragung bei der ein zufällig gewählter Teil einer Grundgesamtheit  $\Omega$  nach zwei Charakteren  $X$  und  $Y$  befragt werden (z.B.  $X(\omega)$  =Körpergrösse und  $Y(\omega)$  =Gewicht einer Person  $\omega \in \Omega$ ). Es ist klar, dass Gewicht und Grösse einer Person nicht unabhängig sind und dass die Kenntnis der Grössenverteilung und Gewichtsverteilung noch nicht reicht, um zum Beispiel die Anzahl der übergewichtigen Personen zu studieren. Wir brauchen die **gemeinsame Verteilung** der Zufallsvariablen  $X$  und  $Y$  und wir bilden dazu den **Zufallsvektor**  $Z = (X, Y)$

$$\begin{aligned} Z : \Omega &\rightarrow \mathbb{R}^2 \\ \omega &\rightarrow Z(\omega) = (X(\omega), Y(\omega)) \end{aligned}$$

Die **gemeinsame Verteilung** von  $(X, Y)$  ist, im Falle diskreter Zufallsvariablen, das Wahrscheinlichkeitsgesetz  $P(X = x, Y = y)$  wobei  $x$  und  $y$  alle möglichen Werte von  $X$  bzw.  $Y$  durchlaufen. Im Falle stetiger Variablen ist die gemeinsame Verteilung durch die (gemeinsame) Verteilungsfunktion  $F_{X,Y}$  mit Werten in  $[0, 1]$  gegeben, so dass für alle  $a, b \in \mathbb{R}$  gilt

$$F_{X,Y}(a, b) := P(\{X \leq a, Y \leq b\})$$

oder gleichbedeutend mit Hilfe der gemeinsamen Dichtefunktion

$$f_{X,Y}(a, b) = \frac{\partial^2 F_{X,Y}(a, b)}{\partial a \partial b} = P(\{a < X \leq a + da, b < Y \leq b + db\}).$$

Die gemeinsame Verteilung ist informationsreicher als die Verteilungen von  $X$  und  $Y$  separat betrachtet. Letztere können in der Tat mit Hilfe der gemeinsamen Verteilung wieder aufgefunden werden. Dies gelingt mit dem Begriff der **marginalen Verteilung** eines Zufallsvektors  $(X, Y)$ .

**Marginale Verteilung eines Vektors**  $(X, Y)$ . Die marginale Verteilung bezüglich  $X$  ist im diskreten Fall

$$P(X = x) = \sum_y P(X = x, Y = y)$$

und im stetigen Fall ist die Verteilungsfunktion gegeben durch

$$F_X(a) := \int_{-\infty}^a \left( \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right) dx.$$

Es wird also einfach über die Variable  $Y$  integriert. Die marginale Verteilung bezüglich  $Y$  wird analog durch Integration über  $X$  erhalten.

Das Mehr an Information der gemeinsamen Verteilung wird offensichtlich mit dem Begriff der bedingten Verteilung. Sie gibt zum Beispiel die Verteilung von  $Y$  unter der Bedingung, dass  $X$  einen speziellen Wert  $x$  annimmt. Wir gehen hier nicht weiter auf den Begriff der bedingten Verteilung ein.

Falls die gemeinsame Verteilung von  $X$  und  $Y$  keine zusätzliche Information bezüglich der beiden

separat betrachteten Verteilungen von  $X$  und  $Y$  bringt, sind die Zufallsvariablen unabhängig! In statistischen Anwendungen ist die Unabhängigkeit der Ziehungen von Stichproben sehr oft erfüllt.

**Unabhängigkeit von Zufallsvariablen.** Zwei Zufallsvariablen  $X$  und  $Y$  sind genau dann unabhängig wenn für alle Zahlenpaare  $a, b$  gilt

$$F_{X,Y}(a, b) := P(\{X \leq a, Y \leq b\}) = P(X \leq a)P(Y \leq b) = F_X(a)F_Y(b).$$

Äquivalent hierzu ist mit Hilfe der gemeinsamen Verteilungsdichte:

$X$  und  $Y$  sind genau dann unabhängig wenn für alle Zahlenpaare  $a, b$  gilt

$$f_{X,Y}(a, b) = f_X(a)f_Y(b)$$

Kurz, **zwei Zufallsvariablen sind unabhängig falls ihre Dichtefunktionen faktorisieren.** Der Begriff der Unabhängigkeit kann (und muss) auf mehrere Zufallsvariablen ausgedehnt werden. Es gilt analog:  $n$  Zufallsvariablen  $X_1, X_2, \dots, X_n$  sind unabhängig, falls die Dichtefunktion faktorisiert:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n).$$

**Beispiel 46** Es werden  $n + m$  Stichproben gemacht alle mit Erfolgswahrscheinlichkeit  $p$ . Sei  $X$  die Zufallsvariable, die die Anzahl Erfolge während den  $n$  ersten Stichproben registriert und  $Y$  die Zufallsvariable die die Anzahl Erfolge während den letzten  $m$  Stichproben zählt.  $X$  und  $Y$  sind unabhängig da der Erfolg einer Stichprobe keine Auswirkung auf die anderen Stichproben hat. Formal ist zu sehen, dass für alle  $x$  und  $y$  gilt

$$P(\{X = x, Y = y\}) = \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y} = P(\{X = x\})P(\{Y = y\})$$

◇

**Folgerung.** Die Faktorisierung der Dichtefunktion hat speziell die Faktorisierung des Erwartungswertes zur Folge. Für zwei unabhängige Zufallsvariablen  $X$  und  $Y$  gilt

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy = E(X)E(Y)$$

Das umgekehrte (nämlich der Schluss  $E(XY) = E(X)E(Y) \Rightarrow X$  und  $Y$  sind unabhängig) ist in der Regel nicht richtig. Das lässt uns folgende zwei Definitionen einführen:

(i) Die Kovarianz zweier Zufallsvariablen  $X$  und  $Y$  mit Erwartungswert  $\mu_X$  resp.  $\mu_Y$  ist

$$\text{Cov}(X, Y) := E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X \mu_Y.$$

(ii) Wir sagen  $X$  und  $Y$  sind korreliert falls

$$\text{Cov}(X, Y) \neq 0$$

Ansonsten sind sie unkorreliert.

## 8.10 Aufgaben

**Übung 21** Sei die Verteilungsfunktion einer Zufallsvariable  $X$  gegeben durch

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

- (i) Geben Sie den Graphen von  $F_X$
- (ii) Berechnen Sie  $P(X < 3)$ ,  $P(X > \frac{1}{2})$ ,  $P(X = 1)$  und  $P(2 < X < 3)$ .

◉

**Übung 22** Gegeben sei eine positive reelle Zahl  $\lambda$  und die Zahlenreihe  $p_i = \exp^{-\lambda} \frac{\lambda^i}{i!}$  für  $i \in \{0, 1, 2, \dots\}$ . Zeigen Sie, dass die Zahlenreihe einer Wahrscheinlichkeitsverteilung einer Zufallsvariablen entspricht und skizzieren Sie die Verteilungsfunktion für  $\lambda = 10$ .

◉

**Übung 23** Geben Sie die Verteilungsfunktionen der in Abbildung (8.4) abgebildeten Verteilungen. ◉

**Übung 24** Machen Sie die Verbindung zwischen dem Erwartungswert und dem gewogenen arithmetischen Mittel eines Datensatzes (siehe Kapitel 4). ◉

**Übung 25** Gesucht ist der Erwartungswert der Zufallsvariablen  $X$ , die die Augenzahl eines (geworfenen) Würfels zählt. ◉

**Übung 26** Berechnen Sie den Erwartungswert einer Zufallsvariable mit Verteilung wie in Übung 25. ◉

**Übung 27** Zeigen Sie, dass für alle Zahlenpaare  $(a, b)$  und für eine beliebige Zufallsvariable  $X$  gilt

$$E(aX + b) = aE(X) + b$$

◉

**Übung 28** Machen Sie die Verbindung zwischen der Varianz (bzw. Standardabweichung) und der empirischen Varianz (bzw. empirischen Standardabweichung) eines Datensatzes (siehe Kapitel 4). ◉

**Übung 29** (1) Gesucht ist  $V(X)$  wo  $X$  die Augenzahl eines (geworfenen) Würfels zählt.

(2) Zeigen Sie

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

(3) Zeigen Sie, dass für alle Zahlenpaare  $(a, b)$  und für eine beliebige Zufallsvariable  $X$  gilt

$$V(aX + b) = a^2V(X)$$

◊

**Übung 30** Berechnen Sie den Erwartungswert und die Varianz einer Bernoulli Zufallsvariable. ◊

**Übung 31** (1) Zeigen Sie für obige Verteilung  $\sum_{i=0}^n p(i) = 1$ .

(2) Zeigen Sie, der Erwartungswert, die Varianz und die Standardabweichung einer binomialverteilten Zufallsvariable  $X$  mit Parameter  $(n, p)$  sind gegeben durch

$$E(X) = np, \quad \text{Var}(X) = np(1-p), \quad \sqrt{\text{Var}(X)} = \sqrt{np(1-p)}.$$

Berechnen Sie anschliessend die Wahrscheinlichkeit, dass  $X$  einen Wert innerhalb der Schranke “Mittelwert  $\pm$  Standardabweichung” annimmt. Dabei kann für  $p$  der Wert  $1/2$  angenommen werden.

◊

**Übung 32** Es ist bekannt, dass die Disketten einer gewissen Firma mit einer Wahrscheinlichkeit von 0.01 fehlerhaft sind (Diskettenzustände sind dabei unabhängig voneinander). Die Firma verkauft die Disketten in Losgrösse 10 und garantiert Zurückzahlung, falls mehr als eine der 10 Disketten kaputt ist. Beim Kauf von drei Losen, wie gross ist die Wahrscheinlichkeit, dass genau ein Los zurückgegeben werden muss? ◊

**Übung 33** Ein Industrieunternehmer kauft Transistoren in Paketen zu 20 Stück. Seine Strategie zur Eingangskontrolle besteht darin, 4 Transistoren pro Los zu testen und akzeptiert das ganze Los nur dann, wenn kein Fehler gefunden wird. Falls die Wahrscheinlichkeit eines einzigen Transistors fehlerhaft zu sein 0.1 beträgt (unabhängig der Zustände der anderen Transistoren), welche Proportion von Losen wird der Industrielle zurückzuweisen haben? ◊

**Übung 34** Berechnen Sie den Erwartungswert und die Varianz einer poissonverteilten Zufallsvariable mit Parameter  $\lambda$ . ◊

**Übung 35** (1) Zeigen Sie,  $f$  ist eine Wahrscheinlichkeitsdichte.

(2) Berechnen Sie den Erwartungswert und die Varianz von  $X$

(3) Zeigen Sie, dass für alle  $a, b$  mit  $0 \leq a < b \leq 1$  gilt

$$P(\{a \leq X \leq b\}) = \int_a^b f(x)dx = b - a.$$

Anders gesagt, die Wahrscheinlichkeit, dass  $X$  einen Wert in einem gegebenen Teilintervall von  $(0, 1)$  annimmt, entspricht der Länge dieses Teilintervalls. ◊

**Übung 36** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu$  und  $\sigma^2$ . Zeige  $\mu = E(X)$  und  $\sigma^2 = V(X)$ .  $\odot$

**Übung 37** Seien  $\alpha$  und  $\beta$  zwei reelle Zahlen und sei  $X \sim N(\mu, \sigma^2)$ . Zeigen Sie, dass  $Y = \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$ .  $\odot$

**Übung 38** (1) Zeigen Sie, dass  $f$  eine Dichtefunktion ist.

(2) Berechnen Sie den Erwartungswert und die Varianz von  $X$ .

(3) Zeichnen Sie den Graphen von  $f$  und die Verteilungsfunktion  $F_X(x) = \int_{-\infty}^x f(x)dx$ .  $\odot$

**Übung 39** (1) Sei  $X$  eine gleichverteilte Zufallsvariable auf  $[0, 10]$ . Berechnen Sie folgende Wahrscheinlichkeiten: a)  $P(\{X < 3\})$ , b)  $P(\{X > 6\})$ , c)  $P(\{3 < X < 8\})$ .  $\odot$

**Übung 40** Zeichnen Sie den Graphen von  $\Phi$  und zeigen Sie

$$\Phi(-x) = 1 - \Phi(x)$$

$\odot$

**Übung 41** Sei  $0 < \alpha < 1$  eine Zahl. Wir definieren  $u_\alpha \in \mathbb{R}$  durch

$$\Phi(u_\alpha) = \alpha$$

Berechnen Sie mit Hilfe der sich im Anhang befindenden Tabelle die Werte  $u_\alpha, u_{\alpha/2}, u_{1-\alpha/2}$  für  $\alpha = 0.01$  und  $\alpha = 0.05$ .  $\odot$

**Übung 42** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu = 0$  und  $\sigma^2 = 1$ . Berechnen Sie

- (a)  $P(X \leq 2.44) = ?$     (d)  $P(X \geq 1) = ?$   
 (b)  $P(X \leq -1.16) = ?$     (e)  $P(X \geq -2.9) = ?$   
 (c)  $P(X \leq 1.923) = ?$     (f)  $P(2 \leq X \leq 10) = ?$

$\odot$

**Übung 43** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu = 0.8$  und  $\sigma^2 = 2$ . Berechnen Sie

- (a)  $P(X \leq 2.44) = ?$     (d)  $P(X \geq 1) = ?$   
 (b)  $P(X \leq -1.16) = ?$     (e)  $P(X \geq -2.9) = ?$   
 (c)  $P(X \leq 1.923) = ?$     (f)  $P(2 \leq X \leq 10) = ?$

$\odot$

**Übung 44** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu = 0$  und  $\sigma^2 = 1$ . Berechnen Sie den Wert der Konstanten  $c$  derart, dass

$$\begin{aligned} (a) P(X \geq c) &= 0.1 & (c) P(0 \leq X \leq c) &= 0.45 \\ (b) P(X \leq c) &= 0.05 & (d) P(-c \leq X \leq c) &= 0.99 \end{aligned}$$

◊

**Übung 45** Sei  $X$  eine normalverteilte Zufallsvariable mit Parameter  $\mu = -2$  und  $\sigma^2 = 0.5$ . Berechnen Sie den Wert der Konstanten  $c$  derart, dass

$$\begin{aligned} (a) P(X \geq c) &= 0.2 & (c) P(-2 - c \leq X \leq -2 + c) &= 0.9 \\ (b) P(-c \leq X \leq -1) &= 0.5 & (d) P(-2 - c \leq X \leq -2 + c) &= 0.996 \end{aligned}$$

◊

**Übung 46** Die Stichprobe einer Bolzenproduktion ergibt für den Durchmesser  $X$  eines Bolzens eine Normalverteilung mit Mittelwert  $\mu = 25.08mm$  und Standardabweichung  $\sigma = 0.05mm$ . Die verlangte nominale Abmessung eines Bolzens ist  $25.00 \pm 0.15mm$ . Berechnen Sie den Anteil der toleranzhaltigen Bolzen an der Gesamtproduktion. Wie ändert sich dieser Anteil, falls durch Justierung der Maschinen der Mittelwert auf den Nominalwert gebracht werden kann (ohne  $\sigma$  zu ändern)? ◊

**Übung 47** Sei  $X$  eine Exponentialverteilung mit Parameter  $\lambda$ . Zeigen Sie

$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$$

◊

**Übung 48** Die Lebensdauer eines elektronischen Elements folge einer Exponentialverteilung mit unbekanntem Parameter  $\lambda$ . Um den Erwartungswert der Lebensdauer abzuschätzen, wird eine grosse Zahl von Elementen getestet und es wird folgendes festgestellt: 1% der Elemente haben eine Lebenserwartung kleiner als 120 Stunden und 36 Minuten. Bestimmen Sie den Erwartungswert der Lebensdauer des elektronischen Elements. ◊

**Übung 49** Sei  $T$  eine Studentverteilte Zufallsvariable mit Freiheitsgrad  $\nu = 10$ . Berechnen Sie mit Hilfe Ihres Taschenrechners die Wahrscheinlichkeit  $P(-1 < T < 1)$ . ◊

**Übung 50** Sei  $T$  eine Studentverteilte Zufallsvariable mit Freiheitsgrad  $\nu = 10$ . Berechnen Sie mit Hilfe der Tabelle im Anhang die Grenze  $t$  so, dass  $P(-\infty < T < t) = 0.995$  ist. ◊



**Übung 51** Gegeben ist eine Urne mit 9 blauen und einer roten Kugel. Wir ziehen zwei Kugeln auf zwei verschiedene Arten 1) **mit Zurücklegen**, 2) **ohne Zurücklegen**. In beiden Fällen nennen wir  $X_i$ ,  $i = 1, 2$  die Zufallsvariable die die Farbe des  $i$ -ten Zuges angibt. Zeigen Sie, im Fall 1 sind die Zufallsvariablen unabhängig und im Fall zwei sind sie abhängig.  $\odot$

**Übung 52** Zeigen Sie, unabhängige Zufallsvariablen sind nicht korreliert.  $\odot$

**Übung 53** Zeigen Sie folgende Eigenschaften

(i)  $Cov(X, Y) = Cov(Y, X)$

(ii)  $Cov(X, X) = Var(X)$

(iii)  $Cov(aX, Y) = aCov(X, Y)$

(iv)  $Cov(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j)$

$\odot$

**Übung 54** Zeigen Sie

(i)  $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + Cov(\sum_{i \neq j}^n X_i, \sum_{j=1}^n X_j)$

$\odot$

# Kapitel 9

## Grenzwertsätze

Die Grenzwertsätze bilden die wichtigsten theoretischen Grundlagen der Wahrscheinlichkeitstheorie. Unter ihnen wird hauptsächlich zwischen zwei Kategorien unterschieden: *Gesetz der grossen Zahlen* auf der einen Seite und *zentraler Grenzwertsatz* auf der anderen Seite. Generell sprechen wir vom Gesetz der grossen Zahlen, falls Bedingungen angegeben werden, unter welchen der Mittelwert von Zufallsvariablen nach ihrem gemeinsamen Erwartungswert strebt.

Die zentralen Grenzwertsätze sagen uns, unter welchen Bedingungen die Summe einer grossen Anzahl von Zufallsvariablen approximativ normalverteilt ist.

### 9.1 Gesetz der grossen Zahlen

Das (starke) Gesetz der grosse Zahlen liefert die sichere Konvergenz des arithmetischen Mittels einer Folge von gleichverteilten Zufallsvariablen gegen den gemeinsamen Erwartungswert.

**Theorem.** Sei  $X_1, X_2, \dots$  eine Folge von unabhängigen und gleichverteilten Zufallsvariablen mit gemeinsamem Erwartungswert  $\mu$ . Mit Wahrscheinlichkeit 1 gilt dann

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mu, \text{ für } n \rightarrow \infty$$

Der interessierte Leser findet einen Beweis z.B. in *Initiation aux probabilités* von S. Ross.

**Anwendungen.** Vom Gesetz der grossen Zahlen wird das theoretische Rüstzeug abgeleitet an Hand von Stichproben eines Merkmals (aus einer Grundmenge), seinen Erwartungswert abzuschätzen. Dieser Problemkreis wird in den folgenden Kapiteln besprochen.

### 9.2 Zentraler Grenzwertsatz

Der zentrale Grenzwertsatz besagt, dass die Summe einer grossen Anzahl von unabhängigen und ähnlichen Zufallsvariablen annähernd einer Normalverteilung folgt. Das gibt uns erstens eine mögliche Näherungsformel für Wahrscheinlichkeiten, die Summen von Zufallsvariablen beinhalten und zweitens, liefert er eine Erklärung, warum viele natürliche Phänomene eine glockenförmige Verteilung besitzen.

**Theorem.** Seien  $X_1, X_2, \dots$  eine Folge von unabhängigen und gleichverteilten Zufallsvariablen mit gemeinsamem Erwartungswert  $\mu$  und gemeinsamer Varianz  $\sigma^2$ . Die Verteilung der Zufallsvariable

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

konvergiert für  $n \rightarrow \infty$  gegen eine standardisierte Normalverteilung, das heisst

$$P\left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx, \text{ für } n \rightarrow \infty$$

**Beispiel.** Es werden 10 Würfel geworfen. Gesucht ist die Wahrscheinlichkeit, dass die Summe aller Augen einen Wert zwischen 30 und 40 annimmt.

*Lösung.* Sei  $X_i$  die Augenzahl des  $i$ -ten Würfels,  $i = 1, 2, \dots, 10$ . Da  $E(X_i) = \frac{7}{2}$  und  $V(X_i) = E(X_i^2) - E(X_i)^2 = \frac{35}{12}$  ist, gilt nach dem zentralen Grenzwertsatz

$$\begin{aligned} P\left(30 \leq \sum_{i=1}^{10} X_i \leq 40\right) &= P\left(\frac{30 - 35}{\sqrt{350/12}} \leq \frac{\sum_{i=1}^{10} X_i - 35}{\sqrt{350/12}} \leq \frac{40 - 35}{\sqrt{350/12}}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-5/\sqrt{350/12}}^{5/\sqrt{350/12}} e^{-x^2/2} dx, \\ &= \Phi(5/\sqrt{350/12}) - \Phi(-5/\sqrt{350/12}) \approx 0.65 \end{aligned}$$

### 9.3 Aufgaben

**Übung 55** Werfen Sie 20 Mal einen Würfel. Beobachten Sie die **Stabilisierung des arithmetischen Mittels** durch Mittelung der ersten 5, 10 und 20 Würfe. Machen Sie sich anhand dieses Beispiels klar, dass es prinzipiell möglich ist, den Erwartungswert einer Zufallsvariable durch unendlich oft Wiederholen experimentell zu bestimmen.  $\odot$

**Übung 56** Visualisieren Sie den zentralen Grenzwertsatz mit Hilfe einer Idee von Francis Galton. Konsultieren Sie dazu die Internetseite

<http://www.stattucino.com/berrie/dsl/Galton.html>. Eine ähnliche Simulation finden Sie unter

<http://www.ms.uky.edu/~mai/java/stat/GaltonMachine.html>.  $\odot$

# Kapitel 10

## Solutions II

7)

a)  $\Omega = \{(i, j) \mid i, j \in \{1, 2, \dots, 6\}, i \neq j\}$ .  $|\Omega| = 30$ .

b)  $\Omega = \{(i, j) \mid i, j \in \{1, 2, \dots, 6\}\}$ .  $|\Omega| = 36$ .

8)

$$A \cup B = \Omega, A \cap B = \emptyset, A \cap C = C \text{ und } \bar{A} = B$$

9)

$$B \cap C = \{2\}, A \cup B = \{2; 4; 6\}, B \cup C = \{1; 2; 4; 6\}, C \cap E = \{\}, \bar{A} = \{1; 3; 4; 5; 6\}, B \setminus C = \{4; 6\}, \bar{A} \cup D = \Omega \text{ and } \bar{A} \cap D = \{1; 4; 6\}.$$

10)

1)  $A$  aber nicht  $B = A \setminus B$ ;

2) weder  $A$  noch  $B = \overline{A \cup B}$

3)  $A$  oder  $B$ , aber nicht beide gleichzeitig  $= (A \cup B) \setminus (A \cap B)$ .

11)

Wir setzen  $k$ =Kopf (=face) und  $z$ =Zahl (=pile).

a)  $\Omega = \{(k, z); (z, k); (k, k); (z, z)\}$

b)  $\Omega = \{(i, j) \mid i, j \in \{1, 2, 3, 4, 5, 6\}\}$

c)  $\Omega = \{(z); (k, z); (k, k, z); \dots; \underbrace{(k, \dots, k, z)}_{n - \text{Mal}} \mid n \in \mathbb{N}\}$

12)

Wir setzen  $k$ =Kopf (=face) und  $z$ =Zahl (=pile).

1)  $\Omega = \{(k, 1); (z, 1); (k, 2); (z, 2); (k, 3); (z, 3); (k, 4); (z, 4); (k, 5); (z, 5); (k, 6); (z, 6)\}$

2)  $A = \{(k, 2); (k, 4); (k, 6)\}$ ,  $B = \{(k, 2); (k, 3); (k, 5); (z, 2); (z, 3); (z, 5)\}$ ,  $C = \{(z, 1); (z, 3); (z, 5)\}$

3)  $A \cap B = \{(k, 2)\}$ ,  $C \cap B = \{(z, 3); (z, 5)\}$ ,  $B \cap \bar{A} \cap \bar{C} = \{(k, 3); (k, 5); (z, 2)\}$ ,

4)  $A$  et  $C$  sont incompatibles.

13)

Wir setzen  $k$ =Kopf (=face) und  $z$ =Zahl (=pile). Weiter sei  $|\Omega|$  die Anzahl Elemente von  $\Omega$  und  $\mathcal{P}(\Omega)$  die Menge der Teilmengen von  $\Omega$ .

a)  $\Omega = \{(k, k); (k, z); (z, k); (z, z)\}$ ,  $|\Omega| = 4$ ,  $|\mathcal{P}(\Omega)| = 16$  (falls der erste Wurf vom zweiten Wurf unterschieden wird).

$\Omega' = \{\{k, k\}; \{k, z\}; \{z, z\}\}$ ,  $|\Omega'| = 3$ ,  $|\mathcal{P}(\Omega')| = 8$  (falls der erste Wurf nicht vom zweiten Wurf unterschieden wird).

b)  $|\Omega| = 3 \Rightarrow |\mathcal{P}(\Omega)| = 8$ .

c)  $|\Omega| = n \Rightarrow |\mathcal{P}(\Omega)| = 2^n$ .

**14)** An der Tafel

**15)**

a)  $\frac{180}{400}$ , b)  $\frac{340}{400}$ , c)  $\frac{60}{400}$ .

**16)**

$P(A \cup B) = \frac{11}{12}$ ,  $P(\bar{A} \cap \bar{B}) = \frac{1}{12}$ ,  $P(B \setminus A) = \frac{1}{4}$  and  $P(A \cup \bar{B}) = \frac{3}{4}$ .

**17)**

a)  $P = \frac{1}{2}$  b)  $P = \frac{7}{8}$  c)  $P = \frac{1}{4}$

**18)**

a)  $P = \frac{1}{18}$  b)  $P = \frac{1}{36}$  c)  $P = \frac{1}{2}$  d)  $P = \frac{1}{6}$  e)  $P = \frac{5}{36}$  f)  $P = \frac{11}{36}$  g)  $P = \frac{25}{36}$

**19)**

Wir definieren (on pose):

TP="test positive", TN="test negative", K="Krebs" (cancer), NK="kein Krebs" (pas de cancer).

Es ist dann,

$$P(TP | K) = 0.95, \quad P(TP | NK) = 0.05,$$

$$P(TN | K) = 0.05, \quad P(TN | NK) = 0.95$$

zusätzlich,  $P(K) = 0.004$ . Gesucht ist (on cherche):  $P(K | TP)$ . On a:

$$P(K | TP) = \frac{P(K, TP)}{P(TP)} = \frac{P(TP | K)P(K)}{P(TP)} = \frac{0.95 \cdot 0.004}{P(TP)}$$

und

$$P(TP) = P(TP | K)P(K) + P(TP | NK)P(NK) = 0.95 \cdot 0.004 + 0.05 \cdot 0.996$$

Total:  $P(K | TP) = 0.07$ .

**20)**

$$P(E_2)P(F) = \frac{6}{36} \frac{6}{36} = \frac{1}{36} = P((4, 3))$$

**21)**

$$P(X < 3) = \frac{11}{12}, \quad P(X > \frac{1}{2}) = \frac{3}{4},$$

$$P(X = 1) = P(X \leq 1) - P(X < 1) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}, \quad P(2 < X < 3) = 0$$

**22)**

$$p_i \geq 0 \text{ und } \sum_{i=0}^{\infty} p_i = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1.$$

**23)**

a) (links) Die zu Grunde liegende Zufallsvariable  $X$  nimmt die Werte 0, 1 und 2 an. Die Wahrscheinlichkeitsverteilung von  $X$  ist: (*à gauche*), *la variable aléatoire sous-jacente  $X$  prend les valeurs 0, 1 et 2. La loi de  $X$  est:*

$$p(0) = P(X = 0) = 0.25, \quad p(1) = P(X = 1) = 0.5, \quad p(2) = P(X = 2) = 0.25.$$

Die Verteilungsfunktion  $F_X(x) = P(X \leq x)$  ist in Figur 10.1 links dargestellt: *La fonction de répartition  $F_X(x) = P(X \leq x)$  est donnée dans la figure 10.1 (à gauche):*

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.25 & 0 \leq x < 1 \\ 0.75 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

b) (rechts) Die zu Grunde liegende Zufallsvariable  $X$  nimmt Werte in  $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$  an. Die Wahrscheinlichkeitsverteilung von  $X$  (d.h.  $p(i) = P(X = i)$ ) ist: *(à gauche), la variable aléatoire sous-jacente  $X$  prend des valeurs dans  $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ . La loi de  $X$  est (càd  $p(i) = P(X = i)$ ):*

$$p(2) = p(12) = \frac{1}{36}, \quad p(3) = p(11) = \frac{2}{36}, \quad p(4) = p(10) = \frac{3}{36}, \\ p(5) = p(9) = \frac{4}{36}, \quad p(6) = p(8) = \frac{5}{36}, \quad p(7) = \frac{6}{36}.$$

Die Verteilungsfunktion  $F_X(x) = P(X \leq x)$  kann zum Beispiel durch den Ausdruck  $F_X(x) = \sum_{i=2}^{\lfloor x \rfloor} p(i)$  gegeben wobei die Summe gleich Null ist falls  $x < 2$ . Die Funktion ist in Figur 10.1 unten rechts als Graph dargestellt.

La fonction de répartition  $F_X(x) = P(X \leq x)$  peut être donnée par l'expression  $F_X(x) = \sum_{i=2}^{\lfloor x \rfloor} p(i)$  où la somme est zéro si  $x < 2$ . Elle est donnée dans la figure 10.1 à gauche.

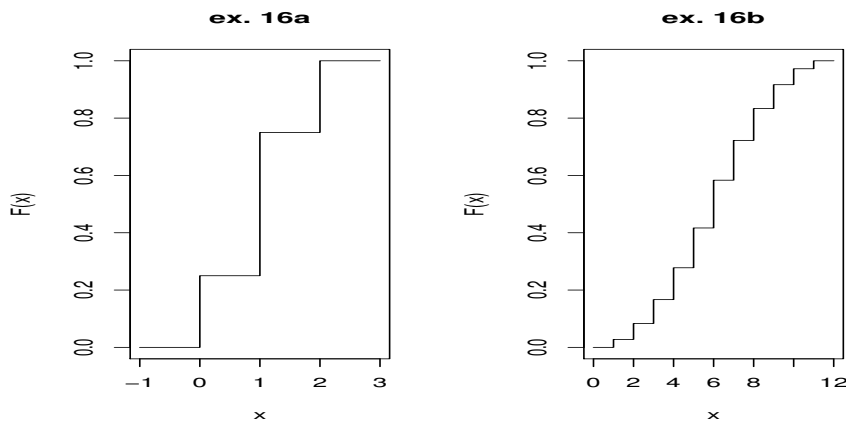


Abbildung 10.1: Verteilungsfunktionen Übung 23.

**25)**

Es ist  $\Omega = \{1, 2, \dots, 6\}$  und für  $\omega \in \Omega$  ist  $X(\omega) = \omega$ . Die Wahrscheinlichkeitsverteilung von  $X$  ist: *La loi de  $X$  est:  $p(i) = P(X = i) = \frac{1}{6}$ , für  $i = 1, \dots, 6$ . Der Erwartungswert berechnet sich zu: L'espérance mathématique est donnée par*

$$E(X) = 1 \frac{1}{6} + 2 \frac{1}{6} + 3 \frac{1}{6} + 4 \frac{1}{6} + 5 \frac{1}{6} + 6 \frac{1}{6} = 3.5$$

**27)**

Seien  $a, b$  zwei reelle Zahlen und  $X$  eine diskrete Zufallsvariable mit Werten  $\{x_1, x_2, \dots, x_n, \dots\}$ . Sei  $p(x_i) = P(X = x_i)$  die Wahrscheinlichkeit, dass  $X$  den Wert  $x_i$  annehme. Da  $E(g(X)) = \sum_i g(x_i)p(x_i)$  ist, gilt nun

$$E(aX + b) = \sum_{i=1}^{\infty} (ax_i + b)p(x_i) = a \sum_{i=1}^{\infty} x_i p(x_i) + b \sum_{i=1}^{\infty} p(x_i)$$

Da aber  $\sum_{i=1}^{\infty} x_i p(x_i) = E(X)$  und  $\sum_{i=1}^{\infty} p(x_i) = 1$  ist folgt  $E(aX + b) = aE(X) + b$ .

Soient  $a, b$  deux nombres réels et  $X$  une variable aléatoire discrète avec des valeurs dans  $\{x_1, x_2, \dots, x_n, \dots\}$ . Soit  $p(x_i) = P(X = x_i)$  la loi de  $X$  (i.e., la probabilité que  $X$  prenne la valeur  $x_i$  est  $p(x_i)$ ). Comme  $E(g(X)) = \sum_i g(x_i)p(x_i)$  il suit

$$E(aX + b) = \sum_{i=1}^{\infty} (ax_i + b)p(x_i) = a \sum_{i=1}^{\infty} x_i p(x_i) + b \sum_{i=1}^{\infty} p(x_i)$$

Mais  $\sum_{i=1}^{\infty} x_i p(x_i) = E(X)$  et  $\sum_{i=1}^{\infty} p(x_i) = 1$  d'où  $E(aX + b) = aE(X) + b$ .

**29)**

1) Mit Übung 25 folgt  $\mu := E(X) = 3.5$ . Es ist (En utilisant les notations de l'exercice 25 on a  $\mu := E(X) = 3.5$ . donc)

$$V(X) = E((X - \mu)^2) = \sum_{i=1}^6 (i - \mu)^2 p(i) = \sum_{i=1}^6 (i - 3.5)^2 \frac{1}{6} \approx 2.9$$

2) Sei  $\mu = E(X)$ . Es folgt mit Hilfe von Übung 27 (On pose  $\mu = E(X)$ . En utilisant l'exercice 27 il suit)

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

3) Seien  $a, b$  zwei reelle Zahlen und  $X$  eine diskrete Zufallsvariable mit Werten  $\{x_1, x_2, \dots, x_n, \dots\}$ . Sei  $p(x_i) = P(X = x_i)$  die Wahrscheinlichkeit, dass  $X$  den Wert  $x_i$  annehme. Übung 27 sagt uns, dass  $E(aX + b) = aE(X) + b$  ist. Es folgt mit der Abkürzung  $\mu = E(X)$  und nochmaligem Anwenden von Übung 27, dass

$$V(aX + b) = E([(aX + b) - a\mu + b]^2) = E(a^2[X - \mu]^2) = a^2 E([X - \mu]^2) = a^2 V(X).$$

Soient  $a, b$  deux nombres réels et  $X$  une variable aléatoire discrète avec des valeurs dans  $\{x_1, x_2, \dots, x_n, \dots\}$ .

Soit  $p(x_i) = P(X = x_i)$  la loi de  $X$  (i.e., la probabilité que  $X$  prenne la valeur  $x_i$  est  $p(x_i)$ ). L'exercice 27 montre que  $E(aX + b) = aE(X) + b$ . En posant  $\mu = E(X)$  il suit, encore en utilisant l'exercice 27

$$V(aX + b) = E([(aX + b) - a\mu + b]^2) = E(a^2[X - \mu]^2) = a^2 E([X - \mu]^2) = a^2 V(X).$$

**30)**

Sei  $X$  eine Bernoulli-Variablen mit Parameter  $p \in [0, 1]$ . Soit  $X$  une variable de Bernoulli de paramètre  $p \in [0, 1]$ .

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$V(X) = E(X^2) - E(X)^2 = (0^2 \cdot (1 - p) + 1^2 \cdot p) - p^2 = p - p^2 = p(1 - p).$$

**31)**

Sei  $p(i) = \binom{n}{i} p^i (1 - p)^{n-i}$  das Gesetz einer binomialverteilten Zufallsvariable  $X$  mit Parameter  $(n, p)$ ,  $p \in [0, 1]$ . Soit  $p(i) = \binom{n}{i} p^i (1 - p)^{n-i}$  la loi binomial de paramètre  $(n, p)$ ,  $p \in [0, 1]$  d'une variable aléatoire  $X$

(1) Die Binomialformel  $(a + b)^n = \sum_i \binom{n}{i} a^i b^{n-i}$  liefert uns sofort das Ergebnis (Utiliser la formule binomial pour voir que)

$$\sum_i \binom{n}{i} p^i (1 - p)^{n-i} = (p + 1 - p)^n = 1$$

(2) Sei  $X$  eine binomialverteilte Zufallsvariable mit Parameter  $(n, p)$ . Per Definition,  $X$  ist die Anzahl erfolgreicher Stichproben aus insgesamt  $n$  Stichproben (Erfolgswahrscheinlichkeit =  $p$ ). Wir stellen nun  $X$  mit Hilfe von  $n$  unabhängigen Bernoullivariablen  $X_1, \dots, X_n$  mit Parameter  $p$  dar. Dabei modelliert  $X_i$  den Ausgang der  $i$ -ten Stichprobe. Ist sie erfolgreich, so ist  $X_i = 1$  wenn nicht, so ist  $X_i = 0$ . Es folgt

$$X = \sum_{i=1}^n X_i.$$

und daher  $E(X) = \sum_{i=1}^n E(X_i) = np$ . Im Mittel werden also  $np$  Stichproben erfolgreich sein. Für die Varianz folgt ähnlich

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = E\left(\left(\sum_{i=1}^n X_i\right)^2\right) - (np)^2 \\ &= E\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right) - (np)^2 \\ &= \sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i)E(X_j) - (np)^2 \\ &= np + n(n-1)p^2 - (np)^2 = np(1-p) \end{aligned}$$

Die Varianz von  $X$  ist also  $n$  mal die Varianz einer Stichprobe.

*Soit  $X$  une variable aléatoire binomial de paramètre  $n, p$ . Par définition,  $X$  est le nombre de succès lors de  $n$  épreuves indépendantes (propa d'un succès =  $p$ ). On représente  $X$  à l'aide de  $n$  variables aléatoires de Bernoulli  $X_1, \dots, X_n$  avec paramètre  $p$ . On interprète  $X_i$  comme l'issue de la  $i$ -ème épreuve. Si c'est un succès  $X_i = 1$  sinon  $X_i = 0$ . Ainsi*

$$X = \sum_{i=1}^n X_i.$$

*et de là on trouve  $E(X) = \sum_{i=1}^n E(X_i) = np$  ce qui veut dire qu'en moyenne  $np$  épreuves parmi  $n$  fournissent succès.*

*Pour la variance on a de façon similaire*

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = E\left(\left(\sum_{i=1}^n X_i\right)^2\right) - (np)^2 \\ &= E\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right) - (np)^2 \\ &= \sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i)E(X_j) - (np)^2 \\ &= np + n(n-1)p^2 - (np)^2 = np(1-p) \end{aligned}$$

*Ainsi la variance de  $X$  est  $n$  fois la variance d'une épreuve.*

### 32)

Sei  $X$  die Zufallsvariable die für ein bestimmtes Los die Anzahl fehlerhafter Disketten zählt. Genau genommen ist  $X$  binomialverteilt mit Parameter  $(10, 0.01)$ . Wir werden jedoch  $X$  durch eine Poisson-Variable mit Parameter  $\lambda = 10 \cdot 0.01 = 0.1$  annähern. Die Wahrscheinlichkeit  $p$ , ein Los zurückkaufen zu müssen, ist dann

$$p := P(X > 1) = 1 - (P(X = 0) + P(X = 1)) = 1 - (e^{-0.1} + 0.1e^{-0.1})$$

Sei nun  $Y$  die Zufallsvariable, die beim Kauf von drei Losen die Anzahl derer modelliert, die zurückgenommen werden müssen.  $Y$  ist binomialverteilt mit Parameter  $(3, p)$ . Es folgt

$$P(Y = 1) = \binom{3}{1} p^1 (1-p)^2.$$

*Soit  $X$  la variable aléatoire comptant le nombre de disquettes dans un lot choisi. A juste titre,  $X$  suit une loi binomiale de paramètre  $(10, 0.01)$ . Nous allons approcher  $X$  par une variable de Poisson avec paramètre  $\lambda = 10 \cdot 0.01 = 0.1$ . La probabilité  $p$  de devoir racheter un lot est alors*

$$p := P(X > 1) = 1 - (P(X = 0) + P(X = 1)) = 1 - (e^{-0.1} + 0.1e^{-0.1}).$$



Si on appelle  $Y$  la variable aléatoire qui compte les lots à racheter en vendant 3, on voit que  $Y$  suit une loi binomiale de paramètre  $(3, p)$ . Ainsi

$$P(Y = 1) = \binom{3}{1} p^1 (1-p)^2.$$

### 33)

Sei  $X$  die Zufallsvariable die für ein bestimmtes Paket die Anzahl fehlerhafter Transistoren zählt. Genau genommen ist  $X$  binomialverteilt mit Parameter  $(4, 0.1)$ . Wir werden jedoch  $X$  durch eine Poisson-Variablen mit Parameter  $\lambda = 4 \cdot 0.1 = 0.4$  annähern. Die Wahrscheinlichkeit  $p$ , ein Paket zurückkaufen zu müssen, ist dann

$$p := P(X > 0) = 1 - P(X = 0) = 1 - e^{-0.4}$$

Sei nun  $Y$  die Zufallsvariable, die beim Kauf von  $n$  Paketen die Anzahl derer modelliert, die zurückgenommen werden müssen.  $Y$  ist binomialverteilt mit Parameter  $(n, p)$ . Es folgt

$$E(Y) = np.$$

Im Mittel müssen also  $p \cdot 100\%$  Prozent der Pakete zurückgenommen werden.

*Soit  $X$  la variable aléatoire comptant le nombre de transistors dans un paquet choisi. A juste titre,  $X$  suit une loi binomiale de paramètre  $(4, 0.1)$ . Nous allons approcher  $X$  par une variable de Poisson avec paramètre  $\lambda = 4 \cdot 0.1 = 0.4$ . La probabilité  $p$  de devoir rejeter un paquet est alors*

$$p := P(X > 0) = 1 - P(X = 0) = 1 - e^{-0.4}.$$

*Si on appelle  $Y$  la variable aléatoire qui compte les paquets à rejeter en vendant  $n$ , on voit que  $Y$  suit une loi binomiale de paramètre  $(n, p)$ . Ainsi*

$$E(Y) = np$$

et donc en moyenne, l'industriel doit rejeter  $p \cdot 100\%$  pourcent des paquets.

### 34)

Sei  $X$  eine Poisson-Variablen mit Parameter  $\lambda > 0$ . *Soit  $X$  une variable de Poisson de paramètre  $\lambda > 0$ .*

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} &= e^{-\lambda} \lambda \sum_{i=0}^{\infty} \frac{d}{d\lambda} \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \lambda \frac{d}{d\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} &= e^{-\lambda} \lambda \frac{d}{d\lambda} e^{\lambda} = \lambda \end{aligned}$$

Für die Varianz benutzen wir  $V(X) = E(X^2) - E(X)^2$ .

$$\begin{aligned} E(X^2) &= \sum_{i=0}^{\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} &= e^{-\lambda} \lambda \sum_{i=1}^{\infty} i \frac{\lambda^{i-1}}{(i-1)!} \\ &= e^{-\lambda} \lambda \sum_{i=1}^{\infty} \frac{d}{d\lambda} \frac{\lambda^i}{(i-1)!} &= e^{-\lambda} \lambda \frac{d}{d\lambda} \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= e^{-\lambda} \lambda \frac{d}{d\lambda} \lambda \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} &= e^{-\lambda} \lambda \frac{d}{d\lambda} \lambda e^{\lambda} \\ &= e^{-\lambda} \lambda (\lambda + 1) e^{\lambda} &= \lambda^2 + \lambda \end{aligned}$$

Il suit pour la variance

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

### 35)

$$(i) \int_{-\infty}^{\infty} f(x) = \int_0^1 1 dx = x \Big|_0^1 = 1$$

$$(ii) E(X) = \int_{-\infty}^{\infty} xf(x) = \int_0^1 x dx = \frac{x^2}{2} \Big|_0^1 = 1/2.$$

$$(iii) V(X) = \int_{-\infty}^{\infty} (x - \frac{1}{2})^2 f(x) dx = \frac{(x-\frac{1}{2})^3}{3} \Big|_0^1 = \frac{1}{12}.$$

**38)**

$$(i) \int_{-\infty}^{\infty} f(x) = \frac{1}{\beta-\alpha} \int_{\alpha}^{\beta} 1 dx = 1$$

$$(ii) E(X) = \frac{\alpha+\beta}{2}.$$

$$(iii) V(X) = \frac{(\beta-\alpha)^2}{12}.$$

**39)**

$$(i) \frac{1}{10} \int_0^3 dx = \frac{3}{10}$$

$$(ii) \frac{1}{10} \int_6^{10} dx = \frac{4}{10}$$

$$(iii) \frac{1}{10} \int_3^8 dx = \frac{5}{10}$$

**36)**

Man mache in den Definitionen für  $E(X)$  und  $V(X)$  die Substitution

$$y = \frac{x - \mu}{\sigma}$$

Für  $a > 0$  gelten nun allgemein die Beziehungen

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}, \quad \int_{-\infty}^{\infty} xe^{-ax^2} dx = 0, \quad \int_{-\infty}^{\infty} x^2 e^{-ax^2} dx = \frac{1}{2a} \sqrt{\frac{\pi}{a}}.$$

Es folgt direkt  $E(X) = \mu$  und  $V(X) = \sigma^2$ .

*Dans les définitions pour  $E(X)$  et  $V(X)$  on fait la substitution*

$$y = \frac{x - \mu}{\sigma}.$$

*Pour  $a > 0$  on a les relations*

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}, \quad \int_{-\infty}^{\infty} xe^{-ax^2} dx = 0, \quad \int_{-\infty}^{\infty} x^2 e^{-ax^2} dx = \frac{1}{2a} \sqrt{\frac{\pi}{a}}.$$

*en les utilisant il suit il suit  $E(X) = \mu$  et  $V(X) = \sigma^2$ .*

**37)** Es genügt zu zeigen, dass  $P(\alpha X + \beta \leq y)$  für alle  $y$  von der Form  $\int_{-\infty}^y f(x) dx$  ist, wobei  $f$  eine gaussische Dichtefunktion mit Mittelwert  $\alpha\mu + \beta$  und Varianz  $\sigma^2\alpha^2$  ist. Für  $\alpha > 0$  gilt

$$P(\alpha X + \beta \leq y) = P(X \leq (y - \beta)/\alpha) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{(y-\beta)/\alpha} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Substitution  $x = \frac{y-\beta}{\alpha}$  liefert

$$P(\alpha X + \beta \leq y) = \frac{1}{\sqrt{2\pi}\sigma\alpha} \int_{-\infty}^y e^{-\frac{(y-(\beta+\alpha\mu))^2}{2\sigma^2\alpha^2}} dy$$

mit Hilfe von Übung 36 identifizieren wir  $E(Y) = \beta + \alpha\mu$  und  $V(X) = \sigma^2\alpha^2$ .

Il suffit de montrer que pour tout  $y$  la probabilité  $P(\alpha X + \beta \leq y)$  est de la forme  $\int_{-\infty}^y f(x)dx$  avec  $f$  une fonction densité d'une loi normale avec paramètres  $\alpha\mu + \beta$  et  $\sigma^2\alpha^2$ . On a pour  $\alpha > 0$

$$P(\alpha X + \beta \leq y) = P(X \leq (y - \beta)/\alpha) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{(y-\beta)/\alpha} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

en utilisant la substitution  $x = \frac{y-\beta}{\alpha}$  on trouve

$$P(\alpha X + \beta \leq y) = \frac{1}{\sqrt{2\pi}\sigma\alpha} \int_{-\infty}^y e^{-\frac{(y-(\beta+\alpha\mu))^2}{2\sigma^2\alpha^2}} dy.$$

Ainsi l'exercice 36 nous dit que  $E(Y) = \beta + \alpha\mu$  et  $V(X) = \sigma^2\alpha^2$ .

**42)**

- a)  $P(X \leq 2.44) = 0.9927$
- b)  $P(X \leq -1.16) = 0.1230$
- c)  $P(X \leq 1.923) = 0.9728$
- d)  $P(X \geq 1) = 0.1587$
- e)  $P(X \geq -2.9) = 0.9981$
- f)  $P(2 \geq X \geq 10) = 0.0227$

**43)**

- a)  $P(X \leq 2.44) = 0.8769$
- b)  $P(X \leq -1.16) = 0.0829$
- c)  $P(X \leq 1.923) = 0.7864$
- d)  $P(X \geq 1) = 0.4438$
- e)  $P(X \geq -2.9) = 0.9956$
- f)  $P(2 \geq X \geq 10) = 0.1981$

**44)**

- a)  $c = 1.282$
- b)  $c = -1.645$
- c)  $c = 1.645$
- d)  $c = 2.58$

**45)**

- a)  $c = -1.404$

- b)  $c = 2.140$   
 c)  $c = 1.163$   
 d)  $c = 2.036$

46)

92% der Bolzen sind im Mittel toleranzhaltig. Wird die Maschine richtig justiert, erhöht sich dieser Wert auf 99.7%.

92% des boulons respectent la tolérance. Cette valeur monte à 99.7% si on ajuste bien la machine.

47)

$$\begin{aligned} E(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx &= \lambda \int_0^{\infty} \frac{d}{d\lambda} e^{-\lambda x} dx \\ &= \lambda \frac{d}{d\lambda} \left( - \int_0^{\infty} e^{-\lambda x} dx \right) &= \lambda \frac{d}{d\lambda} \frac{1}{\lambda} e^{-\lambda x} \Big|_{x=0}^{\infty} = \lambda \frac{d}{d\lambda} \left( 0 - \frac{1}{\lambda} \right) = \frac{1}{\lambda} \end{aligned}$$

Für die Varianz brauchen wir  $V(X) = E(X^2) - E(X)^2$ . On a

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx &= \lambda \frac{d^2}{d\lambda d\lambda} \int_0^{\infty} e^{-\lambda x} dx \\ &= \lambda \frac{d^2}{d\lambda d\lambda} \left( - \frac{1}{\lambda} e^{-\lambda x} \Big|_{x=0}^{\infty} \right) &= \lambda \frac{d^2}{d\lambda d\lambda} \left( \frac{1}{\lambda} \right) = \frac{2}{\lambda^2} \end{aligned}$$

et donc  $V(X) = \frac{1}{\lambda^2}$ .

48)

$X \sim \text{Poisson}(\lambda)$ . Es gilt  $P(X \leq 120.6) = 0.1$  daraus folgt  $\lambda = -\frac{\ln(0.9)}{120.6}$ . Ainsi on trouve pour l'espérance  $E(X) = -\frac{120.6}{\ln(0.9)}$ .

49)

**Vorbemerkung zur Kovarianz.** Durch ausmultiplizieren der Definition erhält man.

**Remarque concernant la covariance.** En distribuant le produit dans la définition on obtient

$$\begin{aligned} \text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) = E(XY - YE(X) - XE(Y) + E(X)E(Y)) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

à voir: Si  $X$  et  $Y$  sont indépendantes alors  $\text{Cov}(X, Y) = 0$ .

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy$$

grâce à l'indépendance, les densités factorisent:  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ , Ainsi

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} (x - \mu_X) f_X(x) dx \int_{-\infty}^{\infty} (y - \mu_Y) f_Y(y) dy = (\mu_X - \mu_X)(\mu_Y - \mu_Y) = 0$$

53)

Alle Eigenschaften folgen sofort aus der Definition der Kovarianz und den Eigenschaften des Integrals. Wir setzen  $E(X) = \mu_X$  und  $E(Y) = \mu_Y$ . Toutes les propriétés suivent de la définition de la covariance et des propriétés de l'intégrale. On pose  $E(X) = \mu_X$  et  $E(Y) = \mu_Y$ .

$$(i) \text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E((Y - \mu_Y)(X - \mu_X)) = \text{Cov}(Y, X)$$

$$(ii) \text{Cov}(X, X) = E((X - \mu_X)(X - \mu_X)) = E((X - \mu_X)^2) = \text{Var}(X)$$

$$(iii) \text{Cov}(aX, Y) = E\left(\left((aX - a\mu_X)(Y - \mu_Y)\right)\right) = aE\left(\left((X - \mu_X)(Y - \mu_Y)\right)\right) = a\text{Cov}(Y, X)$$

$$(iv) \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = E\left(\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)\left(\sum_{j=1}^m Y_j - E\left(\sum_{j=1}^m Y_j\right)\right)\right)$$

Wir haben  $E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \mu_i$  mit  $\mu_i := E(X_i)$

und  $E\left(\sum_{j=1}^m Y_j\right) = \sum_{j=1}^m E(Y_j) = \sum_{j=1}^m \nu_j$  mit  $\nu_j := E(Y_j)$

Also

$$\begin{aligned} \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= E\left(\sum_{i=1}^n \sum_{j=1}^m (X_i - \mu_i)(Y_j - \nu_j)\right) = \\ &= \sum_{i=1}^n \sum_{j=1}^m E\left((X_i - \mu_i)(Y_j - \nu_j)\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}\left((X_i, Y_j)\right). \end{aligned}$$

**54)**

On pose  $X = \sum_{i=1}^n X_i$  et on calcul  $V(X) = E(X^2) - E(X)^2$ .

$$\begin{aligned} E(X)^2 &= \left(\sum_{i=1}^n E(X_i)\right)^2 \\ &= \sum_{i=1}^n E(X_i)^2 + \sum_{i \neq j} E(X_i)E(X_j) \end{aligned}$$

$$\begin{aligned} E(X^2) &= E\left(\left(\sum_{i=1}^n X_i\right)^2\right) = E\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right) \\ &= \sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i X_j) \end{aligned}$$

en regroupant les termes

$$\begin{aligned} V(X) &= \sum_{i=1}^n E(X_i^2) - E(X_i)^2 + \sum_{i \neq j} E(X_i X_j) - E(X_i)E(X_j) \\ &= \sum_{i=1}^n V(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j). \end{aligned}$$

## Teil III

# Schliessende Statistik

# Kapitel 11

## Stichproben

Die Informationsbeschaffung über eine Grundgesamtheit kann schnell zu einer nicht zu bewältigenden Arbeitsflut ausarten. Um sie zu umgehen, werden statistische Erhebungen anhand einer **Stichprobe** (oder Sample), das heisst einer **Teilmenge der Grundgesamtheit**, gemacht. Die Stichprobe soll erlauben, mit einer vorgegebenen Präzision auf gewisse Informationen, die in der Grundgesamtheit stecken, zu schliessen. In diesem kleinen Kapitel sollen die wichtigsten Begriffe angegeben werden.

**Umfang einer Stichprobe.** Generell gilt, je grösser die Stichprobe bezüglich der Grösse der Grundgesamtheit, desto vertrauenswürdiger sind die Aussagen der Studie.

Da die Kosten mit der Stichprobengrösse anwachsen, ist jedoch ein Kompromiss zwischen Kosten und Stichprobengrösse einzugehen. Sehr oft kann folgende Proportionalität angenommen werden:

Die Kosten steigen linear mit dem Stichprobenumfang  $n$ . Die Vertrauenswürdigkeit der Aussage steigt jedoch nur proportional zu  $\sqrt{n}$ .

**Auswahlmethode.** Prinzipiell muss die Stichprobe durch **zufälliges Ziehen** in der Grundgesamtheit so erfolgen, dass alle Elemente der Grundgesamtheit die gleiche Wahrscheinlichkeit haben, in die Stichprobe aufgenommen zu werden. Oft werden dazu Folgen von Zufallszahlen gebraucht. Andere Auswahlmethoden würden weite Teile der hier vorgestellten Theorie ungültig machen! Die im Beispiel 1 vorgeschlagene Methode, jeden hundertsten Bolzen zu testen, ist in diesem Sinne **kein** erlaubtes Auswahlverfahren! Zufallszahlen können mit dem Computer generiert werden, aus speziellen Büchern entnommen werden und seit Kurzem auch vom CERN in Genf bestellt werden (letztere basieren auf quantenmechanischen Experimenten und sind in einem gewissen Sinne die bestmöglichen Zufallszahlen).

**Auswahlverfahren.** Es gibt fundamental zwei Varianten zum (zufälligen) Auswahlverfahren, nämlich eine Ziehung **mit oder ohne Zurücklegen**.

**Die Ziehung mit Zurücklegen** besteht darin, nach der Selektierung und Ziehung eines Elements der Grundgesamtheit, das gezogene Element wieder in die Grundmenge zurückzulegen.

Der Ablauf einer **Ziehung ohne Zurücklegen** versteht sich nun von selbst.

**Auswertung der Stichprobe.** Nehmen wir an, es sei der Charakter  $X$  von fabrikgefertigten Werkstücken (zum Beispiel der Durchmesser eines Bolzens) durch Entnahme einer Stichprobe der Grösse  $n$  zu überwachen. Bevor die Stichprobe ausgewählt ist, sind die potentiellen Werte des Charakters Zufallsvariablen  $X_1, X_2, \dots, X_n$ , die als nichtausgewertete Kopien von  $X$  angesehen werden können. Da wir die Verteilung von  $X$  nicht kennen, beginnen wir die beiden wichtigsten stati-

stischen Grössen abzuschätzen, nämlich den Erwartungswert und die Varianz von  $X$ . Für den Erwartungswert  $\bar{X}$  wird generell die Statistik des arithmetischen Mittels verwendet

$$\bar{X} = f(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n X_i.$$

Zu diesem Zeitpunkt (vor Entnahme der Stichprobe) ist  $\bar{X}$  selber noch eine Zufallsvariable. Für die Varianz wird die Statistik

$$S^2 = f(X_1, \dots, X_n) := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

verwendet, falls der Erwartungswert von  $X$ ,  $\mu = E(X)$ , bekannt ist und

$$S'^2 = f(X_1, \dots, X_n) := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

falls nicht. Beides sind sie Zufallsvariablen.

Sobald die Stichprobe gezogen ist, sind die Werte der Zufallsvariablen  $X_i$  bekannt und wir schreiben sie in der Regel klein als  $x_i$ . Mittelwert und Varianz der Stichprobe ergeben sich zu

$$\bar{x} = \frac{1}{n} \sum_1^n x_i \quad \text{und} \quad s'^2 = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2.$$

Das sind die ersten quantitativen Angaben die von einer Stichprobe zu berechnen sind.

### Bemerkungen.

- (i) **Jede Funktion basierend auf der Stichprobe  $x_1, \dots, x_n$  wird Statistik der Stichprobe genannt.** Speziell sind  $\bar{X}$  und  $S'^2$  Statistiken.
- (ii) Das Gesetz der grossen Zahlen sagt uns, dass für grosse  $n$  der Wert  $\bar{X}$  eine gute Annäherung für den theoretischen Erwartungswert von  $X$ ,  $E(X)$ , ist.
- (iii) Da  $\bar{X}$  eine Zufallsvariable ist kann ihr Erwartungswert ausgerechnet werden. Es ist

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

wobei  $\mu$  der theoretische Erwartungswert von  $X$  ist. Der Mittelwert hängt dabei nicht von der Art der Ziehung (mit oder ohne Zurücklegen) ab. **Für die Varianz von  $\bar{X}$  ist die Art der Ziehung jedoch von Bedeutung.** Es gilt, falls die Auswahlmethode mit Zurücklegen erfolgt:

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

wobei  $\sigma^2$  die theoretische Varianz von  $X$  ist. Falls die Auswahlmethode ohne Zurücklegen erfolgt, sind die Ziehungen nicht mehr unabhängig und es gilt:

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}.$$



- (iv) Ähnlich kann der Erwartungswert von  $S'^2$  errechnet werden. Wir finden, falls die Auswahlmethode mit Zurücklegen erfolgt:

$$E(S'^2) = \sigma^2$$

wobei  $\sigma^2$  die theoretische Varianz von  $X$  ist. Im Falle ohne Zurücklegen finden wir

$$E(S'^2) = \sigma^2 \frac{N}{N-1}$$

wo  $N$  die Grösse der Grundgesamtheit ist. Auf die Angabe der Varianz von  $S^2$  verzichten wir hier.

## 11.1 Aufgaben

**Übung 57** Es werde eine Stichprobe vom Umfang  $n$  konzipiert (Stichprobe mit Zurücklegen). Berechnen Sie den Erwartungswert von  $S^2$  und die Varianz von  $\bar{X}$  falls wie oben  $E(X_i) = \mu$  und  $Var(X_i) = \sigma^2$  ist.

◊

# Kapitel 12

## Schätzmethoden

### 12.1 Punktschätzung

**Allgemeines.** Wir interessieren uns für eine bestimmte Ausprägung  $X$ , die sich bei Individuen einer Grundgesamtheit zeigt. Die Ausprägung sei durch einen (oder mehrere) Parameter  $\lambda$  bestimmt. Wir setzen uns zur Aufgabe,  $\lambda$  mit Hilfe einer Stichprobe der Grösse  $n$  abzuschätzen.

**Beispiel 47** Wir betrachten die Tagesproduktion von Bolzen aus Beispiel 1. Wir ziehen zufällig (sagen wir mit Zurücklegen, damit die Unabhängigkeit gewahrt bleibt)  $n$  Bolzen aus der Tagesproduktion. Da wir die Ausprägung “Durchmesser” abschätzen wollen werden wir entsprechend die Durchmesser der  $n$  Bolzen messen. Dieser Vorgang liefert uns  $n$  unabhängige Zufallsvariablen  $X_1, X_2, \dots, X_n$ . Der Produktionsprozess muss natürlich so angelegt sein, dass der Erwartungswert der  $X_i$  dem Nominalwert des Durchmessers entspricht. Andernfalls wäre ein unakzeptabler systematischer Fehler im Prozess vorhanden! Wir nehmen also an, dass unser Parameter  $\lambda$  gerade dem Erwartungswert  $E(X_i)$  entspricht. Um ihn abzuschätzen bedienen wir uns dem Gesetz der grossen Zahlen welches besagt, dass für  $n \rightarrow \infty$  der arithmetische Mittelwert  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  gegen  $E(X_i) = \lambda$  konvergiert. Praktisch heisst das: Für grosse  $n$  ist  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  ein guter Schätzer für  $\lambda$ .

◇

Kehren wir zurück zu unseren allgemeinen Betrachtungen. Wir nennen

$$\hat{\lambda} = f(x_1, x_2, \dots, x_n)$$

die Schätzung des wirklichen Wertes  $\lambda$  basierend auf der (konkreten) Stichprobe  $x_1, \dots, x_n$ . Wie durch obige Gleichung angedeutet, ist  $\hat{\lambda}$  eine Funktion der Stichprobe  $x_1, \dots, x_n$ . Werden alle möglichen Stichproben der Grösse  $n$  angesprochen, erhalten wir eine Zufallsvariable  $\Lambda$ , den Schätzer von  $\lambda$ :

$$\Lambda = f(X_1, \dots, X_n).$$

Wir sagen, der Schätzer  $\Lambda$  ist erwartungstreu für  $\lambda$  falls  $E(\Lambda) = \lambda$  ist.

**Beispiel 48**  $\bar{X} = f(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$  ist ein erwartungstreuer Schätzer des Mittelwertes. ◇

**Beispiel 49** Die Zufallsvariable  $f(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ , schätzt die Varianz von  $X$ , ist aber nicht erwartungstreu für  $\sigma^2$  (mit Ziehen ohne Zurücklegen) da (Übung!)

$$E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n} \sigma^2$$

gilt. Ein oft verwendeter, erwartungstreuer Schätzer für  $\sigma^2$  ist

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

◇

**Qualität des Schätzers.** Für ein und denselben Parameter gibt es in der Regel mehrere Schätzfunktionen. Es drängt sich die Frage auf, welcher der möglichen Schätzer der Beste ist. Dies ist eine schwierige Frage, da der wirkliche Wert von  $\lambda$  nicht bekannt ist. In diesem Kurs kann darauf nur mit einem “Küchenrezept” eingegangen werden. Es soll der Schätzer gewählt werden, der erstens erwartungstreu ist und zweitens die kleinste Varianz aufweist.

**Beispiel 50** Im Kontext von Beispiel 1 sei versucht, die Proportion  $p$  der toleranzhaltigen Bolzen abzuschätzen. Wir setzen  $X(\omega) = 0$  falls der Bolzen  $\omega$  ausserhalb der Toleranz liegt und  $X(\omega) = 1$  sonst.  $X$  ist eine Bernoulli-Variable von der der Parameter  $p$  geschätzt werden soll. Der Schätzer für den Erwartungswert  $\bar{X} = \frac{1}{n} \sum X_i$  für eine Stichprobe der Grösse  $n$  ist erwartungstreu

$$E(\bar{X}) = \frac{\sum_{i=1}^n (0 \cdot (1-p) + 1 \cdot p)}{n} = p.$$

$\bar{X}$  misst den relativen Anteil der Ausprägung  $X$  in der Stichprobe. Es kann gezeigt werden, dass  $\bar{X}$  der beste Schätzer für  $p$  ist (i.e.,  $\bar{X}$  ist erwartungstreu und hat minimale Varianz). ◇

**Beispiel 51** Obwohl  $S'^2$  ein erwartungstreuer Schätzer für die Varianz  $V(X)$  darstellt, ist die Standardabweichung der Stichprobe, also  $S'$ , kein erwartungstreuer Schätzer für  $\sigma = \sqrt{V(X)}$ . Es kann gezeigt werden, dass folgende Beziehung gilt

$$E(S') = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \sigma := c_4 \sigma.$$

Ein erwartungstreuer Schätzer der Standardabweichung ist entsprechend gegeben durch

$$\hat{\sigma} = \frac{S'}{c_4}$$

mit  $c_4$  wie oben definiert. ◇

**Beispiel 52** (Verallgemeinerung von Beispiel 50). Falls die Individuen der Population  $P$  eine Eigenschaft entweder besitzen oder nicht besitzen, kann der effektive relative Anteil  $p$  der Population mit der gesagten Ausprägung wie folgt abgeschätzt werden.

Sei  $F$  die Zufallsvariable “relativer Anteil der Individuen der Stichprobe mit Ausprägung”.  $F$  ist ein erwartungstreuer Schätzer mit Varianz  $\frac{p(1-p)}{n}$ . Er ist also umso besser, je grösser der Stichprobenumfang  $n$  ist. ◇

## 12.2 Schätzung durch Konfidenzintervalle

Bei der Punktschätzung von  $\lambda$  ändert der Wert des Schätzers  $\hat{\lambda}$  bei jeder Stichprobe, ohne dass wir wissen, ob wir nun nahe oder weit vom wirklichen Wert entfernt sind. Es ist vernünftiger die Fragestellung bescheidener anzusetzen, um eine Antwort in der Form “der wahre Wert  $\lambda$  liegt mit grosser Sicherheit zwischen den Werten  $\hat{\lambda}_1$  und  $\hat{\lambda}_2$ ”, wobei  $\hat{\lambda}_1$  und  $\hat{\lambda}_2$  die Werte zweier zu definierenden Statistiken sind, zu erhalten.

**Beispiel 53** Wir betrachten wiederum die Tagesproduktion von Bolzen aus Beispiel 47. Unser Schätzer für den Durchmesser ist  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Seien für  $n = 10$  folgende zwei Stichproben gezogen worden (Angaben in [mm]):

$$x_1 = 20.1 \quad x_2 = 20.2 \quad x_3 = 20.0 \quad x_4 = 20.1 \quad x_5 = 20.2 \\ x_6 = 19.8 \quad x_7 = 19.9 \quad x_8 = 20.0 \quad x_9 = 20.1 \quad x_{10} = 19.9$$

und

$$x_1 = 20.0 \quad x_2 = 19.7 \quad x_3 = 20.1 \quad x_4 = 20.0 \quad x_5 = 19.9 \\ x_6 = 19.9 \quad x_7 = 19.8 \quad x_8 = 20.1 \quad x_9 = 20.0 \quad x_{10} = 19.9$$

Für die erste Stichprobe folgt  $\hat{\lambda} = 20.03$  und für die zweite  $\hat{\lambda} = 19.94$ . Der Durchmesser wird also irgendwo um 20mm liegen. Aber wie sicher ist diese Aussage? Konfidenzintervalle präzisieren diese Frage und geben entsprechend Antwort.  $\diamond$

**Zweiseitige Konfidenzintervalle.** Wir geben uns eine kleine Wahrscheinlichkeit  $0 < \alpha < 1$  vor (zum Beispiel  $\alpha = 0.01$  oder  $\alpha = 0.05$ ) und suchen zwei Statistiken

$$\Lambda_1 = f_1(X_1, X_2, \dots, X_n) \quad \text{und} \quad \Lambda_2 = f_2(X_1, X_2, \dots, X_n),$$

so dass gilt

$$P(\Lambda_1 \leq \Lambda \leq \Lambda_2) \geq 1 - \alpha.$$

**Interpretation.** Seien  $\hat{\lambda}_1$  und  $\hat{\lambda}_2$  die von  $\Lambda_1$  und  $\Lambda_2$  angenommenen Werte für eine konkrete Stichprobe. Die Wahrscheinlichkeit dass

$$\lambda \in [\hat{\lambda}_1, \hat{\lambda}_2]$$

gilt, ist grösser als  $1 - \alpha$ . Der Wert  $1 - \alpha$  ist ein Mass für das Vertrauen, das wir in die Aussage “ $\lambda$  ist im angegebenen Intervall” haben können.

Das Intervall  $[\hat{\lambda}_1, \hat{\lambda}_2]$  heisst zweiseitiges Konfidenzintervall für  $\lambda$ , mit **Konfidenzkoeffizient**  $1 - \alpha$  oder kurz mit **Niveau**  $\alpha$ .

**Einseitige Konfidenzintervalle.** Einseitige Konfidenzintervalle (rechtsseitig  $[\Lambda_1, \infty[$ , linksseitig  $] - \infty, \Lambda_2]$ ) sind Grenzfälle zweiseitiger Intervalle. Wir setzen

- Ein **einseitiges linkes Konfidenzintervall** erfüllt die Bedingung

$$P(\lambda \leq \Lambda_2) \geq 1 - \alpha.$$

Nimmt  $\Lambda_2(x_1, \dots, x_n)$  für eine Stichprobe  $x_1, \dots, x_n$  den Wert  $\hat{\lambda}_2$  an, kann mit  $(1 - \alpha)100\%$  Sicherheit gesagt werden, der wahre Wert von  $\lambda$  ist im Konfidenzintervall  $] - \infty, \hat{\lambda}_2]$ .

- Ein **einseitiges rechtes Konfidenzintervall** erfüllt die Bedingung

$$P(\Lambda_1 \leq \lambda) \geq 1 - \alpha.$$

Nimmt  $\Lambda_1(x_1, \dots, x_n)$  für eine Stichprobe  $x_1, \dots, x_n$  den Wert  $\hat{\lambda}_1$  an, kann mit  $(1 - \alpha)100\%$  Sicherheit gesagt werden, der wahre Wert von  $\lambda$  ist im Konfidenzintervall  $[\hat{\lambda}_1, \infty[$ .

Wie Konfidenzintervalle gefunden werden, sei nun anhand von einigen klassischen Beispielen aufgezeigt.

### 12.2.1 Konfidenzintervall für den Erwartungswert

Gegeben sei eine Population, deren Individuen eine uns interessierende Ausprägung  $X$  mit Erwartungswert  $E(X) = \mu$  und Varianz  $V(X) = \sigma^2$  tragen. Wir stellen uns die Aufgabe ein zweiseitiges Konfidenzintervall für den Erwartungswert  $\mu$  zum Niveau  $\alpha$  anzugeben. Wir wollen also eine Aussage der Form:

Der (wahre) Erwartungswert  $\mu$  liegt mit einer Sicherheit von  $(1 - \alpha)100\%$  im Konfidenzintervall.

Der Punktschätzer für den Erwartungswert ist  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  wobei  $n$  der Umfang der Stichprobe ist. Das zu suchende Intervall wird sich um  $\bar{X}$  ausdehnen und diese Ausdehnung wird natürlich auf eine gewisse Art von der Standardabweichung  $\sigma$  abhängen. Es sind zwei Fälle zu unterscheiden.

**(1) Fall mit bekanntem  $\sigma$ .** Die entscheidende Zufallsvariable ist

$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Nämlich deshalb, weil wir für grosse  $n$  ( $n \geq 30$ ) die Verteilung von  $U$  annäherungsweise kennen. Dank dem zentralen Grenzwertsatz ist  $U$  in der Tat normalverteilt  $\mathcal{N}(0, 1)$ ! Zwar ist dies strenggenommen nur dann der Fall, wenn  $X$  normalverteilt ist; für grosse  $n$  stimmt das aber auch annähernd ohne diese Annahme. Es ist also für  $u \in \mathbb{R}$ ,

$$P(U \leq u) = \phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-x^2/2} dx$$

Wir definieren nun  $-u_{\alpha/2} = u_{1-\alpha/2}$  durch Festlegen der Fläche unter der Glockenkurve (siehe Abbildung 12.1).

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

Es gilt entsprechend

$$P(u_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq u_{1-\alpha/2}) = 1 - \alpha$$

und wir finden das zweiseitige Konfidenzintervall zum Niveau  $\alpha$

$$\bar{x} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$  ist dabei das beobachtete Mittel der Stichprobe.

Für ein einseitiges Intervall zum Niveau  $\alpha$  finden wir analog

links	$\mu \leq \bar{x} + u_{1-\alpha} \frac{\sigma}{\sqrt{n}}$
rechts	$\mu \geq \bar{x} - u_{1-\alpha} \frac{\sigma}{\sqrt{n}}$

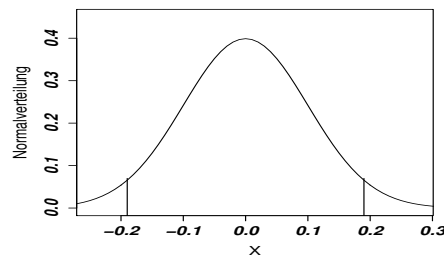


Abbildung 12.1: Dichte der Normalverteilung mit  $\mu = 0$  und  $\sigma = 1$ . Zwischen den beiden vertikalen Strichen und dem Graph befindet sich  $1 - \alpha$  Prozent der Fläche.

Es sei bemerkt, dass für einseitige Intervalle der Wert  $u_{1-\alpha}$  und nicht  $u_{1-\alpha/2}$  einzusetzen ist (warum?!).

**Beispiel 54** Es soll der Erwartungswert einer Zufallsvariablen  $X$  durch eine Stichprobe (Ziehen mit Zurücklegen) der Grösse  $n = 100$  abgeschätzt werden. Das beobachtete Stichprobenmittel ist  $\bar{x} = 43.05$  und die theoretische Varianz sei  $\sigma^2 = 0.25$ .

Der Erwartungswert  $\mu$  wird durch ein zweiseitiges Konfidenzintervall mit  $\alpha = 0.05$  abgeschätzt. Mit Hilfe der Tabelle (siehe Anhang) für die Normalverteilung oder mit Hilfe eines Rechners finden wir

$$u_{1-\alpha/2} = 1.96$$

Das zweiseitige Konfidenzintervall ergibt sich zu:

$$43.05 - 1.96 \frac{0.5}{10} = 42.952 \leq \mu \leq 43.148 = 43.05 + 1.96 \frac{0.5}{10}.$$

Der Erwartungswert von  $X$  befindet sich mit 95% Wahrscheinlichkeit in diesem Intervall.  $\diamond$

**(2) Fall mit unbekanntem  $\sigma$ .** Wir nehmen an, die Zufallsvariablen  $X_1, \dots, X_n$  die den  $n$  Beobachtungen zugehören, seien unabhängig und normalverteilt gemäss  $\mathcal{N}(\mu, \sigma^2)$ , wobei  $\mu$  **und**  $\sigma^2$  unbekannte Parameter sind.

Die Normalverteilung kann nicht herangezogen werden um die Wahrscheinlichkeiten von  $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  zu berechnen, da  $\sigma$  unbekannt ist. Diesem Umstand hat William Gosset, britischer Bierbrauer und Mathematiker, Abhilfe geleistet, da er die Verteilung für  $U$  ausrechnete. Die Idee besteht einfach darin  $\sigma$  durch seinen Schätzer  $S'$  zu ersetzen. Er publizierte sein Resultat unter dem Pseudonym Student. Die Verteilung ist auch unter dem Namen  $t$ -Verteilung bekannt.

**Theorem** (ohne Beweis). Seien  $X_1, \dots, X_n$   $n$  normalverteilte, unabhängige Zufallsvariablen mit Parameter  $\mathcal{N}(\mu, \sigma^2)$ . Die Zufallsvariable  $\frac{\bar{X} - \mu}{S'/\sqrt{n}}$  folgt einer Student-Verteilung (auch  $t$ -Verteilung) mit  $n - 1$  Freiheitsgraden<sup>1</sup>.

Wir erinnern, die Wahrscheinlichkeitsdichte der Student-Verteilung  $T_\nu$  mit  $\nu$  Freiheitsgraden ist dabei gegeben durch

$$f_{T_\nu}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu\pi}} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}, \nu \in \mathbb{N}.$$

<sup>1</sup>Achtung der Ausdruck ist nicht genau die standardisierte Zufallsvariable von  $\bar{X}$ , da  $\sigma$  durch seinen Schätzer  $S'$  ersetzt wurde.

Die Wahrscheinlichkeitsdichte ist ähnlich der einer standardisierten, normalverteilten Dichtefunktion die jedoch seitlich grössere Werte annimmt (siehe Anhang).

**Resultat.** Sei  $-t_{\alpha/2} = t_{1-\alpha/2}$  durch Vorgeben der Fläche unter der Studentverteilung fixiert

$$\int_{-\infty}^{t_{\alpha/2}} f_{T_{\nu-1}}(x)dx = \alpha/2 = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_{1-\alpha/2}} f_{T_{\nu-1}}(x)dx .$$

Es gilt

$$P(t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S'/\sqrt{n}} \leq t_{1-\alpha/2}) = 1 - \alpha$$

und das zweiseitige Konfidenzintervall zum Niveau  $\alpha$  ist entsprechend

$$\boxed{\bar{x} - t_{1-\alpha/2} \frac{s'}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{1-\alpha/2} \frac{s'}{\sqrt{n}}}$$

wobei  $\bar{x}$  der Mittelwert und  $s'$  die Standardabweichung der Stichprobe ist.

Für ein einseitiges Konfidenzintervall zum Niveau  $\alpha$  gilt

$$\boxed{\begin{array}{l} \text{linksseitig} \quad \mu \leq \bar{x} + t_{1-\alpha} \frac{s'}{\sqrt{n}} \\ \text{rechtsseitig} \quad \bar{\mu} \geq \bar{x} - t_{1-\alpha} \frac{s'}{\sqrt{n}} \end{array}}$$

Es sei bemerkt, dass für einseitige Intervalle der Wert  $t_{1-\alpha}$  und nicht  $t_{1-\alpha/2}$  einzusetzen ist (warum?!).

### 12.2.2 Konfidenzintervall für die Varianz

Gegeben sei eine Population, deren Individuen eine uns interessierende Ausprägung  $X$  mit Erwartungswert  $E(X) = \mu$  und Varianz  $V(X) = \sigma^2$  tragen. Wir stellen uns die Aufgabe ein zweiseitiges Konfidenzintervall für die Varianz  $\sigma^2$  zum Niveau  $\alpha$  anzugeben. Wir wollen also eine Aussage der Form:

Die (wahre) Varianz  $\sigma^2$  liegt mit einer Sicherheit von  $(1 - \alpha)100\%$  im Konfidenzintervall.

Der Punktschätzer für die Varianz ist  $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  wobei  $n$  der Umfang der Stichprobe ist. Das zu suchende Intervall wird sich um  $\bar{S}'^2$  ausdehnen.

**Konfidenzintervall für  $\sigma^2$  einer normalverteilten Zufallsvariable.** Die einschneidenste Änderung bezüglich der Schätzung des Mittelwertes ist, dass die Hypothese  $X$  sei normalverteilt, nicht einfach fallen gelassen werden kann, auch wenn  $n \geq 30$  ist! Das kommt daher, dass die entscheidende Zufallsvariable, nennen wir sie  $Y$ , Chi<sup>2</sup>-verteilt ist. In der Tat folgt die Variable

$$Y = \frac{(n-1)S'^2}{\sigma^2}$$

einer Chi<sup>2</sup> Verteilung mit  $n - 1$  Freiheitsgraden. Definieren wir ähnlich wie  $u_\alpha$  und  $t_\alpha$  die beiden Intervallgrenzen  $\chi_{\alpha/2, n-1}^2$  und  $\chi_{1-\alpha/2, n-1}^2$  durch

$$P(\chi_{\alpha/2, n-1}^2 \leq Y \leq \chi_{1-\alpha/2, n-1}^2) = 1 - \alpha$$

so ist das Konfidenzintervall zum Niveau  $\alpha$

$$\boxed{\frac{(n-1)s'^2}{\chi_{1-\alpha/2, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s'^2}{\chi_{\alpha/2, n-1}^2}}$$

Für ein einseitiges Konfidenzintervall zum Niveau  $\alpha$  gilt

linksseitig	$\sigma^2 \leq \frac{(n-1)s'^2}{\chi_{\alpha, n-1}^2}$
rechtsseitig	$\sigma^2 \geq \frac{(n-1)s'^2}{\chi_{1-\alpha, n-1}^2}$

### 12.2.3 Konfidenzintervall für Anteilswert

In diesem Abschnitt gehen wir auf eine häufig auftretende Situation ein. Gegeben sei eine Population deren Individuen eine uns interessierende Ausprägung entweder haben oder nicht haben (man denke an die Bolzen die entweder toleranzhaltig sind oder nicht). Wir wollen den Anteil  $p$  der Individuen abschätzen, die die Ausprägung besitzt. Dazu definieren wir die Bernoulli Zufallsvariable  $X_i$ , die den Wert 1 (resp. 0) annimmt, falls die  $i$ te Ziehung ein Individuum liefert, das die Ausprägung besitzt (resp. nicht besitzt). Erfolgt die Ziehung der Individuen mit Zurücklegen, so sind die  $X_i$  unabhängige, bernoulliverteilte Zufallsvariablen mit Parameter  $p$ . Entsprechend ist

$$S_n := \sum_{i=1}^n X_i$$

binomialverteilt mit Parameter  $(n, p)$ .  $S_n$  zählt die Anzahl Individuen einer Stichprobe vom Umfang  $n$ , die die Ausprägung besitzen.

**Bemerkung.** Erfolgt die Ziehung ohne Zurücklegen, so folgt  $S_n$  einer hypergeometrischen Verteilung. Die Formeln müssen entsprechend modifiziert werden. Wir gehen nicht weiter auf diese Variante ein.

Der Anteilswert  $p$  wird nun durch die Statistik

$$\bar{X} = \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

geschätzt. Wir erinnern: die möglichen Werte von  $\bar{X}$  sind  $\frac{k}{n}$  mit  $k \in \{0, 1, \dots, n\}$  und die Verteilung ist

$$P(\bar{X} = \frac{k}{n}) = P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}.$$

$\bar{X}$  ist der beste Punktschätzer für  $p$  (erwartungstreu und kleinste Varianz). Trotzdem kann natürlich der wahre Wert des relativen Anteils  $p$  sowohl kleiner, als auch größer als der Punktschätzer sein. Mit Sicherheit gilt nur, dass  $p$  jeden Wert zwischen 0 und 1 annehmen kann. Gemäss dem Gedankengang der Konfidenzintervalle suchen wir nun ein Intervall  $[p_u, p_o]$  so, dass die Aussage

$$p \in [p_u, p_o]$$

mit hoher Wahrscheinlichkeit richtig ist (d.h. beim vielfachen Wiederholen des Verfahrens wird das berechnete Konfidenzintervall in den meisten Fällen den Parameter  $p$  enthalten). Wie oft das der Fall sein soll, wird mittels der Vertrauenswahrscheinlichkeit  $1 - \alpha$  ausgedrückt. Wird zum Beispiel  $1 - \alpha = 0.95$  gewählt, so ist in 95% der Fälle die Aussage  $p \in [p_u, p_o]$  richtig.

Für eine bestimmte Stichprobe  $x_1, \dots, x_n$  nimmt der Punktschätzer einen bestimmten Wert  $\hat{p} = \frac{\hat{k}}{n}$  an, wobei  $\hat{k} = S_n(x_1, \dots, x_n)$  ist. Basierend auf dieser Beobachtung lassen sich nun die Grenzen  $p_u$  und  $p_o$  wie folgt bestimmen:



Die **untere Grenze**  $p_u$  ist Lösung der Gleichung

$$P(S_n \geq \hat{k}) = \sum_{i=\hat{k}}^n \binom{n}{i} p_u^i (1 - p_u)^{n-i} = \alpha/2$$

Dies bedeutet folgendes:

Wenn die Wahrscheinlichkeit **mindestens**  $\hat{k}$  “Treffer” zu erzielen für einen fixen Anteilswert  $p$  die Grenze  $\alpha/2$  unterschreitet, so kann mit einer Sicherheit von  $1 - \alpha/2$  **ausgeschlossen** werden, dass  $p$  der gesuchte Anteilswert ist. **Somit ist  $p_u$  der kleinste Wert von  $p$ , bei dem noch angenommen werden kann, dass “mindestens  $\hat{k}$  Treffer” eine “realistische” Beobachtung ist.** Für kleinere Werte von  $p$  erscheint diese Beobachtung zu unwahrscheinlich.

Die **obere Grenze**  $p_o$  ist Lösung der Gleichung

$$P(S_n \leq \hat{k}) = \sum_{i=0}^{\hat{k}} \binom{n}{i} p_o^i (1 - p_o)^{n-i} = \alpha/2$$

Dies bedeutet folgendes:

Wenn die Wahrscheinlichkeit **höchstens**  $\hat{k}$  “Treffer” zu erzielen für einen fixen Anteilswert  $p$  die Grenze  $\alpha/2$  unterschreitet, so kann mit einer Sicherheit von  $1 - \alpha/2$  ausgeschlossen werden, dass  $p$  der gesuchte Anteilswert ist. **Somit ist  $p_o$  der größte Wert von  $p$ , bei dem noch angenommen werden kann, dass “höchstens  $\hat{k}$  Treffer” eine “realistische” Beobachtung ist.** Für größere Werte von  $p$  erscheint dies zu unwahrscheinlich.

**Konkrete Berechnung von  $p_u$  und  $p_o$ .** Sei  $X$  eine binomialverteilte Zufallsvariable mit den Parametern  $n$  und  $p$ . Ganz allgemein muss für ein  $a$  die Gleichung  $P(X \leq k) = a$  gelöst werden. Dies geschieht entweder mit Hilfe von Tabellen oder numerisch. Hier einige gebräuchliche Methoden:

- (i) Die Gleichung  $P(X \leq k) = \alpha/2$  lässt sich mit der **EXCEL** - Funktion “BETAINV” nach  $p$  auflösen. Es ist

$$p_o = \text{BETAINV}(1 - \alpha/2; k + 1; n - k).$$

Mit der Beziehung  $P(X \geq k - 1) = P(X > k) = 1 - P(X \leq k)$  folgt entsprechend

$$p_u = \text{BETAINV}(\alpha/2; k; n - k + 1).$$

- (ii) In **Mathcad** kann man  $p$  mit Hilfe der Funktion “qbeta” berechnen:

$$p = \text{qbeta}(1 - \alpha/2; k + 1; n - k)$$

- (iii) **Maple** und **Mathematika** lösen diese Gleichung auch ohne Probleme mit dem Solve-Befehl. Hilfreich ist auch immer eine graphische Darstellung des Problems (siehe Übungen).

**Beispiel 55** Zu untersuchen ist der Bekanntheitsgrad eines Produktes. Dazu wird eine einfache Zufallsstichprobe vom Umfang 1000 erhoben. Von diesen 1000 Personen ist das Produkt 320 Personen bekannt, 680 unbekannt. Zu bestimmen ist ein Konfidenzintervall zum Niveau  $\alpha = 0.05$ .

**Lösung** Der Punktschätzer liefert

$$\bar{x} = \frac{\hat{k}}{n} = \frac{320}{1000} = 0.32$$

Berechnung des Konfidenzintervalls:  $p_u$  löst

$$\sum_{i=320}^{1000} \binom{1000}{i} p_u^i (1-p_u)^{1000-i} = 0.025$$

Irgend eines der oben genannten numerischen Verfahren liefert  $p_u = 0.291$ . Ähnlich ist  $p_o$  Lösung von

$$\sum_{i=0}^{320} \binom{1000}{i} p_o^i (1-p_o)^{1000-i} = 0.025$$

und wir finden  $p_o = 0.350$ . Folglich ist der wahre Anteilwert  $p$  mit 95%tiger Sicherheit im Intervall  $[0.291; 0.350]$ .  $\diamond$

**Beispiel 56** Es sollen gespritzte Zahnimplantate dynamisch getestet werden und daraus der Anteil  $p$  der fehlerhaften Implantate zu einem Niveau  $\alpha$  abgeschätzt werden.

Aus wirtschaftlichen Gründen können nur 20 Implantate mit dynamischer Wechsellast geprüft werden. Konkret werden dabei 20 Implantate mit 360 Newton während 5 Mio. Zyklen belastet.

Bei einer konkreten Durchführung ergab sich, dass alle Implantate der Belastung stand hielten. Geben Sie die (einseitige) obere Schranke des Konfidenzintervalls für  $p$  zum Niveau  $\alpha = 0.05$  an.

**Lösung** Der Punktschätzer liefert

$$\bar{x} = \frac{\hat{k}}{n} = \frac{0}{20} = 0$$

Es ist offensichtlich, dass der Schätzer zu kleine Werte für  $p$  liefert. Wir suchen also ein Intervall der Form  $[0, p_o]$ . Die obere Schranke  $p_o$  ist Lösung von

$$\sum_{i=0}^{20} \binom{20}{i} p_o^i (1-p_o)^{20-i} = (1-p_o)^{20} = 0.05$$

(Achtung, hier  $\alpha$  verwenden und nicht  $\alpha/2$ ) und wir finden  $p_o = 0.139$ . Folglich ist der wahre Anteilwert  $p$  mit 95%tiger Sicherheit im Intervall  $[0; 0.139]$ .  $\diamond$

**Näherungsformel für grosse  $n$  und nicht zu kleine  $p$ .**

Entsprechend dem zentralen Grenzwertsatz folgt die Zufallsvariable  $\frac{\bar{X}-p}{\sigma/\sqrt{n}}$  mit  $\sigma = \sqrt{p(1-p)}$  für grosse  $n$  annähernd einer standardisierten Normalverteilung. Entsprechend kann das Konfidenzintervall für normalverteilte Zufallsvariablen mit bekanntem  $\sigma$  verwendet werden (da für grosse  $n$  der Ausdruck  $\sqrt{\hat{p}(1-\hat{p})}$  eine gute Näherung für  $\sigma$  ist)!

Ist nun  $\hat{p} = \frac{\hat{k}}{n}$  der Wert von  $\bar{X}$  für eine konkrete Realisierung der Ziehung und gelten die Beziehungen

$$n \geq 30 \quad \text{und} \quad n\hat{p}(1-\hat{p}) \geq 10$$

so kann folgendes Konfidenzintervall für  $p$  verwendet werden:

$$\hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Dabei bedeutet  $u_\alpha$  das  $\alpha$ -Quantil der standardisierten Normalverteilung (i.e.  $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_\alpha} \exp(-x^2/2) dx = \alpha$ ).

**Bemerkung.** Erfolgt die Ziehung ohne Zurücklegen in einer Population der Grösse  $N$ , müssen die Bedingungen und die Intervallformel wie folgt modifiziert werden:

Gilt

$$n \geq 30 \text{ und } n\hat{p}(1 - \hat{p}) \geq 10 \text{ und } n/N \leq 0.5,$$

ist das Konfidenzintervall für  $p$

$$\hat{p} - u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N-1}} \leq p \leq \hat{p} + u_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N-1}}$$

## 12.3 Useful Matlab codes

Confidence intervals can be estimated using bootstrapping. The `bootstrap` is a procedure that involves choosing random samples with replacement from a data set and analyzing each sample the same way. Sampling with replacement means that every sample is returned to the data set after sampling. So a particular data point from the original data set could appear multiple times in a given bootstrap sample. The number of elements in each bootstrap sample equals the number of elements in the original data set. Here a few examples

- Bootstrapping the 0.95 confidence interval (ci) for the mean of a given data set  $y$ :

```
ci = bootci(2000, {@mean,y}, 'alpha', 0.05);
```

the vector  $ci$  contains the upper and lower bound of the confidence interval.

```
data=importdata('filename');
```

- Similarly for the standard deviation and the variance:

```
ci = bootci(2000, {@std,y}, 'alpha', 0.05);
```

```
ci = bootci(2000, {@var,y}, 'alpha', 0.05);
```

- Similarly for the correlation coefficient of two data sets  $x$  and  $y$  the ci-bounds can be estimated using:

```
ci = bootci(2000, {@corr,x,y}, 'alpha', 0.05);
```

## 12.4 Aufgaben

**Übung 58** Bestimmen Sie den Umfang einer Stichprobe, bei der der relative Anteil  $p$  einer Ausprägung in der Grundgesamtheit so abgeschätzt werden soll, dass die Standardabweichung kleiner als 1% des theoretischen Wertes  $p$  ist. ◉

**Übung 59** Ein Unternehmen möchte flächendeckend auf dem Markt ein neues Spülmittel einführen. Um die Käuferakzeptanz auszuloten, wird in einem Supermarkt dieses Produkt mit hohem Werbeaufwand platziert. Es soll mit dieser Aktion der durchschnittliche tägliche Absatz in einem Supermarkt dieser Größe geschätzt werden. Man definiert nun den täglichen Absatz als Zufallsvariable  $X$  [Stück] mit den unbekanntem Parametern Erwartungswert  $\mu$  und der Varianz  $\sigma^2$ . Man geht auf Grund langjähriger Beobachtungen hier davon aus, dass  $X$  annähernd normalverteilt ist. Die

Marktforschungsabteilung hat einen Konfidenzkoeffizienten von 0,95 als ausreichend erachtet. Es wird nun 16 Tage lang der tägliche Absatz erfasst. Es hat sich ergeben

$X$  : 110 112 106 90 96 118 108 114 107 90 85 84 113 105 90 104

Unter der Annahme einer normalverteilter Grundgesamtheit mit bekannter Varianz  $\sigma^2 = 121$  soll das Konfidenzintervall für den Erwartungswert zum Niveau  $\alpha = 0.05$  angegeben werden.  $\odot$

**Übung 60** Ein Unternehmen möchte flächendeckend auf dem Markt ein neues Spülmittel einführen. Um die Käuferakzeptanz auszuloten, wird in einem Supermarkt dieses Produkt mit hohem Werbeaufwand platziert. Es soll mit dieser Aktion der durchschnittliche tägliche Absatz in einem Supermarkt dieser Größe geschätzt werden. Man definiert nun den täglichen Absatz als Zufallsvariable  $X$  [Stück] mit den unbekanntem Parametern Erwartungswert  $\mu$  und der Varianz  $\sigma^2$ . Man geht auf Grund langjähriger Beobachtungen hier davon aus, dass  $X$  annähernd normalverteilt ist. Die Marktforschungsabteilung hat einen Konfidenzkoeffizienten von 0,95 als ausreichend erachtet. Es wird nun 16 Tage lang der tägliche Absatz erfasst. Es hat sich ergeben

$X$  : 110 112 106 90 96 118 108 114 107 90 85 84 113 105 90 104

Unter der Annahme einer normalverteilten Grundgesamtheit mit unbekannter Varianz soll das Konfidenzintervall für den Erwartungswert zum Niveau  $\alpha = 0.05$  angegeben werden.  $\odot$

**Übung 61** Es bezeichne  $p$  die Proportion der nicht toleranzhaltigen Bolzen bezüglich der Gesamtzahl von produzierten Bolzen, wie wir sie im Beispiel 1 eingeführt haben. Die Stichprobe habe einen Umfang  $n = 100$ , (und wir können Dank der grossen Populationsmenge annehmen, dass wir sie mit Zurücklegen ziehen, obwohl dies in der Praxis nicht der Fall ist). In der Stichprobe seien 20 defekte Stücke gefunden worden (dies erlaubt uns die Binomialverteilung, die  $p$  folgt, durch eine Normalverteilung anzunähern). Es soll ein Konfidenzintervall zum Niveau  $\alpha = 0.05$  für den Parameter  $p$  angegeben werden.  $\odot$

**Übung 62** Strassenlärm kann durch gezieltes dotieren von Asphalt mit Kunststoffbeimischungen reduziert werden. Dabei wird das viskose Verhalten des Asphalts verändert. Nachstehende Tabelle zeigt das Viskositätsverhalten von 15 Asphaltmischungen. Unter der Annahme, dass die Viskosität normalverteilt ist, soll sowohl für den Mittelwert wie auch für die Varianz ein Konfidenzintervall zum Niveau  $\alpha = 0.05$  angegeben werden.

<i>Probe</i>	<i>Viskosität</i>	<i>Probe</i>	<i>Viskosität</i>
1	3193	8	2979
2	3124	9	3182
3	3153	10	3227
4	3145	11	3256
5	3093	12	3332
6	3466	13	3204
7	3355	14	3282
		15	3170

**Übung 63** Vor einem Wahllokal werden am Tag der Abstimmung in einer Stichprobe 30 zufällig ausgewählte Personen befragt, ob sie für oder gegen die Vorlage gestimmt haben. Dabei geben 21 Personen an, für die Vorlage gestimmt zu haben. Berechnen sie das 95% Konfidenzintervall für den Anteil aller Personen, welche für die Vorlage gestimmt haben.  $\odot$

**Übung 64** Lösen Sie Beispiel 55 mit Hilfe der Näherungsformel.  $\odot$

**Übung 65** Kann Beispiel 56 mit Hilfe der Näherungsformel gelöst werden? Was ergäbe Übung 63 mit der Näherungsformel (ohne Zurücklegen)?  $\odot$

# Kapitel 13

## Statistischer Test

Der statistische Test (auch Signifikantstest) ist ein wichtiges Werkzeug der schliessenden Statistik. Ein solcher Test liefert als Resultat die Akzeptanz oder das Verwerfen der Annahme die getestet werden soll. Solche Tests finden oft da Anwendung, wo fundierte Entscheidungen getroffen werden müssen. Sie bilden eine wichtige Entscheidungshilfe im wissenschaftlichen und industriellen Alltag.

### 13.1 Parametrischer Test

**Einführung.** Es sei eine Zufallsvariable mitsamt ihrer Verteilung gegeben. Die Verteilungsfunktion hänge nun aber von einem unbekanntem Parameter ab, den es zu ermitteln gilt. Es werden Annahmen über den Parameter gemacht und mittels Stichproben getestet, ob die Annahmen sinnvoll sind.

Dies ist eine klassische Situation wo statistische Tests eingesetzt werden. Der Test heisst parametrisch, da die Verteilung bis auf einen (oder mehrere) Parameter bekannt ist.

**Beispiel 57** Wir wollen wissen ob eine gewisse Münze eines Münzwurfspiels prepariert ist oder nicht. Falls die Münze nicht prepariert ist, ist der Ausgang des Spiels eine bernoulliverteilte Zufallsvariable  $X$  mit Parameter  $p = 1/2$ . Wir wollen also testen, ob  $p$  wirklich  $1/2$  ist oder ob  $p$  einen anderen Wert annimmt. Wir testen die Hypothese  $H_0 : p = 1/2$  gegen die Alternative  $H_1 : p \neq 1/2$ . Der Test basiert nun auf einer Sequenz von Würfeln  $(x_1, \dots, x_n)$  des entsprechenden Würfels und der statistische Test, eine Funktion der Stichprobe, bestehe hier aus der relativen Anzahl der Münzwürfe die Kopf zeigen. Ist dieser Wert innerhalb einer (noch zu bestimmenden) kritischen Region, wird die Annahme  $H_0$  verworfen, ansonsten wird sie akzeptiert (siehe Abbildung 13.1).

◇

#### 13.1.1 Sprachgebrauch.

Sei  $X : \Omega \rightarrow \mathbb{R}$  die Zufallsvariable deren bekannte Verteilung von einem Parameter  $\theta$  abhängt. Die möglichen Werte von  $\theta$  bilden eine Teilmenge von  $\mathbb{R}$  bezeichnet mit  $\Theta$ . Es geht darum, den wirklichen Wert von  $\theta$  innerhalb von  $\Theta$  mit Hilfe von Stichproben des durch  $X$  beschriebenen Zufallsexperiments zu lokalisieren. Die Schätzung von  $\theta$  stützt sich dann auf eine Realisierung  $x_1, \dots, x_n$  der Stichprobe.

**Hypothese.** Die zu testende Hypothese  $H_0$  ist eine Aussage der Form: “ $\theta \in \Theta_0$ ” wobei  $\Theta_0$  eine

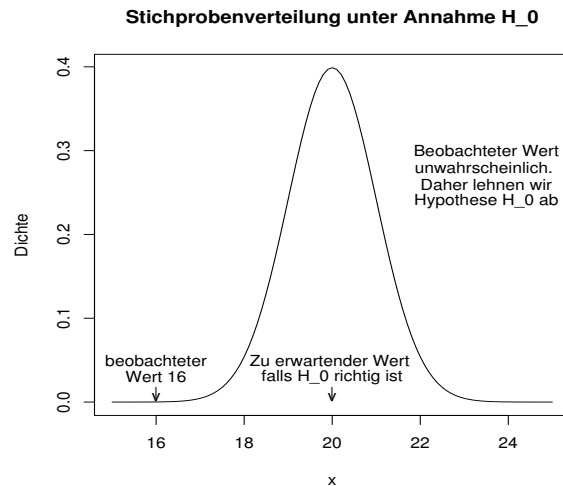


Abbildung 13.1: Grundidee: die Hypothese  $H_0$  wird verworfen, falls der beobachtete Wert unter der Annahme “ $H_0$  ist richtig” zu unwahrscheinlich ist. Hier, 40 Münzwürfe sollten im Mittel 20 Kopf- und 20 Zahlwürfe ergeben. 16 ist zu unwahrscheinlich für eine faire Münze. Schlussfolgerung:  $H_0$  wird verworfen.

Teilmenge von  $\Theta$  ist. Im Eingangs erwähnten Beispiel 57 ist  $\Theta_0 = \{1/2\}$  und der Parameter  $\theta$  ist die Grösse  $p$ .

$H_0$  heisst **Nullhypothese**. Die Negation von  $H_0$  heisst **Alternativhypothese**  $H_1$  und besagt gerade, dass  $\theta$  nicht in  $\Theta_0$  ist.

**Entscheidungsproblem.** Das Problem, das man sich stellt, ist zu entscheiden, ob  $H_0$  eine akzeptable Annahme ist oder nicht. Die Akzeptanz oder der Verwurf von  $H_0$  macht eine Entscheidung notwendig. Die Entscheidung basiert auf einer Stichprobe  $x_1, \dots, x_n$  von gewissem Umfang  $n$ .

**Test.** Es gibt zwei verschiedene Testtypen:

- (i) zweiseitiger Test  $H_0: \theta = \theta_0$  gegen  $H_1: \theta \neq \theta_0$
- (ii) unilateraler Test  $H_0: \theta \geq \theta_0$  gegen  $H_1: \theta < \theta_0$  oder  $H_0: \theta \leq \theta_0$  gegen  $H_1: \theta > \theta_0$ .

Beim zweiseitigen Test ist die Nullhypothese, wie im Beispiel 57, auf einen Punkt reduziert. Beim unilateralen Fall ist die Nullhypothese ein Intervall.

**Testfunktion.** Um zwischen  $H_0$  und  $H_1$  mit Hilfe einer Stichprobe  $x_1, \dots, x_n$  zu entscheiden, bedient man sich einer Testfunktion  $T(X_1, \dots, X_n)$ , die wir auch Teststatistik nennen. Der Test  $T$  wird in der Regel vom Statistiker bzw. der Statistikerin vorgegeben und nimmt Werte in  $\mathbb{R}$  an.

**Akzeptanzregion.** Wir teilen  $\mathbb{R}$  (Zielbereich der Teststatistik) in zwei Teile: einen Teil, der zur Akzeptanz von  $H_0$  führt, und einen Teil, der zur Akzeptanz von  $H_1$  führt.

Die Teilmenge  $A$  von  $\mathbb{R}$ , die zur Akzeptanz von  $H_0$  führt, heisst **Akzeptanzregion**. Das heisst nun, nimmt  $T$  für eine Stichprobe  $(x_1, \dots, x_n)$  einen Wert in  $A$  an, so wird  $H_0$  akzeptiert, ansonsten verworfen. Die komplementäre Teilmenge  $\bar{A}$  von  $A$ , die zum Verwurf von  $H_0$  führt, heisst **kritische Region**. Es gilt

$$H_0 \text{ ablehnen} \Leftrightarrow T(x_1, \dots, x_n) \notin A.$$

Die Akzeptanzregion ist typischerweise ein Konfidenzintervall.

**Beispiel 58** Nehmen wir Beispiel 57 auf.

Wir testen die Hypothese  $H_0 : p = 1/2$  gegen die Alternative  $H_1 : p \neq 1/2$ .

Der Test  $T(X_1, \dots, X_n)$  zähle die Anzahl der Kopfwürfe für  $n$  Würfe. Wird nun eine Sequenz  $(x_1, \dots, x_n)$  gewürfelt, nimmt  $T$  zum Beispiel den Wert 16 an. Ist die Akzeptanzregion um den Erwartungswert (unter der Hypothese  $H_0$  ist der Erwartungswert 20) angesetzt, könnte  $A$  wie folgt definiert sein  $A := [20 - 3, 20 + 3]$ . Der Test  $T$  nimmt für unser Beispiel einen Wert ausserhalb von  $A$  an. Die Annahme  $H_0$  wird entsprechend verworfen.  $\diamond$

**Fehlerrisiko.** Die Entscheidung Akzeptanz/Verwurf von  $H_0$  ist, als statistisches Resultat, mit einer natürlichen Fehlerwahrscheinlichkeit behaftet. Speziell sei angemerkt, dass **ein Nichtverwerfen der Nullhypothese nicht unbedingt bedeutet, dass sie zutrifft, sondern nur, dass sie nicht genügend unplausibel ist, um verworfen zu werden!**

Wir unterscheiden

- (i) **Fehler erster Art** der darin besteht  $H_0$  fälschlicherweise zu verwerfen (falscher Alarm).  
Kurz

$$H_0 \text{ wahr aber } T(x_1, \dots, x_n) \notin A$$

- (ii) **Fehler zweiter Art** der darin besteht  $H_0$  fälschlicherweise zu akzeptieren. Kurz

$$H_1 \text{ wahr aber } T(x_1, \dots, x_n) \in A$$

Das Fehlerrisiko ist die Wahrscheinlichkeit einen Fehler einzugehen. Es gilt

- (i) **Risiko erster Art** (Signifikanzniveau) ist die Wahrscheinlichkeit

$$\alpha = P(T(X_1, \dots, X_n) \notin A \mid H_0 \text{ wahr})$$

Diese Wahrscheinlichkeit kann vom Statistiker festgelegt werden.

- (ii) **Risiko zweiter Art** ist die Wahrscheinlichkeit

$$\beta = P(T(X_1, \dots, X_n) \in A \mid H_1 \text{ wahr})$$

Diese Wahrscheinlichkeit ist meist nicht bekannt und ist schwierig auszurechnen. Sie hängt von  $\alpha$ ,  $n$  und der wahren Grösse des Parameters  $\theta$  ab.

Entscheidung versus Sachlage	$H_0$ wahr	$H_1$ wahr
$H_0$ akzeptieren	$1 - \alpha$	$\beta$
$H_0$ ablehnen	$\alpha$	$1 - \beta$

**Bemerkung.** Die Fehlerwahrscheinlichkeit  $\alpha$  ist genauer gesagt das maximal akzeptable Fehlerisiko erster Art. Man nennt  $\alpha$  auch Signifikanzniveau. Dieses Niveau ist ad hoc vorgegeben und **quantifiziert die Vertrauenswürdigkeit bezüglich falschem Alarm.** Man wählt oft

$$\alpha = 0.05 \text{ oder } \alpha = 0.01$$



Per definitionem des Signifikanzniveaus ist die Wahrscheinlichkeit eines falschen Alarms kleiner oder gleich  $\alpha$ :

$$P(H_0 \text{ wahr aber } T(X_1, \dots, X_n) \in \bar{A}) \leq \alpha$$

Einmal  $\alpha$  festgelegt, geht es darum  $A$  und  $\bar{A}$  zu berechnen mit Hilfe der (oft angenäherten) Verteilung von  $T$ . Ist  $A$  ein Intervall, so entspricht  $\alpha$  gerade dem Konfidenzniveau und  $A$  ist das Konfidenzintervall zum vorgegebenen Niveau.

**Mächtigkeit eines Tests.** Als Macht eines Tests wird die Wahrscheinlichkeit bezeichnet, mit der ein Test eine falsche Nullhypothese entlarvt. **Die Mächtigkeit ist demnach die Wahrscheinlichkeit keinen Fehler zweiter Art zu begehen**, also gleich  $1 - \beta$ .

Ein Test mit geringer Macht hat wenig Wert, obwohl das Signifikanzniveau  $\alpha$  hoch sein kann (die Ergebnisse sind nicht schlüssig, da bei konstantem Nicht-Ablehnen der Nullhypothese diese nur mit geringer Wahrscheinlichkeit auch wirklich wahr ist). Die Macht eines Tests hängt unmittelbar mit dem Signifikanzniveau zusammen: Je größer  $\alpha$ , umso größer auch  $1 - \beta$  und umgekehrt. Man muss also einen geeigneten Kompromiss zwischen möglichst großer Testmacht einerseits (damit möglichst viele falsche Nullhypothesen verworfen werden können) und möglichst kleiner Irrtumswahrscheinlichkeit andererseits finden.

**Beispiel 59** (Zum besseren Verständnis). Ein Schüler muss für eine schriftliche Prüfung die Antworten auf 1000 (gleich schwierige) Fragen lernen. Beim Test werden aus Zeitgründen aber nur 10 Fragen vorgegeben. Der Lehrer wünscht sich, dass der Schüler, falls er "genug" (mindestens 500 Fragen) gelernt hat, durchkommt und dass er, falls er zu wenig gelernt hat (also höchstens 499 Fragen kann) den Test nicht besteht.

Hypothese  $H_0$  : Der Schüler hat genug gelernt ( $\geq 500$  Fragen).

Hypothese  $H_1$  : Der Schüler hat nicht genug gelernt ( $< 500$  Fragen).

**Design des Tests.** Der Lehrer überlegt vor der Prüfung, wie streng er bewerten soll. Er hat vor, den Schüler positiv zu benoten, falls dieser mindestens 5 der 10 Fragen richtig beantwortet. Welche Fehler können ihm bei dieser Bewertung unterlaufen?

$\alpha$ -Fehler: Auch wenn der Schüler eigentlich genug gelernt hat, besteht - sofern er nicht mindestens 995 der 1000 Fragen beantworten kann - die Gefahr, dass zum Test 6 oder mehr Fragen kommen, die er nicht gelernt hat. In diesem Fall würde  $H_0$  zu unrecht verworfen.

$\beta$ -Fehler: Der Schüler hat, sofern er wenigstens 5 Fragen gelernt hat, eine Chance, den Test zu bestehen, auch wenn er nicht die vom Lehrer erwarteten 500 Fragen kann. Tritt dies ein, so nimmt der Lehrer zu unrecht an, dass  $H_0$  gilt.  $\diamond$

### 13.1.2 Test für Mittelwert, Standardabweichung bekannt.

Sei  $X$  eine **gaussverteilte Zufallsvariable** mit unbekanntem Erwartungswert  $\mu$  und bekannter Standardabweichung  $\sigma$ . Es soll die Hypothese

$$H_0 : \mu = \mu_0$$

gegen

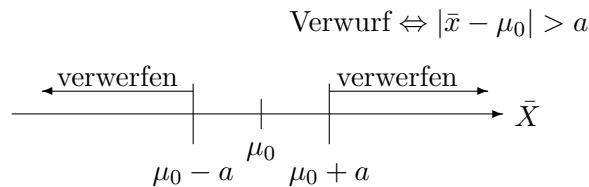
$$H_1 : \mu \neq \mu_0$$

getestet werden. Es wird entsprechend eine Stichprobe mit Umfang  $n$  gezogen und für den Mittelwert der Test

$$T(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

herangezogen.

Es scheint natürlich,  $H_0$  zu verwerfen, falls der Test  $T = \bar{X}$  einen von  $\mu_0$  weit entfernten Wert ergibt:



Es geht nun darum  $a$  (und entsprechend die Akzeptanzregion  $A = ]\mu_0 - a, \mu_0 + a[$ ) nach bestimmten Kriterien zu bestimmen.

$\bar{X}$  ist normalverteilt  $\mathcal{N}(\mu, \sigma^2)$  und wir können die Verwurfswahrscheinlichkeit ausdrücken als

$$\begin{aligned} P_\mu(|\bar{X} - \mu_0| \geq a) &= 1 - P_\mu(\mu_0 - a \leq \bar{X} \leq \mu_0 + a) \\ &= 1 - P_\mu\left(\frac{\mu_0 - a - \mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 + a - \mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

Dabei bedeutet die Schreibweise  $P_\mu$ , dass wir annehmen, der Erwartungswert von  $\bar{X}$  sei  $\mu$ . Die Zufallsvariable  $U := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  folgt (unter dieser Annahme) einer Normalverteilung  $\mathcal{N}(0, 1)$ <sup>1</sup>. Die Verwurfswahrscheinlichkeit, nennen wir sie  $\alpha_\mu$ , variiert als Funktion von  $\mu$  für fixes  $a$ . Es ist also

$$1 - P_\mu\left(\frac{\mu_0 - a - \mu}{\sigma/\sqrt{n}} \leq U \leq \frac{\mu_0 + a - \mu}{\sigma/\sqrt{n}}\right) =: \alpha_\mu$$

und speziell unter der Hypothese  $H_0$  (also falls  $\mu = \mu_0$  ist)

$$P_{\mu_0}(|\bar{X} - \mu_0| \geq a) = \alpha_{\mu_0} = P_{\mu_0}\left(|U| \geq \frac{a\sqrt{n}}{\sigma}\right).$$

Das Risiko  $H_0$  fälschlicherweise zu verwerfen (also  $\alpha_{\mu \neq \mu_0}$ ) ist also umso kleiner je grösser  $a$  ist und verschwindet gänzlich falls  $a$  unendlich gross gemacht wird (dabei wird aber  $1 - \beta$  unzulässig klein). Es ist angebracht zuerst das Signifikanzniveau  $\alpha$ , also das kleinste zu akzeptierende Risiko eines falschen Alarms (Risiko erster Art), unter der Hypothese  $H_0$  festzulegen. Es gilt also

$$\alpha_{\mu_0} = \alpha$$

für ein vorgegebenes  $\alpha$ . Wir bestimmen nun  $a$  so, dass

$$P(|U| \geq \frac{a\sqrt{n}}{\sigma}) = \alpha$$

gilt. Das ist natürlich gerade die Definition eines Konfidenzintervalles zum Niveau  $\alpha$  und wir finden

$$a = u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

<sup>1</sup>dies ist approximativ auch dann der Fall, wenn  $X$  nicht normalverteilt ist aber nur Erwartungswert  $\mu$  und Varianz  $\sigma^2$  hat! **Die Bedingung der Normalität ist also nicht notwendig, falls der Stichprobenumfang genügend gross ist, z.B.  $n > 30$ .**

und schliesslich für die Akzeptanzregion

$$A = \left[ \mu_0 - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu_0 + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Wird  $a$  und somit die Akzeptanzregion  $A$  grösser gewählt, sinkt zwar die Wahrscheinlichkeit eines Fehlers erster Art (alles wird immer akzeptiert, speziell auch  $H_0$ ), der Test wird aber immer weniger Aussagekraft haben, da das Risiko zweiter Art immer grösser wird! In der Tat, das Risiko zweiter Art (fälschliche Akzeptanz) ist definiert als

$$\beta_\mu = 1 - \alpha_{\mu \neq \mu_0} = P_\mu \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - \frac{a}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + \frac{a}{\sigma/\sqrt{n}} \right)$$

Welcher Wert  $\mu$  auch annehmen mag, das Risiko zweiter Art ist umso kleiner je kleiner  $a$  ist. In diesem Sinne ist die Akzeptanzregion das kleinst mögliche Intervall um  $\mu_0$  so, dass das Risiko erster Art (fälschliches Zurückweisen) gerade noch klein genug ist, um das Niveau zu respektieren.

**Resultat.** Die Hypothese  $\mu = \mu_0$  getestet auf einem Signifikanzniveau  $\alpha$  gegen die Alternative  $\mu \neq \mu_0$  ist zu verwerfen falls

$$\begin{array}{l} \bar{x} < \mu_0 - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \\ \text{oder} \\ \bar{x} > \mu_0 + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \end{array}$$

Das Risiko zweiter Art ist dabei

$$\beta_\mu = P \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - u_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + u_{1-\alpha/2} \right)$$

### Numerisches Beispiel.

- 1) Für einen ersten Test sei  $n = 25$ ,  $\mu_0 = 250$  und  $\sigma = 0.5$ .  $H_0$  ist zu akzeptieren falls  $249,80 \leq \bar{x} \leq 250,20$ , sonst muss die Hypothese verworfen werden. Für  $\mu = 250,20$  ist zum Beispiel  $\beta = 0.48$  mit  $\alpha = 0.05$ .
- 2) Für einen zweiten Test sei zum Beispiel  $n = 52$ ,  $\mu_0 = 250$  und  $\sigma = 0.5$ .  $H_0$  ist zu akzeptieren falls  $249,86 \leq \bar{x} \leq 250,14$ , sonst muss die Hypothese verworfen werden. Für  $\mu = 250,20$  ist zum Beispiel  $\beta = 0.18$  mit  $\alpha = 0.05$ .

### Bemerkung zum Risiko zweiter Art und Stichprobenumfang.

Wie angedeutet kann das Risiko zweiter Art oft nicht ausgerechnet werden. In dem einfachen Fall eines Mittelwerttests haben wir jedoch die explizite Formel

$$\beta_\mu = P \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - u_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + u_{1-\alpha/2} \right)$$

$\beta_\mu$  hängt von  $\alpha$ ,  $n$  und  $\mu_0 - \mu$  ab. Dies ist charakteristisch für das Risiko zweiter Art (fälschliche Akzeptanz). Wir haben folgendes Verhalten

- (i) Je weiter der wahre Wert  $\mu$  von der Hypothese  $\mu_0$  entfernt ist (je grösser also  $\mu_0 - \mu$  ist), desto kleiner ist  $\beta_\mu$ . Der Test wird also für gegebenen Stichprobenumfang und gegebenes  $\alpha$  grosse Differenzen leichter aufdecken als kleine, da dann das Risiko zweiter Art kleiner ist.
- (ii) Die Vergrösserung des Stichprobenumfangs verkleinert das Risiko zweiter Art bei gleichbleibendem Signifikanzrisiko und gleichem Abstand  $\mu_0 - \mu$ . Wollen wir also genauer testen, können wir die Macht des Tests durch Erhöhen des Stichprobenumfangs verbessern.

Beispiel: Sei  $\sigma = 1$ ,  $\alpha = 0.05$  und  $H_0 : \mu = 16.00$ . Falls der wahre Wert  $\mu = 16.05$  ist, dann folgt

$$\begin{aligned}\beta_\mu &= P(-0.5\sqrt{n} - 1.96 \leq U \leq -0.5\sqrt{n} + 1.96) \\ &= \phi(-0.5\sqrt{n} + 1.96) - \phi(-0.5\sqrt{n} - 1.96)\end{aligned}$$

für  $n \rightarrow \infty$  geht dieser Ausdruck gegen Null.

### 13.1.3 Test für Mittelwert, Standardabweichung unbekannt.

Sei  $X$  eine gaussverteilte Zufallsvariable (für genügend grosse Stichproben kann diese Annahme auch fallen gelassen werden, z.B.  $n > 30$ ) mit unbekanntem Erwartungswert  $\mu$  und unbekannter Standardabweichung  $\sigma$ . Es soll die Nullhypothese

$$H_0 : \mu = \mu_0$$

gegen die Alternative

$$H_1 : \mu \neq \mu_0$$

getestet werden. Es wird eine Stichprobe der Grösse  $n$  entnommen und wir bilden den statistischen Test

$$T(X_1, X_2, \dots, X_n) = \frac{\bar{X} - \mu}{S'/\sqrt{n-1}}$$

wobei  $S'^2$  die erwartungstreue Varianz der Stichprobe ist,

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$T$  folgt einer Studentverteilung mit Freiheitsgrad  $n-1$  und der zweiseitige Test nimmt folgende Form an :

**Resultat.** Die Hypothese  $H_0 : \mu = \mu_0$  getestet gegen  $H_1 : \mu \neq \mu_0$  ist zu verwerfen, falls

$$\begin{array}{l} \bar{x} < \mu_0 - t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \\ \text{oder} \\ \bar{x} > \mu_0 + t_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \end{array}$$

die Werte  $t_{1-\alpha/2}$  sind gegeben durch die Beziehung

$$P(t_{\alpha/2} \leq T \leq t_{1-\alpha/2}) = 1 - \alpha$$

wobei  $T$  einer Studentverteilung mit Freiheitsgrad  $n-1$  folgt. Sie sind im Anhang tabelliert. Das Risiko zweiter Art ist

$$\beta_\mu = P\left(\frac{\mu_0 - \mu}{S'/\sqrt{n}} - t_{1-\alpha/2} \leq T \leq \frac{\mu_0 - \mu}{S'/\sqrt{n}} + t_{1-\alpha/2}\right)$$

### 13.1.4 Test für Varianz einer normalverteilten Zufallsvariablen.

Sei  $X$  eine gaussverteilte Zufallsvariable<sup>2</sup>. Es soll die Nullhypothese

$$H_0 : \sigma^2 = \sigma_0^2$$

gegen die Alternative

$$H_1 : \sigma^2 \neq \sigma_0^2$$

<sup>2</sup>Es sei angemerkt, dass Tests für die Varianz viel sensibler sind bezüglich der Hypothese "normalverteilt" als Tests für den Mittelwert.

getestet werden. Es wird eine Stichprobe der Grösse  $n$  entnommen, und wir bilden den statistischen Test

$$\chi^2 = \frac{(n-1)S'^2}{\sigma_0^2},$$

wobei  $S'^2$  die erwartungstreue Varianz der Stichprobe ist,

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

$\chi$  folgt einer  $\text{Chi}^2$  Verteilung mit  $n-1$  Freiheitsgraden und der zweiseitige Test nimmt folgende Form an :

**Resultat.** Die Hypothese  $H_0 : \sigma^2 = \sigma_0^2$  getestet gegen  $H_1 : \sigma^2 \neq \sigma_0^2$  ist zu verwerfen falls

$$\begin{array}{l} \chi^2 < \chi_{\alpha/2, n-1}^2 \\ \text{oder} \\ \chi^2 > \chi_{1-\alpha/2, n-1}^2 \end{array}$$

die Werte  $\chi_{\alpha/2, n-1}^2$  und  $\chi_{1-\alpha/2, n-1}^2$  sind gegeben durch die Beziehung

$$P(\chi_{\alpha/2, n-1}^2 \leq \chi \leq \chi_{1-\alpha/2, n-1}^2) = 1 - \alpha$$

wobei  $\chi$  einer  $\text{Chi}^2$  verteilten Zufallsvariable mit Freiheitsgrad  $n-1$  folgt. Sie sind im Anhang tabelliert.

### 13.1.5 Test auf Anteilswert

Gegeben sei eine Population, deren Individuen eine uns interessierende Ausprägung entweder haben oder nicht haben. Wir wollen folgende Hypothese testen:

$H_0$  : der Anteil der Individuen die die Ausprägung besitzen ist  $p_0$ .

Wir begnügen uns mit der Näherungslösung dieses Testproblems basierend auf dem zentralen Grenzwertsatz. Dies ist zulässig, falls der Umfang  $n$  der Stichprobe genügend gross ist und der Punktschätzer der Proportion

$$\hat{p} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n},$$

nicht zu klein ist (typisch:  $n \geq 30$  und  $n\hat{p}(1-\hat{p}) \geq 10$ ).

Wir bilden den statistischen Test

$$Y = \begin{cases} \frac{(\bar{X}+0.5)-np_0}{\sqrt{np_0(1-p_0)}} & \text{falls } \bar{X} < np_0 \\ \frac{(\bar{X}-0.5)-np_0}{\sqrt{np_0(1-p_0)}} & \text{falls } \bar{X} > np_0 \end{cases}$$

$Y$  ist annähernd normalverteilt  $N(0, 1)$  und der zweiseitige Test nimmt folgende Form an :

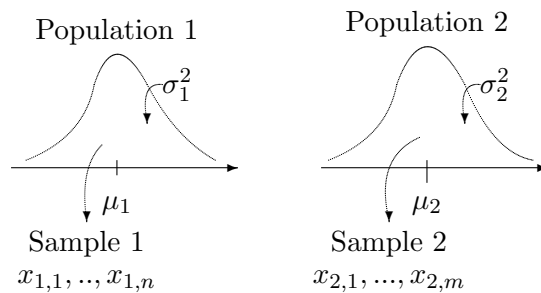
**Resultat.** Die Hypothese  $H_0 : p = p_0$  getestet gegen  $H_1 : p \neq p_0$  auf dem Niveau  $\alpha$  ist zu verwerfen falls gilt

$$|y| > u_{1-\alpha/2}.$$

Der Wert  $u_{1-\alpha/2}$  ist wie gewöhnlich gegeben durch die Beziehung  $P(|Y| \geq u_{1-\alpha/2}) = \alpha$ .

### 13.1.6 Vergleich der Mittel zweier Populationen

Beim Vergleich zweier Gruppen von Beobachtungen betrachtet man im einfachsten Fall ihre Mittelwerte  $\mu_1$  und  $\mu_2$  und untersucht sie auf signifikante Unterschiede (siehe Abbildung). Solche Fragestellungen treten zum Beispiel dann ein, wenn zwei Produkte miteinander verglichen werden sollen (Wirksamkeit zweier Medikamente, Haltbarkeit von Bio-Gemüse versus herkömmlichem Gemüse, mittlerer Benzinverbrauch zweier Fahrzeugklassen etc.). Wir unterscheiden die Tests mit bekannten und unbekanntem Varianzen  $\sigma_1$  und  $\sigma_2$ .



(1) **Test mit bekannter Varianz.** Wir gehen wieder zuerst davon aus, dass wir die zugehörigen theoretischen Varianzen  $\sigma_1$  und  $\sigma_2$  kennen. Zusätzliche Annahmen sind

- (i)  $x_{1,1}, x_{1,2}, \dots, x_{1,n}$  ist eine Stichprobe vom Umfang  $n$  der Population 1.
- (ii)  $x_{2,1}, x_{2,2}, \dots, x_{2,m}$  ist eine Stichprobe vom Umfang  $m$  der Population 2.
- (iii) Die beiden Populationen sind unabhängig.
- (iv) Beide Populationen sind entweder normalverteilt oder können (falls der zentrale Grenzwertsatz greift) als normalverteilt betrachtet werden.

Wir interessieren uns für die Grösse  $\mu_1 - \mu_2$ . Ein natürlicher Punktschätzer ist  $\bar{X}_1 - \bar{X}_2$ . Da

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

gilt, ist  $\bar{X}_1 - \bar{X}_2$  ein erwartungstreuer Schätzer für die Differenz  $\mu_1 - \mu_2$ . Für die Varianz gilt:

$$\text{Var}(\bar{X}_1 - \bar{X}_2) = \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

Dank den zusätzlichen Annahmen (iii) und (iv) ist  $\bar{X}_1 - \bar{X}_2$  normalverteilt. Es gilt also

Die Zufallsvariable

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

ist standardisiert und normalverteilt

Der Hypothesentest für die Differenz der Mittelwerte  $\Delta_0$  zweier Populationen bei bekannter Varianz nimmt folgende Form an:

**Resultat.** Die Hypothese  $H_0 : \mu_1 - \mu_2 = \Delta_0$  getestet gegen  $H_1 : \mu_1 - \mu_2 \neq \Delta_0$  auf dem Niveau  $\alpha$  ist zu verwerfen, falls gilt

$$|z| > u_{1-\alpha/2},$$

wobei  $z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$  der Wert der Teststatistik  $Z$  für eine bestimmte Stichprobe ist. Der Wert  $u_{1-\alpha/2}$  ist wie gewöhnlich gegeben durch die Beziehung  $P(|Z| \geq u_{1-\alpha/2}) = \alpha$ .

**Bemerkung.** Da die Teststatistik  $N(0, 1)$  verteilt ist, ist das Konfidenzintervall für  $\mu_1 - \mu_2$  zum Niveau  $\alpha$  gegeben durch

$$\bar{x}_1 - \bar{x}_2 - u_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + u_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

**Beispiel 60** Zwei identische Produktionsmaschinen liefern Bolzen mit normalverteiltem Durchmesser mit gemeinsamer Varianz  $\sigma_1^2 = \sigma_2^2 = 605 \mu m$ . Eine Messreihe von je 100 Bolzen liefert für die Bolzen der ersten Maschine den Mittelwert  $19.96 mm$  und für die Bolzen der zweiten Maschine den Mittelwert  $20.28 mm$ . Es soll zum Niveau  $\alpha = 0.01$  getestet werden, ob der Sollwert der beiden Produktionsmaschinen verschieden eingestellt ist.

**Lösung** Wir testen  $H_0: \mu_1 = \mu_2$  gegen  $H_1: \mu_1 \neq \mu_2$ . Die Teststatistik  $Z$  liefert unter  $H_0$  den Wert

$$z = \frac{|19.96 - 20.28|}{\sqrt{\frac{0.605}{100} + \frac{0.605}{100}}} = 32/11 \simeq 2.9$$

Da  $u_{1-0.005} = 2.575 < 2.9$  muss die Nullhypothese zum Niveau 0.01 verworfen werden. Die Sollwerte der beiden Maschinen sind signifikant verschieden.  $\diamond$

**Beispiel 61** Farbstoffe sollen getestet werden. Der 1. ist billiger, sodass der 2. nur interessant erscheint, wenn er signifikant besser dem Wetter standhält. Es werden 5 unabhängige Versuche durchgeführt, die zu folgenden Kennzahlen führen:

Farbstoff I	85	87	92	80	84
Farbstoff II	89	89	90	84	88

Dabei bedeutet eine höhere Zahl eine längere Standhaftigkeit. Wir wollen auf einem Niveau  $\alpha = 0.05$  testen, ob der zweite Farbstoff signifikant besser ist als der erste. Wir bezeichnen die Werte für den Farbstoff I mit  $X$  und für den anderen mit  $Y$  und nehmen an, die Kennzahlen seien normalverteilt mit Parameter  $(\mu_X, \sigma_X^2 = 1.5)$  und  $(\mu_Y, \sigma_Y^2 = 1)$ . Die Null-Hypothese  $H_0: \mu_X = \mu_Y$  soll gegen die Alternative  $H_1: \mu_X < \mu_Y$  getestet werden.

**Lösung** Der Wert der Teststatistik berechnet sich zu

$$z = \frac{85.6 - 87.6}{\sqrt{\frac{1.5}{5} + \frac{1}{5}}} = -2/\sqrt{0.5}$$

Da  $u_{1-0.05} = 1.645 < 2/\sqrt{0.5}$  ist, muss die Nullhypothese zum Niveau 0.05 verworfen werden. Der zweite Farbstoff ist signifikant besser als der erste.  $\diamond$

(2) **Test mit unbekannter Varianz.** In einem weiteren Schritt wird die Kenntnis der Varianzen nicht mehr verlangt. Wir unterscheiden die Fälle  $\sigma_1 = \sigma_2$  und  $\sigma_1 \neq \sigma_2$ . Im ersten Fall erhalten wir einen exakten Test. Im zweiten Fall ist der Test eine Näherung.

(a) Sei  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  die unbekannte (aber gleiche) Varianz der beiden normalverteilten Datensätze  $x_{1,1}, \dots, x_{1,n}$  und  $x_{2,1}, \dots, x_{2,m}$  mit Umfang  $n$  und  $m$ . Wir wissen, dass die Zufallsvariable

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$N(0, 1)$  verteilt ist. Darin ersetzen wir nun die unbekannte Varianz  $\sigma^2$  (resp.  $\sigma$ ) durch  $S_p^2$  (resp.  $S_p$ ), ein gewichtetes Mittel der empirischen Varianzen der beiden Datensätze ( $p$  steht für “pooled estimator”):

$$S_p^2 = \frac{n-1}{n+m-2} S_1'^2 + \frac{m-1}{n+m-2} S_2'^2$$

$S_p$  ist  $\chi^2$  verteilt mit  $n + m - 2$  Freiheitsgraden. Daraus folgt die uns interessierende Teststatistik. Es gilt

Die Zufallsvariable

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

ist t-verteilt mit  $n + m - 2$  Freiheitsgraden.

Der Hypothesentest für die Differenz der Mittelwerte  $\Delta_0$  zweier Populationen bei bekannter Varianz nimmt folgende Form an:

**Resultat.** Die Hypothese  $H_0 : \mu_1 - \mu_2 = \Delta_0$  getestet gegen  $H_1 : \mu_1 - \mu_2 \neq \Delta_0$  auf dem Niveau  $\alpha$  ist zu verwerfen falls gilt

$$|t| > t_{1-\alpha/2, n+m-2}$$

Wobei  $t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$  der Wert der Teststatistik  $T$  für eine bestimmte Stichprobe ist. Der Wert  $t_{1-\alpha/2, n+m-2}$  ist wie gewöhnlich gegeben durch die Beziehung  $P(|T| \geq t_{1-\alpha/2, n+m-2}) = \alpha$ .

(b) Wir gehen nun davon aus, dass die unbekannt Varianzen  $\sigma_1^2$  und  $\sigma_2^2$  verschieden sein können. Wir bezeichnen die empirischen Varianzen wie gewöhnlich mit  $S_1'^2$  und  $S_2'^2$  und stellen uns die Aufgabe, die Differenz im Mittel der beiden normalverteilten Datensätze  $x_{1,1}, \dots, x_{1,n}$  und  $x_{2,1}, \dots, x_{2,m}$  mit Umfang  $n$  und  $m$  zu testen. Die uns interessierende Teststatistik

$$T^* = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}}}$$

ist (nur noch) annähernd  $t$ -verteilt mit Freiheitsgrad

$$\nu = \frac{\left(\frac{s_1'^2}{n} + \frac{s_2'^2}{m}\right)^2}{\frac{(s_1'^2/n)^2}{n+1} + \frac{(s_2'^2/n)^2}{m+1}} - 2$$

Es gilt also

Die Zufallsvariable

$$T^* = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}}}$$

ist t-verteilt mit  $\nu$  Freiheitsgraden.

Der Hypothesentest für die Differenz der Mittelwerte  $\Delta_0$  zweier Populationen bei unbekannter



Varianz nimmt folgende Form an:

**Resultat.** Die Hypothese  $H_0 : \mu_1 - \mu_2 = \Delta_0$  getestet gegen  $H_1 : \mu_1 - \mu_2 \neq \Delta_0$  auf dem Niveau  $\alpha$  ist zu verwerfen, falls gilt

$$|t^*| > t_{1-\alpha/2, \nu},$$

wobei  $t^*$  der Wert der Teststatistik  $T^*$  für eine bestimmte Stichprobe ist. Der Wert  $t_{1-\alpha/2, \nu}$  ist wie gewöhnlich gegeben durch die Beziehung  $P(|T| \geq t_{1-\alpha/2, \nu}) = \alpha$ .

## 13.2 Nichtparametrische Tests

**Einführung.** Ein nichtparametrischer Test soll nicht nur einen Parameter einer gegebenen Verteilung validieren, sondern die Verteilung selber als Testgrösse beinhalten. In diesem Abschnitt müssen wir uns aus Platzgründen auf einige wenige graphische Methoden begrenzen, die das Grundproblem zwar eingrenzen, ohne jedoch das Fehlerrisiko zu quantifizieren.

Die Literatur zum Thema ist breit, und wir werden uns auf die einfachsten Tests beschränken. Es sei auch angemerkt, dass das Informatikprogramm “R”, gratis auf Internet verfügbar, viele Tests gebrauchsfertig und gebrauchsfreundlich zur Verfügung stellt.

### 13.2.1 Schätzung einer Verteilungsfunktion

**Direkte Methode.** Gegeben sei eine Zufallsvariable  $X$ , von der wir die Verteilungsfunktion  $F_X(x)$  mit Hilfe einer Stichprobe mit Umfang  $n$  abschätzen wollen. Seien  $x_1, x_2, \dots, x_n$  die Werte der Stichprobe, die wir nun der Grösse nach ordnen:  $\rightsquigarrow x_{(1)}, \dots, x_{(n)}$ . Es gilt entsprechend:  $x_{(1)}$  ist der kleinste Wert der  $x_i$ ,  $x_{(2)}$  der zweitkleinste Wert und so weiter...

Unter der Annahme, dass der Stichprobenumfang und der Umfang der Grundgesamtheit genügend gross sind, kann angenommen werden, dass alle Werte der Stichprobe die gleiche Wahrscheinlichkeit haben, kleiner als ein gewisser Wert  $x$  zu sein, nämlich  $p_x := F_X(x)$ . Die **direkte Methode** besteht darin anzunehmen, dass die Anzahl  $R$  der Werte in der Stichprobe, die kleiner als ein fixiertes  $x$  sind, einer Binomialverteilung mit Parameter  $(n, p_x)$  folgt.

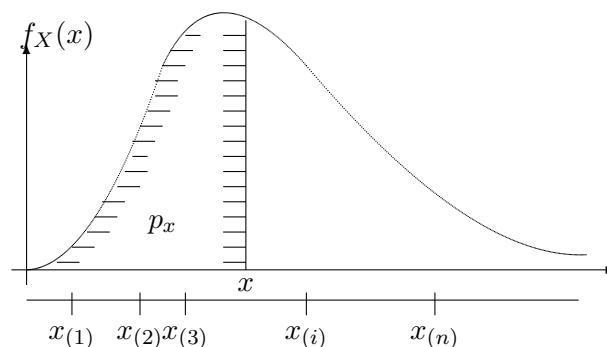


Abbildung 13.2: drei von  $n$  Ziehungen sind kleiner als  $x$ . Das gibt eine erste Idee für  $p_x$ .

Es geht nun darum, den Parameterwert  $p_x$  einer binomialverteilten Zufallsvariablen abzuschätzen. Die direkte Methode schlägt vor,  $p_x$  durch die relative Anzahl der  $x_i$  abzuschätzen mit  $x_i \leq x$  (siehe 65) d.h.,  $p_x = \frac{i}{n}$ .

**Abschätzung von  $F_X(x)$ .** Sei  $x_{(1)}, \dots, x_{(n)}$  die der Grösse nach geordnete Stichprobe. Für  $x_{(i)} \leq$

$x \leq x_{(i+1)}$  ist der Schätzer von  $F_X(x)$  gegeben durch

$$\hat{F}_X(x) = \frac{i}{n}$$

Die Abbildung 13.3 zeigt ein Beispiel eines solchen Schätzers für die Verteilungsfunktion.

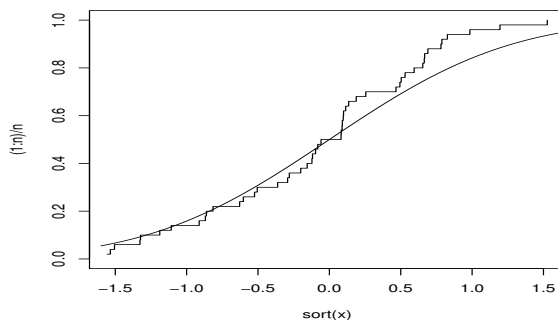


Abbildung 13.3: Schätzung einer Verteilungsfunktion mit 50 Ziehungen einer normalverteilten Zufallsvariablen. Die glatte Kurve ist die richtige Verteilungsfunktion.

Dieser Schätzer, wenn auch erwartungstreu, hat den Nachteil für kleine Stichproben sehr ungenau zu sein, da  $\hat{F}_X(x_n) = 1$  ist, obwohl generell  $F_X(x_n) < 1$  gilt. Es ist angebracht eine Korrektur anzubringen.

**Variante der direkten Methode.** Die Idee bleibt die gleiche wie vorher. Der Schätzer wird aber wie folgt definiert:

$$\hat{F}_X(x_{(i)}) = \frac{i}{n+1}$$

Mehrere Statistikprogramme wie z.B. "R" gebrauchen auch  $\hat{F}_X(x_{(i)}) = \frac{i-0.5}{n}$

### 13.2.2 Annäherung an eine Verteilung via p-p-Plot

Es geht bei diesem Verfahren darum, auf ein und derselben Skala eine theoretische Verteilungsfunktion  $F_{th}$  mit der realen Verteilungsfunktion  $F_X$  zu vergleichen. Die letztere ist nicht bekannt und muss entsprechend für jeden Wert der Stichprobe  $x_i$  abgeschätzt werden. Wie oben werden die Werte in aufsteigender Folge geordnet  $x_{(1)}, \dots, x_{(n)}$ , und wir nehmen für den Schätzer die Funktion  $\hat{F}_X(x_{(i)}) = \frac{i}{n+1}$ .

Falls die reale Verteilung identisch mit der zu testenden theoretischen Verteilung ist (i.e. falls  $H_0 : F_X = F_{th}$  wahr ist), liegen die Punkte  $M_i$  mit Koordinaten

$$M_i(\hat{F}_X(x_{(i)}), F_{th}(x_{(i)}))$$

auf der Hauptdiagonalen des ersten Quadranten.

Die Punkte  $M_i$  bilden eine Punktwolke, deren Form und Disposition bezüglich der Hauptdiagonalen ein Urteil über eine mögliche Akzeptanz oder einen möglichen Verwurf der Hypothese  $H_0$  (gegen  $H_1 : F_X \neq F_{th}$ ) liefert. Diese Punktwolke wird als  $p-p$ -Plot angesprochen.

**Beispiel 62** Sei eine Stichprobe von Umfang  $n = 20$  gegeben. Die  $n$  Werte  $x_i$  der Zufallsvariablen  $X$  seien der Größe nach geordnet.

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	0.08	0.13	0.20	0.26	0.45	0.50	0.53	0.66	0.75	1.00

$i$	11	12	13	14	15	16	17	18	19	20
$x_i$	1.16	1.20	1.33	1.57	1.90	2.10	2.39	2.78	3.39	4.30

Wir testen die Annäherung der Verteilung von  $X$  an eine Exponentialverteilung einmal mit Parameter  $\lambda = 1$  und einmal mit  $\lambda = 0.7$ . Die Punkte sind durch die Koordinaten  $(\frac{i}{n+1}, 1 - e^{-\lambda x})$  gegeben.

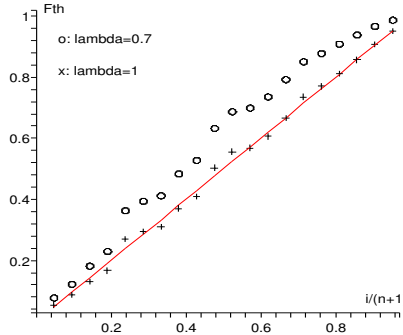


Abbildung 13.4: zwei überlagerte p-p-Plots für  $H_0 = \epsilon_{0.7}$  und  $H_0 = \epsilon_1$ .

Das Diagramm suggeriert eine Akzeptanz von  $F_{th}(x) = 1 - e^{-0.7x}$  und der Verwurf von  $F_{th}(x) = 1 - e^{-x}$ .  $\diamond$

### 13.2.3 Annäherung an eine Verteilung via q-q-Plot

Man vergleicht, unter den gleichen Bedingungen wie für den  $p - p$ -Plot, die Quantile  $x_i$  der geschätzten realen Verteilungsfunktion  $F_n(x_i)$  mit denen der theoretisch vorgeschlagenen Verteilung, nämlich  $a_n(x_i) = F_{th}^{-1}(F_n(x_i))$ .

Falls die reale Verteilung von  $X$  mit der zu testenden theoretischen Verteilung übereinstimmt, liegen die Punkte

$$M_i := (x_i, a_n(x_i)) = (x_i, F_{th}^{-1}(F_n(x_i)))$$

auf der Hauptdiagonalen des ersten Quadranten.

Die Punkte  $M_i$  bilden eine Punktwolke, deren Form und Disposition bezüglich der Hauptdiagonalen ein Urteil über eine mögliche Akzeptanz oder einen möglichen Verwurf der Hypothese  $H_0$  (gegen  $H_1 : F_X \neq F_{th}$ ) liefert. Diese Punktwolke wird als  $q - q$ -Plot angesprochen.

**Beispiel 63** Für dieselbe Stichprobe wie oben wird mit Hilfe eines  $q - q$  Plots eine Annäherung an die Verteilung von  $X$  durch die Exponentialverteilungen mit Parameter  $\lambda = 1$  und  $\lambda = 0.7$  beurteilt. Die gezeichneten Punkte haben die Koordinaten

$$\left(x_i, -\frac{1}{\lambda} \left(1 - \frac{i}{n+1}\right)\right)$$

Das  $q - q$  Diagramm ist:

Es wird bemerkt dass für  $\lambda = 0.7$  die Punktwolke, wie es sein muss, entlang der Hauptdiagonalen verläuft. Die gleiche Entscheidung wie oben drängt sich auf (d.h. Wahl von  $\lambda = 0.7$ ).  $\diamond$

**Bemerkung.** Wird mit dem Logiciel 'R' eine Exponentialverteilung mit geschätztem Parameter ( $\hat{\lambda} = \frac{1}{\bar{x}}$ ) getestet, genügt es zum Beispiel folgendes einzugeben

```
qqplot(qexp(ppoints(n), lambda), sort(x)),
```

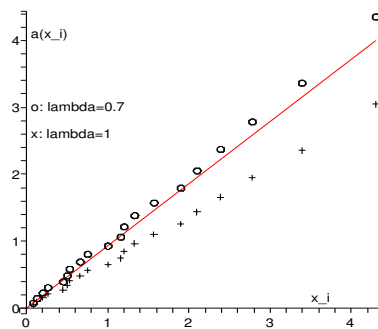


Abbildung 13.5: Zwei überlagerte  $q$ - $q$ -Plots für  $H_0 = \epsilon_{0.7}$  und  $H_0 = \epsilon_1$ .

`qexp` berechnet die Quantile, `sort` sortiert die Daten der Größe nach, `ppoints` berechnet die empirische Verteilungsfunktion basierend auf der Stichprobe mit  $n = 20$ . Wir erhalten direkt den  $q - q$ -Plot für die Exponentialverteilung  $\epsilon_\lambda$ .

2) Der Befehl `qqnorm(x)` von 'R' zeichnet direkt den  $q - q$ -Plot, um den Datenvektor  $x$  mit einer Normalverteilung anzunähern. Die Funktion `qqline(x)` nähert in einem  $q - q$ -Plot die Daten mit einer affinen Abbildung an.

### 13.3 Aufgaben

**Übung 66** Für einen Kunden müssen Schrauben mit einem (mittleren) Durchmesser von  $3.2\text{mm}$  hergestellt werden. Bei einer Kontrollmessung an 10 Schrauben wurde aber ein durchschnittlicher Durchmesser von  $\bar{x} = 3.1\text{mm}$  gemessen. Laut Herstellerangabe arbeitet die Produktionsmaschine mit einer Standardabweichung von  $\sigma = 0.15\text{mm}$ . Hat der Werkmeister die Maschine falsch eingestellt? Testen Sie mit einem Signifikanzniveau von  $\alpha = 5\%$ .  $\odot$

**Übung 67** Die Reaktionszeit eines Computers ist eine wichtige charakteristische Qualitätsgröße. Wir möchten wissen ob die mittlere Antwortzeit für einen spezifischen Befehl  $75\text{ms}$  überschreitet oder nicht. Vorangegangene Untersuchungen zeigten, dass die Standardabweichung der Reaktionszeit bei  $8\text{ms}$  liegt. Testen Sie die Hypothese  $H_0 : \mu = 75$  gegen  $H_1 : \mu > 75$  auf einem Signifikanzniveau  $\alpha = 0.05$  wissend, dass der Mittelwert  $\bar{x}$  der Stichprobe mit Umfang  $n = 25$  gleich  $79.25\text{ms}$  ist.  $\odot$

**Übung 68** Eine Packung Cornflakes enthält laut Verpackung durchschnittlich  $368\text{g}$ . Bei 25 zufällig ausgewählten Packungen werden eine Durchschnittsmenge von  $\bar{x} = 372.5\text{g}$  und eine Standardabweichung von  $s' = 12\text{g}$  festgestellt. Der Hersteller will überprüfen, ob korrekt abgefüllt wird. Testen Sie mit einem Signifikanzniveau von  $\alpha = 5\%$ .  $\odot$

**Übung 69** Sie sind Gemüsegroßekäufer einer Lebensmittelkette. Sie bekommen eine größere Lieferung an verkaufsfertig abgepackten Äpfeln. Der Lieferant behauptet, die durchschnittliche Füllmenge pro Verkaufseinheit beträgt  $3\text{kg}$ . Bei einer Überprüfung an 20 zufällig ausgewählten Einheiten erhalten Sie aber nur eine Durchschnittsmenge von  $\bar{x} = 2.88\text{kg}$  bei einer Standardabweichung von  $s' = 0.3\text{kg}$ .

Hat der Lieferant recht? (Signifikanzniveau  $\alpha = 1\%$ ).  $\odot$

**Übung 70** Im Kontext von Aufgabe 45 soll  $H_0 : \sigma^2 = 10^4$  auf einem Signifikanzniveau  $\alpha = 0.05$  gegen  $H_1 : \sigma^2 \neq 10^4$  getestet werden.  $\odot$

**Übung 71** Formulieren Sie den einseitigen Test  $H_0 : \sigma^2 = \sigma_0^2$  gegen  $\sigma^2 > \sigma_0^2$ .  $\odot$

**Übung 72** Ein berühmter Versuch von Pratt und Woodruff (1938) ging der Frage von übersinnlichen Fähigkeiten durch folgendes Experiment nach:

Ein Versuchsleiter zieht eine von fünf verschiedenen Karten und eine Versuchsperson versucht die Karte zu erraten. Ergebnis: von 60000 Versuchspersonen haben 12489 korrekt geantwortet. Testen Sie auf einem Niveau  $\alpha = 0.05$  die Nullhypothese  $H_0 : p = \frac{1}{5}$  gegen  $H_1 : p \neq \frac{1}{5}$ .  $\odot$

**Übung 73** Formulieren Sie den einseitigen Test  $H_0 : p = p_0$  gegen  $p \geq p_0$  und wenden Sie ihn auf obige Frage an (es sei bemerkt, dass sie dadurch die Möglichkeit von übersinnlichen Fähigkeiten gewisser Versuchspersonen testen).  $\odot$

**Übung 74** Formulieren Sie den einseitigen Test  $H_0 : \mu_1 - \mu_2 = \Delta_0$  gegen  $H_1 : \mu_1 - \mu_2 < \Delta_0$ .  $\odot$

**Übung 75** Sie versuchen die Härtezeit eines Klebers zu reduzieren. Dazu vergleichen Sie die Aushärtezeit eines Standardklebers mit der Aushärtezeit Ihres neuen Klebers der zusätzlich zu der Komposition des Standardklebers einen Aushärter enthält. Die Erfahrung zeigt, dass die Standardabweichung der Aushärtezeit 8 Minuten beträgt und dass sie durch den zusätzlichen Aushärter nicht tangiert wird. Jeweils 10 Klebeverbindungen werden mit dem herkömmlichen und dem neuen Kleber getestet. Die mittleren Aushärtezeiten sind  $\bar{x}_1 = 121min$  für den Standardkleber und  $\bar{x}_2 = 112min$  für den neuen Kleber. Testen Sie auf einem Niveau  $\alpha = 0.05$ , ob der zweite Klebstoff signifikant kürzere Aushärtezeiten hat als der herkömmliche.  $\odot$

**Übung 76** Formulieren Sie den einseitigen Test  $H_0 : \mu_1 - \mu_2 = \Delta_0$  gegen  $H_1 : \mu_1 - \mu_2 > \Delta_0$ .  $\odot$

**Übung 77** Formulieren Sie den einseitigen Test  $H_0 : \mu_1 - \mu_2 = \Delta_0$  gegen  $H_1 : \mu_1 - \mu_2 > \Delta_0$ .  $\odot$

**Übung 78** Lösen Sie Beispiel 60, falls die gemeinsame Varianz nicht bekannt, jedoch empirisch geschätzt wird mit  $s_1'^2 = 600\mu m$  und  $s_2'^2 = 650\mu m$ .  $\odot$

**Übung 79** Lösen Sie Übung 75, falls die gemeinsame Varianz mit jeweils 10 Stichproben empirisch geschätzt wird mit  $s_1'^2 = 10min$  und  $s_2'^2 = 9min$ .  $\odot$

**Übung 80** Formulieren Sie den einseitigen Test  $H_0 : \mu_1 - \mu_2 = \Delta_0$  gegen  $H_1 : \mu_1 - \mu_2 > \Delta_0$ .  $\odot$

**Übung 81** Lösen Sie Beispiel 61, falls die (ungleichen) Varianzen empirisch geschätzt werden.  $\odot$

# Kapitel 14

## Solutions III

57)

$$E(S^2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n} \sum_{i=1}^n E\left((X_i - \bar{X})^2\right)$$

On a par définition  $E(X_i) = \mu$  et  $E(X_i^2) - \mu^2 = \sigma^2$ . En développant:

$$E\left((X_i - \bar{X})^2\right) = E(X_i^2) - 2E(X_i\bar{X}) + E(\bar{X}^2).$$

Der erste Term der Summe ist gleich  $\mu^2 + \sigma^2$  der zweite Term

$$\frac{2}{n}E(X_i^2 + X_i \sum_{j \neq i} X_j) = \frac{2}{n}\left(\mu^2 + \sigma^2 + \mu \sum_{j \neq i} \mu\right) = \frac{2}{n}\left(\mu^2 + \sigma^2 + (n-1)\mu^2\right) = 2\frac{\sigma^2}{n} + 2\mu^2$$

und der letzte Term schreibt sich als

$$\frac{1}{n^2}E\left(\sum_{i=1}^n X_i^2 + 2 \sum_{i < j} X_i X_j\right) = \frac{1}{n^2}\left(n(\sigma^2 + \mu^2) + 2\frac{n(n-1)}{2}\mu^2\right) = \frac{\sigma^2}{n} + \mu^2.$$

Au total on trouve:

$$E(S^2) = \frac{1}{n} \sum_{i=1}^n \left(\mu^2 + \sigma^2 - \frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n}\sigma^2.$$

58) Seien  $X_i \sim \text{Bernoulli}(p)$  verteilt (mit  $0 < p < 1$ ). Relativer Anteil bei  $n$  Ziehungen mit Zurücklegen  $Y := \frac{1}{n} \sum_{i=1}^n X_i \sim \text{Binom}(n, p)$ . Die Standardabweichung von  $Y$  muss nun kleiner als  $0.01p$  sein. Da  $\text{Var}(Y) = \frac{p(1-p)}{n}$  gilt, folgt:

$$\sqrt{\frac{p(1-p)}{n}} < 0.01p \Leftrightarrow n > 10^4 \frac{1-p}{p}.$$

59)

$$102 - 5,39 < \mu < 102 + 5,39$$

60)

$$102 - 5,927 < \mu < 102 + 5,927$$

61) Der Punktschätzer für Anteilswert  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  ist für grosse  $n$  annähernd normalverteilt mit Parameter  $(p, \frac{p(p-1)}{n})$ . Es gilt entsprechend

$$P(u_{\alpha/2} \leq \frac{\bar{X} - p}{\sqrt{\frac{p(p-1)}{n}}} \leq u_{1-\alpha/2}) = 1 - \alpha$$

und wir finden das zweiseitige Konfidenzintervall zum Niveau  $\alpha$

$$\boxed{\bar{x} - u_{1-\alpha/2} \sqrt{\frac{p(p-1)}{n}} \leq p \leq \bar{x} + u_{1-\alpha/2} \sqrt{\frac{p(p-1)}{n}}}$$

$\bar{x}$  ist dabei das beobachtete Mittel der Stichprobe.

**62)**

$$3210.7 - 65.12 < \mu < 3210.7 + 65.12$$

$$7414.3 < \sigma^2 < 34405$$

## Annexe

### Gammafunktion.

Die Tabelle gibt einige Werte der Funktion

$$\Gamma(n/2) = \int_0^{\infty} x^{n/2-1} e^{-x} dx$$

und der Hilfsfunktion

$$c_4(n) = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$$

in Abhängigkeit vom Stichprobenumfang  $n$  an.

$n$	2	3	4	5	10	20	30
$\Gamma(n/2)$	1	.88625	1	1.3294	24	$3.628810^5$	$8.717810^{10}$
$c_4(n)$	0.79785	0.88625	0.92127	0.93999	0.97260	0.98690	0.99137

$n$	40	50	60	70	80
$\Gamma(n/2)$	$1.216510^{17}$	$6.204510^{23}$	$8.841810^{30}$	$2.952310^{38}$	$2.039810^{46}$
$c_4(n)$	0.99358	0.99484	0.99577	0.99635	0.99680

### Normalverteilung.

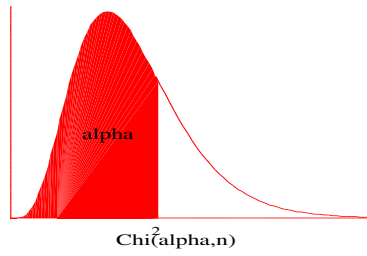
Die Verteilungsfunktion  $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$  einer standardisierten Zufallsvariable  $X \sim N(0; 1)$  ist ein nichtelementares Integral. Man hilft sich mit Tabellen wie der untenstehenden weiter um Näherungswerte zu bestimmen. Zehntelwerte für  $x$  werden in den Linien abgelesen und Hundertstel in den Kolonnen.

Zum Beispiel befindet sich der Wert von  $\phi(1.65)$  an der Schnittstelle der Linie 1.6 und der Kolonne 0.05. Der Wert von  $\phi(1.65)$  ist entsprechend  $\approx 0.9505$ . Für negative  $x$  Werte benutzt man die Beziehung  $\phi(-x) = 1 - \phi(x)$ .



$\phi(x)$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9986	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

$\chi^2$  **Verteilung.** Die  $\chi^2$  Verteilung mit  $\int_0^y f_X(x)dx = \int_0^y \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} dx$  mit Freiheitsgrad  $n$  ist ein nichtelementares Integral. Wir definieren  $\chi_{\alpha,n}^2$  durch  $\int_0^{\chi_{\alpha,n}^2} f_X(x)dx = \alpha$  und geben unten einige Werte für  $\chi_{\alpha,n}^2$ .



	$\alpha = 0.001$	0.005	0.01	0.025	0.05	0.1	0.3	0.5
$n = 1$	$1.57 \cdot 10^{-5}$	$3.93 \cdot 10^{-4}$	$1.57 \cdot 10^{-3}$	$0.0982 \cdot 10^{-3}$	0.0039	0.0158	0.1485	0.4549
2	0.0020	0.0100	0.0201	0.0506	0.1026	0.2107	0.7133	1.3863
3	0.0243	0.0717	0.1148	0.2158	0.3518	0.5844	1.4237	2.3660
4	0.0908	0.2070	0.2971	0.4844	0.7107	1.0636	2.1947	3.3567
5	0.2102	0.4118	0.5543	0.8312	1.1455	1.6103	2.9999	4.3515
6	0.3810	0.6757	0.8721	1.2373	1.6354	2.2041	3.8276	5.3481
7	0.5985	0.9893	1.2390	1.6899	2.1673	2.8331	4.6713	6.3458
8	0.8571	1.3444	1.6465	2.1797	2.7326	3.4895	5.5274	7.3441
9	1.1519	1.7349	2.0879	2.7004	3.3251	4.1682	6.3933	8.3428
10	1.4787	2.1558	2.5582	3.2470	3.9403	4.8652	7.2672	9.3418
11	1.8338	2.6032	3.0535	3.8157	4.5748	5.5778	8.1479	10.3410
12	2.2141	3.0738	3.5706	4.4038	5.2260	6.3038	9.0343	11.3403
13	2.6172	3.5650	4.1069	5.0087	5.8919	7.0415	9.9257	12.3398
14	3.0407	4.0747	4.6604	5.6287	6.5706	7.7895	10.8215	13.3393
15	3.4825	4.6009	5.2294	6.2621	7.2609	8.5468	11.7212	14.3389
16	3.9417	5.1422	5.8122	6.9077	7.9616	9.3122	12.6243	15.3385
17	4.4162	5.6973	6.4077	7.5642	8.6718	10.0852	13.5307	16.3382
18	4.9048	6.2648	7.0149	8.2307	9.3904	10.8649	14.4399	17.3379
19	5.4067	6.8439	7.6327	8.9065	10.1170	11.6509	15.3517	18.3376
20	5.9210	7.4338	8.2604	9.5908	10.8508	12.4426	16.2659	19.3374
21	6.4467	8.0336	8.8972	10.2829	11.5913	13.2396	17.1823	20.3372
22	6.9829	8.6427	9.5425	10.9823	12.3380	14.0415	18.1007	21.3370
23	7.5291	9.2604	10.1957	11.6885	13.0905	14.8480	19.0211	22.3369
24	8.0847	9.8862	10.8563	12.4011	13.8484	15.6587	19.9432	23.3367
25	8.6494	10.5196	11.5240	13.1197	14.6114	16.4734	20.8670	24.3366
26	9.2222	11.1602	12.1982	13.8439	15.3792	17.2919	21.7924	25.3365
27	9.8029	11.8077	12.8785	14.5734	16.1514	18.1139	22.7192	26.3363
28	10.3907	12.4613	13.5647	15.3079	16.9279	18.9392	23.6475	27.3362
29	10.9861	13.1211	14.2564	16.0471	17.7084	19.7677	24.5770	28.3361
30	11.5876	13.7867	14.9535	16.7908	18.4927	20.5992	25.5078	29.3360
40	17.9166	20.7066	22.1642	24.4331	26.5093	29.0505	34.8719	39.3353
50	24.6736	27.9908	29.7067	32.3574	34.7642	37.6886	44.3133	49.3349
60	31.7381	35.5344	37.4848	40.4817	43.1880	46.4589	53.8091	59.3347
70	39.0358	43.2753	45.4417	48.7575	51.7393	55.3289	63.3460	69.3345
80	46.5197	51.1719	53.5400	57.1532	60.3915	64.2778	72.9153	79.3343
90	54.1559	59.1963	61.7540	65.6466	69.1260	73.2911	82.5111	89.3342
100	61.9182	67.3275	70.0650	74.2219	77.9294	82.3581	92.1290	99.3341

Fortsetzung der Tabelle für die  $\chi^2$  Verteilung.

	$\alpha = 0.7$	0.9	0.95	0.975	0.99	0.995	0.99.9
$n = 1$	1.0742	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274
2	2.4079	4.6052	5.9915	7.3778	9.2104	10.5965	13.8150
3	3.6649	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660
4	4.8784	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662
5	6.0644	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147
6	7.2311	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575
7	8.3834	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213
8	9.5245	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239
9	10.6564	14.6837	16.9190	19.0228	21.6660	23.5893	27.8767
10	11.7807	15.9872	18.3070	20.4832	23.2093	25.1881	29.5879
11	12.8987	17.2750	19.6752	21.9200	24.7250	26.7569	31.2635
12	14.0111	18.5493	21.0261	23.3367	26.2170	28.2997	32.9092
13	15.1187	19.8119	22.3620	24.7356	27.6882	29.8193	34.5274
14	16.2221	21.0641	23.6848	26.1189	29.1412	31.3194	36.1239
15	17.3217	22.3071	24.9958	27.4884	30.5780	32.8015	37.6978
16	18.4179	23.5418	26.2962	28.8453	31.9999	34.2671	39.2518
17	19.5110	24.7690	27.5871	30.1910	33.4087	35.7184	40.7911
18	20.6014	25.9894	28.8693	31.5264	34.8052	37.1564	42.3119
19	21.6891	27.2036	30.1435	32.8523	36.1908	38.5821	43.8194
20	22.7745	28.4120	31.4104	34.1696	37.5663	39.9969	45.3142
21	23.8578	29.6151	32.6706	35.4789	38.9322	41.4009	46.7963
22	24.9390	30.8133	33.9245	36.7807	40.2894	42.7957	48.2676
23	26.0184	32.0069	35.1725	38.0756	41.6383	44.1814	49.7276
24	27.0960	33.1962	36.4150	39.3641	42.9798	45.5584	51.1790
25	28.1719	34.3816	37.6525	40.6465	44.3140	46.9280	52.6187
26	29.2463	35.5632	38.8851	41.9231	45.6416	48.2898	54.0511
27	30.3193	36.7412	40.1133	43.1945	46.9628	49.6450	55.4751
28	31.3909	37.9159	41.3372	44.4608	48.2782	50.9936	56.8918
29	32.4612	39.0875	42.5569	45.7223	49.5878	52.3355	58.3006
30	33.5302	40.2560	43.7730	46.9792	50.8922	53.6719	59.7022
40	44.1649	51.8050	55.7585	59.3417	63.6908	66.7660	73.4029
50	54.7228	63.1671	67.5048	71.4202	76.1538	79.4898	86.6603
60	65.2265	74.3970	79.0820	83.2977	88.3794	91.9518	99.6078
70	75.6893	85.5270	90.5313	95.0231	100.4251	104.2148	112.3167
80	86.1197	96.5782	101.8795	106.6285	112.3288	116.3209	124.8389
90	96.5238	107.5650	113.1452	118.1359	124.1162	128.2987	137.2082
100	106.9058	118.4980	124.3421	129.5613	135.8069	140.1697	149.4488

Ablesebeispiel  $\chi_{0.975,16}^2 = 28.8453$ .

**Studentverteilung.** Auch die standardisierte Studentverteilung

$$\phi_{T_\nu}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu\pi}} \int^x dy \left(1 + \frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}$$

mit Freiheitsgrad  $\nu$  ist ein nichtelementares Integral. Im Unterschied zur Tabelle für die Normal-

verteilung wird hier die Umkehrfunktion angegeben.

$\phi_{T_\nu}(t_{1-\alpha/2})$	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
$\nu = 1$	0.32492	1.00000	3.07768	6.31375	12.7062	31.8205	63.6567	636.6192
2	0.28867	0.81649	1.88561	2.91998	4.3026	6.9645	9.9248	31.599
3	0.27667	0.76489	1.63774	2.35336	3.18245	4.5407	5.8409	12.924
4	0.27072	0.74069	1.53320	2.13184	2.7764	3.7469	4.6040	8.610
5	0.26718	0.72668	1.47588	2.01504	2.5705	3.3649	4.0321	6.8688
6	0.26483	0.71755	1.43975	1.94318	2.4469	3.1426	3.7074	5.958
7	0.26316	0.71114	1.41492	1.89457	2.3646	2.9979	3.4994	5.407
8	0.26192	0.70638	1.39681	1.85954	2.3060	2.8964	3.3553	5.041
9	0.26095	0.70272	1.38302	1.83311	2.2621	2.8214	3.2498	4.780
10	0.26018	0.69981	1.37218	1.81246	2.2281	2.7637	3.1692	4.586
11	0.25955	0.69744	1.36343	1.79588	2.2009	2.7180	3.1058	4.437
12	0.25903	0.69548	1.35621	1.78228	2.1788	2.68100	3.0545	4.317
13	0.25859	0.69382	1.35017	1.77093	2.1603	2.6503	3.0122	4.220
14	0.25821	0.69241	1.34503	1.76131	2.1447	2.6244	2.9768	4.140
15	0.25788	0.69119	1.34060	1.75305	2.1314	2.6024	2.9467	4.072
16	0.25759	0.69013	1.33675	1.74588	2.1199	2.5834	2.9207	4.015
17	0.25734	0.68919	1.33337	1.73960	2.1098	2.5669	2.8982	3.965
18	0.25712	0.68836	1.33039	1.73406	2.1009	2.5523	2.8784	3.921
19	0.25692	0.68762	1.32772	1.72913	2.0930	2.5394	2.8609	3.883
20	0.25674	0.68695	1.32534	1.72471	2.0859	2.5279	2.8453	3.849
21	0.25658	0.68635	1.32318	1.72074	2.0796	2.5176	2.8313	3.819
22	0.25643	0.68580	1.32123	1.71714	2.0738	2.5083	2.8187	3.792
23	0.25629	0.68530	1.31946	1.71387	2.0686	2.4998	2.8073	3.767
24	0.25617	0.68485	1.31783	1.71088	2.0639	2.4921	2.7969	3.745
25	0.25606	0.68443	1.31634	1.70814	2.0595	2.4851	2.7874	3.725
26	0.25595	0.68404	1.31497	1.70561	2.0555	2.4786	2.7787	3.706
27	0.25585	0.68368	1.31370	1.70328	2.0518	2.4726	2.7706	3.689
28	0.25576	0.68335	1.31252	1.70113	2.0484	2.4671	2.7632	3.673
29	0.25568	0.68304	1.31143	1.69912	2.0452	2.4620	2.7563	3.659
30	0.25560	0.68275	1.31041	1.69726	2.0422	2.4572	2.7500	3.646
$\infty$	0.25334	0.67449	1.28155	1.64485	1.9599	2.3263	2.5758	3.290

Ist zum Beispiel das Signifikanzniveau  $\alpha$  bei 10%, dann ist  $\phi_{T_\nu}(t_{1-\alpha/2}) = 1 - \alpha/2 = 0.95$ . Entsprechend ist  $t_{1-\alpha/2} = \phi_{T_\nu}^{-1}(0.95)$ . Für einen Freiheitsgrad  $\nu = 5$  finden wir mit Hilfe der Tabelle (Linie mit Freiheitsgrad 5 und Kolonne 0.95) den Wert  $t_{1-\alpha/2} \simeq 2.015048$ .